Final Author Comment

De Vleeschouwer et al. – cp-2021-151

We thank referee R3 for their in-depth and constructive assessment of our manuscript. The first major comment refers to the TEX_{86} – clumped isotope discrepancy and we fully agree with the reviewer in this matter. The second comment contains several questions as to which environmental factors could force the isotopic gradient. To answer these questions, we present new arguments that indicate mixed-layer temperature gradient along the 19-29°S transect as the dominant driver. The third comment raises some legitimate concerns about clumped isotope statistics. In answer to these concerns, we justify certain choices and follow the reviewer in most of their requests. The fourth comment relates to the clumped isotope calibration curve by Peral et al. (2018): The recalculation of this calibration will be published in the next couple of weeks by *Geochimica et Cosmochimica Acta* (GCA). All minor comments raised by referee #3 are adopted.

1. TEX_{86} vs. clumped temperatures

Firstly, I have some concerns about the authors' conclusions that TEX86 likely reflects sea surface temperatures while the clumped isotope analyses are indicative of the temperatures in the lower mixed layer, as put forward in the abstract (lines 24-26). In their discussion about the discrepancy between TEX86 and clumped temperatures, the authors put forward the hypothesis that TEX86 may be seasonally biased (lines 343-350). In fact, previous studies have suggested that a summer bias on TEX86 is likely and explains the consistent difference between TEX86 and stable isotope paleotemperature estimates found in other studies (Jia et al., 2017; O'Brien et al., 2017). While they disregard their other working hypothesis about the southward displacement of sinking particles, it seems that the authors cannot exclude the possibility of seasonal bias in their TEX86 data (see line 387), which could easily explain the ~5°C difference between TEX86 and clumped results (line 283) if TEX86 represents summer SST and clumped records MAT (see Fig. 1). In absence of clear evidence about the living habitat of T. sacculifer (lines 371-385), and with the only other line of evidence being that the TEX86 temperatures "seem reasonable compared to the present-day mean annual temperatures" (lines 341-342), I think the conclusion that clumped isotope temperatures represent the lower mixed layer and TEX86 represents the mean annual SST is not sufficiently supported.

The reviewer is absolutely right: We cannot exclude the possibility of a seasonal warm-bias in the TEX_{86} data. In fact, there is good evidence that a seasonal warm bias is affecting our TEX_{86} data from Site U1459. In lines 345 – 347, we already indicated that observations by Smith et al. (2020) and Benjamin Petrick (pers. comm.) at respectively Sites U1463 and U1460 suggest a ~3°C warm bias using the BAYSPAR surface temperature calibration (see Fig. S6 in Smith et al., 2020, where core-top TEX_{86} temperatures are systematically higher than mean annual temperatures at those localities). This warm bias more than likely reflects a seasonal bias, as the observed ~3°C difference roughly corresponds to the difference between present-day summer sea surface temperature (~25°C) and mean annual sea surface temperature (~22.8°C). Hence, we are on the same page with the reviewer. We admit that lines 24-26 in the abstract are short on detail, and that we shouldn't have labeled this seasonal warm bias as "minor" in line 357. This will be fixed.

However, a ~3°C warm bias on the TEX_{86} temperatures does not fully resolve the discrepancy with the clumped isotope temperatures. A systematic difference of ~2°C remains. The reported clumped-isotope calcification temperatures range between 18.2 – 20.8°C, suggesting a lower mixed layer habitat depth for *G. sacculifer*. On the one hand, we will elaborate on a possible cold-bias on the clumped temperatures in §4.1 in response to comments by Clara Bolton (R1) and R2. On the other hand, we will emphasize that strategies were implemented to minimize the clumped-isotope cold-bias (avoiding partly dissolved specimens and cleaning protocol to remove microcrystalline sparitic cements).

2. Isotopic gradients

While I understand that the authors only measured clumped isotope temperatures in one site, it is a shame that their discussion of the isotopic gradient along the Leeuwin Current does not benefit from the addition of clumped isotope analyses. My very first question on reading this discussion after the discussion on the clumped isotope results is how much of this isotopic gradient reflects temperature gradient and how much reflects the difference in seawater isotopic composition. Would a strengthening or weakening of the Leeuwin Current affect both these variables similarly? Would it somehow be possible to infer from the changes in temperature and seawater oxygen isotope composition over time, which the authors can infer from their clumped isotope record, whether the changes observed in δ18O over time are mostly driven by temperature or water composition? And by extension, could this evidence be used to say something about which factor predominantly forced the changes in isotopic gradient? Finally, if the author's hypothesis that the foraminifera calcify in the lower mixed layer is correct (see previous comment), how does this impact the discussion of isotopic gradient? Can the authors somehow exclude that changes in the calcification depth or the depth of the mixed layer between the two sites which are compared affect the difference in δ18O without the need for a change in the strength of the Leeuwin Current? I feel that there is some untapped opportunity for discussion on this topic which would integrate the clumped isotope analyses more firmly into the main discussion of the manuscript.

This question combines two comments made by Clara Bolton (R1) and R2:

a) In response to Clara Bolton, we provide an estimate of salinity changes on G-IG timescales. We infer a maximum of 0.6 psu salinity difference between phases of strong Leeuwin Current (Interglacial, present-day analogue: austral winter) and weak Leeuwin Current (Glacial, present-day analogue: austral summer). This estimate of salinity change on G-IG timescales is based on the present-day seasonal variability and likely represents an over-estimation because the Leeuwin Current never shut down completely during glacial intervals. Translated to $\delta^{18}O_{sw}$, this corresponds to 0.10 – 0.15‰, which is minor compared to the variability in the isotopic gradient (2‰).

b) In response to R2, we refer to the results of He et al. (2021). The black line in the plot below (i.e. Fig. 3C in He et al., 2021) shows the TEX_{86} gradient between Sites U1461 (also in the northern Carnarvon Basin, just as Sites U1463 and 763) and U1459 (Perth Basin). The low-resolution reconstruction of Leeuwin Current intensity by He et al. (2021) is in excellent agreement with our high-resolution $\delta^{18}O$-based reconstruction (orange line). As the TEX_{86} gradient is solely temperature-driven, the excellent co-variation provides strong support for interpreting the isotopic gradient as mainly temperature-driven. This additional line of argumentation will be implemented in the revised version of the manuscript.

Based on the above points, we make the case that the mixed-layer temperature gradient between 19 – 29°S, in response to Leeuwin Current strength, is the factor predominantly forcing the reported changes in isotopic gradient. We consider changes in the calcification depth or the depth of the mixed layer between the two end-members unlikely, as this would cause changes in the isotopic gradient that are not observed in the TEX_{86} gradient.
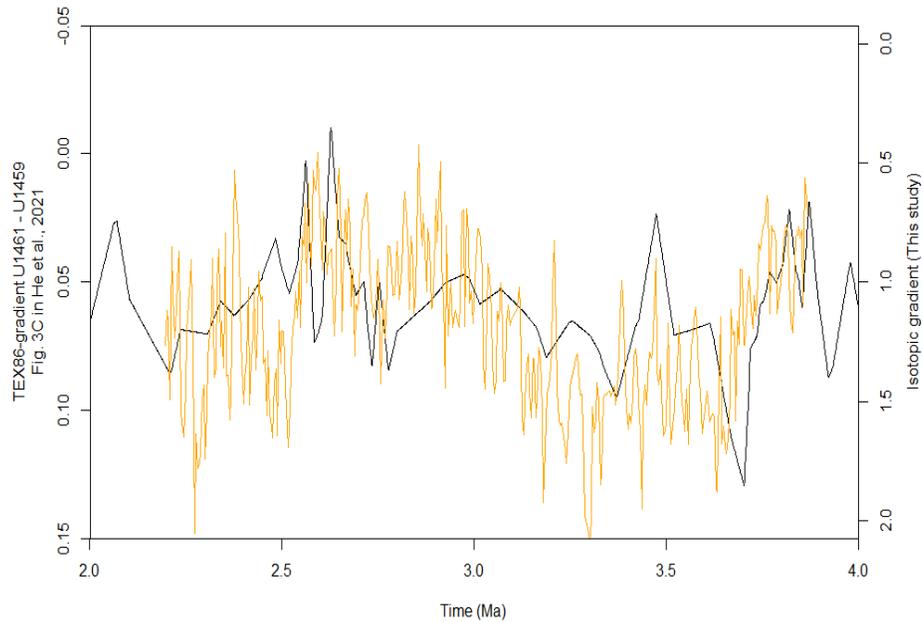
Comparison of the TEX86 gradient between Sites U1461 and U1459 (Fig. 3C in He et al., 2021) and the isotopic gradient from our work. Both proxies are completely independent of each other, yet show similar patterns throughout the Plio-Pleistocene, indicating the robustness of these results.

3. Clumped isotope statistics.

First of all, the caption of Table 2 (line 309) and the methods description (line 277) list different reproducibility errors for the clumped isotope measurements. I assume that the standard deviations cited in line 277 are one order of magnitude too high (e.g. 0.0314‰ instead of 0.314‰, as in line 309).

Indeed: Line 277 should report 0.0314 and 0.0173‰. This will be corrected.

Secondly, I noticed that the authors used the reproducibility of their standards for calculating the standard errors in Table 2 (see lines 307-309) instead of the within-sample reproducibility. This method is likely to underestimate the uncertainty on the D47 values in the samples, as the homogenized ETH-4 standard on which the standard deviation is based will likely reproduce better than the samples consisting of foraminifera pooled from up to four adjacent samples (line 189; up to 60 cm core depth when using the median sampling resolution from line 177). The authors should at least report the reproducibility of clumped isotope analyses within their samples.

The reviewer is correct: because of the low number of repeated measurements on samples, we chose to use the standard deviation on the standards to calculate the standard error. Our reasoning was that, with just a handful of repeated measurements on hand, using the sample standard deviation to calculate the standard error would be a case of *small number statistics*. We would like to point out that the measurement strategy with the NuCarb, compared to the more common MAT253+ setup, is such that relatively few repeated measurements are made with relatively large homogenized samples (500 micrograms vs. 80 - 120 micrograms). For this reason, in this particular case, we considered that the standard error on our sample measurements was best estimated by calculating the standard error from the standard deviation on the standards.

That being said, we will implement the reviewer's suggestion to report the reproducibility of individual sample clumped isotope analyses in an additional column added to Table 2. Furthermore, we will move away from solely using the standard deviation on standards to calculate

the standard error, and instead we will adopt the full propagation of analytical uncertainties in $\Delta 47$ measurements as described by Daëron (2021). He proposes to adopt a "pooled" standardization method that considers constraints on reproducibility available from both standard and sample analyses. This approach has been shown to yield realistic error estimates. In the text, we will add a statement as to why the "pooled" approach is the best way forward to estimate standard errors in this case with relatively small numbers of replicate measurements.

Moreover, additional aliquots of the eight clumped isotope temperatures have been measured to reduce the error bars on those temperatures. These will be included in the revised manuscript.

Thirdly, in the clumped community it is common practice to report uncertainties at the 95% confidence level (e.g. Fernandez et al., 2017). Instead, the authors report uncertainties at ±1 standard error in Tables 2 and 3. The captions of Figures 4 and 5 do not show what the error bars on the clumped datapoints represent, but from comparison with the tables I infer that these are also 1 SE. This reporting makes the uncertainty look smaller than in other studies using 95% confidence level and in my opinion the reporting of ±1 SE ("68% CL") is less intuitive. I realize that calculating 95% CL, or even the within sample standard deviation, of samples with 2 or 3 replicates (PB03, PB05, PB06 and PB08) is challenging due to the lack of statistics. This problem illustrates the risk of analyzing small numbers of replicates of samples and will make it challenging to assess the confidence on these clumped isotope datapoints, or to compare the results amongst themselves (e.g. via a Student's T-test) or with other data. I do not know how this issue can be resolved without adding additional replicates, and I do sympathize with the authors given how much work it is to gather enough foraminifera for these measurements. At the very least, I would therefore urge the authors to add information about their within sample reproducibility (standard deviations) for all samples and calculate 95% confidence levels for those samples for which this is feasible (sample size > 3), in addition to making the clumped isotope results available in an open-access repository (now, only regular stable isotope data is archived).

In the revised manuscript we will include additional measurements, such that all clumped temperatures will be based on a sample size > 5. Following the suggestion of the reviewer, we will show both the 68 and 95% confidence levels in the figures, and the 95% confidence levels in Tables 2 and 3. All figure captions will clearly indicate that the 68% and 95% confidence levels are shown, and that this uncertainty is randomly distributed. For the sake of consistency, we will also show the 95% confidence interval (2.5% - 97.5%) on the TEX$_{86}$ temperatures in Fig. 4

The clumped isotope final results (incl. temperatures) will be made available in the PANGAEA open-access repository upon publication of the manuscript, together with all other proxy data. Moreover, we will upload the raw clumped data (including machine data and correction steps) to the EarthChem repository. This way, EarthChem holds all data to allow future clumped experts to recalculate any data.

Finally, while not (yet) a standard in the clumped isotope community, it would be good practice if the uncertainty on the clumped isotope calibration(s) used in the study were to be propagated on the clumped isotope result. This uncertainty is not contained within the measurement uncertainty and is usually relatively small (<5 ppm). However, given the differences between the sample sizes and temperature ranges between the calibrations cited in Table 3, the differences in uncertainties of these calibrations could be discussed.

Ucertainties on the calibrations will be propagated to the temperature results in the revised version of the manuscript.

4. Recalculated clumped isotope calibration

It is a really nice addition that the authors compare the results of applying difference clumped isotope calibrations on their data (Table 3). This gives the reader a good intuition of the difference using different calibration makes in the study. Unfortunately, the I-CDES scale updated calibration of Peral et al. (2018) is not provided in the paper. This makes it impossible to verify the temperature results of this calibration. For the sake of open science, I urge the authors to make the I-CDES-scale referenced calibration dataset on which the updated Peral et al. (2018) calibration is based available in an open repository (e.g. Pangaea or EarthChem database) and to cite the new calibration formula with uncertainties on slopes and intercepts (sensu Equation 1) in the manuscript text or in Table 1 (as in Meinicke et al., 2021). Providing the calibreation dataset is especially important as the uncertainties on the calibration (see previous comment) cannot be propagated from the errors on the slopes and intercepts of the calibration formula alone, as information about the covariation of slope and intercept are missing from this information.

The I-CDES scale updated calibration of Peral et al. (2018) is currently under review with *Geochemica Cosmochemica Acta* and is expected to be accepted in the coming weeks. This anticipated work Peral et al. (2022) will contain the calibration dataset and will be cited in our manuscript as we did with Meinicke et al. (2021) for the Meinicke et al. (2020) calibration.

We will add the uncertainties on slopes and intercepts on all calibrations in Table 1.

5. Minor Comments
- Line 24-26, 39, 195, 451, 535 will be rephrased
- Line 457: This is an excellent suggestion: cross-spectral analysis will be added to underpin the co-variation between different time-series on different timescales.
- Figure 7: ("permille") will be added to the isotopic gradient axis, and "(unitless)" will be added to the composite axis.
- Figure C1: Pictures are from Gallagher et al. (2017): This will be clarified.

Daëron, M. (2021). Full Propagation of Analytical Uncertainties in Δ47 Measurements. *Geochemistry, Geophysics, Geosystems, 22*(5), e2020GC009592. https://doi.org/10.1029/2020GC009592

Gallagher, S. J., et al. (2017). Site U1459. In S. J. Gallagher, C. S. Fulthorpe, K. Bogus, & the Expedition 356 Scientists (Eds.), *Indonesian Throughflow.* (Vol. 356). College Station, TX: Proceedings of the International Ocean Discovery Program.

He, Y., et al. (2021). Development of the Leeuwin Current on the northwest shelf of Australia through the Pliocene-Pleistocene period. *Earth and Planetary Science Letters, 559*, 116767. https://doi.org/10.1016/j.epsl.2021.116767

Meinicke, N., et al. (2020). A robust calibration of the clumped isotopes to temperature relationship for foraminifers. *Geochimica et Cosmochimica Acta, 270*, 160-183. https://doi.org/10.1016/j.gca.2019.11.022

Meinicke, N., et al. (2021). Coupled Mg/Ca and Clumped Isotope Measurements Indicate Lack of Substantial Mixed Layer Cooling in the Western Pacific Warm Pool During the Last ~5 Million Years. *Paleoceanography and Paleoclimatology, 36*(8), e2020PA004115. https://doi.org/10.1029/2020PA004115

Peral, M., et al. (2018). Updated calibration of the clumped isotope thermometer in planktonic and benthic foraminifera. *Geochimica et Cosmochimica Acta, 239*, 1-16. https://doi.org/10.1016/j.gca.2018.07.016

Smith, R. A., et al. (2020). Plio–Pleistocene Indonesian Throughflow variability drove Eastern Indian Ocean sea surface temperatures. *Paleoceanography and Paleoclimatology, 35*(10), e2020PA003872. 10.1029/2020pa003872