Astron. Astrophys. 81, 344–346 (1980)

ASTRONOMY AND ASTROPHYSICS

Research Note

Accuracy of the Maximum Entropy Method

F. M. Strauss*

Instituto de Física, Universidade Federal do Rio Grande do Sul, 90000 Porto Alegre, RS, Brasil

Received April 10; revised June 26, 1979

Summary. Although the maximum entropy method is capable of giving the period of a pure sine wave with very few data points, it is very sensitive to noise. The half power width of a spectral peak obtained by this method is proportional to $(\sigma_n/A)^2$, where σ_n is the standard deviation of the noise and A is the semiamplitude of the light curve. More serious are the consequences of a systematic shift of the spectral peak, by an amount proportional to the same factor. A simple autocorrelation method that is free of systematic errors is also discussed.

Key words: variable stars – maximum entropy method – autocorrelation

I. Introduction

The maximum entropy (ME) method of spectral analysis was recently developed by Burg (for a detailed discussion see Smylie et al., 1973). Although originally the method was derived, as the name indicates, from considerations of Information Theory, it has been shown by Van den Bos (1971) to be equivalent to obtain from the data, by least squres, the coefficients of a prediction filter. Such a filter has the property that given a time series $\{y_0, y_1, y_2, \ldots, y_{N-1}\}$ with uniform spacing Δt, any y_j is a linear combination of the M preceding values:

$$y_j = \sum_{k=1}^{M} g_k y_{j-k} + a_j, \tag{1}$$

where $\{g_1, g_2, \ldots, g_M\}$ are the coefficients of a prediction filter of order M and a_j represents the prediction error. The *ME* power spectrum is given by

$$|\tilde{y}(f)|^2 \propto \left|1 - \sum_{k=1}^{M} g_k \exp(-2\pi i f k \Delta t)\right|^{-2}. \tag{2}$$

The coefficients of the prediction filter can be found by minimizing the mean squared prediction error by the method of least squares. The resulting normal equations have as coefficients the values of the autocorrelation function which may be estimated from

$$\phi_i = \frac{1}{(N-M)} \sum_{j=M}^{N-1} y_j y_{j-1}, \qquad i = 0, 1, \ldots, M. \tag{3}$$

* Present address: Dr. Federico M. Strauss, CRAAM/ON/CNPq, Rua Pará 277, 01243 São Paulo, SP, Brasil

II. The Case of a Pure Sine Wave with Noise

The order M of the filter depends on the complexity of the data. For instance, a simple sine wave

$$y_j = A \cos(2\pi f_0 j \Delta t + \psi) + n_j \tag{4}$$

(where A, f_0, ψ, and n_j stand for semi-amplitude, frequency, initial phase and noise) can be predicted by a second order filter with coefficients

$$g_1 = 2\cos(2\pi f_0 \Delta t), \qquad g_2 = -1. \tag{5}$$

The presence of more than one sine wave (including harmonics) requires proportionally more terms. The $M=2$ case is sufficient for many astronomical applications; in geophysics however higher orders are necessary, for instance in the analysis of polar motions or seismic data. This leads to large systems of normal equations, which may be solved by Burg's recursive algorithm (Smylie et al., 1973).

To find the effect of noise on the ME power spectrum in the $M=2$ case, we may use the autocorrelation function of the time series defined in Eq. (4), which has the theoretical values

$$\begin{aligned} \phi_0 &= \tfrac{1}{2}A^2 + \sigma_n^2, \\ \phi_1 &= \tfrac{1}{2}A^2 \cos(2\pi f_0 \Delta t), \\ \phi_2 &= \tfrac{1}{2}A^2 \cos(4\pi f_0 \Delta t), \end{aligned} \tag{6}$$

where σ_n is the standard deviation of the noise n_j of Eq. (4). The normal equations are (Smylie et al., 1973):

$$\begin{aligned} \phi_0 g_1 + \phi_1 g_2 &= \phi_1 \\ \phi_1 g_1 + \phi_0 g_2 &= \phi_2 . \end{aligned} \tag{7}$$

Their solution is

$$g_1 = \phi_1(\phi_0 - \phi_2)/D, \qquad g_2 = (\phi_0\phi_2 - \phi_1^2)/D, \qquad D = \phi_0^2 - \phi_1^2. \tag{8}$$

Inserting Eqs. (6) into (8), we obtain in the low noise limit, or more precisely when

$$\sigma_n^2/A^2 \ll \sin^2(2\pi f_0 \Delta t) \tag{9}$$

the coefficients

$$\begin{aligned} g_1 &\simeq 2\cos(2\pi f_0 \Delta t) - \frac{4\cos(2\pi f_0 \Delta t)}{\sin^2(2\pi f_0 \Delta t)} \frac{\sigma_n^2}{A^2} \\ g_2 &\simeq -1 - \left[4 + \frac{2}{\sin^2(2\pi f_0 \Delta t)}\right] \frac{\sigma_n^2}{A^2} \end{aligned} \tag{10}$$

which differ from the exact values given in Eq. (5) by amounts proportional to the ratio $(\sigma_n/A)^2$. This will produce a broadening,

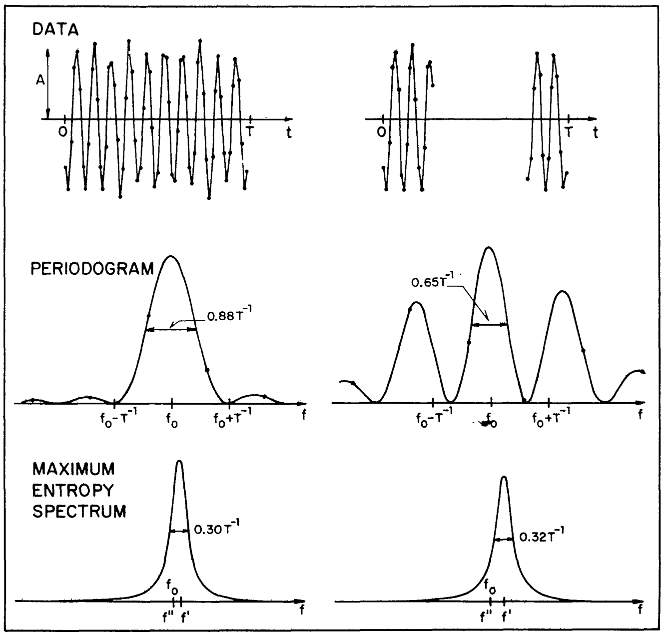

Fig. 1. *Left:* simulated time series consisting of 64 samples of a sine wave of frequency f_0 in 10% noise, as given by Eq. (4), and a portion of its unsmoothed periodogram (the filled circles represent the result of a fast Fourier transform). The maximum entropy spectrum has its peak shifted to f', while f'' represents the estimate obtained from Eq. (13). *Right:* same as above, with data reduced to 32 points with a large eccentric gap

and, worse, a shift of the ME power spectrum peak. Inserting Eqs. (10) into (2), after some manipulation, we obtain a half power width of

$$\Delta f_{1/2} \simeq (\sigma_n/A)^2 (\pi \Delta t)^{-1} [1 + 3\cot^2(2\pi f_0 \Delta t)] \tag{11}$$

and a shift of

$$f' - f_0 \simeq (\sigma_n/A)^2 (\pi \Delta t)^{-1} \cot(2\pi f_0 \Delta t). \tag{12}$$

These expressions are not valid near $f_0 = 0$ and $f_0 = f_N = (2\Delta t)^{-1}$ because of Eq. (9). They do not depend on the length of the record because we used the exact expressions (6). In practice some additional error will be introduced by the estimation of the autocorrelation function from a finite record. As examples, Fig. 1 shows two simulated light curves with a signal to noise ratio $A/\sigma_n = 10$ and their respective periodograms and ME power spectra. While the resolution of the latter is somewhat better, the shift of the peak is readily apparent even for this relatively small amount of noise. The integral of the ME spectrum, which is proportional to A^2, is not greatly affected.

III. A Simplified Autocorrelation Method

In the approach of Blackman and Tukey (1958), the power spectrum is obtained from a Fourier transform of the tapered autocorrelation function. When a single periodicity is present in the data, a short-cut method can be used that gives directly the value of f_0 without having to calculate the complete autocorrelation function. This is based on the three Eqs. (6), which may be solved for A, σ_n, and f_0, as functions of the values of ϕ_0, ϕ_1, and ϕ_2 given by Eq. (3). The solutions ($\pm$ their errors) are:

$$\begin{aligned} A &= [(\phi_2^2 + 8\phi_1^2)^{1/2} - \phi_2]^{1/2} \pm (2/N)^{1/2} \sigma_n \\ \sigma_n &= (\phi_0 - \tfrac{1}{2}A^2)^{1/2} \pm N^{-1/2} \sigma_n \\ f'' &= (2\pi \Delta t)^{-1} \operatorname{arcos}(2\phi_1/A^2) \\ &\quad \pm (2/N)^{1/2} [\pi \Delta t \sin(2\pi f_0 \Delta t)]^{-1} (\sigma_n/A)^2 , \end{aligned} \tag{13}$$

where f'' is a better estimate of f_0 than the peak of a ME spectrum. The noise contained in the data causes an uncertainty in the estimate of the autocorrelation function which propagates to the result giving the standard deviations indicated in Eqs. (13), obtained with the theory of propagation of errors (e.g., Bevington, 1969), for $N \gg 1$. For the example shown on the left side of Fig. 1, the uncertainties in A and σ_n are less than 2%. The error in the frequency is still proportional to $(\sigma_n/A)^2$ as in the ME method, but is no longer systematic and is too small to be noticeable in Fig. 1.

IV. Discussion

The ME method is a new, high resolution technique that provides estimates of power spectra even for short time series. It has been applied widely to geophysical problems and also, to a lesser extent, to astrophysics (e.g., Richer and Ulrych, 1974).

We may compare the accuracy of the ME method with that of the periodogram. Because of the finite length of the time series, the periodogram is the convolution of the Fourier transform of the light curve with that of the time window $W(t)$ (Gray and Desikachary, 1973). For uninterrupted observation during a span T, the amplitude of the Fourier transform of $W(t)$ is $|\tilde{W}(f)| = |\sin(\pi f T)/(\pi f)|$. Thus, if $y(t)$ contains only one frequency f_0 (a more general case is not considered here), the periodogram will be proportional to $|\tilde{W}(f-f_0) + \tilde{W}(f+f_0)|^2$, consisting of two peaks (at $f = \pm f_0$) of half power widths $\Delta f_{1/2} \simeq 0.886\, T^{-1}$. The interference between each peak and the sidebands of the other, and the presence of noise, distort slightly the spectrum causing small shifts of the maxima. The existence of gaps in the data (when several nights are combined) complicates the shape of the time window $W(t)$ and markedly increases the power in the sidebands of $\tilde{W}(f)$, as can be seen in Fig. 1.

In principle, the finite extent T of the data is of no consequence with the ME method, which uses the coefficients of a prediction filter. These contain the information necessary to extend the time series indefinitely in both directions, as well as filling any gap. The drawback is the strong sensitivity to noise, which causes errors in the calculated coefficients of the prediction filter, thus affecting non-linearly the spectrum through Eq. (2). Whereas the resulting broadening [Eq. (11)] is mostly due to the noise, the frequency shift [Eq. (12)] is mostly due to the non-linear interaction between the peaks located (for real data) at $f = \pm f_0$. Equations (11)–(13) show that errors are minimized if the sampling rate is such that $f_0 \simeq (4\Delta t)^{-1}$ (i.e., one half of the Nyquist frequency f_N), where distortion disappears. In the absence of noise both the ME and the correlation method give exact results, independently of T, with $N \geqq 3$.

It is easy to show that for complex data, the corresponding first order complex prediction filter produces unshifted spectra, since only one peak appears in the range $-f_N < f < f_N$. This is the reason why the analysis of Lacoss (1971), based on a complex correlation matrix, resulted in a width $f_{1/2}$ independent of f_0 and failed to reveal the shift, although his numerical experiments showed both effects. These problems are greatly reduced for a periodogram, where the noise power is distributed throughout the whole spectrum so that only a small fraction of the noise affects each peak; also, the effects of interference between peaks are less severe.

With respect to computational speed, the periodogram was about five times slower than the ME method in the examples of Fig. 1, but much more accurate (the frequency shift of the peak was 20 times smaller). The autocorrelation scheme of Section III was several hundred times faster than the periodogram, only slightly less accurate, and may even run on a small calculator. There is little advantage in using the Fast Fourier Transform (FFT) algorithm (Cooley and Tukey, 1965), because of its insufficient resolution: for instance, the 64 data points of Fig. 1 produce a 32 point FFT ($0<f<f_N$) shown as full circles in the periodograms. To obtain enough intermediate spectral points with the FFT one can extend the data with zeros to complete at least 1024 points before transforming, although losing thereby some of the advantage in computing time.

Acknowledgements. This work was partially supported by the Brazilian institutions CNPq and FINEP. Computations were performed at the Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul. We thank an anonymous referee for useful comments.

References

Bevington, P.R.: 1969, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York
Blackman, R.B., Tukey, J.W.: 1958, The Measurement of Power Spectra, Dover, New York
Cooley, J.W., Tukey, J.N.: 1965, *Math. of Comput.* **19**, 297
Gray, D.F., Desikachary, K.: 1973, *Astrophys. J.* **181**, 523
Lacoss, R.T.: 1971, *Geophysics* **36**, 661
Richer, H.B., Ulrych, T.J.: 1974, *Astrophys. J.* **192**, 719
Smylie, D.E., Clarke, G.K.C., Ulrych, T.J.: 1973, *Methods in Computational Physics* **13**, 391
Van den Bos, A.: 1971, IEEE *Trans.* IT-**17**, 493