Remodeling the fossil record

analysis of emergent evolutionary and ecological patterns

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

The Paleobiology Database
revealing the history of life

Macroevolution and macroecology

Structured data and modelling emergent patterns

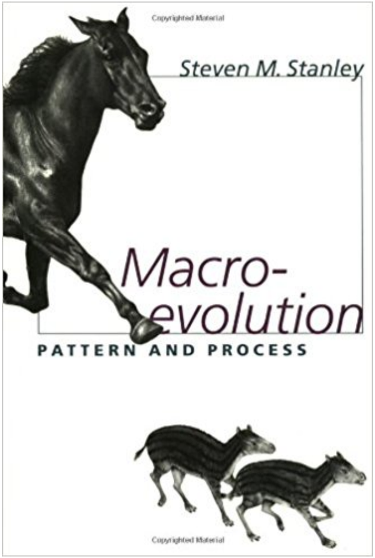
Patterns in survival

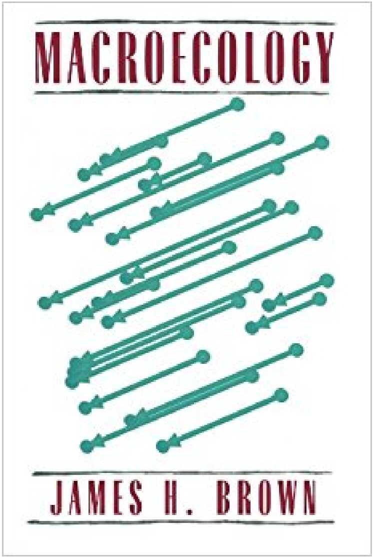
- Background extinction and expected differences in species survival
- Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

- Mammal species pool functional composition

Conclusions and commentary


Copyrighted Material
Steven M. Stanley
Macro-
evolution
PATTERN AND PROCESS
Copyrighted Material


MACROECOLOGY
JAMES H. BROWN

Definition

- macroevolution: study of patterns which emerge when considering the evolutionary history of multiple species.
- macroecology: study of patterns which emerge when considering the ecology of multiple species.
- both in time and space

Traits as conceptual and operational link

Definition

- trait: identifiable property of an organism
  e.g. pelage color, body mass, beak depth, tooth shape
- functional trait: trait that strongly influences performance/means of interacting with environment
- species trait: identifiable property assignable to a species

Species selection

Rabosky and McCune 2010 *TREE*

> *Species selection is the outcome of heritable variation in speciation and extinction rates among taxa.*

- avoids selection versus sorting, "strict" species selection versus effect macroevolution
- no operational structure
  - species inherit traits
  - traits affect fitness

Species fitness

Definition

Expected time till extinction.

(Cooper 1984 *J. Theoretical Biology*)

- logic: if more fit, more likely to be present
- distribution based definition (population)
- other definitions can be derived, just expand definition of extinction
  - lines-of-descent

Extinction

Population decline maybe a common cause of ...extinction, but organisms never die from population decline, and population decline is never caused by organismal death alone ...It's the relative balance between birth, death, and lifespan of organisms that determines ...extinction.

(Simpson 2016 *bioRxiv*)

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 in Evolution p. 143*)

Species pool concept

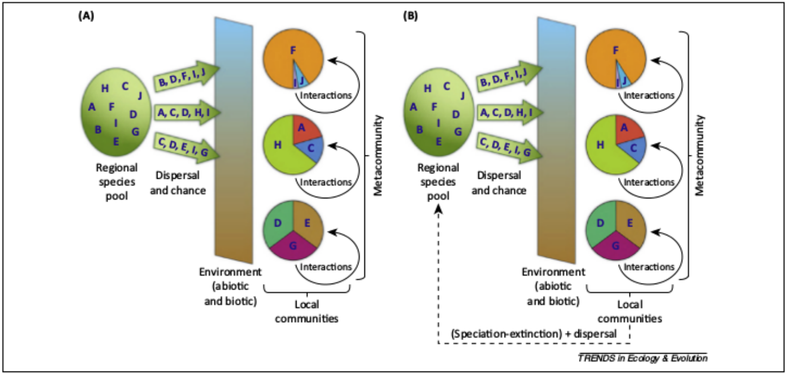
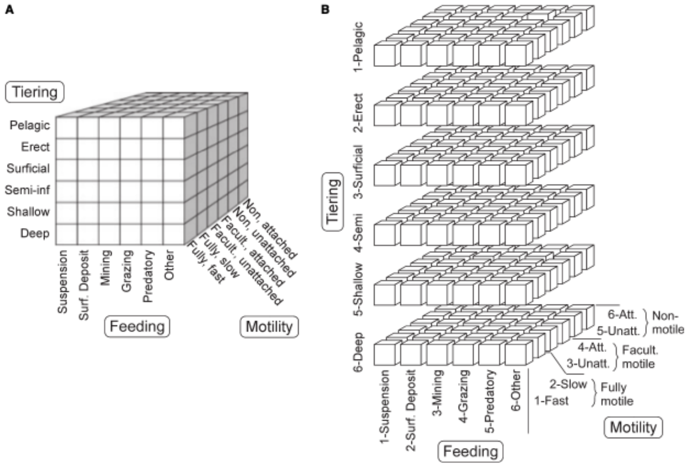


Figure 1. Two models of community assembly. **(A)** Local communities comprise a subset of species from the regional species pool that have passed through environmental filters. There is no feedback from the metacommunity (collection of local communities) to the regional species pool. Adapted from [5]. **(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*)

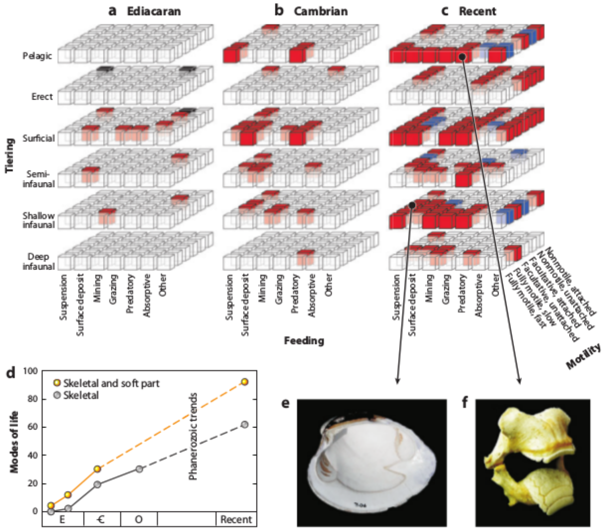
Functional groups



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)

Functional groups over time



(Bush and Bambach 2011 *Annu. Rev. Earth Planet Sci.*)

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in survival

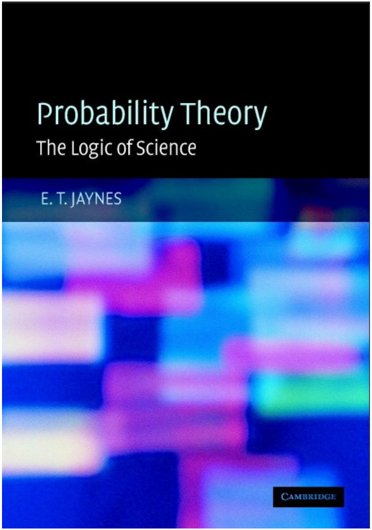
- Background extinction and expected differences in species survival
- Interplay between extinction intensity and extinction selectivity

Patterns in functional diversity

- Mammal species pool functional composition

Conclusions and commentary

Inference



The theory of probability is the only mathematical tool avaliable to help map the unknown and the uncontrollable.

(B. Mandelbrot)

Inference devices and statistical models

statistical model as inference device

engineering approach to analysis: building blocks used to create the device

Structured data in biology and paleontology

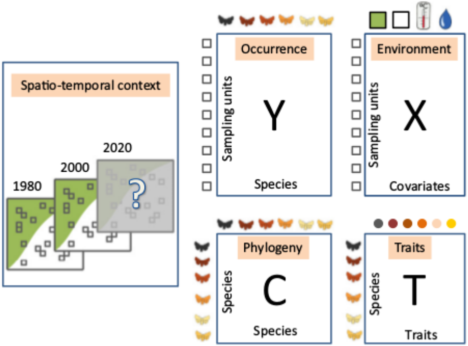


Figure 3 Data typically collected in community ecology. The occurrence data (denoted as the **Y** matrix) includes the occurrences of the species recorded in a set of temporal and/or spatial sampling units. The environmental data (denoted as the **X** matrix) consists of the environmental covariates measured over the sampling units. The traits data (denoted as the **T** matrix) consists of a set of traits measured for the species present in the **Y** matrix. To account for the phylogenetic dependencies among the species, we can include a fourth matrix consisting of the phylogenetic correlations among the species (denoted as the **C** matrix). The spatiotemporal context includes location and time information about the samples.

(Ovaskainen *et al.* 2017 *Ecology Letters*)

Models of structured data

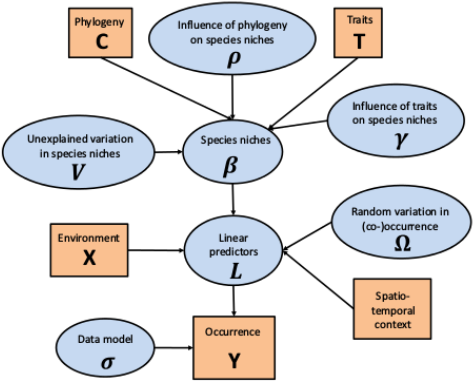


Figure 4 A graphical summary of the HMSC statistical framework. In this Directed Acyclic Graph (DAG), the orange boxes refer to data, the blue ellipses to parameters to be estimated, and the arrows to functional relationships described with the help of statistical distributions.

(Ovaskainen *et al.* 2017 *Ecology Letters*)

Models of macroevolution: birth-death

Models of macroevolution: Brownian motion

$x_t - x_s \sim \mathcal{N}(\mu, \sigma)$

Species distribution models

Goal

Understand relationship between species distribution in space (and time) as function of that species' environmental context or species trait values.

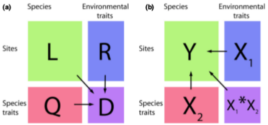


Fig. 1. Graphical representation of the fourth-corner problem and its solution. (a) The problem as posed by Legendre, Galzin & Harmelin-Vivien (1997), where the goal is to combine abundance (**L**), trait (**Q**) and environment (**R**) data in some way, to determine a matrix describing the trait–environment 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.*)

Bayesian statistics

framework used throughout this dissertation

flexible, expressive, intuitive

using HMC and ADVI as implemented in Stan

Macroevolution and macroecology

Structured data and modelling emergent patterns

Patterns in survival

- 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

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?

Cenozoic mammals of North America

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

Hierarchical survival model

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

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

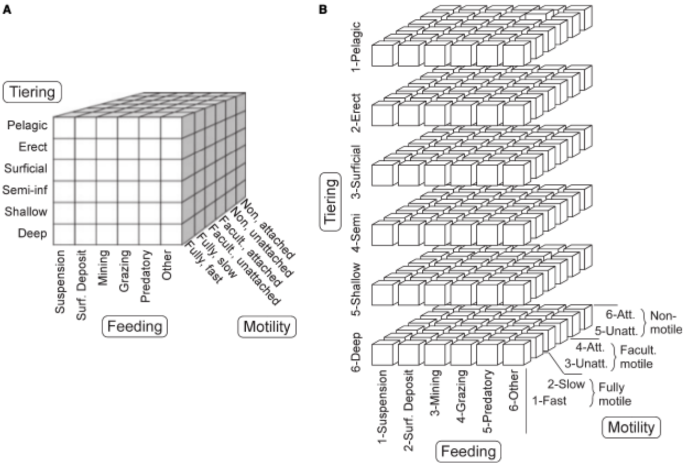
- Effect of geographic range consistent with prior expectations; low variance.
- No effect of body size; low variance.
- Epicontinental environmental preference slightly favored on averaged; high variance.
- Strong support for survival of unspecialized as generalization wrt environmental preference;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

(Smits 2015 *PNAS*)

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)

Fourth-corner modelling problem

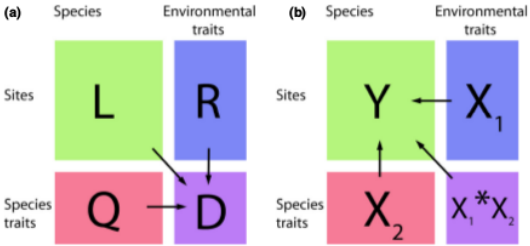


Fig. 1. Graphical representation of the fourth-corner problem and its solution. (a) The problem as posed by Legendre, Galzin & Harmelin-Vivien (1997), where the goal is to combine abundance (**L**), trait (**Q**) and environment (**R**) data in some way, to determine a matrix describing the trait–environment 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.)

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

Paleontological fourth-corner model

Model and sampling statement definition

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

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

Posterior predictive performance

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 survival

- 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

Summary

three studies on emergent patterns and their relation to functional traits

many questions and hypotheses, both domain specific and general

Advances

law of constant extinction

survival of the unspecialized

functional diversity

Future

spatial data

better models of preservation

better hypotheses of relation between trait and target (ezard example)

Final thoughts