A general theory on frequency and time-frequency analysis of irregularly sampled time series based on projection methods. I. Frequency analysis

Guillaume Lenoir[1] and Michel Crucifix[1,2]

[1]Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université catholique de Louvain, BE-1348, Louvain-la-Neuve, Belgium
[2]Belgian National Fund of Scientific Research, rue d'Egmont, 5, BE-1000 Brussels, Belgium

Correspondence to: Guillaume Lenoir (guillaume.lenoir@hotmail.com)

Abstract. We develop a general framework for the frequency analysis of irregularly sampled time series. It is based on the Lomb-Scargle periodogram, but extended to algebraic operators accounting for the presence of a polynomial trend in the model for the data, in addition to a periodic component and a background noise. Special care is devoted to the correlation between the trend and the periodic component. This new periodogram is then cast into the Welch overlapping segment averaging (WOSA) method in order to reduce its variance. We also design a test of significance for the WOSA periodogram, against the background noise. The model for the background noise is a stationary Gaussian continuous autoregressive-moving-average (CARMA) process, more general than the classical Gaussian white or red noise processes. CARMA parameters are estimated following a Bayesian framework. We provide algorithms computing the confidence levels for the WOSA periodogram that fully take into account the uncertainty on the CARMA noise parameters. Alternatively, a theory using point estimates of CARMA parameters provides analytical confidence levels for the WOSA periodogram, which are more accurate than Markov chain Monte Carlo (MCMC) confidence levels and, below some threshold for the number of data points, less costly in computing time. We then estimate the amplitude of the periodic component with least squares methods, and derive an approximate proportionality between the squared amplitude and the periodogram. This proportionality leads to a new extension for the periodogram: the weighted WOSA periodogram, that we recommend for most frequency analyses with irregularly sampled data. The estimated signal amplitude also permits filtering in a frequency band. Our results generalize and unify methods developed in the fields of geosciences, engineering, astronomy and astrophysics. They also constitute the starting point for an extension to the continuous wavelet transform developed in a companion article (Lenoir and Crucifix, 2017). All the methods presented in this paper are available to the reader in the Python package WAVEPAL.

1 Introduction

In many areas of geophysics, one has to deal with irregularly sampled time series. However, most of state of the art tools for the frequency analysis are designed to work with regularly sampled data. Classical methods include the discrete Fourier transform (DFT), jointly with the Welch overlapping segment averaging (WOSA) method, developed by Welch (1967), or the multitaper

method, designed in Thomson (1982) and Riedel and Sidorenko (1995). Given the excellent results they provide, it is tempting to interpolate the data and simply apply these techniques. Unfortunately, interpolation may seriously affect the analysis with unpredictable consequences for the scientific interpretation (Mudelsee, 2010, p. 224).

In order to deal with non-interpolated, astronomical, data, Lomb (1976) and Scargle (1982) proposed what is now known as the Lomb-Scargle periodogram (denoted here LS periodogram). The LS periodogram is at the basis of many algorithms proposed in the literature, in particular, in astronomy, e.g. in Mortier et al. (2015), Vio et al. (2010), or Zechmeister and Kürster (2009), and in geophysics, e.g. in Schulz and Stattegger (1997), Schulz and Mudelsee (2002), Mudelsee et al. (2009), Pardo Igúzquiza and Rodríguez Tovar (2012), or Rehfeld et al. (2011). More specifically, in climate and paleoclimate, the time series are often very noisy, exhibit a trend, and potentially carry a wide range of periodic components (e.g. see Fig. 6). Considering all these properties, we design in this work an operator for the frequency analysis generalizing the LS periodogram. The latter was built to analyze data which can be modeled as a periodic component plus noise. Since the periodic component may not necessarily oscillate around zero, Ferraz-Mello (1981) and Heck et al. (1985) extended the LS periodogram, proposing an operator that is suitable to analyze data which can be modeled as a periodic component plus a constant trend plus noise. Their operator is designed to take into account the correlation between the constant trend and the periodic component, and is now a classic tool for analyzing astronomical irregularly spaced time series. In climate and paleoclimate, the periodic component may oscillate around a more complex trend than just a constant. This is why, in this work, we extend the previous result by proposing an operator that is suitable to analyze data which can be modeled as a periodic component plus a polynomial trend plus noise. Our operator is also designed to take into account the correlation between the trend and the periodic component. Our extended LS periodogram is however not sufficient to deal with very noisy data sets, and it also exhibits spectral leakage, like the DFT. In the world of regularly sampled, very noisy, time series, *smoothing* techniques can be applied to reduce the variance of the periodogram, after tapering the time series in order to alleviate spectral leakage (see Harris, 1978). One of them is the WOSA method (Welch, 1967), which consists in segmenting the time series into overlapping segments, tapering them, taking the periodogram on each segment, and finally taking the average of all the periodograms. This technique was transferred to the world of irregularly sampled time series in the work of Schulz and Stattegger (1997), where they apply the classical LS periodogram to each tapered segment, and take the average. In this article, we generalize their work by applying the tapered WOSA method to our extended LS periodogram. Moreover, we show that it is preferable to weight the periodogram of each WOSA segment before taking the average, in order to get a reliable representation of the squared amplitude of the periodic component. This leads us to define the *weighted WOSA periodogram*, that we recommend for most frequency analyses.

The periodogram is often accompanied by a test of significance for the spectral peaks, which relies on the choice of an additive background noise. Two traditional background noises are used in practice. The first one is the Gaussian white noise, which has a flat power spectral density, and which is a common choice with astronomical data sets, e.g. in Scargle (1982) orHeck et al. (1985). The second one is the Gaussian red noise or Ornstein-Uhlenbeck process, for which the power spectral density is a Lorentzian function centered at frequency zero, and which is a common choice with (paleo)climate time series, e.g. in Schulz and Mudelsee (2002) or Ghil et al. (2002). Arguments in favor of a Gaussian red noise as the background stochastic process for climate time series are given in Hasselmann's influential paper (Hasselmann, 1976). Other background noises are also found

in geophysics, often under the form of an autoregressive-moving-average (ARMA) process (see Mudelsee, 2010, p. 60, for an extensive list). In this work, we consider a general class of background noises, which are the continuous autoregressive-moving-average (CARMA) processes, defined in Sect. 3.2. A CARMA(p,q) process is the extension of an ARMA(p,q) process to a continuous time (Brockwell and Davis, 2016, Sect. 11.5). Gaussian white noise and Gaussian red noise are particular cases of a Gaussian CARMA process, i.e. they are a CARMA(0,0) process and a CARMA(1,0) process respectively. Recent advances now allow to accurately estimate the parameters of an irregularly sampled CARMA process from one of its samples (see Kelly et al., 2014).

Estimating the percentiles of the distribution of the weighted WOSA periodogram of an irregularly sampled CARMA process is the core of this paper. This gives the confidence levels for performing tests of significance at every frequency, i.e. test if the null hypothesis - the time series is a purely stochastic CARMA process - can be rejected (with some percentage of confidence) or not. We aim at developing a very general approach. Let us enumerate some key points:

1. Estimation of CARMA parameters is performed in a Bayesian framework and relies on state of the art algorithms provided by Kelly et al. (2014). In the special case of a white noise, we provide an analytical solution.

2. Based on 1, we provide confidence levels computed with Markov chain Monte Carlo (MCMC) methods, that fully take into account the uncertainty on the parameters of the CARMA process, because we work with a *distribution* of values for the CARMA parameters instead of a unique set of values.

3. Alternatively to 2, if we opt for the traditional choice of a unique set of values for the parameters of the CARMA background noise, we develop a theory providing *analytical* confidence levels. Compared to a MCMC-based approach, the analytical method is more accurate and, if the number of data points is not too high, quicker to compute, especially at high confidence levels, e.g. $99\,\%$ or $99.9\,\%$. Computing high levels of confidence is required in some studies, for example in paleoceanography (Kemp, 2016).

4. Confidence levels are provided for any possible choice of the overlapping factor for the WOSA method, extending the traditional $50\,\%$ overlapping choice (Schulz and Stattegger, 1997; Schulz and Mudelsee, 2002).

5. Under the case of a white noise background, without WOSA segmentation and without tapering, we define the *F-periodogram* as an alternative to the periodogram. It has the advantage of not requiring any parameter to be estimated.

Finally, we note that spectral power and estimated squared amplitude are no longer the same thing if the time series is irregularly sampled. Both quantities may be of physical interest. We estimate the amplitude of the periodic component with least squares methods, and derive an approximate proportionality between the squared amplitude and the periodogram, from which we deduce the weights for the weighted WOSA periodogram. The estimated signal amplitude also gives access to filtering in a frequency band.

The paper is organized as follows: In Sect. 2, we introduce the notations and recall some basics of algebra. In Sect. 3, we define the model for the data and write the background noise term into a suitable mathematical form. Section 4 starts with some

reminders about the Lomb-Scargle periodogram and then extends it to take into account the trend, and a second extension deals with the WOSA tapered case. In Sect. 5, we remind that significance testing is nothing but a statistical hypothesis testing. Under the null hypothesis, we estimate the parameters of the CARMA process and estimate the distribution of the WOSA periodogram, either with Monte-Carlo methods or analytically. In the case of a white noise background, we define the F-periodogram as an alternative to the periodogram. Section 6 aims at computing the amplitude of the periodic component of the signal and the difference between the squared amplitude and the periodogram is explained. Sections 7 and 8 are based on the results of Sect. 6. There, we propose a third extension for the LS periodogram and show how to perform filtering. Section 9 presents an example of analysis on a paleoceanographic time series. Finally, a Python package named WAVEPAL is available to the reader and is presented in Sect. 10.

2 Notations and mathematical background

2.1 Notations

Let us introduce the notations for the time series. The measurements $X_1, X_2, ..., X_N$ are done at the times $t_1, t_2, ..., t_N$ respectively, and we assume there is no error on the measurements as well as on the times. They are cast into vectors belonging to $\mathbb{R}^N$:

$$|t\rangle = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} \quad \text{and} \quad |X\rangle = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}. \tag{1}$$

We use here the bra-ket notation, which is common in physics. In $\mathbb{R}^N$, the transpose of $|a\rangle$ is $\langle a|$, i.e. $\langle a|' = |a\rangle$, and in $\mathbb{C}^N$, $\langle a|$ is the conjugate transpose of $|a\rangle$, i.e. $\langle a|^* = |a\rangle$. The inner product of $|a\rangle$ and $|b\rangle$ is $\langle\, a \,|\, b \,\rangle$.

- Let A be a (m, n) matrix and B be a (n, m) matrix. If A is real, A' denotes its transpose, and if A is complex, A^* denotes its conjugate transpose. The trace of AB is denoted by $\mathrm{tr}(AB)$ and we have $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.
- We use the terminology *Gaussian white noise* or simply *white noise* for a (multivariate) Gaussian random variable with constant mean and covariance matrix $\sigma^2\mathbb{I}$.
- $|Z\rangle$ always denotes a standard multivariate Gaussian white noise, i.e.

  $$|Z\rangle \overset{d}{=} \mathcal{N}(0, \mathbb{I}), \tag{2}$$

  where $\overset{d}{=}$ means "is equal in distribution" and $\mathbb{I}$ is the identity matrix.
- A sequence of independent and identically distributed random variables is denoted by "iid".

2.2 Orthogonal projections in $\mathbb{R}^N$

The orthogonal projection on a vector space spanned by the m linearly independent vectors $|a_1\rangle$, ..., $|a_m\rangle$ in $\mathbb{R}^N$ for some $m \in \mathbb{N}_0$ ($m \leq N$) is

$$P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} = V(V'V)^{-1}V', \tag{3}$$

where $\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}$ is the closed span of those m vectors, i.e. the set of all the linear combinations between them. V is a (N,m) matrix defined by

$$V = \begin{pmatrix} | & & | \\ |a_1\rangle & \dots & |a_m\rangle \\ | & & | \end{pmatrix}. \tag{4}$$

Like for any orthogonal projection, we have the following equalities:

$$P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} = P'_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} = P^2_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}}. \tag{5}$$

The m linearly independent vectors $|a_1\rangle$, ..., $|a_m\rangle$ may be orthonormalized by a Gram-Schmidt procedure, leading to m orthonormal vectors $|b_1\rangle$, ..., $|b_m\rangle$, and the orthogonal projection may then be rewritten as

$$P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} = P_{\overline{\text{sp}}\{|b_1\rangle,...,|b_m\rangle\}} = \sum_{k=1}^{m} |b_k\rangle\langle b_k|. \tag{6}$$

Under that form, we see that the above projection has m eigenvalues equal to 1 and $(N-m)$ eigenvalues equal to 0.

Let $|c_1\rangle$, ..., $|c_q\rangle$ be q linearly independent vectors in $\mathbb{R}^N$, with $q \leq m$, and such that $\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\} \subseteq \overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}$. Then $(P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} - P_{\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\}})$ is an orthogonal projection on $\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\} \cap \overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}^{\perp}$, and

$$P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} P_{\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\}} = P_{\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\}} P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} = P_{\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\}}. \tag{7}$$

Moreover, for any vector $|Y\rangle \in \mathbb{R}^N$, we have

$$||(P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}} - P_{\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\}})|Y\rangle||^2 = ||P_{\overline{\text{sp}}\{|a_1\rangle,...,|a_m\rangle\}}|Y\rangle||^2 - ||P_{\overline{\text{sp}}\{|c_1\rangle,...,|c_q\rangle\}}|Y\rangle||^2. \tag{8}$$

We recommend the book of Brockwell and Davis (1991) for more details.

2.3 Quantifying the irregularity of the sampling

The biggest time step for which t_1, ..., t_N, are a subsample of a regularly sampled time series is the greatest common divisor[1] (GCD) of all the time steps of $|t\rangle$. In formulas:

$$\Delta t_{\text{GCD}} = \text{GCD}(\Delta t_1, ..., \Delta t_{N-1}), \tag{9}$$

[1]The GCD is usually defined on the integers, but we can extend it to rational numbers. In practice, t_1, ..., t_N come from measurements with a finite precision and are thus rational numbers.

where

$$\Delta t_k = t_{k+1} - t_k \ \ \forall k \in \{1, ..., N-1\}, \tag{10}$$

and

$$\forall k \in \{1, ..., N\}, \ \exists m \in \mathbb{Z} \text{ s.t. } t_k = m\Delta t_{\text{GCD}}, \tag{11}$$

where $\mathbb{Z}$ denotes the space of integer numbers. Quantifying the irregularity of the sampling is then straightforward. We define

$$r_t = 100\frac{(N-1)\Delta t_{\text{GCD}}}{t_N - t_1}. \tag{12}$$

This ratio is between $0\ \%$ and $100\ \%$, the latter value being reached with regularly sampled time series.

3 The model for the data

3.1 Definition

A suitable and general enough model to analyze the periodicity at frequency $f = \frac{\Omega}{2\pi}$ is:

$$\begin{aligned} |X\rangle &= |\text{Trend}\rangle + E_\omega \cos(\Omega|t\rangle + \phi_\omega) + |\text{Noise}\rangle \\ &= |\text{Trend}\rangle + A_\omega|c_\Omega\rangle + B_\omega|s_\Omega\rangle + |\text{Noise}\rangle, \end{aligned} \tag{13}$$

with $A_\omega = E_\omega \cos(\phi_\omega)$, $B_\omega = -E_\omega \sin(\phi_\omega)$, and $E_\omega^2 = A_\omega^2 + B_\omega^2$. $|c_\Omega\rangle$ and $|s_\Omega\rangle$ are defined componentwise, i.e. $|c_\Omega\rangle = \cos(\Omega|t\rangle) = [\cos(\Omega t_1), ..., \cos(\Omega t_N)]'$ and $|s_\Omega\rangle = \sin(\Omega|t\rangle) = [\sin(\Omega t_1), ..., \sin(\Omega t_N)]'$. We have added the subscript ω to make the difference between the probed frequency, ω, and the data frequency, Ω. Indeed, the periodogram (defined in Sect. 4), the amplitude periodogram (Sect. 6) and the weighted WOSA periodogram (Sect. 7) do not necessarily probe the signal its its true frequency Ω.

3.2 The background noise

3.2.1 Definition of a CARMA process

We follow here the definitions and conventions of Kelly et al. (2014), and technical details can be found in Brockwell and Davis (2016, Sect. 11.5).

The background noise term, $|\text{Noise}\rangle$, considered in this paper is a zero-mean stationary Gaussian continuous autoregressive-moving-average (CARMA) process sampled at the times of $|t\rangle$. As explained in the following, it generalizes traditional background noises used in geophysics.

A CARMA(p,q) process is simply the extension of an ARMA(p,q) process to a continuous time[2]. A zero-mean CARMA(p,q)

[2] A CARMA(p,q) process sampled at the times of an infinite regularly sampled time series is an ARMA(p,q) process.

process $y(t)$ is the solution of the following stochastic differential equation:

$$\frac{\mathrm{d}^p y(t)}{\mathrm{d}t^p} + \alpha_{p-1}\frac{\mathrm{d}^{p-1} y(t)}{\mathrm{d}t^{p-1}} + ... + \alpha_0 y(t) = \beta_q \frac{\mathrm{d}^q \epsilon(t)}{\mathrm{d}t^q} + \beta_{q-1}\frac{\mathrm{d}^{q-1}\epsilon(t)}{\mathrm{d}t^{q-1}} + ... + \epsilon(t), \tag{14}$$

where $\epsilon(t)$ is a continuous-time white noise process with zero mean and variance σ^2. It is defined from the standard Brownian motion $B(t)$ through the following formula:

$$\sigma \mathrm{d}B(t) = \epsilon(t)\mathrm{d}t \tag{15}$$

The parameters α_0, ... , α_{p-1} are the autoregressive coefficients, and the parameters β_1, ..., β_q are the moving average coefficients. $\alpha_p = \beta_0 = 1$ by definition. When $p > 0$, the process is stationary only if $q < p$ and the roots $r_1, ..., r_p$ of

$$\sum_{k=0}^{p} \alpha_k z^k = 0, \tag{16}$$

have negative real parts. Strictly speaking, the derivatives of the Brownian motion $\frac{\mathrm{d}^k B}{\mathrm{d}t}$, $k > 0$, do not exist, and we therefore interpret Eq. (14) as being equivalent to the following measurement and state equations

$$y(t) = \langle\, b \,|\, w(t) \,\rangle, \tag{17}$$

and

$$\mathrm{d}|w(t)\rangle = A|w(t)\rangle \mathrm{d}t + \mathrm{d}B(t)|e\rangle, \tag{18}$$

where $|b\rangle = [\beta_0, \beta_1, ..., \beta_q, 0, ..., 0]'$ is a vector of length p, $|e\rangle = [0, 0, ..., 0, \sigma]'$, and

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \dots & -\alpha_{p-1} \end{pmatrix}. \tag{19}$$

Equation (18) is nothing else but an Itô differential equation for the state vector $|w(t)\rangle$.

In practice, only CARMA processes of low order are useful in our framework, typically, (p,q)=$(0,0)$, $(1,0)$, $(2,0)$, $(2,1)$, since at higher order, they often exhibit dominant spectral peaks (see Kelly et al., 2014), which is not what we want as a model for the spectral background. Indeed, on the basis of our model, Eq. (13), it is desirable that the spectral peaks come from the deterministic cosine and sine components. We now consider two useful particular cases of a CARMA process before analyzing the general case.

3.2.2 Gaussian white noise

When $p = 0$ and $q = 0$, the process reduces to a white noise, normally distributed with zero-mean and variance σ^2. The $|\mathrm{Noise}\rangle$ term in Eq. (13) is then simply

$$|\mathrm{Noise}\rangle = \sigma|Z\rangle = K|Z\rangle, \tag{20}$$

with $K = \sigma\mathbb{I}$.

3.2.3 Gaussian red noise

When $p = 1$ and $q = 0$, the CARMA(1,0) or CAR(1) process is an Ornstein-Uhlenbeck process or red noise (Uhlenbeck and Ornstein, 1930), which is quite of interest in geophysical and other applications (Mudelsee, 2010). Since we work with a discrete time series, it is necessary to find the solution of Eq. (14) at t_1, ..., t_N. This is done by integrating that equation between consecutive times, i.e. from t_{i-1} to t_i $\forall i \in \{2, ..., N\}$. The components of the $|\text{Noise}\rangle$ vector are then:

$$y(t_1) \stackrel{d}{=} \mathcal{N}(0, \frac{\sigma^2}{2\alpha}),$$
$$y(t_i) = \rho_i y(t_{i-1}) + \eta_i \qquad \forall i \in \{2, ..., N\}, \tag{21}$$

where

$$\rho_i = \exp(-\alpha(t_i - t_{i-1})) \quad \text{and} \quad \eta_i \stackrel{d}{=} \mathcal{N}(0, \frac{\sigma^2}{2\alpha}(1 - \rho_i^2)). \tag{22}$$

See Robinson (1977) and Brockwell and Davis (2016, p. 343) for more details. The requirement on stationarity, Eq. (16), imposes $\alpha > 0$. The generated time series has a constant mean equal to zero and a constant variance equal to $\frac{\sigma^2}{2\alpha}$. The $|\text{Noise}\rangle$ term in Eq. (13) can also be written under a matrix form:

$$|\text{Noise}\rangle = K|Z\rangle, \tag{23}$$

where K is a (N,N) lower triangular matrix whose elements are

$$K_{i,j} = \sqrt{\frac{\sigma^2}{2\alpha}}\sqrt{1 - \rho_j^2}\exp(-\alpha(t_i - t_j)) \quad \forall j \leq i, \tag{24}$$

where we define $\rho_1 = 0$. This matrix form is used in Sect. 5.3.3.

Note that, if the time series is regularly sampled, ρ is a constant and Eq. (21) becomes the equation of a finite-length AR(1) process, which is stationary since $\alpha > 0$ implies $\rho < 1$.

3.2.4 The general Gaussian CARMA noise

The solution of Eq. (14) at the time t_n ($n = 2, ..., N$), that we denote by y_n, is

$$y_n = \langle\, b \,|\, w_n \,\rangle,$$
$$\text{where } |w_n\rangle = \exp(A(t_n - t_{n-1}))|w_{n-1}\rangle + |\eta_n\rangle, \tag{25}$$

$|\eta_n\rangle$ follows a multivariate normal distribution with zero mean and covariance matrix C_n given by

$$C_n = \int_0^{t_n - t_{n-1}} \mathrm{d}t \exp(At)|e\rangle\langle e|\exp(A't), \tag{26}$$

The above formula requires the computation of matrix exponentials and numerical integration. This can be avoided by diagonalizing matrix A, with $A = UDU^{-1}$. D is a diagonal matrix with diagonal elements given by the roots of Eq. (16):

$$D_{kk} = r_k \quad \forall k \in 1, ..., p, \tag{27}$$

and U is a Vandermonde matrix, with

$$U_{lk} = r_k^{l-1} \quad \forall l, k \in 1, ..., p. \tag{28}$$

Now, by defining $|\tilde{w}_n\rangle = U^{-1}|w_n\rangle$, we get

$$y_n = \langle b|U|\tilde{w}_n\rangle, \tag{29a}$$

$$\text{where } |\tilde{w}_n\rangle = \Lambda_n|\tilde{w}_{n-1}\rangle + |\tilde{\eta}_n\rangle. \tag{29b}$$

The matrix exponential $\exp(A(t_n - t_{n-1}))$ has been transformed into $\Lambda_n = U^{-1}\exp(A(t_n - t_{n-1}))U$ which is simply a diagonal matrix with elements $\Lambda_{n_{kk}} = \exp(r_k(t_n - t_{n-1}))$. The covariance matrix of $|\tilde{\eta}_n\rangle$, that we write Σ_n, also takes a relatively simple form:

$$\Sigma_{n_{kl}} = -\sigma^2 \frac{\kappa_k \kappa_l^*}{(r_k + r_l^*)}(1 - \exp((r_k + r_l^*)(t_n - t_{n-1}))) \quad \forall k, l \in \{1, ..., p\}, \tag{30}$$

which is a Hermitian matrix, and where $|\kappa\rangle$ is the last column of U^{-1}. The initial condition y_1 is determined by imposing stationarity, which is fulfilled only if $|w_1\rangle$ has a zero mean and a covariance matrix V whose elements are

$$V_{kl} = -\sigma^2 \sum_{m=1}^{p} \frac{r_m^{k-1}(-r_m)^{l-1}}{2\text{Re}\{r_m\}\prod_{s=1, s\neq m}^{p}(r_s - r_m)(r_s^* + r_m)} \quad \forall k, l \in \{1, ..., p\}. \tag{31}$$

Stationarity implies that the process $y(t)$ has a zero mean and variance $\langle b|V|b\rangle$ $\forall t$. All the above formulas and how to get them can be found in Kelly et al. (2014), Jones and Ackerson (1990) and Brockwell and Davis (2016, Sect. 11.5.2).

Generation of a CARMA(p, q) process can be performed with the Kalman filter since Eq. (29b) and (29a) are nothing but the state and measurement equations respectively (see Kelly et al., 2014, for more details). Alternatively, $|y\rangle$ can be written under a matrix form as in Eq. (23). Matrix formalism is useful in Sect. 5.3.3. Let us start with Eq. (29b):

$$|\tilde{w}_n\rangle = \Lambda_n|\tilde{w}_{n-1}\rangle + U^{-1}|\eta_n\rangle. \tag{32}$$

The covariance matrix of $|\eta_n\rangle$, $C_n = U\Sigma U^*$, is of course real symmetric and positive semi-definite. We thus have the following Schur decomposition:

$$C_n = Q_n Q_n', \tag{33}$$

where Q_n is a real matrix. Consequently,

$$\begin{aligned}|\tilde{w}_n\rangle &= \Lambda_n|\tilde{w}_{n-1}\rangle + U^{-1}Q_n|\epsilon_n\rangle \\ &= \Lambda_n\Lambda_{n-1}|\tilde{w}_{n-2}\rangle + \Lambda_n U^{-1}Q_{n-1}|\epsilon_{n-1}\rangle + U^{-1}Q_n|\epsilon_n\rangle \\ &= ... \\ &= \sum_{i=2}^{n}(\prod_{l=i+1}^{n}\Lambda_l)U^{-1}Q_i|\epsilon_i\rangle + \prod_{l=2}^{n}\Lambda_l|\tilde{w}_1\rangle, \end{aligned} \tag{34}$$

where $|\epsilon_1\rangle$, ..., $|\epsilon_n\rangle$ are iid standard Gaussian white noises. The product of the Λ's can be simplified. Its diagonal elements are:

$$(Y_{in})_{jj} := (\prod_{l=i+1}^{n}\Lambda_l)_{jj} = \exp(r_j(t_n - t_i)). \tag{35}$$

As stated above, $|w_1\rangle$ follows a multivariate normal distribution with zero mean and covariance matrix V. We can use again the Schur decomposition to write $V = WW'$, where W is a real matrix, yielding

$$\begin{aligned}|\tilde{w}_n\rangle &= \sum_{i=2}^{n}Y_{in}U^{-1}Q_i|\epsilon_i\rangle + Y_{1n}U^{-1}W|\epsilon_1\rangle \\ &= \sum_{i=1}^{n}P_{in}|\epsilon_i\rangle, \end{aligned} \tag{36}$$

with $P_{1n} = Y_{1n}U^{-1}W$ and $P_{in} = Y_{in}U^{-1}Q_i$ for $i > 1$. The CARMA process at time t_n is then given by

$$\begin{aligned}y_n &= \langle b|U|\tilde{w}_n\rangle \\ &= \sum_{i=1}^{n}\langle b|U|P_{in}|\epsilon_i\rangle. \end{aligned} \tag{37}$$

Finally, the $|\text{Noise}\rangle$ term in Eq. (13) is

$$|\text{Noise}\rangle = |y\rangle = \begin{pmatrix} \langle b|U|P_{11} & \langle 0| & ... & ... & \langle 0| \\ \langle b|U|P_{12} & \langle b|U|P_{22} & \langle 0| & ... & \langle 0| \\ & & \ddots & & \\ & & & \ddots & \\ \langle b|U|P_{1N} & \langle b|U|P_{2N} & ... & ... & \langle b|U|P_{NN} \end{pmatrix} \begin{pmatrix} |\epsilon_1\rangle \\ |\epsilon_2\rangle \\ \vdots \\ |\epsilon_N\rangle \end{pmatrix} = K|Z\rangle, \tag{38}$$

where K is a $(N, N \times p)$ real matrix and $|Z\rangle$ has a length $N \times p$. Matrix K is triangular if $p = 1$, which is the particular case treated in Sect. 3.2.3.

3.3 The trend

The model for the trend must be as general as possible and compatible with a formalism based on orthogonal projections (see Sect. 4). This is the reason why we choose a polynomial trend of some degree m:

$$|\text{Trend}\rangle = \sum_{k=0}^{m}\gamma_k|t^k\rangle, \text{ where } |t^k\rangle = [t_1^k, ..., t_N^k]', \tag{39}$$

where $|t^k\rangle$ is defined componentwise, i.e. $|t^k\rangle = [t_1^k, ..., t_N^k]'$. Considering or not the presence of a trend in the model for the data is left to the user, given that we can always interpret a polynomial trend of low order as a very low-frequency oscillation.

4 Periodogram & relatives

4.1 Lomb-Scargle periodogram

Consider the orthogonal projection of the data $|X\rangle$ onto the vector space spanned by the vectors cosine and sine, defined by $|c_\omega\rangle = \cos(\omega|t\rangle)$ and $|s_\omega\rangle = \sin(\omega|t\rangle)$. The periodogram at the frequency $f = \frac{\omega}{2\pi}$ is defined as the squared norm of that projection:

$$||P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}|X\rangle||^2. \tag{40}$$

When the time series is regularly sampled with a constant time step Δt, and if we only consider the Fourier angular frequencies, $\omega_k = \frac{2\pi k}{N\Delta t}$ ($k = 0, ..., N-1$), the periodogram defined above is equal to the squared modulus of the discrete Fourier transform (DFT) of real signals.

Now, rescale $|c_\omega\rangle$ and $|s_\omega\rangle$ such that they are orthonormal. This can be done by defining

$$|c_\omega^\sharp\rangle = \frac{\cos(\omega|t\rangle - \beta_\omega)}{\sqrt{\Sigma_{i=1}^N \cos^2(\omega t_i - \beta_\omega)}}, \qquad |s_\omega^\sharp\rangle = \frac{\sin(\omega|t\rangle - \beta_\omega)}{\sqrt{\Sigma_{i=1}^N \sin^2(\omega t_i - \beta_\omega)}}, \tag{41}$$

where β_ω is the solution of

$$\tan(2\beta_\omega) = \frac{\Sigma_{i=1}^N \sin(2\omega t_i)}{\Sigma_{i=1}^N \cos(2\omega t_i)}. \tag{42}$$

The spanned vector space naturally remains unchanged (see Fig. 1). These formulas are nothing but the Lomb-Scargle formulas (Scargle, 1982, Eq. (10)). The periodogram is now

$$||P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}|X\rangle||^2 = \langle\, c_\omega^\sharp \,|\, X\,\rangle^2 + \langle\, s_\omega^\sharp \,|\, X\,\rangle^2. \tag{43}$$

Note that, for any signal $|X\rangle \in \mathbb{R}^N$,

$$0 \leq \frac{||P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}|X\rangle||^2}{\langle\, X \,|\, X\,\rangle} \leq 1, \tag{44}$$

and this is equal to 1 if $|X\rangle = A|c_\omega\rangle + B|s_\omega\rangle$.

Some properties of the LS periodogram are presented in appendix A. Here and for the rest of the article, the frequency $f = \omega/2\pi$ is considered as a continuous variable.

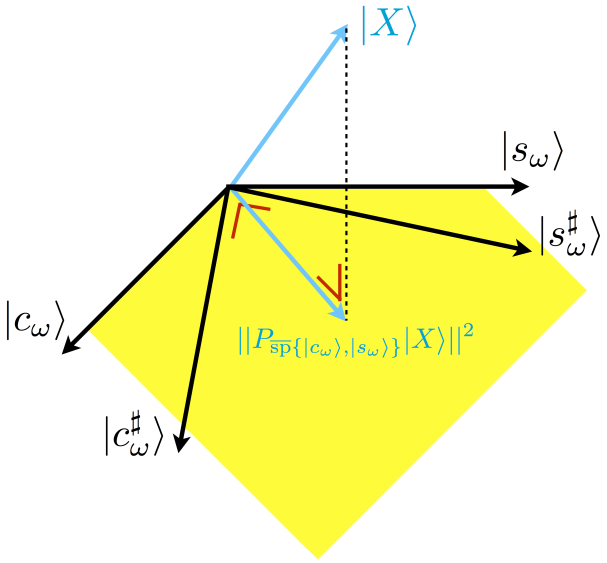


Figure 1. Schematic view of the linear rescaling in $\mathbb{R}^N$ leading to the Lomb-Scargle formulas. In yellow is drawn a subset of $\overline{\mathrm{sp}}\{|c_\omega\rangle, |s_\omega\rangle\}$. A span is invariant under linear combinations of its vectors. The dashed line corresponds to the minimal euclidean distance between the data $|X\rangle$ and $\overline{\mathrm{sp}}\{|c_\omega\rangle, |s_\omega\rangle\}$.

4.2 Periodogram and mean

The LS periodogram applies well to data which can be modeled as

$$|X\rangle = A_\omega|c_\Omega\rangle + B_\omega|s_\Omega\rangle + |\text{Noise}\rangle. \tag{45}$$

However, the periodic components may not necessarily oscillate around zero, and a better model is

$$|X\rangle = \mu|t^0\rangle + A_\omega|c_\Omega\rangle + B_\omega|s_\Omega\rangle + |\text{Noise}\rangle, \tag{46}$$

where $|t^0\rangle = [1, 1, ..., 1]'$. Substracting the average of the data is then often done before applying the LS periodogram. But that mere operation implicitly assumes that $\langle t^0 \,|\, c_\Omega \rangle = \langle t^0 \,|\, s_\Omega \rangle = 0$, which is not necessarily the case. In other words, the data average is not necessarily equal to μ, the process mean. Fig. 2a illustrates that fact. Note that this discrepancy occurs in regularly sampled data as well, at non-Fourier frequencies, i.e. when $N\Delta t$ is not a multiple of the probing period. See Fig. 2b.

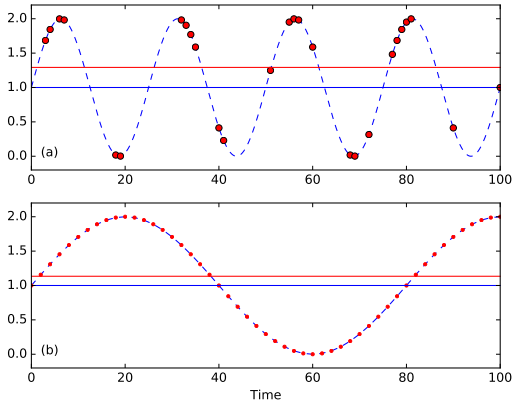


Figure 2. Signal average and sampling. (a) The continuous signal is in dashed blue and it is irregularly sampled at red dots. The continuous signal oscillates around 1 (blue line), which does not correspond to the average of the sampled signal (red line). (b) Same as (a) with a regularly sampled signal.

In order to deal with the mean in a suitable way, we define the periodogram as

$$||(P_{\overline{\text{sp}}\{|t^0,c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}})|X\rangle||^2. \tag{47}$$

Formula (47) is taken from Brockwell and Davis (1991), Ferraz-Mello (1981) or Heck et al. (1985); equivalence between them is shown in appendix B. $[P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}}]$ is also an orthogonal projection. A simple example will justify the principle. Consider the following purely deterministic mono-periodic signal with N data points:

$$|Y\rangle = \mu|t^0\rangle + A|c_\omega\rangle + B|s_\omega\rangle = V_3|\Phi\rangle, \tag{48}$$

with

$$\begin{pmatrix} | & | & | \\ |t^0\rangle & |c_\omega\rangle & |s_\omega\rangle \\ | & | & | \end{pmatrix}, \tag{49}$$

and

$$|\Phi\rangle = \begin{pmatrix} \mu \\ A \\ B \end{pmatrix}. \tag{50}$$

The projection at ω is

$$\begin{aligned}(P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}})|Y\rangle &= (\mathbb{I} - P_{\overline{\text{sp}}\{|t^0\rangle\}})P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}}|Y\rangle \\ &= (\mathbb{I} - P_{\overline{\text{sp}}\{|t^0\rangle\}})V_3|\Phi\rangle \\ &= |Y\rangle - P_{\overline{\text{sp}}\{|t^0\rangle\}}|Y\rangle \\ &= A|c_\omega\rangle + B|s_\omega\rangle - \frac{\langle\, t^0 \,|\, c_\omega \,\rangle}{\langle\, t^0 \,|\, t^0 \,\rangle}A|t^0\rangle - \frac{\langle\, t^0 \,|\, s_\omega \,\rangle}{\langle\, t^0 \,|\, t^0 \,\rangle}B|t^0\rangle. \end{aligned} \tag{51}$$

We see that it is invariant with respect to μ, and we find back the signal minus its average. We thus have

$$||(P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}})|Y\rangle||^2 = N\ \text{Var}(|Y\rangle), \tag{52}$$

where $\text{Var}(|Y\rangle) = \left(\sum_{i=1}^N Y_i^2\right)/N - \left(\sum_{i=1}^N Y_i\right)^2/N^2$. This is a result similar to what we get with regularly sampled data and the DFT[3].

Now, we do a Gram-Schmidt orthonormalization like in Ferraz-Mello (1981), in order to simplify formula (47). To this end, we define the three orthonormal vectors $|h_0\rangle = |t^0\rangle/|||t^0\rangle||$, $|h_1\rangle$ and $|h_2\rangle$ satisfying

$$\overline{\text{sp}}\{|t^0\rangle, |c_\omega\rangle, |s_\omega\rangle\} = \overline{\text{sp}}\{|h_0\rangle, |h_1\rangle, |h_2\rangle\}. \tag{53}$$

Consequently,

$$P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}} = |h_1\rangle\langle h_1| + |h_2\rangle\langle h_2|, \tag{54}$$

and

$$||(P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}})|X\rangle||^2 = \langle\, h_1 \,|\, X \,\rangle^2 + \langle\, h_2 \,|\, X \,\rangle^2. \tag{55}$$

Note that, for any signal $|X\rangle \in \mathbb{R}^N$, we have

$$0 \leq \frac{||(P_{\overline{\text{sp}}\{|t^0\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}})|X\rangle||^2}{N\ \text{Var}(|X\rangle)} \leq 1, \tag{56}$$

and this is equal to 1 for a signal given by $|X\rangle = \mu|t^0\rangle + A|c_\omega\rangle + B|s_\omega\rangle$.

4.3 Periodogram and a polynomial trend

If we want to work with the full model, Eq. (13), which has a polynomial trend of degree m, we can naturally extend the result of Sect. 4.2 and work with

$$||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle\}})|X\rangle||^2 = \langle\, h_{m+1} \,|\, X \,\rangle^2 + \langle\, h_{m+2} \,|\, X \,\rangle^2, \tag{57}$$

[3] If we have $|Y\rangle = \mu|t^0\rangle + A|e_\omega\rangle$, where $|e_\omega\rangle = \exp(i2\pi\omega|t\rangle)$ and ω being a Fourier frequency, then $||DFT_\omega(|Y\rangle)||^2 = ||P_{\overline{\text{sp}}\{|e_\omega\rangle\}}|Y\rangle||^2 = N\,||A||^2 = N\ \text{Var}(|Y\rangle)$. Var is here the biased variance, which is defined as the squared norm of the signal minus its average value, and divided by N.

where $|h_{m+1}\rangle$ and $|h_{m+2}\rangle$ are determined from a Gram-Schmidt orthonormalization starting with the orthonormalization of $|t^0\rangle$, ..., $|t^m\rangle$.

It may happen that, for large m, the correlation matrix in the formula of orthogonal projection be singular. In that case, two options, less optimal, are possible: reduce the degree m, or perform the detrending before the spectral analysis, for example with a moving average.

Similarly to Sect. 4.2, we have, for any signal $|X\rangle \in \mathbb{R}^N$,

$$0 \leq \frac{||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle\}})|X\rangle||^2}{|||X\rangle - P_{\overline{\text{sp}}\{|t^0\rangle,...,|t^m\rangle\}}|X\rangle||^2} \leq 1, \tag{58}$$

and this is equal to 1 for a signal given by $|X\rangle = \sum_{k=0}^{m} \gamma_k |t^k\rangle + A|c_\omega\rangle + B|s_\omega\rangle$. Finally, we have a result similar to Eq. (51), in the sense that the projection given in Eq. (57) is invariant with respect to the parameters of the trend (but it naturally depends on the choice of the *degree* m).

4.4 Tapering the periodogram

A finite length signal can be seen as an infinite length signal multiplied by a rectangular window. This implies, among others, that a mono-periodic signal will have a periodogram characterized by a peak of finite width, with possibly large sidelobes, instead of a Dirac delta function. This is called *spectral leakage*.

The phenomenon has been deeply studied in the case of regularly sampled data. Leakage may be controlled by choosing alternatives to the default rectangular window. This is called *windowing* or *tapering* (see Harris, 1978, for an extensive list of windows). They all share the property to vanish at the borders of the time series.

In the case of irregularly sampled data, building windows for controlling the leakage is a much more challenging task. Even with the default rectangular window, leakage is very irregular, data and frequency dependent, due to the long-range correlations in frequency between the vectors on which we do the projection. To our knowledge, no general and stable solution for that issue is available in the literature. We thus recommend to use the default rectangular window, i.e. do no tapering, if r_t, defined in Eq. (12), is small, and to use simple windows, like the sin^2 or the Gaussian window, for moderately irregularly sampled data (r_t greater than $80\ \%$ or $90\ \%$). With tapering, formula (57) becomes

$$||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|Gc_\omega\rangle,|Gs_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle\}})|X\rangle||^2, \tag{59}$$

where G is a frequency-independent diagonal matrix, which is used to weight the sine and cosine vectors. For example, with a $\sin^2$ window, also called Hanning window, we have

$$G_{kk} = \sin^2\left(\frac{\pi\,(t_k - t_1)}{t_N - t_1}\right) \qquad \forall k \in \{1, ..., N\}. \tag{60}$$

4.5 Smoothing the periodogram with the WOSA method

4.5.1 The consistency problem

Besides spectral leakage, another issue with the periodogram is consistency. Indeed, for regularly sampled time series, the periodogram is known not to be a consistent estimator of the true spectrum as the number of data points tends to infinity (see Brockwell and Davis, 1991, Chap. 10). Another view of the problem is that the periodogram remains very noisy whatever the number of data points we have at our disposal. Smoothing procedures are therefore applied to reduce the variance of the periodogram. The drawback of any smoothing procedure is naturally a decrease of the frequency resolution. Among the smoothing methods available in the literature, two are traditionally used: multitaper methods (MTM), developed by Thomson (1982) and Riedel and Sidorenko (1995), and the Welch overlapping segment averaging (WOSA) method (Welch, 1967). See Walden (2000) for a unified view.

Multitaper methods are certainly not generalizable to the case of irregularly sampled data, except in very specific cases that are not of interest in geophysics, like in Bronez (1988), which deals with band-limited signals, useful in the field of the telecommunications, or Fodor and Stark (2000), which considers regularly sampled time series with some gaps, useful for time series with a ratio r_t, defined in Eq. (12), close to 100. We will then use the WOSA method applied to the LS periodogram, like in Schulz and Stattegger (1997) and Schulz and Mudelsee (2002), or to its relatives (formulas (47), (57), or the most general (59)).

4.5.2 Principle of the WOSA method

Trendless time series

The time series is divided into overlapping segments. The tapered LS periodogram is computed on every segment, and the WOSA periodogram is the average of all these tapered periodograms. This technique relies on the fact that the signal is stationary, as always in spectral analysis[4]. The length of the segments and the overlapping factor need to be chosen depending on how much we want to reduce the variance of the noise. As a general rule, shortening the segments will decrease the frequency resolution. Consequently, there is always a trade-off between the frequency resolution and the variance reduction.

For regularly sampled data, each segment of fixed length has the same number of data points. In the irregularly sampled case, it is not anymore the case and we have two options.

1. Take segments with a fixed number of points and thus a variable length. In the non-tapered case, the periodogram on each segment provides deterministic peaks (coming from the deterministic sin/cos components) with more or less the same height. But variable length segments will give deterministic peaks of variable width.

2. Take segments of fixed length but with a variable number of data points. The periodogram on each segment provides deterministic peaks with more or less the same width, except if there is a big gap at the beginning or at the end of the

[4]Basically, the *spectrum* cannot be defined without that hypothesis. See the Wiener-Khinchin theorem in e.g. Priestley (1981, Chap. 4)

segment, such that its effective length is reduced. But they will have variable height since the number of data points is not constant.

We judge it is better to have peaks with similar width on each segment when averaging the periodograms in a frequency band. Consequently, we recommend the second option. An example of WOSA segmentation is shown on Fig. 8a.

Time series with a trend

The only difference with the previous case is that, for each segment, we consider the projection on $|t^0\rangle$, ..., $|t^m\rangle$ jointly with the tapered cosine and sine components. Formula (59) is applied to each segment with $|Gc_\omega\rangle$ and $|Gs_\omega\rangle$ localized on the WOSA segment, but $|t^0\rangle$, ..., $|t^m\rangle$ are taken on the full length of the time series, because the trend is the one of the whole time series.

4.5.3 The WOSA periodogram in formulas

Two parameters are required: the length of WOSA segments, D, and the overlapping factor, $\beta \in [0, 1[$. $\beta = 0$ when there is no overlap. We denote by Q the number of WOSA segments, which is equal to

$$Q = \left\lfloor \frac{t_N - t_1 - D}{(1-\beta)D} \right\rfloor + 1, \tag{61}$$

where $\lfloor\rfloor$ is the floor function. Because of the rounding, D must be adjusted afterwards:

$$D = \frac{t_N - t_1}{1 + (1-\beta)(Q-1)}. \tag{62}$$

Define τ_q to be the starting time of the q^{th} segment ($q \in \{1, ..., Q\}$). Note that τ_q is not necessarily equal to one of the components of $|t\rangle$. It follows that

$$\tau_q = t_1 + (1-\beta)(q-1)D \quad q = 1, ..., Q. \tag{63}$$

The WOSA periodogram is then

$$\begin{aligned} ||P_{\text{WOSA}}(\omega)|X\rangle||^2 &= \frac{1}{Q}\sum_{q=1}^{Q} ||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|G_q c_{\omega,q}\rangle,|G_q s_{\omega,q}\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle\}})|X\rangle||^2 \\ &= \frac{1}{Q}\sum_{q=1}^{Q} \langle X|(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|G_q c_{\omega,q}\rangle,|G_q s_{\omega,q}\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle\}})|X\rangle. \end{aligned} \tag{64}$$

Note that the sum of these orthogonal projections is not anymore an orthogonal projection. $|G_q c_{\omega,q}\rangle$ and $|G_q s_{\omega,q}\rangle$ are the tapered cosine and sine on the q^{th} segment. For example, with the Hanning ($\sin^2$) window,

$$(G_q c_{\omega,q})_k = g_q(t_k)\cos(\omega(t_k - \tau_q)), \qquad (G_q s_{\omega,q})_k = g_q(t_k)\sin(\omega(t_k - \tau_q)), \tag{65}$$

where

$$g_q(t_k) = \begin{cases} \sin^2\left(\frac{\pi\,(t_k-\tau_q)}{D}\right) & \text{if } 0 \le (t_k-\tau_q) \le D, \\ 0 & \text{otherwise.} \end{cases} \tag{66}$$

It may be shown that $\overline{\text{sp}}\{|t^0\rangle, |t^2\rangle, ..., |t^m\rangle, |G_q c_{\omega,q}\rangle, |G_q s_{\omega,q}\rangle\}$ is invariant with the variable τ_q appearing in the cosine and sine terms, so that we can impose $\tau_q = 0\ \forall q$ inside the cosine and sine terms.

In formula (64), for each orthogonal projection, we apply a Gram-Schmidt orthonormalization (similarly to Sect. 4.3):

$$||P_{\text{WOSA}}(\omega)|X\rangle||^2 = \frac{1}{Q}\sum_{q=1}^{Q}(\langle X\,|\,h_{1,q}(\omega)\rangle\langle h_{1,q}(\omega)\,|\,X\rangle + \langle X\,|\,h_{2,q}(\omega)\rangle\langle h_{2,q}(\omega)\,|\,X\rangle), \tag{67}$$

where, for each q, $|h_{1,q}(\omega)\rangle$ and $h_{2,q}(\omega)\rangle$ are orthonormal. We are now able to write the WOSA periodogram under a simple matrix form:

$$||P_{\text{WOSA}}(\omega)|X\rangle||^2 = \langle X|M_\omega M'_\omega|X\rangle, \tag{68}$$

where

$$M_\omega = \frac{1}{\sqrt{Q}}\begin{pmatrix} | & | & & | & | \\ |h_{1,1}(\omega)\rangle & |h_{2,1}(\omega)\rangle & \dots & |h_{1,Q}(\omega)\rangle & |h_{2,Q}(\omega)\rangle \\ | & | & & | & | \end{pmatrix}. \tag{69}$$

4.5.4 Practical considerations

First, note that the Gram-Schmidt orthonormalization process requires at least $m+3$ data points. WOSA segments with less than $m+3$ points must therefore be ignored in the average of the periodograms.

Second, as we want to get deterministic peaks with more or less the same width on every segment, a WOSA segment is kept in the average if the data cover some percentage of its length D, namely,

$$q^{th} \text{ segment kept if: } 100\frac{t_{q,2}-t_{q,1}}{D} \ge C, \tag{70}$$

where $t_{q,1}$ and $t_{q,2}$ are the times of the first and last data points inside in the q^{th} segment, and C is the coverage factor. Its default value in WAVEPAL is $90\ \%$.

Third, the frequency range on the q^{th} segment is bounded by these two frequencies:

$$f_{\text{min}} = \frac{1}{t_{q,2}-t_{q,1}} \quad \text{and} \quad f_{\text{max}} = \frac{1}{2\overline{\Delta t}_q}. \tag{71}$$

The maximal period ($1/f_{\text{min}}$) corresponds to the effective length on the segment. The maximal frequency in the case of regularly sampled data must be the Nyquist frequency, $f_{\text{max}} = 1/2\Delta t$. For irregularly sampled data, different choices for $\overline{\Delta t}_q$

are possible. As suggested in appendix A, an option is $\overline{\Delta t}_q = \Delta t_{\text{GCD,q}}$, but this choice is insufficient to avoid *pseudo-aliasing* issues. Imagine for example a regularly sampled time series with 1000 data points and $\Delta t = 1$. Add one point at the end with the last time step being 0.1. The resulting irregularly sampled time series will thus have $\Delta t_{\text{GCD}} = 0.1$. If we take $f_{\text{max}} = 5$, it is obvious that some kind of aliasing will occur between $f = 0.5$ and f_{max}. This it what we call *pseudo-aliasing*. A much better choice in this case is of course $f_{\text{max}} = 0.5$. Section 5 of Bretthorst (2001) provides further discussions on this topic.

In practice,

$$\overline{\Delta t}_q = \max\left\{ \frac{\sum_{k=1}^{N} G_{q_{k,k}} \Delta t_{c_k}}{\text{tr}(G_q)}, \frac{\sum_{k=1}^{N-1} H_{q_{k,k}} \Delta t_k}{\text{tr}(H_q)} \right\}, \tag{72}$$

where

$$\begin{aligned}
&\Delta t_k = t_{k+1} - t_k \quad \forall k \in \{1, ...N-1\},\\
&\Delta t_{c_k} = \frac{t_{k+1} - t_{k-1}}{2} \quad \forall k \in \{2, ...N-1\},\\
&\Delta t_{c_1} = t_2 - t_1,\\
&\Delta t_{c_N} = t_N - t_{N-1},
\end{aligned} \tag{73}$$

and H_q is a diagonal matrix with

$$H_{q_{k,k}} = \text{taper at time } \frac{t_k + t_{k+1}}{2},\ k \in \{1, ..., N-1\}, \tag{74}$$

appears to work well. More justification and an example are provided in part II of this study (Lenoir and Crucifix, 2017, Sect. 3.8), where it is shown that such a formula can handle aliasing issues in the case of time series with large gaps. Matrix H_q is similar to matrix G_q, defined in Sect. 4.4, but with elements taken at $(t_k + t_{k+1})/2$ instead of t_k. Quantity $\overline{\Delta t}_q$ is equal to the maximum between the average time step and the average central time step if there is no tapering ($G_q = H_q = \mathbb{I}$), and is equal to Δt in the regularly sampled case. These restrictions on the frequency bounds imply that the total number of WOSA segments, Q, in formula (64), is not the same for all the frequencies. This is illustrated on Fig. 8b.

Fourth, in order the have a reliable average of the periodograms, we only represent the periodogram at the frequencies for which the number of WOSA segments is above some threshold. In WAVEPAL, default value for the threshold at frequency f is

$$\text{Threshold: } \min\{10, \max_{\{f\}} Q(f)\}, \tag{75}$$

where $Q(f)$ is the number of WOSA segments at frequency f. It means that frequency f belongs to the range of frequencies of the WOSA periodogram if $Q(f)$ is greater than or equal to the threshold.

5 Significance testing with the periodogram

5.1 Hypothesis testing

Significance testing allows us to test for the presence of periodic components in the signal. It is mathematically expressed as a hypothesis testing (see Brockwell and Davis, 1991, Chap. 10). Taking our model, Eq. (13), the null hypothesis is

$$H_0 : A_\omega = B_\omega = 0. \tag{76}$$

Therefore, $|X\rangle = |\text{Trend}\rangle + |\text{Noise}\rangle$. The alternative hypothesis is

$$H_1 : A_\omega \text{ and } B_\omega \text{ are not both zero.} \tag{77}$$

The decision of accepting or rejecting the null hypothesis is based on the periodogram evaluated at ω, whose general formula is given in Eq. (64). The test is performed independently for each frequency (*pointwise testing*). Concretely, for each frequency, we compute the distribution of the periodogram under the null hypothesis, and then see if the *data* periodogram at that frequency is above or below a given percentile (e.g. the 95^{th}) of that distribution. The percentile is called *level of confidence*. If the data periodogram is above the X^{th} percentile of the reference distribution, we reject the null hypothesis with X % of confidence. The *level of significance* is equal to $(100 - X)$ %, e.g. a 95 % confidence level is equivalent to a 5 % significance level. Hypothesis testing is, for this reason, often called *significance testing*. See Fig. 8c and 8d for an illustration on paleoclimate data. We recommend Priestley (1981, Chap. 6) for more details on the methodology.

To perform significance testing, we thus need

1. to estimate the parameters of the process under the null hypothesis. This is studied in Sect. 5.2.
2. to estimate the distribution of the periodogram under the null hypothesis. This is studied in Sect. 5.3.

5.2 Estimation of the parameters under the null hypothesis

5.2.1 Introduction

Under the null hypothesis, the signal is $|X\rangle = |\text{Trend}\rangle + |\text{Noise}\rangle$, and we thus need to estimate the parameters of the trend and those of the zero-mean CARMA process. The best statistical approach is to estimate them jointly, and marginalize over the parameters of the trend, since the periodogram is invariant with respect to these parameters, according to Sect. 4.3. But this would imply very involved computations that are way beyond the scope of this work. We are thus forced to a compromise and proceed as follows: data are detrended, $|X_{\text{det}}\rangle = |X\rangle - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle\}})|X\rangle$, and then we estimate the parameters of the CARMA process, based on the model $\mu|t^0\rangle + |\text{Noise}\rangle$, where $|\text{Noise}\rangle$ is a zero-mean stationary Gaussian CARMA process sampled at the times of $|t\rangle$.

Estimation of CARMA parameters is done in a Bayesian framework. We analyze separately the case of the white noise, which is done analytically, and the case of CARMA(p,q) processes with $p \geq 1$, for which Markov-Chain Monte-Carlo (MCMC) methods are required. Bayesian analysis provides a posterior distribution of the parameters based on priors.

5.2.2 Gaussian white noise

We want to estimate the two parameters of the white noise, the mean μ and the variance σ^2. According to the Bayes theorem:

$$\Pi(\mu,\sigma^2|D) = \frac{\Pi(D|\mu,\sigma^2)\Pi(\mu,\sigma^2)}{\Pi(D)} \sim \Pi(D|\mu,\sigma^2)\Pi(\mu,\sigma^2), \tag{78}$$

where Π is the probability density function (PDF) and D is the detrended data $X_{\text{det},1},...,X_{\text{det},N}$. Based on the PDF of a multivariate white noise, the likelihood function is

$$\Pi(D|\mu,\sigma^2) = \left(\sqrt{\frac{1}{2\pi\sigma^2}}\right)^N \exp\left(\frac{-\sum_{i=1}^N (X_{\text{det},i}-\mu)^2}{2\sigma^2}\right). \tag{79}$$

We take *Jeffreys priors* (Jeffreys, 1946) for μ and σ^2:

$$\Pi(\mu,\sigma^2) = \Pi(\mu)\Pi(\sigma^2), \text{ with } \Pi(\mu) \sim 1 \text{ and } \Pi(\sigma^2) \sim \frac{1}{\sigma^2} \tag{80}$$

Jeffreys priors are non-informative and invariant under reparametrization. Note that $\Pi(\sigma^2)$ is log-uniform. Since we do not actually need to estimate μ (see Sect. 4.3 and formula (64)), we marginalize over that variable,

$$\begin{aligned}
\Pi(\sigma^2|D) &= \int_{-\infty}^{+\infty} \mathrm{d}\mu \Pi(\mu,\sigma^2|D) \\
&\sim \frac{1}{\sigma^2} \int_{-\infty}^{+\infty} \mathrm{d}\mu \Pi(D|\mu,\sigma^2) \\
&\sim \frac{1}{\sigma^2} \left(\sqrt{\frac{1}{2\pi\sigma^2}}\right)^N \exp\left(\frac{-\sum_{i=1}^N X_{\text{det},i}^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \mathrm{d}\mu \exp(-(a\mu^2+2b\mu)) \\
&\sim \frac{1}{\sigma^2} \left(\sqrt{\frac{1}{2\pi\sigma^2}}\right)^N \exp\left(\frac{-\sum_{i=1}^N X_{\text{det},i}^2}{2\sigma^2}\right) \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{a}\right),
\end{aligned} \tag{81}$$

with $a = N/2\sigma^2$ and $b = -\sum_{i=1}^N X_{\text{det},i}/2\sigma^2$. Rearranging terms gives

$$\Pi(\sigma^2|D) \sim \left(\frac{1}{\sigma^2}\right)^{\frac{N+1}{2}} \exp\left(-\frac{1}{\beta\sigma^2}\right), \tag{82}$$

with $\beta = 2/N\hat{\sigma}^2$, where $\hat{\sigma}^2 =$ is the (biased) variance of the detrended data[5]. With the variable change $y = 1/\sigma^2$, we have

$$\Pi(y|D) \sim y^{\frac{N-3}{2}} \exp(-y/\beta), \tag{83}$$

which is nothing but a gamma distribution:

$$\frac{1}{\sigma^2} \overset{d}{=} \gamma\left(\frac{N-1}{2}, \frac{2}{N\hat{\sigma}^2}\right). \tag{84}$$

[5] $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^N X_{\text{det},i}^2 - (\frac{1}{N}\sum_{i=1}^N X_{\text{det},i})^2$

Note that the mean of the distribution in Eq. (84) is equal to $(N-1)/(N\hat{\sigma}^2)$, which is the usual unbiased estimator of $1/\sigma^2$. Finally, the PDF of σ^2 is maximum at

$$\sigma^2_{\max} = \frac{N}{N+1}\hat{\sigma}^2. \tag{85}$$

This is obtained from the derivative of Eq. (82).

5.2.3 Gaussian CARMA(p,q) noise with $p \geq 1$

For other cases than the white noise, Kelly et al. (2014) provide robust algorithms to estimate the posterior distribution of the CARMA parameters and of the parameter μ of an irregularly sampled, purely stochastic, time series, which can be modeled as a CARMA process. Those algorithms are based on Bayesian inference and MCMC methods. We recommend to read in particular Sect. 3.3 and 3.6 of Kelly et al. (2014) for a discussion on the choice of the priors and for computational considerations respectively. That paper is accompanied by a Python and C++ package called *CARMA pack*. Some outputs of CARMA pack are shown in Sect. 9.

5.3 Estimation of the distribution of the periodogram under the null hypothesis

5.3.1 Working with a trendless stochastic process

Under the null hypothesis, the signal is $|X\rangle = |\text{Trend}\rangle + |\text{Noise}\rangle = \sum_{k=0}^{m} \gamma_k |t^k\rangle + |\text{Noise}\rangle$. The WOSA periodogram, Eq. (64), being invariant with respect to the parameters of the trend, we can pose $\gamma_k = 0$ for all k and $|X\rangle$ reduces to a zero-mean CARMA process.

5.3.2 Monte-Carlo approach

For each frequency, we need the distribution of the WOSA periodogram, Eq. (68), where $|X\rangle$ is now a CARMA process for which we know the distribution of its parameters, from Sect. 5.2. With Monte-Carlo methods, we are thus able to estimate any percentile of the distribution of the periodogram. If $|X\rangle$ is a zero-mean white noise, $|X\rangle$ is sampled from a standard normal distribution multiplied by the square root of the variance, whose inverse is sampled from the gamma distribution (Eq. (84)). If $|X\rangle$ is a CARMA(p,q) process with $p \geq 1$, $|X\rangle$ is generated with the Kalman filter (from CARMA pack - see Sect. 5.2.3). An example of confidence levels is shown on Fig. 8d.

We are thus able to estimate confidence levels for the WOSA periodogram taking into account the uncertainty on the parameters of the background noise.

5.3.3 Analytical approach

If we consider constant CARMA parameters, we show in this section that analytical confidence levels can be computed, even in the very tail of the distribution of the periodogram of the background noise. An example is given on Fig. 8c. The advantage of the analytical approach is double:

1. It provides confidence levels converging to the exact solution, as the number of conserved moments increases (see below). From a certain number of conserved moments, we can consider that convergence is numerically reached (see Fig. 9). Such an approach is particularly interesting for high confidence levels, as illustrated on Fig. 8c with the $99.9\,\%$ confidence level, for which a MCMC approach would require a huge number of samples to get a satisfactory accuracy.

2. As a consequence, for a given percentile, computing time is usually shorter with the analytical method than with the MCMC method. We note, however, that the MCMC approach generally needs less computing time when the number of data points becomes large, as shown in appendix E.

First approximation

If the marginal posterior distribution of each CARMA parameter is unimodal, we take the parameter value at the maximum of its PDF (white noise case, see Eq. (85)), or the median parameter[6] (other cases). Note that multimodality tends to appear more frequently for CARMA processes of high order. Working with a unique set of parameters allows us to find an analytical formula for the distribution of the WOSA periodogram. Considering the matrix forms of the CARMA noise (Eq. (20) or (38)) and the WOSA periodogram (Eq. (68)), we demonstrate the following theorem.

Theorem 1. *The WOSA periodogram, defined in Eq. (68), under the null hypothesis (76), is*

$$||P_{WOSA}(\omega)|X\rangle||^2 \stackrel{d}{=} \sum_{k=1}^{2Q(\omega)} \lambda_k(\omega)\chi^2_{1_k}, \tag{86}$$

where $|X\rangle = \sum_{k=0}^{m} \gamma_k |t^k\rangle + K|Z\rangle$*, K is the CARMA matrix defined in Eq. (20) or (38), and* $Q(\omega)$ *is the number of WOSA segments at* ω*.*
$\chi^2_{1_1}$*, ...,* $\chi^2_{1_{2Q(\omega)}}$ *are iid chi-square distributions with* 1 *degree of freedom, and* $\lambda_1(\omega)$*, ...,* $\lambda_{2Q(\omega)}(\omega)$ *are the eigenvalues of* $M'_\omega K K' M_\omega$ *and are non-negative. Matrix* M_ω *is defined in Eq. (69).*

Proof. Since the WOSA periodogram, Eq. (68), is invariant with respect to the parameters of the trend, we pose them equal to zero and consider the zero-mean CARMA process

$$|X\rangle = K|Z\rangle. \tag{87}$$

The periodogram is thus

$$||P_{\text{WOSA}}(\omega)|X\rangle||^2 = \langle Z|K' M_\omega M'_\omega K|Z\rangle = \gamma'\gamma, \tag{88}$$

with $\gamma = M'_\omega K|Z\rangle$. Since $|Z\rangle$ is a standard multivariate normal distribution, we have

$$\gamma \stackrel{d}{=} \mathcal{N}(0, M'_\omega K K' M_\omega). \tag{89}$$

[6]For CARMA processes with $p > 0$ and $q \geq 0$, the marginal posterior distribution is obtained by MCMC methods, and determining the maximum of the PDF thus requires some post-processing, such as smoothing the distribution. A simple alternative is to take the median.

$M'_\omega KK'M_\omega$ is a $(2Q(\omega), 2Q(\omega))$ real symmetric positive semi-definite matrix. We can thus diagonalize it:

$$\exists \text{ an orthogonal matrix } U \text{ s.t. } U'M'_\omega KK'M_\omega U = D, \tag{90}$$

with D being a diagonal matrix with the $2Q(\omega)$ non-negative eigenvalues of $M'_\omega KK'M_\omega$. We now have $U'\gamma \stackrel{d}{=} \mathcal{N}(0, D)$, and

$$||P_{\text{WOSA}}(\omega)|X\rangle||^2 = \gamma'\gamma = \gamma' UU'\gamma = \langle Z|\sqrt{D}\sqrt{D}|Z\rangle \stackrel{d}{=} \sum_{k=1}^{2Q(\omega)} \lambda_k(\omega)\chi^2_{1_k}, \tag{91}$$

where the $\chi^2_{1_k}$ distributions are iid. □

The **pseudo-spectrum** is defined as the expected value of the periodogram distribution:

$$\widehat{S}(\omega) = \sum_{k=1}^{2Q(\omega)} \lambda_k(\omega) = \text{tr}(M'_\omega KK'M_\omega). \tag{92}$$

The difference between the *pseudo-spectrum* and the traditional *spectrum* is explained in appendix C.

If the background noise is white, we have $K = \sigma\mathbb{I}$ and this implies that $\text{tr}(M'_\omega KK'M_\omega) = \text{tr}(M'_\omega M_\omega)\sigma^2 = \text{tr}(M_\omega M'_\omega)\sigma^2 = 2\sigma^2$, such that the pseudo-spectrum is

$$\widehat{S}(\omega) = 2\sigma^2, \tag{93}$$

and is thus flat. This is a well-known result of the LS periodogram (Scargle, 1982), generalized here to more evolved periodograms. Moreover, if there is no WOSA segmentation ($Q(\omega) = 1\ \forall\omega$), the periodogram is exactly chi-square distributed with 2 degrees of freedom:

$$||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle,|Gc_\omega\rangle,|Gs_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle\}})\sigma|Z\rangle||^2 \stackrel{d}{=} \sigma^2\chi^2_{1_1} + \sigma^2\chi^2_{1_2} \stackrel{d}{=} \sigma^2\chi^2_2, \tag{94}$$

which is also a generalization of a well-known result of the LS periodogram (Scargle, 1982).

The variance of the distribution of the periodogram, Eq. (86), is equal to $2\sum_{k=1}^{2Q(\omega)} \lambda_k^2(\omega) = 2||M'_\omega KK'M_\omega||_F^2$, where $||\cdot||_F$ is the Frobenius norm. As expected, it decreases with Q, as illustrated on Fig. 3.

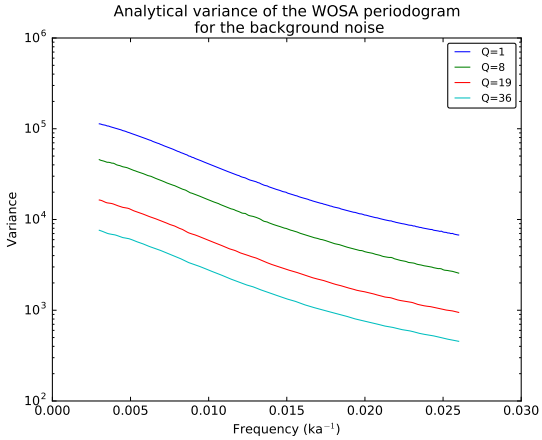


Figure 3. Analytical variance of the WOSA periodogram for a Gaussian red noise with $\sigma = 2$ and $\alpha = 1/20$ (see Sect. 3.2.3 for the definition of a red noise) for different values of Q. The frequency range is chosen such that, for each curve, $Q(\omega)$ is constant all along. The red noise is built on the irregularly sampled times of ODP1148 core (see Sect. 9).

Going back to Eq. (86), it is well-known that a linear combination of (independent) χ^2 distributions is not analytically solvable. Fortunately, excellent approximations are available in Provost et al. (2009), allowing to avoid Monte-Carlo methods.

Second approximation

We approximate the linear combination of independent chi-square distributions, conserving its first d moments. When $d \to \infty$, the approximation converges to the exact distribution. In practice, estimation of a percentile is already very good with a very few moments, as illustrated on Fig. 9. Let us proceed step by step by increasing the number of conserved moments. Define $X = \sum_{k=1}^{2Q(\omega)} \lambda_k(\omega) \chi^2_{1_k}$.

1-moment approximation

We require the expected value of the process to be conserved, which is satisfied with the following approximation:

$$X \overset{d}{\approx} \frac{1}{2Q(\omega)} [\sum_{k=1}^{2Q(\omega)} \lambda_k(\omega)] \chi^2_{2Q(\omega)}, \tag{95}$$

or, equivalently,

$$X \overset{d}{\approx} \frac{1}{2Q(\omega)} \widehat{S}(\omega) \chi^2_{2Q(\omega)}. \tag{96}$$

2-moment approximation

The approximate distribution of the linear combination of the chi-square distributions must have two parameters, and we conserve the expected value and variance. A chi-square distribution with M degrees of freedom provides a good fit:

$$X \overset{d}{\approx} g \chi^2_M. \tag{97}$$

Equating the expected values and variances gives

$$M = \frac{(\text{tr}(A))^2}{||A||_F^2} \text{ and } g = \frac{||A||_F^2}{\text{tr}(A)}, \tag{98}$$

where $A = M'_\omega K K' M_\omega$ and $||A||_F^2$ is the squared Frobenius norm of matrix A, i.e. the sum of its squared eigenvalues. Note that $g\chi^2_M \overset{d}{=} \gamma_{M/2,2g}$, where $2g$ is the *scale* parameter of the gamma distribution, which motivates the following d-moment approximation.

d-moment approximation

We apply here the formulas presented in Provost et al. (2009). Let f_X be the PDF of X. This distribution is approximated by the PDF of a d^{th} degree gamma-polynomial distribution:

$$f_X(x) \approx \gamma_{\alpha,\beta}(x) \sum_{i=0}^{d} \xi_i x^i, \qquad x \geq 0, \tag{99}$$

where the parameters α and β are estimated with the 2-moment approximation detailed above. ξ_0, ..., ξ_d are the solution of

$$\begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_d \end{pmatrix} = \begin{pmatrix} \eta(0) & \eta(1) & \dots & \eta(d-1) & \eta(d) \\ \eta(1) & \eta(2) & \dots & \eta(d) & \eta(d+1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta(d) & \eta(d+1) & \dots & \eta(2d-1) & \eta(2d) \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mu(1) \\ \vdots \\ \mu(d) \end{pmatrix}. \tag{100}$$

$\mu(1)$, ..., $\mu(d)$ are the exact first d moments of X and can be computed analytically by recurrence (see Eq. (5) of Provost et al., 2009). $\eta(h)$ is the h^{th} moment of the gamma distribution, $\eta(h) = \beta^h \Gamma(\alpha+h)/\Gamma(\alpha)$. The approximate cumulative distribution function (CDF) of X, evaluated at c_0, is then

$$F_X(c_0) \approx \frac{1}{\Gamma(\alpha)} \sum_{i=0}^{d} \xi_i \beta^i \gamma(i+\alpha, c_0/\beta), \qquad c_0 > 0, \tag{101}$$

where $\gamma(s,x)$ is the lower incomplete gamma function:

$$\gamma(s,x) = \int_0^x \text{d}t\, t^{s-1} \exp(-t). \tag{102}$$

After all that chain of calculus, we reached our objective, that is, the estimation of a confidence level for the WOSA periodogram. It is given by the solution c_0 of

$$\frac{1}{\Gamma(\alpha)} \sum_{i=0}^{d} \xi_i \beta^i \gamma(i+\alpha, c_0/\beta) - p = 0, \tag{103}$$

for some p-value p, e.g. $p = 0.95$ for a 95 % confidence level.

The gamma-polynomial approximation can be extended to the *generalized* gamma-polynomial approximation. The latter is based on the generalized gamma distribution and is defined in appendix D. It gives percentiles that usually converge faster than

with the gamma-polynomial approximation. However, we observed that the generalized gamma-polynomial approximation is quite sensitive to the quality of the first guess for the three parameters of the generalized gamma distribution (see appendix D). We thus recommend the use of the gamma-polynomial approximation as a first choice. Both options are available in WAVEPAL.

Finally, we mention that there exists an alternative expression to the above development, in terms of Laguerre polynomials (see Provost, 2005). It has the advantage of not requiring the matrix inversion in Eq. (100), the latter possibly being singular at large values of the degree d. However, we have not found any improvement on the stability or computing time using that approach.

5.4 The F-periodogram for the white noise background

We shown in Eq. (94) that the periodogram of a Gaussian white noise is exactly chi-square distributed if there is no WOSA segmentation. Significance testing against a white noise requires the estimation of the white noise variance after having detrended the data. Knowing that a F-distribution is the ratio of independent chi-square distributions, it is possible to get rid of the detrending and variance estimation and deal with a well-known distribution, by working with

$$\frac{(N-m-3)||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}}-P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle\}})|X\rangle||^2}{2||(\mathbb{I}-P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}})|X\rangle||^2}. \tag{104}$$

We call it the *F-periodogram*. We already know that the numerator is invariant with respect to the parameters of the trend of the signal. It is clear that the denominator is invariant with respect to the parameters of the trend as well as with respect to the amplitudes of the periodic components (only the $|\text{Noise}\rangle$ term remains when applying it to Eq. (13)). Moreover, that ratio is invariant with respect to the variance of the signal. Last but not least, the orthogonal projections in the numerator, $[P_{\overline{\text{sp}}\{|t^0\rangle,|t^2\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}}-P_{\overline{\text{sp}}\{|t^0\rangle,|t^2\rangle,...,|t^m\rangle\}}]$, and in the denominator, $[\mathbb{I}-P_{\overline{\text{sp}}\{|t^0\rangle,|t^2\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}}]$, are done on spaces that are orthogonal to each other. Consequently, if we consider the null hypothesis (76) with a white noise, the numerator and the denominator follow **independent** chi-square distributions, and

$$\begin{aligned}\frac{(N-m-3)||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}}-P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle\}})|X\rangle||^2}{2||(\mathbb{I}-P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,...,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}})|X\rangle||^2} &\stackrel{d}{=} \frac{(N-m-3)\chi^2_2}{2\chi^2_{N-m-3}} \\ &\stackrel{d}{=} F(2,N-m-3),\end{aligned} \tag{105}$$

where

$$|X\rangle \stackrel{d}{=} \sum_{k=0}^{m}\gamma_k|t^k\rangle + \mathcal{N}(\mu,\sigma^2) \stackrel{d}{=} |\text{Trend}\rangle + \mathcal{N}(\mu,\sigma^2), \tag{106}$$

and where $F(2,N-m-3)$ is the Fisher-Snedecor distribution with parameters 2 and $N-m-3$. In conclusion, the F-periodogram can be an alternative to the periodogram when performing significance testing. It has the advantage of not requiring any parameter to be estimated and applies under the following conditions

- The background noise is assumed to be white
- There is no WOSA segmentation

– There is no tapering

The F-periodogram is available in WAVEPAL under the above requirements.

With a WOSA segmentation, projections at the numerator and at the denominator are not performed anymore on orthogonal spaces, and this cannot therefore be applied.

The above results are a generalization of formulas in Brockwell and Davis (1991) and Heck et al. (1985). See appendix F for additional details.

6 The amplitude periodogram

6.1 Definition

Going back to Eq. (13), we now look for the amplitude $E_\omega = \sqrt{A_\omega^2 + B_\omega^2}$ at a given frequency $f = \frac{\omega}{2\pi}$. The estimation of E_ω^2 is called the *amplitude periodogram* and is denoted by $\widehat{E}_\omega^2$. We estimate A_ω and B_ω with a least squares approach. We start with a trendless signal, and will show that the amplitude periodogram and the periodogram are approximately proportional.

6.2 Trendless signal

6.2.1 No tapering

The estimated amplitudes we look for, $\widehat{A}_\omega$ and $\widehat{B}_\omega$, are the solution of

$$(\widehat{A}_\omega, \widehat{B}_\omega) = \underset{\{(A,B)\in\mathbb{R}^2\}}{\operatorname{argmin}} \; || \, |X\rangle - (A|c_\omega\rangle + B|s_\omega\rangle)||^2. \tag{107}$$

Since we look for the minimal distance, the solution is given by the orthogonal projection onto the vector space spanned by $|c_\omega\rangle$ and $|s_\omega\rangle$, namely

$$P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}|X\rangle = \widehat{A}_\omega|c_\omega\rangle + \widehat{B}_\omega|s_\omega\rangle. \tag{108}$$

Let us develop this equation:

$$V_{\omega_2}(V'_{\omega_2}V_{\omega_2})^{-1}V'_{\omega_2}|X\rangle = V_{\omega_2}|\widehat{\Phi}_\omega\rangle, \tag{109}$$

where

$$V_{\omega_2} = \begin{pmatrix} | & | \\ |c_\omega\rangle & |s_\omega\rangle \\ | & | \end{pmatrix} \text{ and } |\widehat{\Phi}_\omega\rangle = \begin{pmatrix} \widehat{A}_\omega \\ \widehat{B}_\omega \end{pmatrix}, \tag{110}$$

and we find the well-known expression for the solution of a least squares problem:

$$|\widehat{\Phi}_\omega\rangle = (V'_{\omega_2}V_{\omega_2})^{-1}V'_{\omega_2}|X\rangle. \tag{111}$$

Finally,

$$\widehat{E}_\omega = |||\widehat{\Phi}_\omega\rangle||. \tag{112}$$

In the regularly sampled case, at the Fourier frequencies, the amplitude periodogram is proportional to the periodogram, with a factor $2/N$ (or a factor $1/N$ at $\omega = 0$ and $\omega = \pi/\Delta t$; the projection being done on the single cosine at those frequencies). It is not anymore the case with irregularly sampled time series, and the proportionality is only approximate:

$$\widehat{E}^2_\omega \approx \frac{2}{N}||P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}|X\rangle||^2. \tag{113}$$

To prove the above formula, rewrite the model (13) at $\Omega = \omega$:

$$\begin{aligned}|X\rangle &= E_\omega \cos(\omega|t\rangle + \phi_\omega - \beta_\omega + \beta_\omega) + |\text{Noise}\rangle \\ &= A_\omega \cos(\omega|t\rangle - \beta_\omega) + B_\omega \sin(\omega|t\rangle - \beta_\omega) + |\text{Noise}\rangle,\end{aligned} \tag{114}$$

where β_ω is defined in Eq. (42) and makes the phase-lagged sine and cosine orthogonal. A_ω and B_ω no longer have the same expressions as in Eq. (13), but we still have $E^2_\omega = A^2_\omega + B^2_\omega$. We can rewrite Eq. (111) but this time with V_{ω_2} holding the above phase-lagged sine and cosine. We now make use of the approximation stated in Lomb (1976, p. 449):

$$\sum_{i=1}^{N} \cos^2(\omega t_i - \beta_\omega) \approx \frac{N}{2} \quad \text{and} \quad \sum_{i=1}^{N} \sin^2(\omega t_i - \beta_\omega) \approx \frac{N}{2}. \tag{115}$$

Note that the sum of both is exactly equal to N. Equation (113) is then obtained observing that $V'_{\omega_2} V_{\omega_2} \approx \frac{N}{2}\mathbb{I}$. Basic trigonometry gives the following equalities for the relative error of the above approximations:

$$\left|\frac{\sum_{i=1}^{N} \cos^2(\omega t_i - \beta_\omega) - N/2}{N/2}\right| = \left|\frac{\sum_{i=1}^{N} \sin^2(\omega t_i - \beta_\omega) - N/2}{N/2}\right| = \left|\frac{\sum_{i=1}^{N} \cos(2(\omega t_i - \beta_\omega))}{N}\right|, \tag{116}$$

so that the two approximations of (115) reduce to only one:

$$\frac{\sum_{i=1}^{N} \cos(2(\omega t_i - \beta_\omega))}{N} \approx 0. \tag{117}$$

The quality of this approximation is illustrated on Fig. 4.

6.2.2 With tapering

Like with the periodogram, leakage also appears in the amplitude periodogram. Consequently, it may be better to work with the projection on tapered cosine and sine if the data are not too much irregularly sampled, as explained in Sect. 4.4. Considering the tapered case is also an important mathematical prerequisite for an extension to the continuous wavelet transform. This is developed in part II of this study (Lenoir and Crucifix, 2017).

$\widehat{A}_\omega$ and $\widehat{B}_\omega$ are determined by projecting the data onto tapered cosine and sine:

$$P_{\overline{\text{sp}}\{|Gc_\omega\rangle,|Gs_\omega\rangle\}}|X\rangle = \widehat{A}_\omega|c_\omega\rangle + \widehat{B}_\omega|s_\omega\rangle. \tag{118}$$

Developing the equation gives

$$|\widehat{\Phi}_\omega\rangle = (V'_{\omega_2} G V_{\omega_2})^{-1} V'_{\omega_2} G |X\rangle, \tag{119}$$

and

$$\widehat{E}_\omega = |||\widehat{\Phi}_\omega\rangle||, \tag{120}$$

where V_{ω_2} is defined in Sect. 6.2.1 and G is defined in Sect. 4.4.

Note that the approach we follow does not correspond to the classical least squares problem as above since, in Eq. (118), the cosine and sine are tapered only on the left-hand side of the equality. However, one can reconstruct a signal from its projection coefficients with another function than the one which is used to determine those coefficients (see Torrésani, 1995, Eq. (II.8) p. 15, in which the similarity with $V_{\omega_2}|\widehat{\Phi}_\omega\rangle = V_{\omega_2}(V'_{\omega_2} G V_{\omega_2})^{-1} V'_{\omega_2} G|X\rangle$ is evident). Note that $V_{\omega_2}(V'_{\omega_2} G V_{\omega_2})^{-1} V'_{\omega_2} G$ is a projection, since it is idempotent, but the projection is not orthogonal, because it is not symmetric.

Similarly to the non-tapered case, we now determine an approximate proportionality between the amplitude periodogram and the tapered periodogram. We start with the model (13) evaluated at $\Omega = \omega$ and written under the following form

$$|X\rangle = A_\omega \cos(\omega|t\rangle - \beta_\omega) + B_\omega \sin(\omega|t\rangle - \beta_\omega) + |\text{Noise}\rangle, \tag{121}$$

where β_ω is introduced such that $\langle\, Gc_\omega \,|\, Gs_\omega \,\rangle = 0$, or equivalently, such that $V'_{\omega_2} G^2 V_{\omega_2}$ is diagonal. A little development gives the formula for determining β_ω:

$$\tan(2\beta_\omega) = \frac{\sum_{i=1}^N G_{ii}^2 \sin(2\omega t_i)}{\sum_{i=1}^N G_{ii}^2 \cos(2\omega t_i)}, \tag{122}$$

which is a generalization of Eq. (42). We now make use of the following approximations:

$$\frac{\sum_{i=1}^N G_{ii} \cos(2(\omega t_i - \beta_\omega))}{\text{tr}(G)} \approx 0, \tag{123a}$$

$$\frac{\sum_{i=1}^N G_{ii}^2 \cos(2(\omega t_i - \beta_\omega))}{\text{tr}(G^2)} \approx 0, \tag{123b}$$

which are similar to the approximation made in (117). That implies, with no extra approximation, the following formulas:

$$\begin{aligned}
\sum_{i=1}^N G_{ii} \cos^2(\omega t_i - \beta_\omega) &\approx \frac{\text{tr}(G)}{2}, \\
\sum_{i=1}^N G_{ii} \sin^2(\omega t_i - \beta_\omega) &\approx \frac{\text{tr}(G)}{2},
\end{aligned} \tag{124}$$

and

$$\begin{aligned}
\sum_{i=1}^N G_{ii}^2 \cos^2(\omega t_i - \beta_\omega) &\approx \frac{\text{tr}(G^2)}{2}, \\
\sum_{i=1}^N G_{ii}^2 \sin^2(\omega t_i - \beta_\omega) &\approx \frac{\text{tr}(G^2)}{2}.
\end{aligned} \tag{125}$$

Note that in (124) and (125), the sum of the two members is conserved and we find back Eq. (115) when $G = \mathbb{I}$. Moreover, we approximate the following sum:

$$\frac{\sum_{i=1}^{N} G_{ii} \cos(\omega t_i - \beta_\omega) \sin(\omega t_i - \beta_\omega)}{\mathrm{tr}(G)/2} \approx 0, \tag{126}$$

so that $V'_{\omega_2} G V_{\omega_2}$ is diagonal. The quality of these approximations is illustrated on Fig. 4. Putting all together gives

$$V'_{\omega_2} G V_{\omega_2} \approx \frac{\mathrm{tr}(G)}{2}\mathbb{I}, \quad \text{and} \quad V'_{\omega_2} G^2 V_{\omega_2} \approx \frac{\mathrm{tr}(G^2)}{2}\mathbb{I}, \tag{127}$$

from which we deduce

$$\widehat{E}^2_\omega \approx \frac{2\mathrm{tr}(G^2)}{\mathrm{tr}(G)^2} ||P_{\overline{\mathrm{sp}}\{|Gc_\omega\rangle, |Gs_\omega\rangle\}}|X\rangle||^2. \tag{128}$$

Finally, we mention that the above relation is approximate as well in the case of regularly sampled time series.

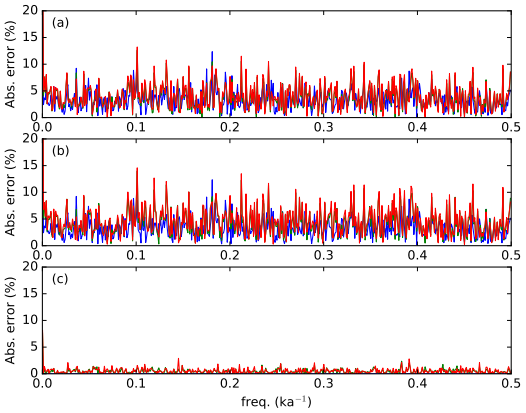


Figure 4. Illustration of the quality of the approximations (a) Eq. (123a) (b) Eq. (123b) and (c) Eq. (126). In blue: No tapering (square taper), in green: $\sin^2$ taper, in red: Gaussian taper. The approximation (117) is thus in blue in (a) or (b). Each panel represents the left-hand side of the equation, multiplied by 100, to express percentage. This indicates how small is the numerator compared to the denominator. The time vector $|t\rangle$ comes from the ODP1148 core (see Sect. 9) for which $\Delta t_{\mathrm{GCD}} = 1$ kyr.

6.3 Signal with a trend

We now work with the full model (13) including the trend. Our aim is again to find the amplitude E_ω, or, equivalently A_ω and B_ω. We proceed in the same way as in Sect. 6.2:

$$P_{\overline{\mathrm{sp}}\{|t^0\rangle, |t^1\rangle, \ldots, |t^m\rangle, |Gc_\omega\rangle, |Gs_\omega\rangle\}}|X\rangle = \sum_{k=0}^{m} \widehat{\gamma_k}|t^k\rangle + \widehat{A}_\omega|c_\omega\rangle + \widehat{B}_\omega|s_\omega\rangle = V_{\omega_{m+3}}|\widehat{\Phi}_\omega\rangle, \tag{129}$$

where

$$V_{\omega_{m+3}} = \begin{pmatrix} | & & | & | & | \\ |t^0\rangle & \dots & |t^m\rangle & |c_\omega\rangle & |s_\omega\rangle \\ | & & | & | & | \end{pmatrix}, \tag{130}$$

and

$$|\widehat{\Phi}_\omega\rangle = \begin{pmatrix} \widehat{\gamma_0} \\ \vdots \\ \widehat{\gamma_m} \\ \widehat{A}_\omega \\ \widehat{B}_\omega \end{pmatrix}. \tag{131}$$

We can write: $P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\dots,|t^m\rangle,|Gc_\omega\rangle,|Gs_\omega\rangle\}} = W_{\omega_{m+3}}(W'_{\omega_{m+3}}W_{\omega_{m+3}})^{-1}W'_{\omega_{m+3}}$, where $W_{\omega_{m+3}}$ is identical to $V_{\omega_{m+3}}$ except in the last two columns, where the cosine and sine are tapered by G. We thus obtain

$$|\widehat{\Phi}_\omega\rangle = (W'_{\omega_{m+3}}V_{\omega_{m+3}})^{-1}W'_{\omega_{m+3}}|X\rangle, \tag{132}$$

and

$$\widehat{E}^2_\omega = \widehat{A}^2_\omega + \widehat{B}^2_\omega = \widehat{\Phi}_\omega(m+2)^2 + \widehat{\Phi}_\omega(m+3)^2, \tag{133}$$

where $\widehat{\Phi}_\omega(m+2)$ and $\widehat{\Phi}_\omega(m+3)$ are the two last components of vector $|\widehat{\Phi}_\omega\rangle$.

6.4 With WOSA

The signal being stationary, we can estimate the amplitude on overlapping segments and take the average. That gives a better estimation, more robust against the background noise, but it has the disadvantage of widening the peaks and thus reducing the resolution in frequency. We simply take Eq. (132), apply it to each segment[7], and compute the average. We have

$$\widehat{E}^2_\omega = \frac{1}{Q(\omega)}\sum_{q=1}^{Q(\omega)}[\widehat{\Phi}_{q,\omega}(m+2)^2 + \widehat{\Phi}_{q,\omega}(m+3)^2]. \tag{134}$$

6.5 Amplitude periodogram or periodogram?

So far, we have studied in detail the periodogram and its confidence levels as well as the estimated amplitude. Of course, confidence levels can also be determined for the amplitude, with Monte-Carlo simulations, or with an analytical approximation similar to Sect. 5.3.3.

In the regularly sampled case, at Fourier frequencies, the cosine and sine vectors are orthogonal, so that, in the non-tapered case

[7] We remind that the vectors $|t^k\rangle$ associated to the trend are taken on the whole time series. Only the (tapered) cosine and sine are taken on the WOSA segment.

and with a constant trend, there is no difference between the periodogram and the amplitude periodogram, up to a multiplicative constant. Even with WOSA segmentation, the number of data points being identical on each segment, that multiplicative constant remains invariant.

In the irregularly sampled case, choosing one or the other depends on what one wants to conserve. The periodogram conserves the flatness of the white noise pseudo-spectrum (see Eq. (93)) and can therefore be of interest to study the background noise of the time series. On the other hand, the amplitude periodogram gives a direct access to the estimated signal amplitude. Another criteria to take into account is the computing time. Indeed, the amplitude periodogram requires matrix inversions (or, equivalently, resolution of linear systems) and is then slower to compute, while the periodogram allows to deal with orthogonal projections and is computationally more efficient. Finally, we mention that, with a trendless signal, difference between both is rather explicit (see Eq. (118)):

$$\text{Periodogram: } ||\widehat{A}_\omega|c_\omega\rangle + \widehat{B}_\omega|s_\omega\rangle||^2 \tag{135}$$

versus

$$\text{Amplitude periodogram: } \widehat{A}_\omega^2 + \widehat{B}_\omega^2. \tag{136}$$

This is variance (multiplied by the number of data points) versus squared amplitude. A compromise between the amplitude periodogram and the periodogram is the *weighted periodogram*, which is defined in the next section.

7 The weighted WOSA periodogram

Taking into account the approximate linearity between the amplitude periodogram and the tapered periodogram, Eq. (128), a possibility is to perform the frequency analysis with a weighted version of the WOSA periodogram. On each WOSA segment, the periodogram is weighted by $w_q = 2\text{tr}(G_q^2)/\text{tr}(G_q)^2$, $q = 1, ..., Q(\omega)$. The advantage of the weighted WOSA periodogram is to provide deterministic peaks (coming from $A_\omega|c_\omega\rangle + B_\omega|s_\omega\rangle$) of more or less equal power on all the WOSA segments, thus alleviating the issue stated in Sect. 4.5.2. The disadvantage is that the pseudo-spectrum of a white noise is not flat anymore (Eq. (93) is not valid anymore, except when $Q = 1$). Working with the weighted version is done by modifying matrix M_ω, Eq. (69), which is now

$$M_\omega = \frac{1}{\sqrt{Q(\omega)}} \begin{pmatrix} | & | & & | & | \\ \sqrt{w_1}|h_{1,1}(\omega)\rangle & \sqrt{w_1}|h_{2,1}(\omega)\rangle & \dots & \sqrt{w_Q(\omega)}|h_{1,Q(\omega)}(\omega)\rangle & \sqrt{w_Q(\omega)}|h_{2,Q(\omega)}(\omega)\rangle \\ | & | & & | & | \end{pmatrix}. \tag{137}$$

Note that the weights w_q are the same on each segment when the time series is regularly sampled, so that the whole WOSA periodogram is, in that case, just multiplied by a constant, and the pseudo-spectrum of a white noise is flat. We observed that the weighted periodogram is often very close to the amplitude periodogram, like in the example presented in Fig. 10. We thus recommend the use of the weighted WOSA periodogram in most analyses.

When filtering is to be performed, the amplitude periodogram must be computed as well. This is the topic of the next section.

8 Filtering

We want to reconstruct the deterministic periodic part, $\widehat{A}_\omega|c_\omega\rangle + \widehat{B}_\omega|s_\omega\rangle$ of our model (13) evaluated at $\Omega = \omega$. From Eq. (132), we can extract $\widehat{A}_\omega = \widehat{\Phi}_\omega(m+2)$ and $\widehat{B}_\omega = \widehat{\Phi}_\omega(m+3)$, and reconstruction at a single frequency is therefore direct. Reconstruction on a frequency range can be done by summing $\widehat{A}_\omega|c_\omega\rangle + \widehat{B}_\omega|s_\omega\rangle$ over ω.

Note that, in theory, reconstruction could be done segment by segment, using the WOSA method. But, in practice, we observe that it does not give good results with stationary signals. Of course, if the signal is not stationary, reconstruction segment by segment is a clever choice, but, with such signals, it is better to use more appropriate tools such as the wavelet transform. See the second part of this study (Lenoir and Crucifix, 2017), in which some examples of filtering are given.

9 Application on paleoceanographic data

The time series we use to illustrate the theoretical results is the benthic foraminiferal $\delta^{18}O$ record from Jian et al. (2003) that holds 608 data points with distinct ages and covers the last 6 million years. An example of frequency analysis is described below.

9.1 Preliminary analysis

We first look at the sampling. $\Delta t_{\text{GCD}} = 1$ kyr, and $r_t = 10.13\ \%$. Following the recommendation of Sect. 4.4, we therefore use the default rectangular window taper. The sampling and its distribution are drawn on Fig. 5. We then choose the degree of the polynomial trend to be $m = 7$, see Fig. 6. This choice for m is justified by a sensitivity analysis performed in Sect. 9.4. We remind that the time series is not detrended before estimating the spectral power of the data, but it is detrended before estimating the confidence levels.

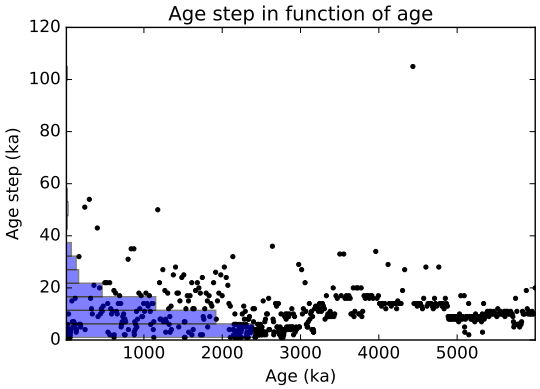


Figure 5. The age step, $(t_k - t_{k-1})\ \forall k \in 2, ..., N$, and its distribution.

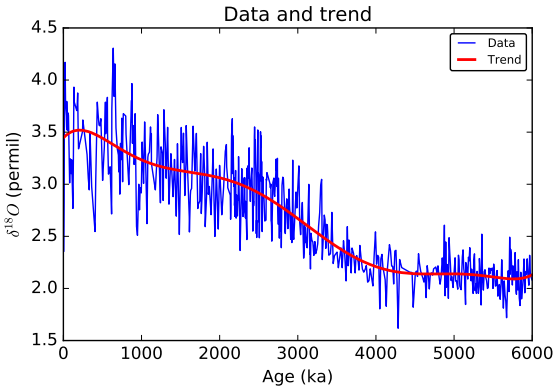


Figure 6. The time series and its 7^{th} degree polynomial trend.

9.2 CARMA(p,q) background noise analysis

We choose the order of the background noise CARMA process. We opt for the traditional red noise background (Hasselmann, 1976), $p = 1$ and $q = 0$. Note that we observe similar confidence levels with other choices (see the sensitivity analysis in Sect. 9.5). We then estimate the parameters of the stationary CARMA process (here, a red noise) on the detrended data. This is done with the algorithm provided by Kelly et al. (2014) (see Sect. 5.2.3). Quality of the fit is analyzed on Fig. 7a, 7c and 7e. Fig. 7a analyzes the residuals. If the detrended data are a Gaussian red noise, the residuals must be distributed as a Gaussian white noise. We see that the distribution is indeed close to a Gaussian. Fig. 7c shows the autocorrelation function (ACF) of the residuals. If the residuals are a Gaussian white noise sequence, they must be uncorrelated at any lag. We can therefore arrange the residuals on a regular grid with a unit step and then take the classical ACF, which can only be applied to regularly sampled data. Fig. 7c is consistent with the assumption that the residuals are uncorrelated. Fig. 7e shows the ACF of the squared residuals. If the residuals are a Gaussian white noise sequence, the squared residuals are a white noise sequence (which is not Gaussian anymore) and must therefore be uncorrelated at any lag. Deviations from the confidence grey zone indicate that the variance is changing with time and the signal is therefore not stationary. This is actually what is happening with our time series. Changes in variance are already visible on the raw time series (Fig. 6). Remember that, at this stage, we are within the world of the null hypothesis, Eq. (76), and slight violation of the goodness of fit may be due to the presence of additive periodic deterministic components, that is the alternative hypothesis.

The marginal posterior distributions of the CARMA parameters are shown on Fig. 7b, 7d and 7f, jointly with the ACF of the MCMC samples. Each distribution is unimodal, and we may therefore use the analytical approach of Sect. 5.3.3 to estimate the confidence levels. Based on the ACFs of the MCMC samples of the three parameters, we skim off the initial joint distribution of the parameters to make their samples almost uncorrelated. In this example, we pick up 1231 samples among the 16000

initial ones. This number of 1231 samples results from the fact that we impose an ACF which is less than 0.2 for each marginal distribution[8].

[8] As explained in Sect. 9.3, these 1231 samples are then used to compute the median parameters, producing the analytical confidence levels of Fig. 8c and 8d and the MCMC confidence levels of Fig. 8c. The MCMC confidence levels of Fig. 8d are computed from 50000 samples of the parameters, after skimming off a distribution with much more samples.

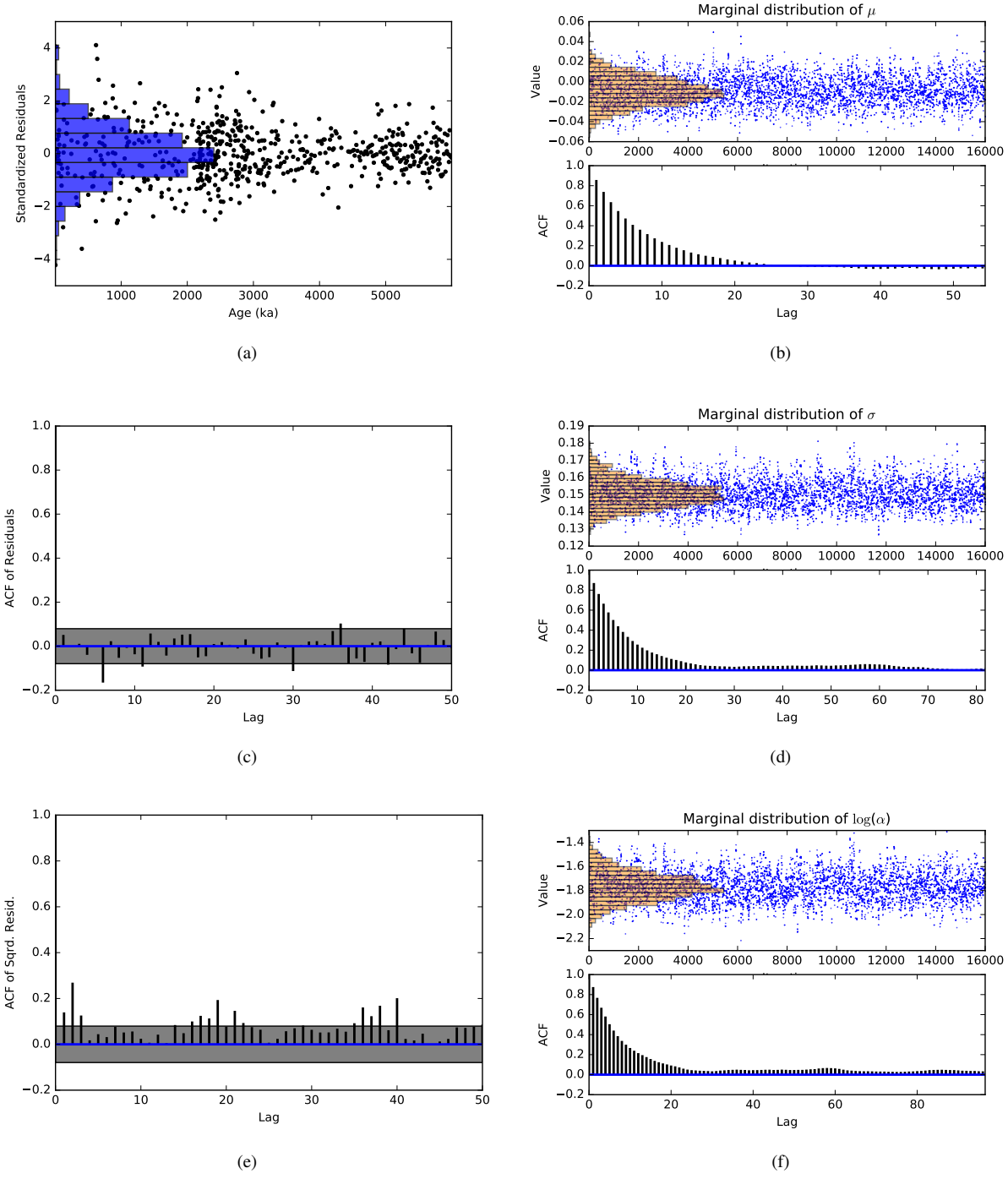


Figure 7. CARMA(1,0) background noise analysis. (a), (c) and (e) assess the fit. (a) Standardized residuals. (c) ACF of the residuals. (e) ACF of the squared residuals. The lag refers to an arbitrary scale on which the data are regularly spaced with a unit step. The grey portion is the 95 % confidence region. (b), (d) and (f) show the samples of the MCMC and the posterior marginal distributions (top panel), jointly with the ACF of the MCMC samples (bottom panel). (b) Mean. (d) Standard deviation of the white noise term. (f) $\log(\alpha)$, where α is defined in Sect. 3.2.3.

9.3 Frequency analysis

We compute the weighted WOSA periodogram of Sect. 7. The frequency range is automatically determined from the results of Sect. 4.5.4. The length of the WOSA segments depends on the required frequency resolution. Here we choose segments of about 600 kyr and a $75\ \%$ overlapping. The WOSA segmentation is presented on Fig. 8a.

The weighted WOSA periodogram and its $95\ \%$ and $99.9\ \%$ confidence levels are presented on Fig. 8c and 8d. Both figures display the analytical confidence levels, which are computed with the median parameters of the red noise process (that is the median of 1231 samples of the distributions shown on Fig. 7b, 7d and 7f) and a 12-moment gamma-polynomial approximation (Sect. 5.3.3). We can check for the convergence of the gamma-polynomial approximation, at some frequencies. This is presented on Fig. 9. Figure 8c also shows the MCMC confidence levels, computed from 50000 red noise time series, all generated with the median red noise parameters. As we can see on Fig. 8c, the matching between the analytical and MCMC confidence levels is excellent, also in the very tail of the distribution, at the $99.9\ \%$ confidence level. We can go a step further and take into account the uncertainty on the CARMA parameters, as explained in Sect. 5.3.2. Figure 8d presents the MCMC confidence levels that are computed from 50000 red noise time series, generated with stochastic parameters, that are taken from the joint posterior distribution of the parameters of the red noise process. The number of WOSA segments per frequency, denoted by $Q(f)$ in Sect. 4 to 7, is on Fig. 8b, and provides an indication of the noise damping per frequency. Indeed, the variability due to the background noise is increasingly damped as the number of WOSA segments grows.

We also compute the amplitude periodogram, Eq. (134), which is actually very close to the weighted periodogram, as shown on Fig. 10. Similar results are obtained using other tapers (not shown). This illustrates the quality of the approximations made in Sect. 6.2.2. Note that the estimation of the amplitude E_ω of the model (13) is always biased by the background noise (we observe on Fig. 10 that the peaks emerge from a baseline which is well above zero).

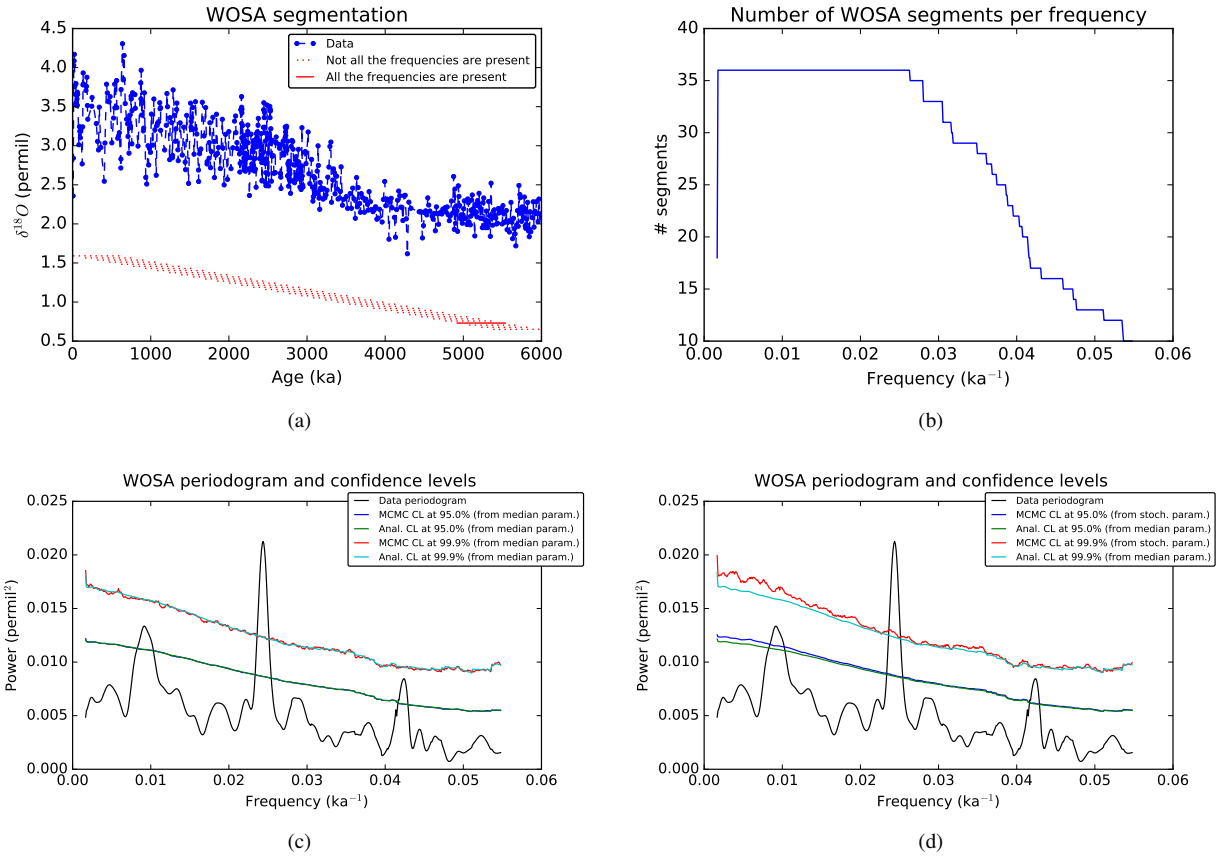


Figure 8. Frequency analysis. (a) The time series, in blue, and the WOSA segments, in red. (b) Number of WOSA segments per frequency. (c) and (d): Weighted WOSA periodogram and the confidence levels (CL) at 95 % and 99.9 %. Analytical CL (Anal. CL) are computed with the median parameters of the red noise process. In (c), the MCMC CL are computed from the MCMC red noise time series, all generated with the median red noise parameters. In (d), the MCMC CL are computed from the MCMC red noise time series, generated with stochastic parameters, that are taken from the joint posterior distribution of the parameters of the red noise process.

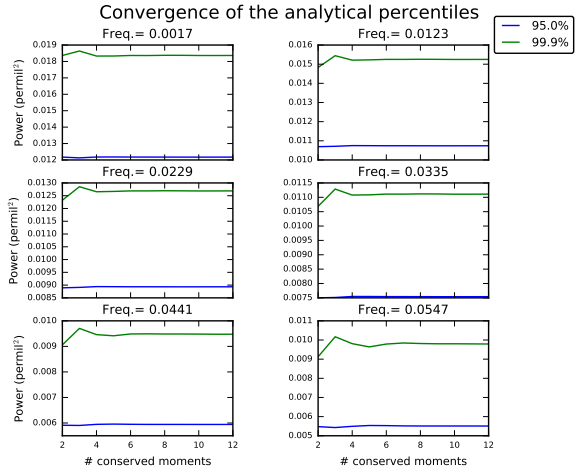


Figure 9. At six particular frequencies, check for the convergence of the analytical percentiles.

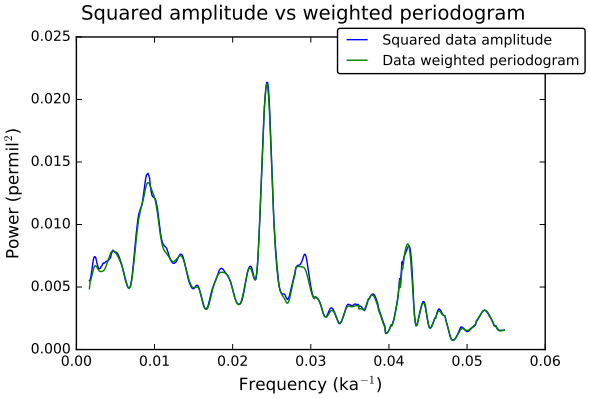


Figure 10. Comparison between the amplitude periodogram (= squared amplitude) and the weighted periodogram. The green curve is the same as the black curve of Fig. 8c and 8d

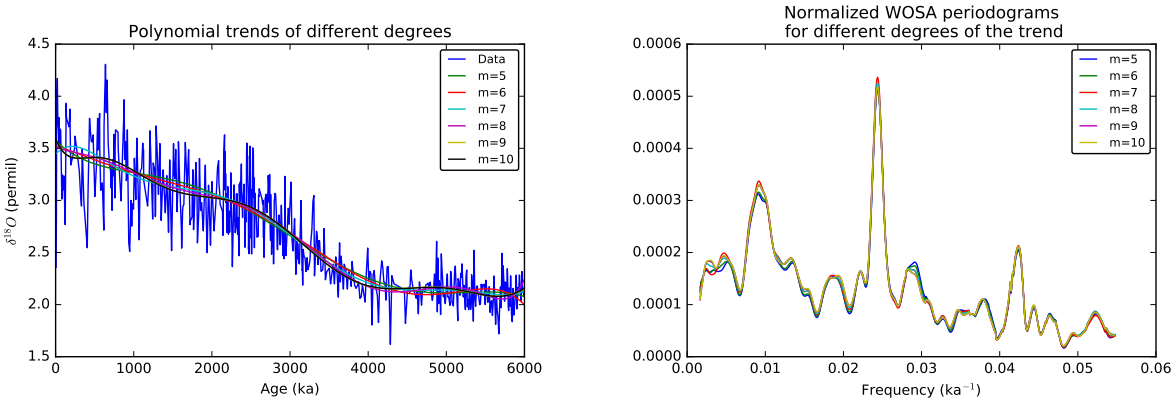
9.4 Sensitivity analysis for the degree of the polynomial trend

We show on Fig. 11 that the degree m of the polynomial trend, taken between 5 and 10, does not influence substantially the WOSA periodogram. Below $m = 5$, the trend no longer fits the data correctly (from a mere visual inspection), while above $m = 10$, spurious oscillations may appear.

Note that we do not apply here the Akaike Information criterion (AIC) (Akaike, 1974). Indeed, defining a stochastic model for the trend and estimating its likelihood is quite tedious in our case, since we work with CARMA stochastic processes. Moreover, at this stage, we do not want to choose yet between the orders of the CARMA process.

9.5 Sensitivity analysis for the order of the CARMA process

Fig. 12 displays the confidence levels for various orders of the CARMA process: $(p,q) = (0,0)$, $(p,q) = (1,0)$, $(p,q) = (2,0)$ and $(p,q) = (2,1)$. It is clear that the CARMA(0,0) (= white noise) does not capture enough spectral variability to perform significance testing and that using a CARMA(2,0) or a CARMA(2,1) is basically equivalent to using a red noise.



(a) (b)

Figure 11. (a) Trends of different degrees for the time series. (b) Weighted WOSA periodograms for different degrees of the trend. Each periodogram is normalized like in Eq. (58), in order to make a meaningful comparison.

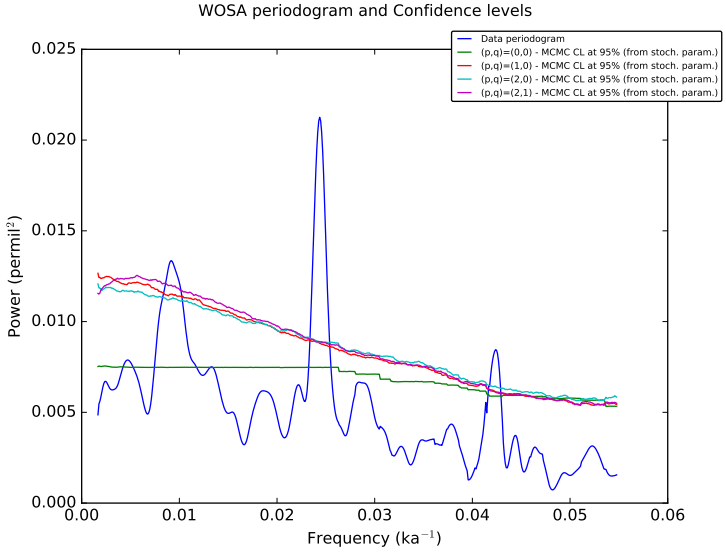


Figure 12. The weighted WOSA periodogram and its 95 % confidence levels for different orders (p,q) of the CARMA process. Note that the marginal posterior distributions of some parameters of the CARMA(2,0) and CARMA(2,1) processes are multimodal, so the analytical approach cannot be applied, and MCMC confidence levels must therefore be used.

10 WAVEPAL Python package

WAVEPAL is a package, written in Python 2.X, that performs frequency and time-frequency analyses of irregularly sampled time series, significance testing against a stationary Gaussian CARMA(p,q) process, and filtering. Frequency analysis is based on the theory developed in this article, and time-frequency analysis relies on the theory developed in part II of this study (Lenoir and Crucifix, 2017). It is available at https://github.com/guillaumelenoir/WAVEPAL.

11 Conclusions

We proposed a general theory for the detection of the periodicities of irregularly sampled time series. This is based on a general model for the data, which is the sum of a polynomial trend, a periodic component and a Gaussian CARMA stochastic process. In order to perform the frequency analysis, we designed new algebraic operators that match the structure of our model, as extensions of the Lomb-Scargle periodogram and the WOSA method. A test of significance for the spectral peaks was designed as a hypothesis testing and we investigated in detail the estimation of the percentiles of the distribution of our algebraic operators under the null hypothesis. Finally, we shown that the least squares estimation of the squared amplitude of the periodic component and the periodogram are no longer proportional if the time series is irregularly sampled. Approximate proportionality relations were proposed and are at the basis of the weighted WOSA periodogram, which is the analysis tool that we recommend for most frequency analyses. The general approach presented in this paper allows an extension to the continuous wavelet transform, which is developed in part II of this study (Lenoir and Crucifix, 2017).

Code availability. The Python code generating the figures of this article is available in the supplementary material.

Appendix A: Some properties of the Lomb-Scargle periodogram

We present some properties of the LS periodogram, defined in Sect. 4.1.

A1 Periodicity of the periodogram

The LS periodogram and all its generalizations (e.g. Eq. (64)) exhibit a periodicity similar to the DFT of regularly sampled real processes: The periodogram over the frequency range $]-1/2\Delta t_{\text{GCD}}, 1/2\Delta t_{\text{GCD}}]$ repeats itself periodically. Moreover, the periodogram at frequency $-f$ is equal to the periodogram at frequency $+f$. Consequently, we must work at most on the frequency range $[0, 1/2\Delta t_{\text{GCD}}[$ to avoid aliasing.

A2 Total reconstruction

Integrating the orthogonal projection $P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}$ between frequency 0 and $1/2\Delta t_{\text{GCD}}$ does not give the identity operator. We only have an approximate equality. Using Lomb's approximation, given in Eq. (115), and no extra approximation, some

algebra gives

$$\int_0^{\pi/\Delta t_{\text{GCD}}} \mathrm{d}\omega \, (|c_\omega^\sharp\rangle\langle c_\omega^\sharp| + |s_\omega^\sharp\rangle\langle s_\omega^\sharp|) \approx \frac{2\pi}{N\Delta t_{\text{GCD}}}\mathbb{I}. \tag{A1}$$

It is interesting to compare it with the integration of complex exponentials, which gives exactly the identity operator:

$$\int_{-\pi/\Delta t_{\text{GCD}}}^{\pi/\Delta t_{\text{GCD}}} \mathrm{d}\omega \, |e_\omega^\sharp\rangle\langle e_\omega^\sharp| = \frac{2\pi}{N\Delta t_{\text{GCD}}}\mathbb{I}, \tag{A2}$$

where $|e_\omega^\sharp\rangle = \frac{1}{\sqrt{N}}\exp(i\omega|t\rangle) = \frac{1}{\sqrt{N}}(|c_\omega\rangle + i|s_\omega\rangle)$. The above formula may be interpreted as a form of Parseval's identity. That property of exact reconstruction is, incidentally, at the basis of the multitaper method (Lenoir, 2017). With that property and the no less interesting mathematical properties of the complex exponentials, it is legitimate to ask why we would not work with the projection on a complex exponential instead of a projection on cosine and sine. The main disadvantage of working with exponentials is the loss of power in the negative frequencies. Indeed, the trendless model (13) at $\Omega = \omega$ can be rewritten as

$$\begin{aligned} |X\rangle &= E_\omega \frac{\exp(i(\omega|t\rangle + \phi_\omega)) + \exp(-i(\omega|t\rangle + \phi_\omega))}{2} + |\text{Noise}\rangle \\ &= C_\omega|e_\omega\rangle + D_\omega|e_{-\omega}\rangle + |\text{Noise}\rangle, \end{aligned} \tag{A3}$$

where $|e_\omega\rangle = \exp(i\omega|t\rangle)$. In the case of irregularly sampled time series, we no longer have, in general, $\langle e_\omega \, | \, e_{-\omega} \rangle = 0$, so that some power is lost in the negative frequencies when projecting on $\overline{\text{sp}}\{|e_\omega\rangle\}$. We could then think about performing the projection on $\overline{\text{sp}}\{|e_\omega\rangle, |e_{-\omega}\rangle\}$, but this does not lead to the identity operator when integrating from frequency $-1/2\Delta t_{\text{GCD}}$ to $+1/2\Delta t_{\text{GCD}}$.

A3 Invariance under time translation

As stated in Scargle (1982), the LS periodogram is invariant under time translation. $P_{\overline{\text{sp}}\{|c_\omega\rangle, |s_\omega\rangle\}}$ is of course invariant under such a transformation. The result can be generalized to more evolved projections. Indeed, $[P_{\overline{\text{sp}}\{|t^0\rangle, |t^1\rangle, \ldots, |t^m\rangle, |c_\omega\rangle, |s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle, |t^1\rangle, \ldots, |t^m\rangle\}}]$ is also invariant under time translation, provided all the powers of $|t\rangle$ from 0 to m are taken into account. That projection is also invariant under time dilatation if the frequency is contracted accordingly.

Appendix B: Periodogram and mean: Equivalence between published formulas

We show here the equivalence between some published formulas, with notations that are a mix between those of the cited articles and those of the present one, in order to facilitate the reading.

Brockwell and Davis (1991, p. 335) work with

$$||(P_{\overline{\text{sp}}\{|t^0\rangle, |c_\omega\rangle, |s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle\}})|X\rangle||^2. \tag{B1}$$

It is defined for regularly sampled time series, and is suitable for irregularly sampled time series as well. That formula is the same as Eq. (47).

Ferraz-Mello (1981) considers irregularly sampled time series and defines the intensity (p. 620) by

$$I(\omega) = c_1^2 + c_2^2, \tag{B2}$$

where $c_1 = \langle\, f \,|\, h_1 \,\rangle$ and $c_2 = \langle\, f \,|\, h_2 \,\rangle$. $|f\rangle$ contains the measurements (this is $|X\rangle$ in the present article) and $|h_1\rangle$ and $|h_2\rangle$ are exactly the same as in Eq. (53). $I(\omega)$ is thus equal to Eq. (55).

Heck et al. (1985) deal with irregularly sampled time series and define (Eq. (1) p. 65):

$$\mathrm{SP}(\nu) = \langle X|F_{1,0}(\nu)|X\rangle = \langle X|A(\nu)[A(\nu)'A(\nu)]^{-1}A(\nu)'|X\rangle, \tag{B3}$$

where ν denotes the frequency ($\nu = \omega/2\pi$) and $A(\nu)$ is a $(N,2)$ matrix whose first column is $|c_\omega\rangle - |t^0\rangle\langle\, t^0 \,|\, c_\omega \,\rangle/N$ and second column is $|s_\omega\rangle - |t^0\rangle\langle\, t^0 \,|\, s_\omega \,\rangle/N$. Equation (B3) is nothing but the squared norm of the orthogonal projection of the data $|X\rangle$ onto the span of those two vectors. By a Gram-Schmidt orthonormalization, it is easy to see that $\overline{\mathrm{sp}}\{|c_\omega\rangle - |t^0\rangle\langle\, t^0 \,|\, c_\omega \,\rangle/N\ ,\ |s_\omega\rangle - |t^0\rangle\langle\, t^0 \,|\, s_\omega \,\rangle/N\} = \overline{\mathrm{sp}}\{|h_1\rangle, |h_2\rangle\}$, where $|h_1\rangle$ and $|h_2\rangle$ are defined in Eq. (53). We thus have the periodogram defined in Eq. (55).

Appendix C: On the pseudo-spectrum

We define the *pseudo-spectrum* as the expected value of the WOSA periodogram under the null hypothesis (see Sect. 5.1):

$$\widehat{S}(\omega) = E\left\{||P_{\mathrm{WOSA}}(\omega)|X\rangle||^2\right\}, \tag{C1}$$

where $|X\rangle = |\mathrm{Trend}\rangle + |\mathrm{Noise}\rangle$, in which $|\mathrm{Noise}\rangle$ is a zero-mean stationary Gaussian CARMA process sampled at the times of $|t\rangle$, and the expectation is taken on the samples of the CARMA noise. With what we have seen in Sect. 5.3.2 and 5.3.3, the periodogram is either obtained with Monte-Carlo methods or analytically with some approximations. In the former case, $\widehat{S}(\omega)$ is estimated by taking the numerical average of the periodogram at each frequency. In the latter case, an analytical formula for the pseudo-spectrum is available. Indeed, the process under the null hypothesis is $|X\rangle = K|Z\rangle + \sum_{k=0}^{m}\gamma_k|t^k\rangle$, where K is defined in Eq. (20) or (38), and we have

$$\widehat{S}(\omega) = \sum_{k=1}^{2Q(\omega)} \lambda_k(\omega) = \mathrm{tr}(M_\omega' K K' M_\omega), \tag{C2}$$

where the different terms are defined in theorem 1.

When dealing with a trendless signal, we can perform the WOSA on the classical tapered periodogram and the pseudo-spectrum becomes

$$\widehat{S}(\omega) = E\left\{||P_{\mathrm{WOSA}}(\omega)|X\rangle||^2\right\} = E\left\{\sum_{q=1}^{Q(\omega)} ||P_{\overline{\mathrm{sp}}\{|G_q c_{\omega,q}\rangle, |G_q s_{\omega,q}\rangle\}}|X\rangle||^2\right\}. \tag{C3}$$

In the case of regularly sampled data, Eq. (C3) converges to the *spectrum* $S(\omega)$ as the number of data points increases (up to a multiplicative factor Δt, the time step). See Walden (2000) where it is shown that $||P_{\text{WOSA}}(\omega)|X\rangle||^2$ is a mean-square-consistent and asymptotically unbiased estimator of the spectrum. The *spectrum* $S(\omega)$, also called *Fourier power spectrum*, of a regularly sampled zero-mean real stationary process $|X\rangle$ is defined by (see[9] Sect. 10.3 of Brockwell and Davis, 1991):

$$S(\omega) = \Delta t \lim_{N\to\infty} E\left\{||P_{\overline{\text{sp}}\{|c_\omega\rangle,|s_\omega\rangle\}}|X\rangle||^2\right\}. \tag{C4}$$

Now, considering Eq. (96), we thus have, for trendless regularly sampled time series, the following 1-moment approximation:

$$||P_{\text{WOSA}}(\omega)|X\rangle||^2 \overset{d}{\approx} \frac{1}{2Q} S(\omega) \chi^2_{2Q}. \tag{C5}$$

With that approximation, the spectrum $S(\omega)$, which is well known for some processes like ARMA processes, gives access to the confidence levels. The above formula is widely used in the literature on regularly sampled time series in the case of one WOSA segment ($Q = 1$), for which the one moment approximation is good enough (see, for instance, Torrence and Compo, 1998, Eq. (17)).

In the case of irregularly sampled data, the spectrum $S(\omega)$ can be defined over the frequency range $[-1/2\Delta t_{\text{GCD}}, 1/2\Delta t_{\text{GCD}}[$. This follows from the spectral representation theorem (Priestley, 1981, Chap. 4) applied to irregularly sampled time series. But $\widehat{S}(\omega)$ usually strongly differs from $S(\omega)$, except in the white noise case where the spectrum is flat. Building estimators of the spectrum $S(\omega)$ in the case of irregularly sampled time series actually seems very challenging, as briefly discussed in Sect. 4.5.1.

Appendix D: The generalized gamma-polynomial distribution as an approximation for the linear combination of chi-square distributions

We extend the gamma-polynomial approximation of Sect. 5.3.3 to the *generalized* gamma-polynomial approximation. Both conserve the first d moments of the distribution X. The generalized gamma-polynomial approximation is based on the generalized gamma distribution, which has three parameters, such that the prerequisite of a d-moment approximation is a 3-moment approximation with the generalized gamma distribution.

D1 3-moment approximation

We work with the generalized gamma distribution, which has 3 parameters,

$$X \overset{d}{\approx} \gamma_{\alpha,\beta,\delta}. \tag{D1}$$

Its PDF is

$$f_\gamma(x;\alpha,\beta,\delta) = \frac{\delta}{\beta^{\alpha\delta}\Gamma(\alpha)} x^{\alpha\delta-1} \exp(-(x/\beta)^\delta) \qquad \alpha,\beta,\delta > 0, \tag{D2}$$

[9] In that book, the authors work with the projection on complex exponentials, $|e_\omega\rangle = |c_\omega\rangle + i|s_\omega\rangle$, instead of a projection on cosine and sine. But this is asymptotically the same since, asymptotically, the cosine and sine are orthogonal at all the frequencies.

where Γ is the gamma function. It reduces to the gamma distribution when $\delta = 1$. Its moments are

$$\mu(k) = \beta^k \frac{\Gamma(\alpha + k/\delta)}{\Gamma(\alpha)} \qquad k \in \mathbb{N}. \tag{D3}$$

Equating the first 3 moments ($k = 1, 2, 3$) of the generalized gamma to the first 3 moments of X gives α, β and δ. But, that requires to find the zeros of a nonlinear 3-dimensional function. We observed that root-finding algorithms may be sensitive to the choice of the first guess, and a particular attention must therefore be dedicated to it.

In Stacy and Mihram (1965), it is shown that, if Y follows a generalized gamma distribution, working with $\ln(Y)$ allows to find easily the parameters α, β, δ. Indeed, it only requires a root-finding for a monotonic unidimensional function. Unfortunately, the distribution of the logarithm of a linear combination of chi-square distributions is not known. We thus use the 2-moment approximation, for which we can find the moments of the logarithm of the distribution. Indeed, if we write $Y \overset{d}{=} g\chi^2_M$, in which g and M are determined from Eq. (98), and $Z = \ln(Y)$, some calculus gives us the cumulant generating function of Z:

$$K(t) = t\ln(2g) + \ln(\Gamma(M/2 + t)) - \ln(\Gamma(M/2)), \tag{D4}$$

from which we obtain the cumulants. The first three are

$$\kappa(1) = \ln(2g) + \psi_0(M/2), \tag{D5a}$$

$$\kappa(2) = \psi_1(M/2), \tag{D5b}$$

$$\kappa(3) = \psi_2(M/2), \tag{D5c}$$

where ψ_i is the polygamma function (ψ_0 is the digamma function). From the cumulants, we have the expected value $\kappa(1)$, the variance $\kappa(2)$, and the skewness $\kappa(3)/\kappa(2)^{3/2}$. Applying Eq. (21) of Stacy and Mihram (1965) gives us the parameters α_0, β_0, δ_0 for Y, parameters that we then use as a first guess for the generalized-gamma approximation of X.

D2 d-moment approximation

We extend here the formulas[10] presented in Provost et al. (2009). Let f_X be the PDF of X. f_X is approximated by the PDF of a d^{th} degree generalized gamma-polynomial:

$$f_X(x) \approx \gamma_{\alpha,\beta,\delta}(x) \sum_{i=0}^{d} \xi_i x^i, \qquad x \geq 0, \tag{D6}$$

where the parameters α, β and δ are estimated with the above 3-moment approximation. ξ_0, ..., ξ_d are the solution of Eq. (100), where $\eta(h) = \beta^h \Gamma(\alpha + h/\delta)/\Gamma(\alpha)$. The estimation of a confidence level for the WOSA periodogram is then the solution c_0 of

$$\frac{1}{\Gamma(\alpha)} \sum_{i=0}^{d} \xi_i \beta^i \gamma(i/\delta + \alpha, (c_0/\beta)^\delta) - p = 0, \tag{D7}$$

for some p-value p, e.g. $p = 0.95$ for a 95 % confidence level. If we pose $\delta = 1$, the *generalized gamma-polynomial* approximation reduces to the *gamma-polynomial* approximation presented in Sect. 5.3.3.

[10] In Provost et al. (2009), formulas are given for the *gamma*-polynomial, but as suggested by the authors, they can easily be generalized to the *generalized gamma*-polynomial.

Appendix E: Computing time: Analytical versus Monte-Carlo significance levels

A comparison between the computing times, for generating the WOSA periodogram, with the analytical and with the MCMC significance levels, based on the hypothesis of a red noise background, are presented on Fig. E1. They are expressed in function of the number of data points, which are disposed on a regular time grid, in order to make a meaningful comparison. Confidence levels with the analytical approach are estimated with a 10-moment approximation, and the number of samples for the MCMC approach is 10000 for the 95^{th} percentiles and 100000 for the 99^{th} percentiles. The other parameters are default parameters of WAVEPAL. All the runs were performed on the same computer[11].

We see that the analytical approach is faster than the MCMC approach as long as the number of data points is below some threshold, the latter increasing with the level of confidence. Indeed, the analytical approach delivers computing times of the same order of magnitude whatever is the percentile (the two blue curves in Fig. E1a and E1b are in the same order of magnitude), unlike the MCMC approach, which must require more samples as the level of confidence increases, in order to keep a sufficient accuracy. The difference between both computing times therefore increases as the level of confidence increases.

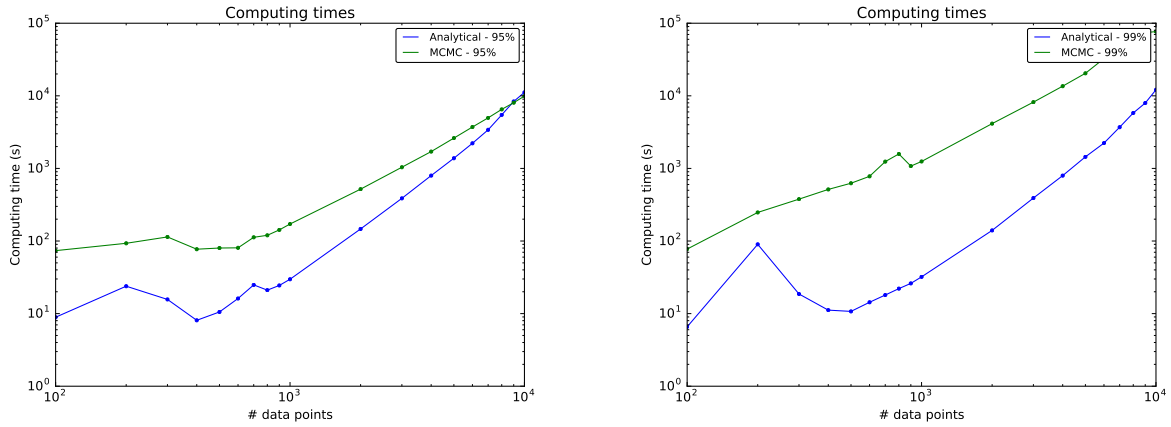


Figure E1. Computing times for generating the WOSA periodogram with analytical (blue) and MCMC (green) confidence levels, in function of the number of data points (disposed on a regular time grid). Log-log scale. Left: 95^{th} percentiles. Right: 99^{th} percentiles.

Appendix F: On the F-periodogram

The formula of the F-periodogram (Eq. (104)) is based on Brockwell and Davis (1991, pp. 335-336). In that book, the authors work with a constant trend. We have generalized the formula in order to deal with a polynomial trend.

A slightly different formula was published in Heck et al. (1985, p. 65), again with a constant trend. The F-periodogram is

[11] CPU type: SandyBridge 2.3 GHz. RAM: 64GB.

denoted by θ_F in their paper. In the case of a generalization to a polynomial trend, their formula becomes

$$\frac{(N-2)||(P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle\}})|X\rangle||^2}{2||[\mathbb{I} - (P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle,|c_\omega\rangle,|s_\omega\rangle\}} - P_{\overline{\text{sp}}\{|t^0\rangle,|t^1\rangle,\ldots,|t^m\rangle\}})]|X\rangle||^2}, \tag{F1}$$

but, unlike Eq. (104), it has a denominator which is not invariant with respect to the parameters of the trend.

Competing interests. The authors declare that they have no conflict of interest.

5 *Acknowledgements.* The authors are very grateful to R. Donner, L. Jacques, and S. Nicolay, for their comments on a preliminary version of the manuscript. This work is supported by the Belgian Federal Science Policy Office under contract BR/12/A2/STOCHCLIM.

References

Akaike, H.: A new look at the statistical model identification, IEEE Transactions on Automatic Control, 19, 716–723, doi:10.1109/TAC.1974.1100705, 1974.

Bretthorst, L.: Nonuniform Sampling: Bandwidth and Aliasing, in: Bayesian Inference and Maximum Entropy Methods in Science and Engineering - AIP Conference Proceedings, edited by Rychert, J., Gary, E., and Smith, R., pp. 1–28, 2001.

Brockwell, P. and Davis, R.: Time Series: Theory and Methods, Springer Series in Statistics, Springer, ISBN: 978-1-4419-0319-8, Second edn., 1991.

Brockwell, P. and Davis, R.: Introduction to Time Series and Forecasting, Springer Texts in Statistics, Springer International Publishing, ISBN: 978-3-319-29854-2, Third edn., doi:10.1007/978-3-319-29854-2, 2016.

Bronez, T.: Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences, IEEE Transactions on Acoustics, Speech and Signal Processing, 36, 1862–1873, doi:10.1109/29.9031, 1988.

Ferraz-Mello, S.: Estimation of Periods from Unequally Spaced Observations, The Astronomical Journal, 86, 619–624, doi:10.1086/112924, 1981.

Fodor, I. and Stark, P.: Multitaper spectrum estimation for time series with gaps, IEEE Transactions on Signal Processing, 48, 3472–3483, doi:10.1109/78.887039, 2000.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, Reviews of Geophysics, 40, 1003, doi:10.1029/2000RG000092, http://dx.doi.org/10.1029/2000RG000092, 2002.

Harris, F.: On the use of windows for harmonic analysis with the discrete Fourier transform, Proceedings of the IEEE, 66, 51–83, doi:10.1109/PROC.1978.10837, 1978.

Hasselmann, K.: Stochastic climate models Part I. Theory, Tellus, 28, 473–485, doi:10.1111/j.2153-3490.1976.tb00696.x, http://dx.doi.org/10.1111/j.2153-3490.1976.tb00696.x, 1976.

Heck, A., Manfroid, J., and Mersch, G.: On period determination methods, Astronomy & Astrophysics Supplement Series, 59, 63–72, 1985.

Jeffreys, H.: An Invariant Form for the Prior Probability in Estimation Problems, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 186, 453–461, doi:10.1098/rspa.1946.0056, http://rspa.royalsocietypublishing.org/content/186/1007/453, 1946.

Jian, Z., Zhao, Q., Cheng, X., Wang, J., Wang, P., and Su, X.: Pliocene-Pleistocene stable isotope and paleoceanographic changes in the northern South China Sea, Palaeogeography, Palaeoclimatology, Palaeoecology, 193, 425–442, doi:http://dx.doi.org/10.1016/S0031-0182(03)00259-1, http://www.sciencedirect.com/science/article/pii/S0031018203002591, 2003.

Jones, R. and Ackerson, L.: Serial correlation in unequally spaced longitudinal data, Biometrika, 77, 721–731, doi:10.1093/biomet/77.4.721, http://biomet.oxfordjournals.org/content/77/4/721.abstract, 1990.

Kelly, B., Becker, A., Sobolewska, M., Siemiginowska, A., and Uttley, P.: Flexible and Scalable Methods for Quantifying Stochastic Variability in the Era of Massive Time-domain Astronomical Data Sets, The Astrophysical Journal, 788, 33, http://stacks.iop.org/0004-637X/788/i=1/a=33, 2014.

Kemp, D.: Optimizing significance testing of astronomical forcing in cyclostratigraphy, Paleoceanography, doi:10.1002/2016PA002963, http://dx.doi.org/10.1002/2016PA002963, 2016PA002963, 2016.

Lenoir, G.: Multitaper spectral estimation for the continuous wavelet transform: a general numerical approach and an application to the Morlet wavelet, in prep., 2017.

Lenoir, G. and Crucifix, M.: A general theory on frequency and time-frequency analysis of irregularly sampled time series based on projection methods. II. Extension to time-frequency analysis, Nonlinear Processes in Geophysics Discussions, https://doi.org/10.5194/npg-2017-27, in review, doi:10.5194/npg-2017-27, https://www.nonlin-processes-geophys-discuss.net/npg-2017-27/, 2017.

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

Mortier, A., Faria, J. P., Correia, C. M., Santerne, A., and Santos, N. C.: BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram, Astronomy & Astrophysics, 573, A101, doi:10.1051/0004-6361/201424908, http://dx.doi.org/10.1051/0004-6361/201424908, 2015.

Mudelsee, M.: Climate Time Series Analysis - Classical Statistical and Bootstrap Methods, vol. 42 of *Atmospheric and Oceanographic Sciences Library*, Springer Netherlands, ISBN: 978-90-481-9481-0, 2010.

Mudelsee, M., Scholz, D., Röthlisberger, R., Fleitmann, D., Mangini, A., and Wolff, E. W.: Climate spectrum estimation in the presence of timescale errors, Nonlinear Processes in Geophysics, 16, 43–56, doi:10.5194/npg-16-43-2009, http://www.nonlin-processes-geophys.net/16/43/2009/, 2009.

Pardo Igúzquiza, E. and Rodríguez Tovar, F.: Spectral and cross-spectral analysis of uneven time series with the smoothed Lomb-Scargle periodogram and Monte Carlo evaluation of statistical significance, Computers & Geosciences, 49, 207–216, doi:10.1016/j.cageo.2012.06.018, http://www.sciencedirect.com/science/article/pii/S0098300412002130, 2012.

Priestley, M.: Spectral Analysis and Time Series, Two Volumes Set, Probability and Mathematical Statistics - A series of Monographs and Textbooks, Academic Press, ISBN: 0-12-564922-3, Third edn., 1981.

Provost, S.: Moment-Based Density Approximants, The Mathematica Journal, 9, 727–756, http://www.mathematica-journal.com/issue/v9i4/DensityApproximants.html, 2005.

Provost, S., Ha, H.-T., and Sanjel, D.: On approximating the distribution of indefinite quadratic forms, Statistics, 43, 597–609, doi:10.1080/02331880902732123, http://dx.doi.org/10.1080/02331880902732123, 2009.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics, 18, 389–404, doi:10.5194/npg-18-389-2011, http://www.nonlin-processes-geophys.net/18/389/2011/, 2011.

Riedel, K. and Sidorenko, A.: Minimum bias multiple taper spectral estimation, IEEE Transactions on Signal Processing, 43, 188–195, doi:10.1109/78.365298, 1995.

Robinson, P.: Estimation of a time series model from unequally spaced data, Stochastic Processes and their Applications, 6, 9–24, doi:http://dx.doi.org/10.1016/0304-4149(77)90013-8, 1977.

Scargle, J.: Studies in astronomical time series analysis II - Statistical aspects of spectral analysis of unevenly spaced data, The Astrophysical Journal, 263, 835–853, doi:10.1086/160554, http://adsabs.harvard.edu/abs/1982ApJ...263..835S, 1982.

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

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

Stacy, E. W. and Mihram, G. A.: Parameter Estimation for a Generalized Gamma Distribution, Technometrics, 7, 349–358, doi:10.1080/00401706.1965.10490268, http://www.tandfonline.com/doi/abs/10.1080/00401706.1965.10490268, 1965.

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

Torrence, C. and Compo, G.: A Practical Guide to Wavelet Analysis, Bulletin of the American Meteorological Society, 79, 61–78, doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2, http://dx.doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2, 1998.

Torrésani, B.: Analyse continue par ondelettes, Savoirs actuels/Série physique, CNRS Editions and EDP Sciences, ISBN: 978-2-271-05364-0, 1995.

Uhlenbeck, G. E. and Ornstein, L. S.: On the Theory of the Brownian Motion, Physical Review, 36, 823–841, doi:10.1103/PhysRev.36.823, http://link.aps.org/doi/10.1103/PhysRev.36.823, 1930.

Vio, R., Andreani, P., and Biggs, A.: Unevenly-sampled signals: a general formalism for the Lomb-Scargle periodogram, Astronomy & Astrophysics, 519, A85, doi:10.1051/0004-6361/201014079, http://dx.doi.org/10.1051/0004-6361/201014079, 2010.

Walden, A. T.: A unified view of multitaper multivariate spectral estimation, Biometrika, 87, 767–788, doi:10.1093/biomet/87.4.767, http://biomet.oxfordjournals.org/content/87/4/767.abstract, 2000.

Welch, P.: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics, 15, 70–73, doi:10.1109/TAU.1967.1161901, 1967.

Zechmeister, M. and Kürster, M.: The generalised Lomb-Scargle periodogram, Astronomy & Astrophysics, 496, 577–584, doi:10.1051/0004-6361:200811296, http://dx.doi.org/10.1051/0004-6361:200811296, 2009.