1 Materials and Methods

1.1 Geological unit information

All information used in this analysis is avaliable freely through Macrostrat `macrostrat.org` and the Paleobiological Database PBDB `paleobiodb.org`. For this analysis, we used direct API calls to pull data from the relational databases underpinning both Macrostrat and the PBDB (Peters and McClennen, 2015, Peters et al., 2018); this means that our analyses are inherently dynamic and can be instantly updated as these databases continue to grow.

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

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 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 the Macrostrat database includes information on the count of Paleobiology Database collections drawn from that unit, the count of unique fossils listed in the Paleobiology Database associated with that unit, the unique taxonomic identifiers for each of those fossils, and the unique identifiers for each of the collections drawn from that unit (cltn_id). This final value acts as the foreign key for extracting fossil occurrence and taxonomic information from the

Paleobiology Database.

A final API call is made, this time to Paleobiology Database, using the collection id foreign key from Macrostrat (e.g. `paleobiodb.org/data1.2/occs/list.txt?coll_id=...`); this API call was 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 drawn from geological units from the Ordovician and Silurian present in Macrostrat.

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 scripts from the project repository `https://github.com/psmits/not_fossil`.

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

An interesting feature of Macrostrat geologic units is that they are ordered according to the underlying continuous-time age model (Peters et al., 2018). This age model increases the overal resolution of the geological record. Unfortunately the fossil 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 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 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 averaging those we get the midpoint age. In total, we divided the data into 20 uniform-duration discrete time intervals.

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

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

First, units that have at least one description including the words igneous, volcanic, metamorphic, chemical, anhydrite, evaporite, or halite were removed 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 become green, mudstone and mud become mudstone). Next, some additional words were removed from descriptions for either being too general or too 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 the final descriptions because our final step is extremely strict.

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

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

Prior to analysis, real values were log transformed by subtracting the mean value from all observations then dividing by twice the standard deviation of the observations. Similarly, positive real valued covariates were log-plus-one transformed and then rescaled in a similar manner. Rescaling the covariates has multiple advantages: 1) regression coefficients now describe the expected change 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 scale (the expected standard deviation of a binary variable is 0.5) (Gelman and Hill, 2006).

In contrast to the other covariations, compositional covariates are constrained to sum to 1 which creates degrees-of-freedom issues when trying to model their possible effects as including these covariates without appropriate transformation creates two or more nonidentifiable parameters. To that end, the composition variables were isometric log-ratio transformed (ilr) (Egozcue et al., 2003) which reduces the total number of variables to one less than original as composition is defined in relation to a baseline (percent non-dolomitic carbonates). Unfortunately, the scale and interpretations of the associated regression coefficients are different from the other covariates, making direct comparison tricky. The ilr transformation was done using the `compositions` package for R (van den Boogaart et al., 2014).

The fossil occurrence and diversity for each of the geological units was determined for each of the following taxonomic groups: Anthozoa, Brachiopoda, Bivalvia, Cephalopoda, Gastropoda, and Trilobita. Fossil membership was

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 (Gelman and Hill, 2006, Gelman et al., 2014). TO model this potential overdisperssion 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:

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

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 and its use in count regression please see Gelman and Hill (2006).

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 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 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 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 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).

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

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 unknown/estimated covariance matrix Σ. In order to improve sampling performance and choice of priors, the covariance matrix was decomposed into a vector of scales τ and a correlation matrix Ω as recommended by the Stan manual (Team, 2017). Their associated priors are as follows:

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

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

Unless otherwise noted, all prior choices reflect our decision to use weakly-informative regularizing priors. Additionally, because all covariates are 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 purpose and are unnecessary. Additionally, more diffuse priors would not reflect our actual expectations regarding the magnitude of covariate effects. Finally, the regularizing property of priors helps constrain our results such that we do not obtain spurious estimates of the covariate effects (Gelman et al., 2014,

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.

2008); see also `http://mc-stan.org/users/documentation/case-studies/weakly_informative_shapes.html`

In total, the complete model is as follows

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

As stated earlier, for all taxonomic groups the intercept term is an estimate of

the expected (log) geologic unit diversity for a geologic unit with mean thickness, area, latitude, and a purely non-dolomitic carbonate lithology. The effects of thickness, area, and latitude correspond to the expected change in (log) geologic unit diversity per change of the coviarate value in standard deviations. The effects of dolomite, fine and coarse siliciclastic correspond to the change associated with unit change to the logration representing the compositional part of interest (Hron et al., 2012).

1.2.1 Implementing model in Stan

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 Sampler as implemented in the probabilistic programming language Stan (Carpenter et al., 2017). The posterior distribution was approximated from four parallel chains run 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 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 approximately stationary and the samples are well mixed (Gelman et al., 2014). After the model was fit to the data, 100 datasets were simulated from the posterior predictive distribution of the model. These simulations were used to test for adequacy of model fit as described below.

Hierarchical models can have very complex posterior geometries which make full exploration of the log-posterior surface difficult Stan manual (Team, 2017). The two strategies for overcoming sampling pathologies associated with sampling an extremely convoluted log-posterior surface are a non-centered parameterization of the normal distribution used to describe hierarchical structure in the model, as well as adjusting some the key parameters governing Stan's sampling

adaptation phase.

Non-centered parameterization help mitigate divergences because this separates 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 coefficient, though this parameter has good sampling behavior is relatively constrained by a regularizing prior. For the details of what that means and how this change in parameterization improves sampling please see Betancourt and Girolami (2013) and the Stan manual (Team, 2017).

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 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; target value of eff/steps<0.0001), if any samples saturated the specified maximum trajectory length for avoiding infinite loops (treedepth; target value of 0 samples), presence of divergent samples which indicate pathological sampling in some neighborhoods of the log-posterior (divergences; target value of 0 samples), and the energy Bayesian Fraction of Missing Information (E-BFMI; target value >0.2). For a further explanation of these diagnostic criteria see Stan manual (Team, 2017).

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.

Stan's adaptation phase can be adjusted to help overcome issues resulting from extremely high curvature of the log-posterior. Ultimately, estimation of the joint posterior distributions for each of the taxonomic datasets required different settings for the number of steps for each chain as well as multiple model adaptation parameters (Table 2).

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 observed data, then data simulated from the posterior predictive distribution should be similar to the observed given the same covariates, etc. (Gelman et al., 2014). Posterior 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 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 simulated from the posterior. Model adequacy is indicated by our simulated values being approximately equal to the observed values.

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 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 checks, highlighting potential differences between our model and the data generating process thus promoting further study (Gelman and Hill, 2006,

Gelman et al., 2014). For example if 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 instead question how or why our model fails and possibly improve our model to better reflect important unmodeled variance.

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 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 predictive checks were done using the `bayesplot` package for R (Gabry and Mahr, 2018).

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 because the addition of a second parameter allows us to model this overdispersion (Gelman and Hill, 2006, Gelman et al., 2014). While our model is not too different from the data, there is room for improvement in modeling the actual dispersion of geologic unit diversity.

Comparisons of the empirical probability density functions for each of the

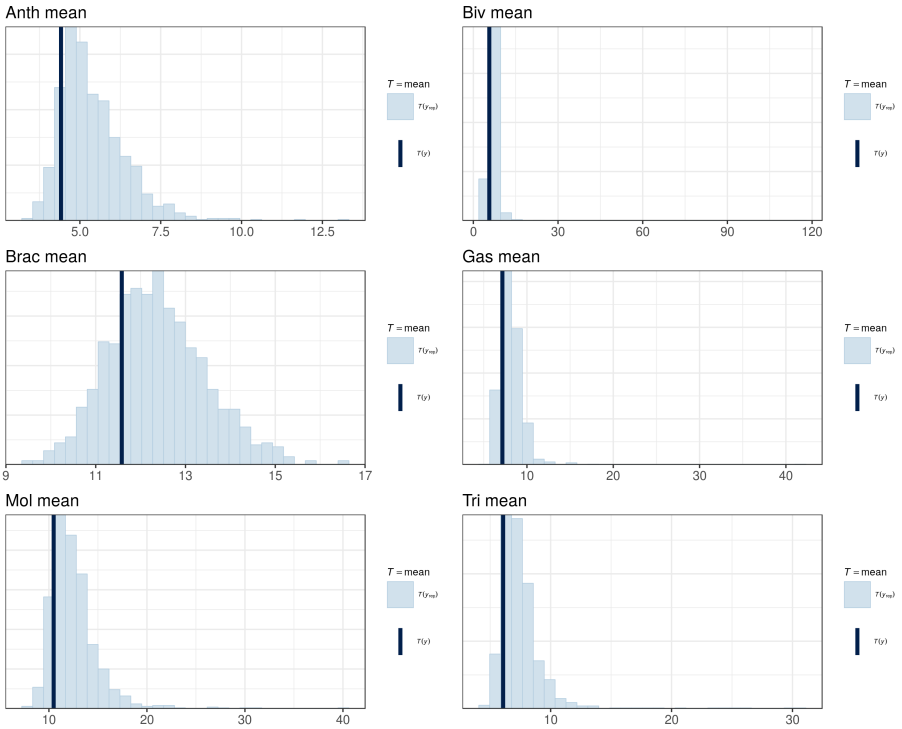

Figure 1: Posterior predictive results comparing the observed mean diversity of a geological unit for each of the studied taxonomic groups to a distribution of 1000 estimates from datasets simulated from the posterior predictive distribution of our models. Model adequacy is determined by how similar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

taxonomic groups to the posterior predictive distribution of density functions generated by our model fits indicate that our model is very capable of recapitulating the observed data for nearly its entire range (Fig. 3). There are, however, minor but noticable differences between the posterior predictive distribution and the empirical data. For example, the posterior predictive distribution for Anthozoa slightly underestimates the number of units with diversity approximately 8. A similar underestimate is observable when comparing the posterior predictive distibution to the empirical data at unit diversity of approximately 11, and Brachiopoda at unit diversity of

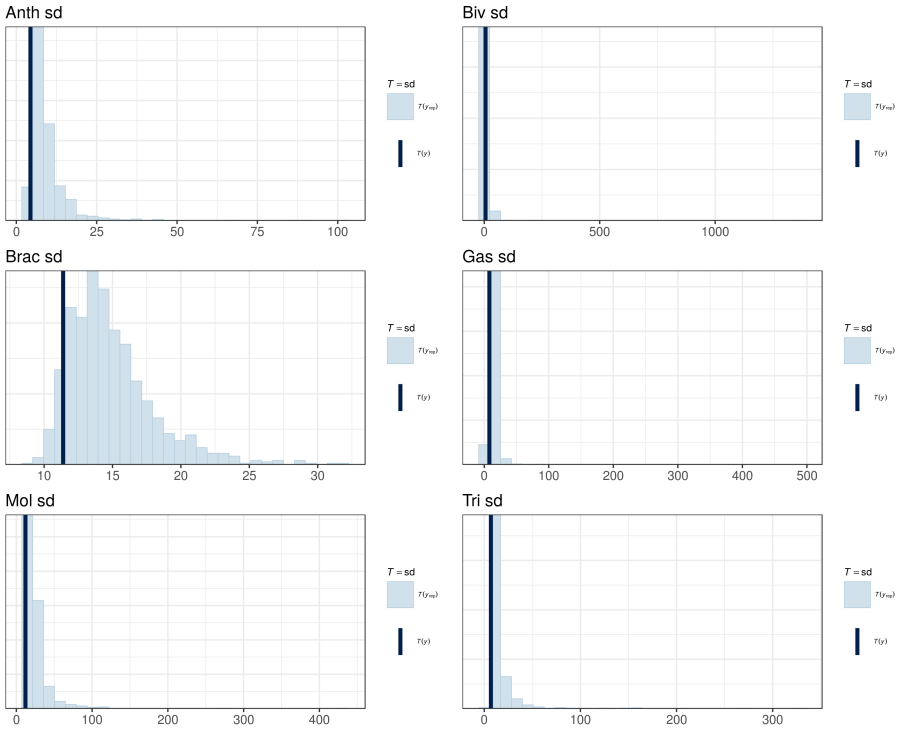

Figure 2: Posterior predictive results comparing the observed standard deviation diversity of a geological unit for each of the studied taxonomic groups to a distribution of 1000 estimates from datasets simulated from the posterior predictive distribution of our models. Model adequacy is determined by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

approximately 11. However, the posterior predictive distributions of our models fit the data well in nearly all cases, indicating that our model is potentially capturing some aspects of the data generating process.

2.2 Estimated versus observed unit diversity

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

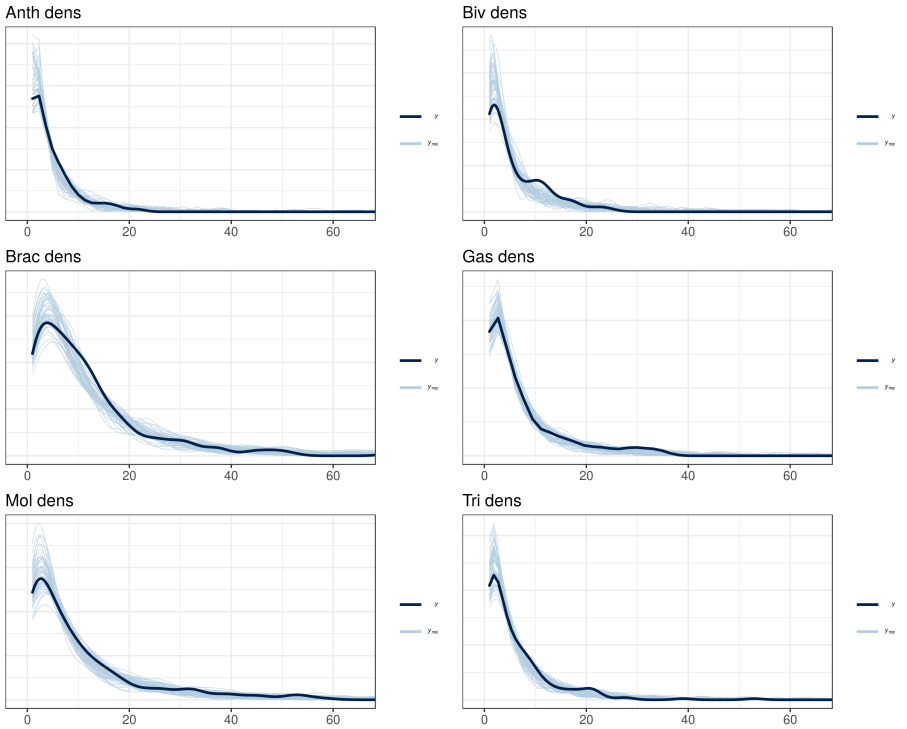

Figure 3: Posterior predictive results comparing the empirical probability density of a geological unit for each of the studied taxonomic groups to a distribution of 1000 probability densities from datasets simulated from the posterior predictive distribution of our models. Model adequacy is determined 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.

mean geologic unit taxonomic diversity. This congruence gives us confidence in estimating the probability of the Hirnantian (1) having lower average unit diversity than the times directly before and after, and (2) having lower average unit diversity than the average unit diversity of the late Ordovician and the Silurian.

We tested the probability that geologic units during the Hirnantian have lower expected unit diversity than the times immediately before and after by testing all advacent time bins (Fig. 5). For each adjacent time bins, we estimated the

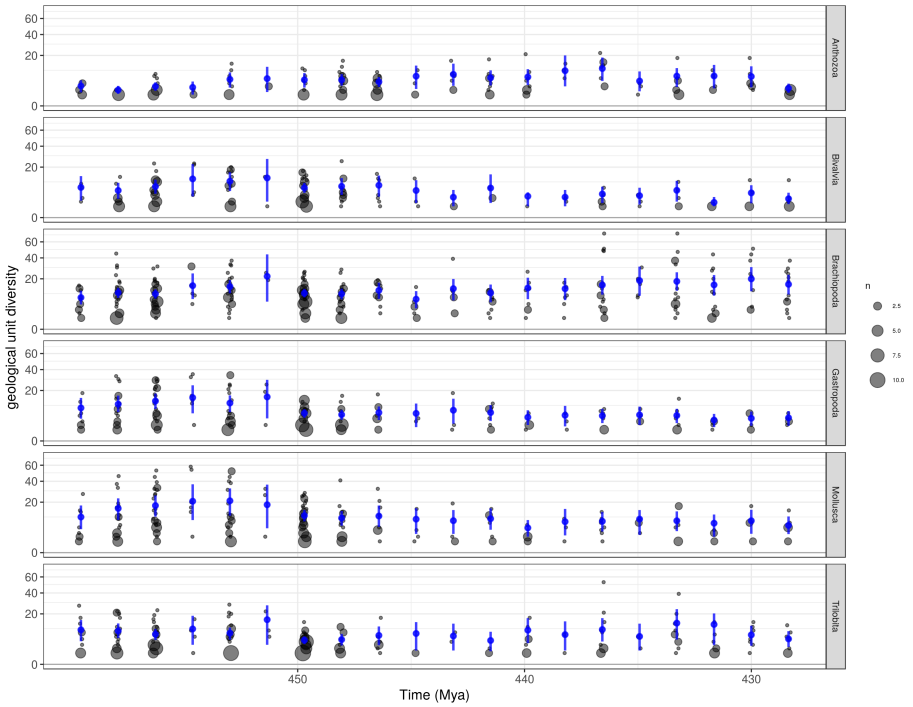

Figure 4: Geological unit diversity though time and the expected diversity (with 80% credible interval) as estimated from our models. Unit diversity is presented as partially transparent points and our jittered in the y-axis to improve readability. Point size is proportional to the number of units in 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.

probability that the earlier time bin (time t) has a greater expected duration than the later time bin (time $t + 1$). Our analysis demonstrates that, for most comparisons, the expected unit diversity of Hirnantian time bin is not expected with high probability to be different from the time units immediately preceeding and following it.

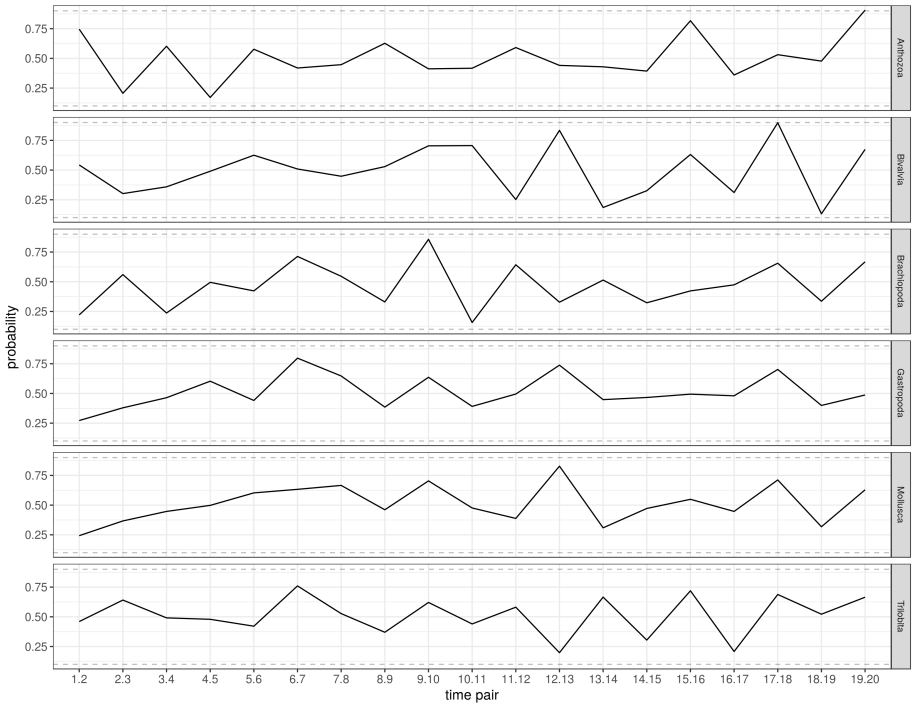

Figure 5: Probability that our estimate of mean unit diversity at time t is greater than the estimate at time $t+1$. The dashed grey horizontal lines correspond to probability of 0.9 and 0.1; these are the thresholds we chose as indicating if a pair-wise difference is potentially larger (or smaller)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

2.3 Effects of geological covariates on estimated diversity

As stated earlier, for all taxonomic groups the intercept term is an estimate of the expected (log) diversity of geologic unit diversity with mean thickness, area, latitude, and a purely non-dolomitic carbonate lithology. The effects of thickness, area, and latitude correspond to the expected change in (log) geologic unit diversity per change of the covariate value in units of standard deviations. The effects of dolomite, fine and coarse siliciclastic correspond to the change associated with unit change to the logratio representing the lithological composition of interest (Hron et al., 2012).

Estimates of covariates effects over time demonstrate a gradual shift in effects over time and not a sudden shift during the Hirnantian or between the Ordovician and the Silurian (Fig. 6). 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 Bivalvia, Gastropoda, and Mollusca.

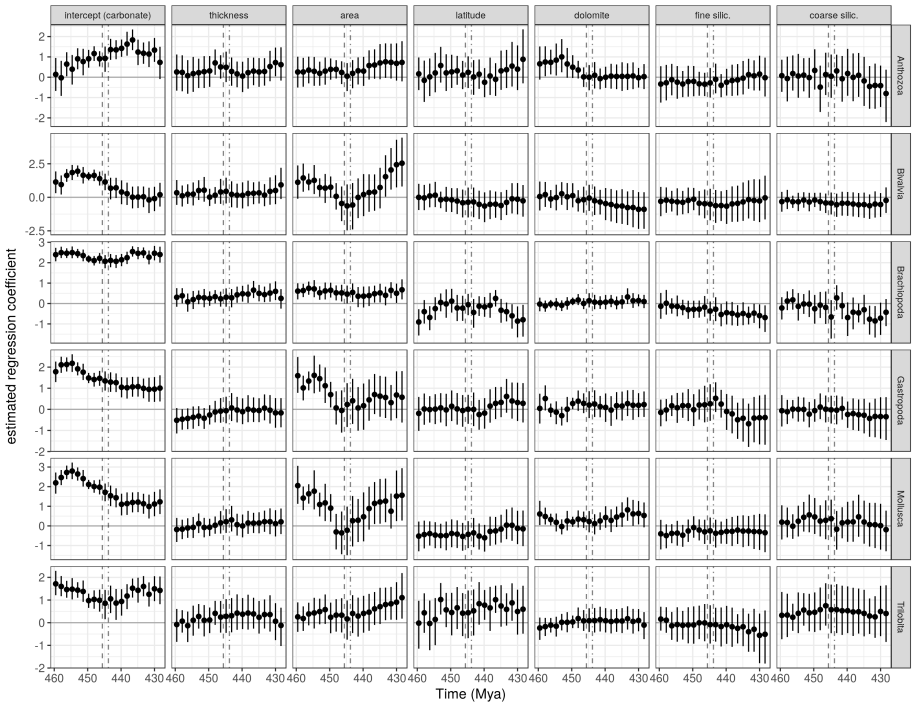

Figure 6: Estimates of all estimated covariate effect time series for each of the analyzed taxonomic groups, including intercept estimates. Points represent mean estimate along with a 80% credible interval. The black horizontal line corresponds to no effect. Points are plotted at the mid-point of the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

Similar to our earlier comparison of expected geologic unit diversity, we tested if any of the covariate effects estimated 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 the Hirnantian time bin is significantly different from the time

bins immediately before and after (Fig. 7). There are very few examples of one time bin having significantly different expected unit diversity than the one proceeding it. The intercept of our model fit to the Bivalvia dataset is expected to be greater during the second time unit than the third, though this difference is not associated with the Hirnantian time unit and thus is not very relevant to the thrust of our analysis. Similarly, we estimated a potentially significant difference the estimated effect of area in Mollusca, where the estimate for the seventh time unit is greater than the estimate for the eighth time unit.

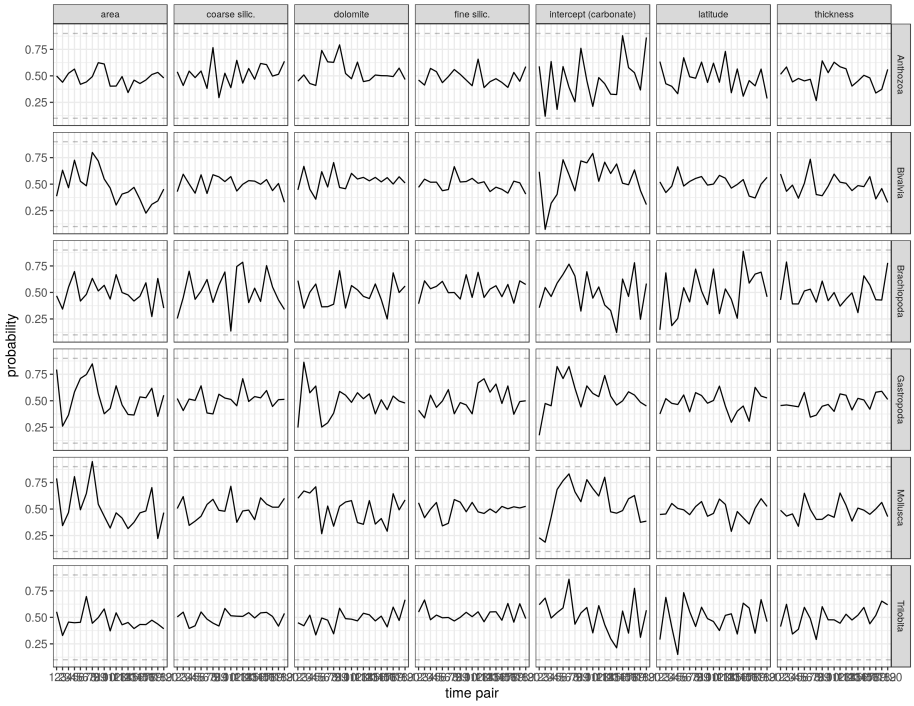

Figure 7: Probability that a parameter estimate at time t is greater than the estimate at time $t+1$. The dashed grey horizontal lines correspond to probability of 0.9 and 0.1; these are the thresholds we chose as indicating if a pair-wise difference is potentially larger (or smaller) than no-difference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

Of particular interest is if the Hirnantian has a lower expected unit diversity

than the average of the Ordovician or Silurian. Additionally, we are interested in if the covariate estimates for the Hirnantian are different than those estimated for the rest of the Ordovician or Silurian. Here we calculate the probability that the expected unit diversity of the Hirnantian is less than the averages for the Ordovician or Silurian and compare those estimates to the probabilities that the covariate effects are less than the averages for the Ordovician or Silurian (Fig. 8). We find that in most cases there is no strong evidence (probability > 0.9) for the Hirnantian being significantly different from the averages of either the Ordovician or the Silurian. However, there is weak evidence (probability > 0.75) for some differences between the Hirnantian and the Ordovician or Silurian.

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

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

We do find evidence that Brachiopods are expected to have greater unit diversity during the Silurian than during the Hirnantian. This result is one of the strongest support results from this study. Bivalvia and Mollusca intercept

weak evidence expected to be greater in Hirnantian than average Silurian. This is also the effect of being a purely non-dolomitic carbonate unit. Bivalvia and Mollusca are estimated to have a lower effect of geologic unit area during the Hirnantian than the average of the Silurian. Finally, there is marginal evidence for the intercept estimate for Brachiopoda during the Hirnantian is expected to be lower than the average estimated intercept for the Silurian.

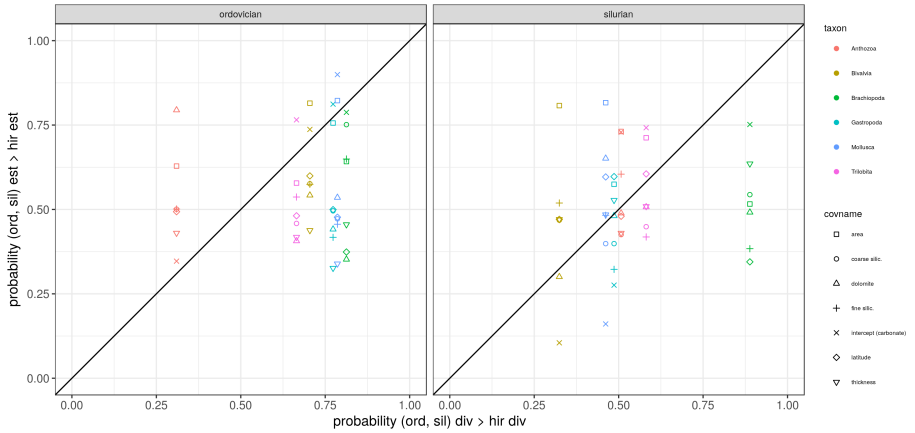

Figure 8: Scatterplot of the estimated probability that geological unit diversity is lower during the Hirnantian than either the Ordovician (left facet) or the Silurian (right facet) vs the estimated probability that a covariate estimate is lower during the Hirnantian than either the Ordovician or the 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.

References

M. J. Betancourt and Mark Girolami. Hamiltonian Monte Carlo for Hierarchical Models. *arXiv*, 2013. doi: 10.1201/b18502-5. URL `http://arxiv.org/abs/1312.0906`.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 2017. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL `http://www.jstatsoft.org/v76/i01/`.

J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, 35(3):279–300, 2003. ISSN 08828121. doi: 10.1023/A:1023818214614.

Jonah Gabry and Tristan Mahr. *bayesplot: Plotting for Bayesian Models*, 2018. URL `https://CRAN.R-project.org/package=bayesplot`. R package version 1.5.0.

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, 2006.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, 3rd edition, 2014.

By Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression

models. *The Annals of Applied Statistics*, 2(4):1360–1383, 2008. doi: 10.1214/08-AOAS191.

K. Hron, P. Filzmoser, and K. Thompson. Linear regression with compositional explanatory variables. *Journal of Applied Statistics*, 39(5):1115–1128, 2012. ISSN 02664763. doi: 10.1080/02664763.2011.644268.

Shanan E. Peters and Michael McClennen. The Paleobiology Database application programming interface. *Paleobiology*, 42(1):1–7, 2015. ISSN 00948373. doi: 10.1017/pab.2015.39.

Shanan E Peters, Jon M Husson, and John J Czaplewski. Macrostrat: a platform for geological data integration and deep-time Earth crust research. *EarthArXiv*, 2018. doi: http://doi.org/10.17605/OSF.IO/YNAXW. URL `https://eartharxiv.org/ynaxw`.

Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2017. URL `http://mc-stan.org`.

K. Gerald van den Boogaart, Raimon Tolosana, and Matevz Bren. *compositions: Compositional Data Analysis*, 2014. URL `https://CRAN.R-project.org/package=compositions`. R package version 1.40-1.