Response to Referee 2 – Brummer et al., "*Modal shift in North Atlantic seasonality during the last deglaciation*"

We thank the reviewer for their time and for both their general and specific comments. In the following reply, reviewer comments are in RED and our own comments are in BLACK.

My main issue with the study is that the number of analyses, i.e. specimens, per sample is too low to give a representative split up in different populations. Up to 20 specimens were picked per sample, and for quite a few samples less than that were successfully analysed. What is the risk that the split into two populations for these samples is not simply due to highly variable values that only give the impression of separate populations?

20 specimens were picked at random, for every sample. This number of specimens represents the optimum number for down core coverage with the number of specimens per isotope run on a GasBench II set-up, considering time and costs. We disagree with the reviewer that if we measured more specimens the populations would necessary coalesce into a single population, though of course because we do not know the original 'shape' of the (total) population (for instance an approximate sine such as SST when plotted as a histogram has a distribution in which there are more data at the two 'end members' than in the middle) it is difficult to assess this. Although we could perform a theoretical test to see whether this is possible. Whilst, recurrence is not proof, it is intriguing that they two populations do reoccur within the sediment, we do have single isotope data from a deeper depth down core that we can add that represents a different climatological setting. Furthermore, our inference regarding picking for pooled specimens – in which the number of specimens used as the basis of a 'mean' signal is small - would still hold (i.e., section 4.3).

Whether the populations result from 'overfitting' the mixture analysis is of course a concern. The Akaike Information Criterion (AIC; Akaike, 1974. PAST manual: https://folk.uio.no/ohammer/past/pastmanual.pdf pg. 129) is a test of best fit of the mixture model for overfitting (AIC is given in column 3 - table 2) which has a small sample correction.

Page 1 Line 24: are you suggesting the deglaciation lasted for 10 kyr?

We agree the referee that this is oddly worded, we will therefore reword for clarity. It was not our intention to imply that, instead we were referring to the approximate time ('ca.') from maximum ice sheet extent until the 'minimum' extent. We will reword as:

"This represents a shift in the timing of the main plankton bloom from late to early summer in a 'deglacial' intermediate mode that persisted from the glacial maximum until the start of the Holocene."

Line 32: many more references could be cited here to better reflective the literature. These references are all from the same lab.

We will add in more references citing other labs.

Page 2, Line 29: delete the first "and"

We will delete the repetition.

Page 3, 2.3 title: add single specimens to it to distinguish from 2.4 where the bulk analyses are described.

Will be changed.

Line 21: the pachydermas weighed >10 μg?

Whilst the specimens were not weighed –the amplitude on mass 44 correlates with the weight of a specimen, allowing us to make an informed guess of the amount of carbonate per analysis. N. pachyderma is an encrusted form and we took 250-300 µm sized specimens these allow for sufficient gas for a signal to be generated.

line 24: how many specimens/what weight were used?

Approximately 1 mg of foraminifera were used – unfortunately we did not count the number of specimens.

Line 37: "varoes"

Will change to 'varies'.

Page 4, line 20: missing year in Jonkers and Kucera

We will add the year.

Line 32: I assume these are the pooled d18O?

We will clarify: "The upper ~290 cm of core T88-3P is Holocene in age as evidenced by near uniform values of pooled specimen $\delta^{18}O$ values"

Line 36: "during IRD events"

We will alter 'at IRD events' to 'during IRD events'.

Page 5, line 4: The striking bimodality is quite difficult to see, it could simply be more variation in the analyses. Why not plot the results also as histograms? And similar for the d13C results; it is not easy to see now how the variations are.

We will consider making histograms, although for the core sections 340 – 380 cm (covering the section where more than on population exists), this will result in 16 histograms for a single analysis (d18O) and for a single species, therefore it would 64 histograms in total. Unless the referee agrees that a single histogram 'lumping' the data together, in combination with figure 4, is suitable.

Additionally, why is the x-axis labelled in x time 10 4 years? This is confusing, just stick to the regular ka.

We apologise for having overlooked this error (the plotting programme added x10^4). We will remove and change kyr to ka.

Page 6, line 7: Is 250-300 μm correct?

Yes, it appears odd to use a smaller than standard (i.e., 300-355 µm) size fraction, but *N. pachyderma* is a small species. Because it is generally a small species there is the concern that by selecting too large (therefore greater mass) or too small sized specimens we could, if size is some indicator of ecology, bias the results. Despite this reduced size it is a 'thick/heavy' species given its compact form and heavy calcification (regardless of encrustment) which produces enough weight for single $\delta^{18}O$ analysis.

Line 8: were any of the sediment-trap pachydermas genetically determined?

Unfortunately, this is not possible for material sinking into deep-moored sediment traps. Generally, foraminiferal shells settling into a trap at 2.5 km water depth are free of original cellular matter for genetic analysis or found infested by bacterial and ciliates consuming any remains. Given that, the trap samples are ashed to isolate the mineral skeletons from the organic matter and leave a clean residue for isotope and chemical analysis. Whilst there has been some suggestion that variance in stable isotope value may relate to genetic factors, only recently (to our knowledge) has a protocol been develop for combined genetic and stable isotopes of small samples:

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213282

Line 35: pachyderma is also unlikely to have lived in this meltwater; they normally stick below this relatively fresh layer.

We agree – and will refer and expand upon this in a revised MS. Here, we are referring to the spike in isotope records that occur in the literature – the so-called 'meltwater spike' in a number of papers including Berger et al. (1977; https://www.nature.com/articles/269301a0), Jones & Ruddiman (1982; https://doi.org/10.1016/0033-5894(82)90056-4) and so forth. We will alter the text accordingly:

"The presence of continental ice-rafted debris (IRD) down core in T88-3P, without a clear concomitant 'spike' in the $\delta^{18}O$, referred to in the literature as a 'meltwater spike' (Berger et al., 1977; Jones and Ruddiman, 1982) of either *N. pachyderma* or *G. bulloides* (Fig. 2) would suggest that the difference in $\delta^{18}O$ between the two populations is dominated by temperature, consistent with previous studies showing no meltwater spike (Duplessy et al., 1996; Straub et al., 2013)."

Page 7, line 31: delete "."

We will delete this.

Page 8, line 7: the Bard, 2001 reference is missing from the References

We will add it accordingly to the reference list.

Section 4.3: the results here show that in a setting like the North Atlantic the pooled specimen analyses may be biased when not enough specimens are being used. Could you provide an estimate how many specimens would be needed to give a reliable estimate?

We thank the reviewer for bringing up this is an important point. Our results are merely showing the potential error or spread between increasing in-group numbers of pooled specimens (figure 6). The aim of pooled analysis is to average out specimen to specimen variability and produce a mean value for the core interval (time-interval) sampled that can be used as a climatological signal. Until Shackleton (1965) the amount of carbonate required for a single measurement was 4.5 mg, reducing to 1 mg and subsequently the amount required has steadily decreased as the technology has evolved. Pooled specimens therefore have steadily decreased from 100's to 10's, or less. It is our opinion that it would not be correct for us to state an exact number for a reliable estimate, as this undoubtedly will change depending on the sedimentation rate, the core, the time interval, the location, the weight limitations of the mass spectrometer (upper or lower), etc. However, if one considers the question of "when not enough specimens are being used" in fact it is not so much the total number of specimens but the proportion between the populations, if there is a single population then fewer specimens may be enough (although one would still need to account for the variance within that population).

In addition, one could argue that replicates rather than group number may be better at reducing associated biases (e.g., keeping the number in group constant and performing several replicates). What we do think, however is that pooled specimens should be considered in light of this 'hidden' variance. Therefore, in a revised MS we will expand upon this section through calculation of the how much the difference in proxy information (e.g. temperature or salinity estimates) may be.

Figure 2b: Is this 14C age of 41900 years used for the age model or not? It seems not, so then it should be deleted from the figure or indicated as such.

Whilst the date is not used for the age model because of the calibration curve's assumptions around this age, we disagree that it should be left out as it (i) has been measured and (ii) gives a general indication of the relative age of this sediment. That being said, we will alter the colour of the text to red / italic to indicate a date we did not use – but we do not find 'error' with.

Figure 5: Add headings of the different areas on top of each "column".

Thank you for this suggestion. We will add both the name, area and latitude of each trap for each 'column'.