Least-squares wavelet analysis of unequally spaced and non-stationary time series and its applications

Ebrahim Ghaderpour* ·
Spiros D. Pagiatakis

Abstract The least-squares spectral analysis, an alternative to the classical Fourier transform, is a method of analyzing unequally spaced and non-stationary time series in their first and second statistical moments. However, when a time series has components with low or high amplitude and frequency variability over time, it is not appropriate to use either the least-squares spectral analysis or Fourier transform. On the other hand, the classical short-time Fourier transform and the continuous wavelet transform do not consider the covariance matrix associated with a time series nor do they consider trends or datum shifts. Moreover, they are not defined for unequally spaced time series. A new method of analyzing time series, namely, the least-squares wavelet analysis is introduced that is a natural extension of the least-squares spectral analysis. This method decomposes a time series to the time-frequency domain and obtains its spectrogram. In addition, the probability distribution function of the spectrogram is derived that identifies statistically significant peaks. The least-squares wavelet analysis can analyze any non-stationary and unequally spaced time series with components of low or high amplitude and frequency variability, including datum shifts, trends, and constituents of known forms, by taking into account the covariance matrix associated with the time series. The outstanding performance of the proposed method on synthetic time series and a very long baseline interferometry series is demonstrated, and the results are compared with the weighted wavelet Z-transform.

Ebrahim Ghaderpour*
Lassonde School of Engineering, York University, Toronto, ON, Canada M3J1P3.
Tel.: 647-879-4343
E-mail: ebig2@yorku.ca

Spiros D. Pagiatakis
Lassonde School of Engineering, York University, Toronto, ON, Canada M3J1P3.
Tel.: 416-736-5484
E-mail: spiros@yorku.ca

1 Introduction

A time series is a sequence of data points measured at discrete time intervals and may not be sampled at equally spaced intervals or it may contain data gaps, and so it is unequally spaced. Sometimes, and depending on the purpose of specific experiments, researchers may intentionally define a variable sampling rate to avoid aliasing (Shapiro and Silverman 1960; Pagiatakis et al. 2007). In certain experiments, measurements may also have uncertainties (variances) based on instrument calibration and thus may also be unequally weighted.

If all statistical properties of a time series (i.e., the mean, variance, all higher order statistical moments and auto-correlation function) do not change in time, then the time series is called stationary. A time series is called non-stationary if it violates at least one of the assumptions of stationarity (Brown and Hwang 2012). A non-stationary time series may contain systematic noise, such as trends (linear or exponential or others) and/or datum shifts (offsets) indicating that its mean value is not constant in time. The second moments (central and mixed moments) of the time series values form a symmetric matrix called the covariance matrix (Vanicek and Krakiwsky 1986). In certain fields, such as geodesy, geophysics, and astronomy, a time series usually has an associated covariance matrix which means

that the time series is non-stationary in its second sta-tistical moments. In geodynamic applications, seismic noise may contaminate the time series of interest or certain components of the time series may exhibit vari-able frequency, such as linear, quadratic, exponential or hyperbolic chirps. The reader is referred to Kay and Marple (1981), Mallat (1999) and Brown and Hwang (2012) for more details.

Time series analyses have almost exclusively been performed using the classical Fourier method and its modifications (Kay and Marple 1981; Stein and Shakarchi 2003), often incorrectly postulating that the time series under consideration is stationary, resulting in erratic results. It is not unusual to see that many re-searchers even attempt to modify the original raw mea-surements by interpolation, editing or offset removal, often using empirical and questionable approaches, sim-ply to satisfy the stringent requirements of the Fourier transform.

The least-squares spectral analysis (LSSA) was in-troduced by Vanicek (1969, 1971) for the purpose of analyzing non-stationary (in the first and second statis-tical moments) and unequally spaced time series. The LSSA has been described in detail in many publica-tions (Lomb 1976; Wells et al. 1986; Craymer 1998; Pa-giatakis 1999) and applied in many experiments (Scar-gle 1982; Hui and Pagiatakis 2004; Abd El-Gelil et al. 2008; Psimoulis et al. 2008; Wu et al. 2010). Concep-tually, this method estimates a spectrum based on the least-squares fit of sinusoids to the entire time series.

In the LSSA, trends, datum shifts and the covari-ance matrix associated with a time series are parame-terized and consequently estimated concurrently with the calculation of the spectrum (Wells et al. 1985). However, the LSSA cannot generally be used for time series whose components change amplitude (e.g., tran-sient signals) and frequency (e.g., doppler-shifted sig-nals) over time because it decomposes a time series to the frequency domain rather than to the time-frequency domain (similar to Fourier analysis). Craymer (1998) compared the LSSA with Fourier analysis (discrete) and need not be repeated here. Suffices to say that the Fourier analysis is a special case of the LSSA when the time series is strictly stationary, equally spaced, equally weighted, and with no gaps.

In order to measure the frequency variation of sounds, Gabor (1946) introduced the short-time Fourier transform (STFT) to analyze piece-wise stationary and equally spaced time series. He basically endeavored to obtain the spectrum of a time series, in short-time inter-vals (or within a window which translates through the entire time series) rather than considering the entire time series at once (Mallat 1999). Since in the STFT

the sinusoidal functions that do not complete an inte-ger number of cycles within a window are not orthogo-nal, the segment of the time series within such window cannot be represented as linear combination of these sinusoids properly. This increases the bandwidth of the spectrum (poor frequency resolution) corresponding to these sinusoidal functions in the frequency domain (Mallat 1999). Note that the STFT does not consider the correlation between the sinusoidal functions when they are not orthogonal within a window. In the STFT, the window length is fixed during translation in time, which is not appropriate for analyzing time series with components of high frequency variability over time.

The continuous wavelet transform (CWT) is a well-established method that computes a scalogram based on the dilations and translations of a (mother) wavelet function (Mallat 1999). One may also convert the scalogram to a spectrogram using different approaches that each has its own advantages and disadvantages (Hlawatsch and Boudreaux-Bartels 1992; Sinha et al. 2005). The CWT spectrogram has high-frequency res-olution (better frequency localization) at low frequen-cies and high-time resolution (better time localization) at high frequencies and vice versa (Mallat 1999; Sinha et al. 2005). In the CWT, if one uses an inappropriate wavelet, then the CWT spectrogram may give mislead-ing and physically meaningless results (Qian 2002).

Since unequally spaced wavelets are not defined, the CWT and discrete wavelet transform are not appro-priate for analyzing an unequally spaced time series unless one uses an interpolation method to fill in the missing data or shrinks the wavelet (Hall and Turlach 1997; Sardy et al. 1999). However, the interpolation is not an acceptable approach as it damages the time se-ries and results in erratic spectrograms in the time-frequency domain especially for non-stationary time se-ries (Hayashi and Yoshida 2005; Rehfeld et al. 2011). Moreover, it is not appropriate to apply the STFT or CWT (based on their definitions) to unequally spaced and/or unequally weighted time series that may also have trends (Foster 1996; Mallat 1999).

There are a number of methods developed and ap-plied to non-stationary time series, such as Wigner-Ville distribution (Classen and Mecklenbrauker 1980; Bay-dar and Ball 2001), Hilbert-Huang transform (Huang et al. 1998; Barnhart 2011), empirical wavelet trans-form (Gilles 2013), tunable-Q wavelet transform (Se-lesnick 2011), and ensemble empirical mode decompo-sition (Huang and Wu 2009; Ren et al. 2014). Each of them has its own advantages and weaknesses (Chen and Feng 2003; Huang and Wu 2008; Guang et al. 2014), and they are not generally appropriate for analyzing

non-stationary and non-ergodic time series that are unequally spaced and unequally weighted.

There are several other methods proposed by researchers for analyzing unequally spaced and non-stationary time series. Foster (1996) proposed a new method of time-frequency analysis, namely, the weighted wavelet Z-transform (WWZ), that performs very well in detecting the periodicities of the signals in both equally and unequally spaced time series. Amato et al. (2006) use a wavelet-based reproducing kernel Hilbert space technique and show that their technique is superior in terms of the mean squared error. Mathias et al. (2004) propose a method that is based on the least-squares method to deal with the undesirable side effects of nonuniform sampling in the presence of constant offsets.

In this contribution, a new method is developed that can analyze non-stationary time series in the first and second statistical moments that are also unequally spaced and comprise constituents of low/high amplitude and frequency variability over time. This method, namely, the least-squares wavelet analysis (LSWA), is a natural extension of the LSSA and can analyze rigorously any type of time series without any modifications or editing (Sect. 2.1). The practical implementation steps of the LSWA for equally and unequally spaced time series are demonstrated, and some of the advantages of the LSWA over the CWT are shown on two synthetic equally spaced time series (Sect. 2.2). In Sect. 2.3, the WWZ is described in detail and compared to the LSWA. In Sect. 2.4, stochastic surfaces (confidence level surfaces) are obtained above which the peaks in the least-squares wavelet spectrograms are statistically significant at certain confidence level (usually 95% or 99%). As an application of the LSWA, a very long baseline interferometry length time series is analyzed in Sect. 3.

2 Methods

2.1 Least-squares Wavelet Analysis

Conceptually, the approach is to decompose a time series to time-frequency domain by an appropriate segmentation of the series and calculate spectral peaks based on the least-squares fit of sinusoids (other functions may be chosen as well) to each segment. The segmentation is similar to the one in the CWT, and computing the spectral peaks is similar to the LSSA, so the approach is a combination of the CWT and the LSSA.

Suppose that $\mathbf{f} = \left[f(t_j) \right]$, $1 \leq j \leq n$, is a discrete time series of n data points; here the t_j's are not necessarily equally spaced. Let $\mathbf{\Omega} = \{\omega_k;\ k = 1, \ldots, \kappa\}$ be

a set of spectral frequencies, and let $\mathbf{y}$ be an arbitrary segment of the time series that contains $L(\omega_k)$ samples such that ω_k can be resolved within it

$$\mathbf{\Phi} = \tag{1}$$
$$\left[\cos\left(2\pi\omega_k t_{i+j-\frac{1}{2}(L(\omega_k)+1)} \right), \sin\left(2\pi\omega_k t_{i+j-\frac{1}{2}(L(\omega_k)+1)} \right) \right],$$

$$\mathbf{y} = \left[f\left(t_{i+j-\frac{1}{2}(L(\omega_k)+1)} \right) \right], \tag{2}$$

where $1 \leq i \leq L(\omega_k)$ and

$$L(\omega_k) := \begin{cases} \left\lfloor \dfrac{L_1 M}{\omega_k} \right\rfloor + L_0 & \text{if } \left\lfloor \dfrac{L_1 M}{\omega_k} \right\rfloor + L_0 \text{ is odd,} \\[2ex] \left\lfloor \dfrac{L_1 M}{\omega_k} \right\rfloor + L_0 + 1 & \text{otherwise,} \end{cases} \tag{3}$$

where symbol $\lfloor * \rfloor$ indicates the largest integer no greater than $*$, L_0 is a fixed number of data points, L_1 is a selected number of cycles of the sinusoidal base functions, M is the average number of data points per unit time, and ω_k is the number of cycles per unit time ($\omega_k \in \mathbf{\Omega}$). For instance, for an equally spaced time series recorded with a rate of one sample per millisecond, if frequency is in Hertz, then $M = 1000$ (data points per second), and if $L_1 = 2$, then two cycles of sinusoidal base functions of frequency ω_k will be fitted to a segment of the time series with $L(\omega_k)$ data points. Also, L_0 is the selected number of additional samples considered in the least-squares fit to achieve the desired time and frequency resolution in the spectrogram (see Sect. 2.2 for more details).

For simplicity of notations, L is used instead of $L(\omega_k)$ from now on. For each k, $1 \leq k \leq \kappa$, and each j, $1 \leq j \leq n$, if $1 \leq i + j - (L+1)/2 \leq n$, then the size of $\mathbf{y}$ in Eq. (2) which is the same as the number of rows in $\mathbf{\Phi}$ given by Eq. (1) is $R(j, k) :=$

$$\begin{cases} (L-1)/2 + j & \text{if } 1 \leq j < (L+1)/2, \\ L & \text{if } (L+1)/2 \leq j \leq n - (L-1)/2, \\ (L+1)/2 + n - j & \text{if } n - (L-1)/2 < j \leq n. \end{cases} \tag{4}$$

Note that when $L_1 = 0$, $R(j, k)$ is independent from the choice of ω_k. Assume that $\mathbf{P} = \mathbf{C_f}^{-1}$ is the associated weight matrix of $\mathbf{f}$, and $\mathbf{P_y}$ is the principal submatrix of $\mathbf{P}$ of dimension $R(j, k)$ (see below for more details).

If $\mathbf{f}$ contains constituents of known forms $(\phi_1, \ldots, \phi_q)$ but of unknown amplitudes (e.g., trends or sinusoids of constant frequencies), then for each pair (t_j, ω_k) let

$$\underline{\mathbf{\Phi}} = \left[\phi_1\left(t_{i+j-\frac{1}{2}(L+1)} \right), \ldots, \phi_q\left(t_{i+j-\frac{1}{2}(L+1)} \right) \right], \tag{5}$$

$$\overline{\mathbf{\Phi}} = \left[\underline{\mathbf{\Phi}}, \mathbf{\Phi} \right], \tag{6}$$

where $\mathbf{\Phi}$ is given by Eq. (1) and use the model $\mathbf{y} = \overline{\overline{\mathbf{\Phi}}}\,\overline{\mathbf{c}}$ to determine $\hat{\overline{\mathbf{c}}}$ using the least-squares method. Therefore,

$$\hat{\overline{\mathbf{c}}} = \overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}, \qquad (7)$$

where T signifies transpose and $\overline{\mathbf{N}} = \overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\overline{\mathbf{\Phi}}$ (matrix of normal equations). Let $\overline{\mathbf{J}} = \overline{\mathbf{\Phi}}\,\overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}$. Define the weighted least-squares wavelet spectrogram (LSWS) as

$$\overline{s}(t_j, \omega_k) = \frac{\mathbf{y}^{\mathrm{T}}\mathbf{P_y}\overline{\mathbf{J}}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}} \in (0, 1), \qquad (8)$$

where $1 \leq j \leq n$ and $1 \leq k \leq \kappa$. The justification of the name comes from the fact that when $L_1 \neq 0$, the size of $\mathbf{y}$ decreases as frequency increases in the calculation of the spectrogram using the least-squares method similar to the CWT. Notice that if $\mathbf{f}$ is equally spaced, equally weighted and $\overline{\overline{\mathbf{\Phi}}} = \mathbf{\Phi}$, then $\hat{\overline{\mathbf{c}}}^{\mathrm{T}}\hat{\overline{\mathbf{c}}}$ has similar properties to the STFT with a rectangular window when $L_1 = 0$, and it has similar properties to the CWT when $L_1 \neq 0$. However, Eq. (8) shows how much of $\overline{\overline{\mathbf{\Phi}}}\,\hat{\overline{\mathbf{c}}}$ is contained in $\mathbf{y}$.

A more practical approach used in this contribution is to first remove (suppress) the known constituents from each segment $\mathbf{y}$, and then analyze the residual segment $\hat{\mathbf{g}}$. More precisely, first use the model $\mathbf{y} = \underline{\mathbf{\Phi}}\,\underline{\mathbf{c}}$ to estimate $\underline{\mathbf{c}}$ as

$$\hat{\underline{\mathbf{c}}} = \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}, \qquad (9)$$

where $\underline{\mathbf{N}} = \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}$, and so $\hat{\mathbf{g}} = \mathbf{y} - \underline{\mathbf{\Phi}}\,\hat{\underline{\mathbf{c}}}$. Then use the model $\mathbf{y} = \overline{\overline{\mathbf{\Phi}}}\,\overline{\mathbf{c}} = \underline{\mathbf{\Phi}}\,\underline{\mathbf{c}} + \mathbf{\Phi}\,\mathbf{c}$ to estimate $\mathbf{c}$ as

$$\hat{\mathbf{c}} = \mathbf{N}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\hat{\mathbf{g}}, \qquad (10)$$

where $\mathbf{N} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\,\underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}$ (see Appendix A). Let $\mathbf{J} = \mathbf{\Phi}\mathbf{N}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}$. Define the weighted LSWS as

$$s(t_j, \omega_k) = \frac{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P_y}\mathbf{J}\hat{\mathbf{g}}}{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P_y}\hat{\mathbf{g}}} \in (0, 1), \qquad (11)$$

where $1 \leq j \leq n$ and $1 \leq k \leq \kappa$. This spectrogram shows how much the sinusoids of frequency ω_k contributes to each residual segment of $\mathbf{f}$ and is called percentage variance when multiplied by 100. This definition is very useful in searching for hidden signals (short duration signals with very low amplitude with respect to the total amplitude of the data points) in a time series (e.g., Example 3). The ordinary LSWS does not consider the weight matrix $\mathbf{P_y}$ in the calculation of the spectrogram.

To better understand how to determine the weight matrix $\mathbf{P_y}$ for each segment $\mathbf{y}$, a simple example is provided. Let $\mathbf{P}$ be of order $n = 10$, and $M = 10$, $\omega_1 = 2$,

$L_1 = 1$ and $L_0 = 0$. It is easy to see that $L(\omega_1) = 5$, and so for $j = 1$ and $j = 10$, the top left and the bottom right principal submatrices of $\mathbf{P}$ (margins) have dimension 3 (green matrices in Fig. 1), and the submatrices shown by blue have dimension 5. Note that the correlations between data points in the entire time series are considered in each $\mathbf{P_y}$ due to the inversion of the entire $\mathbf{C_f}$. In practice, $\mathbf{P}$ is usually diagonal that means the correlations of the time series values are not given. Note that $\mathbf{P_y}$ is a positive definite matrix because $\mathbf{P}$ is positive definite (Harville 2008), and so Eqs. (7), (9) and (10) are optimal.

$$\mathbf{C_f}^{-1} = \mathbf{P} =$$
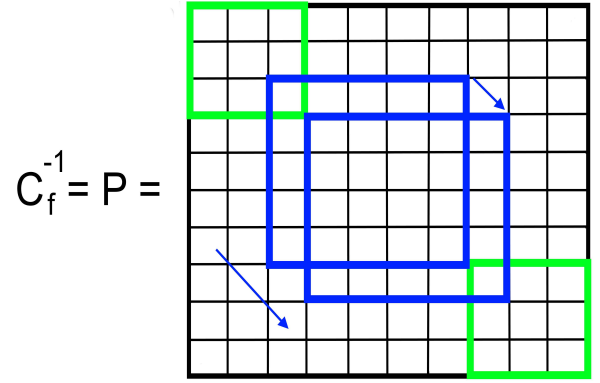
Fig. 1 The weight matrix $\mathbf{P}$ with some of its principal submatrices

The known constituents contaminate the spectrogram given by Eq. (8), but they do not contaminate the spectrogram given by Eq. (11) because they are removed from the segments of the time series. However, Eq. (8) shows how much the constituents of known form and the sinusoids of a particular frequency simultaneously contribute to the segments. Note that the least-squares spectrum (LSS) is a special case of the LSWS that is independent of time and has only one segment (the entire time series). The reader is referred to Wells et al. (1985) and Pagiatakis (1999) for more details.

2.2 Practical Implementation of the Least-squares Wavelet Algorithm

To visualize the concept of the segmentation of a time series better, the term window is used. In this contribution, two different characteristics for the window are used, namely, its size and its length.

For a time series of n data points, a window of size $R(j, k)$ given by Eq. (4) located at t_j, $1 \leq j \leq n$, is a window that contains $R(j, k)$ data points comprising $(L - 1)/2$ data points on either side of t_j. If $1 \leq j < (L + 1)/2$, then a window of size $R(j, k)$ lo-
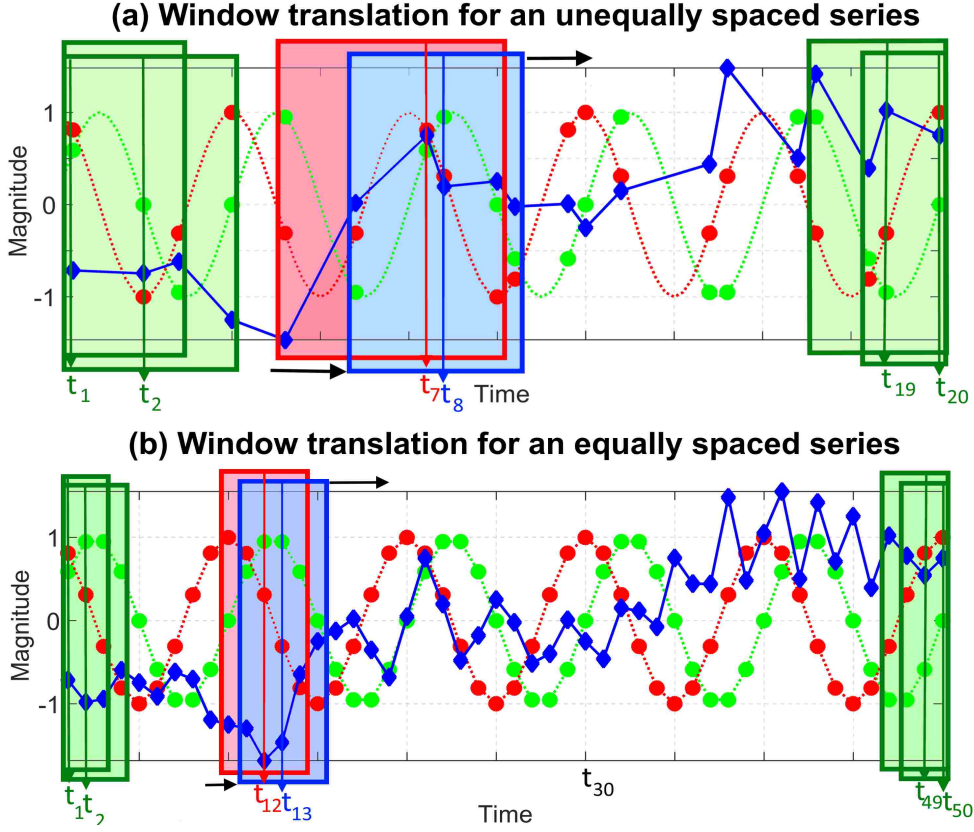
Fig. 2 (a) An unequally spaced and (b) an equally spaced time series (blue diamonds) along with sinusoids (cosine with red circles and sine with green circles) and some windows

cated at t_j is a window that contains $(L-1)/2$ data points on the right hand side and $j-1$ data points on the left hand side of t_j; this window is referred to as the left marginal window. Similarly, one can identify the number of data points for the right marginal window. From Eq. (1), when j runs from j_0 to j_1, a window of size $R(j_0, k)$ located at t_{j_0} translates to a window of size $R(j_1, k)$ located at t_{j_1}. In the LSWA, at each step, the sinusoids and the constituents of known forms are fitted simultaneously to each segment of the time series within a window of size $R(j, k)$ located at t_j to estimate their amplitudes and phases in the segment.

The length of a window located at t_j, denoted by $\ell(t_j)$, can be calculated as follows. If $(L+1)/2 \leq j \leq n - (L-1)/2$, then

$$\ell(t_j) = t_{j+(L-1)/2} - t_{j-(L-1)/2}. \tag{12}$$

If $1 \leq j < (L+1)/2$ (the left marginal windows), then

$$\ell(t_j) = t_{j+(L-1)/2} - t_1, \tag{13}$$

and if $n - (L-1)/2 < j \leq n$ (the right marginal windows), then

$$\ell(t_j) = t_n - t_{j-(L-1)/2}. \tag{14}$$

From Eqs. (12), (13) and (14), it is clear that the length of the window located at t_j depends on the distribution of the t_i's.

In order to illustrate these concepts, an unequally spaced and an equally spaced time series along with the sine and cosine base functions of frequency ω_k are illustrated in Fig. 2a and b, respectively. In Fig. 2a, the red window is of size 5 (includes 5 data points) located at t_7, the median time mark within the window. Similarly, the blue window is of size 5 located at t_8 which is again the median of the window. The length of the red window is $\ell(t_7) = t_9 - t_5$, and the length of the blue window is $\ell(t_8) = t_{10} - t_6$. In this case, $\ell(t_7) > \ell(t_8)$. For the equally spaced time series (Fig. 2b), it is clear that the red and blue windows have the same size and length. The left and right marginal windows along with the green windows and the translations are shown in these figures.

For an unequally spaced time series, the window length varies but the window size remains constant during translation except for the marginal windows. For unequally spaced time series, if the length of the window is fixed during translation rather than its size, then there may not exist enough data points within

(a) Window translation for a lower frequency

(b) Window translation for a higher frequency
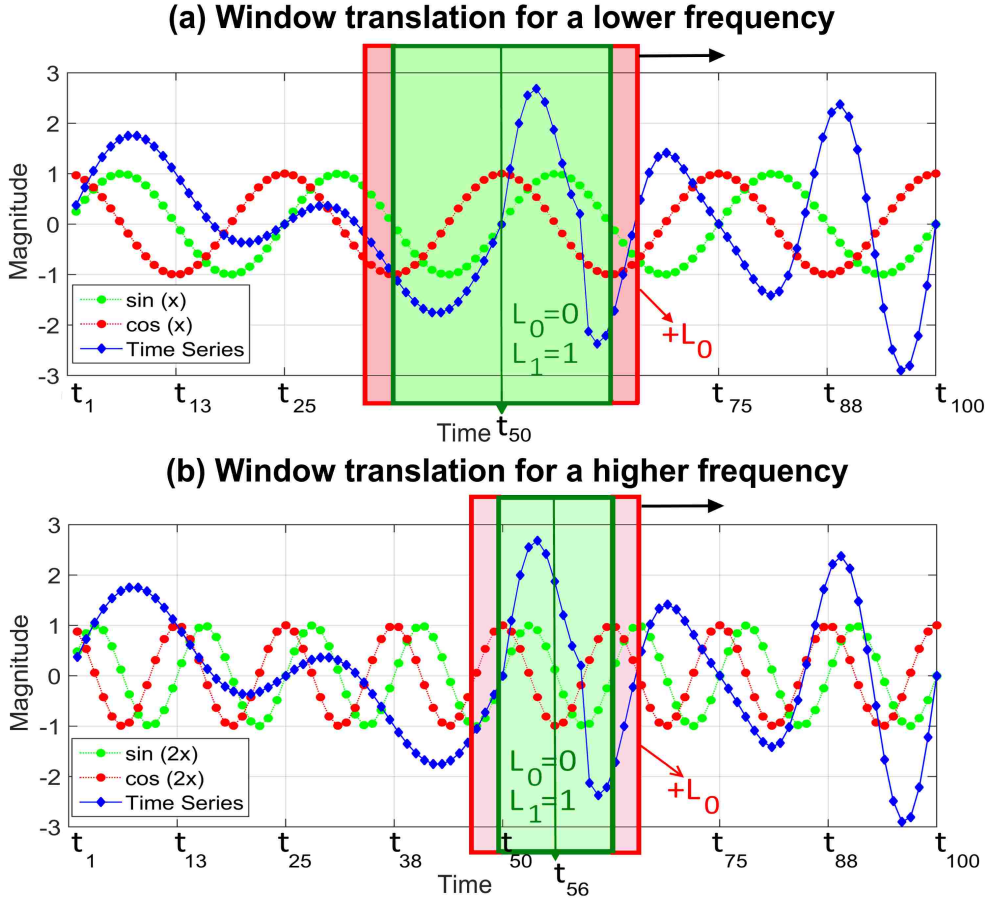
Fig. 3 Window translation of an equally spaced time series (blue) for (a) $\sin(x)$ and $\cos(x)$ and (b) $\sin(2x)$ and $\cos(2x)$

the window located at t_j to calculate $s(t_j, \omega_k)$. Thus, it is more appropriate to fix the size of the translating window. Since in Eq. (1), j runs from 1 to n, one unit at a time, the window located at t_j overlaps with the window located at t_{j+1} by $L-1$ data points when $(L+1)/2 \leq j \leq n-(L-1)/2$ (red and blue windows in Figs. 2a and b). The overlaps for the marginal windows are also shown in Figs. 2a and b (green windows).

Figures 3a and b show the changes in the window size for an equally spaced time series when $L_1 = 1$ cycle with fixed L_0. The window size decreases when the frequency increases allowing the detection of short duration waves (signal, transients, etc.), and L_0 (a constant number of samples) increases the size of the window so that the resolution may be adjusted based on the specific scope of analysis.

For a time series (equally or unequally spaced), the sinusoidal functions (and the constituents of known forms) are generally not orthogonal within a window of size $R(j, k)$ or length $\ell(t_j)$ located at t_j. The LSWA considers their correlations by the off-diagonal elements of the matrix of normal equations (similar to the LSSA). For each t_j, Eq. (11) shows how much the sinusoids of

frequency ω_k contribute to the residual segment of the time series within a window of size $R(j, k)$ or length $\ell(t_j)$ located at t_j.

Puryear et al. (2012) simultaneously include the sinusoidal functions with different frequencies in the columns of design matrix $\boldsymbol{\Phi}$, which makes the system $\mathbf{y} = \boldsymbol{\Phi}\mathbf{c}$ underdetermined, and then they use constraints to calculate the spectrogram. However, the LSWS is calculated by trying each frequency ω_k one at a time, so the system is highly overdetermined and the matrix of normal equations remains regular by selecting appropriate L_0 and L_1, window size parameters.

When analyzing an unequally spaced time series, spectral leakages (peaks that do not correspond to signals) appear in the LSS and LSWS. However, the leakages are far less than the leakages in the Fourier transform because the correlation among the sinusoids is considered for each frequency in the calculation of the LSS and LSWS. Since the frequencies are examined one at a time (out-of-context), the leakages appear in the LSS and LSWS. By considering the correlation among the sinusoids of different frequencies (the frequencies of the constituents of known forms), the spectral leak-

ages will be mitigated (Craymer 1998), and this can be achieved by simultaneously removing the constituents of known forms from the segments.

The selection of parameters L_0 and L_1 for the analysis of a time series mainly depends on whether the time series is weakly or strongly non-stationary. In addition, it depends on the time and frequency resolution as well as the number of constituents being estimated within the window whose size should be adequate to avoid singularity of the matrix of normal equations in a desired frequency band. Examples 1 and 3 show the effect of various selections of L_0 and L_1 on the true signal peaks in the spectrogram for equally and unequally spaced time series. In many applications, for time series with series more than 50 data points per unit time, it is recommended that L_0 be 20 to 30 samples, and L_1 be 3 or 4 cycles to avoid the singularity of $\underline{\mathbf{N}}$ and $\mathbf{N}$ when removing the constituents of known forms (usually less than 10). In practical applications, especially for unequally spaced time series, choosing $\underline{\boldsymbol{\Phi}} = [\mathbf{1}, \mathbf{t}]$ makes the mean values of the segments approximately zero because even if there is no systematic trend present in an unequally spaced time series, the sine and cosine functions no longer have zero mean causing an error in determining the zero point of the signals and reducing the percentage variance of the signal peaks in the spectrogram.

In the following examples, whenever the CWT is applied for analysis, the MATLAB command cwt($\mathbf{f}$, scales, wname) is used, where $\mathbf{f}$ is time series values, scales is vector of positive scales for the CWT scalogram, and wname is name of a wavelet (e.g., morl for Morlet wavelet). One may convert the scales to frequencies using center frequencies of scales by calling MATLAB command scal2frq(scales, wname, DELTA) to obtain a spectrogram, where DELTA is sampling period. This is not a correct approach of analyzing unequally spaced time series. One of the advantages of the LSWA over the CWT is that the LSWA does not convert any scale to frequency to obtain a spectrogram, and it directly decomposes a time series to time-frequency domain.

Example 1.

In this example, the sensitivity of the LSWS for different window size parameters is examined. Consider the following hyperbolic chirp signal (Mallat 1999)

$$f(t_j) = \cos\left(\frac{230}{2.3 - t_j}\right), \tag{15}$$

where $t_j = (0.001j)$s, and j runs from 1 to 2000 (Fig. 4a). Note that the instantaneous frequency of the hy-

perbolic chirp is $230/\left(2\pi(2.3 - t_j)^2\right)$ Hz. Since there are 1000 data points per second, $M = 1000$.

Figure 4b shows the ordinary LSWS with $L_0 = 201$ and $L_1 = 0$. It can be seen that the true peaks at higher frequencies are not well resolved, in other words, the peaks lose their percentage variance in higher frequencies (arrows). The reason is that the window size is fixed to 201 (except for the marginal windows), and the time series is changing its frequency rapidly as time advances. Thus, the sinusoidal base functions do not fit the segments of the time series at higher frequencies as well as they do at lower frequencies.

The ordinary LSWS with $L_0 = 0$ and $L_1 = 1$ is shown in Fig. 4c. Although the percentage variance of the true peaks in the spectrogram does not change across the time-frequency domain, the spectrogram has low-frequency resolution at high frequencies. Furthermore, the spikes appear in the spectrogram at the times when the series has zero magnitude (arrows) because one cycle of sinusoids also fits those segments that cross the horizontal axis. These spikes may also have applications as they show the behavior of the time series.

If one selects $L_0 = 20$ and $L_1 = 2$, then the percentage variance of the spikes reduces because two cycles and some additional data points are incorporated in the fitting process resulting in a better resolution (Fig. 4d). In fact, the selection of the window size parameters is based on the purpose of analysis that is similar to the selection of window size in the STFT. The magnitude of $\hat{\mathbf{c}}$ in Eq. (10) is approximately 1 at the signal peaks.

Figure 4e shows the CWT after applying the MATLAB command cwt($\mathbf{f}$, scales, wname) with wname=bior1.3, a biorthogonal wavelet that has similar structure to the sinusoids (Cohen et al. 1992). Note that the scalogram is converted to spectrogram using the center frequencies of scales. The color bar values represent the absolute values of the CWT coefficients (abs CWT coefficients). The higher frequencies in the CWT spectrogram are not well resolved and lose power (indicating that the amplitude of the chirp signal is decreasing over time) because the spectrogram is computed in terms of frequency bands (i.e., scales) that overlap each other more when frequency increases (Sinha et al. 2005).

The spikes in the CWT spectrogram appear at the same positions as in the LSWS. However, since the LSWS is normalized and the correlation among sinusoids is taken into account, the spikes in the LSWS are sharper. Another advantage of the LSWS is that the true peaks are not scattered as the CWT spectrogram because the least-squares minimization is used in translation. Note that the LSWS shows how much the sinusoids of a particular frequency within a
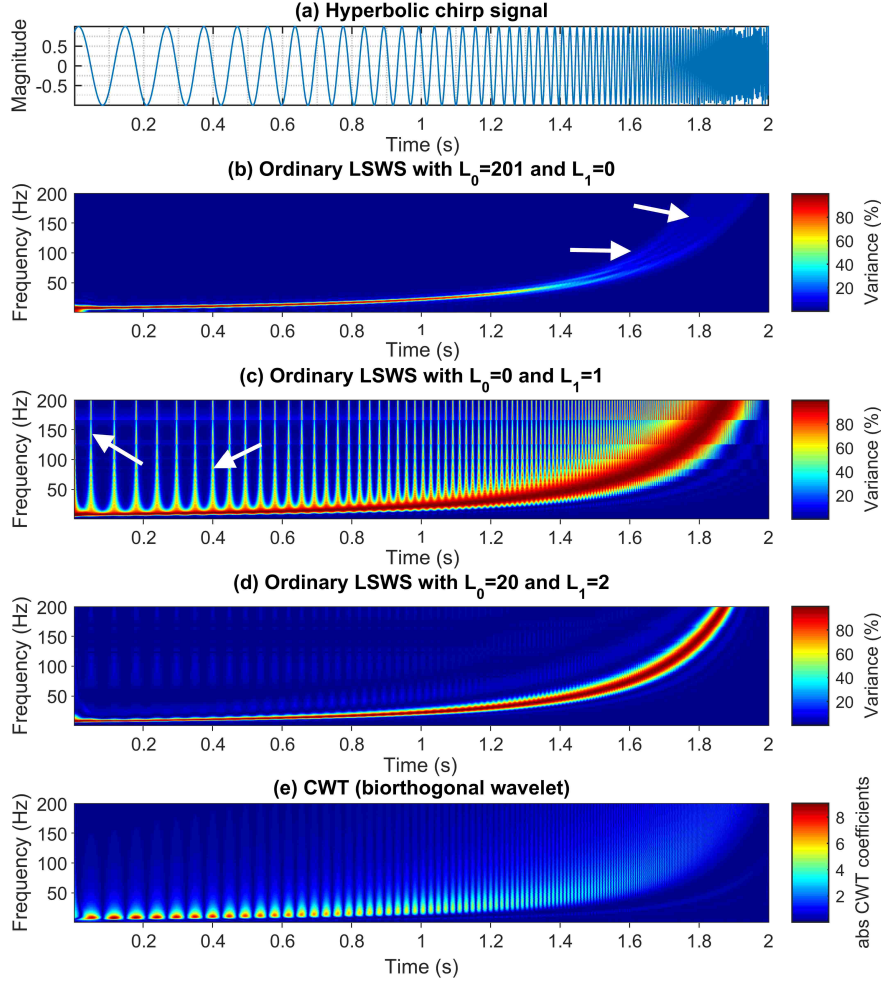
Fig. 4 The equally spaced chirp signal given by Eq. (15) and its analyses

window contributes to the segment of the chirp signal within the window.

Example 2.

Consider the following equally spaced time series which is the sum of two hyperbolic chirp signals

$$f(t_j) = \cos\left(\frac{230}{2.3 - t_j}\right) + \cos\left(\frac{500}{2.3 - t_j}\right), \quad (16)$$

where $t_j = (0.001j)$s, and j runs from 1 to 2000 (Fig. 5a). Note that the instantaneous frequencies of the hyperbolic chirps are $230/\left(2\pi(2.3 - t_j)^2\right)$ Hz and $500/\left(2\pi(2.3 - t_j)^2\right)$ Hz.

The CWT spectrogram with a Morlet wavelet (most commonly used; similar to a sine wave in attenuation) is shown in Fig. 5b. One can see that the CWT peaks lose power toward higher frequencies (white arrows). This might be misinterpreted as the amplitudes of the hyperbolic chirps are decreasing over time (attenuating) which is not true because the hyperbolic chirps

constructed in this example have constant amplitude over time. Comparing Figs. 4e and 5b, one can observe that using the Morlet wavelet mitigated the effect of the spikes in the spectrogram.

Figure 5c shows the ordinary LSWS of the time series given by Eq. (16) for $L_0 = 20$ and $L_1 = 2$. Using these window size parameters, the true signal peaks are very well resolved, and their percentage variances do not change significantly across the time-frequency domain. Similar to Example 1, the effect of the spikes (appears toward the locations in time where the time series has zero magnitude) is also mitigated in the spectrogram (white arrows).

The spikes can be mitigated further by defining an appropriate weight function such as the Gaussian function. For instance, let $\mathbf{P_y}$ be the diagonal matrix of order $R(j, k)$ given by Eq. (4) for the window located at τ (choose $\tau = t_j$) whose diagonal entries are given by the following Gaussian function

$$G(i) = e^{-c(2\pi\omega_k(t_i - \tau))^2}, \quad (17)$$
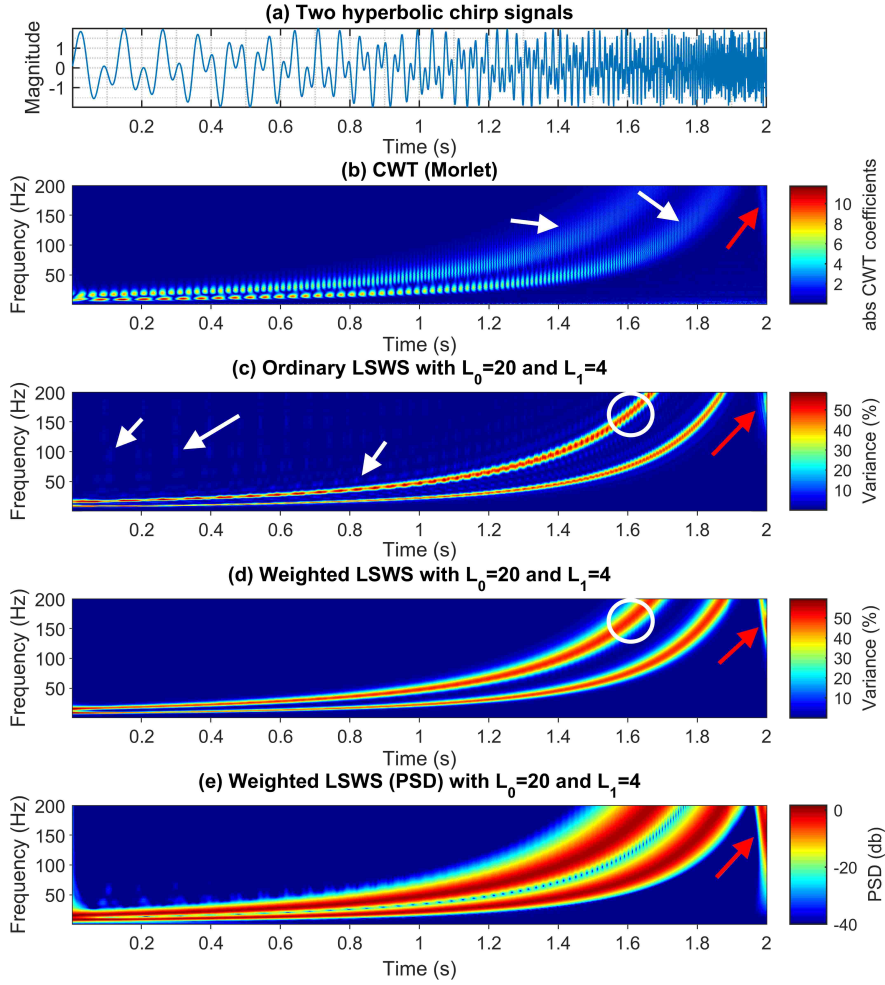
Fig. 5 The time series given by Eq. (16) and its analyses

where c is a window constant, and the t_i's are the times of the data points within the window of size $R(j, k)$. In astronomical applications, a popular choice for c would be 0.0125 (Foster 1996). By this selection of $\mathbf{P_y}$, the sinusoidal basis functions will be adapted to the Morlet wavelet in the least-squares sense for analysis. The weighted LSWS is illustrated in Fig. 5d by considering the weights defined by Eq. (17). It can be seen that the effect of the spikes is mitigated compared to Fig. 5c, but the bandwidth of the true spectral peaks increased (poor frequency resolution, inside the circles in Figs. 5c and d). Defining the weights using Eq. (17) may also change the true peak locations (Foster 1996). Thus, it is recommended that the ordinary LSWS be used when there is no covariance matrix associated with a time series. Figure 5e shows the LSWS given by Eq. (11) in terms of power spectral density (PSD) in decibels (dB) defined by

$$\text{PSD}_{\text{LSWA}} = 10 \log_{10} \left(\frac{s(t_j, \omega_k)}{1 - s(t_j, \omega_k)} \right). \tag{18}$$

The values less than -40 dB are set to -40 dB. Note that the peaks shown by red arrows in Figs. 5b to e are the alias effect of the signal $\cos \left(500/(2.3 - t_j) \right)$.

2.3 Comparison Between the Least-squares Wavelet Spectrogram and the Weighted Wavelet Z-transform

In this section, the same notation is used as in Foster (1996) to define the WWZ. Suppose that $\mathbf{f} = \left[f(t_j) \right]$ is a time series of n data points. For each τ (the window location; can be the t_j's or equally spaced times) and each $\omega_k \in \mathbf{\Omega}$, let $\phi_1 = \mathbf{1}$

$$\phi_2 = \left[\cos \left(2\pi\omega_k(t_1 - \tau) \right), \dots, \cos \left(2\pi\omega_k(t_n - \tau) \right) \right],$$
$$\phi_3 = \left[\sin \left(2\pi\omega_k(t_1 - \tau) \right), \dots, \sin \left(2\pi\omega_k(t_n - \tau) \right) \right], \tag{19}$$

where $\mathbf{1}$ is the all-ones vector of dimension n. The inner product of two vectors $\mathbf{u} = [u(t_1), \dots, u(t_n)]$ and $\mathbf{v} =$
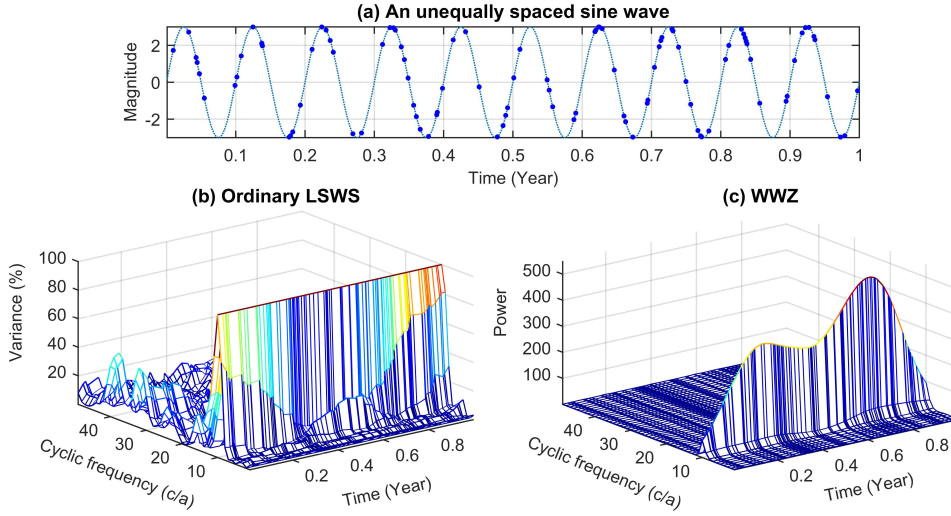
Fig. 6 An unequally spaced pure sine wave of cyclic frequency 10 (c/a) and its analyses

$[v(t_1), \ldots, v(t_n)]$ is defined as

$$\langle \mathbf{u} | \mathbf{v} \rangle = \frac{\sum_{i=1}^{n} w_i u(t_i) v(t_i)}{\sum_{i=1}^{n} w_i}, \tag{20}$$

where w_i is the statistical weight chosen as $w_i = e^{-c(2\pi\omega_k(t_i-\tau))^2}$, and c is a window constant which may be selected as $c = 0.0125$ as discussed in Example 2. The constant c determines how rapidly the analyzing wavelet decays (Foster 1996). Let $\mathbf{S}$ be the square matrix of order 3 whose ab entry is $\langle \phi_a | \phi_b \rangle$ ($1 \le a, b \le 3$). The weighted wavelet Z-transform (WWZ) (power) is defined as

$$Z = \frac{(N_{\text{eff}} - 3)V_y}{2(V_x - V_y)}, \tag{21}$$

where $N_{\text{eff}} = (\sum_{i=1}^{n} w_i)^2 / (\sum_{i=1}^{n} w_i^2)$ is the effective number of data points, $V_x = \langle \mathbf{f} | \mathbf{f} \rangle - \langle \mathbf{1} | \mathbf{f} \rangle^2$ is the weighted variation of the data, and $V_y = \sum_{a,b} \mathbf{S}_{ab}^{-1} \langle \phi_a | \mathbf{f} \rangle \langle \phi_b | \mathbf{f} \rangle - \langle \mathbf{1} | \mathbf{f} \rangle^2$ is the weighted variation of the model function obtained by the least-squares method (Foster 1996). Note that the term $V_y/(V_x - V_y)$ in Eq. (21) is the estimated signal-to-noise ratio (Foster 1996), and the numerator of Eq. (11) is the estimated signal and its denominator is the sum of the estimated signal and noise (Pagiatakis 1999). The least-squares method used in the LSWS and WWZ is a great tool that estimates the amplitude of a physical fluctuation much more accurately than the traditional Fourier transform. The amplitude of signals in the LSWA can be estimated using Eq. (10) after finding the frequency or period of the signals using the LSWS that is similar to the weighted wavelet amplitude (WWA) (Foster 1996). Both the LSWS and WWZ are excellent in frequency localization of the signals; however, the WWZ

is a poor estimator of amplitude (Foster 1996), and the spectral peaks in the WWZ lose their power toward higher frequencies similar to the CWT (Fig. 5b). The latter shortcoming is caused mainly by the effective number of data points that decreases when the frequency increases, making the numerator of Eq. (21) smaller. Subtracting V_y from V_x in the denominator of Eq. (21) alters significantly the true power of the signals that may not allow one to assess the behavior of the time series, especially when searching for possible hidden signals with low power in a time series.

In both the LSWS and WWZ, the locations of the windows can be equally spaced; however, the locations are chosen to be at the times corresponding to the time series values simply because there might be signatures in the time series gaps that cannot be detected using the available data points. An unequally spaced pure sine wave of cyclic frequency 10 cycles per annum (c/a) and amplitude 3 is generated by randomly selecting 100 data points per year ($M = 100$) and demonstrated in Fig. 6a. The ordinary LSWS is shown in Fig. 6b with $L_1 = 4$, $L_0 = 10$ and $\phi_1 = [\mathbf{1}]$. The percentage variance of the peak at 10 (c/a) is 100% at all times. Note that the spectral leakages in the spectrogram are not significant at certain confidence level defined for the distribution of the LSWS (Sect. 2.4). Figure 6b shows the WWZ with $c = 0.0125$. The peaks at the cyclic frequency 10 (c/a) have variable power over time, but the spectral leakages are mitigated (they exist but with lower power) because of the subtraction in the denominator of Eq. (21).

The LSWA can detect and suppress significant peaks (that takes into account the correlations among the sinusoids of different frequencies), and so it can be
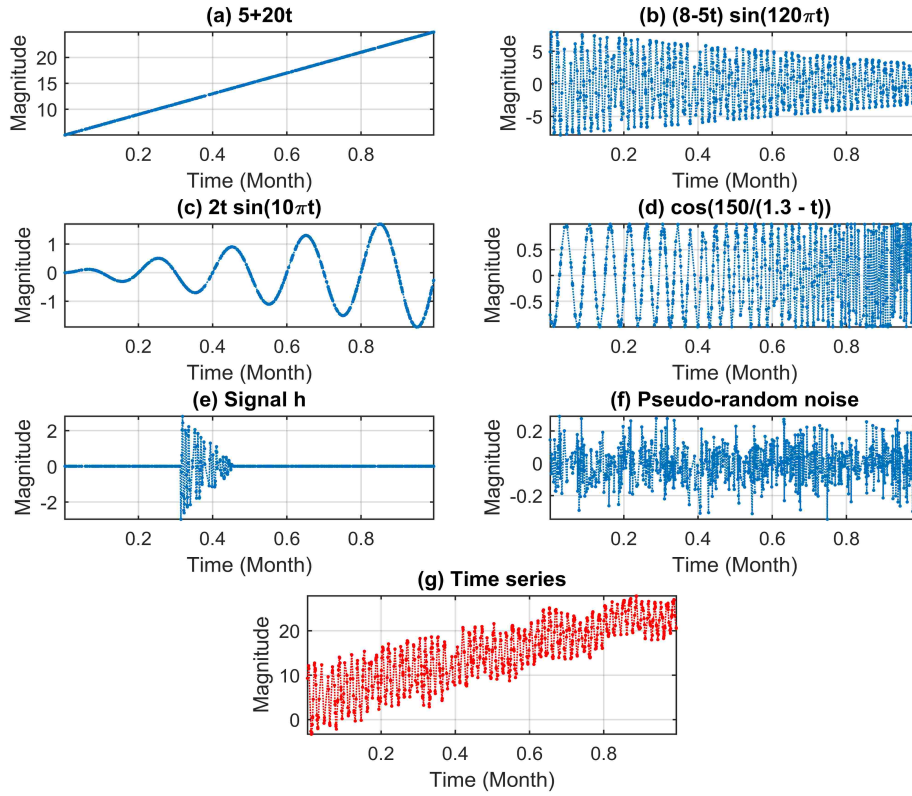
Fig. 7 The unequally spaced time series given by Eq. (22) and its constituents. (a) to (f) the constituents and (g) the time series (the sum of the constituents)

used to search for the possible hidden signals in a time series. Suppressing the significant peaks simultaneously mitigates the spectral leakages in the LSS and LSWS of the unequally spaced residual series and results in strengthening the peaks of possible hidden signals (see the following two examples). Another advantage of using the ordinary LSWS is the computational efficiency that is much better than the WWZ because the weights are not involved in the computation. Moreover, the window size in the LSWS decreases when frequency increases; whereas, the window size in the WWZ is the same at all times and frequencies (Eq. (19)), fact that significantly increases the computational cost.

Example 3.

Consider the following inherently unequally spaced time series simulated for one month

$$
\begin{aligned}
f(t_j) = & 5 + 20t_j + (8 - 5t_j)\sin(2\pi \cdot 60t_j) \\
& + 2t_j \sin(2\pi \cdot 5t_j) + \cos\left(\frac{150}{1.3 - t_j}\right) \\
& + h(t_j) + wgn(t_j),
\end{aligned}
\tag{22}
$$

where $1 \le j \le 720$, $0 \le t_j \le 1$, and wgn is white Gaussian noise generated by the MATLAB function wgn. Also

$$
h(t_j) = \begin{cases} (9 - 0.03j)\sin(2\pi \cdot 120t_j) & \text{if } 200 \le j \le 300, \\ 0 & \text{otherwise.} \end{cases}
\tag{23}
$$

The linear trend, the two sine waves of variable amplitudes and constant cyclic frequencies 60 and 5 cycles per month (c/m), the hyperbolic chirp, the short duration sine wave **h** given by Eq. (23) of variable amplitude and constant cyclic frequency 120 c/m, the white Gaussian noise and the time series **f** are illustrated in Figs. 7a to g, respectively.

In this example, it is intended to search for signals **h**, the hyperbolic chirp, and the sine wave of constant cyclic frequency 5 c/m. Note that the white Gaussian noise, the linear trend and the sine wave of constant cyclic frequency 60 c/m are considered noise in this example that highly contaminate the time series. A notable advantage of the LSSA, LSWA and WWZ is that one may select any set of spectral frequencies based on the purpose of analysis. The set of cyclic frequencies chosen for the analysis in this example is $\Omega = \{1, 2, 3, \ldots, 200\}$ c/m.
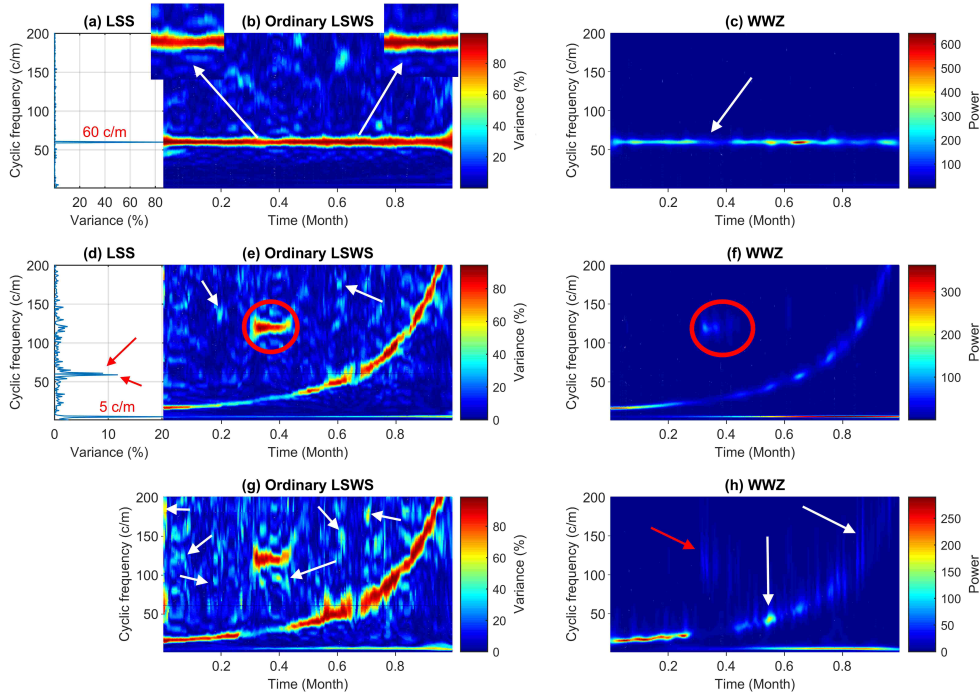
Fig. 8 The analyses of the unequally spaced time series shown in Fig. 7g

Figure 7g shows that the time series has a trend. The LSS is illustrated in Fig. 8a (white panel) after removing the trend. The LSS detects one strong peak at 60 c/m, but it does not explain the nature of the constituent in **f** that creates this peak, in other words, the LSS does not show whether this peak is for a wave of variable or constant amplitude over time, or whether the duration of the wave is short or long.

Since there are 720 samples per month, $M = 720$. Figure 8b shows the ordinary LSWS with $L_0 = 10$ and $L_1 = 4$ after removing the trend (i.e., by setting $\underline{\boldsymbol{\Phi}} = [\mathbf{1}, \mathbf{t}]$). The peaks corresponding to sine wave of constant cyclic frequency 60 c/m can be observed in Fig. 8b (horizontal reddish band in the spectrogram). The percentage variances of the spectral peaks in the spectrogram show how much each residual segment of **f** of size $R(j, k)$ contains a constituent of cyclic frequency ω_k. For instance, the peaks shown by the left zoomed panel have lower percentage variance than the peaks shown by the right zoomed panel indicating that there might be other constituents in the residual segments approximately from time 0.3 to 0.4 month.

Figure 8c shows the WWZ of the time series after removing the linear trend estimated by the least-squares method (approximately $5 + 20t_j$). Similar to the LSWS, the WWZ shows the peaks corresponding to the sine wave of constant cyclic frequency 60 c/m, and the effect of the spectral leakages is mitigated. Although the peaks from 0.3 to 0.4 month lose their power (arrow), it

cannot be ascertained that there might be some hidden components of different frequency in that time period because of the discussion before this example (Fig. 6c).

In the next step, one may suppress the peak at 60 c/m (remove the sinusoidal wave of 60 c/m) to search for other possible signals. The LSS of the residual series after suppressing the peak is shown in Fig. 8d. The LSS now shows several significant peaks. Note that the sine wave of 60 c/m has variable amplitude. Since the sine wave of constant amplitude is used to fit the entire **f**, the wave will not fit well to **f**, and so some peaks appear around 60 c/m in the LSS for the residual series (red arrows). If researchers are not aware of a phenomenon that causes variable amplitude waves in a time series, this can be misinterpreted. Furthermore, the hyperbolic chirp and series **h** cannot be studied from the LSS, and the LSS does not show the amplitude variation for the sine wave of 5 c/m.

Selecting $\underline{\boldsymbol{\Phi}} = [\mathbf{1}, \mathbf{t}, \cos(2\pi \cdot 60\mathbf{t}), \sin(2\pi \cdot 60\mathbf{t})]$ in the LSWA removes the trend and the constituent of cyclic frequency 60 c/m simultaneously from the segments of **f** (Fig. 8e). From Fig. 8e, one can study the spectral peaks corresponding to signals **h** (inside the circle), the hyperbolic chirp with variable frequency over time and the sine wave of variable amplitude and constant cyclic frequency 5 c/m. Note that the actual nature of these signals cannot be studied from Fig. 8d. The spectral peaks corresponding to the sine wave of variable amplitude and constant cyclic frequency 60 c/m are com-

pletely removed from the spectrogram shown in Fig. 8e because sinusoids within the windows fit better to the segments of the series, but this is not achieved in the LSS (arrows in Fig. 8d).

Figure 8f shows the WWZ of the sum of series shown in Figs. 7c to f. One can clearly see that signal **h** is not completely detected (inside the circle). The reason is that the effective numbers of data points are small in certain time periods of **h** that reduced the power significantly, and also the WWZ is defined in terms of signal-to-noise ratio that reduced the power of false peaks (mitigated the spectral leakages). A stochastic surface (to be defined in the next section) can, in fact, flag the peaks shown by arrows in Figs. 8e and g as noise at certain confidence level (usually 99%).

To see the sensitivity of the LSWS and WWZ to the window width, the window size parameters are chosen as $L_1 = 2$ and $L_0 = 20$ with the same $\underline{\Phi}$ as in Fig. 8e, and the ordinary LSWS is shown in Fig. 8g. In addition, the WWZ is shown in Fig. 8h by selecting $c = 0.1$. The frequency localizations in both spectrograms are now poor (poor frequency resolutions), but the WWZ shows the significant variability of power compare to Fig. 8f, and signal peaks are not well localized (arrows).

In Figs. 8e to h, since the window size decreases when frequency increases, the spectral peaks corresponding to signal **h** with cyclic frequency 120 c/m have higher time resolution but poorer frequency resolution, and the spectral peaks corresponding to signal of cyclic frequency 5 c/m have higher frequency resolution but poorer time resolution.

2.4 Stochastic Surfaces in the Least-squares Wavelet Spectrograms

In this section, a similar methodology is used as in Pagiatakis (1999) to determine stochastic surfaces for the least-squares wavelet spectrogram (LSWS) that define confidence levels (usually 95% or 99%) above which spectral peaks in the LSWS are statistically significant. It is also shown that the stochastic surfaces of the LSWS are independent of the a-priori variance factor σ_0^2, that is, the stochastic surfaces are independent of any scale defect of the covariance matrix associated with the time series. The reader is referred to Vanicek and Krakiwsky (1986) for more details on the variance factor.

A remarkable property of the least-squares wavelet spectrogram defined by Eqs. (8) and (11) is that it provides the percent of a specific component(s) (as defined by $\overline{\Phi}$) contained in **f**. This ratio includes the weight matrix in both numerator and denominator and as such,

any scale defect in the covariance matrix cancels out. This leads to the conclusion that the probability distribution function (PDF) of the spectrogram does not depend on either the a-priori or a-posteriori variance factors. Therefore, the derivation of the PDF of the spectrogram will be based on previous work by Pagiatakis (1999) considering only the case of a-priori variance factor being known since the treatment of the case of unknown a-priori variance factor (a-posteriori variance factor used) was in error.

Suppose that **f** has been derived from a population of random variables following the multidimensional normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{C_f})$, where $\mathbf{C_f}$ may not be regular. Following similar techniques in Steeves (1981) and Pagiatakis (1999), it can be seen that the LSWS given by Eq. (11) follows the beta distribution with parameters 1 and $\Re/2$, where $\Re = R(j,k) - q - 2$ (q is the number of constituents of known forms). In other words, $s(t_j, \omega_k)$ follows $\left[1 + (\Re/2) F_{\Re,2} \right]^{-1} = \beta_{1,\Re/2}$, where F and β are the F-distribution and the beta distribution, respectively. Therefore, the stochastic surface at $(1-\alpha)$ confidence level (usually $\alpha = 0.01$ or $\alpha = 0.05$) for the spectrogram given by Eq. (11) is

$$\zeta_{(t_j, \omega_k)} = \left[1 + \frac{\Re}{2} F_{\Re,2,\alpha} \right]^{-1}. \qquad (24)$$

Following the discussion in Pagiatakis (1999), if $s(t_j, \omega_k) > \zeta_{(t_j, \omega_k)}$, then the spectral peak is statistically significant at $(1-\alpha)$ confidence level.

Note that the stochastic surface given by Eq. (24) depends on the window size, and it does not depend on the frequency when $L_1 = 0$. The stochastic surface is also independent of the time (the location of the window) except for the marginal windows.

Using the relation $F_{\Re,2,\alpha} = F_{2,\Re,1-\alpha}^{-1} = (2/\Re)(\alpha^{-2/\Re} - 1)^{-1}$ (Rao and Mitra 1971), Eq. (24) can be written as

$$\zeta_{(t_j, \omega_k)} = \left[1 + \frac{1}{\alpha^{-2/\Re} - 1} \right]^{-1}. \qquad (25)$$

Following similar techniques in Pagiatakis (1999), it is not difficult to see that the spectrogram given by Eq. (8) follows $\beta_{(q+2)/2, \Re/2}$ with the mean $(q+2)/(\Re+q+2)$ and variance $2(q+2)\Re/((\Re+q+2)^2(\Re+q+4))$; however, in practice, the spectrogram given by Eq. (11), following $\beta_{1,\Re/2}$ with the mean $2/(\Re+2)$ and variance $4\Re/((\Re+2)^2(\Re+4))$, provides better results. The mean and variance of random variables following the beta distribution are discussed in Johnson et al. (1995).

The next section demonstrates an application of the method to an unequally spaced and unequally weighted real time series and shows the importance of

the stochastic surfaces in the identification of statistically significant spectral components in a time series.

3 Application

A very long baseline interferometry (VLBI) length time series is used as an application of the LSWA. The VLBI length series is between stations Westford in the United States of America and Wettzell in Germany. The length of series changes over time due to many reasons such as atmospheric temperature (due to deformation of the antennas), plate tectonic movement, and tides (Titov 2002; Campbell 2004). The time series was obtained from www.ccivs.bkg.bund.de. The series comprises 1733 unequally spaced and unequally weighted baseline length estimates via the least-squares method, from January 9th 1984 at 19:12:00 universal time (UT) to September 3rd 2014 at 16:48:00 UT (Fig. 9a).

Since there are 1733 samples in 30.6 years (unit time is a year), there are approximately 57 samples per annum ($M = 57$). The window size parameters chosen for the analyses are $L_1 = 4$ cycles and $L_0 = 30$ samples. The set of cyclic frequencies selected for this analysis is $\boldsymbol{\Omega} = \{0.1, 0.2, 0.3, \ldots, 12\}$ c/a. Note that the linear trend expresses the lengthening of the baseline due to the relative tectonic plate movement on which the two VLBI antennas are mounted.

Figure 9b shows the ordinary LSWS of the VLBI series and its stochastic surface at 99% confidence level (gray surface) after removing the trend (by setting $\underline{\boldsymbol{\Phi}} = [\mathbf{1}, \mathbf{t}]$). The two-dimensional representation of Fig. 9b is shown in Fig. 9d. The constant annual frequency (1 c/a or ~ 365 days, white arrows) and the semiannual frequency (2 c/a or ~ 182.5 days) are significant in certain years (red arrows), where $\sim$ indicates approximation. The ordinary LSS also shows a significant peak of variance 4.5% at period of ~ 365 days (Fig. 9c). The linear trend estimated by the least-squares method as $5998325.356 + 0.017t$ is removed from the time series, and the result of the WWZ for the residual series is shown in Fig. 9e. Similar to the ordinary LSWS, the WWZ shows the annual and semiannual peaks (arrows in Figs. 9d and e).

The weighted LSS and LSWS after removing the trend are shown in Figs. 9f and g. Many peaks that are not significant in Fig. 9d, are now significant at 99% confidence level (e.g., arrows). In the next step, the annual peaks (having the highest percentage variance) are suppressed, and the spectrum and the spectrogram of the residual series are shown in Figs. 9h and i, respectively. It can be seen that peaks shown by arrows from year 1993 to year 1997 are not significant anymore (the spectral leakages). The peaks at cyclic frequency 1.6

c/a (~ 228 days, red arrow) are suppressed, and the spectrum and spectrogram of the new residual series are shown in Figs. 9j and k, respectively. One can observe that the spectral peaks shown by white arrows in Fig. 9i are no longer significant in Fig. 9k. However, the semiannual peaks shown by red arrows and the peaks inside the circles remained significant after simultaneously removing the linear trend and sinusoids of cyclic frequencies 1 and 1.6 c/a.

In order to investigate what phenomena could possibly be causing the significant peaks in the spectrogram, the atmospheric temperature records at both stations are also analyzed. They are obtained from http://ggosatm.hg.tuwien.ac.at/DELAY/SITE/VLBI/ almost at the same time period of the VLBI series (Figs. 10a and d). The least-squares spectra of the Westford and Wettzell temperature series after removing the trend are shown in Figs. 10b and e, respectively. The annual peaks in the spectra show that the annual peaks in Figs. 9d, e and g are mainly due to the temperature variation causing deformation of the antennas.

The annual peak is suppressed for each time series, and the least-squares spectra of the residual temperature series of Westford and Wettzell are shown in Figs. 10c and f, respectively. The peak at 1.6 c/a in Fig. 10c shows that the Westford antenna is being deformed with a period of ~ 228 days (unusual atmospheric temperature variations), and the evidence of its effect was observed in both spectrograms in Figs. 9g and 9i from year 1997 to year 2001. The LSS of the Westford temperature series from year 1997 to year 2001 (not shown here) has a significant peak at 1.6 c/a (~ 228 days). The semiannual peaks in the spectrograms that are significant since 1997 could also be physically linked to the semiannual temperature variation at the Wettzell station as observed in Fig. 10f. Similarly, the peaks shown inside the circles in Fig. 9k at cyclic frequency of 9.5 c/a (~ 38 days) could possibly be caused by unusual temperature variation at the Wettzell station.

4 Conclusions

The LSSA is an appropriate method of analyzing non-stationary and unequally spaced time series. Contrary to Fourier analysis, the LSSA can account for constituents of known forms, such as trends and/or datum shifts simultaneously with periodic constituents. The covariance matrix associated with the time series (if known) can also be accounted for in the analysis. No editing of the time series of any sort is necessary when using the LSSA. If a non-stationary (in the second moments) time series comprises constituents of low/high
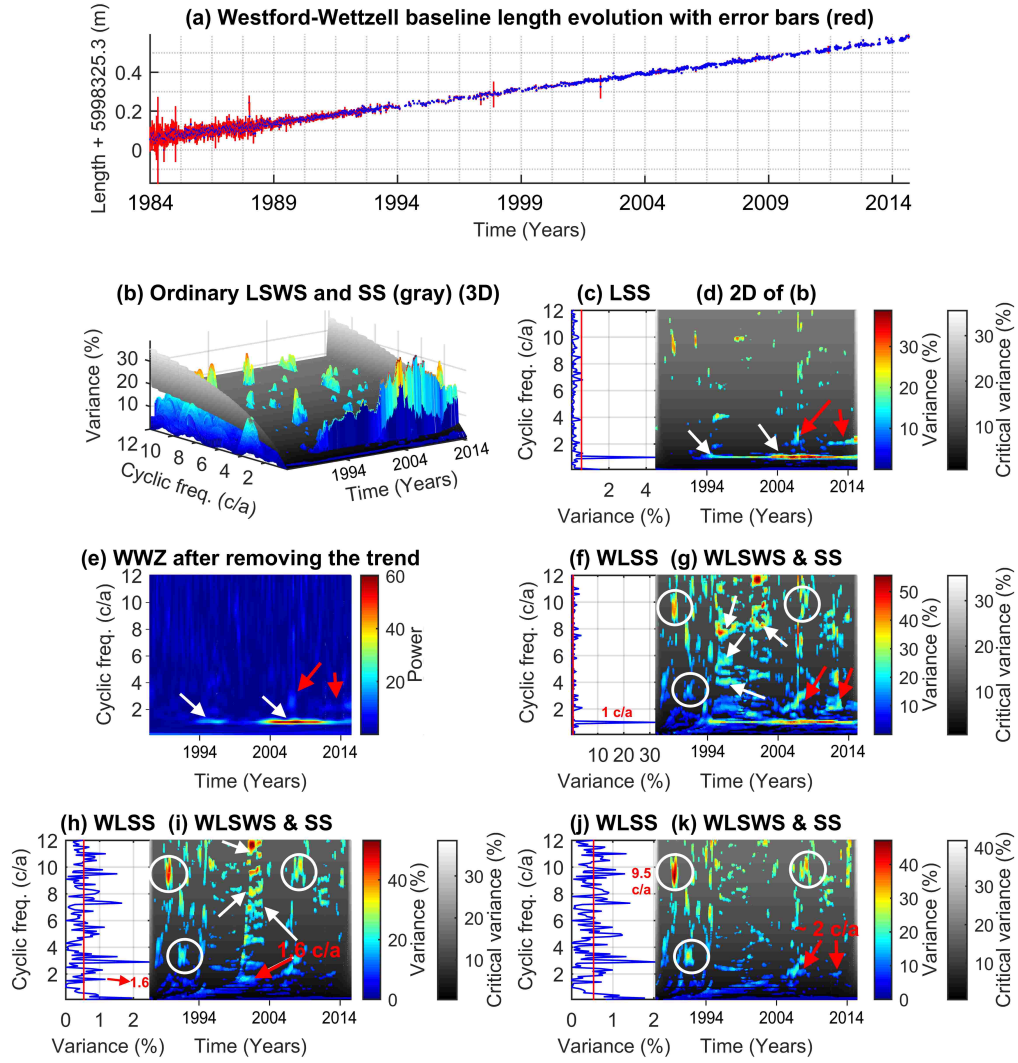
Fig. 9 The VLBI time series and its analyses using the LSS, LSWS and WWZ. Note that WLSS, WLSWS, SS, 3D and 2D are the short forms for the weighted LSS, weighted LSWS, stochastic surface, three-dimensional and two-dimensional representations, respectively
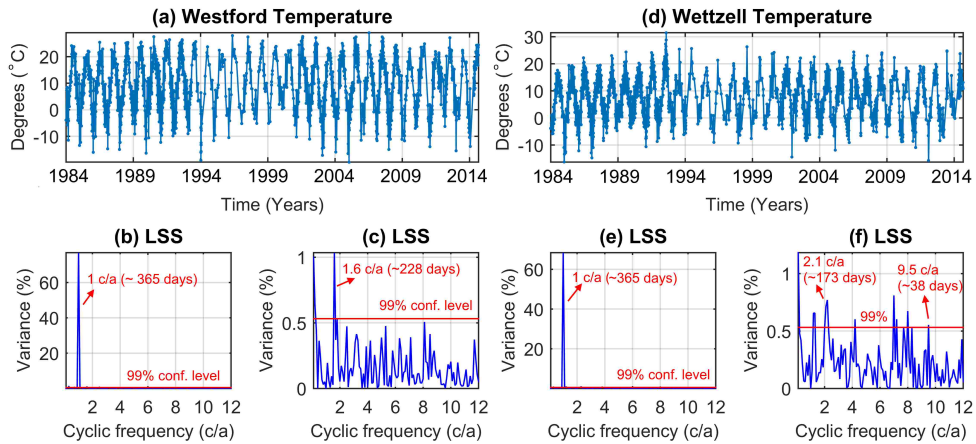
Fig. 10 The unequally spaced temperature records at the VLBI sites (a) Westford (d) Wettzell and their LSSA (b) and (c) for Westford and (e) and (f) for Wettzell

frequency variability over time, neither the LSSA nor the Fourier analysis is appropriate for the analysis because both methods decompose the time series in the frequency domain, disabling the nature of constituents in the time series.

The selection of sinusoidal functions in the LSWA allows one to directly decompose a time series into the time-frequency domain by taking into consideration the correlations among the sine and cosine functions and other constituents of known forms as well as any correlated noise. The proposed method allows one to compute spectrograms rigorously for equally and unequally spaced time series without any modification or editing. The LSWS shows how much each constituent within a window contributes to the segment of the time series within the window, enabling the determination of the nature of constituents in the time series. When the fully populated covariance matrix associated with the time series is provided, the LSWA considers the correlations between the entire time series values in each segment.

For equally spaced and equally weighted time series, when $L_1 = 0$ in Eq. (3), the LSWS has similar time and frequency resolution to the STFT, and it has similar time and frequency resolution to the CWT when $L_1 \neq 0$. In the LSWA, when $L_1 \neq 0$, the size of the translating window decreases when the frequency increases and vice versa, allowing the detection of all short- and long-duration constituents simultaneously, particularly when they are in attenuation (e.g., free oscillations of the Earth, resonant phenomena, etc.). Care must be taken when selecting L_0 and L_1. For instance, if the window location approaches to a large gap in an unequally spaced time series, then the LSWA will borrow data points corresponding to the neighboring time period when selecting a large L_0. Taking these neighboring data points into account may affect the percentage variance of the true signal peaks in the spectrogram and may not be reliable when the time series is strongly non-stationary. Therefore, the researchers need to know some basic information of the time series that are studying for the selection of the window size parameters. The stochastic surfaces defined for the spectrogram show the significant spectral peaks at a certain confidence level (usually 95% or 99%).

In future work, the cross-spectrogram between two (or more) time series will be introduced, and its underlined PDF will be derived. Moreover, the phase information about the two time series with their coherency will be discussed. This method will allow one for instance to estimate the coherency between the VLBI length series and temperature records.

References

Abd El-Gelil M, Pagiatakis S D, El-Rabbany A (2008) Frequency-dependent atmospheric pressure admittance of superconducting gravimeter records using least-squares response method. Phys Earth Planet Inter 170(1-2):24-33. doi:10.1016/j.pepi.2008.06.031

Amato U, Antoniadis A, Pensky M (2006b). Wavelet kernel penalized estimation for non-equispaced design regression. Stat Comput 16(1):3755. doi:10.1007/s11222-006-5283-4

Barnhart B L (2011) The Hilbert-Huang transform: Theory, application, development. Ph.D. Thesis, University of Iowa

Baydar N, Ball A (2001) A comparative study of acoustic and vibration signals in detection of gear failures using Wigner-Ville distribution. Mech Syst Signal Process 15(6):1091-1107. doi.10.1006/mssp.2000.1338

Brown R G, Hwang P Y C (2012) Introduction to random signals and applied Kalman filtering. 383, John Wiley and Sons, Inc, New Jersey, USA

Campbell J (2004) VLBI for geodesy and geodynamics, The role of VLBI in astrophysics, astrometry and geodesy, Chap. 22, NATO Science Series II: Math Phys Chem (135):359-381

Chen Y, Feng M Q (2003) A technique to improve the empirical mode decomposition in the Hilbert-Huang transform. Earthq Eng Eng Vib 2(1):7585. doi:10.1007/BF02857540

Classen T A C M, Mecklenbrauker W F G (1980) The Wigner distribution: A tool for time-frequency analysis (Parts I, II, III). Philips J Res 35

Cohen A, Daubechies I, Feauveau J C (1992) Biorthogonal bases of compactly supported wavelets. Comm Pure Appl Math 45(5):485-560. doi: 10.1002/cpa.3160450502

Craymer M R (1998) The least-squares spectrum, its inverse transform and autocorrelation function: Theory and some application in geodesy. Ph.D. Thesis, University of Toronto, Canada

Foster G (1996) Wavelet for period analysis of unevenly sampled time series. Astron J 112(4):1709-1729

Gabor D (1946) Theory of communication. J IEE 93(26):429-457. doi:10.1049/ji-3-2.1946.0074

Gilles J (2013) Empirical wavelet transform, IEEE Trans Sig Process 61(16):3999-4010. doi:10.1109/TSP.2013.2265222

Guang Y, Sun X, Zhang M, Li X, Liu X (2014) Study on ways to restrain End Effect of Hilbert-Huang transform. J Computers 25(3):22-31

Hall P, Turlach B A (1997) Interpolation methods for nonlinear wavelet regression with irregularly spaced design. Ann Stat 25 (5):1912-25. doi: 10.1214/aos/1069362378

Harville D A (2008) Matrix algebra from a statistician's perspective. Springer Science and Business Media, LLC

Hayashi T, Yoshida N (2005) On covariance estimation of non-synchronously observed diffusion processes. Bernoulli 11(2):359379. doi:10.3150/bj/1116340299

Hlawatsch F, Boudreaux-Bartels F (1992) Linear and quadratic time-frequency signal representations. IEEE Sig Proc Mag 9(2):21-67. doi:10.1109/79.127284

Huang N E, Wu Z (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal doi: 10.1142/S1793536909000047

Huang N E, Wu Z (2008) A review on Hilbert-Huang transform: Method and its applications to geophysical studies. Rev Geophys 46, p. 23. doi:10.1029/2007RG000228

Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C, Liu H H (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 454:903995. doi:10.1098/rspa.1998.0193

Hui Y, Pagiatakis S D (2004) Least-squares spectral analysis and its application to superconducting gravimeter data analysis. Geo Spat Inf Sci 7(4):279-283. doi:10.1007/BF02828552

Johnson N L, Kotz S, Balakrishnan N (1995) Continuous univariate distributions, 2 (2nd ed.). Wiley. ISBN 978-0-471-58494-0

Kay S M, Marple S L (1981) Spectrum analysis-A modern perspective. Proc IEEE 69(11):1380-1419. doi:10.1109/PROC.1981.12184

Lomb N R (1976) Least-squares frequency analysis of unequally spaced data. Astrophys Space Sci 39(2):447-462. doi:10.1007/BF00648343

Mallat S (1999) A Wavelet tour of signal processing. 637 Academic Press, Cambridge UK

Mathias A, Grond F, Guardans R, Seese D, Canela M, Diebner H H (2004) Algorithms for Spectral Analysis of Irregularly Sampled Time Series. J Stat Softw 11(2), p. 27. doi:10.18637/jss.v011.i02

Pagiatakis S D (1999) Stochastic significance of peaks in the least-squares spectrum. J Geodesy 73(2):67-78. doi:10.1007/s001900050220

Pagiatakis S D, Yin H, Abd El-Gelil M (2007) Least-squares self coherency analysis of superconducting gravimeter records in search for the Slichter triplet. Phys Earth Planet Inter 160:108-123. doi: 10.1016/j.pepi.2006.10.002

Psimoulis P, Pytharouli S, Karambalis D, Stiros S (2008) Potential of Global Positioning System (GPS) to measure frequencies of oscillations of engineering structures. J Sound Vib 318(3):606-623 doi: 10.1016/j.jsv.2008.04.036

Puryear C I, Portniaguine O N, Cobos C M, Castagna J P (2012) Constrained least-squares spectral analysis: Application to seismic data. Geophysics 77(5):143-167. doi:10.1190/geo2011-0210.1

Qian S (2002) Introduction to time-frequency and wavelet transforms. Prentice-Hall Inc. Upper Saddle River, NJ

Rao C R, Mitra S K (1971) Generalized inverse of matrices and its applications. John Wiley, New York pp. 240. ISBN: 0471708216, 9780471708216

Rehfeld K, Marwan N, Heitzig J, Kurths J (2011) Comparison of correlation analysis techniques for irregularly sampled time series. Nonlinear Process Geophys 18:389404. doi:10.5194/npg-18-389-2011

Ren H, Wang Y L, Huang M Y, Chang Y L, Kao H M (2014) Ensemble empirical mode decomposition parameters optimization for spectral distance measurement in hyperspectral remote sensing data. Rem Sens 6(3):2069-2083. doi:10.3390/rs6032069

Sardy S, Percival D B, Bruce A G, Gao H, Stuetzle W (1999) Wavelet shrinkage for unequally spaced data. Stat Comput 9 (1):65-75. doi:10.1023/A:1008818328241

Scargle J D (1982) Studies in astronomical time series analysis: II - Statistical aspects of spectral analysis of unevenly spaced data. Astrophys J, Part 1 (263):835-853. doi:10.1086/160554

Selesnick I V (2011) Wavelet transform with tunable Q-factor. IEEE Trans Sig Process 59:3560-3575. doi:10.1109/TSP.2011.2143711

Shapiro H S, Silverman R A (1960) Alias-free sampling of random noise. J Soc Indust Appl Math 8(2)225-248. doi:10.1137/0108013

Sinha S, Routh P S, Anno P D, Castagna J P (2005) Spectral decomposition of seismic data with continuous-wavelet transform. Geophysics 70(6):19-25. doi: 10.1190/1.2127113

Steeves R R (1981) A statistical test for significance of peaks in the least-squares spectrum. Collected Papers of Geodetic Survey, Dept. of Energy, Mines and Resources, Surveys and Mapping, Ottawa 149-166

Stein E M, Shakarchi R (2003) Fourier analysis. Princeton University Press, USA

Titov O (2002) Spectral analysis of the baseline length time series from VLBI data. IVS General Meeting Proc pp. 315-319

Vanicek P(1969) Approximate spectral analysis by least-squares fit. Astrophy Space Sci 4(4):387-391. doi:10.1007/BF00651344

Vanicek P (1971) Further development and properties of the spectral analysis by least-squares. Astrophys Space Sci 12(1):10-73. doi: 10.1007/BF00656134

Vanicek P, Krakiwsky E J (1986) Geodesy the concepts. Amsterdam The Netherlands, University of New Brunswick, Canada

Wells D E, Vanicek P, Pagiatakis S D (1985) Least-squares spectral analysis revisited. Department of Surveying Engineering, University of New Brunswick, Canada

Wu D L, Hays P B, Skinner W R (2010) A Least-squares method for spectral analysis of space-time series. J Atmos Sci 52:3501-3511. doi: 10.1175/1520-0469(1995)

Appendix A: Derivation of Eq. (10)

A similar methodology is used as in Wells et al. (1985) and Craymer (1998) to obtain Eq. (10). From the model $\mathbf{y} = \overline{\boldsymbol{\Phi}}\,\overline{\mathbf{c}} = \underline{\boldsymbol{\Phi}}\,\underline{\mathbf{c}} + \boldsymbol{\Phi}\,\mathbf{c}$, estimate $\overline{\mathbf{c}}$ using the least-squares method as $\hat{\overline{\mathbf{c}}} = \overline{\mathbf{N}}^{-1}\overline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}$, where $\overline{\mathbf{N}} = \overline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\overline{\boldsymbol{\Phi}}$. Now write $\overline{\mathbf{N}}^{-1}$ as

$$\overline{\mathbf{N}}^{-1} = \begin{bmatrix} \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}} & \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}} & \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M_1} & \mathbf{M_2} \\ \mathbf{M_3} & \mathbf{M_4} \end{bmatrix},$$

where

$$\mathbf{M_1} = \left(\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}} - \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi}\,(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}} \right)^{-1},$$

$$\mathbf{M_2} = -\mathbf{M_1}\,\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi}\,(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi})^{-1},$$

$$\mathbf{M_3} = -\mathbf{M_4}\,\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}}\,\underline{\mathbf{N}}^{-1},$$

$$\mathbf{M_4} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}}\,\underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi} \right)^{-1},$$

and $\underline{\mathbf{N}} = \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}}$. Thus, from

$$\begin{bmatrix} \mathbf{x} \\ \hat{\mathbf{c}} \end{bmatrix} = \hat{\overline{\mathbf{c}}} = \overline{\mathbf{N}}^{-1}\overline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} = \begin{bmatrix} \mathbf{M_1} & \mathbf{M_2} \\ \mathbf{M_3} & \mathbf{M_4} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \\ \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \end{bmatrix},$$

one obtains

$$\begin{aligned} \hat{\mathbf{c}} &= \mathbf{M_3}\,\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} + \mathbf{M_4}\,\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \\ &= -\mathbf{M_4}\,\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}}\,\underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} + \mathbf{M_4}\,\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \\ &= \mathbf{M_4}\,\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\left(\mathbf{y} - \underline{\boldsymbol{\Phi}}\,\underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \right) \\ &= \mathbf{N}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\hat{\mathbf{g}}. \end{aligned}$$

where $\mathbf{N} = \mathbf{M_4}^{-1} = \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\boldsymbol{\Phi}}\,\underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\boldsymbol{\Phi}$, and $\hat{\mathbf{g}} = \mathbf{y} - \underline{\boldsymbol{\Phi}}\,\underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}$ is the estimated residual segment. Note that for each frequency, $\mathbf{N}$ is a square matrix of order two, and so its inverse is computationally efficient.