Responses to Manfred Mudelsee:

RC2/RC3 (comments reposted with typos corrected):

An 800-kyr planktonic δ18O stack for the West Pacific Warm Pool

ESSD-2023-335

Review Manfred Mudelse 21 September 2023

The concept of the presented manuscript is fine: (A) stack construction for the WPWP, (B) usage of Bayesian age model algorithms, and (C) comparison of stack with other records (e.g., LR04) in terms of variability and spectral properties. However, there are flaws with data analysis and an amount of minor errors that render the current manuscript not publishable. My advice is to give authors enough time to re-submit a manuscript that overcomes data-analytical flaws and minor errors.

Data analysis

(1) Table 1

The MD97-2141 record stands out against the others in terms of temporal resolution. The manuscript should inform readers that indeed the 0.33-kyr resolution (Oppo et al. 2003) is a reasonable value since that record has a high sedimentation rate (5 to 15 cm/kyr) and also a fine sampling (1 cm). It would further be informative to study the effects of ex- or inclusion of that high-resolution record on results (variability and spectra). This can be done by repeating stack construction and variability and spectrum estimation without MD97-2141. Of course, keep MD97-2141 for calculation of the final stack.

Additional information about the sedimentation rate and sampling of MD97-2141 was added to the data section of the manuscript. For the revised version of the stack, the resolution of MD97-2141 was decreased one-fifth the original resolution (~0.55 kyr) which is closer to the other cores included in the stack (e.g., newly added core KX22-4, which has an average resolution of ~0.57 kyr [Zhang et al., 2021]). Although understanding the impact of record resolution on the stack could be interesting, we did not investigate a version of the stack with MD97-2141 fully excluded because of the high computational cost of producing each version of the stack.

The revised manuscript will read: "Published data for core MD97-2141 has an average sedimentation rate of 5-15 cm/yr and is sampled at 1 cm intervals with a mean sample spacing of 0.11 kyr (Oppo et al., 2003). However, we smoothed the data using a 5-point running mean sampled at every fifth point, which reduces its mean sample spacing to 0.55 kyr, so that this one record does not overly dominate the regional stack."

Zhang, Shuai; Yu, Zhoufei; Gong, Xun; Wang, Yue; Chang, Fengming; Lohmann, Gerrit; Qi, Yiquan; Li, Tiegang (2021): Precession cycles of the El Niño/Southern oscillation-like system controlled by Pacific upper-ocean stratification. Communications Earth & Environment, 2(1), https://doi.org/10.1038/s43247-021-00305-5

Zhang, Shuai; Yu, Zhoufei; Gong, Xun; Wang, Yue; Chang, Fengming; Li, Tiegang (2021): Sable oxygen isotope and Mg/Ca ratios of planktonic foraminifera from KX97322-4 (KX22-4). PANGAEA, https://doi.org/10.1594/PANGAEA.939377

(2) Output resolution of WPWP stack (end of Section 4.1; Bowman et al. 2023)

The stack has 8101 data points covering the interval from 0 to 810 kyr at a constant temporal resolution of 0.1 kyr. While it is fine to present such a stack for visualization purposes, it is not OK to use it for variability or spectrum estimation because the 0.1 kyr resolution is smaller than the individual resolutions (by a factor ranging from 3.3 for core MD97-2141 up to nearly 40 for core MD97-2140). This boost-up of the sample size may lead to significantly overstated claimed statistical uncertainties (for variability or spectrum estimation). One analysis strategy to assess the effect would be to repeat variability and spectrum estimation for various other, coarser stack resolutions (say, from 0.1 kyr up to 1.0 kyr in steps of 0.1 kyr). Such a sensitivity study could make an interesting appendix for other researchers wishing to study time-resolution effects.

> We subsampled the stack to a 1 kyr resolution before performing spectral analysis, which was not stated in the methods section; that has now been clarified. In the revised manuscript, all analyses of the WPWP and North Atlantic stacks will also be calculated at 1 kyr spacing.
>
> Repeating the variability and spectrum estimations at incremented temporal resolutions may be interesting to some readers; however, we consider it beyond the scope of this study. The high resolution version of the stack is available for those who would like to perform their own investigation of the effect of age spacing on the variability and spectra.

(3) Spectrum estimation (Section 4.3)

> Usage of FFT is obsolete since it renders bad (in terms of estimation bias, variance, RMSE, etc.) estimates. This is known for decades (Thomson 1982, Percival and Walden 1993, Mudelsee, 2014). And since the stack is evenly spaced (at 0.1 kyr or up to 1.0 kyr resolution), one needs not invoke the Lomb-Scargle Fourier Transform (Schulz and Mudelsee 2002) but can work with Thomson's multitaper estimation (MTM), which is the method of choice here. See again the mentioned works (Thomson 1982, Percival and Walden 1993, Mudelsee, 2014) and literature cited therein. Mudelsee (2014) lists also software tools for MTM estimation in case there is need for the authors.
>
> For the revision, all spectral analysis will be done using a temporal resolution of 1 kyr. We also agree with the reviewer that more modern methods have statistical advantages, and we will implement the suggested Thomson's multitaper estimation method in place of the FFT results in the revised version of the manuscript, using the Matlab function pmtm(). The new method has minimal impact on our findings.

(4) Uncertainty presentation of stack

Time-varying standard deviation is certainly interesting, but I think that more about stack uncertainty can be learned from calculation of internal and external errors (and hence use weighting for stack calculation). Internal errors refer to individual records, while external errors measure the spread among the various contributing records. Individual records with smaller uncertainties should, hence, stronger contribute to the stack. Of course the challenge is to do justice to the fact that the number of records available depend on the investigated age. Details about weighting, internal and external errors can be found in the paper by Mudelsee et al. (2014), who constructed a Cenozoic δ18O stack.

This manuscript focuses on the application of existing stacking software to planktonic d18O records of the WPWP. An alternate stacking methodology would be required for us to be able to separate internal versus external errors in the stack, which is outside the scope of this study. However, more explanation of the BIGMACS stacking approach, why it is appropriate for this application, and how it weights data across sites will be added to the manuscript. One key point here is the assumption made by BIGMACS that all records in the stack are homogeneous, i.e., that they all share the same underlying signal (with allowance for site-specific shift and scale values). Under this assumption, all residuals/errors are assumed to be internal errors associated with sampling noise and measurement uncertainty. Therefore, when stacking with BIGMACS, it is important to choose records for inclusion in the stack that share the same regional influence, and our analysis of the standard deviation of the new stack suggests that the planktonic d18O records we included meet this criteria because we find a similar spread in values about the mean as two published regional benthic d18O stacks, which each only included sites that shared the same deep water mass composition.

In BIGMACS stack construction, each data point included in the stack is weighted equally because all measurements are assumed to be drawn from the same underlying distribution. Therefore, sites with higher resolution sampling provide more information/samples than sites with less data. The uncertainty of the d18O value of the stack at any point in time also includes the effects of relative age (alignment) uncertainty for that point in time in each core record, an advancement compared to the way age uncertainty was considered in the Cenozoic stack. Sites or time intervals with very noisy data have larger alignment uncertainties in BIGMACS.

The regional stacks constructed in BIGMACS are also different from the Cenozoic stacks in Mudelsee et al (2014) because the BIGMACS stacks include orbital-scale (and in some cases millennial-scale) regional climate variability and do not combine data across different oceanographic settings. The type of uncertainty information provided by BIGMACS is appropriate for our goal of characterizing the regional planktonic d18O variability of the WPWP and providing a tool to improve planktonic d18O-based age models for the region.

Although the BIGMACS stacking software provides no procedure for separating internal and external errors in the stack, we will add a column to Table 2 that reports the standard deviation of the residuals between the mean stack and each site (on its median age model and after applying its estimated shift and scale). This will provide readers with additional insight into how similar the planktonic d18O record of each site is to regional mean.

Minor errors

I refer only to Abstract and References since already there appeared quite a number.

The revised manuscript was edited more thoroughly for phrasing and formatting mistakes.

Abstract, l. 1

The expression "different ... than" may sound strange to British ears.

Rephrased.

Abstract, l. 2

Write "greenhouse gas concentrations".

Corrected.

Abstract, l. 3

Insert a hyphen: "orbital-scale climate response".

Corrected.

Abstract, l. 6

Two commas inserted makes it more readable: "... and benthic δ18O stacks, also constructed using BIGMACS, demonstrate that ...".

Updated.

Abstract, l. 7

Insert a bit information: "Sixty-seven radiocarbon dates from the upper parts of four of the WPWP cores ...".

Additional information has been added, as well as updated values with the addition of new radiocarbon data.

Abstract, l. 11

The expression "0 - 450 ka" (with a hyphen) looks ugly. Either use an en-dash without spaces or else write "0 to 450 ka".

Updated throughout text.

References

(1) Do not capitalize (headline style) titles of listed journal articles (e.g. Huybers & Wunsch 2004 “Uncertainty estimates” ... and not “Uncertainty Estimates”).

The reference titles have been corrected to match the style of ESSD in the revised manuscript.

(2) Do properly use superscripts (e.g., Imbrie et al. 1984 “ ... revised chronology of the marine δ18O ...”, and not “δ18O”; Lee et al. 2022 wrong “d18O”, Lisiecki and Raymo 2005 wrong “δ18O”).

Corrected.

(3) Do give editor names for cited chapters from edited books (e.g., Imbrie et al. 1984).

Corrected.

References cited in review

Mudelsee M (2014) Climate Time Series Analysis: Classical Statistical and Bootstrap Methods. Second edition, Springer, Cham, Switzerland, 454 pp.

Mudelsee M, Bickert T, Lear CH, Lohmann G (2014) Cenozoic climate changes: A review based on time series analysis of marine benthic δ18O records. Reviews of Geophysics 52:333—374.

Percival DB, Walden AT (1993) Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge University Press, Cambridge, 583 pp.

Schulz M, Mudelsee M (2002) REDFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. Computers and Geosciences 28:421—426.

Thomson DJ (1982) Spectrum estimation and harmonic analysis. Proceedings of the IEEE 70:1055—1096.