Biological Rhythm Research
2001, Vol. 32, No. 2, pp. 271–284

0165-0424/01/3202-271$16.00

Rescuing Rhythms from Noise: A New Method of Analysis

S. Hoenen[1], M. Schimmel[2] and M.D. Marques[1]

[1]Museu de Zoologia, University of São Paulo, Brasil; [2]Dept de Geofísica, Instituto Astronômico e Geofísico, University of São Paulo, Brasil

Abstract

The analysis of a temporal series usually begins with a visual inspection of the raw data, from which a proper method for the detection of periodicities is chosen. Some of the methods currently used, as circular statistics, Cosinor, or spectral analyses, are useful when it comes to ascertain the existence of some periods, expected ‘a priori’ or to detect unknown frequencies. Even though some of the methods allow a wide scanning of possibilities, difficulties arise when signals are weak and concealed in larger amplitude noise.

The register of the activity of a cave cricket, *Strinatia brevipennis*, under constant conditions, showed an intricate pattern of small peaks, interspersed with rare ones of much higher amplitudes. Attempts to analyse these data with the usual methods gave inconsistent results and sometimes did not detect rhythms. The results are mostly biased by the large amplitude components which hamper the detection of rhythms from weak signals. Schimmel and Paulssen (1997) proposed a noise-reduction method, which detects weak but coherent signals. This new tool was developed for the analysis of seismic data, being afterwards adapted to the analysis of temporal series of biological data. The method is called phase weighted stack (PWS) and performs a weighted summation of temporal series according to their coherence. The results are stacked time series which are cleaned up from incoherent noise, allowing the detection of weak signals that otherwise would be undistinguishable from noises. The method also enables the identification of the time (hour) of every periodic signal. The use of PWS in the analysis of crickets’ activity data cleared out frequencies, exposing a circadian component in all records.

Keywords: Rhythm detection, data processing, noise reduction, periodicities, cave crickets.

Address correspondence to: S. Hoenen, Museu de Zoologia, USP, C.P. 42694, 04299-970, São Paulo, SP, Brasil. E-mail: smhoenen@usp.br

Introduction

The analysis of a temporal series should begin with a visual inspection of the data, from which one must choose the proper method(s) for the detection of periodicities. A usual representation of activity rhythms, for instance, is the so-called actogram. It is a section of temporal series which displays periodicities by the systematic appearance of signals over the time series employed.

When there is a reason to assume some periodicity 'a priori', there are a few techniques of analysis, among them the Cosinor test method and circular statistics. However, these methods could be applied as well to detect the period of a series, by searching for the best adjustment which will be the frequency one is looking for. However, if the period is unknown and could not be guessed, then a spectral analysis technique, such as the Fourier analysis or the Lomb-Scargle method, should be applied. There are several mathematical approaches to analyse time-series, and many are variants of the Fourier analysis. Three of them are briefly presented here:

(a) Cosinor (Nelson et al., 1979):
This method consists in adjusting a cosine function to a temporal series of data. With this technique, the following parameters are obtained: acrophase (φ), which represents the moment of maximum value of the variable; amplitude (A), which indicates the maximum variation of the values in relation with a mean; mesor (M), which is the mean value of the cosine function; and the p value, which indicates the statistical significance of the fitted curve.

(b) Circular statistics (Batschelet, 1981; Zar, 1996):
This method analyses the data distributed in a circle. Each result represents an angle in the circumference; the mean angle is obtained through a vectorial sum. Similar to Cosinor, the statistical significance of the mean vector should be tested, which will indicate the presence of a rhythm. This evaluation could be made by the Rayleigh test.

(c) Spectral analysis (Percival & Walden, 1993; Benedito-Silva, 1997):
Spectral analysis makes use of the representation of a time series in some fashion as a sum of sines and cosines of different frequencies and amplitudes. The totality of these terms is called the spectrum. It shows the contribution of each frequency component to the total variability of the time series. Sinusoidal rhythms are detected when they have outstanding signal in the amplitude or power spectrum. For uniformly sampled data, the Fourier transform is used. For nonuniformly sampled data, however, the concept of independent frequencies is not valid anymore. The Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982; Ruf, 1999; Van Dongen et al., 1999) is one of the methods developed to overcome the problem of missing data. It removes bias that may arise from forcing unevenly spaced data on grids and interpolation of missing values.

The register of the activity of a cave cricket, *Strinatia brevipennis*, showed an intricate pattern of small peaks, interspersed with rare ones of much higher amplitudes (Hoenen, unpublished data). Attempts to analyse these data with the usual methods gave inconsistent results and sometimes did not detect rhythms. The results are mostly biased by the large amplitude components which hamper the detection of rhythms from weak signals. As non-sinusoidal rhythms might be present in the data and most methods, such as the spectral ones, are more adequate to detect sinusoidal patterns, another method to analyse such data had to be looked for.

The purpose of this paper is to introduce the use of the PWS method to analyse chronobiological series of data, presenting the results of analysis of real data with different methods for comparison. The use of PWS was motivated by the ability to detect non-sinusoidal signals by their wave form coherence and its larger robustness to outlying noise.

Materials and Methods

The activity rhythm of cave crickets was automatically recorded through a system that detects the vibrations produced when the cricket moves. The equipment is described elsewhere (Hoenen & Gnaspini, 1999). The example presented here is from a male cricket, tested in constant conditions of temperature and humidity and under a light/dark cycle (LD 11:13 h), in which dark means red light. Food was available *ad libitum*. This cricket was previously maintained under DD and transfered to LD in the first day of the series.

The results were analysed using the COSINOR test method (modified after Nelson et al., 1979 by Benedito-Silva, 1997), circular statistics (Batschelet, 1981; Zar, 1996), Lomb-Scargle periodogram (Ruf, 1999; Van Dongen et al., 1999) and the PWS method. The complete description of the latter is presented in Schimmel and Paulssen (1997). We present here a brief outline and the adaptation for the analysis of the data of cave crickets.

Determination of rhythms with PWS

PWS stands for phase-weighted stacks and represents a tool for incoherent noise reduction through the weighted summation of time series. The weight is an amplitude-unbiased coherence measure called ‘phase stack’. In subsequent sections we first explain with an hypothetical example how the summations of temporal series can be used to determine the periodical appearance of signals. Then we briefly summarize the principles of the PWS method and show its performance and applicability with further examples.

Method

Figure 1 schematically sketches the detection of rhythms by time shifting and stacking. The record used (Fig. 1a) is 7 days long and contains two periodical events with 24- and 26-h periodicity. This temporal series is systematically shifted to create the data section from Figure 1b. The solid lines mark the actogram which contains the data for each day in a column. Signals with a 24-h periodicity are aligned vertically and a summation of the series will enhance these signals. All other signals sum up

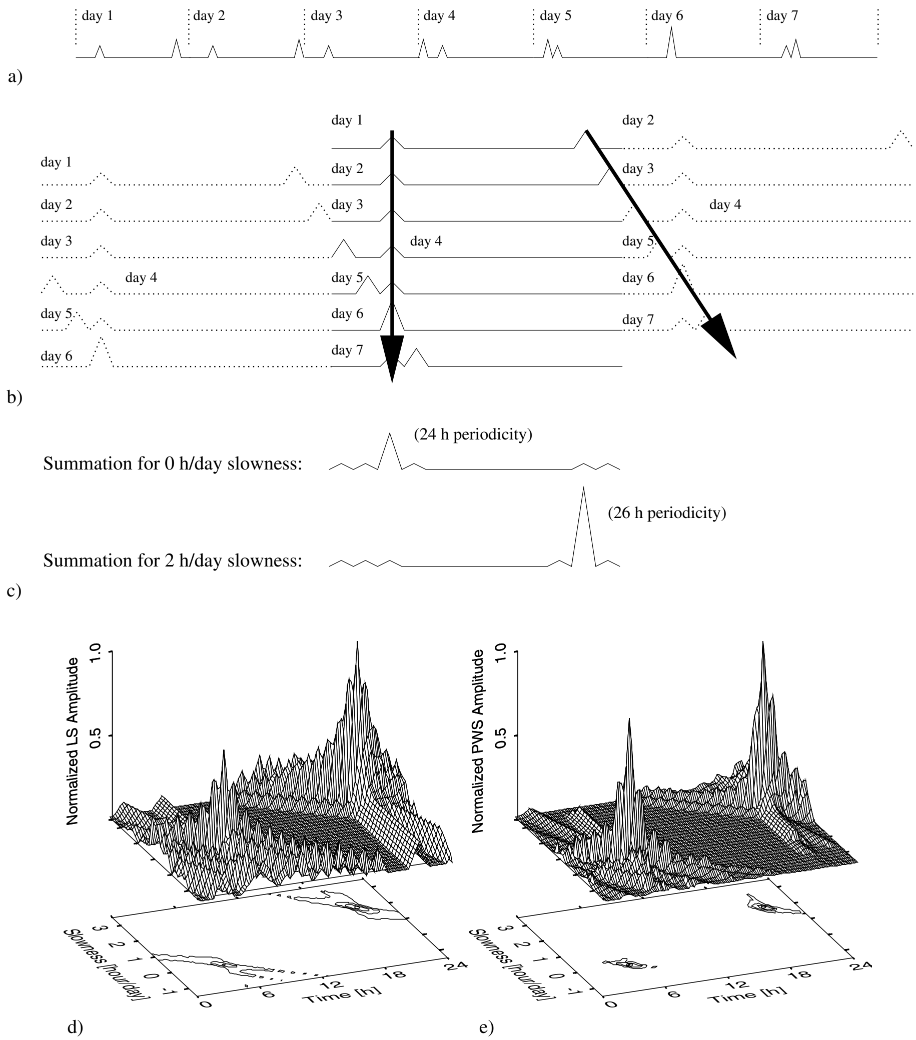

Figure 1. Synthetic example to sketch the detection of rhythms by time shifting and summation. The temporal series (a) with two periodical events is used to create the data section or actogram (solid lines) in (b). The arrows mark the periodical signals. (c) Stacks for the rhythms indicated by the arrows in figure (b). (d) shows the 3-D vespagram and the 2-D contour plot for the linear stack as function of slowness. Contouring starts at 0.1 with 0.2 intervals. (e) Same as (d) but for PWS.

less efficiently. Similarly, one can shift and sum the records for other assumed periodicities to detect other rhythms. For instance, shifting the time series by −2 hours per day will align the signals which are indicated by the second arrow (Fig. 1b). The results for 0 h/day (24-h rhythm) and −2 h/day (26-h rhythm) are shown in Figure 1c.

The stacks are performed with respect to the first day. A repeated and systematical performance of the described procedure can be used to search for the periodical appearance of signals. In seismology this process corresponds to a *slant stack*. Equation (1) represents the described stacking

$$f(t,p) = \sum_{j=1}^{N} s_j(t - \Delta d_j p) \tag{1}$$

Index j enumerates the N time series s(t) used. Δd_j is the corresponding increment of days with respect to a reference record and day. p is called *slowness* and used to determine the time shifts of each record. The stack f(t,p) can be visualized in form of a 3-D vespagram and/or 2-D contour plot as function of time t and slowness p. Figure 1d shows both for the discussed example. LS stands for linear stack and the amplitudes have been normalized to one. The two maxima can be used to determine time and slowness of the periodical signals.

Difficulties in suppressing noise employing linear stacks might occur when the signals are weak and when the amount of data is small. This motivated to non-linear stacking techniques such as PWS. The back bone of PWS is an amplitude unbiased coherence measure designed by Schimmel and Paulssen (1997). This coherence measure is called phase stack (PS) and is used to weight the linear stack. The incoherent signals or noise are suppressed by a low weight, i.e., the weight cleans the linear stack from the signals which summed up incoherently. Figure 1e shows the ability of PWS to increase the signal-to-noise ratio by suppressing the signals which are summed at, say, the wrong slowness.

The weight (PS) is based on complex trace analysis (e.g., Tanner et al., 1979). Therein the real time series s(t) is uniquely transformed to a complex-valued time series S(t) by employing the Hilbert transform H[s(t)].

$$S(t) = s(t) + iH[s(t)] = A(t)e^{i\Phi_j(t)} \tag{2}$$

S(t) is called analytic or complex trace (Bracewell, 1965) and can be expressed by the amplitude, A(t), and phase functions, $\Phi(t)$. The amplitude function is the envelope of s(t). In the complex plain each sample of the analytic trace can be visualized by a vector with length A(t) and angle $\phi(t)$ with the real axis. As time t progresses, the vector rotates in the complex plain around the time axis. The projection of this helix onto the surface spanned by the real axis and the time axis is the real temporal series s(t).

The phase stack $c_{ps}(t)$ is the absolute value of the complex summation of the amplitude normalized analytic traces.

$$c_{ps}(t) = \left| \sum_{j=1}^{N} \frac{1}{N} e^{i\Phi_j(t)} \right| \tag{3}$$

In consequence, $c_{ps}(t)$ does not depend explicitly on amplitudes and is therefore an amplitude unbiased measure. Because the complex summation $c_{ps}(t)$ employs the principles of constructive and destructive interference, coherent signals have the same phasor independently of their amplitudes and consequently sum up completely constructively. In the complex plain, this is visualized by the summation of vectors which all point to the same direction. There is still constructive interference if the phases vary slightly. Strong varying phases, however, cause a destructive summation. Owing to the employed normalization by N $c_{ps}(t)$ ranges between 0 and 1 as function of time. Amplitude 1 is achieved when the signals are perfectly phase coherent. 0 amplitude is caused by the complete destructive summation of incoherent signals. The PWS $f_{PWS}(t,p)$ is obtained by multiplying the weight $c_{ps}(t)$ with the linear stack.

$$f_{PWS}(t,p) = \sum_{j=1}^{N} s_j(t - \Delta d_j p) \cdot \left| \sum_{n=1}^{N} \frac{1}{N} e^{i\Phi_n(t-\Delta d_n p)} \right|^{\nu} \tag{4}$$

The phase stack acts as a filter with a certain sharpness of the transition between coherence and incoherence which is controlled by the power ν. For most applications a power ν = 2 seems to be a good choice and is used throughout this paper. We like to add that the PS can be smoothed by averaging within a moving window centered at each sample. This can increase the robustness of the PS.

Application to synthetic data

Figure 2 shows the performance of PWS using a made up data example which employs different types of non-sinusoidal signals and noise. The actogram (Fig. 2a) contains a constant amplitude signal (signal 1), two signals of different rhythms which interfere (signals 2 and 3), signal 4 with varying amplitudes, and the incoherent signals 5 and 6. Signals 5 and 6 are considered to be noise due to their occasional occurrence and varying polarities, respectively. Figure 2b shows the same data with superimposed random noise of similar frequency contents. Note that now signals 4 and 5 can not be distinguished by eye. The corresponding linear stacks (LS), phase stacks (PS), and phase-weighted stacks (PWS) at zero slowness are illustrated in Figures 2c and 2d. It can be seen that PS is large whenever the signals are coherent. It is important to mention here that an analytic signal with zero amplitude theoretically does not have a phase, but that numerically a phase zero is ascribed. In consequence coherent zero amplitude sections have a PS value of 1. In Figure 2c, PS deviates from 1 owing to the presence of numerical noise in the zero amplitude sections. The noisy actogram yields a PS with decreased coherence values for the signals 1, 2, and 4 which, nevertheless, clearly stand out from the surrounding noise. From the LS it is obvious that one can not distinguish between the signals 1, 2, and 4, and the incoherent signals 5 and 6. Indeed all signals show clearly up and thus might be misinterpreted as coherent signals. The PWS, however, is cleaned from signals 5, 6 and other noise. This is due to the small values of the weight function which indicate the little coherence at zero slowness. More coherent signals are less down-weighted and remain visible in the PWS. The envelopes of the linear stacks and PWS as func-

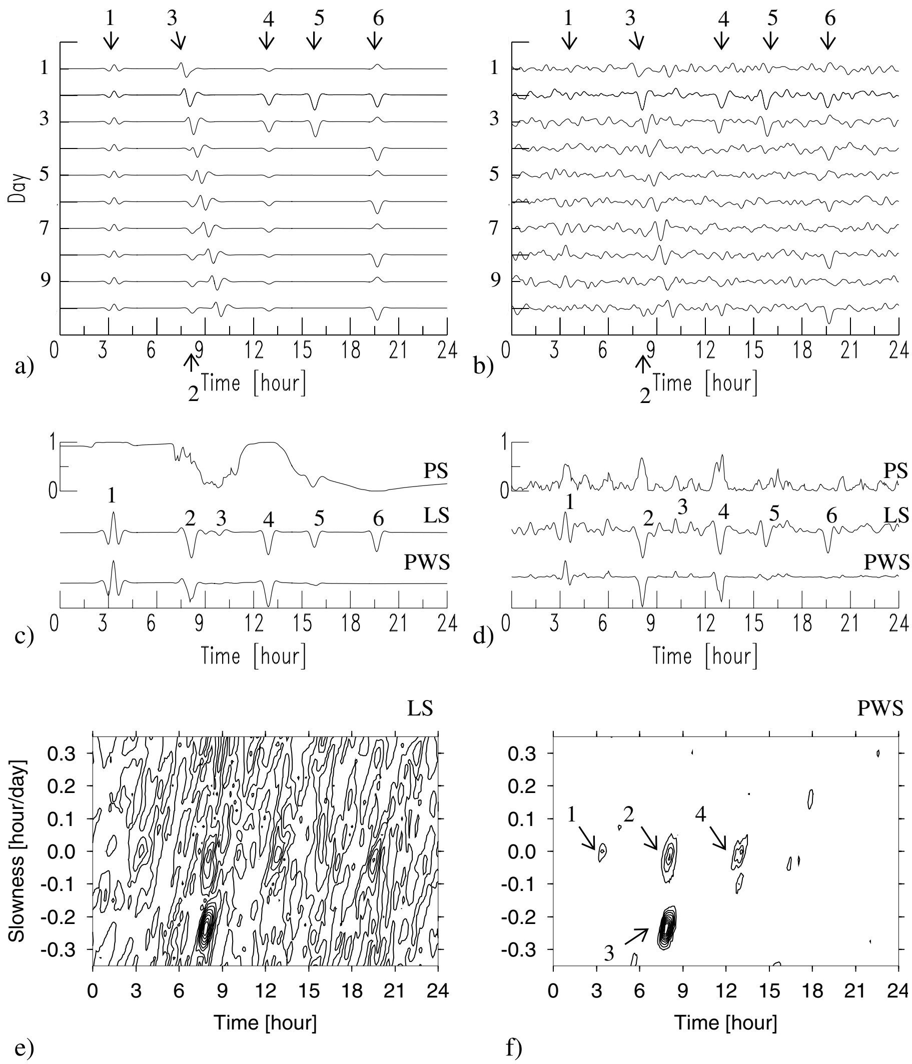

Figure 2. (a) Actogram with different signals labeled by numbers. (b) Same as (a) but with superimposed random noise. (c) From top to bottom: squared phase stack, linear stack, and PWS at zero slowness. (d) same as (c) but for the noisy actogram. (e) Normalized linear stack as function of slowness for the noisy data from (b). Contoured are the envelopes at 0.1 intervals. (f) same as (e) but for PWS. (Figure modified from Schimmel & Paulssen, 1997.)

tion of slowness and time are shown in Figures 2e and 2f. The amplitudes are normalized to one and contoured at 0.1 intervals. It can be seen that PWS has a larger signal-to-noise ratio and enables the correct slowness and time determination of the weak signals.

Data Analysis and Results

The study of temporal patterns of cave crickets showed a somewhat erratic pattern of the activity rhythms, as the raw data of the activity of a male cricket, recorded during 24 days, shows (Fig. 3). Owing to the presence of large amplitude noise the maximum amplitude plotted corresponds to ca 11% of the largest value in the data set.

The time series were equally sampled at 1 min. intervals. Feeding and storing data were carried out at irregular intervals and provoked generally short interruptions in the measurements. The total duration of data gaps reaches almost 14% of the total

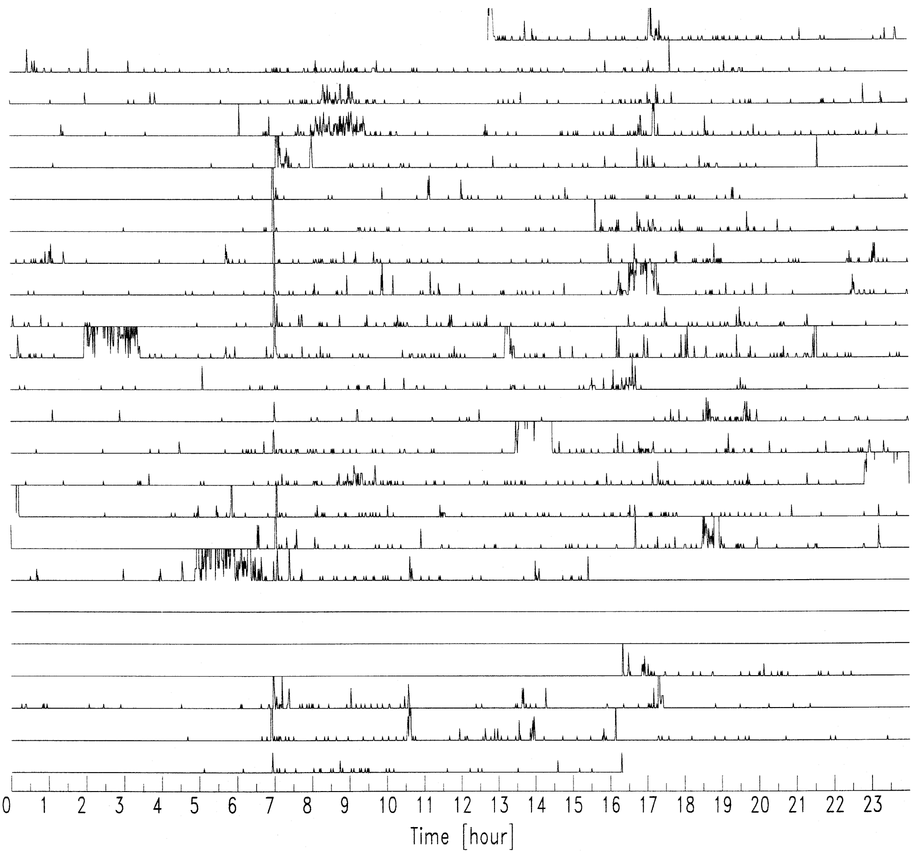

Figure 3. Raw data of the percentage of activity per minute of a male cricket of *Strinatia brevipennis*, tested from 25 November to 18 December, 1997. Conditions of test: LD (11:13-lights on at 07:00), temperature constant at 20°C; humidity oscillating between 90% and 98%. Food *ad libitum*. The maximum amplitude plotted is 8 which corresponds to ca 11% of the maximum value in the data set.

length of the experiment. For the application of the COSINOR, circular statistics, Fourier amplitude spectrum and PWS methods, data gaps were filled by zero amplitudes. This was not necessary for the Lomb-Scargle analysis since it handles missing data. For the Fourier, Lomb-Scargle and PWS analysis the influence of the strong amplitude outliers was decreased using the power 1/3 of the amplitudes. Further, the time series were smoothed by averaging the amplitudes within a moving box car window of 30 min centered at each sample. This decreases the influence of the high frequency noise and increases the robustness of the methods to be applied.

The best adjustment of the analysis employing Cosinor and circular methods pointed out a period of 28 h. The parameters obtained with Cosinor and circular analyses are shown in Table 1. The adjustment of the curves by Cosinor are shown in Figure 4 and the circular distribution can be seen in Figure 5. The Lomb-Scargle periodogram and the Fourier amplitude spectrum are shown in Figure 6. The use of the raw data rather than the smoothed time series does only slightly change the amplitudes of the spectra without inclusion or exclusion of new frequencies.

The graphical results of the analysis with the PWS method are shown in Fig. 7b. For completeness the squared PS and LS are shown in the Figures 7a and 7c, respectively. With PS and PWS, we could establish the period of 24:00 h as the most significative for the series. Moreover, the time of the periodical signals (in the first day) is 07:00 h which is the moment of lights on, and 17:00 h which is one hour before lights off. With these results, we could detect the expected entrainment of the rhythm, with events of anticipation to lights off. Although the time series were smoothed, LS (Fig. 7c) is biased by the large amplitude noise.

Discussion

The advantage of the COSINOR test method is the possibility to obtain rhythm parameters, which are estimated by the least-squares fitting of a cosine function. The

Table 1. Parameters obtained with the Cosinor and circular test methods.

COSINOR	Amplitude	Mesor	p-value	Variance
Period 24 h	1.22	18.92	0.976	7416.92
Period 28 h	15.70	18.71	0.017	7293.89
CIRCULAR	R	r	n	s
Period 24 h	252.29	0.028	9153	05:19
Period 28 h	4045.88	0.442	9153	04:02

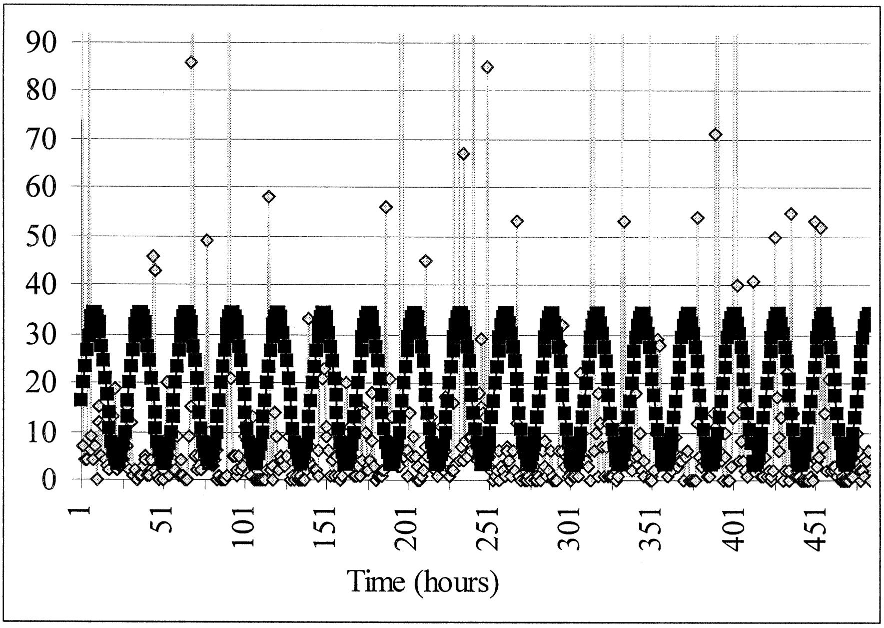

Figure 4. Results of the analysis of the activity rhythm of *Strinatia brevipennis* shown in Figure 3, using the COSINOR method. The lozenges (◇) represent the data obtained, the squares (■) are the adjustment of the results to a cosine curve.

availability of parameters that describe the temporal series is important because they allow comparisons among different series. However, there are some requirements, for instance in relation to residual errors and data distribution, that may not be possible to accomplish. Residual errors should be normally distributed about a fitted cosine curve and the sampling of the data should be equally distributed throughout the studied interval. Further, the least-square fit is strongly biased by large amplitude noise, because these outliers dominate the misfit function.

Circular statistics is more advantageous than COSINOR because it does not demand so many requirements. Nevertheless, it also has disadvantages, mainly when temporal series are long. The assemblage of the data leads to final values so big that the statistical analysis is validated only because of the size of the series and not because of the characteristics of the results. Moreover, the Rayleigh test requires that the data should be distributed according to the von Mises distribution, analogous to the normal distribution for linear data. Hence, data with bi- or multimodal distribution could not be tested by this method.

The Lomb-Scargle periodogram provides a usefull method that handles all types of data collection and solves the missing data problem without requiring too much computational effort. Moreover, it is very easy to apply. However, this technique, as

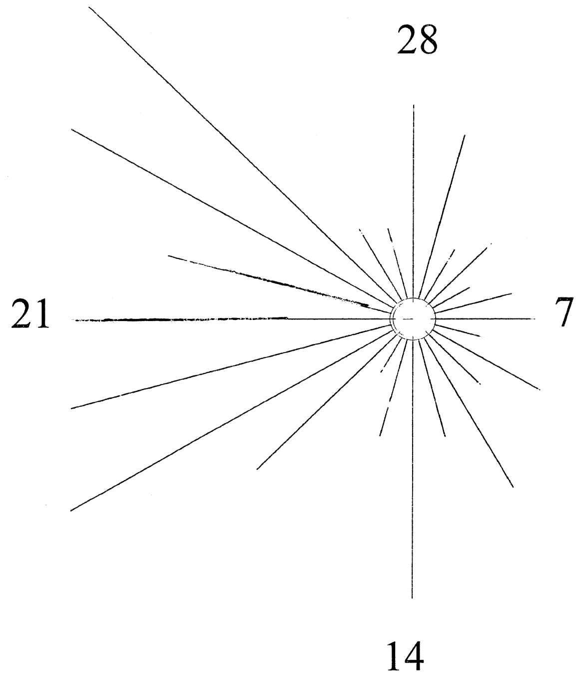

Figure 5. Results of the analysis of the activity rhythm of *Strinatia brevipennis* shown in Figure 3, using circular statistics. Circumference represents 28 h.

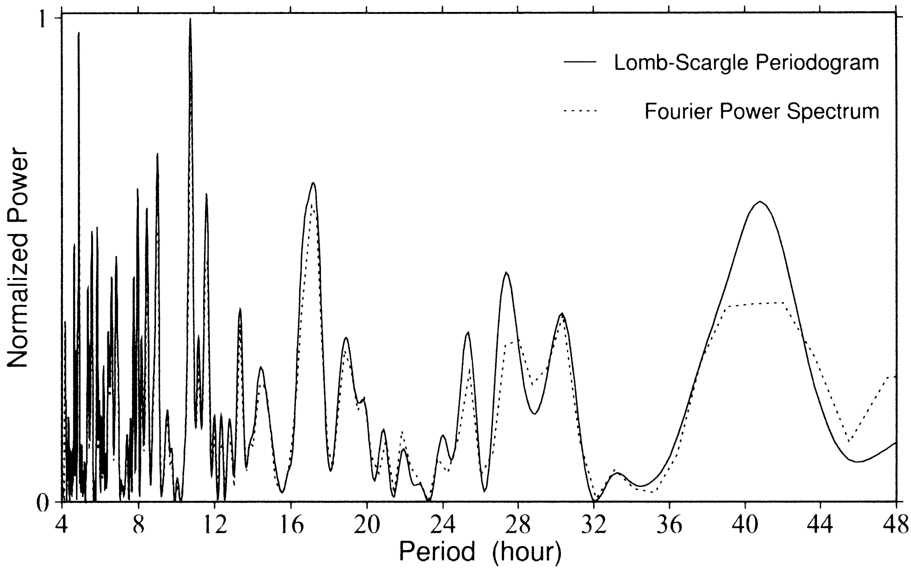

Figure 6. Results of the analysis with the Lomb-Scargle periodogram (10 times oversampled) and the Fourier amplitude spectrum for comparison. For more details, see Ruf (1999) and Van Dongen et al. (1999).

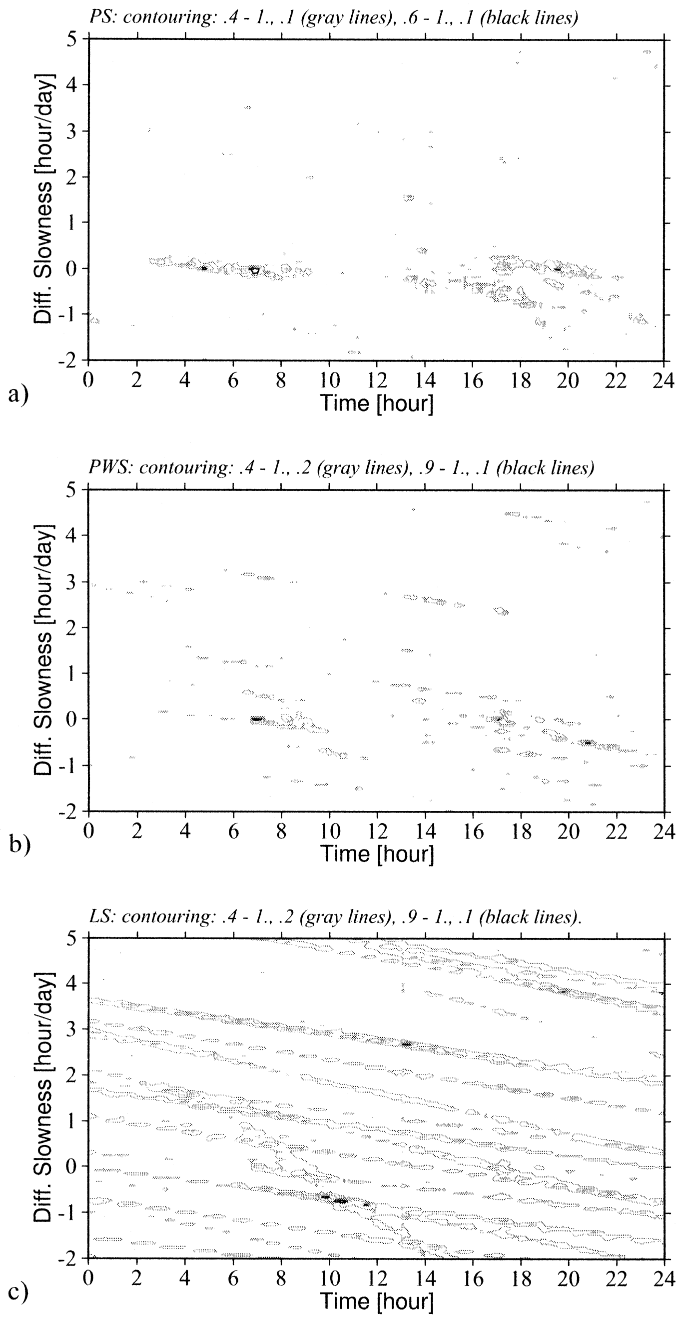

Figure 7. Squared PS (a), PWS (b), and LS (c) of the activity rhythm showed in Fig. 3. Diff. Slowness: means how much the rhythm deviates from 24 h (0 = 24 h). Time: regular clock hour. The gray lines contour the amplitudes from 0.4 to 1 (maximum) in 0.1 intervals. The maximum amplitudes are marked in black.

all the other methods based on fitting sinusoid waves to the data, has the disadvantage of not modeling properly rhythms that present a nonsinusoidal shape, as it seems to be the case of the cave crickets' activity. Because of the poor fit, the goodness of fit should not be used as a measure of the strength of the rhythm (Van Someren et al., 1999). Besides, measurement outliers such as caused by sporadic instrument failures or disturbances in the measuring environment can have strong influences in the power spectrum (Schimmel, subm.). Our data contain outliers of unknown origin which influence the measure.

An important information that the PWS method provides is the time the periodical signals occur in the reference record which in our case is the first trace (first day of measurement). This information, together with the period, allows the tracing of the signal along the series, even in free-running rhythms. The COSINOR also provides this information which, however, is only valid if there is only one signal at the corresponding period. The data presented here contain at least two 24 h signals (at 07:00 h = lights on, and an anticipated activity increase to lights off) which would not be separated by the spectral methods.

However, signals which differ in shape are considered less coherent (depending on the waveform differences) and therefore are difficult to be detected or may even not be detected at all. Another disadvantage of PWS can be that this method was designed for evenly spaced data.

Considering the examples of activity of the cave cricket, a careful examination of the graphical representation leads to notice an organized pattern. However, it is not an easy task to define the range of variation of the rhythm, within the circadian frequencies, in this species, and the understanding of the tendencies of the process is the goal of our studies with cave animals. Thus, the description of the rhythms that this new mathematical tool provides makes worthy the efforts to incorporate this method to those already known.

Conclusions

We presented here another method for the analysis of temporal data, among several that exists nowadays in the literature. This method, however, seems to be useful for those data without very clear rhythmic patterns, and when the amplitude of the signal changes often.

The PWS method enabled the detection of coherent non-sinusoidal signals in cricket data. On one hand, PWS is less affected by strange outlying measurements in the data. On the other hand, it is designed for evenly sampled data. Anyway, this method is a powerful tool for detection of non-sinusoidal rhythms, whereas other very used methods, which effectively detect sinusoidal rhythms, have difficulties in detecting the observed rhythm.

With the results obtained with the PWS, we can say that the cricket species studied seem to have a well developed endogenous clock which is sincronized by the light/dark cycle, since an anticipation to the dark could be observed in terms of an increase in activity one hour before the switch-off of the lights.

Acknowledgements

The authors wish to thank the comments of Dr. G. Ortega that did much to improve the final version of the manuscript. This study was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP # 96/02494-3 to S. Hoenen and # 97/04780-6 to M. Schimmel).

Bibliography

Batschelet E (1981): *Circular statistics in biology*. London, Academic Press.

Benedito-Silva AA (1997): Aspectos metodológicos da cronobiologia. In: Marques N, Menna-Barreto L, eds., *Cronobiologia: Princípios e Aplicações*. São Paulo, EDUSP, pp. 215–238.

Bracewell RN (1965): *The Fourier transform and its applications*. New York, Macgraw-Hill Book Co., Inc.

Hoenen S, Gnaspini P (1999): A new system for recording endogenous activity rhythms of cave arthropods. *Mém Biospeol 26*: 75–83.

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

Nelson W, Tong YL, Lee J-K, Halberg F (1979): Methods for Cosinor-rhythmometry. *Chronobiologia 6(4)*: 305–323.

Percival DB, Walden AT (1993): *Spectral analysis for physical applications (Multitaper and conventional univariate techniques)*. Cambridge, Cambridge University Press.

Ruf T (1999): The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. *Biol Rhy Res 30(2)*: 178–201.

Scargle JD (1982): Studies in astronomical time series analysis. 2. Statistical aspects of unevenly spaced data. *Astrophys J 302*: 757–763.

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

Tanner MT, Koehler F, Sheriff RE (1979): Complex seismic trace analysis. *Geophys 44*: 1041–1063.

Van Dongen HPA, Olofsen E, Van Hartevel JH, Kruy EW (1999): A procedure of multiple period searching in unequally spaced time-series with the Lomb-Scargle method. *Biol Rhy Res 30(2)*: 149–177.

Van Someren EJW, Swaab DF, Colende CC, Cohen W, McCall WV, Rosenquist PB (1999): Bright light therapy: improved sensitivity to its effects on rest-activity rhythms in Alzheimer patients by application of nonparametric methods. *Chronobiol Int 16(4)*: 505–518.

Zar JH (1996): *Biostatistical analysis*. New Jersey, Prentice Hall.