

## Research



**Cite this article:** Smits P, Finnegan S. 2019 How predictable is extinction? Forecasting species survival at million-year timescales. *Phil. Trans. R. Soc. B* **374**: 20190392. <http://dx.doi.org/10.1098/rstb.2019.0392>

Accepted: 25 July 2019

One contribution of 17 to a discussion meeting issue 'The past is a foreign country: how much can the fossil record actually inform conservation?'

### Subject Areas:

palaeontology, ecology, evolution

### Keywords:

conservation, palaeobiology, extinction, forecasting

### Author for correspondence:

Seth Finnegan

e-mail: [sethf@berkeley.edu](mailto:sethf@berkeley.edu)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4704200>.

# How predictable is extinction? Forecasting species survival at million-year timescales

Peter Smits and Seth Finnegan

Department of Integrative Biology, University of California Berkeley, Berkeley, CA, USA

PS, 0000-0003-4074-1447; SF, 0000-0002-6175-6173

A tenet of conservation palaeobiology is that knowledge of past extinction patterns can help us to better predict future extinctions. Although the future is unobservable, we can test the strength of this proposition by asking how well models conditioned on past observations would have predicted subsequent extinction events at different points in the geological past. To answer this question, we analyse the well-sampled fossil record of Cenozoic planktonic microfossil taxa (Foraminifera, Radiolaria, diatoms and calcareous nanoplankton). We examine how extinction probability varies over time as a function of species age, time of observation, current geographical range, change in geographical range, climate state and change in climate state. Our models have a 70–80% probability of correctly forecasting the rank order of extinction risk for a random out-of-sample species pair, implying that determinants of extinction risk have varied only modestly through time. We find that models which include either historical covariates or account for variation in covariate effects over time yield equivalent forecasts, but a model including both is overfit and yields biased forecasts. An important caveat is that human impacts may substantially disrupt range-risk dynamics so that the future will be less predictable than it has been in the past.

This article is part of a discussion meeting issue 'The past is a foreign country: how much can the fossil record actually inform conservation?'

## 1. Introduction

The intensifying biodiversity crisis confronts conservation biologists with the difficult task of trying to predict which species are most threatened with extinction in the near future. Predicting which species will go extinct is difficult because reliable population and geographical range time series are typically known for only the past few decades in even the best-studied groups, and because few modern extinctions have been adequately documented. This has led to the suggestion that some risk assessments might be improved by incorporating palaeontological data [1,2]. The fossil record preserves information about the full histories, including ultimate extinction, of thousands of lineages, and this information can help to augment the shorter-term higher-resolution data used to make risk assessments of extant taxa.

Extinction intensity (average rate) and selectivity (difference in risk among taxa) have varied greatly through time, and the relative risk of extinction exhibited by different taxonomic and ecological groups can provide insights into the drivers of both background and mass extinction [3–8]. Many studies have examined the effects of various potential predictors on extinction risk through time [3,5,9–14] or refined methods for identifying and measuring these effects [15–20]. These studies have produced a growing body of knowledge regarding which factors have been general determinates of extinction risk in the geological past.

A related question that has received much less attention is how successful we might expect to be when using this knowledge to attempt to predict future extinction events.

Because future extinctions are unobservable we cannot directly evaluate the ultimate performance of such predictions. However, we can take a given point in the geological past, develop a predictive model based on extinction patterns prior to that point, and assess the predictive performance of this model on subsequent (e.g. 'future', from the point of view of the model) extinction/survival events. Putting aside the very important question of how human activities will alter future determinants of extinction risk, such an approach provides a framework for evaluating the expected accuracy of future risk assessments based on past extinction events.

Here we take this approach, using as a model system the Cenozoic record of skeletonized marine planktonic microorganisms (Foraminifera, Radiolaria, diatoms and calcareous nannoplankton including coccolithophores). This record has several key strengths for our purposes: planktonic microorganisms are widespread and abundant in pelagic habitats, have high preservation potential, and because of their use in biostratigraphic, palaeoclimatic and oceanographic study they have been the focus of an extensive international coring and study effort [21,22]. A compilation of these data is readily available through the Neptune database, an online repository of species occurrences obtained through the Deep Sea Drilling Program and the Ocean Drilling Project [21,22]. This database provides abundant samples in space and time, a high degree of temporal resolution for the entirety of the Cenozoic, and has a taxonomic synonymization framework for dealing with more than 50 years of taxonomic opinion [21]—as close to ideal data for this analysis as possible. Analysing patterns of extinction and global occurrence at fine temporal scales means we can better elucidate how well we can predict species extinction at human-relevant scales.

The overall question of how well models based on past extinction patterns perform at forecasting future extinctions depends in part on model complexity. Simple models requiring only a few parameters are in general preferable because more complex models run a greater risk of being overfit to the observations on which they are trained. In addition, many traits that might influence extinction risk among extant species are difficult to assign confidently to extinct species. For these reasons, we elect to focus on 'baseline' models which include only a few parameters that have been shown to be important and/or consistent determinants of extinction risk in the marine fossil record. Numerous studies have established that geographical range is one of the most important determinants of extinction risk in the fossil record, and that a species' geographical range can be highly variable over geological time [2,5,23–28]. In addition to geographical range, we also considered global climate state and change in climate state since previous observation in order to evaluate the influence of climate or climate change trajectory on extinction risk. Finally, we included species age, both because previous studies of planktonic taxa have found it to be a determinant of extinction risk and because its inclusion in our models is critical to their nature as survival models (see Model Specifications below). We reiterate that our primary objective is to evaluate the predictive performance of simple models that include only a few general parameters; more complex models including other likely determinants of extinction risk such as skeletal mineralogy, trophic ecology and thermal tolerance range might well perform better.

There are a number of ways in which past extinction patterns might be used to model present risk. The simplest case assumes that relationships between predictors (hereafter covariates) and extinction risk have been constant through time. A more complex but more realistic case allows relationships between covariates and extinction risk to vary through time, consistent with evidence for temporal variation in extinction selectivity regime. Finally, an important consideration is the degree to which species' geographical range trajectories exhibit deterministic versus Markovian behaviour [2,27,29,30]. In the former case, knowledge of the specific past trajectory of a species—whether its range has expanded or contracted from some point in the past to the present—might help to improve assessments of its current risk. In the latter case, only the current geographical range of the species would convey useful information about current and future risk (although that assessment would still be based on the relative extinction risk of species that had similar ranges at different points in the geological past). In all cases we use information about the past to predict the future, the question is whether and how much historical information (e.g. the histories of species over time) improves our ability to forecast future extinction events.

Below, we evaluate four models along a spectrum from simplest (fixed covariate effects, Markovian range dynamics) to most complex (varying covariate effects, deterministic range dynamics). We ask: (i) how well they perform at classifying species as extinctions or survivors in the data they were fitted to, and (ii) how well they perform at classifying species as extinctions or survivors in 'future' data that were not used in fitting the models.

## 2. Material and methods

### (a) Data specifications

We analysed microfossil occurrence information from the Neptune Database <http://www.nsb-mfn-berlin.de/nannotax> [21,22]. This occurrence-based dataset includes calcareous nannoplankton, diatoms, planktonic Foraminifera and radiolarians. Occurrences were filtered to include only those species with first occurrences no earlier than 63 Ma. This filtering criterion excludes taxa that survived the Cretaceous/Palaeogene extinction or arose during this recovery interval, and ensures that our occurrence histories fully overlap with the temperature time-series used as a potential extinction risk predictor (see below).

All fossil occurrences were assigned to 1 Myr bins based on the estimated age of the fossil occurrence as listed in the Neptune Database. After binning, each species' geographical range was calculated for each of the 1 Myr bins in which it occurred. Geographical range was calculated as the minimum spanning tree distance between all observations of that taxon during that temporal bin; this distance was measured in kilometres. Minimum spanning tree distance was calculated using the GeoRange package for R [31].

We also included how a species' geographical range has changed since its last three observation times. We measured this change in geographical range by calculating the difference in geographical range between an observation and that species' three previous occurrences. Change between the most recent and the three previous occurrences was calculated individually for each of those lags. If there were not enough previous observations to calculate, then that value was recorded as a zero. These differences were calculated after minimum spanning tree distance was transformed and standardized (see the electronic supplementary material, S1.1.2).

Average global temperature of each 1 Myr bin was calculated from estimates based on magnesium/calcium (Mg/Ca) of deep

**Table 1.** Models and their definitions.

code	description	covariates	R formula syntax <sup>a</sup>
C	constant effects, no historical cov.	geographical range, temperature	event <sup>b</sup> ~ range <sup>c</sup> +temp <sup>d</sup> + (1   time <sup>e</sup> /phylum <sup>f</sup> ) + (1   age <sup>g</sup> /phylum)
V	varying effects, no historical cov.	geographical range, temperature	event ~ range + temp + (1 + range + temp   time/phylum) + (1   age/phylum)
CP	constant effects, historical cov.	geographical range, change in geographical range, temperature, previous temperature	event ~ +range_diff1 <sup>g</sup> + range_diff2 <sup>h</sup> + range_diff3 <sup>h</sup> + temp + temp_lag <sup>i</sup> + (1   time/phylum) + (1   age/phylum)
VP	varying effects, historical cov.	geographical range, change in geographical range, temperature, previous temperature	event ~ range + range_diff1 + range_diff2 + range_diff3 + temp + temp_lag + (1 + range + range_diff1 + range_diff2 + range_diff3 + temp + temp_lag   time/phylum) + (1   age/phylum)

<sup>a</sup>See the electronic supplementary material, equation S2 for full statistical model definition.

<sup>b</sup>Species observation where variable is 1 if time of last observation, otherwise 0.

<sup>c</sup>'Species' geographical range in log km<sup>2</sup>. Mean centred, scaled to s.d. = 1.

<sup>d</sup>Global temperature in degrees C. Mean centred, scaled to s.d. = 1.

<sup>e</sup>Time of observation.

<sup>f</sup>Taxonomic group of species (i.e. Foraminifera, diatoms, Radiolaria, calcareous nannoplankton).

<sup>g</sup>Age at observation.

<sup>h</sup>Change in geographical range since last observation (number indicates how many lags).

<sup>i</sup>Temperature at previous observation.

sea benthic foraminifera (Cramer *et al.* [32]). We use Mg/Ca rather than oxygen isotopes to avoid the confounding effect of varying ice-volume—this property is of particular importance for this analysis as polar ice-caps developed midway through the Cenozoic. Cramer *et al.* [32], estimated temperature for every 0.1 Myr interval between 0 and 63 Ma. The temperature estimate for each 1 Myr interval was calculated as the mean of all estimates within that interval.

We also included the global temperature from the last previous time interval in which each species was observed. If there were not enough previous observations to calculate, then that value was recorded as a zero. This lag was calculated after global temperature was transformed and standardized (see the electronic supplementary material, S1.1.2).

The Mg/Ca-based temperature estimates we use are measured from benthic Foraminifera, and are an estimate of deep water ocean temperature. Although it would be desirable to have local sea-surface temperature estimates, these data are not available for the great majority of occurrences. We thus use deep-sea temperature as a coarse indicator of mean global climate state.

See the electronic supplementary material, S1.1 for a further explanation on how observations were temporally binned, and how our covariates were standardized and transformed prior to analysis.

## (b) Model specifications

We used a discrete-time survival modelling framework to estimate how well we can predict extinction risk at 1 Myr time scales. **At its core, our model is a multilevel logistic regression with taxon age in millions of years as a varying intercept [33]. We considered four different models involving different permutations of covariate effects (fixed or time-varying) and historical covariates: covariate effects constant over time and no historical covariates included (model C); covariate effects allowed to vary over time but no historical covariates included (model V); covariate effects constant over time and historical covariates included (model CP); and covariate effects allowed to vary over time and historical covariates included (model VP).** The C and V models attempt to predict

based only on present state, whereas the CP and VP models allow for the possibility of non-Markovian behaviour by including change in state from the previous time increment.

**We always included species age at time of observation (i.e. observed prior duration) as a varying-intercept term. This factor may or may not contribute to differences in species extinction risk over time [14,34–38], but its inclusion in our model is critical to its nature as a survival model [33]. The effect of species age is allowed to vary by taxonomic group.**

Similarly, we included time of observation as an additional varying-intercept term to account for changes in average global extinction risk over time that are not related to the covariates included in this model. This varying-intercept is further allowed to vary by taxonomic group. This time-varying intercept term allows us to tease apart the differences in extinction risk associated with time of observation versus age since first observation. An important note is that for our V and VP models, the covariation between this varying-intercept and the varying slopes of our covariates is explicitly modelled (see the electronic supplementary material, S1.2).

See table 1 for further explanation of how the four models we considered differ from each other. A complete description of the statistical model used in this analysis is available in the electronic supplementary material, S1.2. Additionally, the full description of how these models were implemented and coded, including choice of priors, is available in the electronic supplementary material, S1.2.

## (c) In-sample and out-of-sample forecasting

We are interested in model performance (i.e. forecasting) in two distinct contexts: in-sample performance, and out-of-sample predictive performance.

In-sample forecasting is a posterior predictive check in that we are estimating our model's ability to correctly classify the data to which it was fitted. Posterior predictive checks are a type of sensitivity analysis because we are checking the quality of the model's fit to the data. If our models have poor in-sample forecasting performance, then they are not adequate descriptors of the data and will most likely make poor out-of-sample predictions. In-sample forecasting measures,

however, are not the same as understanding our models' ability to forecast future extinctions or if our models are overfit to our data and produce biased out-of-sample estimates [39].

We are particularly interested in understanding how well our model forecasts extinction probability of data from the future that were not used in fitting the model (out-of-sample data). To quantify our ability to forecast species' extinction risk, we estimated average out-of-sample forecasting performance using fivefold time-series cross-validation. For time-series data, the folds (data partitions) are approximately equal segments of time. Each fold represents a sequence of time points. With 63 time points, each of the five folds represents approximately 13 Myr time increments. It is important to bear in mind, however, that each time increment includes many hundreds to thousands of individual observations.

$k$ -fold cross-validation for time series follows a specific sequence of procedures [39–41]. Prior to cross-validation, the data is divided into  $k$  nearly even segments or folds—for a time series, this means the data is divided into  $k$  continuous sequences. Next, the model is fitted to the first fold (time segment), and the posterior estimates of that fit are then used to forecast the extinction probability of the second fold (i.e. the future). Then the model is fitted to the combined first and second folds, and the posterior estimates of that fit are used to forecast the extinction probability of the third fold. This process continues until  $k - 1$  folds are included in fitting the model and the final fold is predicted from this model. When combined, the results from these forecasts are then combined to yield our estimate of expected out-of-sample performance. In five-fold cross validation, the data are divided into five folds and the cross-validation procedure yields predictions for four of the folds.

Cross-validation is a procedure for estimating a model's expected out-of-sample error. Information criteria such as Akaike information criterion (AIC) or Watanabe-Akaike information criterion are approximations of out-of-sample predictive error as estimated by cross-validation [39,42]. Cross-validation implicitly takes into account model complexity because when a model is overfit to its data, out-of-sample predictions will be biased and inaccurate [39]. A high degree of similarity between out-of-sample and in-sample estimates indicates that the model is not overfit to the data (though it is not necessarily an adequate descriptor of the data). Cross-validation is preferable to simple metrics such as AIC because instead of a single value, it produces an entire posterior distribution of estimates.

The relative adequacy of the four model variants was compared using the area under the receiver operating characteristic curve (AUC) [43,44]. This measure is commonly used to measure the performance of classification models as it has the desirable characteristic of comparing the model's true positive rate with its false positive rate, as opposed to accuracy which only considers true positives. AUC ranges between 0.5 and 1, with 0.5 indicating no difference in classification from random and 1 indicating perfect classification. AUC can be interpreted as the probability that our model correctly ranks the relative extinction risks of a randomly selected extinct-extant species pair [43,44]. AUC values of approximately 0.8 or greater can be considered 'good' [45], so we consider values between 0.7 and 0.8 as 'fair,' and values between 0.6 and 0.7 as 'poor.'

To reiterate, the primary focus of this study is on understanding how well our models forecast future extinction events by comparing our in-sample and out-of-sample forecast estimates. Posterior estimates for the regression coefficient estimates from our VP model (table 1) are available in our electronic supplementary material, S2.

See our repository at <https://github.com/psmits/trident> for full code details. The entire analysis was coded in R and uses tidyverse and tidyverse adjacent tools such as `dplyr` [46], `purrr` [47]

and `tidybayes` [48]. All of our models were written using the `brms` [49,50] R package, which implements Stan-based Bayesian models which are fitted via Hamiltonian Monte Carlo [51].

### 3. Results

#### (a) In-sample forecasting adequacy

Comparison between the posterior distributions of in-sample AUC for each of the four models demonstrates that the parameter-rich model VP has the greatest median in-sample AUC when compared with the other three models, while there is substantial overlap in the posterior distributions of the forecasts from the other three models (figure 1a).

However, the actual difference in forecast AUC result between the VP model and the other three models is extremely small (0.01 AUC unit), and all of the in-sample AUC estimates from our models are concentrated between AUC values of 0.775 and 0.795 (figure 1a). This result suggests that the difference in performance between these models might be so small that there is no practical benefit of the VP model over the other three. Ultimately, determining which of these models produces the best forecasts for future extinctions requires comparing these in-sample results to our out-of-sample results (see below).

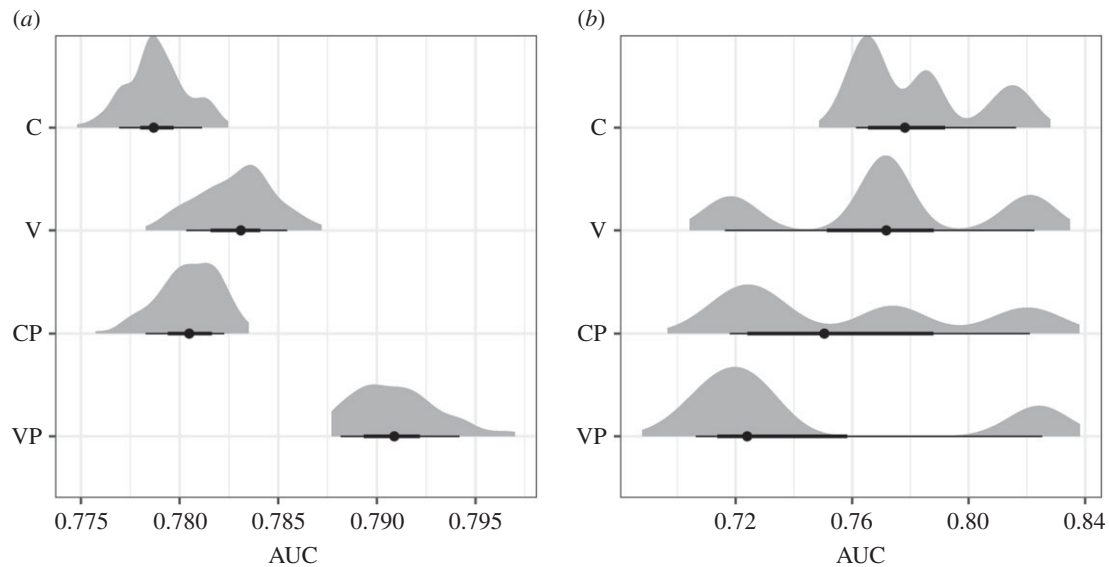
In-sample forecasts from the four models over time are broadly similar among taxonomic groups (figure 2). In-sample forecasts for diatoms are the weakest of the four taxonomic groups as all four models have several intervals with no predictive power (AUC not significantly greater than 0.5). The best in-sample forecast results are for radiolarians, for which all models have at most 1 interval with little predictive power. The pattern of high and low in-sample forecast performance is broadly similar among the four models.

#### (b) Out-of-sample forecasting performance

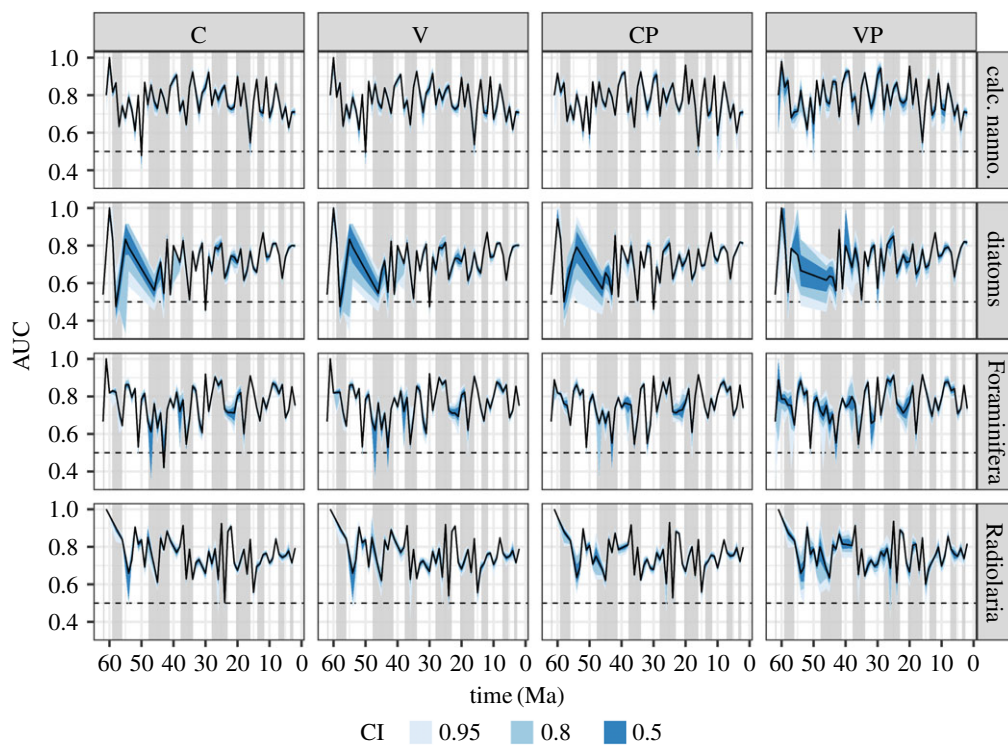
Out-of-sample forecast AUC estimates, based on fivefold cross validation [40,41], exhibit a broader range than in-sample estimates, with AUC ranging between approximately 0.7 and 0.85 (figure 1a,b). While the VP model has the best in-sample forecasting performance (figure 1a) this model performs poorly at out-of-sample forecasting compared to the other models (figure 1b). The poor out-of-sample performance suggests that this most complex model is overfit and that one of the simpler models would be preferable for predicting future extinctions. Thus the models that include both historical covariates (e.g. change in geographical range) and time-varying effects produce biased extinction forecasts. Interestingly, models that include either historical covariates but assume constant effects (CP), or do not include historical covariates but include time-varying effects (V), perform similarly when forecasting future extinction events (figure 1).

As noted above, there were some time intervals in which in-sample forecasts were no better than random (figure 2). Such intervals are generally much rarer for out-of-sample forecasts. The major exception to this pattern are diatoms, which have at least one time interval for all four models in which the median AUC of the out-of-sample forecasts were no better than random. The only other group for which median posterior predictive estimate of out-of-sample AUC reaches 0.5 is calcareous nannoplankton, and then only with the V model.





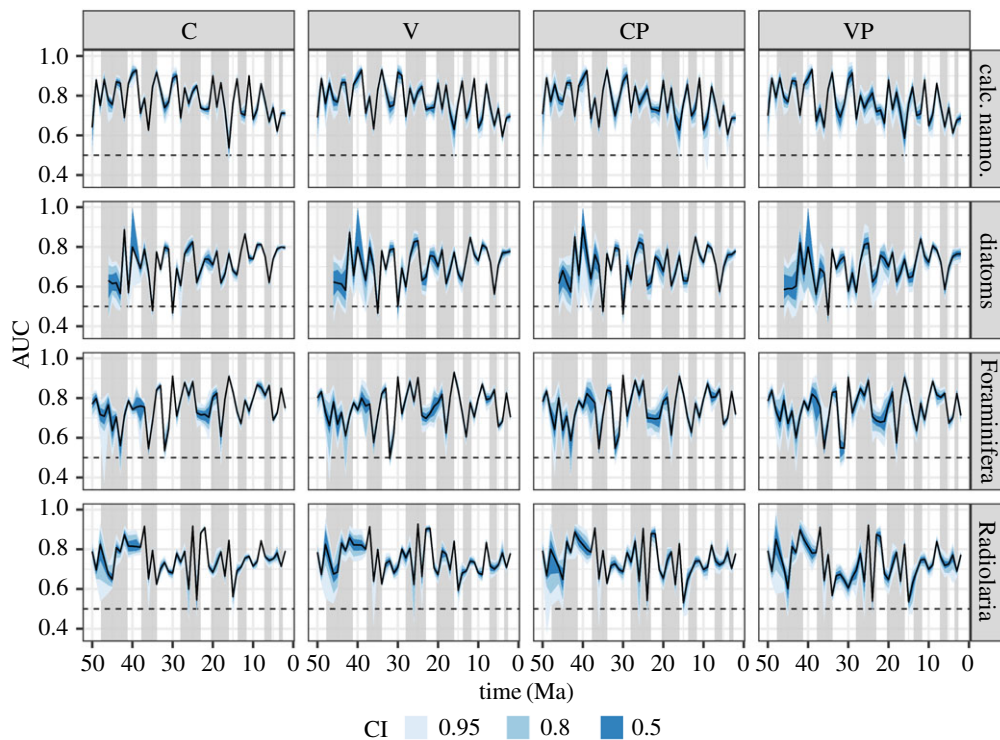
**Figure 1.** Comparisons of measures of model performance for both in-sample (a) and out-of-sample (b) cross-validation. The area under the receiver operating characteristic curve (AUC) was calculated for each model. These estimates are calculated from the model's posterior predictive distribution (a) or from predictions made to new data (b), respectively. Marked below the posterior distributions are the median AUC and 50% and 80% posterior intervals for all observations in the dataset. Models with higher AUC values indicate better performance over models with lower AUC values. AUC is bounded between 0.5 and 1. See table 1 for an explanation of the four models (C, V, CP, VP).



**Figure 2.** Understanding model adequacy over time and taxonomic group by comparing in-sample forecasting performance measured by AUC for each of the four models. These estimates reflect each model's fit to the various taxonomic groups over time. The black line corresponds to the median AUC value, while the envelopes correspond to multiple credible intervals (CI) as indicated in the legend. In all cases, higher AUC values indicate greater predictive performance versus lower AUC values. The grey intervals mark the geological ages of the Cenozoic. See table 1 for a description of each of the four models (C, V, CP, VP).

We compared the difference in AUC estimates from the out-of-sample forecasts to the AUC estimates from in-sample forecasts by subtracting the in-sample AUC estimates from the out-of-sample AUC estimates (figure 4); a difference in AUC close to 0 indicates complete congruence between the in-sample and out-of-sample forecasts. A positive difference indicates that out-of-sample forecasts actually outperform in-sample forecasts, whereas a negative difference indicates poorer out-of-sample performance than

in-sample forecasts. Divergences between out-of-sample and in-sample forecasts are rare and tend not to cluster in time, consistent with the broad visual congruence between the in-sample and out-of-sample performance (figures 2 and 3). An example multimillion year pattern indicating significantly poorer out-of-sample forecast performance than in-sample forecast performance is for Radiolaria based on the VP model between 35 Ma and approximately 28 Ma (figure 4). There exist similar periods of worse



**Figure 3.** Comparison of the four models' ability to forecast future extinction events as measured by out-of-sample AUC values over time as aggregated by taxonomic group for each of the four models. The AUC of the individual million year intervals within each fold is plotted to highlight the heterogeneity in performance within and between folds. This presentation decomposes each of the 12 Myr folds by each of the taxonomic groups into the predictions made for each of the million-year intervals. The black line corresponds to the median AUC estimate, with the envelopes corresponding to multiple credible intervals (CI) as indicated in the legend. The grey intervals mark the geological ages of the Cenozoic. See table 1 for a description of each of the four models (C, V, CP, VP).

out-of-sample forecasting performance for other combinations of taxonomic group and predictive model, for example, the CP model has worse out-of-sample forecasting than in-sample for the last 5 Myr of the Cenozoic. In general, however, most out-of-sample and in-sample forecasts are almost identical.

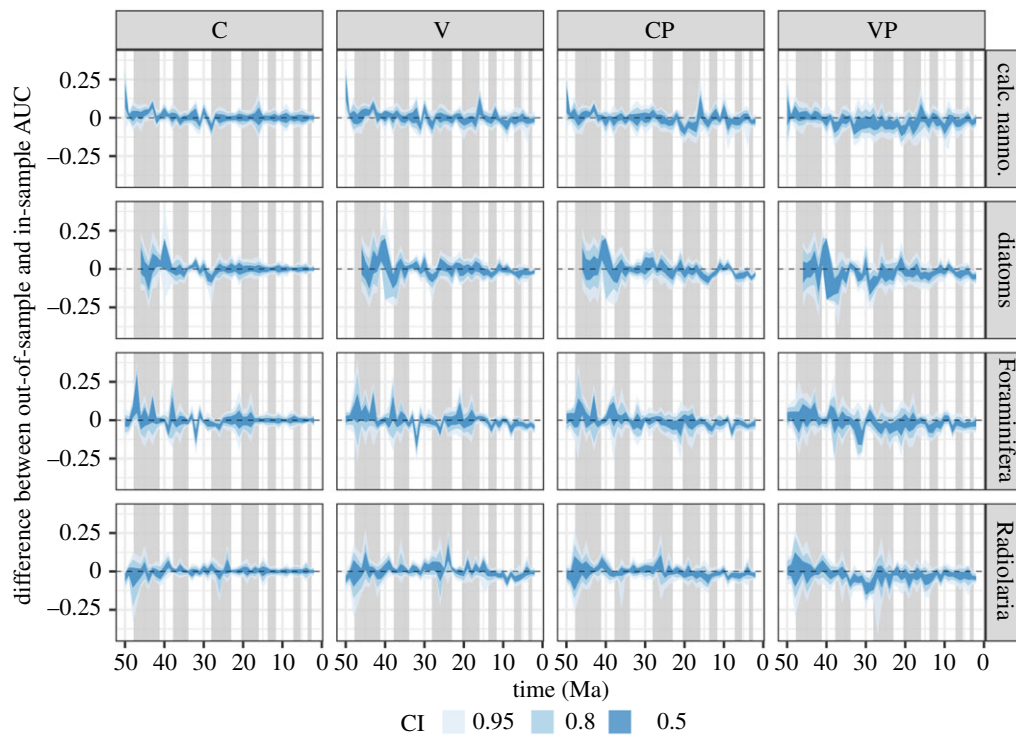
## 4. Discussion

We find that all of our models are expected to correctly forecast which species of a randomly selected extinct-extant pair is more likely to go extinct between 70% to 80% of the time (figure 1*b*). These results confirm that past extinction patterns can provide valuable information about which extant species are most threatened with extinction in the near geological future, and that some historical information does not degrade our ability to forecast future extinction risk. To reiterate, all of our models are fitted to past extinction events from the Cenozoic and covariates like geographical range or global climate state are associated with these extinctions; our models are conditioned on the past. Some of our models, however, encode historical information such as how covariate effects have varied over time or the change in a species geographical range over time (table 1).

Three of the four models we evaluated are practically identical in their ability to make in-sample and out-of-sample forecasts. Although the in-sample AUC estimates differ between models, all of these estimates are in a narrow range of possible AUC values (figure 1*a*). Our VP model had the best in-sample forecasting results including the historical covariates and allows all covariate effects to vary over time. However, the out-of-sample forecasts from this model are

biased, indicating that it is overfit to the data (figure 1*b*). The CP model that includes historical covariates such as geographical range trajectory yields out-of-sample forecasts with nearly identical results to the V model that allows covariate effects to vary over time but does not include historical covariates.

While all of our models are conditioned on past extinction data from the Cenozoic planktonic microfossil record, we used multiple approaches to encode historical information. Models that include the historical covariates (e.g. change in geographical range) but do not allow covariate effects to vary over time (i.e. the CP model) encode the past explicitly but assumes that covariates effects are constant over time. Allowing covariate effects to vary over time, as with our V model, does not explicitly encode the history of individual species into our model but instead models the history of how covariate effects have varied over time, thus implicitly encoding historical information about the species in our models. By modelling the variation in covariate effects over time, forecasts made for future extinction events are conditioned on a wider range of potential covariate effects which can improve model flexibility when forecasting extinction in the novel environmental conditions we might expect in the future. Comparing out-of-sample forecast results indicates that these approaches yield approximately equal forecasting performance (figure 1*b*). Our results contrast somewhat with those of Kiessling & Kocsis [2], who found that including historical range trajectories significantly and substantially improved the performance of extinction risk forecasts. Whether this reflects differences in analytical methodology or study system (planktonic microfossil taxa versus marine macroinvertebrates) is not clear and is worth further investigation.



**Figure 4.** Comparing out-of-sample and in-sample forecasts. Congruence between in-sample and out-of-sample indicates that a model is not necessarily overfit to the data. This value is calculated as the values presented in figure 3 minus those values presented in figure 2. A difference close to 0 indicates complete congruence between in-sample and out-of-sample forecasts, while a positive difference indicates that the out-of-sample forecasts are actually higher performing than the in-sample forecasts, and a negative difference indicates poorer out-of-sample performance than in-sample forecast. See table 1 for a description of each of the four models (C, V, CP, VP).

The relative quality and consistency between in-sample and out-of-sample forecasting performance for three of the four models we considered is encouraging given that these estimates are based on very limited biological and environmental information about the studied taxa. Even our most complex models only account for a few simple aspects of geographical range, prior history and phylogenetic group. The principal reason we were not able to include more biological information in the models used here is because we lack additional life history or ecological information for many of the marine micro- and nannoplankton included in this study. Foraminifera are an exception to this problem as aspects of life history, ecology and physiology are known for many foraminiferan species [3]. However, comparable information does not exist for all Foraminifera species, nor does this type of data exist for the other three taxonomic groups studied here. Future analyses including this type of information and focused more narrowly on the Foraminifera may be informative.

An extremely important caveat, of course, is that human impacts may substantially alter present and future extinction risk dynamics relative to the average Cenozoic state, so that the future may become less predictable than it has been in the past [1,11]. The CP model assumes that the effects of the historical covariates are constant through time, but given growing evidence that human impacts substantially alter extinction risk dynamics [1,5,11], this assumption may not be valid and may limit or bias our ability to predict extinction in truly novel environmental regimes. Thus, it might be preferable to use a model similar to our V model which allows extinction risk and selectivity to vary over time. For this reason, while our CP and V models yield

similar out-of-sample forecasts, we believe the V model offers more practical benefits for predicting extinction risk in future, anthropogenically impacted environments.

On a related note, it is notable that there are no obvious consistent changes in average model performance during episodes of climatic environmental change such as the mid-Eocene to early Oligocene figure 1*b*, although this interval is characterized by elevated extinction among several planktonic groups [52–54]. This suggests that relationships between the simple covariates included in our models and extinction risk were relatively stable though this interval. However, it is likely that more complex models accounting for other aspects of ecology and biogeography might exhibit reduced forecasting performance during this and other climate change episodes.

In summary, our results suggest that models trained on prior extinction/survival patterns do modestly well at predicting relative extinction probability of randomly selected species pairs based on a small number of simple taxonomic, geographical and historical predictors. Although a model that includes historical covariates such as change in geographical range and change in climate between observations while also allowing covariate effects to vary over time performs best at in-sample prediction, this model is overfit to our data and produces less accurate out-of-sample forecasts than our three less complex models. The remaining three models yield nearly equivalent out-of-sample forecasts. The results of this simple exercise suggest that conservation decisions could indeed be bolstered by including fossil data. Additionally, including historical information via explicit modelling of historical covariate effects or modelling how covariate effects have changed over time does not



diminish and may ultimately improve our ability to forecast future extinctions.

**Data accessibility.** Additional data is available at <https://github.com/psmits/trident>.

**Competing interests.** We declare we have no competing interests.

**Funding.** This work was funded by a David and Lucile Packard Fellowship to S.F.

**Acknowledgements.** We thank Jonathan Payne, Thomas Ezard, Erin Saupe and Peter Roopnarine for useful discussions. Wolfgang Kiesling, Matthew Clapham and an anonymous reviewer provided thoughtful reviews that strengthened the manuscript considerably.

## References

- Finnegan S *et al.* 2015 Paleontological baselines for evaluating extinction risk in the modern oceans. *Science* **348**, 567–570. (doi:10.1126/science.aaa6635)
- Kiesling W, Kocsis ÁT. 2016 Adding fossil occupancy trajectories to the assessment of modern extinction risk. *Biol. Lett.* **12**, 20150813. (doi:10.1098/rsbl.2015.0813)
- Ezard THG, Aze T, Pearson PN, Purvis A. 2011 Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science* **332**, 349–351. (doi:10.1126/science.1203060)
- Knoll AH, Bambach RK, Payne JL, Pruss S, Fischer WW. 2007 Paleophysiology and end-Permian mass extinction. *Earth Planet. Sci. Lett.* **256**, 295–313. (doi:10.1016/j.epsl.2007.02.018)
- Payne JL, Finnegan S. 2007 The effect of geographic range on extinction risk during background and mass extinction. *Proc. Natl Acad. Sci. USA* **104**, 10 506–10 511. (doi:10.1073/pnas.0701257104)
- Payne JL, Bush AM, Chang ET, Heim NA, Knope ML, Pruss SB. 2016 Extinction intensity, selectivity and their combined macroevolutionary influence in the fossil record. *Biol. Lett.* **12**, 20160202. (doi:10.1098/rsbl.2016.0202)
- Smits PD. 2019 How macroecology affects macroevolution: the interplay between extinction intensity and trait-dependent extinction in brachiopods. *bioRxiv* 523811.
- Wang SC, Bush AM. 2008 Adjusting global extinction rates to account for taxonomic susceptibility. *Paleobiology* **34**, 434–455. (doi:10.1666/07060.1)
- Foote M. 2006 Substrate affinity and diversity dynamics of Paleozoic marine animals. *Paleobiology* **32**, 345–366. (doi:10.1666/05062.1)
- Harnik PG. 2011 Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proc. Natl Acad. Sci. USA* **108**, 13 594–13 599. (doi:10.1073/pnas.1100572108)
- Harnik PG *et al.* 2012 Extinctions in ancient and modern seas. *Trends Ecol. Evol.* **27**, 608–617. (doi:10.1016/j.tree.2012.07.010)
- Harnik PG, Simpson C, Payne JL. 2012 Long-term differences in extinction risk among the seven forms of rarity. *Proc. R. Soc. B* **279**, 4969–4976. (doi:10.1098/rspb.2012.1902)
- Peters SE. 2008 Environmental determinants of extinction selectivity in the fossil record. *Nature* **454**, 626–629. (doi:10.1038/nature07032)
- Smits PD. 2015 Expected time-invariant effects of biological traits on mammal species duration. *Proc. Natl Acad. Sci. USA* **112**, 13 015–13 020. (doi:10.1073/pnas.1510482112)
- Alroy J. 2000 New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* **26**, 707–733. (doi:10.1666/0094-8373(2000)026<0707:NMFQMP>2.0.CO;2)
- Alroy J. 2010 The shifting balance of diversity among major marine animal groups. *Science* **329**, 1191–1194. (doi:10.1126/science.1189910)
- Alroy J. 2014 Accurate and precise estimates of origination and extinction rates. *Paleobiology* **40**, 374–397. (doi:10.1666/13036)
- Alroy J, Koch PL, Zachos JC. 2000 Global climate change and North American mammalian evolution. *Paleontol. Soc.* **26**, 259–288. (doi:10.1666/0094-8373(2000)26[259:gccana]2.0.co;2)
- Alroy J *et al.* 2001 Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266. (doi:10.1073/pnas.111144698)
- Foote M. 2001 Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology* **27**, 602–630. (doi:10.1666/0094-8373(2001)027<0602:ITPOPO>2.0.CO;2)
- Lazarus D. 1994 Neptune: a marine micropaleontology database. *Math. Geol.* **26**, 817–832. (doi:10.1007/BF02083119)
- Spencer-Cervato C. 1999 the Cenozoic deep sea microfossil record: explorations of the DSDP/ODP sample set using the Neptune database. *Palaeontol. Electron.* **2**, 4–286.
- Foote M. 2007 Symmetric waxing and waning of marine invertebrate genera. *Palaeobiology* **33**, 517–529. (doi:10.1666/06084.1)
- Jablonski D. 2008 Species selection: theory and data. *Annu. Rev. Ecol. Evol. Syst.* **39**, 501–524. (doi:10.1146/annurev.ecolsys.39.110707.173510)
- Jablonski D, Hunt G. 2006 Larval ecology, geographic range, and species survivorship in cretaceous mollusks: organismic versus species-level explanations. *Am. Nat.* **168**, 556–564. (doi:10.1086/507994)
- Jablonski D, Roy K. 2003 Geographical range and speciation in fossil and living molluscs. *Proc. R. Soc. B* **270**, 401–406. (doi:10.1098/rspb.2002.2243)
- Liow LH, Stenseth NC. 2007 The rise and fall of species: implications for macroevolutionary and macroecological studies. *Proc. R. Soc. B* **274**, 2745–2752. (doi:10.1098/rspb.2007.1006)
- Liow LH, Skaug HJ, Ergon T, Schweder T. 2010 Global occurrence trajectories of microfossils: environmental volatility and the rise and fall of individual species. *Paleobiology* **36**, 224–252. (doi:10.1666/08080.1)
- Foote M, Crampton JS, Beu AG, Marshall BA, Cooper RA, Maxwell PA, Matcham I. 2007 Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science* **318**, 1131–1134. (doi:10.1126/science.1146303)
- Pigot AL, Owens IP, Orme CDL. 2012 speciation and extinction drive the appearance of directional range size evolution in phylogenies and the fossil record. *PLoS Biol.* **10**, e1001260. (doi:10.1371/journal.pbio.1001260)
- Boyle D. 2017 GeoRange: calculating geographic range from occurrence data. See <https://CRAN.R-project.org/package=GeoRange>. R package version 0.1.0.
- Cramer BS, Miller KG, Barrett PJ, Wright JD. 2011 Late Cretaceous–Neogene trends in deep ocean temperature and continental ice volume: reconciling records of benthic foraminiferal geochemistry ( $\delta^{18}O$  and Mg/Ca) with sea level history. *J. Geophys. Res.: Oceans* **116**, 1–23. (doi:10.1029/2011JC007255)
- Tutz G, Schmid M. 2016 *Modeling discrete time-to-event data*. New York, NY: Springer International Publishing.
- Crampton JS, Cooper RA, Sadler PM, Foote M. 2016 Greenhouse–icehouse transition in the Late Ordovician marks a step change in extinction regime in the marine plankton. *Proc. Natl Acad. Sci. USA* **113**, 1498–1503. (doi:10.1073/pnas.1519092113)
- Ezard THG, Pearson PN, Aze T, Purvis A. 2012 The meaning of birth and death (in macroevolutionary birth-death models). *Biol. Lett.* **8**, 139–142. (doi:10.1098/rsbl.2011.0699)
- Finnegan S, Payne JL, Wang SC. 2008 The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. *Paleobiology* **34**, 318–341. (doi:10.1666/07008.1)
- Liow LH *et al.* 2011 Pioneering paradigms and magnificent manifestos: Leigh Van Valen's priceless contributions to evolutionary biology. *Evol. Int. J. Org. Evol.* **65**, 917–922. (doi:10.1111/j.1558-5646.2011.01242.x)
- Van Valen L. 1973 A new evolutionary law. *Evol. Theory* **1**, 1–30.
- Hastie T, Tibshirani R, Friedman J. 2009 *The elements of statistical learning*, 2nd edn. New York, NY: Springer.
- Arlot S, Celisse A. 2009 A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79. (doi:10.1214/09-SS054)



41. Bergmeir C, Hyndman RJ, Koo B. 2018 A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.* **120**, 70–83. (doi:10.1016/j.csda.2017.11.003)
42. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013 *Bayesian data analysis*, 3rd edn. Boca Raton, FL: Chapman and Hall.
43. Fawcett T. 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874. (doi:10.1016/j.patrec.2005.10.010)
44. Mason SJ, Graham N. 2002 Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Q. J. R. Meteorol. Soc.* **128**, 2145–2166. (doi:10.1256/003590002320603584)
45. Tang W, He H, Tu XM. 2012 *Applied categorical and count data analysis*. Boca Raton, FL: CRC Press.
46. Wickham H, François R, Henry L, Müller K. 2018 dplyr: a grammar of data manipulation. See <https://CRAN.R-project.org/package=dplyr>. R package version 0.7.8.
47. Henry L, Wickham H. 2018 purrr: functional programming tools. See <https://CRAN.R-project.org/package=purrr>. R package version 0.2.5.
48. Kay M. 2018 Tidybayes: Tidy Data and Geoms for Bayesian Models. See <http://mjskay.github.io/tidybayes/>. R package version 1.0.3.
49. Bürkner P-C. 2017 brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28. (doi:10.18637/jss.v080.i01)
50. Bürkner P-C. 2018 Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395–411. (doi:10.32614/RJ-2018-017)
51. Team SD. 2017 Stan Modeling Language Users Guide and Reference Manual. See <http://mc-stan.org>.
52. Kamikuri S-i, Wade BS. 2012 Radiolarian magnetobiochronology and faunal turnover across the middle/late Eocene boundary at Ocean Drilling Program Site 1052 in the western North Atlantic Ocean. *Mar. Micropaleontol.* **88–89**, 41–53. (doi:10.1016/j.marmicro.2012.03.001)
53. Prothero DR. 1994 The late Eocene-Oligocene extinctions. *Annu. Rev. Earth Planet. Sci.* **22**, 145–165. (doi:10.1146/annurev.ea.22.050194.001045)
54. Wade BS, Pearson PN. 2008 Planktonic foraminiferal turnover, diversity fluctuations and geochemical signals across the Eocene/Oligocene boundary in Tanzania. *Mar. Micropaleontol.* **68**, 244–255. (doi:10.1016/j.marmicro.2008.04.002)