<u>Response to comments from reviewers</u>

Key
 – Reviewers' comments
 – Authors' response

Referee #1

Review of "Global and regional Pleistocene benthic δ18O stacks with a comparison of different age modeling strategies" by Zhou et al.

General comments

The authors propose an update to the LR04 benthic d18O stack (Lisiecki & Raymo 2005) that is widely used to construct age models of marine sediments in paleoceanographic studies. This update, which relies on an extended number of sites (224 cores), includes one global stack presented on 3 different chronologies and two regional stacks (Atlantic, Pacific). The manuscript describes the construction of these stacks and their chronologies. It also discusses the age uncertainties associated with them and the applications for which each stack is best suited.

There is absolutely no doubt that the new stacks proposed by the authors are of high scientific importance and will be widely adopted by the paleoceanographic community. Overall, the manuscript is well written and illustrated. It would have been helpful to include more information on the sites selected for each chronological iteration, as well as a more detailed discussion on specific aspects, which are detailed below.

Response #1 – We thank the reviewer for recognizing the scientific importance of our project. Below, we address the comments as fully as we can in order to take advantage of the input and improve our manuscript.

Specific comments

1. The Methods section starts directly with the description of stacking steps. I think a new subsection should be added first to better describe the marine cores selected for the construction of each stack and the related criteria: number of cores, proportions of new sites vs. cores already included in previous stacks, name of cores, their location (including water-depth), their references, mean temporal resolution of records, which proxies are available for each core apart from benthic d18O (magnetic data, SST, others?). Right now, this information is partly disseminated throughout the text, partly present in the supplementary data or missing. I think presenting an overview of selected cores and criteria in the main text and explicitly referring to the supplementary data for more information would be helpful.

Response #2 – We agree with the reviewer that the Methods section should begin with a new subsection describing the marine cores used for the stacks. This subsection will include all the information the reviewer mentioned: the number of cores, proportions of new sites vs. cores

already included in previous stacks, names of cores, their locations (including water depth), their references, and the mean temporal resolution of the records. For information on cores that are too long to include in the main manuscript, such as their locations, we will provide an aggregated statistical overview. We will also discuss the geomagnetic reversal and biostratigraphic age information when available. However, we note that our stack construction does not use SST data. While SST data are frequently measured in marine sediments and can similarly inform glacial/interglacial trends as benthic $\delta^{18}O$, compiling all the SST records associated with over 200 cores is outside the scope of this study and best suited to a follow-up investigation. See Response #8 for further clarification of this point.

2. I am surprised that the list of geomagnetic reversal and excursions selected for BIGSTACK$_{magrev}$ (Table 1) does not include recent well-dated magnetic events such as the Mono-Lake, Laschamps and Blake events. Reading the discussion in lines 411-416, I guess it is because only geomagnetic reversal/excursions dates derived from astronomical tuning are included. First, I would place this statement earlier in the text, in the Methods section. Second, I think one advantage of using magnetic reversal/excursions is also to benefit from absolute and relatively accurate dating from Ar/Ar. I find it really unfortunate that the attempt to use Ar/Ar dating did not succeed (lines 414-417). I would encourage the authors to further detail and illustrate this attempt. And if difficulties arise in the interval around 1.9-2.1 Ma only, why not reduce the coverage of the stack to the most recent 1.9 Ma or simply omit the period 1.9-2.1 Ma in the produced stack? More discussion on this issue would be helpful.

Response #3 – We agree with the reviewer that using Ar/Ar dated geomagnetic reversal/excursion ages is highly preferable to using astronomically tuned ages. We revisited the BIGSTACK$_{magrev}$ construction process and found a way to use Ar/Ar dated magnetic reversal/excursion ages that still allows the eTimeOpt procedure to work. In brief, we found two cores (ODP 1082 and 1083) that only have data in the oldest segment of the stack (~1.9 Ma and older), which may have biased the magrev age model towards younger ages. After removing these two cores, the auto-tuning improved significantly. See below the comparison of the old and new magrev and auto stacks (Figs. 1 and 2), as well as the time-frequency analysis results from the multi-taper method evolutive harmonic analysis before and after tuning for the new auto stack (Fig. 3). Note that even though ODP 1082 and 1083 only have data in the oldest segment of the stack, their removal impacts younger segments of the stack because of the interpolation between geomagnetic reversals/excursions.

In the revised manuscript, we plan to switch to using the new versions of BIGSTACK$_{magrev}$ and BIGSTACK$_{auto}$. We will include a comparison of the old and new magrev age models (and their tuning results) in the supplementary material to illustrate the age models' sensitivity to estimating magnetic reversal ages in these two different ways (tuned ages vs Ar/Ar ages).

We have found identifications of the Laschamp and Blake events in ODP 983, 984, and 1089 in the literature (Channell et al., 1997; Channell, 1999; Stoner et al., 2003) and incorporated the Ar/Ar ages of these events in the construction of BIGSTACK$_{magrev}$ (Singer, 2014). Mono-Lake is recorded in only one marine sedimentary site (ODP 919) off east Greenland (Channell et al.,
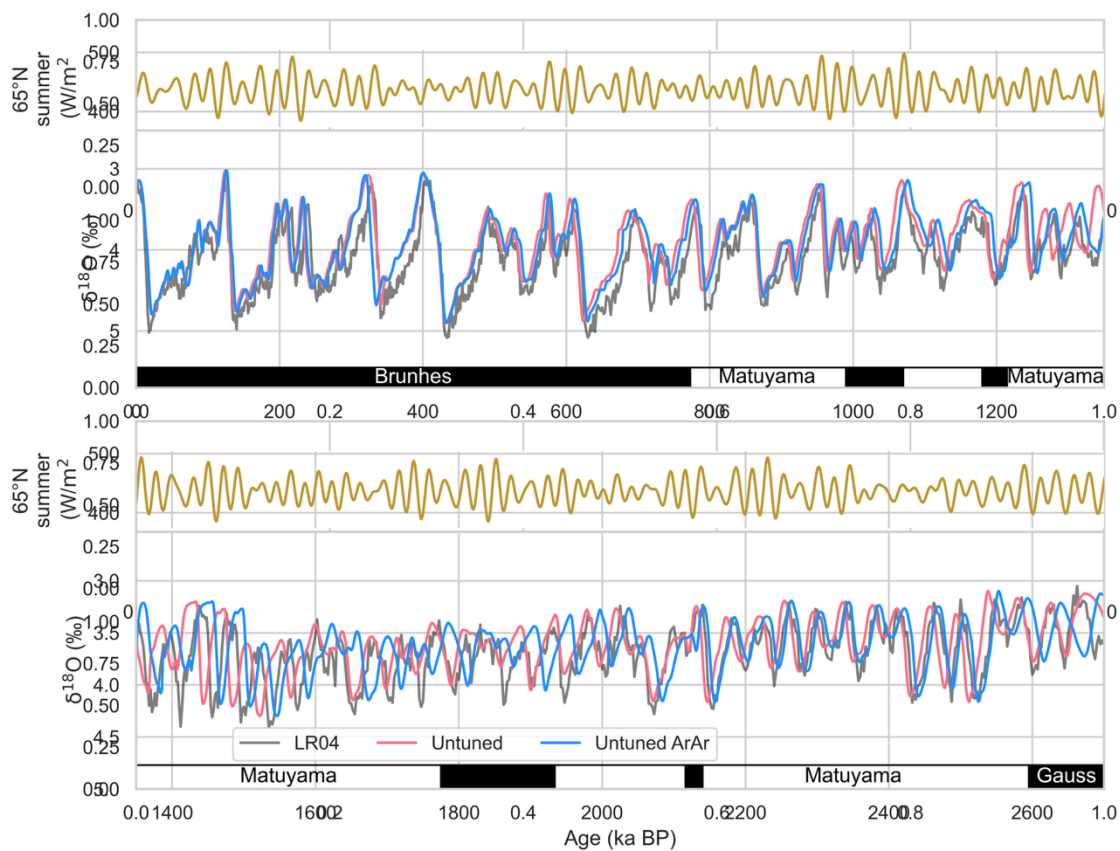
Figure 1. A comparison of the current BIGSTACK$_{magrev}$ (red) and the new version of BIGSTACK$_{magrev}$ (blue) that uses Ar/Ar-dated reversal/excursion ages.

2020), and the core from this site does not have an abundance of benthic foraminifera (Channell, 2006).

Given the high scientific importance of the produced stacks, I think that their relative comparisons are too brief and insufficiently discussed. Because stacks are plotted on different Y-axes in available figures (Figures 6, 7, etc), I find it very difficult to compare the amplitude and the timing of the stacks. I think it is particularly important for the comparison of the global BIGSTACKmixed stack to the regional Atlantic and Pacific stacks. I would invite the authors to plot these stacks on the same Y-axes and develop the discussion on their similarities and differences. Also, I would find it interesting to reinforce the discussion by adding a comparison, in addition to LR04, of the new stacks to ProbStack and HN23 stack, which is currently absent in the manuscript. Finally, more discussion on the reasons and the impacts of a higher smoothing effect in the BIGSTACKs compared to LR04 (lines 402-404) would be helpful. Indeed, I have the impression that the five BIGSTACKs are highly smoothed compared to LR04 (see https://yz3062.github.io/htmls/BIGSTACK_bokeh.html). I already found LR04 quite smoothed when using it to derive age models of sediment cores for the last glacial cycles. I anticipate the
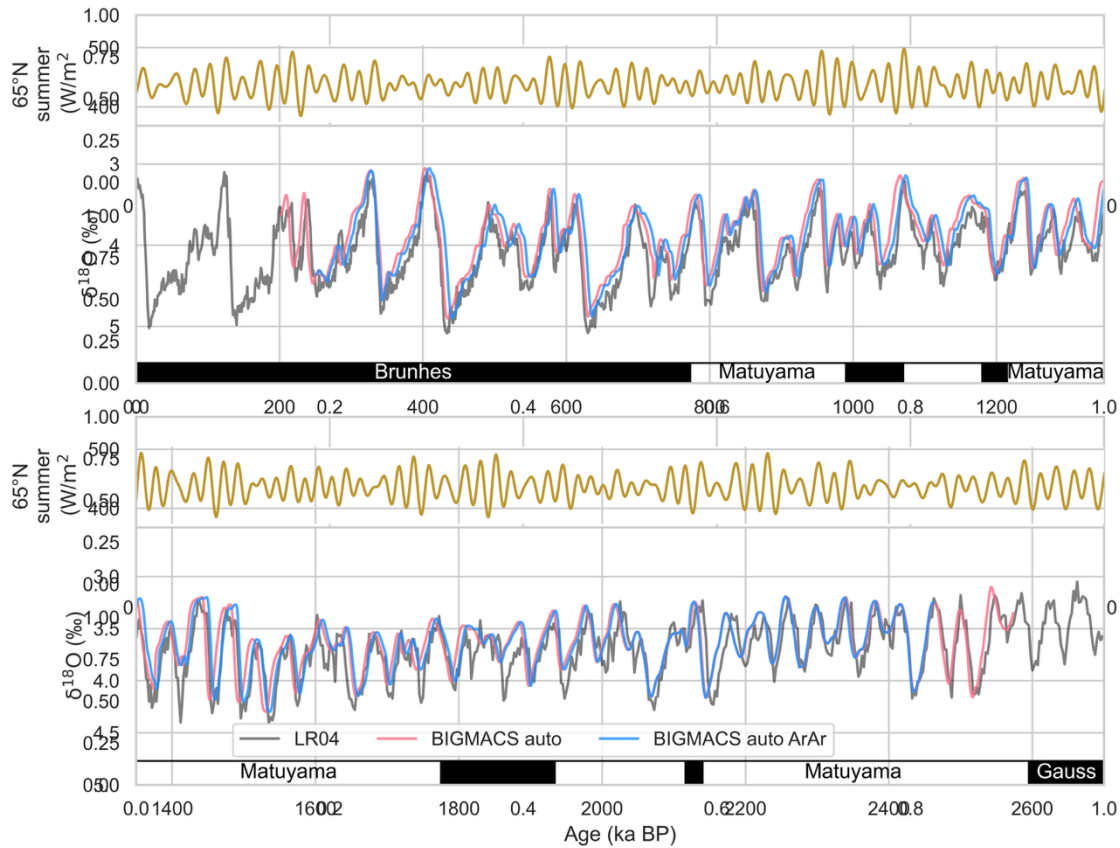
Figure 2. A comparison of the current BIGSTACK$_{auto}$ (red) and the new version of BIGSTACK$_{auto}$ (blue) that derives from a BIGSTACK$_{magrev}$ that uses Ar/Ar-dated reversal/excursion ages (blue curve in Fig. 1). Note that the new version of BIGSTACK$_{auto}$ (blue) uses a longer window size of 500 kyrs in eTimeOpt, which shortens the resulting stack slightly more.

high smoothing of BIGSTACKs may be a major drawback when using them as references to produce marine chronologies. I am wondering to which extent this smoothing could be minimized during the stacking approach and would expect this smoothing to be better discussed in the text, in particular highlighting this potential limitation in scientific applications.

Response #4 – We wholly agree with the reviewer that the comparison between BIGMACS$_{mixed}$, BIGMACS$_{mixedA}$, and BIGMACS$_{mixedP}$ can be improved, both visually and in the text. We will also add a discussion of our stacks with ProbStack and H23NA. Lastly, we agree with the reviewer that the smoothness of the new stacks should be examined more carefully. This point was also brought up by Reviewer 2. We will highlight the degree of smoothing of the new
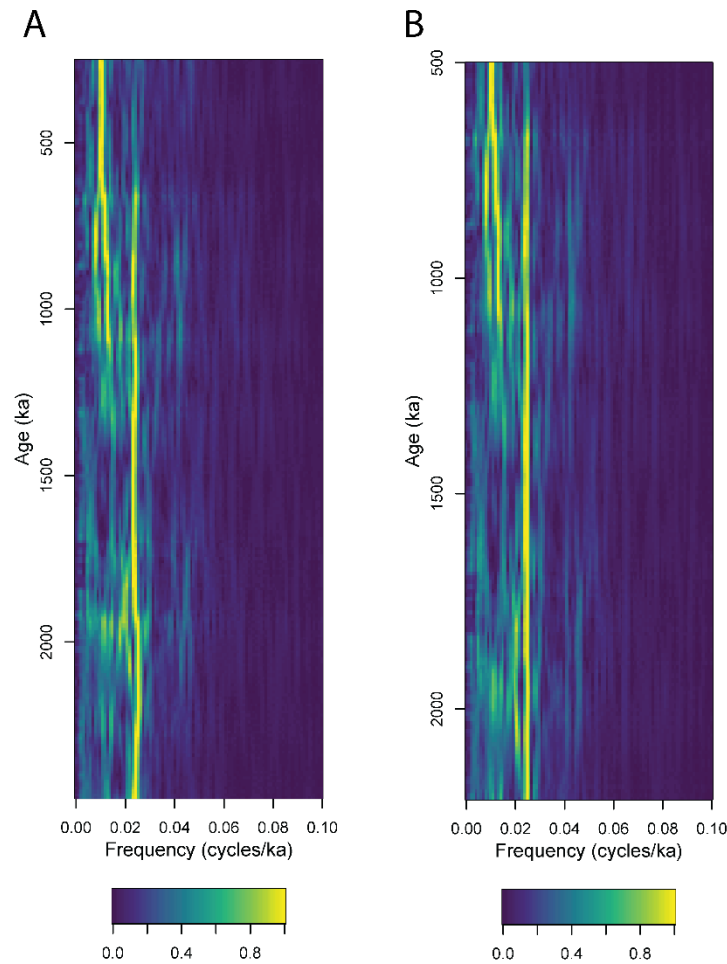
Figure 3. The time-frequency analysis results from the multi-taper method evolutive harmonic analysis, before and after tuning for a new version of BIGMACS$_{auto}$ that uses an input stack that uses Ar/Ar-dated reversal/excursion ages.

stacks, discuss why the stacks are so smooth (due to BIGMACS's use of Gaussian Process Regression), and assess the implications of this smoothing for their scientific applications. We plan to add a statement to the paper explaining that this stack is specifically geared towards orbital-scale age models, and that we advise against using it for millennial-scale stratigraphy of the most recent glacial cycles. For high-resolution regional and global stratigraphy of the last glacial cycle, we will point readers to Lisiecki and Stern (2016). We will also add supplemental figures comparing the Lisiecki and Stern (2016) Deep North Atlantic stack with our BIGSTACK$_{mixedA}$ and comparing the Lisiecki and Stern (2016) volume-weighted global stack with our BIGSTACK$_{mixed}$ to illustrate the difference.

3. I am not convinced by the temporal lag of 1 ka that is applied to derive the Pacific stack compared to the Atlantic one. Even though this issue is already discussed in lines 466-477, I am wondering whether this approach is the most robust. As already mentioned in

the text (line 468), I think the Pacific-Atlantic lag is larger than 1 ka during deglaciations. Have the authors considered using a larger Pacific-Atlantic lag during deglaciations, instead of or in addition to a constant throughout the investigated time period?

Response #5 – We agree with the reviewer's assessment that applying a constant temporal lag of 1 kyr to the Pacific stack compared to the Atlantic is overly simplistic. We plan to increase the Pacific lag to 2 kyr during the middle of each termination, then ramp it down to 1 kyr at the interglacial peak and the glacial maximum. The 2-kyr lag during the middle of each termination is supported by a previous study that estimated the average Pacific lag during Late Pleistocene terminations to be 1600 yr (Lisiecki and Raymo, 2009).

Technical comments

Line 17: The reference Ahn et al. 2017 is missing in the reference list.

Response #6 – We thank the reviewer for the careful reading of our manuscript. The reference is there, but it is hard to see because it seems to be part of the previous item in the bibliography due to the lack of indentation.

Lines 68-70: This section describes the "strengths and weaknesses" of chronological approaches. For radiocarbon dating, I would mention in these lines the difficulty to correctly estimate past marine age reservoir values.

Response #7 – Agreed.

Line 153: please clarify what is meant with "multiproxy age models". Do the authors mean age models that use several proxy-records from the same core?

Response #8 – BIGMACS allows the simultaneous use of radiocarbon and benthic $\delta^{18}O$ to create an age model. It also allows the use of additional age control points, such as geomagnetic reversals/excursions and biostratigraphy. However, it cannot currently stratigraphically align multiple proxy time series such as benthic $\delta^{18}O$ and SST. We agree that this term can be ambiguous and will clarify that, by 'multiproxy,' we mainly mean benthic $\delta^{18}O$ and age control points (e.g., radiocarbon ages, magnetic reversals/excursions, and/or biostratigraphic events).

Lines 227-230: I think this information should arrive earlier in the text.

Response #9 – Agreed.

Line 305: I think the tie-points that have been defined to manually align BIGMACSmagrev to the different targets to produce the mixed BIGSTACK should be provided in the supplementary data files. It may already be the case, but I did not find them to check.

Response #10 – Agreed. We will provide that information as a supplementary data file.

Figure 5: please add the colour code in the figure's caption so that the reader knows which colored curve corresponds to which stacks.

Response #11 – Will do.

Figures 6, 7, 9, 10: please add the name of the stacks directly onto the figure (e.g. along the Y-axes or next the record).

Response #12 – Will do.

Figure 8: similarly please indicate directly onto the figure which age difference is plotted in panels a, b, and c.

Response #13 – Will do.

Supplementary Data file 10: it would be helpful to specify the content of this supplementary data. To which stacking step or chronology do ages given in this excel file correspond for all cores?

Response #14 – This file contains details of input records for BIGSTACK$_{mixedP}$ and BIGSTACK$_{mixedA}$. We will add a new "README" tab at the beginning that explains the file's contents.