PALEOBIOLOGY
A PUBLICATION OF THE
Paleontological
S O C I E T Y

Article

Relative species abundance and population densities of the past: developing multispecies occupancy models for fossil data

Trond Reitan* , Torbjørn H. Ergon , and Lee Hsiang Liow

Abstract.—The number of individuals of species varies, but estimating abundance, given incomplete and biased sampling in both contemporary and fossilized communities, is challenging. Here, we describe a new occupancy model in a hierarchical Bayesian framework with random effects, in which multispecies occupancy and detection are modeled as a means to estimate relative species abundance and relative population densities. The modeling framework is suited for temporal samples of fossil communities with repeated sampling including multiple species with similar preservation potential. We demonstrate our modeling framework using a fossil community of benthic organisms to estimate relative species abundance dynamics and changing relative population densities of focal species in nine (geological) time intervals over 2.3 Myr. We also explore potential explanatory factors (paleoenvironmental proxies) and temporal autocorrelation that could provide extra information on unsampled time intervals. The modeling framework is applicable across a wide range of questions on species-level dynamics in paleoecological community settings.

Trond Reitan. Natural History Museum, University of Oslo, Norway.
  E-mail: trond.reitan@ibv.uio.no

Torbjørn H. Ergon. Centre for Ecological and Evolutionary Synthesis, Department of Biosciences,
  University of Oslo, Norway. E-mail: t.h.ergon@ibv.uio.no

Lee Hsiang Liow. Natural History Museum, University of Oslo, Norway; and Centre for Ecological and Evolutionary
  Synthesis, Department of Biosciences, University of Oslo, Norway. E-mail: trond.reitan@ibv.uio.no

Introduction

Understanding past and contemporary patterns and dynamics of populations and communities requires robust estimates of variation in abundance of organisms (Williams et al. 2002; Sutherland et al. 2013). While it is notoriously difficult to estimate absolute population sizes or densities due to the imperfect detection of individuals (Schwarz and Seber 1999), it is generally much easier to estimate relative differences/changes in population sizes/densities (Williams et al. 2002). Fortunately, such relative estimates are often sufficient for ecological inference. For example, community ecologists and paleoecologists have long been interested in measuring and explaining distributions of relative species abundance (RSA; i.e., the abundance of a species relative to the abundance of all species) in communities (Fisher et al. 1943; MacArthur and Wilson 1967; Kidwell 2001).

Likewise, it is often sufficient to model changes in population density (hereafter "relative population density" [RPD]) over time (Royama 1992; Caswell 2001) relative to the same population from a given reference point, for example, the start of population monitoring.

While contemporary ecological and fossil data reflect ecological and evolutionary processes at vastly different timescales, sampling strategies and data structure may be similar. Like contemporary ecological data, fossil data often consist of detection records (i.e., observations of the presence) of species. Fossil records are often associated with geological formations (time intervals) of different ages, where low or zero detection frequencies in certain formations may be due to low (then) extant population densities and/or low preservation probabilities. When detection and nondetection of focal species in replicated samples have been recorded, it is possible to estimate both

occupancy and the probability of detection given occupancy (MacKenzie et al. 2002). Over the past decades, a rich literature on such "site occupancy models" or "species occupancy models" has emerged in ecology (King 2014; MacKenzie et al. 2017). Parenthetically, note that these models are not equivalent to those involving occupancy in the older paleontological literature, as they do not explicitly model unobserved presences (Jernvall and Fortelius 2004; Raia et al. 2006; Foote et al. 2007). Site occupancy models were developed to estimate the probability of true species presence, for example, as a function of habitat variables. Later developments also linked detection probabilities to species abundance (Royle and Nichols 2003). Multispecies expansions of these models have facilitated studies of community composition and species richness (Dorazio et al. 2006; Yamaura et al. 2011; Iknayan et al. 2014; Devarajan et al. 2020).

Site occupancy models (hereafter "occupancy models") were first used to address paleoecological questions in a study of Ordovician brachiopods (Liow 2013). Briefly, a bare-bones site occupancy model aims at estimating the probability that any randomly selected site in an area where a given species lives (or lived) will be occupied by that species. However, individuals of that species are not always detectable, even if they live in the sites selected for survey. They could be camouflaged (e.g., frogs or stick-insects), stealthy (e.g., lynx), nocturnal (hard to detect in daytime surveys), or in the case of the fossil record, present in the site in question in the deep past, but with no preserved or recognizable remains to be sampled by the paleontologist in the same physical location. The general approach in occupancy modeling is to replicate sampling in each site to tease apart the probability of occupancy (for the area as a whole) and detection probability. Given the extra layers of the challenges of detecting individuals that lived in the deep past, as their past presence may be masked by taphonomic, geological, and sampling processes, occupancy models could benefit from specific tailoring to fossil data.

Here, we develop a multispecies occupancy model, tailored for fossil occupancy data, aimed at estimating temporal patterns (over millions of years) of RSA and RPD. As is typical for fossil data, preservation also influences detection probability, and the preservation can vary substantially among formations (Behrensmeyer et al. 2000). One way of tackling temporal variation in preservation is by incorporating random effects for formations. By incorporating data from multiple species, we aim to reduce the influence of preservation on abundance estimates by "filtering out" formation-specific random effects on detection probability common to all species. Importantly, random effects also allow us to estimate formation-specific RSA and RPD when data consist of multiple samples (sites) and subsamples (replicates). In addition to formation-specific random effects to capture dynamics (perceptible temporal changes) common to all species, we also use those that capture the dynamics of individual species. All of these random effects allow us to "borrow strength" across species (e.g., Zipkin et al. 2010).

Using simulated data, we ask whether our multispecies occupancy model provides more accurate estimates of relative abundance than detection ratios traditionally used in paleoecology as an estimate for relative abundance. Here, "relative abundance" (what we call detection ratios) is estimated as the number of individuals of the focal species observed divided by the total number of all fossil individuals of the focal group observed in the sample or formation. We then apply this model to a dataset of marine invertebrates (cheilostome bryozoans) that attach to hard substrates (shells) over nine time intervals (geological formations) spanning 2.3 Myr from a marine basin in New Zealand. We discuss the general utility of our model in paleoecological settings and suggest venues for further development.

Materials and Methods

Study System

The empirical example we use is a community of fossilized benthic organisms, namely encrusting cheilostome bryozoans found in the Wanganui Basin (Carter and Naish 1998; Proust et al. 2005; Pillans 2017). Cheilostome bryozoans are a common component of both fossil and living marine benthic environments,

and the majority are encrusters on hard substrates, with a minority being attached, erect species, and only a handful being free-living species. We examined subsamples (= shells, typical substrates for bryozoans) for encrusting cheilostomes in 119 sites within 9 geological formations rich in fossil marine deposits representing time intervals from 2.29 to 0.30 Myr (Fig. 1), in which the number of shells varied between 30 and 50 (Supplementary Table S1). We focused on transgressive system track shell beds that reflect similar depositional conditions (facies) to increase the comparability of the biological habitats represented by our fossil material. By assuming that cheilostome species' abundances in sampled sites are representative for the region at the time they were preserved, we can make regional estimates for each time interval. We tabulated the observed presence of any fossilized individuals of three focal cheilostome species, namely, *Antharcthoa tongima*, *Escharoides excavata*, and *Arachnopusia unicornis*

(Supplementary Fig. S1) on each shell. There is ample among-formation, within-formation, and among-species variation in the detection ratio, that is, the number of shells with focal species of encrusting bryozoans observed divided by the total number of shells examined (Supplementary Fig. S2). We also introduce a fourth "species," the superspecies, which represents all other encrusting bryozoan species in the community, excluding the three focal species (e.g., the open circles in Fig. 1 could illustrate the superspecies in our empirical example). In doing so, we can utilize observations from other species in the same community without collecting detailed species-level data in a species-rich system to improve parameter estimates (see "Building the Model"). In addition, including the superspecies ensures that estimated species abundances will be relative to all bryozoan species, rather than only the sum of the included focal species. The formations were chosen because they are known to
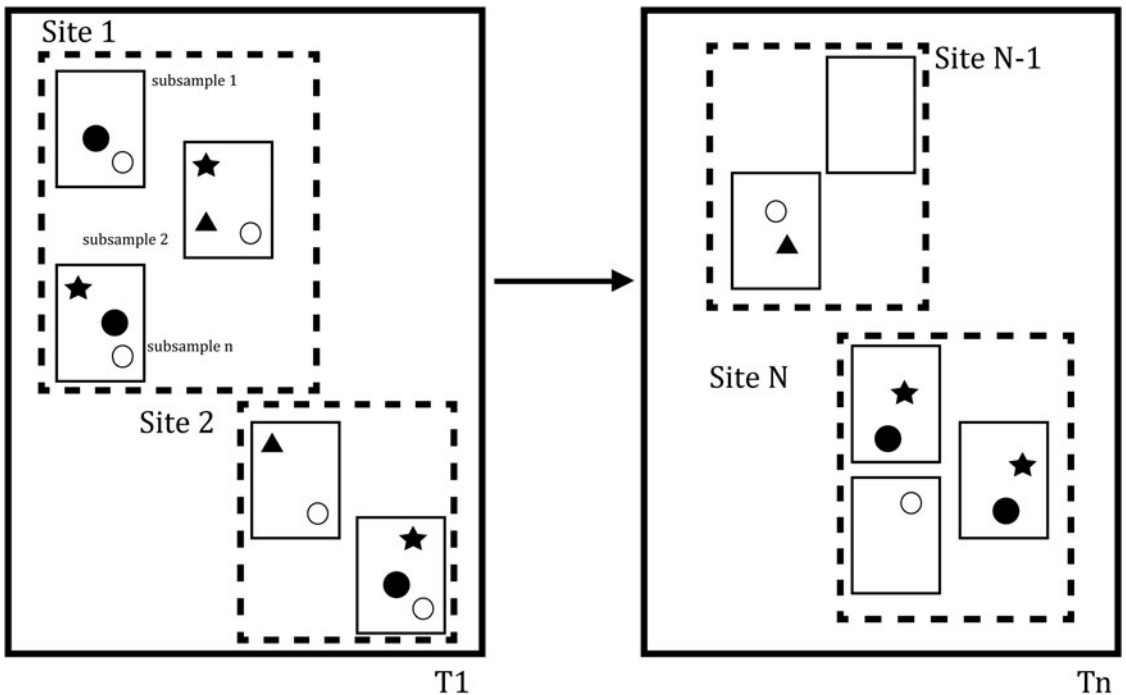

FIGURE 1. A schematic diagram to show the sampling scheme. Each thick-bordered open rectangle represents a time interval (only two are illustrated for the first time interval, *T1*, and the nth time interval, *Tn*). Within each time interval, sites (dotted rectangles, only two represented in each) are sampled. Note that any given site is rarely if ever in the same past physical locations, the common situation in paleontological data, hence the sites have unique labels. Within each site, there are subsamples (smaller, solid-bordered rectangles) in which different species (solid shapes) are observed. The arrow between the time intervals represents the flow of time.

harbor bryozoans, so our superspecies is assumed to always be present, that is, occupancy probability = 1. In other applications, a superspecies can be excluded or the occupancy probability of the superspecies can be estimated within the model.

Overview of Modeling Approach

Our main objective is to estimate the temporal (i.e., formation-to-formation) dynamics of RSA and RPD. We refer to both RSA and RPD as "relative abundance" for short until the "Estimating RSA and RPD" section. Our data are the detection/nondetection observations on subsamples (shells in our empirical example) from different sites (Fig.1, Supplementary Fig. S1). A species has the potential of being observed in a given subsample if it is truly present in a given site. On the contrary, if a given site does not have any evidence of a given species on any of its subsamples, it could mean that the site was truly devoid of that species or that the species was present but just not sampled in any of the subsamples belonging to that site. A basic single-species occupancy model teases apart observations within sites into true occupancy and detection

components by capitalizing on replicate samples within sites. In the modeling approach we describe here, we begin with a basic single-species occupancy model, then add random effects to account for site variation, and finally account for random variation among species and formations. Once this full model has been set up (Fig. 2), we use it to link the inferred detection and occupancy probabilities to relative abundance.

Building the Model

The Basic Occupancy Model for Number of Detections per Site.—A basic occupancy model (MacKenzie et al. 2002) presents the probability that a species occupied a given site, that is, the occupancy probability, Ψ, and the probability that a subsample has at least one observation of the focal species, given occupancy, that is, the detection probability, p. The probability that a species is found on a given subsample is thus Ψp, where Ψ operates on the site level while p operates on the subsample level. The occupancy and detection probabilities will be functions of various parameters and random effects and can be specific to the site i belonging to a specific time interval, f, and the species, s. Thus, we
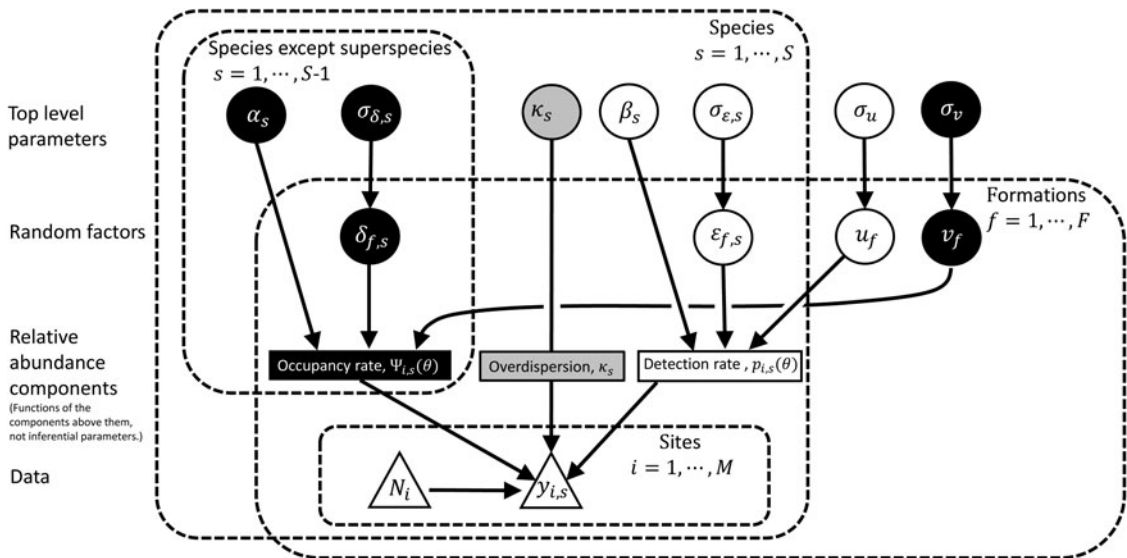

FIGURE 2. This figure summarizes our full hierarchical occupancy model for estimating relative abundance (relative species abundance [RSA] or relative population density [RPD]) composed of top-level parameters and random effects that describe their overdispersion. Data are denoted as triangles where M is the number of sites and y the shells from site i, where species s is observed. Black circles denote occupancy parameters, white circles denote detection parameters, and the gray circle denotes the overdispersion parameter. An arrow from an element A (i.e., circle, triangle, or rectangle) to another B, denotes that B is conditioned on A either by a function or a distribution (see text for details).

write $\Psi_{i,s}(\theta)$ and $p_{i,s}(\theta)$ for the occupancy and detection probabilities, respectively, where θ is the set of top-level parameters and random variables of the model in question (Fig. 2). The relative abundances for each focal species will be derived from these two sets of probabilities.

We need to take non-occupied sites into account when constructing our models. The consideration of non-occupied sites leads to the phenomenon of zero inflation, that is, an increased chance of zeros in our data. We define the zero-inflated binomial distribution as $\quad P_{zbin}(y|n, p, \Psi) = (1 - \Psi)I(y = 0) + \Psi \binom{n}{y}$ $p^y(1 - p)^{n-y}$, where $\binom{n}{y}p^y(1 - p)^{n-y}$ is the binomial probability of independently detecting y out of n with detection probability p. Focusing on the observations, $y_{i,s}$ is the number of subsamples at site i with observations of species s. For subsamples in occupied sites, some may have observations of species s, while others do not (independently of each other), hence, $y_{i,s}$ is a zero-inflated binomial random variable:

$$y_{i,s} \sim zbin\left(N_i, p_{i,s}(\theta) = \text{logit}^{-1}(\beta_s), \Psi_{i,s}(\theta)\right)$$
$$= I(s = S) + I(s < S)\text{logit}^{-1}(\alpha_s)), \quad (1)$$

where N_i is the total number of subsamples examined at site i. $I()$ stands for the indicator function, which takes value 1 when the statement inside is true and 0 if false. S is the total number of species under analysis (in our example, this includes our 3 focal species plus the superspecies). We let the last species be the superspecies. If there is no superspecies assigned in the system, then equation (1) will take the simpler form $y_{i,s} \sim zbin\left(N_i, p_{i,s}(\theta) = \text{logit}^{-1}(\beta_s), \Psi_{i,s}(\theta) = \text{logit}^{-1}(\alpha_s)\right)$.

The unconditional probability of detection is $p_{i,s}(\theta)\Psi_{i,s}(\theta)$, that is, detection, regardless of whether a site is actually occupied (Royle and Nichols 2003). We express both occupancy and detection probabilities using a logit-transform, that is, $\text{logit}(r) \equiv \log\left(\frac{r}{1-r}\right)$, where r is a probability, for the convenience of expanding the model (see next sections). The two parameters, α_s and β_s (Fig. 2), give regional (i.e., within the Wanganui Basin in our application)

occupancy and detection probabilities for each species, regardless of time interval (formation). The parameter set is hence $\theta = \{\alpha_1, \cdots, \alpha_{S-1}, \beta_1, \cdots, \beta_S\}$. The subscript i is included for clarity, although sites are not considered in this section. $\alpha_{s=S}$ does not appear, as we assume that the superspecies is always present.

Now that we have described a basic occupancy model, we continue the model description in a stepwise fashion by considering variation at site, species, and formation levels until the model has enough elements for relative abundance estimates. We do this for three reasons. The first is to put focus on each of the model components. Second, Markov chain Monte Carlo (MCMC) convergence was achieved only when we used the parameter estimates from a simpler model as the starting points for the next, more complex model. Third, we wanted to justify adding model complexity, using the Bayes factor as measure of evidence (Jeffreys 1998).

Including Site-Dependent Random Effects for Number of Detections per Site through Overdispersion.—Variation in abundance among sites is expected in natural systems. Because detection is linked to true abundance, the detection probability of a given species is expected to fluctuate from site to site. Fossil preservation can also influence detection probabilities on the site-level. Observations in our dataset consist of one summary data point per site (tabulated from the subsample replicates) per species. We use overdispersion for modeling sites instead of a per-observation random effect (Harrison 2014), as these two modeling choices are equivalent, but the former is computationally easier. This is because including random effects for each of the sites would radically and unnecessarily increase model complexity. Hence, to facilitate extensive simulations, we use the beta-binomial distribution, which has an analytical expression, namely:

$$P_{\beta bin}(y|n, p, \kappa)$$
$$= \frac{\Gamma(n+1)\Gamma\left(y+\frac{p}{\kappa}\right)\Gamma\left(n-y+\frac{1-p}{\kappa}\right)\Gamma\left(\frac{1}{\kappa}\right)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma\left(n+\frac{1}{\kappa}\right)\Gamma\left(\frac{p}{\kappa}\right)\Gamma\left(\frac{1-p}{\kappa}\right)},$$
$$(2)$$

where y out of n is the outcome, p the detection probability, and κ the overdispersion parameter where $\kappa = 0$ means no overdispersion (see Supplementary Material for details). This specifies the distribution of detections given occupancy. Thus, the zero-inflated (un-conditioned on occupancy) beta-binomial distribution is:

$$P_{z\beta bin}(y|n,p,\kappa,\Psi) = (1-\Psi)I(y=0)$$
$$+ \Psi \frac{\Gamma(n+1)\Gamma\left(y+\frac{p}{\kappa}\right)\Gamma\left(n-y+\frac{1-p}{\kappa}\right)\Gamma\left(\frac{1}{\kappa}\right)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma\left(n+\frac{1}{\kappa}\right)\Gamma\left(\frac{p}{\kappa}\right)\Gamma\left(\frac{1-p}{\kappa}\right)},$$

(3)

where Ψ is the zero inflation of the number of detections per site and the likelihood is:

$$y_{i,s} \sim z_{\beta bin}\left(N_i, p_{i,s}(\theta) = \text{logit}^{-1}(\beta_s), \kappa_s,\right.$$
$$\left.\Psi_{i,s}(\theta) = I(s=S) + I(s<S)\text{logit}^{-1}(\alpha_s)\right)$$

(4)

κ_s describes the species-dependent overdispersion, while the other terms are as in equation (1). The parameter set is now $\theta = \{\alpha_1, \cdots, \alpha_{S-1}, \beta_1, \cdots, \beta_S, \kappa_1, \cdots, \kappa_S\}$. The detection and occupancy probabilities depend on the identity of the time interval that the site belongs to, rather than the site itself, as the beta-binomial distribution accounts for overdispersion among sites. At this point, no time-interval dependency in occupancy or detection has been added.

Including Species- and Formation-Dependent Random Effects.—We now introduce temporal dynamics by using time interval–dependent random effects that are species independent, that is, they summarize dynamics common to all species in the community. For the detection probability, the random effects imply fluctuations in the preservation as well as in average abundance of all species in the community. For occupancy, the random effects allow fluctuations in the overall presence of the set of species in question. The time intervals with richer data, by contributing to estimates of overall probabilities and the standard deviations of the random effects, can thus inform

estimates for those with sparser data. The model is now:

$$y_{i,s} \sim z_{\beta bin}\left(N_i, p_{i,s}(\theta) = \text{logit}^{-1}(\beta_s + u_{f(i)}), \kappa_s,\right.$$
$$\Psi_{i,s}(\theta) = I(s=S) + I(s<S)\text{logit}^{-1}(\alpha_s + v_{f(i)})$$

(5a)

$$u_f \sim N(0, \sigma_u^2), \quad v_f \sim N(0, \sigma_v^2),$$

(5b)

where $f(i)$ is the time interval that site i belongs to, u_f and v_f are the new time interval–dependent random effects, and σ_v and σ_u are the standard deviations of these effects for detection and occupancy respectively. Now, $\theta = \{\alpha_1, \cdots, \alpha_{S-1}, \beta_1, \cdots, \beta_S, \kappa_1, \cdots, \kappa_S, \sigma_u, \sigma_v, u_1, \cdots, u_F, v_1, \cdots, v_F\}$, where F is the number of time intervals.

While equation (5) does allow for dynamics due to time variations in all species in the region, the species probabilities are in sync. To facilitate dynamics that permit fluctuations in the relative abundances for each species, we need random effects that depend on species and formation combinations. When doing so, we have:

$$y_{i,s} \sim z_{\beta bin}(N_i, p_{i,s}(\theta) = \text{logit}^{-1}$$
$$(\beta_s + u_{f(i)} + \varepsilon_{f(i),s}), \kappa_s,$$
$$\Psi_{i,s}(\theta) = I(s=S) + I(s<S)\text{logit}^{-1}$$
$$(\alpha_s + v_{f(i)} + \delta_{f(i),s}))$$

(6a)

$$u_f \sim N(0, \sigma_u^2), \quad v_f \sim N(0, \sigma_v^2), \quad \delta_{f,s}$$
$$\sim N(0, \sigma_{\delta,s}^2), \quad \varepsilon_{f,s} \sim N(0, \sigma_{\varepsilon,s}^2),$$

(6b)

where $\varepsilon_{f,s}$ and $\delta_{f,s}$ are the new time interval– and species-dependent random effects and $\sigma_{\varepsilon,s}$ and $\sigma_{\delta,s}$ are the standard deviations of these effects, for detection and occupancy, respectively. As $p_{i,s}$ and $\Psi_{i,s}$ only depend on the time period, f, we label them as $p_{f,s}$ and $\Psi_{f,s}$ in the following sections.

The parameter set is now $= \{\alpha_1, \cdots, \alpha_{S-1}, \beta_1, \cdots, \beta_S, \kappa_1, \cdots, \kappa_S, \sigma_u, \sigma_v, \sigma_{\delta,1}, \cdots, \sigma_{\delta,S-1}, \sigma_{\varepsilon,1}, \cdots, \sigma_{\varepsilon,S}, u_1, \cdots, u_F, v_1, \cdots, u_F, \delta_{1,1}, \cdots, \delta_{F,S-1}, \varepsilon_{1,1}, \cdots, \varepsilon_{F,S}\}$. We choose independent and wide priors for each parameter (see Supplementary Material section "Prior Distribution for the Full Model"). All positive-value parameters, including the standard deviations and

overdispersion parameters, are log transformed so that on re-parametrization, they fall on the real number line. The top-level parameters (these exclude random variables, because their distribution depends on other parameters) are $\{\alpha_1, \cdots, \alpha_{S-1}, \beta_1, \cdots, \beta_S, \kappa_1, \cdots, \kappa_S, \sigma_u, \sigma_v, \sigma_{\delta,1}, \cdots, \sigma_{\delta,S-1}, \sigma_{\varepsilon,1}, \cdots, \sigma_{\varepsilon,S}\}$; see Fig. 2 (top row). With 3 species and 1 superspecies, we now have 20 ($5 \times S$) top-level parameters and 81 (($2S + 1$) $\times F$) random variables (eq. 6b). We call equation (6) the "full model," because it has all the necessary components for estimating relative abundance dynamics (Fig. 2), which we detail in the "Estimating RSA and RPD" section.

A Stepwise Approach for Improving Estimation.—Because the full model is fairly complex and required hierarchically arranged random effects, we utilized MCMC sampling for inference (see Supplementary Material section "MCMC for Statistical Inference"). We used common estimated parameter values from a simpler model when starting a more complex model, while gradually increasing the model complexity in a stepwise fashion (i.e., from eq. 1 to 4, 5, then 6), as preliminary analyses often failed when starting from a random place in the parameter space. In doing so, we also tested whether each increasingly complex model explained the data better, using Bayes factors.

Incorporating Explanatory Variables.—We expanded equation (6) in two different ways, motivated by our aim to predict relative abundances in unmeasured time intervals with more precision than just using the time interval–independent median values derived from α_s and β_s.

First, we included temporal explanatory variables in our empirical example pertaining to paleoclimate. Specifically, as each species should thrive in a different optimal climate, with a given tolerance width, we impose a quadratic term for our explanatory variables (on detection probability, occupancy probability, or both). We use two related but different paleoclimate proxies, namely the global $\delta^{18}O$ data (data from Lisiecki and Raymo 2005) and the North Atlantic magnesium/calcium (Mg/Ca) ratios (data from Sosdian and Rosenthal 2009), both based on measurements from benthic foraminifera, as explanatory variables. These contain complex signals of

sea-temperature, ice-volume, and sea-level changes, all of which potentially affect both the population growth rates (through optimal temperatures and the availability of substrate species) and detection probabilities (through sea-level changes) of our focal species.

Second, we explored autocorrelated processes, that is, statistical dependencies in a given variable between one time interval and the next time interval, by using an Ornstein–Uhlenbeck process (Supplementary Material sections "Model Expansions That Include Explanatory Variables" and "Introducing Correlations in the Random Effects"). The idea here is that the state of the population will depend on a previous time interval.

Estimating RSA and RPD

In this section, we use the full multispecies occupancy model described earlier (see also Fig. 2) to estimate relative abundance. Detection entails observing a species that is present, so the more sites that are occupied, the higher the regional abundance. Within a site, it is also reasonable to assume that detection probability increases with higher true abundance. In typical fossil data, detection additionally requires preservation and successful sampling and taxonomic identification of fossilized organisms. Preservation and hence taxonomic identifiability are often strongly associated with the formation to which the sample belongs (Behrensmeyer et al. 2000). For the purpose of estimating RSA and RPD in the fossil record, we introduce corrected probabilities $\Psi^*_{f,s}$ and $p^*_{f,s}$. We let $p^*_{f,s}(\theta) \equiv \text{logit}^{-1}(\beta_s + \varepsilon_{f,s})$, where the purely time interval–dependent random factor, u_f, is subtracted from the detection probability estimate. This is done with the assumption that the $u_{f(i)}$ term is mainly affected by preservation rather than by common biological dynamics among species. That is, $p^*_{f,s}(\theta)$ is our reconstruction of how the detection probabilities would be if preservation was the same for all formations, and thus should depend only on the abundance of each species. On the other hand, for our empirical data, preservation is unlikely to affect the time interval–dependent random factor for occupancy, v_f, thus we assume $\Psi^*_{f,s} \equiv \Psi_{f,s}$. When detection probabilities are low, moderate correlations

between detection and occupancy probabilities in the joint posterior distribution could mean that preservation dynamics influenced inferred occupancy probabilities. However, we expect these indirect effects on such inferred occupancy probabilities to be small compared with common occupancy dynamics.

We link the average observable abundance per subsample given occupancy, $\lambda_{f,s}$, to the corrected detection probability. Assuming that individuals are distributed independently of each other, $p_{f,s}^*$ is then given by the Poisson distribution:

$$p_{f,s}^* = P(\text{number of preserved individuals} > 0)$$
$$= 1 - e^{-\lambda_{f,s}} \qquad (7)$$

We can therefore derive $\lambda_{f,s}$ from an estimate of $p_{f,s}^*$ (main analyses) or derive $p_{f,s}^*$ from $\lambda_{f,s}$ (simulations and Supplementary Material). In our case, breaking the Poisson distribution assumption due to overdispersion of number of colonies per subsample only imperceptibly affects the abundance estimates (see Supplementary Material). Note that Yamaura et al. (2011) assumed detection to be the result of sampling from a binomial distribution, and the Poisson distribution is a limit of the binomial distribution, and equation (7) is in fact equivalent to equation (1) in Yamaura et al. (2011), given a re-parametrization, see Supplementary Material.

While we subtract the random factors representing the common preservation dynamics in detection, the average preservation rate over time is unknown. Thus, there is a proportionality coefficient, $k_{f,s}$, between the average true and observable abundance per subsample given occupancy, such that

$$\lambda_{f,s} = k_{f,s}\Lambda_{f,s}, \qquad (8)$$

where $\Lambda_{f,s}$ is the average true abundance per subsample.

We first assign both species- and time-interval dependency on $k_{f,s}$ to make explicit the assumptions we later use. In an ideal world, our subtraction of the effects of preservation dynamics when constructing $p_{f,s}^*$ makes the proportionality coefficient both species- and time interval–independent, that is, $k_{f,s} = k$.

The average true abundance per subsample (unconditioned on occupancy) $A_{f,s}$ is

$$A_{f,s} = \Psi_{f,s}^* \Lambda_{f,s} = \Psi_{f,s}\lambda_{f,s}/k_{f,s} \qquad (9)$$

We can now define the RSA as the true abundance per subsample of a species normalized to the sum over all species of a given time interval ($R_{f,s}$). Under the assumption that preservation is the same for all species in question and that nothing else affects $k_{f,s}$, then the proportionality coefficients will be species independent, that is, $k_{f,s} = k_f$. k_f then drops out when calculating the RSA:

$$
\begin{aligned}
R_{f,s} &\equiv \frac{A_{f,s}}{\sum_{s'=1}^{S} A_{f,s'}} = \frac{\Psi_{f,s}\lambda_{f,s}/k_f}{\sum_{s'=1}^{S}\Psi_{f,s'}\lambda_{f,s'}/k_f} \\
&= \frac{\Psi_{f,s}\lambda_{f,s}}{\sum_{s'=1}^{S}\Psi_{f,s'}\lambda_{f,s'}} \\
&= \frac{\Psi_{f,s}\log(1-p_{f,s}^*)}{\sum_{s'=1}^{S}\Psi_{f,s'}\log(1-p_{f,s}^*)} \qquad (10)
\end{aligned}
$$

For an alternative modeling approach, built up from the average observable abundance per subsample given occupancy, $\lambda_{f,s}$, rather than detection probabilities, see Supplementary Material "Description of the Abundance-focused Model."

We define the RPD, $Q_{f,s}$, as the true abundance for the species at a given time interval relative to the true abundance of the same species averaged over all time intervals. We normalize $Q_{f,s}$ to the temporal mean rather than to a specific time interval (e.g., the first available), as it is less sensitive to uncertainty and estimates near zero. As long as the proportionality coefficients are independent of time interval, $k_{f,s} = k_s$, we can relate this to observed quantities, such that:

$$
\begin{aligned}
Q_{f,s} &\equiv \frac{A_{f,s}}{\frac{1}{F}\sum_{f'=1}^{F} A_{f',s}} = \frac{\frac{\Psi_{f,s}\lambda_{f,s}}{k_s}}{\frac{1}{F}\sum_{f'=1}^{F}\frac{\Psi_{f',s}\lambda_{f',s}}{k_s}} \\
&= \frac{\Psi_{f,s}\lambda_{f,s}}{\frac{1}{F}\sum_{f'=1}^{F}\Psi_{f',s}\lambda_{f',s}} \\
&= \frac{\Psi_{f,s}\log(1-p_{f,s}^*)}{\frac{1}{F}\sum_{f'=1}^{F}\Psi_{f',s}\log(1-p_{f,s}^*)} \qquad (11)
\end{aligned}
$$

$Q_{f,s}$ will vary around the value 1 and is comparable within species, but not among species (unlike $R_{f,s}$). Note that assumptions for RSA and RPD are different and that the choice of RSA or RPD will be context dependent. In cases for which it is desirable to consider RSA and RPD for the same system, we require that $k_{f,s} = k$ for both relative abundances to be valid simultaneously.

Simulations

We performed two sets of simulations. The "abundance-specified simulation study" demonstrates how well occupancy probabilities, abundance per subsample, and other variables (e.g., detection probabilities and relative abundances) can be estimated. The "occupancy dynamics–focused simulation study" presents the sampling regimes under which we might plausibly detect occupancy probability dynamics (i.e., non-overlapping 95% credibility intervals) when the parameters were as estimated in our empirical data.

The simulated data of the abundance-specified simulation study were generated by specifying the $\Psi_{f,s}'s$ and $\lambda_{f,s}'s$. Equation (7) was used for back-transforming into detection probabilities, and the data were then generated using equation (3). We let species 1 have dynamics in Ψ and species 2 have dynamics in λ.

For the occupancy dynamics–focused simulation study, we generated data under different sampling intensities (10, 20, 30, 50, 100, and 1000 sites per formation and 60, 100, 200, 400, and 1000 shells per site) and analyzed these data using the model and parameter estimates from our empirical data. See Supplementary Material for more details on both sets of simulations.

Results

Empirical Findings

We found that including both the time interval–dependent (eq. 5) and the time interval– and species-dependent random effects (eq. 6) improved the description of our empirical data (Supplementary Material Table S2). In other words, the full model (eq. 6, illustrated in Fig. 2) was preferred over simpler models, based on Bayes factors (see Supplementary Material section on "MCMC for Statistical Inference" for details), implying that the occupancy and detectability of the different bryozoan species varied significantly with time intervals (formation). However, including paleoclimate explanatory variables or autocorrelated random effects did not improve our model (Supplementary Material Table S2). In other words, for our current data, we are not able to predict relative abundance for unmeasured time intervals beyond the median. The Bayes factor did not resolve the choice between the alternative "abundance-focused model" and equation (6), and the models gave highly similar estimates of RSAs (see Supplementary Fig. S5).

The overdispersion parameters, $\{\kappa_s\}_{s \in \{1,\cdots,S\}}$, were estimated to 0.09, 0.05, 0.04, and 0.07 for *Antharcthoa tongima*, *Escharoides excavata*, and *Arachnopusia unicornis* and the superspecies, respectively (see Supplementary Table S3 for credibility bands), where $\kappa_s = 0$ means no overdispersion. While the estimates of the overdispersion parameters κ_s are small, they represent overdispersion that effectively doubles the variance of the number of detections per site, compared with no overdispersion (see Supplementary Fig. S6).

The standard deviation parameters of the random effects have large uncertainty (Supplementary Table S3), except for the formation-dependent but species-independent random effect (σ_u) used for detection probability. However, the model testing suggests that all random effects were necessary to obtain an acceptable model fit.

The uncertainty surrounding the occupancy probability of each of the focal species is quite large (Fig. 3); we cannot establish that occupancy is well below 1.0 for any combination of species and formation. We performed a simulation study to check how much data we would need for the uncertainty to be smaller than the estimated dynamics; see "Occupancy Dynamics-focused Simulation Study" in the Supplementary Material for details. Note that the relative changes in modeled detection probabilities (Fig. 3) are similar to the dynamics of detection ratios (Supplementary Fig. S2), implying that the latter, commonly used as an estimate for relative abundance, is not an appropriate estimate for its purpose.
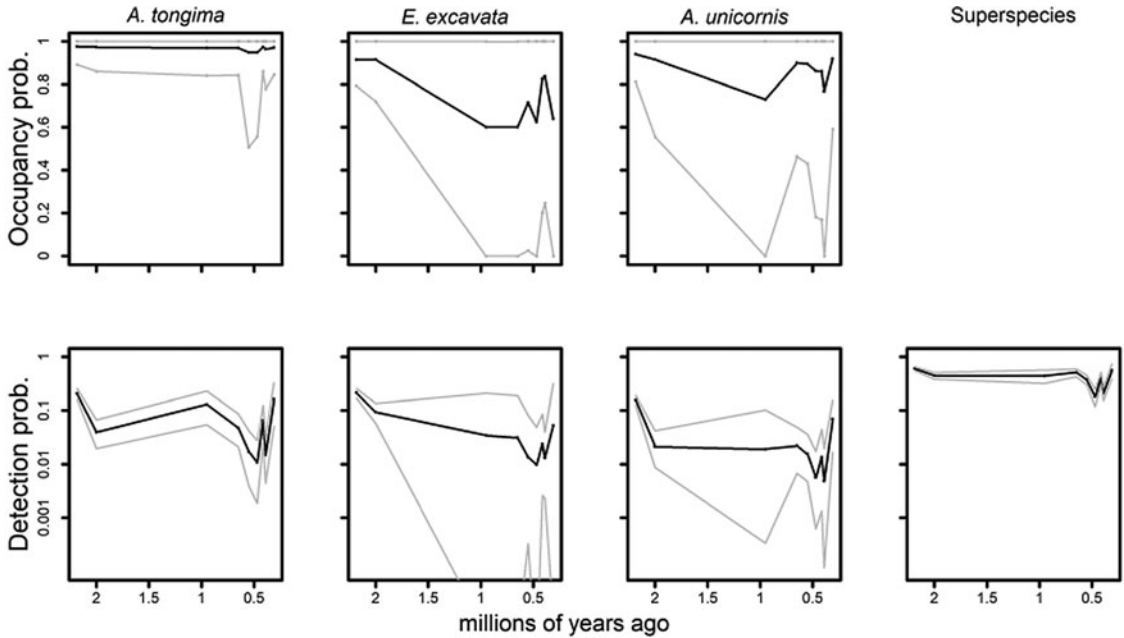
FIGURE 3.   Estimated occupancy and detection probabilities. Estimates are from our full model, where black lines join the species posterior median occupancy for each formation (time interval) plotted in the middle of the age range of the given formation. Gray lines show 95% posterior credibility intervals for the estimates. Note that the occupancy for superspecies is not plotted, as it is assumed to be 1 throughout, and that the y-axes for occupancy and detection are different.

The RSA ($R_{f,s}$) of the superspecies and *A. tongima* are estimated with relatively high precision and vary significantly over time, while those of *E. excavata* and *A. unicornis* are estimated with greater uncertainty (Fig. 4; see Supplementary Figs. S5 and S11 for alternative RSAs). The RPD ($Q_{f,s}$) estimates (Fig. 5) are also fairly uncertain, but some patterns are evident. Although the RSA of the superspecies fluctuates noticeably over time (Fig. 4), its RPD is remarkably constant (Fig. 5). This suggests that even though the abundance of any single species may fluctuate substantially over long timescales, the abundance of the bryozoan community is rather stable, at least during the time frame of this study (spanning ca. 2 Myr). Note that *A. tongima* and *E. excavata* are about equally abundant in the oldest formations (Fig. 4), but *E. excavata* becomes noticeably less abundant in the younger formations, at least relative to its own average abundance over time (Fig. 5). The abundance of *A. unicornis* is reduced from the first to the second time interval and then remains relatively low. For our system, we explicitly assume that $k_{f,s} = k$ (see "Materials and Methods").

Simulation Results

The abundance-specified simulation study shows a spread of the estimates around the true values for the input parameters (Supplementary Figs. S12–S15). As a consequence, there is also a spread in the quantities that, in our modeling, are derived, namely occupancy probabilities (Supplementary Fig. S16), detection probabilities (Supplementary Fig. S17), and RSAs (see Fig. 6 and Supplementary Fig. S18; the latter shows inferred dynamics for a few simulations). These estimates are distributed quite evenly around both sides of the actual values. Minute biases were expected (and found) given our informative priors and nonlinear transformations but were no cause for worry (see "Abundance-specified Simulation Study" in the Supplementary Material). RSA has traditionally been estimated as the number of detections of a species in a formation divided by the sum over species of the number of detections in the formation (e.g., Kidwell 2001; Peters 2004; Harnik 2011). Such detection ratio–based R estimates seem to deviate more from the actual values (Fig. 6), where averaged root mean-squared error (RMSE) for our model estimates equals 0.011 and 0.063 for detection
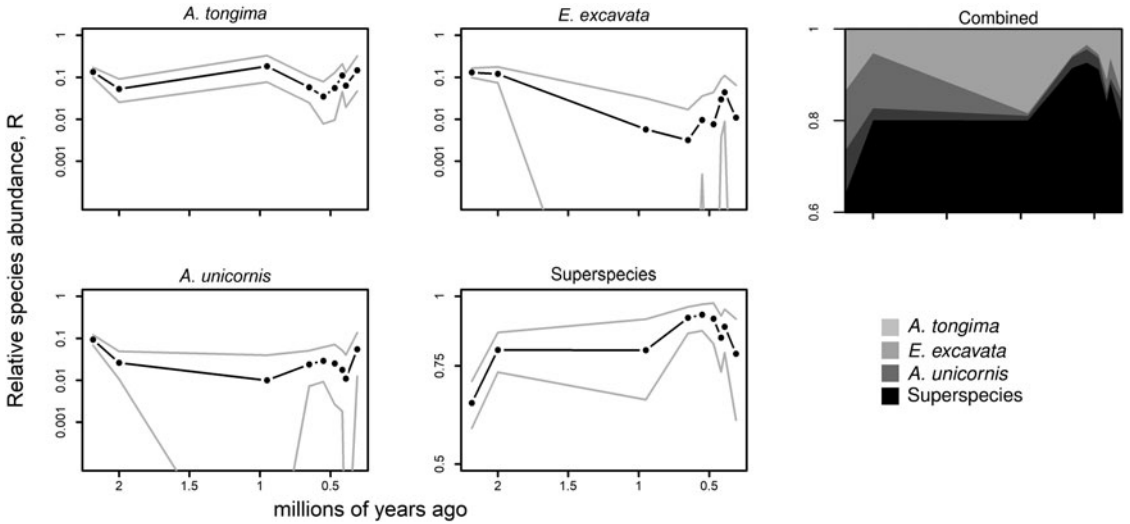
FIGURE 4. Estimated RSA. Estimates are from our full model, where black lines join the species mean RSA, (plotted on a log scale, except for the superspecies for visibility) for each formation (time interval). Gray lines show 95% posterior credibility intervals for the estimates and medians. A relative species abundance of 0.1 (for a given species in a time interval given) means that every 10th bryozoan in the region was of this species. The inset on the right ("Combined") shows the estimates combined for the four species/superspecies from their separate plots (note the different scale used for visual clarity).
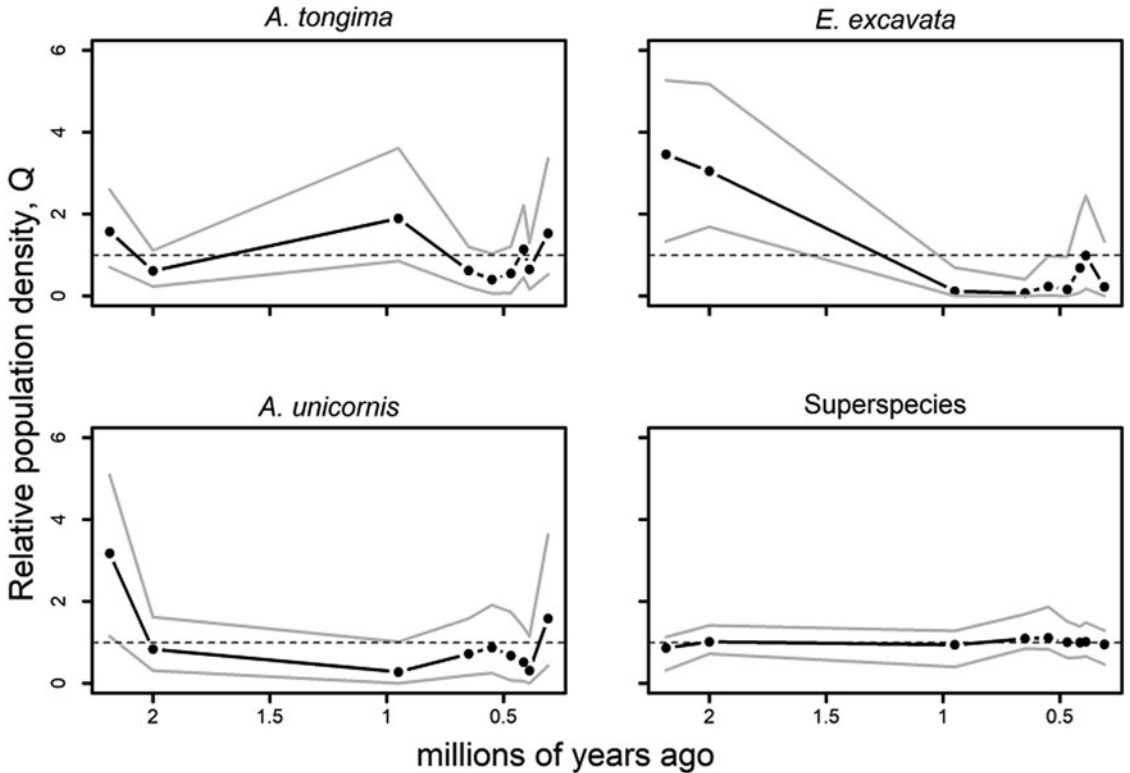

FIGURE 5. Estimated relative population density (RPD). Estimates are from our full model, where black lines join the species mean relative population density, Q, for each formation (time interval). Gray lines show 95% posterior credibility intervals for the estimates. Formation-specific values are divided by the mean across formations. Hence, a value of 0.1 means that the abundance is 10% of the mean across formations for the given species (horizontal stippled lines at value 1).
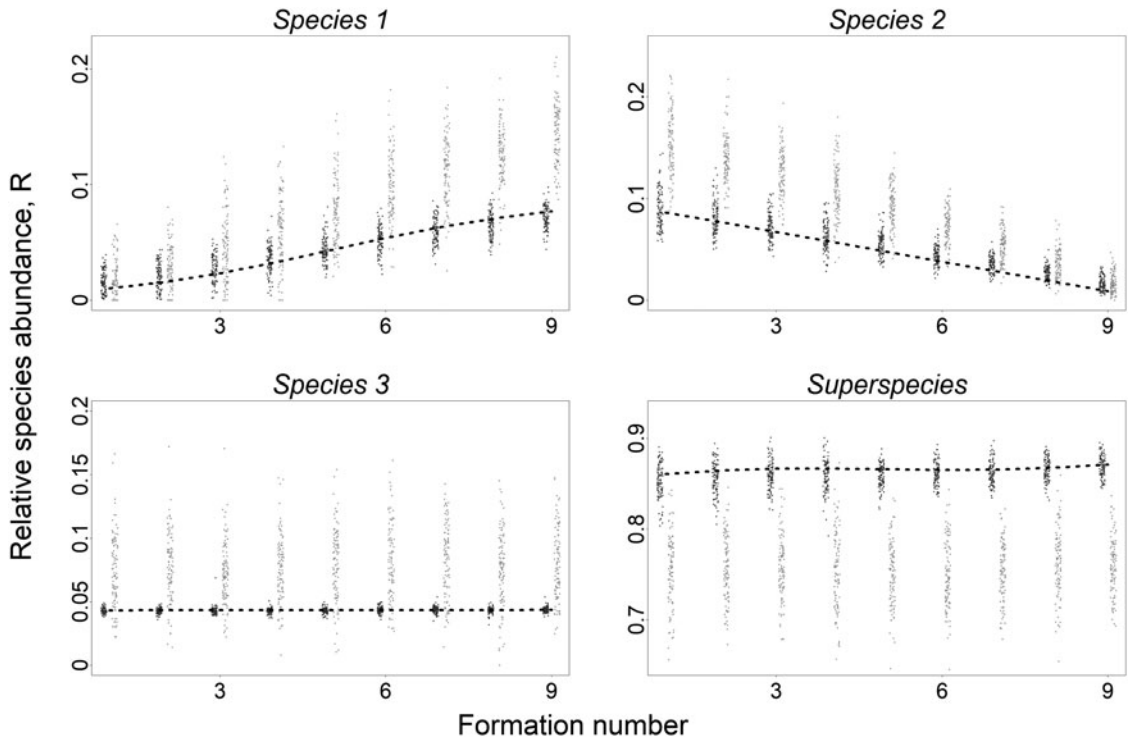
FIGURE 6.   RSA from the abundance-specified simulation study. Dashed black lines show the true relative species abundances for the various species and formations, dark gray dots show the inferred dynamics using the full model estimates, and light gray dots show detection ratio–based estimates. The full model estimates are offset slightly to the left of each formation, and the detection ratio–based estimates are offset slightly to the right.

ratio–based estimates. Thus our impression from the figure is reinforced by these numbers. When decomposing into bias and standard error, we found that biases and standard deviations were smaller for full model estimates than for detection ratio–based estimates for all species–time period combinations.

Discussion

Ecologists are interested in estimating changing RSA and RPD because these values provide a prime window into population dynamics (Sutherland et al. 2013). On a shorter timescale, understanding how environmental attributes and species traits affect population changes within communities are not only key to ecological understanding but also conservation management (Bowler et al. 2018). On a longer timescale, relative abundances of fossil taxa give us baselines for conservation (Barnosky et al. 2017), ecological backdrops

for evolution of phenotypes (Hannisdal 2006), and changing ecological interactions (e.g., Liow et al. 2019) so that we can better link paleoecological dynamics to evolutionary changes. However, estimating numbers of individuals in nature is challenging, regardless of the characteristics of an organism (e.g., sessile or motile, small-bodied or large-bodied), the type of data (e.g., direct counts, capture–recapture data), or the timescale involved (e.g., seasonal, yearly, or paleoecological data). Occupancy modeling, which explicitly models detection probabilities when estimating parameters of biological interest, including changes in relative abundance, is one powerful way of incorporating different sources of data heterogeneity and uncertainty. While occupancy modeling is increasingly widespread in "traditional" ecological studies (Bailey et al. 2014), it has yet to be applied regularly in paleoecology.

To briefly elaborate on the applicability of occupancy modeling in general, including our

modeling framework for relative abundance (RSA) or density (RPD) in paleoecological settings, we emphasize that fossil detection probability is far from perfect (i.e., 1), not least because preservation is far from guaranteed (Kidwell and Holland 2002). Traditionally in paleoecology, however, there is an underlying assumption, usually implicit, that preservation (and hence the detection of preserved organisms) is comparable across samples and sites, sometimes even across time intervals, as long as sampling is standardized. Here, detection ratios (see "Materials and Methods: Study System") are usually presented as estimates of RSA (Kidwell 2002; Currano et al. 2008; Espinosa et al. 2020). However, we know from simulations and ecological studies, and the observations made in this study (compare Fig. 3 and Supplementary Fig. S2), that this assumption is problematic (Iknayan et al. 2014; MacKenzie et al. 2017). Not only is it important to progress beyond tabulations of paleoecological data for improved inferences, but parameters estimated using fossil data should be as comparable as possible with those estimated using living organisms. This will allow us to infer historical baselines for conservation applications and to gain a better understanding of changing biota over longer timescales for which we may have analogous crisis situations (Harnik et al. 2012; Barnosky et al. 2017).

Occupancy modeling is generally applicable to the fossil record where replicates within sites can be surveyed to estimate regional parameters. In the multispecies modeling framework we presented for encrusting bryozoans, we used shells within sites (sections of outcrops) as site replicates. In other applications of occupancy modeling in the fossil record using either more basic occupancy models or our multispecies framework, we envisage the use of subsamples of outcrops (e.g., replicated slabs of outcrops; Liow 2013), bulk material, age-constrained lake or deep-sea core samples as site replicates. If these subsamples and the "sites" from which they are extracted have constant areas/volumes, the estimated occupancy, RSA, and RPD are straightforward to interpret without increasing complexity (unlike in our system, see below). Note that sites (and their subsamples) should be randomly selected and representative of the region that is the focal realm of the given study. In addition, to pick up temporal dynamics rather than habitat dynamics, we explicitly assumed that the samples (i.e., the habitats and environments they represent) are comparable across time. While selected sites will be heterogenous with respect to both preservation and true species occupancy, site-specific covariates can be used to improve parameter estimation (MacKenzie et al. 2017), or when there is no suitable covariate that can be used, using site-specific random effects as demonstrated is also a good solution to site heterogeneity. As already mentioned, the "superspecies" is not required if one's empirical system is not particularly species rich (unlike our empirical example). If all observed individuals are potentially assignable to a species in a group of interest, the estimated RSA and RPD will be relative to the group as a whole. As in all statistical modeling, the idiosyncrasies of the empirical system should be carefully considered, given the assumptions of the given model to help with both modeling and interpretation of the estimates. Below, we discuss some details of our own empirical system to illustrate such considerations and give examples of other empirical data that may be common in paleontology to highlight any differences.

Rather than using the observed detection or nondetection of species, we could have used the counts of the number of individuals of a given species in each subsample. If we used the latter, we would have built a model similar to an N-mixture model (Royle 2004). However, the subsamples in our example (shells or fragments thereof) varied in size, and these differences are expected to affect the number of individuals (colonies in our case). As shell size was not quantified, a random effect for subsamples would be needed to account for this variation. This inclusion would massively increase model complexity while introducing an uncertainty that would make the extra information (counts per subsample in our case) of little use. Because the computational cost would increase dramatically, while the outcome was not expected to improve significantly, we decided against this route for our empirical demonstration. However, in other applications, subsample size can be accounted for.

The accuracy of the RSA and RPD estimates depends on how close the assumptions concerning the proportionality coefficients are to reality. The estimates of RSA assume that the proportionality coefficients do not vary among species, and the estimates of RPD assume that the coefficients do not vary among formations for each species (i.e., both RSA and RPD are only accurate at the same time if the proportionality coefficients are constant across both species and formations, $k_{f,s} = k$). As already mentioned, we assume that both RSA and RPD are valid in our empirical system (as described in our results). However, it might not be necessary to estimate both RPD and RSA in another research context, nor might it be appropriate. For example, if tree pollen preserved in lake sediments is used to estimate relative abundance of tree species, and if some tree species are ecologically rare but highly prolific when producing pollen, RPD will be valid in our model, but RSA will not.

Our estimates of RSA apply only to the shell substrates that we have sampled; likewise the unit for our RPD is density per shell. Hence, if it is desirable to interpret the estimates given a different unit (e.g., per area sea bottom), one would have to make additional assumptions. Such assumptions depend on the application. For our data, the recruitment of encrusting bryozoans to substrates is thought to be largely limited by the availability of adults, although substrate orientation, the presence of biofilms, and substrate types (e.g., hard substrates vs. soft substrates like sea grass or kelp) may also influence larval attachment and subsequent growth (Taylor and Wilson 2003) and may have species specificity. We have purposefully limited our data collection to bivalve shells, the most abundantly available and preserved substrate, which is always represented in our Pleistocene system (Beu 2012). In addition, while bryozoans might be selective of habitats, for example, the strength of currents and coarseness of sediments in the habitat affects their filter-feeding abilities, (Wood et al. 2013), the same bryozoan species can be found on varied substrates, that is, different species of bivalves, rocks, gastropods, and echinoderms (Rust and Gordon 2011). This empirical

knowledge encouraged us to estimate RPD ($Q_{f,s}$) assuming that the availability of suitable substrate for any bryozoan species in our dataset is not limiting.

When estimating RPD, we removed the formation-specific random effects on detection probability belonging to all species. This has a strong impact on the RPD estimates, as the standard deviations of these random effects are estimated to be quite substantial. Our assumption (see the paragraphs preceding eq. 7) is that these random effects reflect variation in preservation in our study system, with similar bryozoan species encrusting similar shells (e.g., Liow et al. 2019). In other applications, however, the time-specific random effects may reflect true fluctuations in the community-level abundance, and hence should not be removed.

For our study, the biggest driver of relative abundance is the dynamics of detection and thus of average observable abundance per subsample given occupancy, while inferred occupancy probability and its estimated uncertainty are estimated to be quite high for all species and formations, thus revealing little dynamics (Fig. 3). This is because site observations are high for all three focal species, even though subsample detection probabilities are relatively low. The occupancy dynamics–focused simulations study (see "Occupancy Dynamics-focused Simulation Study" and Table S4 in the Supplementary Material for details) showed that reliably getting occupancy estimates that vary from formation to formation requires intense sampling protocols in our empirical system for our choice of species, with the possible exception of *Escharoides excavata*. Luckily, detecting occupancy dynamics was not the primary goal of our study. However, this potential issue of data insufficiency illustrates the importance of collecting pilot data and having some prior information about the empirical system in which occupancy modeling, with or without the added wish for relative abundance modeling, is to be performed, for instance, by running simulations such as those we suggested, before extensive data collection.

We note several extensions to our models that can be considered with regard to our empirical system. First, other sources of

system-specific variation might be taken into account. In our example, this includes the species of the shell substrate (e.g., some were cockles and others were scallops) and their body size, both of which may be selected by the bryozoan species involved and/or preferentially preserved. Second, we could potentially handle the number of colonies of each species observed for each subsample, because this could give an extra indication of the local average abundance per shell, although this is demanding data collection-wise as well as computationally for our dataset (as mentioned earlier). Third, in typical paleontological datasets, there are often time intervals in which we are not able to sample fossils because suitable material was not deposited. In our empirical example, we used two paleoenvironmental proxies (δ^{18}O and Mg/Ca ratios) as covariates in expanded models (Supplementary Material) in the hope that they contained predictive information we could use on unsampled time intervals. While neither of the two we had published data for were informative, it is possible that other paleoenvironmental proxies, given other paleoecological occupancy datasets, could be explored for this purpose. Alternatively, if possible, it is more appropriate to use paleoclimate proxies obtained from space and time samples matching the fossil occupancy data to more accurately describe the paleoenvironment in which the organisms lived, an option we did not have in our empirical example.

We hope that more paleoecologists will consider occupancy modeling as a means to estimate relevant ecological parameters and that modelers will pick up where we left off to improve the inference of biologically relevant parameters using a challenging but rich fossil record.

Data Availability Statement

The code and data for all analyses are provided at https://github.com/trondreitan/TRAMPOline. The code package is called TRAMPOline, based on an earlier acronym for the project, Temporal Relative Abundance-focused Multi-sPecies Occupancy model. Supplementary Material is available at https://github.com/trondreitan/TRAMPOline/blob/main/Reitan%20Ergon%20Liow%20occupancy_SI.docx.

Literature Cited

Bailey, L. L., S. M. Kidwell, and J. D. Nichols. 2014. Advances and applications of occupancy models. Methods in Ecology and Evolution 5:1269–1279.

Barnosky, A. D., E. A. Hadly, P. Gonzalez, J. Head, P. D. Polly, A. M. Lawing, J. T. Eronen, D. D. Ackerly, K. Alex, E. Biber, J. Blois, J. Brashares, G. Ceballos, E. Davis, G. P. Dietl, R. Dirzo, H. Doremus, M. Fortelius, H. W. Greene, J. Hellmann, T. Hickler, S. T. Jackson, M. Kemp, P. L. Koch, C. Kremen, E. L. Lindsey, C. Looy, C. R. Marshall, C. Mendenhall, A. Mulch, A. M. Mychajliw, C. Nowak, U. Ramakrishnan, J. Schnitzler, K. Das Shrestha, K. Solari, L. Stegner, M. A. Stegner, N. C. Stenseth, M. H. Wake, and Z. Zhang. 2017. Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. Science 355:eaah4787.

Behrensmeyer, A. K., S. M. Kidwell, and R. A. Gastaldo. 2000. Taphonomy and paleobiology. Paleobiology 26:103–147.

Beu, A. G. 2012. Marine Mollusca of the last 2 million years in New Zealand. Part 5. Summary. Journal of the Royal Society of New Zealand 42:1–47.

Bowler, D. E., H. Heldbjerg, A. D. Fox, R. B. O'Hara, and K. Böhning-Gaese. 2018. Disentangling the effects of multiple environmental drivers on population changes within communities. Journal of Animal Ecology 87:1034–1045.

Carter, R. M., and T. R. Naish. 1998. A review of Wanganui Basin, New Zealand: global reference section for shallow marine, Plio-Pleistocene (2.5–0 Ma) cyclostratigraphy. Sedimentary Geology 122:37–52.

Caswell, H. 2001. Matrix population models: construction, analysis and interpretation. Sinauer, Sunderland, Mass.

Currano, E. D., P. Wilf, S. L. Wing, C. C. Labandeira, E. C. Lovelock, and D. L. Royer. 2008. Sharply increased insect herbivory during the Paleocene–Eocene Thermal Maximum. Proceedings of the National Academy of Sciences USA 105:1960–1964.

Devarajan, K., T. L. Morelli, and S. Tenan. 2020. Multi-species occupancy models: review, roadmap, and recommendations. Ecography 43:1612–1624.

Dorazio, R. M., J. A. Royle, B. Soderstrom, and A. Glimskar. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87:842–854.

Espinosa, B. S., C. D'Apolito, S. A. F. Silva-Caminha, M. G. Ferreira, and M. L. Absy. 2020. Neogene paleoecology and biogeography

of a Malvoid pollen in northwestern South America. Review of Palaeobotany and Palynology 273:104131.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. Journal of Animal Ecology 12:42–58.

Foote, M., J. S. Crampton, A. G. Beu, B. A. Marshall, R. A. Cooper, P. A. Maxwell, and I. Matcham. 2007. Rise and fall of species occupancy in Cenozoic fossil molluscs. Science 318:1131–1134.

Hannisdal, B. 2006. Phenotypic evolution in the fossil record: numerical experiments. Journal of Geology 114:133–153.

Harnik, P. G. 2011. Direct and indirect effects of biological factors on extinction risk in fossil bivalves. Proceedings of the National Academy of Sciences USA 108:13594–13599.

Harnik, P. G., H. K. Lotze, S. C. Anderson, J. E. Byrnes, Z. V Finkel, S. Finnegan, D. M. Lindberg, L. H. Liow, R. Lockwood, C. R. McClain, J. L. McGuire, A. O'Dea, J. M. Pandolfi, C. Simpson, and D. P. Tittensor. 2012. Extinctions in ancient and modern seas. Trends in Ecology and Evolution 27:608–617.

Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ:2:e616.

Iknayan, K. J., M. W. Tingley, B. J. Furnas, and S. R. Beissinger. 2014. Detecting diversity: emerging methods to estimate species diversity. Trends in Ecology and Evolution 29:97–106.

Jeffreys, H. 1998. The theory of probability. Oxford University Press, Oxford.

Jernvall, J., and M. Fortelius. 2004. Maintenance of trophic structure in fossil mammal communities: Site occupancy and taxon resilience. American Naturalist 164:614–624.

Kidwell, S. M. 2001. Preservation of species abundance in marine death assemblages. Science 294:1091–1094.

Kidwell, S. M. 2002. Time-averaged molluscan death assemblages: palimpsests of richness, snapshots of abundance. Geology 30:803–806.

Kidwell, S. M., and S. M. Holland. 2002. The quality of the fossil record: implications for evolutionary analyses. Annual Review of Ecology and Systematics 33:561–588.

King, R. 2014. Statistical ecology. Annual Review of Statistics and Its Application 1:401–426.

Liow, L. H. 2013. Simultaneous estimation of occupancy and detection probabilities: an illustration using Cincinnatian brachiopods. Paleobiology 39:193–213.

Liow, L. H., T. Reitan, K. L. Voje, P. D. Taylor, and E. Di Martino. 2019. Size, weapons, and armor as predictors of competitive outcomes in fossil and contemporary marine communities. Ecological Monographs 89:e01354.

Lisiecki, L. E., and M. E. Raymo. 2005. A Pliocene–Pleistocene stack of 57 globally distributed benthic δ^{18}O records. Paleoceanography 20:PA1003.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, N.J.

MacKenzie, D., J. Nichols, A. Royle, K. Pollock, L. Bailey, and J. Hines. 2017. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Academic Press, San Diego, Calif.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

Peters, S. E. 2004. Relative abundance of Sepkoski's evolutionary faunas in Cambrian–Ordovician deep subtidal environments in North America. Paleobiology 30:543–560.

Pillans, B. 2017. Quaternary stratigraphy of Whanganui Basin—a globally significant archive. Pp. 141–170 in J. Shulmeister, ed. Landscape and Quaternary environmental change in New Zealand. Atlantis Press, Paris.

Proust, J. N., G. Lamarche, S. Nodder, and P. J. Kamp. 2005. Sedimentary architecture of a Plio-Pleistocene proto-back-arc basin: Wanganui Basin, New Zealand. Sedimentary Geology 181:107–145.

Raia, P., C. Meloro, A. Loy, and C. Barbera. 2006. Species occupancy and its course in the past: macroecological patterns in extinct communities. Evolutionary Ecology Research 8:181–194.

Royama, T. 1992. Analytical population dynamics. Chapman & Hall, London.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777–790.

Rust, S., and D. Gordon. 2011. Plio-Pleistocene bryozoan faunas of the Wanganui Basin, New Zealand: stratigraphic distribution and diversity. New Zealand Journal of Geology and Geophysics 54:151–165.

Schwarz, C. J., and G. A. F. Seber. 1999. Estimating animal abundance: review III. Statistical Science 14:427–456.

Sosdian, S., and Y. Rosenthal. 2009. Deep-sea temperature and ice volume changes across the Pliocene–Pleistocene climate transitions. Science 325:306LP–310.

Sutherland, W. J., R. P. Freckleton, H. C. J. Godfray, S. R. Beissinger, T. Benton, D. D. Cameron, Y. Carmel, D. A. Coomes, T. Coulson, M. C. Emmerson, R. S. Hails, G. C. Hays, D. J. Hodgson, M. J. Hutchings, D. Johnson, J. P. G. Jones, M. J. Keeling, H. Kokko, W. E. Kunin, X. Lambin, O. T. Lewis, Y. Malhi, N. Mieszkowska, E. J. Milner-Gulland, K. Norris, A. B. Phillimore, D. W. Purves, J. M. Reid, D. C. Reuman, K. Thompson, J. M. J. Travis, L. A. Turnbull, D. A. Wardle, and T. Wiegand. 2013. Identification of 100 fundamental ecological questions. Journal of Ecology 101:58–67.

Taylor, P. D., and M. A. Wilson. 2003. Palaeoecology and evolution of marine hard substrate communities. Earth-Science Reviews 62:1–103.

Williams, B. K., J. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego.

Wood, A. C. L., A. A. Rowden, T. J. Compton, D. P. Gordon, and P. K. Probert. 2013. Habitat-forming bryozoans in New Zealand: their known and predicted distribution in relation to broad-scale environmental variables and fishing effort. PLoS ONE 8:e75160.

Yamaura, Y., J. Andrew Royle, K. Kuboi, T. Tada, S. Ikeno, and S. Makino. 2011. Modelling community dynamics based on species-level abundance models from detection/nondetection data. Journal of Applied Ecology 48:67–75.

Zipkin, E. F., J. Andrew Royle, D. K. Dawson, and S. Bates. 2010. Multi-species occurrence models to evaluate the effects of conservation and management actions. Biological Conservation 143:479–484.