University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Hautphenne, S;Massaro, M;Turner, K

Title:
Fitting Markovian binary trees using global and individual demographic data

Date:
2019-08-01

Citation:
Hautphenne, S., Massaro, M. & Turner, K. (2019). Fitting Markovian binary trees using global and individual demographic data. Theoretical Population Biology, 128, pp.39-50. https://doi.org/10.1016/j.tpb.2019.04.007.

Persistent Link:
https://hdl.handle.net/11343/249512

Fitting Markovian binary trees using global and individual demographic data

Sophie Hautphenne[1], Melanie Massaro[2], and Katharine Turner [3]

Abstract

We consider a class of continuous-time branching processes called *Markovian binary trees* (MBTs), in which the individuals lifetime and reproduction epochs are modeled using a *transient Markovian arrival process* (TMAP). We develop methods for estimating the parameters of the TMAP by using either age-specific averages of reproduction and mortality rates, or age-specific individual demographic data. Depending on the degree of detail of the available information, we follow a weighted non-linear regression or a maximum likelihood approach. We discuss several criteria to determine the optimal number of states in the underlying TMAP. Our results improve the fit of an existing MBT model for human demography, and provide insights for the future conservation management of the threatened Chatham Island black robin population.

Keywords: Markov process, branching process, non-linear regression, maximum likelihood, *Petroica traversi*.

1. Introduction

Classical continuous-time *linear birth-and-death processes* assume constant birth and death rates over the lifetime of individuals. While mathematically convenient, this assumption rarely holds in practice. In particular, the memoryless property inherent to these processes implies that individuals

[1]The University of Melbourne, School of Mathematics and Statistics, 3010 Melbourne, Victoria, Australia, sophiemh@unimelb.edu.au; and Ecole polythechnique fédérale de Lausanne, sophie.hautphenne@epfl.ch. Corresponding author.

[2]Charles Sturt University, Institute for Land, Water and Society and School of Environmental Sciences, mmassaro@csu.edu.au

[3]Australian National University, katharine.turner@anu.edu.au; and Ecole polythechnique fédérale de Lausanne, katharine.turner@epfl.ch

do not age, and therefore they do not offer enough flexibility to model real biological populations in which the age of individuals impacts on their fertility and mortality rates. Yet, linear birth-and-death processes have been frequently used due to their tractability and amenability to efficient parameter estimation, see for instance [16, 28]. In contrast, *age-dependent* (or *Bellman-Harris*) *branching processes* assume general lifetime distributions, but they are much less tractable. They have been analysed from a statistical point of view under some particular assumptions: for example in [15], the full tajectory of the population is supposed to be observed, while in [19] the population can be observed partially but the birth rates are assumed to be constant (not age-dependent).

In this paper, we consider an extension of the linear birth-and-death process, allowing for non-constant birth and death rates, that is both tractable and suitable for an age-specific analysis. We model the lifetime and reproduction epochs of individuals in a population using a transient *Markovian arrival process* (transient MAP or TMAP, for short). In brief, a TMAP is a point process in which the event rate depends on the state of an underlying transient continuous-time Markov chain with n transient states, also called *phases*, and one absorbing state. Each arrival in the TMAP corresponds to the birth of a child, and absorption in phase 0 corresponds to the individual's death. The phase of a child at birth can potentially depend on the parent's phase at the time of reproduction. Under these assumptions, the lifetimes follow phase-type distributions, which are known to be dense in the set of non-negative real-valued distributions; similarly, Markovian arrival processes are dense in the set of point processes. The resulting continuous-time branching processes, called *Markovian binary trees* (MBTs), are the matrix generalisation of linear birth-and-death processes, to the same extent that phase-type distributions are the matrix generalisation of exponential distributions, and Markovian arrival processes are the matrix generalisation of Poisson processes [20]. As a consequence, MBTs offer much more flexibility than linear birth-and-death processes, while keeping an excellent computational tractability due to their Markovian nature.

The MBT model has already been used to efficiently compare demographic properties of female families in different countries: in [9] the authors assume that each phase of the underlying TMAP corresponds to an age-class and use the available age-specific fertility and mortality rates directly as the transition rates in the model. In [8], an Expectation-Maximisation algorithm is developed to estimate the parameters of MBTs from the continuous observation of some populations when the phases of the underlying process

are not observed; the method is then used to estimate the speciation rates in phylogenetic trees for various values of the number of phases n.

We note that, despite their similar terminology, MBTs are not to be confused with *Markov branching models* developed in the context of phylogenetics, such as the *beta-splitting model*, which are discrete recursive structures (cladograms) specified through symmetric split distributions, and are motivated by the study of phylogenetic tree imbalance [1, 2, 4]. Yet, MBTs would also be suitable models for phylogenetic trees, as they allow for speciation and extinction rate variations among species, and cover a wide range of imbalance [18]. Following the terminology in Stadler's survey [30], MBTs are *species-non-exchangeable models* with *asymmetric speciation*: they account for *trait-dependent* speciation and extinction rates, and therefore some coexisting species may be more likely to speciate (give birth) or become extinct (die) at a given point in time than others. MBTs encompass many models in which speciation rates can be inherited from the parental species rather than being constant over the phylogeny. Statistical methods for estimating the trait-dependent speciation and extinction rates, based on observed phylogenetic trees (with branch lengths), are discussed in [30], however the techniques generally require some knowledge of the traits. In an MBT, traits may be modelled by the phases of the TMAP, which we assume are not observable.

Here, we adopt a fitting approach for MBTs different from [8] and [9]. Our motivation is to develop statistical tools to fit MBTs to populations of species for which detailed information about age-specific survival and reproductive rates of individuals is available. The model can then be used to calculate age-dependent demographic properties and to assess the viability of the population, which aids in the conservation management of endangered species. Examples of MBT outputs include the probability of extinction of the population, the distribution of the time until extinction, and the distribution of the population size at any given time, conditional on the age structure of the population at initial time [10, 11, 12].

We fit our model to different types of datasets which may be available from demographic databases or from field studies of animal populations. These datasets may have different degrees of detail. We distinguish between:

- *Global population data*, consisting of the *average* age-specific fertility and mortality *rates* over an entire population. This type of data is usually provided in databases on human fertility and mortality.

- *Individual demographic data*, consisting of data on age-specific fertility

and mortality *counts* for each individual in a population. This type of data often exists for closely monitored animal species. Here we will use a rich dataset on a highly threatened bird species, the Chatham Island black robin (*Petroica traversi*) [5, 17, 24].

To fit an MBT to these types of data, we will assume that the phase of a child at birth is independent from the parental phase at reproduction time. The problem therefore is reduced to fitting the underlying TMAP.

The parameter estimation method depends on the type of data which are available: in the global population data case, we use a weighted non-linear regression technique to fit the parameters, and in the individual demographic data case, we use a maximum likelihood approach; this leads to the first direct parameter estimation method for TMAPs based on the number of arrivals falling in successive time intervals. We consider different validation methods to determine the optimal number of phases n in the TMAP, and we discuss the construction of empirical and theoretical confidence intervals for the model outputs.

We apply our results to two real-world examples. The weighted non-linear regression method is applied to human demography, and leads to an MBT model which better fits the data than the one considered in [9]. The maximum likelihood method is applied to the black robin population, and provides insights about the species demography. Complimentary to these results, a parallel paper on the asymptotic age frequency in Markovian population models [13] makes use of an MBT model for the black robin population whose parameters are estimated using the methodology developed in Section 3, and refers to the present paper for the details of the method.

The paper is organised as follows: in Section 2, we define the TMAP and the MBT, and we describe the special case that we shall focus on. In Section 3, we develop a method to estimate the model parameters based on the average age-specific fertility and mortality rates, and in Section 4, we do the same based on individual age-specific fertility and mortality counts. In Section 5 we briefly discuss the goodness of fit of our methods through some artificial examples (we refer to Supplementary material for more details), and we apply each method to a real-world example. Three appendices gather proofs and extensions of some of our results.

Unless stated otherwise, we implicitly assume that all vectors are column vectors. We let $\mathbf{1}$ and $\mathbf{0}$ denote, respectively, vectors of ones and zeros, whose size will be clear from the context.

2. Transient Markovian arrival processes and Markovian binary trees

The lifetime and reproduction epochs of an individual are controlled by a *transient Markovian arrival process* (TMAP), which is a two-dimensional continuous-time Markov process $\{(X(t), \varphi(t)) : t \in \mathbb{R}^+\}$ on the state space $\mathbb{N}_0 \times \{0, 1, \dots, n\}$, where n is finite. It combines the *level* process $\{X(t)\}$, which counts the number of arrivals (births) in $[0, t]$, with the *phase* process $\{\varphi(t)\}$, which is an underlying Markov chain with a single absorbing phase 0. The states $(k, 0)$ of $\{(X(t), \varphi(t))\}$ are absorbing for all $k \geq 0$; the other states are transient.

A TMAP is characterised by two $n \times n$ transition rate matrices D_0 and D_1, and a non-negative $n \times 1$ absorption (death) rate vector $\boldsymbol{d}$. For $k \geq 0$, feasible transitions are

(i) from (k, i) to (k, j), for $1 \leq i \neq j \leq n$, at rate $(D_0)_{ij}$, or

(ii) from (k, i) to $(k + 1, j)$, for $1 \leq i, j \leq n$, at rate $(D_1)_{ij}$, or

(iii) from (k, i) to $(k, 0)$, for $1 \leq i \leq n$, at rate d_i.

Transitions *(i)* are *hidden*: the phase of the individual changes but the level is not incremented. Transitions *(ii)* are *observable*: a birth is recorded, and the phase of the individual may or may not change. Transitions *(iii)* indicate the termination of the individual's life.

The matrix D_1 and the vector $\boldsymbol{d}$ are nonnegative, D_0 has nonnegative off-diagonal elements and strictly negative elements on the diagonal such that $D\,\mathbf{1} + \boldsymbol{d} = \mathbf{0}$, where the matrix $D := D_0 + D_1$ corresponds to the transient generator of the underlying phase process $\{\varphi(t)\}$. We also define the initial probability row vector $\boldsymbol{\alpha}^\top = (\alpha_i)_{1 \leq i \leq n}$, where $\alpha_i = \mathbb{P}(\varphi(0) = i)$, and we assume that $\boldsymbol{\alpha}^\top \mathbf{1} = 1$, so that $\varphi(0) \neq 0$ a.s. In a TMAP, the intervals of time between successive birth events are not necessarily independent nor identically distributed. The lifetime of a TMAP corresponds to the time until absorption of $\{\varphi(t)\}$, and therefore has a phase-type distribution $\mathrm{PH}(\boldsymbol{\alpha}^\top, D)$, see [21]. More details on TMAPs can be found in [21].

To define a *Markovian binary tree* (MBT), it remains to specify an $n \times n^2$ birth rate matrix B, whose entry $B_{i;jk}$ records the rates at which a parent in phase i makes a transition to phase k while giving birth to a child in phase j. To be consistent with the definition of D_1, B should be such that $\sum_{j=1}^{n} B_{i;jk} = (D_1)_{ik}$. One particular example is given by $B = \boldsymbol{\alpha}^\top \otimes D_1$, where $\otimes$ denotes the Kronecker product between matrices; componentwise, this

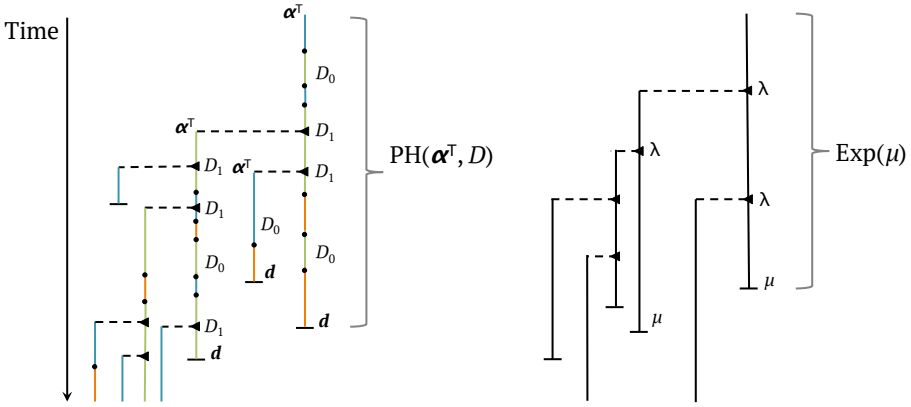

Figure 1: A graphical representation of the evolution of an MBT (left) and a linear birth-and-death process (right) with their corresponding transition rates. Vertical lines represent the lifetime of individuals. In an MBT, each individual makes transitions within unobservable phases (coloured segments in the graph); an individual's lifetime and reproduction events are modelled by a TMAP with parameters $\boldsymbol{\alpha}^\top, D_0, D_1$ and $\boldsymbol{d}$, and therefore the lifetime has a phase-type distribution PH$(\boldsymbol{\alpha}^\top, D)$, where $D = D_0 + D_1$. In a linear birth-and-death process, each individual reproduces following a Poisson process with parameter λ and has an exponentially distributed lifetime with parameter μ.

means $B_{i;jk} = \alpha_j (D_1)_{ik}$. This corresponds to the case where the child's initial phase is independent from the parent's phase transition at reproduction time; due to the nature of the data, we will assume that we are in this case throughout the paper. We refer to Figure 1 for a graphical representation of the lifetime and reproduction process of individuals in the corresponding MBT. Once properly fitted to the data, the MBT can be used to study demographic properties of the whole population (see Section 5.3).

Markovian arrival processes are known to be dense in the set of point processes, that is, they can approximate any point process to an arbitrary degree of accuracy [3]. This property, together with their tractability, make TMAPs a convenient tool to model the reproduction process and lifetime of individuals. If no assumption is made on their structure, the matrices $\boldsymbol{\alpha}^\top, D_0, D_1$ and $\boldsymbol{d}$ contain a total of $p = 2n^2 + n - 1$ entries. A special class of TMAPs, called *acyclic transient Markov modulated Poisson processes* (which we shall denote as TMAP*s for short), assumes

- individuals start their lifetime in phase 1 with probability one,
- they can only move from phase i to phase $i+1$ or to phase 0, with

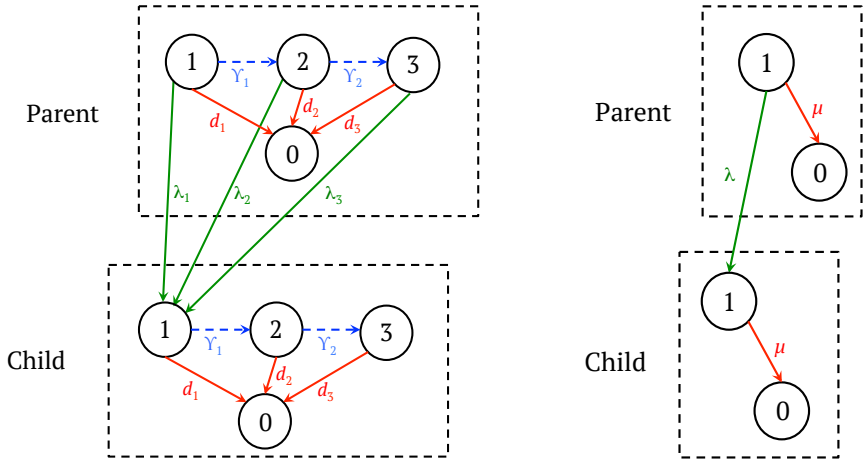

Figure 2: The transitions rates for a typical parent in a TMAP* with $n = 3$ transient phases (left), and in a linear birth-and-death process (right) which corresponds to a TMAP* with $n = 1$. The absorbing phase 0 corresponds to the death of the individual. The observable transitions associated with a birth or a death event are represented by plain arrows, and the hidden transitions are represented by dashed arrows. Note that phases are never observed; the only knowledge on the phases is that every child starts in phase 1. These diagrams illustrate the parameters to be estimated.

respective rates γ_i for $1 \le i \le n-1$ and d_i for $1 \le i \le n$,

- while in phase i, they reproduce at rate λ_i and do not make any simultaneous phase transition at reproduction time.

Under these assumptions, $\boldsymbol{\alpha}^\top = [1, 0, \dots, 0]$, $D_1 = \text{diag}(\lambda_1, \dots, \lambda_n)$, and the only non-zero entries of D_0 are $(D_0)_{i,i+1} = \gamma_i$ and

$$(D_0)_{ii} = \begin{cases} -\lambda_i - d_i - \gamma_i, & 1 \le i \le n-1, \\ -\lambda_i - d_i, & i = n. \end{cases}$$

We denote by $\boldsymbol{\theta} := \{\boldsymbol{\alpha}^\top, D_0, D_1, \boldsymbol{d}\}$ the full set of parameters. There is a total of $p = 3n - 1$ parameters in a TMAP*. Figure 2 shows the possible transitions for a typical individual (parent) in a TMAP* with $n = 3$, compared with the linear birth-and-death process with birth rate λ and death rate μ. Due to the structure of D_0 and D_1 in the TMAP* case, the phase-type distributed lifetime corresponds more precisely to a *Coxian distribution*. Such distributions are important as any acyclic phase-type distribution has an equivalent Coxian representation. Therefore, in terms of the lifetime dis-

tribution, the TMAP* does not impose much restriction compared to the general TMAP.

We emphasize that the n transient phases do not necessarily have a physical interpretation and do not generally correspond to age-classes. They are fictitious stages of exponential length during an individual's lifetime, and their purpose is to make the model more accurate and flexible than a linear birth-and-death process. In that view, we will not be interested in the specific value of the model parameters, but rather in the model performance measures (outputs) which have a biological interpretation.

3. Global population data

3.1. Available data and model equivalent

In this section, we assume the available data are estimates of the *expected* age-specific fertility rates, $\hat{b}_x$, and of the *expected* age-specific mortality rates, $\hat{d}_x$, where $x \in \{0, 1, 2, \ldots, M\}$ denotes the age, that is, the period of time $[x, x+1)$ during the lifetime, and M is the maximal age for which data are available. The method developed in this section can be generalised to ℓ-year age classes, $\ell > 1$; details are provided in Appendix B.

The rates $\hat{b}_x$ and $\hat{d}_x$ are interpreted, respectively, as the expected number of offspring per year from a parent at age x, and the probability that an individual who reached age x dies within the year. We denote by $\bar{d}(x, \boldsymbol{\theta})$ and $\bar{b}(x, \boldsymbol{\theta})$ the equivalent quantities computed from the TMAP with parameters $\boldsymbol{\theta}$ (in which the age x takes non-negative real values). These functions have the following analytic expression, the proof of which is provided in Appendix A.

Proposition 3.1. *The age-specific mortality and fertility rates at age $x \in \mathbb{R}^+$ in a TMAP are respectively given by*

$$\begin{aligned} \bar{d}(x, \boldsymbol{\theta}) &= \frac{\boldsymbol{\alpha}^\top e^{Dx}(I - e^D)\mathbf{1}}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}, \\ \bar{b}(x, \boldsymbol{\theta}) &= \frac{\boldsymbol{\alpha}^\top e^{Dx}(I - e^D)(-D)^{-1}D_1\mathbf{1}}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}, \end{aligned}$$

where $\boldsymbol{\alpha}^\top$ is the initial probability vector and $D = D_0 + D_1$ is the transition rate matrix of the phase process $\{\varphi(t)\}$, as defined in Section 2.

3.2. Parameter estimation

Since the data do not correspond to raw observations but to expected values, a maximum likelihood approach is not appropriate here. The objective is to find the TMAP whose corresponding age-specific mortality and fertility curves $\bar{d}(x,\boldsymbol{\theta})$ and $\bar{b}(x,\boldsymbol{\theta})$ are at minimal distance from the data $\hat{d}_x$ and $\hat{b}_x$. We extend the regression approach of Lin and Liu [22], who used death rates only to fit phase-type lifetime distributions. The model parameters are estimated by minimizing the sum of weighted squared errors for both the birth and the death rates: $\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} F(\boldsymbol{\theta})$, where

$$F(\boldsymbol{\theta}) = \sum_{x=0}^{M} \left[(\hat{d}_x - \bar{d}(x,\boldsymbol{\theta}))^2 + (\hat{b}_x - \bar{b}(x,\boldsymbol{\theta}))^2 \right] \hat{S}_x, \tag{3.1}$$

and the weights $\hat{S}_x$ are the observed probabilities of survival until age x,

$$\hat{S}_x = (1-\hat{d}_0)(1-\hat{d}_1)\cdots(1-\hat{d}_{x-1}).$$

As age increases, there may be less available data, leading to higher variance in the age-specific mortality and fertiliy rates. The weights $\hat{S}_x$, which were already used in [22], balance the potential resulting heteroscedasticity. If the estimated age-specific rates $\hat{d}_x$ and $\hat{b}_x$ are computed as averages of n_x uncorrelated raw observations, another simple choice of weights would be the values n_x themselves (see [14, pp.95–96]). Note that n_x is not always available in practice, but it is proportional to the observed probability of survival until age x if the latter is estimated as $\hat{S}_x = n_x/N$ from a cohort starting with N individuals.

Since the functions $\bar{d}(x,\boldsymbol{\theta})$ and $\bar{b}(x,\boldsymbol{\theta})$ are non-linear in both the input variable x and in the parameters $\boldsymbol{\theta}$ of the TMAP, we are dealing with a weighted non-linear regression. If there is missing information in the data and no estimate exists for $\hat{d}_x$ or $\hat{b}_x$ for some age x, then we set the corresponding term $(\hat{d}_x - \bar{d}(x,\boldsymbol{\theta}))^2$ or $(\hat{b}_x - \bar{b}(x,\boldsymbol{\theta}))^2$ to zero in the sum.

Remark. The function $\bar{d}(x,\boldsymbol{\theta})$ corresponds to the hazard rate at age x in survival analysis. Several hazard models have been considered to fit mortality data, such as the Gompertz-Makeham, the Siler, and the Heligman-Pollard models [7]. Similarly, several age-specific fertility models have been studied, including the Hadwiger, the Beta, and the Gamma models [26]. Our aim here is not to compare these models with the fits provided by the functions $\bar{d}(x,\boldsymbol{\theta})$ and $\bar{b}(x,\boldsymbol{\theta})$, and we do not claim our results to be more accurate. However, we stress that the novelty of our approach lies in the

fact that, as opposed to the known mortality and fertility models which are generally studied *separately*, $\bar{d}(x,\boldsymbol{\theta})$ and $\bar{b}(x,\boldsymbol{\theta})$ are performance measures coming from *the same Markovian model with parameters* $\boldsymbol{\theta}$, and are thus optimised together. The estimated Markovian model then corresponds to the best model fitting *both* the mortality and fertility data, and can be used to make a complete demographic study of the population, as shown in [9]. In particular, the MBT model allows us to compute properties of the population such as its extinction probability, the distribution of its size at a given time, etc., which cannot be obtained using the separate models for mortality and fertility.

3.3. Goodness of fit and optimal value of n

The *mean squared error* of the estimator $\hat{\boldsymbol{\theta}}$ is given by

$$\mathrm{MSE} = \mathbb{E}\left[\sum_{x=0}^{M}\left[(\bar{d}(x,\boldsymbol{\theta})-\bar{d}(x,\hat{\boldsymbol{\theta}}))^2+(\bar{b}(x,\boldsymbol{\theta})-\bar{b}(x,\hat{\boldsymbol{\theta}}))^2\right]\bar{S}(x,\boldsymbol{\theta})\right],$$

where $\bar{d}(x,\boldsymbol{\theta})$, $\bar{b}(x,\boldsymbol{\theta})$, and $\bar{S}(x,\boldsymbol{\theta})$ are respectively the age-specific mortality function, the age-specific fertility function, and the age-specific survival function corresponding to the *true* model, and $\bar{d}(x,\hat{\boldsymbol{\theta}})$ and $\bar{b}(x,\hat{\boldsymbol{\theta}})$ are the equivalent functions corresponding to the *estimated* model. If the true model is known, then the MSE can be estimated through resampling. Alternatively, it could be estimated when we are given a collection of datasets, each containing global population information.

The value of the MSE indicates the goodness of fit of the model. When the true model is unknown, a way to determine the optimal number of phases n is to minimise the MSE. When the true model is known, comparing the MSE for different values of n informs us on the sensitivity of the output with respect to the number of phases, as illustrated in Supplementary material.

4. Individual demographic data

4.1. Available data

In this section, we assume the data are *individual* age-specific fertility and mortality counts in successive age-classes of length $\ell > 0$ (these age-classes are assumed of equal length, but though computationally convenient, this is not essential). The sample then consists of N vectors (one for each individual) of the type

$$\boldsymbol{v} = [6,\ 8,\ -2,\ 9,\ 0,\ 3,\ 3,\ -1], \tag{4.1}$$

of variable length, whose entries $v_i, i \geq 1$, are interpreted as follows:

- $v_i = k \in \{0, 1, 2, \ldots\}$ if the individual had k offspring while in the age-class $[(i-1)\ell, i\ell)$,
- $v_i = -1$ if the individual died in the previous age-class $[(i-2)\ell, (i-1)\ell)$, possibly after producing some offspring, and
- $v_i = -2$ if the individual was alive at the beginning of the age-class $[(i-1)\ell, i\ell)$ but there is no (or incomplete) information on her progeny in that age-class.

4.2. Parameter estimation

We denote by $p(\boldsymbol{v}|\boldsymbol{\theta})$ the probability of observing the individual life vector $\boldsymbol{v}$, under the model parameters $\boldsymbol{\theta}$. As we describe below, this probability can be computed analytically for any life vector $\boldsymbol{v}$ using the properties of TMAPs. We therefore follow a maximum likelihood approach here: based on a sample of i.i.d. individual life vectors $\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(N)}\}$, we maximize the log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(N)}) = \sum_{j=1}^{N} \log p(\boldsymbol{v}^{(j)}|\boldsymbol{\theta}). \tag{4.2}$$

Let $K = \max_{i,j}\{v_i^{(j)} : 1 \leq i, 1 \leq j \leq N\}$ be the maximum number of offspring over all age-classes and all individuals in the sample. The probabilities $p(\boldsymbol{v}^{(j)}|\boldsymbol{\theta})$ can be written as matrix products involving the matrices and vectors $P(k) = (P_{ij}(k))$, $\boldsymbol{p}(k) = (p_i(k))$, $P = (P_{ij})$, and $\boldsymbol{p} = (p_i)$, defined as

$$\begin{aligned}
P_{ij}(k) &:= \mathbb{P}(X(\ell) = k, \varphi(\ell) = j | X(0) = 0, \varphi(0) = i), && (4.3)\\
p_i(k) &:= \mathbb{P}(X(\ell) = k, \varphi(\ell) = 0 | X(0) = 0, \varphi(0) = i), && (4.4)\\
P_{ij} &:= \mathbb{P}(\varphi(\ell) = j | \varphi(0) = i) = \sum_{k \geq 0} P_{ij}(k), && (4.5)\\
p_i &:= \mathbb{P}(\varphi(\ell) = 0 | \varphi(0) = i) = \sum_{k \geq 0} p_i(k), && (4.6)
\end{aligned}$$

for $1 \leq i, j \leq n$ and $1 \leq k \leq K$. As an illustrative example, consider the four life vectors

$$\boldsymbol{v}^{(1)} = [2, 3, 1, -1], \quad \boldsymbol{v}^{(2)} = [2, -2, 1, -1], \quad \boldsymbol{v}^{(3)} = [2, 3], \quad \boldsymbol{v}^{(4)} = [2, -2].$$

By conditioning on the phases of the TMAP at the boundaries of the successive ℓ-year intervals, the probability of observing these vectors is

$$\begin{aligned} p(\boldsymbol{v}^{(1)}|\boldsymbol{\theta}) &= \boldsymbol{\alpha}^\top P(2)P(3)\,\boldsymbol{p}(1), & p(\boldsymbol{v}^{(2)}|\boldsymbol{\theta}) &= \boldsymbol{\alpha}^\top P(2)P\,\boldsymbol{p}(1), \\ p(\boldsymbol{v}^{(3)}|\boldsymbol{\theta}) &= \boldsymbol{\alpha}^\top P(2)\,[P(3)\mathbf{1}+\boldsymbol{p}(3)], & p(\boldsymbol{v}^{(4)}|\boldsymbol{\theta}) &= \boldsymbol{\alpha}^\top P(2)\mathbf{1}. \end{aligned}$$

Note that if $\boldsymbol{v}$ is a vector of size $M+1$ with all entries equal to -2, then $p(\boldsymbol{v}|\boldsymbol{\theta}) = \boldsymbol{\alpha}^\top P^M\mathbf{1}$ is the probability that the individual survives at least the first M age-classes.

The quantities defined in (4.3)–(4.6) can be computed explicitly, as shown in the next proposition, whose proof is provided in Appendix C.

Proposition 4.1. *For $1 \le k \le K$, the matrix $P(k)$ and the vector $\boldsymbol{p}(k)$ are given by*

$$\begin{aligned} P(k) &= (1/k!)(\boldsymbol{e}_k^\top \otimes I)\exp(\mathcal{M}\ell)(\boldsymbol{e}_1 \otimes I), \\ \boldsymbol{p}(k) &= (1/k!)(\boldsymbol{e}_k^\top \otimes I)[I-\exp(\mathcal{M}\ell)](-\mathcal{M})^{-1}\,(\boldsymbol{e}_1 \otimes I)\,\boldsymbol{d}, \end{aligned}$$

where $\boldsymbol{e}_k$ is the kth unit column vector of size K, and

$$\mathcal{M} = \begin{bmatrix} D_0 & & & & \\ D_1 & D_0 & & & \\ & 2D_1 & D_0 & & \\ & & \ddots & & \\ & & & KD_1 & D_0 \end{bmatrix}.$$

The matrix P and vector $\boldsymbol{p}$ are given by

$$P = \exp(D\ell), \qquad \boldsymbol{p} = [I-\exp(D\ell)](-D)^{-1}\boldsymbol{d}.$$

Our maximum likelihood method generalises the results in Davison and Ramesh [6] who consider the parameter estimation of Markov modulated Poisson processes in the binary data case where $v_i = 1$ if there is at least one event in the interval $[(i-1)\ell), i\ell)$, and $v_i = 0$ otherwise. To our knowledge, other parameter estimation methods for these processes are based on the observation of the successive inter-event times rather than on the number of events within successive time intervals, see for instance [27] and [29]. As confirmed in Supplementary material, Figure S5, when the length of the time intervals decreases to zero, the estimates obtained with our method converge to those obtained with the method based on the observation of the

successive inter-event times.

4.3. Goodness of fit and optimal number of phases

We consider three different criteria for choosing the optimal value of the number n of phases. These criteria are compared on numerical examples in Supplementary material.

Akaike Information Criterion (AIC). We choose the value of n which minimises the AIC defined as

$$\mathrm{AIC} = 2p - 2\mathcal{L}(\hat{\boldsymbol{\theta}}; \boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(N)}),$$

where the number of parameters $p = 3n - 1$ for a TMAP* model with n phases. This criterion deals with the trade-off between goodness of fit of the model and its complexity. One advantage is that it does not rely on the knowledge of the true model: the value of AIC can be directly computed from the log-likelihood of the estimated model given the data.

Cross-validation (CV). We perform a K-fold cross-validation over the data sample of individual life vectors (with typical value $K = 5$). The idea is to randomly divide the data into K equal-sized parts. We leave out part k, fit the model to the other $K - 1$ parts (combined), and then evaluate the likelihood of the left-out kth part (test set) under the estimated parameters. We choose the model maximising the mean test log-likelihood obtained by averaging the results for $k = 1, 2, \ldots, K$. Similar to the AIC, this method does not require us to know the true model.

Mean squared integrated loss (MSIL). Let $\mathcal{V}$ be the set of all life vectors $\boldsymbol{v}$ with entries in $\mathbb{N}_0 \cup \{-1\}$. Any life vector with at least one entry equal to -2 is interpreted as a (disjoint) union of vectors in $\mathcal{V}$: indeed if $v_i = -2$ for some $\boldsymbol{v}$ and i, then $\boldsymbol{v} := \{\boldsymbol{v}' \in \mathcal{V} : v'_i \geq 0 \text{ and } v'_j = v_j \text{ for all } j \neq i\}$. For any fixed number of phases n, and a given sample of life vectors $\{\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots \boldsymbol{v}^{(N)}\}$, the parameters $\boldsymbol{\theta}_n$ are estimated using the maximum likelihood method. From the estimate $\hat{\boldsymbol{\theta}}_n$ we define a corresponding probability mass function $\hat{f}_n(\cdot)$ over $\mathcal{V}$ as

$$\hat{f}_n(\boldsymbol{v}) := p(\boldsymbol{v}|\hat{\boldsymbol{\theta}}_n), \quad \boldsymbol{v} \in \mathcal{V},$$

where $p(\cdot|\hat{\boldsymbol{\theta}}_n)$ is the probability measure defined in Section 4.2. The optimal number of phases is the value of n minimising the *mean squared integrated*

loss, defined as

$$\begin{aligned} \text{MSIL} &= \mathbb{E}\left[\sum_{\boldsymbol{v}\in\mathcal{V}}(f(\boldsymbol{v})-\hat{f}_n(\boldsymbol{v}))^2\right] \\ &= \sum_{\boldsymbol{v}\in\mathcal{V}} f(\boldsymbol{v})^2 - 2\mathbb{E}[\sum_{\boldsymbol{v}\in\mathcal{V}} f(\boldsymbol{v})\hat{f}_n(\boldsymbol{v})] + \mathbb{E}[\sum_{\boldsymbol{v}\in\mathcal{V}} \hat{f}_n(\boldsymbol{v})^2], \end{aligned} \tag{4.7}$$

where $f(\boldsymbol{v}) := p(\boldsymbol{v}|\boldsymbol{\theta})$ denotes the true probability mass function over $\mathcal{V}$. Since the first term is independent of n, the value of n minimising $\text{MSIL}^*(n) := \mathbb{E}[\sum_{\boldsymbol{v}\in\mathcal{V}} \hat{f}_n(\boldsymbol{v})^2] - 2\mathbb{E}[\sum_{\boldsymbol{v}\in\mathcal{V}} f(\boldsymbol{v})\hat{f}_n(\boldsymbol{v})]$ also minimises the MSIL. The problem therefore reduces to estimating $\text{MSIL}^*(n)$ for each $n \geq 1$.

If the true model is known, then $f(\cdot)$ is known, and the expectations in $\text{MSIL}^*(n)$ can be estimated through resampling. If the true model is unknown, then a K-fold cross-validation method can be applied to estimate $\text{MSIL}^*(n)$. Let A_k and B_k be the kth training set and test set, respectively. Let $\hat{f}_n^k(\cdot)$ denote the probability mass function estimator using n phases and training set A_k. We have

$$\mathbb{E}\left[\sum_{\boldsymbol{v}\in\mathcal{V}} \hat{f}_n(\boldsymbol{v})^2\right] \approx \frac{1}{K}\sum_{k=1}^{K}\sum_{\boldsymbol{v}\in\mathcal{V}} \hat{f}_n^k(\boldsymbol{v})^2, \tag{4.8}$$

and since the sets B_k are all drawn from the true distribution $f(\cdot)$, we have

$$\mathbb{E}\left[\sum_{\boldsymbol{v}\in\mathcal{V}} f(\boldsymbol{v})\hat{f}_n(\boldsymbol{v})\right] \approx \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|B_k|}\sum_{\boldsymbol{v}\in B_k} \hat{f}_n^k(\boldsymbol{v}).$$

The set of life vectors $\mathcal{V}$ being infinite, the sums in (4.7) and (4.8) need to be modified in practice. For a given pair of integers L and M, we partition the set $\mathcal{V}$ to form a new *finite* set $\tilde{\mathcal{V}}_{L,M}$ such that $\sum_{\boldsymbol{v}\in\mathcal{V}} f(\boldsymbol{v}) = \sum_{\tilde{\boldsymbol{v}}\in\tilde{\mathcal{V}}_{L,M}} f(\tilde{\boldsymbol{v}}) = 1$. The vectors $\tilde{\boldsymbol{v}} \in \tilde{\mathcal{V}}_{L,M}$ are of length M and have their entries in the finite set $\{-1, 0, 1, 2, \ldots, L, L+1\}$, so that

$$|\tilde{\mathcal{V}}_{L,M}| = \sum_{i=1}^{M}(L+2)^i = (L+2)\frac{(L+2)^M - 1}{(L+1)}.$$

They define equivalence classes in $\mathcal{V}$ as follows:

- if $-1 \leq \tilde{v}_i \leq L$ for all $1 \leq i \leq M$, then

$$\tilde{\boldsymbol{v}} := \big\{\boldsymbol{v} \in \mathcal{V} : v_i = \tilde{v}_i, \text{ for all } i \in \{1, \ldots, M\}\big\},$$

  in which case $f(\tilde{\boldsymbol{v}}) = p(\tilde{\boldsymbol{v}}|\boldsymbol{\theta})$;

- if $\tilde{v}_{i_1} = \ldots = \tilde{v}_{i_\ell} = L+1$ for some $1 \leq i_1, \ldots i_\ell \leq M$, $\ell \geq 1$, then

$$\begin{aligned}\tilde{\boldsymbol{v}} \;:=\; \big\{\boldsymbol{v} \in \mathcal{V} : v_i = \tilde{v}_i \text{ for all } i \in \{1, \ldots, M\} \setminus \{i_1, \ldots i_\ell\},\\ \text{and } v_i \geq L+1 \text{ for all } i \in \{i_1, \ldots i_\ell\}\big\},\end{aligned}$$

  in which case $f(\tilde{\boldsymbol{v}})$ is computed from the probability measure $p(\cdot|\boldsymbol{\theta})$ as given in the next Lemma.

Lemma 4.1. *For any $\tilde{\boldsymbol{v}} \in \tilde{\mathcal{V}}_{L,M}$ such that $\tilde{v}_{i_1}, \ldots, \tilde{v}_{i_\ell} = L+1$ for some indices $1 \leq i_1, \ldots i_\ell \leq M$, $\ell \geq 1$, we have*

$$f(\tilde{\boldsymbol{v}}) = \sum_{k_1,\ldots,k_\ell \in \{-2,0,1,\ldots,L\}} (-1)^{\ell + \sum_{i=1}^{\ell} \mathbb{1}\{k_i = -2\}}\, p(\boldsymbol{v}^*(k_1, \ldots, k_\ell)|\boldsymbol{\theta}), \qquad (4.9)$$

where the vector $\boldsymbol{v}^*(k_1, \ldots, k_\ell)$ is such that, for $1 \leq i \leq M$,

$$v_i^*(k_1, \ldots, k_\ell) = \begin{cases} \tilde{v}_i & \text{if } i \notin \{i_1, \ldots i_\ell\} \\ k_j & \text{if } i = i_j \text{ for some } 1 \leq j \leq \ell. \end{cases}$$

Proof. We know $f(\tilde{\boldsymbol{v}}) = \sum_{k_1,\ldots,k_\ell \geq L+1} p(\boldsymbol{v}^*(k_1, \ldots, k_\ell)|\boldsymbol{\theta})$ by the definition of $\tilde{\boldsymbol{v}}$ and $\boldsymbol{v}^*(k_1, \ldots, k_\ell)$. This sum contains ℓ embedded sums of the form $\sum_{k_j \geq L+1}$, which we rewrite as $\sum_{k_j \geq L+1} = \sum_{k_j \geq 0} - \sum_{0 \leq k_j \leq L}$. Using

$$\sum_{k_j \geq 0} p(\boldsymbol{v}^*(k_1, \ldots, k_j, \ldots, k_\ell)|\boldsymbol{\theta}) = p(\boldsymbol{v}^*(k_1, \ldots, -2, \ldots, k_\ell)|\boldsymbol{\theta}),$$

and rearranging the terms then leads to (4.9). □

Replacing $\mathcal{V}$ by $\tilde{\mathcal{V}}_{L,M}$ results in a different version of the MSIL criterion which selects the best model capturing differences in the number of children less than or equal to L over the first M age-classes. It is clear that the larger the age-class length ℓ, the smaller M and the larger L should be chosen in order for $\mathcal{V}$ to be well approximated by $\tilde{\mathcal{V}}_{L,M}$. When $\ell = 1$, a possible choice of the partitioning parameters is taking M as the ceiling of the expected lifetime plus one, and $L+1$ as the maximal expected number of children per

age-class, that is,

$$M = \lceil \sum_{x \geq 1} \hat{S}_x \rceil + 1, \quad \text{and} \quad L + 1 = \lceil \max_{1 \leq x \leq M} \hat{b}_x \rceil. \tag{4.10}$$

Another choice leading to smaller equivalence classes in $\mathcal{V}$ is

$$M = \min\{x \geq 0 : \hat{S}_x < p\} + 1, \quad \text{and} \quad L + 1 = \lceil \max_{1 \leq x \leq M} (\hat{b}_x + \hat{\sigma}_x) \rceil, \tag{4.11}$$

where $1 - p$ is a covering probability, and $\hat{\sigma}_x$ is the standard error of the age-specific fertility rate at age x. Formulae for $\ell-$year age-classes are analogous.

4.4. Confidence intervals for the model outputs

For any real-valued performance measure $g(x, \boldsymbol{\theta})$ of the model, such as the mortality and fertility functions $\bar{d}(x, \boldsymbol{\theta})$ and $\bar{b}(x, \boldsymbol{\theta})$ at age $x \in \mathbb{R}^+$, we can construct empirical percentiles and theoretical confidence intervals for $g(x, \hat{\boldsymbol{\theta}})$.

If the true model is known, the pointwise mean $\mu_{\hat{\boldsymbol{\theta}}}(x)$ and standard deviation $\sigma_{\hat{\boldsymbol{\theta}}}(x)$ of the estimator $g(x, \hat{\boldsymbol{\theta}})$ can be estimated by drawing independent samples from the true model. For large values of the sample size N, a 95% empirical confidence interval for $g(x, \hat{\boldsymbol{\theta}})$ can then be approximated by $\mu_{\hat{\boldsymbol{\theta}}}(x) \pm 1.96\, \sigma_{\hat{\boldsymbol{\theta}}}(x)$ for each value of $x \in \mathbb{R}^+$; the width of the resulting band gives us an indication of the variability of the estimated model, given the true model. If the true model is unknown, bootstrapping from a single data sample substitutes independent redrawing from the true model, and provides bootstrap percentile confidence intervals.

For every fixed x, asymptotic theoretical confidence intervals for $g(x, \hat{\boldsymbol{\theta}})$ are found using the delta method,

$$g(x, \hat{\boldsymbol{\theta}}) \sim \mathcal{N}(g(x, \boldsymbol{\theta}), \nabla g(x, \boldsymbol{\theta})^\top\, J(\hat{\boldsymbol{\theta}})^{-1}\, \nabla g(x, \boldsymbol{\theta})), \quad \text{as } N \to \infty,$$

where $\nabla g(x, \boldsymbol{\theta})$ is the gradient of $g(x, \boldsymbol{\theta})$, and $J(\hat{\boldsymbol{\theta}}) := (\partial^2 \mathcal{L}(\boldsymbol{\theta})/(\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top))|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$ is the observed information matrix. A 95% pointwise confidence interval for $g(x, \boldsymbol{\theta})$ is then given by

$$g(x, \hat{\boldsymbol{\theta}}) \pm 1.96 \sqrt{\nabla g(x, \hat{\boldsymbol{\theta}})^\top\, J(\hat{\boldsymbol{\theta}})^{-1}\, \nabla g(x, \hat{\boldsymbol{\theta}})}. \tag{4.12}$$

5. Numerical applications

In this section, the two parameter estimation methods are applied to three types of illustrative examples. We first analyse artificial examples in which we simulate TMAP*s, we estimate their parameters based on the simulations, we make a goodness of fit analysis and compare the different criteria for choosing the optimal number of phases. We provide a summary of the results here and refer to Supplementary material for details. Next, we use real global female population data in different countries to estimate the model parameters. Finally, we fit an MBT using real individual demographic data on the black robin population, and we give a biological interpretation to our results.

We use the Matlab function `fmincon` to minimise the sum of weighted squared errors (3.1) in the global population data case, or to maximise the log-likelihood function (4.2) in the individual demographic data case, under the constraint of positive parameters. The algorithm requires an initial value (seed) for the parameters. A reasonable guess for the model parameters is

$$\gamma_i = \frac{n}{M+1} \qquad \text{for } 1 \le i \le n-1,$$

and

$$\lambda_i = \frac{\sum_{x=0}^{M} \hat{b}_x}{M+1}, \quad \mu_i = \frac{\sum_{x=0}^{M} \hat{d}_x}{M+1}, \qquad \text{for } 1 \le i \le n.$$

To minimise the risk of convergence to local extrema, we start with 25 different seeds obtained by adding independent random noises to each of the above values, and we choose the optimal solution among all.

5.1. Artificial examples

We consider three examples of TMAP*s with $n = 3$ or $n = 4$ phases, and we apply the two parameter estimation methods on data samples constructed by simulating trajectories of these models. As expected, the fits corresponding to the individual demographic data are much closer to the real model, and are associated to smaller empirical percentile intervals, than those corresponding to the global population data.

We observe in Figure S7 that the real value of n never minimises the MSE on the examples. This indicates that the MSE is not a satisfactory criterion to determine the optimal number of phases in our model. Finding this optimal value based solely on global population data remains an open problem. In contrast, based on individual demographic data, we see in

Figure S8 that in all examples the AIC provides the correct answer most of the time, while the CV and MSIL show similar trends and slightly underestimate the true value of n. The parameters L and M in the MSIL were chosen according to (4.10) and the criterion turned out not to be sensitive to this choice as the optimal value of n is the same for neighbouring values of L and M. Note that for a specific simulation it may happen that the AIC returns the correct value of n while the CV (or the MSIL) returns a wrong value, and vice versa. There may therefore be discrepancies between the optimal values provided by the different criteria. However, based on our analysis, it is fair to conclude that the AIC is a satisfactory criterion to determine the optimal number of phases in our model. All the details and figures can be found in Supplementary material.

5.2. Female families in different countries

In Hautphenne and Latouche [9], real global population data on female human mortality and fertility rates corresponding to five-year age-classes from different countries are used to fit MBTs with 22 phases. In contrast to the present model, in that paper a one-to-one correspondence between the phases and the age-classes is assumed, so the age-specific fertility and mortality rates are directly used as the entries of the matrix D_1 and the vector $\boldsymbol{d}$.

Here we use the weighted non-linear regression method described in Section 3.2 to estimate the parameters of new MBTs with 22 phases, and we compare the age-specific mortality and fertility curves resulting from these new models with those resulting from the models in [9]. Besides facilitating the comparison, considering the same number of phases as in [9] allows us to start the optimisation algorithm with the most realistic initial parameter values given by the model values in that paper. We show the results for a supercritical country (Congo), an almost-critical country (USA) and a subcritical country (Japan) in Figure 3. We see that the new fits have substantially improved: the MSE is divided by a factor of 5.21 for Congo, 1.89 for the USA and 1.38 for Japan. We observe that the fits of the mortality curves are less satisfactory for the older ages; note that removing the weights (which are decreasing with age) does not improve the fits.

5.3. Black Robin population

The black robin is an endangered songbird species endemic to the Chatham Islands, an isolated archipelago located 800km East of New Zealand, see Figure 4. By 1980, the population of black robins had declined to five

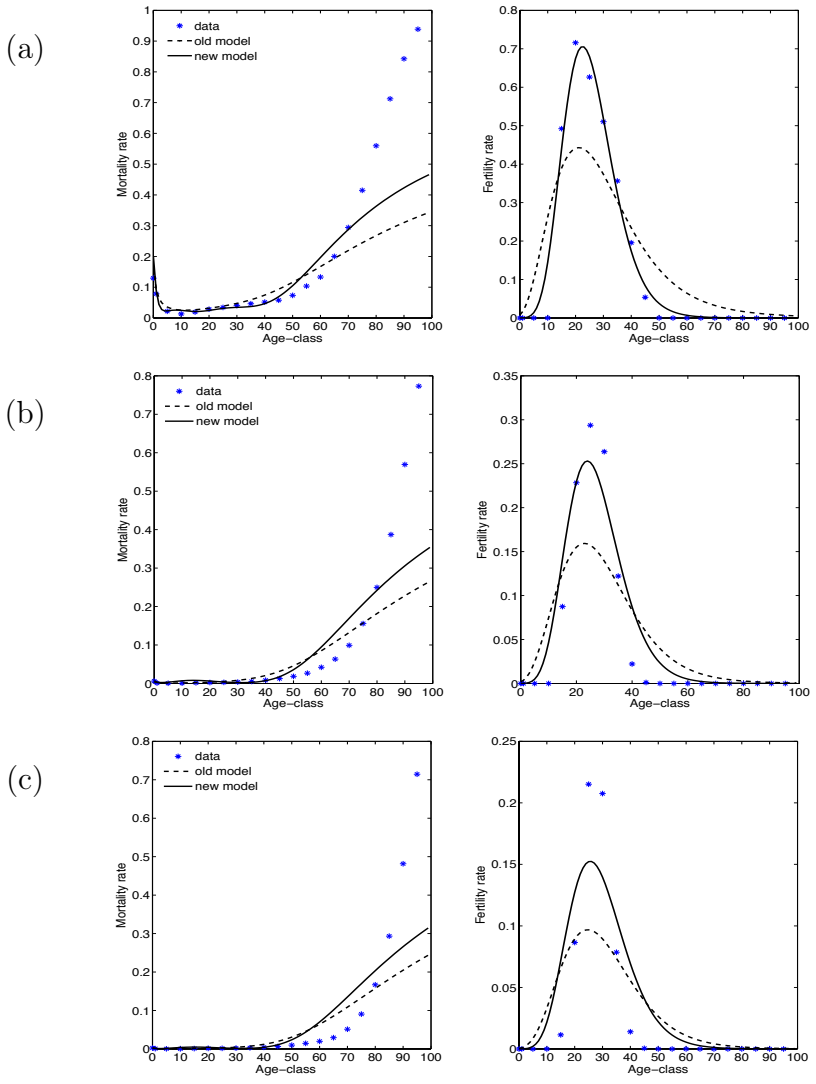

Figure 3: **Female families.** Optimal model with $n = 22$ phases estimated using global population data (new model) corresponding to Congo (a), USA (b) and Japan (c), compared to the model used in [9] (old model).

birds, including only a single successful breeding pair, on Mangere Island [5]. Through intensive conservation efforts in 1980–1989 by the New Zealand Wildlife Service (now the Department of Conservation), the population recovered to 93 birds by spring 1990 [17]. Over the next decade (1990–1998), the population was closely monitored, but without human intervention. Nev-

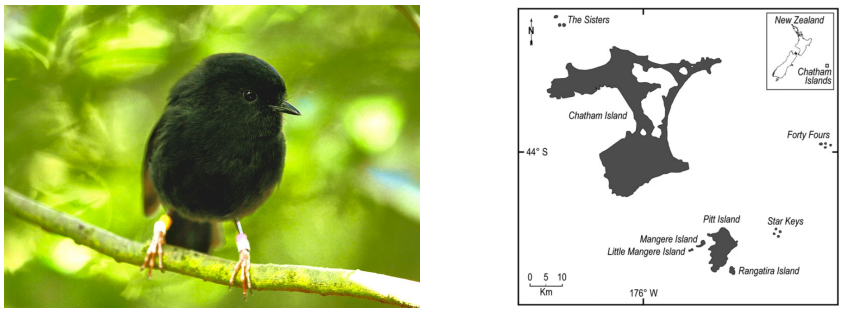

Figure 4: Left: The endangered black robin *Petroica traversi* (photo by Melanie Massaro). Right: The map of the Chatham Islands, New Zealand (image from [25]). Nowadays, black robins only occur on Mangere and Rangatira Islands.

ertheless the population continued to grow rapidly to 197 adults by 1998, but after this period, the population growth slowed considerably and it only reached 239 adults in 2011 [25] and 298 in 2014 [23].

For the conservation management of highly threatened species it is important to know the potential future viability, or survival probability, of a population, because if a population is not viable (i.e. fertility and survival rates are low), it will eventually become extinct. Population viability depends on reproductive rates and survival of individuals, but these rates may vary between sexes and across an individual's life span (with age). Hence, the exact male-to-female ratio and the ages of each individual within a population will influence a population's future viability. Reintroduction of a species into previously occupied parts of its former natural range is nowadays a common hands-on conservation method. In order to maximise the survival chances of the new population, it is necessary to know the optimal age distribution of the reintroduced population, which can only be designed based on a complete age-specific demographic analysis of the species. The black robin is an ideal species for which to develop these new statistical tools, because biologists have been collecting complete raw datasets on this bird species for several decades. An age-specific reorganisation of these data leads to a total of 433 life vectors corresponding to female individuals and containing female offspring counts for the monitoring period 2007–2014.

We performed different tests to determine the optimal number of phases to fit the black robin female data. The results are shown in Figure 5, where we see that the optimal value is $n = 8$ according to the AIC, $n = 13$ according to the CV criterion, and $n = 7$ according to the MSIL criterion

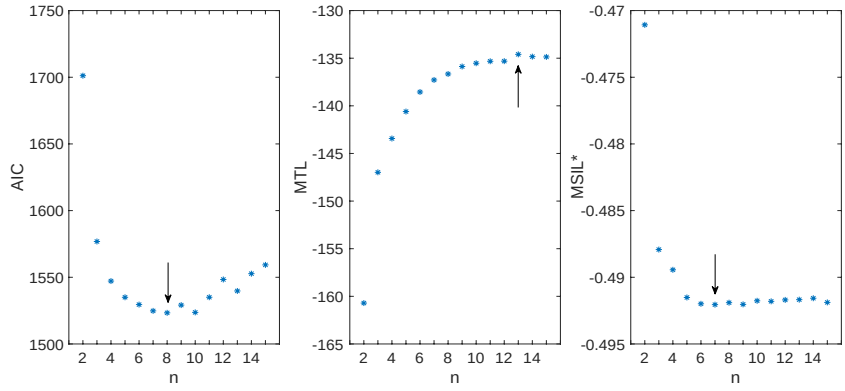

Figure 5: **Black robins.** Left: Result of the Akaike Information Criterion (AIC): the optimal number of phases is $n = 8$; Middle: Result of the cross-validation method, showing the mean test log-likelihood (MTL): the optimal number of phases is $n = 13$; Right: Result of the mean squared integrated loss (MSIL) criterion with $M = 3$ and $L = 1$: the optimal number of phases is $n = 7$.

Table 1: **Black robins.** The optimal number of phases according to the mean squared integrated loss (MSIL) criterion for different values of truncation values M and L ($M = 3$ and $L = 1$ correspond to the values obtained using (4.10)).

$M \setminus L$	0	1	2	3
2	5	9	13	13
3	5	**7**	14	14
4	5	15	14	14

with $M = 3$ and $L = 1$ (determined using (4.10)). In this case, the MSIL criterion is quite sensitive to the choice of M and L, as indicated in Table 1.

The model fits based on the female global population data and on the female individual demographic data (life vectors) with $n = 8$ are compared in Figure 6. Since individual demographic data are available, the model based on these data is the most informative. Black robins reach maturity at 1 or 2-years of age. The age-specific mortality curves obtained from the estimated model show that mortality of black robins is the highest before they reach maturity and the lowest when they are 1 to 2 years old. Once birds reach 3 years of age, mortality rates do not increase dramatically with age, nor do fertility rates decline, which would support the hypothesis that there is no senescence. However, as few birds reach the old ages, the accuracy of the estimates obtained using global population data declines with age. Figure 7

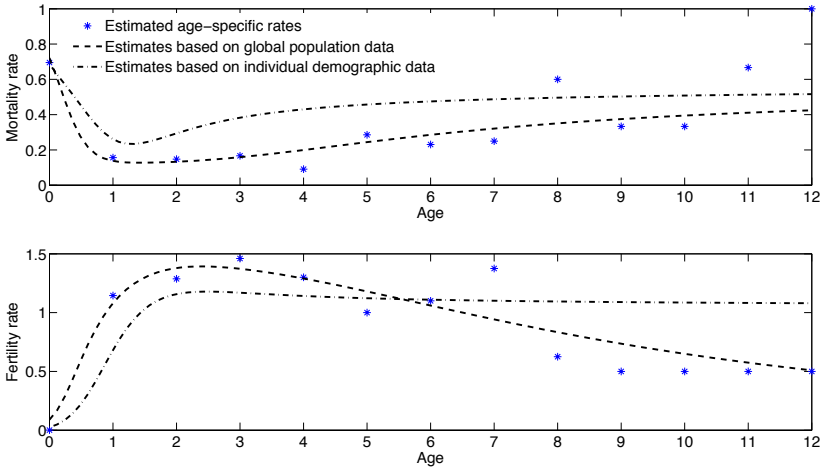

Figure 6: **Black robins.** Age-specific mortality and fertility curves for the model with $n = 8$ phases, whose parameters are estimated based on global population data and on individual demographic data.

shows the 95% bootstrap confidence intervals for the estimators of the model outputs, obtained by bootstrapping 25 samples from the original sample of life vectors. We see that the intervals are quite narrow, especially for the estimators obtained using the individual demographic data, and become wider as age increases.

One of the most informative model output is the probability of extinction of a female family (that is, consisting only of female descendants) generated by a single female, as a function of the age of that first female. This probability can be computed from the MBT model using any of the available algorithms (see for instance [10, 12]) and is shown in Figure 8. The curve highlights the combined effect of the age-specific mortality and fertility rates on the viability of the female family, hence of the whole population, by extension. We see that, if we were to found a new population starting with a single female bird, in order to maximise the survival chance of the population, the optimal age of the initial female should be the age minimising the probability of extinction; in this case it is 1.52 years. A 95% bootstrap confidence interval for the optimal age, obtained by bootstrapping 25 samples from the original sample of life vectors, is $[1.34, 1.86]$.

Other relevant outputs of the model include the asymptotic age frequency in the black robin population, that is, the proportion of birds in the different

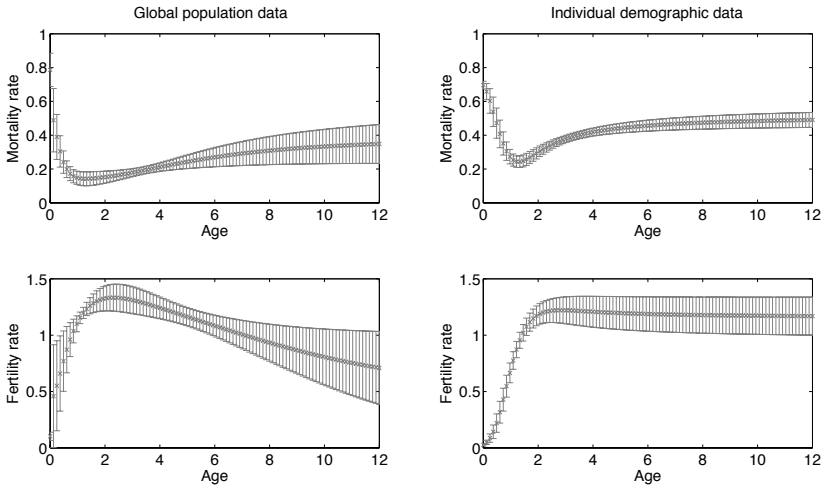

Figure 7: **Black robins.** Mean and 95% pointwise bootstrap confidence intervals of the model fits corresponding to 25 bootstrapped datasets generated from the individual dataset containing $N = 433$ life vectors, for $n = 8$ phases, using global population data (right) and individual demographic data (left).

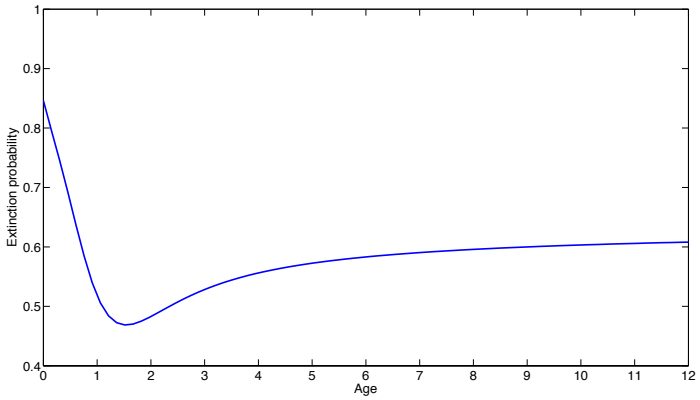

Figure 8: **Black robins.** The extinction probability of a female population as a function of the age of the initial female computed from the model with $n = 8$ phases estimated using individual demographic data.

age classes if we let the population evolve for a long period of time with the same demographic rates. This is studied in detail in a parallel paper [13], which uses an MBT model with $n = 8$ phases, whose parameters are

estimated by the method developed in Section 3.

6. Conclusion and future directions

We considered a two-dimensional continuous-time Markov chain, the *transient Markovian arrival process* (TMAP), as a model for the lifetime and reproduction of individuals in a population. For inference and tractability purposes, we made some assumptions on the structure of the transition rate matrices associated with this process (we call the simpler process a TMAP*). TMAPs allow us to define a generalisation of linear birth-and-death processes called *Markovian binary trees* (MBTs), in which birth and death rates are not constant over time, and dependences may exist between inter-reproduction time intervals and between the lifetimes of a parent and its children. MBTs have a high level of flexibility and are amenable to efficient parameter estimation. Here, we estimated the parameters of the TMAP* based on two types of datasets containing information about age-specific fertility and mortality rates, which we referred to as *global population data* and *individual population data*. We also suggested several criteria to determine the optimal number of phases (or states) in our Markovian model. We carried out a performance analysis and comparison of the proposed estimation methods and criteria, and we applied our techniques to human demography and conservation biology using real datasets.

Our analysis suggests a number of directions for future research, particularly for the study of global population data, which still lacks a satisfactory method to determine the optimal number of phases. Cross-validation is delicate in that case as there are generally few data points. Leave-one-out cross-validation could be used, where we leave one age-class out; possible complications would then include the choice of the weights $\hat{S}_x$ as these explicitly use the death rates for all age-classes.

Further methods for analysing global population data could be developed when the age-specific mortality and fertility rates follow some specific distributions. For example, we may know that the birth rate and death rate at each age lie in an exponential family of distributions. The analysis could then involve a parallel process of estimating these distributions and using them to simulate new samples of global population data. Each new sample can then be used to do parameter estimation and construct confidence intervals. Alternatively, the theory developed for finding confidence intervals for weighted non-linear regression methods could be used in this case; unfortunately, this process is not computationally straightforward.

Appendix A. Proof of Proposition 3.1

To ease the notation, we drop the dependence in the parameters $\boldsymbol{\theta}$ in the model outputs in this Appendix and Appendix B.

Let L be the lifetime of an individual and $\bar{S}(x) = \mathbb{P}(L > x)$ be the survival function. Since L follows a phase-type distribution $\mathrm{PH}(\boldsymbol{\alpha}^\top, D)$, we have $\bar{S}(x) = 1 - \mathbb{P}(L \leq x) = \boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}$. The probability of death at age x, $\bar{d}(x)$, can thus be calculated as

$$\begin{aligned}\bar{d}(x) &= \mathbb{P}(x < L \leq x+1 | L > x) = \frac{\mathbb{P}(L > x) - \mathbb{P}(L > x+1)}{\mathbb{P}(L > x)} \\ &= \frac{\bar{S}(x) - \bar{S}(x+1)}{\bar{S}(x)} = \frac{\boldsymbol{\alpha}^\top e^{Dx}(I - e^D)\mathbf{1}}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}.\end{aligned}$$

It is shown in [21] that the mean number of events until time t in a TMAP started in phase i at time 0 is given by $\mathbb{E}(X(t)|\varphi(0) = i) = [(I - e^{Dt})(-D)^{-1}D_1\mathbf{1}]_i$. Let $X([x, x+t))$ denote the number of events in the TMAP in the time interval $[x, x+t)$. By time-homogeneity of the TMAP, we have $\mathbb{E}(X([x, x+t))|\varphi(x) = i) = \mathbb{E}(X(t)|\varphi(0) = i)$. The mean number of offspring generated by an individual at age x is thus

$$\begin{aligned}\bar{b}(x) &= \mathbb{E}(X([x, x+1))|L > x) \\ &= \sum_{1 \leq i \leq n} \mathbb{P}(\varphi(x) = i | L > x)\mathbb{E}(X([x, x+1))|L > x, \varphi(x) = i) \\ &= \sum_{1 \leq i \leq n} \frac{\mathbb{P}(\varphi(x) = i, L > x)}{\mathbb{P}(L > x)}\mathbb{E}(X(1)|\varphi(0) = i) \\ &= \sum_{1 \leq i \leq n} \frac{[\boldsymbol{\alpha}^\top e^{Dx}]_i}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}[(I - e^D)(-D)^{-1}D_1\mathbf{1}]_i \\ &= \frac{\boldsymbol{\alpha}^\top e^{Dx}(I - e^D)(-D)^{-1}D_1\mathbf{1}}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}.\end{aligned}$$

This completes the proof. □

Appendix B. Global population data with $\ell-$year age classes

In demography, the available data often consist of age-specific fertility and mortality rates over $\ell-$year age-classes with $\ell > 1$, that is,

- the expected number of offspring *per year* from a parent in age-class $[x, x+\ell)$, denoted by $\hat{\beta}_{[x,x+\ell)}$, and
- the probability that an individual who reached the age-class $[x, x+\ell)$ dies *within the year*, denoted by $\hat{\mu}_{[x,x+\ell)}$.

We extend the definition of $\bar{d}(x)$ and $\bar{b}(x)$ given in Section 3.1 to $\ell-$year age-classes, and we define the functions $\bar{d}(x,\ell)$ and $\bar{b}(x,\ell)$, computed from the TMAP model, as follows:

$$\begin{aligned}\bar{d}(x,\ell) &= \mathbb{P}(x < L \leq x+\ell | L > x)\\ \bar{b}(x,\ell) &= \mathbb{E}(X([x,x+\ell))|L > x).\end{aligned}$$

It is a simple matter to generalise Proposition 3.1 to obtain

Corollary Appendix B.1. *The age-specific mortality and fertility rates $\bar{d}(x,\ell)$ and $\bar{b}(x,\ell)$ are given by*

$$\begin{aligned}\bar{d}(x,\ell) &= \frac{\boldsymbol{\alpha}^\top e^{Dx}(I-e^{D\ell})\mathbf{1}}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}\\ \bar{b}(x,\ell) &= \frac{\boldsymbol{\alpha}^\top e^{Dx}(I-e^{D\ell})(-D)^{-1}D_1\mathbf{1}}{\boldsymbol{\alpha}^\top e^{Dx}\mathbf{1}}.\end{aligned}$$

The correspondence between $\bar{d}(x,\ell)$ and $\bar{b}(x,\ell)$ and the mortality and fertility rates in age-class $[x, x+\ell)$ is given in the next Lemma.

Lemma Appendix B.1. *The age-specific mortality and fertility rates $\bar{d}(x,\ell)$ and $\bar{b}(x,\ell)$ satisfy*

$$\begin{aligned}\bar{d}(x,\ell) &\equiv 1-(1-\hat{\mu}_{[x,x+\ell)})^\ell\\ \bar{b}(x,\ell) &\equiv \hat{\beta}_{[x,x+\ell)}\,\frac{1-(1-\hat{\mu}_{[x,x+\ell)})^\ell}{\hat{\mu}_{[x,x+\ell)}},\end{aligned}$$

where the symbol $\equiv$ has to be interpreted as "is the model equivalent of".

Proof. The model function $1-\bar{d}(x,\ell)$ is the probability that an individual who reached age x survives at least until age $x+\ell$, that is, survives ℓ successive one-year age intervals, which occurs with probability $(1-\hat{\mu}_{[x,x+\ell)})^\ell$.

The model function $\bar{b}(x,\ell)$ can be rewritten as

$$\begin{aligned}\bar{b}(x,\ell) &= \mathbb{E}(X([x,x+1))|L>x) + \mathbb{E}(X([x+1,x+2))|L>x) + \dots \\ &\quad +\mathbb{E}(X([x+\ell-1,x+\ell))|L>x) \\ &\equiv \hat{\beta}_{[x,x+\ell)} + (1-\hat{\mu}_{[x,x+\ell)})\hat{\beta}_{[x,x+\ell)} + \dots + (1-\hat{\mu}_{[x,x+\ell)})^{\ell-1}\hat{\beta}_{[x,x+\ell)} \\ &= \hat{\beta}_{[x,x+\ell)}\,\frac{1-(1-\hat{\mu}_{[x,x+\ell)})^{\ell}}{\hat{\mu}_{[x,x+\ell)}},\end{aligned}$$

which completes the proof. □

In order to estimate the model parameters in this case, the objective function given by Equation (3.1) then needs to be modified according to Corollary Appendix B.1 and Lemma Appendix B.1.

Appendix C. Proof of Proposition 4.1

In order to compute the matrix $P(k)$, we compute the matrix $P(k,t)$ for $0 \le k \le K$ and for any $t \ge 0$, where

$$P_{ij}(k,t) := \mathbb{P}(X(t)=k, \varphi(t)=j | X(0)=0, \varphi(0)=i),$$

and observe that $P(k) = P(k,\ell)$. It is well known from the theory of Markovian arrival processes that the probability generating function $P^*(z,t) := \sum_{k\ge 0} P(k,t)z^k$, is given by the matrix exponential

$$P^*(z,t) = \exp[D(z)t], \quad \text{where } D(z) := D_0 + z\,D_1.$$

Since $P(k,t) = (1/k!)[\partial^k P^*(z,t)/(\partial z)^k]\big|_{z=0}$ for any $k \ge 0$, we need to take the derivatives of the matrix exponential $\exp[D(z)t]$ with respect to z. The scalar rule of exponential differentiation only holds here if D_0 and D_1 commute, which is generally not the case. Instead, we first differentiate $P^*(z,t)$ with respect to t,

$$\partial P^*(z,t)/\partial t = D(z)P^*(z,t),$$

and we then take successive derivatives of this equation with respect to z:

$$\begin{aligned}
\partial^2 P^*(z,t)/(\partial t)(\partial z) &= D_1 P^*(z,t) + D(z)\partial P^*(z,t)/\partial z \\
\partial^3 P^*(z,t)/(\partial t)(\partial z)^2 &= 2D_1 \partial P^*(z,t)/\partial z + D(z)\partial^2 P^*(z,t)/(\partial z)^2 \\
&\vdots \\
\partial^{(K+1)} P^*(z,t)/(\partial t)(\partial z)^K &= KD_1 \partial^{(K-1)} P^*(z,t)/(\partial z)^{(K-1)} \\
&\quad + D(z)\partial^K P^*(z,t)/(\partial z)^K.
\end{aligned}$$

This system of partial derivative equations can be rewritten as an ordinary differential equation for the unknown matrix containing the partial derivatives of $P^*(z,t)$ with respect to z,

$$\frac{d}{dt}\begin{bmatrix} P^*(z,t) \\ \partial P^*(z,t)/\partial z \\ \partial^2 P^*(z,t)/(\partial z)^2 \\ \vdots \\ \partial^K P^*(z,t)/(\partial z)^K \end{bmatrix} = \begin{bmatrix} D(z) & & & & \\ D_1 & D(z) & & & \\ & 2D_1 & D(z) & & \\ & & \ddots & & \\ & & & KD_1 & D(z) \end{bmatrix} \begin{bmatrix} P^*(z,t) \\ \partial P^*(z,t)/\partial z \\ \partial^2 P^*(z,t)/(\partial z)^2 \\ \vdots \\ \partial^K P^*(z,t)/(\partial z)^K \end{bmatrix},$$

whose solution is

$$\begin{bmatrix} P^*(z,t) \\ \partial P^*(z,t)/\partial z \\ \partial^2 P^*(z,t)/(\partial z)^2 \\ \vdots \\ \partial^K P^*(z,t)/(\partial z)^K \end{bmatrix} = \exp\left(\begin{bmatrix} D(z) & & & & \\ D_1 & D(z) & & & \\ & 2D_1 & D(z) & & \\ & & \ddots & & \\ & & & KD_1 & D(z) \end{bmatrix} t\right) \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Taking $z = 0$ and denoting

$$\mathcal{M} = \begin{bmatrix} D_0 & & & & \\ D_1 & D_0 & & & \\ & 2D_1 & D_0 & & \\ & & \ddots & & \\ & & & KD_1 & D_0 \end{bmatrix},$$

we obtain $P(k) = P(k,\ell) = (1/k!)(\boldsymbol{e}_k^\top \otimes I)\exp(\mathcal{M}\ell)(\boldsymbol{e}_1 \otimes I)$.

Then, $\boldsymbol{p}(k)$ is obtained by conditioning on the time $u \in [0,\ell]$ when the

individual dies,

$$\begin{aligned}\boldsymbol{p}(k) &= \int_0^\ell P(k,u)\,\boldsymbol{d}\,du \\ &= (1/k!)(\boldsymbol{e}_k^\top \otimes I)\int_0^\ell \exp(\mathcal{M}u)\,du\,(\boldsymbol{e}_1 \otimes I)\,\boldsymbol{d} \\ &= (1/k!)(\boldsymbol{e}_k^\top \otimes I)[I-\exp(\mathcal{M}\ell)](-\mathcal{M})^{-1}\,(\boldsymbol{e}_1 \otimes I)\,\boldsymbol{d}.\end{aligned}$$

Next, $P = P^*(1,\ell) = \exp(D\ell)$, where $D = D_0 + D_1$, and finally

$$\boldsymbol{p} = \int_0^\ell \exp(Du)\boldsymbol{d}\,du = [I-\exp(D\ell)](-D)^{-1}\boldsymbol{d}.$$

This completes the proof. □

Supplementary material

Additional information for this article is available online in Supplementary material, referenced in Sections 3.3, 4.2, 4.3, and 5.

Acknowledgements

We thank the referee and associate editor for constructive comments which helped us to improve the paper. Sophie Hautphenne thanks the Australian Research Council (ARC) for support through the Discovery Early Career Researcher Award DE150101044. The black robin research was funded by the New Zealand Foundation for Research, Science and Technology (UOCX0601) to Melanie Massaro, by the School of Biological Sciences, University of Canterbury, by the Brian Mason Scientific and Technical Trust, and by the Mohamed bin Zayed Species Conservation Fund. This research was only possible with permission from the Chatham Island Conservation Board and the logistic help of the Department of Conservation.

References

[1] D. Aldous. Probability distributions on cladograms. In *Random Discrete Structures*, pages 1–18. Springer, 1996.

[2] D. J. Aldous. Stochastic models and descriptive statistics for phylogenetic trees, from yule to today. *Statistical Science*, 16(1):23–34, 2001.

[3] S. Asmussen and G. Koole. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, 30(02):365–372, 1993.

[4] M.G.B. Blum and O. François. Which random processes describe the tree of life? a large-scale study of phylogenetic tree imbalance. *Systematic Biology*, 55(4):685–691, 2006.

[5] D. Butler and D. Merton. *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press, Auckland, 1993.

[6] A.C. Davison and N.I. Ramesh. Some models for discretized series of events. *Journal of the American Statistical Association*, 91(434):601–609, 1996.

[7] T.B. Gage and C.J. Mode. Some laws of mortality: how well do they fit? *Human Biology*, pages 445–461, 1993.

[8] S. Hautphenne and M. Fackrell. An EM algorithm for the model fitting of Markovian binary trees. *Computational Statistics & Data Analysis*, 70:19–34, 2014.

[9] S. Hautphenne and G. Latouche. The Markovian binary tree applied to demography. *Journal of Mathematical Biology*, 64(7):1109–1135, 2012.

[10] S. Hautphenne, G. Latouche, and M.-A. Remiche. Newton's iteration for the extinction probability of a Markovian binary tree. *Linear Algebra and its Applications*, 428(11-12):2791–2804, 2008.

[11] S. Hautphenne, G. Latouche, and M.-A. Remiche. Transient features for Markovian binary trees. In *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, page 18. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.

[12] S. Hautphenne, G. Latouche, and M.-A. Remiche. Algorithmic approach to the extinction probability of branching processes. *Methodology and Computing in Applied Probability*, 13(1):171–192, 2011.

[13] S. Hautphenne, M. Massaro, and P. Taylor. How old is this bird? The age distribution under some phase sampling schemes. *Journal of Mathematical Biology*, 75(6-7):1319–1347, 2017.

[14] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[15] G. Jones. Calculations for multi-type age-dependent binary branching processes. *Journal of Mathematical Biology*, 63(1):33–56, 2011.

[16] N. Keiding. Maximum likelihood estimation in the birth-and-death process. *The Annals of Statistics*, 3(2):363–372, 1975.

[17] E.S. Kennedy, C.E. Grueber, R.P. Duncan, and I. G. Jamieson. Severe inbreeding depression and no evidence of purging in an extremely inbred wild species – the Chatham Island black robin. *Evolution*, 68(4):987–995, 2014.

[18] N. Kontoleon. *The Markovian binary tree: a model of the macroevolutionary process*. PhD thesis, 2006.

[19] A. Lambert, H.K. Alexander, and T. Stadler. Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. *Journal of Theoretical Biology*, 352:60–70, 2014.

[20] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. SIAM, 1999.

[21] G. Latouche, M.-A. Remiche, and P.G. Taylor. Transient Markov arrival processes. *Annals of Applied Probability*, pages 628–640, 2003.

[22] X. S. Lin and X. Liu. Markov aging process and phase-type law of mortality. *North American Actuarial Journal*, 11(4):92–109, 2007.

[23] M. Massaro, A. Chick, E.S. Kennedy, and R. Whitsed. Post-reintroduction distribution and habitat preferences of a spatially limited island bird species. *Animal Conservation*, 21(1):54–64, 2018.

[24] M. Massaro, R. Sainudiin, D. Merton, J.V. Briskie, A.M. Poole, and M.L. Hale. Human-assisted spread of a maladaptive behavior in a critically endangered bird. *PloS One*, 8(12):e79066, 2013.

[25] M. Massaro, M. Stanbury, and J.V. Briskie. Nest site selection by the endangered black robin increases vulnerability to predation by an invasive bird. *Animal Conservation*, 16(4):404–411, 2013.

[26] P. Peristera and A. Kostaki. Modeling fertility in modern populations. *Demographic Research*, 16(6):141–194, 2007.

[27] N.I. Ramesh. Statistical analysis on Markov-modulated Poisson processes. *Environmetrics*, 6(2):165–179, 1995.

[28] J. F. Reynolds. On estimating the parameters of a birth-death process. *Australian Journal of Statistics*, 15(1):35–43, 1973.

[29] T. Rydén. Parameter estimation for Markov modulated Poisson processes. *Stochastic Models*, 10(4):795–829, 1994.

[30] T. Stadler. Recovering speciation and extinction dynamics based on phylogenies. *Journal of Evolutionary Biology*, 26(6):1203–1219, 2013.