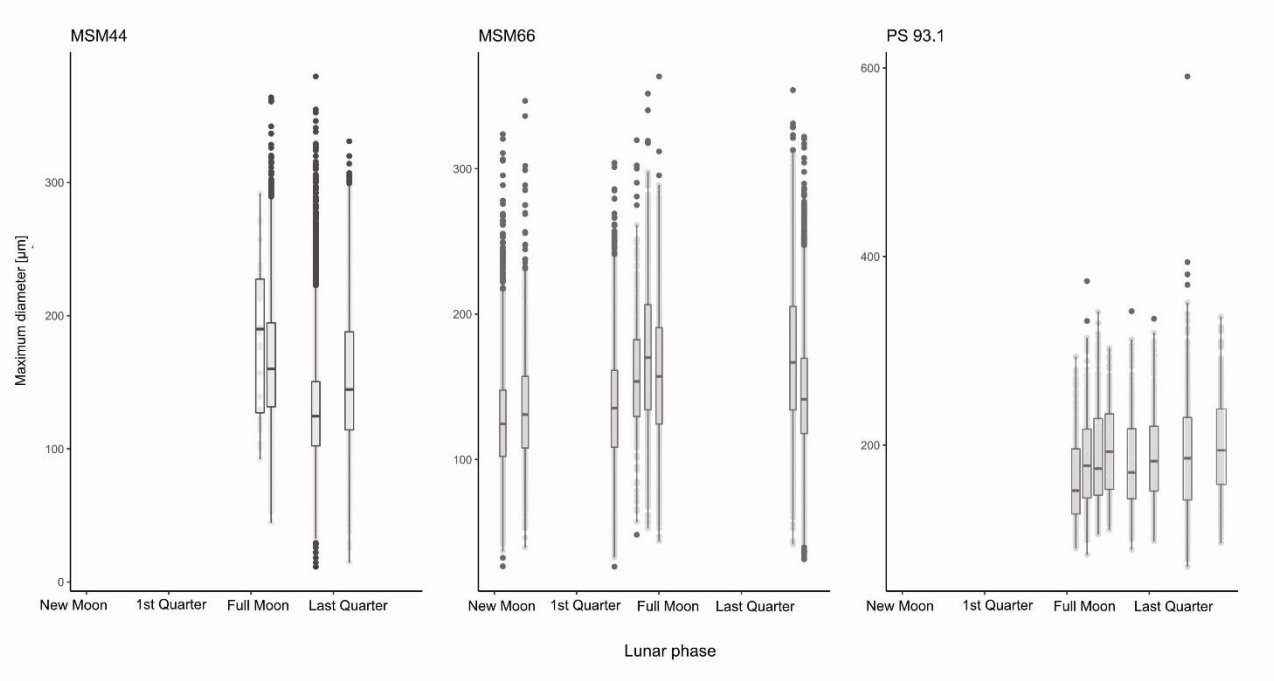
Reply to **RC2**: 'Comment on bg-2022-59', Ralf Schiebel, 13 Jun 2022
Please find our detailed responses to all raised aspects (mainly in the annotated pdf file) and how to incorporate revisions in an updated version of the manuscript below, written in italics.

Overall, the manuscript on by Tell and coauthors reads goods, and adds useful information on the population dynamics and flux of N. pachyderma. In addition to the comments and suggestion made by another reviewer, Robert Spielhagen, I would suggest amendments, which I have detailed in an annotated pdf file I send along with this review. In particular, wording needs to be revised in places. Some of the figures are not up to publication standard, and labels need to be revised and/or added. Most importantly, the manuscript needs to be rearranged in places, and some sections need to be moved from the “Results” to the “Discussion”. I would also suggest to analyze the data on test size for any relationship with the synodic lunar cycle, which would strongly support (or not) the population dynamics and pulsed flux discussed in the manuscript.

Thank you very much for the suggestion to analyse the possible linkage between foraminifera test size and the lunar cycle. As a large part of the analysed data is from a region with polar day throughout the main growing period, we have not considered an effect from the lunar cycle thus far. However, since in other parts of the ocean this effect has been documented, and we cannot exclude it upfront, we have had a closer look at a selection of the observations that would be suitable to detect this effect. Specifically, we tested the relationship between shell size and lunar day using the size throughout all depth intervals, divided into the different time intervals and regions of sampling, represented by the three different cruises where size data have been generated. See the figure below to see the spread of all values, represented as box plots with the horizontal bar in the middle giving the median value of the size measurement, plotted against the lunar day. The data from PS93.1 represents sizes of filled (living) shells. We do not have the data on shell status in connection to the size measurements from MSM44 and MSM66, so sizes of all shell types are used there. Unfortunately, none of the analysed data sets covers a whole lunar cycle. In connection to the possibility of superimposed spatial shell size variability also within the sampled regions, drawing conclusions from patterns in sizes on the impact of the lunar cycle across samples from different regions and years would be too difficult, if not impossible, so we feel it is more appropriate to remain with the presented analysis. These provide no evidence for a strong (detectable) systematic change in size with lunar day, indicating that the reproduction of the studied species was either not synchronised or that only a small part of the population reproduced in synchrony.



In a revised version of the manuscript, we will describe this analysis in the results section on the shell size, stating that analysis in connection to the lunar cycle, as it was done in previous studies as shown in Schiebel et al. (2017), is strongly limited due to the data at hand and, in the portion of the data where it was possible, did not allow us to draw a conclusion in terms of a connection between shell size and the lunar cycle.

L. 7: that's a very anthropocentric point of view, and varies between species
We will instead write "could become less challenging to pelagic calcifiers,".

L. 44: and biological carbon pump
We will add this.

L. 52: deeper than what?
We will add "than the productive zone", to clarify what "deeper" refers to.

L. 53: hm, maybe or maybe not; this is not a scientific argument, and I would suggest to skip it
We will change the sentence accordingly: Also, sediment trap records are scarce in the Arctic.

L. 68: and? ... for the sedimentary assemblage while non-encrusted tests have been removed by dissolution...?
We will change the second half of the sentence as follows to make this more precise: "because most individuals add a crust and because encrusted shells are more resistant to dissolution

L. 72: How do you define a "crust"? Gametogenetic calcification?
We will add in the methods that the crust is defined based on their (larger) weight and the shell texture, as shown in Fig. 2.

L. 86: an individual?
We here refer to overall abundances, as we cannot follow individuals by sampling. We understand that it is not clear from the current sentence, and will change it accordingly (changes marked in bold): "To distinguish the production and export zones ***and to determine*** *the* ***average*** *depth of calcification* ***of N. pachyderma****, …"*

L. 104: processed two times? processed for what?
We agree that it is unclear and will change the wording and write more precisely: "Samples from the Baffin Bay were either processed on board or stored at -80°C until processed onshore."

L. 106: usually, the minimum diameter is assessed, because it is comparable to sieve-size intervals
We unfortunately only have the measurement of the maximum diameter for the samples from the Baffin Bay. We agree that using the maximum diameter does not enable a direct comparison of absolute values to sieve-size intervals. The most important aspect, nevertheless, is to not mix different types of size parameters in one analysis, which we do not do. We are therefore confident that our conclusions are valid. We propose to stick with the parameter as done in the submitted version of the manuscript to be able to work with the datasets together.

L. 111-112 / Fig. 2 (f): To my impression, the specimen shown in (f) has not been alive when sampled, and Rose Bengal staining may not be suited for live-staining of planktic foraminifers.
Please see the paper of Lutze and Altenbach (1991) Technik und Signifikanz der Lebendfärbung benthischer Foraminiferen mit Bengalrot.

We are aware of the fact that rose Bengal might be staining recently dead specimens because of cell degradation (see Schönfeld et al., 2014). But visually separating shells based on the presence or absence of cytoplasm without staining also leads to some ambiguity. This is because recently dead foraminifera can still contain cytoplasm. Therefore, the proportion of dead foraminifera represents in most cases lower bound/slight underestimation. This also becomes clear by our results shown in the manuscript in fig. 3a. We will include remarks on this in the method section (new part marked in bold):
l. 105-106: "The counts were made separately for cytoplasm-bearing shells and empty shells, differentiated during the processing of the wet samples. ***As recently dead foraminifera can still contain cytoplasm, this leads to a bias in the numbers in favour of shells interpreted as being alive upon sampling.****"*
l. 110-112: "In accordance with data from earlier studies, fully white shells were classified as dead ("empty") (e.g. Fig. 2e), all other (pink) shells as cytoplasm-bearing (e.g. Fig. 2f), assumed to represent specimens that were alive during retrieval. ***As rose Bengal might be staining recently dead specimens because of remaining cytoplasm in the shells (Schönfeld et al., 2014), there is a possible bias towards too high numbers of cytoplasm-bearing shells.****"*

L. 160: 37 profile is a lot, and a different method than Loncaric's may be applied for a more representative result
It is a valid point that the method has disadvantages. We nevertheless suggest keeping it, as we think that our overall estimates on the BPZ are satisfying enough to see what changes are present within the productive zone and where it ends to be able to calculate carbonate fluxes below. We will add a sentence in the manuscript (l. 163-164) to make the disadvantages clear: "For those 40 profiles, the BPZ was defined as the bottom depth of the transition zone (Fig. 2a, Z_{BPZ} (range end)). ***This can result in a bias towards the estimated BPZ being located below the actual position. This bias is restricted by the overall sampling interval (median: 50 m) and has no effect on our flux estimates which are based on average shell abundances below the BPZ.***

L. 180 ff: this section reads like a Discussion, not Materials and Methods
The intention of this section is to explain why we investigated the named parameter and what aspects we looked for in our analysis, which is why we put it into the methods section. We think it is necessary to have it in this section to help the reader understand the purpose of the analysis. Therefore, we would like to keep it at its current position.

L. 198/199: this may not possibly give realistic results, because the population dynamics and flux is too dynamic and too fast
We agree and are aware of the fact that using average values could be rather far from reflecting the actual situation. Nevertheless, as long as no further measurements are available, we think this is the simplest way to estimate mass fluxes in different regions of the Arctic realm. We will clarify in the manuscript that this method is likely to underestimate variability.

L. 220: subsurface is possibly not defined by depth alone, but by surface water stratification, i.e., subsurface under the seasonal pycnocline
We agree, but we are here not referring to a defined subsurface in terms of the water masses. We just intend to express that some profiles show a maximum directly at the 0-50 m depth interval, and some only further below (50-150 m depth). To avoid confusion, we will change the sentence as follows: "Shell ***abundances*** *show either a* ***maximum within the***

upper 50 m, or in the depth zone below reaching down to 150 m *(exemplarily shown in Fig. A1)."*

L. 274: is this rounding, or is it 0.010 ? Please use the same number of digits for the same purpose
Thank you, it is 0.010, we will change that.

L. 281 do you mean "less lobate"?
Yes, we will write this in the text.

L. 285 in "roundness"?
Yes; we can add this clarification.

Fig.7: change labels of both x- and y-axis to meaningful information
We would like to keep the label of the y-axis as it is, because another label could be even more confusing. However, we understand the point and will therefore in the caption clarify even more what is shown here, explaining that in this situation, 100 % equals the productive zone and 50 % equals half of the depth of the productive zone.
The x-axis in the plot, as proposed by Robert Spielhagen, will be changed such that actual values are shown but plotted on a logarithmic scale so they can be interpreted directly.

L. 315 in what?
A decrease in mass flux, we will specify in the text.

L. 317: what do these numbers show?
As specified in the brackets where the first time several values are given (in line 313-314), the first value is based on encrusted/empty shells only and the other on weights of non-encrusted/filled shells only. For clarity, we will repeat this information in the manuscript.

L. 320: this not "Results" but "Discussion"
We will delete this part here.

Fig. 8: What does "Count" include? Any unit? Please add numbers to the "open ends" of the x- and y-axis. The "floating bar" at the very left end of the x-axis looks strange. Please think about re-designing the figure.
We will change the label to "Number of observations" (it is not the count of foraminifera, but sampled stations).
Regarding the open ends of the plot, we will increase the number of breaks (as also suggested by Robert Spielhagen) and make sure that there are labels at the end of the plot.

Fig. 9:
- "Change" of what?
Change in mass flux in %/100 m (as indicated on the x-axis of 9b, we will also write it on top of 9a). We will delete the legend for the coloured dots and include the information in the caption to decrease the information in the figure and therefore increase its readability.
- What are the scales and units of the dots and boxes in the top panel?
*The box plots relate to the scale below which is also the scale for the scatter plot; we will add this explanation in the figure caption. As it was also raised in the review by Robert Spielhagen, we will also implement in the caption that the line represents the median and the points values outside the 1.5*IQR (outliers).*
is this the orange color?
Yes, "coral" is referring to the orange-like colour. We will change it to "orange" and "blue" to make it clear.

Fig 10: what is the width of the intervals?
The width of depth that was sampled in the nets differs among all the stations included, as shown in Tbl. 2 for the added profiles and for all others in the linked references. We will clarify this in the caption.

Fig 12: how realistic are residence times of >80 days?
It is not possible for us to exactly predict what residence times are really realistic for N. pachyderma in the Arctic realm. Based on culture experimentations (Spindler, 1996), however, we know that the genotype from the Southern Ocean can live several months and up to more than a year. It is equally not implausible that the extremely long implied residence times are a product of partly unprecise flux estimations. We will add a comment regarding the range of the residence times in the discussion section. We will also make clear that the lower end has to be interpreted considering that we analyse a lot of shells that look like they have reached their final life stage already. This means that the few days of residence only represent the time they spent in the water as "mature" organisms. We will change the paragraph accordingly (change marked in bold):
l. 457ff: "Nevertheless, the estimated residence time of about 4 days in our data indicated that the life span of the sampled N. pachyderma is either too short to be strongly affected by environmental condition changes, or that the population size is constant at least across a short time scale. The latter would make huge changes in the environmental conditions unlikely. ***When interpreting the lower end of estimated residence time, it has to be considered that this might not represent the overall lifespan of the foraminifera, but the days it stays alive after having reached maturity, which is what, based on optical parameters, we mainly analyse. Based on results from culture experiments on N. pachyderma from the Southern Ocean, a lifespan of several months, the upper end of our estimated residence time, seems to be possible (Spindler, 1996). It is equally possible that unprecise flux estimations in some samples lead to those extreme values. With the majority of all samples showing a residence time of only a couple of days, we*** *conclude that the possible blurring of signs of OVM would be rather small, and the lack of a clear trend indicating OVM at all stations can be seen as a reliable result. "*

L. 376ff: please discuss also the results presented by Carstens, Wefer, Volkmann, Pados, Spielhagen, Schiebel, etc
We can indeed include the results from Carstens and Wefer in the discussion. As we use all the data from the other studies in our manuscript, it is trivial to make a comparison. We will add the following sentences in the discussion section in l.377ff: "Greco et al. (2019) have shown that the habitat depth of N. pachyderma varies substantially. ***A variation in the depth interval of maximum abundances of N. pachyderma is also presented by Carstens & Wefer (1992) and Carstens et al. (1997), where a connection between distinct water masses and temperature regimes is drawn.*** *Our dataset corroborates* ***these*** *observations and indicates that the base of the productive zone of N. pachyderma is also highly variable and reflects the habitat depth (vertical distribution of living specimens). Like Greco et al. (2019), we observe that even if there would be a general pattern of habitat depth and BPZ position being driven by environmental factors,* ***as also proposed by Carstens et al. (1997),*** *it is overlain by considerable variability, even among profiles collected in the same region and around the same time."*

L. 402 by dissolution?
Yes, we will specify in the text.

L. 404-405: i.e., statistically not significant? Please show in the "Results" section!
L. 407-408: please present the statistical significance, and, if significant, discuss the results
We did not do statistical analysis exactly because of the low numbers. Therefore, we will change the wording to make clear that it is a descriptive result.

L. 409: These observations can help to calculate fluxes - please do
We will change the sentence as follows: "These observations helped us to calculate fluxes of shells of N. pachyderma from plankton net samples at a more realistic depth in the Arctic Ocean, even where the productive zone has not been explicitly constrained. It can serve as a base for further studies."

L. 413-425 what about the Schiebel et al. (2017) data on depth distribution and reproduction of N. pachyderma? Please discuss
See our comment on our analysis regarding the lunar cycle at the beginning.

L. 455-456: any proof?
There is no proof, which is why we stayed speculative with our wording: "Depending on the life span of N. pachyderma, which could be longer than one or two months (Carstens and Wefer, 1992; Kohfeld et al., 1996), it is possible that the samples contain individuals from multiple generations that were produced during different environmental conditions."

L. 457-458: do you suggest a life span of N. pachyderma of 4 days?
We understand the confusion, see our proposed addition to the discussion section here at the above stated change in connection to Fig.12.

L 483: what?
It refers to the shell weight, we will clarify in the manuscript.

L 483 ff please rephrase this paragraph to make it concise and unequivocally understandable
We will do that. We will explain that the sampling mesh size creates a bias, as a 63 µm net samples different material than a 100 µm one.

L. 495: sorry, I've got no idea what's meant here
In an idealised scenario, the % of empty (= dead) shells, and the % of encrusted (= those that have been at the clearly visible end of their life cycle) would be 100 % in the area below the productive zone. This is not the case, there are still shells that are cytoplasm-bearing and shells that do not have a thick crust present below the productive zone. Therefore, it makes sense to use the overall average to estimate fluxes instead of only using weights from the empty and encrusted ones. We will clarify the text accordingly.

L. 498: to calculate flux. (this should be movded to the "Results" chapter.
Because it was explained in the method section, we will refer to it in the text here.

L. 542: not on average. However, following deep mixing, e.g., by storms, the surface mixed layer and productive zone of planktic foraminifers, may be deeper than 100 m, and which may have a huge effect in particular during times of enhanced test production in spring. Please see Schiebel et al. (1995).
We agree, which is why we did not limit our flux calculation to 100 m. We will change the sentence accordingly to make clear that we cannot judge the effect of a different productive zone depth in the North Atlantic in terms of the comparison with this data set:
"We know that 100 m can be too shallow to estimate the fluxes in the Arctic, but cannot ***judge the effect of a possibly deeper or varying productive zone in the North Atlantic (Schiebel et al., 1995) on flux estimates."***

L. 595: this is not far away from the 100 m used as export layer by Schiebel (2002), isn't it ;-)
L. 598: if this is true, how do you arrive at a 5-fold overestimation when comparing your data to Schiebel's (2002)? I guess that you are may be right with your assumptions, but you should present it in the right way

Thank you for pointing out that the current formulation in the concluding chapter is not clear enough to well present the main results of our study.

Even though we agree that the median calculated BPZ in our analysis of 113 m is not far away from the 100 m used in Schiebel (2002), we show that it is important to make the differentiation. The abundance profiles show that there is a rapid decrease at the BPZ. Therefore, using 100 m to estimate fluxes can lead to a strong overestimation (fivefold, in our data), even if the actual BPZ is only located closely below the 100 m mark. This is why we want to point out here that for more precise flux estimation on a regional basis, it is important to either know (calculate) the BPZ, or set it lower, as we show that it can reach down to 300 m for N. pachyderma in this region.

Regarding the estimated loss with depth, this is not connected to the before stated fivefold overestimation when using a different depth to calculate fluxes. The given average loss with depth represents the loss in flux from the sampling position below the BPZ towards the deepest position of sampling. This loss is also decreasing with further depth, being highest directly below the BPZ and lowest at depth, as shown in Fig. 9.

To make those concluding results of our study clear, we will change the paragraph in the discussion as follows (as we detected an error in our calculation, the value for the change with depth will be changed from 1.5 % to 6.6% per 100 m):

"Our compilation of vertically resolved data on the dominant Arctic planktonic foraminifera N. pachyderma reveals that the base of the productive zone of this species is on ***median*** *located at 113 m depth, but shows large regional variability and locally reaches down to 300 m. Our analyses show that it is important to constrain the base of the productive zone to estimate fluxes in the export zone: using a constant 100 m depth to estimate fluxes leads to a fivefold flux overestimation* ***in contrast to the flux at the top of the export zone. Below the BPZ, the shell flux is decreasing on average by 6.6 % per 100 m, with highest losses directly below the BPZ and no further change after 300 m below the BPZ is observed. Therefore,*** *we can conclude that in the absence of knowledge on the position of the BPZ, using 300 m depth should provide a conservative, yet more realistic estimate* ***of the N. pachyderma export flux in the Arctic realm*** *than using the formerly often used depth of 100 m. Within the productive zone …"*