Least Squares Wavelet Analysis

Ebrahim Ghaderpour and Spiros Pagiatakis

Lassonde School of Engineering, York University, Canada
Emails: ebig2@yorku.ca, spiros@yorku.ca

Introduction

In certain applications, such as electrical engineering, geodesy, and wireless networks, time series may not be sampled at equally spaced intervals or they may contain data gaps, and we refer to these time series as unequally spaced. Measurements have variances, so the time series may also be unequally weighted and come with their associated covariance matrices. Time series may also contain systematic noise, such as trends (linear or exponential or other) 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 and so on (e.g., Mallat, 1999). It is not unusual to see that some researchers attempt to modify the time series by interpolation, "editing" or removing offsets to satisfy the stringent requirements of the Fourier transform. The Least Squares Spectral Analysis (LSSA), introduced by Vanicek (1969), is a powerful method of analyzing non-stationary and unequally spaced time series (Wells et al., 1985), however, it cannot generally be used for time series with constituents that have variable amplitude and frequency because the LSSA transforms the time series to the frequency domain rather than the time-frequency domain. 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 (e.g., Mallat, 1999). However, the STFT and the CWT are not defined for unequally spaced time series, nor do they consider the first and second statistical moments of non-stationary time series. In this paper, we introduce a new method that can analyze non-stationary and unequally spaced time series exhibiting low/high frequency and amplitude variability over time. This new method, namely, 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.

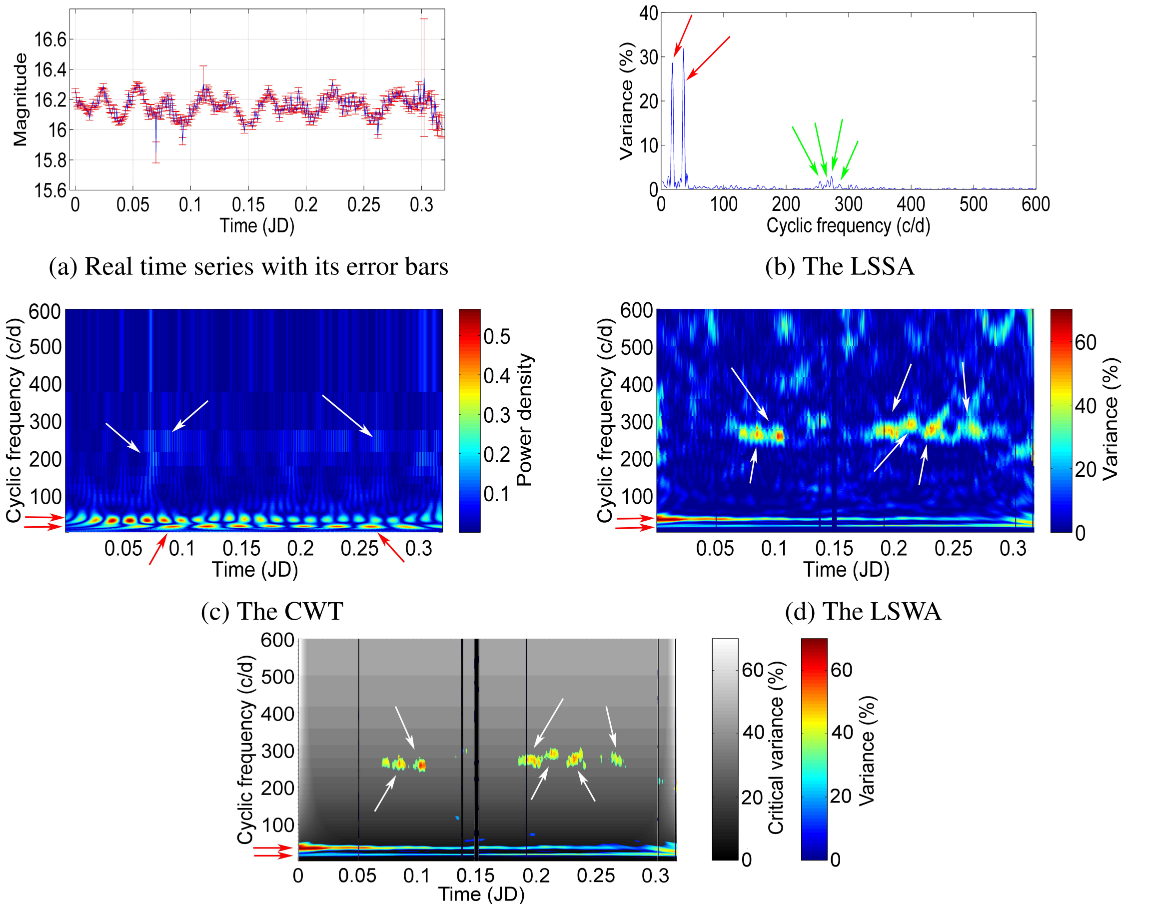
Least squares wavelet analysis

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 (usually sinusoid, but other wavelet functions are possible) to segments of the time series rather than to the whole series at once as it is done by the LSSA or similarly by Fourier analysis. The segmentation of the time series is achieved by a sliding (translating) window whose size is characterized by the number of data points it includes. Its size depends on the series inverse sampling interval (M; in units of data points per sampling interval), the number of cycles of the base function to be fitted in the segment (L_1), the frequency ω_k of the base function (dilation) and on the desired time and frequency resolution of the final analysis. Practically, the (time) length of the window is variable when the series is unequally spaced so as to maintain its size (number of data points) for a specific time and frequency resolution. Realistically, 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 sliding window as follows:

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

where 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. For instance, for an equally spaced time series recorded in milliseconds, 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. Notable differences between our approach and the classical wavelet analysis are in the segmentation of the series via the translating window to achieve maximum resolution in time and frequency (Ghaderpour and Pagiatakis, 2015a). After the segmentation of the series is defined through the window (translation, dilation), we apply the least squares method to fit of the base functions (wavelets) on the time series segments, using exactly the algorithms of LSSA (Wells et al., 1985), and the result is presented in terms of percentage variances. Clearly, this approach takes into consideration all the desirable properties of the weighted LSSA by extracting the appropriate principal submatrix of the inverse covariance matrix of the series pertaining to the segment analysed. In addition, the statistics of the LSSA spectrum can be used in each segment (Pagiatakis, 1999) producing stochastic surfaces above which spectrogram peaks are statistically significant (usually above 95% or 99% confidence level). We refer to Ghaderpour and Pagiatakis (2015b) for more details.

Example: In order to demonstrate the predominance of the LSWA over the LSSA and CWT, we analyze an astronomical, unequally spaced and unequally weighted time series representing the magnitude of the brightness of $V455$ Andromeda (www.aavso.org). Due to the obstruction of the star by sunlight, weather conditions and availability of telescope time, the magnitude of the star is measured at unequal time intervals. This time series comprises 400 data points obtained on September 10^{th}, 2013 (cf., Fig. 1a). First, we apply the LSSA (cf., Fig. 1b), and we observe two strong peaks in lower frequencies (red arrows) and several weak peaks in higher frequencies (green arrows). However, we do not know which part of the time series these peaks are coming from. Applying the CWT is not correct because the time series is inherently unequally spaced and unequally weighted. To see the advantages of the LSWA over the CWT, however, we apply the CWT with a Morlet wavelet (cf., Fig. 1c). The CWT peaks are not in the same location in the time-frequency domain because the CWT treats the time series as equally spaced (white arrows). The CWT does not consider the errors in the time series values, and the two constant cyclic frequencies appear as discontinuous lines of low resolution in the CWT spectrogram (red arrows). Now we apply



(a) Real time series with its error bars

(b) The LSSA

(c) The CWT

(d) The LSWA

(e) The LSWA with its stochastic surface (99% confidence level)

Figure 1: An unequally spaced and unequally weighted time series along with three different analyses. Abscissa values in panels (a), (c), (d) and (e) are Julian Day (JD) since 2456545.717593.

the LSWA by choosing $M = 1250$ (we choose M as the inverse of the average sampling interval), $L_0 = 10$ and $L_1 = 4$ cycles, and we suppress the datum shift and trend (cf., Fig. 1d). The low frequencies are very clearly resolved (red arrows), and the presence of short duration high frequencies is clear and distinct (white arrows). Only the spectrograms shown by the white arrows and the two constant cyclic frequencies (red arrows) will be statistically significant at 99% confidence level defined by the gray surface in Fig. 1e.

Conclusions

The above example 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, strongly non-stationary and non-ergodic time series.

References

[1] Ghaderpour, E. and Pagiatakis, S. Least squares wavelet analysis, Under review, Digital Signal Process. Elsevier (2015a)

[2] Ghaderpour, E. and Pagiatakis, S. Stochastic surfaces in the least squares wavelet analysis, Under review, Digital Signal Process. Elsevier (2015b)

[3] Mallat, S., A wavelet tour of signal processing, 637. Academic Press, Cambridge UK (1999)

[4] Pagiatakis S., Stochastic significance of peaks in the least-squares spectrum, Journal of Geodesy, 73, 67-78 (1999)

[5] Vanicek, P., Approximate spectral analysis by least squares fit, Astrophy Space Sci, 4, 387-391 (1969)

[6] Wells, D. E. and Vanicek, P. and Pagiatakis, S. D., Least squares spectral analysis revisited, 68, Tech Rep 84, University of New Brunswick, Canada (1985)