Comment on cp-2021-87
Anonymous Referee #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.

(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 d18O seawater 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).

(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 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?

(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???

(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’.

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”.

**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.

**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’.

**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.

**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.

**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.

**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.

**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.

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

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

**Line 127:** I think it shall read “regressions” (plural)

**Line 130:** what does “available as climatology” mean? Same line: Use “spatial resolution” instead of “same level of detail”?

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

**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.

**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'?

**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

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

**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”...

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

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

**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”

**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..

**Line 230:** suggestion: “...regarding these initial observations…”

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

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

**Line 240:** prior to the sampling

**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

**Line 246:** for clarity: please mention again: what are the two scenarios? (1) Variable
calcification depth, and (2) calcification during summer?

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

**Line 286:** suggestion: "when variations in temperature and...“

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

**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
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.

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.

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.

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 inaccuracies 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...)

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.

"...that has so far been..."

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?

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 d18Oequilibrium.

References:


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