Biological Rhythm Research
2001, Vol. 32, No. 3, pp. 341–345

0929-1016/3203-341$16.00

Emphasizing Difficulties in the Detection of Rhythms with Lomb-Scargle Periodograms

Martin Schimmel

Department of Geophysics, University of São Paulo, Brazil.

Abstract

The Lomb-Scargle periodogram was introduced in astrophysics to detect sinusoidal signals in noisy unevenly sampled time series. It proved to be a powerful tool in time series analysis and has recently been adapted in biomedical sciences. Its use is motivated by handling non-uniform data which is a common characteristic due to the restricted and irregular observations of, for instance, free-living animals. However, the observational data often contain fractions of non-Gaussian noise or may consist of periodic signals with non-sinusoidal shapes. These properties can make more difficult the interpretation of Lomb-Scargle periodograms and can lead to misleading estimates. In this letter we illustrate these difficulties for noise-free bimodal rhythms and sinusoidal signals with outliers. The examples are aimed to emphasize limitations and to complement the recent discussion on Lomb-Scargle periodograms.

Keywords: Lomb-Scargle periodogram, rhythm detection, outlying measures, non-sinusoidal signals.

Introduction

The Lomb-Scargle periodogram analysis (Lomb 1976; Scargle, 1982) was introduced in astronomy for non-uniform data and is based on the least-square fitting of sine waves to the data. The obtained least-square spectrum provides the measurement of power as function of frequency which explains best the overall variance of the data. A detailed description of the method and its performance on biomedical data can be found by Ruf (1999) and Van Dongen et al. (1999). An important aspect of the measure is that if the data consists of pure Gaussian noise then the power follows an exponential distribution which is used to test the significance of detected events against random noise (Hernandez, 1999; Ruf, 1999; Van Dongen et al., 1999; among others).

Address correspondence to: M. Schimmel, Dept. of Geophysics, IAG-USP, Rua do Matão 1226,05508-900 São Paulo, SP, Brazil. E-mail: martin@iag.usp.br

Due to the least-square fitting of sine waves the Lomb-Scargle method effectively detects sinusoidal rhythms in time series. The underlying assumption of the least-square optimization is that the noise in the data is normal distributed. Deviations from normal distributed noise or sinusoidal signal waveforms are common and can difficult or even inhibit the detection of rhythms with the Lomb-Scargle method. In this letter this is illustrated in two theoretical examples: a bimodal rhythm and a sinusoidal signal with two discordant data points. The examples are hold simple for a better understanding of the principles which can apply in more complicated data. This letter is aimed to complement the recent discussion about the Lomb-Scargle method and to emphasize limitations of this method and the importance of additional analyses.

Non-Sinusoidal Rhythm: Bimodal Signal

Spectral methods such as the Lomb-Scargle analysis effectively decompose time series in some way into sine waves to estimate the power at each frequency. Equidistant data can be decomposed into a set of equally spaced frequency components which ensure that the powers in the spectrum are independent due to orthogonality properties (e.g., Hernandez, 1999; Van Dongen et al., 1999). Although, for non-equally spaced data the concept of orthogonal sine functions at independent frequencies is not valid anymore, the idea of decomposition is approximately fulfilled under the assumption of small spectral leakage. Inherent to the decomposition with sine waves the power spectra of non-sinusoidal rhythms are generally more difficult to interpret.

A simple example with synthetic data is shown in Figure 1. The time series used are 7 days long and evenly sampled at 8-min intervals with 1261 samples. The solid trace (Figure 1a) contains a noise free bimodal 24-h rhythm. The corresponding Lomb-Scargle periodogram is presented with the solid line in Figure 1c. The independent frequencies (equation 10, Van Dongen et al., 1999) are marked by crosses. It can be seen from the periodogram that the largest two powers are obtained for sine waves with periods of 12 h and 8 h, respectively. The dashed trace in the Figure 1a is the sum of the two dominant sine waves. It shows that the decomposition in the two sine waves with periods of 8 h and 12 h already approximates the main characteristics of the bimodal pattern with 24-h periodicity. Summing all sine waves yields the solid line. Only the harmonics in the power spectrum indicate the non-sinusoidal rhythm of 24 hours. However, in more complicated situations, for example in the presence of noise or other signals, such periodicity might be completely hidden in the spectrum. This is exemplified with the time series of Figure 1b which contains the same bimodal rhythm but contaminated with normally distributed noise. Its periodogram is plotted with dashed line style in Figure 1c. It can be seen that the 24-h periodicity is now well hidden in the power spectrum.

Non-Gaussian Noise: Isolated Outliers

It is common that observational data contain a small fraction of strange measurements or isolated outliers such as spikes caused by instrument failure or sporadical distur-

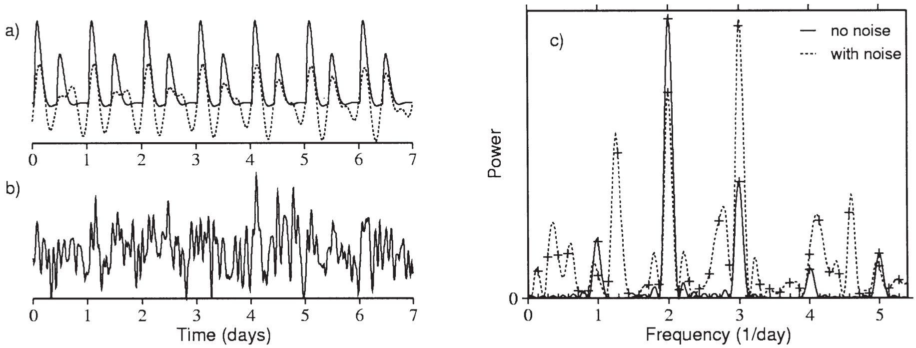

Figure 1. (a) The solid trace contains a bimodal 24-h rhythm. The superimposed dashed time series show the sum of two sine waves with periods of 8 h and 12 h which correspond to the two strongest frequency components of a Fourier decomposition; (b) The time series is the solid trace from (a) but contaminated with normal distributed noise; (c) Lomb-Scargle periodograms for the noise free (a) and noisy trace (b). The crosses mark the independent frequencies.

bances in the measuring environment. These strange isolated measurements usually lead to Gaussian shaped distributions around the center but heavier tails, i.e., the outlying data points become more important. This increased importance is illustrated in the following example. Figures 2a–c show the time series used. For simplicity they contain only one single sine wave of 24-h periodicity and two discordant data points. The time series are 7 days long, evenly sampled at 1 h, 2 h, and 3 h 30 min intervals and contain 169, 85, and 49 samples. Their Lomb-Scargle periodograms are shown in Figure 2d. The independent frequencies are marked by crosses. The power spectra of the traces with n = 85 and n = 49 samples are multiplied by 3 and 6, respectively, to increase their visibility. The false alarm probabilities (p) of the highest power at an independent frequency (equation 2, Ruf, 1999; equation 20, Van Dongen, 1999) are 1.e-12, 0.01, and 0.97. 1-p is the probability that the data contain a signal with corresponding power. However, these values assume normal noise distribution and are therefore not valid (especially for low n) in our example.

From Figure 2d (and the false alarm probabilities) it is obvious that the two spikes have an increasing influence in time series with decreasing number of samples. Although the data contain an evident sine wave no clear 24-h rhythm is detected in the Lomb-Scargle periodogram for the third trace. This is because the squared amplitudes of the spikes become dominant in the least-square fitting of sine waves. The so-called misfit function is minimised by trying to fit the large amplitude outliers first. As consequence, the highest power is found at frequencies which correspond to the 28-h separation of the discordant data points and at their higher harmonics. If one increases the amplitudes of the spikes then also the other power spectra will hide the 24-h periodicity.

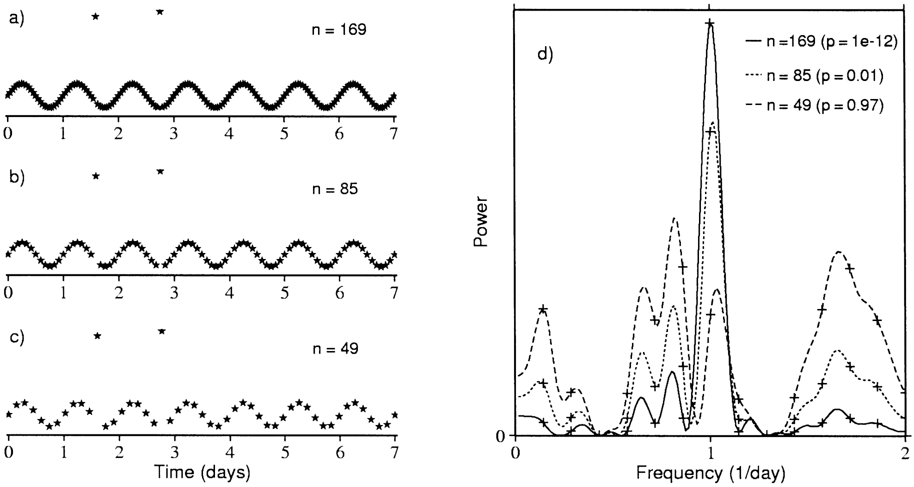

Figure 2. (a–c) The time series show an equally sampled sine wave with 24-h period and two outlying samples. The time series differ in sampling interval and number n of samples. (d) The corresponding Lomb-Scargle periodograms are presented with different line styles. The crosses mark the independent frequencies.

Discussion and Conclusions

The examples illustrate that non-sinusoidal signals or outlying large amplitude features can make more difficult the interpretation of Lomb-Scargle periodograms and eventually lead to misleading conclusions. The fact that the frequencies in Figure 1c (solid line) are harmonics of each other might help the interpretation but is generally obscured by the presence of noise or other signals. Whether bimodal signals can be detected depend on their shape and power spectrum. Alternative approaches to detect especially non-sinusoidal rhythms can be based on signal coherence which commonly adapt data summation (stacking) or cross-correlation techniques. Besides the periodicity one can obtain the average waveform and occurrence time of each signal (e.g., Hoenen et al., 2001) which might be important to understand whether events are triggered. The use of alternative methods which are not based on the resemblance of sine waves has also been stressed by Van Someren et al. (1999).

The second example further shows that the Lomb-Scargle method can be sensitive to discordant data points such as isolated large amplitude outliers caused by instrument failures or disturbances in the measure environment. The underlying statistic for the significance test is not robust in the sense that it remains insensitive to outliers. False alarm probabilities can take wrong values. For more complicated and realistic data these phenomena might go undetected and one should be aware of possible biases in the results.

The Lomb-Scargle method remains a powerful tool, but it must be emphasized to use additional analyses whenever difficulties are suspected due to violations which often are difficult to control. It might be useful to perform a parallel analysis on a theoretical model which, as closely as possible, approximates the experiment. Variations of unknowns within the expected bounds can indicate the sensitivity of the expected analysis due to outliers, employed sample distribution, and/or other factors.

Acknowledgements

The author is grateful to Mirian D. Marques for critical reading of the manuscript and financial support by the FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) grant 97/04780-6.

References

Hernandez G (1999): Time series, periodograms, and significance. *J Geophys Res 104*: 10355–10368.

Hoenen S, Schimmel M, Marques MD (2001): Rescuing rhythms from noise: a new method of analysis. *Biol Rhythm Res*: in press.

Lomb NR (1976): Least-squares frequency analysis of unequally spaced data. *Astrophys Space Sci 39*: 447–462.

Ruf T (1999): The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time series. *Biol Rhythm Res 30*: 178–201.

Scargle JD (1982): Studies in astronomical time series analysis. II. Statistical aspects of unevenly spaced data. *Astrophys J 302*: 757–763.

Van Dongen HPA, Olofsen E, Van Hartevelt JH, Kruyt EW (1999): A procedure of multiple period searching in unequally spaced time-series with the Lomb-Scargle method. *Biol Rhythm Res 30*: 149–177.

Van Someren EJW, Swaab DF, Colenda CC, Cohen W, McCall WV, Rosenquist PB (1999): Bright light therapy: improved sensitivity to its effects on rest-activity in Alzheimer patients by application of nonparametric methods. *Chronobiol Int 16*: 505–518.