Author changes and response point by point round 2

Referee comment in black

Author answer in blue

Author Changes in red

Anonymous Referee #1

The revised manuscript addresses all of my comments and provides adequate explanation of the methodology, including sensitivity analyses of different parameters, such as the use of different climate reanalysis datasets, mixed-layer depth, thermocline depth, and the limitations associated with the steps used in the forward modelling. The authors mention that ENSO and IOD events recorded in the unperturbed ORAS5 reanalysis show increased amplitude and frequency variability, both in SST and thermocline temperature (Line no.: 261-263; 578 – 580). The addition of a supplementary figure showing increased variability of SST and thermocline temperature for all four sites during ENSO and IOD periods, based on the ORAS5 reanalysis, would further substantiate these statements in the manuscript. Overall, the manuscript has addressed my concerns, and I recommend the manuscript for publication.

We thank the referee for this suggestion and for the positive assessment of the revised manuscript. We have added this new supplementary figure to complete our argumentation of our 75 % sensitivity test in the variability. This figure shows for all four sites, the interannual SST and thermocline temperature anomalies together with ENSO (Oceanic Niño Index) and IOD (Dipole Mode Index) event periods over 1958–2018. It confirms that the strongest thermocline temperature changes at each site (ranging from 2.78°C at MD96-2060 to 7.11°C at SO189-39KL) coincide with independently identified ENSO or IOD events, which is consistent with our statement.

Author's changes in the manuscript: We have added it as a new supplementary Figure S6. We put the call in the text and updated the values from 2.69 °C - 6.58 °C to 2.78°C - 7.11°C **in section 3.1.2, line 278.**

Referee #2 Casey Saenger

In their manuscript "Assessing the potential of Individual Foraminifera Analyses (δ18O) to reconstruct variability, seasonality and extremes in the tropical Indian Ocean" Lichterfeld and co-authors present a comparison of core-top individual foraminifera oxygen isotope data to instrumental data. Using four sites with variable seasonal forcing and two species that ostensibly live at different depths, the authors are able to compare observations and data across a fairly wide range. I am highly supportive of the large scale goals of this research, and believe individual foraminifera analysis is an exciting new approach for considering climate extremes. I also think the community needs to invest more energy in the modern calibration of IFA techniques to best interpret down core variations in distributions. The core-top data in this manuscript has some potential to do that, but unfortunately I think falls far short of that potential.

We thank the referee for this supportive assessment of the study and for his time invested in this detailed review. We answer every point below individually. We would like to note that, because of the timing issue described by the editor, this review was conducted and refers mainly to the original submitted version of the manuscript, before the major revisions addressed in response to both referees in the first round of review. Several of the methodological points described below (reanalysis validation, depth-habitat treatment, paleotemperature equations, flux-weighting sensitivity, uncertainty propagation) have already been addressed in the current revised version. For the remaining points, we either provide additional analyses and manuscript changes, or a detailed rebuttal.

My major criticism of this work is that the forward modelling of foraminifera oxygen isotopes (some would call it a proxy system model) is overly simplistic and the statistical comparisons between measured and modeled data is insufficient. As the authors note in their discussion, the ecology of G. ruber and N. dutertrei is not simple. Even if the species are present year round, they might have higher or lower abundances in certain months that produce a seasonal bias in the environmental conditions they record. By not considering any seasonality and assuming that all months are equally likely for forams to live, it seems like the authors effectively do a lot of Monte Carlo and bootstrapping to generate modeled oxygen isotope distributions that could be calculated algebraically since they are just linear transformations of the observed SST and SSS using a single Bemis relationship. This approach is presented as conservative ("avoiding any assumption of a seasonally phase-locked life cycle" lines 205-206), but then is cited as a potential complication ("the seasonal timing of shell production and the ecological preferences of each species should be considered" line 586). Since we have some constraints on seasonality, I would like to see the authors consider it more realistically in their modeling from the start. At risk of being the reviewer that mentions his own work, I dealt with this is Saenger et al. (2024, https://doi.org/10.3389/fmars.2024.1321254) by using PLAFOM2.0 (Kretschmer et al., 2018; https://doi.org/10.5194/bg-15-4405-2018) to weight timeseries based on foram seasonality. All months can still be selected in the Monte Carlo, but the months when forams are likely to be more abundant are preferentially selected. Other estimates of seasonality like the Lombard et al. (2011) paper already cited seem reasonable, but there should be some consideration of seasonality.

We thank the referee for this detailed methodological point on the forward-modelled framework. We first clarify an apparent inconsistency that the referee identified between lines 205–206 and line 586. These two statements refer to two distinct assumptions, which we have now clarified in the text. The framing at lines 205–206 concerns only the absence of a fixed phase-locked calcification month for individual foraminifera, so we do not assume that every individual calcifies within the same calendar month across years. This differs from the point raised at line 586, which concerns a seasonal bias in production abundance, which is exactly the point the referee raises below, and which we now address with two independent seasonality-weighting tests.

Regarding the forward-modelling, we note the equations used are no longer a single Bemis et al., (1998) relationship. Since the last version of the revised manuscript, species-specific paleotemperature equations (Farmer et al., 2007) are used separately for *G. ruber* and *N. dutertrei* (Eq. 1–2). Moreover, for the Monte Carlo iteration the paleotemperature and δ^{18}Ow–S equation coefficients are themselves resampled from their reported uncertainty distributions, propagating environmental and calibration-equation uncertainty which is not reducible to a single algebraic linear transformation.

About the seasonality, we agree that this is an important consideration, we have tested it in two ways. The first one, refers to the previous manuscript version (Section 5.3.2, Fig. S12), where we performed a flux-weighting sensitivity test using local FORCIS sediment-trap data (Chaabane et al., 2023). This changes the simulated IFA SD by only −0.037 to +0.039‰ across sites and species, with no change to the site ranking or conclusions. Then, following the referee's suggestion, we additionally use PLAFOM2.0 (Kretschmer et al., 2018) as done in Saenger et al. (2024) for monthly *G. ruber* production estimates. PLAFOM2.0 does not simulate *N. dutertrei* so this test is restricted to the surface species.

Interestingly, PLAFOM2.0 and FORCIS disagree on which months are favoured for *G. ruber* production at three of the four sites, reflecting that a mechanistic global ecosystem model and a local empirical flux measurement capture conceptually different quantities (modelled standing concentration vs. measured exported flux). Despite this disagreement in the underlying monthly pattern, PLAFOM2.0 weighting changes simulated *G. ruber* IFA SD by only 0.002–0.028 ‰, comparable to the FORCIS-based weighting, with no change in the site ranking. This convergence between two independent - and mutually disagreeing - seasonality estimates on a negligible net effect gives us confidence that our conclusions are not sensitive to the choice of seasonal production weighting.

We therefore retain uniform Monte Carlo sampling as justified in the previous revised manuscript and now further supported by this second sensitivity test. Also, the reference to Lombard et al. (2011), is already cited in the manuscript as a source of seasonal growth-rate estimates (Section 5.3.2). Finally, regarding the referee's broader concern that "statistical comparisons between measured and modelled data is insufficient": we address this point specifically and quantitatively in our responses to the related specific comments below.

We have also updated our forward-modelling framework (publicly available on GitHub) to include PLAFOM2.0-based flux weighting as an additional option alongside the existing FORCIS-based one, for transparency and reproducibility, and to allow this refinement to be readily applied in future downcore applications of the method.

Author's changes in the manuscript: We clarified the distinction between the phase locking assumption and the seasonality assumption in Section 3.1.2. We also added the citation of Saenger et al. (2024) and Kretschmer et al. (2018), and a paragraph in **Section 5.3.2 lines 874-881** reporting the PLAFOM2.0-based cross-check alongside the existing FORCIS-based sensitivity test and a new supplementary figure (Fig. S12). We finally updated the forward-modelling framework with PLAFOM2.0 weighting as a user-selectable option.

Similarly simplistic is this manuscript's consideration of depth habitat. G. ruber is assumed to live right at the ocean surface (0.5 m) while N. dutertrei lives only at 97m. This ignores considerable work that has been done to estimate foram depth habitat. For example, while G. ruber is often considered a "surface" dweller, Stainbank et al. (2019; doi: 10.1371/journal.pone.0222299) suggests a depth of 60-70m in the Maldives that is not very different from the 80-100 m the suggest for N. dutertrei. The authors' modeled pseudoforams should use more realistic depth ranges, and randomly sample across a range of depths consistent with prior work within the reanalysis data. This would generate a more realistic comparison between modeled and observed foram oxygen isotopes and I expect would alter some of the manuscript's conclusions.

We thank the referee for raising this point. We note an important nuance in the reference about the ACDs. They report isotope ACDs of 77±4 m for *G. ruber* and 97±4 m for *N. dutertrei*, but attribute this depth deeper than the ~30 m classically assigned to *G. ruber* to the fact that "geochemical signatures are weighted by the final few precipitated chambers, which generally form when they are deeper in the water column". We agree this citation should be qualified to reflect this ACD-specific caveat, and have revised the text accordingly.

We re-emphasize, as already pointed by Referee 1 during the first round of review, that the integrated layers (0–50 m for *G. ruber*, 70–125 m for *N. dutertrei*), better reflect the vertical range over which these species are known to calcify. We made a sensitivity test to compare fixed-depth versus layer-integrated extraction on simulated IFA SD. The sensitivity test leads to a change in the estimated standard deviation was < 0.065‰ at three of four sites (Section 3.1.1, Supplementary Fig. S2). Notably, at the fourth site, SO189-39KL, we specifically retained the fixed 0.5 m depth rather than the layer-integrated average, because layer-averaging was shown to dilute the near-surface, upwelling-driven interannual extremes that are central to this site's signal. We have also now discussed at length the FORCIS-based and literature-based depth-habitat constraints for both species in Section 5.3.2 according the the former review of Referee 1, including the ~40 m average occurrence depth for *N. dutertrei* from Chaabane et al. (2026), and explicitly acknowledged the uncertainty inherent to assigning a single calcification depth to species with broad or variable vertical ranges.

Then, following the referee's suggestion, we also implemented and tested the probabilistic depth-sampling approach requested, in which each simulated individual now draws its own calcification depth with temperature and salinity interpolated from the full ORAS5 vertical profile. Unlike the fixed-versus-layer test above, this comparison reveals a substantial effect: it changes simulated IFA SD by +8 to +55 % for six of the eight site/species combinations, by +159 % for *G. ruber* at SO189-39KL, and by −36 % for *G. ruber* at MD96-2060 (paired Wilcoxon test, 1000 Monte Carlo repetitions, $p < 0.001$ in all cases), and largely reverses the *G. ruber* site ranking. We consider this as an important finding, unlike the fixed-versus-layer test, the choice of depth-sampling method is not a negligible methodological detail here, and we do not want to understate it. Especially, looking at the ranking of the site by SD, which remains largely the same for *N. dutertrei* but changes substantially for *G. ruber*, with SO189-39KL moving from the lowest to the highest SD and MD96-2060 from the second highest to the lowest.

We would like to remind that our forward-modelling framework was deliberately designed as a simple tool to interpret the Indian Ocean IFA results potentially contained downcore, and that the SO189-39KL case described above already showed that depth assignment can matter at individual sites. Importantly, the measured core-tops do not support these deeper calcification depths: depth sampling shifts the simulated G. ruber means by +0.43 to +0.87 ‰, several times the offset observed between synthetic and measured core-tops (+0.08 to +0.36 ‰), and shifts N. dutertrei means in the opposite direction to what the measurements require. As in our first-round reply to Referee 1, we retain the depth-integrated layer approach (which was already an improvement over the originally reviewed fixed-depth version, and already justified via the sensitivity tests above), as our working framework for

this manuscript, and explicitly flag probabilistic depth sampling as a priority for future applications of this forward-modelling framework.

Author's changes in the manuscript: We added a paragraph in **Section 5.3.2 in lines 882-890**, reporting a new sensitivity test (Fig. S13) comparing the depth-integrated layer approach used in this study with a probabilistic, individual-level depth-sampling scheme based on the apparent calcification depths of Stainbank et al. (2019), applied to all sites. This paragraph reports the resulting change in simulated IFA SD for each site and species, and its effect on the *G. ruber* site ranking. We also identify improved depth sampling as a priority for future developments of the forward-modelling framework.

Continuing with my concerns about the forward foram model, I would appreciate some sensitivity testing of the relationship used to estimate foram δ18O from SST and SSS. There are many choices here at it is not obvious that the Bemis et al. (1998) equation is appropriate. For example, Bemis is based on G. bulloides and O. universa, neither of which are used in this study. This is made worse by the fact that the actual Bemis et al. equation used isn't provided and there is inconsistency about whether the authors use a high or low light relationship. The authors should evaluate relationships based on G. ruber and N. dutertrei, for which papers like Multiza et al. (2003; https://doi.org/10.1016/S0031-0182(03)00633-3) and Wejnert et al. (2013; 10.1016/j.marmicro.2013.03.001) may be helpful. While the temperature dependence (slope) of oxygen isotope fractionation between foram and seawater is usually fairly constant, there are differences in intercept that might explain the systematic offset the authors observe.

We agree that the choice of equation may contribute to the offset and have tested it. To answer this point, we already ran a sensitivity test comparing our currently-used species-specific equations (Farmer et al., 2007) from the last revised version of the manuscript against Mulitza et al. (2003) for *G. ruber*, and against the only equation developed specifically for *N. dutertrei* from Bouvier-Soumagnac & Duplessy (1985), as reported in Wejnert et al. (2013).

For *G. ruber*, the two equations have a comparable slope (Farmer: −4.78±0.27; Mulitza: −4.44±0.23) and (15.4 vs. 14.2 °C), and switching to Mulitza et al. (2003) shifts the simulated mean by a consistent −0.43 to −0.49 ‰ across all four sites, a magnitude comparable to the systematic offset already reported between synthetic and measured core-tops. We emphasize that such shift moves the synthetic distribution away from the real measured core-top value. We find no evidence that adopting Mulitza et al. (2003) would resolve the observed offset, it shows that it will approximately double it. Moreover, the effect on the simulated SD is very small in absolute terms (+0.006 to +0.029‰, i.e. 2–7% relative to the Farmer-based SD), though statistically significant (paired Wilcoxon, $p<0.001$ at all sites), consistent with our earlier Bemis-vs-Farmer comparison (Fig. S10).

For *N. dutertrei*, we note that Wejnert et al. (2013) found that the only equation developed specifically for this species (Bouvier-Soumagnac & Duplessy, 1985) yields unrealistically deep calcification depths (average ~165 m), and that generic equations calibrated on other species perform better for *N. dutertrei*. Its very large reported parameter uncertainty (±12.1 °C on the intercept, ±2.64 on the slope) reflects this limited calibration. Using it in our framework shifts the simulated mean by −0.24 to −0.29 ‰, again increasing the offset with the measured core-tops, and reduces simulated SD by 16–23 % relative to Farmer et al. (2007), as expected from its steeper slope.

Author's changes in the manuscript: We have updated a supplementary figure (Fig S10) reporting the new sensitivity test showing the simulated IFA SD for the reference of the latest version of the manuscript (Farmer et al., 2007) and alternative equations from (Mulitza et al., 2003 for *G. ruber*; Bouvier-Soumagnac & Duplessy, 1985 for *N. dutertrei*). We also test the influence of the mean shift and have added a paragraph **in Section 5.3.1, lines 856-861** which further supports our retention of Farmer et al. (2007) for both species.

The handling of the measured data also seems problematic and may be causing misinterpretation when comparing measured and modeled foraminifera δ18O. First, measuring individual foraminifera is not trivial and one cannot simply pop a 4 microgram foraminifera into a Kiel device and expect good data. Loading small volumes of reference gas

and making sure sample and reference gas remain balanced can be a challenge, yet all of the analytical methods of this paper are described in two sentences (Lines 235-239)! The authors should significantly elaborate on their method and include the masses and counts of standards used to determine precisions. The masses of standards should be small and consistent with the masses of the foraminifera samples. The resulting uncertainty should be incorporated into the forward foram model to produce more of an apples-apples comparison.

The mass spectrometer sensitivity is monitored daily with measurements of 14 NBS-19 international standards, that alternate between the samples for a batch of 32 samples in total. The isotopic standard deviations are then monitored on each batch using this protocol. For IFA, we measure isotopic ratios on foraminifera with weight between 15 and 20 micrograms (up to 40 µg for thermocline species). We also pay attention to weigh small quantities standards, usually between 30-50 µg. The standard deviation on multiple specimens using weights between 30 and 70 micrograms is 0.08 permil, as reported in the methods. Also, the lower limit of the mass 44 signal is monitored, and samples producing less than 1500 mV signal are systematically removed to keep only samples producing large enough CO2 quantities. Aside monitoring for good vacuum prior to each analysis, data screening includes the checking of the standard deviation of the dual inlet measuring cycle of the $\delta^{18}O$ (that is measured 6 times per sample), and data are taken exclusively for samples having an internal standard deviation lower than 0.05, ensuring the integrity of the data quality during analysis.

Author's changes in the manuscript: Added methodological description in **section 3.2, lines 327 and 331**

Furthermore, the authors' removal of data that fall outside 95% (Line 493) seems problematic and poorly justified. When measuring bulk, homogenized foraminifera with the goal of identifying a mean, this approach may be justified. However, when considering distributions these values are to be expected and (if you are confident in your analytical approach) are the extreme geochemical values that might be telling you about extreme climate events. Indeed, having 5 analyses out of 139 that fall outside 95% is about what you should expect, and unless there is a clear analytical justification for removing these values they should stay. After all, the authors did not remove the SST and SSS data that fell outside 95% and thus have unnecessarily created an apples-oranges comparison.

We thank the referee for raising this point, it touches on a burning question that has been debated in the IFA community for over 15 years. For example, in Ganssen et al., 2011, one of us reported the same issue in the original publication of the round of review of Ganssen (Guillaume Leduc, review at https://cp.copernicus.org/preprints/6/C1298/2011/cpd-6-C1298-2011.pdf). The reasoning developed that we revisit here is that an outlier detection aims to robustly estimate the range of a distribution rather than the single maximum-minimum difference, which is highly sensitive to sample size and to potential anomalous measurement.

In our dataset, we checked the robustness of that method across all sites comparing the IQR method and a median-absolute-deviation-based modified z-score (threshold 3.5), which does not rely on quartile positions and is itself robust to the presence of extreme values. For the new data presented in this study, both methods identify exactly the same five outliers, indicating a robust feature of the data rather than an artefact of the 1.5 × IQR threshold.

We also question the plausibility of these five values by converting their deviations from each site mean into a temperature anomaly using the paleotemperature equation slope for each species. The most extreme case, a *N. dutertrei* at U1467, implies a thermocline temperature ~12 °C colder than the site mean, far outside the range of any documented ENSO/IOD event there (the largest interannual thermocline anomaly at this site being 4.61 °C in ORAS5 over 1958–2018). Even allowing for a slightly cooler Late Holocene mean state, we consider it very unlikely that this reflects a genuine environmental extreme, and retain its removal on this basis.

Author's changes in the manuscript: We have added a supplementary figure about both methods of the **outlier** detection check (IQR vs. modified z-score) and the temperature-equivalent and discussed it in the method **section 3.2 Lines 339-343**, by adding more justification (Fig. S9).

My last major concern with the forward modeled foraminfera is how casually it is compared to observations. The authors say that they would use many metrics including Shapiro-Wilk, skewness, kurtosis and quantile-quantile plots (Lines 239-241) to compare measured foram data to forward-modelled data, but the only comparison actually done is using standard deviation. Even this comparison is qualitative. Values of 0.34‰ and 0.28‰ are described as "statistically similar" but a value of 0.21‰ differs significantly (lines 440-441). The authors must state what their measure of significance is and rely on these measures rather than visual comparison. They should also discuss metrics other than just standard deviation and I encourage them to apply something like a two-sample Kolmogorov Smirnov test to see if measured and model distributions are significantly different.

We thank the referee for this methodological point. We hope anyway that we will convince the readers that the variability in IFA of G. ruber and N. dutertrei as tracked by the s.d. is helpful to extract seasonal and interannual variabilities, i.e. the initial scope of the paper. We hence don't want to further explore what extra metrics could bring since the article is now rather long. We now report this comparison with an explicit significance criterion (α=0.05) using a two-sample Kolmogorov-Smirnov test, as suggested by the referee, alongside Shapiro-Wilk normality tests for each distribution.

The KS test results show that measured and synthetic distributions differ significantly ($p<0.05$) at 5 of 7 site/species combinations tested (all except G. ruber at NIOP 905 and U1467). This significance is not always driven by dispersion (some SD are nearly identical, 0.333 ‰ vs. 0.353 ‰, and 0.213‰ vs. 0.187‰) but the test is also sensitive to the full shape of the distribution, including the mean offset already discussed in Section 5.3.1.

We have revised Section 5.3 to report these results of the different tests and update the previous terms used ("reasonable agreement", "statistically similar") based on the new results. We clarify that our SD-based comparison and interpretation specifically, concern the amplitude of variability which is a distinct question from whether the two distributions are identical in mean and shape

Author's changes in the manuscript: Section 5.3 revised to report Kolmogorov-Smirnov test results (statistic and p-value, α=0.05) and Shapiro-Wilk normality test results for each site/species comparison in Fig. 8, replacing the previous qualitative "reasonable agreement"/"statistically similar" language. Added a new supplementary figure (Fig. S9) summarizing these statistics for all site/species combinations.

I also think this manuscript is trying to make two different points that are only tangentially related. As discussed above, the first centers on comparing measured foraminifera data against pseudoforam data to evaluate if real forams can reproduce the expected variability. The other aspect is artificially manipulating observed timeseries to identify how these changes might be manifested in IFA studies. I think both could be expanded upon, and if done well, could be separate manuscripts. Alternatively, the connection between the two goals needs to be made clearer.

This point was also raised by Referee 1 in the first round of review (General Comment 3), and which we addressed at that stage by clarifying the two-step structure of the framework and adding a dedicated flowchart (Supplementary Fig. S4) in our latest version of the revised manuscript, illustrating the two distinct objectives and how they relate to one another.

To summarize what we argued, the step 1 is an idealized sensitivity analysis where we apply a uniform climate perturbation (±75%) across all four sites under identical parameters, designed specifically to compare the relative sensitivity of each site and species to seasonal versus interannual variability. The second step 2 is a realistic, site-specific model-data comparison, using regional δ^{18}Ow–S relationships and modern reanalysis data, designed to evaluate whether real core-top IFA measurements with what the forward model predicts with reanalysis. The connection between the two is that the step 1 establishes which sites and species are theoretically most diagnostic of seasonal or interannual change, providing the basis for interpreting the Step 2 model-data comparison and, ultimately, for selecting appropriate targets for future downcore (e.g. LGM) applications of this method.

We agree that either step could be expanded into two different studies. However, we think that presenting them together is useful and adapted., it allows the reader to see directly how a site's theoretical sensitivity (Step 1)

translates, or not, into an ability to detect that sensitivity in real sedimentary material (Step 2) which is precisely the kind of diagnostic the IFA community needs when selecting sites and species for future applications of this proxy.

Author's changes in the manuscript: No major changes as we made them in response to Referee 1 (General Comment 3). The framework and its rationale are described in the Introduction, formalized in Section 3.1.2, and illustrated in Supplementary Fig. S4 in the last version of the revised manuscript.

A concern I have related to the second goal stems from the simplicity of the foram forward model. Since forams are assumed to be equally likely in every month, what is presented as pseudoforam "Results" are really just ways of describing the various study sites (i.e. the magnitude of temperature and salinity variability at each site) or the "Method" of imposing 75% changes in variability. Perhaps I am missing something, but Figure 5 is an example; should it be surprising that imposing a 75% change results in a corresponding increase in δ18O? Figure 7 is another example; doesn't this just reflect the relative magnitude of temperature vs salinity variability at different sites? When temperature variability is larger, the 75% increase exerts a larger influence on absolute SST values relative to absolute SSS values, thereby steepening the slope. It seems like this outcome could be predicted from the reanalysis data without doing any pseudoforam modeling at all.

We thank the referee for this point. Yet we think it is necessary to translate the impact of temperature and salinity variability into the $\delta^{18}O$ modelled to better interpret the likely mechanism we may observe downcore. For example, if an upwelling region or a low-salinity, highly stratified region will have different $\delta^{18}O$, it is still important to interpret changes in $\delta^{18}O$ in terms of temperature or salinity if one wants to comment on climatic variability phenomena such as e.g. the monsoon (and its associated impact on rainfall and salinity of surface seawater) or the upwelling (and its associated impact on temperature of surface or subsurface waters) to make the IFA a useful climatic variability proxy, a transformation that is nonlinear, stochastic and therefore not analytically reducible from the reanalysis time series alone, as we have also discussed in our response to the referee's Major Comment 1 above.

The statistics applied on temperature and salinity data of reanalysis can predict which site should be more sensitive to one climatic parameter we would want to track. Regarding the Figure 7, the referee's description of what the QQ-plot slope reflects is precisely the diagnostic value of this approach. The QQ-plot slope allows us to directly compare scenario by scenario (temperature increase/decrease, salinity increase/decrease), which variable is driving $\delta^{18}Oc$ variability at a given site and depth in a way that is directly comparable across sites and depths using a single consistent metric.

Author's changes in the manuscript: We have added a clarifying sentence in Section 3.1.2, lines 270-272 and Section 5.2, line 701 to explain the utility and advantage of this sensitivity analysis is to quantify, in proxy ($\delta^{18}O$) space and with propagated uncertainty and that this quantitative translation is required as the baseline against which the Step 2 model-data comparison is evaluated.

Additionally, while the authors provide some rationale for their chosen 75% as mimicking IOD or ENSO regimes, I question whether they are physically realistic. An example of how this aspect of the study could be expanded would be to consider large ensembles of historical climate models and/or future projections run with different radiative forcings. This would result in a range of SST and SSS states that would be physically realistic and therefore more helpful in interpreting any paleo reconstruction.
On the topic of possible future paleo reconstructions, that is a place where the disconnect between the two components of this study is most obvious. Q-Q plots are helpful for comparing two populations such as the authors' original and perturbed states. But since there is only core-top data for the actual forams, there can be no equivalent comparison between core-top and some past interval. I think a logical way that the manipulated timeseries portion

of this work could be expanded would be to consider a wider range of perturbations, including sensitivity testing to other reanalysis products. Then present those expanded results in a future manuscript that also presents paleo data. In that way, you might identify which perturbations are most consistent with the paleo observations during different times at different sites.

We thank the referee for these two related points. On the physical realism of the ±75% perturbation, we agree that a large ensemble of climate model simulations would provide a valuable estimate of SST/SSS and it could be an interesting and useful direction for further work. We note however, that our current justification is not arbitrary, as we detailed in our response to Referee 1 (General Comment 3) and in the new supplementary figure prepared for the Round 2 comment from Referee 1, we quantified the peaks of ENSO/IOD event in the thermocline temperature anomalies directly from 60 years of ORAS5 reanalysis at each of the four sites (2.78–7.11°C, representing ~50–75% of the mean seasonal amplitude), and compared it to ENSO (ONI) and IOD (DMI) event records. We consider this an empirical estimate of realistic variability, more directly linked to the oceanographic regime of each site than a generic CMIP-type ensemble range would be.

We also note that the following applies regardless of which physical SST/SSS ensemble is used to define realistic perturbations. For example, a species with a fixed thermal ecological preference may track a preference across different mean climate states by shifting which season they are predominantly produced in. A warm species, recorded preferentially in winter under modern (already warm) conditions might instead be recording preferentially summers under a colder mean state such as the LGM if that is when temperatures reach its optimal range.

On the point of the two components of the study, regarding the future paleo application. We agree on this point and it is in fact what we are doing with a future manuscript which will present new data for the LGM. This future study is designed along the lines suggested by the referee, by testing the range of perturbation scenarios against paleo IFA observations. Overall, we view the present manuscript as establishing the baseline of the methodology site and species sensitivity, forward-model and data agreement under modern conditions and have clarified this intended trajectory in the Conclusion.

Author's changes in the manuscript: Added a precision in the **Conclusion, line 1054** with future perspectives and noting that a large-ensemble climate-model approach could be used to test a wider range of perturbations.

Specific comments:

Title: I would include the species considered

We prefer to keep the title concise; both species are named in the abstract.

Line 43: modes not mode

Done **line 46**

Line 48: Italicize in situ.

Done, line 51

Lines 31-48: For an introductory paragraph, I find this sprawling. Make sure it has a clear purpose and every sentence is written to achieve that purpose.

We had already revised this paragraph in the revised manuscript, trying to be clear.

Line 45: The authors should do sensitivity testing using other reanalysis products such as Hadley EN4 and SODA.

This validation was already requested by Referee 1 in the first round of review (General Comment 1) and has been done. We validated ORAS5 against HadISST1.1 (Rayner et al., 2003), GODAS (Behringer & Xue, 2004), and the EN4 instrumental dataset (Good et al., 2013) for surface and thermocline temperature and salinity at all four sites (Section 3.1.1, Supplementary Fig. S1) from the latest version. As explained in our first-round response, SODA could not be included because the download links for the SODA3 and SODA4 products on the University of Maryland website were not available at the time of the analysis. This validation confirms that ORAS5 reliably reproduces both the seasonal cycle and interannual variability of temperature at the four proxy sites, with some site-dependent differences in salinity noted and discussed.

Author's changes in the manuscript: No further changes required, it is already addressed in **Section 3.1.1 Line 177** and Supplementary Fig. S1 in response to Referee 1 (General Comment 1).

Line 90: Typo, southwesterlies winds

Done Section 2, line 99

Line 102: Phrases like "during a year" aren't necessary and can be removed.

Done, Section 2, line 108

Line 163: Provide the actual Bemis equation used. Do sensitivity testing with other paleotemperature equations. Note that in other places in this manuscript it is stated that a high light equation was used.

This has been resolved, the forward model no longer uses the Bemis et al. (1998) equation at all, having been replaced with species-specific equations from Farmer et al. (2007) for both G. ruber and N. dutertrei, with coefficients explicitly given (Eq. 1–2) in the latest version. We additionally performed the sensitivity testing requested against alternative species-specific paleotemperature equations (Mulitza et al., 2003; Bouvier-Soumagnac & Duplessy, 1985) in the previous major comments.

Author's changes in the manuscript: No change to the equations used (Farmer et al., 2007 retained for both species). We have added a new supplementary Figure S10, to report the paleotemperature-equation sensitivity test requested (Farmer et al., 2007 vs. Mulitza et al., 2003 for *G. ruber*; vs. Bouvier-Soumagnac & Duplessy, 1985 for *N. dutertrei* and compared to Bemis et al, 1998)

Lines 168-171: Include n and r-squared for all equations. The uncertainty in these relationships should be propagated to estimates of foram δ18O. Also consider using the new CoralHydro2K database, which may have additional data. https://earthchem.org/communities/seawater-oxygen-isotopes

We have added the sample size (n) for both δ^{18}Ow–S equations used, the R^2 and uncertainties are already reported in Section 3.1.1 since the last version. For Singh et al, 2010, n = 152 and for our thermocline equation n = 2073. Regarding uncertainty propagation: this is already implemented. At each Monte Carlo iteration, both the slope and intercept of these δ^{18}Ow–S equations are resampled from their reported standard errors, jointly propagating this uncertainty into the simulated foraminiferal δ^{18}O (Section 3.1.2).

We also thank the referee for pointing us to the CoralHydro2k Seawater δ18O Database (Atwood et al., 2026). We consider it a valuable resource for future refinement of this framework.

Author's changes in the manuscript: Added sample size **Section 3.1.1 lines 210 and 214** (n=152 surface; n=2073 subsurface) alongside the R^2 and parameter uncertainties already reported for both δ^{18}Ow–S equations in Section 3.1.1.

Line 200: It is confusing to provide δ18O -SSS relationships and then not use them. I would prefer consistency in drawing all relationships from a common database. At minimum, these region-specific relationships need to be provided in the same way as the equations above.

The inconsistency introducing a δ^{18}Ow–S relationship and then using a different one reflects the two-step structure of our forward-modelling framework. The region-specific relationships used in Step 2 were referenced only by citation, without being stated explicitly at the point they are introduced, unlike Eqs. 1–4. They are now given explicitly, with their sources, in Section 5.3.3; R^2, n and parameter uncertainties are not reported in all of the original publications and are therefore not given for these equations. Regarding use all the relationships from a common database, we considered this, but no single common source is available that provides validated, region-specific δ^{18}Ow–S relationships for all four sites, so we consider preferable to applying a single non-regional relationship uniformly in this second step, for the realism.

Author's changes in the manuscript: Region-specific δ^{18}Ow–S equations now referenced and given with their sources in Section 5.3.3. No further change to the two-step logic itself, which was already clarified in Section 3.1.2 in response to Referee 1 (Round 1, Specific Comment 8).

Line 206: "cycle that is, we do not" is awkward

Done, **Section 3.1.2 line 264**

Line 211: Please add discussion of if the +/- 75% is dynamically realistic. If not, consider more realistic ways to amplify/damp reanalysis data.

This point duplicates the referee's related Major concern above regarding the physical realism of the ±75% perturbation, where we quantify peak event ENSO/IOD related thermocline temperature anomalies from ORAS5 reanalysis.

Author's changes in the manuscript: See response to the related Major comment above; no additional changes beyond those already described there (new supplementary figure and Conclusion text added)

Line 213: Provide the test that determines if a standard deviation is "statistically robust"

"Statistically robust" refers to the fact that a sample size of 60 individuals is sufficient to obtain a stable estimate of the population SD following the estimation of total precision on the estimated variance in IFA (Leduc et al., 2009; Thirumalai et al., 2013). We perform 1000-repetition Monte Carlo bootstrap of the same 60-individual sampling scheme provides an empirical measure of this robustness, since the standard error of the resulting SD

estimate across repetitions (typically ±0.02–0.05‰, reported throughout Figs. 5 and 8) is small relative to the SD values themselves (0.19–0.70‰). We then consider that a SD estimate is robust in this sense, we have clarified this in the text.

Author's changes in the manuscript: Clarified in Section 3.1.2 lines 293–295) that "statistically robust" refers to the standard error of the SD estimate across the 1000-repetition Monte Carlo bootstrap.

Line 213: State if sampling is with or without replacement.

Done, with replacement in section 3.1.2, lines 287

Line 215: I would benefit from clarification here. I do not understand what it means for the "distribution of this ensemble of resampling" to not overlap. Overlap with what? I am also not clear how you do 1000 iterations of 60 and then somehow combine those 1000 iterations into a population of 60 data points (i.e. that are shown in Figure 5). Please elaborate on how your Monte Carlo sampling is performed.

We agree that the description was insufficient and have revised it. To clarify the MC procedure, for each site, species and scenarios we perform two steps.

First a single draw of 60 individuals as shown in Fig. 5 with its SD reported directly in the figure. Second, to assess whether this SD is a reliable representation of the underlying population rather than an artefact of that particular draw we repeat the same 60-individual sampling procedure 1000 times (the Monte Carlo bootstrap), each time recording the resulting SD.

This produces a distribution of 1000 SD estimates for the original scenario and a separate distribution of 1000 SD estimates for the altered scenario. "Do not overlap" refers to these two distributions of bootstrapped SD estimates: even accounting for the sampling variability inherent to drawing only 60 individuals at a time, the SD values obtained under the two scenarios are clearly separated, with a difference larger than their spread, confirming that the single-draw SD values shown in Fig. 5 are not sensitive to which particular 60 individuals happen to be drawn, and that the reported difference between scenarios reflects a genuine effect rather than sampling noise. We formalise this further with a paired Wilcoxon signed-rank test on these two sets of 1000 SD estimates ($p<0.001$ at all sites and scenarios; see also our response to the referee's comment on the "statistically robust SD" wording above).

Author's changes in the manuscript: Update of the description of the Monte Carlo procedure in **Section 3.1.2 at line 289** to clarify the two distinct steps.

Line 225: Statements like "In this study" are not necessary

Done and removed, lines 15, 76, 81, 317 and 872

Line 227: Add more information about the seasonality and depth preference of these species consistent with my comment above. Some of this information is already in the Discussion section.

No separate addition needed here; addressed by the PLAFOM2.0 paragraph already added in Section 5.3.2 (in response to Major comment 1), positioned immediately after the FORCIS February-peak sentence at NIOP 905, which notes the convergence at that site specifically and the disagreement at the other three.

Line 235: Please check this size fraction is correct. I have 212 and 300 micron sieves, but 215 and 315 are not typical in my experience.

Done, Corrected "215–315 µm" to "250-315 µm" **Section 3.2, line 327**

Lines 235-238: Expand significantly on your analytical method.

Added the range of weight used for individual specimens' analysis.

Lines 242-245: I suggest not removing outliers unless you have a clear analytical reason for doing so.

For this comment, same as related Major concern above on outlier removal. We refer the referee to our detailed response there, where we demonstrate that our outlier flagging is robust across two independent detection methods (IQR and a median-absolute-deviation-based modified z-score, which agree exactly on all five flagged individuals), and that the most extreme flagged value implies a temperature ~12 °C colder than the site mean at U1467, far outside anything documented at that site over 60 years of reanalysis, providing an analytical/physical basis for their removal beyond the statistical criterion alone.

Author's changes in the manuscript: See response to the related Major comment above; no additional changes beyond those already described there (Section 3.2, corrected outlier count wording, Fig. S9).

Line 246: Chapter 5?

Author's changes in the manuscript: The sentence referring to Chapter 5 has been removed, as the treatment of outliers is now fully described in Section 3.2

Line 262: This paragraph seems tacked on and out of place. I would put it earlier in the Methods section.

Author's changes in the manuscript: Moved the dissolution/preservation paragraph, before the Table 1.

Line 276: "variabilities" is awkward

Author's changes in the manuscript: *Corrected "These variabilities translate" to "This variability translates"(Section 4.1.1, Line 411*

Line 285: delete "values"

Done , **lines 420**

Line 332: Check Fig 4A-B. I think Fig 4A-D are include "Altered" scenarios

Done, line 486

Line 336: I don't think this paragraph says much. Either elaborate on the distinct patterns that emerge or consider deleting.

Done, we remove the paragraph

Line 366: "synthetics" is awkward

Already done in previous version

Line 367: phrases like "what we said" are not appropriate in formal writing.

Already done in previous version

Line 367: "assimilate to G. ruber" is awkward

Done, Section 4.1.2, line 534

Lines 408-410. The first 3 sentences are just restating what's been said and are not Results. I suggest they be delted.

Done, Section 4.2, lines 591

Lines 410-414: Given that none of the actual measured core-tops have 60 data points it would be more appropriate to forward model pseudoforams with the same number as actually measured.

We used a fixed N=60 for all synthetic core-tops to ensure a consistent statistical basis across sites, following standard IFA sampling design (Leduc et al., 2009; Thirumalai et al., 2013). We note that the standard error of an SD estimate scales approximately as $\sigma/\sqrt{(2n)}$, meaning that precision gains diminish rapidly beyond $n \approx 40–50$ individuals. However, while sample size affects the precision of an SD estimate, it does not systematically bias its expected value, typically our estimate of SD of the precision is ±0.02–0.05‰ as we explain in the response to the referee's comment on statistically robust SD above.

No change to the sampling design

Section 4.2: These results should also provide the other statistics such as Shapiro-Wilk, skewness, kurtosis, etc.

See response to the related Major comment above; no additional changes beyond those already described there (Section 5.3, lines 793-798 revised text and Fig. S9, reporting SD, skewness, kurtosis, Shapiro-Wilk, and KS statistics for all site/species comparisons).

Line 415: Somewhere in this manuscript please add a sentence summarizing the method of Ganssen et al. (2011). Was it exactly identical? Same size fraction? Measured in the same lab on the same instrument?

Added a supplementary Fig. S15 summarizing the analytical protocol (size fraction, instrument, laboratory, external reproducibility) for all externally-sourced IFA datasets used in this study, for reader reference.

Lines 440-441: Please provide the statistical test used to determine "statistically similar" versus "differ significantly"

Added change in ***Section 5.1, lines 659*** *revised to report explicit F-test p-values for all pairwise SD comparisons between sites.*

Line 445: Please provide the test for NIOP 905 SD being significantly higher

Author's changes in the manuscript: See response to lines 440-441 above, no additional changes needed here.

Line 463: "most reduced" is awkward

Done, section 5.1, lines 688

Lines 466-467: "multiple dynamical and ecological processes" is vague. Can the authors be more specific about which dynamical or ecological processes?

Done, **Section 5.1 line 692-693**

Line 504: events not event

Done, **Section 5.2 line 741**

Line 551: This systematic offset may reflect the paleotemperature equation used. Sensitivity testing with additional appropriate equations would help evaluate this.

Author's changes in the manuscript: See response to the related comment above; the paleotemperature-equation sensitivity test and its result are reported in the Fig. S10 (Section 5.3.1).

Line 577: This line of argument seem problematic. If the reanalysis data isn't a viable calibration target since it already shows too much anthropogenic influence, should I believe any of this manuscript?

Author's changes in the manuscript: Revised **Section 5.3.1 at lines 825-827 and 850-854** to restrict the anthropogenic-warming argument to the offset in mean values, removing the statement that the forward model overestimates natural baseline variability. The possible recent intensification of ENSO, IOD and monsoon extremes is now presented as a limitation rather than as an explanation of the offset. Our comparison does not detect it in the SD, but it cannot be excluded below the precision of a 60-specimen estimate (±0.02–0.05 ‰). Yet, the mean $\delta^{18}O$ is not the target of the study, whereas the SD is.

Line 585-586: Rather than saying these biological complexities should be considered in the future, the authors should consider them in this study. It will make this work more impactful.

Author's changes in the manuscript: Added a short paragraph in **Section 5.3.2 lines 882-890** reporting the probabilistic depth-sampling sensitivity test directly (magnitude of effect on SD, and its impact on the *G. ruber* site ranking; Fig. S13), rather than only flagging depth habitat as a future consideration, and retained the depth-integrated layer approach as the primary method. Seasonal shell production is now also tested directly, with FORCIS- and PLAFOM2.0-based weightings (Fig. S12).

Lines 598-601: Here is the type of information you already know and should incorporate into your pseudoforam modelling.

We thank the referee for this comment. Much of the information referenced in this passage has already been incorporated into the forward-modelling methodology since the version reviewed, rather than only discussed: the fixed 0.5 m surface depth has been replaced with a layer-integrated extraction (0–50 m), with an accompanying sensitivity test (Supplementary Fig. S14) showing minimal effect at NIOP 905 and FORCIS-based seasonal flux data are now used directly to weight the Monte Carlo sampling (Section 5.3.2), rather than only being cited descriptively. We have further incorporated the PLAFOM2.0-based seasonality test and the probabilistic depth-sampling test in this same section, in direct response to the referee's related Major comments above.

Author's changes in the manuscript: No additional change, already made in response to Referee 1 (depth methodology, FORCIS weighting) and those described in response to the referee's Major comments above (PLAFOM2.0, probabilistic depth sampling), all located **in Section 5.3.2.**

Lines 664-665: Would an ensemble of mid-Holocene or last millennium simulations help evaluate this?

Author's changes in the manuscript: Added a sentence noting that an ensemble of mid-Holocene/last-millennium simulations would provide a more direct test of this hypothesis **in Section 5.3.4, lines 990.**

Table 1: The caption attached to this seems too long. Tables typically only have a title.

Done

Figure 1: Consider adding how δ18Oc was calculate, how the thermocline is defined, etc. to the caption.

Done

Figure 5: Consider making this a violin plot, which will better show the overlapping data. Similarly, add some jitter to the points so they don't overlap.

Done

Figure 6: Consider using a violin plot instead of a box plot. If using a box plot clarify what the box and whiskers are showing (inner two quartiles? 1 standard deviation?). Instead of just showing the mean of the reanalysis data include the equivalent of the box and whiskers at all depths. This will help evaluate how the measured forams capture the distribution of data, not just its mean. Note that this figure caption says a high light calibration of Bemis was used. I suggest removing the yellow star. It could be confused as data. I like the jitter here!

Done, we updated the caption in the Figure 6, and add an uncertainty range around the profile of estimated δ^{18}Oc. We kept the yellow stars, which identify the newly generated datasets, and defined them explicitly in the caption.

Figure 8: I'm not sure this plot is necessary. It seems as if this same information could be incorporated into Figure 6.

We think that this plot is necessary, because it also show the synthetic generated data and compare them directly. Moreover, it helps to follow the last part of the discussion section which contains an important amount of text.

No changes are necessary.

References:

Atwood, A. R., Moore, A. L., DeLong, K. L., Long, S. E., Sanchez, S. C., Hargreaves, J. A., Morris, C. A., Pauly, R. E., Dassié, É. P., Felis, T., Voelker, A. H. L., Murty, S. A., and Cobb, K. M.: The PAGES CoralHydro2k Seawater δ18O Database: a FAIR-aligned compilation of seawater δ18O data to uncover "hidden" insights from the global ocean, Earth Syst. Sci. Data, 18, 1921–1941, https://doi.org/10.5194/essd-18-1921-2026, 2026.

Behringer, D. and Xue, Y.: Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean, in: Eighth Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, 2004.

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, 150–160, 1998.

Bouvier-Soumagnac, Y. and Duplessy, J.-C.: Carbon and oxygen isotopic composition of planktonic foraminifera from laboratory culture, plankton tows and Recent sediment: implications for the reconstruction of paleoclimatic conditions and of the global carbon cycle, J. Foraminifer. Res., 15, 302–320, https://doi.org/10.2113/gsjfr.15.4.302, 1985.

Chaabane, S., de Garidel-Thoron, T., Giraud, X., Schiebel, R., Beaugrand, G., Brummer, G.-J., et al.: The FORCIS database: A global census of planktonic Foraminifera from ocean waters, Sci Data, 10, 354, https://doi.org/10.1038/s41597-023-02264-2, 2023.

Chaabane, S., Schiebel, R., Meilland, J., Brummer, G.-J. A., Mortyn, P. G., Sulpis, O., Chalk, T. B., Giraud, X., Howa, H., Kuroyanagi, A., Beaugrand, G., and Garidel-Thoron, T. D.: Reassessment of the global distribution and diversity of modern planktonic Foraminifera from the FORCIS database, https://doi.org/10.22541/au.173991286.60773031/v1, 2026.

Farmer, E. C., Kaplan, A., De Menocal, P. B., and Lynch-Stieglitz, J.: Corroborating ecological depth preferences of planktonic foraminifera in the tropical Atlantic with the stable oxygen isotope ratios of core top specimens, Paleoceanography, 22, PA3205, https://doi.org/10.1029/2006PA001361, 2007.

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, Clim. Past, 7, 1337–1349, https://doi.org/10.5194/cp-7-1337-2011, 2011.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Kretschmer, K., Jonkers, L., Kucera, M., and Schulz, M.: Modeling seasonal and vertical habitats of planktonic foraminifera on a global scale, Biogeosciences, 15, 4405–4429, https://doi.org/10.5194/bg-15-4405-2018, 2018.

Leduc, G.: Interactive comment on "Quantifying sea surface temperature ranges of the Arabian Sea for the past 20 000 years" by G. M. Ganssen et al., Clim. Past Discuss., 6, C1298, https://cp.copernicus.org/preprints/6/C1298/2011/cpd-6-C1298-2011.pdf, 2011.

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, Paleoceanography, 24, PA3202, https://doi.org/10.1029/2008PA001701, 2009.

Lombard, F., Labeyrie, L., Michel, E., Bopp, L., Cortijo, E., Retailleau, S., Howa, H., and Jorissen, F.: Modelling planktic foraminifer growth and distribution using an ecophysiological multi-species approach, Biogeosciences, 8, 853–873, https://doi.org/10.5194/bg-8-853-2011, 2011.

Mulitza, S., Boltovskoy, D., Donner, B., Meggers, H., Paul, A., and Wefer, G.: Temperature:$\delta^{18}O$ relationships of planktonic foraminifera collected from surface waters, Palaeogeogr. Palaeocl., 202, 143–152, https://doi.org/10.1016/S0031-0182(03)00633-3, 2003.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, https://doi.org/10.1029/2002JD002670, 2003.

Saenger C, Jimenez-Diaz C, Gagnon A, Mix A, Ross A and Xu T (2024) A framework for reconstructing marine heatwaves from individual foraminifera in sedimentary archives. Front. Mar. Sci. 11:1321254. doi: 10.3389/fmars.2024.1321254

Singh, A., Jani, R. A., and Ramesh, R.: Spatiotemporal variations of the $\delta^{18}O$–salinity relation in the northern Indian Ocean, Deep-Sea Res. Pt. I, 57, 1422–1431, https://doi.org/10.1016/j.dsr.2010.08.002, 2010.

Stainbank, S., Kroon, D., Rüggeberg, A., Raddatz, J., de Leau, E. S., Zhang, M., and Spezzaferri, S.: Controls on planktonic foraminifera apparent calcification depths for the northern equatorial Indian Ocean, PLOS ONE, 14, e0222299, https://doi.org/10.1371/journal.pone.0222299, 2019.

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, 401–412, https://doi.org/10.1002/palo.20037, 2013.

Wejnert, K. E., Thunell, R. C., and Astor, Y.: Comparison of species-specific oxygen isotope paleotemperature equations: Sensitivity analysis using planktonic foraminifera from the Cariaco Basin, Venezuela, Mar. Micropaleontol., 101, 76–88, https://doi.org/10.1016/j.marmicro.2013.03.001, 2013.