

# Materials and Methods

## Taxon occurrences and species-level information

All fossil occurrence information used in this analysis was downloaded from the Paleobiology Database (PBDB). The initial download restricted all occurrences to all Mammalia observed in North America between the Maastriichtian (72-66 Mya) and Gelasian (2.58-1.8 Mya) stages [? ]. Occurrences were then further limited to those occurring between 64 and 2 million years ago (Mya). Taxonomic, stratigraphic, and ecological metadata for each occurrence and species was also downloaded. A new download for a raw, unfiltered PBDB datafile following the same criterion used here is available at <http://goo.gl/2slgeU>. The raw datafiles used as a part of this study, along with all code for filtering and manipulating this download is available at <http://github.com/psmits/coping>.

After being downloaded, the raw occurrence data was then sorted, cleaned, and manipulated programmatically before analysis. Many species taxonomic assignments as present in the raw PBDB data were updated for accuracy and consistency. For example, species classified in the order Artiodactyla were reclassified as Cetartiodactyla. These re-assignments follow [? ] which were based on taxonomies present in the Encyclopedia of Life (<http://eol.org>) and [? ]. All taxa whose life habit was classified as either volant (i.e. Chiroptera) or aquatic (e.g. Cetacea) were excluded from this analysis because of both differences in fossilization potential and environmental context as well as their lack of direct applicability to the study of terrestrial species pools.

Species ecotype is defined here as the interaction between locomotor category and diet categories, the goal of which is to classify species based on the manner with which they interact with their environment. Most mammal species records in the PBDB have life habit (i.e. locomotor category) and dietary category assignments. In order to simplify interpretation, analysis, and per ecotype sample size these classifications were coarsened in a similar manner to [? ] following Table 1. Additionally for this study, the life history category was further broken up to better reflect the diversity of mammal locomotor modes. Ground dwelling species locomotor categories were reassigned based on their ankle posture associated with their taxonomic group as described in Table ?? [? ]. Ecotype categories with less than 10 total species of that

combination were excluded, yielding a total of 18 observed ecotypes out of a possible 24.

Table 1: Species trait assignments in this study are a coarser version of the information available in the PBDB. Information was coarsened to improve per category sample size and uniformity and followed this table.

This study		PBDB categories
Diet	Carnivore	Carnivore
	Herbivore	Browser, folivore, granivore, grazer, herbivore.
	Insectivore	Insectivore.
	Omnivore	Frugivore, omnivore.
Locomotor	Arboreal	Arboreal.
	Ground dwelling	Fossorial, ground dwelling, semifossorial, saltatorial.
	Scansorial	Scansorial.

Table 2: Ankle posture assignment as based on taxonomy. Assignments are based on [? ]. Taxonomic groups are presented alphabetically and without reference for the nestedness of families in orders. As such, do not infer higher-order structure from this table.

	Order	Family	Stance
Ailuridae		plantigrade	
Allomyidae		plantigrade	
Amphicyonidae		plantigrade	
Amphilemuridae		plantigrade	
Anthracotheriidae		digitigrade	
Antilocapridae		unguligrade	
Apheliscidae		plantigrade	
Aplodontidae		plantigrade	
Apternodontidae		scansorial	
Arctocyonidae		unguligrade	
Barbourofelidae		digitigrade	
Barylambdidae		plantigrade	
Bovidae		unguligrade	
Camelidae		unguligrade	
Canidae		digitigrade	
Continued on next page			

**Table 2 – continued from previous page**

Order	Family	Stance
Dinocerata	Cervidae	unguligrade
	Cimolodontidae	scansorial
	Coryphodontidae	plantigrade
	Cricetidae	plantigrade
	Cylindrodontidae	plantigrade
	Cyriacotheriidae	plantigrade
	Dichobunidae	unguligrade
		unguligrade
	Dipodidae	digitigrade
	Elephantidae	digitigrade
	Entelodontidae	unguligrade
	Eomyidae	plantigrade
	Erethizontidae	plantigrade
	Erinaceidae	plantigrade
	Esthonychidae	plantigrade
	Eutypomyidae	plantigrade
	Felidae	digitigrade
	Florentiamyidae	plantigrade
	Gelocidae	unguligrade
	Geolabididae	plantigrade
	Glyptodontidae	plantigrade
	Gomphotheriidae	unguligrade
	Hapalodectidae	plantigrade
	Heteromyidae	digitigrade
	Hyaenidae	digitigrade
	Hyaenodontidae	digitigrade
	Hypertragulidae	unguligrade
	Ischyromyidae	plantigrade
	Jimomyidae	plantigrade
Lagomorpha		digitigrade
	Leptictidae	plantigrade
	Leptochoeridae	unguligrade
	Leptomerycidae	unguligrade
	Mammutidae	unguligrade
	Megalonychidae	plantigrade
Continued on next page		

**Table 2 – continued from previous page**

Order	Family	Stance
Mesonychia	Megatheriidae	plantigrade
	Mephitidae	plantigrade
	Merycoidodontidae	digitigrade
		unguligrade
	Mesonychidae	digitigrade
	Micropternodontidae	plantigrade
	Mixodectidae	plantigrade
	Moschidae	unguligrade
	Muridae	plantigrade
	Mustelidae	plantigrade
	Mylagaulidae	fossorial
	Mylodontidae	plantigrade
	Nimravidae	digitigrade
	Nothrotheriidae	plantigrade
Notoungulata		unguligrade
	Oromerycidae	unguligrade
	Oxyaenidae	digitigrade
	Palaeomerycidae	unguligrade
	Palaeoryctidae	plantigrade
	Pampatheriidae	plantigrade
	Pantolambdidae	plantigrade
	Periptychidae	digitigrade
		unguligrade
	Phenacodontidae	unguligrade
Perissodactyla		plantigrade
Primates	Procyonidae	plantigrade
	Proscalopidae	plantigrade
	Protoceratidae	unguligrade
	Reithroparamyidae	plantigrade
	Sciuravidae	plantigrade
	Sciuridae	plantigrade
	Simimyidae	plantigrade
	Soricidae	plantigrade
	Suidae	digitigrade
	Talpidae	fossorial
Continued on next page		

**Table 2 – continued from previous page**

Order	Family	Stance
	Tayassuidae	unguligrade
	Tenrecidae	plantigrade
	Titanoideidae	plantigrade
	Ursidae	plantigrade
	Viverravidae	plantigrade
	Zapodidae	plantigrade

Estimates of species mass used as a covariate in this study were sourced from multiple databases and papers, especially those focusing on similar macroevolutionary or macroecological questions [? ? ? ? ? ]; this is similar to what was done in ? ]. When species mass was not available, proxy measures were used and then transformed into estimates of mass. For example, given a measurement of a mammal tooth size, it is possible and routine to estimate its mass given some regression equation. The PBDB has one or more body part measures for many species. These were used as body size proxies for many species, as was the case in ? ]. Mass was log-transformed and then mean-centered and rescaled by dividing by two-times its standard deviation; this insures that the magnitude of effects for both continuous and discrete covariates are directly comparable [? ? ].

All fossil occurrences from 64 to 2 million years long (Mya) were binned into 31 2 million year (My) bins. This temporal length was chosen because it is approximately the resolution of the North American mammal fossil record [? ? ? ].

## Environmental and temporal covariates

The environmental covariates used in this study are collectively referred to as group-level covariates because they predict the response of a “group” of individual-level observations (i.e. species occurrences of an ecotype). Additionally, these covariates are defined for temporal bins and not the species themselves; as such they predict the parts of each species occurrence history. The group-level covariates in this study are two global temperature estimates

Table 3: Regression equations used in this study for estimating body size. Equations are presented with reference to taxonomic grouping, part name, and reference.

Group	Equation	log(Measurement)	Source
General	$\log(m) = 1.827x + 1.81$	lower m1 area	[? ]
General	$\log(m) = 2.9677x - 5.6712$	mandible length	[? ]
General	$\log(m) = 3.68x - 3.83$	skull length	[? ]
Carnivores	$\log(m) = 2.97x + 1.681$	lower m1 length	[? ]
Insectivores	$\log(m) = 1.628x + 1.726$	lower m1 area	[? ]
Insectivores	$\log(m) = 1.714x + 0.886$	upper M1 area	[? ]
Lagomorph	$\log(m) = 2.671x - 2.671$	lower toothrow area	[? ]
Lagomorph	$\log(m) = 4.468x - 3.002$	lower m1 length	[? ]
Marsupials	$\log(m) = 3.284x + 1.83$	upper M1 length	[? ]
Marsupials	$\log(m) = 1.733x + 1.571$	upper M1 area	[? ]
Rodentia	$\log(m) = 1.767x + 2.172$	lower m1 area	[? ]
Ungulates	$\log(m) = 1.516x + 3.757$	lower m1 area	[? ]
Ungulates	$\log(m) = 3.076x + 2.366$	lower m2 length	[? ]
Ungulates	$\log(m) = 1.518x + 2.792$	lower m2 area	[? ]
Ungulates	$\log(m) = 3.113x - 1.374$	lower toothrow length	[? ]

and the Cenozoic “plant phases” defined by [? ]. Global temperature across most of the Cenozoic was calculated from Mg/Ca isotope record from deep sea carbonates [? ]. Mg/Ca based temperature estimates are preferable to the frequently used  $\delta^{18}\text{O}$  temperature proxy [? ? ? ? ] because Mg/Ca estimates do not conflate temperature with ice sheet volume and depth/stratification changes. The former is particularly important to this analysis as the current polar ice-caps appeared and grew over the Cenozoic. These properties make Mg/Ca based temperature estimates preferable for macroevolutionary and macroecological studies [? ]. Two aspects of the Mg/Ca-based temperature curve were included in this analysis: mean and range. Both were calculated as the mean of all respective estimates for each 2 My temporal bins. Both mean and range were then rescaled by subtracting the mean and then dividing by twice the standard deviation.

The second set of environmental factors included in this study are the Cenozoic plant phases defined in [? ]. Graham’s plant phases are holistic descriptors of the taxonomic composition of 12 ecosystem types, which plants are present at

Table 4: Definitions of the start and stop times of the three plant phases used this study as defined by [? ].

Plant phase	Phase number	Start	Stop
Paleocene-Eocene	1	66	50
Eocene-Miocene	2	50	16
Miocene-Pleistocene	3	16	2

a given time, and the relative modernity of those plant groups with younger phases representing increasingly modern taxa [? ]. [? ] defines four intervals from the Cretaceous to the Pliocene, though only three of these intervals are included in this analysis. Graham’s plant phases was included as a series of “dummy variables” encoding the three phases included in this analysis [? ]; this means that the first phase is synonymous with the intercept and subsequent phases are defined by their differences from the first phase. The temporal boundaries of these plant phases are defined in Table ??.

## Modelling species occurrence

Two different models were used in this study: a pure-presence model and a birth-death model. Both models at their core are a hidden Markov model where the latent process has an absorbing state [? ]. The difference between these two models is if the probabilities of a species originating or surviving are considered equal or different (Table ??). Something that is important to realize is that while there are only two state “codes” in a presence-absence matrix (i.e. 0/1), there are in fact three states in a birth-death model: not having originated yet, extant, and extinct. The last of these is the absorbing state, as once a species has gone extinct it cannot re-originate [? ]; this is made obvious in the transition matrices as the probability of an extinct species changing states is 0 (Table ??). See below for parameter explanations (Tables 3, 4, and 5).

## Data augmentation

All empirical presence/absence observations are potentially incomplete or observed with error. The hidden Markov model at the core of this analysis

		State at $t + 1$					State at $t + 1$		
		$0_{never}$	1	$0_{extinct}$			$0_{never}$	1	$0_{extinct}$
State at $t$	$0_{never}$	$1 - \theta$	$\theta$	0	State at $t$	$0_{never}$	$1 - \phi$	$\phi$	0
	1	0	$\theta$	1		1	0	$\pi$	$1 - \pi$
	$0_{extinct}$	0	0	1		$0_{extinct}$	0	0	1

(a) Pure-presence

(b) Birth-death

Table 5: Transition matrices for the pure-presence (??) and birth-death (??) models. Both of these models share the core machinery of discrete-time birth-death processes but make distinct assumptions about the equality of originating and surviving (Eq. 2, and 3). Note also that while there are only two state “codes” (0, 1), there are in fact three states: never having originated  $0_{never}$ , present 1, extinct  $0_{extinct}$  [? ].

allows for observed absences to be used meaningfully to estimate the number of unobserved species. Of concern in this analysis is the unknown “true” size of the dataset; how many species could have actually been observed? While many species have been observed, the natural incompleteness of all observations, especially in the case of paleontological data, there are obviously many species which were never sampled [? ? ].

Let  $N$  be the total number of observed species,  $M$  be the upper limit of possible species that could have existed given some model of species occurrence, and  $N^*$  is the number of all-zero histories added to the presence absence matrix  $y$  where  $N^* = M - N$ . This approach assumes that  $\hat{N} \sim \text{Binomial}(M, \psi)$  where  $\hat{N}$  is the estimated “true” number of species and  $\psi$  is the probability that any augmented species should actually be “present.” Because  $M$  is user defined, this approach effectively gives  $\psi$  a uniform prior over  $N$  to  $M$  [? ]. For this study,  $M = \lfloor 1.25 \times N \rfloor$ .

Data imputation is the process of estimating missing data for partially observed covariates given the other fully-observed observations and some model [? ? ], this is simple in a Bayesian context because data are also parameters [? ]. Augmented species are fully imputed species and thus have no known mass so a mass estimate must be imputed for each possible species [? ]. Assuming that mass values for augmented species are from the same distribution as observed species, the distribution of observed mass values are estimated as part of the model and new mass values are then generated from this distribution. This



Table 6: Parameters associated with the observation process part of the hidden Markov model.

Parameter	dimensions	explanation
$y$	$N \times T$	observed species presence/absence
$z$	$N \times T$	“true” species presence/absence
$p$	$T$	probability of observing a species that is present at time $t$
$m$	$N$	species log mass, rescaled
$\alpha_0$	1	average log-odds of $p$
$\alpha_1$	1	change in average log-odds of $p$ per change mass
$r$	$T$	difference from $\alpha_0$ associated with time $t$
$\sigma$	1	standard deviation of $r$

approach is an example of imputing covariate information that is missing completely at random [? ?]. Because log mass values are rescaled as a part of this study, the body mass distribution is already known ( $\mathcal{N}(0, 0.5)$ ) the body mass of the augmented species are generated by simple random draws from this distribution. In addition to body mass information, the augmented species need an ecotype classification. Because these species are completely unknown, they were all classified as “augmented” to indicate their unknown biology. This classification has no biological interpretation.

### Observation process

The type of hidden Markov model used in this study has three characteristic probabilities: probability  $p$  of observing a species given that it is present, probability  $\phi$  of a species surviving from one time to another, and probability  $\pi$  of a species first appearing [? ]. In this formulation, the probability of a species going extinct is  $1 - \pi$ . For the pure-presence model  $\phi = \pi$ , while for the birth-death model  $\phi \neq \pi$ .

The probability of observing a species that is present  $p$  is modeled as a logistic regression was a time-varying intercept and species mass as a covariate. The effect of species mass on  $p$  was assumed linear and constant over time and given a prior reflecting a possible positive relationship; these assumptions are

Table 7: Parameters for the model of presence in the pure-presence model

Parameter	dimensions	explanation
$z$	$N \times T$	“true” species presence/absence
$\theta$	$N \times T - 1$	probability of $z = 1$
$a$	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of $\theta$
$m$	$N$	species log mass, rescaled
$b_1$	1	effect of species mass on log-odds of $\theta$
$b_2$	1	effect of species mass, squared, on log-odds of $\theta$
$U$	$T \times D$	matrix of group-level covariates
$\gamma$	$U \times D$	matrix of group-level regression coefficients
$\Sigma$	$D \times D$	covariance matrix of $a$
$\Omega$	$D \times D$	correlation matrix of $a$
$\tau$	$D$	vector of standard deviations for each ecotype $a_d$

reflected in the structure of the model Equation 1, presented here:

$$\begin{aligned}
 y_{i,t} &\sim \text{Bernoulli}(p_{i,t}z_{i,t}) \\
 p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) \\
 r_t &\sim \mathcal{N}(0, \sigma).
 \end{aligned} \tag{1}$$

The parameters associated with Equation 1 are described in Table 3.

### Pure-presence process

For the pure-presence model there is only a single probability dealing with the presence of a species  $\theta$  (Table ??). This probability was modeled as multi-level logistic regression with both species-level and group-level covariates [? ?]. The parameters associated with pure-presence model are presented in Table 4 and the full sampling statement in Equation 2.

The species-level of the model (Eq. 2) is a logistic regression where the intercept varies by ecotype. Additionally, species mass was included as a covariate associated with two regression coefficients allowing a quadratic relationship with log-odds of occurrence. This assumption is based on the known distribution of mammal body masses where species with intermediate mass values are more common than either small or large bodied species.

These assumptions are also reflected in the choice of priors for these regression coefficients.

The values of each ecotype’s intercept are themselves modeled as regressions using the group-level covariates associated with environmental context. Each of these regressions has an associated variance of possible values of each ecotype’s intercept [? ]. In addition, the covariances between ecotype intercepts, given this group-level regression, are modeled [? ].

All parameters not modeled elsewhere were given weakly informative priors [? ? ]. Weakly informative means that priors do not necessarily encode actual prior information but instead help regularize or weakly constrain posterior estimates. These priors have a concentrated probability density around and near zero; this has the effect of tempering our estimates and help prevent overfitting the model to the data [? ? ? ]. The general line of thinking behind this approach is that a result of 0 or “no effect” is more preferable to a wrong or extremely weak result. The sampling statement, excluding the imputation of body mass associated with the augment species, is as follows

$$\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(p_{i,t}z_{i,t}) & \alpha_0 &\sim \mathcal{N}(0, 1) \\
p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) & \alpha_1 &\sim \mathcal{N}(1, 1) \\
r_t &\sim \mathcal{N}(0, \sigma) & \sigma &\sim \mathcal{N}^+(1) \\
z_{i,1} &\sim \text{Bernoulli}(\rho) & b_1 &\sim \mathcal{N}(0, 1) \\
z_{i,t} &\sim \text{Bernoulli}(\theta_{i,t}) & b_2 &\sim \mathcal{N}(-1, 1) \\
\theta_{i,t} &= \text{logit}^{-1}(a_{t,j[i]} + b_1 m_i + b_2 m_i^2) & \gamma &\sim \mathcal{N}(0, 1) \\
a &\sim \text{MVN}(u\gamma, \Sigma) & \tau &\sim \mathcal{N}^+(1) \\
\Sigma &= \text{diag}(\tau)\Omega\text{diag}(\tau) & \Omega &\sim \text{LKJ}(2)
\end{aligned} \tag{2}$$

### Birth-death process

In the birth-death version of the model,  $\phi \neq \pi$  and so each of these probabilities are modeled separately but each is handled in a similar manner to how  $\theta$  is modeled in the pure-presence model (Eq. 2, Table ??). The parameters associated with the birth-death presence model are presented in Table 5 and the full sampling statement, including observation (Eq. 1), is described in

Equation 3:

$$\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(p_{i,t}z_{i,t}) & \Sigma^\phi &= \text{diag}(\tau^\phi)\Omega^\phi\text{diag}(\tau^\phi) \\
p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) & \Sigma^\pi &= \text{diag}(\tau^\pi)\Omega^\pi\text{diag}(\tau^\pi) \\
r_t &\sim \mathcal{N}(0, \sigma) & \rho &\sim \text{U}(0, 1) \\
\alpha_0 &\sim \mathcal{N}(0, 1) & b_1^\phi &\sim \mathcal{N}(0, 1) \\
\alpha_1 &\sim \mathcal{N}(1, 1) & b_1^\pi &\sim \mathcal{N}(0, 1) \\
\sigma &\sim \mathcal{N}^+(1) & b_2^\phi &\sim \mathcal{N}(-1, 1) \\
z_{i,1} &\sim \text{Bernoulli}(\phi_{i,1}) & b_2^\pi &\sim \mathcal{N}(-1, 1) \\
z_{i,t} &\sim \text{Bernoulli}\left(z_{i,t-1}\pi_{i,t} + \sum_{x=1}^t (1 - z_{i,x})\phi_{i,t}\right) & \gamma^\phi &\sim \mathcal{N}(0, 1) \\
\phi_{i,t} &= \text{logit}^{-1}(a_{t,j[i]}^\phi + b_1^\phi m_i + b_2^\phi m_i^2) & \gamma^\pi &\sim \mathcal{N}(0, 1) \\
\pi_{i,t} &= \text{logit}^{-1}(a_{t,j[i]}^\pi + b_1^\pi m_i + b_2^\pi m_i^2) & \tau^\phi &\sim \mathcal{N}^+(1) \\
a^\phi &\sim \text{MVN}(U\gamma^\phi, \Sigma^\phi) & \tau^\pi &\sim \mathcal{N}^+(1) \\
a^\pi &\sim \text{MVN}(U\gamma^\pi, \Sigma^\pi) & \Omega^\phi &\sim \text{LKJ}(2) \\
& & \Omega^\pi &\sim \text{LKJ}(2).
\end{aligned} \tag{3}$$

Similar to the pure-presence model, both  $\phi$  and  $\pi$  are modeled as logistic regressions with varying-intercept and one covariate associated with two parameters. The possible relationships between mass and both  $\phi$  and  $\pi$  are reflected in the parameterization of the model and choice of priors (Eq. 3).

The intercepts of  $\phi$  and  $\pi$  both vary by species ecotype and those values are themselves the product of group-level regression using environmental factors as covariates (Eq. 3); this is identical to the pure presence model (Eq. 2).

## Posterior inference and model adequacy

Computer programs that implement joint posterior inference for the above models (Eqs. 2, 3) were written in the probabilistic programming language Stan [? ]. The models used here both feature latent discrete parameters in the large matrix  $z$  (Tables 3, 4, 5; Eqs. 1, 2, 3). All methods for posterior inference implemented in Stan are derivative based which causes complications for actually implementing the above models because integers do not have

Table 8: Parameters for the model of presence in the pure-presence model

Parameter	dimensions	explanation
$z$	$N \times T$	“true” species presence/absence
$\phi$	$N \times T$	probability of $z_{-,t} = 1   z_{-,t-1} = 0$ ; origination
$\pi$	$N \times T - 1$	probability of $z_{-,t} = 1   z_{-,t-1} = 1$ ; survival
$a^\phi$	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of $\theta$
$a^\pi$	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of $\theta$
$m$	$N$	species log mass, rescaled
$b_1^\phi$	1	effect of species mass on log-odds of $\phi$
$b_1^\pi$	1	effect of species mass on log-odds of $\pi$
$b_2^\phi$	1	effect of species mass, squared, on log-odds of $\phi$
$b_2^\pi$	1	effect of species mass, squared, on log-odds of $\pi$
$U$	$T \times D$	matrix of group-level covariates
$\gamma^\phi$	$U \times D$	matrix of group-level regression coefficients
$\gamma^\pi$	$U \times D$	matrix of group-level regression coefficients
$\Sigma^\phi$	$D \times D$	covariance matrix of $a^\phi$
$\Sigma^\pi$	$D \times D$	covariance matrix of $a^\pi$
$\Omega^\phi$	$D \times D$	correlation matrix of $a^\phi$
$\Omega^\pi$	$D \times D$	correlation matrix of $a^\pi$
$\tau^\phi$	$D$	vector of standard deviations for each ecotype $a_d^\phi$
$\tau^\pi$	$D$	vector of standard deviations for each ecotype $a_d^\pi$

derivatives. Instead of implementing a latent discrete parameterization, the log posterior probabilities of all possible states of the latent parameters  $z$  were calculated and summed (i.e. marginalized).

Species durations at minimum range-through from the FAD to the LAD, but the incompleteness of all observations means that the actual time of origination or extinction is unknown. The marginalization approach used here means that the probabilities all possible histories for a species are calculated, from the end members of the species having existed for the entire study interval and the species having only existed between the directly observed FAD and LAD to all possible intermediaries (Fig ??) [? ]. This process is identical, language-wise, to assuming range-through and then estimating the possibility of range extension due to incomplete sampling.

	Time Bin							
	1	2	3	4	5	6	7	8
Observed	0	0	0	1	0	1	1	0
Certain	?	?	?	1	1	1	1	?
Potential	0	0	0	1	1	1	1	0
Potential	0	0	1	1	1	1	1	0
Potential	0	1	1	1	1	1	1	0
Potential	1	1	1	1	1	1	1	0
Potential	0	0	0	1	1	1	1	1
Potential	0	0	1	1	1	1	1	1
Potential	0	1	1	1	1	1	1	1
Potential	1	1	1	1	1	1	1	1

Figure 1: Conceptual figure of all possible occurrence histories for an observed species. The first row represents the observed presence/absence pattern for a single species at eight time points. The second row corresponds to the known aspects of the “true” occurrence history of that species. The remaining rows correspond to all possible occurrence histories that are consistent with the observed data. The process of parameter marginalization described in the text

The combined size of the dataset and large number of parameters in both

models (Eqs. 2, 3), specifically the total number of latent parameters that are the matrix  $z$ , means that stochastic approximate posterior inference is computationally very slow even using NUTS based HMC implemented. Instead, an approximate Bayesian approach was used: variational inference. A recently developed automatic variational inference algorithm called “automatic differentiation variational inference” (ADVI) is implemented in Stan and was used here [? ?]. ADVI assumes that the posterior is Gaussian but still yields a true Bayesian posterior; this assumption is similar to quadratic approximation of the likelihood function commonly used in maximum likelihood based inference [?]. The principal limitation of assuming the joint posterior is Gaussian is that the true topology of the log-posterior isn’t estimated; this is a particular burden for scale parameters which are bound to be positive (e.g. standard deviation).

Of additional concern for posterior inference is the partial identifiability of observation parameters  $p_{t=1}$  and  $p_{t=T}$  [?]. This issue means that the estimates of sampling probabilities at the “edges” of the time series cannot fully be estimated because there are no known “gaps” in species occurrence histories that are guaranteed to be filled. Instead, the values of the first and final columns of the “true” presence-absence matrix  $z$  for those observations that do not already have presences in the observed presence-absence matrix  $y$  cannot be estimated [?]. The hierarchical modeling approach used here helps mitigate this problem by pulling the values of  $p_{t=1}$  and  $p_{t=T}$  towards the overall mean of  $p$  [?], and in fact this approach might be more analytically sound than the more ad-hoc approaches that are occasionally used to overcome this hurdle [?]. Additionally, because  $p_{t=1}$  and  $p_{t=T}$  are only partially identifiable, estimates of occurrence  $\theta$  and origination  $\phi$  at  $t = 1$  and estimates of  $\theta$ ,  $\phi$  and survival  $pi$  at  $t = T$  may suffer from similar edge effects. Again, the hierarchical modeling approach used here may help correct for this reality by drawing these estimates towards the overall means of those parameters.

After fitting both models (Eqs. 2, 3) using ADVI, model adequacy and quality of fit was assessed using a posterior predictive check [?]. Because all Bayesian models are inherently generative, simulations of new data sets is “free” [? ?]. By simulating 100 theoretical data sets from the posterior estimates of the model parameters and the observed covariate information the congruence between predictions made by the model and the observed empirical data can be assessed. These datasets are simulated by starting with the observed states of the presence-absence matrix at  $t = 1$ ; from there, the time series roll forward

as stochastic processes with covariate information given from the empirical observations. Importantly, this is fundamentally different from observing the posterior estimates of the “true” presence-absence matrix  $z$ . The posterior predictive check used in this study is to compare the observed average number of observations per species to a distribution of simulated averages; if the empirically observed value sits in the middle of the distribution than the model can be considered adequate in reproducing the observed number of occurrences per species.

The ADVI assumption of a purely Gaussian posterior limits the utility and accuracy of the posterior predictive checks because parameter estimates do not reflect the true posterior distribution and are instead just an approximation [? ]. Because of this, posterior predictive estimates are themselves only approximate checks of model adequacy. The posterior predictive check that is used in this study focuses on mean occurrence and not to any scale parameters that might be most affected by the ADVI assumptions.

Given parameter estimates, diversity and diversification rates are estimated through posterior predictive simulations. Given the observed presence-absence matrix  $y$ , estimates of the true presence-absence matrix  $z$  can be simulated and the distribution of possible occurrence histories can be analyzed. This is conceptually similar to marginalization where the probability of each possible occurrence history is estimated (Fig. ??).

The posterior distribution of  $z$  gives the estimate of standing diversity  $N_t^{stand}$  for all time points as

$$N_t^{stand} = \sum_{i=1}^M z_{i,t}. \quad (4)$$

Given estimates of  $N_t^{stand}$  for all time points, the estimated number of originations  $O_t$  are be estimated as

$$O_t = \sum_{i=1}^M z_{i,t} = 1 | z_{i,t-1} = 0 \quad (5)$$

and number of extinctions  $E_t$  estimated as

$$E_t = \sum_{i=1}^M z_{i,t} = 0 | z_{i,t-1} = 1. \quad (6)$$



Per-captia growth  $D^{rate}$ , origination  $O^{rate}$  and extinction  $E^{rate}$  rates are then calculated as

$$\begin{aligned} O_t^{rate} &= \frac{O_t}{N_{t-1}^{stand}} \\ E_t^{rate} &= \frac{E_t}{N_{t-1}^{stand}} \\ D_t^{rate} &= O_t^{rate} - E_t^{rate}. \end{aligned} \tag{7}$$