

Review of Hochfeld et al, “How sediment archives can improve model projections of marine ecosystem change”

Summary

This manuscript is a perspective piece outlining an approach to validate evolutionary ecosystem models using sediment archives. The authors discuss the limitations of current models and the advantages of using long-term integrated records such as sediments to constrain evolutionary changes. They introduce a framework in which this can be applied and discuss remaining challenges.

General comments

The main concept of the paper is well articulated and convincing. However, the actual application of the framework to real models and archives is not as clear. I suggest adding a section that contains a more quantitative, worked example with a specific model and sedimentary record, including details such as what model structural aspects and/or parameters were considered for calibration, what proxies were used to constrain the model, what adjustments were made upon inclusion of the proxy data, and how much it reduced uncertainty.

I would also like to see a more robust discussion of uncertainties in both models and proxies. For example, how does the age model uncertainty and temporal resolution of sediment archives affect derived rates of evolutionary change? Do different rates of preservation under different environmental conditions limit the conclusions you can make about relative abundance and evolutionary change?

A definition of precisely what you mean by “evolutionary ecosystem model” would also be helpful (and contrast with ecosystems that are not evolutionary). I suggest including a conceptual diagram of how these models generally work.

Specific comments

L26: Include more recent references for projected changes in marine ecosystems e.g. CMIP6 Fish-MIP results: Tittensor, D.P., Novaglio, C., Harrison, C.S. *et al. Nat. Clim. Chang.* **11**, 973–981 (2021). <https://doi.org/10.1038/s41558-021-01173-9>

L41: be more specific about “positive feedbacks” and “tipping points”

L46-48: “Since models hardly agree ...” This statement is overly simplistic; the validity of model projections depends on much more than inter-model agreement, for example model complexity, ability to match present-day observations, etc. It is the ability to

evaluate *which* model projections are valid that is the challenge. Give specific numbers of the range of predicted change.

L49-71: This paragraph would be a good place to discuss different types and complexities of ecosystem models and what is meant by evolutionary adaptation in this context

L49-50: “ ... to verify that all relevant processes are considered ...” This is extremely unlikely with current models and computing capacity; a better statement would be “*the most relevant processes*”

L60: “... allow ~~reconstructing~~ the reconstruction of long-term ...”

L68-69: what time resolution is needed to constrain evolutionary changes?

Section 2.1: add discussion of age model uncertainties (particularly marine reservoir ages for Radiocarbon) and how this could impact the ability to validate models with sediment archives

L117: “... allow ~~reconstructing~~ the reconstruction of ...”

L151: Don't begin a sentence with “also”

L161: “resting stage” should be described and/or defined when this term is introduced (also in Figure 1 caption)

L186 “This approach potentially allows one to obtain ...”

L203: “ ... and the mechanisms behind the changes”

L209: How would you derive grazing rates from these experiments?

L245: define “perennial time scales”

Section 3.1: This section could be expanded to talk more about different kinds of evolutionary models and how they integrate adaptation

L261: What are “evolution experiments”?

L330-331: be more specific about the location of the 5 deg C temperature anomaly during the Holocene thermal maximum, e.g. high latitudes rather than global. Why not other past warm intervals, such as the Last Interglacial or the Pliocene, which are potentially more extreme?

L345: How do you know that the problem is structural and not just due to a few parameters that are incorrect? This example needs more detail

L389-390: Assuming that the fittest individuals are the most abundant is questionable; “fitness” is an abstract concept and rates of preservation in the sediment archive can heavily bias derived abundance rates. How might this affect your conclusions?