

Spring 2016 committee meeting for Peter Smits

How macroecology affects macroevolution: the interplay between extinction intensity and trait-dependent extinction in brachiopods.

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

Collaboration projects

Moving forward

How macroecology affects macroevolution: the interplay between extinction intensity and trait-dependent extinction in brachiopods.

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

Collaboration projects

Moving forward

Question and analysis

- ▶ How do the effect of emergent traits on duration (**extinction selectivity**) vary with expected duration (**extinction intensity**) and each other?
- ▶ What is the relationship between environmental affinity and duration wrt specialists versus generalists?
- ▶ **Approach:** hierarchical Bayesian survival model, varying intercepts and slopes, Weibull distribution, imputed gap statistic for taxa with duration < 3.

Analysis of post-Cambrian Paleozoic brachiopod genus durations

- ▶ stage as time unit; duration measured in stages
- ▶ multiple emergent traits analyzed
 - ▶ geographic range
 - ▶ used JADE to handle effect of biased sampling
(Chao et al. 2015 *Ecology*)
 - ▶ body size
 - ▶ from Payne et al from the *Treatise*
 - ▶ environmental preference (x , x^2)
 - ▶ gap statistic as measure of sampling (Foote and Raup 1996 *Paleobio*)
- ▶ factors vary by cohort (except gap statistic)

New measure of taxon's environmental affinity

(# epicontinental / total # occurrences) is what quantile of the distribution of all other background occurrences $\text{Beta}(\alpha, \beta)$.

- ▶ α is the # epicontinental background occurrences (+ 1).
- ▶ β is the # open ocean background (+ 1).

Measure of sampling and imputed values

Sampling is measured as the gap statistic r :

(number of bins with an occurrence - 2) / (duration in bins - 2)

Can only be estimated for taxa with duration of three or more.

Have to impute (e.g. fill-in) the values for all other taxa r^* .

$$s \sim \text{Beta}(\phi, \lambda)$$

$$\phi = \text{logit}^{-1}(W\gamma)$$

$$s^* \sim \text{Beta}(\phi^*, \lambda)$$

$$\phi^* = \text{logit}^{-1}(W^*\gamma)$$

Note: Beta distribution parameterized in terms of mean ϕ and total count λ .
Also, this presentation excludes final (hyper)priors.

Sampling statement for the joint posterior probability

$$y_{i,t} \sim \text{Weibull}(\sigma_{i,t}, \alpha)$$

$$\log(\sigma_{i,t}) = \frac{X_i B_{j[i],t} + \delta s_i}{\alpha}$$

$$B_j \sim \text{MVN}(\mu, \Sigma)$$

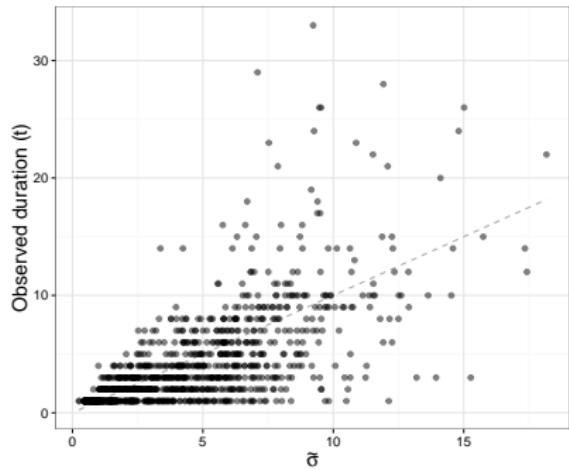
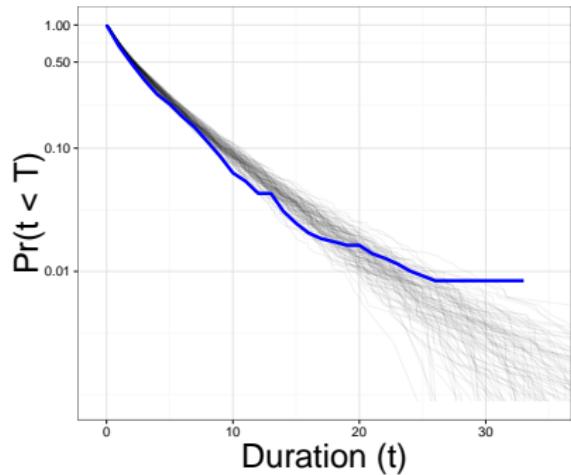
$$\Sigma = \text{diag}(\tau) \Omega \text{diag}(\tau)$$

$$s_i \sim \text{Beta}(\phi_i, \lambda)$$

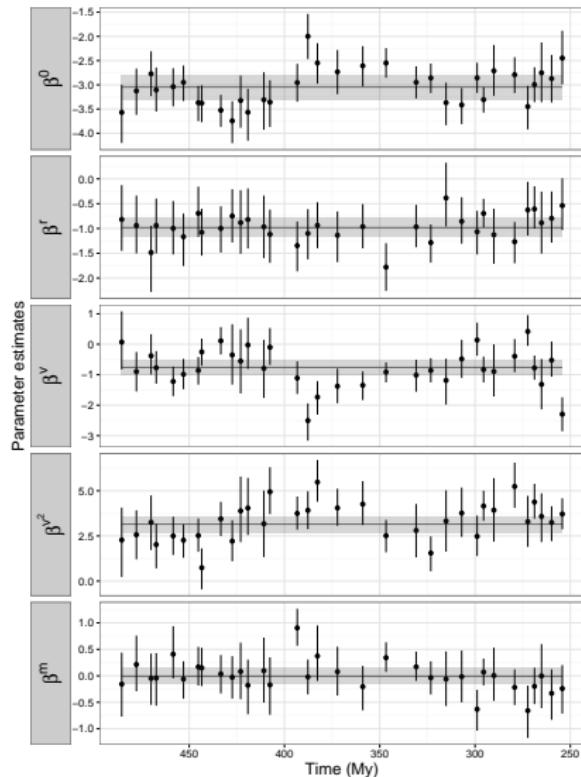
$$\phi_i = \text{logit}^{-1}(W_i \gamma)$$

Note: Calculation of log probability of right and left censored observations is modified from the above. Also, presentation excludes final (hyper)priors.

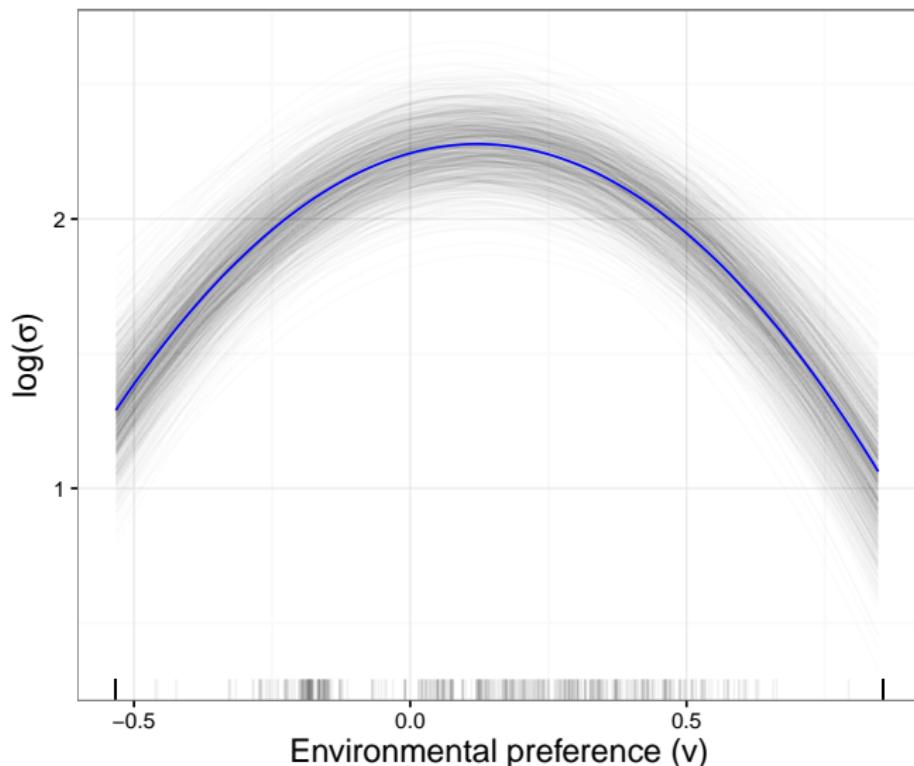
Model adequacy



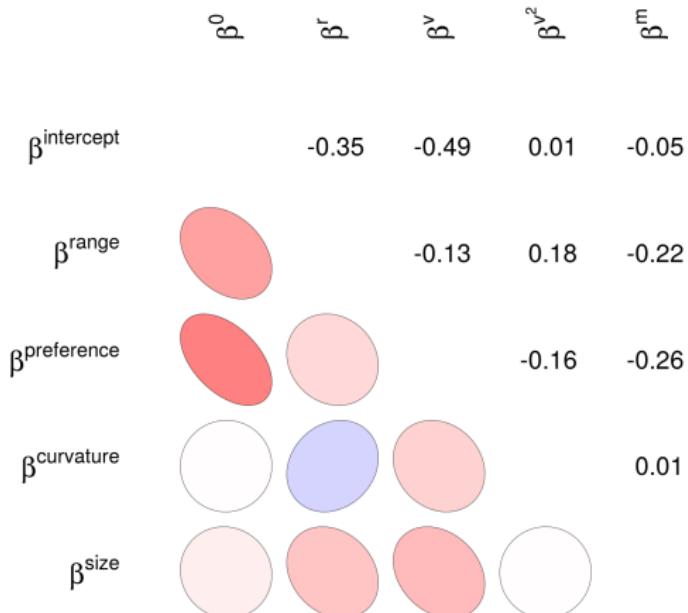
Effects of biological traits on survival



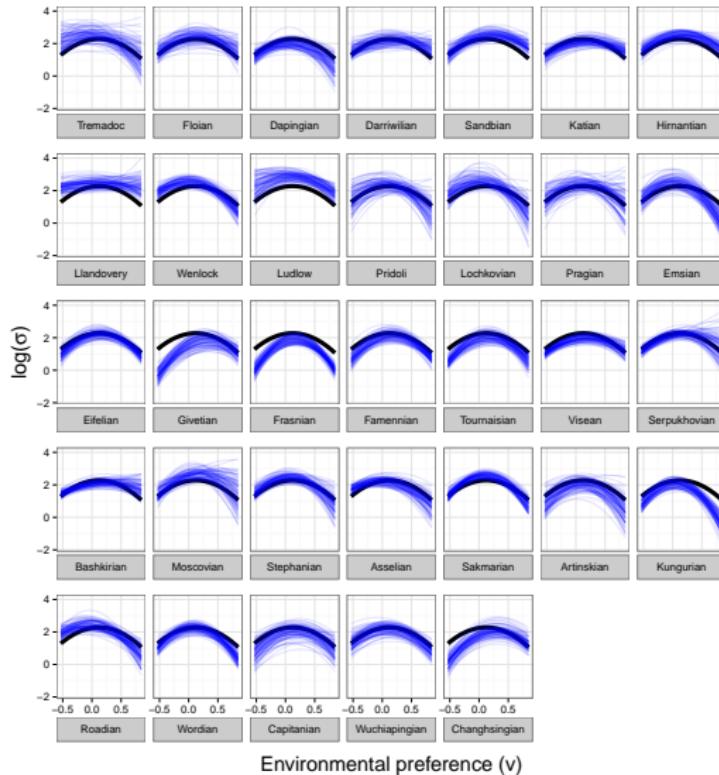
Parabolic effect of environmental preference on survival



Correlation between cohort trait effects



Parabolic effect of environmental preference on survival by stage



Conclusions

- ▶ model fit is ok, I wish it was better
- ▶ strong support for nonlinear relationship between environmental preference and survival
 - ▶ intermediate preference expected to have greater duration than either end members
- ▶ correlation between extinction intensity and selectivity
 - ▶ as average duration increases, effect of geographic range decreases **medium**
 - ▶ as average duration increases, y-intercept of the nonlinear effect of environmental preferences increases (e.g. from slight epicontinental to complete middle) **strong**
- ▶ increasing extinction risk with taxon age

Updates

- ▶ presented at GSA 2015
- ▶ rejected from *Evolution*
 - ▶ encouraged resubmit
 - ▶ audience issues
 - ▶ difficult but transformational reviews
 - ▶ resubmitted 3 March
 - ▶ still “under review” as of this morning (70 days)

How macroecology affects macroevolution: the interplay between extinction intensity and trait-dependent extinction in brachiopods.

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

Collaboration projects

Moving forward

Initial empirical setup

- ▶ arboreal taxa are expected to have greatest extinction risk of the locomotor categories (Smits 2015 PNAS)
- ▶ two possible causes
 1. this effect is constant for all time
 2. Paleogene-Neogene transition
 - ▶ neutral effect of arboreality during Paleogene
 - ▶ strong selection against arboreality during Neogene
 - ▶ Neogene effect is stronger than Paleogene effect
 - ▶ means that overall mean effect is closer to Neogene
- ▶ former implies no appreciable demographic differences over Cenozoic
- ▶ latter implies a difference in demography between Paleogene and Neogene

The ground-dwelling problem

- ▶ “ground-dwelling” too broad a category
- ▶ following a suggestion from Graham, split by foot posture
 - ▶ based primarily on Carrano 1997 *ZJLS*
- ▶ allows for estimation of subtler shifts in demography within “ground-dwelling”

Other covariates of interest

- ▶ traits
 - ▶ dietary category (shifts in trophic structure)
 - ▶ body mass
- ▶ environmental factors
 - ▶ floral phase following Graham 2010
A Natural History of the New World
 - ▶ global temperature, both mean and interquartile range

traits vary with time;
environmental factors predict average occurrence probability

can estimate probability of expected enrichment or depletion effect
of individual traits for each time bin wrt all other analyzed traits

Theoretical underpinning

- ▶ changes in demographic structure of regional species pool
 - ▶ I'm not modeling the diversification process, **only presence**
 - ▶ no difference between origination, survival, extinction in effect
- ▶ intersection of macroecology and macroevolution
- ▶ fourth-corner type problem
 - ▶ trait- **and** environment-based assembly

Conceptual diagram of fourth-corner problem

GLMs for trait-based community assembly 397

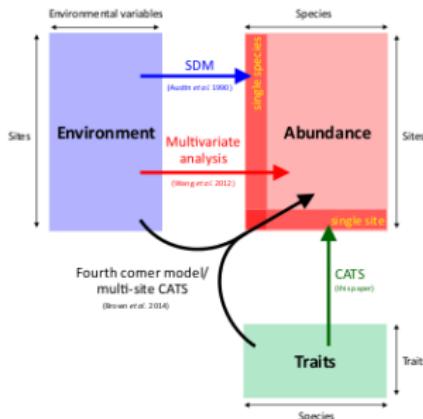
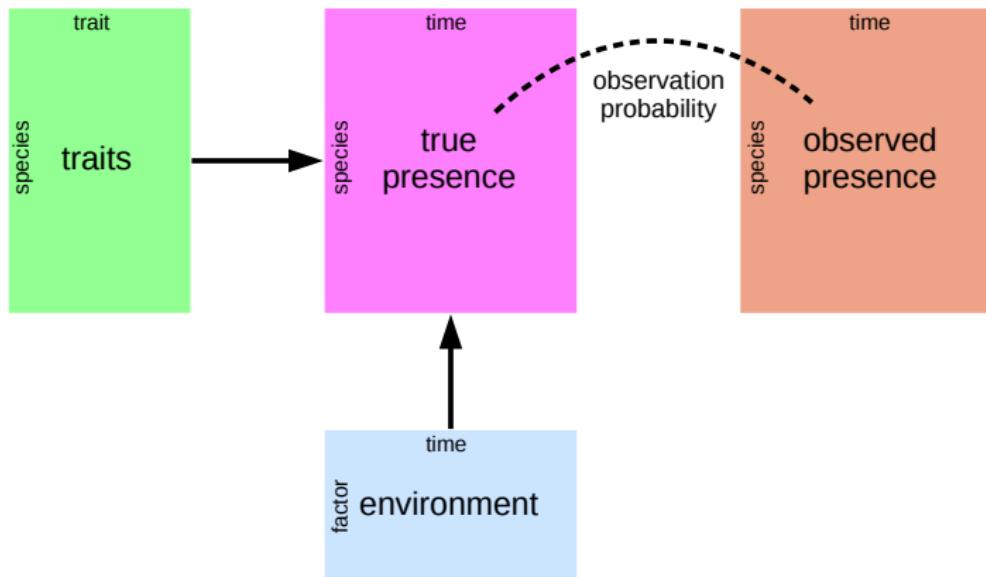


Fig. 4. Schematic diagram of the relationship of CATS and multisite CATS to species distribution modelling, multivariate analysis and fourth-corner models. All these methods can be understood as fitting predictive models for the abundance ('Abundance' table) of one or more species at one or more sites as a function of environmental variables ('Environment' table) and/or species traits ('Traits' table). In principle, any predictive modelling framework could be used to fit any of these models, but in each case, an example reference has been included which used generalized linear models.

(Warton, Shipley, and Hastie. 2015. *Methods in Ecology and Evolution.*)

Conceptual diagram of my model



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
- ▶ 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
- ▶ the latent discrete “true” presence modeled as a multi-level logistic regression
 - ▶ individual- and group-level

Sampling statement for the joint posterior probability

$$y_{i,t} \sim \text{Bernoulli}(\rho_t z_{i,t})$$

$$\text{logit}(\rho_t) \sim \mathcal{N}(\rho', \sigma_\rho)$$

$$z_{i,t} \sim \text{Bernoulli}(\theta_{i,t})$$

$$\text{logit}(\theta_{i,t}) = z_{i,t-1}(\alpha_t + X_i \beta_{t-}) + \left(\prod_{k=1}^{t-1} 1 - z_{i,k} \right) (\alpha_t + X_i \beta_{t-})$$

$$\beta_{t,d} \sim \mathcal{N}(\mu_d, \sigma_d)$$

$$\alpha_t \sim \mathcal{N}(\mu + \phi_p[t] + U_t \gamma, \sigma_\mu)$$

$$\phi_p \sim \mathcal{N}(0, \sigma_\phi)$$

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.

Assessing model fit from the posterior predictive distribution

- ▶ simulate fossil record given only $y_{t=1}$ given all covariates, θ
 - ▶ where θ is the set of all parameters
- ▶ cross-validation for time series
 - ▶ Bayesian statement is $p(\tilde{y}_{-(t+1)}|y_{-t}, \theta)$
 - ▶ issues because of high structured data (e.g. floral phase)

Current project issues

- ▶ fit with Stan (NUTS sampling/MCMC of posterior)
- ▶ crocorock CEB server
- ▶ 4 parallel chains
 - ▶ 10 000 warm-up, 10 000 sampling, thinned by 10
- ▶ slow to converge
 - ▶ simple model with no individual- or group-level effects takes a week (same settings)
- ▶ solutions?
 - ▶ restrict taxonomic or temporal scope?
 - ▶ decrease model complexity: where and how?
 - ▶ other reparameterizations?

How macroecology affects macroevolution: the interplay between extinction intensity and trait-dependent extinction in brachiopods.

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

Collaboration projects

Moving forward

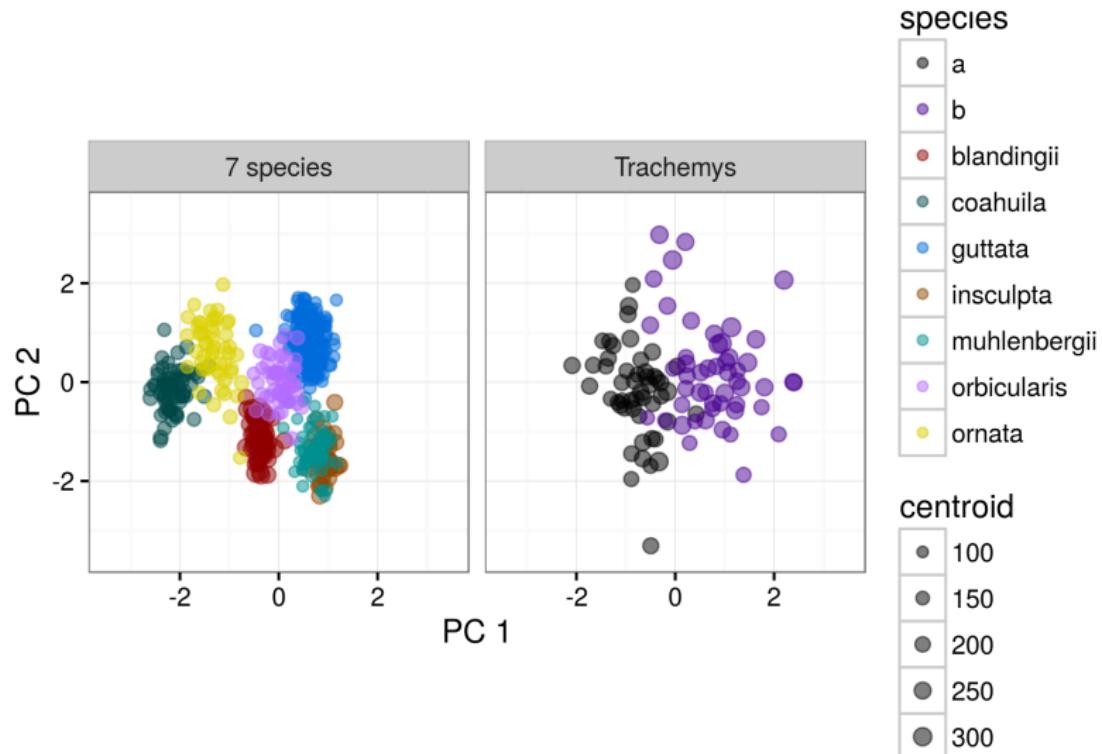
How cryptic is cryptic diversity? Machine learning approaches to classifying morphological variation in the Pacific Pond Turtle (*Emys marmorata*)

- ▶ estimate which species classification is best supported by morphology
 - ▶ multiple machine learning approaches
 - ▶ three datasets; special attention to *Emys*
 - ▶ comparison of in- and out-of-sample model performance
- ▶ collaboration with Ken, Jim Parham, and Bryan Stuart
- ▶ submitted to then rejected from Systematic Biology

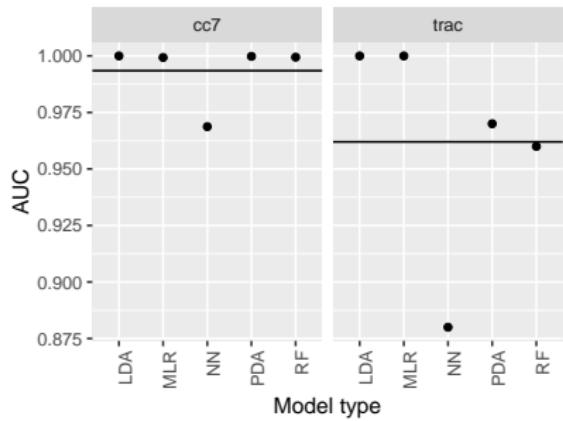
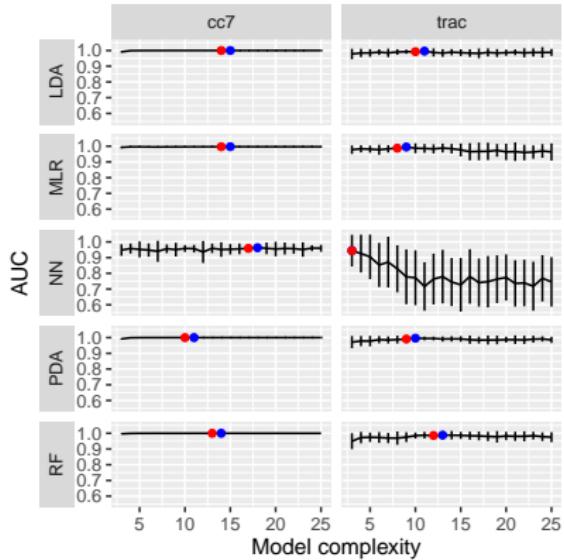
Statistical learning approach

- ▶ split data into training and testing datasets (80/20)
- ▶ training dataset
 - ▶ multiple models w/ between 3-28 predictors
 - ▶ 10 rounds, 5-fold cross-validation
 - ▶ estimate out-of-sample AUC of ROC
 - ▶ compare for model selection
 - ▶ follows Hastie et al. 2009 “Elements of Statistical Learning”
- ▶ testing dataset
 - ▶ using selected model, predict classification
 - ▶ AUC of ROC as measure of success
 - ▶ compare average from multiple models to compare hypotheses

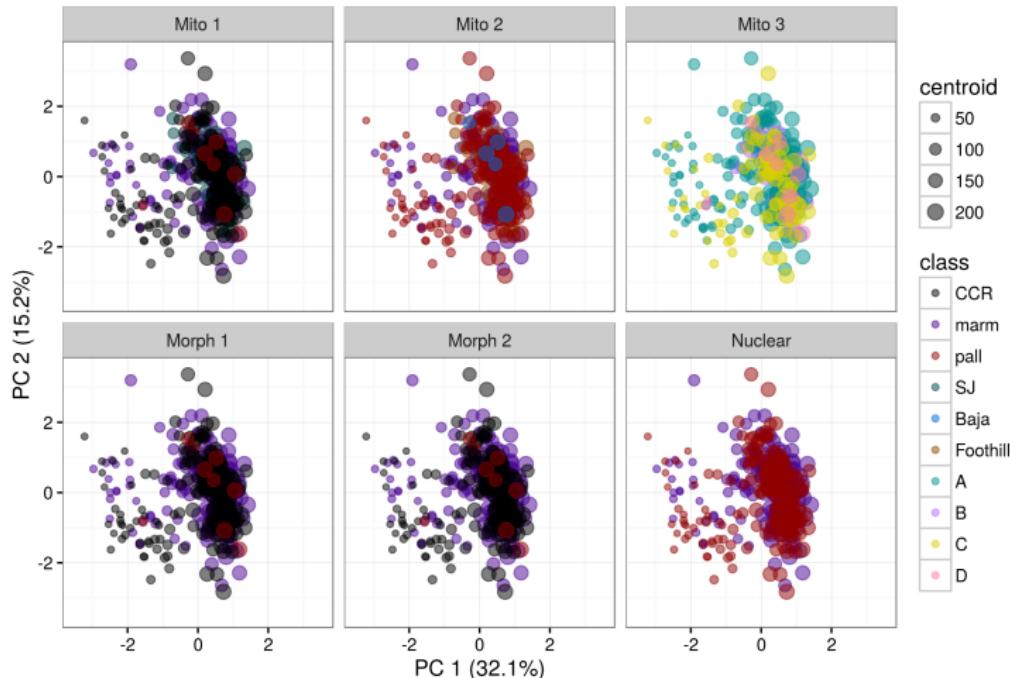
PC-s for the 7 turtle and *Trachemys* species sets



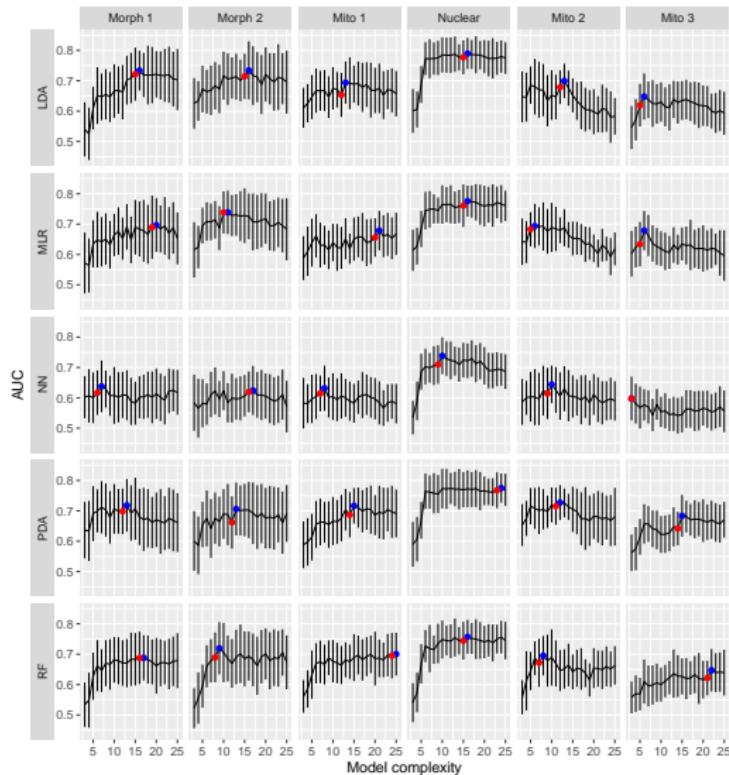
7 turtle, and *Trachemys* species sets



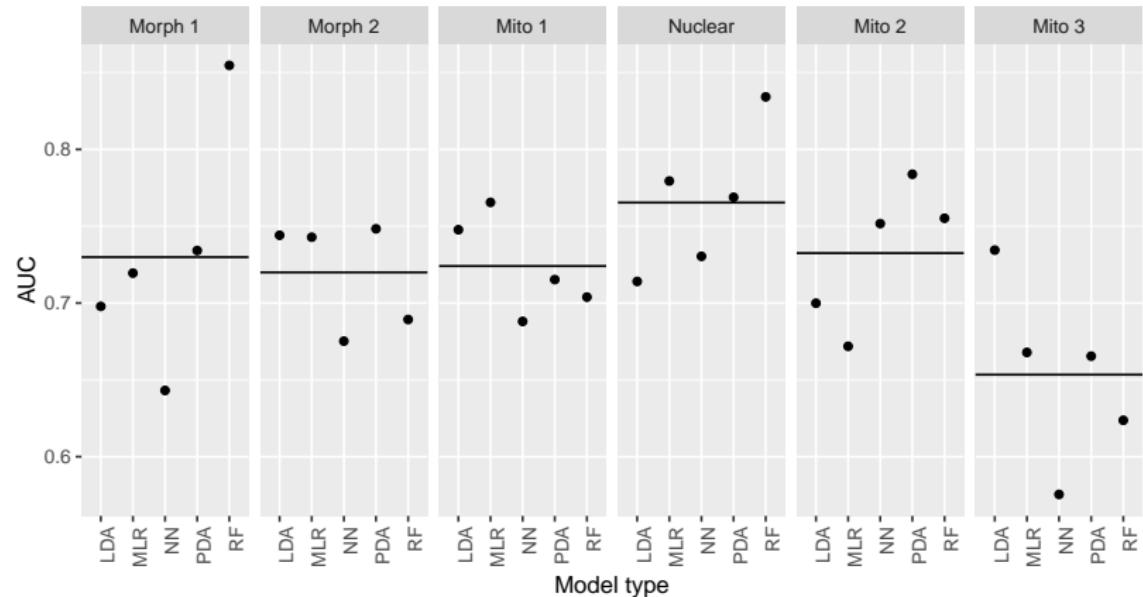
PC-s for the *Emys* species sets



Emys: CV of model using training data



Emys: Results of prediction of testing data



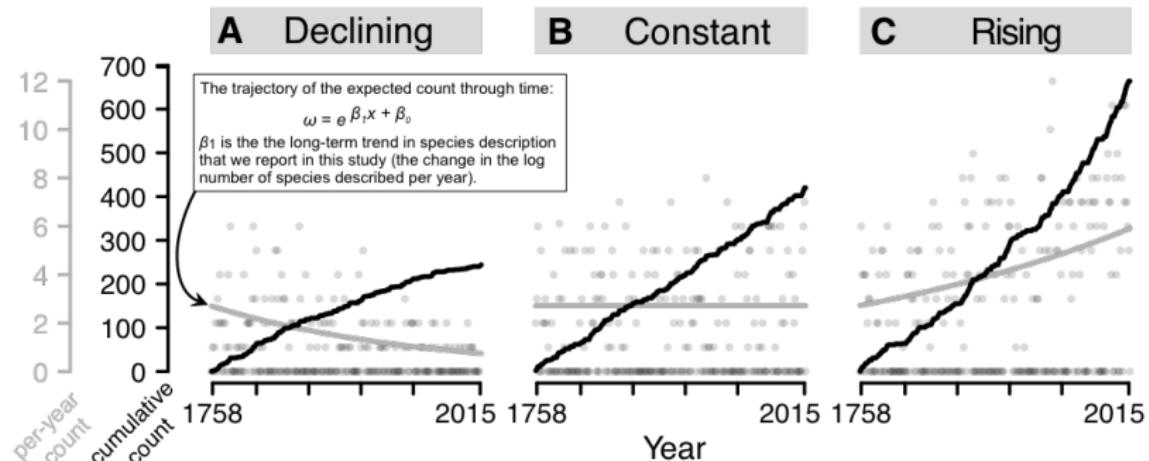
Modeling the rate at which new species are named.

Goal

Predictive model of number of bivalve species named (per publication) per year by biogeographic province.

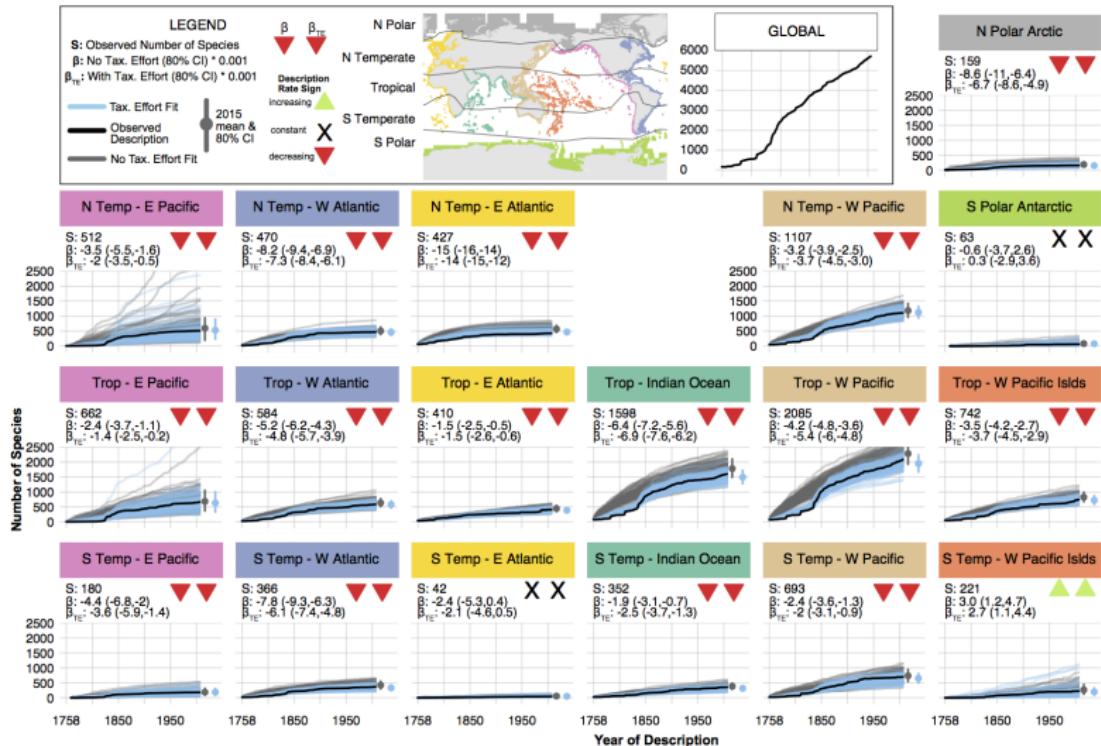
- ▶ collaboration with Stewart Edie
- ▶ he developed the question, I developed the model
- ▶ hierarchical zero-inflated Poisson time-series model
- ▶ forecast rank order of provincial diversity in approx. 15 years

Conceptual model



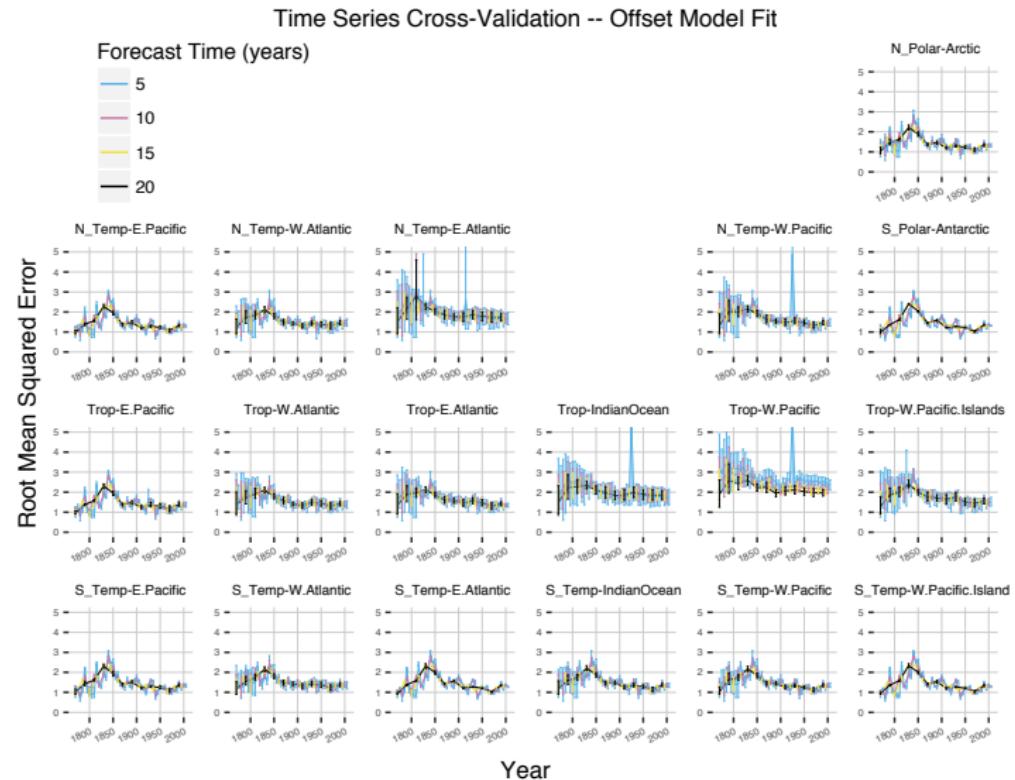
(Edie, Smits, et al. *in. prep.*)

Regional accumulation and prior predictive comparison



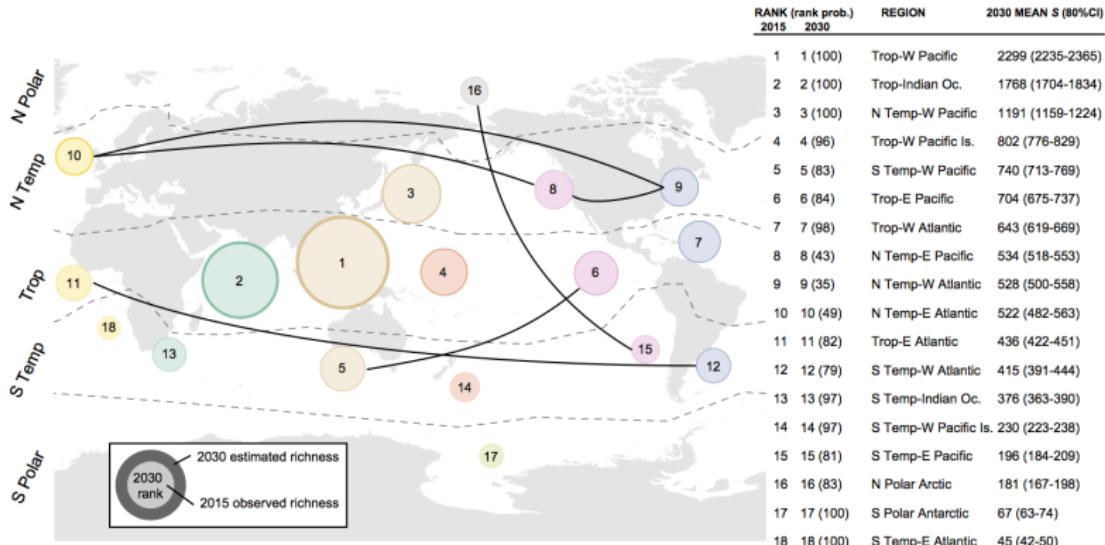
(Edie, Smits, et al. *in. prep.*)

Cross-validation of forecasting error



(Edie, Smits, et al. *in. prep.*)

Forecasted regional ranks



(Edie, Smits, et al. *in. prep.*)

How macroecology affects macroevolution: the interplay between extinction intensity and trait-dependent extinction in brachiopods.

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

Collaboration projects

Moving forward

- 1. Miller Fellowship at Berkeley with Charles Marshall**
 - ▶ Charles has met me a couple times.
- 2. 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 him
- 4. NIMBioS Post-doc with Brian O'Meara**
 - ▶ I don't know him.

My “research program”

- ▶ macroevolutionary macroecology
 - ▶ research at the intersection of disciplines
- ▶ hypothesis driven
 - ▶ empirical tests of macroevolutionary theories
- ▶ modeling (Bayesian, hierarchical/GLMM)
 - ▶ emphasis on model checking and validation