

Title: The interplay between extinction intensity and selectivity:
2 correlation in trait effects on taxonomic survival

Running title: Variation in trait effects on taxonomic survival

4 **Author:** Peter D Smits, psmits@uchicago.edu, Committee on Evolutionary
Biology, University of Chicago, IL, USA.

6 **Keywords:** macroevolution, extinction, macroecology, Bayesian, brachiopods

Word count:

8 **Table count:** 2

Figure count: 6

10 **Data archival location:** If accepted, all data and code necessary to duplicate
this analysis will be made available on DRYAD.

Abstract

While the effect of geographic range on extinction risk is well documented, how other traits may increase or decrease extinction risk is less well known. I analyze patterns of Paleozoic brachiopod genus durations and their relationship to geographic range, affinity for epicontinental seas versus open ocean environments, and body size. Additionally, I allow for environmental affinity to have a nonlinear effect on duration. Using a hierarchical Bayesian approach, I also model the interaction between the effects of biological traits and a taxon's time of origination. My analysis framework eschews the traditional distinction between background and mass extinction, and instead the entire time period is analyzed as parts of the same continuum.

These results support the hypothesis that as extinction intensity increases, overall extinction selectivity decreases.

1 Introduction

How do biological traits affect extinction risk? Biological traits are defined here as descriptors of a taxon's adaptive zone, which is the set of all biotic–biotic and biotic–abiotic interactions that a taxon can experience (Simpson, 1944). In effect, these are descriptors of a taxon's broad-sense ecology. Jablonski (1986) observed that during a mass extinction event, the effects of biological traits on taxonomic survival decreased in size. However, this pattern was not the case for the effect of geographic range on survival (Jablonski, 1986).

Jablonski (1986) phrased his conclusions in terms of background versus mass extinction, but this scenario is readily transferable to a continuous variation framework as there is no obvious distinction in terms of extinction rate between these two states (Wang, 2003). Additionally, the Jablonski (1986) scenario has

38 strong model structure requirements in order to test its proposed
macroevolutionary mechanism; not only do the taxon trait effects need to be
40 modeled, but the correlation between trait effects need to be modeled as well.

There are two end-member macroevolutionary mechanisms which may underlie
42 the pattern observed by Jablonski (1986): the effect of geographic range on
predictive survival remains constant and those of other biological traits decrease,
44 or the effect of geographic range in predicting survival increases and those of
other biological traits stay constant. Reality, of course, may fall somewhere
46 along this continuum.

Conceptually, taxon survival can be considered an aspect of “taxon fitness”
48 along with expected lineage specific branching/origination rate (Cooper, 1984,
Palmer and Feldman, 2012). A taxon with a beneficial trait should persist for
50 longer, on average, than a taxon without that beneficial trait. Here I model
brachiopod taxon durations because trait based differences in extinction risk
52 should manifest as differences in taxon durations. Brachiopods are an ideal
group for this study as they are well known for having an exceptionally
54 complete fossil record (Foote, 2000). I focus on the brachiopod record from most
of the Paleozoic, from the start of the Ordovician (approximately 485 My)
56 through the end Permian (approximately 252 My) as this represents the time of
greatest global brachiopod diversity (Alroy, 2010).

58 The analysis of taxon durations, or time from origination to extinction, falls
under the purview of survival analysis, a field of applied statistics commonly
60 used in health care (Klein and Moeschberger, 2003) but has a long history in
paleontology (Crampton et al., 2016, Simpson, 1944, 1953, Smits, 2015, Van
62 Valen, 1973, 1979). I adopt a hierarchical Bayesian survival modeling approach,
which represents both a conceptual and statistical unification of the
64 paleontological dynamic and cohort survival analytic approaches (Baumiller,

1993, Crampton et al., 2016, Ezard et al., 2012, Foote, 1988, Raup, 1975, 1978,
66 Simpson, 2006, Smits, 2015, Van Valen, 1973, 1979). By using a Bayesian
framework I am able to quantify the uncertainty inherent in the estimates of the
68 effects of biological traits on survival.

1.1 Factors affecting brachiopod survival

70 Geographic range is widely considered the most important taxon trait for
estimating differences in extinction risk at nearly all times, with large
72 geographic range associated with low extinction risk (Jablonski, 1986, 1987,
Jablonski and Roy, 2003, Payne and Finnegan, 2007), though Foote and Miller
74 (2013) find that this generalization does not hold in the Mesozoic. For the
Paleozoic, however, I expect this to hold true for the entire period analyzed.

76 Epicontinental seas are a shallow-marine environment where the ocean has
spread over the surface of a continental shelf with a depth typically less than
78 100m. In contrast, open-ocean coastline environments have much greater
variance in depth, do not cover the continental shelf, and can persist during
80 periods of low sea level. Because of this, it is strongly expected that taxa which
favor epicontinental seas would be at great risk during periods of low sea levels,
82 such as during glacial periods, when epicontinental seas are drained. During the
Paleozoic (approximately 541–252 My), epicontinental seas were widely spread
84 globally but declined over the Mesozoic (approximately 252–66 My) and have
nearly disappeared during the Cenozoic (approximately 66–0 My) as open-ocean
86 coastlines became the dominant shallow-marine setting (Johnson, 1974, Miller
and Foote, 2009, Peters, 2008).

88 Miller and Foote (2009) demonstrated that during several mass extinctions taxa
associated with open-ocean environments tend to have a greater extinction risk

90 than those taxa associated with epicontinental seas. During periods of
background extinction, however, they found no consistent difference between
92 taxa favoring either environment. These two environment types represent the
primary environmental dichotomy observed in ancient marine systems (Miller
94 and Foote, 2009, Peters, 2008, Sheehan, 2001). Given these findings, I predict
that as extinction risk increases, the extinction risk associated with favoring
96 open-ocean environments should generally increase.

Because environmental preference is defined as the continuum between
98 occurring exclusively in open-ocean environments versus epicontinental
environments, intermediate values are considered “generalists” in the sense that
100 they favor neither end member. A long standing hypothesis is that generalists or
unspecialized taxa will have greater survival than specialists (Baumiller, 1993,
102 Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson, 1944). Because
of this, the effect of environmental preference was modeled as a quadratic
104 function where a concave down relationship between preference and expected
duration indicates that generalists are favored over specialists end-members.

106 Body size, measured as shell length, is also considered as a potential trait that
influences extinction risk (Payne et al., 2014). Body size is a proxy for metabolic
108 activity and other correlated life history traits (Payne et al., 2014). Given no
strong evidence that body size effects extinction risk in brachiopods, the effect
110 is most likely small and could be either positive, negative, or even absent.

It is well known that, given the incompleteness of the fossil record, the observed
112 duration of a taxon is an underestimate of that taxon’s true duration
CITATIONS. Because of this, the concern is that a taxon’s observed duration
114 may reflect its relative sampling ability and not because of the covariates of
interest. In this case, for sampling ability to be a confounding factor in this
116 analysis there must be consistent relationship between sampling ability and

duration (i.e. greater sampling, longer duration or *vice versa*). If there is no
118 relationship, positive or negative, between sampling ability and duration than
interpretation can be made clearly; while observed durations are obviously
120 truncated true durations, a lack of a relationship would indicate that the
amount and form of this truncation is independent of the taxon's duration (e.g.
122 taxa with better sampling also happen to be longer lived).

2 Materials and Methods

124 2.1 Fossil occurrence information

The brachiopod dataset analyzed here was sourced from the Paleobiology
126 Database (<http://www.paleodb.org>) which was then filtered based on taxonomic
(Rhynchonellata), temporal (post-Cambrian Paleozoic), stratigraphic, and other
128 occurrence information used in this analysis. These filtering criteria are very
similar to those from Foote and Miller (2013) with an additional constraint of
130 being present in the body size data set from Payne et al. (2014). Epicontinental
versus open-ocean assignments for each fossil occurrence are partially based on
132 those from Miller and Foote (2009), with additional occurrences assigned
similarly (Miller and Foote, personal communication). Analyzed occurrences
134 were restricted to those with paleolatitude and paleolongitude coordinates,
assignment to either epicontinental or open-ocean environment, and belonging
136 to a genus present in the body size dataset (Payne et al., 2014).

Fossil occurrences were analyzed at the genus level which is common for
138 paleobiological, macroevolution, or macroecological studies of marine
invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2012, Kiessling
140 and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,

Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

142 While species diversity dynamics is frequently of much greater interest than
those of higher taxa (though see, for example, Foote 2014, Hoehn et al. 2015),
144 the nature of the fossil record makes accurate and precise taxonomic
assignments at the species level for all occurrences. In particular, the simplicity
146 of brachiopod external morphology and the quality of preservation makes it very
difficult for assignments below the genus level. As such, the choice to analyze
148 genera as opposed to species was in order to assure a minimum level of
confidence and accuracy in the data analyzed here.

150 Genus duration was calculated as the number of geologic stages from first
appearance to last appearance, inclusive. Durations were based on geologic
152 stages as opposed to millions of years because of the inherently discrete nature
of the fossil record; dates are not assigned to fossils themselves but instead
154 fossils are known from a geological interval which represents some temporal
range. Stages are effectively irreducible temporal intervals in which taxa may
156 occur. Genera with a last occurrence in or after Changhsingian stage were right
censored at the Changhsingian; genera with a duration of only one stage were
158 left censored (Klein and Moeschberger, 2003). The covariates used to model
genus duration were geographic range size (r), environmental preference (v, v^2),
160 body size (m), and sampling (s).

Geographic range was calculated using an occupancy approach. First, all
162 occurrences were projected onto an equal-area cylindrical map projection. Each
occurrence was then assigned to one of the cells from a 70×34 regular raster
164 grid placed on the map. Each grid cell represents approximately $250,000 \text{ km}^2$.
The map projection and regular lattice were made using shape files from
166 <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
2015). For each stage, the total number of occupied grid cells was calculated.

168 Then, for each genus, the number of grid cells occupied by that genus was
 calculated. Dividing the genus occupancy by the total occupancy gives the
 170 relative occupancy of that genus. Mean relative genus occupancy was then
 calculated as the mean of the per stage relative occupancies of that genus.

 172 Sampling was calcualted as the average “gap ratio” of all stages in which a
 taxon existed. The basis of the gap ratio are range-through taxa, which are taxa,
 174 for any stage t , that are present in stages $t - 1$ and $t + 1$ CITATION FOOTE
 RAUP ALROY. A gap is when a range-through taxon is not present in stage t .

 176 The gap ratio for a stage is the number of gaps divided by the total number of
 range-through taxa, and the gap ratio for an individual taxon is the average of
 178 all gap ratios for each stage the taxon is present (inclusive). Additional
 occurrence information for the late Cambrian and earliest Trassic were used to
 180 calculate the gap ratios of the first and last stages included in this analysis.

Environmental preference was defined as probability of observing the ratio of
 182 epicontinental occurrences to total occurrences ($\theta_i = e_i/E_i$) or greater given the
 background occurrence probability θ'_i as estimated from all other taxa occurring
 184 at the same time (e'_i/E'_i). This measure of environmental preference is
 expressed.

$$\begin{aligned}
 p(\theta'_i|e'_i, E'_i) &\propto \text{Beta}(e'_i, E'_i - e'_i)\text{Beta}(1, 1) \\
 &= \text{Beta}(e'_i + 1, E'_i - e'_i + 1)
 \end{aligned} \tag{1}$$

$$v = Pr(\theta_i > \theta'_i)$$

186 Body size, measured as shell length, was sourced directly from Payne et al.
 (2014).

188 Prior to analysis, geographic range was logit transformed and body size was
 natural-log transformed. All covariates were then standardized by subtracting
 190 the mean from all values and dividing by twice its standard deviation, which

follows Gelman and Hill (2007). This standardization means that all regression
192 coefficients are comparable as the expected change per 1-unit change in the
rescaled covariates. In total there are D covariates.

194 2.2 Analytical approach

Hierarchical modelling is a statistical approach which explicitly takes into
196 account the structure of the observed data in order to model both the within
and between group variance (Gelman et al., 2013, Gelman and Hill, 2007). The
198 units of study (e.g. genera) each belong to a single group (e.g. origination
cohort). Each group is considered a draw from a shared probability distribution
200 of all cohorts, observed and unobserved. The group-level parameters are then
estimated simultaneously as the other parameters of interest (e.g. covariate
202 effects) (Gelman et al., 2013). The subsequent estimates are partially pooled
together, where parameters from groups with large samples or effects remain
204 large while those of groups with small samples or effects are pulled towards the
overall group mean. All covariate effects (regression coefficients), as well as the
206 intercept term (baseline extinction risk), were allowed to vary by group
(origination cohort). The covariance between covariate effects was also modeled.

208 Genus durations were assumed to follow a Weibull distribution which allows for
age-dependent extinction (Klein and Moeschberger, 2003): $y \sim \text{Weibull}(\alpha, \sigma)$.
210 The Weibull distribution has two parameters: scale σ , and shape α . When
212 $\alpha = 1$, σ is equal to the expected duration of any taxon. α is a measure of the
effect of age on extinction risk where values greater than 1 indicate that
extinction risk increases with age, and values less than 1 indicate that
214 extinction risk decreases with age. Note that the Weibull distribution is
equivalent to the exponential distribution when $\alpha = 1$.

²¹⁶ The scale parameter σ was modeled as a regression following Kleinbaum and Klein (2005) with both varying intercept and varying slopes; this is expressed

$$\sigma_i = \exp\left(\frac{-\mathbf{X}_i B_{j[i]}}{\alpha_{j[i]}}\right) \quad (2)$$

²¹⁸ where i indexes across all observations, $j[i]$ is the cohort membership of the i th observation, X is a $N \times (D + 1)$ matrix of covariates along with a column of 1's for the intercept term, and B is a $J \times D$ matrix of cohort-specific regression coefficients.

²²² Each of the rows of matrix B are modeled to be realizations from a multivariate normal distribution with length D location vector μ and $J \times J$ covariance matrix Σ : $B_j \sim \text{MVN}(\mu, \Sigma)$. The covariance matrix was then decomposed into a length J vector of scales τ and a $J \times J$ correlation matrix Ω , defined

$$\Sigma = \text{diag}(\tau)\Omega\text{diag}(\tau) \quad (3)$$

²²⁶ where “diag” indicates a diagonal matrix.

The elements of μ were given independent normally distributed priors. The effects of geographic range size and the breadth of environmental preference were given informative priors reflecting the previous findings while the others were given weakly informative favoring that covariate having no effect. The correlation matrix Ω was given an almost uniform LKJ distributed prior

²³² following CITATION STAN manual. These priors are defined

$$\begin{aligned}\mu_0 &\sim \mathcal{N}(0, 5) \\ \mu_r &\sim \mathcal{N}(-1, 1) \\ \mu_v &\sim \mathcal{N}(0, 1) \\ \mu_{v^2} &\sim \mathcal{N}(1, 1) \\ \mu_m &\sim \mathcal{N}(0, 1) \\ \mu_s &\sim \mathcal{N}(0, 1) \\ \tau &\sim \text{C}^+(1) \\ \Omega &\sim \text{LKJ}(2).\end{aligned}\tag{4}$$

The shape parameter α was allowed to vary by cohort where

²³⁴ $\log(\alpha_j) \sim \mathcal{N}(\alpha' + a_j, \sigma_\alpha)$. The effect of origination cohort a_j is modeled a
shared as normal distribution such that $a_j \sim \mathcal{N}(0, \sigma_a)$. α , σ_α and σ_a were given
²³⁶ weakly informative normal and half-Cauchy priors defined

$$\begin{aligned}\alpha' &\sim \mathcal{N}(0, 1) \\ \sigma_a &\sim \text{C}^+(1) \\ \sigma_\alpha &\sim \text{C}^+(1).\end{aligned}\tag{5}$$

The joint posterior was approximated using a Markov-chain Monte Carlo
²³⁸ routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
²⁴⁰ programming language Stan (Stan Development Team, 2014). The posterior
distribution was approximated from four parallel chains run for 10,000 steps
²⁴² each, split half warm-up and half sampling and thinned to every 10th sample for
a total of 4000 posterior samples. Chain convergence was assessed via the scale

²⁴⁴ reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate convergence. Convergence means that the chains are approximately stationary
²⁴⁶ and the samples are well mixed (Gelman et al., 2013).

²⁴⁸ The fit of the above model (the “full” model) was compared to the fits of three other sub-models: constant α across cohorts, no sampling as a covariate, or both
²⁵⁰ constant α and no sampling covariate. These models were compared for predicted out-of-sample predictive accuracy using both the widely-applicable information criterion (WAIC) and leave-one-out cross-validation estimated via
²⁵² Pareto-smoothed importance sampling (PSIS-LOO) CITATIONS. Both of these are estimates of the out-of-sample predictive accuracy or the expected quality of
²⁵⁴ fit of the model to new data.

²⁵⁶ WAIC is a more fully Bayesian alternative to AIC or DIC (Gelman et al., 2013, Watanabe, 2010); comparisons of WAIC values are useful for better understanding the effect of model complexity on out-of-sample predictive
²⁵⁸ accuracy. The calculation of WAIC used here corresponds to the “WAIC 2” formulation recommended by Gelman et al. (2013). Lower values of WAIC
²⁶⁰ indicate greater expected out-of-sample predictive accuracy than higher values.

²⁶² PSIS-LOO is similar to WAIC in that it is an approximation of out-of-sample predictive accuracy except its calculation is completely different CITATION GELMAN VEHTARI GABRY. Models comparison is done using a
²⁶⁴ leave-one-out crossvalidation information criterion (LOOIC), which is simply the PSIS-LOO estimate multiplied by -2 so that it is on the deviance scale. As
²⁶⁶ with WAIC, models with lower values are expected to have a greater out-of-sample predictive accuracy than models with greater values.

²⁶⁸ Calculations of WAIC and PSIS-LOO for a model fit using Stan were done using the R package “loo” CITATION PACKAGE. See VEHTARI GELMAN GABRY

270 for detailed explanations of the calucations for both WAIC and PSIS-LOO.

Model adequacy was evaluated using a couple of posterior predictive checks.

272 Posterior predictive checks are a means for understanding model for or
274 adequacy where the basic idea is that replicated data sets simulated from the
fitted model should be similar to the original data and systematic differences
between the simulations and observations indicate weaknesses of the model fit
276 (Gelman et al., 2013). For both approaches used here, 1000 posterior predictive
datasets were generated from 1000 unique draws from the posterior distribution
278 of each parameter. The two posterior predictive checks used in this analysis are
a comparison of a non-parameteric estimate of the survival function $S(t)$ from
280 the empirical dataset to the non-parameteric estimates of $S(t)$ from the 1000
posterior predictive datasets, and comparison of the observed genus durations to
282 the average posterior predictive estimate of $\log(\sigma_i)$ (Eq. 2). The former is to see
if simulated data has a similar survival pattern to the observed, and the latter is
284 to see if the model systematically over- or under- estimates taxon survival.

3 Results

286 The model with greatest estimated out-of-sample predictive accuracy, as
estimted via both LOOIC and WAIC, has both constant α and no sampling
288 covariate (Table 1). Because of this, only the results from this model will be
presented here. The effects of either of these potential model aspects are
290 inferred to be too small to account for the additional model complexity
necessary to estimate them.

292 Comparison of the distribution of posterior predictive estimates of $S(t)$ to the
empirical estimate reveal few obvious biases except the case of estimates of the
294 far right tail of observed durations (Fig. 1). This result is supported by the

Table 1: Widely applicable information criterion (WAIC) and leave-one-out cross-validation information criterion (LOOIC) values for the four models compared in this analysis. The WAIC and LOOIC values of the “best” model are in bold.

model	WAIC	LOOIC
constant α	9080.39	4823.13
constant α , no sampling	4676.84	4621.95
no sampling	8250.22	5563.97
full model	759172.46	109417.64

additional posterior predictive comparison where most estimates are not
 systematically biased, though there is a consistent under-estimate of $\log(\sigma)$ for
 older taxa (Fig. 2). The results of both posterior predictive checks indicate that,
 for the majority of observations, model fit is generally not biased towards over-
 or under-estimates of duration.

The cohort-level estimate of the effect of geographic range size indicates that as
 a taxon’s geographic range increases, that taxon’s duration is expected to
 increase (Table 2). Given the estimates of μ_r and τ_r , there is a less than 0.008%
 (± 0.05) probability that this relationships would be reversed
 $(\Pr(\mathcal{N}(\mu_r, \tau_r) > 0))$. The between cohort variance in the cohort-specific
 estimates of the effect of geographic range β_r are the lowest of all the regression
 coefficients (Table 2).

Body size is estimated to have no effect on taxon duration, with the estimate
 nearly centered at 0 (Table 2). The variance between the cohort-specific
 estimates of the effect of body size τ_m is found to be greater than the variance
 of between-cohort estimates of the effect of geographic range size τ_r .

The group-level estimate of the effect of environmental preference is estimated
 from both μ_v and μ_{v^2} .

The estimate of μ_v indicates that epicontinental favoring taxa are expected to
 have a greater duration than open-ocean favoring taxa. Additionally, given the

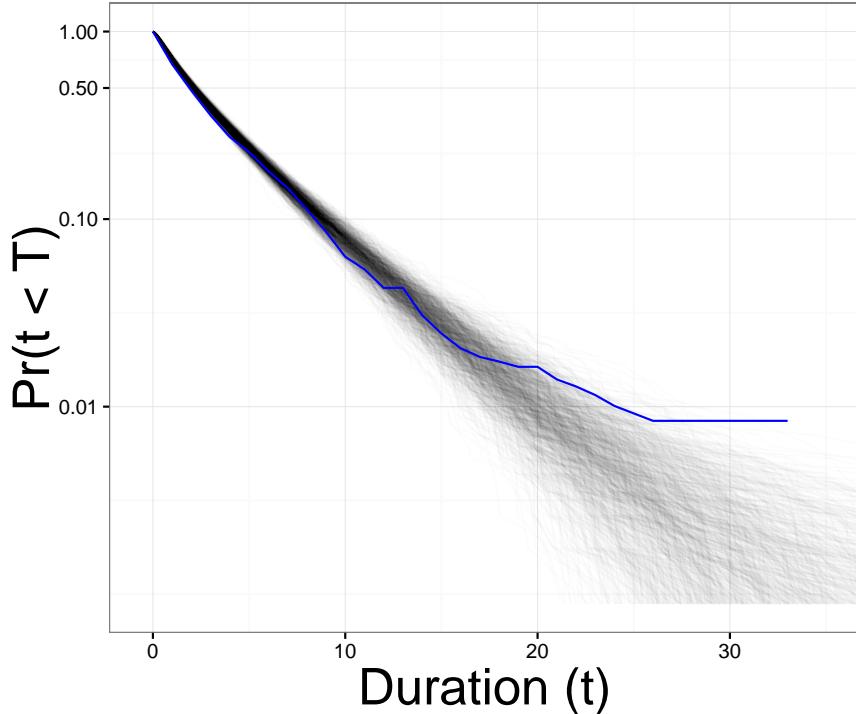


Figure 1: Comparison of the empirical estimate of $S(t)$ (highlighted) versus estimates from 1000 posterior predictive data sets (black). $S(t)$ corresponds to the probability that the age of a genus t is less than the genus' ultimate duration T . The vertical axis is log10 transformed.

estimate of between cohort variance τ_v , there is approximately 19% ($\pm 8SD$)
³¹⁶ probability that taxa favoring open-ocean environments would have a greater
 expected duration than taxa favoring epicontinental environments
³¹⁸ ($\Pr(\mathcal{N}(\mu_v, \tau_v) > 0)$). Notice the direction of the inequality given the negative
 size in equation 2. There is a high amount amount of between-cohort variation
³²⁰ in estimates of β_v (Table 2).

The estimate of μ_{v^2} indicates that the overall relationship between
³²² environmental preference and $\log(\sigma)$ is concave down (Fig. 3), with only a 1.9%
 (± 2.4) probability that any given cohort is not concave down
³²⁴ ($\Pr(\mathcal{N}(\mu_{v^2}, \tau_{v^2}) < 0)$). As above, notice the direction of the inequality given the

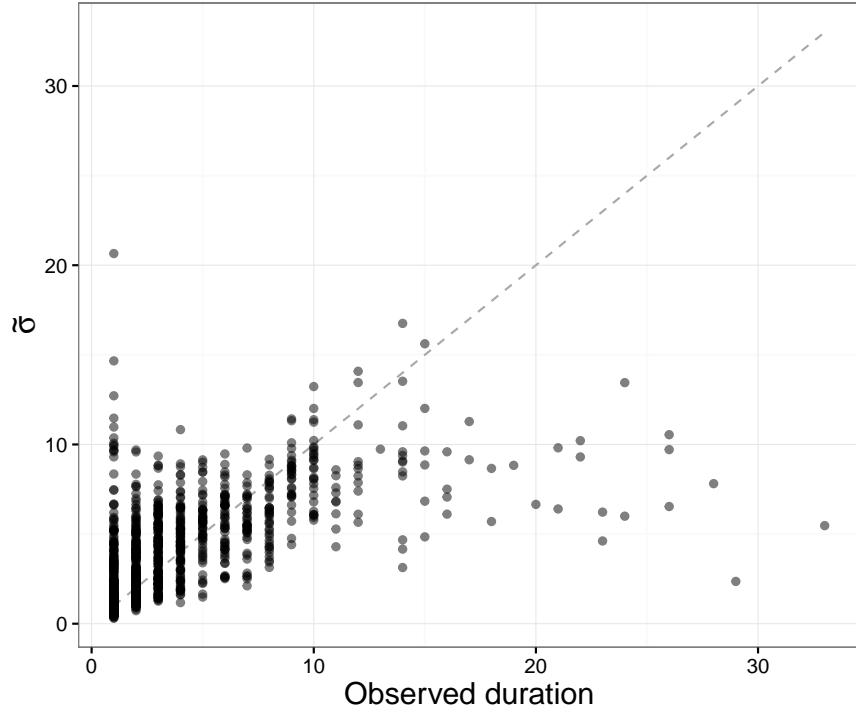


Figure 2: Comparison of all observed genus durations in number of geological stages to the average posterior predictive estimates of $\log(\sigma)$. The dashed, diagonal line corresponds to $x = y$.

negative size in equation 2. Given the estimate of τ_{v^2} , there is an expected high
 326 amount amount of between-cohort variation in estimates of β_{v^2} (Table 2).

The cohort-specific estimates of all the regression coefficients demonstrate a lot
 328 of cohort to cohort variance, with no obvious trends. As indicated in Table 2
 and detectable visually (Fig. 4), the between-cohort estimates for β_0 , β_r , and
 330 β_m all have much lower variance than the between-cohort estimates of both β_v
 and β_{v^2} .

332 While most cohort-specific estimates are very similar to the overall cohort-level
 estimate there are a few notable excursions away from the overall mean (Fig. 4).
 334 There are simultaneous excursions in both β_0 and β_v for cohorts originating in

Table 2: Group-level estimates of the effects of biological traits on brachiopod generic survival. μ values are the location parameters of the effects, while τ values are the scale terms describing the variation between cohorts. The mean, standard deviation, 10th, 50th, and 90th quantiles of the posterior are presented.

parameter	mean	standard deviation	10%	50%	90%
μ_i	-2.32	0.14	-2.50	-2.32	-2.15
μ_r	-0.76	0.11	-0.91	-0.76	-0.62
μ_v	-0.66	0.17	-0.88	-0.66	-0.43
μ_{v^2}	2.88	0.31	2.48	2.88	3.27
μ_m	0.04	0.12	-0.12	0.04	0.19
τ_i	0.50	0.10	0.37	0.49	0.63
τ_r	0.27	0.13	0.11	0.26	0.45
τ_v	0.76	0.16	0.56	0.74	0.97
τ_{v^2}	1.24	0.33	0.84	1.21	1.67
τ_m	0.47	0.12	0.33	0.47	0.63

the Givetian (387-382 My) and Frasnian (382-372 My) stages; both of which

³³⁶ directly precede the end-Devonian mass extinction event at the

Frasnian/Famennian boundary. These excursions indicate notably high

³³⁸ extinction intensity faced by these cohorts along with an increase in expected duration for taxa favoring epicontinental environments over open-ocean ones.

³⁴⁰ Cohorts originating from the Silurian through the Early Devonian have a

slightly lower extinction intensity than the overall mean; these are the from the
³⁴² Llandovery (443-443 My) through the Emsian (407-393 My). This is also a time period where there is the lowest probability that epicontinental favoring taxa

³⁴⁴ expected to have greater duration than open-ocean favoring taxa. Both the

Silurian and Devonian periods are notable for having been periods with a mostly
³⁴⁶ “hothouse” climate, with no polar icecaps and a high sea-level CITATION.

The cohort-specific relationships between environmental preference and $\log(\sigma)$

³⁴⁸ were calculated from the estimates of β_0 , β_v , and β_{v^2} (Fig. 5) and reflect how

these three parameters act in concert and not just individually (Fig. 4). Beyond

³⁵⁰ results already discussed above in the context of individual parameters, it is

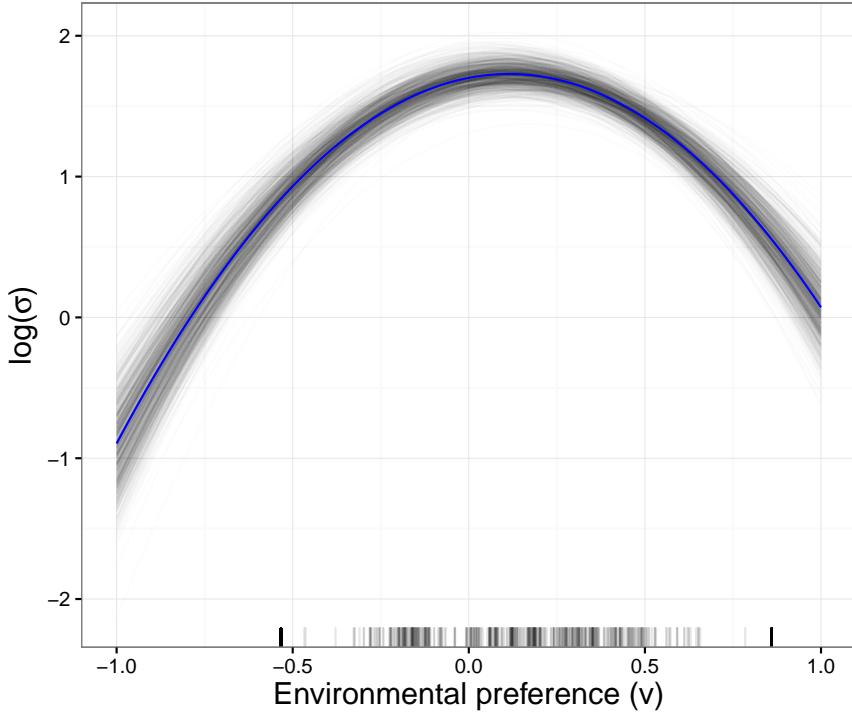


Figure 3: The overall expected relationship between environmental affinity v_i and a $\log(\sigma)$ when $r = 0$ and $m = 0$. Each grey line corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior predictive distribution. The overall relationship is concave down with an optimum greater than 0, which means that taxa favoring epicontinental environments are expected to have longer durations than those favoring open-ocean environments.

notable that the cohort originating in the Kungurian (279-272 My) has the most
 352 sharply curved appearance due to a high estimate β_{v^2} (Fig. 4). This cohort
 expected to have the biggest different in extinction risk between environmtal
 354 generalists and specialists. The cohorts originating during the Emsian (407-393
 My) and Frasnian (382 - 372 My) are tied for second for sharpness of curvature.
 356 The least sharply curved cohorts include those orignating during Tremadocian
 (484-477 My), Hirnantian (445-443 My), Llandovery (443-433 My), and Ludlow
 358 (427-423 My). Except for the Tremadocian cohort, all of these cohorts originate

during the Silurian through the Early Devonian range identified earlier as
360 having lower expected extinction intensity than what is expected from the group-level estimate.

362 The correlation in between cohort-specific estimates of the regression coefficients are estimated as the off-diagonal elements of the correlation matrix Ω . Only two
364 of the elements of Ω are distinguishable from 0: the correlation of β_0 (extinction intensity) with both β_r and β_v (Fig. 6).

366 There is an approximate 86% probability that the cohort-specific estimates of baseline extinction intensity β_0 and the effect of geographic range β_r are
368 negatively correlated; this means that it is expected that for cohorts experiencing a lower extinction intensity the effect of geographic range increases,
370 and *vice versa*.

Similarly, there is an approximate 99.9% probability that the cohort-specific
372 estimates of β_0 and β_v are negatively correlated; this means that as extinction intensity increases it is expected that epicontinental taxa become less favored
374 over open-ocean environments. There is a 19% ($\pm 8.1SD$) probability that, for any given cohort, open-ocean environments will be favored.

376 Estimates of β_r and β_v are themselves not correlated, as there is only an
approximate 68% probability of a positive correlation. The lack of
378 cross-correlation may be caused by the much higher between-cohort variance of the effect of environmental preference τ_v than the very small between-cohort
380 variance in the effect of geographic range τ_r (Table 2). The effect of geographic range might simply not vary enough relative to the much noisier environmental
382 preference to detect a correlation.

4 Discussion

³⁸⁴ The generating observation behind this study was that the only biological trait
that was found to affect extinction risk was geographic range while traits that
³⁸⁶ had previously been beneficial had no effect (Jablonski, 1986). This observation
raises two linked questions: how does the effect of geographic range change with
³⁸⁸ changing extinction intensity, and how does the effect of other biological traits
change with changing extinction intensity?

³⁹⁰ I find that as intensity increases, the effect of geographic range is expected to
decreases. I also find that as intensity increases, the effect of environmental
³⁹² preference is expected to be smaller. There is no evidence for a correlation
between the effect of geographic range and environmental preference. However,
³⁹⁴ the between-cohort variance in effect of geographic range is much less than the
between-cohort variance of the effect of environmental preference. As stated
³⁹⁶ above, this may underlie the lack of correlation between these two effects.

Additionally, the lower between-cohort variance in the effect of geographic range
³⁹⁸ versus that higher between-cohort variance implies that for cohorts with a
greater extinction intensity than the average cohort, the difference in the effect
⁴⁰⁰ geographic range and the group-level effect of geographic range will be smaller
than the difference between the effect of environmental preference and the
⁴⁰² group-level effect of environmental preference.

I find consistent support for the “survival of the unspecialized,” with respect for
⁴⁰⁴ environmental preference, as a time-invariant generalization of brachiopod
survival; taxa with intermediate environmental preferences were found to have
⁴⁰⁶ lower extinction risk than taxa specializing in either epicontinental or
open-ocean environments (Fig. 3), though the curvature of the relationship
⁴⁰⁸ varies from nearly flat to very peaked (Fig. 5). However, this relationship is not

symetric about 0 as taxa favoring epicontinental environments are expected to
410 have a greater duration than taxa favoring open-ocean environments.

Hopkins et al. (2014), in their analysis of niche conservatism and substrate
412 preference in marine invertebrates, found that brachiopods were among the
least “conservative” groups; taxa were found change preference on short time
414 scales. While substrate preference is not the same as environmental preference,
a question arises: do “generalists” have a greater duration because they are
416 “true” environmental generalists from orgination (e.g. can inhabit both
environments) or have the taxa with longer durations evolved in preference over
418 their duration in the face of changing envrionments (i.e. niche evolution and
adaptation)? The nature of the analysis done here assumes that traits are
420 constant over the duration of the taxon; future work should explore how
environmental preference changes over lineage duration.

422 In this study I used the average quality of the record for the duration of a taxon
as a measure of the effect of sampling; the idea being to determine if taxa with
424 shorter or longer durations are associated with differences in sampling regime.
This covariate, however, is not included in the WAIC/LOOIC best model (Table
426 1). I conclude then that the effect of sampling on duration is at most small and
is not acting as a confounding factor.

428 The Cormack-Jolly-Seber (CJS) model is an alternative approach for modeling
survival that can take into account imperfect observation CITATION. This
430 model is a type of hidden Markov model with an absorbing state (i.e.
extinction). In this model, survival is defined as the probability of surviving
432 from time t to time $t + 1$. Additionally, the effect of preservation and sighting is
estimated as probability of observing a taxon that is present; this can extend
434 the duration of a taxon beyond its last occurrence. This approach is a very
different from the method used here; instead of modeling persistance from time

⁴³⁶ t to time $t + 1$, I model the distribution taxon durations (lifetime) which then
⁴³⁷ decomposes into a function describing survival ($S(t)$) and a function describing
⁴³⁸ instantaneous extinction risk CITATION.

⁴⁴⁰ The use of genera as the unit of the study and how to exactly interpret the
⁴⁴¹ effects of the biological traits is a remaining concern. For example, if any of the
⁴⁴² traits analyzed here are associated with increases in speciation rates, this might
⁴⁴³ increase the duration of genera through self-renewal (Raup, 1991, 1994), which
⁴⁴⁴ would be an example of the difference in biological pattern between species and
⁴⁴⁵ genera (Jablonski, 1987, 2007, 2008). This could lead to a trait appearing to
⁴⁴⁶ decrease generic level extinction risk by increasing species level origination rate
⁴⁴⁷ instead of decreasing species level extinction risk. However, given the nature of
⁴⁴⁸ the fossil record and maintaining a minimum level of data consistency/quality,
⁴⁴⁹ there is no simple solution to decreasing this uncertainty in the interpretations
⁴⁵⁰ of how the biological traits studied at the genus-level may translate to the
species-level.

⁴⁵² The model used here could be improved through either increasing the number of
⁴⁵³ analyzed taxon traits, expanding the hierarchical structure of the model to
⁴⁵⁴ include other major taxonomic groups of interest, and the inclusion of explicit
phylogenetic relationships between the taxa in the model as an additional
⁴⁵⁵ hierarchical effect. An example taxon trait that may be of particular interest is
⁴⁵⁶ the affixing strategy or method of interaction with the substrate of the taxon,
⁴⁵⁷ which has been found to be related to brachiopod survival where, for
⁴⁵⁸ cosmopolitan taxa, taxa that are attached to the substrate are expected to have
a greater duration than those that are not (Alexander, 1977).

⁴⁶⁰ It is theoretically possible to expand this model to allow for comparisons within
and between major taxonomic groups. This approach would better constrain the
⁴⁶² brachiopod estimates while also allowing for estimation of similarities and

differences in cross-taxonomic patterns. The major issue surrounding this
464 particular expansion involves finding an similarly well sampled taxonomic group
that is present during the Paleozoic. Example groups include Crinoidea,
466 Ostracoda, and other members of the “Paleozoic fauna” (Sepkoski Jr., 1981).

Taxon traits like environmental preference or geographic range (Hunt et al.,
468 2005, Jablonski, 1987) are most likely heritable, at least phylogenetically
(Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this
470 analysis assumes that differences in extinction risk between taxa are
independent of the shared evolutionary history of those taxa (Felsenstein, 1985).
472 In contrast, the origination cohorts only capture shared temporal context. For
example, if taxon duration is phylogenetically heritable, then closely related
474 taxa may have more similar durations as well as more similar biological traits.
Without taking into account phylogenetic similarity the effects of these
476 biological traits would be inflated solely due to inheritance. The inclusion of
phylogenetic context as an additional individual-level hierarchical effect
478 independent of origination cohort would allow for determining how much of the
observed variability is due to shared evolutionary history versus shared temporal
480 context versus actual differences associated with biological traits (Smits, 2015).

Acknowledgements

482 I would like to thank K. Angielczyk, M. Foote, P. D. Polly, R. Ree, and G.
Slater for helpful discussion during the conception of this study. I'd also like to
484 thank D. Bapst, N. Pierrehumbert and M. Villarosa Garcia for additional
comments. Additionally, thank you A. Miller for the epicontinental versus
486 open-ocean assignments. This entire study would not have been possible
without the Herculean effort of the many contributors to the Paleobiology

⁴⁸⁸ Database. In particular, I would like to thank J. Alroy, M. Aberhan, D. Bottjer,
⁴⁸⁹ M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland, L. Ivany, W. Kiessling,
⁴⁹⁰ B. Kröger, A. McGowan, T. Olszewski, P. Novack-Gottshall, M. Patzkowsky, M.
Uhen, L. Villier, and P. Wager. This work was supported by a NASA
⁴⁹² Exobiology grant (NNX10AQ446) to A. Miller and M. Foote. I declare no
conflicts of interest. This is Paleobiology Database publication XXX.

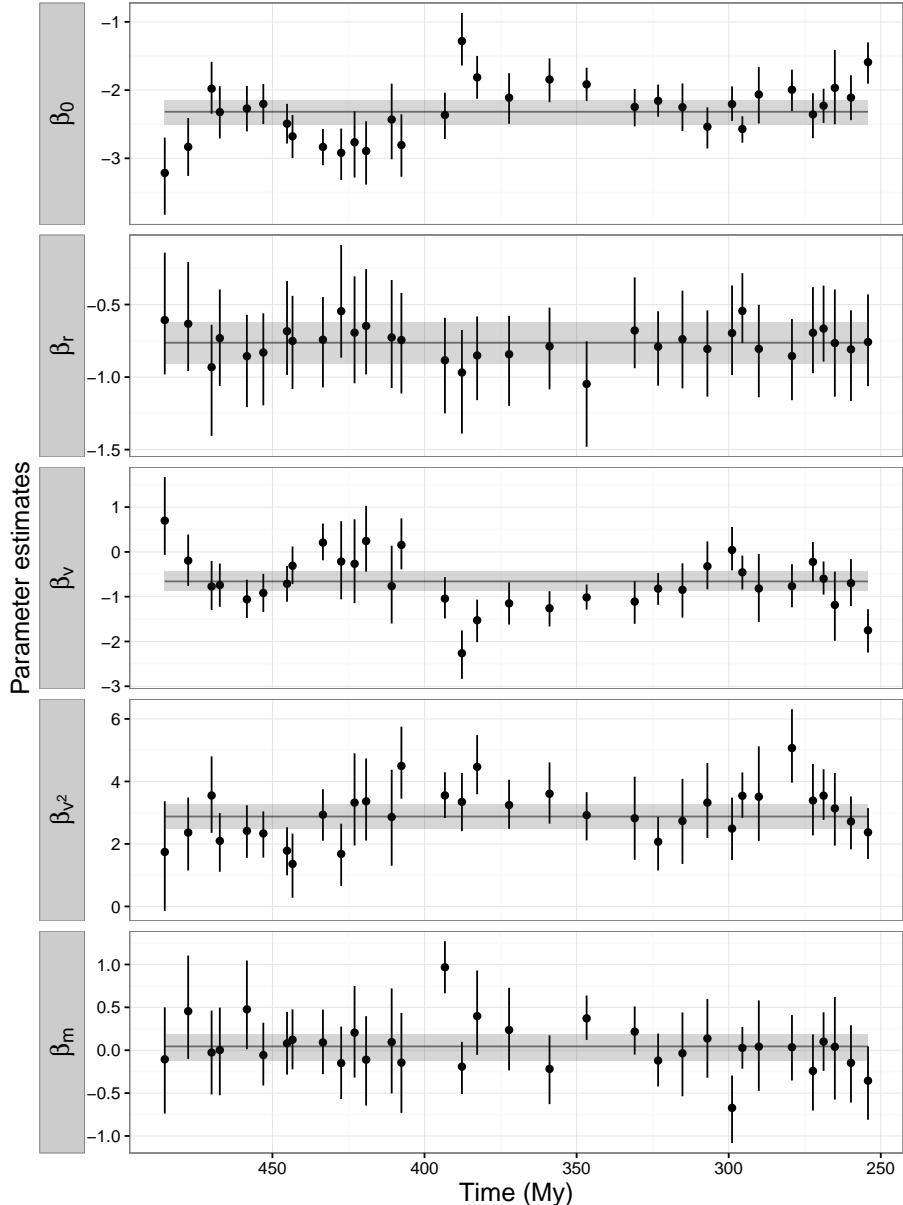


Figure 4: Comparison of cohort-specific estimates of β_0 , the effect of geographic range on extinction risk β_r , the effect of environmental preference β_v and β_{v^2} , and body size β_m . Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. Points are plotted at the midpoint of the cohorts stage of origination in millions of years before present (My). Black, horizontal lines are the overall estimates of covariate effects along with 80% credible intervals (shaded).

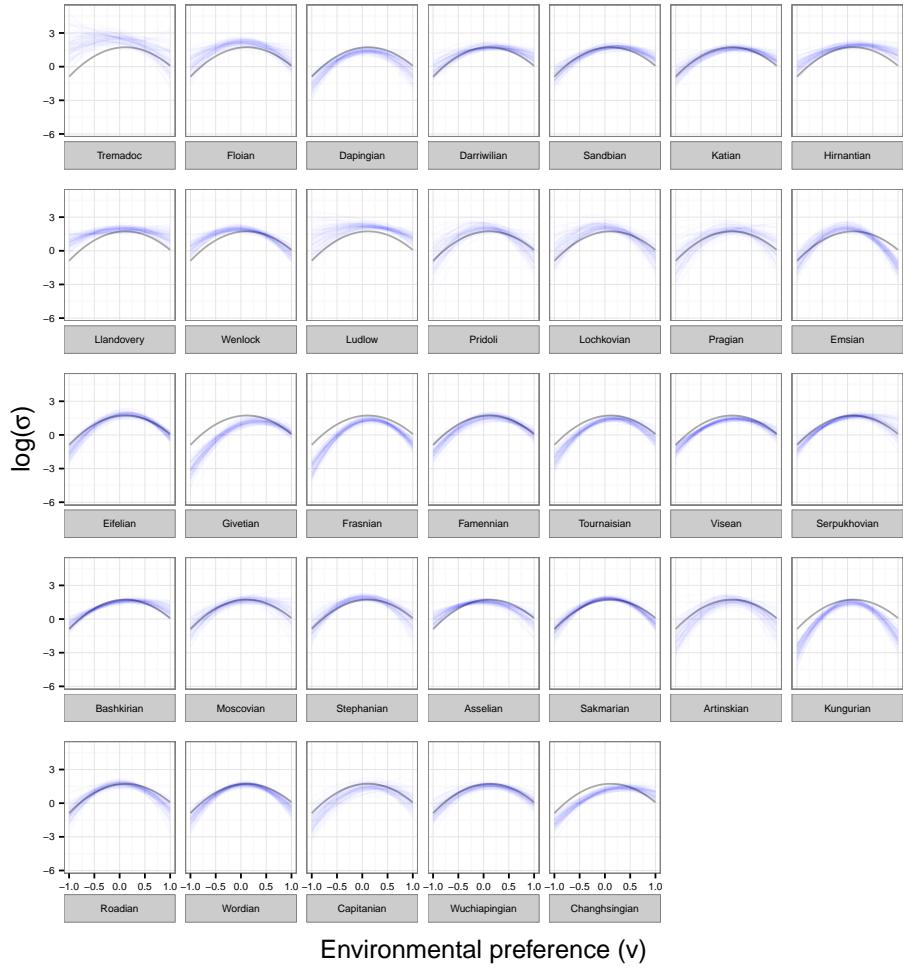


Figure 5: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on $\log(\sigma)$ to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range of (transformed and rescaled) observed values of environmental preference. The oldest cohort is in the top-right and younger cohorts proceed left to right, with the youngest cohort being the right-most facet of the last row. Facet names correspond to the name of the stage in which that cohort originated.

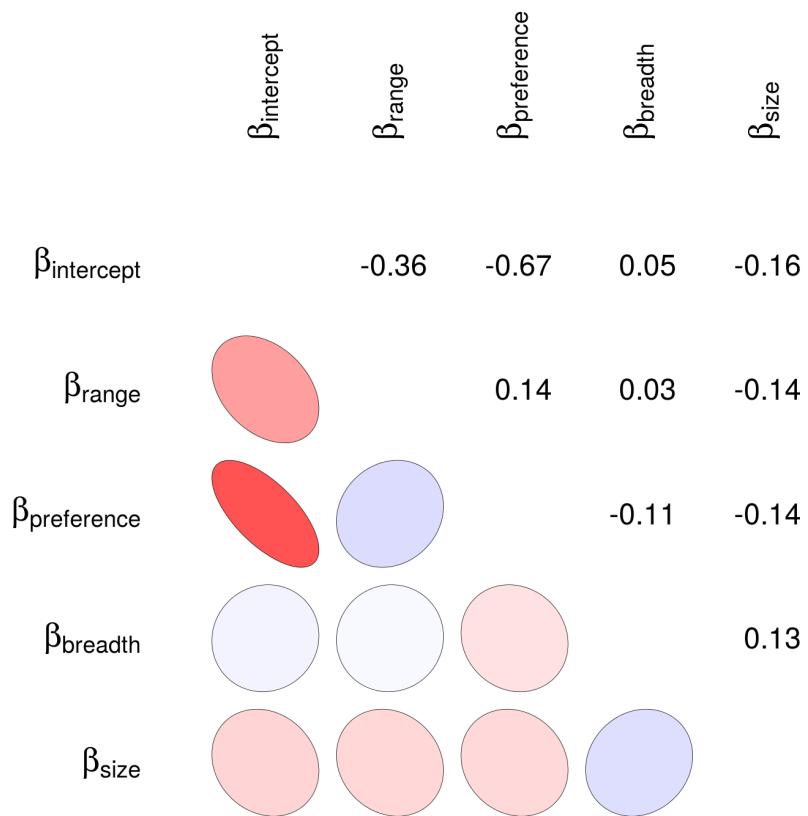


Figure 6: Mixed graphical and numerical representation of the correlation matrix Ω of variation in cohort-specific covariate estimates. These correlations are between the estimates of the cohort-level effects of covariates, along with intercept/baseline extinction risk. The median estimates of the correlations are presented numerically (upper-triangle) and as idealized ellipses representing that much correlation (lower-triangle).

References

- Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
496 to the mode of stabilization on the substrate. *Palaeogeography, Palaeoclimatology, Palaeoecology* 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal Groups. *Science* 329:1191–1194.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of
500 filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- Cooper, W. S., 1984. Expected time to extinction and the concept of fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- Crampton, J. S., R. A. Cooper, P. M. Sadler, and M. Foote, 2016. Greenhouseicehouse transition in the Late Ordovician marks a step change in
504 extinction regime in the marine plankton. *Proceedings of the National Academy of Sciences P*. 201519092.
- Ezard, T. H. G., P. N. Pearson, T. Aze, and A. Purvis, 2012. The meaning of birth and death (in macroevolutionary birth-death models). *Biology letters*
508 8:139–42.
- Felsenstein, J., 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites.
512 *Paleobiology* 14:258–271.
- , 2000. Origination and extinction components of taxonomic diversity:
516 Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.

- , 2014. Environmental controls on geographic range size in marine animal genera. *Paleobiology* 40:440–458.
- Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine animal genera. *Paleobiology* 39:171–192.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013. Bayesian data analysis. 3 ed. Chapman and Hall, Boca Raton, FL.
- Gelman, A. and J. Hill, 2007. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Harnik, P. G., C. Simpson, and J. L. Payne, 2012. Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences* 279:4969–4976.
- Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL <http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- Hoehn, K. B., P. G. Harnik, and V. L. Roth, 2015. A framework for detecting natural selection on traits above the species level. *Methods in Ecology and Evolution* .
- Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15:1351–1381.
- Hopkins, M. J., C. Simpson, and W. Kiessling, 2014. Differential niche dynamics among major marine invertebrate clades. *Ecology Letters* 17:314–323.
- Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed Model. *The American Naturalist* 163:84–96.

- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
542 comment on "On the heritability of geographic range sizes". American
Naturalist 166:129–135.
- 544 Jablonski, D., 1986. Background and mass extinctions: the alternation of
macroevolutionary regimes. Science 231:129–133.
- 546 ———, 1987. Heritability at the species level: analysis of geographic ranges of
cretaceous mollusks. Science 238:360–363.
- 548 ———, 2007. Scale and hierarchy in macroevolution. Palaeontology 50:87–109.
- , 2008. Species Selection: Theory and Data. Annual Review of Ecology,
550 Evolution, and Systematics 39:501–524.
- Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
552 living molluscs. Proceedings. Biological sciences / The Royal Society
270:401–6.
- 554 Johnson, J. G., 1974. Extinction of Perched Faunas. Geology 2:479–482.
- Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
556 benthic biodiversity dynamics through Triassic Jurassic time. Paleobiology
33:414–434.
- 558 Klein, J. P. and M. L. Moeschberger, 2003. Survival Analysis: Techniques for
Censored and Truncated Data. 2nd ed. Springer, New York.
- 560 Kleinbaum, D. G. and M. Klein, 2005. Survival analysis: a self-learning text. 2
ed. Springer, New York, NY.
- 562 Liow, L. H., 2004. A test of Simpson's "rule of the survival of the relatively
unspecialized" using fossil crinoids. The American naturalist 164:431–43.
- 564 ———, 2007. Does versatility as measured by geographic range, bathymetric

- range and morphological variability contribute to taxon longevity? Global
566 Ecology and Biogeography 16:117–128.
- Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
568 biology. Evolution 45:1065–1080.
- Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
570 the kinetics of mass extinction and origination. Science 326:1106–9.
- Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
572 predict diversity dynamics in marine Mesozoic bivalves. Paleobiology
39:360–372.
- 574 ———, 2015. Interdependence of specialization and biodiversity in Phanerozoic
marine invertebrates. Nature communications 6:6602.
- 576 Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
than evolvability. PloS one 7:e38025.
- 578 Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on
extinction risk during background and mass extinction. Proceedings of the
580 National Academy of Sciences 104:10506–11.
- Payne, J. L., N. A. Heim, M. L. Knope, and C. R. McClain, 2014. Metabolic
582 dominance of bivalves predates brachiopod diversity decline by more than 150
million years. Proceedings of the Royal Society B 281:20133122.
- 584 Peters, S. E., 2008. Environmental determinants of extinction selectivity in the
fossil record. Nature 454:626–9.
- 586 Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law.
Paleobiology 1:82–96.
- 588 ———, 1978. Cohort Analysis of generic survivorship. Paleobiology 4:1–15.

- _____, 1991. *Extinction: Bad Genes or Bad Luck?* Norton, New York.
- 590 _____, 1994. The role of extinction in evolution. *Proceedings of the National Academy of Sciences* 91:6758–6763.
- 592 Sepkoski Jr., J. J., 1981. A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology* 7:36–53.
- 594 Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of Earth and Planetary Sciences* 29:331–364.
- 596 Simpson, C., 2006. Levels of selection and large-scale morphological trends. Ph.D. thesis, University of Chicago.
- 598 Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- 600 Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York.
- 602 _____, 1953. *The Major Features of Evolution*. Columbia University Press, New York.
- 604 Smits, P. D., 2015. Expected time-invariant effects of biological traits on mammal species duration. *Proceedings of the National Academy of Sciences* .
- 606 Stan Development Team, 2014. Stan: A c++ library for probability and sampling, version 2.5.0. URL <http://mc-stan.org/>.
- 608 Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- _____ , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- 610 Vilhena, D. A., E. B. Harris, C. T. Bergstrom, M. E. Maliska, P. D. Ward, C. A. Sidor, C. A. E. Strömberg, and G. P. Wilson, 2013. Bivalve network reveals

- 612 latitudinal selectivity gradient at the end-Cretaceous mass extinction.
Scientific Reports 3:1790.
- 614 Wang, S. C., 2003. On the continuity of background and mass extinction.
Paleobiology 29:455–467.
- 616 Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and
Widely Applicable Information Criterion in Singular Learning Theory.
- 618 Journal of Machine Learning Research 11:3571–3594.