2

Statistical Analysis of Biological Rhythm Data

Harold B. Dowse

Summary

The author has developed an ensemble of digital signal analysis techniques applicable to biological time series containing circadian and ultradian periodicities that is of very high resolution and functions well even in the presence of extreme noise and trend. A method for quantifying the significance, strength, and regularity of the rhythmic process is included. To illustrate these techniques, the author presents analyses of artificial periodic data containing varying amounts of noise, trend, and multiple periodicities. The periods and amplitudes of circadian and, where included, ultradian periodicities, and all other components of the test signals are known exactly. Analyses are illustrated in a step-by-step manner and the results are compared with the known input parameters. Trends are removed; spectra, autocorrelation functions, and rhythmicity indices are produced and discussed. References covering theory and details of all analyses are supplied. All programs employed are available from the author free of charge.

Key Words: Autocorrelation; biological rhythms; circadian; maximum entropy spectral analysis; MESA; noise; rhythmicity index; RI; spectral analysis; trend; ultradian.

1. Introduction

Extraction of reliable information on period, phase, and signal robustness from biological rhythm data is central to interpretation of the data. Advances in digital signal analysis in the fields of engineering, geology, and astrophysics have yielded a good selection of high-resolution algorithms, a number of which are applicable to analysis of biological time series. A combination of the older autocorrelation analysis with the modern technique of maximum entropy spectral analysis (MESA), adapted for the short, noisy, nonstationary time series typical of biological systems, has proven especially useful. MESA provides an extremely reliable estimate of period with arguably the highest resolution and sensitivity available, whereas autocorrelation can be used to test for the statistical significance of any periodicity in the data and, at the same time, provides

From: *Methods in Molecular Biology, vol. 362: Circadian Rhythms: Methods and Protocols*
Edited by: E. Rosato © Humana Press Inc., Totowa, NJ

a useful measure of the strength of the signal being analyzed. In order to maximize the reliability and sensitivity of these analyses, it is common to remove digitally any nonstationary trends, as well as other periodicities not of interest. As an example of the latter, ultradian periodicity in deep body temperature in the hourly (circhoral) range may be obscured by the much stronger circadian cycle. Removal of the near 24-h period enables analysis of the ultradian oscillations ***(1–14)***.

2. Materials

1. PC running Windows® operating system capable of executing programs in a DOS window, or any system that can run MATLAB® (*see* **Note 1**) or similar software.
2. Text data editor with sufficient capacity to handle the required data stream. MS Word in text mode is convenient in Windows OS.
3. Any plotting system with sufficient capacity for large files. MATLAB is used here throughout.
4. SIGGEN: This program can create data sets containing up to two periodicities with sinusoidal, square, triangle, or sawtooth wave forms with any desired percentage of white noise added (*see* **Note 2**).
5. HRMES: This is a version of MESA adapted for biological rhythms by the author. It has a low-pass filter that can optionally be activated at run time, which digitally removes noise in the very short period range (*see* **Note 3**).
6. AUTOCO: This is a straightforward autocorrelation analysis program written by the author that contains the same optional digital filter as that found in MESA.
7. FILCON (*see* **Note 4**): This routine is used here for high-pass filtering to eliminate long-range trends or for removal of circadian periodicities to accentuate ultradian rhythms.
8. CROSSCO: A variation on autocorrelation that allows an objective estimate of phase and, additionally, a test of whether two data sets have periodicity in common.
9. AUTPEAK: A routine to find the height of the third autocorrelation peak to yield the rhythmicity index (RI; *see* **Note 5**). RI is an objective measure of how robust and regular a periodic signal is.
10. MESPEAK: A Turbo Basic routine to find the height of the three most prominent peaks in a MESA spectrum (*see* **Note 5**).

3. Methods

The methods outlined below describe the editing/formatting, preconditioning, analysis, and plotting of biological rhythm data in the circadian and ultradian range. The author has used similar programs to analyze *Drosophila melanogaster* heartbeat as well, but this aspect is not covered here (*see* **Note 6**). For this work, it was considered desirable to use artificial data, as periodicity, signal-to-noise ratio, amplitude, and waveform are controlled and hence are known exactly beforehand.

3.1. Editing and Preparation of Data Vector

3.1.1. Creation of Data Sets to Be Analyzed

The program SIGGEN was employed to create data sets similar to biological data. A simple sinusoid was chosen, but other waveforms can be produced by this program. A circadian period of 24.75 h was chosen for all data sets. This noninteger value provides a realistic test of the capabilities of the analysis programs. In all examples, the amplitude was set at unity and was positive everywhere except as noted. Both the MESA and autocorrelation routines in the suite always fit a straight line to the data by regression and subtract it to eliminate any linear trend. This has the added benefit of adjusting the mean of the series to zero. Eliminating this so-called "DC offset" emphasizes the periodic portion of the signal in the subsequent analysis. In general, amplitude will affect spectral power, but not RI. All data sets consisted of 10 d of readings with a sampling rate at twice per hour with one exception, as noted (SET III). A more rapid sampling rate has been shown to be unnecessary for periodicities in the circadian range and very high rates may actually be undesirable if behavioral data are being summed into "bins" ***(15)***.

In SET I, the data were produced with amounts of white noise added in the following percentages: 10, 20, 30, 40, 50, 60, 70, and 80. The raw data set generated with 80% noise is shown in **Fig. 1** (*see* **Note 7**).

In SET II, data were produced identically to SET I, except that a long-range sinusoidal trend was added with a period of 40 d, with the phase angle adjusted so that the data have an approximately monotonically increasing envelope from $t = 0$ to 10 d. The amplitude of this sinusoid was set at 10-fold that of the circadian periodicity. **Figure 2** depicts data containing 80% noise with the long-period trend added.

In SET III, data were produced as in SET I and additionally had a 1-h circhoral ultradian component added. The peak-to-peak amplitude of the circadian rhythm was 10-fold that of the ultradian, which was unity. Sampling rate was set to 12/h (i.e., at 5-min intervals) to characterize more adequately this relatively high-frequency component ***(7)***. Such circhoral rhythms have been uncovered in human deep-core body temperature using the sensitive methods outlined here ***(14)***. **Figure 3** shows the data containing the standard circadian periodicity along with an ultradian periodicity of 1 h. In this example 80% noise was added.

For SET IV, a series of files created with data sampled twice per hour and with a 24.75-h periodicity was generated. In these examples, phase was delayed by 0, 2, 4, 6, 8, and 22 h, respectively, to show the output of CROSSCO.

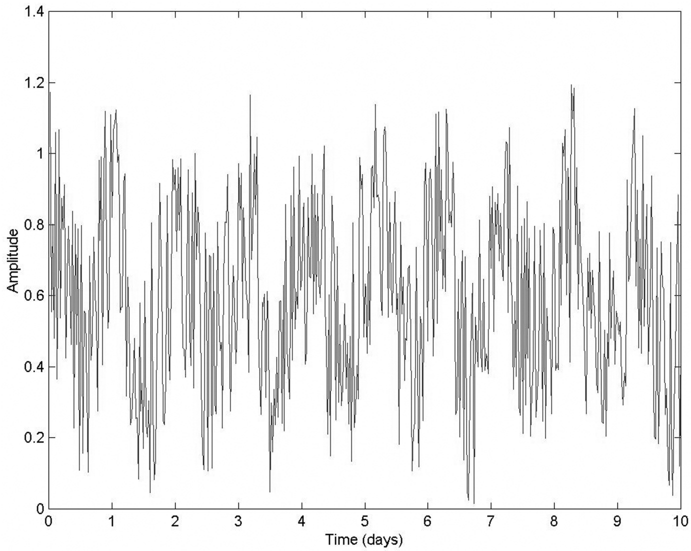


Fig. 1. An artificial data set consisting of a unit amplitude sinusoid and periodicity of 24.75 h with an added 80% white noise. Data were sampled at half-hour intervals over the equivalent of 10 d.

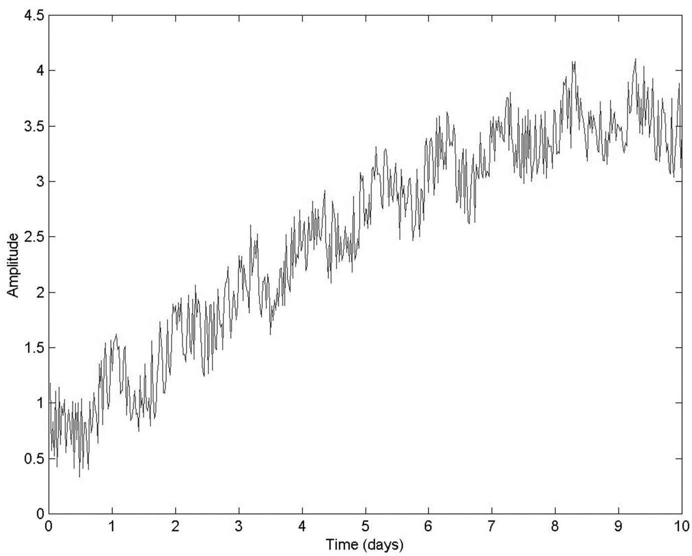


Fig. 2. The data set depicted in **Fig. 1**, but in this case with a monotonic trend added with amplitude functionally five times that of the sinusoid. Period and amplitude of the signal remain in **Fig. 1**.

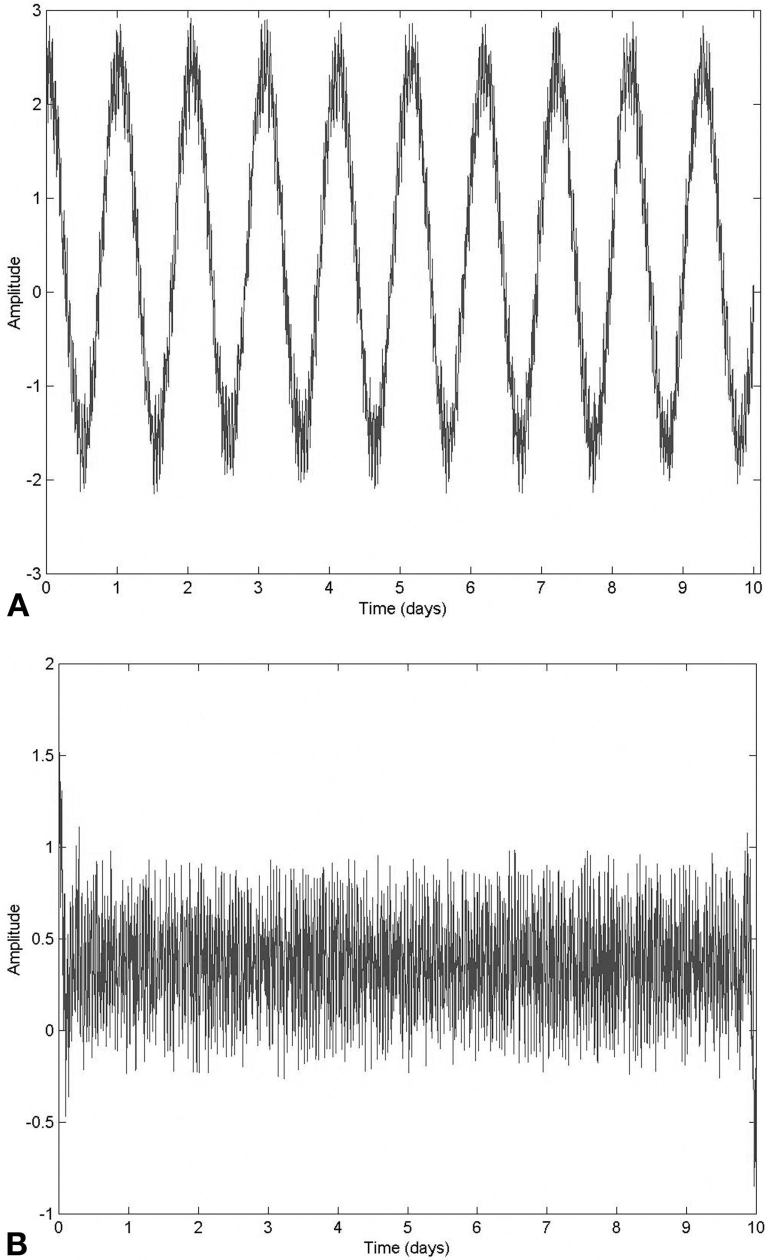


Fig. 3. (**A**) Data as **Fig. 2**, containing a periodicity of 24.75 h and 80% noise, but with a unit amplitude ultradian rhythm having a period of 1 h added. This cannot be seen in the unfiltered data. (**B**) The same data set after removal of the circadian range periodicity using FILCON.

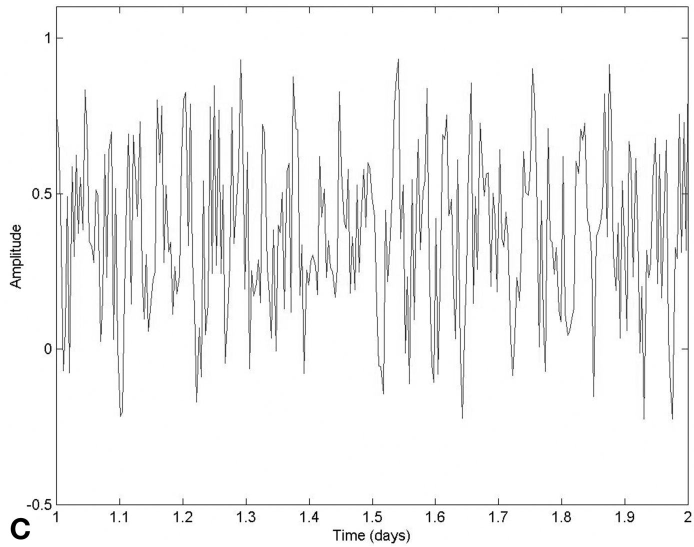


Fig. 3. (*continued*) (**C**) Here, a 1-d interval of the data is shown at a magnified scale to illustrate the noisy hourly rhythm.

3.1.2. Format of Data

The analysis software is configured to allow batch processing of concatenated files. Input data are placed in a single column. Each separate file is given a descriptive, reasonably short title (here, fewer than 15 characters) at the start of the vector. The end of each individual file is signaled by appending –5000. Batch processing is terminated with the word "END." Output of both spectra and autocorrelations is written file by file in two columns, with either spectral power or autocorrelation values in the first column and either period or lag in the second. In autocorrelation output, the 95% confidence value indicating "significant rhythmicity" is given below the title (*see* **Note 8**). For actual rhythmicity data, the columns of input are concatenated and formatted as above using an editor. Output for individual spectra and correlograms excised from the concatenated file were plotted using MATLAB (*see* **Note 9**). An example of the raw data format follows:

```
I80
1.173287
5.549589E-01
8.012864E-01
4.805624E-01
1.060405
```

```
...
3.837983E-02
2.867427E-01
4.349197E-01
6.987450E-01
8.844982E-01
1.210237E-01
4.635225E-01
-5000.000000
END
```

3.2. Estimating Periodicity and Strength of Rhythmicity

3.2.1. Estimates of Period

To estimate period, the concatenated data sets were analyzed using MESA and the results plotted using MATLAB software. Data were analyzed both with and without the use of a 4-h-cutoff low-pass two-pole Butterworth digital filter (*see* **Note 10** and **ref.** ***16***). To report the best estimate, the three highest spectral peaks were read from the output files with MESPEAK. These results were then checked against the output plots and the autocorrelations to ensure that the peak best representing the periodicity was chosen.

The output of MESA analysis for the data file from SET I, containing 80% noise with no low-pass filtering, is shown in **Fig. 4**. The period, as retrieved from the output data file, is 24.62, corresponding closely to the known programmed-in value. Despite the very low signal-to-noise ratio in the raw data, the MESA spectral peak is sharp and little of the noise appears in the region of the circadian periodicity. **Table 1** contains the results for periodicity in SET I.

For the data in SET II, with the long-period trend, the data were either run as is, with and without the low-pass filter, or after first being conditioned by removal of all periodicity greater than 2 d using FILCON, set to null out the coefficients for periodicities greater than 2 d. Analysis proceeded as usual from that point on. **Table 1** summarizes the output for these data. An improvement in the estimates is notable.

For SET III (with the circhoral component), the data were analyzed first untreated, and then again after conditioning with FILCON set to remove all periods greater than 10 h. **Figure 5** depicts the spectrum derived from the data containing a circadian period of 24.75 h, a 1-h ultradian periodicity of much lower amplitude, and 80% noise after de-trending with FILCON. **Table 1** shows the periodicities extracted by MESA before and after removal of the circadian periodicity. For data before filtering, both the primary and secondary peaks found are reported.

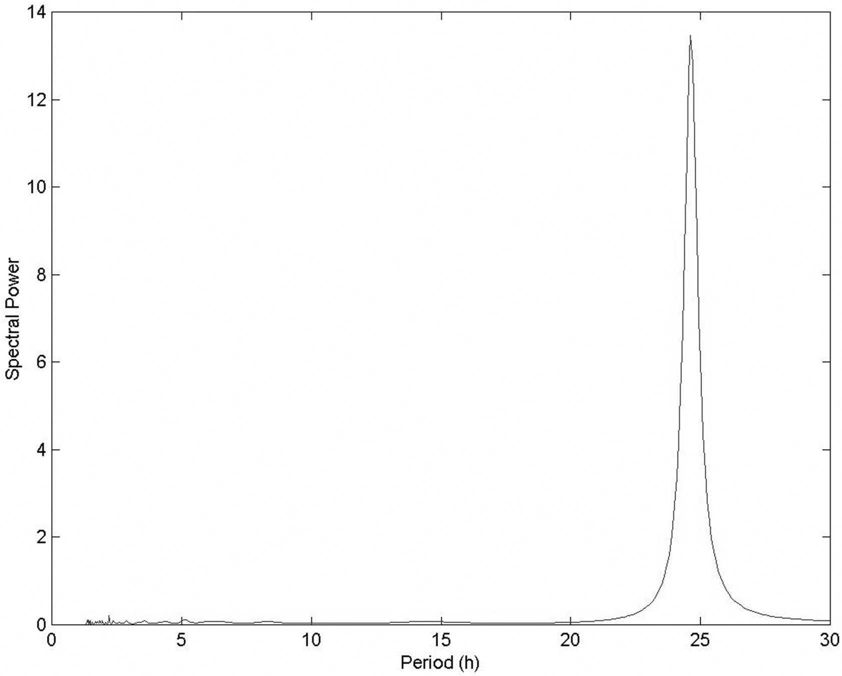


Fig. 4. The MESA spectrum from the data depicted in **Fig. 1**. No filtering was done, yet the spectrum is extremely clean, even given the high proportion of noise and attendant low signal-to-noise ratio.

Table 1
Primary Periods Found by MESA for Each Level of Noise

% Noise	I	II U	II F	III U 1°	III U 2°	III F
10	24.78	25.26	24.81	24.8	1.00	1.00
20	24.78	25.34	24.81	24.72	1.00	1.00
30	24.78	25.34	24.88	24.72	1.00	1.00
40	24.78	25.43	24.88	24.80	1.00	1.00
50	24.78	25.43	24.88	24.72	1.00	1.01
60	24.78	25.51	24.97	24.80	1.00	1.01
70	24.78	25.43	24.97	24.80	1.00	1.01
80	24.62	25.10	24.88	24.80	1.00	1.01

The data set analyzed is noted by set number (*see* **Subheading 3.2.1.**). U = Raw (1° and 2° indicate primary and secondary peaks uncovered), and F = output after program FILCON was run on the data. Note that even before the strong circadian rhythm was removed, MESA found the hourly peak even in 80% noise.

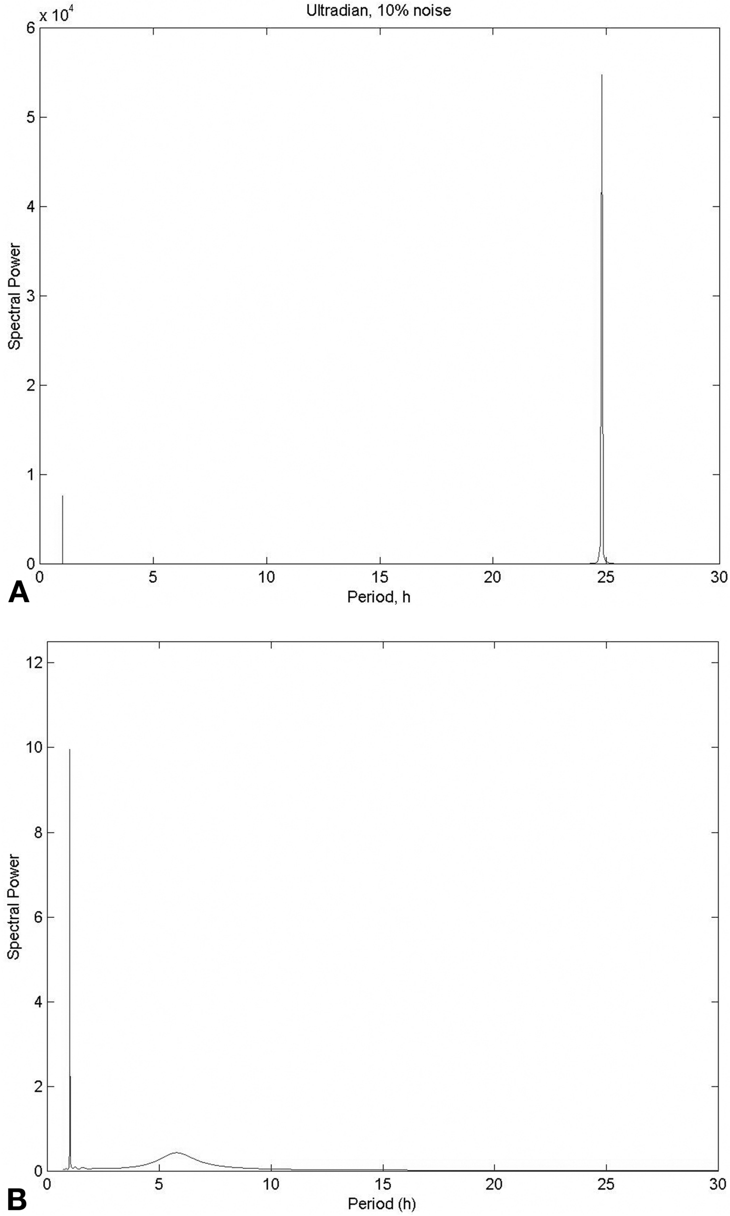


Fig. 5. MESA plot produced from the data shown in **Fig. 3** (**A**) before removal of the circadian component with FILCON and (**B**) after removal. The large circadian peak has been entirely eliminated, and the remaining ultradian periodicity is clear, constituting the only peak in the spectrum. The small bump at approx 6 h is an artifact of the sharp cutoff of the Fourier filtering.

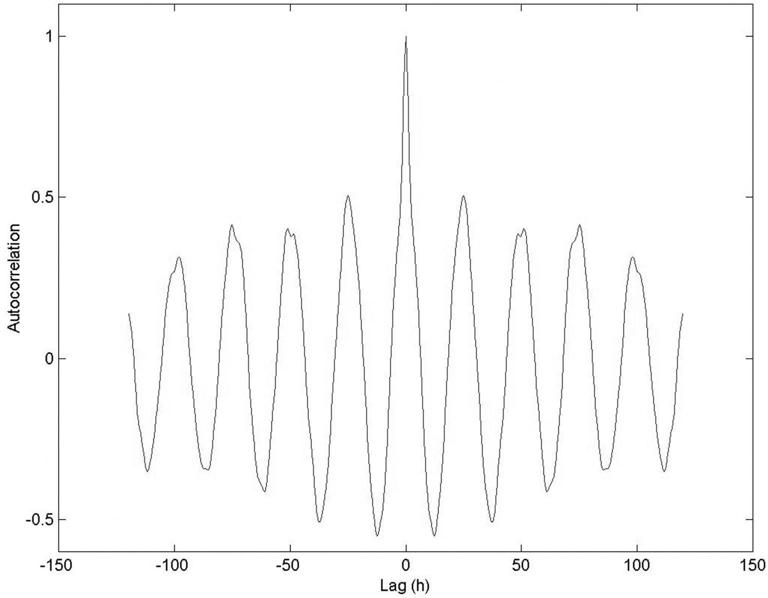


Fig. 6. Autocorrelation plot from the data set depicted in **Fig. 1**. The decay of the envelope in the function is a result of the large amount of noise added. Nonetheless, strong rhythmicity is evident and corroborates the periodicity reported by MESA.

3.2.2. Autocorrelation Analysis

To test for significance of periodicity and provide a crosscheck on the period estimates, the program AUTOCO was run on all data as above. All data were treated as had been done with MESA with regard to filtering and conditioning with FILCON. Plotted output was compared with the spectral analysis results to check for agreement. The autocorrelogram was shown in **Fig. 6**. computed for the same data set that was analyzed by MESA in **Subheading 3.2.1.** For all figures in this section, correlograms shown will be for the same files as the MESAs in the previous section. Note that the peaks of autocorrelation are robust and repeat regularly, clearly verifying the periodicity reported by MESA in **Fig. 4**.

The autocorrelation for the data initially containing a trend are shown before and after operation of FILCON (**Fig. 7**). Note that the peaks of autocorrelation are superimposed on the strong trend, but that this disappears after its removal. It is substantially easier to verify the MESA peak using correlograms from de-trended data.

An autocorrelation of the data from SET III, containing the ultradian rhythm in the presence of the strong circadian rhythm and 80% noise, is depicted in

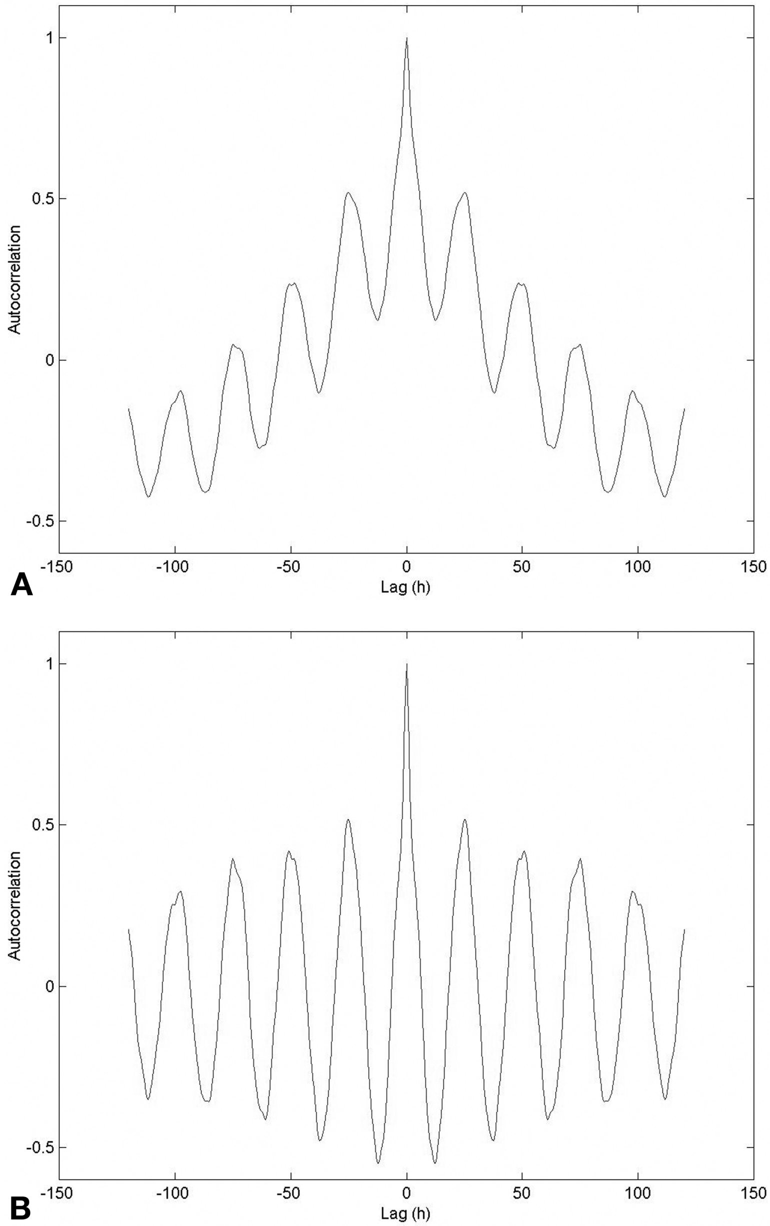


Fig. 7. Autocorrelation analysis of the data depicted in **Fig. 2** with a strong monotonic trend shown before (**A**) and after (**B**) de-trending with FILCON.

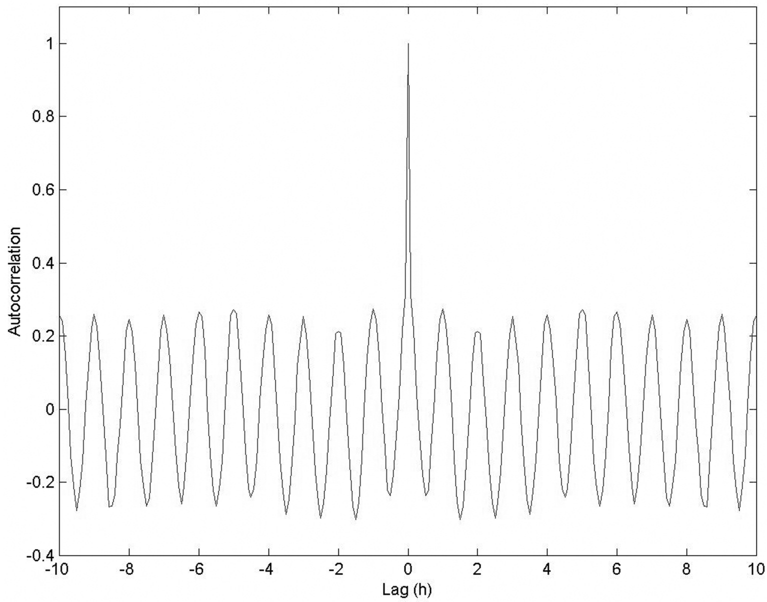


Fig. 8. Autocorrelation produced from data containing a 1-h ultradian periodicity, as depicted in **Fig. 3**, after the circadian component had been removed. Before removal, there was no visible evidence of this period. The scale has been shortened to ±10 h of lags. This corroborates the MESA spectrum for both the filtered and unfiltered data.

Fig. 8. Here, the correlograms before and after removal of the circadian rhythm are shown to illustrate the dramatic change in the character of the signal. The hourly rhythm is unequivocally verified by the regular peaks in the autocorrelation function (*see* **Note 11**).

3.2.3. Assessing Robustness of Rhythmicity

The output files from the autocorrelation analysis were further scrutinized for RI with a BASIC program set to find the height of the third peak (counting the peak at lag zero as 1). As has been discussed elsewhere ***(7–11)***, the decay envelope of the autocorrelation function is a measure of the long-range regularity of the rhythmicity, as well as its robustness. These numbers could then be compared with the known (programmed-in) signal-to-noise ratios.

Table 2 contains the extracted RIs for SETS I through III. Note that the RI covaries with the amount of noise added. Filtering the data produces a striking increase in the RI at the higher noise levels. For the data containing the ultradian component, the RIs before filtering are representative of the circadian periodicity, whereas numbers after filtering are for the remaining ultradian rhythm.

Table 2
Rhythmicity Index (RI) for Each Percentage of Noise

% Noise	SET I	SET II U	SET II F	SET III U	SET III F
10	0.79	0.73	0.79	0.79	0.66
20	0.79	0.71	0.78	0.79	0.66
30	0.78	0.68	0.77	0.79	0.65
40	0.77	0.65	0.76	0.79	0.63
50	0.74	0.59	0.73	0.79	0.59
60	0.70	0.50	0.68	0.79	0.52
70	0.60	0.38	0.58	0.78	0.39
80	0.41	ARR	0.40	0.77	0.21

Data set numbers are indicated (*see* text). U = raw data, F = output after FILCON program operated on the data. ARR means the signal was deemed arrhythmic. For data in SET III, the RI for the circadian component is reported for the raw data, and for the data treated with FILCON in the F group.

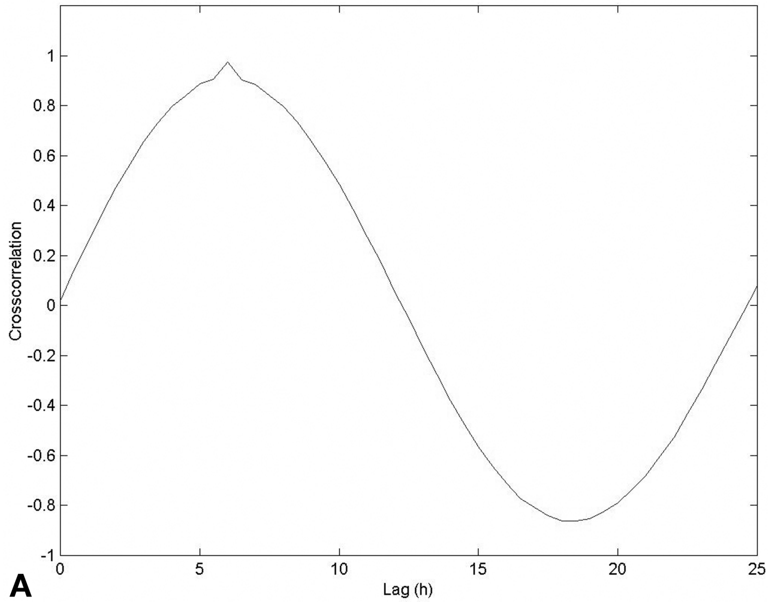
3.3. Use of Cross-Correlation Analysis to Estimate Phase and Periodicity in Common

Phase was assessed in data SET IV by running the program CROSSCO. This program uses a reference data set of known phase and amplitude as a base to measure the phase of an experimental vector. The algorithm is identical to standard autocorrelation with the exception that two data sets are being compared, instead of a single data set being compared with itself *(7)*. If the data are in phase, there will be a peak at lag zero, although it will likely not be unity, as is always seen in autocorrelograms, given that the two sets will differ. Displacement of the first peak from lag zero in either direction is a direct measure of the phase difference between the reference and experimental periodicities (*see* **Note 12**).

Figure 9 depicts the output of CROSSCO, comparing a reference data set with phase defined as zero and a data set that is phase-delayed by 6 h. All data had 80% noise added. The peak in correlation occurs at the lag corresponding to a 6-h difference. This analysis also serves to illustrate what two rhythms with a periodicity in common would look like when analyzed in this manner. The robust crosscorrelation, even in the presence of considerable noise, indicates the periods of the two data sets are close. (*See* **Note 12)**.

4. Notes

1. MATLAB is a numeric computation software package offered by The Math Works in Natick, MA (www.mathworks.com). It can be configured to compile and run software written in other languages—in this case, FORTRAN. Other



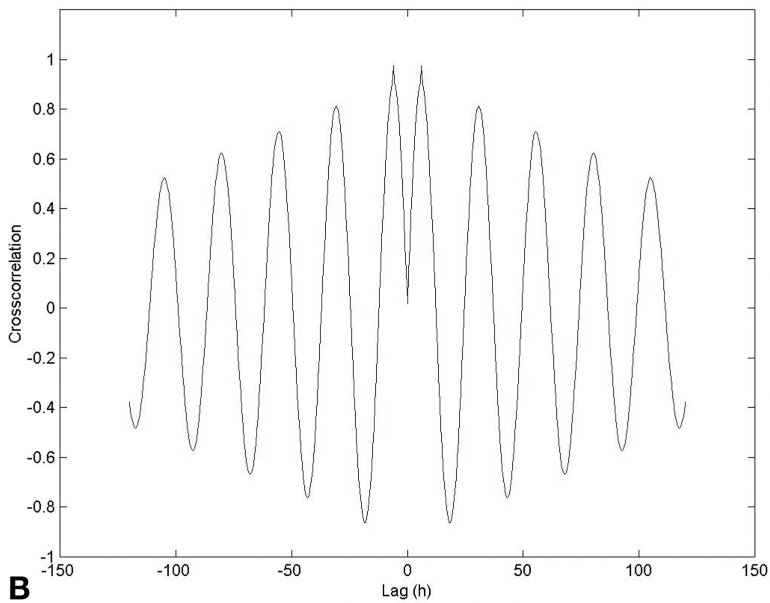


Fig. 9. (**A**) A short segment of the crosscorrelation of two data vectors differing in phase by 6 h. The periods in the two data sets are the same and both have 80% noise added. (**B**) The entire crosscorrelation shown to emphasize the periods held in common between the two vectors. If the periods were different, there would be a concomitant decay in the envelope of the function depending on the difference in the periods.

systems have this capability, and the author and others have converted the programs here to run in these systems ***(8)***.

2. The programs employed here were run in their DOS window executable file format, as it was desired to maintain the maximum flexibility in producing the examples. In general, this would not be necessary for normal work. Except as noted, all programs were written in FORTRAN.
3. MESA has proven its worth in analysis of biological time series over a period of 20 yr. It was instrumental in uncovering ultradian periodicities in the behavioral rhythms in strains of *D. melanogaster* having no circadian rhythmicity, such as *period^{01}* ***(17,18)***. MESA was set to a very high resolution here, which might not be necessary for normal usage.
4. FILCON: This program takes the discrete Fourier transform of the data and may then be directed to zero out coefficients in a frequency range to be eliminated. Owing to the great sensitivity of MESA, FILCON was not always necessary here to demonstrate the periodicities, but in practice, the author has found it essential as a first step when trends or strong confounding rhythms are present.
5. The programs used to extract the spectral peaks and the RIs were written in Turbo Basic, but could be implemented in several languages. They involve simple bubble sort analysis at the core. In this algorithm, values from a set are ordered in a column by magnitude. The RI program incorporates criteria to ensure that the proper peak is reported. If the autocorrelation function is sufficiently weak, the program reports out arrhythmicity.
6. Heartbeat of *Drosophila* is in the range of 1 to 4 Hz and is monitored in several ways, including optically. The data are best presented as frequency rather than period ***(9–11)***.
7. Noise was added by incorporating the output of a white noise generator. The noise file is simply added to the output of the signal generating function in a proportion that reflects the percentage and reflects the signal-to-noise ratio.
8. The determination of significance in rhythmicity may be based on a number of methods. MESA lacks any way of inherently testing for significance of the peaks, as would be possible with the Fourier transform, but the strength of the system described here is that it uses an entirely different algorithm, the autocorrelation function, to assess significance. One may calculate a 95% confidence limit to apply to peaks in the autocorrelation functions, namely $2/\sqrt{N}$; however, it is common simply to look for regularly recurring peaks in the correlogram to determine if a genuine periodicity is present ***(7)***.
9. The author uses a batch plotting routine written specifically in MATLAB that accommodates this format for heartbeat (frequency output) data, but currently has none for circadian rhythms (period output).
10. The Butterworth is a recursive filter that can be configured in a high- or low- pass form. The one used here was two-pole low-pass with a 3-db cutoff period of 4 h at a sampling rate of two per hour ***(16)***. Recursive filters use a combination of raw and previously filtered data in computing the output. Two pole filters have three coefficients in the formula. A 3-db cutoff means a power reduction in the

signal of 50% at the transition period, here 4 h. None of the output shown in the figures had been filtered first, owing to MESA's power, but normally the author looks at data both with and without filters, and actual rhythm data commonly are improved greatly by the process. This filter will induce an approx 4-h phase shift. If this is a problem, one simply runs the filter twice, sending the filtered data set through in reverse order to cancel out the shift ***(13)***.

11. The sensitivity and resolution of MESA, coupled with the ability of FILCON to remove strong circadian rhythmicity, was essential to uncovering a cirhoral rhythm in human core body temperature. Other methods, including fast Fourier transform, failed to detect it despite the periodicity being clearly visible in the raw data plots ***(14)***.
12. When assessing phase in this manner with actual biological data, the test data set must have the same period as the experimental set. The periods of each experimental series are first estimated with MESA, and test sets of the same period are created using SIGGEN. The program automatically adjusts and normalizes amplitude before crosscorrelation is estimated. This method is particularly useful for very noisy and irregular data sets because it estimates phase based on the entire signal rather than just one identifiable phase marker, which by itself may not be reliable from cycle to cycle.
13. All programs written by the author are available free of charge by e-mail or FTP either as FORTRAN source code, executable files, or as files executable from MATLAB.

References

1. Dowse, H. B., and Ringo, J. M. (1989) The search for hidden periodicities in biological time series revisited. *J. Theoret. Biol.* **139,** 487–515.
2. Dowse, H. B., and Ringo, J. M. (1991) Comparisons between "periodograms" and spectral analysis: apples are apples after all. *J. Theoret. Biol.* **148,** 139–144.
3. Burg, J. P. (1978) Maximum entropy spectral analysis. In: *Modern Spectrum Analysis* (Childers, D.G., ed.), Wiley, New York.
4. Burg, J. P. (1978) A new analysis technique for time series data. In: *Modern Spectrum Analysis* (Childers, D.G., ed.), Wiley, New York.
5. Ulrych, T., and Bishop, T. (1975) Maximum entropy spectral analysis and autoregressive decomposition. *Rev. Geophys. Space Physics* **13,** 183–300.
6. Ables, J. G. (1974) Maxiumum entropy spectral analysis. *Astron. Astrophys. Suppl. Series* **15,** 383–393.
7. Chatfield, C. (1989) *The Analysis of Time Series: An Introduction*. Chapman and Hall, London.
8. Levine, J., Funes, P., Dowse, H., and Hall, J. (2002) Signal analysis of behavioral and molecular cycles. *Biomed. Central. Neurosci.* **3,** 1.
9. Dowse, H. B., Ringo, J. M., Power, J., Johnson, E., Kinney, K., and White, L. (1995) A congenital heart defect in *Drosophila* caused by an action potential mutation. *J. Neurogenet.* **10,** 153–168.

10. Johnson, E., Ringo, J., and Dowse, H. (1997) Modulation of *Drosophila* heartbeat by neurotransmitters. *J. Comp. Physiol. B.* **167,** 89–97.
11. Johnson, E., Ringo, J., Bray, N., and Dowse, H. (1998) Genetic and pharmacological identification of ion channels central to *Drosophila*'s cardiac pacemaker. *J. Neurogenet.* **12,** 1–24.
12. Krishnan, B., Levine, J., Sisson, K., et al. (2001) A new role for cryptochrome in a *Drosophila* circadian oscillator. *Nature* **411,** 313–317.
13. Levine, J., Funes, P., Dowse, H., and Hall, J. (2002) Advanced analysis of a *cryptochrome* mutations's effects on the robustness and phase of molecular cycles in isolated peripheral tissues of *Drosophila. Biomed. Central Neurosci.* **3,** 5.
14. Lindsley, G., Dowse, H., Burgoon, P., Kilka, M., and Stephenson, L. (1999) A persistent circhoral ultradian rhythm is identified in human core temperature. *Chronobiology Intl.* **16,** 69–78.
15. Dowse, H. B., and Ringo, J. M. (1994) Summing locomotor activity into "bins": How to avoid artifact in spectral analysis. *Biol. Rhythm Res.* **25,** 2–14.
16. Hamming, R. W. (1983) *Digital Filters.* Prentice-Hall, London.
17. Dowse, H. B., Hall, J. C.,and Ringo, J. M. (1987) Circadian and ultradian rhythms in *period* mutants of *Drosophila melanogaster*. *Behav. Genet.* **17,** 19–35.
18. Dowse, H. B., Dushay, M. S., Hall, J. C., and Ringo, J. M. (1989) High resolution analysis of locomotor activity rhythms in *disconnected*, a visual system mutant of *Drosophila melanogaster*. *Behav. Genet.* **19,** 529–542.

http://www.springer.com/978-1-58829-417-3

Circadian Rhythms
Methods and Protocols
Rosato, E. (Ed.)
2007, XIV, 578 p. 115 illus., Hardcover
ISBN: 978-1-58829-417-3
A product of Humana Press