Comparing estimation techniques for ~~timescale-dependent~~ temporal scaling ~~of climate variability~~ in paleoclimate ~~time series~~timeseries

Raphaël Hébert [1,2], Kira Rehfeld [3], and Thomas Laepple [1,4]

[1]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Telegrafenberg A45, 14473 Potsdam, Germany
[2]Institute of Geosciences, University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany
[3]Institut für Umweltphysik, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany
[4]University of Bremen, MARUM – Center for Marine Environmental Sciences and Faculty of Geosciences, 28334 Bremen, Germany

Correspondence: Raphaël Hébert, raphael.hebert@awi.de

Abstract. Characterizing the variability across timescales is important to understand the underlying dynamics of the Earth system. It remains challenging to do so from paleoclimate archives since they are more than often irregular and traditional methods to produce timescale-dependent estimates of variability such as the classical periodogram and the multitaper spectrum generally require regular time sampling. We have compared those traditional methods using interpolation with interpolation-free methods, namely the Lomb-Scargle periodogram and the first-order Haar structure function. The ability of those methods to produce timescale-dependent estimates of variability when applied to irregular data was evaluated in a comparative framework using surrogate paleo-proxy data generated with realistic sampling. The metric we chose to compare them is the scaling exponent, i.e. the linear slope in log-transformed coordinates, since it summarizes the behaviour of the variability across timescales. We found that for scaling estimates in irregular timeseries, the interpolation-free methods are to be preferred over the methods requiring interpolation as they allow for the utilization of the information from shorter timescale which are particularly affected by the irregularity. In addition, our results suggest that the Haar structure function is the safer choice of interpolation-free method since the Lomb-Scargle periodogram is unreliable when the underlying process generating the timeseries is not stationary. Given that we cannot know a priori what kind of scaling behaviour is contained in a paleoclimate timeseries, and that it is also possible that this changes as a function of timescale, it is a desirable characteristic for the method to handle both stationary and non-stationary cases alike.

1 Introduction

Paleoclimate archives are crucial to improve our understanding of climate variability on decadal to multi-centennial timescales and beyond (Braconnot et al., 2012). They provide independent data for the evaluation of the climate models which are used to project future climate change. To this end, quantitative estimates of variability based on palaeoclimate records allow to compare past changes in variance over a wide range of timescales (e.g. Laepple and Huybers, 2014b, a; Rehfeld et al., 2018;

Zhu et al., 2019) and how variance is distributed across different timescales (e.g. Mitchell et al., 1976; Huybers and Curry, 2006; Lovejoy, 2015; Nilsen et al., 2016; Shao and Ditlevsen, 2016).

We generally refer to scaling as the statistical characterization of the changes in climate variability as a function of timescale τ, or equivalently as a function of frequency f such that $f = \tau^{-1}$. The term scaling often implies, although not necessarily,

25 power-law scaling: a stochastic process is said to be power-law scaling over a range of timescales $[\tau_1, \tau_2]$ if a ~~time-scale~~ timescale dependent statistical metric $S(\tau)$ approximately follows a power-law relationship such that $S(\tau) \propto \tau^a$, where a is a general power-law scaling exponent. Therefore, ~~processes that fulfill this property show a log-linear relationship in~~ the power spectrum ~~(Schuster, 1898; Percival and Walden, 1993)~~ of such processes will appear linear on a log-log plot over a given range of timescales (Percival and Walden, 1993). The corresponding power-law scaling exponent can be informative of

30 the underlying dynamics of the system, such as the degree of temporal auto-correlation, i.e. the system's memory (Mandelbrot and Wallis, 1968; Lovejoy et al., 2015; Graves et al., 2017; Fredriksen and Rypdal, 2015; Del Rio Amador and Lovejoy, 2019). While assuming power-law scaling may be a simplification, it is a rather accurate first-order description for a vast range of geophysical processes (Cannon and Mandelbrot, 1984; Pelletier and Turcotte, 1999; Malamud and Turcotte, 1999; Fedi, 2016; Corral and González, 2019).

35 Methods used for scaling analysis generally assume that the process under investigation has been sampled at regular time steps. This is appropriate for some instrumental observations and annually-resolved paleoclimate archives such as tree- and ~~coral- rings~~coral-rings. However, since most paleoclimate archives are the product of slow and intermittent accumulation in sediments or ice sheets, sampling them at regular depth intervals translates to irregular time intervals (Bradley, 2015). In addition, the irregular accumulation process usually has to be reconstructed and necessarily introduces age-uncertainty (Rehfeld

40 and Kurths, 2014), although it does not affect the scaling estimates strongly (Rhines and Huybers, 2011).

Therefore, the primary challenge is that scaling analysis methods need to be adapted for the analysis of sparse and irregular series. Two approaches to this problem can be distinguished.

Firstly, an interpolation routine can be employed prior to the analysis in order to regularize the series. Once the series are regular, traditional methods such as the classical periodogram ~~(CPG~~(CPG; Schuster, 1898), see Table 1 for a list of all the

45 acronyms used in this paper) or the multitaper spectrum method (MTM; Thomson, 1982) can be used~~(See this approach used in a paleoclim~~. See this approach used in a paleoclimatology context in Huybers and Curry (2006), Laepple and Huybers (2014a, b) and Rehfeld et al. (2018), and an implementation of these methods in R for regular climate data, including functions for statistical testing, scaling exponent estimation and trend estimations for different residual models, provided by Vyushin et al. (2009).

Secondly, the estimator can be adjusted for arbitrary sampling times. The so-called Lomb-Scargle Periodogram ~~(LSP; Lomb, 1976; Scargl~~

50 (LSP; Lomb, 1976; Scargle, 1982; Horne and Baliunas, 1986; VanderPlas, 2018) was developed in the field of astronomy to identify periodic components in noisy astronomical timeseries with sampling hiatus and was often used to analyze Laser Doppler Velocimetry experiments which produce irregularly sampled velocity data (Benedict et al., 2000; Munteanu et al., 2016; Damaschke as well as for the detection of biomedical rhythms (Schimmel, 2001). The LSP has sometimes been used in paleoclimatological context (Schulz and Stattegger, 1997; Trauth, 2020), although ~~the LSP~~ it may introduce additional bias and variance (Schulz

55 and Stattegger, 1997; Schulz and Mudelsee, 2002; Rehfeld et al., 2011). More recently, Lovejoy and Schertzer (2012) advo-

Table 1. Table of acronyms used in this paper and references.

Acronym	Long Name	Key Reference
CPG	Classical Periodogram	Schuster (1898)
LSP	Lomb-Scargle Periodogram	Lomb (1976); Scargle (1982)
MTM	Multitaper Spectrum Method	Thomson (1982)
HSF	first-order Haar Structure Function	Lovejoy and Schertzer (2012)
DFA	Detrended Fluctuations Analysis	Peng et al. (1995)
LGM	Last Glacial Maximum	CLIMAP Project Members (1976)
fGn	fractional Gaussian noise	Mandelbrot and Van Ness (1968)
fBm	fractional Brownian motion	Mandelbrot and Van Ness (1968)

cated for the use of the Haar Structure Function (HSF), based on Haar wavelets (Haar, 1910), in geophysics due to its ease of interpretation and accuracy. Incidentally, the HSP can easily be adapted for irregular timeseries.

Another method which is often used for scaling analysis is the detrended fluctuations analysis ~~(Peng et al., 1995; Kantelhardt et al., 2002,~~ (DFA Peng et al., 1995; Kantelhardt et al., 2001, 2002) which has also been applied to climatic and paleoclimatic timeseries

60 (Koscielny-Bunde et al., 1996; Rybski et al., 2006; Shao and Ditlevsen, 2016). However, a pre-study showed it was less efficient for irregular timeseries and we decided to omit it for clarity. In addition, it underestimates variance at any given timescale because of the necessary detrending (Nilsen et al., 2016) and it is therefore of limited use beyond the estimation of scaling exponents.

In the present work, we compare the different methods for the scaling of variability and make them accessible in a single

65 software package. Our main aim is to assess their ability to estimate variability across timescales in a paleoclimatological context which often entails scarce and irregular timeseries. In order to benchmark the methods, we evaluate them on surrogate data with known properties similar to those of palaeoclimate records without abrupt transitions. Finally, we apply the methods to a database comprising ~~LGM~~ Last Glacial Maximum (LGM) and Holocene records to evaluate their performances on real data.

70 **2   Data and Methods**

In this section, we outline the different methods considered to evaluate scaling, and how they can be compared. We also provide a method to generate "paleo-proxy" surrogate data with realistic variability and sampling, which is then used to test and compare the scaling methods.

2.1 Scaling estimation methods

75 Scaling generally refers to the behaviour of a quantity $S(\tau)$ as a function of scale τ (or frequency f such that $f = \tau^{-1}$) for a given process $X(t)$. In the current work, we exclusively consider timeseries, but the same methods can be used to investigate spatial scaling relationships. The quantity $S(\tau)$ considered can be the statistical moments of an appropriately defined fluctuation ΔX such as the power spectral density or Haar fluctuations. It is usual to assume such form to define a structure function for the statistical moments (Schertzer and Lovejoy, 1987; Lovejoy and Schertzer, 2012) of the process under investigation:

80
$$S_{\Delta X, q}(X, \tau) = \underline{<} \langle \Delta X_{j,\tau}^{q} \underline{>} \rangle \propto \tau^{\xi(q)} \tag{1}$$

where "$\underline{<...>}\langle...\rangle$" denotes ensemble averaging over all fluctuations j available at the given scale τ, q is the statistical moment (i.e. $q = 1$ corresponds to the mean, $q = 2$ to the variance, $q = 3$ to the skewness and so on), and $\xi(q) = qH - K(q)$ is the exponent function where H is the fluctuation scaling exponent and $K(q)$ is the moment scaling function. $K(q)$ is zero for Gaussian processes and thus, for this specific case, all statistical moments scale similarly, i.e. $\xi(q) \propto qH$.

85   The equivalent metric for the power spectrum method is the spectral scaling exponent β (see below section 2.1.1 for a formal definition). The two scaling exponents can be related by the following:

$$\beta = 1 + 2H - K(2) \tag{2}$$

where $K(q)$ at $q = 2$ is used since the spectrum is a second-order statistic. This relation can be understood intuitively since β describes the scaling of the power spectral density obtained via the Fourier transform of the auto-covariance function whereas 90 H describes the scaling of the real space fluctuations. Therefore, since the auto-covariance is proportional to the expectation value of the (zero-mean) timeseries squared, the exponent H is multiplied by two, while the integration in the definition of the Fourier transform increases the scaling exponent by $+1$.

In this work, we will focus on the quasi-Gaussian case, i.e. when statistics approximately follow the Gaussian distribution, in order to minimize the number of estimated parameters; paleoclimate archives often lack the resolution and/or length to 95 estimate many parameters with confidence. This assumption is rather well justified for temperature and precipitation timeseries at timescales longer than the turbulent weather regime (i.e. timescales longer than weeks or a few years depending on the location and medium (Lovejoy and Schertzer, 2012)). On the other hand, highly intermittent archives which clearly display multifractality, such as volcanic series (Lovejoy and Varotsos, 2016) or dust flux (Lovejoy and Lambert, 2019), would require the "intermittency correction" from the moment scaling function $K(q)$.

100   Under the quasi-Gaussian assumption, we can simplify equation 2 since this implies $K(2) \approx 0$ and therefore:

$$\beta \approx 1 + 2H \tag{3}$$

This allows to convert estimated $\hat{\beta}$ (where "$\hat{}$" denotes an estimator for the given quantity) into their equivalent estimated fluctuation scaling exponents $\hat{H}$. We use the fluctuation scaling exponent H for inter-comparison between the methods in this study. H is a natural choice as it describes directly the behaviour of fluctuations in real space and is the usual parameter

in functions to generate fractional noises (Mandelbrot and Van Ness, 1968; Mandelbrot, 1971; Molz et al., 1997). While the fluctuation exponent H takes its origin in the so-called Hurst exponent (Hurst, 1956), its meaning and definition has evolved and changed over time, see Graves et al. (2017) and Lovejoy et al. (2021) for historical summaries.

The process to estimate the scaling exponents can be divided into three steps. First, if the series is irregularly spaced, it requires to be regularized in order to be usable for the CPG and the MTM. The regularization is not necessary for the HSF and the LSP which have the advantage that they can be calculated directly on the irregular timeseries. Second, the fluctuations proper to each method are calculated as a function of timescale, and finally, the scaling exponent is fitted on the result.

2.1.1 Power Spectrum

For an ergodic, weakly stationary stochastic process X, the power spectrum P is given by the Fourier transform of the auto-covariance function γ, and equivalently the squared Fourier transform of the signal (Storch and Zwiers, 1984),

$$P(\tau) = |\mathfrak{F}\{X\}|^2 = \mathfrak{F}(\gamma\{X\}). \tag{4}$$

where τ is the timescale and $\mathfrak{F}$ denotes the Fourier transform operator. Equivalently, the power spectrum can be given as a function of the frequency $f = \tau^{-1}$ instead of the timescale τ. We choose to write it as a function of timescale to allow for visual comparison with the Haar method below, and also because it is more intuitive: a non-expert can easily grasp what the 1000-year timescale means rather than the equivalent 0.001-year^{-1} frequency.

Classical Periodogram

The power spectrum for a discrete process X with N timesteps at regular intervals can be estimated using the classical periodogram (Chatfield, 2013):

$$\hat{P}_C(\tau) = \frac{1}{\pi N} \left| \sum_{t=1}^{N} X_t e^{\frac{-2\pi i t}{\tau}} \right|^2 \tag{5}$$

If $\log(P(\tau))$ behaves linearly over a range of timescales τ, the timeseries is considered to be power-law scaling over this timescale band, i.e. $P(\tau) \approx A\tau^{\beta}$ for an arbitrary constant A.

Multitaper Spectrum

The MTM method improves upon the CPG by producing independent estimates using a set of orthogonal functions: the prolate spheroidal wave functions $h_{t,k}$ (Slepian and Pollak, 1961), also known as the Slepian tapers, which have the desirable property of minimizing spectral leakage (Thomson, 1982). The spectral individual estimates for the kth taper can be written in a form similar to the periodogram above:

$$\hat{P}_k(\tau) = \frac{1}{\pi N} \left| \sum_{t=1}^{N} h_{k,t} X_t e^{\frac{-2\pi i t}{\tau}} \right|^2 \tag{6}$$

The estimator can then be expressed as the mean of the K tapered estimates:

$$\hat{P}_{MT}(\tau) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{P}_k(\tau) \tag{7}$$

Interpolation

135 To produce the power spectrum with the methods above, it is necessary to interpolate the series at a regular resolution. Following Laepple and Huybers (2013), the data were first linearly interpolated to 10 times the mean resolution, then low-pass filtered using a finite response filter with a cut-off frequency of 1.2 divided by the target mean resolution in order to avoid aliasing. Linear interpolation corresponds to a convolution in the temporal domain with a triangular window. The Fourier transform of the triangular window is $sinc^2$, where the $sinc$ function is defined as $sinc = x^{-1} sin(x)$, and therefore, a linear interpolation

140 multiplies the power spectrum by $sinc^4$ modulated by the resolution of the interpolation (Smith, 2011), resulting in a power loss near the Nyquist frequency.

Lomb-Scargle Periodogram

As an alternative to the classical periodogram which requires regular data, Scargle (1982) introduced the LSP as a generalized form of the CPG (equation 5):

145 $$\hat{P}_{LS}(\tau) = \frac{A^2}{2} \left[\sum_{j}^{N} X(t_j) \cos \frac{2\pi(t_j - T)}{\tau} \right]^2 + \tag{8}$$

$$\frac{B^2}{2} \left[\sum_{j}^{N} X(t_j) \sin \frac{2\pi(t_j - T)}{\tau} \right]^2 \tag{9}$$

where A, B and T are arbitrary functions of timescale τ and sampling times t_j which can be irregular. We see that if $A = B = \sqrt{\frac{2}{N}}$ and $T = 0$, we would recover the classical periodogram in case of regular sampling, but the reduction is not unique and other choices of A and B can be made. The periodogram estimates are chi-squared distributed with 2 degrees of freedom

150 (i.e. an exponential distribution) and therefore, A and B will be chosen to retain this property. For independent and identically distributed white noise, it is the case when:

$$A(\tau, t_j) = \sqrt{\sum_{j}^{N} \cos^2 \frac{2\pi t_j}{\tau}} \tag{10}$$

$$B(\tau, t_j) = \sqrt{\sum_{j}^{N} \sin^2 \frac{2\pi \nu t_j}{\tau}} \tag{11}$$

The CPG is invariant to time translation since shifting the timesteps by a constant value only affects the phase of the complex exponential inside the absolute value. The function T is thus introduced for the generalized periodogram to retain that property, which is the case when:

$$T(\tau, t_j) = \frac{\tau}{4\pi} \tan^{-1}\left(\frac{\sum_j^N \sin 4\pi\tau^{-1}t_j}{\sum_j^N \cos 4\pi\tau^{-1}t_j}\right) \tag{12}$$

The LSP has been mostly used in astronomy to detect periodic components in noisy irregular data. As outlined above, the functions A and B are chosen following the assumption that the process X is approximately white noise with no temporal correlation. This assumption is of course problematic from the perspective of climate timeseries which usually exhibit long-range temporal correlation, but, as we will see later, good estimates of scaling exponents can nonetheless be recovered over a wide range of H.

2.1.2 Haar Structure Function

The HSF allows to perform the scaling analysis in real space, i.e. without performing a Fourier transform. When used to define a structure function, Haar wavelets are appropriate to estimate the fluctuation exponent of processes with $H \in (-1, 1)$ (Lovejoy and Schertzer, 2012), a range covering most geophysical processes, and therefore, also because of its readiness to be applied to irregular series and its ease of interpretation, Lovejoy and Schertzer (2012) argued it makes it a convenient choice for the scaling analysis of geophysical timeseries.

Haar fluctuations $\mathcal{H}$ are simple to implement for irregular sampling. They can be defined for a given interval of length τ as the difference between the mean of the first half of the interval with the second half:

$$\mathcal{H}_{\tau,j}(X(t_i)) = \frac{2}{\tau}\left|\sum_{t_j+\frac{\tau}{2}<t_i<t_j+\tau} X(t_i) - \sum_{t_j<t_i<t_j+\frac{\tau}{2}} X(t_i)\right| \tag{13}$$

The discrete sampling t_i does not have to be regular since the average of the available fluctuations in the intervals is taken. This is of course an approximation since the fluctuations available might not exactly correspond to the expected timescale.

The first-order HSF $S_{\mathcal{H},q=1}$ can now be defined by using the Haar fluctuation defined above in equation 1 and letting $q = 1$:

$$S_{\mathcal{H},q=1}(X(t),\tau) = \underline{\leq}\langle\mathcal{H}_{\tau,j}(X(t))\underline{\geq}\rangle \approx A\tau^H \tag{14}$$

Higher-order structure functions can be defined by considering other values of q and analyze multifractal processes, but since we are only dealing with the quasi-Gaussian case in this manuscript, $q = 1$ is sufficient. For simplicity, we refer to the first-order HSF simply as HSF.

Since we are taking the average value of the absolute of the fluctuations, if the fluctuations are distributed in a quasi-Gaussian fashion, then we expect the distribution of Haar fluctuations at each scale to be a half-normal distribution, i.e. a zero-mean Gaussian distribution folded along zero. The mean μ_{HG} of such half-Gaussian distribution, which can be obtained by taking the absolute of a Gaussian distribution, is proportional to the standard deviation σ_G of the original Gaussian distribution, i.e. $\mu_{HG} = \sqrt{\frac{2}{\pi}}\sigma_G$. Since the sign of the Haar fluctuation before taking the absolute is arbitrary, we can also consider the

negative of those fluctuations to create an ensemble which is even closer to a Gaussian distribution and has mean equal zero by construction. We then estimate the mean absolute Haar fluctuation using the definition with the standard deviation given above. This procedure marginally improves the estimation procedure over directly taking the average of the absolute of the Haar fluctuations available.

2.1.3 Slope Estimation

For a power-law relationship between variables x and y such as $y = Ax^B$, it is usual to use standard least-square fitting to find the coefficient A and the exponent B as the linear coefficients of the equivalent linear relationship after taking the logarithm of the equation such that: $\log y = B \log x + \log A$. Least-square fitting assumes the residuals are normally distributed which is often a good approximation for the logarithm of the power spectral density of a stochastic process at a given frequency.

In the case of stationary Gaussian processes, it can be shown that the CPG and LSP estimates at a given timescale are distributed like a chi-square distribution with degrees of freedom equal to two (i.e. an exponential distribution), and for the MTM it is approximately twice the number of tapers, albeit slightly less depending on their degree of dependence. The logarithm of the distributions mentioned above is similar enough to a normal distribution to obtain reasonable estimates with an ordinary least-square fit. Another option is to use a generalized linear model with a gamma distribution model which the chi-square is a special case of. While very similar results are obtained for both fitting methods, we chose the generalized linear model method since it is theoretically more justified.

For a Gaussian process, the first-order Haar fluctuations will follow a normal distribution, but as we are fitting the absolute value of the fluctuations, the estimates at a given timescale will follow a half-normal distribution as mentionned above. Although the half-normal distribution is not a specific case of the gamma distribution, but rather the generalized gamma distribution, we also use the generalized linear model to estimate its scaling exponent for practical purposes.

Timescale Band for Slope Estimation

For all methods, we need to select a range of timescales over which to estimate the scaling exponent using the generalized linear model. The maximum fitting timescale τ_{max} used was always one third of the largest scale available in order to maximize the range of timescales used while avoiding the longer timescales which have poor statistics in the case of the HSF (Lovejoy and Schertzer, 2012), and underestimate variance in the case of the MTM because of its known bias (Prieto et al., 2007) and the usual linear detrending. Since the impact of irregularity is not important at long timescales, we chose this same τ_{max} for all methods for consistency. The optimal minimum timescale τ_{min} on the other hand can vary depending on the method used and requires a careful treatment.

For the CPG and the MTM, we fitted above the timescale corresponding to the Nyquist frequency, i.e twice the resolution, for regular series. In the case of irregular data, since the interpolation brings about a power loss at small timescales (Schulz and Stattegger, 1997; Rhines and Huybers, 2011; Kunz et al., 2020), fitting over the small timescales produces a positive bias on the slope estimation. Therefore, when estimating the scaling exponents for the irregular series, we consistently fitted above

three times the mean resolution τ_μ as a compromise (i.e. 1.5 times the Nyquist frequency). This choice of τ_{min} was informed (as for the methods below) from the results using the paleoclimate database (see section 3.3).

For both the HSF ans LSP, we used twice the resolution as τ_{min} for the regular cases and we used $\tau_{min} = 2\tau_\mu$ for the irregular cases.

2.2 Evaluation of the estimators

2.2.1 Surrogate data

To test the methods, we produce surrogate data with the same characteristic as the paleoclimate archives, namely a given scaling behaviour and an irregular sampling in time.

Simulation of power-law noise

A classical example of a non-stationary process ($H > 0$) is normal Brownian motion ($H = \frac{1}{2}$) which is produced by (integer order) integration of a normal Gaussian white noise ($H = -\frac{1}{2}$). A process with a given scaling behaviour can be obtained by fractional, rather than integer, order integration (or differentiation) of a given set of innovations, i.e. a random series of uncorrelated values obtained from a certain distribution. To generate our surrogate data, we consider the simplest case when the innovations are drawn from a normal Gaussian distribution. This leads to the classical fractional Gaussian noise (fGn) and fractional Brownian motion (fBm) (Mandelbrot and Van Ness, 1968; Mandelbrot, 1971; Molz et al., 1997). For any fGn process there is a related fBm process which can be obtained by (integer-order) integration, which increases the scaling exponent of the process by $+1$. It is usual to define the associated fGn and fBm processes by the same scaling exponent $h \in (0, 1)$ which directly describes the scaling behaviour of the HSF for the fBm, or for the fGn after (integer order) integration. However, it is inconvenient to use the same parameter for both fGn and fBm when considering them in a common framework and we prefer to also identify the fGn by the scaling behaviour of its HSF, rather than that of its integral, such that it has $H \in (-1, 0)$. In the current paper, we generally refer to both fGn and fBm as "fractional noise" described by an exponent $H \in (-1, 1)$. They are generated using an algorithm developed by Franzke et al. (2012).

Series with $H \in (-0.5, -0.3)$ are typical of monthly land air-surface temperature up to decadal timescales, while series with $H \in (-0.3, 0)$ are more typical of sea-surface temperature over similar timescales. Non-stationary behaviour with $H > 0$ is typically observed in pre-Holocene series comprising Dansgaard-Oeschger events (Nilsen et al., 2016).

Generation of irregular paleoseries

To produce irregularly sampled series akin to paleoclimate archives, an "annual resolution" series with a given scaling exponent is first produced with the above method, and then degraded at the desired resolution using two different methods. The first one is to simply block average, and the second one is to sub-sample a low-pass filtered version of the series.

For the block averaging method, boundaries are determined as the midpoint between subsequent timesteps, and all data in between is averaged. This corresponds, in the temporal domain, to a convolution with a square window, or equivalently in

the frequency domain, multiplying the Fourier transform by the $sinc$ function ($sinc(x) = x^{-1}sin(x)$). Therefore, the power spectrum is multiplied by a $sinc^2$ and this brings about a loss of power at the high frequencies. For the second method, the timescale of the low-pass filter is taken as twice the mean resolution of the series, and the filtered series is simply sub-sampled at the desired timesteps. In the frequency domain, the filtering corresponds to multiplying the spectrum by a square function which cuts off variability below the specified timescale, and therefore, this is useful to reduce the aliasing of higher-frequencies.

The first method would correspond to an archive sampled without gaps and where there is no smoothing of the signal, for example speleothem and varved sediments if samples containing several layers are taken, or marine records when the sample distance is smaller than the typical mixing depth in the sediment (Berger and Heath, 1968). The second method would correspond to archives with spaced out sampling (for example 1 cm every 10 cm) and including processes such as bio-turbation and diffusion which smooth out the high-frequency signal, for example biochemical and ecological data extracted from sediment cores (Dolman and Laepple, 2018; Dolman et al., 2020). We show the second method as our main result since it is more common to have archives with such smoothing processes, but also because it removes most aliasing effect from our results. It therefore leaves a clearer picture of the other effects inherent to each methods. It is however an idealization since the smoothing timescale of the physical processes involved is not related to the resolution; it should be independently estimated and accounted for in applied studies, although it is seldom reported. All the results were also computed for the block average method and are shown in the supplementary information (Fig. S1, Fig. S2, Fig. S3, Fig. S4, Fig. S5, Fig. S6S7, Fig. S8, Fig. S9, Fig. S10).

Previous studies have shown that the distribution of sampling times for typical sedimentary records can be approximated by a gamma distribution (Reschke et al., 2019a). To test systematically the impact of increasingly irregular sampling, we thus draw the time steps from a gamma distribution defined by its shape parameter k (or skewness ν such that $k = \nu^{-1}$) and mean parameter μ which corresponds to the mean time-step τ_μ. When the skewness parameter is $\nu = 0$, then we have regular sampling of width μ. Figure 1 shows an example of such surrogate series at annual resolution and an irregularly degraded version (with timesteps drawn from a gamma distribution with $\nu = 1$), along with the results of applying the three scaling estimation methods considered for the main analysis: the MTM, the HSF and the LSP. We omit a discussion of the results computed using the CPG as they are generally very similar to those of the MTM, albeit with higher variance.

2.2.2 Performance metrics and performance plots

Our aim is to evaluate how the different methods perform in the estimation of scaling exponents for irregularly sampled paleoclimate data $X(t)$ with different scaling behaviour, and of variable length and irregularity. In order to assess the accuracy and precision of our scaling exponent estimator $\hat{H}$ for a given set of parameters, we generate an ensemble of surrogate data $\hat{X}(t)$ and analyze its statistics. We evaluate three measures: the bias B for the accuracy of the estimator $\hat{H}$ with respect to the input H, the standard deviation σ for the precision of the estimator, and the root-mean-square error $RMSE$ which combines both previous measures:

$$B = \leq \langle \hat{H} \geq \rangle - H \tag{15}$$
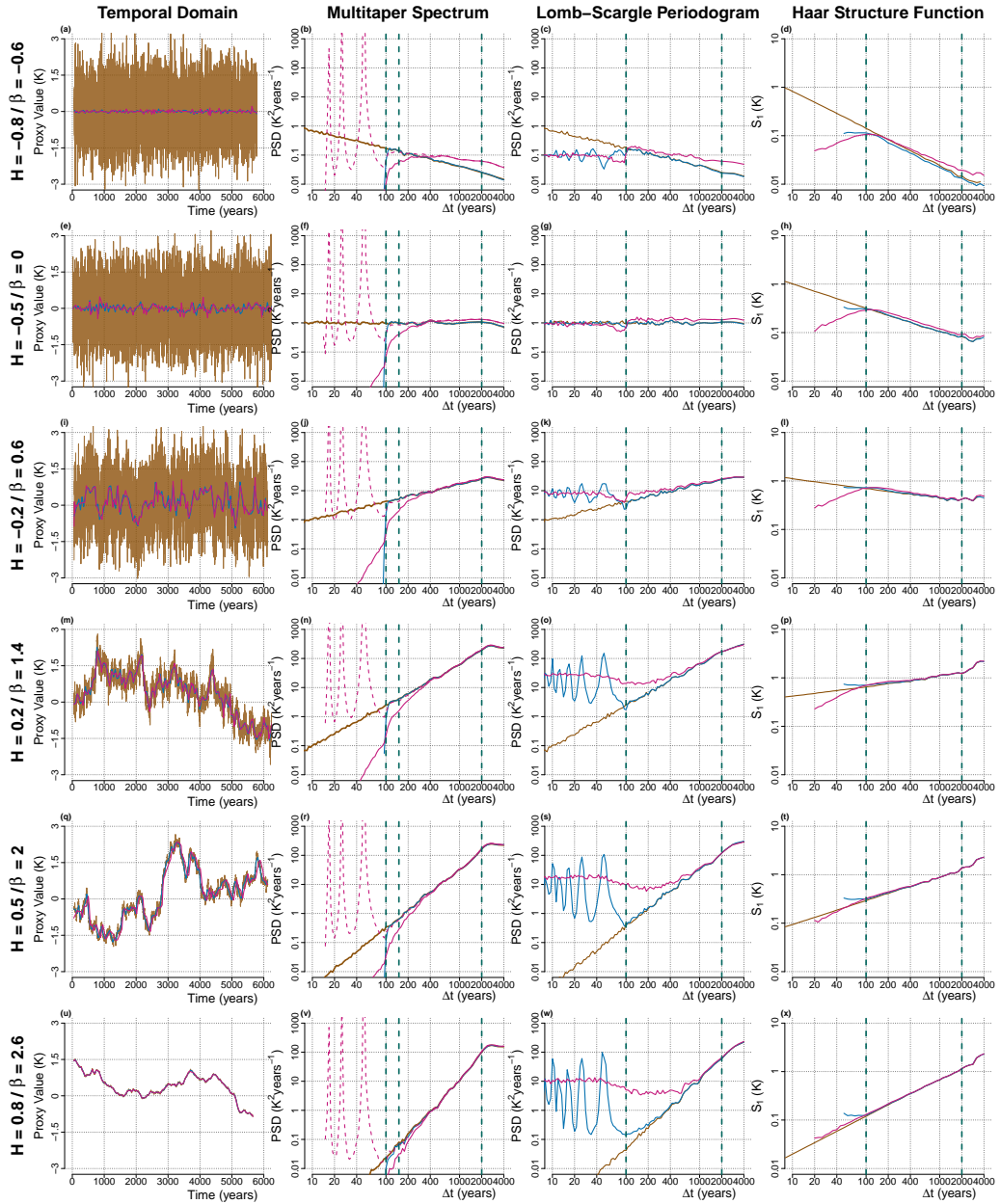
Figure 1. (a,e,i,m,q,u) Surrogate timeseries generated with a given H are shown at annual resolution (brown), degraded to a regular 50-year resolution (blue) and degraded to an irregular and random spacing drawn out of a gamma distribution with skewness $\nu = 1$ and mean 50 years. **(b,f,j,n,r,v)** Shown are the mean power spectra, estimated using the MTM, of 100 realizations of surrogate timeseries generated as in (a,e,i,m,q,u), respectively, and shown with the same colour scheme. The irregular case is also shown after dividing the power spectra by the expected $sinc^4$ bias due to interpolation (dashed pink). Also shown are the bounds for the fitting range considered later (vertical dashed blue): at the Nyquist frequency corresponding to 100 years, at 1.5 times the Nyquist frequency corresponding to 150 years and at one third the length of the timeseries at 2000 years. **c,g,k,o,s,w (c,g,k,o,s,w)** Same as (b,f,j,n,r,v), respectively, but for the LSP instead of the MTM. **d(d,h,l,p,t,x)**. Same as (b,f,j,n,r,v), respectively, but showing HSF instead of MTM power spectra.

$$\sigma = \sqrt{< \hat{H}^2 > - < \hat{H} >^2}\sqrt{\langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2} \tag{16}$$

$$\mathrm{RMSE}^2 = B^2 + \sigma^2 \tag{17}$$

We exploited the geometric relation between the three to easily visualize the results. We summarize the results for a set of parameters by one point on a bias-standard deviation diagram (Fig. 4) where the x-axis gives the bias, the y-axis the standard deviation and the distance from the origin then corresponds to the root-mean-square error. The bias and standard deviation at for each combination is calculated from a large ensemble of surrogate data (10 000 realizations).

2.3 Data

In order to test the methods with the sampling properties of typical paleo-proxy data, we consider an available database of marine and terrestrial proxy records for temperature (Rehfeld et al., 2018), which was also used for signal-to-noise ratio assessments by Reschke et al. (2019b). Each of the 99 sites covers at least 4000 years in the interval of the Holocene (8–0 kyr ago, 88 time series in total) and/or the LGM (27–19 kyr ago, 39 time series total) at a mean sampling interval of 225 years or lower. These records are irregularly sampled in time, and come with different sampling characteristics (Fig. 2).

3 Results

Using the methods described above to generate synthetic data, we can evaluate the ability of each method to recover the input scaling exponents. For each method, we calculate the bias and standard deviation over the ensembles of surrogate data, for input H between -0.9 and 0.9 in increments of 0.1. This wide range of H values covers all values encountered later in the multi-proxy database; the vast majority of climatic series even fall within an even more restricted range of $H \in [-0.5, 0.5]$. First, we consider the ideal case of regular sampling, then we study the effect of irregular sampling and finally, the impact of the length of the time series.

3.1 Effect of Regular and Irregular Sampling

We evaluate the estimators for four cases pertaining to the resolution of the data and always a fixed length of 128 data points (Fig. 3, Fig.4). The first case considers "annual data" which was directly simulated and not degraded afterwards; it is therefore regular. It is shown for comparison with the second case where we simulate 5120-year long series and then degrade them to 40-year resolution. This allows us to study the impact of the degrading method which is necessary for producing irregular series. The third and fourth cases deal with series of 128 irregular timesteps drawn from a gamma distribution with skewness $\nu = 0.5$ and $\nu = 1$ respectively, and a mean parameter of 40-year so that the resulting series have the same mean resolution as the second (regular) case. To illustrate the contribution of different frequency ranges to the precision and accuracy of the
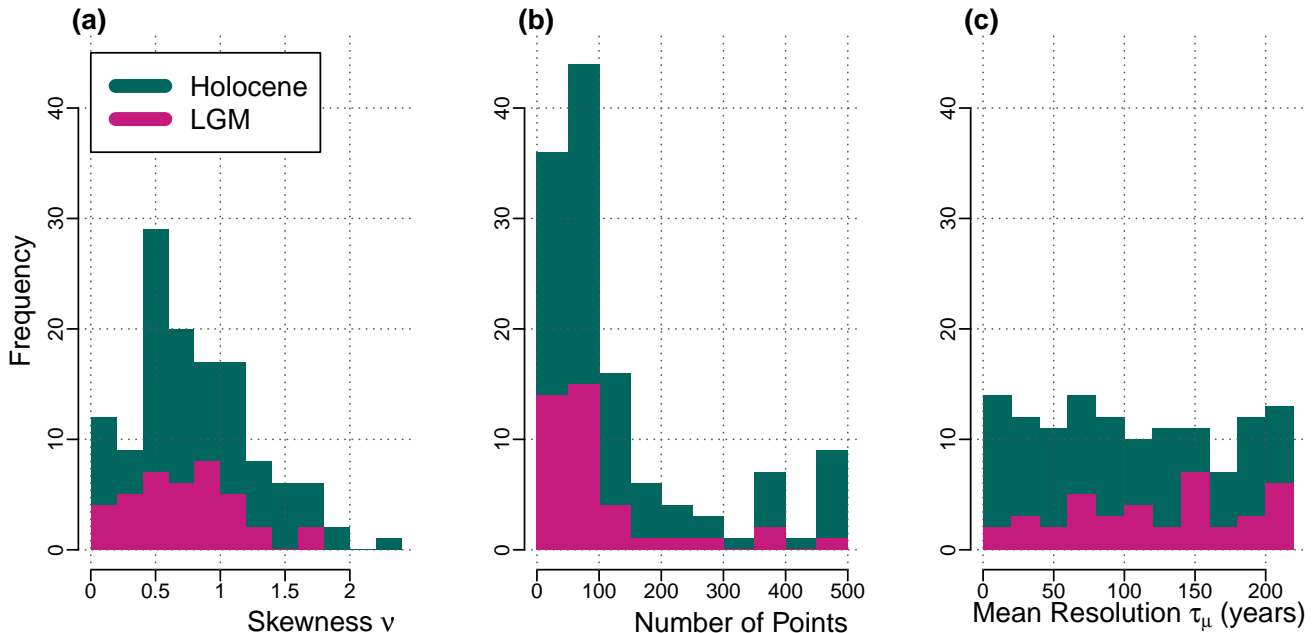
Figure 2. Sampling characteristics of the 127 paleoclimate timeseries (Rehfeld et al., 2018). The 88 Holocene series (teal) and 39 Last Glacial Maximum series (pink) are evaluated along the following characteristics: **(a)** the skewness ν of the distribution of time steps , **(b)** the number of points, i.e. samples , and **(c)** the mean resolution τ_μ of the time steps.

estimators, the scaling exponents were also fit on sub-ranges of equal width in the log of the timescales which we refer to as the shorter timescales, the intermediate timescales and the longer timescales, corresponding to, respectively, 2-9.2 τ_μ, 4.3-19.9 τ_μ and 9.2-42.7 τ_μ, where τ_μ is the mean resolution of the timeseries (Figure 5).

For the "annual data" series, the scaling exponent H can generally be recovered for all values of $H \in [-0.9, 0.9]$ with a standard devation below 0.13 and an absolute bias less than 0.06 (Fig. 4, top left), except for the LSP when $H > -0.1$ as it increasingly underestimates higher values of H. The small bias observed in the MTM and LSP estimates stems from the shorter timescales (Fig. 5) which are sensitive to aliasing of the power below the Nyquist frequency, and therefore the measured scaling slope of series with negative β (i.e. $H < -0.5$) are biased high whereas increasingly negative bias is obtained for increasingly positive values of β ($H > -0.5$). It follows that the flat case of $\beta = 0$ ($H = -0.5$) is unbiased by aliasing since the aliased power is likewise flat. The HSF based estimates also have a significant bias in the other direction when the series considered have an input H decreasing below $H = -0.5$ (Fig. 3).

The second case of regular 40-year series yields similar results as the previous case, although there is no more aliasing since we sub-sampled the 5120-year long series at 40-year intervals after an 80-year low-pass filter was applied (Fig. 4, top right). The MTM returns a consistent estimate with a variance which remains almost constant over the range of H. The estimator has little bias for stationary series with $H < 0$, but it steadily increases for higher values as the lower-frequencies bend upward
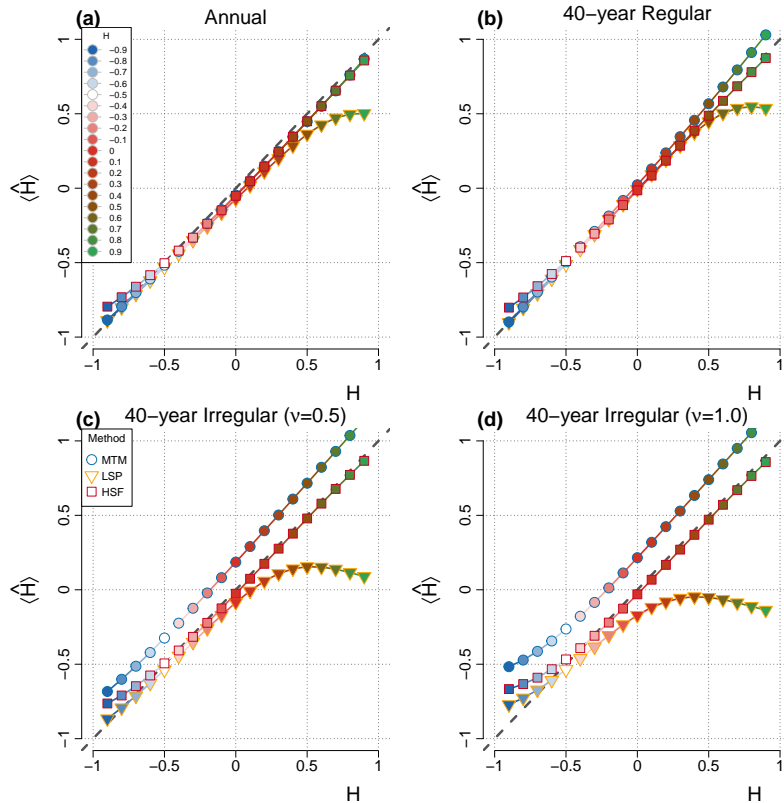
Figure 3. Mean estimate $\langle \hat{H} \rangle$ from the surrogate timeseries with input values of $H \in (-1,1)$ (i.e. $\beta \in (-1,3)$). Deviations from the 1-to-1 line (dashed gray) correspond to the the bias B of the mean estimate $\langle \hat{H} \rangle$. Different types of data are evaluated: **(a)** regular "annual data", i.e. it was directly simulated and not degraded after, **(b)** regular 'surrogate data degraded at regular 40-year interval, and irregular surrogate data with timesteps drawn from a gamma distribution with **(c)** skewness $\nu = 0.5$ or **(d)** skewness $\nu = 1$.

(Fig. 5). This seems to be related to the bias low characteristic of the MTM at the longest timescales (Prieto et al., 2007) which creates an inflection point (around $\Delta t \approx 3000$ years on Fig. 1) and the power lost to its right appears redistributed to its left. Therefore, as H increases, the amount of power lost is more important and the inflection stronger. The LSP estimates are also largely unbiased until about $H = 0.5$, and above it a strong negative bias is developed, particularly on the side of the smaller timescales (Fig. 5, Fig. 3, Fig. 1).

The problem of irregularity is considered using irregular surrogate data skewness parameter $\nu = 0.5$ and $\nu = 1$ (Fig. 4, bottom row). In the case of the MTM, we observe a consistent bias for all H of about 0.5 and 0.1 for the shorter and intermediate timescales, respectively, for the weakly irregular case ($\nu = 0.5$), and practically no bias (0.02) at the longer timescales (Fig. 5). This corresponds very closely to the expected bias from the linear interpolation step, and it could potentially be accounted and corrected for. While it is outside of the scope of this study to formally develop such method, if we apply a first-order correction
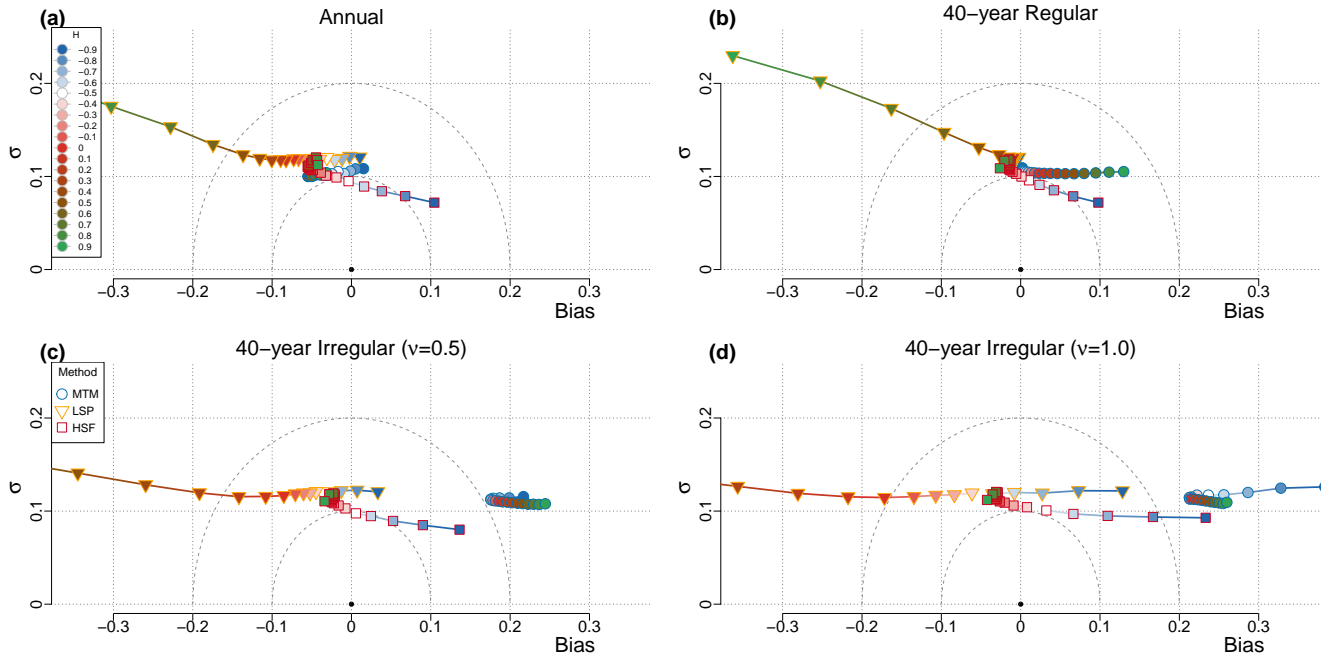
Figure 4. Bias-Standard Deviation diagram for ensembles of H estimates surrogate timeseries with input values of $H \in (-1, 1)$ (i.e. $\beta \in (-1, 3)$). Different types of data are evaluated: **(a)** regular "annual data", i.e. it was directly simulated and not degraded after, **(b)** regular 'surrogate data degraded at regular 40-year interval, and irregular surrogate data with timesteps drawn from a gamma distribution with **(c)** skewness $\nu = 0.5$ or **(d)** skewness $\nu = 1$.

by dividing the MTM spectra by a $sinc^4$ with time constant corresponding to the mean sampling resolution, we can greatly

335   reduce the bias down to the Nyquist frequency (Fig. 1, Fig. ~~S7~~S11).

The negative bias for the LSP evoked above for $H > 0.5$ amplifies with irregularity and even affects significantly a wider range of H values, down to $H = -0.5$ for the most irregular case ($\nu = 1$.). This is a consequence of higher than expected variance over the smaller timescales in the LSP when the slope is steep, as already identified by Schulz and Mudelsee (2002) using red noise. The HSF estimates on the other hand remain mostly insensitive to the irregular sampling for all $H > -0.5$, but

340   the positive bias we already identified in the regular case when $H < -0.5$ amplifies for more negative values of H. Similarly, all methods yield increasing positive bias for such anti-persistent timeseries (i.e. with $H < -0.5$) when the sampling is irregular. This conjuncture seems to indicate that the anti-correlation characteristic of $H < -0.5$ processes is lost to some extent when degraded in an irregular manner to produce the surrogate data, as they all show a similar increase of their bias (by about 0.15 for the $H = -0.9$ case) with respect to their bias for the $H = -0.5$ case (Fig. 3, Fig. 4).
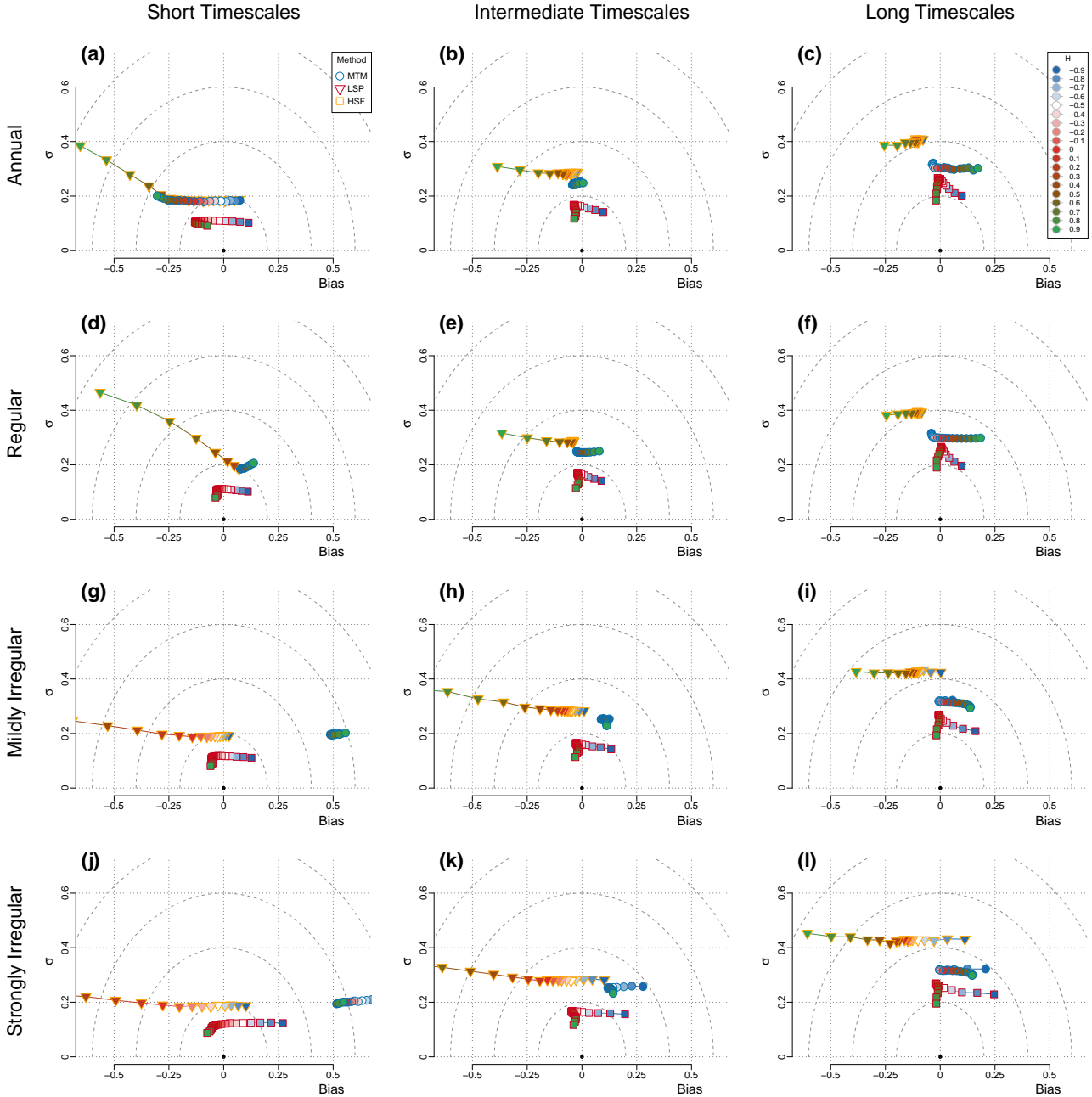
Figure 5. Timescale dependence of the bias and variance for regular and irregular series. We evaluate the three methods: MTM (circles), HSF (square), and LSP (triangles). The colors corresponds to the input H value for each simulation, ranging from -0.9 to 0.9. in increments of 0.1. The rows correspond to different types of surrogate series: **(a-c)** "annual data", **(d-f)** regular data , **(g-i)** mildy irregular data ($\nu = 0.5$), **(j-l)** strongly irregular data ($\nu = 1.0$); see section 2.2.1. The columns correspond to three different fitting ranges in terms of the mean resolution τ_μ, **(a,d,g,j)** the shorter timescales: 2-9.2 τ_μ, **(b,e,h,k)** the intermediate timescales: 4.3-19.9 τ_μ and **(c,f,i,l)** the longer timescales: 9.2-42.7 τ_μ.

3.2 Effect of Time series length

While the irregularity had a larger impact on the bias of the estimator, the length of the time series mostly influences the variance of the estimator (Fig. S86, Fig. S12).

In the case of the MTM estimates, all values of H result in a similar standard deviation for a given resolution. When the resolution is doubled, the standard deviation increases by a factor close to the expected $\sqrt{2}$ (by $\sim$1.4-1.7), going from $\sigma \approx 0.27$ at the 160-year resolution to $\sigma \approx 0.07$ at the 20-year resolution (Fig. S8S12). The bias also improves when the resolution increases, particularly for the higher values of H such as for $H = 0.9$ which goes from $B \approx 0.27$ at the 160-year resolution to $B \approx 0.10$ at the 20-year resolution, while for negative H values it goes from $B \approx 0.05 - 0.07$ to $|B| < 0.01$ for the same resolution change (Fig. 6, Fig. S6).

In the case of the LSP, up to $H \approx 0.1$, the standard deviation of the estimates also decrease by a factor close to $\sqrt{2}$ (by $\sim$1.4-1.8, Fig. S8S12) with each doubling of resolution, going from $\sigma \approx 0.35$ at the 160-year resolution to $\sigma \approx 0.08$ at the 20-year resolution, and the bias improves only slightly since it was already small ($|B| < 0.04$). For the higher H values, the standard deviation improves less (by $\sim$1.2-1.4 at $H = 0.9$, Fig. S8S12), going from $\sigma \approx 0.43$ at the 160-year resolution to only $\sigma \approx 0.19$ at the 20-year resolution. At the same time, the bias change becomes more positive, thus compensating slightly better the overall strong negative bias characteristic of the LSP estimates for very high H values.

In the case of the HSF, when increasing the resolution, the standard deviation of the estimates improves more for the lower H values than for the higher values, improving by a factor of $\sim$1.4-1.8 for $H = -0.9$ to $\sim$1.2-1.4 for $H = 0.9$ (Fig. S8S12). Overall, the standard deviation improves from $\sigma \approx 0.20 - 0.25$ at the 160-year resolution to $\sigma \approx 0.06 - 0.09$ at the 20-year resolution. The slight negative bias found for the $H > -0.5$ estimates ($|B| < 0.04$) practically vanishes, but not the positive one for $H < -0.5$ which only decreases slightly.

3.3 Application to database

In order to see how these results translate to typical proxy records, we perform the analysis with surrogate data with sampling characteristic and scaling behaviour directly extracted from the database of Holocene and LGM records (see section 2.3). We make the assumption that the series are approximately power-law scaling and that they can therefore be modelled by fractional noise described by a scaling exponent H. Since the best H to describe the approximate scaling behaviour of the series is unknown, we make an initial approximation with each method and take the median of the three results as the reference value for the given series, with which we generate an ensemble of surrogate with the same sampling scheme as the given series. We use the real timesteps in order to evaluate the impact of each specific sampling scheme. Again, we consistently fitted up to one third of the length (see section 2.1.3), but determined empirically the best minimum fitting scale τ_{min} such that it minimizes the RMSE of the estimator with respect to the reference H used to generate the surrogate data.

We found the HSF to yield the smallest τ_{min} of the three methods, with τ_{min} below twice the mean resolution τ_μ for 121 out of 127 series and even below τ_μ for 43 out of 127. The LSP yielded similar results with 109 out of 127 series having τ_{min} below twice τ_μ, and 41 out of 127 having τ_{min} even below τ_μ. This contrasts with the MTM which almost never suggests best
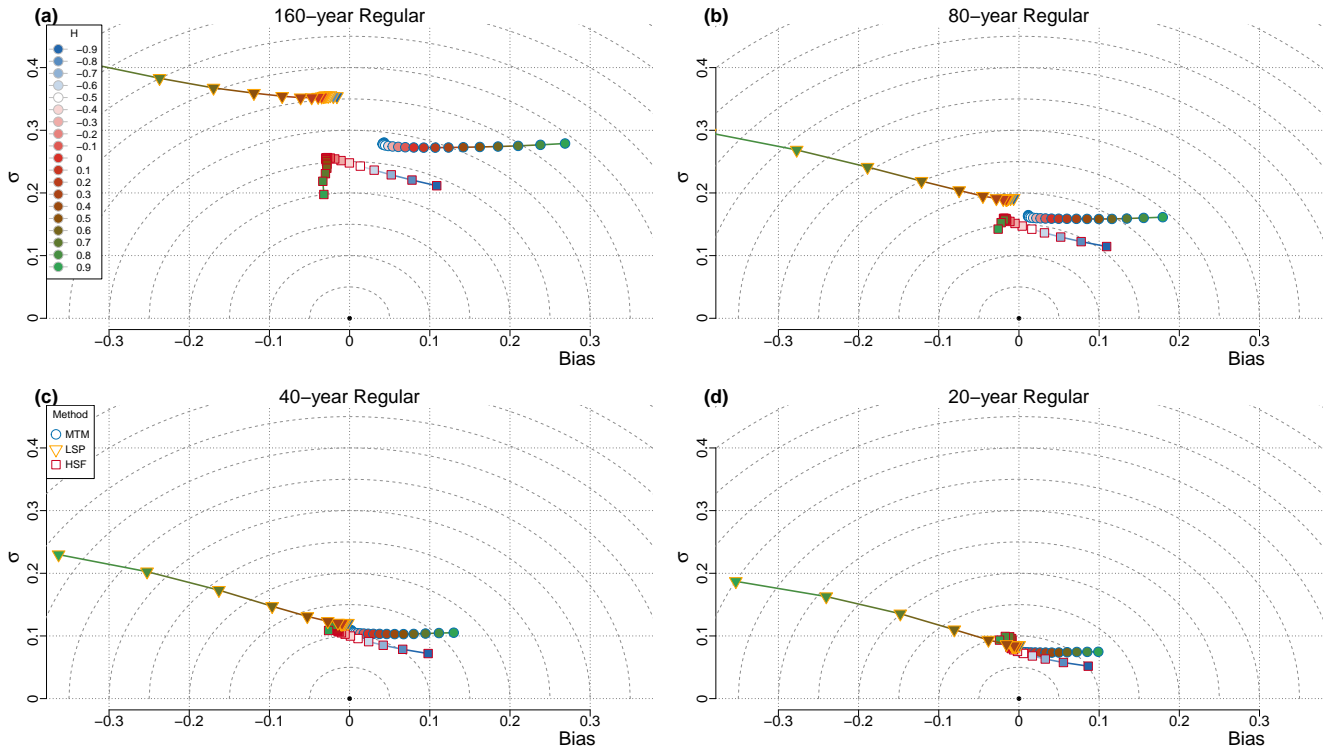
Figure 6. Resolution dependence of the estimators. The performance of the estimators is shown for regular surrogate data of different resolutions: **(a)** 160 years , **(b)** 80 years, **(c)** 40 years and **(d)** 20 years. For each case, the series spanned 5120 years, and therefore each case contains 32, 64, 128 and 256 data points, respectively.

results for τ_{min} below twice τ_μ, but rather an average minimum resolution $\bar{\tau}_{min} = 3.3\tau_\mu$, compared to $\bar{\tau}_{min} = 1.2\tau_\mu$ for the HSF and $\bar{\tau}_{min} = 1.3\tau_\mu$ for the LSP. This underlines that because interpolation-free methods give more reliable estimates at

380 shorter timescales, they allow for a better usage of the full data.

The HSF yields the best results of all three methods both in terms of variance of the estimates, and in terms of bias, with a mild tendency towards positive biases. The MTM also tends to show positive biases, albeit higher, while the LSP tends to show negative biases. Both the MTM and the LSP generally show higher variance of their estimators than the HSF (Fig. 8), although for different reasons. The MTM has higher variance because the higher τ_{min} does not allow to use the smaller timescales in the

385 estimation procedure, while the LSP shows higher variance because of the timeseries with high input H for which it performs poorly, especially when the data are irregular (Fig. 4).

On average, the HSF method gave the lowest RMSE $= 0.11 \pm 0.04$ compared to RMSE $= 0.30 \pm 0.15$ and RMSE $= 0.19 \pm 0.15$ for the MTM and the LSP, respectively. The poor performance of the LSP compared to the HSP stems from the higher H since it performs as well as the Haar when $H < 0$, with RMSE $= 0.12 \pm 0.07$ (Fig. 9).
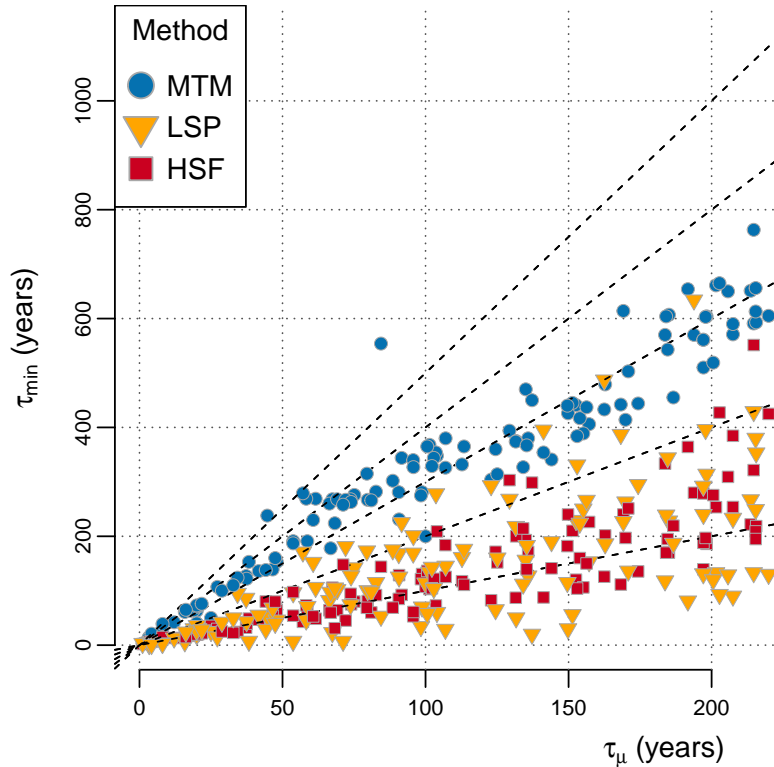
Figure 7. Best minimum fitting timescales. For each methods are shown the best minimum fitting timescales τ_{min} (minimizing the RMSE in H) as a function of the mean resolution τ_μ for ensembles of surrogate data generated with the same sampling scheme as the corresponding paleoclimate timeseries from the database.

4 Discussion

Our comparative study indicates that for irregular timeseries the methods with interpolation, i.e. the CPG and the MTM, are less efficient to evaluate variability across timescales than the methods without interpolation, i.e. the LSP and the HSF. In the case of regular timeseries however, all methods were found to perform similarly (Fig. 4b) for a wide range of input H, and only showed bias on the fringes of the H range considered. In addition, we found that the choice of method should also be informed by the characteristics of the underlying process measured since there was an observed dependence between the performance metrics and the scaling exponent H which generated the surrogate data. As such, the LSP may be only appropriate for irregular series suspected of having H near or below $H = -0.5$, while the HSF shows better reliability even when H is near or above $H = 0$. It might not be so surprising that the LSP performed poorly for higher H values since it has been developed to approximate the CPG for stationary noise processes (Scargle, 1982) and processes with $H > 0$ are non-stationary, i.e. their mean is not stable and changes with time.
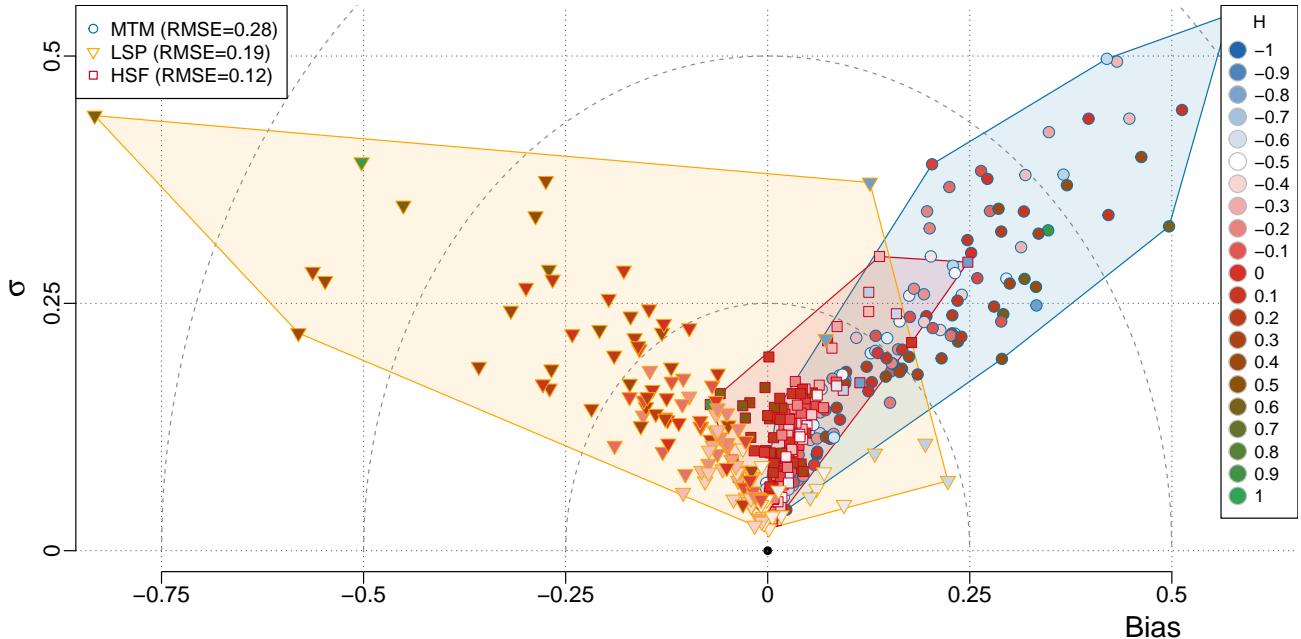
Figure 8. Bias-standard deviation diagram for the surrogate timeseries generated using timesteps from the paleoclimate database. The input H used to generate the timeseries is indicated by the colour inside the markers. The shaded polygons contain all the points for a given method (see legend for colours).

In addition to the irregularity, we also compared how the resolution of the timeseries affects the estimators of H. While making the series more irregular mostly affected the accuracy of the estimates, i.e. their bias, without affecting their precision, i.e. their standard deviation, increasing the resolution mostly improved the precision and the accuracy only to a lesser extent. Increasing the resolution naturally improves the estimates as it provides more information and also because it decreases the relative weight of the most uncertain timescales: the smaller timescales, near the sampling resolution, and the longer timescales, near the length of the series.

Our conclusions however rely on the real data having similar properties to the surrogate data used to validate the methods, and the conclusions may be different if the real data have properties different than the assumed power-law scaling. It is difficult to say for example which method would perform best if the data analyzed would contain two different scaling regimes, one with low H and the other with high H (Lovejoy and Schertzer, 2013). Another limitation of our surrogate data is that it is Gaussian-distributed, whereas real paleoclimate data can exhibit multifractality (Schmitt et al., 1995; Shao and Ditlevsen, 2016). Furthermore, in order to degrade the simulated annual timeseries into irregular paleoclimate-like data we had to use simplified numerical methods to mimic the physical processes and manipulations leading to the recording and retrieval of paleoclimate archives. For our purpose, we chose to low-pass filter the data and sub-sample since this is very common in sedimentary data and ice cores because of, respectively, bio-turbation and diffusion (Dolman and Laepple, 2018; Dolman et al., 2020; Kunz
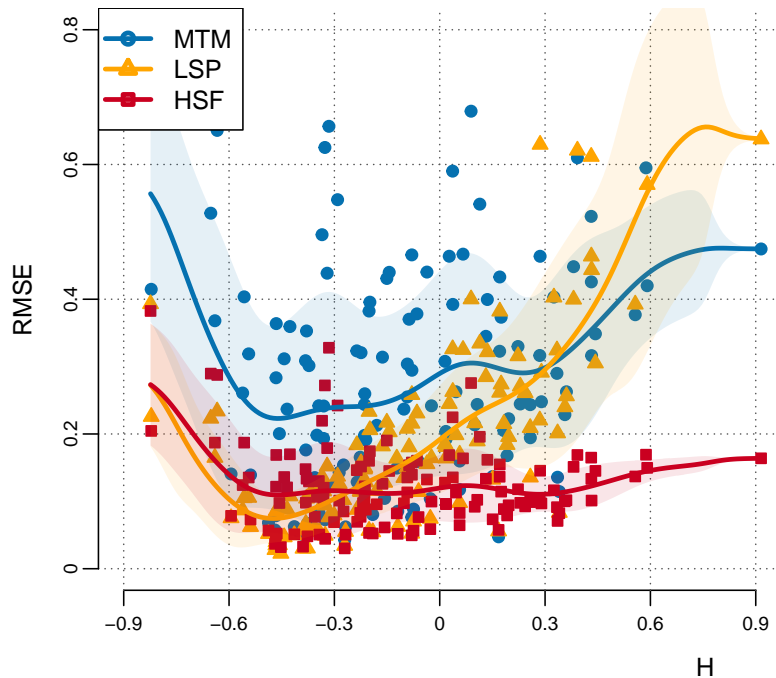
Figure 9. RMSE of the H estimates as a function of the input H for the surrogate timeseries based on the proxy database. The RMSE are given for the three estimation methods as a function of the H used to generate the surrogates. Also shown is a Gaussian smoothing of the points for each method (thick line) and the one standard deviation confidence interval (shaded).

et al., 2020). However, reality is more complex and other processes could significantly alter the recorded signal and bias the variability estimates in ways not accounted for by our surrogate model. For specific case studies, it is advisable to develop forward models (Stevenson et al., 2013; Dee et al., 2017; Dolman and Laepple, 2018; Casado et al., 2020) of the studied paleoclimate archives in order to understand potential biases generated by the recording and specific sampling.

420   In summary, it is difficult to ascertain which method is the "best" since the answer depends on the irregularity, the resolution and, most importantly, the underlying H values. In this respect, the HSF is possibly the safer method for paleoclimate applications since it only performs poorly for series with $H < -0.5$, an almost unseen behaviour in climatic timeseries at timescales longer than decadal. Although the LSP gives equally good result for $H \lesssim 0$ and even better near $H = -0.5$, i.e. white noise, its increasing bias and standard deviation for higher H values makes it an uncertain choice for timescales longer than centennial

425   since climate timeseries at those timescales often show $H \approx 0$ and greater. The MTM produces good estimates for relatively regular series, but struggles with highly irregular timeseries since it must rely on interpolation. While this produces a large bias on the shorter timescales which we have thus discarded, the bias is rather consistent and methods could be developed to correct for it (e.g. Fig. ~~S7~~S11). On the other hand, the longer timescales of the MTM are relatively well estimated and it is appropriate to estimate the absolute variance over a timescale band well above the mean resolution (Rehfeld et al., 2018;

430 Hébert et al., 2021). Further, the effect of proxy biases are well studied for the power spectrum (Kunz et al., 2020; Dolman et al., 2020) mainly due to the known properties of the Fourier transform. However, the MTM, and the LSP, do not allow for the characterization of intermittency through the study of all statistical moments, contrary to the HSF which is thus better suited for the analysis of timeseries displaying multifractality, such as paleoclimate timeseries at glacial-interglacial timescales comprising Dansgaard–Oeschger events (Schmitt et al., 1995; Shao and Ditlevsen, 2016)

435 Our study is not comprehensive and other methods such as the bias corrected version of Lomb-Scargle (Schulz and Mudelsee, 2002, REDFIT) and the z-transform wavelet (Zhu et al., 2019) exist. More precise estimates of the scaling exponent are obtained by parametric methods based on maximum likelihood, but they require strong assumptions about the underlying process (Del Rio Amador and Lovejoy, 2019). We aimed to cover the most commonly used methods that also allow for a simple interpretation of variability across timescales, either as they represent the well studied power spectrum or the Haar
440 fluctuations that directly provide changes in amplitude.

5 Conclusions

Characterizing the variability across timescales is important to understand the underlying dynamics of the Earth system. It remains challenging to do so from paleoclimate archives since they are more than often irregular and traditional methods to produce timescale-dependent estimates of variability such as the classical periodogram and the multitaper spectrum generally
445 require regular time sampling.

We have compared those traditional methods using interpolation with interpolation-free methods, namely the Lomb-Scargle periodogram and the first-order Haar structure function. The ability of those methods to produce timescale-dependent estimates of variability when applied to irregular data was evaluated in a comparative framework. The metric we chose to compare them is the scaling exponent, i.e. the linear slope in log-transformed coordinates, since it summarizes the behaviour of the
450 variability over a given timescale band. Doing so assumes power-law scaling, a behaviour which is often observed in geophysical timeseries at least approximatively (Mandelbrot and Wallis, 1968; Cannon and Mandelbrot, 1984; Pelletier and Turcotte, 1999; Malamud and Turcotte, 1999; Fedi, 2016; Corral and González, 2019). To evaluate our estimators, we generated fractional noise as surrogate annual timeseries characterized by a given scaling exponent H, also known as fractional Gaussian noise when they are stationary ($H < 0$) and fractional Brownian motion when they are non-stationary ($H > 0$). The annual
455 timeseries were then degraded to resolutions characteristics of paleoclimate archives for the recent Holocene.

We found that for scaling estimates in irregular timeseries, the interpolation-free methods are to be preferred over the methods requiring interpolation as they allow for the utilization of the information from shorter timescales without introducing additional bias. In addition, our results suggest that the Haar structure function is the safer choice of interpolation-free method since the Lomb-Scargle periodogram is unreliable when the underlying process generating the timeseries is not stationary.
460 This conclusion was reinforced by the application to the proxy database which indeed showed the Haar structure function to give more reliable estimates over a wide range of H. Given that we cannot know a priori what kind of scaling behaviour is

contained in a paleoclimate timeseries, and that it is also possible that this changes as a function of timescale, it is a desirable characteristic for the method to handle both stationary and non-stationary cases alike.

Code and data availability. The data is available as a supplement to Rehfeld et al. (2018) and the code is publicly available on BitBucket: https://bitbucket.org/RphlHbrt/rscaling/src/master/.

Author contributions. All authors participated in the conceptualization of the research and the methodology. RH developed the software and visualization, and conducted the formal analysis and investigation. KR and TL provided supervision. RH prepared the original draft and all authors contributed to the review and editing of the final manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. This study was undertaken by members of CVAS following discussions at several workshops. CVAS is a working group of the Past Global Changes (PAGES) project, which in turn received support from the Swiss Academy of Sciences and the Chinese Academy of Sciences. We particularly thank T. Kunz and S. Lovejoy for in depth discussions, as well as T. Graves and C. Franzke for freely sharing software for generating fractional noise. We also acknowledge discussions with M. Casado, L. del Rio Amador, A. Dolman, I. Kröner and T. Münch. We thank all original data contributors who made their proxy data available. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project no. 395588486 and the German Federal Ministry of Education and Research (BMBF) through the PalMod project (grant no. 01LP1926C) is acknowledged. This is a contribution to the SPACE and GLACIAL LEGACY ERC projects; these projects have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 716092 and no. 772852).

References

Benedict, L. H., Nobach, H., and Tropea, C.: Estimation of turbulent velocity spectra from laser Doppler data, Measurement Science and Technology, 11, 1089–1104, https://doi.org/10.1088/0957-0233/11/8/301, https://doi.org/10.1088/0957-0233/11/8/301, publisher: IOP Publishing, 2000.

Berger, W. H. and Heath, G. R.: Vertical mixing in pelagic sediments, Journal of Marine Research, 26, 134–143, 1968.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nature Climate Change, 2, 417–424, https://doi.org/10.1038/nclimate1456, https://www.nature.com/articles/nclimate1456, 2012.

Bradley, R. S.: Paleoclimatology, Elsevier, https://doi.org/10.1016/C2009-0-18310-1, https://linkinghub.elsevier.com/retrieve/pii/C20090183101, 2015.

Cannon, J. W. and Mandelbrot, B. B.: The Fractal Geometry of Nature., The American Mathematical Monthly, 91, 594, https://doi.org/10.2307/2323761, https://www.jstor.org/stable/2323761?origin=crossref, 1984.

Casado, M., Münch, T., and Laepple, T.: Climatic information archived in ice cores: impact of intermittency and diffusion on the recorded isotopic signal in Antarctica, Climate of the Past, 16, 1581–1598, https://doi.org/https://doi.org/10.5194/cp-16-1581-2020, https://cp.copernicus.org/articles/16/1581/2020/, 2020.

Chatfield, C.: The Analysis of Time Series., Springer, New York, NY, https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5575991, oCLC: 1066190229, 2013.

CLIMAP Project Members: The Surface of the Ice-Age Earth, Science, 191, 1131–1137, https://doi.org/10.1126/science.191.4232.1131, https://www.sciencemag.org/lookup/doi/10.1126/science.191.4232.1131, 1976.

Corral, A. and González, A.: Power Law Size Distributions in Geoscience Revisited, Earth and Space Science, 6, 673–697, https://doi.org/https://doi.org/10.1029/2018EA000479, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018EA000479, 2019.

Damaschke, N., Kühn, V., and Nobach, H.: A fair review of non-parametric bias-free autocorrelation and spectral methods for randomly sampled data in laser Doppler velocimetry, Digital Signal Processing, 76, 22–33, https://doi.org/10.1016/j.dsp.2018.01.018, https://www.sciencedirect.com/science/article/pii/S1051200418300307, 2018.

Dee, S. G., Parsons, L. A., Loope, G. R., Overpeck, J. T., Ault, T. R., and Emile-Geay, J.: Improved spectral comparisons of paleoclimate models and observations via proxy system modeling: Implications for multi-decadal variability, Earth and Planetary Science Letters, 476, 34–46, https://doi.org/10.1016/j.epsl.2017.07.036, http://www.sciencedirect.com/science/article/pii/S0012821X1730420X, 2017.

Del Rio Amador, L. and Lovejoy, S.: Predicting the global temperature with the Stochastic Seasonal to Interannual Prediction System (Stoc-SIPS), Climate Dynamics, 53, 4373–4411, https://doi.org/10.1007/s00382-019-04791-4, https://doi.org/10.1007/s00382-019-04791-4, 2019.

Dolman, A. M. and Laepple, T.: Sedproxy: a forward model for sediment-archived climate proxies, Climate of the Past, 14, 1851–1868, https://doi.org/https://doi.org/10.5194/cp-14-1851-2018, https://www.clim-past.net/14/1851/2018/, 2018.

Dolman, A. M., Kunz, T., Groeneveld, J., and Laepple, T.: Estimating the timescale-dependent uncertainty of paleoclimate records – a spectral approach. Part II: Application and interpretation, Climate of the Past Discussions, pp. 1–22, https://doi.org/https://doi.org/10.5194/cp-2019-153, https://cp.copernicus.org/preprints/cp-2019-153/, publisher: Copernicus GmbH, 2020.

Fedi, M.: Scaling Laws in Geophysics: Application to Potential Fields of Methods Based on the Laws of Self-similarity and Homogeneity, in: Fractal Solutions for Understanding Complex Systems in Earth Sciences, edited by Dimri, V., Springer Earth System Sciences, pp. 1–18,

Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-24675-8_1, https://doi.org/10.1007/978-3-319-24675-8_1, 2016.

Franzke, C. L. E., Graves, T., Watkins, N. W., Gramacy, R. B., and Hughes, C.: Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,
520   370, 1250–1267, https://doi.org/10.1098/rsta.2011.0349, https://royalsocietypublishing.org/doi/10.1098/rsta.2011.0349, 2012.

Fredriksen, H.-B. and Rypdal, K.: Spectral Characteristics of Instrumental and Climate Model Surface Temperatures, Journal of Climate, 29, 1253–1268, https://doi.org/10.1175/JCLI-D-15-0457.1, http://journals.ametsoc.org.ezp-prod1.hul.harvard.edu/doi/abs/10.1175/JCLI-D-15-0457.1, 2015.

Graves, T., Gramacy, R., Watkins, N., and Franzke, C.: A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA,
525   1951–1980, Entropy, 19, 437, https://doi.org/10.3390/e19090437, https://www.mdpi.com/1099-4300/19/9/437, 2017.

Haar, A.: Zur Theorie der orthogonalen Funktionensysteme, Mathematische Annalen, 69, 331–371, https://doi.org/10.1007/BF01456326, https://doi.org/10.1007/BF01456326, 1910.

Horne, J. H. and Baliunas, S. L.: A prescription for period analysis of unevenly sampled time series, The Astrophysical Journal, 302, 757, https://doi.org/10.1086/164037, http://adsabs.harvard.edu/doi/10.1086/164037, 1986.

530 Hurst, H. E.: Methods of using long-term storage in reservoirs., Proceedings of the Institution of Civil Engineers, 5, 519–543, https://doi.org/10.1680/iicep.1956.11503, http://www.icevirtuallibrary.com/doi/10.1680/iicep.1956.11503, 1956.

Huybers, P. and Curry, W.: Links between annual, Milankovitch and continuum temperature variability, Nature, 441, 329–332, https://doi.org/10.1038/nature04745, http://www.nature.com/doifinder/10.1038/nature04745, 2006.

Hébert, R., Herzschuh, U., and Laepple, T.: Land temperature variability driven by oceans at millennial timescales, Nature Geoscience,
535   Submitted, 2021.

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., and Bunde, A.: Detecting long-range correlations with detrended fluctuation analysis, Physica A: Statistical Mechanics and its Applications, 295, 441–454, https://doi.org/10.1016/S0378-4371(01)00144-3, https://www.sciencedirect.com/science/article/pii/S0378437101001443, 2001.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., and Stanley, H.: Multifractal detrended fluctuation anal-
540   ysis of nonstationary time series, Physica A: Statistical Mechanics and its Applications, 316, 87–114, https://doi.org/10.1016/S0378-4371(02)01383-3, https://linkinghub.elsevier.com/retrieve/pii/S0378437102013833, 2002.

Koscielny-Bunde, E., Bunde, A., Havlin, S., and Goldreich, Y.: Analysis of daily temperature fluctuations, Physica A: Statistical Mechanics and its Applications, 231, 393–396, https://doi.org/10.1016/0378-4371(96)00187-2, https://linkinghub.elsevier.com/retrieve/pii/0378437196001872, 1996.

545 Kunz, T., Dolman, A. M., and Laepple, T.: A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records – Part 1: Theoretical concept, Climate of the Past, 16, 1469–1492, https://doi.org/https://doi.org/10.5194/cp-16-1469-2020, https://cp.copernicus.org/articles/16/1469/2020/, publisher: Copernicus GmbH, 2020.

Laepple, T. and Huybers, P.: Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability, Earth and Planetary Science Letters, 375, 418–429, https://doi.org/10.1016/j.epsl.2013.06.006, http://www.sciencedirect.com/science/
550   article/pii/S0012821X13003221, 2013.

Laepple, T. and Huybers, P.: Global and regional variability in marine surface temperatures, Geophysical Research Letters, 41, 2528–2534, https://doi.org/10.1002/2014GL059345, http://doi.wiley.com/10.1002/2014GL059345, 2014a.

Laepple, T. and Huybers, P.: Ocean surface temperature variability: Large model–data differences at decadal and longer periods, Proceedings of the National Academy of Sciences, 111, 16 682–16 687, https://doi.org/10.1073/pnas.1412077111, http://www.pnas.org/lookup/doi/10.1073/pnas.1412077111, 2014b.

Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, Astrophysics and Space Science, 39, 447–462, https://doi.org/10.1007/BF00648343, https://doi.org/10.1007/BF00648343, 1976.

Lovejoy, S.: A voyage through scales, a missing quadrillion and why the climate is not what you expect, Climate Dynamics, 44, 3187–3210, https://doi.org/10.1007/s00382-014-2324-0, https://doi.org/10.1007/s00382-014-2324-0, 2015.

Lovejoy, S. and Lambert, F.: Spiky fluctuations and scaling in high-resolution EPICA ice core dust fluxes, Climate of the Past, 15, 1999–2017, https://doi.org/10.5194/cp-15-1999-2019, https://cp.copernicus.org/articles/15/1999/2019/, 2019.

Lovejoy, S. and Schertzer, D.: Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, Nonlinear Processes in Geophysics, 19, 513–527, https://doi.org/https://doi.org/10.5194/npg-19-513-2012, https://www.nonlin-processes-geophys.net/19/513/2012/, 2012.

Lovejoy, S. and Schertzer, D.: Low-Frequency Weather and the Emergence of the Climate, in: Extreme Events and Natural Hazards: The Complexity Perspective, pp. 231–254, American Geophysical Union (AGU), https://doi.org/10.1029/2011GM001087, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GM001087, 2013.

Lovejoy, S. and Varotsos, C.: Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings, Earth System Dynamics, 7, 133–150, https://doi.org/10.5194/esd-7-133-2016, https://esd.copernicus.org/articles/7/133/2016/, 2016.

Lovejoy, S., del Rio Amador, L., and Hébert, R.: The ScaLIng Macroweather Model (SLIMM): using scaling to forecast global-scale macroweather from months to decades, Earth System Dynamics, 6, 637–658, https://doi.org/https://doi.org/10.5194/esd-6-637-2015, https://esd.copernicus.org/articles/6/637/2015/, 2015.

Lovejoy, S., Procyk, R., Hébert, R., and del rio Amador, L.: The Fractional Energy Balance Equation, Quarterly Journal of the Royal Meteorological Society, accepted, 2021.

Malamud, B. D. and Turcotte, D. L.: Self-Affine Time Series: I. Generation and Analyses, in: Advances in Geophysics, edited by Dmowska, R. and Saltzman, B., vol. 40 of *Long-Range Persistence in Geophysical Time Series*, pp. 1–90, Elsevier, https://doi.org/10.1016/S0065-2687(08)60293-9, http://www.sciencedirect.com/science/article/pii/S0065268708602939, 1999.

Mandelbrot, B. B.: A Fast Fractional Gaussian Noise Generator, Water Resources Research, 7, 543–553, https://doi.org/10.1029/WR007i003p00543, http://doi.wiley.com/10.1029/WR007i003p00543, 1971.

Mandelbrot, B. B. and Van Ness, J. W.: Fractional Brownian Motions, Fractional Noises and Applications, SIAM Review, 10, 422–437, https://www.jstor.org/stable/2027184, 1968.

Mandelbrot, B. B. and Wallis, J. R.: Noah, Joseph, and Operational Hydrology, Water Resources Research, 4, 909–918, https://doi.org/https://doi.org/10.1029/WR004i005p00909, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR004i005p00909, 1968.

Mitchell, J. et al.: An overview of climatic variability and its causal mechanisms, Quaternary Research, 6, 481–493, 1976.

Molz, F. J., Liu, H. H., and Szulga, J.: Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions, Water Resources Research, 33, 2273–2286, https://doi.org/https://doi.org/10.1029/97WR01982, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97WR01982, 1997.

Munteanu, C., Negrea, C., Echim, M., and Mursula, K.: Effect of data gaps: comparison of different spectral analysis methods, Annales Geophysicae, 34, 437–449, https://doi.org/10.5194/angeo-34-437-2016, https://angeo.copernicus.org/articles/34/437/2016/, 2016.

Nilsen, T., Rypdal, K., and Fredriksen, H.-B.: Are there multiple scaling regimes in Holocene temperature records?, Earth System Dynamics, 7, 419–439, https://doi.org/10.5194/esd-7-419-2016, https://www.earth-syst-dynam.net/7/419/2016/, 2016.

Pelletier, J. D. and Turcotte, D. L.: Self-Affine Time Series: II. Applications and Models, in: Advances in Geophysics, edited by Dmowska, R. and Saltzman, B., vol. 40 of *Long-Range Persistence in Geophysical Time Series*, pp. 91–166, Elsevier, https://doi.org/10.1016/S0065-2687(08)60294-0, http://www.sciencedirect.com/science/article/pii/S0065268708602940, 1999.

Peng, C., Havlin, S., Stanley, H. E., and Goldberger, A. L.: Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos: An Interdisciplinary Journal of Nonlinear Science, 5, 82–87, https://doi.org/10.1063/1.166141, http://aip.scitation.org/doi/10.1063/1.166141, 1995.

Percival, D. B. and Walden, A. T.: Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques, Cambridge University Press, Cambridge, UK, 1993.

Prieto, G., Parker, R., Thomson, D., Vernon, F., and Graham, R.: Reducing the bias of multitaper spectrum estimates, Geophysical Journal International, 171, 1269–1281, 2007.

Rehfeld, K. and Kurths, J.: Similarity estimators for irregular and age-uncertain time series, Clim. Past, 10, 107–122, https://doi.org/10.5194/cp-10-107-2014, http://www.clim-past.net/10/107/2014/, 2014.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics, 18, 389–404, https://doi.org/10.5194/npg-18-389-2011, https://npg.copernicus.org/articles/18/389/2011/, 2011.

Rehfeld, K., Münch, T., Ho, S. L., and Laepple, T.: Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene, Nature, 554, 356–359, https://doi.org/10.1038/nature25454, http://www.nature.com/doifinder/10.1038/nature25454, 2018.

Reschke, M., Kunz, T., and Laepple, T.: Comparing methods for analysing time scale dependent correlations in irregularly sampled time series data, Computers & Geosciences, 123, 65–72, https://doi.org/10.1016/j.cageo.2018.11.009, http://www.sciencedirect.com/science/article/pii/S0098300418304308, 2019a.

Reschke, M., Rehfeld, K., and Laepple, T.: Empirical estimate of the signal content of Holocene temperature proxy records, Climate of the Past, 15, 521–537, https://doi.org/10.5194/cp-15-521-2019, https://www.clim-past.net/15/521/2019/, 2019b.

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 $\delta18O$ record, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/https://doi.org/10.1029/2010JD014764, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014764, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD014764, 2011.

Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL025591, http://www.agu.org/pubs/crossref/2006/2005GL025591.shtml, 2006.

Scargle, J. D.: Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data, The Astrophysical Journal, 263, 835, https://doi.org/10.1086/160554, http://adsabs.harvard.edu/doi/10.1086/160554, 1982.

Schertzer, D. and Lovejoy, S.: Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, Journal of Geophysical Research, 92, 9693, https://doi.org/10.1029/JD092iD08p09693, http://doi.wiley.com/10.1029/JD092iD08p09693, 1987.

Schimmel, M.: Emphasizing Difficulties in the Detection of Rhythms with Lomb-Scargle Periodograms, Biological Rhythm Research, 32, 341–346, https://doi.org/10.1076/brhm.32.3.341.1340, https://doi.org/10.1076/brhm.32.3.341.1340, publisher: Taylor & Francis _eprint: https://doi.org/10.1076/brhm.32.3.341.1340, 2001.

Schmitt, F., Lovejoy, S., and Schertzer, D.: Multifractal analysis of the Greenland Ice-Core Project climate data, Geophysical Research Letters, 22, 1689–1692, https://doi.org/https://doi.org/10.1029/95GL01522, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95GL01522, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/95GL01522, 1995.

Schulz, M. and Mudelsee, M.: REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series, Computers & Geosciences, 28, 421–426, https://doi.org/10.1016/S0098-3004(01)00044-9, https://www.sciencedirect.com/science/article/pii/S0098300401000449, 2002.

Schulz, M. and Stattegger, K.: Spectrum: spectral analysis of unevenly spaced paleoclimatic time series, Computers & Geosciences, 23, 929–945, https://doi.org/10.1016/S0098-3004(97)00087-3, https://www.sciencedirect.com/science/article/pii/S0098300497000873, 1997.

Schuster, A.: On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, Terrestrial Magnetism, 3, 13–41, https://doi.org/10.1029/TM003i001p00013, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/TM003i001p00013, 1898.

Shao, Z.-G. and Ditlevsen, P. D.: Contrasting scaling properties of interglacial and glacial climates, Nature Communications, 7, 10 951, https://doi.org/10.1038/ncomms10951, http://www.nature.com.ezp-prod1.hul.harvard.edu/ncomms/2016/160316/ncomms10951/full/ncomms10951.html, 2016.

Slepian, D. and Pollak, H. O.: Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - I, Bell System Technical Journal, 40, 43–63, https://doi.org/10.1002/j.1538-7305.1961.tb03976.x, https://ieeexplore.ieee.org/document/6773659, 1961.

Smith, J. O.: Physical audio signal processing: for virtual musical instruments and audio effects, W3K Publishing, Lexington, oCLC: 774174525, 2011.

Stevenson, S., McGregor, H. V., Phipps, S. J., and Fox-Kemper, B.: Quantifying errors in coral-based ENSO estimates: Toward improved forward modeling of δ^{18}O: IMPROVED CORAL FORWARD MODELING, Paleoceanography, 28, 633–649, https://doi.org/10.1002/palo.20059, http://doi.wiley.com/10.1002/palo.20059, 2013.

Storch, H. v. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, 1 edn., https://doi.org/10.1017/CBO9780511612336, https://www.cambridge.org/core/product/identifier/9780511612336/type/book, 1984.

Thomson, D.: Spectrum estimation and harmonic analysis, Proceedings of the IEEE, 70, 1055–1096, https://doi.org/10.1109/PROC.1982.12433, 1982.

Trauth, M. H.: MATLAB® recipes for earth sciences, Springer, oCLC: 1230144419, 2020.

VanderPlas, J. T.: Understanding the Lomb–Scargle Periodogram, The Astrophysical Journal Supplement Series, 236, 16, https://doi.org/10.3847/1538-4365/aab766, https://doi.org/10.3847/1538-4365/aab766, publisher: American Astronomical Society, 2018.

Vyushin, D., Mayer, J., and Kushner, P.: Spectral Analysis of Time Series, https://www.atmosp.physics.utoronto.ca/people/vyushin/mysoftware.html, 2009.

Zhu, F., Emile-Geay, J., McKay, N. P., Hakim, G. J., Khider, D., Ault, T. R., Steig, E. J., Dee, S., and Kirchner, J. W.: Climate models can correctly simulate the continuum of global-average temperature variability, Proceedings of the National Academy of Sciences, 116, 8728–8733, https://doi.org/10.1073/pnas.1809959116, https://www.pnas.org/content/116/18/8728, 2019.