ANNALS OF GEOPHYSICS, 62, 5, DM566, 2019; doi: 10.4401/ag-7923

"*LSTPERIOD* SOFTWARE: SPECTRAL ANALYSIS OF MULTIPLE IRREGULARLY SAMPLED TIME SERIES"

George Caminha-Maciel[*,1,2] and Marcia Ernesto[3]

[(1)] University of Hawai'i at Manoa, SOEST – Hawaii Institute of Geophysics and Planetology (HIGP), Petrofabrics and Paleomagnetics Laboratory, Honolulu, Hawaii, USA (current address).
[(2)] Universidade Federal de Santa Catarina, Departamento de Geociências, COMPUTAGEO – Laboratório de Geofísica Computacional, Campus Universitário – Trindade, Florianópolis, SC, Brasil.
[(3)] Universidade de São Paulo, Instituto de Astronomia Geofísica e Ciências Atmosféricas, Departamento de Geofísica, Cidade Universitária, São Paulo, Brasil.

Article history
Received September 11, 2018; accepted April 2, 2019.
Subject classification:
Paleoceanography and paleoclimatology, Paleoclimate; Inverse methods; Statistical analysis.

ABSTRACT

Irregularly sampled time series are common in several different areas, such as astronomy, meteorology, biology, oceanography, cyclostratigraphy, and others. The periodogram is a primary tool to extract meaningful information from irregularly spaced and noisy time series. It is an element of decision theory, meaning the periodogram usually transforms the data, and its ordinates are subsequently submitted to a statistical test compared to a population originating from a known stochastic model (white Gaussian noise). If some ordinate f_0 (usually a local maximum, a peak) fails in this test, we declare that it is a 'periodicity' at a frequency f_0. Besides its full usage, this method until now suffer from numerous theoretical difficulties in adapting to real case situations and shows lack of usefulness for very poorly sampled and high noise cases. All of it implies low usefulness for applying in most sedimentary sequences at our disposal nowadays. The *LSTperiod* is an application, written in Matlab, conceived to perform spectral analysis of multiple irregularly sampled time series. It combines information from Lomb-Scargle periodogram estimates over different time series sampling the same phenomenon, enabling the recovering of signals from very poorly sampled and noisy time series. The software comprises a set of four Graphical User Interfaces (GUIs) that allow the user to:

1) Have broad choices of the frequency-domain range and density for spectral estimation;
2) Select possible spectral features (i.e., pick "T") for testing as a model $[A*\sin(\frac{2\pi}{T}\,t-\theta)]$ through the visualization of several goodness-of-fit statistics;
3) Visualize the fitting residuals in the time domain, for each time series, for the chosen sinusoidal model.

These tools help the user to identify and analyze any suspected feature in the estimated spectra through its related linear system responses. All estimated parameter can be saved on worksheets and the visualizations in several different figure formats. We illustrate the use of the software with a set of Ocean Drilling Program (ODP) data series that show long-period Milankovitch-related spectral features and demonstrate its performance using synthetic time series.

1. INTRODUCTION

Large amounts of data in the form of irregularly sampled time series have emerged from several different areas, such as astronomy, meteorology, biology, oceanography and cyclostratigraphy [Baldysz et al., 2016; Bowdalo et al., 2016; Dawidowicz and Krzan, 2016; Jalón-Rojas et al., 2016; Péron et al., 2016;

Mortier and Cameron, 2017]. In all these fields, the time series are very often the product of irregular sampling. For example, in astronomy, the time series are regularly sampled but usually present inevitable gaps due to interruptions on observations, due to daylight time or bad weather conditions. Additionally, in cyclostratigraphy, it is not possible to control the sampling times at all, leaving us with a set of averaged data points defined by the deposition rate. These issues pose a significant challenge to spectral analysis since it prevents the building of an equivalent set of independent points in the frequency domain similar to that obtained from applying a Discrete Fourier Transform (DFT) to a regularly spaced time series. Therefore, over the last years, many efforts have been devoted to understanding the advantages, as well as limitations, of the methods dedicated to extracting meaningful spectral information from irregularly sampled time series, and vast literature on the subject has been produced [e.g., Schwarzenberg-Czerny, 1989; Babu & Stoica, 2010; Baluev, 2013; Vio et al., 2013; Munteanu et al., 2016; Jalón-Rojas et al., 2016; VanderPlas, 2018].

Deep-sea sedimentary columns have revealed valuable records of climatic fluctuations, and they are currently abundant due to global ocean drilling projects such as the DSDP, ODP, and IODP. A great deal of literature has been devoted to Quaternary cyclostratigraphy [see Hinnov, 2013 for a review], and its recording of the climate response to the external forcing (i.e., the Milankovitch theory of orbital forcing) has become a leading issue [e.g., Berger, 2013]. Global analyses and comparisons of sedimentary systems at different latitudes and in different environments aim to evaluate the sensitivities of sedimentary records (climatic proxies) to orbital forcing over time [e.g., Lisiecki and Raymo, 2007]; however, as stressed by Berger [2013], there is not a simple link between orbital forcing and climate response. Also, climatic components, as expressed in sedimentary records, are not usually stationary, showing regional modulated expressiveness, and correlation with the sampling process itself (sedimentary deposition rate).

There is currently a multitude of different methods available to analyze time series – both in time and frequency domains. Among them, Fourier methods accomplish a decomposition of the time series in an orthogonal basis (very frequently sines and cosines). Especially the so-called Discrete Fourier Transform (DFT) - could be highlighted mainly for being numerically fast (through Fast Fourier Transform - FFT algorithm), physically insightful (simple physical interpretation for its decomposition terms), and have straightforward extensions to non-equally sampled data (as in the periodogram, classical or the least squares).

Considering the inevitably irregularly sampling of deep-sea records (as for records from many other sources), the usual methods of spectral decomposition based on the DFT, as the Fast Fourier Transform (FFT) algorithm, are not directly applicable. Therefore, the least squares periodogram, also called the Lomb-Scargle (LS) periodogram [Lomb, 1976; Scargle, 1982; VanderPlas, 2018] appears as a natural extension of FFT methods. The LS periodogram and its variants, which add statistical significance tests for spectral peaks [Stoica et al., 2009; Vio et al., 2010; Baluev et al., 2013; Vio et al., 2013; Mortier et al., 2015], have been the tool of choice for these cases. The lack of orthogonality requires several strong assumptions for the time series (i.e., regarding the high total length compared to signal period, high average sampling rate, and noise frequency domain characterization) to conduct these peak significance tests. Also, the recovery of spectral information from time series data also introduces artifacts such as aliasing (power leakage to distant frequencies), spectral leakage (power leakage to nearby frequencies), mixing with red noise, mixing with sampling noise, and other distortions of the original spectral content. Additional issues such as the lack of a sufficient frequency resolution and the presence of statistical biases remain unsolved [e.g., Zechmeister and Kürster, 2009; Mortier et al., 2015]. Moreover, as stressed by Hernandez [1999], some procedures, including de-trending, filtering or pre-whitening, routinely applied to data subject to spectral analysis may modify the signal power and displace the original spectral peaks.

Besides that, real-world dynamical systems do not always exhibit sharp lines since they generally evolve. Therefore, stationary frequency components for all times are not very common on well-sampled natural time series, especially in cyclostratigraphy. Therefore, new methods of analysis are required; particularly those devoted to properly analyzing and displaying these signals, as well as to access uncertainties on the estimated spectral content from irregularly sampled time series.

Caminha-Maciel and Ernesto [2013] presented a method to address these issues through a Bayesian combination of independent experimental information derived from multiple time series as a stacking procedure applied within the frequency domain aimed to smooth the periodogram and to enhance the signal. We developed this method to study weak signals in

short and arbitrarily sampled time series with spectral distortions caused by noise and sampling deficiencies. We tested the method numerically for synthetic Gaussian noisy and poorly sampled time series.

This estimation process is based on the LS periodogram, but instead of decision theory method (statistical significance test), it incorporates the inverse problem approach envisioned by Tarantola and Valette [1982], Tarantola and Mosegaard [2000] and Tarantola [2006] for a combination of experimental and theoretical information. For this reason, we will designate it as the Lomb-Scargle-Tarantola (LST) periodogram - a process not based on optimized quantities such as averages or standard deviations (i.e., confidence limits) but in smoothing the periodogram estimates through a combination of information.

Indeed, in this approach, we assume that we can express the information for the estimated parameter using freely normalizable probability functions called "state of information" functions, defined over a finite set of domain points. Then, the solution of the inverse problem can be stated as the result of the logical operations OR and AND applied among these independent state of information functions. The results of the OR and AND operators can be shortly explained in the following way: the OR operator represents the generalized creation of histograms and is related to the arithmetic average of the curves; while the AND operator represents the generalized concept of conditional probabilities and suggests an initial null probability distribution conditioned to several independent experimental probability distributions. The AND curve enhances the common features between each of the individual spectra primarily and represents the set of 'surviving models', and it is assumed to be the main result of the inversion process.

Here we summarise the main steps of the algorithm:

1) For each time series, define the minimal time interval between two consecutive points as $t^{(i)}{}_{min}$, and define $t^{(i)}{}_{max}$ as the length of each time series.
2) The Nyquist frequency is set as $f_{Nyquist} = \frac{1}{2} * \max_i\{t^{(i)}{}_{min}\}$, which defines the minimum periodicity $T_{min} = 1/f_{Nyquist}$ (or the highest frequency) tested in the analysis. The maximum permitted periodicity (or lowest frequency) is set as $T_{max} = 1.5 * \min\{t^{(i)}{}_{max}\}$.
3) The program calculates the Lomb-Scargle periodogram for each series between T_{min} and T_{max} using a pre-selected first grid density N_0 over the interval. We set the initial N_0 as 100, but the user can later change this number to an adequate number.
4) Each periodogram is normalized by its total power. The resulting state of information functions have total areas equal to one and retain the estimated signal/noise ratio (S/N) for each time series, i.e., it maintain $P(f_1)/P(f_2)$ for any grid points f_1 and f_2, where $P(f)$ stands for the power of f.
5) Combine these state of information functions with the OR and AND operators [see Tarantola and Mosegaard, 2000] and then conduct the normalization step again for a total area (i.e., power) equals one.

In this paper, we present the software *LSTperiod* for spectral analysis according to the Caminha-Maciel and Ernesto [2013] method. The software also provides complementary information to help during the interpretation process including a set of four Graphical User Interfaces (GUIs) allowing for broad frequency-domain visualization of periodogram estimates and selection of spectral features regardless of any interpretation of its dynamical origin (i.e., signal or noise). We also illustrate the use of the program with an application to a set of ODP data series and some synthetic data series.

The remainder of this paper is organized as follows: Section 2 contains the essential background to introduce the LST periodogram; Section 3 contains a description of the LSTperiod software with its main features e mode of operation; Section 4 shows the application to a real set of cyclostratigraphic series with very well known results, and its performance when applied to synthetic time series; and Section 5 presents the final considerations.

2. THEORETICAL BACKGROUND

This section contains a brief review of the main mathematical background needed to introduce the LSTperiod, briefly discuss the Lomb-Scargle periodogram and the Tarantola's combination of information approach for inverse problems. We also present the new idea of periodogram analysis: linear fitting on multiple time series of a known (pre-chosen) physical model followed by statistical diagnosis metrics (goodness-of-fit statistics).

The Fourier Transform (e.g., DFT) is a function that gives a sequence of $N_0/2$ (different) complex coefficients from a time series $X(t)$, with N_0 equally time spaced points. In this case, the modulus of those coefficients makes a function known as the "spectrum" (its estimator is the periodogram). This function gives information about the splitting of energy among frequencies (or

periods) and has central physical meaning in the study of dynamical systems. These frequencies are spaced by an increment of 1/T, where T stands for the total time covered by the series (this is called "resolution" of the spectrum). Moreover, another important parameter is the Nyquist frequency, which is the highest frequency about what there is, theoretically, any recoverable information on the time series. The Nyquist frequency is defined as 1/(2*Δt), where Δt stands for the sampling interval. One important result in this area, the so-called Shannon-Nyquist theorem states that if the time series is equally spaced, and it does not contain information above the Nyquist frequency, then all the information in the time series is in its DFT. When the series are unevenly-spaced, this theorem no longer applies.

2.1 PERIODOGRAM ANALYSIS

Since a very long time periodograms have been used to make estimates of times series spectral distribution and to search for hidden periodicities in experimental data. At first, a formulation similar to the squared absolute values of DFT was used, the so-called "classical periodogram". This periodogram shows very noisy results even for a slightly noisy time series. One of the reasons for this is the lack of orthogonality on Fourier basis for irregular spaced data.

After the Lomb and Scargle's works [Lomb, 1976; Scargle, 1982; Vio et al., 2013], a new form of periodogram takes place, the now-called "Lomb-Scargle periodogram". Scargle introduced a shift parameter τ, calculated for each frequency, that plays the role of minimizing the inter-cross terms in the Fourier basis, $\sum_j sin(\omega tj)cos(\omega tj)$, improving overall separation between the basis. This periodogram, which was proved equivalent to a least square fitting of a sinusoidal signal, shows much smoother results:

$$P(\omega)=\frac{1}{2}\left\{\frac{[\sum_j X_j cos\omega(t_j-\tau)]^2}{\sum_j cos^2\omega(t_j-\tau)}+\frac{[\sum_j X_j sin\omega(t_j-\tau)]^2}{\sum_j sin^2\omega(t_j-\tau)}\right\} \quad (1)$$

Where the parameter τ is given by:

$$\tau(\omega)=\frac{1}{2\omega}\, arc\, tan\left\{\sum_j sin2\omega t_j \Big/ \sum_j cos2\omega t_j\right\} \quad (2)$$

Scargle also introduced a statistical test for a periodicity on the ordinates of a periodogram (a decision theory). This theory suggests testing the maximum ordinate (also called Fisher criteria) of the periodogram against the hypothesis of an entire stochastic generated population of ordinates. What means to test if a white Gaussian noise process, which implies an exponential distribution of the ordinates, could generate this maximum ordinate. Hence the probability α that at least one of P(ω) would exceed a certain level is given by:

$$\alpha = 1-\left[1-e^{-Pk/\sigma_n^2}\right]^{N^*} \quad (3)$$

Throughout this equation, it is possible to determine a "false alarm level" (or detection threshold) given by:

$$L_{th}=-\sigma_n^2\, ln\left[1-(1-\alpha)^{1/N^*}\right] \quad (4)$$

A critical aspect of this statistical test is that it can be applied only over independent ordinates, what bring limitations to poorly sampled time series (with a small number of points):

a) The maximum number of independent points ideally allowable over the frequency interval would be N/2 – what may not coincide with the interest signals frequencies;
b) Since there is no orthogonality on the basis, there is no way to re-establish independence between periodogram ordinates.

A large part of the subsequent literature on the subject is dedicated to finding conditions to establish independence between the periodogram ordinates. Most of this literature indicates conditions of the statistical regularity of the time series – a long time length of the series compared to periodicities of interest, uniformly random sampling, stationary periodic components and stochastic regularity of the noise (a closed model for it). However, rarely, in real data cases, we can verify these premises beforehand.

2.2 STATE OF INFORMATION FUNCTIONS

Inverse problems is a set of techniques developed in geophysics to deal with broad classes of statistical problems related to incomplete information – incomplete theoretical formulation or incomplete data (ill-posed problems). Albert Tarantola's theory of inverse problems also called "combination of information" has some advantages such as:

a) Simplicity - being physically intuitive,
b) It allows incorporating in a natural way multi-modal (or weakly defined) distributions for data as well as for the solution.

This technique has, however, the drawback of being computationally costly, although not so acute for low-dimensional problems. Also, this computational issue is continuously waived as modern computers improve.

Here we summarise this approach (Tarantola and Valette, 1982; Tarantola and Mosegaard, 2007; Tarantola, 2009):

a) Replace analytical theories with optimized quantities like as means and variances, for operations between probability distributions over the parameter space. These probability distributions could be freely normalizable and multi-modal, and should express our actual knowledge on the variable;
2) Operate these probability distributions (also called "state of information functions") with the two logical operands "AND" and "OR". The operator OR can be seen as a generalization of doing histograms and is defined as an arithmetic average of the individual distributions $\sum_i P_i(f)$. The operator AND represents the generalization of the idea of conditional probabilities and is defined as $\frac{\prod_i P_i(f)}{\mu}$, where μ is the null information function – depending on the geometry of the problem.

The null information function μ is usually assumed as constant over the whole interval. In some cases, for example in physical problems involving the variable *frequency,* it has been shown that it is better if it takes another functional form as $1/f$. In our case however, these frequencies (or periodicities) are only labels for a wide class of eigen-functions (sines and cosines) and we can use a constant function as null information distribution.

2.3 LST PERIODOGRAM

Due to Shannon's theorem, we already know that an irregularly sampled time series may not contain complete information on the frequency components of the original series. Any attempt to recover these components should apply some additional *a priori* information – explicitly or implicitly. Here we consider all the information that could potentially be resolved by each time series.

Then we choose searching bandwidths that start with a period which is one and a half times the length of the shortest time series, up to the shortest period (or maximum frequency, the Nyquist frequency) taken as twice the shortest distance between two consecutive points, considering all series.

The basic idea of LSTperiod is, for each frequency, to define for each time series a linear factor proportional to the estimated power in that frequency. To do this, we need to find invariant statistics that capture the S/N (signal to noise ratio) without being much sensitive to each series total variance – which depends on noise variance, length of the series, sampling pattern, among other aspects. The solution found was to normalize each periodogram by the total power (area under the periodogram), i.e., all periodograms exhibits total variance equals one. This procedure is equivalent to stretching the X variable in the time domain (i.e., X(t) versus t) and does not alter $P(f_i)/P(f_j)$, with $i \neq j$.

After this normalization, these distributions are operated with the OR and AND logical operators described above, and so normalized again. These two distributions represent the stacking of all experimental data periodograms and incorporate within all available information on S/N.

The density of calculation is initially set as 100 points on the interval, but it is allowable to change depending on the user's needs and computational hardware.

2.3.1 LINEAR SYSTEMS AND GOODNESS-OF-FIT STATISTICS

Any point in these curves (including the original periodograms) can be set as a possible candidate for a sinusoidal model fitting. Several visualizations are possible to the complete set of parameters originated from this fitting – powers, amplitudes and phases. Another window also displays some residuals visualizations and statistics (commonly used goodness-of-fit metrics).

In this way, the results (amplitudes and phases), as well as the quality of fitting (residuals metrics), can be compared among all-time series. This method allows us to study the coherent signals present in times series as well as the sampling process itself. Moreover, it could be more enlightening, from the physical standpoint, than merely to determine some confidence level for a periodicity, regardless of critical analysis over the time series and the sampling process underlined.

3. THE LSTPERIOD SOFTWARE

The LSTperiod software was developed within the MATLAB platform, and is freely available (http://www.iag.usp.br/paleo/sites/default/files/LSTperiod-files.zip) as a Matlab (R2012b or newer) m-code (with some graphical ".fig" files and ".dat" test files) to run within the Matlab environment, or as an executable ".exe" Windows 10 version (which does not requires the Matlab previously installed). The software comprises four GUIs: *LSTgui*, *guiFitFreq*, *guiResiduals,* and *LST_Table*. With the software, we also deliver a brief manual and a set of synthetic data to help users to

format their time series and have an initial experience.

To run *LSTperiod*, the program (executable file, or source code plus "fig" files) and all of the data files must be in the same folder. Data files should be low-level text files saved with ".dat" extension and should contain only two columns of numbers separated by space: time (t) and the measured variable X(t). For the files, we highly recommend adopting short names starting with a letter; otherwise, *LSTperiod* may display chunks in some visualization labels.

As the software starts, it will read all data files in the same folder and exhibit the periodograms. However, three blank panels will appear, because the frequency range must be selected. After that, the user can select the wanted files and run again. The user may run as many files as wanted; the only restriction is the processing time, which will increase for a large number of files running simultaneously. It is widely known that the periodogram (as well as the DFT) is a slow algorithm, and the processing time will quickly rise as the number of points used in the calculation of the spectra increases. For this reason, *LSTperiod* comes with two specific tools to manage processing time: an adjustable grid density slider and a small *wait-bar* window to display the evolution of any actual calculation going on (Figure 1).

At the first run, *LSTperiod* opens three panels – one (bottom of the window) containing all the spectra for all series, the OR spectrum (middle), and the AND spectrum (top) (Figure 1). However, at this point, the panels will be blank as an adjustment of the frequency range is necessary, and this is done by sliding the bars. The next step is to set the analysis bandwidth: the user can type a numerical value into the *'central BW'* (central bandwidth) editable box, which corresponds to the central value of the desired frequency range. The number of points (density) in the calculation can be set using the bottom (larger) slider. After the adjustments, the program must be rerun by pressing the *Run* button. This operation can be repeated as many times as necessary until the features of interest are all highlighted. If wanted, one or more data files may be removed from the batch of data files by pressing the file icon (*'open file'*) at the top left of the window. The software will list all data files, and the desired files must be simultaneously marked (i.e., selected using the mouse and pressing the *Ctrl* button on the keyboard) for a new run.

3.1 GUIFITFREQ: LINEAR SYSTEM – PARAMETER ANALYSIS

A thorough investigation of any peak found in the combined AND or OR periodograms or any of the individual periodogram may be performed by marking the peak with the *'datatip'* tool at the top of the window and then pressing the *Fit(T)* button. This action will

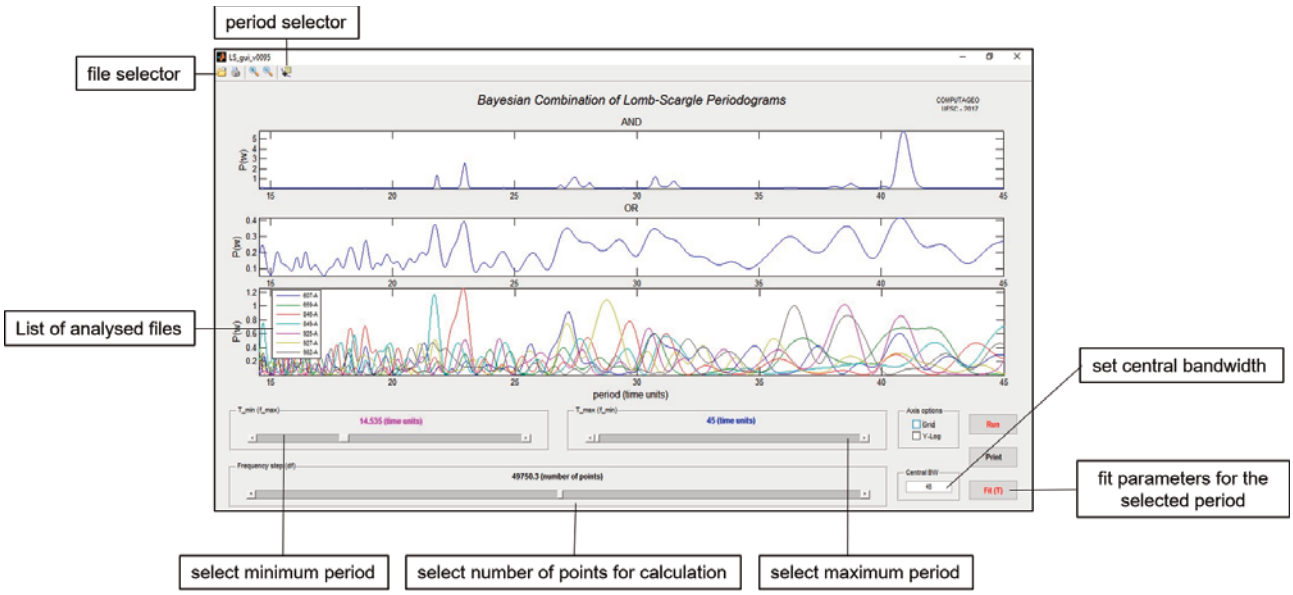

FIGURE 1. First LSTperiod window showing the periodogram calculated for each data file separately (bottom), and the combined results for the OR (middle) and AND (top) spectra. The sliding bars at the bottom of the window set the analyzed range of frequencies and the number of points in the calculation. The Fit(T) push button opens a second window once the pe–riod has been selected with the selection (period selector – datatip) tool.

FIGURE 2. Second LSTperiod window (guiFitFreq) for the selected parameter analysis, which is chosen using the error analysis menu box. The estimated parameters can be exported and saved to a ".xls" table.

open a new window showing a bar plot displaying the amplitudes of the selected period among each of the series. The pop-up menu in the bottom right (Figure 2) will display several commonly related parameters, including the phase, mean squared error, the coefficient of correlation, and coefficient of determination. All of the results for each file can be automatically sent to a table by pressing the *'Table'* button. The software will open a new GUI called *LST_table* where the numeric results can be seen, selected and copied (using the Ctrl+c and Ctrl+v keys), and they can be sent to a worksheet file (".xls" – what only works with Microsoft Excel installed). For each analyzed period the corresponding parameters will be saved separately in a different sheet of a spreadsheet so that the data will be available for further analysis according to the user. This spreadsheet will be saved with the temporary file name *'temp_LST_table.xls'*.

3.2 GUIRESIDUALS: LINEAR SYSTEM – RESIDUAL ANALYSIS

We can see the deviations (or errors) in the fit for a selected period for each data file just by pressing the *Residuals* button (Figure 2). A new window will open displaying the graphics for the selected data file and a goodness-of-fit parameter. In the *Residuals* pop-up menu, we can select several options from a list of possible visualizations, including the standard error, error against the model, error against the data, Gaussian fit for the errors, and quantile-quantile statistics. The definitions of these characteristic parameters are available in statistical books and texts [e.g., Walpole et al., 2007; Rodgers and Nicewander, 1988]. Each window can be printed and saved as a Matlab figure (".fig" files) or in any of several other formats, such as ".jpeg", ".tiff", ".eps" or ".pdf".

4. EXAMPLES

4.1 REAL DATA

To illustrate the performance of *LSTperiod*, we used a set of series showing the variation of benthic ^{18}O along sedimentary cores drilled by the Ocean Drilling Program (ODP) and reported by Lisiecki and Raymo [2005]. The time series were already submitted to spectral analyses [Lisiecki and Raymo, 2007] showing climatically modulated periods of approximately 19, 23 and 41 kyr, the last one being relevant only before 1.4 Myr. For this reason, we limited the maximum length of the time series to 1 Myr, although some of the time series are longer than 5 Myr. The data were organized merely into two-column (i.e., time - t and data values - $X(t)$) text files (".dat" extension) and were processed without any pre-treatment (e.g., filtering or de-trending). The selected bandwidth was 14.5 - 45 kyr, although longer periodicities could also be investigated. The results are displayed in Figure 1. The several individual spectra (bottom of the figure) are noisy, and the real periodicity may be challenging to identify. The AND spectrum (top of the figure) is far smoother than

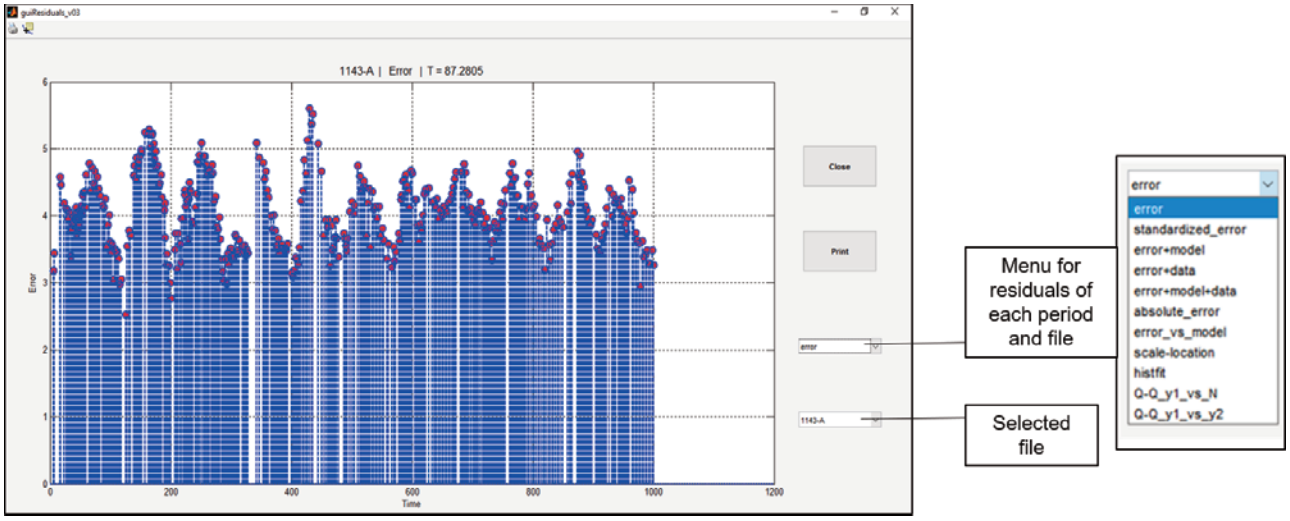

FIGURE 3. LSTperiod window for one of the options in the guiResiduals. The error analysis refers to the modeling of the selected file displayed in the bottom box. Each error type displayed in the menu has a graphical output for visual inspection.

the individual spectra, and it expresses features that are common to all series.

Mainly, we are interested in better evaluating the ~19, ~23 and ~41 kyr periods. The period of 41 kyr seems to be the most prominent feature in the spectrum as stressed by previous work [Lisiecki and Raymo, 2007]. The 23 kyr period is also detected, whereas the 19 kyr period, which is thought to be persistent throughout the last 5 Ma, does not appear clearly within our AND results. However, the 19 kyr period is present in the OR spectrum and is clearly defined in many of the individual spectra. It is worth mentioning that these data are derived from distant sites in the Atlantic and Pacific Oceans and that they might not show the same components of the forcing system.

Furthermore, Lisiecki and Raymo [2007] mentioned a low signal-to-noise ratio at this frequency. Other peaks at approximately 27 and 31 kyr are also present in the AND spectrum and seem unrelated to orbital forcing. However, it is beyond the scope of this paper to discuss these results further.

4.2 SYNTHETIC DATA

Here, we illustrate the performance of the method and software by analyzing a set of synthetic data series with known harmonic content (Figure 4). The synthetic time

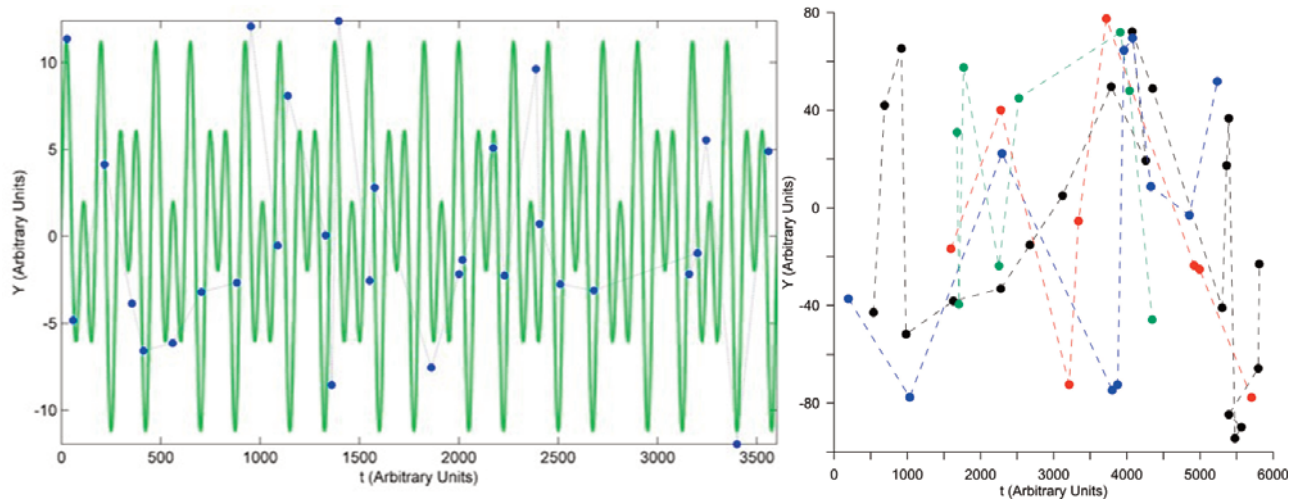

FIGURE 4. Synthetic series produced as described in the text. The green line (left) represents the original series corresponding to a superposition of sinusoidal components; blue dots and line (left) are one of the inhomogenous sampling series; the other four are plotted on the right diagram.

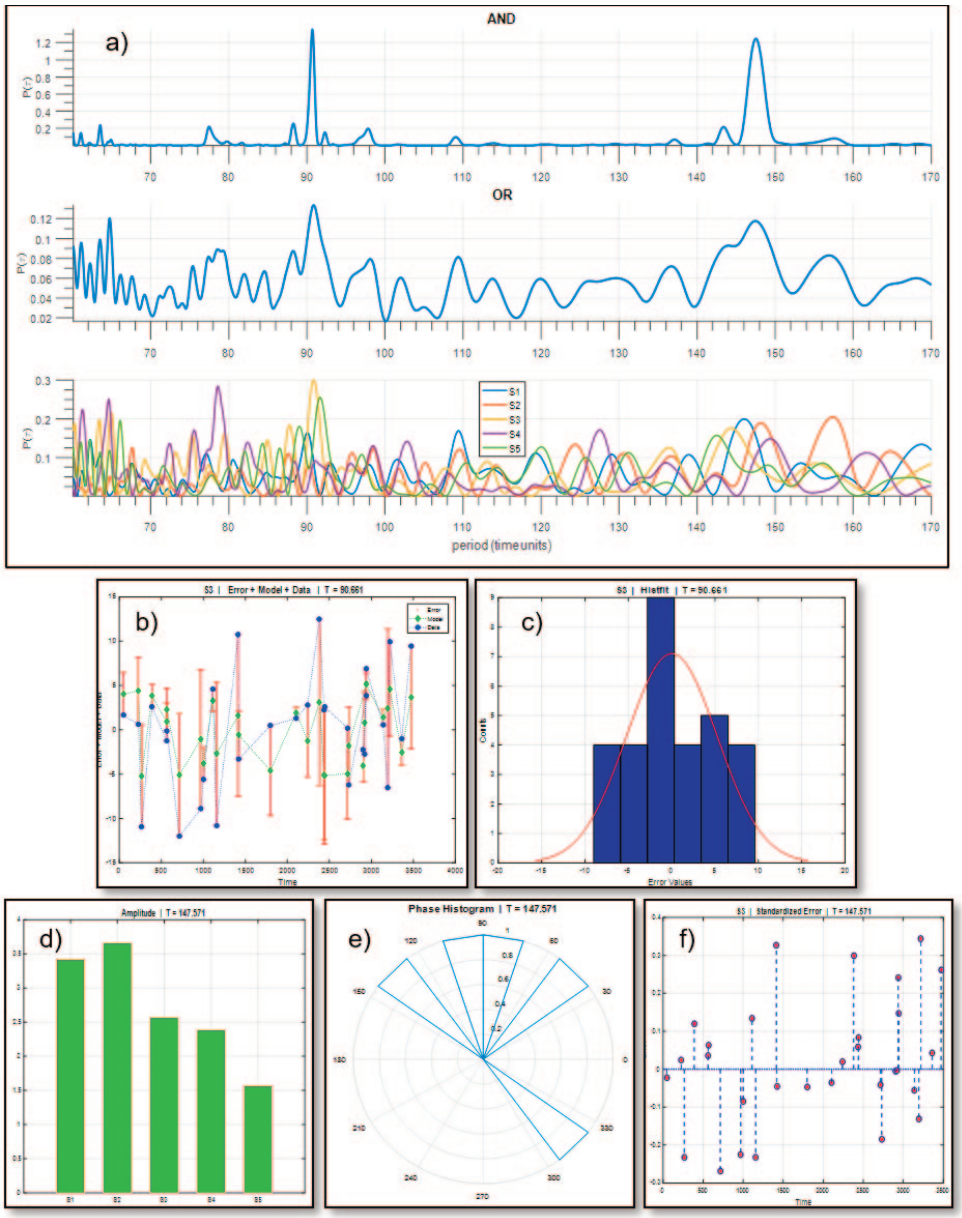

FIGURE 5. Spectral results for the synthetic time series with periods of 90 and 150 arbitrary time units. Small boxes show some examples of graphical outputs corresponding to analyses of the spectral peaks selected: a) original generated series (green line) and the sampling points (blue dots and line); b) calculated model plus error and data; c) error histogram; d) peak amplitude for each time series; e) rose diagram for the phase; and f) standardized errors for a particular time series.

series is the result of ten cycles of the superposition of two components with periods of 90 (with an amplitude of 6 and a phase of zero) and 150 (with an amplitude of 4 and a phase of $\pi/4$) time units. The resulting time series was then randomly sampled (30 to 50 points) over the entire length of the original series, thereby producing five different short time series named S1 through S5. To each point of the five series, we added a non-constant Gaussian noise level of 5% of its amplitude.

Considering the poor quality of the final series (Figure 4), the recovery of the periodicities was excellent. Figure 5a showed periods of ~90.78 and ~147.52; a very smooth AND spectrum was produced compared to the noisy individual spectra. Regarding amplitudes and phases (Figure 5d and 5e), it was possible to recover amplitudes of ~4 for the 147 period and phases of approximately $\pi/4$ for at least two series. However, it is not surprising that the amplitude and phase were not rigorously recovered considering the noisy and poorly sampled time series. The other diagrams in Figure 5 show the comparison of the original and modeled data along with the error bars (5b), the error histogram (5c), and the standardized errors for the series.

5. FINAL REMARKS

The LS periodogram is a well-known algorithm for detecting and characterizing signals in unevenly sampled data. However, the statistical significance of the detected periods may be difficult to ascertain, especially if the series has no characteristic sampling interval, and is deficient in accurately recording the periodic or quasi-periodic event. For this reason, many records of this type that carry some valuable information are often discarded as considered inappropriate for spectral analysis.

LSTperiod is a software specially designed to overcome problems in the calculation of the periodograms (e.g., poor data window) of short, noisy and irregularly sampled time series through the selective choice of the frequency range and grid density. The combination of various poor data sets that represent the sampling of the same phenomenon allows a reduction of the noise through smoothing and signal-noise gain, and consequently the extraction of relevant information with a low S/N ratio. *LSTperiod* runs on MatLab (R2012b or newer) environment and provides ample possibilities to investigate the reliabilities of the periodicities in time series data. The examples presented in section 4 show the power of *LSTperiod* in comparing variances of different records and statistically analyzing each suspected spectral feature. We demonstrated with synthetic examples that periodicities could be reliably identified in a set of poorly sampled time series. We illustrated the use of the *LSTperiod* method with cyclostratigraphic data because this is one prevalent issue in Geosciences, but the software can be applied to any data set that can benefits from the combination of information as proposed here. Time series with periodic gaps, as often found in astronomy, are also troublesome because they can generate unpredictably common patterns (like leakage and aliasing) within the spectral windows. In this case, the stacking procedure will enhance not only the *common* signals but also the *common* spectral anomalies.

Another significant advantage of *LSTperiod* is that it offers a statistical diagnosis for models derived from any suspected spectral feature, and thus, we can test as many frequency points as desired. On the other hand, as periodograms are highly time-consuming [Deeming, 1975; Hernandez, 1999; Townsend, 2010], *LSTperiod* may become slow for long time series (i.e., exceeding 10,000 readings), which may be a disadvantage.

This software does not intend to be a new tool for implementing old statistical methods. Indeed, there is much room for improvement in the user's interface and graphical capabilities (we count on the community's feedback to improve future versions).

Instead, it is a proof-of-concept tool, to demonstrate de very idea that spectral analysis of irregularly sampled natural time series is a process of extracting *incomplete* information. As so, there is always a need to input some *a priori* information. Nowadays, in the literature, this has been performed through the assumption of obscure statistical hypotheses (as white noise), and the so-called "time series carpentry" – meaning the cutting and manipulation of the time series to obtain desired results. As a consequence, most of the data obtained as time series (especially in cyclostratigraphy) is considered non-suitable to spectral analysis. Here we made a clear choice for the input of a priori information: starting for a broadest possible spectral bandwidth, we show graphics of invariant estimates and let to the user select and highlight features that seemed consistent with the previous physical knowledge of the system. Any suspected features can be tested not only relating to its amplitude or S/N (signal to noise ratio) but also related to its coherence properties – along with the time series and in comparison with other samples of the system's variability (other times series with common frequency components).

Acknowledgements. This work was funded by the FAPESP (grant 2004/05363-5) and the CNPq (grant 308475/2015) Brazilian funding agencies. The authors thank L.E. Lisiecki for kindly provided access to her DSDP and ODP databases. Two anonymous reviewers were very helpful in improving the manuscript. *LSTperiod* is a software registered at the INPI Brazilian innovation agency (number BR5120170008612).

REFERENCES

Babu, P., and Stoica, P., 2010. Spectral analysis of nonuniform sampled data – a review. Dig. Signal Process., 20: 359-378.

Baldysz, Z., Nykiel, G., Araszkiewicz, A., Figurski, M., and Szafranek, K., (2016). Comparison of GPS tropospheric delays derived from two consecutive EPN reprocessing campaigns from the point of view of climate monitoring. Atmos. Meas. Tech., 9, 4861–4877.

Baluev, R., (2013). Detecting multiple periodicities in observational data with the multi-frequency periodontal – I. Analytic assessment of the statistical significance. MNRAS 436, 807–818.

Berger, W.H., (2013). On the Milankovitch sensitivity of the Quaternary deep-sea record. Clim. Past., 9, 2003-2011.

Bowdalo, D.R., Evans, M.J., and Sofen, E.D., (2016). Spectral analysis of atmospheric composition: application to surface ozone model–measurement comparisons. Atmos. Chem. Phys., 16, 8295–8308.

Caminha-Maciel, G., Ernesto, M., (2013), Characteristic wavelengths in VGP trajectories from magnetostratigraphic data of the Early Cretaceous Serra Geral lava piles, southern Brazil. In "Magnetic methods and the Timing of Geological Processes" (L. Jovane, E. Herrero-Bervera, L. Hinnov, B.A. Housen, eds.), The Geological Society of London, Special Publications, 373.

Dawidowicz, K., and Krzan, G., (2016). Analysis of PCC model dependent periodic signals in GLONASS position time series using Lomb-Scargle periodogram. Acta Geodyn. Geomater., 13, 3 (183), 299–314.

Deeming, T.J., (1975). Fourier analysis with unequally spaced data. Astrophys. Space Sci., 36, 137-158.

Hernandez, G., (1999). Time series, periodograms, and significance. J. Geophys. Res. 104: 10,355-10,368.

Hinnov, L.A., (2013). Cyclostratigraphy and its revolutionizing applications in the Earth and planetary sciences. GSA Bulletin125: 1703-1734.

Jalón-Rojas, I., Schmidt, S., and Sottolichio, A., (2016). Evaluation of spectral methods for high-frequency multi-annual time series in coastal transitional waters: advantages of combined analyses. Limnol. Oceanogr.: Methods, 14, 381-396.

Lisiecki, L.E., and Raymo, M.E., (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic ^{18}O records. Paleocean., 20, PA1003.

Lisiecki, L.E., and Raymo, M.E., (2007). Plio–Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. Quaternary Sci. Rev., 26, 56–69.

Lomb, N.R., (1976), Least-squares frequency analysis of unequally spaced data. Astrophys. Space Sci., 39: 447-462.

Mortier, A., Faria, J.P., Correia, C.M., Santerne, A., and Santos, N.C., (2015). BGLS: A Bayesian formalism for the generalized Lomb-Scargle periodogram. Astron. Astrophys., 573, A101.

Mortier, A. and Cameron, A.C., (2017). Stacked Bayesian general Lomb-Scargle periodogram: Identifying stellar activity signals. Astron. and Astrophys., 601, A110.

Munteanu, C., Negrea, C., Echim, M., and Mursula, K., (2016). Effect of data gaps: comparison of different spectral analysis methods. Ann. Geophys., 34, 437–449.

Pardo-Igúzquiza, E., and Rodríguez-Tovar, F.J., (2011). Implemented Lomb-Scargle periodogram: a valuable tool for improving cyclostratigraphic research on unevenly sampled deep-sea stratigraphic sequences. Geo-Mar. Lett., 31, 537-545.

Pardo-Igúzquiza, E., and Rodríguez-Tovar, F.J., (2012). Spectral and cross-spectral analysis of uneven time series with the smoothed Lomb-Scargle periodogram and Monte Carlo evaluation of statistical significance. Comput. Geosci., 49, 207-216.

Péron, G., Fleming, C.H., de Paula, R.C., and Calabrese, J., (2016). 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, 4:19.

Rodgers, J., and Nicewander, W.A., (1988). Thirteen Ways to Look at the Correlation Coefficient. Am. Stat., 42: 59-66.

Scargle, J.D., (1982), Studies in astronomical time series analysis, II, Statistical aspects of spectral analysis of unevenly spaced data. Astrophys. J. 263: 835-853.

Schwarzenberg-Czerny, A., 1999. Optimum period search : quantitative analysis. Astrophys. J., 516: 315-323.

Stoica, P., Li, J., and He, H., (2009). Spectral analysis of non-uniformly sampled data: a new approach versus the periodogram. IEEE Trans. Signal Proc., 57, 3, 843-858.

Tarantola, A., (2006), Popper, Bayes and the inverse problem. Nature, 2: 492-494.

Tarantola, A. and Mosegaard, K., (2000), Mathematical basis for physical inference. Cornell University Library, arXiv:math-ph/0009029v1.

Tarantola, A. and Valette, B., (1982), Inverse problems = Quest for information. J. Geophysics, 50: 159-170.

Townsend, R.H.D., (2010). Fast calculation of the Lomb-Scargle periodogram using graphic processing units. Astrophys. J. Supplement Series, 191, 247–253.

VanderPlas, J.T., 2018. Understanding the Lomb-Scargle Periodogram. Astrophys. J. Sup. Ser., 236: 1-28.

Vio, R., Andreani, P., Biggs, A., (2010). Unevenly-

sampled signals: a general formalism for the Lomb-Scargle periodogram. Astron. Astrophys., 519, A85.

Vio, R., Diaz-Trigo, M., and Andreani, P., (2013). Irregular time series in astronomy and the use of the Lomb-Scargle periodogram. Astron. Computing, 1, 5-16.

Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K., (2007). Probability & Statistics for Engineers & Scientists. Pearson Prentice Hall, NJ, USA.

Zechmeister, M. and Kürster, M., (2009). The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. Astron. Astrophys., arXiv:0901.2573v1 [astro-ph.IM].

*CORRESPONDING AUTHOR: George CAMINHA-MACIEL,
University of Hawai'i at Manoa, SOEST,
Hawaii Institute of Geophysics and Planetology (HIGP),
Petrofabrics and Paleomagnetics Laboratory, Honolulu,
Hawaii 96822, USA;
email: caminha.maciel@ufsc.br