Searching for Biological Rhythms: Peak Detection in the Periodogram of Unequally Spaced Data

Hans P. A. Van Dongen,*[,1] Erik Olofsen,[†] Jan H. VanHartevelt,[‡] and Erik W. Kruyt[‡]
*Division of Sleep and Chronobiology, Department of Psychiatry, and Center for Sleep and Respiratory Neurobiology, University of Pennsylvania Health System, USA,
[†]Department of Anaesthesiology, Leiden University Medical Centre, the Netherlands,
[‡]Department of Physiology, Leiden University Medical Centre, the Netherlands

Abstract The classical power spectrum, computed in the frequency domain, outranks traditionally used periodograms derived in the time domain (such as the χ^2 periodogram) regarding the search for biological rhythms. Unfortunately, classical power spectral analysis is not possible with unequally spaced data (e.g., time series with missing data). The Lomb-Scargle periodogram fixes this shortcoming. However, peak detection in the Lomb-Scargle periodogram of unequally spaced data requires some careful consideration. To guide researchers in the proper evaluation of detected peaks, therefore, a novel procedure and a computer program have recently become available. It is recommended that the Lomb-Scargle periodogram be the default method of periodogram analysis in future biomedical applications of rhythm investigation.

Key words time series analysis, periodogram analysis, power spectrum, Lomb-Scargle method, peak detection, multiple peaks, biological rhythms, unequally spaced data

Periodogram analysis plays a major role in the study of biological rhythms, and for good reasons. The periods (or frequencies) and the statistical significance of multiple rhythms in a time series can be assessed, whether these rhythms are a priori known to exist or not. In fact, a periodogram is often used to detect and evaluate possible rhythms, a procedure that is known as "peak detection." In many biomedical applications, data-folding techniques are traditionally used to obtain a periodogram in the time domain, such as the χ^2 periodogram (Sokolove and Bushell, 1978). These periodograms give measures of "power" (i.e., intensity) of rhythms in the time series for a range of periods T. Peaks in the periodogram indicate relatively dominant rhythms at the corresponding periods. These rhythms can be identified as statistically significant if their powers exceed a certain preestablished threshold.

Unfortunately, the χ^2 periodogram and other periodograms based on data folding suffer a number of undesirable properties. These include the following: the time series must be equally spaced (i.e., regular sampling and no missing values); the resolution of detectable periods is limited, and equal to the sampling interval Δt (i.e., $T = j \prec \Delta t$ with $j = 1, 2, \ldots$); the powers at adjacent periods (say, T and $T + \Delta t$) are correlated; the threshold for statistical significance varies with period; there is no straightforward relationship between the power of a peak and the amplitude of the corresponding rhythm; identification of the most dominant rhythm in a time series may be ambiguous; detectability is relatively poor under high noise conditions (i.e., low signal-to-noise ratio); and peaks reappear at multiples of the actual period of a rhythm (i.e., "subharmonics"). As a consequence, investigators must apply such periodograms with great care and

1. To whom all correspondence should be addressed.

JOURNAL OF BIOLOGICAL RHYTHMS, Vol. 14 No. 6, December 1999 617-620

should rely on additional methods of time series analysis to reconfirm their findings.

A good alternative to periodogram analysis in the time domain is periodogram analysis in the frequency domain, also called power spectral analysis. A power spectrum gives measures of power for a range of frequencies f, which are the inverse of periods T, for example $f = 1 / T$. Well-known is the "classical" power spectrum that can be obtained with a famous computer algorithm called Fast Fourier Transform (FFT). For the classical power spectrum, only the following three undesirable properties of time domain periodograms apply: the time series must be equally spaced; the resolution of detectable periods (frequencies) is limited, in this case by the total duration T_{tot} of the time series (i.e., $f = j / T_{tot}$ with $j = 1, 2, \ldots$); and peaks reappear, in this case beyond a frequency limit called the Nyquist frequency $f_{Ny} = 1 / (2 \times \Delta t)$ and referred to as "aliasing". Fortunately, in most biomedical applications, sampling is sufficiently rapid for rhythms of interest to be well below the Nyquist frequency (although it is important to recognize that the undersampling of rhythmicities, whether of interest or not, renders any type of periodogram analysis invalid, because power ends up at the wrong frequencies).

The limitation on the frequency resolution, and especially the necessity of equally spaced sampling, restrict the applicability of the classical power spectrum in the study of biological rhythms. It is often difficult to obtain entirely equally spaced sampling, either because of the nature of the measured variables or because external factors disturb the measurements. To give a few examples, data collection on free-living animals may only be possible at times when a researcher is able to track the animals, and particular behaviors can only be studied if such behaviors actually take place. In these cases, the investigator is forced to draw unequally spaced samples. Also, while human thermoregulatory variables may be measured with precise regularity, for instance, occasional movement or feeding artifacts must often be rejected from the data. Similarly, equipment failure may cause the loss of samples, resulting in time series with missing values.

Thus, there is a need for a periodogram that can handle unequally spaced data, and preferably that also allows for a better frequency resolution. A direct generalization of the classical power spectrum to accept unequally spaced sampling times is possible by using a Discrete Fourier Transform instead of the Fast Fourier Transform algorithm, but the result is unsatisfactory, because this generalized power spectrum is not invariant to time translation. This means that there is an unpredictable component in the results that depends on the (arbitrary) choice of the origin of time (i.e., time "zero") or, equivalently, on the phases of the rhythms in the time series. This property clearly defeats the purpose of periodogram analysis.

PEAK DETECTION IN THE LOMB-SCARGLE PERIODOGRAM

Independently, Lomb (1976) and Scargle (1982) recognized the problems associated with periodogram analysis of unequally spaced data. Based on harmonic regression (i.e., cosinor analysis), they developed a least-squares power spectrum that fixes the invariance to time translation problem of the generalized power spectrum. This Lomb-Scargle periodogram, as it is currently called, was made available to the public at large by the work of Press et al. (1992), who also implemented an optimized algorithm that makes computation feasible for up to about a million data points on modern computers. The Lomb-Scargle periodogram has excellent properties with well-defined statistical characteristics, is capable of properly handling unequally spaced data, and also allows—in principle—for infinite frequency resolution (obviously, the noise in the time series determines the actual accuracy for frequency estimates of detected rhythms). Furthermore, for equally spaced data and at frequencies $f = j / T_{tot}$ (with $j = 1, 2, \ldots$), the results are exactly the same as for the classical power spectrum.

Ruf (1999) discussed the superiority of the Lomb-Scargle periodogram compared to the χ^2 periodogram in biomedical applications and emphasized this point with simulated examples. Earlier, a review of the Lomb-Scargle periodogram, with extensive background information and many relevant formulae, appeared in Van Dongen et al. (1997). Here it was considered that the properties of the Lomb-Scargle periodogram make it, in any case, the method of choice for periodogram analysis of *equally* spaced data. For unequally spaced data, however, the advantages of the Lomb-Scargle periodogram may be concealed by effects of the unequally spaced sampling per se, as can be particularly conspicuous if there is more than one rhythm in the time series (note also that it is often impossible to rule out that there are multiple rhythms in a time series, even if only one rhythm is of interest).

Unequal spacing of the data, therefore, may lead to interpretative challenges, due to the added information embedded in (specific patterns in) the *timing* of the samples. This causes additional peaks to appear in the periodogram for rhythms that merely exist by virtue of the sampling times t_i. Naturally, since there is only a finite number of samples, there can be only a finite amount of information in the periodogram; hence, peaks due to the sampling times t_i and peaks associated with the data points y_i may get mixed up. The consequence is that the powers at all frequencies in the Lomb-Scargle periodogram may be correlated with each other if the data are unequally spaced.

To a large extent, however, the information embedded in the sampling times can be disentangled from the information contained in the actual measurements. This is done by considering the "periodogram window," which is the periodogram of the sampling times t_i alone, with no consideration of the measured values y_i. The periodogram window provides insight into the correlation of peaks in the Lomb-Scargle periodogram caused by the unequal spacing of the sampling times. This information can be used to interpret each peak in the Lomb-Scargle periodogram, one by one (Ferraz-Mello, 1981), in a statistically reliable manner and with great accuracy and frequency resolution. A novel procedure of (multiple) peak detection in the Lomb-Scargle periodogram of unequally spaced data, that takes advantage of the information in the periodogram window, was first presented (and illustrated with examples) in Van Dongen et al. (1999). A free computer program[1] for this procedure was also made available.

DISCUSSION

The Lomb-Scargle periodogram is a modern tool for periodogram analysis of unequally spaced data. When applied to equally spaced data, it yields the same results as the classical power spectrum, although the Lomb-Scargle periodogram allows for a much higher frequency resolution. Because of its desirable mathematical and statistical properties, its outstanding performance relative to the χ^2 periodogram, its capability of handling unequally spaced data, and its availability and ease of implementation on modern computers, it is recommended that the Lomb-Scargle periodogram be the default method of periodogram analysis in the search for biological rhythms.

A careful approach must be taken when analyzing unequally spaced data. Power in the Lomb-Scargle periodogram (and any other type of periodogram) is confounded by the additional information embedded in the unequally spaced sampling times per se. The interpretation of peaks in the periodogram must therefore be performed on a one-at-a-time basis, with mindful consideration of the power contribution due to the sampling times. A step-by-step procedure and a corresponding computer program were recently made available to guide investigators in this process. With this novel procedure, the Lomb-Scargle periodogram is believed to be the most versatile method currently available for the detection and evaluation of biological rhythms.

ACKNOWLEDGMENTS

This work was supported by the Dutch Organization for Scientific Research grant 575-65-068, and in part by NASA cooperative agreement NCC 9-58 with NSBRI, AFOSR grant F49620-95-1-0388, and the Institute for Experimental Psychiatry Research Foundation.

NOTE

1. This Matlab® computer program, which includes computation of the Lomb-Scargle periodogram as well as the periodogram window, can be obtained for free by e-mail to e.olofsen@anesthesiology.medfac.leidenuniv.nl.

REFERENCES

Ferraz-Mello S (1981) Estimation of periods from unequally spaced observations. Astron J 86:619-624.

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

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (1992) *Numerical Recipes in C: The Art of Scientific Computing*, 2d ed, Cambridge University Press, Cambridge UK.

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

Scargle JD (1982) Studies in astronomical time series analysis: II. Statistical aspects of spectral analysis of unevenly spaced data. Astrophys J 263:835-853.

Sokolove PG and Bushell WN (1978) The chi square periodogram: Its utility for analysis of circadian rhythms. J Theor Biol 72:131-160.

Van Dongen HPA, Olofsen E, VanHartevelt JH, and Kruyt EW (1997) Periodogram analysis of unequally spaced data: The Lomb method. Department of Physiology, Leiden University, Leiden. ISBN 90-803851-1-5.

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