

Author's response

Occupancy history influences extinction risk of fossil marine microplankton groups

We would like to again thank the Editor for the work on this manuscript, as well as both reviewers for the insightful feedback. We address each point in more detail below, but a short summary of the key changes follows:

- Several of the functions and scripts have been reorganized and/or rewritten to make them easier to follow and more computationally-efficient.
- The sampling term was removed from the models altogether to allow for more direct comparison with extant species models.
- Stepwise model selection was used to select model formulas based on AIC values.
- Figure 1 was updated to reflect a correction in the flow of time.
- Table 2 and Table 3 were created to summarize each selected model.
- Data was tested for multicollinearity (the results of which are now in the supplement).
- Models were run with a “paired-cell” approach to determine the impact of spatial sampling biases on the results. These results are reported in the supplement.
- Some of the data set descriptive statistics were updated to reflect a correction in the trimming script.
- Mixed effect modeling was explored at the suggestion of Reviewer 2, the results of which are now in the supplement.
- Various discussion points were updated based on reviewer feedback
- Various stylistic points were updated based on reviewer feedback.

RC1:

Overview

This is an intriguing manuscript of a rich data set appropriately analyzed with multivariate logistic regression. It makes a significant point for four major planktonic groups: that change in occupancy is a good predictor of extinction risk. It substantially adds to the ecological literature on extinction risk, emphasizing a factor that can be addressed only through the fossil record. As

the authors point out, the importance of occupancy for extinction risk has been shown for marine invertebrates. Still, it is encouraging to see that change in occupancy has importance that extends to the plankton and is likely to matter more broadly.

Response: We thank the reviewer for this encouraging feedback, as well as the following constructive suggestions.

The manuscript is well-written, and the figures and tables are well-constructed. I have a few general points that should be addressed before publication, plus a few minor comments. In short, the authors should:

* Simplify the models so that they include only statistically significant terms.

Response: We have simplified the models accordingly.

* Present a table of model coefficients for each 16 taxon/time-bin combination, displaying only the statistically significant fitted coefficients.

Response: We have taken the Reviewer's advice, and created the suggested table showing coefficients for each model (Table 3).

* Present a table of LMG relative importance for each of the statistically significant coefficients, as well as the overall percent variance explained for each of the 16 taxon/time-bin combinations.

Response: We have also created a table showing the LMG values for each statistically significant coefficient, and overall percent variance explained (Table 2).

* Update the text to emphasize patterns in LMG relative importance across the models.

Response: We have updated the text accordingly.

* Delete the tables (2 and 3) and figures (3 and 4) that are no longer needed.

These are easily accomplished, so I consider these moderate revisions. I would be happy to review a revised manuscript.

Response: We thank the Reviewer for taking the time to review this manuscript.

Principal comments

Multiplicative vs. additive models

There's a mixup about the meaning of the + and * symbols in specifying models in R, and this is not unique to the authors; many are misled by these symbols. These symbols are overloaded (in programmer speak); that is, in the context of specifying a linear model, they do not have their commonplace meanings of add and multiply. In R's formulation of a linear model, $y \sim x + z$

means "y as a function of x and z," not "y as a function of x plus (or added to) z." Likewise, $y \sim x * z$ means "y as a function of x and z and the interaction x and z," not "y as a function of x times (or multiplied by) z."

Confusing these meanings has led the authors to describe some models as additive and others as multiplicative; instead, these should be described as "without interactions" and "with all interactions." This is a small matter that can be quickly dealt with; I raise it because it feeds into the following more critical comments on model selection and reporting.

Response: Indeed, there was a mixup in the previous version of our manuscript, but only in terminology and not in meaning. We have updated the text following the suggestions.

Model selection

A surprising result of this study is that the full-interaction model (extinction $\sim o * oc * s$, where o is occupancy, oc is occupancy change, and s is sampling) is selected in 15 out of the 16 group/bin-size combinations (Table 2). When alternative models — Table S2, where 16/16 favor the full-interaction model, and Table S4, where 13/16 favor the full-interaction model — are considered, this rises to 44/48 (92%) of all models. This full-interaction model (extinction $\sim o * oc * s$) expands to the formula extinction $\sim o + oc + s + o:oc + o:s + oc:s + o:oc:s$ (each of the terms with a colon indicates the interaction of predictor variables), so it is also called the saturated model.

The fact that the saturated model was selected in 92% of the cases is highly unusual, and I've never come across such a situation in my work, including all the student projects in my data analysis class. By selecting these models, the authors are, in effect, claiming that all seven terms (o, oc, s, o:oc, o:s, oc:s, and o:oc:s) are statistically significant (i.e., the coefficients are demonstrably non-zero). That's what is unusual; seldom are all the interaction terms significant, and often, one or more non-interaction terms are not significant.

Examining the supplemental material gives a glimpse of what is going on. For example, using the foraminifera / 1 m.y. combination, where the full-interaction ("multiplicative" in table 2) model was selected, a summary of the model [using `summary(foram_1000_mod1)`] shows that four of the predictors are not statistically significant (o, o:oc, o:s, o:oc:s). The model must be simplified to include only the significant terms. Although one might be tempted to discard all the non-significant terms at this step, that's inadvisable because removing one term can cause another term to become significant. One must use model simplification so that significant terms are included, and non-significant terms are excluded (see the excellent coverage in Crawley's *The R Book*).

With seven predictor variables, exploring all possible (128) combinations manually is not feasible. The `stepAIC()` function provides a way to simplify the process. It starts with the saturated model (e.g., $ext \sim o * oc * s$), removes non-significant terms using AIC, and stops when a simplified model containing only statistically significant (i.e., demonstrably non-zero) coefficients is left. Calling `summary()` on that model shows the coefficients. What it produces is the simplest model that fits the data best. The authors can do this easily by calling three commands for each model; here's an example using the foraminifera/1 m.y. data:

```
saturatedModel <- glm(extinction ~ raw_prop_sampled * delta_1 * sampling, family =  
binomial(link = "logit"), data = F_1000000[[1]])
```

```
simplifiedModel <- stepAIC(saturatedModel)
```

```
summary(simplifiedModel)
```

Using AIC for model selection, as the authors have done, is good, but there are many more models to consider than the five the authors picked (o, oc, s, o+oc+s, and o*oc*os). StepAIC() facilitates finding the best models. The saturated model may be better than the other four in most cases, but this is likely because the four simple models miss significant predictors. Using stepAIC() also eliminates the need to test the four simple cases they do explicitly. Instead, start with the saturated model and let stepAIC() simplify it. If the best model is one of the authors' four simple cases (o, oc, s, o+oc+s), that will be discovered by using stepAIC().

Response: We updated the analytical pipeline to include the stepwise model selection, which does not substantially impact the final results of this study.

```
## Model reporting
```

The results of the models need to be reported, something that's largely missing from the current manuscript. For any mathematical model, two aspects need to be reported: the fitted coefficients and their effect sizes, that is, how necessary each coefficient is for modeling the response variable (extinction here). Although the authors touch on effect size partly by reporting some LMG relative importance values, they don't do it comprehensively, which makes comparing the relative importance of the predictor variables impossible. LMG relative importance is one of the most informative parts of the modeling because it's not just whether occupancy change matters but how much it matters relative to the other predictors (occupancy, sampling, and the various interactions).

Two tables are needed in the main body of the manuscript and not in the supplemental material; both should have 16 rows, one each for the 16 taxon/bin-size combinations. The first should have seven columns for the model coefficients for o, oc, s, o:oc, o:s, oc:s, and o:oc:s, obtained with summary(model). Where a coefficient is not statistically significant (i.e., not demonstrably different from zero), it should be left blank, or a dash should be used instead of its value. The caption should indicate that these are the statistically significant (i.e., statistically non-zero) coefficients for each model. Given the rampant confusion over what statistically significant means, it's advisable to add the "i.e., statistically non-zero" to clarify what significance means.

Response: We have added the suggested table to the manuscript summarizing the model coefficients for each of the statistically significant terms.

The second table is similar, but it should present effect size through LMG relative importance for each of the statistically significant model coefficients (all seven). Again, leave blanks or just a dash for the non-significant coefficients in any model. The table should also have an additional

column for overall explained variance or something similar to show how completely the entire model can explain extinction.

Response: We have also added a table summarizing LMG relative importance values. We have reported the raw LMG values for each term, instead of the standardized percentage with respect to the total of the LMG values for the three analyzed model terms.

These two tables will convey what's important about the models, and they allow Tables 2 & 3 and Figures 3 & 4 to be deleted as they become unnecessary. It is unnecessary to report the model AIC values in the main body of the text; they're helpful in model selection but not for comparing among the models owing to differences in sample size. They're also not as relevant as the model coefficients and relative importance.

Adding these tables will also necessitate new text to discuss the patterns in relative importance among the taxon/bin-size contributions. Those will be an essential take-home from this research and one of broad interest, especially for the comparative importance of occupancy, occupancy change, and sampling. If specific interactions are common and meaningful, that will also be worth discussion.

Response: We thank the Reviewer for these suggestions. The updated analysis does not substantially change our conclusions relating to temporal grain and siliceous versus calcareous plankton, and we have updated the text to discuss these nuances in more depth.

Why history might matter

It's fascinating the occupancy change matters, as has been shown for several marine invertebrate groups, especially so at the time scales of 100 kyr, 200 kyr, 500 kyr, and 1 myr. Organisms have no way of looking backward; their populations go up or down based on conditions in the moment, so the importance of occupancy change must lie in the state of the ecological system as a whole. In other words, if occupancy change matters, it must mean that something about the present ecological system is unfavorable, and that it has been persistently unfavorable.

It does make me wonder if the increasing importance of occupancy at longer time scales is just a reflection of a species being on the declining side of Foote/Liow occupancy curves. In other words, is all that's being said is that a species in a long-term decline is more likely to go extinct than one that is not? Even if it is, that's distinct from rarity, which ecologists focus on, and it's an insight that can be gained only from the fossil record.

Response: We have explored this idea further in the discussion section.

Minor matters, keyed to line numbers

48: replace "looking at" with "of".

Response: We have made this update.

52: By disappearance, is extinction meant? Extirpation, too?

Response: We meant global extinction and have updated the manuscript for clarity.

63: Insert a short explanation of what is meant by sampling. Is it the proportion of cells having some data?

Response: Although we no longer include the sampling term in the model, we have improved the explanation of the sampling metrics used to assess sampling completeness.

96: No change is needed, but I am surprised by the low proportion of extant species with usable records — at best 12%?!

Response: This value refers to the percent of species that are both extant and also have adequate spatial and temporal data. The remaining records are mostly extinct organisms (mostly with usable records). We have clarified this in Table 1.

121-123: No change is needed, but I am surprised that pacman profiling trims so much of the data, particularly from the top of the column. I would have guessed that last occurrences would be better constrained and more reliable in a core, given that downhole collapse affects first occurrences.

Response: We have double-checked this.

128–129: Most methods are admirably well-explained, with only two exceptions: simple completeness and three-timer completeness. Both could use a single-sentence definition for readers not familiar with them.

Response: We have added text to clarify the methods.

146–148: Delete simply and actually (lines 146–148, for example) from the methods; they're not needed.

Response: We have updated the manuscript accordingly.

206–211: I'm wondering about multicollinearity in this model, that is, the extent to which the three predictions (o, oc, and s) are correlated. A table of the correlations of the predictors for each data set could be added to the supplemental material. If these are large, though, the authors will need to address the implications of multicollinearity for model interpretation, which can be difficult.

Response: We calculated the Pearson correlation for occupancy and occupancy change (the sampling term was removed from the analysis as per the suggestion in RC2). The correlation values for the 16 taxon/bin-size combinations had a mean of 0.23, a median of 0.23, and a maximum of 0.32. Although there exists some correlation, it is relatively weak. We have included these results in the supplemental material.

266, 359: There should be a space between a value and its units.

Response: We have updated the manuscript accordingly.

314: It seems improbable that three of these values are all 0.053. Is that correct, or are there typos here?

Response: Due to updates in the analytical pipeline, the values reported in this table are now obsolete.

333, elsewhere: Results should be reported in the present tense, not the past tense.

Response: We have updated the manuscript accordingly.

541: Burnham and Anderson (2002) isn't cited. It's worth checking the other references as well.

Response: We have corrected this and have checked the other references.

Supplemental 116–117: replace "approaching the present" with "from the Pliocene to the Pleistocene".

Response: We have made this update

RC2

Thank you for the opportunity to read this manuscript, 'Occupancy history influences extinction risk of fossil marine microplankton groups,' by Smith, Kocsis, and Kiessling. I was extremely intrigued by the premise and results of the study. The prediction of modern extinction was interesting. The addition of robustness testing in analysis made this work particularly convincing to me, and as such, I have relatively few/minor comments.

I apologise for my delay returning these comments, and I hope that these suggestions prove helpful in refining the work.

Response: The authors appreciate the time and input of the Reviewer. We are grateful for the useful comments provided below.

Minor comments

1. The result that 'the temporal scale by which we analyze these data can influence our understanding of extinction risk in marine microplankton' stands out as a strong finding. I recommend adding this to the abstract.
 - **We have added this finding to the abstract.**
2. Line 47: Saulsbury et al. (2023) in PNAS is a recent paper to add. Although it is couched in terms of age-dependent extinction risk, it is equally about abundance-dependent and therefore range-size-dependent extinction risk. <https://doi.org/10.1073/pnas.2307629121>
 - **Response: We appreciate the paper recommendation, and have included it in the revised version of the manuscript.**

3. Line 47: I would also ask to rewrite the manuscript sentence on this line. It's unnecessary to claim the topic is understudied - I am convinced of its inherent importance regardless! To say range trajectory dynamics are 'sparingly' considered and 'many studies exclude [it]' feels as though it could slight those who have indeed worked on the topic.
 - **We have updated this to better reflect the current state of the research.**
4. Line 50: I suggest omitting this sentence (beginning 'increasing anthropogenic impact'), as it reads as a non sequitur. The topic is already well-justified as worthy of study without this.
 - **We have updated the manuscript accordingly.**
5. Line 60: I don't follow this sentence ('variations in the material...') and therefore recommend removing it for concision.
 - **Response: We have made the recommended change.**
6. Line 67: Rather than pose the possible results as a binary set of options ('whether the trajectory of geographic occupancy actually influences extinction risk [or not]'), I would reword this to reflect the nuance of your methods to investigate the degree to which geographic occupancy explains extinction risk.
 - **Response: We have reworded this to better reflect the nuances of our methods.**
7. Line 179: I assume the authors mean the final data table, i.e. dataset?
 - **Response: This is correct, we have updated for clarity.**
8. Line 279: Name specific examples of the 'various biotic events in the Cenozoic' that 'can be detected'.
 - **Response: We have updated the manuscript to include specific examples.**
9. Line 318: The first use of 'relative importance value' in this paragraph needs explaining – this is the relative importance of occupancy change to standing occupancy, correct?
 - **Response: That is correct: we reported the importance of the occupancy change term with respect to the sum of the importance values for all three of the analyzed model terms. Based on feedback from RC1, we have now reported the raw LMG values (without converting to a standardized percentage). We have made an effort to clarify what exactly is being reported.**
10. Line 348: What was the D^2 and the relative importance of the occupancy change term in this robustness-test version of analysis?
 - **Response: We have added these values to the text.**
11. Line 372: Changes in circulation and stratification are key influences on plankton biodiversity distributions, yes, but I am less confident to say they are THE key influences. There are additional influences beyond these two factors as well.
 - **Response: We have updated the text to reflect other influences.**
12. Line 374: There is a lot of contrary evidence to this. For instance, plankton lateral shifts are documented in two Nature papers out this year - Ying et al.

[<https://doi.org/10.1038/s41586-024-08029-0>] and Chaabane et al.

[<https://doi.org/10.1038/s41586-024-08191-5>] A separate factor that matters is edge effects: species at poles and the equator are at the edge of the world, climatically speaking, as there is nowhere for them to migrate that is colder than found at poles or warmer than found at equator, respectively. I would consider this geographic constraint different to a mobility/biological constraint.

- **Response: We have updated our discussion based on these comments and references.**

13. Line 443: There are some logical steps in this sentence that are currently missing and need to be spelled out (or remove the sentence). What does nutrient limitation have to do with this?

- **Response: We have updated the manuscript accordingly**

14. Line 478: Rather than ‘across the tree of life,’ I would say ‘across marine life’. Neither this work nor Kiessling and Kocsis (2016) analysed terrestrial species, which have substantially different migration capacities and population sizes.

- **Response: We have made this update in the text.**

15. Section 5 (Conclusion): While I agree historic data are relevant to informing future predictions of biodiversity loss, I am (regretfully) unconvinced conservation biologists would take the findings of this study as important to influence management decisions on ‘where to spend limited conservation funding in the future’—something that is sadly dictated more often by what is feasible to achieve in local sociopolitical contexts than what is forecast by scientific models. What about substituting the second sentence in this paragraph and instead mentioning how these findings are empirical contributions to help resolve open evolutionary theory about how range reductions end in extinction over the long timescales of species ‘lifetimes’, i.e. age-dependency of extinction rates?

- **Response: We appreciate the feedback, and have updated the conclusion to focus more on the broader implications to evolutionary theory to which this study contributes.**

16. Table 2: Were the top models always only (a) fully multiplicative or (b) fully additive? That is, were there never cases of a model such as $ex \sim oc * ch + p$? Or were such models not considered? This is hard to interpret based on how Table 2 is presented.

- **Response: They were not considered in the initial version of the manuscript (they were, however, in previous iterations of this study). Upon reading the feedback from both Reviewers, we updated the way we select model formulas and present the results. Now, various model combinations are assessed using the stepAIC() function.**

17. Cite version numbers for all software packages: NSBcompanion (line 76), divDyn (line 131), taxize (line 221), and any others.

- **Response: We have added version numbers.**

Modelling suggestions

I have three suggestions about the analysis framework, none of which I expect to substantively change the findings but which may increase the informativeness (for the first), rigor (for the second), and accuracy (for the third).

My first observation and the one that would require more work is about the definition of 'completeness.' Section 2.2, beginning line 126, describes metrics that measure only temporal completeness. I agree range-through completeness is relevant and interesting as a parameter; I found myself repeatedly wondering what an additional measure of spatial completeness would show. For example, if the analysis were repeated with a count of global occupied grid cells (already used to derive proportional occupancy) as a measure of 'completeness,' what would be its explanatory power?

Response: Although we did examine temporal completeness, the Reviewer has correctly pointed out that we did not explicitly examine spatial completeness in this manuscript. We have re-run these models with and without a “paired-cell approach”, whereby we calculated the proportional occupancy and the change in occupancy using only the cells that were common to both bin_i and bin_{i+1} (see Kiessling and Kocsis 2016). This paired-cell approach was used to help account for variability in spatial sampling between temporal bins. With this approach, heavily (spatially) sampled intervals are trimmed down for each pairwise analysis such that only the common geographic cells between both time bins are considered as the denominator of the proportional occupancy when calculating standing occupancy or the change in occupancy between bins. The drawback to such an approach is that it reduces the amount of data that can be used to fit a model. Using this paired-cell approach shows minimal changes in the results and serves as a robustness test and check against biases caused by variations in spatial sampling. These results have been included in the supplementary material.

A different thought about the models is whether a mixed-effects framework might be appropriate. My understanding of the input data is that individual species contribute multiple observations to the dataset (from different points in their 'lifespan'). Is it possible to add species identity as a random effect, reducing the estimated degrees of freedom to account for pseudo-replicate observations, and still calculate deviance? Maybe adding a random effect to the GLMs will make them fail to converge – it might not be feasible – but then again it might be simple to implement. I wouldn't require this change but do think it worth investigating.

Response: We have re-run the models using a mixed-effects framework, and found that this update did not substantially change the LMG values, although several of the models did fail to converge. Given the updates in how we will report results (see the two tables requested in RC1), we did not apply the mixed-effects framework to the main analysis. Nonetheless, we see the utility of and rationale behind a mixed-effects framework, so we have included these results in the supplementary material.

Lastly, since the model used to predict modern taxa (section 2.7) couldn't include a three-timer sampling value, the entire model selection and model fitting process should omit three-timer completeness as a possible variable. Perhaps this was already done, but the paragraph beginning line 228 made it seem as though the process was to take the best-fit model and then leave out the completeness covariate when predicting on the modern data. It would be better to

find the best-fit model out of the model set that doesn't have this term to begin with, then make predictions.

Response: We appreciate the Reviewer pointing this out, and agree with this point. When fitting the models on extant data, the sampling term was not included at all (extant-only models were fit separately from full-data models). Given this suggestion, we have removed the sampling term completely from all models, thus making comparisons between extinct- and extant-species models more straightforward.