Béjard and colleagues present data on the calcification intensity of three species of planktonic foraminifera from the western Mediterranean Sea. They investigate intra- and inter-annual changes using an exceptionally long sediment trap time series and compare their findings with data from nearby sediments. They find that the calcification intensity variations are species and time-scale specific. They assess environmental and biotic variables that could potentially affect variations in the calcification intensity and conclude that optimum growth conditions, temperature and carbonate system parameters are most likely responsible for seasonal variation, whereas on longer time scales carbonate system parameters are more important. They also find a consistent difference in calcification intensity between sediment trap and sediment samples, which they attribute to global change.

The manuscript is interesting and timely and presents a unique data set. However, I have many questions about the analyses that need clarification. I am in particular not convinced that all the patterns and trends are significant and fear that the authors overinterpret the data. It seems that on seasonal time scales the only marked change is in the species truncatulinoides and that in the other two species the variability is much smaller and unlikely to rise above noise levels. This is in itself a really interesting observation given the environmental changes during the year, but the authors take the values at face value and attempt to correlate each seasonal pattern to environmental variability. As said, I am not convinced that this is meaningful. The same holds for the interannual variability. The observed changes are small and I am unsure if all sources of uncertainty have been considered appropriately. The authors also find a difference between the sediment trap and sediment samples, which may be significant for truncatulinoides. However, their attribution to global change, which is vague in itself, is not convincing and alternative explanations have not been convincingly ruled out. Hence, I cannot but recommend major revisions before the manuscript is suitable for publication, but I hope that the comments are helpful to improve the manuscript.

Main comments
Several aspects of the data handling are unclear and need clarification:

- Please show the raw data (individual measurements, not the mean values) and the shell fluxes before showing/analysing the intra and interannual variability. At first I was under the impression that the entire time series was analysed, but I think the authors only analysed selected samples for each species (and not the same samples for each species).
- To the best of my knowledge Rigual-Hernandez et al. (2012) presented shell fluxes >150 µm. How do the shell fluxes in the specific size fractions analysed here relate to those? In other words, how was the flux-weighting done? Based on size-specific shell flux data, or by assuming that the shell size remained constant and that weights for the calculations of the weighted averages are independent of size?
- Please provide the depths of the analysed sediment samples. I assume that the dated sediments have also been analysed, but it's not obvious from the text. Why were the only replicate analyses done on the core top (PLA CT)? What was the spread?
- Details on the statistics are scarce. What tests were used for what reason? Are they applicable (the data appear log-normally distributed) and are predictor and response variables really independent (Fig. 2, shows area against area normalised weight).
- Along the same lines, what is the meaning of the error bars in the various figures? In Fig. 3 it seems that they reflect the uncertainty (standard deviation?) of the sample averages for each month. My most important point is that this hides a large part of the variability (see Table 2) and I recommend to show the individual values (i.e. for each measured shell, normalised to average weight). In addition, how were samples assigned to a month, by start date, mid date, or by taking into account the collection interval? And are the mean and the standard deviation appropriate summary statistics for these data that are far from normally distributed? The same applies to Fig. 4? And in the latter figure, are the mean values flux-weighted?
- Why do none of the environmental variables have error bars? This is especially relevant for the analysis of the seasonal patterns.
- What is the rationale to infer linear trends in Fig. 4? And are the calcification intensity values flux-weighted? And how much of the variability is due to changes in the shell flux?
- Please provide the data in accordance with the Copernicus data policy: https://www.biogeosciences.net/policies/data_policy.html (preferrably also for review)

Additional points:

Separate analysis of intra and inter-annual variability. The authors first analyse the intra-annual variability in calcification intensity and then proceed with the inter-annual. What this means is that some of the interannual variability is incorporated in the composite seasonal cycle, perhaps obscuring or suggesting trends that are not there. Why not model the calcification intensity as a function of month/day of the year and time? This will allow obtaining more meaningful estimates of the seasonal cycle and the long-term trend and can for instance be done using GAMs.
Significance of the trends/patterns. Irrespective of the approach the authors take (although I recommend modelling the seasonal and interannual data together) the authors need to establish whether or not the trends/differences are significant. I am not convinced this is the case for any of the time scales considered (seasonal, interannual and decadal/centennial; with the possible exception of G. truncatulinoides on intra-annual and centennial time scales) and the authors run the risk of trying to explain noise (and derive meaningless conclusions) in the sections where they attempt to explain what drives variability in calcification intensity.

Drivers of variation in calcification intensity. I agree with the authors that identifying the exact drivers of calcification intensity is not trivial because of collinearity in the environmental variables. However, there are ways to better handle this problem than assessing the correlation of each variable individually. I can think of multiple linear regression or more flexible GAMs.

Differences among species. One of the most interesting observations is that different species show different patterns in the calcification intensity (the latter still needs to be demonstrated and I suspect that only G. truncatulinoides shows a significant pattern on seasonal time scales). Right now it is hidden in the text, but it would be interesting to discuss in greater depth and with more focus why this might be the case and make a clearer link to the ecology of the species. The relationship between calcification intensity and shell flux is particularly intriguing in this respect and the authors go some way to discussing this for truncatulinoides (allocation of energy), but not for the other species. I recommend that they do.

Related to this is the concept of optimum growth conditions and whether shell flux (mortality in fact for sediment traps) is a good measure (perhaps not), which should be discussed in greater depth. Perhaps the relative abundance of a species is a better indication of optimum growth conditions? Or the first derivative of the shell flux/abundance as the fastest growth occurs when the population size increases most.

Differences among time scales. Another intriguing finding is that the environmental drivers of calcification intensity seem to vary with time scale. I am not sure if this is a robust finding (see above), but at face value the calcification intensity of truncatulinoides appears positively correlated with temperature and carbonate ion concentration on seasonal time scales, negatively correlated on interannual time scales and on longer time scales positively with temperature and negatively with CO32-. If real, why could this be the case? The authors hardly touch one this, but it is interesting as it suggests that different mechanisms than considered might be responsible.

Ages of the core tops. There are several issues with paragraph 4.4 where the calcification intensity from the sediment trap time series is compared to the core top data. First of all, it is unclear what is actually compared. Fig. 6, which is referred to in this section, shows the mean (I presume) values for each sample (cup or depth), not the flux-weighted calcification intensity from the sediment trap time series and the core top data as is suggested in the text (which would mean 1 data point from the trap and one for each core
top). This needs to be clarified. It makes of course most sense to compare what the authors write (but they need to explain how exactly they derive the flux-weighted average, right now it is unclear if this is based on the monthly averages or on the entire time series), but the figure shows something else. Notwithstanding the calculation, it is interesting that the differences are so consistent, I think the authors should emphasise that more. Second, the authors have not calibrated the 14C ages to calendar years. This means that the age/durations mentioned in the text are simply wrong. The authors need to correct for the local reservoir effect and calibrate the ages before they can make any statements about ages or rates. I realise that this does not affect the difference in calcification, but it is nevertheless important for the interpretation (as the core tops might on average not be post-industrial, but see below). Thirdly, the 14C ages are average values only. Due to time integration within each sample and because of bioturbation the real ages of the foraminifera vary (Dolman et al., 2021; Lougheed and Metcalfe, 2021). The authors need to acknowledge this in their interpretation.

Attribution of difference with sedimentary shells. With regards to the line of argumentation to exclude the effect of dissolution I disagree with the authors, particularly on the third point (“if partial dissolution was to take place here, MBWs from the seabed sediments and core tops would be lighter than the ones from the sediment traps” L762-764). (Seafloor) Dissolution tends to affect the thinnest shells first, leaving the sediment enriched in thick/more heavily calcified shells (Berger, 1970 and many more studies since), thus exactly what the authors observe. The difference in calcification intensity thus seems to be in agreement with dissolution affecting the sedimentary shells instead of providing evidence against it.

Similarly, the fourth argument (“SEM observations of all 3 species in samples from the sediment traps showed no sign of dissolution and foraminifera were well preserved” L765) is also unconvincing. Shells spend orders of magnitude more time on the sediment-water interface than in the water column and dissolution in the water column is therefore less of a problem. The authors therefore need SEM images of the sedimentary shells to prove that they were not affected by dissolution. Ideally, they also include SEM pictures of cracked sedimentary shells to demonstrate that the cleaning was sufficient to remove any sediment from within the shells.

In addition, can the authors rule out that the difference in calcification intensity is not due to changes in the seasonality of the flux (because of the seasonal variability in calcification intensity, it is after all possible to change the flux-weighted mean calcification intensity by just changing the flux; cf de Moel et al., 2009) or due to addition of calcite below the depth of the trap (G. truncatulinoides)? So, even if the difference in calcification intensity between the core top and the sediment trap shell is significant (which needs to be demonstrated first), there are still other options than ocean acidification (based on the seasonal pattern, temperature seems an obvious candidate) that could at least in theory explain the difference. The authors therefore need to provide more convincing arguments to support their conclusions.

The authors may want to include the paper by Weinkauf et al (Weinkauf et al., 2016) in their discussion.
Finally, the text needs tidying up and should be checked for spelling and grammar. The figures also need to be improved, I have provided suggestions below.

**Specific comments**

The title does not seem consistent with the objective as specified in the first sentence of the abstract.

L59: the industrial period started about 170 years ago (although somewhere else in the text the authors suggest it started around 1800 CE).

L60: unprecedented for what time frame?

L77: Jonkers et al did not discuss ocean acidification.

L91: “bearing” not “wearing”. I think the word depending here is also ambiguous, it may suggest that ecology and feeding strategy evolved first and that the species added symbionts or spines. If anything, I think it was the other way around. Better phrase this neutrally.

L113-124: Is this section needed here? A lot of the information is also presented in the section “study area”

L124: “MedECC”?

L129: insert “can” before “provide”, short (< 1 year) deployments cannot provide annually integrated fluxes.

L146-193: can this not be condensed to the information that is relevant to the study?

Fig. 1: the line representing NWM lacks an arrow to indicate the direction of the flow. It also seems oddly placed as if it connects Minorca with Sicily.
L206: “the characterisation and quantification” … “were analysed” reword

Table 1: provide reservoir age (R or delta R) and calibrated age (include details on calibration in the methods)

L237-241: please be clearer that the number of unique samples processed (Table 2) is lower. Also explain why species were analysed in size fractions, it is not immediately obvious since the size is also measured, and that this may lead to an underestimation of the variability in calcification intensity in the entire population of the planktonic foraminifera species.

L242-247: please explain why this number of shells is sufficient to characterise the variability/mean within a sample.

L267-270: were the morphometric analyses done on the same shells that were weighted (i.e. the 10-30 shells in specific size fractions)?

L275: I understand the goal of this normalisation, but please describe better what the difference is between “mean parameter size fraction” and “mean parameter sample”. And what is the advantage of this method compared to the area density (they seem highly correlated).

L288: SeaWiFS started in 1997, please mention this and explain why you analyse composites instead of analysing each sample using the corresponding environmental variable.

Table 2: is standard deviation meaningful for data that are distributed like this?

L376: it would be good to explain the rationale for testing this better and earlier. Size and weight are obviously correlated and it is not a priori clear why that is an issue.

The names MBWarea and MBWdiameter are a bit confusing, especially since the unit is microgram. It is interesting to see that there is still a size weight relationship within a narrow size fraction, but not entirely surprising. What are the implications?

Fig. 2: please try to make this figure clearer. Allow more space between the subplots and
show the axes next to the points. Make sure that the decimal separator is a point and not a comma. Add space between genus abbreviation and species name.

L409-411: and variability within the size range.

L417-418: an R2 of 1 seems an artefact. It would only occur if all shells have the same shape, which I doubt.

L431: “Mean annual MBW area and roA values were calculated ... to illustrate the seasonal variability” does not make sense. It seems that monthly values were calculated. And why the average? It does not seem a robust indicator.

Fig. 3: (in addition to the comments above): use point instead of comma, add space in species name, add number of observations per month. Why are there no error bars for roA? (and what is the advantage of showing both metrics?)

L445-447: These two sentences mean the same (and I do not agree that all species show clear patterns at all).

L469: are these values flux-weighted?

Fig. 4: comma and space. It is unclear to me why the carbonate system parameters have not been averaged. The linear trends do not make sense, especially for the nutrients and salinity.

L481: knit-picking, but there can only be one minimum value.

L485: the increase in calcification intensity is not constant.

L491-492: I am sure that the other environmental variables also showed interannual variability. Why treat the carbonate system differently?

L495-496: This sentence is irrelevant here in the results section. Move to discussion if relevant at all.
L500-501: this sentence seems to come too early, you have not yet established that there is a reduction in calcification intensity. In addition, the figure does not allow to infer much about any Holocene trends since there is no time axis.

L502: how was the flux-weighting done? Using the monthly averages, or the observed data? If the latter, is it not biased to the times (fluxes) when the observations were made?

L506: “in the last 489 years” is not correct, see comments above.

Table 3: what does seasonal mean in this case? Monthly, three-monthly (which months?)? How was significance determined? And is that appropriate? Why provide the entire matrix, it contains a lot of redundant information. And what is the purpose of correlating everything with everything (how relevant is it to show that phosphate and carbonate ion concentration (use correct notation) are correlated)? Simply show what is relevant and discussed.

L558-561: this is interesting. Perhaps add some information about the life cycle of the different species here to discuss why truncatulinoides shows a unique pattern.

Fig. 5: commas. Not all environmental variables are from the DYFAMED site (or the caption of table 2 is incorrect).

L591-594: the use of the term OGC seems confused. Obviously, the niche of the species is multi-dimensional, i.e. the species are likely to have food and other preferences (temperature for instance), but it is not likely that OGC are linearly related to any environmental variable (e.g. it may be too hot for a species). OGC occur within a range of environmental variables, and there is hence no “proxy” for it, they can only be described (e.g, OGC are between x and y). Without establishing what OGC actually are, it is therefore difficult to use them as a predictor of growth/calcification intensity (L665).

L617-618: if this is true, what would the sensitivity be (how much less calcification with how much more phosphate) and how does this compare with the studies cited?

L666: what aspects of seasonal changes?
L685: presumable Fig 4. Are the trends significant?

L716: figure 4 instead?

L718-720: “the recent SST decrease”. Can this not be tested explicitly?

Fig 6: commas. What is compared (see above). The different colours for the species are redundant here as they are in separate graphs. Consider changing them and give more distinct colours to the sites to improve the clarity.

L797: see Bird (Bird et al., 2017) for bulloides.

L859: what is in the supplement and where can it be found?

References


