Clim. Past Discuss.,
https://doi.org/10.5194/cp-2019-9-RC1, 2019

Interactive comment on "On the validity of foraminifera-based ENSO reconstructions" *by* Brett Metcalfe et al.

Anonymous Referee #1

Received and published: 25 March 2019

SUMMARY

This manuscript by Metcalfe et al. tries to understand where in the tropical Pacific Ocean the analysis of distributions of individual foraminifera geochemistry results (d18Oc, T(MgCa)) could be used to find changes in the El Niño-Southern Oscillation. This type of analysis has been done before (Thirumalai, Ford, White), with a focus on the inverse problem of estimating ENSO change from individual foraminifera distributions. Here the novelty is the inclusion of a forward model of foraminifera growth rate. This model is used to estimate the biased sampling in depth and time that different foraminiferal species have, and how this contributes to the analysis of the ENSO signal. However, that part of the model is not really validated and it is unclear how much it adds to the analysis. Furthermore, the statistical analysis focuses on a forward prob-

lem rather than the inverse problem that is the real challenge for detecting changing ENSO from individual foraminiferal analysis. The forward problem is whether El Nino, neutral, and La Nina months have different distributions and requires that each individual d18Oc or T value be assigned beforehand to one of those three states. The inverse problem is to determine from comparison of two different d18Oc or T distributions (as would be measured in two sediment samples) whether any change in their distributions occurred and whether it can be ascribed to changes in the statistics of ENSO events (frequency, magnitude). Finally, there are also additional questions in the author's methodology that are opaque and need to be clarified.

As it stands, the focus on the forward problem and on statistical approaches not used for paleo-IF analysis make the manuscript in its present form not a good evaluation of the IF approach for ENSO reconstruction. The title is misleading and the abstract mis-states the conclusions from the study. For the reasons above and detailed below it is difficult to evaluate the utility and applicability of the manuscript to the questions the authors raise. With different analyses the authors could address the questions they pose. However, it could be very different from the manuscript in its current form and in my opinion would need to independently evaluated and reviewed.

GENERAL COMMENTS

1. Make sure that the abstract and conclusions follow from the analyses and are properly stated (see specific notes below).

2. Validate growth rate forward model

Validate the growth rate calculation through comparison with sedimentary relative abundances. This was done to some extent in the paper cited for the foraminifera model (Roche et al., 2018) but in that paper no clear assessment of the errors was presented. The model in Roche et al., 2018 is a simplification of earlier growth rate modeling of foraminifera. In Roche 2017 all parameters besides temperature are discarded. How well then does the model work? I think the authors should use the modeled growth rate for the species they are targeting and calculate the relative abundance of those three species in a sediment sample. This can be compared to the measured relative abundance of those three species (summing to one) recalculated from their relative abundance amongst all species counted in coretop datasets. This should be shown as a scatterplot of observed vs. predicted on x- and y-axes rather than on a map as is shown in the supplement to Roche et al., 2018.

3. Validate the d18Oc predictions

Validate the d18Oc predictions from the growth rate and geochemistry model. This was done in Roche et al., 2017 but is also somewhat circular because the sedimentary d18O values were used to determine the depth of production. I admit I am not sure how to actually validate the approach except from an additional validation dataset not used for determining production depth.

4. Include the analysis of Tc for Mg/Ca reconstructions

Inexplicably the authors refuse to analyze the temperature distributions even though those are the data from the common Mg/Ca method of individual foraminifera analysis (Sadekov, Ford, White). The author's stated reason is due to "…the complexity in reconstructions of trace metal geochemistry…and the potential error associated with determining which carbonate phase is first used when foraminifera biomineralise…". While there are ongoing methodological and calibration efforts for this and other proxies (including d18Oc), to ignore such a widespread type of analysis seems very short-sighted. If the authors do not want to forward model the Mg/Ca proxy itself they can simply analyse the temperatures in their dataset. Either way this is something that should be included in the manuscript.

5. Remove maps of carbonate preservation/depth

It is fine for the authors to state the general problem in the text, but there are regions of shallow depth were carbonate is preserved that are not captured in the coarse DEM

used.

6. Remove map of sedimentation rate

Either quantitatively discuss the role of sedimentation rate and bioturbation or remove this map. The sedimentation rate threshold is intimately tied to the secular and non-ENSO variability and thus is a much more complicated analysis than the general discussion in the text. I think the discussion is a starting point but the author's miss that the important factors are really the magnitude of other, non-ENSO sources of variability at the timescale of a sediment sample (plus bioturbation) compared to the magnitude of the ENSO change signal and the non-ENSO variability.

7. Focus in the inverse problem

It is really the inverse problem of detecting a change in ENSO from a change in the distribution of foraminifera d18Oc or T that is the focus of IF ENSO reconstructions. The analysis in this paper basically asks the question: are the distributions from El Niño months different from neutral or La Niña months? This is a useful first step in the inverse problem but it doesn't really answer the question stated in the title about the validity of foraminifera-based ENSO reconstructions.

8. Apply statistical tests on parameters used on paleo-IF distributions

The author's use Anderson-Darling tests for differences in distributions. They should demonstrate how this might be useful for paleo-IF analysis. It would also be greatly to their advantage to test the approaches actually used for paleo-IF analysis (1-sigma, quantiles) to see how they perform in this framework. A welcome contribution would be demonstration that a new/different type analysis from those typically applied to paleo-IF distributions is better. As it stands, the focus on the forward problem and on statistical approaches not used for paleo-IF analysis make the manuscript in its present form not a good evaluation of the IF approach for ENSO reconstruction.

9. Definition of El Niño, neutral, and La Niña months

There is a large body of literature and accepted methods for defining El Niño, neutral, and La Niña periods. In the text the authors take a simplistic approach, but there is no reason for this. Why not actually use the societal and dynamically important definitions of these events including the requirement of a minimum consecutive number of months of anomalies and changing baseline for anomalies (to account for secular warming of the ocean)? This definition has a basis in theory as an El Nino (La Nina) event unfolds over a length of time and thus a single month anomaly may not be associated with the dynamics that are part of the coupled ENSO system.

10. Clearly separate the role that the growth model and (T,S) timeseries play in identifying ENSO change.

To what degree are the outcomes and conclusions of this paper depending upon the modeled growth rates versus the sea water properties (T,S,d18Ow)? Many prior workers have analyzed in different ways the reconstruction of ENSO from IF analysis. These approaches include summary statistics like the standard deviation (Thirumalai; Koutavas; Leduc; Sadekov; Rustic), as well as examination of changes in the quantiles of IF distributions (Ford; White). What is added here is the foraminifera growth rate weighting. What effect does this have? From the histograms in Roche et al. (2018) it appears that the growth-rate weighting does not have major consequences for the mean d18Oc value of a sediment sample. It may have consequences for the IF variability though. The authors could show a map that quantifies the growth-rate weighting effect with respect to the non-weighted results (ratio, difference).

11. Examine how ENSO amplitude vs. frequency change IF distributions

The authors raise an interesting point in their conclusion that has not been well addressed, namely how do changes in the statistics of ENSO (frequency, amplitude) affect IF distributions and reconstructions of ENSO variability. Evaluating these two different questions would be an important contribution to IF analysis of ENSO change. But, introducing the idea in the conclusions without a previous discussion in the manuscript

is not a good idea in my opinion.

SPECIFIC COMMENTS ON TEXT

– Abstract –

Page 1 line 15 “Our results show that it is possible to use d18Oc from foraminifera to disentangle the ENSO signal only in certain parts of the Pacific Ocean.” This line in the abstract is sharply in contrast to the line in the conclusion at Page 12 line 21 “Overall, our results suggest that foraminiferal d18O for a large part of the Pacific Ocean can be used to reconstruct ENSO.” Which is it?

Page 1 line 17 – “Furthermore, a large proportion of these areas coincide with sea-floor regions exhibiting a low sedimentation rate and/or water depth below the carbonate compensation depth, thus precluding the extraction of a temporally valid palaeoclimate signal using long-standing palaeoceanographic methods.” The role of sedimentation rate in IF analysis is important but there is not any investigation of this effect in the present manuscript so it is not really a conclusion or finding. This statement should not be included in the paper in its present form.

Page 1 line 17 – “Furthermore, a large proportion of these areas coincide with sea-floor regions exhibiting a low sedimentation rate and/or water depth below the carbonate compensation depth, thus precluding the extraction of a temporally valid palaeoclimate signal using long-standing palaeoceanographic methods.” The role of water depth and carbonate preservation is also important. But, there is not any investigation of the sedimentation rate effect in the present manuscript so it is not really a conclusion or finding. Furthermore, there are seamounts and other shallow sites not captured in the gridded dataset that can contain records for palaeoceanographic investigations. This statement should not be included in the paper in its present form.

– Main Text –

Page 3 line 23 – Here the authors introduce the 1-sigma d18Oc parameter than has been used in some studies to look at changes in ENSO variance. But, they never really address whether this parameter is useful and can detect changes in ENSO. Thirumalai et al. (2013) took this question on already. More discussion of what has been done previously is needed. Also, why not test the actual way that IF analysis is used (e.g. 1-sigma, quantiles etc.) rather than a new method as introduced here (Anderson-Darling test)?

Page 4 line 1 – The new model for foraminifera growth only uses the temperature component of the previous model. Why? How different are the results?

Page 4 line 15 – Allowing symbiont-bearing foraminifera to possibly grow to 400 m simply based upon optimal temperatures seems not correct. They need to be in the photic zone.

–Methods–

Page 5 line 5 – The conversion of VSMOW to VPDB looks to be in error. The correct formula for this conversion is [d18O_VSMOW+1]/[d18O_VPDB+1] = 1.03091 where d18O does not include the 10ˆ3 term. Thus d18O_VPDB = d18O_VSMOW/1.03091 +(1/1.03091)-1 or d18O_VPDB = 0.97002*d18O_VSMOW - 0.02998. In d18O expressed with the 10ˆ3 term, the equation would read: d18O_VPDB = 0.97002*d18O_VSMOW - 29.98.

Page 5 line 10 – Why was growth rate arbitrarily constrained to these different depths? First, foraminifera with algal symbionts should be in the photic zone. Second, didn't the Roche et al., 2018 paper try to identify the depth-production relationship for the different species from the predicted d18Oc and measured MARGO d18Oc? Why not use those depths?

Page 6 line 1 – "…these for now can be ignored." Why can the other factors determining foraminifera growth be ignored? This cannot be a statement unless it is backed up. Or, the authors use only temperature but then go through an appropriate validation

process (more than what is shown in Roche et al., 2018) as suggested above.

Page 6 line 11 – Starting here, it is very unclear how and why the particular set of conditions for El Niño, La Niña, and neutral periods were chosen. What time series of sea surface temperatures were chosen for computing anomalies (in each grid square, Nino 3.4, Nino 3, Nino 4, etc.)? Were the anomalies based upon a 3-month running mean? Were the anomalies computed relative to a fixed period or, as is now the accepted approach, relative to 5-year interval means? Why not use the definition of El Nino etc. events that include the requirement for consecutive months of anomalies? This definition has a basis in theory as an El Nino (La Nina) event unfolds over a length of time and thus a single month anomaly may not be associated with the dynamics that are part of the coupled ENSO system.

Page 6 line 18 – Why and how was the pdf/cdf from the actual data fitted and smoothed with an Epanechnikov kernel? What impact did this fitting and smoothing (particularly the choice of bandwidth) have on the Anderson-Darling test and the results overall?

Page 6 line 24 – This paragraph is very unclear and the errors associated with binning prior to analysis of the pdf seem avoidable. For example, why not take the growth rate in each of the 696 months in each grid at each depth, and scale the growth rate to calculate an effective # of individuals such that they sum to 1000 across all months? Round those numbers to integers and then use the integer # of individuals for each month to replicate that actual months Tc or d18Oc value. The resulting ordered list of values can then be binned/smoothed etc. and represents a pseudo-distribution that one might find in a sediment sample?

–Results–

Page 7 line 3 – It says that the mean d18Oc for El Nino and neutral months are compared. How? Earlier and later it is stated that the A-D test is applied to compare distributions. What is meant by these lines?

Page 7 line 5 – "...ENSO events can potentially be detected by paleoceanographers and unmixed using, for example, a simple mixing algorithm with individual foraminiferal analysis..." This is not really practicable because it assumes complete stationarity in the El Nino, La Nina, and neutral distribution. This is unlikely as all are expected to change, and do in models and data (e.g. coral time series from middle Holocene show changed seasonal amplitude and ENSO cycles).

Page 7 line 7 –"In cases where FPEN and FPNEU do not exhibit significantly different means, then the chosen species and/or location represent a poor choice to study ENSO dynamics." This may not always be the case because the mean values could be similar but the distributions wildly different (such as long tails with different signs). Changing numbers of El Nino and neutral and La Nina events could that quite dramatically change the shape of the combined distribution that is ultimately preserved in sediments. And, it may be possible to find regions of such a distribution that can be used to diagnose changing ENSO.

Page 7 line 20 – Why is Anderson-Darling test done here but the mean values are discussed above? If the A-D test shows that the El Nino and Neutral distributions are different (at some statistical level) then that means alteration of those distributions (more/fewer, stronger/weaker events) would alter the summed distributions that one gets from a sediment sample. But, how would this actually be detected in the sediment sample? That the AD test demonstrates the El Nino, Neutral, and/or La Nina distributions are different is helpful but it does not get at whether ENSO change could actually be detected in a sediment sample.

Page 7 line 26 – Applying a 1-sigma value from modeled minus coretop comparisons to the AD test value does not seem appropriate. This value assesses the accuracy of the model in predicting the absolute value of the mean of a coretop sample. But it is not an appropriate estimate for the significance of the difference between two different IF values or the difference in the AD statistic.

Page 8 line 1 – This paragraph is rather confusing to understand. It sounds like the authors are comparing a depth-weighted reconstruction and non-depth weighted reconstructions at fixed depths (Fig. 3 vs. 4)?

Page 8 line 9 – Unclear what "on the low-end" means.

Page 8 line 16 – "...a large percentage of the tropical Pacific remains accessible to palaeoclimate studies." This is very much not the message in the Abstract and from the title of the paper. Those sections should reflect this finding.

Page 8 line 25 – "Indeed, one should view discrete sediment intervals, and the foraminifera contained within them, as representative of an integrated multi-decadal or even multi-centennial signal..." This is exactly how foraminifera paleo-IF studies have viewed them and should be stated up front (start of paragraph for instance).

Page 8 line 28 – "Therefore, in order to reliably extract short-term environmental information from foraminiferal-based proxies, the signal that one is testing or aiming to recover must exhibit a large enough amplitude in order to perturb the population by a significant degree from the background signal, otherwise it will be lost due to the smoothing effect of bioturbation..." This statement does not make sense to me. The background signal IS the signal, i.e. the seasonal cycle, ENSO etc. Changes in ENSO must be such that they alter that signal (the distribution of IF analyses), but bioturbation etc. should not erase the signal unless one is looking for short periods of change less than the time integrated into the sample.

Page 9 line 6 – "...a series of high magnitude, but low frequency El Niño events could be smoothed out of the downcore, discrete-depth record." They will not be smoothed out as the authors state. Those anomalous IF values may be rare, but will be present in the sediment sample and if measured can be used to examine changing ENSO.

Page 9 line 7 – The sediment accumulation rate needed to observe/reconstruct changes in ENSO is not fixed. It depends upon the magnitude and duration of secular trends, and variability with respect to both the time integrated in a sediment sample and the magnitude of the ENSO signal and its change. This is a quite interesting but also complicated subject and arbitrarily cutting the sedimentation rate at 5 cm/ky is not justified.

Page 9 line 9 – The map of water depth is quite coarse and misses important locations that are above the CCD, accumulate carbonate (and foraminifera), and can be used for palaeoceanographic reconstructions. Thus, while the overall point is true, the map as shown is misleading.

–Discussion–

Page 9 line 20 – Why discount trace metal temperatures (Mg/Ca)? At least use the Tc analysis as a stand-in for Mg/Ca.

Page 9 line 23 – The focus of this paper is on IF analysis. Why are the Koutavas and Lynch- Stieglitz, 2003; Koutavas and Lynch-Stieglitz, 2003 etc. cited here? The whole discussion in this paragraph, lines 16-31 feels out of place.

Page 10 line 3 – The references to Cole and Tudhope, 2017; White et al., 2018 seem to be in error. These papers do not discuss lake core colour etc.

Page 10 line 3 – "If the number and magnitude of ENSO events were reduced, the relatively low downcore resolution of marine records may not accurately capture the dynamics of such lower amplitude ENSO events using existing methods." This statement is not justified by the author's analysis or a citation.

Page 10 line 5 – "The possibility of a marine sediment archive being able to reconstruct ENSO dynamics comes down to several fundamentals: the time-period captured by the sediment intervals (a combination of SAR and bioturbation), the frequency and intensity of ENSO events, as well as the foraminiferal abundance during ENSO and non-ENSO conditions." Also included is the magnitude of change in ENSO statistics and resulting foramifera Tc or d18Oc, sampling uncertainty on the IF distribution. See

also note above on the role of sedimentation rate.

Page 10 line 9 – "The results presented here imply that much of the Pacific Ocean is not suitable for reconstructing ENSO studies using palaeoceanography, yet several studies have exposed shifts within σ(d18Oc) of surface and thermocline dwelling foraminifera. One can, therefore, question what is being reconstructed in such studies." The results presented here don't really test whether individual foraminifera d18Oc (or Tc) studies can reconstruct ENSO. Furthermore, the water depth and sedimentation rate constraints are the reason for excluding much of the Pacific. This statement is therefore incorrect and the search for other explanations does not follow.

Page 10 line 19 – This second part of the paragraph is interesting and has been commented on before. But, at no point do the authors actually evaluate any of these effects or approaches so they can't really assess the different factors they raise here.

Page 10 line 28 – The discussion of model limitations does not ask what would seem to be the most important questions: Does the modeled growth rate actually reflect the real ocean (and the sampling bias for what is recorded in sediments)? Do the modeled growth-rate weighted d18O distributions match actual measured individual foraminifera d18O distributions (such as in Koutavas and Joanides or Rustic)? If no growth-rate weighting is applied are the results better or worse?

Page 11 line 14 – Why are the authors so dismissive of Mg/Ca analyses? The list of possible complications is important but it remains a fundamental observation that the Mg/Ca of foraminiferal calcite changes with growth temperature and has been validated in many different ways.

–Conclusion–

Page 12 line 17 – "Previous work. . ." The only citation here is to Zhu et al., 2017. There has been a lot of work comparing IFA different time slices (both d18Oc and Mc/Ca) that should be cited here (Koutavas et al., Leduc, Koutavas and Joanides, Sadekov et al, Ford et al, Rustic et al, White et al). Furthermore, they have not all used 1-sigma d18Oc as the metric for detecting change.

Page 12 line 21 – "Overall, our results suggest that foraminiferal δ18O for a large part of the Pacific Ocean can be used to reconstruct ENSO. . ." This contradicts what is said in the abstract and in some places in the text (but is similar to in other places in the text). Which is it?

Page 12 line 22 – What is meant by ". . .especially if an individual foraminiferal analysis. . . approach is used. . ." I thought the whole analysis in the paper was on whether individual foraminifera analysis can be used? Was there another method tested (for example the means analysis referred to at Page 7 line 3)?

Page 12 line 24 – "However, the sedimentation rate of ocean sediments in the region is notoriously slow (Olson et al., 2016) and much of the ocean floor is under the CCD. These factors reduce the size of the area available for reconstructions considerably (Lougheed et al., 2018), thus precluding the extraction of a temporally valid palaeoclimate signal using long-standing methods." This is generally true, but there are seamounts and other regions that may actually preserve carbonate. Furthermore, the sedimentation rate constraint is also somewhat arbitrary and depends upon secular trends and non-ENSO variability encompassed in a particular sample.

Page 12, line 27 – "We further highlight that the conclusions drawn from foraminiferal reconstructions should consider both the frequency and magnitude of El Niño events during the corresponding sediment time interval (with full error) to fully understand whether or not a strengthening or dampening occurred." While this is true, nowhere in the manuscript is this issue addressed. Inclusion as a conclusion to the paper is therefore not warranted.

Page 12 line 30 – "The use of ecophysiological models. . .are not limited to foraminifera and provide an important way to test whether proxies used for palaeoclimate reconstructions are suitable for the given research question." This is not really a conclusion

of the study. And, given the uncertainties and lack of rigorous testing of the foraminifera model in this study, this is a questionable statement overall.

–Figures–

Figure 3 – Why are there white and grey areas that mean the same thing?

Figure 4 – Are the temperature data growth weighted? What species? If not, why not analyze the Tc data in parallel to the d18Oc data to evaluate what advantage/disadvantage the two different signals have (e.g. from S).

Figure 5 – Why are the white and grey areas grouped together? What do they mean? Are these panels based upon growth-rate weighted values?

Figure 6 – Are these panels based upon growth-rate weighted values?

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2019-9, 2019.