Least Squares Wavelet Analysis

Ebrahim Ghaderpour (ebig2@yorku.ca) and **Spiros D. Pagiatakis** (spiros@yorku.ca)
Lassonde School of Engineering, York University, Toronto, Ontario, Canada

Introduction

In many applications, time series may not be sampled at equally spaced intervals or they may contain data gaps. Measurements have variances, so the time series may also be unequally weighted. Time series may also contain systematic noise, such as trends and/or datum shifts (offsets). In certain geodynamic applications, seismic noise may contaminate the time series of interest or certain components of the time series may exhibit variable frequency, such as linear, quadratic, exponential or hyperbolic chirps.

Many researchers attempt to modify or edit the time series to satisfy the stringent requirements of the Fourier transform. The Least Squares Spectral Analysis (LSSA) is a powerful method of analyzing non-stationary and unequally spaced time series, however, it cannot be used for time series with constituents of variable amplitude and frequency.

The Short Time Fourier Transform (STFT) and the Continuous Wavelet Transform (CWT) are useful for the analysis of equally spaced and non-stationary time series with constituents of variable amplitude and frequency. However, the STFT and the CWT are not defined for unequally spaced time series, nor do they consider non-stationarity.

We develop a new method that can analyze non-stationary and unequally spaced time series exhibiting low/high frequency and amplitude variability over time. The Least Squares Wavelet Analysis (LSWA) is an extension of the LSSA and can analyze rigorously any type of time series superseding any current time series analysis method.

Methodology

Conceptually, the LSWA is a combination of the classical wavelet analysis (with variations) and the LSSA. The LSWA attempts to fit via least squares a base function (sinusoid or other wavelet) to segments of the time series. The segmentation of the time series is achieved by a translating (sliding) window whose size is characterized by the number of data points it includes. Its size depends on the series inverse sampling interval ($\boldsymbol{M}$), the number of cycles of the base function to be fitted in the segment ($\boldsymbol{L_1}$), the frequency $\boldsymbol{\omega_k}$ of the base function (dilation) and on the desired time and frequency resolution of the final analysis. The (time) length of the window is variable when the series is unequally spaced to maintain its size (number of data points) for a specific time and frequency resolution. The size of the window must include a minimum number of data points to achieve a reasonable redundancy for the least-squares fit. We define the size of the translating window as follows:

$$L(\omega_k) = \left\lfloor \frac{L_1 M}{\omega_k} \right\rfloor + L_0, \tag{1}$$

where $\boldsymbol{L_0}$ is the additional number of data points that we consider in the segmentation of the time series to achieve the desired time and frequency resolution in the LSWA spectrogram.

We illustrate the translating windows on an equally spaced time series and two different frequencies when $\boldsymbol{L_1 = 1}$ and $\boldsymbol{L_0 = 0}$ (Fig. 1). The window size decreases when frequency increases. The mid points of the windows coincide with $\boldsymbol{t_j}$ of the time series (shown by vertical red lines).

Fig. 1: Translating windows for an equally spaced time series for two frequencies

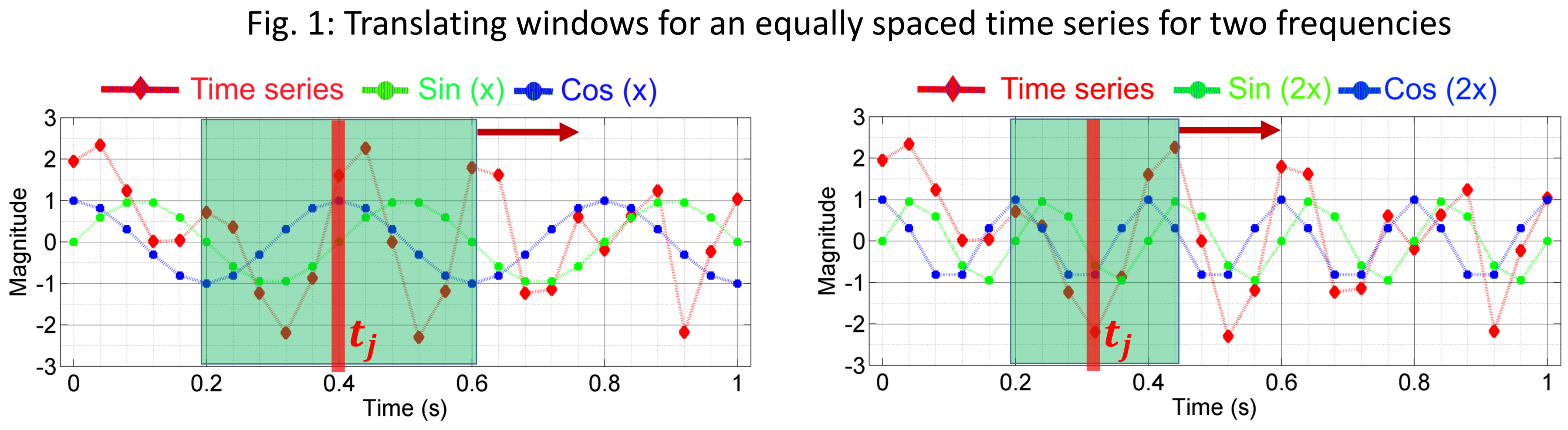
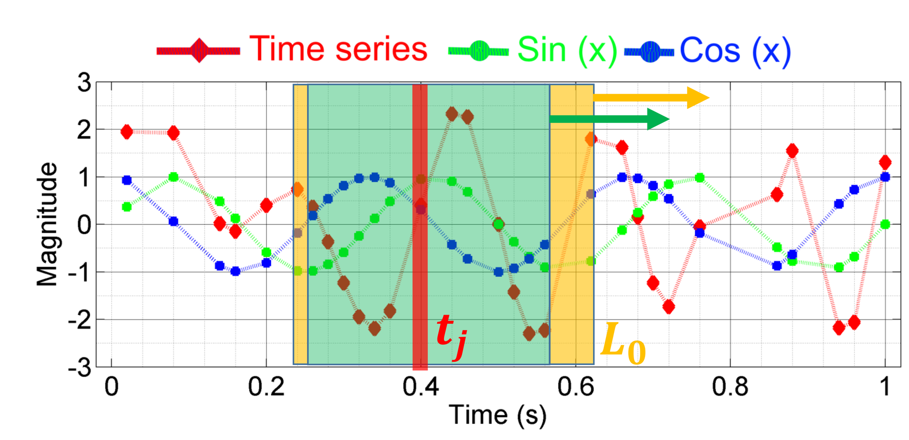


Fig. 2 shows the translating window when $\boldsymbol{L_1 = 1}$ and $\boldsymbol{L_0 = 2}$ for an unequally spaced time series. The window size does not change in translation, but the window length does. The location of the window is defined by the median point and not by the middle of the window. Notable differences between our approach and the classical wavelet analysis are in the segmentation of the time series to achieve maximum resolution in time and frequency.

Fig. 2: Translating window for an unequally spaced time series for one frequency



Mathematically, if $\boldsymbol{f}$ is the time series of $\boldsymbol{n}$ data points, $\boldsymbol{y}$ is the segment of the time series within the window of size $L(\omega_k)$ located at $\boldsymbol{t_j}$, and $\boldsymbol{\Phi} = [\cos 2\pi\omega_k t,\ \sin 2\pi\omega_k t]$ is the design matrix of dimension $L(\omega_k) \times 2$, then the LSWA spectrogram is calculated by:

$$s(t_j, \omega_k) = \frac{y^T P_y \Phi (\Phi^T P_y \Phi)^{-1} \Phi^T P_y y}{y^T P_y y} \in (0, 1), \tag{2}$$

where $\boldsymbol{P_y}$ is the principal submatrix of $\boldsymbol{P}$ (the inverse of the associated covariance matrix of f) with dimension $L(\omega_k)$. When the time series has constituents of known forms, we can consider them as additional columns in the design matrix. We may also suppress them first before the LSWA is applied.

Assume that $\boldsymbol{f}$ has been derived from a population of random variables following the multi-dimensional normal distribution. The LSWA spectrogram, Eq. (2), follows the **beta** distribution (similar to the LSSA), and a stochastic surface can be obtained above which the spectral peaks in the LSWA spectrogram are statistically significant.

Example 1:

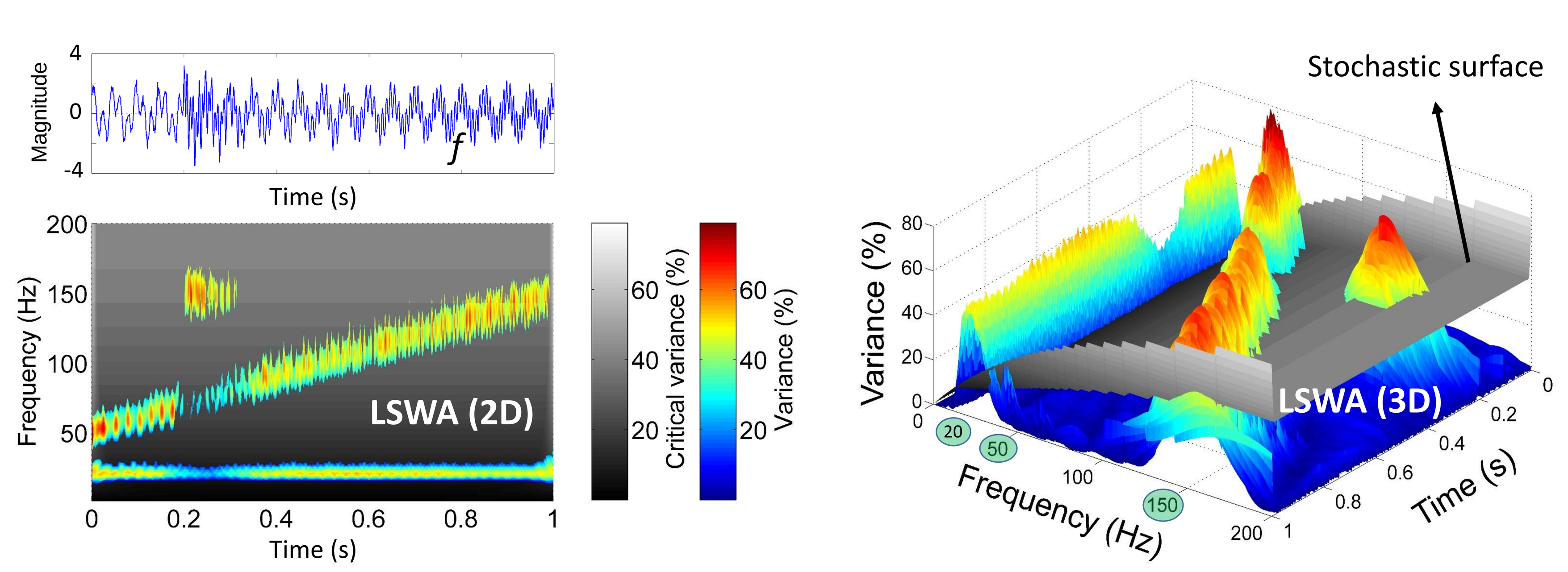
Suppose that

$$f(t) = \sin(50 \times 2\pi(t + t^2)) + \cos(20 \times 2\pi t) + g(t) + \varepsilon,$$ where ε is white noise and

$$g(t) = \begin{cases} (4 - 10t)\cos(150 \times 2\pi t) & 0.2 \le t \le 0.4 \\ 0 & O.w. \end{cases}$$

The LSWA spectrogram of f with the stochastic surface (gray) at 95% confidence level in Fig. 3 indicates that f contains constituents with variable amplitude and frequency.

Fig. 3: f and its LSWA



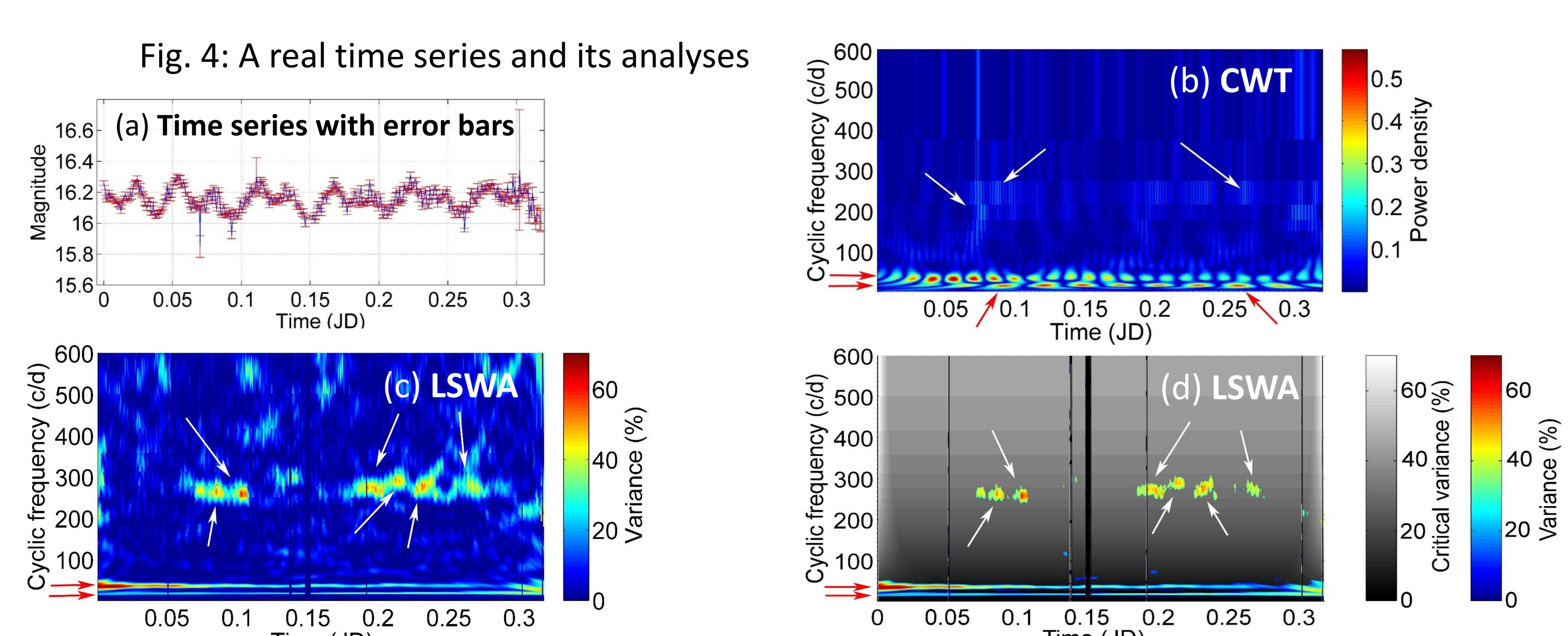
Example 2:

We analyze an unequally spaced and weighted time series (Fig. 4a) representing the magnitude of the brightness of V455 Andromeda (www.aavso.org).

The CWT (Fig. 4b) cannot account for errors in the time series nor can it considers that the time series is unequally spaced. The two constant cyclic frequencies (from physics) appear as discontinuous lines of low resolution in the CWT spectrogram (red arrows).

In the LSWA (Fig. 4c), the low frequencies are very clearly resolved (red arrows), and the presence of short duration high frequencies is clear and distinct (white arrows). The peaks shown by the white arrows and the two constant cyclic frequencies (red arrows) are statistically significant at 99% confidence level defined by the gray surface in Fig. 4d.

Fig. 4: A real time series and its analyses



Conclusions

The above examples and many other tests we performed on a large variety of synthetic time series (but not presented here) exemplify the power of the LSWA to analyze any time series in a rigorous manner and demonstrate its predominance over any spectral and classical wavelet analyses methods. The LSWA is particularly suitable for analyzing unequally spaced and strongly non-stationary and non-ergodic time series.

Bibliography

(2015) Ghaderpour E and Pagiatakis S D, Least squares wavelet analysis
(1999)Mallat S, A wavelet tour of signal processing, 637. Academic Press, Cambridge UK
(1999) Pagiatakis S D, Stochastic significance of peaks in the least-squares spectrum, J. Geodesy, 73, 67-78
(1985) Pagiatakis S D, Vanicek, P. and Wells, D. E., Least squares spectral analysis revisited, 68, University of New Brunswick, Canada
(1969) Vanicek P, Approximate spectral analysis by least squares fit, Astrophy Space Sci, 4, 387-391