Response letter_2nd revision

We appreciate the remarks and suggestions of the reviewers and are grateful for the effort the they have invested. We have modified the text substantially in response to reviewer comments. Below we respond to each comment individually and indicate the revisions accordingly. For clarity, referee comments are indicated in bold and Authors' comments are indicated in italics. All the changes in the revised manuscript are highlighted with yellow.

Reply to Anonymous Referee #2:

METHODS:

In the manuscript, the authors still do not present arguments why a relationship of living benthic foraminifera abundance with dissolved oxygen concentration can be applied to a dead assemblage. In the response to reviewers comments the authors agree there are substantial differences between the two populations.

It seems like there is a misunderstanding. We cannot compare living (rose Bengal stained) benthic foraminiferal faunas and dead (empty tests, not stained) benthic foraminiferal assemblages from surface sediment samples because the respective census data are not available from any of the Recent distributional studies in the Peruvian OMZ which are considered in this paper. In the revised version, we indicated this under section 2.2 by adding a text. Living assemblages are the best means to correlate with prevailing environmental conditions measured at the time of collection, whereas the dead assemblages are an integral of many generations and may include specimens that were living under different conditions than those measured at the time of collection. In most of the surface samples, however, and in particular in those from the upper OMZ, nearly all benthic foraminiferal specimens were stained. With the given high accumulation rates, it appears to be less likely, that any dead assemblage will contain old, relictic specimens. Therefore, it is justified to assume that 1) living and dead foraminifera assemblages are not that different from each other in our samples, that 2) the dead assemblage accurately mirrors the prevailing environmental conditions at which the living fauna was thriving, and 3) that only well-constrained taphonomical bias inferred an alteration of the dead assemblage during fossilisation (see below).

To eliminate the confusion in the rest of this response letter, we will use the following words for different groups of foraminifera mentioned in this study: "living" for stained specimens and "dead" for specimens that were not stained in surface sediments, whereas "fossil" will refer to benthic foraminifera observed in the long sediment cores (Table 3 of the MS; M77/2-50-4, 52-2 and 59-1).

In our first response letter, we indeed mentioned a comparison between living species and fossil ones observed in the cores. It has to be emphasized, however, that a reduced living benthic foraminiferal dataset (the most abundant 16 species) is not identical with the abundant species (>5%) observed in at least one of the three sediment cores (see Table 1 below). In particular there is one species, Cancris carmenensis, is absent from any samples of concerned time intervals in any sediment core. The remaining 15 species are indeed observed in sediment cores. Seven of these species (indicated by bold in Table 1 below) are common with >5% in relative abundance in at least one sediment core. We hypothesize this substantial difference between living reference dataset and the fossil dataset reflects differences in paleoenvironmental conditions. Considering where sediment cores were obtained and how surface sediment samples distributed (Figure 2 of the MS), such a contrast in species lists is not

surprising. It is also not to wonder that this discrepancy is potentially reflected in the transfer function, estimated values and errors. It is possible that the overall estimated values are somewhat low, because the majority of reference surface samples are from depths shallower than the sediment cores investigated, hence at lower oxygen concentrations. However, without additional samples and further study, we see no way to correct this potential bias. If we knew that a particular species appears only above 45 µmol/kg we could potentially correct estimated by taking that value as a baseline or check point. However, the limited understanding of benthic foraminiferal species relationships with bottom water characteristics does not enable us to make such assumptions. In our opinion, to make such specific assertions about species oxygen tolerances would be wrong and misleading. Therefore, we abstain from using the estimates at face value but rather take the more conservative and realistic approach of using changes through time as calculated for each sediment core and time period. We are aware of the limitations and indicated so in the text. This approach is very new, and we believe that it is a worthwhile methodology to pursue. As with any new approach, precision will improve as additional samples from mid-depths of the Peruvian Margin and more species' information becomes available.

Table 1. The most abundant species (>5%) observed in both surface sediments (reduced living benthic foraminifera dataset) and in downcore records (fossil assemblages observed sediment cores considering the last 25 kyr).

Living dataset 16 species	in cores		"Fossil" >5% in at least one core
Bolivina costata	common		*Alabaminella weddellensis*
Bolivina interjuncta			*Anomalinoides minimus*
Bolivina plicata			***Bolivina costata***
Bolivina seminuda			*Bolivina interjuncta* var. *bicostata*
Bolivina spissa	common		*Bolivina pacifica*
Bolivinita minuta	common		*Bolivina quadrata*
Cancris carmenensis	absent		*Bolivina seminuda var. humilis*
Cassidulina crassa			***Bolivina spissa***
Cassidulina delicata	common		***Bolivinita minuta***
Epistominella obesa	common		*Buccella peruviana*
Epistominella pacifica	common		*Bulimina exilis*
Fursenkoina fusiformis			*Bulimina pagoda*
Gyroidina soldanii			*Cassidulina auka*
Suggrunda porosa			*Cassidulina carinata*
Uvigerina peregrina	common		***Cassidulina delicata***
Valvulineria glabra			*Cassidulina laevigata*
			Cassidulina minuta
			Cibicides mckannai
			Epistominella afueraensis
			Epistominella exigua
			Epistominella obesa
			Epistominella pacifica
			Epistominella smithi
			Fursenkoina cornuta
			Gyroidina rothwelli
			Pseudoparella subperuviana
			Pseudoparella sp.
			Uvigerina auberiana
			Uvigerina bifurcata
			Uvigerina peregrina
			Uvigerina semiornata
			Virgulina spinosa

ACTION:

The authors need to include statements in their manuscript highlighting that the dead and living populations are intrinsically different at the sites, and discuss how this influences the results. For example, could this explain the huge differences in dissolved oxygen in Holocene versus present day oxygen concentrations? Or do the authors think there has been a huge improvement over the last ca 5 kyr (which would be a big thing?) Either way this needs to be thoroughly explained in the main text.

We hope that we have cleared up the misunderstanding on dead vs living foraminifera contrasts. Reasons for relatively low Holocene values vs present values will remain elusive until additional work is undertaken. As indicated in previous comments, the low estimated values of this time frame may or may not be an artefact of the limited data set. Thus, bottom waters might be more oxygenated at the later stage of the early Holocene. Since there are not many quantitative investigations on that time scale allowing a comparison, and the scope of our observations is limited, we cannot further comment on this. Additional samples and study are needed to confirm or modify the paleoceanographic implications of our results. This issue is now detailed in the Discussion of the revised manuscript.

Would it not make more sense to develop a relationship between recent dead populations with dissolved oxygen concentrations to apply down-core? If the authors believe this is not a valid approach, this should be discussed in the manuscript.

We cannot proceed with such an approach because this information is not available. We argue that using dead populations rather than living populations introduces significant artefacts into a transfer function. As noted by a number of workers, wherever possible, it is best to use specimens that were living in the environmental conditions measured at the time of collection. We added text about this under section 2.2. Nevertheless, since the majority of the specimens in surface sediments were stained, it is not likely that there would be a substantial difference in the statistical analysis if the dead assemblage was used.

I am not sure what the difference is between making quantitative statements about specific periods compared with making quantitative statements at one location between specific periods?

We perceive this comment as the reviewer rather quotes for describing the temporal variability at a standard record and to abstain from a differentiated consideration of each locality. The difference of the proxy values among the sediment cores accounts for a spatial variability in oxygenation, as it is observed today and most likely also prevailed in the geological past. Today's spatial variability in oxygenation on the scale of several hundreds of kilometres and hundreds of metres in depth is affected by an interplay between surface ocean productivity and the remineralisation of particulate organic matter at depth, and the advection of more oxygenated waters from the North. Therefore, any variability in hydrographic or chemical properties through time, e.g. oxygen or temperature, has to be regarded in a spatial context. It is a given challenge of paleoceanographic studies in general to approach an understanding or at least the recognition of this dynamics. We have discussed this already in an earlier paper (Erdem et al., 2016), to which reference is given in the revised version of the manuscript.

How were standard deviations calculated?

Below we describe the calculation in detail. This information will be provided in supplementary material.

The polynomial transfer function is given in the form of eq.1:

Eq.1: $$x = C + \sum RCo_n \cdot \%_n$$

where x is the environmental variable that should be reconstructed with the transfer function (in this case $[O_2]_{BW}$ or RRPOC); C is a constant from the multiple regression (tab. 5); RCo_n is the regression coefficient for the foraminiferal species n (tab. 5); and $\%_n$ is the percentage of the foraminiferal species n within the assemblage.

For the calculation of the errors for x (i.e. $[O_2]_{BW}$ or RRPOC) a complete error propagation has been done including the 1σ errors of all species within the polynomial transfer function. The error propagation has been applied to the polynomial transfer function (Eq. 1) in the form of equation Eq. 2:

Eq.2: $$\sigma_x = \sqrt{\left(\frac{\partial x}{\partial C} \cdot \sigma_C\right)^2 + \sum\left(\frac{\partial x}{\partial RCo_n} \cdot \sigma_{RCo_n}\right)^2 + \sum\left(\frac{\partial x}{\partial \%_n} \cdot \sigma_{\%_n}\right)^2}$$

where σ_x is the standard deviation (1sd) of the environmental variable x (i.e. $[O_2]_{BW}$ or RRPOC) σ_C is the error of the constant C (tab.5); σ_{RCo_n} is the standard error for the regression coefficient RCo for the foraminiferal species n (tab.5); and $\sigma_{\%_n}$ is the standard error for $\%_n$. Solution of the derivatives in Eq.2 results in Eq.3:

Eq.3: $$\sigma_x = \sqrt{{\sigma_C}^2 + \sum\left(\%_n \cdot \sigma_{RCo_n}\right)^2 + \sum\left(RCo_n \cdot \sigma_{\%_n}\right)^2}$$

We have to state that we neglect the last term in equations 2 and 3 because we do not know $\sigma_{\%_n}$, since three replicates would be necessary for each sample to determine this error. In this study all downcore samples were counted by a single investigator. Previous studies showed that population densities of sample replicates, which have been picked dry by a single investigator had an accuracy (1σ) of ±2 % (Schönfeld et al., 2013). Thus, in comparison with the high σ_C and σ_{RCo_n} (tab.5) the error $\%_n$ should be negligible. Nevertheless, the proportions of frequent species within the same assemblage may differ by 2–7 % between different picking modes or laboratories (Schönfeld et al., 2013).

AGE MODEL:

The information provided is not satisfactory.

ACTION:

The authors need to visually show how benthic d18O records were correlated to stacked records of Antarctic ice cores (this information is not in the referenced publications). I am especially curious to find out how this was done for the Holocene and how the early and mid Holocene were differentiated for core 52-2. In addition, the authors need to describe how radiocarbon ages and oxygen isotope dates were determined within the hiatuses of 47-2 and 50-4.

For the shallow core, there are only two 'ages' from the late deglaciation and early Holocene. How can you confidently constrain ages between 13 and 22, including the LGM interval?

Following both reviewers' suggestion and in order to improve this issue, we decided to skip the record of the shallow core 47-2 for the time being. Accordingly, all the information on core 47-2 is removed from figures and text of the revised version.

The authors need to show the evidence (d18O graph for each site and correlation to which record) in the supplementary information. Stern and Lisiecki (2014) provide stacked records for different ocean basins and water depths, it is more appropriate to use these for correlation purposes.

We are following the suggestion and we add another figure introducing the age model of concerned sediment cores. The figure is originally published as supplementary material to Erdem et al. (2016), for this publication we modified it and added additional information, in particular benthic stack record from Stern and Lisiecki (2014). We used Pacific Intermediate water benthic isotope stack (supplementary dataset stored in Pangaea (Stern and Lisiecki, 2018)) for correlation with the sediment cores concerned. Indeed, we realized that a better fine tuning was necessary, particularly for the deglaciation section of the core 52-2 (see below graph, Fig. 1). We did not tune the other two core records (50-4 and 59-1) as their age models were seemingly better established. Figure 6 and 7 of the manuscript are modified accordingly considering the new age model.

The reservoir ages at our sampling locations likely varied over the last deglaciation. Unfortunately, data about changes in reservoir ages is scarce. The closest records of changing planktic ^{14}C reservoir ages are located in the equatorial Pacific, close to Panama (Zhao and Keigwin, 2018) and off Chile (Sarnthein et al., 2019; Siani et al., 2013). These ^{14}C reservoir ages differ significantly from each other during some periods and likely cannot be applied to our study sites. Further uncertainty is added to cal. ages based on planktic ^{14}C ages by distinct plateaus in atmospheric ^{14}C ages that can last for up to 1,000 yr (Sarnthein et al., 2015, 2019). Therefore, it is not possible to assign rapid short time changes to distinct short events in our records and we focus more on the long-term trends in our record. The reservoir ages are likely not significantly different between our three study sites, due to their regional proximity. Thus, the relative trends between our sediment records are supposed to be valid. Future studies on changing ^{14}C reservoir ages in the ocean will improve age models of existing paleorecords.

The current age model of core 52-2 is based on five 14C dating, three tie points by correlation to core 59-1 for the Holocene part, three new tie points by correlation to benthic isotope stack for the deglaciation period and three tie points to EPICA ice core for the later part of the record. Overall, sedimentation rates throughout the core is relatively stable (around 30 cm/kyr).

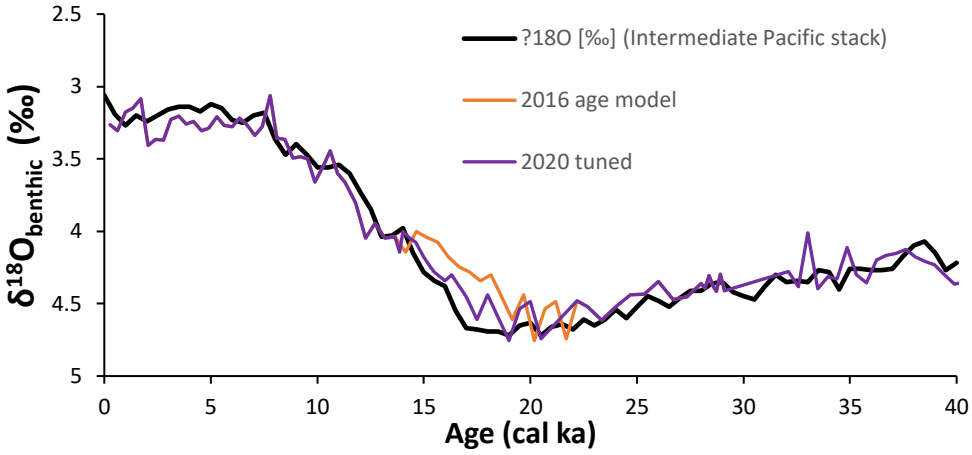

Figure 1. Before (orange) and after (purple) fine tuning of core 52-2 age model by correlating benthic $\delta^{18}O$ from Stern and Lisiecki (2014) with $\delta^{18}O$ of core 52-2.

Other comments and observations that need actioning:

In the abstract it is stated that 'Each core displayed a similar trend of decreasing oxygen levels since the LGM.' This is not true, the shallow core is an exception.

Shallow core (M77/2-47-2) is removed from this manuscript due to reasonable concerns on the validity of the age model (see above). The Abstract is modified accordingly.

The statement about the changes being time transgressive is strange and does not seem corroborated by the data. The shallower cores (at 1013 m and 997 m from 8 degrees South and 4 degrees South) start changing around the same time, whereas the deeper core seems to change later. In addition, for the deeper core the age model is based on two AMS ages around 13 and 22 cal ka, so exact timing of change is not constrained.

With the new tuned age model for the deep core (52-2) this pattern is a bit more visible. However, we now do not have information on the full HS1 period. Considering the comments from both reviewers we modified our interpretation and limit the speculative tone of discussion on time transgressive expansion of the OMZ. Accordingly, we removed a substantial part from section 4.2. Interpretation of the results also changed. All new sentences are highlighted with yellow.

Does the last sentence of the abstract refer to data in the article? If not it needs a reference.

Considering the modifications in the interpretation, discussion part, we removed the last sentence from the abstract.

Page 13 1st sentence: what does this mean, 'a slight recovery of the OMZ'? Did it become less or more oxygenated?

More oxygenated. We rephrased this section accordingly.

Figure 1 and 2 need legends for oxygen concentrations.

Done

What does Figure 4 show? Discuss in detail in the main text.

With the age model figure included we decided to move this figure to supplementary information. A broader explanation is given in supplementary by comparison of CCA application to census dataset and reduced dataset.

What is the purpose of Table 1? Generally oxygen threshold are related to redox reactions, not biota abundance.

The threshold values mentioned in this table have been used in benthic foraminifera related publications. We decided to include this comparison in order make aware of current classifications used in the literature and their differences.

Reply to Anonymous Referee #3:

The authors have detailed how to account for different assemblages downcore than the living ones. But I think the reviewers were also concerned on changes in preservation through time. The depositional setting off Peru is pretty harsh towards the preservation of foraminiferal shells and changes a lot through time, i.e. with the changes in upwelling. So the assemblages may be biased by dissolution of more dissolution-susceptible species.

Additionally to this, I have two further points, firstly Bengal Rose was used to identify living specimens, but especially in low-oxygen settings this may largely over-estimate the population of living specimens;

We used a conservative approach to assess rose Bengal stained specimens, which has been shown to be a reasonable means to evaluate living populations (Murray and Bowser, 2000; Schönfeld, 2012). This staining method is also widely used in benthic foraminifera studies from oxygen minimum settings to determine the living specimens (e.g., (Jannink et al., 1998; Caulle et al., 2014; Koho et al., 2015). Since the work of Figueira et al. (2012), it is now generally accepted.

and secondly, I would like to see some information on how oxygen changes at these core locations through the year, i.e. seasonal changes. Could this be a reason for the scatter or apparent offset between reconstructed values and values at the time of sampling? I.e. the forams may have been alive during sampling but maybe the conditions were not ideal such that they were "hibernating" and actually have their active life phase during another time of the year with different oxygen concentrations.

Bottom water oxygen concentrations around the lower boundary in the region is relatively stable throughout the year whereas the upper boundary is quite dynamic (Paulmier and Ruiz-Pino, 2009). Our living benthic foraminifera compilation dataset was reduced to stations below 300m water depth. Multiple regression used for the downcore application concerns only deeper sampling locations therefore living foraminifera from these deeper than 300m water depth are not expected to experience seasonal $[O_2]_{BW}$ variability as much as stations along the shelf (50-100m) would experience. Nevertheless, we compared the differences between our datasets and between each other for deeper stations since sampling period and seasons showed differences (Perez et al., sampled during December – January 1998; M77 took place during October to December in 2008 and latest expedition M137 was in May 2013). We compared the $[O_2]_{BW}$ from stations close to each other (e.g., around 12°S similar depths) and the only difference was observed at two 820 m stations, the other stations did not show significant contrast between different sampling periods (Table 4 in the MS), the other stations did not show significant contrast between different sampling periods. Therefore, we consider the living assemblage data as being representative and in equilibrium with the prevailing environmental conditions. We added text about this under section 2.2.2.

The age model is indeed a necessary point to be added. I agree that it appears that there is a prograding trend in changing oxygen values but unless the age models are very precise I would tone this down a bit.

We are following the suggestion and we add another figure introducing the age model of concerned sediment cores as mentioned earlier (see above).

The data used in this manuscript should be reported as a dataset in for example Pangaea. Currently there are only references to the different papers where assemblage, core top and water data have been presented.

The datasets are currently stored in Pangaea. Upon publication they will be publicly available.

References:

Caulle, C., Koho, K. A., Mojtahid, M., Reichart, G. J., and Jorissen, F. J.: Live (Rose Bengal stained) foraminiferal faunas from the northern Arabian Sea: faunal succession within and below the OMZ, Biogeosciences, 11, 1155-1175, 10.5194/bg-11-1155-2014, 2014.

Erdem, Z., Schönfeld, J., Glock, N., Dengler, M., Mosch, T., Sommer, S., Elger, J., and Eisenhauer, A.: Peruvian sediments as recorders of an evolving hiatus for the last 22 thousand years, Quaternary Science Reviews, 137, 1-14, 10.1016/j.quascirev.2016.01.029, 2016.

Figueira, B.O., Grenfell, H.R., Hayward, B.W. and Alfaro, A.C.: Comparison of Rose Bengal and CellTracker Green staining for identification of live salt-marsh foraminifera. *The Journal of Foraminiferal Research*, *42*(3), 206-215, doi.org/10.2113/gsjfr.42.3.206, 2012.

Jannink, N. T., Zachariasse, W. J., and van der Zwaan, G. J.: Living (Rose Bengal stained) benthic foraminifera from the Pakistan continental margin (northern Arabian Sea), deep-sea research I, 45, 1483—1513, 1998.

Koho, K., de Nooijer, L., and Reichart, G.: Combining benthic foraminiferal ecology and shell Mn/Ca to deconvolve past bottom water oxygenation and paleoproductivity, Geochimica et Cosmochimica Acta, 165, 294-306, 2015.

Murray, J. W., and Bowser, S. S.: Mortality, protoplasm decay rate, and reliability of staining techniques to recognize 'living'foraminifera: a review, Journal of Foraminiferal Research, 30, 66-70, 2000.

Paulmier, A., and Ruiz-Pino, D.: Oxygen minimum zones (OMZs) in the modern ocean, Progress in Oceanography, 80, 113-128, 2009.

Sarnthein, M., Balmer, S., Grootes, P. M. and Mudelsee, M.: Planktic and Benthic 14C Reservoir Ages for Three Ocean Basins, Calibrated by a Suite of 14C Plateaus in the Glacial-to-Deglacial Suigetsu Atmospheric 14C Record, Radiocarbon, 57(1), 129–151, doi:DOI: 10.2458/azu_rc.57.17916, 2015.

Sarnthein, M., Küssner, K., Grootes, P. M., Ausin, B., Eglinton, T., Muglia, J., Muscheler, R. and Schlolaut, G.: Plateaus and jumps in the atmospheric radiocarbon record – Potential origin and value as global age markers for glacial-to-deglacial paleoceanography, a synthesis, Clim. Past Discuss., 2019, 1–63, doi:10.5194/cp-2019-127, 2019.

Siani, G., Michel, E., De Pol-Holz, R., Devries, T., Lamy, F., Carel, M., Isguder, G., Dewilde, F. and Lourantou, A.: Carbon isotope records reveal precise timing of enhanced Southern Ocean upwelling during the last deglaciation, Nat. Commun., 4(May), 1–9, doi:10.1038/ncomms3758, 2013.

Schönfeld, J.: History and development of methods in Recent benthic foraminiferal studies, Journal of Micropalaeontology, 31, 53-72, 2012.

Stern, J. V., and Lisiecki, L. E.: Termination 1 timing in radiocarbon-dated regional benthic δ18O stacks, Paleoceanography, 29, 1127-1142, 2014.

Stern, J.V; Lisiecki, L., E: Regional benthic $\delta^{18}O$ stacks and their d18O uncertainties. PANGAEA, https://doi.org/10.1594/PANGAEA.891137, 2018.

Zhao, N. and Keigwin, L. D.: An atmospheric chronology for the glacial-deglacial Eastern Equatorial Pacific, Nat. Commun., 9(1), 3077, doi:10.1038/s41467-018-05574-x, 2018.