UNIVERSITY OF SOUTHAMPTON

Faculty of Environmental and Life Sciences
School of Ocean and Earth Science

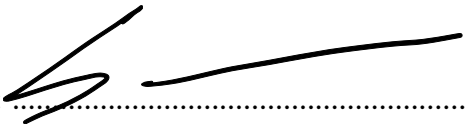
QPID: A new palaeoclimate database, and an exploration of ocean remineralisation depth

Thomas A. J. Arney
B.Sc.
ORCiD: 0000-0003-4380-4079

A dissertation submitted in partial fulfilment of the requirements for the degree of M.Sc. (Oceanography) by instructional course

September 2020

As the nominated University supervisor of this M.Sc. project by Thomas Arney, I confirm that I have had the opportunity to comment on earlier drafts of the report prior to submission of the dissertation for consideration of the award of M.Sc. Oceanography.

Signed:  .....

Supervisor's name: Gavin L. Foster

Abstract

The ocean carbon cycle is a key driver of Earth's climate, and yet there are still uncertainties about its influences and effects, and how it changes over time. One important remaining question is whether the temperature dependency of the rate of microbial respiration in the ocean has a significant impact on the efficiency of the biological pump. This process, which sequesters CO_2 away from the atmosphere for thousands of years, is a key uncertainty in models forecasting climate change into the future. Organic matter is decomposed by microbes as it sinks through the ocean, recycling nutrients and carbon back into the water column (a process termed *remineralisation*). The depth at which this remineralisation occurs has a direct effect on surface water carbon content and therefore atmospheric CO_2 concentrations. To investigate this, a new large database was compiled: the Quaternary Palaeoclimate Isotope Database (QPID), containing $>$140 000 measurements of planktic and benthic foraminiferal stable oxygen and carbon isotopes from 396 globally distributed marine sediment cores. This dataset was used to reconstruct the $\delta^{13}C$ gradient, a proxy for nutrient content of the upper water column, in the Holocene and the LGM. No significant differences in gradient were found, either as a global average or in the Atlantic or Pacific Oceans. A lack of data due to sparse coverage, vital effects, and small expected signals may account for this null result. This work also includes reflections on the state of "big data" in palaeoceanography, and the current efforts aiming to improve the lack of standardisation and data availability which currently hinders large syntheses and meta analyses of palaeoclimate data.

Contents

1 Introduction

Anthropogenic climate change is leading us into a significantly warmer future: the addition of ~2000 Gt of CO_2 to Earth's carbon cycle has already disturbed the climate system, and will continue to do so with increasing effect (IPCC, 2014). Understanding the response of the ocean system to such forcing is vital in our attempts at predicting our warm future.

Identifying the important drivers and feedbacks in the climate system is an active area of research, and an important source of data about the behaviour of such a complex system is how it has changed in the geological past. On the timescale of glacial-interglacial cycles, the ocean is a key control on atmospheric carbon dioxide concentrations ($pCO_{2_{atm}}$), since the carbon reservoir is orders of magnitude larger than the terrestrial biosphere or atmosphere, and the only one capable of changes on this timescale (Broecker, 1982). Although the orbital dynamics of precession, obliquity and eccentricity are considered the primary control on Earth's climate variations on a millennial timescale, they cannot explain the full variation of timing or extent of glacial-interglacial cycles (Sigman and Boyle, 2000); to do so, the role of other controls on the carbon cycle and their relative importance must be established. One such control may be the proportion of organic matter exported from the upper water column, and how and why this varies (Matsumoto, 2007; Chikamoto et al., 2012).

1.1 The ocean carbon cycle and remineralisation

The ocean carbon cycle is a complex system of interacting chemical, physical and biological processes. The carbon sinks (those that remove carbon away from the atmosphere for geologically significant periods of time) can be simplified to three downward fluxes: first defined in Volk and Hoffert (1985) as the carbonate pump, physical (solubility) pump, and biological carbon pump (see Figure 1).

The surface waters of the ocean are the interface between the important reservoirs of oceanic and atmospheric carbon. At this boundary, gaseous and aqueous CO_2 equilibrate, the first step in the inorganic carbon chemistry of seawater (Equation 1). The bicarbonate ion is introduced in the intermediate step as the aqueous CO_2 dissociates into carbonic acid, which dissociates further into the carbonate ion, forming the other

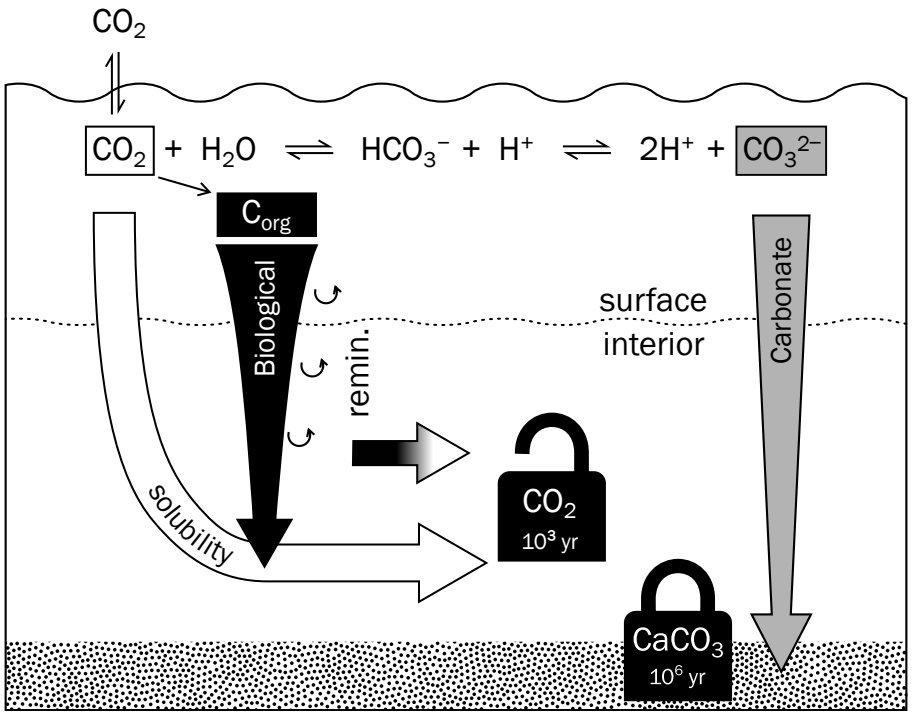


Figure 1: A simplified view of the three carbon pumps, and their effect on ocean carbon storage. Each pump moves carbon from the surface to the deep ocean interior, and sequesters it away from the atmosphere on different timescales: the carbonate or inorganic carbon pump for millions of years in carbonate sediments, the physical or solubility pump for thousands of years in ventilated bottom water, and the biological or organic carbon pump for thousands of years, dependent on the depth of remineralisation.

end-member of the equilibrium series, and the important material that foraminifera, corals and other calcifiers require for shell building.

$$CO_2(g) \rightleftharpoons CO_2(aq) + H_2O \rightleftharpoons HCO_3^- + H^+ \rightleftharpoons 2\,H^+ + CO_3^{2-} \qquad (1)$$

Once these planktic calcifiers die, they sink through the water column and if they reach the seafloor before dissolving, they form sediments, the carbon then being sequestered for millions of years. This, along with coral reef deposition and the abiotic deposition of carbonate rocks, is the inorganic carbon pump.

The physical (solubility) pump results from the increased solubility of CO_2 in cold downwelling water, forming CO_2-enriched deep water masses which sequester the carbon for thousands of years, until the upwelling and degassing of old water.

The aqueous CO_2 is used by phytoplankton in photosynthesis to form organic matter. Like the calcifiers, these organisms eventually die and as this organic carbon (C_{org}) sinks, it is oxidised and transferred back into the dissolved inorganic carbon pool

(DIC; the carbon-containing species in Equation 1). This process of converting particulate organic carbon (POC) or large polymerised molecules of dissolved organic carbon (DOC) into DIC is *remineralisation*, and is a result of the respiration carried out largely within the microbial loop. This remineralisation only forms part of the biological pump if it occurs in the ocean interior: the 98.9 % of carbon that is remineralised in the upper water column (Lutz et al., 2002) is not sequestered, and is largely able to re-equilibrate with the atmosphere. If the C_{org} reaches the deep ocean (>1.5 km depth), this carbon is sequestered for thousands of years, effectively removing it from the atmosphere on millennial timescales, and therefore reducing the greenhouse effect. Whether C_{org} is exported (remineralised at depth) or not is dependent on environmental conditions which may include temperature (Henson et al., 2015). This remineralisation depth is therefore an important source of uncertainty in climate models and predictions, particularly since a relatively small change can have an outsized effect on $pCO_{2_{atm}}$ (Kwon et al., 2009).

The remineralisation of C_{org} is observed in the modern ocean to obey a power law relationship (Equation 2) with depth: the 'Martin curve,' named after John Martin, the lead author of the paper to first describe it (Martin et al., 1987):

$$f_z = f_{z_0} \left(\frac{z}{z_0} \right)^b \tag{2}$$

where f is the carbon flux at depth z, normalised to the rate of production using a reference depth z_0, and b is the attenuation coefficient, widely used as a metric to describe the depth of remineralisation. However, power laws like this tend to overestimate the export to depth (Lutz et al., 2002), and more recent work (Marsay et al., 2015) applies an exponential relationship instead:

$$f_z = f_{z_0} \exp\left(-\frac{z - z_0}{z^*} \right) \tag{3}$$

where z^* is the *length scale of remineralisation*, the depth over which the flux of POC reduces by a factor of e, and roughly equivalent to the original b in Equation 2.

Normalising to the rate of production accounts for the variation in *absolute* strength of the biological carbon pump, but the *efficiency* (the proportion of C_{org} removed from the surface to the deep) is controlled by the balance of two factors: the settling velocity of POC or DOC and the rate at which microbial respiration occurs. The former is affected primarily by the density of the settling particle, and the viscosity of the water through which it sinks (Komar et al., 1981), which varies strongly with temperature

(Rumble et al., 2018): as the ocean warms, the viscosity of water decreases, allowing a faster settling velocity, and less time for remineralisation to occur given the same rate of respiration. The overall effect, all else being equal, would therefore be an increase in biological pump efficiency.

The second factor, the rate of microbial respiration, also varies with temperature (Brown et al., 2004), and although all biochemical reactions are temperature-dependent to some extent (Cossins and Bowler, 1987), López-Urrutia et al. (2006) found that while the rates of respiration and photosynthesis both increase with temperature, respiration increases by a relatively larger amount. They conclude the balance between the two in any particular oceanic ecosystem is profoundly sensitive to environmental temperature, and that an increase in temperature would reduce the relative size of the biological carbon sink, a positive feedback for climate change.

Both factors, then, are temperature-dependent: the settling velocity and rate of respiration should both increase, but their effects on the efficiency of the biological pump are opposed. In order to assess where the balance lies, we may look back at what effect temperature changes in the geological past have had.

In the modern ocean, the depth of remineralisation is generally explored using the flux of POC to various depths as a proportion of primary production, using sediment traps to measure accumulation rates. This direct approach is largely unavailable in the geological record, and so we turn to a well used proxy: stable carbon isotopes.

1.2 The stable carbon isotope proxy

Carbon has three naturally occurring isotopes, two of which are stable: carbon-12 and carbon-13. Although the latter accounts for only 1.1 % of all naturally occurring carbon (Rumble et al., 2018), its varying proportion with respect to ^{12}C provides a useful tracer of carbon in the environment. This proportion is usually expressed in delta notation:

$$\delta^{13}C = \left[\frac{\left({}^{13}C \Big/ {}^{12}C \right)_{sample}}{\left({}^{13}C \Big/ {}^{12}C \right)_{standard}} - 1 \right] \times 1000 \tag{4}$$

The small difference in atomic mass between the isotopes causes a mass fractionation to occur in a number of situations, including the photosynthesis reaction: the carbon-fixing enzyme rubisco has a higher affinity for ^{12}C than ^{13}C, meaning C_{org} is ^{12}C-enriched ("life prefers the light"), and as such has a lighter (more negative) $\delta^{13}C$

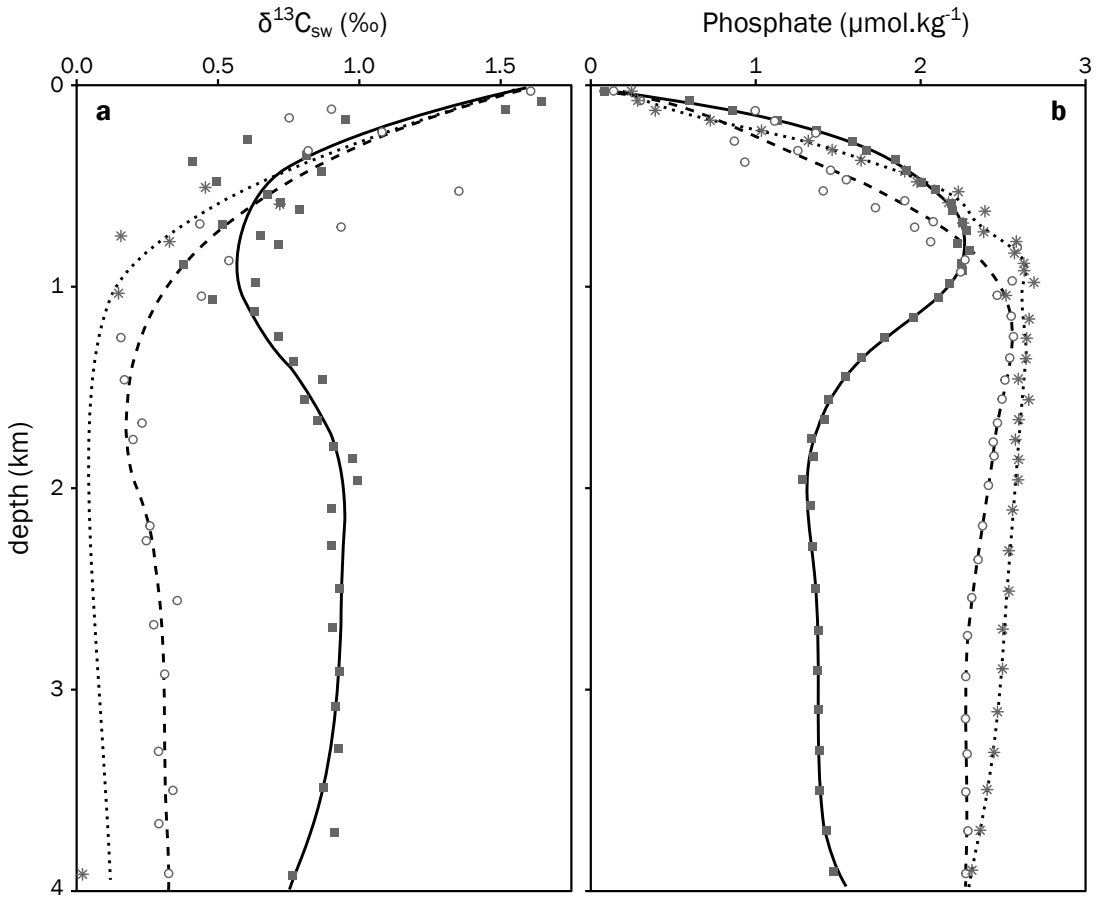


Figure 2: Observed depth-averaged profiles from GEOSECS data of **a**, $\delta^{13}C$ of seawater, and **b**, the concentration of phosphate, in the Atlantic Ocean (solid line and squares), Pacific Ocean (dotted line, asterisks), and Indian Ocean (dashed line, open circles). Phosphate is remineralised along with carbon, and shows a typical nutrient-like non-conservative profile, with which the $\delta^{13}C_{sw}$ profile is anti-correlated for the reasons outlined in the text. Data averaged in depth intervals of 50 m until 1 km depth, 100 m until 2 km, and 200 m thereafter using the box-averaging function in Ocean Data View (Reiner, 2020).

signature. The $\delta^{13}C$ of dissolved inorganic carbon ($\delta^{13}C_{DIC}$) in productive euphotic surface waters is correspondingly heavy.

As POC sinks, it is oxidised and the C_{org} is remineralised – that is, it is converted back to DIC. Since the original C_{org} was ^{12}C-rich, the resultant new DIC is too, reducing the heavy signature left behind after photosynthesis. The exponential increase in the proportion of remineralised versus unaltered C_{org} with depth (Equation 3) causes the rate of change of $\delta^{13}C_{sw}$ with depth to slow, resulting in the characteristic profile of $\delta^{13}C$ in the upper water column (Figure 2a).

This profile is not recorded directly in the geological record, but since different species of foraminifera live at different depths in the ocean, and build their calcite tests using

carbon predominantly from the DIC pool, the $\delta^{13}C$ of foraminiferal calcite can be used as a proxy, assuming the $\delta^{13}C$ of foraminiferal calcite accurately records the $\delta^{13}C_{DIC}$ (see Section 1.2.1). Specifically, planktic foraminifera reflect the $\delta^{13}C$ of the upper water column, while benthic species will reflect deep water signatures.

The less time between two different climate states, the less likely that other variables will change. As such, glacial-interglacial cycles provide an excellent interval for study: the last glacial maximum (19–23 ka) was around 6 °C cooler than the pre-industrial (PI) (Tierney et al., 2020), but largely similar in palaeogeography and faunal assemblages. A warming of 6 °C over <20 kyr should provide a large enough temperature change while oceanic conditions and biological processes remain largely similar.

The effect of an increase in temperature is hypothesised to act to decrease the depth of remineralisation: in the LGM (Figure 3a), cold water temperatures mean respiration rates are lower (López-Urrutia et al., 2006; Mazuecos et al., 2015) and (given the same settling velocity of POC) remineralisation occurs more slowly, allowing more C_{org} to be exported from the surface layers. Conversely, the warmer Holocene (Figure 3b) should have higher rates of respiration and remineralisation, leading to the decay of C_{org} in shallower waters, hence the rapid change with depth of the phosphate curve. Since the $\delta^{13}C$ curve is the mirror image of the phosphate curve (Figure 2), and this can be reconstructed from the $\delta^{13}C_{calcite}$ of foraminifera with different calcification depths, the effect of temperature on remineralisation should be detectable by analysing the reconstructed $\delta^{13}C$ gradients in the Holocene and LGM.

1.2.1 Complications and uncertainties

The ubiquity of carbon in oceanic systems, combined with its low mass and large mass difference between isotopes, means the carbon isotope proxy is sensitive to a number of environmental drivers which can affect both planktic and benthic foraminiferal $\delta^{13}C$. Both atmospheric CO_2 disequilibria and wind-blown dust can both affect the $\delta^{13}C_{DIC}$ of surface water even before it is incorporated into planktic foraminiferal calcite. Likewise, $\delta^{13}C_{sw}$ is a tracer of ocean circulation and as such the signal that benthic foraminifera record will also reflect not just the $\delta^{13}C$ as altered by the water column immediately above them, but also any changes in deep water mass circulation (e.g. the deep water profile of the Atlantic in Figure 2a).

Related to both of these points is the effect of the anthropogenic release of thousands of gigatonnes of organic carbon-derived, isotopically light CO_2 in the atmospheric carbon reservoir, which equilibrates with the modern ocean and therefore superimposes

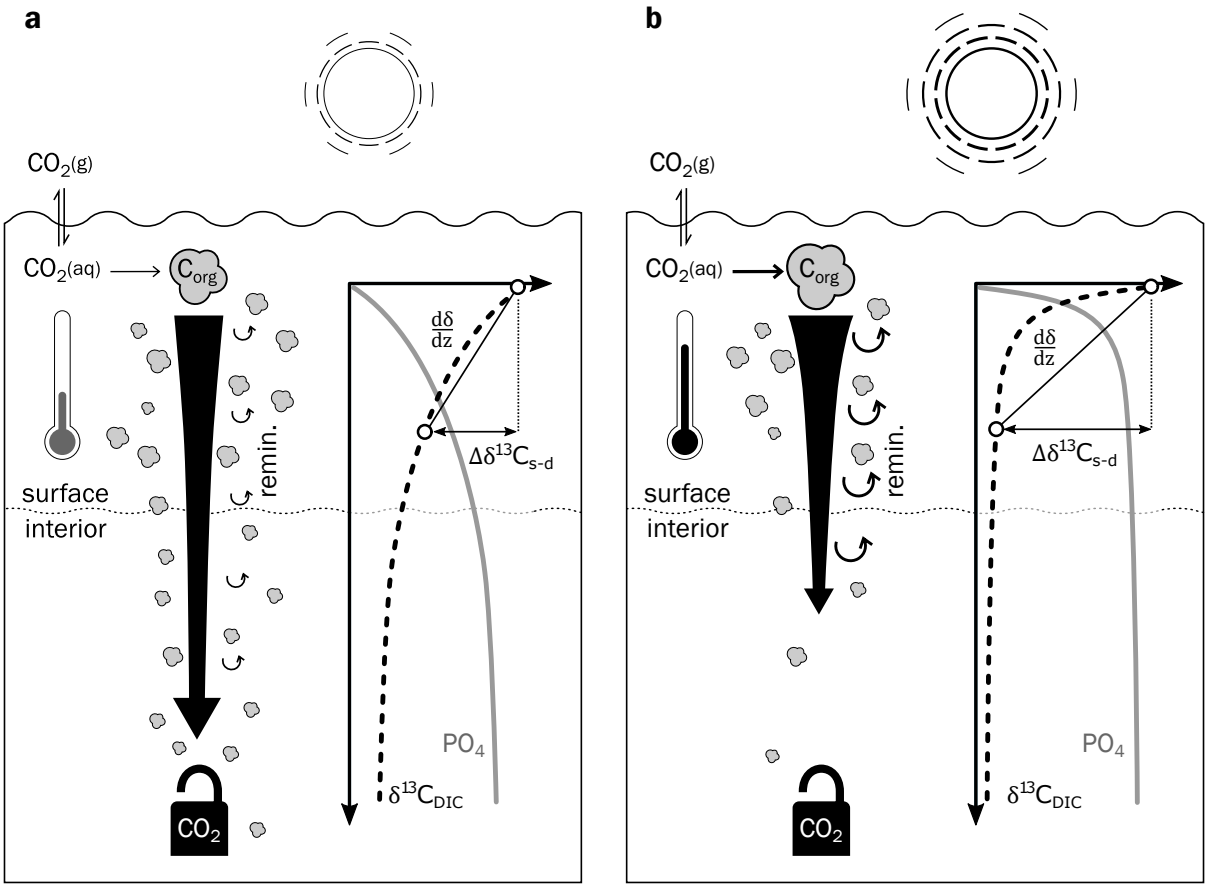


Figure 3: The hypothetical difference in $\delta^{13}C_{sw}$ and nutrient (phosphate) gradients in **a**, cooler climate states (e.g. LGM) and **b**,warmer climate states (e.g. Holocene, relative to LGM). The warmer the ocean, the faster the rates of microbial respiration and therefore remineralisation. A faster rate of remineralisation (given the same sinking speed of POC) causes a decrease in the length scale of remineralisation, a faster rate of change of $\delta^{13}C_{DIC}$ with depth z ($\frac{d\delta}{dz}$), and a larger difference between the $\delta^{13}C_{DIC}$ of surface and deep water ($\Delta\delta^{13}C_{s\text{-}d}$).

an isotopically light signature on $\delta^{13}C$ data starting around the industrial revolution (Olsen and Ninnemann, 2010).

In addition to extraneous changes in the background $\delta^{13}C$, foraminiferal calcite is not a perfect recording medium. Since the tests are mineralised by an organism which maintains its own very local chemical environment, respires, and sometimes maintains a community of symbionts, the $\delta^{13}C$ of the immediate environment does not necessarily reflect the ocean environment (Spero et al., 1991). Different modes of calcification, sources of food, and growth rates can also affect the offset of an individual's $\delta^{13}C_{calcite}$ from the background $\delta^{13}C_{sw}$ (e.g. Schmidt et al., 2008). These are collectively known as *vital effects*, and in some species can affect the $\delta^{13}C$ value by 2 or 3 ‰ (Ravelo and Fairbanks, 1995; McCorkle et al., 1997).

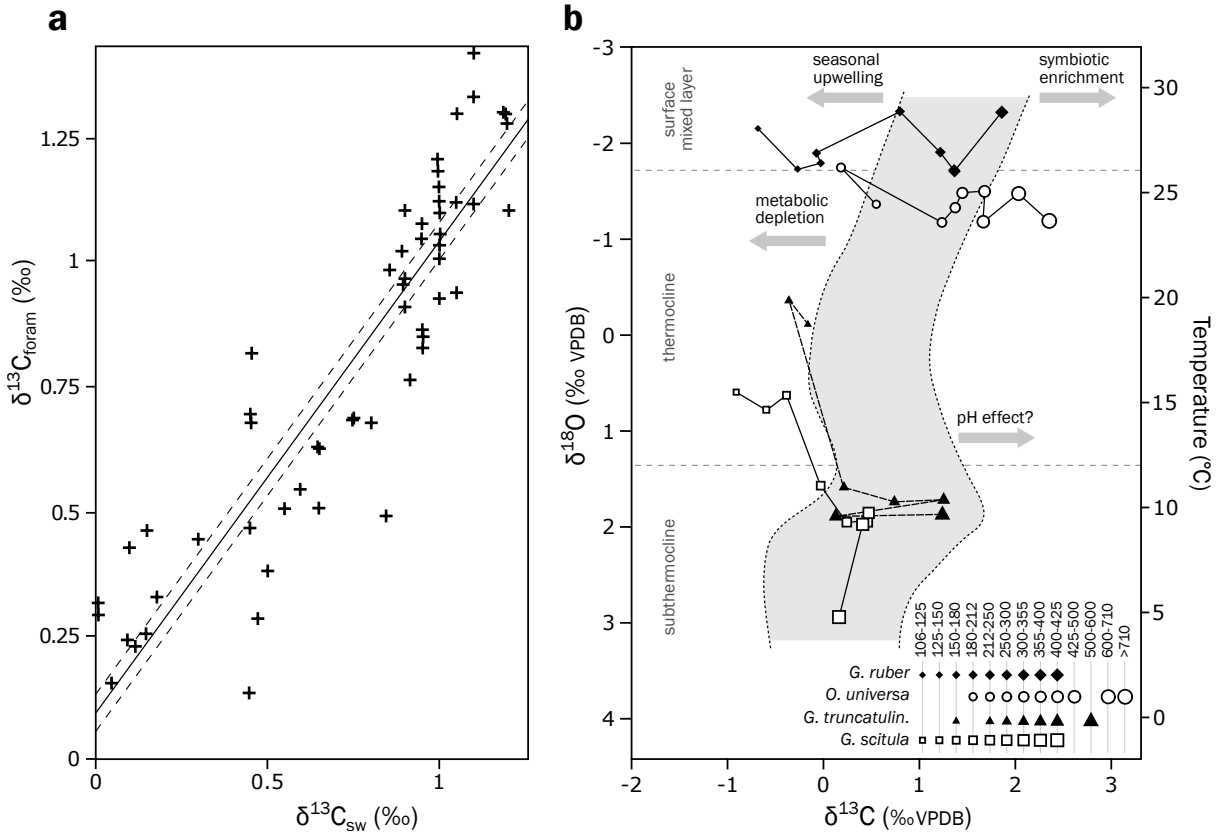


Figure 4: Vital effects in foraminiferal isotope signatures. **a**, $\delta^{13}C$ of recent (core-top) *Cibicidoides* and *Planulina* foraminifera showing a very small offset of $+0.07 \pm 0.04$ ‰ from the $\delta^{13}C_{DIC}$ of the water column above the core location. After Duplessy et al. (1984). **b**, The effect of test size on oxygen and carbon isotope values in four planktic foraminifera species: the surface mixed-layer dweller *Globigerinoides ruber* (diamonds), the deeper mixed layer/upper thermocline dweller *Orbulina universa* (circles), the thermocline dweller *Globorotalia truncatulinoides* (triangles), and the sub-thermocline dweller *Globorotalia scitula* (squares). The size of each symbol is scaled to size of the test, and lines join increasing sizes to show the effect on isotope signatures. Grey shading is the expected water column $\delta^{13}C_{DIC}$ value. Adapted from Birch et al. (2013).

These vital effects are smaller in epibenthic foraminifera (those whose habitat is on or just above the sediment-water interface): the genera *Cibicidoides* and *Planulina* in particular record $\delta^{13}C_{sw}$ with very little offset (Figure 4a).

Planktic foraminifera, with their very different modes of life, are much more variable. In algal symbiont-bearing species, the offset of $\delta^{13}C_{calcite}$ varies with the size of the test. Failing to account for size can lead to vital effects of ±1 ‰ being superimposed (Birch et al., 2013) on the true value of $\delta^{13}C_{sw}$.

For any calcitic microfossil, a diagenetic overprint via dissolution and/or recrystallisation can alter the isotopic signature. Better preserved and younger sediments are less prone to this effect, and recent exceptionally preserved deposits (e.g. John et al., 2013)

show that the most reliable records will come from shallow, clay-rich depositional environments.

Successful interpretation of the stable carbon isotope proxy is complicated, but by analysing one size fraction of one species over time, the effect shown in Figure 4b can be mitigated, though this is not always feasible. In whole-ocean studies, local or regional environmental variation could be reduced either by careful selection of representative sites, or by the aggregation of a large number of less representative sites, with the aim of averaging out any local and site specific deviations.

1.3 Evidence for the temperature dependence of remineralisation

While some evidence for the temperature dependence of remineralisation has been observed in the modern ocean via differing export ratios (e.g. Guidi et al., 2015; Henson et al., 2015; Henson et al., 2019) and modelled using Earth system models (ESM; e.g. Wilson et al., 2019) relatively few studies have tested the hypothesis in the geological record.

John et al. (2013) studied Eocene sediments and observed a greater rate of change with depth ($\frac{d\delta}{dz}$) in this warmer climate state, as hypothesised in Figure 3b. They followed this by successfully modelling the observed Eocene $\delta^{13}C_{DIC}$ profiles using the cGENIE ESM (John et al., 2014).

Studies searching for a cause for lower $pCO_{2_{atm}}$ during the last glacial maximum (LGM) which only include some form of general circulation model (GCM) omit the influence of the biological pump, despite it being an integral part of the ocean carbon cycle. By coupling biogeochemical models to the GCM, this omission may be addressed (Yamamoto et al., 2018). However, the complexity of the biological components of these models varies widely, and only a minority include the temperature dependence of biochemical reactions such as respiration and therefore remineralisation (Segschneider and Bendtsen, 2013). This probably makes them too simple (Henson et al., 2015), and if included in a coupled biogeochemical model, the temperature dependence of remineralisation can explain a considerable proportion of the glacial pCO_2 reduction (e.g. Matsumoto et al., 2007; Chikamoto et al., 2012).

As outlined earlier (see Section 1.2), if the LGM and the late Holocene (0–6 ka) are compared, the length scale of remineralisation should shorten as the climate warms by $\sim$6 °C.

Using the carbon isotope proxy, the alternative hypothesis H_a would be:

> The difference between shallow and deep $\delta^{13}C$ values ($\Delta\delta^{13}C_{s\text{-}d}$) is larger in the warm Holocene (Figure 3b) than in the cold LGM (Figure 3a),

and the null hypothesis H_0:

> There is no difference between $\Delta\delta^{13}C_{s\text{-}d}$ in the Holocene and the LGM, or the $\Delta\delta^{13}C_{s\text{-}d}$ of the Holocene is smaller than that of the LGM.

1.4 Data collection in palaeoclimate science

Given the complications with the stable carbon isotope proxy described in Section 1.2.1, large datasets could provide enough data to average out variation and noise from the local environmental drivers. Combined with restricting the analysis to a small number of well-constrained species, and comparing different climate states from the most recent geological past, most of the uncertainties and complications of the proxy could be mitigated. This would, however, be dependent on having a sufficiently large dataset, since using data from too small a collection would introduce all the environmental effects without averaging them out sufficiently.

Generally, new databases are compiled using one of two methods: expansion or reduction (Figure 5; Jonkers et al., 2020). In the more traditional expansion method, expert knowledge of the available data or a literature search is required, and new data is successively added to the new database. The reduction method, on the other hand, relies on large repositories (e.g. NOAA's NCEI, `https://www.ncdc.noaa.gov/data-access/paleoclimatology-data`, or PANGAEA: `https://pangaea.de`) since all data that might prove relevant is downloaded in bulk and then filtered down according to set inclusion criteria. Although the reduction method is potentially a much faster way to achieve the same database size, the expansion method is not limited to data hosted in a public repository, and therefore may include additional data which meet the criteria for inclusion, but are not generally publicly available (for instance, only available via personal communication). Jonkers et al. (2020) describe this as *dark data*.

The data themselves are not the only consideration in compiling large datasets. The metadata (that is, data about the data) and storage formats may form central parts of inclusion criteria. In the expansion method, since the compilation is largely done manually, there is a lot of leeway for different data formats which can be manipulated

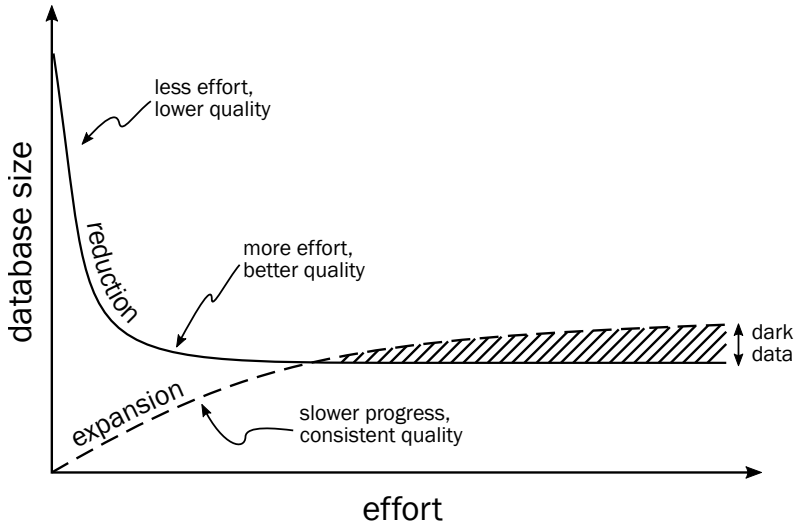


Figure 5: Two methods of database compilation: expansion (dashed line), in which data are gradually added as they are sourced; and reduction (solid line), in which all relevant data from a large repository are reduced quickly as data that do not meet inclusion criteria are filtered out. The expansion method, if enough effort is invested, allows the inclusion of *dark data* – suitable but not publicly available data. After Jonkers et al. (2020).

individually to extract the necessary information, and the metadata does not have to reside with the data, since the compiler can usually fairly easily refer to the original publication. The reduction method, on the other hand, is necessarily scripted, in order to deal with the large amounts of data. As such, the records must be in an identical format, or one of a small set of formats that the compiler can account for in their code. Similarly, the metadata should be machine readable and co-located with the data, or else lose the advantage of automation.

Palaeoclimate data is stored in a huge variety of formats, both in terms of the digital file types and the format of the data within them. The file types can be split into two categories: simple text-based files, and binary formats. The former encompasses a variety of ways of formatting basic tabulated data: comma-separated (CSV) files, tab- or space-delimited files, and other non-standard methods. Binary formats are usually proprietary and linked to a programming language or framework (their native format), and include R data store (RDS) files, MATLAB's MAT-files, and Python's "pickle" serialisation format. One non-proprietary format is netCDF, used widely in meteorology, climate modelling, and geophysics.

While text files are universally accessible (even compressed, as for example in ZIP directories), suitable for use in version control systems, and very flexible, binary formats are generally checked on writing and therefore more reliable storage mechanisms, and

are usually optimised for the quick reading and writing of large amounts of data. They are however only accessible to users who have the knowledge and tools to deserialize them, and this may be restricted by the terms of a commercial license (e.g. MATLAB).

LiPD (Linked PaleoData, `http://lipd.net`; McKay and Emile-Geay, 2016) is a new data format specifically designed for palaeoclimate data, flexible enough in theory to support any kind of proxy data along with their age models and metadata. As an open format, it is not proprietary, though specialised tools are needed to deserialize it.

1.4.1 Existing compilations

By combining two datasets, compiled using opposing methods of expansion and reduction, the resultant database could be significantly larger than any previous palaeoclimate database. Two such compilations are the PalMod database (Jonkers et al., 2020) and the data from a synthesis performed by Oliver et al. (2010).

The PalMod 130k marine palaeoclimate database (Jonkers et al., 2020) was compiled using the reduction method, and the result is a large (896 time series from 143 sites) and detailed compilation, including eight palaeoclimate parameters and full age model details for the last 130 kyr. Their inclusion criteria included strictly those records which had robust chronology, and so records which might have been usable in this study were excluded.

The Oliver et al. (2010) synthesis had less strict inclusion criteria and was compiled largely with the expansion technique. As such, it includes many records which PalMod excludes, either because of poorly defined chronology or because they fall under the dark data category. This database covers the last 150 ka and includes data from over 300 cores.

Merging these two databases will inevitably result in some overlap and duplication of records given the finite size of all palaeoclimate data; whether the size of this overlap could be reduced will be explored by using two source databases with different compilation methods and different inclusion criteria.

2 Methods

To maximise the number of data points, two existing quaternary data sets (Oliver et al., 2010, hereafter 'the Oliver et al. data'; and PalMod, Jonkers et al., 2020; see also Section 1.4.1) were merged. Duplicate records were removed, records with no age or depth information discarded, and parameter and taxon names standardised. The new compilation, the Quaternary Palaeoclimate Isotope Database (QPID; available at `https://github.com/t-arney/QPID`), is focused on machine readability and accessibility.

2.1 Data compilation procedure

The Oliver et al. data was downloaded with the supplementary data as a CSV file; the PalMod data was downloaded in R Data Store (RDS) format. Both data sets were imported into R for processing, where it was possible to correct or standardise each record while merging.

2.1.1 Database design

Ideally, large amounts of data would be input and maintained in a full database engine which would allow the checking of data validity on entry (for example, checking that IDs are unique, required fields are non-null, etc.) and provide optimised access routines. Although Structured Query Language (SQL) interfaces exist for data science programming languages such as R, proficiency is not widespread and easy accessibility is key in allowing users to expend the least effort and retain as much time as possible for answering the scientific question.

As such, and to address the data formatting issues discussed in Section 1.4, a compromise relational database-style system of CSV files was used since this provides the best compromise between accessibility, file size, and ease of use. Relational databases separate data into semantically-themed tables – in this case data and metadata (Figure 6). These tables are linked by unique IDs which allow cross-referencing between two tables. If information from more than one table is required, a *join* is performed and the resultant table can then be used in analysis. The functions `outerjoin()` in MATLAB, `merge()` in R, and `VLOOKUP()` in Microsoft Excel perform this task. Structur-

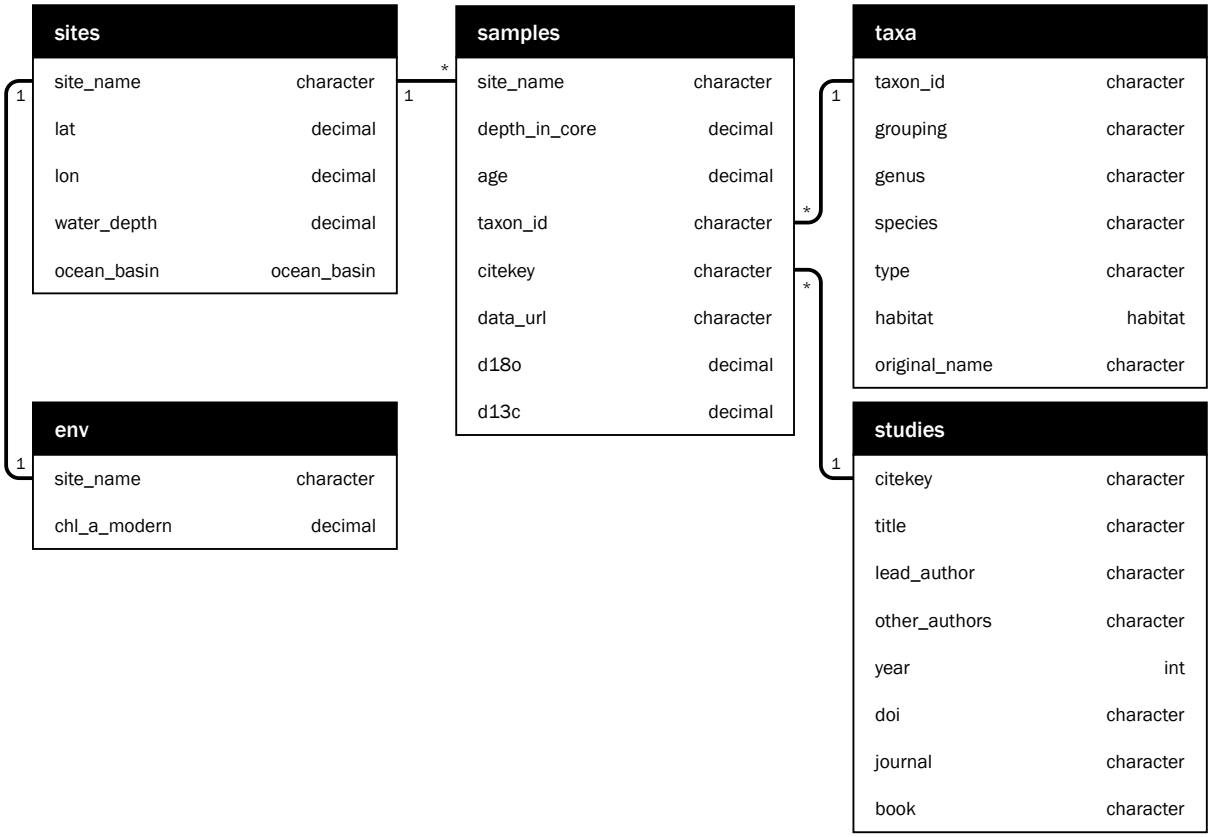


Figure 6: QPID's relational database-style design. Each box is a separate table and CSV file, storing the data (samples table) and metadata (other tables). Lines connecting variables are IDs used for cross-referencing (primary and foreign keys in a relational database). Next to each variable is its data type, where *habitat* is one of `Pl` (planktic) or `Bn` (benthic) and *ocean_basin* is discussed in the main text.

ing information in this way avoids the common pitfall of CSV files in which metadata is repeated for each data point, resulting in a much larger file.

2.1.2 Standardisation

A common issue in palaeoclimate science is the lack of standardisation among parameter names, or even references to core names or sample labels. As such, locating duplicate records is not straightforward. The first level of duplication which must be accounted for, and a useful example for the more general issue of standardisation, is the site name and position. Table 1 contains pairs of sites (a-f) which are each an example of a different form of duplication.

Rows (a) and (b) have duplicated names but different positional metadata. Row (b) should be site V22-193, as evidenced by the original core descriptions (Szatmari, 1968). This also provides an example of the non-standardised naming of geological samples:

Table 1: forms of record duplication in source databases

	site name	latitude	longitude	water depth	source
(a)	V22-108	−43.18	−3.25	4171	Oliver et al.
(b)	V22-108	9.917	−20.982	4956	Oliver et al.
(c)	ODP1145	19.583 33	117.6333	−3175	PalMod
(d)	1145	19.583	117.633	3175	Oliver et al.
(e)	EW9209_1JPC	5.9067	−44.195	−4056	PalMod
(f)	EW9209-1JPC	5	−43	4056	Oliver et al.

The original report names the core "VEMA 22-193", but apparent informal conventions usually abbreviate Vema, the name of the research vessel, to either "V" or "VM". This is also shown in rows (c) and (d), where the Ocean Drilling Project (ODP) core ODP 1145 is listed with only the core number in the Oliver et al. data. QPID has been populated with the more specific name in such cases.

Rows (e) and (f) show two issues: first, that the two names differ only by the punctuation used to separate two halves – symptomatic of converting an original name with spaces into an ID – which to a human are obvious duplicates but to a machine are unique. The second issue is the different levels of reporting accuracy concerning positional data. Row (f) is truncated to the integer values of latitude and longitude, as presented in the text of the original publication (Curry and Oppo, 1997), while the more precise information in row (e) comes from the metadata contained in the PANGAEA record. In these cases, the more precise information was applied to the data.

It is interesting to note that the water depth values between both sources were much more closely aligned than the site names, and could almost be used to distinguish duplicated cores (see the water depth column of Table 1). Other than no convention for the sign of the value, it is almost always reported to the nearest metre, and therefore has an unambiguous format of three or four numeric digits (unlike the core name, which lacks any consistent structure).

During the data wrangling stage, sites were also given a code indicating the ocean basin, region, or sea in which they sit, formed using the first two letters of the major ocean name, preceded by a modifier comprising either a compass point abbreviation or the letter T for tropical (equatorial) areas, with the exception of the Mediterranean Sea which was assigned the abbreviation `Med`. The boundaries, based on those defined by the International Association for the Physical Sciences of the Ocean (IAPSO Smithson, 1992) are shown in Figure 7, and allow regions of interest to be isolated: for instance, the Gulf Stream in the `NWAt` (northwest Atlantic) region, the high nutri-

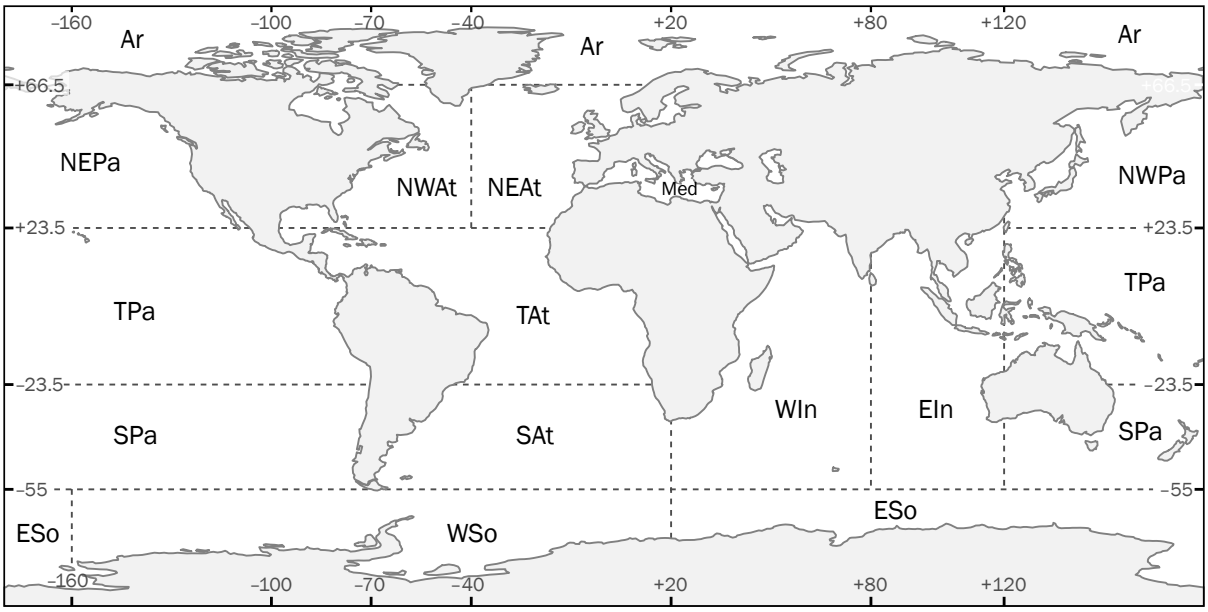


Figure 7: Boundaries of the the regions used in the database field `ocean_basin`, based on the boundaries defined by IAPSO (Smithson, 1992).

ent/low chlorophyll (HNLC) northeast Pacific (`NEPa`), or upwelling in the equatorial Pacific (`TPa`).

Taxon names were also standardised: although the PalMod data had already been standardised, genus names had been abbreviated, so they were expanded and the Oliver et al. data aligned using the World Register of Marine Species (WoRMS) database (`https://www.marinespecies.org`). The abbreviated names were searched using a wildcard character after the first letter of the genus (e.g. `G% ruber` to match the genus *Globigerinoides*), and the search restricted to species rank or lower within the phylum Foraminifera (d'Orbigny, 1826). For ambiguous taxa, the original publications of cores which contained measurements on the species in question were consulted for references to the correct genus, and if still unknown, the data URL was checked. For species that have been reclassified, the accepted name (according to WoRMS) was used (e.g. *Trilobatus sacculifer* instead of the outdated designation *Globigerinoides sacculifer*). Taxa were also assigned a code based on their habitat: planktic (`Pl`) or benthic (`Bn`).

Original publication information metadata was also included where known, since this is important for traceability. Oliver et al. included a large table in their article text giving references for most of the cores used in their synthesis. However, this format is not machine-readable and assimilating the data into QPID meant manually copying the references and linking them to the correct core. 15 % of the cores in the data were not included in the table; these are referenced to Oliver et al. (2010) to provide some traceability. PalMod did include references directly in the metadata, making their in-

clusion much easier. However, Jonkers et al. used the digital object identifier (DOI) as the ID, and for those works without a DOI (often old PhD theses), they included full bibliographic data. This means that not all records are guaranteed to be referenced in the same way, which is less machine-readable. In QPID, each record has a unique 'citekey' assigned, based on the author-year style of citation. This ID can be cross-referenced to the studies table, where full bibliographic data is provided (including a DOI if available). The citekey means any resource can be referenced – online or not – from personal communication to a CD-ROM, and has enough meaning on its own to provide some traceability if the samples table is viewed without the studies table.

2.2 Tests of database quality

Before proceeding to use the new QPID database to investigate the temperature dependence of remineralisation, it was necessary to check that the data was as expected, and that it could reflect known or theoretical oceanographic and geological phenomena. First, the coverage, both geographical and temporal, was found and plotted. Any extreme values were flagged and checked for validity. To test the veracity of the oxygen isotope data, the $\delta^{18}O$ of *Cibicidoides wuellerstorfi* was plotted against the LR04 benthic stack (Lisiecki and Raymo, 2005) and visually compared. The carbon isotope data were assessed by comparing their water column profile to the expected profile for each ocean (Figure 2a), by plotting $\delta^{13}C$ against water depth for each ocean.

2.3 Exploring ocean remineralisation changes using the database

In order to investigate the temperature dependence of remineralisation, global data was used to normalise any environmental effects, with the exception of high latitudes (ocean basin codes `Ar`, `WSo`, and `ESo`) since faunal assemblages and habitat ranges are likely to have changed significantly with the retreat of sea ice after the LGM. The Mediterranean was also excluded on the basis of unrepresentative oceanographic conditions.

Data were extracted from the LGM (19–23 ka; Peterson et al., 2014) and the late Holocene, defined as the last 5 kyr but excluding the most recent 250 years to avoid anthropogenic influences on the $\delta^{13}C$ record. Species which reliably record $\delta^{13}C_{sw}$ were extracted: the epibenthics *Cibicidoides*, *Cibicides*, and *Planulina* (Duplessy et al., 1984, referred to as *Cib.* spp.), and the planktic species *Globigerinoides ruber* (Birch et al., 2013). An averaged correction of +0.94 ‰ (Spero et al., 2003) was applied to *G. ru-*

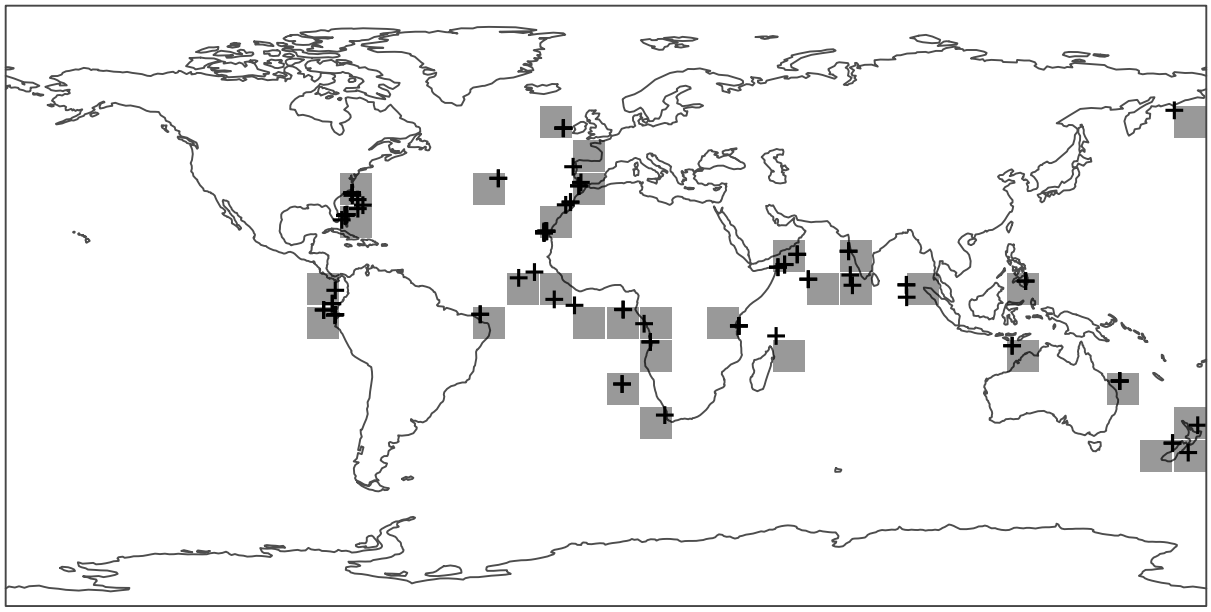


Figure 8: 10° × 10° latitude/longitude grid cells used to aggregate data and remove spurious weighting from multiple closely spaced cores (e.g. in the north Atlantic). Plus symbols indicate original core locations, grey squares are populated grid cells.

ber values, and the pink chromotype was discarded (`G_ruber` and `G_ruber_w` taxon IDs selected). A calcification depth of 25 ± 25 m was assigned to *G. ruber* (mid-point between shallow and deep bounds from the literature, e.g. Hemleben and Spindler, 1983; Kemle-von Mücke and Oberhänsli, 1999; Niebler et al., 1999; Anand et al., 2003; Spero et al., 2003; Farmer et al., 2007; Steph et al., 2009; see also supplementary Table S1), and the calcification depths of the benthic taxa were defined as the water depth at the core site, corrected by -125 ± 5 m for LGM records to account for sea level change (Murray-Wallace and Woodroffe, 2014). Only *Cib.* spp. where the corrected calcification depth was <1000 m were used to restrict the analysis to the upper water column, where the rate of change of the δ^{13}C profile is greatest (Figure 2a). The script used to filter the input data can be viewed in full in the Appendix.

These filtered data (n = 923) were assigned to a 10° × 10° latitude-longitude grid cell to reduce the weighting of areas with multiple closely spaced cores, as occurs at the western and eastern edges of the north Atlantic. 32 of these grid cells contained at least one core with suitable data (Figure 8). Within each grid cell, the mean δ^{13}C values of *G. ruber* and *Cib.* spp. were calculated, along with the mean calcification depth and associated upper and lower bounds.

Paired data is necessary to calculate differences, but the number of shallow and deep values is not equal within each time period. A random sample of the larger set was therefore taken in order to create paired values which could be subtracted. The re-

sultant sample sizes were 6 paired data points in the Holocene and 9 in the LGM. The shallow-deep differences, $\Delta\delta^{13}C_{s\text{-}d}$, were then calculated for each pair, and the mean of these taken for each time period. 95 % confidence intervals were calculated by multiplying the standard error of the mean by 1.96. The sampling and difference calculations were run multiple times so that all potential pairs were included and all the data used. A reproducible average value could then be reported, since the random sampling of a small number of data points from a larger set would otherwise introduce variation in the result. Finally, the unpaired data were plotted and a mean gradient calculated for visual comparison.

3 Results

The Quaternary Palaeoclimate Isotope Database (QPID) is a key output of this work, and its validity and coverage is described first. As a large collection of carbon and isotope data, its potential uses could include the constraining vital effects, assessing whole-ocean $\delta^{13}C$ change (Peterson et al., 2014), exploring glacial-interglacial or even millenial-timescale events based on the benthic $\delta^{18}O$ stack (Figure 10). It is applied in this work to exploring the role of temperature in the variability of the length scale of remineralisation.

3.1 QPID: The Quaternary Palaeoclimate Isotope Database

QPID includes 141 236 measurements on 83 138 foraminiferal calcite samples from 396 globally distributed deep-sea sediment cores (Figure 9a). The Atlantic is over-represented relative to the Pacific Ocean, despite the Pacific occupying around twice the surface area (Cotter et al., 2019), likely due to a research focus in the Atlantic because of its importance in the global ocean circulation, and its proximity to countries which fund and carry out research. Similarly, the Southern Ocean is very unrepresented, partly due to the difficulties in carrying out drilling in the region's hostile conditions, and partly due to the lower occurrence of carbonate deposits at high latitudes.

Table 2 details the distribution of the sites and samples, and the proportion of the samples which have a $\delta^{18}O$ measurement, a $\delta^{13}C$ measurement, or both. Overall, 97 % of samples have a $\delta^{18}O$ measurement, 73 % have $\delta^{13}C$, and 70 % have both, which largely reflects the fact that $\delta^{13}C$ measurements are collected (though not always published) simultaneously with $\delta^{18}O$, since CO_2 is usually analysed (Coplen et al., 1983).

The temporal distribution of the samples (Figure 9b) shows a similar research focus on the late Pleistocene and early Holocene, around the time of the last deglaciation (5–20 ka), though this is probably also a result of more recent (and therefore near-surface) sediments being easier to recover. 87 % of samples have ages $<$150 ka, but some cores with samples in that range also extend beyond 150 ka, and the remaining 13 % of samples are up to 3 Myr old. The proportion of samples with both isotope measurements remains very stable at around 70 %, both geographically and temporally over the range 0–150 kyr (grey bars in Figure 9b; Table 2).

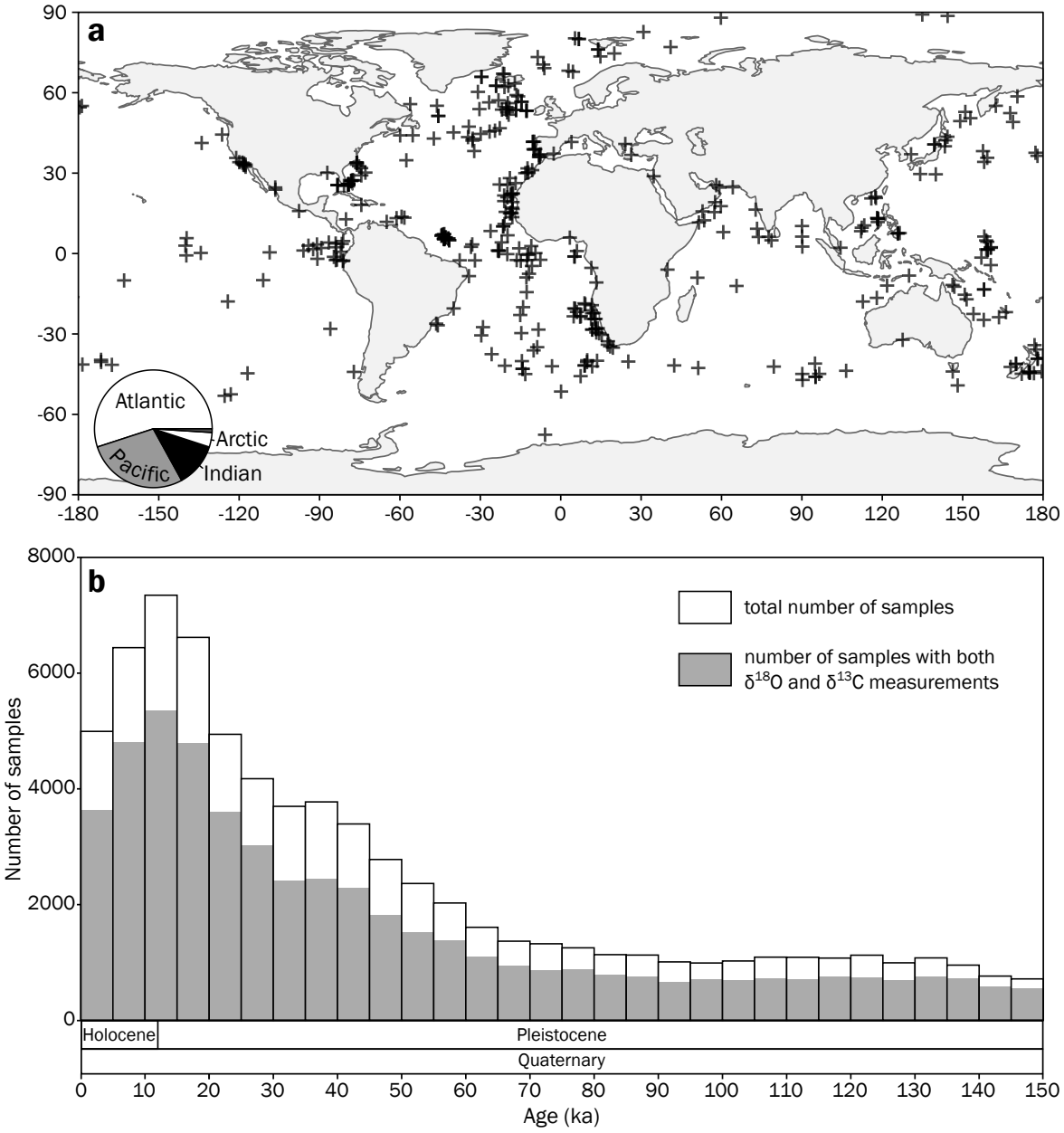


Figure 9: Geographical and temporal coverage of the QPID database. **a**, locations of all sites. Inset: proportion of sites located in each ocean. Note the prevalence of sites in the Atlantic (55 %) and the relative lack of sites in the Pacific (28 %) and Indian (12 %) Oceans. Few sites are present in the Southern Ocean due to the general lack of carbonate sediments below 45° S and the difficult conditions for expeditions. **b**, number of samples per 5 kyr, over the 150 kyr coverage of the database (white bars) and the number of those samples with both $\delta^{18}O$ and $\delta^{13}C$ measurements (grey bars). Data are skewed towards the recent, with a maximum around the time of the last deglaciation.

Table 2: Distribution of samples and measurements by ocean in the QPID database. The absolute number of measurements is given with the proportion of all samples with that measurement for $\delta^{18}O$, $\delta^{13}C$, and both isotope systems.

		$\delta^{18}O$		$\delta^{13}C$		both	
Ocean/sea	samples	data pts	%	data pts	%	data pts	%
Atlantic	43 553	41 879	96	33 588	77	31 914	73
Pacific	27 553	27 033	98	17 734	64	17 214	63
Indian	5543	5522	100	4258	77	4237	76
Arctic	5491	4909	89	4733	86	4151	76
Mediterranean	810	810	100	395	49	395	49
Southern	188	188	100	187	100	187	100
Total	83 138	80 341	97	60 895	73	58 098	70

In order to assess the validity of the oxygen isotope data, the $\delta^{18}O$ of *Cibicidoides wuellerstorfi* was corrected by +0.64 ‰ (Shackleton and Hall, 1984), plotted against age, and a 1 kyr moving average (median) applied (Figure 10a). Compared to the LR04 benthic stack (Lisiecki and Raymo, 2005), the two records are in excellent agreement, despite the QPID data containing a small number of measurements with unusually low $\delta^{18}O$ values between 110 and 140 ka.

The carbon isotope data was assessed by plotting the $\delta^{13}C$ of Holocene (0–10 ka) *C. wuellerstorfi* against water depth at the core site for the major oceans. Though not as pronounced, the data largely fit the expected profiles (from GEOSECS data, see Figure 2a): that is, isotopically heavy in the lower surface waters (∼1 ‰), decreasing with depth to a value of ∼0.5 ‰ for the Atlantic and ∼0 ‰ for the Pacific due to the influence of the thermohaline circulation on the $\delta^{13}C_{sw}$ of the deep Pacific. The Indian Ocean has too little data for a full profile, though the mid-depth values that are recorded also fit with the GEOSECS observations.

3.2 Application: The temperature dependence of remineralisation

Surface and deep $\delta^{13}C$ values for *G. ruber* and *Cibicidoides* spp., *Cibicides* spp., and *Planulina* spp. (*Cib.* spp.), are presented in Figure 11. Average gradients between the surface (0–50 m) and the deep (500–1000 m) are also plotted for visual comparison. As identified in the validation of database contents, the Indian Ocean has little data, and no *Cib.* spp. between 500 and 1000 m in either the Holocene or LGM, which precludes any useful analysis at the ocean level. The Atlantic and Pacific Oceans (Figure 11a) do have enough data to plot coarse gradients (Figure 11b). Quantitative results are

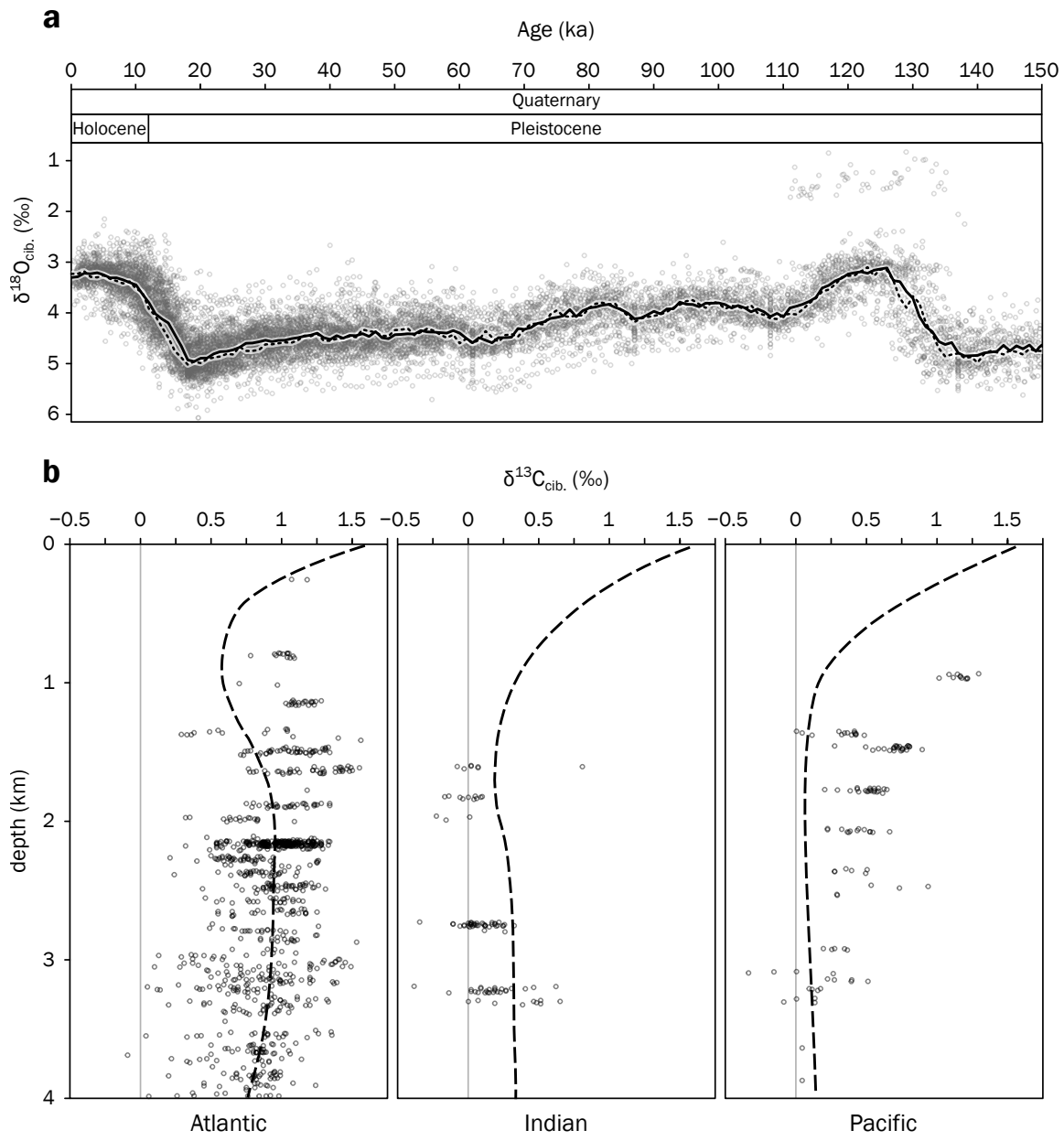


Figure 10: Synthesis of QPID isotope data for comparison to known oceanographic patterns. **a**, $\delta^{18}O$ of *C. wuellerstorfi* for the last 150 kyr (black open circles) with a 1 kyr moving average (solid black line), corrected using the same offset (Shackleton and Hall, 1984) as the standard LR04 benthic stack with which it is compared (Lisiecki and Raymo, 2005; dashed line). **b**, $\delta^{13}C$ of Holocene *C. wuellerstorfi* versus core depth, with the best fit lines to the GEOSECS data from Figure 2a for comparison.

Table 3: Shallow-deep $\delta^{13}C$ differences ($\Delta\delta^{13}C_{s\text{-}d}$) in the Holocene and LGM with 95 % confidence intervals.

region	$\Delta\delta^{13}C_{s\text{-}d}$ (‰)		Holocene - LGM
	Holocene	LGM	difference (‰)
Atlantic	1.2 ± 0.8	0.5 ± 0.6	0.7 ± 1.0
Pacific	1.1 ± 0.5	1.7 ± 0.6	-0.7 ± 0.8
Global	1.2 ± 0.5	1.1 ± 0.5	0.1 ± 0.7

presented in Table 3.

Visually, the profiles in Figure 11 do not appear to resemble the hypothetical profiles in Figure 3. While the hypothetical profiles are exponential, the data from *G. ruber* and *Cib.* spp. are roughly linear. This is expected for those cases with only two clusters of points, but those with a wider spread (e.g. the LGM Atlantic and global aggregate) also show a linear trend, though the depth resolution is low, particularly in the mid-depths between 50 and 500 m where the largest non-linear changes could be expected.

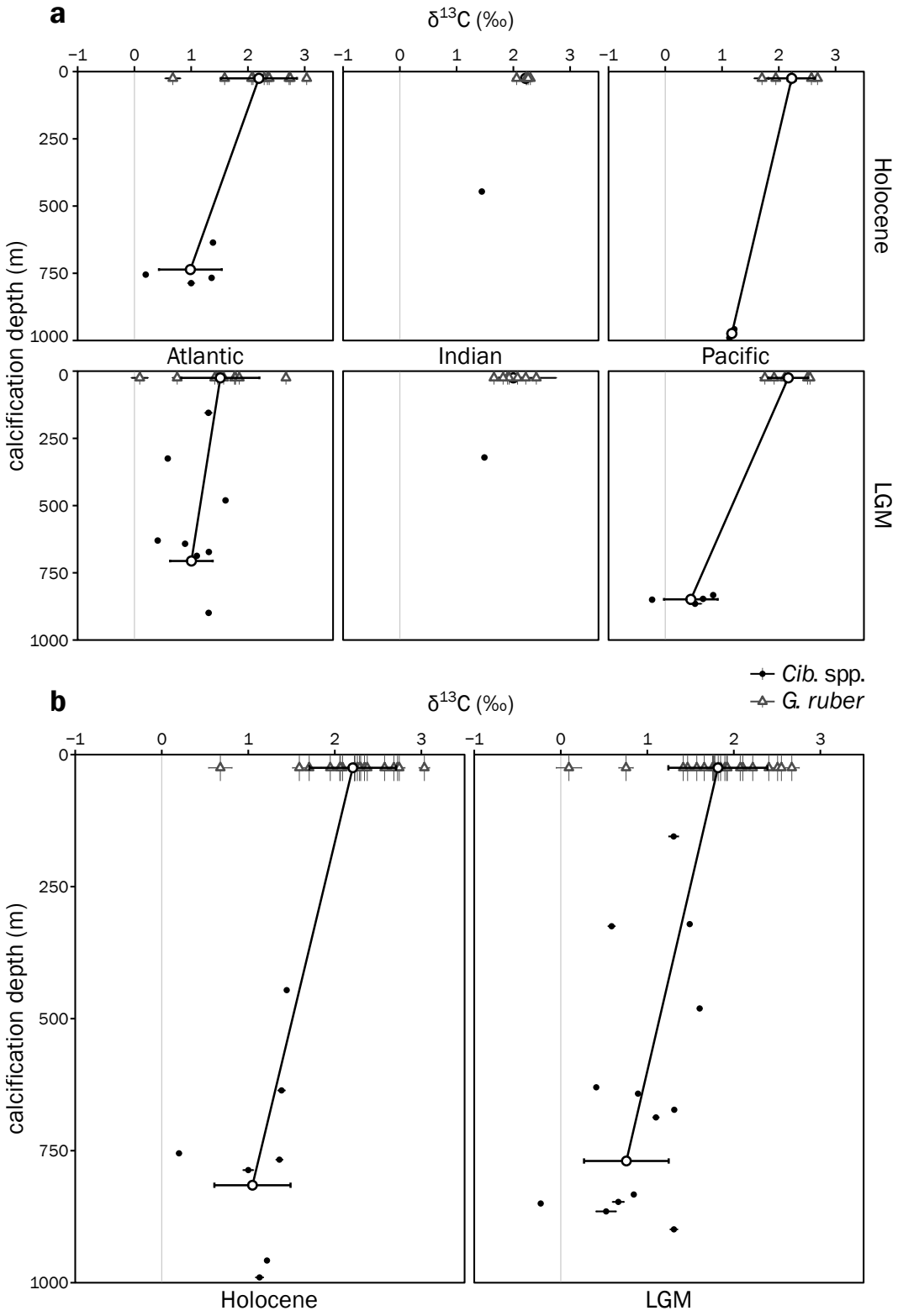


Figure 11: δ^{13}C values in the surface (*G. ruber*, triangles) and deep (*Cib.* spp., solid circles) in the Holocene and LGM **a**, by ocean and **b**, global average. Planktics have been corrected for vital effects and benthics for LGM sea level change. The average surface (0–50 m) to deep (500–1000 m) gradient is also plotted (open circles, black lines).

4 Discussion

4.1 Database coverage and quality

QPID is, within the realm of palaeoclimate data, a large database. Merging two existing databases built using opposing but complementary compilation methods resulted in relatively little overlap and duplication. QPID has global coverage, with multiple cores from every ocean apart from the Southern Ocean, but 55 % of its cores and 52 % of its samples are from the Atlantic, largely reflecting research focus on the Atlantic as an important component of the thermohaline circulation, and as a convenient location for many science-funding countries. For applications involving only one ocean—particularly the Atlantic—this inequality would be immaterial, but it introduces spurious weighting of the data on a global scale. One solution is to grid the data, but at the expense of sample sizes. Temporally, the data is more homogenous, though another research focus is apparent: nearly half the samples date from the most recent quarter of the 150 kyr coverage, and a quarter of the samples date from around the last deglaciation (5–20 ka).

Although QPID has >140 000 data points, the nature of this study required extracting multiple subsets of the data, all acting to compound the reduction in sample size of the previous step (Table 4). After six rounds of reducing the sample size by an average (geometric mean) of 72 % each time, the remaining sample size is only 62 data points, which must then split further into each test case (combinations of shallow and deep in the LGM and Holocene) to arrive at final sample sizes in each test case of around 10, four orders of magnitude smaller than the full database. This underlines how im-

Table 4: The compound effect of extracting subsets of the original dataset

sample size	reduction factor	subsetting step	
141 236			(all measurements)
60 895	0.43	1:	discard $\delta^{18}O$ measurements
7615	0.13	2:	select data from LGM & Holocene
6336	0.83	3:	discard polar regions & Mediterranean
2566	0.40	4:	select *G. ruber* and *Cib.* spp.
923	0.36	5:	discard deep benthics
62	0.07	6:	grid & aggregate

portant it is for a researcher to be able to include as much data as possible in a new compilation, in order to maximise the sample size remaining after compound subsetting. Some level of subsetting, likely multiple steps each compounding the reduction of sample sizes, is inevitable in a system as complex as Earth's climate. If the field of palaeoclimate is to benefit from the syntheses and meta analyses common in other disciplines, it is therefore critical that all future data (and as much legacy data as possible) is archived in an open, accessible, and reliable format, with as much metadata as possible.

In compiling any amount of palaeoclimate data however, it will become clear that the data and file formats are very varied and the metadata patchy, leading to avoidable work which must be carried out before a suitable dataset is ready for analysis. By using a standard data format and including all essential metadata, a great deal of time and effort could be saved and instead redirected towards research goals. Efforts to introduce such a standard are underway: the LiPD (Linked PaleoData) format aims to provide a reliable, easy-to-use, platform-independent file format capable of storing proxy data, age models, and metadata (McKay and Emile-Geay, 2016). Since this format is not standard, reliable and simple tools are necessary to enable users to easily access data from a LiPD file. For example, the netCDF format is supported by a suite of low-level tools which can be used from the command line or integrated into other frameworks – MATLAB has integrated support for netCDF files, and established packages are available on the Comprehensive R Archive Network (CRAN) repository for R. LiPD does currently have immature equivalents to the netCDF tools, but these versions still require work to be usable, which requires investment. Similarly, and in parallel to LiPD, a reporting standard for palaeoclimate data is being developed: PaCTS (Khider et al., 2019) aims to maximise the reuse value of palaeoclimate data by providing guidelines for which metadata should be reported with the data. This should allow any new data to be easily reused in a new database or synthesis with much less effort expended in searching for necessary metadata.

4.2 Temperature and remineralisation

While the Atlantic appears to have the hypothesised slower rate of change of $\delta^{13}C$ with depth, the Pacific appears to have the opposite. In the global aggregation (Figure 11b), the effects seem to cancel – the two gradients look very similar. Neither the global ocean nor the Atlantic or Pacific alone show significant differences in $\Delta\delta^{13}C_{s\text{-}d}$ between the LGM and Holocene, and therefore the null hypothesis that there is no difference between the LGM and Holocene $\delta^{13}C$ gradients is upheld, and the alternative rejected.

However, given the uncertainties involved in carbon isotopes, vital effects, depth constraints, and sparseness and unequal weighting of data, the effect may not in fact be absent but could simply be unresolvable by this study. In particular, the data is very sparse in the mid-depths (∼50–500 m) of all but the glacial Atlantic, and in the deep Indian Ocean.

One factor not accounted for in this analysis is the possibility of changes in primary productivity compensating for any change in remineralisation depth. The export efficiency (the proportion of organic carbon from primary productivity exported to the deep ocean, linked to the depth of remineralisation via Equations 2 and 3) has been found in some cases to be inversely related to levels of primary productivity (Henson et al., 2019). One method for controlling for this would be to restrict the analysis to oligotrophic regions, since the ocean gyres are relatively stable and they can be assumed to have remained oligotrophic through the last glacial cycle. However, just as there is a research focus in the Atlantic and in the last 30 ka, there is also a dearth of cores from the gyres. Subsetting the data to the least productive one third of sites (using modern chlorophyll values as a proxy for oligotrophic conditions) results in another subsetting step in Table 4, with a very large reduction factor of ∼90 %, and a final sample size before splitting into test cases of only 15 data points. As such, this factor could not be controlled for in this study, but it points to the need for more data in oligotrophic regions to address issues such as this.

4.3 Further work

This analysis could be improved by adding more sediment cores to the dataset, to provide more global coverage. By focussing on shallow sites (<1000 m water depth) in the Pacific and Indian Oceans, the most significant shortcomings of the current database would be addressed. Alternatively, more utility could be extracted from the current data: by carefully choosing suitable species of planktic foraminifera other than *G. ruber* (e.g. *Globorotalia truncatulinoides*), many more sites could be used for the 'deep' value, instead of just the shallow benthics from water depths of less than 1000 m. This route would require much better constrained calcification depths, and could be calculated individually using $\delta^{18}O_{sw}$, combined with rigorous carbon isotope correction values based on size fraction and location. Finally, some of the issues with the carbon isotope proxy could be avoided by using a different proxy such as the boron isotope proxy, which is a reliable indicator of the pH of seawater. This system is less affected by environmental conditions, but still requires careful attention to vital effects and species offsets.

5 Conclusions

Merging two existing databases compiled using opposing but complementary methods results in a significantly bigger database—QPID—due to little overlap between the two sources.

QPID is large by palaeoclimate standards, with global coverage over the last 150 kyr. The effect of research focus means that QPID is weighted towards the late Pleistocene–early Holocene Atlantic, with the best timeseries being that of *C. wuellerstorfi*. Topics requiring data outside these values are at risk of small smaple sizes due to the influence of compound subsetting, and it is essential that barriers to science in the form of excessive data wrangling be reduced or removed.

Currently, palaeoceanography is hindered by the lack of an open, accessible, and reliable data format combined with a lack of standardised metadata. Both areas are under active development, but must continue to be funded and recruit support from current researchers. Data generated in future studies should be shared openly with at least the minimal metadata described in the PaCTS reporting standard (`http://wiki.linked.earth/PaCTS_v1.0`).

The disadvantage of the over-weighting of the recent Atlantic in QPID—and palaeoclimate and palaeoceanography in general—is apparent in the sparseness of data used to investigate the temperature dependence of remineralisation, and possibly the reason that the alternative hypothesis was rejected, and no temperature effect on the depth of remineralisation was found.

Acknowledgements

Many thanks to Gavin Foster and Thomas Chalk for being so generous with their time and providing endless help, direction and patient advice; Toby Tyrell for his advice concerning statistical methods; and last but certainly not least, thanks to Danielle Pearson for her proofreading and limitless support and encouragement.

Supplementary materials

The QPID database (version 1.0) is available online at `https://github.com/t-arney/QPID`, along with the code used to compile it.

The code specific to the analyses in this work, along with supplementary Table S1 is available separately at `https://github.com/t-arney/msc-thesis` :

- `samples-filter.R`
- `LGM-Hol-difference-calculations.R`
- Table S1: Planktic foraminifera calcification depths

References

Anand, P., Elderfield, H., and Conte, M. H. (2003), "Calibration of Mg/Ca Thermometry in Planktonic Foraminifera from a Sediment Trap Time Series", *Paleoceanography* 18 (2), DOI: 10.1029/2002PA000846.

Birch, H., Coxall, H. K., Pearson, P. N., Kroon, D., and O'Regan, M. (2013), "Planktonic Foraminifera Stable Isotopes and Water Column Structure: Disentangling Ecological Signals", *Marine Micropaleontology* 101, pp. 127–145, DOI: 10.1016/j.marmicro.2013.02.002.

Broecker, W. S. (1982), "Glacial to Interglacial Changes in Ocean Chemistry", *Progress in Oceanography* 11 (2), pp. 151–197, DOI: 10.1016/0079-6611(82)90007-6.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and West, G. B. (2004), "Toward a Metabolic Theory of Ecology", *Ecology* 85 (7), pp. 1771–1789, DOI: 10.1890/03-9000.

Chikamoto, M. O., Abe-Ouchi, A., Oka, A., and Smith, S. L. (2012), "Temperature-Induced Marine Export Production during Glacial Period", *Geophysical Research Letters* 39 (21), DOI: 10.1029/2012GL053828.

Coplen, T. B., Kendall, C., and Hopple, J. (1983), "Comparison of Stable Isotope Reference Samples", *Nature* 302 (5905), pp. 236–238, DOI: 10.1038/302236a0.

Cossins, A. and Bowler, K. (1987), *Temperature Biology of Animals*, London: Chapman and Hall, 346 pp., Google Books: s6m1BwAAQBAJ.

Cotter, C. H., Bardach, J. E., and Morgan, J. R. (2019), *Pacific Ocean*, *Encyclopaedia Britanica (Online)*, URL: https://www.britannica.com/place/Pacific-Ocean.

Curry, W. B. and Oppo, D. W. (1997), "Synchronous, High-Frequency Oscillations in Tropical Sea Surface Temperatures and North Atlantic Deep Water Production during the Last Glacial Cycle", *Paleoceanography* 12 (1), pp. 1–14, DOI: 10.1029/96PA02413.

Duplessy, J.-C., Shackleton, N. J., Matthews, R. K., Prell, W., Ruddiman, W. F., Caralp, M., and Hendy, C. H. (1984), "^{13}C Record of Benthic Foraminifera in the Last Interglacial Ocean: Implications for the Carbon Cycle and the Global Deep Water Circulation", *Quaternary Research* 21 (2), pp. 225–243, DOI: 10.1016/0033-5894(84)90099-1.

Farmer, E. C., Kaplan, A., Menocal, P. B. de, and Lynch-Stieglitz, J. (2007), "Corroborating Ecological Depth Preferences of Planktonic Foraminifera in the Tropical Atlantic with the Stable Oxygen Isotope Ratios of Core Top Specimens", *Paleoceanography* 22 (3), DOI: 10.1029/2006PA001361.

Guidi, L., Legendre, L., Reygondeau, G., Uitz, J., Stemmann, L., and Henson, S. A. (2015), "A New Look at Ocean Carbon Remineralization for Estimating Deepwater Sequestration", *Global Biogeochemical Cycles* 29 (7), pp. 1044–1059, DOI: 10.1002/2014GB005063.

Hemleben, C. and Spindler, M. (1983), "Recent Advances in Research on Living Planktonic Foraminifera", *Reconstruction of Marine Paleoenvironments*, ed. by J. E. Meulenkamp, Utrecht Micropaleontological Bulletins 30, Utrecht University, URL: https://dspace.library.uu.nl/handle/1874/205889.

Henson, S. A., Le Moigne, F., and Giering, S. (2019), "Drivers of Carbon Export Efficiency in the Global Ocean", *Global Biogeochemical Cycles* 33 (7), pp. 891–903, DOI: 10.1029/2018GB006158.

Henson, S. A., Yool, A., and Sanders, R. (2015), "Variability in Efficiency of Particulate Organic Carbon Export: A Model Study", *Global Biogeochemical Cycles* 29 (1), pp. 33–45, DOI: 10.1002/2014GB004965.

IPCC (2014), *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. by Core Writing Team, R. K. Pachauri, and L. Mayer, Geneva: IPCC, 151 pp.

John, E. H., Pearson, P. N., Coxall, H. K., Birch, H., Wade, B. S., and Foster, G. L. (2013), "Warm Ocean Processes and Carbon Cycling in the Eocene", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2001), p. 20130099, DOI: 10.1098/rsta.2013.0099.

John, E. H., Wilson, J. D., Pearson, P. N., and Ridgwell, A. (2014), "Temperature-Dependent Remineralization and Carbon Cycling in the Warm Eocene Oceans", *Palaeogeography, Palaeoclimatology, Palaeoecology*, Selected Papers, Geologic Problem Solving with Microfossils3 413, pp. 158–166, DOI: 10.1016/j.palaeo.2014.05.019.

Jonkers, L., Cartapanis, O., Langner, M., McKay, N., Mulitza, S., Strack, A., and Kucera, M. (2020), "Integrating Palaeoclimate Time Series with Rich Metadata for Uncertainty Modelling: Strategy and Documentation of the PalMod 130k Marine Palaeoclimate Data Synthesis", *Earth System Science Data* 12 (2), pp. 1053–1081, DOI: 10.5194/essd-12-1053-2020.

Kemle-von Mücke, S. and Oberhänsli, H. (1999), "The Distribution of Living Planktic Foraminifera in Relation to Southeast Atlantic Oceanography", *Use of Proxies in Paleoceanography: Examples from the South Atlantic*, ed. by G. Fischer and G. Wefer, Berlin, Heidelberg: Springer, pp. 91–115, DOI: 10.1007/978-3-642-58646-0_3.

Khider, D., Emile-Geay, J., McKay, N. P., Gil, Y., Garijo, D., Ratnakar, V., Alonso-Garcia, M., Bertrand, S., Bothe, O., Brewer, P., Bunn, A., Chevalier, M., Comas-Bru, L., Csank, A., Dassié, E., DeLong, K., Felis, T., Francus, P., Frappier, A., Gray, W., Goring, S., Jonkers, L., Kahle, M., Kaufman, D., Kehrwald, N. M., Martrat, B., McGregor, H., Richey, J., Schmittner, A., Scroxton, N., Sutherland, E., Thirumalai, K., Allen, K., Arnaud, F., Axford, Y., Barrows, T., Bazin, L., Pilaar Birch, S. E., Bradley, E., Bregy, J., Capron, E., Cartapanis, O., Chiang, H.-W., Cobb, K. M., Debret, M., Dommain, R., Du, J., Dyez, K., Emerick, S., Erb, M. P., Falster, G., Finsinger, W., Fortier, D., Gauthier, N., George, S., Grimm, E., Hertzberg, J., Hibbert, F., Hillman, A., Hobbs, W., Huber, M., Hughes, A. L. C., Jaccard, S., Ruan, J., Kienast, M., Konecky, B., Le Roux, G., Lyubchich, V., Novello, V. F., Olaka, L., Partin, J. W., Pearce, C., Phipps, S. J., Pignol, C., Piotrowska, N., Poli, M.-S., Prokopenko, A., Schwanck, F., Stepanek, C., Swann, G. E. A., Telford, R., Thomas, E., Thomas, Z., Truebe, S., von Gunten, L., Waite, A., Weitzel, N., Wilhelm, B., Williams, J., Williams, J. J., Winstrup, M., Zhao, N., and Zhou, Y. (2019), "PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data", *Paleoceanography and Paleoclimatology* 34 (10), pp. 1570–1596, DOI: 10.1029/2019PA003632.

Komar, P. D., Morse, A. P., Small, L. F., and Fowler, S. W. (1981), "An Analysis of Sinking Rates of Natural Copepod and Euphausiid Fecal Pellets", *Limnology and Oceanography* 26 (1), pp. 172–180, DOI: 10.4319/lo.1981.26.1.0172.

Kwon, E. Y., Primeau, F., and Sarmiento, J. L. (2009), "The Impact of Remineralization Depth on the Air–Sea Carbon Balance", *Nature Geoscience* 2 (9), pp. 630–635, DOI: 10.1038/ngeo612.

Lisiecki, L. E. and Raymo, M. E. (2005), "A Pliocene-Pleistocene Stack of 57 Globally Distributed Benthic δ^{18}O Records", *Paleoceanography* 20 (1), DOI: 10.1029/2004PA001071.

López-Urrutia, Á., Martin, E. S., Harris, R. P., and Irigoien, X. (2006), "Scaling the Metabolic Balance of the Oceans", *Proceedings of the National Academy of Sciences* 103 (23), pp. 8739–8744, DOI: 10.1073/pnas.0601137103, pmid: 16731624.

Lutz, M., Dunbar, R., and Caldeira, K. (2002), "Regional Variability in the Vertical Flux of Particulate Organic Carbon in the Ocean Interior", *Global Biogeochemical Cycles* 16 (3), DOI: 10.1029/2000GB001383.

Marsay, C. M., Sanders, R. J., Henson, S. A., Pabortsava, K., Achterberg, E. P., and Lampitt, R. S. (2015), "Attenuation of Sinking Particulate Organic Carbon Flux through the Mesopelagic Ocean", *Proceedings of the National Academy of Sciences* 112 (4), pp. 1089–1094, DOI: 10.1073/pnas.1415311112, pmid: 25561526.

Martin, J. H., Knauer, G. A., Karl, D. M., and Broenkow, W. W. (1987), "VERTEX: Carbon Cycling in the Northeast Pacific", *Deep Sea Research Part A. Oceanographic Research Papers* 34 (2), pp. 267–285, DOI: 10.1016/0198-0149(87)90086-0.

Matsumoto, K. (2007), "Biology-Mediated Temperature Control on Atmospheric pCO_2 and Ocean Biogeochemistry", *Geophysical Research Letters* 34 (20), DOI: 10.1029/2007GL031301.

Matsumoto, K., Hashioka, T., and Yamanaka, Y. (2007), "Effect of Temperature-Dependent Organic Carbon Decay on Atmospheric pCO_2", *Journal of Geophysical Research: Biogeosciences* 112 (G2), DOI: 10.1029/2006JG000187.

Mazuecos, I. P., Arístegui, J., Vázquez-Domínguez, E., Ortega-Retuerta, E., Gasol, J. M., and Reche, I. (2015), "Temperature Control of Microbial Respiration and Growth Efficiency in the Mesopelagic Zone of the South Atlantic and Indian Oceans", *Deep Sea Research Part I: Oceanographic Research Papers* 95, pp. 131–138, DOI: 10.1016/j.dsr.2014.10.014.

McCorkle, D. C., Corliss, B. H., and Farnham, C. A. (1997), "Vertical Distributions and Stable Isotopic Compositions of Live (Stained) Benthic Foraminifera from the North Carolina and California Continental Margins", *Deep Sea Research Part I: Oceanographic Research Papers* 44 (6), pp. 983–1024, DOI: 10.1016/S0967-0637(97)00004-6.

McKay, N. P. and Emile-Geay, J. (2016), "Technical Note: The Linked Paleo Data Framework – a Common Tongue for Paleoclimatology", *Climate of the Past* 12 (4), pp. 1093–1100, DOI: 10.5194/cp-12-1093-2016.

Murray-Wallace, C. V. and Woodroffe, C. D. (2014), "Pleistocene Sea-Level Changes", *Quaternary Sea-Level Changes: A Global Perspective*, Cambridge: Cambridge University Press, pp. 256–319, DOI: 10.1017/CBO9781139024440.

Niebler, H.-S., Hubberten, H.-W., and Gersonde, R. (1999), "Oxygen Isotope Values of Planktic Foraminifera: A Tool for the Reconstruction of Surface Water Stratification", *Use of Proxies in Paleoceanography*, ed. by G. Fischer and G. Wefer, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 165–189, DOI: 10.1007/978-3-642-58646-0_6.

Oliver, K. I. C., Hoogakker, B. A. A., Crowhurst, S., Henderson, G. M., Rickaby, R. E. M., Edwards, N. R., and Elderfield, H. (2010), "A Synthesis of Marine Sediment Core

δ^{13}C Data over the Last 150 000 Years", *Climate of the Past* 6 (5), pp. 645–673, DOI: 10.5194/cp-6-645-2010.

Olsen, A. and Ninnemann, U. (2010), "Large δ^{13}C Gradients in the Preindustrial North Atlantic Revealed", *Science* 330 (6004), pp. 658–659, DOI: 10.1126/science.1193769.

Peterson, C. D., Lisiecki, L. E., and Stern, J. V. (2014), "Deglacial Whole-Ocean δ^{13}C Change Estimated from 480 Benthic Foraminiferal Records", *Paleoceanography* 29 (6), pp. 549–563, DOI: 10.1002/2013PA002552.

Ravelo, A. C. and Fairbanks, R. G. (1995), "Carbon Isotopic Fractionation in Multiple Species of Planktonic Foraminifera from Core-Tops in the Tropical Atlantic", *The Journal of Foraminiferal Research* 25 (1), pp. 53–74, DOI: 10.2113/gsjfr.25.1.53.

Reiner, S. (2020), *Ocean Data View*, version 5.2.1, URL: https://odv.awi.de/.

Rumble, J. R., Lide, D. R., and Bruno, T. J. (2018), *CRC Handbook of Chemistry and Physics*, 99th ed., Boca Raton: CRC Press.

Schmidt, D. N., Elliott, T., and Kasemann, S. A. (2008), "The Influences of Growth Rates on Planktic Foraminifers as Proxies for Palaeostudies – a Review", *Geological Society, London, Special Publications* 303 (1), pp. 73–85, DOI: 10.1144/SP303.6.

Segschneider, J. and Bendtsen, J. (2013), "Temperature-Dependent Remineralization in a Warming Ocean Increases Surface pCO_2 through Changes in Marine Ecosystem Composition", *Global Biogeochemical Cycles* 27 (4), pp. 1214–1225, DOI: 10.1002/2013GB004684.

Shackleton, N. and Hall, M. (1984), "Oxygen and Carbon Isotope Stratigraphy of Deep-Sea Drilling Project Hole-552a - Plio-Pleistocene Glacial History", *Initial Reports of the Deep Sea Drilling Project* 81 (DEC), pp. 599–609, DOI: 10.2973/dsdp.proc.81.116.1984.

Sigman, D. M. and Boyle, E. A. (2000), "Glacial/Interglacial Variations in Atmospheric Carbon Dioxide", *Nature* 407 (6806) (6806), pp. 859–869, DOI: 10.1038/35038000.

Smithson, M. (1992), *Pelagic Tidal Constants 3*, Publication Scientifique 35, IAPSO, 191 pp., URL: http://iapso.iugg.org/images/stories/pdf/IAPSO_publications/Publications_Scientifiques/Pub_Sci_No_35.pdf (visited on 06/09/2020).

Spero, H. J., Lerche, I., and Williams, D. F. (1991), "Opening the Carbon Isotope 'vital Effect' Black Box, 2, Quantitative Model for Interpreting Foraminiferal Carbon Isotope Data", *Paleoceanography* 6 (6), pp. 639–655, DOI: 10.1029/91PA02022.

Spero, H. J., Mielke, K. M., Kalve, E. M., Lea, D. W., and Pak, D. K. (2003), "Multispecies Approach to Reconstructing Eastern Equatorial Pacific Thermocline Hydrography during the Past 360 Kyr", *Paleoceanography* 18 (1), DOI: 10.1029/2002PA000814.

Steph, S., Regenberg, M., Tiedemann, R., Mulitza, S., and Nürnberg, D. (2009), "Stable Isotopes of Planktonic Foraminifera from Tropical Atlantic/Caribbean Core-Tops: Implications for Reconstructing Upper Ocean Stratification", *Marine Micropaleontology* 71 (1), pp. 1–19, DOI: 10.1016/j.marmicro.2008.12.004.

Szatmari, P. (1968), *VEMA 22-193: Megascopic Description of a Split Core*, URL: https://www.ngdc.noaa.gov/mgg/geology/data/vema/vm22/193/vm22_193pc_description.pdf (visited on 09/15/2020).

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., and Poulsen, C. J. (2020), "Glacial Cooling and Climate Sensitivity Revisited", *Nature* 584 (7822) (7822), pp. 569–573, DOI: 10.1038/s41586-020-2617-x.

Volk, T. and Hoffert, M. I. (1985), "Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO_2 Changes", *The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present*, ed. by E. Sundquist and W. S. Broecker, American Geophysical Union (AGU), pp. 99–110, DOI: 10.1029/GM032p0099.

Wilson, J. D., Barker, S., Edwards, N. R., Holden, P. B., and Ridgwell, A. (2019), "Sensitivity of Atmospheric CO_2 to Regional Variability in Particulate Organic Matter Remineralization Depths", *Biogeosciences* 16 (14), pp. 2923–2936, DOI: 10.5194/bg-16-2923-2019.

Yamamoto, A., Abe-Ouchi, A., and Yamanaka, Y. (2018), "Long-Term Response of Oceanic Carbon Uptake to Global Warming via Physical and Biological Pumps", *Biogeosciences* 15 (13), pp. 4163–4180, DOI: 10.5194/bg-15-4163-2018.