

Biogeosciences Discuss., referee comment RC1
<https://doi.org/10.5194/bg-2021-202-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2021-202

Anonymous Referee #1

Referee comment on "Testing the effect of bioturbation and species abundance upon discrete-depth individual foraminifera analysis" by Bryan C. Lougheed and Brett Metcalfe, Biogeosciences Discuss., <https://doi.org/10.5194/bg-2021-202-RC1>, 2021

General comments

In this manuscript, Lougheed and Metcalfe use transient model outputs (Trace21k output as input for SEAMUS) to assess whether discrete-depth Individual Foraminifera Analysis (IFA) can faithfully reflect temperature distribution. Within the idealized model environment, the authors are able to simulate pre-depositional, post-depositional and post-retrieval processes that may affect the temperature distribution recorded by foraminiferal tests becoming a part of the sediment, including sea surface temperature (SST), foraminiferal abundance in response to SST, sediment accumulation rate, bioturbation, number of foraminifera picked (sample size), and machine error. They assess the sensitivity of IFA-derived SST distribution by varying the aforementioned parameters in the model environment. The output is of course best-case scenario – as the reality is a lot more chaotic and noisy. Despite this, the idealized simulations show that the IFA-derived SST distribution show extremely low reproducibility with the typical sample size adopted by users of IFA (50-100 picked specimens). The reproducibility is especially poor near the edge of the distribution – which is the region of interest to paleoceanographers. Another important finding is that varying species abundance in response to climate change may also bias IFA-derived reconstructions, and this bias cannot be avoided if one were to pick 10000 specimens for the IFA-based reconstruction.

IFA has become increasingly popular as a tool for reconstructing past climate variability, thus the scientific questions explored by the authors are timely and of broad appeal. The manuscript is clear, generally well-written and accessible even to readers who have no strong background in numerical modelling. I expect the paper to be of great interest to users of IFA and proxy system modelling, and to a lesser degree also to those who study foraminiferal ecology. The scope of the study also fits the remit of the journal. I find the conclusions convincing, but think that the paper may benefit from some clarification here and there, and more discussion on how to apply the knowledge derived from these idealized simulations to actual sediment records, or at least some concrete suggestions on what (not) to do when using IFA in reconstruction. After all, the community that will benefit the most from this paper is likely paleoceanographers who apply IFA (I certainly hope so), thus the more reason to make it as accessible as possible to this community. In

this regard, the reader could use some elaboration on what would be the minimum requirement in terms of SAR, sample size / number of specimen picked. Any regions where IFA would work nicely or should be avoided? Some comments on how realistic the idealized model output is, considering that one of the largest sources of uncertainty in IFA-derived SST distribution, i.e. the vertical migration of planktic foraminifera, is not considered in the simulation? I think it might also help maximize the impact of the work if the authors can make the outcome accessible in the form of a web GUI for users who are fluent in programming language – but this is more of a would-be-nice-to-have kind of suggestion.

That said, I am happy to recommend publication once these concerns have been addressed by the authors. Altogether this should amount to minor to moderate revision. Below I outline a few specific comments / suggestions that I hope the authors will find helpful in revising their manuscript.

Specific comments

Line 29-32: Most studies are based on 50-100 specimens, so I'd add a sentence saying under what conditions can this sample size yield meaningful reconstruction. The results clearly indicate that one would be safe if 10000 specimens are picked, but alas this is not something that is realistic. Even 500 specimens are not always possible if one tries very hard.

Line 155-156: "... differs in model execution..." please elaborate more in what way it is different that makes it suitable for use in this IFA experiment.

Line 163-165: This is a bit confusing. The model is run at monthly time-step, so the foraminifera in the model do not record daily temperature but only the monthly mean? Also, it would be helpful for the reader to follow the manuscript if the authors could provide more information on how the recording process is simulated in SEAMUS. Is the temperature value recorded by foraminiferal test an average of several weeks of daily temperature?

Line 168: 10000 foraminifera per cm of sediment (at a single site presumably) sounds a lot. Is this value based on some ecological studies? If yes please add the references here. I also wonder if this value affects the model output? Say, for example, if one were to assume that only 1000 foraminifera are produced per cm of sediment, how would the smaller number of foraminifera affects the simulation of SST distribution.

Line 255-260: Why the criterion of $r^2 > 0.6$ on top of $p < 0.05$? Any reason why both criteria are needed for this study? I note that $p < 0.05$ is a more commonly adopted criterion in paleoclimatology when assessing correlation between time series. How does the result

change when only the $p < 0.05$ criterion is applied?

Line 310-314: As this is one of the main results, please provide more detail on the calculation of over-sampling (e.g. what does it mean with $> 500\%$ oversampling).

Line 356-357: I applaud the authors for being candid but this could be rephrased to sound a bit more positive. Something along the line of "results are associated with large uncertainties due to unconstrained model parameters". This will also set up nicely the next sentence about future work to quantify the errors associated with IFA-based reconstruction.

Section 4.0 (alternatively, add a new sub-section before section 4.0): See my general comments above. After going through the rather negative results based on the idealized simulations, one is left wondering what does this mean for real-world reconstruction. Can we indeed apply what we learn from this idealized simulation to actual records? Is it too early to tell, since the parameters used in the simulation are unconstrained? I think the reader, especially users of IFA, would like to have more details in this regard, as well as some concrete suggestions on what to do/ not to do when interpreting IFA-based reconstruction beyond the rather general suggestions already offered by the authors.