Paleoceanography and Paleoclimatology

RESEARCH ARTICLE
10.1029/2023PA004663

Shell Reworking Impacts on Climate Variability Reconstructions Using Individual Foraminiferal Analyses

Natalia Bienzobas Montávez[1], **Kaustubh Thirumalai**[2], **and Gianluca Marino**[1]

[1]Centro de Investigación Mariña, Palaeoclimatology Lab, Universidade de Vigo, GEOMA, Vigo, Spain, [2]Department of Geosciences, University of Arizona, Tucson, AZ, USA

Key Points:

- A new software (TurbIFA) quantifies the uncertainties in IFA data sets tied to sampling, analytical errors, and bioturbation
- TurbIFA assesses the sensitivity of IFA reconstructions to bioturbation parameters and propagates analytical and sampling uncertainties
- Application to existing data sets showcases its ability to assess statistical significance of IFA-based climate variability reconstructions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

G. Marino,
gianluca.marino@uvigo.es

Citation:

Bienzobas Montávez, N., Thirumalai, K., & Marino, G. (2024). Shell reworking impacts on climate variability reconstructions using individual foraminiferal analyses. *Paleoceanography and Paleoclimatology*, *39*, e2023PA004663. https://doi.org/10.1029/2023PA004663

Received 25 APR 2023
Accepted 30 APR 2024

Author Contributions:

Conceptualization: Natalia Bienzobas Montávez, Kaustubh Thirumalai, Gianluca Marino
Formal analysis: Natalia Bienzobas Montávez
Funding acquisition: Kaustubh Thirumalai, Gianluca Marino
Investigation: Natalia Bienzobas Montávez, Kaustubh Thirumalai, Gianluca Marino
Methodology: Natalia Bienzobas Montávez, Kaustubh Thirumalai, Gianluca Marino
Project administration: Gianluca Marino

Abstract Particle mixing by benthic fauna beneath the sediment-water interface (or bioturbation) fundamentally challenges the proxy based retrieval of past climatic conditions from deep-sea sediment cores. Previous efforts targeted the impacts of bioturbation on the nature of paleoceanographic changes gleaned from the proxy record, whereas impacts on seasonal and/or interannual variability reconstructions have received less attention. We present TurbIFA (Tracking uncertainty of reworking & bioturbation on IFA), a software that adapts and combines existing algorithms to quantitatively estimate the impact of sediment reworking and other uncertainties and assess significance of ocean and climate variability reconstructions based on individual foraminiferal analyses (IFA). Building upon previous idealized investigations of bioturbation using hydroclimate-sediment simulations, TurbIFA advances the IFA proxy system modeling such that users may directly assess the sensitivity of their data to various local parameters related to shell reworking across the global ocean. Using the output of state-of-the-art coupled atmosphere-ocean general circulation models, TurbIFA simulates planktic foraminiferal $\delta^{18}O$ or Mg/Ca-temperature signal carriers and evaluates uncertainties in the sample size, analytical protocols along with as those arising from bioturbation. Application of TurbIFA to synthetic and existing data sets indicates that the significance of IFA-based reconstructions can be assessed once the impacts of sediment accumulation rates, sediment mixed layer depths, length of time integrated by the chosen IFA sampling interval, and changes in the amplitude of climate variability (i.e., the targeted environmental signal) are comprehensively evaluated. We contend that TurbIFA can aid quantitative assessments of past seasonal and interannual variability gleaned from the paleoceanographic record.

Plain Language Summary Planktic foraminifera are marine zooplankton with a life cycle of approximately a month. They dwell in the upper ocean, where they precipitate calcium carbonate shells that, postmortem, settle through the water column and become a main constituent of deep-ocean sediments. The geochemical composition of individual planktic foraminiferal shells provides snapshots of past surface ocean temperatures during their lifespan. Statistical distributions of these data inform on the past seasonal and/or interannual climate variability, but post-depositional processes may bias these reconstructions. Here we quantify the effects of the vertical displacement of foraminiferal shells in sediments by benthic microfauna, which contaminates sediment core samples with older shells deposited under colder or warmer climates. The code routines are available in an open-source repository. The final user can define their study-specific variables as inputs, thereby allowing the application of this method to a broad spectrum of climate and sedimentary settings.

1. Introduction

Over the past 15 years, the analysis of stable oxygen isotope ($\delta^{18}O$) and magnesium to calcium (Mg/Ca) ratios in individual shells of planktic foraminifera—termed individual foraminiferal analyses (IFA, Thirumalai et al., 2013)—has gained momentum as a powerful tool to reconstruct seasonal and/or interannual oceanic (and climatic) variability in various oceanographic settings. IFA data have laid the foundations for a deeper understanding of high-frequency climate variability in the past, for example, the El Niño–Southern Oscillation (ENSO) system (Ford et al., 2015; Khider et al., 2011; Koutavas & Joanides, 2012a; Koutavas et al., 2006; Leduc et al., 2009; Rustic et al., 2020; Scroxton et al., 2011; White et al., 2018), seasonality in the North Atlantic Ocean and Arabian Sea (Brummer et al., 2020; Metcalfe et al., 2019; Naidu et al., 2019), the recently hypothesized equatorial Indian Ocean mode of variability during the Last Glacial Maximum (LGM, Thirumalai et al., 2019b). One emerging paradigm from IFA studies points to the mean state of global (and regional) climate as a modulator

Resources: Kaustubh Thirumalai, Gianluca Marino
Software: Natalia Bienzobas Montávez, Kaustubh Thirumalai
Supervision: Kaustubh Thirumalai, Gianluca Marino
Validation: Natalia Bienzobas Montávez, Kaustubh Thirumalai, Gianluca Marino
Visualization: Gianluca Marino
Writing – original draft: Natalia Bienzobas Montávez
Writing – review & editing: Natalia Bienzobas Montávez, Kaustubh Thirumalai, Gianluca Marino

of the superimposed, shorter-term climate variability (Ford et al., 2015; Rustic et al., 2020; Thirumalai et al., 2019b).

These findings build on the assumption that the statistics of IFA distributions reflect ocean and climate variability within the time span archived in a discrete sedimentary sample. An IFA data set typically consists of $\delta^{18}O$ and/or Mg/Ca measurements performed on individual planktic foraminiferal shells, which capture the geochemical signature of past seawater conditions over their lifespan of a few weeks to a month (Kucera, 2007). A key tenet to this assumption is that modern IFA distributions are commonly determined from core-top (late Holocene, LH) samples and compared to their downcore counterparts (IFA time slices) to characterize paleoclimate variability. Toward this, studies have demonstrated that IFA performed on modern core-top sediments accurately match the patterns of climate variability observed in the instrumental record (Figure 1). For instance, early work from the eastern equatorial Pacific Ocean documented remarkable similarity between the mean and variance in the core-top IFA $\delta^{18}O$ data of two nearby sediment cores (both core-tops with an age of ∼1 ka BP) and their agreement with observations (Koutavas & Joanides, 2012a, 2012b). Likewise, late Holocene (LH, <2 ka BP) IFA $\delta^{18}O$ metrics from the eastern Indian Ocean closely agree with modern observational time series (Thirumalai et al., 2019a, 2019b). Finally, evidence from the equatorial Pacific Ocean demonstrates that the variance of IFA Mg/Ca distributions for the surface-dwelling planktic foraminifer *Globigerinoides ruber* in LH (∼4.5 ka BP to ∼0.4 ka BP) core-tops (n = 9) reflect site-specific mixed layer temperature variability (Rongstad et al., 2020).

However, such field-based evidence is at odds with statistical tests performed using a transient hydroclimate-sediment model that incorporates the impact of bioturbation on the reproducibility of IFA data (Lougheed & Metcalfe, 2022). The study hinges on a bioturbation model (Lougheed, 2020) specifically designed for IFA studies and suggests that accurate reconstructions of the amplitude of climate variability and the likelihood of extreme events can only be achieved through the analysis of several hundreds of individual foraminiferal shells per sample (Lougheed & Metcalfe, 2022). This is much larger than the typical IFA sample sizes utilized in recent studies, which range from ∼40 to ∼100 specimens. It may correspond to all the individuals available within a discrete depth of sediment and/or time slice, notably if taxonomic consistency (at the morphotype level) and relatively narrow sampling windows are sought. The findings of Lougheed and Metcalfe (2022) implicate that IFA studies would have to be limited to continental margin settings (∼7% of the global ocean by surface) that generally feature high sediment accumulation rates (SARs) (Restreppo et al., 2020) and high export fluxes of planktic foraminifera (Kiss et al., 2021; Žarić et al., 2006) that compensate for the "dilution effect" tied to high terrigenous supply. Hence, to ensure reliable interpretations of (past) climate variability and broader geographic coverage of IFA-based reconstructions it is important to accurately quantify the impact of bioturbation on IFA-based reconstructions, particularly in tropical open-ocean locations where SARs are lower, surface waters are

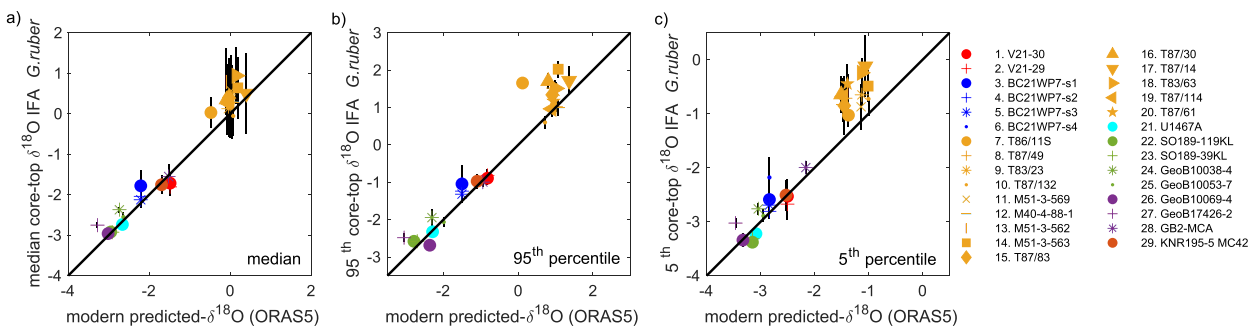

Figure 1. Comparison of published core-top stable oxygen isotope ($\delta^{18}O$) IFA (*Globigerinoides ruber*) data and predicted-$\delta^{18}O$ (ocean surface) from ORAS5 reanalysis data (Zuo et al., 2019): (a) median; (b) 95th percentile; (c) 5th percentile. For each study, we extracted ORAS5 sea surface temperatures and sea surface salinity data spanning from year 1978 to year 2018, calculated the seawater $\delta^{18}O$ ($\delta^{18}O_{sw}$) from the regional $\delta^{18}O_{sw}$-S relationship, and made the conversion to predicted-$\delta^{18}O$, using the paleotemperature equations of Bemis et al. (1998). IFA databases: 1–2. (Koutavas & Joanides, 2012a, 2012b); 3–6. Individuals picked at 250–300, 300–355, 355–400, and 400–500 μm size fractions De Moel et al. (2009), de Moel et al. (2010), 7–20. Wit et al. (2010), 21. Stainbank et al. (2021a, 2021b), 22–25. Thirumalai et al. (2019a, 2019b); 26–28. Groeneveld et al. (2019a, 2019b), 29. Rustic et al. (2015, 2016). $\delta^{18}O_{sw}$-S relationships are based on LeGrande and Schmidt (2006) for: the tropical Pacific Ocean (Koutavas & Joanides, 2012a; Rustic et al., 2015); the Indian Ocean (De Moel et al., 2009; Thirumalai et al., 2019b); the North Atlantic Ocean (cores T86/11S and T87/132, Wit et al., 2010); and the tropical Atlantic Ocean (core GB2-MCA, Groeneveld et al., 2019b). They are based on Singh et al. (2010) for the Northern Indian Ocean (Stainbank et al., 2021b) and on Conroy et al. (2017) for the Manus temporal data set (core GeoB17426-2, Groeneveld et al., 2019a). Error bars ($\pm 2\sigma$) account for the noise that stems from randomly sampling (1,000 times) the predicted-$\delta^{18}O$ time series, based on the sample size and a Gaussian analytical uncertainty (specific to each data set).

typically oligotrophic (Buitenhuis et al., 2013; Müller & Suess, 1979), and export fluxes of planktic foraminifera are relatively low.

Sediment mixing occurs in the few centimeters of sediment below the sediment-water interface, usually extending down to 5 cm (Solan et al., 2019; Teal et al., 2008). It results in the vertical displacement of foraminiferal shells (and other sediment particles) through the sediment mixed layer depth (SMLD), which increases the age variance within the discrete depth samples targeted by IFA studies (Dolman et al., 2021; Lougheed et al., 2020; Zuhr et al., 2022) and distorts traditional "multi-test" reconstructions (e.g., Slowey & Curry, 1995). This implies that a $\delta^{18}O$ or a Mg/Ca-temperature IFA data set potentially contains individuals that could have precipitated their shells when climatic mean state and variability were different, thereby generating noise that biases the distribution and interpretation of the IFA data (Lougheed & Metcalfe, 2022).

The comparison of core-top and downcore IFA data—a fundamental tenet of the IFA approach (e.g., Ford et al., 2015; Thirumalai et al., 2019b; White et al., 2018)—relies to a first order on the stationarity of the ecological preferences of the selected foraminiferal species. For example, in regions with pronounced oceanographic variability, the variance of core-top IFA Mg/Ca and $\delta^{18}O$ distributions may be somewhat muted because the targeted species (*Neogloboquadrina dutertrei*) tends to restrict its growing season and/or depth habitat (Groeneveld et al., 2019a). On the other hand, if the (paleo)ecology of a certain species is well constrained, the number of individual shells analyzed at each location (in the absence of bioturbation) should be ideally similar and enough (~60–80) to faithfully reconstruct local oceanographic variability (Thirumalai et al., 2013). This factor was tackled in previous studies that developed proxy system models such as the INFAUNAL (Thirumalai et al., 2013) and QUANTIFA (Glaubke et al., 2021) algorithms to calculate confidence limits, based on bootstrap resampling methods for forward-modeled IFA $\delta^{18}O$ and Mg/Ca-temperature distributions, respectively.

Previous work has also simulated the mixing of individual particles (e.g., foraminiferal shells) by bioturbation in sediment core archives, for example, the TURBO2 (Trauth, 2013), *sedproxy* (Dolman & Laepple, 2018), SEAMUS (Lougheed, 2020), and iTURBO2 (Hülse et al., 2022) algorithms. Bioturbated sediment archives feature an apparent heterogeneity in the ages of individual foraminifera (Dolman et al., 2021; Lougheed et al., 2018), calling for IFA reconstructions that account for sediment-mixing processes in their noise estimation routine and interpretative framework. Although existing algorithms (Dolman & Laepple, 2018; Lougheed, 2020; Lougheed & Metcalfe, 2022) model the vertical mixing of foraminiferal shells below the sediment-water interface, they do not quantify the uncertainties and/or assess the statistical significance of oceanic and climate variability reconstructions that are obtained from $\delta^{18}O$ and/or Mg/Ca-temperature IFA data. Moreover, they do not allow users to input their own (core-top and downcore) data sets to provide a quantitative platform to investigate the sensitivity and significance of their IFA measurements with respect to shell reworking processes. Here we build upon and advance previous work (Dolman & Laepple, 2018; Glaubke et al., 2021; Lougheed, 2020; Thirumalai et al., 2013) by presenting a software that combines and adapts existing algorithms (i.e., SEAMUS, QUANTIFA, INFAUNAL) that is designed to aid paleoceanographers that use IFA in their research to: (a) quantify the cumulative uncertainties arising from methodological protocols (number of individuals measured, instrumental precision) in IFA-$\delta^{18}O$ and Mg/Ca-based sea surface temperature (SST) data sets collected from a bioturbated sediment archive; and (b) assess statistical significance of the climate variability reconstructions based on IFA. The software uses the outputs of a simulation of the transient climate evolution of the last 21 Kyr (TraCE-21ka, He, 2011), whereas other time series can be used where models appear to underestimate the amplitude of variability at the regional to local scales (Laepple et al., 2023).

2. Methods

This study provides a new software dubbed TurbIFA (Tracking uncertainty of reworking & bioturbation on IFA) that adapts and combines a set of user-friendly MATLAB code routines (Glaubke et al., 2021; Lougheed, 2020; Thirumalai et al., 2013) and is freely available to download (Bienzobas Montavez et al., 2024).

The bioturbation simulation routine implemented in TurbIFA builds on the same protocols outlined in SEdiment AccuMUlation Simulator (SEAMUS, Lougheed, 2020), a model that simulates the accumulation and mixing of individual foraminiferal shells in the sediments over time. Following Berger and Heath (1968), SEAMUS and TurbIFA simulate a homogeneous random step-by-step mixing of "foraminiferal individuals" through the active sediment mixed layer. TurbIFA's mixing model reuses and rewrites parts of the SEAMUS source code and implements a series of technical edits, among which the monthly values (of e.g., temperature) are simulated as

individual foraminifera signal carriers and not individuals themselves, an approach that more faithfully mirrors the actual IFA analytical protocols and general philosophy (Thirumalai et al., 2013). Briefly, the number of individuals archived in an IFA discrete depth slice is equal to the size of the IFA data set, whereas the monthly temperature values (foraminiferal signal carriers) archived in a depth slice are "picked" to calculate the noise associated to the IFA sample-size (see Supporting Information S1 for further information about the implementation of coding routines in TurbIFA). Given that bioturbation is a random process, the age distribution of the "individual shells" archived in a given discrete depth sample (and hence the e.g., predicted-$\delta^{18}O$ values) slightly changes each time the model is run (Lougheed, 2020). Accordingly, by default, TurbIFA generates an ensemble of 50 bioturbated synthetic sediment core archives.

Within this framework to model the impact of shell reworking on "picked" foraminifera, TurbIFA implements the QUANTIFA proxy system model (Glaubke et al., 2021) to compute the attendant uncertainties of the IFA distribution associated with the data set size (sampling uncertainty) and instrumental precision (analytical uncertainty). Accordingly, TurbIFA is capable of assessing the statistical significance of changes in the sample distributions (quantile-quantile [Q-Q] plots) of IFA data sets. For $\delta^{18}O$ IFA data sets, we forward-model planktic foraminiferal-$\delta^{18}O$ within a framework of monthly resolved time series. In contrast, TurbIFA directly requests temperature inputs for Mg/Ca-temperature IFA data, such that the calibration choice and other Mg/Ca-related uncertainties are not assessed. This is because there are several calibration algorithms in the published literature (Anand et al., 2003; Dekens et al., 2002; Gray & Evans, 2019; Kısakürek et al., 2008; Rosenthal et al., 2022; Thirumalai et al., 2016; Tierney et al., 2019). Alternatively, TurbIFA allows for the calibration error to be propagated as a Gaussian noise at the time of "picking" the temperatures (i.e., the foraminiferal signal carrier) so that users may test the sensitivity of the results by altering the magnitude of the noise terms added.

We use monthly resolved surface air temperatures (SAT) and decadal-resolution sea surface salinity (SSS) time series from TraCE-21ka (He, 2011)) as the basis for TurbIFA. Considering that this simulation spans from 21 ka to present, the full applications of the current version of TurbIFA are limited to IFA data sets within this timescale. We note that TurbIFA has been designed to utilize other reconstructions as (if) they become available, in line with other forward modeling packages (Dolman & Laepple, 2018; Lougheed, 2020). For Mg/Ca-SST IFA data, we extract SAT at the nearest grid point to the user-defined core location, whereas for the predicted-$\delta^{18}O$ we: (a) use the nearest grid point in the TraCE-21ka monthly SAT and decadal SSS time series; (b) resample the TraCE-21ka SSS at monthly resolution; (c) apply the regional seawater $\delta^{18}O$ ($\delta^{18}O_{sw}$) to salinity (S) relationship(s) and the (0.27‰) $\delta^{18}O_{sw}$ to foraminiferal (calcite) $\delta^{18}O$ scale $CO_2 - H_2O$ conversion (Hut, 1987); (d) add the glacial-interglacial influence of sea level (ice volume) on the synthetic $\delta^{18}O_{sw}$ time series, based on the latest sea-level reconstruction (Lambeck et al., 2014), interpolated at a monthly resolution, assuming that such a scaling is constant through time (Thirumalai et al., 2016); and (e) use established paleotemperature equations (Bemis et al., 1998) to forward-model foraminifera $\delta^{18}O$. This study targets the SAT (rather than SST) time series from TraCE-21ka, because it is the only temperature output available at the monthly resolution needed to mimic foraminiferal lifespan and evaluate seasonal and/or interannual climate variability changes. A caveat pertains to the assumption that no depth habitat changes affect the species targeted by the study because long-term monthly resolved climate model outputs are to date only available for the ocean surface (He, 2011).

In this study, we showcase how TurbIFA can aid paleoceanographic interpretations and apply it to real-world IFA data sets. First, we perform sensitivity tests using long-term (30,000 years) sediment mixing simulations that contextualize the impacts of site-specific parameters (SAR, SMLD, amplitude of climate variability, mean state changes) on synthetic foraminiferal Mg/Ca-based sea surface temperature (SST) reconstructions archived in a discrete depth sample (Figures 2–4). We extract a 1-Kyr-long surface air temperature time slice from TraCE-21ka (He, 2011) at a tropical Pacific oceanic location (−1.85°S, 202.5°E) and generate a 30-Kyr-long SST record by repeatedly splicing it. This provides a surrogate SST record archived in a hypothetical sediment core of the length needed to explore the impacts of reworking on IFA data for a given input time slice and various sediment mixing scenarios. These include cases where high SMLD (i.e., 10 cm) and low SAR (i.e., 2 cm Kyr^{-1}) coexist (note that TurbIFA users can use the local SAR and SMLD at their core site as input; see Supporting Information S1 for more details). The sensitivity test targets 1 Kyr time slice(s), in which the amplitude of variability (seasonal, interannual) and the mean temperature(s) are altered to be +2×, 0.5×, and +2°C, −2°C, respectively. We then apply TurbIFA to previously published $\delta^{18}O$ and Mg/Ca-temperature IFA measurements in mixed-layer dwelling planktic foraminifera (e.g., *G. ruber*) along three tropical oceanic sites (detailed below), which are located in the wider Indo-Pacific Ocean in continental margin and "open ocean" settings. These sites are contrasted for their

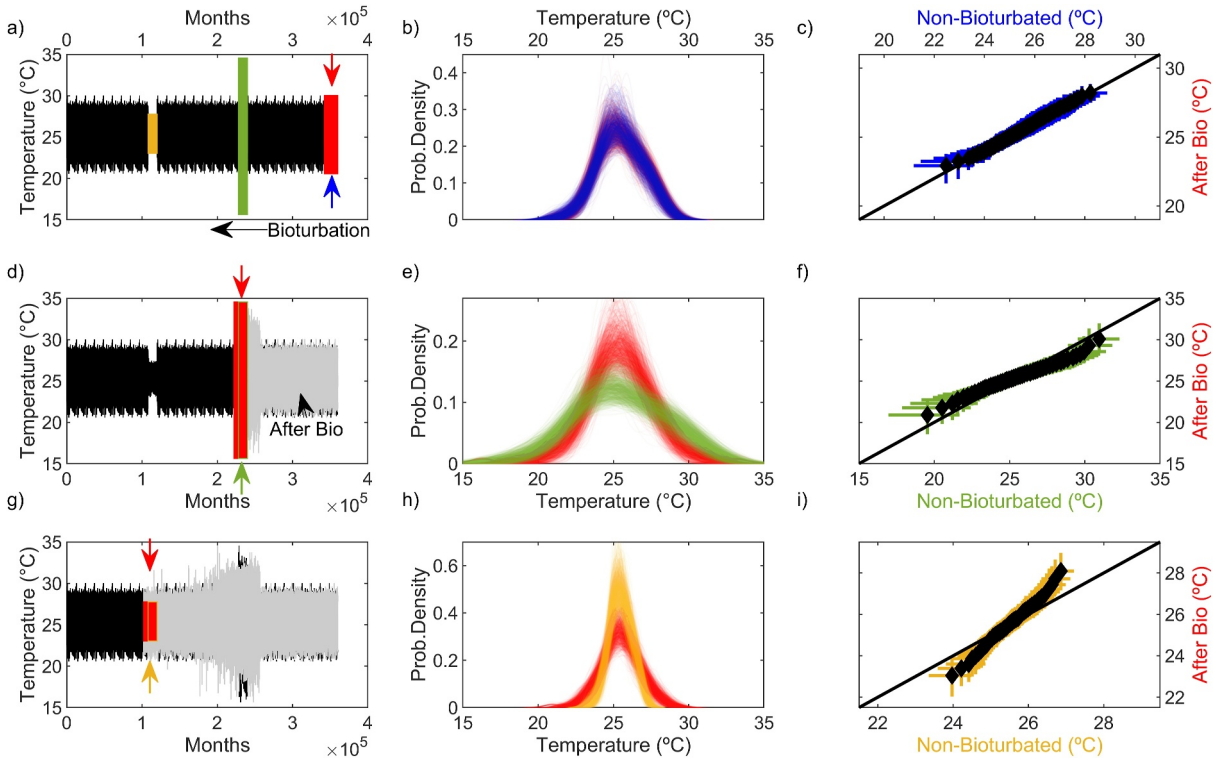

Figure 2. Impact of bioturbation on an idealized sea surface temperature (SST) distribution, which is archived in a synthetic discrete depth sediment sample that deposited over a time span of 100 years within the active sediment mixed layer depth (SMLD). Sediment particle mixing smears the signal through intervals that differ in the amplitude of their SST variability. The mixing of SST time series progresses upward in the core (i.e., toward the core top). (a, d, and g): Monthly SST time series before (black) and after (gray) the bioturbation front has reached a certain core depth, whereby green and yellow rectangles indicate times at which the amplitude of variability was altered (by −2× and +2×, respectively) and red rectangle indicates the upward progression of the bioturbation front. (b, e, and h): Kernel density distribution of the resampled SST (pseudo) IFA distributions before and after (red line) bioturbation. (c, f, and i): Quantile-quantile plots of the SST values before (X axes, Non-Bioturbated) and after bioturbation (Y axes, After Bio) at each time step (discrete depth). Quantiles ($n = 50$) were calculated as the median of the resampled SST data sets in a quantile-quantile space. The error bars are calculated using the QUANTIFA (Glaubke et al., 2021) bootstrap resampling and binning scheme. Arrows indicate the position of the synthetic core where SST data before and after bioturbation is resampled. 1:1 line in black.

respective SMLD, SAR, and nature and amplitude of climate variability, thereby providing key examples for assessing the impacts of sediment-mixing processes on paleoclimate IFA reconstructions. We show how TurbIFA can help to interpret and assess the significance of IFA results for studies targeting the last ~21 Kyr (i.e., over the TraCE-21ka timespan; see Examples 1 and 2 below) and beyond (see Example 3 below).

2.1. Uncertainty Evaluation

We study the impact of sediment-mixing processes on $\delta^{18}O$ and Mg/Ca-SST IFA data sets using a series of MATLAB code routines that: (a) evaluate the age heterogeneity of the individuals archived in a sediment depth slice (Lougheed, 2020); (b) quantify the uncertainties arising from methodological IFA protocols (Glaubke et al., 2021; Thirumalai et al., 2013); (c) provide confidence intervals and statistical significance of the associated downcore IFA climate variability reconstructions (Glaubke et al., 2021; Thirumalai et al., 2013) and; (d) assess whether bioturbation distorts its paleoclimatic interpretation. We perform a bootstrap resampling routine by repeatedly "picking" ($n = 1{,}000$) the foraminiferal signal carriers archived within the sedimentary depth slice and perform the following operations:

1. We assess the probability of picking 1–20 individuals that precipitated their shells during time slices 1-to-10 Kyr older than the chosen IFA time slice representing a given discrete sample depth by running TurbIFA core model or *bioprob* script (see the software documentation in Bienzobas Montavez et al. (2024) for further information). The ranges of reworked individuals and of the ages of the "potentially bioturbated" time slices ensure that the impact of bioturbation on the tails of the IFA distribution is rigorously assessed with changes in the mean climatic state as well. Toward this, we calculate the difference in age between each simulated single

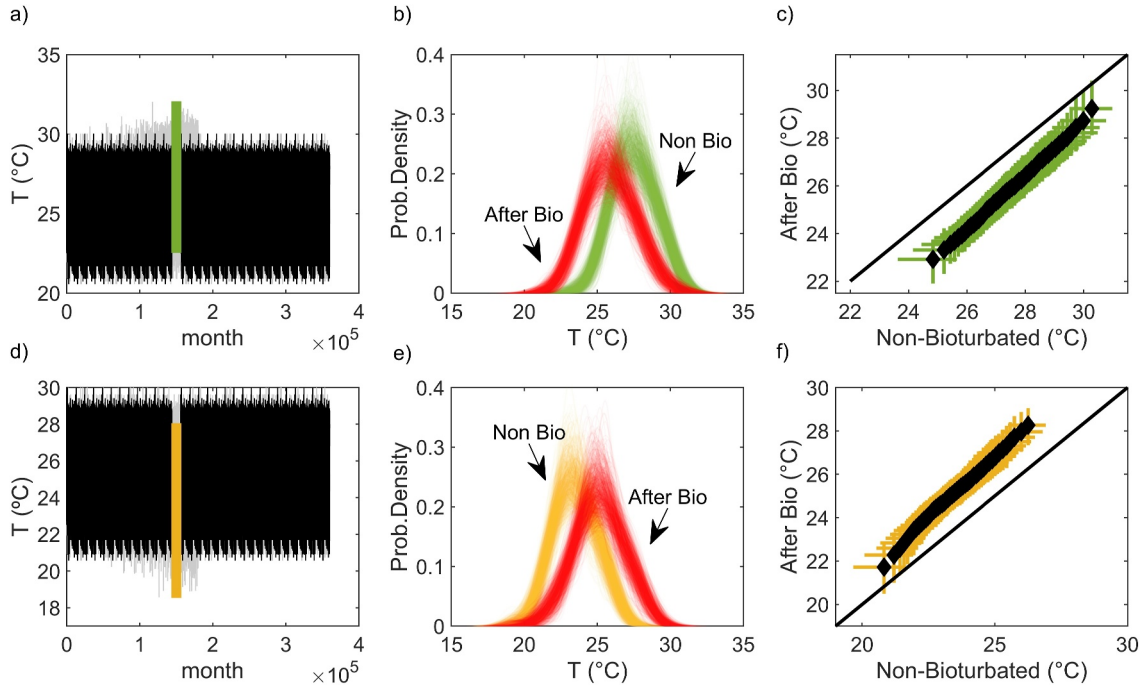

Figure 3. Bioturbation impact on an idealized SST signal archived in a synthetic discrete depth sediment sample that spans 500 years in duration. Sediment particle mixing has affected intervals where there is an abrupt shift toward warmer (+2°C) (a–c) or cooler (−2°C) (d–f) mean temperatures. Monthly resolved SST time series before (black) and after (gray) bioturbation. The green rectangle covers the time slice with altered (+2 or −2°C) temperature. Kernel density functions of the resampled SST data sets (60 "foraminifera" each) before (green: +2°C; yellow: −2°C) and after bioturbation (red). Quantiles ($n = 50$) were calculated as the median of the resampled SST data sets after conversion to a quantile-quantile space.

"foraminiferal shell" and the mean age of their depth slice. Note that the oldest age of the TraCE-21ka time series imposes the age (21 ka) of the oldest individual that can be found at a discrete depth. By fixing SAR, SMLD, and the abundance of the individuals to remain constant through time, TurbIFA samples "foraminifera" from an exponential age distribution calculated using a scale parameter equal to SMLD/SAR (Berger & Heath, 1968). Parameters are also allowed to vary over time (based on user input), wherein TurbIFA adjusts the times series length to be bioturbated before performing the random step-by-step sediment-mixing simulation (Lougheed, 2020). After the simulation of bioturbation, the picking of "foraminiferal individuals" is performed on the sediment depth slice where its mean age is closest to the mean age of the IFA depth.

2. We calculate the uncertainty (95% confidence intervals) in the IFA $\delta^{18}O$ or Mg/Ca-SST distribution standard deviation (1σ), and the probability that a downcore change in the IFA-1σ is significant. Similarly, we pick simulated "foraminiferal individuals" from the bioturbated discrete depth of sediment where its mean age (derived from TraCE-21ka) is closest to the mean age of the IFA depth. TurbIFA uses the INFAUNAL algorithm (Thirumalai et al., 2013) to calculate the probability of a significant change, where we use the root-mean square error of the IFA core-top and downcore standard errors as our threshold for significance (Lanzante, 2005). Note that the use of the IFA-1σ parameter to discern changes in oceanographic variability is only adequate when the measured IFA data sets are normally distributed.
3. We calculate the uncertainties (95% confidence interval) associated with contrasting the core-top versus downcore IFA distributions in the quantile-quantile space, using the QUANTIFA algorithm (Glaubke et al., 2021). For each "bioturbated" synthetic core, the "picking" routine is performed on the bioturbated discrete depth of sediment where its mean age is closest to the mean age of the IFA data set depth slice. The pseudo-IFA data sets are transformed into quantile-quantile space, and the uncertainty for each quantile is computed by assessing the 2.5th and 97.5th percentiles. The "final" cumulative uncertainty associated with the IFA distribution are determined as the median of the errors observed across the ensemble of synthetic cores ($n = 50$).
4. We evaluate the probability that a significant downcore change in the shape (increase or decrease) of the IFA distribution can be explained by sediment-mixing processes rather than by changes in climate variability.

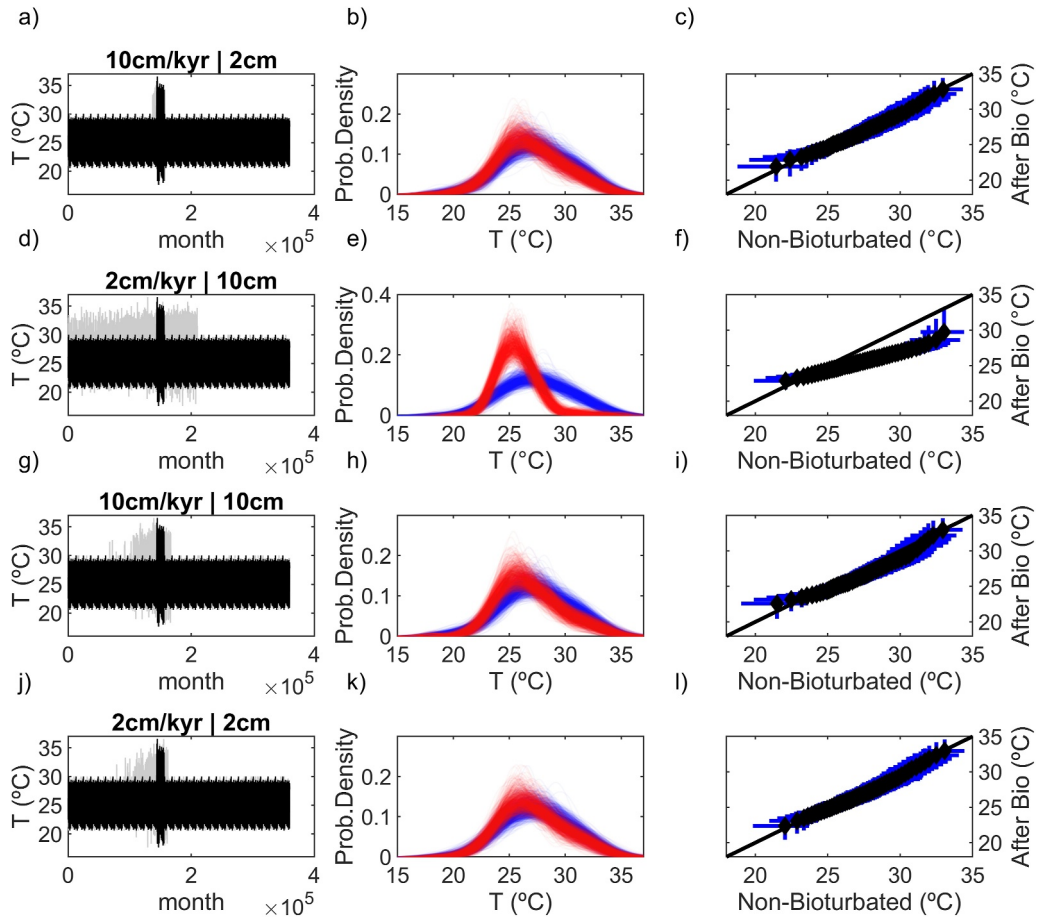

Figure 4. Impact of bioturbation on the SST distribution as a function of sediment accumulation rate (SAR) and sediment mixed layer depth (SMLD). SST data before (blue) and after (red) bioturbation.

Toward this, we compare the results obtained via "picking foraminiferal shells" from a bioturbated versus non-bioturbated (or unperturbed) sediment archive within quantile-quantile space. We designate significant changes for each quantile if the reconstructed change is larger than the calculated uncertainty (Glaubke et al., 2021). The results are averages over 1,000 realizations of the sampling routine and the ensemble of synthetic cores for the bioturbated scenario. Note that for the non-bioturbated archive, the resampling is performed within the IFA time slice itself (years archived in a depth slice according to the SAR).

2.2. Application to Real-World IFA Data

Illustrative "use-case examples" are presented by applying TurbIFA to IFA conducted at three locations where published IFA $\delta^{18}O$ or Mg/Ca-SST data sets are available (Stainbank et al., 2021b; Thirumalai et al., 2019b; White et al., 2018). The targeted locations in the Indo-Pacific Ocean feature different SAR, SMLD, and climatic controls (ENSO, Indian Ocean Dipole [IOD], Indian Ocean Basin Mode [IOBM]) on surface-ocean variability.

In the first example, we use *G. ruber* IFA Mg/Ca data from core MGL1208-14MC/12GC (White et al., 2017, 2018), which is located in the central equatorial Pacific Ocean, where ENSO drives interannual SST variability, the mean SAR is low (~3 cm Kyr^{-1}), and the SMLD is assumed to be high (see below), consistent with the relatively old age of the core-top (4.03 ka B.P.). We used their IFA Mg/Ca-based temperatures (°C) for the LH core top and the Younger Dryas (YD; 12.17 ka B.P.) as input to TurbIFA. As the reported SMLD (Solan et al., 2019) is too shallow to explain the age of the core top, and there are no organic carbon data available in the region, we computed the SMLD as the core-top age multiplied by the SAR, yielding an SMLD value of ~12 cm. The time slice covered within each IFA discrete depth is calculated according to the calibrated age and mean SAR

Table 1
TurbIFA Input Data for the Examples Used in This Study

Core	MGL1208-14MC/12GC	SO189-119KL	359-U1467C
Location	0°13'S 204°E	3.52°N 96.32°E	3.20°N 73.22°E
Depth (m)	3049	808	487
Core-top IFA age (ka)	4,030	1,889	–
Core-top IFA time slice (ka)	3,864–4,196	1,861–1,917	–
Paleo-IFA age (ka)	12,170	8,028	331
Paleo-IFA time slice (ka)	12,004–12,336	8,000–8,056	330,892–331,108
Proxy	Mg/Ca	$\delta^{18}O$	$\delta^{18}O$
SAR (cm Kyr^{-1})	3[a]	18[b]	Dynamic
SMLD (cm)	12	34	5
Num_p	68	62	83
$Season_p$	Year	Year	Year
$\delta^{18}O_{sw}$-S	–	Indian Ocean[c]	–
Analytical uncertainty (1σ)	0.5	0.05	–

Note. SAR = sedimentation accumulation rate; SMLD = sediment mixed layer depth; Num_p = number of individuals measured per IFA sample; $Season_p$ = seasonal preference (growing months). [a]Calculated between coretop and the Younger Dryas (YD). [b]Calculated over the last 21 Kyr. [c]LeGrande and Schmidt (2006).

(Table 1). For this example, the SAR, SMLD, and abundance of IFA individuals are set to be constant through time–including for the periods of comparison (i.e., LH, YD).

The second example targets a site located in the eastern tropical Indian Ocean, SO189-119KL (Mohtadi et al., 2014). This is a continental margin location with remarkably high SAR (~18 cm Kyr^{-1}) in a region where interannual SST variability is largely driven by the IOD. To obtain an extreme, idealized "high SAR/deep SMLD" scenario, we use a very deep value of 34 cm (Solan et al., 2019) as input SMLD, in spite of a young core top age of ~0.6 ka B.P. (Mohtadi et al., 2014), which points to a realistically shallower SMLD. The large simulated SMLD value implies that the mean age of the core top time slice is ~1,889 ka (i.e., calculated from SMLD/SAR) after bioturbation, which is comparable to the authors' inference of a 2-Kyr-long time slice. Hence, we input it as the core top IFA age and calculated the time represented within the sampling interval according to the mean SAR of 18 cm Kyr^{-1}. Whereas *G. ruber* $\delta^{18}O$ IFA data are available for the LH (Thirumalai et al., 2019a, 2019b), we generate a synthetic IFA distribution (Table 1) for the Early Holocene (EH) period (8,000–8.056 ka B.P., as 1 cm spans ~56 years) as a period of downcore (time slice) comparison. We do so by extracting the predicted-$\delta^{18}O$ from TraCE-21ka (He, 2011). Similar to the first example, we hold the SAR, SMLD, and abundance to be constant for both periods.

In the third example, we calculate the probability of picking older planktic foraminiferal shells which would potentially "contaminate" a pristine IFA signal at a core site in the central tropical Indian Ocean where the targeted interglacial time slices are older than the TraCE-21ka period of 21 ka (Stainbank et al., 2021b). In the original study, IFA $\delta^{18}O$ was obtained from two species of planktic foraminifera (*G. ruber*, *Trilobatus sacculifer*), where both ensuing sample distributions featured heavier than expected $\delta^{18}O$. Stainbank et al. (2021b) interpreted the heavy $\delta^{18}O$ signatures in their IFA samples in terms of reworking of glacial shells into interglacial intervals. Our analysis focuses on the *G. ruber* IFA-$\delta^{18}O$ data set from Marine Isotope Stage (MIS) 9e (U14467C-2H-6, 0–1 cm; 15–16 cm; and 18–19 cm, Stainbank et al., 2021a) and we assume that all individuals were picked in the oldest MIS9e IFA discrete depth sample (U14467C-2H-6, 18–19 cm) that is, the discrete depth closest to the MIS 10 glacial period. We apply TurbIFA to the period between 400 ka B.P. and 300 ka B.P., to ensure coverage of MIS 9e as well as of the previous glacial period (MIS 10). We then make adjustments to the site-specific SAR and SMLD and test if mixing processes could introduce reworked individuals from the nearest glacial period. To estimate the SMLD, we use the average organic carbon flux data measured at an equatorial Indian Ocean sediment trap station (Rixen et al., 2019) after applying the equation by Trauth et al. (1997) and using the dynamical SAR for the simulation time slice (Table 1). Toward this, TurBIFA uses the depths of the core where the ages are

known (Stainbank et al., 2022) and interpolates them to obtain a sedimentary record resolved at a monthly resolution, following the SEAMUS approach (Lougheed, 2020).

3. Results

3.1. Sensitivity Experiments on the Basis of Synthetic SST Time Series

In this section, we document results arising from sensitivity experiments using a 30 Kyr-long synthetic temperature time series (see Section 2). We altered the amplitude of variability within two synthetic time slices of the duration of 1 Kyr each, such that the amplitude of variability is +2× and −2× relative to the unaltered time series (Figure 2a, green and yellow rectangles). Note that as the duration of the altered event decreases, it will become more prone to mixing with time slices where the climate signal changes. The time series is then progressively mixed upwards (Figures 2a, 2d, and 2g), and consequently, the SST data archived in a discrete depth of sediment inside the active SMLD is randomly resampled (Figures 2b, 2c, 2e, 2f, 2h, and 2i). For this exercise, we use a constant SAR of 10 cm Kyr^{-1} and an SMLD of 15 cm. The simulated bioturbation front, which is represented by the contemporaneous shift of the SMLD (red rectangle) toward the present (i.e., toward the left in Figure 2), progressively marches from the bottom of the hypothetical sedimentary interval to the present and actively "bioturbates" the SST signal (regions colored in gray). We note that the impact of mixing serves to smear the signal of altered temperature intervals and modifies the signal's variance while keeping the mean intact across both altered intervals. The ensuing temperature distributions are then contrasted via Kernel density function fits (Figures 2b, 2e, and 2h) and quantile-quantile plots (Figures 2c, 2f, and 2i) after resampling (60 "foraminifera" 1,000 times) the bioturbated (red) versus the non-bioturbated (blue, green, and yellow, respectively) cases. The top panel of Figure 2 represents how the time series appears after mixing the deepest (oldest) 15 cm of the synthetic sediment core archive and resampling the SST data archived in a discrete depth of sediment within it. The ensuing distributions are virtually identical in this scenario (Figures 2a–2c), while the impact of bioturbation and smearing becomes apparent when the simulations reach the time steps at which the amplitude of the synthetic SST time series has been altered (by +2× and −2×, Figures 2c–2i). Mixing processes under the −2× scenario drive a larger change in the SST distribution values (Figures 2g–2i). This occurs because the extreme SST values are homogeneously distributed along the SMLD, increasing the spread of SST values (and relative outliers) in the discrete depths of reduced climate variability (note the departure from the 1:1 line in the Q-Q plots).

To explore the impact of sediment mixing over a time slice that features large and rapid (abrupt) change in the mean climate (temperature) state, we altered the synthetic SST time series by increasing (+2°C, Figure 3 top panel) and decreasing (−2°C, Figure 3 bottom panel) the mean temperature over two discrete time intervals, spanning 1 Kyr each. In this experiment, we use an SMLD of 5 cm and a SAR of 2 cm Kyr^{-1}. We find that the distribution of temperatures affected by sediment mixing underestimates the imposed rapid temperature changes, with apparently cooler (warmer) temperatures in the +2°C (−2°C) case (Figures 3a–2f). This occurs because bioturbation affects the IFA distribution (see mean change between distributions in Figures 3c and 3f) by mixing values from before and after the rapid (abrupt) change in the mean temperature state (Figures 3a and 3d).

Next, we explore the relationship between site-specific SAR and SMLD and the degree to which the SST in the bioturbated time interval departs from the pristine ("non-bioturbated") SST signal. To this end, we modify both the amplitude (+2×) of SST variability and the mean temperature state (+2°C) over a 1-Kyr time interval and apply TurbIFA's mixing model to time series under the following scenarios: (a) high SAR and shallow SMLD (Figures 4a–4c); (b) low SAR and deep SMLD (Figures 4d–4f); high SAR and deep SMLD (Figures 4g–4i); and (c) low SAR and shallow SMLD (Figures 4j–4l). Note that the blue Kernel distributions show resampled SST data (60 "foraminifera" each) before bioturbation, whereas the red Kernel distributions show resampled SST data after bioturbation. Across these scenarios, the impact of mixing decreases as the SAR increases relative to the SMLD (i.e., sediment accumulation overshadows the impact of sediment particle mixing). In the low SAR/deep SMLD scenario, the mean of the bioturbated SST distribution is biased toward colder temperatures, and the tails are skewed (Figures 4e and 4f). The occurrence of low SAR and deep SMLD implies that the climate conditions of both mean states are homogenized along the SMLD. Hence, the reworking of individuals from an older (colder) mean state biases the mean temperature toward apparently cooler conditions. Given that individuals are reworked from a time slice of reduced SST variability, the probability of picking extreme warm and cold individuals

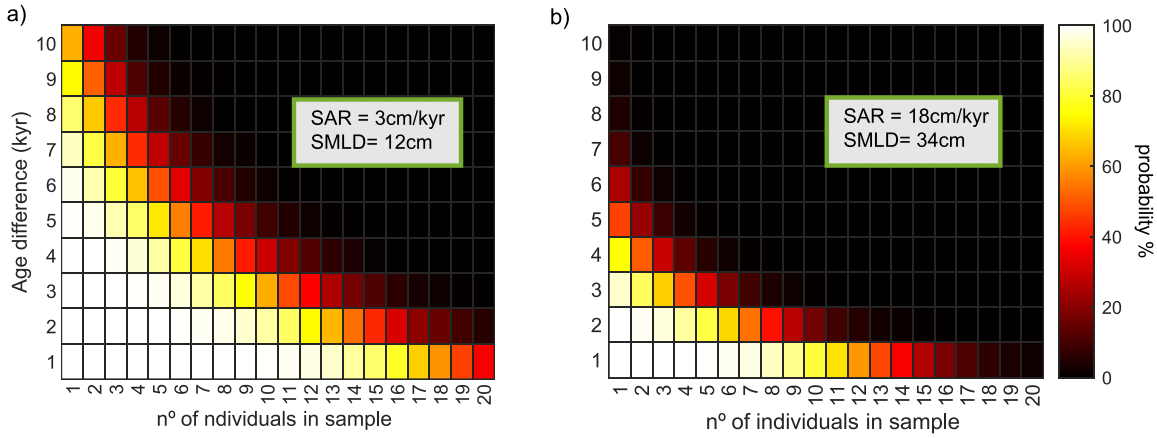

Figure 5. Heatmap charts showing the probability of picking individuals from mean states older than the IFA time slice in sediment cores from the central equatorial Pacific Ocean (MGL1208-14MC/12GC, White et al., 2017, 2018) and eastern Indian Ocean (SO189-119KL, Thirumalai et al., 2019a, 2019b). In both examples, constant SAR, SMLD, and IFA abundance are simulated. Hence foraminifera are "picked" from an exponential age distribution using a scale parameter equal to SMLD/SAR and the difference in age between each simulated "foraminiferal shell" and the mean age of the exponential distribution is computed. The probability (%) is calculated after averaging the results of 1,000 simulations of the picking process. In addition, as the number of foraminifera picked does not vary, core-top results can be used downcore (i.e., Younger Dryas [YD] and Early Holocene [EH] IFA time slices, respectively). The number of "foraminifera" picked is 68 and 62, respectively.

decreases, that is, the spread of the bioturbated distribution (Figure 4e, red Kernel density functions) is smaller than in the absence of mixing (Figure 4e, blue Kernel density functions). Therefore, in such a scenario, bioturbation may lead to an underestimation of the amplitude of paleoclimate variability.

3.2. Results From Applications of TurbIFA

Example 1 Tropical central Pacific Ocean: LH versus YD IFA-Mg/Ca-SST distributions

We applied TurbIFA to *G. ruber* IFA Mg/Ca data set from sediment core MGL1208-14MC/12GC (White et al., 2018), retrieved in the central Pacific Ocean (Table 1). We compared the core-top IFA distribution (°C) with that generated for the YD timeslice at 12.17 ka. The heatmap output provides quantitative insights into the degree of mixing of the foraminiferal shells (reflected in their age variance) in the core-top and YD IFA time slices selected for our analysis (Figure 5a). A constant SAR of 3 cm Kyr^{-1}, an SMLD of 12 cm, and an IFA sample size of 68 individuals at this location leads to a ≥80% probability of picking 3 individuals that grew during time slices 5 Kyr older than the age of the IFA sample and a ~90% of picking 1 individual that is 7 Kyr older (Figure 5a). The increase in the IFA 1σ value (°C) during the YD is significant in ~64% of the runs (Table 2), suggesting an increase in the amplitude of SST variability during the YD relative to the LH. However, the uncertainty of the technique itself can likely explain the overall changes in the shape of the IFA distribution during the YD (Figure 6a).

Table 2
Uncertainty in IFA 1 − σ Value and Probability That Reconstructed Changes Are Significant (at the 95% Confidence Level) for the Case Studies of White et al. (2018) and Pseudo-EH (Thirumalai et al., 2019b)

	Core top (°C)	Paleo (°C)	Core top (‰)	Paleo (‰)
	MGL1208-14MC /12GC	MGL1208-14MC /12GC (°C)	SO189 −119KL	SO189 −119K
IFA (1σ)	1.395	1.782	0.232	0.223
Uncertainty (2.5th/97.5th)	−0.235/0.236	−0.233/0.239	−0.029/0.038	−0.074/0.085
Prob. increase (%)		63.6		1.79
Prob. decrease (%)		0		3.47

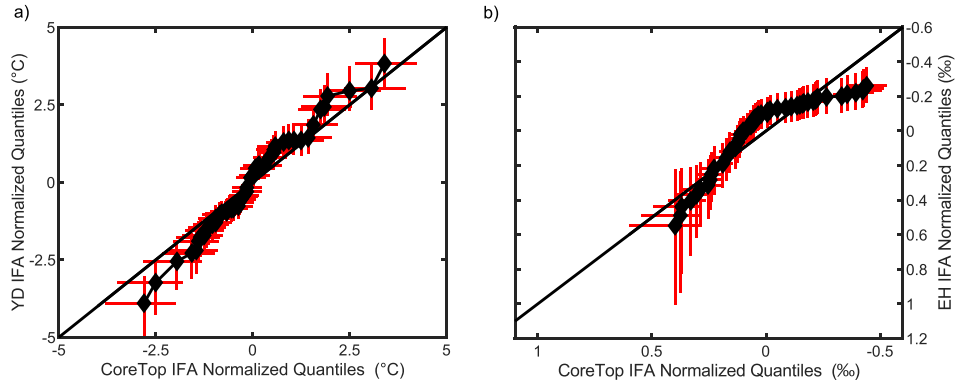

Figure 6. Normalized (mean subtracted) quantile-quantile plots of paleo and core-top distributions for (a) Mg/Ca (°C) IFA data from the central equatorial Pacific Ocean (White et al., 2017, 2018) and (b) $\delta^{18}O$ IFA data from the eastern Indian Ocean (Thirumalai et al., 2019a, 2019b). Red bars show the uncertainty (95% confidence level) related to the number of individuals analyzed (68 and 62, respectively) and instrumental error (0.5°C and 0.05‰, respectively) after performing QUANTIFA's bootstrap resampling and binning routine (Glaubke et al., 2021). Note that we have not propagated the Mg/Ca calibration error, as it not appears to be an extra source of noise (White et al., 2018).

Example 2 Tropical eastern Indian Ocean: LH versus EH IFA$\delta^{18}O$ distributions

We applied TurbIFA to $\delta^{18}O$ IFA data from a high sediment accumulation location (core SO189-119KL) in the Indonesian Margin, eastern Indian Ocean. We use the modern (LH) IFA $\delta^{18}O$ data set from Thirumalai et al. (2019a, 2019b) and compare it with simulated data for the EH, extracted from TraCE-21ka (He, 2011). In this case, we impose an extremely deep (unrealistic) SMLD of 34 cm as input for a sample size of 62 individuals (see Section 2 for more details). This leads to a ~90% probability of picking at least four individuals 2-Kyr older than the analyzed IFA time slice (Figure 5b). According to the TraCE-21ka output, the amplitude of climate variability during the EH was similar (~0.22‰, $1-\sigma$) relative to that inferred from the measured LH IFA data (~0.23‰, $1-\sigma$). TurbIFA demonstrates that changes in the spread (1σ) of the EH "IFA distribution" are unlikely to be significant (Table 2). In contrast with the White et al. (2018) example above, the uncertainty at the cold tail (heavier $\delta^{18}O$ values) of EH IFA distribution is quite larger than in the warm tail (Figure 6b). This suggests that sediment-mixing processes have the potential to introduce some enriched $\delta^{18}O$ values that contaminate the sampled IFA interval for the EH discrete depth.

Example 3 Tropical central Indian Ocean: bioturbation during Marine Isotope Stage 9

Here we focus on calculating the probability of glacial foraminiferal shells being reworked into an IFA-$\delta^{18}O$ distribution generated for an interglacial (MIS9) time-slice at a location in the central Indian Ocean IODP Site U1467C. We applied a subroutine of the TurbIFA package (employing the *bioprob* script; see Supporting Information S1 for details) as the targeted time slices of the original study (Stainbank et al., 2021b) are older than the TraCE-21ka timespan. We simulated a constant SMLD of 5 cm and dynamical SAR by imposing the long-term age-depth values used for pooled *G. ruber* $\delta^{18}O$ reconstructions for cores U1467C-B between 300 and 400 ka. The $\delta^{18}O$ stratigraphy of U1467 suggests that the end of glacial MIS10 occurred at ~340 ka, an age that is ~9 Kyr older than the MIS9e IFA discrete depth considered here. Our results suggest that the reworking of glacial-aged individuals is unlikely to explain the heavy $\delta^{18}O$ values observed during MIS 9e (Figure 7b), with virtually zero probability of mixing individuals from beyond ~8 Kyrs.

4. Discussion

Particle mixing is site- and time-dependent, which implies that the signal distortion in a bioturbated versus a non-bioturbated sedimentary archive can be neither interpolated between locations and time slices nor inferred from the regional oceanographic and sedimentary conditions (Dolman et al., 2021; Lougheed & Metcalfe, 2022). Through the analysis of idealized scenarios (Section 3.1) and application of TurbIFA to real-world IFA data sets (Section 3.2), we show that the sensitivity of $\delta^{18}O$ and/or Mg/Ca-SST IFA distributions to local changes in SAR and SLMD at a given site can be quantitatively evaluated (Figures 2–5).

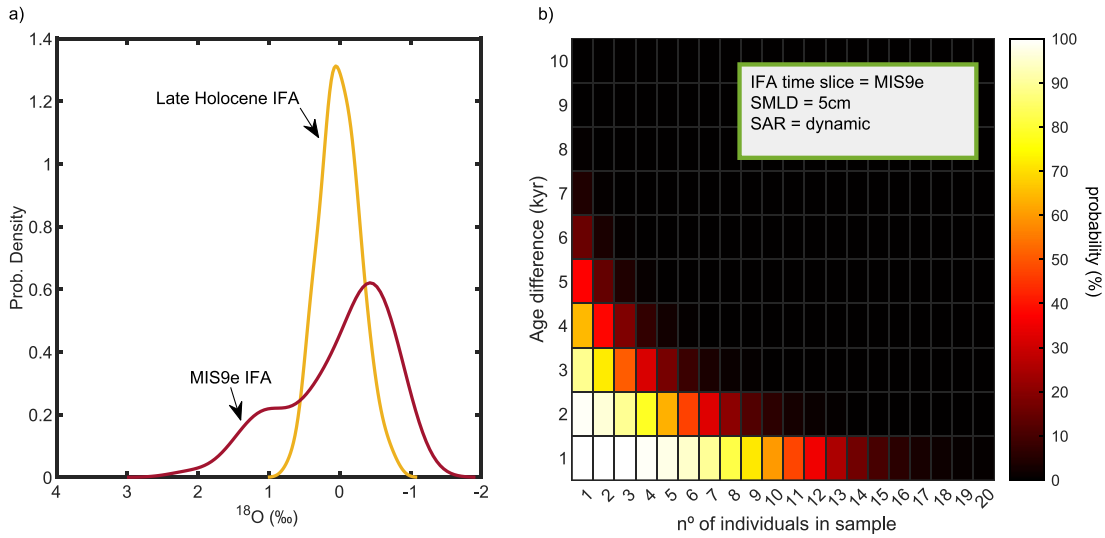

Figure 7. Probability of picking reworked individuals in a time slice older than the last glacial maximum (LGM). The study targeted a core located in the central equatorial Indian Ocean and marine isotope stage (MIS) 9e interglacial (Stainbank et al., 2021b, foraminiferal shells picked, $n = 83$). Kernel density functions (mean subtracted) represent the Late Holocene (LH, yellow line) and MIS9e (red line) IFA-δ^{18}O distributions (Stainbank et al., 2021a). We generated a monthly resolved time series spanning the 400 ka B.P.–300 ka B.P. interval, bioturbated it, and repeated the picking process 1,000 times to then calculate the difference in age between each simulated "foraminiferal shell" and the mean age of the discrete depth from which they were picked. The input data used in this example are reported in Table 1. Statistical analysis was performed using the *bioprob* script (see Text S4 in Supporting Information S1 for further information about the use of *bioprob* MATLAB code routine).

Our sensitivity experiments rely on bioturbation simulations using synthetic long-term (30 Kyr) SST time series. Three idealized scenarios in which the amplitude of climate variability, the mean climate state, and SMLD/SAR are modified to asses their separate and cumulative impacts on IFA distributions. These experiments point to an increase (decrease) in the bias tied to sediment particle mixing when the SAR decreases (increases) relative to the SMLD (Figure 4). In low-latitude, open-ocean settings that generally feature low SAR (Restreppo et al., 2020), bioturbation is thus expected to mix foraminiferal shells that grew under conditions characterized by either different amplitude of climatic variability (Figure 2) or different mean (underlying) climate state (Figure 3).

In a central equatorial Pacific Ocean location (White et al., 2018) where deep SMLD (12 cm) and low SAR (3 cm Kyr^{-1}) occur, we find that both core-top (approximating modern conditions) and EH IFA data are potentially contaminated by a mixture of foraminiferal shells from different mean climate states. We evaluate a "high SAR/deep SMLD" scenario by simulating a high SAR (18 cm Kyr^{-1}) and an unrealistically deep SMLD (34 cm) to a site in the eastern Indian Ocean (SO189-119KL core). These input data favor the displacement of deglacial "shells" into the EH time slice. Given that many IFA-based reconstructions are interpreted in quantile-quantile space (e.g., Ford et al., 2015; White et al., 2018), we have evaluated the probability for each quantile of the IFA distribution at these two locations to document an increase (or a decrease) of variability before and after bioturbating the synthetic sediment core archives (Figure 8), which also exemplifies the potentially different impacts of biturbation on Mg/Ca-SST (°C) and δ^{18}O IFA data sets. We note that future users of TurbIFA may generate similar site-specific statistical interpretations by using their own IFA data and assessing the significance and sensitivity of their core-top and downcore signals.

In the Mg/Ca-SST IFA data set of White et al. (2018), the ability of the IFA distribution to reflect "true" variability changes during the YD appears to be unbiased by bioturbation (Figures 8a–8c). This can be plausibly attributed to the combination of (a) subdued glacial-interglacial (mean state) temperature changes in (non-upwelling) equatorial regions, such as the central equatorial Pacific Ocean (Monteagudo et al., 2021) and (b) small changes in temperature variability overtime at the core location as simulated by TraCE-21ka (He, 2011). It suggests that IFA-based reconstructions in low SAR areas can be reliable, while underscoring the importance of exploring and understanding the potential impacts of bioturbation and of rigorously propagating the errors associated with the IFA method at each specific site. In the δ^{18}O IFA data set of Thirumalai et al. (2019a, 2019b), bioturbation introduces heavier δ^{18}O values from older-than-EH time slices in the IFA discrete depth (Figure 8d),

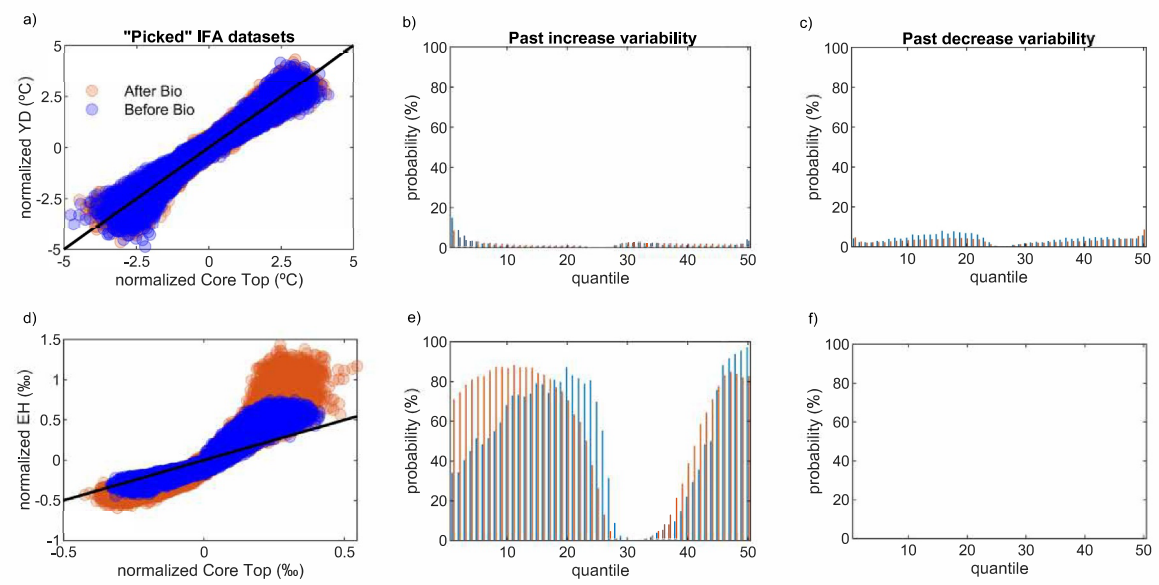

Figure 8. Probability for a given quantile to document a significant increase or decrease in ocean and climate variability after resampling pseudo-IFA data sets before and after bioturbation in a synthetic sediment core archive. (a–c): core location of White et al. (2017, 2018). (d–f): core location of Thirumalai et al. (2019a, 2019b). Non-bioturbated data (blue) is resampled within the time slice covered by the IFA data set (Table 1). Bioturbated data (in orange) is resampled at discrete sediment depths, where the mean age after bioturbation closely matches the mean age of IFA sample (Table 1). The statistical threshold that we used corresponds to the error (at the 95% confidence level) obtained from random resampling (following QUANTIFA's routine) the SST or $\delta^{18}O$ data archived at the IFA time slice (non-bioturbated) or discrete depth (bioturbated). Subsequently, a deviation from a specific quantile (targeted IFA time slice vs. core top) is considered significant if it exceeds the 95% confidence limit. The probability values displayed in the plot represent the median of the results from 50 runs of the sediment-mixing model.

which results in a larger spread (hence error, Figure 6b). Note that $\delta^{18}O$ IFA data sets are more prone (than Mg/Ca) to be biased by bioturbation due to glacial-interglacial changes in the $\delta^{18}O_{sw}$ of seawater associated with changes in ice volume (Schrag et al., 2002). Despite bioturbation, the overall results do not change and indicate significant, extremely heavier $\delta^{18}O$ values during the EH (Figure 8e, higher quantiles) compared to modern conditions with a probability of ~80% (~100% before bioturbation). However, the magnitude of this change is notably greater when bioturbation processes are accounted for, which challenges the assessment of changes in the amplitude of climate variability over time, for example, when comparing IFA data sets collected at glacial (or interglacial) maxima. The spread of the warm tail (lighter $\delta^{18}O$) notably increases after bioturbation. This leads to an apparent increase in climate variability (Figure 8e, lower quantiles), with the ability of IFA to reflect the true amplitude of warm events during the EH period being potentially biased due to bioturbation.

The IFA-$\delta^{18}O$ data set for *G. ruber* at interglacial marine isotope stage (MIS) 9e feature some considerably heavier (by ~2‰) than expected $\delta^{18}O$ values (Figure 7a), leading to the hypothesis that substantial contamination by glacial specimens (>20 individuals out of the 83 analyzed) may have occurred (Stainbank et al., 2021b). We find that it is virtually impossible to pick one or more individuals with an age ~9 Kyr older than the MIS9e IFA time slice (Figure 7b). Given the proximity of the selected IFA sample with the end of Termination IV (~1–2 Kyr), it seems more likely (~80% probability) that some ($n = 8$) deglacial (rather than glacial) individuals may have contributed to the extreme values. Changes in the foraminiferal depth habitat and/or changes in the $\delta^{18}O_{sw}$ under monsoon intensification at interglacial maxima (Wang et al., 2014) may be invoked to explain almost a part of the unexpectedly heavy $\delta^{18}O$ IFA values found in the low SAR location.

5. Conclusions

The TurbIFA software that we present here combines existing algorithms (SEAMUS, QUANTIFA, INFAUNAL) to quantify the impacts of sediment particle mixing (bioturbation) on $\delta^{18}O$ and/or Mg/Ca-temperature IFA data sets obtained from mixed-layer planktic foraminifera for time slices between the LGM and present. The software simulates the process of bioturbation by mixing synthetic time series of temperature (Mg/Ca) and $\delta^{18}O$ obtained from forward-modeled surface ocean parameters to quantitatively determine how changes in the amplitude of

climate variability, mean climate state, sedimentation rate, and in the depth of the sediment mixing process alter the "true" SST signal archived in an IFA sample. Next, TurbIFA propagates the uncertainty inherent to the IFA method (i.e., size of the data set, analytical) to determine the controls and statistical significance of the IFA-based climate variability reconstructions in the face of bioturbation. Application of TurbIFA to existing IFA data sets quantifies the degree to which sediment mixing impacts the uncertainty of the climate variability reconstructions at a given site, suggesting that uncertainties also depend on the climate variability within and adjacent to the targeted time slice. TurbIFA, therefore, provides a key tool to obtain a more comprehensive (and rigorous) analysis of uncertainties associated with the IFA approach. We contend that this has the potential to deepen our quantitative understanding of the relationship between the amplitude and mode of climate variability and the underlying climate state.

Data Availability Statement

TurbIFA model, the code routines used for the sensitivity experiments, and the IFA data used as case studies are available in Zenodo (Bienzobas Montavez et al., 2024) and GitHub https://github.com/Nbienzobas/TurbIFA. The individual foraminiferal data sets used to reproduce the figures were downloaded from the: PANGAEA database (de Moel et al., 2010; Groeneveld et al., 2019b; Stainbank et al., 2021a); NOAA Paleoclimate database (Koutavas & Joanides, 2012b; Rustic et al., 2016; Thirumalai et al., 2019a; White et al., 2017); and article (Wit et al., 2010). The age model for core 359-U1467C was retrieved from the PANGAEA database (Stainbank et al., 2022). TraCE-21ka surface air temperature and sea surface salinity simulation outputs were downloaded from the Climate Data Gateway portal (https://www.earthsystemgrid.org/project/trace.html). ECMWF's ORAS5 sea surface temperature and sea surface salinity reanalysis data were downloaded from the Integrated Climate Data Centre, Hamburg University (https://icdc.cen.uni-hamburg.de/thredds/catalog/ftpthredds/EASYInit/oras5/catalog.html).

Acknowledgments
We acknowledge the comments of Bryan Lougheed, three anonymous reviewers, the Associate Editor, and the Editor (Matthew Huber), who have contributed to improve this manuscript. This work contributes to the project PID2019-109653RB-I00 (N.B. M., K.T., and G.M., TRIOPACC, Ministerio de Ciencia e Innovación, Spain). G.M. acknowledges support from the Universidade de Vigo's programme to attract excellent research talent (RR04092017), a Beatriz Galindo Fellowship (BG20-00157, Ministerio de Ciencia e Innovación, Spain), and a generous start-up package. KT acknowledges support from National Science Foundation Grant 1903482. Funding for open access charge: Universidade de Vigo/CISUG.

References

Anand, P., Elderfield, H., & Conte, M. H. (2003). Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography*, *18*(2), 1050. https://doi.org/10.1029/2002pa000846

Bemis, B. E., Spero, H. J., Bijma, J., & Lea, D. W. (1998). Reevaluation of the oxygen isotopic composition of planktonic foraminifera: Experimental results and revised paleotemperature equations. *Paleoceanography*, *13*(2), 150–160. https://doi.org/10.1029/98pa00070

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

Bienzobas Montavez, N., Thirumalai, K., & Marino, G. (2024). TurbIFA model (version v3.1) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.11091286

Brummer, G.-J. A., Metcalfe, B., Feldmeijer, W., Prins, M. A., Van'T Hoff, J., & Ganssen, G. M. (2020). Modal shift in North Atlantic seasonality during the last deglaciation. *Climate of the Past*, *16*(1), 265–282. https://doi.org/10.5194/cp-16-265-2020

Buitenhuis, E. T., Hashioka, T., & Quéré, C. L. (2013). Combined constraints on global ocean primary production using observations and models. *Global Biogeochemical Cycles*, *27*(3), 847–858. https://doi.org/10.1002/gbc.20074

Conroy, J. L., Thompson, D. M., Cobb, K. M., Noone, D., Rea, S., & Legrande, A. N. (2017). Spatiotemporal variability in the δ^{18}O-salinity relationship of seawater across the tropical Pacific ocean. *Paleoceanography*, *32*(5), 484–497. https://doi.org/10.1002/2016pa003073

Dekens, P. S., Lea, D. W., Pak, D. K., & Spero, H. J. (2002). Core top calibration of Mg/Ca in tropical foraminifera: Refining paleotemperature estimation. *Geochemistry, Geophysics, Geosystems*, *3*(4), 1–29. https://doi.org/10.1029/2001gc000200

De Moel, H., Ganssen, G., Peeters, F., Jung, S., Kroon, D., Brummer, G., & Zeebe, R. (2009). Planktic foraminiferal shell thinning in the Arabian Sea due to anthropogenic ocean acidification? *Biogeosciences*, *6*(9), 1917–1925. https://doi.org/10.5194/bg-6-1917-2009

de Moel, H., Ganssen, G. M., Peeters, F. J. C., Jung, S. J. A., Kroon, D., Brummer, G.-J. A., & Zeebe, R. E. (2010). Stable isotope measurements on *G. ruber* from the core top of BC21WP7 [Dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.746071

Dolman, A. M., Groeneveld, J., Mollenhauer, G., Ho, S. L., & Laepple, T. (2021). Estimating bioturbation from replicated small-sample radiocarbon ages. *Paleoceanography and Paleoclimatology*, *36*(7), e2020PA004142. https://doi.org/10.1029/2020pa004142

Dolman, A. M., & Laepple, T. (2018). Sedproxy: A forward model for sediment-archived climate proxies. *Climate of the Past*, *14*(12), 1851–1868. https://doi.org/10.5194/cp-14-1851-2018

Ford, H. L., Ravelo, A. C., & Polissar, P. J. (2015). Reduced El Niño–Southern Oscillation during the last glacial maximum. *Science*, *347*(6219), 255–258. https://doi.org/10.1126/science.1258437

Glaubke, R. H., Thirumalai, K., Schmidt, M. W., & Hertzberg, J. E. (2021). Discerning changes in high-frequency climate variability using geochemical populations of individual foraminifera. *Paleoceanography and Paleoclimatology*, *36*(2), e2020PA004065. https://doi.org/10.1029/2020pa004065

Gray, W. R., & Evans, D. (2019). Nonthermal influences on Mg/Ca in planktonic foraminifera: A review of culture studies and application to the last glacial maximum. *Paleoceanography and Paleoclimatology*, *34*(3), 306–315. https://doi.org/10.1029/2018pa003517

Groeneveld, J., Ho, S. L., Mackensen, A., Mohtadi, M., & Laepple, T. (2019a). Deciphering the variability in mg/ca and stable oxygen isotopes of individual foraminifera. *Paleoceanography and Paleoclimatology*, *34*(5), 755–773. https://doi.org/10.1029/2018pa003533

Groeneveld, J., Ho, S. L., Mackensen, A., Mohtadi, M., & Laepple, T. (2019b). Mg/Ca and stable carbon and oxygen isotopes on individual and pooled foraminifera [Dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.904559

He, F. (2011). Simulating transient climate evolution of the last deglaciation with CCSM 3. *72*(10).

Hülse, D., Vervoort, P., van de Velde, S. J., Kanzaki, Y., Boudreau, B., Arndt, S., et al. (2022). Assessing the impact of bioturbation on sedimentary isotopic records through numerical models. *Earth-Science Reviews*, *234*, 104213. https://doi.org/10.1016/j.earscirev.2022.104213

Hut, G. (1987). Consultants' group meeting on stable isotope reference samples for geochemical and hydrological investigations.
Khider, D., Stott, L., Emile-Geay, J., Thunell, R., & Hammond, D. (2011). Assessing El Niño Southern Oscillation variability during the past millennium. *Paleoceanography*, *26*(3), PA3222. https://doi.org/10.1029/2011pa002139
Kısakürek, B., Eisenhauer, A., Böhm, F., Garbe-Schönberg, D., & Erez, J. (2008). 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. https://doi.org/10.1016/j.epsl.2008.06.026
Kiss, P., Jonkers, L., Hudáčková, N., Reuter, R., Donner, B., Fischer, G., & Kucera, M. (2021). Determinants of planktonic foraminifera calcite flux: Implications for the prediction of intra-and inter-annual pelagic carbonate budgets. *Global Biogeochemical Cycles*, *35*(9), e2020GB006748. https://doi.org/10.1029/2020gb006748
Koutavas, A., Demenocal, P. B., Olive, G. C., & Lynch-Stieglitz, J. (2006). Mid-Holocene El Niño–Southern Oscillation (ENSO) attenuation revealed by individual foraminifera in eastern tropical pacific sediments. *Geology*, *34*(12), 993–996. https://doi.org/10.1130/g22810a.1
Koutavas, A., & Joanides, S. (2012a). El Niño–Southern Oscillation extrema in the Holocene and last glacial maximum. *Paleoceanography*, *27*(4), PA4208. https://doi.org/10.1029/2012pa002378
Koutavas, A., & Joanides, S. (2012b). NOAA/WDS paleoclimatology—Eastern equatorial Pacific LGM-Holocene *G.ruber* δ^{18}O data [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.25921/qznc-m246
Kucera, M. (2007). Chapter six planktonic foraminifera as tracers of past oceanic environments. *Developments in Marine Geology*, *1*, 213–262. https://doi.org/10.1016/s1572-5480(07)01011-1
Laepple, T., Ziegler, E., Weitzel, N., Hébert, R., Ellerhoff, B., Schoch, P., et al. (2023). Regional but not global temperature variability underestimated by climate models at supradecadal timescales. *Nature Geoscience*, *16*(11), 958–966. https://doi.org/10.1038/s41561-023-01299-9
Lambeck, K., Rouby, H., Purcell, A., Sun, Y., & Sambridge, M. (2014). Sea level and global ice volumes from the last glacial maximum to the Holocene. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), 15296–15303. https://doi.org/10.1073/pnas.1411762111
Lanzante, J. R. (2005). A cautionary note on the use of error bars. *Journal of Climate*, *18*(17), 3699–3703. https://doi.org/10.1175/jcli3499.1
Leduc, G., Vidal, L., Cartapanis, O., & Bard, E. (2009). Modes of eastern equatorial Pacific thermocline variability: Implications for ENSO dynamics over the last glacial period. *Paleoceanography*, *24*(3), PA3202. https://doi.org/10.1029/2008pa001701
LeGrande, A. N., & Schmidt, G. A. (2006). Global gridded data set of the oxygen isotopic composition in seawater. *Geophysical Research Letters*, *33*(12), L12604. https://doi.org/10.1029/2006gl026011
Lougheed, B. C. (2020). Seamus (v1. 20): A δ^{14}C-enabled, single-specimen sediment accumulation simulator. *Geoscientific Model Development*, *13*(1), 155–168. https://doi.org/10.5194/gmd-13-155-2020
Lougheed, B. C., Ascough, P., Dolman, A. M., Löwemark, L., & Metcalfe, B. (2020). Re-evaluating ^{14}C dating accuracy in deep-sea sediment archives. *Geochronology*, *2*(1), 17–31. https://doi.org/10.5194/gchron-2-17-2020
Lougheed, B. C., & Metcalfe, B. (2022). Testing the effect of bioturbation and species abundance upon discrete-depth individual foraminifera analysis. *Biogeosciences*, *19*(4), 1195–1209. https://doi.org/10.5194/bg-19-1195-2022
Lougheed, B. C., Metcalfe, B., Ninnemann, U. S., & Wacker, L. (2018). 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. https://doi.org/10.5194/cp-14-515-2018
Metcalfe, B., Feldmeijer, W., & Ganssen, G. M. (2019). Oxygen isotope variability of planktonic foraminifera provide clues to past upper ocean seasonal variability. *Paleoceanography and Paleoclimatology*, *34*(3), 374–393. https://doi.org/10.1029/2018pa003475
Mohtadi, M., Prange, M., Oppo, D. W., Pol-Holz, D., Merkel, U., Zhang, X., et al. (2014). North Atlantic forcing of tropical Indian Ocean climate. *Nature*, *509*(7498), 76–80. https://doi.org/10.1038/nature13196
Monteagudo, M. M., Lynch-Stieglitz, J., Marchitto, T. M., & Schmidt, M. W. (2021). Central equatorial Pacific cooling during the last glacial maximum. *Geophysical Research Letters*, *48*(3), e2020GL088592. https://doi.org/10.1029/2020gl088592
Müller, P. J., & Suess, E. (1979). Productivity, sedimentation rate, and sedimentary organic matter in the oceans—I. Organic carbon preservation. *Deep-Sea Research, Part A: Oceanographic Research Papers*, *26*(12), 1347–1362. https://doi.org/10.1016/0198-0149(79)90003-7
Naidu, P. D., Niitsuma, N., Thirumalai, K., & Naik, S. S. (2019). Significant seasonal contrast in the Arabian Sea during deglaciation: Evidence from oxygen isotopic analyses of individual planktic foraminifera. *Quaternary International*, *503*, 163–169. https://doi.org/10.1016/j.quaint.2018.08.005
Restreppo, G. A., Wood, W. T., & Phrampus, B. J. (2020). Oceanic sediment accumulation rates predicted via machine learning algorithm: Towards sediment characterization on a global scale. *Geo-Marine Letters*, *40*(5), 755–763. https://doi.org/10.1007/s00367-020-00669-1
Rixen, T., Gaye, B., & Emeis, K.-C. (2019). The monsoon, carbon fluxes, and the organic carbon pump in the northern Indian Ocean. *Progress in Oceanography*, *175*, 24–39. https://doi.org/10.1016/j.pocean.2019.03.001
Rongstad, B. L., Marchitto, T. M., Serrato Marks, G., Koutavas, A., Mekik, F., & Ravelo, A. C. (2020). Investigating ENSO-related temperature variability in equatorial Pacific core-tops using Mg/Ca in individual planktic foraminifera. *Paleoceanography and Paleoclimatology*, *35*(2), e2019PA003774. https://doi.org/10.1029/2019pa003774
Rosenthal, Y., Bova, S., & Zhou, X. (2022). A user guide for choosing planktic foraminiferal Mg/Ca-temperature calibrations. *Paleoceanography and Paleoclimatology*, *37*(6), e2022PA004413. https://doi.org/10.1029/2022pa004413
Rustic, G. T., Koutavas, A., Marchitto, T. M., & Linsley, B. K. (2015). Dynamical excitation of the tropical Pacific Ocean and ENSO variability by little ice age cooling. *Science*, *350*(6267), 1537–1541. https://doi.org/10.1126/science.aac9937
Rustic, G. T., Koutavas, A., Marchitto, T. M., & Linsley, B. K. (2016). NOAA/WDS paleoclimatology—Eastern tropical Pacific 1,000 year foraminferal δ^{18}O and Mg/Ca SST [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.25921/9ngn-ch10
Rustic, G. T., Polissar, P. J., Ravelo, A. C., & White, S. M. (2020). Modulation of late Pleistocene ENSO strength by the tropical Pacific thermocline. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-19161-6
Schrag, D. P., Adkins, J. F., McIntyre, K., Alexander, J. L., Hodell, D. A., Charles, C. D., & McManus, J. F. (2002). The oxygen isotopic composition of seawater during the last glacial maximum. *Quaternary Science Reviews*, *21*(1–3), 331–342. https://doi.org/10.1016/s0277-3791(01)00110-x
Scroxton, N., Bonham, S. G., Rickaby, R. E., Lawrence, S., Hermoso, M., & Haywood, A. M. (2011). Persistent El Niño–Southern Oscillation variation during the Pliocene epoch. *Paleoceanography*, *26*(2), PA2215. https://doi.org/10.1029/2010pa002097
Singh, A., Jani, R., & Ramesh, R. (2010). Spatiotemporal variations of the δ^{18}O–salinity relation in the northern Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *57*(11), 1422–1431. https://doi.org/10.1016/j.dsr.2010.08.002
Slowey, N. C., & Curry, W. B. (1995). Glacial-interglacial differences in circulation and carbon cycling within the upper western North Atlantic. *Paleoceanography*, *10*(4), 715–732.

Solan, M., Ward, E. R., White, E. L., Hibberd, E. E., Cassidy, C., Schuster, J. M., et al. (2019). Worldwide measurements of bioturbation intensity, ventilation rate, and the mixing depth of marine sediments. *Scientific Data*, *6*(1), 1–6. https://doi.org/10.1038/s41597-019-0069-7

Stainbank, S., Kroon, D., de Leau, E. S., & Spezzaferri, S. (2021a). Individual foraminiferal analysis (IFA) geochemical data generated from IODP Site 359-U1467 [Dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.933025

Stainbank, S., Kroon, D., de Leau, E. S., & Spezzaferri, S. (2021b). Using past interglacial temperature maxima to explore transgressions in modern Maldivian coral and Amphistegina bleaching thresholds. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-89697-0

Stainbank, S., Spezzaferri, S., De Boever, E., Bouvier, A.-S., Chilcott, C., de Leau, E. S., et al. (2022). Long-term, high-resolution foraminiferal geochemical records ($\delta^{18}O$, $\delta^{13}C$) from IODP site 359-U1467 [Dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.939911

Teal, L., Bulling, M. T., Parker, E., & Solan, M. (2008). Global patterns of bioturbation intensity and mixed depth of marine soft sediments. *Aquatic Biology*, *2*(3), 207–218. https://doi.org/10.3354/ab00052

Thirumalai, K., DiNezio, P., Tierney, J., Puy, M., & Mohtadi, M. (2019a). NOAA/WDS paleoclimatology—Eastern tropical Indian Ocean *G. ruber* stable isotope data from the Late Holocene and LGM [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.25921/b2w8-3929

Thirumalai, K., DiNezio, P. N., Tierney, J. E., Puy, M., & Mohtadi, M. (2019b). An El Niño mode in the glacial Indian Ocean? *Paleoceanography and Paleoclimatology*, *34*(8), 1316–1327. https://doi.org/10.1029/2019pa003669

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

Thirumalai, K., Quinn, T. M., & Marino, G. (2016). Constraining past seawater $\delta^{18}O$ and temperature records developed from foraminiferal geochemistry. *Paleoceanography*, *31*(10), 1409–1422. https://doi.org/10.1002/2016pa002970

Tierney, J. E., Malevich, S. B., Gray, W., Vetter, L., & Thirumalai, K. (2019). Bayesian calibration of the Mg/Ca paleothermometer in planktic foraminifera. *Paleoceanography and Paleoclimatology*, *34*(12), 2005–2030. https://doi.org/10.1029/2019pa003744

Trauth, M. H. (2013). Turbo2: A Matlab simulation to study the effects of bioturbation on paleoceanographic time series. *Computers & Geosciences*, *61*, 1–10. https://doi.org/10.1016/j.cageo.2013.05.003

Trauth, M. H., Sarnthein, M., & Arnold, M. (1997). Bioturbational mixing depth and carbon flux at the seafloor. *Paleoceanography*, *12*(3), 517–526. https://doi.org/10.1029/97pa00722

Wang, P., Wang, B., Cheng, H., Fasullo, J., Guo, Z., Kiefer, T., & Liu, Z. (2014). The global monsoon across timescales: Coherent variability of regional monsoons. *Climate of the Past*, *10*(6), 2007–2052. https://doi.org/10.5194/cp-10-2007-2014

White, S. M., Ravelo, A. C., & Polissar, P. J. (2017). NOAA/WDS Paleoclimatology—Line Islands Mg/Ca and sea surface temperature reconstruction in the early and mid-Holocene [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.25921/7254-qv56

White, S. M., Ravelo, A. C., & Polissar, P. J. (2018). 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. https://doi.org/10.1002/2017gl075433

Wit, J., Reichart, G.-J., A Jung, S., & Kroon, D. (2010). Approaches to unravel seasonality in sea surface temperatures using paired single-specimen foraminiferal $\delta^{18}O$ and Mg/Ca analyses. *Paleoceanography*, *25*(4), PA4220. https://doi.org/10.1029/2009pa001857

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

Zuhr, A. M., Dolman, A. M., Ho, S. L., Groeneveld, J., Löwemark, L., Grotheer, H., et al. (2022). Age-heterogeneity in marine sediments revealed by three-dimensional high-resolution radiocarbon measurements. *Frontiers in Earth Science*, *10*, 871902. https://doi.org/10.3389/feart.2022.871902

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, *15*(3), 779–808. https://doi.org/10.5194/os-15-779-2019