Appendix for *Robustly estimating the demographic contribution of immigration: simulation, sensitivity analysis and seals*

Murray Christian[1,2,*], W. Chris Oosthuizen[2], Marthán N. Bester[1], and P.J. Nico de Bruyn[1]

[1] Department of Zoology and Entomology, Mammal Research Institute, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

[2] Centre for Statistics in Ecology, the Environment and Conservation, Department of Statistical Sciences, University of Cape Town, Cape Town 7701, South Africa

[*]Corresponding author: murraychristian@live.co.za

Contents

1 Multi-event model

1.1 Transition and emission matrices

Transition matrices

The transition matrices of the multi-event model are stochastic matrices whose $(i,j)^{\text{th}}$ entry is the probability that an individual in state i at time t will be in state j at time $t+1$ (thus, source states in rows, destination states in columns). Without considering tag loss there are eight states: pre-breeders aged $0, 1, 2, 3, 4$, breeders, non-breeders and 'dead'. The transition matrices are time-dependent, but we suppress this in the notation:

$$\begin{array}{r} \\ Pb_0 \\ Pb_1 \\ Pb_2 \\ Pb_3 \\ Pb_4 \\ Br \\ Nb \\ D \end{array} \begin{array}{c} \begin{array}{cccccccc} Pb_0 & Pb_1 & Pb_2 & Pb_3 & Pb_4 & Br & Nb & D \end{array} \\ \begin{pmatrix} 0 & s_0 & 0 & 0 & 0 & 0 & 0 & 1-s_0 \\ 0 & 0 & s_N & 0 & 0 & 0 & 0 & 1-s_N \\ 0 & 0 & 0 & s_N(1-f_3) & 0 & s_N f_3 & 0 & 1-s_N \\ 0 & 0 & 0 & 0 & s_N(1-f_4) & s_N f_4 & 0 & 1-s_N \\ 0 & 0 & 0 & 0 & 0 & s_N & 0 & 1-s_N \\ 0 & 0 & 0 & 0 & 0 & s_B bb & s_B(1-bb) & 1-s_B \\ 0 & 0 & 0 & 0 & 0 & s_N nb & s_N(1-nb) & 1-s_N \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Incorporating tag loss results in each of the first seven states being duplicated and assigned one or two tags. We write these using superscripts on the original variable to denote the number of tags e.g. Pb_1 becomes $Pb_1^{(1)}$ and $Pb_1^{(2)}$. In addition to deceased and permanently emigrated individuals, the 'dead' state now includes individuals that have lost all tags. The 15×15 transition matrices can be written in terms of block sub-matrices as:

$$\begin{array}{r} \\ \mathbf{Pb_0}^{(1,2)} \\ \mathbf{Pb_1}^{(1,2)} \\ \mathbf{Pb_2}^{(1,2)} \\ \mathbf{Pb_3}^{(1,2)} \\ \mathbf{Pb_4}^{(1,2)} \\ \mathbf{Br}^{(1,2)} \\ \mathbf{Nb}^{(1,2)} \\ D \end{array} \begin{array}{c} \begin{array}{cccccccc} \mathbf{Pb_0}^{(1,2)} & \mathbf{Pb_1}^{(1,2)} & \mathbf{Pb_2}^{(1,2)} & \mathbf{Pb_3}^{(1,2)} & \mathbf{Pb_4}^{(1,2)} & \mathbf{Br}^{(1,2)} & \mathbf{Nb}^{(1,2)} & D \end{array} \\ \begin{pmatrix} \mathbf{0}_{2\times2} & s_0 \cdot \mathbf{tl} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{d}(s_0) \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & s_N \cdot \mathbf{tl} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{d}(s_N) \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & s_N(1-f_3) \cdot \mathbf{tl} & \mathbf{0}_{2\times2} & s_N f_3 \cdot \mathbf{tl} & \mathbf{0}_{2\times2} & \mathbf{d}(s_N) \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & s_N(1-f_4) \cdot \mathbf{tl} & s_N f_4 \cdot \mathbf{tl} & \mathbf{0}_{2\times2} & \mathbf{d}(s_N) \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & s_N \cdot \mathbf{tl} & \mathbf{0}_{2\times2} & \mathbf{d}(s_N) \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & s_B bb \cdot \mathbf{tl} & s_B(1-bb) \cdot \mathbf{tl} & \mathbf{d}(s_B) \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & s_N nb \cdot \mathbf{tl} & s_N(1-nb) \cdot \mathbf{tl} & \mathbf{d}(s_N) \\ \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & 1 \end{pmatrix} \end{array},$$

where $\mathbf{0}_{2\times2}$ is the two-by-two zero matrix, $\mathbf{0}_{1\times2}$ is a zero row of length two, $\mathbf{d}(s_0), \mathbf{d}(s_N), \mathbf{d}(s_B)$ are 2×1 column vectors

$$\mathbf{d}(s_0) = \begin{pmatrix} 1-s_0(1-\tau) \\ 1-s_0 \end{pmatrix}, \quad \mathbf{d}(s_N) = \begin{pmatrix} 1-s_N(1-\tau) \\ 1-s_N \end{pmatrix}, \quad \mathbf{d}(s_B) = \begin{pmatrix} 1-s_B(1-\tau) \\ 1-s_B \end{pmatrix},$$

and **tl** is a matrix of tag-state transition probabilities:

$$\begin{array}{c} \\ \text{1 tag} \\ \text{2 tags} \end{array} \begin{array}{c} \begin{array}{cc} \text{1 tag} & \text{2 tags} \end{array} \\ \begin{pmatrix} 1-\tau & 0 \\ \tau & 1-\tau \end{pmatrix} \end{array}. \tag{1}$$

Thus, for example, the 2×2 sub-matrix corresponding to rows indexed by $\mathbf{Pb_2}^{(1,2)}$ and columns indexed by $\mathbf{Pb_3}^{(1,2)}$ is

$$\begin{array}{c} \\ Pb_2^{(1)} \\ Pb_2^{(2)} \end{array} \begin{array}{c} \begin{array}{cc} Pb_3^{(1)} & Pb_3^{(2)} \end{array} \\ \begin{pmatrix} s_N(1-f_3)(1-\tau) & 0 \\ s_N(1-f_3)\tau & s_N(1-f_3)(1-\tau) \end{pmatrix} \end{array}.$$

Tag loss probabilities

We assume that the probability of losing exactly one tag from one year to the next (τ) does not vary in time, but does depend on a factor for the tag location. From $1983-1999$ tags were placed on the inner inter-digital webbing of the hind flippers, from $2000-2014$ they were placed on the outer inter-digital webbing, and from $2015-2020$ they were moved back to the inner inter-digital webbing. Posterior estimates for tag-loss probabilities from our multi-event model are plotted in Fig. 1.

Emission matrices

The emission matrices of the multi-event model are stochastic matrices whose $(i,j)^{\text{th}}$ entry is the probability that an individual in state i at time t will be observed in observation state j at time t. Without considering the number of tags an individual has, there are eight observation states for each tagged individual in each year. These eight states are defined as the observation (or not) of an individual in the two secondary periods within the breeding season, and outside the breeding season. We denote these three periods with 'O' (for odd weeks of the breeding season), 'E' (for even weeks) and 'M' (for moult; the main haul-out period outside the breeding season). Using a bar to denote the complement of a probability, i.e. $\bar{p} = 1 - p$:

$$\begin{array}{c} \\ Pb_0 \\ Pb_1 \\ Pb_2 \\ Pb_3 \\ Pb_4 \\ Br \\ Nb \\ D \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{MOE} & \text{MO} & \text{ME} & \text{M} & \text{OE} & \text{O} & \text{E} & \text{unobs} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_N & 0 & 0 & 0 & \bar{q}_N \\ 0 & 0 & 0 & q_N & 0 & 0 & 0 & \bar{q}_N \\ 0 & 0 & 0 & q_N & 0 & 0 & 0 & \bar{q}_N \\ 0 & 0 & 0 & q_N & 0 & 0 & 0 & \bar{q}_N \\ q_B p_B^o p_B^e & q_B p_B^o \bar{p}_B^e & q_B \bar{p}_B^o p_B^e & q_B \bar{p}_B^o \bar{p}_B^e & \bar{q}_B p_B^o p_B^e & \bar{q}_B p_B^o \bar{p}_B^e & \bar{q}_B \bar{p}_B^o p_B^e & \bar{q}_B \bar{p}_B^o \bar{p}_B^e \\ 0 & 0 & 0 & q_N & 0 & 0 & 0 & \bar{q}_N \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

We assume that the number of tags on a detected individual is always observed correctly and that it does not vary between the three detection periods of the year. Therefore the full emission matrix

including tag number is

$$
\begin{array}{r}
\\
\mathbf{Pb_0}^{(1,2)} \\ \mathbf{Pb_1}^{(1,2)} \\ \mathbf{Pb_2}^{(1,2)} \\ \mathbf{Pb_3}^{(1,2)} \\ \mathbf{Pb_4}^{(1,2)} \\ \mathbf{Br}^{(1,2)} \\ \mathbf{Nb}^{(1,2)} \\ D
\end{array}
\begin{array}{cccccccc}
\text{MOE}^{(1,2)} & \text{MO}^{(1,2)} & \text{ME}^{(1,2)} & \text{M}^{(1,2)} & \text{OE}^{(1,2)} & \text{O}^{(1,2)} & \text{E}^{(1,2)} & \text{unobs} \\
\end{array}
$$

$$
\begin{pmatrix}
\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 1} \\
\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & q_N \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \bar{q}_N \mathbf{1}_{2\times 1} \\
\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & q_N \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \bar{q}_N \mathbf{1}_{2\times 1} \\
\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & q_N \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \bar{q}_N \mathbf{1}_{2\times 1} \\
\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & q_N \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \bar{q}_N \mathbf{1}_{2\times 1} \\
q_B p_B^o p_B^e \mathbf{I}_{2\times 2} & q_B p_B^o \bar{p}_B^e \mathbf{I}_{2\times 2} & q_B \bar{p}_B^o p_B^e \mathbf{I}_{2\times 2} & q_B \bar{p}_B^o \bar{p}_B^e \mathbf{I}_{2\times 2} & \bar{q}_B p_B^o p_B^e \mathbf{I}_{2\times 2} & \bar{q}_B p_B^o \bar{p}_B^e \mathbf{I}_{2\times 2} & \bar{q}_B \bar{p}_B^o p_B^e \mathbf{I}_{2\times 2} & \bar{q}_B \bar{p}_B^o \bar{p}_B^e \mathbf{1}_{2\times 1} \\
\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & q_N \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \bar{q}_N \mathbf{1}_{2\times 1} \\
\mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & 1
\end{pmatrix},
$$

where $\mathbf{I}_{2\times 2}$ is the identity matrix and $\mathbf{1}_{2\times 1}$ is a column vector of length two whose entries are 1. Thus, for example, the 2×2 sub-matrix corresponding to rows indexed by $\mathbf{Br}^{(1,2)}$ and columns indexed by $MO^{(1,2)}$ is

$$
\begin{array}{c} \\ Br^{(1)} \\ Br^{(2)} \end{array}
\begin{array}{c}
\begin{array}{cc} MO^{(1)} & MO^{(2)} \end{array} \\
\begin{pmatrix} q_B p_B^o \bar{p}_B^e & 0 \\ 0 & q_B p_B^o \bar{p}_B^e \end{pmatrix}.
\end{array}
$$

Initial state probability vectors

All seals in our study were tagged as pups with two tags, or rarely one tag. Therefore the initial state is known, and the initial state probability vectors were

$$
\boldsymbol{\delta} = \begin{cases} (1, 0, 0, \cdots, 0) & \text{for one tag,} \\ (0, 1, 0, \cdots, 0) & \text{for two tags.} \end{cases}
$$

1.2 Priors on vital rates

In this section we give the full specification of the 'Random Effects' and 'Gaussian Process' priors on the vital rates. Throughout, we let ν denote a generic vital rate, one of $s_0, s_N, s_B, f_3, f_4, bb, nb$. We use boldface for the entire vector of values over the study period, $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_T)$.

Random Effects priors

The random effects priors on the vital rates are

$$
\begin{aligned}
\text{logit}(\nu_t) &\sim \text{Normal}\,(\mu_\nu, \sigma_\nu) \quad \text{for all } t, \\
\mu_\nu &\sim \text{Normal}(0, 1.25), \\
\sigma_\nu &\sim \text{Exponential}(1.5).
\end{aligned}
$$

These hyperprior distributions on μ_ν and σ_ν imply approximately uniform priors on the vital rates $\nu_t, t = 1, \cdots, T$. (The commoner choices $\mu_\nu \sim \text{Normal}(0, 1)$ and $\sigma_\nu \sim \text{Exponential}(1)$ would work just as well here too).

Gaussian Process priors

In the random effects prior, the annual estimates ν_t are all drawn from the same distribution, with parameters μ_ν and σ_ν. In contrast, our Gaussian Process priors allow the logit-scale mean and standard

deviation to vary in time:

$$\text{logit}(\nu_t) \sim \text{Normal}\left(\mu_{\nu,t}, \sigma_{\nu,t}\right) \quad \text{for all } t.$$

The time-dependent mean $\boldsymbol{\mu}_\nu = (\mu_{\nu,1}, \cdots, \mu_{\nu,T})$ and standard deviation $\boldsymbol{\sigma}_\nu = (\sigma_{\nu,1}, \cdots, \sigma_{\nu,T})$ are then modelled using Gaussian Processes:

$$\begin{aligned} \boldsymbol{\mu}_\nu &\sim \text{Multivariate Normal}\left(\mathbf{0}, \mathbf{K}_{\boldsymbol{\mu}_\nu}\right), \\ \log\left(\boldsymbol{\sigma}_\nu\right) &\sim \text{Multivariate Normal}\left(\mathbf{0}, \mathbf{K}_{log(\boldsymbol{\sigma}_\nu)}\right). \end{aligned}$$

Thus there are two Guassian process used in this formulation, one for each of the logit-scale mean $\boldsymbol{\mu}_\nu$ and standard deviation $\boldsymbol{\sigma}_\nu$ governing the marginal distributions on the vital rates. This extra flexibility in the standard deviation allows the uncertainty in annual vital rate estimates to increase as information becomes sparser going back towards the start of the study (see Appendix 1.3, Fig. 2).

Each of the covariance matrices $\mathbf{K}_{\boldsymbol{\mu}_\nu}, \mathbf{K}_{log(\boldsymbol{\sigma}_\nu)}$ are determined by exponentiated quadratic kernel functions,

$$\mathbf{K}_{i,j} = \sigma_K^2 \exp\left(-\frac{|t_i - t_j|^2}{2\eta_K^2}\right) \text{ for all } i, j,$$

where σ_K^2 is the marginal variance and η_K is the length-scale parameter (and this pair of parameters differ between the mean process and log standard deviation process). The correlation between components i and j of the random vector ($\boldsymbol{\mu}$ or $\log(\boldsymbol{\sigma})$) governed by such a covariance matrix is

$$\exp\left(-\frac{|t_i - t_j|^2}{2\eta_K^2}\right).$$

Therefore the length-scale η_K controls the rate at which these correlations decay: for small values of η_K the correlations are close to zero (assuming $i \neq j$ hence $t_i \neq t_j$), whilst for large values of η_K the correlations are close to one.

We standardised the times $t = 1, \cdots T$ before defining the covariance matrices, and placed the following priors on the marginal variances and length-scales of each Gaussian Process; for the mean process $\boldsymbol{\mu}_\nu$:

$$\begin{aligned} \sigma_{K_{\boldsymbol{\mu}_\nu}} &\sim \text{Half-Normal}(0, 2), \\ \eta_{K_{\boldsymbol{\mu}_\nu}} &\sim \text{Lognormal}(\log(0.3), 0.2), \end{aligned}$$

and for the log standard deviation process $\log(\boldsymbol{\sigma}_\nu)$:

$$\begin{aligned} \sigma_{K_{\log(\boldsymbol{\sigma}_\nu)}} &\sim \text{Half-Normal}(0, 2), \\ \eta_{K_{\log(\boldsymbol{\sigma}_\nu)}} &\sim \text{Lognormal}(\log(0.5), 0.2). \end{aligned}$$

We have followed Vehtari 2023 in our use of half-normal priors for the marginal standard deviations σ_K and lognormal priors for the lengthscales η_K. It is important that the priors for the η_K preclude

values near zero, as the model cannot distinguish between correlations that decay to zero in less time than the smallest difference in the (standardised) time covariate. The precise parameter values in the lognormal distributions follow Vehtari 2023; their appropriateness relies in the time covariate being standardised. The precise parameter values for the half-normal distribution are chosen so that prior probabilities are likely to be in the region $(-4, 4)$ on the logit scale, which corresponds to roughly $(0.02, 0.98)$ on the probability scale. Prior vs. posterior plots of all Gaussian process hyperparameters are presented in Fig 3.

1.3 Supplementary figures for multi-event models

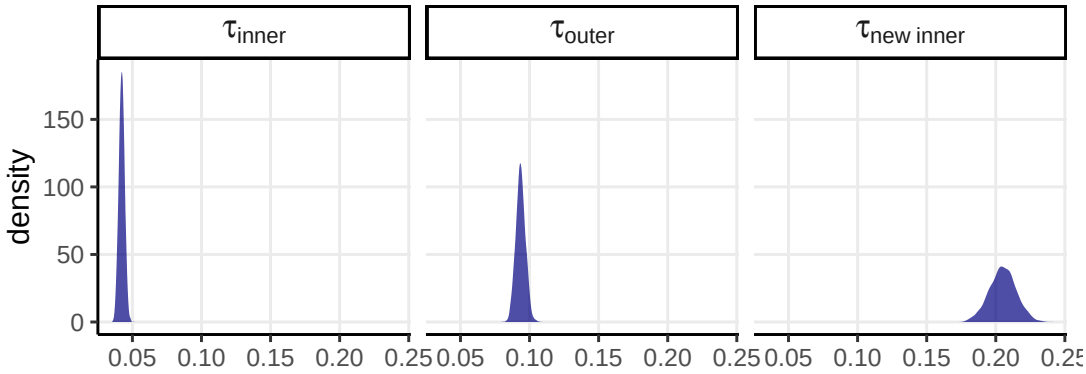


Figure 1: Posterior estimates for tag loss probabilities from the multi-event model with Gaussian Process priors on the vital rates. The tag loss probabilities are: τ_{inner} for tags used $1983 - 1999$, which were placed on the inner inter-digital webbing; τ_{outer} for tags used $2000 - 2014$, which were placed on the outer inter-digital webbing; $\tau_{\text{new inner}}$ for tags used $2015 - 2021$, which were again placed on the inner inter-digital webbing. Uniform$(0, 1)$ priors were used for all tag loss probabilities.

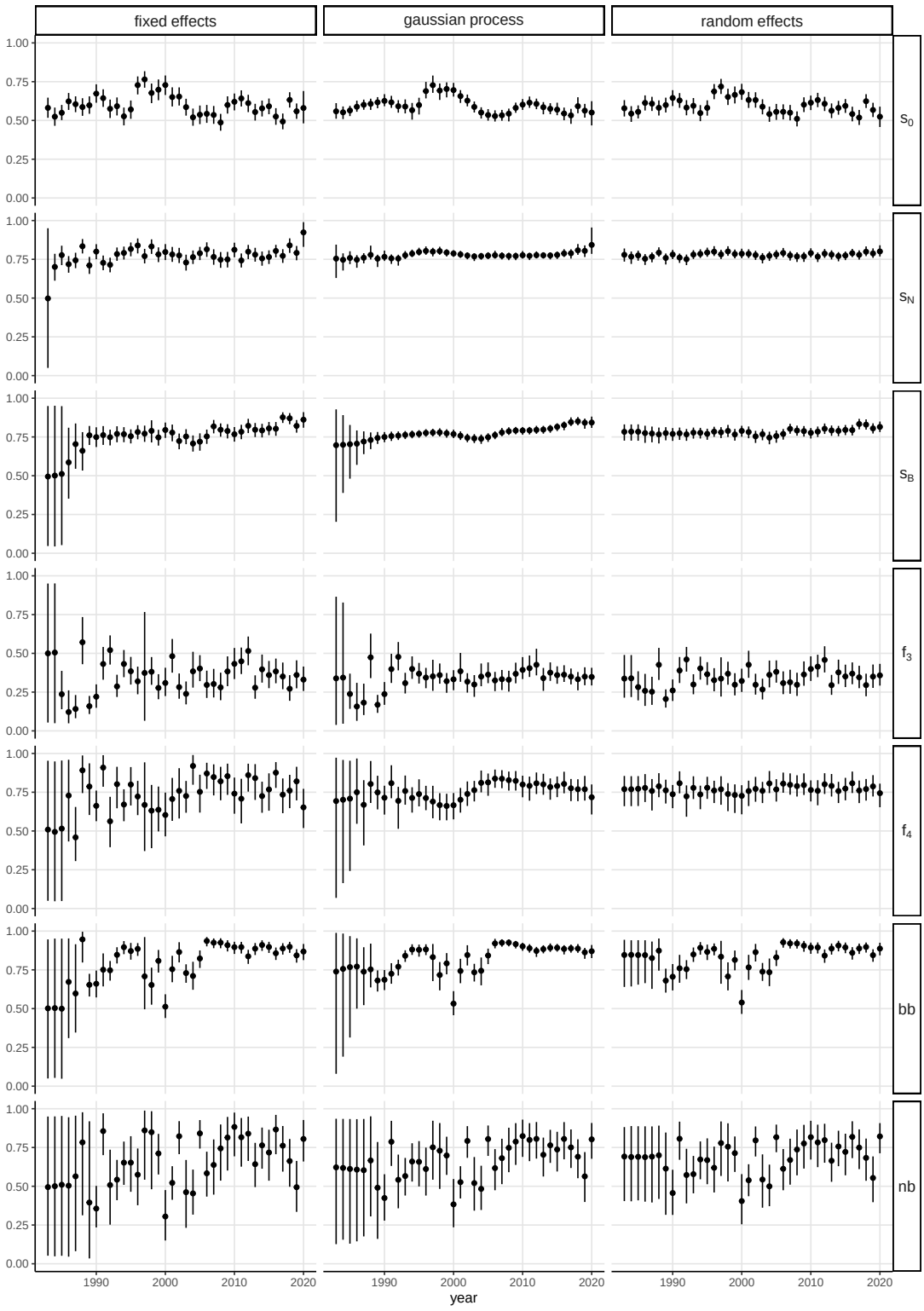


Figure 2: Comparison of vital rate estimates for the multi-event model with 'fixed effects', 'gaussian process' and 'random effects' priors. Shown are posterior medians and 90% credible intervals.

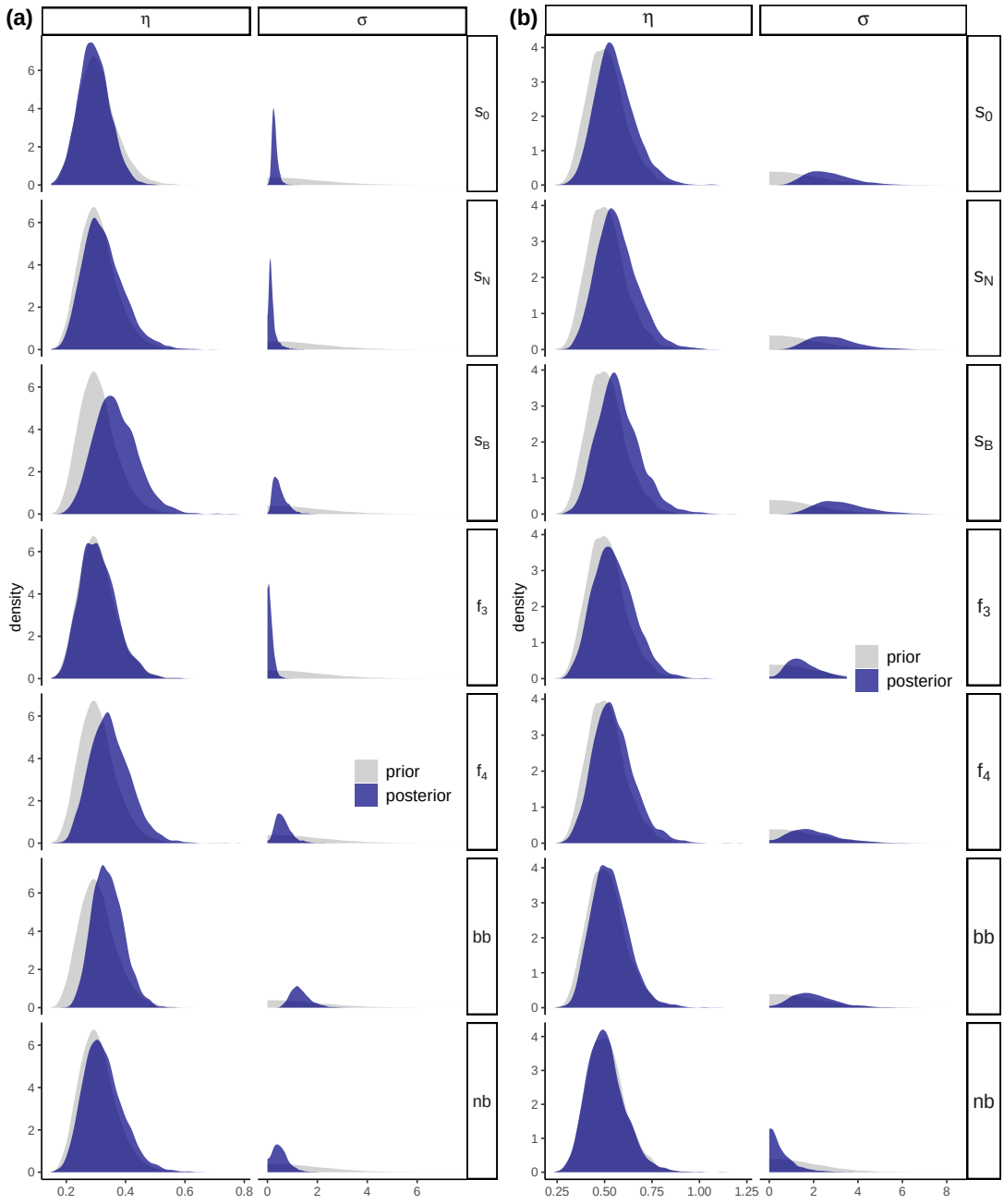


Figure 3: Prior vs. posterior plots for the Gaussian Process hyperparameters for each vital rate. The lengthscale η and marginal sd σ for the mean process $\boldsymbol{\mu}_\nu$ are shown in the left panel (a) and for the log standard deviation process $\log(\boldsymbol{\sigma}_\nu)$ in the right panel (b).

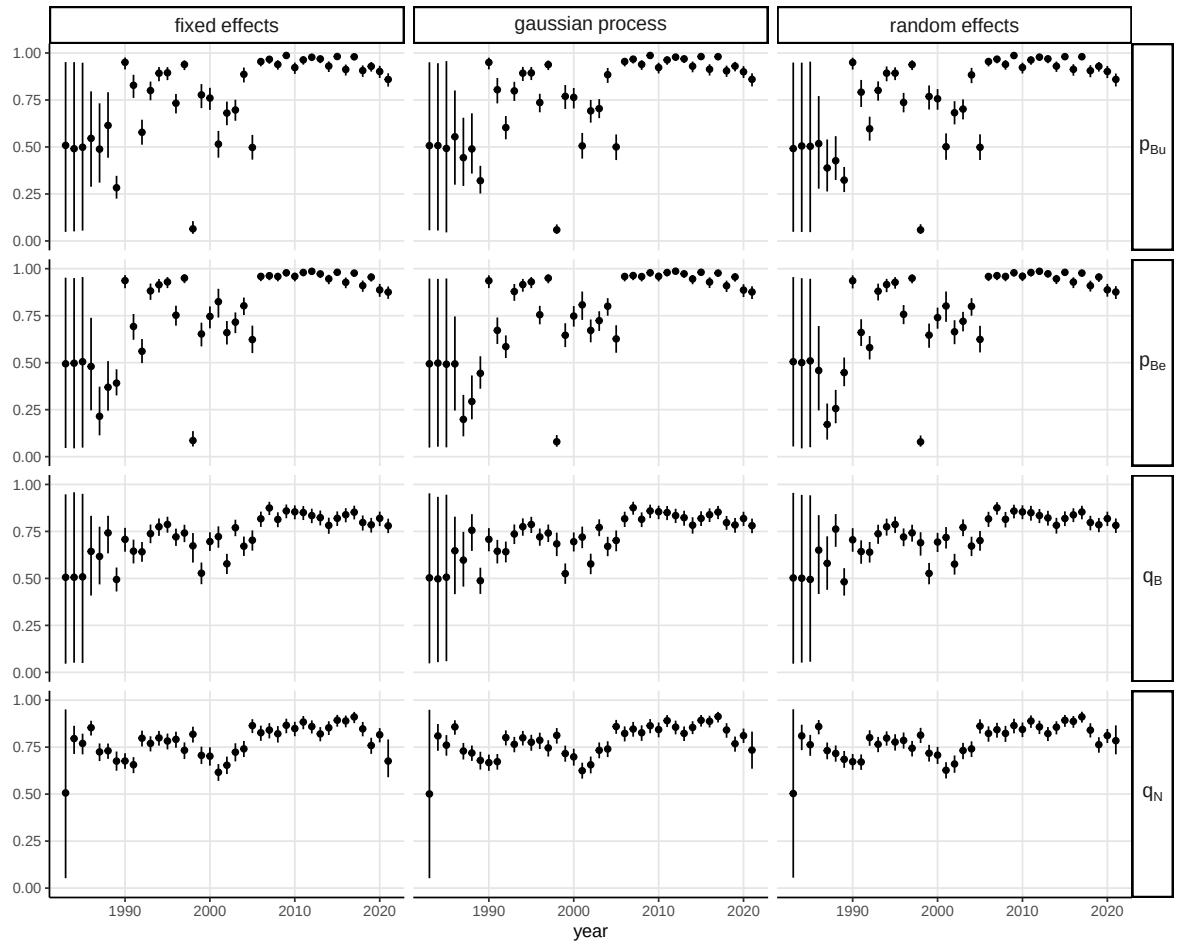


Figure 4: Comparison of detection probabilities for the multi-event model with 'fixed effects', 'random effects' and 'gaussian process' priors on the vital rates. All three models place independent Uniform$(0, 1)$ priors on all detection probabilties. Shown are posterior medians and 90% credible intervals.

2 Population model

The population model defines the joint distribution of all population stage sizes in each year from 1986 to 2021. This is achieved by

(i) placing a prior distribution on the population stage sizes in 1986, the initial year, and

(ii) defining the conditional distributions of the population stage sizes in year $t+1$ given those in year t (and given immigration and the vital rates).

First we describe the latter part of the population model, and we discuss the priors on initial stage sizes in the following section.

2.1 Conditional distributions

In expectation, our population model defines a matrix population model

$$\mathbf{P}_{t+1} = \mathbf{A}_t \mathbf{P}_t,$$

where the vector of population stage class sizes $\mathbf{P}_t$ and the projection matrix $\mathbf{A}_t$ are given by

$$\mathbf{P}_t = \begin{pmatrix} Pb_1 \\ Pb_2 \\ Pb_3 \\ Pb_4 \\ Br \\ Nb \\ In \end{pmatrix}_t, \qquad \mathbf{A}_t = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2}s_W s_0 & 0 & \frac{1}{2}s_W s_0 \\ s_N & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_N(1-f_3) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_N(1-f_4) & 0 & 0 & 0 & 0 \\ 0 & s_N f_3 & s_N f_4 & s_N & s_B bb & s_N nb & s_B bb \\ 0 & 0 & 0 & 0 & s_B(1-bb) & s_N(1-nb) & s_B(1-bb) \\ \omega & \omega & \omega & \omega & \omega & \omega & \omega \end{pmatrix}_t. \quad (2)$$

To incorporate demographic stochasticity we used Binomial distributions to model the population stage sizes at time $t+1$ given those at time t:

$$
\begin{aligned}
Pb_{1,t+1} &\sim \text{Binomial}\left(Br_t + In_t, 0.5 s_{W,t} s_{0,t}\right), \\
Pb_{2,t+1} &\sim \text{Binomial}\left(Pb_{1,t}, s_{N,t}\right), \\
Pb_{3,t+1} &\sim \text{Binomial}\left(Pb_{2,t}, s_{N,t}(1-f_{3,t})\right), \\
Pb_{4,t+1} &\sim \text{Binomial}\left(Pb_{3,t}, s_{N,t}(1-f_{4,t})\right), \\
Br_{t+1} &= Pb_2Br_{t+1} + Pb_3Br_{t+1} + Pb_4Br_{t+1} + BrBr_{t+1} + NbBr_{t+1} + InBr_{t+1}, \text{ where} \\
&\quad Pb_2Br_{t+1} \sim \text{Binomial}\left(Pb_{2,t}, s_{N,t} f_{3,t}\right), \\
&\quad Pb_3Br_{t+1} \sim \text{Binomial}\left(Pb_{3,t}, s_{N,t} f_{4,t}\right), \\
&\quad Pb_4Br_{t+1} \sim \text{Binomial}\left(Pb_{4,t}, s_{N,t}\right), \\
&\quad BrBr_{t+1} \sim \text{Binomial}\left(Br_t, s_{B,t} bb_t\right), \\
&\quad NbBr_{t+1} \sim \text{Binomial}\left(Nb_t, s_{N,t} nb_t\right), \\
&\quad InBr_{t+1} \sim \text{Binomial}\left(In_t, s_{B,t} bb_t\right), \\
Nb_{t+1} &= BrNb_{t+1} + NbNb_{t+1} + InNb_{t+1}, \text{ where} \\
&\quad BrNb_{t+1} \sim \text{Binomial}\left(Br_t, s_{B,t}(1-bb_t)\right), \\
&\quad NbNb_{t+1} \sim \text{Binomial}\left(Nb_t, s_{N,t}(1-nb_t)\right), \\
&\quad InNb_{t+1} \sim \text{Binomial}\left(In_t, s_{B,t}(1-bb_t)\right).
\end{aligned}
$$

The exception is immigration, which does not depend on the other population stage sizes, only on the parameters of the multivariate prior we place on it.

2.2 Priors on initial stage class sizes

In this section we explain how we set priors on initial stage class sizes and how we evaluated the sensitivity of our inferences to the priors.

Setting the priors

The variables that require informative priors are $Pb_{1,1}, Pb_{2,1}, Pb_{3,1}, Pb_{4,1}, Br_1$ and Nb_1, which represent the stage class sizes in 1986, the first year in which breeding counts are available. We gave these normal priors with standard deviation equal to the (positive) mean divided by 3.1, a factor that ensures the prior probability of a negative value is 0.001. To set the prior means we

(i) sourced published estimates of Southern Elephant Seal vital rates from populations other than Marion Island (see Table 1), with the exception of survival to weaning, where we used an estimate from the Marion Island population (but not from the data used in our study),

(ii) evaluated the projection matrix $\mathbf{A}$ of Equation (2) using the estimated values of Table 1 and an assumed immigration rate of $\omega = 0.1$, then

(iii) computed the stable stage distribution of this matrix, then

variable	value	reference	source
s_W	$1 - 0.038$	Pistorius et al. 2001	Table 1
s_0	0.72	Desprez et al. 2013	Fig. 3
s_N	0.83	Desprez et al. 2013	Fig. 3
s_B	0.83	Desprez et al. 2013	Fig. 3
f_3	0.18	Desprez et al. 2013	Fig. 5, MEM
f_4	0.46	Desprez et al. 2013	Fig. 5, MEM
nb	$0.7 \times 0.69 + 0.3 \times 0.38$	Desprez et al. 2013 2017	Fig. 1
bb	$0.7 \times 0.79 + 0.3 \times 0.52$	Desprez et al. 2013 2017	Fig. 1

Table 1: Southern Elephant Seal vital rate estimates sourced from the literature.

(iv) scaled the stable stage distribution so that the total number of breeding females ($Br_1 + In_1$) equals 641, the 15 October breeding female count for 1986.

Prior sensitivity analysis

Since there are no estimates of the immigration rate for Marion Island, and it would be inappropriate to use such an estimate from a different population, we chose $\omega = 0.1$ in setting the priors on the initial stage class sizes. To demonstrate that our results do not depend on this choice we fitted the IPM with four different sets of priors obtained using $\omega = 0.01, 0.05, 0.1, 0.15$. Whilst the value used for immigration rate does affect the priors (Fig. 5a), it has little effect on the posterior distributions of the population stage sizes in 1990, the first year included in the tLTREs (Fig. 5b).

2.3 Informative prior on survival to weaning

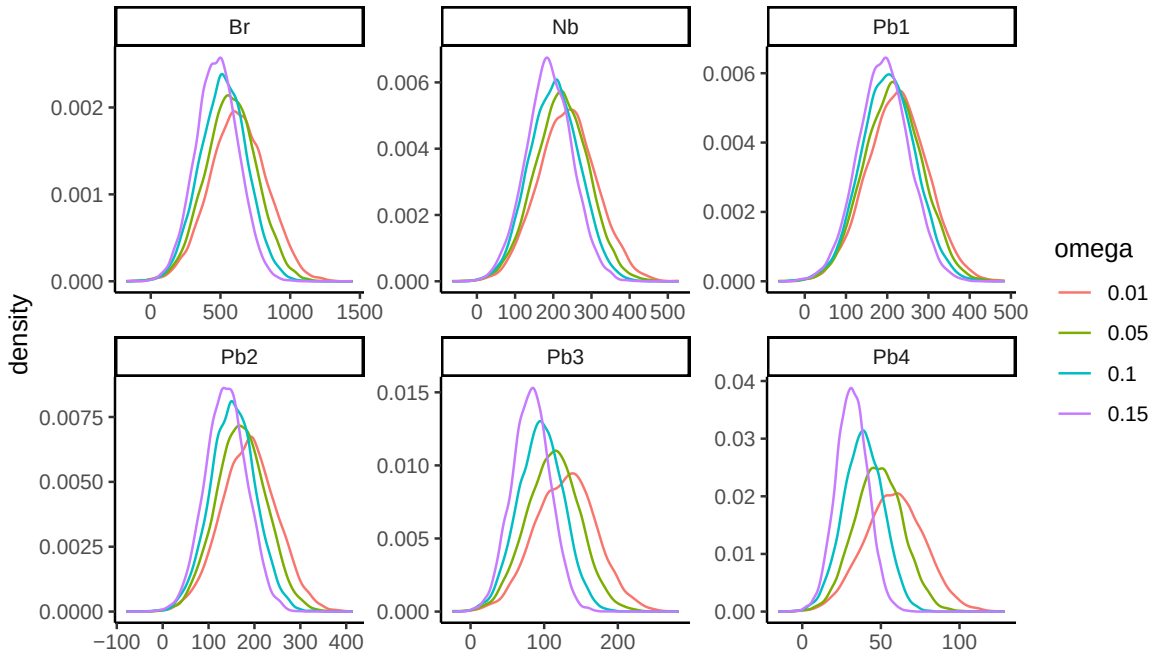
Our fecundity model consists of one parameter, the probability that the single pup born to each breeding female survives until weaning. Data on survival of un-weaned pups was not systematically collected over the entire study period, so we could not estimate annual survival probabilities as we have for our other vital rates. Pistorius et al. 2001 estimated the total annual mortality of pre-weaning and early post-weaning pups on Marion Island from 1990 - 1999. The average total mortality was 3.8%, with a range of $1.6\% - 7.3\%$, and most of the total mortality was pre-weaning mortality (Pistorius et al. 2001). Translating these results into survival probabilities, we defined an informative prior for survival to weaning as

$$s_{W,t} \sim \text{Beta}(96 * 5, 4 * 5), \quad \text{independently for all } t.$$

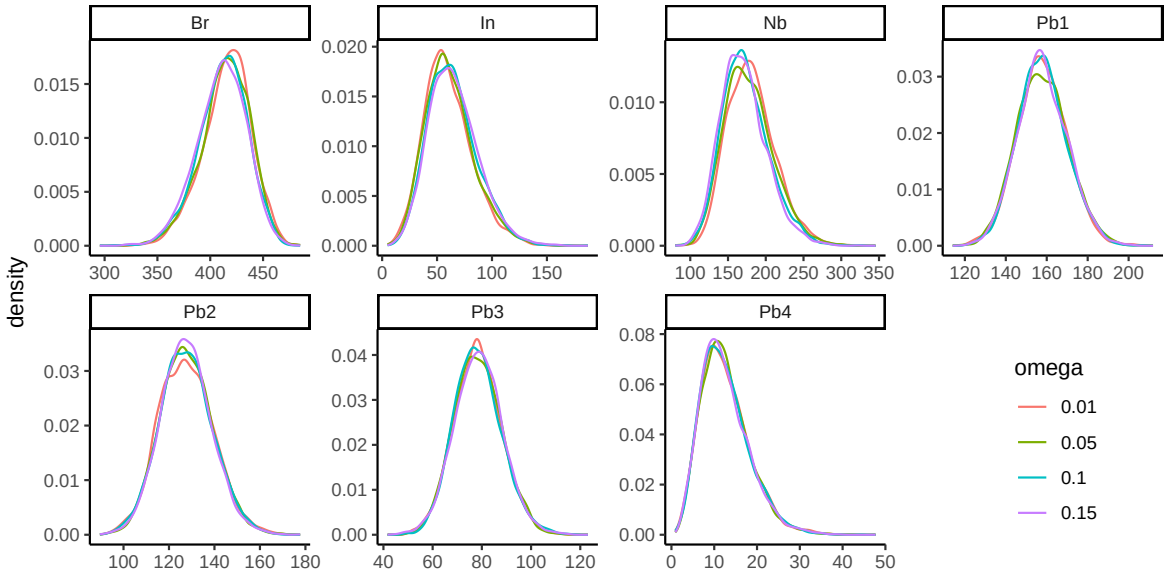
Through this prior, this extra source of mortality and the uncertainty in its estimate are accounted for in our population model.

2.4 Priors on immigration

Here we give the full specifications of our 'Random Effects' and 'Gaussian Process' priors on immigration.



(a) Prior distributions on population stage sizes in 1986. The prior distribution on immigration is not shown here because it does not depend on the assumed immigration rate ω used in calculating the priors.



(b) Posterior distributions on population stage sizes in 1990.

Figure 5: Sensitivity of population stages sizes in 1990 (the first year incorporated into the tLTRE analysis) to priors on initial population stage sizes.

Random Effects prior

We followed Paquet et al. 2021 in the formulation of a 'Random Effects' prior using a Poisson-Lognormal distribution:

$$\begin{aligned} In_t &\sim \text{Poisson}(\Lambda_t) \text{ for all } t, \\ \Lambda_t &\sim \text{Lognormal}\left(\mu_{\log(\Lambda)}, \sigma_{\log(\Lambda)}\right), \\ \mu_{\log(\Lambda)} &\sim \text{Normal}\left(\log(50), 0.4\right), \\ \sigma_{\log(\Lambda)} &\sim \text{Half-normal}(0, 0.4). \end{aligned}$$

We opted for a half-normal prior on the standard deviation paramater $\sigma_{\log(\Lambda)}$ as it has shorter tails than an exponential, and this helps the MCMC sampling. The expected value log(50) for the prior on $\mu_{\log(\Lambda)}$ comes from a mean of 50 estimated using much wider priors in an initial model run. However the sampling is poor with wider priors (we get many divergent transitions after warmup), hence we adopted more informative priors. The precise values for the hyperprior standard deviations are based on prior predictive checks, that is, a trial-and-error process of setting values for these parameters and checking that of the number of immigrants simulated under them was reasonable. For these choices, we were careful that the prior did not dominate the posterior, that is, that the posteriors for these parameters still concentrated within the priors; see the prior vs. posterior plots of Fig 6.

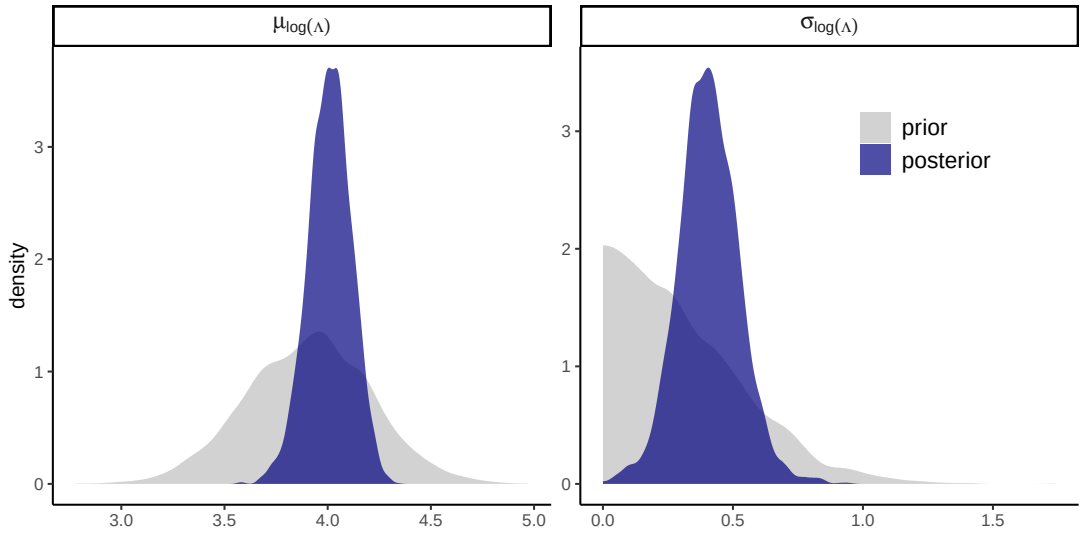


Figure 6: Prior vs. posterior plots for the hyperparameters of the 'Random Effects' Poisson-Lognormal prior on immigration.

Gaussian Process prior

Let $\boldsymbol{In}$ denote the vector of immigrant numbers over the study. We placed a Gaussian Process prior on immigration as

$$\log\left(\boldsymbol{In}\right) \sim \text{Multivariate Normal}\left(\boldsymbol{\mu}_{\log(\boldsymbol{In})}, \boldsymbol{K}_{\log(\boldsymbol{In})}\right),$$

where the mean vector is constant

$$\boldsymbol{\mu}_{\log(\boldsymbol{In})} = \begin{pmatrix} \mu_{\log(\boldsymbol{In})} \\ \vdots \\ \mu_{\log(\boldsymbol{In})} \end{pmatrix},$$

and the covariance matrix is once again determined by an exponentiated quadratic kernel

$$\boldsymbol{K}_{\log(\boldsymbol{In}),i,j} = \sigma^2_{\log(\boldsymbol{In})} \exp\left(-\frac{|t_i - t_j|^2}{2\eta^2_{\log(\boldsymbol{In})}}\right) \text{ for all } i, j.$$

We standardised time before constructing the covariance matrix, and placed the following priors on the mean, marginal variance and lengthscale hyper-parameters:

$$\begin{aligned} \mu_{\log(\boldsymbol{In})} &\sim \text{Normal}(\log(50), 0.4), \\ \sigma_{\log(\boldsymbol{In})} &\sim \text{Half-Normal}(0, 2), \\ \eta_{\log(\boldsymbol{In})} &\sim \text{Lognormal}(\log(0.2), 0.3). \end{aligned}$$

The parameter $\mu_{\log(\boldsymbol{In})}$ is the GP analog of the parameter $\mu_{\log(\Lambda)}$ in the RE prior on immigration above, hence the use of a similar prior for it here. As with GP priors on the vital rates, we use a zero-avoiding lognormal prior on the lengthscale $\eta_{\log(\boldsymbol{In})}$, having standardised the time covariate. Although we used a relatively short-tailed half-normal prior on the marginal standard deviation $\sigma_{\log(\boldsymbol{In})}$, it is probably wider than necessary given that it applies to the log scale number of immigrants. Prior vs. posterior plots of these hyperparamters are presented in Fig. 7.

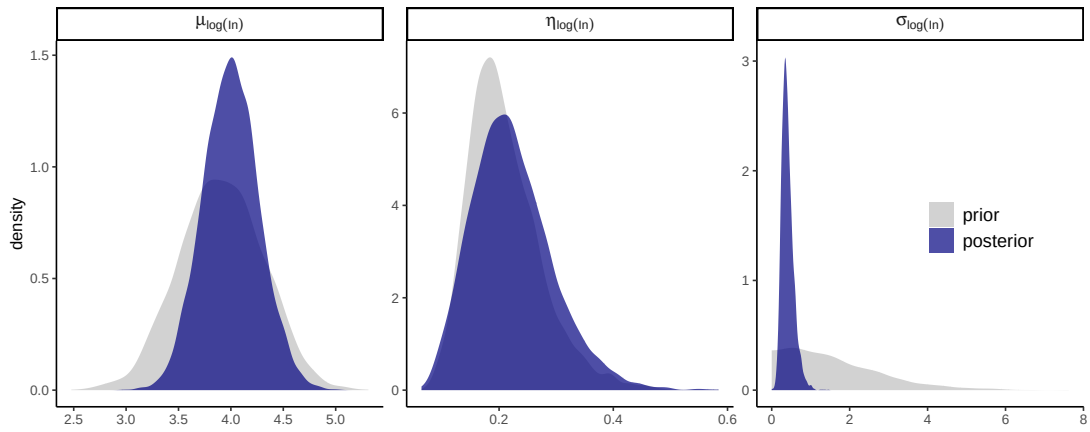


Figure 7: Prior vs. posterior plots for the hyperparameters of the Gaussian Process prior on immigration.

2.5 Supplementary figure for population models

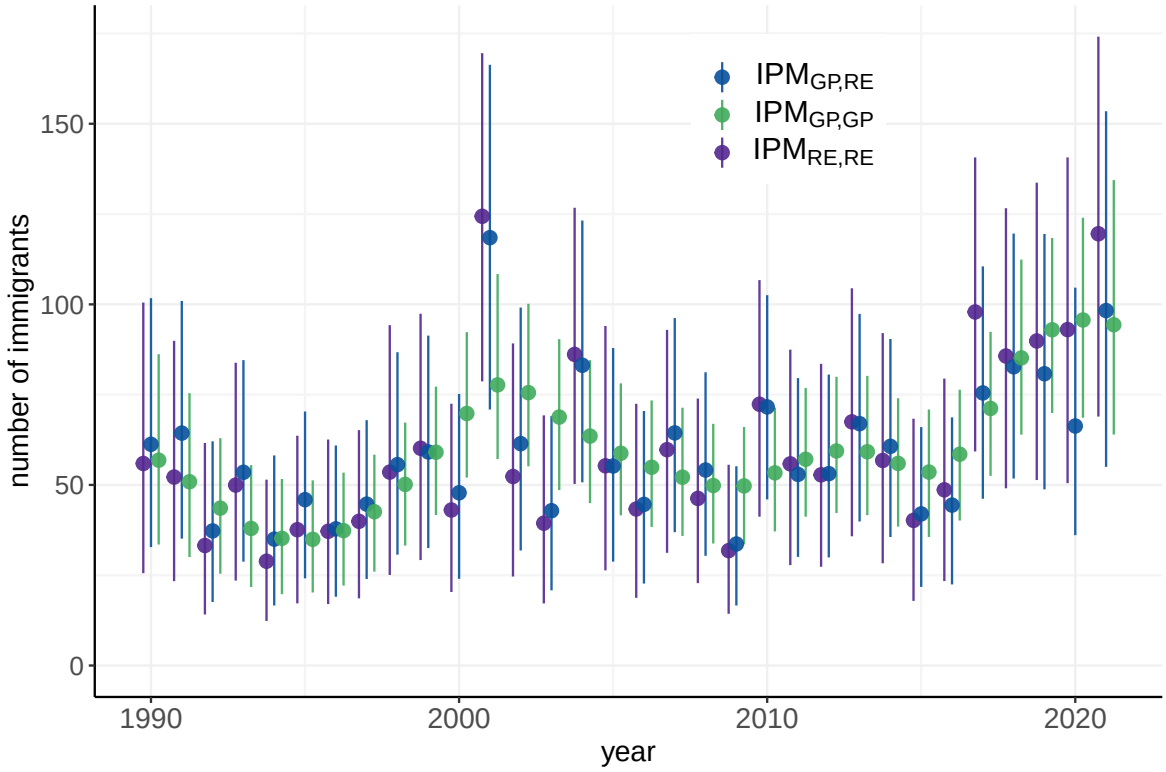


Figure 8: Comparison of annual estimates of the number of immigrants for three IPMs with different priors on the vital rates and on immigration. Shown are posterior medians and 90% credible intervals.

2.6 Estimating the count standard deviation σ_c in the breeding count model

In our breeding count model we fixed the count standard deviation to ten:

$$\text{count}_t \sim \text{Normal}(Br_t + In_t, \sigma_c = 10).$$

Because the estimates of immigration and its demographic contribution are likely sensitive to the count model (Schaub and Kéry 2021), we fitted two extra versions of $IPM_{GP,RE}$ in which we estimated σ_c using either an informative Normal$(10, 2.5)$ or vague Half-normal$(0, 20)$ prior.

The informative prior acknowledges that the exact value $\sigma_c = 10$ is somewhat arbitrary; values between five or fifteen might also be fine, but setting σ_c to one or twenty is probably inappropriate. On the other hand, the vague Half-normal$(0, 20)$ prior does not utilise the external information we have about the count accuracy and estimates σ_c using only the information in the IPM data. Although σ_c is informed by this data, in the sense that it's posterior concentrates within the prior, the estimate does not come close to capturing the information we can add to the model through setting an informative prior or a specific value as we did (Fig. 9a cf. 9b).

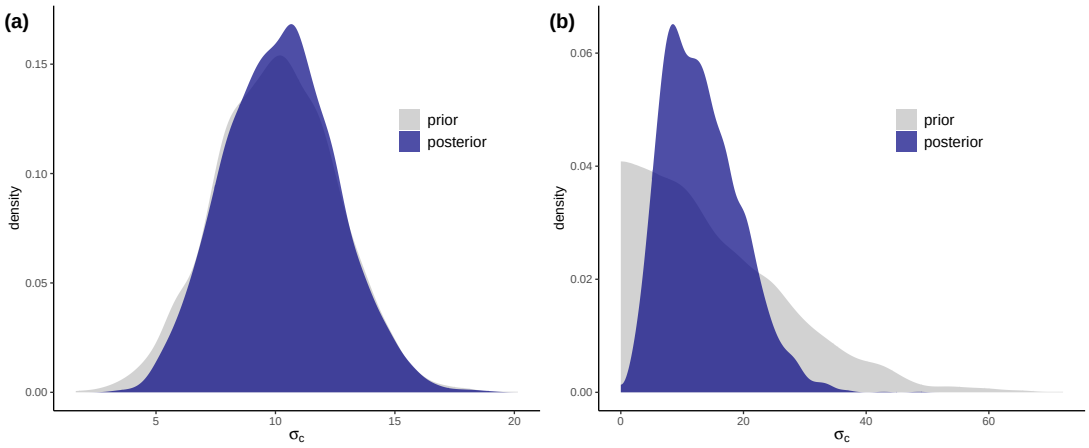


Figure 9: Prior vs. posterior plots for the count standard deviation σ_c under (a) an informative prior and (b) a vague prior.

Predictably, the estimates for immigration and its contribution under the informative prior are similar to those when we set $\sigma_c = 10$. Under the vague prior the estimates are only slightly different and continue to support our conclusion that immigration has been one of the main drivers of population growth in this study; see Figure 10.

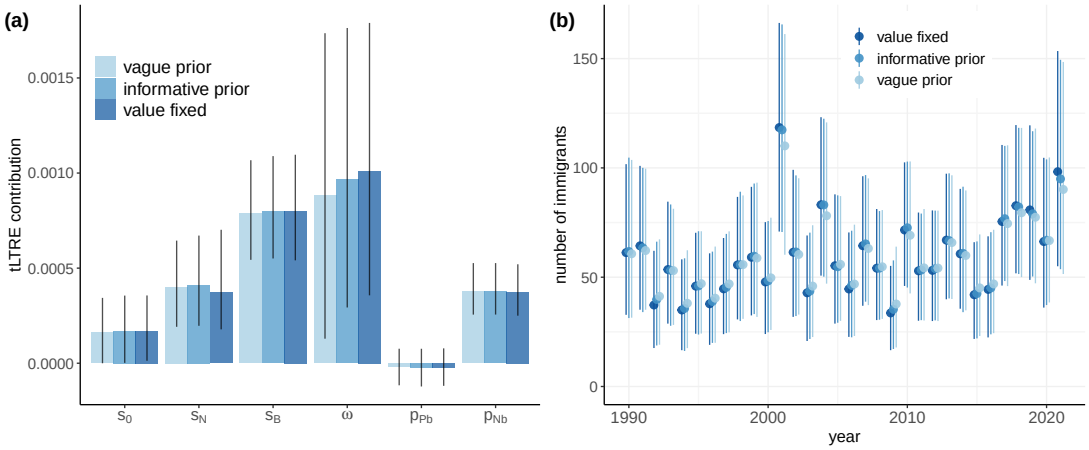


Figure 10: Sensitivity of annual estimates of immigration (a) and its demographic contribution to the treatment of the count standard deviation σ_c in $IPM_{GP,RE}$. The value was either fixed $\sigma_c = 10$, estimated with an informative Normal$(10, 2.5)$ prior, or estimated with a vague Half-normal$(0, 20)$ prior.

In this section we have relied on our interpretation of σ_c as the standard deviation for the count errors alone, and not for residual errors in the IPM as a whole (cf. Schaub and Kéry 2021). Section 7 contains results and discussion to support this.

One final consideration when estimating σ_c in Stan is the presence of divergent transitions after warmup. These diagnostics flag issues with the Markov chains' ability to explore the posterior distribution, particularly in regions where the posterior is highly curved (Monnahan et al. 2017). Whilst one can often resolve the issue by reparametrising the model, we have not been able to do so here or

with other simpler state-space models in which one estimates a variance parameter for both a state and observation process. Thus, our sample from the posterior in these models may be slightly biased (due to a failure to draw samples from some regions of the posterior that are consistent with the data), but we expect the bias to be minimal.

3 Goodness-of-fit

We examined goodness-of-fit of our IPM using posterior predictive checks applied to the component multi-event model and to the IPM as a whole. Here we present goodness-of-fit only for our 'main' $IPM_{GP,RE}$, which was also used in the simulation study. In the rest of this section, 'the multi-event model' refers to the multi-event model with (cheap) Gaussian process priors on the vital rates.

3.1 Goodness-of-fit to the mark-recapture data

We defined two population-level summaries of the mark-recapture data, aimed at testing the model's fit:

- *total annual detections* $D_{\text{tot},t}$: the total number of seals detected at any time of year t,
- *annual breeder detections* $D_{\text{br},t}$: the number of seals detected in the breeding season of year t.

Each of these was used as the basis for a graphical posterior predictive check and for computing a Bayesian p-value using a Freeman-Tukey discrepancy measure.

Generating replicate datasets

Let $\boldsymbol{\theta}$ denote a vector of all the vital rates, detection probabilities and tag loss probabilities in the multi-event model. We randomly selected 100 posterior samples $\boldsymbol{\theta}^{(s)}, s = 1, \cdots, 100$ from the 4000 posterior samples of our multi-event model. Then we generated 100 replicate mark-recapture datasets $y^{(s)}_{\text{cmr}}, s = 1, \cdots, 100$ by simulating data from each of the posterior samples $\boldsymbol{\theta}^{(s)}$. Each replicate dataset has exactly the same structure as the real data, that is, the same number of tagged pups per year, the same tag locations, and the same number of tags (usually two) as were placed on each seal.

Graphical posterior predictive checks

For each replicate dataset, $y^{(s)}_{\text{cmr}}$, the population-level summaries $D^{(s)}_{\text{tot},t}$ and $D^{(s)}_{\text{br},t}$ are calculated by simply counting the number of detections of each type over all individuals in each year t. The same summaries are calculated from the real data (denoted $D^{\text{real}}_{\text{tot},t}$ and $D^{\text{real}}_{\text{br},t}$) and the resulting time series for the real data and each posterior replicate are compared graphically (Fig 11).

Bayesian p-values for the Freeman-Tukey discrepancy measure

For a quantitative assessment of fit, we used the Freeman-Tukey discrepancy measure to compare expected and observed numbers of detections of each type (total and breeder detections) over the whole study period. For two vectors $\boldsymbol{u}, \boldsymbol{v}$ of length T, the Freeman-Tukey discrepancy is defined as

$$FT(\boldsymbol{u}, \boldsymbol{v}) = \sum_{t=1}^{T} \left(\sqrt{u_t} - \sqrt{v_t}\right)^2 .$$

Here we only describe the calculation of a Bayesian p-value for the breeder summary $D_{br,t}$. The calculation for the total detections summary is identical, simply replacing $D_{br,t}$ with $D_{tot,t}$ throughout.

For each posterior sample $s = 1, \cdots, 100$ we let

- $\boldsymbol{D}_{\text{br}}^{(s)} = \left(D_{\text{br},1}^{(s)}, \cdots, D_{\text{br},T}^{(s)}\right)$, the vector of observed breeder detections in each year calculated from the replicate data $y_{\text{cmr}}^{(s)}$ as above, and
- $\mathbb{E}\boldsymbol{D}_{\text{br}}^{(s)} = \left(\mathbb{E}D_{\text{br},1}^{(s)}, \cdots, \mathbb{E}D_{\text{br},T}^{(s)}\right)$, the vector of expected breeder detections in each year, calculated from the posterior sample $\boldsymbol{\theta}^{(s)}$ (we describe the calculation of expected detections below).

In addition we have $\boldsymbol{D}_{\text{br}}^{\text{real}}$ the vector of observed breeder detections calculated from the real data. Then, for each posterior replicate s, we calculate the Freeman-Tukey discrepancy between the expected and observed number of detections, for the replicate and real data:

$$FT\left(\boldsymbol{D}_{\text{br}}^{(s)}, \mathbb{E}\boldsymbol{D}_{\text{br}}^{(s)}\right), \quad FT\left(\boldsymbol{D}_{\text{br}}^{\text{real}}, \mathbb{E}\boldsymbol{D}_{\text{br}}^{(s)}\right).$$

The Bayesian p-value is the proportion of replicates for which the discrepancy between observed and expected detections in the real data exceeds that of the replicate data:

$$p_B = \frac{1}{100}\sum_{s=1}^{100} I\left(FT\left(\boldsymbol{D}_{\text{br}}^{(s)}, \mathbb{E}\boldsymbol{D}_{\text{br}}^{(s)}\right) < FT\left(\boldsymbol{D}_{\text{br}}^{\text{real}}, \mathbb{E}\boldsymbol{D}_{\text{br}}^{(s)}\right)\right).$$

It remains to describe how we calculated the expected number of breeder detections used for the posterior predictive check.

Calculating expected detections

Once again we will focus on the calculation of expected *breeder* detections, and point out the one small modification required for the calculation of the expected *total* number of detections. To simplify notation, we drop the breeder subscript from the number of detected breeders in year t, thus $D_t = D_{\text{br},t}$. We define

- $D_{c,t}$ to be the number of detected breeders in year t that were tagged as pups in year c (i.e. belong to cohort c),

Then, decomposing D_t by cohort and taking expectations one has

$$\mathbb{E}D_t = \mathbb{E}\left(\sum_{c=1}^{t-3} D_{c,t}\right) = \sum_{c=1}^{t-3} \mathbb{E}D_{c,t}, \tag{3}$$

and the calculation reduces to calculating the $\mathbb{E}D_{c,t}$. For this we let

- $\boldsymbol{\Theta}_t^{(s)}$ to be the transition probability matrix from time t to $t+1$, evaluated at the posterior sample parameters $\boldsymbol{\theta}^{(s)}$, and
- $\boldsymbol{\Phi}_t^{(s)}$ to be the emission matrices at time t, evaluated at the posterior sample parameters $\boldsymbol{\theta}^{(s)}$.

Then the probability that an individual from cohort c is detected as a breeder in year t is

$$\sum_{j \in J_{\text{Br}}} \delta\left(\prod_{\tau=c}^{t-1} \boldsymbol{\Theta}_\tau^{(s)}\right) \boldsymbol{\Phi}_{t,\cdot,j}^{(s)}, \tag{4}$$

where

- δ is row-vector of zeros with one '1' in the place corresponding to the individual's initial number of tags as a pup,
- J_{Br} is the set of column indices in the emission matrices $\mathbf{\Phi}$ that correspond to being detected at least once in the breeding season, and
- $\mathbf{\Phi}^{(s)}_{t,\cdot,j}$ is the j^{th} column of the emission matrix $\mathbf{\Phi}^{(s)}_t$.

The expected number of breeder detections in year t from cohort c simply sums the above over individuals in that cohort:

$$\mathbb{E}D_{c,t} = \sum_{i \text{ in cohort } c} \sum_{j \in J_{\text{Br}}} \delta_i \left(\prod_{\tau=c}^{t-1} \mathbf{\Theta}^{(s)}_{\tau} \right) \mathbf{\Phi}^{(s)}_{t,\cdot,j}. \tag{5}$$

To calculate the expected number of *total* detections in each year (rather than breeder detections), the only modification that needs to be made to this is to replace J_{Br}, the column indices corresponding to breeder detection events, with J_{Tot}, the column indices correspoding to any detection event. (Thus J_{Tot} is the set of all column indices except the last, because the last column corresponds to 'undetected'.)

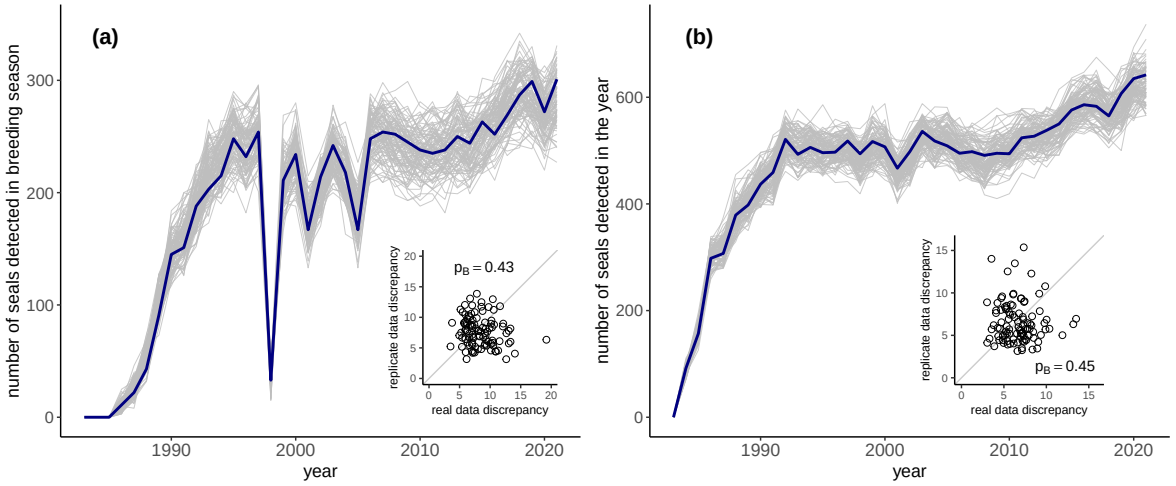


Figure 11: Posterior predictive check for breeder detections (a) and all detections (b).

3.2 Goodness-of-fit to count data

As for our mark-recapture data, we assessed the goodness-of-fit to our count data using graphical posterior predictive checks and accompanying Bayesian p-values.

As before, we randomly sampled 100 posterior draws from the joint posterior of our IPM. We simulated replicate count data by simulating the entire population under our population model (which is specified above in Section 2.1), using the values of vital rates and initial population stage sizes in each of these posterior draws. Then we drew the count data in each year according to our breeding count model:

$$\text{count}_t \sim \text{Normal}(Br_t + In_t, \sigma = 10).$$

We use a Freeman-Tukey discrepancy measure to compute a Bayesian p-value in an identical manner to that described above. The only aspect that differs is the calculation of the expected breeder counts, which are obtained by calculating expected population stage sizes in all years from the population model, and then summing the expected number of local and foreign-breeders in each year for an expected breeder count.

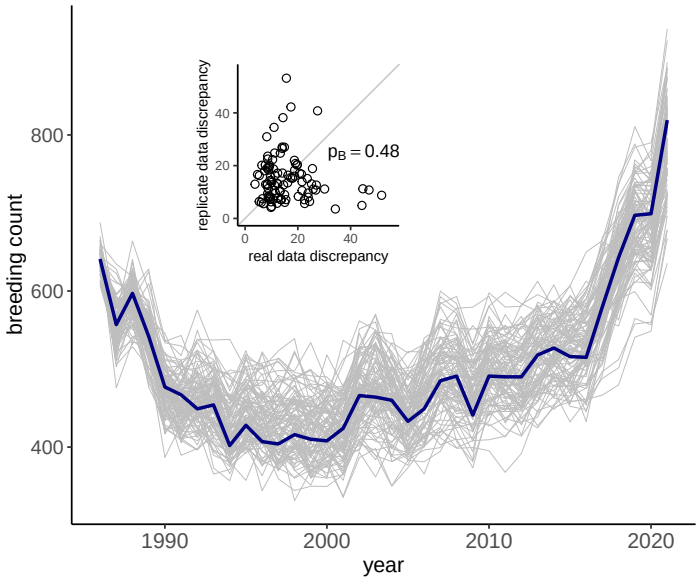


Figure 12: Posterior predictive check for the 15 October breeder count data.

4 tLTRE analysis

4.1 Calculating sensitivities with respect to population structure

The population structure is captured by the proportions p_{Br}, p_{Nb} ,p_{In} and p_{Pb} which are constrained to sum to one. This constraint complicates the calculation and interpretation of sensitivities (partial derivatives) with respect to population structure. A partial derivative asks how much a function varies as one variable changes whilst all others remain fixed. However, when the variables are constrained by the sum-to-one requirement, any change in one variable must be accompanied by a change in at least one other variable. To deal with this we parametrise our expression for λ_t in terms of p_{Nb} and p_{Pb} alone:

$$\lambda_t = (0.5 s_{W,t} s_{0,t} + s_{B,t})\,(1 - p_{Nb,t} - p_{Pb,t}) + s_{N,t}\,(p_{Nb,t} + p_{Pb,t}) + \omega_t.$$

Now the sensitivities represent the rate of change of λ as each of p_{Nb} and p_{Pb} change whilst holding the other fixed, but allowing $p_{Br} + p_{In}$ to vary according to the sum-to-one constraint.

4.2 Evaluating the approximation

The tLTRE approximates the temporal variance in realised population growth with the sum of contributions from different demographic parameters. In the notation of Koons et al. 2016, where θ represents a generic demographic parameter:

$$\mathrm{var}(\lambda_t) \approx \sum_{i,j} \left.\frac{\partial \lambda}{\partial \theta_i}\right|_{\bar{\Theta}} \left.\frac{\partial \lambda}{\partial \theta_j}\right|_{\bar{\Theta}} \mathrm{cov}(\theta_i, \theta_j).$$

The partial derivatives are evaluated at the temporal mean of all parameters, denoted $\bar{\Theta}$. Summing over j computes the contribution due to θ_i, so the right side represents the sum of contributions from all parameters.

To evaluate this approximation we calculate a relative difference by dividing the difference between the left and right sides by $\mathrm{var}(\lambda_t)$. Repeating this for each posterior sample from $IPM_{GP,RE}$ gives a posterior distribution over these relative differences, which is plottted in Fig. 13.

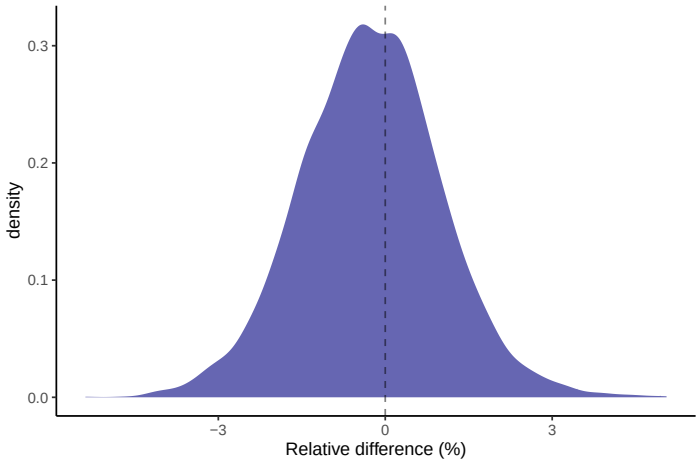


Figure 13: Posterior distribution of relative differences between temporal variance in realised population growth rates and the sum of demographic contributions from all parameters in the tLTRE analysis.

5 Simulations

Repeatedly fitting the IPM is computationally expensive, so we used an approximate, iterative procedure when fitting simulated data. In this procedure we first fit the multi-event model to the mark-recapture data, then use the posterior to define an informative prior on vital rates in the population model and fit it to the count data. This is similar to the 'continuous updating' of general Bayesian inference, in which the posterior is updated as new data are obtained. Here we show that the tLTRE estimates from the iterative fit are very similar to those from the full fit, with all differences of negligible size relative to the standard errors (Fig. 14).

5.1 Full and iterative fit of the IPM

Fitting our IPM to all sets of data simultaneously takes $17-20$ hours using a 24-core machine, running four chains in parallel, and using six cores per chain for within-chain parallelisation of the multi-event likelihood calculations. In order to fit the IPM hundreds of times to simulated data, we developed an "iterative" procedure that takes less than 6 hours on the same hardware. In this procedure we

(i) fit the multi-event model to the capture-mark-recapture data, thus obtaining posterior samples for all the vital rates,

(ii) approximate the joint posterior of all vital rates with a multivariate normal distribution (on the logit scale), where the means and pairwise covariances are computed from the posterior samples, and

(iii) use this logit-multinormal distribution as a prior on the vital rates for the population model that is fit to the count data.

We justify this procedure theoretically and empirically. Firstly, we show that in the absence of the approximation (ii), the posterior obtained from the "iterative" fit is the same as that from the "full" fit. Secondly, we show that the tLTRE contributions estimated from the two fitting procedures are very similar.

5.2 The iterative posterior is the full posterior

In this section, we use θ to denote the full set of vital rates, η the set of tag loss and detection probabilities, ζ the stage-class sizes, immigrants, and survival to weaning, y_{cmr} the mark recapture data, and y_{count} the count data. We seek the full posterior $[\theta, \eta, \gamma | y_{cmr}, y_{count}]$, whilst the posterior obtained from the multi-event model fit to the capture-mark-recapture data is $[\theta, \eta | y_{cmr}]$. Using the facts that the count data is conditionally independent of everything else given the population variables

ζ, and ζ is conditionally independent of everything else given the vital rates θ, we get

$$\begin{aligned}
[\theta, \eta, \gamma | y_{cmr}, y_{count}] &\propto [y_{cmr}, y_{count}, \theta, \eta, \zeta] \\
&= [y_{cmr}, y_{count} | \theta, \eta, \zeta] \cdot [\theta, \eta, \zeta] \\
&= [y_{cmr}, y_{count} | \theta, \eta, \zeta] \cdot [\theta] \cdot [\eta] \cdot [\zeta | \theta] \\
&= [y_{count} | \zeta] \cdot [\zeta | \theta] \cdot [y_{cmr} | \theta, \eta] \cdot [\theta] \cdot [\eta] \\
&\propto [y_{count} | \zeta] \cdot [\zeta | \theta] \cdot [\theta, \eta | y_{cmr}].
\end{aligned}$$

Therefore the full posterior is the same as the posterior obtained by fitting the count data to the population model, using as prior for the vital rates the posterior obtained from the multi-event model.

5.3 tLTRE sensitivity to iterative fit

Although the full posterior and the iterative posterior are theoretically the same, our representation of the posterior $[\theta, \eta | y_{cmr}]$ from the multi-event model is only approximate. To show that this approximation is adequate for the purpose of tLTRE analyses, we compared tLTRE estimates obtained from the full and iterative fits applied to real data, and data simulated under the model. The discrepancies between the two fits are of negligible size relative to the standard errors of the estimates (see Fig. 14).

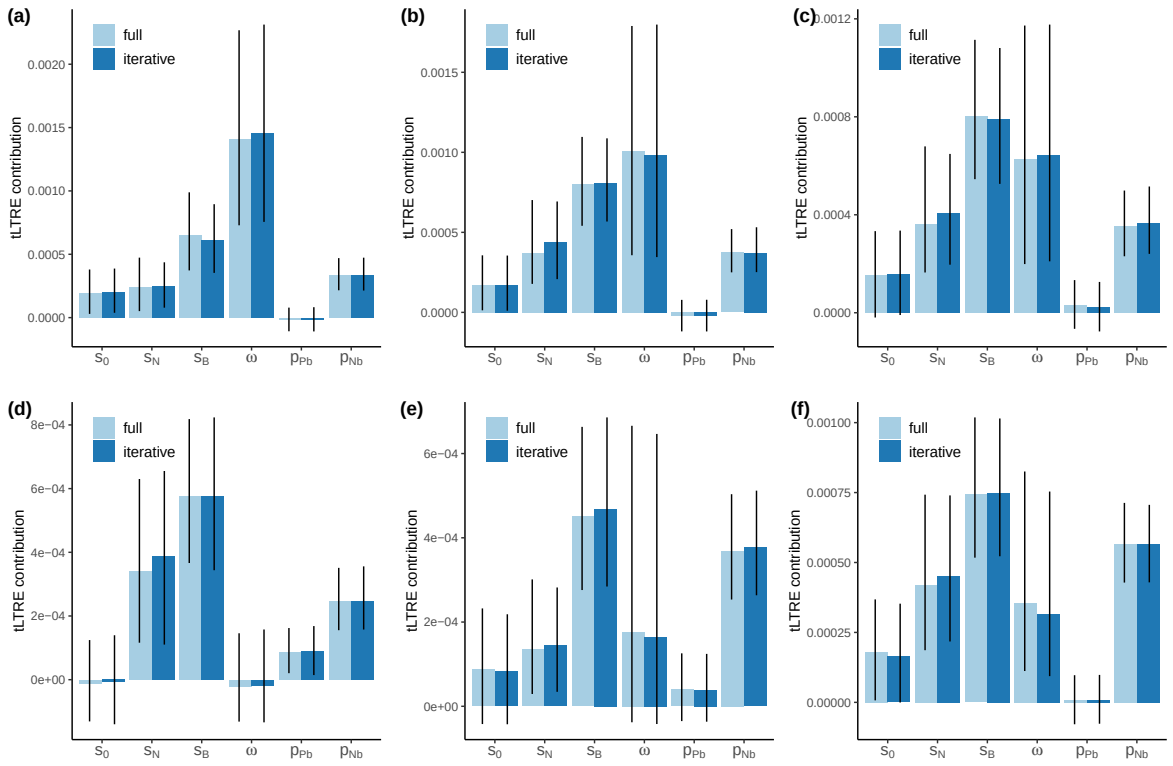


Figure 14: Comparison of tLTRE estimates based on full vs. iterative IPM fits, for six IPMs fit to real and simulated data. (a) $IPM_{RE,RE}$ fit to real data, (b) $IPM_{GP,RE}$ fit to real data, (c) $IPM_{GP,GP}$ fit to real data, (d)-(f) $IPM_{GP,RE}$ fit to simulated data with (d) low, (e) medium or (f) high temporal variance in immigration.

6 Stan implementation

Stan uses Hamiltonian Monte Carlo (HMC), which generates samples that are typically much less autocorrelated than those produced by Gibbs or Metropolis-Hastings samplers (Monnahan et al. 2017). Therefore fewer iterations are required to achieve a given Effective Sample Size (ESS), and for all our models we ran four chains, each with 1000 warmup iterations and 1000 sampling iterations. To assess convergence and mixing of the chains and validity of the posterior sample, we checked that all parameters had split-$\hat{R}$ statistic less than 1.01 (Vehtari et al. 2021) and large ESSs, and that divergent transitions after warmup were absent or rare.

HMC does not accommodate models with discrete parameters because it requires the likelihood to be differentiable. Therefore we marginalised out the discrete, latent states of seals in the multi-event model, and approximated Binomial and Poisson distributions for population stage sizes with moment-matched lognormal distributions (see sections 6.1 and 6.2 below for details).

To improve efficiency of the likelihood calculation for the multievent model we used a reduced representation of the mark-recapture data, in which the 9500 individual capture histories are represented by 2727 unique capture histories and the number of times each occurs (Turek et al. 2016).

The Gaussian Process priors on each vital rate are computationally expensive, requiring inversion of covariance matrices at every iteration as new samples of the GP hyper-parameters (the marginal variance and lengthscale) are drawn. For this reason we used a 'cheap', approximate GP when fitting the IPMs, where we fixed the GP hyper-parameters to their posterior medians from a fit of the multi-event model with full GP priors. The vital rate estimates from the full GP and 'cheap' GP models are practically indistinguishable (Fig. 15), and the covariance matrices are only inverted once upfront in the cheap version, which makes it twice as fast to fit the IPM.

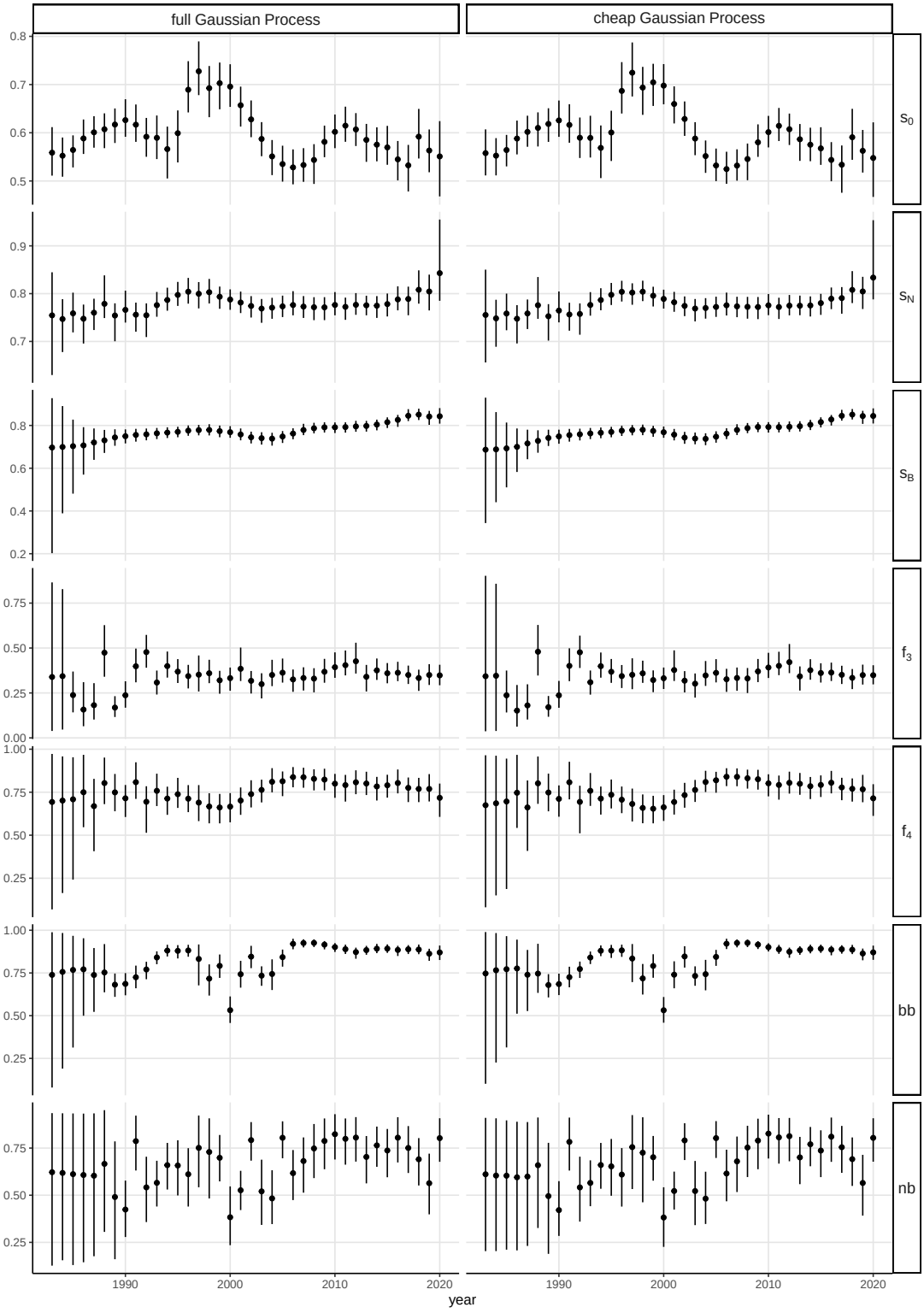


Figure 15: Comparison of vital rate estimates for the multi-event model with the 'full' Gaussian Process prior, and the approximate, 'cheap' Gaussian Process prior. Shown are posterior medians and 90% credible intervals.

6.1 Marginalising discrete latent states in the multi-event model

The multi-event model is a Hidden Markov Model, so the calculation of the marginalised likelihood for each individual capture history is a standard calculation using the forward algorithm (Zucchini et al. 2016). Let

- $\mathbf{y} = (y_{fc}, \cdots, y_T)$ denote an individual capture history,
- $\mathbf{z} = (z_{fc}, \cdots, z_T)$ denote the individual's latent state encoding it's biological state and tag number,

where $t = fc$ is the year when they were first tagged ("first captured") and T is the final year in which resightings are made. Since every seal is tagged as a pup, z_{fc} is known and the marginalised likelihood we calculate is conditional on this: $p(y_{fc+1}, \cdots, y_T | z_{fc})$. This likelihood depends on all the survival, breeding, tag loss and detection probabilities, which are the quantities we want to estimate, but for the purposes of describing the forward algorithm we supress this in the notation. Then the marginalised likelihood can be expressed as a matrix product (Zucchini et al. 2016)

$$\delta \Theta_{fc} D(\Phi_{\cdot, y_{fc+1}}) \cdots \Theta_{T-1} D(\Phi_{\cdot, y_T}) \mathbf{1},$$

where

- δ is a row vector of zeros with one '1' in the position corresponding to the Pb_0 state and the number of tags the pup was tagged with,
- Θ_t is the transition probability matrix from year t to $t+1$,
- $D(\Phi_{\cdot, y_t})$ is a diagonal matrix, whose diagonal entries are given by the y_t^{th} column of the emission matrix Φ_t, and
- $\mathbf{1}$ is a column vector of ones.

To implement this we define a sequence of row vectors $\boldsymbol{\gamma}_{fc}, \cdots, \boldsymbol{\gamma}_{\boldsymbol{T}}$, called 'forward variables', which are calculated recursively:

(i) $\boldsymbol{\gamma}_{fc} = \delta$,

(ii) $\boldsymbol{\gamma}_{t+1} = \boldsymbol{\gamma}_t \Theta_t D(\Phi_{\cdot, y_{t+1}})$ for $t = fc, \cdots, T-1$.

Then the marginalised likelihood equals $\boldsymbol{\gamma}_T \mathbf{1}$, or equivalently, the sum of the entries of $\boldsymbol{\gamma}_T$.

6.2 Moment-matching Binomial and Poisson distributions in the population model

Our population model treats the number of individuals in each stage class as a non-negative integer, and the conditional distributions of these population stage sizes are given as Binomial distributions. In addition, in $IPM_{GP,RE}$, the prior on the number of immigrants involves a Poisson distribution. To implement this in Stan we need to treat the population stages sizes and the number of immigrants as continuous, non-negative values. Therefore we replace the Binomial and Poisson distributions with moment-matched lognormal disibtrutions, as follows.

Binomial distributions. Instead of

$$P \sim \text{Binomial}(N, p),$$

we use

$$P \sim \text{Lognormal}(\mu, \sigma^2),$$

where the values for μ, σ^2 are chosen so that the mean and variance of the lognormal distribution are equal to those of the Binomial distribution:

$$\begin{aligned} \exp\left(\mu + \frac{\sigma^2}{2}\right) &= Np, \\ (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2) &= Np(1-p). \end{aligned}$$

Solving these equations gives

$$\begin{aligned} \mu &= \frac{3}{2}\log(Np) - \frac{1}{2}\log(Np + (1-p)), \\ \sigma^2 &= \log(Np + (1-p)) - \log(Np). \end{aligned}$$

Poisson distributions. Similarly, instead of

$$P \sim \text{Poisson}(\lambda),$$

we use

$$P \sim \text{Lognormal}(\mu, \sigma^2),$$

where the values for μ, σ^2 are chosen so that the mean and variance of the lognormal distribution are equal to those of the Poisson distribution:

$$\begin{aligned} \exp\left(\mu + \frac{\sigma^2}{2}\right) &= \lambda, \\ (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2) &= \lambda. \end{aligned}$$

Solving these equations gives

$$\begin{aligned} \mu &= \frac{3}{2}\log(\lambda) - \frac{1}{2}\log(\lambda + 1), \\ \sigma^2 &= \log(\lambda + 1) - \log(\lambda). \end{aligned}$$

7 Interpretation of count standard deviation σ_c

As noted in the manuscript, in setting the value $\sigma_c = 10$ in our population count model we have relied on our interpretation of this parameter as the standard devation for the count errors. However, Schaub and Kéry 2021 (see especially sections 5.5.2 and 6.6.5) argue that this parameter represents the residual standard deviation in the IPM as a whole. Their interpretation is supported by simulations which show that the estimate for σ_c increases when the IPM is misspecified relative to a correctly specified IPM (e.g. Schaub and Kéry 2021, section 6.6.5). However, we think this behavior arises because the state-space model struggles to separate variance in the state process from that in the observation process, a problem that has been noted in the literature (e.g. Auger-Méthé et al. 2016; Hobbs and Hooten 2015, Chapter 6.3). When there is just one count per year, it appears that σ_c is the most weakly-informed parameter in the model and is overestimated to accommodate lack of fit. (When we suggest that σ_c is the most weakly informed parameter in the model we are assuming there is no additional parameter, as is the case in Schaub and Kéry 2021, section 6.6.5).

To identify observation variance (and thus state process variance as well) it is recommended to replicate counts each year (e.g. Hobbs and Hooten, 2015, Chapter 6.3, pg.141). If one simulates data with replicate counts, then the estimate for σ_c does *not* increase when the model is misspecified, and the 'true' data-generating value is recovered. To illustrate this, we modified the Woodchat Shrike simulation of Schaub and Kéry (2021, Section 6.6.5) and compared estimates for σ_c between correctly specified and misspecified IPMs, fit to IPM data with single or repeat (two) counts within each year. The estimates for σ_c are shown in Fig. 16.

Returning to our elephant seal IPM, we do not include repeat counts each year, but the accuracy with which we can count seals at the breeding aggregations allows us to have strong prior beliefs about the value of σ_c. Crucially, this information about σ_c is not in the data; it needs to be supplied from the outside, either through an informative prior or setting the value as we have done (see Section 2.6).

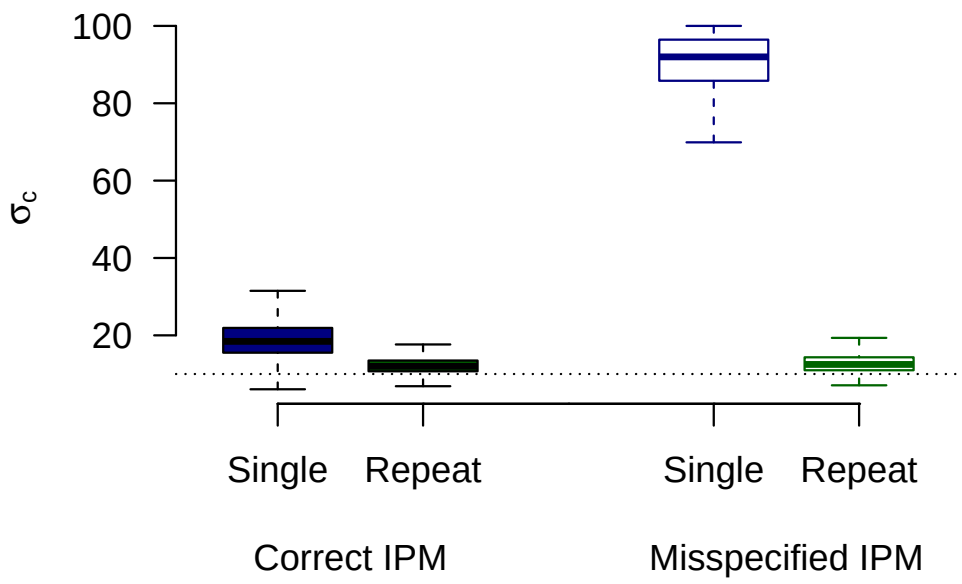


Figure 16: Posterior estimates for σ_c for two IPMs fit to two related datasets. Both IPMs use a Normal count model with standard deviation σ_c. In the correct IPM the inferential model is the same as the data-generating model. In the misspecified IPM, the population model is deliberately misspecified by assuming that first-year birds cannot breed, when in fact they do in the data-generating model. In the 'single' dataset, only one count is simulated per year, whereas in the 'repeat' dataset, two counts are simulated per year. The mark-recapture and fecundity data in these datasets are the same. The dotted line indicates the 'true' value of σ_c under which count data where sampled.

References

Auger-Méthé, M. et al. (2016). "State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems". *Scientific reports* 6.1, p. 26677.

Desprez, M. et al. (2013). "Known unknowns in an imperfect world: incorporating uncertainty in recruitment estimates using multi-event capture–recapture models". *Ecology and Evolution* 3.14, pp. 4658–4668.

Hobbs, N.T. and M.B. Hooten (2015). *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press.

Koons, D.N. et al. (2016). "A life-history perspective on the demographic drivers of structured population dynamics in changing environments". *Ecology Letters* 19.9, pp. 1023–1031. DOI: 10.1111/ele.12628.

Monnahan, C. C., J. T. Thorson, and T. A. Branch (2017). "Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo". *Methods in Ecology and Evolution* 8.3, pp. 339–348. DOI: 10.1111/2041-210X.12681.

Paquet, M. et al. (2021). "Integrated population models poorly estimate the demographic contribution of immigration". *Methods in Ecology and Evolution* 12.10, pp. 1899–1910. DOI: 10.1111/2041-210X.13667.

Pistorius, P. et al. (2001). "Pup mortality in southern elephant seals at Marion Island". *Polar Biology* 24.11, pp. 828–831. DOI: 10.1007/s003000100285.

Schaub, M. and M. Kéry (2021). *Integrated population models: theory and ecological applications with R and JAGS*. Academic Press.

Turek, D., P de Valpine, and C.J. Paciorek (2016). "Efficient Markov chain Monte Carlo sampling for hierarchical hidden Markov models". *Environmental and Ecological Statistics* 23.4, pp. 549–564. DOI: 10.1007/s10651-016-0353-z.

Vehtari, A. (2023). *Gaussian Process demonstration with Stan*. URL: https://avehtari.github.io/casestudies/Motorcycle/motorcycle_gpcourse (visited on 04/06/2023).

Vehtari, A. et al. (2021). "Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion)". *Bayesian Analysis* 16.2. DOI: 10.1214/20-BA1221.

Zucchini, W., Macdonald I.L., and R. Langrock (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. 2nd. Chapman and Hall/CRC.