# Algorithms for Spectral Analysis of Irregularly Sampled Time Series

**Adolf Mathias**
ZKM - Center for Art and Media

**Florian Grond**
ZKM - Center for Art and Media

**Ramon Guardans**
Medialab Madrid

**Detlef Seese**
University of Karlsruhe (TH)

**Miguel Canela**
University of Barcelona

**Hans H. Diebner**
ZKM - Center for Art and Media

## Abstract

In this paper, we present a spectral analysis method based upon least square approximation. Our method deals with nonuniform sampling. It provides meaningful phase information that varies in a predictable way as the samples are shifted in time. We compare least square approximations of real and complex series, analyze their properties for sample count towards infinity as well as estimator behaviour, and show the equivalence to the discrete Fourier transform applied onto uniformly sampled data as a special case. We propose a way to deal with the undesirable side effects of nonuniform sampling in the presence of constant offsets. By using weighted least square approximation, we introduce an analogue to the Morlet wavelet transform for nonuniformly sampled data. Asymptotically fast divide-and-conquer schemes for the computation of the variants of the proposed method are presented. The usefulness is demonstrated in some relevant applications.

*Keywords*: spectral analysis, irregularly sampled time series, wavelets, fast algorithms, economic time series, geologic time series.

## 1. Introduction

Despite the vast number of methods for time series analysis, there is still a lack in manageable methods dealing with nonuniformly sampled data. In this paper, we present a spectral analysis method based upon least square approximation, applied to different types of data. Our method is based upon assumptions similar to those of the well-known Lomb-Scargle

periodogram (Lomb 1976; Scargle 1976; Press, Teukolsky, Vetterling, and Flannery 1992; Bretthorst 2000), and deals with nonuniform sampling. Beyond that, it has the advantage of a meaningful phase that varies in a predictable way as the samples are shifted in time, while the Lomb-Scargle periodogram was designed to compute the power spectrum. We study estimator properties and the convergence of our results for the sample counts going towards infinity. We compare least square approximations of real and complex series and show the equivalence to the discrete Fourier transform applied onto uniformly sampled data as a special case. We analyze how constant offsets affect the obtained spectral coefficients and propose a way to deal with the undesirable side effects of nonuniform sampling.

We then reanalyze the approach with focus on weighted least square approximation and introduce an analogue to the Morlet wavelet transform (Chui 1992; Daubechies 1987; Holschneider 1995) for nonuniformly sampled data. Next, we construct a family of asymptotically fast divide-and-conquer schemes for the evaluation of the presented analysis methods. We demonstrate the functioning of our methods by applying them onto the paleoclimatic $CO_2$ and Deuterium concentrations covering a time span of approx. 420000 years that were obtained from the ice core (Petit JR *et al.* 1999) extracted from the ice cover of Lake Vostok, Antarctica. The occupation with the nonuniformly sampled data presented here were the main reason for the work that lead to this publication. The paper and the data are available from `http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v399/n6735/abs/399429a0_fs.html`. Finally, we apply the methods to tick data obtained on one day at the London stock exchange. Since the trade occurs at irregular times, one also has to deal with nonuniformly sampled data.

## 2. Spectral transforms for unevenly sampled time series

Given a time series $\mathbf{x} = (x_0 \ x_1 \ \ldots \ x_{n-1})$ sampled at nonuniform time instants $\mathbf{t} = (t_0 \ t_1 \ \ldots \ t_{n-1})$, the approach

$$x_k = (c_\omega \ s_\omega) \begin{pmatrix} \cos(\omega t_k) \\ \sin(\omega t_k) \end{pmatrix} + e_k, \ \forall k \in \{0, \ldots, n-1\} \tag{1}$$

can be used to derive least square trigonometric approximations for a given angular frequency $\omega$. Here, $e_k \in \mathbb{R}$ is a zero-mean error term. In the following, $\sum$ always means $\sum_{k \in K}$ with the index set $K = \{0, \ldots, n-1\}$, and $\Omega_{\mathbf{t},\omega}$ denotes the matrix

$$\begin{pmatrix} \cos \omega t_0 & \cos \omega t_1 & \cdots & \cos \omega t_{n-1} \\ \sin \omega t_0 & \sin \omega t_1 & \cdots & \sin \omega t_{n-1} \end{pmatrix}.$$

The coefficients $c_\omega$ and $s_\omega$ can be obtained by multiplying Eq. 1 with the Moore-Penrose pseudo-inverse (Penrose 1955),

$$(c_\omega \ s_\omega) \ = \ \mathbf{x} \, \Omega_{\mathbf{t},\omega}^{\mathrm{T}} \left( \Omega_{\mathbf{t},\omega} \Omega_{\mathbf{t},\omega}^{\mathrm{T}} \right)^{-1} .$$

Using trigonometric identities, we compute

$$(c_\omega \; s_\omega) = \sum \left( x_k \left( \begin{array}{c} \cos \omega t_k \\ \sin \omega t_k \end{array} \right)^{\mathrm{T}} \right) \left( \begin{array}{cc} \sum \cos^2 \omega t_k & \sum \cos \omega t_k \sin \omega t_k \\ \sum \cos \omega t_k \sin \omega t_k & \sum \sin^2 \omega t_k \end{array} \right)^{-1}$$

$$= 2 \sum \left( x_k \left( \begin{array}{c} \cos \omega t_k \\ \sin \omega t_k \end{array} \right)^{\mathrm{T}} \right) \frac{\left( \begin{array}{cc} n - \sum \cos 2\omega t_k & -\sum \sin 2\omega t_k \\ -\sum \sin 2\omega t_k & n + \sum \cos 2\omega t_k \end{array} \right)}{n^2 - (\sum \cos 2\omega t_k)^2 - (\sum \sin 2\omega t_k)^2} \qquad (2)$$

and, based on that, we introduce

$$P_\omega = c_\omega^2 + s_\omega^2, \qquad (3)$$

a value that will be used later in considerations about power spectra.

Lomb (1976) made similar assumptions about the nature of the analyzed data and introduced a time shift $\theta$ of the series that makes the vectors $\{\cos(\omega(t_k - \theta))\}_{k \in K}$ and $\{\sin(\omega(t_k - \theta))\}_{k \in K}$ orthogonal, dependent on $\omega$:

$$\sum \cos(\omega(t_k - \theta)) \sin(\omega(t_k - \theta)) = \frac{1}{2} \sum \sin 2(\omega(t_k - \theta))$$

$$= \frac{1}{2} \sum \sin 2\omega t_k \cos 2\omega\theta - \cos 2\omega t_k \sin 2\omega\theta$$

$$= 0$$

$$\sum \sin 2\omega t_k \cos 2\omega\theta = \sum \cos 2\omega t_k \sin 2\omega\theta$$

$$\theta = \frac{1}{2\omega} \arctan \frac{\sum \sin 2\omega t_k}{\sum \cos 2\omega t_k}$$

$$= \frac{1}{2\omega} \arg \left( \sum e^{2i\omega t_k} \right). \qquad (4)$$

The representation using arctan is the common way to express $\theta$. Note that we silently assume it to be equivalent to the representation using arg, i.e. taking into account the signs of numerator and denominator separately and dealing with 0 in the denominator.

Applying Eq. 2 onto the original time series with changed sample times $t_k - \theta$, $k \in K$, yields

$$(c_\omega \; s_\omega) = \sum \left( x_k \left( \begin{array}{c} \cos(\omega(t_k - \theta)) \\ \sin(\omega(t_k - \theta)) \end{array} \right)^{\mathrm{T}} \right) \left( \begin{array}{cc} \frac{1}{\sum \cos^2(\omega(t_k - \theta))} & 0 \\ 0 & \frac{1}{\sum \sin^2(\omega(t_k - \theta))} \end{array} \right)$$

$$= \left( \frac{\sum x_k \cos(\omega(t_k - \theta))}{\sum \cos^2(\omega(t_k - \theta))} \; \frac{\sum x_k \sin(\omega(t_k - \theta))}{\sum \sin^2(\omega(t_k - \theta))} \right).$$

The absolute square value is

$$|(c_\omega \; s_\omega)|^2 = \frac{(\sum x_k \cos(\omega(t_k - \theta)))^2}{(\sum \cos^2(\omega(t_k - \theta)))^2} + \frac{(\sum x_k \sin(\omega(t_k - \theta)))^2}{(\sum \sin^2(\omega(t_k - \theta)))^2},$$

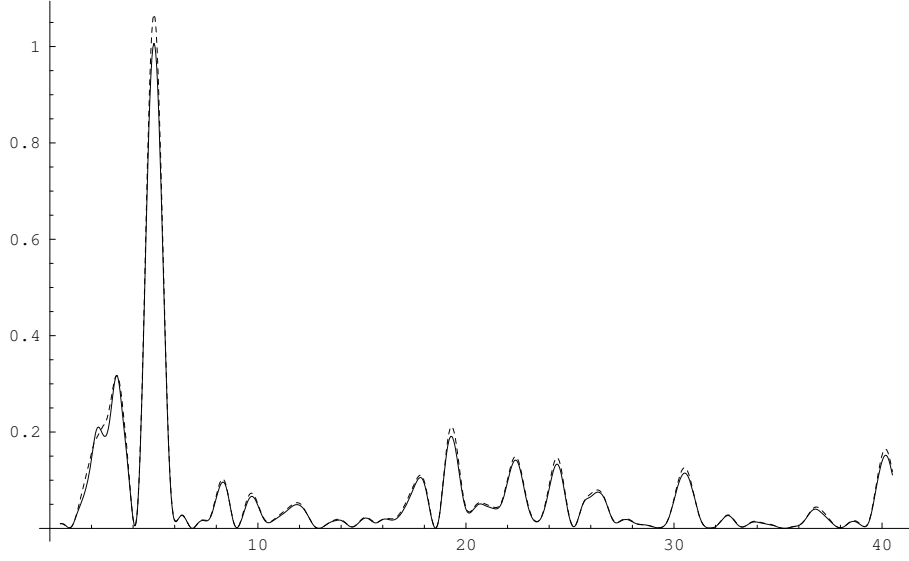which, apart from the squared denominators, is equivalent to the Lomb periodogram,

Figure 1: The solid line shows the absolute square result of Eq. 3 applied to $\sin(4.8t) + X_{0,2}$ , sampled at 400 values of $t$ that are uniformly randomly distributed over $[0, 2\pi)$. $X_{0,2}$ denotes a normally distributed random variable with mean 0 and variance 2. The dashed line shows the Lomb periodogram applied to the same data, scaled with $\frac{1}{182}$. This factor has to be adapted for different sample counts.

$$P_{\text{lomb}}(\omega) = \frac{\left(\sum x_k \cos\left(\omega\left(t_k - \theta\right)\right)\right)^2}{\sum \cos^2\left(\omega\left(t_k - \theta\right)\right)} + \frac{\left(\sum x_k \sin\left(\omega\left(t_k - \theta\right)\right)\right)^2}{\sum \sin^2\left(\omega\left(t_k - \theta\right)\right)} . \tag{5}$$

Using the mean $\langle x \rangle$ and the variance $\sigma^2$ of $\{x_k\}$, the Lomb normalized periodogram is defined as

$$P_{\text{lombnorm}}(\omega) = \frac{1}{2\sigma^2}\left(\frac{\left(\sum\left(x_k - \langle x\rangle\right)\cos\left(\omega\left(t_k - \theta\right)\right)\right)^2}{\sum \cos^2\left(\omega\left(t_k - \theta\right)\right)} + \frac{\left(\sum\left(x_k - \langle x\rangle\right)\sin\left(\omega\left(t_k - \theta\right)\right)\right)^2}{\sum \sin^2\left(\omega\left(t_k - \theta\right)\right)}\right) .$$

The replacement of the $x_k$ with $x_k - \langle x\rangle$ is one way to deal with time series with constant offset, for which we propose an alternative solution below. This replacement is also commonly used for the unnormalized Lomb periodogram.

The amplitudes of both versions of the Lomb periodogram appear to depend on the sampling density, as can be seen in in Figure 2.

In the case of a complex time series $\xi = \{\xi_k\}_{k \in K}$ and an approximation by complex functions $e^{i\omega t}$, the result is even simpler. We use the matrix $\Psi_{\mathbf{t},\omega}$,

$$\Psi_{\mathbf{t},\omega} = \begin{pmatrix} e^{i\omega t_0} & e^{i\omega t_1} & \ldots & e^{i\omega t_{n-1}} \end{pmatrix}$$

to formulate the expression

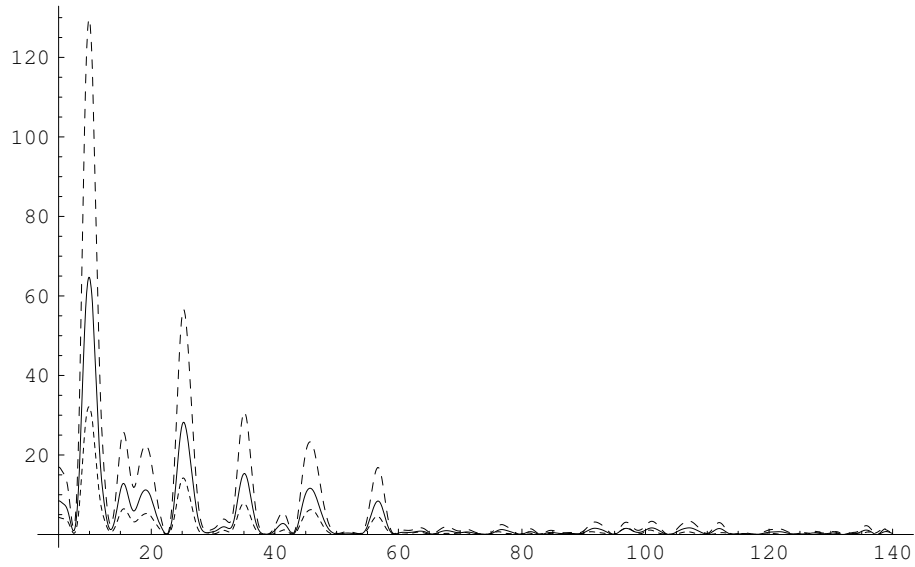$$\xi = \zeta_\omega \Psi_{\mathbf{t},\omega} + \epsilon \tag{6}$$

Figure 2: The solid line shows the Lomb normalized periodogram applied to the $CO_2$ series from the Vostok data with subtracted mean value. The densely dashed curve shows the spectrum of the same data set with odd-numbered samples removed, the widely dashed curves were obtained by doubling each sample. The abscissae show frequencies in 1/Myr.
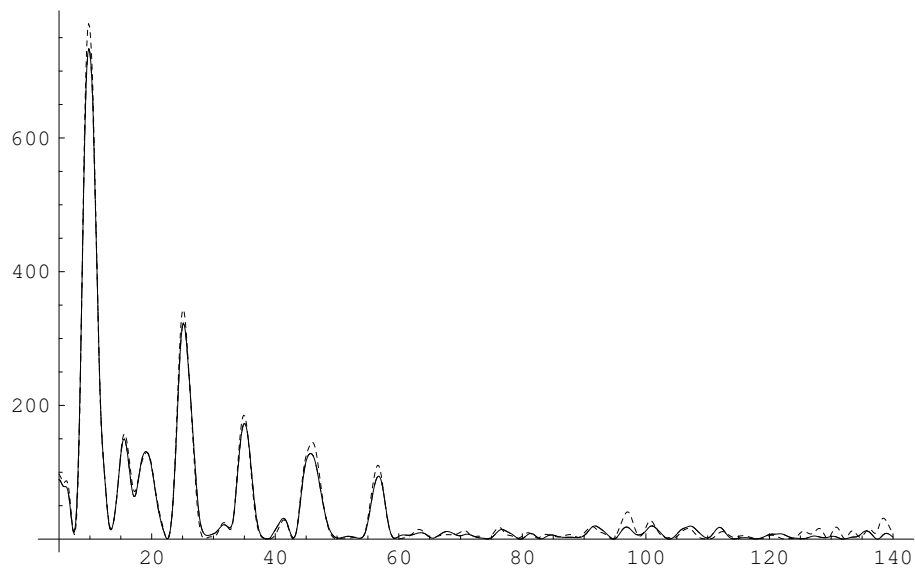


Figure 3: The solid line shows the result of Eq. 3 applied to the $CO_2$ series from the Vostok data with subtracted mean value. As in Figure 2, the dashed curves were obtained from of the same data set with odd-numbered samples removed and with each sample doubled, respectively. The abscissae show frequencies in 1/Myr.
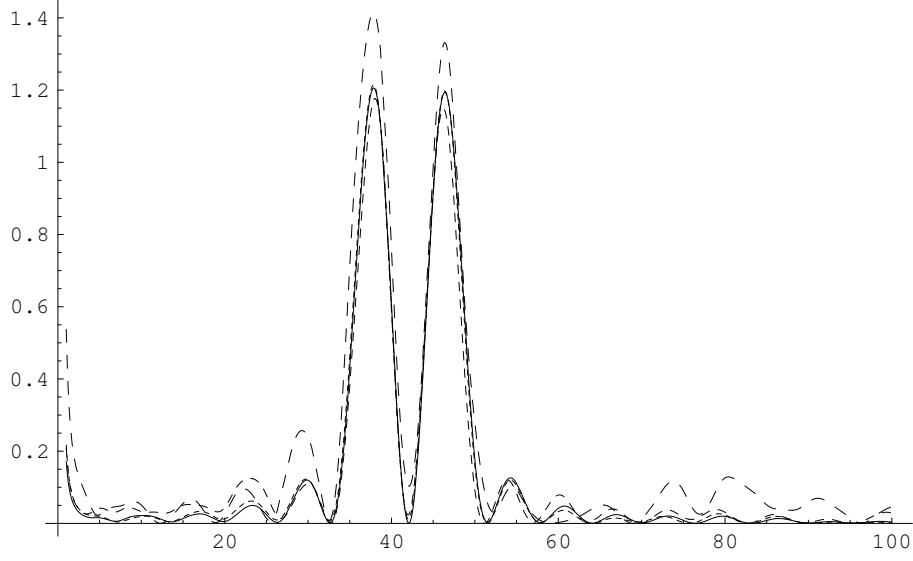
Figure 4: Eq. 3 applied to $\sin 39t + \sin(39 + 2\pi)t$ randomly sampled over $[0, 1]$ with increasing density (100, 400, 2000 and 10000 points, curve plot density increasing with sample density).

analogously to Eq. 1 with a complex coefficient $\zeta_\omega$ and a vector $\epsilon$ of complex zero average error terms.

The pseudo-inverse solution yields in this case

$$\zeta_\omega = \xi \Psi_{\mathbf{t},\omega}^* \left( \Psi_{\mathbf{t},\omega} \Psi_{\mathbf{t},\omega}^* \right)^{-1} = \frac{1}{n} \sum \xi_k e^{-i\omega t_k} \,. \tag{7}$$

Using $\Re z = \frac{1}{2}(z + \overline{z})$ and $\Im z = \frac{1}{2i}(z - \overline{z})$, we rewrite the real approximation of Eq. 2 in terms of the complex approximation. Here, $\zeta_\omega$ denotes the complex approximation of the real series $\{x_k\}_{k \in K}$, $\iota_{2\omega}$ denotes the complex approximation at frequency $2\omega$ of the unit time series $\{1\}_{k \in K}$ over $\{t_k\}_{k \in K}$, and $\Re c$ and $\Im c$ mean real and imaginary parts of a complex number $c$, respectively. We then obtain

$$
\begin{aligned}
c_\omega + i s_\omega &= \frac{2}{1 - |\iota_{2\omega}|^2} \begin{pmatrix} \Re\zeta_\omega & -\Im\zeta_\omega \end{pmatrix} \begin{pmatrix} 1 - \Re\iota_{2\omega} & \Im\iota_{2\omega} \\ \Im\iota_{2\omega} & 1 + \Re\iota_{2\omega} \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} \\
&= \frac{2\left( \zeta_\omega - \overline{\iota_{2\omega}}\zeta_\omega \right)}{1 - |\iota_{2\omega}|^2} \,.
\end{aligned}
\tag{8}
$$

In the uniform case of $t_k = k\Delta t$, $\Delta t \neq 0$, $k \in K = \{0, \ldots, n-1\}$ and $\omega = \frac{l\pi}{n\Delta t}$, $l \in K \setminus \{0\}$, the sums $\sum \cos 2\omega t_k$ and $\sum \sin 2\omega t_k$ vanish in the real approximation given by Eq. (2), and the result is equivalent to the complex conjugate of one spectral coefficient of the discrete Fourier transform. For $\omega = 0$, the average $c_\omega = \langle \mathbf{x} \rangle$ is obtained. Each term of the sum in Eq. 6 can be written as a matrix multiplication,

$$
\begin{pmatrix} c_\omega & s_\omega \end{pmatrix} \begin{pmatrix} \cos \omega t_k & \sin \omega t_k \\ -\sin \omega t_k & \cos \omega t_k \end{pmatrix} = \begin{pmatrix} \Re x_k & \Im x_k \end{pmatrix} \,. \tag{9}
$$

Approximating such a uniform real-valued series with the complex approximation of Eq. 7 leaves the imaginary part undetermined, whereas the strictly real approximation of Eq. 2 can be interpreted as a complex equation with imaginary part set to zero, $\Im x_k = 0$. Eq. 2 can be used to approximate both time series $\Re x_k$ and $\Im x_k$ of Eq. 9, whereby $\Im x_k = 0$, as mentioned before, which obviously leads to $c_\omega = s_\omega = 0$. The complex approach of Eq. 9 chooses the average between the results from Eq. 2 and $c_\omega = s_\omega = 0$, which is $\frac{1}{2}(c_\omega \; s_\omega)$. This explains the factor 2 in Eq. 2.

In the general case of nonuniformly sampled time series, one usually estimates power spectra by means of periodograms, as for example the frequently used Lomb-Scargle periodogram (Lomb 1976; Scargle 1976; Press *et al.* 1992). When however phase information is desired, e.g. for the reconstruction of time series out of spectral coefficients and for comparison of signals, Eq. (2) does the job. In the sequel, we also use it for the extension of wavelet analysis onto nonuniform time series.

The amplitude of our result is invariant with respect to time shift, whereas the phase is covariant, as can be verified easily. By introducing a time shift $\Delta t$, we have

$$
\begin{aligned}
\Omega_{\mathbf{t}+\Delta t,\omega} &= \begin{pmatrix} \cos \omega \left( t_0 + \Delta t \right) & \cos \omega \left( t_1 + \Delta t \right) & \cdots & \cos \omega \left( t_{n-1} + \Delta t \right) \\ \sin \omega \left( t_0 + \Delta t \right) & \sin \omega \left( t_1 + \Delta t \right) & \cdots & \sin \omega \left( t_{n-1} + \Delta t \right) \end{pmatrix} \\
&= \begin{pmatrix} \cos \omega \Delta t & -\sin \omega \Delta t \\ \sin \omega \Delta t & \cos \omega \Delta t \end{pmatrix} \Omega_{\mathbf{t},\omega} \\
&= \mathbf{R}_{\omega \Delta t} \Omega_{\mathbf{t},\omega} \, .
\end{aligned}
\tag{10}
$$

The pseudo-inverse of $\Omega_{\mathbf{t}+\Delta t,\omega}$ becomes then

$$
\left( \mathbf{R}_{\omega \Delta t} \Omega_{\mathbf{t},\omega} \right)^* \left( \mathbf{R}_{\omega \Delta t} \Omega_{\mathbf{t},\omega} (\mathbf{R}_{\omega \Delta t} \Omega_{\mathbf{t},\omega})^* \right)^{-1} = \Omega_{\mathbf{t},\omega}^* \left( \Omega_{\mathbf{t},\omega} \Omega_{\mathbf{t},\omega}^* \right)^{-1} \mathbf{R}_{\omega \Delta t}^{\mathrm{T}}
$$

and the coefficients after time shift

$$
\left( c_\omega \; s_\omega \right)' = \left( c_\omega \; s_\omega \right) \; \mathbf{R}_{\omega \Delta t}^{\mathrm{T}}.
$$

Obviously, for the complex approximation (Eq. 7) the time shift yields

$$
\zeta_\omega' = \zeta_\omega e^{i\omega \Delta t} \, .
$$

Both the real and the complex approximation's amplitudes are thus shift invariant, and a meaningful phase is obtained. The well-known advantage of periodograms applied to nonuniformly sampled time series is also given here, namely their insensitivity to aliasing phenomena when the differences between the $t_i$ obey a normal distribution.

We note in passing that the reconstruction of a signal from the spectral analysis of a time series requires a full matrix inverse as a consequence of the nonuniform sampling.

# 3. Convergence

We base our following considerations on a fixed real infinite sequence $\{t_k\}_{k \in \mathbb{N}_0}$ with

$$\varphi(t) = \lim_{n \to \infty} \lim_{\epsilon \to +0} \frac{1}{n\epsilon} \left| \left\{ t_k \, \middle| \, t - \frac{\epsilon}{2} \le t_k < t + \frac{\epsilon}{2} \right\} \right|$$

a generalized function describing the empirical density that the $t_k$ obey, and a real function $x(t)$ where $x(t)\varphi(t)$ is absolutely integrable. For sequences with noncontinuous empirical density, e.g. $t_k = k \bmod m$ for a fixed $m \in \mathbb{N}$, $\varphi(t)$ may become infinite for some values of $t$, a case which is handled by the theory of generalized functions or distributions. In most cases it suffices however to consider the limits of ordinary functions, e.g. $\varphi(t) = \frac{1}{m} \lim_{\sigma \to 0} \frac{1}{\sqrt{\pi}\sigma} \sum_{h=0}^{m-1} e^{-\left(\frac{t-h}{\sigma}\right)^2}$ for the previously mentioned series.

Under these circumstances, we can write

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} x(t_k) = \int_{-\infty}^{\infty} \varphi(t) x(t) dt. \tag{11}$$

We next consider a complex function $\xi(t)$, where $\xi(t)\varphi(t)$ is absolutely integrable, and a measurement error $\rho_k$ obeying an independent complex symmetric distribution with zero mean. The values $\xi(t_k)$ are identified with the $\xi_k$. The complex approximation of Eq. 7 applied to $\xi(t_k) + \rho_k$ becomes then

$$
\begin{aligned}
\lim_{n \to \infty} \zeta_\omega &= \lim_{n \to \infty} \frac{1}{n} \sum (\xi(t_k) + \rho_k) e^{-i\omega t_k} \\
&= \lim_{n \to \infty} \frac{1}{n} \sum \xi(t_k) e^{-i\omega t_k} + \lim_{n \to \infty} \frac{1}{n} \sum \rho_k e^{-i\omega t_k} \\
&= \int_{-\infty}^{\infty} \xi(t)\varphi(t) e^{-i\omega t} dt \\
&= \mathcal{F}_t \left( \xi(t)\varphi(t) \right),
\end{aligned}
$$

where $\left( \mathcal{F}_t f(t) \right)(\omega)$ denotes the Fourier transform of $f(t)$ with respect to the variable $t$ at frequency $\omega$. We note that when $\mathcal{F}_t \xi(t)$ exists, the complex approximation converges to $\mathcal{F}_t \xi(t) * \mathcal{F}_t \varphi(t)$. The nonconstant empirical sample density $\varphi(t)$ thus makes the complex approximation result a biased spectral estimator. The bias converges to 0 as $\mathcal{F}_t \varphi(t)$ approaches $\delta_{0,t}$, i.e. components of $\varphi(t)$ with non-zero frequency vanish. We note that

$$
\begin{aligned}
\lim_{n \to \infty} \mathrm{Var} \frac{1}{n} \sum \rho_k e^{-i\omega t_k} &= \lim_{n \to \infty} E \left( \frac{1}{n^2} \sum \rho_k e^{-i\omega t_k} \overline{\sum \rho_k e^{-i\omega t_k}} \right) \\
&= \lim_{n \to \infty} E \left( \frac{1}{n^2} \left( \sum_{k=0}^{n-1} |\rho_k|^2 + 2 \sum_{k=0}^{n-1} \sum_{j=0}^{k-1} \Re \left( \rho_k \overline{\rho_j} e^{-i\omega(t_k - t_j)} \right) \right) \right) \tag{12} \\
&= \lim_{n \to \infty} \frac{1}{n} \mathrm{Var} \, \rho_k = 0,
\end{aligned}
$$

which means that the complex approximation is a consistent estimator.

The rewritten real approximation in Eq. 8 becomes

$$\lim_{n\to\infty} c_\omega + is_\omega = \lim_{n\to\infty} \frac{2\left(\overline{\zeta_\omega} - \overline{\iota_{2\omega}}\zeta_\omega\right)}{1 - |\iota_{2\omega}|^2}$$

$$= \frac{2\left(\overline{\mathcal{F}_t x(t)\varphi(t)} - \overline{(\mathcal{F}_t\varphi(t))(2\omega)}\left(\mathcal{F}_t x(t)\varphi(t)\right)\right)}{1 - |\mathcal{F}_t\varphi(t)(2\omega)|^2}. \tag{13}$$

We consider $\omega \neq 0$. As $\mathcal{F}_t\varphi(t)$ approaches $\delta_{0,t}$, $(\mathcal{F}_t\varphi(t))(2\omega)$ vanishes, leaving a result that, except for the factor 2 explained above, is equivalent to the conjugate of the complex approximation. Under the assumed circumstances, the real approximation thus converges to something different from the complex approximation. If $\varphi$ is symmetric, then $\mathcal{F}_t\varphi(t)$ is real, and the real and imaginary parts of $\overline{\mathcal{F}_t x(t)\varphi(t)}$ are scaled individually. In the general case, the real approximation converges to a complex linear combination of the complex approximation's limit and its complex conjugate.

Since the real approximation is based on the complex approximation, it is, for $\mathcal{F}_t\varphi(t) \to \delta_{0,t}$, an unbiased estimator. Assuming a real symmetric distribution for the error term $\rho_k$ in Eq. 12, we also obtain zero variance for $\zeta_\omega$ which means that the real approximation is a consistent estimator as well.

Concluding, we can state that when the Fourier transform of the randomly sampled signal exists, i.e. the signal is a superposition of sinusoids, both the complex and the real approximation converge and provide consistent and for $\mathcal{F}_t\varphi(t) \to \delta_{0,t}$ unbiased estimators.

In order to study the convergence of the Lomb periodogram, we first introduce

$$u = \lim_{n\to\infty} e^{-i\omega\theta} = \lim_{n\to\infty} e^{-\frac{i}{2}\arg\sum e^{2i\omega t_k}}$$

$$= e^{-\frac{i}{2}\arg(\mathcal{F}_t\varphi(t))(2\omega)}.$$

Rewriting Eq. 5 and applying Eq. 11, we obtain

$$\lim_{n\to\infty} P_{\text{lomb}}(\omega) = \lim_{n\to\infty}\left(\frac{\left(\frac{1}{n}\sum x_k\cos\left(\omega\left(t_k-\theta\right)\right)\right)^2}{\frac{1}{n}\left(\frac{1}{n}\sum\cos^2\left(\omega\left(t_k-\theta\right)\right)\right)} + \frac{\left(\frac{1}{n}\sum x_k\sin\left(\omega\left(t_k-\theta\right)\right)\right)^2}{\frac{1}{n}\left(\frac{1}{n}\sum\sin^2\left(\omega\left(t_k-\theta\right)\right)\right)}\right)$$

$$= \lim_{n\to\infty}\left(\frac{\left(\frac{1}{n}\sum x_k\Re\left(e^{-i\omega\theta}e^{i\omega t_k}\right)\right)^2}{\frac{1}{n}\left(\frac{1}{2n}\sum\left(1+\Re\left(e^{-i2\omega\theta}e^{i2\omega t_k}\right)\right)\right)} + \frac{\left(\frac{1}{n}\sum x_k\Im\left(e^{-i\omega\theta}e^{i\omega t_k}\right)\right)^2}{\frac{1}{n}\left(\frac{1}{2n}\sum\left(1-\Re\left(e^{-i2\omega\theta}e^{i2\omega t_k}\right)\right)\right)}\right)$$

$$= \lim_{n\to\infty} 2n\left(\frac{\left(\frac{1}{n}\sum x_k\Re\left(ue^{i\omega t_k}\right)\right)^2}{\left(\frac{1}{n}\sum\left(1+\Re\left(u^2 e^{i2\omega t_k}\right)\right)\right)} + \frac{\left(\frac{1}{n}\sum x_k\Im\left(ue^{i\omega t_k}\right)\right)^2}{\left(\frac{1}{n}\sum\left(1-\Re\left(u^2 e^{i2\omega t_k}\right)\right)\right)}\right)$$

$$= \lim_{n\to\infty} 2n\left(\frac{\left(\Re\mathcal{F}_t\left(u\varphi(t)x(t)\right)\right)^2}{1+\left(\mathcal{F}_t\left(u^2\varphi(t)\right)\right)(2\omega)} + \frac{\left(\Im\mathcal{F}_t\left(u\varphi(t)x(t)\right)\right)^2}{1-\left(\mathcal{F}_t\left(u^2\varphi(t)\right)\right)(2\omega)}\right)$$

$$= \infty$$

unless $\mathcal{F}_t\left(u\varphi(t)x(t)\right) = 0$.

# 4. Wavelet transforms

We now combine the approximations derived above with a weighted least square approximation to formulate a windowed transform. In our case, a weighted least square term that should be minimized with respect to $(c_\omega \ s_\omega)$ reads

$$\sum \beta_k^2 \left( (c_\omega \ s_\omega) \left( \begin{array}{c} \cos \omega t_k \\ \sin \omega t_k \end{array} \right) - x_k \right)^2 \ = \ \sum \left( (c_\omega \ s_\omega) \left( \begin{array}{c} \beta_k \cos \omega t_k \\ \beta_k \sin \omega t_k \end{array} \right) - \beta_k x_k \right)^2, (14)$$

where $\beta_k^2 \equiv w(t_k)$ is an appropriate weight function.

We obtain the following pseudo-inverse solution for Eq. 14:

$$\begin{aligned}
(c_\omega \ s_\omega) \ = \ & 2 \sum \left( \beta_k^2 x_k \left( \begin{array}{c} \cos \omega t_k \\ \sin \omega t_k \end{array} \right)^{\mathrm{T}} \right) \cdot \\
& \frac{\left( \begin{array}{cc} \sum \beta_k^2 - \sum \beta_k^2 \cos 2\omega t_k & -\sum \beta_k^2 \sin 2\omega t_k \\ -\sum \beta_k^2 \sin 2\omega t_k & \sum \beta_k^2 + \sum \beta_k^2 \cos 2\omega t_k \end{array} \right)}{(\sum \beta_k^2)^2 - (\sum \beta_k^2 \cos 2\omega t_k)^2 - (\sum \beta_k^2 \sin 2\omega t_k)^2}.
\end{aligned} (15)$$

The complex version of this is again much simpler,

$$\begin{aligned}
\zeta_\omega \ &= \ \frac{\sum \beta_k^2 \, e^{-i\omega t_k} \, \xi_k}{\sum \beta_k^2} \\
&= \ \frac{\sum w(\omega t_k) \, e^{-i\omega t_k} \, \xi_k}{\sum w(\omega t_k))}.
\end{aligned} (16)$$

For continuous functions as well as for evenly sampled data, the Morlet wavelet transform is an instance of a weighted Fourier spectrum, whereby the window width scales with $\omega$. It uses the Gaussian $e^{-(\sigma\omega t)^2}$ with an additional scaling parameter $\sigma$ as weight function.

We may now define Morlet-like wavelet transforms for unevenly sampled data by using the windowed approximations from Eq. (15) and Eq. (16). The convolution $\frac{1}{a} \int \psi^*(\frac{b-\tau}{a}) f(\tau) d\tau$ uses the mother wavelet $\psi(t)$, which we choose to be $\psi(t) = w(\sigma t)e^{-i\omega t}$ in accordance with our approach. It remains to introduce a time shift. The complex case then becomes,

$$\zeta_{t,\omega} = \frac{\sum w(\sigma \omega (t_k - t)) \, e^{-i\omega(t_k - t)} \, \xi_k}{\sum w(\sigma \omega (t_k - t)))}. (17)$$

In order to accelerate the computation, we use a window function with compact support in time that does not employ transcendental functions,

$$w(t) = \left\{ \begin{array}{cc} (1 - |t|)^3 + 3|t|(1 - |t|)^2 = 1 - 3|t|^2 + 2|t|^3, & t \in [-1 \ldots 1] \\ 0 & \text{otherwise} \end{array} \right. ,$$

depicted in Figure 5.
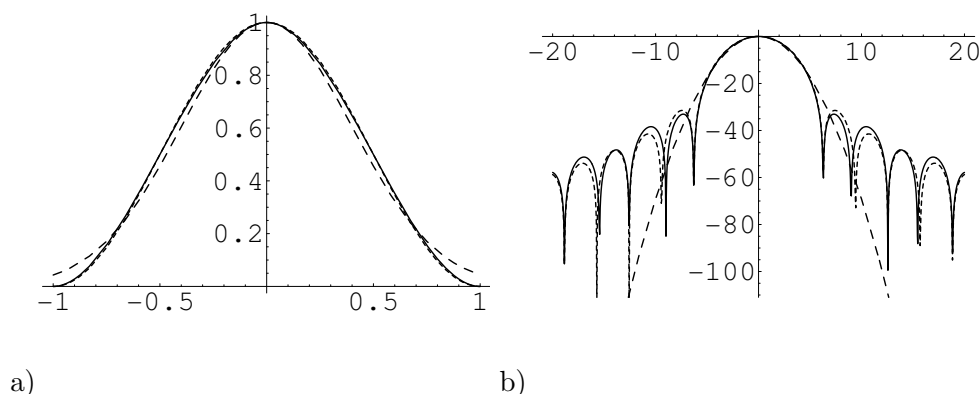
a)                                          b)

Figure 5: Shown are (a) $w(t)$ (solid) compared with the Hanning window (fine-dashed) and the appropriately scaled Morlet wavelet window $e^{-\pi t^2}$ (dashed), and (b) their logarithmic power spectra, measured in dB. The Hanning window and $w(t)$ nearly coincide in (a).

This window function has a shape that is very similar to the popular Hanning window (see Figure 5)

$$h(t) = \begin{cases} (1 + \cos(\pi t))/2, & |t| \leq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{18}$$

Its Fourier transform is

$$\hat{w}(\omega) = 12 \frac{2 - 2\cos(\omega) - \omega\sin(\omega)}{\omega^4},$$

and it produces comparable sideband ripples (see Figure 5). The first sideband ripple is even lower than the Hanning window's one and not substantially worse than the sideband falloff of the Gaussian's spectrum. All three power spectra are shown in Figure 5. An appropriate choice for the value of $\sigma$ is 0.05. This means that $\frac{2}{\sigma}$ radians or $\frac{1}{\pi\sigma}$ periods of $e^{-i\omega t}$ fit into the whole support of the weight function.

In order to estimate the spectral phase differences between two complex time series $\xi_k, t_k, k \in 0, \ldots, K-1$ and $\eta_l, u_l, l \in 0, \ldots, L-1$, we use the complex conjugate product of their approximations according to Eq. 17,

$$\frac{\sum_{k=0}^{K-1} w(\sigma\omega(t_k - t)) e^{-i\omega(t_k - t)} \xi_k}{\sum_{k=0}^{K-1} w(\sigma\omega(t_k - t))} \frac{\sum_{l=0}^{L-1} w(\sigma\omega(u_l - t)) e^{i\omega(u_l - t)} \overline{\eta_l}}{\sum_{l=0}^{L-1} w(\sigma\omega(u_l - t))} =$$
$$\frac{\sum_{k=0}^{K-1} w(\sigma\omega(t_k - t)) e^{-i\omega t_k} \xi_k}{\sum_{k=0}^{K-1} w(\sigma\omega(t_k - t))} \frac{\sum_{l=0}^{L-1} w(\sigma\omega(u_l - t)) e^{i\omega u_l} \overline{\eta_l}}{\sum_{l=0}^{L-1} w(\sigma\omega(u_l - t))} . \tag{19}$$

Since $e^{-i\omega t_k} \xi_k$ and $e^{i\omega u_l} \eta_l$ do not vary with t, they can be precomputed, which accelerates the computation in comparison to Eq. 17 when the result is computed with increasingly fine time resolution.

# 5. Fast approximation

We use the power series approximation of $e^{it}$ to formulate a divide-and-conquer scheme for the evaluation of Eq. 7 consisting of 4 fundamental steps described below. A fast algorithm for the real approximation that makes use of the complex approximation is given thereafter.

## 5.1. Complex approximation

*Subdivision*

In a first step, the index set $K = \{0, \ldots, n-1\}$ is subdivided into subsets $K_h$ with associated $\tau_h$ such that for a given highest analysis frequency $\omega_{max}$, when $e^{-i\omega_{max}(t_k - \tau_h)}$ is approximated by a power series, the error is sufficiently small.

The sampling times for $k \in K_h$ obey $-\Delta\tau \leq t_k - \tau_h < \Delta\tau$, which is chosen as $\Delta\tau = \frac{c}{\omega_{max}}$, i.e. a constant multiple of the period length defined by $\omega_{max}$ (Floating point processors often use $c = \frac{\pi}{2}$ for computing the sine and cosine functions). The initial $\tau_0$ is chosen such that $-\Delta\tau \leq t_0 - \tau_0$, and $\tau_{h+1} = 2\Delta\tau + \tau_h$. The subdivision yields $H = \left\lceil \frac{t_{n-1} - t_0}{2\Delta\tau} \right\rceil$ subsets.

*Precomputation*

Next, for each $K_h$, the vector $\mathbf{s}_h$ of the $P+1$ first terms of the power series expansion of $\frac{1}{n} \sum_{k \in K_h} x_k e^{-i(t_k - \tau_h)}$ is precomputed as

$$\mathbf{s}_h = \{s_{h,p}\}_{0 \leq p \leq P} = \left\{ \sum_{k \in K_h} \frac{x_k \left(-i\mu \left(t_k - \tau_h\right)\right)^p}{p!} \right\}_{0 \leq p \leq P} .$$

The factor $\mu$ plays a role in the improvement of numeric accuracy. For now, we can assume $\mu = 1$.

*Summation for a given $\omega$*

Using the $\mathbf{s}_h$, we now compute the spectral coefficients for $\frac{1}{2}\omega_{max} < \omega \leq \omega_{max}$.

$$
\begin{aligned}
\frac{1}{n} \sum_{k=0}^{n-1} x_k e^{-i\omega t_k} &= \frac{1}{n} \sum_{h=0}^{H-1} e^{-i\omega\tau_h} \sum_{k \in K_h} x_k e^{-i\omega(t_k - \tau_h)} \\
&= \frac{1}{n} \sum_{h=0}^{H-1} e^{-i\omega\tau_h} \sum_{k \in K_h} \sum_{p=0}^{\infty} \left(\frac{\omega}{\mu}\right)^p \frac{x_k \left(-i\mu \left(t_k - \tau_h\right)\right)^p}{p!} \\
&\approx \frac{1}{n} \sum_{h=0}^{H-1} e^{-i\omega\tau_h} \sum_{p=0}^{P} \left(\frac{\omega}{\mu}\right)^p \sum_{k \in K_h} \frac{x_k \left(-i\mu \left(t_k - \tau_h\right)\right)^p}{p!} \\
&= \frac{1}{n} \sum_{h=0}^{H-1} e^{-i\omega\tau_h} \sum_{p=0}^{P} \left(\frac{\omega}{\mu}\right)^p s_{h,p}.
\end{aligned}
$$

Using the fact that the $\tau_h$ are evenly spaced, the $e^{-i\omega\tau_h}$ do not have to be calculated explicitly but can be obtained from $e^{-i\omega\tau_0}$ by repetitive multiplication of $e^{-2i\omega\Delta\tau}$.

*Merging of $\mathbf{s}_{2h}$ and $\mathbf{s}_{2h+1}$ into $\mathbf{s}'_h$*

For the upper limit of the next frequency interval $\omega'_{max} := \frac{1}{2}\omega_{max}$, the power series approximation has the required exactness over an interval of double length. It follows that the $\mathbf{s}_{2h}$ and $\mathbf{s}_{2h+1}$ can be combined to $\mathbf{s}'_h$, using the offset $\tau'_h := \frac{1}{2}(\tau_{2h} + \tau_{2h+1})$ and size $\Delta\tau' := 2\Delta\tau$ belonging to the index set $K'_h$ with $-\Delta\tau' \leq t_k - \tau'_h < \Delta\tau'$ for all $k \in K'_h$. The index $h$ now runs from 0 to $H' := \lceil \frac{H}{2} \rceil$. The vectors of power series expansion terms $\mathbf{s}_{2h}$ and $\mathbf{s}_{2h+1}$ are combined into $\mathbf{s}'_h$ using reexpansion:

$$
\begin{aligned}
\mathbf{s}'_h &= \left\{ \sum_{k \in K'_h} \frac{x_k (-i\mu(t_k - \tau'_h))^p}{p!} \right\}_{0 \leq p < P} \\
&= \left\{ \sum_{k \in K_{2h}} \frac{x_k (-i\mu(t_k - \tau_{2h} + \tau_{2h} - \tau'_h))^p}{p!} \right\}_{0 \leq p < P} + \\
&\quad \left\{ \sum_{k \in K_{2h+1}} \frac{x_k (-i\mu(t_k - \tau_{2h+1} + \tau_{2h+1} - \tau'_h))^p}{p!} \right\}_{0 \leq p < P} \\
&= \left\{ \sum_{k \in K_{2h}} x_k \sum_{q=0}^{p} \binom{p}{q} \frac{(-i\mu(t_k - \tau_{2h}))^q (-i\mu(\tau_{2h} - \tau'_h))^{p-q}}{p!} \right\}_{0 \leq p < P} + \\
&\quad \left\{ \sum_{k \in K_{2h+1}} x_k \sum_{q=0}^{p} \binom{p}{q} \frac{(-i\mu(t_k - \tau_{2h+1}))^q (-i\mu(\tau_{2h+1} - \tau'_h))^{p-q}}{p!} \right\}_{0 \leq p < P} \\
&= \left\{ \sum_{q=0}^{p} \frac{(-i\mu(\tau_{2h} - \tau'_h))^{p-q}}{(p-q)!} \sum_{k \in K_{2h}} x_k \frac{(-i\mu(t_k - \tau_{2h}))^q}{q!} \right\}_{0 \leq p < P} + \\
&\quad \left\{ \sum_{q=0}^{p} \frac{(-i\mu(\tau_{2h+1} - \tau'_h))^{p-q}}{(p-q)!} \sum_{k \in K_{2h+1}} x_k \frac{(-i\mu(t_k - \tau_{2h+1}))^q}{q!} \right\}_{0 \leq p < P} \\
&= \left\{ \sum_{q=0}^{p} s_{2h,q} \frac{(-i\mu(\tau_{2h} - \tau'_h))^{p-q}}{(p-q)!} + \sum_{q=0}^{p} s_{2h+1,q} \frac{(-i\mu(\tau_{2h+1} - \tau'_h))^{p-q}}{(p-q)!} \right\}_{0 \leq p < P} \\
&= \left\{ \sum_{r=0}^{p} s_{2h,p-r} \frac{(-i\mu(\tau_{2h} - \tau'_h))^{r}}{r!} + \sum_{r=0}^{p} s_{2h+1,p-r} \frac{(-i\mu(\tau_{2h+1} - \tau'_h))^{r}}{r!} \right\}_{0 \leq p < P}. \quad (20)
\end{aligned}
$$

With the introduction of $\Delta\tau = \tau_{2h+1} - \tau'_h = -(\tau_{2h} - \tau'_h)$, this can be further simplified:

$$
\mathbf{s}'_h = \left\{ \sum_{r=0}^{p} s_{2h,p-r} \frac{(-i\mu(-\Delta\tau))^{r}}{r!} + \sum_{r=0}^{p} s_{2h+1,p-r} \frac{(-i\mu\Delta\tau)^{r}}{r!} \right\}_{0 \leq p < P}
$$

$$= \left\{ \sum_{r=0}^{p} \frac{\left((-1)^r s_{2h,p-r} + s_{2h+1,p-r}\right)(-i\mu\Delta\tau)^r}{r!} \right\}_{0 \le p < P} \tag{21}$$

$$= \left\{ \sum_{r=0}^{p} \frac{\left(s_{2h,p-r} + (-1)^r s_{2h+1,p-r}\right)(i\mu\Delta\tau)^r}{r!} \right\}_{0 \le p < P} \tag{22}$$

$$= \left\{ \sum_{q=0}^{\lfloor \frac{p}{2} \rfloor} \frac{\left(s_{2h,p-2q} + s_{2h+1,p-2q}\right)(-1)^q (\mu\Delta\tau)^{2q}}{(2q)!} + \right.$$

$$\left. i \sum_{q=0}^{\lfloor \frac{p-1}{2} \rfloor} \frac{\left(s_{2h,p-2q-1} - s_{2h+1,p-2q-1}\right)(-1)^q (\mu\Delta\tau)^{2q+1}}{(2q+1)!} \right\}_{0 \le p < P} \tag{23}$$

This process is repeated until either $\omega_{max}^{(d+1)} = \frac{1}{2}\omega_{max}^{(d)}$ passes the desired minimal analysis frequency $\omega_0$, or the merging of adjacent subsets leaves just one subset behind (which means that at frequencies below the $\omega_{max}^{(d)}$ where this occurs, the time range of the time series covers merely a fraction of a full period of $e^{-i\omega t}$, which is of questionable value, except for $\omega = 0$). When the number of subsets in one level is odd, the last remaining vector $\mathbf{s}_{H^{(d)}}^{(d)}$ is merged with an empty one.

## 5.2. Real approximation

Based on the scheme for complex approximation and the complex representation of the real approximation in Eq. 8, a computation scheme for the real approximation can be constructed,

$$c_\omega + is_\omega = \frac{2\left(\overline{\zeta_\omega} - \overline{\iota_{2\omega}}\zeta_\omega\right)}{1 - |\iota_{2\omega}|^2} \, .$$

As in Eq. 7, $\zeta_\omega$ denotes the complex approximation result for the considered time series, whereas $\iota_{2\omega}$ denotes the complex approximation at frequency $2\omega$ of the unit time series $(1 \ldots 1)$ over $(t_0 \ldots t_{n-1})$ that can be computed in a separate subdivision scheme with a few obvious simplifications stemming from the fact that the sample values equal 1.

We note that in the real case, the $s_{h,p}$ are real when $p$ is even and imaginary otherwise. In Eq. 23, it can be seen that most of the computation steps for the merging process can be performed with purely real arithmetic, which further speeds up the computation.

## 5.3. Complexity considerations

We analyze the application of the fast scheme onto logarithmic and linear ranges of frequencies. In the logarithmic case, the algorithm provides the highest acceleration compared to the straighforward implementation of 7. Assuming that $n_{\text{oct}}$ octaves of $n_{\text{coeff}}$ spectral coefficients per octave shall be computed, the following steps, each shown with their complexity, have to be performed:

- Subdivision and precomputation: $O(nP)$,

- Summation for $n_{\text{coeff}}$ spectral coefficients in each of the $n_{\text{oct}}$ octaves with $\omega_{\max}^{(d+1)} < \omega \le \omega_{\max}^{(d)}$, $0 \le d < n_{\text{oct}}$: $O\left(n_{\text{coeff}} P \frac{H}{2^d}\right)$,

- Merging of the $\mathbf{s}_h$ and $\mathbf{s}_{h+1}$: $O\left(P^2 \frac{H}{2^d}\right)$.

Summing everything up and applying the sum of geometric series, we obtain a total complexity of

$$
O \;=\; O\left(nP + \sum_{d=0}^{n_{oct}-1} \left(n_{\text{coeff}} + 1\right) P^2 \frac{H}{2^d}\right) = O\left(P\left(n + n_{\text{coeff}} HP\right)\right).
$$

When a linear frequency range $\omega_l = \frac{l\omega_{\max}}{L}$, $l \in \{0, \dots, L\}$ is desired, the number of coefficients that can be computed with subdivision level $d$ is $\left\lfloor \frac{L}{2^d} \right\rfloor - \left\lfloor \frac{L}{2^{d+1}} \right\rfloor = O\left(\frac{L}{2^d}\right)$. For $\omega = 0$, the coefficient is computed using the $0^{\text{th}}$ term of the $\mathbf{s}_h$. Summation of the above and substitution of geometric series yields

$$
O = O\left(nP + \sum_{d=0}^{d_{\max}} \left(\frac{L}{2^d} + 1\right) P^2 \frac{H}{2^d}\right) = O\left(P\left(n + H + LH\right)\right) = O\left(P\left(n + LHP\right)\right).
$$

### 5.4. Weighted complex approximation

Here, we use the Hanning window from Eq. 18. We build upon subdivision and precomputation steps similar to the ones described above, with

$$
\mathbf{s}_h = \{s_{h,p}\}_{0 \le p \le P} = \left\{ \sum_{k \in K_h} \frac{x_k \left(-i\left(t_k - \tau_h\right)\right)^p}{p!} \right\}_{0 \le p \le P}
$$

and

$$
\mathbf{j}_h = \{j_{h,q}\}_{0 \le q \le Q} = \left\{ \sum_{k \in K_h} \frac{\left(i\left(t_k - \tau_h\right)\right)^q}{q!} \right\}_{0 \le q \le Q},
$$

where $Q$ can be chosen smaller than $P$.

We then rewrite Eq. 17 as follows,

$$
\frac{\displaystyle\sum_{k \in K} x_k w\left(\sigma\omega\left(t_k - t\right)\right) e^{-i\omega(t_k - t)}}{\displaystyle\sum_{k \in K} w\left(\sigma\omega\left(t_k - t\right)\right)} \;=\; \frac{\displaystyle\sum_{k \in K(t)} x_k \left(1 + \cos\left(\sigma\omega\left(t_k - t\right)\right)\right) e^{-i\omega(t_k - t)}}{\displaystyle\sum_{k \in K(t)} \left(1 + \cos\left(\sigma\omega\left(t_k - t\right)\right)\right)}
$$

$$
\approx \frac{\displaystyle\sum_{h \in H(t)}\sum_{k \in K_h} x_k \left(e^{-i\omega(t_k - t)} + \frac{1}{2}e^{-i\omega_-(t_k - t)} + \frac{1}{2}e^{-i\omega_+(t_k - t)}\right)}{\displaystyle\sum_{h \in H(t)}\sum_{k \in K_h} \left(1 + \frac{1}{2}e^{i\sigma\omega(t_k - t)} + \frac{1}{2}e^{-i\sigma\omega(t_k - t)}\right)}
$$

$$\approx \frac{\sum_{h \in H(t)} \left( e^{-i\omega(\tau_h - t)} \sum_{p=0}^{P} \omega^p s_{h,p} + \frac{1}{2} e^{-i\omega_-(\tau_h - t)} \sum_{p=0}^{P} \omega_-^p s_{h,p} + \frac{1}{2} e^{-i\omega_+(\tau_h - t)} \sum_{p=0}^{P} \omega_+^p s_{h,p} \right)}{\sum_{h \in H(t)} \left( 1 + \frac{1}{2} e^{i\sigma\omega(\tau_h - t)} \sum_{q=0}^{Q} (\sigma\omega)^q j_{h,q} + \frac{1}{2} e^{-i\sigma\omega(\tau_h - t)} \sum_{q=0}^{Q} (-\sigma\omega)^q j_{h,q} \right)}$$

using the abbreviations $\omega_- = \omega(1 - \sigma)$ and $\omega_+ = \omega(1 + \sigma)$. $K(t)$ denotes the range $k_a \ldots k_b$ of indices of samples that lie within the translated and scaled Hanning window, i.e. $\sigma\omega(t_{k_a} - t) < \pi$ and $\sigma\omega(t_{k_b} - t) < \pi$. $H(t)$ denotes the range of indices of subdivision intervals that cover the Hanning window. The factors $e^{-i\omega_-(\tau_h - t)}$, $e^{-i\omega(\tau_h - t)}$, $e^{-i\omega_+(\tau_h - t)}$ and $e^{i\sigma\omega(\tau_h - t)}$ can be multiplied up iteratively from their respective start value for $\tau_{h_a}$. The pairwise merging of the $\mathbf{s}_h$ and $\mathbf{j}_h$ is performed similarly to the complex approximation. It has to be noted that the process of covering the periodic extension of the Hanning window used with precomputed sample ranges introduces small errors since the sample range boundaries almost never coincide with the window's boundary.

## 6. Approximation for time series with constant offset

When analyzing time series with a constant offset, e.g. $\mathbf{x} + d(1 \ldots 1)$, $d \in \mathbb{R}$, using Eq. 2, we have $d(1 \ldots 1)\Omega_{\mathbf{t}}^{\mathrm{T}} \left( \Omega_{\mathbf{t}} \Omega_{\mathbf{t}}^{\mathrm{T}} \right)^{-1} \neq 0$ in the general case, due to the nonorthogonality of the row vectors of $\Omega_{\mathbf{t}}$ and the unit vector. This leads to increasingly wrong spectral coefficients for low frequencies.

We therefore compute

$$(c_\omega \ s_\omega \ d_\omega) \begin{pmatrix} \cos\omega t_0 & \cos\omega t_1 & \ldots & \cos\omega t_{n-1} \\ \sin\omega t_0 & \sin\omega t_1 & \ldots & \sin\omega t_{n-1} \\ 1 & 1 & \ldots & 1 \end{pmatrix} = (x_0 \ x_1 \ \ldots \ x_{n-1}) , \qquad (24)$$

by means of the pseudoinverse,

$$(c_\omega \ s_\omega \ d_\omega) = 2 \sum \left( x_k \begin{pmatrix} \cos\omega t_k \\ \sin\omega t_k \\ 1 \end{pmatrix}^{\mathrm{T}} \right) \begin{pmatrix} n + \sum\cos 2\omega t_k & \sum\sin 2\omega t_k & 2\sum\cos\omega t_k \\ \sum\sin 2\omega t_k & n - \sum\cos 2\omega t_k & 2\sum\sin\omega t_k \\ 2\sum\cos\omega t_k & 2\sum\sin\omega t_k & 2n \end{pmatrix}^{-1}$$

$$= 2n \left( \Re\zeta_\omega \ -\Im\zeta_\omega \ \langle\mathbf{x}\rangle \right) \left( n \begin{pmatrix} 1 + \Re\iota_{2\omega} & -\Im\iota_{2\omega} & 2\Re\iota_\omega \\ -\Im\iota_{2\omega} & 1 - \Re\iota_{2\omega} & -2\Im\iota_\omega \\ 2\Re\iota_\omega & -2\Im\iota_\omega & 2 \end{pmatrix} \right)^{-1} \qquad (25)$$

$$= \left( \Re\zeta_\omega \ -\Im\zeta_\omega \ \langle\mathbf{x}\rangle \right) \frac{\begin{pmatrix} 2 - 4(\Im\iota_\omega)^2 - 2\Re\iota_{2\omega} & 2\Im\iota_{2\omega} - 4\Re\iota_\omega\Im\iota_\omega & 2\Re\iota_\omega(\Re\iota_{2\omega}-1)+2\Im\iota_{2\omega}\Im\iota_\omega \\ 2\Im\iota_{2\omega} - 4\Re\iota_\omega\Im\iota_\omega & 2 - 4(\Re\iota_\omega)^2 + 2\Re\iota_{2\omega} & 2\Im\iota_\omega(\Re\iota_{2\omega}+1)-2\Im\iota_{2\omega}\Re\iota_\omega \\ 2\Re\iota_\omega(\Re\iota_{2\omega}-1)+2\Im\iota_{2\omega}\Im\iota_\omega & 2\Im\iota_\omega(\Re\iota_{2\omega}+1)-2\Im\iota_{2\omega}\Re\iota_\omega & 1 - |\iota_{2\omega}|^2 \end{pmatrix}}{1 - |\iota_{2\omega}|^2 - 2|\iota_\omega|^2 + 2\Re\iota_{2\omega}\left( (\Re\iota_\omega)^2 - (\Im\iota_\omega)^2 \right) + 4\Re\iota_\omega\Im\iota_\omega\Im\iota_{2\omega}} .$$
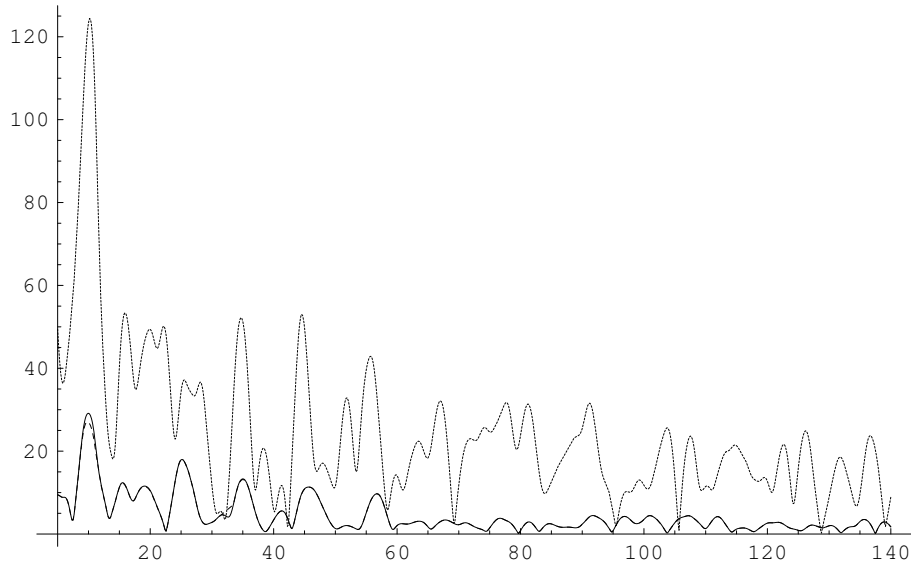
Figure 6: The absolute value from Eq. 25 applied to the Vostok $CO_2$series (solid) compared with the absolute value from Eq. 2 applied to the $CO_2$ series (densely dashed) and to the $CO_2$ series with subtracted mean value (widely dashed, nearly coincident with the solid curve).

using the symbols $\zeta_\omega$ and $\iota_\omega$ as above to denote complex approximations for the time series and the unit series. The $\zeta_\omega$, $\iota_\omega$ and $\iota_{2\omega}$ can be computed in divide-and-conquer schemes, whereby the results for $\iota_\omega$ can be reused as $\iota_{2\omega}$ in subsequent octaves.

When looking at Eqs. 1 and 24, it becomes obvious that at low frequencies, the approximation becomes problematic. In Figure 7 which displays results down to a frequency corresponding to a third of one period over the full sample time span of 422000 years, it can be seen that the low frequency coefficients start to diverge. As a rule of thumb, we note that approximation of time series over a time span of less than half a period becomes undefined.

Looking for periodic components in a signal, we did some promising experiments with the simultaneous approximation of sinusoids with $m$ integer multiple frequencies of a given $\omega_0$, using the pseudoinverse solution of

$$
(d_\omega \ c_\omega \ s_\omega \ldots c_{m\omega} \ s_{m\omega})
\left\{
\begin{array}{c}
1 \\
\cos \omega t_k \\
\sin \omega t_k \\
\vdots \\
\cos m\omega t_k \\
\sin m\omega t_k
\end{array}
\right\}_{k \in K}
= \mathbf{x} \, .
$$

It is possible to construct a divide-and-conquer scheme for these approximations as well. It becomes impracticable however to perform the matrix inversion algebraically.
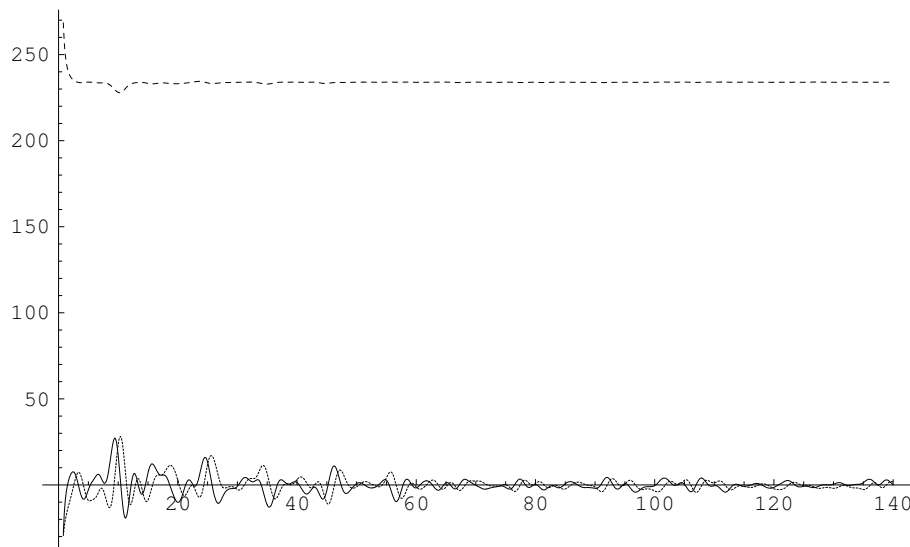
Figure 7:  Cosine (solid), sine (densely dashed) and constant coefficients (widely dashed) of Eq. 25 applied to the Vostok $CO_2$ series.

# 7. Application samples

The software accompanying this paper is written as an R module, and was developed under Linux. Besides the functions written in native R in order to provide a comprehensive sample, the real and complex spectral transform algorithms and the weighted transform, as well as their non-accelerated counterparts have been implemented in C as a loadable library. We give an overview and a couple of examples using the implemented functions here; please refer to the online documentation in R for details concerning their application and further examples.

The R package is distributed as a compressed installation directory.  Please use the shell command

```
tar xzvf nuspectral.tgz; R CMD install nuspectral
```

to install the package into the appropriate directories of your R installation. You may have to login as super-user first before performing the installation.

After having started R, please use

```
library("nuspectral")
```

to load the library. For the function names mentioned below and the data sets provided with the package, help pages are available that can be viewed using the R help command, e.g.

```
help("fastnureal")
```

The following functions, each operating on the time series defined by the abscissa vector X and the ordinate vector Y , are provided:

- `lombcoeff(X, Y, o)` computes a coefficient of the Lomb periodogram for the given circular frequency.

- `lombnormcoeff(X, Y, o)` computes a coefficient of the Lomb normalized periodogram for the given circular frequency.

- `nurealcoeff(X, Y, omegamax, ncoeff, noctave)` nurealcoeff

- `nuwaveletcoeff(X, Y, t, o, wgt = cubicwgt, wgtrad = 1)` computes a coefficient of the nonuniform wavelet transform in Eq. 17 at the given time and circular frequency. The weight function is by default the cubic polynomial weight function shown in Figure 5, and the radius beyond which the weight function is truncated to 0 is by default 1.

- `nucorrcoeff(X1, Y1, X2, Y2, t, o, wgt = cubicwgt, wgtrad = 1)` computes a coefficient of the wavelet correlation of the time series given by abscissa vectors X1 and X2 and ordinate vectors Y1 and Y2 respectively.

The following functions have been implemented in C.

- `nucomplex(X, Y, omegamax, ncoeff, noctave)` computes the logarithmically scaled complex spectral approximation of Eq. 7 for the given maximum circular frequency, number of coefficients and coefficients per octave.

- `fastnucomplex(X, Y, omegamax, ncoeff, noctave)` is the accelerated version of the above according to section 5.1.

- `nureal(X, Y, omegamax, ncoeff, noctave)` computes the logarithmically scaled real spectral approximation of Eq. 3 for the given maximum circular frequency, number of coefficients and coefficients per octave.

- `fastnureal(X, Y, omegamax, ncoeff, noctave)` is the accelerated version of the above according to section 5.2.

- `nurealwavelet(X, Y, omegamax, ncoeff, noctave, tmin, tmax, tsubdiv, sigma=0.1)` computes the wavelet spectrum of Eq. 17 for real Y, the given maximum circular frequency, number of coefficients, coefficients per octave, time interval and time subdivision.

- `fastnurealwavelet(X, Y, omegamax, ncoeff, noctave, tmin, tmax, tsubdiv, sigma=0.1)` computes the accelerated wavelet spectrum using the Hanning window according to section 5.4.

The data we will be using stem from

- the Vostok Lake ice drilling core (Petit JR *et al.* 1999); the ice column extracted from a more than 3 km deep hole drilled into the antarctic ice over the huge frozen lake near the Russian Vostok research station provides invaluable information about climate, chemical composition of the atmosphere and other factors within the last 422000 years. The time series are all nonuniform, and almost none of the sample times of different series match.

- tick data from the 11. July 2001 at the London stock exchange. Given are the rates of all stocks sampled at the transaction times. Since the rates change upon trade transactions that occur more or less randomly, these time series are also sampled nonuniformly.

Distributed in R format with the package are the Vostok data series

- `co2` – the $CO_2$ content of the ice.

- `ch4` – the Methane content of the ice.

- `deut` – the Deuterium content of the ice; the ratio of Deuterium to total Hydrogen is a proxy for the atmospheric temperature.

- `o18` – the $^{18}O$ (a heavy Oxygen isotope) content of the air enclosed in the ice, a proxy for the solar irradiation.

- `dust` – the dust content of the ice.

The stock data example has been downloaded from the internet, and has not been included into the R package because of copyright issues.

The wavelet transform diagrams below haven't been generated with R. Some exemplary invocations of R functions that perform the equivalent transformations are presented, however, in the respective figure captions.

# 8. Summary

We have presented a way to compute spectra for nonuniformly sampled data based on least square approximation and the Moore-Penrose pseudoinverse, both for real and complex approximations. The presented methods are shift-covariant in time and provide phase information. We have analyzed the correspondence of our methods with the discrete Fourier transform and the impact of constant signal offsets onto the spectral coefficients, and studied estimator properties and the convergence of our results for the sample counts going towards infinity. We further have introduced weighted versions of these spectral transforms that provide wavelet transforms for nonuniformly sampled data.

As a most remarkable result, we have presented asymptotically fast algorithms for the computation of the real and complex spectral analysis, which makes our methods comparably as powerful as the fast Fourier transform. These fast and robust methods to evaluate the spectral signal in time series open up new possibilities for real-time applications of increasing importance in many fields, e.g. in medicine, physiology as well as video and audio processing. The methods have also been applied in an interactive work of art where the user can explore animated 3D renderings in image and sound of a spectral analysis of the Vostok data. Our methods also promise to be useful for reconstruction and interpolation of missing and distorted data. The potentiality of our methods has been hinted by the presented relevant applications.

Future work will include an improvement of the numerical basis of our fast algorithms, and the development of other fast algorithms based on the presented scheme.
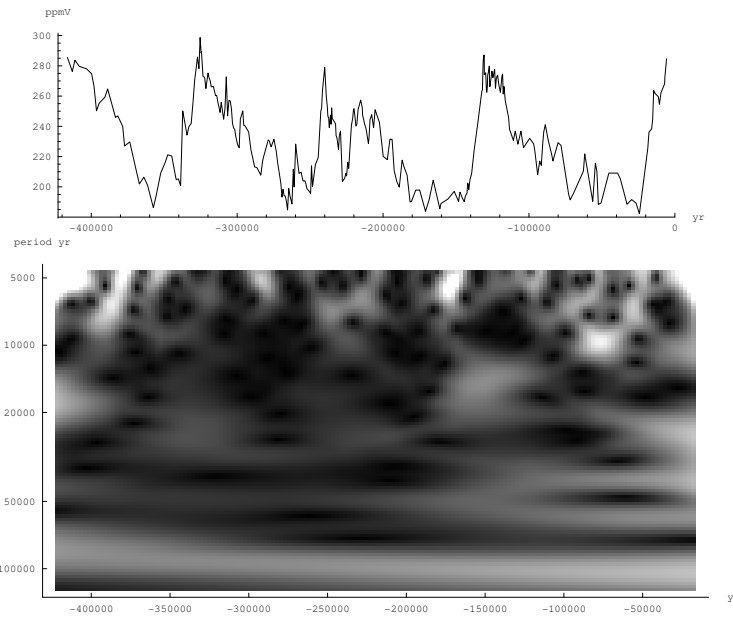
Figure 8: The Vostok $CO_2$ series and the absolute square value of Eq. 17 applied to it and displayed as intensity. The prominent peak at a period of approx. 100000 years appears to have a diminishing frequency. The R table whose absolute square value has been visualized was obtained with

```
nurealwavelet(co2[[2]],co2[[4]],0.0015,100,20,0,420000,10000)
```
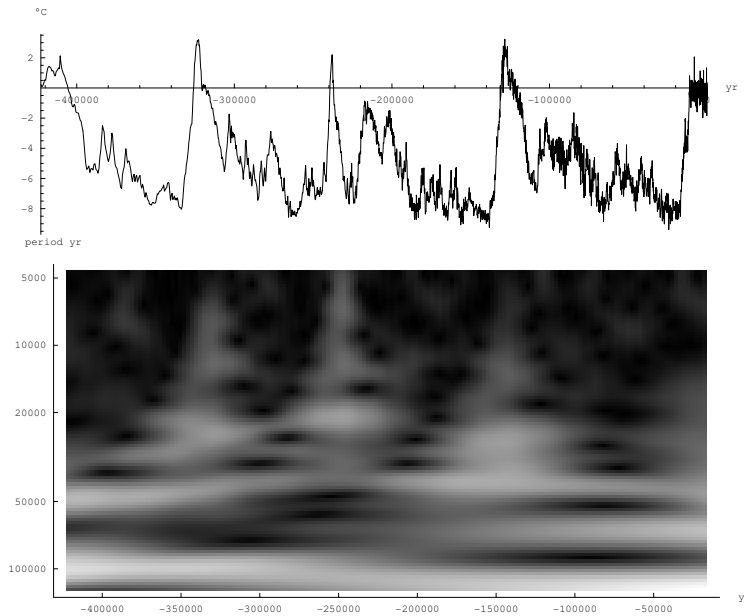


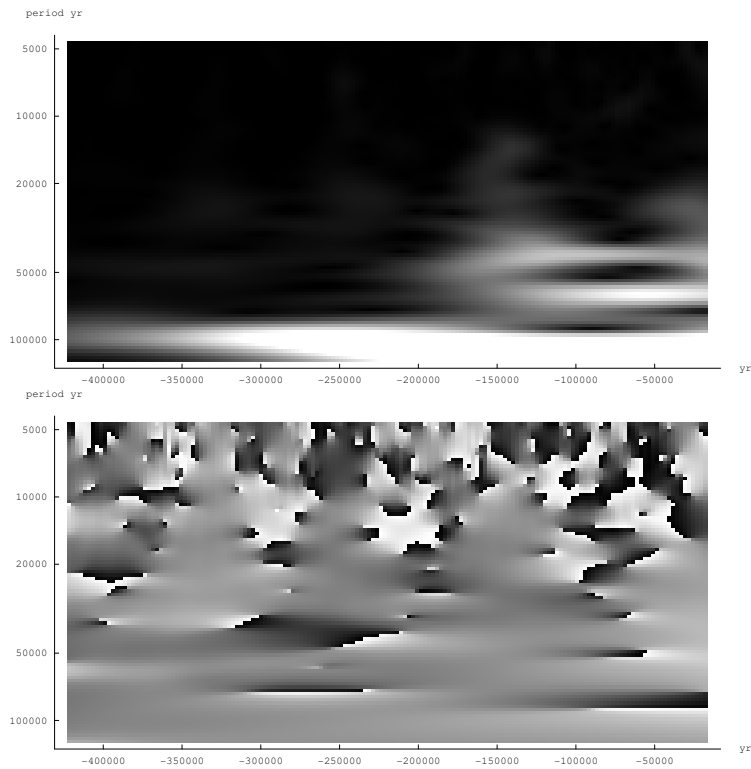Figure 9: The Vostok Deuterium series and the absolute square value of Eq. 17 applied to it.

Figure 10: The absolute value and the phase of Eq. 19 applied to the Vostok $CO_2$ and Deuterium series, showing an interdependence of the two values over time and frequency. Medium gray encodes in-phase, white and black encode opposite phase, light gray indicates that the $CO_2$ value is ahead of the Deuterium value. Further studies of results of this kind could provide a deeper understanding of climatic mechanisms. The R table whose absolute square value and phase have been visualized was obtained with

```
nurealwavelet(co2[[2]],co2[[4]],0.0015,100,20,0,420000,10000) *
Conj(nurealwavelet(deut[[2]],deut[[4]],0.0015,100,20,0,420000,10000))
```
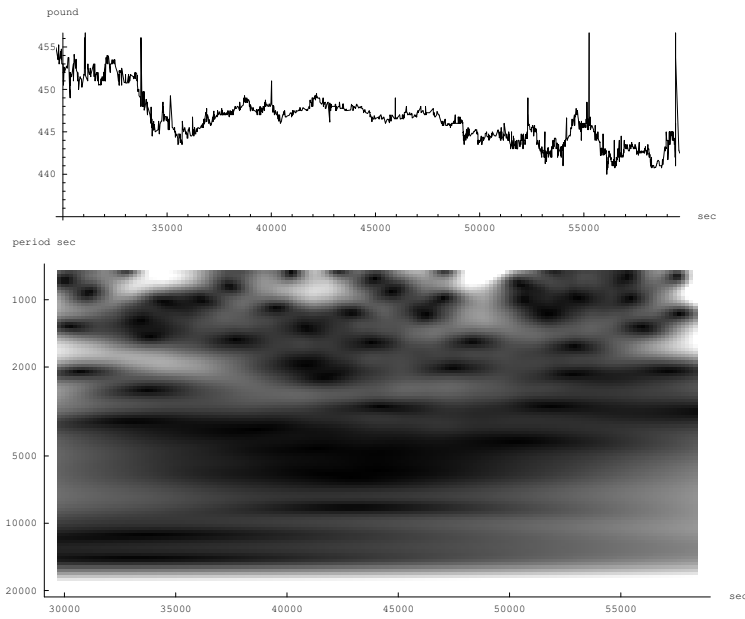
Figure 11: Tick data for British Telecom on 11/07/2001 at the London stock exchange and the absolute value of Eq. 17 applied to it. The abscissa values are in seconds past midnight.
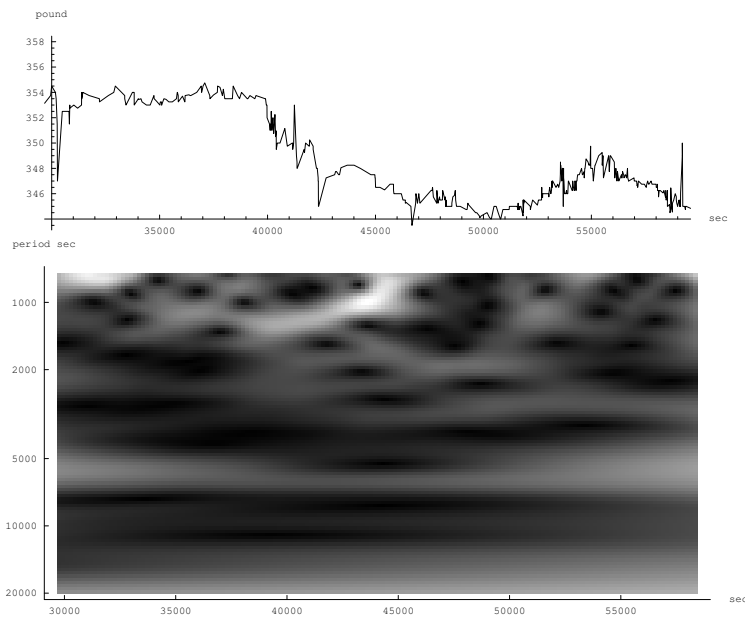


Figure 12: Tick data for British Airways on 11/07/2001 at the London stock exchange and the absolute value of Eq. 17 applied to it. The abscissa values are in seconds past midnight.
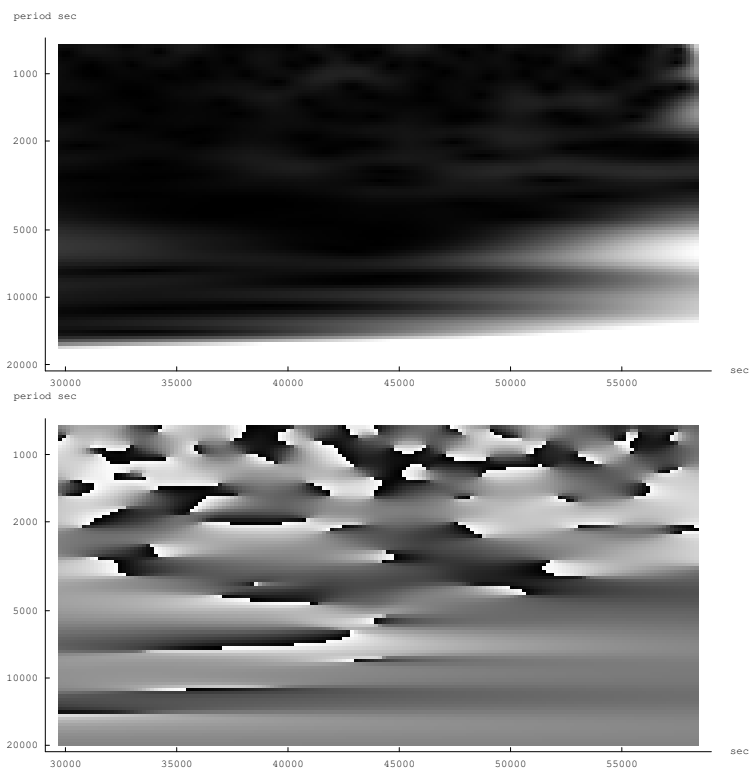
Figure 13: Absolute value and phase of Eq. 19 applied to the tick data of British Airways and British Telecom as shown above.
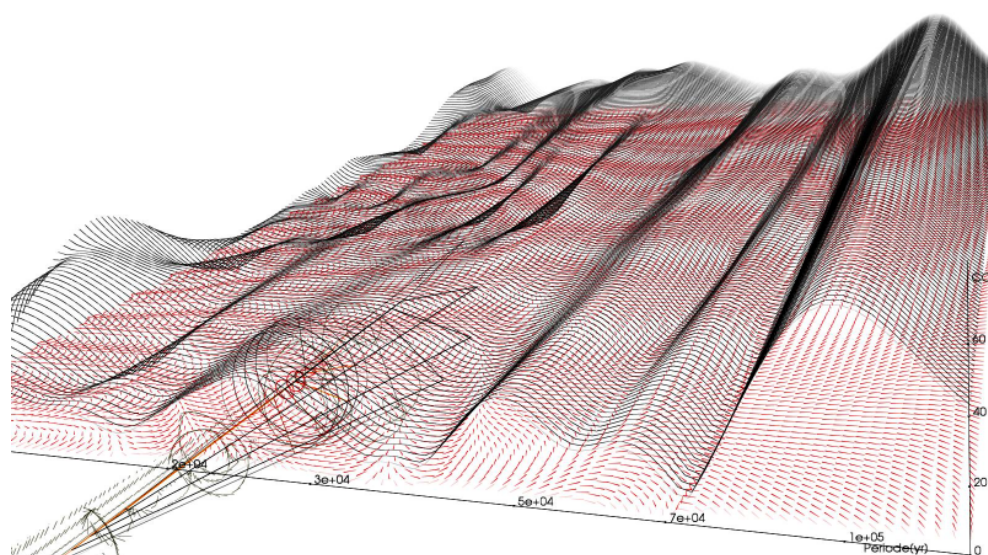
Figure 14: A screen shot from "Algorithmic Echolocation", an interactive installation by two of the authors (R.Guardans and A.Mathias) and others, shown at the exhibition cycle *Banquete* (2003; Palacio de la Virreina, Barcelona; ZKM, Karlsruhe; Medialab, Madrid). The installation allows the visitor to navigate through the Vostok data material (here the $CO_2$ data set) in a way that is inspired by a sound sampler, using loops and variable playback speed. The data is analyzed using Eq. 17, and the absolute values are displayed as elevation and the phase as vector field. Further, an animated graphic representation of the equation acting on the data is shown. Beyond graphics, the original data and the spectral analysis result are used to directly control sound parameters of a software synthesizer, a process called audification.

# Acknowledgements

# References

Bretthorst G (2000). "Nonuniform Sampling: Bandwidth and Aliasing." In G Erickson (ed.), "Maximum Entropy and Bayesian Methods," Kluwer.

Chui C (1992). *An Introduction to Wavelets*. Academic Press, Boston.

Daubechies I (1987). "The Wavelet Transform, Time-Frequency Localization and Signal Analysis." *IEEE Transactions on Information Theory*, **36**, 961–1005.

Holschneider M (1995). *Wavelets – An Analysis Tool*. Oxford University Press, New York.

Lomb N (1976). "Least-Squares Frequency Analysis of Unequally Spaced Data." *Astrophysics and Space Science*, **39**, 447–462.

Penrose R (1955). "A Generalized Inverse for Matrices." In "Proceedings of the Cambridge Philosophical Society 51," pp. 406–413.

Petit JR *et al* (1999). "Climate and Atmospheric History of the Past 420 000 Years from the Vostok Ice Core, Antarctica." *Nature*, **399**, 429–436.

Press W, Teukolsky S, Vetterling W, Flannery B (1992). *Numerical Recipes in C: The Art of Scientific Computing*, pp. 575–584. Cambridge University Press, 2 edition.

Scargle J (1976). "Studies in Astronomical Time Series Analysis II. Statistical Aspects of Spectral Analysis of Unevenly Spaced Data." *Astrophysics and Space Science*, **302**, 757–763.

**Affiliation:**

Adolf Mathias
ZKM - Center for Art and Media
Institute for Basic Research
76135 Karlsruhe, Germany
E-mail: amathias@zkm.de

Florian Grond
ZKM - Center for Art and Media
Institute for Basic Research
76135 Karlsruhe, Germany
E-mail: fgrond@zkm.de

Hans Diebner
ZKM - Center for Art and Media
Institute for Basic Research
76135 Karlsruhe, Germany
E-mail: hdiebner@zkm.de

Detlef Seese
University of Karlsruhe (TH)
Institute AIFB
76128 Karlsruhe, Germany
E-mail: seese@aifb.uni-karlsruhe.de

Ramon Guardans
Alvarez de Castro 12, Madrid 28010
Medialab Madrid, Madrid 28015
Spain
E-mail: ramon.guardans@soundplots.com

Miguel Canela
University of Barcelona
Department of Applied Mathematics and Analysis
Barcelona 08007, Spain
E-Mail: miguelcanela@ub.edu