

# PLACEHOLDER TITLE

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

November 9, 2015

## **Abstract**

## **1 Introduction**

The spatial context of how taxonomic diversity changes over time

Key to this study is the idea that any given taxon might leave a region and then re-enter it at a later time. This fact means that standard approaches which assume constant/range-through presence following initial occurrence until final occurrence (Alroy, 2010b, Foote, 2000a,b, 2003, Liow et al., 2008, 2015, Silvestro et al., 2014, Smits, 2015) are inappropriate as they do not capture the macroevolutionary process of interest. Additionally, given the imperfect record that is the fossil record, distinguishing between true absence and unobserved presence becomes an extremely goal when trying to understand the macroevolutionary processes of interest here.

Here I choose to model regional diversity as a hidden Markov Model (HMM), which is how occupancy is modeled in ecology (Royle and Dorazio, 2008). In a

HMM, the observed state of an individual is considered imperfectly observed, such that actually observing the “true” or latent state is done with some sampling probability. The actual diversification process affects the latent state of a taxon, which is itself estimated. A HMM is similar to a Jolly-Seber capture-mark-recapture model (Liow et al., 2015, Royle and Dorazio, 2008) except the transition from 1 to 0 is not an absorbing state and taxa can “return to life.”

## 2 Materials and Methods

### 2.1 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 Foote and Miller (2013).

Fossil occurrences were analyzed at the genus level which is common for paleobiological, macroevolutionary, or macroecological studies of marine invertebrates (Alroy, 2010a, Foote and Miller, 2013, Harnik et al., 2013, Kiessling and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015, Payne and Finnegan, 2007, Simpson and Harnik, 2009, Villhena et al., 2013). 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 (Jablonski, 1987). 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.

## 2.2 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 (Gelman and Hill, 2007, Gelman et al., 2013).

Note that the following model is strongly inspired by the dynamic occupancy model presented in (Royle and Dorazio, 2008).

$y_{i,j,t}$  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:

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

### 2.3 Posterior inference

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 each run for 100000 steps, thinned to every 100th 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 (Gelman et al., 2013).

Given the estimate of the joint posterior distribution, some downstream metapopulation summary statistics can be calculated. Given the estimates of  $z$  it is trivial to calculate the number of taxa that newly enter or exit any region  $j$  at any time  $t$ . Additionally, the net change in taxonomic diversity (entrances minus exits) can be calculated for any region  $j$  at any time  $t$ .

Turnover  $\tau$  defined as the probability that the occurrence of a taxon at time  $t$  is new ( $Pr(z_{t-1} = 0 | z_t = 1)$ ) (Royle and Dorazio, 2008). First, the occupancy probability  $\psi$  at times  $t = 2, \dots, T$  can be calculated recursively as

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

Turnover can then be calculated as

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

### 3 Results

### 4 Discussion

### Acknowledgements

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for helpful discussion during the conception of this study. Additionally, thank you A.

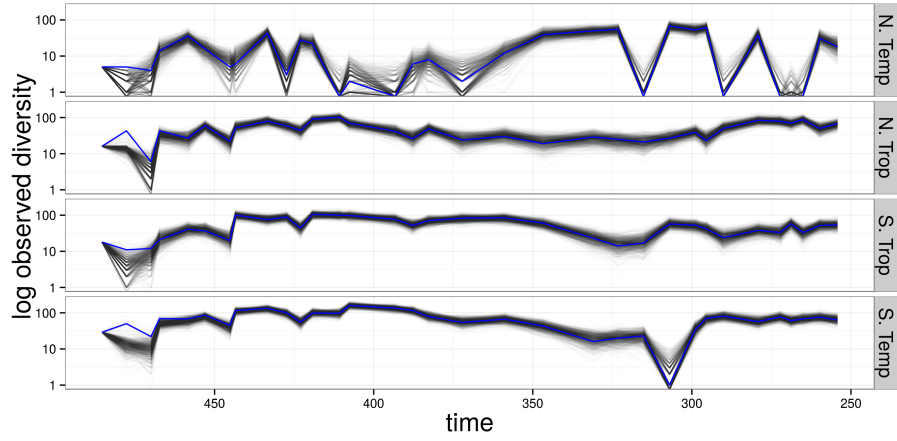


Figure 1: CAPTION

Miller for the epicontinental versus open-ocean assignments. This entire study would not have been possible without the Herculean effort of the many contributors to the Paleobiology Database. In particular, I would like to thank J. Alroy, M. Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland, L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P. Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P. Wager. This work was supported by a NASA Exobiology grant (NNX10AQ446) to A. Miller and M. Foote. I declare no conflicts of interest. This is Paleobiology Database publication XXX.

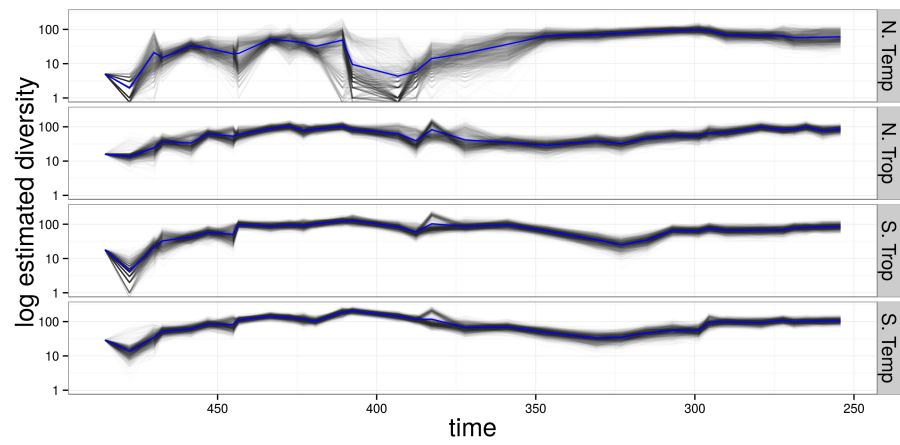


Figure 2: CAPTION

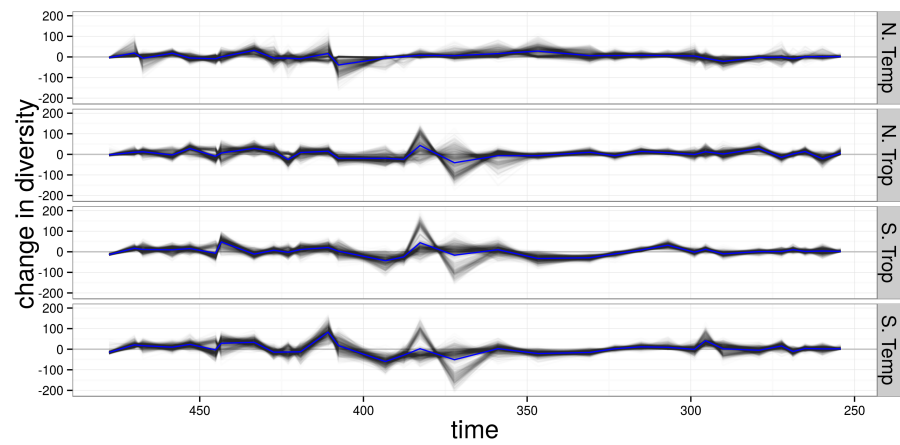


Figure 3: CAPTION



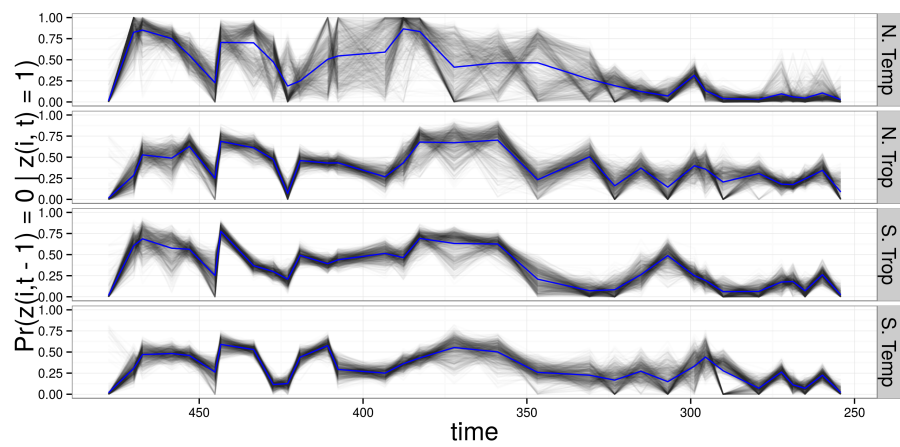


Figure 4: CAPTION

## References

- J. Alroy. The Shifting Balance of Diversity Among Major Marine Animal Groups. *Science*, 329(5996):1191–1194, 2010a.
- J. Alroy. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. In J. Alroy and G. Hunt, editors, *Quantitative Methods in Paleobiology*, pages 55–80. The Paleontological Society, 2010b.
- M. Foote. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology*, 26(sp4):74–102, Dec. 2000a. ISSN 0094-8373. doi: 10.1666/0094-8373(2000)26[74:OAECOT]2.0.CO;2. URL <http://www.bioone.org/doi/abs/10.1666/0094-8373%282000%2926%5B74%3A0AECOT%5D2.0.CO%3B2>.
- M. Foote. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology*, 26(4):578–605, Dec. 2000b. ISSN 0094-8373. doi: 10.1666/0094-8373(2000)026;0578:OAECOT;2.0.CO;2. URL <http://www.bioone.org/doi/abs/10.1666/0094-8373%282000%29026%3C0578%3A0AECOT%3E2.0.CO%3B2>.
- M. Foote. Origination and extinction through the Phanerozoic: a new approach. *Journal of Geology*, 111:125–148, 2003.
- M. Foote and A. I. Miller. Determinants of early survival in marine animal genera. *Paleobiology*, 39(2):171–192, Mar. 2013. ISSN 0094-8373. doi: 10.1666/12028. URL <http://www.bioone.org/doi/abs/10.1666/12028>.
- A. Gelman and J. Hill. *Data Analysis using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, New York, NY, 2007.
- A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall, Boca Raton, FL, 3 edition, 2013.
- P. G. Harnik, C. Simpson, and J. L. Payne. Long-term differ-

- ences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences*, 282(1801), Oct. 2013. ISSN 0962-8452. doi: 10.1098/rspb.2012.1902. URL <http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2012.1902>.
- D. Jablonski. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. *Science*, 238(4825):360–363, Oct. 1987. ISSN 0036-8075. doi: 10.1126/science.238.4825.360. URL <http://www.ncbi.nlm.nih.gov/pubmed/17837117>.
- W. Kiessling and M. Aberhan. Environmental determinants of marine benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology*, 33(3): 414–434, 2007.
- L. H. Liow, M. Fortelius, E. Bingham, K. Lintulaakso, H. Mannila, L. Flynn, and N. C. Stenseth. Higher origination and extinction rates in larger mammals. *Proceedings of the National Academy of Sciences*, 105(16):6097, 2008. URL [papers2://publication/uuid/5A93DDA3-204F-4D9C-AFAC-A15722A18C61](http://papers2://publication/uuid/5A93DDA3-204F-4D9C-AFAC-A15722A18C61).
- L. H. Liow, T. Reitan, and P. G. Harnik. Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. *Ecology letters*, pages 1030–1039, Aug. 2015. ISSN 1461-0248. doi: 10.1111/ele.12485. URL <http://www.ncbi.nlm.nih.gov/pubmed/26293753>.
- A. I. Miller and M. Foote. Epicontinental seas versus open-ocean settings: the kinetics of mass extinction and origination. *Science*, 326(5956):1106–9, Nov. 2009. ISSN 1095-9203. doi: 10.1126/science.1180061. URL <http://www.ncbi.nlm.nih.gov/pubmed/19965428>.
- S. Nürnberg and M. Aberhan. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*, 39

- (3):360–372, Apr. 2013. ISSN 0094-8373. doi: 10.1666/12047. URL <http://www.bioone.org/doi/abs/10.1666/12047>.
- S. Nürnberg and M. Aberhan. Interdependence of specialization and biodiversity in Phanerozoic marine invertebrates. *Nature communications*, 6:6602, Jan. 2015. ISSN 2041-1723. doi: 10.1038/ncomms7602. URL <http://www.ncbi.nlm.nih.gov/pubmed/25779979>.
- J. L. Payne and S. Finnegan. The effect of geographic range on extinction risk during background and mass extinction. *Proceedings of the National Academy of Sciences*, 104:10506–11, June 2007. ISSN 0027-8424. doi: 10.1073/pnas.0701257104. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1890565&tool=pmcentrez&rendertype=abstract>.
- J. A. Royle and R. M. Dorazio. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Elsevier, London, 2008.
- D. Silvestro, J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Systematic biology*, 63(3):349–67, May 2014. ISSN 1076-836X. doi: 10.1093/sysbio/syu006. URL <http://www.ncbi.nlm.nih.gov/pubmed/24510972>.
- C. Simpson and P. G. Harnik. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology*, 35(4):631–647, Dec. 2009. ISSN 0094-8373. doi: 10.1666/0094-8373-35.4.631. URL <http://www.bioone.org/doi/abs/10.1666/0094-8373-35.4.631>.
- P. D. Smits. Expected time-invariant effects of biological traits on mammal species duration. *Proceedings of the National Academy of Sciences*, 2015.
- D. A. Vilhena, E. B. Harris, C. T. Bergstrom, M. E. Maliska, P. D. Ward,

C. A. Sidor, C. A. E. Strömberg, and G. P. Wilson. Bivalve network reveals latitudinal selectivity gradient at the end-Cretaceous mass extinction. *Scientific Reports*, 3:1790, May 2013. ISSN 2045-2322. doi: 10.1038/srep01790.