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

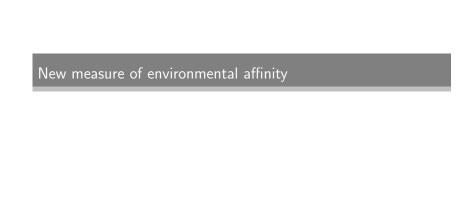
Other projects

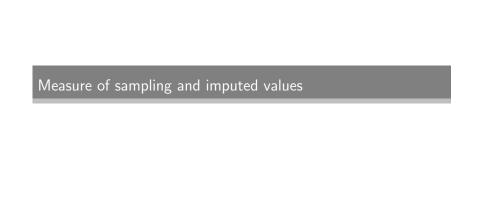
Taxon occurrence as a function of both emergent biological traits and its environmental context

Other projects

#### History

- presented at GSA 2015
- rejected from Evolution
  - encouraged resubmit
  - audience issues
  - difficult and transformative reviews
  - resubmitted 3 March





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

Other projects

### Analysis of Cenozoic mammal fossil record for NA

# individual-level (genus 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
- body size (rescaled log body mass)

#### group-level (2 My time unit t)

- overall mean of log-odds of occurrence probability
- temperature record
  - 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
- the latent discrete "true" presence modeled as a multi-level logistic regression
  - ▶ individual- and group-level

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

Note: My implementation in Stan marginalizes over all possible values of z and takes advantage of noncentered parameterizations of the hierarchical effects for better posterior sampling behavior.

## Posterior predictive model checking

- ▶ simulate fossil record given only  $y_{t=1}$ , all its covariates, and  $\theta$ 
  - where  $\theta$  is the set of all parameters
- leave-one-out cross-validation for time series
  - ▶ Bayesian statement is  $p(\tilde{y}_{-(t+1)}|y_{-t}\theta)$
- ROC as measure of performance

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

#### Other projects

# How cryptic is cryptic diversity? Machine learning approaches to classifying morphological variation in the Pacific Pond Turtle (*Emys marmorata*)

- estimate which species classification is best supported by morphology
  - multiple machine learning approaches
  - focus on one turtle species complex
  - results compared against results from two other turtle datasets
  - comparison of in- and out-of-sample model performance
- collaboration with Ken, Jim Parham, and Bryan Stuart
- submitted to then rejected from Systematic Biology
- resubmitted soon

# Modeling the rate at which new species are named.

- collaboration with Stewart Edie; he's lead
- I developed the statistical model
  - zero-inflated Poisson model
  - both Bernoulli and Poisson modeled as time series
  - response is the number of species named per publication per year for each biogeographic province
  - increasing, decreasing, or level?
- draft phase
- targets seem to be PNAS or Systematic Biology

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

Other projects

#### Post-doc ideas

- 1. Miller Fellowship at Berkeley with Charles Marshall
  - Charles has met me a couple times.
- Peter Buck Fellowship at Smithsonian with Gene Hunt (and Peter Wagner and Kate Lyons)
  - ► Gene, Pete, and Kate all know who I am.
- 3. Michigan Fellowship at University of Michigan with Matt Friedman
  - ▶ I don't know if he's actually moving there.
- 4. NIMBiOS Post-doc with Brian O'Meara
  - I don't know him.

#### Research statement

Modeling sources of variation in global and regional patterns of biodiversity over time.

- Intersection of macroevolution and macroecology.
- Species pool demography as focus.
- Paleontological data.