



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

Thesis Draft

GEOGRAPHIC VARIABILITY OF DUST AND TEMPERATURE IN MACROWEATHER AND CLIMATE SCALING REGIMES.

NICOLÁS ACUÑA REYES

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
ELWIN VAN'T WOUT

Santiago de Chile, June 2023

© MMXV, NICOLÁS ACUÑA REYES



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

Thesis Draft

GEOGRAPHIC VARIABILITY OF DUST AND TEMPERATURE IN MACROWEATHER AND CLIMATE SCALING REGIMES.

NICOLÁS ACUÑA REYES

Members of the Committee:

ELWIN VAN'T WOUT

FABRICE LAMBERT

ANASTASIOS MATZAVINOS

CRISTIAN TEJOS

SHAUN LOVEJOY

WERNHER BREVIS

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, June 2023

© MMXV, NICOLÁS ACUÑA REYES

*Gratefully to life and his
mysterious ways of caring for us*

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	x
ABSTRACT	xi
RESUMEN	xii
1. INTRODUCTION	1
2. Hypothesis and objectives	2
3. Mathematical formulation	3
3.1. Haar fluctuation	3
3.2. Missing data	3
3.3. Spearman correlation	4
3.4. Non-uniform Fast Fourier Transform (NUFFT)	6
3.5. Haar fluctuation and spectra relation	8
3.6. Cubic smoothing spline	8
4. Algorithms and functions	10
4.1. Haar function for equidistant datasets.	10
4.2. Haar function for non-equidistant datasets.	12
4.3. Non-uniform Fast Fourier Transform	15
5. Synthetic datasets and algorithm validation	16
5.1. Sinusoidal dataset	16
5.2. Composite paleotemperature dataset	22
5.2.1. Datalength test	26
5.2.2. Noise test	27
5.2.3. Components amplitude test	29

Thesis Draft

5.2.4. Extra LF component test	31
5.2.5. Linear interpolation test	32
6. Paleoclimate datasets	34
6.1. Paleoclimate dataset description	34
6.2. Datasets summary	35
6.3. Dataset preprocessing	39
7. Haar fluctuation analysis for equidistant datasets	43
7.1. Complete datasets	43
7.2. Last Glacial Cycle (LGC)	46
7.3. Holocene and Last Glacial Maximum	49
7.4. Sensitivity analysis	52
8. Comparison between Haar functions for equidistant and non-equidistant datasets	55
9. Climatic interpretation	57
9.1. Dataset correlation	57
9.2. LGC Haar fluctuation analysis: Non-equidistant case	58
9.3. Spectra analysis: Non-Uniform Fast Fourier Transform	62
9.4. Timescale transition and latitude	65
10. CONCLUSIONS	69
REFERENCES	70
Appendix A	75

LIST OF FIGURES

4.1	Haar function for equidistant datasets outcome	11
4.2	Calibration parameter	13
4.3	Lake Tanganyika fluctuations as scatter cloud and its cubic smoothing spline .	14
5.1	Sinusoidal data sampling	16
5.2	Sinusoidal functions.	17
5.3	Noisy sinusoidal functions.	17
5.4	Fluctuations RMS from sinusoidal function 1	18
5.5	Fluctuations RMS from sinusoidal function 2	19
5.6	Fluctuations RMS from sinusoidal function 3	20
5.7	Fluctuations RMS from sinusoidal function 4 and 5	21
5.8	HF components of the synthetic paleotemperature signal	23
5.9	LF components of the synthetic paleotemperature signal	23
5.10	200 Kyear synthetic paleotemperature wave	24
5.11	Non-uniform sampling of the synthetic paleotemperature signal	25
5.12	Five segments fluctuations RMS comparison	27
5.13	Near-white and near-red noise components used in the <i>Noise test</i>	28
5.14	Original and noisy paleotemperature signal fluctuations RMS	28
5.15	High and low frequency results from the <i>Component amplitude test</i>	29
5.16	Paleotemperature fluctuations RMS after the addition of a 4th LF component	31

5.17	Fluctuations RMS from the first 10ky of the paleotemperature signal under uniform, non-uniform, and interpolated sampling	32
6.1	Global proxies distribution	35
6.2	Normalized resolution: Dust/Dust flux datasets.	37
6.3	Normalized resolution: Temperatures datasets.	38
6.4	<i>Bottom resolution</i> estimation for Vostok dust dataset	40
6.5	Original, resampled and interpolated ML1208-17PC dataset	41
7.1	Complete uniform temperature datasets Haar fluctuations.	44
7.2	Complete uniform dust datasets Haar fluctuations.	45
7.3	Central California original non-uniform resolution comparison: Complete dataset and Last Glacial Cycle	47
7.4	Last Glacial Cycle uniform temperature datasets Haar fluctuations.	48
7.5	Last Glacial Cycle uniform dust datasets Haar fluctuations.	48
7.6	Holocene uniform temperature datasets Haar fluctuations.	50
7.7	Holocene uniform dust datasets Haar fluctuations.	50
7.8	Last Glacial Maximum uniform temperature datasets Haar fluctuations.	51
7.9	Last Glacial Maximum uniform dust datasets Haar fluctuations.	52
7.10	Uniform Haar function sensitivity test 1: Resolution test	53
7.11	Uniform Haar function sensitivity test 2 and 3: Resolution & length test	54
8.1	Matlab and Python fluctuations RMS comparison	56
9.1	Spearman correlation coefficients	57

9.2	Vostok, EDC and Sanbao cave cycle comparison	58
9.3	Last Glacial Cycle temperatures Haar fluctuations RMS from non-equidistant datasets.	60
9.4	Last Glacial Cycle dust Haar fluctuations from non-equidistant datasets.	61
9.5	NUFFT from non-equidistant temperatures datasets.	63
9.6	NUFFT from non-equidistant dust datasets.	64
9.7	Normalised RMS fluctuations from dust and temperature by latitude.	66
9.8	RMS fluctuations from dust and temperature with slopes, clustered by latitude.	67
10.1	Detailed figure of fluctuations RMS from sinusoidal function 1	75
10.2	Detailed figure of fluctuations RMS from sinusoidal function 2	76
10.3	Detailed figure of fluctuations RMS from sinusoidal function 3	77
10.4	Detailed figure of fluctuations RMS from sinusoidal function 4	78
10.5	Detailed figure of fluctuations RMS from sinusoidal function 5	79

LIST OF TABLES

5.1	Non-uniform synthetic paleotemperature sampling	24
6.1	Datasets summary	36
6.2	Proxy Counter	37
6.3	Datasets DOI	39
6.4	Datasets bottom resolution	42
7.1	Last Glacial Cycle bottom resolutions	46
7.2	Holocene and Last Glacial Maximum bottom resolution	49
7.3	Uniform Haar function sensitivity test	53
9.1	τ_c and maximum summary	68

ABSTRACT

To understand the numerous events happening in our atmosphere, we need to separate and study them into different time scales or regimes, from which the *macroweather* regime covers time scales between 20 days and 50 years, while the *climate* regime goes from 50 to 80.000 years. In this work, we conducted a Haar fluctuation analysis using temperature and dust deposition datasets from 24 locations of the world, thus being able to assess the global atmospheric variability for the Last Glacial Cycle. To do so, we developed an algorithm to work with non uniform datasets since this is the raw format of paleoclimatic data.

We also compared the outcomes of our interpolation-free algorithm with those from an interpolation-dependent function, which biased the results as it filters the data. In addition, we executed a spectral analysis using a Non-Uniform Fast Fourier Transform function to complement our study but concluded that our real-time approach was more suitable to give us more insights, especially on spiky data sets.

Thus we found that the atmosphere variability presents a geographical dependence related to the latitude but also to its marine or continental location, where Antarctica presented a macroweather to climate transition (τ_c) at timescales around 200-600y, while in Greenland it was around 100-300y. This value increases for the mid-latitudes up to 1800-2500y, whereas in the tropics it decreases again to around 630y.

Keywords: Haar function, Atmosphere, Climate variability, Multi-scale analysis.

RESUMEN

Para entender los numerosos eventos que ocurren continuamente en la atmósfera, es necesario separarlos y estudiarlos en diferentes escalas temporales o regímenes, dentro de los cuales el régimen de *macrotiempo* abarca entre 20 días y 50 años, mientras que el régimen de *clima* va de 50 a 80.000 años. En este trabajo, realizamos un análisis de fluctuación de Haar utilizando la temperatura y la deposición de polvo de 24 sitios del mundo, pudiendo así evaluar la variabilidad atmosférica global para el Último Ciclo Glacial. Para hacer esto, desarrollamos un algoritmo capaz de trabajar con conjuntos de datos no equidistantes, ya que este es el formato bruto de los datos paleoclimáticos.

También comparamos los resultados de nuestro algoritmo libre de interpolación con los de una función dependiente de interpolación, el cual sesga los resultados a medida que filtra los datos. Además, ejecutamos un análisis espectral utilizando una Función de Transformada Rápida de Fourier No Uniforme para complementar nuestro estudio, pero dada la naturaleza altamente fluctuante de los datos, no logramos sacar ninguna conclusión complementaria.

Así, encontramos que la variabilidad de la atmósfera presenta una dependencia geográfica relacionada con la latitud pero también con su ubicación marina o continental, donde Antártica presentó una transición de macrotiempo a clima (τ_c) en escalas de tiempo alrededor de los 200-600 años, mientras que en Groenlandia esta fue alrededor de 100-300 años. Este valor aumenta para las latitudes medias hasta 1800-2500y, mientras que en los trópicos vuelve a disminuir hasta alrededor de 630y.

Palabras Claves: Función de Haar, Atmósfera, Variabilidad climática, Análisis multi-escala.

1. INTRODUCTION

Understanding the systematic changes in the Earth’s atmosphere is a complex task to address. First, the number of relevant variables involved is large. Second, there are still many open questions regarding land, ocean, and atmosphere interactions, which makes the latter highly turbulent and chaotic. While climate systems exhibit variability across all spatial and temporal ranges, decadal to millennial scales are of special interest since their natural variability interacts and overlaps with anthropogenic forcing, thus being of great importance for actual epoch understanding and future climate projections.

Temperature variability has been shown to scale according to power law relationships, leading to the distinction between five distinct power law scaling regimes (weather, macroweather, climate, macroclimate, and megaclimate) based on different scaling exponents for the relationship $E(\omega) \sim \omega^{-\beta}$, where E denotes the spectral energy and ω its frequency. Naturally, spectral analysis has become a pretty common technique to study fluctuations, but the highly non-Gaussian spikiness of the datasets typically implies strong spectral spikes whose origin is purely random, thus making its physical interpretations nontrivial.

To work with spiky data, in this work we study atmospheric variability in time domain rather than frequency domain by using Haar fluctuations. The latter quantify the absolute difference of the mean over the first and second halves of an interval, allowing us to define variability as a function of scale $\langle |\Delta F(\Delta t)| \rangle \propto \Delta t^H$, where the H exponent indicates if fluctuations increase ($H > 0$) or decreases ($H < 0$) with scale Δt . To do so, we used temperature and dust fluxes datasets from different sites of the world since both are well correlated at orbital frequencies and hence should both respect the same scale symmetry.

Nevertheless, Haar wavelets until now have been used on uniform datasets, while paleoclimatic datasets are non-uniform since their resolution decreases as we move to the past. Therefore, we have designed an algorithm to develop Haar wavelet analysis on non-uniform datasets, thus improving the results.

2. HYPOTHESIS AND OBJECTIVES

This work hypothesizes that the timescale transition from macroweather to climate (τ_c) has a spatial variability. In particular, we expect lower τ_c in the poles rather than the equator or the tropics since the ocean presents a slower reaction to climate forcings.

Under this hypothesis, our general objective is to implement an algorithm for studying non-equidistant datasets via Haar fluctuation analysis, allowing us to identify (τ_c) for different locations of the world.

The specific objectives are as follows:

- (i) To validate the algorithm of non-uniform Haar wavelet analysis on synthetic datasets.
- (ii) To create a suitable global paleoclimatic database to study its atmospheric variability on a decennial-millennial timescale.
- (iii) To find and compare the τ_c 's obtained from frequency analysis and fluctuation analysis.
- (iv) To compare the τ_c 's for different latitudes of the world.

3. MATHEMATICAL FORMULATION

3.1. Haar fluctuation

Haar fluctuations are a useful tool to characterize variability in time domain of a series $F(t)$, both because of the simplicity of its definition and interpretation. To determine the Haar fluctuation over a time interval Δt , we need to compute the absolute difference of the mean over the first and second halves of an interval:

$$|\Delta F(\Delta t)| = \left| \frac{1}{\frac{\Delta t}{2}} \left(\int_{t-\Delta t/2}^t F(t') dt' - \int_{t-\Delta t}^{t-\Delta t/2} F(t') dt' \right) \right|. \quad (3.1)$$

If over a range of timescales Δt there is a power law relation ($\Delta F(\Delta t) \propto \Delta t^H$), then the mean absolute fluctuation also varies as

$$\langle |\Delta F(\Delta t)| \rangle \propto \Delta t^H \quad (3.2)$$

where “ $\langle \rangle$ ” indicates ensemble average, i.e., average over all the available disjointed intervals. Here, $H > 0$ means that the average fluctuations increase with scale, while $H < 0$ implies the opposite. However, this exponent is veracious only within the range $H \in [-1, 1]$, which is valid for almost all geodata analysed to date (Lovejoy & Lambert, 2019).

3.2. Missing data

To apply the standard Haar wavelet analysis, we need a uniform dataset. However, paleoclimate archive may have missing data due to possible sample loss, contamination or bad core quality. Linear interpolation is one common alternative to fill the missed values, both because of its ease of implementation and its good approximation to the average value.

If the two missed value neighbors' points are known and given by the coordinates (x_0, y_0) and (x_1, y_1) , we can estimate the intermediate missing point (x, y) from the equation of slopes,

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}, \quad (3.3)$$

which is a special case of polynomial interpolation with $n = 1$. Solving this equation for y gives the formula for linear interpolation in the interval (x_0, x_1) ,

$$y = \frac{y_0(x_1 - x) + y_1(x - x_0)}{x_1 - x_0}. \quad (3.4)$$

This technique can also be applied to a set of missing data points $(x_0, y_0), \dots, (x_n, y_n)$ by doing a concatenation of linear interpolations between each pair of data points. Nevertheless, since linear interpolation is a low-pass filter, it will always imply some degree of information loss, which is especially important when working with highly intermittent datasets.

3.3. Spearman correlation

To compare different datasets, we use the Spearman correlation. This coefficient measures how well the relationship between two variables can be described using a monotonic function. While Pearson's correlation (the usual correlation) assesses linear relationships, Spearman's correlation assesses monotonic relationships, being equal to the Pearson correlation between the rank values of those two variables. This coefficient can be computed as

$$r_s = \frac{\text{cov}(R(X)R(Y))}{\sigma_{R_X}\sigma_{R_Y}}, \quad (3.5)$$

where $\text{cov}(R(X)R(Y))$ and $\sigma_{R_X}\sigma_{R_Y}$ denotes the covariance and standard deviations of the rank variables, respectively.

As the datasets covers different ages using different number of samples, we needed to define a set of rules to compare them pairwise over a common timescale. Thus, let's $\mathcal{X} =$

$\{D_{1,L_1}, \dots, D_{n,L_n}\}$ be the set of available datasets in the form $D_{i,L_i} = (A_{i,l}, V_{i,l})_{1 \leq l \leq L_i}$, where $A_{i,l}$ and $V_{i,l}$ denote the age and value measured of the set i at the sample l . Then:

1. Select 2 datasets $D_{i,L_i}, D_{j,L_j} \in \mathcal{X}$ to work with.

2. Store the minimum maximum age as A_{\max} :

$$A_{\max} = \min(A_{i,L_i}, A_{j,L_j}), \quad (3.6)$$

3. Truncate the sets leaving only the rows which age is lower or equal to A_{\max} , i.e., for $z = \{i, j\}$ and $l \in \{1, \dots, L_z\}$:

$$D_{z,l} = \begin{cases} \text{Holds,} & \text{if } A_{z,l} \leq A_{\max}, \\ \text{Cut out,} & \text{any other case.} \end{cases} \quad (3.7)$$

We will call $l_{z-\max}$ the new maximum age index, i.e. $l_{z-\max} = \max(l, \text{ such that } A_{z,l} \leq A_{\max})$.

4. Since it is highly probable that for all row k , $A_{i,l_k} \neq A_{j,l_k}$, we will interpolate the set with higher resolution and evaluate it at the ages of the lower resolution set. To do so, we define a common interval \mathcal{I} (in time) over which we will compute the correlation.

Let's define $\mathcal{I} = [\text{starting age}, \text{stopping age}]$ and suppose $D_{j,l_{j\max}}$ is the higher resolution set after truncation, then:

- if $A_{i,1} < A_{j,1} \rightarrow \text{Starting age} = \min\{l, \text{ such that } A_{i,l} \geq A_{j,1}\}$
- if $A_{i,1} \geq A_{j,1} \rightarrow \text{Starting age} = A_{i,1}$
- if $A_{i,l_{i\max}} \leq A_{j,l_{j\max}} \rightarrow \text{Stopping age} = A_{i,l_{i\max}}$
- if $A_{i,l_{i\max}} > A_{j,l_{j\max}} \rightarrow \text{Stopping age} = \max\{l, \text{ such that } A_{i,l} \leq A_{j,l_{j\max}}\}$

5. Use linear interpolation (\mathcal{L}) to evaluate the higher resolution set at the ages of the lower resolution one, i.e., $V_{j,k} = \mathcal{L}_{D_{j,l_{j\max}}}(A_{i,k}), \forall A_{i,k} \in \mathcal{I}$.

6. Normalize both datasets by doing $\frac{V_z - \mu_z}{\sigma_z}$, where μ_z and σ_z represent the mean and standard deviation of the truncated datasets values $(V_z)_{z=\{i,j\}}$, respectively.
7. Compute the Spearman's correlation coefficient between the sets $(V_z)_{z=\{i,j\}}$ following equation (3.5).

3.4. Non-uniform Fast Fourier Transform (NUFFT)

Under a scaling regime, the spectrum follows a power law form as,

$$E(\omega) \approx \omega^{-\beta}, \quad (3.8)$$

where the spectral exponent β characterizes the spectral density. Even though the physical interpretation of spectra is not obvious (Lovejoy, 2015) it can be a good complementary tool to better understand fluctuations. Nevertheless, as paleoclimatic datasets are usually not equidistant in time, the widely known Fast Fourier Transform must be modified to work with nonuniform datasets.

Hence, for a given $N \in \mathbb{N}$, let's take the index set:

$$I_N = \{\mathbf{k} \in \mathbb{Z} : -\frac{N}{2} \leq \mathbf{k} \leq \frac{N}{2}\}.$$

We are interested in evaluating the 2π -periodic trigonometric polynomial

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in I_N} \hat{f}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}}, \quad \hat{f}_{\mathbf{k}} \in \mathbb{C}, \quad (3.9)$$

at arbitrary knots \mathbf{x}_j , $j \in I_M$ for given arbitrary coefficients $\hat{f}_{\mathbf{k}} \in \mathbb{C}$, $\mathbf{k} \in I_N$. To do so, the idea is to approximate $f(\mathbf{x})$ by a linear combination of translates of a convenient window function, to then evaluate the obtained approximation at the knots \mathbf{x}_j . Thus, we can define a 2π -periodic function such that it can be expressed as the Fourier series:

$$\tilde{\varphi}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}} c_{\mathbf{k}}(\tilde{\varphi}) e^{i\mathbf{k} \cdot \mathbf{x}}. \quad (3.10)$$

Now, let's take an oversampling factor $\sigma \geq 1$ such that $\sigma N \in \mathbb{N}$, and determine the coefficients $g_{\mathbf{l}} \in \mathbb{C}$, $\mathbf{l} \in I_{\sigma N}$, of the linear combination:

$$s_1(\mathbf{x}) = \sum_{\mathbf{l} \in I_{\sigma N}} g_{\mathbf{l}} \tilde{\varphi}\left(\mathbf{x} - \frac{2\pi\mathbf{l}}{\sigma N}\right) \quad (3.11)$$

such that s_1 is an approximation of the polynomial (3.9). It is shown in (Plonka, Potts, Steidl, & Tasche, 2018) that the Fourier series of the 2π -periodic function s_1 has the discrete Fourier coefficients:

$$\hat{g}_{\mathbf{k}} = \sum_{\mathbf{l} \in I_{\sigma N}} g_{\mathbf{l}} e^{-2\pi i \mathbf{k} \cdot \mathbf{l}/(\sigma N)}, \quad (3.12)$$

hence, by applying the inverse FFT we can compute the coefficients $g_{\mathbf{l}}$ obtaining:

$$g_{\mathbf{l}} = \frac{1}{(\sigma N)} \sum_{\mathbf{k} \in I_N} \hat{g}_{\mathbf{k}} e^{2\pi i \mathbf{k} \cdot \mathbf{l}/(\sigma N)}, \quad \mathbf{l} \in I_{\sigma N}. \quad (3.13)$$

In addition, if we assume that the function φ can be well approximated by its truncation $\psi = \varphi|Q$ on $Q = [-\frac{2\pi m}{\sigma N}, \frac{2\pi m}{\sigma N}]$ where $m \in \mathbb{N}$, that is:

$$\psi(\mathbf{x}) = \begin{cases} \varphi(\mathbf{x}) & \mathbf{x} \in Q, \\ 0 & \mathbf{x} \in \mathbb{R} \setminus Q, \end{cases} \quad (3.14)$$

hence, we can now approximate s_1 by the 2π -periodic function

$$s(\mathbf{x}) = \sum_{\mathbf{l} \in I_{\sigma N}} g_{\mathbf{l}} \tilde{\psi}\left(\mathbf{x} - \frac{2\pi\mathbf{l}}{\sigma N}\right), \quad (3.15)$$

which for a fixed knot \mathbf{x}_j contains at most $(2m + 1)$ nonzero terms, obtaining:

$$f(\mathbf{x}_j) \approx s_1(\mathbf{x}_j) \approx s(\mathbf{x}_j). \quad (3.16)$$

Thus, equation (3.9) can be approximately computed for all \mathbf{x}_j , $j \in I_M$ by the double approximations, s_1 and s .

3.5. Haar fluctuation and spectra relation

If we expand the definition of equation (3.2) to a more general form, we can consider other statistical moments as generalized structured functions $S_q(\Delta t)$:

$$S_q(\Delta t) = \langle |\Delta F(\Delta t)|^q \rangle \propto \Delta t^{\xi(q)} \quad \text{Thesis Draft} \quad (3.17)$$

where $\xi(q)$ is concave and non-linear due to fluctuation intermittency. By decomposing $\xi(q)$ into a linear and a nonlinear component, we can write it as

$$\xi(q) = qH - K(q). \quad (3.18)$$

As we mentioned above, in a scaling process of frequencies ω the spectrum follows a power law:

$$E(\omega) \approx \omega^{-\beta}, \quad (3.19)$$

where the spectral exponent $\beta = 1 + \xi(2) = 1 + 2H - K(2)$ characterizes the spectral density. If we ignore the intermittency, i.e. $K(2) = 0$, then we get $H = (\beta - 1)/2$, showing that $\beta > 1$ imply $H > 1$, hence fluctuations growth with scale, while if $\beta < 1$, $H < 1$ and fluctuation decrease with scale.

For that reason, fluctuation and frequency analysis are good complements and may help each other to better understand the process behavior. Nevertheless, we should be always aware of the approximations done in the analysis (e.g. $K(2) = 0$) and hence take it as tool to be interpreted by an expert eye. For instance, if we assess a process with strong intermittency (typically $K(2) \approx 0.2$) we could get a high frequency dominance in the variance ($\beta < 1$) but a low frequency dominating the mean fluctuations ($H > 1$).

3.6. Cubic smoothing spline

In the final step of our Haar function algorithm, we use a cubic smoothing spline (CSAP) to smooth out the results and facilitate its interpretation. Although it could be

done by different techniques (e.g., binning the fluctuation by time lags and taking the mean at each interval), we empirically decided to stay with the CSAP as it seemed to deliver more accurate results.

Thus, we used the CSAP package available on Python, which is an implementation of the algorithm proposed in (Boor, 1978). Hence for a set (x, y) of length $n \in \mathbb{N}$, the smoothing spline f minimizes the expression:

$$p \sum_{j=1}^n w_j |y_j - f(x_j)|^2 + (1-p) \int \lambda(t) |D^2 f(t)|^2 dt, \quad (3.20)$$

where the sum (left term) is the error measure and the integral (right term) is the roughness measure. The default value of the error weight w_j is 1, while the default value of $\lambda(t)$ is the constant function 1. $D^2 f(t)$ denotes the second derivative of the function f .

The smoothing parameter p should be in the range $[0, 1]$ where:

- $p = 0$: The smoothing spline f is the line that fits the data by least-squares.
- $p = 1$: The smoothing spline f is the interpolation by natural cubic splines.

Usually the best choice of p is close to $1/(1 + h^3/6)$ where h is the mean distant between samples (Boor, 1978). We found that $p = 1$ results in very intermittent fluctuations impossible to analyse and interpret by eye , while values lower or equal to 0.999 smoothed the curves too much. Hence after some iterations we fixed $p = 0.9999$ which, despite being quite close to 1, was the more accurate value to smooth the data to visually interpret fluctuations without loosing too much information.

4. ALGORITHMS AND FUNCTIONS

4.1. Haar function for equidistant datasets.

To perform a Haar fluctuation analysis with equidistant datasets, we used the Matlab function available in <https://www.physics.mcgill.ca/~gang/software/index.html>, which computes the convolution between a Haar wavelet and the uniformly spaced datasets. This function takes as inputs a structured data string (“*field*” input), a calibration parameter (“*calibcoord*” input) needed to adjust the time-space axis, and the number of points at low and high time lags to drop before computing some cascade exponents (“*lowpts*” and “*highpts*” input).

As output, the Haar function finds the exponent $C1$ and α (codimension and multifractality index) and the scaling exponent H , among other cascade parameters that are beyond the scope of this work. In addition, the function displays five figures by default, from which we are interested in the 4th one which shows the $q=1$ and $q=2$ (root means square (RMS)) structure-functions (see equation 3.17).

To illustrate this, let’s use the EPICA Dome C dust flux dataset available in <https://doi.pangaea.de/10.1594/PANGAEA.778213>, which contains the 25-year constant time-step averages of the high-resolution 1 cm data. Thus, the *field* input will be a .csv file containing only the EDC dust flux column, without any depth or age information. Then, as the data has a regular 25 years resolution the *calibcoord* parameter will be 1/25, while *lowpts* and *highpts* needs to be adjusted to the range over which $C1$, α , and H will be estimated. These inputs are shown in the script of Example box 4.1.

Example box 4.1: Matlab Haar function script

```
>> Data = readtable('EDC_DustFlux_25yr.csv')
>> Array = table2array(Data)
>> Epica = transpose(Array)
>> Haar(Epica, 1/25, 1, 25)
```

Hence by following the lines of the example script, we can obtain C_1 , α , and H from a portion of the structured EDC dataset. Figure 4.1 shows the $q=1$ and $q=2$ curves displayed by the function, where the straight lines (of slope H) have been fitted over the working range [$lowpts$: Fluctuation length - $highpts$]. If it is not our range of interest, $lowpts$ and $highpts$ should be adjusted.

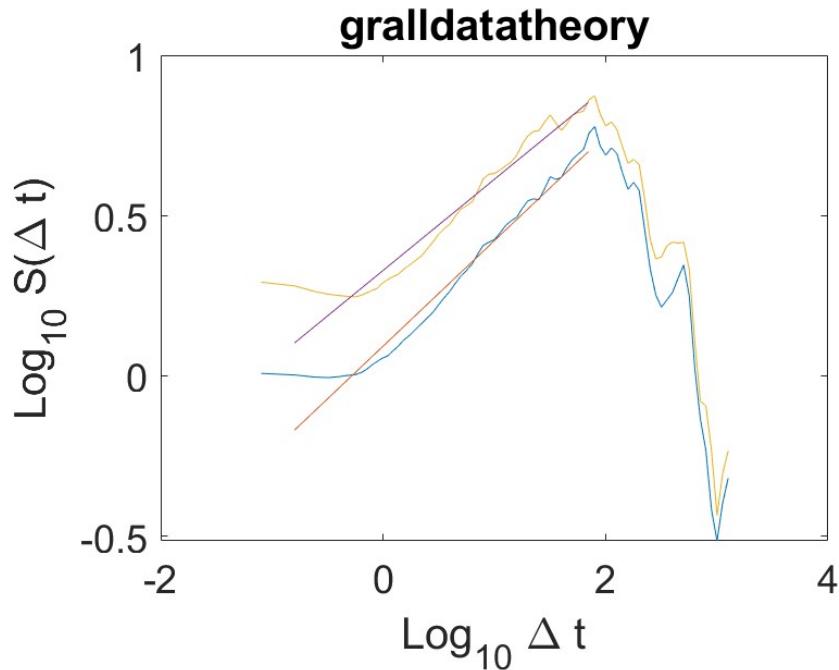


Figure 4.1. Haar function for equidistant datasets outcome: $q=1$ structure-function (blue) and $q=2$ RMS (yellow). Straight lines are slopes references fitted over the working range [$lowpts$: Fluctuation length - $highpts$].

4.2. Haar function for non-equidistant datasets.

To avoid linear interpolation and works with raw non-equidistant datasets, we implemented an interpolation-free algorithm in *Python* to perform the Haar function. We start from the hypothesis that over scaling ranges the statistics of the ensemble follows the scaling form:

$$\langle \Delta T(\lambda \Delta t)^q \rangle = \lambda^{\xi(q)} \langle \Delta T(\Delta t)^q \rangle, \quad (4.1)$$

where $\langle \Delta T(\Delta t)^q \rangle$ is the q^{th} order structure function, λ is a dimensionless scale ratio, “ $\langle \rangle$ ” is the ensemble average and $\xi(q)$ is the structure function exponent. According to the law of large numbers we can estimate the q^{th} order moment via the sample mean:

$$\overline{\Delta T(\Delta t)^q} = \frac{1}{n(\Delta t)} \sum_{i=1}^{n(\Delta t)} (\Delta T(\Delta t)_i)^q \quad (4.2)$$

where $n(\Delta t)$ is the number of disjoint intervals at lag Δt , and $(\Delta T(\Delta t)_i)$ is the i^{th} fluctuation of T at resolution Δt . Thus, at the large n limit, we obtain $\overline{\Delta T(\Delta t)^q} \approx \langle \Delta T(\Delta t)^q \rangle$.

Yet, as we are working with samples from a single realization (thus, not completely statistically independent) the law of large numbers will apply only if the process $T(t)$ is ergodic¹, which is generally not the case in multifractal processes. Nevertheless, as nonergodicity is only important for high enough order moments² (Lovejoy & Schertzer, 2013), we will restrict the analysis to statistics near the mean in order to use the sample means as estimator of the ensemble mean (eq. 4.2).

While the mean absolute fluctuation ($q = 1$) could be used for the analysis, the root mean square (RMS) exponent ($\xi(2)/2$) have become conventional as for Gaussian processes it suffices to fully characterize the scaling. Thus, in the following, we will use the RMS fluctuation as an approximation to the absolute mean ($H \approx \xi(2)/2$).

¹i.e. if temporal averages of a single sample of the process are the same as the average at one time over an ensemble of identical processes.

²As the dressed moments generally diverge for q -moments such that $q \geq q_D$, where q_D is the solution to the implicit codimension equation: $C(q_D) = D$.

Let assume now that there are N measurements of temperature $T(t_i)$, where t_i is time and i is an index from 1 through N . Let's define a running sum S_i as

$$S_i = \sum_{j \leq i} T(t_j), \quad (4.3)$$

and consider an index j and an even number k , which defines the interval $[t_j, t_{j+k}]$ where the fluctuation $\Delta T_{j,k}$ will be estimated. First, we determine the sums of $T(t_i)$ over the first and second halves interval:

$$\Delta S^{(1)} = S_{j+k/2} - S_j, \quad \Delta S^{(2)} = S_{j+k} - S_{j+k/2}, \quad (4.4)$$

then, the Haar fluctuation will correspond to the average of the first half minus the average of the second half of the interval, estimated as:

$$\Delta T_{j,k} = 2 \left(\frac{S^{(1)}}{t_{j+k/2} - t_j} - \frac{S^{(2)}}{t_{j+k} - t_{j+k/2}} \right), \quad (4.5)$$

where the canonical 2 multiplying the difference is a calibration parameter (C) necessary to expand the working range from anomaly (when $H < 0$) and differences (when $H > 0$) to the Haar range $-1 \leq H \leq 1$. Even though we could use a theoretic value of C (Figure 4.2), it might not be advantageous in real world applications as there may be zones of both positive and negative H present, thus, the compromise of $C = 2$ is an accurate decision (Lovejoy, Varotsos, & Lambert, 2018).

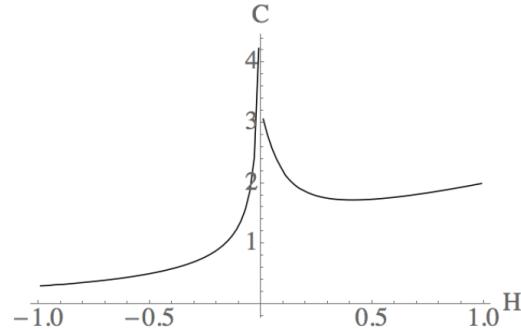


Figure 4.2. The theoretical calibration constant C for the RMS second moment. Adapted from Appendix A.2 in (Lovejoy et al., 2018)

Then, for a series $T(t_i)$ of length n , the algorithm will estimate and store $(\Delta T_{j,k})^2$ for all $k = \{2, 4, \dots, \text{int}(n/2)\}$ and for all $j = \{1, 2, \dots, n - k\}$. As this algorithm moves in the index space (j, k) and not in the time space, it is necessary to set a threshold regarding whether the fluctuation $\Delta T_{j,k}$ should be considered or not. A good option is to assess the ratio between time samples (ϵ) and to define a parameter ϵ_{\min} that can be adjusted to make the condition as restrictive as we like. In this way, we define the ratio

$$\epsilon = \frac{t_{j+k/2} - t_j}{t_{j+k} - t_{j+k/2}}, \quad (4.6)$$

and only keep the fluctuations within the range $\epsilon_{\min} \leq \epsilon \leq (1 - \epsilon_{\min})$. Decreasing ϵ_{\min} implies a precision loss and hence it smooths the resulting curve, on the other hand, taking ϵ_{\min} too close to 1/2 (uniform dataset case) can result in the rejection of too many fluctuations, with the consequence that the statistics will be poor. In our case, we have fixed ϵ_{\min} at 1/4.

Once fluctuations have been estimated and filtered, we finish with a scattered cloud of its squared values. To visualize the result in a log-log scale (easier to interpreted), we take the logarithm of the square root of the sample mean, and apply the cubic smoothing spline(CSAP) described in section 3.6 to reduce its spikiness. For instance, Figure 4.3 show fluctuations from Lake Tanganyika as a scatter cloud and its CSAP.

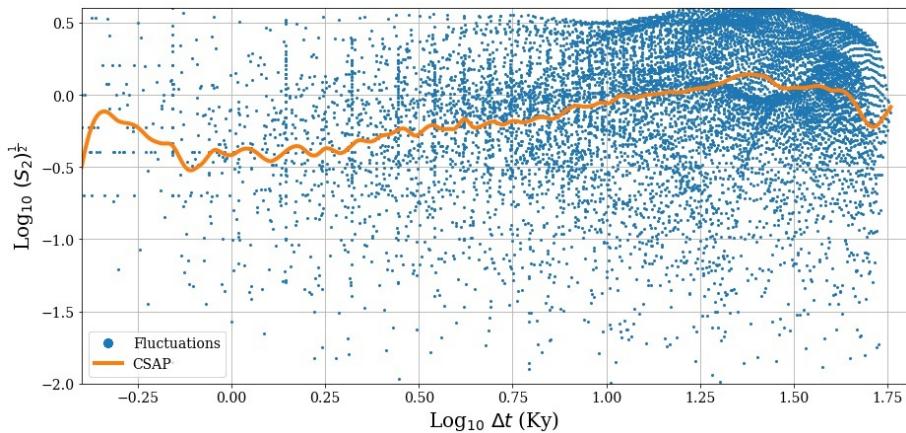


Figure 4.3. Lake Tanganyika fluctuations as scatter cloud (Blue) and its cubic smoothing spline (Orange).

We can also compute the uncertainties in fluctuations for fixed lags Δt_i by taking the standard deviation of the sample fluctuations, which quantifies the width of their probability distributions. Again, if the samples were statistically independent of each other, and assuming that the second moment exists, the central limit theorem implies that a measure of the spread of the distribution about the mean will be a Gaussian random variable with variance

$$V(\Delta t)^2 = \frac{1}{n(\Delta t) - 1} \sum_{i=1}^{n(\Delta t)} \left((\Delta T(\Delta t)_i) - \overline{\Delta T(\Delta t)} \right)^2, \quad (4.7)$$

thus, the ensemble mean associated at each interval Δt_i is estimated from the samples as:

$$\langle \Delta T(\Delta t) \rangle_i \approx \overline{\Delta T(\Delta t_i)} \pm \frac{V(\Delta t_i)}{n(\Delta t_i)^{1/2}}, \quad (4.8)$$

where the $n(\Delta t_i)^{1/2}$ normalization is from the Gaussian case of the central limit theorem. Nevertheless, as we take the assumption of statistical independence of the sample fluctuations (not the real case) all the estimations done are not strictly valid, and thus they should be considered as a helpful tool in a qualitative manner. For that reason, we will not show the uncertainties intervals in figures of the fluctuation analysis section.

4.3. Non-uniform Fast Fourier Transform

To perform the spectral analysis for non-equidistant datasets, we have used the Non-uniform Fast Fourier Transform function available in Matlab, which uses the syntax:

$$Y = \text{nufft}(X, t, f) \quad (4.9)$$

to compute the nonuniform discrete Fourier transform at the query points f using the sample points t . The variable X denotes the quantity of interest and can be specified as a vector, matrix, or multidimensional array. t are the sample points (here, ages), specified as a vector of length n . Finally, f correspond to the query points, specified as a vector. By default, the query points vector is $(0:(n-1)) / n$.

5. SYNTHETIC DATASETS AND ALGORITHM VALIDATION

5.1. Sinusoidal dataset

To validate our non-uniform Haar function algorithm, we tested it on synthetic data so we could easily judge if it was working correctly. Therefore, we defined a working range of 50 time units and sampled it at 0.1 units resolution, hence obtaining a 500 samples linear array to be used as equidistant time set (Figure 5.1a). From this array we randomly chose 300 samples to be used as non-uniform time set (Figure 5.1b), so we could compare the algorithm outcome under both structured and unstructured datasets.

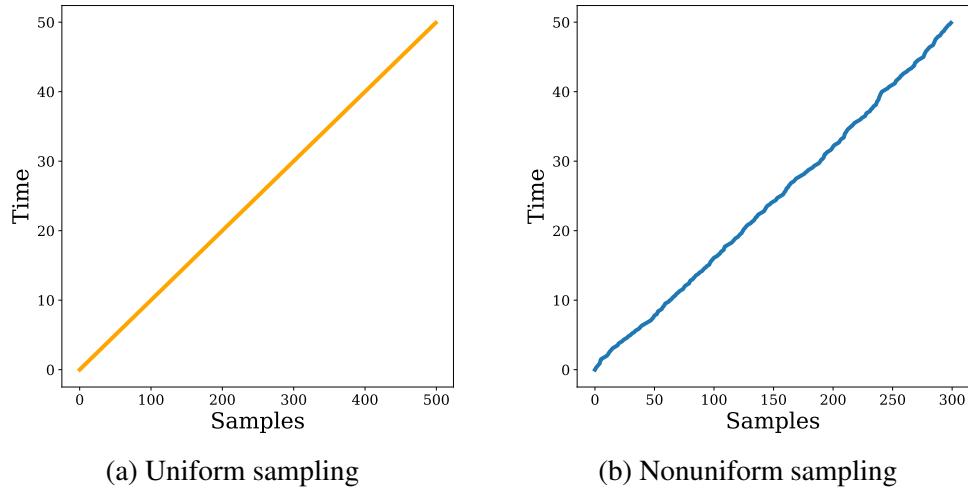


Figure 5.1. Sinusoidal data sampling: The non-uniform time set (right) has been randomly chosen from the uniform time set (left).

Using these time sets we computed the three following sinusoidal functions, which are shown in Figure 5.2:

$$x(t) = \sin(\pi t) \quad (5.1)$$

$$x(t) = \sin(\pi t) + \sin\left(\frac{\pi}{2}t\right) \quad (5.2)$$

$$x(t) = \sin(\pi t) + \sin\left(\frac{\pi}{2}t\right) + \sin\left(\frac{\pi}{3}t\right) \quad (5.3)$$

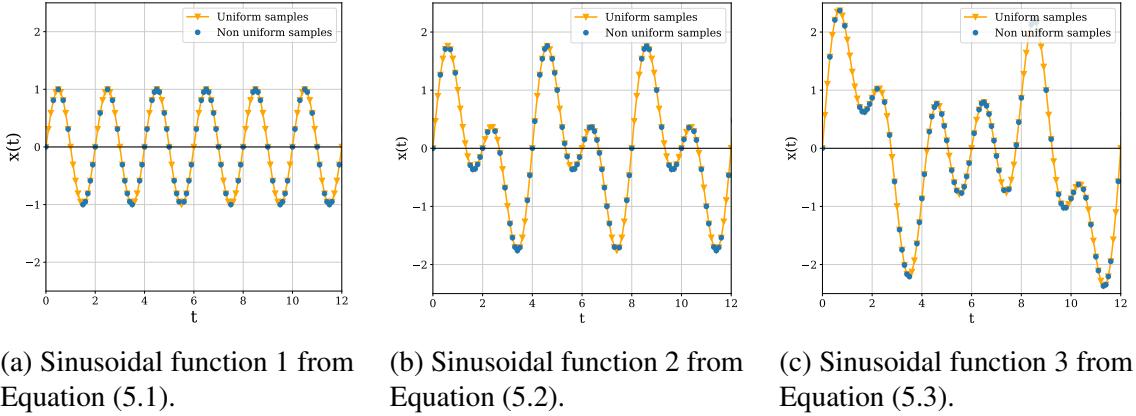


Figure 5.2. Sinusoidal functions used to test the non-uniform algorithm performance. Twelve time units (function 3 period) are shown for easier viewing. Orange triangles and blue circles correspond to uniform and non-uniform sampling, respectively.

Thesis Draft

To also test the effects of working with noisy data, we added to Equation (5.3) some Gaussian white noise of 0 mean and 0.5 standard deviations (eq.(5.4)) and then increased it to one standard deviation (eq.(5.5)). These two noisy functions are shown in Figure 5.3:

$$x(t) = \sin(\pi t) + \sin\left(\frac{\pi}{2}t\right) + \sin\left(\frac{\pi}{3}t\right) + \mathcal{N}(0, 0.5) \quad (5.4)$$

$$x(t) = \sin(\pi t) + \sin\left(\frac{\pi}{2}t\right) + \sin\left(\frac{\pi}{3}t\right) + \mathcal{N}(0, 1) \quad (5.5)$$

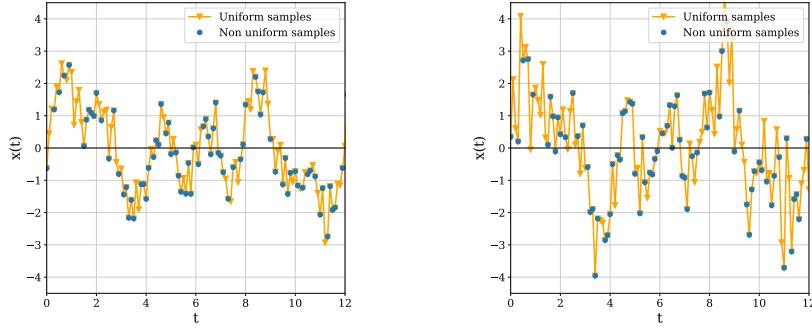
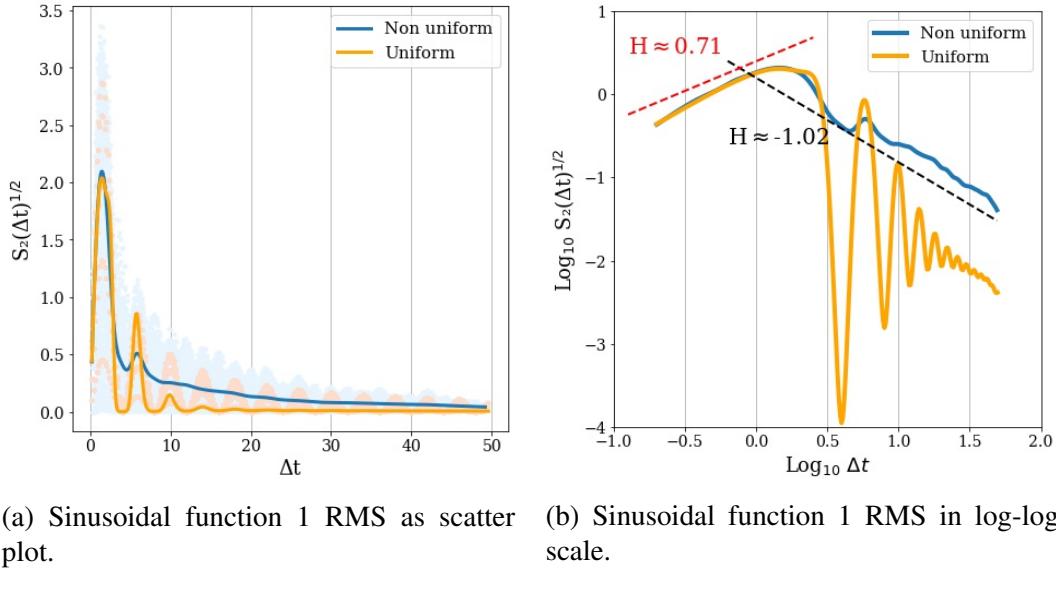


Figure 5.3. Noisy sinusoidal functions used to test the non-uniform algorithm performance. Orange triangles and blue circles correspond to uniform and non-uniform sampling, respectively.

Once defined the sinusoidal data, we started testing the algorithm. Figure 5.4 show the fluctuations RMS from the sinusoidal function 1. The left side figure show fluctuations as a scatter plot with its ensemble averaged values remarked as continuous lines, while the right side figure show the ensemble averaged fluctuations in a log-log plot, where dashed lines have been added as slope references (H). Once again orange and blue colours refers to data with uniform and non-uniform sampling, respectively.



(a) Sinusoidal function 1 RMS as scatter plot. (b) Sinusoidal function 1 RMS in log-log scale.

Figure 5.4. Fluctuations RMS from sinusoidal function 1 as scatter and log-log plots. Orange and blue colours correspond to uniform and non-uniform sampling, respectively. Dashed lines have been added as slope reference (H). For more detailed figures, see Appendix A.

So first, we notice in Figure 5.4a a main peak of magnitude $S_2(\Delta t)^{1/2} = 2$ located at $\Delta t = 2$, which matches with the function amplitude and periodicity. This is followed by harmonics peaks every two wave lengths, which magnitudes decreases as Δt becomes bigger and a larger number of wave length are covered. Of course harmonics are less distinguishable when the non-uniform sampling is used, as the Haar function assess more time lags combinations in these cases. At the same time, we can see in Figure 5.4b a positive (red) slope of $H \approx 0.7$ which fits the curve until $\log_{10} \Delta t \approx 0.3$, followed by a

negative (black) slope of $H \approx -1$, showing that fluctuations increase with time lags until one wave length, to then decrease with scale sharply.

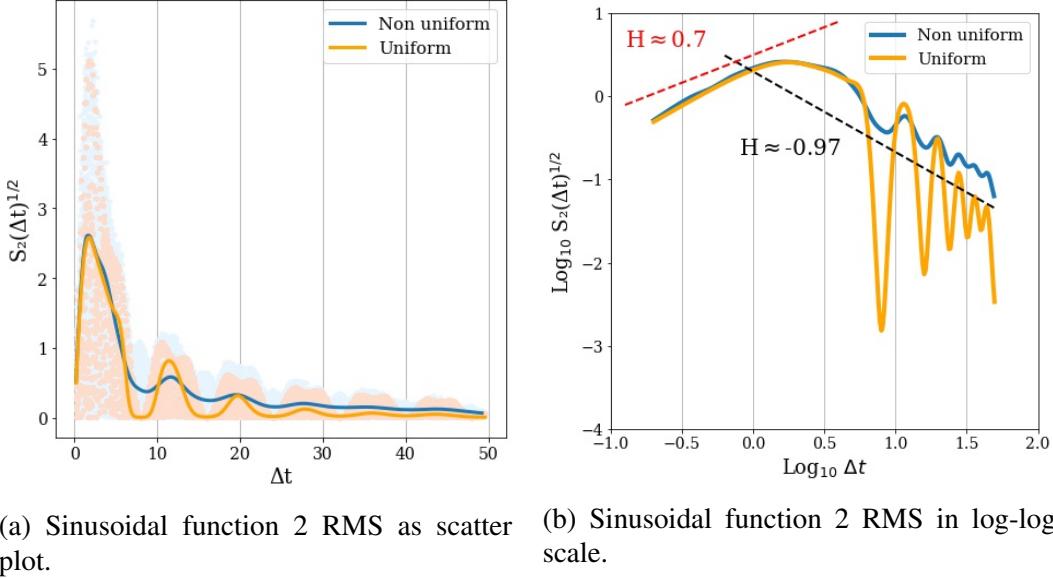
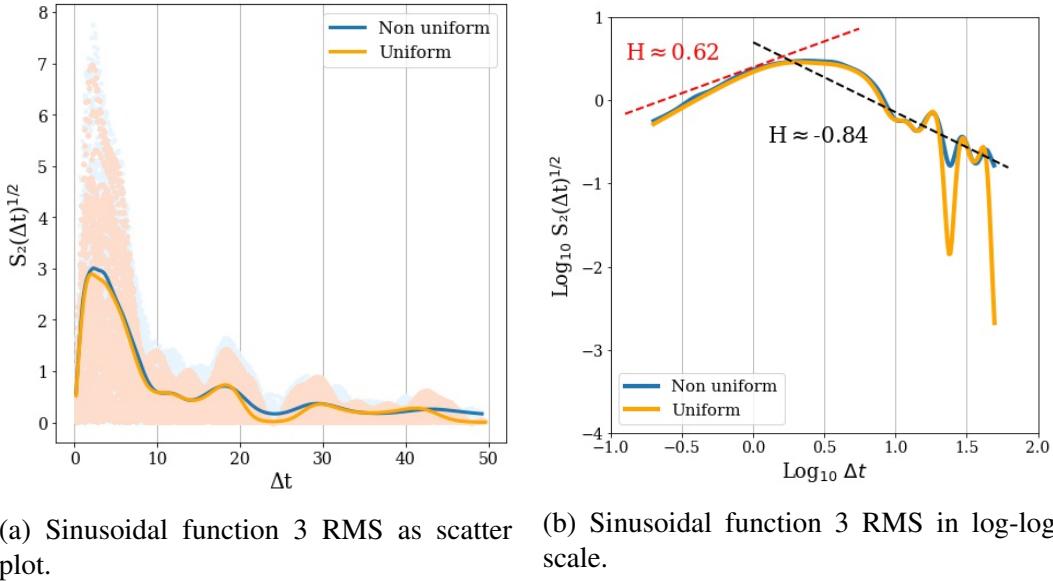


Figure 5.5. Fluctuations RMS from sinusoidal function 2 as scatter and log-log plots. Orange and blue colours correspond to uniform and non-uniform sampling, respectively. Dashed lines have been added as slope reference (H). For more detailed figures, see Appendix A.

Similarly, by looking at Figure 5.5a we notice how the scatter plot have become more spread and the averaged fluctuations from uniform and non-uniform sampling started to getting closer to each other. Harmonics were still visible in the orange cloud, but its peaks were wider then before. On the other hand, Figure 5.5b shows that both positive and negative slopes remained around the same values ($H \approx 0.7$ and -1), but since the main peaks were wider now, the transition between scale symmetries was more diffuse.

Next, we computed fluctuations for the sinusoidal function 3. Figure 5.6 shows again its RMS in linear and log-log scale, where we can notice how fluctuations from both uniform and non-uniform sampling were almost equal now. The main peak continued increasing and becoming more spread, which make sense since the function contains one extra lower frequency compared with the two cases from above, therefore its slopes were



(a) Sinusoidal function 3 RMS as scatter plot.
(b) Sinusoidal function 3 RMS in log-log scale.

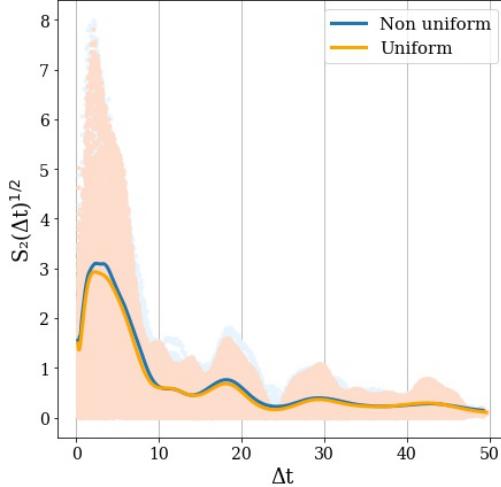
Figure 5.6. Fluctuations RMS from sinusoidal function 3 as scatter and log-log plots. Orange and blue colours correspond to uniform and non-uniform sampling, respectively. Dashed lines have been added as slope reference (H). For more detailed figures, see Appendix A.

less steep ($H \approx 0.6$ and -0.8). Nevertheless, it is still visible how fluctuations decrease with scale after one wave length, which was expected.

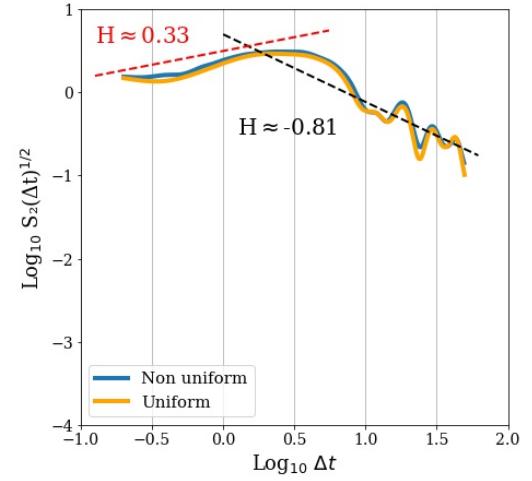
Once it was done, we repeat the analysis for the noisy datasets. Since white noise is a random signal with equal intensity at different frequencies (giving it a constant power spectral density), we expected to notice its effects only at high-frequency ranges. Figure 5.7 show the results from sinusoidal functions 4 and 5, where we see even more scatter clouds now. This cloud dispersion is specially noticeable around $\Delta t \approx [0, 8]$ where fluctuations reached values close to 8. Nevertheless, the peaks position of the ensemble averages were around the same time lags as in the function 3 case.

On the contrary, by looking fluctuations in the log-log scale, we notice a major decrease in the positives (red) slopes which have dropped to $H \approx [0.1, 0.3]$. This flattening in the high-frequency range means that fluctuations at $\Delta t \leq 0.3$ (i.e. smaller than a wave length) have increased its variability by the addition of white noise, which was expected.

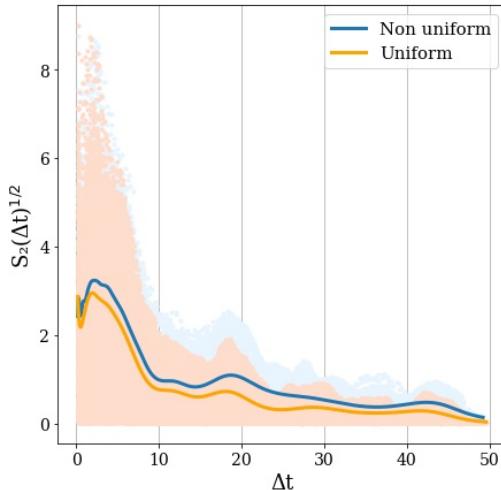
On the other hand, lower frequencies do not seem to be significantly affected and its variability continued decreasing with scale at the same rate as before (i.e., $H \approx -0.8$).



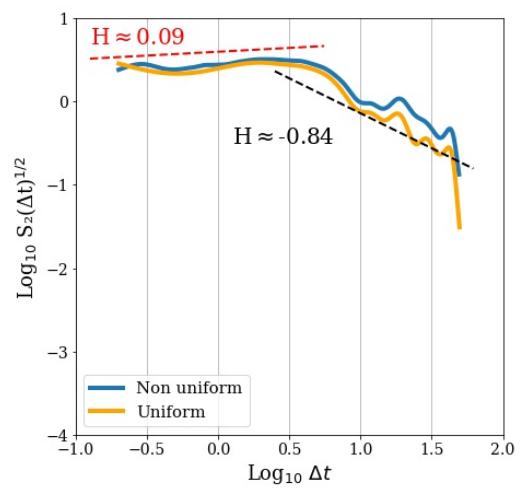
(a) Sinusoidal function 4 RMS as scatter plot.



(b) Sinusoidal function 4 RMS in log-log scale.



(c) Sinusoidal function 5 RMS as scatter plot.



(d) Sinusoidal function 5 RMS in log-log scale.

Figure 5.7. Fluctuations RMS from sinusoidal function 4 and 5 as scatter and log-log plots. Orange and blue colours correspond to uniform and non-uniform sampling, respectively. Dashed lines have been added as slope reference (H). For more detailed figures, see Appendix A.

So, in summary, this first test showed us that:

- Fluctuations decreased with scale after one wavelength in all 5 cases.
- The ensemble average peaks (and valleys) were located at consistent time lags.
- The ensemble average magnitude looked coherent with the sinusoidal functions amplitudes.
- The white noise added to the signal made it more difficult to detect the high-frequency fluctuations.
- Fluctuation differences from uniform and non-uniform data diminished as the sinusoidal functions got more complex in composition.

For these reasons, our algorithm looks robust enough to work with non-uniform datasets since its outcomes are correct when compared to those from uniform data. Nevertheless, we decided to run one more synthetic test before working with real climatic data, this time emulating a “standard” paleotemperature data set.

5.2. Composite paleotemperature dataset

After running our Haar function algorithm over sinusoidal datasets, we extended the validation test to a more realistic scenario by defining a “Paleotemperature signal” composed by the following six waves:

- High frequency components (HF):
 - Wave 1: 5 year-period wave of amplitude 0.1°C
 - Wave 2: 10 year-period wave of amplitude 0.2°C
 - Wave 3: 100 year-period wave of amplitude 0.1°C
- Low frequency components (LF):
 - Wave 4: 20 Kyear-period wave of amplitude 3°C
 - Wave 5: 40 Kyear-period wave of amplitude 2°C
 - Wave 6: 100 Kyear-period wave of amplitude 1.5°C

These amplitudes were selected in order to obtain a high frequency variation of around $\pm 0.4^\circ\text{C}$ and a low frequency variation around 10 times bigger. Figures 5.8 and 5.9 show the high and low frequency waves used in the composition, respectively. The resulting waves (the sum of the HF and LF components) are visualized in black.

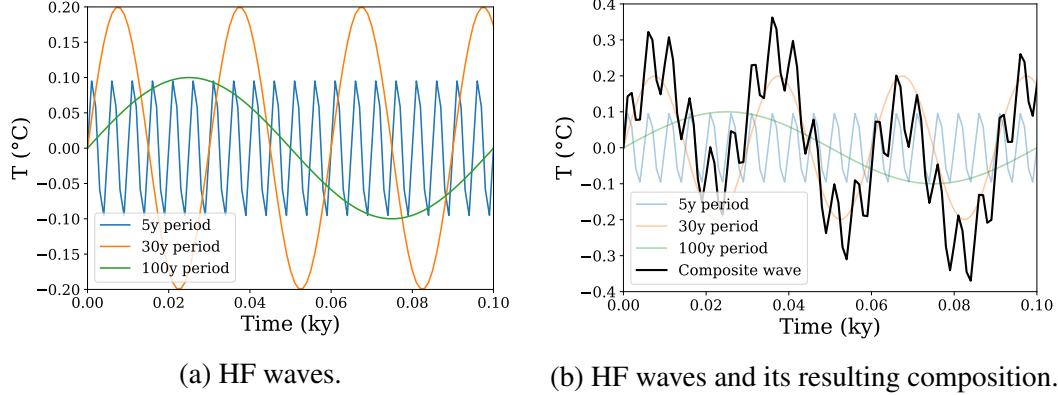


Figure 5.8. HF components of the synthetic paleotemperature signal. Left side figure show the three HF waves used, while right side figure shows the resulting high frequency wave in black.

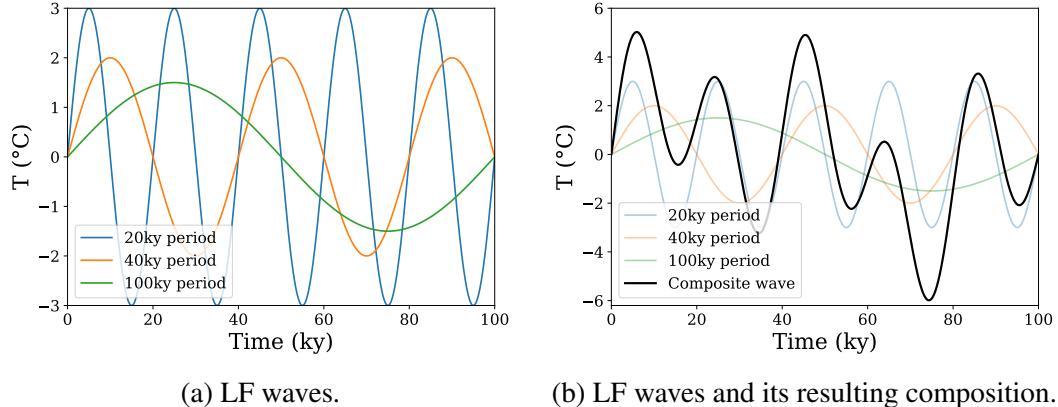


Figure 5.9. LF components of the synthetic paleotemperature signal. Left side figure show the three LF waves used, while right side figure shows the resulting high frequency wave in black.

Using these, we defined the temperature signal as a 200 Kiloyears wave since it was an acceptable paleoclimatic extension. Also, this length would allow us to sample the signal

at high resolution and compute its fluctuations with a reasonable computational cost. The resulting 200 kyear wave is shown in Figure 5.10.

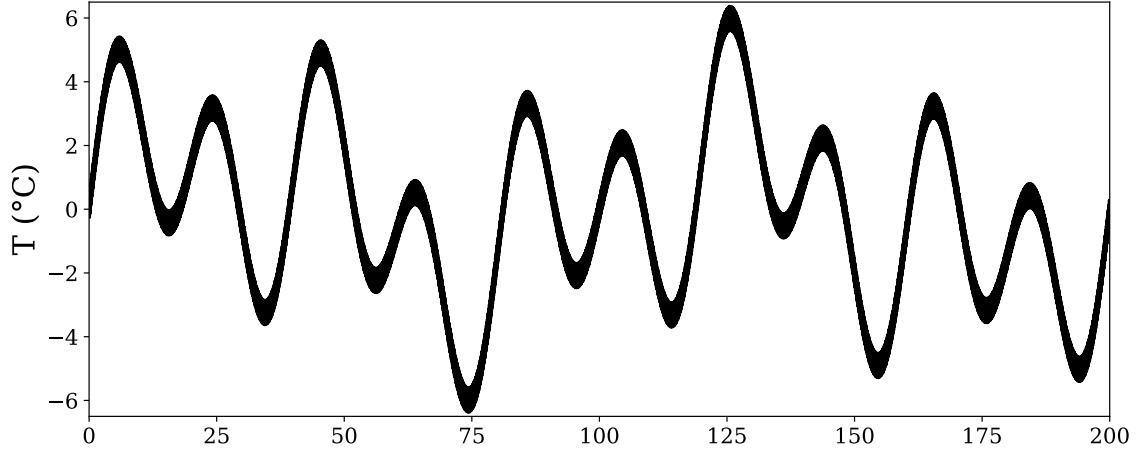


Figure 5.10. 200 Kyear synthetic paleotemperature wave.

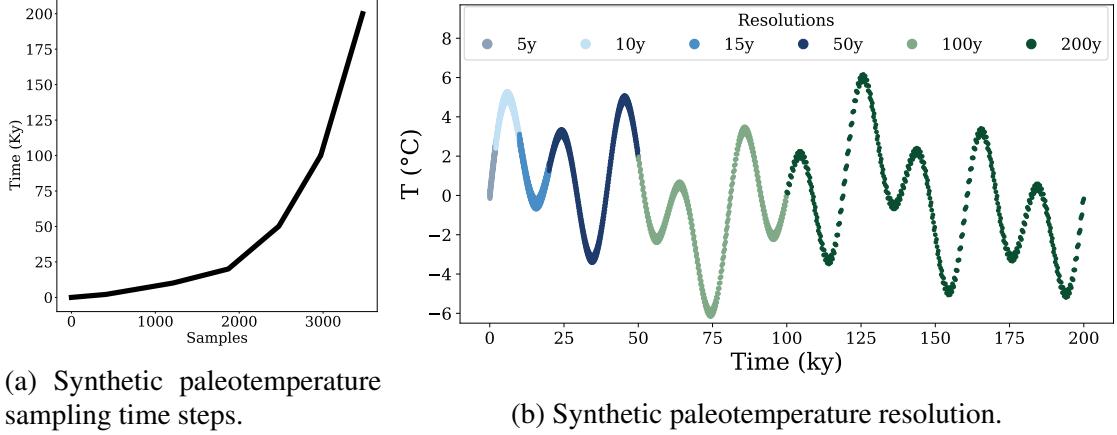
To make it more realistic, we sampled the working wave at different rates, going from a 200 years resolution on the “bottom” up to a 5 year resolution on the “top” as is shown in Table 5.1. This non-uniform sampling is also shown in Figure 5.11.

Table 5.1. Non-uniform synthetic paleotemperature sampling

Time span (Ky)	Resolution (y)	N°of samples
2	5	400
8	10	800
10	15	666
30	50	600
50	100	500
100	200	500

Once it was done, we ran the following five tests to assess the algorithm sensitivity:

- (i) **Data length test:** We computed and compared fluctuations RMS for segments of 200ky (complete paleotemperature signal), 150 ky, 100 ky, 50 ky, and 10 ky. Each segment was considered from the “top” portion of the signal.



(a) Synthetic paleotemperature sampling time steps.

(b) Synthetic paleotemperature resolution.

Figure 5.11. Non-uniform sampling of the synthetic paleotemperature signal. Left side figure show the time steps used in the sampling, while right side figure shows the final wave coloured by resolution.

- (ii) **Noise test:** Since it has been found that temperatures time series from the EPICA ice core record present near-red noise on subglacial timescales, and near-white noise on glacial timescales (Keyes, Giorgini, & Wettlaufer, 2022), we added near-red and near-white noise to our synthetic paleotemperature wave and compared the RMS fluctuations between the initial and noisy signal. Two scenarios were tested: first using a 0 mean and 0.05 standard deviation noise (Noise 1), and then a 0 mean and 0.1 standard deviation noise (Noise 2).
- (iii) **Components amplitude test:** We modified the paleotemperature signal by taking the component *Wave 1* and amplifying its magnitude by a factor of 2, 5, and 10. For each case, we computed and compared its fluctuations RMS with the original ones. Once it was done, we returned *Wave 1* to its initial value and repeated the exercise for all the other five components waves.
- (iv) **Extra LF component test:** We modify three times the paleotemperature signal by adding to it a 4th LF component of 3Ky, 5Ky, and 10ky periodicity, respectively. At each time, we computed and compared its fluctuations RMS with the original ones. The new LF component amplitude was always fixed at 1°C, thus being comparable with the rest of the low-frequency components.

(v) **Linear interpolation test:** We took the first 10 Kyears of the paleotemperature signal and resampled it at a 5-year uniform resolution. From there, we randomly choose 800 samples (nonuniform dataset) and brought them back to a uniform 5-y resolution by linear interpolation. Hence we finished with one uniformly sampled set, one nonuniform sampled set, and one uniform (after interpolation) set to compute and compare its fluctuations RMS.

5.2.1. Datalength test

Figure 5.12 compares the fluctuations RMS obtained from the *Data length test*. At first glance, all five segments look to present pretty much the same fluctuations, with a global maximum at lower frequencies and a smaller peak at higher frequencies. This difference was expected since, by construction, the LF components had higher amplitudes than the HF ones.

However, we notice that shortening the segments causes minor differences at the end of each RMS curve. This is because, for a given Δt , shorter segments averages less fluctuations and hence differences arises (e.g. yellow right tail ends with a peak at $\log_{10} \Delta t = 1.6$, while the rest decreases at the same time step). Likewise, and for the same reason, the 10ky segment behaves differently at higher frequencies but catches up the other segments as the time steps increases.

By looking at the peaks, it is possible to identify the 100y, 20Ky, 40Ky, and 100Ky-period components appearing at $\log_{10} \Delta t = [-1, 1.3, 1.6, 2]$, respectively, whereas the 5y and 10y-period components are not distinguishable because of the resolution. In addition, we noticed a clear peak at $\log_{10} \Delta t = -1.5$ (or 30y) which appears as the combination of the 10y and 5y-period components harmonics.

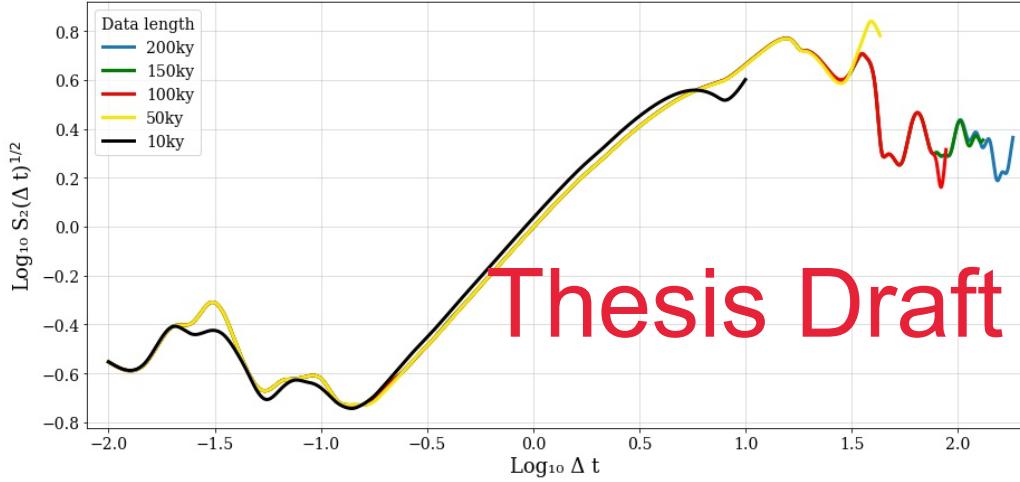


Figure 5.12. Five segments fluctuations RMS comparison.

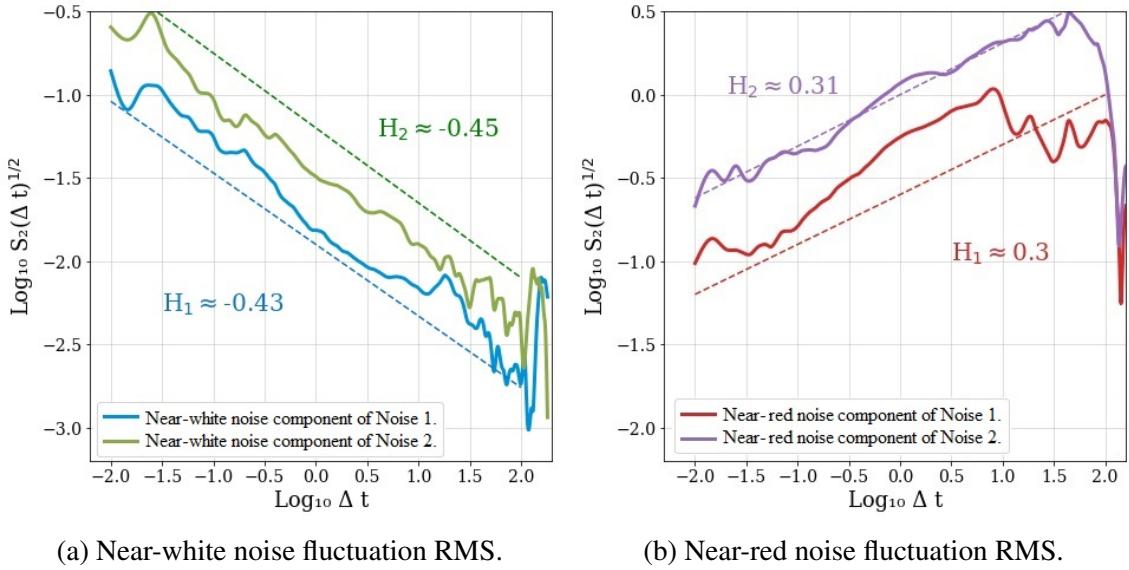
5.2.2. Noise test

We show in Figure 5.13 the near-white and near-red noise components used in the *Noise test*. Theoretically, in a log-log plot, white noise fluctuations RMS should be fitted by a straight line of -0.5 slope, whereas perfect red noise should be close to +0.5, however, because of the bounded number of samples and the non-equidistant sampling, we did not obtain straight lines. Nevertheless, the reference dashed lines added to Figures 5.13a and 5.13b show slopes of around -0.44 and +0.3, respectively, which is sufficiently close to perfect white and red noise for our validation.

Hence, we added the components³ to the paleotemperature signal and computed its fluctuations RMS, which are show in Figure 5.14. Here we can see that fluctuations still presented a global maximum at lower frequencies, which is almost the same in the three cases, while the smaller peak at high frequencies was notoriously shifted when Noise 2 was added.

We anticipated these smaller differences between the original and noisy signal 1 since it had a smaller standard deviation than the noisy signal 2. For this same reason, and since the latter had a standard deviation comparable with the magnitude of the high-frequency

³Noise 1 = near-white + near-red components 1; Noise 2 = near-white + near-red components 2.



(a) Near-white noise fluctuation RMS.

(b) Near-red noise fluctuation RMS.

Figure 5.13. Near-white and near-red noise components used in the *Noise test*. Dashed lines have been added as slopes references. Theoretically, in a log-log plot, white and red noise fluctuation RMS should be fitted by a straight line of -0.5 and +0.5, respectively.

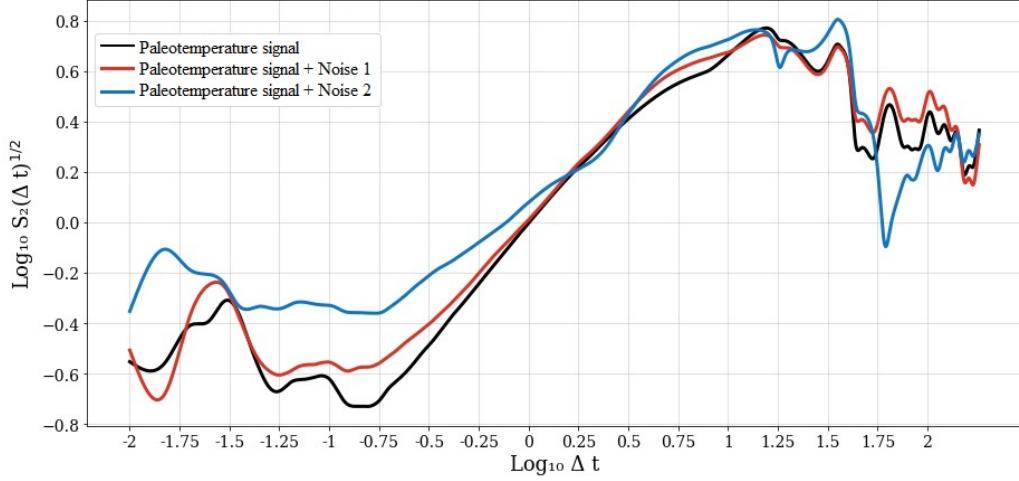


Figure 5.14. Original and noisy paleotemperature signal fluctuations RMS.

components, fluctuations from the noisy signal 2 were notoriously higher up to $\log_{10} \Delta t = 0.23$ or 1700y. Nevertheless, high frequencies fluctuations from the noisy signal 1 have also increased with the addition of white noise to the original signal.

On the other hand, fluctuations after $\log_{10} \Delta t = 0.23$ seems to be pretty much similar between the three signals, with the exception of a big drop in the noisy signal 2 at $\log_{10} \Delta t = 1.76$, which is presumably because of the randomness of the components added.

5.2.3. Components amplitude test

Figure 5.15 shows the results from the *Components amplitude test*. First, we notice that since the higher resolution used in the paleotemperature signal sampling was five years (see Figure 5.11b), the amplifications of the 5y-period component did not cause any visible effect in the fluctuations RMS curve, as is shown in Figure 5.15a.

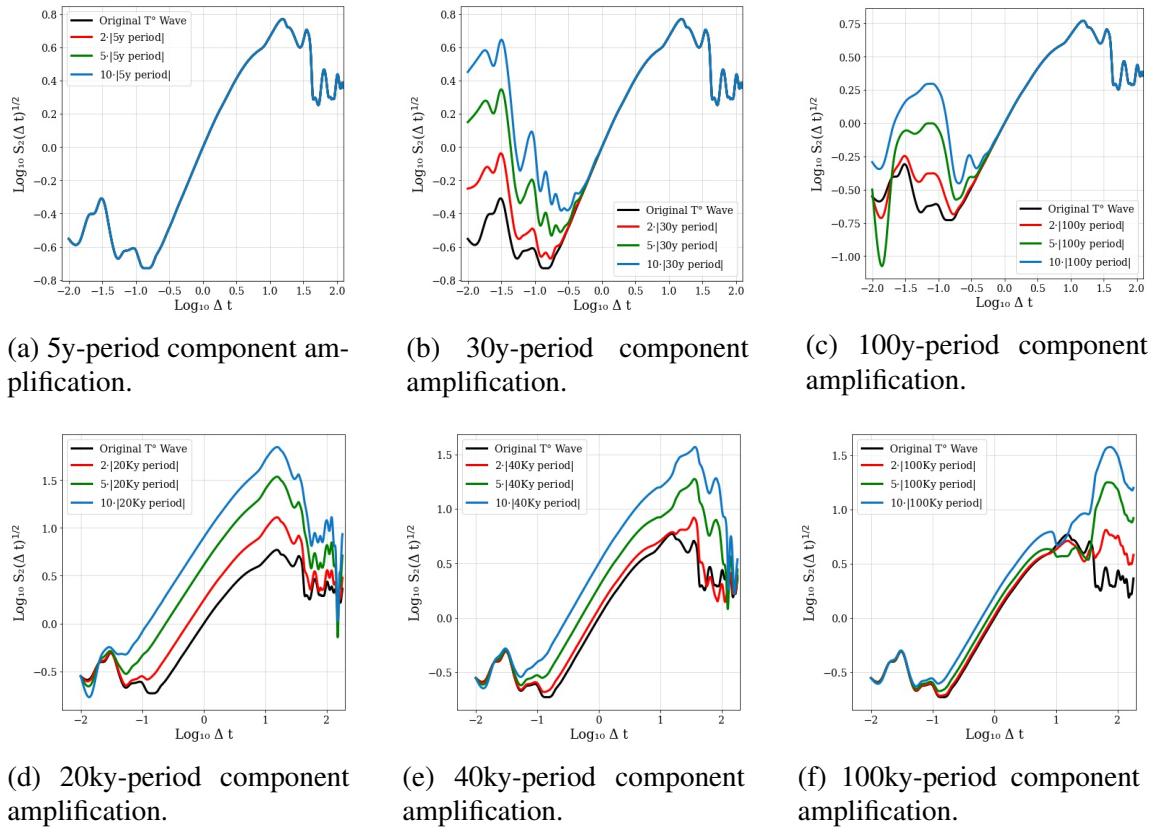


Figure 5.15. High and low frequency results from the *Component amplitude test*. Every component was individually multiplied by a factor of 2, 5, and 10.

Regarding the other HF amplifications, we notice it only had an effect over fluctuations in the range $\log_{10} \Delta t = [-2, -0.25]$ or $[10y, 560y]$, becoming then indistinguishable. We can see in Figure 5.15b how the respecting 30y peak ($\log_{10} \Delta t = -1.5$) and its neighbourhood have increased in all three runs, reaching a magnitude close to the low-frequency maximum. As the component got stronger, hidden harmonics became visible and the global minimum got shifted to the right. In parallel, we find the same scenario in Figure 5.15c, with the slight difference that fluctuations have dropped in frequencies under 20y (or $\log_{10} \Delta t = -1.7$) when amplified by 2 and 5. This time, only one harmonic arises with amplification.

On the other hand, we can see in Figures 5.15d, 5.15e, and 5.15f that LF amplifications had an impact on a much wider frequency range, reaching to $\log_{10} \Delta t = -1.5$ in the 20ky-period case to $\log_{10} \Delta t = -1$ in the 100Ky-period case. Nevertheless, the magnitude of these impact on high and mid frequencies is not the same in the three cases, being easily visible in the 20ky and 40ky-period cases but much more subtle in the 100ky-period one. Peaks have increased consistently in all cases once again, and the global minimum tends to be shifted to the left.

It is clear that every components amplification will impact the resulting paleotemperature signal modifying either the position or magnitude of its fluctuations RMS peaks. This results in a change of the slope of the curve (i.e., a change of the H exponent) which we can interpret as:

- HF amplification: The H exponent in the range $\log_{10} \Delta t = [-1.5, -0.5]$ decreases with the amplification, meaning that even if now the HF variability is higher, the process rapidly “forgets” its effects.
- LF amplification: The H exponent in the range $\log_{10} \Delta t = [-1, 1.5]$ increases with the amplification, meaning that the process has more memory of its effects.

5.2.4. Extra LF component test

Now, we show the results from the *Extra LF component test*. So far we have repeatedly seen that paleotemperature fluctuations RMS presents a HF maximum at 30 years, which is a combination of the harmonics from the 5y and 10y-period components, but the LF maximum located at $\log_{10} \Delta t = 1.19$ or 15.5ky was not straightforward to infer. Thus, the idea of the test was to see how the curve changes when an extra LF component was added.

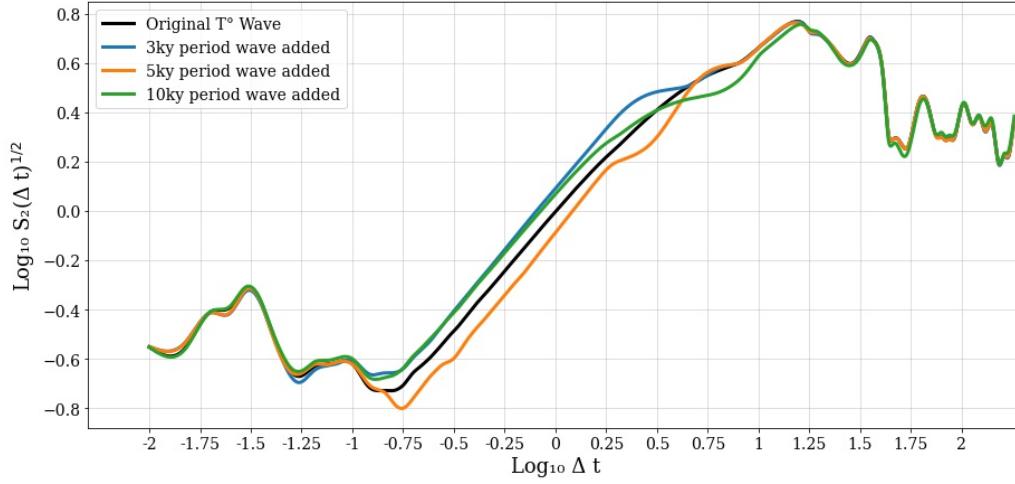


Figure 5.16. Paleotemperature fluctuations RMS after the addition of a 4th LF component.

Figure 5.16 compares fluctuations RMS after the addition of a 4th LF component. The blue curve shows an increase in fluctuations at $\log_{10} \Delta t = 0.48$, corresponding to one wavelength of the 3ky-period component, while orange and green curves do not show any increase in the time steps corresponding to their respective periods. Although fluctuations in the mid frequencies have changed (increasing when a 3ky and a 10ky-period components were added, and decreasing for the 5ky-period case), the high and low-frequency maxima have not shifted, suggesting that the peak found at $\log_{10} \Delta t = 1.19$ actually corresponds to the most unstable time lag of the synthetic paleotemperature dataset.

5.2.5. Linear interpolation test

Lastly, we show the results from the *Linear interpolation test*, which objective was to check that the algorithm was in fact computing improved outcomes compared to the interpolation approach. Thus, Figure 5.17 shows the resulting fluctuations RMS from the uniform, non-uniform (randomly sampled), and uniform-after interpolation sets.

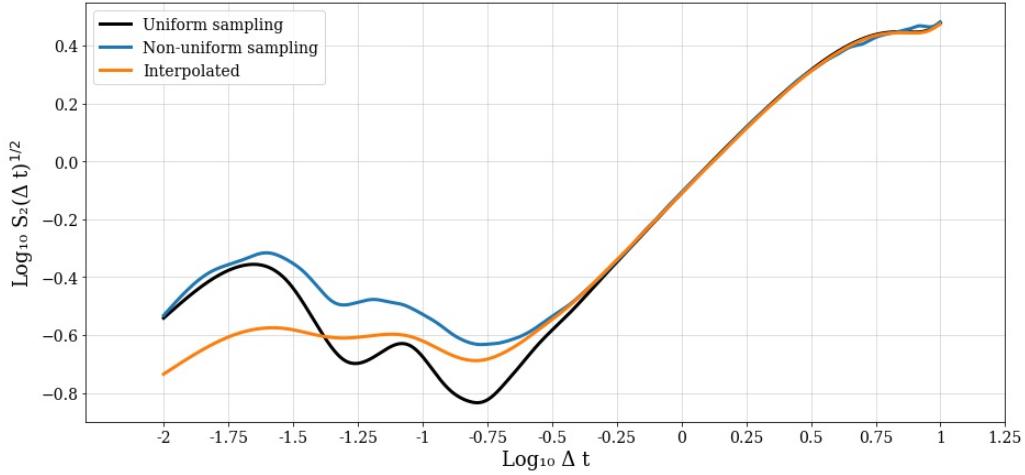


Figure 5.17. Fluctuations RMS from the first 10ky of the paleotemperature signal under uniform, non-uniform, and interpolated sampling.

First, we notice that peaks located at $\log_{10} \Delta t = -1.65$ and -1.07 under uniform sampling are no longer visible in the orange curve, while at the same time the global minimum at $\log_{10} \Delta t = -0.79$ has been overestimated. This came as linear interpolation is a low-pass filter and thus it implies high-frequency information loss, resulting either in an over or under estimation of fluctuations.

On the other hand, fluctuations from the non-uniform set also differs in high-frequency time lags compared to those from the uniform set, showing an overestimation in the range $\log_{10} \Delta t = [-2, -0.4]$. We attributed this overestimation to the fact that the non-uniform set contains only 40% of the information (800 out of 2000 samples), and thus, fewer fluctuations have been averaged. Even so, both peaks are still clearly distinguishable, showing that the algorithm does deliver more accurate results than its interpolated counterpart.

This way, we finished validating our Haar function algorithm for non-uniform datasets, where we saw that it responds well to all previous study scenarios. However, even if the algorithm is reliable, the precise values of fluctuations may, to some degree, vary from case to case because of the implicit approximations done, for example, fluctuation smoothing by cubic spline. Hence we should not take them as an analytical reference but rather as a highly probable estimation. In other words, to properly assess the stability of a process, it is better to focus on both the position (in time) of the fluctuations RMS peaks and the time lags where the scale symmetry holds and breaks (i.e., changes in the H exponent).

6. PALEOCLIMATE

6.1. Paleoclimate dataset description

One of the aims of this work was to create a suitable global paleoclimatic database to study its atmospheric variability on a decennial-millennial timescale. To do so, we take as a reference the work done by (Lovejoy & Lambert, 2019), where they found that the Antarctic timescale transition from macro-weather to climate (τ_c) was around 250 years. Hence, we decided to work only with datasets with at least centennial resolution, as otherwise, we could be outside the scale where τ_c presumably occurs. On the other hand, in (Lovejoy & Schertzer, 2013) and (Lovejoy, 2015) is postulated the existence of five different climatic regimes, with a macroweather regime covering from around 20 days to dozens of years, and a climate regime going from dozens to thousands of years. Therefore, we selected datasets of at least 60k years (roughly a half-glacial cycle) since it would allow us to have enough fluctuations to assess both the macroweather and climate regime.

Thus, after looking in open data publishers³ we finished with a global database containing 11 sets of dust flux or concentration and 13 sets of temperature reconstructions, from which 16 came from ice cores, five from marine sediments, one from lake sediment, one from speleothem, and one from loess. In general, dust values were obtained from Ca_2^+ and nonsea-salt Ca_2^+ concentration, while temperature values came from Oxygen isotopic ratios δO^{18} .

Both dust and temperature measurements are affected by geophysical phenomena such as volcano eruptions, which cause sudden peaks in dust fluxes and rises the atmospheric temperatures. Since those are random and isolated events, it is impossible to distinguish them from climatic variations just by eye, so we decided to keep all the peaks of the datasets without filtering any outliers. Likewise, although the datasets present trends over the years, we decided not to remove them since any high pass filter would smooth fluctuations, and hence high-frequency information would be lost.

³In particular: <https://www.ncei.noaa.gov/> and <https://www.pangaea.de/>

Usually, paleoclimatic dataset sampling is done by chemically analyzing small layers of samples from proxy cores, with the drawback that the same sample size from the bottom or the top does not represent the same time step, thus being sensitive to different natural phenomena such as ice core compaction. For this reason, we expected datasets with monotonic decreasing resolutions. Nevertheless, because of this resolution loss, some datasets have been previously reconstructed by linear interpolation and were available in a structured form with constant resolution.

6.2. Datasets summary

As mentioned, we finished with 24 paleoclimatic datasets that fulfilled our analysis constraints, which locations are shown in Figure 6.1. We notice that almost half of the sets came from the poles (either from Greenland or Antarctica), which is one limitation in our objective of comparing τ_c between different latitudes. Also, since we can not presume that different regions would have the same atmospheric behavior, fewer proxies by latitude imply a more local analysis.

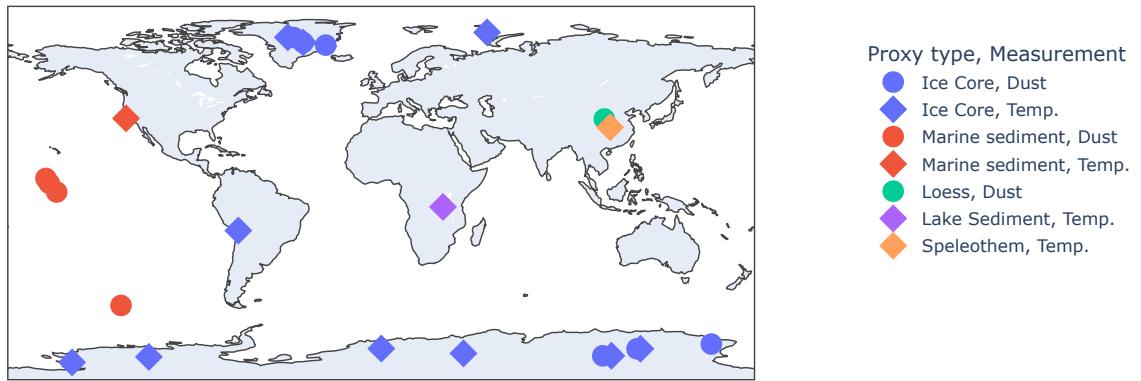


Figure 6.1. Global proxies distribution

Table 6.1 summarizes the coordinates, time span and data length of the datasets. It is worth mentioning that even though Sajama ice core covers only 25ky B.P (and not 60ky as we defined it before), we decided to keep it as it could still contribute to a better

understanding of the analysis overview, especially under the consideration of its isolated location.

Table 6.1. Datasets summary: Dust/Dust fluxes (upper part), Temperatures (bottom part).

Name	Latitude (°)	Longitude (°)	Time span (Ky)	Samples
Vostok	-78.46	106.87	422	522
EPICA Dome C	-75.10	123.35	801	31838
Talos Dome	-72.82	159.18	150	64998
Central South Pacific (PS75/59-2)	-54.22	-125.43	474	2384
Central Pacific Ocean (ML1208-17PC)	0.48	-156.45	149	192
Central Pacific Ocean (ML1208-31BB)	4.68	-160.05	141	254
Central Pacific Ocean (ML1208-37BB)	7.04	-161.63	144	105
Xifeng - China	35.70	107.60	800	722
RECAP	71.30	-26.72	121	2317
GRIP	72.59	-37.64	248	3038
NGRIP	75.00	-42.30	108	2774
Siple Dome	-81.65	-149.00	94.5	1879
WAIS	-79	-112	67.7	6375
Vostok	-78.46	106.87	422	3311
Dome Fuji	-77.3	39.7	340	1189
Epica Dome C	-75.1	123.35	801	5787
EPICA EDML	-75.00	0.07	150	2303
Sajama - Bolivia	-18.1	-68.88	25	251
Lake Tanganyika - Eastern Africa	-6.65	29.80	59.4	209
Sanbao Cave - China	31.7	110.4	641	6111
Central California SST	37	-123	161	284
GRIP	72.59	-37.64	104	4993
NGRIP	75.1	-42.3	123	5651
NEEM	77.45	51.06	111	19173

On the other hand, Table 6.2 specifies the number of proxies by latitude and time span. We can see that the majority of the sets cover more than the last glacial cycle (≈ 130 Ky).

Table 6.2. Proxy Counter: By latitude (upper part), by timespan (bottom part)

Latitude/Age	Ice Core	Marine Sediment	lake	loess	speleothem
[90,60]°	6	0	0	0	0
[60,20]°	0	1	0	1	1
[20,-20]°	1	3	1	0	0
[-20,-60]°	0	1	0	0	0
[-60,-90]°	9	0	0	0	0
[60 ≤ Age < 130] ky	7	0	1	0	0
[130 ≤ Age] ky	8	5	0	1	1

Regarding the datasets resolutions, we show in Figures 6.2 and 6.3 the normalized resolutions of the sets clustered by type of measure (dust or temperature), where we notice both uniform (straight lines) and non-uniform samplings.

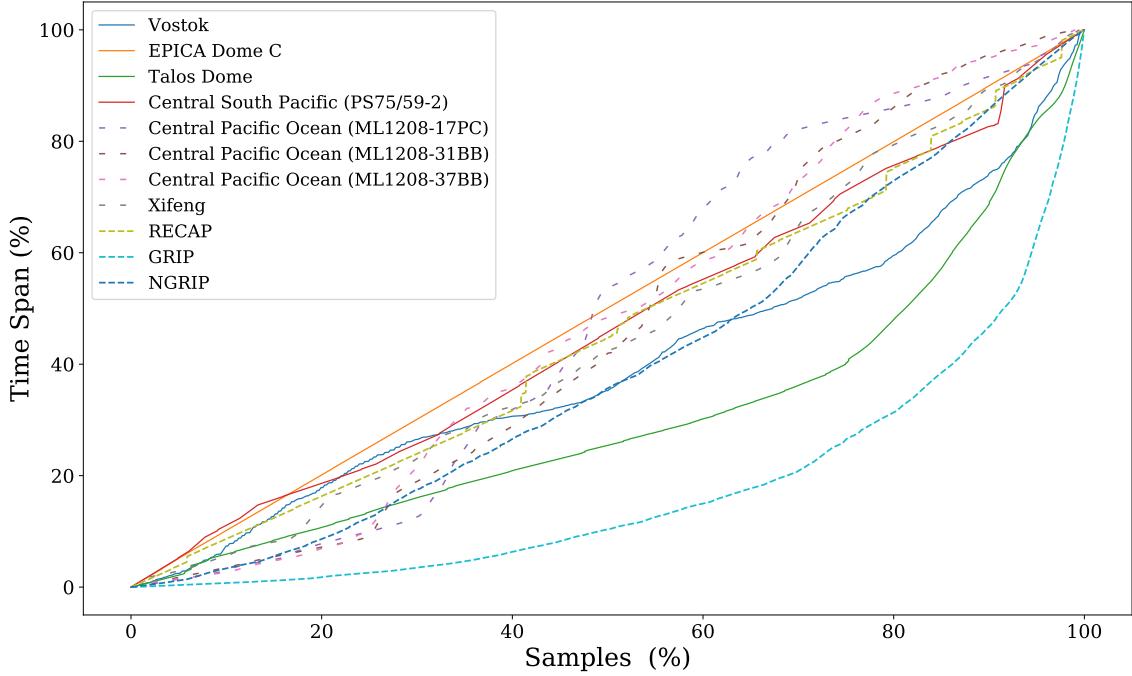


Figure 6.2. Normalized resolution: Dust/Dust flux datasets.

In general, proxies that cover larger timespans have the tradeoff of presenting lower resolutions, which is the case of loess or marine sediments. On the contrary, since ice

cores can be sampled continuously or every few centimeters while melting the ice, they present the advantage of covering larger timespans at high resolution.

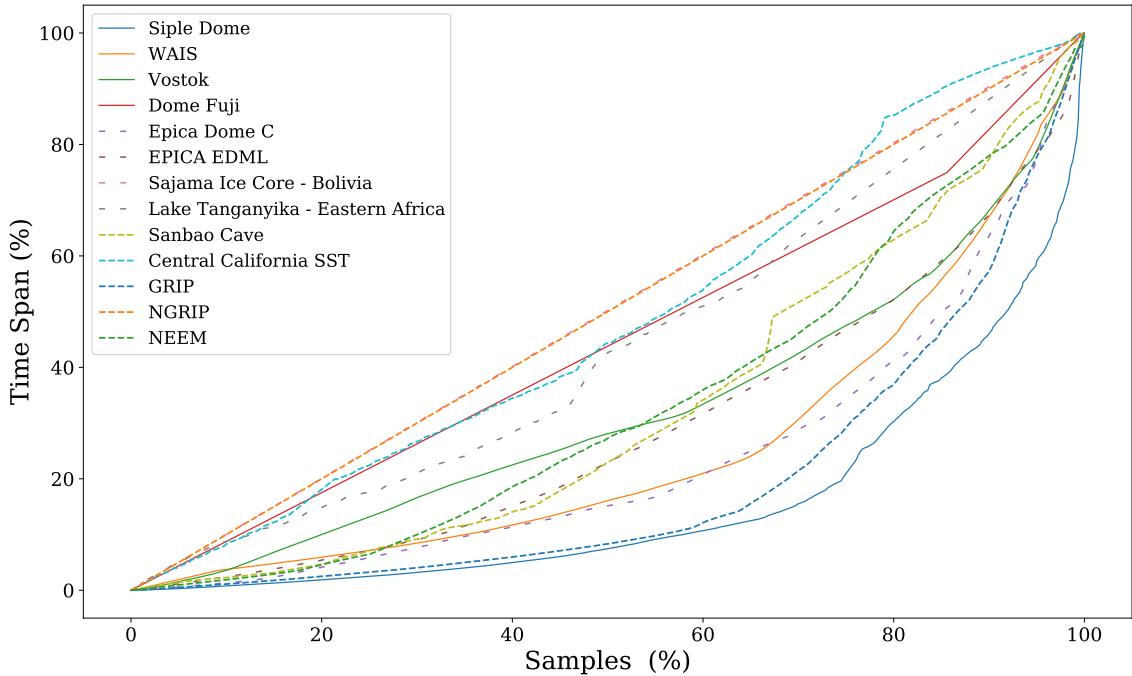


Figure 6.3. Normalized resolution: Temperatures datasets.

Finally, Table 6.3 presents the references to all the respective datasets scientific papers.

Table 6.3. Datasets DOI: Dust/Dust flux samples (upper part), Temperatures samples (bottom part)

Name	DOI
Vostok	https://doi.org/10.1594/PANGAEA.55502
EPICA Dome C	https://doi.org/10.1594/PANGAEA.778213
Talos Dome	https://doi.org/10.1594/PANGAEA.833044
Central South Pacific (PS75/59-2)	https://doi.org/10.1126/science.1245424
Central Pacific Ocean (ML1208-17PC)	https://doi.org/10.1038/ncomms10449
Central Pacific Ocean (ML1208-31BB)	https://doi.org/10.1038/ncomms10449
Central Pacific Ocean (ML1208-37BB)	https://doi.org/10.1038/ncomms10449
Xifeng - China	https://doi.org/10.5194/cp-5-21-2009
RECAP	https://doi.org/10.1038/s41467-019-12546-2
GRIP	https://doi.org/10.1594/PANGAEA.54557
NGRIP	https://doi.org/10.1594/PANGAEA.587836
Siple Dome	https://doi.org/10.1016/j.quascirev.2005.02.002
WAIS	https://doi.org/10.1038/nature14401
Vostok	https://doi.org/10.1594/PANGAEA.55505
Dome Fuji	https://doi.org/10.4236/acs.2013.32020
Epica Dome C	https://doi.org/10.1126/science.1141038
EPICA EDML	https://doi.org/10.1594/PANGAEA.601866
Sajama - Bolivia	https://doi.org/10.1126/science.282.5395.1858
Lake Tanganyika - Eastern Africa	https://doi.org/10.1126/science.1160485
Sanbao Cave - China	https://doi.org/10.1038/nature18591
Central California SST	https://doi.org/10.1029/2009PA001836
GRIP	https://doi.org/10.1016/j.quascirev.2014.10.032
NGRIP	https://doi.org/10.5194/cp-10-887-2014
NEEM	https://doi.org/10.1594/PANGAEA.925552

6.3. Dataset preprocessing

Before analyzing the datasets by Haar fluctuations, we needed to make some corrections. First, we noticed that GRIP temperatures, Sanbao Cave, NEEM, Talos Dome, and the three ML1208 sets sometimes presented a sub-annual resolution. In these cases, we kept one item with the averaged value of the repeated measurements in that year. Similarly, Lake Tanganyika had one year repeated (29.855y BP), so we kept only one of them.

On the other hand, WAIS dataset presented two items with a value of 999.999 which were clearly incorrect, so we decided to set the values to Nan. Finally, NEEM had a non-monotonic time steps with a decreasing time in the 111-120ky B.P range, which is why we cut the set at 110.8Ky. Please notice that length shown in Table 6.1 consider all these corrections.

Since in the beginning, we did not have the Haar function for non-equidistant datasets, our first approach was to convert all resolutions to constant. Hence, the first step was to define for each dataset a desired resolution to attain, which was an arbitrary and iterative decision as the sampling rates were different from case to case. Therefore, we took the median of the lower (in depth) 10% of each dataset as desired resolution and called it *bottom resolution*. Notice that we took the median and not its averaged value since the latter is more sensitive to extremes, so the highly spiky nature of the sets would lead to underestimating the resolution. We exemplify it in Figure 6.4, which shows the bottom resolution of Vostok dust dataset, where despite fluctuations, the median resolution of its lower 10% samples is 1697 years.

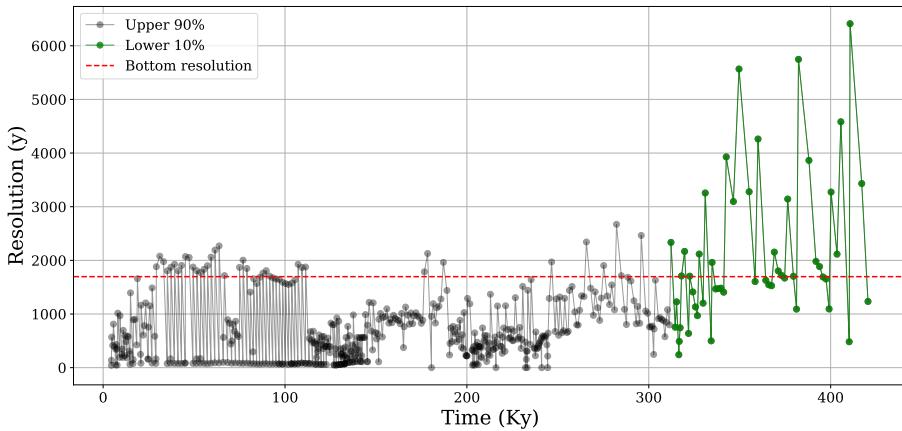


Figure 6.4. *Bottom resolution* estimation for Vostok dust dataset. Grey dots correspond to the upper 90% of the samples, while green dots marks its lower 10%. The dashed red line show the resulting bottom resolution.

Once we had the constant time-distances for the sets, we needed to map a value for each age t_i . To do so, we stored at every time t_i the average value of dust or temperature of

the samples within the interval $[t_i, t_i + \text{bottom res}]$, which sometimes led to the appearance of Nan values. Hence, once the mapping was complete, we closed the gaps using linear interpolation.

Following this procedure, we converted all datasets to equidistant sets, being then able to analyse them by using the uniform Haar function. Since these transformations inevitably implied information loss, we always compared the raw and equidistant datasets to verify that we were not highly underestimating fluctuations. To illustrate this, we show in Figure 6.5 the original, resampled and interpolated ML1208-17PC dataset.

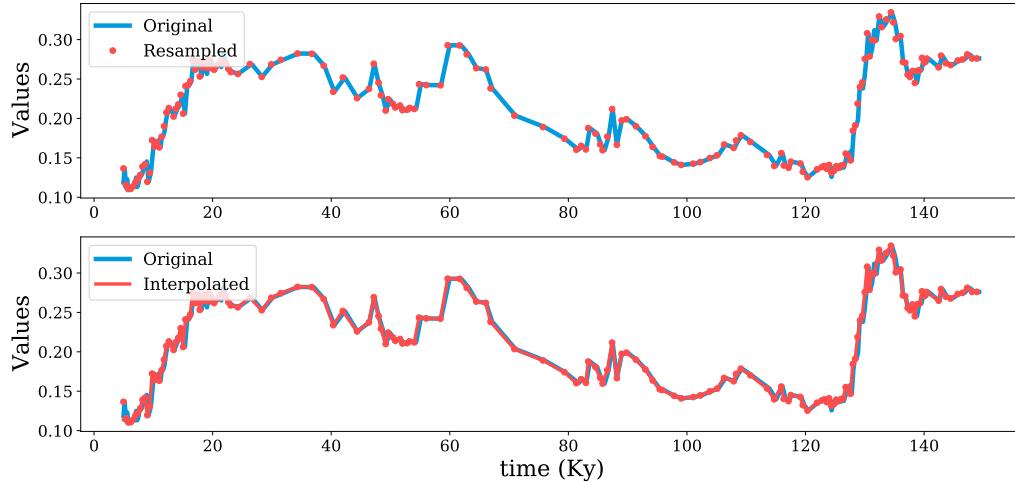


Figure 6.5. Original, resampled and interpolated ML1208-17PC dataset: The upper figure shows the resampled and original set, while the lower figure shows the same but with the gaps (Nan's) closed by linear interpolation. Red dots in the upper figure may look non-equidistant due to the appearance of Nan values in the set.

Thus, Table 6.4 show the final resolutions (bottom resolutions) used to resampled the datasets. Here we can see a wide range of resolutions, going from 6y in Talos Dome, up to 1697y in Vostok dust dataset.

Table 6.4. Datasets bottom resolution: Dust/Dust flux resolutions are shown in the upper half, while temperatures resolution are shown in the bottom half.

Name	Uniform resolution (Ky)
Vostok	1.697
EPICA Dome C	0.025
Talos Dome	0.006
Central South Pacific (PS75/59-2)	0.222
Central Pacific Ocean (ML1208-17PC)	0.402
Central Pacific Ocean (ML1208-31BB)	0.196
Central Pacific Ocean (ML1208-37BB)	0.713
Xifeng - China	1.18
RECAP	0.05
GRIP	0.414
NGRIP	0.051
Siple Dome	0.146
WAIS	0.033
Vostok	0.468
Dome Fuji	0.5
Epica Dome C	0.51
EPICA EDML	0.16
Sajama - Bolivia	0.1
Lake Tanganyika - Eastern Africa	0.348
Sanbao Cave - China	0.2
Central California SST	0.344
GRIP	0.091
NGRIP	0.02
NEEM	0.008

7. HAAR FLUCTUATION ANALYSIS FOR EQUIDISTANT DATASETS

In this section, we show the Haar fluctuations of the uniform datasets. We started by analyzing the complete sets, regardless of their data length differences, but we quickly noticed the need to restrict the analysis to shorter and shorter periods as the bottom resolutions do not seem fine enough to cover t_c . As we will show in section 9.1, Sanbao Cave time series presented a behavior highly unrelated to the rest of the database, so we did not consider it in the analysis.

7.1. Complete datasets

We show in Figure 7.1 and 7.2 the fluctuations RMS of the complete temperature and dust datasets, respectively. To start having an idea of the spatial variability, we sorted the figures by latitude, i.e., with (a) being the northernmost proxy and (l) or (k) (temperature and dust, respectively) being the southernmost. In general terms, and regardless of the latitude, we notice that fluctuations eventually tend to increase with scale.

Looking at Figure 7.1, we notice that NEEM presented a clear decrease in its centennial variability, showing a minimum around $\log_{10} \Delta t = -0.7$ or 200 years. Subtle decreases are also visible in WAIS and Lake Tanganyika at $\log_{10} \Delta t = -0.5$ and 0.4, respectively, but the time scales covered look to be too short to draw strong conclusions. Regarding the other locations, fluctuations presented straight left-side tails and did not show clear scale symmetry transitions.

We also noticed that some datasets had a too low resolution to catch centennial-scale fluctuations, such as Vostok, Dome Fuji, EDC, or Central California, which makes sense if we consider the bottom resolutions used in the resampling (see Table 6.4) and the fact that the smaller time lag assess by the uniform Haar function is two times the data resolution.

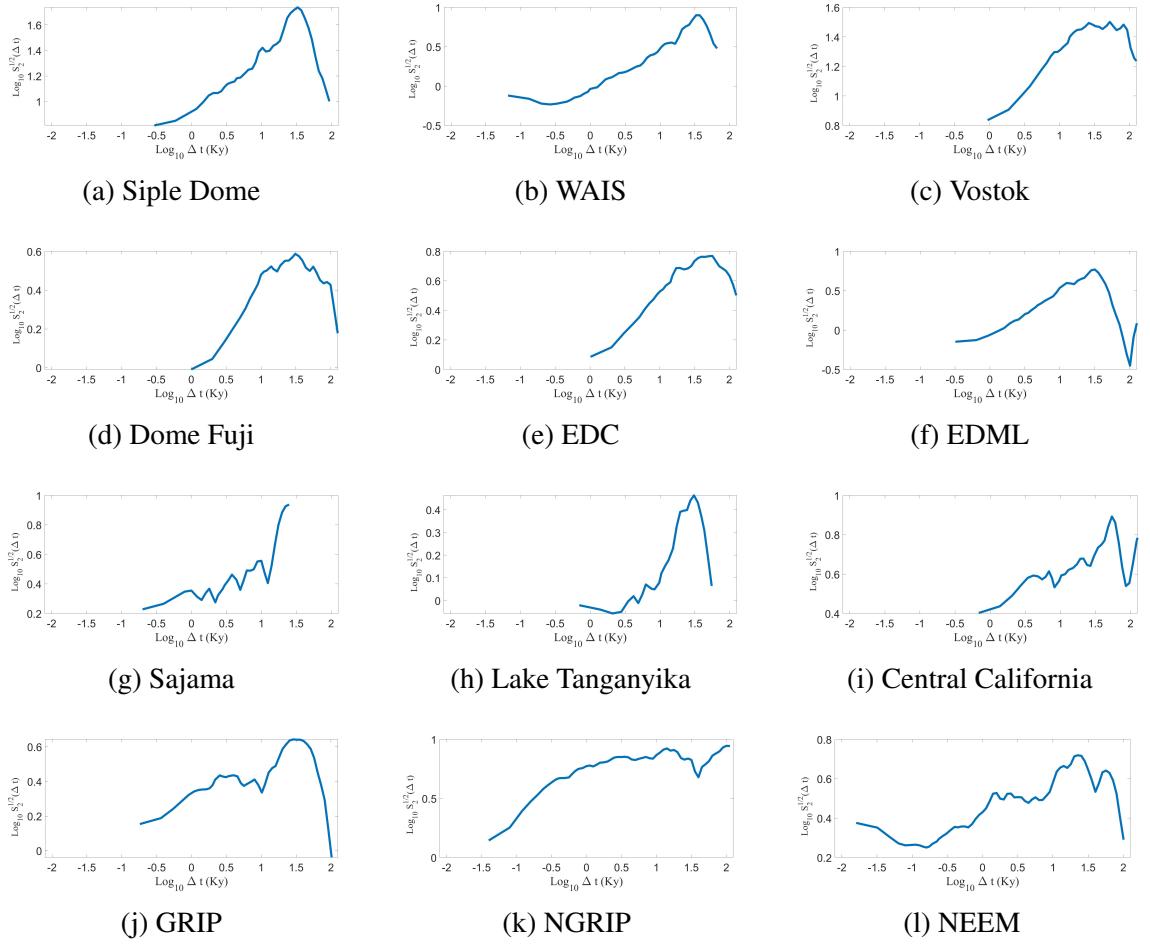


Figure 7.1. Complete uniform temperature datasets Haar fluctuations. Figures presents the RMS curves on a log-log plot, with $\log_{10} \Delta t$ in the x-axis, and $\log_{10} S_2^{1/2}(\Delta t)$ in the y-axis. Time is displayed in kiloyears.

On the other hand, dust datasets fluctuations in Figure 7.2 show almost only straight left-side tails and do not cover enough time scales to see macroweather fluctuations, with the exception of Talos Dome and EDC. The latter presented a centennial variability decrease around $\log_{10} \Delta t = -0.5$, whereas Talos Dome, even covering a larger range, did not present a clear minimum in its fluctuations.

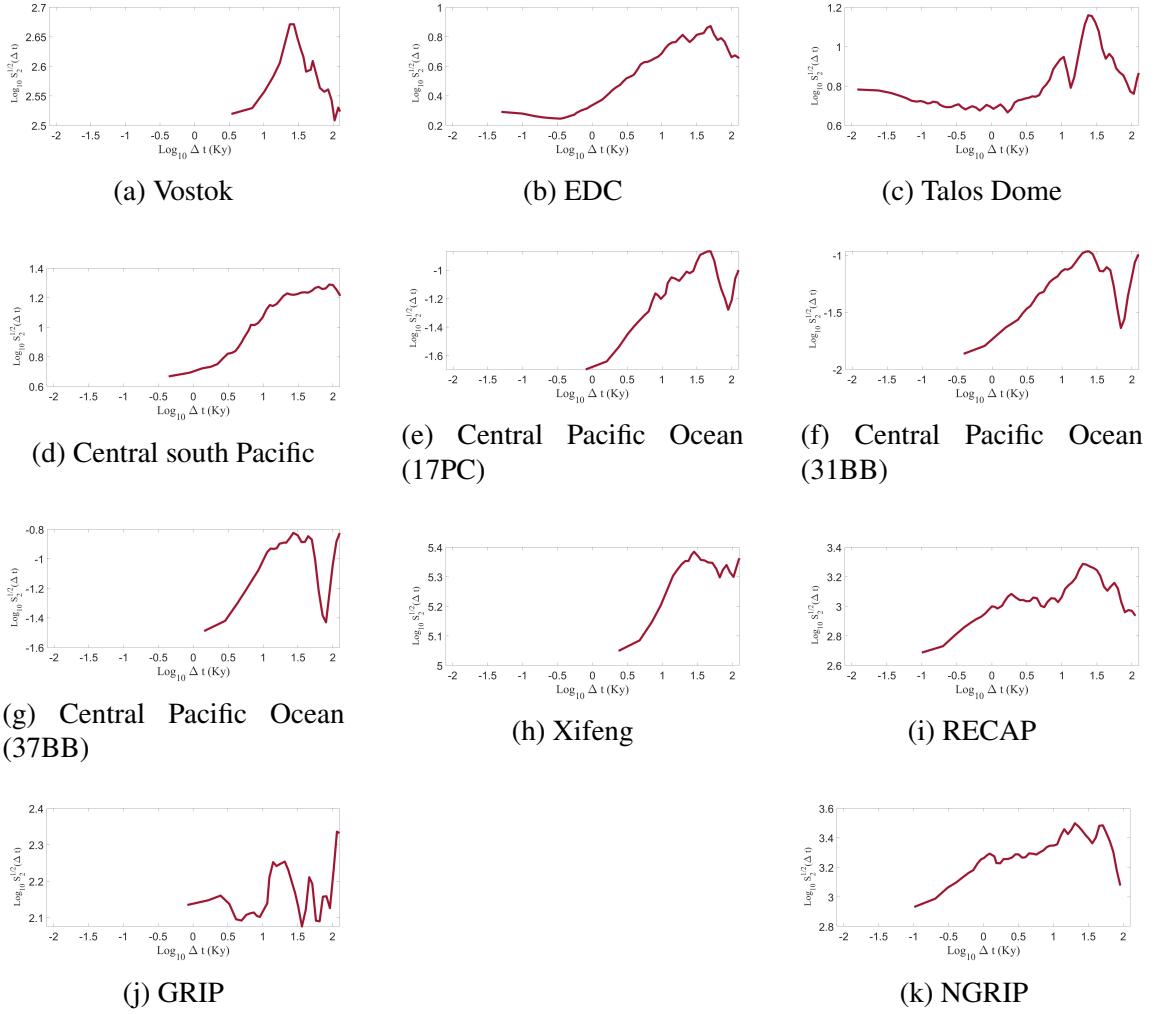


Figure 7.2. Complete uniform dust datasets Haar fluctuations. As in Figure 7.1, the graph shown the RMS in a log-log plot.

We presumed that resolution was the main limitation in the analysis since sometimes it was not enough to cover the macroweather regime, but also because lower resolutions imply fewer samples and hence less fluctuation to assess. For these reasons, we decided to restrict the analysis to the Last Glacial Cycle (130Ky) since it could improve the resolutions.

7.2. Last Glacial Cycle (LGC)

For this analysis, we restricted its time extension to the Last Glacial Cycle and re-estimated the bottom resolutions for all datasets. We also decided to work only with the sets whose extension was 130ky and no shorter, as then we could make more accurate comparisons between fluctuations. The new bottom resolutions are shown in Table 7.1.

Table 7.1. Last Glacial Cycle bottom resolutions: Dust/Dust flux resolutions are shown in the upper half, while temperatures resolution are shown in the bottom half. Values in red represent losses in resolution.

Name	Whole dataset Uniform resolution (Ky)	130Ky Uniform resolution (Ky)
Vostok	1.697	0.392
EPICA Dome C	0.025	0.025
Talos Dome	0.006	0.006
Central South Pacific (PS75/59-2)	0.222	0.148
Central Pacific Ocean (ML1208-17PC)	0.402	0.521
Central Pacific Ocean (ML1208-31BB)	0.196	0.439
Central Pacific Ocean (ML1208-37BB)	0.713	2.264
Xifeng - China	1.18	2.26
GRIP	0.414	0.124
<hr/>		
Vostok	0.468	0.06
Dome Fuji	0.5	0.25
Epica Dome C	0.51	0.045
EPICA EDML	0.16	0.133
Central California SST	0.344	0.883

As is remarked in Table 7.1, some bottom resolutions in fact decreased with the time shortening, which was not expected beforehand. To understand this, we show in Figure 7.3 the Central California original (non-uniform) resolutions for its complete dataset and its last 130 Ky. We colored in gray the bottom 10% of the samples, from where the bottom resolution is estimated. As we can see, the resolution close to the 130ky is worse than the one close to 160ky (complete set extension), which explains why some resolutions decreased.

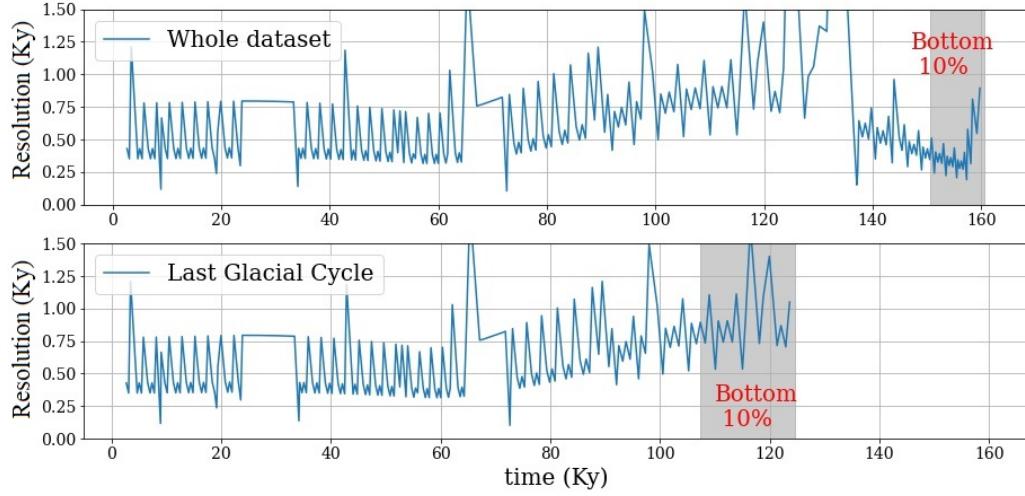


Figure 7.3. Central California original non-uniform resolution comparison: (Upper figure) Complete dataset, (lower figure) Last Glacial Cycle. The gray areas represent the bottom 10% of the samples.

Therefore, when the new resolutions got worse, we still cut the data to 130ky but resampled it with its previous bottom resolution, i.e., the resolutions obtained by using the complete datasets.

Hence, Figure 7.4 and 7.5 show fluctuations RMS from the temperature and dust dataset, respectively. Since the resolutions improved, many figures covered wider timescale ranges than before. By looking at the temperatures, we notice that EDC now catches centennial timescales and shows a minimum around $\log_{10} \Delta t = -0.4$ or 400y, whereas the other locations do not show any macroweather scale symmetry. Similarly, fluctuations from dust datasets do not present any clear improvement with the exception of EDC, which presented a slightly more pronounced minimum now.

As the results were still not good enough to characterize the centennial variability, we decided to once again narrow down the analysis to improve the resolutions, this time restricting it to the Holocene and the Last Glacial Maximum.

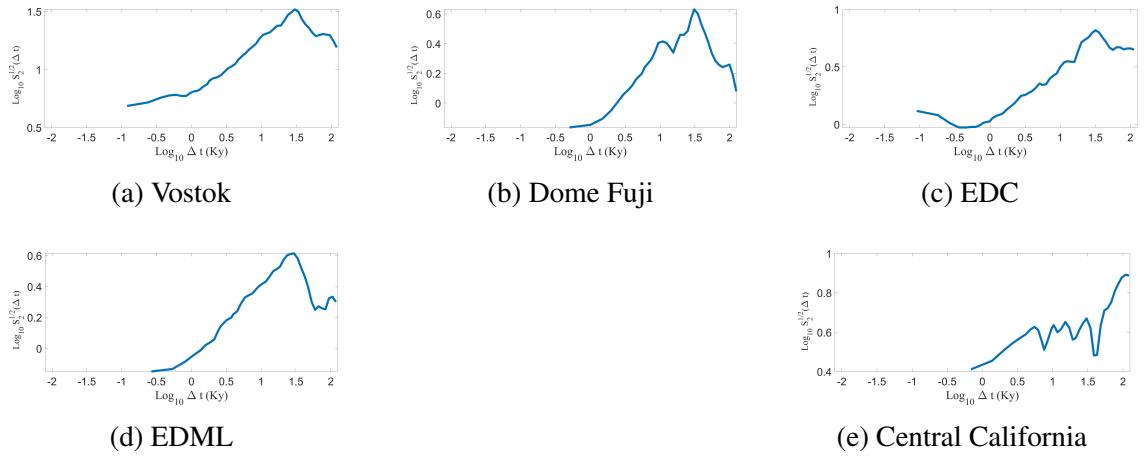


Figure 7.4. Last Glacial Cycle uniform temperature datasets Haar fluctuations.

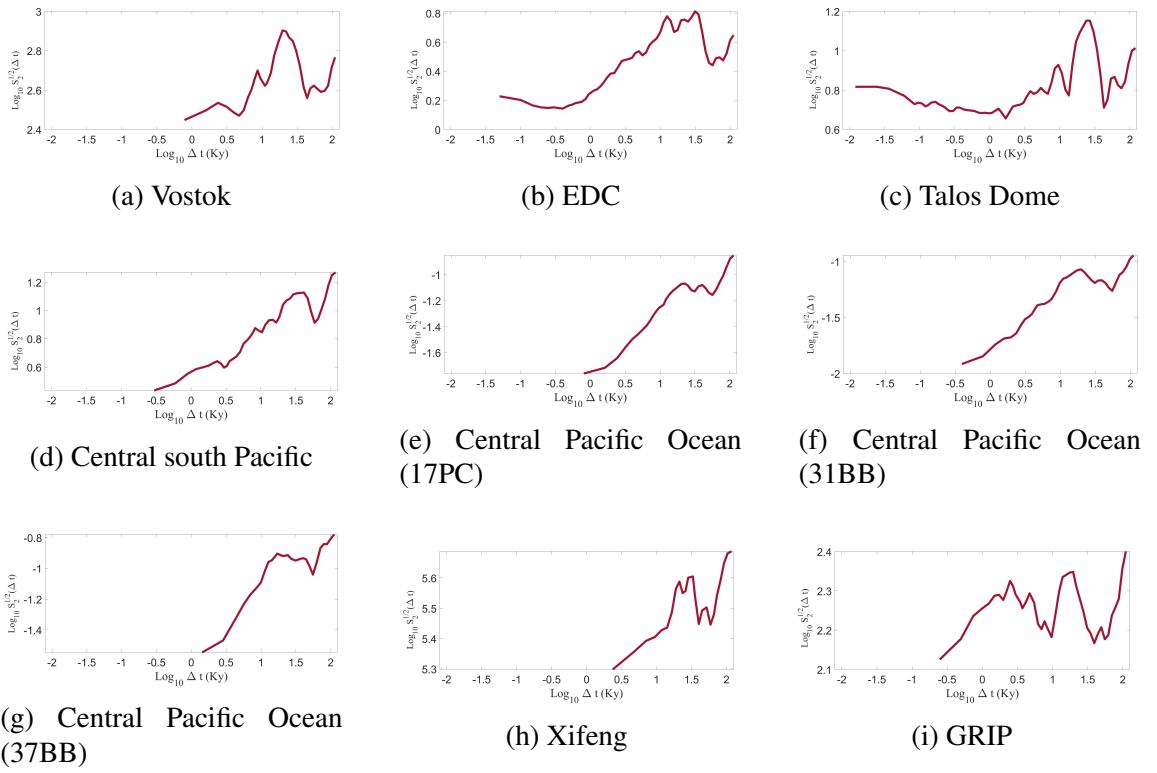


Figure 7.5. Last Glacial Cycle uniform dust datasets Haar fluctuations.

7.3. Holocene and Last Glacial Maximum

This time we worked with two 10 ky periods:

- The Holocene, which is the current geological epoch and covers approximately the last 10 ky B.P.
- The Last Glacial Maximum (LGM), which corresponds to the most recent time that the global ice sheets were at their greatest extent. We considered it as the last [18 - 28] ky B.P.⁴

Table 7.2 compares the original (whole dataset), Holocene and LGM bottom resolutions. It is important to notice that the new time shortening led to a notorious decrease in the data lengths, sometimes going from hundreds to a dozen samples. Then, to ensure at least a centennial resolution, we considered only the sets with no less than 100 samples.

Table 7.2. Holocene and Last Glacial Maximum bottom resolution: Dust/Dust flux resolutions are shown in the upper half, while temperatures resolution are shown in the bottom half.

Name	Whole dataset Uniform resolution (Ky)	Holocene Uniform resolution (Ky)	LGM Uniform resolution (Ky)
EPICA Dome C	0.025	0.025	0.025
Talos Dome	0.006	0.001	0.001
RECAP	0.05	0.05	0.05
GRIP	0.414	0.006	0.038
NGRIP	0.051	—	0.031
Siple Dome	0.146	0.016	0.09
WAIS	0.033	0.005	0.014
Vostok	0.468	0.049	0.089
Epica Dome C	0.51	0.019	0.056
EPICA EDML	0.16	0.022	0.045
Sajama - Bolivia	0.1	0.1	—
GRIP	0.091	0.006	0.033
NGRIP	0.02	—	0.02
NEEM	0.008	0.001	0.004

⁴A more accurate LGM reference timing would be [19 - 26.5] ky B.P. according to (Clark et al., 2009).

Figure 7.6 shows the Holocene temperature fluctuations RMS. This time, millennial variability is much lower than before since we are covering only 10Ky, and thus we manage to see a global minimum around the range $\log_{10} \Delta t = [-0.5, 0]$ in Siple Dome and WAIS but not in the other locations.

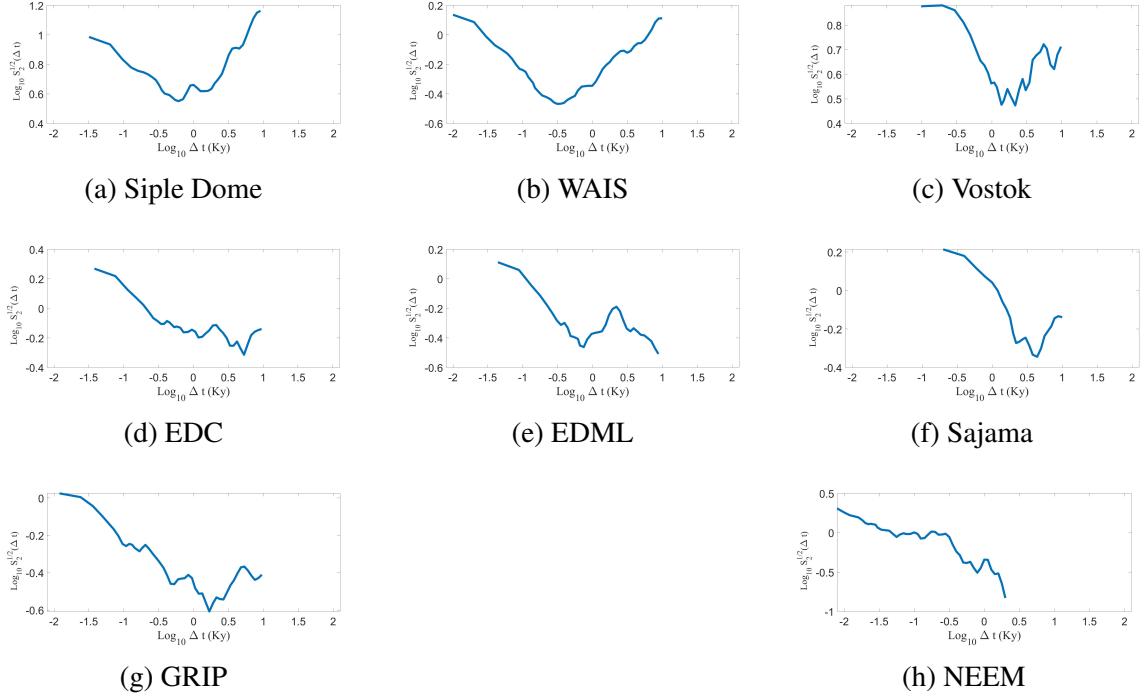


Figure 7.6. Holocene uniform temperature datasets Haar fluctuations.

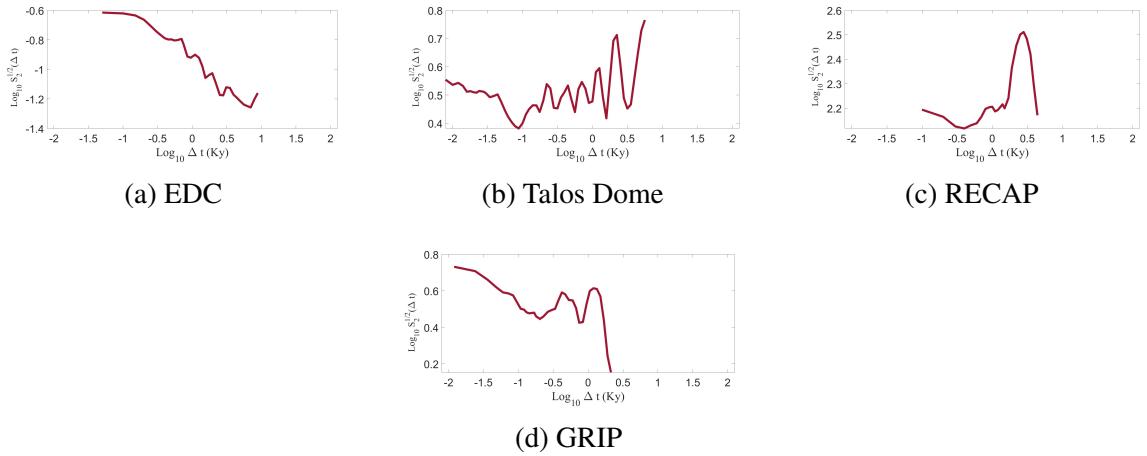


Figure 7.7. Holocene uniform dust datasets Haar fluctuations.

Similarly, Figure 7.7 shows the Holocene dust fluctuations RMS. In this case, both EDC and GRIP presented a decreasing variability with scale, while Talos Dome and RE-CAP seem to eventually increment their fluctuation with scale. Since all four figures presented different behaviors, we can not point to any presumable regime shift or conclusion.

On the other hand, Figure 7.8 presents the LGM temperature fluctuations RMS. This time we did not find similarities between locations since some of them increased with scale (e.g., Siple Dome), others decreased with scale (e.g., EDC and NEEM), while others did not show a consistent time scale symmetry (e.g., GRIP or Vostok). Nevertheless, we noticed that WAIS and EDML presented a decrease in their variability around the range $\log_{10} \Delta t = [-0.5, 0]$, which is the same found in the Holocene.

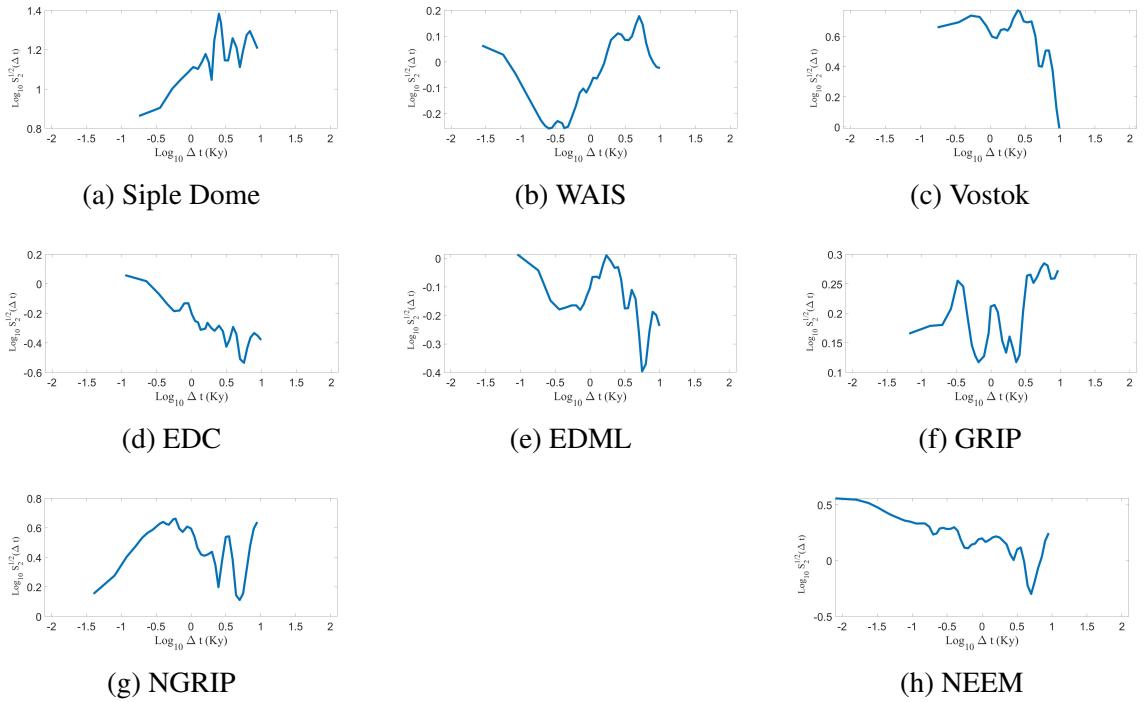


Figure 7.8. Last Glacial Maximum uniform temperature datasets Haar fluctuations.

Finally, Figure 7.9 shows the LGM dust fluctuations RMS. Once again, the results are very different between themselves and do not draw any clear conclusion. Only EDC seems

to present a climate regime shift around $\log_{10} \Delta_t = -0.5$, which would be consistent with the Last Glacial Cycle EDC fluctuations, but the figure is still too ambiguous to conclude.

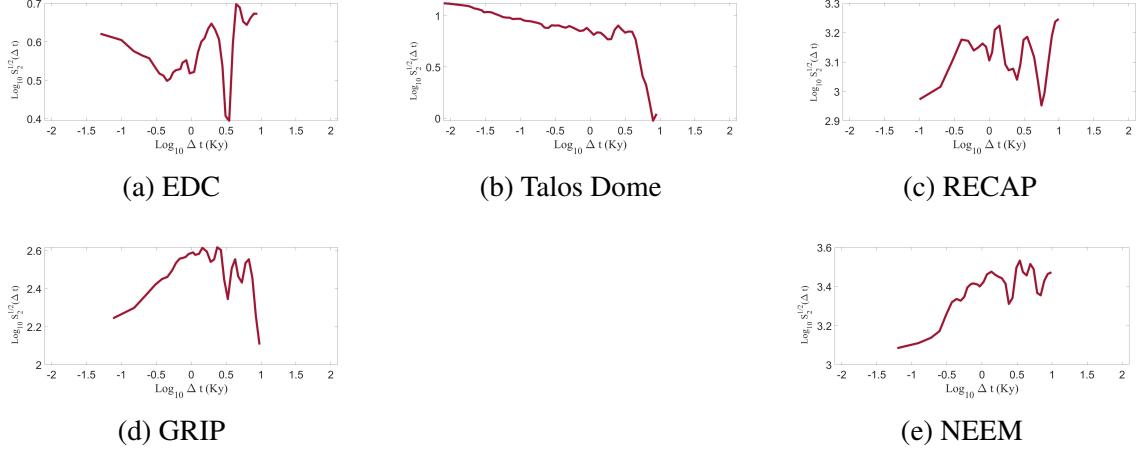


Figure 7.9. Last Glacial Maximum uniform dust datasets Haar fluctuations.

So far, we have used the uniform Haar function to analyze the datasets without being able to identify either climate regimes or timescales symmetries. In addition, whilst restricting the analysis to different periods, we noticed that the function outcomes were sensitive to data length and resolution. For this reason, we decided to apply a sensitivity analysis to verify how reliable the results were.

7.4. Sensitivity analysis

To test the uniform Haar function sensitivity to data length and resolution, we used the “EDC_DustFlux_25yr.tab” file⁵ to compute its fluctuations under three different scenarios, which are summarized in Table 7.3. For each case, we computed and compared the Haar fluctuations outcomes, hence testing the effects of: only resolution variations (Test 1), only data length variations (Test 2), and the combined effect of both (Test 3).

⁵The same file used in 4.1, which contains the EDC 25-year constant time-step averages of the high-resolution 1 cm data.

Table 7.3. Uniform Haar function sensitivity test.

Test id	Type	Data length (Ky)	Resolution (y)
1	Resolution	120	1
1	Resolution	120	10
1	Resolution	120	25
2	Data length	60	25
2	Data length	120	25
2	Data length	240	25
3	Combined	60	1
3	Combined	120	10
3	Combined	240	25

Figure 7.10 shows the outcomes from *Test 1*, where we computed the EDC fluctuations RMS using three different resolutions. We can see how the mid and low frequency range of the curves are practically the same, whereas time lags under $\log_{10} \Delta t \approx -0.5$ (i.e., $\Delta t \leq 316$ y) shown big differences.

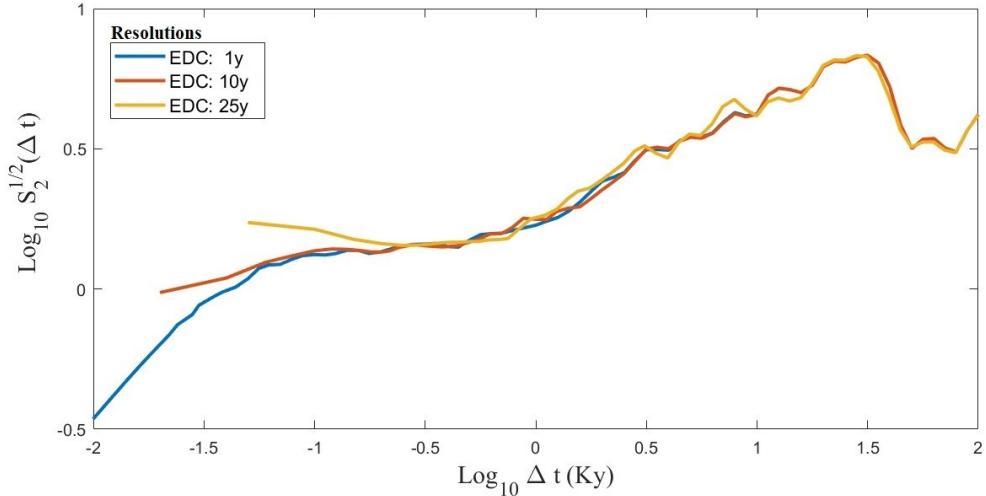


Figure 7.10. Uniform Haar function sensitivity test 1: Resolution test.

First, we can see the differences in the time lag coverage, where finer resolutions covers fluctuations at smaller time lags and thus its RMS curves have longer left tails. More remarkable is the fact that finer resolutions resulted in an underestimation of fluctuations

at high frequencies, showing a “pulling down” effect over the curves. This effect is because we used linear interpolation to re-sample the sets, which filtrates the high frequency components.

Then we continue with the *Test 2*, where we tested the effects of shortening the dataset while maintaining its resolution. Figure 7.11a shows how the general features of the fluctuations RMS hold, but the values themselves differ from case to case. These differences comes as we are taking the fluctuation averages over each disjoint set (at each Δt), hence, at smaller data lengths, we will inevitably take into account less fluctuations and the final averaged value will change. On the other hand, Figure 7.11b shows the combined effect of changing both resolution and data length (*Test 3*). As expected at this point, this resulted in three curves with a similar shape but with some differences over the whole curve (fluctuations over or underestimation), while the pulling-down effect on the left side was still visible.

This test pointed out the need for an interpolation-free algorithm that allows us to analyse the Haar fluctuations using climatic datasets in its raw format, i.e., with non-uniform resolution.

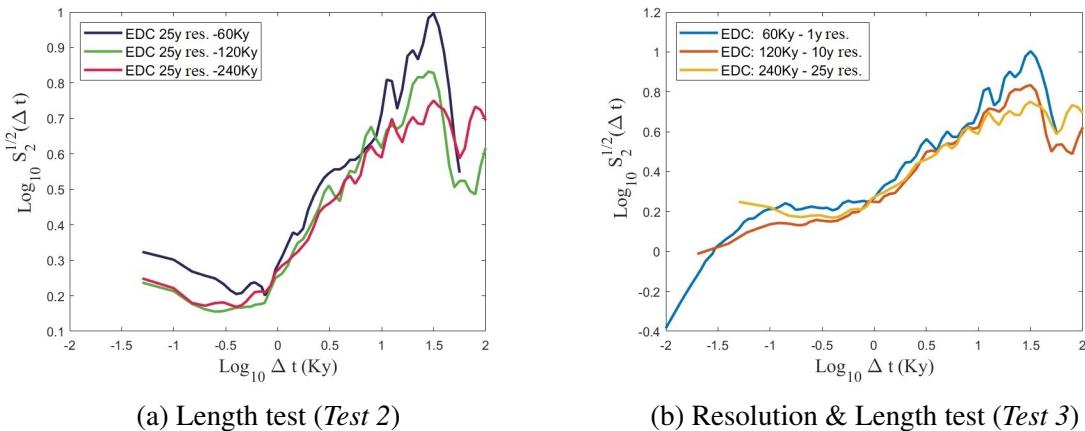


Figure 7.11. Uniform Haar function sensitivity test 2 and 3: Resolution & length test.

8. COMPARISON BETWEEN HAAR FUNCTIONS FOR EQUIDISTANT AND NON-EQUIDISTANT DATASETS

In this section, we compared the performance of the Haar function for equidistant datasets (detailed in section 4.1) and our Haar function for non-equidistant datasets (detailed in section 4.2). To do so, we used the real data shown in chapter 6 to compute and compare its fluctuations RMS using the three following combinations:

- (i) Matlab Haar functions on uniform dataset.
- (ii) Python Haar function on uniform dataset.
- (iii) Python Haar function on non-uniform dataset.

This way, we could also check if the Python algorithm improves the results when working with previously interpolated data sets. For instance, we show in Figure 8.1 the outcomes obtained from two temperature data set (EPICA EDML and Vostok), and two dust data set (EPICA EDC and Talos Dome), where notorious differences can be easily remarked. Columns from left to right correspond to scenarios (i), (ii), and (iii), respectively.

First, by comparing the left and middle columns, we see how the fluctuations RMS from Python are notoriously smoother than the ones from Matlab in all four locations, which eases the identification of peaks and valleys. This improvement is presumably because the Python code avoids data binning and takes averages for every assessed time step ($\langle |\Delta F(\Delta t)| \rangle$). For the same reason, Matlab fluctuations magnitudes are slightly smaller (bigger) at higher (at lower) frequencies compared to the ones from Python.

On the other hand, if we compare the middle and right side columns, we can see that the fluctuations RMS from non-uniform datasets reach much shorter time steps than before since the non-equidistant sampling allows the function to assess a broader combination of time steps. At the same time, this results in more accurate fluctuations estimations (as the number of disjoint sets has grown) and increases the chances of getting the macro-weather to climate transition at centennial time scale.

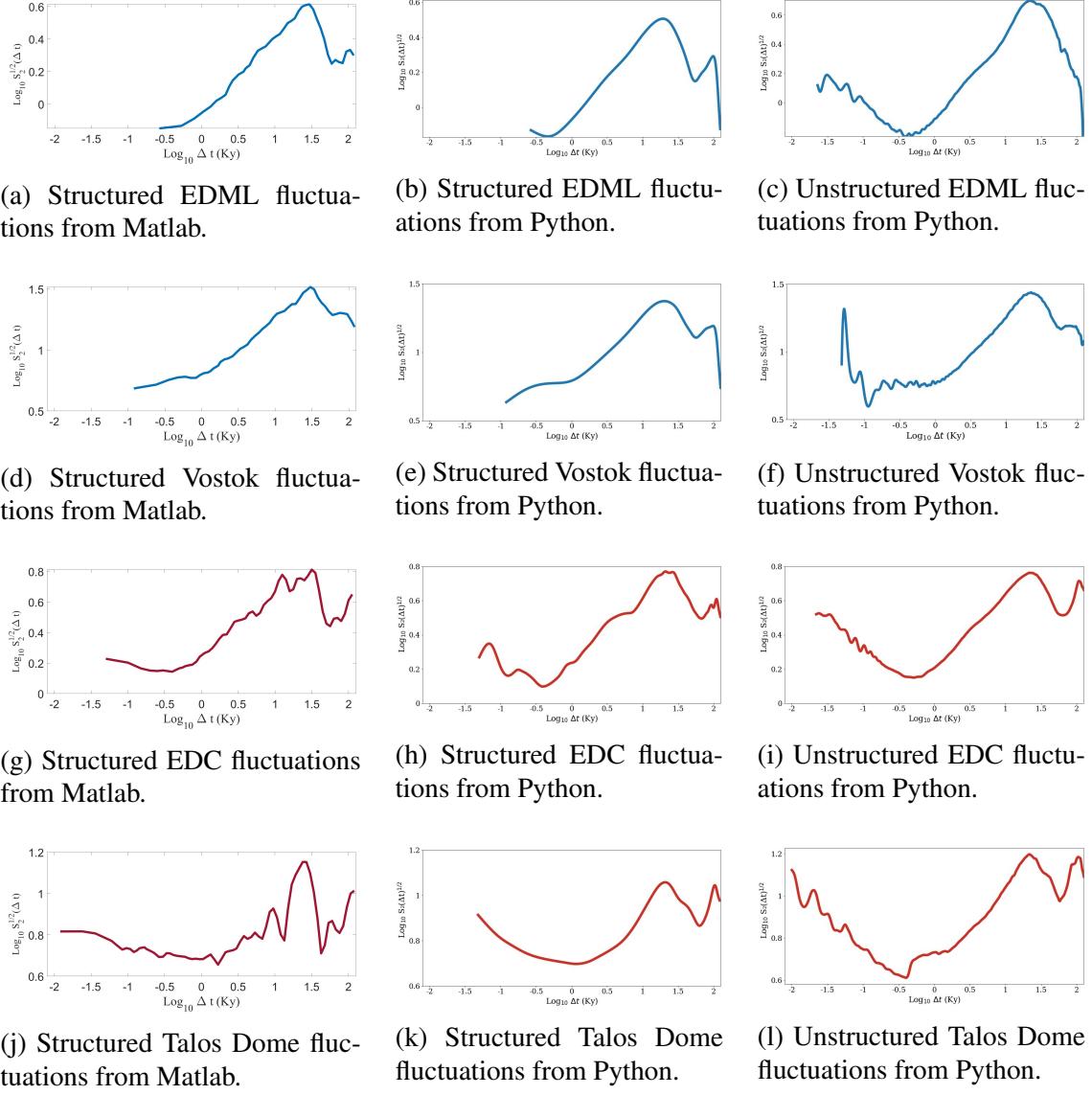


Figure 8.1. Matlab and Python fluctuations RMS comparison. Left and middle columns came from structured (interpolated) data sets, while the right column used the originals (non-equidistant) data sets. Blue and red curves refer to temperature and dust measurements, respectively.

9. CLIMATIC INTERPRETATION

9.1. Dataset correlation

Before comparing the atmospheric variability between different locations, we start by verifying how related its time series are. If there were no linkage between them, the τ_c comparison would not be a helpful study approach. For this reason, we compute the Spearman Correlation coefficient between datasets following the steps detailed in section 3.3. Figure 9.1 show the coefficients for both dust/dust flux and temperatures measurements, where we can notice almost only positive correlations.

We expected positive correlations since each proxy type is sensitive to specific Earth processes, so the same phenomena that drive an increase or decrease in temperature (or dust) in the north should also do it in the south. While this is not always straightforward at shorter periods (given the many local interactions), the patterns become visible when working with long enough time series as the ones used here, which is why we can identify, for instance, the Glacial and Interglacial Cycles.

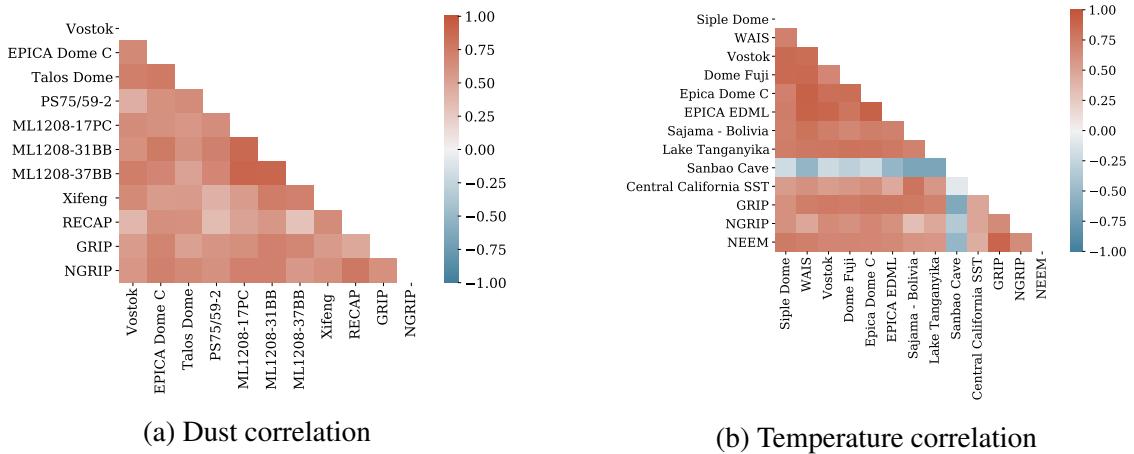


Figure 9.1. Spearman correlation coefficients

On the other hand, Sanbao Cave was the only dataset with no positive correlations, presumably because its values were from Speleothem measures which are sensitive to the

monsoon rain cycles over China. Figure 9.2 compares the Vostok, EDC and Sanbao cave cycle patterns, where the Y-axis has been adapted to ease the visual comparison. The gray areas represent roughly the global interglacial intervals.

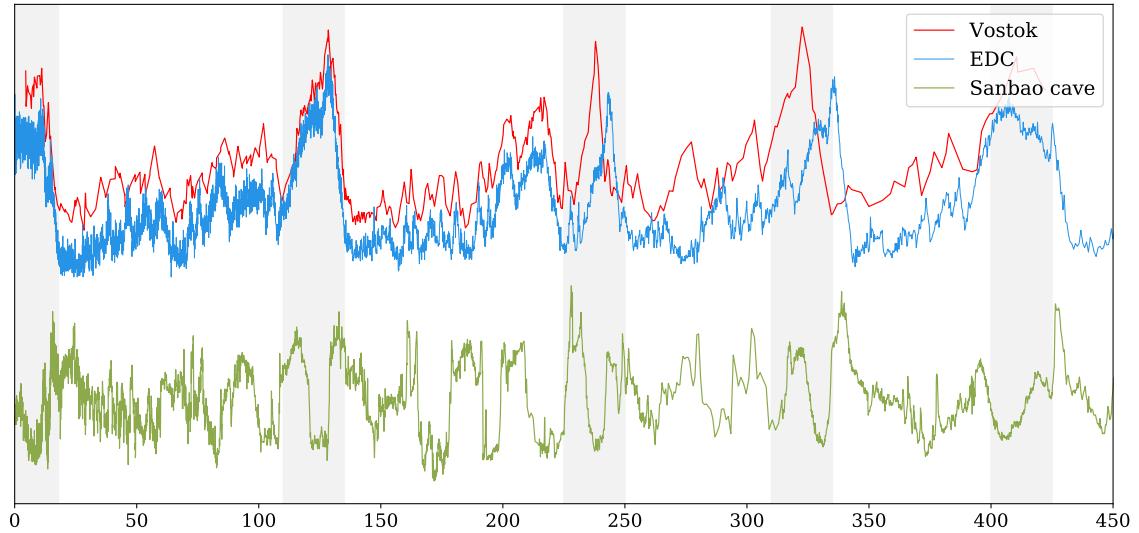


Figure 9.2. Vostok, EDC and Sanbao cave cycle comparison. We can easily notice the difference between the 120 Ky cycle followed by Vostok and EDC, and the 30 Ky-period cycles of Sanbao cave.

It is clear how dust deposition and temperature variations are sensitives to glaciation cycles, following approximately a 100-130 ky period, whereas Sanbao cave looks to follow much shorter cycles patterns. For this reason, we decided to not used Sanbao cave in the rest of this work, as we are mainly interested in the fluctuation analysis over glacial cycles.

9.2. LGC Haar fluctuation analysis: Non-equidistant case

In this section we analyse fluctuations over the Last Glacial Cycle the Haar function for non-equidistant datasets. As we aimed to make a comparison between different locations and taking into account the wide variety of data extensions (from 25ky to 800ky), we decided to restrict ourselves to a maximum range of 130 ky (i.e., the LGC). Unlike section 7.2, this time we did consider sets whose extension was within the Last Glacial Cycle

range. Also, we must mention that RECAP, Dome Fuji, Sajama, and NGRIP temperatures were not available in raw format, but we decided to keep and use them as mere references since their information could (or not) complement our broader analysis.⁶

Figure 9.3 shows the final RMS fluctuations from all temperature datasets. In general, figures are smooth and spikeless, showing more stable states at mid frequencies (centennial-millennial scale) and higher fluctuations to the lower ones. In many cases (e.g., Siple Dome, WAIS, or GRIP), fluctuations at centennial-millennial scale are also the global minimum of the plot, giving us an initial idea of the scale where the macro-weather to climate transition occurs. Also, we notice in Lake Tanganyika (Figure 9.3h) a significant drop at $\log_{10} \Delta t = -0.44$, which looks odd compared with its other time lags, but since there are no other lake sediments nor African datasets, we cannot presume the origin of this fall. However, if we overlook this drop, we notice how this site also has higher fluctuations towards the extremes and lower variability at centennial-millennial scale.

On the contrary, even if Sajama and Central California show an increasing variability with scale (Figures 9.3g and 9.3i, respectively), they do not present any clear peak nor minimum, presumably because of the low amount of samples or the linear interpolation in the Sajama case. Similarly, NGRIP (Figure 9.3k) also presented a different RMS behavior with an increasing variability from early time lags, showing an equally unstable state for both centennial and millennial scales.

On the one hand, Figure 9.4 shows the final RMS fluctuations from all dust datasets. This time, identifying a pattern between sites is not straightforward, and fluctuations look more spiky and irregular than before. On the other hand, we have EDC and Talos Dome showing clearly a minimum at centennial scale and a maximum at millennial scale, thus comparable with the plots from Figure 9.3. Then, we notice that RECAP, GRIP, and NGRIP fluctuations rapidly rise to an unstable state and sustain it from decennial to millennial scales, being in some way similar to Figure 9.3k. Finally, except for Xifeng, the

⁶See the comparison of scenarios (ii) and (iii) in chapter 8.

remaining sites fluctuations look to somehow increase with scale but, given its irregularities, it is not possible to draw conclusions yet.

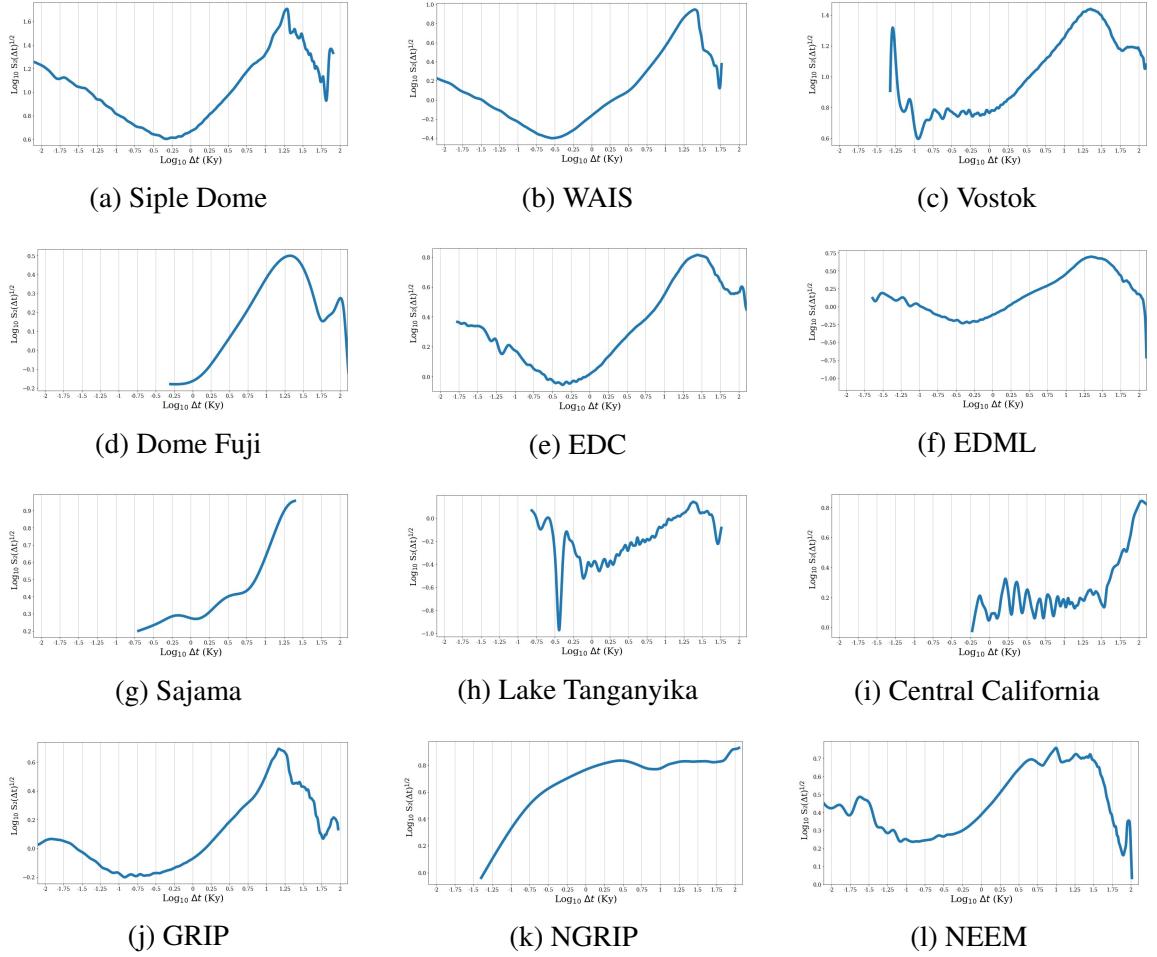


Figure 9.3. Last Glacial Cycle temperatures Haar fluctuations RMS from non-equidistant datasets. Figures presents the RMS curves on a log-log plot, with $\log_{10} \Delta t$ in the x-axis, and $\log_{10} S_2^{1/2}(\Delta t)$ in the y-axis. Time is displayed in kiloyears.

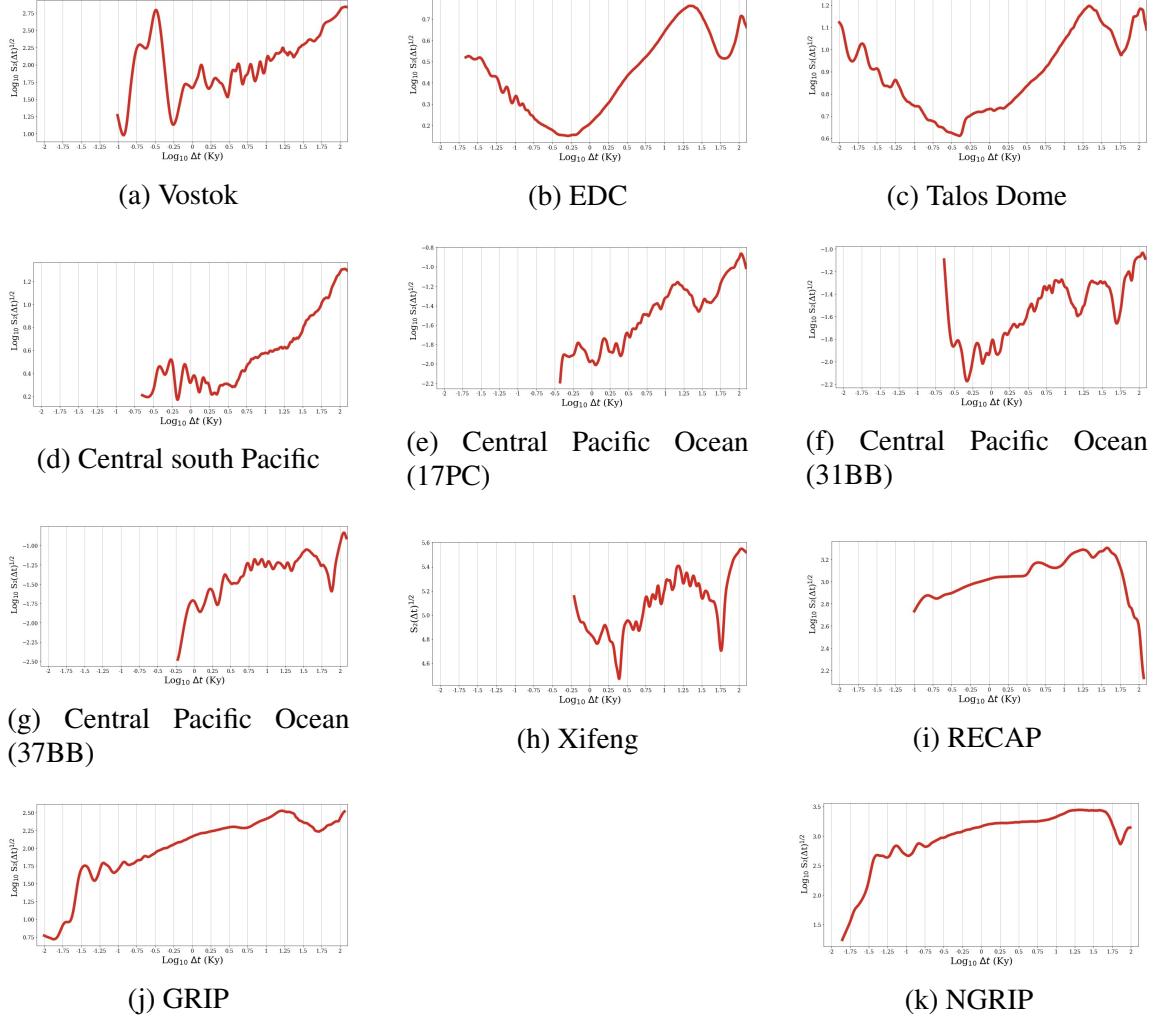


Figure 9.4. Last Glacial Cycle dust Haar fluctuations from non-equidistant datasets. As in Figure 9.3, the graph shown the RMS in a log-log scale.

Even though the resulting RMS fluctuations from the Haar function could already allow us to conclude about the global atmospheric stability, we decided to complement the study by exploring other techniques, in particular, by using one of the most common techniques in signal processing: Spectra analysis.

9.3. Spectra analysis: Non-Uniform Fast Fourier Transform

As we were now analysing non-equidistant datasets, we used the NUFFT function detailed in section 4.3 to perform the spectral analysis. Figures 9.5 and 9.6 show the temperatures and dust results obtained with the NUFFT function in a log-log scale; the x-axis is display in time lags instead of frequencies to ease the comparison with the Haar fluctuations already presented in the previous section.

Once again, the results from temperature are roughly similar, showing a decrease in spectra with scale and small spikes along the curve. Only WAIS and Central California differ from the other sites, with some odd results around 3.2ky and 7Ky, respectively. To have a rough idea of the processes behind it, we identified some of the most notorious peaks in the figures, finding that the most repeated frequencies are around 42Ky^{-1} , 11Ky^{-1} , and 4.7Ky^{-1} . On the contrary, dust figures are much more irregular and spiky, sometimes decreasing in spectra with scale (e.g., EDC) while other times presenting some global minimum (e.g., Vostok or Central Pacific Ocean-37BB). Comparing its most notorious peaks, we found that 21.7Ky^{-1} , 11Ky^{-1} , and 1.4Ky^{-1} are approximately the most repeated frequencies. These differences between temperatures and dust (either in real or frequency space) is expected since both proxies are driven by different processes, where temperatures are accurately obtained from δO^{18} isotopes measurements while dust/dust flow respond to a suite of mixed geophysical phenomena such as rainfall, winds, vegetation density, among others.

Because of the spikiness and highly variable results of the NUFFT, fitting and comparing the slopes of the curves becomes a nontrivial task, even if we cluster the analysis by latitude or proxy type. Also, while the spectral density tells us how quickly the variance changes per frequency interval (i.e., the β exponent in eq. 3.19), its physical meaning is again nontrivial and will always require some expert interpretation. For this reason, we will no longer continue analysing the spectra leaving this task as an open subject for futures researches.

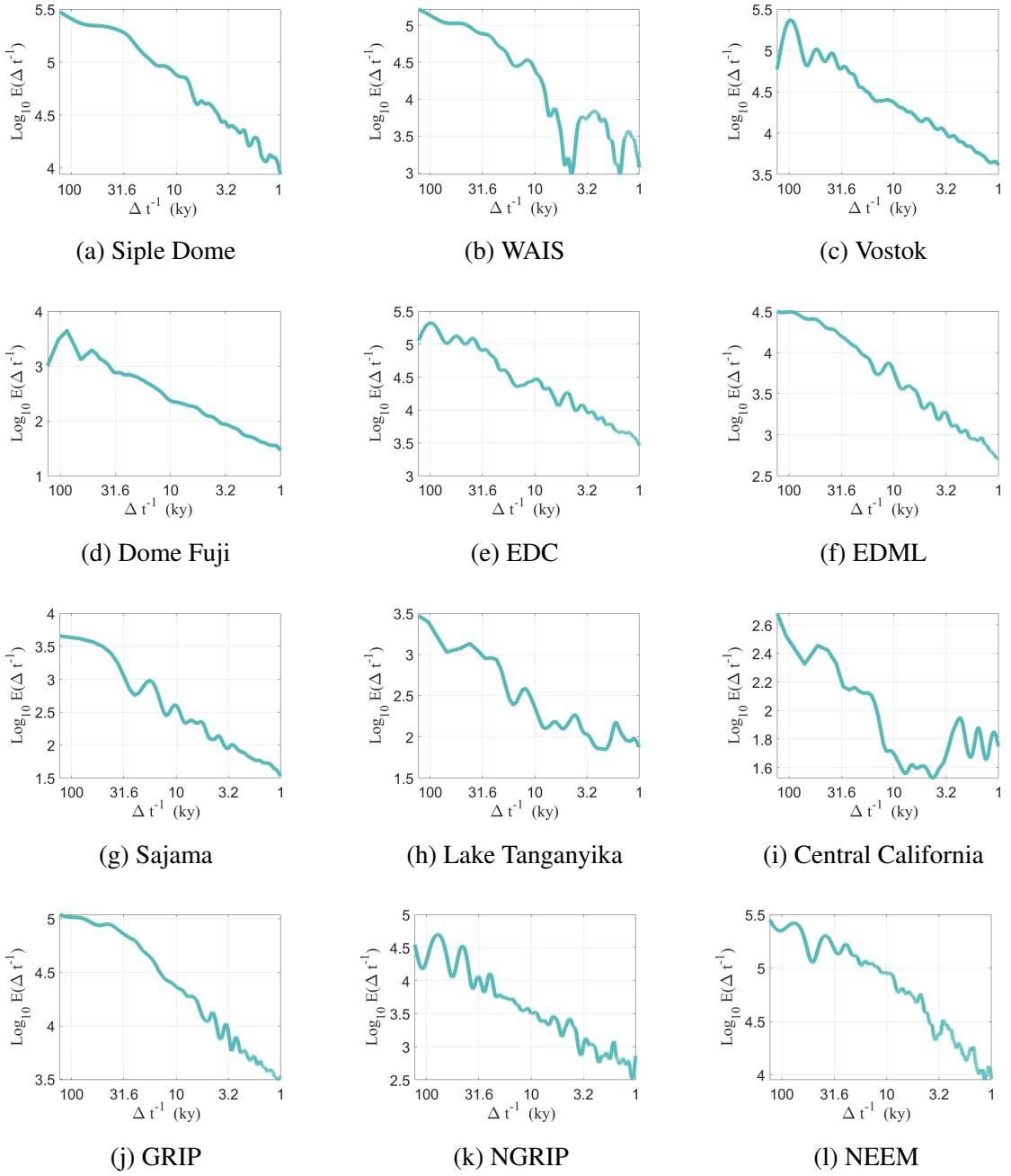


Figure 9.5. NUFFT from non-equidistant temperatures datasets. Figures are shown in a log-log scale, with the x-axis display as time lag⁻¹ in Ky.

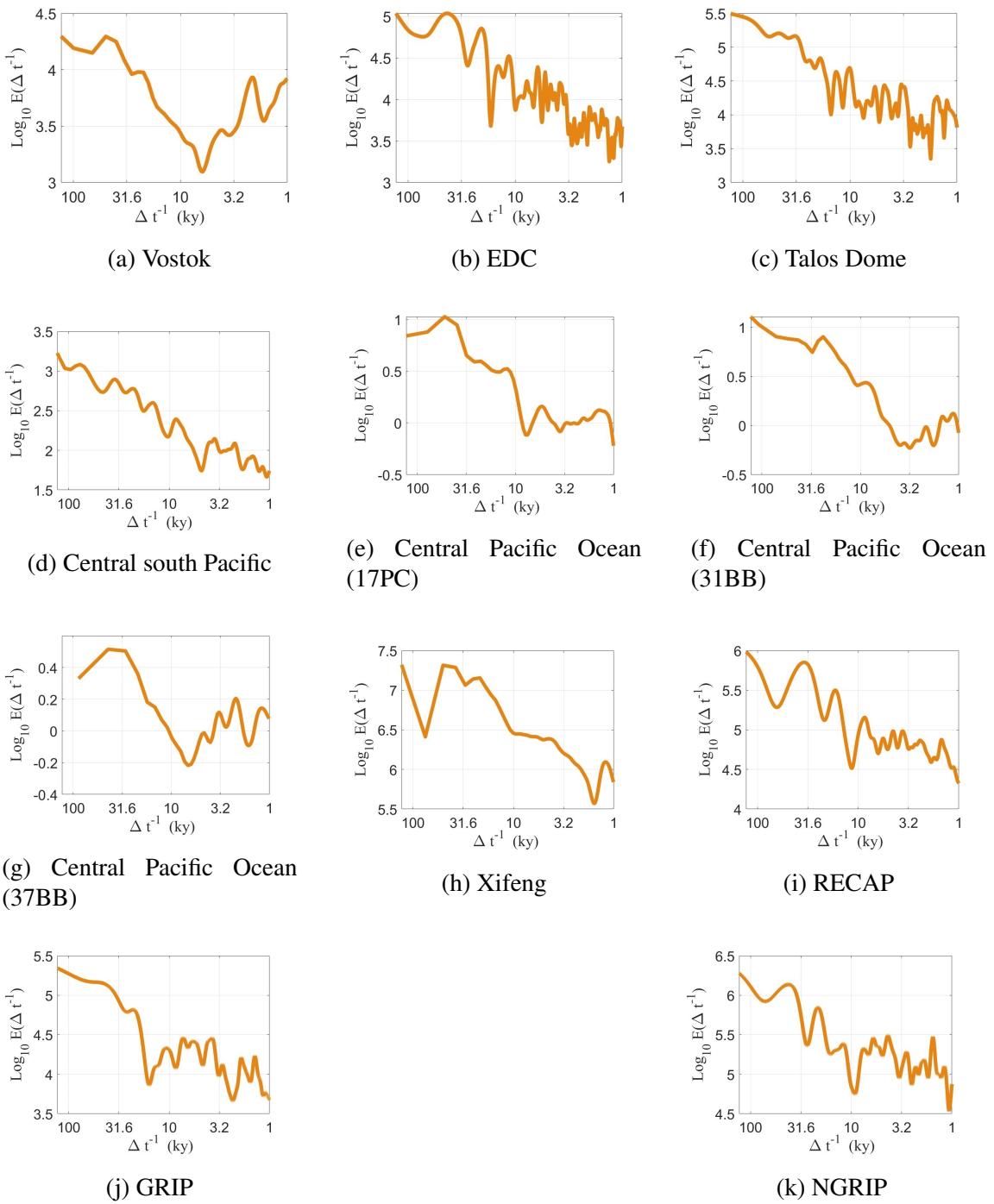


Figure 9.6. NUFFT from non-equidistant dust datasets.

9.4. Timescale transition and latitude

In this final section, we return to the Haar fluctuation RMS to further extend its analysis. This time, we clustered and analysed fluctuations by latitude, intending to produce a global mapping of atmospheric variability. Hence, we divided the Earth's 180° latitudes into five segments (see Table 6.2 upper part) under the following titles:

- Antarctica: From -90° to -60°.
- Southern mid latitudes: From -60 to -20°.
- The tropics: From -20° to 20°.
- Northern mid latitudes: from 20° to 60°.
- Greenland: From 60° to 90.

As datasets came from different sites, techniques, and times, we started by plotting the normalised form of its Haar fluctuation RMS, thus avoiding any measurement unit problem (Figure 9.7). At the same time, and to have a rough idea of the macro-weather to climate transition (τ_c), we took the original RMS fluctuations curves and fit some reference lines to analyse how slopes change with scale (Figure 9.8). Hence, by complementing both the original and normalised RMS fluctuations we were able to delve deeper into the geographical variability of dust and temperature. It is worth mentioning that because of their odd behaviour, both Sajama and Vostok (from dust) were removed from the figures to facilitate their visual interpretation.

First, we see in Figures 9.7a and 9.8a consistent results between all datasets, where both dust and temperatures present a global minimum around $\log_{10} \Delta t = -0.25$ and a peak at $\log_{10} \Delta t = [1.25, 1.5]$, suggesting a macro-weather to climate transition at centennial scale. On the contrary, Figures 9.7b, 9.7d, 9.8b, and 9.8d shows no consistent climate regime shift between dust and temperature at mid-latitudes, with a minimum around $\log_{10} \Delta t = 0.3$ for Xifeng and Central South Pacific, but with no distinguishable τ_c for Central California ($H_{10} = 0$).

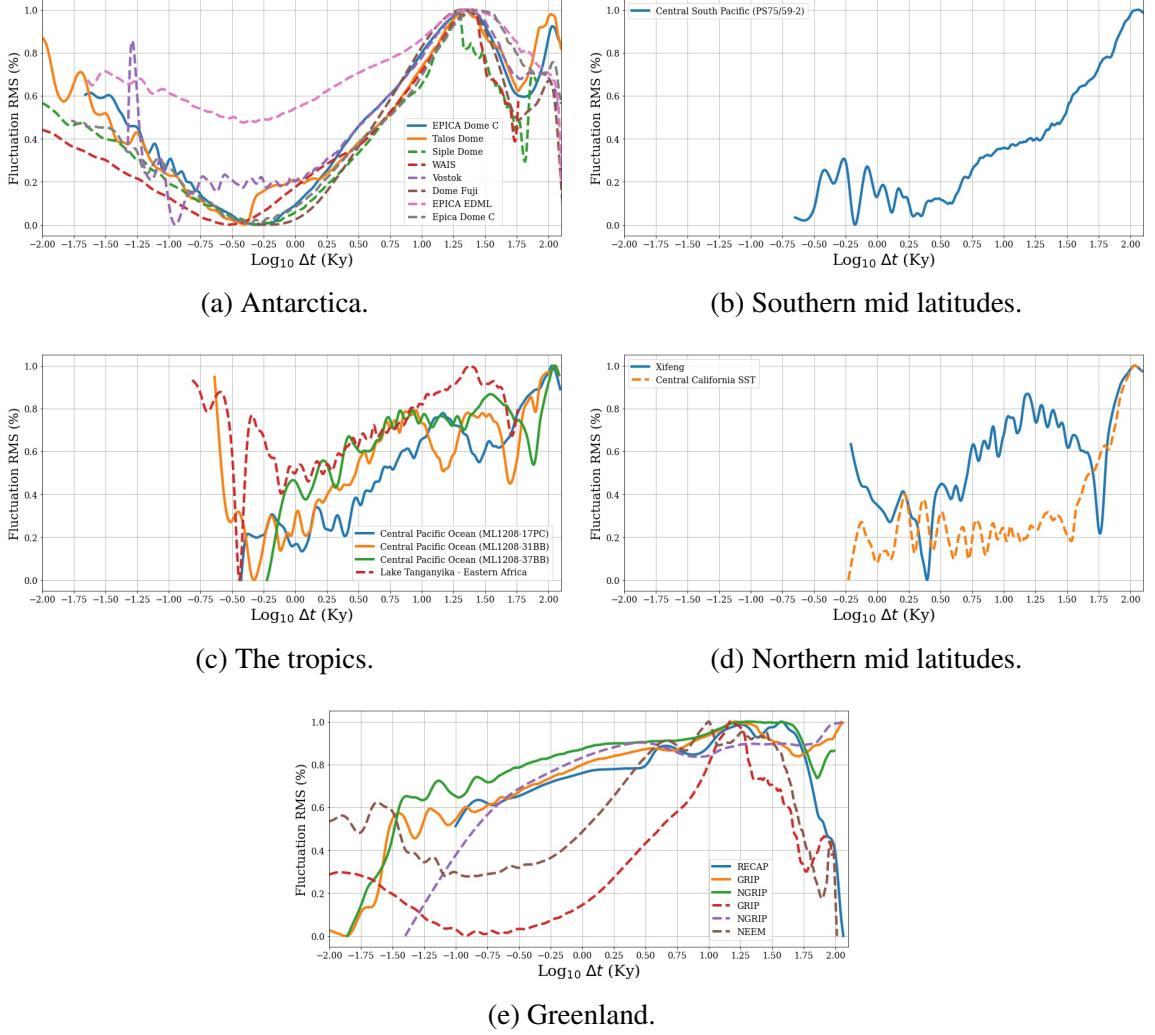


Figure 9.7. Normalised RMS fluctuations from dust and temperature clustered by latitude. Dashed and solid lines represent temperature and dust datasets, respectively.

Additionally, Xifeng does present a peak at $\log_{10} \Delta t = 1.25$, but despite being located at the same time lag, we can not presume it is related to the one from Antarctica since H_9 and H_2 are quite far from each other.

On the other hand, we found differences between land and ocean variability in the Tropics (Figures 9.7c and 9.8c), where Lake Tanganyika has a minimum at around $\log_{10} \Delta t = -0.24$ and a peak at $\log_{10} \Delta t = 1.4$, while Central Pacific fluctuations, even if increasing

with scale ($H_7 = 0.64$), do not show either any distinguishable minimum or maximum. Finally, Figures 9.7e and 9.8e presents some interesting results where all temperatures but NGRIP have a minimum at centennial scale and a maximum at millennial scale, thus similar to the results from Antarctica. But at the same time, dust fluctuations have a completely different behaviour, with a sustained variability from decennial to millennial scales ($H_{12} = 0.33$), which makes not possible to identify any climate regime transition. This fast increase in dust variability presented a slope higher than 1, which is out of the range of veracious values of the Haar function, thus, we did not added it to the figure.

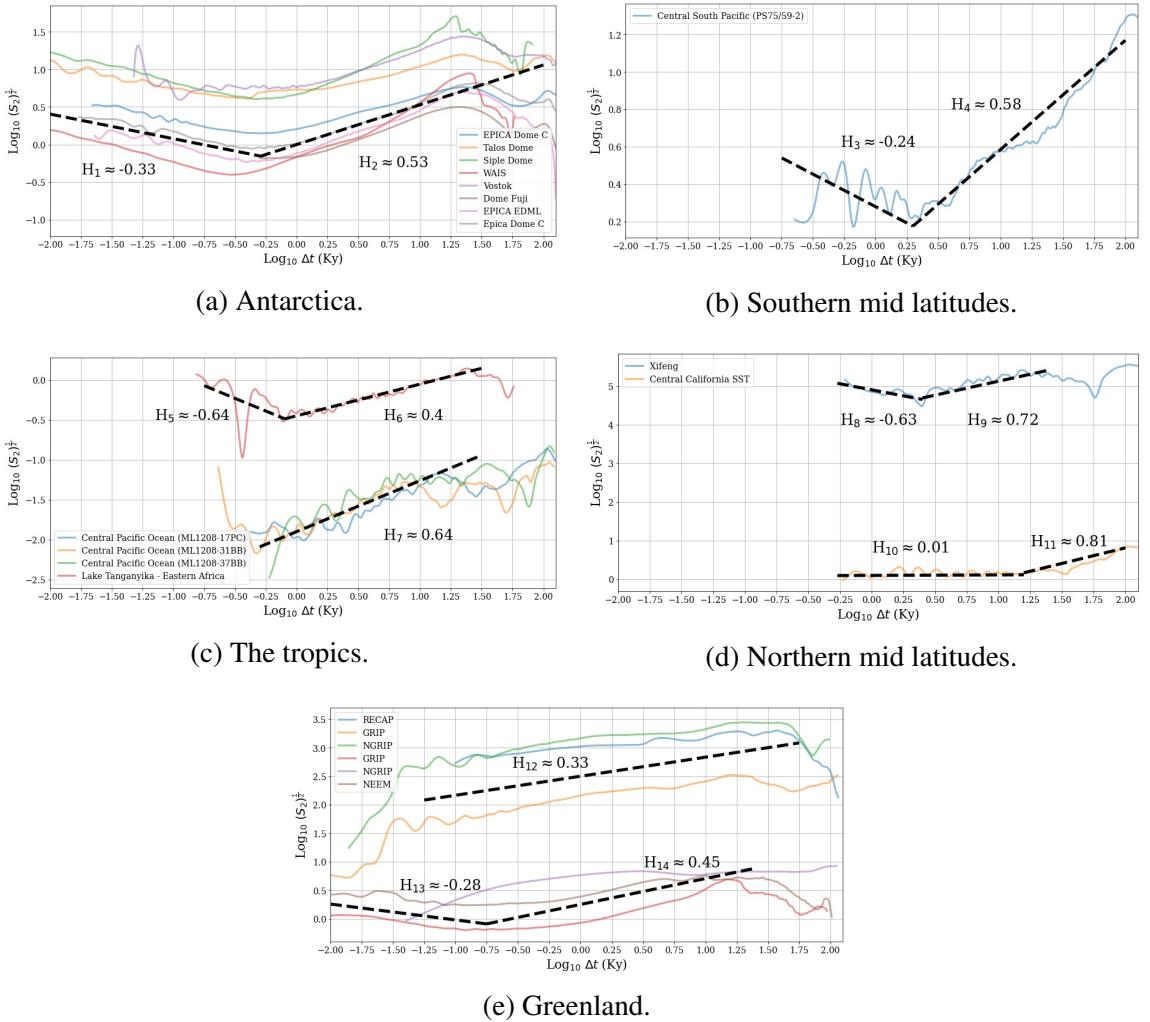


Figure 9.8. RMS fluctuations from dust and temperature clustered by latitude, where dashed lines have been added as slopes references.

We summarised all these results in Table 9.1. Here, *Slope 1* denotes the H exponent found at decennial/centennial time scales (e.g., H_1 , H_3 , or H_5 in Figure 9.8), and *Slope 2* denotes the H exponent found at centennial/millennial time scales (e.g., H_2 , H_4 , or H_6 in Figure 9.8).

Table 9.1. τ_c and maximum summary

Location	Slope 1	Slope 2	τ_c (y)	Maximum (y)
Greenland dust	<1	0.33	-	20.000
Greenland temp	-0.17	0.45	100-300	20.000
Northern mid latitudes (Loess)	-0.63	0.75	1.800-2.500	-
Northern mid latitudes temp	0	0.75	-	-
Tropic dust (Central Pacific)	>-1	0.67	-	30.000
Tropic temp. (East Africa Lake)	-0.64	0.4	630	20.000
Southern mid latitudes dust	-0.35	0.58	1.800-2.000	-
Southern mid latitudes temp	-	-	-	-
Antarctica dust	-0.3	0.5	200-600	20.000
Antarctica temp	-0.3	0.5	200-600	20.000

10. CONCLUSIONS

Until now, many methods have been developed with the aim of analyzing multiscale time series such as paleoclimatic datasets, where the cyclic nature of the Earth's process has led to a major interest in spectral techniques despite the difficulties of physically interpreting spectra. For this reason, in this work we developed and validated a Haar function algorithm to perform fluctuations analysis in the real domain. This algorithm also allows us to work with non-equidistant sets, which is the raw format of paleoclimatic dataset, hence avoiding linear interpolation and reducing the biases of the analysis.

Hence, we were able to assess the atmospheric variability in 24 locations across the world by working with temperature and dust deposition datasets of the Last Glacial Cycle, focusing on characterizing its variability at decennial to centennial timescale, which is of special interest since it interacts and overlaps with anthropogenic forcing. By clustering the locations by latitude, we studied the geographic atmosphere variability by identifying where the timescale symmetries break and the transition from macroweather to climate occurs, which shows changes with latitude but also between land and oceans.

Nevertheless, we identify some limitations in our analysis that can be improved in future studies. First, most of the datasets were from the poles (either Antarctica or Greenland), which made the analysis more local and becomes a drawback to globally characterizing atmosphere variability. Also, even if spectra interpretation is not trivial, it may be a helpful tool to complement the analysis, so we encourage future research to delve into the proper way of analyzing global atmospheric variability via spectral analysis. Finally, we did not consider uncertainty margins in our fluctuation estimations since there is no full theory available yet, which could be a great improvement in our algorithm and analysis itself.

REFERENCES

- Boor, C. d. (1978). *A practical guide to splines*. Springer New York. Retrieved from <https://link.springer.com/book/9780387953663>
- Brook, E. J., White, J. W., Schilla, A. S., Bender, M. L., Barnett, B., Severinghaus, J. P., ... Steig, E. J. (2005). Timing of millennial-scale climate change at siple dome, west antarctica, during the last glacial period. *Quaternary Science Reviews*, 24(12), 1333-1343. Retrieved from <https://www.sciencedirect.com/science/article/pii/S027737910500082X> doi: <https://doi.org/10.1016/j.quascirev.2005.02.002>
- Buizert, C., Adrian, B., Ahn, J., Albert, M., Alley, R. B., Baggenstos, T. K., Daniel Bauska, ... Woodruff, T. E. (2015). Precise interpolar phasing of abrupt climate change during the last ice age. *Nature*, 520, 661–665. Retrieved from <https://doi.org/10.1038/nature14401> doi: 10.1038/nature14401
- Cheng, H., Edwards, R. L., Sinha, A., Spötl, C., Yi, L., Chen, S., ... Zhang, H. (2016). The asian monsoon over the past 640,000 years and ice age terminations. *Nature*, 534, 640–646. Retrieved from <https://doi.org/10.1038/nature18591> doi: 10.1038/nature18591
- Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., ... McCabe, A. M. (2009). The last glacial maximum. *Science*, 325(5941), 710–714. doi: 10.1126/science.1172873
- De Angelis, M., Steffensen, J. P., Legrand, M., Clausen, H., & Hammer, C. (1997). Primary aerosol (sea salt and soil dust) deposited in greenland ice during the last climatic cycle: Comparison with east antarctic records. *Journal of Geophysical Research: Oceans*, 102(C12), 26681-26698. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97JC01298> doi: <https://doi.org/10.1029/97JC01298>
- Gkinis, V., Vinther, B. M., Popp, T. J., Faber, A.-K., Holme, C. T., Jensen, C. M., ...

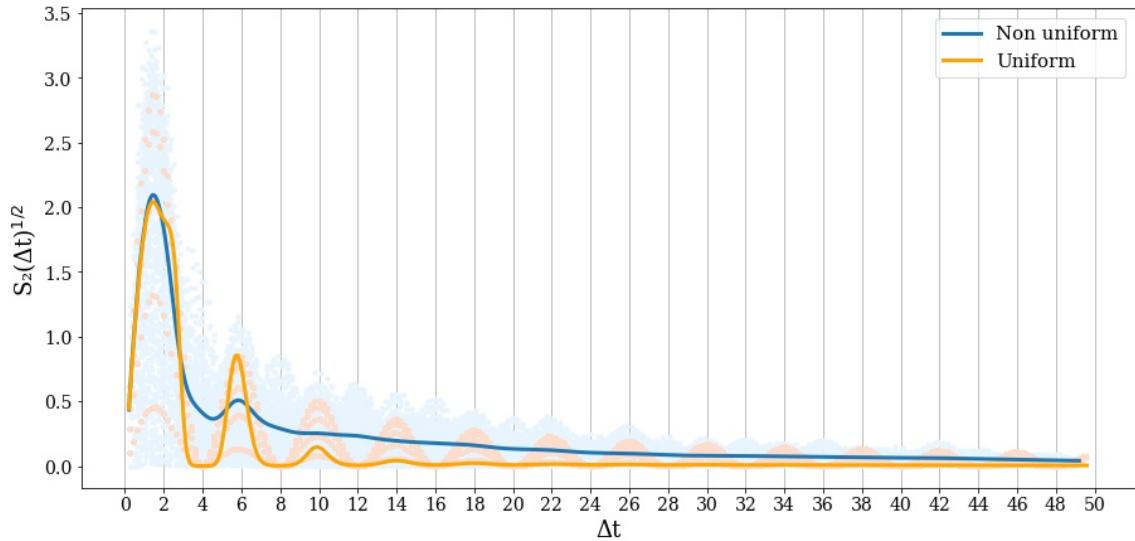
- White, J. (2021). *A 120,000-year long climate record from a nw-greenland deep ice core at ultra-high resolution* (Vol. 8) (No. 141). Retrieved from <https://doi.org/10.1038/s41597-021-00916-9> doi: 10.1038/s41597-021-00916-9
- Guo, Z. T., Berger, A., Yin, Q. Z., & Qin, L. (2009). Strong asymmetry of hemispheric climates during mis-13 inferred from correlating china loess and antarctica ice records. *Climate of the Past*, 5(1), 21–31. Retrieved from <https://cp.copernicus.org/articles/5/21/2009/> doi: 10.5194/cp-5-21-2009
- Jacobel, A., McManus, J., Anderson, R., & et al. (2016). Large deglacial shifts of the pacific intertropical convergence zone. *Nature Communications*, 7(10449). Retrieved from <https://doi.org/10.1038/ncomms10449> doi: 10.1038/ncomms10449
- Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., ... Wolff, E. W. (2007). Orbital and millennial antarctic climate variability over the past 800,000 years. *Science*, 317(5839), 793-796. Retrieved from <https://www.science.org/doi/abs/10.1126/science.1141038> doi: 10.1126/science.1141038
- Keyes, N. D. B., Giorgini, L. T., & Wettlaufer, J. S. (2022). *Stochastic paleoclimatology: Modeling the epica ice core climate records*. arXiv. Retrieved from <https://arxiv.org/abs/2210.00308> doi: 10.48550/ARXIV.2210.00308
- Kindler, P., Guillevic, M., Baumgartner, M., Schwander, J., Landais, A., & Leuenberger, M. (2014). Temperature reconstruction from 10 to 120 kyr b2k from the ngrip ice core. *Climate of the Past*, 10(2), 887–902. Retrieved from <https://cp.copernicus.org/articles/10/887/2014/> doi: 10.5194/cp-10-887-2014
- Lambert, F., Bigler, M., Steffensen, J. P., Hutterli, M., & Fischer, H. (2012). Centennial mineral dust variability in high-resolution ice core data from dome c, antarctica. *Climate of the Past*, 8(2), 609–623. Retrieved from <https://cp.copernicus.org/articles/8/609/2012/> doi: 10.5194/cp-8-609-2012
- Lamy, F., Gersonde, R., Winckler, G., Esper, O., Jaeschke, A., Kuhn, G.,

- ... Kilian, R. (2014). Increased dust deposition in the pacific southern ocean during glacial periods. *Science*, 343(6169), 403-407. Retrieved from <https://www.science.org/doi/abs/10.1126/science.1245424> doi: 10.1126/science.1245424
- Lovejoy, S. (2015, Jun). A voyage through scales, a missing quadrillion and why the climate is not what you expect. *Climate Dynamics*, 44(11), 3187-3210. Retrieved from <https://doi.org/10.1007/s00382-014-2324-0> doi: 10.1007/s00382-014-2324-0
- Lovejoy, S., & Lambert, F. (2019). Spiky fluctuations and scaling in high-resolution epica ice core dust fluxes. *Climate of the Past*, 15(6), 1999–2017. Retrieved from <https://cp.copernicus.org/articles/15/1999/2019/> doi: 10.5194/cp-15-1999-2019
- Lovejoy, S., & Schertzer, D. (2013). *The Weather and Climate: Emergent Laws and Multifractal Cascades*. Cambridge Press Univeristy. Retrieved from <https://hal-enpc.archives-ouvertes.fr/hal-00805839>
- Lovejoy, S., Varotsos, C., & Lambert, F. (2018). A.2 fluctuations as convolutions, filters and the h limits. In *Atmospheric scaling and climate variability across scales* (1st ed., Vol. 1, p. 98–108).
- Lyle, M., Heusser, L., Ravelo, C., Andreasen, D., Olivarez Lyle, A., & Diffenbaugh, N. (2010). Pleistocene water cycle and eastern boundary current processes along the california continental margin. *Paleoceanography*, 25(4). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009PA001836> doi: <https://doi.org/10.1029/2009PA001836>
- Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J.-M., Basile, I., ... Stievenard, M. (1999). Climate and atmospheric history of the past 420,000 years from the vostok ice core, antarctica. *Nature*, 399, 429–436. Retrieved from <https://doi.org/10.1038/20859> doi: 10.1038/20859

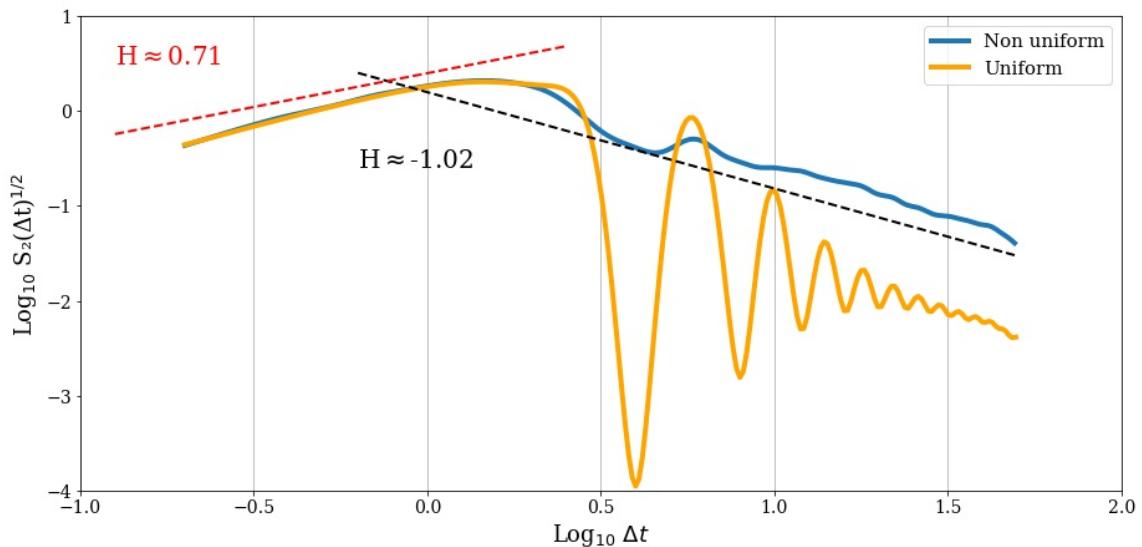
- Plonka, G., Potts, D., Steidl, G., & Tasche, M. (2018). Fast fourier transforms for nonequidistant data. *Numerical Fourier Analysis*, 377–419. doi: 10.1007/978-3-030-04306-3_7
- Ruth, U., Barnola, J.-M., Beer, J., Bigler, M., Blunier, T., Castellano, E., ... Wolff, E. (2007). “edml1”: a chronology for the epica deep ice core from dronning maud land, antarctica, over the last 150 000 years. *Climate of the Past*, 3(3), 475–484. Retrieved from <https://cp.copernicus.org/articles/3/475/2007/> doi: 10.5194/cp-3-475-2007
- Ruth, U., Bigler, M., Röhlisberger, R., Siggaard-Andersen, M.-L., Kipfstuhl, S., Goto-Azuma, K., ... Steffensen, J. P. (2007). Ice core evidence for a very tight link between north atlantic and east asian glacial climate. *Geophysical Research Letters*, 34(3). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL027876> doi: <https://doi.org/10.1029/2006GL027876>
- Sato, H., Suzuki, T., Hirabayashi, M., Iizuka, Y., Motoyama, H., & Fujii, Y. (2013). Mineral and sea-salt aerosol fluxes over the last 340 kyr reconstructed from the total concentration of al and na in the dome fuji ice core. *Atmospheric and Climate Sciences*, 3(2), 186-192. Retrieved from <https://www.scirp.org/journal/paperinformation.aspx?paperid=30768> doi: 10.4236/acs.2013.32020.
- Schüpbach, S., Federer, U., Kaufmann, P. R., Albani, S., Barbante, C., Stocker, T. F., & Fischer, H. (2013). High-resolution mineral dust and sea ice proxy records from the talos dome ice core. *Climate of the Past*, 9(6), 2789–2807. Retrieved from <https://cp.copernicus.org/articles/9/2789/2013/> doi: 10.5194/cp-9-2789-2013
- Seierstad, I. K., Abbott, P. M., Bigler, M., Blunier, T., Bourne, A. J., Brook, E., ... Vinther, B. M. (2014). Consistently dated records from the greenland grip, gisp2 and ngrip ice cores for the past 104 ka reveal regional millennial-scale $\delta^{18}\text{O}$ gradients with possible heinrich event

- imprint. *Quaternary Science Reviews*, 106, 29-46. Retrieved from <https://www.sciencedirect.com/science/article/pii/S027737911400434X> doi: <https://doi.org/10.1016/j.quascirev.2014.10.032>
- Simonsen, M. F., Baccolo, G., Blunier, T., Borunda, A., Delmonte, B., Frei, R., ... Valelonga, P. (2019). East greenland ice core dust record reveals timing of greenland ice sheet advance and retreat. *Nature Communications*, 10(4494). Retrieved from <https://doi.org/10.1038/s41467-019-12546-2> doi: 10.1038/s41467-019-12546-2
- Thompson, L. G., Davis, M. E., Mosley-Thompson, E., Sowers, T. A., Henderson, K. A., Zagorodnov, V. S., ... Francou, B. (1998). A 25,000-year tropical climate history from bolivian ice cores. *Science*, 282(5395), 1858-1864. Retrieved from <https://www.science.org/doi/abs/10.1126/science.282.5395.1858> doi: 10.1126/science.282.5395.1858
- Tierney, J. E., Russell, J. M., Huang, Y., Damsté, J. S. S., Hopmans, E. C., & Cohen, A. S. (2008). Northern hemisphere controls on tropical southeast african climate during the past 60,000 years. *Science*, 322(5899), 252-255. Retrieved from <https://www.science.org/doi/abs/10.1126/science.1160485> doi: 10.1126/science.1160485

Appendix A: Detailed figures of fluctuations RMS from sinusoidal function.

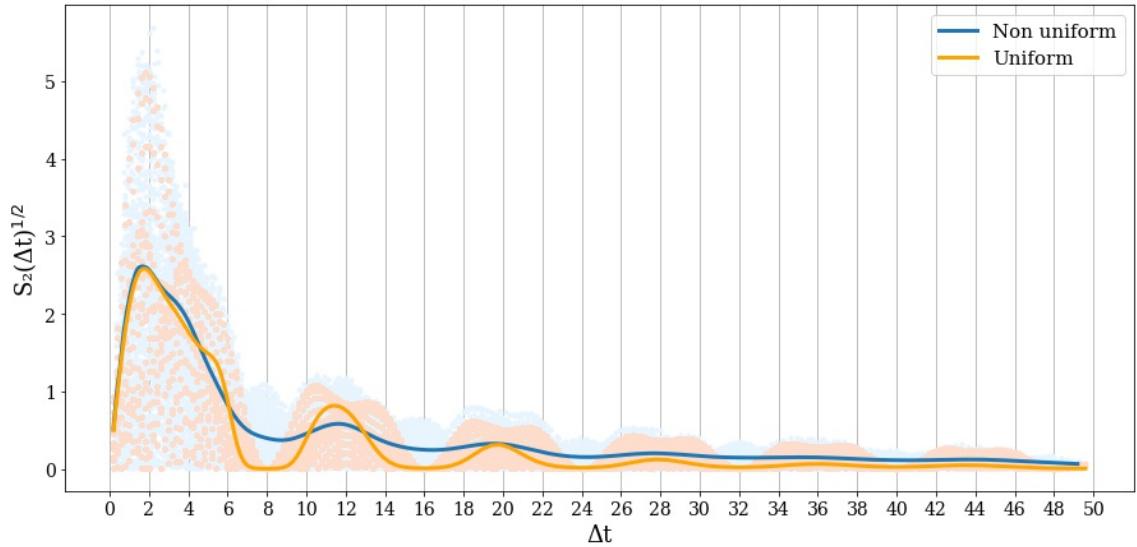


(a) Sinusoidal function 1 RMS as scatter plot.

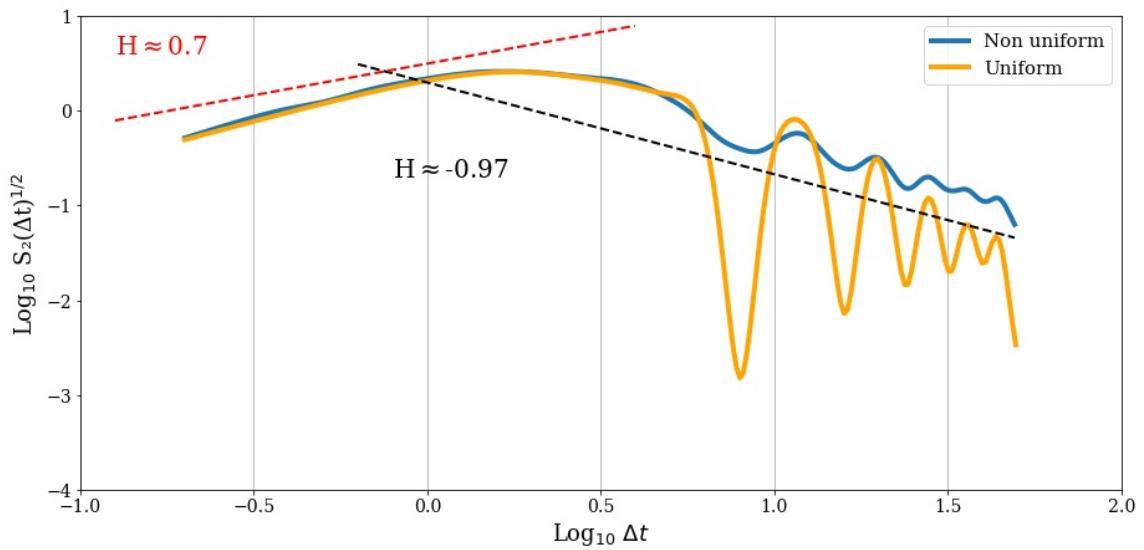


(b) Sinusoidal function 1 RMS in log-log scale.

Figure 10.1. Detailed figure of fluctuations RMS from sinusoidal function 1.

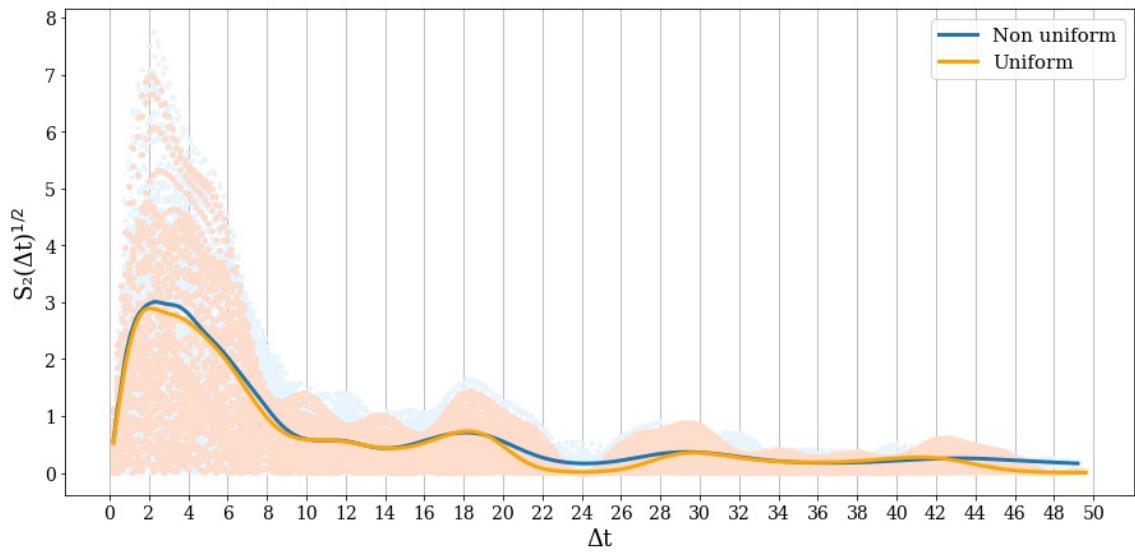


(a) Sinusoidal function 2 RMS as scatter plot.

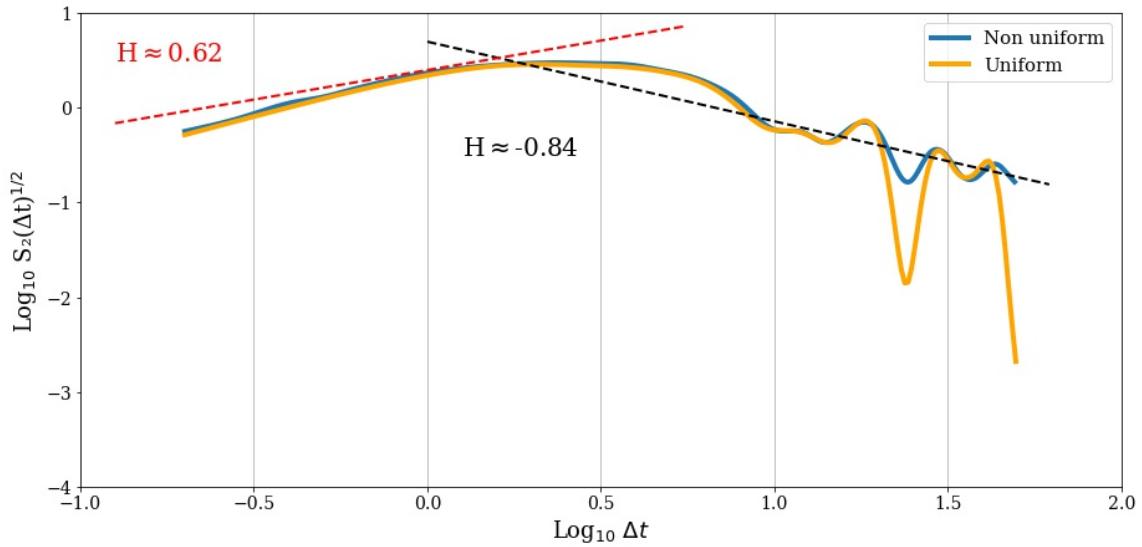


(b) Sinusoidal function 2 RMS in log-log scale.

Figure 10.2. Detailed figure of fluctuations RMS from sinusoidal function 2.

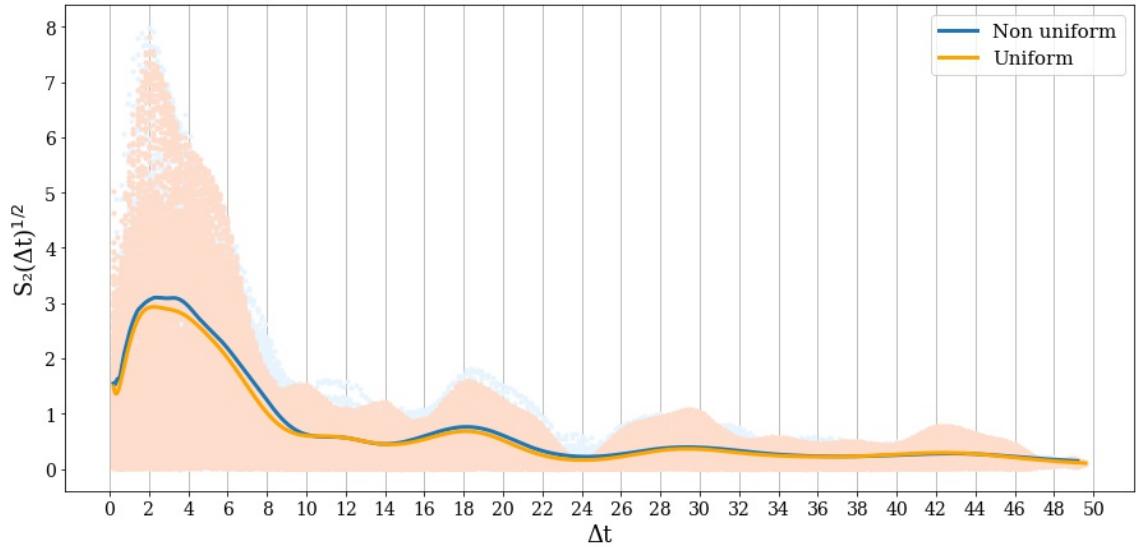


(a) Sinusoidal function 3 RMS as scatter plot.

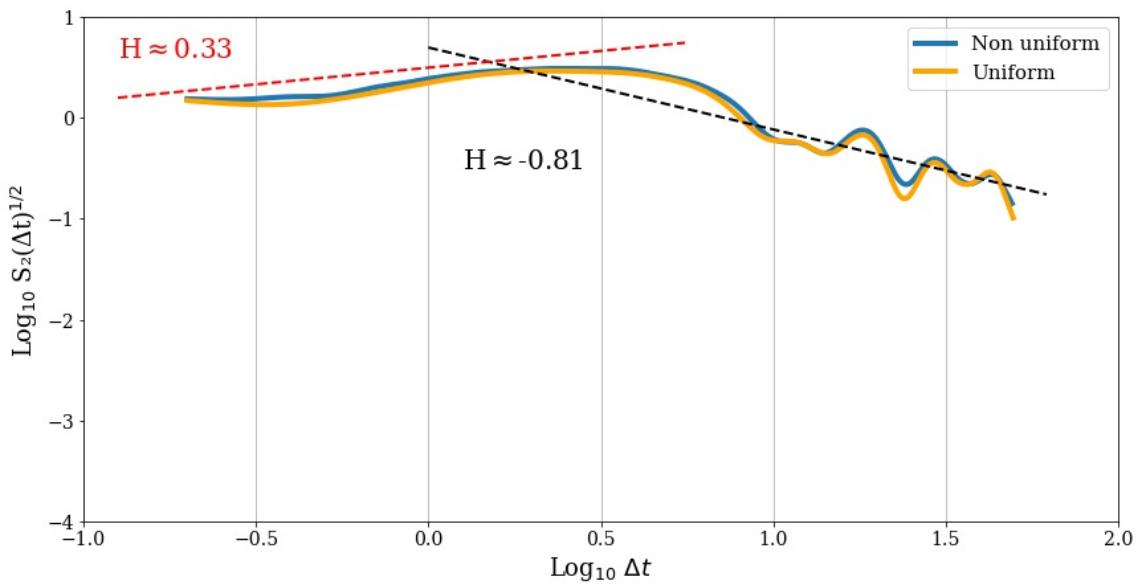


(b) Sinusoidal function 3 RMS in log-log scale.

Figure 10.3. Detailed figure of fluctuations RMS from sinusoidal function 3.

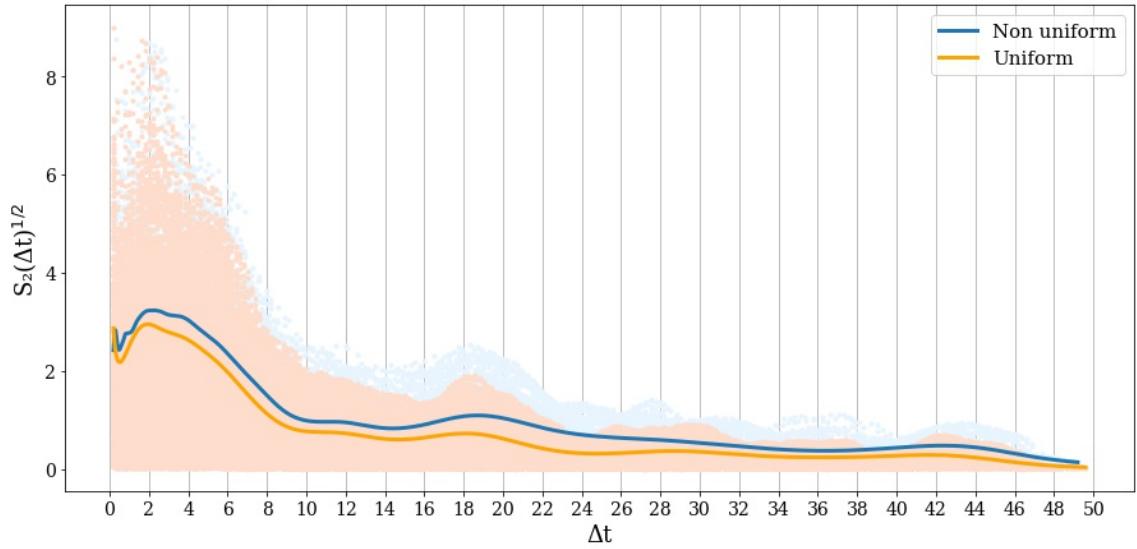


(a) Sinusoidal function 4 RMS as scatter plot.

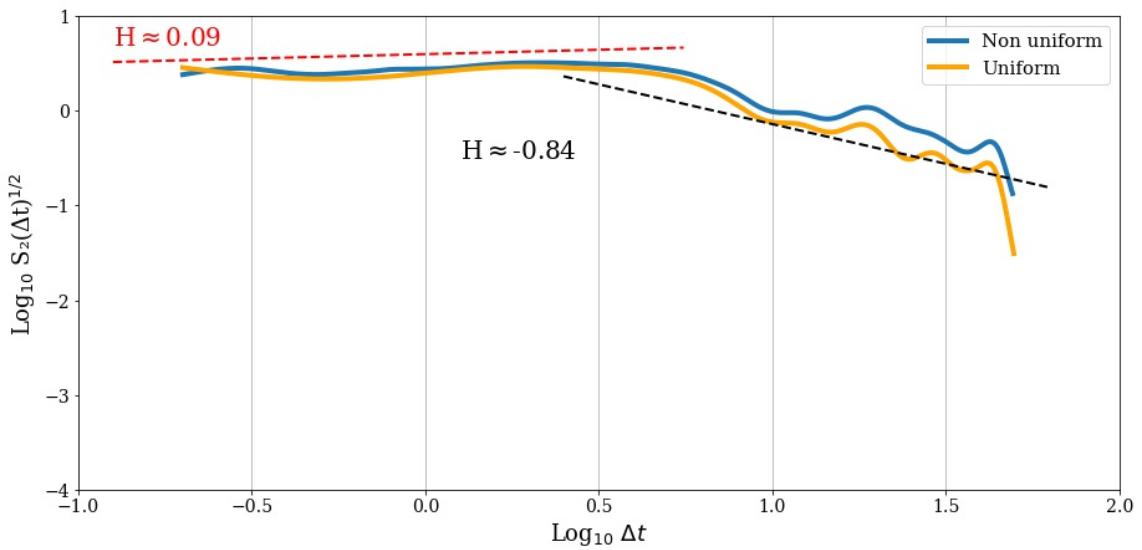


(b) Sinusoidal function 4 RMS in log-log scale.

Figure 10.4. Detailed figure of fluctuations RMS from sinusoidal function 4.



(a) Sinusoidal function 5 RMS as scatter plot.



(b) Sinusoidal function 5 RMS in log-log scale.

Figure 10.5. Detailed figure of fluctuations RMS from sinusoidal function 5.