

# Importance sampling and Bayesian model comparison in ecology and evolution

Dave W. Hudson<sup>1</sup>  | Dave J. Hodgson<sup>1</sup>  | Michael A. Cant<sup>1</sup> | Faye J. Thompson<sup>1</sup>  |  
Richard Delahay<sup>2</sup> | Robbie A. McDonald<sup>3</sup>  | Trevelyan J. McKinley<sup>4</sup>

<sup>1</sup>Centre for Ecology and Conservation,  
University of Exeter, Penryn, UK

<sup>2</sup>National Wildlife Management Centre,  
Animal and Plant Health Agency, Sand  
Hutton, UK

<sup>3</sup>Environment and Sustainability Institute,  
University of Exeter, Penryn, UK

<sup>4</sup>Faculty of Health and Life Sciences,  
Exeter, UK

## Correspondence

Dave W. Hudson

Email: [d.hudson2@exeter.ac.uk](mailto:d.hudson2@exeter.ac.uk)

Trevelyan J. McKinley

Email: [t.mckinley@exeter.ac.uk](mailto:t.mckinley@exeter.ac.uk)

## Funding information

Animal and Plant Health Agency; Natural  
Environment Research Council, Grant/  
Award Number: NE/M010260/1;  
University of Exeter

**Handling Editor:** Chris Sutherland

## Abstract

1. Bayesian approaches to the modelling of ecological systems are increasingly popular, but there are competing methods for formal model comparisons. Here, we focus on the task of performing multimodel inference through estimating posterior model weights, which encompasses uncertainties in the choice of competing model structure into the inference outputs.
2. Model-based approaches such as reversible-jump Markov chain Monte Carlo (RJ-MCMC) are flexible and allow multimodel inference, but can be complex to implement and optimise, and so we translate a model-based approach for ecological applications using Importance Sampling to estimate the marginal likelihood of the data given a particular model. This approach allows for model comparison through the estimation of Bayes' Factors or interpretable posterior model probabilities, yielding model weights that facilitate multimodel inference through Bayesian model averaging.
3. We demonstrate Importance Sampling with two case study investigations in animal demography: censused analysis of banded mongoose (*Mungos mungo*) survival where missing data are uncommon, and capture–mark–recapture analysis of European badger (*Meles meles*) survival where data are commonly missing.
4. We compare outcomes of the model comparison using the Importance Sampling approach to those obtained through single-model inference approaches using Deviance information criteria and the Watanabe–Akaike information criteria. The results of the Importance Sampling method aligns with RJ-MCMC model comparisons while often being more straightforward to fit and optimise, particularly if the competing models are non-nested.

## KEYWORDS

Bayesian, importance sampling, marginal-likelihood estimation, model-comparison, multimodel inference, survival analysis

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

## 1 | INTRODUCTION

Ecological research is often based on data from field observations, which can be plagued by varying degrees of unknown measurement error and missing information (Cressie et al., 2009; Martin et al., 2005). If these imperfections are not accounted for correctly, this can result in biased estimates of ecological parameters and potentially spurious inference (Williams et al., 2002). Bayesian inference is regularly advocated as a powerful approach when dealing with missing data, since it provides a coherent framework to account for and characterise the uncertainties associated with the missing information (Daniels & Hogan, 2008). It has thus become the default for many applied problems (Fragoso et al., 2018), particularly in the analysis of ecological systems (Clark et al., 2005).

Statistical analysis in ecology now relies heavily on model comparison techniques (Johnson & Omland, 2004) and multimodel inference (Burnham & Anderson, 2002; Harrison et al., 2018), which allow the researcher to formulate a set of competing statistical models and evaluate the relative strength of evidence in the data in support of alternative hypotheses (Plummer, 2008). As a result of the growth in popularity of Bayesian modelling for ecological research various approaches to model choice and comparison have been proposed, each with advantages and disadvantages. Popular methods generally fall into two categories: (1) penalised-loss functions such as the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002) and the Watanabe–Akaike Information Criterion (WAIC; Watanabe, 2010), which produce a rank score for each competing model; and (2) model-based approaches (Hooten et al., 2015) that focus on estimating the *marginal likelihood* of the data for a given model—this referring to the expected value of the likelihood function with respect to the prior(s). In the latter case, model comparisons can then be conducted by calculating either Bayes' Factors or posterior model weights (Hoeting et al., 1999; Jeffreys, 1961; Kass & Raftery, 1995). These latter approaches offer robust multimodel inference yet are often overlooked due to complexities in calculating the marginal likelihood.

Here, we highlight an approach to model comparison that estimates the marginal likelihood through a combination of Markov chain Monte Carlo (MCMC) and Importance Sampling (IS). This combined approach has been shown to work well compared with competing methods, particularly when dealing with high prevalence of missing data (Mckinley et al., 2020; Touloupou et al., 2018; Tran et al., 2014), a situation which has drawn concern when using the more established model-comparison methods (Celeux et al., 2006; Daniels & Hogan, 2008). Importance Sampling is a statistical method used to estimate properties of a specific probability distribution (the target distribution) by drawing weighted samples from an approximating distribution (referred to as the proposal or importance distribution), which is chosen to be more straightforward to sample from. Samples drawn from the importance distribution allow for the approximation of integrals or expected values that may be challenging to evaluate otherwise. We follow the two-stage approach of Touloupou et al. (2018), in which a model is fitted to observed

data via MCMC to collect posterior samples, which are then used to inform a tractable proposal distribution from which we can estimate the marginal likelihood. This allows for efficient estimation because samples are more likely to be drawn from regions of the parameter space with higher posterior density through the importance distribution (Tokdar & Kass, 2010). We can then perform a comprehensive analysis of the relative quality of competing models and perform multimodel inference of population parameters and their associations with predictor variables.

There are many methods for estimating marginal likelihoods (e.g. Gelfand & Dey, 1994; Zhou et al., 2016) and for a comparison between approaches please see, for example Touloupou et al. (2018). We present IS here because it is relatively straightforward to implement and has been shown to work well in situations with missing data (Touloupou et al., 2018). We also implement an alternative approach using reversible-jump MCMC (RJ-MCMC)—an extension to more conventional MCMC methods which allows a Markov chain to jump between different model configurations while simultaneously exploring parameter values within each model. RJ-MCMC is powerful, but can be difficult to code and optimise. It is best suited to situations where competing models are nested, such as for variable selection scenarios. In this latter case, it can be implemented in some general-purpose Bayesian software packages such as NIMBLE (de Valpine et al., 2017). The two-stage IS approach that we explore here can work well with non-nested models but requires each model to be fitted separately, and so is less-suited to variable selection scenarios with large numbers of variables, where the number of competing models can be very large. As statisticians/modellers we feel that it is useful to have a tool-box of techniques that can be useful in different contexts, and it is with that philosophy that we present the IS method as an alternative to RJ-MCMC, here. We also present DIC and WAIC scores which, although optimising different measures, serve to highlight that the choice of model comparison method is often context dependent and can favour different models.

We provide a straightforward linear regression example of the IS approach using simulated data in [Supplementary Material](#) that serves as a simple tutorial, but focus here on implementation of the approach to ecological data in the form of two case study investigations of survival in mammal populations.

The modelling and accurate estimation of survival is critical to many areas of wildlife research (e.g. evolutionary pressures in the wild—Roulin et al., 2010; conservation of populations—Morris & Doak, 2002; wildlife disease—Benton et al., 2018) but is reliant upon the analysis of repeated observations of individuals (Dey et al., 2019), which can be challenging in dynamic wild populations (Delahay et al., 2009). Generally, researchers gathering data at the individual level will strive for census information where missing data are rare; however, such situations are infrequent. A common approach is to use capture–mark–recapture (CMR), the repeated sampling of a population in which individuals are first marked and released, and, at each subsequent occasion they are either recaptured, not detected, or recovered dead (Catchpole et al., 1998). The resulting data are punctuated with missing information due to low

detection/recapture probabilities and/or censoring—when measurements or observations are only partially known. For example, if some deaths occur after the study period has ended, then those individuals provide only partial information about mortality by surviving at least until the end of the study (so-called right-censoring).

In survival analysis, problems with missing data and censoring are generally overcome by fitting a parametric model that adequately describes the data (Wilson, 1994), which can then be used to interpolate or extrapolate probabilities of events happening at some point during the censoring period. Many models have been developed for this purpose, including (Gompertz, 1825; Makeham, 1867; Siler, 1979). Historically, the Gompertz model was assumed to provide an adequate fit to data describing mortality for most mammalian species (Kirkwood, 2015), but there is now a growing understanding that mortality trajectories of wild animals can depart from this standard form (e.g. Colchero et al., 2019; Jones et al., 2014; Ronget et al., 2020). Best-fit parameter estimates for a single, preferred model will often be compared between populations to determine mortality and survival differences associated with genetic or environmental variation but the data from these different groups may be more suited and better fit by altogether different underlying functions (Wilson, 1994). There is a growing need for efficient and reliable methods to compare the fit of competing survival models (e.g. Larson et al., 2016) or perform multimodel inference.

In this paper, we aim to: (1) evaluate the performance of Bayesian model comparison and multimodel inference via marginal likelihood estimation (specifically using a combination of MCMC and IS) using two common formats of ecological data (census and CMR); and (2) broaden the appeal of this approach to ecologists for whom imperfect detection and sampling methodologies often result in incomplete data.

## 2 | MATERIALS AND METHODS

Mortality trajectories describe the pattern of mortality through an organism's lifespan. We demonstrate the IS method by comparing the results with the well-established model comparison metrics DIC and WAIC, as well as RJ-MCMC. DIC is defined as the difference between the expected deviance of the model and the effective number of parameters (Spiegelhalter et al., 2002). Watanabe-Akaike Information Criterion is an estimate of the expected log pointwise predictive density, corrected for overfitting (Watanabe, 2010). RJ-MCMC is an extension of standard MCMC methods, which operates by proposing simultaneous changes to the parameters and the model structure, yielding posterior inclusion probabilities for nested variants of the model itself (Green, 1995). RJ-MCMC can be used for non-nested models but often suffers with poor mixing, particularly as the models differ in terms of complexity and parameter definitions. For further details, see the [Supplementary Materials](#).

Prior to any model comparisons an important step is to ensure all candidate models are capable of giving rise to the data. Various approaches to model-checking are available (Conn et al., 2018; Morey

et al., 2013), but we note that validating models can be difficult when inferring the parameters of latent or partially observed processes like survival. With census data (where censoring is rare we are able to compare predictive survival curves with Kaplan-Maier plots of actual survival (see [Figure 4a](#)) but this is not possible with CMR data. We checked the validity of our models by simulating lifespans and capture histories using the inferred posterior distributions of parameters from our models, then comparing these posterior distributions to our observed data. We also check the mixing and convergence of the MCMC chains (see [Supplementary Material](#)).

Both case studies involve two implementations of our approach to model comparison: (1) compare the fits of four different mortality models to establish an underlying mortality pattern within each data set ([Table 1](#)); and (2), using the model(s) established in (1) to then investigate sex-specific variation. We note that these investigations could be combined (i.e. compare sex-specific variations of all models) but to facilitate the implementation of the RJ-MCMC in NIMBLE without the use of customised samplers, we chose to separate the steps here.

The exponential model assumes constant mortality throughout life independent of age; the Gompertz model (Gompertz, 1825) describes mortality as exponentially increasing with age; the Gompertz-Makeham (Makeham, 1867) is an extension of the Gompertz model with an additional age-independent mortality hazard parameter; and the Siler model (Siler, 1979) describes a 'bathtub' shaped mortality curve with an initial decline in mortality from a high intercept, then near-constant early- to mid-life mortality, followed by exponentially increasing mortality due to actuarial senescence. We structure competing models such that time of death is distributed according to the given mortality model(s) with each parameter then allowing (or not allowing) sex-specific variations. For example, when fitting the Siler model to census data:

$$t_j^D \sim \text{Siler}(a_{1j}^M, a_{2j}^M, b_{1j}^M, b_{2j}^M, c_j^M) \quad (1)$$

where  $t_j^D$  is time of death for individual  $j$  and each parameter is then model as:

$$\begin{aligned} \log(a_{1j}^M) &= \log(a_1) + \beta_1 \times \text{sex}_j \\ \log(a_{2j}^M) &= \log(a_2) + \beta_2 \times \text{sex}_j \\ \log(b_{1j}^M) &= \log(b_1) + \beta_3 \times \text{sex}_j \\ \log(b_{2j}^M) &= \log(b_2) + \beta_4 \times \text{sex}_j \\ \log(c_j) &= \log(c) + \beta_5 \times \text{sex}_j \end{aligned} \quad (2)$$

TABLE 1 Mortality functions used as proposal models to fit to the data.

Model	Mortality rate $\mu(x \theta)$	Parameters
Exponential	$r$	$r > 0$
Gompertz	$ae^{bx}$	$a, b > 0$
Gompertz-Makeham	$ae^{bx} + c$	$a, b, c > 0$
Siler	$a_1 e^{-b_1 x} + c + a_2 e^{b_2 x}$	$a_1, a_2, b_1, b_2, c > 0$

We code sex as a binary variable, taking the value 0 if an individual is female, and 1 if they are male; and  $\beta_{1,5}$  represent the coefficients for the effect of being male on each of the Siler parameters. In the above example (Equation 1 and Equation 2), the model would allow sex-specific variation on all parameters of the Siler model. For complete model definitions, see [Supplementary Material](#).

## 2.1 | Estimating the marginal likelihood

For a given model, Bayesian statistical inference estimates a posterior probability distribution,  $f(\theta | \mathbf{y})$ , for parameters  $\theta$  given data  $\mathbf{y}$ . The posterior distribution satisfies

$$f(\theta | \mathbf{y}) = \frac{f(\mathbf{y} | \theta) f(\theta)}{f(\mathbf{y})}, \quad (3)$$

where  $f(\mathbf{y} | \theta)$  is the likelihood function (the distribution of the data given the parameters) and  $f(\theta)$  is the prior distribution (representing our belief in the values of the parameters in the absence of data). The denominator,  $f(\mathbf{y})$ , is the *marginal likelihood*, and is defined as the (multidimensional) integral of the numerator of Equation (3) with respect to all parameters (or equivalently, it can be interpreted as the expected value of the likelihood function with respect to the prior distribution):

$$f(\mathbf{y}) = \int_{\theta} f(\mathbf{y} | \theta) f(\theta) d\theta. \quad (4)$$

For clarity, we can make explicit that all inferences from Equation (3) are in fact dependent on some model  $M$  i.e.

$$f(\theta | \mathbf{y}, M) = \frac{f(\mathbf{y} | \theta, M) f(\theta | M)}{f(\mathbf{y} | M)}. \quad (5)$$

In the context of model comparisons, we now consider a set of competing proposal models  $M_1, \dots, M_K$ .

**Step 1:** Fit the proposal model  $M_1$  to the data  $\mathbf{y}$ , estimating the posterior distribution  $f(\theta | \mathbf{y}, M_1)$  using MCMC.

**Step 2:** Find a suitable importance distribution for  $f(\theta | \mathbf{y}, M_1)$ . We do this by fitting a series of multivariate finite Gaussian mixture models of increasing complexity to the posterior samples from Step 1, select the best-fitting mixture model using, for example the Bayesian information criterion (BIC), and check that this gives a good approximation of the posterior density (Figure 1). To do this step, we used the R package 'mclust' (Scrucca et al., 2016). We use BIC because we only need to find a suitable, but tractable, approximation to the posterior, and BIC is easy to compute and favours simpler models than the more common Akaike's Information Criterion for this purpose. Letting  $q_{FMM}(\theta | M_1)$  be the probability density function of the finite Gaussian mixture model described above, we then define a 'defence mixture' (Hesterberg, 1995) importance distribution of the form:

$$q(\theta | M_1) = pq_{FMM}(\theta | M_1) + (1 - p)f(\theta | M_1) \quad (6)$$

where  $f(\theta | M_1)$  is the prior distribution. We set the mixing proportion,  $p = 0.95$  which is at the conservative end of the recommended range (Hesterberg, 1995). With a slight abuse of notation, we then

draw  $n$  random samples from the importance distribution  $q(\theta | M_1)$ , for  $i = 1, \dots, n$ . The defence mixture is used to ensure that the importance distribution is overdispersed with respect to the target distribution, and thus ensuring that the variance of the importance sample estimator is finite.

**Step 3:** We then estimate the marginal likelihood of the model as the mean, over sampled posterior values, of:

$$\hat{f}(\mathbf{y} | M_k) = \frac{1}{n} \sum_{i=1}^n \frac{f(\mathbf{y} | \theta_i, M_k) f(\theta_i | M_k)}{q(\theta_i | M_k)} \quad (7)$$

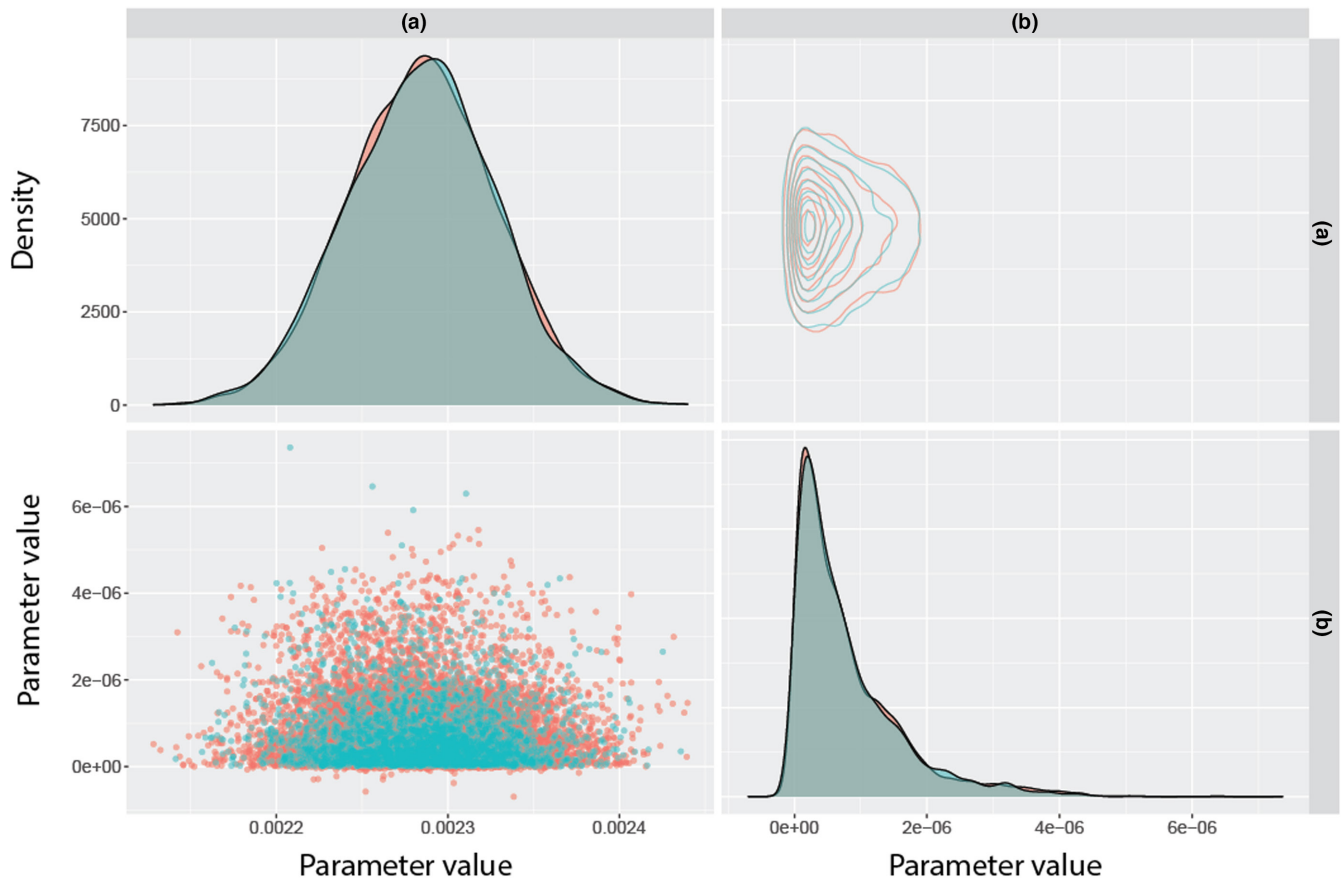
**Step 4:** We then repeat Steps 1–3 for any number of alternate proposal models  $M_2, \dots, M_K$ . We select the 'best' fitting model (i.e. the one with the highest log-marginal likelihood) and apply *Occam's Window*—a model selection approach which takes full account of the true model uncertainty by selecting a subset of models within a specific threshold of the 'best' model to then be used in subsequent Bayesian model averaging (Madigan & Raftery, 1994). In our analyses we select any model with a log-marginal likelihood value within  $\log(20)$  of the 'best' model (Kass & Raftery, 1995). We then calculate the posterior model probabilities of all models within the threshold as:

$$P(M_k | \mathbf{y}) = \frac{\hat{f}(\mathbf{y} | M_k) P(M_k)}{\sum_{i=1}^K \hat{f}(\mathbf{y} | M_i) P(M_i)} \quad (8)$$

where  $P(M_k)$  is the prior probability weight for model  $M_k$  such that  $\sum_{k=1}^K P(M_k) = 1$ . It is common to set  $P(M_k) = \frac{1}{K}$ , thus assuming that there is no a priori reason to prefer one model over another (although this can be changed according to prior knowledge). Hence, posterior model weights are normalised to sum to one over the set of selected models, see Kass and Raftery (1995) and Madigan and Raftery (1994).

## 2.2 | Case study 1: Census data (banded mongoose survival)

The data collection is licensed by the Uganda National Council for Science and Technology and approved by both the Uganda Wildlife Authority and the University of Exeter's Ethical Review Committee. We analysed life history data from a habituated population of wild banded mongooses (*Mungos mungo*) living on and around the Mweya Peninsula in Queen Elizabeth National Park, western Uganda (for further details see the [Supplementary Materials](#)). We only included individuals with a known birth date (removing 140 individuals from a total of 3380), and all death dates were modelled as right- or interval-censored. Mongoose pups that died prior to being sexed were included in the model, with the unknown sex becoming an additional latent variable inferred by the model. In this instance, it is reasonable to assume that the probability of a pup not being sexed was independent of its sex, and as such we used a Bernoulli distribution to capture whether individuals were male or female, and used a Uniform prior distribution, bounded by zero and one, on the probability of being male. The ability to include missing covariate information and



**FIGURE 1** Example pairs plot of the posterior samples from MCMC (blue) showing sample values and distribution density from the importance distribution  $q_{FM}(\theta | M_1)$  (red) generated from the selected mixture model overlaid. This example is for the Gompertz model fitted to the banded mongoose case study data.

robustly characterise the uncertainty associated with the missing information emphasises the flexibility of the Bayesian approach and avoids data being discarded unnecessarily.

### 2.3 | Case study 2: Capture–mark–recapture data (European badger survival)

The capture, examination and sampling of live badgers was carried out under Home Office Project Licence PP3493437 and preceding versions of this licence. We made use of CMR data from a long-term monitoring project of a population of wild European badgers (*Meles meles*) in Woodchester Park, Gloucestershire (for further details see McDonald et al., 2018). The badger population is sampled using live traps on (usually) four occasions per year with all trapped badgers anaesthetized and samples taken for several diagnostic tests for infection with *Mycobacterium bovis* before being released. On first capture, each badger is given a unique tattoo so it can be identified without error when subsequently captured. We used badgers of known age (i.e. badgers caught and identified as cubs or yearlings, thus removing 45 individuals captured at age >1 year from a total of 2786). The difference in handling CMR data as opposed to the mongoose census data lies in the definition of

the likelihood function—as a result of the sampling process, each individual will contribute different information to the survival estimation. Figure 2 represents the different individual scenarios that appear in our analysed dataset: individuals 1, 3 and 4 are captured and identified as cub; individual 2 dies prior to being captured and so contributes nothing to survival estimation; individuals 1 and 4 are both considered right censored as we only have information about survival up to a certain point (last captured alive); individual 3 contributes the most information and is considered interval censored (died between interval 6 and 7).

### 2.4 | Priors

We specified weakly informative exponential distributions (rate=1) for the priors of the model parameters. Bayes' Factors are sometimes criticised for their sensitivity to the priors used so we repeated the analyses using more diffuse exponential distributions (rate=0.1); this had negligible effect on any outcomes of the model comparisons, so we present the Exp(1) results here. For the RJ-MCMC analysis (see Supplementary Material), we set weakly informative exponential (rate=1) priors on the model parameters; the recapture rate had a uniform (0, 1) prior; the inclusion indicators for each parameter were



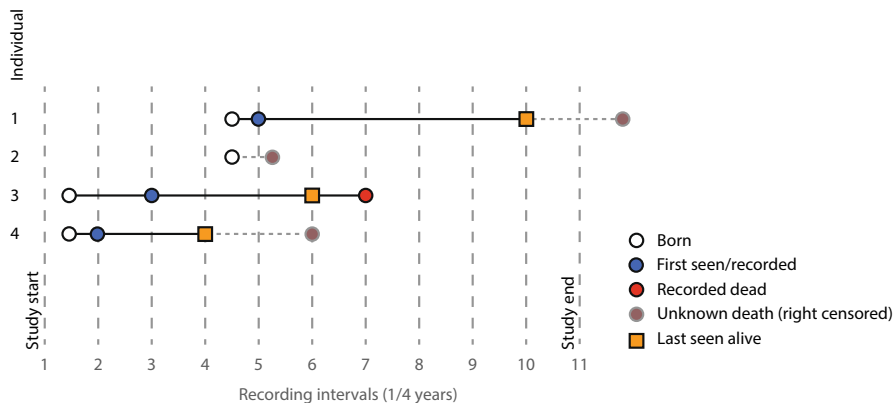


FIGURE 2 Representation of the types of censorship evident in the badger dataset.

set with a Bernoulli ( $p=0.5$ ) distribution; and all beta coefficients were set with a weakly informative Normal (mean=0, SD=1) priors.

### 3 | RESULTS

#### 3.1 | Case study 1: Banded mongoose

The IS approach selected the Siler model as the 'best' fitting underlying mortality model (Figure S1), which we carried forward; no other model was within the Occam's Window threshold (Kass & Raftery, 1995) to be included. DIC and WAIC scores were consistent with the IS approach (Table S1). We were unable to achieve satisfactory mixing in an RJ-MCMC analysis for the initial model comparisons: We believe this could be achieved by writing custom sampling algorithms but such methods fall beyond the scope of this paper.

We compared the 32 possible variants of the Siler model allowing (and not allowing) sex-specific variation on every combination of parameters. The IS approach selected a model with sex-specific variation on parameters  $b_1$  and  $b_2$  as the 'best' model but with nearly all other variants within the threshold to be considered (Figure 3a). We calculated posterior model probabilities for the competing models, which we compared to the results of the RJ-MCMC analysis and found almost identical agreement between the methods (Figure 3b; Table 2). Finally, we calculated DIC and WAIC (Table S2) and found mixed levels of agreement (Table 2). The WAIC scores selected the same model as the IS approach as the most supported (sex variation on  $b_1$ ,  $b_2$ ), with four additional models within 2 units (Gelman et al., 2014) although only one of these featured in the top five models according to the IS approach and the RJ-MCMC. DIC scores were a lot less consistent: the model allowing sex-specific variation on parameters  $a_2$  and  $b_1$  was the most supported with the next best model (sex variation on  $a_2$ ,  $b_2$ ) 2.06 units away. The 'best' model identified by all the other approaches (sex variation on  $b_1$ ,  $b_2$ ) ranked third. The individual differences in the majority of DIC and WAIC scores between competing models are consistently low suggesting there is little to choose between the models in terms of these measures. To represent the final stage of such an analysis, we removed the models that fell outside the log(20) threshold and recalibrated the posterior

model probabilities to then produce model-averaged, posterior-predictive survival and mortality trajectories, which can be compared with Kaplan–Maier survival plots of the data (Figure 4).

The census style data of the banded mongooses analysed here contains minimal missing information and the IS approach, RJ-MCMC and WAIC all chose the same model as the 'best' performing. The results of these three techniques also agreed there was little to choose between the majority of the other proposal models which strengthens the argument for a multimodel inference approach to predictions. DIC selected a different 'best' model outright, if using established convention, but the differences between scores for the remaining models were small.

#### 3.2 | Case study 2: European badger

In the initial comparison of the four potential baseline mortality models the IS approach selected the Gompertz–Makeham model as the most suitable, with the Siler model lying just outside the log(20) threshold (Figure S2). WAIC agreed with the IS approach but DIC preferred the Siler model (Table S3). As with the first case study, we were unable to achieve satisfactory mixing in an RJ-MCMC analysis for the initial model comparisons. We note that early-life mortality is unobserved if pups die underground, and so it could be that the preference for the GM over the Siler may be due to identifiability issues with the latter, caused by missing early-life mortality data (Hudson et al., 2023).

We constructed the eight variants of the Gompertz–Makeham model allowing (and not allowing) sex-specific variation on every combination of the model parameters. The IS approach chose the model allowing sex-specific variation on  $c$  as the best model with three of the other proposal models within the log(20) threshold to be included (Figure 5a).

We calculated posterior model probabilities for the competing models, which we compared with the results of the RJ-MCMC analysis and found almost identical agreement between the methods (Figure 5b, Table 3), the four models that returned log marginal likelihood values outside the log(20) threshold were not successfully visited by the RJ-MCMC analysis (the chains do visit these models but negligible probabilities means the jumps are not

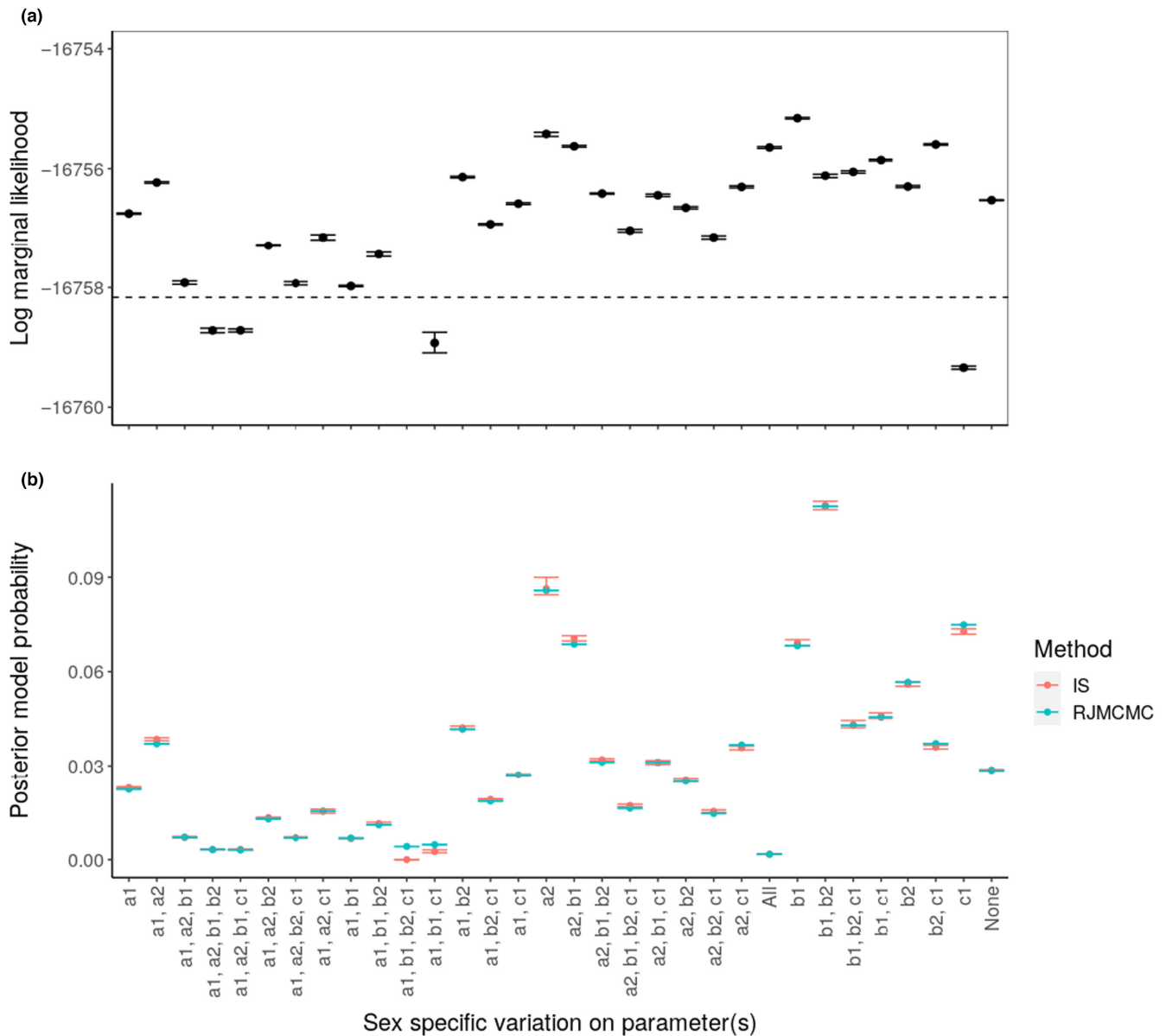


FIGURE 3 (a) Log-marginal likelihood values generated by a method utilising Importance Sampling; (b) Comparison of posterior model probabilities generated by a method utilising IS (red) and RJ-MCMC (blue) of 32 competing Siler mortality models allowing sex-specific variation on different combinations of parameters fitted to survival data. Dashed line in a. represents a value  $\log(20)$  less than the most supported model (Kass & Raftery, 1995). Samples were bootstrapped 1000 times to provide 95% confidence intervals.

retained) corresponding to an estimated posterior model probability of zero. We found identical results with the IS approach for WAIC (Table 3) but again DIC scores were less consistent—DIC suggesting the most supported model allowed sex-specific variation on  $a$  and  $b$ , a model completely dismissed by all other approaches (Table 3; Table S4).

We follow the Occam's window approach (Kass & Raftery, 1995) to construct posterior-predictive model-averaged survival and mortality trajectories using all of the proposal models within the  $\log(20)$  threshold of the best fitting model (Figure 6). The posterior model probabilities are recalibrated with only the four selected models included.

## 4 | DISCUSSION

We have translated an accurate and accessible method to carry out Bayesian model comparisons for ecological analyses using an efficient two-stage approach to the estimation of the marginal likelihood based on Importance Sampling. This approach can be used to generate posterior model probabilities thus making multi-model inference more accessible for Bayesian models. Our analyses have focussed on survival trajectory analysis where data sets will often include missing or incomplete data but the method is equally applicable in any Bayesian investigation requiring model comparison.

**TABLE 2** Rank order of models analysing sex-specific mortality variation in banded mongoose using census data and four different model comparison approaches: marginal likelihood estimation via IS, RJ-MCMC (variable selection), Watanabe-Akaike Information Criterion (WAIC) and Deviance Information Criterion (DIC).

Model	Importance sampling	RJ-MCMC	WAIC	DIC
Sex diff on $b_1$ $b_2$	1	1	1	3
Sex diff on $a_2$	2	2	19	19
Sex diff on $c_1$	3	3	23	24
Sex diff on $a_2$ $b_1$	4	4	4	1
Sex diff on $b_1$	5	5	13	15
Sex diff on $b_2$	6	6	26	12
Sex diff on $b_1$ $c_1$	7	7	8	13
Sex diff on $b_1$ $b_2$ $c_1$	8	8	2	4
Sex diff on $a_1$ $b_2$	9	9	7	7
Sex diff on $a_1$ $a_2$	10	11	5	9
Sex diff on $b_2$ $c_1$	11	10	25	25
Sex diff on $a_2$ $c_1$	12	12	21	21
Sex diff on $a_2$ $b_1$ $b_2$	13	13	3	8
Sex diff on $a_2$ $b_1$ $c_1$	14	14	9	5
No sex diff	15	15	28	28
Sex diff on $a_1$ $c_1$	16	16	18	22
Sex diff on $a_2$ $b_2$	17	17	24	2
Sex diff on $a_1$	18	18	22	6
Sex diff on $a_1$ $b_2$ $c_1$	19	19	20	26
Sex diff on $a_2$ $b_1$ $b_2$ $c_1$	20	20	10	14
Sex diff on $a_2$ $b_2$ $c_1$	21	22	27	27
Sex diff on $a_1$ $a_2$ $c_1$	22	21	17	20
Sex diff on $a_1$ $a_2$ $b_2$	23	23	11	18
Sex diff on $a_1$ $b_1$ $b_2$	24	24	6	10
Sex diff on $a_1$ $a_2$ $b_1$	25	25	12	16
Sex diff on $a_1$ $a_2$ $b_2$ $c_1$	26	26	14	17
Sex diff on $a_1$ $b_1$	27	27	16	11
Sex diff on $a_1$ $a_2$ $b_1$ $c_1$	28	31	15	23
Sex diff on $a_1$ $a_2$ $b_1$ $b_2$	29	30	31	32
Sex diff on $a_1$ $b_1$ $c_1$	30	28	29	29
Sex diff on all parameters	31	32	30	31
Sex diff on $a_1$ $b_1$ $b_2$ $c_1$	32	29	32	30

The mongoose data are derived from census and hence do not suffer a great deal of missingness. We found support for multiple models that allow sex-specific variation in survival but the model that did not allow any sex-specific variation was also within the threshold for inclusion and therefore cannot be dismissed. Having produced model-averaged, posterior predictive survival curves, our analysis suggests that males live slightly longer than females on average, which supports preliminary analyses from Cant et al. (2016). In the analysis of the CMR badger data, the best fitting model chosen

by IS was one allowing sex-specific variation on the age-independent parameter  $c$  which serves to raise or lower the overall trajectory. We found considerable evidence for sex-specific variation in mortality with all of the models within the threshold of the best fitting model allowing sex-specific variation on at least one parameter and the Gompertz–Makeham model with no sex variation dismissed with a posterior model probability less than one—matching previous analyses of the population (e.g. Hudson et al., 2019).

We compared results of the IS approach from both case-studies to DIC and WAIC scores and where possible RJ-MCMC. Although results were broadly the same as that achieved with RJ-MCMC, WAIC and in particular DIC were less consistent.

DIC has faced some criticism in the literature and can perform poorly for some types of model—for example mixture models (Celeux et al., 2006) or latent state models (Pooley & Marion, 2018) such as those analysed here (summarised in Spiegelhalter et al., 2014). This may be due to its reliance on point estimates of the parameters (Celeux et al., 2006), we investigated the posterior parameter distributions of the competing models and found some evidence of multimodality suggesting that a single point estimate (i.e. the posterior mean) is unlikely to be a robust choice in this case.

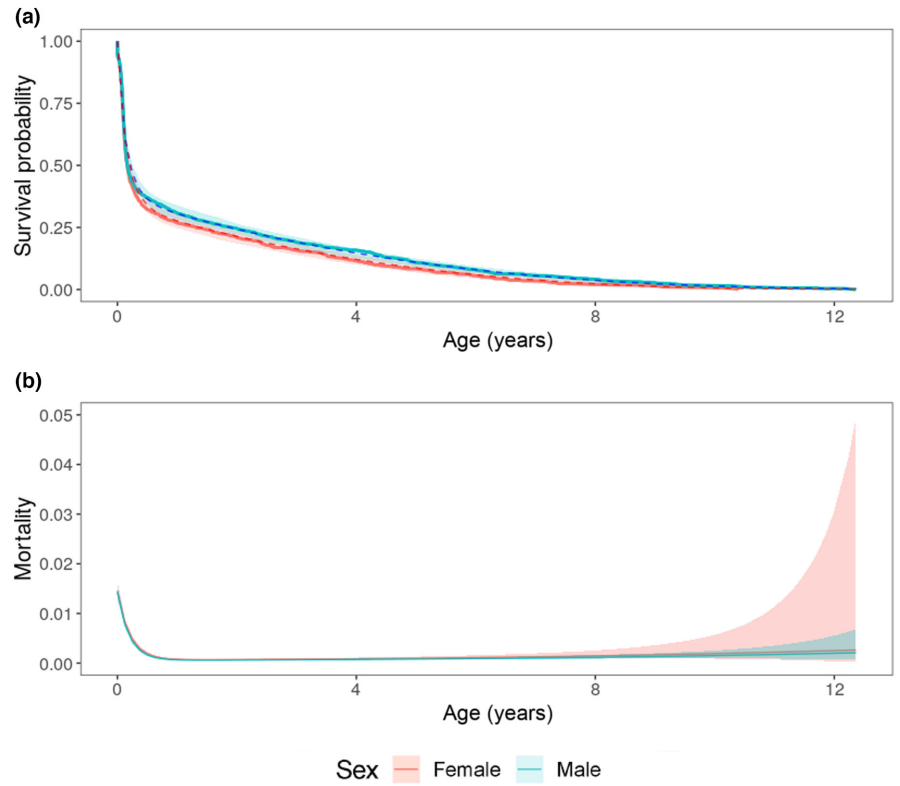
WAIC is regarded as a fully Bayesian approach because it employs measures that average across the posterior distribution and is often considered the preferred information criterion in Bayesian analyses (Gelman et al., 2013). However, it can still encounter difficulties in specific scenarios as it depends on a data partition that may pose challenges for structured models, such as those involving spatial or network data (Gelman et al., 2013). The estimates also contain a random error term which can have a large variance if the dataset is small which could lead to overfitting in the selection process (Cawley & Talbot, 2010). This becoming particularly problematic when comparing large numbers of proposal models such as in a variable selection scenario as we have presented here (Piironen & Vehtari, 2017).

Our analyses, and all of the model comparison approaches, have shown that often there are multiple models that can adequately describe the data-generating processes. To select one model outright would ignore this model uncertainty and potentially lead to specious inference which is why multimodel inference is becoming more commonly used (Piironen & Vehtari, 2017; Raftery et al., 2011).

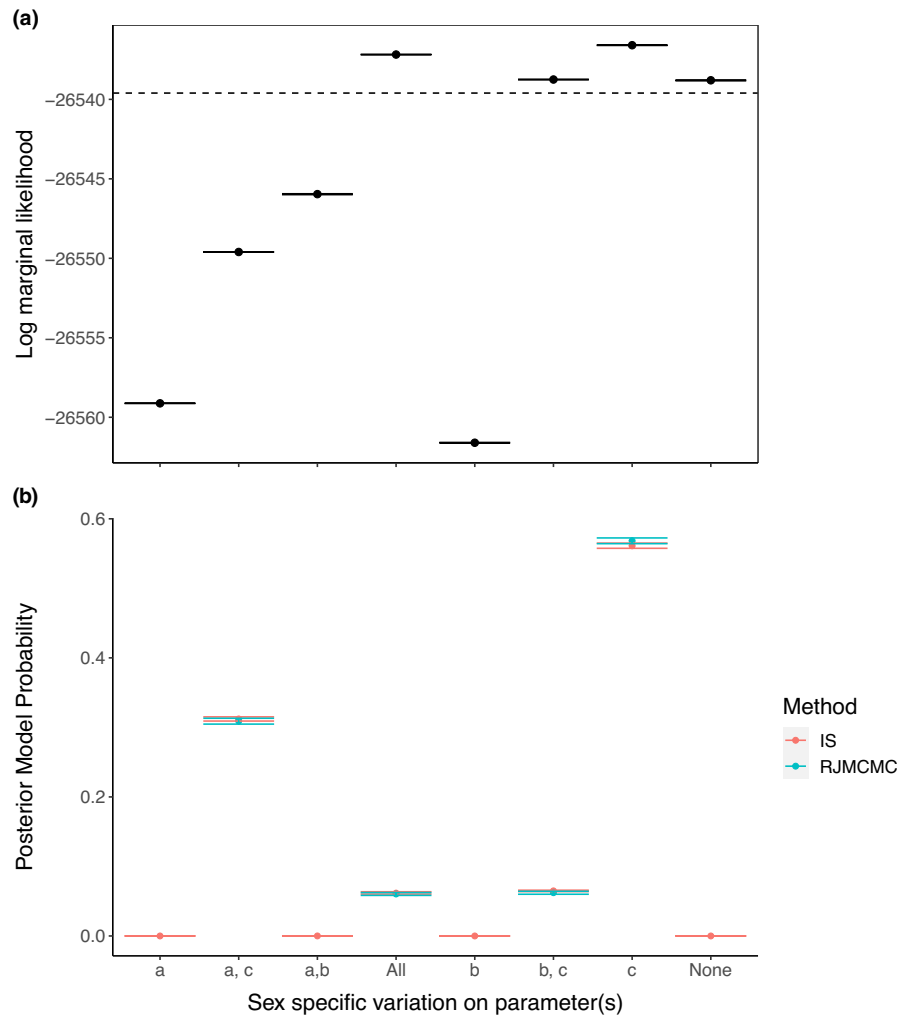
RJ-MCMC offers an alternative approach to IS for calculating posterior model weights (Green, 1995) but can be challenging to implement (see Brooks et al., 2003; Han & Carlin, 2001 for a review) particularly when competing models are very distinct in terms of the number and interpretation of parameters: in such cases, the MCMC can be difficult to tune and can suffer poor mixing, resulting in the need for very long runs—particularly for complex mixtures of models. This was problematic in our initial model comparisons even though in this instance the four models we compared can be considered nested—we failed to achieve satisfactory mixing and so were not able to use the RJ-MCMC approach for this set of comparisons. We could have spent time developing customised samplers to solve this problem, but we intended this manuscript to focus on the IS method



**FIGURE 4** (a) Kaplan–Maier plots of recorded sex-specific survival (solid line) with model-averaged, posterior-predictive survival trajectories overlaid (dashed lines), shaded areas indicating 95% credible intervals and (b) Model-averaged posterior-predictive mortality trajectories for a sex-specific analysis of census data from a wild population of banded mongooses fitted to sex-specific variations of the Siler mortality model. Shaded areas represent 95% credible intervals.



**FIGURE 5** (a) Log-marginal likelihood values generated by a method utilising Importance Sampling; (b) Comparison of posterior model probabilities generated by a method utilising Importance Sampling (red) and reversible-jump Markov chain Monte Carlo (blue) of eight competing Gompertz–Makeham mortality models allowing sex-specific variation on different combinations of parameters fitted to survival data. Dashed line in a. represents a value  $\log(20)$  less than the most supported model. Samples were bootstrapped 1000 times to provide 95% confidence intervals.



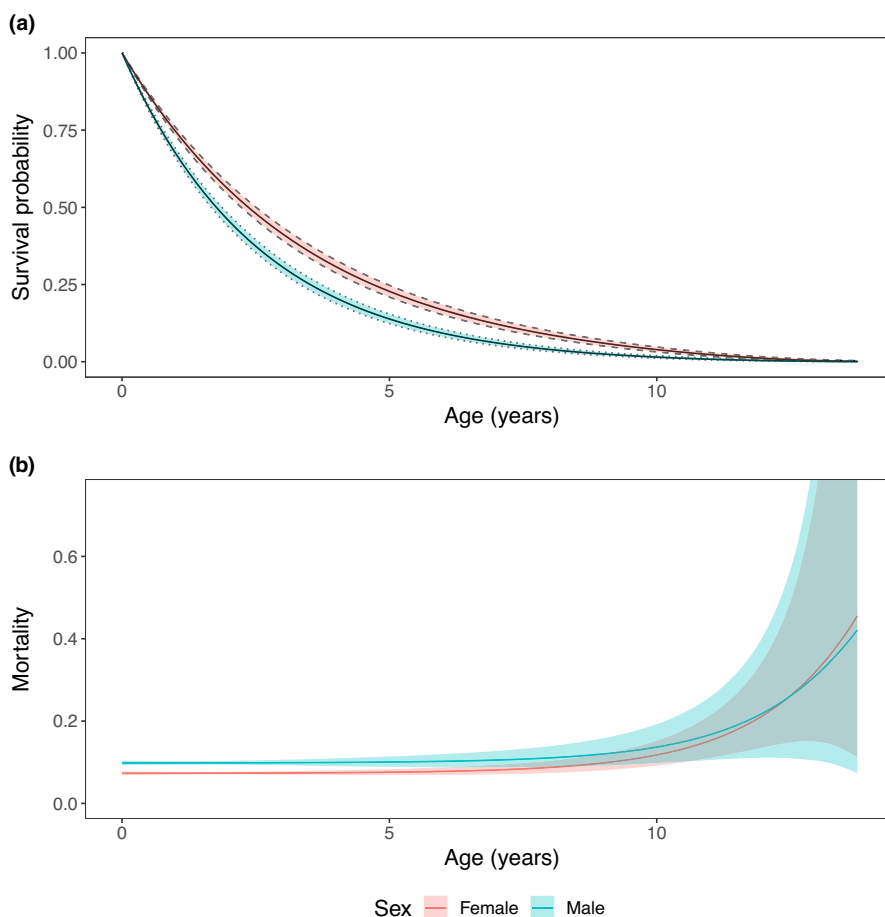
and so did not pursue this here. Model selection problems can often reduce to a simpler framework of variable selection such as in our sex-specific analyses, where the question becomes: which subset of variables should be included within a model? RJ-MCMC performs

**TABLE 3** Rank order of models analysing sex-specific mortality variation in a population of European badgers (*Meles meles*) using capture-mark-recapture data and four different model comparison approaches: Marginal likelihood estimation via Importance Sampling, reversible-jump Markov chain Monte Carlo, Widely Applicable Information Criterion (WAIC) and Deviance Information Criterion (DIC) (for WAIC and DIC scores see [Supplementary Materials](#)).

Model	Importance sampling	RJ-MCMC	WAIC	DIC
Sex diff on c	1	1	1	2
Sex diff on a,c	2	2	2	4
Sex diff on b, c	3	3	3	3
Sex diff on a, b, c	4	4	4	5
Sex diff on a, b	5	Dismissed	5	1
Sex diff on a	6	Dismissed	6	6
No sex diff	7	Dismissed	7	7
Sex diff on b	8	Dismissed	8	8

particularly well in these instances where models are nested and transition rules are simpler to define (O'Hara & Sillanpaa, 2009; Touloupou et al., 2018). But, IS does not require the competing models to be nested—had either of the initial model comparisons selected two models within threshold meaning they should both be considered in the sex-specific analyses then RJ-MCMC would have likely failed custom samplers. In contrast, the IS approach is applied post hoc to the separate analysis of each competing model meaning the situation could easily be considered as long as the number of competing models is not prohibitive to evaluate.

Criticisms of marginal likelihood estimation approaches to model comparison using Bayes' Factors often focus on their sensitivity to the choice of priors when analysing small data sets (Kass, 1993) and challenges in their calculation (Xie et al., 2011). We found negligible differences with alternative priors when investigating both datasets but would recommend this safety check. Efficient calculation of the marginal likelihood remains an ongoing area of research in statistics (Wang et al., 2018) but here we have demonstrated the application of a versatile two-stage approach, which allows for flexible and robust model comparisons. A recently proposed alternative approach is to employ *posterior predictive stacking* that uses out-of-sample predictive measures, most notably Leave-One-Out Cross Validation measures, as an alternative way to derive model weights, and this is an interesting avenue for future research (Yao et al., 2018).



**FIGURE 6** Model-averaged posterior-predictive (a) survival and (b) mortality trajectories for a sex-specific analysis of capture-mark-recapture data from a wild population of European badgers (*Meles meles*) fitted to sex-specific variations of the Gompertz-Makeham mortality model. Shaded areas represent 95% credible intervals.

## 5 | CONCLUSIONS

Advancements in computing and software now enable Bayesian approaches to fit complex models routinely (Friel & Wyse, 2012). Neglecting model uncertainty can result in overconfident inferences (Parrish et al., 2012), strengthening the case for model averaging. We have described a clear and flexible method to estimate the marginal likelihood of any Bayesian model using Importance Sampling which facilitates straightforward model comparison through Bayes' Factors or posterior model probabilities. It proves especially valuable in ecological studies, notably in survival analysis, where missing data and censoring are common.

### AUTHOR CONTRIBUTIONS

Dave W. Hudson, Dave J. Hodgson and Tevelyan J. McKinley conceived the ideas; Dave W. Hudson and Tevelyan J. McKinley designed methodology and ran the analyses; Michael A. Cant, Faye J. Thompson, Richard Delahay and Robbie A. McDonald are owners/curators of the data; Dave W. Hudson, Dave J. Hodgson and Tevelyan J. McKinley led the writing of the manuscript. All authors reviewed the manuscript and gave final approval for publication.

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

### PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14237>.

### DATA AVAILABILITY STATEMENT

All code and data presented in this manuscript are available via Zenodo (Hudson, 2023).

### ORCID

Dave W. Hudson  <https://orcid.org/0000-0001-5150-0150>

Dave J. Hodgson  <https://orcid.org/0000-0003-4220-2076>

Faye J. Thompson  <https://orcid.org/0000-0001-7581-2204>

Robbie A. McDonald  <https://orcid.org/0000-0002-6922-3195>

### REFERENCES

- Benton, C. H., Delahay, R. J., Smith, F. A. P., Robertson, A., McDonald, R. A., Young, A. J., Burke, T. A., & Hodgson, D. (2018). Inbreeding intensifies sex- and age-dependent disease in a wild mammal. *Journal of Animal Ecology*, 87(6), 1500–1511. <https://doi.org/10.1111/1365-2656.12878>
- Brooks, S. P., Giudici, P., & Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 3–39. <https://doi.org/10.1111/1467-9868.03711>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed.). Springer.
- Cant, M. A., Nichols, H. J., Thompson, F. J., & Vitikainen, E. (2016). Banded mongooses: Demography, life history, and social behavior. In W. Koenig & J. Dickinson (Eds.), *Cooperative breeding in vertebrates: Studies of ecology, evolution, and behavior* (pp. 318–337). Cambridge University Press. <https://doi.org/10.1017/CBO9781107338357.019>
- Catchpole, E. A., Freeman, S. N., Morgan, B. J. T., & Harris, M. P. (1998). Integrated recovery/recapture data analysis. *Biometrics*, 54, 33. <https://doi.org/10.2307/2533993>
- Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11, 2079–2107. <https://www.jmlr.org/papers/volume11/cawley10a/cawley10a>
- Celeux, G., Forbes, F., Robert, C. P., & Titterton, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4), 651–674.
- Clark, J. S., Ferraz, G., Oguge, N., Hays, H., & DiCostanzo, J. (2005). Hierarchical Bayes for structured, variable populations: From recapture data to life-history prediction. *Ecology*, 86, 2232–2244. <https://doi.org/10.1890/04-1348>
- Colchero, F., Jones, O. R., Conde, D. A., Hodgson, D., Zajitschek, F., Schmidt, B. R., Malo, A. F., Alberts, S. C., Becker, P. H., Bouwuis, S., Bronikowski, A. M., Vleeschouwer, K. M. D., Delahay, R. J., Dummermuth, S., Fernández-Duque, E., Frisenvænge, J., Hesselsøe, M., Larson, S., Lemaître, J. F., ... Gaillard, J. M. (2019). The diversity of population responses to environmental change. *Ecology Letters*, 22(2), 342–353. <https://doi.org/10.1111/ELE.13195>
- Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, 88(4), 526–542. <https://doi.org/10.1002/ecm.1314>
- Cressie, N., Calder, C. A., Clark, J. S., Hoef, J. M. V., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, 19(3), 553–570. <https://doi.org/10.1890/07-0744.1>
- Daniels, M. J., & Hogan, J. W. (2008). Missing data in longitudinal studies: Strategies for bayesian modeling and sensitivity analysis.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413. <https://doi.org/10.1080/10618600.2016.1172487>
- Delahay, R. J., Smith, G. C., & Hutchings, M. R. (Eds.). (2009). The science of wildlife disease management. *Management of disease in wild mammals* (pp. 1–8). Springer. [https://doi.org/10.1007/978-4-431-77134-0\\_1](https://doi.org/10.1007/978-4-431-77134-0_1)
- Dey, S., Delampady, M., & Gopalaswamy, A. M. (2019). Bayesian model selection for spatial capture–recapture models. *Ecology and Evolution*, 9(20), 11569–11583. <https://doi.org/10.1002/ece3.5551>
- Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86, 1–28. <https://doi.org/10.1111/insr.12243>
- Friel, N., & Wyse, J. (2012). Estimating the evidence—A review. *Statistica Neerlandica*, 66(3), 288–308. <https://doi.org/10.1111/j.1467-9574.2011.00515.x>
- Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56, 501–514. <https://doi.org/10.1111/j.2517-6161.1994.tb01996.x>
- Gelman, A., Carlin, J. B. B., Stern, H. S. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. B. (2014). Bayesian data analysis, third edition (texts in statistical science). In Book (third). Taylor and Francis Inc. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Gelman, A., Hwang, J., & Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*

- 2013 24:6, 24(6), 997–1016. <https://doi.org/10.1007/S1122-013-9416-2>
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new model of determining life contingencies. *Philosophical Transactions of the Royal Society of London*, 115, 513–585.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711–732. <https://doi.org/10.1093/biomet/82.4.711>
- Han, C., & Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors. *Journal of the American Statistical Association*, 96(455), 1122–1132. <https://doi.org/10.1198/016214501753208780>
- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794. <https://doi.org/10.7717/peerj.4794>
- Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37, 185–194. <https://doi.org/10.1080/00401706.1995.10484303>
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4), 382–417. <https://doi.org/10.1214/ss/1009212519>
- Hooten, M. B., Hobbs, N. T., & Ellison, A. M. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, 85, 3–28. <https://doi.org/10.1890/14-0661.1>
- Hudson, D. W. (2023). davehudson67/ImportanceSamplingCode: v1.0 (v1.0). *Zenodo*. <https://doi.org/10.5281/zenodo.8414128>
- Hudson, D. W., Delahay, R., McDonald, R. A., McKinley, T. J., & Hodgson, D. J. (2019). Analysis of lifetime mortality trajectories in wildlife disease research: BaSTA and beyond. *Diversity* 2019, 11(10), 182. <https://doi.org/10.3390/D11100182>
- Hudson, D. W., McKinley, T. J., Benton, C. H., Delahay, R., McDonald, R. A., & Hodgson, D. J. (2023). Multi-locus homozygosity promotes actuarial senescence in a wild mammal. *Journal of Animal Ecology*, 92, 1881–1892. <https://doi.org/10.1111/1365-2656.13979>
- Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford University Press.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*, 19(2), 101–108. <https://doi.org/10.1016/j.tree.2003.10.013>
- Jones, O. R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C. G., Schaible, R., Casper, B. B., Dahlgren, J. P., Ehrlén, J., García, M. B., Menges, E. S., Quintana-Ascencio, P. F., Caswell, H., Baudisch, A., & Vaupel, J. W. (2014). Diversity of ageing across the tree of life. *Nature*, 505(7482), 169–173. <https://doi.org/10.1038/nature12789>
- Kass, R. E. (1993). Bayes factors in practice. *The Statistician*, 42, 551. <https://doi.org/10.2307/2348679>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kirkwood, T. B. L. (2015). Deciphering death: A commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140379. <https://doi.org/10.1098/rstb.2014.0379>
- Larson, S. M., Colchero, F., Jones, O. R., Williams, L., & Fernandez-Duque, E. (2016). Age and sex-specific mortality of wild and captive populations of a monogamous pair-bonded primate (*Aotus azarae*). *American Journal of Primatology*, 78, 315–325. <https://doi.org/10.1002/ajp.22408>
- Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89(428), 1535–1546. <https://doi.org/10.1080/01621459.1994.10476894>
- Makeham, W. M. (1867). On the law of mortality. *Journal of the Institute of Actuaries*, 13(6), 325–358. <https://doi.org/10.1017/s204616660003238>
- Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J., & Possingham, H. P. (2005). Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8, 1235–1246. <https://doi.org/10.1111/j.1461-0248.2005.00826.x>
- McDonald, J. L., Robertson, A., & Silk, M. J. (2018). Wildlife disease ecology from the individual to the population: Insights from a long-term study of a naturally infected European badger population. *Journal of Animal Ecology*, 87(1), 101–112. <https://doi.org/10.1111/1365-2656.12743>
- McKinley, T. J., Neal, P., Spencer, S. E. F., Conlan, A. J. K., & Tiley, L. (2020). Efficient Bayesian model choice for partially observed processes: With application to an experimental transmission study of an infectious disease. *Bayesian Analysis*, 15, 839–870. <https://doi.org/10.1214/19-BA1174>
- Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2013). The humble Bayesian: Model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, 66(1), 68–75. <https://doi.org/10.1111/j.2044-8317.2012.02067.x>
- Morris, W. F., & Doak, D. F. (2002). *Quantitative conservation biology. In Theory and practice of population viability analysis*. Sinauer Associates.
- O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, How and which. *Bayesian Analysis*, 4(1), 85–118.
- Parrish, M. A., Moradkhani, H., & Dechant, C. M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research*, 48, W03519. <https://doi.org/10.1029/2011WR011116>
- Piironen, J., & Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27(3), 711–735. <https://doi.org/10.1007/S11222-016-9649-Y/FIGURES/12>
- Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3), 523–539. <https://doi.org/10.1093/biostatistics/kxm049>
- Pooley, C. M., & Marion, G. (2018). Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science*, 5, 171519. <https://doi.org/10.1098/RSO5.171519>
- Raftery, A. E., Zheng, Y., Hjort, N. L., & Claeskens, G. (2011). Discussion. *Journal of the American Statistical Association*, 98(464), 931–938. <https://doi.org/10.1198/016214503000000891>
- Ronget, V., Lemaître, J. F., Tidière, M., & Gaillard, J. M. (2020). Assessing the diversity of the form of age-specific changes in adult mortality from captive mammalian populations. *Diversity*, 12(9), 354. <https://doi.org/10.3390/D12090354>
- Roulin, A., Altwegg, R., Jensen, H., Steinsland, I., & Schaub, M. (2010). Sex-dependent selection on an autosomal melanic female ornament promotes the evolution of sex ratio bias. *Ecology Letters*, 13, 616–626. <https://doi.org/10.1111/j.1461-0248.2010.01459.x>
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R Journal*, 8(1), 289–317. <https://doi.org/10.32614/rj-2016-021>
- Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 60, 750–757. <https://doi.org/10.2307/1936612>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639. <https://doi.org/10.1111/1467-9868.00353>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3), 485–493. <https://doi.org/10.1111/rssb.12062>
- Tokdar, S. T., & Kass, R. E. (2010). Importance sampling: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 54–60.

[https://doi.org/10.1002/WICS.56@10.1002/\(ISSN\)1939-0068.STATISTICS](https://doi.org/10.1002/WICS.56@10.1002/(ISSN)1939-0068.STATISTICS)

- Touloupou, P., Alzahrani, N., Neal, P., Spencer, S. E. F., & McKinley, T. J. (2018). Efficient model comparison techniques for models requiring large scale data augmentation. *Bayesian Analysis*, 13(2), 437–459. <https://doi.org/10.1214/17-BA1057>
- Tran, M.-N., Scharth, M., Pitt, M. K., & Kohn, R. (2014). Importance sampling squared for Bayesian inference in latent variable models. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2386371>
- Wang, Y. B., Chen, M. H., Kuo, L., & Lewis, P. O. (2018). A new Monte Carlo method for estimating marginal likelihoods. *Bayesian Analysis*, 13(2), 311–333. <https://doi.org/10.1214/17-BA1049>
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
- Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations: Modeling, estimation, and decision making*. Academic Press.
- Wilson, D. L. (1994). The analysis of survival (mortality) data: Fitting Gompertz, Weibull, and logistic functions. *Mechanisms of Ageing and Development*, 74(1–2), 15–33. [https://doi.org/10.1016/0047-6374\(94\)90095-7](https://doi.org/10.1016/0047-6374(94)90095-7)
- Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M. H. (2011). Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Systematic Biology*, 60(2), 150–160. <https://doi.org/10.1093/sysbio/syq085>
- Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with Discussion), 13(3), 917–1007. <https://doi.org/10.1214/17-BA1091>
- Zhou, Y., Johansen, A. M., & Aston, J. A. D. (2016). Toward automatic model comparison: An adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3), 701–726. <https://doi.org/10.1080/10618600.2015.1060885>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Figure S1.** Log marginal likelihood values comparing the fit of 4 different mortality models to census data of a population of banded Mongoose (*Mungos mungo*).

**Figure S2.** Log marginal likelihood values comparing the fit of 4 different mortality models to capture-mark-recapture data of a wild population of European badger (*Meles meles*).

**Table S1.** DIC and WAIC scores comparing the fit of 4 different mortality models to census data of a population of banded Mongoose (*Mungos mungo*).

**Table S2.** DIC and WAIC comparing the fit of 32 different variations of the Siler model allowing every combination of sex-specific variation across parameters fit to data from a population of banded Mongoose (*Mungos mungo*).

**Table S3.** DIC and WAIC scores comparing the fit of 4 different mortality models to capture-mark-recapture data of a wild population of European badger (*Meles meles*).

**Table S4.** DIC and WAIC comparing the fit of 8 different variations of the Gompertz–Makeham model allowing every combination of sex-specific variation across parameters fit to data from a wild population of European badger (*Meles meles*).

**How to cite this article:** Hudson, D. W., Hodgson, D. J., Cant, M. A., Thompson, F. J., Delahay, R., McDonald, R. A., & McKinley, T. J. (2023). Importance sampling and Bayesian model comparison in ecology and evolution. *Methods in Ecology and Evolution*, 00, 1–13. <https://doi.org/10.1111/2041-210X.14237>