Does a proxy measure up?: A framework to assess and convey proxy reliability

F. Garrett Boudinot[1], Joseph Wilson[2]

[1] Department of Geological Sciences and Institute of Arctic and Alpine Research, University of Colorado Boulder,

5   Boulder, Colorado, 80309, USA

[2] Department of Philosophy and Center for the Study of Origins, University of Colorado Boulder, Boulder,

Colorado, 80309, USA

Correspondence to: F. Garrett Boudinot (garrett.boudinot@colorado.edu)

Abstract. Earth scientists describe a wide range of observational measurements as "proxy measurements." By referring to such a vast body of measurements simply as "proxy," workers dilute significant differences in the various ways that measurements relate to the phenomena they intend to describe. The limited language around these measurements makes it difficult for the non-specialist to assess the reliability and uncertainty of data generated from "proxy" measurements. Producers and reviewers of proxy data need a common framework for conveying proxy measurement methodology, uncertainty, and applicability for a given study.

We develop a functional distinction between different forms of measurement based on the different ways that their outputs (values, interpretations) relate to the phenomena they intend to describe (e.g., temperature). Paleotemperature measurements, which are used to estimate temperatures of systems in Earth's past, serve as a case study to examine and apply this new functional proxy definition. We explore the historical development and application of two widely-used paleotemperature proxies, calcite $\delta^{18}O$ and TEX_{86}, to illustrate how different measurements relate to the phenomena they intend to describe. Both proxies are vulnerable to causal factors that interfere with their relationship with temperature, but address those "confounding causal factors" in different ways. While the goal of proxy development is to fully identify, quantify, and calibrate to all confounding causal factors, the reality of proxy applications, especially for past systems, engenders unavoidable and potentially significant uncertainties. We propose a framework that allows researchers to be explicit about the limitations of their proxies, and identify steps for further development. This paper underscores the ongoing effort and continued need for critical examination of proxies throughout their development and application, particularly in Earth history, for reliable proxy interpretation.

1 Introduction

Proxy measurements can provide information about otherwise elusive properties of systems in Earth's past, present, and worlds beyond. With a growing interest in quantitatively measuring these properties more precisely and in new environments, the diversity of proxies has increased dramatically. While "proxy" is often used to differentiate "indirect" (e.g., geochemical, physical, etc.) measurements from more "direct" forms of observational measurement, neither of those terms provide insight into the reliability or applicability of different measurements. Even "direct" forms of measurement can be considered proxy in this sense; all involve some level of observational "indirectness" (Wilson and Boudinot, *in review*). Earth scientists are particularly aware of the nuances of measurement applicability – as workers look farther back in time, the reliability of a measurement (i.e., our understanding of what that measurement represents) typically becomes less certain. A standardized framework for conveying how proxy measurements relate to different systems and phenomena would be widely useful for describing these complex associations to non-specialists, students, modelers, and other proxy users.

The goal of this paper is to describe how methods of observational measurements differ in the ways their outputs (values, data, interpretations) relate to the phenomena they intend to describe. All forms of observational measurement are influenced by factors that are not the property being measured. We provide insight into the assumptions behind the interpretation and development of different forms of measurement, with the goal of more clearly describing those assumptions and uncertainties in the context of data interpretations.

We examine paleotemperature measurements, which are used to estimate temperatures of systems in Earth's past, as a case study given the growing interest in paleoclimate, the diversity of measurements available, and the field's relationship to unknown changes in the Earth-climate system through time. We propose a theoretical framework and language that can more accurately distinguish different measurement-property relationships, which we hope will lead to more robust measurement calibrations, more transparent measurement outputs, and stronger interpretations. While paleoclimate is the focus below, the ideas described here apply to observational measurements across many fields of science.

2 Functional distinctions for proxy measurements

The placement of measurements in two overarching groups, proxy and direct, is particularly common in climate sciences (NOAA National Centers for Environmental Information; Jansen et al., 2007). Philosophical work (Wilson and Boudinot, *in review*) has pointed out the need for clarification behind the definition of proxy measurements as "indirect" and non-proxy measurements as "direct," and questioned how proxies can provide reliable measurements in spite of such perceived indirectness. While many have referred to oxygen isotopes in calcite ($\delta^{18}O_{calcite}$) as a proxy for temperature, and the mercury thermometer as a direct measurement of temperature (NOAA National Centers for Environmental Information; Jansen et al., 2007), both scientists and philosophers of science have pointed out that neither measurement technique truly represents "direct" observation (e.g., Ruddiman, 2008; Wilson and Boudinot, *in review*). The mercury thermometer measures temperature via the observable thermal expansion of mercury as a function of temperature, while the $\delta^{18}O_{calcite}$ measures paleotemperature via observable variation of ^{18}O incorporation into calcite ($CaCO_3$) as a function of temperature, resulting from the differences in vibrational energies of different oxygen isotopes (i.e., ^{16}O , ^{17}O, ^{18}O). In other words, neither produces a "direct" measurement of temperature; both rely on the observation of some effect of temperature in a system.

Each of these measurements are also influenced by other non-temperature causal factors. Mercury expansion is not only a function of temperature, but also of the partial pressure of the atmosphere and expansion dynamics of liquid mercury. Similarly, $\delta^{18}O_{calcite}$ is influenced by the $\delta^{18}O$ of the surrounding water ($\delta^{18}O_{H2O}$; Urey, 1948), the pH of the surrounding water (Spero et al., 1997), and if biomineralized by calcifying organisms, biological kinetic effects on ^{18}O incorporation (Bemis et al., 1998; Ravelo and Hillaire-Marcel, 2007). Philosophers attuned to the conceptual and epistemic issues regarding different forms of scientific measurement (e.g. Suppes, 1951; Franklin, 1990; Chang, 2004; Van Fraassen, 2010; Wilson and Boudinot, *in review*) have recently proposed that proxies differ from other forms of measurement in how they account for these *confounding causal factors* (CCFs; see glossary of terms).

Under this definition, non-proxy measurements are those that have been designed and manufactured to eliminate all of the potential effects of known CCFs on the measurement output. Because these non-proxy measures control which parts of the system contribute to the final measurement outputs, we refer to them as *controlled* measurements (see glossary of terms).

Mercury thermometers, for example, are manufactured with a glass casing that controls the atmospheric pressure within the thermometer. The glass case eliminates variation in non-temperature CCFs (e.g., changes in atmospheric pressure, potential for fluid exchange) such that the measured signal can only represent the phenomena in question, temperature. The lines on the thermometer are calibrated to the thermodynamic properties of mercury, such that a specific volumetric expansion of mercury is a causal result of the specific local temperature. In this way, the mercury thermometer is used to perform a controlled measurement.

While the process is more sophisticated, the digital thermometers more commonly used today also control all known CCFs within the instrument to provide a single, calibrated temperature value. For those digital thermometers that use electrical resistance, for example, the built-in computer immediately converts an electrical resistance reading to temperature, and is calibrated to effectively remove the influence of non-temperature effects on such resistance, including the composition, length, and width of the metal probe used in the thermometer. Because digital thermometers account for all CCFs that influence the relationship between electrical resistance and temperature in real-time, digital thermometers, too, are used to perform controlled measurements.

Proxy measurements are distinct because their process of measurement does not rule out all CCFs (see glossary of terms). This means that the original signal from the analytical measurement must be subject to further manipulation, such as incorporation into a calibration. Those calibrations are based on the field's best understanding of the drivers of that measured property, and quantitatively attempt to minimize the influence of CCFs to produce a value that represents the phenomena in question (Fig. 1). For example, $\delta^{18}O_{calcite}$ is a proxy measurement because $\delta^{18}O_{calcite}$ is measured simply as a ratio of ^{18}O to ^{16}O of a calcite sample compared to an isotopic standard, and alone that analytical measurement does not reflect temperature. To measure temperature using $\delta^{18}O_{calcite}$, researchers must use a calibration that incorporates information about other parts of the system that influence the inclusion of ^{18}O into calcite, such as the $\delta^{18}O_{H2O}$ of the surrounding water, and any potential biological effects of calcification. Because most proxy applications do not allow the researcher to produce controlled measurements of each of those CCFs, the output from a proxy is at best an "estimate" (i.e., the $\delta^{18}O_{calcite}$ proxy measurement produces paleotemperature estimates).

The term "indicator" is often used synonymously with "proxy" or even "measurement" (e.g., *"Application of the Ce anomaly as a paleoredox indicator,"* German and Elderfield, 1990; "Using fossil leaves as paleoprecipitation indicators,' Wilf et al., 1998; "Stomatal density and stomatal index as indicators of paleoatmospheric CO_2 concentrations," Royer, 2001; "indicator of relative changes in sea surface temperature," Hollis et al., 2019; "Palaeoecological proxies…include crustacean Ostracoda…their indicator species…are sensitive to deoxygenation and eutrophication," Yasuhara et al., 2019). The use of this term for such wide range of applications highlights the lack of clarity in the existing literature, which eventually leads to a lack of clarity in the dissemination of resulting information. While all measurements do "indicate" the quality of some property, they do so in different ways, and are accompanied by quite different levels of reliability and uncertainty. The proposed distinction between proxy and controlled measurements, and within proxy measurements (see below), is aimed to provide clarity to the discussion of measurements and their outputs – and CCFs provide such clarification.

The importance of CCFs for proxy measurements was recognized in the development of the first quantitative paleotemperature proxy, $\delta^{18}O_{calcite}$. Harold Urey first described the thermodynamic relationship between $\delta^{18}O_{calcite}$ and calcite formation temperatures through a simple linear calibration that relates $\delta^{18}O_{calcite}$ to temperature in degrees Celsius (Urey 1948). Urey discussed two important CCFs influencing the $\delta^{18}O_{calcite}$ relationship with temperature that could have changed significantly through geologic time and space, namely the $\delta^{18}O_{H2O}$ of the (mean) global ocean, and $\delta^{18}O_{H2O}$ of local waters surrounding the precipitating carbonate. While the early reports posited that global $\delta^{18}O_{H2O}$ changed on long timescales (millions of years) as a result of rock weathering, later work showed that global $\delta^{18}O_{H2O}$ had varied significantly on much shorter timescales (tens of thousands of years) due to fluctuations of global ice volume (Emiliani, 1955). The uncertainty of mean ocean $\delta^{18}O_{H2O}$ is greater farther back in Earth history, due to currently unconstrained conditions such as ancient ocean latitudinal gradient effects (i.e., reduced latitudinal temperature gradient and resultant local $\delta^{18}O_{H2O}$ ~100 million years ago) and silicate weathering rates (Urey et al., 1951). Most Earth systems have experienced variability through Earth history, contributing to increased uncertainty associated with CCFs moving farther back in geologic time. As such, different temporal applications of a single proxy can dramatically change that proxy estimate's uncertainty.

The potential for unknown CCFs exists even for well-calibrated proxy systems and control measurements (Wilson and Boudinot, *in review*). While the mercury thermometer successfully controls for its relevant CCFs, a hypothetical application that reveals a theretofore unknown CCF would lead us to no longer consider the thermometer a controlled measurement, at least until it were manufactured in a way to also remove the effects of that CCF. The potential for the existence of unknown CCFs necessitates cautious interpretations of all measurements, particularly those in development or under new applications. But how exactly are CCFs incorporated into proxies?

3 Assessing a proxy

3.1 Situating proxies on a spectrum

CCFs are incorporated into proxy measurements through a calibration equation (Fig. 1), which provides a quantitative representation of the relative influence of each causal factor that contributes to the measured property. Using the calibration, researchers can account for the influence of CCFs, and produce an estimate of the phenomenon in question. However, the extent to which calibrations identify and address CCFs differs greatly between proxies and proxy applications.

We place proxy measurements along a spectrum that can illustrate the diversity of how proxies relate to CCFs (Fig. 2a). Controlled measurements, with all CCFs known and controlled for (e.g., mercury thermometer), occupy one end of the spectrum. On the other end of the spectrum are proxy measures that are not (yet) calibrated to directly account for their CCFs, such that only a correlation is proposed (*correlation-constrained proxy;* see glossary of terms), carrying uncertainty regarding the nature and precise causal influence of associated CCFs. In between the two ends of the spectrum are proxies which have a calibration that accounts for the CCFs' influence on the measurement output, and are accompanied by a quantitative measurement (*observation-constrained proxy*) or quantitative inference (*inference-constrained proxy*) of those CCFs (Fig. 2a; see glossary of terms). By situating any measurement along this spectrum, one can assess how much the measured value is affected by CCFs as opposed to the property in question (i.e., the *potential uncertainty*, Fig. 2b, see below), such as $\delta^{18}O_{H2O}$ instead of temperature.

Controlled measurements work the same across locations and through time. A mercury thermometer should have the same level of accuracy and precision in a high-altitude, low-humidity study site as in a low-altitude, high-humidity site. Ideally, all proxy measurements would eventually develop into controlled measurements. Unfortunately, and particularly in paleo applications, the certainty ascribed to the mercury expansion calibration is not easily attainable or validated. Furthermore, even controlled measurements can be complicated by work in "extreme" environments, where temperatures may exceed the minimum or maximum range to which the thermometer is calibrated (e.g., beyond the boiling point of mercury). Thus, how a measurement's calibration is developed and utilized determines the situations and uncertainty for that measurement's application.

To illustrate the proxy range of the spectrum, we situate $\delta^{18}O_{calcite}$ as either an *observation-constrained proxy* or an *inference-constrained proxy* depending on how CCFs are quantitatively accounted for (Fig. 2a). When the $\delta^{18}O_{H2O}$ value in the temperature calibration derives from an independent measurement (proxy or controlled) of the $\delta^{18}O_{H2O}$ of the water from which the calcite precipitated, then the proxy is an *observation-constrained proxy;* values to account for the CCFs in the calibration derive from empirical observations (Fig. 2a). These components of the calibration can be accounted for with information from proxy or controlled measurements, with the latter contributing less uncertainty given the constraints on CCFs in controlled measurements.

On the other hand, in instances where $\delta^{18}O_{H2O}$ cannot be measured, such as in deeper-time applications, the researcher must provide an *inference* (i.e., reasoned approximation) of local $\delta^{18}O_{H2O}$. Based on the extrapolation of a well-known system to a lesser-known system, *inference-constrained proxy* measurements inherently present a more biased estimate, due to biases in the researchers' inference of that system, rather than empirical evidence (Fig. 2b). For example, some researchers have inferred 100 million year-old $\delta^{18}O_{H2O}$ for the $\delta^{18}O_{calcite}$ paleotemperature calibration by applying a first-order estimate of $\delta^{18}O_{H2O}$ based on certain characteristics of the system in question, such as a mean $\delta^{18}O_{H2O}$ value that applies to any "non-glacial world" (O'Brien et al., 2017). Workers modified this mean value to represent $\delta^{18}O_{H2O}$ of local waters (where calcite was precipitated) by adjusting the mean $\delta^{18}O_{H2O}$ based on modern latitudinal $\delta^{18}O_{H2O}$ variability (e.g., O'Brien et al., 2017). This inference is still based on quantitative measurements (e.g., modern $\delta^{18}O_{H2O}$ latitudinal trends), but requires several

inferences that assume that two systems are similar (i.e., all ice free oceans in Earth history are

230   isotopically similar; latitudinal $\delta^{18}O_{H2O}$ variability is similar between 100 million years ago and

the present). Because that inference is accompanied by uncertainty that is not easily quantifiable

(e.g., uncertainty associated with assumptions made by the researcher, rather than analytical

uncertainty, see below), the potential uncertainty for inference-constrained proxies is larger than

those that are observation-constrained.

235   Importantly, many calibrations require a combination of inference and observation to

produce a final estimate of the target property, as CCFs differ in how they can be accounted for.

In other words, many proxy applications use both observation- and inference-constraints to

satisfy a calibration.

Moving further away from controlled measurements on our spectrum, we find proxy

240   measurements that are *correlated* with temperature, but the CCFs are not fully or quantitatively

accounted for in a calibration; here, the CCFs are unknown (or roughly understood), though a

corollary relationship is identified. It is functionally impossible to accurately assess the

uncertainty of estimates produced by these measurements (Fig. 2b), as the causal factors

influencing the measurement are not quantitatively represented in a calibration. Not only could

245   the signal from such *correlation-constrained proxy* be partially driven by some unknown CCF, it

could even be entirely driven by CCFs (e.g., Junium et al., 2018), but would interpreted as driven

by the property in question.

An example of a correlation-constrained proxy is the present incarnation of the TEX_{86}

paleotemperature proxy. In 2002, workers identified a suite of sedimentary hydrocarbons that

250   shared a similar structure, but contained a different number of cyclic moieties (Schouten et al.,

2002; Fig. 2). Relative abundances of these isoprenoidal glycerol diether glycerol tetraether

(isoGDGT) compounds with different cyclic moieties were represented by a ratio (Table 1).

When these compounds were recovered from modern sediments and this ratio was calculated, a

clear correlation with the surface water temperature at the sample location was identified. In

255   other words, the number of cyclic moieties in the sedimentary isoGDGTs were correlated with

the surface water temperatures at the location that they were found. Using statistical (regression)

analyses of a suite of modern sediments and sea surface temperature measurements, a calibration

was produced, and the authors proposed this molecular ratio as a quantitative paleotemperature

proxy (Schouten et al., 2002). A physiological response was posited to explain the relationship –

260   less cyclic moieties contributed to a more malleable lipid membrane, which would be

advantageous in cooler waters.

In the ensuing years, several questions about the origin and implications of these

molecules were raised. They seemed to be produced predominantly by Thaumarcheota, a type of

marine archaea that live well below the sea surface (Schouten et al., 2000), where the

265   temperature correlation was strongest. Additionally, field and culture observations from variable

environments produced different calibrations (i.e., different slopes and y-intercepts to describe

the correlation between the isoGDGT ratio and temperature; Table 1) and even different ratios

(e.g., TEX_{86}^{L} for low-temperature regions; Table 1). If the ratio of isoGDGT cyclicity directly

represented temperature, then why would that ratio be different depending on the study design,

270   location, and time period? And if the calibration accurately accounted for the CCFs contributing

to the effect of temperature on isoGDGT cyclicity, why would it be different from place to

place?

These questions are driving fundamental research in understanding the *mechanistic*

relationships between TEX_{86} and temperature. Several important advances in this mechanistic

275   understanding have already been produced: culture and field experiments have shown that the

cyclic moieties represent a metabolic response to energy demands, growth phase, nutrient

availability, and ecosystem composition, rather than solely a physiological response to

temperature (Elling et al., 2014; Qin et al., 2015; Hurley et al., 2016; Polik et al., 2018). These

studies advance TEX_{86} beyond the corollary relationship (i.e., colder temperatures makes more

280   cyclic moieties) into a nuanced, yet more accurately representative, understanding of all causal

factors and their mechanisms (i.e., relationship between sea surface temperatures and nutrient

and oxygen availability, which impacts archaeal metabolic energy demands). However, while

work on TEX_{86} drivers suggest that non-temperature factors cause variations in isoGDGT

cyclization, TEX_{86} application studies continue to report a specific temperature value. The

285   argument behind continued TEX_{86} applications is the correlation of ammonia oxidation rates and

temperature in *most modern settings* (Hurley et al., 2016). However, many studies have

suggested that ammonia or oxygen concentrations in past environments likely varied in a way

that did not correlate with temperature (e.g., Liu et al., 2009; Polik et al., 2018). This proxy's

CCFs need full consideration in experimental design and interpretation for it to be truly

290   quantitative – and its uncertainty appropriately reported.

3.2 Discussing proxy data

A clear distinction should be made between various forms and degrees of uncertainty related to proxy measurements (see glossary of terms). All proxy measurements are the result of some analysis (e.g., $\delta^{18}O_{calcite}$ as the normalized ratio of ^{18}O to ^{16}O in a sample) and incorporation into a calibration (e.g., $\delta^{18}O_{calcite}$ as a function of temperature, $\delta^{18}O_{H2O}$, and biological effects; Fig. 1), from which derives three forms of uncertainty. The first is *analytical uncertainty*, which is simply the uncertainty associated with the precision and accuracy of the analytical measurement. For oxygen isotopes in calcite, this would include the isotope ratio mass spectrometer's precision and accuracy when determining the ratio of ^{18}O to ^{16}O of a sample normalized to a standard. We argue that analytical uncertainty can always be quantified using standards, and is distinct from unquantifiable uncertainties. Unquantifiable uncertainties associated with calibration (including unknown CCFs), as well as sample preparation and analysis, and are grouped into *potential uncertainties* (Fig. 2b). The distinction between factors that fall into potential versus analytical uncertainty is defined by quantitation. Researchers take many steps to quantify errors and uncertainties associated with sample preparation and analysis. When employed, such efforts reduce the potential uncertainty and more accurately reflect that analytical uncertainty. For example, hydrocarbon standards might be incorporated into a sedimentary sample before hydrocarbon extraction, such that the researcher can quantify if any hydrocarbons, including isoGDGTs, are lost or altered throughout the in-lab processing. Researchers could report or normalize to that loss and alteration, more transparently reflecting the uncertainty in the analysis. However, some potential uncertainties will always exist in a non-quantifiable manner, such as unknown CCFs or un-measureable changes in CCFs through time. Because the error in an inference-constrained proxy might not be quantifiable (i.e., logical deductions might not have a quantifiable uncertainty), its potential uncertainty will always be higher than an observation-constrained proxy, where the analytical uncertainty of the CCF measurement can be quantified (Fig. 2b).

The final type of uncertainty is the *reported uncertainty*, which should ideally cover (either quantitatively or in discussion) both analytical and potential uncertainties. However, for many proxies, the reported uncertainty varies widely in practice. For example, the variety of isoGDGT ratios and calibrations (Table 1), and the lack of codified reporting standards used in the expression of TEX_{86}-derived paleotemperatures, leads to notable variability in the reported

295
300
305
310
315
320

uncertainty associated with TEX_{86}. Some TEX_{86}-derived paleotemperature estimates are plotted without error bars and accompanied by an in-text discussion of the analytical uncertainty from calibration and replicate analyses (e.g., Woelders et al., 2017), while the analytical uncertainty

325 for others are not discussed (e.g., Slujis et al., 2006). For some estimates, the analytical uncertainty derived from only the calibration is provided (e.g., Hollis et al., 2012; Ho et al., 2014). Analytical uncertainties from replicate analyses have been combined with the analytical uncertainties of calibration statistics as error windows on plots (e.g., Tierney et al., 2010; Shevenell et al., 2011), while discussion of potential uncertainties, such as changes in the known

330 (but not calibrated-to) CCFs, varies greatly between reports (e.g., Tierney et al., 2010; Shevenell et al., 2011). Because potential uncertainty is by-definition unquantifiable, it might not be incorporated into quantitative data presentation styles such as Cartesian plots, but can certainly be discussed in light of the existing work on TEX_{86} CCFs.

Importantly, researchers have taken steps to communicate the reliability of proxy data

335 relative to other measurements in reviews, conference sessions, and proxy assessment compilations (e.g. Ravelo and Hillaire-Marcel, 2007; Newman et al., 2016; Hollis et al., 2019; Wilson and Boudinot, 2019). For example, the Paleoclimate Modelling Intercomparison Project (PMIP)'s appraisal of proxy data for the Intergovernmental Panel on Climate Change (IPCC) reports (Hollis et al., 2019) provides an in-depth description of the paleotemperature proxies

340 used to inform the IPCC reports. The appraisal describes each proxy's theoretical background, which gives data generators and modelers a better understanding of the biogeochemical processes that relate each proxy to temperature. The assessment then describes strengths and weaknesses of each proxy relative to the other measurements, which can guide users in determining which proxy may be best suited for a given study, as well as providing

345 considerations for the interpretation of the resulting data. Finally, the assessment provides "recommended methodologies," which includes analytical recommendations, a single recommended calibration, and other best-practices for reporting proxy data and interpretations. By providing a consensus presentation of recommended methodologies particularly, the PMIP proxy assessment and similar projects constitute an important means for standardizing data

350 assessment and reporting, and guiding proxy users in developing study designs. The framework presented here will improve those methods by providing direct language (e.g., CCFs, types of uncertainty) to more clearly navigate discussions of proxy assessments.

A complete outline of potential uncertainties and the often complex phenomena-measurement relationships is difficult to incorporate into grants, peer-reviewed manuscripts, and

355   educational programs. The lack of extensive discussion of a proxy's uncertainty can lead to an over-simplification of these relationships (i.e., an under-consideration for CCFs and uncertainties). However, detailing how proxies might relate to some unknown CCFs (as is done here) can make any proxy seem subject to countless unknown CCFs, which may engender an unwarranted dismissal of proxy data interpretations. Because proxy data informs models,

360   manuscripts, and educational lessons, there needs to be a more universally accepted and functional means of discussing and conveying proxy uncertainty that is honest yet robust. Our spectrum of proxy measurements relates measurements to their CCFs, and thus the spectrum and language provide such a means of conveying uncertainty in a universal way.

Many studies, for example, have shown that TEX_{86} trends were driven by changes in

365   nitrogen availability and marine ecology in some paleo environments (Liu et al., 2009; Hurley et al., 2016; Junium et al., 2018; Polik et al., 2018). How can workers be sure that TEX_{86} is not driven by these dynamics in other settings, unless those CCFs of nitrogen availability and marine ecology changes are directly assessed? Because uncertainty in estimating these environmental characteristics are often not incorporated (as they are not incorporated in the current litany of

370   quantitative TEX_{86} calibrations; Table 1), we have described the *potential uncertainty* of TEX_{86} (and other correlation-constrained proxies) as much higher than is often reported (Fig. 2b). By referring to TEX_{86} as a *correlation-constrained proxy*, modelers, reviewers, and researchers can immediately be aware of this under-reporting of uncertainty, which would inform their interpretation of the temperature estimates produced by TEX_{86} in a meaningful yet succinct way.

375

3.3 Development of a proxy

Proxy development is the production and improvement of a calibration which quantitatively accounts for all CCFs that contribute to the measured signal. The controlled characteristic of a mercury thermometer allows the measurement of temperature without needing an external

380   calibration, as the temperature lines are calibrated to the exact expansion of mercury within the glass walls. Prior to the full calibration of the lines on the mercury thermometer, mercury might have served as a proxy: a gram of mercury on a table would expand and contract with fluctuating

temperatures, which could be a qualitative, correlation-constrained proxy for temperature (the mercury expanded, so the temperature likely got hotter).

385   Because proxy measurements do not account for the influence of all known CCFs, quantitative proxy measurements require some external calibration equation to produce reliable estimates. Calibrations express the relative effect of each causal factor (Fig. 1), and provide insight into the applicability of a proxy by addressing the range in which the calibration is useful, and the natural variability (uncertainty) associated with that calibration. Proxy applications are

390   limited to the range in which that proxy has been studied and calibrated; applications outside that range do not produce reliable estimates.

         Harold Urey's first description of the thermodynamic relationship between $\delta^{18}O_{calcite}$ and calcite formation temperatures was simply "The calculated slope, 4.4 per mil between 0°C and 25°C" (Urey, 1948). More complex calibrations now exist for the $\delta^{18}O_{calcite}$ paleotemperature

395   proxy, which accounts for its numerous CCFs including $\delta^{18}O_{H2O}$ and biological effects (Ravelo and Hillaire-Marcel, 2007; Hollis et al., 2019). While the $\delta^{18}O_{calcite}$ proxy is far from a controlled measurement, its historical development exemplifies the consistent work to make proxies more like a controlled measurements, i.e., to eliminate or limit the influence of CCFs. But what does such proxy development look like in practice?

400   The first step of proxy development is the identification of some corollary relationship between a measurable property (e.g., $\delta^{18}O$ of calcite) and property unable to be measured in a controlled fashion (e.g., temperature of a past environment). At first order, these are usually qualitative and based on some hypothesis to describe a system. Mercury expands with increasing temperature due to general fluid dynamics; ^{18}O is more favorably incorporated into calcite at

405   lower temperatures due to differences in vibrational energies between ^{18}O and ^{16}O; some organisms alter their cell membranes to maintain homeostasis in variable environments.

         Proxies that are based on such a corollary relationship can serve as *qualitative proxy measures*, which provide useful comparative or relative information. This is the case for some paleotemperature proxies: geological evidence of glacial expansion and retreat in a certain

410   location can indicate relative local temperature change, but variability in numerous (difficult or impossible to constrain) CCFs prohibits a calibration to quantitative temperature changes in degrees Celsius. Such comparative information is appropriate for many paleo studies, where the question is focused on trends and relative changes through time or differences between sites.

This corollary relationship can lead researchers into an "optimism phase," where the assumption

415   of a direct, cause-effect relationship between a phenomena and an observation makes users

optimistic that a proxy can be used with confidence (Elderfield, 2002).

If researchers aim to use a proxy quantitatively, the relationship between the target

property (e.g., temperature), the observable property (e.g., $\delta^{18}O_{calcite}$), and all CCFs must be

accounted for in a calibration (Fig. 1). Quantitative proxies require an (empirically derived)

420   estimation or (logically deduced) inference of the influence of all CCFs represented in a

calibration. Calcite precipitation experiments with variable pH, $\delta^{18}O_{H2O}$, salinity, and

biomineralizing organisms have contributed to calibrations that include those CCFs, and

represent how they contribute to ^{18}O incorporation into calcite (Ravelo and Hillaire-Marcel,

2007). Studies using those calibrations must account for those CCFs. For example, calcite-

425   producing organisms live in either bottom waters or surface waters – the temperature from the

two will not only have slightly different CCFs, but will also reflect temperature from different

parts of the water column. Workers would identify the type of organisms to know where it lived,

and would address the CCFs specific to that organism (e.g., Bemis et al., 1998). The process of

testing CCFs must be extensive to provide confidence in the proxy. Often, this phase of

430   development unearths unforeseen CCFs, such as the role of water column oxygenation on

isoGDGT cyclicity (Qin et al., 2015; Hurley et al., 2016). While some have argued that this can

lead to a "pessimism phase," where proxy users might no longer have confidence in that proxy's

utility (Elderfield, 2002), in fact these revelations are essential to proxy development – it is the

scientific method at work, and such exhaustive testing of CCFs is a prerequisite for the confident

435   use of a proxy.

The identification and testing of CCFs is inherently an iterative processes. Urey and

others provided serious consideration of CCFs before applying the $\delta^{18}O_{calcite}$ paleotempearture

proxy. It was proposed that the proxy be used only "if the isotopic composition of the water is

known not to differ from the mean of the present seas, or…in the case that it does [differ], if both

440   the isotopic composition of the carbonate and water are determined" (Urey et al., 1951). Urey

described local variability in $\delta^{18}O_{H2O}$ due to evaporation and salinity as "the greatest difficulty"

for accurate temperature measurements, but promised, "this problem is being studied from

several angles and it is hoped that corrections can be applied in the future" (Urey et al., 1951).

Urey's careful consideration of CCFs, and the subsequent and ongoing investigations into those

445 CCFs, serves as an exemplar for proxy discussion, interpretation, and development.

Sometimes, the development of one proxy can constrain a CCF for another proxy by providing a new means of estimating that CCF. The development of the Mg/Ca paleotemperature proxy, based on the incorporation of magnesium relative to calcium in foraminiferal calcite, provided an independent constraint on temperature at the same time (i.e., mid-1990s) that

450 $\delta^{18}O_{calcite}$ was being developed as a paleotemperature proxy (Hastings et al., 1998). By using Mg/Ca to estimate temperature in the same setting as $\delta^{18}O_{calcite}$, researchers were able to independently constrain temperature, and thus use $\delta^{18}O_{calcite}$ to estimate $\delta^{18}O_{H2O}$ (Mashiotta et al., 1999). The development of two independent paleothermometers, each with their own CCFs, provided researchers new opportunities and greater confidence in applying those proxies;

455 $\delta^{18}O_{calcite}$ and Mg/Ca combined helped to identify the degree to which $\delta^{18}O_{H2O}$ influenced the $\delta^{18}O_{calcite}$ proxy, and resulted in a new means to constrain the CCF of $\delta^{18}O_{H2O}$ for future studies. Similarly, multiple studies have compared temperature estimates from TEX_{86} with other organic (e.g., alkenones; Huguet et al., 2006; Lee et all., 2008; Li et al., 2013) and inorganic (e.g., Mg/Ca and $\delta^{18}O_{calcite}$; e.g., Hollis et al., 2012; Hetzberg et al., 2016; O'Brien et al., 2017) proxies in the

460 same settings. While those multi-proxy comparative studies are helping to identify CCFs related to TEX_{86} and other paleotemperature proxies, the numerous unconstrained CCFs related to TEX_{86} make direct testing of CCFs difficult for even those comparative studies. For example, are deviations between $\delta^{18}O_{calcite}$ and TEX_{86} due to depth of production in the water column (e.g., Li et al., 2013; Hetzberg et al, 2016), production season (Huguet et al., 2006), or some other CCF

465 like nutrient availability (Hurley et al., 2016)? Some TEX_{86} applications have used independent proxies to constrain CCFs related to the environment, such as the use of the BIT index (Hopmans et al., 2004) to estimate changes in the input of isoGDGTs from non-marine sources (e.g., Weijers et al., 2006; Hollis et al., 2012). Future work integrating the *physiological* CCFs associated with TEX_{86}, such as changes in water column oxygenation (Qin et al., 2015) and

470 nutrient availability (Hurley et al., 2016) into such multi-proxy comparisons could further constrain the role of different CCFs on TEX_{86} paleotemperature estimates.

Alternatively, the use of statistical methods can elucidate CCFs and their impact on proxy measurements. One example is the Bayesian statistical modelling approach, which uses existing data (usually field-produced calibrations) over a wide range of environments to produce a "best-

475   fit" calibration for the range of values measured in a given study. The resulting model allows
        workers to identify which environments/locations produce a calibration that best fits their data,
        and thus provides a means for workers to investigate environmental conditions, and the related
        CCFs, that more fully express the relationship between, for example, TEX_{86} and temperature
        (Tierney and Tingley, 2014). In fact, the PMIP proxy assessment (Hollis et al., 2019)
480   recommends TEX_{86} users utilize the Bayesian calibration fit as the best current means to estimate
        paleotemperatures (Hollis et al., 2019), demonstrating how the field may use these statistical
        methods to provide best-practices for measurement applications. Similarly, stochastic modelling
        approaches are used in hydrological data interpretations as a means to estimate the partial effects
        (or confounding effects) of different causal factors contributing to a given signal (Yevjevich,
485   1987), and such approaches could be utilized by the paleotemperature community.

        Additionally, the application of transfer functions, including Proxy System Models, are
        used to make inferences about CCFs. Transfer functions provide a theoretical (rather than
        empirical) constraint on a systems' properties in an attempt to predict the quality of properties
        rather than observe them (Telford and Birks, 2005). While the reliability of transfer functions
490   itself is an area of active discussion (e.g., Telford et al., 2004; Telford et al., 2013), transfer
        functions represent yet another statistical approach used to account for CCFs in lieu of empirical
        observations, and are employed by some to reduce uncertainty for correlation- and inference-
        constrained proxies. For example, Proxy System Models use transfer functions to provide an
        assessment of proxy-phenomenon relationships and the driving mechanisms behind proxy
495   measurement outputs (e.g., Dee et al., 2016; Dee et al., 2018; Okazaki and Yoshimura, 2019).
        These statistical methods are an important aid in the determination of CCFs on observational
        signals, and can be powerful in the development of proxy calibrations.

        Ultimately, a mix of variable-controlled laboratory experiments, statistical analyses, and
        field validation experiments all contribute to proxy development. The identification and
500   expression of corollary relationships in a statistical regression is only the first step. Comparisons
        between laboratory (e.g., culture) experiments and field measurements might produce different
        calibrations; causes for differences in the regression should be investigated. For TEX_{86}, the
        recognition of significant variability amongst field calibrations led workers to investigate non-
        temperature properties, such as physiological effects of Thaumarchaeota, in variable-controlled
505   in-laboratory culture experiments (e.g., Elling et al., 2014; Qin et al., 2015; Hurley et al., 2016).

In response, field studies of isoGDGT cyclization were performed in modern and paleo settings (e.g., Hurley et al., 2016; Junium et al., 2018; Polik et al., 2018), and compared with those CCFs identified in culture experiments. These studies together suggest that TEX_{86} users should aim to measure changes in water column oxygenation, ammonia availability, and ecosystem structure,

510   and incorporate those measurements quantitatively into a calibration to develop TEX_{86} as an observation-constrained proxy. Unfortunately, the current limitation (and area of most research) concerns the production of a calibration which accurately reflects all CCFs (Table 1). Many researchers have moved forward with applying TEX_{86} in paleo studies, providing an in-text inference of some CCFs, often concluding that the CCFs do not affect the temperature estimate

515   (e.g., O'Brien et al., 2017), or independently measuring a select number of CCFs (such as changes in the input of isoGDGts using the BIT index; e.g., Weijers et al., 2006). The lack of a unifying calibration that quantitatively accounts for those CCFs implies that these applications exemplify *correlation-constrained proxy* measurements, and the associated reported uncertainty should aim to reflect the accompanying potential uncertainties (Fig. 2b).

520   Because an ideal calibration reflects all contributing pieces of a system (Fig. 1), *a single calibration is necessary for a proxy to be reliably quantitative.* It should be verifiable and applicable in a wide variety of locations, times, and situations. If the calibration is inadequate for some situation, then the calibration does not account for *all* potential CCFs. We consider these calibrations incomplete; for some systems, the unknown CCF does not change, and the

525   calibration explains the corollary relationship, but for other systems, the unknown CCF is introduced or changes, such that the calibration no longer adequately represents the relationship between the measured entity and the property in question. This is the state of current TEX_{86}– each different calibration purports a different quantitative description of the relationship between causal factors (e.g, temperature) and isoGDGT cyclicity (Table 1), and none quantitatively

530   account for CCFs (Table 1; Fig. 2a). Ongoing work to better constrain what CCFs are at play, and how they can be quantified, can move TEX_{86} towards a more observation- or inference-constrained proxy, and lead to more reliable TEX_{86} paleotemperature estimates.

While we use TEX_{86} as an exemplar here, we recognize that limitations in quantitative proxy development and calibration exist across all fields of study, and particularly in the Earth

535   sciences. Not all proxies need be quantitative, and all quantitative proxies present uncertainty. But for a measurement to be most effective (broad applications, less uncertainty), it should be

developed as close to a controlled measurement as possible. This means developing a causal, mechanistic understanding of the relevant system (i.e., a single calibration) as a means to adequately control for the influence of CCFs and produce reliable proxy estimates.

540

4 Conclusions

The distinction between controlled and proxy measurements, and within proxy measurements, serves a more functional role for interpreting, assessing, and developing proxies than previous distinctions between proxy and "direct" measurements. The language proposed here concerning

545   proxy calibrations (e.g., observation- versus inference-constrained proxy) and uncertainty (e.g., analytical versus potential) succinctly and directly addresses the relationship between measurements and the property they intend to describe, and more clearly directs proxy calibration development. Using this language, modelers can more confidently appropriate proxy data outputs into their models, researchers can more efficiently design studies to produce robust

550   measurements, reviewers can more easily assess the reporting of uncertainty and interpretations, and educators can more clearly convey the differences in measurements available for students to learn from, apply, and improve. Readers may find that observational measurements not typically considered proxy measurements in their field may in fact fall on the proxy end of our spectrum. We hope that such realizations might drive workers to investigate what has been taken for

555   granted in previous interpretations, or how future study designs can more accurately assess and account for CCFs. Ultimately, we propose that as much can be learned about a system by developing a proxy as can be learned by applying it.

560

565

Glossary of terms

- **Confounding Causal Factors (CCFs):** Characteristics of an environment that affect the output of a measurement, but are not the property being measured.

- **Controlled measurement:** Measurement that has been manufactured/designed to eliminate the potential effects of all known CCFs on the measurement output.

- **Proxy measurement:** Measurement that does not eliminate the influence of all known CCFs on the intended/targeted property.

- **Observation-constrained proxy:** Proxy measurement where the CCFs are quantitatively incorporated into a calibration, and are accounted for with values produced by other proxy measurement estimates or controlled measurements.

- **Inference-constrained proxy:** Proxy measurement where the CCFs are quantitatively incorporated into a calibration, and are qualitatively accounted for using a reasoned approximation (inference) of the value based on comparisons to similar systems, rather than values produced by measurements of the system in question.

- **Correlation-constrained proxy:** Proxy measurement that does not account for known CCFs, but is based on a hypothesized relationship between a certain property and a measurement output. Uses a calibration that does not quantitatively represent the causal structure of the system.

- **Analytical uncertainty:** The uncertainty associated with the precision and accuracy of the analytical instrument.

- **Potential uncertainty:** The degree to which the measurement/estimated value is affected by something other than the property being measured.

- **Reported uncertainty:** A textual and/or numerical representation of the combined analytical and potential uncertainties associated with a measurement.
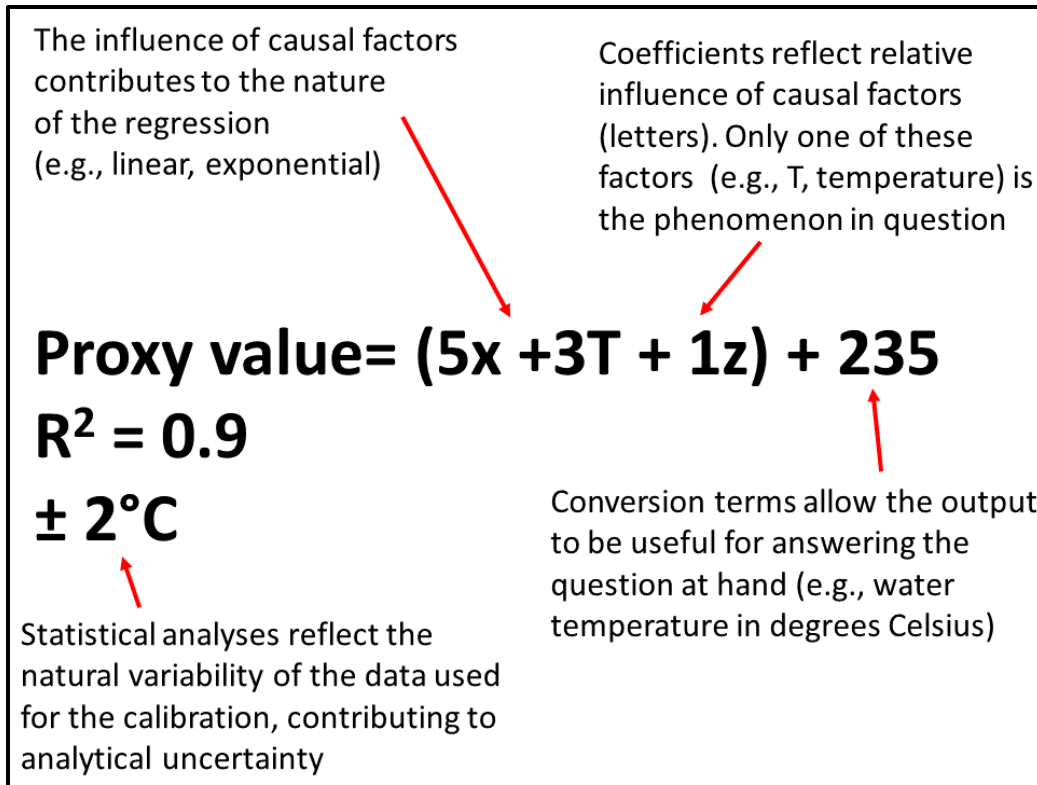
Figure 1: Schematic and description of an idealized calibration for a hypothetical paleotemperature proxy.
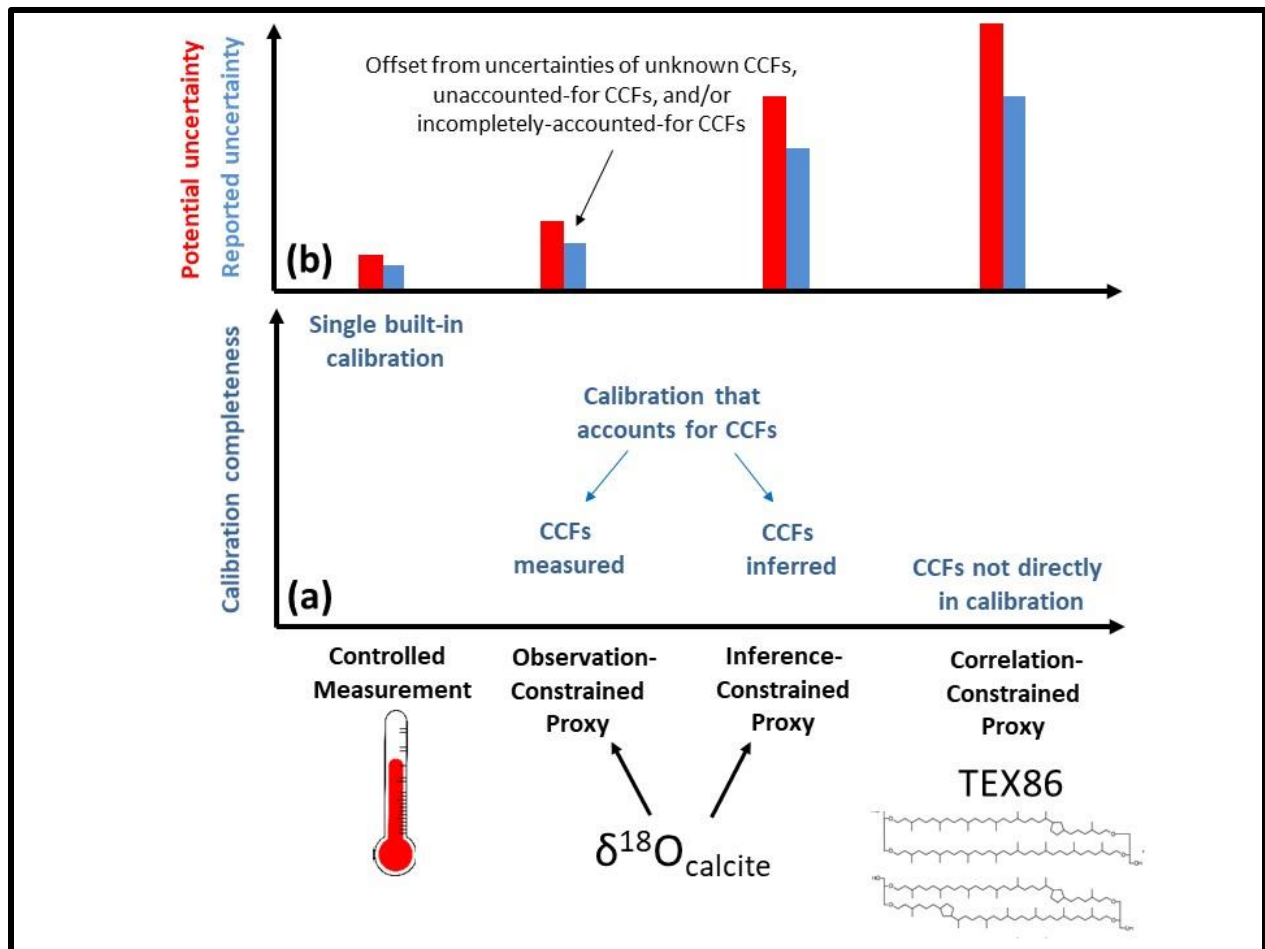
Figure 2: A spectrum (X axis) of observational measurements as function of their incorporation of confounding causal factors and related uncertainty. (a) Bottom Y axis describes the completeness of a measurement's calibrations (i.e., how completely a calibration accounts for all causal factors). Controlled measurements on the left have full control of all causal factors. Observation-constrained proxies have a calibration that quantitatively accounts for CCFs, and allows the researcher to measure those CCFs. Inference-constrained proxies also have a calibration that quantitatively accounts for CCFs, but the researcher cannot measure the CCFs, so the quantitative values for CCFs used in the calibration must be inferred from other evidence. On the right, correlation-constrained proxies have the least direct (quantitative) control of the causal factors, with calibrations that do not quantitatively account for CCFs. (b) Top Y axis represents uncertainty of each measurement, with the red line signifying potential uncertainty and the blue bar showing range of reported uncertainty in literature. Because analytical uncertainty varies greatly between proxies, instruments, and users, we have excluded its representation. Arrow and description of offset applies to all measurements.

Range (°C)	Equation	Reference
0-30	$T = (TEX_{86} - 0.27) / 0.015$	Schouten et al. (2002)
22-30	$T = (TEX_{86} - 0.016) / 0.027$	Schouten et al.(2003)
10-28	$T = (TEX_{86}' - 0.2) / 0.016$	Slujis et al. (2006)
5-30	$T = -10.78 + 56.2 \times TEX_{86}$	Kim et al. (2008)
25-28	$T = (TEX_{86} + 0.09) / 0.035$	Trommer et al. (2009)
-3-30	$T = 50.475 - 16.332 \times (1/TEX_{86})$	Liu et al. (2009)
-3-30	$T = 81.5 \times TEX_{86} - 26.6$	Kim et al. (2010)
-3-30	$T = -19.1 \times (1/TEX_{86}) + 54.5$	Kim et al. (2010)
-3-30	$T = 49.9 + 67.5 \times$ (GDGT index-1)	Kim et al. (2010)
5-30	$T = 38.6 + 68.4 \times$ (GDGT index-2)	Kim et al. (2010)
10-40	$T = 48.2 \times TEX_{86} + 1.04$	Kim et al. (2010)
10-40	$T = -9 \times (1/TEX_{86}) + 45.2$	Kim et al. (2010)
10-40	$T = 42.9 \times$ (GDGT index-1) + 46.5	Kim et al. (2010)
10-40	$T = 52 \times$ (GDGT index-2) + 42	Kim et al. (2010)
4-30	$T = -14 + 55.2 \times TEX_{86}$	Powers et al. (2010)
10-30	$T = 3.5 + 38.9 \times TEX_{86}$	Tierney et al (2010)
-2-30	$T = (TEX_{86} - 0.3038) / 0.0125$	Shevenell et al. (2011)
14-34	$T = 32.873 \times \ln(\text{GDGT index-1}) + 50.771$	Hollis et al. (2012)
14-34	$T = 39.036 \times \ln(TEX_{86}) + 36.455$	Hollis et al. (2012)
15-35	$T = (TEX_{86} - 0.21) / 0.015$	Qin et al. (2015)
10-30	$TEX_{86} = -0.0006T^2 + 0.023T + 0.33$	Qin et al. (2015)
10-25	$TEX_{86} = -0.0017T^2 + 0.054T + 0.11$	Qin et al. (2015)
2-10	$T = 27.898(TEX_{86}^L) + 22.723$	Harning et al. (2019)

Name	Calculations	Reference
TEX_{86}	[GDGT-2]+[GDGT-3]+[Cren']/[GDGT-1]+[GDGT-2]+[GDGT-3]+[Cren']	Schouten et al. (2002)
TEX_{86}'	[GDGT-2]+[GDGT-3]+[Cren']/[GDGT-1]+[GDGT-2]+[Cren']	Slujis et al. (2006)
TEX_{86}^L	$-\log($[GDGT-2]/[GDGT-1]+[GDGT-2]+[GDGT-3]$)$	Kim et al. (2010)
TEX_{86}^H	$0.99 \times TEX_{86}^L + 0.12$	Kim et al. (2010)
GDGT index-1	$\log($[GDGT-2]/[GDGT-1]+[GDGT-2]+[GDGT-3]$)$	Kim et al. (2010)
GDGT index-2	$\log(TEX_{86})$	Kim et al. (2010)

Table 1: Compilation of TEX_{86} calculations and calibrations as of 2020. Modified from Tierney (2012).

620

Author contribution

JW and FGB designed the study. FGB wrote the manuscript, and JW and FGB edited the manuscript.

625 **Data Availability**

All data described here is presented in previously published literature and is cited as such.

Competing interests

The authors declare that they have no conflict of interest.

630

Acknowledgements

We thank the Center for the Study of Origins for their support. We thank T. Marchitto, G. Miller, B. Johnson, M. Huber, and F.D. Boudinot for their helpful comments that improved the manuscript, and H. Spero, J. Sepúlveda, and C. Cleland for their discussions that improved the

635 manuscript. FGB acknowledges the Department of Geological Sciences at the University of Colorado Boulder, NSF Division of Earth Sciences Earth-Life Transitions (ELT) program grant #1338318, and the American Chemical Society Petroleum Research Fund (ACS-PRF) - Doctoral New Investigator Award #58815-DNI2 for their support. JW acknowledges the Department of Philosophy and the Graduate School at the University of Colorado Boulder for their support.

640

References

Ahm, A.C., Bjerrun, C.J., Blattler, C.L., Swart, P.K., and Higgins, J.A.: Quantifying early

marine diagenesis in shallow-water carbonate sediments, Geochimica et Cosmochimica

Acta, 236, 140-159, https://doi.org/10.1016/j.gca.2018.02.042, 2018.

645 Bemis, B.E., Spero, H.J., Bijma, J., and Lea, D.W.: Reevaluation of the oxygen isotopic

composition of planktonic foraminifera: Experimental results and revised

paleotemperature equations, Paleoceanography, 13, 150-160,

https://doi.org/10.1029/98PA00070, 1998.

Broecker, W.S. and van Donk, J.: Insolation changes, ice volumes, and the O^{18} record in Deep-

650   Sea Cores, Reviews of Geophysics and Space Physics, 8, 169-198,

https://doi.org/10.1029/RG008i001p00169, 1970.

Chang, H: Inventing temperature: Measurement and scientific progress. Oxford University Press,

Oxford, United Kingdom, 2004.

Dee, S.G., Russell, J.M., Morrill, C., Chen, Z., and Neary, A.: PRYSM V2.0: A proxy system

655   model for lacustrine archives, Paleoceanography and Paleoclimatology, 33, 1250-1269,

https://doi.org/10.1029/2018PA003413, 2018.

Dee, S.G., Steiger, N.J., Emile-Geay, J., Hakim, G.J.: On the utility of proxy system models for

estimating climate states over the common era, Journal of Advances in Modeling Earth

Systems, 8, 1164-1179, https://doi.org/10.1002/2016MS000677, 2016.

660   Elderfield, H.: Foraminiferal Mg/Ca paleothermometry: expected advances and unexpected

consequences, Goldschmidt Conference, Davos, Switzerland, August 17-23,

Geochmimica et Cosmochimica Acta, 66, A213, 2002.

Emiliani, C.: Pleistocene temperatures, The Journal of Geology, 63, 538-578,

https://doi.org/10.1086/626295, 1955.

665   German, C.R., and Elderfield, H.: Application of the Ce Anomaly as a paleoredox indicator: the

ground rules, Paleoceanography, 5, 823-833, https://doi.org/10.1029/PA005i005p00823,

1990.

Franklin, A.: Experiment, right or wrong. Cambridge University Press, Cambridge, United

Kingdom, 1990.

670   Harning, D., Andrews, J.T., Belt, S.T., Babedo-Sanz, P., Geirsdottir, A., Dildar, N., Miller, G.H.,

and Sepúlveda, J.: Sea ice control on winter subsurface temperatures of the north

Iceland shelf during the Little Ice Age: a TEX$_{86}$ calibration case study,

Paleoceanography and Paleoclimatology, 34, 1-16,

https://doi.org/10.1029/2018PA003523, 2019.

675   Hetzberg, J.E., Schmidt, M.W., Bianchi, T.S., Smith, R.W., Shields, M.R., and Marcantonio, F.:

Comparison of eastern tropical Pacific TEX$_{86}$ and *Globigerinoides ruber* Mg/Ca derived

sea surface temperature: Insights from the Holocene and Last Glacial Maximum, Earth

and Planetary Science Letters, 434, 320-332, https://doi.org/10.1016/j.epsl.2015.11.050,

2016.

680   Hollis, C.J., Taylor, K.W.R., Handley, L., Pancost, R.D., Huber, M., Creech, J.B., Hines, B.R.,

Crouch, E.M., Morgans, H.E.G., Crampon, J.S., Gibbs, S., Pearson, P.N., and Zachos,

J.C.: Early Paleogene temperature history of the Southwest Pacific Ocean:

Reconciling proxies and models, Earth and Planetary Science Letters, 349-350, 53-

66, https://doi.org/10.1016/j.epsl.2012.06.024, 2012.

685   Hollis, C.J., Jones, T.D., Anagnostou, E., Bijl, P.K., Cramwinckel, M.J., Cui, Y., Dickens, G.R.,

Edgar, K.M., Eley, Y., Evans, D., Foster, G.L., Frieling, J., Inglis, G.N., Kennedy, E.M.,

Kozdon, R., Lauretano, V., Lear, C.H., Littler, K., Lourens, L., Meckler, A.N., Naafs,

B.D.A., Pälike, H., Pancost, R.D., Pearson, P.N., Röhl, U., Royer, D.L., Salzmann, U.,

Schubert, B.A., Seebeck, H., Sluijs, A., Speijer, R.P., Stassen, P., Tierney, J., Tripati, A.,

690   Wade, B., Westerhold, T., Witkowski, C., Zachos, J.C., Zhang, Y.G., Huber, M., and

Lunt, D.J.: The DeepMIP contribution to PMIP4: methodologies for selection,

compilation and analysis of latest Paleocene and early Eocene climate proxy data,

incorporating version 0.1 of the DeepMIP database, Geoscientific Model Development,

12, 3149-3206, https://doi.org/10.5194/gmd-12-3149-2019, 2019.

695   Huguet, C., Kim, J-H., Sinninghe Damsté, J.S., and Schouten, S.: Reconstruction of sea surface temperature variations in the Arabian Sea over the last 23 kyr using organic proxies (TEX_{86} and U^K_{37}'), Paleoceanography, 21, PA3003, https://doi.org/10.1029/2005PA001215, 2006.

Hurley, S.J., Elling, F.J., Konneke, M., Buchwald, C., Wankel, S.D., Santoro, A.E., Lipp, J.S.,
700   Hinrichs, K-U, and Pearson, A.: Influence of ammonia oxidation rate on thaumarchaeal lipid composition and the TEX_{86} temperature proxy, Proceedings of the National Academy of Sciences, 113, 28, 7762-7767, https://doi.org/10.1073/pnas.1518534113, 2016.

Jansen, E., J. Overpeck, K.R. Briffa, J.-C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B.
705   Otto-Bliesner, W.R. Peltier, S. Rahmstorf, R. Ramesh, D. Raynaud, D. Rind, O. Solomina, R. Villalba and D. Zhang: Palaeoclimate. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)].
710   Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Junium, C.K., Meyers, S.R., and Arthur, M.A.: Nitrogen cycle dynamics in the Late Cretaceous greenhouse, Earth and Planetary Science Letters, 481, 404-411, https://doi.org/10.1016/j.epsl.2017.10.006, 2018.

715   Kim, J-H, Schouten, S., Hopmans, E.C., Donner, B., and Sinninghe Damsté, J.S.: Global sediment core-top calibration of the TEX_{86} paleothermometer in the ocean, Geochimica et Cosmochimica Acta, 72, 1154-1173, https://doi.org/10.1016/j.gca.2007.12.010, 2008.

Kim, J-H, van der Meer, J., Schouten, S., Helmke, P., Willmott, V., Sangiorgi, F., Koc, N., Hopmans, E.C., and Sinninghe Damsté, J.S.: New indices and calibrations derived from the distribution of crenarchaeal isoprenoid tetraether lipids: Implications for past sea surface temperature reconstructions, Geochimica et Cosmochimica Acta, 74, 4639-4654, https://doi.org/10.1016/j.gca.2010.05.027, 2010.

Lee, K.E., Kim, J-H., Wilke, I., Helmke, P., and Schouten, S., A study of the alkenone, TEX_{86}, and planktonic foraminifera in the Benguela Upwelling System: Implications for past sea surface temperature estimates, Geochemistry Geophysics Geosystems, 9, https://doi.org/10.1029/2008GC002056, 2008.

Li, D., Zhao, M., Tian, J., and Li., L.: Comparison and implication of TEX_{86} and U^K_{37}' temperature records over the last 356 kyr of ODP Site 1147 from the northern South China Sea, Palaeogeograpy, Palaeoclimatology, Palaeoecology, 376, 213-223, DOI: 10.1016/j.palaeo.2013.02.031, 2013.

Liu, Z., Pagani, M., Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Shah, S.R., Leckie, R.M., and Pearson, A.: Global cooling during the Eocene-Oligocene climate transition, Science, 233, 1187-1190, DOI: 10.1126/science.1166368, 2009.

Mashiotta, T.A., Lea, D.W., Spero, H.J.: Glacial-interglacial changes in Subantarctic sea surface temperature and δ^{18}O-water using foraminiferal Mg, Earth and Planetary Science Letters, 170, 417-432, https://doi.org/10.1016/S0012-821X(99)00116-8, 1999.

Newman, D.K., Neubauer, C., Ricci, J.N., Wu, C-H., and Pearson, A.: Cellular and molecular biological approaches to interpreting ancient biomarkers, Annual Reviews of Earth and Planetary Sciences, 44, 493-522, https://doi.org/10.1146/annurev-earth-050212-123958, 2016.

NOAA National Centers for Environmental Information: "What are proxy data?",

https://www.ncdc.noaa.gov/news/what-are-proxy-data, last access: 1 February 2020.

Powers, L., Werne, J.P., Vanderwoude, A.J., Sinninghe Damsté, J.S., Hopmans, E.C., and

Schouten, S.: Applicability and calibration of the TEX_{86} paleothermometer in lakes,

745   Organic Geochemistry, 41, 404-413, https://doi.org/10.1016/j.orggeochem.2009.11.009,

2010.

O'Brien, C.L., Robinson, S.A., Pancost, R.D., Sinninghe Damsté, J.S., Schouten, S., Lunt, D.J.,

Alsenz, H., Bornemann, A., Bottini, C., Brassell, S.C., Farnsworth, A., Forster, A.,

Huber, B.T., Inglis, G.N., Jenkyns, H.C., Linnert, C., Littler, K., Markwick, P.,

750   McAnena, A., Mutterlose, J., Naafs, B.D.A., Püttmann, W., Sluijs, A., van Helmond,

N.A.G.M., Vellekoop, J., Wagner, T., and Wrobel, N.E.: Cretaceous sea-surface

temperature evolution: Constraints from TEX_{86} and planktonic foraminiferal oxygen

isotopes, Earth-Science Reviews, 172, 224-247,

https://doi.org/10.1016/j.earscirev.2017.07.012, 2017.

755   Okazaki, A. and Yoshimura, K.: Global evaluation of proxy system models for stable water

isotopes with realistic atmospheric forcing, JGR Atmospheres, 124, 8972-8993,

https://doi.org/10.1029/2018JD029463, 2019.

Qin, W., Carlson, L.T., Armbrust, E.V., Devol, A.H., Moffett, J.W., Stahl, D.A., and Ingalls,

A.E.: Confounding effects of oxygen and temperature on the TEX_{86} signature of marine

760   Thaumarchaeota, Proceedings of the National Academy of Sciences, 112, 10979-10984,

DOI: 10.1073/pnas.1501568112, 2015.

Ravelo, A.C., and Hillaire-Marcel, C.: The use of oxygen and carbon isotopes of foraminifera in paleoceanography, in Developments in Marine Geology, v. 1, 735-764, https://doi.org/10.1016/S1572-5480(07)01023-8, 2007.

765   Royer, D.L.: Stomatal density and stomatal index as indicators of paleoatmospheric CO_2 concentrations. Review of Palaeobotany and Palynology, 114, 1-28, https://doi.org/10.1016/S0034-6667(00)00074-9, 2001.

Ruddiman, W.F.: Earth's climate: Past and Future, second edition, W.H. Freeman and Company, New York, USA, 2008.

770   Schouten, S., Hopmans, E.C., Pancost, R.D., and Sinninghe Damsté, J.S.: Widespread occurrence of structurally diverse tetraether membrane lipids: Evidence for the ubiquitous presence of low-tempearture relatives of hyperthermophiles, Proceedings of the National Academy of Sciences, USA, 97, 14421-14426.

Schouten, S., Hopmans, E.C., Schefuβ, E., and Sinninghe Damsté, J.S.: Distributional variations

775   in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures?, Earth and Planetary Science Letters, 204, 265-274, https://doi.org/10.1016/S0012-821X(02)00979-2, 2002.

Schouten, S., Hopmans, E.C., Forster, A., van Breugel, Y., Kuypers, M.M.M., and Sinninghe Damsté, J.S.: Extremely high sea-surface temperatures at low latitudes during the

780   middle Cretaceous as revealed by archaeal membrane lipids, Geology, 31, 1069- 1072, https://doi.org/10.1130/G19876.1, 2003.

Shevenell, A.E., Ingalls, A.E., Domack, E.W., and Kelly, C.: Holocene Southern Ocean surface temperature variability west of the Antarctic Peninsula, Nature, 470, 250-254, https://doi.org/10.1038/nature09751, 2011.

785   Slujis, A., Schouten, S., Pagani, M., Woltering, M., Brinkhuis, H., Sinninghe Damsté, J.S.,

Dickens, G.R., Huber, M., Reichart, G-J, Stein, R., Matthiessen, J., Lourens, L.J.,

Pedentchouk, N., Backman, J., Moran, K., & the Expedition 302 Scientists:

Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum,

Nature, 441, 610-613, https://doi.org/10.1038/nature04668, 2006.

790   Spero, H.J., Bijma, J., Lea, D.W., and Bemis, B.E.: Effect of seawater carbonate concentration

on foraminiferal carbon and oxygen isotopes, Nature, 390, 497-500,

https://doi.org/10.1038/37333, 1997.

Suppes, P. A set of independent axioms for extensive quantities, Portugaliae Mathematica, 10,

163-172, 10.1007/978-94-017-3173-7_3, 1951

795   Telford, R.J., and Birks, H.J.B.: The secret assumption of transfer functions: problems with

spatial autocorrelation in evaluating model performance, Quaternary Science Reviews,

24, 2173-2179, 10.1016/j.quascirev.2005.05.001, 2005.

Telford, R.J., Andersson, R., Birks, H.J.B., and Juggins, S.: Biases in the estimation of transfer

function prediction errors, Paleoceanography, 19, PA4014,

800   https://doi.org/10.1029/2004PA001072, 2004.

Telford, R.J., Li, C., Kucera, M.: Mismatch between the depth habitat of planktonic foraminifera

and the calibration depth of SST transfer functions may bias reconstructions, Climate of

the Past, 9, 859-870, https://doi.org/10.5194/cp-9-859-2013, 2013.

Tierney, J.E.: GDGT thermometry: Lipid tools for reconstructing paleotemperatures, in

805   Reconstructing Earth's Deep-Time Climate – The State of the Art in 2012,

Paaleontological Society Short Course, The Paleontological Society Papers, 18, Linda

V. Ivany and Brian T. Huber (eds.), 115-131,

https://doi.org/10.1017/S1089332600002588, 2012.

Tierney, J.E., Mayes, M.T., Meyer, N., Johnson, C., Swarzenski, P.W., Cohen, A.S., and Russell,

810   J.M.: Late-twentieth-century warming in Lake Tanganyika unprecedented since AD 500,

Nature Geoscience, 3, 422-425, https://doi.org/10.1038/ngeo865, 2010.

Tierney, J.E., and Tingley, M.P.: A Bayesian, spatially-varying calibration model for the TEX_{86}

proxy, Geochmimica et Cosmochimica Acta, 127, 83-106,

https://doi.org/10.1016/j.gca.2013.11.026, 2014.

815   Trommer, G., Siccha, M., van der Meer, M.T.J., Schouten, S., Sinninghe Damsté, J.A., Schulz,

H., Hemleben, C., and Kucera, M.: Distribution of Crenarchaeota tetraether

membrane lipids in surface sediments from the Red Sea: Organic Geochemistry, 40,

724-731, https://doi.org/10.1016/j.orggeochem.2009.03.001, 2009.

Urey, H.C.: Oxygen isotopes in nature and in the laboratory, Science, 108, 2810, 489-496, DOI:

820   10.1126/science.108.2810.489, 1948.

Urey, H.C., Lowenstam, H.A., Epstein, S., and McKinney, C.R.: Measurement of

paleotempeatures and temperatures of the upper Cretaceous of England, Denmark, and

the southeastern United States, Bulletin of the Geological Society of America, 62, 399-

416, https://doi.org/10.1130/0016-7606(1951)62[399:MOPATO]2.0.CO;2, 1951.

825   Van Fraassen, B.C. Scientific representation: Paradoxes of perspective, Analysis, 70, 511-514,

https://doi.org/10.1093/analys/anq042, 2010.

Wilf, P., Wing, S.L., Greenwood, D.R., and Greenwood, C.L.: Using fossil leaves as

paleoprecipitaiton indicators; an Eocene example. Geology, 26, 203-206,

https://doi.org/10.1130/0091-7613(1998)026<0203:UFLAPI>2.3.CO;2, 1998.

830   Wilson, J., and Boudinot, F.G., Proxy measurement in paleoclimatology and the TEX_{86}

paleothermometer (*in review).*

Wilson, J., and Boudinot, F.G., The reliability of proxy measurements and the proxy/non-proxy

distinction, American Geophysical Union Fall Meeting, abstract ID PP14B-07, San

Francisco, California, December 9-13 2019.

835   Woelders, L., Vellekoop, J., Kroon, D., Smit, J., Casadio, S., Pramparo, M.B., Dinares-Turell, J.,

Peterse, F., Slujis, A., Lenaerts, J.T.M., and Speijer, R.P.: Latest Cretaceous

climatic and environmental change in the South Atlantic region, Paleoceanography, 32,

466-483, https://doi.org/10.1002/2016PA003007, 2017.

Yasuhara, M., Rabalais, N.N., Conley, D.J., Gutierrez, D.: Palaeo-records of histories of

840   deoxygenation and its ecosystem impact, in Ocean deoxygenation: Everyone's problem –

Causes, impacts, consequences, and solutions, Laffoley, D., Baxter, J.M. (eds.). IUCN,

Gland, Switzerland, 213-224, http://dx.doi.org/10.1007/s00338-019-01765-0, 2019.

Yevjevich, V.: Stochastic models in hydrology, Stochastic Hydrology and hydraulics, 1, 17-36,

https://doi.org/10.1007/BF01543907, 1987.

845   Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K.: Trends, rhythms, and

aberrations in global climate 65 Ma to present, Science, 292, 686-693, DOI:

10.1126/science.1059412, 2001.