Listy Biometryczne - Biometrical Letters
Vol. 34 (1997), No. 1, 1-8

Detecting synchronous rhythms

Erhard Reschenhofer

Department of Statistics, Operations Research and Computer Science,
University of Vienna, Universitätsstr 5, A-1010 Vienna, Austria

SUMMARY

In a previous paper of the author (Reschenhofer, 1995), a simple method for the detection of a hidden periodicity in a set of biological time series has been proposed which requires some prior information about the frequency and the phase of the suspected periodicity. The present note propagates a similar, but more flexible method which gets along without any prior information about the phase.

KEY WORDS: chronobiology, mixed spectrum, periodogram.

1. Introduction

Many biological time series can be adequately described as a sum of a highly autocorrelated series with a purely continuous spectrum and a linear combination of sine and cosine waves (with random amplitudes and typically with periods of one year, one month or one day caused by seasonal variations, lunar phases, light intensity, etc.) which has a purely discrete spectrum. The main problem in analyzing such a time series with a mixed spectrum is to separate the continuous spectral components (in the language of communications engineering the "noise") from the discrete spectral components (the "signal"). As is well known (see, e.g., Bartlett, 1957), given only a finite sample and no further information it is impossible to distinguish discrete components from very narrow peaks in the continuous spectrum. Hence all standard procedures for the detection of discrete components rely on certain hardly verifiable assumptions concerning the bandwidth of the continuous spectrum (see Priestley, 1981, chapter 8). An alternative approach is to examine only the phases of the suspected periodic terms, i.e., the ratios of the amplitudes of the sine waves and the amplitudes of the corresponding cosine waves. For the case where one is able to make a guess of the frequency and the phase of a possible discrete component, Reschenhofer (1995) proposed to reject the hypothesis of a purely continuous spectrum whenever a sinusoid with

the specified frequency and phase fits the data much better than another sinusoid with the same frequency and a phase which differs from the specified phase by $\pi/2$. This test procedure can easily be extended to the case where several replicate series are available (which are obtained from different subjects) and it then allows to detect a discrete component with probability one as the number of replicate series goes to infinity (provided that the initial guess of the phase does not differ by more than $\pi/4$ from the true phase).

However, in finite samples the power of this test procedure depends crucially on the goodness of the initial guess of the phase. Since the prior specification of the phase is often a delicate problem, an alternative method is proposed in the next section which is easier to use since it does not require any prior information about the phase. Moreover, the analysis of a typical chronobiological data set in section 3 shows that even in those cases where such prior information is available, the new method may still be competitive.

2. The distribution of the phase estimates

We assume that the observations $x_1, ..., x_n$ can be described by the equation

$$x_t = \mu + A\cos(\omega_o t + \varphi) + z_t, \qquad t = 1, ..., n,$$

where $0 < \omega_o < \pi$ is a known constant, μ and A are unknown parameters, φ is a random variable with a uniform distribution on the interval $(-\pi, \pi)$, and (z_t) is a zero-mean stationary process with a purely continuous spectrum. The phase φ is treated as a random variable rather than as a constant in order to ensure that the series (x_t) is stationary. However, for any particular realization of the series, φ is a constant and hence may be estimated. The null-hypothesis to be tested is that the spectrum of the series (x_t) does not contain a discrete component, i.e., $H_o : A = 0$. To obtain estimates for A and φ we

(i) rewrite the above equation as

$$x_t = \mu + \cos(\omega_o t) + b\sin(\omega_o t) + z_t,$$

where

$$a = A\cos(\varphi) \quad \text{and} \quad b = -A\sin(\varphi),$$

(ii) estimate a and b in the usual way by

$$\hat{a} = \frac{2}{n}\sum x_t \cos(\omega_o t) \quad \text{and} \quad \hat{b} = \frac{2}{n}\sum x_t \sin(\omega_o t),$$

and

(iii) solve the equations

$$\hat{a} = \hat{A}\cos(\varphi) \quad \text{and} \quad \hat{b} = -\hat{A}\sin(\hat{\varphi})$$

for $\hat{A}$ and $\hat{\varphi}$.

We have:

$$\hat{A} = (\hat{a}^2 + \hat{b}^2)^{1/2}$$

and

$$\hat{\varphi} = \begin{cases} \arctan(-\hat{b}/\hat{a}) & a > 0 \\ \arctan(-\hat{b}/\hat{a}) - \pi & a < 0, \quad b > 0 \\ \arctan(-\hat{b}/\hat{a}) + \pi & a < 0, \quad b \leq 0 \\ -\pi/2 & a = 0, \quad b > 0 \\ \pi/2 & a = 0, \quad b < 0 \\ \text{arbitrary} & a = 0, \quad b = 0 \end{cases}$$

If ω_o is a Fourier-frequency, i.e., if $\omega_0 = 2\pi j/n$, $\hat{a}$ and $\hat{b}$ are just the least squares estimates of the parameters a and b from the regression

$$x_t = \mu + a\cos(\omega_0 t) + b\sin(\omega_0 t) + \text{error}.$$

If, in addition, (z_t) is Gaussian white noise with zero mean and variance σ_z^2, $\hat{a}$ and $\hat{b}$ are also normally distributed under H_0 (with mean zero and variance $\sigma^2 = \sigma_z^2 n/2$). Using the orthogonality relation

$$\sum_{t=1}^{n} \sin(2\pi jt/n)\cos(2\pi jt/n) = 0$$

we have

$$\operatorname{cov}(\hat{a}, \hat{b}) = 0.$$

Hence the joint density of $\hat{a}$ and $\hat{b}$ is given by

$$(2\pi\sigma^2)^{-1}\exp[-(\hat{a}^2 + \hat{b}^2)/(2\sigma^2)]$$

and that of $\hat{A}$ and $\hat{\varphi}$ by

$$\begin{aligned} & (2\pi\sigma^2)^{-1}\exp[-\hat{A}^2/(2\sigma^2)] \begin{vmatrix} \cos(\hat{\varphi}) & -\hat{A}\sin(\hat{\varphi}) \\ \sin(\hat{\varphi}) & \hat{A}\cos(\hat{\varphi}) \end{vmatrix} \\ = \; & (2\pi\sigma^2)^{-1}\exp[-\hat{A}^2/(2\sigma^2)]\hat{A}, \quad 0 < \hat{A} < \infty, \quad -\pi < \hat{\varphi} < \pi. \end{aligned}$$

Integration with respect to $\hat{A}$ yields the marginal density for $\hat{\varphi}$ which does not depend on $\hat{\varphi}$. Thus $\hat{\varphi}$ is uniformly distributed on the interval $(-\pi, \pi)$.

For real data, the assumption that the z_t are serially uncorrelated is much too strong. Fortunately, the result that $\hat{\varphi}$ is uniformly distributed holds approximately also in a more general setting. This is due to the fact that under H_0 (and some standard assumptions, see, e.g., Priestley, 1981, p.397 and p.422),

$$\hat{a}\left(\frac{2}{n}\right)^{-1/2} \quad \text{and} \quad \hat{b}\left(\frac{2}{n}\right)^{-1/2}$$

asymptotically are i.i.d. $N(0, f_x(\omega_0)/(2\pi))$, where f_x is the spectral density of (x_t).

In situations where several replicate series are available the problem of testing for a jump in the spectral distribution function at a fixed frequency ω_0 can therefore be reduced to that of testing the uniformity of the phase estimates obtained from the different series. The standard test for the latter problem is the Kolmogorov-Smirnov test. This test will be very powerful if the phase estimates cluster around a single value. But since it is well known that the Kolmogorov-Smirnov test may have extremely low power in case of multimodality, alternative goodness-of-fit tests should be used in this case (see Reschenhofer and Bomze, 1991). For example, multimodality will typically occur if shift-workers belong to the group of test persons from which the different time series are obtained.

3. Application to real data

To examine the fluctuations of the hormone gastrin over time, blood samples were taken every 4 hours over three days from 15 healthy test persons. For each person, the first blood sample was taken at 6:00 and the last one 72 hours later. Thus $n = 19$. Graphs of the observed time series are given in Figures 1 and 2. Since we suspect the existence of a periodic 24 hours pattern we apply the test proposed in the previous section to our data. One unit of time covers 4 hours, hence the period of 24 hours corresponds to the frequency $4(2\pi/24) = \pi/3$. The phase estimate $\hat{\varphi}$ corresponding to the frequency $\omega_0 = \pi/3$ is calculated for each of the 15 series of measurements. The results are given in Table 1. Under the null hypothesis that the time series do not contain a strictly periodic component, the theoretical distribution function of the phase estimates approximately rises linearly from 0 at $\varphi = -\pi$ to 1 at $\varphi = \pi$. However, for our data all estimates are positive. The value of the Kolmogorov-Smirnov test statistic, which is just the maximum deviation of the empirical distribution function from the theoretical one, must therefore be greater than 0.5. Consequently it exceeds the critical value $c_{0.01} = 0.404$ of the Kolmogorov-Smirnov one-sample test for $n = 15$. Thus the null hypothesis can be rejected at the 0.01 level of significance. Moreover, since the smallest phase estimate is 0.415 the value of the test statistic cannot be

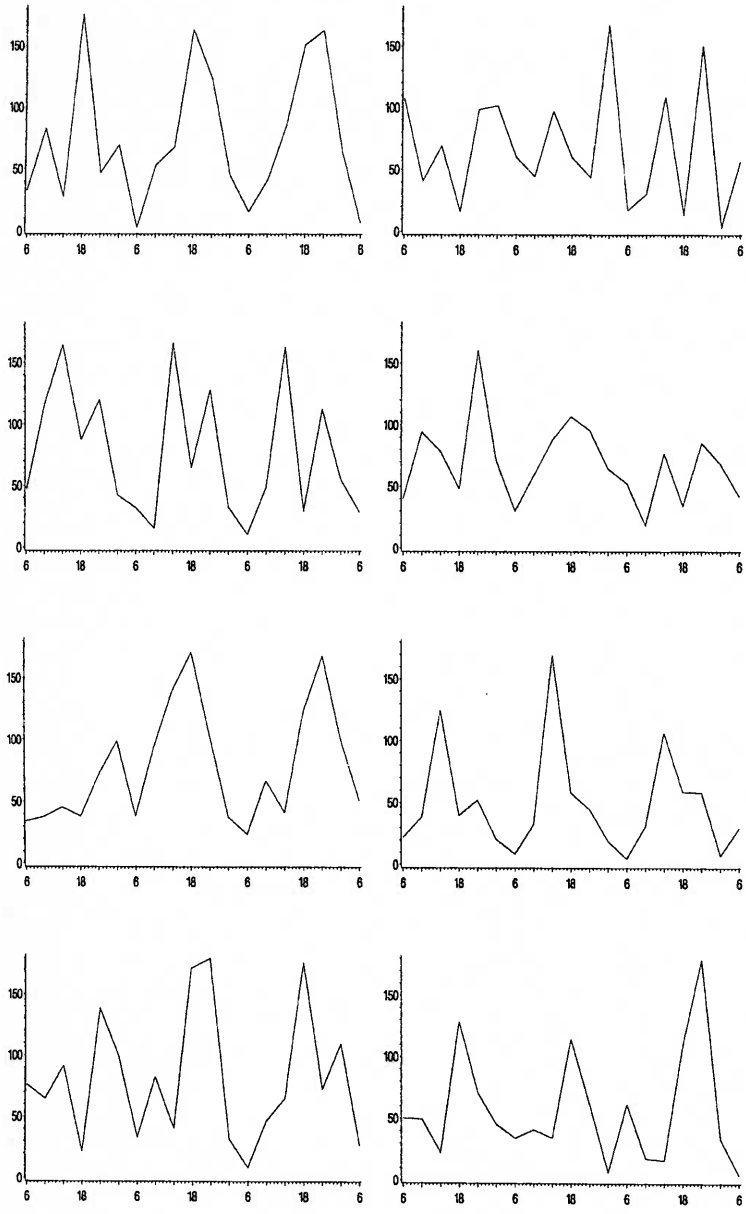

Fig. 1. Graphs of the blood gastrin values for test persons 1-8

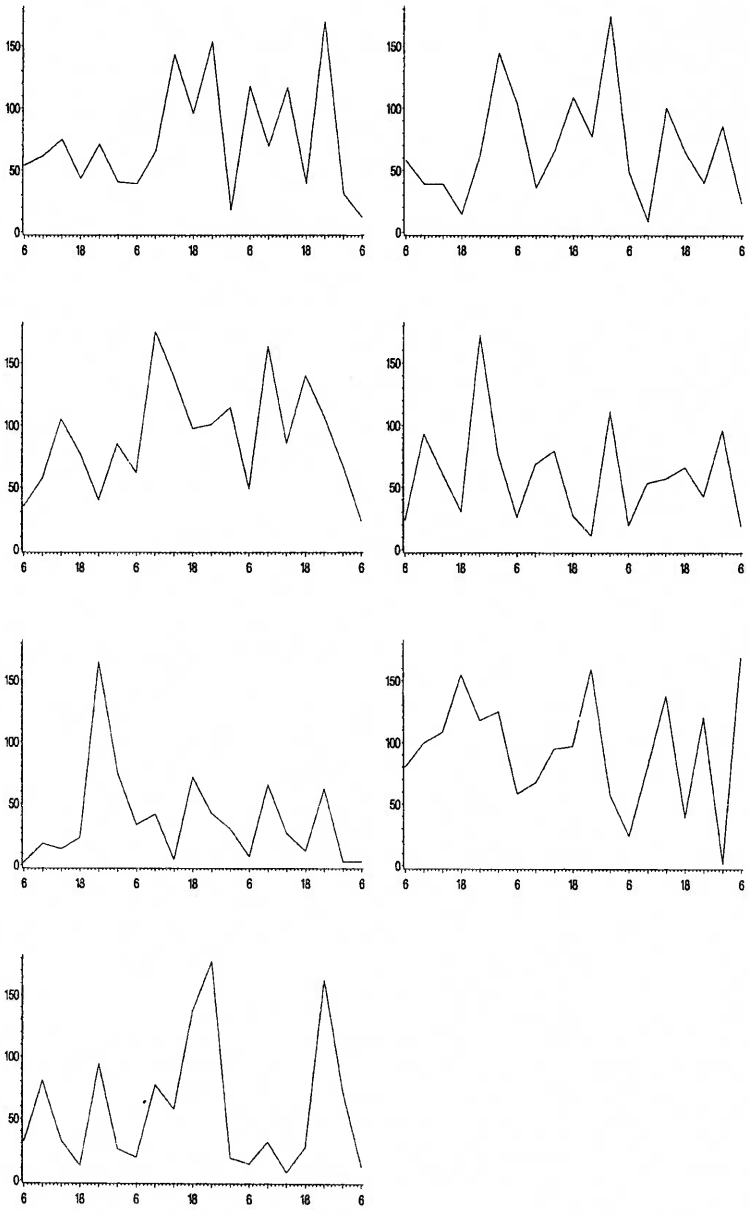

Fig. 2. Graphs of the blood gastrin values for test persons 9-15

smaller than $(\pi + 0.415)/(2\pi) \simeq 0.566$. This argument will still be valid if the test statistic is evaluated for an arbitrary subsample of replicate series. Noting that for $n = 5$ the critical value $c_{0.05}$ equals 0.563 we see that our test would also have been able to reject the null hypothesis (at the 0.05 level) for any subsample of size $n \geq 5$.

Table 1. Estimated amplitudes and phases for the fifteen gastrin series

$\hat{A}$	$\hat{\varphi}$
8.936	2.254
19.582	0.415
9.356	2.843
4.796	1.135
11.946	1.172
6.067	2.118
14.531	1.394
49.128	1.869
6.176	0.562
23.085	2.462
10.686	1.577
10.426	1.714
35.199	2.743
28.199	1.609
12.696	1.596

Based on prior information about the phase which originates from earlier studies (the results of Ganguli and Forrester, 1972; Moore and Wolfe, 1974; Tarquini et al., 1984, suggest that the highest gastrin values occur in the evening), the test proposed by Reschenhofer (1995) has been applied to the same data set. Interestingly, using the whole sample of 19 series this test yields a result which is just significant at the 5% level. This lower performance is mainly due to the large variability of the phase estimates. Thus we conclude that in view of the large interindividual differences typically occurring in biological and medical investigations the test considered in this paper is competitive even in situations where prior information about the phase is available.

Acknowledgements

I wish to thank Werner Ploberger and an anonymous referee for helpful comments, Christian Ivad for graphical programming and Wolfgang Marktl for leaving the data to me.

References

Bartlett, M.S. (1957). Discussion on 'Symposium on spectral approach to time series'. *J. Roy. Statist. Soc. Ser. B* **19**, 1-63.

Ganguli, P.C. and Forrester, J.M. (1972). Circadian rhythm in plasma level of gastrin. *Nature* **236**, 127-128.

Moore, J.G. and Wolfe, M. (1974). Circadian plasma gastrin patterns in feeding and fasting man. *Digestion* **11**, 226-231.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*, Academic Press, London.

Reschenhofer, E. (1995). Use of phase for the detection of hidden periodicities. *Biometrical Journal* **37**, 957-964.

Reschenhofer, E. and Bomze, I.M. (1991). Length tests for goodness of fit. *Biometrika* **78**, 207-216.

Tarquini, B., Benvenuti, M., Cavallini, V., Cecch Ettin, M., Romano, S. and Sensi, G. (1984). Pepsinogen gastrin mesor ratio: a potential simple marker for gastric secretory function. *Chronobiologia* **11**, 11-17.

Received 22 October 1996; revised 2 April 1997

Wykrywanie synchronicznych rytmów

Streszczenie

W jednej z poprzednich prac autor zaproponował prostą metodę wykrywania ukrytej okresowości w zbiorze szeregów czasowych. Metoda ta wymaga pewnej informacji a priori dotyczącej częstotliwości i fazy tej okresowości. Obecna praca podaje podobną, lecz bardziej uniwersalną metodę, nie wymagającą żadnej informacji a priori o fazie.

Słowa kluczowe: chronobiologia, periodogram, spektrum mieszane.

References

Bartlett, M.S. (1957). Discussion on 'Symposium on spectral approach to time series'. *J. Roy. Statist. Soc. Ser. B* **19**, 1-63.

Ganguli, P.C. and Forrester, J.M. (1972). Circadian rhythm in plasma level of gastrin. *Nature* **236**, 127-128.

Moore, J.G. and Wolfe, M. (1974). Circadian plasma gastrin patterns in feeding and fasting man. *Digestion* **11**, 226-231.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*, Academic Press, London.

Reschenhofer, E. (1995). Use of phase for the detection of hidden periodicities. *Biometrical Journal* **37**, 957-964.

Reschenhofer, E. and Bomze, I.M. (1991). Length tests for goodness of fit. *Biometrika* **78**, 207-216.

Tarquini, B., Benvenuti, M., Cavallini, V., Cecch Ettin, M., Romano, S. and Sensi, G. (1984). Pepsinogen gastrin mesor ratio: a potential simple marker for gastric secretory function. *Chronobiologia* **11**, 11-17.

Received 22 October 1996; revised 2 April 1997

Wykrywanie synchronicznych rytmów

Streszczenie

W jednej z poprzednich prac autor zaproponował prostą metodę wykrywania ukrytej okresowości w zbiorze szeregów czasowych. Metoda ta wymaga pewnej informacji a priori dotyczącej częstotliwości i fazy tej okresowości. Obecna praca podaje podobną, lecz bardziej uniwersalną metodę, nie wymagającą żadnej informacji a priori o fazie.

Słowa kluczowe: chronobiologia, periodogram, spektrum mieszane.