Dear editor,

Please find a revised version of our manuscript “Variability in Neogloboquadrina pachyderma stable isotope ratios from isothermal conditions: implications for individual foraminifera analysis”. We have made the changes we proposed in our reply to the reviewers and also added more information from our rebuttal to the manuscript as you requested.

Specifically, we elaborated on the offsets from equilibrium that we observe in spring (lines 294-298 and point 3 raised by reviewer 2) and added some information about the degree of encrustation of the shells to the discussion in section 3.4.

Below we have copied the reviewers comments and provided a detailed response to all concerns. References to line numbers refer to the version with tracked changes, which is appended at the end of our response.

We hope that our revised manuscript is now suitable for publication in Climate of the Past,

Lukas Jonkers
On behalf of all authors.

We would like to thank the reviewers for their careful reading of our manuscript and their constructive comments. Below we have copied the review in full and provide our response in orange text.

Text quoted from the original manuscript is in grey and our changes based on the review are in blue.

We feel that thanks to these suggestions the manuscript has improved considerably and hope that our proposed revision now meets the criteria for publication in Climate of the Past.

Lukas Jonkers
On behalf of all authors.

REVIEWER 1

The manuscript by Jonkers and colleagues compares multiple samples of the stable isotopes from the shells of the planktic foraminifer N. pachyderma from the same sediment trap samples. They then use a combination of nearby hydrographic records, modeling, and statistical analyses to assess the variability within a population not attributable to environmental factors, primarily temperature. They find a substantial amount of variability in multiple samples from the same cups, which is used to illustrate the inherent “excess” variability of reconstructions using very few shells. With increasing use of high resolution instrumentation making use of small samples and individual foraminifera analysis (IFA) more frequent, the implications of these findings are important.

I have a few suggestions which I hope the authors will find useful. My primary suggestion for the manuscript is to do with framing. From line 1 of the abstract, the rationale of the study is laid out to be an estimate of excess variability in individual shells measurements and therefore utility of IFA. The catch is that the methodology used here is not IFA but rather multiple pooled samples. Several assumptions are required to make the leap from environmental data and pooled measurements to an estimate of excess variability by a theoretical IFA measurement, some of which require additional justification. My comments include a few specific suggestions of where this may be helpful. However, it is also my opinion that the framing of this manuscript as a quantification of IFA excess variability may be slight overreach drawing from this particular dataset. There are certainly implications for IFA, and the rough calculation done here are useful in illustrating that. However, given the number of assumptions required and the use of pooled rather than individual shells in the analyses, overemphasis on a quantification of “noise” in IFA analyses specifically, may do a disservice to the really important findings of large excess variability.

We thank the reviewer for their constructive comments. We agree with the reviewer that our quantification of the excess variability requires more discussion and added the following paragraph in section 3.3: “Whereas our modelling approach provides an estimate that is likely closer to reality than assuming that foraminifera reflect environmental conditions averaged over a single (calendar) month, our estimate could be evaluated by simulating other calcification trajectories. We found that our results are insensitive to the duration of chamber formation and experiments where we allowed complete shell formation within one day, equivalent to assigning all weight to the last chamber, yielded an expected 0.09 ‰ standard deviation of individual foraminifera δ18O. Therefore, the assumption of equal weight of the four chambers has little bearing on our results. Ultimately, the modelled foraminifera δ18O depends on the hydrographic data used to estimate δ18Oequilibrium. By using data from the surface and from great depth, we have obtained two end-member scenarios of vertical δ18Oequilibrium variability that implicitly encompass ontogenetic vertical migration. However, future estimates of expected individual foraminifera δ18O variability could be improved by explicitly incorporating horizontal δ18Oequilibrium variability and advection during shell growth in the modelling strategy.

Apart from being sensitive to our modelling design and data availability, our estimate of excess δ18O variability among individual shells is also sensitive to the quantification of variability among shells. To obtain a conservative estimate we excluded potential outliers. Were we to consider all measurements, the average standard deviation among groups would be 0.15 ± 0.11 ‰ (0.17 ± 0.09 ‰ during spring) and the resulting excess δ18O

variability 0.25 ± 0.19 ‰. Thus our approach yields a conservative and better constrained estimate of the excess variability."

We also made sure to be more careful with our wording regarding the estimate of the excess noise in the abstract and in the conclusions. However, we think that our phrasing in the main text (e.g. "Assuming that our simulations are a reasonable approximation of reality, the excess variability (s.d.) that cannot be explained by variability in temperature and δ18Oseawater is therefore 0.11 ± 0.06 ‰.") is not overselling the results and we would prefer to keep the original text here.

Minor/specific points:

111: Why were outliers removed? Points that deviate farther from the mean would seem particularly valuable for this dataset, unless there is specific justification for their removal. Perhaps there is a reason for this data treatment that just needs to be better explained?
This point was also raised by reviewer 2. The only reason to apply this filtering was to ensure that our analysis is insensitive to potential outliers, without making statements about the reliability of the removed data points. One could therefore view the variability in N. pachyderma stable isotope ratios that we use as a minimum, rendering our estimate of the magnitude of the excess variability conservative. We made this reasoning clearer, both in the method section and in the discussion (see e.g. our suggested change above).

146: The assumption of chamber formation over one day in pachyderma is a bit misleading. While initial chamber formation may occur over one day (as in the referenced studies), calcification is likely more prolonged in this species. A better model than the spinose foraminifera observed in the Spindler and Be papers, might be congener N. dutertrei, where laboratory labelling experiments affirm that much of the calcite is added over a period of several days and nights as evidenced by banding and the apparently continuous uptakes of 'spikes' added in culture (see Fehrenbacher et al., 2017).
We agree with the reviewer that our modelled chamber formation is an absolute minimum. It is, however, in agreement with the data from Spindler (1996) on *N. pachyderma*. We nevertheless checked what the effect is of longer chamber formation and reran our simulations with a four day duration of chamber formation as suggested for *N. dutertrei* (Fehrenbacher et al., 2017). The effect is negligible because of the high temporal autocorrelation in the d18Oequilibrium time series that renders the effect of smoothing insignificant. The expected standard deviation of foraminifera d18O based on our model is in both cases 0.08 permille. (Note that in our original submission we modelled chamber formation within at most one day and that yielded an expected standard deviation of 0.09 permille.) To clarify we added the following text to section 2.3: "The assumed duration of chamber formation is based on culture studies (Bé et al., 1979; Spindler, 1996). However, culture studies in the closely related species N. dutertrei have shown that chamber formation may take up to four days (Fehrenbacher et al. 2017). Longer chamber formation could in theory reduce the variability foraminifera δ18O because of increased smoothing of the environmental signal. In practice this effect is however negligible because of strong temporal autocorrelation in the δ18Oequilibrium time series that renders the effect of smoothing of up to four days insignificant. Our approach thus yields an estimate of variability that is robust against the likely range of chamber formation duration."

281: I am struggling with this calculation, on which so much of the interpretation relevant to IFA rests. While this estimation accounts for the N term, it makes two assumptions. The first is that the sample mean would have been the same if IFA had been carried out rather that multiple pooled analyses – this is probably a reasonable assumption, if one has on minimal instrumental error and near identical calcite contribution from all shells. However, the other assumption is that the stable isotope value of an individual shell would be the same as the value of the pooled analyses. This is a less robust assumption, belied by even the conclusions of this paper. Individual shells would be expected to represent a greater range of values, and therefore overall greater deviation from the sample mean. I think the argument for calculating excess of theoretical IFA as such could benefit from a statement of these underlying assumptions.

The obvious rebuttal to the caveat(s) raised above is that these are necessary assumptions given the sample set and/or that once again the estimate of unexplained variance is highly conservative. This might be the case, but if so perhaps there is too much emphasis on the quantification of this speculative 0.19 per mill (and therefore 0.11 per mil) number as a noise threshold.

We appreciate the concerns by the reviewer and will better explain the way we performed the calculation. The reviewer is right about the first assumption that we assume an identical contribution to the total calcite mass for each shell (and hence identical mean values). We will state this more clearly. However, we do not make the second assumption. Instead, we explicitly derive the standard deviation among individual shells from the standard deviation of the pooled measurements, the former is - as the reviewer correctly notes - indeed larger (double in our case) than the latter. To clarify these issues we changed the sentence: "Since our measurements are based on groups of four shells, the standard deviation of individual shells is double (√4) the observed standard deviation." to: "Since our measurements are based on groups of four shells the observed standard deviation is an underestimate of the standard deviation among individual shells. Assuming that each shell in the group contributed equally to the total mass, the degree of underestimation of the standard deviation scales with the square root of the group size (Groeneveld et al. 2019). Thus we multiply the observed standard deviation by two (sqrt(4)) to obtain an estimate of the standard deviation of individual shells." (see lines 356-360)

333-335: My reading of Livsey et al. (2020) is that lamellar and crust calcite were indistinguishable in d18O space

Good point, we accidentally mixed up Mg/Ca and d18O. This makes the likelihood that variable encrustation could explain the observed variability even smaller. We deleted the sentence and added: "However, the difference between crust and lamellar calcite δ18O of N. pachyderma intercepted in spring when the water column was well-mixed is not significant (Livsey et al. 2020). Variable encrustation can therefore not be the explanation for the excess δ18O variability observed during the isothermal conditions in spring.", see lines 449-453.

Other minor points: I was curious about the lack of shell measurements here, as stable isotope values are well known to correlate with size, something that the authors discuss. I understand that this is a reanalysis and such measurements may no longer be available, but it is a point potentially worth addressing.

The reviewer rightly points out that size of individual shells would be an interesting parameter to have at our disposal. However, as the reviewer also correctly infers such measurements are unfortunately not available. We would like to highlight though that we have analysed larger scale pattern in shell size and its influence on sedimentary stable isotope ratios in a previous paper (Jonkers et al., 2013).

References:

Fehrenbacher, J. S., Russell, A. D., Davis, C. V., Gagnon, A. C., Spero, H. J., Cliff, J. B., ... & Martin, P. (2017). Link between light-triggered Mg-banding and chamber formation in the planktic foraminifera Neogloboquadrina dutertrei. Nature communications, 8(1), 1-10.

Livsey, C. M., Kozdon, R., Bauch, D., Brummer, G. J. A., Jonkers, L., Orland, I., ... & Spero, H. J. (2020). Highâ resolution Mg/Ca and δ18O patterns in modern Neogloboquadrina pachyderma from the Fram Strait and Irminger Sea. Paleoceanography and Paleoclimatology, 35(9), e2020PA003969.

References

Fehrenbacher, J. S., Russell, A. D., Davis, C. V., Gagnon, A. C., Spero, H. J., Cliff, J. B., Zhu, Z., and Martin, P.: Link between light-triggered Mg-banding and chamber formation in the planktic foraminifera *Neogloboquadrina dutertrei*, Nature communications, 8, 15441, 2017.

Jonkers, L., van Heuven, S., Zahn, R., and Peeters, F. J. C.: Seasonal patterns of shell flux, $\delta^{18}O$ and $\delta^{13}C$ of small and large *N. pachyderma* (s) and *G. bulloides* in the subpolar North Atlantic, Paleoceanography, 28, 164–174, 2013.

Spindler, M.: On the salinity tolerance of the planktonic foraminifer *Neogloboquadrina pachyderma* from Antarctic sea ice, Proceedings of the NIPR Symposium on Polar Biology, 9, 85–91, 1996.

REVIEWER 2

The manuscript submitted by Jonkers et al. describes a study (based on existing data from earlier publications) that aims to assess whether planktic foraminifera of the genus Neogloboquadrina pachyderma accurately record environmental parameters (here: temperatures deduced from d18O, and d13C). Shells of N. pachyderma were derived from a sediment trap, moored in the Irminger Sea. The trap collected sinking plankton during multiple years, and the collection intervals were roughly 2.5 weeks. For analysis, Jonkers et al. pooled four N. pachyderma shells from each sample vial, and multiple groups of four shells were analyzed for each collection interval. A within-sample variability of 0.11‰ for d18O and of 0.10‰ for d13C was found, independent of the season or month of sampling. Furthermore, the variability in d18O and d13C exceeds water column variability in spring when the water column is isothermal.

In order to assess potential sources for this variability, the authors run simulations (main parameters are the potential timespan of chamber formation, calcification depth, and delay due to settling), and conclude that the observed variability in d18O can only partially be explained by environmental variability. The authors estimate an “excess noise” on d18O of about 0.11‰ (biological or other yet unknown origin), which, as the authors postulate, needs to be taken into account when interpreting geochemical variability among individual foraminifera.

This is an interesting study/manuscript that is certainly an important contribution, however, there are certain issues that the authors should address:

(1) Jonkers et al. is linking this study to Individual Foraminifera Analysis (IFA), which is increasingly common with the rapid development of new or improved analytical approaches. However, IFA are, senso stricto, measurements of single, individual foraminifera shells. However, the authors were analyzing groups of four shells. I am not sure to what extent the findings of Jonkers et al. can be interpolated to ‘true’ single-shell IFA, but I would prefer to remove all references to IFA or soften the wording. However, Jonkers et al. raise an important question: We need to decide between the “reliability” of individual planktic foraminifera shells as a proxy recorder, and the potential attenuation of high-frequency or short-lived climate signals due to the measurement of populations that are too large to record these short-term signals. Instead of referring to IFA, I recommend to include a short discussion about sample sizes for paleoclimate records (built upon Schiffelbein and Hills, 1984, and subsequent studies). There is no simple answer – but the new data presented by Jonkers et al. provide the opportunity to discuss this topic from a new/different perspective.
The reviewer raises some important points. We understand the doubts by the reviewer, but do not agree that our analyses have no implications for the interpretation of IFA results. The replicate measurements on groups of four shells of course cannot directly provide information about the stable isotope variability among individual shells. However, what these replicate measurements can provide is an estimate of variability within the population of planktonic foraminifera that cannot be obtained from a single measurement (whether on a sample containing many foraminifera, or single shells). This estimate of the variability within the population, even though an underestimation of the variability among individuals, is valuable knowledge for the interpretation of IFA results. We are therefore convinced that the framing of our study along the lines of “implications for IFA” is justified. We made our reasoning clearer in the revised version and added the following sentence to section 2.2:

"Even though the measurements were done on groups of four shells, the replicate measurements on small numbers of shells allow us to obtain a first order estimate of the minimum stable isotope variability within the population of N. pachyderma. Our analyses are therefore meaningful for the interpretation of IFA results."

We would also like to stress that all our modelling exercises are consistent with the measurements on groups of shells, not on individual shells. It is only in section 3.3 that we provide an estimate of the inter-individual variability. This estimate is based on the mathematical relationship between the standard deviation among groups (of shells) and the standard deviation among individuals (shells) that make up those groups. Assuming that each shell contributes equally to the total mass of the group, the standard deviation in the d18O of individual shells scales with the square root of the group size, in our case sqrt(4) = 2. We realise that this calculation was not described clearly enough and changed the wording in section 3.3. Specifically, we changed: "Since our measurements are based on groups of four shells, the standard deviation of individual shells is double (√4) the observed standard deviation." to: "Since our measurements are based on groups of four shells the observed standard deviation is an underestimate of the standard deviation among individual shells. Assuming that each shell in the group contributed equally to the total mass, the degree of underestimation of the standard deviation scales with the square root of the group size. Thus we multiply the observed standard deviation by two (sqrt(4)) to obtain an estimate of the standard deviation of individual shells."
The argument above also explains why our estimate of IFA is more robust than what could be obtained from analyses that pooled more specimens. For example, variability obtained from replicates of pooled analyses of 25 shells scales to the IFA variability by a factor of 5. This means that to constrain the IFA variability as well as in analyses of 4 shells, one would need 2.5 x as many replicates.

The reviewer suggests that instead of focussing on the implications for IFA, we should consider discussing the number of foraminifera that should be analysed for robust palaeoceanographic reconstructions. We agree that this is an important topic. However, the variability among sedimentary foraminifera stable isotope ratios depends on many more factors than we can assess from our time series. For instance, it depends on the seasonal amplitude of temperature (and d18Osw) variation, the seasonality in the flux of foraminifera, the sedimentation rate as well as the bioturbation depth. The effect of these factors has been discussed extensively in the literature (e.g. Dolman and Laepple, 2018; Jonkers and Kučera, 2017; Lougheed and Metcalfe, 2021). In our opinion such a discussion would go beyond our original question about the reliability of single planktonic foraminifera shells as recorders of environmental conditions.

(2) On purpose, the authors excluded the possibility of horizontal drifting -which is okay. Including horizontal drifting will add new layers of complexity and uncertainties, and potentially raise a whole new set of open questions and challenges. Still, horizontal drifting should be discussed as a potential source of the large measured d18O variability in N. pachyderma shells that exceeds the annual range in "d18O equilibrium" values at the location of the sediment trap. In quickly checking the velocities within the Irminger Gyre (e.g., Våge et al., 2011), the shells can be transported to the sediment trap over significant distances and "import" proxy-signals from a very different location. Basically, the authors exclude horizontal drifting, run the model, and postulate that the measured d18O (and d13C)

data in the shells cannot be reproduced with local temperature and d18Osewater data, independent of the selected calcification depth. Thus, the authors ascribe the 'excess' variability in foraminifera d18O and d13C to biological (and/or other) factors. Latest at this point, horizontal drifting should be again included into the discussion (although it was not included in the model, which is okay).

The reviewer raises a fair point that indeed deserves more discussion. The possibility of advection is real, even though the influence on the stable isotope variability is not directly scalable with the advection distance because it also depends on the spatial temperature and d18Osw variability (simply said, advection only has influence on the foraminifera proxy signal in regions where there is temperature variability). We tried to allude to this in the last paragraph of section 2.3, but we now also include a more extensive discussion about our estimate of excess variability. We added: "Whereas our modelling approach provides an estimate that is likely closer to reality than assuming that foraminifera reflect environmental conditions averaged over a single (calendar) month, our estimate could be evaluated by simulating other calcification trajectories. We found that our results are insensitive to the duration of chamber formation and experiments where we allowed complete shell formation within one day, equivalent to assigning all weight to the last chamber, yielded an expected 0.09 ‰ standard deviation of individual foraminifera δ18O. Therefore, the assumption of equal weight of the four chambers has little bearing on our results. Ultimately, the modelled foraminifera δ18O depends on the hydrographic data used to estimate δ18Oequilibrium. By using data from the surface and from great depth, we have obtained two end-member scenarios of vertical δ18Oequilibrium variability that implicitly encompass ontogenetic vertical migration. However, future estimates of expected individual foraminifera δ18O variability could be improved by explicitly incorporating horizontal δ18Oequilibrium variability and advection during shell growth in the modelling strategy.

Apart from being sensitive to our modelling design and data availability, our estimate of excess δ18O variability among individual shells is also sensitive to the quantification of variability among shells. To obtain a conservative estimate we excluded potential outliers. Were we to consider all measurements, the average standard deviation among groups would be 0.15 ± 0.11 ‰ (0.17 ± 0.09 ‰ during spring) and the resulting excess δ18O variability 0.25 ± 0.19 ‰. Thus our approach yields a conservative and better constrained estimate of the excess variability." to section 3.3.

(3) A puzzling observation is the fact that some group of four shells feature significantly higher d18O values than we would expect at sample location, even when we assume calcification during the coldest season and at a large water depth (see Fig. 2). This is an interesting finding and should be discussed. Low d18O values in N. pachyderma are often observed, and some previous studies (e.g., Bauch 1997, Ravelo and Hillaire-Marcel (2007), Simstich et al., (2003)…) postulated that either vital effects or the effect of low-d18O meltwater lenses cause low d18O values in N. pachyderma shells. However, reports of N. pachyderma shells that are "too heavy" in their d18O composition are rare. Were the shells transported from colder waters to the location of the sediment trap? This should be further discussed. In particular, it needs to be emphasized that each data point integrated the composition of four shells. Thus, the spread of individual shells in d18O (and d13C) is likely larger, and single shells may feature even higher d18O values than the group of four. If it is not possible to reconstruct these high d18O from the water column profile – what is the explanation, if we exclude horizontal drifting?

The reviewer touches on an interesting point. Apart from d18O values lower than equilibrium, which we attributed to remnants of the summer population that survived without calcifying (Jonkers et al., 2010) there are indeed also samples with a d18O higher than d18Oeq. We agree that these data points are puzzling.
The reviewer suggests that advection from colder waters could be an explanation. We agree that this could be the case. However, advection from colder water likely means advection from the East Greenland Current, which is also considerably fresher and hence has lower d18Oseawater (around -2.5 permille VSMOW). The d18Oseawater effect would therefore overwhelm the temperature effect and advection from the East Greenland Current is therefore likely to lead to lower foraminifera d18O.

Importantly, we note that during the time when d18Ocalcite higher than d18equilibrium is observed, some samples also show lower than equilibrium values (Fig. 2B). This large spread in the d18Ocalcite is entirely consistent with the hypothesis that foraminifera d18O contains additional, temperature and d18O seawater-independent noise and are therefore individually not reliable as environmental indicators. We added the following to make our line of reasoning clearer: "This variability arises from apparently random positive and negative offsets from δ18Oequilibrium, suggesting that it does not result from a mechanism that would cause a systematic bias in the foraminifera δ18O. Advection or long foraminifera life spans, which could theoretically cause foraminifera from the previous summer to survive until spring, are therefore unlikely to provide a full explanation for the observed variability." (lines 294-298).

(4) For this study, defining criteria for outliers is very important and critical. The authors defined outliers as being more than 1.5 times the interquartile range away from the overall mean. Was this selection arbitrary? Do we know whether the "outliers" provide a true signal? Four shells are measured together, thus, one or two shells within this group of four must feature very different d13C or d18O values to shift the averaged composition of four shells sufficient to trigger the 'outlier' criterion. Jonkers et al. removed 6% of the d18O data. This is a high number. In other words: It seems the authors believe that 6% of all d18O measurements conducted within the framework of this study are not trustworthy. This needs to be discussed in more detail. The sample material was clean and well preserved (sediment trap, no issues with clay contamination or diagenesis), and standard procedures/equipment was used for sample preparation and analysis. We have many decades of experience with this analytical approach. Thus, in theory, the quality of the data should be as good as it can get. But 6% removed???
We agree with the reviewer that this is a point that requires further explanation. Our rationale to apply a strict outlier criterion was to avoid discussion about the influence of potential outlier effects on our interpretation. We did not want to imply that measurements identified as outliers are unreliable. Please also note that this approach is fairly common in IFA studies (Ganssen et al., 2011; Groeneveld et al., 2019). To make our rationale clearer we added the following sentence to section 2.2 where the outlier removal is described: "In order to obtain a conservative estimate of the variability among the measured groups of N.pachyderma shells we remove possible outliers."

Importantly, even when reducing the variability by removing samples outside 1.5 times the interquartile range, the remaining variability in stable isotope ratios is still larger than what

could be expected. Our conclusions on the excess variability are therefore conservative. We will elaborate more on this in the discussion on the quantification of the excess variability (see the proposed text under point 2 above).

(5) General comment regarding the figures: Many labels in the figures are too small. It is okay when reading the publication as PDF (which most of us will do), but much information will be lost when the figures are printed. In addition, the manuscript would greatly benefit from some careful 'wordsmithing'.
We increased the font size in the figure labels and do our best to improve the writing. We would like to thank the reviewer again for their careful reading and the many helpful suggestions to improve the text.

Some minor suggestions:

Line 54: The sentence seems to be incomplete. Suggestion for completion: "…and only few consider calibration issues associated with individual planktic foraminifera (Glaubke et al., 2021) as a source of uncertainty".
Done.

Line 56, 57: "geochemistry is too generic". Temperature exerts a first order control on Mg/Ca and d18O (when d18Osw is accounted for). There are several other foraminifera-based proxies that are not primarily controlled by temperature.
Good point; we replaced "geochemistry" with "Mg/Ca and d18O".

Line 67: consider rewording: a proxy is only approximating a parameter of interest. It is not a "precise" environmental indicator. Precise implies precision. Better choices are 'robust', or 'reliable'.
Wel changed to "reliable".

Lines 78-83- the last paragraph of the introduction describes results or conclusions (…"We observe marked variability… … and find that the observed variability….. … we argue that this biological…"). I leave it up to the authors, however, I strongly recommend keeping the introduction descriptive, without mentioning the results or even some interpretation
We would prefer to keep this as it is as we think it makes for more interesting reading if the editor agrees.

Line 101: Can the authors provide more detail? 45 samples (= collection intervals) were analyzed, most of them were measured at least twice. However, it follows from Section 2.1 that the sediment trap provided much more than 45 samples (or collection intervals). It would be nice if the authors could provide more information about the criteria for sample selection.
To make it clearer that not all samples (collection intervals) contained foraminifera we added the following sentence: "Not all samples from the time series contained enough shells of N. pachyderma (Fig. 1), so the complete data set consists of 172 measurements from 45 samples, of which 163 are from 36 samples with at least two measurements." See line 113-114

Line 106: I am a bit confused. I thought IFA stands for "Individual Foraminifera Analysis", which means individual shells. However, according to Section 2.2, groups of four N.

pachyderma shells were analyzed. Thus, the number of shells is high compared to IFA, not low, as stated by the authors. I am not even sure if groups of four shells can or should be considered as IFA.
We are sorry that the reviewer got the impression that we performed individual foraminifera analyses. We never intended to claim that we did, and made it clearer that our measurements were done on small groups of shells. To thi send, we added the following sentence "Even though the measurements were done on groups of four shells, the replicate measurements allow us to obtain a first order estimate of the minimum stable isotope variability within the population of N. pachyderma. Our analyses are therefore meaningful for the interpretation of IFA results." to section 2.2. We understand the confusion about the number of shells per sample, the crucial difference is in the number of replicates: for IFA usually in the order of 50-70, whereas we have used up to six replicates per sample. To avoid confusion we will replace "The number of shells measured per sample …" with "The number of replicate measurements per sample…" see line 120

This comment echoes the first comment by the reviewer and we would like to emphasise that we do not consider our measurements equivalent to single shell analysis, but that the conclusions we derive from our data are still important for the interpretation of IFA data, precisely because the observed variability among groups of four shells represents a minimum estimate of the variability among individual shells. In the revised section 3.3 we also elaborate further on how we derive an estimate of the variability among individual shells from the measurements on groups of shells.

Line 107: "weeks to month" – does this refer to the collection intervals, or the combination of collection interval + life span of the foraminifera (in particular the time when they grew their shells)? I think this should be mentioned for clarity.
To make this issue clearer we changed "This is however justified given the short integration time of sediment trap samples (in our case 16-19 days) compared to sediment samples (at least decades to centuries)." to "This is however justified given the short collection intervals of sediment trap samples (weeks to months) compared to the long integration time of sediment samples (at least decades to centuries)." (lines 120-122).

Line 120: Yes, but there are also studies postulating that N. pachyderma features a (negative) vital effect in d18O (Bauch, Simstich, Hillaire-Marcel, and many others). Although I am okay how this is written, adding a short discussion – emphasizing why the authors believe that N. pachyderma calcifies without a vital effect for d18O – would be helpful
The reviewer touches on an interesting point. The offset, or lack thereof, from d18Oequilibrium is important to constrain when interpreting the d18O of foraminifera. However, for the present study, the issue is only of limited relevance. This is because in the temperature range investigated here, the slopes of the different palaeotemperature equations are nearly identical (Jonkers et al., 2013) and the effect of using a different equation on the d18O variability is thus negligible. For this reason we deem a discussion about a possible offset from d18Oequilibrium more a distraction than an addition in this manuscript. The reader is referred to two publications where this issue is elaborated. A potential reason for the difference with other studies (including those mentioned by the reviewer) is the use of different palaeotemperature equations.

Line 121: It shall read "Jonkers et al., 2010, 2013". Same in line 121

OK.

Lines 122-125: please reword the sentence – overuse of ‘because’ (we use these because…and because)
Done.

Line 127: I think it shall read “regressions” (plural)
We disagree, multiple (linear) regression is regression with multiple predictor variables, not multiple regressions with single predictor variables.

Line 130: what does “available as climatology” mean? Same line: Use “spatial resolution” instead of “same level of detail”?
Climatology data means the long-term average as for instance provided in the world ocean atlas. The lack of detail is hence not only in space, but more importantly in time (see fig. 2). To clarify, we replaced: “δ13CDIC data are available as climatology only and can hence not be used to the same level of detail as δ18O” with “Since the δ13CDIC data are derived from data that represent long-term average conditions (climatology), they cannot be used to the same level of detail as δ18O” See line 149.

Lines 130, 131: measured variability in foraminiferal d13C (to make it clearer)?
Done.

Line 135: It sounds as if the formation of the entire shell takes place in the same water depth. Most planktic foraminifera (also N. pachyderma) migrate to deeper waters as part of their ontogenetic development. Earlier chambers are typically formed in shallower waters than the later chambers (and crust, if present). This should be mentioned here.
The reviewer is right about the possibility of ontogenetic vertical migration in N. pachyderma. However, we want to discuss - and model - the temporal and spatial integration aspects separately. We want to first highlight the temporal aspects of calcification (this sentence) and we discuss variability in calcification depth, including vertical migration in lines 170 to 180.

Line 142: What does “survival’ in the water column (without calcification)” mean? The last chamber is formed, the organism is not further calcifying (end of life cycle), and the finished shell is sinking without further modification (calcification or dissolution) to the trap. Why ‘survival’?
The reviewer assumes that the foraminifera die immediately after formation of the last chamber, this need not to be the case in culture studies (Spindler, 1996) and we explicitly take this possibility into account. We also model this behaviour because of indications for the presence of a non-calcifying population in the data themselves. In our first paper describing these data (Jonkers et al., 2010) we attributed the low d18O values in autumn to remnants of the summer population that were no longer actively calcifying. However, we do see that the term “surviving” may be confusing and changed “add a delay between formation of the final chamber and arrival at the sediment trap that reflects survival in the water column (without calcification) and sinking time.” to “add a delay between formation of the final chamber and arrival at the sediment trap that reflects time spent in the water column without calcification as well as sinking time.” (see line). We also changed the wording elsewhere in the manuscript (section 3.2).

Line 158: For clarification: The authors mean the time span between the formation of the first chamber of the final whorl, and the last chamber of the final whorl? – please reword for clarity
We made it clearer that this refers to “modelled” foraminifera that, as stated, consist of only four chambers. In this sense, this has no bearing on real foraminifera shells that consist of multiple whorls. We will change the text to: “calcification spans (the time it takes to form the four-chambered synthetic shell chambers)” See line 199-200.

Line 181: “ignore” sounds very harsh. What about: “…was not considered…”
Changed.

Line 186: For clarity: What about: “In order to approximate the measured d18O values with our model simulation, we average the d18O of four simulated shells”…
Changed.

Line 191: For clarity: … if the standard deviation of the measured d18O values (correct?) is higher than the observed…
Changed.

Lines 195, 205, 209: please do not use “ignore”
Reworded.

Line 203: please reword “foraminifera would see”. What about: “the additional variability in temperature the individual planktic foraminifera would be exposed during its life cycle”
Changed.

Line 224: suggestion: “…and the range in measured d18O is, in all cases, smaller than the….”. However, this is a bit confusing. If I understand correctly, the range in measured d18O is consistently smaller than the seasonal range in surface d18O equilibrium. However, the range in measured d18O exceeds the range of d18O equilibrium during time intervals with an isothermal water column (see lines 243-245). The authors may consider to put these information together for clarity..
We changed “There is no relationship between the number of measurements within a sample and the range in δ18O and it is always smaller than the seasonal range in surface δ18Oequilibrium and most of the time also smaller than the vertical gradient in δ18Oequilibrium (Fig. 4).” to “There is no relationship between the number of measurements within a sample and the range in δ18O (Fig. 4). The within sample range is always smaller than the seasonal range in surface δ18Oequilibrium. Most of the time the observed δ18O range is also smaller than the vertical gradient in δ18Oequilibrium, except during isothermal conditions in spring when it exceeds the δ18Oequilibrium range (Fig. 4).”

Line 230: suggestion: “…regarding these initial observations…”
OK

Line 235: suggestion: “The fact that this cannot be seen in the data…”
OK

Line 239: “if the observed variability in foraminifera d18O is higher…. expected from temperature and d18O seawater at the time…”
Reworded.

Line 240: prior to the sampling
We are unsure what the reviewer is commenting about. These words appear exactly like this in the text.

Line 241: delays (plural). Please see my earlier comment regarding ‘survival’. I still don’t know what it actually means. I assume the authors would like to say that the ‘finished’ shells remains in the water column without any further modification (of course, these are assumptions for the model, nature is more complex), until it is collected in the sediment trap
We added the “s”. Please see our response above regarding survival.

Line 246: for clarity: please mention again: what are the two scenarios? (1) Variable calcification depth, and (2) calcification during summer?
We changed “Our simulations are thus sensitive to the choice of calcification depth and it is important to assess if both scenarios are equally realistic. We can do so by determining the prediction error in the mean δ18O across all samples (Fig. 6).” to: “Our simulations are thus sensitive to the choice of calcification depth and it is important to assess if the scenario with variable depth habitat is more realistic than the scenario with constant, near-surface habitat. We can compare both scenarios by determining the prediction error in the mean δ18O across all samples (Fig. 6).” See lines 310-314.

Line 273: It shall read “Davis et al., 2017, 2020a”
OK

Line 286: suggestion: “when variations in temperature and…”
We changed the wording.

Line 296: “In the first study, the range in …amounts to 0.15‰ (Leduc et al., 2009). In the second study,…”
OK

Line 321-326: Please also add a few sentences explaining that N. pachyderma features no symbionts, thus, we can exclude the effect of symbiont activity on shell-d13C
The reviewer rightly points out that symbiont activity cannot affect the d13C in N. pachyderma. However, in this section we discuss possible causes for variability in both d18O and d13C. Since symbiont activity does not affect d18O we see no merit in mentioning factors that could only affect d13C.

Line 332-325: This is an important discussion – the proportion of crust to lamellar calcite. The authors are discussing that the crust calcite has a different d18O value than the lamellar calcite (lines 339-340). Yes, but this is because the crust is typically formed in deeper waters. Livsey et al. (2020) has shown that both the crust and the lamellar calcite likely form in equilibrium with ambient temperature and seawater d18O.

Therefore, the difference between lamellar calcite d18O and crust calcite d18O can only be explained by downward migration in the water column. However, in this manuscript, the authors postulate that the calcification depth of N. pachyderma is limited to a well-defined, narrow band. There is the risk that this discussion is contradicting previous statements from the authors.

We agree with the reviewer that this is an important point that should be clarified. Reviewer 1 also pointed to an inconsistency that we inadvertently included. We apologise for the confusion.

Livsey et al measured the d18O on lamellar and crust calcite from N. pachyderma shells from the same sediment trap time series. They performed measurements on shells intercepted in spring, i.e. formed during isothermal conditions and found a small, but not significant, difference between crust and lamellar calcite d18O. The conclusion put forward by the reviewer ("Therefore, the difference between lamellar calcite d18O and crust calcite d18O can only be explained by downward migration in the water column") is not supported by the data and therefore not in conflict with the inferred narrow band of calcification. We changed the section to: "The excess variability could also arise from differences in the proportion of crust to lamellar calcite. Variable crust to lamellar calcite ratios among foraminifera could therefore add temperature-independent noise, similar to what has been suggested for Mg/Ca (Jonkers et al., 2021, 2016). However, the difference between crust and lamellar calcite δ18O of N. pachyderma intercepted in spring when the water column was well-mixed is not significant (Livsey et al. 2020). Variable encrustation can therefore not be the explanation for the excess δ18O variability observed during isothermal conditions in spring. In addition, this explanation would require that the crust and lamellar calcite also have different carbon isotope ratios. However, previous work is inconclusive in this regard. Observations from plankton hauls suggest that encrusted and crust-free N. pachyderma have systematically different δ13C, but that the effect of encrustation is not as strong as on δ18O (Kohfeld et al., 1996). A larger dataset from the sediment on the other hand, indicates no effect of encrustation (Healy-Williams, 1992). Whether or not variable encrustation is the cause of the observed excess variability in δ18O and δ13C therefore remains an open question." to avoid confusion. See lines 444-458.

In addition, there is no discussion whether the authors have carefully investigated the shells by binocular microscope. N. pachyderma shells collected by sediment traps typically feature only a thin crust, or the crust is entirely absent (in contrast, fossil shells typically feature a thick crust). I believe that some information about the degree of encrustation of the investigated N. pachyderma shells would help to bolster the discussion regarding the potential impact of crust calcite on the variability in d18O.

We have not systematically investigated the degree of encrustation in the time series, however, we never observed any specimens without a crust (see also Jonkers (2016)). To make this clearer we have added "We did not perform a systematic analysis of the degree of encrustation of N. pachyderma in the sediment trap samples,but in the many years of work on this time series we have never come across a crust-free specimen. It is nevertheless likely that the degree of encrustation varies among individuals and variable crust to lamellar calcite ratios among foraminifera could therefore add temperature-independent noise,..." (line 444-447). Note however, that we cannot draw firm conclusions about the influence of crusting on d13C because there is no consensus in literature. So, even though it is highly unlikely that the degree encrustation was constant among the individuals, it does not affect

the conclusion. We also made this clear in the last sentence of the paragraph "Whether or not variable encrustation is the cause of the observed excess variability in δ18O and δ13C therefore remains an open question."

Line 350: I prefer to be careful and not implying that this is the case for all planktic foraminifera. So far, we only have data for N. pachyderma. For other species, there are only indirect indications.
We disagree and think our findings have broader implications than for N. pachyderma alone, especially because of the indications for similar variability in other species (see lines 366-368 in the original text). We have also stressed the need for more research (lines 369-360) and feel that our phrasing ("we therefore presume…") is sufficiently careful. Thus we prefer to keep to the original wording.

Line 356: Thus, for now, it needs to be assumed that N. pachyderma forms it shell in equilibrium with seawater d18O and ambient temperature, superimposed by a noise of 0.11‰? I still would be a bit more cautious. The model simplifies very complex natural processes, and some of the apparent excess noise may reflect inabilities of the model to accurately reflect nature. Culture studies would help to provide more confidence (of course, there is the issue of culturing N. pachyderma successfully...)
We agree that the quantification of the noise level model-dependent and will make this clearer in the text (see also the comment by reviewer 1). We also agree that culture studies may help and would like to highlight that there has been tremendous progress in culturing this species recently (Davis et al., 2020). We are therefore happy to mention this again in the text.

Line 373: Please add more information regarding Mg/Ca (temperature proxy, why could it be useful in future studies to elucidate the cause of variability). Without additional information, this may not be clear to some readers.
We added this information.

Line 385: "…that has so far been…"
Ignored? We changed to " … has so far not been considered…"

Fig. 2: Although mentioned in the figure caption, it would be nice to have a legend, explaining yellow points and green bars. Please add a description of Panels B) and D) to the figure caption. These enlarged plots show the sampling interval April 2006 – March 2007, correct?
We added the following sentence to the figure caption to better explain what is shown in the different panels: "Panels A and C show the time series of δ18O and δ13C, respectively. Panels B and D highlight the annual pattern, they show the same data collapsed onto a single year.". We also moved the sentence explaining the meaning of the colours, so the explanation appears earlier in the caption.

Fig. 4: I cannot see any difference between the lines in grey color, depicting the difference in d18O between the surface and 200-250 m water depth, and the (same?) line in Fig. 2 depicting surface d18Oeqilibrium.

The difference between surface and deep d18Oeq shown in figure 4A is indeed very similar, but not identical, to the line showing surface d18Oeq in Fig. 2. This is because of the lack of substantial variability of d18Oeq at depth (dark line in Fig 2).

References:

Bauch, D. (1997). "Oxygen isotope composition of living Neogloboquadrina pachyderma (sin.) in the Arctic Ocean." Earth and Planetary Science Letters 146: 47-58.

Ravelo, A. C. and C. Hillaire-Marcel (2007). The Use of Oxygen and Carbon Isotopes of Foraminifera in Paleoceanography. Developments in Marine Geology. H. M. Claude and V. Anne De, Elsevier. Volume 1: 735-764.

Simstich, J., et al. (2003). "Paired d18O signals of Neogloboquadrina pachyderma (s) and Turborotalita quinqueloba show thermal stratification structure in Nordic Seas." Marine Micropaleontology 48: 107-125.

Våge, K., Pickart, R.S., Sarafanov, A., Knutsen, Ø., Mercier, H., Lherminier, P., van Aken, H.M., Meincke, J., Quadfasel, D., Bacon, S., 2011. The Irminger Gyre: Circulation, convection, and interannual variability. Deep Sea Research Part I: Oceanographic Research Papers, 58(5): 590-614.

References

Davis, C. V., Livsey, C. M., Palmer, H. M., Hull, P. M., Thomas, E., Hill, T. M., and Benitez-Nelson, C. R.: Extensive morphological variability in asexually produced planktic foraminifera, Science Advances, 6, eabb8930, 2020.

Dolman, A. M. and Laepple, T.: Sedproxy: a forward model for sediment-archived climate proxies, Climate of the Past, 14, 1851–1868, 2018.

Ganssen, G. M., Peeters, F. J. C., Metcalfe, B., Anand, P., Jung, S. J. A., Kroon, D., and Brummer, G. J. A.: Quantifying sea surface temperature ranges of the Arabian Sea for the past 20 000 years, Climate of the Past, 7, 1337–1349, 2011.

Groeneveld, J., Ho, S. L., Mackensen, A., Mohtadi, M., and Laepple, T.: Deciphering the Variability in Mg/Ca and Stable Oxygen Isotopes of Individual Foraminifera, Paleoceanography and paleoclimatology, 34, 755–773, 2019.

Jonkers, L. and Kučera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, Climate of the Past, 13, 573–586, 2017.

Jonkers, L., Brummer, G.-J. A., Peeters, F. J. C., van Aken, H. M., and De Jong, M. F.: Seasonal stratification, shell flux, and oxygen isotope dynamics of left-coiling *N. pachyderma* and *T. quinqueloba* in the western subpolar North Atlantic, Paleoceanography, 25, PA2204; 10.1029/2009PA001849, 2010.

Jonkers, L., Jiménez-Amat, P., Mortyn, P. G., and Brummer, G.-J. A.: Seasonal Mg/Ca variability of *N. pachyderma* (s) and *G. bulloides*: Implications for seawater temperature reconstruction, Earth and planetary science letters, 376, 137–144, 2013.

Jonkers, L., Buse, B., Brummer, G.-J. A., and Hall, I. R.: Chamber formation leads to Mg/Ca

banding in the planktonic foraminifer *Neogloboquadrina pachyderma*, Earth and planetary science letters, 451, 177–184, 2016.

Lougheed, B. and Metcalfe, B.: Testing the effect of bioturbation and species abundance upon discrete-depth individual foraminifera analysis, , https://doi.org/10.5194/bg-2021-202, 2021.

Spindler, M.: On the salnity tolerance of the planktonic foraminifer *Neogloboquadrina pachyderma* from Antarctic sea ice, Proceedings of the NIPR Symposium on Polar Biology, 9, 85–91, 1996.

Formatted: Header

Variability in *Neogloboquadrina pachyderma* stable isotope ratios from isothermal conditions: implications for individual foraminifera analysis

Lukas Jonkers[1], Geert-Jan A. Brummer[2], Julie Meilland[1], Jeroen Groeneveld[3] and Michal Kucera[1]

Correspondence to: Lukas Jonkers (ljonkers@marum.de)

[1] MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany
[2] Royal Netherlands Institute for Sea Research, Texel, the Netherlands
[3] Department of Geology, Hamburg University, Hamburg, Germany

Abstract. Individual foraminifera analysis (IFA) holds promise to reconstruct seasonal to interannual oceanographic variability. Even though planktonic foraminifera are reliable recorders of environmental conditions on a population level, whether they also are on the level of individuals is unknown. Yet, one of the main assumptions underlying IFA is that each specimen records ocean conditions with negligible noise. Here we test this assumption using stable isotope data measured on groups of four shells of *Neogloboquadrina pachyderma* from a 16-19 days resolution sediment trap time series from the subpolar North Atlantic. We find a within-sample variability of 0.11 and 0.10 ‰ for $\delta^{18}O$ and $\delta^{13}C$ respectively that show no seasonal pattern and exceed water column variability in spring when conditions are homogeneous down to 100s of metres. We assess the possible effect of life cycle characteristics and delay due to settling on foraminifera $\delta^{18}O$ variability with simulations using temperature and $\delta^{18}O_{seawater}$ as input. These simulations indicate that the observed $\delta^{18}O$ variability can only partially be explained by environmental variability. Individual *N. pachyderma* are thus imperfect recorders of temperature and $\delta^{18}O_{seawater}$. Based on these simulations, we estimate the excess noise on $\delta^{18}O$ to be 0.11 ± 0.06 ‰. The origin and nature of the recording imprecision require further work, but our analyses highlight the need to take such excess noise into account when interpreting the geochemical variability among individual foraminifera.

Short summary. The variability in the geochemistry among individual foraminifera is used to reconstruct seasonal to interannual climate variability. This method requires that each foraminifera shell accurately records environmental conditions, which we test here using a sediment trap time series. Even in the absence of environmental variability, planktonic foraminifera display variability in their stable isotope ratios that needs to be considered in the interpretation of individual foraminifera data.

Formatted: Header

Deleted: We

1 Introduction

Planktonic foraminifera hold the promise to provide palaeo-environmental information at high temporal resolution, owing to their life cycle, which is in the order of weeks to months and calcification that takes place over hours to days. This potential is exploited in individual foraminifera analysis (IFA), when instead of measuring groups of shells, shells are measured individually and the variability among the individual shells is used to reconstruct environmental variability during deposition of the sample. This approach has been applied to reconstruct changes in intra- and inter-annual ocean variability across time scales (Ganssen et al., 2011; Leduc et al., 2009; Rustic et al., 2015).

The use of IFA to reconstruct past oceanographic variability implicitly assumes that each foraminifera shell is a perfect recorder of environmental conditions during calcification and that there is no, or negligible, biological noise in this recording. The assumption of perfect recording seems reasonable because at population level temperature exerts a dominant control on foraminifera $\delta^{18}O$ and Mg/Ca (Bemis et al., 1998; Elderfield and Ganssen, 2000). Analytical issues aside (Fehrenbacher et al., 2020), the uncertainty associated with IFA is often viewed from the perspective of whether the population is well enough characterised, how habitat tracking may affect the results or how variability at different time scales (seasonality/ENSO) can be distinguished (Glaubke et al., 2021; Leduc et al., 2009; Metcalfe et al., 2020; Thirumalai et al., 2013) and only few consider calibration issues associated with individual planktonic foraminifera as a source of uncertainty (Glaubke et al., 2021).

However, there are several indications suggesting that whilst temperature exerts a first order control on the Mg/Ca and $\delta^{18}O$ of foraminifera, other factors (biotic and/or abiotic) also play a role. For instance, the variability in Mg/Ca and $\delta^{18}O$ in foraminifera populations from sediment samples often exceeds the variability that can be expected based on local hydrography (Groeneveld et al., 2019; Leduc et al., 2009). Whilst such evidence from sediment may be ambiguous due to uncertainty in the age of the sample and the exact habitat of the foraminifera analysed, laboratory studies also suggest that foraminifera geochemistry is affected by temperature-independent variability (Dueñas-Bohorquez et al., 2011; de Nooijer et al., 2014; Spero and Lea, 1993). Laboratory-based calibrations of $\delta^{18}O$-temperature relationships hint at a similar non-temperature related noise (Bemis et al., 1998; Erez and Luz, 1982). Observations from plankton nets and sediment traps also demonstrate marked variability (Davis et al., 2020b; Haarmann et al., 2011; Livsey et al., 2020). These observations are not conclusive in their own right, but together they suggest that there are reasonable grounds to assess if the composition of individual foraminifera can be used as a reliable environmental indicator.

Here we assess the variability in $\delta^{18}O$ and $\delta^{13}C$ among shells of *Neogloboquadrina pachyderma* collected in the subpolar North Atlantic Ocean using a moored sediment trap. The advantage of using sediment trap material is that the temporal origin of the shells is much better constrained than in sedimentary material (days to weeks compared to years to centuries) and that seasonal variability in the abundance of foraminifera does not affect the geochemical variability within each sample. Previous work on this time series has shown that on a population level *N. pachyderma* faithfully tracks the seasonal cycle in upper ocean temperature at this location (Jonkers et al., 2010). The site in the Irminger Sea serves as a natural laboratory because of deep wintertime mixing that makes the water column homogeneous down to 100s of metres. In this study we reanalyse the previously

Formatted: Header

Deleted: Leduc et al., 2009;

Deleted: Leduc et al., 2009; Thirumalai et al., 2013;

Deleted: geochemistry

Deleted: Leduc et al., 2009;

Deleted: ; Dueñas-Bohorquez et al., 2011

Deleted: Livsey et al., 2020;

Deleted: precise

Deleted: in

published data with the specific aim to assess the variability in the stable isotope ratios and to what degree the observed variability can be explained by variability in the environment. We observe marked variability in $\delta^{18}O$ and $\delta^{13}C$ even at times when the water column was thoroughly mixed. We use a simple model to evaluate the influence of life cycle characteristics on foraminifera $\delta^{18}O$ variability and find that the observed variability exceeds predictions. Our simulations provide a first-order quantification of the excess $\delta^{18}O$ variability and we argue that this biological noise should be considered when interpreting the variability in $\delta^{18}O$ among individual foraminifera.

2 Material and methods

2.1 Sediment trap mooring setting

We analyse stable oxygen and carbon isotope data from *N. pachyderma* from a 2.5-year long sediment trap time series from the centre of the Irminger Gyre (ca. 59.25° N, 38.66° W; Fig. 1). The sediment trap was positioned at a water depth of 2750 m, 250 m above the bottom. Collecting intervals were 19 days from autumn 2003 to autumn 2004 and 16 days from autumn 2005 to summer 2007. During the year, temperature, which is the main control on $\delta^{18}O$ at this location (Jonkers et al., 2010), varies between approximately 5 and 10 °C near the surface (Fig. 1). There is no marked seasonal cycle in temperature from around 200 m depth, where temperatures remain at approximately 5 °C year-round. Deep convective mixing, resulting in isothermal conditions, takes place in winter time (de Jong et al., 2012). The time series of *N. pachyderma* stable isotopes we analyse here captures these isothermal conditions three times.

2.2 Data

Stable isotope measurements were performed on groups of four *N. pachyderma* shells (150-250 µm) with up to six measurements per collection interval. In Jonkers et al. (2010) we presented average stable isotope data, but here we return to the raw data and assess the variability within each sample. Even though the measurements were done on groups of four shells, the replicate measurements on small numbers of shells allow us to obtain a first order estimate of the minimum stable isotope variability within the population of *N. pachyderma*. Our analyses are therefore meaningful for the interpretation of IFA results. Not all samples from the time series contained enough shells of *N. pachyderma* (Fig. 1), so the complete data set consists of 172 measurements from 45 samples, of which 163 are from 36 samples with at least two measurements. All measurements were done using a Thermo MAT253 mass spectrometer coupled to a Kiel IV device. The analytical error (1 s.d.), determined from repeat measurements of the NBS-19 standard, amounts to 0.05 ‰ for $\delta^{18}O$ and 0.03 ‰ for $\delta^{13}C$. Further details about the mooring and the analytical procedure are presented in Jonkers et al. (2010).

The number of replicate measurements per sample is relatively low compared to what is used for IFA on sedimentary material. This is however justified given the short collection intervals of sediment trap samples (in our case 16-19 days) compared to the long integration time of sediment samples (at least decades to centuries). Moreover, with low numbers of measurements we are likely to underestimate the variability at population level and our inferences will therefore be conservative.

Formatted: Header

Deleted: The

Deleted: shells measured

Deleted: integration time

Deleted: weeks to months

In order to obtain a conservative estimate of the variability among the measured groups of *N. pachyderma* shells we remove possible outliers. Given the small sample sizes, outliers were identified using all data in Fig. 2, and excluded from our analysis to avoid unnecessary inflation of inter-specimen variability. We calculated the residual from the mean for each sample and defined outliers as being more than 1.5 times the interquartile range away from the overall mean (Fig. 3). This approach resulted in the removal of 10 (6%) and 4 (2%) measurements of $\delta^{18}O$ and $\delta^{13}C$, respectively.

We compare the observations to expected $\delta^{18}O$ equilibrium values and estimates of the $\delta^{13}C$ of dissolved inorganic carbon ($\delta^{13}C_{DIC}$). We calculate equilibrium $\delta^{18}O$ ($\delta^{18}O_{eq}$) using the Kim and O'Neil (1997) palaeotemperature equation because *N. pachyderma* calcifies without an offset from this equation (Jonkers et al., 2010, 2013). For the deployments from 2003-2004 and 2005-2006 we use the same temperature and salinity data as in previous work (2010, 2013). However, for the deployment from 2006-2007 temperature and salinity data at 10 and 266 m are available from the nearby CIS mooring (59.66° N; 39.66° W) and we use these as it allows using *in-situ* surface salinity measurements and because of better temporal coverage at depth (Jonkers et al., 2016). Seawater $\delta^{18}O$ ($\delta^{18}O_{seawater}$) was derived from salinity, using the regional salinity-$\delta^{18}O_{seawater}$ relationship used in Jonkers et al. (2010).

Estimates of $\delta^{13}C_{DIC}$ are the same as in Jonkers et al. (2013) and based on multiple-linear regression of temperature, salinity and nutrients within the wider subpolar North Atlantic. Since the $\delta^{13}C_{DIC}$ data are derived from data that represent long-term average conditions (climatology), they cannot be used to the same level of detail as $\delta^{18}O$. We compare the measured variability in $\delta^{13}C$ to the seasonal range in $\delta^{13}C_{DIC}$ and the seasonal range in expected foraminifera $\delta^{13}C$ by taking into account a temperature-dependent offset from $\delta^{13}C_{DIC}$ (Jonkers et al., 2013).

2.3 Predicting *N. pachyderma* $\delta^{18}O$ variability

Planktonic foraminifera intermittently add chambers during their life cycle and start sinking towards the ocean floor upon death. The signal contained in their stable isotope ratios is therefore a reflection of the environmental conditions during a certain time prior to arrival in the sediment trap. To assess if the observed variability in $\delta^{18}O$ can be explained by temperature and $\delta^{18}O_{seawater}$ alone, we predict $\delta^{18}O$ calcite ($\delta^{18}O_{equilibrium}$) using a model that is more complex in its representation of calcification than what is usually attempted when interpreting results of individual foraminifera analyses (Glaubke et al., 2021; Groeneveld et al., 2019; Thirumalai et al., 2013). We simulate foraminifera $\delta^{18}O$ as an average of chamber $\delta^{18}O$ and add a delay between formation of the final chamber and arrival at the sediment trap that reflects time spent in the water column without calcification and sinking to the depth of the trap. In this way we represent calcification during the foraminifera life cycle more realistically and allow for more variability than when assuming that each foraminifera shell represents environmental conditions averaged over one (calendar) month. Our approach is based on the following assumptions: 1) foraminifera build their chambers at random times during their life cycle; 2) chamber formation takes one day; 3) each foraminifera shell consists of four chambers with equal mass and 4) all shells have the same mass.

Formatted: Header

Deleted: (Jonkers et al., 2013, 2010).

Deleted: Jonkers et al.,

Deleted: , 2010

Deleted: because

Deleted: available as

Deleted: only and can hence not

Deleted: observed

Deleted: build their skeleton over a certain calcification time

Deleted: Thirumalai

Deleted: 2013

Deleted: Glaubke

Deleted: 2021

Deleted: survival

Deleted: (

Deleted:)

Deleted: time.

Deleted: maximum

The first assumption is reasonable in light of the limited amount of information available on the (temporal aspects of the) ontogeny of *N. pachyderma* (Bé et al., 1979; Spindler, 1996). The assumed duration of chamber formation is based on culture studies (Bé et al., 1979; Spindler, 1996). However, culture studies in the closely related species N. dutertrei have shown that chamber formation may take up to four days (Fehrenbacher et al., 2017). Longer chamber formation could in theory reduce the variability foraminifera $\delta^{18}O$ because of increased smoothing of the environmental signal. In practice this effect is however negligible because of strong temporal autocorrelation in the $\delta^{18}O_{equilibrium}$ time series that renders the effect of smoothing of up to four days insignificant. Our approach thus yields an estimate of variability that is robust against the likely range of chamber formation duration. In *N. pachyderma* the last whorl of the shell makes up most of the mass and generally consists of four chambers that are of similar size. The assumed number and equal mass of the chambers is thus reasonable. The last assumption is out of convenience.

For each sample we simulate $\delta^{18}O$ for different calcification spans (the time it takes to form a four-chambered synthetic shell) and delays (the time between formation of the last chamber and arrival at the trap). We vary the calcification span between 4 and 168 days and the delay between 5 and 180 days. The minimum value for the delay is based on estimates of sinking velocity of planktonic foraminifera (Takahashi and Bé, 1984). We exclude scenarios where the sum of calcification span and delay is more than 181 days because of the clear seasonal pattern in mean $\delta^{18}O$. This pattern indicates that long delays are unlikely because minimum $\delta^{18}O$ values are observed shortly after peak temperatures. Very long calcification spans are also unlikely as these would result in small seasonal $\delta^{18}O$ variation. We allow for some variability in the calcification span and delay by varying the calcification span in each scenario within a lognormal distribution with the mode equal to the calcification span and a standard deviation of 0.3. The delay is varied using a normal distribution with a standard deviation that is the square root of the delay.

To investigate the effect of calcification depth we run two groups of simulations, one where we assume that calcification takes place exclusively at the surface and another where we allow for variable calcification depth, either near the surface or at depth (ca. 250 m), within each sample. We include the possibility that shells were formed at depth because *N. pachyderma* is known to inhabit a wide depth range (Greco et al., 2019) and previous studies indicated a large and variable apparent calcification depth (Kohfeld et al., 1996; Simstich et al., 2003). However, the real range of apparent calcification depth of *N. pachyderma* in the Irminger Sea is probably narrower than the 200-250 m assumed in the simulations. This is because the average $\delta^{18}O$ of *N. pachyderma* shows a seasonal trend with a magnitude that suggests an apparent calcification depth around 50 m (Jonkers et al., 2010, 2013). This scenario thus likely overestimates variability, especially during the summer season when the water column is stratified. We do not simulate calcification exclusively at depth because this is clearly at odds with observed seasonal amplitudes of $\delta^{18}O$ and $\delta^{13}C$.

We do not consider the possibility of ontogenetic vertical migration in our simulations. This is partly an assumption out of necessity because we do not have temperature and salinity data between the surface and 200 m depth for the entire time series. We however stress that our approach is conservative because ontogenetic migration would decrease the variability in foraminifera stable isotope ratios.

Formatted: Header

Deleted: Spindler, 1996;

Deleted: in agreement with

Deleted: Spindler, 1996;

Deleted: .

Deleted: would

Deleted: more

Deleted: too high than too low

Deleted: in which the four chambers are formed

Deleted: 1

Deleted: We

Deleted: In all simulations we ignore

Deleted: this

To be consistent with the measurements on groups of four shells, we average the $\delta^{18}O$ of four simulated shells. We add measurement uncertainty (white noise with a standard deviation of 0.05 ‰) to the averaged $\delta^{18}O$ and calculate the standard deviation of the $\delta^{18}O$ of 2-6 groups (depending on the sample) of four shells. We repeated this process 300 times for each sample and for each combination of delay and calcification span. We consider cases significant when the predicted standard deviation is higher than the observed standard deviation in 95 % of the simulations.

Estimates of $\delta^{18}O_{equilibrium}$ are not available for the entire time series and our simulations are therefore restricted to the spring of 2004, the spring to autumn of 2006 and the spring of 2007. Because we lack detailed data on $\delta^{13}C_{DIC}$ we did not simulate foraminifera $\delta^{13}C$. We, however, do not ignore foraminifera $\delta^{13}C$ in our analysis.

Modelling is by definition a simplification of reality. Even though important aspects of our model (variable depth, faster calcification) yield estimates of expected variability that are higher than in previous work, we follow previous work and consider local temperature and $\delta^{18}O_{seawater}$ as the only predictors of $\delta^{18}O_{equilibrium}$ (Glaubke et al., 2021; Thirumalai et al., 2013). For simplicity we do not consider advection of foraminifera because it is not directly clear how advection within the Irminger Gyre, where temperatures are spatially rather uniform, would influence the temperature variability that planktonic foraminifera would be exposed to during calcification. Assessing the influence of advection can only be done using lagrangian modelling (van Sebille et al., 2015) and ultimately relies on the accuracy with which the model captures spatial and temporal temperature variability. Such modelling is beyond the scope of this study. We neither consider the effect of the carbonate ion concentration ($[CO_3^{2-}]$) on foraminifera stable isotopes (Spero et al., 1997). Because of the positive correlation between temperature and $[CO_3^{2-}]$(Jonkers et al., 2013) and a negative correlation between $[CO_3^{2-}]$and foraminifera $\delta^{18}O$ (Spero et al., 1997) the $[CO_3^{2-}]$ effect would slightly increase the seasonal range $\delta^{18}O_{equilibirum}$. Assuming that the sensitivity of *N. pachyderma* $\delta^{18}O$ is similar to that of *G. bulloides*, the increase would be in the order of 0.15 ‰. Since we do not consider this possible source of variability, our simulations are likely to provide conservative estimates of foraminifera $\delta^{18}O$ variability.

3 Results and discussion

3.1 Raw data

The $\delta^{18}O$ of *N. pachyderma* varies between 0.93 ‰ in early winter 2006 and 2.88 ‰ in spring 2004 (Fig. 2). The overall seasonal amplitude is around 1 ‰, with a minimum in $\delta^{18}O$ that lags the maximum temperatures by one to two months. Stable oxygen isotope ratios are in general within the range of predicted $\delta^{18}O_{equilibrium}$. The $\delta^{13}C$ values show a smaller amplitude (-0.37 to 0.58 ‰) and are always offset from $\delta^{13}C_{DIC}$ (Fig. 2). The $\delta^{13}C$ values generally decrease from spring to winter. For both $\delta^{18}O$ and $\delta^{13}C$ the observed within sample variability exceeds the analytical uncertainty (Fig. 3).

After outlier removal, the within-sample range of $\delta^{18}O$ varies between 0.05 and 0.51 ‰ (mean 0.24 ‰) and does not show a consistent pattern during the year (Fig. 4). There is no relationship between the number of measurements within a sample and the range in $\delta^{18}O$ (Fig. 4). The within sample range is always smaller than

Formatted: Header

Deleted: To approximate the measured data

Deleted: ; Glaubke et al., 2021

Deleted: ignore

Deleted: see.

Deleted: also ignore

Deleted: ignore

Deleted: and it

the seasonal range in surface $\delta^{18}O_{equilibrium}$. Most of the time the observed $\delta^{18}O$ range is also smaller than the vertical gradient in $\delta^{18}O_{equilibrium}$, except during isothermal conditions in spring when it exceeds the $\delta^{18}O_{equilibrium}$ range (Fig. 4). The range in $\delta^{13}C$ is similar to $\delta^{18}O$ and varies between 0.06 and 0.46 ‰ (mean 0.21 ‰) and neither shows a clear seasonal pattern (Fig. 4). Compared to $\delta^{18}O$, the range of foraminifera $\delta^{13}C$ is more often above the expected range (Fig. 4).

There are two important points regarding these initial observations. The first is that the observed range in foraminifera stable isotope values exceeds the expected range in spring (April - May) when the water column is well-mixed down to 800 m depth. This variability arises from apparently random positive and negative offsets from $\delta^{18}O_{equilibrium}$, suggesting that it does not result from a mechanism that would cause a systematic bias in the foraminifera $\delta^{18}O$. Advection or long foraminifera life spans, which could theoretically cause foraminifera from the previous summer to survive until spring, are therefore unlikely to provide a full explanation for the observed variability. This is the first indication that the variability in foraminifera isotope ratios does not solely result from environmental variability. The second observation is the apparent lack of a seasonal cycle in the range in $\delta^{18}O$ and $\delta^{13}C$ even though stratification develops as the sea surface warms. In theory, the variability in foraminifera stable isotope ratios could therefore increase towards the warm season. The fact that this cannot be seen in the data indicates that *N. pachyderma* calcifies in a relatively narrow and constant vertical range.

3.2 Predicted foraminifera $\delta^{18}O$ variability

To assess if observed variability in $\delta^{18}O$ of *N. pachyderma* is higher than the variability expected from temperature and $\delta^{18}O$ of seawater at the time of sampling because the foraminifera calcified prior to the sampling we carried out simulations using a range of possible calcification spans and delays. These simulations indicate that the standard deviation of *N. pachyderma* $\delta^{18}O$ in spring when the water column is virtually isothermal (IRM-1 A-14, IRM-3 A-13, IRM-3 A-14, IRM-4 A-14 and IRM-4 A-15) exceeds what can be expected based on reasonable calcification histories and delays (Fig. 5). The predicted variability only significantly exceeds the observations during summer and almost exclusively in the simulations that allow variable calcification depth. Our simulations are thus sensitive to the choice of calcification depth and it is important to assess if the scenario with variable depth habitat is more realistic than the scenario with constant, near-surface habitat. We can compare both scenarios by determining the prediction error in the mean $\delta^{18}O$ across all samples (Fig. 6). The minimum prediction error is, in both scenarios, distributed along an arc shape, with lower errors at longer calcification spans and delays up to about a month or at short calcification spans and delays in the order of one to two months. However, the errors reach markedly lower values in the scenario where calcification only occurs near the surface. Because the seasonal peak in temperature is reached earlier at the surface than at depth, it remains difficult to determine precisely which combination of calcification depth, calcification span and delay is most realistic, but the amplitude of the mean seasonal $\delta^{18}O$ indicates that the surface only scenario is closer to what the foraminifera actually experienced than the variable depth scenario. This indicates that even when taking reasonable calcification histories and delays into account, the observed variability in foraminifera $\delta^{18}O$ is unlikely to reflect environmental (temperature) variability alone.

Formatted: Header

Deleted: and most

Deleted: first

Deleted: it does not

Deleted: delay due to survival and settling.

Deleted: only

Deleted: both scenarios are equally realistic. We can do so

Our simulations also permit us to put some constraints on the calcification span and delay that characterises *N. pachyderma* at this location. The hardest constraints can be put on the possibility of long delays between formation of the last chamber and arrival at the trap. Sinking speed measurements suggest that the delay due to sinking at this location is likely to be between 5 and 19 days (Takahashi and Bé, 1984). We obtain minimum prediction errors for delays up to approximately two months (Fig. 6). Subtracting the sinking time estimates from these delays implies that *N. pachyderma* is unlikely to spend more than one month in the water column without calcifying after the last chamber has formed. This means that the simulations with delays >100 days are not realistic.

Our simulations indicate that calcification spans under two weeks yield smaller errors when associated with delays in the order of 30 - 60 days and similarly low prediction errors are obtained using longer calcification spans and shorter delays. Based on our data it is difficult to ascertain which cases are more realistic. However, such long delays would require long intervals spent in the water column without calcification. A single culture study using Antarctic *N. pachyderma* showed intermittent chamber formation over a period of about two months and a single case of gametogenesis approximately two weeks after the formation of the final chamber (Spindler, 1996). Other studies also suggest an approximately two month life span (Davis et al., 2017, 2020a). This suggests that delays of up to approximately one month (including settling) and calcification of the final four chambers over the course of about two months are most probable.

3.3 Excess foraminifera $\delta^{18}O$ variability

The mean observed standard deviation for of $\delta^{18}O$ is 0.11 ± 0.05 ‰ for the complete time series and 0.10 ± 0.03 ‰ for the samples from the time when the water column was isothermal (IRM-1 A-14, IRM-3 A-13, IRM-3 A-14, IRM-4 A-14 and IRM-4 A-15). As noted above, the fact that the variability in $\delta^{18}O$ does not show a consistent pattern during the year, suggests that we have captured the full range of within-sample variability even though the number of measurements per sample is relatively low. Since our measurements are based on groups of four shells the observed standard deviation is an underestimate of the standard deviation among individual shells. Assuming that each shell in the group contributed equally to the total mass, the degree of underestimation of the standard deviation scales with the square root of the group size (Groeneveld et al., 2019). Thus we multiply the observed standard deviation by two ($\sqrt{4}$) to obtain an estimate of the standard deviation of individual shells. That means that the $\delta^{18}O$ of individual foraminifera at this location is likely to have a standard deviation of 0.19 ± 0.07 ‰ (0.21 ± 0.11 ‰ when considering all observations).

For the samples from the times when the water column was deeply mixed, i.e. when variations in temperature, salinity and hence $\delta^{18}O_{equilirbium}$ were negligible, our simulations predict a standard deviation for individual shells of 0.08 ‰. This prediction is identical for both depth scenarios. It includes a 0.05 ‰ measurement uncertainty and is based on all considered scenarios with a delay less than 100 days, which is reasonable given the low model skill at longer delays. Assuming that our simulations are a reasonable approximation of reality, the excess variability (s.d.) that cannot be explained by variability in temperature and $\delta^{18}O_{seawater}$ is therefore 0.11 ± 0.06 ‰, which in terms of temperature roughly translates to a standard deviation of 0.4 °C.

Formatted: Header

Deleted: due to survival and/or settling.

Deleted: survival of

Deleted: for more than a month

Deleted: calcification is unlikely

Deleted: survival

Deleted: , 2017

Deleted: , the standard deviation of individual shells is double ($\sqrt{4}$)

Deleted: .

Deleted: likely also other environmental variability was

Deleted: 09

Whereas our modelling approach provides an estimate that is likely closer to reality than assuming that foraminifera reflect environmental conditions averaged over a single (calendar) month, our estimate could be evaluated by simulating other calcification trajectories. We found that our results are insensitive to the duration of chamber formation and experiments where we allowed complete shell formation within one day, equivalent to assigning all weight to the last chamber, yielded an expected standard deviation of individual foraminifera $\delta^{18}O$ of 0.09 ‰. Therefore, the assumption of equal weight of the four chambers has little bearing on our results. Ultimately, the modelled foraminifera $\delta^{18}O$ depends on the hydrographic data used to estimate $\delta^{18}O_{equilibrium}$. By using data from the surface and from great depth, we have obtained two end-member scenarios of vertical $\delta^{18}O_{equilibrium}$ variability that implicitly encompass ontogenetic vertical migration. However, future estimates of expected individual foraminifera $\delta^{18}O$ variability could be improved by explicitly incorporating horizontal $\delta^{18}O_{equilibrium}$ variability and advection during shell growth in the modelling strategy.

Apart from being sensitive to our modelling design and data availability, our estimate of excess $\delta^{18}O$ variability among individual shells is also sensitive to the quantification of variability among shells. To obtain a conservative estimate we excluded potential outliers. Were we to consider all measurements, the average standard deviation among groups would be 0.15 ± 0.11 ‰ (0.17 ± 0.09 ‰ during spring) and the resulting excess $\delta^{18}O$ variability 0.25 ± 0.19 ‰. Thus our approach yields a conservative and better constrained estimate of the excess variability.

We compare this estimate of unexplained $\delta^{18}O$ variability to two studies that used individual foraminifera $\delta^{18}O$ from cores in the eastern equatorial Pacific Ocean to infer changes in the El-Niño Southern Oscillation. In the first study, the range in the standard deviations of *N. dutertrei* $\delta^{18}O$ shells in eight time slices across the past 50,000 years amounts to 0.15 ‰ (Leduc et al., 2009). In the second study, Rustic et al. (2015) interpreted changes in the standard deviation of *G. ruber* $\delta^{18}O$ over the last millennium that were smaller than 0.45 ‰ (variance of 0.20 ‰²). Forward modelling studies also indicate that changes in the amplitude (doubling or halving) in the central equatorial Pacific would translate to changes in the standard deviation of IFA of maximum 0.15 ‰ (Thirumalai et al., 2013). In all cases, the unexplainable $\delta^{18}O$ variability we observe makes up a substantial part of the signal. Thus, non-temperature effects on individual foraminifera $\delta^{18}O$ need to be considered when interpreting the results of IFA.

3.4 Possible causes of excess variability

The relatively constant variability in $\delta^{18}O$ and $\delta^{13}C$ within the *N. pachyderma* population in the Irminger Sea during the year argues against a direct environmental influence on the variability. This is because on seasonal time scales environmental variability is strongly correlated to temperature and/or stratification. The observed variability could therefore be random or reflect biological processes within the population of foraminifera, where each shell, or each chamber, records the environment with a small offset. As long as the excess variability remains random or uncorrelated with the environment, the average stable isotope composition of (large enough subsample of) a foraminifera population will accurately reflect environmental conditions. On a population level, planktonic foraminifera $\delta^{18}O$ is indeed a reliable indicator of seawater temperature and $\delta^{18}O_{seawater}$ (e.g. Bemis et

Formatted: Header

Deleted: 04 ‰ (squared

Deleted: ‰).

Deleted: signal is similar to, or even smaller than, the

Deleted: variance

al., 1998; Erez and Luz, 1982), suggesting that the excess variability among individual specimens is cancelled out within populations.

Alternatively the excess variability could arise from environmental or biotic forcing that we did not consider in our simulations. Crucially, any possible mechanism needs to explain the approximately equal variability in $\delta^{18}O$ and $\delta^{13}C$ that we observe in the time series.

Shell size is likely to affect metabolic rates and the observed excess variability could therefore be related to differences in shell size (Spero and Lea, 1993, 1996). However, in such a scenario, the effect would be expected to be much stronger on $\delta^{13}C$ than on $\delta^{18}O$, as is the case for *G. bulloides* (Spero and Lea, 1996). The comparable variability in both carbon and oxygen isotope ratios thus suggests that size differences within the foraminifera population are unlikely to explain the observed excess variability.

Along similar lines, growth rate may vary among individual foraminifera and thereby influence the stable isotope composition, as has for instance been shown for corals (McConnaughey, 1989). However, in corals, $\delta^{13}C$ is, like with the size effect above, more sensitive to changes in the growth rate than $\delta^{18}O$. Therefore, if such an effect were to occur among (non-symbiotic) planktonic foraminifera, growth rate differences would neither be the likely cause of the excess variability in stable isotope ratios.

The excess variability could also arise from differences in the proportion of crust to lamellar calcite. We did not perform a systematic analysis of the degree of encrustation of *N. pachyderma* in the sediment trap samples,but in the many years of work on this time series we have never come across a crust-free specimen. It is nevertheless likely that the degree of encrustation varies among individuals and variable crust to lamellar calcite ratios among foraminifera could therefore add temperature-independent noise, similar to what has been suggested for Mg/Ca (Jonkers et al., 2016, 2021). However, the difference between crust and lamellar calcite $\delta^{18}O$ of *N. pachyderma* intercepted in spring when the water column was well-mixed is not significant (Livsey et al., 2020). Variable encrustation can therefore not be the explanation for the excess $\delta^{18}O$ variability observed during the isothermal conditions in spring. In addition, this explanation would require that the crust and lamellar calcite also have different carbon isotope ratios. However, previous work is inconclusive in this regard. Observations from plankton hauls suggest that encrusted and crust-free *N. pachyderma* have systematically different $\delta^{13}C$, but that the effect of encrustation is not as strong as on $\delta^{18}O$ (Kohfeld et al., 1996). A larger dataset from the sediment on the other hand, indicates no effect of encrustation (Healy-Williams, 1992). Whether or not variable encrustation is the cause of the observed excess variability in $\delta^{18}O$ and $\delta^{13}C$ therefore remains an open question.

Notwithstanding the fact that the exact cause of the excess variability in *N. pachyderma* stable isotope ratios needs to be constrained in future studies, our analysis shows that individual planktonic foraminifera record environmental conditions with less precision than average populations. Our study thus confirms earlier indications (Groeneveld et al., 2019; Livsey et al., 2020), but we have attempted a first quantification of this noise for $\delta^{18}O$, which has up to now been ignored in the interpretation of individual foraminifera data.

Formatted: Header

Deleted: , 1993

Deleted: Crust and lamellar calcite have different $\delta^{18}O$ even in *N. pachyderma* shells formed in isothermal conditions

Moved down [1]: (Livsey et al., 2020).

Deleted: Variable

Deleted: , 2016

Moved (insertion) [1]

Deleted: . This

3.5 Implications for reconstructions of environmental variability based on individual foraminifera

The possibility that individual planktonic foraminifera record seawater conditions with limited precision has up to now been overlooked when using the geochemistry of individual planktonic foraminifera to reconstruct climate variability. Our analyses provide evidence that the $\delta^{18}O$ of individual *N. pachyderma* shells may reflect seawater temperature and $\delta^{18}O$ with a precision of only 0.11 ‰. For now we assume that the cause of this lack of precision is random biological noise, but future studies are needed to verify that this is indeed the case, or if the recording precision is dependent on environmental or biological factors.

Our observations strengthen the case to use large numbers of foraminifera, not just for IFA. Depending on instrumental precision the biological recording noise doubles or triples the variability that can be expected in (individual) planktonic foraminifera $\delta^{18}O$, even when temperatures were constant during calcification. Any study using individual foraminifera $\delta^{18}O$ to infer past environmental variability, thus needs to cross this noise threshold in order to obtain meaningful results. Lack of recording precision will also influence the shape of the distribution of IFA results (Fig. 7), especially at the tails of the distribution that are often used to infer changes in upper ocean dynamics (Glaubke et al., 2021).

There are no reasons to believe that the existence of biological recording noise is unique to *N. pachyderma* or to stable oxygen and carbon isotopes alone. In fact, most of the indications for excess variability are based on other species (Bemis et al., 1998; Erez and Luz, 1982; Leduc et al., 2009; Spero and Lea, 1996). We therefore presume that a similar noise characterises other species and proxies as well. However, more research is needed to constrain the nature and causes of this lack of precision in the recording by individual foraminifera. Future research, including culturing, needs to consider different species in different environmental settings. Including Mg/Ca as an independent temperature-sensitive parameter may also help to elucidate the cause of the excess variability. Notwithstanding, our data clearly show that the assumption that individual planktonic foraminifera are perfect recorders of (monthly mean) temperature is not valid. Biology cannot be ignored in the interpretation of planktonic foraminifera proxies.

4 Conclusions

Stable isotope measurements on groups of four shells of *N. pachyderma* from a 16-19 day resolution sediment trap time series in the subpolar North Atlantic show large within sample variability. Stable oxygen and carbon isotope ratios within the time series have a mean standard deviation of 0.11 and 0.10 ‰, respectively and show no relationship with the seasonal trend in temperature ($\delta^{18}O_{eq}$) or the $\delta^{13}C$ of dissolved inorganic carbon. This lack of a seasonal pattern in the variability suggests that at this location *N. pachyderma* has a seasonally rather stable apparent calcification depth, which based on the amplitude of the sample mean $\delta^{18}O$ is around 50 m. Due to deep mixing the site is characterised by homogeneous water column conditions at the start of the spring foraminifera flux pulse. *Neogloboquadrina pachyderma* stable isotope variability at this time exceeds the variability that can be expected from the local hydrography, indicating that an additional source of variability that has so far not been considered in the interpretation of records of the geochemistry of individual foraminifera. Predictions of the observed variability in *N. pachyderma* $\delta^{18}O$ from temperature and $\delta^{18}O_{seawater}$

Formatted: Header

Deleted: Leduc et al., 2009;

Deleted: , but also

Deleted: , which

Deleted: ignored

using realistic calcification and settling histories fail to match the observed variability. We therefore conclude that the $\delta^{18}O$ of individual *N. pachyderma* imperfectly record temperature and $\delta^{18}O_{seawater}$. Whether random, or controlled by environmental or biological factors, *N. pachyderma* records environmental variability with some degree of noise.

Our first-order estimate of the recording noise of individual specimens amounts to 0.11 ‰ (1 sd), which is approximately double the typical analytical noise. Whilst more studies are needed to constrain the origin and variability in this recording noise, there are no reasons to believe it is a feature exclusive to *N. pachyderma*. The considerable recording noise should therefore be considered when interpreting geochemical variability among individual foraminifera.

Acknowledgements

L.J. acknowledges funding through the German climate modelling initiative PALMOD, funded by the German Ministry of Science and Education (BMBF). J.M. is funded by the Cluster of Excellence „The Ocean Floor – Earth's Uncharted Interface" funded by the German Research Foundation (DFG). The CIS mooring temperature and salinity data were collected and made freely available by the OceanSITES project (www.oceansites.org) and the national programs that contribute to it. We acknowledge funding within the VAMOC (RAPID) program NWO grant 854.00.020 for deployment of the sediment trap moorings.

Data availability statement

The stable isotope data have been submitted to pangaea.de

Author contributions

LJ concept, analysis. Modelling concept with feedback from JM

GJB mooring, funding

LJ led the writing of the manuscript and created the figures. All authors reviewed and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Formatted: Header

Deleted: The estimated

Deleted: Stable

Deleted: will be made available on

Formatted: Header

Figures

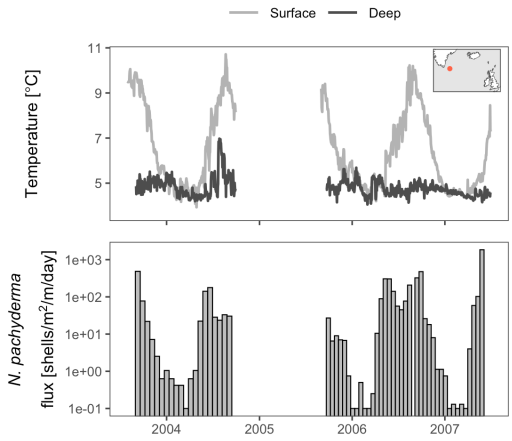


Fig.1: Temperature at the surface and at 200-250 m water depth at the Irminger Sea sediment trap mooring (red dot in map inset). In winter and spring the water column is mixed to great depths
Bottom panel shows the evolution of the shell flux of *N. pachyderma* (150-250 µm from Jonkers et al. (2010); zero fluxes are shown as 0.1 shells/m^2/day); stable isotope data are available for all but lowest flux intervals (Fig. 2). No data is available for the deployment from 2004 to 2005 because of failure of the sediment trap.

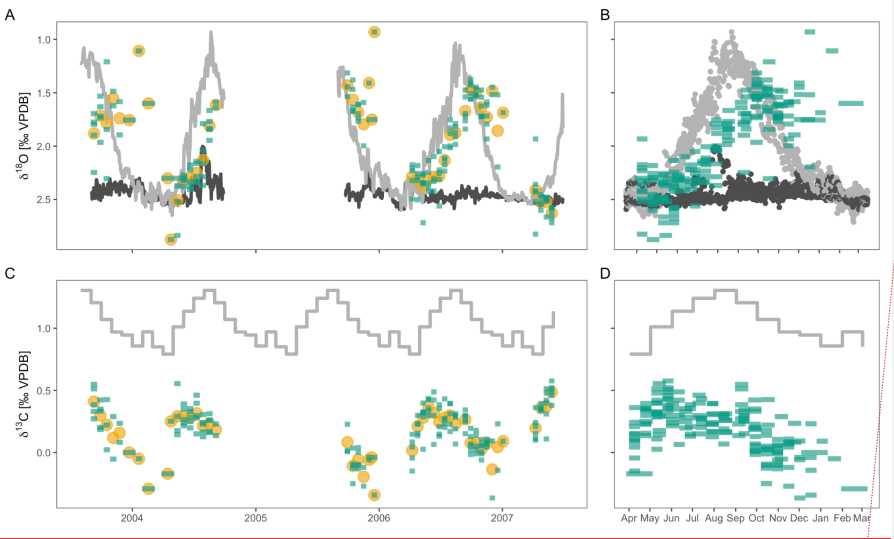
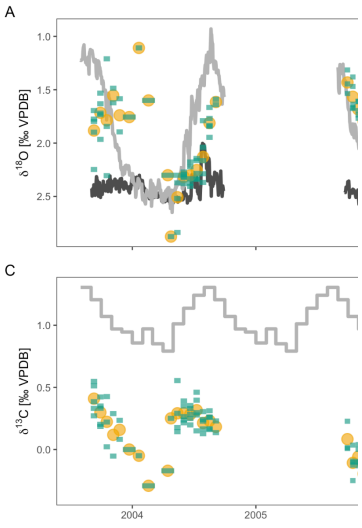


Fig. 2: *Neogloboquadrina pachyderma* stable isotopes in the Irminger Sea sediment trap time series. Panels A and C show the time series of $\delta^{18}O$ and $\delta^{13}C$, respectively. Panels B and D highlight the annual pattern, they show the same data collapsed onto a single year. Green bars extend over the collection interval and show individual measurements for groups of four shells; yellow points are average values per sample. The light grey lines depict surface $\delta^{18}O_{eq}$ and $\delta^{13}C_{DIC}$; dark grey lines in A and B are $\delta^{18}O_{eq}$ at 200-250 m depth. The oxygen and carbon isotopes show considerable variability within each sample, also when the water water column is mixed in April - May, suggesting stable isotope variability in excess of what can be explained based on environmental variability alone. The average oxygen isotope ratios track the seasonal cycle of near surface $\delta^{18}O_{eq}$ (light grey line in A and B) with an offset due to a slightly deeper calcification depth and/or a delay. Stable carbon isotopes also show a clear seasonal cycle, but with a marked offset from the $\delta^{13}C$ of DIC (grey line in C and D).

Formatted: Header

Deleted:



Moved (insertion) [2]

Moved up [2]: Green bars extend over the collection interval and show individual measurements for groups of four shells; yellow points are average values per sample. The light grey lines depict surface $\delta^{18}O_{eq}$ and $\delta^{13}C_{DIC}$; dark grey lines in A and B are $\delta^{18}O_{eq}$ at 200-250 m depth.

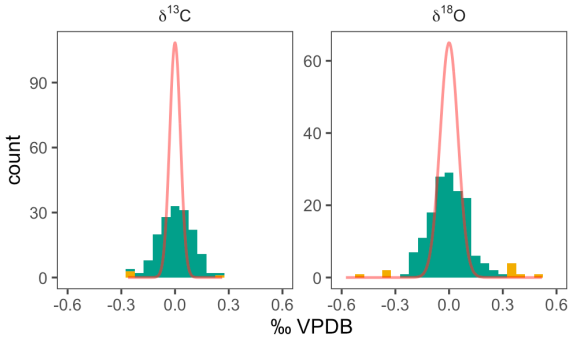
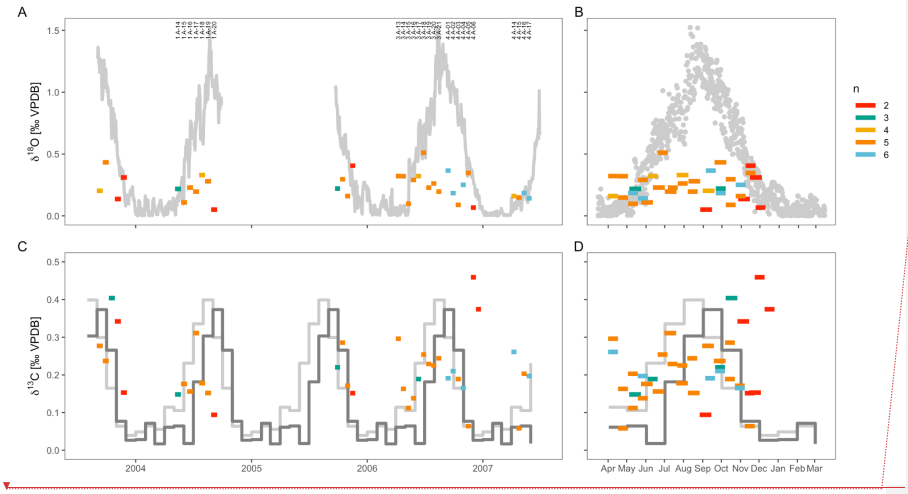


Fig. 3: Within-sample variability in *N. pachyderma* stable isotopes exceeds analytical noise. Histograms of residual $\delta^{18}O$ and $\delta^{13}C$ compared to expected density distribution if variability were due to analytical uncertainty alone (red line). Yellow colours indicate outliers (see methods).

Formatted: Header

Deleted: <*object*>

Formatted: Header



Deleted:

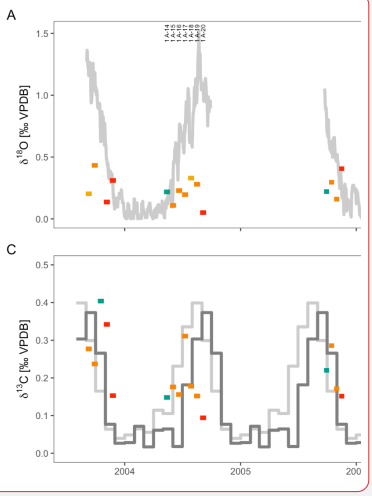


Fig. 4: The within-sample stable isotope range of *N. pachyderma* exceeds expected variability in spring when water column conditions are homogeneous and shows no consistent seasonal pattern. Note difference scales for $\delta^{18}O$ and $\delta^{13}C$. Bars extend to the collection intervals, colours indicate number of measurements per sample. Grey colours in A and B depict the difference in $\delta^{18}O$ between the surface and 200-250 m water depth. Light grey lines in C and D show the seasonal range in $\delta^{13}C_{DIC}$ and dark grey lines the seasonal range in foraminifera $\delta^{13}C$ calculated using a temperature-dependent offset from $\delta^{13}C_{DIC}$ (see methods). Samples for which the $\delta^{18}O$ variability is simulated (Fig. 5) are indicated in A.

Formatted: Header

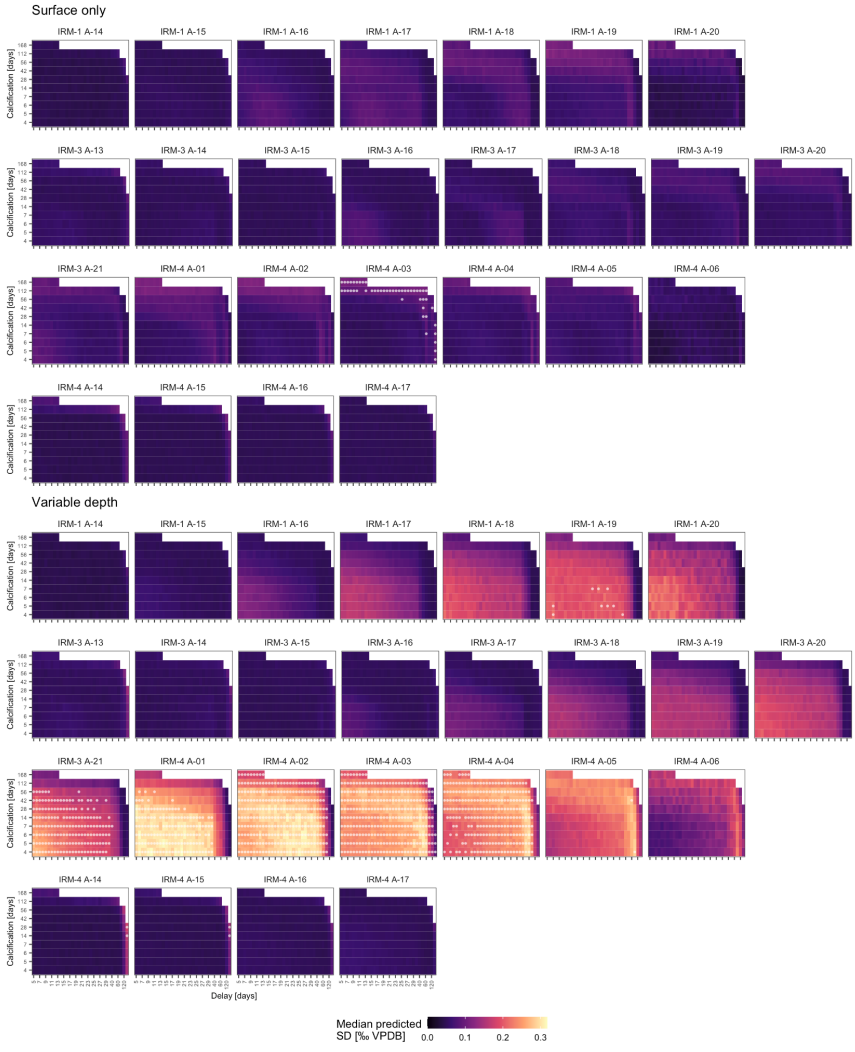
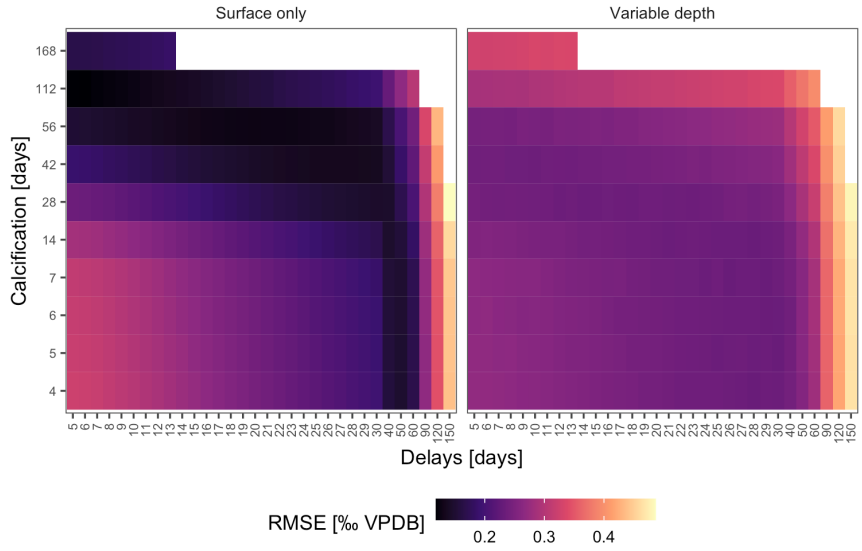
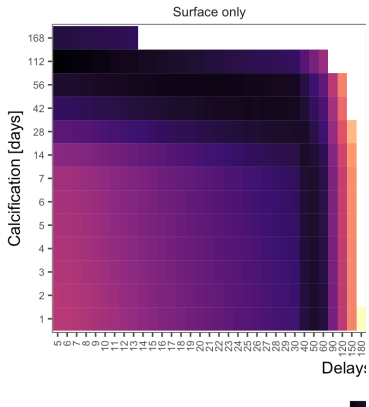


Fig. 5: Observed $\delta^{18}O$ variability in *N. pachyderma* generally exceeds expectations. Simulated $\delta^{18}O$ variability as a function of calcification span and delay for the surface only and variable depth scenarios for each sample indicated in Fig. 4. White dots indicate scenarios where the simulated variability significantly exceeds the observed variability, note that this only occurs when a variable calcification depth is assumed. Samples are ordered by year (with two rows for the 2005 - 2006 period), such that springtime samples are shown on the left. Note that for clarity x axis ticks and labels are only shown for every second tick, all steps are shown in Fig. 6.

Deleted:

Formatted: Header

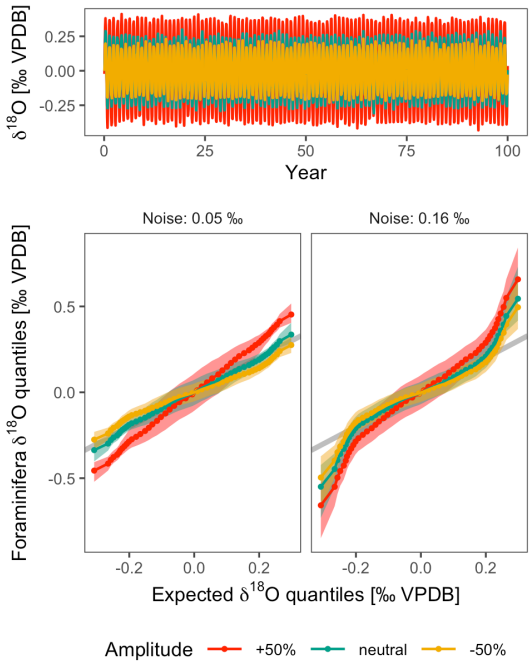


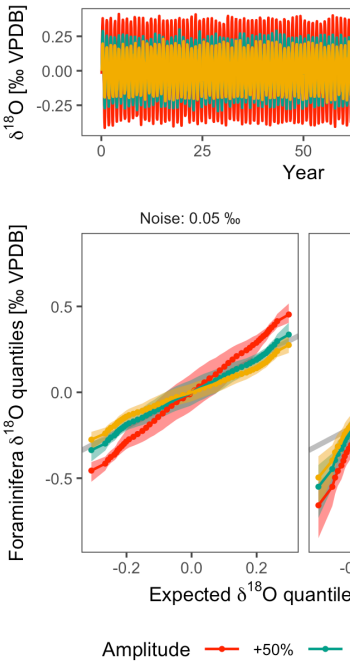


Deleted:

Fig. 6: Mean foraminifera $\delta^{18}O$ constrains simulations. Prediction errors for sample mean $\delta^{18}O$ reach markedly lower values for the surface-only simulations, indicating that this scenario is more likely to characterise *N. pachyderma* in the Irminger Sea. This means that the observed variability (Fig. 4) is unlikely a reflection of temperature and $\delta^{18}O_{seawater}$ variability alone and that the $\delta^{18}O$ of individual *N. pachyderma* shells is not a precise indicator of environmental conditions during calcification.

Formatted: Header





Deleted:

Fig. 7: Excess $\delta^{18}O$ variability mostly affects tails of $\delta^{18}O$ distribution within individual foraminifera. This simple simulation shows the effect of excess variability on capability to reconstruct changes in the amplitude of the seasonal cycle. The input consists of a synthetic $\delta^{18}O_{eq}$ time series with a seasonal amplitude of 0.25 ‰ that is not atypical of conditions in the central equatorial Pacific. The monthly time series is constructed using a sine wave with 0.02 ‰ random noise and is sampled 100 times at random to crudely represent planktonic foraminifera $\delta^{18}O$. This is an optimistic scenario as fewer foraminifera are usually used for IFA. The Q-Q plots show the effect of a change in the seasonal amplitude of $\delta^{18}O_{eq}$ for a scenario that only accounts for analytical noise (assumed to be 0.05 ‰) and for another that incorporates the excess variability found in this study. Higher noise levels affect the tails of the distribution and make it harder to detect changes in the seasonality.

References

Bé, A. W. H., Hemleben, C., Anderson, O. R., and Spindler, M.: Chamber formation in planktonic foraminifera, Micropaleontology, 25, 294–307, 1979.

Bemis, B. E., Spero, H. J., Bijma, J., and Lea, D. W.: Reevaluation of the Oxygen Isotopic Composition of Planktonic Foraminifera: Experimental Results and Revised Paleotemperature Equations, Paleoceanography, 13, 150–160, 1998.

Davis, C. V., Fehrenbacher, J. S., Hill, T. M., Russell, A. D., and Spero, H. J.: Relationships Between Temperature, pH, and Crusting on Mg/Ca Ratios in Laboratory-Grown *Neogloboquadrina* Foraminifera, Paleoceanography, https://doi.org/10.1002/2017PA003111, 2017.

Davis, C. V., Livsey, C. M., Palmer, H. M., Hull, P. M., Thomas, E., Hill, T. M., and Benitez-Nelson, C. R.: Extensive morphological variability in asexually produced planktic foraminifera, Science Advances, 6, eabb8930, 2020a.

Davis, C. V., Fehrenbacher, J. S., Benitez-Nelson, C., and Thunell, R. C.: Trace Element Heterogeneity Across Individual Planktic Foraminifera from the Modern Cariaco Basin, Journal of foraminiferal research, 50, 204–218, 2020b.

Dueñas-Bohorquez, A., da Rocha, R. E., Kuroyanagi, A., de Nooijer, L. J., Bijma, J., and Reichart, G.-J.: Interindividual variability and ontogenetic effects on Mg and Sr incorporation in the planktonic foraminifer *Globigerinoides sacculifer*, Geochimica et cosmochimica acta, 75, 520–532, 2011.

Elderfield, H. and Ganssen, G.: Past temperature and $\delta^{18}O$ of surface ocean waters inferred from foraminiferal Mg/Ca ratios, Nature, 405, 442–445, 2000.

Erez, J. and Luz, B.: Temperature control of oxygen-isotope fractionation of cultured planktonic foraminifera, Nature, 297, 220–222, 1982.

Fehrenbacher, J., Marchitto, T., and Spero, H. J.: Comparison of laser ablation and solution-based ICP-MS results for individual foraminifer Mg/Ca and Sr/Ca analyses, Geochemistry, Geophysics, Geosystems, 21, https://doi.org/10.1029/2020gc009254, 2020.

Fehrenbacher, J. S., Russell, A. D., Davis, C. V., Gagnon, A. C., Spero, H. J., Cliff, J. B., Zhu, Z., and Martin, P.: Link between light-triggered Mg-banding and chamber formation in the planktic foraminifera *Neogloboquadrina dutertrei*, Nature communications, 8, 15441, 2017.

Ganssen, G. M., Peeters, F. J. C., Metcalfe, B., Anand, P., Jung, S. J. A., Kroon, D., and Brummer, G. J. A.: Quantifying sea surface temperature ranges of the Arabian Sea for the past 20 000 years, Climate of the Past, 7, 1337–1349, 2011.

Glaubke, R. H., Thirumalai, K., Schmidt, M. W., and Hertzberg, J. E.: Discerning changes in high-frequency climate variability using geochemical populations of individual Foraminifera, Paleoceanography and paleoclimatology, 36, https://doi.org/10.1029/2020pa004065, 2021.

Greco, M., Jonkers, L., Kretschmer, K., Bijma, J., and Kucera, M.: Depth habitat of the planktonic foraminifera *Neogloboquadrina pachyderma* in the northern high latitudes explained by sea-ice and chlorophyll concentrations, Biogeosciences , 16, 3425–3437, 2019.

Groeneveld, J., Ho, S. L., Mackensen, A., Mohtadi, M., and Laepple, T.: Deciphering the Variability in Mg/Ca and Stable Oxygen Isotopes of Individual Foraminifera, Paleoceanography and paleoclimatology, 34, 755–773, 2019.

Haarmann, T., Hathorne, E. C., Mohtadi, M., Groeneveld, J., Kölling, M., and Bickert, T.: Mg/Ca ratios of single planktonic foraminifer shells and the potential to reconstruct the thermal seasonality of the water column, Paleoceanography, 26, PA3218, 10.1029/2010pa002091, 2011.

Healy-Williams, N.: Stable isotope differences among morphotypes of *Neogloboquadrina pachyderma* (Ehrenberg): implications for high-latitude palaeoceanographic studies, Terra nova, 4, 693–700, 1992.

Formatted: Header

Formatted: Line spacing: single

Deleted: J. Foraminiferal Res.,

Deleted: Geochim. Cosmochim. Acta

Deleted: Geochem. Geophys. Geosyst.,

Deleted: Ganssen, G. M., Peeters, F. J. C., Metcalfe, B., Anand, P., Jung, S. J. A., Kroon, D., and Brummer, G. J. A.: Quantifying sea surface temperature ranges of the Arabian Sea for the past 20 000 years, Clim. Past, 7, 1337–1349, 2011.¶
Glaubke, R. H., Thirumalai, K., Schmidt, M. W., and Hertzberg, J. E.: Discerning changes in high-frequency climate variability using geochemical populations of individual Foraminifera, Paleoceanogr. paleoclimatology, 36, https://doi.org/

Formatted: Line spacing: single

Deleted: Paleoceanogr Paleoclimatol

Deleted: Neogloboquadrina pachyderma (Ehrenberg): implications for high-latitude palaeoceanographic studies, Terra Nova, 4, 693–700, 1992.

Deleted: ¶
de Jong, M. F., van Aken, H. M., Våge, K., and Pickart, R. S.: Convective mixing in the central Irminger Sea: 2002-2010, Deep Sea Res. Part I, 63, 36–51, 2012.¶

de Jong, M. F., van Aken, H. M., Våge, K., and Pickart, R. S.: Convective mixing in the central Irminger Sea: 2002-2010, Deep Sea Research Part I: Oceanographic Research Papers, 63, 36–51, 2012.

Jonkers, L., Brummer, G.-J. A., Peeters, F. J. C., van Aken, H. M., and De Jong, M. F.: Seasonal stratification, shell flux, and oxygen isotope dynamics of left-coiling *N. pachyderma* and *T. quinqueloba* in the western subpolar North Atlantic, Paleoceanography, 25, PA2204; 10.1029/2009PA001849, 2010.

Jonkers, L., van Heuven, S., Zahn, R., and Peeters, F. J. C.: Seasonal patterns of shell flux, $\delta^{18}O$ and $\delta^{13}C$ of small and large *N. pachyderma* (s) and *G. bulloides* in the subpolar North Atlantic, Paleoceanography, 28, 164–174, 2013.

Jonkers, L., Buse, B., Brummer, G.-J. A., and Hall, I. R.: Chamber formation leads to Mg/Ca banding in the planktonic foraminifer *Neogloboquadrina pachyderma*, Earth and planetary science letters, 451, 177–184, 2016.

Jonkers, L., Gopalakrishnan, A., Weßel, L., Chiessi, C. M., Groeneveld, J., Monien, P., Lessa, D., and Morard, R.: Morphotype and crust effects on the geochemistry of *Globorotalia inflata*, Paleoceanography and paleoclimatology, https://doi.org/10.1029/2021pa004224, 2021.

Kim, S.-T. and O'Neil, J. R.: Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates, Geochimica et cosmochimica acta, 61, 3461–3475, 1997.

Kohfeld, K. E., Fairbanks, R. G., Smith, S. L., and Walsh, I. D.: *Neogloboquadrina pachyderma* (sinistral coiling) as Paleoceanographic Tracers in Polar Oceans: Evidence from Northeast Water Polynya Plankton Tows, Sediment Traps, and Surface Sediments, Paleoceanography, 11, 679–699, 1996.

Leduc, G., Vidal, L., Cartapanis, O., and Bard, E.: Modes of eastern equatorial Pacific thermocline variability: Implications for ENSO dynamics over the last glacial period, Paleoceanography, 24, PA3202, 2009.

Livsey, C. M., Kozdon, R., Bauch, D., Brummer, G.-J. A., Jonkers, L., Orland, I., Hill, T. M., and Spero, H. J.: High-Resolution Mg/Ca and $\delta^{18}O$ Patterns in Modern *Neogloboquadrina pachyderma* From the Fram Strait and Irminger Sea, Paleoceanography and Paleoclimatology, 35, e2020PA003969, 2020.

McConnaughey, T.: ^{13}C and ^{18}O isotopic disequilibrium in biological carbonates: I. Patterns, Geochimica et cosmochimica acta, 53, 151–162, 1989.

Metcalfe, B., Lougheed, B. C., Waelbroeck, C., and Roche, D. M.: A proxy modelling approach to assess the potential of extracting ENSO signal from tropical Pacific planktonic foraminifera, Climate of the Past, 16, 885–910, 2020.

de Nooijer, L. J., Hathorne, E. C., Reichart, G. J., Langer, G., and Bijma, J.: Variability in calcitic Mg/Ca and Sr/Ca ratios in clones of the benthic foraminifer *Ammonia tepida*, Marine micropaleontology, 107, 32–43, 2014.

Rustic, G. T., Koutavas, A., Marchitto, T. M., and Linsley, B. K.: Dynamical excitation of the tropical Pacific Ocean and ENSO variability by Little Ice Age cooling, Science, https://doi.org/10.1126/science.aac9937, 2015.

van Sebille, E., Scussolini, P., Durgadoo, J. V., Peeters, F. J. C., Biastoch, A., Weijer, W., Turney, C., Paris, C. B., and Zahn, R.: Ocean currents generate large footprints in marine palaeoclimate proxies, Nature communications, 6, https://doi.org/10.1038/ncomms7521, 2015.

Simstich, J., Sarnthein, M., and Erlenkeuser, H.: Paired $\delta^{18}O$ signals of *Neogloboquadrina pachyderma* (s) and *Turborotalita quinqueloba* show thermal stratification structure in Nordic Seas, Marine micropaleontology, 48, 107–125, 2003.

Spero, H. J. and Lea, D. W.: Intraspecific stable isotope variability in the planktic foraminifera *Globigerinoides sacculifer*: Results from laboratory experiments, Marine micropaleontology, 22, 221–234, 1993.

Spero, H. J. and Lea, D. W.: Experimental determination of stable isotope variability in *Globigerina bulloides*: implications for paleoceanographic reconstructions, Marine micropaleontology, 28, 231–246, 1996.

Spero, H. J., Bijma, J., Lea, D. W., and Bemis, B. E.: Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes, Nature, 390, 497–500, 1997.

Spindler, M.: On the salinity tolerance of the planktonic foraminifer *Neogloboquadrina pachyderma* from

Formatted: Header

Formatted: Line spacing: single

Deleted: Planet. Sci. Lett.,

Deleted: Paleoceanogr.

Deleted: Geochim. Cosmochim. Acta

Deleted: Geochim. Cosmochim. Acta

Deleted: Clim.

Deleted: Mar. Micropaleontol.,

Deleted: Nat. Commun.,

Deleted: Mar. Micropaleontol.,

Deleted: Mar. Micropaleontol.,

Deleted: Mar. Micropaleontol.,

Deleted: salnity

Deleted: foraminfer

Antarctic sea ice, Proceedings of the NIPR Symposium on Polar Biology, 9, 85–91, 1996.

Takahashi, K. and Bé, A. W. H.: Planktonic foraminifera: factors controlling sinking speeds, Deep-sea research. Part A, Oceanographic research papers, 31, 1477–1500, 1984.

Thirumalai, K., Partin, J. W., Jackson, C. S., and Quinn, T. M.: Statistical constraints on El Niño Southern Oscillation reconstructions using individual foraminifera: A sensitivity analysis, Paleoceanography, 28, 401–412, 2013.

Formatted: Header

Deleted: Sea Res.