

# 1 Materials and Methods

## 1.1 Taxon occurrence information

The dataset analyzed here was sourced from the Paleobiology Database (<http://www.paleodb.org>) which was then filtered based on taxonomic, temporal, stratigraphic, and other occurrence information that was necessary for this analysis. These filtering criteria are very similar to those from [? ].

Fossil occurrences were analyzed at the genus level which is common for paleobiological, macroevolutionary, or macroecological studies of marine invertebrates [? ? ? ? ? ? ? ? ? ]. Although species diversity dynamics tend to be of much greater interest than those of higher taxa, the nature of the fossil record makes accurate and precise taxonomic assignments at the species level for all occurrences extremely difficult if not impossible. Additionally, there is evidence of real differences in biological patterns at the genus level versus the species level [? ]. As such, the choice to analyze genera as opposed to species was in order to assure a minimum level of confidence and accuracy in the data analyzed here.

Geographic regions were defined by dividing the globe into four latitudinal bands at the equator and both the tropics of Cancer and Capricorn ( $\pm 23.5^\circ$ ). These boundary lines were chosen because they are defined independent of taxonomic occurrence and are constant throughout they are defined independent of taxonomic occurrence and are spatially constant throughout time. Tectonic plates, for example, while geologically constant are not spatially constant. Given that latitudinal diversity gradients are one of the focuses of this study, using spatially variable regions is inappropriate given that a given plate may transition from tropical to temperate or vice-versa.

## 1.2 Model specification

Taxon presence was modeled has a hierarchical hidden Markov model (HMM) where the three “process parameters” of gain/newly entering a province ( $\gamma$ ), persistence/survival ( $\phi$ ), and observation ( $p$ ). For each province, each of these process parameters were modeled hierarchically so that estimates were allowed to vary over time but in cases of little information those estimates were drawn

to the overall mean for that province. The estimates for each province were also estimated hierarchically in relation to each other; this way all estimates were relative to each other. The hierarchical structure of this model helps control for both overfitting and multiple comparisons during posterior analysis [? ? ].

Note that the following model is strongly inspired by the dynamic occupancy model presented in [? ].

$y_{i,j,t}$  is the observed occurrence of taxon  $i$  in province  $j$  at time  $t$ , where  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, J$ , and  $t = 1, 2, \dots, T$ .  $y = 1$  is occupied while  $y = 0$  is unoccupied.  $z_{i,j,t}$  is the “true” occurrence of taxon  $i$  in province  $j$  at time  $t$ , given the estimate of sampling. Just as with  $y$ ,  $z = 1$  is occupied while  $z = 0$  is unoccupied.

$\phi_{j,t}$  is the probability of surviving, in province  $j$ , from time  $t$  to time  $t + 1$  ( $Pr(z_{t+1} = 1 | z_t = 1)$ ).  $\gamma_{j,t}$  is the probability of newly entering province  $j$  at time  $t + 1$  ( $Pr(z_{t+1} = 1 | z_t = 0)$ ).  $p_{j,t}$  is the probability of observing a true occurrence ( $Pr(y = 1 | z = 1)$ ) in province  $j$  at time  $t$ .

$\psi$  is probability of sit occupancy/probability of occurrence ( $Pr(z_{i,t=1} = 1)$ ). The first time point is defined in terms of  $\psi$  because there is (assumed) no previous time points.

The parameters  $\phi$ ,  $\gamma$ , and  $p$  are then all defined hierarchically within time bins as samples from the shared mean between provinces.  $\Phi_t$ ,  $\Gamma_t$ , and  $P_t$  are the mean probabilities for a given point in time  $t$ .  $M_\phi$ ,  $M_\gamma$ , and  $M_p$  are the overall mean estimates of survival, origination, and preservation probabilities.

And finally, I use independent uniform priors for  $\psi_j$  by province  $j$ :  $\psi_j \sim U(0, 1)$ .

In total, the model can be summarized by the following statements:

$$\begin{aligned}
y_{i,t,j} &\sim \text{Bern}(p_{t,j}z_{i,t,j}) \\
z_{i,t=1,j} &\sim \text{Bern}(\psi_j) \\
z_{i,t,j} &\sim \text{Bern}(\phi_{j,t-1}z_{i,t-1,j} + \gamma_{j,t-1}(1 - z_{i,t-1,j})) \\
\text{logit}(\phi_{j,t}) &\sim \text{N}(\Phi_t, \sigma_{\phi,t}) \\
\text{logit}(\gamma_{j,t}) &\sim \text{N}(\Gamma_t, \sigma_{\gamma,t}) \\
\text{logit}(p_{j,t}) &\sim \text{N}(P_t, \sigma_{p,j}) \\
\Phi_i &\sim \text{N}(M_{\phi}, \sigma_{\Phi}) \\
\Gamma_i &\sim \text{N}(M_{\gamma}, \sigma_{\Gamma}) \\
P_i &\sim \text{N}(M_p, \sigma_P) \\
\sigma_{\phi,i} &\sim \text{C}^+(1) \\
M_{\phi} &\sim \text{N}(0, 1) \\
\sigma_{\Phi} &\sim \text{C}^+(1) \\
\sigma_{\gamma,j} &\sim \text{C}^+(1) \\
M_{\gamma} &\sim \text{N}(0, 1) \\
\sigma_{\Gamma} &\sim \text{C}^+(1) \\
\sigma_{p,j} &\sim \text{C}^+(1) \\
M_p &\sim \text{N}(0, 1) \\
\sigma_P &\sim \text{C}^+(1)
\end{aligned} \tag{1}$$

The joint posterior distribution of the HMM model was approximated using a Gibbs sampling MCMC routine as implemented in the JAGS probabilistic programming language CITATION. Four chains were run for 20000 steps, thinned to every 10th sample, and split evenly between warm-up and sampling phases. Chain sampling convergence was assessed using the  $\hat{R}$  statistic with values close to 1 (less than 1.1) indicating approximate convergence [? ].

Given the estimate of the joint posterior distribution, some downstream metapopulation summary statistics can be calculated. The first of these is turnover  $\tau$  defined as the probability that the occurrence of a taxon at time  $t$  is new ( $Pr(z_{t-1} = 0 | z_t = 1)$ ) [? ]. First, the occupancy probability  $\psi$  at times  $t = 2, \dots, T$  can be calculated recursively as

$$\psi_t = \psi_{t-1}\phi_{t-1} + (1 - \psi_{t-1})\gamma_{t-1}. \tag{2}$$

Turnover can then be calculated as

$$\tau_t = \frac{\gamma_{t-1}(1 - \gamma_{t-1})}{\gamma_{t-1}(1 - \psi_{t-1}) + \phi_{t-1}\psi_{t-1}} \quad (3)$$

An additional summary is the growth rate  $\lambda$  [? ? ] which is calculated as

$$\lambda_t = \frac{\psi_{t+1}}{\psi_t}. \quad (4)$$