Response to Referee #1:

Ref.: Ms. No. bg-2017-429

Title: Modeling seasonal and vertical habitats of planktonic foraminifera on a global scale

We would like to thank reviewer Inge van Dijk for her constructive comments and suggestions, which will help us to greatly improve our manuscript. Based on the comments of all four reviewers we will prepare a new version of our manuscript as outlined below.

However, during the review process, we discovered an error in the underlying ocean model. Unfortunately, the ocean circulation is not correctly represented in the used coarse resolution (i.e., ~3º) model configuration. For a correct representation of the ocean and to yield scientifically consistent results, we had to perform a new model run with a higher horizontal resolution (i.e., ~1º) on a supercomputing system. This model run takes ca. 5 weeks and is currently in the final production phase. At a first glance, the new results will not differ that much from our previous results as the representation of the upper ocean, where the analyzed foraminiferal species live, was actually reasonably well simulated in the coarse resolution model configuration compared to, e.g., the World Ocean Atlas 2013. We expect that the distribution of only a few species might be affected, when using the higher resolution model configuration with a more realistic representation of the ocean physics. Since we have not yet obtained the final results, we were not always able to provide detailed answers to your comments and had to keep our responses rather general.

Please find, in the following, the original comments in black and our responses in light blue; the indicated page and line numbers refer to the previously submitted manuscript.

Referee #1 comments:

I have carefully read the manuscript 'Modeling seasonal and vertical habitats of planktonic foraminifera on a global scale' by Kretschmer and coauthors, which presents a model to predict global concentrations of five species of planktonic foraminifera and their depth habitat. This model could aid paleoclimatologists to correct for habitat depth when using shells of planktonic foraminifera to reconstruct ocean conditions. I need to remark that I have no experience using PLAFOM, or any practical experience with either the BEC model or CESM1.2(BGC) configuration. Therefore, my comments are rather general and an experienced user should review e.g. the use of model parameters and choice of configuration. I only have a couple of remarks that mainly focus on the usability and applicability of the model to reconstruct past depth habitats.

General comments

In general the authors should avoid certain 'model jargon', if they want to convince the broad

foraminiferal society to use and apply this model. It is sometimes difficult to follow which steps are taken and assumptions were made to test or simulate certain scenarios (e.g. page 6, lines 23-25).

Thank you for pointing this out. We will change parts of the method section, also according to the higher resolution model configuration, and will include more or delete redundant information, when appropriate, for a better understanding. However, to ensure reproducibility of our study, we cannot avoid using a certain 'model jargon' to explain the applied modeling approach and the used model setup. We already tried to use as little model jargon as possible and provided in all conscience a comprehensible model description.

Even though habitat tracking is very important when using shells of planktonic foraminifera to reconstruct ocean conditions, it is still (more?) crucial to pinpoint the actual calcification depth within the depth habitat, since this is where the calcite is formed. Even though the model can reasonably well predict (globally) the vertical distribution, this does not mean that at this specific depth the environmental signal was 'logged' into the shell. Please include somewhere a couple of sentences on the reconstructed depth habitat compared to the actual calcification depth. Could this be the next step for PLAFOM3.0?

This is a valid point and in a next step, we would like to combine PLAFOM2.0 with a module, which specifically takes this into account and calculates species-specific isotope compositions of the modeled foraminiferal species, such that we could directly infer information about the calcification depth of each species. However, without any information about the species-specific habitats, it is difficult to provide a statement regarding the calcification depths of the individual foraminiferal species. Therefore, we at first intended to simulate realistic species-specific habitat depths and next we plan on obtaining realistic calcification depths. We will include a paragraph in the discussion (section 4.1.3) regarding a comparison of the reconstructed depth habitat with the actual calcification depth of the individual species after evaluating the new results.

Section 2.3.1. What about other ocean parameters that vary over geological timescales which might influence growth rates? Like $[PO_4^{3-}]$ (Aldridge et al., 2012, BG) on SNW or the effect of carbonate chemistry on calcification rates? For instance Lombard et al., 2010 found lower growth rates of several species with lowered $[CO_3^{2-}]$ conditions and Davis et al., 2017 (Sci. Rep.) observed lower calcification rates with decreasing pH. Why are these parameters not taken into account in the model? Are these effect minor compared to temperature and food availability?

This is a valid point again, but we are not attempting to model species-specific growth rates (as opposed to Lombard et al., 2011). Rather we aim to more directly estimate foraminifera abundance, which can be compared to the sediment record more directly. The relationship

between growth rate and abundance is far from straightforward (cf. Lombard et al., 2011) and we are not aware of studies that have investigated the effect of those parameters on the abundance of planktonic foraminifera. We are aware that other ocean parameters might influence species-specific growth rates. The aim of this study, however, was to test if the existing planktonic foraminifera model is able to reproduce species-specific habitats when combined with a model configuration that resolves the vertical. One has to bear in mind that a model is only a simplification of reality and including more parameters would likely introduce more degrees of freedom and could lead to more model uncertainty and could additionally increase the computational costs. However, for a future model development it is worth considering those parameters. Here it is beyond the scope of this study to include more parameters to determine growth rates.

Section 2.5.2. and 2.5.3. The authors use the sediment trap/plankton tow samples to test the accuracy of the model in predicting seasonality & depth habitats. However, the amount of data used for this comparison is not covering the total range of oceanic settings, since big parts of the ocean are underrepresented. Is it possible to extend this database by adding other published sediment trap data? This way you can show your model can predict depth habitat in a wider range of ocean conditions, which will make it more robust for application in deep time. Just some quick suggestions: Mediterranean Sea: Mallo et al., 2017 BG; SW Atlantic: Venancio et al., 2016 Marine Micropaleontology; Mozambique channel: Steinhardt et al., 2014 Marine Micropaleontology; Panama basin: Thunell et al., 1983 EPSL; Indian Ocean: Guptha et al., 1997 JFR.

The reviewer rightly points out that our data compilation is not comprehensive. However, we pursued the strategy to acquire sediment trap and plankton tow data at more or less the same region to guarantee a consistent model-data-comparison throughout the manuscript when analyzing species-specific seasonal and vertical habitat patterns (see Figure 1b). We agree that this prerequisite limits the number of studies that can be used to evaluate the model, but the underlying data base covers all provinces and provides good estimates of the different species-specific habitats and their variability on a global scale that is sufficient to show the strength and weaknesses of our model.

Figure 2. Is it possible to add an 'offset map', in which you correlate e.g. the coretop data with the model data, to see where the model exactly over-/underestimates the data? This way you would be able to perform some (correlation) statistics, and this would clearly show the areas where the model did not predict the correct distribution. I understand you are trying to capture the global signal (as stated several times in the manuscript), but paleooceanographers are more interested in

specific areas when correction for e.g. depth habitat, and these are often also in more complicated oceanic settings (for example coastal/upwelling/river run off areas).

We will include an additional map in Figure 2 that provides a more thorough comparison between modeled and observed assemblages. Therefore, we will calculate the Bray-Curtis index of similarity between the model data and the core-top data, such that we can provide a measure of confidence. Note for the calculation, we will account for the different sizes of each species by using a mean size for each species based on the results of Schmidt et al. (2004) and recalculate the modeled relative abundances accordingly. We will add this analysis to the manuscript (i.e., to sections 3.1 and 4.1.1) to provide a thorough model-data-comparison. Nevertheless, the used model configuration consisting of three different models (i.e., POP2, BEC, PLAFOM2.0) could hamper a thorough statistical analysis as it is not unequivocally possible to differentiate which component might actually lead to a possible over-/underestimation of the data. Even the now used higher model resolution could likely lead to misrepresentations of small-scale processes, oceanic fronts, river runoff areas, and coastal upwelling regions, and could, thus, account for the model-data-mismatch. In addition, it is not possible to correlate the core-top data with the model data directly, because PLAFOM2.0 calculates foraminiferal concentrations via carbon biomass (i.e., in mmol C/m^3) and the core-top samples provide foraminiferal concentrations via number of specimens.

Page 11, line 27-31 and page 12, line 20-21. The authors state that part of the mismatch between the model and coretop data might stem from different genotypes having varying ecological preferences, and therefore their own unique model parameters. If so, does did not create a major bias for the whole model, especially when reconstructing depth habitats in deep time? For geological samples it is not possible to distinguish between genotypes, and therefore certain species might respond different in terms of depth habitat than the model will predict? Also, could it be that certain ecological preferences have changed over time? Can the authors predict how far in geological time you could still use this model to obtain reliable data on global distribution and depth habitat?

The reviewer points out two important considerations: i) cryptic species with different ecological preferences and ii) the question of stationarity. We would argue that both hold for all attempts to use planktonic foraminifera to reconstruct the past ocean. The assumption of stationarity of any proxy is fundamental to all paleoclimate reconstructions. The model can of course only be used for the time that the species have been present and for as long as we have indications that their ecology remained constant (cf. Huber et al., 2000 for *N. pachyderma*). The primary intended use of the model is to apply it to climate conditions covering the Last Glacial Maximum and/or the last couple of glacial-interglacial cycles, but not to deep time, when different species existed or extant

species may have had different ecological preferences.

With respect to cryptic species the reviewer is right to point out that this forms an important caveat. However, as the reviewer also mentions, it is often impossible to distinguish between cryptic species in the fossil record, so this caveat applies to any reconstruction using planktonic foraminifera. This is exactly the reason why ecological preferences of cryptic species need to be resolved, so that reconstructions and modeling efforts can be improved. To clarify this point, we will add this issue to the end of section 4.1.1:

"[...] In addition, the discrepancies between the model and core-top data might also partly stem from the underlying model parameterizations applied on a global scale, which do not distinguish between distinct genotypes of the different species with potentially varying ecological preferences. However, the recognition of cryptic species remains challenging, if not impossible, and these species are therefore rarely separated in sediment samples. This complicates a direct comparison between geological samples and model data, even if the ecological preferences were perfectly constrained. However, the model skill/performance suggests that ecological differences between cryptic species are limited and that the model provides a useful first-order approximation of global species distribution."

Minor comments

Page 2, line 18, 32; Page 6, line 16; page 11, line 23: Some problem with bracketing, e.g. double bracketing etc.

We checked for the double bracketing and, where possible, we will delete the unnecessary brackets. However, for some cases (i.e., Page 6, line 16; page 11, line 23) we will not change the bracketing as this would potentially cause a misunderstanding with the referencing.

Page 6, line 24: quasi-steady

Done.

Page 7, line 15: space missing between '(Figure 1a).' and 'We'

Done.

Page 8, line 5 and page 11, line 17: Arctic Circle

Done.

Page 12, line 10-14. Can you explain the underestimation of the model in scenarios were assemblages are dominated by two species?

Here, we actually meant that the model is not able to capture the full extent of the observed relative abundances in certain areas where a dominance of some species is actually expected. We will change this sentence accordingly after evaluating the new results.

Page 12, line 21: change or remove 'see'
Done.

References:

- Huber, R., H. Meggers, K.-H. Baumann, M. E. Raymo, and R. Henrich (2000), Shell size variation of the planktonic foraminifer *Neogloboquadrina pachyderma* sin. in the Norwegian-Greenland Sea during the last 1.3 Myrs: implications for paleoceanographic reconstructions, *Palaeogeography, Palaeoclimatology, Palaeoecology*, *160*, 193-212.
- Lombard, F., L. Labeyrie, E. Michel, L. Bopp, E. Cortijo, S. Retailleau, H. Howa, and F. Jorissen (2011), Modeling planktic foraminifer growth and distribution using an ecophysiological multi-species approach, *Biogeosciences*, *8*, 853-873.
- Schmidt, D. N., S. Renaud, J. Bollmann, R. Schiebel, and H. R. Thierstein (2004), Size distribution of Holocene planktic foraminifer assemblages: biogeography, ecology and adaptation, *Marine Micropaleontology*, *50*, 319-338.
- Thunell, R. C., W. B. Curry, and S. Honjo (1983), Seasonal variation in the flux of planktonic foraminifera: time series sediment trap results from the Panama Basin, *EPSL*, *64*, 44-55.