Response to Reviewers for "How large are temporal representativeness errors in paleoclimatology?"

Dan Amrhein

August 31, 2019

I appreciate the thoughtful reviews on this manuscript. I have addressed them point-by-point below, with reference to changes in the paper when appropriate.

Reviewer 1

This study addresses an important issue, which often is forgotten in the paleodata comparisons. Paleoclimatic measurements are usually dated and mostly dating uncertainties are communicated, too. However, depending on the archive and sample methods, measurements are often time averages, integrated over a specific period. This averaging period needs to be taken into account if these measurements are compared to simulations or other observations with different averaging periods. This study makes an important contribution to quantifying this error source.

The topic is highly relevant for Climate of the Past, ideas are novel and substantial conclusion reached. I found the figures, which introduce the basics concepts, very clear and helpful illustration. However, I assume the level of text and presented mathematical background goes beyond the average reader of this journal. Therefore, I suggest major revision, which should mainly simplify the entire text, to make it clearer and more accessible to a broad audience.

I appreciate the comments of Reviewer 1 and their support for the relevance of the work. One of my goals in writing this paper is to make some of the machinery of time series analysis, which I found important and insightful for thinking about problems of what proxies represent, relevant and accessible to the paleo community, so I welcome ideas on how to improve this.

Reflecting suggestions from both reviewers, I have overhauled the paper to emphasize a simplified description of the errors motivated by a minimal working example (Fig. 1). The bulk of the derivations have been moved to the appendix

and the number of equations in the main text reduced from 28 to 7.

- Line 11: Many expressions have not been introduced, yet and are therefore not clear to the reader. What is the "target interval". Better avoid abbreviations like "tau" in the abstract.

This term has been replaced by "climate interval of interest," and the abbreviations removed from the abstract.

- Line 13: What is meant by "archive smoothing" and "anti-aliasing"?

I have replaced "archive sampling" with an explanation that "the paleoclimate archive has been smoothed in time prior to sampling" and saved discussions of anti-aliasing filters until later.

- Page 2, lines 3ff.: I would suggest bullet points for the various error sources

Done, see new page 3.

Fig. 1: Great, this makes the problem easily accessible.

Table 1: This looks more like the variable list of the 500-page book than for a CP article. It may help to have reduced (basic) version of the mathematical background in the main part of the paper and the derivations in a supplement.

As suggested, the great bulk of the derivations was moved to the appendix.

Page 6, line 1: "boxcar" needs to be explained

This content has been moved to Appendix A, where it is explicitly defined.

Page 6, line 6: over over

Fixed.

Fig. 3: Labeling too small

Thank you, I oriented the figure in landscape to make fonts larger.

References: Latex code remained in the pdf version.

Fixed, thank you.

Reviewer 2

The paper by Armhein addresses an important topic in palaeoclimatology,

namely that of time representativeness of proxy data and the subsequent use of that data to compare between models and other data sets. I found the paper very hard to read and very technical, especially for COTP, though I would consider myself at the more technical end of the audience for this type of paper. I wonder if this work might be served better by publishing the mathematics in a more theoretical journal and subsequently interpretation and case studies in COTP.

A key goal of this paper is to communicate some results from time series analysis, which give us a powerful set of tools for thinking about paleo sampling issues, to the paleo community. I welcome the feedback given on how to streamline the presentation of the mathematical details so that the paper is well-suited to the readership of COTP. As part of restructuring the paper, I moved most of the equations and technical details to the appendices and tried to focus on results that are useful for constructing and interpreting paleo records.

I found the figures well-presented but hard to interpret, and I found the table of notation (Table 1) particularly unhelpful; it looks like it was put together in 5 minutes.

I worked to streamline figure captions and eliminated the table, reflecting the simpler descriptions now in the main text.

The key concept seems to revolve around the idea that there is a target measurement x which is desired to be averaged over temporal period τ_x, but the scientist only has access to an observation y which is averaged over an interval of τ_y. The origin of both the target and the observation is the 'true' time series $r(t)$. The paper's goal seems to be to estimate the error in using y to estimate x. I found this slightly strange, for two reasons. The first is that x and y are assumed to be centred at different time points. It's not clear to me why the centring needs to be (that) different.

I included the possibility of an offset between x and y both to make the problem more general (as such offsets inevitably arise, and I was curious what their effect was) and to permit computing errors from age model uncertainty, which are common in paleoclimate and which are modeled here as a stochastic offset. The zero-offset instance is a specific case of the general treatment.

The second is that, if I were doing this, I would try to estimate $r(t)$ as well as I possibly could, and then create smooths from this to allow me to compare with other time series. Perhaps this isn't possible, but I couldn't see why.

It is challenging to estimate $r(t)$ (one has to use paleo data). In particular, we have effectively no information about $r(t)$ at frequencies higher than data sampling rate, and yet this information is important for estimating representational errors. In the paper I argue that we don't need $r(t)$, but rather an estimate of its statistics, which simplifies the problem somewhat. This point is

now discussed starting on p3 l14.

I wonder if some of these problems might be solved by having a maximally simple running example at the start that makes clear the novelties of the approach using the bare minimum of maths/notation. That seems to be what has been attempted in Figure 1, though this just raises further questions. The light grey line appears to be $r(t)$ (this isn't mentioned?) which is unobserved. Labelling the vertical axis as $r(t)$ is slightly confusing since y and x are not values from $r(t)$ but averages from it. We would like to get x from y where y is centred on t minus Δ over period τ_y and x is centred on t over period τ_x. The key quantity theta is defined as the different between x and y, and most of the maths proceeds with the assumption that $r(t)$ is weakly stationary.

Figure 1 has been expanded to give more intuition for different kinds of error that may arise and to serve as a more complete maximally simple example. I refer back to it repeatedly in the text.

At this point I find myself a bit stuck as to the appropriateness of how this works. I don't think I've ever seen a situation similar to this. Aside from the stationarity assumption (which might hold locally?) in all the cases I've worked we have multiple y observations covering the spanned period. (I suppose it might be more realistic if $\tau_x < \tau_y$ as often we're interested in estimating a climate value at a more precise time point than the observations).

I agree that usually more than one observation is being used. One of the motivations for this work, which I have emphasized here (p7, l13), is this problem of computing mean values from multiple observations. The minimum variance solution for a mean value depends not just on the values of the multiple observations, but also on their uncertainties (see e.g. Wunsch (2006), 2.7.5). The problem of determining uncertainties individually (what I do in the paper) is thus crucial for determining how measurements should be weighted to estimate time means and uncertainties of paleoclimate intervals.

I'm further confused by the lower panel of Fig 1 which I initially assumed was the distribution of the error described by τ_x (they nearly match in magnitude) but is actually a completely different measurement uncertainty. I don't think that panel is helpful here. However, the legend points out that the uncertainty in theta is characterised by: sam- pling procedures, variability of r, archive smoothing, and chronological uncertainty. I totally agree with these fantastically important points. It's a strangely important sen- tence to appear in a figure caption. Having the lower panel display just one of these is confusing.

Figure 1 has been modified to remove the chronological uncertainty PDF. A sentence describing the different origins of uncertainty appears now in the abstract (l. 9).

In Section 2 the paper dives into a lot of technical detail which I followed mathematically but quickly lost all the interpretation. I was hoping to pick things up again in Figure 2 which shows the frequency representation of the boxcar function at the top and the power transfer function at the bottom for assumed values of Δ, τ_x, and τ_y. This seems to show which frequencies are contributing most to the TR error. However I've read the bullet points in Section 2.3 about 5 times now and I can't follow them. They talk about certain summary statistics which aren't shown in the Figure (226 and 1325 years?).

Thank you for the feedback. I have moved the technical description to the appendix to make less of a roadblock for the reader. Rather than describing explicitly the cutoff frequencies of the transfer function, I discuss qualitatively the fact that one frequency band is primarily responsible for error variance.

Figure 3 rescues things a little bit by being a bit clearer but even then it doesn't intro- duce the 4 panels or the colours so I can only make good guesses as to what some these plots are showing. The top two panels are particularly helpful. Unfortunately without any further interpretation of Figures 2 and 3 I was completely lost beyond Sec- tion 3.1, which is a big shame. I really wanted to follow this.

A more detailed description of all of the panels in this figure now appears on p7, l26. I have worked to smooth the transition between the technical setup and the subsequent interpretation to help the reader connect the dots. Hopefully the paper is now easier to follow.
Some minor points:

** I got very confused about the role of τ_a. I thought the archive smoothing was represented by τ_y as that is what we are observing?*

The archive smoothing (now defined on p3, l25) is meant to stand in for processes (e.g. bioturbation in sediment cores) that can smooth a record prior to sampling it. Sampling also has a smoothing effect, so that a sampled record has potentially been smoothed twice.

** Section 2 I thought was mis-named. It's not really a statistical model in a generative sense. It's more a collection of useful summary statistics that can indicate problems in temporal representativeness. **

In the new version, this (now less technical) section has been renamed "Origins of temporal representativeness error"

*The glossary in Table 1 needs at least a sentence of interpreta- tion for the more complicated quantities. Saying e.g. $H(v)$ represents the power transfer function is fine, but what would be more useful is to say that high values of this for fre- quencies v indicate a large contribution to the variance of theta **

Thank you for this suggestion. This table has been removed.

*P6L11 "archives". Also this sentence is quite unclear and where my confusion about τ_a stems from **

This sentence has been removed and archive sampling defined on p3, l25.

Eq 18 starts to get very confusing when the double square brackets are used to indicate expectations over different random variables. Perhaps use a subscript?

Done; thank you.

Cited: Wunsch, C. (2006). *Discrete inverse and state estimation problems: with geophysical fluid applications.* Cambridge University Press.

How large are temporal representativeness errors in paleoclimatology?

Daniel E. Amrhein[1]

[1]University of Washington School of Oceanography and Department of Atmospheric Sciences

Correspondence: Dan Amrhein (amrhein@uw.edu)

Abstract. Ongoing work in paleoclimate reconstruction prioritizes understanding the origins and magnitudes of errors that arise when comparing models and data. One class of such errors arises from assumptions of proxy temporal representativeness (TR), i.e. how accurately proxy measurements represent climate variables at particular times and time intervals. Here we consider effects arising when 1) the time interval over which data average and the climate interval of interest have different durations, 2) those intervals are offset from one another in time (including when those offsets are unknown due to chronological uncertainty), and 3) the paleoclimate archive has been smoothed in time prior to sampling. Because all proxy measurements are time averages of one sort or another, and it is challenging to tailor proxy measurements to precise time intervals, such errors are expected to be common in model-data and data-data comparisons, but how large and prevalent they are is unclear. This work provides a first-order quantification of temporal representativity errors and to studies the interacting effects of sampling procedures, archive smoothing, chronological offsets and errors (e.g. arising from radiocarbon dating), and the spectral character of the climate process being sampled.

Experiments with paleoclimate observations and synthetic time series reveals that TR errors can be large relative to paleoclimate signals of interest, particularly when the time duration sampled by observations is very large or small relative to the target time duration. Archive smoothing can reduce sampling errors by acting as an anti-aliasing filter, but destroys high-frequency climate information. The contribution from stochastic chronological errors is qualitatively similar to that when an observation has a fixed time offset from the target. An extension of the approach to paleoclimate time series, which are sequences of time-average values, shows that measurement intervals shorter than the spacing between samples lead to errors, absent compensating effects from archive smoothing. Nonstationarity in time series, sampling procedures, and archive smoothing can lead to changes in TR errors in time. Including these sources of uncertainty will improve accuracy in model-data comparisons and data comparisons and syntheses. Moreover, because sampling procedures emerge as important parameters in uncertainty quantification, reporting salient information about how records are processed and assessments of archive smoothing and chronological uncertainties alongside published data is important to be able to use records to their maximum potential in paleoclimate reconstruction and data assimilation.

1 Introduction

Paleoclimate records provide important information about the variability, extremes, and sensitivity of Earth's climate to greenhouse gases on time scales longer than the instrumental period. As the number of published paleoclimate records has grown and the sophistication of numerical model representations of past climates has improved, it has become increasingly important to understand the uncertainty with which paleoclimate observations represent climate variables so that they can be compared to one another and to model output. Additionally, quantifying uncertainty is important for ongoing efforts to assimilate paleoclimate data with numerical climate models (e.g., Hakim et al., 2016; Amrhein et al., 2018).

Paleoclimate records can have errors arising from many different sources: biological effects (e.g., Elderfield et al., 2002; Adkins et al., 2003), aliasing onto seasonal cycles (Wunsch, 2000; Fairchild et al., 2006; Dolman and Laepple, 2018), spatial representativeness (Van Sebille et al., 2015), proxy-climate calibrations (e.g., Tierney and Tingley, 2014), and instrument errors, to name a few. This paper focuses on errors from temporal representativeness (TR), which we define as the degree to which a measurement averaging over one time interval can be used to represent a second, target time interval. For instance, in a data assimilation procedure that fits a model to observations at every year, it is important to know the uncertainty associated with relating a decadal-average proxy observation to an annual-average target interval. Furthermore, computing a mean is often the implicit goal of binning procedures that combine observations from within a target time period such as a marine isotope stage, and we expect those observations to have errors that vary with their averaging duration and offsets from the target. Importantly, the term "error" is not meant to connote a procedural error on behalf of a collector or user of observations: Given the sparsity of data and the nature of geophysical time series, there is often a good rationale to use one time period to approximate another that is adjacent or has a different duration. Our goal is to understand the uncertainty arising in such a representation in a general framework.

Much of the previous study of errors arising from sampling in time has focused on aliasing, whereby variability at one frequency in a climate process appears at a different frequency in discrete samples of that process. Pisias and Mix (1988) described consequences of aliasing in the study of deterministic peaks in climate spectra due to Milankovich forcing. Wunsch and Gunn (2003) described criteria for choosing sample spacing so as not to alias low-frequency variability in sediment cores, and Wunsch (2000) demonstrated how aliasing can lead to spurious spectral peaks in ice core records. Beer et al. (2012) and von Albedyll et al. (2017) describe how running means can reduce aliasing of solar cycle variability in ice core records. In paleoclimate, measurements are often unevenly spaced in time due to changes in archive deposition rates; Jones (1972) showed that aliasing is present and even exacerbated in unevenly-sampled records relative to regularly sampled ones. Anderson (2001) and McGee et al. (2013) describe how bioturbation and other diagenetic processes smooth records in time and may reduce aliasing errors.

A second area of previous focus stems from chronological uncertainties, whereby times assigned to measurements may be biased or uncertain. In some cases, such as for radiocarbon dating, estimates of these uncertainties are available from Bayesian approaches that incorporate sampling procedures (Buck, 2004; Buck and Millard, 2004; Bronk Ramsey, 2009); practices for incorporating this information into model-data or data-data comparisons vary, and developing tools for analyzing

chronological uncertainty is an active area of research. Huybers and Wunsch (2004) include the effect of uncertainties in tie points in order to align multiple records of Pleistocene oxygen isotopes, and Haam and Huybers (2010) developed tools for estimating the statistics of time-uncertain series. The effect of time uncertainty on estimates of signal spectra is modest in some cases (Rhines and Huybers, 2011), in part because time uncertainty acts to smooth high-frequency variability when computed as an expectation over a record (Moore and Thomson, 1991).

This paper synthesizes effects contributing to TR errors in an analytical model and explores their amplitudes and dependence on signal spectra and sampling time scales. Extending results from time-mean measurements to time series demonstrates how sampling practices can lead to aliasing errors when records are not sampled densely, e.g. when an ocean sediment core is not sampled continuously along its accumulation axis. While we do not claim that TR error is the most important source of uncertainty in paleoclimate records, it does appear to be large enough to affect results in some cases. Moreover, this work is a step towards reducing the number of "unknown unknowns" in paleoclimate reconstruction.

2 Origins of temporal representativeness error

Our focus is first on errors arising when a mean value computed over one time period is used to represent another time period – for instance, when a time average over 20-19 kya (thousand years ago) is used to represent an average over 23-19 kya, the nominal timing of the Last Glacial Maximum (Clark et al., 2012). We define the TR error θ as the difference between x and y

$$\theta = x - y. \tag{1}$$

where y is affected by one or more type of TR error. As illustrated using a synthetic time series in Figure 1, our focus is on TR errors arising when:

- The time interval over which an observation averages in time (τ_y) has a different length from that of the targeted time interval (τ_x; Figure 1a).
- The time interval τ_y is offset in time from τ_x by a time Δ (Figure 1b). These offsets can be either known or, in the presence of chronological uncertainty of observations, stochastic and unknown.
- The paleoclimate archive was smoothed prior to sampling, whether by by bioturbation, diagenesis, residence times in karst systems upstream of speleothems (Fairchild et al., 2006), or other effects. In order to perform a first-order exploration of smoothing effects, we represent archive smoothing moving average over a time scale τ_a. Figure 1c illustrates how smoothing introduces errors for the case where $\tau_a = 2\tau_x$.

Visual inspection of Figure 1 yields some intuitive expectations. As the observational time interval τ_y grows small relative to τ_x, one expects TR errors to grow as the observation "feels" more of the variability at high frequencies. TR errors could also be expected to grow as a measurement is increasingly offset from the target in time. But interactions between different types of errors complicate the picture: for instance, in some cases a measurement interval that is short relative to τ_x might have smaller

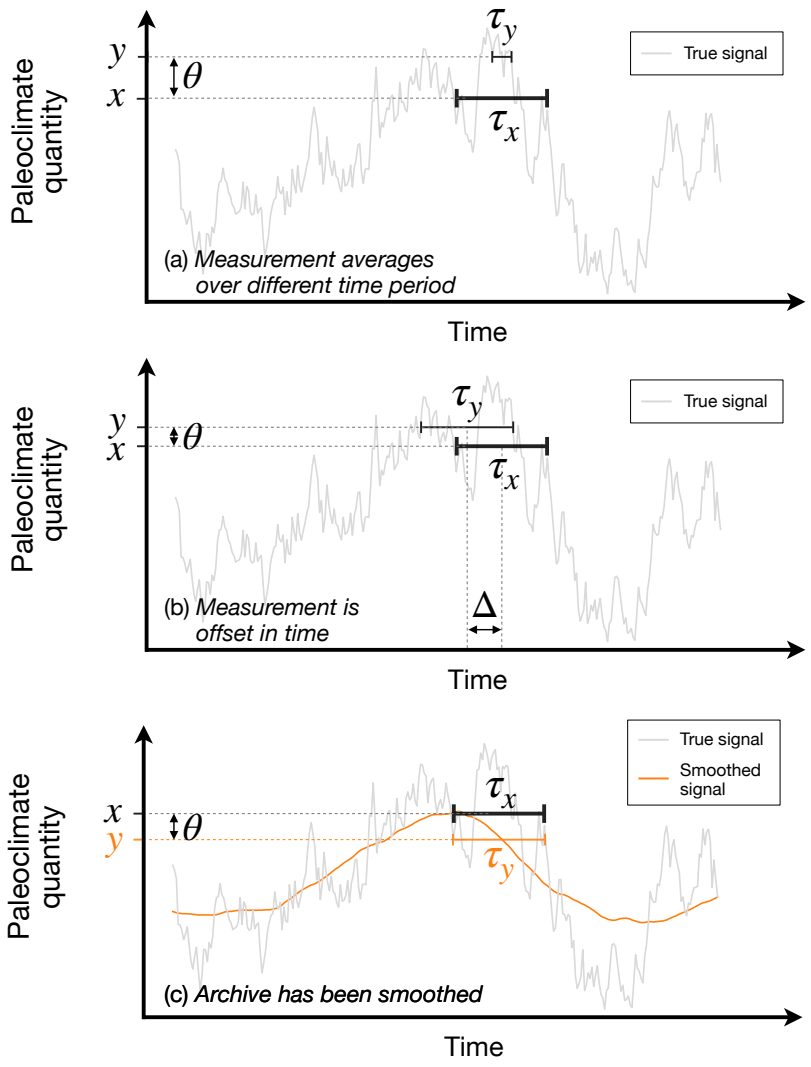

Figure 1. Several factors can contribute to temporal representativeness errors, defined here as the difference θ between a true time-average paleoclimate quantity x and a measurement y that averages over a different time interval. These effects are illustrated using a synthetic autoregressive time series. In each panel, the true quantity x is the same. Panel (a) shows the difference when the y averages over a time duration τ_y that is 5 times shorter than the averaging interval τ_x of the target value. Panel (b) shows the error when the observed and target averaging intervals are the same, but the observation is centered on a different value in time. Additional uncertainties, not shown here but discussed in the text, arise if the time offset is a stochastic random variable, as can occur e.g. with chronological uncertainties from radiocarbon dating. Panel (c) illustrates effects when the observation spans the correct time interval, but when the paleoclimate archive being sampled stores a smoothed version of the true signal; here that smoothing has a timescale of $\tau_a = 2\tau_x$. These errors are merely examples and are not meant to argue, e.g., that offset errors are always greater than errors from different averaging periods.

error if it is also offset in time, or if it samples an archive that stores a smoothed version of the climate signal. Subsequent sections examine interactions between various TR error sources.

This list is not exhaustive and neglects, for instance, effects from small numbers of foraminifera in sediment core records and other errors that are inherited from the construction of $r(t)$ from proxy observations. To isolate TR errors we assume that observations directly sample the true climate process, $r(t)$. This approach is intended to be complementary to proxy system models (PSMs; e.g., Evans et al. (2013)) that relate proxy quantities to climate variables ("forward operators" in the language of data assimilation). The procedures described may be used to estimate TR uncertainty when PSMs do not; when they do, the model can provide a theoretical grounding for understanding those results. Variances from multiple error sources can be added together under the approximations that they are independent and Gaussian. When these assumptions fail, more holistic forward modeling of errors in PSMs may be necessary.

3 Estimating temporal representativeness error

Because in paleoclimatology we do not have complete knowledge of the underlying climate signal $r(t)$ (it is what we are trying to sample), it is impossible to infer what the TR error is for each measurement as done in the synthetic example (Figure 1). Instead, our aim is to determine typical values for errors, which are important for data assimilation and for comparing models and observations and observations to one another. We will characterize TR error θ by estimating its variance, $\left\langle (\theta - \langle\theta\rangle)^2 \right\rangle$, where angle brackets denote statistical expectation. To do this, we approximate $r(t)$ as being weakly statistically stationary, meaning that its mean and variance do not change in time; caveats surrounding this assumption are addressed later in the paper. Under the weak stationarity assumption, the mean error $\langle\theta\rangle$ is zero, and we take the expectation by evaluating θ^2 at all the times in $r(t)$ to compute the variance,

$$\left\langle \theta^2 \right\rangle = \frac{1}{\tau_0} \int_{t_0}^{t_f} (x-y)^2 dt, \tag{2}$$

where t_0 and t_f are the initial and final times in $r(t)$, and $\tau_0 = t_f - t_0$. Intuitively, we are estimating the error in representing x by y *(at a single time)* as the time-mean squared difference of running means of $r(t)$ *(over all times)*. In practice, though we do not know $r(t)$, knowledge of its statistics is adequate to estimate $\left\langle \theta^2 \right\rangle$.

Representing TR error in the frequency domain (Appendix A) emerges as an intuitive way to describe errors that also provides closed-form expressions that can be readily integrated to explore the effects of different sampling and time series parameters. A basic result (see Equation A13) is that in the frequency domain, TR errors are represented as

$$\left\langle \theta^2 \right\rangle = \frac{1}{\tau_0} \int_0^{\infty} H(\nu, \tau_x, \tau_y, \tau_a, \Delta) \left|\hat{r}(\nu)\right|^2 d\nu. \tag{3}$$

where ν denotes frequency, H is a so-called transfer function, and $\left|\hat{r}(\nu)\right|^2$ is the power spectral density of the true signal $r(t)$. In effect, the error variance is a weighted sum of the power at different frequencies in $r(t)$, where the weights in frequency space (given by H) depend on how the paleoclimate record has been sampled and smoothed. This behavior is typical of aliasing,

where variance in the signal at one frequency appears erroneously in a measurement at a different frequency (in this case, at the zero frequency, which is the time mean).

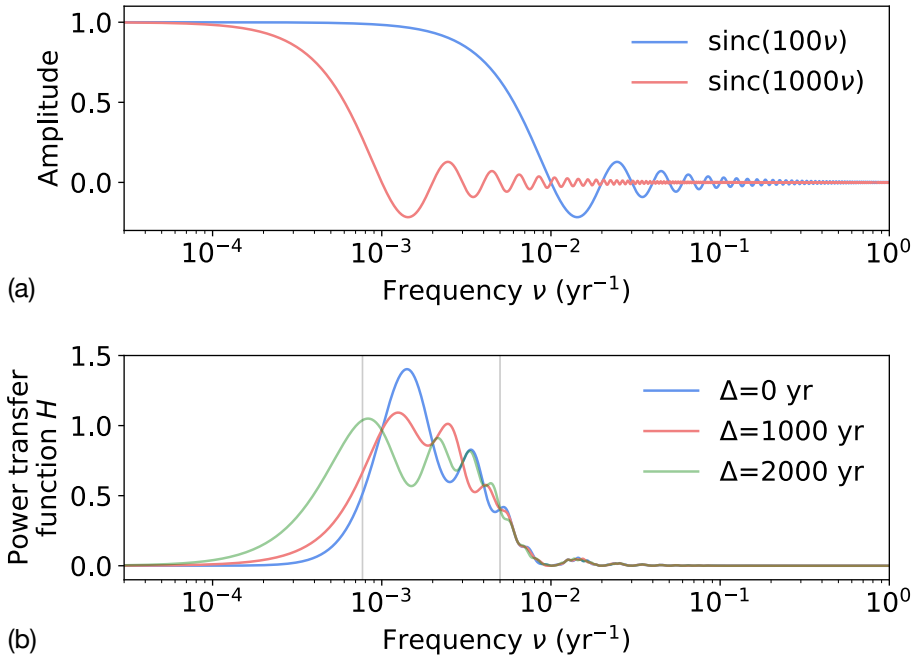

Figure 2. The power transfer function H (Equation 3) illustrates the dependence of temporal representativeness errors on frequencies in the climate signal and on sampling time scales. In the case where the offset Δ between measurement and target is 0, H is a squared difference of sinc functions $\mathrm{sinc}(\tau\nu)$ (A13), illustrated here for $\tau = 100$ years and $\tau = 1000$ years. (b) Transfer functions for three different values of the time offset Δ. Grey bars indicate the 1/200 and 1/1300 yr^{-1} frequencies, which approximately bound the frequencies contributing to TR errors in the $\Delta = 0$ case.

While the details are left to the appendix, it is noteworthy that in many practical cases, TR errors can be straightforwardly attributed to signal variability within a particular frequency band. This frequency-band behavior emerges because H is a squared difference of sinc functions (Figure 2a), which has a bump-like shape (Figure 2b). For instance, if a centennial mean is used to represent a millennial mean, in the absence of archive smoothing, the expected error variance is roughly equal to the variance in $r(t)$ at periods between 200 and 1300 years; the error is the same if a centennial mean is used to represent a decadal mean. Thus the difference between the sample and target averaging intervals (τ_y and τ_x) sets the frequency band that is aliased onto the mean. These effects are modulated in the presence of archive smoothing, and when there is a time offset in the measurement relative to the target, additional variability is aliased onto errors.

4 Illustrating TR error quantification at the Last Glacial Maximum by sampling a high-resolution paleoclimate archive

Here we explore the procedure for estimating TR errors described in the previous section in the context of estimating mean properties at the Last Glacial Maximum (LGM), the period roughly 20,000 years ago that is associated with the greatest land ice extent during the last glacial period. Following MARGO Project Members (2009) and others, LGM target quantities are defined to be estimates of time means over the 4000-year-long period from 23,000 to 19,000 years ago (23-19 kya),

$$x_{LGM} = \frac{1}{4000} \int_{-23,000}^{-19,000} r(t)\, dt.$$

We will consider TR errors arising from representing x_LGM by an observed 1000-year time-mean value that is centered on 21 kya,

$$y_{LGM} = \frac{1}{1000} \int_{-20,500}^{-21,500} r(t)\, dt. \tag{4}$$

Qualitatively, errors from this representation have the form illustrated in Figure 1a. Such an estimate – dated to within the LGM, but averaging over only a subset – could reasonably be included in a binned-average compilation of LGM data. However, because statistically robust averaging procedures must downweight uncertain observations according to observational error, including TR errors, is important to avoid biasing any binned averages. Similarly, were we to compare y_{LGM} to an LGM-mean estimate of $r(t)$ from a model without taking TR errors into account, we might erroneously conclude that the model did not fit the data.

How large is the TR error in representing x_{LGM} by y_{LGM}? We will illustrate the procedure proposed in Section 3 by taking a high-resolution climate record to be a true climate signal $r(t)$ and sampling it at longer time averages than the record spacing. Here we will use the the North Greenland Ice Core Project (NGRIP; Andersen et al. (2004)) 50-year average time series of oxygen isotope ratios (δ^{18}O). Equation 2 states that TR error variance is equal to the squared difference between running means of $r(t)$, averaged over the record length. Smoothing the NGRIP record with running means of length $\tau_x = 4000$ and $\tau_y = 1000$ yields time series of potential target and observation values x and y (black and red lines, Figure 3a). Their difference is the error θ (thick black line, 3a), and their squared difference is the blue line in 3b. The time mean $\langle \theta^2 \rangle$ (red line, Figure 3b) is 0.7 $\left(‰ \delta^{18}\mathrm{O}\right)^2$ and is the estimate of the error variance.

A prominent feature in Figure 3b is the time variability of TR error: in some time periods (including the LGM) errors are relatively small, whereas they are markedly larger at, e.g., 80-70 kya. This time-variability in errors arises from nonstationarities in the NGRIP oxygen isotope record. The transfer function (blue line, 3c), shows that for our choices of τ_x and τ_y, variability in the frequency band lying between roughly 2200 and 5300 year periods is responsible for TR errors. A wavelet analysis (Figure 3d) shows that increased variability in this band is coincident with increases in TR error variance: note, e.g., the the correspondence of high wavelet values in that band near -75000 years with contemporaneous large values in the blue line in 3b. Evidently, in the presence of nonstationary climate variability, TR errors can vary in time. They may also vary due to changes

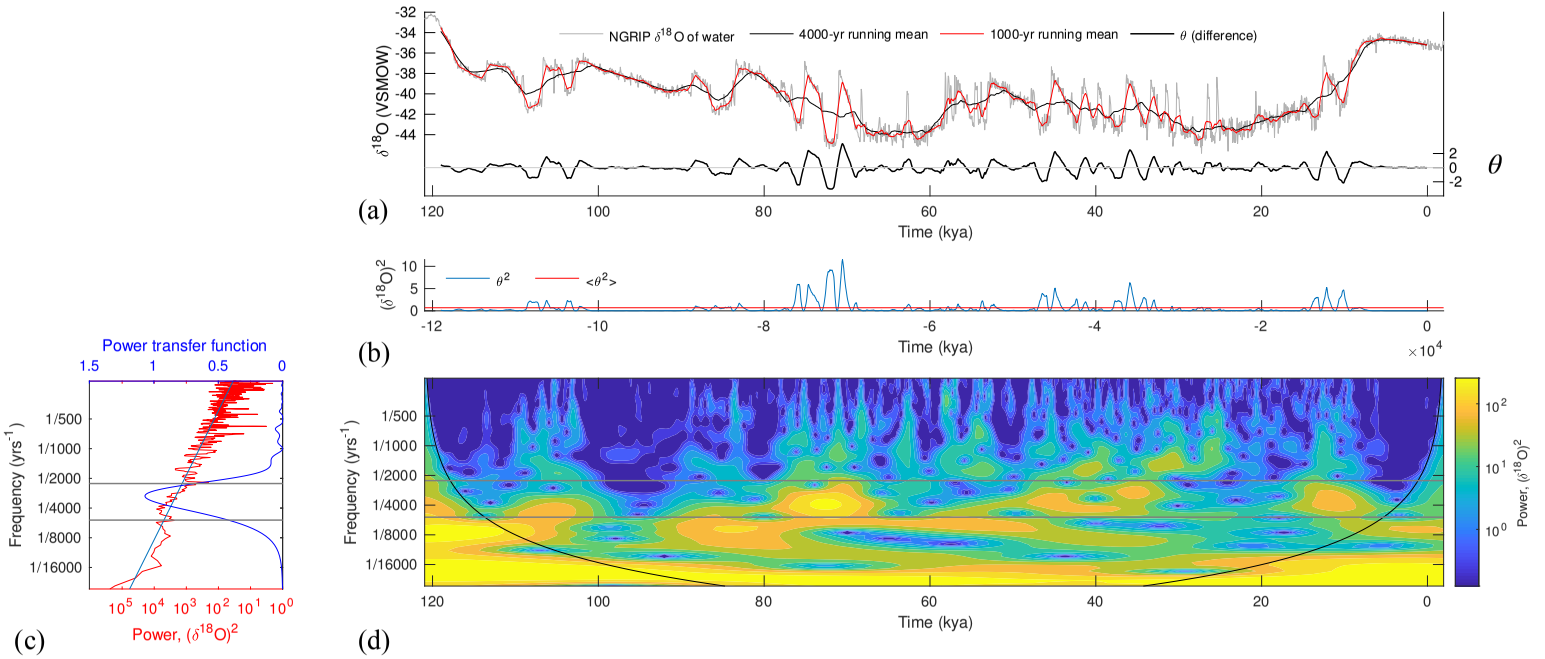

Figure 3. Temporal representativeness error in the time and frequency domains. Errors in representing a 4000-year mean by a 1000-year mean are estimated by computing the difference θ ((a), thick black line) between a 4000-year (red line) and 1000-year (thin black line) running mean of the NGRIP $\delta^{18}O_{ice}$ record (grey). The time average (red line, (b)) of θ^2 (blue line) is an estimate (0.7, units of $\left(‰\delta^{18}O\right)^2$) of the temporal representativeness error variance. Large values in θ^2 correspond to time periods with increased variability, as diagnosed by a wavelet analysis (d), particularly in the band between 2257 and 5298 year periods (grey lines). The light blue line in panel (c) is the power spectrum $\nu^{-\beta}$ with $\beta = 1.53$ derived by a least-squares fit to the NGRIP spectrum.

in sampling procedures over the course of constructing a time series, as discussed in Section 6. Observations of intervals with less variability in the TR error frequency band (e.g., the LGM) will be less susceptible to TR errors, an additional quantitative justification for the long-held process of focusing study on time-means of periods with relatively less variability.

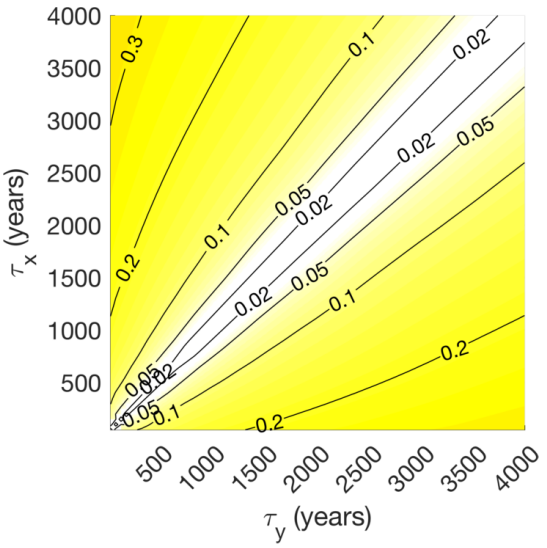

Figure 4. Error-to-signal variance fractions f (Equation 5) for estimates of time-mean values computed from the NGRIP record of Pleistocene oxygen isotopes contoured as a function of target averaging interval τ_x and observation averaging interval τ_y. A value of 0.1 means that TR error amplitudes are 10% of the "signal," which defined as the typical difference between two time averages over durations τ_x separated by 21,000 years.

Next, we will extend our analysis of NGRIP to cover a range of different values of τ_x and τ_y. To compare the NGRIP results to synthetic time series in the following sections with arbitrary units, we will analyze the unitless noise-to-signal standard deviation ratio,

$$f = \frac{\sqrt{\langle \theta^2 \rangle}}{\sigma}. \tag{5}$$

Because one motivation of studying the LGM is inferring differences from modern climate, we adopt as our "signal" amplitude the typical anomaly σ between two mean intervals of length τ_x separated by $= 21{,}000$. This quantity is estimated from the NGRIP time series for each value of τ_x.

Recall that for $\tau_x = 4000$ and $\tau_y = 1000$, the estimated error variance was 0.7 $\left(‰\delta^{18}\mathrm{O}\right)^2$. Figure 4 contours the results of the same calculation (now expressed as the noise-to-signal ratio f) for every combination of τ_x and τ_y between 10 and 4000 years. TR errors depend jointly on values of τ_x and τ_y. Errors are zero for $\tau_x = \tau_y$ (corresponding to an ideal sampling scheme) and increase monotonically away from those values. Errors are greatest (up to 30% of signal amplitudes) for small values of τ_y relative to τ_x, where TR error dwarfs the relatively small signal amplitudes that are typical of 21,000-year differences in

long-term time averages[1]. Subsequent sections extend this analysis to a broader set of sampling parameters (including archive smoothing and time offsets) as well as records with different spectral characteristics.

5 Exploring interactions between sampling parameters and signal spectra

The succinct expression of TR errors in terms of power spectra (3) is a clue that the spectral character of paleoclimate processes are an important factor for the amplitude of TR errors. To investigate how errors depend on the spectrum of $r(t)$, we will shift our focus away from observations and consider climate processes with power-law spectra, i.e. those whose power spectral densities $|\hat{r}(\nu)|^2$ have the form

$$|\hat{r}(\nu)|^2 \propto \nu^{-\beta}, \tag{6}$$

where β is termed the spectral slope (when plotted in log-log space, $\nu^{-\beta}$ is a straight line with slope $-\beta$). We choose this idealized form because spectra consistent with a power-law description are common in climate (Wunsch, 2003). White noise, which partitions variance equally across frequencies, has a spectral slope of 0; signals with a steeper slope (larger β) have a larger fraction of their variance originating from low-frequency variability. Here we consider spectral slopes $\beta = 0.5$ and $\beta = 1.5$, motivated by Huybers and Curry (2006), who fit paleoclimate records to spectral slopes between $\beta = 0.3$ and $\beta = 1.6$. Climatological spectral features that are not described by power laws, such as peaks due to deterministic astronomical forcing from annual or Milankovich variability, also contribute to errors (Pisias and Mix, 1988; Wunsch, 2000) but are not considered specifically in these examples. All calculations are performed by numerical integration of Equation (A13) by global adaptive quadrature.

5.1 Effects from archive smoothing and spectra

Similar to Figure 4, Figure 5 contours the noise-to-signal ratio f as a function of τ_x and τ_y, but now for four cases spanning two values of the archive smoothing time scale τ_a (0 and 1000 years) and two values of spectral slope β. Signals with steeper spectral slopes ($\beta = 1.5$ rather than $\beta = 0.5$), show smaller f values because TR errors, which originate at relatively high frequencies (Figure 2), are smaller relative to the greater amount of low-frequency variability, as discussed also by Wunsch (1978) and Wunsch (2003). The close resemblance between Figure 5b the equivalent figure computed from NGRIP (4), which has spectral slope of 1.53 (Figure 3c), is partly coincidental; analysis of synthetic records with spectral slopes of 1.5 (not shown) reveals variability in f because of variations about the power law distribution in finite-length, stochastically generated time series, of which NGRIP is arguably one realization. Nevertheless, the agreement shows correspondence between results from paleoclimate data and our idealized approach.

Dependencies on τ_x and τ_y change when we include archive smoothing (Figures 5a and 5; schematized in Figure 1a and 1c). These effects are evident primarily for $\tau_y < \tau_a$. In that "smoothed" regime, the largest values of f for small τ_y are reduced because archive smoothing removes some of the high-frequency variability that would otherwise be felt by observations and

[1]Error variances are equal if τ_x and τ_y are interchanged, but asymmetry in f arises because σ depends on τ_x.

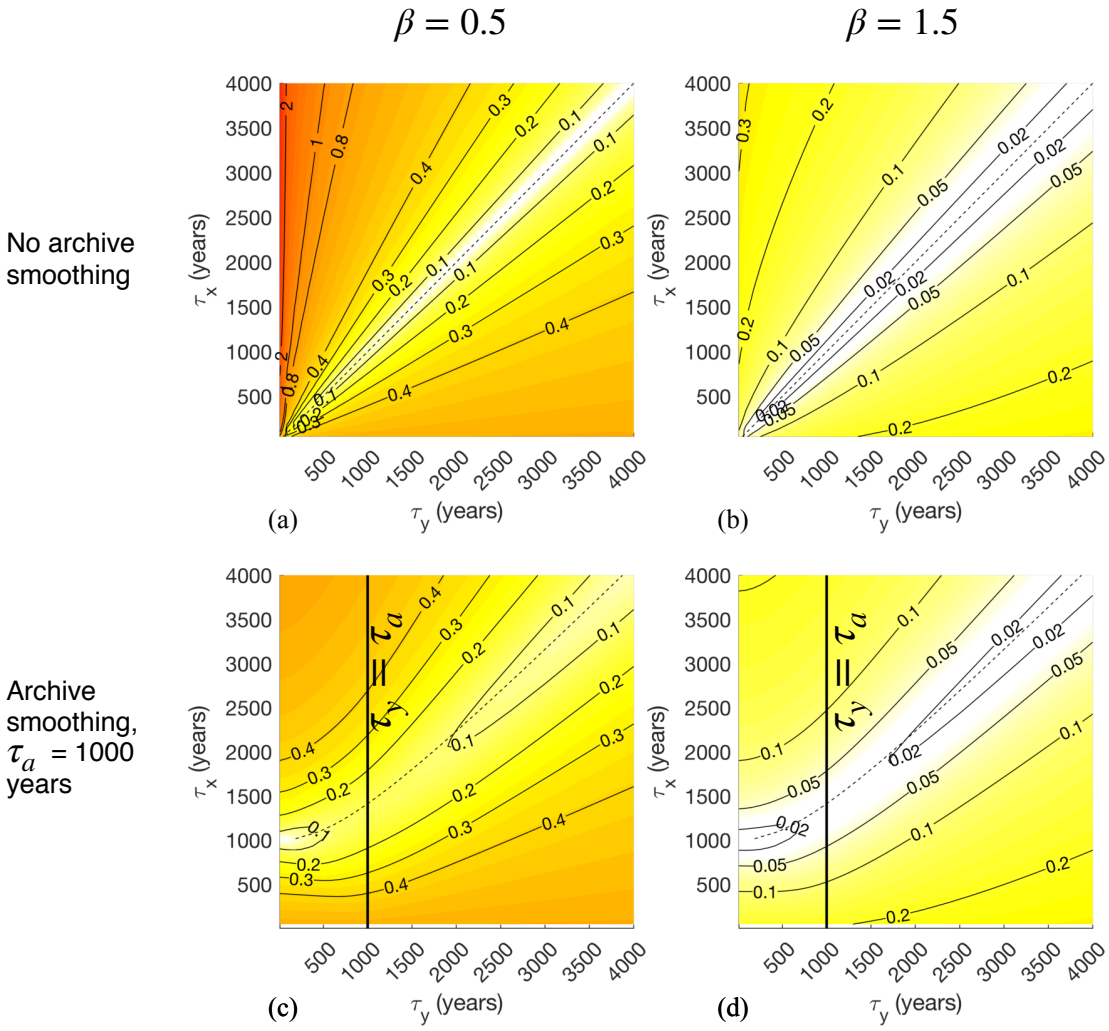

Figure 5. Error-to-signal fractions f for time-mean estimates plotted as a function of target averaging interval τ_x and observation averaging interval τ_y. Climate signal spectra are approximated as power law functions of frequency ($|\hat{r}(\nu)|^2 \propto \nu^{-\beta}$) with spectral slopes β equal to 0.5 (left column) and 1.5 (right column). The top row corresponds to a case with no archive smoothing ($\tau_a = 0$) while the bottom row corresponds to a case where the signal $r(t)$ is smoothed by a running mean over $\tau_a = 1000$ years. Values to the left of the bold line at $\tau_y = \tau_a$ lie in the "smoothed regime" where archive smoothing qualitatively affects results. Time scales were chosen to be relevant to the problem of time-mean estimation at the Last Glacial Maximum, ca. 20 kya. Dotted lines show values of $\tilde{\tau}_y$ derived to minimize error estimated according to Equation (A20).

erroneously aliased onto the mean. Another effect is that $\tau_y = \tau_x$ no longer minimizes f everywhere; in the smoothed regime, smaller values of τ_y lead to reduced TR error. This is because archive smoothing already provides a measure of time averaging; so that when $\tau_x = 1000$, the value of τ_y that minimizes error is close to zero, because anything longer would be "oversmoothing" the record and effectively giving a longer time average than τ_x. Archive smoothing also reduces the sensitivity of errors to the choice of τ_y for $\tau_y < \tau_a$. Finally, the presence of smoothing means that arbitrarily short choices of τ_x can no longer be resolved without error, as evidenced by the monotonic growth of error as τ_x decreases from τ_a.

To the extent that these simple experiments reflect actual paleoclimate sampling procedures, one could attempt to sample time-mean intervals to avoid TR errors. In the absence of archive sampling, the (trivial) result is that τ_y should be equal to τ_x. But the danger of oversmoothing means that this rule is not always appropriate for $\tau_a \neq 0$. Appendix A derives (A20) an approximate expression for the error-minimizing value, $\tilde{\tau}_y = \sqrt{\tau_x^2 - \tau_a^2}$, that is a function of both the target interval length and smoothing time scale. These values (dotted lines, Figure 5) are in good qualitative agreement with minimum TR error values.

5.2 Effects from known time offsets

Having explored how choices of τ_x and τ_y contribute to TR errors, we next illustrate effects from chronological offsets when $\Delta \neq 0$ (schematized in 1b). Motivated by the LGM time scale, we focus again on the case where τ_x is fixed to 4000 years and integrate A13 varying τ_y between 10 and 6000 years and Δ between 10 and 4000 years for values of $\beta = (0.5, 1.5)$ and $\tau_a = (0, 1000)$.

For all values of τ_a and β, errors grow monotonically away from the values $\Delta = 0$, $\tau_y = \tau_x$, which corresponds to the case with no TR error[2] (Figure 6). As in the previous section, a "smoothed regime" is evident for $\tau_y \leq \tau_a$ across all values of Δ shown: because archive smoothing damps variability in time, the errors from shifting an observation relative to the target become less severe. Another qualitative difference emerges for values of Δ that are greater or less than $\left|\tau_x - \tau_y\right|/2$ (blue line, Figure 6a and 6c). This boundary designates when the observed time period is sufficiently offset that it begins to fall outside the target interval; at that point, errors grow rapidly as Δ increases. As before, errors are more pronounced for $\beta = 0.5$ than for $\beta = 1.5$, with errors larger than the signal ($f > 1$) for small values of τ_y at all values of Δ for $\beta = 0.5$.

5.3 Effects from probabilistic time offsets

When the dating of a measurement is uncertain, a range of Δ values may be possible, as specified by a probability distribution function $p(\Delta)$. To explore effects from chronological uncertainty, we assume that $p(\Delta)$ is Gaussian about zero with standard deviation equal to the time scale σ_Δ. We include this uncertainty in TR error variance by taking a second expectation (denoted by $\langle\rangle_\Delta$, in addition to the time expectation in Equation 2) to give

$$\left\langle \left\langle \theta^2 \right\rangle \right\rangle_\Delta = \int_{-\infty}^{\infty} p(\Delta) \left\langle \theta^2 \right\rangle d\Delta. \tag{7}$$

[2] A small amount of oversmoothing is present at $\tau_y = \tau_x$ in the $\tau_a = 1000$ case, but is not qualitatively important.

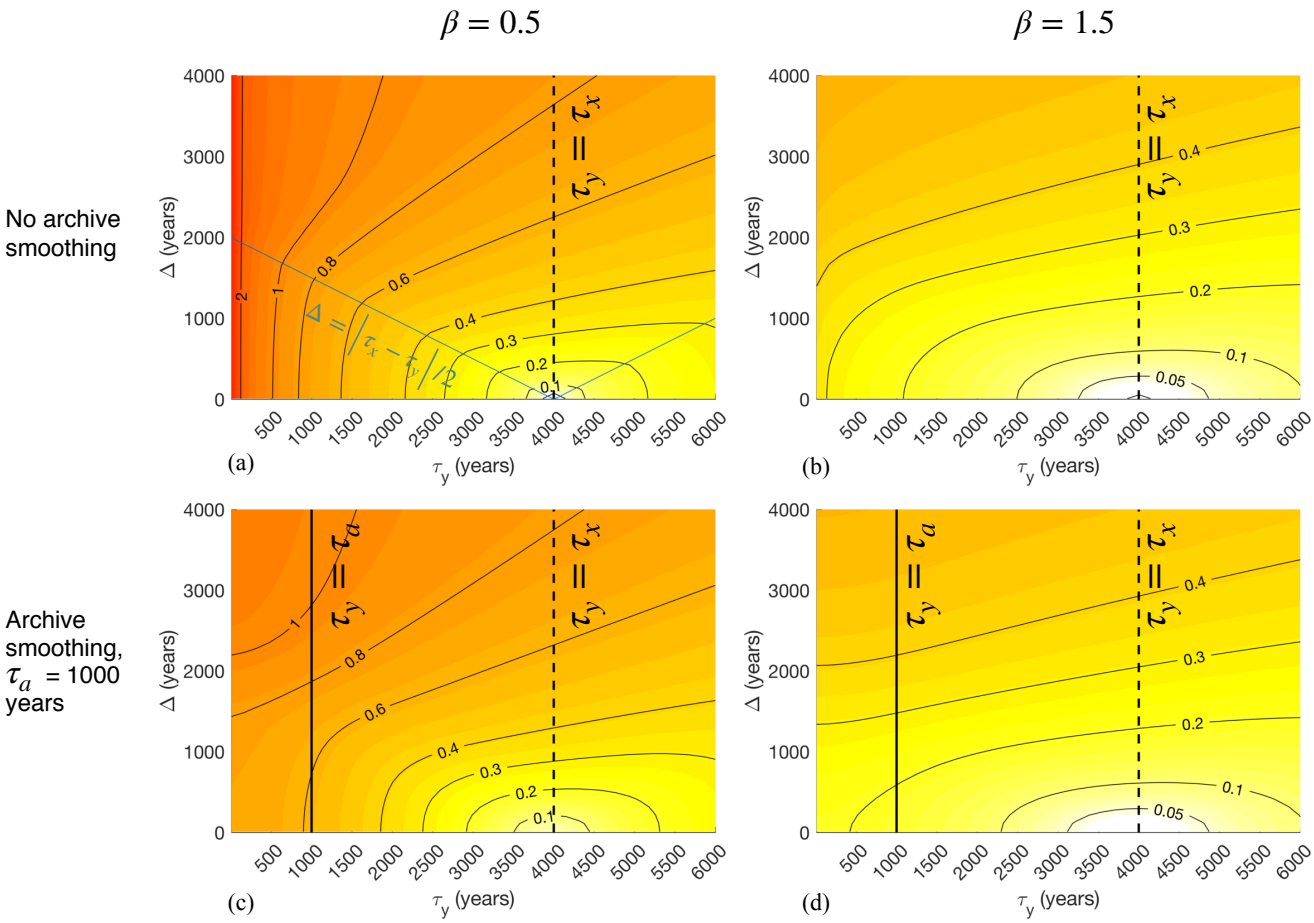

Figure 6. Same as Figure 5, but illustrating effects of offsets Δ between target and observational intervals on noise-to-signal ratios. Error fractions f are plotted as a function of the observational averaging interval τ_y and the standard deviation σ_Δ of a Gaussian distribution of observational offset centered on zero. In all cases, the target averaging interval is $\tau_x = 4000$, reflecting the nominal length of the Last Glacial Maximum. Values along the line $\tau_y = \tau_x$ strictly reflect the influence of chronological offsets. The blue line in panel (a) denotes values for which $\Delta = |\tau_x - \tau_y| / 2$, indicating the maximum values of Δ for which τ_x and τ_y completely overlap.

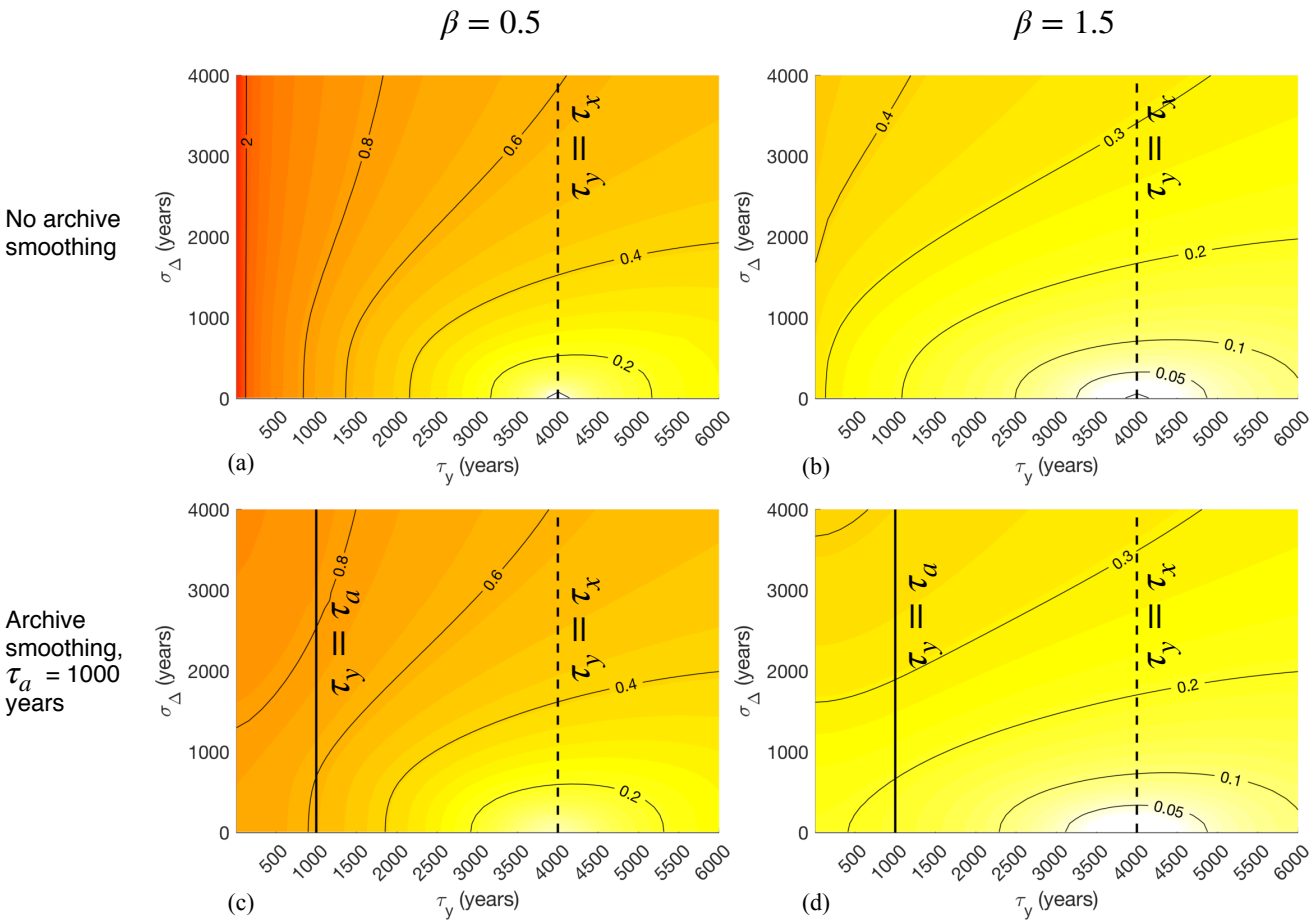

Figure 7. Same as Figure 5, but illustrating effects of chronological uncertainties in observations on noise-to-signal ratios. Error fractions f are plotted as a function of the observational averaging interval τ_y and the standard deviation σ_Δ of a Gaussian distribution of time offsets centered on zero. In all cases, the target averaging interval is $\tau_x = 4000$, reflecting the nominal length of the Last Glacial Maximum. Values along the line $\tau_y = \tau_x$ strictly reflect the influence of chronological uncertainty, which is zero when the observational offset is exactly known to be zero, (i.e., $\sigma_\Delta = 0$).

In practice, $p(\Delta)$ can be multimodal or otherwise non-Gaussian (e.g., from radiocarbon ages; Telford et al. (2004)) which could qualitatively change results. While not explored here, such errors can be investigated by integrating Equation (7) with different choices of $p(\Delta)$.

Integrating (7) and varying σ_Δ and τ_y shows that TR errors in the presence of Gaussian chronological uncertainty $p(\Delta)$ with standard deviation σ_Δ are qualitatively similar to those from a fixed offset $\Delta = \sigma_\Delta$ (cf. Figures 6 and 7). The transition in sensitivity to σ_Δ across $\sigma_\Delta = \left|\tau_x - \tau_y\right|/2$ is less pronounced than for the equivalent in Figure 6, consistent with the range of lags that is possible for any nonzero σ_Δ. Nevertheless, the intuitively sensible conclusion is that chronological errors will be gravest when uncertainties tend to place measurements outside of target intervals. Reduced errors in the smoothed regime $\tau_y \leq \tau_a$ indicate that archive smoothing can reduce effects from chronological errors in some cases.

6 Extension to time series

Paleoclimate time series are sequences of time-mean values; here, we discuss how the TR errors discussed for time-mean estimation affect transient records of climate variability. We show that in the absence of archive smoothing, dense sampling (i.e., setting the averaging interval equal to the spacing between measurements) is a nearly optimal approach to minimize TR errors.

The sampling theorem of Shannon (1949) states that sampling $r(t)$ instantaneously at times separated by a fixed time interval τ_s unambiguously preserves signal information only when $r(t)$ does not contain any spectral power at frequencies greater than $1/2\tau_s$ (called the Nyquist frequency, ν_{Nyq}). When this criterion is not met, the discrete signal is corrupted by aliasing, whereby variability in $r(t)$ at frequencies greater than ν_{Nyq} appears artificially at lower frequencies in the discrete signal. To mitigate aliasing, one can either increase the sampling rate or apply a low-pass "anti-aliasing" filter to $r(t)$ to attenuate power at frequencies higher than ν_{Nyq}.

In the process of constructing a paleoclimate time series, sampling time-mean values serves a moving average and, thereby, an anti-aliasing filter. Thus we expect sample averaging procedures to affect aliasing errors in time series, as also discussed by von Albedyll et al. (2017). Appendix B uses Shannon's theorem to obtain a frequency-domain expression for the TR errors for individual time series measurements. The procedure is to 1) define local (in time) values of τ_s^i and ν_{Nyq}^i for the i^{th} observation and 2) compute the expected errors if an entire time series were sampled using those local properties. To do this, we make the assumption that the sampling interval τ_s^i is locally constant: that is, for the i^{th} measurement y^i taken at time t^i, y^{i-1} was taken at time $t^i - \tau_s^i$, and y^{i+1} was taken at time $t^i + \tau_s^i$. If the sampling interval changes rapidly, conclusions from this approach might not apply. Again leaving the details to the appendix, we note that similar to the time-mean case, the error variance $\left\langle \theta^{i2} \right\rangle$ for the i^{th} observation is a weighted integral over the power density spectrum of $r(t)$ (B6). Unlike in the mean estimation case, where TR errors can be zero, nonzero error is unavoidable with uniform sampling because of differences between the shape of the sinc function and the ideal, abrupt frequency cutoff that minimizes error according to Shannon's theorem.[3]

[3]Sampling a paleoclimate archive nonuniformly in time could better approximate the ideal filter and reduce errors, but this may not be practical given the challenges of recovering and sampling paleoclimate data.

To demonstrate sensitivities to sampling parameters we again compute noise-to-signal ratios. In keeping with our local measure of TR error, we take the signal strength to be the standard deviation of the time series that would result if $r(t)$ were sampled with the same averaging and sampling interval as the i^{th} observation over 21,000 years, the approximate duration of the last deglaciation. The results are qualitatively similar to those for the time-mean case, with two main distinctions (cf. Figure 8 and 5). First, as discussed above, errors are always 10% or more of signal amplitudes because of errors arising from constructing a time series as a sequence of time mean values. Second, values of τ_y that minimize errors do not obey $\tau_y = \tau_s$, but are larger by a factor of roughly 1.2, suggesting that, absent considerations from archive smoothing, the ideal sample would span an interval slightly longer than the sampling interval to minimize errors. In practice, sampling densely (that is, without space between observations) is a good approximation of this error minimizing strategy.

As in the time-mean case, the effects of archive smoothing are large in a regime of sampling parameter space ($\tau_y \leq \tau_a$), implying that knowledge of τ_a is important for informing choices of τ_s and τ_y. Clearly, sampling at intervals $\tau_s < \tau_a$ will result in errors because some of the variability of interest will have been destroyed. Choosing a τ_s that is larger than both τ_y and τ_a will result in aliasing errors. Finally, within the smoothed regime $\tau_y \leq \tau_a$, TR errors are less sensitive to choices of τ_y than they are for $\tau_y \geq \tau_a$, meaning that sampling discontinuously (i.e., not densely) may not be problematic.

7 Discussion

This paper presents a framework for quantifying temporal representativeness (TR) errors in paleoclimate, broadly defined as resulting when one time average is represented by another. A simple model illustrates interacting effects from record sampling procedures, chronological errors, and the spectral properties of the climate process being sampled.

We find that TR errors for time mean estimates can be large relative to climate signals, with noise-to-signal standard deviation ratios greater than 1 in some cases, particularly where the observational interval τ_y is smaller than the target interval τ_x. These errors result from aliasing climate variability onto time mean estimates and can be mitigated to some degree by choices of sampling procedures and by archive smoothing, both of which act as anti-aliasing filters. However, archive smoothing can also destroy information about climate variability, and the combined effects of sampling and smoothing can over-smooth a record and lead to increased errors. TR errors due to sampling are not independent of chronological errors, but interact, for instance in the way that errors grow more quickly as a function of chronological uncertainty amplitude when that uncertainty is likely to place a measurement outside of a target interval (Figure 7). Given that these error variances were estimated using parameters representative of the LGM, it seems possible that TR errors may explain some of the disagreement among proxy measurements within that time period (e.g., MARGO Project Members, 2009; Caley et al., 2014), though nonstationarities may cause TR errors to be overestimated for climate intervals like the LGM that appear to be quiescent relative to other time periods. While we do not claim that TR errors are the largest source of error for any particular proxy type or reconstruction problem, they may be in some cases. The tools presented can be used to assess likely error amplitudes.

Though not the principal goal, these analyses provide a basis for sampling practices that minimize errors, for instance for avoiding oversmoothing that can arise through the combined effects of sampling and archive smoothing. When constructing

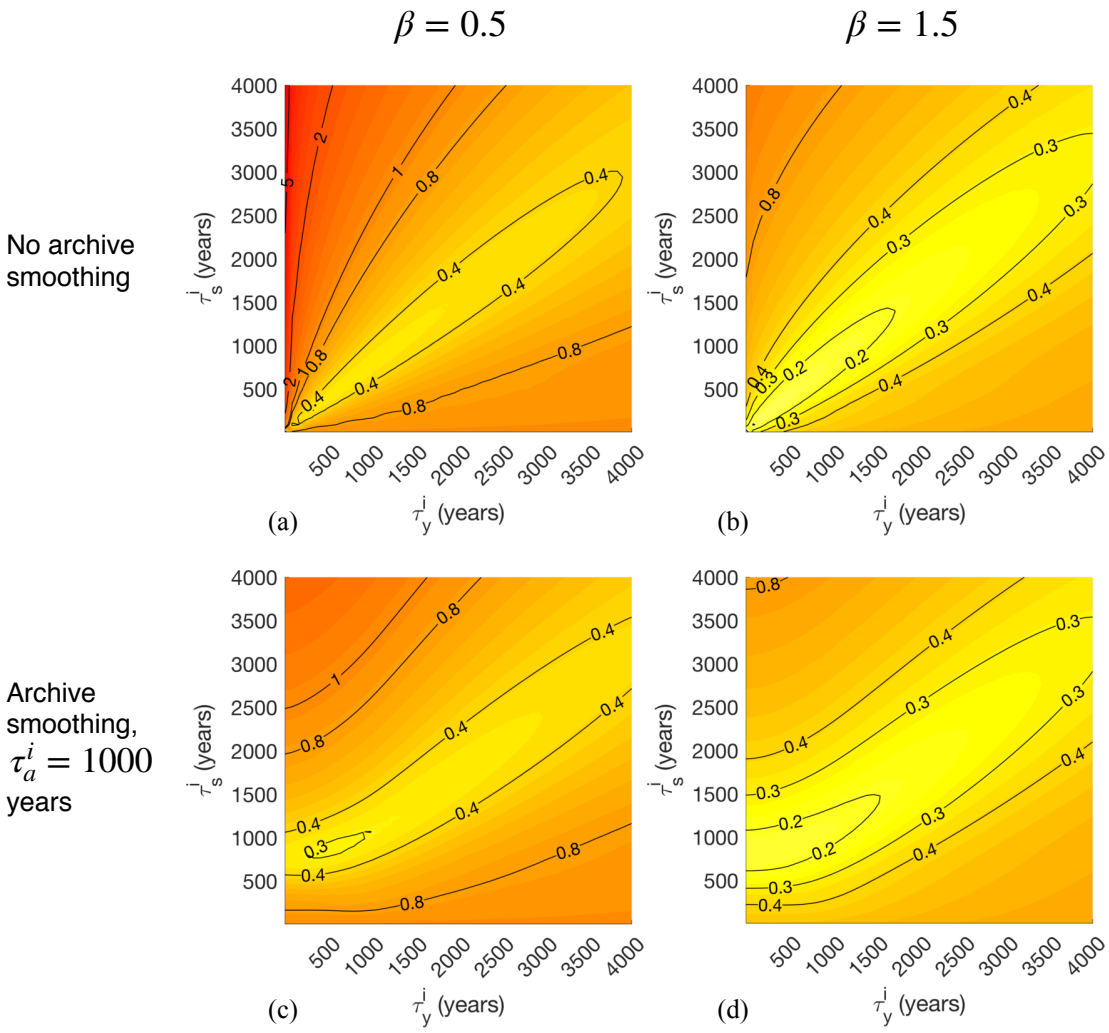

Figure 8. Same as Figure 5, but illustrating the dependence of the error-to-signal standard deviation ratio for individual measurements in a time series as a function of local time series spacing (τ_s^i) and the observational averaging time interval τ_y^i.

paleoclimate time series, it is important to bear in mind not just the Nyquist frequency but the role of sampling and smoothing time scales as anti-aliasing filters; these considerations point to dense sampling (i.e., without space between contiguous samples) in order to minimize error in the absence of effects from archive smoothing (Section 6). However, many practical considerations motivate paleoclimate sampling strategies, and may outweigh the concerns described here. For instance, records sampled densely cannot be used as a starting point for subsequently constructing higher-resolution records. Moreover, preservation of natural archives for subsequent analyses is important for reproducibility and for sharing resources between laboratories, and may be complicated by continuous sampling.

To some extent, the simple model for TR error can be generalized to more complex scenarios. If samples are nonuniform in time – for instance, due to large changes in chronology, or because material was sampled using a syringe or drill bit with a circular projection onto an archive – then the sinc function in (3) can be replaced by Fourier transforms of the relevant function. Similarly, a more complex pattern of archive smoothing can be accommodated by substituting a different smoothing kernel. Non-Gaussian age uncertainties can be incorporated by substituting a different distribution in (7). Changes in sampling properties through time (as might arise from non-constant chronologies or sampling procedures) can readily be accommodated because all computations are performed on a point-by-point basis. If sampling or smoothing time scales are unknown, a similar procedure can be adopted as was used for Δ in (7), whereby a second integration is performed to compute the expectation over an estimated probability distribution of one or more time scales.

Several caveats apply to the uncertainty estimates given. First, the model neglects some effects that may be important, such as inhomogeneities in preserved climate signals owing to e.g. diagenesis or scarcity of biological fossils. Second, nonstationarity in record spectra leads to time variations in errors, as illustrated in Figure 3. Third, in the analysis of time series errors, we ignore the possibility that errors covary in time, which can result from chronologies constructed by interpolating ages between tie points; more complete characterizations could be achieved by Monte Carlo sampling of age model uncertainty (Anchukaitis and Tierney, 2013). More broadly, there is clear need for comprehensive approaches in uncertainty quantification that can reveal interactions among the various sources of uncertainty in paleoclimate records. Forward proxy system models (e.g., Evans et al., 2013; Dee et al., 2015; Dolman and Laepple, 2018) are a promising way forward to assess uncertainties holistically.

Results for time series (Section 6) hold when record spacing and chronologies do not change too rapidly and where the goal is to obtain a discrete representation of a continuous process. For other objectives, other sampling procedures may be preferred. For instance, "burst sampling," whereby rapid sequences of observations are taken at relatively long intervals, is used in modern oceanographic procedures to estimate spectral nonstationarities in time (Emery and Thomson, 2014), and unevenly spaced paleoclimate observations can be leveraged to give a range of frequency information using variogram approaches (Amrhein et al., 2015) or the Lomb-Scargle periodogram (e.g., Schulz and Stattegger, 1997).

Representativity errors due to aliasing are not limited to the time domain, and similar procedures may be useful for quantifying errors due to spatial representativeness by considering how well proxy records can constrain the regional and larger scales typically of interest in paleoclimatology. An analogous problem is addressed in the modern ocean by Forget and Wunsch (2007), and Zhao et al. (2018) considered spatial representativeness in choosing how to weight deglacial radiocarbon time series in spatial bin averages. A challenge of any such approach is that the spatial averaging functions (analogous to our τ_y,

but occupying three spatial dimensions) represented by proxy records are often not well known; Van Sebille et al. (2015), for instance, explores how ocean advection determines three-dimensional patterns represented by sediment core observations. Because spatial patterns and time scales of ocean and climate variability are linked, it may ultimately be necessary to consider the full, four-dimensional spatiotemporal aliasing problem.

The hope is that these procedures may prove useful for first-order practical uncertainty quantification. A challenge is estimating the signal spectrum $|\hat{r}|^2$, which itself can be affected by aliasing (Kirchner, 2005). One approach is to use spectra from other records that are more highly-resolved or were sampled densely, e.g. from a sediment core at an adjacent site, or a record believed to record similar climate variability. Alternately, measurements of archive properties that can be made cheaply and at high resolution – such as magnetic susceptibility, wet bulk density, and other proxy properties that are routinely made on sediment cores – could prove useful for estimating $|\hat{r}|^2$ if those properties are related linearly to $r(t)$ (Herbert and Mayer, 1991; Wunsch and Gunn, 2003). Another challenge is that time scales that we have shown affect errors are often not published alongside paleoclimate datasets, thus turning systematic errors (where parameters like τ_y are known) into stochastic errors because a range of possible values must be explored. Publishing all available information about sampling practices, age model construction, and assessments of archive smoothing will greatly aid uncertainty quantification efforts.

Appendix A: Expressing time-mean temporal representativeness errors in the frequency domain

This appendix describes an analytical approach for estimating temporal representativity errors in the context of estimating time means. These errors have a compact representation in the frequency domain that rationalizes interactions between sampling procedures, time uncertainty, and signal spectra in contributing to errors. Fore more on the theorems and properties of Fourier analysis that are referenced see e.g. Bracewell (1986).

A1 Derivation

Define a mean value $m(t,\tau)$ of a climate variable $r(t)$ as a function of the duration τ and the time t on which that duration is centered,

$$m(t,\tau) = \int_{-\infty}^{\infty} \Pi\left(t',\tau\right) r\left(t+t'\right) dt', \tag{A1}$$

where $\Pi(t,\tau)$ is a normalized "boxcar" function centered on $t=0$ with width τ,

$$\Pi(t,\tau) = \begin{cases} 1/\tau & |t| \le \tau/2 \\ 0 & |t| > \tau/2. \end{cases} \tag{A2}$$

The operation in (A1) defines a moving average $m(t,\tau)$ and is known as a convolution, hereafter denoted as a star,

$$m(t,\tau) = \Pi(t,\tau) \star r(t). \tag{A3}$$

Then let the target quantity x be a mean of $r(t)$ over an interval of length τ_x centered on t, and an observation y to be an average over a different duration τ_y centered on a different time $t+\Delta$,

$$x = m(t, \tau_x) \tag{A4}$$

$$y = m(t+\Delta, \tau_y). \tag{A5}$$

The Fourier transform will be written both using the operator $\mathscr{F}$ and by a hat. Denoting frequency by ν, it is defined as

$$\mathscr{F}(x(t)) \equiv \hat{x}(\nu) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi i \nu t} dt.$$

Parseval's theorem states that the integral of a squared quantity in the time domain is equal to the integral of the squared amplitude of the Fourier transform of that quantity, so that after substituting (A5) we can write (2) as

$$\left\langle \theta^2 \right\rangle = \frac{1}{\tau_0} \int_{-\infty}^{\infty} \left(m(t, \tau_x) - m(t+\Delta t, \tau_y)\right)^2 dt \tag{A6}$$

$$= \frac{1}{\tau_0} \int_{0}^{\infty} \left| \mathscr{F}\left[m(t, \tau_x) - m(t+\Delta t, \tau_y)\right] \right|^2 d\nu. \tag{A7}$$

By the Fourier shift theorem,

$$\mathscr{F}\left[m(t+\Delta, \tau_y)\right] = e^{-2\pi i \nu \Delta} \mathscr{F}\left[m(t, \tau_y)\right]. \tag{A8}$$

Then, by the linearity of the Fourier transform,

$$\left\langle \theta^2 \right\rangle = \frac{1}{\tau_0} \int_{0}^{\infty} \left| \hat{m}(\nu, \tau_y) - e^{-2\pi i \nu \Delta} \hat{m}(\nu, \tau_x) \right|^2 d\nu. \tag{A9}$$

By the convolution theorem, convolution in the time domain is equivalent to multiplication in the frequency domain. Thus, the Fourier transform of a time mean as defined in (A3) is

$$\hat{m}(\nu, \tau) = \mathscr{F}\left[\Pi(t, \tau) \star r(t)\right] \tag{A10}$$

$$= \hat{\Pi}(\nu, \tau) \cdot \hat{r}(\nu). \tag{A11}$$

Substituting this relation into (A9) yields

$$\left\langle \theta^2 \right\rangle = \frac{1}{\tau_0} \int_{0}^{\infty} \left| \hat{\Pi}(\nu, \tau_x) - e^{-2\pi i \nu \Delta} \cdot \hat{\Pi}(\nu, \tau_y) \right|^2 |\hat{r}(\nu)|^2 d\nu. \tag{A12}$$

Finally, we represent smoothing prior to sampling by defining a new climate signal, $r'(t)$, that has had a running mean applied,

$$r'(t) = \Pi(t, \tau_a) \star r(t).$$

Substituting $\hat{r}'(\nu)$ into (A12) and applying the convolution theorem gives

$$\left\langle \theta^2 \right\rangle = \frac{1}{\tau_0} \int_{0}^{\infty} \left| \hat{\Pi}(\nu, \tau_x) - e^{-2\pi i \nu \Delta} \cdot \hat{\Pi}(\nu, \tau_a) \cdot \hat{\Pi}(\nu, \tau_y) \right|^2 |\hat{r}(\nu)|^2 d\nu. \tag{A13}$$

Numerical integration of (A13) is used in the text to illustrate dependencies of TR error on sampling parameters.

A2 Interpretation

The integrand of (A13) is the product of two components. The second, $|\hat{r}(\nu)|^2$, is the power spectral density of $r(t)$, which describes the variance contained at frequencies in $r(t)$. The first component is a power transfer function,

$$H(\nu, \tau_x, \tau_y, \tau_a, \Delta) = \left| \hat{\Pi}(\nu, \tau_x) - e^{-2\pi i \nu \Delta} \cdot \hat{\Pi}(\nu, \tau_a) \cdot \hat{\Pi}(\nu, \tau_y) \right|^2, \tag{A14}$$

which describes how power at different frequencies in $r(t)$ contributes to $\langle \theta^2 \rangle$. The Fourier transform of the boxcar function is a sinc function,

$$\hat{\Pi}(\nu, \tau) = \mathrm{sinc}(\tau\nu) = \frac{\sin(\pi\tau\nu)}{\pi\tau\nu}, \tag{A15}$$

which converges towards 1 at frequencies below $1/\tau$ and oscillates with decreasing amplitude about 0 at higher frequencies (Figure 2a).

When τ_x and τ_y are adequately separated so that the transfer function has a simple bandpass shape as seen in Figure X, the "cutoff frequencies" $\nu^\dagger_{low}$ and $\nu^\dagger_{high}$ are useful to diagnose how sampling procedures contribute to TR error. These are the frequencies on either side of the band at which the transfer function is equal to 0.5. In the presence of zero time offsets, the cutoff frequencies can be estimated by solving

$$H\left(\nu^\dagger_{low}\right) \approx \left| \mathrm{sinc}^2\left(\tau_x \nu^\dagger_{low}\right) - 1 \right|^2 = \tfrac{1}{2} \tag{A16}$$

$$H\left(\nu^\dagger_{high}\right) \approx \left| \mathrm{sinc}\left(\tau_y \nu^\dagger_{high}\right) \right|^2 = \tfrac{1}{2}. \tag{A17}$$

which yields $\nu^\dagger_{low} = 0.755\tau_x^{-1}$ and $\nu^\dagger_{high} = 0.443\tau_y^{-1}$. (In the case where τ_x is less than τ_y, then $\nu^\dagger_{low} = 0.755\tau_y^{-1}$ and $\nu^\dagger_{high} = 0.443\tau_x^{-1}$).

We can expect the presence of archive smoothing might reduce errors originating from high frequencies in $r(t)$, thereby reducing $\nu^\dagger_{high}$ and narrowing the band of aliased frequencies. In the presence of archive smoothing, the expression for $\nu^\dagger_{high}$ becomes

$$H\left(\nu^\dagger_{high}\right) = \left| \mathrm{sinc}\left(\tau_a \nu^\dagger_{high}\right) \mathrm{sinc}\left(\tau_y \nu^\dagger_{high}\right) \right|^2 = \frac{1}{2}. \tag{A18}$$

An approximate solution using a Taylor series representation is

$$\nu^\dagger_{high} \approx \frac{0.443}{\sqrt{\tau_a^2 + \tau_y^2}}, \tag{A19}$$

which illustrates a combined effect from sampling and archive smoothing for determining which frequencies contribute to TR errors. Thus when τ_y and τ_a are small relative to τ_x, archive smoothing reduces TR errors, consistent with numerical integrations (comparing Figures 5a and 5b with 5c and 5d).

Using (A19), we can estimate an ideal sampling interval $\tilde{\tau}_y$ in the presence of archive smoothing by minimizing the width of the frequency band that contributes to TR error. Setting $0.443\tilde{\tau}_x^{-1}$ (i.e., the cutoff frequency in the case where the combined

averaging effect of sampling and smoothing gave the desired averaging interval τ_x) equal to $0.443(\tau_y^2 + \tau_a^2)^{-\frac{1}{2}}$ and solving yields

$$\tilde{\tau}_y = \sqrt{\tau_x^2 - \tau_a^2} \text{ for } \tau_x > \tau_a. \tag{A20}$$

Numerical experiments (see dotted lines in all panels of Figure 5) support the robustness of this approximation for two different signal spectra.

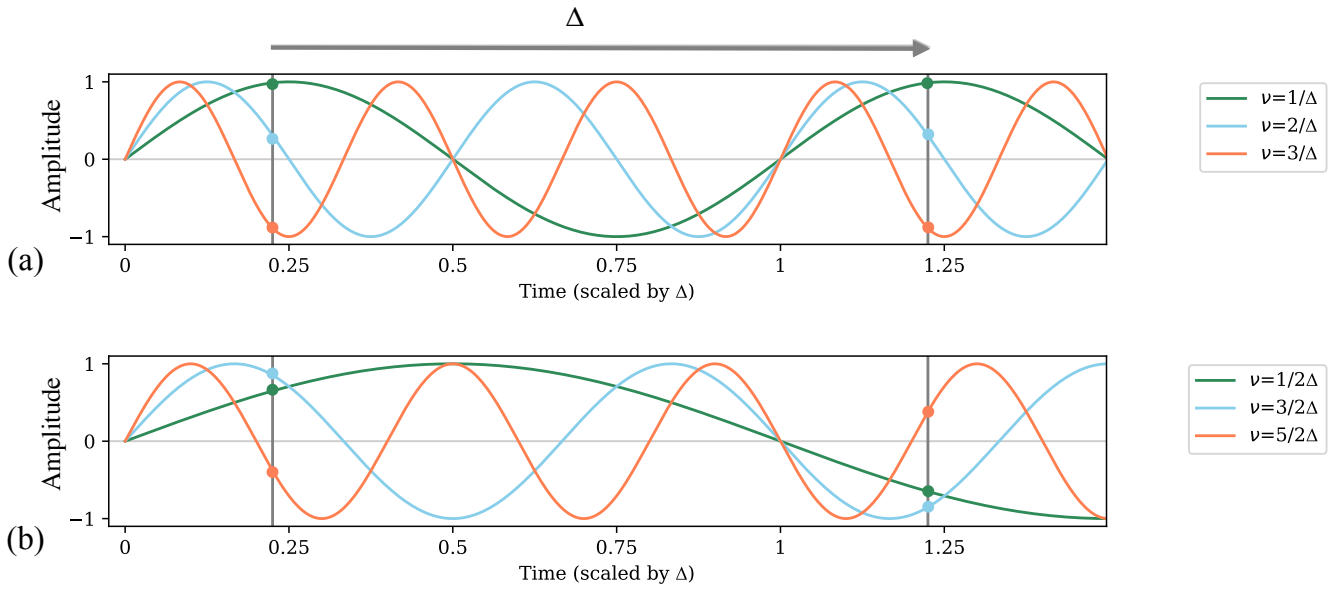

Figure A1. Illustration of the frequency dependence of errors in representing an instantaneous measurement of a process $r(t)$ at a time t by another measurement $r(t+\Delta)$. Each line represents a different frequency component of $r(t)$, grey vertical lines represent sampling times, and colored circles represent values of components at those times. At frequencies $\nu = \frac{n}{\Delta}$ for $n = 0, 1, 2, \ldots$, (a), the Fourier components of $x(t)$ will be exactly in phase when sampled at a time lag Δ, so these components do not contribute to the error variance $\left\langle (r(t) - r(t+\Delta))^2 \right\rangle$. By contrast, at frequencies $\nu = \frac{n}{\Delta} + \frac{1}{2\Delta}$ (b), the Fourier components are exactly out of phase, so these components tend to contribute most to the error variance. At intermediate frequencies, contributions lie between the two extremes, leading to a cosine function of error contribution as a function of frequency (Equation A22).

To study the error contribution from a time offset Δ, consider the limit where τ_x, τ_y, and τ_a approach zero, corresponding to instantaneous observations in time, so that $\langle\theta^2\rangle$ approaches

$$\langle\theta^2\rangle = \frac{1}{\tau_0}\int_0^\infty \left|1-e^{-2\pi i\nu\Delta}\right|^2 |\hat{r}(\nu)|^2 d\nu. \tag{A21}$$

Expanding $\left|1-e^{-2\pi i\nu\Delta}\right|^2$ and simplifying gives

$$\langle\theta^2\rangle = \frac{1}{\tau_0}\int_0^\infty (2-2\cos(2\pi\nu\Delta))\,|\hat{r}(\nu)|^2 d\nu \tag{A22}$$

so that the power transfer function is $H = 2-2\cos(2\pi\nu\Delta)$ and the expected error due to Δ is a cosinusoidally-weighted function of the signal power spectrum. H takes a minimum value of 0 at frequencies

$$\nu_{min} = 0, \frac{1}{\Delta}, \frac{2}{\Delta}, \ldots \frac{n}{\Delta}$$

for integer values of n; at these frequencies, measurements spaced by Δ in time are in phase and are therefore exactly correlated (Figure A1a). The weights take a maximum value of 4 at frequencies

$$\nu_{max} = \frac{1}{2\Delta}, \frac{3}{2\Delta}, \frac{5}{2\Delta}, \ldots \frac{n}{\Delta} + \frac{1}{2\Delta}$$

where measurements separated by Δ are always exactly out of phase (Figure A1b). At those frequencies, the underlying signal $r(t)$ is projected twofold onto the error, so that its variance contribution is multiplied fourfold. These variations in frequency contributions modulate effects from smoothing and sampling timescales (Figure 2b).

Appendix B: Expressing time series temporal representativeness errors in the frequency domain

This appendix extends the analytical approach for estimating temporal representativity errors from estimating time means to time series. Define the moving average time series associated that would result if all of $r(t)$ were sampled as the i^{th} observation y^i to be

$$y^i(t) = \Pi\left(t,\tau_y^i\right) \star \Pi\left(t,\tau_a^i\right) \star r(t) \tag{B1}$$

where we have included a contribution from archive smoothing, so that its Fourier transform is

$$\hat{y}^i(\nu) = \hat{\Pi}\left(\nu,\tau_y^i\right)\cdot\hat{\Pi}\left(\nu,\tau_a^i\right)\cdot\hat{r}(\nu). \tag{B2}$$

By Shannon's sampling theorem, an accurate discrete representation of $r(t)$ results from sampling all frequencies in $r(t)$ less than or equal to the local Nyquist frequency $\nu_{Nyq}^i = 1/\left(2\tau_s^i\right)$. As such, the target value x^i for the i^{th} measurement y^i is the value of $r(t)$ sampled at t^i after filtering $r(t)$ to remove high-frequency variability. The Fourier transform of a time series of values of x^i is

$$\hat{x}^i(\nu) = G\left(\nu,\tau_s^i\right)\hat{r}(\nu) \tag{B3}$$

where the "ideal" transfer function $G(\nu, \tau_s)$ is the Heaviside function

$$G(\nu, \tau_s) = \begin{cases} 1 & \nu < 1/(2\tau_s^i) \\ 0 & \nu \geq 1/\left(2\tau_s^i\right) \end{cases} \tag{B4}$$

that is ideal in the sense that it eliminates variability at frequencies greater than $\nu_{Nyq}^i = 1/\left(2\tau_s^i\right)$. Then we define TR error at the i^{th} measurement to be

$$\theta^i = x^i - y^i. \tag{B5}$$

As in the previous section, we estimate the variance of θ^i by taking the expected value as if the entire record had been sampled using the local values τ_s^i and τ_y^i. Then, equivalent to (A13),

$$\left\langle \theta^{i2} \right\rangle = \frac{1}{\tau_0} \int_0^\infty \left| G\left(\nu, \tau_s^i\right) - \hat{\Pi}\left(\nu, \tau_a^i\right) \cdot \hat{\Pi}\left(\nu, \tau_y^i\right) \right|^2 |\hat{r}(\nu)|^2 \, d\nu. \tag{B6}$$

Acknowledgements. Thanks to LuAnne Thompson, Greg Hakim, Lloyd Keigwin, Cristi Proistecescu, Carl Wunsch, and Thomas Laepple for useful conversations. Comments from two anonymous reviewers in a previous round of reviews helped improve the manuscript. Support came from NOAA grant NA14OAR4310176 and an NSF postdoctoral fellowship. Wavelet software was provided by C. Torrence and G. Compo, and is available at URL: http://paos.colorado.edu/research/wavelets/.

References

Adkins, J. F., Boyle, E. A., Curry, W. B., and Lutringer, A.: Stable isotopes in deep-sea corals and a new mechanism for "vital effects", Geochimica et Cosmochimica Acta, 67, 1129–1143, https://doi.org/10.1016/S0016-7037(00)01203-6, 2003.

Amrhein, D. E., Gebbie, G., Marchal, O., and Wunsch, C.: Inferring surface water equilibrium calcite δ 18 O during the last deglacial period from benthic foraminiferal records: Implications for ocean circulation, Paleoceanography, 30, 1470–1489, 2015.

Amrhein, D. E., Wunsch, C., Marchal, O., and Forget, G.: A Global Glacial Ocean State Estimate Constrained by Upper-Ocean Temperature Proxies, Journal of Climate, 31, 8059–8079, 2018.

Anchukaitis, K. J. and Tierney, J. E.: Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records, Climate Dynamics, 41, 1291–1306, https://doi.org/10.1007/s00382-012-1483-0, 2013.

Andersen, K. K., Azuma, N., Barnola, J.-M. M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H., and Others: {H}igh-resolution record of {N}orthern {H}emisphere climate extending into the last interglacial period, Nature, 431, 147, 2004.

Anderson, D. M.: Attenuation of millennial-scale events by bioturbation in marine sediments, Paleoceanography, 16, 352–357, 2001.

Beer, J., McCracken, K., and Von Steiger, R.: {C}osmogenic radionuclides: theory and applications in the terrestrial and space environments, Springer Science & Business Media, 2012.

Bracewell, R. N.: The Fourier transform and its applications, vol. 31999, McGraw-Hill New York, 1986.

Bronk Ramsey, C.: Bayesian Analysis of Radiocarbon Dates, Radiocarbon, 51, 337–360, https://doi.org/10.1017/S0033822200033865, https://www.cambridge.org/core/product/identifier/S0033822200033865/type/journal{_}article, 2009.

Buck, C. E.: Bayesian Chronological Data Interpretation: Where Now?, LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-, pp. 1–24, 2004.

Buck, C. E. and Millard, A.: Tools for constructing chronologies: crossing disciplinary boundaries, Springer Verlag, 2004.

Caley, T., Roche, D. M., Waelbroeck, C., and Michel, E.: {O}xygen stable isotopes during the {L}ast {G}lacial {M}aximum climate: perspectives from data–model (i{LOVECLIM}) comparison, Climate of the Past, 10, 1939–1955, 2014.

Clark, P. U., Shakun, J. D., Baker, P. A., Bartlein, P. J., Brewer, S., Brook, E., Carlson, A. E., Cheng, H., Kaufman, D. S., Liu, Z., and Others: {G}lobal climate evolution during the last deglaciation, Proceedings of the National Academy of Sciences, 109, E1134—-E1142, 2012.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220–1247, https://doi.org/10.1002/2015MS000447, http://doi.wiley.com/10.1002/2015MS000447, 2015.

Dolman, A. M. and Laepple, T.: Sedproxy: a forward model for sediment archived climate proxies, Climate of the Past Discussions, pp. 1–31, https://doi.org/10.5194/cp-2018-13, https://www.clim-past-discuss.net/cp-2018-13/, 2018.

Elderfield, H., Vautravers, M., and Cooper, M.: The relationship between shell size and Mg/Ca, Sr/Ca, δ 18O, and δ13C of species of planktonic foraminifera, Geochemistry, Geophysics, Geosystems, 3, 1–13, https://doi.org/10.1029/2001GC000194, http://doi.wiley.com/10.1029/2001GC000194, 2002.

Emery, W. J. and Thomson, R. E.: {D}ata analysis methods in physical oceanography, Newnes, 2014.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: {A}pplications of proxy system modeling in high resolution paleoclimatology, Quaternary Science Reviews, 76, 16–28, 2013.

Fairchild, I. J., Smith, C. L., Baker, A., Fuller, L., Spötl, C., Mattey, D., and McDermott, F.: Modification and preservation of environmental signals in speleothems, Earth-Science Reviews, 75, 105–153, https://doi.org/10.1016/j.earscirev.2005.08.003, 2006.

Forget, G. and Wunsch, C.: {E}stimated global hydrographic variability, Journal of Physical Oceanography, 37, 1997–2008, 2007.

Haam, E. and Huybers, P.: {A} test for the presence of covariance between time-uncertain series of data with application to the {D}ongge {C}ave speleothem and atmospheric radiocarbon records, Paleoceanography, 25, 2010.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research, 121, 6745–6764, https://doi.org/10.1002/2016JD024751, 2016.

Herbert, T. D. and Mayer, L. A.: {L}ong climatic time series from sediment physical property measurements, Journal of Sedimentary Research, 61, 1991.

Huybers, P. and Curry, W.: {L}inks between annual, {M}ilankovitch and continuum temperature variability, Nature, 441, 329–332, 2006.

Huybers, P. and Wunsch, C.: A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change, Paleoceanography, 19, PA1028, 2004.

Jones, R. H.: {A}liasing with unequally spaced observations, Journal of Applied Meteorology, 11, 245–254, https://doi.org/10.1175/1520-0450(1972)011<0245:AWUSO>2.0.CO;2, 1972.

Kirchner, J. W.: Aliasing in 1/f(alpha) noise spectra: origins, consequences, and remedies., Physical review. E, Statistical, nonlinear, and soft matter physics, 71, 066 110, https://doi.org/10.1103/PhysRevE.71.066110, https://link.aps.org/doi/10.1103/PhysRevE.71.066110http://www.ncbi.nlm.nih.gov/pubmed/16089823, 2005.

MARGO Project Members: {C}onstraints on the magnitude and patterns of ocean cooling at the {L}ast {G}lacial {M}aximum, Nature Geoscience, 2, 127–132, 2009.

McGee, D., Winckler, G., Stuut, J. B. W., Bradtmiller, L. I., and Others: {T}he magnitude, timing and abruptness of changes in {N}orth {A}frican dust deposition over the last 20,000 yr, Earth and Planetary Science Letters, 371, 163–176, 2013.

Moore, M. I. and Thomson, P. J.: Impact of jittered sampling on conventional spectral estimates, Journal of Geophysical Research, 96, 18 519, https://doi.org/10.1029/91JC01623, 1991.

Pisias, N. G. and Mix, A. C.: {A}liasing of the geologic record and the search for long-period {M}ilankovitch cycles, Paleoceanography, 3, 613–619, 1988.

Rhines, A. and Huybers, P.: Estimation of spectral power laws in time uncertain series of data with application to the Greenland ice sheet project 2 δ¹⁸O record, Journal of Geophysical Research Atmospheres, 116, 1–9, https://doi.org/10.1029/2010JD014764, 2011.

Schulz, M. and Stattegger, K.: SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series, Computers and Geosciences, 23, 929–945, https://doi.org/10.1016/S0098-3004(97)00087-3, 1997.

Shannon, C. E.: {C}ommunication in the presence of noise, Proceedings of the IRE, 37, 10–21, 1949.

Telford, R. J., Heegaard, E., and Birks, H. J. B.: The intercept is a poor estimate of a calibrated radiocarbon age, Holocene, 14, 296–298, https://doi.org/10.1191/0959683604hl707fa, 2004.

Tierney, J. E. and Tingley, M. P.: A Bayesian, spatially-varying calibration model for the TEX86proxy, Geochimica et Cosmochimica Acta, 127, 83–106, https://doi.org/10.1016/j.gca.2013.11.026, http://dx.doi.org/10.1016/j.gca.2013.11.026, 2014.

Van Sebille, E., Scussolini, P., Durgadoo, J. V., Peeters, F. J., Biastoch, A., Weijer, W., Turney, C., Paris, C. B., and Zahn, R.: Ocean currents generate large footprints in marine palaeoclimate proxies, Nature Communications, 6, 1–8, https://doi.org/10.1038/ncomms7521, http://dx.doi.org/10.1038/ncomms7521, 2015.

von Albedyll, L., Opel, T., Fritzsche, D., Merchel, S., Laepple, T., and Rugel, G.: 10 {B}e in the {A}kademii {N}auk ice core–first results for {CE} 1590–1950 and implications for future chronology validation, Journal of Glaciology, pp. 1–9, 2017.

Wunsch, C.: {T}he {N}orth {A}tlantic general circulation west of 50˜{W} determined by inverse methods, Reviews of Geophysics, 16, 583–620, 1978.

Wunsch, C.: 0n sharp spectral lines in the climate record and the millennial peak, Paleoceanography, 15, 417–424, 2000.

Wunsch, C.: {Greenland}–{A}ntarctic phase relations and millennial time-scale climate fluctuations in the {Greenland} ice-cores, Quaternary Science Reviews, 22, 1631–1646, 2003.

Wunsch, C. and Gunn, D. E.: {A} densely sampled core and climate variable aliasing, Geo-Marine Letters, 23, 64–71, 2003.

Zhao, N., Marchal, O., Keigwin, L., Amrhein, D., and Gebbie, G.: A Synthesis of Deglacial Deep-Sea Radiocarbon Records and Their (In)Consistency With Modern Ocean Ventilation, Paleoceanography and Paleoclimatology, 33, 128–151, 2018.