A flexible Bayesian method for estimating stratigraphic intervals and their co-occurrence in time

Gustavo A. Ballen[1,2,3,4,*]

[1] Instituto de Biociências de Botucatu, Universidade Estadual Paulista "Júlio de Mesquita Filho", Botucatu, SP, Brazil
[2] School of Biological and Behavioural Sciences, Queen Mary University of London, London, United Kingdom
[3] Center for Tropical Paleoecology and Archaeology, Smithsonian Tropical Research Institute, Ancón, Balboa, Panamá
[4] Museu de Zoologia da Universidade de São Paulo, Universidade de São Paulo, São Paulo, SP, Brazil
∗ Correspondence: gustavo.a.ballen@gmail.com, gustavo.ballen@unesp.br
Orcid https://orcid.org/0000-0001-5424-8608

1 Abstract

1. The fossil record is both incomplete and full of biases, which hinders making inferences about the past of life on earth. Stratigraphic intervals help us estimate two important quantities from the fossil record of a given lineage, its origination and extinction times.

2. Many models are available in the literature. This toolset is complemented here by proposing a flexible method which allows to estimate origination and extinction times, as well as model preservation potential in flexible ways.

3. A method is presented for representing the stratigraphic interval as a continuous distribution by integrating over parameter uncertainty throughthe use of the posterior predictive distribution. It is possible then to combine these distributions using conflation, to build a time probabilistic model for the co-occurrence of these intervals.

4. Empirical examples using Palynomorphs and the origination time of the marine Barracudas are used for illustrating the potential of the method. The methods are implemented in the `StratIntervals.jl` Julia package, which allows a big amount of possible prior distributions, and MCMC samplers.

2 Introduction

There is a powerful yet blurry window into the past of life on earth, which is the fossil record. Many insights into how the life came to be in its present form, and caveats on the possible directions that we can go regarding climate change and the biodiversity crisis are but few of the many applications of the knowledge we can gather from studying fossils (Smith et al., 2025). It has long been recognised that the fossil record is very incomplete, and multiple sources of bias hinder our ability to use it for inferences about the environments in deep time (Signor and Lipps, 1982). However, statistical tools have proved to be effective in helping us deal with the imperfections of this record, and gather insights from it through quantitative palaeobiology and its integration with evolutionary biology (Holland et al., 2024). One of the most difficult questions to ask about any extinct lineage are when they appeared, and then when they disappeared, that is, to estimate origination and extinction times (Strauss and Sadler, 1989).

There are several methods available for inference on stratigraphic intervals, with particular focus on confidence intervals on endopoint parameters. These methods vary in assumptions (e.g., whether preservation potential is uniform), statistical paradigm (e.g., whether using frequentist or Bayesian approaches), motivation for development (e.g., applied to palaeobiology or conservation biology). Comprehensive reviews are available in the literature (Marshall, 1990; Solow, 2005; Rivadeneira et al., 2009; Marshall, 2010; Boakes et al., 2015; McCrea et al., 2024), and recommended for details. Rivadeneira et al. (2009) arranged the methods know at that time into three classes: Class 1 methods, which assume constant preservation potential (e.g., Strauss and Sadler, 1989; Solow, 1993a; Ferraz et al., 2003; McInerny et al., 2006), Class 2 methods, which relax that assumption (e.g., Solow, 1993b; Marshall, 1997; McCarthy, 1998; Ferraz et al., 2003), and Class 3 methods, which are distribution-free (e.g., Marshall, 1994; Solow and Roberts, 2003). Later methods, especially that of Wang et al. (2016) and the one presented herein, fall in Class 2 because preservation is not assumed to be constant, but still rely on distributions.

Bayesian approaches were proposed since the first attempts at estimation on stratigraphic intervals. Strauss and Sadler (1989) already recognised the advantages of Bayesian formulations of their method, and provided a specific one using a uniform prior and a likelihood function constrained to the interval $[0, 1]$. Weiss and Marshall (1999) assumes a constant preservation potential and was the first to consider discrete sampling and therefore a non-continuous likelihood function. Later on, Weiss et al. (2004) extended this framework for dealing with non-constant preservation by using abundance data. Both methods are different from any other Bayesian one in that the likelihood function is not continuous. Ferraz et al. (2003) provided two other Bayesian formulations, one assuming linearly-decreasing preservation potential, and one uniform, in which suddenly extinction happens. They favoured the former model claiming that it is more realistic. Wang

et al. (2016) proposed a flexible method where different types of preservation potentials may fit, although they specified specific priors for the free endpoint and the preservation potential parameter. Then they used the analytical formulae for calculating exactly the posterior distributions of the model parameters. The method however conditions on one of the endpoints being zero. Alroy and Solow discussed in detail several aspects of Bayesian methods starting with a pair of papers by Alroy (2014, 2015) applied to the estimation of extinction, including the specification of priors and justification for likelihood functions (Alroy, 2016b,a; Solow, 2016b,a). Finally, Kodikara et al. (2020) proposed a method for estimating year of extinction using sighting data, and implemented the hierarchical Bayesian model in JAGS for sampling from the posterior parameter distributions. It is clear from the diversity of approaches using Bayesian techniques, that both flexible models and their implementation in computational tools are necessary to explore and extend these Bayesian methods for stratigraphic intervals. The goal of the present paper is to provide both.

3 Materials and methods

3.1 Implementation

The Bayesian implementation of the three-parameter Beta distribution as a model for stratigraphic intervals has been implemented in the `StratIntervals.jl` package, written in pure `Julia` (Bezanson et al., 2017). The package uses several dependencies (`DataFrames`, `Distributions`, `Interpolations`, `KernelDensity`, `QuadGK`, `Random`, `SpecialFunctions`, and `StatsPlots`; Bouchet-Valat and Kamiński, 2023; Besançon et al., 2021; Johnson, 2013) which provide data structures, plotting engines, or implement specific mathematical and statistical procedures not available in the `Base Julia` library. In particular, `StratIntervals` uses `Turing` (Ge et al., 2018), a package providing a probabilistic modeling language which allows for flexible implementation of MCMC algorithms.

The stable version of `StratIntervals` can be found in the official Julia Registry and installed with:

```
# in the Julia command line
using Pkg
# install StratIntervals
Pkg.add("StratIntervals")
# then load the package
using StratIntervals
```

The development version can be found at `https://github.com/gaballench/StratIntervals.jl`.

3.2 Model evaluation

Three sets of simulations for verifying the properties of the method under increasing sample size and different prior specifications were carried out as follows. For all simulations, a true interval with $\theta_1 = 150.0$ Ma and $\theta_2 = 100.0$ Ma was chosen, and a true preservation parameter $\lambda = 0.0$. All simulations were implemented in pure Julia code, using GNU parallel (Tange, 2011, 2024) when approrpiate, and run in the server `gymnotus` at the Instituto de Biociências, Universidade Estadual Paulista, Botucatu. All the simulation code is available at `https://github.com/gaballench/bayesian_collection_stratintervals` as well as on zenodo (DOI: 10.5281/zenodo.14868381).

3.2.1 Maximum likelihood estimation

Increasing datasets of 10, 50, 100, 1.000, and 10.000 occurrences were simulated to plot the joint likelihood surface for combination of (θ_1, λ), (θ_2, λ), and (θ_1, θ_2), while keeping the remaining parameter fixed to the true value. The true value of each parameter was then compared to the likelihood surface to see if parameter estimation is consistent.

3.2.2 Bayesian inference estimation

Increasing sample size of 10, 50, 100, 200, and 500 occurrences were simulated under the true stratigraphic interval. Then, MCMC sampling was used to calculate the posterior distributions of all three parameters through 10000 iterations and using the no-U-turn sampler (NUTS) (Hoffman and Gelman, 2014).

We are here mostly interested in the estimation of endpoint parameters θ_1, θ_2, and following the results suggesting an appropriate behaviour of λ estimation during simulations and maximum likehood estimation (see below), it was fixed to its true value instead and no prior specified for it.

Normal priors $N(\mu, \sigma)$ were set in four different combinations: Correct–incorrect, and informative–uninformative. A correct prior is defined as one centered at the true parameter value, whereas an incorrect prior is centered elsewhere and outside the stratigraphic interval (e.g. 20 units away from the true value). An informative prior (i.e., one with low variance) was set to $\sigma = 1.0$, whereas an uninformative one to $\sigma = 20.0$. Thus, for θ_1 the four combinations of priors are $N(150.0, 1.0)$, $N(150.0, 20.0)$, $N(150.0 + 20.0, 1.0)$, and $N(150.0 + 20.0, 20.0)$, whereas for θ_2 these are $N(100.0, 1.0)$, $N(100.0, 20.0)$, $N(100.0 - 20.0, 1.0)$, and $N(100.0 - 20.0, 20.0)$.

The effect of co-estimating θ_1 and θ_2 was assessed by running the simulations with priors set as described above, whereas the effect of estimating just one at a time was measured by fixing the other parameter to its true value and sampling just one at a time.

A total of 1000 simulations were run for each of these eight combinations: prior correctness, informativeness, and co-estimation.

A separate set of simulations was carried out for assessing the impact of the value of λ on the estimation of endpoint θ parameters. Here, a dataset of 75 occurrences was simulated from the Three-Parameter Beta distribution with the true value of λ varying (-1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5), whereas the true value of θ_1 and θ_2 were set to 550.0 and 500.0 Ma respectively, for a total of 7 simulated datasets. Priors on model parameters were set to $\theta_1 \sim N(\mu = 550.0, \sigma = 10.0)$, $\theta_2 \sim N(\mu = 500.0, \sigma = 10.0)$ and $\lambda \sim N(0.0, 2.0)$. The posterior distributions of model parameters were sampled using 100.000 iterations and the no-U-turn sampler.

3.3 Empirical examples

3.3.1 Empirical example 1: The palynomorphs of the Cerrejón formation

The Cerrejón formation records the sedimentary dynamics during the Palaeocene in northern South America. The nearly 1000 m stratigraphic section is rich in plant macro and microfossils, preserved in a coastal forest environment. Also, it is very important in economic terms, becausef it preserves multiple coal seams of very high quality, which has been extensively exploited in the Cerrejón open-pit coal mine. Jaramillo et al. (2007) studied the palynomorphs of a long section from the Cerrejón coal mine spanning the Manantial, Cerrejón, Tabaco, and Palmito formations from base to top. About 195 samples are available in the combined 4752 and 4774 sections, which were identified, counted, and subject to palaeoecological analyses. The data on the 216 different palynomorphs found were herein analysed to illustrate the application of stratigraphic interval estimation, as well as the use conflation for estimating the distribution for a collection of stratigraphic intervals.

Stratigraphic intervals of assorted palynomorphs. Three palynomorphs were chosen for estimating the stratigraphic interval parameters using depth occurrences. Most of the 216 taxa were discarded because they (1) spanned the whole sequence, or (2) had very few occurrences. From the selected palynomorphs, three were chosen which range from 26 to 46 occurrences, and varying in the presence of occurrences close to the top, thus used for fixing the parameter θ_2 instead of co-estimating both endpoint parameters.

The priors on $\theta_1 \sim N(\mu = \max(\text{occurrences}), \sigma = 140.0)$, $\theta_2 \sim Exp(\theta = \min(\text{occurrences})/2)$, and $\lambda \sim N(\mu = 0.0, \sigma = 2.0)$ were used for each stratigraphic interval. These priors are rather uninformative, following the results herein on the possible sensitivity to the prior in the presence of small sample sizes. The parameter θ_2 was fixed to 0.0 in *Ischyosporites problematicus* because the latest occurrences were very close to the top of the section, thus estimating only $\theta 1$ and λ.

The posterior predictive distribution was not calculated in this setting, only the posterior parameter distribution. MCMC sampling was carried out with the no-U-turn sampler, with 10.000 iterations. All parameter traces attained effective sample sizes above 600 (Table 1).

The lost sample of Coal Seam 115. While processing samples collected at the Cerrejón coal mine in northern Colombia, a doctoral student, found out that one important sample was labeled without information for depth and stratigraphic section, the only information available was "Coal Seam 1-something". We do not have first-hand stratigraphic information for the sample, but still want to figure out where in the generalised column the sample came from, as important fossil leaves were also found just at the top of that stratigraphic level and therefore a finer provenance information is key. What is the best estimate of stratigraphic position we can have?

The palynomorphs which were present at the stratigraphic position corresponding to Coal Seam 115 (285.99 m) were filtered, and found six of them: *Psilatricolporites* sp. ($n = 70$), *Retitricolporites* sp. ($n = 61$), *Psilamonocolpites grandis* ($n = 44$), *Echinatisporis minutus* ($n = 41$), *Retitricolpites* sp. ($n = 23$), and *Matonisporites* sp ($n = 23$). All the occurrences for these palynomorphs were collected, the occurrence at level 258.88 was removed in order to mimic the absence of information for the position of that level, which reduced in 1 the number of occurrences.

The priors were set as in the previous section. The parameter θ_2 was again fixed to 0.0 in *Psilatricolporites* sp., *Retitricolporites* sp., and *Matonisporites* sp., for the reasons discussed above.

The posterior predictive distribution was calculated here in order to conflate such distribution for each stratigraphic interval into a combined distribution for Coal Seam 115. MCMC sampling was carried out again with the no-U-turn sampler, but this time using 100.000 iterations because 10.000 did show issues of convergence for three out of the six species. Afterwards, parameter traces attained effective sample sizes above 600 again. Quantiles were calculated with the area under the curve of the conflated PDF and then compared with the true position for Coal Seam 115.

3.3.2 Empirical example 2: The time of origin of Barracudas

The origination time of the Barracudas (family Sphyraenidae) will be estimated using a curated dataset of fossil occurrences of the family (Ballen, 2020). Because the group is still living, the extinction time θ_2 will be fixed to zero during Bayesian posterior parameter estimation of θ_1 and λ. The effect of using different priors was assessed by setting $\theta_1 \sim N(\mu = 100.0, \sigma = 25.0)$, $\text{RefOffExp}(\theta = 50.0, o = 50.0, \rho = 1)$, and $U(50.0, 300.0)$. The prior $\lambda \sim N(\mu = 0.0, \sigma = 2.0)$ was used in all cases. Also, comparisons between removing and leaving duplicate time occurrences were carried out. It is important to note that these duplicates are occurrences from different localities, not redundancies form the same locality, and therefore have the same

age just because of uncertainty in dating, and thus are not true duplicates. The effect of this remotion is to reduce sample size.

The posterior predictive distribution was not calculated in this setting, only the posterior parameter distribution. MCMC sampling was carried out with the no-U-turn sampler, with 10.000 iterations. All parameter traces attained effective sample sizes above 600 (Table 3). Simulations were carried out in parallel for efficiency using GNU parallel (Tange, 2011, 2024).

4 Results

Each stratigraphic interval represents our known occurrences of a given lineage in geologic time. These lineages have unobserved origination and extinction times $\theta_{\cdot}$ The occurrences of the lineage are preserved in the fossil record in a way which has been described by different models in the literature, from uniform to distribution-free ways. It is reasonable to consider that preservation is most likely non-uniform across geologic time for a given lineage. This can be accomplished by using a generalised beta distribution which has a very flexible set of shapes governed by the parameters α and β. Together these can describe a non-uniform preservation pattern through geologic time. A reparametrisation in order to describe an equally rich set of preservation possibilities is applied here in order to lower the free parameters in the model while reianing flexibility.

The data are the known sets of fossil occurrence for a given fossil assemblage, that is, a fossil community composed of different species which coexist in space and time. Each of these species is expected to be also known from other points in time (and maybe space) and thus each of the species can be termed a stratigraphic series τ composed of all the points in time at which we know the lineage to have existed (e.g., blue squares in a given lineage of Figure 1).

Besides providing a flexibe Bayesian method for stratigraphic intervals, the goal of the present method is to construct a distribution for the time of co-occurrence of stratigraphic intervals. We can think of these co-occurrences as assembalges of species as being sampled at some point in time co-occurring in time and space. We want to use their time model, that is, the stratigraphic interval for each component of the assemblage, to construct the co-occurrence time.

4.1 The model: Likelihood function of a stratigraphic interval

Let τ be a random variable representing an occurrence event in time. A stratigraphic interval (Strauss and Sadler 1989) is a segment of time bound by an origination time θ_1 and an extinction time θ_2, where a lineage is existing on earth. Althoug in principle such interval in continuous, we can only access or sample it through

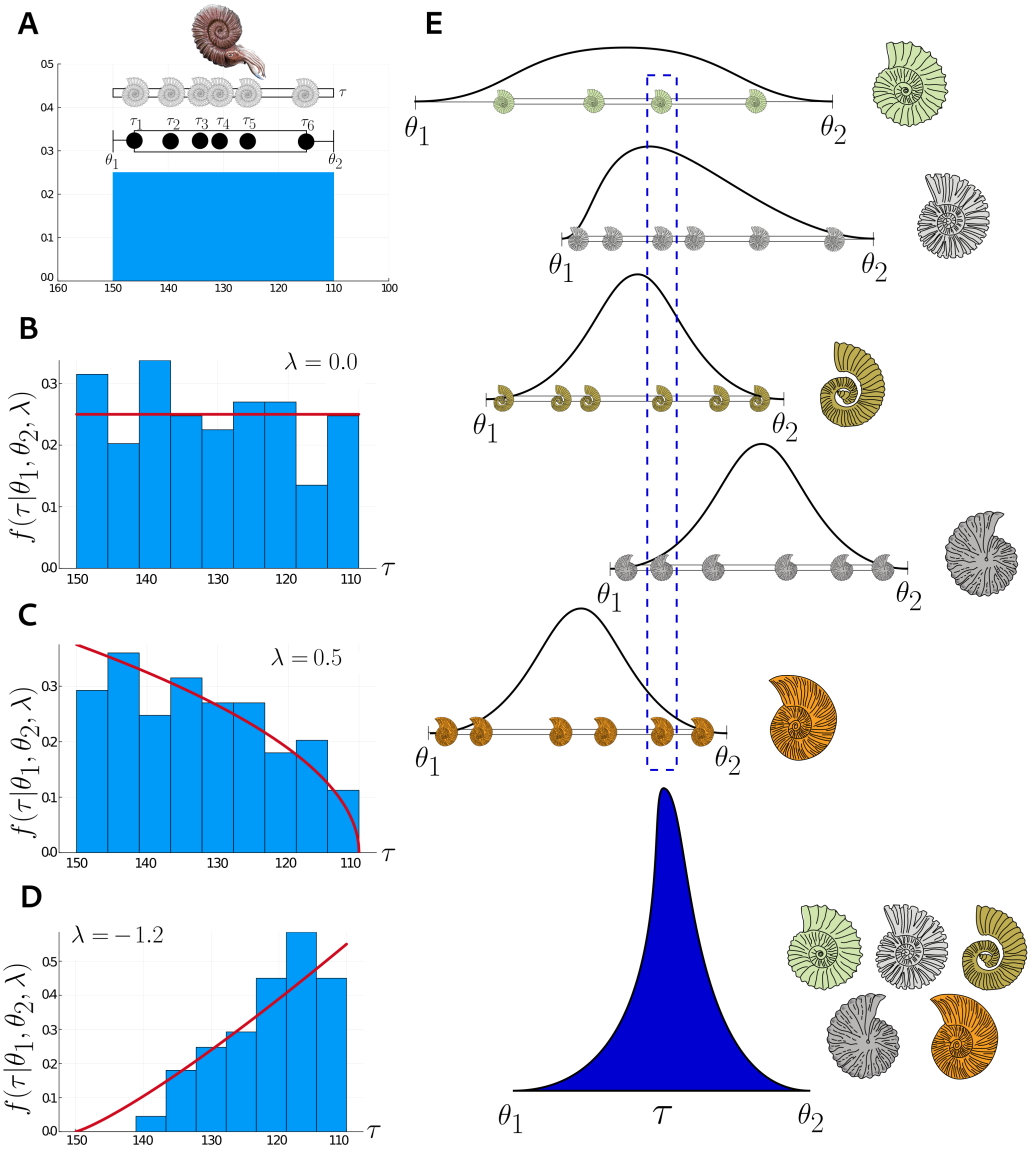

Figure 1: The model of a stratigraphic interval of time occurrences given the parameters. A) The fossil lineage (exemplified by the Ammonite) spans the time interval $[\theta_1, \theta_2]$, and then fossilised occurences are preserved and discovered at time points shown by the fossil figures. These define the stratigraphic interval with time occurrences $[\tau_1, ..., \tau_6]$ which are realisations of the underlying distribution $f(\tau|\theta_1, \theta_2, \lambda)$ (uniform in this example), whose probability is defined by the recovery potential governed by λ. That last parameter describe the relative frequency of time occurrences along the stratigraphic interval. B) When $\lambda = 0.0$, the distribution of occurrences is uniform through time. C) When it is negative the trend is decreasing in time, and D) the opposite when it is positive. The independent variable is time, depicted by τ in all figures, whereas the dependent variable is the probability density function $f(\tau|\theta_1, \theta_2, \lambda)$. E) Assemblage generating process. Each lineage has an origination and extinction time designed by $\theta_.$. Coloured ammonites represent occurrences in time, and we assume that all except those inside the blue dashed region have their age given, so that we are interested in estimating the age of thise in that region, which represent the coexistence time of all five lineages. Such time is given by the conflation of all the porbability density functions representing the stratigraphic interval of each species. Ammonite reconstructions available under Creative Commons license (in A) thanks to Kouhei Futaka and Izolende, and thanks to Jorge W. Moreno-Bernal (in E).

point time occurrences inside the bound interval, whose distribution is governed by a parameter or set of parameters describing the potential of preservation. We construct a function that describes the likelihood of observing a given time occurrence with the four-parameter Beta distribution (e.g., Wang, 2005) defined over $x \in [a, c]$ as

$$f(\tau|\theta_1, \theta_2, \alpha, \beta) = \frac{(\tau - \theta_1)^{\alpha-1}(\theta_2 - \tau)^{\beta-1}}{(\theta_2 - \theta_1)^{\alpha+\beta-1}B(\alpha, \beta)} \tag{1}$$

where $B(\alpha, \beta)$ is the Beta function. We have renamed here the variables $y = \tau$, $a = \theta_1$, and $c = \theta_2$ in order to describe the process above.

Now let us reparametrise the function so that it depends only on a parameter λ instead of α and β in a similar way as (Wang et al., 2016). This has the advantage of requiring less data while still giving the likelihood function enough flexibility to describe different preservation scenarios. We shall call this function the three-parameter Beta distribution. Note however that here we do not fix any of the interval parameters, which gives the likelihood function:

$$f(x|\theta_., \lambda) \sim Beta(\theta_., 1, 1 - \lambda) \text{ if } \lambda \leq 0 \tag{2}$$

and

$$f(x|\theta_., \lambda) \sim Beta(\theta_., 1 + \lambda, 1) \text{ if } \lambda > 0 \tag{3}$$

which gives

$$\begin{aligned} f(\tau|\theta_1, \theta_2, \lambda) &= \frac{(\theta_2 - \tau)^{-\lambda}}{(\theta_2 - \theta_1)^{1-\lambda}B(1, 1 - \lambda)} \text{ if } \lambda \leq 0 \\ \text{where } B(1, 1 - \lambda) &= \int_0^1 t^{1+\lambda-1}(1 - t)^{1-1}\mathrm{d}t = \frac{1}{1 - \lambda} \end{aligned} \tag{4}$$

and

$$\begin{aligned} f(\tau|\theta_1, \theta_2, \lambda) &= \frac{(\tau - \theta_1)^{\lambda}}{(\theta_2 - \theta_1)^{1+\lambda}B(1 + \lambda, 1)} \text{ if } \lambda > 0 \\ \text{where } B(1 + \lambda, 1) &= \int_0^1 t^{1-1}(1 - t)^{1-\lambda-1}\mathrm{d}t = \frac{1}{1 + \lambda} \end{aligned} \tag{5}$$

Assuming that time occurrences are independent and identically distributed, we can also calculate the likelihood of observing a vector representing the collection of N time occurrences given a stratigraphic interval by the product of likelihoods of individual time occurrences:

$$f(\boldsymbol{\tau}|\theta_1, \theta_2, \lambda) = \prod_{i=1}^{N} f(\tau_i|\theta_1, \theta_2, \lambda) \tag{6}$$

4.2 The posterior distribution of parameters and posterior predictive of τ

We now apply the Bayes' theorem in order to construct the posterior densities for these parameters given the observed vector of occurrences in time:

$$f(\theta_1, \theta_2, \lambda|\boldsymbol{\tau}) = \frac{f(\theta_1)f(\theta_2)f(\lambda)\prod_{i=1}^{N} f(\tau_i|\theta_1, \theta_2, \lambda)}{\int f(\theta_1)f(\theta_2)f(\lambda)\prod_{i=1}^{N} f(\tau_i|\theta_1, \theta_2, \lambda)\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}\lambda} \tag{7}$$

Depending on the specific choice of priors for θ_1, θ_2, and λ, this expression may have a closed form (e.g., using conjugate priors). However, for several prior functions this expression may not be closed and therefore we can resort to MCMC sampling in order approximate the integral in the denominator.

Now that we have a posterior of parameters, we can sample from this conditional distribution in order to build a posterior predictive distribution of τ by integrating over the parameter uncertainty. This posterior predictive distribution represents the probabilistic model for the occurrences through time while accounting for uncertainty in endpoint and preservation parameter estimation.

$$p(\tilde{\tau}|\boldsymbol{\tau}) = \int p(\tilde{\tau}|\theta_1, \theta_2, \lambda)p(\theta_1, \theta_2, \lambda|\boldsymbol{\tau})\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}\lambda \tag{8}$$

4.3 The distribution of the co-occurrence of multiple stratigraphic intervals

Combining different distributions is not straitghforward (Genest and Zidek, 1986). However, the conflation of probability density functions is a useful procedure which combines them provided that each of them is independent (Hill, 2011; Hill and Miller, 2011). Assuming that each distribution describing the posterior predictive distribution of each interval is independent, and simplifying notation so that $\tilde{\tau} = \tilde{\tau}|\boldsymbol{\tau}$, we can define the composite distribution of $\tilde{\tau}$ for M intervals as the conflation of individual posterior predictive distributions:

$$Q(\tilde{\tau}) = \frac{\prod_{i=1}^{M} p_i(\tilde{\tau})}{\int_{-\infty}^{\infty}\prod_{i=1}^{M} p_i(x)\mathrm{d}x} \tag{9}$$

Such distribution is useful e.g. when we want to build credible intervals for the time τ of co-occurrence of stratigraphic intervals. The conceptual interpretation as follows: M different lineages as represented by their stratigraphic intervals should coexist at most for some time interval when they all were alive. As the conflation of densities is a density itself, it can be used for asking questions on the probability of co-existence

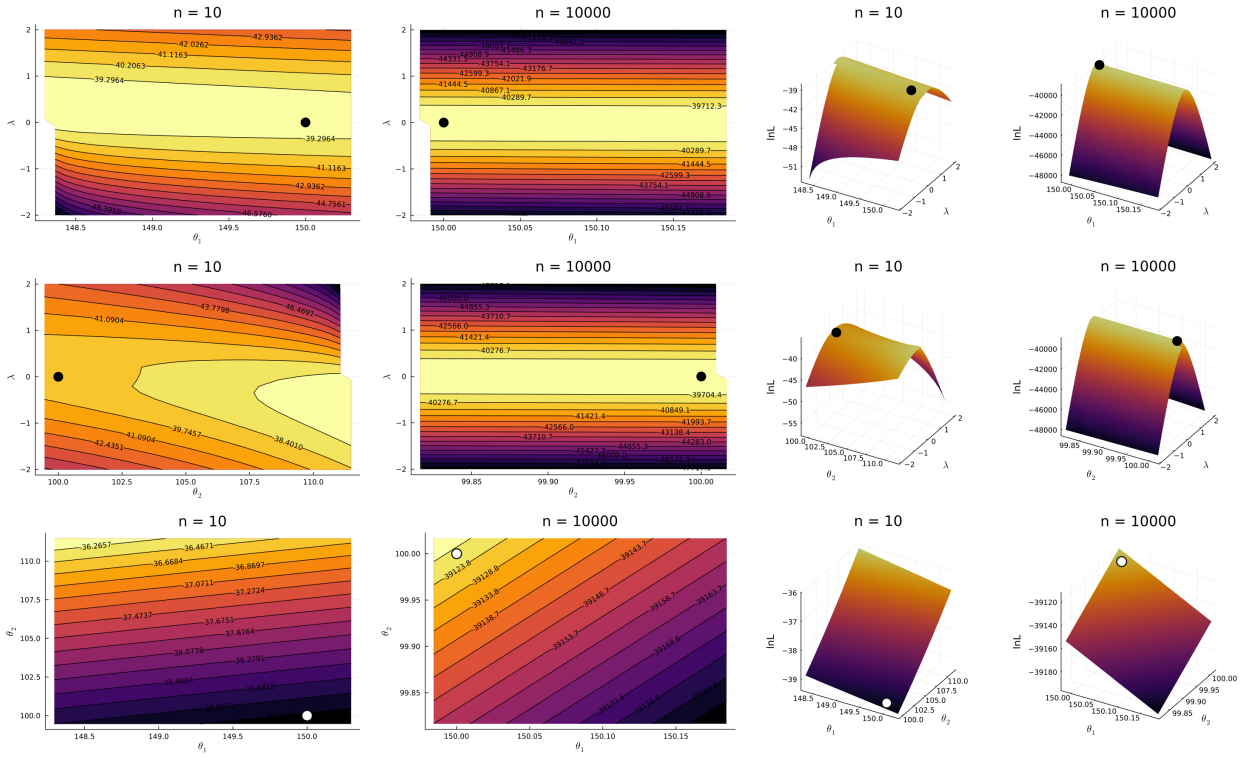

Figure 2: Log-likelihood surface of different pairs of parameters from the ThreeParBeta$(\tau; \theta_1, \theta_2, \lambda)$ distribution as a function of a fixed dataset of varying size ($n = 10$ or $n = 10.000$). Left-most columns are 2-dimensional representations of the 3-dimensional surfaces to the right. Isoclines represent log-likelihood, the brighter the higher. The x and y axes are the parameter pairs (θ_1, λ), (θ_2, λ), and (θ_1, θ_2) from top to bottom. In all cases, the remaining parameter was fixed to its true value. The pair of true values is plotted as a dot, which changes in colour just to aid in visualisation. The estimation of λ is unbiased, whereas the estimation of both θ_1 and θ_2 is biased and only converges to the true value asymptotically.

of lineages during some arbitrary time interval given the distribution.

5 Testing the model

5.1 Asymptotic properties and parameter estimation under maximum likelihood and Bayesian inference

The likelihood surface was found to be flat with respect to the endpoint parameters, whereas it it well-behaved for the preservation parameter. The endpoint parameter estimators are biased, which means that the first appearance datum (FAD) and last appearance datum (LAD), instead of the true value, always maximise the likelihood surface. For small sample size, they are also unidentifiable. However, as n rises and the LAD and FAD tend to approach the true endpoint values, and therefore the MLE converges to the true value (Figure 2).

The Bayesian version of the method however does not show bias in estimation and is better behaved in terms of estimation accuracy (Figure 3). The cost of co-estimating the both endpoint parameters in simulations as compared to fixing one of them to the true value is negligible (Supplementary Figure S1) and therefore we will discuss here only simulations estimating one endpoint parameter while fixing the other to the true value.

When simulating multiple datasets (1000) of varying sample size, we find that the distribution of posterior medians concentrate around the true value (Figure 4), and the variance of this sample of medians decreases as we increase in sample size. This measures the precision of estimation of the model in general, over multiple datasets drawn from the same distribution.

The inter-quartile range (IQR) of the posterior distribution, which measures better the spread in asymmetrical ones, also reduces as we grow in sample size (Figure 5), therefore showing a gain in precision within each posterior estimation. This measures the precision of estimation conditional to a specific dataset. However, when a prior is incorrect and informative, the precision actually decreases with sample size. This is not the case for incorrect–uninformative, correct–informative, and correct-uninformative priors, all of which show an increase in precision with increas in sample size.

The mean squared error (MSE) lowers by raising sample size (Figure 6), which is consistent with the results of the IQR, and highlights both the increase in accuracy and precision during estimation across multiple simulated datasets from the same model. The decrease in MSE in consistent regardless of the presence of incorrect and/or uninvormative priors.

The value of λ and therefore the shape of the Three-Parameter Beta distribution have an influence on the posterior variance of the θ endpoint parameters (also see Supplementary Figure S2). When $\lambda = 0.0$, and therefore the model has a uniform shape, the posterior variance of both θ_1 and θ_2 is the same (Figure 7A). However, when λ is negative (e.g., = -1.5) there is a tail towards the past (Figure 7B), whereas if it is positive (e.g., 1.5), the tail goes to the present (Figure 7C).

5.2 The effect of priors

The bias and semi-identifiability of endpoint parameters in maximum likelihood estimation is not seen however when using Bayesian inference. By applying priors to the $\theta_{\cdot}$ parameters we end up with a posterior surface which allows proper statistical inference, that is, parameter identifiability (Figure 3).

Because of this important role of the priors, we explore now the behaviour of the Bayesian model under two situations: Priors for θ_1, θ_2 with relative low and high variance, and priors which are consistent or inconsistent with the true parameter value, that is, whether the true value is near the distribution mean, or

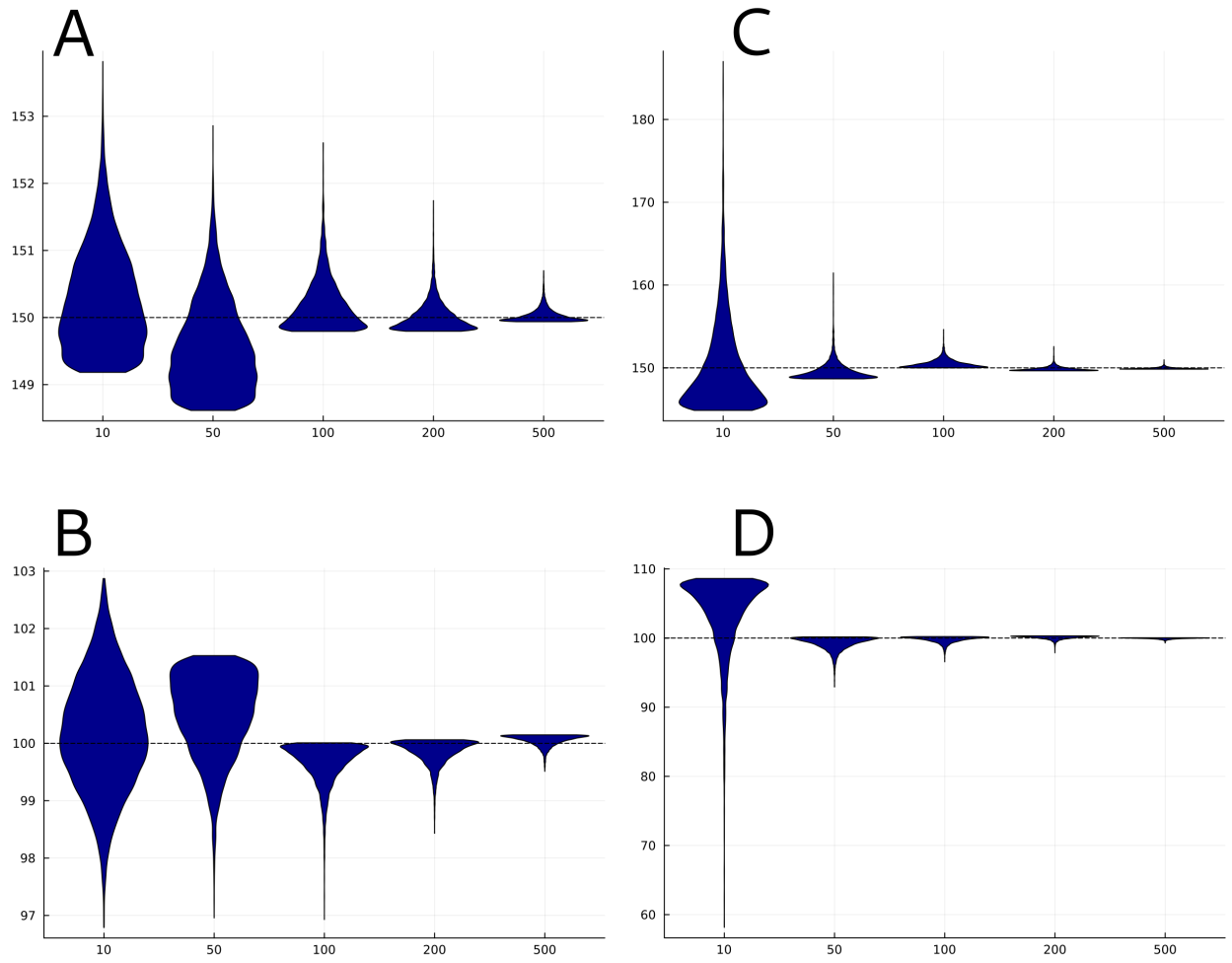

Figure 3: Posterior distribution of a single simulated dataset of varying sample size from the ThreeParBeta$(\tau; \theta_1, \theta_2, \lambda)$ distribution to illustrate the effect of sample size (x axis, in Ma) and prior variance on θ_1 and θ_2. A) Posterior distribution of θ_1 with prior $N(\mu = 150.0, \sigma = 1.0)$. B) Posterior distribution of θ_2 with prior $N(\mu = 100.0, \sigma = 1.0)$. C) Posterior distribution of θ_1 with prior $N(\mu = 150.0, \sigma = 20.0)$. D) Posterior distribution of θ_2 with prior $N(\mu = 100.0, \sigma = 20.0)$. The dashed line represents the true value of the parameter in all plots.

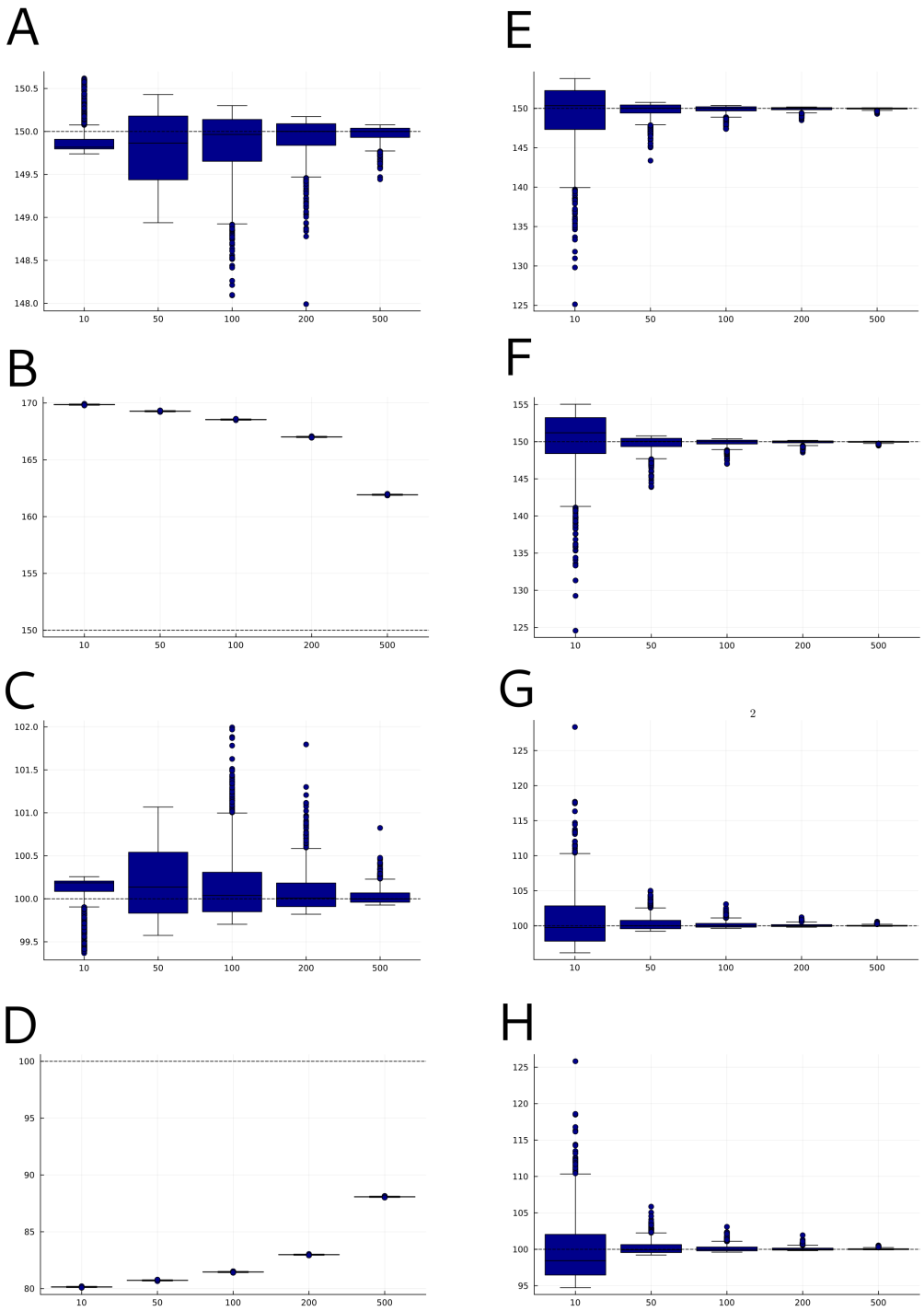

Figure 4: Distribution of posterior medians (y axis) across 1000 simulations for combinations of correct and incorrect prior, and informative and uninformative prior, as a function of sample size (x axis). A) Correct and informative prior ($\theta_1 \sim N(150.0, 1.0)$). B) Incorrect and informative prior ($\theta_1 \sim N(150.0 + 20.0, 1.0)$). C) Correct and informative prior ($\theta_2 \sim N(100.0, 1.0)$). D) Incorrect and informative prior ($\theta_2 \sim N(100.0 - 20.0, 1.0)$). E) Correct and uninformative prior ($\theta_1 \sim N(150.0, 20.0)$). F) Incorrect and uninformative prior ($\theta_1 \sim N(150.0 + 20.0, 20.0)$). G) Correct and uninformative prior ($\theta_2 \sim N(100.0, 20.0)$). H) Incorrect and uninformative prior ($\theta_2 \sim N(100.0 - 20.0, 20.0)$). The dashed line represents the true value in all cases.

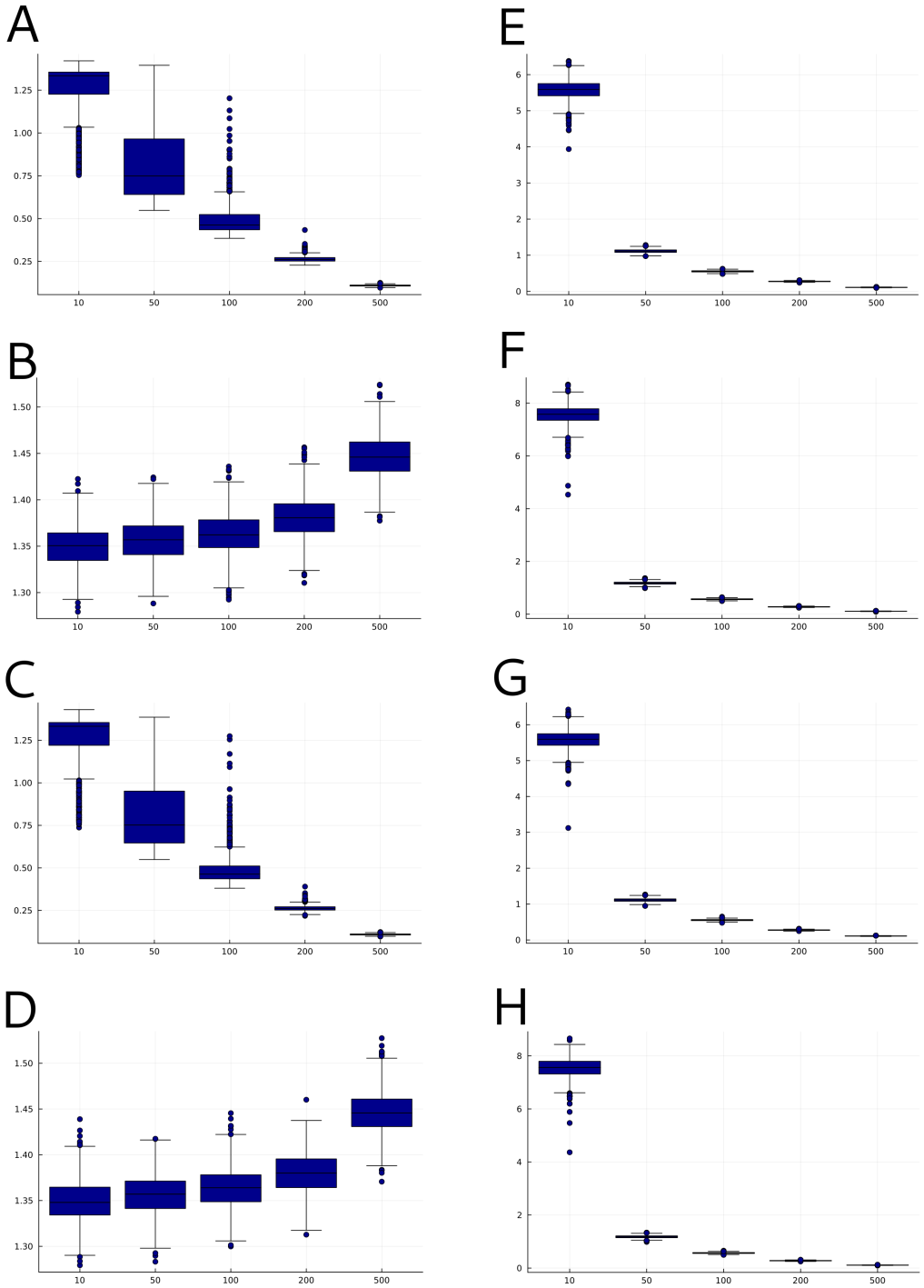

Figure 5: Distribution of posterior inter-quartile range (IQR, y axis) across 1000 simulations for combinations of correct and incorrect prior, and informative and uninformative prior, as a function of sample size (x axis). A) Correct and informative prior ($\theta_1 \sim N(150.0, 1.0)$). B) Incorrect and informative prior ($\theta_1 \sim N(150.0 + 20.0, 1.0)$). C) Correct and informative prior ($\theta_2 \sim N(100.0, 1.0)$). D) Incorrect and informative prior ($\theta_2 \sim N(100.0 - 20.0, 1.0)$). E) Correct and uninformative prior ($\theta_1 \sim N(150.0, 20.0)$). F) Incorrect and uninformative prior ($\theta_1 \sim N(150.0 + 20.0, 20.0)$). G) Correct and uninformative prior ($\theta_2 \sim N(100.0, 20.0)$). H) Incorrect and uninformative prior ($\theta_2 \sim N(100.0 - 20.0, 20.0)$).

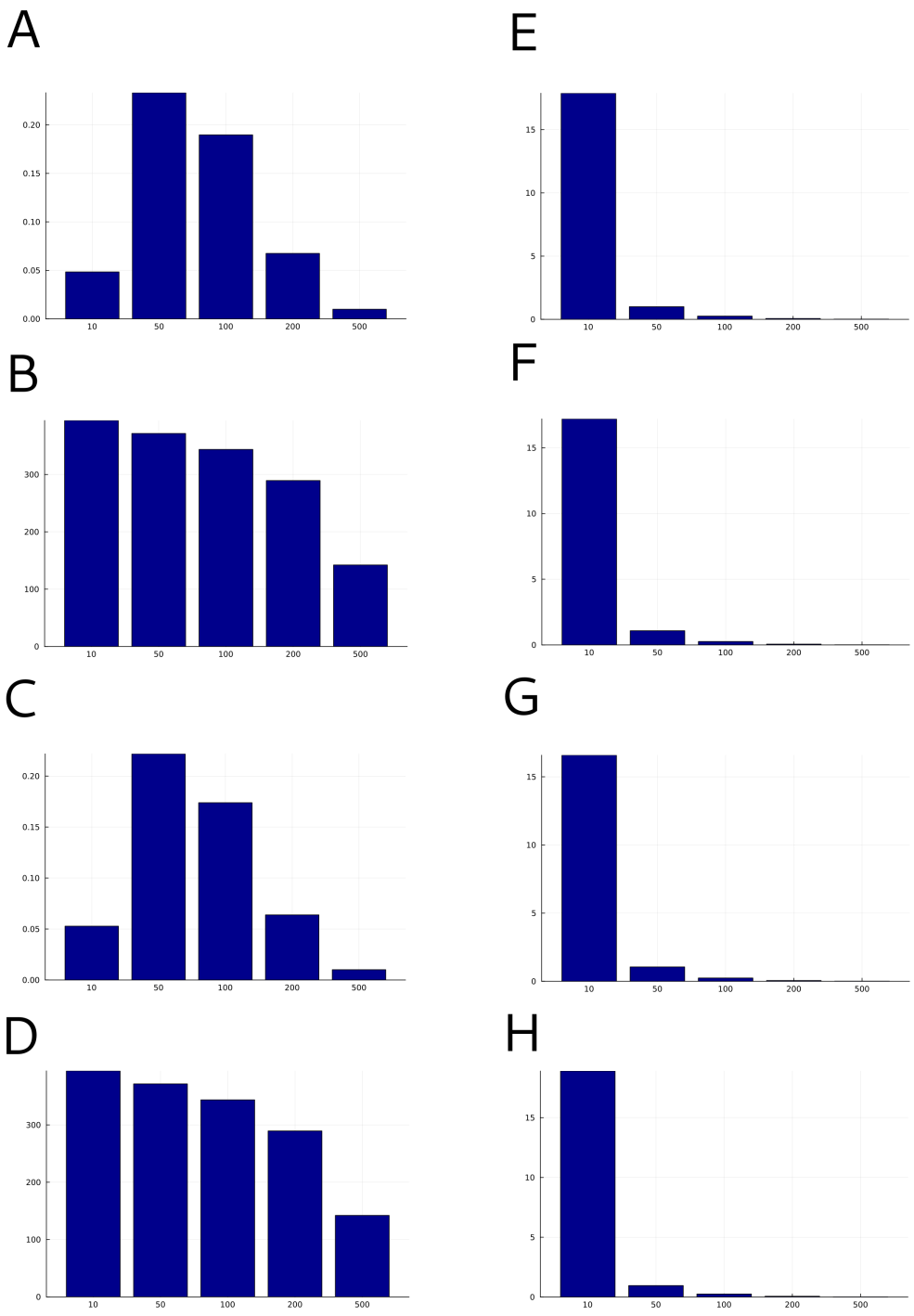

Figure 6: Mean squared error (MSE, y axis) across 1000 simulations for combinations of correct and incorrect prior, and informative and uninformative prior, as a function of sample size (x axis). A) Correct and informative prior ($\theta_1 \sim N(150.0, 1.0)$). B) Incorrect and informative prior ($\theta_1 \sim N(150.0 + 20.0, 1.0)$). C) Correct and informative prior ($\theta_2 \sim N(100.0, 1.0)$). D) Incorrect and informative prior ($\theta_2 \sim N(100.0 - 20.0, 1.0)$). E) Correct and uninformative prior ($\theta_1 \sim N(150.0, 20.0)$). F) Incorrect and uninformative prior ($\theta_1 \sim N(150.0 + 20.0, 20.0)$). G) Correct and uninformative prior ($\theta_2 \sim N(100.0, 20.0)$). H) Incorrect and uninformative prior ($\theta_2 \sim N(100.0 - 20.0, 20.0)$).

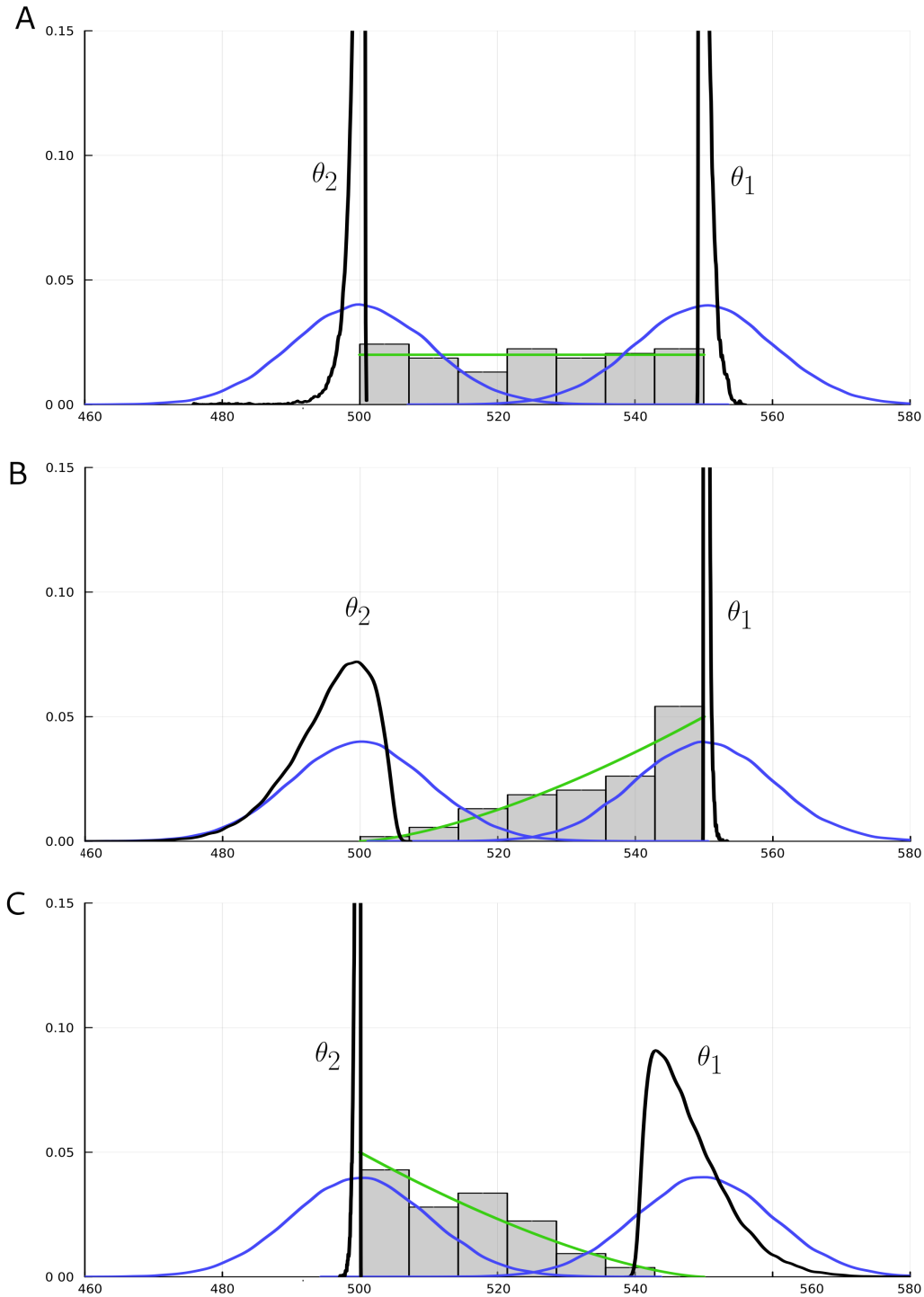

Figure 7: Effect of λ on θ_1 and θ_2 for a sample size of 75 occurrences sampled from different parameter values of λ. A) $\lambda = 0.0$. B) $\lambda = 1.5$. C) $\lambda = -1.5$. In all plots the green line represents the ThreeParBeta$(\tau; \theta_1 = 550.0, \theta_2 = 500.0, \lambda)$ from where the 75 occurrences were sampled. Blue lines represent the rather uninformative prior centered at the true value on each ($\theta_1 \sim N(\mu = 550.0, \sigma = 10.0)$ and $\theta_2 \sim N(\mu = 500.0, \sigma = 10.0)$). The black lines represent the posterior distribution. In all plots the x axis is time (Ma) whereas the y axis is density.

very off from it. This allows to verify how sensitive to the prior it is as a function of sample size. When applying priors that are centered on the true parameter values, both low- and high-variance priors are capable of estimating the posterior distributions that include the true value. The effect of the prior variance on the posterior variance however reduces as we grow in sample size (Figures 3,5).

If an incorrect prior is chosen, one could potentially run into issues where the posterior do not include the true value even for large sample sizes. By applying a prior which is incorrect in the sense that it is centered away from the true values, we can assess the effect of informative and uninformative priors on the posterior distriutions. We find that a low-variance prior which are centered away from the true values have difficulties in converging to the true values even for large sample sizes. This effect is however not seen in high-variance priors, which quickly are capable of converging to the region around the true value as sample size grows (Figures 4,5,6 for wrong prior). Overall the method is well behaved as regardless of whether priors are correct or incorrect, informative or uninformative, the mean squared error decreases with an increase in sample size (Figure 6). The interquartile ranges tend to increase with sample size only for priors which are incorrect and informative, whereas they always decrease for correct informative and uninformative priors, as well as for incorrect uninformative priors (Figure 5)

5.3 Empirical examples

5.3.1 Empirical example 1: The palynomorphs of the Cerrejón formation

Stratigraphic intervals of different palynomorphs. Posterior sampling showed good efficiency and convergence (Table 1), and recovered behaviours also seen in simulations. First, striatigraphic intervals where λ created a tail towards θ_1 also made the inference of this parameter to show higher variance than in cases where λ was closer to 0.0 (compare Figure 8A and 8B). All posterior distributions of θ_1 were mostly contained into the measured sections (Figure 8). Posterior distribution of θ_2 was bound between 0.0 and the latest occurrence, however, it did not tend to follow the shape of the prior distribution for *Arecipites regio*, thus suggesting a stronger influence of more data (71 occurrences, Table 2, Figure 8A). In contrast, the posterior was similar to the prior in *Scabratriporites triangularis*, presumably due to the smaller amount of data (26 occurrences, Table 2, Figure 8C). The posterior distribution of θ_2 was consistently different from the prior for all species, regardless of sample size. The posterior distribution of λ were always quite different from 0.0, which means that the stratigraphic interval model is different from uniform.

The lost sample from Coal Seam 115. The use of six palynomorphs for building a conflated distribution for their co-occurrence restricts the possible depth of the mystery sample to the 95% interval 341.8m–54.4m, with a width of 287.4 m. The median of the conflated distribution was 193.8. This interval

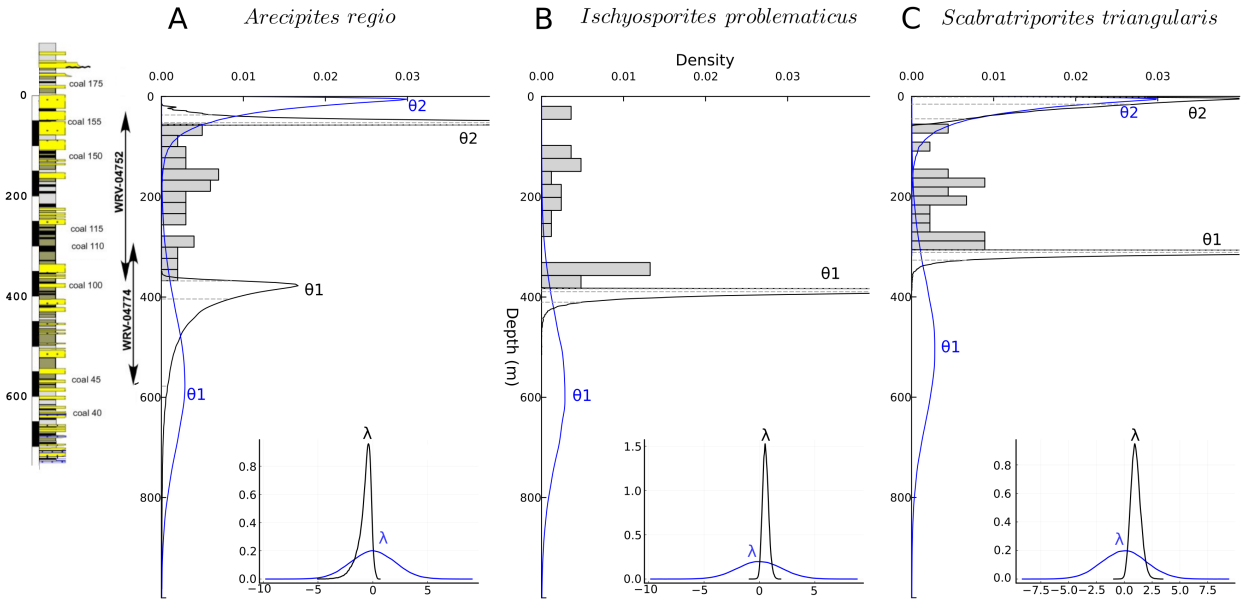

Figure 8: Stratigraphic intervals of A) *Arecipites regio* ($n = 46$), B) *Ischyosporites problematicus* ($n = 33$), and C) *Scabratriporites triangularis* ($n = 26$). The parameter θ_2 was fixed for *Ischyosporites problematicus* as its latest occurrences were very close to the uppermost samples in the stratigraphic column. Offset left is the generalised stratigraphic column of the Cerrejón formation starting from the top of core WVR-0475 and finishing below core WVR-4774. Subplots A–C are aligned with the generalised colum on the depth (m) axis. Lowermost densities correspond to θ_1, uppermost ones to θ_2, and inlay plots to λ. Prior distributions are blue, whereas posterior distributions are black. Data are the grey bars in the histogram. Dashed grey lines are the posterior quantiles 0.025, 0.5, and 0.975.

Table 1: MCMC efficiency and convergence measures for the analyses of the Cerrejón formation palynomorphs. Abbreviations are Monte-Carlo standard error (MCSE), effective sample size (ESS), effective sample size per second (ESS s-1).

Species	**Parameter**	**MCSE**	**ESS (bulk)**	**ESS (tail)**	**$\hat{R}$**	**ESS s-1**
Arecipites regio	θ_1	1.4890	2683.4	3350.8	1.0007	37.1
Arecipites regio	θ_2	0.1878	1871.2	1068.0	1.0008	25.9
Arecipites regio	λ	0.0133	2291.7	3232.8	1.0009	31.7
Ischyosporites problematicus	θ_1	0.0968	11013.8	13622.9	1.0001	566.1
Ischyosporites problematicus	λ	0.0021	18007.2	18124.6	1.0000	925.5
Scabratriporites triangularis	θ_1	0.0820	8153.0	10892.7	1.0001	226.2
Scabratriporites triangularis	θ_2	0.3586	1024.6	1070.8	1.0005	28.4
Scabratriporites triangularis	λ	0.0077	3344.0	5446.6	1.0000	92.8

Table 2: Summary statistics for the posterior sampling of three palynomorphs from the Cerrejón formation. Percentage columns are quantiles at that given probability value. Abbreviations are standard deviation (SD).

Species	**Parameter**	**2.5%**	**25.0%**	**50.0%**	**75.0%**	**97.5%**	**Mean**	**SD**
Arecipites regio	θ_1	368.8	381.2	403.5	452.3	625.8	429.6	70.1032
Arecipites regio	θ_2	31.0	46.9	51.2	53.7	55.2	49.2	6.3193
Arecipites regio	λ	-2.4	-1.0	-0.6	-0.3	0.0	-0.7	0.6192
Ischyosporites problematicus	θ_1	383.3	385.5	388.9	394.8	416.8	391.8	9.2678
Ischyosporites problematicus	λ	0.1	0.4	0.5	0.7	1.1	0.6	0.2697
Scabratriporites triangularis	θ_1	306.5	308.1	310.6	314.9	331.2	312.7	6.8307
Scabratriporites triangularis	θ_2	0.8	6.4	14.4	25.7	47.9	17.3	13.1489
Scabratriporites triangularis	λ	0.3	0.7	1.0	1.3	2.0	1.0	0.4396

includes the true depth of the Coal seam 115 which is 285.99 m (Figure 9). Four coal seams are found in this interval, 110, 115, 150, and 155. The 95% interval of the conflated distribution was in general restricted to the stratigraphic span of section WRV-4752.

Most of the palynomorph age models, represented by their posterior predictive distributions, agreed more or less on the interval which determined in great part the shape of the conflation. However, *Psilatricolporites* sp. showed a wider distribution, not contributing much to the final shape because of its greater variance when compared to the posterior predictive distributions of the remaining palynomorphs.

5.3.2 Empirical example 2: The time of origin of Barracudas

Leaving the original dataset with time duplicates (i.e., occurrences with the same time but from different localities, therefore not complete duplicates) results in 75 occurrences. In this scenario the shape of the posterior distribution is quite insensitive to the choice of the prior. Removing duplicates leaves 33 occurrences, for which the choice of the prior impacts somewhat the shape of the posterior distribution.

The most important factors influencing the sensitivity of the posterior seem to be both the prior mean and variance (Fig. 10). The Normal prior tends to concentrate more density around mean ($\sigma^2 = 625.0$), and therefore tends to push more the posterior towards its mean, in this case, into the past. In contrast, the priors RefOffExponential and Uniform, with more variance ($\sigma^2 = 2500.0$ and $\sigma^2 = 5208.33$ respectively), tend to have a slighter impact on the posterior distribution.

The posterior distribution of θ_1, the origination time, was estimated roughly between 54.3–53.1 Ma (mean age 54.2–54.4 Ma depending on the prior) when using the complete dataset, regardless of the prior used (Table 4). It grew to 113.8–53.8 Ma (narrowest, RefOffExponential prior) and even to 146.4–54.1 Ma (widest, Uniform prior) when using the reduced dataset, this time showing more influence of the prior. In this case, the posterior mean was 70.1 Ma (RefOffExponential prior), 79.1 Ma (Uniform prior), or 83.5 Ma (Normal prior).

The posterior distribution of λ included zero but showed prevalence of values on the negative side of the

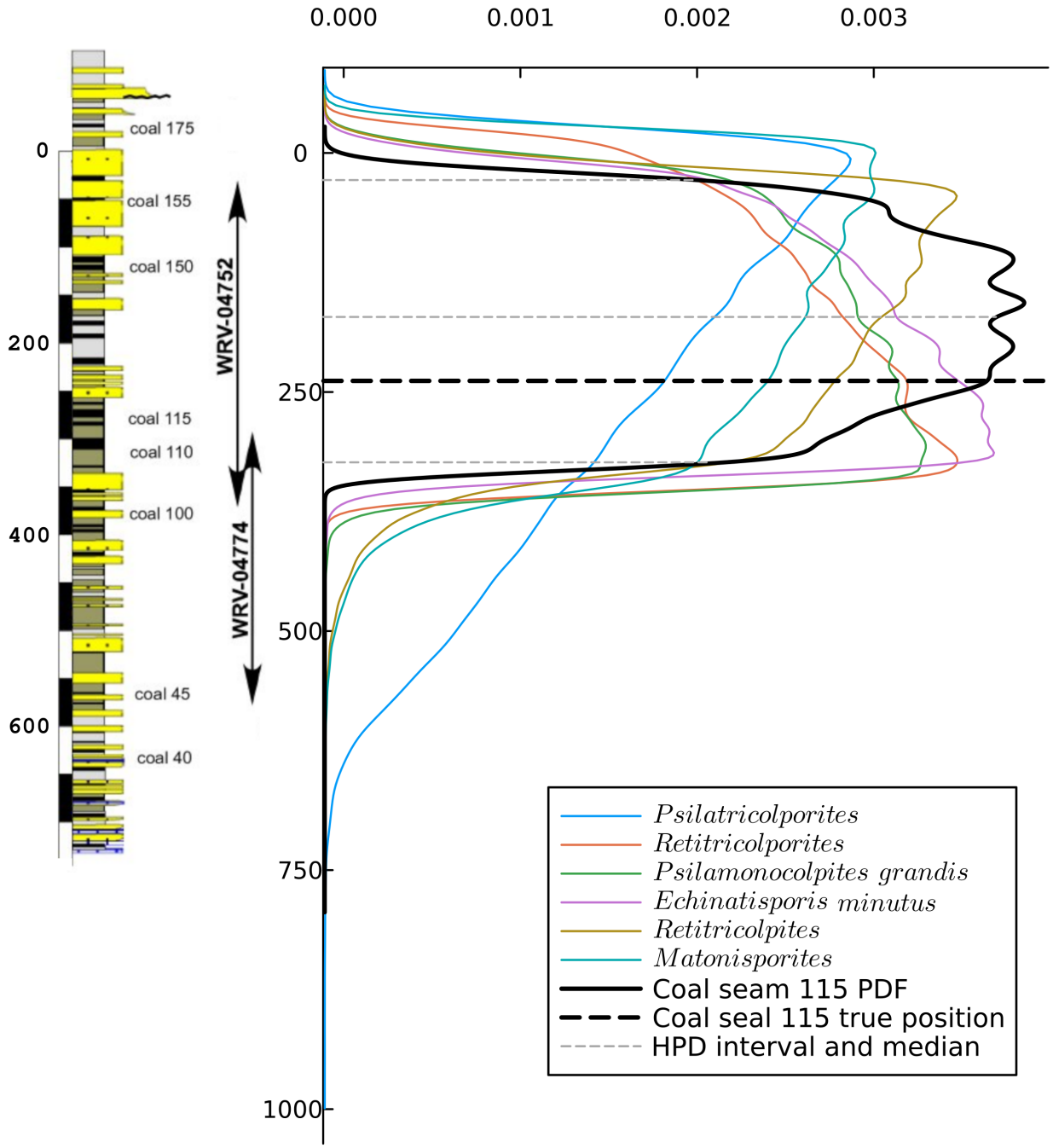

Figure 9: The most probable depth for a mysterious sample labeled Coal Seam 115 as inferred from the conflation of the posterior predictive distributions of the palynomorphs recovered from that sample. Offset left is the generalised stratigraphic column of the Cerrejón formation starting from the top of core WVR-0475 and finishing below core WVR-4774. The true depth falls within the highest density interval of the conflated distribution, although its median is not aligned to the true value.

Table 3: MCMC sampling efficiency measures for the analyses of the origination time of Barracudas. Abbreviations are Monte-Carlo standard error (MCSE), effective sample size (ESS), effective sample size per second (ESS s-1). Clean represents whether duplicates were removed or not.

Prior	**Clean**	**Parameter**	**MCSE**	**ESS (bulk)**	**ESS (tail)**	**$\hat{R}$**	**ESS s-1**
N(100.0, 25.0)	No	θ_1	0.0828	816.7	817.8	1.0000	80.1
N(100.0, 25.0)	No	λ	0.0047	1089.7	1078.0	1.0002	106.8
RefOffExp(50.0, 50.0, 1)	No	θ_1	0.0346	1250.1	2268.2	1.0000	61.5
RefOffExp(50.0, 50.0, 1)	No	λ	0.0028	1539.6	2141.4	1.0009	75.7
U(550.0, 300.0)	No	θ_1	0.0419	1095.2	1362.9	1.0000	53.9
U(550.0, 300.0)	No	λ	0.0032	1413.5	1910.8	1.0003	69.5
N(100.0, 25.0)	Yes	θ_1	0.4962	1287.4	889.1	1.0001	128.8
N(100.0, 25.0)	Yes	λ	0.0252	1246.6	785.7	1.0002	124.8
RefOffExp(50.0, 50.0, 1)	Yes	θ_1	0.4003	1702.3	1661.0	1.0026	95.1
RefOffExp(50.0, 50.0, 1)	Yes	λ	0.0212	1599.1	1768.0	1.0031	89.3
U(550.0, 300.0)	Yes	θ_1	0.7271	1297.7	1806.3	1.0003	73.8
U(550.0, 300.0)	Yes	λ	0.0322	1370.2	1548.8	1.0000	77.9

Table 4: Summary statistics for the posterior sampling of the origination time of Barracudas. Percentage columns are quantiles at that given probability value. Clean represents whether duplicates were removed or not. Abbreviations are standard deviation (SD).

Prior	**Clean**	**Parameter**	**2.5%**	**25%**	**50%**	**75%**	**97.5%**	**Mean**	**SD**
N(100.0, 25.0)	No	θ_1	53.1	53.4	53.8	54.8	58.8	54.4	2.0286
N(100.0, 25.0)	No	λ	-0.4	-0.1	-0.1	0.0	0.1	-0.1	0.1348
RefOffExp(50.0, 50.0, 1)	No	θ_1	53.1	53.4	53.8	54.5	57.5	54.2	1.2295
RefOffExp(50.0, 50.0, 1)	No	λ	-0.3	-0.1	0.0	0.0	0.1	-0.1	0.1090
U(550.0, 300.0)	No	θ_1	53.1	53.4	53.8	54.6	57.7	54.2	1.3448
U(550.0, 300.0)	No	λ	-0.3	-0.1	0.0	0.0	0.1	-0.1	0.1133
N(100.0, 25.0)	Yes	θ_1	55.3	68.3	81.1	95.8	126.4	83.5	19.2489
N(100.0, 25.0)	Yes	λ	-3.9	-2.5	-1.7	-1.1	-0.3	-1.8	0.9564
RefOffExp(50.0, 50.0, 1)	Yes	θ_1	53.8	58.5	64.9	76.3	113.8	70.1	16.3802
RefOffExp(50.0, 50.0, 1)	Yes	λ	-3.4	-1.6	-1.0	-0.6	-0.1	-1.2	0.8516
U(550.0, 300.0)	Yes	θ_1	54.1	60.9	71.3	89.3	146.4	79.1	25.0567
U(550.0, 300.0)	Yes	λ	-4.4	-2.2	-1.3	-0.8	-0.2	-1.6	1.1182

axis (e.g., -0.4–0.1) for the complete dataset, and was consistently restricted to the negative side of the axis for the reduced dataset (e.g., -3.4– -1.2). This means that the distribution of occurrences tended to have a tail towards the past, which makes the posterior of θ_1 wider than if it was uniform (cf. Figure 7).

6 Discussion

This is one of the first Bayesian method for stratigraphic intervals to be available in code, and the first Julia package dedicated to estimation of stratigraphic intervals. It is the first to allow flexible specification of a large number of prior distributions, both via the `Distributions.jl` package or implementing custom ones. Also, the possibility of specifying custom MCMC samplers helps attain convergence very easily. The choice of the no-U-turn sampler has proven to be quite useful in general cases, and more efficient than the

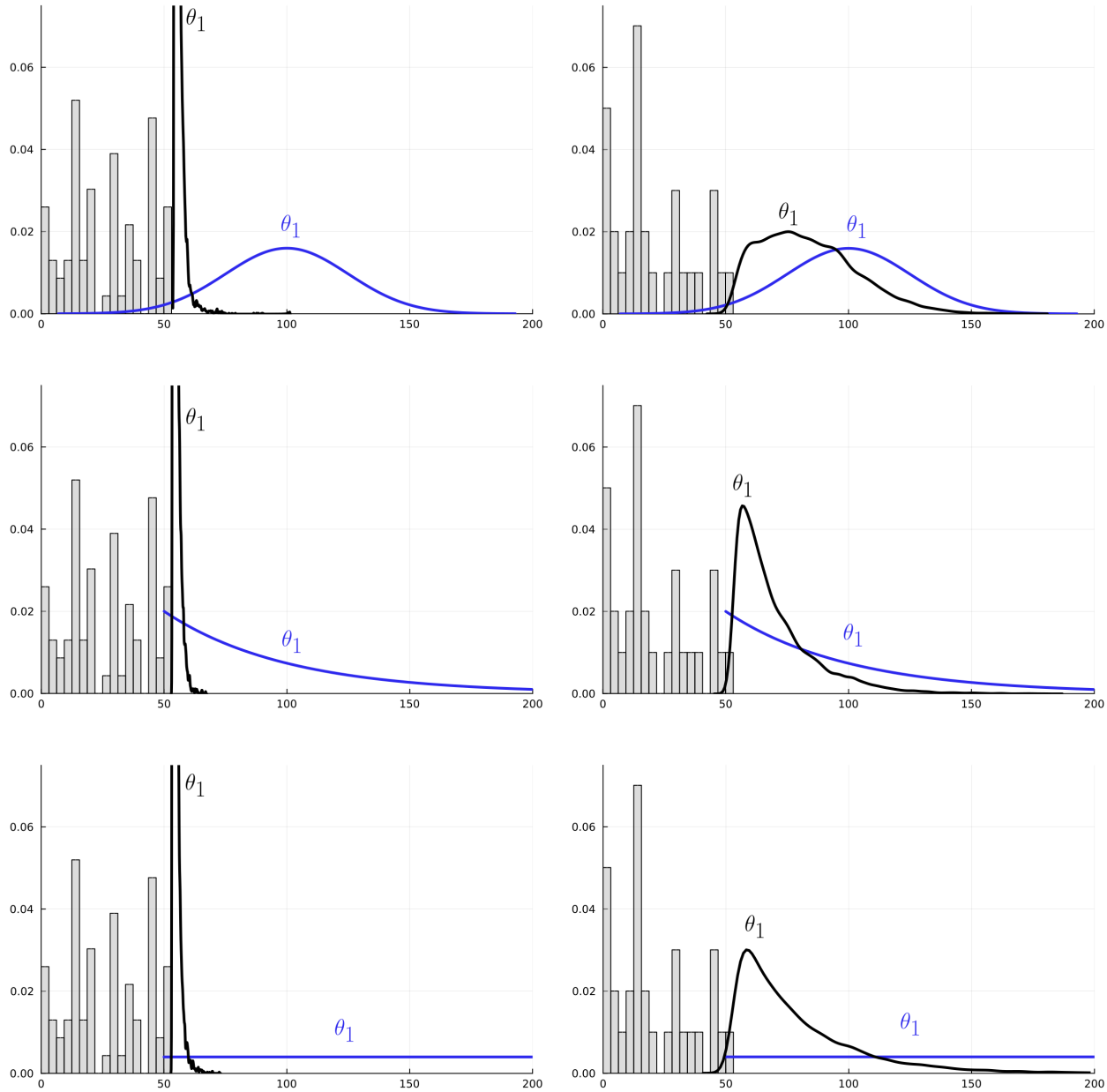

Figure 10: Origination time of the Barracudas. Left column are datasets without removing time duplicates. Right column are datasets removing time duplicates. Top row is a prior $\theta_1 \sim N(100.0, 25.0)$, middle row is a prior $\theta_1 \sim \text{RefOffExp}(50.0, 50.0, 1)$, and bottom row is a prior $\theta_1 \sim U(50.0, 300.0)$. Axes represent time (in Ma) on the horizontal, and density on the vertical. Prior distribution in blue, posterior distribution in black. Data are the grey bars in the histogram.

Hamiltonian Monte Carlo, which requires to set more arguments and may run into issues of convergence. The `Turing.jl` package also provides other samplers including Random-walk Metropolis-Hastings, Gibbs, and compositional Gibbs, thus extending the possibilities of fine-grain control over MCMC sampling. Finally, allowing to fix parameters to given values allows to use stratigraphic intervals in situations where e.g. the lineages in question are extant, but not restricting the user to this as happens with the Adaptive Beta of Wang et al. (2016).

Maximum likelihood and Bayesian estimation. The bias in estimation where the FAD and LAD are the maximum likelihood estimates appear to be issues present in other methods for stratigraphic intervals, not just the present one (Strauss and Sadler, 1989; Marshall, 1990; Rivadeneira et al., 2009; Wang et al., 2016, T. Quental, pers. comm.). This makes the use of Bayesian methods more convenient to for avoiding biased estimates of origination and extinction times. Wang (2005) studied the maximum likelihood properties of the Four-Parameter Beta distribution and found similar issues with parameter estimation, also advocating for constraining the likelihood surface using uninformative priors on the endpoint parameters θ_1, θ_2. He discussed the conditions under which parameter estimation is possible, through the use of Bayesian methods. The approach herein is in line with his findings, but is of more general application as it does not enforce the use of a specific prior on endpoint parameters. Wang (2005) examined the correlation between parameters and concluded that endpoint ones tend to be correlated. This is an issue for inference using Metropolis-Hasting samplers, and it was unnecessary to examine it here because both the Hamiltonian and no-U-turn samplers are very efficient traversing sufaces in these situations.

Priors. The behaviour of different priors was in general terms similar in that given sufficient data, the likelihood function is capable of dominating the information content, thus determining to a great extent the parameter posterior distributions. However, it is clear that even fairly uninformative priors can exert some influence when sample size is small (Figure 10). It is important to note that priors which are very off form the true values (herein called incorrect priors) have an undesirable influence on the posterior distributions, which is proportional to the information content in the prior distribution, and therefore inversely proportional to the variance. Thus, even if a prior is incorrect because its mean is off from the true value, we can make it behave more properly by setting a larger variance for it, that is, making it fairly uninformative. These findings suggest that the method is robust to prior specification, and it is recommended that empirical applications with small sample size (around order to 50 occurrences or less) specify wide priors in order to avoid any undesirable influence of the prior on the posterior. Although priors wehich are both incorrect and informative behave very badly in the sense that they do not recover the true value in simulations, they tend to approach the true values which may eventually converge if letting simulations to run for very long and with very large sample sizes (above 500 in this study). This behaviour can be avoided by following the guidelines

on prior specification above.

Comparison with other methods. The Three-Parameter Beta model, in the Bayesian formulation, shares features with several methods already available. On one side, it assumes continuous sampling, as is common to available tools, but in contrast with methods for discrete sampling (Weiss and Marshall, 1999; Weiss et al., 2004). It can accomodate both uniform (Strauss and Sadler, 1989; Solow, 1993a; Ferraz et al., 2003; Weiss and Marshall, 1999; McInerny et al., 2006; Wang et al., 2016) and non-uniform (Solow, 1993b; Marshall, 1997; McCarthy, 1998; Weiss et al., 2004; Ferraz et al., 2003; Wang et al., 2016) preservation potentials, thus being similar to many Class 1 and Class 2 methods, and in contrast with non-parametric or distribution free methods (Marshall, 1994; Solow and Roberts, 2003). The method can condition on a fixed endpoint parameter like many do (e.g., Wang et al., 2016), in particular those taylored for sighting data in conservation biology, which assume that the oldest occurrence is an arbitrary zero point, and then estimate the extinction time (Ferraz et al., 2003; Solow, 2005; Kodikara et al., 2020), but it is not forced to such conditioning and therefore can estimate both the origination and extinction times like some methods from palaeobiology (Strauss and Sadler, 1989; Marshall, 1994). It is unique however in allowing the user to specify any prior they want which is supported by the `Distributions.jl` package or implemented in `StratIntervals.jl` (e.g., the RefOffExponential distribution). Also, it is unique among many Bayesian methods in that it is fully implemented in a software package which can be easily integrated with others from the Julia language ecosystem. This helps in allowing th user to extend the models available into three directions: Choice of priors, which multiple works have found to matter (Alroy, 2016b,a; Solow, 2016b,a), choice of samplers, therefore allowing to improve sampling efficiency, and the possibility to extend methods by using different likelihood functions.

Empirical applications. The combined effect of the tail of the distribution towards the past, as well as the lower sample size in the reduced dataset explains why the posterior distribution of θ_1 was considerably wider, and should be considered when assessing the results of the application of the present method to empirical datasets. Although the choice of the prior for θ_1 has some influence on the posterior distribution, this is greatly reduced by sample size, under the same inferential settings. If the distribution of occurrences is concentrate towards any of the endpoints, the posterior estimation of the opposite endpoint will suffer from wider variance. This effect is seen both in the Palynomorph and the Barracudas datasets. This is a feature of the data and the distribution rather than a pathological behaviour of the method. It is comparable to trying to fit a curve to an asymmetric distribution, variations on the tailed side can vary to a larger extent than the part attached to the other side. This makes sense as more of the density of the information in the data are concentrated to the other side of the distribution, and therfore we have less information on the tailed side (Figures 8 and 10).

The conflation of distributions showed desirable properties in real-world applications. The fact that we recovered a reasonable position for the Coal seam 115 just by observing the occurrences of only six palynomorphs present in the sample is exciting, as normally much more taxa are available in real-world samples for biostratigraphic analyses (Jaramillo et al., 2007). Its potential however is not restricted to micropalaeontology but instead to any application where combinations of distributions is of interest. For instance, Ballen and Reinales (2024) showed its application to the construction of secondary calibrations in divergence time estimation. Although this family of methods was born from the field of quantitative palaeobiology, its application to questions in evolutionary biology and related fields can be much wider, especially in the field of divergence time estimation.

Calculating the co-occurrence of multiple lineages in time has shown to be useful in estimating the time of co-occurrence of collections of objects in time. This method does not require the intervals to represent living organisms, and therefore can be easily applied to historical objects such as musical pieces which are preserved in multiples sources, e.g. Gregorian chant melodies in manuscripts (Ballen et al., 2024). These scenarios are quite common in the study of cultural evolution and digital humanities, which are interesting fields where data sometimes correspond to series of occurrences in time, the methods presented herein are promising.

7 Conflict of interest

The author declares no conflict of interest.

8 Acknowledgments

This research was supported by FAPESP through a doctoral scholarship (2014/11558-5), an internship abroad (2016/02253-1), and a postdoctoral fellowship (2023/07838-1), as well as by the BBSRC (grant BB/T01282X/1 awarded to Mario dos Reis). I thank Mario dos Reis and Ziheng Yang for valuable feedback during early phases of this project, I could not have finished it without their guidance and criticism in methods development. I thank Phil Donoghue, Davide Pisani and the Molecular Palaeobiology Group at University of Bristol for inviting me to deliver a talk on the topic of this paper when it was still in early phases of development. I thank Tiago Quental and Gustavo Burin for discussions on methods for palaeobiology and macroevolution throughout the years, and for early discussions on methods for stratigraphic intervals since my first year of doctoral studies. I thank Carlos Jaramillo for supporting my career in palaeontology, and for encouraging me and several other Colombian students to build analytical and computational

skills in palaeobiology. Discussions on palaeontolgy and geology with many colleagues (Javier Luque, Jorge W. Moreno, Camila Martinez, Maria Camila Vallejo, Federico Moreno, Andrés Cárdenas, Edwin Cadena, Camilo Montes, Catalina Suarez, and Sandra Reinales) have been a valuable source of inspiration. Living ammonite reconstruction in Figure 1A thanks to Kouhei Futaka, available under Creative Commons license (https://doi.org/10.7875/togopic.2020.08), and ammonite drawing in Figure 1A thanks to Izolende, also available under Creative Commons (https://commons.wikimedia.org/wiki/User:Ilzolende). Special thanks to Jorge W. Moreno-Bernal for the ammonite reconstructions in Figure 1E. Finally I thank Sandra Reinales for supporting me personally and professionally in this and many other projects throughout the years.

References

Alroy, J. (2014). A simple bayesian method of inferring extinction. *Paleobiology*, 40(4):584–607.

Alroy, J. (2015). Current extinction rates of reptiles and amphibians. *Proceedings of the National Academy of Sciences*, 112(42):13003–13008.

Alroy, J. (2016a). Reply to solow: Sense and nonsense in the choice of extinction priors. *Proceedings of the National Academy of Sciences*, 113(9):E1133–E1133.

Alroy, J. (2016b). A simple bayesian method of inferring extinction: reply. *Ecology*, 97(3).

Ballen, G. A. (2020). New records of the genus *Sphyraena* (Teleostei: Sphyraenidae) from the Caribbean with comments on dental characters in the genus. *Journal of Vertebrate Paleontology*, 40(6):e1849246.

Ballen, G. A., Mühlová, K. H., and Hajič jr., J. (2024). What did the dove sing to pope gregory? ancestral melody reconstruction in gregorian chant using bayesian phylogenetics. *BioRxiv*, pages 1–17.

Ballen, G. A. and Reinales, S. (2024). tbea: tools for pre-and post-processing in bayesian evolutionary analyses. *bioRxiv*, pages 2024–06.

Besançon, M., Papamarkou, T., Anthoff, D., Arslan, A., Byrne, S., Lin, D., and Pearson, J. (2021). Distributions.jl: Definition and modeling of probability distributions in the JuliaStats Ecosystem. *Journal of Statistical Software*, 98(16):1–30.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98.

Boakes, E. H., Rout, T. M., and Collen, B. (2015). Inferring species extinction: the use of sighting records. *Methods in Ecology and Evolution*, 6(6):678–687.

Bouchet-Valat, M. and Kamiński, B. (2023). DataFrames.jl: Flexible and fast tabular data in Julia. *Journal of Statistical Software*, 107(4):1–32.

Ferraz, G., Russell, G. J., Stouffer, P. C., Bierregaard Jr, R. O., Pimm, S. L., and Lovejoy, T. E. (2003). Rates of species loss from amazonian forest fragments. *Proceedings of the National Academy of Sciences*, 100(24):14069–14073.

Ge, H., Xu, K., and Ghahramani, Z. (2018). Turing: A language for flexible probabilistic inference. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 1682–1690.

Genest, C. and Zidek, J. V. (1986). Combining probability pistributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–135.

Hill, T. P. (2011). Conflations of probability distributions. *Transactions of the American Mathematical Society*, 363(6):3351–3372.

Hill, T. P. and Miller, J. (2011). How to combine independent data sets for the same quantity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(3):033102.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Holland, S. M., Patzkowsky, M. E., and Loughney, K. M. (2024). Stratigraphic paleobiology. *Paleobiology*, page 1–18.

Jaramillo, C. A., Bayona, G., Pardo-Trujillo, A., Rueda, M., Torres, V., Harrington, G. J., and Mora, G. (2007). The palynology of the cerrejón formation (upper paleocene) of northern colombia. *Palynology*, 31(1):153–189.

Johnson, S. G. (2013). QuadGK.jl: Gauss–Kronrod integration in Julia. https://github.com/JuliaMath/QuadGK.jl.

Kodikara, S., Demirhan, H., Wang, Y., Solow, A., and Stone, L. (2020). Inferring extinction year using a bayesian approach. *Methods in Ecology and Evolution*, 11(8):964–973.

Marshall, C. R. (1990). Confidence intervals on stratigraphic ranges. *Paleobiology*, 16(1):1–10.

Marshall, C. R. (1994). Confidence intervals on stratigraphic ranges: Partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology*, 20(4):459–469.

Marshall, C. R. (1997). Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology*, 23(2):165–173.

Marshall, C. R. (2010). Using confidence intervals to quantify the uncertainty in the end-points of stratigraphic ranges. *Quantitative Methods in Paleobiology*, 16:291–316.

McCarthy, M. A. (1998). Identifying declining and threatened species with museum data. *Biological conservation*, 83(1):9–17.

McCrea, R. S., Cheale, T., Campillo-Funollet, E., and Roberts, D. L. (2024). Inferring species extinction from sighting data. *Cambridge Prisms: Extinction*, 2:e19.

McInerny, G. J., Roberts, D. L., Davy, A. J., and Cribb, P. J. (2006). Significance of sighting rate in inferring extinction and threat. *Conservation biology*, 20(2):562–567.

Rivadeneira, M. M., Hunt, G., and Roy, K. (2009). The use of sighting records to infer species extinctions: an evaluation of different methods. *Ecology*, 90(5):1291–1300.

Signor, P. W. and Lipps, J. H. (1982). Sampling bias, gradual extinction patterns and catastrophes in the fossil record. *Geological Society of America Special Paper*, 190:291–296.

Smith, J., Dowding, E., Abdelhady, A., Abondio, P., Araujo, R., Aze, T., Balisi, M., Buatois, L., Carvajal-Chitty, H., Chattopadhyay, D., Coiro, M., Dietl, G., Gonzalez Arango, C., Kevrekidis, C., Kimmig, J., Mychajliw, A., Pimiento, C., Regalado Fernandez, O., K.M., S., Warnock, R., Yang, T., Yasuhara, M., Akita, L., Allen, B., Anderson, B., Andreoletti, J., Archuby, F., Ballen, G., Bari, M., Benton, M., Bergh, E., Brambilla, L., Brombacher, A., Chan, Y., Chiarenza, A., Chinzorig, T., Coates, K., Cordie, D., Cortes-Sanchez, M., Cruz-Vega, E., Cybulski, J., De Baets, K., De Entrambasaguas, J., Dillon, E., Du, A., Dunhill, A., Erlandson, J., Forel, M., Foster, W., Gates, T., Gavryushkina, A., Grace, M., Grossart, H., Hansel, P., Harnik, P., Hopkins, M., Hopkins, S., Hu, K., Huang, H., Irmis, R., Jaques, V., Jenkins, X., Jukar, A., Kelley, P., Kihn, R., Klompmaker, A., Kocsis, A., Kriwet, J., Lazarus, D., Liao, C., Lin, C., Louys, J., Lozano-Fernandez, J., Lozano-Francisco, M., Lueders-Dumont, J., Malve, M., Martindale, R., Mazzini, I., Modenini, G., Mondal, S., Mondini, M., Monferran, M., Mulvey, L., Nanglu, K., Nguyen, J., Norris, R., O'Dea, A., Ollendorf, A., Orihuela, J., Pandolfi, J., Pereira, T., Piro, A., Plotnick, R., Plaza-Torres, S., Porto, A., Prieto-Marquez, A., Punyasena, S., Quental, T., Raja, N., Ranaivosoa, V., Ribas-Deulofeu, L., Rivals, F., Roden, V., Rosso, A., Saleh, F., Salvador, R., Saupe, E., Schneider, S., Sclafani, J., Smith, M.,

Souron, A., Steinbauer, M., Stewart, M., Tambussi, C., Thomas, E., Tschopp, E., Tutken, T., Varela, S., Vezzosi, R., Villasenor, A., Weinkauf, M., Zanno, L., Zhang, C., Zhao, Q., and Kiessling, W. (2025). Big questions in paleontology: A community-driven project to motivate new insights about the history of life on earth. *Paleobiology (In press)*, XX:XX–XX.

Solow, A. R. (1993a). Inferring extinction from sighting data. *Ecology*, 74(3):962–964.

Solow, A. R. (1993b). Inferring extinction in a declining population. *Journal of Mathematical Biology*, 32:79–82.

Solow, A. R. (2005). Inferring extinction from a sighting record. *Mathematical biosciences*, 195(1):47–55.

Solow, A. R. (2016a). On bayesian inference about extinction. *Proceedings of the National Academy of Sciences*, 113(9):E1132–E1132.

Solow, A. R. (2016b). A simple bayesian method of inferring extinction: comment. *Ecology*, 97(3).

Solow, A. R. and Roberts, D. L. (2003). A nonparametric test for extinction based on a sighting record. *Ecology*, 84(5):1329–1332.

Strauss, D. and Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology*, 21(4):411–427.

Tange, O. (2011). Gnu parallel–the command-line power tool. *USENIX Mag*, 36(1):42.

Tange, O. (2024). GNU Parallel 20240522 ('Tbilisi'). https://doi.org/10.5281/zenodo.11247979.

Wang, J. Z. (2005). A note on estimation in the four-parameter beta distribution. *Communications in Statistics - Simulation and Computation*, 34(3):495–501.

Wang, S. C., Everson, P. J., Zhou, H. J., Park, D., and Chudzicki, D. J. (2016). Adaptive credible intervals on stratigraphic ranges when recovery potential is unknown. *Paleobiology*, 42(2):240–256.

Weiss, R. E., Basu, S., and Marshall, C. R. (2004). A framework for analysing fossil record data. In *Tools for constructing chronologies: crossing disciplinary boundaries*, pages 213–230. Springer.

Weiss, R. E. and Marshall, C. R. (1999). The uncertainty in the true end point of a fossil's stratigraphic range when stratigraphic sections are sampled discretely. *Mathematical Geology*, 31:435–453.