Remodeling the fossil record

analysis of emergent evolutionary and ecological patterns

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

- Macroevolution and macroecology
- Structured data and modelling emergent patterns
- Patterns in extinction
  - Background extinction and expected differences in species survival
  - Interplay between extinction intensity and extinction selectivity
- Patterns in functional diversity
  - Mammal species pool functional composition
- Conclusions and commentary

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in extinction

- Background extinction and expected differences in species survival
- Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

- Mammal species pool functional composition

Conclusions and commentary

levels of organisation and emergence

Definition

- Macroevolution
- Macroecology

trait: identifiable property of an organism

species trait: idetifiable property of the entire species

functional trait: class of traits describing interaction with environment species selection

Rabosky and McCune: due to heritability of speciation/extinction rates

operationalize through traits while paying attention to definition of fitness species fitness

logic: if more fit, more likely to be present

Cooper 84 fundamental defintion: expected time till extinction

derived definitions depend on definition of extinction (lines-of-descent)

Extinction

fundamentally emergent phenomenon; all members must die for collective to die.

Law of Constant Extinction

Extinction risk, in a given adaptive zone, is taxon–age independent.

(Van Valen 1973 *Evol. Theory*)

Survival of the unspecialized

When related phyla die out ... more specialized phyla tend to become extinct before less specialized. This phenomenon is also far from universal, but it is so common that it does deserve recognition as a rule or principle in evolutionary studies: **the rule of the survival of the relatively unspecialized.**

(Simpson, 1944, Tempo and Mode of Evolution, p. 143*)*

Species pool concept

(Mittelbach and Schemske, 2015, *TREE*)

functional composition of species pool

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in extinction

Background extinction and expected differences in species survival

Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

Mammal species pool functional composition

Conclusions and commentary

Inference

Learning from incomplete information.

statistical model as inference device

engineering approach to analysis: building blocks used to create the device structure in biological and paleontological data

Models of macroevolution

birth-death for diversity

Brownian motion for trait

Species distribution models

aboitic associations – traditional, maxent

biotic associations – cats/maxent assembly

fourth-corner – where we are

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in extinction

- Background extinction and expected differences in species survival
- Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

- Mammal species pool functional composition

Conclusions and commentary

Question

Why do taxa go extinct at different rates?

Motivating questions

- How do mammal species traits affect extinction risk?
  - How do shared time of origination or evolutionary history relate to extinction risk?
- How do my findings compare to current risk factors?
- Is species extinction risk age-independent?

Relationship between range size and extinction risk

(Harnik and Simpson 2013 *Proc B*)

Hypotheses of effects of dietary category

Hypotheses of effects of locomotor category

Hypotheses of effects of locomotor category

Survival model diagram

Pattern of species survival under two models

Effect of dietary category on extinction risk

Effect of locomotor category on extinction risk

Difference in risk between origination cohorts

Three sources of variance

Conclusions

- Survival of the unspecialized as time-invariant generalization.
- Decrease in extinction risk with time.
  - Both cohort/temporal and phylogenetic effect.
- Some incongruence with risk factors in the Recent.
  - e.g. effect of body size, trophic category, phylogenetic clustering.

Observation

At K/Pg mass extinction, biological traits (except geographic range) have no effect on bivalve taxonomic survival.

(Jablonski, 1986, *Science*)

Questions and analysis

- How do the effect of emergent traits on duration (extinction selectivity) vary with expected duration (extinction intensity)?
- **Approach:** hierarchical Bayesian survival model;
  effect estimates vary with origination cohort;
  correlation btw effects modeled.

Intensity and selectivity

Brachiopods

(ComputerHotline, wikimedia CC BY 2.5; Dwergenpaartje, wikimedia CC BY-SA 3.0)

Post-Cambrian Paleozoic brachiopod genera and covariates

- time range approx. 488-252 Mya.
- stage as time unit; duration measured in stages (2-5 My each)
- multiple emergent traits analyzed; estimates vary by origination cohort
  - geographic range
  - body size
  - environmental preference (v, v^2)
- gap statistic as measure of sampling (Foote and Raup 1996 *Paleobio*)
  imputed for taxa with short durations

New measure of taxon's environmental affinity

(# epicontinental / total # occurrences) is what quantile of the distribution of all other background occurrences Beta(α, β).

- 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^*.

$$
\begin{aligned}
s &\sim \text{Beta}(\phi, \lambda) \\
\phi &= \text{logit}^{-1}(W\gamma) \\
s^* &\sim \text{Beta}(\phi^*, \lambda) \\
\phi^* &= \text{logit}^{-1}(W^*\gamma)
\end{aligned}
$$

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

$$
\begin{aligned}
y_{i,t} &\sim \text{Weibull}(\sigma_{i,t}, \alpha) \\
\log(\sigma_{i,t}) &= \frac{X_i B_{j[i],t} + \delta s_i}{\alpha} \\
B_j &\sim \text{MVN}(\mu, \Sigma) \\
\Sigma &= \text{diag}(\tau)\Omega\text{diag}(\tau) \\
s_i &\sim \text{Beta}(\phi_i, \lambda) \\
\phi_i &= \text{logit}^{-1}(W_i \gamma)
\end{aligned}
$$

$$
\begin{aligned}
\mu_{intensity} &\sim \mathcal{N}(0, 5) \\
\mu_{range} &\sim \mathcal{N}(-1, 1) \\
\mu_{envpref} &\sim \mathcal{N}(0, 1) \\
\mu_{envcurve} &\sim \mathcal{N}(1, 1) \\
\mu_{size} &\sim \mathcal{N}(0, 1) \\
\delta &\sim \mathcal{N}(0, 1) \\
\tau &\sim \text{C}^{+}(1) \\
\Omega &\sim \text{LKJ}(1) \\
\lambda &\sim \text{Pareto}(0.1, 1.5) \\
\gamma &\sim \mathcal{N}(0, 1)
\end{aligned}
$$

Note: Calculation of log probability of right and left censored observations is modified from the above

Hierarchical survival model

Model adequacy

Variation in trait effects between cohorts

Overall effect of environmental preference

Change in effect of environment between cohorts

Change in effect of environment between cohorts

Correlation of effects between cohorts

Effect summary

- Effect of geographic range consistent with prior expectations; low variance.
- No effect of body size; low variance.
- Epicontinental environmental preference slightly favored on averaged; high variance.
- Strong support for survival of unspecialized as generalization wrt environmental preference; medium variance.

Macroevolutionary process

- Magnitude of effect of geographic range and environmental preference increase with extinction intensity.
- As extinction risk decreases, the differences between taxa matter less.

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in extinction

- Background extinction and expected differences in species survival
- Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

- Mammal species pool functional composition

Conclusions and commentary

(Smits, 2015, *PNAS*)

Eco-cube and ecotypes

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

Fourth-corner modelling problem

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

Paleontological 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 2011

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 covariates

Paleontological fourth-corner model

Model and sampling statement definition

$$
\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(p_{i,t} z_{i,t}) \\
p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) \\
r_t &\sim \mathcal{N}(0, \sigma) \\
\alpha_0 &\sim \mathcal{N}(0, 1) \\
\alpha_1 &\sim \mathcal{N}(1, 1) \\
\sigma &\sim \mathcal{N}^{+}(1) \\
z_{i,1} &\sim \text{Bernoulli}(\phi_{i,1}) \\
z_{i,t} &\sim \text{Bernoulli}\left(z_{i,t-1}\pi_{i,t} + \sum_{x=1}^{t}(1 - z_{i,x})\phi_{i,t}\right) \\
\phi_{i,t} &= \text{logit}^{-1}(a^{\phi}_{t,j[i]} + b^{\phi}_1 m_i + b^{\phi}_2 m_i^2) \\
\pi_{i,t} &= \text{logit}^{-1}(a^{\pi}_{t,j[i]} + b^{\pi}_1 m_i + b^{\pi}_2 m_i^2) \\
a^{\phi} &\sim \text{MVN}(U\gamma^{\phi}, \Sigma^{\phi}) \\
a^{\pi} &\sim \text{MVN}(U\gamma^{\pi}, \Sigma^{\pi})
\end{aligned}
$$

$$
\begin{aligned}
\Sigma^{\phi} &= \text{diag}(\tau^{\phi})\Omega^{\phi}\text{diag}(\tau^{\phi}) \\
\Sigma^{\pi} &= \text{diag}(\tau^{\pi})\Omega^{\pi}\text{diag}(\tau^{\pi}) \\
\rho &\sim \text{U}(0, 1) \\
b^{\phi}_1 &\sim \mathcal{N}(0, 1) \\
b^{\pi}_1 &\sim \mathcal{N}(0, 1) \\
b^{\phi}_2 &\sim \mathcal{N}(-1, 1) \\
b^{\pi}_2 &\sim \mathcal{N}(-1, 1) \\
\gamma^{\phi} &\sim \mathcal{N}(0, 1) \\
\gamma^{\pi} &\sim \mathcal{N}(0, 1) \\
\tau^{\phi} &\sim \mathcal{N}^{+}(1) \\
\tau^{\pi} &\sim \mathcal{N}^{+}(1) \\
\Omega^{\phi} &\sim \text{LKJ}(2) \\
\Omega^{\pi} &\sim \text{LKJ}(2).
\end{aligned}
$$

Note: Product term ensures taxon-loss is permanent. Implementation in Stan marginalizes over all possible (range-through) values of z instead of estimating the

Parameter estimation and inference

- full HMC/MCMC slow
- Automatic Differentiation Variational Inference (ADVI)
    - approximate Bayesian inference
    - assumes posterior is Gaussian
    - true but approximate Bayesian posterior

Posterior predictive performance

Effect of mass on log-odds of observation

Effect of mass on log-odds of occurrence

Probability of ecotype origination

Probability of ecotype survival

Group-level effects (plant phase, climate)

Total species pool diversity and diversification

Ecotype-specific diversity

Ecotype-specific origination

Ecotype-specific extinction

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
  - second-order Markov process?
  - 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
- plant phase has small, idiosyncratic effects

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in extinction

- Background extinction and expected differences in species survival
- Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

- Mammal species pool functional composition

Conclusions and commentary