

1 Materials and Methods

2 1.1 Geological unit information

All information used in this analysis is available freely through Macrostrat
4 `macrostrat.org` and the Paleobiological Database PBDB `paleobiodb.org`.
For this analysis, we used direct API calls to pull data from the relational
6 databases underpinning both Macrostrat and the PBDB (??); this means that
our analyses are inherently dynamic and can be instantly updated as these
8 databases continue to grow.

Macrostrat geological units that have some amount of their sediments within
10 the Ordovician or Silurian; this requires two API calls to Macrostrat, one for
each of the periods (e.g.
12 `macrostrat.org/api/v2/units?interval_name=Ordovician`). These data
frames were merged using a left-join with geological unit identification code
14 (`unit_id`) as the key value which prevents double counting of geologic units that
range through both periods.

16 The key value of geological unit metadata associated with this data frame are
the unique identifiers for each geological unit. These values can be used as the
18 key value for linking these geological units to the fossils found within those
units (e.g. `macrostrat.org/api/v2/fossils?unit_id=...`). This aspect of
20 the Macrostrat database includes information on the count of Paleobiology
Database collections drawn from that unit, the count of unique fossils listed in
22 the Paleobiology Database associated with that unit, the unique taxonomic
identifiers for each of those fossils, and the unique identifiers for each of the
24 collections drawn from that unit (`cltn.id`). This final value acts as the foreign
key for extracting fossil occurrence and taxonomic information from the

26 Paleobiology Database.

A final API call is made, this time to Paleobiology Database, using the
28 collection id foreign key from Macrostrat (e.g.
`paleobiodb.org/data1.2/occs/list.txt?coll_id=...`); this API call was
30 technically done as two calls and the resulting union of those data frames is a
data frame with all the metadata for all of the fossil occurrences in collections
32 drawn from geological units from the Ordovician and Silurian present in
Macrostrat.

34 For a more explicit and exacting description of the data gathering and
preparing process, please see the `./R/download_scrap.r` and `/R/prepare_data.r`
36 scripts from the project repository https://github.com/psmits/not_fossil.

The above series of API calls produces three data frames: Macrostrat geological
38 units and their metadata, fossil counts and collection information for those
Macrostrat geological units, and the unique fossils present in these units and
40 their metadata from the Paleobiology database.

An interesting feature of Macrostrat geologic units is that they are ordered
42 according to the underlying continuous-time age model (?). This age model
increases the overall resolution of the geological record. Unfortunately the fossil
44 collection information for each unit does not include within-unit superposition
data; this means that the diversity within a geologic unit cannot be tracked over
46 the duration of the unit but only as a function of the complete unit. Because of
this, the most precise unit of our analysis is the geologic unit. Specifically, we
48 assign each geologic units to a single temporal bins based on which bin
contained their midpoint. Macrostrat provides a top and bottom age and by
50 averaging those we get the midpoint age. In total, we divided the data into 20
uniform-duration discrete time intervals.

52 The geological unit metadata values that are relevant to this analysis include
areal extent of the unit (positive, real values), maximum unit thickness
54 (positive, real values), average paleo latitudinal position (real values), and the
lithological description of the unit. Lithology is expressed as one or more
56 natural language statements (e.g. siliclastic sedimentary) and the percentage of
the unit associated with that lithology.

58 Lithological description, being made of natural language statements, requires
some standardization and simplification to make it amenable to analysis.
60 Ultimately, we described lithology as some combination of fine siliciclastics,
coarse siliciclastics, dolomitic carbonates, and non-dolomitic carbonates. The
62 multi-step process to reduce the original natural language descriptions is
detailed here, but for the complete and explicit process the `strict.lithology`
64 function in `./R/rock_mung.r` script file from the project repository
https://github.com/psmits/not_fossil.

66 First, units that have at least one description including the words igneous,
volcanic, metamorphic, chemical, anhydrite, evaporite, or halite were removed
68 prior to analysis. After this step, descriptive terms were unified so that the
natural language descriptions are easier match together (e.g. green and greenish
70 become green, mudstone and mud become mudstone). Next, some additional
words were removed from descriptions for either being too general or too
72 specific (e.g. sedimentary, dark, etc.). Each of these choices are inherently
arbitrary means of simplifying text, but these ultimately have little to no effect
74 the final descriptions because our final step is extremely strict.

This final step in assigning the final lithological descriptions to simpler values
76 that are amenable to analysis. Descriptions were classified as “fine siliciclastics”
were those containing at least one of the terms siltstone, claystone, mudstone,
78 shale, and argillite. In contrast, descriptions were classified as “coarse

siliciclastics” if they were a siliciclastic lithology that did not include one of the
80 terms keyed to “fine siliciclastics”. Descriptions were classified as “dolomitic
carbonates” if those lithological descriptions contained the word “dolomite”.
82 Finally, non-dolomitic carbonates were all other carbonate lithologies.

Prior to analysis, real values were log transformed by subtracting the mean
84 value from all observations then dividing by twice the standard deviation of the
observations. Similarly, positive real valued covariates were log-plus-one
86 transformed and then rescaled in a similar manner. Rescaling the covariates has
multiple advantages: 1) regression coefficients now describe the expected change
88 in unit diversity per change in standard deviation of covariate, and 2) regression
coefficients are comparable across covariates because they are all on the same
90 scale (the expected standard deviation of a binary variable is 0.5) (?).

In contrast to the other covariations, compositional covariates are constrained
92 to sum to 1 which creates degrees-of-freedom issues when trying to model their
possible effects as including these covariates without appropriate transformation
94 creates two or more nonidentifiable parameters. To that end, the composition
variables were isometric log-ratio transformed (ilr) (?) which reduces the total
96 number of variables to one less than original as composition is defined in
relation to a baseline (percent non-dolomitic carbonates). Unfortunately, the
98 scale and interpretations of the associated regression coefficients are different
from the other covariates, making direct comparison tricky. The ilr
100 transformation was done using the `compositions` package for R CITATION.

The fossil occurrence and diversity for each of the geological units was
102 determined for each of the following taxonomic groups: Anthozoa, Brachiopoda,
Bivalvia, Cephalopoda, Gastropoda, and Trilobita. Fossil membership was
104 determined based on the metadata for phylum or class from the Paleobiology
database. We fit our model separately to each of these datasets.

1.2 Modeling of the fossil diversity found in a geologic units

All geologic units we're analyzing have at least one species occurrence associated with it; this explicit observation restriction means that instead of a full distribution of counts from 0 to positive infinity, we instead have a truncated distribution ranging from 1 to positive infinity.

A natural statistical distribution for discrete data is the Poisson distribution. The Poisson distribution makes strong assumptions about the mean-variance relationship of the data which is rarely found in life as data frequently has much larger variance in counts than the mean count; this variance is described as overdispersion as the data has a greater scale than expected from a Poisson distribution CITATION. TO model this potential overdispersion in the data, we opted for using the Negative Binomial distribution instead of the Poisson. The Negative Binomial distribution can be derived as a mixture a Gamma and a Poisson distribution where the Gamma accounts for increased variance.

The Negative Binomial distribution parameterized in terms of of mean or expected count μ and a description of the dispersion of the data ϕ is formulated as:

$$\text{Negative Binomial}(y|\mu, \phi) = \binom{y + \phi - 1}{y} \frac{\mu}{\mu + \phi} \frac{\phi}{\mu + \phi}^{\phi} . \quad (1)$$

We chose this parameterization of the Negative Binomial distribution because it has one of the simplest interpretations; the mean μ is the expected taxonomic diversity of a given geologic unit, and the amount of overdispersion in counts is equal to the inverse of ϕ scaled by the square of the mean μ . Our hierarchical/multi-level model can be characterized as a type of GLMM with varying-intercept and varying-slopes where the assumed data distribution is a zero-truncated Negative Binomial distribution and our regression uses a log-link

function. For a more detailed description of the Negative Binomial distribution
 132 and its use in count regression please see CITATION CITATION and
 CITATION.

134 The effects of the unit covariates are expressed as the regression coefficients β
 which were allowed to vary over time t . The temporal structure of the covariates
 136 was modeled as a random walk prior on the matrix of time-level means γ ; a
 random-walk prior is a simple way of constraining the estimates for β_t given the
 138 estimate of β_{t-1} . Additionally, the scale parameters σ for each of the K
 coefficients are related to the rate of change over time; a low value of σ_k
 140 corresponds to little between time variance in the effect of that covariate on
 diversity while a large value of σ_k indicates that the effect of that covariate is
 142 inconsistent through time. The values I and S are hyperprior values that we
 specified based on our prior expectations of average unit diversity (Table 1).

$$\begin{aligned}
 \mu_i &= \exp(X_i \beta_{t[i]}) \\
 \beta_t &\sim \text{MVN}(\gamma_t, \Sigma) \\
 \gamma_{t,k} &\sim \begin{cases} \mathcal{N}(I, S) & \text{if } t = 1 \\ \mathcal{N}(0, 1) & \text{if } t = 1, k \neq 1 \\ \mathcal{N}(\gamma_{t-1,k}, \sigma_k) & \text{if } t > 1, k \neq 1 \end{cases} \quad (2) \\
 \sigma_k &\sim \mathcal{N}^+(1) \\
 \frac{1}{\phi} &\sim \mathcal{N}^+(1).
 \end{aligned}$$

144 The additional covariance between variation in the regression coefficients β over
 time that not accounted for by the random-walk prior on γ are modeled by the
 146 unknown/estimated covariance matrix Σ . In order to improve sampling
 performance and choice of priors, the covariance matrix was decomposed into a

148 vector of scales τ and a correlation matrix Ω as recommended by the Stan
Manual CITATION. Their associated priors are as follows:

$$\begin{aligned}\Sigma &= \text{diag}(\tau)\Omega\text{diag}\tau \\ \Omega &\sim \text{LKJ}(1) \\ \tau &\sim \mathcal{N}^+(1).\end{aligned}\tag{3}$$

150 The correlation matrix Ω was given a LKJ distribution prior based on
recommendations in Stan Manual. This distribution has a single parameter
152 where values close to 0 correspond to a uniform distribution across all possible
correlation matrices, and as values increase this distribution convergences on an
154 identity matrix. This weakly-informative prior nudges our estimates towards a
result of no correlation between covariate effects over time though is not
156 sufficiently strong enough to prevent us inferring a possible correlation if there
is there is enough evidence.

158 Unless otherwise noted, all prior choices reflect our decision to use
weakly-informative regularizing priors. Additionally, because all covariates are
160 on approximately unit scale and we do not expect any of our regression
coefficients to have magnitude greater than 2, more diffuse priors would serve no
162 purpose and are unnecessary. Additionally, more diffuse priors would not reflect
our actual expectations regarding the magnitude of covariate effects. Finally,
164 the regularizing property of priors helps constrain our results such that we do
not obtain spurious estimates of the covariate effects. Further statistical and
166 philosophical backing for these prior choices is available HERE HERE AND
HERE CITATION.

Taxonomic group	Intercept prior mean I	Intercept prior scale S	Dispersion scale H
Anthozoa	1	2	3
Brachiopoda	2	2	5
Bivalvia	1	2	3
Cephalopoda	2	2	5
Gastropoda	2	2	5
Trilobita	2	2	5

Table 1: Key prior choices for each of the taxonomic groups included in this analysis. Prior choice reflects our expectations of the average diversity of that group in a geologic unit.

¹⁶⁸ In total, the complete model is as follows

$$\begin{aligned}
y &\sim \text{Negative Binomial}(\mu, \phi)T[1,] \\
\mu_i &= \exp(X_i \beta_{t[i]}) \\
\beta_t &\sim \text{MVN}(\gamma_t, \Sigma) \\
\gamma_{t,k} &\sim \begin{cases} \mathcal{N}(I, S) & \text{if } t = 1 \\ \mathcal{N}(0, 1) & \text{if } t = 1, k \neq 1 \\ \mathcal{N}(\gamma_{t-1,k}, \sigma_k) & \text{if } t > 1, k \neq 1 \end{cases} \quad (4) \\
\frac{1}{\phi} &\sim \mathcal{N}^+(1). \\
\sigma_k &\sim \mathcal{N}^+(1) \\
\Sigma &= \text{diag}(\tau)\Omega\text{diag}\tau \\
\Omega &\sim \text{LKJ}(2) \\
\tau &\sim \mathcal{N}^+(1).
\end{aligned}$$

As stated earlier, for all taxonomic groups the intercept term is an estimate of
¹⁷⁰ the expected (log) geologic unit diversity for a geologic unit with mean
thickness, area, latitude, and a purely non-dolomitic carbonate lithology. The

172 effects of thickness, area, and latitude correspond to the expected change in
(log) geologic unit diversity per change of the covariate value in standard
174 deviations. The effects of dolomite, fine and coarse siliciclastic correspond to the
change associated with unit change to the logratio representing the
176 compositional part of interest (?).

1.2.1 Implementing model in Stan

178 The joint posterior was approximated using a Markov-chain Monte Carlo
routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
180 Sampler (?) as implemented in the probabilistic programming language Stan
(?). The posterior distribution was approximated from four parallel chains run
182 for 40,000 steps, split half warm-up and half sampling and thinned to every
20th sample for a total of 4000 posterior samples. Chain convergence was
184 assessed via the scale reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$)
indicate approximate convergence. Convergence means that the chains are
186 approximately stationary and the samples are well mixed (?). After the model
was fit to the data, 100 datasets were simulated from the posterior predictive
188 distribution of the model. These simulations were used to test for adequacy of
model fit as described below.

190 Hierarchical models can have very complex posterior geometries which make full
exploration of the log-posterior surface difficult Stan Manual CITATIONS. The
192 two strategies for overcoming sampling pathologies associated with sampling an
extremely convoluted log-posterior surface are a non-centered parameterization
194 of the normal distribution used to describe hierarchical structure in the model,
as well as adjusting some the key parameters governing Stan's sampling
196 adaptation phase.

Non-centered parameterization help mitigate divergences because this separates
198 the location from the scale, thus “opening” up the log-posterior surface. The
cost of this reparameterization is the addition of one parameter per regression
200 coefficient, though this parameter has good sampling behavior is relatively
constrained by a regularizing prior. For the details of what that means and how
202 this change in parameterization improves sampling please see Betancourt and
Girolami 2013 CITATION and the Stan manual CITATION.

204 The above model specifications (Eq. 4) were modified as follows:

$$\begin{aligned}
\beta_t &= \gamma_t + z\Sigma \\
\gamma'_{t,1} &\sim \mathcal{N}(I, S) \quad \text{if } t = 1 \\
\gamma'_{t,-} &\sim \mathcal{N}(0, 1) \quad \text{if } t = 1 \text{ and } k > 1 \\
\gamma'_{t,-} &\sim \mathcal{N}(0, 1) \quad \text{if } t > 1 \\
z &\sim \mathcal{N}(0, 1)
\end{aligned} \tag{5}$$

We used five different diagnostic criteria to determine if our chains were well
206 mixed and if our posterior estimates were based on unbiased samples: the scale
reduction factor \hat{R} (target value of <1.1), effective number of samples (eff;
208 target value of $\text{eff}/\text{steps} < 0.0001$), if any samples saturated the specified
maximum trajectory length for avoiding infinite loops (treedepth; target value
210 of 0 samples), presence of divergent samples which indicate pathological
sampling in some neighborhoods of the log-posterior (divergences; target value
212 of 0 samples), and the energy Bayesian Fraction of Missing Information
(E-BFMI; target value >0.2). For a further explanation of these diagnostic
214 criteria see Stan Manual CITATIONS.

Stan’s adaptation phase can be adjusted to help overcome issues resulting from
216 extremely high curvature of the log-posterior. Ultimately, estimation of the joint

Taxonomic group	chain steps (half warm-up, half sample)	thinning	adapt delta	max tree dep
Anthozoa				
Bivalvia				
Brachiopoda				
Gastropoda				
Trilobita				
(Bivalvia + Gastropoda)				

Table 2: Settings for the Stan sampler for estimating model posterior densities.

posterior distributions for each of the taxonomic datasets required different
218 settings for the number of steps for each chain as well as multiple model
adaptation parameters (Table 2).

220 Model adequacy was evaluated using a series of posterior predictive checks. The
concept of model adequacy is that if our model is an adequate descriptor of the
222 observed data, then data simulated from the posterior predictive distribution
should be similar to the observed given the same covariates, etc. (?). Posterior
224 predictive checks can take many forms but the basic idea is to compare some
property of the empirical data to that property estimated from each of the
226 simulated datasets. For each check, the value of a test statistic from the data is
compared to a distribution of that test statistic estimated from datasets
228 simulated from the posterior. Model adequacy is indicated by our simulated
values being approximately equal to the observed values.

230 We used a variety of posterior predictive checks to assess the quality of model’s
fit to each of the different datasets. The goal of using this many and variety of
232 checks is understand the quality and nature of model fit. For example, our
model may have good fit to many aspects of the data but “fail” one or more
234 checks, highlighting potential differences between our model and the data
generating process thus promoting further study CITATION. For example if
236 there was obvious divergence between our models and the data we would not
have confidence in any downstream analyses or hypothesis tests, and we would

238 instead question how or why our model fails and possibly improve our model to
better reflect important unmodeled variance.

240 The checks used here are comparisons of the overall mean unit diversity, the
overall standard deviation of unit diversity, the mean unit diversity for each time
242 step, and the standard deviation of unit diversity for each time step to those
test statistics from 1000 posterior predictive datasets. Most of these posterior
244 predictive checks were done using the `bayesplot` package for R CITATION.

2 Results

2.1 Posterior predictive results

Overall, the expected taxonomic diversity of geological units for each taxonomic group is adequately described by the fitted models, where adequacy means that the posterior predictive distribution of our models resemble the empirical data. While there are aspects of misfit when considering their entire distribution, the aspects of the distribution which are critical to our analysis are well fit by our models.

A point comparison between the observed mean geologic unit diversities and the posterior predictive distributions for each taxonomic group indicate that our fitted models are able to recapitulate this aspect of the observed data (Fig. 1). This result is reassuring because our model is specifically a model of expected geologic unit diversity, and a good fit to mean diversity means our model fits are at least capturing this basic aspect of the data.

Comparison of the observed standard deviation estimates for each of the taxonomic datasets to the posterior predictive distributions of our model fits show that our model is slightly over estimating the scale of our data, though not to a necessarily concerning degree (Fig. 2). Count data can frequently be over dispersed and have a standard deviation to mean ratio greater than 1; this reality is the reason we chose to use a truncated Negative-Binomial as opposed to a Poisson distribution

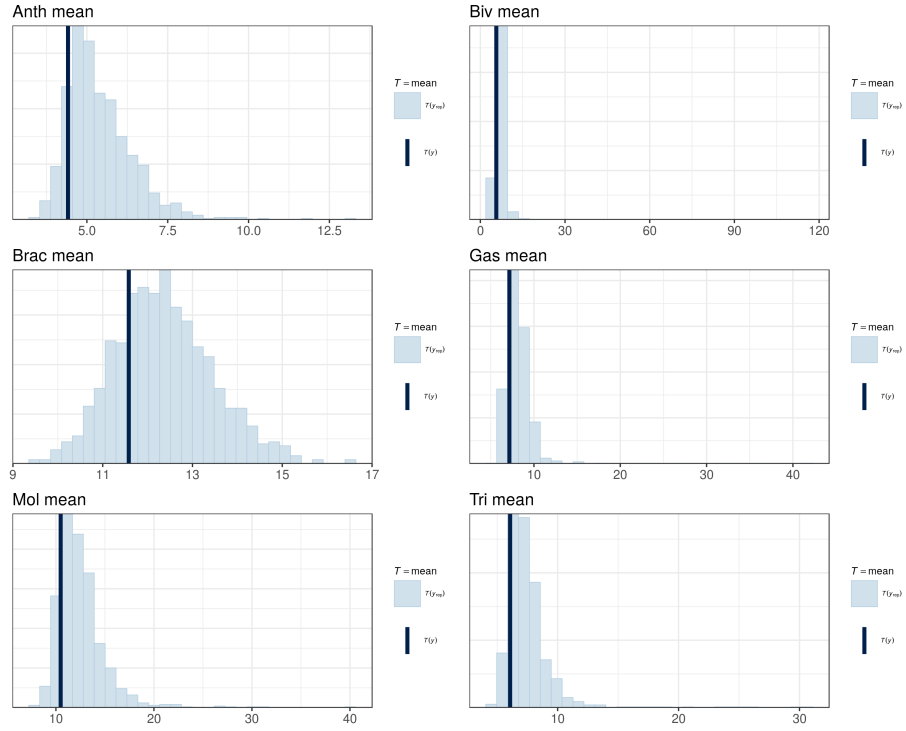


Figure 1: Posterior predictive results comparing the observed mean diversity of a geological unit for each of the studied taxonomic groups to a distribution of 1000 estimates from datasets simulated from the posterior predictive distribution of our models. Model adequacy is determined by how similar the posterior predictive distribution is to the observed value. In all cases, our models appear able to reproduce to observed means.

270 because the addition of a second parameter allows us to model
 this overdispersion CITATIONS. While our model is not too
 272 different from the data, there is room for improvement in
 modeling the actual dispersion of geologic unit diversity.

274 Comparisons of the empirical probability density functions for
 each of the taxonomic groups to the posterior predictive
 276 distribution of density functions generated by our model fits
 indicate that our model is very capable of recapitulating the

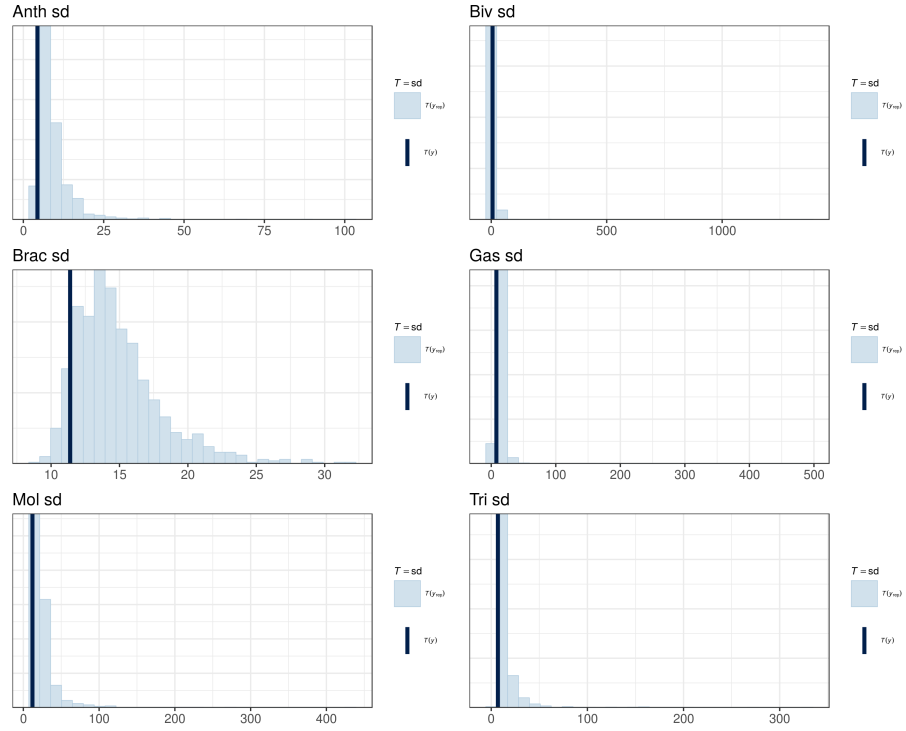


Figure 2: Posterior predictive results comparing the observed standard deviation diversity of a geological unit for each of the studied taxonomic groups to a distribution of 1000 estimates from datasets simulated from the posterior predictive distribution of our models. Model adequacy is determined by how similar the posterior predictive distribution is to the observed value. In all cases, our models appear able to reproduce to observed standard deviations.

278 observed data for nearly its entire range (Fig. 3). There are,
however, minor but noticable differences between the posterior
280 predictive distribution and the empirical data. For example, the
posterior predictive distribution for Anthozoa slightly
282 underestimates the number of units with diversity approximately 8.
A similar underestimate is observable when comparing the
284 posterior predictive distribution to the empirical data at unit
diversity of approximately 11, and Brachiopoda at unit diversity
286 of approximately 11. However, the posterior predictive

distributions of our models fit the data well in nearly all cases,
 288 indicating that our model is potentially capturing some aspects
 of the data generating process.

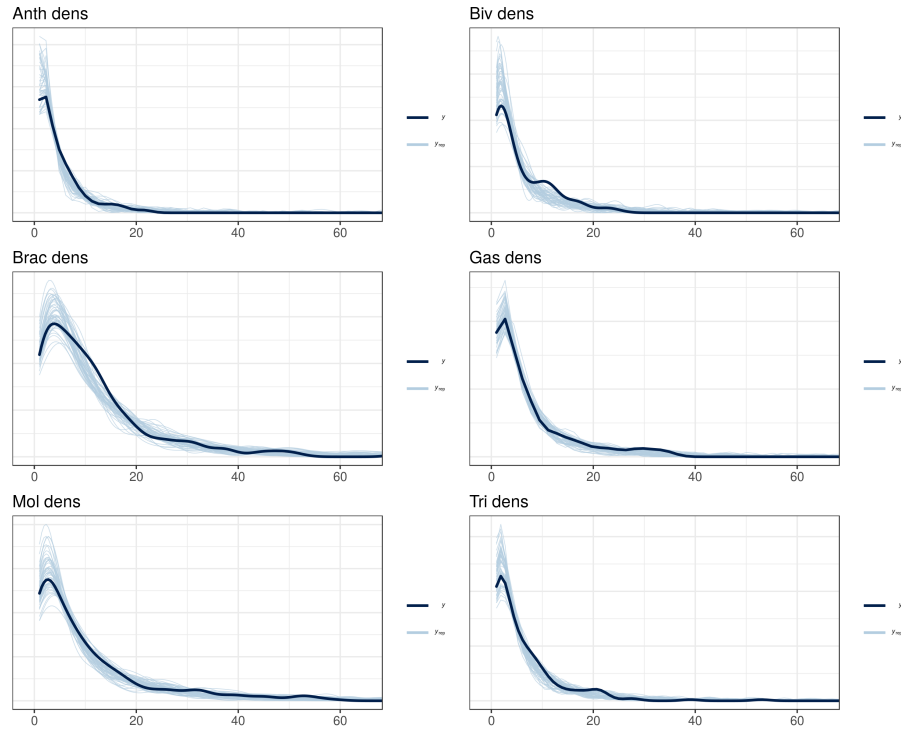


Figure 3: Posterior predictive results comparing the empirical probability density of a geological unit for each of the studied taxonomic groups to a distribution of 1000 probability densities from datasets simulated from the posterior predictive distribution of our models. Model adequacy is determined by how similar the posterior predictive distribution is to the observed value. In all cases, our models appear able to almost reproduce to observed ecdf-s.

2.2 Estimated versus observed unit diversity

Comparison between observed unit diversity over time and our
 292 models' estimates of mean unit diversity for those time steps
 reveals broad congruence (Fig. 4). At no point does our model

294 have a spurious or unrealistic estimate of mean geologic unit
 taxonomic diversity. This congruence gives us confidence in
 296 estimating the probability of the Hirnantian (1) having lower
 average unit diversity than the times directly before and after,
 298 and (2) having lower average unit diversity than the average unit
 diversity of the late Ordovician and the Silurian.

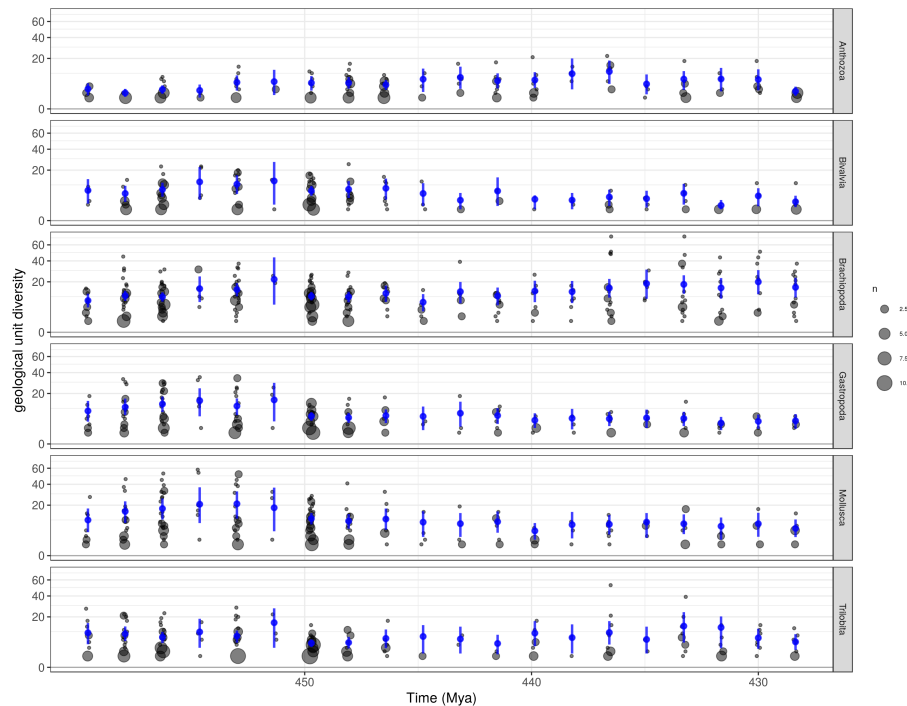


Figure 4: Geological unit diversity through time and the expected diversity (with 80% credible interval) as estimated from our models. Unit diversity is presented as partially transparent points and jittered in the y-axis to improve readability. Point size is proportional to the number of units in that interval that have identical unit diversity. The dashed grey line corresponds to the onset of the Hirnantian geological stage, while the dashed-dotted grey line corresponds to the end of the Ordovician epoch and the start of the Silurian epoch.

300 We tested the probability that geologic units during the
 Hirnantian have lower expected unit diversity than the times

302 immediately before and after by testing all adjacent time bins
 (Fig. 5). For each adjacent time bins, we estimated the
 304 probability that the earlier time bin (time t) has a greater
 expected duration than the later time bin (time $t + 1$). Our
 306 analysis demonstrates that, for most comparisons, the expected
 unit diversity of Hirnantian time bin is not expected with high
 308 probability to be different from the time units immediately
 preceeding and following it.

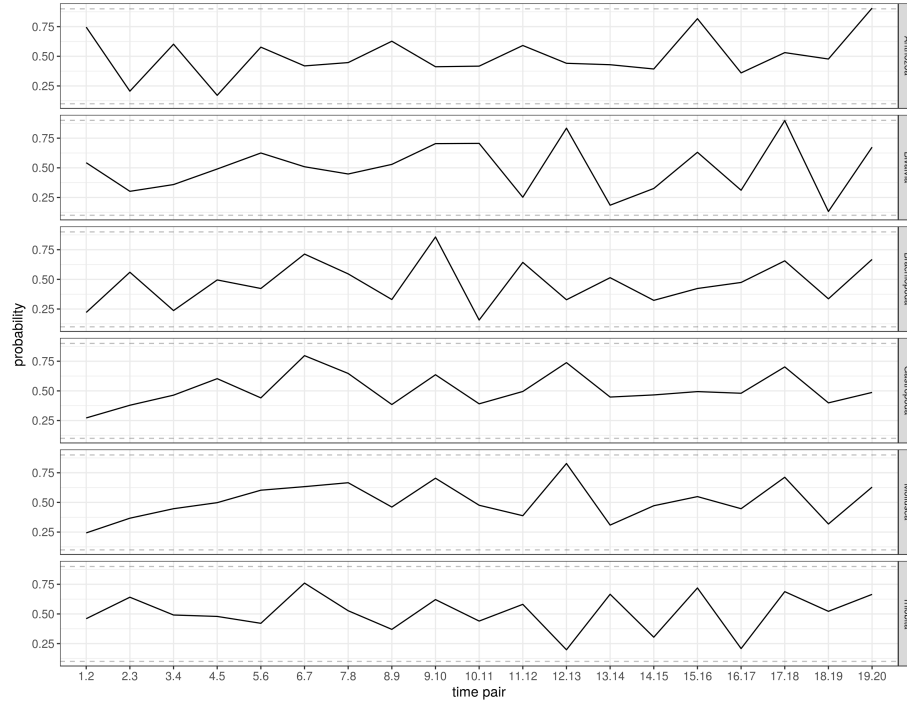


Figure 5: Probability that our estimate of mean unit diversity at time t is greater than the estimate at time $t + 1$. The dashed grey horizontal lines correspond to probability of 0.9 and 0.1; these are the thresholds we chose as indicating if a pair-wise difference is potentially larger (or smaller) than no-difference ($P = 0.5$), and worthy of further inspection.

310 2.3 Effects of geological covariates on estimated diversity

As stated earlier, for all taxonomic groups the intercept term is
312 an estimate of the expected (log) diversity of geologic unit
diversity with mean thickness, area, latitude, and a purely
314 non-dolomitic carbonate lithology. The effects of thickness, area,
and latitude correspond to the expected change in (log) geologic
316 unit diversity per change of the covariate value in units of
standard deviations. The effects of dolomite, fine and coarse
318 siliciclastic correspond to the change associated with unit
change to the logratio representing the lithological composition
320 of interest (?).

Estimates of covariates effects over time demonstrate a gradual
322 shift in effects over time and not a sudden shift during the
Hirnantian or between the Ordovician and the Silurian (Fig. ??).
324 Interestingly, the covariate that may demonstrate the biggest
pattern associated with the Hirnantian is the effect of geologic
unit areal extent which appears to decrease in effect for
326 Bivalvia, Gastropoda, and Mollusca.

328 Similar to our earlier comparison of expected geologic unit
diversity, we tested if any of the covariate effects estimated
330 for one time bin (time t) were greater than the estimates from the
following time bin (time $t+1$). We find no strong evidence that
332 the Hirnantian time bin is significantly different from the time
bins immediately before and after (Fig. 7). There are very few
334 examples of one time bin having significantly different expected
unit diversity than the one proceeding it. The intercept of our

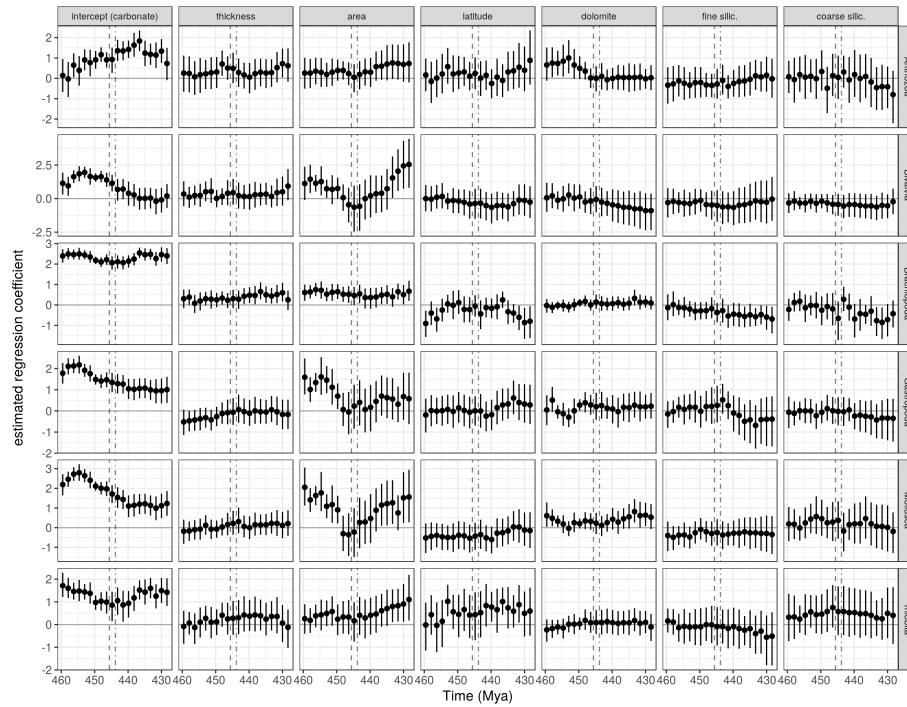


Figure 6: Estimates of all estimated covariate effect time series for each of the analyzed taxonomic groups, including intercept estimates. Points represent mean estimate along with a 80% credible interval. The black horizontal line corresponds to no effect. Points are plotted at the mid-point of the discrete time interval.

336 model fit to the Bivalvia dataset is expected to be greater
during the second time unit than the third, though this
338 difference is not associated with the Hirnantian time unit and
thus is not very relevant to the thrust of our analysis.
340 Similarly, we estimated a potentially significant difference the
estimated effect of area in Mollusca, where the estimate for the
342 seventh time unit is greater than the estimate for the eighth
time unit.
344 Of particular interest is if the Hirnantian has a lower expected
unit diversity than the average of the Ordovician or Silurian.

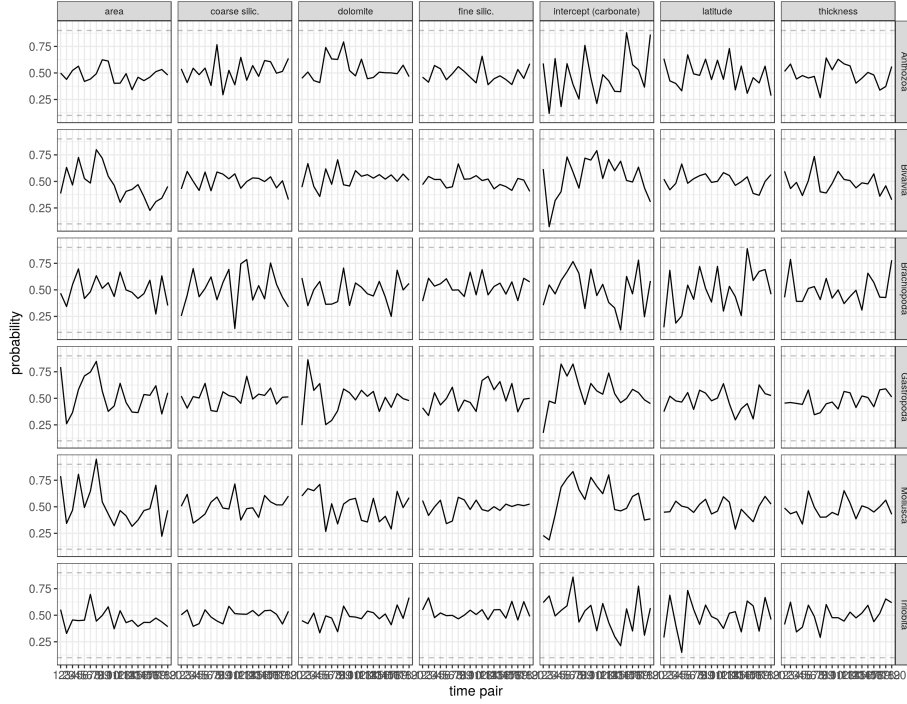


Figure 7: Probability that a parameter estimate at time t is greater than the estimate at time $t + 1$. The dashed grey horizontal lines correspond to probability of 0.9 and 0.1; these are the thresholds we chose as indicating if a pair-wise difference is potentially larger (or smaller) than no-difference ($P = 0.5$), and worthy of further inspection.

346 Additionally, we are interested in if the covariate estimates for
the Hirnantian are different than those estimated for the rest of
348 the Ordovician or Silurian. Here we calculate the probability
that the expected unit diversity of the Hirnantian is less than
350 the averages for the Ordovician or Silurian and compare those
estimates to the probabilities that the covariate effects are
352 less than the averages for the Ordovician or Silurian (Fig. 8).
We find that in most cases there is no strong evidence
354 (probability > 0.9) for the Hirnantian being significantly
different from the averages of either the Ordovician or the

356 Silurian. However, there is weak evidence (probability > 0.75) for
some differences between the Hirnantian and the Ordovician or
358 Silurian.

All of the following results are supported with only weak
360 evidence and are thus of interest for future study of
lithological and diversity differences associated with the
362 Hirnantian.

For Brachiopoda, Gastropoda, and Mollusca the average diversity
364 of geologic units in the Ordovician is estimated with weak
support to be greater than the expected unit diversity of the
366 Hirnantian. We find weak support for a lower effect of dolomite
composition on Anthozoan unit diversity during the Hirnantian
368 than the average of the Ordovician. Similarly, we find weak
support for a lower intercept term and the effect of coarse
370 siliciclastics for Brachiopoda, Gastropoda, Mollusca, and
Trilobita in the Hirnantian than the average of the Ordovician.
372 The intercept is also the effect of being a purely non-dolomitic
carbonate unit. We also find marginal support for the effect of
374 area on unit diversity of Bivalvia, Gastropoda, and Mollusca
being greater for the average of the Ordovician than those units
376 from the Hirnantian.

We do find evidence that Brachiopods are expected to have greater
378 unit diversity during the Silurian than during the Hirnantian.

This result is one of the strongest support results from this
380 study. Bivalvia and Mollusca intercept weak evidence expected to
be greater in Hirnantian than average Silurian. This is also the
382 effect of being a purely non-dolomitic carbonate unit. Bivalvia

and Mollusca are estimated to have a lower effect of geologic
 384 unit area during the Hirnantian than the average of the Silurian.
 Finally, there is marginal evidence for the intercept estimate
 386 for Brachiopoda during the Hirnantian is expected to be lower
 than the average estimated intercept for the Silurian.

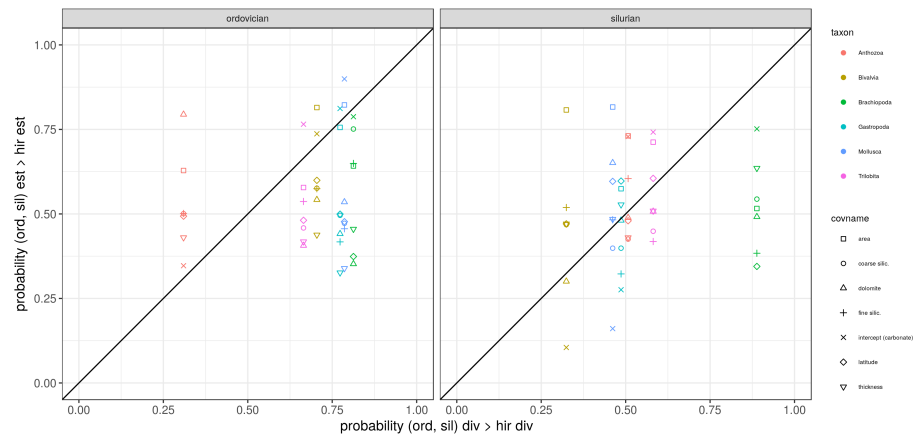


Figure 8: Scatterplot of the estimated probability that geological unit diversity is lower during the Hirnantian than either the Ordovician (left facet) or the Silurian (right facet) vs the estimated probability that a covariate estimate is lower during the Hirnantian than either the Ordovician or the Silurian. For each of the taxonomic groups there is only one estimate for the probability of difference in diversity, but there are six probability estimates for each of the covariate effect parameters.