Response to Reviewers for Manuscript:

"Evaluating manual versus automated benthic foraminiferal d18O alignment techniques for developing chronostratigraphies in marine sediment records"

By Jennifer L. Middleton, Julia Gottschalk, Gisela Winckler, Jean Hanley, Carol Knudson, Jesse R. Farmer, Frank Lamy, Lorraine E. Lisiecki, and the Expedition 383 Scientists

https://egusphere.copernicus.org/preprints/2023/egusphere-2023-2906/#discussion

Response to Reviewer 1:

(reviewer comments copied for clarity, author responses in red, *italicization* indicates new text added to the manuscript)

The generation of chronostratigraphies in paleoceanography by tuning benthic δ18O records to a target curve is a widespread practice. However, a lack of standardization in the tuning procedure and a limited assessment of resultant age model errors persist. This paper addresses these issues by offering a comprehensive overview of commonly used tuning targets (LR04, Prob-stack, CENOGRID, LR09 Atlantic and Pacific), comparing manual and automatic tuning (utilizing the HMM Match algorithm), and providing a best practice guide for developing age models based on benthic foraminiferal δ18O records.

In summary, the paper is well-written and contributes timely insights toward standardizing benthic foraminiferal δ18O stratigraphy. It serves the paleoceanography community by presenting a balanced overview of the strengths and weaknesses of each tuning target. The authors adeptly discuss errors associated with both tuning targets and the target curve itself. In this regard it is apparent that errors derived from different tuning targets are hard to compare due to the different ways in which the target curves have been constructed (e.g., CENOGRID not being a stack like LR04).

We thank Reviewer 1 for their supportive summary of our work.

While the proposed best practice guidelines seem reasonable, I think the authors underappreciate the human brain's proficiency in pattern recognition. Experienced paleoceanographers may discern irregularities indicating stratigraphic issues or site-specific influences on the δ18O signal, a capability not easily replicated by automated algorithms. Hence, although it might somewhat self-explanatory, I would suggest to include into the best practice guide that the results generated by algorithms should be critically evaluated by visual inspection. The error provided e.g. by HMM Match is certainly a helpful means for such post-tuning evaluation.

We appreciate this point and will revise the manuscript accordingly. Indeed, there are some sedimentary environments (especially where sedimentation rates are highly variable) and some benthic foraminiferal oxygen isotope records (especially where temporal resolution varies and/or large data gaps occur) for which the guiding assumptions of the HMM-Match algorithm are not

well suited. As such, we will add the following text to the manuscript in the discussion of manual vs. automated alignments (Section 5.1):

> *"While automated alignment algorithms like HMM-Match and BIGMACS provide many advantages in the generation of benthic foraminiferal $\delta^{18}O$ chronostratigraphies, we note that depositional environments with highly variable sedimentation rate changes and benthic foraminiferal $\delta^{18}O$ records with long data gaps or temporal variations in sampling resolution may not abide by the assumptions included in these automated algorithms. Specifically, the HMM-Match algorithm is designed to minimize large sedimentation rate changes between data points based on the probability distribution calculated from a compilation of 37 radiocarbon-dated sediment cores (Lin et al., 2014). This guiding principle may hinder the success of HMM-Match-generated alignments for benthic foraminiferal $\delta^{18}O$ records with irregular sampling frequencies or from regions like the Antarctic Southern Ocean where very large sedimentation rate changes are expected across a deglaciation (e.g., Hasenfratz et al., 2019). Consequently, a close visual evaluation of the automated alignment outputs against their designated targets should be completed manually (i.e., by the user) to quality check the resulting benthic foraminiferal $\delta^{18}O$ chronostratigraphies."*

We will also include the following text in the bullet point summary of best practices in Section 5.4:

> *"Visually assess the quality of alignment outputs generated by automated algorithms for alignment mismatches or other irregularities, especially across glacial terminations."*

The authors highlight the benefits of HMM Match estimates in aligning every data point of a high-resolution time series, asserting its superiority over manual wiggle matching, especially evident in the strong mismatch during MIS 6 in Site 980/981. However, several concerns arise:

1. Aligning every data point may lead to overconfidence, neglecting the natural noise in the data, including e.g. bioturbation.

We thank the reviewer for providing us the opportunity to clarify that the HMM-Match algorithm accounts for the possibility of natural noise within a benthic foraminiferal oxygen isotope data set by assuming that the difference between each oxygen isotope data point of an input core and the target will fall along a normal probability distribution. We will include the following text in Section 3.2 to clarify this point to the general reader:

> *"The HMM-Match algorithm accounts for natural variance in a benthic foraminiferal $\delta^{18}O$ data set (e.g., due to bioturbation, spatial variability, and measurement uncertainty) by assuming that the residual $\delta^{18}O$ value between each input record and the target will fall along a normal probability distribution (Lin et al., 2014)."*

2. Considering line 555-557: if visual matching is hindered by high resolution, a smoothed version of the record could be considered.

We will update the text in lines 555-557 to state:

> "This observation *demonstrates* that age uncertainties associated with manual tuning cannot *always* be reduced by increasing the temporal resolution of the input data *and suggests that a smoothing of a high-resolution input record prior to alignment may be beneficial for tuning*."

3. It should be acknowledged that the manual tuning presented in the paper is just one (!) possible solution. Comparing manual tunings from different (experienced paleoceanographers), while beyond the paper's scope, would offer valuable insights into the variability of visual wiggle matching.

We agree that evaluation of a suite of manual alignment outcomes generated by many distinct and experienced users would be very interesting and is beyond the scope of this work. To address this comment, will update the text in Section 3.1 to include the following lines:

> *"Due to the subjectivity inherent in user-defined alignments, the manual alignments presented in Supplemental Tables S2-S4 represent just one of many possible manual alignment outcomes for each record."*

A further side note: While automated algorithms like HMM Match have advantages, it's crucial to consider accessibility, especially for institutions with limited funding for MATLAB licenses. Providing standalone versions or developing tools in license-free environments like R would benefit the broader scientific community.

We thank the reviewer for bringing up this important point. We will include the following text at the end of our suggested best practices in Section 5.4:

> *"On a broader note, given the financial burden of programming platforms like MATLAB, we encourage the development or translation of automated alignment algorithms into license-free coding languages such as Python or R so that they may become more accessible to the research community."*

Additional comments: Regarding Fig. 2, consider using colors that offer better discernibility than purple and dark blue.

We will change the dark blue record in Fig. 2 to black to improve the distinction between each record. We will check the resulting color palette and the other figures for sensitivity to color-blindness and color-deficiency using the Coblis color blindness simulator tool (https://www.color-blindness.com/coblis-color-blindness-simulator/) and adjust their palettes accordingly.

Apart from the issue above, the paper is a very valid contribution and should be published after minor corrections.

We thank Reviewer 1 again for their helpful comments and suggestions that helped to improve our manuscript.

Response to Reviewer 2:

This paper presents an evaluation of the automated HMM-Match alignment method versus manual alignment methods for establishing chronostratigraphies in marine sediment cores. Three benthic foraminifer $\delta^{18}O$ records covering the past 3.5 My are dated by alignment to various targets, using a manual approach and the automated HMM-Match algorithm. These three benthic foraminifer records are from distinct oceanic regions (i.e. ODP site 980/981 in the North Atlantic at 55°N and ~2200 m depth, ODP site 1090 in the South Atlantic at ~43°S and 3700 m depth, and IODP site U1541 in the South Pacific at ~54°S and 3600 m) thus allowing the authors to test possible biases resulting from temporal offsets between regional target curves. The authors conclude that automated alignment methods like the HMM-Match algorithm yield lower age uncertainties than the manual alignment approach and recommend their use to perform stratigraphic alignments.

However, as explained in detail below, the comparisons presented in this paper seem to indicate systematic biases in the age models produced by the automated HMM-Match alignment algorithm.

I thus recommend that the article be rejected unless the authors can provide satisfying answers to the main concerns 1 and 2 listed below.

We thank Reviewer 2 for their review and for giving us the opportunity to clarify some aspects of our study that help address their primary concerns listed below. We also wish to reiterate that our study investigates and discusses the minimum level of uncertainty estimates on age models derived from benthic foraminiferal oxygen isotope alignment (as stated in the manuscript), while most authors do not consider or include any such uncertainty estimates when presenting such age models. We therefore consider our work to be an important contribution towards overcoming this common discrepancy.

Main concerns

- The HMM-Match algorithm is used throughout this study to perform automated probabilistic alignment of $\delta^{18}O$ records to chosen dated $\delta^{18}O$ This algorithm (published in Lin et al., 2014) is presented in section 3.2. The author explain that the algorithm "checks the implied sedimentation rate changes against their natural likelihood based on an independent compilation of radiocarbon-derived sediment core chronologies over the last deglaciation" (l. 300-302). However, due to compression of the sedimentary column, "natural sedimentation rates" cannot be assessed based on the last deglaciation for an entire record: the apparent sedimentation rates that can be derived from radiometric dating of the last deglaciation in mid and low latitude regions are necessarily higher than further deep in the sedimentary column. Moreover, in high latitude regions (north of ~ 40°N, i.e. the case of ODP980/981, and south of ~40°S, i.e. the case of ODP1090),

radiocarbon dating cannot be used to derive realistic sedimentation rates across the last deglaciation, due to highly variable surface reservoir ages. Even though there may be additional reasons, these two reasons could explain why the automated HMM-Match algorithm generates systematic age biases: the average age offset between the manual and HMM-Match based alignment to the LR04 target curve is positive for all three cores. For IODP core U1541 this average age offset ranges from 3.2 to 4.6 ky (l. 435-455), while it is 3.7 ky (l. 506-508) and 2.8 ky for ODP1090 and ODP980/981 (l. 551-553), respectively. Without a thorough discussion of why this is the case, such a systematic positive age offset casts a serious doubt on the reliability of the automated HMM-Match alignment algorithm.

Reviewer 2 raises two concerns within this comment that we address separately.

Apparent systemic bias:

As indicated by Figures 4c, 5c, and 7c, this is a misconception by Reviewer 2 because there is no systematic positive age bias between the HMM-Match-derived and the manually-derived alignments of each core to LR04. Rather, the HMM-Match and manually-derived alignments lead to age models for which some of the HMM-Match-derived ages are younger than the manually-derived ages (yielding negative offset values) and some are older (yielding positive offset values) and the magnitude of these offsets varies throughout each record.

In order to describe the average magnitude of age offset between HMM-Match-derived and manually-derived ages, we reported the average of the absolute values of the offset for each core in lines 435-455, 506-508, and 551-553 (which is positive by definition). We use this approach to avoid underreporting the magnitude of the offsets between alignment methods, although the sign of this offset changes throughout the record. We believe that this is a transparent way to compare manual versus automated tuning techniques.

When the mean offset value for each record is computed at face value (e.g., the sign of the offset is considered), the resulting mean offset between the HMM-Match-derived and manually-derived alignments to LR04 for IODP Site U1541, ODP Site 1090 (over the last 1.4 Myr only), and ODP Site 980/981 are 2.8 kyr, 0.2 kyr, and -0.9 kyr, respectively. These values are significantly smaller than the absolute average values we reported because the positive and negative offsets partially cancel each other out.

We will clarify our approach of calculating age offsets between alignment techniques by updating the text within the results section as follows:

> *"The direction of age model offset (e.g., whether HMM-Match-based ages are younger or older than the manually-derived ages) varies throughout the record. There is no systematic lead or lag associated with the age model of the HMM-Match-derived alignment relative to the age model of the manually-derived alignment.* Including (excluding) the two intervals of complete cycle offsets, the *mean value of the absolute* difference between the manually aligned and HMM-Match aligned LR04 ages for IODP Site U1541 averages 4.6 kyr (3.2 kyr) across the entire record." (Lines 453-455)

"Over the last 1.4 Myr, the manual and automated HMM-Match-based alignments of the benthic foraminiferal $\delta^{18}O$ record from ODP Site 1090 to the LR04 stack yield similar chronostratigraphies with an average *of the absolute values of age offsets equal to* 3.7 kyr between them (Fig. 5)." (Line 506)

"The *absolute* difference in age models between these two alignment approaches reaches a maximum value of 24 kyr at 650 ka (leading into MIS 16) and has an average value of 2.8 kyr across the 1.8 Myr record at this site." (Line 551)

HMM-Match's assessment of natural sedimentation rates:

We thank Reviewer 2 for the opportunity to clarify this point as well. The radiocarbon records compiled by Lin et al. (2014) are used to assess the likelihood of relative (not absolute) changes in sedimentation rates within the HMM-Match algorithm, including those associated with sediment compression and expansion. By targeting the likelihood of sedimentation rate changes and not absolute sedimentation rates, the HMM-Match algorithm is not biased by sediment compaction in deeper records. This is demonstrated in our own results. While the Lin et al. (2014) radiocarbon compilation only includes records with mean sedimentation rates >8 cm/kyr, the HMM-Match algorithm generates alignments that include sedimentation rates as low as 0.5 cm/kyr (as indicated by the HMM-Match-derived sedimentation rate records of IODP Site U1541, ODP Site 1090, and ODP Site 980/981 in Figure 8 of the manuscript).

We will update the description of the HMM-Match algorithm in Section 3.2 in the following manner to address these points:

"The algorithm checks the implied *relative* sedimentation rate changes associated with each alignment fit against their natural likelihood based on *the distribution of relative sedimentation rate changes observed in* an independent compilation of 37 radiocarbon-derived sediment core chronologies over the last *40 kyr* (Lin et al., 2014)."

In addition, the radiocarbon compilation comprises 37 cores with closely spaced radiocarbon measurements (0.5 – 4 kyr intervals), of which 16 records extend to 40 ka. This compilation includes some Southern Ocean records (north of ~46°S), but intentionally excludes records from the high latitude North Atlantic because of radiocarbon reservoir age variations. While the radiocarbon compilation is not globally representative, we note that the HMM-Match algorithm does not attempt to spatially constrain sedimentation rate changes within an input core based on its location (the latitude and longitude of input records is not considered). Rather, the algorithm considers what sedimentation rate changes are most unlikely between two data points based on the distribution of sedimentation rate changes observed in the 37 cores included. Thus, although this radiocarbon compilation does not include cores from the same regions as at IODP Site U1541 and ODP Site 980/981, the distribution of sedimentation rate changes included within the radiocarbon compilation appears to provide sufficient flexibility for the HMM-Match algorithm to generate reasonable alignments for both sites (e.g., Figures 4 and 7). Nonetheless, we recognize that there may be some depositional environments that exhibit natural sedimentation rate changes that fall outside the likely range estimated by Lin et al. (2014) and used in HMM-Match, as the reviewer suggests.

We will therefore include the following caveat regarding the use of HMM-Match in specific depositional regions and environments in the discussion Section 5.1 (which is identical to the text inserted for our response to the first comment of Reviewer 1):

"While automated alignment algorithms like HMM-Match and BIGMACS provide many advantages in the generation of benthic foraminiferal $\delta^{18}O$ chronostratigraphies, we note that depositional environments with highly variable sedimentation rate changes and benthic foraminiferal $\delta^{18}O$ records with long data gaps or temporal variations in sampling resolution may not abide by the assumptions included in these automated algorithms. Specifically, the HMM-Match algorithm is designed to minimize large sedimentation rate changes between data points based on the probability distribution calculated from a compilation of 37 radiocarbon-dated sediment cores (Lin et al., 2014). This guiding principle may hinder the success of HMM-Match-generated alignments for benthic foraminiferal $\delta^{18}O$ records with irregular sampling frequencies or from regions like the Antarctic Southern Ocean where very large sedimentation rate changes are expected across a deglaciation (e.g., Hasenfratz et al., 2019). Consequently, a close visual evaluation of the automated alignment outputs against their designated targets should be completed manually (i.e., by the user) to quality check the resulting benthic foraminiferal $\delta^{18}O$ chronostratigraphies."

- Moreover, the age offsets computed between the HMM-Match alignment of each studied core to the Prob-stack and to the LR04 or Cenogrid target curves are systematically positive as well. This suggests that there are systematic offsets between the different targets curves themselves. This seems extremely strange in the case the Prob-stack and LR04 target curves because these two curves are based on the same absolute age constraints. It is thus absolutely necessary to discuss why this is the case in the article.

As above, there is a misunderstanding of our reported values. The persistently positive values we report for the age offsets between the HMM-Match alignments to Probstack and to LR04, LR09, and CENOGRID for each record under investigation are also the result of taking the average of the absolute value of the observed offsets between each alignment (which vary between positive and negative values throughout the record, as indicated in Figures 4f, 5f, and 7f in the manuscript). We will clarify this point in the manuscript by revising the text as follows:

"However, there is a clear difference between alignments between 2220-2110 ka, overlapping with the period of high age model uncertainty for the benthic foraminiferal $\delta^{18}O$ alignment to Prob-stack reported by HMM-Match, when *the magnitude of absolute* age model offsets between the Prob-stack-based alignment and the IODP Site U1541 alignments to LR04 and to CENOGRID reach up to 60 kyr (Fig. 4). For the entire 3.5 Myr record, the average *magnitude of absolute* age offsets between the Prob-stack-based age model and those based on alignments to LR04 and to CENOGRID are 5.2 and 5.7 kyr, respectively (Fig. 4, Supplemental Table S5). Over the last 800 kyr, the period for which the LR09 benthic $\delta^{18}O$ stack is available for the Pacific Ocean, *absolute* age offsets between target alignment outputs are lower. Specifically, the average *absolute* age offsets between Prob-stack and LR04, LR09, and CENOGRID-based alignment ages are 2.5, 2.3, and 4.7 kyr, respectively, across this interval (Supplemental Table S5)." (Lines 475-483)

"The absolute values of the age offsets between the benthic foraminiferal δ^{18}O alignment of IODP Site U1541 to Prob-stack and to the LR09 Pacific stack vary from near zero up to 16 kyr, with an average of 2.3 kyr (Fig. 4)." (Line 488)

"Over the past 800 kyr, the average *magnitude of absolute* age offsets between the HMM-Match alignment of the benthic foraminiferal δ^{18}O record at ODP Site 1090 to Prob-stack and similar alignments to the LR04, LR09 Atlantic, and CENOGRID tuning targets are 3.8 kyr, 3.5 kyr, and 5.7 kyr, respectively." (Line 526)

"In comparison, the *magnitude of absolute* age offsets between the age models generated by automated HMM-Match-based benthic foraminiferal δ^{18}O alignments to the LR04, LR09 Atlantic, and CENOGRID targets at ODP Site 980/981 are larger, with a maximum value of 18 kyr between the alignments to Prob-stack and to LR04 at 1.42 Ma (Fig. 7). Over the last 800 kyr, the average *absolute* age offsets between automated HMM-Match-based benthic foraminiferal δ^{18}O alignments to Prob-stack and to LR04, LR09 Atlantic, and CENOGRID at ODP Site 980/981 are 2.0 kyr, 1.4 kyr, and 4.7 kyr, respectively (Fig. 7)." (Lines 563-569)

- The authors acknowledge that stratigraphic alignment of benthic δ18O records to a preferred dated target curve yields age model uncertainties associated with the assumption that the undated record and target experienced synchronous changes in benthic foraminiferal δ^{18}O values (e.g., l. 30-31). However, they should make clear that this source of age uncertainty is irreducible and that using regional stacks rather than global stacks only reduces this problem but does not eliminate it as long as the evolution of the water masses geometry across climate transitions is not precisely known. Therefore, a number of sentences should be amended to make this clear: e.g. l. 229-231, 420-425, l. 436-439. The latter sentence is misleading as such: "We employ Prob-stack as our benchmark tuning target because it is the most globally representative target and therefore contains the most realistic estimate of glacial/interglacial seawater δ^{18}O variations at any given point in the deep ocean that may be reflected in benthic foraminiferal δ^{18}O values" should be changed into "We employ Prob-stack as our benchmark tuning target because it is the most globally representative target, assuming synchronous changes in benthic δ^{18}O anywhere in the entire ocean".

We will clarify our language on this point. However, we note that we already dedicate many lines to the discussion of regional variability throughout the manuscript (including lines 105-122, 497-504, 646-662, and 770-772, in addition to those highlighted by Reviewer 2) and consider further elaboration on the importance of and limited constraints on this source of uncertainty beyond the scope of this work.

"This site also enables investigation into the age uncertainties associated with benthic foraminiferal δ^{18}O stratigraphies under favorable alignment conditions, where regional seawater *temperature and* δ^{18}O differences and associated temporal offsets in benthic foraminiferal δ^{18}O changes between the undated record and the *global LR04 and Prob-*

stack targets, which are heavily weighted by North Atlantic sampling, are expected to be minimal." (Lines 229-231)

"We employ Prob-stack as our benchmark tuning target because it is the most globally representative target, assuming globally synchronous changes in benthic $\delta^{18}O$ (Ahn et al., 2017)." (Lines 436-439)

On the same topic, the first sentence of section 5.2 (l. 633-635) is actually untrue, rather than misleading, and should be discarded: because benthic $\delta^{18}O$ changes are not synchronous anywhere in the entire ocean, Prob-stack does not include the most holistic estimate of spatial variance in benthic foraminiferal $\delta18O$ values across the last 5 Myr, and does not allow one to calculate realistic estimates of chronostratigraphic alignment uncertainties.

We will revise this sentence in the text as follows to clarify this point.

"We encourage the utilization of Prob-stack as a global tuning target because Prob-stack includes the most holistic estimate *to date* of spatial variance in benthic foraminiferal $\delta^{18}O$ values across the last 5 Myr, *under the assumption of global synchronicity in benthic foraminiferal $\delta^{18}O$ variability*, and therefore is best suited for calculating *more* realistic estimates of chronostratigraphic alignment uncertainties (Ahn et al., 2017)." (Lines 633-635)

- Also, l. 229-231 and l. 625-628, regional seawater $\delta^{18}O$ differences is not the only reason for temporal offsets between a given benthic $\delta^{18}O$ record and a $\delta^{18}O$ target: bottom water temperature is also very variable from one region and water depth to another and does also have a major impact on benthic δ^{18} Both sentences have to be corrected to make this clear.

Reviewer 2 is correct and we will update the text to reflect this point accordingly:

"In addition, temporal variations in bottom water temperature can introduce further discrepancies between the $\delta^{18}O$ values of benthic foraminifera and the bottom water (e.g., Marchitto et al., 2014; Elderfield et al., 2012)." (Introduction, Lines 75-80)

"This site also enables investigation into the age uncertainties associated with benthic foraminiferal $\delta^{18}O$ stratigraphies under favorable alignment conditions, where regional seawater *temperature and* $\delta^{18}O$ differences and associated temporal offsets in benthic foraminiferal $\delta^{18}O$ changes between the undated record and the *global LR04 and Prob-stack targets, which is heavily weighted by North Atlantic sampling,* are expected to be minimal." (Lines 229-231, as mentioned above)

"However, neither HMM-Match-based nor BIGMACS-based alignment age uncertainties account for the uncertainty associated with the absolute chronology of the target reference or regional variability in *bottom water temperature and/or* the timing of local changes in seawater (and resulting benthic foraminiferal) $\delta^{18}O$ between the input record and the selected tuning target." (Lines 625-628)

"The age model uncertainties caused by regional asynchronicity in *temperature and/or* seawater (i.e., benthic foraminiferal) $\delta^{18}O$ changes across glacial cycles must also be considered when comparing event timing among benthic foraminiferal $\delta^{18}O$-tuned sediment records from different basins and water masses." (Lines 646-648).

More minor comment:

- Abstract lines 40-43: as such, the sentence is unclear and groups information of different levels as if they were on the same level. It should be simplified and divided into several sentences.

We will rephrase this sentence for clarity as follows:

"Our analysis suggests average age uncertainties of 3 to 5 kyr *associated with manually-derived versus automated alignment, 1 to 3 kyr associated with automated probabilistic alignment itself, and 2 to 6 kyr associated with the choice of tuning target.*"

- There have been many studies on the relationship between benthic foraminifer $\delta^{18}O$ and bottom water temperature since the founding paper of N. Shackleton in 1974. Marchitto et al. (2014) have shown that *Cibicidoides* and *Planulina*, rather than *Uvigerina*, fractionate at equilibrium, and that *Uvigerina* is isotopically heavier than *Cibicidoides* and *Planulina* by 0.47‰ (that is actually very close to the offset found in the present study between *Cibicidoides* and *Uvigerina* in core U1541), in contrast to the historically used 0.64‰.

Reviewer 2 raises an important point. We will update the Introduction to include the results of more recent studies accordingly:

"Benthic foraminifera of the genus Uvigerina are considered an ideal benthic foraminiferal species for the generation of $\delta^{18}O$ stratigraphies because *they are* believed to calcify in equilibrium with seawater $\delta^{18}O$, although it occupies a shallow infaunal habitat (Shackleton, 1974). Other widely used species for these efforts include epibenthic foraminifera of the genus *Cibicidoides* or *Cibicides*, whose stable oxygen isotope composition is generally corrected by +0.64 ‰ to match presumable equilibrium seawater $\delta^{18}O$ values (Shackleton and Opdyke, 1973). *More recent studies have found disequilibrium effects between the $\delta^{18}O$ values of Uvigerina and Cibicidoides species that range between 0.47 ‰ (Marchitto et al., 2014) and 0.73 ‰ (Jöhnck et al., 2012), depending on local bottom water and pore water pH conditions. In addition, regional and temporal variations in bottom water temperature can introduce further discrepancies between the $\delta^{18}O$ values of benthic foraminifera and the bottom water (e.g., Marchitto et al., 2014; Elderfield et al., 2012). However, the premise of benthic $\delta^{18}O$ stratigraphy hinges on the representation of bottom water $\delta^{18}O$ by the $\delta^{18}O$ of benthic foraminiferal species of the genus Uvigerina (such as U. peregrina and U. hispida) and Cibicidoides (such as C. wuellerstorfi and C. kullenbergi), with a constant correction factor between the two (Lisiecki and Raymo, 2005). In other words, benthic $\delta^{18}O$ stratigraphy assumes the effect of bottom water and temperature variations in space and time to be minimal (Lisiecki and Raymo, 2005)."*

And will add the following studies to the reference section:

Jöhnck, J., Holbourn, A., Kuhnt, W., Andersen, N.: Oxygen isotope offsets in deep-water benthic foraminifera. Journal of Foraminiferal Research, 51, 225-244, https://doi.org/10.2113/gsjfr.51.3.225, 2021.

Marchitto, T., Curry, W., Lynch-Stieglitz, J., Bryan, S., Cobb, K., and Lund, D.: Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera. Geochemica et Cosmochimica Acta, 130, 1-11, 2014.

- The sentences line 199-202, and 215-219 are pointless with respect to the topic of the paper.

We will remove these lines from the text to streamline the paper.

- L. 225: "consistent" should be changed into "constant"

We will change this word to "constant."

- The approach used to align the XRF signals of core PS75/059-2 and U1541 should be specified: is it manual or automated? If an automated approach has been used, which one is it?

We will clarify this point by updating the text in Section 2.2 to say:

"To combine the benthic foraminiferal $\delta^{18}O$ records from both cores, the PS75/059-2 data were mapped onto a common U1541 depth scale via *a manual* stratigraphic alignment of high-resolution X-ray fluorescence (XRF) Fe intensity variations in PS75/059-2 (Lamy et al., 2014) and U1541 that result in 22 tie points (Supplementary Fig. S1 and Supplementary Table S1)."

- L. 277-281: I do not see what is the point of describing the original stratigraphic alignment of core ODP980/981. Also, why speak of the original stratigraphic alignment of that core and not of core ODP1090 ?

We thank the reviewer for catching this error. We had included the original stratigraphic alignment details for both ODP Site 1090 and ODP Site 980/981 in Section 2.1 and we will remove lines 277-281 in Section 2.2 from the manuscript, accordingly. These details were included in Section 2.1 because they form the basis of the segment start and end tie points input into the HMM-Match algorithm for the automated alignments of ODP Site 1090 and ODP site 980/981.

- L. 463-464: "in the upper 85 m" is not a relevant piece of information in this article.

We respectfully disagree with this assessment. The full length of the IODP Site U1541 sediment record is over 137 m CCSF-A and we wish to clarify that the benthic foraminiferal oxygen isotope data that we measured for this site only span the upper 85 m CCSF-A.

- The sentence l. 557-559, which is repeated l. 601-603, is so obvious that it does not seem to deserve inclusion in a publication.

We respectfully disagree with this assessment as well. We intend for our paper to be informative and accessible to students and cross-disciplinary geoscience readers as well multidisciplinary paleoclimatologists.

- L. 654 should be rephrased: a stack cannot be "radiocarbon dated", only individual d18O records can be radiocarbon dated.

We will clarify this important point by updating the text at this point to the following:

"Regional stacks of individual benthic foraminiferal $\delta^{18}O$ records with high-resolution radiocarbon dates over the last glacial cycle suggest that the spatial variability in seawater $\delta^{18}O$ and its response to changing climate may induce systematic age offsets up to 1.7 kyr between the deep South Pacific IODP Site U1541 record and ODP Sites 1090 and 980/981 in the deep South Atlantic and deep North Atlantic, respectively, during glacial terminations (Stern and Lisiecki, 2014)."

Reference

Marchitto, T., W. Curry, J. Lynch-Stieglitz, S. Bryan, K. Cobb, and D. Lund (2014), Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera, *Geochimica et Cosmochimica Acta*, *130*, 1-11.

We will add this reference to the reference section, as mentioned above.