Taxon occurrence as a function of both biological traits and environmental context

the changing North American species pool

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

The Paleobiology Database
revealing the history of life

Question

When are certain ecologies/ecotypes enriched or depleted?

Cenozoic mammals of North America

Differences in extinction risk

(Smits, 2015, *PNAS*)

Species pool concept

Figure 1. Two models of community assembly. **(A)** Local communities comprise a subset of species from the regional species pool that have passed through environmental filters. There is no feedback from the metacommunity (collection of local communities) to the regional species pool. Adapted from [5]. **(B)** Local communities are assembled as in (A), but speciation adds new species to the pool, extinction removes others, and dispersal allows the persistence of species that might otherwise go extinct.

(Mittelbach and Schemske, 2015, *TREE*)

Eco-cube and ecotypes

TEXT-FIG. 1. Ecospace as defined by the three axes of tiering, motility level and feeding strategy. A, the ecospace cube with categories on each axis labelled. B, the ecospace cube 'exploded', showing 216 'bins' or modes of life specified by the combination of the categories on each ecospace axis.

(Bambach *et al.*, *2007*, Palaeontology)

The fourth-corner problem

Fig. 1. Graphical representation of the fourth-corner problem and its solution. (a) The problem as posed by Legendre, Galzin & Harmelin-Vivien (1997), where the goal is to combine abundance (**L**), trait (**Q**) and environment (**R**) data in some way, to determine a matrix describing the trait–environment relationship (**D**). (b) The proposed model-based solution to the fourth-corner problem, where the goal is to predict abundance (**Y**) as a function of predictor variables for environment ($\mathbf{X}_1$), species traits ($\mathbf{X}_2$) and their interaction ($\mathbf{X}_1$*$\mathbf{X}_2$). The matrix of coefficients for the interaction between $\mathbf{X}_1$ and $\mathbf{X}_2$ is the fourth corner.

(Brown *et al.*, *2014*, Methods Ecol. Evol.)

Paleo-fourth corner model

Covariates of interest

individual-level
(species i at time unit t)

- log-odds of occurrence probability at time t
- effect of locomotor type
  - arboreal, digitigrade, plantigrade, unguligrade, fossorial, scansorial
- effect of dietary type
  - carnivore, herbivore, insectivore, omnivore
- effect body size (rescaled log body mass)

group-level (2 My time unit t)

- overall mean of log-odds of occurrence probability
- temperature record based on Mg/Ca estimates
  - mean and interquartile range of rescaled value
- plant community phase following Graham

Model of taxon occurrence

- response is p/a of genus in NA at time t
  - Bernoulli variable
  - probability is (observation prob) times ("true" presence)
- observation probability is effect of sampling/fossil record
  - basic model does not model sampling
- the latent discrete "true" presence modeled as a multi-level logistic regression
  - individual- and group-level

Paleo-fourth corner model

Model and sampling statement definition

$$
\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(\rho_t z_{i,t}) \\
\text{logit}(\rho_t) &\sim \mathcal{N}(\rho^{'}, \sigma_\rho) \\
z_{i,t} &\sim \text{Bernoulli}(\theta_{i,t}) \\
\text{logit}(\theta_{i,t}) &= z_{i,t-1}(\alpha_t + X_i\beta_{t_}) + (\prod_{k=1}^{t-1} 1 - z_{i,k})(\alpha_t + X_i\beta_{t_}) \\
\beta_{t,d} &\sim \mathcal{N}(\mu_d, \sigma_d) \\
\alpha_t &\sim \mathcal{N}(\mu + \phi_{p[t]} + U_t\gamma, \sigma_\mu) \\
\phi_p &\sim \mathcal{N}(0, \sigma_\phi)
\end{aligned}
$$

Note: Product term ensures taxon-loss is permanent. Implementation in Stan marginalizes over all possible (range-through) values of z instead of estimating the discrete parameters. I also use a noncentered parameterization of the hierarchical effects for better posterior sampling behavior. This presentation excludes final (hyper)priors.

Parameter estimation and inference

- full HMC/MCMC slow
- Automatic Differentiation Variational Inference (ADVI)
  - approximate Bayesian inference
  - assumes posterior is Gaussian, no correlation between parameters
  - true Bayesian posterior

Posterior predictive performance

Effect of mass on log-odds of occurrence

Basic model

Full model

Probability occurrence is of ecotype (basic model)

Probability occurrence is of ecotype (full model)

Group-level effects (climate, plant phase)

Concerns and conclusions

- basic and full models have similar results until Neogene
- posterior predictive simulations disimilar to observed; poor model adequacy
  - previous work has never evaluated model adequacy, so no baseline for comparison
  - second-order Markov process?
  - try for full posterior inference
- decreasing ability to discern arboreal taxa over time (absence/increased rarity)
- increase in scansorial taxa over time
- increase in herbivorous taxa over time
- group-level covariates (climate, plants) have very small effects

Acknowledgements

- Advising
  - Kenneth D. Angielczyk, Michael J. Foote, P. David Polly, Richard H. Ree, Graham Slater
- Angielczyk Lab
  - David Grossnickle, Dallas Krentzel, Jackie Lungmus
- Foote lab
  - Marites Villarosa Garcia, Nadia Pierrehumbert
- David Bapst, Ben Frable, Graeme Lloyd, Matt Pennell
- UChicago CEB

The Paleobiology Database
revealing the history of life