

## Track changes MS

### Comments to the Author:

Editor Decision: Reconsider after major revisions (03 Jun 2019) by Eric Wolff

*Your paper received a number of quite negative reviews. Nonetheless, I have decided to allow you to prepare a new version (major revisions) for consideration by reviewers.*

*I accept that there is value in a paper that forward models the proxy record to assess whether ENSO characteristics are even recorded in forams, as opposed to an inverse model that seeks to determine ENSO from the foram record. Nonetheless, you should recognise that several well-qualified reviewers felt that your study had not met the targets expressed in the title and abstract. Rather than (as your responses tended to) berate them, I would like to see you improve your presentation and explanations so that readers are not left disappointed.*

*This applies also to other aspects of the paper: as an example, like some reviewers I found the figures and the different shadings hard to understand; please improve either the figures or the captions, which is more constructive than implying that the reviewers have been deficient in not being able to follow them.*

*In your responses you have raised a number of issues that you state were not part of your paper - for example bioturbation, and alternative statistical methods. While I do not expect you to change your model or methodology completely to meet these concerns, you clearly do need to add paragraphs to your paper to state what might be the effect if these issues were covered, and why their lack does not affect your main findings.*

*I look forward to receiving a substantially improved paper for further review.*

### Response to Editor

Dear Prof. Eric Wolff,

Thank you for considering our manuscript for revisions. Please find attached our revised MS with track changes. As the 'track changes' document is heavily altered it may obscure the content, therefore we outline here the major changes in brief. We also include some rationale as to why we made the changes.

- We have changed the Title.
- We have altered the text of the abstract and conclusions to better reflect the fact that whilst 'foraminifera in the water' may 'sense' ENSO, the sedimentology of the Pacific reduces the useful area.
- We have edited the methodology and streamlined the explanation, as this is where the text may have led the reviewers to misinterpret our intentions.
- We have calculated the standard deviation/variance of each model run and each species, we have also calculated the interannual std. dev. vs. the total std. dev. We include a figure of *G. sacculifer* at 60 m and a supplementary table 1 in which we compare our results with three previous papers data and one model (ICESM)
- Differences with our model and data may reflect: Smaller standard deviation for larger n; the fact our dataset is in the "Anthropocene" whilst the data is from ~1000 years ago; there are reversals or evidence for bioturbation in the cores; subsample picking variation; and or the fact that our model may have limitations in terms of food, light, etc. (as stated in the discussion and methodology).

- We have run 4 picking experiments (picking 60 months, 10000 times at each grid cell: 170\*40\*60\*10000). The four picking experiments highlight different ways of subsampling the model dataset.
- Unfortunately, it is not possible as the reviewers suggested to use QQ-plots as it would require 6800 plots; and interpolation of the skewed climate state data. We have included an explanation, as per your request, as to why we have not done this.
- We have moved the test of the input data from the supplement into the main text (figure 4) – this is not FAME instead we perform the Anderson-Darling on fixed depths. We have expanded and altered the original text to highlight that we had in fact tested the input signal (the reviewer had requested to see what the difference is between FAME and the input)
- Water depth map – as suggested by the reviewers, we have plotted seamount locations from the only scientifically verified (as opposed to text-mining Wikipedia entries for seamounts) available database of seamounts in the Pacific. An estimated 50,000 seamounts exist in the Pacific, but unfortunately this database contains only 291 values (and these include duplicate locations pertaining to be different parts of the same seamount system). Complementary to this, we have also added a new figure, to highlight that the GEBCO dataset is not a low resolution dataset, allowing us to show where the different water depth cutoffs for CCD alter or expand the available area suitable for sediment core retrieval.
- We have plotted the FAME overlay (figure 7) only with the water depth (previously it was with the SAR plus water depth). This does not alter our results. Instead we now contrast the water depth with different variations of SAR (figure 9) and highlight the potential altered signal through bioturbation of each scenario (further expanded upon in figure 10). This is intended to better explain our results.
- Sedimentation rate map - We have expanded the figures to include the inferred sediment accumulation rate, the oxygen saturation of the bottom water (unfortunately the 2018 pre-release of WOA as yet does not include this oxygen parameter in netcdf format although we will happily revise this if it becomes available) in response to a reviewers comment that oxygen will alter the bioturbation depth, and bioturbation model output. The main text was expanded to reflect those changes.
- We have run a bioturbation model, SEAMUS (model code and outline has been submitted by co-author Bryan Loughheed to Geoscientific Model Discussions and is online: Loughheed, B. C.: SEAMUS (v1.0): a  $\Delta 14\text{C}$ -enabled, single-specimen sediment accumulation simulator, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2019-155>, in discussion, 2019.), using the Uppsala Supercomputer and added corresponding new figures and text. As outlined in our responses to the reviewers, a 60-year ocean reanalysis dataset is too short for a bioturbation model (due to thousands of years of model spin-up time being necessary to simulate a sediment archive). We have, therefore, used a Greenland ice core record as an input dataset at 10-year intervals for several thousand years, simply to understand the effects of bioturbation. The use of this dataset is in no way meant to infer some connection between ENSO and ice cores. It was used because it represented a temporally high-resolution climate input dataset with which to demonstrate the effects of bioturbation. The equivalent marine record, the benthic stack LR04, is already bioturbated, seeing as it is a stack of sediment cores with an average SAR of  $\sim 2.5 \text{ cm ka}^{-1}$
- We have, as suggested by the reviewer, included a discussion point of the potential amplitude changes, a point we had raised in the conclusions regarding future work.
- Figures - As outlined in our response to reviewer 3, we agree that some figures could have been clearer (As we had stated in the reply: *"We will correct this – the label on the figure is correct (left G. sacculifer; mid G. ruber; right N. dutertrei), we will correct the figure captions. The reviewer is correct that some are white and some grey when we originally made the figure everything was white, black and hashed. Unfortunately, these hashes draw the eye (and can hide some small locations) away from the white only locations we decided to make it grey to highlight this (The top panel of Figure 5 'G. sacculifer 60 m' is missing the grey which is our mistake but demonstrates the drawing of the eye). The N. dutertrei dataset does not have the hashing so we decided to make it white only."*). We have corrected the error in the caption and one panel of figure 5. I hope the addition of an explanatory legend in our figures can solve this problem.
- As explained to the reviewer the rationale for having white vs. black OR grey and hashed vs. black is threefold: white background with a hashed section can draw the eye; the two schemes are meant to be distinct if we used the grey (means are greater than the error) for N. dutertrei one could infer that those results are comparable with G. ruber and G. sacculifer grey regions (though we do not know that); and the binary grayscale approach was used for its contrast.

- The color scheme white, grey, black, hashed can be changed with only a minor amount of work – however, if our explanatory legends are still not clear we would appreciate input from the reviewers in terms of what colour scheme to use. Likewise, the color schemes of 7 and 8 (i.e., yellow/blue-purple) can be altered as well, however we felt that this nicely picks out the bathymetric highs. In addition, Figure 9C; 9E and 9G color scheme can also be altered, however the red grey and yellow color scheme works well with dark and light contrast. This allows for 6 combinations to be plotted on one figure.

*On behalf of all-co-authors*

Brett Metcalfe

Bryan Lougheed

Claire Waelbroeck

Didier Roche

On the validity of foraminifera-based ENSO reconstructions Using a foraminiferal ecology model to test if tropical Pacific planktonic foraminifera are suitable recorders of ENSO

5 Brett Metcalfe<sup>1,2</sup>, Bryan C. Lougheed<sup>1,3</sup>, Claire Waelbroeck<sup>1</sup> and Didier M. Roche<sup>1,2</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>2</sup>Earth and Climate Cluster, Department of Earth Sciences, Faculty of Sciences, VU University Amsterdam, de Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

10 <sup>3</sup>Department of Earth Sciences, Uppsala University, Villavägen 16, 75236 Uppsala, Sweden

Correspondence to: Brett Metcalfe ([brett.metcalfe@lsec.ipsl.fr](mailto:brett.metcalfe@lsec.ipsl.fr) or [b.metcalfe@vu.nl](mailto:b.metcalfe@vu.nl))

A complete understanding of past El Niño-Southern Oscillation (ENSO) fluctuations is important for the future predictions of regional climate using climate models. Reconstructions of One approach to reconstructing past ENSO dynamics uses carbonate oxygen isotope ratios ( $\delta^{18}\text{O}_c$ ) and trace metal geochemistry (Mg/Ca) recorded by planktonic foraminifera as recorders of past climate to assess reconstruct past spatiotemporal changes in upper ocean conditions. In this paper we utilise a model of planktonic foraminifera populations, *Foraminifera as Modelled Entities* (FAME), to forward model the potential monthly average  $\delta^{18}\text{O}_c$  and temperature signal proxy values for *Globigerinoides ruber*, *Globigerinoides sacculifer*, and *Neogloboquadrina dutertrei* from input variables covering the period of the instrumental record. We tested whether the modelled foraminifera population  $\delta^{18}\text{O}_c$  and  $T_e$  associated with El Niño events statistically differ from the values associated with other climate states. For these foraminiferal species our results suggest that the values of El Niño events can be differentiated from other climate states. Our model computes the proxy values of foraminifera in the water, suggesting that, in theory, water locations for a large proportion of the Tropical Pacific should be suitable for differentiating El Niño events from other climate states. However, in practice it may not be possible to differentiate climate states in the sediment record. Specifically, comparison of our model results with the sedimentological features of the Pacific Ocean shows that a large proportion of the hydrographically/ecologically suitable regions, coincide with low sediment accumulation rate at the sea floor and/or regions that lie below the critical water depths for calcite preservation (lysocline and CCD).

We investigate whether planktonic foraminifera-based proxies offer sufficient spatiotemporal continuity with which to reconstruct past ENSO dynamics. Concentrating upon the period of the instrumental record, we use the *Foraminifera as Modelled Entities* model to statistically test whether or not  $\delta^{18}\text{O}_c$  and the Temperature signal ( $T_e$ ) in planktonic foraminifera directly records the ENSO cycle. Our results show that it is possible to use  $\delta^{18}\text{O}_c$  from foraminifera to disentangle the ENSO signal only in certain parts of the Pacific Ocean. Furthermore, a large

Formatted: Font: Italic

Formatted: Not Highlight

Formatted: Font: Italic, Not Highlight

Formatted: Not Highlight

Formatted: Font: Italic, Not Highlight

Formatted: Not Highlight

Formatted: Font: Italic, Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Subscript, Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

~~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.~~

## 1.0 Introduction

### 1.1 El Niño-Southern Oscillation (ENSO)

Predictions of short-term, abrupt changes in regional climate are imperative for improving ~~the~~ spatiotemporal precision and accuracy when forecasting future climate. Coupled ocean-atmosphere interactions (wind circulation and sea surface temperature) in the tropical Pacific, collectively known as the El Niño-Southern Oscillation (ENSO) on interannual timescales and the Pacific Decadal Oscillation on decadal timescales, represent global climate's largest source (Wang et al., 2017) of inter-annual climate variability (Figure 1). Due to ENSO's major socio-economic impacts upon pan-Pacific nations, which, depending on the location, can include flooding, drought and fire risk, it is imperative to have an accurate understanding of both past and future behaviour of ENSO (Trenberth and Otto-Bliesner, 2003; Rosenthal and Broccoli, 2004; McPhaden et al., 2006). The instrumental record of the past century provides important information (~~is~~ that can be translated into the Southern Oscillation Index; SOI), however, detailed oceanographic observations of the components of ENSO (both the El Niño and Southern Oscillation), however, detailed observations of ENSO, such as the Tropical Oceans Global Atmosphere (TOGA; 1985-1994) experiment only provide information from the latter half of the twentieth century (Wang et al., 2017). To acquire longer records, researchers must, ~~therefore~~, turn to the geological record using various archives that are available from the (pan-)Pacific region, including such as: corals (Cole and Tudhope, 2017); foraminifera (Ford et al., 2015; Garidel-Thoron et al., 2007; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Leduc et al., 2009; White et al., 2018); stalagmite (Asmerom et al., 2007; Zhu et al., 2017b); fish detritus (Patterson et al., 2004; Skrivaneck and Hendy, 2015); lake (Anderson et al., 2005; Bennett et al., 2001; Benson et al., 2002; Conroy et al., 2008; Enzel and Wells, 1997; Higley et al., 2018; Loubere et al., 2013; Zhang et al., 2014); terrestrial (Barron et al., 2003; Barron and Anderson, 2011; Caramanica et al., 2018; Hendy et al., 2015; Staines-Urías et al., 2015); ~~and~~ sedimentological parameters (Moy et al., 2002) including varves (Du et al., 2018; Nederbragt and Thurow, 2001, 2006) to reconstruct long-term variations in proxies, linked to climate, that may provide clues to ENSO and its impact upon both regional and global sea surface temperature climate (SST). An integrated approach combining palaeoclimate proxies (Ford et al., 2015; Garidel-Thoron et al., 2007; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Leduc et al., 2009; White et al., 2018) and computer models (Zhu et al., 2017a) can help shed light on the triggers of past ENSO events, their magnitude and their spatiotemporal distribution. Yet, the ~~reconstruction-simulation~~ of past ENSO using climate models has been fraught with difficulties due to the associated feedbacks of ENSO upon model boundary conditions (e.g., SST, pCO<sub>2</sub>) (Ford et al., 2015). One way to deduce the relative impact and importance of various feedbacks and, in turn, reduce model-dependent noise in our predictions, is to compare model output with proxy data (Roche et al.,

Formatted: Heading 2

Formatted: Font: Not Bold

Formatted: Highlight

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

2017; Zhu et al., 2017a). Such an approach, however, requires an abundance of reliable spatiotemporal proxy data from the entire Pacific Ocean. Moreover, such proxy reconstructions are subject to several unknowns, uncertainties and biases, especially given that foraminiferal species preferentially flourish in a particular ecological niche. A major factor governing the spatiotemporal distribution of a given planktonic foraminiferal species is the presence of an ideal water temperature.

5 Here, we investigate whether planktonic foraminifera (The use of ecophysiological models (Kageyama et al., 2013; Lombard et al., 2009, 2011) 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.

10 ~~-based proxies offer sufficient spatiotemporal continuity with which to reconstruct past changes of ENSO. Firstly, we investigate if planktonic foraminiferal proxies have the theoretical capacity to reconstruct ENSO dynamics (Figure 2), and secondly, if the sedimentation rate (Berger, 1970a, 1971; Boltovskoy, 1994; Loughheed et al., 2018; Olson et al., 2016) or the water depth (Berger, 1967, 1970b; Boltovskoy, 1966; Loughheed et al., 2018) (and therefore potential for dissolution of carbonate sediments) in the Pacific Ocean enables the capture and preservation of the foraminiferal signal (Figure 3).~~ 1.2 Foraminiferal Proxies

15 For the specific case of foraminifera populations in the water, it particularly arises from the species-specific ecological niche. The mapping of proxy value to climate value can therefore be skewed, a major factor governing the spatiotemporal distribution of a given planktonic foraminiferal species is the presence of an ideal water temperature. Proxies of past ENSO and Pacific SST (Ford et al., 2015; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Leduc et al., 2009; Sadekov et al., 2013; White et al., 2018) are based upon the biomineralisation of the calcite, or a polymorph such as verite (Jacob et al., 2017), and shells of foraminifera (Emiliani, 1955; Evans et al., 2018; Zeebe and  
20 Wolf-Gladrow, 2001). In general, there are three major types of foraminifera-based palaeoceanographic proxies:

(1) those associated with the faunal composition and their abundance within deep-sea sediments that utilises either a qualitative approach (Phleger et al., 1953; Schott, 1952); a weighted average (Berger and Gardner, 1975; Jones, 1964; Lynts and Judd, 1971); a selected species approach (e.g. coiling direction, or warm-water species presence; Ericson et al., 1964; Ericson and Wollin, 1968; Hutson, 1980b; Parker, 1958; Peeters et al., 2004; Ruddiman, 1971; Schott, 1966); a regression analysis (Hecht, 1973; Imbrie and Kipp, 1971; Williams and Johnson, 1975); or, a transfer function (CLIMAP Project Members, 1976; McIntyre et al., 1976; Williams, 1976; Williams and Johnson, 1975) that compares the down-core records with a dataset of 'modern' values and their associated water column parameters (Hutson, 1977, 1978);

(2) those associated with the stable oxygen isotope composition of a whole shell analysed either individually (Feldmeijer et al., 2015; Ganssen et al., 2011; Koutavas et al., 2006; Koutavas and Joanides, 2012; Leduc et al., 2009; Metcalfe et al., 2015, 2019; Pracht et al., 2018) or pooled (Garidel-Thoron et al., 2007; Koutavas et al., 2002; Stott et al., 2002, 2004), herein  $\delta^{18}\text{O}_c$  (c = calcite), which can ideally be used to reconstruct SST

Formatted: Heading 2

Formatted: Indent: Left: 1 cm, Right: 1.2 cm

and past oxygen isotope values in seawater  $\delta^{18}\text{O}_{\text{sw}}$  (sw = seawater) when paired with a proxy that can either reconstruct temperature or salinity;

and **(3)** those associated with the trace metal geochemistry (e.g., Ford et al., 2015; Sadekov et al., 2013; Stott et al., 2002, 2004; White et al., 2018)(Stott et al., 2002, 2004), more specifically the natural logarithm of the relative concentration of Mg and Ca ( $\ln(\text{Mg}/\text{Ca})$ ), of the shell, based upon the temperature dependent (Elderfield and Ganssen, 2000; Nürnberg et al., 1996) incorporation and substitution of a Mg cation into the calcite lattice (Branson et al., 2013, 2016).

However, ~~t~~the interpretation of these ~~proxies~~proxies, however, is not straightforward, for example, calibration of foraminiferal assemblage based transfer functions with surface temperatures as opposed to a deeper temperature signal may skew the reconstructed temperature (Telford et al., 2013);  $\delta^{18}\text{O}_c$  is further can be affected by species-specific size effects (Feldmeijer et al., 2015; Metcalfe et al., 2015; Pracht et al., 2018), disequilibria or vital effects, which clouds the accurate reconstruction of past SST and  $\delta^{18}\text{O}_{\text{sw}}$ . There is no simple bijective function between  $\delta^{18}\text{O}_c$  and the oceanic variables  $\delta^{18}\text{O}_{\text{sw}}$

and temperature used in its calculation, with variability in  $\delta^{18}\text{O}_{\text{sw}}$  limiting the use of  $\delta^{18}\text{O}_c$  as a pure temperature proxy. Likewise, researchers have not been able to discount the impacts of the ambient salinity (Allen et al., 2016; Gray et al., 2018; Groeneveld et al., 2008; Kısakürek et al., 2008) and carbonate ion concentration (Allen et al., 2016; Evans et al., 2018; Zeebe and Sanyal, 2002) on the Mg/Ca content of foraminifera, nor biological effects such as growth banding (Eggins et al., 2003; Hori et al., 2018; Sadekov et al., 2008, 2009; Vetter et al., 2013). Foraminifera are also not passive recorders of environmental conditions such as SST, in that the very ambient environment that researchers wish to reconstruct also modifies the foraminiferal population as well (Mix, 1987; Mulitza et al., 1998). Sensitivity to the variable being reconstructed may increase or decrease the relative contribution of individual events through modulation of the flux to the seafloor, increasing or decreasing the chance of sampling such occurrences

~~Indeed, e~~ (Mix, 1987; Mulitza et al., 1998). ~~—e~~ Culture experiments have identified temperature (Lombard et al., 2009, 2011), light (Bé et al., 1982; Bé and Spero, 1981; Lombard et al., 2010; Rink et al., 1998; Spero, 1987; Wolf-Gladrow et al., 1999), carbonate ion concentration [ $\text{CO}_3^{2-}$ ] (Bijma et al., 2002; Lombard et al., 2010) and ontogenetic changes (Hamilton et al., 2008; Wycech et al., 2018) as variables that drive, alter or induce changes in foraminiferal growth.

~~omputation of the influence of biological and vital effects upon physiochemical proxies should be a fundamental consideration for any accurate data-model comparison. Recent attempts at circumnavigating these problems have employed isotope-enabled models (Caley et al., 2014; Roche et al., 2014; Zhu et al., 2017a) or forward-proxy models (Dolman and Laepple, 2018; Jonkers and Kučera, 2017; Roche et al., 2017) to predict potential  $\delta^{18}\text{O}_c$  values in foraminifera. However, the underlying foraminiferal population can be highly sensitive to the particular parameter being tested (Mix, 1987; Mulitza et al., 1998) — culture experiments have identified temperature (Lombard et al., 2009, 2011), light (Bé et al., 1982; Bé and Spero, 1981; Lombard et al., 2010; Rink et al., 1998; Spero, 1987; Wolf-Gladrow et al., 1999) carbonate ion concentration~~

Formatted: Indent: Left: 1 cm, Right: 1.2 cm

Formatted: Font: Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Indent: Left: 1 cm, Right: 1.2 cm

Formatted: Subscript

[CO<sub>2</sub>'] (Bijma et al., 2002; Lombard et al., 2010) and ontogenetic changes (Hamilton et al., 2008; Wycech et al., 2018) as variables that drive, alter or induce changes in foraminiferal growth. Such sensitivities limit the experimental independence of a data-model comparison using the aforementioned parameters (Roche et al., 2017).

A chief parameter that has been employed in previous ENSO proxy work using individual foraminifera analysis (IFA) is the  $\sigma(\delta^{18}\text{O}_c)$  parameter. This parameter  $\sigma(\delta^{18}\text{O}_c)$  is used to reconstruct the downcore individual foraminifera variance in  $\delta^{18}\text{O}_c$ , and increased  $\sigma(\delta^{18}\text{O}_c)$  is considered to correlate to increased variation in SST and, in turn, increased ENSO incidence and/or magnitude (Leduc et al., 2009; Zhu et al., 2017a) or even increased inter-annual variance in temperature. However, foraminifera are not passive recorders of environmental conditions such as SST, in that the very the ambient environment that researchers wish to reconstruct also modifies the foraminiferal population as well (Mix, 1987). Here, we use the recently developed *Foraminifera as Modelled Entities* (FAME) model (Roche et al., 2017) to take into account potential modulation of  $\delta^{18}\text{O}_c$  and the Temperature recorded in the calcite, herein  $T_c$ , by foraminifera growth, an estimate of the proxy Mg/Ca. Computation of the influence of biological and vital effects upon physiochemical proxies, such as those based on foraminifera should be a fundamental consideration for any accurate data-model comparison. Recent attempts at circumnavigating proxy related problems have employed isotope-enabled models (Caley et al., 2014; Roche et al., 2014; Zhu et al., 2017a), proxy models (Dolman and Laepple, 2018; Jonkers and Kučera, 2017; Roche et al., 2018) or uncertainty analysis (Thirumalai et al., 2013; Fraass and Lowery, 2017; Dolman and Laepple, 2018) to predict both the potential  $\delta^{18}\text{O}_c$  values in foraminifera and/or the probability of detection of a climatic event. The use of ecophysiological models (Kageyama et al., 2013; Lombard et al., 2009, 2011) can help circumvent some of the problems associated with a purely mathematical approximation (e.g., Caley et al., 2014) of the translation of an ambient signal into a palaeoclimate proxy. They are not limited to foraminifera and can provide an important way to test whether proxies used for palaeoclimate reconstructions are suitable for the given research question. Several studies have investigated the response of planktonic foraminifera from core material or computed pseudo foraminiferal distributions, their proxy values, and the resultant (likely) distribution of these proxy values with respect to ENSO (e.g., Leduc et al., 2009; Thirumalai et al., 2013; Ford et al., 2015; Zhu et al., 2017).

However, foraminifera are not passive recorders of environmental conditions such as SST, in that the very the ambient environment that researchers wish to reconstruct also modifies the foraminiferal population as well (Mix, 1987). Here, we use the recently developed *Foraminifera as Modelled Entities* (FAME) model (Roche et al., 2017) to take into account potential modulation of  $\delta^{18}\text{O}_c$  and the Temperature recorded in the calcite, herein  $T_c$ , by foraminifera growth, an estimate of the proxy Mg/Ca.

Here, we investigate whether living planktonic foraminifera can be theoretically used in ENSO reconstructions, differing from previous research by using a foraminiferal growth model, *Foraminifera as modelled entities* (FAME; Roche et al., 2018), to tackle the dynamic seasonal and depth habitat of planktonic foraminifera (Wilke et al., 2006; Steinhardt et al., 2015; Mix, 1987; Multitza et al., 1998). To be a valid proxy for the reconstruction of ENSO, the proxy values of populations of planktonic foraminifera associated to different climatic states (*i.e.*, El Nino, Neutral, La Nina) should be significantly different from one another. In order to test our research question, *'are the distributions of proxy values associated with El*

Formatted: Not Highlight

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic



*Niño months statistically different from distributions of proxy values associated with neutral or La Niña months?*, our methodology follows a forward modelling approach in which the computed values of the temperature recorded by calcite ( $T_c$  - a pseudo temperature aimed at mimicking Mg/Ca albeit one uninfluenced by secondary factors) and  $\delta^{18}O_c$  are assigned to one of these climatological states. This forward modelling approach does not pre-suppose foraminifera can record ENSO variability ('Can we detect?') i.e., what is done when inverting the core top pooled  $\delta^{18}O$  or individual foraminiferal  $\delta^{18}O$  distributions and infer changes in ENSO ('How could we detect?'). A secondary objective is to compare the output of this approach with secondary factors that further modulate the climatic signal through post mortem processes. We identify regions in the Pacific Ocean where the sedimentation rate (Berger, 1970a, 1971; Boltovskoy, 1994; Lougheed et al., 2018; Olson et al., 2016) may be too low or the water depth (Berger, 1967, 1970b; Boltovskoy, 1966; Lougheed et al., 2018) too deep (causing dissolution of carbonate sediments) thus preventing the capture and preservation of the foraminiferal signal.

A number of models exist to determine the foraminiferal responses to present (Fraile et al., 2008, 2009; Kageyama et al., 2013; Kretschmer et al., 2017; Lombard et al., 2009, 2011; Roy et al., 2015; Waterson et al., 2016; Žarić et al., 2005, 2006), past (Fraile et al., 2009; Kretschmer et al., 2016) and future (Roy et al., 2015) climate scenarios. FAME uses the associated temperature and  $\delta^{18}O_{eq}$  at each grid cell to compute a time averaged  $\delta^{18}O_c$  and  $T_c$  for a given species. This model utilises the temperature component of the Lombard et al. (2009) equations to simulate temperature derived growth (Kageyama et al., 2013; Lombard et al., 2009, 2011) and subsequently growth rate weighted  $\delta^{18}O_c$  (Roche et al., 2017) and  $T_c$  for *Globigerinoides ruber*, *Globigerinoides sacculifer* and *Neogloboquadrina dutertrei*, forced by fifty eight years of monthly ORAS S4 ocean reanalysis temperature and salinity data (Balmaseda et al., 2013). The original Lombard et al. (2009, 2011) equations are based upon a synthesis of culture studies, pooled together irrespective of experimental design or rationale, therefore they can be considered to conceptually represent the fundamental niche of a given foraminiferal species, i.e. the range in environment that the species can survive. Using the MARGO core top  $\delta^{18}O_c$  database (MARGO Project Members\*, 2009), Roche et al. (2018) computed the optimum depth habitat (the depth habitat that exhibits the strongest correlation when comparing FAME  $\delta^{18}O_c$  and MARGO  $\delta^{18}O_c$ ) for each species in the MARGO database (MARGO Project Members\*, 2009). The MARGO database does not include *N. dutertrei*, meaning that we concentrate our efforts mainly on *G. ruber* and *G. sacculifer*. By identifying the optimum depth habitat, Roche et al. (2018) established the realised niche, i.e. the range in environment that the species can be found, for these species for the late Holocene. Unlike some foraminiferal models, FAME does not include limiting factors such as competition, respiration or predation variables as no reliable proxy exists for such parameterisation in the geological record, therefore aspects such as interspecific competition that may limit the niche width of a species are not computed. Consequently, we allow all the species of foraminifera to grow down to -400 m (depending if optimal temperature conditions are met) to capture the total theoretical niche width. As the optimised depths of Roche et al. (2018) are shallower, and upper ocean water is more prone to temperature variability, our approach likely dampens both the modelled  $\delta^{18}O_c$  and  $T_c$ . Therefore, the sensitivity of the model was tested by applying the same procedure but with the limitation of the depth set to 60; 100 and 200 m.

Formatted: Subscript

## 2. Methods

### 2.1 Input variables (Temperature, Salinity and $\delta^{18}\text{O}_{\text{sw}}$ )

For input variables, The temperature and salinity of the ocean reanalysis data product (Universiteit Hamburg, DE) ORA S4 (Balmaseda et al., 2013) were extracted at one-degree resolution for the tropical Pacific (-20°S to 20°N and 120°E to -70°W), with each single grid cell comprised of data for 42 depth intervals (5 – 5300 m water depth) and 696 months (January 1958 – December 2015). For computation of the oxygen isotope of seawater ( $\delta^{18}\text{O}_{\text{sw}}$ ), A global 1-degree grid was generated, and each grid cell was classified as belonging to one of 27 distinct ocean regions, as defined by either societal and scientific agencies, for identifying regional  $\delta^{18}\text{O}_{\text{sw}}$  – salinity relationships (LeGrande and Schmidt, 2006). Using the  $\delta^{18}\text{O}_{\text{sw}}$  database of LeGrande and Schmidt (2006) a regional  $\delta^{18}\text{O}_{\text{sw}}$  – salinity relationship was defined, of which the salinity is the salinity measured directly at the isotope sample collection point (included within the database). Two matrices were computed; one giving values of the slope ( $m$ ) and the other of intercept ( $c$ ) of the resultant linear regression equations, these were used as look-up tables to define the monthly  $\delta^{18}\text{O}_{\text{sw}}$  from the monthly salinity Ocean reanalysis product ORAS S4 (Balmaseda et al., 2013), which was used for the calculation of  $\delta^{18}\text{O}_{\text{eq}}$  *i.e.* the expected  $\delta^{18}\text{O}$  for foraminiferal calcite formed at a certain temperature (Kim and O’Neil, 1997). The  $\delta^{18}\text{O}_{\text{eq}}$  is calculated from a rearranged form of the following temperature equation:

$$T = T_0 - b \cdot (\delta^{18}\text{O}_c - \delta^{18}\text{O}_{\text{sw}}) + a \cdot (\delta^{18}\text{O}_c - \delta^{18}\text{O}_{\text{sw}})^2, (1)$$

Specifically, we used the quadratic approximation (Bemis et al., 1998) of Kim and O’Neil (1997), where  $T_0 = 16.1$ ,  $a = 0.09$ ,  $b = -4.64$  and converted from V-SMOW to V-PDB using a constant of -0.27 ‰ (Hut, 1987; Roche et al., 2017):

$$\Delta = b^2 - 4a \cdot (T_0 - T_{\text{sw}}), (2)$$

$$\delta^{18}\text{O}_{c,\text{eq}} = \frac{-b - \sqrt{\Delta}}{2a} + \delta^{18}\text{O}_{\text{sw}} - 0.27, (3)$$

The difference between the constant of Hut (1987) and the dynamic value (Brand et al., 2014) is minor.

### 2.2 Foraminifera as modelled entities (FAME)

For each latitude and longitude grid, a monthly growth rate was calculated using the ORA S4 temperature and a Michaelis-Menton kinetics to predict growth calculated from culture experiments, described in Lombard et al. (2009). Growth rate was artificially constrained to arbitrary values of the upper 60, 100, 200 and 400 m for all species to reflect the symbiotic nature of various foraminiferal species. Foraminifera as modelled entities has been developed as a tool for translating, a climatic input (typically a reanalysis dataset or climate model output) into a (simulated-) climatic signal, a signal that aims to approximate the depth integrated growth of foraminifera (e.g., Pracht et al., 2019; Wilke et al., 2006; Steindhardt et al., 2015). Data-model comparison studies suffer from an inability to directly compare like with like so that there are differences in (i) the units used *i.e.*, most proxies reconstructing temperature do not give values of temperature in degrees °C or K but in their own proxy units (e.g., per mil ‰; mmol/mol; species abundance or ratio) necessitating a conversion; and (ii) there is a

Formatted: Centered

Formatted: Centered

Formatted: Not Highlight

reduction in scales, *i.e.*, models give a wealth of information (multiple depth layers and high resolution time slices) in the time-depth domain. A number of models and modelling studies exist to determine the foraminiferal responses to present (Fraile et al., 2008, 2009; Kageyama et al., 2013; Kretschmer et al., 2017; Lombard et al., 2009, 2011; Roy et al., 2015; Waterson et al., 2016; Žarić et al., 2005, 2006), past (Fraile et al., 2009; Kretschmer et al., 2016) and future (Roy et al., 2015) climate scenarios, FAME uses the associated temperature and  $\delta^{18}\text{O}_{\text{eq}}$  at each grid cell to compute a time averaged  $\delta^{18}\text{O}_c$  and  $T_c$  for a given species. FAME was produced as an attempt to reduce the error associated with data-model comparisons by (i) generating simulated-proxy time-series from model runs that can be compared with age-depth values down core; and (ii) to reduce the model information for a given time-slice into a manageable and relevant value using an integration that would make sense on a biological point of view (Roche et al., 2018).

The FAME model utilises the temperature-growth rate equations of Lombard *et al.* (2009) to simulate temperature-derived growth rate (Kageyama et al., 2013; Lombard et al., 2009, 2011), this growth rate is then used as a weighing to produce a growth rate-weighted proxy value (Roche et al., 2018). The original Lombard et al. (2009, 2011) equations are based upon a synthesis of culture studies, pooled together irrespective of experimental design or rationale, therefore they can be considered to conceptually represent the fundamental niche of a given foraminiferal species, *i.e.* the range in environment that the species can survive. The basic structure of FAME is based upon temperature based Michaelis-Menton kinetics to predict growth rate, described in Lombard et al. (2009), without using the parameters (e.g., light, respiration, food) associated with FORAMCLIM (Lombard et al., 2011). The absence of known values or proxy values for the full set of parameters associated with FORAMCLIM has led us to an Occam's Razor favoured approach in model parameterisation for FAME (Roche et al., 2018). Although other processes may also impact species such as mixed layer depth and nutrients these variables for now can be set aside, as temperature provides the dominant signal, it is worth noting that in all probability some variance will arise from these processes and deviation between observed and expected values should consider this.

Using the MARGO core top  $\delta^{18}\text{O}_c$  database (MARGO Project Members\*, 2009), Roche *et al.* (2018) validated and computed the optimum depth habitat (the depth habitat that exhibits the strongest correlation when comparing FAME  $\delta^{18}\text{O}_c$  and MARGO  $\delta^{18}\text{O}_c$ ) for each species in the MARGO database (MARGO Project Members\*, 2009). Whilst, both models, FAME and FORAMCLIM, can compute the growth rate of eight foraminiferal species from culture studies (Kageyama et al., 2013; Lombard et al., 2009, 2011; Roche et al., 2017), the limited number of species available for a global core top comparison led to a reduction in the number of species modelled (Roche et al., 2018). Here the output of FAME is further restricted to three species that have been the main focus of foraminifera-based studies that have been used to infer ENSO variability, namely the upper ocean dwelling *Globigerinoides sacculifer* and *Globigerinoides ruber*, as well as the thermocline dwelling *Neogloboquadrina dutertrei* (Ford et al., 2015; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Leduc et al., 2009; Sadekov et al., 2013). The MARGO database does not include *N. dutertrei*, meaning that we concentrate our efforts mainly on *G. ruber* and *G. sacculifer*.

In this study, ORA S4 temperature was used as the input variable, with the growth rate computations artificially constrained to arbitrary values of the upper 60; 100 and 200 m to reflect the presence of photosymbiotic algae in the various

Formatted: Not Highlight



temperature sensitivity (Zachary and Kucera, 2017; Liu, 1997). Although other processes may also impact species such as mixed layer depth and nutrients these for now can be ignored. Therefore, both the resultant  $\delta^{18}\text{O}_e$  and  $T_e$  are influenced/affected by  $T_c$ . But for a weighted distribution each monthly contribution is  $\frac{\text{count}_i}{\sum \text{count}_i}$ . The same procedure was employed for the  $T_e$ , where the computed  $\delta^{18}\text{O}_e$  is replaced with the value of temperature profile for a given latitude and longitude grid point.

#### 2.4.4 Statistical analysis

The tropical Pacific Ocean is divided into four Niño regions based on historical ship tracks, from east to west: Niño 1 and 2 (0° to -10°S, 90°W to 80°W), Niño 3 (5°N to -5°S, 150°W to 90°W), Niño 3.4 (5°N to -5°S, 170°W to 120°W) and Niño 4 (5°N to -5°S, 160°E to 150°W). Pan-Pacific meteorological agencies differ in their definition (An and Bong, 2016, 2018) of an El Niño, with each country's definition reflecting socio-economic factors. Therefore, for simplicity we use One-index-for-ENSO, the Oceanic Niño Index (ONI), based upon the Niño 3.4 region (because of the region's importance for interactions between ocean and atmosphere) which is a 3-month running mean of SST anomalies in ERSST.v5 (Huang et al., 2017). However, Pan-Pacific meteorological agencies differ in their definition (An and Bong, 2016, 2018) of an El Niño, with each country's definition reflecting socio-economic factors, therefore, for simplicity we- We utilise a threshold of  $\chi \geq +0.5^\circ\text{C}$  (where  $\chi$  is the value of ONI) as a proxy for El Niño,  $-0.5^\circ\text{C} \leq \chi \leq +0.5^\circ\text{C}$  for neutral climate conditions and  $-0.5^\circ\text{C} \leq \chi$  for a La Niña in the Oceanic Niño Index. Many meteorological agencies consider that five consecutive months of  $\chi \geq +0.5^\circ\text{C}$  must occur for the classification of an El Niño event. However, here it is the only difference is that we considered, that any single month falling within our threshold values as representative of El Niño, neutral or La Niña conditions (grey bars in Figure 1). This simplification reflects the lifecycle of planktonic foraminifera (~4 weeks) seeing that the population at time step  $t$  does not record what happened at  $t-1$ , or what will happen at  $t+1$ . As we are producing the mean population growth weighted  $\delta^{18}\text{O}_e$  values, 'almost' El Niño or 'almost' La Niña would be indistinguishable from the build-up and subsequent climb-down of actual El Niño and La Niña events. Therefore, these 'almost' El Niño or 'almost' La Niña are placed within their respective climatological pools as El Niño or La Niña.

-Each time-step for the entirety of the Pacific was classified as one of three climate states (El Niño; Neutral; and La Niña), where after By using this threshold, three weighted histograms for the resultant each  $\delta^{18}\text{O}_e$  and  $T_e$  at each timestep produced by FAME and their resultant distributions (El Niño; Neutral; and La Niña) were computed for every month and for every latitude and longitude grid-point for the 1958-2015 period for each grid-point were binned into their respective categories. Monthly output representing different climate conditions were compared against each another for  $\delta^{18}\text{O}_e$  and  $T_e$  separately. An Epanechnikov-kernel distribution was first fitted to the binned monthly output of a single climate state, the bandwidth varies between grid-points to provide for an optimal kernel distribution. The use of an Epanechnikov-kernel distribution to fit the data, as opposed to other types of distribution, represents a trade-off between keeping as many parameters constant whilst mimicking the underlying dataset for a large number of grid points. The conversion of the data from dataset to distribution

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Not Bold

Formatted: Highlight

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold, Italic

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold, Italic

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold, Italic

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Superscript

Formatted: Font: Not Bold

may induce some small error induced by: rounding to whole integers; the use of a  $\delta^{18}\text{O}_{\text{mid-point}}$  which gives an error associated with the bin size ( $\pm 0.05\%$ ) that is symmetrical close to the distributions measures of central tendency but asymmetrical at the sides; and finally, the associated rounding error at the bin edges within a histogram ( $\pm 0.005\%$ ). Subsequently after which any two desired distributions could be compared for (dis)similarity using an Anderson-Darling test (1954). Here, all values, i.e., the population, associated with a climatological state are compared with the other populations representing the different climatological state, the results plotted here are Neutral climate state vs. El Niño climate state. The areas where the two distributions are found to agree, at the 5% significance level, are shown as e.g., the grey/white area in Figure's 3 to 6.

Comparison, between the weighted ( $p(\delta^{18}\text{O}_{\text{err}}) = \frac{\sum_{z=0}^{z=1} \epsilon_{z, \text{err}}}{\sum_{z=0}^{z=1} \epsilon_{z, \text{err}} + \sum_{z=0}^{z=1} \epsilon_{z, \text{err}}}$ ) and unweighted ( $p(\delta^{18}\text{O}_{\text{err}}) = \frac{\epsilon_{z, \text{err}}}{\epsilon_{z, \text{err}}}$ ) distributions were made for El Niño climate; combined Neutral and La Niña climates; only Neutral climate; and only La Niña climate.

As the weighted distributions are effectively probability distributions, in order to fit a distribution, we multiplied the bin counts by 1000, effectively converting probability into a hypothesised distribution.

Using the repeat matrix function (MatLab function: *repmat*), a matrix of  $\delta^{18}\text{O}_e$  was produced using each bin's mid-point ( $\delta^{18}\text{O}_{\text{mid-point}}$ ) there is a threefold error combined with this methodology which may account for minor variation between discrete runs of the model: first the counts values were rounded to whole integers so an exact number of cells could be added to a matrix; secondly the  $\delta^{18}\text{O}_{\text{mid-point}}$  was used which gives an error associated with the bin size ( $\pm 0.05\%$ ) that is symmetrical close to the distributions measures of central tendency but asymmetrical at the sides; and finally, the associated rounding error at the bin edges within a histogram ( $\pm 0.005\%$ ). 2.5 Test of input data (Temperature and calculated  $\delta^{18}\text{O}_{\text{eq}}$ )

Foraminifera as modelled entities produces a modulated response that seeks to replicate how foraminifera modify the climate signal, several studies have approximated the foraminiferal signal in a different way (e.g., Thirumalai et al., 2013; Zhu et al., 2017a). In order to understand how FAME has altered the signal, and the degree to which the conclusions drawn depend upon the modelled growth rates, the input datasets of the sea water properties (Temperature and calculated  $\delta^{18}\text{O}_{\text{eq}}$ ), underwent a similar statistical test (Figure 4). Unlike FAME, which integrates over several depth levels using the computed growth rate, the test of the input datasets was with fixed depths without any growth rate weighting. These fixed depths are 5, 149 and 235 m, giving a Eulerian view (Zhu et al., 2017a) in which to observe the implications of FAME's dynamic depth habitat. As per the FAME output, each timestep value was placed into its climate state and an Anderson-Darling test performed to compare the (dis)similarity of on the resultant distributions.

### 2.6 Alternative statistical tests

In order to compare our results with previously published studies using planktonic foraminifera we employed a series of simple statistical tests, mimicking those applied to sediment archives by the palaeoclimate community. A chief parameter that has been employed in previous ENSO proxy work using individual foraminiferal analysis (more specifically, individual foraminiferal analysis; IFA) is the measure of individual foraminifera downcore standard deviation ( $\sigma(\delta^{18}\text{O}_c)$ ) parameter. This parameter  $\sigma(\delta^{18}\text{O}_c)$  is used to reconstruct the downcore individual foraminifera variance in  $\delta^{18}\text{O}_c$ , and increased

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

$\sigma(\delta^{18}\text{O}_c)$  is considered to correlate to increased variation in SST and, in turn, increased ENSO incidence and/or magnitude (Leduc et al., 2009; Zhu et al., 2017a) or even increased inter-annual variance (Thirumalai et al., 2013) in temperature. The variance ( $\sigma^2(\delta^{18}\text{O}_c)$ ) of the timeseries were computed both as the total variance and as the interannual variance, the latter is computed as outlined in Zhu et al. (2017a). For the interannual variance, the mean monthly climatology is subtracted from the dataset, producing monthly anomalies and a linear trend removed (using the detrend function of MatLab 2019a) – the resultant data was left unfiltered (*i.e.*, Zhu et al., 2017a used a 1-2-1 filter). Four ‘picking’ experiments were performed, as FAME computes the average value for a given time step and given the single foraminiferal isotope variance for an equivalent time step (e.g., weeks; Steinhardt et al., 2015) it is more than likely that this computation reduces the real spread in values. Therefore, rather than use the terminology specimen we prefer to use months. Given the complexity in reconstructions of trace metal geochemistry (Elderfield and Ganssen, 2000; Nürnberg et al., 1996); the potential error associated with determining which carbonate phase is first used when foraminifera biomineralise (Jacob et al., 2017); growth-band integration; secondary factors (e.g., salinity, carbonate ion) the focus of the picking here has been on the  $\delta^{18}\text{O}_c$ . Irrespective of which experiment, 60 months were drawn, with replacement, and the number of Monte Carlo iterations is set at 10,000. We assume that the ‘picker’ is taxonomically well-trained and/or has a procedure in which species can be checked taxonomically post-analysis (e.g. photographing all specimens prior to analysis, Pracht et al. (2019)) and therefore do not include an error that deals with incorrect identification. Although we note that parameterisation of misidentification would be difficult, as it requires understanding of the variability in both standard deviation and absolute values for species co-occurring downcore (Feldmeijer et al., 2015; Metcalfe et al., 2015; 2019). For each run of experiment’s (i) to (iii) the drawn months were saved in order to perform (iv):

(i) In Picking Experiment-I (Figure 31), the months drawn for each iteration of the Monte Carlo were selected and each grid-point was sampled (*i.e.*, there are 10000·60 selected months). This assumes that the same months are selected at grid point A as point B.

(ii) In Picking Experiment-II (Figure 31), at each grid-point a Monte Carlo was run (*i.e.*, there are 170·40·10000·60 selected months). This assumes that different months could be selected between grid point A and point B.

(iii) In Picking Experiment-III (Figure 31), at each grid-point a Monte Carlo was run using the growth rate weighting for each month (*i.e.*, there are 170·40·10000·60 selected months), this assumes that in periods of higher growth there will be a higher flux of the species and therefore a greater chance of selecting that month. The rationale being that not only are different months selected between grid point A and point B, but if A and B differ climatologically there may be an over subscription of ecologically beneficial habitats in one core location compared to the other.

Formatted: Superscript

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Font: Italic

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Italic

Formatted: Highlight

Formatted: Font: Italic

(iv) In Picking Experiment-IV (Figure 30 to 31), the experiment of (ii) was re-run but with the addition of two sources of error: The first error is based upon FAME producing the average value for a given time slice, therefore short-term variability in temperature and/or the spread in the population (*i.e.*, variance in depth of an individual; variance in chamber growth per individual), as evidenced by single foraminiferal analysis of sediment trap samples (*e.g.*, Steinhardt et al., 2015), is potentially lost. Therefore, for each picked month between -0.40 and 0.40 ‰ is added to the picked month value (in intervals of 0.02 ‰), this is approximately  $\pm 2^{\circ}$  C (*i.e.*,  $-4^{\circ}$  C). The second error is the analytical error that an individual measurement will have. Machine measurement error is assumed to lie between -0.12 and 0.12 ‰ (in intervals of 0.005 ‰ – the 3<sup>rd</sup> decimal place is an exaggeration of machine capabilities although it will have repercussions for rounding) the 1 $\sigma$  of within run (as opposed to long-term average) of international stable isotope standards. The intervals of both errors (0.02 ‰ and 0.005 ‰) were chosen to give a similar number ( $n = 41$  and 49) of potential randomly selected error for each picked month. Each picked month has their own randomly selected error for both of these errors, *i.e.*, each value is the sum of the month picked and their own error. The values for within month variability (Figure 30) and machine error (Figure 31) are calculated separately and then combined (Figure 31), as they may have a corresponding or conflicting signs, either ‘cancelling’ out each other or amplifying the difference.

- Formatted: Highlight
- Formatted: Not Highlight
- Formatted: Not Highlight
- Formatted: Font: (Default) +Headings (Times New Roman), Not Highlight
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: Italic
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Superscript
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: (Default) +Headings (Times New Roman)

- Formatted: Highlight
- Formatted: Highlight
- Formatted: Font: (Default) +Headings (Times New Roman)

An associated statistical methodology is the graphical summary (as opposed to a numerical summary via a test value) of plotting the quantiles of two probability or the quantiles of sample probability distribution against a theoretical distribution distributions also referred to as a Quantile-Quantile, or Q-Q plot (*e.g.*, Ford et al., 2015; White et al., 2018). A complimentary (*i.e.*, used in association with, not as replacement, Filliben 1975) test metric, the Probability plot correlation coefficient (Filliben, 1975) can be used as a numerical summation of this approach, which bases its rationale on near linearity between the two tested distributions. This graphical technique is not used here for the following reasons, (i) the climatic categories (*i.e.*, El Nino, Neutral, La Nina) imposed upon the data give uneven sized sample distributions requiring an interpolated quantile estimate; and (ii) the large graphical computation required (170-40).

- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Font: Italic

### 2.7 An approximation of sedimentary archives: Water depth & Sedimentation Rate

The resolution of the ocean reanalysis data for the time period 1958-2015 would essentially be analogous with a sediment core representing  $50 \text{ yr}^{-1} \text{ cm}^{-1}$  (or  $20 \text{ cm}^{-1} \text{ kyr}^{-1}$ ). Based on our analysis, such a hypothetical core with a rapid sediment accumulation rate (SAR) could allow for the possible disentanglement of El Niño related signals from the climatic signal using IFA, but only in a best-case scenario involving minimal bioturbation, which is unlikely in the case of oxygenated waters. Indeed, one should view discrete sediment intervals retrieved from systematically bioturbated deep-sea sediment cores contain, and the foraminifera with ages spanning many centuries contained within them, as representative of an integrated multi-decadal or even multi-centennial signal (Lougheed et al., 2018; Peng et al., 1979). This is in contrast to;

- Formatted: Font: Italic



as opposed to other proxies such as corals (Cole and Tudhope, 2017), speleotherms (Chen et al., 2016) and molluscs (Butler et al., 2013; Milano et al., 2017), or varves in which distinct layers correspond to discrete time intervals where distinct time-specific banding is present (true 'time-series' proxies). The ambient signal following translation into a foraminiferal signal within the water is therefore further modulated by several post mortem processes, which include: the latitudinal-longitudinal shift in position of sinking foraminifera - the so-called 'funnel affect' (van Sebille et al., 2015; Deuser et al., 1981); dissolution of calcium carbonate either in the water (Schiebel et al., 2007), at the seafloor, or due to pore fluids; and bioturbation. As mentioned, mixing by bioturbation, results in an apparent smoothing of the downcore, discrete-depth multi-specimen signal (Hutson, 1980a; Löwemark, 2007; Löwemark et al., 2005, 2008; Löwemark and Grootes, 2004; Cole and Tudhope, 2017; Mix, 1987), thus leading to the possibility of interpreting single outlying foraminifera values within a specific depth as representing an 'extreme' climate, when they may in fact represent climate from a different time or epoch. This is especially apparent in  $\delta^{18}\text{O}_c$  where there is a difference temporally of  $\delta^{18}\text{O}_{sw}$  (e.g., the ice volume effect in glacial and interglacial cycles  $\sim 1.25\%$ ) meaning that the same temperature can have radically different  $\delta^{18}\text{O}_c$  values, a consequence of this is that a series of high magnitude, but low frequency El Niño events could be disturbed in a discrete-depth record. 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 have a large enough magnitude, be largely unaffected by dissolution (*i.e.*, above the lysocline) so as not to adversely affect the population and the sedimentation rate must be high enough to give sufficient temporal coverage and rule out upwards bioturbation of single foraminifera from significantly different climate periods. In our first step in consideration of post-mortem signal alteration we focus on dissolution. The lysocline, the depth at which dissolution first becomes apparent (Berger, 1968; 1970), and the Calcite (or Calcium Carbonate) Compensation Depth (CCD; Bramlette, 1961) vary between the different ocean basins; the Atlantic Ocean in which deep water forms has a relatively deep CCD as a by-product of 'young' well ventilated bottom waters whereas the Pacific Ocean the final section of the thermohaline circulation conveyor belt, has a shallower CCD. In order to highlight the potential for dissolution, the bathymetry of the Pacific was extracted from the General Bathymetric Chart of the Oceans GEBCO 2014 30 arc-second grid (version 20150318, www.gebco.net) between  $-20^\circ\text{S}$  to  $20^\circ\text{N}$  and  $120^\circ\text{E}$  to  $-70^\circ\text{W}$  (Figure 8). Depths of 3500 m below sea-level (bsl), 4000 m bsl and 4500 m bsl are used here as cut-off values, these depths represent multiple possible depths under which there is the potential for noticeable dissolution (*i.e.*, lysocline) or be dissolved (*i.e.*, CCD). Whilst our intention here is a generalised view to be used as an approximate guide, it is important to note that the Pacific Ocean has the largest proportion globally of  $>1$  km tall seamounts that are smaller than  $<100$  km (Wessel, 1997). Which may have important, relatively shallow-water sedimentary sequences, which may also be of sufficient sediment accumulation rate, therefore we supplement the GEBCO bathymetric data with the locations of seamounts. However, whilst exhibit a large enough amplitude in order to perturb the population by a significant degree from the background signal, there are an estimated 50,000 seamounts in the Pacific that are taller than a km (Menard, 1964; Wessel and Lyons, 1997), only 12,000 have been documented on charts (Batiza, 1982), and approximately 291 have been dated (Koppers et al., 2003; Clouard and Bonneville,

Formatted: Highlight

Formatted: Highlight

Formatted: Subscript

Formatted: Font: Italic

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

2005; Hillier, 2007). It is these 291, <1% of the estimated seamounts, we have overlain onto the bathymetric data (Figure 8b), although this number is further reduced as we only plot between 20°S and 20°N.

The second step when considering post mortem signal alteration is the sediment accumulation rate (SAR). We first plot the time-averaged deep-sea SAR (Olson et al., 2016), adapted by Lougheed *et al.* (2018) for the Tropical Pacific (Figure 3). New geochronological tools, such as dual <sup>14</sup>C-<sup>δ</sup><sup>18</sup>O measurements on single foraminifera (Lougheed et al., 2018), show that low sedimentation rate cores can have large variances in age between individual foraminifera present within a discrete 1 cm depth interval (Berger and Heath, 1968; Lougheed et al., 2018). In order to model bioturbation, a number of papers have used a diffusion style approach that reduces the parameters down to sediment mixing intensity and sediment mixing depth (herein referred to as bioturbation depth, BD), although this may be an artificial division purely driven by mathematical need rather than biological constraints (Boudreau, 1998). The bioturbation depth has been shown to have a global average of 9.8 cm (1σ: ± 4.5 cm) that is independent of both water depth and sedimentation rate (Boudreau, 1998), though likely controlled as a result of the energy efficiency of foraging, *e.g.* deeper burrows may cost more energy to produce than can be offset in extracted food resources, and potential decay in labile food resources with sediment depth. It is not possible to carry out a transient bioturbation model upon the temperature and salinity ocean reanalysis data that we used for FAME, as it only covers half a century of data, whereas thousands of years of input data are required to force a transient bioturbation model. To investigate how much temporal signal is integrated into discrete-depth intervals for typical tropical Pacific SAR, we, therefore, utilised the single foraminifera sediment accumulation simulator (SEAMUS, Lougheed, 2019) to bioturbated, as the input climate signal (Figures 9 to 11), the 0-40,000 year <sup>δ</sup><sup>18</sup>O<sub>w</sub> of NGRIP (North Greenland Ice Core Project Members, 2004; Rasmussen et al., 2014; Seierstad et al., 2014). The ice core time series is an ideal input for a bioturbation simulator, as it represents a highly temporally resolved climate input signal. SEAMUS simulates foraminifera in 10-year timesteps. The use of the NGRIP timeseries here is purely as an input parameter to investigate the effect of bioturbation upon a given climate signal -, it is important to stress that by using NGRIP as an input signal for SEAMUS we are neither implying that tropical Pacific cores should have signal similar to NGRIP or inferring some kind of causal relationship. As we seek to investigate the effect of bioturbation, no attempt has been made to modulate the input signal's absolute values to mimic expected <sup>δ</sup><sup>18</sup>O<sub>e</sub> values and why each plot of the synthetic down core time series retains the use of V-SMOW, despite carbonates being required to be V-PDB (Coplen 1995). Keeping all things constant, and varying a single parameter between experiments with SEAMUS, the sediment accumulation rate (SAR) was varied to fixed values of either 1, 2, 5 or 10 cm kyr<sup>-1</sup> (representative of typical Pacific SAR) and a bioturbation depth (BD) of either 5, 10 or 15cm based upon the global estimate and it's error bounds (Boudreau, 1998). For each experiment, the selected values of SAR and BD were kept constant for the entire SEAMIS model run (i.e., the intensity and magnitude of bioturbation was not varied). In reality, SAR and BD may vary temporally depending on local conditions (*e.g.*, food, oxygen). Finally, the FAME results for the three species are overlaid with a water depth mask of above and below 3500 m below sea-level (mbsl), to also show seafloor areas under the CCD depth, where carbonate material is not preserved (Berger, 1967, 1970b). A comparison between water depth and time-averaged deep-sea SAR (Olson et al., 2016), adapted by Lougheed *et al.* (2018) is shown in Figure 7 and 9.

Formatted: Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Italic

Formatted: Superscript

Formatted: Subscript

Formatted: Highlight

Formatted: Subscript

Formatted: Highlight

Formatted: Superscript

Formatted: Font: Italic

Formatted: Not Highlight

Field Code Changed

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

otherwise it will be lost due to the smoothing effect of bioturbation (Hutson, 1980a; Löwemark, 2007; Löwemark et al., 2005, 2008; Löwemark and Grootes, 2004) upon the downcore, discrete-depth signal (Cole and Tudhope, 2017; Mix, 1987). Similarly, the individual characters of El Niño events, which are very short in duration, become lost in the bioturbated sediment record. New geochronological tools, such as dual  $^{14}\text{C}$ - $\delta^{18}\text{O}$  measurements on single foraminifera (Lougheed et al., 2018), show that low sedimentation rate cores can have large variances in age between individual foraminifera present within a discrete 1 cm depth interval (Berger and Heath, 1968; Lougheed et al., 2018). In the case of a high SAR core ( $20 \text{ cm}^+ \text{ kyr}^{-1}$ ), assuming a sediment mixing layer of 10 cm, benthic seafloor processes would result in a minimum 700-year 1 $\sigma$  confidence interval for a 1 cm discrete depth age (Lougheed et al., 2018). A consequence of bioturbation in sediment cores is that a series of high-magnitude, but low-frequency El Niño events could be smoothed out of the downcore, discrete-depth record. To investigate which areas of the sea floor can potentially preserve a foraminiferal downcore signal, we overlay our results with a map of time-averaged deep-sea SAR (Olson et al., 2016), adapted by Lougheed et al. (2018) to also show seafloor areas under the CCD depth, where carbonate material is not preserved (Berger, 1967, 1970b).

### 3.0 Results

The results of the forward model (Figures 2 and 4) are compared with their input (Figure 3) in order to identify regions in which the values are statistically distinct (Figure 5-6). These results are then shown against the water depth (Figure 7 to 9) and the SAR (Figure 9-11) for the region. The results utilise Foraminifera as Modelled Entities (FAME; Figures 2, 3, 4, 6, and 7); the original Ocean Reanalysis data with computed  $\delta^{18}\text{O}_{\text{eq}}$  (Figure 4); the General Bathymetric Chart of the Oceans (GEBCO; Figures 7 to 9); and the single foraminifera sediment accumulation simulator (SEAMUS; Figures 9 to 11).

#### 3.14 FAME Output: Variance

We compute growth-weighted  $\delta^{18}\text{O}_c$  (Figure 35 and 74) and temperature (Figure 6 and 7) distributions for each  $1^\circ$  by  $1^\circ$  grid cell in the fifty-eight year simulation using FAME (Roche et al., 2018), constraining the calculation to the Tropical Pacific Ocean (between  $-20^\circ\text{S}$  to  $20^\circ\text{N}$  and  $120^\circ\text{E}$  to  $-70^\circ\text{W}$ ). Our model produces 696 individual monthly maps for all three species (Figure 3). While two of the three species (*G. ruber* and *G. sacculifer*) have similar ecologies, they show differences in their resultant  $\delta^{18}\text{O}_c$  for the same ocean conditions (Figure 3). A comparison of our computed variance with measured data (Supplementary Table 1) is made, we compare both the value of the nearest grid-cell and because of the size of the grid and drift of foraminifera (van Sebille et al., 2015) an average of a 3 by 3 grid in which the nearest grid-cell to the core location is in the center. A comparison is made with both the iCESM model output and the core's that match this output (Zhu et al., 2017a). For the Late Holocene sample (~1.5 ka) MD02-2529 ( $08^\circ12.33'\text{N}$   $84^\circ07.32'\text{W}$ ; 1619 m) in which *N. dutertrei* individual foraminifera were analysed from  $>250 \mu\text{m}$  (Leduc et al., 2009) giving a calculated standard deviation of measured foraminifera of 0.38‰. Whereas, the full ~60 year time series ( $n = 696$ ) of FAME presented here, gives a standard deviation

- Formatted: Highlight
- Formatted: Highlight
- Formatted: Normal
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Highlight
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: (Default) +Headings (Times New Roman), Highlight
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Font: (Default) +Headings (Times New Roman)
- Formatted: Not Highlight
- Formatted: Not Highlight
- Formatted: Font: Italic
- Formatted: Not Highlight
- Formatted: Font: (Default) +Headings (Times New Roman), Not Highlight
- Formatted: Font: (Default) +Headings (Times New Roman), Not Highlight
- Formatted: Not Highlight
- Formatted: Font: (Default) +Headings (Times New Roman), Not Highlight
- Formatted: Not Highlight
- Formatted: Font: (Default) +Headings (Times New Roman), Not Highlight
- Formatted: Not Highlight
- Formatted: Font: (Default) +Headings (Times New Roman), Not Highlight
- Formatted: Not Highlight

for all species, at depth cut off 60 m between 0.26 and 0.32 ‰; at depth cut off 100 m between 0.20 and 0.29 ‰; at depth cut off 200 m between 0.20 and 0.25 ‰; and at depth cut off 400 m between 0.20 and 0.24 ‰ (see Supplementary Table 1). Although these values vary if the average of the surrounding grid cells is used (see Supplementary Table 1). In comparison the iCESM results have the following standard deviation values, for a Eulerian (fixed) depth of 50 m: 0.4 ‰; Eulerian 100 m: 0.6 ‰; and Lagrangian value of 0.49 ‰. There are three samples (Koutavas and Joanides, 2012; Sadekov et al., 2013) located south of core site MD02-2529, these are the Late Holocene (~1.6 ka) samples of V21-30 (01°13'S 89°41'W; 617 m) and (~1.1 ka) V21-29 (01°03'S 89°21'W; 712 m) in which *G. ruber* was measured individually (Sadekov et al., 2013). For these two sites the measured standard deviation is 0.507 ‰ and 0.510 ‰ for V21-30 and V21-29 respectively (Koutavas and Joanides, 2012). The third core site at a similar location is (~1.6ka) CD38-17P (01°36'04 S 90°25'32W; 2580 m) was not analysed individually, instead replicates of pooled samples of 2 or 3 shells of *N. dutertrei* (Sadekov et al., 2013) were made these measured values give a standard deviation of 0.28 ‰. The full ~60 year time series (n = 696) of FAME presented here, gives a standard deviation for all species, at depth cut off 60 m between 0.33 and 0.41 ‰; at depth cut off 100 m between 0.27 and 0.40 ‰; at depth cut off 200 m between 0.25 and 0.35 ‰; and at depth cut off 400 m between 0.25 and 0.34 ‰ (see Supplementary Table 1). Although these values vary if the average of the surrounding grid cells is used (see Supplementary Table 1). In comparison the iCESM results have the following standard deviation values, for a Eulerian (fixed) depth of 50 m: 0.53 ‰; Eulerian 100 m: 0.75 ‰; and Lagrangian value of 0.35 ‰.

The study of ENSO has focused on whether the variability is entirely in response to ENSO or whether it is dominated by interannual variability (Xie, 1994, 1995; Wang et al 1994, 2010), here the interannual (Figure 3C) and total variance (Figure 3A) was computed and a ratio between the two calculated (Figure 3B; see Supplementary Table 1). Like the same analysis of interannual and total variance computed for iCESM and SODA reanalysis (Carton et al., 2000), outlined in Zhu et al. (2017a), there is also high ratio of interannual to total variance in our computed FAME dataset (Figure 3B). Although there are regions in the Eastern Equatorial Pacific wherein this ratio reduces. Despite this reduction, the ratio between total and interannual variance is still above > 0.5.

The Monte-Carlo experiments (Figure 3D-I) highlight the variation in picking a subset of the months, here 60, from the full timeseries. The FAME- $\delta^{18}\text{O}_{\text{cc}}$  *G. sacculifer* with a depth cut-off of 60 m is plotted here, the values for each grid point is the range in standard deviation (i.e., the maximum standard deviation minus the minimum standard deviation) between iterations of the Monte-Carlo (n= 10,000). The range in standard deviations between iterations is plotted instead of the mean of the standard deviations. With increasing *n* the mean converges toward the sample mean, however as the point of the Monte-Carlo is to generate plausible 'samples' it is more important to take into account the range in possible values which would help to establish the potential variability of subsampling. For the most part, regions with high total variance (Figure 3A) also have a larger range in standard deviations between the iterations 'picked'. It is interesting to note that by changing from the same months picked for each grid-point (Monte-Carlo I: Figure 3D) to varying the months picked between grid-points (Monte-Carlo II: Figure 3E or Monte-Carlo III: 3F) the range goes from 'smooth' to a more noisy dataset. Whilst the values plotted here are not the absolute values (as they are the range in standard deviation for a given grid point for the entire

Formatted: Not Highlight	
Formatted: Not Highlight	
Formatted: Not Highlight	
Formatted: Not Highlight	
Formatted: Highlight	
Formatted: Not Highlight	
Formatted: Not Highlight	
Formatted	... [1]
Formatted	... [2]
Formatted	... [3]
Formatted	... [4]
Formatted	... [5]
Formatted	... [6]
Formatted	... [7]
Formatted	... [8]
Formatted	... [9]
Formatted	... [10]
Formatted	... [11]
Formatted	... [12]
Formatted	... [13]
Formatted	... [14]
Formatted	... [15]
Formatted	... [16]
Formatted	... [17]
Formatted	... [18]
Formatted	... [19]
Formatted	... [20]
Formatted	... [21]
Formatted	... [22]
Formatted	... [23]
Formatted	... [24]
Formatted	... [25]
Formatted	... [26]
Formatted	... [27]
Formatted	... [28]
Formatted	... [29]
Formatted	... [30]
Formatted	... [31]
Formatted	... [32]

10,000 iterations), it can be seen that some of the inter-core comparisons could in essence relate to differences in picking, *i.e.* different 'months' picked between grid-points may exacerbate or accentuate differences. Likewise, adding random variability, between -0.4 and 0.4 ‰ (Figure 3G and 3I), may also reduce the differences between areas of high Total variance and low Total variance. Though the values associated with machine error (-0.12 to 0.12 ‰) appear to do little to affect the range (Figure 3H and 3J). Whilst again the values plotted are not the absolute values, variability added in an attempt to mimic biological variation of a given time slice, increases the range of possible standard deviations in regions with low Total variance (Figure 3K and 3L), therefore understanding the biological variability on shorter timescale which here may be over exaggerated may be crucial (e.g., Steinhardt et al., 2015; Mikis et al., 2019).

### 3.2 FAME Output: Anderson-Darling test

Using a basin-wide statistical test, we examine whether the mean- $\delta^{18}\text{O}_c$  values of a given El Niño foraminifera population ( $FP_{EN}$ ) and a given non-El Niño ('Neutral conditions') foraminifera population ( $FP_{NEU}$ ) can be expected to be significantly different at any given specific location. Where  $FP_{EN}$  and  $FP_{NEU}$  exhibit significantly different distributions mean- $\delta^{18}\text{O}_c$  values, ENSO events can potentially be detected by paleoceanographers and unmixed using, for example, a simple mixing algorithm with individual foraminiferal analysis (Wit et al., 2013). In cases where  $FP_{EN}$  and  $FP_{NEU}$  do not exhibit significantly different means values, then the chosen species and/or location represent a poor choice to study ENSO dynamics. Each simulation time step was placed into a climate states: identification of timesteps that represent El Niño (EN), Neutral (NEU), and La Niña climate conditions was done using the Oceanic Nino Index (ONI) derivative (Huang et al., 2017) (Figure 3). Comparison, for each species, FAME's predicted growth-weighted  $\delta^{18}\text{O}_c$  and  $T_e$  distributions associated with each climate event was done using an Anderson-Darling (AD) test. This statistical test can be used to determine whether or not two distributions can be said to come from the same population. The results of this test are presented in the following way, in which there are four criteria: areas where the population distributions of the two climate states are found to be statistically similar have black grid cells in all panels referring to the Anderson-Darling test results (Figure 5A-F); the results in which the areas where the populations distributions of two climate states are found to be statistically distinct are shown with two distinct colour schemes depending on whether a computable error can be included (Grey and Hashed) or not (White). For FAME- $\delta^{18}\text{O}_c$  the results where the populations are dissimilar are either plot as grey and hashed for *G. ruber* and *G. sacculifer* or white for *N. dutertrei* (Figure 4). This is because for these two species (*G. ruber* and *G. sacculifer*) we have the possibility to determine how robust these results are. We use the  $1\sigma$  values of the observed (FAME) minus expected (MARGO), as computed by Roche et al. (2018) with the MARGO core top  $\delta^{18}\text{O}_c$  database, as the potential error associated with the FAME model. Regions in which the difference between the two populations are larger than the potential error are associated with grey, whereas those less than the potential error as hashed regions (Figure 5A and B), these errors should be seen as a guide rather than a rejection of a site. Because the MARGO database does not contain *N. dutertrei* we have given the panels concerning this species a separate colour scheme, black represents grid-cells for which the two populations cannot

Formatted: Font: Italic

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Font: (Default) +Headings (Times New Roman), Highlight

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Highlight

Formatted: Subscript

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Highlight

Formatted: Font: Italic

be said to be statistically different, white grid-cells are those in which the two populations can be said to be statistically different (Figure S0). As we do not have a similar way to calculate the error for  $T_e$ , FAME- $T_e$  results are shown (in Figure 4) with this same binary pattern (i.e., white grid-cells are those in which the two populations can be said to be statistically different and black are those in which the two populations can be said to be similar). To reduce the complexity, the overlay of the species Anderson-Darling results (Figure 7) also uses the binary colour scheme (white or black).

Our results show that much of the Pacific Ocean can be considered to have statistically different population between  $FP_{EN}$  and  $FP_{NEU}$  for both  $\delta^{18}O_e$  (Figure 5) and  $T_e$  (Figure 6). We consider that the likely cause for such a remarkable result is due to FAME computing a weighted average and, therefore, the lack of a signal found exclusively within the regions demarked in Figure 5 as El Niño regions could represent how the temperature signal is integrated via an extension of the growth rate; growing season and depth habitat of distinct foraminiferal populations. Taking into account the FAME- $\delta^{18}O_e$  error for *G. ruber* and *G. sacculifer*. We compute growth-weighted  $\delta^{18}O_e$  (Figure 3 and 4) and temperature (Figure 4) distributions for each 1° by 1° grid cell in the fifty-eight year simulation using FAME (Roche et al., 2018), constraining the calculation to the Tropical Pacific Ocean (between -20°S to 20°N and 120°E to -70°W).

we have computed regions in which the difference in oxygen isotopes between the two populations ( $\Delta\delta^{18}O_e$ ) compared with the AD-test is smaller than the aforementioned error (Hatching in Figure 5), i.e. where the mean difference between  $FP_{EN}$  and  $FP_{NEU}$  is within the error. The hatched regions in Figure 5 considerably reduce the areal extent of significant difference between  $FP_{EN}$  and  $FP_{NEU}$ , with the remaining regions aligning with the El Niño 3.4 region (Figure 1). It is important to note that this error relates to the model and in reality, the difference between the climate states could be larger or smaller. No such test was performed on the *N. dutertrei* dataset, because of its absence from the MARGO dataset. Our model produces 696 individual monthly maps for all three species (Figure 3), a series of species-specific histograms for different climatic states for each grid point (Figure S1) and their underlying cumulative distribution functions (Figure S2). While two of the three species (*G. ruber* and *G. sacculifer*) have similar ecologies, they show differences in their resultant  $\delta^{18}O_e$  for the same ocean conditions. These results can subsequently be compared to the recorded climate conditions associated with each grid cell and simulation time step: identification of timesteps that represent El Niño (EN), Neutral (NEU), and La Niña climate conditions was done using the Oceanic Niño Index (ONI) derivative (Huang et al., 2017) (Figure 2). Comparison, for each species, FAME's predicted growth-weighted  $\delta^{18}O_e$  distributions associated with each climate event was done using an Anderson-Darling (AD) test. This statistical test can be used to determine whether or not two distributions can be said to come from the same population. Our results show that much of the Pacific Ocean can be considered to have statistically different population between  $FP_{EN}$  and  $FP_{NEU}$ . We consider that the likely cause for such a remarkable result is due to FAME computing a weighted average and, therefore, the lack of a signal found exclusively within the regions demarked in Figure 5 as El Niño regions could represent how the temperature signal is integrated via an extension of the growth rate, growing season and depth habitat of distinct foraminiferal populations. To determine how robust these results are we use the 1 $\sigma$  values of the observed (FAME) minus expected (MARGO), as computed by Roche et al. (2018) with the MARGO core top  $\delta^{18}O_e$  database, as the potential error associated with the FAME model. We have computed regions in which the difference in

- Formatted: Highlight
- Formatted: Subscript
- Formatted: Subscript
- Formatted: Highlight
- Formatted: Font: Italic
- Formatted: Highlight
- Formatted: Font: Times New Roman, Font color: Custom Color(0,0,10), Pattern: Clear (White)
- Formatted: Highlight
- Formatted: Subscript
- Formatted: Highlight
- Formatted: Font: Italic
- Formatted: Font: Italic
- Formatted: Highlight
- Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

oxygen isotopes between the two populations ( $\Delta\delta^{18}\text{O}_e$ ) compared with the AD-test is smaller than the aforementioned error (Hatching in Figures 3 and 4), i.e. where the mean difference between  $FP_{EN}$  and  $FP_{NEU}$  is within the error. The hatched regions in Figure 3 considerably reduce the areal extent of significant difference between  $FP_{EN}$  and  $FP_{NEU}$ , with the remaining regions aligning with the El Niño 3.4 region (Figure 3). No such test was performed on the *N. dutertrei* dataset, because of its absence from the MARGO dataset. However, *N. dutertrei* areas may be unsuitable for the extraction of a palaeoclimate signal for sedimentological reasons, as will be outlined later. To further test the model-driven results and to assess if they are still consistent when the depth limitation (400 m) is shoaled is varied (to 60, 100, 200 m), the analysis was rerun with depths of 60, 100, 200 and an extreme value of 400 m with the aforementioned range of depths (Figure 5 and 6-7). Whilst it is possible to discern differences between the depths, it is important to note that a large percentage of the tropical Pacific remains accessible to palaeoclimate studies. A shallower depth limitation in the model increases the area for the 'warm' species, suggesting that the influence of a reduced variability in temperature or  $\delta^{18}\text{O}_{eq}$  with a deeper depth limit causes the differences between  $FP_{EN}$  and  $FP_{NEU}$  to be reduced. Overlaying the results of the Anderson-Darling test for all three species (Figure 7) per depth for 60, 100 and 200 m highlights the areas where multi-species comparisons could be made, as a further test we combine a combination of shallower depth for *G. ruber* and deeper depths for *G. sacculifer* and *N. dutertrei* (Pracht et al., 2019). A shallower depth limitation in the model increases the area for the 'warm' species, suggesting that the influence of a reduced variability in temperature or  $\delta^{18}\text{O}_{eq}$  with a deeper depth limit causes the differences between  $FP_{EN}$  and  $FP_{NEU}$  to be reduced.

### 3.3 Test of input parameters (fixed depth: temperature and $\delta^{18}\text{O}_{eq}$ )

The model-driven results were assessed with the underlying observational input dataset, to check if they consistent the input data (temperature and  $\delta^{18}\text{O}_{eq}$ ), these underwent the same statistical test (Figure 4 and Figure 5), although. Instead of a variable depth, we opted for with fixed depths of 5 m, 149 m and 235 m (see section 2.5), giving a Eulerian view (Zhu et al., 2017a) in which to observe the implications of a dynamic depth habitat. The results for each grid point are presented as either black, grey or hashed, areas where the population distributions of the two climate states are found to be statistically similar have black grid cells. Regions in which the difference between the two populations are larger than the potential error are associated with grey, whereas those less than the potential error as hashed regions. The threshold error (i.e., the difference between the means of each distribution) is for temperature (Figure 3A-C) 0.5 °C and for  $\delta^{18}\text{O}_{eq}$  (Figure 3D-F) 0.10 ‰, these errors should be seen as a guide rather than a rejection of a site. By using a fixed depth, non-FAME, these results show that the shallowest depths produce populations that are significantly different both in terms of their mean values and their distributions. In the upper panel of Figure 4, the canonical El Niño 3.4 region is clearly visible at 5 m depth. Whilst differences exist between Anderson-Darling results for the temperature input data (Figure 4) and the FAME AD results ( $\delta^{18}\text{O}$  (Figure 5) and  $T_e$  (Figure 6) Figure 4), for instance close to the Panama isthmus, there are significant similarities between the plots.

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Heading 2

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Subscript

Formatted: Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Subscript

Formatted: Highlight

These plots also show that our FAME data (Figure 5.7), in which we allow foraminiferal growth down deeper than the depths in Roche et al. (2018), are a conservative estimate and thus are on the low-end (Figure 4), to account for potential discrepancies with depth habitats. In the original paper on depth habitats based upon temperatures derived from  $\delta^{18}\text{O}_c$ , Emiliani (1954) cautioned that the depth habitats obtained would represent a weighted average of the total population, and while foraminiferal depth habitats are likely to vary spatiotemporally, the average depth habitat is skewed toward the dominant signal (Mix, 1987).

To further test the model-driven results and to assess if they are still consistent when the depth limitation (400 m) is shoaled (to 60, 100, 200 m), the analysis was rerun with the aforementioned range of depths (Figure 5 and 6). Whilst it is possible to discern differences between the depths, it is important to note that a large percentage of the tropical Pacific remains accessible to palaeoclimate studies. A shallower depth limitation in the model increases the area for the 'warm' species, suggesting that the influence of a reduced variability in temperature or  $\delta^{18}\text{O}_{eq}$  with a deeper depth limit causes the differences between  $FP_{EM}$  and  $FP_{NEU}$  to be reduced.

### 3.42 Water depth and SAR

Our analysis uses reanalysis data for the time period 1958-2015, a hypothetical core that had a comparable resolution would essentially be analogous with a sediment core with a rapid sediment accumulation rate (SAR), representing  $50 \text{ yr cm}^{-1}$  (or  $20 \text{ cm kyr}^{-1}$ ). Based on our analysis, such a hypothetical core could allow for the possible disentanglement of El Niño related signals from the climatic signal, but only in a best-case scenario involving minimal bioturbation, which is unlikely in the case of oxygenated waters. Extracting the oxygen saturation ( $\text{SO}_2$ ) state, of the Pacific Ocean bottom waters from the Annual Climatology WOA13 give values that are predominantly  $>40 \%$  (Figure 9B). Oxygen saturation is the concentration of Oxygen in a medium against the maximum that can be dissolved in the same medium. Whilst annual variability may exist, it is unlikely that bioturbation would be prevented by low oxygen. Therefore, using a cut off value that has been considered sufficiently high enough to outpace bioturbation (e.g., Koutavas and Lynch-Stieglitz, 2003) of  $5 \text{ cm kyr}^{-1}$  (Figure 9A) it can be demonstrated that much of the Pacific has an inferred lower sedimentation rate ( $< 5 \text{ cm kyr}^{-1}$ ; Figure 9C) than this cut off value. To test the influence of bioturbation, the bioturbation simulator SEAMUS was run using the NGRIP time series. The results of SEAMUS highlight the potential single foraminifera depth displacement that low sedimentation rates can result in (Figure 9). Following the current available geochronological method (i.e., age-depth method) such specimens that are displaced in depth are assigned the average age of the depth that they were displaced to, which could result in erroneous interpretations of climate variability when analysis such as IFA is applied (Lougheed et al., 2018). The results of SEAMUS are plotted both as time series of the bioturbated NGRIP signal (Figure 10) and as histograms of the probability of finding a particularly pseudo-foraminifera with a given age in the bioturbation depth (Figure 11). As the bioturbation depth varies between 5, 10 and 15 cm for the different simulations of SAR, the histogram in each panel (in Figure 11) represent different thicknesses of sediment, i.e., for Figure 11 panels a, d, g, and j histograms represent data with a BD thickness of 5 cm. Likewise, the timeseries is plotted with the discrete 1 cm depth median age; the median age of the bioturbation depth (Figure

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Font: 9.5 pt

Formatted: Subscript

Formatted: Font: Not Bold

Formatted: Superscript

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight



11) is the reason why each timeseries does not 'start' at 0 age (Keigwin and Guilderson, 2009). The variance within a single depth in a core largely represents the integrated time signal for that depth, as opposed to the variance of a climatic signal for an inferred (or measured) average age for the depths in question. The proxy variance will be based both upon a non-uniform distribution in temporal frequency of specimens, i.e., older specimens are few compared to younger specimens. A large proportion of the specimens in the BD come from years that are 'proximal' this may give undue confidence that the probability of picking a specimen from these years is higher, however the long-tail of the distribution means that there is an equally high chance of picking a specimen that has come from several thousand years earlier than the discrete-depth's median age. If we consider for the moment this as picking specimens from a box, there is a high chance of picking from a single box that represents the age you want however there is an equally high chance of picking from numerous boxes with varying age. If the spread in the climatic variable is uniform throughout this time then it can be possible to reproduce a similar signal, although this would not by definition represent the actual spread in the actual climatic variable for a given time, however the spread in the climate variable is unlikely to be constant. With a varying spread in the climatic signal bioturbation can introduce the possibility of spurious interpretations, but it is of course more obvious where the measured distributions over-exaggerate the climate signal (e.g., Wit et al., 2013). Furthermore, if we consider that researchers do not pick as randomly as they profess, there is both a size and preservation bias to specimens selected, and size is not constant down-core (e.g., Metcalfe et al., 2015) we can further introduce bias within the dataset. The SEAMUS output that corresponds with our chosen SAR cut-off value of  $5 \text{ cm kyr}^{-1}$  (Figure 10 and 11), the lower limit of our mask (Figure 9), is shown in panels in Figure 10 and 11, much of the region for which FAME is calculated upon has inferred sedimentation rates lower than this cut-off value (Figure 9C to 9H).

Formatted: Highlight

20 An additional factor in the post mortem preservation of the oceanographic signal in foraminiferal shells is whether the shells can be preserved. The GEBCO bathymetry data is binned into 250 m wide bins, and the data normalised to 1.0. As the data contains both bathymetric and topographic (below and above 0 m), the grey area in each histogram represent  $> 0 \text{ m}$  (Figure 11). Whilst, there are differences depending on the cut-off value (Figure 8C to 8E) it is important to note that much of the canonical El Niño 3.4 region (Wang et al., 2017) used in oceanography (Figure 11) is also excluded from these suitable areas. 25 Overlaying the water depth and the SAR with the Anderson-Darling results (Figure 11) highlights that The resolution of the ocean reanalysis data for the time period 1958-2015 would essentially be analogous with a sediment core representing  $50 \text{ yr}^+ \text{ cm}^+$  (or  $20 \text{ cm}^+ \text{ kyr}^+$ ). Based on our analysis, such a hypothetical core with a rapid sediment accumulation rate (SAR) could allow for the possible disentanglement of El Niño related signals from the climatic signal using IFA, but only in a best case scenario involving minimal bioturbation, which is unlikely in the case of oxygenated waters. 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 (Lougheed et al., 2018; Peng et al., 1979), as opposed to other proxies such as corals (Cole and Tudhope, 2017), speleotherms (Chen et al., 2016), or varves in which distinct layers correspond to discrete time intervals (true 'time-series' proxies). 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

Formatted: Highlight

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

population by a significant degree from the background signal, otherwise it will be lost due to the smoothing effect of bioturbation (Hutson, 1980a; Löwemark, 2007; Löwemark et al., 2005, 2008; Löwemark and Grootes, 2004) upon the downcore, discrete-depth signal (Cole and Tudhope, 2017; Mix, 1987). Similarly, the individual characters of El Niño events, which are very short in duration, become lost in the bioturbated sediment record. New geochronological tools, such as dual- $^{14}\text{C}$ - $^{18}\text{O}$  measurements on single foraminifera (Lougheed et al., 2018), show that low sedimentation rate cores can have large variances in age between individual foraminifera present within a discrete 1 cm depth interval (Berger and Heath, 1968; Lougheed et al., 2018). In the case of a high SAR core ( $20\text{ cm}^+\text{kyr}^+$ ), assuming a sediment mixing layer of 10 cm, benthic seafloor processes would result in a minimum 700-year  $1\sigma$  confidence interval for a 1 cm discrete depth age (Lougheed et al., 2018). A consequence of bioturbation in sediment cores is that a series of high magnitude, but low frequency El Niño events could be smoothed out of the downcore, discrete depth record. To investigate which areas of the sea floor can potentially preserve a foraminiferal downcore signal, we overlay our results with a map of time averaged deep-sea SAR (Olson et al., 2016), adapted by Lougheed et al. (2018) to also show seafloor areas under the CCD depth, where carbonate material is not preserved (Berger, 1967, 1970b). Of the total area where  $FP_{EN}$  is significantly different from  $FP_{NEU}$  (i.e. those areas where planktonic foraminiferal flux is suitable for reconstructing past ENSO dynamics), only a small proportion corresponds to areas where the sea floor is both above the CCD ( $< 3500\text{ mbsl}$ ) and SAR is at least  $5\text{ cm/ka}$  (Figure 3I). It is important to note that at certain locations, near islands or seamounts, the SAR and water depth may be high enough to allow for a signal to be preserved (Figure 3I). Also of note is that fact that much of the canonical El Niño 3.4 region (Wang et al., 2017) used in oceanography (Figure 1) is also excluded from these suitable areas in Figure 3D.

## 4.0 Discussion

### 4.1 From Life to Sedimentary Assemblages

Whilst we are principally interested in understanding whether living foraminifera can theoretically reconstruct ENSO, comparison with data requires an additional analysis, data-model comparisons are subjective, nominally supposing that the data is the value to be achieved by the model. However, if the foraminifera modulate the original climate signal, then preservation selectively filters which specimens are conserved whereas bioturbation acts to reorder, transposing the order in which they are recovered from the depth domain. Once the sediment is recovered, the researcher acts as a final filter, which is in essence a random picking. Although technically most researchers will pick whole shells so alongside size selectivity (e.g., Metcalfe et al., 2015) there is also preservation bias associated with picking of foraminifera (e.g. Koutavas and Lynch-Stieglitz, 2003). Whilst the presence of depths in the ocean whereupon calcite is absent from sediments was described in the earliest work (e.g., Murray and Renaud, 1891; Sverdrup, 1942), overlaying maps of measured surface sediment carbonate percentage with water depth in the Pacific Ocean led Bramlette (1961) to coin the term 'compensation depth' (Wise, 1978). This work highlighted the 'narrow' depths (4-5000 m) in the Central Pacific of the CCD. Conceptually Berger (1971) placed

Formatted: Highlight

Formatted: Highlight

Formatted: Normal

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Highlight

Formatted: Highlight

three levels in the Pacific ocean that were descriptive of the aspects (e.g., chemical, palaeontological and sedimentological) of the calcite budget; the saturation depth, demarking supersaturated from undersaturated; the lysocline, the depth at which dissolution becomes noticeable (Berger 1968, 1971); and compensation depth (Bramlette, 1961), in which supply is compensated through dissolution. The aspects of the lysocline was estimated by the faunal assemblages of Parker and Berger (1971, figure's 14 and 15 of that publication), for much of the equatorial Pacific the lysocline is estimated at ~3800 m. As the lysocline is where dissolution becomes apparent, ergo it is a sample already visibly degraded, we therefore set the limit of the water depth mask shallower. In fact, in regions of high fertility, such as the Eastern Equatorial Pacific, the lysocline was estimated to be present at ~2800 m (Thunell et al., 1981) or ~3000 m (Berger, 1971; Parker and Berger, 1971) for instance core V21-28 close to the Galapagos Islands (01°05'N, 87°17'W) has a shallower dissolution affect even at 2714 m (Luz, 1973). A comparison between the hydrographic and sedimentary lysocline, using a mooring in the Panama Basin showed that the sedimentary lysocline is a product of where the hydrographic lysocline meets the seafloor (Thunell et al., 1981), this could lead to dissolution within the water (e.g., Schiebel et al., 2007). In the EEP region the shallower lysocline has an equally shallower CCD, at ~3600 m, here the high fertility is considered responsible for its shoaling, lowering the pH through increased CO<sub>2</sub> (Berger et al., 1976). The correspondence between lysocline depth and CCD depth does not hold true for the entirety of the Pacific, plotting a N-S cross-section from 50°N to 50°S Berger (1971) noted that in the Central Equatorial Pacific, the high fertility region generates a larger zone of dissolution resistant facies even with a shoaled lysocline. If we factor in the sedimentation rate of the Pacific, which has been estimated to be considerably lower than 1 cm (Blackman and Somayajulu, 1966; Berger, 1969; Menard, 1964), then dissolution may become further exacerbated. The longer a shell remains at the sediment-water interface the greater the prospects for it to be dissolved become, therefore low SAR increases the chance of dissolution (Bramlette, 1961). For instance in 15 equatorial Pacific cores, below 4000 m, the average SAR was presented (Hays et al., 1969; here calculated) at 0.96 cm kyr<sup>-1</sup> (1σ ± 0.43 cm kyr<sup>-1</sup>). Although there are regions and/or core locations in which the SAR is higher, for instance eight EEP cores shallower than the lysocline depth (Thunell et al., 1981) of ~2800 m were presented by Koutavas and Lynch-Stieglitz (2003) which have an average SAR, calculated, at 7.20 cm kyr<sup>-1</sup> (1σ ± 2.82 cm kyr<sup>-1</sup>). The average age for these same core's 0 cm core depth is 2184 years (1σ ± 1521 yrs), whilst it cannot be assumed that there has been no loss during recovery (i.e., core top is not sediment-water interface), a non-zero core top age is expected for both bioturbation (Keigwin and Guilderson, 2009) and dissolution. Alongside, the potential for dissolution there is the also the mixing of ocean sediments by the benthos (Bramlette and Bradley, 1942). For instance, Arrhenius (1961) noted that ash beds present in cores of the EEP (Worzel, 1959; Ewing et al., 1959) had a 2-3 cm layer above and below what should have originally been a sharp boundary in which they estimated that ~50% of the material originated from the other side of the boundary. If one assumes 1 cm kyr<sup>-1</sup> sedimentation rates, then the range in age of the obvious 6 cm mixed sediments is minimally ~6000 years per cm, comparison with an analogous SEAMUS simulation (bioturbation depth 5 cm; SAR 1cm) highlights the considerable spread in age, placing the 95.45% range between 110 and 18954 years (Ewing, 1959). Much of this temporal variability will be hidden, especially when proxy values correspond with the expected values, and more obvious when the values are larger than expected (e.g., Wit et al.,

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Subscript

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Superscript

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Font: Italic

2013). Owing to the lack of absolute variability during the Holocene the apparent confirmation of similarity between proxy values and modern distributions of the 'to be reconstructed' variable is not a confirmation of proxy reliability. Especially in the tropics wherein seasonal variability is limited. The effects of both bioturbation and dissolution are further amplified when combined with finite sampling strategies. Therefore, the results of the sedimentological features, presented here, imply that much of the Pacific Ocean is not suitable for preserving (Figure 7-9) the ENSO signal, despite the possibility of the species of foraminifera having unique values for different climate states (Figure 4.7). ENSO studies using palaeoceanography have exposed shifts, one can, therefore, question what is being reconstructed in such studies.

#### 4.1.2 Palaeoceanographic Implications

##### 4.2.1 Pacific climate reconstructions

One artefact of sampling (Dolman and Laepple, 2018) is the potential occurrence of aliasing (Pisias and Mix, 1988; Wunsch, 2000; Wunsch and Gunn, 2003): a fundamental problem with proxy records is that they can be confounded by local regional climate, and/or ENSO's teleconnections, that mimic ENSO changes albeit at a different temporal frequency. Our own analysis using our FAME  $\delta^{18}\text{O}_c$  and  $T_c$  output mimics foraminiferal sedimentary archives, pooling several decades worth of data in which the resolution is coarse enough to obscure and prevent individual El Niño events being visible but allowing for some kind of long-term mean state of ENSO activity to be reconstructed (Cole and Tudhope, 2017). The results of our Anderson-Darling test may be unduly influenced by the Pacific decadal variability (PDV), also referred to as the Pacific Decadal Oscillation (PDO) (Pena et al., 2008). In much of the tropical Pacific the ratio of decadal to interannual  $\sigma\text{SST}$  suggests that they are comparable in magnitude, therefore fluctuations in SST are more obviously apparent outside of the purely canonical regions of ENSO (Wang et al., 2017). It could be that the areas outside of these canonical ENSO regions (Figure 4) reflect the PDO (Pena et al., 2008; Wang et al., 2017). The use of the variance  $\sigma^2(\delta^{18}\text{O}_c)$ , or standard deviation  $\sigma(\delta^{18}\text{O}_c)$ , as an indicator of ENSO is dependent on whether the original climate signal's variance was or was not dominated by interannual variance. Zhu *et al.* (2017) computed the total variance change with and without the annual cycle suggesting that, for some cores the increased assumed ENSO variability at the LGM as deduced by proxy records (Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003) may be purely a by-product of the annual cycle or dominated by it. Although the values of El Niño can be considered significantly different from other climate states, our own analysis using the ratio of total to interannual variance also suggests that much of the variance in the simulated foraminiferal signal is dominated by interannual variance. There are differences in the ratio of total to interannual variance between species and in different regions of the tropical Pacific, however, even with a dynamic depth habitat the ratio is still high (Figure 4; Supplementary Table 1).

Several authors have focussed on individual foraminifera analysis (IFA) or pooled foraminiferal analysis in the Pacific region, either for trace metal or stable isotope geochemistry. Given the complexity in reconstructions of trace metal geochemistry (Elderfield and Ganssen, 2000; Nürnberg et al., 1996), and the potential error associated with determining which carbonate phase is first used when foraminifera biomineralise (Jacob et al., 2017), the focus here has been on the  $\delta^{18}\text{O}_c$ .

Formatted: Heading 3

Formatted: Superscript

Formatted: Highlight

**Commented [BM1]:** authors have focussed on individual foraminifera analysis (IFA) or pooled foraminiferal analysis in the Pacific region, either for trace metal or stable isotope geochemistry. Given the complexity in reconstructions of trace metal geochemistry (Elderfield and Ganssen, 2000; Nürnberg et al., 1996), and the potential error associated with determining which carbonate phase is first used when foraminifera biomineralise (Jacob et al., 2017), the focus here has been on the  $\delta^{18}\text{O}_c$ .

Formatted: Heading 3

has been on the  $\delta^{18}\text{O}_c$ . Palaeoceanographic (Ford et al., 2015; Garidel-Thoron et al., 2007; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Leduc et al., 2009; Sadekov et al., 2013; White et al., 2018) and palaeoclimatological archives (Moy et al., 2002) have been used to indirectly and directly study past ENSO. Several authors have focussed on individual foraminifera analysis (IFA) or pooled foraminiferal analysis in the Pacific region, either for trace metal or stable isotope geochemistry. Given the complexity in reconstructions of trace metal geochemistry (Elderfield and Ganssen, 2000; Nürnberg et al., 1996), and the potential error associated with determining which carbonate phase is first used when foraminifera biomineralise (Jacob et al., 2017), the focus here has been on the  $\delta^{18}\text{O}_c$ . The resultant data of such studies have been used to infer a relatively weaker Walker circulation, a displaced ITCZ and equatorial cooling (Koutavas and Lynch-Stieglitz, 2003); both a reduction (Koutavas and Lynch-Stieglitz, 2003) and intensification (Dubois et al., 2009) in eastern equatorial Pacific upwelling; and both weakened (Leduc et al., 2009) and strengthened ENSO variability (Koutavas and Joanides, 2012; Sadekov et al., 2013) during the LGM. A number of these results are contentious, for instance the reduction in upwelling in this region (Koutavas and Lynch-Stieglitz, 2003) is contradicted by Dubois et al. (2009), who used alkenones (i.e.,  $U_{37}^{K'}$  ratios) to suggest an upwelling intensification. Whilst the  $U_{37}^{K'}$  proxy has problems within coastal upwelling sites (Kienast et al., 2012) it does not discount their claim, especially considering that  $\delta^{18}\text{O}$  records can themselves be influenced by salinity upon the  $\delta^{18}\text{O}_{sw}$  component (Rincón-Martínez et al., 2011) and the potential influence of  $[\text{CO}_3^{2-}]$  upon foraminiferal  $\delta^{18}\text{O}_c$  (de Nooijer et al., 2009; Spero et al., 1997; Spero and Lea, 1996). 4.2.2 The use of models in reconstructions

Formatted: English (United Kingdom)

Reconstructions of the past climate of the Pacific have inferred a relatively weaker Walker circulation, a displaced ITCZ and equatorial cooling (Koutavas and Lynch-Stieglitz, 2003); both a reduction (Koutavas and Lynch-Stieglitz, 2003) and intensification (Dubois et al., 2009) in eastern equatorial Pacific upwelling; and both weakened (Leduc et al., 2009) and strengthened ENSO variability (Koutavas and Joanides, 2012; Sadekov et al., 2013) during the LGM. However, a number of the inferences are contentious, for instance the reduction in upwelling in this region (Koutavas and Lynch-Stieglitz, 2003) is contradicted by Dubois et al. (2009), who used alkenones (i.e.,  $U_{37}^{K'}$  ratios) to suggest an upwelling intensification (Zhang et al., 2017). Whilst the  $U_{37}^{K'}$  proxy has problems within coastal upwelling sites (Kienast et al., 2012) it does not discount their claim, especially considering that  $\delta^{18}\text{O}$  records can themselves be influenced by salinity upon the  $\delta^{18}\text{O}_{sw}$  component (Rincón-Martínez et al., 2011) and the potential influence of  $[\text{CO}_3^{2-}]$  upon foraminiferal  $\delta^{18}\text{O}_c$  (de Nooijer et al., 2009; Spero et al., 1997; Spero and Lea, 1996). Furthermore, the discrepancies in reconstructed climate between marine cores' is worth noting, as ultimately it is from proxies that inferences are made about past climate (Trenberth and Otto-Bliesner, 2003; Rosenthal and Broccoli, 2004). Such inferences have suggest that the past climate of the Pacific region (from the geologically recent too deep time) has been in an: -El Niño state (Koutavas et al., 2002; Stott et al., 2002; Koutavas and Lynch-Stieglitz, 2003); permanent El Niño state (Huber and Caballero, 2003) or Super El Niño state (Stott et al., 2002); La Niña state (Andreasen et al., 2001; Beaufort et al., 2001; Martínez et al., 2003); or a different climatic state (Pisias and Mix, 1997; Feldberg and Mix, 2003).

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

The possibility of a marine sediment archive being able to reconstruct ENSO dynamics comes down to several fundamentals besides whether the signal can or cannot be preserved (i.e., whether the core site has either too low SAR, too high BD or a water depth not conducive to calcite preservation); 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; as well as what the proxy is recording. There is also the presumption that a particular climate

Formatted: Font: Italic

event should be recorded, our Anderson-Darling test for instance highlights that there are locations that cannot discern the difference between El Niño and other climate states whilst for the same time period there are locations where the different climate states can be differentiated. Whilst our analysis is a statistical treatment of the data, each species, and different types of phyto- or zooplankton preserved in ocean sediments, are likely to record the same set of environmental conditions differently (Mix, 2006). This is, in brief, the rationale for the development of FAME, the same climate signal seen through the view of species-specific proxies will give a fractured view constrained by each species particular ecophysiological constraints (Mix, 1987; Roche et al., 2018). A dynamic depth habitat in which the environmental signal becomes a weighted average of the water column can further confound the original signal (Wilke et al., 2006). What can be seen as contradictory reconstructions can therefore be viewed as the prevailing or dominant conditions at a given location at the time when environmental conditions overlap ecological constraints for a given species.

especially considering that Terrestrial records suggest the number of El Niño events per century in the early Holocene (8-6 ka BP) was minimal (Moy et al., 2002), with between 0 and 10 events occurring per century. This dampened ENSO is observed within lake core colour intensity and records driven primarily by precipitation - although like other proxies this can also be interpreted differently, *i.e.* as a large changes in the hydrological cycle shifting precipitation away regionally (Trenberth and Otto-Bliiesner, 2003) (Cole and Tudhope, 2017; White et al., 2018). If we assume for now that 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. The sensitivity and probability of detecting a change in IFA with changes in frequency and amplitude, has been dealt with before (Thirumalai et al., 2013), although without considering bioturbation. 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.

The synthesis of pseudo-timeseries to discern the potential distribution for different scenarios, whilst a necessary approximation, is nonetheless one that is free of cause and causality. Modulating a timeseries for events with enhanced or weakened amplitude or fewer or greater number of events assumes in essence that there is limited feedback both regionally (between two sites) and internally within the timeseries (*i.e.*, a process that operates on a higher level). Reconstructions of the past can benefit from inclusion within conceptual frameworks that incorporate both data and modelling studies (*e.g.*, Trenberth and Otto-Bliiesner, 2003; Rosenthal and Broccoli, 2004; McPhaden et al., 2006). The use of coupled ocean-atmosphere models (*e.g.*, Clement et al., 1999; Zebiak and Cane, 1987); isotope enabled Earth system models (*e.g.*, iCESM; Zhu et al., 2017); or multi-model ensembles with prescribed boundary conditions can be used for the generation of timeseries in which the physics of atmospheric and oceanic circulation are constrained and feedbacks between sites can occur. The perceived failure of several climate models to resolve ENSO adequately, resulting in variable ENSO frequency and amplitude between models, could therefore be used to determine the proxy signal from model derived timeseries at different frequencies and intensities of ENSO. Albeit a timeseries of variable ENSO that is grounded in ocean-atmosphere coupling.

Formatted: Font: Italic

Formatted: Highlight

Formatted: Not Highlight

Formatted: Highlight

Such analysis could also provide information on a secondary assumption, in which time slices from the same core inherently assume that where a particular oceanographic feature exists now is also where it may have existed before. This gives a somewhat binary view, the feature either occurs or does not occur, and if it occurs then it has either enhanced or weakened. Yet this can (though not always) preclude a scenario in which the feature has shifted. Analysis of the El Nino patterns suggests that there are two types of El Nino that are spatially delineated: the dateline Central Pacific El Nino and the Eastern Pacific El Nino. The expansion, contraction or shift of certain large scale oceanographic features (e.g., Polar Front, Upwelling) during periods of warmer than average (e.g., the last interglacial) or colder than average temperatures (e.g. the LGM) can complicate the comparison of two down core samples, i.e., a static core continuously recording a particular climate event as opposed to a shifting oceanographic regime passing over or beyond a core site (Weyl, 1978). Climate models could therefore also be used to determine applicable core locations for comparison of proxy values with 'like with like' oceanographic features (similar to the analysis of Evans et al. (1998) for predicting coral sites), without necessarily the cost of a time-slice project (e.g., CLIMAP, MARGO).

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  $\sigma(\delta^{18}O_e)$  of surface and thermocline dwelling foraminifera. One can, therefore, question what is being reconstructed in such studies. One artefact of sampling is the potential occurrence of aliasing (Pisias and Mix, 1988; Wunsch, 2000; Wunsch and Gunn, 2003): a fundamental problem with proxy records is that they can be confounded by local regional climate, and/or ENSO's teleconnections, that mimic ENSO changes albeit at a different temporal frequency (Dolman and Laepple, 2018). Our own analysis using our FAME  $\delta^{18}O_e$  and  $T_e$  output mimics foraminiferal sedimentary archives, pooling several decades worth of data in which the resolution is coarse enough to obscure and prevent individual El Niño events being visible but allowing for some kind of long term mean state of ENSO activity to be reconstructed (Cole and Tudhope, 2017). The results of our Anderson-Darling test may be unduly influenced by the Pacific decadal variability (PDV), also referred to as the Pacific Decadal Oscillation (PDO) (Pena et al., 2008). In much of the tropical Pacific the ratio of decadal to interannual  $\sigma$ SST suggests that they are comparable in magnitude, therefore fluctuations in SST are more obviously apparent outside of the purely canonical regions of ENSO (Wang et al., 2017). It could be that the areas outside of these canonical ENSO regions (Figure 1) reflect the PDO (Pena et al., 2008; Wang et al., 2017). The use of  $\sigma(\delta^{18}O_e)$  as an indicator of ENSO is dependent on whether the variance is dominated by interannual variance (Ford et al., 2015; Thirumalai et al., 2013; Zhu et al., 2017a). Zhu et al. (2017) computed the total variance change with and without the annual cycle suggesting that, for some cores the increased assumed ENSO variability at the LGM as deduced by proxy records (Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003) may be purely a by-product of the annual cycle.

#### 4.2 Limitations of the methods applied and assessment of model uncertainties

For simplicity we have assumed that our model is 'perfect', of course that is inaccurate, there are four potential sources of error: the input variables (temperature, salinity and their conversion into  $\delta^{18}O_{sw}$  and  $\delta^{18}O_{eq}$ ); the model's error with respect to

Formatted: Font: Italic

Formatted: Highlight

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) +Headings (Times New Roman)

real world values (Roche et al., 2018); the statistical test's errors (associated Type I – in which attribution of significance is given to an insignificant random event, a false 'positive' – and Type II – in which a significant event is attributed to be insignificant, a false 'negative' - errors); and reducing the complexities of foraminiferal biology via parameterization. The input variables can have errors associated with both the absolute values of temperature and salinity used here; and the limitation of input values to a single value per month. Whilst it is possible to interpolate to a daily resolution, this is problematic for two reasons: (1) daily temperature records have much more high frequency oscillations than the data here and (2) the lifecycle of a single foraminifera is approximately monthly, therefore by using monthly data it provides an estimate of the average population signal. Conversion of salinity and temperature into  $\delta^{18}\text{O}_{\text{sw}}$  and  $\delta^{18}\text{O}_{\text{eq}}$  uses a quadratic approximation, one source of error is the unknown influence of carbonate ion concentration on both the Kim and O'Neil (1997) equation and the foraminiferal microenvironment (de Nooijer et al., 2008, 2009; Spero et al., 1997; Spero and DeNiro, 1987; Spero and Lea, 1996) which has further implications due to the upwelling of cool, low pH, waters in the eastern Tropical Pacific (Cole and Tudhope, 2017; Raven et al., 2005). The spatial variability in salinity, particularly within regions underlying the intertropical convergence zone (ITCZ) and the moisture transport from the Caribbean into the eastern Pacific along the topographic low that represents Panama Isthmus, the resultant conversion of salinity to  $\delta^{18}\text{O}_{\text{sw}}$  and then  $\delta^{18}\text{O}_{\text{eq}}$  may contain further error. If such errors are independent of the absolute value of the variable, *i.e.* the error on cold temperature is the same and not larger than warm temperatures, then the error terms effectively cancel one another out. A point of note, is that the  $\delta^{18}\text{O}$  to  $^{\circ}\text{C}$  conversion of Kim and O'Neil (1997) is considered to be marginally larger at the cold end than at the warm end (0.2 ‰ per  $1^{\circ}\text{C}$  to 0.22 ‰ per  $1^{\circ}\text{C}$ ) than that originally discerned (O'Neil et al., 1969).

The comparison of the pseudo-Mg/Ca temperature signal produced here ( $T_c$ ) to a value corresponding to that reconstructed from measurements of Mg/Ca should be done with caution. Computation of pseudo-foraminiferal  $\delta^{18}\text{O}$  in FAME is aided by the ability to compute an initial  $\delta^{18}\text{O}$  equilibrium value for a given latitude-longitude grid-point and timestep. The weighting of  $\delta^{18}\text{O}$  value used in FAME is an approximation of the foraminiferal shell, chambers are generally homogenous in  $\delta^{18}\text{O}$  value excluding either terminal features such as crust or gametogenic calcite which can lead to chamber heterogeneities (e.g., Wycech et al., 2018) although this can be approximated with an additional parameter (Roche et al., 2018). The same cannot be said for Mg/Ca, alongside heterogeneities in the shell which may be the result of diurnal processes (terminal features in the computation of  $\delta^{18}\text{O}$  are technically simpler to model), there are differences in both sample preparation and measurement techniques. Whilst, the change in Mg/Ca with temperature has been validated (e.g., Elderfield and Ganssen, 2000) the computation of a pseudo-proxy value for and from model parameters remains enigmatic. Construction of a matrix of equilibrium Mg/Ca would ideally be the most logical step in a second generation of the FAME model. Whilst, simply solving the Mg/Ca palaeotemperature equation for an input of T and an output Mg/Ca is a first approximation, aAs stated previously, several other parameters can alter this technique, this includes -abiotic effects such as salinity (Allen et al., 2016; Gray et al., 2018; Groeneveld et al., 2008; Kısakürek et al., 2008) or carbonate ion concentration (Allen et al., 2016; Evans et al., 2018; Zeebe and Sanyal, 2002); biotic effects such as diurnal calcification (Eggins et al., 2003; Hori et al., 2018; Sadekov et al., 2008, 2009; Vetter et al., 2013); or additional factors such as sediment (Fallet et al., 2009; Feldmeijer et al.,

Formatted: Not Highlight

Formatted: Font: Italic



2013) or specimen (Barker et al., 2003; Greaves et al., 2005) ‘cleaning’ techniques. Given the role of Mg in inhibiting calcium carbonate formation, the manipulation of seawater similar to the modification of the cell’s pH (de Nooijer et al., 2008, 2009) may aid calcification and explain the formation of low-Mg by certain foraminifera (Zeebe and Sanyal, 2002). ~~S~~ ~~however~~ scaling these processes up to a basin-wide model is beyond the remit of this current paper.

5 Our modelling results also depend upon the species symbiotic nature and potential genotypes. For instance, mixotrophs, those organisms that utilise a mixture of sources for energy and carbon (planktonic foraminifera such as *G. ruber*; and/or *G. sacculifer*) can outcompete heterotrophic (or photoheterotrophic) organisms (planktonic foraminifera such as *Neogloboquadrina pachyderma*; *N. eogloboquadrina incompta*) especially in stratified-oligotrophic waters. FORAMCLIM has placed both *G. bulloides* and *N. dutertrei* as symbiont barren for the purposes of ecophysiological parameters used for computing expected potential habitats for both future and past scenarios. However, recent culture studies have shown that *G. bulloides* Type II-d instead of being symbiont barren actually plays host to endobiotic cyanobacteria (Bird et al., 2017), whilst *N. dutertrei* Type I-e contains what are reputed to be algal symbionts (Bird et al., 2018). They do however, confirm that *N. incompta* is symbiont barren (Bird et al., 2018). The authors have suggested that ecophysiological models, such as FORAMCLIM, should be updated to include both the photosynthetic changes and resultant respiration parameters (Bird et al., 2018). Whilst FAME uses only the temperature component of FORAMCLIM (Roche et al., 2018) and we have only modelled *N. dutertrei*, it is important to note that there are distinctions between the fundamental niche that FAME computes, *i.e.* the conditions that an organism can survive, and the realised niche, *i.e.* what an organism actually occupies given limiting factors within the environment. Likewise, FAME and FORAMCLIM are based upon the original culture experiments that assumed that both species (*G. bulloides* and *N. dutertrei*) are non-symbiotic (see Bird et al., 2018, 2017). A species that hosts symbionts will likely have a restricted temperature that is associated with the temperature tolerance of their symbionts, given that the next generation of a species of planktonic foraminifera must be re-infected with their symbionts. Likewise, Variation in cryptic speciation may exist of *G. bulloides* do exist, lead to foraminiferal given that genotypes of morphotypes exhibiting distinct environmental preferences (Bird et al., 2018, 2017; Darling et al., 2004, 2000, 1999; Huber et al., 1997; Morard et al., 2013; de Vargas et al., 1999, 2002), especially between genotypes Type Ia, Ib, Ha, Ib, and He and the genotype Type II-d that has a restricted geographic distribution (Darling et al., 2004). Incorporation of both a theoretical genotype abundance (Morard et al., 2013) and ecophysiological tolerances of different genotypes (Bird et al., 2018) within an ecophysiological model could further reduce error within modelling of planktonic foraminiferal habitats, and thus reduce data-model comparison error. For instance, Morard et al. (2013) simulated the impact of genotypes upon palaeoceanographic reconstructions (in particular transfer functions) using a theoretical abundance, calculated with a best-fit gaussian response model, depending upon SST later using a similar approach (Morard et al., 2016) to deduce the impact upon  $\delta^{18}\text{O}$ . Incorporation of both a theoretical genotype abundance (Morard et al., 2013) and ecophysiological tolerances of different genotypes (Bird et al., 2018) within an ecophysiological model could further reduce error within modelling of planktonic foraminiferal habitats, and thus reduce data-model comparison error.

## Conclusion

Previous work has focussed on comparing either the difference between time slices or between downcore and modern IFA variance, using a reduction in  $\sigma(\delta^{18}\text{O}_c)$  as an indicator of reduced ENSO (Zhu et al., 2017a). Concentrating on the period spanning the instrumental record by using the FAME module, we present forward modelled the species-specific (*i.e.*, *G. ruber*, *G. sacculifer* and *N. dutertrei*) oxygen isotope values ( $\delta^{18}\text{O}_c$ ) and pseudo-Temperature output ( $T_c$ ), computed from for ocean reanalysis data. The aim of this study was to determine whether to validate the modelled values from different climate states are statistically different, our results suggest for large expanses of the Tropical Pacific the climate states do have different values. Whilst, the results show that the values between El Niño states and Neutral climate states are statistically different for a large proportion of the Tropical Pacific, the total variance is dominated by the interannual variance for much of the region. Overlaying our computed foraminiferal distributions with the characteristics of the Pacific Ocean we infer that much of the signal recorded in foraminifera corresponds to areas where several processes will alter the preservation of the foraminiferal signal. First, the inferred SAR for much of the region is critically low, and a simulation of bioturbation for different bioturbation depths and SAR, typical for the Pacific indicates that discrete core depths can have a large temporal spread in single foraminifera, possibly precluding the extraction of ENSO-related climate variability. Second, a large proportion of the seafloor lies below the lysocline, the depth at which dissolution of foraminifera becomes apparent. These factors reduce the size of the area available for reconstructions considerably, thus arguably precluding the extraction of a temporally valid palaeoclimate signal using long-standing methods. It is our inference that only at exceptional ocean sediment core sites is it possible to determine the variability in ENSO, which makes it difficult to build a Pacific basin-wide understanding of past ENSO dynamics. use of foraminiferal proxies in reconstructing ENSO dynamics in the Pacific Ocean.

Overall, our results suggest that foraminiferal  $\delta^{18}\text{O}$  for a large part of the Pacific Ocean can be used to reconstruct ENSO, especially if an individual foraminiferal analysis (Lougheed et al., 2018; Wit et al., 2013) approach is used (Ford et al., 2015; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Sadekov et al., 2013; White et al., 2018), contrary to previous analysis (Thirumalai et al., 2013). 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. 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. The use of ecophysiological models (Kageyama et al., 2013; Lombard et al., 2009, 2011) 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.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Subscript

Formatted: Not Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

### Code and data availability

The ocean reanalysis data used in this paper are available from the Universiteit Hamburg. ~~An~~ open source version of the FAME code is available from Roche et al. (2018). ~~S,~~ ~~and the~~ statistical routines are available as part of the Statistical package of MATLAB R2018a; mapping tools (including the topographic colormap) are part of the Mapping Toolbox. The function to retrieve GEBCO bathymetry (data available at [www.gebco.net](http://www.gebco.net)) from netcdf format, gebconetcdf(FILE,Wlon,Elon,Slat,Nlat), is available from the MATLAB Central File Exchange (<https://mathworks.com/matlabcentral/fileexchange/46669-gebconetcdf-file-wlon-elon-slat-nlat>). The single foraminifera sediment accumulation simulator (SEAMUS) is published in Lougheed (2019), available at [\[waiting for doi from Geoscientific Model Discussions\]](#). A video of the  $\delta^{18}\text{O}_{\text{shell}}$  output has been archived online (<https://doi.org/10.5281/zenodo.2554843>, Metcalfe et al., 2019).

Formatted: Highlight

Formatted: Highlight

### Author Contributions

B.M. and D.M.R. designed the study. B.M. analysed the data. B.C.L. processed ocean SAR and depth data, and ran the bioturbation model. B.M. drafted the manuscript with contributions from all authors.

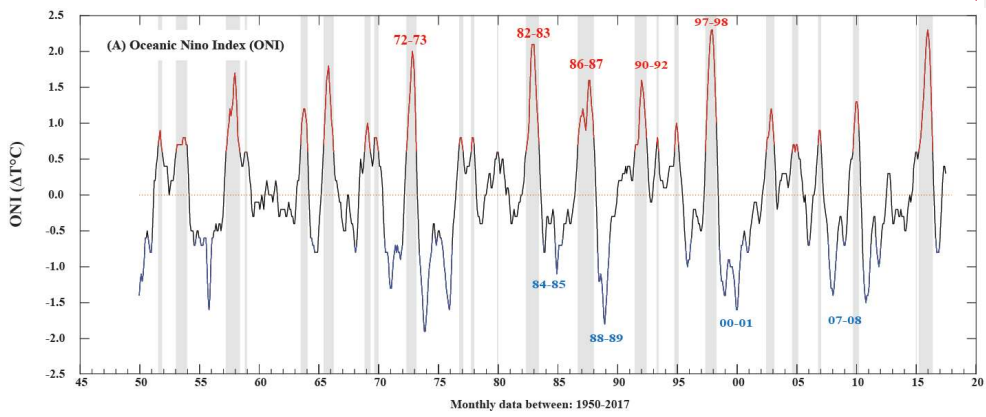
### Competing Interests

The authors declare no competing interests.

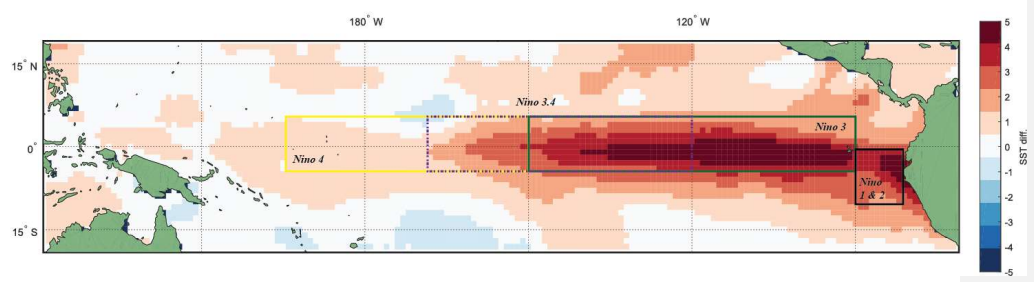
### Acknowledgements

B.M. ~~was~~ supported by a Laboratoire d'excellence (LabEx) of the Institut Pierre-Simon Laplace (Labex L-IPSL), ~~;~~ funded by the French Agence Nationale de la Recherche (grant no. ANR-10-LABX-0018). B.M. thanks both LSCE and the VU University Amsterdam for guest status. D.M.R. is supported by the French agency Centre National de la Recherche Scientifique (CNRS) and the VU University Amsterdam. ~~We thank the Universiteit Hamburg for their online access server for ocean reanalysis data.~~ This is a contribution to the ACCLIMATE ERC project. The research leading to these results has received funding to C.W. from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013 Grant agreement n° 339108). B.C.L. acknowledges Swedish Research Council (Vetenskapsrådet – VR) grant 2018-04992, The Swedish National Infrastructure for Computing (SNIC) at the Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) provided computer resources for running the SEAMUS model. ~~We thank the Universiteit Hamburg for their online access server for ocean reanalysis data.~~





Formatted: Left, Line spacing: single



5 Figure 1. Oceanic Niño Index and the temperature anomaly for a single El Niño event. (Top) Oceanic Niño Index (ONI), sourced from a 3-month running mean of SST anomalies in ERSST.v5 of the Niño 3.4 region (Huang et al., 2017). Grey vertical bars represent the periods in which El Niño-like conditions exist using a simple one-month threshold. (Bottom) The sea surface temperature difference between week beginning 1st December 1997 minus the long-term climatic mean (1971 – 2000) for December. The 1997 – 1998 El Niño represents an EP-ENSO. The long term monthly climatology, the NOAA optimum interpolation (OI) SST V2, based upon the methodology of Reynolds and Smith (Reynolds and Smith, 1995) using two distinct  
10 climatologies for 1971 - 2000 and 1982 – 2000 (Reynolds et al., 2002). Boxes represent the Niño region: Niño 1 and 2 (0° - -10°S, 90°W - 80°W), Niño 3 (5°N - -5°S, 150°W - 90°W), Niño 3.4 (5°N - -5°S, 170°W - 120°W) and Niño 4 (5°N - -5°S, 160°E - 150°W).

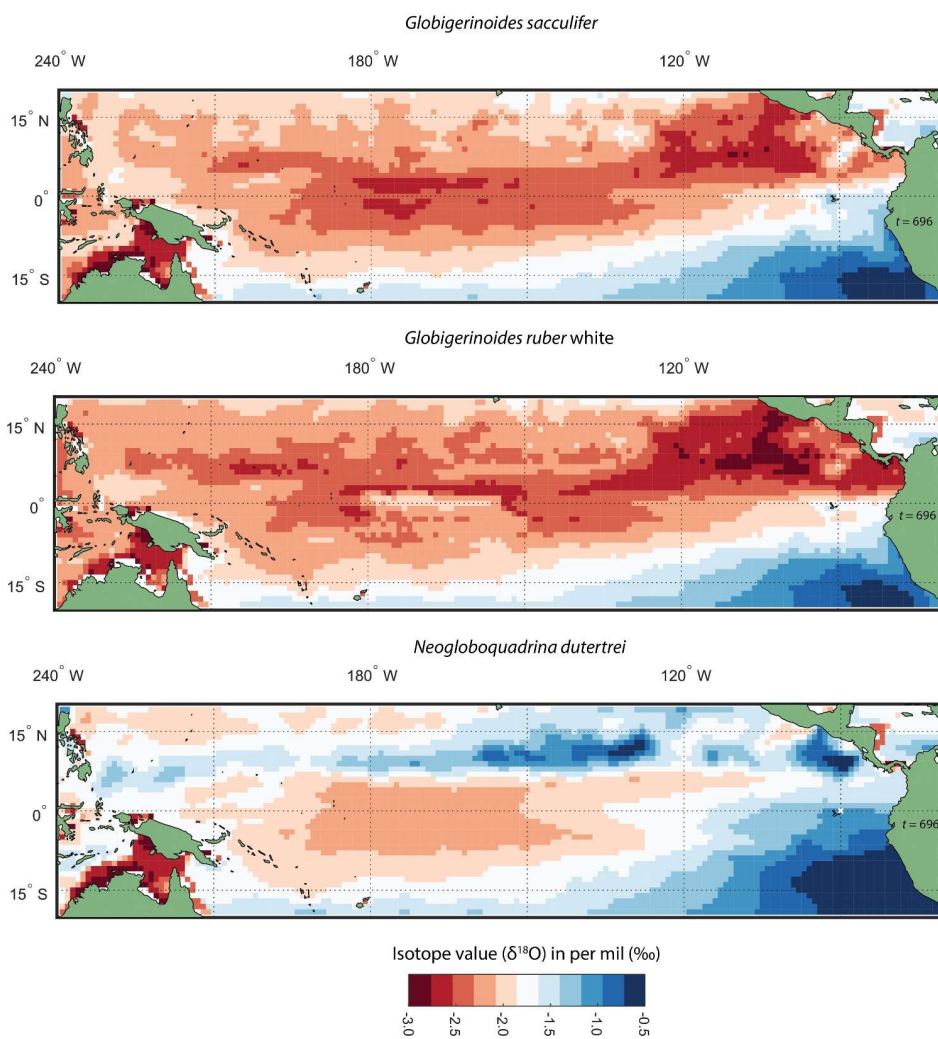
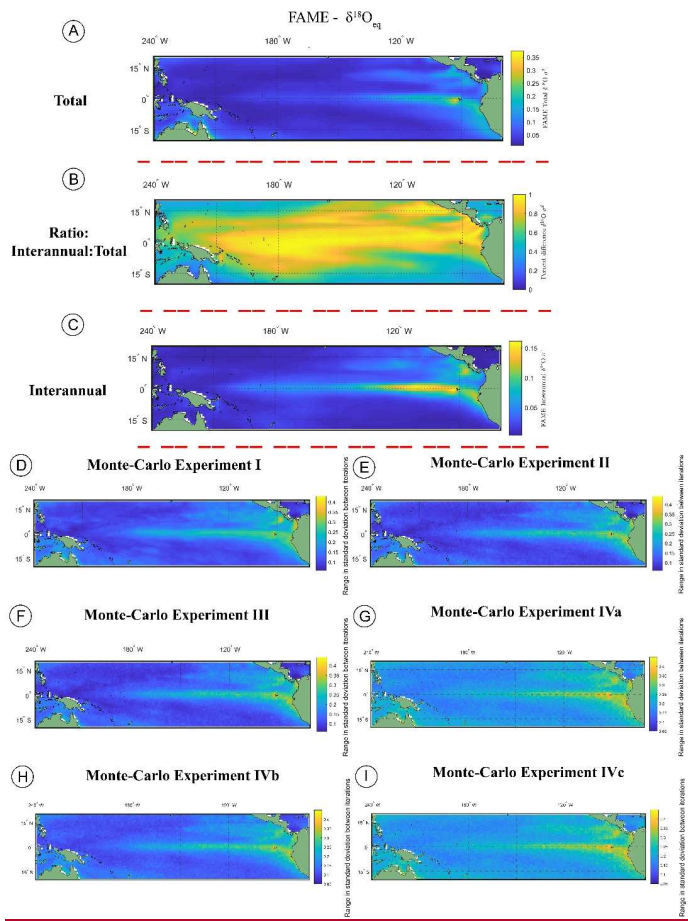


Figure 2. A snapshot of the output of FAME. Each panel represents an individual species (Top Panel *G. sacculifer*; Middle Panel *G. ruber* and Bottom Panel *N. dutertrei*)  $\delta^{18}\text{O}$  for a single time step ( $t = 696$ ). The species  $\delta^{18}\text{O}$  for each grid-point is based upon the integrating the  $\delta^{18}\text{O}_{\text{eq}}$  values using a growth-rate based weighting (FAME; Roche et al., 2018), a computed integrated signal based upon the growth rate and the  $\delta^{18}\text{O}_{\text{eq}}$  values. Values are in per mil (‰ V-PDB).

Formatted: Font: Italic



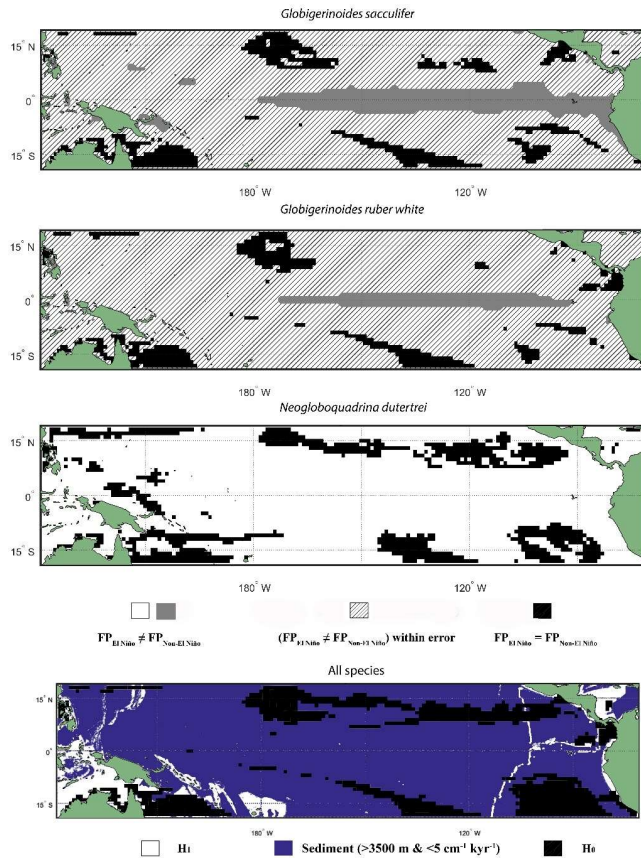


Figure 3. Total variance and Interannual variance and the range in standard deviation of the Monte-Carlo experiments. (a) Total variance of *Globigerinoides sacculifer*  $\delta^{18}\text{O}_{cc}$ , using FAME- $\delta^{18}\text{O}_{cc}$  for a cut-off value of 60 m. (b) The ratio of (a) and (c), where (c) is the Interannual variance of the timeseries of (a). In (d-i) we plot the range in standard deviation obtained by picking 60 months with replacement with 10,000 iterations, the experiments are as follows: (d) the same months were chosen for all grid-points for each iteration of the Monte-Carlo; (e) each grid-point has its own randomly selected months for each iteration of the Monte-Carlo; (f) the same as (e) but we weight the values by the total amount of growth per month; (g) the months selected for (e) were re-run but a random variability is added to each month; (h) the months selected for (e) were re-run but a random measurement error is added to each month; and (i) the months selected for (e) were re-run but the (g) random variability and (h) measurement error were added. Note the scale change between (d-f) and (g-i). Maps of those locations that can reconstruct El Niño events for *G. ruber*, *G. sacculifer*, *N. dutertrei* and all species. Upper (three) panels represent the Anderson-Darling test results for the species *G. ruber*, *G. sacculifer*, and *N. dutertrei* where the black areas reflect latitudinal and longitudinal grid-points that failed to reject the null

Formatted: Font: Italic

Formatted: Superscript

Formatted: Subscript

Formatted: Superscript

Formatted: Subscript



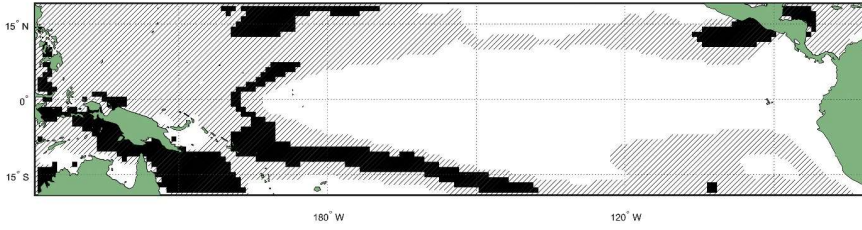
hypothesis ( $H_0$ ) and therefore the foraminiferal population (FP) of the El Niño is similar to the Non-El Niño, and therefore the distribution between the neutral climate and El Niño cannot be said to be different ( $FP_{El\ Niño} = FP_{Non-El\ Niño}$ ). White/grey areas reflect regions in which the  $H_1$  hypothesis is accepted, therefore the distributions can be said to be unique ( $FP_{El\ Niño} \neq FP_{Non-El\ Niño}$ ). Hatching represents the FAME species specific values of 0.32 % and 0.20 %. The final, bottom, panel represents a synthesis of the upper panels areas in which reconstructions are possible, represented by both the blue and white area. The black area represents those areas where the null hypothesis for all three species cannot be rejected. Blue represents sedimentological limitations that preclude the extraction of a foraminiferal signal from the sediment: deeper than 3500 mbsl (below CCD) and time averaged sediment accumulation rate (SAR) of less than  $<5\text{ cm/ka}$  (Lougheed et al., 2018; Olson et al., 2016) (see Figure S7 for the different sedimentological components).

5

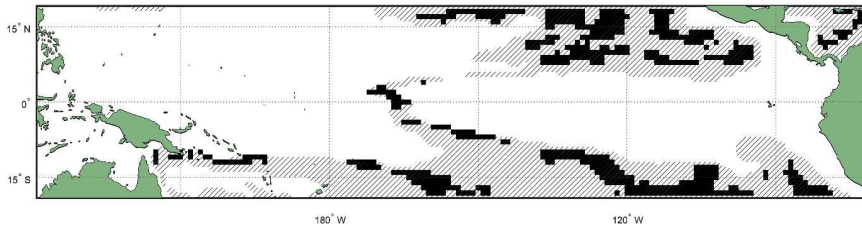
10

Formatted: Normal

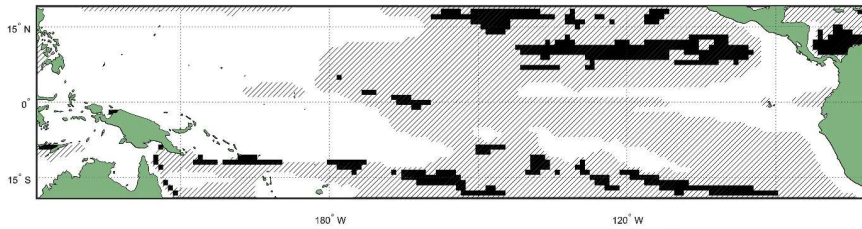
Temperature at 5 m




Temperature at 149 m



Temperature at 235 m



  $FP_{El Niño} \neq FP_{Non-El Niño}$         $(FP_{El Niño} \neq FP_{Non-El Niño})$  within error        $FP_{El Niño} = FP_{Non-El Niño}$

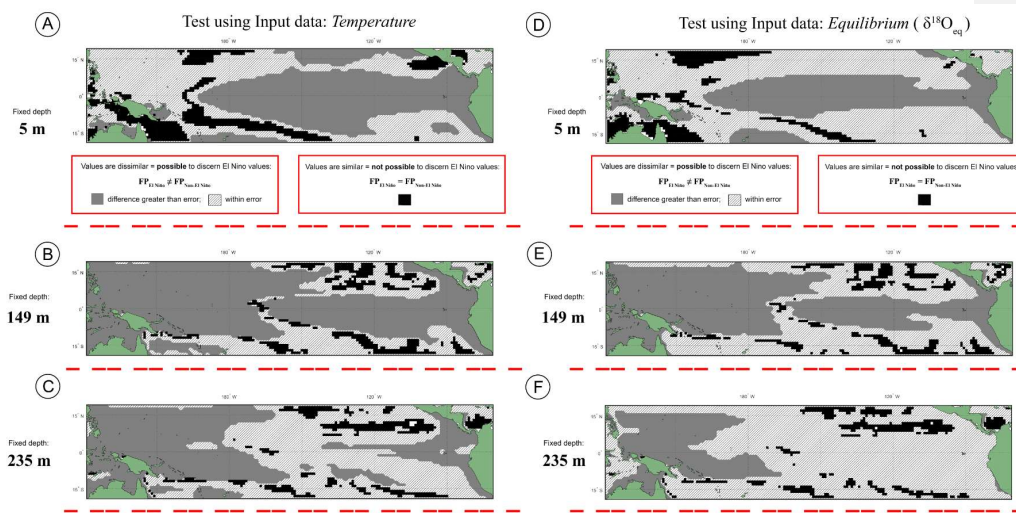


Figure 4. Anderson-Darling Results for Input datasets of Temperature and Equilibrium  $\delta^{18}\text{O}$  ( $\delta^{18}\text{O}_{eq}$ ). Results of the test in which input variables underwent the same statistical procedure (see section 2.0) as the modelled data for (A-C) temperature and (D-F)  $\delta^{18}\text{O}_{eq}$  values. Here, model input data was extracted for three fixed depths (A & D) 5 m; (B & E) 149 m; (C & F) and 235 m) without any growth weighting applied. Black regions are those grid points in which the null hypothesis ( $H_0$ ), that the El Niño and Non- El Niño populations are not statistically different ( $FP_{El\ Niño} = FP_{Non-EI\ Niño}$ ), cannot be rejected. Gray regions Results of an Anderson-Darling test between El Niño and Neutral climate conditions based upon the Temperature input data: Fixed depth. White represent grid points where the  $H_1$  hypothesis is accepted, therefore the distributions of the foraminiferal population (FP) for El Niño and Non- El Niño can be said to be unique ( $FP_{El\ Niño} \neq FP_{Non-EI\ Niño}$ ). Black that the null hypothesis ( $H_0$ ), that the El Niño and Non- El Niño populations are not statistically different ( $FP_{El\ Niño} = FP_{Non-EI\ Niño}$ ), cannot be rejected. The hatched regions represent areas where the  $H_1$  hypothesis can be accepted, therefore the distributions of the foraminiferal population (FP) for El Niño and Non- El Niño can be said to be unique ( $FP_{El\ Niño} \neq FP_{Non-EI\ Niño}$ ), though the difference between the means of tested distribution are less than (A-C) The hatching represents those locations where the difference between the means of both populations is less than  $0.5^\circ\text{C}$  or (D-F) 0.1 ‰. Each panel represents a single depth (5, 15049 and 23550 m) sampled equally through time.

Formatted: Subscript

Formatted: Normal

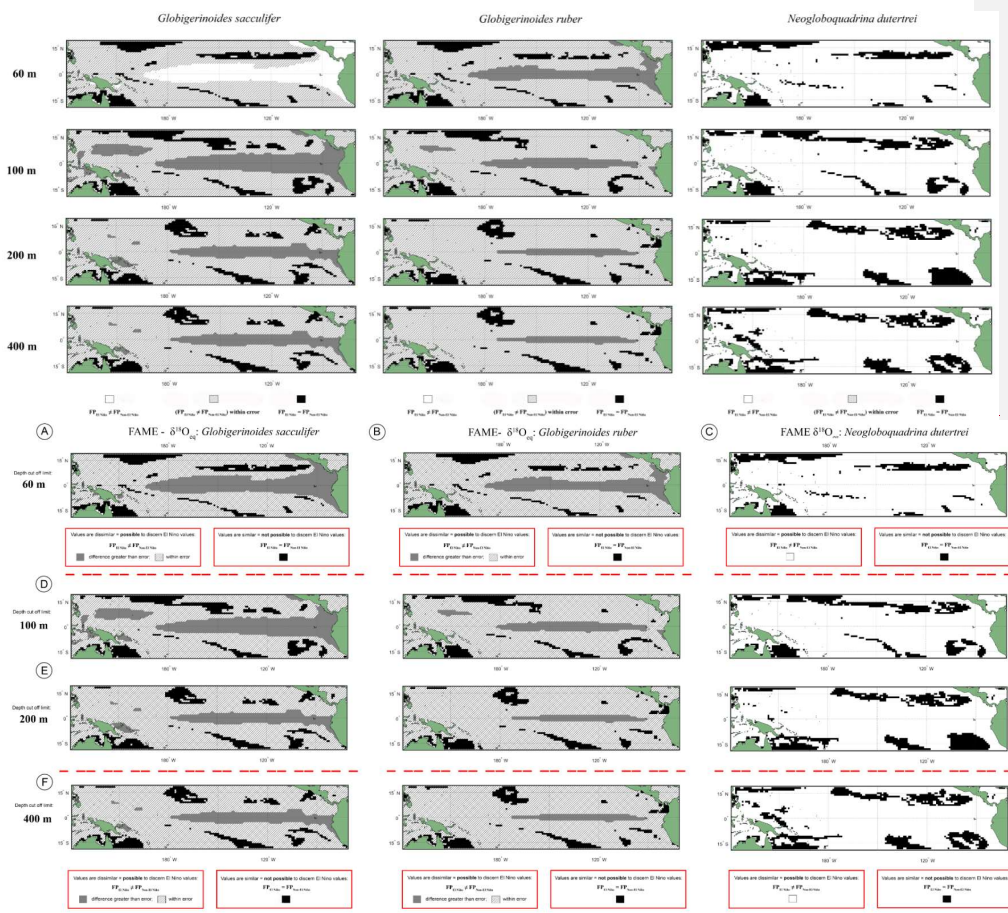


Figure 5. Anderson-Darling Results for modelled FAME-Oxygen isotope ( $\delta^{18}O_{org}$ ): Panels representing Maps of those locations of where dissimilar and similar values of FAME modelled species  $\delta^{18}O$  occur between climate states, for (columns) particular species and (rows) particular model depth cut-off limits. Each panel represents the Anderson-Darling test result, which are plotted with ([A] *Globigerinoides sacculifer* and [B] *Globigerinoides ruber*) and without ([C] *N. dutertrei*) model derived error. For all panels black areas reflect latitudinal and longitudinal grid points that failed to reject the null hypothesis ( $H_0$ ) and therefore the foraminiferal population (FP) of the El Niño is similar to the Non-El Niño ( $FP_{El\ Niño} = FP_{Non-El\ Niño}$ ). The results in which the  $H_1$  hypothesis is accepted, in which the, therefore the distributions can be said to be unique ( $FP_{El\ Niño} \neq FP_{Non-El\ Niño}$ ), are plotted as either: (A – *G. sacculifer*, B – *G. ruber*) grey and hatched or (C – *N. dutertrei*) solely as white regions. For species with calculatable error, grey regions represent values where the difference between the two means of the population is greater than species-specific

Formatted: Font: Italic  
 Formatted: Font: Italic  
 Formatted: Font: Italic  
 Formatted: Font: Italic  
 Formatted: Font: Italic  
 Formatted: Font: Italic  
 Formatted: Font: Italic

standard deviation of the FAME model and hatched regions represent those in which the means are less than this standard deviation (Roche et al., 2018). For species without a calculatable error, the regions are plotted in white. The rows represent the model runs with a depth cut-off limit at: (A-C) 60 m; (D) 100 m; (E) 200 m; and (F) 400 m.

5 that can reconstruct El Niño events for (Left Panel) *G. ruber*; (Middle panel) *G. saeculifer*; and (Right panel) *N. dutertrei*. Panels  
represent the Anderson-Darling test results for the species *G. ruber*, *G. saeculifer*, and *N. dutertrei* where the black areas reflect  
latitudinal and longitudinal grid points that failed to reject the null hypothesis ( $H_0$ ) and therefore the foraminiferal population  
(FP) of the El Niño is similar to the Non-El Niño, and therefore the distribution between the neutral climate and El Niño cannot be  
said to be different ( $FP_{El Niño} = FP_{Non-El Niño}$ ). White/grey areas reflect regions in which the  $H_1$  hypothesis is accepted, therefore the  
distributions can be said to be unique ( $FP_{El Niño} \neq FP_{Non-El Niño}$ ). Hatching represents the FAME species specific values of 0.32% and  
10 0.20%. Each row represents a particular depth limit for the model, from top to bottom: 60 m; 100 m; 200 m and 400 m.

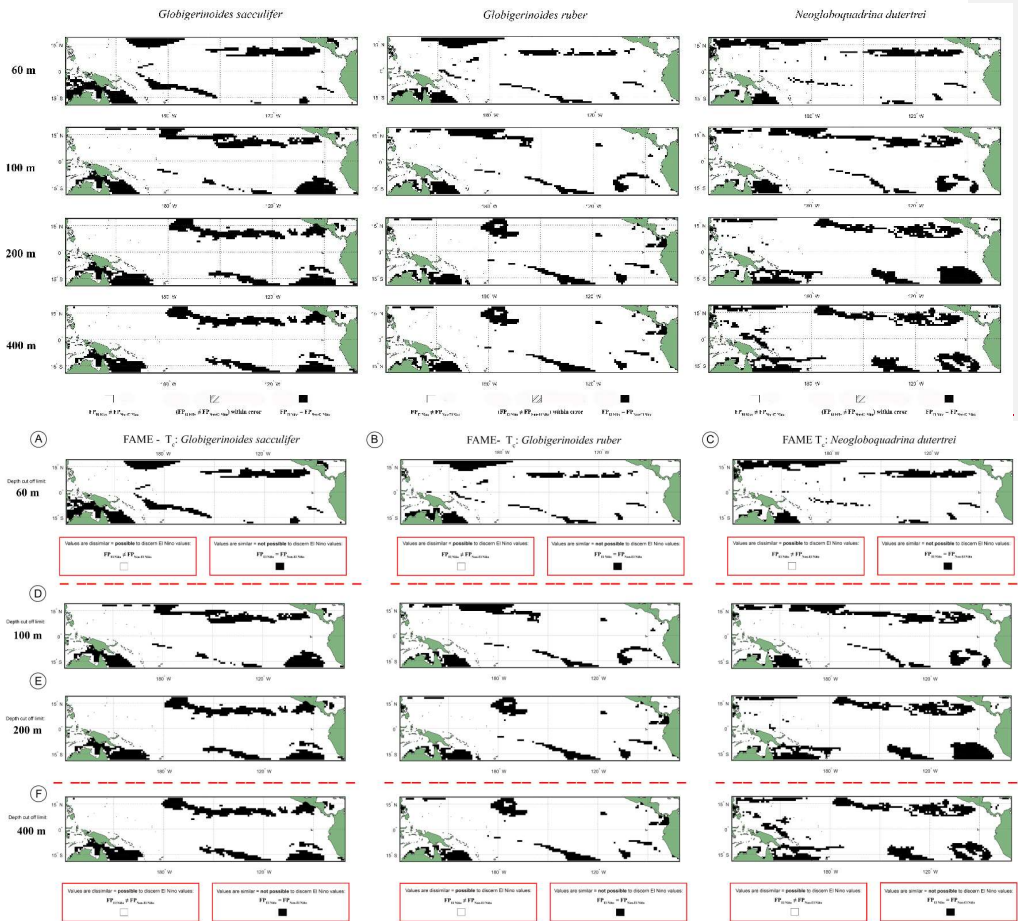
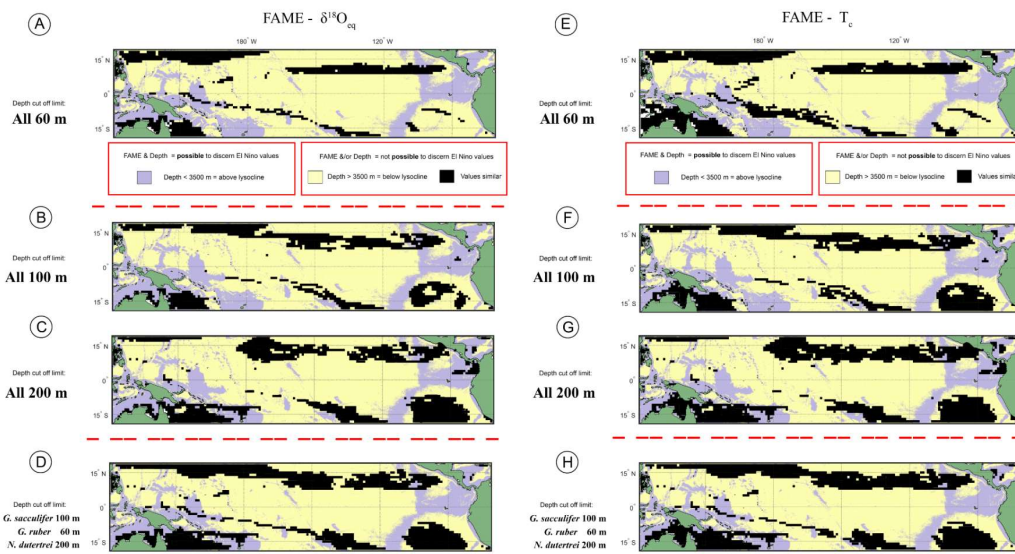


Figure 6. Anderson-Darling Results for modelled FAME-T<sub>c</sub>: Panels representing locations of where dissimilar and similar values of FAME modelled temperature recorded in the calcite shells (T<sub>c</sub>) occur between climate states, for (columns) particular species and (rows) particular model depth cut-off limits. Each panel represents the Anderson-Darling test result, which are plotted with (A) *Globigerinoides sacculifer* and (B) *Globigerinoides ruber* and without (C) *N. dutertrei* model derived error. For all panels black areas reflect latitudinal and longitudinal grid points that failed to reject the null hypothesis (H<sub>0</sub>) and therefore the

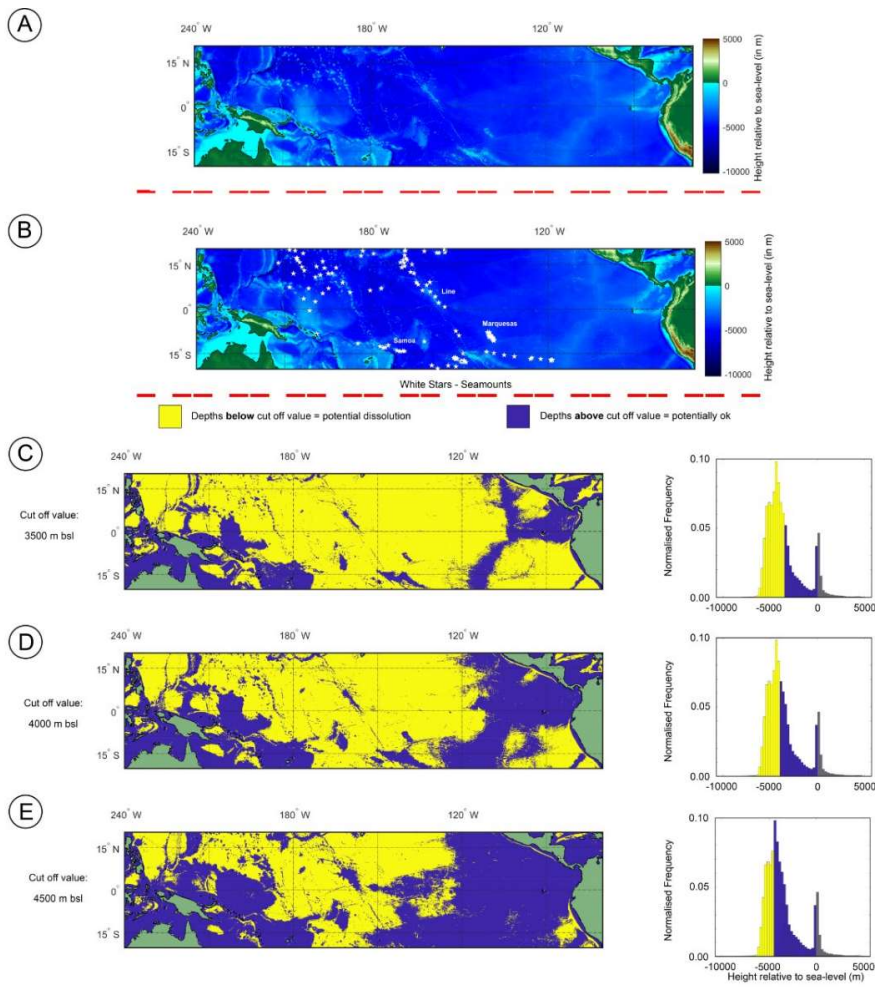
foraminiferal population (FP) of the El Niño is similar to the Non-El Niño, and therefore the distribution between the neutral climate and El Niño cannot be said to be different ( $FP_{El\ Niño} = FP_{Non-El\ Niño}$ ). The results in which the  $H_1$  hypothesis is accepted, in which the distributions can be said to be different ( $FP_{El\ Niño} \neq FP_{Non-El\ Niño}$ ), are plotted as white regions. The rows represent the model runs with a depth cut-off limit at: (A-C) 60 m; (D) 100 m; (E) 200 m; and (F) 400 m.



Temperature signal ( $T_c$ ): Maps of those locations that can reconstruct El Niño events for (Left Panel) *G. ruber*; (Middle panel) *G. sacculifer*; and (Right panel) *N. dutertrei*. Panels represent the Anderson-Darling test results for the species *G. ruber*, *G. sacculifer*, and *N. dutertrei* where the black areas reflect latitudinal and longitudinal grid points that failed to reject the null hypothesis ( $H_0$ ) and therefore the foraminiferal population (FP) of the El Niño is similar to the Non-El Niño, and therefore the distribution between the neutral climate and El Niño cannot be said to be different ( $FP_{El\ Niño} = FP_{Non-El\ Niño}$ ). White/grey areas reflect regions in which the  $H_1$  hypothesis is accepted, therefore the distributions can be said to be unique ( $FP_{El\ Niño} \neq FP_{Non-El\ Niño}$ ). Each row represents a particular depth limit for the model, from top to bottom: 60 m; 100 m; 200 m and 400 m.

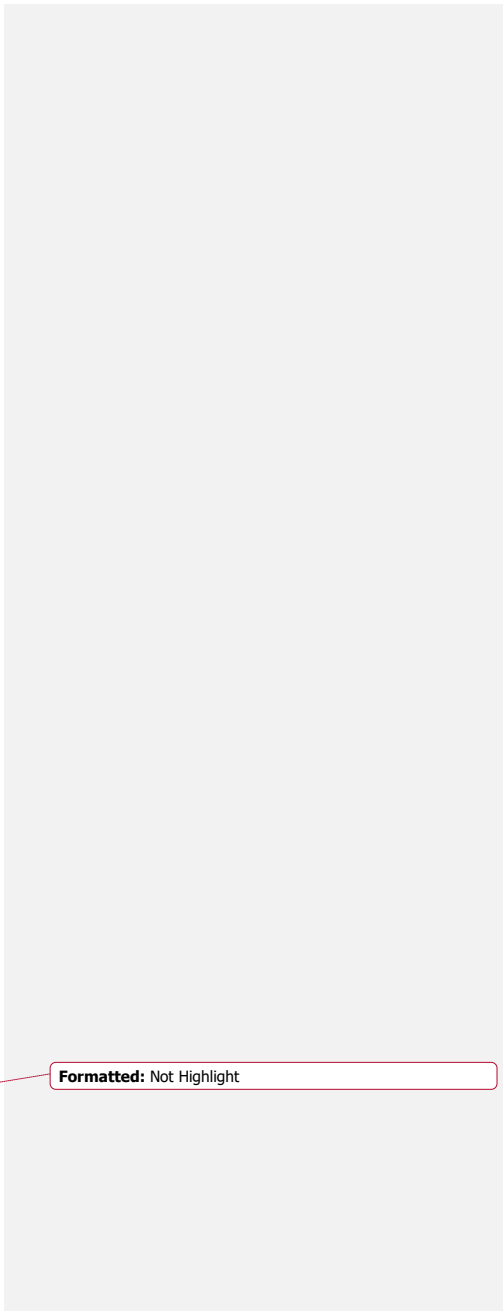
Figure 7. Combined A-D plots. As figure 5 and figure 6, in that panels represent locations of where dissimilar and similar values for the two climate states for (a-d) FAME- $\delta^{18}O_{eq}$  modelled oxygen isotope values or (e-h) FAME- $T_c$  modelled temperature recorded in the calcite shells ( $T_c$ ) occur. Each panel represents the Anderson-Darling test result, the results for *Globigerinoides sacculifer*, *Globigerinoides ruber* and *N. dutertrei* are overlaid. For all panels black areas reflect latitudinal and longitudinal grid points that failed to reject the null hypothesis ( $H_0$ ) and therefore the foraminiferal population (FP) of the El Niño is similar to the Non-El Niño, and therefore the distribution between the neutral climate and El Niño cannot be said to be different ( $FP_{El\ Niño} = FP_{Non-El\ Niño}$ ). The results in which the  $H_1$  hypothesis is accepted, in which the distributions can be said to be different ( $FP_{El\ Niño} \neq FP_{Non-El\ Niño}$ ), are plotted as yellow where the depth is deeper than 3500 m bsl or purple where the depth is shallower than 3500 m bsl (see Figure 8). Purple locations are where our results suggest that the signal of ENSO has different values and the water depth allows for preservation – although this purple region will be smaller when inferred SAR is taken into account (see Figure 9). The rows represent the model runs with a depth cut-off limit at: (A and E) 60 m; (B and F) 100 m; (C and G) 200 m; and (D and H) where a combination of depths were utilised (Pracht et al., 2019).

Formatted: Subscript



**Figure 8.** Bathymetric map of the Tropical Pacific Ocean highlighting the areas above and below the Lysocline and/or Calcite compensation depth (CCD). (A) GEBCO map of height relative to 0 m; (B) same as (A) with location of seamounts plotted (white stars); (C-E) binary colour map of GEBCO data, yellow is values below cut-off depth value ([C] 3500 m below sea-level (bsl); [D] 4000 m bsl; and [E] 4500 m bsl respectively) and purple above the cut-off depth value. The histograms represent the normalised frequency of grid cell height in bins of 250 m wide, yellow is values below cut off value ([C] 3500 m below sea-level (bsl); [D] 4000 m bsl; and [E] 4500 m bsl respectively), purple above cut off value. The grey bins in each histogram are those above 0 m.

5



Formatted: Not Highlight



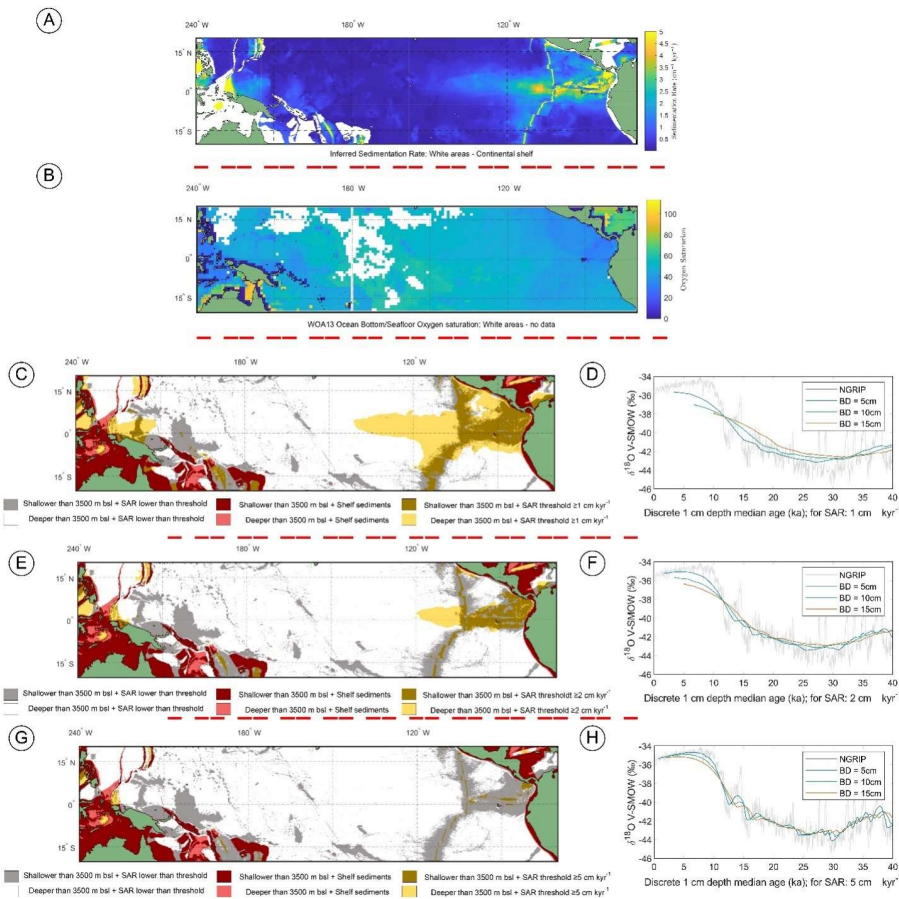
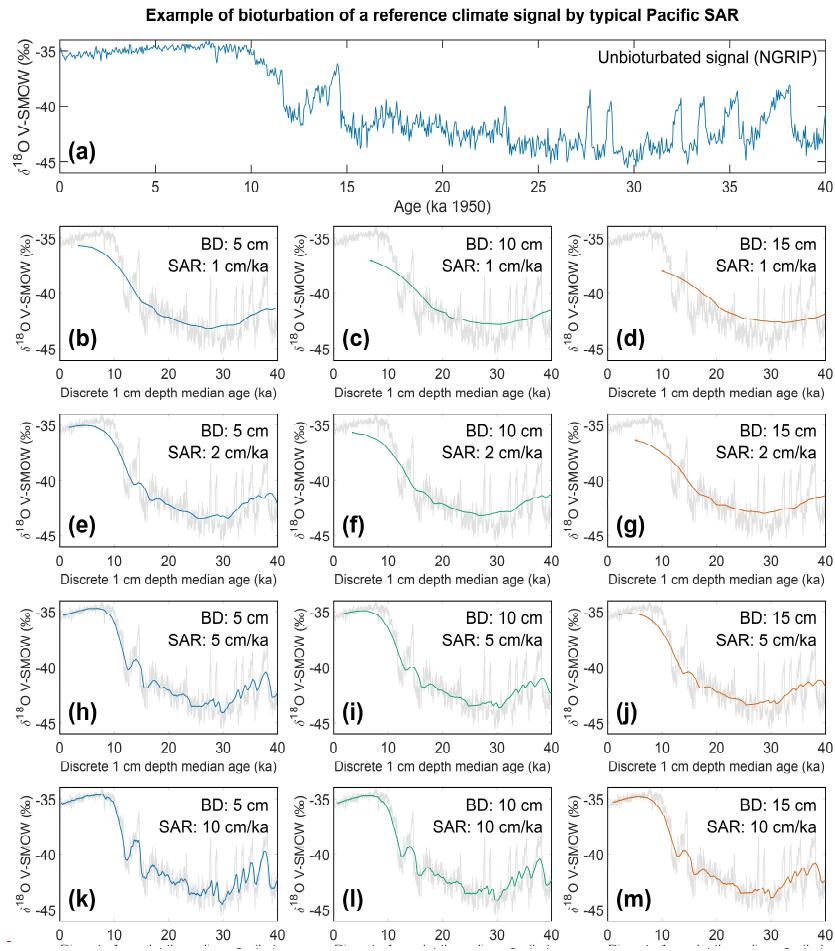


Figure 9. Map of the sedimentation rate and oxygen saturation for the Tropical Pacific. (A) Inferred sedimentation rate (Olsen et al 2016). White regions represent continental shelf. (B) Oxygen saturation of the bottom grid layer of World Ocean Atlas 2013 (data from: <https://www.nodc.noaa.gov/cgi-bin/OC5/woa13/woa13oxnu.pl>). (C, E, G) Overlay between water depth and inferred SAR. Red / Pink: Continental shelf sediments that are (Red) shallower or (Pink) deeper than 3500 mbsl; Gray / White: SAR lower than SAR threshold and the seafloor depth is (grey) shallower or (white) deeper than 3500 mbsl; Light Yellow/Gold: Light yellow represents areas where the SAR is above the threshold but the water depth is deeper than 3500 mbsl in comparison Gold represents areas where the SAR is above the threshold and the water depth is deeper than 3500 mbsl. The ideal locations are therefore plotted as Gold. Cut-off limits for SAR are (C)  $>1 \text{ cm kyr}^{-1}$ ; (E)  $>2 \text{ cm kyr}^{-1}$  and (G)  $>5 \text{ cm kyr}^{-1}$ . (D, F, H) bioturbation simulations for the minimum threshold for each SAR (see Figure 10 and Figure 11 for the output of SEAMUS). Each plot gives the

Formatted: Superscript  
Formatted: Superscript  
Formatted: Superscript

input values of NGRIP (grey) and for each SAR three analysis were performed with different bioturbation depths (BD) of: (Blue) 5 cm; (Green) 10 cm; and (Orange) 15 cm.



**Figure 10.** Output of the bioturbation model SEAMUS. (A) The unbioturbated input signal, NGRIP (North Greenland Ice Core Project Members, 2004; Rasmussen et al., 2014; Seierstad et al., 2014), used in our simulation of bioturbation for different SAR with SEAMUS (Lougheed, 2019). Sediment mixed layer referred to here as bioturbation depth (BD) is fixed at (B, E, H, K) 5 cm, (C, F, I, L) 10 cm and (D, G, J, M) 15 cm for sedimentation accumulation rates (SAR) of (B-D) 1 cm kyr<sup>-1</sup>; (E-G) 2 cm kyr<sup>-1</sup>; (H-J)

Formatted: Normal

Formatted: Superscript

5 cm kyr<sup>-1</sup> and (K-M) 10 cm kyr<sup>-1</sup>. The output is plotted as the discrete 1 cm depth median age. In (B-M) grey values represent the unbioturbated input signal, NGRIP. Note, we retain the original units (V-SMOW) of the original timeseries used, no inference between Pacific climate and Greenland is intended by the use of NGRIP (see section 2.7).

Simulated single specimen age distribution for typical Pacific SAR

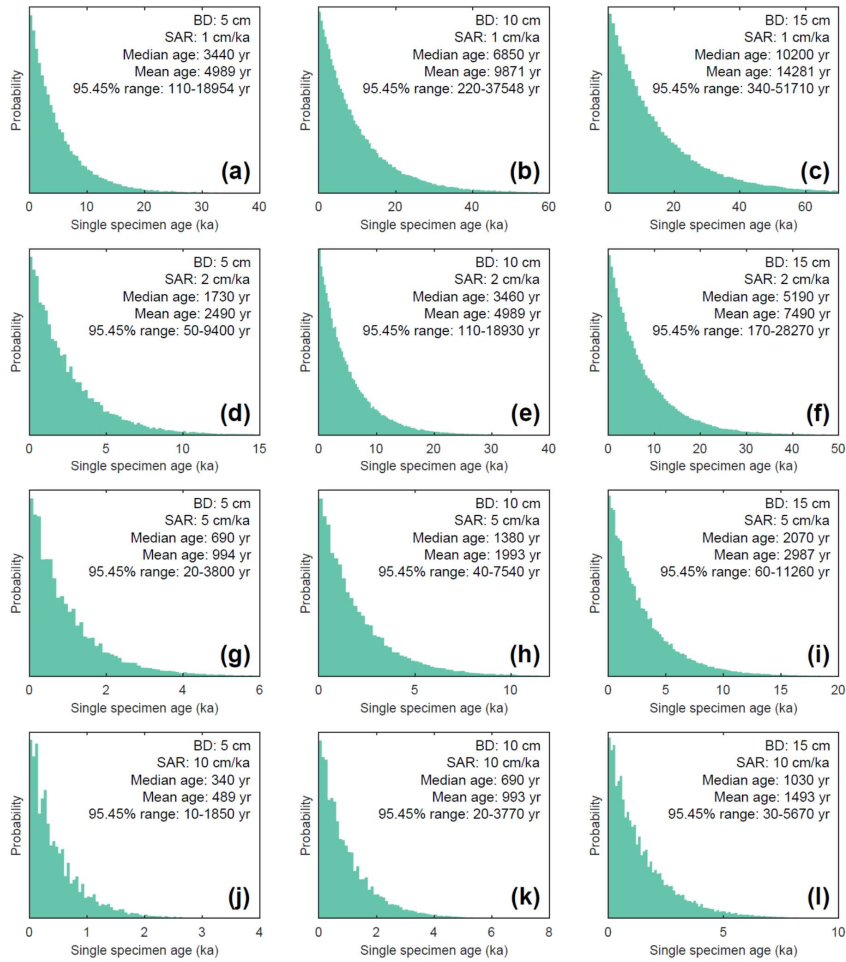


Figure 11. Histograms of simulated specimen age within the bioturbation depth. The simulated age distribution present within the sediment mixed layer, referred to here as bioturbation depth (BD). BD is fixed at (A, D, G, J) 5 cm, (B, E, H, K) 10 cm and (C, F, I, L) 15 cm for sedimentation accumulation rates (SAR) of (A-C) 1 cm kyr<sup>-1</sup>; (D-F) 2 cm kyr<sup>-1</sup>; (G-I) 5 cm kyr<sup>-1</sup> and (J-L) 10 cm

Formatted: Caption

**kvr!**. The output is plotted as the discrete 1 cm depth median age. Note the size of the BD varies, therefore the simulated age distribution comes from a varying 'core depth'.

Formatted: Font: 10 pt

## References

- 5 Allen, K. A., Hönisch, B., Eggins, S. M., Haynes, L. L., Rosenthal, Y. and Yu, J.: Trace element proxies for surface ocean conditions: A synthesis of culture calibrations with planktic foraminifera, *Geochimica et Cosmochimica Acta*, 193, 197–221, doi:10.1016/j.gca.2016.08.015, 2016.
- 10 An, S.-I. and Bong, H.: Inter-decadal change in El Niño-Southern Oscillation examined with Bjerknæs stability index analysis, *Climate Dynamics*, 47(3–4), 967–979, doi:10.1007/s00382-015-2883-8, 2016.
- An, S.-I. and Bong, H.: Feedback process responsible for the suppression of ENSO activity during the mid-Holocene, *Theoretical and Applied Climatology*, 132(3–4), 779–790, doi:10.1007/s00704-017-2117-6, 2018.
- 15 Anderson, L., Abbott, M. B., Finney, B. P. and Edwards, M. E.: Palaeohydrology of the Southwest Yukon Territory, Canada, based on multiproxy analyses of lake sediment cores from a depth transect, *The Holocene*, 15(8), 1172–1183, doi:10.1191/0959683605hl889rp, 2005.
- [Anderson, T. W., and Darling, D. A.: A test of goodness of fit. \*Journal of the American Statistical Association\*, 49, 765-769. http://dx.doi.org/10.2307/2281537, 1954](http://dx.doi.org/10.2307/2281537)
- 20 [Andreasen, D.H., Ravelo, A.C., Broccoli, A.J.: Remote forcing at the last glacial maximum in the tropical Pacific Ocean. \*Journal of Geophysical Research - Oceans\* 106, 879–897, 2001](http://dx.doi.org/10.1029/2001JC001879)
- [Arrhenius, G.: Pelagic Sediments, In: Sears M \(Ed\) \*Oceanography. Publication 67, American Association Advancement of Science, Washington, D.C.\*, https://doi.org/10.5962/bhl.title.34806, 1961](https://doi.org/10.5962/bhl.title.34806)
- 25 Asmerom, Y., Polyak, V., Burns, S. and Rasmussen, J.: Solar forcing of Holocene climate: New insights from a speleothem record, southwestern United States, *Geology*, 35(1), 1–4, doi:10.1130/G22865A.1, 2007.
- Balmaseda, M. A., Mogensen, K. and Weaver, A. T.: Evaluation of the ECMWF ocean reanalysis system ORAS4, *Quarterly Journal of the Royal Meteorological Society*, 139(674), 1132–1161, 2013.
- 30 Barker, S., Greaves, M. and Elderfield, H.: A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry, *Geochem. Geophys. Geosyst.*, 4(9), 8407, 2003.
- Barron, J. A. and Anderson, L.: Enhanced Late Holocene ENSO/PDO expression along the margins of the eastern North Pacific, *Quaternary International*, 235(1), 3–12, doi:https://doi.org/10.1016/j.quaint.2010.02.026, 2011.
- Barron, J. A., Heusser, L., Herbert, T. and Lyle, M.: High-resolution climatic evolution of coastal northern California during the past 16,000 years, *Paleoceanography*, 18(1), doi:10.1029/2002PA000768, 2003.
- 35 [Batiza, R.: Abundances, distribution and sizes of volcanoes in the Pacific Ocean and implications for the origin of non-hotspot volcanoes. \*Earth and Planetary Science Letters\*, 60, 195-206, 1982](https://doi.org/10.1016/j.epsl.2011.08.015)

Formatted: Normal

Formatted: Normal

Bé, A. W. H. and Spero, H. J.: Shell regeneration and biological recovery of planktonic foraminifera after physical injury induced in laboratory culture, *Micropaleontology*, 27(3), 305–316, 1981.

Bé, A. W. H., Spero, H. J. and Anderson, O. R.: Effects of symbiont elimination and reinfection on the life processes of the planktonic foraminifer *Globigerinoides sacculifer*, *Marine Biology*, 70, 73–86, 1982.

5 [Beaufort, L., Garidel-Thoron, T., Mix, A.C., Pisias, N.G.: ENSO like forcing on oceanic primary production during the late Pleistocene. \*Science\* 293, 2440–2444, 2001](#)

Formatted: Normal

10 Bemis, B. E., Spero, H. J., Bijma, J. and Lea, D. W.: Reevaluation of the oxygen isotopic composition of planktonic foraminifera: Experimental results and revised paleotemperature equations, *Paleoceanography*, 13(2), 150–160, doi:10.1029/98PA00070, 1998.

Bennett, J. R., Cumming, B. F., Leavitt, P. R., Chiu, M., Smol, J. P. and Szeicz, J.: Diatom, Pollen, and Chemical Evidence of Postglacial Climatic Change at Big Lake, South-Central British Columbia, Canada, *Quaternary Research*, 55(3), 332–343, doi:10.1006/qres.2001.2227, 2001.

15 Benson, L., Kashgarian, M., Rye, R., Lund, S., Paillet, F., Smoot, J., Kester, C., Mensing, S., Meko, D. and Lindström, S.: Holocene multidecadal and multicentennial droughts affecting Northern California and Nevada, *Quaternary Science Reviews*, 21(4), 659–682, doi:https://doi.org/10.1016/S0277-3791(01)00048-8, 2002.

Berger, W. H.: Foraminiferal Ooze: Solution at Depths, *Science*, New Series, 156(3773), 383–385, 1967.

20 [Berger W. H.: Planktonic Foraminifera: selective solution and paleoclimatic interpretation. \*Deep-Sea Research\*, 15, 31-43, 1968](#)

Formatted: Normal

Berger, W. H.: Planktonic Foraminifera: Differential Production and Expatriation Off Baja California, *Limnology and Oceanography*, 15(2), 183–204, 1970a.

Berger, W. H.: Planktonic Foraminifera: Selective solution and the lysocline, *Marine Geology*, 8(2), 111–138, doi:10.1016/0025-3227(70)90001-0, 1970b.

25 Berger, W. H.: Sedimentation of planktonic foraminifera, *Marine Geology*, 11(5), 325–358, doi:10.1016/0025-3227(71)90035-1, 1971.

Berger, W. H. and Gardner, J. V.: On the determination of Pleistocene temperatures, *Journal of Foraminiferal Research*, 5(2), 102–113, 1975.

30 [Berger, W.H., Adelseck, C., and Mayer, L.A.: Distribution of carbonate in surface sediments of the Pacific. \*Journal of Geophysical Research\*, 81 \(15\), 2617-2627, doi:10.1029/JC081i015p02617, 1976](#)

Formatted: Normal

Berger, W. H. and Heath, G. R.: Vertical mixing in pelagic sediments, *Journal of Marine Research*, 26(2), 134–143, 1968.

35 Bijma, J., Hönisch, B. and Zeebe, R. E.: Impact of the ocean carbonate chemistry on living foraminiferal shell weight: Comment on “Carbonate ion concentration in glacial-age deep waters of the Caribbean Sea” by W. S. Broecker and E. Clark: COMMENT, *Geochemistry, Geophysics, Geosystems*, 3(11), 1–7, doi:10.1029/2002GC000388, 2002.

Bird, C., Darling, K. F., Russell, A. D., Davis, C. V., Fehrenbacher, J., Free, A., Wyman, M. and Ngwenya, B. T.: Cyanobacterial endobionts within a major marine planktonic calcifier (*Globigerina bulloides*, Foraminifera) revealed by 16S rRNA metabarcoding, *Biogeosciences*, 14(4), 901–920, doi:10.5194/bg-14-901-2017, 2017.

5 Bird, C., Darling, K. F., Russell, A. D., Fehrenbacher, J. S., Davis, C. V., Free, A. and Ngwenya, B. T.: 16S rRNA gene metabarcoding and TEM reveals different ecological strategies within the genus *Neogloboquadrina* (planktonic foraminifer), edited by F. Frontalini, *PLOS ONE*, 13(1), e0191653, doi:10.1371/journal.pone.0191653, 2018.

[Blackman, A. and Somayajulu, B.L.K.: Pacific Pleistocene Cores: Faunal Analyses and Geochronology, \*Science\*, 154 \(3751\), 886-889, 1966](#)

Formatted: Normal

10 Boltovskoy, D.: The Sedimentary record of Pelagic Biogeography, *Progress in Oceanography*, 34(2–3), 135–160, 1994.

Boltovskoy, E.: 314. Depth at which foraminifera can survive in sediments, *Contributions from the Cushman Foundation for Foraminiferal Research*, XVII(2), 43–45, 1966.

[Boudreau, B. P. 1998: Mean mixed depth of sediments: The wherefore and the why, \*Limnology and Oceanography\*, 43 \(3\), 524-526, 1998.](#)

15 [Bramlette M.N.: Pelagic sediments. In: Sears M \(Ed\) \*Oceanography. Publication 67. American Association Advancement of Science\*, Washington, D.C., <https://doi.org/10.5962/bhl.title.34806>, 1961](#)

20 [Bramlette, M.N. and Bradley W.H.: \*Geology and Biology of North Atlantic Deep-Sea Cores between Newfoundland and Ireland: I. Lithology and Geologic Interpretation\*, U.S. Geological Survey Professional Paper, 196-A, 1-34, <https://pubs.usgs.gov/pp/0196/report.pdf>, 1941](#)

Formatted: Normal

25 [Brand, W.A., Coplen, T.B., Vogl, J., Rosner, M., and Prohaska, T.: Assessment of international reference materials for isotope-ratio analysis \(IUPAC Technical Report\), \*Pure and Applied Chemistry\*, 86\(3\), 425-467, doi:10.1515/pac-2013-1023, 2014](#)

Branson, O., Redfern, S. A. T., Tyliczszak, T., Sadekov, A., Langer, G., Kimoto, K. and Elderfield, H.: The coordination of Mg in foraminiferal calcite, *Earth and Planetary Science Letters*, 383, 134–141, doi:10.1016/j.epsl.2013.09.037, 2013.

30 Branson, O., Bonnin, E. A., Perea, D. E., Spero, H. J., Zhu, Z., Winters, M., Hönisch, B., Russell, A. D., Fehrenbacher, J. S. and Gagnon, A. C.: Nanometer-Scale Chemistry of a Calcite Biomineralization Template: Implications for Skeletal Composition and Nucleation, *Proceedings of the National Academy of Sciences*, 113(46), 12934–12939, doi:10.1073/pnas.1522864113, 2016.

Caley, T., Roche, D. M., Waelbroeck, C. and Michel, E.: Oxygen stable isotopes during the Last Glacial Maximum climate: perspectives from data–model (iLOVECLIM) comparison, *Climate of the Past*, 10(6), 1939–1955, doi:10.5194/cp-10-1939-2014, 2014.

35 Caramanica, A., Quilter, J., Huaman, L., Villanueva, F. and Morales, C. R.: Micro-remains, ENSO, and environmental reconstruction of El Paraiso, Peru, a late preceramic site, *Journal of Archaeological Science: Reports*, 17, 667–677, doi:<https://doi.org/10.1016/j.jasrep.2017.11.026>, 2018.

40 Chen, S., Hoffmann, S. S., Lund, D. C., Cobb, K. M., Emile-Geay, J. and Adkins, J. F.: A high-resolution speleothem record of western equatorial Pacific rainfall: Implications for Holocene ENSO evolution, *Earth and Planetary Science Letters*, 442, 61–71, doi:<https://doi.org/10.1016/j.epsl.2016.02.050>, 2016.

CLIMAP Project Members: The Surface of the Ice-Age Earth, *Science*, 191(4232), 1131–1137, 1976.

[Clement, A.C., Seager, R. and Cane, M.A.: Orbital controls on the El Niño/Southern Oscillation and the tropical climate, \*Paleoceanography\*, 14 \(4\), 441–456, 1999](#)

5 [Clouard, V. and Bonneville, A.: Ages of seamounts, islands, and plateaus on the Pacific plate, \*Geological Society of America Special Paper\* 388, 71–90, doi: 10.1130/2005.2388\(06\), 2005](#)

Cole, J. and Tudhope, A. W.: Reef-Based Reconstructions of Eastern Pacific Climate Variability, in *Coral Reefs of the Eastern Tropical Pacific: Persistence and Loss in a Dynamic Environment*, edited by P. W. Glynn, D. P. Manziello, and I. C. Enochs, pp. 535–548, Springer Netherlands, Dordrecht., 2017.

10

Conroy, J. L., Overpeck, J. T., Cole, J. E., Shanahan, T. M. and Steinitz-Kannan, M.: Holocene changes in eastern tropical Pacific climate inferred from a Galápagos lake sediment record, *Quaternary Science Reviews*, 27(11), 1166–1180, doi:<https://doi.org/10.1016/j.quascirev.2008.02.015>, 2008.

15

Darling, K., Wade, C. M., Steward, I. A., Kroon, D., Dingle, R. and Leigh Brown, A. J.: Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers, *Nature*, 405, 43–47, 2000.

Darling, K. F., Wade, C. M., Kroon, D., Brown, A. J. L. and Bijma, J.: The Diversity and Distribution of Modern Planktic Foraminiferal Small Subunit Ribosomal RNA Genotypes and their Potential as Tracers of Present and Past Ocean Circulations, *Paleoceanography*, 14(1), 3–12, doi:10.1029/1998PA900002, 1999.

20

Darling, K. F., Kucera, M., Pudsey, C. J. and Wade, C. M.: Molecular evidence links cryptic diversification in polar planktonic protists to Quaternary climate dynamics, *Proceedings of the National Academy of Sciences*, 101(20), 7657–7662, doi:10.1073/pnas.0402401101, 2004.

Dolman, A. M. and Laepple, T.: Sedproxy: a forward model for sediment archived climate proxies, *Climate of the Past Discussions*, 1–31, doi:10.5194/cp-2018-13, 2018.

25

Du, X., Hendy, I. and Schimmelmann, A.: A 9000-year flood history for Southern California: A revised stratigraphy of varved sediments in Santa Barbara Basin, *Marine Geology*, 397, 29–42, doi:<https://doi.org/10.1016/j.margeo.2017.11.014>, 2018.

Dubois, N., Kienast, M., Normandeau, C. and Herbert, T. D.: Eastern equatorial Pacific cold tongue during the Last Glacial Maximum as seen from alkenone paleothermometry, *Paleoceanography*, 24(4), doi:10.1029/2009PA001781, 2009.

30

Eggins, S. M., Sadekov, A. and de Deckker, P.: Modulation and daily banding of Mg/Ca in *Orbulina universa* tests by symbiont photosynthesis and respiration: a complication for seawater thermometry, *Earth and Planetary Science Letters*, 225, 411–419, 2003.

Elderfield, H. and Ganssen, G. M.: Past temperature and  $\delta^{18}\text{O}$  of surface ocean waters inferred from foraminiferal Mg/Ca ratios, *Nature*, 405(6785), 442–445, 2000.

35

Emiliani, C.: Depth habitats of some species of pelagic foraminifera as indicated by oxygen isotope ratios, *American Journal of Science*, 252(3), 149–158, doi:10.2475/ajs.252.3.149, 1954.

Emiliani, C.: Pleistocene Temperatures, *The Journal of Geology*, 63(6), 538–578, 1955.

Formatted: Normal



- Enzel, Y. and Wells, S. G.: Extracting Holocene paleohydrology and paleoclimatology information from modern extreme flood events: An example from southern California, *Geomorphology*, 19(3), 203–226, doi:https://doi.org/10.1016/S0169-555X(97)00015-9, 1997.
- Ericson, D. B. and Wollin, G.: Pleistocene climates and chronology in deep-sea sediments, *Science*, 162, 1227–1234, 1968.
- 5 Ericson, D. B., Ewing, M. and Wollin, G.: The Pleistocene Epoch in Deep-Sea Sediments, *Science, New Series*, 146(3645), 723–732, 1964.
- Evans, D., Müller, W. and Erez, J.: Assessing foraminifera biomineralisation models through trace element data of cultures under variable seawater chemistry, *Geochimica et Cosmochimica Acta*, 236, 198–217, doi:10.1016/j.gca.2018.02.048, 2018.
- 10 [Evans, M. N., Kaplan, A. and Cane, M. A.: Optimal sites for coral-based reconstructions of global sea surface temperature, \*Paleoceanography\*, 13 \(5\), 502-516, 1998](#)
- [Ewing, M., Heezen, B.C., and Ericson, D.: Significance of the Worzel deep-sea ash, \*PNAS\*, 45, 355, 1959](#)
- Fallet, U., Boer, W., van Assen, C., Greaves, M. and Brummer, G. J. A.: A novel application of wet-oxidation to retrieve carbonates from large, organic-rich samples for application in climate research, *Geochemistry Geophysics Geosystems*, 10(8), 2009.
- 15 [Feldberg, M.J., Mix, A.C.: Planktonic foraminifera, sea surface temperatures, and mechanisms of oceanic change in the Peru and south equatorial currents, 0–150 ka BP. \*Paleoceanography\* 18, 1–16, 2003](#)
- 20 Feldmeijer, W., Metcalfe, B., Scussolini, P. and Arthur, K.: The effect of chemical pretreatment of sediment upon foraminiferal-based proxies, *Geochemistry, Geophysics, Geosystems*, 14(10), 3996–4014, doi:10.1002/ggge.20233, 2013.
- Feldmeijer, W., Metcalfe, B., Brummer, G. J. A. and Ganssen, G. M.: Reconstructing the depth of the permanent thermocline through the morphology and geochemistry of the deep dwelling planktonic foraminifer *Globorotalia truncatulinoides*, *Paleoceanography*, 30(1), 1–22, doi:10.1002/2014PA002687, 2015.
- 25 [Filliben, J.J.: The Probability Plot Correlation Coefficient Test for Normality, \*Technometrics\*, 17 \(1\), 111-117, 1975](#)
- Ford, H. L., Ravelo, A. C. and Polissar, P. J.: Reduced El Nino-Southern Oscillation during the Last Glacial Maximum, *Science*, 347(6219), 255–258, doi:10.1126/science.1258437, 2015.
- 30 Fraile, I., Schulz, M., Mulitza, S. and Kucera, M.: Predicting the global distribution of planktonic foraminifera using a dynamic ecosystem model, *Biogeosciences*, 5, 891–911, 2008.
- Fraile, I., Schulz, M., Mulitza, S., Merkel, U., Prange, M. and Paul, A.: Modeling the seasonal distribution of planktonic foraminifera during the Last Glacial Maximum, *Paleoceanography*, 24, doi:10.1029/2008PA001686, 2009.
- 35 Ganssen, G. M., Peeters, F. J. C., Metcalfe, B., Anand, P., Jung, S. J. A., Kroon, D. and Brummer, G.-J. A.: Quantifying sea surface temperature ranges of the Arabian Sea for the past 20 000 years, *Climate of the Past*, 7(4), 1337–1349, doi:10.5194/cp-7-1337-2011, 2011.
- Garidel-Thoron, T. de, Rosenthal, Y., Beaufort, L., Bard, E., Sonzogni, C. and Mix, A. C.: A multiproxy assessment of the western equatorial Pacific hydrography during the last 30 kyr, *Paleoceanography*, 22(3), doi:10.1029/2006PA001269, 2007.

Formatted: Normal

Formatted: Normal

Formatted: Font: Italic

Formatted: Normal

- Gray, W. R., Weldeab, S., Lea, D. W., Rosenthal, Y., Gruber, N., Donner, B. and Fischer, G.: The effects of temperature, salinity, and the carbonate system on Mg/Ca in Globigerinoides ruber (white): A global sediment trap calibration, *Earth and Planetary Science Letters*, 482, 607–620, doi:10.1016/j.epsl.2017.11.026, 2018.
- 5 Greaves, M., Barker, S., Daunt, C. and Elderfield, H.: Accuracy, standardization, and interlaboratory calibration standards for foraminiferal Mg/Ca thermometry, *Geochem. Geophys. Geosyst.*, 6(2), Q02D13, 2005.
- Groeneveld, J., Nürnberg, D., Tiedemann, R., Reichert, G.-J., Steph, S., Reuning, L., Crudeli, D. and Mason, P.: Foraminiferal Mg/Ca increase in the Caribbean during the Pliocene: Western Atlantic Warm Pool formation, salinity influence, or diagenetic overprint?, *Geochemistry Geophysics Geosystems*, 9, Q01P23, doi:doi:10.1029/2006GC001564, 2008.
- 10 Hamilton, C. 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(3–4), 256–267, 2008.
- [Hays, J.D., Saito, T., Opdyke, N., and Burckle, L.H.: Pliocene-Pleistocene Sediments of the Equatorial Pacific: Their Paleomagnetic, Biostratigraphic, and Climatic Record, \*Geological Society of America Bulletin\*, 80, 1481-1514, 1969](#)
- 15 Hecht, A. D.: A model for determining Pleistocene paleotemperatures from planktonic foraminiferal assemblages, *Micropaleontology*, 19, 68–77, 1973.
- Hendy, I. L., Napier, T. J. and Schimmelmann, A.: From extreme rainfall to drought: 250 years of annually resolved sediment deposition in Santa Barbara Basin, California, *Quaternary International*, 387, 3–12, doi:https://doi.org/10.1016/j.quaint.2015.01.026, 2015.
- 20 Higley, M. C., Conroy, J. L. and Schmitt, S.: Last Millennium Meridional Shifts in Hydroclimate in the Central Tropical Pacific, *Paleoceanography and Paleoclimatology*, 33(4), 354–366, doi:10.1002/2017PA003233, 2018.
- [Hillier, J. K.: Pacific seamount volcanism in space and time, \*Geophys. J. Int.\*, 168, 877-889, doi: 10.1111/j.1365-246X.2006.03250.x, 2007](#)
- 25 Hori, M., Shirai, K., Kimoto, K., Kurasawa, A., Takagi, H., Ishida, A., Takahata, N. and Sano, Y.: Chamber formation and trace element distribution in the calcite walls of laboratory cultured planktonic foraminifera (*Globigerina bulloides* and *Globigerinoides ruber*), *Marine Micropaleontology*, 140, 46–55, doi:10.1016/j.marmicro.2017.12.004, 2018.
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S. and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons, *Journal of Climate*, 30(20), 8179–8205, doi:10.1175/jcli-d-16-0836.1, 2017.
- 30 Huber, B. T., Bijma, J. and Darling, K.: Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny), *Paleobiology*, 23(01), 33–62, doi:10.1017/S0094837300016638, 1997.
- [Huber, M. and Caballero, R.: Eocene El Niño: Evidence for Robust Tropical Dynamics in the “Hothouse”, \*Science\*, 299 \(5608\), 877-881, 2003](#)
- 35 Hut, G.: Consultants group meeting on stable isotope reference samples for geochemical and hydrological investigations, Int. At. Energy Agency, Vienna., 1987.

Formatted: Normal

Formatted: Normal

Formatted: Normal

- Hutson, W. H.: Transfer functions under no-analog conditions: experiments with Indian Ocean planktonic foraminifera, *Quaternary Research*, 8(3), 355–367, 1977.
- Hutson, W. H.: Applications of transfer functions to Indian Ocean planktonic foraminifera, *Quaternary Research*, 9, 87–112, 1978.
- 5 Hutson, W. H.: Bioturbation of deep-sea sediments: Oxygen isotopes and stratigraphic uncertainty, *Geology*, 8(3), 127–130, doi:10.1130/0091-7613(1980), 1980a.
- Hutson, W. H.: The Agulhas Current during the Late Pleistocene-analysis of modern faunal analogs, *Science*, 207(4426), 64–66, 1980b.
- 10 Imbrie, J. and Kipp, N. G.: A new Micropaleontological method for paleoclimatology: Application to a Late Pleistocene Caribbean core, in *The Late Cenozoic Glacial Ages*, edited by K. K. Turekian, pp. 71–181, Yale University Press, New Haven, Connecticut., 1971.
- Jacob, D. E., Wirth, R., Agbaje, O. B. A., Branson, O. and Eggins, S. M.: Planktic foraminifera form their shells via metastable carbonate phases, *Nature Communications*, 8(1), 1265, doi:10.1038/s41467-017-00955-0, 2017.
- Jones, J. I.: Distribution of living planktonic foraminifera of the West Indies and adjacent waters, PhD Thesis, University Wisconsin-Madison, Madison, Wisconsin USA., 1964.
- 15 Jonkers, L. and Kučera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, *Climate of the Past*, 13(6), 573–586, doi:10.5194/cp-13-573-2017, 2017.
- Kageyama, M., Braconnot, P., Bopp, L., Caubel, A., Foujols, M.-A., Guilyardi, E., Khodri, M., Lloyd, J., Lombard, F., Mariotti, V., Marti, O., Roy, T. and Woillez, M.-N.: Mid-Holocene and Last Glacial Maximum climate simulations with the IPSL model—part I: comparing IPSL\_CM5A to IPSL\_CM4, *Climate Dynamics*, 40(9), 2447–2468, doi:10.1007/s00382-012-1488-8, 2013.
- 20 [Keigwin, L. D. and Guilderson, T. P.: Bioturbation artifacts in zero-age sediments, \*Paleoceanography\*, 24, PA4212, doi: 10.1029/2008PA001727, 2009](#)
- 25 Kienast, M., MacIntyre, G., Dubois, N., Higginson, S., Normandeau, C., Chazen, C. and Herbert, T. D.: Alkenone unsaturation in surface sediments from the eastern equatorial Pacific: Implications for SST reconstructions, *Paleoceanography*, 27(1), doi:10.1029/2011PA002254, 2012.
- Kim, S.-T. and O'Neil, J. R.: Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates, *Geochimica et Cosmochimica Acta*, 61(16), 3461–3475, doi:10.1016/S0016-7037(97)00169-5, 1997.
- 30 Kisakürek, B., Eisenhauer, A., Böhm, F., Garbe-Schönberg, D. and Erez, J.: Controls on shell Mg/Ca and Sr/Ca in cultured planktonic foraminiferan, *Globigerinoides ruber* (white), *Earth and Planetary Science Letters*, 273(3–4), 260–269, doi:10.1016/j.epsl.2008.06.026, 2008.
- 35 [Koppers A.A.P. Staudigel H. Pringle M.S. Wijbrans J.R.: Short-lived and discontinuous intraplate volcanism in the South Pacific: hot spots or extensional volcanism, \*Geochem. Geophys. Geosys.\*, 4 \(1089\), 2003](#)
- Koutavas, A. and Joanides, S.: El Niño–Southern Oscillation extrema in the Holocene and Last Glacial Maximum, *Paleoceanography*, 27(4), doi:10.1029/2012PA002378, 2012.

Formatted: Normal

Formatted: Normal

- Koutavas, A. and Lynch-Stieglitz, J.: Glacial-interglacial dynamics of the eastern equatorial Pacific cold tongue-Intertropical Convergence Zone system reconstructed from oxygen isotope records: GLACIAL PACIFIC COLD TONGUE, *Paleoceanography*, 18(4), n/a-n/a, doi:10.1029/2003PA000894, 2003.
- 5 Koutavas, A., Lynch-Stieglitz, J., Marchitto, T. M. and Sachs, J. P.: El Niño-Like Pattern in Ice Age Tropical Pacific Sea Surface Temperature, *Science*, 297(5579), 226, doi:10.1126/science.1072376, 2002.
- Koutavas, A., deMenocal, P. B., Olive, G. C. and Lynch-Stieglitz, J.: Mid-Holocene El Niño–Southern Oscillation (ENSO) attenuation revealed by individual foraminifera in eastern tropical Pacific sediments, *Geology*, 34(12), 993, doi:10.1130/G22810A.1, 2006.
- 10 Kretschmer, K., Kucera, M. and Schulz, M.: Modeling the distribution and seasonality of *Neogloboquadrina pachyderma* in the North Atlantic Ocean during Heinrich Stadial 1, *Paleoceanography*, 31(7), 986–1010, doi:10.1002/2015PA002819, 2016.
- Kretschmer, K., Jonkers, L., Kucera, M. and Schulz, M.: Modeling seasonal and vertical habitats of planktonic foraminifera on a global scale, *Biogeosciences Discussions*, 1–37, doi:10.5194/bg-2017-429, 2017.
- Leduc, G., Vidal, L., Cartapanis, O. and Bard, E.: Modes of eastern equatorial Pacific thermocline variability: Implications for ENSO dynamics over the last glacial period: ENSO DYNAMICS OVER THE LAST 50 KA, *Paleoceanography*, 24(3), doi:10.1029/2008PA001701, 2009.
- 15 LeGrande, A. N. and Schmidt, G. A.: Global gridded data set of the oxygen isotopic composition in seawater, *Geophysical Research Letters*, 33(12), L12604, 2006.
- Lombard, F., Labeyrie, L., Michel, E., Spero, H. J. and Lea, D. W.: Modelling the temperature dependent growth rates of planktic foraminifera, *Marine Micropalaeontology*, 70, 1–7, 2009.
- 20 Lombard, F., da Rocha, R. E., Bijma, J. and Gattuso, J. P.: Effect of carbonate ion concentration and irradiance on calcification in planktonic foraminifera, *Biogeosciences*, 7(1), 247–255, 2010.
- Lombard, F., Labeyrie, L., Michel, E., Bopp, L., Cortijo, E., Retailliau, S., Howa, H. and Jorissen, F.: Modelling planktic foraminifer growth and distribution using an ecophysiological multi-species approach, *Biogeosciences*, 8(4), 853–873, doi:10.5194/bg-8-853-2011, 2011.
- 25 Loubere, P., Creamer, W. and Haas, J.: Evolution of the El Niño-Southern Oscillation in the late Holocene and insolation driven change in the tropical annual SST cycle, *Global and Planetary Change*, 100, 129–144, doi:https://doi.org/10.1016/j.gloplacha.2012.10.007, 2013.
- [Lougheed, B. C.: SEAMUS \(v1.0\): a  \$\Delta^{14}\text{C}\$ -enabled, single-specimen sediment accumulation simulator. \*Geoscientific Model Development\*, 2019.](#)
- 30 Lougheed, B. C., Metcalfe, B., Ninnemann, U. S. and Wacker, L.: Moving beyond the age depth model paradigm in deep-sea palaeoclimate archives: dual radiocarbon and stable isotope analysis on single foraminifera, *Climate of the Past*, 14(4), 515–526, 2018.
- Löwemark, L.: Importance and Usefulness of Trace fossils and Bioturbation in Paleocyanography, in *Trace Fossils: Concepts, Problems, Prospects*, edited by W. Miller, pp. 413–427, Elsevier, Amsterdam., 2007.

Löwemark, L. and Grootes, P. M.: Large age differences between planktic foraminifers caused by abundance variations and Zoophycos bioturbation, *Paleoceanography*, 19, PA2001, doi:10.1029/2003PA000949, 2004.

Löwemark, L., Hong, W.-L., Yui, T.-F. and Hung, G.-W.: A test of different factors influencing the isotopic signal of planktonic foraminifera in surface sediments from the northern South China Sea, *Marine Micropaleontology*, 55(1–2), 49–62, doi:10.1016/j.marmicro.2005.02.004, 2005.

Löwemark, L., Konstantinou, K. I. and Steinke, S.: Bias in foraminiferal multispecies reconstructions of paleohydrographic conditions caused by foraminiferal abundance variations and bioturbational mixing: A model approach, *Marine Geology*, 256, 101–106, 2008.

Luz, B.: Stratigraphic and Paleoclimatic Analysis of late Pleistocene Tropical Southeast Pacific Cores, *Quaternary Research*, 3, 56-72, 1973

Lynts, G. W. and Judd, J. B.: Late Pleistocene paleotemperatures at Tongue of the Ocean, Bahamas, *Science*, 171, 1143–1144, 1971.

MARGO Project Members\*: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nature Geoscience*, 2(2), 127–132, doi:10.1038/ngeo411, 2009.

McIntyre, A., Kipp, N. G., Bé, A. W. H., Crowley, T., Kellogg, T., Gardner, J. V., Prell, W. L. and Ruddiman, W. F.: Glacial North Atlantic 18,000 years ago: A CLIMAP reconstruction, *Geological Society of America Memoirs*, 145, 43–76, 1976.

McPhaden, M.J., Zebiak, S.E., and Glantz, M.H.: ENSO as an Integrating Concept in Earth Science, *Science*, 314 (5806), 1740-1745, 2006

Menard H.W.: *Marine Geology of the Pacific*, McGraw-Hill, New York, 1964

Metcalfé, B., Feldmeijer, W., de Vringer-Picon, M., Brummer, G.-J. A., Peeters, F. J. C. and Ganssen, G. M.: Late Pleistocene glacial–interglacial shell-size–isotope variability in planktonic foraminifera as a function of local hydrography, *Biogeosciences*, 12(15), 4781–4807, doi:10.5194/bg-12-4781-2015, 2015.

Metcalfé, B., Lougheed, B.C., Waelbroeck, C., and Roche D.M.: DATA: Data set, Zenodo, <https://doi.org/10.5281/zenodo.2554843>, 2019

Metcalfé, B., Feldmeijer, W., and Ganssen, G.M.: Oxygen isotope variability of planktonic foraminifera provide clues to past upper ocean seasonal variability, *Paleoceanography and Paleoclimatology*, doi: 10.1029/2018PA003475, 2019

Mikis, A., Hendry, K. R., Pike, J., Schmidt, D. N., Edgar, K. M., Peck, V., Peeters, F. J. C., Leng, M. J., Meredith, M. P., Todd, C. L., Stammerjohn, S., and Ducklow, H.: Temporal variability in foraminiferal morphology and geochemistry at the West Antarctic Peninsula: a sediment trap study, *Biogeosciences Discuss.*, <https://doi.org/10.5194/bg-2019-19>, in review, 2019.

Mix, A. C.: The oxygen-isotope record of deglaciation, in: *North America and adjacent oceans during the last deglaciation*, in *The Geology of America*, vol. K-3, edited by W. F. Ruddiman and H. E. J. Wright, pp. 111–135, Geological Society of America, Boulder, Colorado., 1987.

Mix, A.C.: Running hot and cold in the eastern equatorial Pacific, *Quaternary Science Reviews*, 25, 1147-1149, doi:10.1016/j.quascirev.2006.03.008, 2006

Formatted: Normal

Formatted: Normal

Formatted: Normal

- Morard, R., Quillévéré, F., Escarguel, G., de Garidel-Thoron, T., de Vargas, C. and Kucera, M.: Ecological modeling of the temperature dependence of cryptic species of planktonic Foraminifera in the Southern Hemisphere, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 391, 13–33, doi:10.1016/j.palaeo.2013.05.011, 2013.
- 5 Morard, R., Reinelt, M., Chiessi, C. M., Groeneveld, J. and Kucera, M.: Tracing shifts of oceanic fronts using the cryptic diversity of the planktonic foraminifera *Globorotalia inflata*, *Paleoceanography*, 31(9), 1193–1205, doi:10.1002/2016PA002977, 2016.
- Moy, C. M., Seltzer, G. O., Rodbell, D. T. and Anderson, D. M.: Variability of El Niño/Southern Oscillation activity at millennial timescales during the Holocene epoch, *Nature*, 420, 162, doi:10.1038/nature01194, 2002.
- 10 Mulitza, S., Wolff, T., Pätzold, J., Hale, W. and Wefer, G.: Temperature sensitivity of planktic foraminifera and its influence on the oxygen isotope record, *Marine Micropaleontology*, 33(3–4), 223–240, doi:10.1016/S0377-8398(97)00040-6, 1998.
- [Murray, J.; Renard, AF: Deep-sea deposits \(based on the specimens collected during the voyage of HMS Challenger in the years 1872 to 1876\). Report on the scientific results of the voyage of H.M.S. Challenger during the years 1873-76; John Menzies and Co., Edinburgh, United Kingdom, 1891](#)
- 15 Nederbragt, A. J. and Thurow, J. W.: A 6000yr varve record of Holocene climate in Saanich Inlet, British Columbia, from digital sediment colour analysis of ODP Leg 169S cores, *Marine Geology*, 174(1), 95–110, doi:https://doi.org/10.1016/S0025-3227(00)00144-4, 2001.
- Nederbragt, A. J. and Thurow, J. W.: Amplitude of ENSO cycles in the Santa Barbara basin, off California, during the past 15,000 years, *Journal of Quaternary Science*, 20, 447–456, 2006.
- 20 [North Greenland Ice Core Project Members: High-resolution record of Northern Hemisphere climate extending into the Last Interglacial period, \*Nature\*, 431, 147-151, 2004](#)
- de Nooijer, L. J., Toyofuku, T., Oguri, K., Nomaki, H. and Kitazato, H.: Intracellular pH distribution in foraminifera determined by the fluorescent probe HPTS: Intracellular pH of foraminifera, *Limnology and Oceanography: Methods*, 6(11), 610–618, doi:10.4319/lom.2008.6.610, 2008.
- 25 de Nooijer, L. J., Toyofuku, T. and Kitazato, H.: Foraminifera promote calcification by elevating their intracellular pH, *Proceedings of the National Academy of Sciences*, 106(36), 15374–15378, doi:10.1073/pnas.0904306106, 2009.
- Nürnberg, D., Bijma, J. and Hemleben, C.: Assessing the reliability of magnesium in foraminiferal calcite as a proxy for water mass temperatures, *Geochimica et Cosmochimica Acta*, 60(5), 803–814, 1996.
- 30 Olson, P., Reynolds, E., Hinnov, L. and Goswami, A.: Variation of ocean sediment thickness with crustal age, *Geochem. Geophys. Geosy.*, 17, 1349–1369, 2016.
- O’Neil, J. R., Clayton, R. N. and Mayeda, T. K.: Oxygen Isotope Fractionation in Divalent Metal Carbonates, *The Journal of Chemical Physics*, 51(12), 5547–5558, 1969.
- Parker, F. L.: Eastern Mediterranean foraminifera, *Rep. Swedish Deep Sea Exped.*, 8(4), 219–283, 1958.
- 35 [Parker, F.L. and Berger, W. H.: Faunal and solution patterns of planktonic Foraminifera in surface sediments of the South Pacific, \*Deep-Sea Research\*, 18, 73-107, 1971](#)

Formatted: Normal

Formatted: Normal

Formatted: Normal

- Patterson, D. T., Prokoph, A., Wright, C., Chang, A. S., Thomson, R. E. and Ware, D. M.: Holocene Solar Variability and Pelagic Fish Productivity in the NE Pacific., *Palaeontologia Electronica*, 7(4), 17, 2004.
- Peeters, F. J. C., Acheson, R., Brummer, G. J. A., de Ruijter, W. P. M., Schneider, R., Ganssen, G. M., Ufkes, E. and D., K.: Vigorous exchange between the Indian and Atlantic oceans at the end of the past five glacial periods, *Nature*, 430, 661–665, 2004.
- Pena, L. D., Cacho, I., Ferretti, P. and Hall, M. A.: El Niño–Southern Oscillation–like variability during glacial terminations and interlatitudinal teleconnections, *Paleoceanography*, 23(3), doi:10.1029/2008PA001620, 2008.
- Peng, T.-H., Broecker, W. S. and Berger, W. H.: Rates of benthic mixing in deep-sea sediment as determined by radioactive tracers, *Quaternary Res.*, 11, 141–149, 1979.
- 10 Phleger, F. B., Parker, F. L. and Pierson, J. F.: North Atlantic Foraminifera, Swedish Deep-Sea Expedition 1947-1948, 7, 3–9, 1953.
- Pisias, N. G. and Mix, A. C.: Aliasing of the geologic record and the search for long-period Milankovitch cycles, *Paleoceanography*, 3(5), 613–619, doi:10.1029/PA003i005p00613, 1988.
- 15 [Pisias, N.G., Mix, A.C., 1997. Spatial and temporal oceanographic variability of the eastern equatorial Pacific during the late Pleistocene: evidence from Radiolaria microfossils. \*Paleoceanography\* 12, 381–393.](#)
- Pracht, H., Metcalfe, B. and Peeters, F. J. C.: Oxygen isotope composition of final chamber of planktic foraminifera provides evidence for vertical migration and depth integrated growth, *Biogeosciences Discussions*, 1–32, doi:10.5194/bg-2018-146, 2018.
- 20 [Rasmussen, S.O., M. Bigler, S. Blockley, T. Blunier, B. Buchardt, H. Clausen, I. Cvijanovic, D. Dahl-Jensen, S. Johnsen, H. Fischer, V. Gkinis, M. Guillevic, W. Hoek, J. Lowe, J. Pedro, T. Popp, I. Seierstad, J. Steffensen, A. Svensson, P. Vallelonga, B. Vinther, M. Walker, J.J. Wheatley, M. Winstrup: A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. \*Quaternary Science Reviews\*, 106, 14-27, 2014](#)
- 25 [Raven, J., Caldeira, K., Elderfield, H., Hoegh-Guldberg, O., Liss, P., Riebesell, U., Shepherd, J., Turley, C. and Watson, A.: Ocean acidification due to increasing atmospheric carbon dioxide., The Royal Society., 2005.](#)
- Reynolds, R. W. and Smith, T. M.: A high-resolution global sea surface temperature climatology, *Journal of Climate*, 8, 1571–1583, 1995.
- 30 Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. and Wang, W.: An Improved In Situ and Satellite SST Analysis for Climate, *Journal of Climate*, 15(13), 1609–1625, doi:10.1175/1520-0442(2002)015<1609:aiisas>2.0.co;2, 2002.
- Rincón-Martínez, D., Steph, S., Lamy, F., Mix, A. and Tiedemann, R.: Tracking the equatorial front in the eastern equatorial Pacific Ocean by the isotopic and faunal composition of planktonic foraminifera, *Marine Micropaleontology*, 79(1), 24–40, doi:10.1016/j.marmicro.2011.01.001, 2011.
- 35 Rink, S., Kühl, M., Bijma, J. and Spero, H. J.: Microsensor studies of photosynthesis and respiration in the symbiotic foraminifer *Orbulina universa*, *Marine Biology*, 131, 583–595, 1998.

Formatted: Normal

Formatted: Normal

Roche, D. M., Paillard, D., Caley, T. and Waelbroeck, C.: LGM hosing approach to Heinrich Event 1: results and perspectives from data–model integration using water isotopes, *Quaternary Science Reviews*, 106, 247–261, doi:10.1016/j.quascirev.2014.07.020, 2014.

5 Roche, D. M., Waelbroeck, C., Metcalfe, B. and Caley, T.: FAME (v1.0): a simple module to simulate the effect of planktonic foraminifer species-specific habitat on their oxygen isotopic content, *Geoscientific Model Development*, 1–22, doi:10.5194/gmd-2017-251, 2018.

[Rosenthal, Y. and Broccoli, A. J.: In search of Paleo-ENSO, \*Science\*, 304 \(5668\), 219-221, 2004](#)

10 Roy, T., Lombard, F., Bopp, L. and Gehlen, M.: Projected impacts of climate change and ocean acidification on the global biogeography of planktonic Foraminifera, *Biogeosciences*, 12(10), 2873–2889, doi:10.5194/bg-12-2873-2015, 2015.

Ruddiman, W. F.: Pleistocene Sedimentation in the Equatorial Atlantic: Stratigraphy and Faunal Paleoclimatology, *Geological Society of America Bulletin*, 82, 283–302, 1971.

15 Sadekov, A., Eggins, S. M., De Deckker, P. and Kroon, D.: Uncertainties in seawater thermometry deriving from intratest and intertest Mg/Ca variability in *Globigerinoides ruber*: UNCERTAINTIES Mg/Ca SEAWATER THERMOMETRY, *Paleoceanography*, 23(1), n/a-n/a, doi:10.1029/2007PA001452, 2008.

Sadekov, A., Eggins, S. M., De Deckker, P., Ninnemann, U., Kuhnt, W. and Bassinot, F.: Surface and subsurface seawater temperature reconstruction using Mg/Ca microanalysis of planktonic foraminifera *Globigerinoides ruber*, *Globigerinoides sacculifer*, and *Pulleniatina obliquiloculata*: SEAWATER TEMPERATURE RECONSTRUCTION, *Paleoceanography*, 24(3), doi:10.1029/2008PA001664, 2009.

20 Sadekov, A. Y., Ganeshram, R., Pichevin, L., Berdin, R., McClymont, E., Elderfield, H. and Tudhope, A. W.: Palaeoclimate reconstructions reveal a strong link between El Niño-Southern Oscillation and Tropical Pacific mean state, *Nature Communications*, 4, 2692, doi:10.1038/ncomms3692, 2013.

[Schiebel, R., Barker, Stephen, Lendt, R., Thomas, H. and Bollmann, J.: Planktic foraminiferal dissolution in the twilight zone. \*Deep-Sea Research. Part II. Topical studies in oceanography\* 54 \(5-7\), 676-686. 10.1016/j.dsr2.2007.01.009, 2007](#)

25 Schott, W.: Zur klimaschichtung der tiefseesedimente im aquatorialen atlantischen ozean, *Geologische Rundschau*, 40(1), 20–31, 1952.

Schott, W.: Foraminiferenfauun und stratigraphie der tiefsee-sedimente in Nordatlantischen ozean, *Rep. Swedish Deep Sea Exped.*, 7(8), 357–469, 1966.

30 [van Sebille, E., Scussolini, P., Durgadoo, J. V., Peeters, F.J.C., Biastoch, A., Weijer, W., Turney, C., Paris, C.B., and Zahn R.: Ocean currents generate large footprints in marine palaeoclimate proxies, \*Nature Communications\*, 6, 6521 doi: 10.1038/ncomms7521, 2015](#)

35 [Seierstad et al.: Consistently dated records from the Greenland GRIP, GISP2 and NGRIP ice cores for the past 104 ka reveal regional millennial-scale  \$\delta^{18}\text{O}\$  gradients with possible Heinrich event imprint, \*Quaternary Science Reviews\*, 106, 29-46, doi:10.1016/j.quascirev.2014.10.032, 2014](#)

Formatted: Normal

Formatted: Normal

Formatted: Normal



Skrivanek, A. and Hendy, I. L.: A 500 year climate catch: Pelagic fish scales and paleoproductivity in the Santa Barbara Basin from the Medieval Climate Anomaly to the Little Ice Age (AD 1000–1500), *Quaternary International*, 387, 36–45, doi:<https://doi.org/10.1016/j.quaint.2015.07.044>, 2015.

5 Spero, H. J.: SYMBIOSIS IN THE PLANKTONIC FORAMINIFER, *ORBULINA UNIVERSA*, AND THE ISOLATION OF ITS SYMBIOTIC DINOFLAGELLATE, *GYMNODINIUM BÉII* SP. NOV. <sup>1</sup>, *Journal of Phycology*, 23, 307–317, doi:[10.1111/j.1529-8817.1987.tb04139.x](https://doi.org/10.1111/j.1529-8817.1987.tb04139.x), 1987.

Spero, H. J. and DeNiro, M. J.: The influence of symbiotic photosynthesis on the  $\delta^{18}\text{O}$  and  $\delta^{13}\text{C}$  values of planktonic foraminiferal shell calcite, *Symbiosis*, 4(1–3), 213–228, 1987.

10 Spero, H. J. and Lea, D. W.: Experimental determination of stable isotope variability in *Globigerina bulloides*: implications for paleoceanographic reconstructions, *Marine Micropaleontology*, 28(3–4), 231–246, doi:[10.1016/0377-8398\(96\)00003-5](https://doi.org/10.1016/0377-8398(96)00003-5), 1996.

Spero, H. J., Bijma, J., Lea, D. W. and Bemis, B. E.: Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes, *Nature*, 390(6659), 497–500, doi:[10.1038/37333](https://doi.org/10.1038/37333), 1997.

15 Staines-Urías, F., González-Yajimovich, O. and Beaufort, L.: Reconstruction of past climate variability and ENSO-like fluctuations in the southern Gulf of California (Alfonso Basin) since the last glacial maximum, *Quaternary Research*, 83(3), 488–501, doi:<https://doi.org/10.1016/j.yqres.2015.03.007>, 2015.

[Steinhardt, J., Cleroux, C. de Nooijer, L.J., Brummer, G.-J.A., Zahn, R., Ganssen, G.M., and Reichert, G.-J.: Reconciling single-chamber Mg / Ca with whole-shell  \$\delta^{18}\text{O}\$  in surface to deep-dwelling planktonic foraminifera from the Mozambique Channel, \*Biogeosciences\*, 12, 2411-2429, <https://doi.org/10.5194/bg-12-2411-2015>, 2015](#)

20

Stott, L., Poulsen, C., Lund, S. and Thunell, R.: Super ENSO and Global Climate Oscillations at Millennial Time Scales, *Science*, 297(5579), 222, doi:[10.1126/science.1071627](https://doi.org/10.1126/science.1071627), 2002.

Stott, L., Cannariato, K., Thunell, R., Haug, G. H., Koutavas, A. and Lund, S.: Decline of surface temperature and salinity in the western tropical Pacific Ocean in the Holocene epoch, *Nature*, 431(7004), 56–59, doi:[10.1038/nature02903](https://doi.org/10.1038/nature02903), 2004.

25 [Sverdrup, H.U., Johnson, M.W. and Fleming, R.H.: The Oceans, Their Physics, Chemistry, and General Biology. Prentice-Hall, New York, 1942](#)

[Telford, R., Li, C., and Kucera, M.: Mismatch between the depth habitat of planktonic foraminifera and the calibration depth of SST transfer functions may bias reconstructions. \*Climate of the Past\*, 9, 859-870, <https://doi.org/10.5194/cp-9-859-2013>, 2013](#)

30

Thirumalai, K., Partin, J. W., Jackson, C. S. and Quinn, T. M.: Statistical constraints on El Niño Southern Oscillation reconstructions using individual foraminifera: A sensitivity analysis, *Paleoceanography*, 28(3), 401–412, doi:[10.1002/palo.20037](https://doi.org/10.1002/palo.20037), 2013.

35 [Thunell, R.C., Kier, R.S., and Honjo, S.: Calcite dissolution: an in situ study in the Panama Basin. \*Science\*, 212 \(4495\), 659-661, 1981](#)

Formatted: Normal

Formatted: Normal

Formatted: Normal

[Trenberth, K.E. and Otto-Bliesner, B.L.: Toward Integrated Reconstruction of past Climates, Science, 300 \(5619\), 589-591, 2003](#)

5 de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. and Pawlowski, J.: Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces, Proceedings of the National Academy of Sciences, 96(6), 2864–2868, doi:10.1073/pnas.96.6.2864, 1999.

10 de Vargas, C., Bonzon, M., Rees, N. W., Pawlowski, J. and Zaninetti, L.: A molecular approach to biodiversity and biogeography in the planktonic foraminifer *Globigerinella siphonifera* (dâ€™Orbigny), Marine Micropaleontology, 45(2), 101–116, 2002.

Vetter, L., Kozdon, R., Mora, C. I., Eggins, S. M., Valley, J. W., Hönisch, B. and Spero, H. J.: Micron-scale intrashell oxygen isotope variation in cultured planktic foraminifers, Geochimica et Cosmochimica Acta, 107, 267–278, doi:10.1016/j.gca.2012.12.046, 2013.

15 [Wang, J.: The coupling of the Annual Cycle and ENSO over the Tropical Pacific, Journal of the Atmospheric Sciences, 51 \(8\), 1115-1136. https://doi.org/10.1175/1520-0469\(1994\)051<1115:TCOTAC>2.0.CO:21994, 1994](#)

Wang, C., Deser, C., Yu, J.-Y., DiNezio, P. and Clement, A.: El Niño and Southern Oscillation (ENSO): A Review, in Coral Reefs of the Eastern Tropical Pacific: Persistence and Loss in a Dynamic Environment, edited by P. W. Glynn, D. P. Manzello, and I. C. Enochs, pp. 85–106, Springer Netherlands, Dordrecht., 2017.

20 Waterson, A. M., Edgar, K. M., Schmidt, D. N. and Valdes, P. J.: Quantifying the stability of planktic foraminiferal physical niches between the Holocene and Last Glacial Maximum, Paleooceanography, doi:10.1002/2016PA002964, 2016.

[Wessel, P.: Sizes and Ages of Seamounts using remote sensing: Implications for Intraplate Volcanism, Science, 277, 802-805, 1997](#)

25 [Wessel P. Lyons S.: Distribution of large Pacific seamounts from Geosat/ERS-1, Journal of Geophysical Research, 102, 459–475, 1997](#)

[Weyl, P.K.: Micropaleontology and Ocean Surface Climate, Science, 202 \(4367\), 475-481, 1978](#)

30 White, S. M., Ravelo, A. C. and Polissar, P. J.: Dampened El Niño in the Early and Mid-Holocene Due To Insolation-Forced Warming/Deepening of the Thermocline, Geophysical Research Letters, 45(1), 316–326, doi:10.1002/2017GL075433, 2018.

[Wilke, I., Peeters, F. J. C., Bickert, T.: The influence of seawater carbonate ion concentration \[CO<sub>3</sub><sup>2-</sup>\] on the stable isotope composition of planktic foraminifera species \*Globorotalia inflata\*, Marine Micropaleontology, 58\(4\), 243-258, https://doi.org/10.1016/j.marmicro.2005.11.005, 2006](#)

35 Williams, D. F.: Late Quaternary fluctuations of the polar front and the subtropical convergence in the southeast Indian Ocean, Marine Micropaleontology, 1, 363–375, 1976.

Williams, D. F. and Johnson, W. C.: Diversity of Recent planktonic foraminifera in the southern Indian Ocean and Late Pleistocene paleotemperatures, Quaternary Research, 5, 237–250, 1975.

Formatted

Formatted: Normal

Formatted: Normal

Formatted: Normal

Formatted: Subscript

Formatted: Superscript

Formatted: Font: Italic

Formatted: Normal

Wise S.: Calcite compensation depth. In: Sedimentology. Encyclopedia of Earth Science. Springer, Berlin, Heidelberg, doi: 10.1007/3-540-31079-7, 1978

Formatted: Normal

5 Wit, J. C., Reichart, G. J. and Ganssen, G. M.: Unmixing of stable isotope signals using single specimen  $\delta^{18}\text{O}$  analyses, *Geochemistry, Geophysics, Geosystems*, 14(4), 1312–1320, doi:10.1002/ggge.20101, 2013.

Wolf-Gladrow, D. A., Bijma, J. and Zeebe, R. E.: Model simulation of the carbonate chemistry in the microenvironment of symbiont bearing foraminifera, *Marine Chemistry*, 64(3), 181–198, doi:10.1016/S0304-4203(98)00074-7, 1999.

Worzel, J.L.: Extensive deep sea sub-bottom reflections identified as white ash, PNAS, 45, 349, 1959

Formatted: Normal

10 Wunsch, C.: On sharp spectral lines in the climate record and the millennial peak, *Paleoceanography*, 15(4), 417–424, doi:10.1029/1999PA000468, 2000.

Wunsch, C. and Gunn, D. E.: A densely sampled core and climate variable aliasing, *Geo-Marine Letters*, 23(1), 64–71, doi:10.1007/s00367-003-0125-2, 2003.

15 Wycech, J. B., Kelly, D. C., Kitajima, K., Kozdon, R., Orland, I. J. and Valley, J. W.: Combined effects of gametogenic calcification and dissolution on  $\delta^{18}\text{O}$  measurements of the planktic foraminifer *Trilobatus sacculifer*, *Geochem. Geophys. Geosyst.*, doi:10.1029/2018GC007908, 2018.

Xie, S-P.: On the Genesis of the Equatorial Annual Cycle, Journal of Climate, 7, 2008-2013, 1994

20 Xie, S-P.: Interaction between the annual and interannual variations in the equatorial Pacific, Journal of Physical Oceanography, 25, 1930-1941, 1995

Formatted: Normal

Žarić, S., Donner, B., Fischer, G., Mulitza, S. and Wefer, G.: Sensitivity of planktic foraminifera to sea surface temperature and export production as derived from sediment trap data, *Marine Micropaleontology*, 55(1–2), 75–105, doi:10.1016/j.marmicro.2005.01.002, 2005.

25 Žarić, S., Schulz, M. and Mulitza, S.: Global prediction of planktic foraminiferal fluxes from hydrographic and productivity data, *Biogeosciences*, 3(2), 187–207, doi:10.5194/bg-3-187-2006, 2006.

Zebiak, S.E. and M.A. Cane: A Model El Niño-Southern Oscillation, Monthly Weather Review, 115, 2262–2278, https://doi.org/10.1175/15200493(1987)115<2262:AMENO>2.0.CO;2, 1987

Formatted: Normal

30 Zeebe, R. E. and Sanyal, A.: Comparison of two potential strategies of planktonic foraminifera for house building:  $\text{Mg}^{2+}$  or  $\text{H}^{+}$  removal?, *Geochimica et Cosmochimica Acta*, 66(7), 1159–1169, doi:10.1016/S0016-7037(01)00852-3, 2002.

Zeebe, R. E. and Wolf-Gladrow, D. A.: *CO<sub>2</sub> in seawater: Equilibrium, Kinetics, Isotopes*, 1st ed., Elsevier, Amsterdam., 2001.

35 Zhang, Z., Leduc, G. and Sachs, J. P.: El Niño evolution during the Holocene revealed by a biomarker rain gauge in the Galápagos Islands, *Earth and Planetary Science Letters*, 404, 420–434, doi:https://doi.org/10.1016/j.epsl.2014.07.013, 2014.

Zhang, S., Li, T., Chang, F., Yu, Z., Xiong, Z., and Wang, H., Correspondence between the ENSO-like state and glacial-interglacial condition during the past 360 kyr, Chinese Journal of Oceanology and Limnology, 35 (5), 1018-103, http://dx.doi.org/10.1007/s00343-017-6082-9, 2017

Zhu, J., Liu, Z., Brady, E., Otto-Bliesner, B., Zhang, J., Noone, D., Tomas, R., Nusbaumer, J., Wong, T., Jahn, A. and Tabor, C.: Reduced ENSO variability at the LGM revealed by an isotope-enabled Earth system model, *Geophysical Research Letters*, 44(13), 6984–6992, doi:10.1002/2017GL073406, 2017a.

- 5 Zhu, Z., Feinberg, J. M., Xie, S., Bourne, M. D., Huang, C., Hu, C. and Cheng, H.: Holocene ENSO-related cyclic storms recorded by magnetic minerals in speleothems of central China, *Proc Natl Acad Sci USA*, 114(5), 852, doi:10.1073/pnas.1610930114, 2017b.

<b>Page 17: [1] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Font: (Default) +Headings (Times New Roman), Not Highlight		
<b>Page 17: [2] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [3] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Font: (Default) +Headings (Times New Roman), Not Highlight		
<b>Page 17: [4] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [5] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Font: (Default) +Headings (Times New Roman), Not Highlight		
<b>Page 17: [6] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [7] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Font: (Default) +Headings (Times New Roman), Not Highlight		
<b>Page 17: [8] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [9] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Font: (Default) +Headings (Times New Roman), Not Highlight		
<b>Page 17: [10] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [11] Formatted</b>	<b>Brett Metcalfe</b>	<b>19/07/2019 18:07:00</b>
Font: Italic, Not Highlight		
<b>Page 17: [12] Formatted</b>	<b>Brett Metcalfe</b>	<b>19/07/2019 18:07:00</b>
Not Highlight		
<b>Page 17: [13] Formatted</b>	<b>Brett Metcalfe</b>	<b>19/07/2019 18:08:00</b>
Not Highlight		
<b>Page 17: [14] Formatted</b>	<b>Brett Metcalfe</b>	<b>19/07/2019 18:08:00</b>
Font: Italic, Not Highlight		
<b>Page 17: [15] Formatted</b>	<b>Brett Metcalfe</b>	<b>19/07/2019 18:08:00</b>
Not Highlight		
<b>Page 17: [16] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [17] Formatted</b>	<b>Brett Metcalfe</b>	<b>21/07/2019 14:49:00</b>
Not Highlight		
<b>Page 17: [18] Formatted</b>	<b>Brett Metcalfe</b>	<b>19/07/2019 14:48:00</b>

Highlight

▲  
**Page 17: [19] Formatted**                      **Brett Metcalfe**                      **19/07/2019 18:08:00**

Not Highlight

▲  
**Page 17: [20] Formatted**                      **Brett Metcalfe**                      **09/07/2019 14:04:00**

Highlight

▲  
**Page 17: [21] Formatted**                      **Brett Metcalfe**                      **09/07/2019 14:04:00**

Highlight

▲  
**Page 17: [22] Formatted**                      **Brett Metcalfe**                      **10/07/2019 11:23:00**

Highlight

▲  
**Page 17: [23] Formatted**                      **Brett Metcalfe**                      **20/07/2019 19:09:00**

Highlight

▲  
**Page 17: [24] Formatted**                      **Brett Metcalfe**                      **21/07/2019 10:41:00**

Superscript

▲  
**Page 17: [25] Formatted**                      **Brett Metcalfe**                      **21/07/2019 10:41:00**

Subscript

▲  
**Page 17: [26] Formatted**                      **Brett Metcalfe**                      **21/07/2019 10:43:00**

Font: Italic

▲  
**Page 17: [27] Formatted**                      **Brett Metcalfe**                      **20/07/2019 19:09:00**

Font: Italic

▲  
**Page 17: [28] Formatted**                      **Brett Metcalfe**                      **21/07/2019 16:17:00**

Font: Italic

▲  
**Page 17: [29] Formatted**                      **Brett Metcalfe**                      **21/07/2019 15:06:00**

Highlight

▲  
**Page 17: [30] Formatted**                      **Brett Metcalfe**                      **10/07/2019 12:25:00**

Highlight

▲  
**Page 17: [31] Formatted**                      **Brett Metcalfe**                      **10/07/2019 12:25:00**

Highlight

▲  
**Page 17: [32] Formatted**                      **Brett Metcalfe**                      **10/07/2019 12:25:00**

Highlight

▲

## **Comments to the Author:**

*Editor Decision: Reject (27 May 2019) by Helen McGregor*

*Comments to the Author:*

*Dear Dr Metcalfe*

*Thank you for your manuscript titled 'On the validity of foraminifera-based ENSO reconstructions' (manuscript #cp-2019-9) submitted for publication in *Climate of the Past*. The manuscript has received three referee reviews and one short comment during the Discussion phase.*

*After careful consideration of your Author Responses I regret to inform you that we are unable to accept the manuscript for publication at this time. The three referee reports and short comment are supportive of the FAMES approach, however they bring up similar issues on this specific manuscript, and fundamentally, there appears to be a mismatch between the intention of the study as read by the referees (and myself) in the submitted manuscript, and the intention of your study as stated in the Author Responses. The proposals for revision in the Author Response do not to my mind reconcile this key issue, nor adequately address the other major issues, and therefore I cannot accept the manuscript.*

*The detailed referee reports are made available for your reference and I encourage you to engage with these comments constructively in your future studies.*

*I wish to clarify for future reference that *Climate of the Past* referees are not required to cross-reference each other's reports on the Discussions forum and are free to develop their reviews independently.*

*Again, thank you for your interest in the journal and I regret that I cannot be more positive on this occasion.*

*Yours sincerely,*

*Helen McGregor*

### Further comments intended to clarify our work towards all reviewers and the Topical Editor

Having received reviews from all three anonymous referees we would like to make some general comments before posting a line by line response to referees #2 and #3.

We note that all three referees have submitted very similar reviews. We invested quite some effort in responding to the points raised by the first referee as quickly as possible, so that referees #2 and #3 would be able to consult our response before submitting their own personal reviews. It seems, however, that these two referees submitted very similar reviews as referee #1, which would suggest that referees #2 and #3 have not taken our response to referee #1 into account. It is unfortunate that the full potential of the Climate of the Past Discussions forum format was not exploited. It would indeed have been much more profitable to discuss the comments of referee #2 & #3 if they had been written in the light of our initial response. We also note that even though we strongly disagree in our initial response with the first reviewer comment, at time of writing, there has been no further comments from the reviewers on the discussion page on this first response although the discussion period is not ended. In the interest of stimulating such a discussion, we will briefly synthesise and address the three main criticisms raised by all three reviewers here and our arguments against them. There are several other, minor criticisms/corrections raised by the referees which we find very valid, for which we thank the reviewers and which we will gladly address. The main criticisms raised by the three reviews can be distilled into the following four points:

*(1) The study is a forward model and, therefore, does not inverse model ENSO recorded by sediments.*

This is correct. Our study uses 60 years of observed monthly climate data to forward model planktonic foraminifera populations accounting for habitat water depths and growth season for the entire Pacific Ocean, allowing us to investigate whether or not ENSO dynamics are recorded by foraminifera populations *in the water*. We think that this investigation is in itself a fundamentally interesting subject for research/publication, since it is the first step/prerequisite to investigate whether it is possible to distinguish ENSO from non-ENSO conditions from the planktonic foraminifer fossil record. This question has not been investigated before and is, therefore, new and has the potential for interesting conclusions.

Our model approach is indeed not an inverse model. It is also not a sediment model. The manuscript that we have submitted is a forward model of planktonic foraminifera populations in the water. Clearly, the reading of the referees' comments shows that they would have preferred if we had submitted a completely different manuscript; as they have reviewed our manuscript as if it were yet another study on the inverse problem of detecting a change in ENSO from a change in the distribution of sediment archive planktonic foraminifer  $\delta^{18}\text{O}$ . We are disappointed and strongly disagree with being reviewed in the view of a different manuscript that could have been written, and not evaluated based on the actual manuscript submitted and the science contained therein. We feel this is against the basic principles of a standard review process and strongly object to this treatment of our work.

*(2) The FAME output in this study has not been tested against core tops / sediment records, etc.*

This is correct. In the present manuscript, we have indeed used a 60 years monthly record of climate input to run our FAME foraminifera population model. Core tops and sediment records in the Pacific integrate foraminifera populations from centuries/millennia of time, and are therefore totally unsuitable to compare to our model output in the context of evaluating the impact of ENSO. The best thing that we could think of to compare our FAME output to, in the interest of this particular study, would be data from monthly plankton tows spanning multiple decades, but we couldn't find any. There is thus simply no data we know of that can be used meaningfully in the context of our current work on ENSO. We note, however, that the FAME model itself has been validated against



core tops in Roche et al. (2018), including core tops from the Pacific Ocean. The claim that FAME has not been validated in the Pacific Ocean is therefore unfounded.

*(3) The statistics are not applicable to palaeo records*

This is correct. We did indeed not select a statistical test on the basis of it being applicable to palaeo records. We chose the statistical test (Anderson Darling test) that was most suitable for our purpose and hypothesis: to examine the (dis)similarity of different high-resolution probability distributions produced by our FAME output driven by 60 years of monthly observational data. This approach allows us to use the full potential of our model output to directly test our hypothesis. The standard deviation (as suggested by all three reviewers), is unsuitable for comparing the dissimilarity of distributions, as it can return spurious results: for example, it is possible that two distributions of differing shape could have the same standard deviation, whereas an A-D test would detect a difference. The research question and its associated hypotheses are the central tenant upon which we based our experimental design. We note that we are not developing a statistical toolbox for palaeorecords and readers should of course always assess, on a case by case basis, statistics that are appropriate for their own studies.

*(4) The authors have not run a bioturbation model.*

This is correct. Since we are not modelling sediments in this manuscript, we did not carry out a transient bioturbation model. It would not be suitable to carry out a transient bioturbation model upon 60 years' worth of input data, seeing as bioturbation models require, due to the interplay between sediment accumulation rate (SAR) and bioturbation depth, many millennia of data/spin up time in order to produce valid output.

We were, however, curious as to whether the parts of the Pacific Ocean where FAME predicts that foraminifera in the water can record ENSO dynamics coincide with high SAR sediments. Therefore, as an extra, we included a rough map of regions of the Pacific Ocean where SAR < 5 cm/ka. The reviewers asked why 5 cm/ka was used as a cut-off point. We intentionally used a very generous cut-off point of 5 cm/ka and assumed that the reader would understand that 5 cm/ka is a very slow SAR that would severely hamper the detection of average ENSO dynamics for specific time periods (i.e. less than one millennia). It is not, however, necessary to carry out a transient bioturbation model to understand the time periods integrated by a 5 cm/ka SAR. We can simply carry out a calculation following established understanding of the influence of bioturbation upon the age dispersal of single foraminifera (Berger and Heath, 1968; Berger and Johnson, 1978; Berger and Killingley, 1982), the same understanding that is included in transient bioturbation models themselves (e.g. Trauth, 2013; Dolman and Laepple, 2018; Lougheed et al., 2018). In such a case, assuming the common bioturbation depth of 10 cm (Peng et al., 1979; Trauth et al., 1997; Boudreau, 1998), we can calculate that the 1 sigma age value of foraminifera contained within a single cm of a 5 cm/ka core is:

$$10 \text{ [cm]} / (5 \text{ [cm/ka]} / 1000) = 2000 \text{ years (from which follows that 2 sigma = 4000 years).}$$

Our map may of course miss limited parts of the Pacific Ocean with SAR higher than 5 cm/ka. We will mention this and/or highlight such places in the final version of the manuscript.

Kind Regards,

Brett Metcalfe

Bryan Lougheed

Claire Waelbroeck

Didier Roche

## B. Metcalfe on behalf of the co-authors, Response to Reviewer 1

*[reviewer comments as quoted blocks]*

We appreciate the effort the reviewer has taken in reviewing our paper and thank them for their interest in our work. The rapid publishing of a review on the same day as the reviewer has been nominated for review, especially given the level of detail contained within the review, must have involved great effort, and we thank the reviewer for making time for us. The reviewer raises some important points, but we believe that the reviewer may have misunderstood the goals of our paper, or that we have may not have explained them clearly enough. We would, therefore, like to take this opportunity to clear up these matters.

*["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 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."].*

We should make clear that we are principally not undertaking an evaluation of sediment-based individual foraminifera analysis, but rather an evaluation of the foraminifera populations in the water, before they have been incorporated into the sediment. We will endeavour to make this clearer in the manuscript, as we may not have done so (: we do not explicitly say ‘we are modelling individual foraminifera distributions’ as the reviewer assumes we are, however we also do not explicitly say ‘we are not modelling individual foraminifera distributions’ therefore, a clarifying sentence we will be added to a revised manuscript). Although, clearly, the question of whether foraminifera populations in the water are themselves able to record ENSO or not is important for sediment-based reconstructions, so we have included some minor discussion of sediment dynamics. Furthermore, the FAME methodology we apply does not simulate individual foraminifera, rather, it produces what the likely  $\delta^{18}\text{O}_c$ ,  $T_c$  value for a time-step using a function that ‘weights’ water depths by foraminiferal growth. In other words, FAME is producing mean population values for a given time slice.

The reviewer has suggested that we carry out analysis that is outside the scope and purposes of our manuscript, for example:

*"Furthermore, the statistical analysis focuses on a forward problem rather than the inverse problem that is the real challenge for detecting changing ENSO from individual foraminiferal analysis".*

We should make clearer, and will be glad to do so in the final version, that the purpose of our study is to determine whether or not the foraminiferal population produced under El Nino conditions are statistically different from the Neutral and La Nina conditions during the period of the observed climate record. Such a difference is an important prerequisite for any analysis: Can we detect the change we are searching for? In order to address this question, we choose a forward model as the most suitable tool. In the following reply to the reviewer, we will expand upon why the inverse method may not be suitable for our application/research question; discuss the use of the statistical methods in our paper which will hopefully address the reviewers concerns about palaeo-applications; discuss the reviewer’s comments regarding validation; and answer specific questions.

### **Inverse Problem**

*[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  $d^{18}\text{O}_c$  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.]*

While interesting, we believe that the inverse approach is not suitable to our particular research question. We will make this clearer in the manuscript. The inverse problem, as its very name

suggests, flips the question: “we have this data what variables must have occurred to produce them”. One is inverting the scientific method to explain causal factors from observations rather than explaining observations with causal factors. The reviewer points out several papers that have done this approach, but there is a lack of large-scale analysis beyond single cores or forcing a climate model with these boundary conditions to explain inter basin variability. So, why did we not use an inverse problem, well, firstly, it has been done before, as the reviewer suggests:

*(“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.”)*

And, secondly, it would not address the central question we are asking. Using the inverse problem would change our fundamental question from ‘*determining whether the foraminiferal population produced under El Nino conditions are statistically different from the Neutral and La Nina conditions*’ to an entirely different question, namely, ‘*with this dataset what magnitude and frequency of ENSO would have to have occurred to produce these observations*’. The reviewer seems to partly realise that these are not the same question:

*(“The forward problem is whether El Nino, neutral, and La Nina months have different distributions and requires that each individual  $d18O_c$  or  $T$  value be assigned beforehand to one of those three states.”)*

Crucially and in contrast to the inverse methodology, our research question does not exclude the possibility of there being no detectable change, while the reviewer’s proposed question forces the ENSO parameters to contort into those that generate a particular dataset. In conclusion (of this point), the inverse problem approach would not give us an answer to the hypothesis and/or research question that we have chosen: **with a chosen set of input parameters (temperature, salinity) using an ecological model what would the theoretical observations of  $T_c$  or  $\delta^{18}O$  be? And would the populations of El Nino, Neutral and La Nina climate be similar or different?**

In addition, for the purposes of carrying out the inverse problem, there is a lack of sediment-based data to do a large, basin-wide inverse-analysis. As we have already shown in our study (namely in the SAR and water depth/CCD plots), the seafloor of the Pacific is not conducive to providing samples with which to perform an inverse analysis on a basin-wide scale, and there is also a sampling bias, as the reviewer correctly alludes to: (“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)?”).

A further problem is that we would need to vary temperature AND salinity to realistically produce an inverse model, and not just temperature, as is the case when producing an accurate  $\delta^{18}O_c$ .

*[ Furthermore, the statistical analysis focuses on a forward problem 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  $d18O_c$  or  $T$  value be assigned beforehand to one of those three states. The inverse problem is to determine from comparison of two different  $d18O_c$  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). ]*

We would like to point out that neither in our response to the reviewer, nor in the paper, are we being critical of inverse modelling. We believe that it can be an appropriate and valuable technique, but it is not the suitable technique for answering our central research question. There is a fundamental difference in what the reviewer would like us to produce and what we have done (“*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*”), namely that our paper sets out to use FAME to produce distributions using an input of temperature and salinity. The

reviewer would like us to produce temperature and salinity from the distributions. However, how we should create these distributions without the input temperature parameter required for FAME is not clear. Indeed, as should be clear from the reading of the FAME methodology already published (Roche et al., 2018), there is no simple bijective relationship between the  $\delta^{18}\text{O}_c$  and the oceanic variables (T,  $\delta^{18}\text{O}_{sw}$ ).

## Statistics

*[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.*

*[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. 1sigma, quantiles etc.) rather than a new method as introduced here (Anderson-Darling test)?]*

To reiterate, the reason we use the long-established Anderson-Darling (AD) (1954; doi:10.2307/2281537) approach is because it is the most suitable method for the computer modelling study that we are carrying out, for the following reasons: The FAME model, coupled to the observational climate data that we have inputted, can produce high-resolution probability density functions (PDF) associated with El Nino, La Nina and neutral conditions. An Anderson-Darling test allows us to directly test if these PDFs are significantly different from one another or not. Obviously, an Anderson-Darling test may be more difficult to apply to foraminifera sampled from natural archives, where workers are limited in data resolution by the number of foraminifera that can be picked for analysis, by bioturbation of the natural archive, etc. Subsequently, in those cases, it might indeed make sense to use simpler statistics such as standard deviation and quantiles. Since we are not analysing natural archives, but rather data produced by a model for which we control what is generated, it makes more sense, in our case, to use a more powerful method such as the Anderson-Darling test.

Nonetheless, we appreciate that workers analysing natural archives are accustomed to using more straightforward statistical analyses, and would also like to see the 1 sigma and quantile intervals, so we will additionally report those for comparison in a revised version of the manuscript. Of course, these statistical parameters may not answer our research question and will not impact the answer to our research question (as AD is the appropriate test). These tests may however have flaws, given that the standard deviation is not the best descriptor of non-normal data and outside of the realms of statistics its usage assumes that the data will significantly impact the standard deviation or that standard deviation can be used as a measure of ENSO. If for instance one follows the reasoning of Mix (1987), a species may actually calcify solely in the anomaly regions (the la nina or el nino), and such species may not have a standard deviation that due to ENSO. This argument can be used for quantile-quantile, if the species does not calcify for the full year, an assumption of such an approach, it will mean that the data does not reflect the full year but a subset.

## Validation of the forward model

*[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... 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?]*

The model is indeed a simplification of the earlier FORAMCLIM model, because light, food etc. are not easily parameterised in models or validated with proxies. However, to claim that Roche et al. has “no clear assessment of the errors [was] presented” is not correct in our opinion. For the question relating on how the model works, the reviewer is referred to the initial publication where all the equations are described in detail; the code is itself also made available in the supplementary online material so that the reviewer can even try the model itself. While the reviewer’s comments on Roche et al. (2018) relate to another publication, in the interest of discussion let us focus on how the reviewer would validate the model.

*“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.”*

Unfortunately, this would not work due to the closed sum problem. Relative abundance between species is based upon a closed sum calculation, i.e. not the true abundance as a fraction of the total foraminifera flux. Therefore, variation caused by other species not being considered/modelled has the potential to alter the relative abundance of the species being considered. In other words, if you take the relative abundance of *G. ruber*, *G. sacculifer* and *N. dutertrei* and sum them to 1 that would not get rid of the closed sum problem, because it fails to consider other species which are not being modelled. In fact, you would not only magnify the counting error, but you would also be basing your data on a small percentage of the total foraminifera flux. Additionally, as clearly stated in Roche et al., 2018, the FAME model is constructed so as to produce a  $\delta^{18}\text{O}_c$  value where the given species of foraminifera is assumed to be able to grow: “It should be clearly understood that this approach is not able to and does not attempt to determine the relative abundances of the different species. Instead FAME provides a simplified approach to compute the  $\delta^{18}\text{O}_c$  of a generic population of foraminifers if environmental conditions permit its growth. From a model–data perspective, this approach enables one to compute the calcite  $\delta^{18}\text{O}$  for a given species, were it to exist in the sedimentary record “. A further useful reference in this instance is the actual formulation of the model in equation 8, page 3590 of Roche et al. (2018).

*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)?*

As the reviewer states in this question, there is a sampling bias within sediments hence the selection of a forward model and why in this instance it is more logical than an inverse model.

*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?*

As discussed earlier, FAME is not attempting to produce the measured IF distributions from natural archives.

*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).*

Thank you for the comment. Figure 4 and supplementary figure S3 are already aiming at this, and we will endeavour to make this clearer. We will expand upon the section that is already included in the paper: **“The model-driven results were assessed with the underlying observational dataset, to check how the dataset alters with FAME the input data (temperature and  $\delta^{18}\text{O}_{\text{eq}}$ ) underwent statistical testing (Figure 4 and Figure S3). Instead of a variable depth, we opted for fixed depths at 5, 149 and 235 m, giving a Eulerian view (Zhu et al., 2017a) in which to observe the implications of a dynamic depth habitat. By using a fixed depth, these results show that the shallowest depths produce populations that are significantly different both in terms of their mean values and their PDF. In the upper panel of Figure 4, the canonical El Niño 3.4 region is clearly visible at 5 m depth. Whilst differences exist between the temperature (Figure 4) and the FAME Anderson Darling results (Figure 3), for instance close to the Panama isthmus, there are significant similarities between the plots. These plots also show that our FAME data (Figure 3), in which we allow foraminiferal growth down deeper than the depths in Roche et al. (2018), are a conservative estimate and thus are on the low-end (Figure 4), to account for potential discrepancies with depth habitats.”**

We do not fully understand the following comment of the reviewer:

*“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)?”*

In the way FAME is set up, the weighting for depth is based upon growth, without FAME we would be unable to integrate the required weighting function. Hence why it necessitated fixing the depth for the analysis of the input temperature and  $\delta^{18}\text{O}_{\text{eq}}$  values.

### **Validation of $\delta^{18}\text{O}_c$**

*[“Validate the  $\delta^{18}\text{O}_c$  predictions from the growth rate and geochemistry model. This was done in Roche et al., 2017 but is also somewhat circular because the sedimentary  $\delta^{18}\text{O}$  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.*

*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  $\delta^{18}\text{O}_c$  and measured MARGO  $\delta^{18}\text{O}_c$ ? Why not use those depths?”]*

Thank you for your comment. It is important to stress that there is no geochemistry model in our approach. The optimisation procedure of Roche et al. (2018), gives the maximum allowed growth depth of each species. However, as the reviewer discussed previously, we should test how this influences the resultant distributions we generate, therefore we constrained the model to four depths including to a depth known to be below the photic zone. This is what is stated in page 4 line 7-19; page 5 line 10. We will rephrase this with clarity in mind.

### **Dismissive of Mg/Ca-Temperature?**

*[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 (Sadokov, 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  $\delta^{18}O_c$ ), to ignore such a widespread type of analysis seems very shortsighted. 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.]*

*Page 9 line 20 – Why discount trace metal temperatures (Mg/Ca)? Include the analysis of Tc for Mg/Ca reconstructions.*

*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.]*

The reviewer is alluding to Figure 6, the Tc or calcite/recorded temperature, which is essentially our pseudo-Mg/Ca\* produced with FAME (\* = It is a weighting of temperature rather than  $\delta^{18}O_{eq}$ ). This is discussed in the dataset, we also ran the temperature (Figure 4) of the dataset by itself (without the foraminiferal growth rates). It is true that we don't go into too much detail and we will expand our section discussing the FAME produced temperature (Tc).

However, it is not as the reviewer states as us being 'so dismissive of Mg/Ca analyses' (we are sorry if we created such an impression). A great many researchers are dedicating their time to this valuable geochemical analysis. However, given that the species-specific conversion from temperature to Mg/Ca is not as straight-forward as  $\delta^{18}O_c$  and  $\delta^{18}O_{eq}$ , it would therefore require more parameters, which we are not confident in modelling at this stage. A pseudo Mg/Ca would also need to be validated (yet techniques are not standardised nor cross calibrated to a sufficient degree, with users using laser ablation; pooled specimens and/or whole shell) and the problems associated with dissolution, cleaning for analysis, are not easily parameterised in a model. We do welcome discussions on the computation of pseudo Mg/Ca and consider it something that could be included in the future, possibly in a second generation of the FAME model.

## **Removal of maps**

*[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 where carbonate is preserved that are not captured in the coarse DEM used; 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.]*

As stated in the paper: - "The resolution of the ocean reanalysis data for the time period 1958-2015 would essentially be analogous with a sediment core representing 50 yr-1 cm-1 (or 20 cm-1 kyr-1). Based on our analysis, such a hypothetical core with a rapid sediment accumulation rate (SAR) could allow for the possible disentanglement of El Niño related signals from the climatic signal using IFA, but only in a best-case scenario involving minimal/no bioturbation, which is unlikely in the case of oxygenated sediments". This is an important caveat to communicate to the reader. We believe that removing the maps would remove this valuable piece of information.

## **Definition of ENSO components**

*["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 Niño (La Niña) 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.”]*

What we said: “The tropical Pacific Ocean is divided into four Niño regions based on historical ship tracks, from east to west: Niño 1 and 2 (0° to -10°S, 90°W to 80°W), Niño 3 (5°N to -5°S, 150°W to 90°W), Niño 3.4 (5°N to -5°S, 170°W to 120°W) and Niño 4 (5°N to -5°S, 160°E to 150°W). One index for ENSO, the Oceanic Niño Index (ONI), based upon the Niño 3.4 region (because of the region’s importance for interactions between ocean and atmosphere) is a 3-month running mean of SST anomalies in ERSST.v5 (Huang et al., 2017). However, Pan-Pacific meteorological agencies differ in their definition (An and Bong, 2016, 2018) of an El Niño, with each country’s definition reflecting socio-economic factors, therefore, for simplicity we utilise a threshold of  $\chi \geq +0.5^{\circ}\text{C}$  as a proxy for El Niño,  $-0.5^{\circ}\text{C} \leq \chi \leq +0.5^{\circ}\text{C}$  for neutral climate conditions and  $\chi \leq -0.5^{\circ}\text{C}$  for a La Niña in the Oceanic Niño Index. Many meteorological agencies consider that five consecutive months of  $\chi \geq +0.5^{\circ}\text{C}$  must occur for the classification of an El Niño event. However, here it is considered that any single month falling within our threshold values as representative of El Niño, neutral or La Niña conditions (grey bars in Figure 1). By using this threshold, three weighted histograms for each  $\delta^{18}\text{O}_{\text{c}}$  and  $T_{\text{c}}$  and their resultant distributions (El Niño; Neutral; and La Niña) were computed for every month and for every latitude and longitude grid-point for the 1958-2015 period.”

Why did we do this? Because if a foraminifer lives for 30 days then how appropriate would “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)” be? Because sediments can’t resolve annual/sub-annual resolution (like corals or molluscs), therefore the periods where the threshold passes 0.5 would in the sediment be mixed with with El Niño or La Niña (as in they would have potentially similar values as an El Niño, and as time cannot be resolved they would be considered as El Niño). In the sediment the minimum consecutive months is not used to define an El Niño, as it is impossible, an arbitrary value or any kind of quantile- or sigma distribution is.

### **Next paper.**

*["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.”]*

We are not attempting, at this stage, to specifically reproduce single foraminifera analysis. How ENSO amplitude and frequency impact foraminiferal distributions is a separate paper we are working upon, because it actually cannot be dealt within as a simple discussion topic (and would require a specific and different dataset from the current paper’s dataset), it is something that we thought about as we worked on this manuscript. Therefore, we suggested this approach in our conclusions / perspectives as something that could be worked on in the future, which is something that we believe is normal in scientific manuscripts. We will attempt to make clearer that we are referring to possible future work.

### **Specific comments**

*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*

We disagree, we believe that the inclusion of SAR and water depth (CCD) adds important context to our paper. Those are the two main factors that allow for the carbonate signal to be preserved in sufficient temporal resolution. We can consider adding the location of sea mounts to the map, thank you for this idea.

*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?*

The ‘why’ have been dealt within in Roche et al. (2018). The ‘how different’ is comparing apples and oranges, FAME requires temperature as an input whereas FORAMCLIM needs temperature, light, and organic carbon (food). Light and food are not included in many datasets, nor are they parameterised or have proxies. A validation step of the two different models using the same observational input data is thus not simply attainable. The input data here does not have either of these additional variables (we considered for example using a long term chlorophyll record from satellite data, but such datasets ignore the deep chlorophyll maximum).

*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.*

Four different depths (60; 100; 200 and 400 m) have been used in the model, the use of the shallow and deeper depths likely don’t capture one or more of the species actual ecologies, however that is why we ran it with different depths to understand how chosen depth alters (or doesn’t alter) the results.

What we said: - “Consequently, we allow all the species of foraminifera to grow down to ~ 400 m (depending if optimal temperature conditions are met) to capture the total theoretical niche width. As the optimised depths of Roche et al. (2018) are shallower, and upper ocean water is more prone to temperature variability, our approach likely dampens both the modelled  $\delta^{18}\text{O}_c$  and  $T_c$ . Therefore, the sensitivity of the model was tested by applying the same procedure but with the limitation of the depth set to 60; 100 and 200 m.”

### **–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  $10^{-3}$  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  $10^{-3}$  term, the equation would read:  $d18O\_VPDB = 0.97002*d18O\_VSMOW - 29.98$ .*

The reviewer is referring to the fractionation difference between water and carbonate for which the given expressions are indeed correct. However, what is referred to in the manuscript is the conversion between two scales with measured  $\delta^{18}\text{O}_{sw}$ ; those can be converted from V-SMOW to V-PDB by  $-0.27\text{‰}$  (please see Figure 1 in Hut, 1987 for the original reference).

*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.*

We agree with the reviewer that our initial sentence was somewhat ill-formulated. What we meant here is that the major driver of foraminiferal growth is temperature and hence taking it (temperature) into account will provide the first order signal, as was discussed already in Roche et al. (2018). A revised version of the manuscript will include a modified statement as follow: “these variables for now can be set aside as temperature provides the dominant signal, it is worth noting that in all probability some variance will arise from these processes and deviation between observed and expected values should consider this.” Regarding the validation process we refer the reviewer to the discussion already given 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.*

Pg. 6, Line’s 8 to 10. The ONI dataset was used, this is based upon the 3 month smoothed anomaly in the El Nino 3.4 region, we decided to set thresholds including anything above 0.5 and below -0.5 in their respective EL Nino and La Nina bins because unlike, e.g. corals, a palaeoceanographer using sediment core foraminifera cannot discern a specific year. An ‘almost El Nino’ anomaly won’t be discernible from a full El Nino period in the fossil record, because unlike coral records we cannot determine what the previous 3 months were like.

*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?*

The data was fitted using a fit distribution procedure in MatLab because the statistical function requires a distribution to test. We chose to use the kernel distribution because it mimics the underlying dataset well and we were testing a large number of grid points, therefore we decided to keep numerous parameters constant (for instance we could have decided to change the distribution using a find the best fit distribution but this would have made intercomparison problematic), however to allow our fitted distribution to better mimic the underlying distribution we allowed the programme to vary the bandwidth between grid points for an optimal kernel distribution.

*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?*

The reviewer is correct that it would solve the minor binning error – but it wouldn’t solve the rounding error, if you round these numbers....

What we wrote:- “As the weighted distributions are effectively probability distributions, in order to fit a distribution, we multiplied the bin counts by 1000, effectively converting probability into a hypothesised distribution. Using the repeat matrix function (MatLab function: repmat), a matrix of  $\delta^{18}O_c$  was produced using each bin’s mid-point ( $\delta^{18}O_{mid}$ -point) there is a threefold error combined with this methodology which may account for minor variation between discrete runs of the model: first the counts values were rounded to whole integers so an exact number of cells could be added to a matrix; secondly the  $\delta^{18}O_{mid}$ -point was used which gives an error associated with

the bin size ( $\pm 0.05\%$ ) that is symmetrical close to the distributions measures of central tendency but asymmetrical at the sides; and finally, the associated rounding error at the bin edges within a histogram ( $\pm 0.005\%$ ).”

### –Results–

*Page 7 line 3 – It says that the mean  $\delta^{18}O_c$  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?*

We will reword this sentence for clarity.

*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).*

Here we are discussing an unmixing analysis for a single time slice, if enough foraminifera are measured then it can be possible to disentangle mathematically the various components that go into a single distribution. However, this is only possible if the values of El Nino, La Nina etc. have a different absolute  $\delta^{18}O$  value, our point here. This is true regardless of an unmixing analysis or and holds true for any proxy.

*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.*

True, that is why we tested the distributions as well. Here we are discussing the fact that similar values would be impossible to unravel – we will make this clear. This (“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”) is why we chose to use a statistical test that looks into the distribution.

*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.*

We disagree, if the model has some measurable error, it is appropriate to advertise the fact to readers that at some locations the distributions whilst significantly different with one test fall within the model ‘error’. Hence the use of hashing.

*Page 8 line 9 – Unclear what “on the low-end” means.*

We will clarify this.

*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.*

First the title is “On the validity of foraminifera-based ENSO reconstructions” is ambiguous as to whether they are or are not valid. Second the abstract is referring to the entire discussion, the calculated distributions and the SAR/depth. We will clarify this statement.

*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).*

Yes, we will make this clearer. We note that we cite our individual foram paper (Lougheed et al, 2018) using single shell  $^{14}\text{C}$  to show that a single cm can be not just multi-centennial, but multi-millennial.

*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.*

The reviewer has stated what we are stating in that sentence, we will rephrase for clarity. However, ENSO is not the background signal otherwise it would not be detectible through a temperature anomaly. It is a short term climatic event when one considers that a single cm in ocean sediments can reflect hundred to thousands of years.

*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.*

Thank you for pointing this sentence out. We will clarify this sentence to explain in better detail what we mean. Firstly, the absolute magnitude of events would obviously be smoothed out if one were to be apply discrete, multi-specimen sample downcore analysis (i.e. not single foram analysis), as the reviewer is obviously aware of. Were single foram analysis to be applied, the single foram values corresponding to high-magnitude ENSO events would indeed still be present in the sediment record, as the reviewer correctly points out. However, single foraminifera from multiple ENSO events, and non-ENSO climate, would all be mixed into the same discrete interval, meaning that a time-series of ENSO is essentially not possible to produce, and therefore: (1) the frequency of ENSO events becomes difficult to detect, and (2) that one is forced to make *a priori* assumptions regarding the behaviour of background climate and ENSO climate in past times in order to differentiate between ENSO and non-ENSO single foraminifera in the palaeo record.

*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.*

We agree with the reviewer it is not fixed, the ‘sediment accumulation rate needed to observe/reconstruct changes’ ideally would reflect the percentage of foraminifera within the sediment growing during ENSO events and the magnitude of the events not just the number of events. We furthermore note that, to avoid using a SAR cut off that could be considered arbitrary, we intentionally used a very generous cut-off of 5 cm/ka. Were we to set the cut-off to be higher, following the more traditional

lower cut-off of 10 cm/ka (Bard et al, Shackleton et al), then the areas of the Pacific basin that could be considered suitable would be even more limited.

*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.*

We used the latest GEBCO, but we would be more than happy to include higher resolution data.. However, we are obviously not saying a seamount would not be useable. We can consider adding sea mounts to the map.

*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.*

We will rephrase this sentence. Here we referring to the interpretation, for instance figure 19.3 of Cole and Tudhope (2017)

*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.*

We will add a citation(s).

*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 foraminifera Tc or d18Oc, sampling uncertainty on the IF distribution. See also note above on the role of sedimentation rate.*

At the reviewer’s suggestion we will add in ‘sampling-bias’ into the sentence

*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  $\sigma(d18Oc)$  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.*

This sentence may have led the reviewer to misinterpret our results as sediment-based individual foraminiferal analysis centric, we will rephrase this sentence for clarity and suitability.

*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.*

Here we are discussing other’s findings, for instance, Zhu computed the variance and found that some of the signals detected could be a by-product of the annual cycle.

### –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.*

The reviewer is right, “they have not all used 1-sigma d18Oc as the metric for detecting change”, that is why they are not cited. The reviewer would be justified in suggesting these references here, had we not repeatedly cited them throughout our paper. Whilst we will attempt to make this clearer for the reader, it is worth noting that a few sentences later, we directly refer to the papers the reviewers cites (see comment below:).

Page 12 line 21 – “Overall, our results suggest that foraminiferal  $\delta^{18}\text{O}$  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?

We will add clarity to this statement, however we would like to note that this quote neglects the second part, the first word overall here being ‘generally speaking’ eludes to the fact that it’s a sentence that has a follow up: “Overall, our results suggest that foraminiferal  $\delta^{18}\text{O}$  for a large part of the Pacific Ocean can be used to reconstruct ENSO, **especially if an individual foraminiferal analysis (Lougheed et al., 2018; Wit et al., 2013) approach is used (Ford et al., 2015; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Sadekov et al., 2013; White et al., 2018)**, contrary to previous analysis (Thirumalai et al., 2013). **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.**”

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.

As the reviewer states our statement is ‘generally true’, therefore the difference is our interpretation vs. the reviewer’s interpretation, we politely disagree.

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.

Conclusions do not have to include the main focus of the study but can include information that present the findings in a different light (as in what it means to the readers) or what can be next done. We don’t agree with the reviewer’s suggestion that it is an untested model, but also doubting that Ecophysiological models are not limited to foraminifera – this is a factual statement and they can be of use.

### –Figures–

Where we have used FAME they are growth rate weighted values – this is explained in the methodology and Roche et al. (2018). Figure 4 and supplementary figure 3 use the input data therefore they are non growth rate weighted.

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

As the key shows they represent where the populations are statistically different, the hashing draws the eye too much so for those panels with hashes we make it grey. As one of our species does not have these hashes it remains white.

*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).*

What the caption says: - “Figure 4. Results of an Anderson-Darling test between El Nino and Neutral climate conditions based upon the Temperature input data: Fixed depth.” This is the temperature input data

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

As the key shows they represent where the populations are statistically different, the hashing draws the eye too much so for those panels with hashes we make it grey. As one of our species does not have these hashes it remains white.

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

Where we have used FAME they are growth rate weighted values

## B. Metcalfe on behalf of the co-authors, Response to Reviewer 2

*[reviewer comments as red text in blocks]*

We thank Reviewer 2 for their time in submitting this review of our work, before we address the main and minor comments we like to (re)address a point raised by Reviewer 1 and expand upon application of our study.

In their manuscript Metcalfe et al. present a forward modeling approach through FAME to investigate the use of individual foraminifera analysis (IFA) for ENSO reconstruction. Based on the modeling results, they conclude that this proxy is only valid in part of the Pacific Ocean. However, these regions are often characterized by low sedimentation rate, therefore limiting the use of this proxy.

The reviewer correctly identifies that we have carried out a forward model of foraminifera  $\delta^{18}\text{O}$ . Specifically, we used the FAME model, driven by observed climate data, to predict foraminifera population  $\delta^{18}\text{O}$  accounting for their habitat water depths and growth season, to test whether foraminifera populations in the water can record the ENSO signal. We are not modelling IFA in sediment cores. We carry out a minor discussion of sediment core IFA, because, clearly, whether or not ENSO dynamics can be recorded by foraminifera populations in the water itself has consequences for sediment core records (which obviously source their foraminifera tests from the water column).

While the effort to incorporate forward models into paleoceanographic studies is commendable, I fail to see the practical application of this study.

The reviewer might prefer if our study was more similar to an inverse model approach, however, the manuscript that we submitted is a forward model. We regret if we have not elucidated enough in our paper the practical applications of our forward model. We see several applications:

**Research question / proxy validation** – Pg. 12 ln 30-32 “The use of ecophysiological models (Kageyama et al., 2013; Lombard et al., 2009, 2011) 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.” One application can be validating the research question, i.e., whether foraminifera in the water will in the first place (even before they become fossilised, disturbed, winnowed away) record the signal we are looking (e.g., similar to Leduc et al., 2009; Ford et al., 2015). A pre-screening using the available information (temperature, salinity, etc.) and our understanding of certain processes can be used to focus our research, determine species selection. Which leads on to...

**Site selection** – In an ideal world the location of (palaeo-) data would be at least one datapoint every 1x1 degree latitude longitude (or a higher resolution) so that a direct comparison between model and data could be obtained. However, this is neither practical nor feasible in the real world where access to time and funding (etc.) is limited. The time and effort (from personal experience) that goes into sampling, washing, picking, measuring (and dependent upon pooled or individual, replicates or no replicates, etc) foraminifera is a lot. Forward models can therefore be used to select sites not on geographic proximity to existing published records but based on their strategic location, which can provide critical information about certain climatological processes for data-model comparison. Naturally, a bioturbation module would further elucidate such an approach. There are three models/analytical tools that exist that deal with picking and signal modulation: Sedproxy; FIRM or INFAUNAL. Therefore, we focussed upon the construction of a signal, which leads on to....

**Sandbox modelling** - Pg. 12 ln 26-27: “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. FAME is intended for climate models where boundary conditions can be varied.

Inverse modeling would be impossible and the lack of comparison between the pseudo-proxy distributions and actual distributions of foraminifera prevents validation of the method.

### Major comments

#### Inverse problem

The manuscript focuses on forward modeling of IFA analysis. Although definitely a valuable exercise for data-model comparison (assuming that the climate model can make use of the forward model), it doesn't solve the inverse problem. It would be almost impossible to evaluate the growth factor in the  $\delta^{18}\text{O}$  record.



Our commentary regarding the inverse problem is available with respect to Reviewer 1, we will surmise our key points here: (i) It is not suitable/applicable to our research question; (ii) a lack of a large scale dataset to perform a basin-wide analysis; and (iii) lack of a bijective relationship between  $\delta^{18}\text{O}_c$  and the oceanic variables ( $T$ ,  $\delta^{18}\text{O}_{sw}$ ).

“(assuming that the climate model can make use of the forward model)” - FAME is built with climate models in mind. Data-model comparison studies suffer from an ‘apples and oranges’ problem, of which there are two key problems: (i) the conversion of units i.e., most proxies reconstructing temperature do not give values of temperature in degrees C or K but in their own proxy units (per mil, mmol/mol, etc) requiring a conversion, and (ii) a reduction in scales, i.e., models give a wealth of information (multiple layers) in the time-depth domain. FAME was produced (Roche et al., 2018) to (i) to generate pseudo-proxy time-series from model runs that can be compared with age-depth values down cores. Naturally, including a bioturbation, or mixing, module into down-core work is prudent for core datasets (please see comment below; and to (ii) reduce the information for a given time-slice into a manageable value using an integration that would make sense on a biological point of view (integrating various depths with equal weightings might seem logical, but foraminifera for instance grow at different rates depending on their temperature) and).

It’s also not visually obvious what the difference between the output of a non-weighted model is vs FAME in Figure S1. Some statistics would help, or plotting the resulting kernel distributions on a separate panel,

We will elaborate further on these plots to make them more understandable, Figure S1 are produced solely by FAME, the difference is one is weighted for a larger proportion of growth per month and the other bins the total number of months, these plots were picked at random from the dataset.

Further, bioturbation is also likely to have a large impact on IFA, especially in areas of low sediment accumulation. Why not connect FAME to a bioturbation model and disentangle the influence of these factors?

There are three points we would like to address with this comment, the first is the reviewers admission that IFA (and by association pooled) distributions can be impacted by bioturbation. The second point is that it asks the follow-up question to our own research question, i.e. “Are the populations between different climatological end members significantly different, **and can this difference be resolved in the sediment record**”, this second question would need different input parameters, as discussed next. The third point we would like to address is that our input dataset is ~60 years, in the grand scheme of things a relatively small contribution to the sediment. We choose not to link the dataset to a bioturbation module because the time series is not long enough, a 1cm kyr<sup>-1</sup> SAR with a 10 cm mixed layer depth would need several thousand years of input data (please see comment further below). Furthermore, the application of bioturbation to a monthly time-series can be done in several ways, for instance should one could bioturbate the ‘settling flux’, i.e. the monthly signal, as soon as it is encoded or at the end of the time-series.

#### Statistical analysis

Page 6, Line 25: Multiplying the bin counts will effectively skewed the results of a significance test. In practice, it would be impractical if not impossible to obtain 1000 samples in each bin.

We will reword this section as whilst we state ‘total bin counts’ we add a clarification: - “As the weighted distributions are effectively probability distributions, in order to fit a distribution, we multiplied the total bin counts by 1000 (i.e. so that the total sum is 1000).

Similarly, page 7, line 4, how many foraminifera were artificial picked to produce these maps?

We did not artificially pick foraminifera (here we are referring to population rather than a simulated sample). We are testing the water populations that would be El Nino or La Nina or Neutral. As we are not modelling IFA in sediment cores, we therefore also do not include a picking routine.

#### IFA model - data comparison

There are a number of recent studies with IFA results from the past 1000 years (some of them cited in the current manuscript). How do these distributions compare to the statistical ones?

We will address this in a revised form of the MS.

#### Effect of SAR

Since a model of bioturbation was not implemented here, it’s hard to examine the effect of bioturbation on the IFA.

We have indeed not specifically modelled the effect of bioturbation upon our data, because it is simply not possible to carry out a transient bioturbation model run upon 60 years data. Therefore, to avoid appearing biased against sediment archives in our SAR map, we intentionally used a very generously low SAR cutoff of 5 cm/ka. However, it is not necessary to run a transient bioturbation model to demonstrate the limitations of a SAR that is less than 5 cm/ka. We can carry out simple calculation following established understanding of the influence of bioturbation upon the age dispersal of single foraminifera (Berger and Heath, 1968; Berger and Johnson, 1978; Berger and Killingley, 1982), the same understanding that is included in transient bioturbation models themselves (e.g. Trauth, 2013; Dolman and Laepple, 2018; Lougheed et al., 2018). In such a case, assuming a bioturbation depth of 10 cm (Peng et al., 1997; Trauth et al., 1997; Boudreau, 1998), we can calculate that the 1 sigma age value of foraminifera contained within a single cm of a 5 cm/ka core is  $10 / (5/1000) = 2000$  years (from which follows that 2 sigma = 4000 years). – We will include a version of this paragraph in a revised manuscript

Furthermore, rapid accumulation rates should be possible around islands. The coarse map overlaid here fails to account for these. I would suggest adding to the text that in strategic locations (in the blue areas), sedimentation rates may still be high enough.

We agree seamounts and islands are places where the depth is shallower than our prescribed carbonate compensation cut-off depth and could also potentially have higher SAR, we will add a clarifying statement to the figures and the main-text.

#### Improper referencing

This is not the first study to use pseudo-proxy to examine whether IFA can be used for ENSO reconstruction. Thirumalai et al. present a model that can be more easily applied to a real application. First, reference this study (and others) at the beginning of the manuscript and second, why not extend their “picking” model to also evaluate the contributions of sample size?

We reiterate that our forward model seeks to model the  $\delta^{18}O$  of the full foraminifera populations in the water. We are not modelling sediment core IFA.

#### Minor comments

Abstract: Should state that this is an IFA technique.

See answer to your previous point.

Page 1, Line 23: specify that the interaction on interannual timescale is known as ENSO. On decadal, it's known as the PDO.

We will rephrase to:

“Predictions of short-term, abrupt changes in regional climate are imperative for improving the spatiotemporal precision and accuracy when forecasting future climate. Coupled ocean-atmosphere interactions (wind circulation and sea surface temperature) in the tropical Pacific, collectively known as the El Niño-Southern Oscillation (ENSO) **on interannual timescales and the Pacific Decadal Oscillation on decadal timescales**, represent global climate’s largest source (Wang et al., 2017) of inter-annual climate variability (Figure 1).”

Page 1, Line 27: SO is part of ENSO. Should rephrase as we have long instrumental records of the atmospheric variability but not the ocean.

We will rephrase to:

“The instrumental record of the past century provides important information (*i.e.* the Southern Oscillation Index; SOI), however, detailed **oceanographic observations of the components of ENSO (both the El Niño and Southern Oscillation)**, such as the Tropical Oceans Global Atmosphere (TOGA; 1985-1994) experiment only provide information from the latter half of the twentieth century (Wang et al., 2017).”

Page 3, line 3: Stott et al is not the only reconstruction in the Western Pacific, either use e.g., or as done previously cite multiple sources.

We will include multiple sources and an e.g., so it will read as follows: “and (3) those associated with the trace metal geochemistry (e.g., Ford et al., 2015; Sadekov et al., 2013; Stott et al., 2002, 2004; White et al., 2018), more specifically the natural logarithm of the relative concentration of Mg and Ca ( $\ln(\text{Mg}/\text{Ca})$ ), of the shell, based upon the temperature dependent (Elderfield and Ganssen, 2000; Nürnberg et al., 1996) incorporation and substitution of a Mg cation into the calcite lattice (Branson et al., 2013, 2016).”

However, we would like to clarify that the rationale for the citation here is not referring exclusively to the Western Pacific pool, but rather citing papers (including earlier papers) around the topic of 3 proxy types as noted on Pg. 2 Ln23:

“Proxies of past ENSO and Pacific SST (Ford et al., 2015; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Leduc et al., 2009; Sadekov et al., 2013; White et al., 2018) are based upon the biomineralisation of the calcite, or a polymorph such as verite (Jacob et al., 2017), and shells of foraminifera (Emiliani, 1955; Evans et al., 2018; Zeebe and Wolf-Gladrow, 2001). There are three major types of foraminifera-based palaeoceanographic proxies: (1)... (2)... and (3)...”

**Page 3, Line 30: Mg/Ca is not a simple function of temperature. There is a growing body of evidence that suggests that Mg/Ca is also sensitive to salinity and pH. In addition, the calcite saturation of the bottom waters on post-depositional preservation of the signal.**

We agree that is why on Pg. 11 Ln 16 – 23 we state : “several other parameters can alter this technique, this includes abiotic effects such as salinity (Allen et al., 2016; Gray et al., 2018; Groeneveld et al., 2008; Kısakürek et al., 2008) or carbonate ion concentration (Allen et al., 2016; Evans et al., 2018; Zeebe and Sanyal, 2002); biotic effects such as diurnal calcification (Eggins et al., 2003; Hori et al., 2018; Sadekov et al., 2008, 2009; Vetter et al., 2013); or additional factors such as sediment (Fallet et al., 2009; Feldmeijer et al., 2013) or specimen (Barker et al., 2003; Greaves et al., 2005) ‘cleaning’ techniques. Given the role of Mg in inhibiting calcium carbonate formation, the manipulation of seawater similar to the modification of the cell’s pH (de Nooijer et al., 2008, 2009) may aid calcification and explain the formation of low-Mg by certain foraminifera (Zeebe and Sanyal, 2002)”

As such we are producing a pseudo proxy of what the proxy aims to test, i.e. temperature, we will add some text with regards to these later lines, where the reviewer has pointed it out, as follows:

“Here, we use the recently developed *Foraminifera as Modelled Entities* (FAME) model (Roche et al., 2018) to take into account potential modulation of  $\delta_{18}O_c$  and the Temperature recorded in the calcite, herein  $T_c$ , by foraminifera growth.  $T_c$  can thus be considered as an estimate of the proxy Mg/Ca (**albeit one uninfluenced by secondary factors**).”

**Page 5, Line 3: Why not used species- specific equations?**

A first reason is that the use of equilibrium as opposed to species-specific equations places all foraminiferal species against an equal benchmark. More importantly, most species-specific equations have been produced in culture (though some others in tows and pump samples) so that they contain variability in both growth and environmental (the same data that produced the growth functions in FORAMCLIM/FAME) conditions.

As an equilibrium oxygen isotope matrix (time-depth) is used to produce the FAME weighted distributions, said equilibrium can be switched out for ‘equilibrium species values’, however, such a terminology does not clearly outline the influence of time (growth season) and depth on the different species’ oxygen isotopes.

**Page 5, Line 12: Not sure what is meant by “Which can compute eight foraminiferal species”. Do you mean growth?**

The sentence will be altered as follows:

“FAME is based upon FORAMCLIM which can compute **the growth of** eight foraminiferal species (Kageyama et al., 2013; Lombard et al., 2009, 2011; Roche et al., 2017), however comparison with a core top database has been limited to five foraminiferal species (Roche et al., 2017).”

**Page 6, line 11: There is an abundant body of literature dealing with the definition of an ENSO event. Why not start there?**

We are not attempting to challenge or redefine ENSO or discount ONI or the BJ-Index (etc.) derived event chronologies, instead our simplification is based upon two things: (1) the simplicity of the input variables and (2) whether foraminifera (~4 weeks life) would ‘sense’ the event (as stated in our previous comment to reviewer 1). To clarify we will alter the paragraph as follows:

Pg. 6 Ln 10 – 15: “However, Pan-Pacific meteorological agencies differ in their definition (An and Bong, 2016, 2018) of an El Niño, with each country’s definition reflecting socio-economic factors, therefore, for simplicity we utilise a threshold of  $\chi \geq +0.5^\circ\text{C}$  (**where  $\chi$  is the value of ONI**) as a proxy for El Niño,  $-0.5^\circ\text{C} \leq \chi \leq +0.5^\circ\text{C}$  for neutral climate conditions and  $-0.5^\circ\text{C} \leq \chi$  for a La Niña in the Oceanic Niño Index. Many meteorological agencies consider that five consecutive months of  $\chi \geq +0.5^\circ\text{C}$  must occur for the classification of an El Niño event. However, here it is considered that any single month falling within our threshold values as representative of El Niño, neutral or La Niña conditions (grey bars in Figure 1). **This simplification reflects the lifecycle of planktonic foraminifera (~4 weeks) as the population at time step  $t$  knows not what happened at  $t-1$  or will happen at  $t+1$ . As we are producing the mean population growth weighted  $\delta^{18}O$  values, ‘almost’ El Niño or ‘almost’ La Niña would be indistinguishable from the build-up and subsequent climb-down of actual El Niño and La Niña events. Therefore, these ‘almost’ El Niño or ‘almost’ La Niña are placed within their respective climatological pools as El Niño or La Niña.**”

Page 9, line 20-25: Most of these studies are based on pooled samples and were referencing to an ENSO-like signal rather than the interannual mode of variability that IFA is targeting.

We state that these datasets are from pooled and individual (Pg. 9 ln 18-20 “Several authors have focussed on individual foraminifera analysis (IFA) or pooled foraminiferal analysis in the Pacific region, either for trace metal or stable isotope geochemistry.”) and later we point out that these are the authors inference (Pg. 9 ln 22-23 “The resultant data of such studies have been used to infer a relatively”). We agree that depending upon the method authors can be referencing ‘ENSO-like’ and/or ‘interannual mode of variability’, we make reference to this point at Pg 10 ln 9 -26 (e.g., “Our own analysis using our FAME  $\delta_{18}O_c$  and  $T_c$  output mimics foraminiferal sedimentary archives, pooling several decades worth of data in which the resolution is coarse enough to obscure and prevent individual El Niño events being visible but allowing for some kind of long-term mean state of ENSO activity to be reconstructed”). As we already take into consideration the reviewers point, we can assume that this comment merely reflects the opinion we are performing IFA.

## B. Metcalfe on behalf of the co-authors, Response to Reviewer 3

*[reviewer comments as red text in blocks]*

We thank reviewer 3 for their time in reviewing our paper. However, it is unfortunate that the reviewer did not take the time to read our response to reviewer 1 as many of the same concerns have been addressed there and we have responded to those questions in detail. It is also unfortunate because it would have, hopefully, aided the reviewer in realising that this paper is about testing whether the given foraminiferal populations are statistically different so that they could potentially unravel different climatic states. Our approach is not about IFA research; IFA research is referred to because it provides an excellent sample dataset (and we highlight that it is one of the ways to understand a climate history more thoroughly). It is however not the only dataset and hence we make reference to other studies as well. We also find it disappointing that once more we are having to discuss inverse and forward modelling, as well as the statistical tests used; all aspects that were already explained in our previous response.

*In this study, Metcalfe et al. aim to test whether the approach of using individual foraminifera analysis (IFA) can be used to assess ENSO variability. In order to accomplish this, they use the Foraminifera as Modeled Entities (FAME) model to calculate idealized foraminifera distributions across the tropical Pacific. These results are then combined with seafloor/ CCD depth and sedimentation rate to determine which regions of the Pacific Ocean are suitable targets for IFA approaches. Modeling of foraminifera populations in order to determine if ENSO change is detectable has been done before (e.g., Thirumalai 2013, White 2018), although these studies focus on the detection of ENSO from paleoclimate proxy records. This study's novel contribution is the inclusion of the FAME model and foraminiferal growth rates to the analysis of modeled response of biological calcite to tropical variability.*

We once more state we are not testing IFA or have a model which tests IFA approaches. IFA is a suitable dataset to compare results with because it gives a lot more information than pooled analysis.

*However, the FAME portion of the model is not validated against core-top data from the tropical Pacific, precluding assessment of its utility.*

The FAME model is validated against the whole MARGO dataset which includes the tropical Pacific. Please read Roche et al. (2018).

*The application of these results is likewise problematic, as it focuses on determining whether ENSO events (El Niño, La Niña) and neutral conditions have distinct distributions (forward modeling) rather than on how one could detect ENSO change (inverse modeling).*

There are a number of applications of this method, which we have outlined in response to reviewer 2. However, we consider that the reviewer comment is not an argument against what we have done – the first basic principle of understanding a proxy is 'can we detect', not 'how could we detect', as the how implies we know we can. Our manuscript is devoted to answer the question “can we detect”.

*Further, the discussion on sedimentation rate and CCD is broad-based and does not take into consideration local changes in seafloor topography, changes in bottom-water oxygen availability that may alter bioturbation depths, and the variability characteristics of different regions with regard to the seasonal cycle, decadal-centennial variability, and ENSO change (e.g., Thirumalai 2013, Ford 2015, White 2018).*

It is true we have used broad based and conservative estimators ( $5 \text{ cm}^{-1} \text{ kyr}^{-1}$ ) as the bioturbation mixed layer is known to vary from 1 to 35 cm depending upon the various aspects the reviewer states. However, we fail to understand how “and the variability characteristics [?] of different regions with regard to the seasonal cycle, decadal-centennial variability and ENSO change” would somehow relate to SAR and dissolution depth.

*Finally, there are aspects of the model that are unrealistic (e.g., a 400m depth for symbiont-bearing foraminifera; assuming sample sizes of 1000 for binning) or unrealized (e.g., how many individuals were selected for generating these estimates and a lack of model-data comparison) that present significant issues to the overall utility of this model for paleoceanographic reconstruction of ENSO from IFA.*

In the paper we state we use more than one depth, we first apply a CUT-OFF value of 400 m then progressively shallower (hence the reason for multiple panels) – we know this CUT-OFF value is deeper than symbiotic species (e.g., Pracht et al., 2019, Biogeosciences) hence the use of shallower depths (the reviewer is arguing against a simple test of our model here). It should be pointed out, whilst the model can in principle run down to 400 m it will only register a value if the temperature is applicable for that species (i.e., temperatures outside of the temperature window will not give growth as highlighted by the equations in Roche et al., 2018). Second, we multiply the TOTAL BIN COUNTS by 1000 to convert it into a simple distribution to test the distributions of the various climate - this is not the picking / assumed sample size. Were we to test sample picking we would need (and would have stated) to have done a larger test (although the computation required would be enormous [samples in group \* replications \* resampling] \* [lat \* lon], i.e., an individual foraminifera picking would be  $[1*40*10,000]*[40*120] = \sim 192$  million computations).

*The title of the article does not represent the content or main goals of the study, and the conclusions stated in the abstract are different than those in the main paper.*

We disagree. If we break down the title:

**Validity – is the quality or state of something being valid (valid - being logically correct or well-grounded/justifiable).** We are testing whether the distributions of different climate events are statistically different (this is a fundamental test)

**Foraminifera-based** – we use a foraminiferal model. An alternative could include the word “populations” here, as in “foraminiferal populations”.

**ENSO reconstructions** - The reviewer, we assume, is arguing that as we say ENSO reconstructions, and judging from their previous point (“as it focuses on determining whether ENSO events (El Niño, La Niña) and neutral conditions have distinct distributions (forward modeling) rather than on how one could detect ENSO change (inverse modeling).”) that we haven’t focused on the 'how one could detect', yet we are testing whether the foraminiferal distributions of different climate events are statistically different, hence we have carried out a test on a more fundamental level (i.e. foraminifera in the water, before they are incorporated in the sediment archive).

We will rephrase the abstract and conclusions for clarity.

*The questions the authors raise are valid and useful, but the results as stated do not support their conclusions. In fact, the stated conclusions of the article are, in several places, contradicted within the paper itself. These contradictions are not well-explained, and thus a clear summary of the findings is difficult to parse.*

We will rephrase those sections the reviewer elaborates on in the general comments.

#### **General Comments**

*The study here focuses on forward modeling using FAME for IFA. However, the authors fail to prove whether existing IFA-ENSO reconstructions are valid or provide the tools for evaluating proxy data (e.g., the “inverse problem”, as mentioned in other reviews, whereby foraminifera records are analyzed to infer ENSO). Thus the application of these results to the paleodata world is limited.*

We are not forward modelling for IFA in sediment records, we are forward modelling foraminifera populations in the water.

The Paleodata world exists to answer questions regarding our understanding of past climate, therefore understanding if our proxies work for the period covering the observed climate record represents a fundamental test. The use of foraminiferal records to infer ENSO starts off with the prerequisite that what you are recording is ENSO, so our current research asked whether the values of the different climatic states for two proxies (calcite  $\delta^{18}\text{O}$  and temperature of calcification) are statistically different. The application of these results downcore or the provision of tools for evaluating palaeoproxy data were not the stated aim of our particular research question.

*The more relevant application here is in targeting locations for performing IFA studies, but this is limited as well, as the sedimentological and bioturbation properties of regions across the Pacific are much more variable captured here.*

We included a rough SAR/CCD map for the benefit of the reader. However, if a reader or potential researcher in palaeoceanography has access to much better data regarding SAR/sedimentological properties, our research would still be of value because they can compare their chosen location with the FAME results that we have generated. The fact that the FAME results and FAME-SAR-DEPTH results are plotted separately allows users to pick and choose whichever plot they find necessary.

*The authors use their own definition of ENSO events, despite significant previous literature and established definitions that are commonly used. The use of single month anomalies does not adequately represent the actual ENSO phenomenon, which relies on ocean-atmosphere feedbacks expressed over a period of months, and thus their analysis of differences between El Niño, La Niña, and neutral conditions may be flawed and biased toward non-ENSO SST anomalies.*

We extensively discussed the definition of ENSO events in the paper and our answer to reviewer 1. In the latter we explain the rationale for our choice of definition (that foraminifera life cycle is shorter than a month and therefore several populations would exist that could conceivably have the same isotopic value as a true event).

*This study does not compare the results of their FAME analysis with existing IFA reconstructions of variability from the tropical Pacific. In the eastern Pacific, Rustic 2015 used  $\delta^{18}\text{O}$  IFA on modern-era sediments to show close correspondence with calculated  $\delta^{18}\text{O}$  from reanalysis data; in the central Pacific, White 2018 showed that the distributions of Mg/Ca-based SSTs from individual foraminifera in a 4ky coretop are statistically similar to modern reanalysis data.*

As per our comment to a similar question by reviewer 2, we will address this in a revised form of the MS.

## Specific Comments

*The authors focus on  $\delta^{18}O$  proxies for IFA, and discount Mg/Ca reconstruction and the modeling efforts done with those (White 2018, Ford 2015). To discount Mg/Ca ratios as a paleoproxy without the kind of analysis provided for  $\delta^{18}O$  seems premature. While changes in carbonate concentration, salinity, and preservation environment can indeed alter Mg/Ca ratios, significant study has been done and is underway to understand these roles. Species-specific calibrations and various corrections exist that are well quantified. Not using Mg/Ca for the Tc seems rather limited.*

This point was discussed in our answer to reviewer 1. We use integrated temperature as a pseudo proxy for Mg/Ca – we did not attempt (though it would be interesting to perform such an analysis in another dedicated paper) to convert the input temperature into proxy values, i.e., a pseudo equilibrium Mg/Ca, like how the oxygen isotope values are calculated. Most ocean reanalysis and model datasets do not include the full variables required, although some models do (e.g., Grey and Evans 2019, Paleocean. Paleoclim.).

The section of the paper that deals with and discusses Mg/Ca, is not “to discount” the proxy but was written to preempt comment(s) regarding our paper about why we did not attempt to model the full variables of this proxy (as we state in the paper why it would be beyond the remit of the paper).

Nor have we, as the reviewer states, ‘discounted’ the modelling efforts done with Mg/Ca (e.g. White et al., 2018; Ford et al., 2015) – if the reviewer would like to clarify this point, we would gratefully alter the text. However, this would seem to be the reviewer’s own projection on to our paper and not something we categorically stated. We will try to make this clearer.

*The number of foraminifera picked from a given sediment interval is an important component of IFA. Increasing bin counts to 1000 artificially (Page 6) does not represent the numbers typically used in such analyses; the numbers used for other analyses (Page 7) are not specified.*

They are not specified because we didn’t used the number of foraminifera picked, we artificially convert the distribution into testable values by multiplication. However, as we have stated throughout we are testing the distribution of the *population* not a *sample*.

*In the results, the first statistical test is to test whether the means of the FPen and FP-neu  $\delta^{18}O$  distributions are different and use this to determine whether ENSO events can be detected. Comparison of the population means does not necessarily reflect differences in the population distributions, and only provides a measure of mean conditions that may or may not be related to ENSO variability. The use of the Anderson-Darling test to assess differences in distribution is used later. It is unclear how these two different tests were related, and how the mean  $\delta^{18}O$  FPen/neu was utilized.*

As per a similar comment as reviewer 1, we will reword this sentence for clarity. We agree it does not necessarily reflect differences hence why we used the AD test.

*The author’s use of the Anderson-Darling test to assess differences in distributions is novel, but results of this test are not compared to those that have been used to assess IFA results in previous studies (e.g., std dev (Thirumalai 2013, Koutavas and Joanides 2012, Rustic 2015) or Q-Q (White 2018, Ford 2015)). Is this more sensitive, less sensitive, or does it measure different aspects of the distribution change NOT captured in the other analyses? Without such comparison, the ability to assess the validity of IFA reconstruction (the purported goal of this paper) is limited.*

Different statistical techniques plot, test or validate different aspects of a sample or dataset being used. There seems to be some confusion by all three reviewers as to why the Anderson-Darling is being used. As we have already stated this before (in our reply to reviewer 1), a statistical test should be chosen by a study’s author that can be used to test the research question devised by that author. We seek to investigate the (dis)similarity of distributions, therefore the Anderson-Darling test is the suitable test.

*The specifics of sedimentation rate and bioturbation vary greatly across the tropical Pacific and rely on multiple processes. The role that oxygen plays in bioturbation is important, especially as bottom-water oxygen levels vary across the tropical Pacific. Likewise, seafloor topography is highly variable, with ridges and sea mounts that are not apparent at the resolution used.*

*On P8: “Similarly, the individual characters of El Niño events, which are very short in duration, become lost in the bioturbated sediment record “ The purpose of IFA is not to discern the properties of an individual event. Change in frequency or amplitude of events over a period of time can be statistically detected using various means to compare the distribution of integrated conditions over the period of sedimentation. Bioturbation serves, then, to extend that integrated time and the range of conditions experienced.*

Bioturbation causes, in the case of low SAR sediment record, thousands of years of time to become mixed into a single interval of sediment core. Hence, this may serve to mix values associated with the long-term climate signal and the ENSO signal.

With respect to the comment regarding IFA, we understand (e.g. Ganssen et al., 2011) that IFA is not a way to deduce a particular event (e.g. monsoon) but a way to characterize the samples. However, our sentence on page 8 is not saying that individual events are to be reconstructed, but that the characteristics of single event get muddled up in time. Now this would not be a problem (fundamentally) if  $\delta^{18}\text{O}$  did not have a  $\delta^{18}\text{O}_{\text{sw}}$  component. But  $\delta^{18}\text{O}$  does have a  $\delta^{18}\text{O}_{\text{sw}}$  component, thus, shells that are anomalous in one time period (e.g. LGM) may - with a change in the ice volume effect on the  $\delta^{18}\text{O}_{\text{sw}}$  - have a value that is similar to 'background signal' in another time period (e.g. Holocene). Hence, the use of the word 'lost'.

*Bioturbation will also not remove anomalous values (page 9) – rather, such values may be present as part of a distribution representing more integrated time. Likewise, bioturbation has the effect of smoothing the signal, but the “signal” is a function of all sources of variability (ENSO, annual, decadal, centennial). The relative expression of these forms of variability along with the amount of time integrated by a sample are both important in terms of the ability to capture ENSO signals.*

Anomalous values are only anomalous in relation to the rest of the dataset. As our answer to the comment above explains, if a samples anomalous values are moved into a sample with similar values then they will no longer be anomalous. We will clarify what we meant by smoothing the signal in a revised form of the MS.

*On P.10, Cole and Tudhope (corals) and White et al (IFA) are cited in error when discussing lake colour intensity and precipitation-driven records.*

We were referring to some of the analysis within those papers (as per our comment to a similar point of reviewer 2 we will alter this sentence).

*Also on P10, the authors claim: “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.” – Which methods? Q-Q, std. dev, event counting, others? It’s not entirely clear this is even referring to IFA reconstructions, as the records discussed previous are sedimentary, coral, and IFA (but noted as “precipitation driven”, see above).*

As our comment to reviewer 1 very clearly explained that we are not modelling IFA it is a shame that reviewer 3 did not have the time to read our replies. Here, “methods” is generally referring to proxies – we will elaborate in the revised version to reduce the confusion.

*P.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.**” This statement leaves out other key elements, including the relative expression of ENSO events, the seasonal cycle, and decade-and-longer variability. These elements are (arguably) more important for inverse modeling, where the ability to disentangle growth rates from other sources of variability is impossible, and thus the signatures of ENSO in such records need to be discerned.*

[Bold is to clarify what part of the reviewers comment is a quote] – The reviewer suggests we have left out key elements. We agree that we missed out ‘*seasonal cycle, and decade-and-longer variability*’ which we will add. But the ‘*relative expression of ENSO events*’ is already included within the reviewer’s chosen quote (underlined).

*A key point in the paper (P10) says “The results presented here imply that much of the Pacific Ocean is not suitable for reconstructing ENSO studies using paleoceanography, yet several studies have exposed shifts within std dev( $\delta^{18}\text{O}$ ) of surface and thermocline dwelling foraminifera. One can, therefore, question what is being reconstructed in such studies.”. This study has, at this point, not tested whether the Std.dev of  $\delta^{18}\text{O}$  from individual foraminifera have reconstructed ENSO (also, the wording of this sentence is odd).*

If the two populations are statistically similar (as in La Nina and El Nino have statistically indistinguishable distributions) then it is logical to question what the measure of dispersion (std dev) of the measured sample is linked to (i.e. it is seasonality, or species depth habitat change). We will rephrase in the revised version of the MS to improve readability.

*The first paragraph of the discussion (p9) purports to be about paleoclimatological archives that “have been used to indirectly and directly study past ENSO”. However, the discussion is on mean-state reconstructions (Koutavas 2003, Dubois 2009). Koutavas 2003 is non-IFA mean-state reconstruction; likewise, the Dubois 2009 paper notes that “we prefer not to invoke any ENSO-like state for the glacial EEP based solely on our UK ’37 SST.” While it may be true that this result and Koutavas 2003 are at*



*odds, this is not an issue of IFA or ENSO reconstruction, but rather aggregate analysis and mean -state reconstruction. Discussion of std.dev ENSO studies (modeled by Thirumalai, Koutavas 2006, Koutavas and Joanides 2012, Leduc 2009, Sadekov 2013, Rustic 2015) is not found, yet the following paragraph (see above) is largely about this approach. Further, significant discussion and analysis of IFA reconstructions of ENSO during the LGM is found in Ford 2015, which is not discussed here.*

The quote says “to indirectly and directly study past ENSO”, one could pool papers that are non-IFA mean state reconstructions into the ‘indirectly’ and studies that utilise IFA into ‘directly’. However, in this paragraph we are cataloguing changes (“The resultant data of such studies have been used to infer”) around the Pacific. In the example the reviewer gives of Dubois, we are referring to upwelling intensification. Here is the section the reviewer is referring to in its entirety: “The resultant data of such studies have been used to infer a relatively weaker Walker circulation, a displaced ITCZ and equatorial cooling (Koutavas and Lynch-Stieglitz, 2003); both a reduction (Koutavas and Lynch-Stieglitz, 2003) and **intensification (Dubois et al., 2009) in eastern equatorial Pacific upwelling**; and both weakened (Leduc et al., 2009) and strengthened ENSO variability (Koutavas and Joanides, 2012; Sadekov et al., 2013) during the LGM. A number of these results are contentious, for instance **the reduction in upwelling in this region (Koutavas and Lynch-Stieglitz, 2003) is contradicted by Dubois et al. (2009), who used alkenones (i.e., U37K’ ratios) to suggest an upwelling intensification.**” We can include a section in this paragraph discussing only the std dev (reviewer comment: *Discussion of std.dev ENSO studies (modeled by Thirumalai, Koutavas 2006, Koutavas and Joanides 2012, Leduc 2009, Sadekov 2013, Rustic 2015) is not found, yet the following paragraph (see above) is largely about this approach*) although it is mentioned on page 10 lines 19-26.

*The main analysis uses an unrealistic mixed-layer depth of 400m for the models foraminifera. Symbiont -bearing forams (G. ruber and G. sacculifer) live in the photic zone, and thus modeling and analysis of these organisms should be constrained to these depths.*

There isn’t really a ‘main analysis’ – we just chose the 400 m cut-off to start with. Irrespective, we know that symbiont species don’t live so deep, however, the computation of FAME is such that if the temperature is appropriate a growth rate will be calculated (see Roche et al., 2018). Therefore, as we state in the paper, we varied the cut-off depths to see how these would alter the distribution. This is intended as a sensitivity study.

*The model results using the shallower depths and specific, photic zone depths (Figure 4, figure 5, Figure 6) show that much of the tropical Pacific is suitable for such analyses, provided adequate carbonate preservation. This is very much in contrast with the point made previously in the paper that much of the tropical Pacific is unsuitable.*

Here is the crux of our comments – we state that it is possible but, as the reviewer points out (in Bold hereafter), provided the signal can be recovered (i.e. SAR/Depth). “*The model results [...] show that much of the tropical Pacific is suitable for such analyses, **provided adequate carbonate preservation.***”. We will make this clearer in the revised version

*In these figures, confusingly, some figures show significant areas in white while others use gray for no discernable reason. The figures are also improperly labeled, according to the captions – in each figure, G. sacculifer is on the left, G. ruber is in the middle, and this is reversed in the caption. Which is which?*

We will correct this – the label on the figure is correct (left *G. sacculifer*; mid *G. ruber*; right *N. dutertrei*), we will correct the figure captions. The reviewer is correct that some are white and some grey when we originally made the figure everything was white, black and hashed. Unfortunately, these hashes draw the eye (and can hide some small locations) away from the white only locations we decided to make it grey to highlight this (The top panel of Figure 5 ‘*G. sacculifer* 60 m’ is missing the grey which is our mistake but demonstrates the drawing of the eye). The *N. dutertrei* dataset does not have the hashing so we decided to make it white only.

*The conclusions are at odds with what is presented at various points in the paper. Specifically: “Overall, our results suggest that foraminiferal  $\delta^{18}O$  for a large part of the Pacific Ocean can be used to reconstruct ENSO, especially in an individual foraminifera Analysis approach is used, contrary to previous analysis (Thirumalai et al. 2013). This conclusion is contradicted in the abstract, and in various parts of the study (e.g., P10 – “the results presented here imply that much of the Pacific Ocean is not suitable for reconstruction ENSO studies with paleoceanography. . .”) Which is it?*

We will rephrase these sections for clarity: reconstruction is possible when considering foraminifera only, but not necessarily possible if SAR/depth is taken into account.

*Again, Koutavas 2003 is cited here, but that is not an IFA study. In general, clearly noting which studies are IFA/ENSO and which are mean state / aggregate / non-IFA*

*studies will clarify the discussion surrounding the use of IFA and IFA techniques to identify ENSO signals.*

Again we are not doing an IFA paper, so our citations are based upon a mixture of analyses. We will consider the reviewers suggestion of stating which studies are and are not based upon IFA more clearly in the revised text.

*This study does not directly address the Thirumalai 2013 study, as presented. The role of seasonality does not appear to be well addressed in this study (a key factor of Thirumalai 2013), the questionable definition of ENSO events confounds direct comparison, and the lack of clarity on sampling rates and other facts precludes a direct comparison. If this was a goal in this analysis, the Thirumalai study should be discussed in detail at the beginning (and should be discussed, in any case, earlier when discussing approaches for quantifying the suitability of locations for ENSO reconstruction), and the differences between their approaches (e.g., forward vs. inverse modeling). Suitable criteria for comparison should be noted (e.g., std. dev. Vs A-D tests).*

Our goal as the research question states was not to compare our results with those of Thirumalai et al. (2013) but to *test whether the foraminiferal ( $\delta^{18}\text{O}$  and Tc) distributions for climate state A and B are statistically different*. If they are different then it is theoretically possible to discern A and B when the two are pooled into the same ‘bag’. However, a secondary consideration is whether at any location (irrespective of picking etc) would the signal likely be preserved, hence the SAR and water depth are two simple characteristics that have been known to palaeoceanographers. We are using a forward model, because (as explained in our answer to reviewer 1) it is not a bijective function (i.e., reversing the equation does not give you one solution) which makes it hard to do inverse modelling. We left out a discussion of inverse and forward modelling techniques because it’s not really an appropriate discussion – in fact it would only be appropriate if the study had the ability to do both. With this goal in mind we decided to use a statistical method that *tests whether the distributions for climate state A and B are statistically different*. We could have tested the dataset with various other statistical tests but those would not directly answer our research question hence they would fail the reviewers “suitable criteria for comparison”. Standard deviation and Anderson-Darling test different things. (AD makes a comparison of two distributions and whether they can be considered the same, whereas std dev. is a value for the spread of data around a mean). Regarding confounding and precluding comparisons: The definition of ENSO events is Oceanic Nino Index except we decided to reduce it to one that is appropriate and compatible with the foraminiferal lifecycle (as stated in our comment to reviewer 1) because of the short lifecycle they have (~1 month) and the impossibility to discern ENSO-lite (i.e. those near El Nino or La Nina events – as discussed in the reply to the reviewers comment about ENSO definition) from ENSO (this will bias both our results and any work using foraminifera). As we are testing the population, we did not therefore include a sampling/picking monte carlo style parameter, as that would be testing a sample.

# ***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-

[Printer-friendly version](#)

[Discussion paper](#)



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  $\delta^{18}O_c$  or T value be assigned beforehand to one of those three states. The inverse problem is to determine from comparison of two different  $\delta^{18}O_c$  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

[Printer-friendly version](#)[Discussion paper](#)

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 (Sadkov, 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 where carbonate is preserved that are not captured in the coarse DEM

[Printer-friendly version](#)[Discussion paper](#)

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  $\delta^{18}O_c$  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 Niño (La Niña) 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

[Printer-friendly version](#)

[Discussion paper](#)



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  $\delta^{18}O_c$  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  $\delta^{18}O$  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  $\delta^{18}O_c$  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  $\frac{[d18O\_VSMOW+1]}{[d18O\_VPDB+1]} = 1.03091$  where d18O does not include the  $10^3$  term. Thus  $d18O\_VPDB = \frac{d18O\_VSMOW}{1.03091} + (1/1.03091) - 1$  or  $d18O\_VPDB = 0.97002 * d18O\_VSMOW - 0.02998$ . In d18O expressed with the  $10^3$  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

[Printer-friendly version](#)

[Discussion paper](#)



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, Niño 3.4, Niño 3, Niño 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 Niño etc. events that include the requirement for consecutive months of anomalies? This definition has a basis in theory as an El Niño (La Niña) 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  $T_c$  or  $d_{18O_c}$  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  $d_{18O_c}$  for El Niño 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?

[Printer-friendly version](#)

[Discussion paper](#)



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.

[Printer-friendly version](#)[Discussion paper](#)

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

[Printer-friendly version](#)[Discussion paper](#)

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 foraminifera Tc or d18Oc, sampling uncertainty on the IF distribution. See

[Printer-friendly version](#)[Discussion paper](#)

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  $\sigma(d18Oc)$  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,

[Printer-friendly version](#)

[Discussion paper](#)



Ford et al, Rustic et al, White et al). Furthermore, they have not all used 1-sigma  $\delta^{18}\text{O}_{\text{c}}$  as the metric for detecting change.

Page 12 line 21 – “Overall, our results suggest that foraminiferal  $\delta^{18}\text{O}$  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

[Printer-friendly version](#)[Discussion paper](#)

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.

CPD

Interactive  
comment

Printer-friendly version

Discussion paper





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

### **Anonymous Referee #2**

Received and published: 4 April 2019

In their manuscript Metcalfe et al. present a forward modeling approach through FAME to investigate the use of individual foraminifera analysis (IFA) for ENSO reconstruction. Based on the modeling results, they conclude that this proxy is only valid in part of the Pacific Ocean. However, these regions are often characterized by low sedimentation rate, therefore limiting the use of this proxy.

While the effort to incorporate forward models into paleoceanographic studies is commendable, I fail to see the practical application of this study. Inverse modeling would be impossible and the lack of comparison between the pseudo-proxy distributions and actual distributions of foraminifera prevents validation of the method.

Major comments

[Printer-friendly version](#)

[Discussion paper](#)



## Inverse problem

The manuscript focuses on forward modeling of IFA analysis. Although definitely a valuable exercise for data-model comparison (assuming that the climate model can make use of the forward model), it doesn't solve the inverse problem. It would be almost impossible to evaluate the growth factor in the d18O record.

It's also not visually obvious what the difference between the output of a non-weighted model is vs FAME in Figure S1. Some statistics would help, or plotting the resulting kernel distributions on a separate panel,

Further, bioturbation is also likely to have a large impact on IFA, especially in areas of low sediment accumulation. Why not connect FAME to a bioturbation model and disentangle the influence of these factors?

## Statistical analysis

Page 6, Line 25: Multiplying the bin counts will effectively skewed the results of a significance test. In practice, it would be impractical if not impossible to obtain 1000 samples in each bin. Similarly, page 7, line 4, how many foraminifera were artificial picked to produce these maps?

## IFA model - data comparison

There are a number of recent studies with IFA results from the past ~1000 years (some of them cited in the current manuscript). How do these distributions compare to the statistical ones?

## Effect of SAR

Since a model of bioturbation was not implemented here, it's hard to examine the effect of bioturbation on the IFA. Furthermore, rapid accumulation rates should be possible around islands. The coarse map overlaid here fails to account for these. I would suggest adding to the text that in strategic locations (in the blue areas), sedimentation

[Printer-friendly version](#)[Discussion paper](#)

rates may still be high enough.

### Improper referencing

This is not the first study to use pseudo-proxy to examine whether IFA can be used for ENSO reconstruction. Thirumalai et al. present a model that can be more easily applied to a real application. First, reference this study (and others) at the beginning of the manuscript and second, why not extend their “picking” model to also evaluate the contributions of sample size?

### Minor comments

Abstract: Should state that this is an IFA technique. Page 1, Line 23: specify that the interaction on interannual timescale is known as ENSO. On decadal, it's known as the PDO. Page 1, Line 27: SO is part of ENSO. Should rephrase as we have long instrumental records of the atmospheric variability but not the ocean. Page 3, line 3: Stott et al is not the only reconstruction in the Western Pacific, either use e.g., or as done previously cite multiple sources. Page 3, Line 30: Mg/Ca is not a simple function of temperature. There is a growing body of evidence that suggests that Mg/Ca is also sensitive to salinity and pH. In addition, the calcite saturation of the bottom waters on post-depositional preservation of the signal. Page 5, Line 3: Why not used species-specific equations? Page 5, Line 12: Not sure what is meant by “Which can compute eight foraminiferal species”. Do you mean growth? Page 6, line 11: There is an abundant body of literature dealing with the definition of an ENSO event. Why not start there? Page 9, line 20-25: Most of these studies are based on pooled samples and were referencing to an ENSO-like signal rather than the interannual mode of variability that IFA is targeting.

---

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

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

### **Anonymous Referee #3**

Received and published: 11 April 2019

In this study, Metcalfe et al. aim to test whether the approach of using individual foraminifera analysis (IFA) can be used to assess ENSO variability. In order to accomplish this, they use the Foraminifera as Modeled Entities (FAME) model to calculate idealized foraminifera distributions across the tropical Pacific. These results are then combined with seafloor/ CCD depth and sedimentation rate to determine which regions of the Pacific Ocean are suitable targets for IFA approaches. Modeling of foraminifera populations in order to determine if ENSO change is detectable has been done before (e.g., Thirumalai 2013, White 2018), although these studies focus on the detection of ENSO from paleoclimate proxy records. This study’s novel contribution is the inclusion of the FAME model and foraminiferal growth rates to the analysis of modeled response of biological calcite to tropical variability. However, the FAME portion of the model is

[Printer-friendly version](#)

[Discussion paper](#)



not validated against core-top data from the tropical Pacific, precluding assessment of its utility. The application of these results is likewise problematic, as it focuses on determining whether ENSO events (El Niño, La Niña) and neutral conditions have distinct distributions (forward modeling) rather than on how one could detect ENSO change (inverse modeling). Further, the discussion on sedimentation rate and CCD is broad-based and does not take in to consideration local changes in seafloor topography, changes in bottom-water oxygen availability that may alter bioturbation depths, and the variability characteristics of different regions with regard to the seasonal cycle, decadal-centennial variability, and ENSO change (e.g., Thirumalai 2013, Ford 2015, White 2018). Finally, there are aspects of the model that are unrealistic (e.g., a 400m depth for symbiont-bearing foraminifera; assuming sample sizes of 1000 for binning) or unrealized (e.g., how many individuals were selected for generating these estimates and a lack of model-data comparison) that present significant issues to the overall utility of this model for paleoceanographic reconstruction of ENSO from IFA. The title of the article does not represent the content or main goals of the study, and the conclusions stated in the abstract are different than those in the main paper. The questions the authors raise are valid and useful, but the results as stated do not support their conclusions. In fact, the stated conclusions of the article are, in several places, contradicted within the paper itself. These contradictions are not well-explained, and thus a clear summary of the findings is difficult to parse.

### General Comments

The study here focuses on forward modeling using FAME for IFA. However, the authors fail to prove whether existing IFA-ENSO reconstructions are valid or provide the tools for evaluating proxy data (e.g., the “inverse problem”, as mentioned in other reviews, whereby foraminifera records are analyzed to infer ENSO). Thus the application of these results to the paleodata world is limited. The more relevant application here is in targeting locations for performing IFA studies, but this is limited as well, as the sedimentological and bioturbation properties of regions across the Pacific are much

[Printer-friendly version](#)[Discussion paper](#)

more variable captured here. The authors use their own definition of ENSO events, despite significant previous literature and established definitions that are commonly used. The use of single month anomalies does not adequately represent the actual ENSO phenomenon, which relies on ocean-atmosphere feedbacks expressed over a period of months, and thus their analysis of differences between El Niño, La Niña, and neutral conditions may be flawed and biased toward non-ENSO SST anomalies.

This study does not compare the results of their FAME analysis with existing IFA reconstructions of variability from the tropical Pacific. In the eastern Pacific, Rustic 2015 used  $\delta^{18}\text{O}$  IFA on modern-era sediments to show close correspondence with calculated  $\delta^{18}\text{O}$  from reanalysis data; in the central Pacific, White 2018 showed that the distributions of Mg/Ca-based SSTs from individual foraminifera in a 4ky coretop are statistically similar to modern reanalysis data.

### Specific Comments

The authors focus on  $\delta^{18}\text{O}$  proxies for IFA, and discount Mg/Ca reconstruction and the modeling efforts done with those (White 2018, Ford 2015). To discount Mg/Ca ratios as a paleoproxy without the kind of analysis provided for  $\delta^{18}\text{O}$  seems premature. While changes in carbonate concentration, salinity, and preservation environment can indeed alter Mg/Ca ratios, significant study has been done and is underway to understand these roles. Species-specific calibrations and various corrections exist that are well quantified. Not using Mg/Ca for the Tc seems rather limited.

The number of foraminifera picked from a given sediment interval is an important component of IFA. Increasing bin counts to 1000 artificially (Page 6) does not represent the numbers typically used in such analyses; the numbers used for other analyses (Page 7) are not specified.

In the results, the first statistical test is to test whether the means of the FPen and FPneu  $\delta^{18}\text{O}$  distributions are different and use this to determine whether ENSO events can be detected. Comparison of the population means does not necessarily reflect

[Printer-friendly version](#)[Discussion paper](#)

differences in the population distributions, and only provides a measure of mean conditions that may or may not be related to ENSO variability. The use of the Anderson-Darling test to assess differences in distribution is used later. It is unclear how these two different tests were related, and how the mean  $\delta^{18}\text{O}$  FPen/neu was utilized. The author's use of the Anderson-Darling test to assess differences in distributions is novel, but results of this test are not compared to those that have been used to assess IFA results in previous studies (e.g., std dev (Thirumalai 2013, Koutavas and Joanides 2012, Rustic 2015) or Q-Q (White 2018, Ford 2015)). Is this more sensitive, less sensitive, or does it measure different aspects of the distribution change NOT captured in the other analyses? Without such comparison, the ability to assess the validity of IFA reconstruction (the purported goal of this paper) is limited.

The specifics of sedimentation rate and bioturbation vary greatly across the tropical Pacific and rely on multiple processes. The role that oxygen plays in bioturbation is important, especially as bottom-water oxygen levels vary across the tropical Pacific. Likewise, seafloor topography is highly variable, with ridges and sea mounts that are not apparent at the resolution used.

On P8: “Similarly, the individual characters of El Niño events, which are very short in duration, become lost in the bioturbated sediment record “ The purpose of IFA is not to discern the properties of an individual event. Change in frequency or amplitude of events over a period of time can be statistically detected using various means to compare the distribution of integrated conditions over the period of sedimentation. Bioturbation serves, then, to extend that integrated time and the range of conditions experienced.

Bioturbation will also not remove anomalous values (page 9) – rather, such values may be present as part of a distribution representing more integrated time. Likewise, bioturbation has the effect of smoothing the signal, but the “signal” is a function of all sources of variability (ENSO, annual, decadal, centennial). The relative expression of these forms of variability along with the amount of time integrated by a sample are both

[Printer-friendly version](#)[Discussion paper](#)

important in terms of the ability to capture ENSO signals.

On P.10, Cole and Tudhope (corals) and White et al (IFA) are cited in error when discussing lake colour intensity and precipitation-driven records.

Also on P10, the authors claim: “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.” – Which methods? Q-Q, std. dev, event counting, others? It’s not entirely clear this is even referring to IFA reconstructions, as the records discussed previous are sedimentary, coral, and IFA (but noted as “precipitation driven”, see above).

P.10: 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. This statement leaves out other key elements, including the relative expression of ENSO events, the seasonal cycle, and decade-and-longer variability. These elements are (arguably) more important for inverse modeling, where the ability to disentangle growth rates from other sources of variability is impossible, and thus the signatures of ENSO in such records need to be discerned.

A key point in the paper (P10) says “The results presented here imply that much of the Pacific Ocean is not suitable for reconstructing ENSO studies using paleoceanography, yet several studies have exposed shifts within std dev( $\delta^{18}O_c$ ) of surface and thermocline dwelling foraminifera. One can, therefore, question what is being reconstructed in such studies.”. This study has, at this point, not tested whether the Std.dev of  $\delta^{18}O_c$  from individual foraminifera have reconstructed ENSO (also, the wording of this sentence is odd).

The first paragraph of the discussion (p9) purports to be about paleoclimatological archives that “have been used to indirectly and directly study past ENSO”. However, the



discussion is on mean-state reconstructions (Koutavas 2003, Dubois 2009). Koutavas 2003 is non-IFA mean-state reconstruction; likewise, the Dubois 2009 paper notes that “we prefer not to invoke any ENSO-like state for the glacial EEP based solely on our UK’37 SST.” While it may be true that this result and Koutavas 2003 are at odds, this is not an issue of IFA or ENSO reconstruction, but rather aggregate analysis and mean -state reconstruction. Discussion of std.dev ENSO studies (modeled by Thirumalai, Koutavas 2006, Koutavas and Joanides 2012, Leduc 2009, Sadekov 2013, Rustic 2015) is not found, yet the following paragraph (see above) is largely about this approach. Further, significant discussion and analysis of IFA reconstructions of ENSO during the LGM is found in Ford 2015, which is not discussed here.

The main analysis uses an unrealistic mixed-layer depth of 400m for the models foraminifera. Symbiont-bearing forams (*G. ruber* and *G. sacculifer*) live in the photic zone, and thus modeling and analysis of these organisms should be constrained to these depths. The model results using the shallower depths and specific, photic zone depths (Figure 4, figure 5, Figure 6) show that much of the tropical Pacific is suitable for such analyses, provided adequate carbonate preservation. This is very much in contrast with the point made previously in the paper that much of the tropical Pacific is unsuitable. In these figures, confusingly, some figures show significant areas in white while others use gray for no discernable reason. The figures are also improperly labeled, according to the captions – in each figure, *G. sacculifer* is on the left, *G. ruber* is in the middle, and this is reversed in the caption. Which is which?

The conclusions are at odds with what is presented at various points in the paper. Specifically: “Overall, our results suggest that foraminiferal  $\delta^{18}\text{O}$  for a large part of the Pacific Ocean can be used to reconstruct ENSO, especially in an individual foraminifera Analysis approach is used, contrary to previous analysis (Thirumalai et al. 2013). This conclusion is contradicted in the abstract, and in various parts of the study (e.g., P10 – “the results presented here imply that much of the Pacific Ocean is not suitable for reconstruction ENSO studies with paleoceanography. . .”) Which is it?

[Printer-friendly version](#)[Discussion paper](#)

Again, Koutavas 2003 is cited here, but that is not an IFA study. In general, clearly noting which studies are IFA/ENSO and which are mean state / aggregate / non-IFA studies will clarify the discussion surrounding the use of IFA and IFA techniques to identify ENSO signals.

This study does not directly address the Thirumalai 2013 study, as presented. The role of seasonality does not appear to be well addressed in this study (a key factor of Thirumalai 2013), the questionable definition of ENSO events confounds direct comparison, and the lack of clarity on sampling rates and other facts precludes a direct comparison. If this was a goal in this analysis, the Thirumalai study should be discussed in detail at the beginning (and should be discussed, in any case, earlier when discussing approaches for quantifying the suitability of locations for ENSO reconstruction), and the differences between their approaches (e.g., forward vs. inverse modeling). Suitable criteria for comparison should be noted (e.g., std. dev. Vs A-D tests).

---

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

Printer-friendly version

Discussion paper



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

**Kaustubh Thirumalai**

kaustubh\_thirumalai@brown.edu

Received and published: 30 April 2019

I have not been assigned to be a reviewer on this manuscript, yet a previously published paper authored by my colleagues and I entitled "Statistical constraints on El Niño Southern Oscillation reconstructions using individual foraminifera: A sensitivity analysis and me." (Thirumalai et al. 2013) is quite pertinent to this study. I wanted to point out some flawed rationale in this discussion paper especially considering the explicit lack of utilizing subsampling in their arguments concerning ENSO skill, seasonality, and individual foraminiferal reconstructions. I hope the authors are open to my comments and I would like to state that I am a big proponent of studies such as this one and that I support the FAMES approach.

[Printer-friendly version](#)

[Discussion paper](#)



- The authors' conception of reconstructing ENSO variability versus reconstructing mean state conditions in areas influenced by ENSO is critically flawed, especially with regards to commenting on and inferring changes in variability using individual foraminiferal analysis (IFA). Regardless of using Anderson-Darling statistical tests to assess whether subsampling occurs from their forward-modeled  $\delta^{18}\text{O}$  and temperature histograms, their analyses completely discount that ENSO events are seasonally synchronized to the annual cycle, and much of the variance of subsampled IFA distributions across the Pacific Ocean represents this power. Foraminiferal reconstructions are sensitive to absolute temperatures and NOT to monthly anomalies of temperature. This has been clearly demonstrated previously (e.g., see Thirumalai et al. 2013) and also underpins that discussion of IFA is incomplete without discussing uncertainties in sampling (e.g., White et al. 2018). Thus, their arguments and results (as well as the abstract) need to be significantly revised with this in mind. If the authors want to comment on ENSO and foraminifera (as in their title), they MUST incorporate subsampling uncertainties and how this interacts with the seasonal cycle.

- By definition, the authors state that "FAME uses the associated temperature and  $\delta^{18}\text{O}_{\text{eq}}$  at each grid cell to compute a time averaged  $\delta^{18}\text{O}_{\text{c}}$  and  $T_{\text{c}}$  for a given species". In other words, the authors have shown in their analyses that ENSO events strongly alter the temperature and  $\delta^{18}\text{O}$  in much of the tropical Pacific and that foraminiferal histograms (or foraminiferal distributions) are able to capture mean state conditions by sampling from these altered distributions (i.e., based on the utility of the Anderson-Darling test to account for histogram subsampling). Both of these aspects are well known. This, by no means, demonstrates a calculation of skill or validity of IFA-based ENSO reconstructions. Thus the title of the manuscript is inaccurate and misleading. For a demonstration of calculating ENSO skill in reconstructions, please read the literature: Carré et al. 2013, Emile-Geay and Tingly, 2017, Ford et al., 2014, Hereid et al. 2013, Khider et al. 2011, Tindall et al. 2017, Thirumalai et al. 2013, and so forth (only two of which are cited and not discussed). If this paper is proposed for revisions in this journal, I would strongly contend that the title of this paper should be

[Printer-friendly version](#)[Discussion paper](#)

revised in addition to the recalculation or revision of their text wherein IFA-based ENSO skill is referred to (as opposed to mean state conditions that are influenced by ENSO.)

- The manuscript mischaracterizes Thirumalai et al. 2013 and does not refer to the advances contained therein appropriately. As one example, the authors write in their conclusions: "Overall, our results suggest that foraminiferal  $\delta^{18}\text{O}$  for a large part of the Pacific Ocean can be used to reconstruct ENSO, especially if an individual foraminiferal analysis (Lougheed et al., 2018; Wit et al., 2013) approach is used (Ford et al., 2015; Koutavas et al., 2006; Koutavas and Joanides, 2012; Koutavas and Lynch-Stieglitz, 2003; Sadekov et al., 2013; White et al., 2018), contrary to previous analysis (Thirumalai et al., 2013). " Firstly, considering that their analyses do not account for sampling uncertainty or ENSO skill in IFA-based reconstructions due to the lack of separation from the seasonal cycle as well as decadal and higher forms variability, their conclusion is not supported by their findings. Second, we demonstrate that, in fact, ENSO sensitivity is high (with minimal influences from seasonality) according to forward-modeled IFA (with uncertainty) in the subsurface Eastern equatorial Pacific and the surface-ocean in the central tropical Pacific Ocean (see Figs. 5-6 as well as Discussion). From the conclusions of Thirumalai et al. 2013: "Our results show that the IFA approach is insensitive to ENSO frequency changes (<20% probability) but nevertheless indicate that changes in ENSO amplitude or seasonal cycle amplitude (or a combination of both) can be detected depending on the ratio of interannual-to-annual variability at the location of the study."

#### References:

Carré, M., J. P. Sachs, A. J. Schauer, W. E. Rodríguez, and F. C. Ramos. 2013. Reconstructing El Niño-Southern Oscillation activity and ocean temperature seasonality from short-lived marine mollusk shells from Peru. *Palaeogeography, Palaeoclimatology, Palaeoecology* 371: 45-53.

Emile-Geay, J., and M. P. Tingley. 2015. Inferring climate variability from nonlinear

[Printer-friendly version](#)[Discussion paper](#)

proxies: application to paleo-ENSO studies. *Climate of the Past* 11: 2763-2809.

Ford, H. L., A. C. Ravelo, and P. J. Polissar. 2015. Reduced El Niño–Southern Oscillation during the Last Glacial Maximum. *Science* 347: 255-258.

Hereid, K. A., T. M. Quinn, and Y. M. Okumura. 2013. Assessing spatial variability in El Niño–Southern Oscillation event detection skill using coral geochemistry. *Paleoceanography* 28: 14-23.

Khider, D., L. D. Stott, J. Emile-Geay, R. C. Thunell, and D. E. Hammond. 2011. Assessing El Niño Southern Oscillation variability during the past millennium. *Paleoceanography* 26: PA3222.

Tindall, J. C., A. M. Haywood, and K. Thirumalai. 2017. Modeling the stable water isotope expression of El Niño in the Pliocene: Implications for the interpretation of proxy data. *Paleoceanography* 19: 191-122.

Thirumalai, K., J. W. Partin, C. S. Jackson, and T. M. Quinn. 2013. Statistical constraints on El Niño Southern Oscillation reconstructions using individual foraminifera: A sensitivity analysis. *Paleoceanography* 28: 401-412.

White, S. M., A. C. Ravelo, and P. J. Polissar. 2018. Dampened El Niño in the Early and Mid-Holocene Due To Insolation-Forced Warming/Deepening of the Thermocline. *Geophysical Research Letters* 45: 316-326.

---

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

Printer-friendly version

Discussion paper



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

**Brett Metcalfe et al.**

b.metcalfe@vu.nl

Received and published: 8 May 2019

Dear Dr. Kaustubh Thirumalai, Thank you for taking the time to make known your opinion of our paper and for volunteering your comments. We addressed many of these issues in the reply to reviewer 1, reply to reviewer 2, reply to reviewer 3 and the general reply to all reviewers. However, we will gladly address your points here. We must stress, as we have done so in the previous replies, that we are modelling foraminifera populations in the water, and not in the sediment archive.

COMMENT: “Regardless of using Anderson-Darling statistical tests to assess whether subsampling occurs from their forward-modeled  $\delta^{18}\text{O}$  and temperature histograms, their analyses completely discount that ENSO events are seasonally synchronized to

[Printer-friendly version](#)

[Discussion paper](#)



the annual cycle, and much of the variance of subsampled IFA distributions across the Pacific Ocean represents this power. Foraminiferal reconstructions are sensitive to absolute temperatures and NOT to monthly anomalies of temperature.”

- By forcing the model with 65 years of monthly ocean reanalysis temperature and salinity data we are not discounting seasonality (an odd suggestion), it is exactly why we chose such a dataset to run the analysis rather than producing a theoretical or hypothetical temperature distribution to which to produce FAME modulated signals. We produce 696 values (1 for each month) of  $\delta^{18}\text{O}$  and  $T_c$ , that are then pooled into their respective climatological states (using a modified definition). Naturally this has repercussions. We use absolute temperature (FAME requires absolute temperature, see Roche et al., 2018) and not monthly anomalies for calculation (ONI – the definition of events is however a monthly anomaly).

- If the suggestion from this comment is that we should be more flexible with our definition of El Nino, or how we pool such values, we refer to the arguments of reviewers’ 1-3 against such flexibility.

- We do not use “ Anderson-Darling statistical tests to assess whether subsampling occurs from their forward-modeled  $\delta^{18}\text{O}$  and temperature histograms” but to determine whether the different climate states could be considered statistically different (i.e., if foraminifera populations from event A and event B can be considered to be statistically different or not). The commentator’s interpretation of the Anderson-Darling test does not reflect what we did (or stated). If the commentator is suggesting that second order variability (short -term changes in the seasonal cycle) should be included, then it implicitly implies that foraminifera don’t record El Nino events.

COMMENT: “If the authors want to comment on ENSO and foraminifera (as in their title), they MUST incorporate subsampling uncertainties and how this interacts with the seasonal cycle.”

- As we made clear in the reply to reviewer 1, reply to reviewer 2, reply to reviewer

[Printer-friendly version](#)[Discussion paper](#)



3 and our general reply to all reviewers, we are using FAME, driven by six decades of monthly observed climate data, to model foraminifera populations in the water. We are not modelling foraminifera retrieved from the sediment archive, so we therefore do not model processes such as bioturbation, fabric disturbance during core retrieval, species misidentification, machine error, subsampling-induced noise, sample vial mislabelling, or indeed anything else pertaining to the generation of palaeodata in a practical sense. We seek to understand whether or not foraminifera populations in the water are intrinsically capable of recording ENSO dynamics, and at which locations. We consider this to be the most fundamental consideration for foraminifera-based ENSO studies, and therefore, we fail to see how our title is inappropriate. Assessment of sediment/laboratory processes essentially becomes a moot point if the events to be reconstructed are not recorded by foraminifera in the water.

- Further studies should address whether or not the signal produced is an artefact of subsampling (i.e., that the signal is introduced because of picking and not a real climatological signal) – a point we suggest in the text, that should definitely be of concern for palaeoreconstructions, as the commentator is aware of.

COMMENT: “By definition, the authors state that "FAME uses the associated temperature and  $\delta^{18}\text{O}_{\text{eq}}$  at each grid cell to compute a time averaged  $\delta^{18}\text{O}_{\text{c}}$  and  $T_{\text{c}}$  for a given species". In other words, the authors have shown in their analyses that ENSO events strongly alter the temperature and  $\delta^{18}\text{O}$  in much of the tropical Pacific and that foraminiferal histograms (or foraminiferal distributions) are able to capture mean state conditions by sampling from these altered distributions (i.e., based on the utility of the Anderson-Darling test to account for histogram subsampling). Both of these aspects are well known. This, by no means, demonstrates a calculation of skill or validity of IFA-based ENSO reconstructions. Thus, the title of the manuscript is inaccurate and misleading.”

- We were testing whether it is possible for foraminifera in the water to record ENSO. The skill or validity of IFA-based ENSO reconstructions has been studied before (we

[Printer-friendly version](#)[Discussion paper](#)

have made this point previously to more than one reviewer). Here we are testing whether (in a perfect world) the foraminifera population in the water detect ENSO dynamics.

- 'Thus the title of the manuscript is inaccurate and misleading' – we broke down the title in our response to reviewer 3 as follows: If we break down the title: Validity – is the quality or state of something being valid (valid - being logically correct or well-grounded/justifiable). We are testing whether the distributions of different climate events are statistically different (this is a fundamental test) Foraminifera-based – we use a foraminiferal model. An alternative could include the word “populations” here, as in “foraminiferal populations”. ENSO reconstructions - The reviewer, we assume, is arguing that as we say ENSO reconstructions, and judging from their previous point (“as it focuses on determining whether ENSO events (El Niño, La Niña) and neutral conditions have distinct distributions (forward modeling) rather than on how one could detect ENSO change (inverse modeling).”) that we haven't focused on the 'how one could detect', yet we are testing whether the foraminiferal distributions of different climate events are statistically different, hence we have carried out a test on a more fundamental level (i.e. foraminifera in the water, before they are incorporated in the sediment archive)." We do not think that we are being misleading or inaccurate.

COMMENT: “In other words, the authors have shown in their analyses that ENSO events strongly alter the temperature and  $\delta^{18}\text{O}$  in much of the tropical Pacific and that foraminiferal histograms (or foraminiferal distributions) are able to capture mean state conditions by sampling from these altered distributions (i.e., based on the utility of the Anderson-Darling test to account for histogram subsampling).”

-The first half of this is correct, the signal will be modulated in response to water column conditions (and potential depth habitat or growth related changes). However, we are not referring to mean state conditions, we are testing whether the El Niño foraminifera population is statistically different from the foraminifera population for other climate states. The comment keeps referring to 'histogram subsampling' which is not

[Printer-friendly version](#)[Discussion paper](#)

at Anderson-Darling test does. The A-D test compares two probability distribution functions for (dis)similarity.

COMMENT: “For a demonstration of calculating ENSO skill in reconstructions, please read the literature: Carré et al. 2013, Emile-Geay and Tingly, 2017, Ford et al., 2014, Hereid et al. 2013, Khider et al. 2011, Tindall et al. 2017, Thirumalai et al. 2013, and so forth (only two of which are cited and not discussed).”

- We will endeavour to incorporate those papers we feel match our discussion, however, as pointed out to Reviewer 3, we did mention that our ‘validity’ is referring to a fundamental validity of foraminifera populations in the water to record ENSO dynamics, and not a test of the skill of sediment-based reconstructions.

COMMENT: “The manuscript mischaracterizes Thirumalai et al. 2013 and does not refer to the advances contained therein appropriately” - We have already mentioned rephrasing this section in the replies. We apologise for mischaracterisation.

COMMENT: “ Firstly, considering that their analyses do not account for sampling uncertainty or ENSO skill in IFA-based reconstructions due to the lack of separation from the seasonal cycle as well as decadal and higher forms variability, their conclusion is not supported by their findings.”

- Although our results are condensed into a simple binary 0 and 1, that is an aid to the reader, such simplicity does not reflect a lack of complexity. As we point out above, we have utilised over six decades of monthly ocean reanalysis data, which means that for every grid cell we produce 696 monthly values and divided them into their various climatological states. That means the low and high orders of variability are intrinsically included in our approach. However, foraminifera are a filter of the signal, which means that much of this variability will be altered (e.g., changes in flux, the growth weighting or even the depth habitat). This is why we decided to use a simple foraminiferal module to produce our time series.

[Printer-friendly version](#)[Discussion paper](#)

-Separating out seasonal and annual values would be unrealistic, as foraminifera are pooled in the sediment, therefore it would be a falsifiable test. Separation also implies that it is possible to reconstruct seasonal or decadal time series within the sediment, which is not possible.

Finally, we would like to raise an important point, reviewer 1, 2 and 3 have stressed repeatedly that our study fails to address the inverse problem instead using a forward model, to which we have in our replies to reviewers 1, 2 and 3 repeatedly had to discuss inverse modelling. But it is worth noting that here you state ("according to forward-modeled IFA (with uncertainty)" ) that your previous research (Thirumalai et al., 2013) is also in part based upon a forward model - which suggests the reviewers have not appreciated this fact.

Kind regards,

dr. Brett Metcalfe

---

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

Printer-friendly version

Discussion paper

