Biological Rhythm Research
2001, Vol. 32, No. 3, pp. 355–360
0929-1016/3203-355$16.00

The Issue of Significant Features in Random Noise

Martin Schimmel

Department of Geophysics, University of São Paulo, Brazil

Abstract

With respect to the first example in Schimmel (2001), Van Dongen et al. (2001) conclude from their Lomb-Scargle analysis that the noise I used 'contains new periodicities that are added to the signal (these periodicities by themselves resemble a harmonic series of a 38-hour rhythm).' They infer that 'the variance of the added noise is about five times as large as the variance of the signal' causing the detection of the new significant periodicities in the noise prior to the 24-h bimodal rhythm. Moreover the 'example reflects a combination of an extremely non-sinusoidal signal with noise that is not independent, which results in a time series that is difficult to analyze with virtually any known method.' In the following, I briefly examine these concerns to avoid misunderstandings and to alert that with an inadequate use of the statistical significance test, misleading conclusions can be obtained. Although this paper further emphasizes difficulties in the detection with Lomb-Scargle periodograms, this should not be used as de-motivation. As stated in Schimmel (2001) Lomb-Scargle is a powerful technique but such as any other method one should be aware about its limitations, and use additional tools to constrain the true data characteristics.

Keywords: Lomb-Scargle periodogram, rhythm detection, noise.

Result with an Alternative Method

The noisy data from the first example in Schimmel (2001) have been analyzed with the Phase Weighted Stack (PWS) technique (Schimmel & Paulssen, 1997; Hoenen et al., 2001). With PWS signals are detected by their waveform coherence and their repeated occurrence in time series. In contrast to the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982; Ruf, 1999; Van Dongen et al., 1999), PWS is not a spectral decomposition method. Depending on the data it can be used as an additional or alternative technique to detect rhythms concealed in noise. The lower panel of Figure 1 shows the result for the noisy data from Figure 1b in Schimmel (2001). Contoured are the PWS amplitudes as function of time of the first day of data and rhythm period. For a better visibility, the contouring starts at amplitudes which equal 50% of the

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

largest PWS amplitude. The two outstanding signals at about 1440 min (= 24 h) period and 120–240 min time, and 1440 min period and 720 min time, respectively, are due to the bimodal rhythm. The first day of the noise free bimodal signal is shown in the upper panel of Figure 1 for comparison. The contour plot of Figure 1 shows no other concurrent signals. That is the features detected by Van Dongen et al. (2001) (see their Table 1) are not proven as signals by the PWS analysis. In the Lomb-Scargle analysis the bimodal rhythm is hidden in the noise inherent to the least squares fitting of sine waves which is strongly influenced by the signal and noise amplitudes. PWS is also sensitive to amplitudes, but it is not a variance estimation method and employs an amplitude unbiased coherence measure to weight the results. Further it is not based on the decomposition in sine waves or other functions. Periods larger than 30 h have not been analyzed since the number of cycles within the time series should be at least in the order of 5 to obtain significant results.

Is the Noise Used Too Large?

The variance of the time series with signal & noise is about five times larger than the variance of the data with the noise free signal. This might give the impression that the noise amplitudes are about 5 times larger than the signal amplitudes. The noise free signal (see top panel in Figure 1) has a very small variance since its amplitude is fairly constant most of the time. As consequence, the variance of the noise is larger than the variance of the signal, however, the maximum amplitudes are of the same order. Indeed, the standard deviation of the noise is less than 60% of the maximum bimodal rhythm amplitude and therefore not unrealistic.

Detection of Similar Rhythms in the Noise and Signal & Noise Data

Van Dongen et al. (2001) list in their Table 1 significant rhythms which have been detected in the noise and signal & noise data. Common features in the noise and signal & noise data are expected indeed. This is best understood when making the transition from the Lomb-Scargle power spectrum to the Fourier spectrum. For equally spaced data, such as used in the example 1, the Lomb-Scargle periodogram provides the same spectral information as the Fourier transform. The Fourier amplitude spectrum is proportional to the square root of the power spectral density of the periodogram. A constant factor accounts, among others, for the normalization with the total data variance. The Fourier transform is a linear system and as consequence the spectrum of the signal & noise series is the same as the sum of the signal spectrum and the noise spectrum. The summation is complex employing at each frequency amplitudes and phases. Figure 2 shows the amplitude spectra of the noise, signal, and signal & noise. The symbols in Figure 2 mark rhythms from Table 1 in Van Dongen et al. (2001). It is obvious that the noise amplitude spectrum resembles the signal & noise amplitude spectrum at frequencies where signal amplitudes decreased to very small values. At these frequencies the signal has only a small influence in the summation of signal and noise.

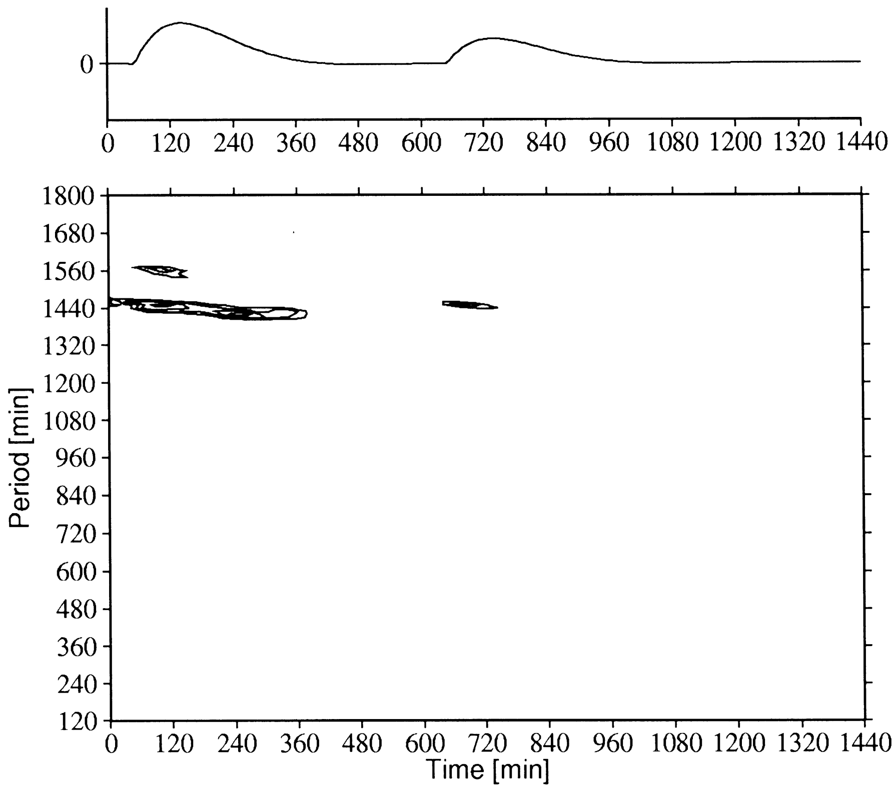

Figure 1. The first day of the noise free bimodal rhythm used in the first example is shown in the upper panel. The contour plot below is the result of a PWS analysis using the random noise contaminated bimodal signal. The data is shown in Figure 1b in Schimmel (2001).

Are the Detected Features in the Noise Really Significant?

The noise used in my example comes from a tested random sequence (Gaussian white noise) generator. Its white amplitude spectrum is shown in Figure 3a. White means that the amplitude spectrum is flat. Random constructive interferences in the Fourier summation cause the small individual peaks in the white spectra. They can not be attributed to harmonic signals. Before adding the noise to the signal the very high frequency components were suppressed using a zero phase low-pass filter. Figure 3b shows the amplitude spectrum of the alternated series. This modification did not change the noise at periods larger than about 2 hours, that is, the peaks in the amplitude spectra of Figures 2 and 3 can not be considered to be harmonic significant signals. The Lomb-Scargle periodograms for the noise of Figures 3a and 3b have similar shapes at periods larger than 2 h, however, the absolute power differs. This is since the total data variance which is used in the normalization of the Lomb-Scargle periodograms changed due to the suppression of the very high frequency components.

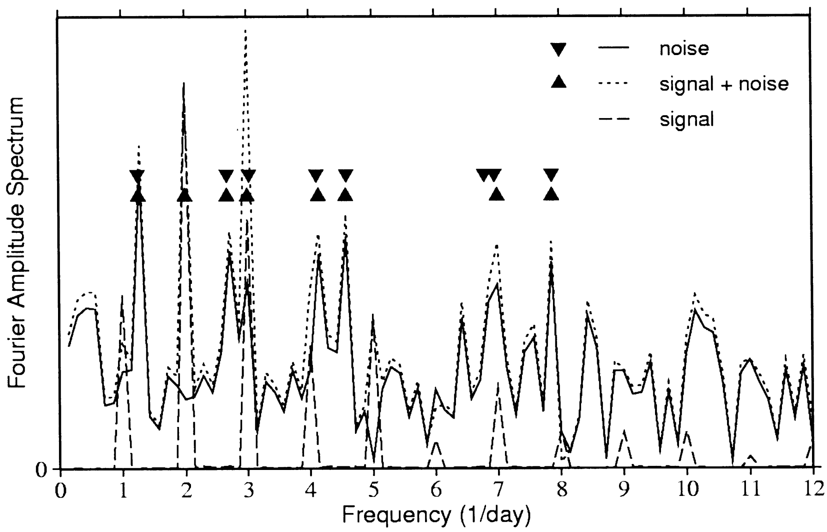

Figure 2. Fourier amplitude spectra of the noise, signal, and signal & noise data of example 1. The amplitude scale is linear. The symbols indicate the features detected by Van Dongen et al. (2001) in the noise and signal & noise data, respectively.

The statistics used by Van Dongen et al. (2001) are invalid for this data and misled them in their conclusions. A glance at the noise or signal & noise spectrum should have revealed that the noise is not white up to the Nyquist frequency. In general, filtering, de-trending or tapering will alter the relative Lomb-Scargle power and can change the number of degrees of freedom in the data. This must be considered in the significance tests. The fact that the noise is not white at all frequencies can happen in natural processes, thus is not unrealistic. The data characteristics need to be investigated before concluding from statistical significance tests.

Complementing Remarks about Outliers

Measurements which lead to a Gaussian distribution with heavier tails contain outliers. These can be due to point defects or local instabilities (nonstationarities). In my example 2 they are unambiguously identified since there is no other noise or signal. In the Gaussian model their occurrence probability is so small that least square (or maximum likelihood) estimates are distorted to account for the heavy tails of the nearly Gaussian shaped distribution. As consequence, conventional least-square based techniques can give inefficient or seriously misleading estimates. This motivated to the different techniques where small deviations from the underlying distributional assumptions are less sensitive in the optimization. The so-called *robust model estimation* or *robust statistics* deal with such methods (e.g., Huber, 1981; Chave et al., 1987). Such techniques should be employed when the conventional least-square

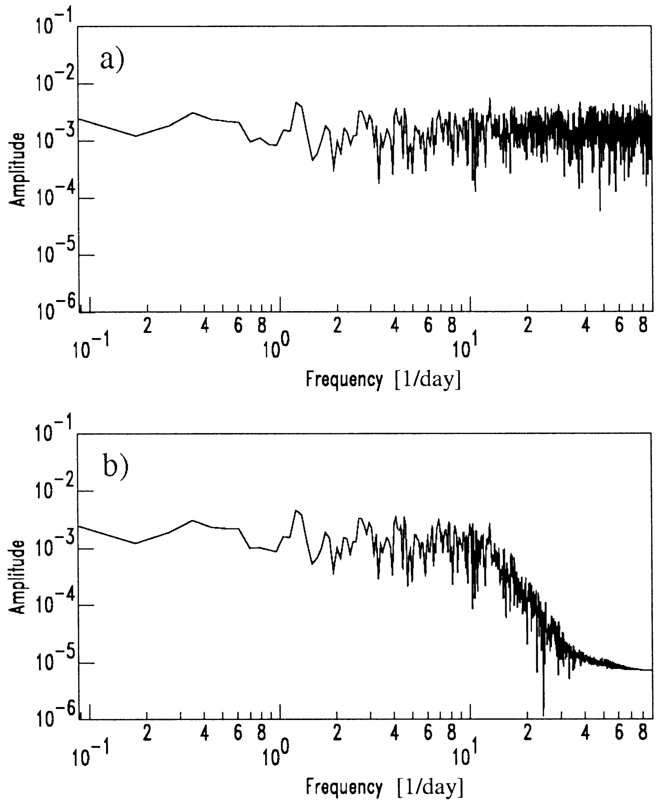

Figure 3. (a) Amplitude spectrum of the time series obtained with a tested random sequence (white Gaussian noise) generator; (b) Amplitude spectrum of the noise used in example 1. The high frequent noise components have been suppressed. Note that the spectrum is the same as in Figure 2 but the scaling is logarithmic and the Nyquist frequency is the largest frequency.

estimates are expected to break down. Minimizing the mean absolute deviation rather than the mean square deviation in the Lomb-Scargle approach is already a step forward to a robust method. The examples and discussions presented here are aimed to point to the limitations of the Lomb-Scargle approach. This should not be understood as de-motivation. The Lomb-Scargle periodogram is a powerful tool but as with any other method one should be aware about problems which can occur when inadequately used.

Acknowledgements

The author appreciates the critical reading by Mirian Marques and financial support by the FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) grant 97/04780-6.

References

Chave AD, Thomson DJ, Ander ME (1987): On the robust estimation of power spectra, coherences, and transfer functions. *J Geophy Res 92*: 633–648.

Hoenen S, Schimmel M, Marques MD (2001): Rescuing rhythms from noise: A new method of analysis. *Biol Rhythm Res 32(2)*: 271–284.

Huber PJ (1981): *Robust statistics*. New York, Wiley.

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.

Schimmel M (2001): Emphasizing difficulties in the detection of rhythms with Lomb-Scargle periodograms. *Biol Rhythm Res 32(3)*: 341–348.

Schimmel M, Paulssen H (1997): Noise reduction and detection of weak, coherent signals through phase weighted stacks. *Geophys J Int 130*: 497–505.

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 Dongen HPA, Ruf T, Olofsen E, Van Hartevelt JH, Kruyt EW (2001): Analysis of problematic time series with the Lomb-Scargle method: A reply to 'Emphasizing difficulties in the detection of rhythms with Lomb–Scargle periodograms'. *Biol Rhythm Res 32(3)*: 347–354.