Biogeosciences Discuss.,
doi:10.5194/bg-2016-266-RC2, 2016

Interactive comment on "Low planktic foraminiferal diversity and abundance observed in a 2013 West-East Mediterranean Sea transect" *by* Miguel Mallo et al.

M. F. G. Weinkauf (Referee)

manuel.weinkauf@unige.ch

Received and published: 29 July 2016

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

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 presented figures and tables are suitable for the most part, but I think some additional figures are needed to present key results.

Having said that, I must unfortunately conclude that in its current state I do not see this manuscript fit for publication, because it shows a multitude of problems, which I will explain in detail in the sections below. Leaving aside that the manuscript definitely needs improvement in terms of style (especially the first sections), I have severe problems with the applied methodology, which is compromising the results reached by the authors. This starts with the application of possibly an inferior method of shell calcification intensity quantification (in fact I cannot say if that is the case, because the authors never explain which method they use or make even clear that they know that different methods exist). Much more severe even is the application of a completely unfit set of analytical techniques, which are fully inappropriate for the data at hand and the questions the authors want to answer. It is furthermore hard to follow the manuscript in parts, because it contains thematic leaps, and it is often not clear about why the analysis described in a certain passage is needed or what the authors want to show with it. The authors are furthermore missing out of the chance to perform a proper assemblage analysis and broadly footed comparison with earlier studies.

I must therefore state that possibly part of the data acquisition and certainly virtually the entire data analysis must be repeated by the authors using adequate methods, leading to my suggestion of a major revision. After this has been done, and the manuscript has been streamlined (after which in my opinion it should be reviewed again), 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. While reading the manuscript I noticed that the writing style needs some attention. The manuscript is understandable, but there are plenty of orthographical and grammatical errors or weird phrasing throughout. Those should be dealt with (and I noted some suggestions in the detailed comments), to make the manuscript more accessible for the reader.

2. The manuscript is partly missing important information, diverts from the topic, or promises undelivered results. Some examples: Parts of the manuscript, especially the section 'Oceanographic Setting', are lacking citations of information sources. Information on data sources and methods are largely missing. The temperature and salinity might come from the mentioned CTD casts, but the carbonate saturation values most certainly not: Have they been calculated on the

basis of water samples (on board or in the lab) or calculated on the basis of database oceanographic data? Which of the several existing methods to calculate size-normalized shell weight has been used? Which software has been used for statistical analyses? All this information belongs in the far too short Material and Methods section! The reason for several analyses (e.g. the correlation between shell size and shell weight) is not properly explained, thus leaving the reader guessing why the authors deem this necessary. A comparison of assemblage data with earlier studies to study long-term trends is promised but never really delivered (not on a reasonable analytical level at least).

3. The existing images are OK, for the most part (labels might be a bit small in several of them). However, several key findings of the study are not presented in any suitable graphical manner, instead referring to figures which cannot present these data in a suitable way. Most notably amongst these, while there are several claims made about the influence of environmental factors on abundance and SNW of the species, not a single such relationship is graphically shown in a cross plot.

4. The manuscript uses several wrong species names and species concepts. The most prominent one is the unfortunate use of the terms *Globigerinoides ruber* sensu stricto and *Globigerinoides ruber* sensu lato, which are pooled, together with *Globigerinoides ruber* (pink), within the same species. This is blatantly wrong. Aurahs et al. (2011) has established that *Globigerinoides ruber* (pink), *Globigerinoides ruber* (white) (your sensu stricto), and *Globigerinoides elongatus* (your sensu lato) are distinctly different species, both biologically and in terms of morphology; and has therefore rehauled their Linnean taxonomy. Could we please all agree that 5 years after this publication we could at last all start to call them by their proper names and abandon this unfortunate sensu stricto/sensu lato distinction. It would be one thing if it would only be about names (I would still request to use up-to-date terminology, but it would be a minor mistake). Rather,

G. elongatus is not even the adelphotaxon to *G. ruber* (white), but is more closely related to *Globigerinoides conglobatus*. Pooling them together under the same species name thus produces a polyphylum. If you want to pool them for some purposes (which can make sense) you can call them '*G. ruber*/*G. elongatus* plexus', or something along those lines. Second, the species *Globigerinella siphonifera* is reported from the samples. However, it is not clear whether this means that only *G. siphonifera* is present, or whether this is a collective term for the entire *Globigerinella siphonifera*/*Globigerinella calida*/*Globigerinella radians* plexus (Weiner et al., 2015), within which species have not been separated by the authors. Third, but less serious because this really is only a naming issue, the former *Globigerinoides sacculifer* should be referred to as *Trilobatus sacculifer* meanwhile (Spezzaferri et al., 2015). Furthermore, in that species your 'quadrocameratus' morphotype is correctly referred to as 'quadrilobatus' morphotype to my knowledge.

5. **The most important issue** is with the statistical analytical approach. According to lines 146–147 you are using a Pearson product moment correlation to test the relative abundances and shell calcification intensities of several species against environmental parameters. This is horribly wrong on a multitude of levels, as I will summarise hereafter. For further details you may have a look at Dytham (2011), Legendre and Legendre (2012), Faraway (2006), and McDonald (2009). **I**—You assume a causal relationship between environmental factors and SNW/species abundance. Correlation analyses are not appropriate here, regression analyses with SNW/abundance as the dependent variable against the independent environmental factors must be used. Occasionally this makes only a cosmetical difference (i.e. type I linear regression vs. Pearson product moment correlation), but even then it is of methodological and implicational importance (compare Legendre and Legendre, 2012, box 10.1). In this case, however, it is even more important because of the points below. **II**—Type I regression (as well as its correlation

equivalent for that matter) is only applicable under certain circumstances, one of which is that x-values are measured without errors (McDonald, 2009; Dytham, 2011; Legendre and Legendre, 2012). It is therefore nearly only usable for laboratory experiments. As long as you are testing for the influence of parameters that you actually measured on board (temperature, salinity, pH), you might this this still works with a lot of good will, but I would argue that even then you have an error on those values, because you only have a snapshot image, and not a mean (let alone constant) value covering the entire life-time of your specimens. Further, I assume (you never state that) that at least part of the data you needed to calculate the carbonate system comes from averaged database data anyway!? And at least then, and in my opinion under all circumstances, you have to use robust type II or type III regression methods. **III**—You cannot simply test the same dependent variable against several independent parameters in different tests. The simple reason is that each of those test has its own type I error chance, and those are summing up until (after a sufficient number of tests) you are guaranteed to get at least one type I error in your analyses (compare Dytham, 2011; Legendre and Legendre, 2012). It is imperative that under such conditions at the very least all multiple tests (i.e. all tests for the influence of individual environmental factors on SNW or abundance per species) are corrected for this problem. Either using a correction for the family-wise error rate (e.g. Bonferroni correction), or a correction for the false discovery rate (e.g. Benjamini and Hochberg, 1995). **IV**—Making several such analyses and correcting them per species is still not the ideal solution, mainly because (as usual in marine environments) all independent variables show a large degree of multicollinearity (just have a look at your own Fig. 1). This means that such simple parameter-wise tests may detect an influence of several parameters, but only because they are highly correlated, and it is unclear which factor influences the dependent variable the most (or at all, for that matter). For the case of SNW in particular it might be much better to use an approach that can test for all independent variables at once, while reducing the influence of the multicollinearity between different environmental factors (Dormann et al., 2013). Such methods could for instance be generalized linear models (GLM) or generalized additive models (GAM), both of which have the added benefit over multiple linear regression that they are invariant to the order in which independent variables are added to the model (compare Faraway, 2006). For relative abundances you face the additional problem, that y-values are not independent of each other within a sample (e.g. if *G. ruber* already represents 50 % of the assemblage, then *G. bulloides* cannot be more abundant than 50 % anymore in that same sample). While there are ways around this (most notably, using absolute abundances with an appropriate link function in a GLM, or applying any of the methods described in van den Boogart and Tolosana-Delgado (2013)) you may also prefer to analyse the assemblage data using suitable ordination techniques (compare for instance Hammer and Harper, 2006). This would have the added benefit that such ordination techniques can also be adapted to properly compare your assemblage with that of earlier studies, in this way delivering on a promise made in the introduction and never fulfilled in the manuscript.

2 Detailed comments

Line 33, 'calcareous zooplankton': I would be very careful talking about zooplankton here. While it is true that all planktonic Foraminifera can live heterotrophic, many are also able to harbour photosymbionts.

Line 35, 'Hembelen et al., 1989': Should be 'Hemleben et al., 1989'.

Line 36, 'due to': Should be 'and show'.

Lines 36–37, 'The species are adapted [...] spines and test shape.': They are certainly adapted to different environments, because naturally there cannot be any two species which occupy exactly the same niche, but implying such a trivial form of adap-

tation is far too oversimplified.

Line 37, 'test shape': Should be 'shape, which are partly related to those adaptations'.

Lines 37–39, 'The distribution of foraminifera [...] which shifts during ontogeny.': A citation for this statement is needed.

Lines 42–45, 'Ecological tolerance limits [...] departure from optimum conditions (Arnold and Parker, 1999).': Which is basically true for every organism, so what is the point here? Plus, this is hardly the best citation for this statement. What about Bé (1977) for example?

Lines 48–50, 'The first modern study of planktic foraminifera [...] expedition of 1947–48.': Was this study published? Cite a source.

Line 54, 'at 250 m depth': Should be 'of the upper 250 m water column'.

Line 57, 'that': Should be 'that the'.

Lines 57–61, 'Thunell (1978) studied samples [...] inside the Mediterranean.': Break up this sentence.

Line 65, 'wide': Should be 'large'.

Lines 65–66, 'They concluded [...] variable foraminifera assemblages,': This is not entirely correct. Pujol and Vergnaud Grazzini (1995) only state that the observed assemblage patterns 'cannot be entirely explained by the general temperature and salinity differences among the different Mediterranean Basins' and are also strongly correlated to more regional hydrogeographic patterns.

Lines 70–72, 'The calcification of foraminifera [...] (Schiebel and Hemleben, 2005).': Those are neither the only factors influencing shell calcification intensity in planktonic Foraminifera, nor are all of the stated relationships universally true. Compare Marshall et al. (2013, tab. 1) and Weinkauf et al. (2016, tab. 7) for a summary of this matter.

Lines 70–77: I think the cited literature for calcification studies is by far not exhaustive. What about Broecker and Clark (2001b), Barker and Elderfield (2002), de Villiers (2004), Manno et al. (2012), and Marshall et al. (2013), to name but a few.

Line 76, 'building': Should be 'formation'.

Lines 82–83, 'In addition, few size-normalized weight (SNW) studies from water column foraminifera are available in the literature.': Then please provide such examples here in the form of citations.

Line 91, 'more unbreakable tests': Should be 'tests with thicker walls'.

Line 92, 'empty tests are passive particles that ocean currents may displace.' Which is perfectly true for living Foraminifera as well; hence they are plankton, not nekton.

Lines 97–98, '(2) characterize, at the species level their ecology through their seasonal and geographical distribution and abundance by comparison with previous studies,': This point is not really present in the paper, at least not above a relatively comparative level. The interpretation why abundances might be different now than they were 20 years ago, and any reliable analysis and graphical presentation that shows that in the first place, is largely missing.

Line 103, 'with a strong thermohaline and wind-driven circulation,': Citation needed!

Lines 105–106, 'These basins are composed of different sub-basins due to partial isolation caused by sills that influence the water circulation, and by different water properties.': Citation needed!

Lines 106–107, 'World Ocean': Should be 'worlds oceans'.

Lines 107–109, 'where the nutrient-rich Atlantic surface waters [...] (evaporation exceeding precipitation).': Citation needed!

Line 111, 'until the': Should be 'and reach as far as the'.

Lines 113–116, 'In the eastern basin, [...] and fresher toward the western basin.': Citation needed!

Lines 117–118, 'Waters returning to the Atlantic through the Strait of Gibraltar at depth are cooler and saltier than the inbound waters, and compensate for the inflow from the Atlantic.': Citation needed!

Line 135, 'at 200 m depth': Should be 'from 200 m depth to the surface'.

Line 141, 'counted and separated by species and size': Should be 'split into fractions by size'.

Line 142, 'to determine the absolute and relative abundances': Should be 'and planktonic Foraminifera were counted on the species level'. Furthermore, it is not mentioned which taxonomic system is used. It is most certainly not up-to-date (compare general comments).

Lines 144–145, 'Individuals of the same station and species within a 50 μm diameter size constraint were weighed with a Mettler Toledo XS3DU microbalance ($\pm$ 1 μg of error).': So I assume they were weighed together (single shell measurements would require a more precise balance). But were the measurements afterwards actually corrected for mean shell size per sample (MBW approach, Barker and Elderfield (2002)), or was the simple SBW approach used (Lohmann, 1995; Broecker and Clark, 2001a). The main problem is that in the latter case, Beer et al. (2010a) has shown that the SBW method is not fully effective in eliminating the shell size effect. Additionally, results cannot be independently replicated and tested when the exact methodology is not sufficiently described. Also, 'error' is the wrong term in this context, and 'nominal precision' should be used instead.

Lines 146–149: It is not mentioned anywhere which software was used to carry out statistical analyses.

Lines 147–149, 'Absolute abundances [...] observed within the environmental parameters.': This is no valid reason at all to skip this. It can be that you are not interested in this, then state why, or that you are concerned about the validity of the results, then state why. A large difference in values does not compromise such an analysis at all if the correct techniques are applied.

Line 171, '*Globigerinoides ruber* sensu strict (s.s.)': As mentioned in the general comments, this species is correctly referred to as *Globigerinoides ruber* (white). Please change in the entire manuscript.

Line 174, 'Globigerinella siphonifera': Your species list contains only *Globigerinella siphonifera*, but neither *G. calida* nor *G. radians* (compare Weiner et al., 2015, and the general comments above). This could mean that either you checked and the other two species are not present at all, or you lumped the entire plexus into one category. Please explain what is the case here.

Lines 176–178, 'In addition, a higher percentage [...] and may not be generalized.': Given the fact that in plankton tows you have only little control over the growth stage of your individuals, one may wonder to what degree this size trend over time may represent a reproduction event.

Line 180, 'sample': Should be 'assemblage'.

Lines 183, 187, 191, 197, '(Fig. 3; Fig 4)': The referred information is illustrated by neither of these figures, because Fig. 3 does not give shell sizes and Fig. 4 does not distinguish between species. Unless Fig. 3 would only represent the fraction $> 350\,\mu$m, but then this is stated nowhere in the figure caption.

Line 192, '*Globigerinoides sacculifer*': This should be *Trilobatus sacculifer* (compare Spezzaferri et al., 2015, and general comments).

Line 198, 'quadrocameratus': Should be 'quadrilobatus' in the entire manuscript.

Line 218, 'A Pearson test': This is the wrong method for the question that should be

answered (compare general comments). By the way, even if correlation per species was the correct approach, abundance data are by default not normally distributed but follow a Poisson distribution. This rules out any parametric test in the first place, and would leave Spearman rank-order correlation or Kendall rank-order correlation as the only reasonable alternative. Compare the general comments section, however, why neither of these is appropriate here.

Lines 220–222: All p-values reported here are invalid, because they have not been corrected for multiple testing on the species level. The general comments section gives more discussion about this. Additionally, why is nothing of that presented in a graphical form?

Lines 222–223, 'Relative abundance was selected instead of absolute abundance to avoid bias due to the big differences between stations' results in absolute abundance.': This approach, however, introduces new problems because now the abundances per station are not independent; and the given reason for this decision is invalid anyway. Compositional regression (van den Boogart and Tolosana-Delgado, 2013) or other adequate approaches would be needed. Compare general comments section.

Lines 223–225, 'The remaining species [...] abundance and environmental parameters.': This is no reasonable explanation. The mere lack of the species at some stations would not rule out such an analysis, if there are still enough stations with values > 0 left.

Lines 229–230, 'The high two-dimensional (silhouette) area-to-diameter correlation is best fitted by a power regression (Fig. S2).': As would be expected. But why is this important in the context of that paper? Additionally, from a purely modelling-point-of-view I might argue that the regression should be fitted so that they are forced to have a zero intercept (everything else seems wrong).

Lines 230–235, 'Comparing the average values [...] northwestern Mediterranean

(Fig. S2).': If the idea is to compare shell sizes between different basins, then this is hardly the best method of presentation. A boxplot or barplot would be much more appropriate here. Further, it is stated nowhere which statistical techniques were used to test the shell size differences between basins. I assume an ANOVA followed by post-hoc tests, but this is explained nowhere.

Lines 237–239, 'The diameter-to-weight relation [...] ($r^2 = 0.516$; Fig. S3).': If you want to imply a dependency relationship (which can make sense, depending on your intention), then it would probably be more logically to assume that weight is dependent on size, so you should exchange the axis in your Fig. S3. Otherwise, here a correlation would be more appropriate. Furthermore, the question is again what is the sense of this analysis in the context of that paper. It should be made clear for the reader, why this analysis is performed.

Lines 239–240, '*O. universa* was finally discarded for comparisons between SNWs at different locations due to a low area–weight correlation, while data from *G. ruber* s.s. correlate well (Fig. S4a).': I do not really see the reason for this. 1) The weight–size relationship is not that bad (p-values are not given, interestingly). 2) I do not understand why the authors would insist in such a relationship to be a necessity for the interpretation of SNW. Sure, if there is no good relationship it would be difficult to predict shell size from shell weight or vice versa. But especially if you imply a relationship between calcification intensity and the environment you would expect to see deviations from this relationship. Otherwise, shell weight would be a function purely of shell size, and size-normalized shell weight would not have any value in environmental interpretations. Now, a lower R^2 value in *O. universa* in my opinion only means, that its shell weight is to an even lower extant controlled by shell size than it is in other species. This could mean, that *O. universa* is more susceptible to environmental protrusions in regard to its ability to control calcification, which would by some standard make it an even better proxy species. I can think of no reason why a low correlation value itself would make SNW interpretations invalid, however.

Lines 240–242, 'The eastern Mediterranean [...] *G. ruber s.s.* (Fig. S4d-e).': This is again not an appropriate way of presenting those results. Use a boxplot/barplot instead.

Lines 243–244, 'The eastern Mediterranean individuals have the lowest median SNW': Is this just eyeballing or has it actually been tested somehow, which regions are different and which are not concerning SNW?

Line 245, 'μg $\cdot$ μm^{-2}': So from this unit I assume the authors yet used the MBW approach, instead of SBW!? It is imperative that this is made clear in the Methods section.

Lines 248–251, 'A Pearson correlation test [...] correlation with fluorescence ($p = 0.01$).': Apart from the fact that this technique is again inappropriate for the data (compare general comments and discussion for the abundance data) it is interesting that this important result is not graphically presented in any form. If such relations really exist, you should show them in the form of a figure.

Lines 252–253, 'The Atlantic has [...] opposite trend as in *G. ruber* s.s.': Again, eyeballing or tested?

Lines 256–257, '*G. bulloides* is positively correlated with pH and [CO_3^{2-}] ($p = 0.05$) in the Pearson test.': Which is again not shown in any graphical representation.

Line 280, 'occurs in a': Should be 'come from'.

Line 280, 'season of the year': Should be 'seasons'.

Line 283: Delete 'eastern'.

Line 284: Delete 'both'.

Lines 285–286, 'no significant differences are observed between samples collected during day and night.': Is this a subjective impression or was it tested statistically, because only in the latter case you should use 'significantly'. Further, why is this not presented graphically somewhere?

Lines 287, 'accounting for a single species': Which is blatantly wrong for virtually every perceivable species concept.

Lines 288–289, '*G. ruber*: sensu stricto, sensu lato (containing different cryptic species; Aurahs et al., 2009a),': This is no up-to-date information in this regard anymore. Furthermore, the references contain only one 'Aurahs et al. (2009)', not an 'a' and 'b' version; please correct this.

Line 292, 'with Cifelli (1974)', 'with Pujol and Grazzini (1995)': 'with' should in both cases be 'by'.

Line 294, 'reach': Should be 'reached'.

Lines 294–295, '*Turborotalita quinqueloba*, *Neogloboquadrina pachyderma*, and *Globorotalia truncatulinoides*.': Another problem for some species (certainly not *G. truncatulinoides*, but probably *T. quinqueloba* and potentially *N. pachyderma*) is that you used a 150 μm mesh size. Most studies by default use 100 μm for plankton net hauls, and part of the discrepancy you see (also in terms of general abundances) might be that you missed a lot of the small specimens. From my experience (compare Weinkauf et al. (2016) vs. Storz (2006)/Storz et al. (2009)) you can miss the majority of specimens in some species by just switching from 125 μm to 150 μm. In this regard, Pujol and Vergnaud Grazzini (1995) used 120 μm, potentially explaining a lot of your observed differences.

Lines 297–298, '*G. sacculifer* type quadrocameratus was not found in previous studies': A potential problem with this statement is whether in those previous studies *T. sacculifer* has been consequently subdivided. While most studies I am aware of distinguish between the sacculifer- and trilobus-morphotypes, it is often unclear whether the quadrilobatus- (or immaturus-) morphotypes would be counted separately if discovered and truly are absent in the samples, or if they are by default pooled in with the

trilobus-morphotype.

Lines 300–302, 'The lower absolute abundance [...] recent years': Yes, it could. But again, given that Pujol and Vergnaud Grazzini (1995) definitely used a finer mesh size, this could simply be the result of you missing a lot of specimens. I would therefore be very cautious with this interpretation. Berger (1969) provides equations with which observed abundances could be calibrated for different hypothetical mesh sizes, and such a correction of your data might provide a much better comparability with earlier studies.

Line 311, '(Fig. 4).': Again, as this figure does not distinguish between species it cannot illustrate the trends you describe here.

Section '5.2. Factors controlling the abundance of the main species': The authors try to interpret each individual species' abundance in terms of seasonality and compare it with other studies. However, it is not fully clear what the purpose of this is supposed to be. Many of the described trends are not new, and while it is always good to replicate results, this should not be the main purpose of the manuscript. Rather, the comparison of abundances with studies from several years ago, and the interpretation of potential reasons for changes (as promised in the introduction) is largely missing.

Line 314, 'results': Should be 'samples'.

Line 324, 'Both varieties *G. ruber* sensu stricto (s.s.) and sensu lato (s.l.)': Those are not varieties but distinctly different species, *G. ruber* (white) and *G. elongatus* respectively. Moreover, they are not even sister-taxa, but *G. elongatus* is the adelphotaxon to *G. conglobatus*. While they have comparable environmental preferences, and might thus be pooled for such an analysis as you intend to do, they should under no circumstances treated in a way that implies they are remotely the same species.

Lines 324–325, 'share similar habitats': Yet they have different environmental preferences, with *G. elongatus* living deeper (Steinke et al., 2005; Numberger et al., 2009) and showing different seasonality (Weinkauf et al., 2016).

Lines 331–332, 'as demonstrated by positive significant correlations with temperature in the *G. ruber* s.s. variety ($\mathrm{p} = 0.01$).': Not that I would oppose this interpretation, but is is yet derived from inappropriate analytical techniques.

Line 338, 'strong positive correlation with salinity ($\mathrm{p} = 0.01$)': Derived from invalid methods!

Lines 340–341, 'The findings of Watkins et al. (1996) are supported by the negative correlations of standing stocks': Are they? If Watkins was right, would you not expect no correlation at all between nutrient availability and abundance of *G. ruber*? Rather, it seems that *G. ruber* is faring less well in regions with more nutrients (if this trend is supported by proper statistical analyses, this is). This means that higher nutrient availabilities are negative for the species, maybe because it loses its competitive advantage against other species, or the higher nutrient concentration reduces light levels, thus hampering the photosymbiont activity.

Lines 341–342, '*G. ruber* s.s. and fluorescence data of our study ($\mathrm{p} = 0.05$).': Derived from invalid methods!

Lines 352–353, 'Hydrographic conditions and consequently food availability seem to be the factors limiting more its abundance once it has reached its habitable temperature range.': Yes, but is this not what would be expected? Liebig's law of the minimum is equally valid for protists and animals as it is for plants.

Line 359, 'shows': Should be ', the species shows'.

Line 368, 'positive correlation with fluorescence ($\mathrm{p} = 0.05$),': Derived from invalid methods!

Line 372, 'Raden et al., 2012': Should be 'van Raden et al., 2012'.

Line 377, 'specie': Should be 'species'.

Line 383, 'opportunistic species': Opportunistic species are such species which can cope with highly unstable and/or unfavourable conditions better than other species can do. They thus massively dominate environments where few other species can live, resulting in very low diversities in those environments. This is often a transitional process until the environment becomes more stable/habitable, after which the opportunistic species are replaced by a more diverse community, because in developed environments they are at a competitive disadvantage to such species. I therefore do not believe that 'opportunistic' is the correct term to describe *G. bulloides*, which is cosmopolitan and often occurs in rather diverse assemblages.

Line 384, 'It correlates with fluorescence peaks since it feeds on phytoplankton': Probably correct interpretation, but derived from invalid methods!

Line 408, 'Its negative correlation with temperature ($\mathrm{p} = 0.01$)': Derived from invalid methods!

Line 417, 'only absent from': Should be 'being absent from only'.

Line 428, 'even if this is not supported by our Pearson correlation.': Which is an inappropriate method anyways!

Lines 438–439, 'The size-normalized weight (SNW) of tests of both *G. ruber* s.s. and *G. bulloides* are statistically significant': This statement is nonsensical, a value itself cannot be significant, it can only be significant in regard to a null hypothesis. I assume you refer to the fact reported in the Results section and Suppl. Fig. 4, that size and weight are not perfectly correlated in *O. universa* (otherwise I do not even know what you want to imply). However, as already mentioned above, this is in my opinion no prerequisite for the SNW to have a meaning. This is even leaving aside, that it is never established whether this relationship is really insignificant ($p > .05$) or if the R^2 value is simply to small for the authors taste.

Line 439, 'follow a systematic change from the Atlantic towards the eastern

Mediterranean': This might be, but it was never properly tested or depicted graphically.

Lines 441–443, 'In contrast, [...] environmental effects.': Incorrect! The strict correlation between size and weight may not exist, but this only means that especially in this species there must be other factors influencing calcification intensity.

Section '5.3.1 Unknown control of the SNW of *O. universa*': OK, now your regression between shell size and shell weight makes more sense, and it would have been good to explain this in the beginning already. I do appreciate that you discuss this possibility of cryptic diversity and gametogenic calcite meddling with your data. However, what André et al (2014) detected are subtypes, they do not even rank on the species level. On that level you have also several subtypes in *G. ruber* and *G. inflata*. To be honest, it could be that the lack of strong correlation between size and weight in *O. universa* results from such an effect that the subtypes react differently. But it can still as well be, that this species simply reacts more heavily towards environmental factors concerning its calcification. I would thus not go so far as to categorically rule out that species for a calcification analysis, because you simply do not know what is the case here. It is still interesting to see SNW values for that species as well, although they might suffer from higher uncertainty. Even more so since despite a large spread, the correlation between size and weight does not seem to show bimodality (indicative for the cryptic species problem), and possibly the SNW data would not do so either. Conversely, the values seem to show a wider spread for larger shells, which can mean that gametogenic calcite is more of a problem, or simply truly that this species is more variable in calcification intensity (then presumably influenced by environmental factors). After thoroughly discussing why this might be a less reliable signal, I would therefore still want to see how SNW in *O. universa* scales with environmental factors.

Lines 448–449, 'Weight-area relation data do not show any statistically significant systematic distribution (Fig. S4c).': You probably mean 'correlation', not 'distribution'.

Lines 453–454, 'their pore-size is also affected by environmental conditions including water temperature (e.g., Bé et al., 1973).': This statement is critical. Bé et al. (1973) did not know about different cryptic species. It might be that pore size is indeed influenced by environmental factors across all cryptic species, but also that cryptic species prefer different water temperatures and what Bé et al. (1973) interpreted as pore size changes within the species is simply the result of different species (with inherently different pore sizes) dominating different water masses.

Line 476, 'nutrient concentration and food availability.': Which is basically the same thing in the context if this study, isn't it?

Lines 476–478, 'However, in contrast to *O. universa*, the SNW data of *G. ruber* and *G. bulloides* follow systematic distributions, which are statistically significant.': It is again not clear what you mean with 'distributions'. All data have a distribution, and values themselves cannot be significant or insignificant. I assume you refer to a significant correlation between size and weight in those species.

Lines 478–480, 'High SNW in the Atlantic [...] also noticeable in Fig. S2d-e and Fig. S4d-e).': Those graphs are all not appropriate to show that. Rather, an actual crossplot between SNW and the individual environmental factors must be shown. Interestingly, this trend is reversed to what has been reported from the Azores Front (Weinkauf et al., 2016).

Lines 480–482, 'At the same sites, [...] interpretation of the data (Fig. 6).': Which could be shown effectively by calculating and presenting the coefficient of variation at those stations. Additionally, how could this trend then be interpreted?

Lines 485–486, 'The relationship between food availability and SNW in *G. bulloides* is opposite to that in *G. ruber* s.s. (Fig. 6)': A better figure is needed to illustrate this.

Lines 488–489, 'In both species *G. ruber* s.s. and *G. bulloides* larger IQRs are

found toward higher absolute SNW.': Which is perfectly normal stochastic behaviour. This is why it is important to normalize variation for expected value by reporting the coefficient of variation instead of raw variation under such circumstances.

Lines 490–492, 'An opposite trend in SNW [...] growth conditions.': I assume this refers to Beer et al. (2010b). Please cite your sources properly.

Line 494, 'Köhler-Rink and Kühl, 2005': This citation is missing in the list of references.

Lines 496–497, 'additional calcite layers might be added to the proximal text surface before reproduction, similar to the process described for *O. universa* (see above).': Yet to my knowledge, those two species are not known for excessive amounts of gametogenic calcite (e.g. Deuser, 1987; Hamilton et al., 2008). Also, the alternative interpretation would be that more optimal conditions trigger faster growth and earlier reproduction, resulting in a trade-off for calcification intensity of each individual chamber already during growth (i.e. before gametogenic calcite is added). Additionaly, 'text' should be 'test' (Line 496)

Lines 505–506, 'However, the comparison might be biased by the fact that *G. ruber* s.s. and s.l. morphotypes were analyzed together in the study of de Moel et al. (2009).': It most certainly is. Compare Weinkauf et al. (2016).

Lines 514–516, 'All of these [...] in an increased SNW': They also support the interpretation, that a multitude of factors influences shell calcification in planktonic Foraminifera.

Line 517, 'given that carbonate chemistry does not limit calcite formation in planktic foraminifera.': This is a blatant misrepresentation of basically the entirety of existing literature (compare Marshall et al. (2013, tab. 1) and Weinkauf et al. (2016, tab. 7)).

Line 522, 'reflect high': Should be 'show large'.

Line 526, 'ten morphospecies in total.': This is wrong since at least the individual species *G. ruber* (white), *G. ruber* (pink), and *G. elongatus* have been pooled together. Furthermore, it is unclear whether *G. calida* and *G. radians* also occur and have been pooled into *G. siphonifera*.

Line 548, 'These observations highlight the need for more interdisciplinary studies on the causes of changing foraminiferal assemblages and decreasing shell production': If this is supposed to hint at the promised comparison with earlier studies then I must state again that 1) since you used a larger mesh size without correcting your data for that fact you cannot compare your abundances with those of earlier studies and 2) you never presented a thorough discussion whether species compositions have been significantly changing during the last 20 years and if so, why.

Lines 588–589: There is no Bijma et al., 1990a, so remove the 'b' after the year.

Lines 625–626: Ivanova et al. (2003) is not cited anywhere in the manuscript. Remove from list of references.

Lines 650–651: 'Grazzini' should be 'Vergnaud Grazzini'.

Lines 682–683: '*Orbulina universa*' should be set in italics.

Caption Fig. 1, '(a) Temperature (°C), (b) salinity, (c) fluorescence (μg $\cdot$ l^{-1}), (d) pH, and (e) [CO_3^{2-}] (μmol $\cdot$ kg^{-1}): Information where these data come from are missing completely. Additionally, the software used for plotting (I assume Ocean Data View, Schlitzer (2014)) has not been cited. Especially Section 2, and to a lesser extant Section 3 involves a huge amount of interpolation due to the large spatial distance between measurement profiles. This makes the reconstructions very unreliable.

Caption Fig. 2, 'First leg: 1 to 13, second leg: 14 to 22.': It might be nice to distinguish the cruise-tracks of the two legs by colour.

Caption Fig. 2, 'MODIS Aqua (L2),': What is this? This source has not been cited in any way (published article, url, ...) and was not mentioned in the Material and Methods section.

Caption Fig. 2, 'from the closest day as possible': Which means exactly what? 1 day, 10 days, 100 days,...? Also, I would have assumed the dates given in the map are the dates for which chlorophyll *a* data have been plotted, or what else is displayed there?

References

Aurahs, R., Treis, Y., Darling, K., and Kučera, M.: A Revised Taxonomic and Phylogenetic Concept for the Planktonic Foraminifer Species *Globigerinoides ruber* Based on Molecular and Morphometric Evidence, Marine Micropaleontology, 79, 1–14, doi:10.1016/j.marmicro.2010.12.001, 2011.

Barker, S. and Elderfield, H.: Foraminiferal Calcification Response to Glacial–Interglacial Changes in Atmospheric CO_2, Science, 297, 833–836, doi:10.1126/science.1072815, 2002.

Bé, A. W. H.: An Ecological, Zoogeographic and Taxonomic Review of Recent Planktonic Foraminifera, in: Oceanic Micropalaeontology, edited by Ramsay, A. T. S., 1, chap. 1, pp. 1–100, Academic Press, London and New York and San Francisco, 1977.

Beer, Ch. J., Schiebel, R., and Wilson, P. A.: Technical Note: On Methodologies for Determining the Size-Normalised Weight of Planktic Foraminifera, Biogeosciences, 7, 2193–2198, doi:10.5194/bg-7-2193-2010, 2010a.

Beer, Ch. J., Schiebel, R., and Wilson, P. A.: Testing Planktic Foraminiferal Shell Weight as a Surface Water $[CO_3^{2-}]$ Proxy using Plankton Net Samples, Geology, 38, 103–106, doi:10.1130/G30150.1, 2010b.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society, Series B: Methodological, 57, 289–300, doi:10.2307/2346101, 1995.

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.

Broecker, W. and Clark, E.: An Evaluation of Lohmann's Foraminifera Weight Dissolution Index, Paleoceanography, 16, 531–534, doi:10.1029/2000PA000600, 2001a.

Broecker, W. S. and Clark, E.: Reevaluation of the $CaCO_3$ Size Index Paleocarbonate Ion Proxy, Paleoceanography, 16, 669–671, doi:10.1029/2001PA000660, 2001b.

de Villiers, S.: Optimum Growth Conditions as Opposed to Calcite Saturation as a Control on the Calcification Rate and Shell-Weight of Marine Foraminifera, Marine Biology, 144, 45–49, doi:10.1007/s00227-003-1183-8, 2004.

Deuser, W. G.: Seasonal Variations in Isotopic Composition and Deep-Water Fluxes of the Tests of Perennially Abundant Planktonic Foraminifera of the Sargasso Sea: Results from Sediment-Trap Collections and their Paleoceanographic Significance, Journal of Foraminiferal Research, 17, 14–27, doi:10.2113/gsjfr.17.1.14, 1987.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: A Review of Methods to Deal with it and a Simulation Study Evaluating their Performance, Ecography, 36, 27–46, doi:10.1111/j.1600-0587.2012.07348.x, 2013.

Dytham, C.: Choosing and Using Statistics: A Biologist's Guide, Wiley–Blackwell, Oxford and Chichester and Hoboken, 3 edn., 2011.

Faraway, J. J.: Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Texts in Statistical Science, Chapman & Hall and CRC Press, Taylor & Francis Group, Boca Raton, 2006.

Hamilton, Ch. P., Spero, H. J., Bijma, J., and Lea, D. W.: Geochemical Investigation of Gametogenic Calcite Addition in the Planktonic Foraminifera *Orbulina universa*, Marine Micropaleontology, 68, 256–267, doi:10.1016/j.marmicro.2008.04.003, 2008.

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.

Lohmann, G. P.: A Model for Variation in the Chemistry of Planktonic Foraminifera due to Secondary Calcification Selective Dissolution, Paleoceanography, 10, 445–457, doi:10.1029/95PA00059, 1995.

Manno, C., Morata, N., and Bellerby, R.: Effect of Ocean Acidification and Temperature Increase on the Planktonic Foraminifer *Neogloboquadrina pachyderma* (sinistral), Polar Biology, 35, 1311–1319, doi:10.1007/s00300-012-1174-7, 2012.

Marshall, B. J., Thunell, R. C., Henehan, M. H., 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.

McDonald, J. H.: Handbook of Biological Statistics, Sparky House Publishing, Baltimore, 2 edn., http://www.biostathandbook.com/, 2009.

Numberger, L., Hemleben, Ch., Hoffmann, R., Mackensen, A., Schulz, H., Wunderlich, J.-M., and Kučera, M.: Habitats, Abundance Patterns and Isotopic Signals of Morphotypes of the Planktonic Foraminifer *Globigerinoides ruber* (d'Orbigny) in the Eastern Mediterranean Sea since the Marine Isotopic Stage 12, Marine Micropaleontology, 73, 90–104, doi:10.1016/j.marmicro.2009.07.004, 2009.

Pujol, C. and Vergnaud Grazzini, C.: Distribution Patterns of Live Planktic Foraminifers as Related to Regional Hydrography and Productive Systems of the Mediterranean Sea, Marine Micropaleontology, 25, 187–217, doi:10.1016/0377-8398(95)00002-I, 1995.

Schlitzer, R.: Ocean Data View, Alfred-Wegener-Institut, http://odv.awi.de, 2014.

Spezzaferri, S., Kučera, M., Pearson, P. N., Wade, B. S., Rappo, S., Poole, Ch. R., Morard, R., and Stalder, C.: Fossil and Genetic Evidence for the Polyphyletic Nature of the Planktonic Foraminifera *"Globigerinoides"*, and Description of the New Genus *Trilobatus*, PLOS ONE, 10, e0128 108, doi:10.1371/journal.pone.0128108, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128108, 2015.

Steinke, S., Chiu, H.-Y., Yu, P.-S., Shen, C.-C., Löwemark, L., Mii, H.-S., and Chen, M.-T.: Mg/Ca Ratios of two *Globigerinoides ruber* (white) Morphotypes: Implications for Reconstructing Past Tropical/Subtropical Surface Water Conditions, Geochemistry, Geophysics, Geosystems, 6, Q11 005, doi:10.1029/2005GC000926, 2005.

Storz, D.: Die Saisonalität planktischer Foraminiferen im Bereich einer Sinkstoffallenstation in subtropischen östlichen Nordatlantik zwischen Februar 2002 bis April 2004 [The Seasonality of Planktonic Foraminifera in Vicinity of Sediment Trap in the Suptropical North Atlantic between February 2002 and April 2004], Diploma thesis, Eberhard–Karls Universität Tübingen, Tübingen, 2006.

Storz, D., Schulz, H., Waniek, J. J., Schulz-Bull, D. E., and Kučera, M.: Seasonal and Interannual Variability of the Planktic Foraminiferal Flux in the Vicinity of the Azores Current, Deep-Sea Research, Part I: Oceanographic Research Papers, 56, 107–124, doi:10.1016/j.dsr.2008.08.009, 2009.

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.

Weiner, A. K. M., Weinkauf, M. F. G., Kurasawa, A., Darling, K. F., and Kučera, M.: Genetic and Morphometric Evidence for Parallel Evolution of the *Globigerinella calida* Morphotype, Marine Micropaleontology, 114, 19–35, doi:10.1016/j.marmicro.2014.10.003, 2015.

Weinkauf, M. F. G., Kunze, J. G., Waniek, J. J., and Kučera, M.: Seasonal Variation in Shell Calcification of Planktonic Foraminifera in the NE Atlantic Reveals Species-Specific Response to Temperature, Productivity, and Optimum Growth Conditions, PLOS ONE, 11, e148 363, doi:10.1371/journal.pone.0148363, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148363, 2016.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-266, 2016.