Review by Manuel F. G. Weinkauf Manuel.Weinkauf@unige.ch

I have been reviewing the manuscript by Mallo et al. entitled 'Low planktic for aminiferal diversity and abundance observed in a 2013 West–East Mediterranean Sea transect', and submitted to the journal Biogeosciences, in its first revised version.

This paper studies planktonic Foraminifera, sampled with plankton nets in the upper 200 m water column during spring/summer 2013 across the entire Mediterranean Sea. It reports abundance patterns of several species across a Mediterranean transect which is characterized by large differences in physical ocean properties (e.g. temperature, salinity). It further tries to infer the influence of those environmental parameters on the abundance and shell calcification intensity of selected (abundant) species. The study finds that the species composition changes across the Mediterranean, with *Globigerina bulloides* and *Trilobatus sacculifer* dominating in the western part, *Globorotalia inflata* in the central part, and *Globigerinoides ruber* (white)/*Globigerinoides elongatus* in the east. The species investigated for their abundance and calcification intensity show distribution and calcification patterns that differ between regions in the Mediterranean Sea, and can partly be correlated with environmental factors.

I appreciate this study for its large potential in filling in gaps in our current knowledge about species distribution in the Mediterranean and their changes both seasonally and across longer timespans by comparison of their results with earlier studies. It can also be a significant contribution to the still relatively scarce set of literature about shell calcification in planktonic Foraminifera. The sections are logically ordered, and the abstract gives a sufficient and well structured overview over the manuscript, but some information is lacking throughout the manuscript (especially the Material and Methods section). Otherwise the manuscript has an appropriate length (although the discussion is rather long I do not think it is excessive). The figures and tables are suitable.

The manuscript have been significantly improved since the original version in most regards. There are still some problems however, that in my opinion make it unpublishable in its current state: (1) The studies compared used a variety of different sampling techniques which necessarily lead to different results. Why the authors inisist on interpreting them as they are instead of correcting their data to make them comparable is beyond me. (2) The PCA applied by the authors is a downgrade from the faulty but more eloquent approach in the first manuscript version. The statistics is still not appropriate for the data, and many trends are extracted by eyeballing instead of proper hypothesis testing (see General comments for details). I think, many of the conclusions reached by the authors are too bold in light of the very basic data analysis. The manuscript must be either toned down in terms of interpretation, or the quantitative analysis must be improved considerably. After this has been done, I would very much appreciate to see this study published in Biogeosciences.

1 General comments

In the section below, I give detailed comments (including line numbers) about very specific issues. However, in this section I already want to summarise some major points that are more relevant for the entire manuscript than at any specific place.

- 1. The work does still not normalize its data for the consistent differences in sampling employed by the other studies, with which comparisons of assemblages are anticipated. Cifelli (1974) sampled the upper 250 m water depth, while Pujol and Vergnaud-Grazzini sampled the upper 350 m. This study uses mainly the association at 200 m and partly an integrated column of the upper 200 m. Furthermore, mesh sizes have been different between most studies. In addition, the authors now state that their net had a diameter of only $40 \,\mathrm{cm} \,(0.12 \,\mathrm{m}^2 \,\mathrm{opening})$, in contrast to the $0.5 \,\mathrm{m}^2$ common with most plankton nets (e.g. Pujol and Vergnaud Grazzini 1995). While absolute abundances are certainly normalized for filtered water volume, this much smaller net opening means that the authors have much larger errors in their assemblage data than the compared studies, because of the much lower volume of filtered sea water. All this has already been criticised in my first review, but the authors did not change anything, although I for instance suggested already there to use equations provided by Berger (1969) to normalize all studies concerning mesh sizes. The authors try to argue that Cifelli (1974), who actually used a comparable mesh size, argue in favour of their interpretation of changing abundances due to changing environments. However, they totally ignore that Cifelli (1974) used another depth range in their studies, so certainly they found other abundances. In my opinion, the authors cannot succesfully show, that the assemblage differences they observe between studies with employing such different sampling techniques are not an artefact of the data, but a real trend.
- 2. The systematics are still not consistent. Why is quadrilobatus designated as belonging to the genus Globigerinoides? From André et al. (2013), which the authors cite themselves, it is very clear that the species genetically belongs to the trilobus-sacculifer plexus (at least as long as recent specimens are concerned). It makes absolutely no sense to not only treat it as a separate species from Trilobatus sacculifer, but even put it into another genus. It should instead be correctly categorized as another morphotype of T. sacculifer.
- 3. The statistical analyses is still a huge problem. The authors state they applied a principal components analysis (PCA), which by the way is data visualization and no proper statistics (because it lacks any possibility to infer significance), and thus a step back from the faulty approach the authors applied in the first iteration of this paper. However, PCA does not include explanatory variables such as environmental parameters. So it is

first not clear to me what have been done, i.e. what are Factors 1 and 2 in Fig. 7? Have samples (as it seems) been ordinated by environment, and then somehow overlain by assemblages? Or is it indeed a redundancy analysis that have been applied, and if so, constrained for which environmental parameters? Furthermore, since PCA is using euclidean distances for ordination, it is very unsuitable for abundance data, and other methods like principal coordinates analysis are much more suitable for comparing assemblages (Hammer and Harper, 2006; Legendre and Legendre, 2012). The authors also still do not use proper techniques to interpret their findings in relation to the hefty multicollinearity in their data. I suggested some techniques in my first review (e.g. GLM, GAM). The authors may also use any of the techniques applied by the Thunell-work group, who also do an excellent job in that (e.g. Marshall et al., 2013; Osborne et al., 2016). As it is now, however, the authors only visually interpret trends in the PCA by eye, which is no proper and robust method when reliable interpretations should be reached.

2 Detailed comments

Line 50, 'Pujol and Vergraud-Grazzini, 1995': This work is consistently misspelled. It should be Pujol and Vergnaud Grazzini, 1995!

Line 52, 'bottom sediments': Should be 'surface sediments'.

Line 63, 'prominent differenced': Should be 'prominently different'.

Line 65, 'retrieved in different sites': Should be 'retrieved from different sites'.

Line 69, 'hydrographis': Should be 'hydrographic'.

Line 79, 'study being carried out': Should be 'study have been carried out'. Lines 97f, 'For further studies that relate foraminiferal calcification with environmental parameters see Weinkauf et al. (2016); Table 7.': You should also cite Marshall et al. (2013) in this regard.

Lines 106f, 'In addition, few size-normalized weight (SNW) and area density (ρ_A) studies from water column foraminifera are available in the literature': Area density is a form of size-normalized weight.

Line 112, 'spring2013': Should be 'spring 2013'.

Lines 120–122, 'In addition, empty tests are passive particles that ocean currents may displace horizontally, but that displacement is negligible due to their quick settling velocities (Caromel et al., 2014).': This is not always correct, and it might be good to show that drift distances in the Mediterranean are actually very low (van Sebille et al., 2015).

Line 146, 'become': Should be 'becomes'.

Line 166. 'primarily 200 m depth': Should be 'primarily from 200 m depth'. Line 179, 'MODIS Aqua L2 satellite': Should be 'MODIS Aqua L2 satellite data'.

Lines 186f, 'Samples were studied from the collecting bottles and the bottom collector, the latter representing 52.33% of the total sample

were treated in aliquots of 1/2, 1/4, 1/6, until 1/8.': I do not understand this sentence.

Line 188, ' \geq 350–500 μ m': Should be '350–500 μ m'.

Line 199, 'Globigerinella siphonifera/G. calida/G. radians plexus': Should be 'The Globigerinella siphonifera/G. calida/G. radians plexus'.

Line 204f, 'the individuals were weighed together by triplicate with a Mettler Toledo XS3DU microbalance': Which means the authors were actually applying the mean area density approach as described in Weinkauf et al. (2013) instead of the more advanced area density approach as described by Marshall et al. (2013).

Lines 216, 'The PCA was performed on the environmental parameters:': So how to understand this? The samples were ordinated by environmental parameters? What then are the scores of the black axes, passively projected assemblage scores? Or is this indeed a redundancy analysis instead of PCA? Compare also general comments why PCA is unsuitable anyways.

Line 218, '(Fig. 7)': What happened to Figs 3–6, which should be cited in the text before Fig. 7?

Lines 218–228, 'The first factor exhibited positive loadings... are shown in Figure 7).': This entire passage belongs into the Results section.

Line 244, 'The exceptions are at Station 3...': And what about stations 1 and 6?

Lines 246f, 'The 350–500- μ m size fraction dominates in the western Mediterranean and is progressively reduced eastwards (Fig. 4)': I do not see this trend. This could be due to the bad layout of figs 3 and 4 (see below).

Line 272, 'G. quadrilobatus': Incorrect genus (see General Comments).

Lines 274–276, 'The PCA performed on the environmental parameters and the sample scores on the two first components clearly shows a separation, regarding Factor 1, between the western and eastern Mediterranean stations (Fig. 7).': I do not understand how this 'PCA' was performed. Did it ordinate the samples on environmental data (as seems the case), then what are the black factors in fig. 7? Or is it indeed an RDA, then constrained for which environmental factors?

Line 278, 'station 10 is an exception': But stations 1, 6, 20, 21, and 22 (all Western Mediterranean) all have low a abundances as well.

Line 279, 'Factor 2': Should be 'principal component 2' or 'PC 2'.

Lines 283–285, 'Overall, the highest absolute abundance of all foraminifera seems related to food availability and only secondarily to the carbonate system (Fig. 7a).': While it makes the impression to be true, as it is this is eyeballing, because PCA cannot yield any significance but is only ordinating datapoints. Since many of your environmental factors show multicollinearity (as I already pointed out in my first revision) you need much more advanced, real statistical methods to say exactly whith which factors correlation is greatest. At the very least, you should use a more appropriate ordination method for abundances (probably constrained ordination, which at least delivers a significance for the overall correlation of data with environmental factors) than PCA, which uses euclidean distances.

Lines 286–292, 'With the exception ... path of Atlantic waters (Fig. 7b)': Where do you see this? *Globigerinoides ruber (white)* shows a peak (the richest sample) on the cold side of the ordination space, and *G bulloides* seems to be more correlated with pH. To convince me that those trends are true, you would have to show me something more robust than just a PCA impression (i.e. a compositional multiple regression as described by van den Boogart and Tolosana-Delgado (2013), as I also already suggested last time).

Lines 298f, 'The Atlantic and the Ionian–Adriatic–Aegean grouping have similar proportions of species.': Except that from Atlantic to Ionian–Adriatic–Aegean grouping dominances are completely shifted: *G. ruber* becomes much more dominant, *G. bulloides* and *T. sacculifer* are strongly reduced in abundance, *O. universa* is much more prevalent, and *G. inflata* is hardly there anymore.

Lines 313f, 'The high two-dimensional (silhouette) area-to-long axis correlation is best fitted by a power regression (Fig. S2).': Which, as I already argued in the first review, should be forced to have zero offset. The authors argued concerning this 'Size and mass of foraminifers relationship does not start at the origin (zero). The proloculus of planktic foraminifera measures between $15-30 \,\mu$ m in average, and has a certain calcite mass, which has so far not been determined (see Hemleben et al., 1989).'. This, however, only means that the model should stop short of zero. Especially when the authors argue that a zero-intercept model would not make sense because it would imply the existence of individuals with zero mass and size, is it not logical to them that non-zero-intercept model which allows a foraminifer to have mass at size zero or have a certain size without mass is even more problematic!

Lines 314f, 'The same growth pattern can be seen in *G. ruber* (white), *G. bulloides*, and *O. universa*': But this assumption is wrong at least in *O. universa*. There, size increase cannot be growth, because the spherical form is the terminal form and cannot grow considerably anymore.

Lines 318f, 'The specimens of G. ruber (white) from the Atlantic have the largest size followed by individuals from the Tyrrhenian Sea, and those from the eastern Ionian Sea.': If this statement is made, I already requested a statistical proof in the last review, to which the authors responded 'We do not need a statistical test to know which is the smallest value'. Since this shows a complete lack of understanding for the nature of any quantitative analysis, here is a short Statistics 101 (I again refer the authors to basic introductory literature such as Hammer and Harper (2006) or Dytham (2011): When dealing with natural values, one value will **always** be larger than the other when measured accurately enough. The question you want to answer is not, is one value larger, to which you know the answer beforehand, but is one value **significantly** larger. This means, is the difference you observe between the values in two random samples large enough that, taking into account uncertainty from the fact that you only sampled a couple of randomly selected specimens from the population, you can be reasonably sure that the populations the samples were drawn from differ in this value. An easy example:

I measure a difference of 0.3 cm between two samples. Do the populations from which those samples have been drawn differ in size? Well, when I use the variation in the samples to estimate the uncertainty in the estimate of the mean, I can tell with a certain probability. When the standard deviation in both samples (of, say, 100 specimens each) is 0.2 cm, then the 95% confidence interval is ± 0.02 cm, so the two populations do differ in size with a probability of more than 95%. If the standard deviation is 5 cm, in contrast, the 95% confidence interval is ± 0.5 cm, so the two populations do not show a significant difference in size. This is, what statistics is for, and in this sense, yes, you do need statistics to know which value is smaller!

Line 337f, 'higher ρ_A are related to slightly lower pH and higher food availability in the western Mediterranean and Atlantic stations': This must be proven, and from the PCA I doubt the pH relationship.

Line 340, 'opposite trend as in *G. ruber* (white)': Should be 'opposite trend than *G. ruber* (white)'.

Line 367f, 'Within the Mediterranean, a previous study with results comparable to ours, sampled the upper 350 m (Pujol and Vergraud-Grazzini, 1995).': They also sampled with another mesh size, for which still no corrections have been applied.

Line 401, 'smallesr': Should be 'smaller'.

Lines 409–411, 'The lower absolute abundance of individuals in our study compared to Pujol and Vergraud-Grazzini (1995), together with low species diversity in the Mediterranean, may indicate a trend of changing conditions over the last decades, ...': I still believe that this has to do more with the different mesh-sizes. The size fraction between $120 \,\mu\text{m}$ and $150 \,\mu\text{m}$ in my experience contains a lot of the standing stock of foraminifers. Section Factors controlling the abundance of the main species: All trends described here are purely derived from the PCA by eye, without any appropriate test. While their explanation can be valuable, their interpretation should be toned down considerably.

Lines 445f, 'The increasing dominance of G. ruber (white) from the western to the eastern Mediterranean Basin coincides with the east-ward increasing salinity (Fig. 7d).': Or Temperature, or CO_2 . It is hard to say without proper analytical techniques under this degree of multicollinearity. Line 537: Remove second 'its'.

Line 548, 'but abundances are slightly higher in the western basin to than the east.': I highly doubt that from the PCA alone. You could prove it though.

Line 569f, 'In contrast, the ρ_A of *O. universa* does not show any change between the western and eastern basins (Fig. 7i), and cannot be used to identify and quantify particular environmental effects.:' I also doubt that there is a difference between basins in *G. bulloides*, and since the authors still refuse to use proper quantitative techniques to prove it ...

Line 615, 'larger IQR indicates ...': This is only true, when the variation in the sample is normalized for expected value (i.e. mean). This means, calculating the coefficient of variation, which I already requested in the first review. The authors replied 'As also described above, in our comment to the reviewer comment about lines 480–482, we are unsure about what statistical method and/or calculation the reviewer is referring to here. Is there a distinct suggestion of some kind, with a reference? We are not sure how to calculate a "coefficient of variation" with regard to box plots and their statistics.'. No, I do not have a reference for it, because the coefficient of variation is such a basic and old method that its origins are lost in the mist of time, and you would not cite a reference as you would not cite a reference when calculating a mean value. Rather, the coefficient of variation is explained (and listed in the index) in every basic statistics book I suggested the authors to consult in my first review. It is also very easily found using Google and the search term 'coefficient of variation'. Again, in short, variation is always correlated to mean value, so variations of samples which mean value differs must be corrected for this stochastic effect. An example: Let's say you measured the length of twenty mice and found it to be 3 ± 0.5 cm. You also measured the length of 20 elephants and found it to be 4 ± 0.5 m. Which species has the higher variation? The absolute value is much larger for elephants (0.5 m) than for mice (0.5 cm), but when calculating the coefficient of variation you actually find mice to be more variable in size (0.166) than elephants (0.125). Since none of the IQRs in the manuscript are corrected (and I would recommend to use the standard deviation instead of the IQR anyways) all conclusions drawn by the authors concerning variation in their samples are invalid.

Line 624, 'variability in ρ_A data increases with increasing absolute ρ_A ': Exactly as stochastically predicted. Calculate the coefficient of variation and compare again.

Line 633, 'retarded': Should be 'hampered'.

Line 636, 'seems': Should be 'seem'.

Line 640, 'suited conditions': Should be 'suitable conditions'.

Line 648, ''heavier average': Should be 'steeper average', maybe.

Line 651f, 'All of these findings support our idea of an effect of limited alimentation on calcification.': I do not understand this sentence.

Caption Fig 4 'Sample size is indicated by n below each station code.': This information is not present in the figure.

Figs 3 and 4: A lot of the interpretation by the authors in concerned with east-west trends. Then why are the graphs not ordered west-east, instead of by station number?

References

André, A., Weiner, A., Quillévéré, F., Aurahs, R., Morard, R., Douady, Ch. J., de Garidel-Thoron, Th., Escarguel, G., de Vargas, C., and Kučera, M.: The Cryptic and the Apparent Reversed: Lack of Genetic Differentiation within the Morphologically Diverse Plexus of the Planktonic Foraminifer *Globigerinoides sacculifer*, Paleobiology, 39, 21–39, doi:10.1666/0094-8373-39.1.21, 2013.

- Berger, W. H.: Ecologic Patterns of Living Planktonic Foraminiferal, Deep-Sea Research and Oceanographic Abstracts, 16, 1–24, doi:10.1016/0011-7471(69)90047-3, 1969.
- Dytham, C.: Choosing and Using Statistics: A Biologist's Guide, Wiley–Blackwell, Oxford and Chichester and Hoboken, 3 edn., 2011.
- Hammer, Ø. and Harper, D.: Paleontological Data Analysis, Blackwell Publishing, Malden and Oxford and Carlton, 2006.
- Legendre, P. and Legendre, L.: Numerical Ecology, no. 24 in Developments in Environmental Modelling, Elsevier, Amsterdam and Oxford, 3 edn., 2012.
- Marshall, B. J., Thunell, R. C., Henehan, M. J., Astor, Y., and Wejnert, K. R.: Planktonic Foraminiferal Area Density as a Proxy for Carbonate Ion Concentration: A Calibration Study using the Cariaco Basin Ocean Time Series, Paleoceanography, 28, 1–14, doi:10.1002/palo.20034, 2013.
- Osborne, E. B., Thunell, R. C., Marshall, B. J., Holm, J. A., Tappa, E. J., Benitez-Nelson, C., Cai, W.-J., and Chen, B.: Calcification of the Planktonic Foraminifera *Globigerina bulloides* and Carbonate Ion Concentration: Results from the Santa Barbara Basin, Paleoceanography, 31, 1083–1102, doi: 10.1002/2016PA002933, 2016.
- van den Boogart, K. G. and Tolosana-Delgado, R.: Analyzing Compositional Data with R, Use R!, Springer-Verlag, Berlin and Heidelberg, doi: 10.1007/978-3-642-36809-7, 2013.
- van Sebille, E., Scussolini, P., Durgadoo, J. V., Peeters, F. J. C., Biastoch, A., Weijer, W., Turney, Ch., Paris, C. B., and Zahn, R.: Ocean Currents Generate Large Footprints in Marine Palaeoclimate Proxies, Nature Communications, 6, Article 6521, doi:10.1038/ncomms7521, 2015.
- Weinkauf, M. F. G., Moller, Т., Koch, M. С., and Kučera, Planktonic M.: Calcification Intensity inForaminifera Reflects Ambient Conditions Irrespective of Environmental Stress. Bio-6639-6655, doi:10.5194/bg-10-6639-2013, URL geosciences, 10,http://www.biogeosciences.net/10/6639/2013/bg-10-6639-2013.html, 2013.