We are IntechOpen,
the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Periodogram Analysis under the Popper-Bayes Approach

George Caminha-Maciel

Abstract

In this chapter, we discuss the use of the Lomb-Scargle periodogram, its advantages, and pitfalls on a geometrical rather than statistical point of view. It means emphasizing more on the transformation properties of the finite sampling – the available data – rather than on the ensemble properties of the assumed model statistical distributions. We also present a brief overview and criticism of recent literature on the subject and its new developments. The whole discussion is under the geophysical inverse theory point of view, the Tarantola's combination of information or the so-called Popper-Bayes approach. This approach has been very successful in dealing with large ill-conditioned, or under-determined complex problems. In the case of periodogram analysis, this approach allows us to manage more naturally the experimental data distributions and its anomalies (uncorrelated noise, sampling artifacts, windowing, aliasing, spectral leakage, among others). Finally, we discuss the **Lomb-Scargle-Tarantola (LST) periodogram**: an estimator of spectral content existing in irregularly sampled time series that implements these principles.

Keywords: Lomb-Scargle, periodogram, irregular sampling, inverse theory, spectral analysis, cyclostratigraphy, paleoclimatology, pattern recognition

1. Introduction

Although being old, periodogram analysis until nowadays represents the main workhorse for the studying of irregularly sampled time series from a vast majority of scientific branches. Since its introduction more than a century ago by A. Schuster, the periodogram has evolved and gained widespread use, even if sometimes without a complete understanding of its more subtle aspects. Its popularity comes from its relatively simple statistical behavior, easy implementation, and easy interpretation of the results. In summary, the Lomb-Scargle periodogram is an estimation method that emulates the power of Fourier decomposition – in a case when it is not possible to apply it – for data series irregularly sampled in time.

Being unnecessary to advocate for Fourier analysis, we want here to remember one of its main advantages – its *simplicity*. Fourier basis, sines and cosines, are the most basic functions that exhibit periodic behavior. Then it is very natural to use a Fourier basis to compare and detect periodic patterns in experimental data.

Usually applied in areas as diverse as astronomy, biology, meteorology, ocean-ography, and cyclostratigraphy [1–7], the Lomb-Scargle periodogram has not, however, a unique direct rule of use. We should always consider the subtle

differences in the time series from each of these areas to explore better the spectral content in the data set and to improve understanding of the results.

There are fundamental questions that still permeate the whole subject of periodogram analysis:

- What are the necessary conditions on the original continuous function for the periodogram to analyze the irregularly sampled time series?

- What is the relationship between the Lomb-Scargle periodogram [8–14] and the Discrete Fourier Transform (DFT)?

- What is the appropriate discrete domain of frequency numbers for which an irregularly sampled time series has information? What is the minimum frequency allowable? Is there a maximum frequency (Nyquist limit) to the analysis of an unevenly sampled time series, and which it would be? What is the proper density of frequency points?

- What is the source of the several spurious peaks arising in the periodogram, besides the original peak on the proper periodicity frequencies?

- What is the uncertainty in the frequency of the periodicity found?

In this chapter, we discuss these questions and present a Popper-Bayes point of view of the periodogram, comparing it with the more traditional approach. See [15] for a comprehensive review (see also https://jakevdp.github.io/blog/2017/03/30/practical-lomb-scargle/). The traditional approach of the Lomb-Scargle method was developed mainly by astronomers and was adapted to applying to the characteristics of the astronomical data. Here we point out that the techniques devoted to astronomy are non-unique and not necessarily appropriate to other areas. We also show examples from cyclostratigraphy, our subject of study, which has some typical sampling anomalies.

Then we will present the **Lomb-Scargle-Tarantola (LST) periodogram** [16, 17], a more general technique for the use of the periodogram. The LST periodogram applies the Popper-Bayes perspective to the periodogram of irregularly sampled time series: incorporate the *a priori* variance (sampled data with all of its anomalies) directly into the *a posteriori* (periodogram) variance, and analyze the ill-defined, possibly multi-modal, complex obtained distribution.

2. The Popper-Bayes approach to inverse problems

In geophysics, it is usual to deal with high-dimensional and ill-conditioned problems. This happens because geophysicists are always trying to understand and image subsurface structures with data generally obtained from the surface. Furthermore, the measurements we want to interpret are, in general, indirectly related to the structure we want to model.

In gravimetry, for example, we get measurements of the gravitational field at a spatial grid on the surface and try to figure out what possible subsurface density anomalies could produce the observed gravitational field anomaly. This problem is highly ill-conditioned since an infinite number of configurations of subsurface bodies and density contrasts could originate the same set of field measurements at the surface. It cannot be solved without adding an *a priori* information or some kind of regularization.

To solve this class of problems, geophysicists developed statistical techniques collectively called *inverse theory*. In the inverse theory, we deal with two main difficulties: first, to find at least one model that satisfies the measurements; second, to qualify the set of obtained models. Usually, we find the **best model** approach to attack this problem – where we pick a model from a subset of possibilities by maximizing some kind of measure over the whole set of possibilities. After that, we attach some uncertainty to the best model chosen.

This approach is favored by statisticians and can be mathematically formalized. Generally, it works well when the involved variables, data and noise, follow regular statistical distributions. However, in the case of very irregularly sampled short time series, this assumption departs from reality.

2.1 Progressing by falsification

A Bayesian attack on the problem would be by posing the following question: how does this newly obtained data set modify our previous knowledge about the system?

Furthermore, it means to think about what we already know about the system – usually put in the form of probability distributions on the dynamical variables – and how to incorporate new information through the means of the constitutive equations on the dynamics. After what, we arrive on the *a posteriori* distributions (or models) "assimilating" the new data.

Here we present a more radical idea of physical inference, which is called the Popper-Bayes approach, which departs entirely from the idea of finding the best model, the mean model, or the maximum likelihood model. Professor Albert Tarantola started this idea: observations might not be used to produce models; they should be used only to falsify models [18–20].

He proposed that physical inference could be set in principle as:

1. Using the available *a priori* information to create all possible models on the system – potentially an infinite number of them;

2. For each model, solve the direct problem – assuming as true, calculate a measure (or probability) for this model in comparison with the actual observations;

3. Use some criteria to define which models are acceptable based on these measures (or probabilities) and the physical theory on the system. The unacceptable models should be dropped or falsified;

4. The set of surviving models constitute **the solution** for the physical inference. Uncertainties on these models should consider the properties of these *a posteriori* distributions over the variables subspace.

Then we have a natural interpretation of multi-modal probability distributions or ill-defined final models – what is very useful in periodogram analysis of unevenly sampled times series.

3. Periodogram analysis of irregularly sampled time series

What are the needed assumptions for properly analyzing a continuous signal by the Fourier method through a discrete sampling?

First of all, it has to be a *single function* of the time variable t (t can also be a spatial variable). It means that the signal needs to be a unique sequence of values $(x(t), t)$, where, for each t, there is one and only one assigned value $x(t)$. Besides that, the values $x(t)$ cannot "explode" (as the exponential function) – they all have to be limited by two real numbers: a maximum and a minimum. Furthermore, the function $x(t)$ cannot oscillate too fast; it has to be relatively smooth. This last condition means that the function cannot have "jumps"; in Fourier analysis, this would mean that the function has limited informational content or is "bandlimited." Our discussion here uses real-valued functions, but the same ideas can be easily extended for complex-valued cases.

The Fourier transform is a mathematical tool that for which a real function $x(t)$ relates another, a complex function – $X(f)$. As a complex number, each $X(f)$ value can be described by a pair of real numbers – an amplitude $A(f)$ and a phase $\theta(f)$. These values $A(f)$ represent the relative importance of each time scale T ($f = 1/T$). The squared values $A^2(f)$ are proportional to the *relative energy* of each of the frequencies f in the signal. The function that gives the relative energy of each of these frequency components is called *power spectral density (PSD)* $P_X(f)$, where $P_X(f) = A^2(f)$. It is usual, in the literature, to find these expressions in terms of the variable ω, called *angular frequency*, instead of f, where $\omega = 2\pi f$.

3.1 The periodogram

The Fourier transform has an analytical form applicable to continuous functions and also has a discrete form applicable to discrete functions – as the sampled time series we intend to study. That one is called *Discrete Fourier Transform (DFT)*. Being computationally intensive, particularly for large data sets, the DFT has a fast implementation called *Fast Fourier Transform (FFT)* algorithm. The FFT algorithm dramatically reduces the time and the computational cost of calculating Fourier transforms for real-time series. It is worth mentioning that if the time series is irregularly sampled, the FFT algorithm cannot be applied.

From the DFT, we can obtain estimates for the PSD of experimental real-time series. The statistics that gives an estimation of the relative energy among the different frequencies present in a signal is called the *periodogram*.

The classical periodogram is simply the squared modulus of the DFT. In the exponential form, it can be written as

$$P_X(\omega) = \frac{1}{N_0} \left| \sum_j \left(x_j \cdot e^{-i\omega t_j} \right) \right|^2 \tag{1}$$

Or in the trigonometric (equivalent) form as

$$P_X(\omega) = \frac{1}{N_0} \left[\left(\sum_j x_j \cos \omega t_j \right)^2 + \left(\sum_j x_j \sin \omega t_j \right)^2 \right] \tag{2}$$

where N_0 is the number of data points in the time series.

One main statistical property of the classical periodogram is that for a time series constituted solely of evenly spaced Gaussian noise, the values of the periodogram are exponentially distributed. Unfortunately, when the time series is irregularly sampled in time, this property no longer holds. This statistical behavior also only applies to observations of uncorrelated white noise.

Lomb [9] and Scargle [10] addressed the problem of finding a generalized form of periodogram in order to:

- Reduces to the classical form in the case of uniformly sampled time series;

- Its statistics are computable;

- It is invariant to global time-shifts in the series.

The classical periodogram is very noisy even for time series only slightly noisy. The Scargle's modified periodogram, called *Lomb-Scargle periodogram*, is much smoother and differs from the classical periodogram in at least two aspects:

- It adds a time-shift term τ. This time-shift is calculated to minimize independence between the two trigonometric basis $(\sin \omega t_j)$ and $(\cos \omega t_j)$. In other words, it minimizes the crossing term $\left(\sum_j \sin \omega t_j \cos \omega t_j\right)$. It is an attempt to improve orthogonality in the equations.

- It adds the denominators $\left(\sum_j \sin^2 \omega(t_j - \tau)\right)$ and $\left(\sum_j \cos^2 \omega(t_j - \tau)\right)$ to the terms in periodogram. These denominators differ from $N_0/2$, which is the expected value in the limiting case of complete phase sampling at each frequency (as in the uniformly sampled series).

The Lomb-Scargle periodogram is given by

$$P_X(\omega) = \frac{1}{2}\left\{\frac{\left[\sum_j x_j \cos \omega(t_j - \tau)\right]^2}{\sum_j \cos^2 \omega(t_j - \tau)} + \frac{\left[\sum_j x_j \sin \omega(t_j - \tau)\right]^2}{\sum_j \sin^2 \omega(t_j - \tau)}\right\} \tag{3}$$

where τ is given by

$$\arctan(2\omega\tau) = \frac{\left(\sum_j \sin 2\omega t_j\right)}{\left(\sum_j \cos 2\omega t_j\right)} \tag{4}$$

With this formulation, the periodogram of uncorrelated irregularly spaced Gaussian noise is also exponentially distributed (sum of squares of two zero-mean Gaussian variables).

3.1.1 The least-squares periodogram and its extensions

The Lomb-Scargle periodogram is equivalent to the least-squares fitting of a sinusoidal model to the data at each frequency ω. The Lomb-Scargle periodogram power relates to the $\chi^2(\omega)$ goodness-of-fit, at the frequency ω.

Let us consider a sinusoidal model at the frequency ω,

$$y(t; \omega) = A_\omega \sin(\omega(t - \phi_\omega)) \tag{5}$$

The χ^2 goodness-of-fit can be defined as

$$\chi^2(\omega) \equiv \sum_j \left(y_j - y(t_j; \omega)\right)^2 \tag{6}$$

We can find the "best" model $\hat{y}(t; \omega)$ by minimizing χ^2. Let $\hat{\chi}^2$ be the minimum and $\left(\hat{A}_\omega; \hat{\phi}_\omega\right)$ the optimal value, then we can write

$$P_{LS}(\omega) \sim \hat{A}_\omega^2 \tag{7}$$

For data sets with errors, we can consider introducing them into the periodogram. Vio and others [16, 17] studied a more general model, including a $N \times N$ error covariance matrix Σ, for N timely observations.

$$\chi^2(\omega) = \left(\vec{y} - \vec{y}_{model}\right)\Sigma^{-1}\left(\vec{y} - \vec{y}_{model}\right) \tag{8}$$

In the case of uncorrelated zero-mean colored noise, this expression reduces to

$$\chi^2(\omega) = \sum_j \left(\frac{y_j - y(t_j; \omega)}{\sigma_j^2}\right)^2 \tag{9}$$

where σ_j^2 are the gaussian errors.

For practical applications, there are some additional issues to consider when introducing data errors in the periodogram calculations, such as unaccountable uncertainties in error estimates, correlated noise, and the dependence in the signal slope. All of that makes the use of error estimates not very advisable [13].

3.2 Periodograms and significance

The Lomb-Scargle periodogram keeps most of the optimal analytical properties of the Fourier transform and its power spectrum:

- Linearity;

- The periodogram of a pure sinusoidal at ω_0 is a sum of the Dirac delta functions at $(\pm\omega_0)$;

- The periodogram, just as the power spectrum, is insensitive to translations in time (only reflects on the phase spectrum);

- It is a real-valued even function (that is the reason why we only calculate the positive part);

- The transform of a gaussian is another (different) gaussian;

- The "Heisenberg uncertainty principle" of Fourier transforms (usual in quantum mechanics) applies: a narrow feature in time becomes a broader peak in frequency and vice-versa.

3.2.1 Spectral windows

The pointwise product of the underlying infinite periodic signal with a rectangular window function usually describes the observed signal; its length is the time series duration T.

In the Fourier transform, the windowing replaces each Dirac delta function at some frequency ω_i with a *sinc* function centered at that same frequency ω_i, in the Fourier transform. This behavior is a direct consequence of the inverse relationship between the time window width and the width of its Fourier transform.

An infinite $\sin(\omega_i t)$ function runs from $-\infty$ to $+\infty$ and has two delta functions at $\pm \omega_i$ as its Fourier transform. The finite signal, the windowing version, has a broader transform – the *sinc* function. The *sinc* function, besides having a broader and lower *central peak* (delta function has infinite height), also has *side lobes*. It spreads the power at the frequency ω_i to the adjacent frequencies. This phenomenon is called *spectral leakage*, and it is more pronounced as the shorter is the duration of the time series (narrower window).

There is another essential aspect of spectral windows: its smoothness. The more abrupt (less smooth) is the window, the more spectral leakage happens in the Fourier transform, lowering the central peak and heightening the side lobes. Instead of a rectangular function, we can use a smoother function, like a $\sin$ bell function, for example, and the resulting spectrum will exhibit much less leakage.

Though we made this discussion about the spectrum power, the Fourier transform squared modulus, it equally applies to its actual estimate from data – the periodogram.

3.2.2 Frequencies for periodogram analysis

There are three parameters to consider when choosing the appropriate frequency grid for periodogram analysis of a particular time series:

1. The minimum frequency, f_{min};

2. The maximum frequency, f_{max};

3. The frequency spacing, Δf.

Recommendations for the choice of each of these parameters vary in the literature. Here we discuss some points to consider in the case of regularly as well as to irregularly sampled time series.

Minimum frequency: The minimum frequency, f_{min}, is the easiest to define. It relates to the largest period of a wave we can investigate in the time series. We usually set it as the inverse of T – the length of the time series, or as zero – where its value virtually equals to the frequency spacing, Δf.

Maximum frequency: The maximum frequency, f_{max}, represents the shortest period of a wave we can investigate in the time series. For evenly sampled time series, the *Nyquist theorem or Sampling theorem* defines this maximum frequency – called *Nyquist frequency*, $f_{Nyquist}$.

This theorem states that if we have a regularly sampled function with the sampling rate of $f_\delta = 1/\delta t$, we can only recover full frequency information if the signal is *band-limited* between frequencies $f_\delta/2$. This theorem states that if we have a regularly sampled function with the sampling rate of $f_\delta = 1/\delta t$, we can only recover full frequency information if the signal is *band-limited* between frequencies $\pm f_\delta/2$.

Putting in another way, the theorem says that to fully represent the content of a band-limited signal whose Fourier transform is zero outside the range of $\pm B$, we must sample the signal with a rate at least $f_\delta = 2B$. Then, for evenly sampled time series, $f_{max} = f_{Nyquist} = f_\delta/2$.

Frequency spacing: The frequency spacing, Δf, has only general guidelines: too small frequency spacing can lead to unnecessarily long computation times, which adds up fastly for large data sets. Too coarse frequency spacing can risk missing narrow peaks in the periodogram – which would fall between adjacent grid points. However, there is a controversy when considering these frequency grids as independent points when applying statistical significance tests in the periodogram ordinates (testing for true periodicities).

An evenly sampled time series represents a pointwise product of the original continuous signal with a sequence of Dirac delta functions (a *Dirac comb*) at the sampling times. The Nyquist limit is a direct consequence of the symmetry in this Dirac comb window. Beyond this limit, the spectrum becomes a periodic repetition of itself – that is why the periodogram is unique between the limits $\pm f_{Nyquist}$. The rise of power in the spectrum beyond the Nyquist limit is called *aliasing* since these peaks are not real but "alias" of the real power inside the Nyquist interval in the original signal.

For unevenly sampled time series:

- The structure in the observing window can lead to *partial aliasing* in the periodogram;

- The non-structured spacing of observations also leads to the arising of *non-structured peaks* in the window transform;

- The maximum frequency limit *might or might not exist*, and if it exists, it tends to be *far higher* than for the evenly sampled case.

For irregularly sampled time series, if there is a *periodic pattern in the observation times gaps*, this can lead to a peak in the periodogram indicating a periodicity. For example, the daily pattern of measurements in astronomy: an observation in time t_0 is likely to be followed by other observation only at time $t_0 + np$ (p is an integer number of days, and n is an integer). Therefore, it can generate a peak at the frequency p in the periodogram.

We find in the literature some proposals for the maximum frequency (Nyquist-like) limit for irregular sampling [15, 21–26]. These estimates are easy to calculate and reduce to the Nyquist frequency limit in the evenly sampled case:

- Arithmetic average of the time intervals.

- Harmonic average of the time intervals.

- Median of the time intervals.

- Minimum of the time intervals.

There are also, in the literature, some Nyquist-like limits based on not-so-simple statistics of the time intervals [15]:

- Greatest common divisor (gcd) of the time intervals. We should consider sampling times as integer numbers (what can always be done by re-scaling the time values).

- Frequency limit due to *time windowing* when observations are not pointwise instantaneous, but instead, they consist of *short-time (δt) integrations* of the

continuous signals, around the sampling times t_j. This kind of sampling is typical in several applications, including cyclostratigraphy. In that case, again, the Fourier transform is the product of the original signal transform and the transform of the time window, which has the width proportional to $1/\delta t$. Then, $f_{max} = f_{Nyquist} = 1/(2\delta t)$. However, in this case, this frequency limit does not imply aliasing. Instead, it is about a frequency limit beyond which all signal is *attenuated to zero*.

- Frequency limit based on *a priori* knowledge on the expected signals.

Finally, for irregularly sampled time series, the maximum frequency limit can be set by the *precision of time measurements*.

3.2.3 Statistics of the periodogram

The classical periodogram has a fundamental statistical property for evenly sampled time series: when the signal consists solely of pure Gaussian noise, the values of the periodogram are exponentially distributed – for irregularly sampled times, this property no longer holds.

The Scargle's generalized form of the periodogram brings back that statistical simplicity for the irregularly sampled case: for time series consisting solely of pure Gaussian noise, the unnormalized periodogram has its ordinates exponentially distributed.

This statistical property is used to test for what would be a "true" periodicity in periodogram ordinates. The standard procedure is to assume that the periodogram maximum ordinate represents a true periodicity, called *Fisher criteria*, and to test this value against all others ordinates – supposedly arising from the background noise.

Scargle defined a *False Alarm Probability (FAP)* that, based on the assumed distribution of Gaussian noise, simply measures the probability that a time series without any signal would arise, due to stochastic fluctuations only, an ordinate of the observed magnitude in the periodogram. Following Scargle [15], the *detection threshold*, z_0, is a magnitude level above which, if we claim that a peak is due a real signal, we would only be wrong a small fraction p_0 (FAP) of the time:

$$z_0 = -\ln\left[1 - (1 - p_0)^{1/N}\right], \tag{10}$$

where p_0 (FAP) is a small number, and N is the number of *independent* frequencies tested.

It is worth noting that this statistical analysis answers the question: "What is the probability that a time series without any periodic component would make arise a peak of that magnitude in the periodogram?" It *does not* answer the utterly more physically significant, more direct question: "*What is the probability that this periodogram feature comes from a periodic phenomenon?*"

The ability to analytically quantify the relationship between peak height and statistical significance of a feature in the periodogram has been one of the main reasons for the widespread use of the Lomb-Scargle periodogram [10–13, 23–25]. However, the independence of the tested frequencies remains an open issue.

Data quality (and quantity) generally reflects on the peak height related to the background noise, which gives peak *significance*, as discussed above. Neither the number of points in a time series or the signal-to-noise ratio affects the peak frequency determination nor its *precision*. The uncertainty in the frequency value of

a peak is related to the peak width, usually in Fourier analysis defined as the peak half-width. For this reason, in periodogram analysis, Gaussian error bars should be avoided as a way to report uncertainties in frequency determinations.

4. The Lomb-Scargle-Tarantola (LST) periodogram

In geophysical inverse theory, there are high mature methods to deal with large uncertainties in data and ill-conditioned models. It comes from the constant necessity; geophysicists have to build and evaluate physical models of the subsurface (or Earth's deep interior) based mostly on data acquired from the surface.

In Seismics, for example, geophysicists developed a procedure called *stacking* where a set of different signals, acquired by distinct geophones, are gathered under specific geometrical settings. The procedure aims to amplify the signal and to cancel noise, improving overall information about the same subsurface points. The stacking of seismic signals is one of the main reasons for the success of this technique on the oil industry nowadays, allowing the discovery of new oil fields and improving oil/gas recovering on reservoir structures.

Caminha-Maciel and Ernesto [16, 17] created a method to analyze spectral content (and its uncertainties) in irregularly sampled times series applying some principles from the geophysical inverse theory – the Popper-Bayes approach, as developed by Albert Tarantola [18–20]. This approach uses freely normalizable probability distributions to encapsulate the information and afterward operates with these distributions (the data information and the model/geometrical information on the problem). The results of these operations constitute a Bayesian physical inference method, and its *a posteriori* probability distributions are proper solutions to our inverse problem. In the following section, we will see how this applies to the Lomb-Scargle periodogram.

4.1 State of information periodogram-based functions

There is a vast diversity of methods to analyze time series, both in time and in frequency domains. Fourier-derived methods still show continued interest since they are fast (due to FFT algorithm), intuitive, and have a straightforward extension to irregularly sampled time series.

However, in some applications, as the high-resolution deep-sea stratigraphic records, there are novel challenges for the extraction and interpretation of meaningful information. It is worth to mention the uncertainties in the measured times (plus the usual uncertainties on the other variables), non-stationarity of the dynamical systems observed, and shortness of the records compared to wavelengths of interest. In the stratigraphic time series, there is also a general dependence of the recording-sampling process on the unknown-climatic signal itself. These issues contribute to breaking the orthogonality among periodogram ordinates, which is necessary to perform appropriately statistical significance tests on supposed-independent ordinates.

Nevertheless, there are some less-known interesting properties of the Lomb-Scargle periodogram:

- Analytical independence among ordinates – non-statistical independence. Since we have a fixed set of points $(X(t_j), t_j)$ and a fixed frequency point ω_0, $P(\omega_0)$ gives the same result regardless of how many different frequencies ω_i we use in the analysis. The periodogram does not "see" other ordinates. The problem is when we try to compare different ordinates (statistical independence).

- The maximum frequency (Nyquist-like) allowed can be much higher than in the evenly sampled case. We can define a Nyquist-like frequency related to the inverse of the minimum time interval, or even the gcd of the time intervals [8].

- The signal-noise (S/N) ratio does not depend on the periodogram normalization factor. In other words, $P(\omega_a)/P(\omega_b)$ is independent of the chosen normalization factor (there are several different options in the literature).

- There is a low false-negative probability for the periodogram for a real signal present in the time series. Then, if we guess correctly a frequency ω_0 present in the time series, we have a very low probability of obtaining a small value for $P(\omega_0)$. The problem is that we also obtain high periodogram ordinates for other frequencies besides the true ones – this makes it difficult to identify them if we do not know beforehand.

In the LST periodogram, a freely normalized version of the Lomb-Scargle periodogram is initially defined over the broadest possible range of frequencies. The frequency set is chosen to include all wavelengths about which could exist information on the time series. The next step is to choose the minimum frequency as zero, the frequency grid spacing δf, or the total length of time series, or some fraction in between these values. We can choose the maximum frequency as the highest frequency about which we believe there is any information in the time series, as some *a priori* known Nyquist limit, up to the limit of the inverse of the gcd of time intervals. Above this limit, we probably have to deal we some folding in the periodogram. The frequency grid δf has to be chosen to allow the calculation to be computationally feasible.

4.1.1 Normalizing by the bandwidth total content

We define

$$P_{LST}(\omega) = K \cdot P_{LS}(\omega) \tag{11}$$

where the normalizing constant K is set as $K = \left(\sum_{\omega} P(\omega)\Delta\omega\right)^{-1}$. This values of the constant K is such that the function P_{LST} normalizes to total area under the curve equals to 1 over the whole set of frequencies ω_i. This area represents the total power in the bandwidth and reflect the times series total variance.

Note that the S/N ratio, as well as the ratio between any two distinct frequencies $P(\omega_a)/P(\omega_b)$, does not depend on the times series total power.

This procedure is equivalent to a stretching of the data series variable $X(t)$ in the time domain and also has the property of making comparable the total power of the various periodograms.

4.2 Periodogram analysis by combination of information

The two main ideas of the Lomb-Scargle-Tarantola periodogram are

1. Smoothing the periodogram.

2. Stacking independent periodogram estimates.

Since its proposition, the periodogram is recognized as high noisy statistics, even for less noisy data. Smoothing the periodogram is not a new idea. There are several

attempts in this direction by averaging adjacent estimates (as in Daniel's averaged periodogram, for example) and in many Bayesian formulations.

A new idea on working with periodograms is stacking in the frequency domain: we can consider two or more distinct time series with information about the same dynamical variable as independent observations of the same phenomenon. For example, several stratigraphic sections covering the same time interval; different variables related to the same dynamics – sediment accumulation and $\delta^{18}O$ time series, both related to seawater surface temperature; independent observations of the same variable – as astronomical observations from distinct geographic locations; and biological circadian rhythms from different organisms.

After that, we operate these periodogram distributions with two logical operators "OR" and "AND." The OR operator can be described as a generalization of doing histograms and is mathematically defined as the arithmetic average of the individual LST periodograms (very similar to usual stacking of seismic signals). The AND operator is a non-linear operator that represents the generalization of conditional probability and is mathematically defined as $\frac{\Pi_i P_i(\omega)}{\mu}$, where μ is the null information function – characterizing the geometry of the physical problem.

In some physical problems involving the dynamical variable *frequency*, f, itself, the null information function can be better written as $1/f$. However, in spectral analysis, the variable frequency only means labels for some general class of eigenfunctions – as the Fourier basis $\sin$ and $\cos$. Then we can fairly consider the null information function, μ, as a *constant function* over the entire domain.

4.2.1 Using LSTperiod software

We have published [17] a proof-of-concept software to implement the LST periodogram. This software, the *LSTperiod* (Download it at http://www.iag.usp.br/paleo/sites/default/files/LSTperiod-files.zip), exemplify one possible implementation of the LST periodogram. We have made a set of choices for the frequency grid, normalizations of the state of information functions, evaluation of the resulting models (candidates periodicities), and visualizations of the results. Those choices are for no means unique [27].

To illustrate the use of the LSTperiod software, we show a set of five stratigraphic series of benthic $\delta^{18}O$ from the sedimentary core drilled by the Ocean
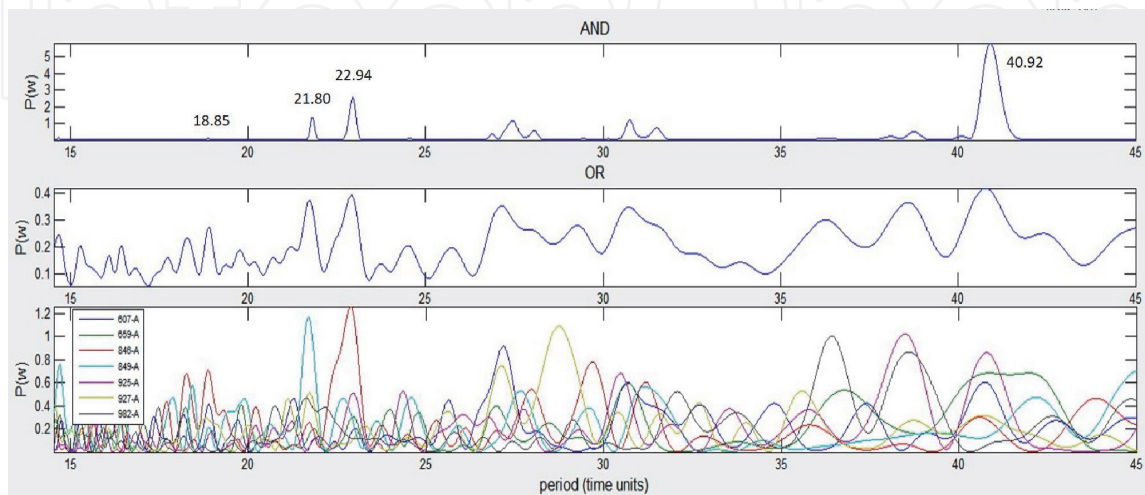

Figure 1.
LST periodograms for the benthic $\delta^{18}O$ series from ocean drilling project (ODP) cores. The windows show the periodogram calculated for each data file separately (bottom) and the combined results for the OR (middle) and AND (top) spectra.
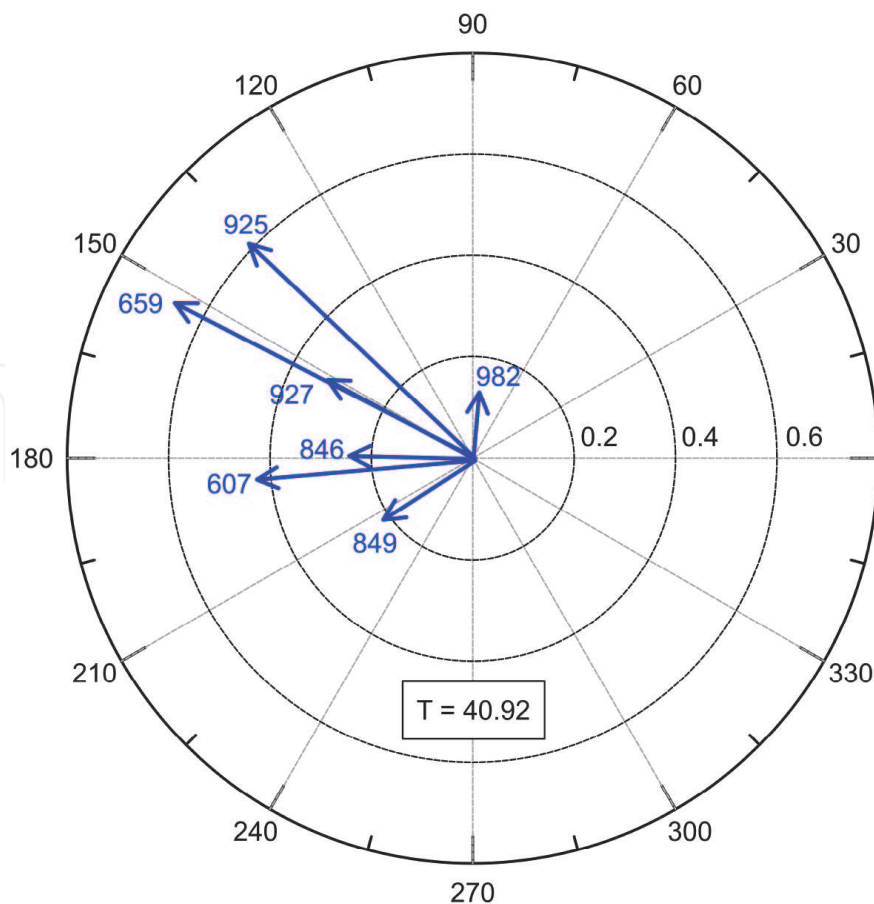
Figure 2.
Rose diagram showing the phases and amplitudes of the calculated period T = 40.92kyr for each analyzed series.

Drilling Project (ODP) [28]. These time series were subjected to spectral analysis and other statistical methods and show Milankovitch climatic cycles around 19, 23, and 41 kyr [28–30].

In **Figure 1**, we can see the periodograms for these time series – individual and combined (OR/AND) periodograms. **Figure 2** shows the amplitude and phase analysis for the $\sim 41kyr$ Milankovitch period found, for each analyzed series.

5. Conclusions

With the recent advances in experimental sciences, there is an increased need to analyze and statistically evaluate the information contained in time series. In some areas, as paleoclimatic studies, the study of the information on this kind of data constitutes the main body of physical evidence to understand and solve fundamental today's problems – as the origin and development of the *climatic change* observed in recent years.

Fourier methods still constitute an updated tool since they are simple and offer easy-to-understand results. However, in several of these applications, as in strati-graphic data, these time series come with severe sampling anomalies. These anomalies ultimately prevent the use of standard Fourier techniques, as the FFT algorithm and periodograms. The Lomb-Scargle periodogram has been very useful to attack this sort of problem. However, its use, as seen in the literature, lacks a proper analysis of the uncertainties associated and, worse, is unfit to be applied in the most poorly sampled time series.

The spectral analysis of irregularly sampled times series represents a problem of studying a physical system from incomplete information. As we know, from geophysical inverse theory, this kind of problem cannot be solved without some input of *a priori* information – explicitly or not.

The Popper-Bayes approach for physical inference proposes to solve inference problems by a combination of information: theoretical usually expressed through the functional form of statistical distributions and by combining independent experimental data. The LST periodogram is a development of the Lomb-Scargle periodogram that brings these inference principles to the periodogram analysis of irregularly sampled time series. The main idea of the LST periodogram is to smooth the periodogram of a dynamical variable through the stacking of spectral information from multiple irregularly sampled times series.

The periodogram of an irregularly sampled times series cannot, by any means, become a set of independent ordinates for being submitted to a proper statistical test. With the LST periodogram, we propose to change the use of the periodogram: from an auxiliary tool to statistical decision theory (define a periodicity) to a dimension reduction problem – from a broad set of possible frequencies to very narrow set of periodogram local maxima (peaks).

Acknowledgements

Note

The Lomb-Scargle-Tarantola (LST) periodogram was very much inspired by the seminal work of Professor Albert Tarantola (1949–2009) on Physical Inference and Geophysical Inverse Problems.

Author details

George Caminha-Maciel
Computational Geophysics Laboratory (COMPUTAGEO), Geology Department, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brazil

*Address all correspondence to: caminha.maciel@ufsc.br

References

[1] Péron G, Fleming CH, de Paula RC, Calabrese J. Uncovering periodic patterns of space use in animal tracking data with periodograms, including a new algorithm for the Lomb-Scargle periodogram and improved randomization tests. Movement Ecology. 2016;**4**:19

[2] Baldysz Z, Nykiel G, Araszkiewicz A, Figurski M, Szafranek K. Comparison of GPS tropospheric delays derived from two consecutive EPN reprocessing campaigns from the point of view of climate monitoring. Atmospheric Measurement Techniques. 2016;**9**: 4861-4877

[3] Berger WH. On the Milankovitch sensitivity of the quaternary deep-sea record. Climate of the Past. 2013;**9**: 2003-2011

[4] Bowdalo DR, Evans MJ, Sofen ED. Spectral analysis of atmospheric composition: Application to surface ozone model–measurement comparisons. Atmospheric Chemistry and Physics. 2016;**16**:8295-8308

[5] Dawidowicz K, Krzan G. Analysis of PCC model dependent periodic signals in GLONASS position time series using Lomb-Scargle periodogram. Acta Geodynamics et Geomaterialia. 2016;**13**(3):299-314

[6] Nielsen T et al. Are there multiple scaling regimes in Holocene temperature records? Earth System Dynamics. 2016;**7**:419-439

[7] Hinnov LA. Cyclostratigraphy and its revolutionizing applications in the earth and planetary sciences. GSA Bulletin. 2013;**125**:1703-1734

[8] Deeming TJ. Fourier analysis with unequally spaced data. Astrophysics and Space Science. 1975;**36**:137-158

[9] Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophysics and Space Science. 1976; **39**:447-462

[10] Scargle JD. Studies in astronomical time series analysis, II, statistical aspects of spectral analysis of unevenly spaced data. The Astrophysical Journal. 1982; **263**:835-853

[11] Vio R, Andreani P, Biggs A. Unevenly-sampled signals: A general formalism for the Lomb-Scargle periodogram. Astronomy and Astrophysics. 2010;**519**:A85

[12] Vio R, Diaz-Trigo M, Andreani P. Irregular time series in astronomy and the use of the Lomb-Scargle periodogram. Astronomy and Computing. 2013;**1**:5-16

[13] Hernandez G. Time series, periodograms, and significance. Journal of Geophysical Research. 1999;**104**(10): 368

[14] Stoica P, Li J, He H. Spectral analysis of non-uniformly sampled data: A new approach versus the periodogram. IEEE Transactions on Signal Processing. 2009;**57**(3):843-858

[15] Vander Plas JT. Understanding the Lomb-Scargle periodogram. The Astrophysical Journal Supplement Series. 2018;**236**:16

[16] Caminha-Maciel G, Ernesto M. Characteristic wavelengths in VGP trajectories from magnetostratigraphic data of the early cretaceous Serra Geral lava piles, southern Brazil. In: Jovane L, Herrero-Bervera E, Hinnov L, Housen BA, editors. Magnetic Methods and the Timing of Geological Processes. London: The Geological Society of London. Special Publications; 2013. p. 373

[17] Caminha-Maciel G, Ernesto M. LSTperiod software: Spectral analysis of multiple irregularly sampled time series. Annals of Geophysics. 2019;**62**:5: DM566. DOI: 10.4401/ag-7923

[18] Tarantola A. Popper, Bayes and the inverse problem. Nature. 2006;**2**: 492-494

[19] Tarantola A, Mosegaard K. Athematical basis for physical inference. Cornell University Library. arXiv:math-ph/0009029v1; 2000

[20] Tarantola A, Valette B. Inverse problems = quest for information. Journal of Geophysics. 1982;**50**:159-170

[21] Mortier A, Faria JP, Correia CM, Santerne A, Santos NC. BGLS: A Bayesian formalism for the generalized Lomb-Scargle periodogram. Astronomy and Astrophysics. 2015;**573**:A101

[22] Mortier A, Cameron AC. Stacked Bayesian general Lomb-Scargle periodogram: Identifying stellar activity signals. Astronomy and Astrophysics. 2017;**601**:A110

[23] Munteanu C, Negrea C, Echim M, Mursula K. Effect of data gaps: Comparison of different spectral analysis methods. Annales de Geophysique. 2016;**34**:437-449

[24] Pardo-Igúzquiza E, Rodríguez-Tovar FJ. Implemented Lomb-Scargle periodogram: A valuable tool for improving cyclostratigraphic research on unevenly sampled deep-sea stratigraphic sequences. Geo-Marine Letters. 2011;**31**:537-545

[25] Pardo-Igúzquiza E, Rodríguez-Tovar FJ. Spectral and cross-spectral analysis of uneven time series with the smoothed Lomb-Scargle periodogram and Monte Carlo evaluation of statistical significance. Computers and Geosciences. 2012;**49**:207-216

[26] Zechmeister M, Kürster M. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. Astronomy & Astrophysics. arXiv: 0901.2573v1 [astro-ph.IM]; 2009

[27] Townsend RHD. Fast calculation of the Lomb-Scargle periodogram using graphic processing units. The Astrophysical Journal Supplement Series. 2010;**191**:247-253

[28] Lisiecki LE, Raymo ME. A Pliocene-Pleistocene stack of 57 globally distributed benthic delta-18-O records. Paleoceanography. 2005;**20**:PA1003

[29] Jalón-Rojas I, Schmidt S, Sottolichio A. Evaluation of spectral methods for high-frequency multi annual time series in coastal transitional waters: Advantages of combined analyses. Limnology and Oceanography: Methods. 2016;**14**:381-396

[30] Lisiecki LE, Raymo ME. Plio–Pleistocene climate evolution: Trends and transitions in glacial cycle dynamics. Quaternary Science Reviews. 2007;**26**:56-69