Review “Southern Ocean bottom water cooling and ice sheet expansion during the middle Miocene climate transition” by Leutert et al.

Response to Referee #1

Please find below the referee’s comments in blue font and the authors’ response in black font.

General comments:

Leutert et al present an interesting new record of bottom water temperatures from the Kerguelen Plateau during the middle Miocene – a time of substantial ice-sheet growth and cooling. The record will be a valuable contribution to our understanding of ice volume vs temperature changes in this interval. A revised age model for ODP Site 747 is presented and seems to be robust. New benthic stable isotope data match well with existing records. The paper is overall well-written; however, I suggest a substantial overhaul of the discussion.

Reply: We are sincerely grateful for the thoughtful and constructive comments of Referee #1 on our manuscript. Importantly, we will follow the referee's advice and substantially revise the discussion of our new record adding more details about possible climate mechanisms and water circulation during the middle Miocene (see below).

The stand-out feature of the new temperature record is a large, transient (0.8 Myr-long) cooling of 3-5°C during the middle Miocene climatic transition, between ~14.5 and 13.7 Ma. The fact that cool temperatures are recorded in three consecutive intervals (each made up of ~30 analyses) suggests it is a robust signal. Because this large cooling occurs during an interval with only a small increase in benthic $\delta^{18}O$, the implication is that it was accompanied by significant de-glaciation lasting ~0.8 Myr (shown by the large decrease in bottom water $\delta^{18}O$). This aspect of the record (its plausibility and implications, possible mechanisms that might have caused it, whether there is any other evidence for deglaciation at this time) are not discussed in enough detail in the paper. For example, the large step decrease in bottom water $\delta^{18}O$ at ~14.5 Ma is barely mentioned. No clear explanation for the cooling is given (although the sub-sequent warming is discussed).

Reply: We agree that we have given this early cooling too little attention in the previous version and will discuss it more prominently. We will put forward two possible drivers for such an early cooling: a relationship to expanding ice sheets or circulation changes in the deep ocean caused by tectonic processes accompanying the opening of Drake Passage and Scotia Sea (e.g., Lagabrielle et al., 2009; Pérez et al., 2021) and/or the closing of the eastern Tethys gateway (e.g., Hamon et al., 2013; Steinthorsdottir et al., 2020; Woodruff and Savin, 1989). However, large uncertainties in the exact timing of these ocean gateway changes, which may have affected Southern Ocean bottom waters and Antarctic ice volume to different extents, hamper an unambiguous correlation. The similar early MMCT cooling observed at Sites 747 (Kerguelen Plateau, Southern Ocean) and 806 (tropical Pacific; Lear et al., 2015) suggests that the Southern Ocean bottom water signal was at least transferred into the Pacific Ocean basin.

We will also extend the discussion of possible mechanisms that may explain the observation of low bottom water $\delta^{18}O$ in times of comparably low BWTs. We propose that a possible alternative interpretation to a transient deglaciation could be a regional bottom water freshening and destratification event, explaining the concurrence of low bottom water $\delta^{18}O$ and low BWT at Site 747, even in the case of only limited deglaciation on Antarctica. The latter is difficult to examine in a conclusive manner here, as the section from 14.4 Ma to 13.8 Ma is missing in the AND-2A core (Levy et al., 2016). Although records from two sites offshore East Antarctica, Wilkes Land IODP Site U1356 and Prydz Bay ODP Site 1165 indicate low ice-rafted detritus values (Pierce et al., 2017) and thus indirectly point to limited ice at that time, more proxy records from Antarctica and its continental shelves as well as additional BWT and $\delta^{18}O_{bw}$ records from different sites and water depths in the Southern Ocean may allow for a better understanding of the features of our Site 747 bottom water record in the future (this will be pointed out in the Discussion of the updated manuscript version).

There is very little discussion of bottom/intermediate water circulation, which water masses might have bathed the site and how this might have changed over the study interval, deep-water formation (e.g. proposed Miocene onset of Antarctic Bottom Water Formation in the Weddell Sea, Pérez et al., 2020), changes in Antarctic gateways, etc. that may have influenced the temperature record.

Reply: See our proposed changes above.

Importantly, the reader does not know what the Miocene paleodepth of the site was and to what extent benthic forams at this site might record local versus global temperature signals.

Reply: The benthic foraminiferal species composition from the middle Miocene sequence from Site 747 is characterized by common deep-sea faunal components and strongly resembles the corresponding middle Miocene sequences from Holes 689B and 690C (Maud Rise), as pointed out by the shipboard scientific party. This evidence supports a lower bathyal to abyssal depth at Site 747 during the middle Miocene (Schlich et al., 1989). We will add a sentence with this information in the main manuscript, but are hesitant to include this more prominently, as we could not find any more precise quantitative estimates of middle Miocene paleodepths for Site 747 and also no evidence for a middle Miocene location of Site 747 in a shallow water environment (e.g., Abrajevitch et al., 2014; Billups and Schrag, 2002; Majewski et al., 2010; Verducci et al., 2009). To the contrary, water depths at Site 747 during the middle Miocene may have even been even larger than today (Schlich et al., 1989). In absence of contrary indications, we interpret Site 747 as recording signals which at the very least are reflective of the Indian Ocean sector of the Southern Ocean, but likely are reflective of processes on a larger scale. In any case, we will point to the uncertainty in the scale of the signal at various positions in the text.

The bottom-water temperature trends at Site 747 (based on Δ_{47}) are quite similar to those seen at Site 806 based on Mg/Ca but not other sites, which is really interesting. Is there a water mass/circulation-related explanation for this?

Reply: The similar BWT patterns reconstructed at Sites 747 (Kerguelen Plateau, Southern Ocean) and 806 (tropical Pacific; Lear et al., 2015) suggests that the Southern Ocean bottom water signal was transferred into the Pacific Ocean basin. This interpretation may imply deep water formation in the Southern Ocean and an ocean gateway configuration similar to today, with an active Antarctic Circumpolar Current and continuous export of deep ocean water masses formed in the Southern Ocean to lower latitudes. This interpretation will be included in our discussion. Possible causes for the differences to other records may include (but are not limited to): Regional differences between water masses bathing these sites, variable pore water chemistry (e.g., bottom water carbonate ion saturation effects on benthic foraminiferal Mg/Ca when measured on epifaunal species as in some of these records), diagenetic effects, data gaps in proxy records based on only one hole (such as those from ODP Sites 747, 761 and 1171) and aliasing due to low-resolution sampling.

Specific comments:

I have a couple of suggestions to improve Figure 1: Firstly, I would use a different (more inclusive) colour scale for the temperature map, as the rainbow colour scale is now widely known to be a poor choice both for colour-blind people and also for reproduction in grayscale.

Reply: We thank Referee #1 for pointing this out. We will change the colour scale to a colour scale going from blue over white to red. In addition to avoiding rainbow colours scales and the introduction of false perceptual thresholds (e.g., Hawkins, 2015), this type of colour scale also appears to be a better choice for colour-blind people (see "https://colorbrewer2.org").

Secondly, I find the plate tectonic reconstruction shown in this figure difficult to interpret, because it shows tectonic plates including ridges and continental shelves, rather than a land-sea mask or reconstructed bathymetry. I suggest that the authors use instead a paleogeographic map which would more clearly show the distribution of continents and oceans and the paleodepths of sites; e.g. the Scotese paleogeographic reconstruction maps (Paleomap project); Straume et al. 2020 (paleobathymetry reconstructions available at 1 Ma resolution: https://zenodo.org/record/4193576#.YAb_heB7lXh); or Cai et al 2017 (which includes digital global paleogeographic maps in the supplement, including a 14 Ma reconstruction).

Reply: We will replace the plate tectonic reconstruction in Fig. 1 with the more recent paleogeographic map of Cao et al. (2017), which more clearly shows the distribution of continents and oceans, and provides some (limited) information about the paleodepths of the sites.

Introduction

"The middle Miocene geographic position of Site 747 relative to Antarctica was similar to today"; I found this statement a bit lacking in detail on paleolatitude, setting, etc., so I suggest expanding on this.

Reply: Following the referee's advice, we will add an estimated paleolatitude range for Site 747 from 16 Ma to 12 Ma.

Also the paleodepth of the site is not discussed – could a shallower paleodepth contribute to the relatively warm temperatures you reconstruct compared to modern, and the relatively large changes?

Reply: An effect of a shallower paleodepth on our bottom water temperature record is in principle possible and will be indicated in the main manuscript. However, we would consider such an effect as minor, as we did neither find any evidence for a middle Miocene Site 747 water depth that was shallower than today nor for changes in paleodepth (see also our reply above). In addition, the 2013 World Ocean Atlas dataset (Locarnini et al., 2013) indicates comparably small changes in (annual mean) temperature with depth below ~1000 m (e.g., by around +0.5–1°C from 2500 m to 1500 m water depth) around Site 747. Last, the good agreement of the $\delta^{18}O$, $\delta^{13}C$ and the clumped isotope BWT values from Site 747 with those from Site 761 (Modestou et al., 2020) supports the interpretation that a substantial paleodepth effect on our Site 747 temperature estimates is unlikely.

The Δ_{47} temperature proxy is well introduced, however given that you list all the potential caveats of the Mg/Ca paleothermometer as applied to benthic foraminifera, I feel the Δ_{47} proxy gets off quite lightly. A brief summary of the potential impact of diagenesis (dissolution, recrystallization, and overgrowth), burial, or other known non-thermal processes on Δ_{47} in benthic foraminifera and their effect on reconstructed temperatures would be useful, even though you discuss this in detail later.

Reply: We fully agree that equal skepticism to all proxies is critical and wish to clarify here that we have attempted to treat all used temperature proxies in an equally critical manner, discussing existing complications and limitations that are relevant for our conclusions thoroughly. However, we will modify the introduction to once more prominently point out the comparably large analytical uncertainties of the clumped isotope thermometer, which in our view pose the main limitation of this technique at the moment. As implied by Referee #1, it is correct that the Δ_{47} proxy can be susceptible to post-depositional diagenetic processes in certain settings, similar to other more traditional geochemical proxies such as Mg/Ca and $\delta^{18}O$. However, when we investigated diagenetic effects (Leutert et al. 2019), to the best of our knowledge the only published study specifically investigating the impact of post-depositional diagenesis on foraminiferal Δ_{47}, we found no detectable effects of diagenesis on the Δ_{47} signatures of middle Eocene benthic foraminifera. This was the case even at pelagic carbonate-rich sites (similar to Site 747) and despite visible signs of diagenetic alteration (e.g., overgrowths of coarse inorganic crystallite), and is supported by evidence based on modeling the effect of diagenesis with reasonable boundary conditions. Compared to the middle Eocene benthic foraminiferal specimens analysed in that diagenesis study, the benthic foraminiferal tests analysed here are from the middle Miocene and thus much younger, making a diagenetic bias even less likely. Of course, there is no absolute certainty when interpreting climate signals from chemical signatures of foraminiferal carbonates as old as the middle Miocene. Therefore, we carefully assessed and documented preservation states of representative specimens (e.g., scanning electron microcopy) and acknowledge this potential source of uncertainty in our proxy discussion. In addition, we will include a sentence on diagenetic effects in the introduction to point this possibility out already earlier, as suggested by Referee #1. However, in the light of the findings from our diagenesis study and the lack of evidence for a diagenetic effect on benthic foraminiferal Δ_{47} in burial settings comparable to that of Site 747, we think it is reasonable to not put a main focus on diagenetic effects on benthic foraminiferal Δ_{47} (or Mg/Ca) in the introduction of our study, and instead focus on non-thermal effects on Mg/Ca during biogenic calcite precipitation. These non-thermal effects are clearly much less of a complicating factor for the clumped isotope thermometer in the setting of this study (see also the recent commentary of Evans (2021)).

Methods/Results & Discussion: I think it would be clearer if the Results and Discussion were separated.

Reply: Results and Discussion will be separated.

Age model: I would move the Age Model section up so that it follows the Site Details section.

Reply: Will be done.

In addition, an age-depth plot for Site 747 (in the supplement if necessary) showing all of the different tie points used (magnetostratigraphy, isotope-based, biotratigraphy) and the described hiatus would be very useful.

Reply: We will add an age-depth plot as a supplementary figure summing up the used tie points, in addition to showing sampling, hiatus and core transitions.

Is the assumption that Site 806 sedimentation rates were constant and similar either side of the orbitally-tuned record between 14.1 and 13.3 Ma supported by shipboard magnetostratigraphic and biostratigraphic datums? I would verify this if you have not already, especially given that this is the record that has the most similar trends to your new record. With this assumption, the comparison is not very robust. Presumably the original publication of the Mg/Ca record had age constraints that covered the whole interval?

Reply: We will change our strategy for the Site 806 age model, updating biostratigraphic events from Kroenke et al. (1991) and Chaisson and Leckie (1993) to the GTS2012 timescale (Gradstein et al., 2012) to complement the orbitally tuned Holbourn et al. (2013) age model (instead of assuming constant sedimentation rates from ~16.4 to ~14.1 Ma and from ~13.3 Ma to ~12.3 Ma). We prefer this approach over just taking the original age model of Lear et al. (2015), as the latter age model is not on the GTS2012 timescale and, more importantly, has been optimized for a much longer time interval (~18–0 Ma) and thus not specifically for the middle Miocene sequence at this site. Encouragingly, the updated Mg/Ca-based BWT record from Site 806 shows an even better fit with the clumped isotope BWT record from Site 747. Also, we note that the interpretation of an early cooling across the MMCT (compared to the stepped benthic $\delta^{18}O$ increase) is robust and unaffected by the choice of the Site 806 age model.

The calculation of uncertainties should be briefly described, rather than just referring to the supplement of another paper.

Reply: We will add a brief description of the error propagation in Appendix A.

"Results from adjacent samples are pooled to achieve this number of measurements" Please be more precise about how many adjacent 2-cm samples were pooled together (mean, min, max depth/age intervals over which results were averaged).

Reply: Mean, min and max depths as well as the number of adjacent samples will be added in a supplementary table to make sure that all information requested by the referee is provided (in addition to the supplementary table showing individual replicate measurements with the corresponding depths and ages).

Samples were run on two different machines, but as far as I can see we cannot tell from the figures which data were run on which machine. It might be useful to colour code data points in Figure S3 to show that there are no machine offsets.

Reply: Data points will be coded machine-specific in the supplementary figure. In any case, we note that significant machine offsets are extremely unlikely with our data processing procedure using and normalizing to an identical set of carbonate standards (see also colour-coded data points in Fig. 1 of this response).

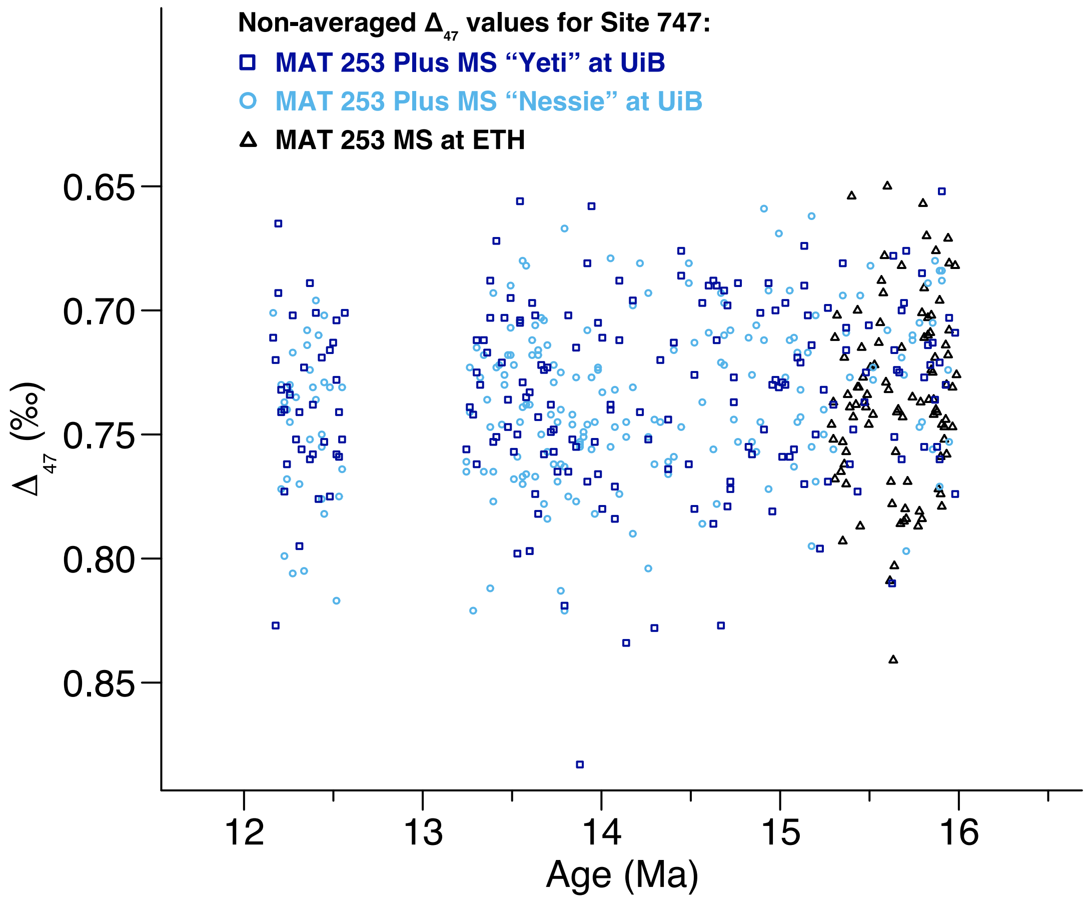

Fig. 1: Comparison of non-averaged Site 747 Δ_{47} values that are colour-coded for each used mass spectrometer at the University of Bergen (UiB) and ETH Zurich.

Are the cited external reproducibilities for both of these machines?

Reply: The cited external reproducibilities refer to the performances of all machines that have been used for this study. In the supplementary tables, separate external reproducibilities for each machine and for each standard are listed for all relevant measuring intervals.

As a side note, I feel like Figure S3a should be shown in the main text (maybe as a top panel in Fig. 3), as it shows the raw data upon which all your subsequent data averaging and interpretations are based.

Reply: We are hesitant to move Fig. S3a in the main text, as we do not think that these "raw" Δ_{47} values can be interpreted in terms of paleoclimate (at least not without further processing/averaging, as pointed out in the main manuscript). Having Fig. S3a in the main text may thus be misleading. We note that this figure will be prominently referenced in the main manuscript.

Fig. 3: horizontal solid lines: averaging intervals; it is not clear to me why the points are not plotted in the middle of the averaging intervals. Is the age of the points weighted towards the highest data density?

Reply: This is correct. As pointed out in the caption of Fig. 3, the position of a plot on the x-axis simply shows the average age over all replicates that were used in a bin. Of course, this implies that for example a sample that has been measured twice (2 replicates) was weighted double. For clarification, we will add this information once more also in the methods (Chapter "2.4 Isotope measurements and data processing").

Why was a 400-kyr moving window approach used rather than a Gaussian-Weighted Filtering approach, as in Modestou et al 2020? I am not sure which method is most appropriate, but the Gaussian-Weighted Filtering approach does seem to smooth out the small-scale features noted by the authors to be caused by scatter in measurements.

Reply: We have tested a lot of different approaches to visually guide the eye including LOESS-based techniques. A LOESS fit has been previously applied to smooth a similar type of clumped isotope record (Leutert et al., 2020). Having weighed up the advantages and disadvantages of all approaches, we have finally decided for the 400 kyr-moving window approach here, which is comparably simple and easy to understand. This type of smoothing does not only allow for a straightforward comparison between records from different sites and minimizes artefacts caused by uneven sampling, parameter selection and edge effects (LOESS) but also allows for temporally shorter averaging intervals in comparison to other approaches such as Gaussian window filters. Note that Modestou et al. (2020) used a 1000 kyr window size for their Gaussian window filter, which is 2.5 times larger than the 400 kyr window size use here. In our setting, a 1000 kyr window size would make potential biases in the timing of the changes more likely. Using a relatively "simple" moving average without any Gaussian weighing also makes it possible to transparently point to the parts of the record that are based on fewer (<30) measurements and thus less certain (e.g., Fernandez et al., 2017). These advantages of the 400 kyr-moving window approach are weighted more heavily here, than the artefact of minor small-scale features that are not smoothed out. In any case, we note that we provide replicate-level clumped isotope data to allow any reader to reproduce our smoothing or adjust the smoothing for other applications, in addition to applying an alternative approach to visualize the clumped isotope BWT timeseries (binning). Most importantly, our interpretation of the temperature record (early cooling of ~3–5°C, transient smaller warming) appears robust toward different smoothing approaches such as different LOESS fits; we will add a new supplementary figure for illustration.

Add an error bar for Mg/Ca-based temperatures.

Reply: We will add an error bar for Mg/Ca-based temperatures in Fig. 3 illustrating the typical uncertainty introduced by sample reproducibility and calibration errors (±1°C; Lear et al., 2015). In this context, however, it is critical to distinguish between random and systematic errors. Random errors can be relatively easily quantified by comparing multiple measurements. In contrast, the quantitative estimation of systematic errors can be difficult or even impossible with available knowledge, as the cause of the error must be identified and quantified for error estimation. We previously propagated errors and included confidence intervals wherever we considered it possible, meaningful and potentially relevant for interpretation. For Mg/Ca-based temperatures, we had avoided plotting error bars due to known systematic non-thermal influences (such as seawater Mg/Ca or the error in Mg/Ca-based temperature estimates caused by saturation state effects) limiting the informative value of such an error bar. In contrast to Miocene Mg/Ca-based temperature errors, the error in clumped isotope temperatures is mostly caused by random analytical errors and thus much easier to understand, propagate and quantify.

On Figures 2 and 3, it would be helpful to highlight the middle Miocene climatic optimum and transition intervals, and also the hiatus.

Reply: MMCT, MCO and hiatus will be highlighted in Figs. 2 and 3.

Line 192 – again please specify how large/variable the intervals over which data were averaged are in the text.

Reply: See our previous comment on Page 8 of this reply.

"We note that small-scale features in the moving average curves are likely caused by the scatter in the underlying individual Δ_{47} measurements, and should not be interpreted as real climate signals" For clarity, please quantify small- scale (<X ◦C) in this sentence.

Reply: "(around 1°C or less)" will be added, as suggested.

Lines 200-203 (and throughout the results and discussion): I suggest citing temperature confidence intervals (± x◦C at x CI) when describing absolute values, this will help to emphasise which trends are significant given the large error bars on Δ_{47} temperatures (e.g. a 3-5°C cooling is larger than 68% CI).

Reply: The corresponding lines will be adjusted following the advice of Referee #1. At these lines, we will also add confidence intervals for relative changes, whereas in the abstract we prefer to list BWT values without uncertainties, as the exact uncertainty range depends on the exact time interval (whose exact definition is beyond the scope of the abstract). Furthermore, we note that "substantially (~3–9°C) warmer bottom waters" will be changed to "substantially (by up to ~9°C) warmer bottom waters" to be more conservative.

Line 218: How do the recalculated bottom-water temperatures from Site 761 compare to the originally published values?

Reply: The recalculated values are well within uncertainty, and are truly essentially indistinguishable. We will add the values based on the Kele et al. (2015) calibration (updated in Bernasconi et al. (2018), originally used by Modestou et al. (2020)) to a supplementary figure, highlighting the good agreement (well with uncertainty) between the original (Modestou et al., 2020) and the recalculated clumped isotope-based BWT values from Site 761.

Line 229: What artefacts could result from comparing a low-resolution record of discrete samples (each representing maybe 1-2000 years, without knowing if it is a glacial or an interglacial) with a record where each sample integrates hundreds of thousands of years?

Reply: We minimize aliasing in our new clumped isotope temperature record from Site 747, using at least nine adjacent sediment samples and even more separate measurements for each clumped isotope temperature estimate. The large number of foraminiferal tests used for each temperature thus largely prevents aliasing in our Site 747 clumped isotope temperature record. In addition, we calculated clumped isotope temperature using two independent averaging approaches (described in Material and Methods), making our observations for this site even more robust. However, we cannot exclude some degree of aliasing in the Site 806 Mg/Ca-based temperature record (Lear et al., 2015), due to a much lower sampling density and much smaller numbers of foraminiferal tests per temperature estimate (limited temporal resolution of Site 806 Mg/Ca record is cautioned in the Discussion).

Line 269: do the authors have any suggestions as to how to investigate this?

Reply: The specific effect of dissolution on benthic foraminiferal Δ_{47} could be assessed by laboratory experiments or by analysing samples from a depth transect including sites at different distances from the carbonate compensation depth, similar as has been done in the equatorial Pacific for benthic $\delta^{18}O$ (Edgar et al., 2013).

Line 288: include d18Obw errors in the text. "For the later MCO (15.6–13.9 Ma), our estimates of $\delta^{18}Obw$ range from around -0.3 ‰ to 0.7 ‰´' This statement doesn't really adequately describe the large step changes in reconstructed bottom water δ18Obw at ~14.5 Ma and 13.7 Ma.

Reply: In the previously submitted version, we intended to begin our discussion on bottom water $\delta^{18}O$ with a broad overview of the observations from both Sites 747 and 761, and then focus on a more detailed discussion of Site 747 bottom water $\delta^{18}O$ and its evolution in the following sentences and paragraphs. We will restructure our discussion of bottom water $\delta^{18}O$ including a more detailed description of its temporal evolution and specifically the stepped changes in reconstructed $\delta^{18}O_{bw}$ at ~14.5–13.7 Ma (including more details on possible effects on bottom water $\delta^{18}O$, water masses and mechanisms). However, given the possibility of additional biases on $\delta^{18}O$ (such as pH or other physiological effects in foraminifera), we prefer to discuss only three approximate $\delta^{18}O_{bw}$ ranges without $\delta^{18}O_{bw}$ errors bars in the text (with detailed error bars given in Fig 5e).

Line 294: due to their temporal resolution and also due to averaging of many samples probably mixing glacial and interglacial climate states.

Reply: We agree that this addition may make the sentence easier to understand, and will modify the corresponding lines following the suggestion of Referee #1.

Line 326: what was the interpretation of this change in vertical gradient?

Reply: Majewski and Bohaty (2010) interpret this change in vertical $\delta^{18}O$ gradient as reflecting a significant decrease in surface water salinity (freshening) across the stepped main increase in benthic $\delta^{18}O$ during the MMCT. This interpretation is also supported by our study and previous studies (e.g., Leutert et al., 2020). For clarification, we will include the interpretation of Majewski and Bohaty (2010) more prominently and closer to the text passage, where we are referring to the change in vertical gradient observed by these authors.