

# Paleoclimate Proxies and the Benefits of Disunity

Aja Watkins<sup>1\*</sup>

<sup>1</sup>University of Wisconsin-Madison

\*Corresponding author: Aja Watkins; [aja.watkins@wisc.edu](mailto:aja.watkins@wisc.edu)

## Abstract

Measuring the climates of the deep past requires use of paleoclimate proxies. I describe two proxy data and measurement practices, regarding proxy calibration and proxy data infrastructure. I document how at least some data and measurement practices in paleoclimatology are disunified: these practices do not involve intercalibration or otherwise statistical combination of multiple proxy records, and metadata necessary for proxy data to be reused or intercompared is often not provided. I argue that, perhaps counter-intuitively, this lack of standardization and unification of proxy data and measurements has several benefits, especially related to the management of error and uncertainty.

## 1. Introduction

Reconstructing the climates of Earth's deep past is epistemically challenging, yet of the utmost importance in a contemporary scientific context where climate scientists are increasingly turning to paleoclimatology to help understand the Earth's climate and apply that understanding to our current climate crisis.<sup>1</sup> Paleoclimatologists rely on "paleoclimate proxies" to measure past climatic variables such as temperature. Paleoclimate proxies include ice cores, tree rings, sediment cores, coral growth rings, fossilized pollen, fossilized leaves, and more. Their use involves measuring some attribute of modern-day traces of the past and reconstructing the most likely environmental conditions under which that trace was formed. For example, the elemental composition of the calcareous shells of marine microorganisms depends on the temperatures at which the relevant chemical reactions involved occurred. Knowledge of this relationship allows paleoclimatologists to reconstruct past temperatures, within some degree of uncertainty.

In this paper, I provide an analysis of some of the key data and measurement practices scientists deploy in relation to paleoclimate proxies. I argue that the development and use of paleoclimate proxies to reconstruct past climates demonstrate some of the benefits of disunity or lack of standardization in data and measurement practices, especially practices which indicate that paleoclimatologists are not looking to produce one, definitive record of Earth's past climate but are satisfied with several (possibly conflicting)

---

<sup>1</sup>For philosophical discussion of these so-called "paleoclimate analogues," see Watkins (2023); Wilson (2023).

records. The claim that disunified data and measurement practices are beneficial contrasts with the intuitive view that data are necessarily more useful if they are reusable or interoperable, or suggestions that measurement procedures need to be standardized to be successful. Importantly, I do *not* claim that paleoclimatologists are intentionally deploying disunified practices as a strategy, nor that they are even aware of the benefits of those practices, nor that the benefits of these practices necessarily outweigh the risks of disunity (which I think remains to be seen in this case). The claim made in this paper is just that the practices paleoclimatologists currently utilize do, in fact, produce some surprising benefits, specifically related to error and uncertainty management.

My argument builds on recent work in the philosophy of the Earth sciences and medical sciences which has demonstrated that similar intuitions concerning the benefits of coherence and coordination, including in measurement and data practices, do not always hold. In particular, Teru Miyake (2011, 2017a,b) has analyzed various scientific debates in the history of the geosciences, especially geophysics and seismology, and has shown how conflicting models or theories about the unobservable processes beneath the Earth's surface ultimately helped these sciences to progress. Alisa Bokulich (2020) has also argued in the context of measurements of geologic time that disagreement between different radiometric dating methods has been an important source of understanding about sources of error in these different measurements. Miguel Ohnesorge (2021, 2022) has argued that disagreeing measurements of the ellipticity of the Earth by nineteenth-century geodesists can nonetheless be seen as epistemically successful. Likewise, I demonstrate in the context of paleoclimate proxies that there are reasons that count in favor of data and measurement practices scientists implement that do *not* produce a single, unified record of Earth's past climates – even if the scientists themselves are unaware of these reasons or are not using these reasons to guide their practice. Finally, in the history and philosophy of medicine, Rebecca Jackson (2021) has argued that in the case of “drop” measurements in anesthesiology, reasons in favor of non-standardized measurement units (e.g., patient experience and safety) outweighed reasons in favor of standardized units (e.g., consistency and ease of communication). Although, unlike Jackson, I will not claim to weigh the reasons in favor of or against standardization – I am dealing with a live, contemporary case rather than a historical one, so I don't have the benefits of hindsight that Jackson has – I will, like her, call attention to some surprising benefits of disunity and lack of standardization.

To do so, I focus on two paleoclimatological practices through which proxy measurements or data are kept disunified. The first relates to the calibration of paleoclimate proxy measurements (section 2). Here I show that paleoclimatologists tend not to *intercalibrate* their proxy measurements, nor otherwise statistically combine them, and I discuss several reasons why doing so might be beneficial, including preserving the independence of these multiple lines of evidence and preventing the production of unquantifiable or unidentifiable sources of error and uncertainty.

Second, I address data infrastructure, including especially how paleoclimate proxy data are stored and norms or requirements surrounding metadata (section 3). Both data storage and metadata practices are remarkably nonstandardized, with authority over these practices spread very thinly across multiple individuals and organizations. Rather than see these consequences of the relevant data infrastructure decisions as a

failure, I instead argue that a structure which subsequently and at least temporarily minimizes data travel and reuse actually has the important benefit of preventing inadvertent compounding of sources of error and uncertainty.

In conclusion, I reiterate that disunity in both practices is helpful in mitigating or managing error and uncertainty (section 4). In a non-ideal epistemic situation such as that found in nearly any study of the deep past, dealing with error and uncertainty appropriately is paramount. Furthermore, in a science that potentially has massive implications for action – as, in this historical moment, all of the Earth and environmental sciences do – scientists are prone to be cautious and take seriously considerations of inductive risk (Oreskes, 2015). Other scientific areas that are similarly non-ideal and risky may also benefit from managing error and uncertainty by preserving disunity, rather than assuming that standardization, coherence, and statistical integration is necessarily a superior epistemic or pragmatic strategy.

## 2. Disunity in Proxy Calibration

In this section, I argue that there are benefits to disunified measurement practices in paleoclimatology, namely, the decision not to intercalibrate different proxies with one another. First, I use existing philosophical literature on calibration to explain how paleoclimate proxies are calibrated to the instrumental record. I then suggest some potentially beneficial consequences of not *intercalibrating* or otherwise statistically combining proxy records, having to do with error and uncertainty management and maintaining multiple, independent lines of evidence.

In order to use paleoclimate proxies to reconstruct past climates, these measurements need to be appropriately calibrated; in other words, a model of the measurement process needs to be developed and refined. In general, calibration of any measurement involves settling on a way to convert a measurement *indication* into a measurement *outcome*. A measurement indication is the reading of a measurement instrument after a measurement procedure has been performed, whereas a measurement outcome is a value actually attributed to the relevant measurand. Often, measurement indications and measurement outcomes are not the same; some conversion process may be required to translate from one to the other.<sup>2</sup> Sometimes, this conversion process is embedded within a measurement instrument itself. In any case, determining how to convert between a measurement indication and a measurement outcome is referred to as calibration. Calibration of paleoclimate proxies involves connecting the attributes of present-day traces of the past with the relevant characteristics of the past climate under which those traces were formed.

In Hasok Chang (2004)'s influential discussion of the “problem of nomic measurement” – how do we know that our measurement tools are adequately representing the quantities being measured when those measurement tools are the only way of accessing the very same quantities? – Chang introduces the idea of “metrological extension.”

---

<sup>2</sup>In the terminology of Parker (2017), the measures where the measurement indication needs to be converted into another quantity are called “derived,” whereas those that don’t require conversion (but might require corrections, etc.) are called “direct.” Paleoclimate proxies are often referred to as “indirect” measures of Earth’s climate (e.g., Masson-Delmotte et al. 2013; NOAA 2018). However, this terminology is confusing in the context of philosophy of measurement, where it is broadly understood that few measurements are as “direct” as they may seem, and are instead extensively model- and theory-laden. I will therefore avoid use of the term “direct” herein.

Metrological extension involves applying a measurement technique beyond the context in which it was developed (in his case, to higher or lower temperatures). Metrological extension gives us one way of understanding proxy calibration; in the case of paleoclimatology, the measurements are being extended to further time periods. In other words, paleoclimatologists have a way to ground their measurements of temperatures of the past that was not available to early thermometrists attempting to develop temperature measurements of the present for the first time: proxy-based measures of temperature can be compared to instrumental (thermometer-based) temperature records as a means of initial calibration. For example, attributes of tree rings or organismal remains or layers of ice can be initially correlated with thermometer-based temperature measurements in a laboratory or field work setting. Once proxy-based temperature measurements are adequately calibrated to instrumental readings, the proxy-based measurements can be extended further back in time than the instrumental readings go.

As proxies are used to extend measures of the climate back in time, additional sources of error and uncertainty are introduced and need to be accounted for in order to isolate the climatic signal. For example, changes in abundance of different types of pollen or the chemical composition of microorganisms in sediment layers might be signals of changes in climate, but, over long enough time scales, might instead be signals of something else, such as changes in range or adaptation (both of which *can* be indicative of climatic change but also of other ecological changes). Extrapolating climate measurements back in time requires an understanding of the ways in which many processes may affect the proxy-based signal.

The practice of calibrating paleoclimate proxies to the instrumental record accords with Eran Tal (2017)'s influential account of calibration. The instrumental measurements (which might come from laboratory settings, the historical record, or in situ measurements) are taken as "fixed," and serve epistemically as a sort of measurement standard to which the other, proxy-based measurements must be compared. Then, a proxy-based record that temporally coincides with the instrumental record is compared to the instrumental record, in order to develop a correlation between the proxy-based measurement and the instrumental measurement. This correlation is used to develop a calibration function that takes as inputs features of the proxy (e.g., stable isotope ratios in the sediment or ice core layer) and outputs climatic features (e.g., temperature).

Other than these explicit comparisons between proxy-based measurements and instrumental measurements, paleoclimatologists also use what are called "proxy system models" (PSMs) in order to constrain the relationship between the proxy-based measurement and a measurement of the climate (e.g., Dolman and Laepple 2018; Lawman et al. 2020). PSMs are also known as "forward models": they take as inputs features of the object to be measured, i.e., the climate, and output the expected measurement indication, i.e., features of the proxy system, that would obtain in those circumstances. In the context of paleoclimatology, PSMs are simulations of how the trace left of the past climate will be affected over time before it is collected and analyzed by us. These simulations are extremely important for helping paleoclimatologists understand and vicariously control for conflating factors that might influence the proxy signal.<sup>3</sup>

---

<sup>3</sup>Regarding the distinction between physical and vicarious control, see Norton and Suppe (2001); in the context of proxies, see Wilson and Boudinot (2022).

PSMs are a crucial part of proxy calibration. Recall that I've argued that proxy-based measurements of the past climate should be seen as a case of metrological extension of climate measurements into new contexts – further back in time. Extending climate measurements further back in time introduces new sources of error and uncertainty. PSMs allow researchers to simulate the effect of these processes, and consequently to account for them in the calibration of the proxy measurement. Of course, our understanding of these processes and how they have changed over time is itself imperfect and subject to revision, but the important thing is that the use of a PSM allows for development of a calibration function that does not mindlessly extrapolate back in time a calibration function developed on present-day correlations between measurement indications and measurement outcomes. Developing a calibration function that is not time-invariant is more felicitous for the research context.

Let's take stock. So far, I have argued that the development of paleoclimate proxies is best analyzed as a case of metrological extension of climate measurements into new temporal contexts. Successful metrological extension in this case requires calibrating paleoclimate proxies to instrumental data and extrapolating the relevant calibration function back in time. However, extrapolation back in time is nontrivial. Metrological extension into new temporal domains thus highlights how different sources of error and uncertainty in proxy-based measurement contexts might change over time, since the relationships between the measurement indication and the measurement outcome are not time-invariant. Paleoclimatologists have thus come to rely on PSMs, a type of forward model that allows for confounding factors to be simulated explicitly and for the ultimate proxy calibration to take into account processes other than climatic ones which may change the proxy-based signal over time. The result is a growing list of paleoclimate proxy measurement procedures, each calibrated to varying levels of accuracy and precision and all used for different periods in Earth's history.

The thesis of this paper, though, is disunity or lack of standardization within various paleoclimatological data and measurement practices has some beneficial consequences (whether or not paleoclimatologists themselves are aware of these benefits, and whether or not these beneficial effects are outweighed by detrimental ones). I suggest that the benefits of disunity are evidenced by a calibration strategy that paleoclimatologists do *not* deploy: intercalibration between proxies.

By and large, different proxies are not intercalibrated. For example, in principle one could constrain tree-ring based climate reconstructions using sediment-based climate reconstructions, or vice versa (one could do this for any pair of paleoclimate proxies that overlap in temporal and geographic range). This strategy has the potential to yield a more precise calibration function for any given proxy; the more times the proxy is intercalibrated with other proxies, the more tightly constrained its calibration function becomes, as it would need to cohere with all of the other proxy records. However, paleoclimatologists generally do not use intercalibration to constrain their proxy calibration functions; proxies are calibrated to the instrumental record, and that's it. I suggest that there are two possible benefits of this practice (again, benefits of which paleoclimatologists may not be aware).

First, intercalibrating paleoclimate proxy measurements would cause them to become *dependent* on one another, in the sense that a change to one measurement process would result in a corresponding change to the other measurement process. If two measurement procedures (or other sources of evidence) are dependent on one

another, it is not surprising if and when the evidence they produce agrees in its support of a claim. For example, two, *dependent* paleoclimate proxies would be expected to agree in support of any given claim about climate trends in particular times and places. Philosophers of the historical sciences have often argued that maintaining independent lines of evidence is a useful strategy for historical scientists, exactly because agreement between multiple, independent lines of evidence is *prima facie* surprising (e.g., Wylie 1989, 2002, 2011; Cleland 2011, 2013; Forber and Griffith 2011; Currie 2018; Bokulich 2020).<sup>4</sup> Likewise, Martin Vezér (2015, 2017) says it is unlikely that multiple different proxy-based records of past climates would agree about claims if those claims were false; so, he argues, when multiple proxy-based climate reconstructions agree on a particular claim, that claim is more likely to be true than if the claim were only supported by one proxy-based reconstruction.<sup>5</sup>

These kinds of arguments that use multiple, independent lines of evidence to support a claim are called *consilience* arguments, and philosophers of the historical sciences broadly agree that lines of evidence need to be independent (in the right sort of way, whatever that is) to be used in such arguments. More contentious is the idea that multiple, independent lines of evidence can also be used to support *robustness* arguments. Rather than using multiple lines of evidence to support a claim, robustness arguments involve using multiple lines of evidence to demonstrate whether a claim is (in)sensitive to various background assumptions. According to Vezér (2015), convergence of multiple paleoclimate proxies shows that the claims different proxies make about the past are insensitive to the details of a particular proxy-based measurement process; this is a robustness argument. However, two philosophical debates about robustness arguments complicate analysis of the paleoclimate case. First, there is debate about whether using multiple, independent lines of evidence really does result in the sensitivity test that robustness arguments purport to provide (e.g., Levins 1966; Orzack and Sober 1993; Staley 2004).<sup>6</sup> Second, there is debate about whether independence of multiple lines of evidence really is necessary to make robustness arguments (for discussion and argument that multiple lines of evidence don't need to be independent in a strict, formal sense, see Schupbach 2018).

Regrettably, this leaves us with limited or unsatisfactory answers to several philosophical questions about independent lines of evidence. These include: What kind of independence, exactly, is required to use multiple lines of evidence in consilience or robustness arguments? Do consilience and robustness arguments really require *independent* lines of evidence, at all? And are consilience or robustness arguments even possible, perhaps given some of the difficulties answering the other two questions? Answering all of these is outside of the scope of this paper. However, I am able to say

<sup>4</sup>Wylie (2011) calls the relevant sense of independence “horizontal” independence, which applies to lines of evidence such that “when used in conjunction with one another, it is implausible that images produced by such different means should converge as a consequence of confounding influences that generate competing error in each distinct line of evidence” (p. 387). Horizontal independence thus enables what Mayo (2018) calls a *severe test* of a claim.

<sup>5</sup>Vezér says that multi-proxy compilations combine relevantly independent sources of evidence when these sources of evidence “derive from different studies that use a wide range of methods and data sets” (2017, p. 5) or when “inferences drawn from one source can be maintained without relying on the same set of assumptions required to draw an inference from another source” (2015, p. 55).

<sup>6</sup>This debate has been very active in philosophy of climate science; see, e.g., Knutti et al. (2008); Lloyd (2009, 2010, 2015); Parker (2010, 2018); Winsberg (2018); Sherwood et al. (2020); O’Loughlin (2021).

that *if* consilience and/or robustness arguments are possible and require a certain kind of independence (e.g., the kind that would make it unlikely for different lines of evidence to agree if what they agreed upon wasn't true), *then* the current practice of not intercalibrating paleoclimate proxies enables consilience and/or robustness arguments, because not intercalibrating preserves the different measurements' independence. This conditional at least indicates a *possible* benefit to not intercalibrating different paleoclimate proxies.

A second benefit of not intercalibrating multiple paleoclimate proxies, I argue, is that doing so allows paleoclimatologists to mitigate the inadvertent compounding of various sources of error and uncertainty. As described above, proxy-based reconstructions of past climates have to contend with many sources of error and uncertainty, a problem which is exacerbated as these records get extended further back in time. The main way of estimating the extent of the error is to use PSMs or other model-based strategies. However, when these techniques aren't available, or when they are underdeveloped, paleoclimatologists have few means of estimating the magnitude of error and uncertainty, and, especially, of teasing out different sources of error and uncertainty (e.g., Carré et al. 2012 distinguish between calibration error, sampling error, and variability-induced error). More philosophical work needs to be done on error and uncertainty in proxy-based measurement contexts. For our purposes, suffice it to say that the difficulties paleoclimatologists have in quantifying their error and uncertainty for each proxy would be compounded if these proxies were intercalibrated. I therefore think that, by refusing to intercalibrate, paleoclimatologists are able to prevent further aggravation of the effects of the uncertainties inherent in their measurement processes.

One legitimate counterexample to this trend of not intercalibrating – and an example that gives a little more insight into paleoclimatologists' explicit understanding of the risks of intercalibration – is found in van Dam and Utescher (2016).<sup>7</sup> van Dam and Utescher compare plant- and mammal-based reconstructions of precipitation in present-day Europe during the Neogene (about 23–20 million years ago). As the authors note, proxies for precipitation have been some of the least successful, especially when compared with temperature proxies. van Dam and Utescher carefully selected fossil sites and samples in order to perform their comparison of the two proxies for paleoprecipitation, then suggesting the possibility of intercalibrating the two in order to more tightly constrain the calibration curve for each. Importantly, intercalibration is only suggested because there are no clear modern analogue species for some of the mammals used, making calibration with the instrumental record especially difficult. This counterexample indicates not only that the practices of paleoclimatologists are varied and difficult to make sweeping generalizations about, but also the conditions which have to be met before paleoclimatologists consider intercalibration a plausible way to go. In this case, intercalibration is presented as a last resort; constraining the paleoprecipitation record by other means has proven unsuccessful, so paleoclimatologists are willing to give intercalibration a try. In other words, this study is evidence for paleoclimatologists' hesitancy about intercalibrating multiple paleoclimate proxies.

In addition to not intercalibrating paleoclimate proxies, paleoclimatologists also tend not to combine reconstructions based on multiple proxies. There has long been recognition of the benefits of so-called “multiproxy” studies. These benefits include the fact

<sup>7</sup>In fact, this is the only counterexample I have found.

that multiple proxies in combination can cover a broader geographic range. For example, Michael Mann (2002) suggests that tree rings, corals, and ice cores are best used in combination, in part because these different proxies cover different geographic regions (terrestrial temperate zone, marine tropics, and polar regions, respectively). In addition to serving as complementary records in a geographic sense, different proxies may have different strengths and weaknesses, which can balance each other out.

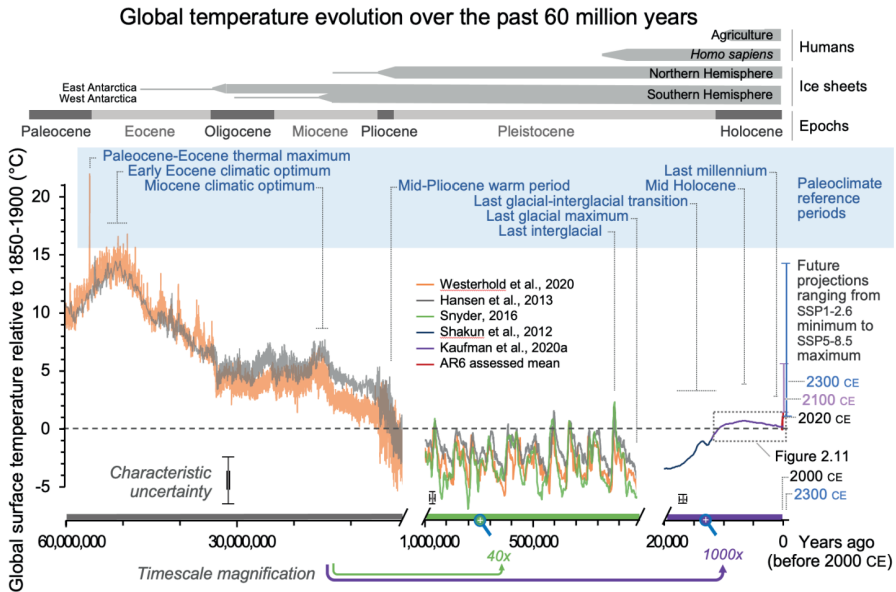
However, despite this acknowledgement that multiproxy methods are superior to any method which relies too heavily on a single proxy, there is no consensus on *how* to best combine information from different proxy records. One obvious approach would be to average multiple reconstructions of the same climatic variable, e.g., temperature. However, doing so risks illicitly combining substantially different measurands as though they were one; for example, polar and tropical temperatures cannot be very meaningfully averaged. Relatedly, averages can be easily skewed by uneven sampling, a problem which plagues paleoclimatology as well as historical sciences and climate science more broadly (e.g., Raja et al. 2022; Brönnimann and Wintzer 2019). As a result of these well-founded hesitations to average different paleoclimate reconstructions together, “the interpretations of multi-proxy datasets [often] rely on visually matching several proxy records” (Schroeter et al., 2020, p. 1). In other words, multiple proxy-based reconstructions are simply placed on one graph of, for instance, temperature over time, and agreement or disagreement between these proxies is just indicated by visual agreement between the resulting lines.

Take, for example, the 60 million year global surface temperature reconstruction reproduced in the second chapter of the most recent IPCC report (Gulev et al., 2021). Specifically, for the period ranging from 60 million to 1 million years ago, the IPCC used two proxy-based temperature reconstructions, one from Hansen et al. (2013) and the other from Westerhold et al. (2020); the data in Hansen et al. (2013) is likewise based on data presented in Zachos et al. (2008). Zachos et al. (2008) and Westerhold et al. (2020) both purport to give a global picture of temperature, using data from several ocean sediment cores collected during successor drilling programs: the Deep Sea Drilling Project (1968-1983), the Ocean Drilling Program (1985-2004), and the Integrated Ocean Drilling Program (2004-present). These studies both use benthic foraminifera tests (shells), calculating a proxy called  $\delta^{18}\text{O}$  (which looks at the ratio of heavy to light oxygen isotopes) for these and converting that to temperature (correcting for different foram genera). In Hansen et al. (2013) and Westerhold et al. (2020), these “raw” data were then smoothed over different temporal resolutions. Both Hansen et al. (2013) and Westerhold et al. (2020) have performed statistical processes which integrate different data sources in order to reconstruct temperatures since the Paleocene. However, more interesting for our purposes is the fact that then, when the IPCC decided to use these two studies, they decided to portray them as two *different* reconstructions of the past. The two temperature profiles are plotted on the same graph, in different colors, and the viewer can see that they broadly agree on Earth’s climate history (see figure 1). Yet – even though these two studies used cores obtained by the same drilling programs and used the same proxy for temperature ( $\delta^{18}\text{O}$ ), the two records were kept separate.<sup>8</sup>

---

<sup>8</sup>The climate reconstruction of the last 2,000 years reproduced in the fifth assessment report similarly kept reconstructions from different past publications separate; see Masson-Delmotte et al. (2013, p. 409).





**Figure 1:** Global temperature reconstruction reproduced in the IPCC’s sixth assessment report (Chapter 2 of the Working Group 1 report; Gulev et al. 2021). Notice that different temperature reconstructions from past publications are kept separate. This is even the case for the 1-60 million years ago reconstruction, which is based on two studies that use very similar methods. Reprinted with permission from the Secretary of the IPCC.

Why? I speculate that the benefits of keeping multiple proxy records visually separate are analogous to the benefits of not intercalibrating different proxies offered above. First, keeping separate proxy-based reconstructions separate preserves, at least to some extent, their independence. Independent lines of evidence might be important for consilience and robustness arguments, i.e., arguments which are intended to enhance our confidence in or security of particular hypotheses. Visually distinguishing multiple proxy-based reconstructions makes it clearer to the viewer that they are intended to play these roles. Second, refusal to statistically combine different reconstructions – or, due caution about when to do so – helps paleoclimatologists to manage error and uncertainty, by preventing them from unintentionally compounding these by combining proxies with different sources and magnitudes of error and uncertainty.

In summary, I have shown that paleoclimatologists at least have a tendency to both (a) not intercalibrate multiple proxies with one another and also (b) not otherwise statistically combine multiple proxy-based records, such as by averaging these reconstructions together. I do not wish to suggest that these practices are *never* pursued, nor that they never should be. But, I have attempted to explain that there are benefits of keeping different proxies and the reconstructions based on them separate, rather than integrating them with one another – even if the paleoclimatologists themselves are not aware of these benefits, and even if, all-things-considered, these benefits are outweighed by various risks. Keeping the separate proxies disunified can help to cope with various sources of error and uncertainty, an important task for a measurement technique that is so nascent.

Additionally, preserving the independence of these multiple lines of evidence may be crucial if paleoclimatologists want to use them in consilience or robustness arguments. Preserving the disunity of proxies in these cases can, then, have important benefits.

### 3. Disunity in Data Infrastructure

Using paleoclimate proxies requires an elaborate data infrastructure, a term I use to encompass practices pertaining to data collection, storage, and travel. I argue that there are important components of the paleoclimate proxy data infrastructure that are disunified, in ways that have the surprising potential to benefit paleoclimatological practice. I focus on how data storage practices are disunified, especially how the use of metadata is not standardized, even within particular archives or databases. I argue that this lack of standardization (predictably) prevents data reuse and repurposing, but that, perhaps counter-intuitively, this is actually a feature and not a bug. I will focus on the Paleoclimatology Database, operated by the National Oceanic and Atmospheric Administration (NOAA) through the National Centers for Environmental Information (NCEI).

The Paleoclimatology Database involves proxy data that can be filtered based on proxy type (e.g., ice core, pollen, coral, tree ring), investigator (i.e., who published the dataset), location, time period, and more. Individual datasets matching the filtering criteria can then be downloaded and analyzed. The Paleoclimatology Database has been immensely useful for proxy-based paleoclimate research in recent years. For example, one relatively high-profile project has been the Past Global Changes (PAGES) 2k Consortium, a collaborative effort to compile multiple proxy records from the Common Era. The PAGES2k Consortium (2017) has utilized the NCEI database to make their records usable by other researchers.

In order to contribute to the Paleoclimatology Database, contributors must follow a template (they then submit the template-compliant dataset via email, after which it is reviewed and posted online).<sup>9</sup> The requirements provided by the template are as follows: First, researchers must use terminology defined explicitly in the Paleoenvironmental Standard Terms (PaST) Thesaurus.<sup>10</sup> The PaST Thesaurus standardizes the terms used for different variables, such as “material” (the material on which the measurements are made) or “data type” (the category in which the dataset will be housed; this variable has only a set list of possible options, e.g., Borehole, Ice Cores, Pollen). Second, there is a required set of metadata that must be provided with the dataset. These include: details about the publication (if any) associated with the dataset, the date of the contribution, funding information, the latitude and longitude coordinates of the study location, the dates over which the analysis was performed, the length of the core (if relevant), the species involved in biogenic proxy measurements (if relevant), a section on “chronology” (i.e., the dates obtained from the material studied, and, perhaps, information about how those dates were obtained – although this is *not* required), and labels for the columns of the dataset itself (using the PaST Thesaurus).

---

<sup>9</sup>The template can be accessed here: <https://www.ncei.noaa.gov/pub/data/paleo/templates/noaa-wds-paleo-template.txt>.

<sup>10</sup>The thesaurus is accessible here: <https://www.ncei.noaa.gov/products/paleoclimatology/paleoenvironmental-standard-terms-thesaurus>.

Ostensibly, the fact that the Paleoclimatology Database requires use of a template with metadata requirements at all may be seen to indicate paleoclimatologists' attempt at standardization of digital data storage. However, there are some key pieces of metadata that are *not* required by the Paleoclimatology Database template, and these indicate that paleoclimatologists may be reluctant to require that their data be interoperable and reusable in particular ways. Most importantly, the database template does not require metadata about which calibration function was used and where that calibration function came from. To be fair, some datasets *do* include this information; for example, data that convert the ratio between magnesium and calcium (Mg/Ca) in samples of *Globigerinoides ruber* (a species of planktic foraminifera) to sea surface temperature (SST) reconstructions almost always are explicit that they used the calibration curve established in Anand et al. (2003) to perform the calibration. However, this is not true of many proxy-based reconstructions, especially those for which there is less of a consensus about which calibration function to use. For example, the Paleoclimatology Database has 17 datasets that use benthic foraminifera to produce climate reconstruction data (usually, bottom water temperature [BWT]) based on Mg/Ca. Of these 17, only six datasets listed in their metadata the calibration function they used to convert Mg/Ca into temperature (some of these list the equation itself, others just refer the user to another study to find the calibration curve).<sup>11</sup> Likewise, of 23 datasets that use the ratio of strontium to calcium (Sr/Ca) in corals as a proxy for SST, only eight include calibration-related metadata.<sup>12</sup> Examples of metadata with and without calibration information are reproduced in figure 2.

Metadata about the calibration function used is necessary for these proxy data to be interoperable (i.e., compared to one another or otherwise integrated) or reusable in new research contexts. In order to be interoperable, for instance, reconstructions that are based on the same material should probably be calibrated using the same calibration function (or there should be some justification for using a different function; for instance, maybe some calibration functions are better suited to certain locations over others). If data are not appropriately calibrated for intercomparison, or if it is not possible to tell whether they are or not, then it is not clear what hypotheses about the agreement or disagreement of data from different datasets are appropriate. For example, if identical elemental ratio data (e.g., Mg/Ca or Sr/Ca) are calibrated to temperature differently, then we wouldn't expect the temperature reconstructions to agree, whereas if we are testing for consistency or robustness between different samples we would expect the reconstructions to agree. Likewise, reuse of data is hindered by lack of calibration metadata. For example, calibration metadata can be used to "reverse engineer" the original proxy-based measurements if these are otherwise unavailable, which could then be re-calibrated if new and improved calibration functions are developed in the future.

At first blush, then, the lack of standardization about calibration-related metadata appears to be a failure on the part of the database organizers, or at least individual contributors. However, I argue that actually the decision not to include or require this metadata might actually have some (perhaps unintended) benefits, exactly

<sup>11</sup>These numbers are based on a search done on July 18, 2022.

<sup>12</sup>These numbers are based on searches done in February and March, 2021.

```
# Variables
# Data variables follow that are preceded by "##" in columns one and two.
# Variables list, one per line, shortname-tab-longname components (9 components:
what, material, error, units, seasonality, archive, detail, method, C or N for
Character or Numeric data)
## age_AD age,,,years AD,,,,,N
## d18Oporite_delta 18O,Porites sp.,,per mil VPDB,,Corals and
Sclerosponges,,analytical technique: isotope ratio mass spectrometry,N
## Sr/Ca-por Sr/Ca,Porites sp.,,mmol/mol,,Corals and Sclerosponges,,analytical
technique: ICP-OES,N
## SrCa-SST-por sea surface temperature,Porites sp. Sr/Ca ,,degrees Celsius,,Corals
and Sclerosponges,,analytical technique: ICP-OES,N
## Dd18O-sw delta 18O seawater,Porites sp.,,per mil VPDB,,Corals and
Sclerosponges,,analytical technique: inductively coupled plasma atomic emission
spectroscopy,N
```

(a) Variables metadata without calibration information.

```
# Variables
# Data variables follow that are preceded by "##" in columns one and two.
# Variables list, one per line, shortname-tab-longname components (9 components:
what, material, error, units, seasonality, archive, detail, method, C or N for
Character or Numeric data)
## age_AD age,,,year Common Era,,Corals and Sclerosponges,,,N
## Sr/Ca strontium/calcium,Porites lutea,,millimole per mole,Monthly,Corals and
Sclerosponges,,inductively-coupled plasma atomic emission spectroscopy,N,Age model
(name: Lawman_2019_06-SB-A1) determined using a MATLAB algorithm that identified
peaks/troughs in the geochemical data and interpolated to monthly resolution
## sst sea surface temperature,strontium/calcium,,degree Celsius,Monthly,Climate
Reconstructions; Corals and Sclerosponges,,regression analysis,N,Calibration
equation: SST = -20.73 x (Sr/Ca) + 210.53
```

(b) Variables metadata with calibration information (highlighted).

**Figure 2:** Example screenshots of datasets from the Paleoclimatology Database where the metadata includes or doesn't include information about the calibration function used. Figure 2a is from the dataset associated with Carilli et al. (2014); 2b is from the dataset associated with Lawman et al. (2020). Notice that Lawman et al. (2020) had to add another column of information about their variables in order to include the calibration information, above and beyond what is required by the database template.

*because* it prevents interoperability and reuse. In particular, I propose that preventing interoperability and reuse contributes beneficially to management of error and uncertainty.

Recall that, in the context of proxy calibration, I argued that not intercalibrating multiple proxies with one another prevents compounding of sources of error and uncertainty in ways that are difficult to detect or quantify. Appropriate interoperability and reuse of data also requires a level of understanding of sources of error and uncertainty that is presently not available for most proxy-based records of Earth's past climates. Data reuse, for instance, requires by definition that data can be exported from one research context to another. Doing so involves some attempt to show that data that were adequate for research purposes in the first context are also adequate for research purposes in the second context (Bokulich and Parker, 2021). And, in many cases, whether data are adequate or not for a given research purpose will depend on their associated degrees of error and uncertainty; if estimates of these are unavailable (or themselves highly uncertain), it is often difficult to tell whether data can satisfactorily travel between research contexts. Likewise, interoperability requires data to be able to be integrated with other data. Absent reliable indications of error and uncertainty, it is difficult to know whether and when two datasets can be appropriately integrated. Furthermore, integration or inter-comparison of different datasets may, like intercalibration, compound the relevant errors and uncertainties. Compounding sources of error and uncertainty is ill-advised unless researchers have systematic ways of doing so, which paleoclimatologists (as of yet) do

not. Of course, there are also risks associated with lack of standardization, and further work needs to be done to weigh the risks and benefits in this case.

These considerations indicate that many proxy-based datasets are not ready for reuse or interoperability, but that this isn't necessarily a bad thing, and instead contributes beneficially to paleoclimatologists' ability to manage error and uncertainty. Proscriptions of interoperability and reuse violate the so-called "FAIR" data principles; FAIR stands for "findability," "accessibility," "interoperability," and "reusability." Proponents of the FAIR data principles make claims such as that "FAIRness is a prerequisite for proper data management and data stewardship" (Wilkinson et al., 2016, p. 6). Cases, such as, I suggest, in the context of paleoclimate proxy data, where there are benefits to preventing interoperability and reusability of data put pressure on these claims in the data ethics and data management literature. Further philosophical work is needed in the nascent field of data ethics to better articulate the scope of standards such as the FAIR principles, or to replace them altogether.<sup>13</sup>

Some other features of the Paleoclimatology Database also indicate that paleoclimate data are not interoperable and reusable on a broad scale. First, the user interface of the database does not make it possible to perform any comparison of multiple datasets at once. For instance, one might want to generate a single graph that tracks SST over time based on Sr/Ca in corals. This functionality is not possible within the database itself, so a user wanting to generate this kind of graph would have to do so by manually downloading all of the relevant datasets and compiling them together. (This process itself might highlight to the user the lack of information about calibration or other ways in which data from different datasets are dissimilar, such as their temporal resolution; this is roughly what happened to me!) By contrast, another widely-used database of historical data, The Paleobiology Database (<https://paleobiodb.org/>), has greater functionality along these lines, integrating all of the data into one interface, as well as "benchmarking" tools to prevent misuse of the data. Second, the Paleoclimatology Database does not purport whatsoever to have data that are "cleaned," for example by removing duplicates or checking for typographical errors, nor does it provide much transparency on how or whether individual contributors have cleaned their own data. This, again, is left up to individual researchers to figure out. Overall, the fact that the Paleoclimatology Database's structure makes these practices of reuse and interoperability so labor-intensive disincentivizes researchers from pursuing these projects (whether or not this was the intention of the database managers). And, as I have argued in the context of intercalibration (in section 2) and calibration metadata, the fact that lack of standardization in this case prevents reuse and interoperability actually has beneficial consequences, namely for error and uncertainty management.

In summary, I have suggested that disunity or lack of standardization in paleoclimatologists' data infrastructure, especially paleoclimatologists' decisions with respect to database design, has some surprising benefits. In particular, I have argued that the lack of required metadata about proxy calibration – as well as, perhaps, the lack of an integrated user interface or transparency about dataset "cleaning" practices – serves to

<sup>13</sup>For example, one important criticism of the FAIR data principles has come from Indigenous data scholarship, in which researchers have noted that the FAIR principles neglect to condemn some of the most atrocious unethical data-related practices perpetrated by colonizers and settlers on Indigenous peoples. One suggestion is to supplement the FAIR principles with the CARE principles; CARE stands for "collective benefit," "authority to control," "responsibility," and "ethics" (e.g., Carroll et al. 2020.)

prevent multiple proxy datasets from being reused or intercompared. Contrary to claims such as those by proponents of the FAIR data principles that interoperability and reuse are *necessarily* positive qualities of datasets, I suggest that data infrastructures which may discourage or at least not require data to be interoperable or reusable may actually help researchers mitigate rather than compound the various sources of error and uncertainty that affect their data. Similar to my argument concerning the lack of proxy intercalibration offered in section 2, then, I propose that disunification in paleoclimatologists' data practices can be beneficial, even in ways of which paleoclimatologists may be unaware and, admittedly, in ways that might, in this case, be outweighed by the risks of non-standardization. Both philosophers and scientists should think more about strategic database design, including required metadata, in terms of what effect this design will have on scientific practice. For example, perhaps it is the case that even if the metadata that would be needed to permit reuse and interoperability shouldn't be required, other metadata (concerning sources of error and uncertainty) could be required in order to help facilitate eventual reuse of these datasets.

#### 4. Conclusion

I have argued that there are data and measurement practices that paleoclimatologists perform that demonstrate the benefits of disunity or lack of standardization. First, I explained how paleoclimate proxies are calibrated to the instrumental record, and tend not to be intercalibrated with one another. Not intercalibrating multiple proxies preserves the independence of the climate reconstructions based on these different measurements, and also prevents inadvertent compounding of multiple sources of error and uncertainty that affect different proxies differently. Second, I showed how even a shared Paleoclimatology Database does not enable reuse or interoperability of proxy data, because there are key pieces of metadata (e.g., about calibration methods) that are not required by the database. Contrary to claims that suggest detailed metadata is required for the data to be useful, I argued that at least temporarily preventing reuse or integration of datasets might actually have the beneficial consequence of mitigating error and uncertainty.

Both of these ways in which proxy data and measurements are disunified or non-standardized have in common that they enable paleoclimatologists to handle myriad sources of error and uncertainty. Paleoclimatology is currently still enormously error- and uncertainty-ridden, and so error and uncertainty management is at the forefront of researchers' concerns. Future work in philosophy of paleoclimatology should investigate whether the benefits described herein outweigh or are outweighed by the risks associated with these same practices. Additionally, future philosophical work on other areas of science where error and uncertainty are central should investigate whether or how these areas might preserve or construct disunity in their data and measurement practices.

#### Acknowledgments

This paper has benefited greatly due to feedback from Alisa Bokulich, Wendy Parker, Michaela McSweeney, Rachell Powell, Wally Fulweiler, Greg Lusk, Meghan Page,

Miguel Ohnesorge, Federica Bocchi, Leticia Castillo Brache, and two anonymous referees, as well as audiences at the Cambridge Early-Career Workshop in the Philosophy of Measurement in 2023, The Future of the Past: Philosophical Issues in the Historical Sciences at the Hebrew University of Jerusalem in 2022, and Measurement at the Crossroads in 2022. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1840990. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

## 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). [10.1029/2002PA000846](https://doi.org/10.1029/2002PA000846).
- Bokulich, A. (2020) Calibration, Coherence, and Consilience in Radiometric Measures of Geologic Time. *Philosophy of Science*. 87(3), 425–456. [10.1086/708690](https://doi.org/10.1086/708690).
- Bokulich, A. and Parker, W. (2021) Data models, representation, and adequacy-for-purpose. *European Journal for Philosophy of Science*. 11(31). [10.1007/s13194-020-00345-2](https://doi.org/10.1007/s13194-020-00345-2).
- Brönnimann, S. and Wintzer, J. (2019) Climate data empathy. *WIREs Climate Change*. 10(2), e559. [10.1002/wcc.559](https://doi.org/10.1002/wcc.559).
- Carilli, J. E., McGregor, H. V., Gaudry, J. J., Donner, S. D., Gagan, M. K., Stevenson, S., Wong, H. and Fink, D. (2014) Equatorial Pacific coral geochemical records show recent weakening of the Walker Circulation: Coral shows weakening Walker Circulation. *Paleoceanography*. 29(11), 1031–1045. [10.1002/2014PA002683](https://doi.org/10.1002/2014PA002683).
- Carré, M., Sachs, J. P., Wallace, J. M. and Favier, C. (2012) Exploring errors in paleoclimate proxy reconstructions using Monte Carlo simulations: paleotemperature from mollusk and coral geochemistry. *Climate of the Past*. 8(2), 433–450. [10.5194/cp-8-433-2012](https://doi.org/10.5194/cp-8-433-2012).
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J. and Hudson, M. (2020) The CARE Principles for Indigenous Data Governance. *Data Science Journal*. 19, 43. [10.5334/dsj-2020-043](https://doi.org/10.5334/dsj-2020-043).
- Chang, H. (2004) *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press, USA.
- Cleland, C. E. (2011) Prediction and Explanation in Historical Natural Science. *The British Journal for the Philosophy of Science*. 62(3), 551–582. [10.1093/bjps/axq024](https://doi.org/10.1093/bjps/axq024).
- Cleland, C. E. (2013) Common cause explanation and the search for a smoking gun. *The Geological Society of America*. [10.1130/2013.2502\(01\)](https://doi.org/10.1130/2013.2502(01)).
- Currie, A. (2018) *Rock, bone, and ruin: an optimist's guide to the historical sciences*. Life and mind: philosophical issues in biology and psychology. The MIT Press. Cambridge, Massachusetts.
- Dolman, A. M. and Laepple, T. (2018) Sedproxy: a forward model for sediment-archived climate proxies. *Climate of the Past*. 14(12), 1851–1868. [10.5194/cp-14-1851-2018](https://doi.org/10.5194/cp-14-1851-2018).
- Forber, P. and Griffith, E. (2011) Historical Reconstruction: Gaining Epistemic Access to the Deep Past. *Philosophy and Theory in Biology*. 3(e203). [10.3998/ptb.6959004.0003.003](https://doi.org/10.3998/ptb.6959004.0003.003).
- Gulev, S., Thorne, P. W., Ahn, J., Dentener, F., Domingues, C., Gerland, S., Gong, D., Kaufman, D., Namchi, H., Quaas, J., Rivera, J., Sathyendranath, S., Smith, S., Trewin, B., von Schuckmann, K. and Vose, R. (2021) Chapter 2: Changing State of the Climate System In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (eds). Cambridge University Press. Cambridge, UK and New York, NY, USA. pp. 287–422.

- Hansen, J., Sato, M., Russell, G. and Kharecha, P. (2013) Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 371(2013), 20120294. [10.1098/rsta.2012.0294](https://doi.org/10.1098/rsta.2012.0294).
- Jackson, R. L. (2021) “The Uncertain Method of Drops”: How a Non-Uniform Unit Survived the Century of Standardization. *Perspectives on Science*. 29(6), 802–841. [10.1162/posc.a.00395](https://doi.org/10.1162/posc.a.00395).
- Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G. A., Meinshausen, M., Murphy, J. M., Plattner, G.-K., Raper, S. C. B., Stocker, T. F., Stott, P. A., Teng, H. and Wigley, T. M. L. (2008) A Review of Uncertainties in Global Temperature Projections over the Twenty-First Century. *Journal of Climate*. 21(11), 2651–2663. [10.1175/2007JCLI2119.1](https://doi.org/10.1175/2007JCLI2119.1).
- Lawman, A. E., Partin, J. W., Dee, S. G., Casadio, C. A., Di Nezio, P. and Quinn, T. M. (2020) Developing a Coral Proxy System Model to Compare Coral and Climate Model Estimates of Changes in Paleo-ENSO Variability. *Paleoceanography and Paleoclimatology*. 35(7). [10.1029/2019PA003836](https://doi.org/10.1029/2019PA003836).
- Levins, R. (1966) The strategy of model building in population biology. *American Scientist*. 54(4), 421–431.
- Lloyd, E. A. (2009) Varieties of Support and Confirmation of Climate Models. *Aristotelian Society Supplementary Volume*. 83(1), 213–232. [10.1111/j.1467-8349.2009.00179.x](https://doi.org/10.1111/j.1467-8349.2009.00179.x).
- Lloyd, E. A. (2010) Confirmation and Robustness of Climate Models. *Philosophy of Science*. 77(5), 971–984. [10.1086/657427](https://doi.org/10.1086/657427).
- Lloyd, E. A. (2015) Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science Part A*. 49, 58–68. [10.1016/j.shpsa.2014.12.002](https://doi.org/10.1016/j.shpsa.2014.12.002).
- Mann, M. E. (2002) The Value of Multiple Proxies. *Science*. 297(5586), 1481–1482. [10.1126/science.1074318](https://doi.org/10.1126/science.1074318).
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Fidel, J., Rouco, G., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Ramesh, R., Rojas, M., Shao, X., Anchukaitis, K., Arblaster, J., Bartlein, P. J., Benito, G., Clark, P., Comiso, J. C., Crowley, T., Deckker, P. D., de Vernal, A., Delmonte, B., DiNezio, P., Dowsett, H. J., Edwards, R. L., Fischer, H., Fleitmann, D., Foster, G., Fröhlich, C., Hall, A., Hargreaves, J., Haywood, A., Hollis, C., Krinner, G., Landais, A., Li, C., Lunt, D., Mahowald, N., McGregor, S., Meehl, G., Mitrovica, J. X., Moberg, A., Mudelsee, M., Muhs, D. R., Mulitza, S., Müller, S., Overland, J., Parrenin, F., Pearson, P., Robock, A., Rohling, E., Salzmann, U., Savarino, J., Sedláček, J., Shindell, D., Smerdon, J., Solomina, O., Tarasov, P., Vinther, B., Waelbroeck, C., Wolf, D., Yokoyama, Y., Yoshimori, M., Zachos, J., Zwartz, D., Gupta, A. K., Rahimzadeh, F., Raynaud, D. and Wanner, H. (2013) Information from Paleoclimate Archives. *Intergovernmental Panel on Climate Change. Assessment Report 5*, 383–464.
- Mayo, D. G. (2018) *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press.
- Miyake, T. (2011) *Underdetermination and Indirect Measurement*. Ph.D. Stanford University. United States – California.
- Miyake, T. (2017a) Magnitude, moment, and measurement: The seismic mechanism controversy and its resolution. *Studies in History and Philosophy of Science Part A*. 65-66, 112–120. [10.1016/j.shpsa.2017.02.002](https://doi.org/10.1016/j.shpsa.2017.02.002).
- Miyake, T. (2017b) Uncertainty and Modeling in Seismology. In *Reasoning in Measurement*. Routledge.
- National Oceanic and Atmospheric Administration (NOAA) Paleoclimatology | National Centers for Environmental Information (NCEI).
- NOAA (2018) What Are Proxy Data?
- Norton, S. and Suppe, F. (2001) Why Atmospheric Modeling Is Good Science In *Changing the Atmosphere*, Miller, C. A. and Edwards, P. N. (eds). The MIT Press. Cambridge, MA. pp. 88–133. [10.7551/mitpress/1789.003.0006](https://doi.org/10.7551/mitpress/1789.003.0006).
- Ohnesorge, M. (2021) How incoherent measurement succeeds: Coordination and success in the measurement of the earth’s polar flattening. *Studies in History and Philosophy of Science*. 88, 245–262. [10.1016/j.shpsa.2021.06.006](https://doi.org/10.1016/j.shpsa.2021.06.006).
- Ohnesorge, M. (2022) Pluralizing measurement: Physical geodesy’s measurement problem and its resolution. *Studies in History and Philosophy of Science*. 96, 51–67. [10.1016/j.shpsa.2022.08.011](https://doi.org/10.1016/j.shpsa.2022.08.011).
- O’Loughlin, R. (2021) Robustness reasoning in climate model comparisons. *Studies in History and Philosophy of Science Part A*. 85, 34–43. [10.1016/j.shpsa.2020.12.005](https://doi.org/10.1016/j.shpsa.2020.12.005).
- Oreskes, N. (2015) How Earth Science Has Become a Social Science. *Historical Social Research*. 40(2), 246–270. [10.12759/HSR.40.2015.2.246-270](https://doi.org/10.12759/HSR.40.2015.2.246-270).



- Orzack, S. H. and Sober, E. (1993) A Critical Assessment of Levins's The Strategy of Model Building in Population Biology (1966). *The Quarterly Review of Biology*. 68(4), 533–546. [10.1086/418301](https://doi.org/10.1086/418301).
- PAGES2k Consortium (2017) A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*. 4, 170088. [10.1038/sdata.2017.88](https://doi.org/10.1038/sdata.2017.88).
- Parker, W. S. (2010) Whose Probabilities? Predicting Climate Change with Ensembles of Models. *Philosophy of Science*. 77(5), 985–997. [10.1086/656815](https://doi.org/10.1086/656815).
- Parker, W. S. (2017) Computer Simulation, Measurement, and Data Assimilation. *The British Journal for the Philosophy of Science*. 68(1), 273–304. [10.1093/bjps/axv037](https://doi.org/10.1093/bjps/axv037).
- Parker, W. S. (2018) The Significance of Robust Climate Projections In *Climate Modelling*, Lloyd, E. A. and Winsberg, E. (eds). Springer International Publishing. pp. 273–296. [10.1007/978-3-319-65058-6\\_9](https://doi.org/10.1007/978-3-319-65058-6_9).
- Raja, N. B., Dunne, E. M., Matiwane, A., Khan, T. M., Nätscher, P. S., Ghilardi, A. M. and Chatopadhyay, D. (2022) Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature Ecology & Evolution*. 6(2), 145–154. [10.1038/s41559-021-01608-8](https://doi.org/10.1038/s41559-021-01608-8).
- Schroeter, N., Toney, J. L., Lauterbach, S., Kalanke, J., Schwarz, A., Schouten, S. and Gleixner, G. (2020) How to Deal With Multi-Proxy Data for Paleoenvironmental Reconstructions: Applications to a Holocene Lake Sediment Record From the Tian Shan, Central Asia. *Frontiers in Earth Science*. 8.
- Schupbach, J. N. (2018) Robustness Analysis as Explanatory Reasoning. *The British Journal for the Philosophy of Science*. 69(1), 275–300. [10.1093/bjps/axw008](https://doi.org/10.1093/bjps/axw008).
- Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., von der Heydt, A. S., Knutti, R., Mauritsen, T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B. and Zelinka, M. D. (2020) An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence. *Reviews of Geophysics (Washington, D.C.: 1985)*. 58(4), e2019RG000678. [10.1029/2019RG000678](https://doi.org/10.1029/2019RG000678).
- Staley, K. (2004) Robust Evidence and Secure Evidence Claims. *Philosophy of Science*. 71(4), 467–488. [10.1086/423748](https://doi.org/10.1086/423748).
- Tal, E. (2017) Calibration: Modelling the measurement process. *Studies in History and Philosophy of Science Part A*. 65-66, 33–45. [10.1016/j.shpsa.2017.09.001](https://doi.org/10.1016/j.shpsa.2017.09.001).
- The Paleobiology Database (????) The Paleobiology Database (PBDB).
- van Dam, J. A. and Utescher, T. (2016) Plant- and micromammal-based paleoprecipitation proxies: Comparing results of the Coexistence and Climate-Diversity Approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 443, 18–33. [10.1016/j.palaeo.2015.11.010](https://doi.org/10.1016/j.palaeo.2015.11.010).
- Vezér, M. A. (2015) *Aggregating Evidence in Climate Science: Consilience, Robustness and the Wisdom of Multiple Models*. Ph.D. thesis. University of Western Ontario.
- Vezér, M. A. (2017) Variety-of-evidence reasoning about the distant past: A case study in paleoclimate reconstruction. *European Journal for Philosophy of Science*. 7(2), 257–265. [10.1007/s13194-016-0156-y](https://doi.org/10.1007/s13194-016-0156-y).
- Watkins, A. (2023) Using Paleoclimate Analogues to Inform Climate Projections. *Perspectives on Science*. pp. 1–58. [10.1162/posc\\_a.00622](https://doi.org/10.1162/posc_a.00622).
- Westerhold, T., Marwan, N., Drury, A. J., Liebrand, D., Agnini, C., Anagnostou, E., Barnett, J. S. K., Bohaty, S. M., Vleeschouwer, D. D., Florindo, F., Frederichs, T., Hodell, D. A., Holbourn, A. E., Kroon, D., Lauretano, V., Littler, K., Lourens, L. J., Lyle, M., Pälike, H., Röhl, U., Tian, J., Wilkens, R. H., Wilson, P. A. and Zachos, J. C. (2020) An astronomically dated record of Earth's climate and its predictability over the last 66 million years. *Science*. 369(6509), 1383–1387. [10.1126/science.aba6853](https://doi.org/10.1126/science.aba6853).
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 3(1), 160018. [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- Wilson, J. (2023) Paleoclimate analogues and the threshold problem. *Synthese*. 202(1), 17. [10.1007/s11229-023-04202-6](https://doi.org/10.1007/s11229-023-04202-6).

- Wilson, J. and Boudinot, F. G. (2022) Proxy measurement in paleoclimatology. *European Journal for Philosophy of Science*. 12(1), 14. [10.1007/s13194-021-00444-8](https://doi.org/10.1007/s13194-021-00444-8).
- Winsberg, E. B. (2018) *Philosophy and climate science*. Cambridge University Press. Cambridge, UK.
- Wylie, A. (1989) Archaeological Cables and Tacking: The Implications of Practice for Bernstein's 'Options Beyond Objectivism and Relativism'. *Philosophy of the Social Sciences*. 19(1), 1–18. [10.1177/004839318901900101](https://doi.org/10.1177/004839318901900101).
- Wylie, A. (2002) *Thinking from Things: Essays in the Philosophy of Archaeology*. University of California Press.
- Wylie, A. (2011) Critical Distance: Stabilising Evidential Claims in Archaeology In *Evidence, Inference and Enquiry*, Dawid, P., Twining, W., and Vasilaki, M. (eds). British Academy.
- Zachos, J. C., Dickens, G. R. and Zeebe, R. E. (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*. 451(7176), 279–283. [10.1038/nature06588](https://doi.org/10.1038/nature06588).