INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392-I

H4.SMR.203 - 14

" SPRING COLLEGE ON GEOMAGNETISM AND AERONOMY "

(2 - 27 March 1987)

" Maximum entropy power spectral estimation "

presented by :

P.F. FOUGERE
Ionospheric Physics Division
Department of Air Force
Air Force Geophysics Laboratory
Air Force Base
Hanscom, Mass. 01731
U.S.A.

A COLLECTION OF REPRINTS ON

MAXIMUM ENTROPY POWER SPECTRAL

ESTIMATION

by

PAUL F. FOUGERE

HENRY R. RADOSKI

EDWARD J. ZAWALICK

AND CO-AUTHORS

AFCRL-TR-74-0088
AIR FORCE SURVEYS IN GEOPHYSICS, NO. 282

A High Resolution Power Spectral Estimate: The Maximum Entropy Method

HENRY R. RADOSKI
PAUL F. FOUGERE
EDWARD J. ZAWALICK

13 February 1974

SPACE PHYSICS LABORATORY PROJECT 8601
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
L. G. HANSCOM FIELD, BEDFORD, MASSACHUSETTS 01730

AIR FORCE SYSTEMS COMMAND, USAF

A High Resolution Power Spectral Estimate: The Maximum Entropy Method

I. INTRODUCTION

This paper discusses three methods of spectral analysis which we denote Blackman-Tukey, Levinson, and Burg. The first is an improvement on the usual Blackman-Tukey method[1]: (1) it provides a higher density of calculated points, and (2) it can be carried out to large lag numbers, that is, over 90% of the sample length. The application of this technique results in a smooth spectrum and higher resolution. The second type of spectrum is called a Levinson spectrum. This is a maximum entropy spectral estimate, calculated by using an algorithm due to Norman Levinson.[2] The Levinson spectrum provides an improvement in resolution over the modified Blackman-Tukey method. The third spectrum is also called a BPEC spectrum, an acronym formed from Burg Prediction Error Coefficients. This last spectrum is also a maximum entropy

(Received for publication 12 February 1974)

1. Blackman, R. B. and Tukey, J. W. (1958) The Measurement of Power Spectra, Dover Publications, Inc., New York, pp. 30-37.
2. Levinson, N. (1947) The Wiener RMS (root mean square) error criterion in filter design and prediction, J. Math. Phys. 25:261.

estimate but is calculated according to a technique developed by John Parker Burg.[3] The BPEC spectrum is an improvement over the Levinson spectrum and produces a sharp spectrum with high frequency resolution.

2. MAXIMUM ENTROPY METHOD

The maximum entropy spectral estimate is relatively new and was first presented by Burg.[4] It offers a novel technique of spectral estimation, which should be of considerable value not only for application to geomagnetic micropulsations and indices of magnetic activity but to any time series. At present, the maximum entropy method seems to be the best method of spectral analysis yet devised.

Entropy is a concept that is useful in statistical mechanics, where it is a measure of the randomness of a system or of the uncertainty of our knowledge about the state of a system. Entropy also arises in information theory where it measures the average information content in a message.[5] Problems in which the entropy of a distribution is maximized, subject to constraints, are well known. For example, if the constraint is that a distribution has a given mean square, the distribution with maximum entropy is a Gaussian. In statistical mechanics, this corresponds to a Maxwell-Boltzman distribution at a given temperature. A similar problem in the calculus of variations is solved by obtaining the maximum entropy spectrum. Here the constraint is that the power density spectrum must agree with Wiener's theorem: the Fourier transform of the spectral density is equal to the autocorrelation function of the observed time series. The entropy function to be maximized is the logarithm of the spectral density integrated over the frequency range. The specific form of the entropy function can be obtained from the change of entropy that occurs when a signal passes through a filter with a given frequency response. In this case, change in entropy is the integral of the logarithm of the square of the filter response.[6] The maximum entropy spectrum is formulated by determining the spectral density P(f) that maximizes the quantity

3. Burg, J. P. (1968) A new analysis technique for time series data. Paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, Netherlands.

4. _____ (1967) Maximum entropy spectral analysis. Paper presented at 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma.

5. Woodward, P. M. (1964) Probability and Information Theory, with Applications to Radar, Pergamon Press, Oxford, p. 49.

6. Shannon, C. E. and Weaver, W. (1964) The Mathematical Theory of Communication, The University of Illinois Press, Urbana, p. 93.

$$E = \int_{-f_o}^{+f_o} \ln P(f)\, df \tag{1}$$

subject to the constraints that the autocorrelation $C(\tau)$ is given by

$$C(\tau) = \int_{-f_o}^{+f_o} P(f) \cdot e^{i2\pi f\tau}\, df \tag{2}$$

where $f_o = (2\Delta t)^{-1}$ is the Nyquist frequency; Δt is the sampling time increment; $\tau = n\Delta t$, where $n = 0, 1, 2 \ldots N$; N measures the maximum lag time of the autocorrelation and the order of the spectrum.

A Blackman-Tukey power spectrum is determined by the Fourier transform of the autocorrelation

$$P(f) = \int_{-\infty}^{+\infty} C(\tau)\, e^{-i2\pi f\tau}\, d\tau . \tag{3}$$

In the maximum entropy method, the constraint is that the autocorrelation must be given by the Fourier transform of the power spectrum. Since the autocorrelation function and the power spectral density are Fourier transform pairs, the Blackman-Tukey and the maximum entropy spectrum would be identical for an infinitely long data sample. However, for finite data samples, they are often very different, and the difference favors the maximum entropy method by producing higher resolution spectra.

Maximizing the integrated logarithm of the spectral density is equivalent to constructing a filter which whitens the signal and maximizes the entropy change through the filter. In other words, the input signal is converted to noise. Ideally, the filter has removed all of the predictable portion of the signal. Therefore, it is not surprising that the maximum entropy filter is related to the unit time prediction filter and is, in fact, a unit time prediction error filter.[7] The output of this filter is not a prediction of the input signal but is the error between the known and the predicted signal. Because the filter

7. Peacock, K. L. and Treitel, S. (1969) Predictive deconvolution: theory and practice, Geophysics 34:155.

involves finding prediction coefficients or autoregressive coefficients, the spectrum also has been called an autoregressive spectral estimator.[8] If we know the coefficients of the whitening filter, then the spectrum is logically determined by the reciprocal of the square response of that filter. Heuristically, if the response of the filter is $G(\omega)$ and the amplitude spectrum of the signal is $X(\omega)$, then the whitening filter has the effect of

$$G(\omega) \cdot X(\omega) = C \tag{4}$$

where C is a constant noise amplitude. Therefore, the power spectral density of the signal is proportional to $X^2(\omega) = C^2/G^2(\omega)$, as seen in Eq. (5) below. The problem reduces to determining the set of prediction error coefficients. The coefficients are determined by a set of N+1 simultaneous equations for N+1 coefficients. This set of equations can be put in the form of a matrix equation involving the elements of the autocorrelation function in the matrix and the prediction error coefficients as a vector. The spectral density and the prediction error coefficients G_j are given by

$$P(f) = \frac{\Delta t\, P_{N+1}}{\left| \left(G_{k+1} Z^k\right) \right|^2} \tag{5}$$

$$\phi_{kj}\, G_{j+1} = P_{N+1}\, \delta_{k,0} \tag{6}$$

where $Z = e^{i2\pi f \Delta t}$; $G_j = 1$; $k = 0, 1, 2, \ldots, N$.[9] Note that the summation convention is used throughout the text: that is, the right side of Eq. (5) is summed from 0 to N. The matrix ϕ_{kj} is constructed from the autocorrelation function $C(\tau) = C(n\Delta t)$ by setting $n = k - j$ and $\phi_{kj} = C(|k - j|)$. P_{N+1} represents the mean output power of the prediction error filter containing N+1 coefficients.

Eq. (6) can be solved directly by inverting the autocorrelation matrix. However, orders of magnitude in time are saved by using Levinson's recursion relations between the coefficients for the matrices of order N and N+1. If $G_j^{(N)}$ is the jth coefficient obtained from a matrix of order N, then the coefficients of the next case are given by

8. Parzen, E. (1969) Multiple time series modeling in Multivariate Analysis-II, P.R. Krishnaiah, Ed., Academic Press, New York, p. 404.

9. Barnard, T.E. (1969) Analytical studies of techniques for the computation of high-resolution wavenumber spectra, Advanced Array Research, Special Report No. 9, Texas Instruments, Inc.

$$G_j^{(N+1)} = G_j^{(N)} + G_{N+1}^{(N+1)}\, G_{N+2-j}^{(N)} \tag{7}$$

where $j = 2, \ldots N$. Ultimately, only the last coefficient, $G_{N+1}^{(N+1)}$ need be specified; the rest are defined in terms of this single coefficient. The determination of this last coefficient greatly influences the resulting spectra. Currently, there are two methods for calculating this value. The Levinson method involves estimating the autocorrelation function. Poor estimates can lead to very poor spectra. The autocorrelation function we have used with success is a biased estimate which involves dividing the sum of the lag products by the total number of data points rather than by the number of lag products. These spectra provide an improvement of resolution over Blackman-Tukey spectra. In the Levinson method, the final coefficient is determined by

$$G_{N+1}^{(N+1)} = -\sum_{j=0}^{N-1} C(j+1)\, G_{N-j}^{(N)} \,/\, P_N . \tag{8}$$

The recursion relation for the mean error power is

$$P_{N+1} = P_N \left[1 - (G_{N+1}^{(N+1)})^2 \right] . \tag{9}$$

Burg's method of calculating the final prediction error coefficient has the advantage of not using any estimate of the autocorrelation function. This technique formally minimizes the error power between the actual signal and a prediction of the signal. The prediction error filter is applied in both a forward and a backward direction. The results of Eqs. (7) and (9) are used to calculate the spectral estimate as given in Eq. (5). The function to be minimized has the form

$$P_N = (2(M-N+1))^{-1} \left[\sum_{j=N}^{M} \left(\sum_{i=1}^{N} G_i X_{j-i+1} \right)^2 + \sum_{j=1}^{M-N+1} \left(\sum_{i=1}^{N} G_i X_{j+i-1} \right)^2 \right] \tag{10}$$

Burg's technique produces spectra of extremely high resolution and large dynamic range. The sharpness of the spectra may require some initial acclimatization after one has become accustomed to working with lower resolution spectra.

3. COMPARISON OF SPECTRA FOR A SYNTHETIC SIGNAL

This section contains examples of maximum entropy and Blackman-Tukey spectra to illustrate the enhanced resolution of the maximum entropy method. Other examples of maximum entropy spectra are given by Lacoss[10] and Ulrych.[11] Figure 1 shows a trace of the ideal signal which was analyzed. The signal was sampled at 257 equally spaced points and consisted of a sum of four sine waves of equal amplitude to which was added a small amount of white noise. In the case to be presented, the noise amplitude was one-half of the sine wave amplitude. This corresponds to a signal-to-noise ratio of 24. The contribution from each sine wave to the power is $S^2/2$, where S is the amplitude of the signal; the white noise contribution is $N^2/3$, where N is the amplitude of the noise. Thus, the signal-to-noise ratio is $6(S/N)^2$. Several other cases, whose results are available, considered the same signal and larger noise amplitudes down to S/N = .25. The frequencies of the four sine waves were selected so that there would be one component at a high frequency, with respect to the Nyquist frequency, one component at a low frequency and the remaining two components close together at a mid-frequency. The mid-frequency doublet was to serve as a measure of the resolution of the spectral analysis. If we set the Nyquist frequency at 100, the high frequency is 97; the low frequency is 3; and the middle frequencies are at 50 and 51. Looked at in another way: if the time represented by the sampled signal shown in the figure is 1 sec, then the Nyquist frequency is 128 Hz, and the four sine waves have frequencies of 4, 64, 65, and 124.5 Hz.

Figure 2 shows a Blackman-Tukey spectrum for the case of 20 lags, corresponding to about 8 percent of the sample length. At this moderate lag number, the Blackman-Tukey spectrum indicates considerable energy at mid-band and at low and high frequencies, but the latter are unresolved to specific frequencies.

Figure 3 shows the maximum entropy spectrum for the corresponding case of 20 prediction error coefficients calculated according to the Levinson algorithm. The correspondence between lag number and coefficient number arises from the fact that the same autocorrelation functions would be involved in calculating the

10. Lacoss, R.T. (1971) Data adaptive spectral analysis methods, Geophysics 36:661.

11. Ulrych, T.J. (1972) Maximum entropy power spectrum of truncated sinusoids, J. Geophys. Res. 77:1496.

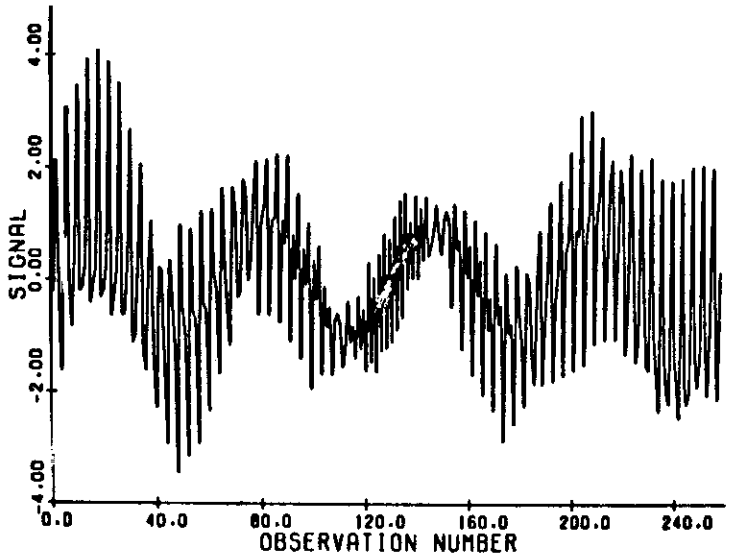


Figure 1. Synthetic Signal Used for the Spectrum Analysis

spectral densities. The locations of the maxima of the high and low-frequency components have been determined to an accuracy of about 1 percent. Although the central doublet has not been resolved, the frequency midway between the unresolved components is determined to about .01 percent.

Figure 4 shows the Burg or BPEC maximum entropy spectrum for the same case of 20 coefficients. The results are qualitatively similar to those obtained for the Levinson spectrum. However, it should be noted that the peaks are considerably sharper in the Burg spectrum.

It is obvious that for 20 coefficients the maximum entropy method results are superior to the corresponding Blackman-Tukey results. To obtain Blackman-Tukey spectra equivalent to these maximum entropy results we must use higher lag numbers. Figure 5 shows the 23 percent lag case (60 lags out of a total of 256). The 23 percent Blackman-Tukey results are very similar to 8 percent maximum entropy results, although the peaks lack the sharpness of the maximum entropy analysis. At this level, the central doublet is still not resolved by the Blackman-Tukey method. However, the power under the central peak is approximately twice that of either the high or low-frequency peaks, as it should be.

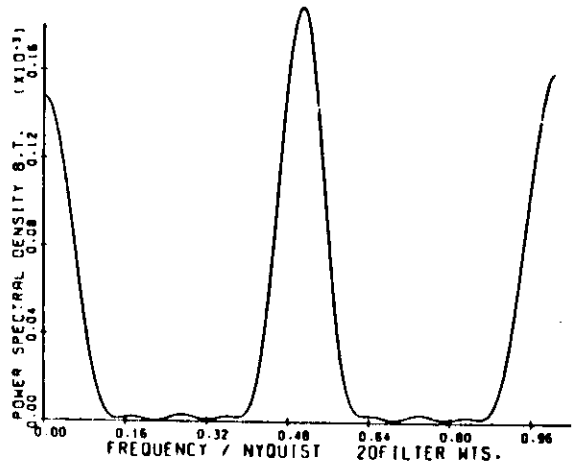


Figure 2. Improved Blackman-Tukey Spectrum: 8 Percent Lag Case or 20 Lags

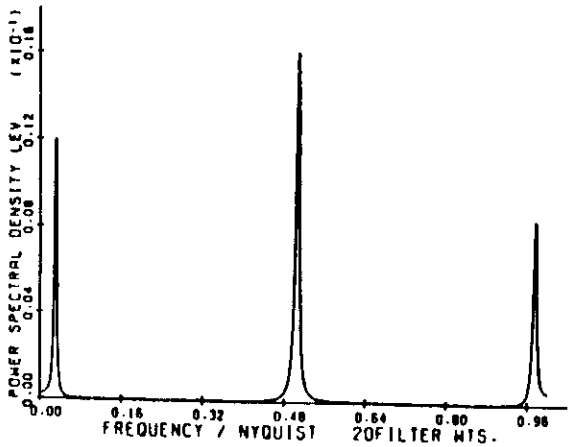


Figure 3. Levinson Maximum Entropy Spectrum for 20 Prediction Error Coefficients

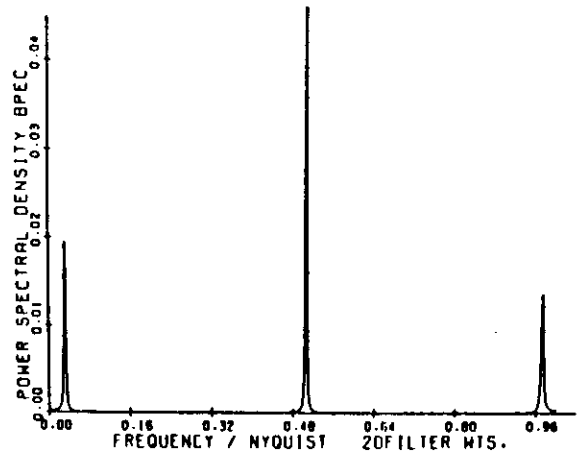


Figure 4. Burg Maximum Entropy Spectrum for 20 Coefficients

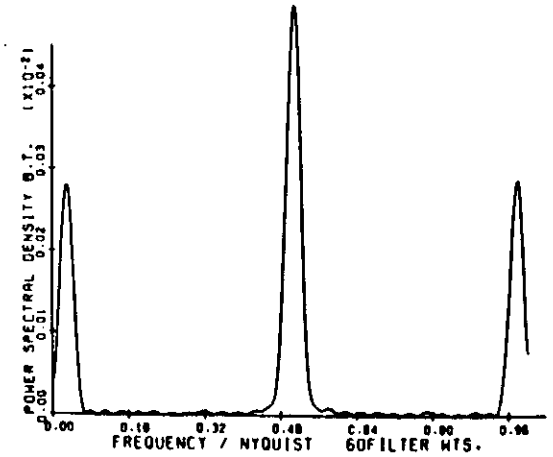


Figure 5. Improved Blackman-Tukey Spectrum: 23 Percent Case or 60 Lags

Figure 6 is the Levinson spectrum for the corresponding case of 60 coefficients. The improved resolution of the maximum entropy method has been maintained. The central peak is beginning to show signs of splitting into its components. This is indicated by an increase in breadth and the slight flattening at the top of the peak.

Figure 7 shows the Burg spectrum for the same case. The central peak has now been resolved quite obviously into its two components. This result clearly reveals the superiority of the Burg technique for calculating the final prediction error coefficient over the direct recursive solution of the matrix equation (6) which we have called the Levinson technique. In this Figure, the amplitudes of the four peaks are approximately equal. However, this case is an appropriate place to stress that the spectrum amplitude is not necessarily a good measure of the power for the maximum entropy method. The amplitude of a peak in the maximum entropy spectrum tends to become higher and the bandwidth narrower as the number of coefficients is increased. We have noted that the amplitude of

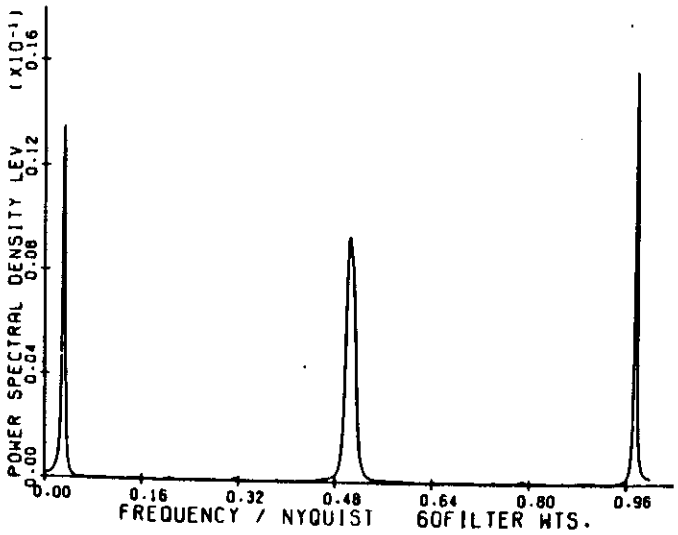


Figure 6. Levinson Maximum Entropy Spectrum for 60 Prediction Error Coefficients

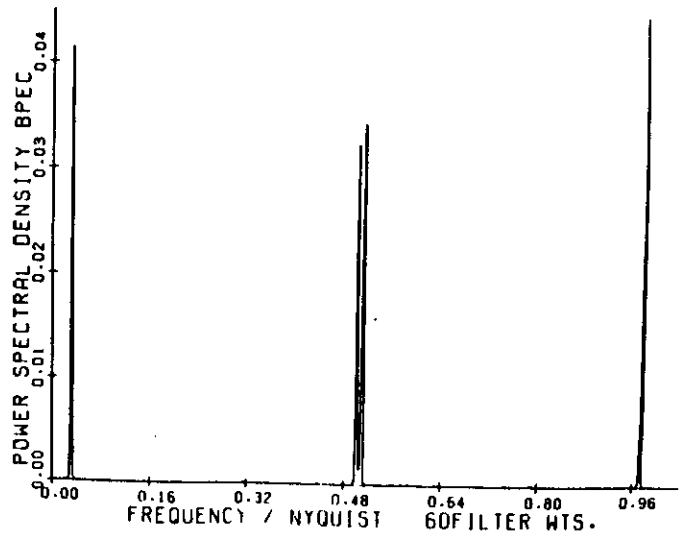


Figure 7. Burg Maximum Entropy Spectrum for 60 Coefficients

a peak may change considerably as the number of filter coefficients changes, but the curves become much steeper and the area under a peak remains approximately constant. Therefore, for power comparisons in the maximum entropy method, the area under a particular peak is the significant quantity.

One more set of spectra will emphasize the hierarchy of resolution among the Blackman-Tukey, Levinson and Burg spectral estimates. Figure 8 shows a Blackman-Tukey spectrum for 100 lags or about 39 percent of the sample. The results have improved over the previous 23 percent result but the central doublet still cannot be seen. Figure 9 shows the Levinson spectrum for 100 filter weights, and the mid-frequency peak has now split into its components. But the cleft is not as deep as in the Burg spectrum for the same number of coefficients. This is shown in Figure 10. The Burg spectrum has remained very stable as the number of weights has increased. This spectrum for 100 coefficients is almost identical to that for 60 coefficients.

One question naturally arises: Does the Blackman-Tukey method ever resolve the central doublet? The answer is "yes" but it does it poorly and not well until about the 98 percent lag case. This is shown in Figure 11. In the

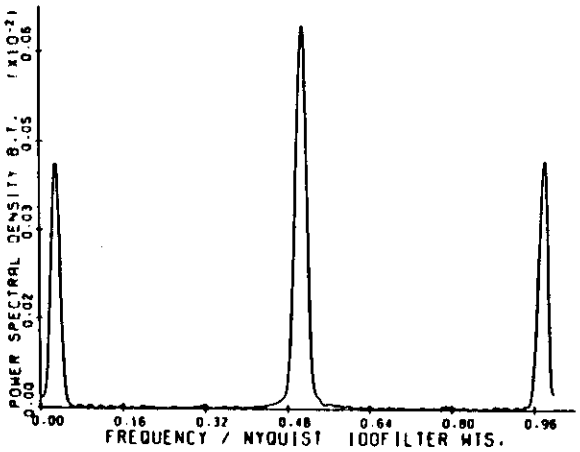


Figure 8. Improved Blackman-Tukey Spectrum: 30 Percent Case or 100 Lags

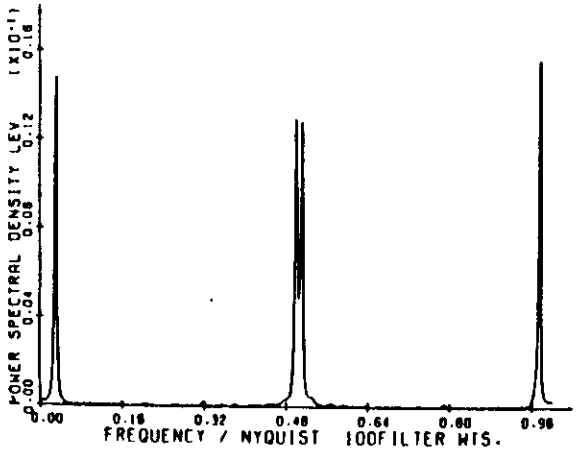


Figure 9. Levinson Maximum Entropy Spectrum for 100 Prediction Error Coefficients

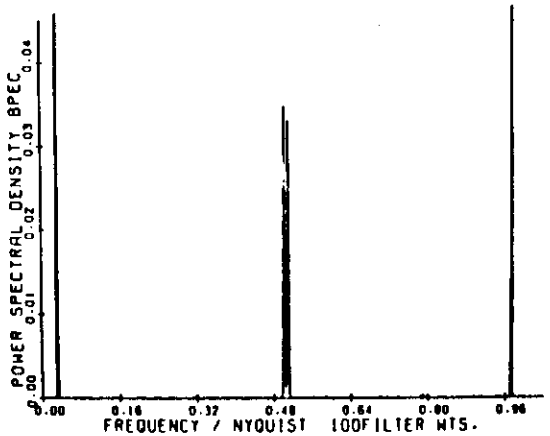


Figure 10. Burg Maximum Entropy Spectrum for 100 Coefficients

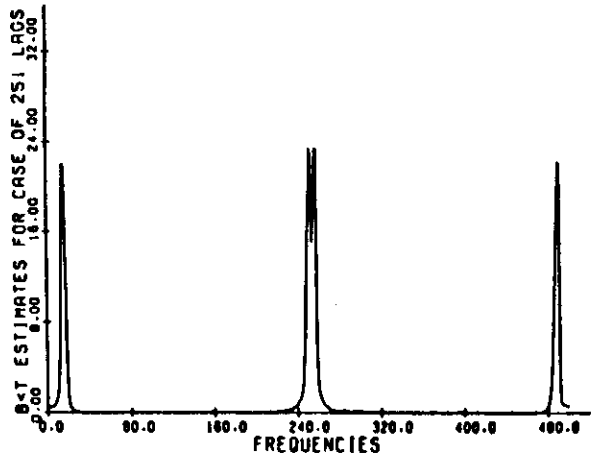


Figure 11. Improved Blackman-Tukey Spectrum: 98 Percent Lag Case or 251 Lags

figure, the frequency scale has been chosen to contain 500 intervals (501 points). The relation between the sequence number and original frequency in the sampling period is: frequency = 0.256 (sequence number). In their book, Blackman and Tukey recommend using a relatively small ratio, up to about 10 percent maximum lag number to total sample size. We see that the improved Blackman-Tukey spectrum finally shows signs of revealing the real signal only well beyond that point — almost an order of magnitude beyond. Thus, in some cases the standard Blackman-Tukey spectrum estimate may be providing only a fuzzy, out-of-focus view of the "real" world.

It is important to emphasize the high resolving power of the Burg maximum entropy method. This can be done with Figure 12, which represents one of our earlier calculations of the BPEC maximum entropy spectrum for 101 coefficients or about 40 percent of the sample length. The central doublet is highly resolved and the four frequencies are accurately determined. However, the amplitudes and the power under the four peaks appear to be considerably in error, since the

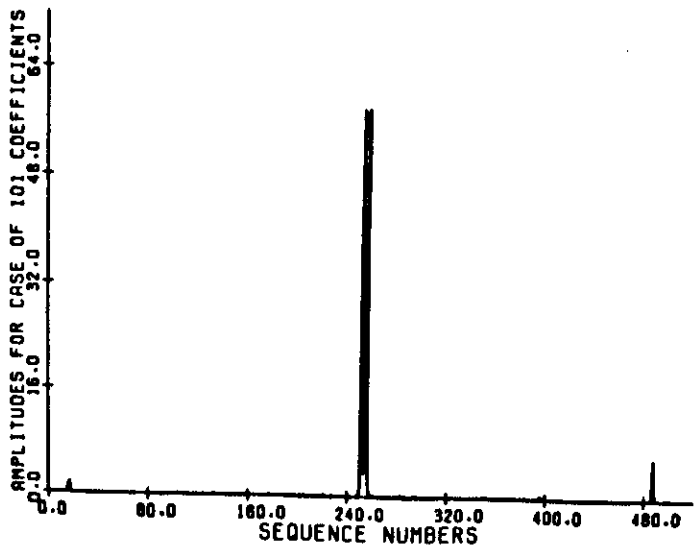


Figure 12. Burg Maximum Entropy Spectrum for 101 Coefficients Plotted on an Equally Spaced Grid of 501 Points

high and low frequency components are almost insignificant compared to the mid-frequency components. In fact, the powers are correct. As in the previous figure, this spectrum was plotted on an equally spaced grid of 501 points. However, in this case, the actual maxima are so sharp that they fall between grid points and do not appear on the plots. For ideal signals, such as we are analyzing, the Burg technique produces extremely high resolution spectra. For example, if the actual maximum amplitude for the high frequency component near sequence number 488 had been plotted on the same scale as shown, it would have a value of 31,000; that is, it would be approximately 4000 times higher than the peak in the figure.

4. CONCLUSION

The results of analyses using the maximum entropy spectral estimate certainly reveal it to be a promising method of signal analysis. The maximum entropy method can produce high resolution spectra with accurate frequency and power estimates even for short data samples. In resolving power, the maximum entropy spectra are superior to those calculated by the Blackman-Tukey technique; also they have the advantage of not requiring the use of window functions. Among methods of calculating maximum entropy spectra, Burg's prescription for calculating the prediction error coefficients produces higher resolution spectra than calculations based on Levinson's algorithm, which requires an estimate of the autocorrelation function. Clearly the maximum entropy method should have many applications both on old and new time series. In conclusion, some examples of this analysis applied to geomagnetic micropulsations and the Zürich sunspot number follow.

Figure 13 is a three-dimensional figure which shows eight modified Blackman-Tukey micropulsation spectra obtained from a 15-minute data sample, as the lag time is varied between 10 and 80 percent in steps of 10. The data is a sample of the $\hat{Y}$ component measured in Eastern Massachusetts on 4 March 1970 at 1800 UT. The three principal peaks are at 66, 100, and 160 seconds. The first curve on the figure, the 10 percent case, is similar to the one at which most Blackman-Tukey analyses would stop.

Figure 14 shows a similar 3-D representation of BPEC spectra for the same data. The number of coefficients used varies from 80 to 140 in steps of 10. In percent of the data sample this is equivalent to about 31 to 55 percent. Multiplet structure never resolved by the Blackman-Tukey method does become

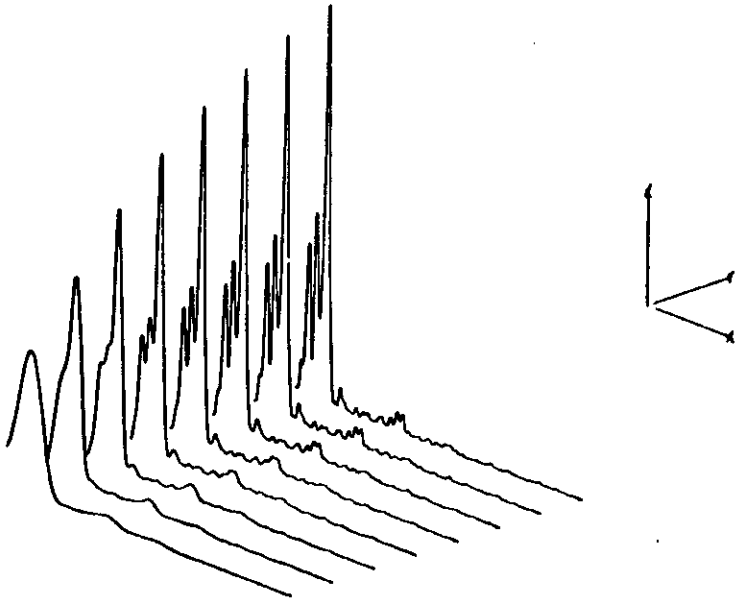


Figure 13. Blackman-Tukey Spectra of Geomagnetic Micropulsations

evident. This sub-structure may be important since some current micropulsation theories are predicting rather complex series of multiplet spectra.[12]

Figure 15 shows Burg spectra computed from 89 years of R_Z, the Zürich sunspot number, using annual means from 1884 to 1972. The eight curves are spectra for cases with 10 to 80 coefficients in steps of 10 or from 12 to 90 percent of the sample length. The frequency axis extends from 6 to 12 cycles/100 years. Therefore, the figure shows a region around 11 years between about 8 and 17 years. The structure of the 11-year line can be seen more clearly in Figure 16. This figure shows one of the previous cases, which

12. McClay, J. F. (1973) On the Asymmetric nature of micropulsations-I. The spectrum, Planet. Space Sci. 21:2193.

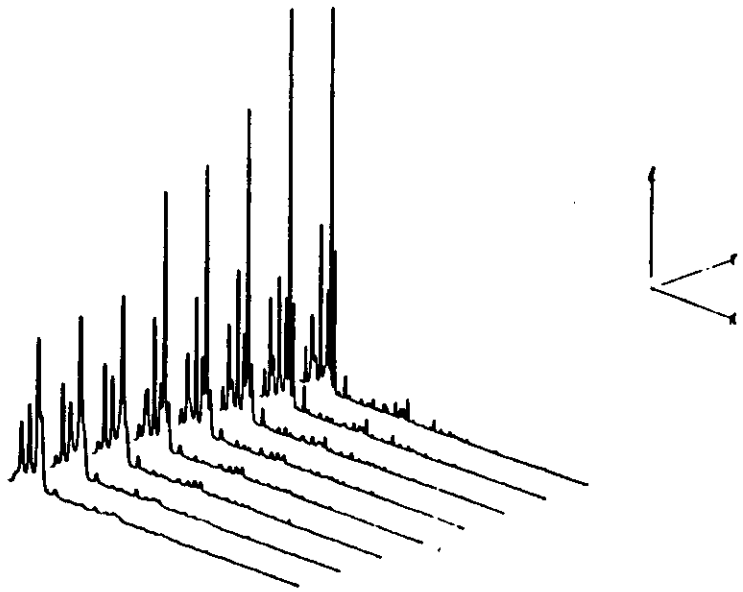


Figure 14. Burg Maximum Entropy Spectra of Geomagnetic Micropulsations

corresponds to 70 coefficients on an extended frequency scale. The well-known 11-year line is revealed as a triplet with periods of 12.9, 10.7, and 9.5 years; there is evidence of harmonic structure at 5.3 years near 19 cycles/100 years (CPHY). The low frequency spike is located at about 129 years.

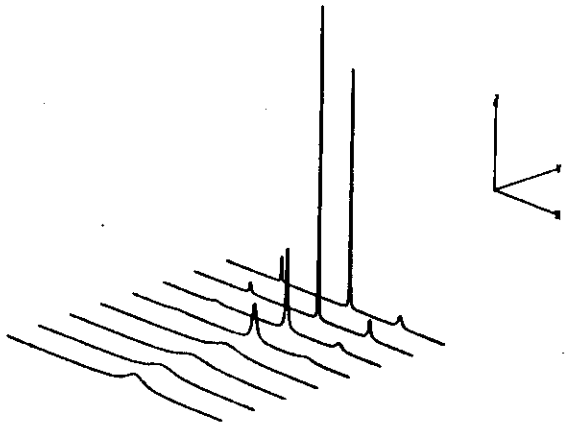


Figure 15. Burg Maximum Entropy Spectra of the Zürich Sunspot Number

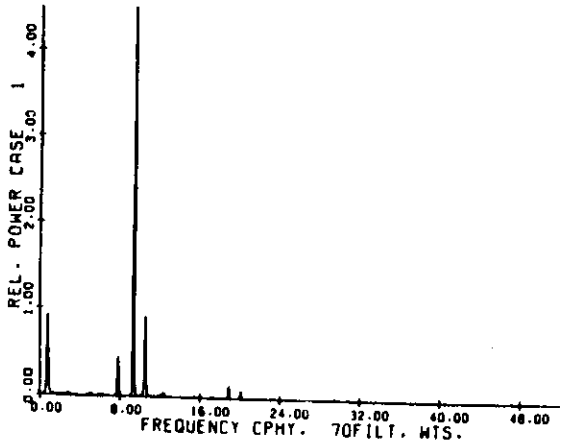


Figure 16. Burg Spectrum of the Sunspot Number for 70 Coefficients

References

1. Blackman, R.B. and Tukey, J.W. (1958) The Measurement of Power Spectra, Dover Publications, Inc., New York, pp. 30-37.
2. Levinson, N. (1947) The Wiener RMS (root mean square) error criterion in filter design and prediction, J. Math. and Phys., 25:261.
3. Burg, J.P. (1968) A new analysis technique for time series data. Paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, Netherlands.
4. _____(1967) Maximum entropy spectral analysis. Paper presented at 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma.
5. Woodward, P.M. (1964) Probability and Information Theory, with Applications to Radar, Pergamon Press, Oxford, p.49.
6. Shannon, C.E. and Weaver, W. (1964) The Mathematical Theory of Communication, The University of Illinois Press, Urbana, p. 93.
7. Peacock, K.L. and Treitel, S. (1969) Predictive deconvolution: theory and practice, Geophysics 34:155.
8. Parzen, E. (1969) Multiple time series modeling in Multivariate Analysis-II, P.R. Krichnaiak, Ed., Academic Press, New York, p. 404.
9. Barnard, T.E. (1969) Analytical studies of techniques for the computation of high-resolution wavenumber spectra, Advanced Array Research, Special Report No. 9, Texas Instruments, Inc.
10. Lacoss, R.T. (1971) Data adaptive spectral analysis methods, Geophysics 36:661.
11. Ulrych, T.J. (1972) Maximum entropy power spectrum of truncated sinusoids, J. Geophys. Res. 77:1396.
12. McClay, J.F. (1973) On the asymmetric nature of micropulsations-I. The spectrum, Planet. Space Sci. 21:2193.

A Comparison of Power Spectral Estimates and Applications of the Maximum Entropy Method

HENRY R. RADOSKI, PAUL F. FOUGERE, AND EDWARD J. ZAWALICK

Air Force Cambridge Research Laboratories, Hanscom Air Force Base, Bedford, Massachusetts 01731

A new spectral estimate, called the maximum entropy method, is described. This estimate was originated by John Parker Burg for use in seismic wave analysis. In the maximum entropy method the entropy, or information, of a signal is maximized under the constraint that the estimated autocorrelation function of the signal is the Fourier transform of the spectral power density. The spectral estimates are calculated in two ways: (1) by minimization of the error power to obtain the coefficients of the prediction error filter, as suggested by Burg, and (2) by a direct solution of the matrix equation using an algorithm due to Norman Levinson. For comparison a Blackman-Tukey technique, calculated with a Hamming window, is used also. We illustrate these three methods by applying them to a composite signal consisting of four sinusoids of unit amplitude: one each at high and low frequencies and two at moderate frequencies with respect to the Nyquist frequency, to which is added white noise of 0.5 amplitude. Results are shown to indicate that the best correspondence with the input spectrum is provided by the Burg technique. Applications of the maximum entropy method to geomagnetic micropulsations reveal complex multiplet structure in the Pc 4, 5 range. Such structure, not previously resolved by conventional techniques, has been predicted by a recent theory of magnetospheric resonances. In a period range 7 orders of magnitude longer than micropulsation periods, analysis of annual sunspot means shows that the 11-yr band is composed of at least three distinct lines. With each of these lines is associated a harmonic sequence. Long periods of the order of 100 yr also are revealed.

This paper discusses three methods of spectral analysis that we call Blackman-Tukey, Levinson, and Burg. The first is actually a modification of the usual *Blackman and Tukey* [1958] method on two counts: (1) it provides a higher density of calculated points, and (2) it can be carried out to large lag numbers, i.e., over 90% of the sample length. The application of this technique results in a smooth spectrum and higher resolution. In our Blackman-Tukey results a Hamming lag window was applied to the autocorrelation function. The second type of spectrum is called a Levinson spectrum. This is a maximum entropy spectral estimate that is calculated by using an algorithm due to *Levinson* [1947]. The Levinson spectrum provides an improvement in resolution over the modified Blackman-Tukey method. The third spectrum is also called a BPEC spectrum, an acronym formed from Burg prediction error coefficients. This spectrum is also a maximum entropy estimate but is calculated according to a technique developed by *Burg* [1968]. The BPEC spectrum turns out to be an improvement on the Levinson spectrum, producing a sharp spectrum with excellent frequency resolution. In the following section the basic concepts and computational techniques used in the maximum entropy method are discussed. Then the accuracy and resolving power of the three spectral methods are illustrated with a synthetic signal. In the final sections the geophysical applications to geomagnetic micropulsations and the Zürich sunspot number are presented.

MAXIMUM ENTROPY METHOD

The maximum entropy spectral estimate is a relatively new method, first presented by *Burg* [1967]. It offers a novel technique of spectral estimation that should be of considerable value not only for application to geomagnetic micropulsations and indices of magnetic activity but to any time series. At present the maximum entropy method seems to be the best method of spectral analysis yet devised.

Entropy, which is a thermodynamic state function, is defined in statistical mechanics as the logarithm of the volume of phase-space accessible or the number of quantum states available to a system [e.g., *Kittel*, 1958, p. 16; *Mayer and Mayer*, 1959, p. 92]. This volume or number measures our lack of precise knowledge of the actual state of a system, since a given macrostate may derive from a multitude of nonidentical microstates. Hence entropy also is a measure of our uncertainty or of the randomness of a system, since a system that is more disordered occupies more volume in phase space. The concept of entropy is also used in information theory. In fact, in terms of probabilities it has the same form as the physical entropy for the canonical ensemble [*Shannon and Weaver*, 1949, p. 20]. Now entropy measures the average information content in a message [*Woodward*, 1964, p. 49]. Here information has the special meaning of measuring the freedom of choice in selecting a message. With such freedom of choice of an event goes an uncertainty of the outcome of the event. Hence in both the physical and the information theoretical case, entropy is a measure of randomness.

Problems in which the entropy of a distribution is maximized subject to constraints are well known. For example, if the constraint is that a distribution has a given mean square, the distribution with maximum entropy is a Gaussian. In statistical mechanics this case corresponds to a Maxwell-Boltzmann distribution at a given temperature. A similar problem in the calculus of variations is solved in obtaining the maximum entropy spectrum. In this case the constraint is that the power density spectrum must agree with *Wiener's* [1949, p. 43] theorem that the Fourier transform of the spectral density is equal to the autocorrelation function of the observed time series. The specific functional form of the entropy function can be obtained from the change of entropy that occurs when a band-limited signal passes through a filter with a given frequency response. In this case the change in entropy is the integral of the logarithm of the square of the filter response [*Shannon and Weaver*, 1949, p. 60; *Lacoss*, 1971]. This form also may be obtained from the entropy of a stationary Gaussian process having the same autocorrelation function as the signal. The entropy is proportional to the quantity E in (1)

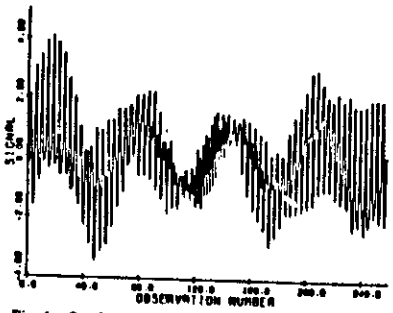


Fig. 1. Synthetic signal used for the spectrum analysis.

[*Smylie et al.*, 1973, p. 414]. The entropy function to be maximized is the logarithm of the spectral density integrated over the frequency range:

$$E = \int_{-f_c}^{+f_c} \ln P(f)\, df \tag{1}$$

where $f_c = (2\Delta t)^{-1}$ is the Nyquist frequency and Δt is the sampling time increment. The mathematical formulation of the maximum entropy spectrum is that we must determine the spectral density $P(f)$ that maximizes the quantity E, subject to the $N + 1$ constraints that the autocorrelation $C(\tau)$ is given by

$$C(\tau) = \int_{-f_c}^{+f_c} P(f)\, e^{i2\pi f\tau}\, df \tag{2}$$

where $\tau = n\Delta t$ for $n = 0, 1, 2, \cdots, N$ and N measures the maximum lag time of the autocorrelation and the order of the spectrum.

A Blackman-Tukey power spectrum is determined in essence by the Fourier transform of the autocorrelation

$$P(f) = \int_{-\infty}^{+\infty} C(\tau) e^{-i2\pi f\tau}\, d\tau \tag{3}$$

In the maximum entropy method the constraint is that the autocorrelation must be given by the Fourier transform of the power spectrum. Since the autocorrelation function and the power spectral density are Fourier transform pairs, the Blackman-Tukey spectrum and the maximum entropy spectrum would be identical for an infinitely long data sample. However, for finite data samples they are often very different, and the difference favors the maximum entropy method by producing higher-resolution spectra.

Maximizing the integrated logarithm of the spectral density can be interpreted as determining that spectral density which makes the entropy of the signal stationary with respect to variations in those unknown values of the autocorrelation function at lag times beyond the length of the original signal [*McGee*, 1971; *Smylie et al.*, 1973, p. 415]. Alternatively, it is equivalent to constructing a filter that whitens the signal and leaves the entropy unchanged. In other words, the input signal is converted to noise. Ideally, the filter has removed all of the predictable portion of the signal. Therefore it is not surprising that the maximum entropy filter is related to the unit time prediction filter and is, in fact, a unit time prediction error filter [*Peacock and Treitel*, 1969]. This is a filter whose output is not a prediction of the input signal but the error between the known signal and the predicted signal. Because the filter involves finding prediction coefficients or autoregressive coefficients the spectrum also has been referred to as an autoregressive spectral estimator [*Parzen*, 1967]. If we know the coefficients of the whitening filter, then the spectrum is logically determined by the reciprocal of the squared response of that filter. Heuristically, if the response of the filter is $G(\omega)$ and the amplitude spectrum of the signal is $X(\omega)$, then the whitening filter has the effect of

$$G(\omega) \cdot X(\omega) = Q \tag{4}$$

where Q is a constant noise amplitude. Therefore the power spectral density of the signal is proportional to $X^2(\omega) = Q^2/G^2(\omega)$, as in (5) below. The problem reduces to determining the set of prediction error coefficients. The coefficients are determined by a set of $N + 1$ simultaneous equations for $N + 1$ coefficients. This set of equations can be put in the form of a matrix equation involving the elements of the autocorrelation function in the matrix and the prediction error coefficients as a vector. The spectral density and the prediction error coefficients G_i are given by

$$P(f) = \frac{\Delta t P_{N+1}}{\left|\sum_{i=0}^{N} (G_{i+1} Z^i)\right|^2} \tag{5}$$

$$\sum_{i=0}^{N} \phi_{ki} G_{i+1} = P_{N+1}\delta_{k,0} \qquad k = 0, 1, \cdots, N \tag{6}$$

where $Z = e^{i2\pi f\Delta t}$ and $G_1 = 1$ [*Barnard*, 1969; *Edward and Fitelson*, 1973]. The matrix with elements ϕ_{kj} is constructed from the autocorrelation function $C(\tau) = C(n\Delta t)$ by setting $n = k - j$ and $\phi_{kj} = C(|k - j|\Delta t)$, and P_{N+1} represents the mean output power of the prediction error filter containing $N + 1$ coefficients.

Equation (6) can be solved directly by inverting the autocorrelation matrix. However, orders of magnitude in time are saved by using Levinson's recursion relations between the coefficients for the matrices of order N and $N + 1$. If $G_j^{(N)}$ is the jth coefficient obtained from a matrix of order N, then the coefficients of the next case are given by

$$G_j^{(N+1)} = G_j^{(N)} + G_{N+1}^{(N+1)} G_{N+2-j}^{(N)} \tag{7}$$

where $j = 2, \cdots, N$. Ultimately, only the last coefficient, $G_{N+1}^{(N+1)}$, need be specified, the rest being defined in terms of this single coefficient. The determination of this last coefficient greatly influences the resulting spectra. Currently, there are

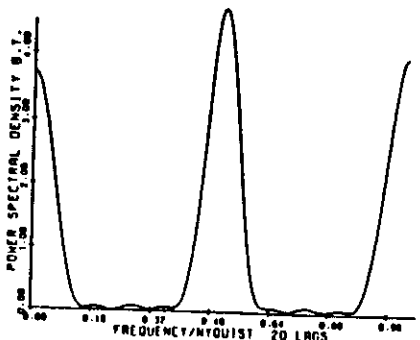


Fig. 2. Blackman-Tukey spectrum: 8% lag case, or 20 lags.

two methods for calculating this value. The Levinson method involves estimating the autocorrelation function. Poor estimates can lead to very poor spectra. The autocorrelation function that we have used with success is a biased but consistent estimate [*Jenkins and Watts*, 1969, p. 175]. First, the mean of the time series is subtracted from the elements of the series, and then the sum of the lag products is divided by the total number of data points rather than by the number of lag products, as is suggested by *Blackman and Tukey* [1958, p. 34]. These spectra provide an improvement of resolution over Blackman-Tukey spectra. In the Levinson method the final coefficient is determined by

$$G_{N+1}^{(N+1)} = -\sum_{i=0}^{N-1} \frac{C(j+1)G_{N-i}^{(N)}}{P_N} \tag{8}$$

The recursion relation for the mean error power is

$$P_{N+1} = P_N[1 - (G_{N+1}^{(N+1)})^2] \tag{9}$$

where the initial term P_1 is the mean power of the input time series. Burg's method of calculating the final prediction error coefficient has the advantage of not needing any independent estimate of the autocorrelation function. In fact, the method can be used to determine simultaneous estimates of the autocorrelation function and of the power spectrum directly from the data. This method formally minimizes the error power between the actual signal and a prediction of the signal. The prediction error filter is applied in both a forward and a backward direction in time to guarantee that the final coefficient has a magnitude of less than 1. Then the results of the coefficient recursion relation equation (7) and the error power recursion relation (9) are used to calculate the spectra according to the maximum entropy spectral estimate as given in (5). An extensive mathematical treatment can be found in the paper by *Smylie et al.* [1973], and a flow diagram for the recursive procedure is given by *Anderson* [1974].

Burg's technique produces spectra of extremely high resolution and large dynamic range. The sharpness of the spectra may require some initial acclimatization after one has become accustomed to working with lower-resolution spectra, as can be seen in the following sections.

Comparison of Spectra for a Synthetic Signal

This section contains examples of maximum entropy and Blackman-Tukey spectra to illustrate the enhanced resolution of the maximum entropy method. Other examples of maximum entropy spectra are given by *Lacoss* [1971], *Ulrych* [1972], and *Currie* [1973*a*, *b*]. Figure 1 shows a trace of the ideal signal that was analyzed. The signal was sampled at 257 equally spaced points and consists of a sum of four sine waves of equal amplitude, to which was added a small amount of white noise. The noise has a uniform probability distribution in the range $-N$ to $+N$. For convenience, N is called the noise amplitude. In the case to be presented, the noise amplitude was one half of the sine wave amplitude. This corresponds to a signal to noise ratio of 24. The contribution from each sine wave to the power is $S^2/2$, where S is the amplitude of the signal, and the contribution from the white noise is $N^2/3$. Thus the signal to noise ratio is $6(S/N)^2$. Several other cases, whose results are available, considered the same signal and larger noise amplitudes down to $S/N = 0.25$. The frequencies of the four sine waves were selected so that there would be one component at a high frequency with respect to the Nyquist frequency, one component at a low frequency, and the remaining two components close together at a midfrequency. The midfrequency doublet was to serve as a measure of the resolution of the spectral analysis. If we set the Nyquist frequency at 100, the high frequency is 97, the low frequency is 3, and the middle frequencies are 50 and 51. Looked at in another way: if the time represented by the sampled signal shown in the figure is 1 s, then the Nyquist frequency is 128 Hz, and the four sine waves have frequencies of 4, 64, 65, and 124.5 Hz.

For the analysis the data have been transformed to standard variables; i.e., each datum has been divided by the standard deviation of the signal after the mean was subtracted. Thus the mean square of the input power of the transformed signal is unity. It follows by Parseval's theorem that the output power or the area under the spectral density curve must be unity also. We have calculated the output spectra on a grid of 501 points (f_i, $i = 1, 2, \cdots, 501$); i.e., there are 500 elementary frequency bands between zero and the Nyquist frequency. When the signal to noise ratio is high, the resolution afforded by 500 elements can become grossly insufficient in determining the actual location of the maximums of the spectra calculated by the maximum entropy method. As the number of filter weights increases, the spectra become sharper and tend to approach line spectra. These spikes can become so sharp that even if each elementary band were subdivided into 500 subbands, the correct shape still could elude detection. We have solved this difficulty and obtained the actual position and amplitudes of the spectral peaks by a systematic search procedure, as follows.

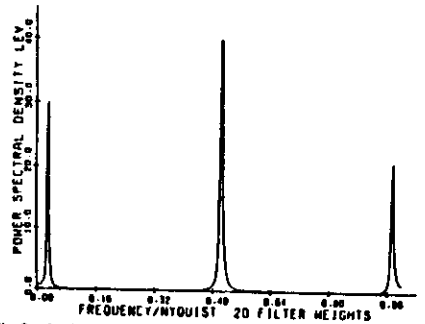


Fig. 3. Levinson maximum entropy spectrum for 20 prediction error coefficients.

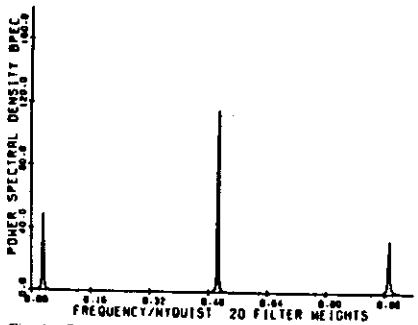


Fig. 4. Burg maximum entropy spectrum for 20 coefficients.

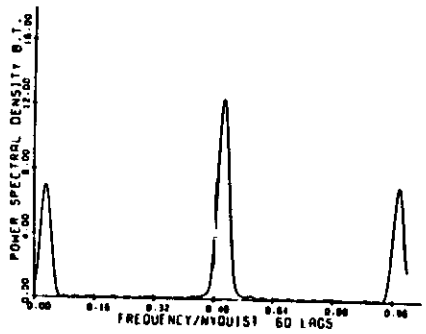


Fig. 5. Blackman-Tukey spectrum: 23% case, or 60 lags.

Each value of i from 2 to 500 is examined. Whenever an i is found such that

$$P(f_{i-1}) < P(f_i) > P(f_{i+1})$$

then an approximate peak has been found. The assumption is made that the peak is single. Halfway between adjacent points, new ordinates are calculated. The above criterion is applied to the set of five points in order to select a new triplet. The process is iterated until the ordinate of the peak is within 1% of the ordinate of both the nearest neighbors. The integral of the power spectral density function was compared to unity to determine if any peak had been missed. To display simply the actual values of the power spectral density would be misleading for power comparisons in these cases. For power comparisons the area under a particular peak is the significant quantity and not the amplitude alone. Therefore we have decided to display the average power density in each of the 500 elementary frequency bands to facilitate power comparisons among and within the various spectral methods. For each value of i the limits of integration were $f_{i\pm1/2}$ with half cells at the end points. The integration was performed by using Simpson's rule n times with $2^{n+1}+1$ ordinates. If two successive approximations to the integral agreed to within 1%, the procedure was terminated. If convergence did not occur by $n = 8$, the elementary region was halved, and the procedure was repeated for each half. This subdivision proceeded until convergence occurred. Such careful numerical integration procedures are not required and have not been used for the Blackman-Tukey spectra.

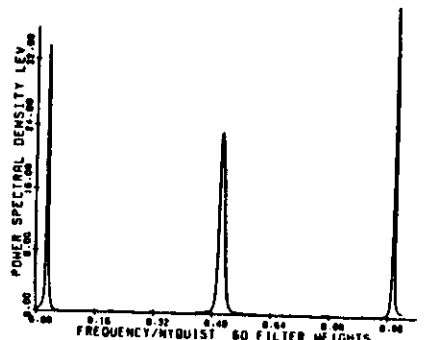


Fig. 6. Levinson maximum entropy spectrum for 60 prediction error coefficients.

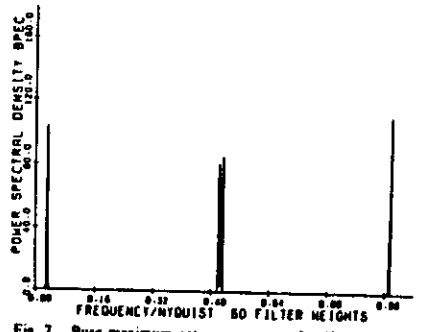


Fig. 7. Burg maximum entropy spectrum for 60 coefficients.

Figure 2 shows a Blackman-Tukey spectrum for the case of 20 lags, corresponding to about 8% of the sample length. At this moderate lag number the Blackman-Tukey spectrum indicates considerable energy at midband and at low and high frequencies, but the latter two ranges are unresolved to specific frequencies. The low-amplitude oscillations appearing in the spectrum are Gibbs' phenomena caused by the sharp commencement and termination of the signal.

Figure 3 shows the maximum entropy spectrum for the corresponding case of 20 prediction error coefficients calculated according to the Levinson algorithm. The correspondence between lag number and coefficient number arises from the fact that the same autocorrelation functions would be involved in calculating the spectral densities. The frequencies of the high and low components agree with the input frequencies to an accuracy of about 1%. Although the central doublet has not been resolved, the frequency midway between the unresolved components is accurate to about a hundredth of a percent.

Figure 4 shows the Burg, or BPEC, maximum entropy spectrum for the same case of 20 coefficients. The results are qualitatively similar to those obtained for the Levinson spectrum, and the frequencies are as accurately determined. However, note that the peaks are considerably sharper in the Burg spectrum.

It is obvious that for 20 coefficients the maximum entropy method results are superior to the corresponding Blackman-

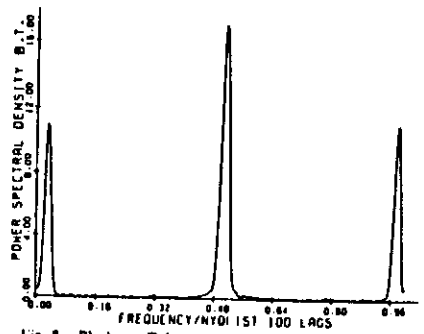


Fig. 8. Blackman-Tukey spectrum: 39% case, or 100 lags.

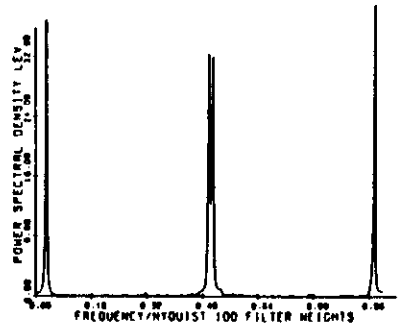


Fig. 9. Levinson maximum entropy spectrum for 100 prediction error coefficients.

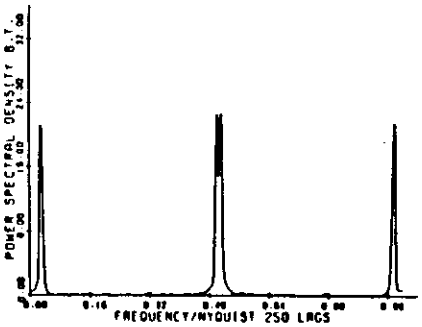


Fig. 11. Blackman-Tukey spectrum: 98% lag case, or 250 lags.

Tukey results. To obtain Blackman-Tukey spectra equivalent to these maximum entropy results, we must use higher lag numbers. Figure 5 shows the 23% lag case, i.e. 60 lags out of a total of 256. The 23% Blackman-Tukey results are very similar to 8% maximum entropy results, although the peaks lack the sharpness of the maximum entropy analysis. At this level the central doublet is still not resolved by the Blackman-Tukey method. However, the power under the central peak is approximately twice that of either the high- or low-frequency peak, as it should be.

Figure 6 is the Levinson spectrum for the corresponding case of 60 coefficients. The improved resolution of the maximum entropy method has been maintained. The central peak is just beginning to show signs of splitting into its components, as is indicated by an increase in breadth and the slight flattening at the top of the peak.

Figure 7 shows the Burg spectrum for the same case. The central peak has now been resolved quite obviously into its two components. This result clearly reveals the superiority of the Burg technique for calculating the final prediction error coefficient over the direct recursive solution of the matrix equation (6), which we have called the Levinson technique.

One more set of spectra will emphasize the hierarchy of resolution among the Blackman-Tukey, Levinson, and Burg spectral estimates. Figure 8 shows a Blackman-Tukey spectrum for 100 lags, or about 39% of the sample. The results have improved over the previous 23% result, but the central doublet still cannot be seen. By comparing Figures 2, 5, and 8 the Gibbs' oscillations can be observed to increase in frequency and decrease in amplitude, as would be consistent with an analytic result. Figure 9 shows the Levinson spectrum for 100 filter weights, and the midfrequency peak has now split into its components. But the cleft is not as deep as it is in the Burg spectrum for the same number of coefficients, as is shown in Figure 10. The Burg spectrum has remained very stable as the number of weights has increased. This spectrum for 100 coefficients is almost identical to the one for 60 coefficients.

One question that naturally arises is whether the Blackman-Tukey method can ever resolve the central doublet. The answer is yes but not until a high lag number is reached. The 98% lag case is shown in Figure 11. In their book, Blackman and Tukey recommend using a relatively small ratio, up to about 10% of maximum lag time to total sample size. We see that the modified Blackman-Tukey spectrum finally shows signs of revealing the real signal only way beyond that, almost an order of magnitude beyond. Despite the large lag number, no indication of instability in the spectrum is observed. Our analyses of geophysical time series also remain stable at large lag numbers.

Conclusions

The results of analyses using the maximum entropy spectral estimate certainly reveal it to be a promising method of signal analysis. The maximum entropy method can produce high-

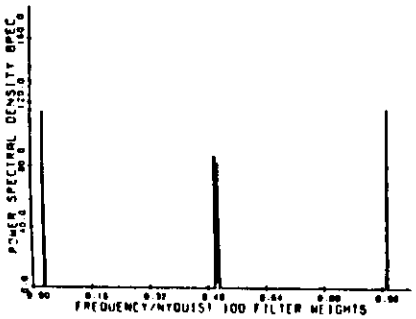


Fig. 10. Burg maximum entropy spectrum for 100 coefficients.

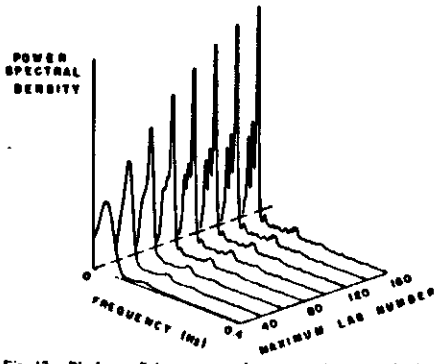


Fig. 12. Blackman-Tukey spectra of geomagnetic micropulsations.

1972. The eight curves are spectra for cases with 10–80 coefficients in steps of 10, or from 12 to 90% of the sample length. The frequency axis extends from 6 to 12 cycles/100 yr; the periods range between about 8 and 17 yr. The prominent peak in the figure is the 11-yr line. Beginning at about 50 coefficients, the spectrum becomes more complex with the appearance of additional lines. The location of these lines is stable as the number of coefficients is increased. In addition, the location of the central peak is unaffected by the appearance of these satellite peaks. In our opinion this series of spectra shows no evidence of instability at coefficient numbers larger than 50% of the sample length. The structure of the entire spectrum can be seen more clearly in Figure 16. This figure shows one of the previous cases, which corresponds to 70 coefficients on an extended frequency scale. The 11-yr line is revealed as a triplet with periods of 12.9, 10.7, and 9.5 yr, and there is evidence of harmonic structure at 5.3 yr near 19 cycles/100 yr. The low-frequency spike is located at about 129 yr.

The primary purpose of this paper was to compare three methods of spectral analysis and to demonstrate the improved resolution inherent in the Burg maximum entropy technique. As illustrations of the geophysical application of Burg's method, spectra of geomagnetic micropulsations and the Zürich sunspot number were given. Detailed presentations on these subjects are in preparation.

Acknowledgments. The authors wish to thank John Parker Burg of Stanford University for his generous assistance, Tad J. Ulrych of the University of British Columbia for a copy of his BPEC program and for several comments that improved our presentation, and Robert G. Currie of the Magnetic Observatory of the South African Council for Scientific and Industrial Research for many valuable discussions.

The Editor thanks R. T. Lacoss and T. J. Ulrych for their assistance in evaluating this paper.

References

Anderson, N., On the calculation of filter coefficients for maximum entropy spectral analysis, *Geophysics*, *39*, 69, 1974.

Barnard, T. E., Analytical studies of techniques for the computation of high-resolution wavenumber spectra, Advanced Array Research, *Spec. Rep. 9*, Tex. Instrum., Inc., Dallas, Tex., 1969.

Blackman, R. B., and J. W. Tukey, *The Measurement of Power Spectra*, Dover, New York, 1958.

Burg, J. P., Maximum entropy spectral analysis, paper presented at 37th meeting, Soc. of Explor. Geophys., Oklahoma City, Okla., Oct. 1967.

Burg, J. P., A new analysis technique for time series data, paper presented at the Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, NATO, Enschede, Netherlands, Aug. 1968.

Currie, R. G., Fine structure in the sunspot spectrum—2 to 70 years, *Astrophys. Space Sci.*, *20*, 509, 1973a.

Currie, R. G., Geomagnetic line spectra—2 to 70 years, *Astrophys. Space Sci.*, *21*, 425, 1973b.

Currie, R. G., The period and Q_c of the Chandler wobble, submitted to *Geophys. J. Roy. Astron. Soc.*, *70*, 1974.

Edward, J. A., and M. M. Fitelson, Notes on maximum-entropy processing, *IEEE Trans. Inform. Theory*, *19*, 232, 1973.

Jenkins, G. M., and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, 1969.

Kittel, C., *Elementary Statistical Physics*, John Wiley, New York, 1958.

Lacoss, R. T., Data adaptive spectral analysis methods, *Geophysics*, *36*, 661, 1971.

Levinson, N., The Wiener RMS (root mean square) error criterion in filter design and prediction, *J. Math. Phys. Cambridge Mass.*, *25*, 261, 1947.

Mayer, E. M., and M. G. Mayer, *Statistical Mechanics*, John Wiley, New York, 1959.

McClay, J. F., On the asymmetric nature of micropulsations, I, The spectrum, *Planet. Space Sci.*, *21*, 2193, 1973.

McGee, T., An exposition of the maximum entropy criterion in spectral analysis, report, Dep. of Geol., Univ. of Brit. Columbia, Dec. 1971.

Parzen, E., Multiple time series modeling, in *Multivariate Analysis*, vol. 2, edited by P. R. Krishnaiah, p. 404, Academic, New York, 1969.

Peacock, K. L., and S. Treitel, Predictive deconvolution: Theory and practice, *Geophysics*, *34*, 155, 1969.

Shannon, C. E., and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

Smylie, D. E., G. K. C. Clarke, and T. J. Ulrych, Analysis of irregularities in the earth's rotation, in *Methods in Computational Physics*, vol. 13, *Geophysics*, pp. 391–430, Academic, New York, 1973.

Ulrych, T. J., Maximum entropy power spectrum of truncated sinusoids, *J. Geophys. Res.*, *77*, 1396, 1972.

Wiener, N., *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Technology Press, Cambridge, Mass., 1949.

Woodward, P. M., *Probability and Information Theory, With Applications to Radar*, Pergamon, New York, 1964.

(Received May 6, 1974;
accepted October 22, 1974.)

14

Physics of the Earth and Planetary Interiors, 12 (1976) 208–216

THE SUPERIORITY OF MAXIMUM ENTROPY POWER SPECTRUM TECHNIQUES APPLIED TO GEOMAGNETIC MICROPULSATIONS

HENRY R. RADOSKI, EDWARD J. ZAWALICK and PAUL F. FOUGERE

Air Force Cambridge Research Laboratories, Hanscom Air Force Base, Bedford, Mass. (U.S.A.)

(Accepted for publication January 22, 1976)

Radoski, H.R., Zawalick, E.J. and Fougere, P.F., 1976. The superiority of maximum entropy power spectrum techniques applied to geomagnetic micropulsations. Phys. Earth Planet. Inter., 12: 208–216.

Several methods of spectrum analysis have been applied to a micropulsation event associated with a magnetospheric substorm. The micropulsations were detected and recorded at Sudbury, Massachusetts, U.S.A. (53.9°N, 357.1°E, geomagnetic coordinates). The data sample is for the period 0630–0900 UT on September 1, 1970 which includes most of an isolated substorm which was preceded by low levels of magnetic activity (K_p = 1°, 0300–0600 UT). Spectra were obtained using 15-min segments of low pass filtered and digitized data for the two horizontal magnetic field components. A dynamic spectrum was obtained by moving the data window by increments of 1 min through the data set. The results show that methods based on the maximum-entropy principle yield higher resolution than the traditional techniques, such as Blackman-Tukey or the discrete Fourier transform using the fast Fourier transform algorithm. Multiplet structure, predicted by recent micropulsation theories, is sharply resolved by maximum-entropy techniques.

1. Introduction

Geomagnetic micropulsations are low-frequency signals superimposed upon the earth's magnetic field (Jacobs, 1970). Their period range is defined somewhat arbitrarily to be between 0.2 and 600 sec. Their amplitude may be a few hundred gammas to milligammas, while the earth's field is of the order of $5 \cdot 10^4$ γ at the earth's surface (1 γ = nT). There is a similarity in period between these small magnetic disturbances and microseisms. While microseisms are associated with coastal disturbances in the weather (or storms), micropulsations often are associated with structural disturbances (or storms) in the magnetosphere. The main magnetic field has its source within the earth, but micropulsations have external sources. They are generally considered to be manifestations of hydromagnetic waves propagating at the Alfvén speed in the magnetosphere. A characteristic time for magnetospheric oscillations may be obtained by dividing a scale size of the magnetosphere, e.g. 10 R_e, by an average Alfvén speed, e.g. 0.1 R_e/sec. This would give a characteristic period of 100 sec, which is in the micropulsation range. Thus, micropulsations may represent the resonant vibrations of the magnetosphere either in its entirety or of various of its structural features (McClay, 1970; Chen and Hasegawa, 1974). Micropulsations are potentially a sensitive tool for diagnosing the properties of the magnetosphere. As the magnetosphere changes because of its varying interaction with the solar wind, the properties of micropulsations should respond. A fundamental property of micropulsation signatures is their frequency content and its variation with time. Therefore, it is important to have a reliable method of spectrum analysis. In recent years there has been considerable interest in a technique of spectrum analysis called the maximum-entropy method (MEM), developed about 1966 by J.P. Burg (Ulrych and Bishop, 1975). This method appears to offer results which have higher resolution and accuracy than standard methods (Radoski et al., 1975). It is the purpose of this paper to investigate both micropulsations and the maximum-entropy method by calculating and comparing spectra of a particular micropulsation event.

2. Instrumentation

The micropulsations are detected and recorded at Sudbury, Massachusetts, U.S.A. by measuring the time rate of change of the earth's magnetic field. The coordinates of this station are 42.4°N, 288.7°E (geographic); 53.9°N, 357.1°E (geomagnetic); 56.9°N, 3.2°E corrected geomagnetic). The horizontal components, X and Y, are directed towards geographic north and east. However, at this location, the gradient to the lines of constant corrected geomagnetic latitude is within a few degrees of the geographic meridian (ComEort, 1970). Therefore, the horizontal geographic components for the Sudbury station also can be considered as corrected geomagnetic components. The horizontal instruments are solid mu-metal core induction coils and the vertical instrument is an air core induction coil consisting of several 1-m square sections. The horizontal components are recorded on both high- and low-sensitivity tracks to yield a total dynamic range of 70 dB. At 0.1 Hz, the recorded horizontal magnetic field can span a range from about 1 mγ to 3 γ. Commercial solid state preamplifiers are used and the signals are FM recorded on magnetic tape at 0.1 in./sec. The response of the system in measuring B is flat between 0.001 and 0.1 Hz. Six analog tracks are recorded simultaneously: high- and low-sensitivity X and Y, the vertical component Z and a time channel.

3. Data reduction

The data were digitized at a rate of approximately 17 points per two seconds. The actual time increment was calculated to be Δt = 0.1168573 sec from 77,017 data points in exactly 2.5 h. The corresponding Nyquist frequency was ν_N = 4.2787 Hz. It was assumed that there was no significant power recorded above the Nyquist frequency. The results indicate that this was a conservative estimate. Since we were interested in periods greater than 10 sec the data were filtered and "decimated" by taking every 30th point. The new Nyquist period became 7.0114 sec. The filter was a 141-point symmetrical numerical filter of the least-square type (Belhannon and Ness, 1966). This low-pass filter has a cosine shape. To avoid any possibility of aliasing, the cut-off period was set at 8.183 sec. Subsequent spectral analysis indicated that very little power was present for periods shorter than 15–20 sec. The Y data were normalized by subtracting the mean of the 2.5-h sample and dividing by the standard deviation. The X data were similarly treated except that the Y-standard deviation was used in order to permit amplitude comparisons between the horizontal components. In addition, the mean was removed from each 15-min sample before its spectrum was calculated. This was done to smooth the long-period end of each spectrum by removing the dc component and periods longer than the time window. Since the Z data had been contaminated by local noise during this period, it was decided to forego a detailed spectral analysis of this component.

4. Data and spectra

The data sample is for the period 0630–0900 UT on September 1, 1970. This period includes most of an isolated sub-storm which began about 0655 UT (Kisabeth and Rostoker, 1974). The substorm was preceded by low levels of magnetic activity: K_p = 1°, 1° for 0000–0600 UT. The data samples for the horizontal components are shown in Fig. 1, in which

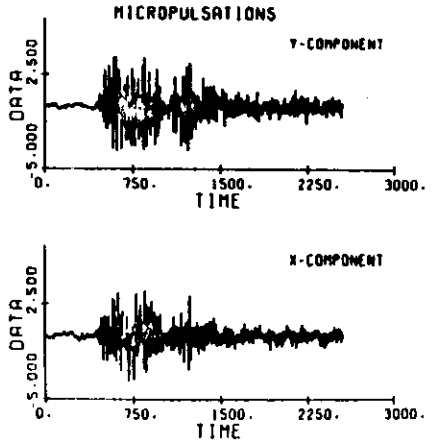


Fig. 1. Horizontal variometer components showing the micropulsation event on September 1, 1970 at 0630–0900 UT.

the time unit is about 3.5 sec (half of Nyquist period). Spectra were obtained from 15-min segments of the digitized low-pass filtered data of the X and Y components. Dynamic spectra, representing the change in time of the spectra, were constructed by moving this 15-min data window in steps of 1 min through the data, giving 136 spectra in all. The various spectral estimates will be compared by using a particular 15-min portion of the data, corresponding to the 35th record. This record is shown in Fig. 2. Once again the time unit is 3.5 sec.

The discrete Fourier transform (DFT) estimate using the popular fast Fourier transform algorithm (FFT) is shown in Fig. 3. The FFT algorithm is simply a fast method of calculating the DFT (Cooley and Tukey, 1965). The spectra shown here have not been smoothed. PSD stands for power spectral density. For the Blackman-Tukey and maximum entropy power spectra, which are to follow, we found that a 501-point spectrum or 500 elementary frequency bands between zero and the Nyquist frequency produced a reasonably smooth spectrum. We wished to obtain an equivalent number of elementary frequency bands with the DFT. In order to calculate a spectrum of comparable smoothness using

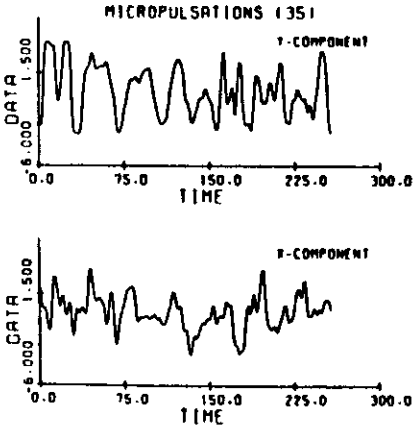


Fig. 2. The 15-min record of the micropulsation event beginning at 0704 UT.

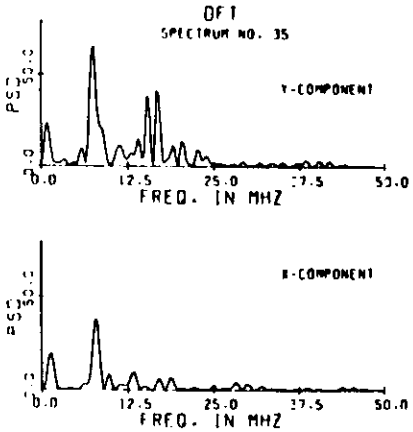


Fig. 3. Discrete Fourier transform spectral estimate. MHZ stands for millihertz.

the FFT algorithm, we require a power of 2 such that half of this value is approximately 500: namely, $2^{10} = 1{,}024$. Thus, we have 512 elementary frequency bands. Since there were 257 observations in the 15-min data sample, 767 zeroes were added to fill out the sample. The procedure of adding zeroes has been a subject of debate especially in cases where the number of added zeroes greatly outweighs the number of data points. In our case the ratio of zeroes to data points is 3 : 1, while other workers have used even larger ratios (Fraser-Smith, 1972). We are not necessarily advocating the zero extension of the data, although such a procedure is defensible as an interpolation technique (J. Bruce, private communication, 1975). However, we are using whatever means of improving the resolution of standard spectral analyses are available in order to provide the fairest possible comparison with maximum-entropy spectra. From the fact that the frequencies at which maxima occur in the spectra calculated from the DFT and Blackman-Tukey methods agree so well with those derived from such a radically different method as maximum entropy, it seems reasonable to conclude that no significant spurious peaks have been created.

Ideally, the resulting spectra should be identical whether we calculate the power spectral density from

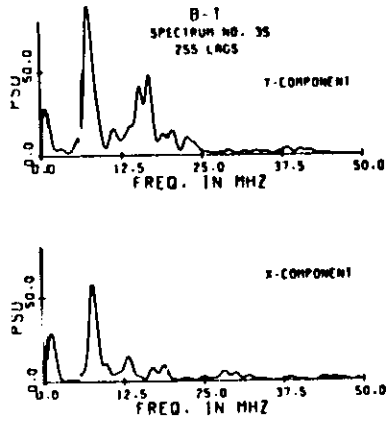


Fig. 4. Blackman-Tukey spectral estimate, 99% lag case.

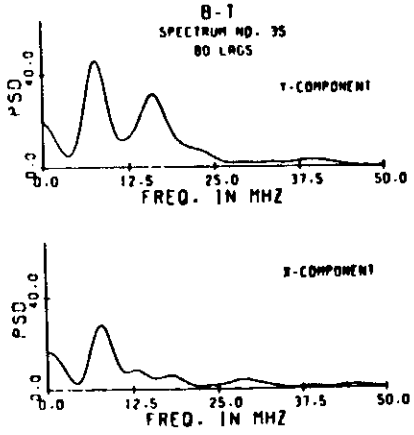


Fig. 6. Blackman-Tukey spectral estimate, 31% lag case.

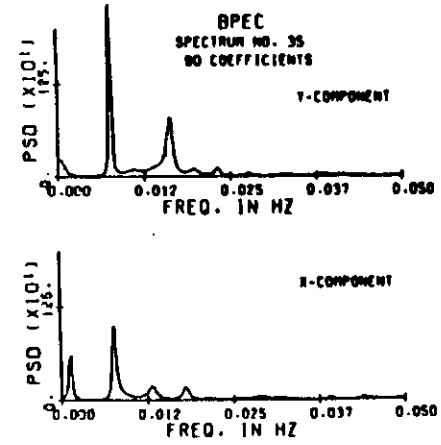


Fig. 5. Maximum entropy spectral estimate using 80 (31%) prediction-error coefficients.

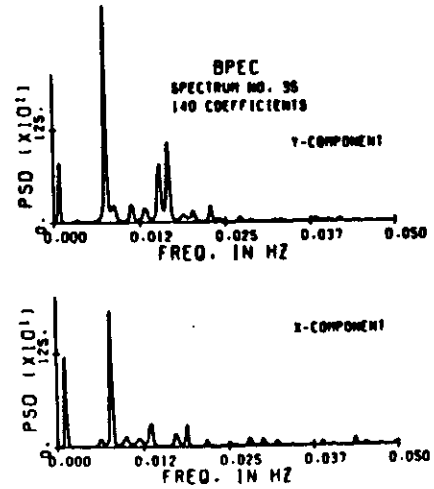


Fig. 7. Maximum entropy spectral estimate using 140 (55%) prediction-error coefficients.

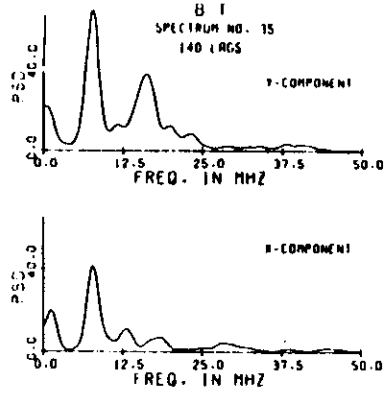


Fig. 8. Blackman-Tukey spectral estimate, 55% lag case.

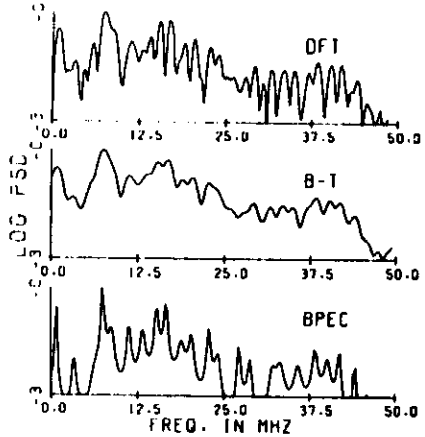


Fig. 9. Comparison of spectra of *Y* component using a logarithmic scale. The lower limit of log (PSD) was set at −3, to display only the principal maxima.

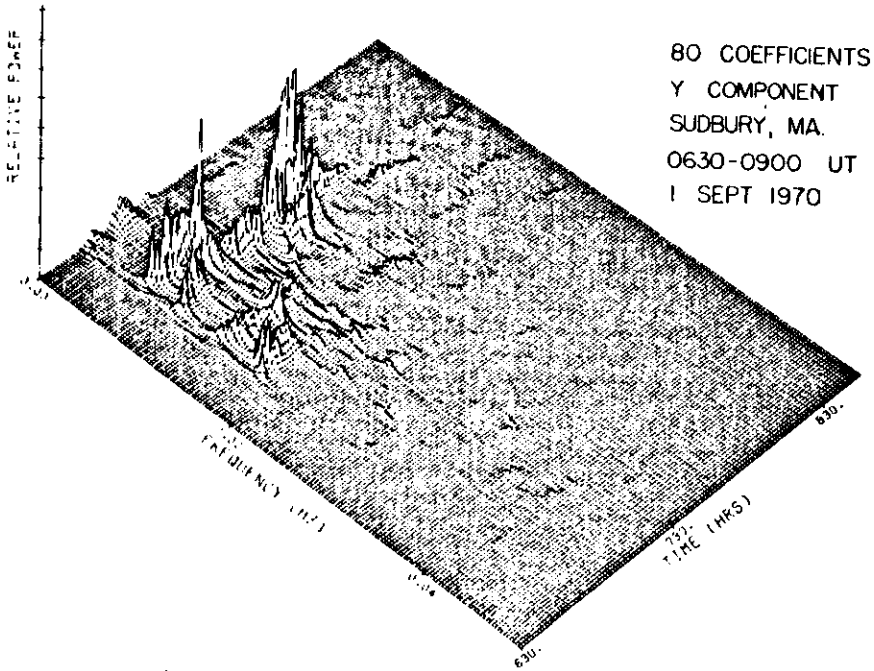


Fig. 10 Dynamic maximum entropy spectrum for the *Y* component: front view.

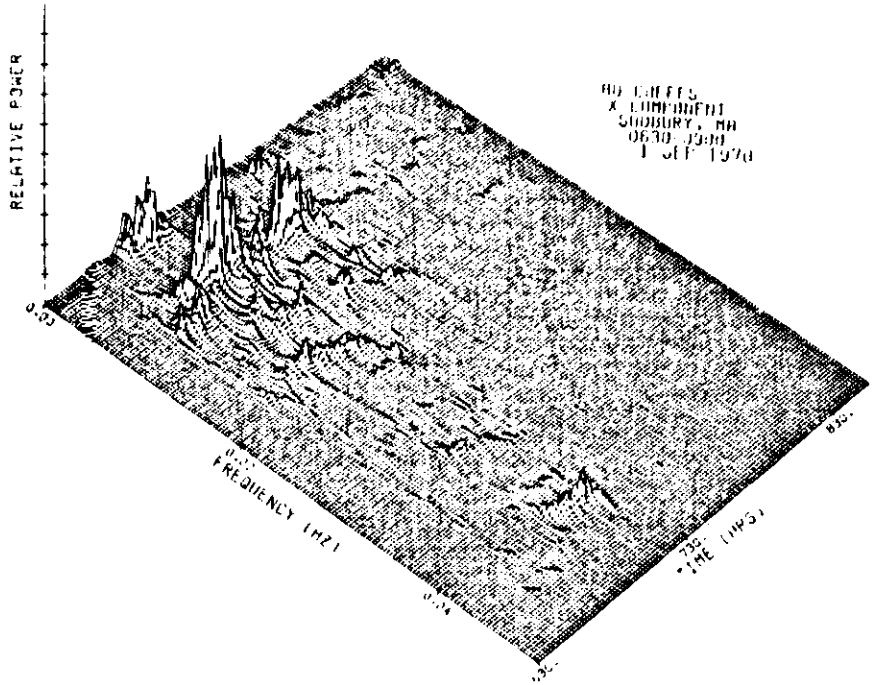


Fig. 11. Dynamic maximum entropy spectrum for the *X* component: front view.

the square of the absolute value of the Fourier transform as in the DFT method or from the Fourier transform of the autocorrelation function as in the Blackman-Tukey method. Thus, the Blackman-Tukey spectra given in Fig. 4 are similar to the DFT results. These spectra were calculated using a very large lag number (99% of the sample length) and a Hamming window. This high lag number produces the best resolution and the closest correspondence with the DFT estimates. Such a high lag number is far beyond the number usually recommended for calculating Blackman-Tukey spectra, which is about 10% of the sample length. In test cases on both micropulsations and synthetic signals we have found no instabilities using large-lag numbers. On the contrary the spectra we have examined change smoothly as the lag number is increased (Radoski et al., 1975).

The improved resolution afforded by the maximum entropy method (MEM) of spectral analysis may be seen by a study of Fig. 5. This spectral technique is also referred to as BPEC, which is an acronym for Burg prediction-error coefficients. These spectra were calculated using only 80 prediction-error coefficients which corresponds to 31% of the record length. For the sake of comparison, the corresponding Blackman-Tukey case with a lag number equal to 31% of the

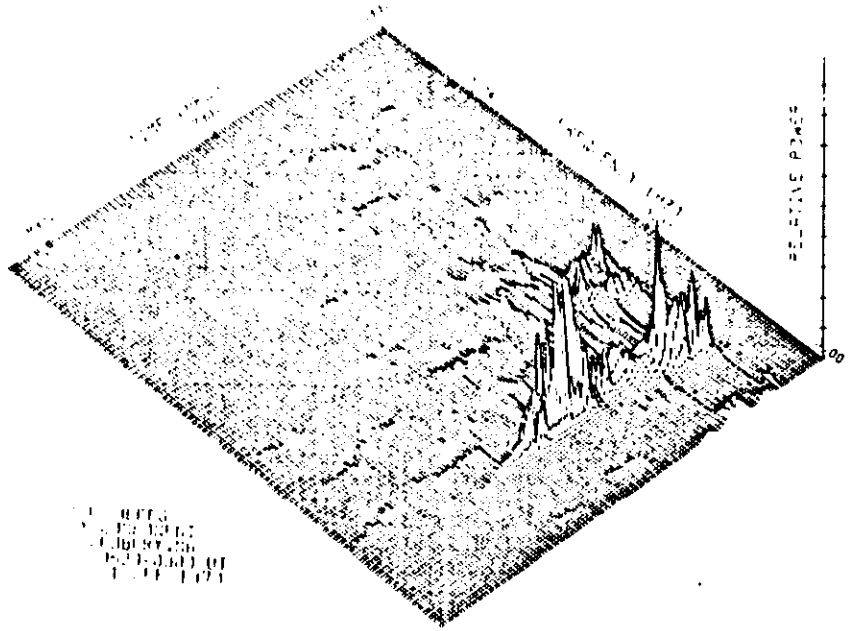


Fig. 12. Dynamic maximum entropy spectrum of the Y component: back view.

record length is shown in Fig. 6.

Several more peaks were resolved in the maximum-entropy spectra when the number of prediction-error coefficients was increased to 140, or 55% of the record length, in Fig. 7. For example, the low-frequency energy has been resolved into sharp peaks and the Y-component peak near 17 mHz has been split into a doublet. Such complex line spectra are predicted by a recent theory of magnetospheric resonances (McClay, 1973).

It has been suggested that the number of prediction-error coefficients should not exceed 50% of the record length because spectral instability may occur (Currie, 1974; Ulrych and Bishop, 1975). The question of the optimum number of coefficients to use is still unresolved. However, in this case, as in the Blackman-Tukey calculations, we have not observed any unstable results arising from the use of a large number of coefficients. We have often obtained excellent results for signals with known spectra using coefficients corresponding to almost 100% of the filter length.

The Blackman-Tukey spectra for a lag number equal to 55% of the record length are given in Fig. 8. In essence, there is no comparison between the two methods at equivalent levels. All three methods (the zero-extended DFT, 99% Blackman-Tukey, 55% maximum entropy) give practically identical frequencies of their maxima but the resolution is considerably sharper in the maximum-entropy method. This is emphasized in Fig. 9 which shows *logarithmic* plots of the Y component. The extra detail comes from the fact that we are plotting three orders of magnitude. The excellent agreement among the numerical values of the principal frequencies calculated by three different techniques indicates that no appreciable instabilities were introduced by the methods.

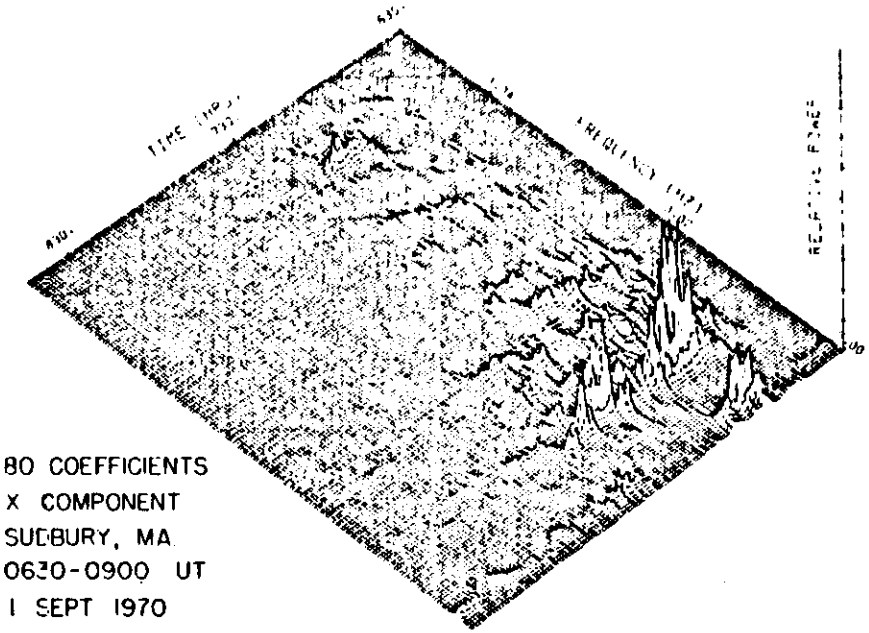


Fig. 13. Dynamic maximum entropy spectrum of the X component: back view.

The development in time of the maximum-entropy spectra of the micropulsation event is shown in Figs. 10–13. Fig. 10 is the dynamic spectrum of the Y component. The two-fold structure of the event is indicated by the twin bursts of energy at about 0700 and 0730 UT. The period is approximately 130 sec for the most prominent peak near 8 mHz. If it is assumed that the observed 130-sec period is produced by a resonant oscillation of the Sudbury field line at $L = 3.2$, then the calculated Alfvén speed would be about 670 km/sec, a reasonable value. In the earlier pulse there is a well-defined secondary peak near 15 mHz and minim structure above 20 mHz. Fig. 11 shows the X component. The predominant X component frequency is approximately the same as that of the Y component. However, there is clearly an island of "high-frequency activity near 40 mHz at 0715 UT. This energy burst i unique to the X component. Other obvious differenc are that the secondary peak around 15 mHz is considerably less well-structured in X than in Y and the event near 0730 UT is much weaker. These three-dimensional plots can be manipulated to give a view the topography of the event from any elevation and angle desired. As an example of this change in perspe tive, the far side of the event is shown in Fig. 12 for Y and in Fig. 13 for X.

A study of simple theoretical models of coupled hydromagnetic waves indicates that the characteristic

field line frequency associated with a given latitude should be excited. There also should be a preferential transfer of energy from the X to the Y component and the spectra of the two components can be different (Radoski, 1974, 1975). Although it obviously will require a bit more than the study of a single event at a single station to interpret micropulsation signatures, the results obtained for this event are encouraging. A further study is in progress comparing the Burg spectra obtained at Sudbury with a similar analysis using data from the meridian line of magnetometers operated by the University of Alberta along 301°E corrected geomagnetic longitude in northwest Canada. We hope to develop a model to explain the features and variations of the spectra and to characterize the general state of the magnetosphere and the level of magnetic activity. It should be mentioned that the Sudbury station is the first in an AFCRL magnetometer network to be established throughout the United States. A longitude chain of five stations near 55°N geomagnetic latitude will be operating in early 1976.

5. Conclusions

In studying the maximum-entropy method and comparing its results to those of other spectral techniques, our principal application has been to geomagnetic micropulsation signatures. We also have applied MEM to various other geophysical time series, such as: the international magnetic character figure, C_i, the Svalgaard AC index of inferred solarwind sector structure, the Zürich sunspot number R_z and data on the Chandler wobble. We have also made intensive studies of ideal signals such as groups of sinusoids, both with and without damping, rectangular pulses and Gaussian pulses. The only difficulty that has been encountered involves cases of single sine waves in the presence of low or no noise in which the application of BPEC sometimes yields multiple peaks depending on the phase of the input (Fougere, 1975). Except for this special case, all of the results we have obtained lead us to conclude that the high resolution of the maximum-entropy method and its independence of window functions make it the best method of spectrum analysis yet devised.

Acknowledgment

The authors wish to thank Elwood Maple of the Geomagnetism Branch of AFCRL for providing the experimental data analyzed in this report.

References

Behannon, K.W. and Ness, N.F., 1966. The design of numerical filters for geomagnetic data analysis. NASA, Rep. No. TN D-3441, 35 pp.

Che, L. and Hasegawa, A., 1974. A theory of long period magnetic pulsations, I. Steady state excitation of field line resonance. J. Geophys. Res., 79 (7): 1024–1032.

Comfort, R.H., 1970. Empirical analytic transformations between geographic and corrected geomagnetic coordinates. Northrop Corp., Huntsville, Ala., Rep. No. M-793-765.

Cooley, J.W. and Tukey, J.W., 1965. An algorithm for the machine computation of complex Fourier series. Math. Comput., 19: 297–301.

Currie, J.R., 1974. Period and Q_w of the Chandler wobble. Geophys. J.R. Astron. Soc., 38: 179–185.

Fougere, P.F., 1975. Spontaneous line splitting of short signals in maximum entropy power spectra. EOS (Trans. Am. Geophys. Union), 56 (6): 439 (abstract).

Fraser-Smith, A.C., 1972. Spectrum of the geomagnetic activity index A_p. J. Geophys. Res., 77 (22): 4209–4220.

Jacobs, J.A., 1970. Geomagnetic Micropulsations. In: J.G. Roederer (Editor), Physics and Chemistry in Space, Vol. 1, Springer, Berlin, 179 pp.

Kisabeth, J.L. and Rostoker, G., 1974. The expansive phase of magnetospheric substorms, 1. Development of the auroral electrojets and auroral arc configurations during a substorm. J. Geophys. Res., 79 (7): 972–984.

McClay, J.F., 1970. On the resonant modes of a cavity and the dynamical properties of micropulsations. Planet. Space Sci., 18: 1673–1690.

McClay, J.F., 1973. On the asymmetric nature of micropulsations, 1. The spectrum. Planet. Space Sci., 21: 2193–2211.

Radoski, H.R., 1974. A theory of latitude dependent geomagnetic micropulsations: the asymptotic fields. J. Geophys. Res., 79 (4): 596–603.

Radoski, H.R., Fougere, P.F., Zawalick, E.J., 1975. A comparison of power spectral estimates and applications of the maximum entropy method. J. Geophys. Res., 80 (4): 619–625.

Radoski, H.R., 1975. Hydromagnetic waves: Temporal development of coupled modes. EOS (Trans. Am. Geophys. Union), 56 (6): 423 (abstract).

Ulrych, R.J. and Bishop, T.N., 1975. Maximum entropy spectral analysis and autoregressive decomposition. Rev. Geophys. Space Phys., 13 (1): 183–200.

Physics of the Earth and Planetary Interiors, 12 (1976) 201–207

SPONTANEOUS LINE SPLITTING IN MAXIMUM ENTROPY POWER SPECTRUM ANALYSIS

PAUL F. FOUGERE, EDWARD J. ZAWALICK and HENRY R. RADOSKI

Space Physics Laboratory, Air Force Cambridge Research Laboratories, Hanscom Air Force Base, Bedford, Mass. (U.S.A.)

(Accepted for publication January 30, 1976)

Fougere, P.F., Zawalick, E.J. and Radoski, H.R., 1976. Spontaneous line splitting in maximum entropy power spectrum analysis. Phys. Earth Planet. Inter., 12: 201–207.

The Burg maximum entropy power spectral estimate yields extremely sharp spectra with high resolution. If the noise level is low, the power spectrum closely resembles a line spectrum. We have discovered, however, that under certain conditions, spontaneous line splitting occurs; lines which should be single split into two or more components. A series of computer experiments using artificial times series was performed with the following results. If the signal length is any multiple of 0.5 cycle there is no splitting for any initial phase of the sine wave; in fact, the spectrum is extremely sharp and stable. The same result applies to a signal length which is an odd multiple of 0.25 cycle but only with an initial phase of 0° or 90°. The splitting is most severe if the signal length is an odd multiple of quarter cycles and the initial phase is an odd multiple of 45°. In this case splitting persists even for signals as long as 49.25 cycles.

1. Introduction

In physical time series analysis, a power spectrum may be estimated for the purpose of finding hidden periodicities in the data sample. Two popular methods for making such an estimate are called the Blackman-Tukey method and the Cooley-Tukey method.

In the Blackman-Tukey (1959) method, the data outside the observation window – that is before the data sample begins and after it ends – are set equal to zero. In the Cooley-Tukey method – which is really nothing more than a Fourier analysis carried out using the fast Fourier transform algorithm – the data sample repeats itself in both time directions – forever.

In the maximum entropy method (MEM) conceived by Burg (1975), no assumptions about extending the data are required – in fact, the method explicitly is non-committal about the extension of the data. Because MEM does not require assumptions about data extension or zeros, it extends them in both time directions. It does this by generating a prediction-error filter – which when applied to the signal gives the error in a one step prediction – technically an innovation – in both time directions.

The method, described more fully in Smylie et al. (1973), Ulrych and Bishop (1975) and Radoski et al. (1975), is sometimes referred to as BPEC for Burg prediction-error coefficients. The key independent variable is the number of filter coefficients which corresponds to the number of lags in a Blackman-Tukey spectrum.

2. Synthetic signals

Our analysis of synthetic signals was motivated in part by the work of Chen and Stegen (1974) who studied the MEM spectra of sinusoids with additive Gaussian noise. They termed certain spectra "unacceptable" if the spectral peak was very broad or if there were two or more significant peaks. They attributed multiple peaks as due to an amplification of the added Gaussian noise by an overly large number of filter weights.

However, we will show that a large number of filter weights is not the critical parameter. Furthermore, the

effect is not caused by too high a noise level; on the contrary the multiplet structure shows up most clearly when the noise level is quite low.

As can be observed in Chen and Stegen's figs. 3 and 4, maximum degradation of the spectrum is achieved when the signal length is an odd number of quarter cycles and the initial phase is an odd multiple of 45°. Our results confirm this finding.

Case 1. Comparison of resolution: MEM vs. Blackman-Tukey. To give a feeling for the difference in results produced by Blackman-Tukey and by MEM, both methods have been applied to an artificial time series. Fig. 1 shows a trace of the signal which was analyzed. The signal was sampled at 257 equally spaced points and consisted of a sum of four sine waves of equal amplitude to which was added a small amount of white noise with a flat distribution. In the case to be presented, the noise amplitude was one-half of the sine wave amplitude. This corresponds to a signal-to-noise ratio of 24. The contribution from each sine wave to the power is $S^2/2$, where S is the amplitude of the signal; the white-noise contribution is $N^2/3$, where N is the amplitude of the noise. Thus, the signal-to-noise ratio is $6(S/N)^2$. Several other cases, whose results are available, considered the same signal and larger noise amplitudes down to $S/N = 0.25$. The frequencies of the four sine waves were selected so that there would be one component at a high frequency with respect to the Nyquist frequency, one component at a low frequency and the remaining two components close together at a mid-frequency. The mid-frequency doublet was to serve as a measure of the resolution of the spectral analysis. If the time represented by the sampled signal shown in Fig. 1 is 1 sec, then the Nyquist frequency is 128 Hz, and the four sine waves have frequencies of 4, 64, 65 and 124.5 Hz.

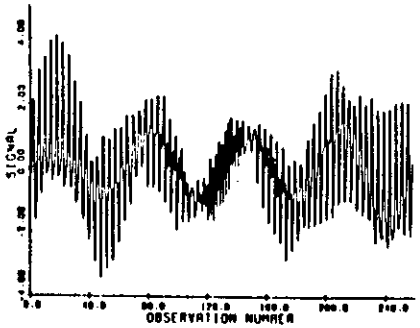


Fig. 1. Artificial time series composed of four, unit amplitude sine waves of frequencies 4, 64, 65 and 124.5 Hz, sampled 257 times at 1-sec intervals. Flat white noise with an amplitude of 0.5 has been added.

Fig. 2 shows the Blackman-Tukey spectrum estimated using 251 lags or 98% of the sample length. This is a good spectrum – but 98% is far beyond the number of lags usually recommended – about 20% or less. The doublet was resolved only in the vicinity of the large lag number.

Fig. 3 shows the MEM spectrum with only 60 prediction-error coefficients, about 23% of the sample length. It is apparent that the spectrum is much sharper than the Blackman-Tukey spectrum. Also, the four output frequencies reproduce the input frequencies very accurately. In this case, the maximum-entropy method is superior to the traditional techniques as exemplified by Blackman-Tukey.

Case 2. Single sine wave, variable phase, one full cycle. In the first test case, we used a flat-noise distribution; the remaining cases all use Gaussian noise with zero mean and unit standard deviation. Case 2 consists of a single sine wave of unit amplitude with 10^{-4} noise amplitude. The signal is one cycle of a 1-Hz sine wave sampled every 0.05 sec, giving 21 data points. The initial phase varied from 0 to 180° in 20° steps. Fig. 4 shows the ten MEM spectra plotted in a three-dimensional perspective plot using the logarithm of the power spectral density as height. The small peaks visible in the figure are many orders of magnitude below the huge, 1-Hz spikes. If the resulting imaginary solid figure has hidden lines, these are suppressed. For these spectra we used 20 weights – 95% of sample length and the spectrum remains stable as the initial phase is changed from zero to 180°. This shows th the use of a large number of weights is not necessa harmful.

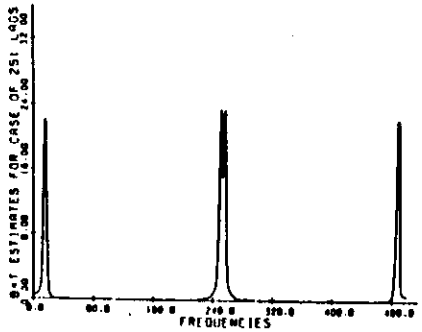


Fig. 2. Blackman-Tukey spectral estimate of data from Fig. 1.

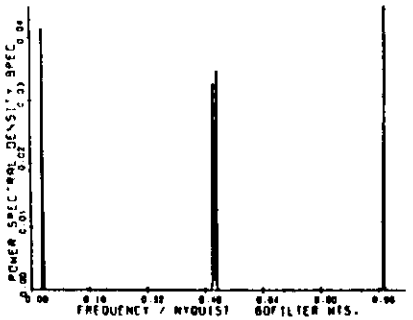


Fig. 3. BPEC maximum entropy spectral estimate of data from Fig. 1.

Case 3. Splitting with an odd number of quarter cy versus initial phase. The test signal used here was a unit amplitude 5-Hz sinusoid, 1.25 cycles long wit noise amplitude = 10^{-4}, sampled six times with Δ 0.05 sec. Maximum-entropy spectra were obtained using six weights with the initial phase varying bet 0 and 180° in 2° increments. The 91 spectra, plot in Fig. 5, display a figure "8" pattern with no split at 0, 90 and 180° initial phase. For all other phase the spectrum is a sharp doublet with maximum sep tion obtained at 45 and 135°. In this case, as well in case 2, the large number of weights used has not introduced difficulties.

Case 4. Effect of signal length. Using the worst com bination of phase (45°) and signal length (odd num of quarter cycles), the frequency was varied betwe 1.25 Hz and 49.25 Hz in steps of 2 Hz. The sampli interval was 0.01 sec giving a Nyquist frequency of Hz, and noise of amplitude 10^{-4} was added. For t case only 25 weights of a possible 101 were used i prediction-error filter. Fig. 6 shows splitting in eve

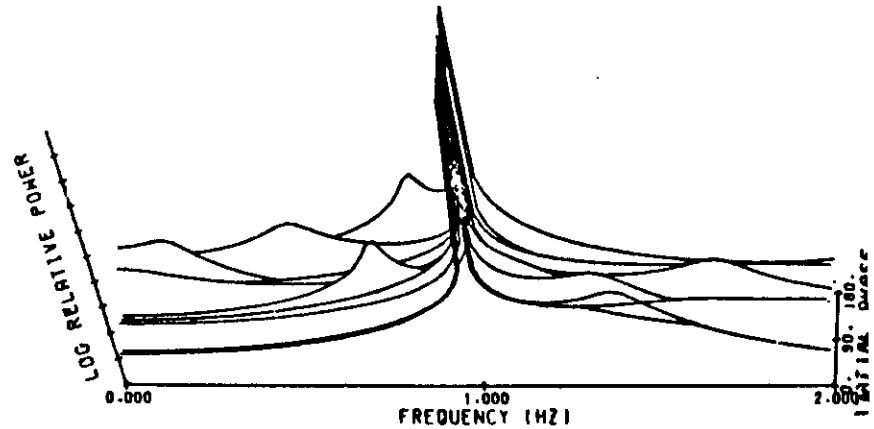


Fig. 4. 10 BPEC spectra of one cycle of a unit amplitude, 1-Hz sine wave to which Gaussian noise of amplitude 10^{-4} has been sampled 21 times. The initial phase varies from 0 to 180° in steps of 20°. 20 filter weights were used.

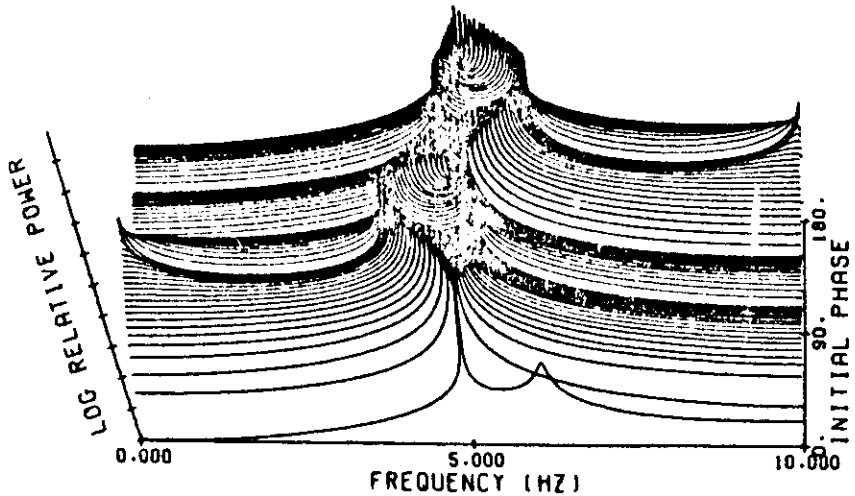


Fig. 5. Ninety-one BPEC spectra of 1.25 cycles of a unit amplitude, 5-Hz sine wave (with 10^{-4} amplitude of additive Gaussian noise) sampled six times using six filter weights. Initial phase varies from 0 to 180° in steps of 2°. Note that there is no splitting at 0, 90 and 180° and maximum splitting at 45 and 135°.

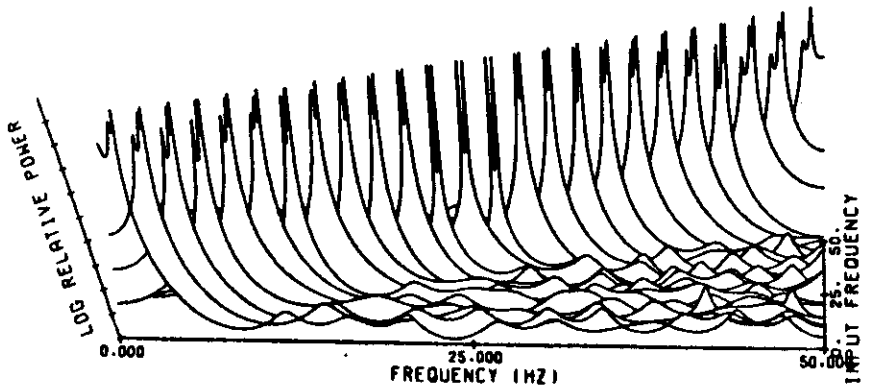


Fig. 6. BPEC spectra of unit amplitude sinusoidal signals with 45° initial phase sampled 101 times in one second. The frequencies of the signals are 1.25, 3.25, 5.25, ..., 49.25 [redacted] Gaussian white noise of amplitude 10^{-4} has been added to each signal.

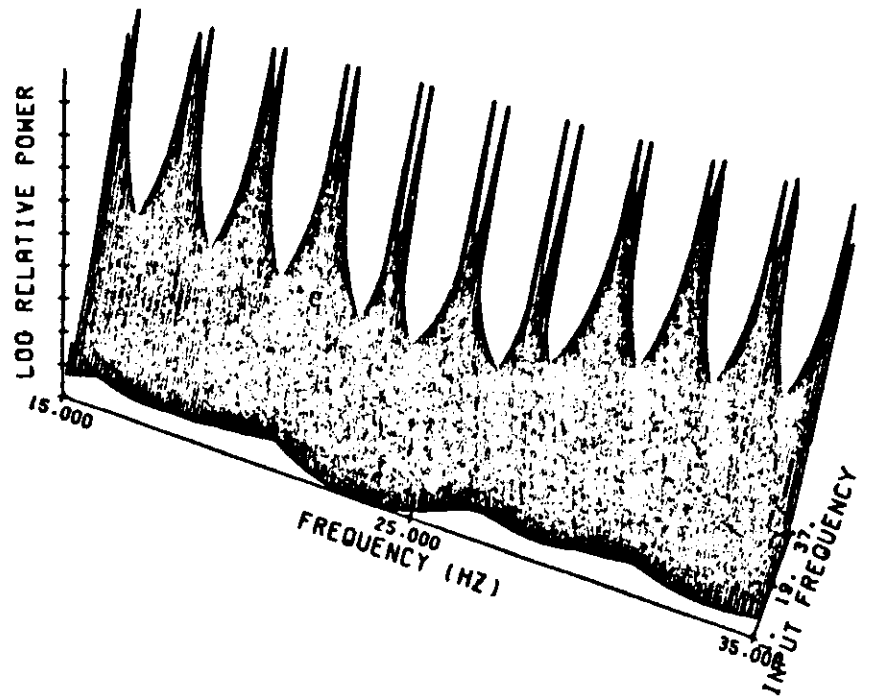


Fig. 7. An expanded version of Fig. 6.

case. At both low and high frequencies the spectra are sharp triplets while at intermediate frequencies we have sharp doublets. Fig. 7 shows an expanded plot of the intermediate frequency region from 15 to 35 Hz. In both Figs. 6 and 7, the fluctuating noise peaks are approximately eight orders of magnitude below the sharp multiplets.

3 Summary of test cases

The preceding test cases have shown that:

(1) MEM spectra have higher resolution than Blackman-Tukey spectra and they achieve their high resolution using a smaller number of weights.

(2) MEM spectra of signals containing an integral number of half-cycles are correct and display no splitting for any initial phase. This is true despite the use of a very large number of weights (20 of 21 in case 2).

(3) The splitting problem is most severe when the signal is an odd number of quarter cycles and the initial phase is an odd multiple of 45°. For an initial phase of 0° or any multiple of 90°, there is no splitting for any length of signal.

(4) When the splitting problem is most severe, lengthening the signal does not ameliorate the situation. Even for signals as long as 49.25 cycles, the spectrum still splits.

4. Application to Zürich sunspot numbers

Thus far we have studied ideal signals; the next logical problem concerns the MEM spectrum of a real geophysical time series. One such time series, which is quite well known and which has been studied extensively is the series of monthly mean Zürich sunspot numbers. A maximum entropy spectral analysis of the sunspot numbers has been given by Currie (1973). Fig. 8 shows a smoothed and decimated version

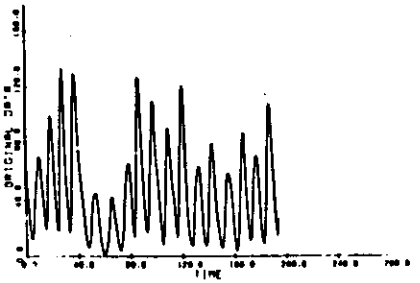


Fig. 8. The sunspot numbers smoothed using a 49-point least-squares low-pass filter applied to the monthly means of R_z, and decimated (sampled) once per year. Initial year is 1751.

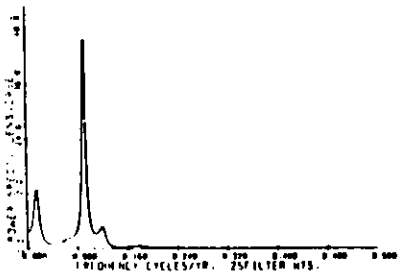


Fig. 9. BPEC spectrum of data in Fig. 8 using 25 filter weights. First 195 observations (approximately 17.75 cycles of the nominal 11-year line) were used.

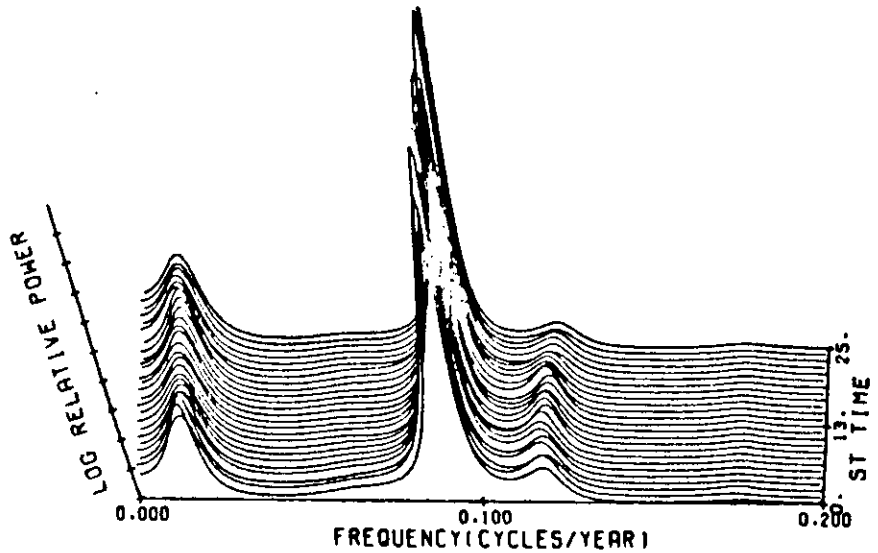


Fig. 10. 25 BPEC spectra of the signal of Fig. 8 using 195 years with a start time of year 1 = 1751, year 2 = 1752, year 25 = 1775.

of this sunspot series, with the time given in years. The smoothing was accomplished by using a low-pass symmetric 49-point least-squares filter (Behannon and Ness, 1966). The smoothed series was decimated by selecting each 12th point to yield a Δt of 1 year, giving a Nyquist frequency of 0.5 cycles per year. MEM was applied to the first 195 years of this data sample. A length of 195 years was chosen to be as close as possible to 17.75 cycles of the nominal 11-year line. The odd number of quarter cycles was chosen in an attempt to induce spurious splitting in accordance with the findings of test case 3 above.

Fig. 9 shows a linear BPEC spectrum with 25 filter weights. The 11-year line indeed appears to split; the major peak is at 10.94 years and the smaller peak to its right is at 8.40 years. If we change the initial phase of the nominal 11-year line by moving the data window ahead one year at a time, we might expect the splitting to vary as it did in test case 3.

Fig. 10 shows no such behavior. Here we see 25 spectra of the data samples 195 years long with start time of year 1 = 1751, year 2 = 1752 and so on up to year 25 = 1775. The 11-year line and the 8.4-year companion are quite stable.

Thus far, the phenomenon of spontaneous splitting has been identified *positively* only in ideal test cases. However, we feel strongly that MEM should be designed to work correctly in all cases, easy cases as well as difficult ones. The problem was solved by one of the authors (PFF) just as the present paper was being completed. A preliminary account of this solution has been presented (Fougere, 1975), from whom the interested reader may obtain a copy. A full account of the solution, being prepared for publication, will be submitted to the *Journal of Geophysical Research*.

References

Behannon, K.W. and Ness, W.F., 1966. The design of numerical filters for geomagnetic data analyses. NASA, Rep. No. TN-D-3341.

Blackman, R.B. and Tukey, J.W., 1959. The Measurement of Power Spectra from the Point of View of Communication Engineering. Dover, New York, N.Y., 190 pp.

Burg, J.P., 1975. Maximum entropy spectral analysis. Ph.D. Thesis, Stanford University, Palo Alto, Calif., 123 pp.

Chen, W.Y. and Stegen, G.R., 1974. Experiments with maximum entropy power spectra of sinusoids. J. Geophys. Res., 79: 3019–3022.

Currie, R.G., 1973. Fine structure in the sunspot spectrum – 2 to 70 years. Astrophys. Space Sci., 20: 509–518.

Fougere, P.F., 1975. A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis. EOS (Trans. Am. Geophys. Union), 56: 1054 (abstract).

Radoski, H.R., Fougere, P.F. and Zawalick, E.J., 1975. A comparison of power spectral estimates and applications of the maximum entropy method. J. Geophys. Res., 80: 619–625.

Smylie, D.E., Clarke, G.K.C. and Ulrych, T.J., 1973. Analysis of irregularities in the earth's rotation. Methods Comput. Phys., 13: 391–430.

Ulrych, T.J. and Bishop, T.N., 1975. Maximum entropy spectral analysis and autoregressive decomposition. Rev. Geophys. Space Phys., 13: 183–200.

A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis

PAUL F. FOUGERE

Air Force Geophysics Laboratory, Hanscom Air Force Base, Bedford, Massachusetts 01731

Under certain conditions, Burg maximum entropy spectra of sampled sine waves, in the presence of additive Gaussian white noise, show either spontaneous line splitting (at low noise levels) or appreciable frequency shifting (at moderate noise levels). This difficulty arises because an unnecessary constraint is imposed during the minimization of the prediction error power. When the constraint is relaxed and a lighter one imposed, the error power decreases and the problem is solved. The nature of the constraint is discussed, and the mathematical details of the new method are presented. The new method is verified by using a few simple test cases in which spontaneous line splitting is healed or frequency shifting is reduced drastically (Fougere, 1975).

INTRODUCTION

In a recent paper, hereinafter referred to as paper I, *Fougere et al.* [1976] have shown that under certain conditions the maximum entropy method using *Burg*'s [1975] prediction error coefficients (BPEC) produces power spectra which display spurious line splitting in the presence of low noise. For general information on the maximum entropy technique, see also the papers by *Smylie et al.* [1973], *Ulrych and Bishop* [1975], and *Radoski et al.* [1975]. It will be shown here that splitting occurs only in the low-noise case; when the noise level is gradually increased, the spectrum is broadened and the multiple peaks coalesce into a single peak shifted substantially away from the correct value.

For a single sinusoidal signal in the presence of Gaussian noise the conditions most favorable to splitting-shifting are an initial phase of 45° and a length equal to an odd number of quarter cycles. Splitting does not occur as a result of using an overly large number of filter weights, as is frequently claimed in the literature [e.g., *Chen and Stegen*, 1974; *Currie*, 1974]. For example, it was shown in paper I by using 20 filter weights with a data sample consisting of one cycle of a sine wave sampled 21 times with initial phase varying between 0 and 180° that the spectra are extremely sharp, stable, and accurate single lines. *Ulrych and Bishop* [1975] claim on intuitive grounds that a 50% ratio of the number of weights to the largest possible number should be considered an absolute upper limit. Conversely, when only 25 of a possible 101 weights were used, splitting occurred in every case for sine waves sampled 101 times in 1 s with an initial phase of 45° and a frequency varying from 1.25 Hz through 49.25 Hz in steps of 2 Hz.

Chen and Stegen [1974] claim that splitting occurs as a result of amplification of noise when the number of filter weights is too large. In fact, the determining parameter is not a high number of filter weights, and in addition, splitting occurs only in the low-noise case. *Chen and Stegen* [1974] also claim that the frequency shifting sometimes observed in BPEC spectra is similar to that described by *Jackson* [1967]. In actual fact, Jackson's worst cases were sine waves an even number of cycles long with either 0° or 90° initial phase. For these two cases, BPEC spectra are neither split nor shifted but are extremely accurate. Thus the problem of splitting or shifting is unique to the maximum entropy method and has very little in common with problems besetting other power spectral estimation techniques.

The following section will present the mathematical details of the method which corrects the splitting-shifting problem but the general idea will be presented here.

In Burg's method a prediction error filter is obtained as follows. The highest-order coefficient is determined by minimizing the error power output from the filter in both time directions. The remaining coefficients are determined by using Levinson's algorithm [*Levinson*, 1947], which involves previously determined prediction error filters. This procedure is sufficient to ensure that the highest-order coefficient has a magnitude less than 1 and that the Z transform of the filter has all its zeroes outside the unit circle. This procedure is not, however, necessary. When we consider the triangular matrix containing the prediction error filters as rows, it is necessary and sufficient that all diagonal elements (called the reflection coefficients) have a magnitude less than 1 and that the remaining elements are determined by using the Levinson algorithm. If the entire matrix is redetermined each time a new prediction error filter is found and if the diagonal elements are thought of as independent variables restricted to have magnitude less than 1, then the minimization problem has more degrees of freedom, and a lower minimum can be found iteratively by beginning with some set of coefficients as a first approximation. The Burg coefficients will usually serve as reasonable starting values for the iteration. The diagonal elements are made to lie in the correct range by setting each one equal to a constant times the sine of a real angle. The use of this procedure has solved every problem of spontaneous line splitting thus far attempted in the low-noise case and has corrected all excessive line shifting in the moderate-noise case.

In most practical cases with real data the Burg coefficients afford an excellent starting point. Convergence to the final nonlinear solution occurs quite quickly after a very small number of iterations. However, there are situations, for example, sinusoidal signals in the presence of low noise, in which the Burg coefficients are not even good first approximations, and convergence from them may be quite slow, requiring many tens of iterations. In these cases it may be better to start with a previously converged nonlinear set and to increase the number of filter coefficients gradually. This procedure has increased the speed of convergence in many cases by a factor of 2 or more.

Another procedure has been suggested independently by *Ulrych and Clayton* [1976]. They remove all constraints and solve the normal equations for the prediction error filters. This procedure is much faster than the present nonlinear procedure. However, as *Burg* [1975] has pointed out, the procedure does not necessarily produce a true prediction error filter. As the number of filter weights is increased, it becomes more and more likely that one or more of the roots will lie inside the unit circle. When that happens, the filter is unstable and gives an incorrect spectrum. If, on the other hand, the roots all lie outside the unit circle, the unconstrained results should be identical (except for roundoff and other numerical errors) to the results of the new nonlinear technique. The nonlinear technique guarantees that the solution is always a true prediction error filter having all its roots lying outside the unit circle. This will be true even if the method fails to converge because by construction the diagonal elements (the reflection coefficients) all have magnitudes less than 1.

In a sense, the new method lies between the unconstrained least squares method of Ulrych and Clayton and the overly constrained method of Burg. The constraint in the new method is precisely necessary and sufficient to produce a true prediction error filter which yields the lowest possible error power.

Paper number 6A0907.

DETAILED MATHEMATICS

We are given an n-point data sample $(x_1, x_2, \cdots, x_n)$ measured at equally spaced values of an independent variable, usually considered to be time. Define an $(m+1)$-point prediction error filter (PEF) $(1, g_{m1}, g_{m2}, \cdots, g_{mm})$ such that the kth prediction errors are

$$\epsilon_{1k} = \sum_{i=0}^{m} x_{k+m-i} g_{mi}$$

$$\epsilon_{2k} = \sum_{i=0}^{m} x_{k+i} g_{mi} \qquad k = 1, 2, 3, \cdots, l-m \tag{1}$$

where $g_{m0} = 1$ and ϵ_{1k} and ϵ_{2k} are the forward and backward prediction errors, respectively. Note that while the prediction filter associated with the PEF can be used to make predictions into the future beyond x_n and into the past before x_1, it is here used only over the data sample as given.

Now the mean square prediction error, or mean error power, in both time directions is

$$P_m = 0.5(n-m)^{-1} \sum_{j=1}^{2} \sum_{k}^{n-m} \epsilon_{jk}^2 \tag{2}$$

If the PEF's (with the leading '1' suppressed) of all orders 1, 2, $\cdots$, m are gathered in one matrix G_m, we may write

$$G_m = \begin{bmatrix} g_{11} & & & \\ g_{21} & g_{22} & & \\ & \cdot & & \\ & \cdot & & \\ & \cdot & & \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{bmatrix} \tag{3}$$

It has been shown by *Levinson* [1947] that the off-diagonal elements of G_m are determined from the diagonal elements by using the Levinson algorithm:

$$g_{jk} = g_{j-1,k} + g_{jj} g_{j-1,j-k} \tag{4}$$

Burg [1975] has shown that if the diagonal elements of G_m lie in the range $-1 < g_{jj} < 1$, then the PEF is minimum phase, that is, its Z transform has all its zeroes outside the unit circle. In order to enforce this condition, set

$$g_{jj} = U \sin \theta_j \tag{5}$$

where θ_j is any real angle and U is a positive constant slightly less than unity. The actual value of U to be employed in practice depends upon the word length of the computer on which the calculation is carried out. The present calculations were made by using the CDC 6600 with a word length of 60 bits; $U = 0.999999$ was found to be satisfactory. If now the entire set of $(g_{jj}, j = 1, m)$ is varied, the PEF will always be minimum phase, and we can find an absolute minimum of P_m.

A satisfactory value of U is a positive number, close to but less than 1, such that the roots of the Z transform of the PEF will, in fact, all lie outside the unit circle. If the roots are determined numerically in the case of a sine wave with low noise, then U can be adjusted to keep the root closest to unity outside the unit circle. For example, if $U = 0.999999$ were to yield a root equal to 1 in magnitude, then a smaller value, say, $U = 0.99999$, would be tried. The procedure would be repeated until a small enough value of U were found such that all of the roots would always lie outside the unit circle. The satisfactory value of U is obviously a sensitive function of the number of bits in each computer word.

By following a method used by *Cain et al.* [1967], variations in θ_j will be written as

$$\theta_j = \theta_j^0 + \Delta\theta_j \tag{6}$$

Now expand the prediction errors ϵ_{jk} in a Taylor series about θ_j and retain only the first two orders:

$$\epsilon_{jk} = \epsilon_{jk}^0 + \sum_{j=1}^{m} \frac{\partial \epsilon_{jk}}{\partial \theta_j} \Delta\theta_j \tag{7}$$

Substitute (7) into (2) to get

$$P_m = 0.5(n-m)^{-1} \sum_{j=1}^{2} \sum_{k}^{n-m} \left(\epsilon_{jk}^0 + \sum_{j=1}^{m} \frac{\partial \epsilon_{jk}}{\partial \theta_j} \Delta\theta_j \right)^2 \tag{8}$$

Set $(\partial P_m / \partial \Delta\theta_\alpha) = 0$ to find the minimum error power and then rearrange the resulting equation slightly to get

$$\sum_{j=1}^{m} \sum_{i=1}^{2} \sum_{k}^{n-m} \frac{\partial \epsilon_{ik}}{\partial \theta_\alpha} \frac{\partial \epsilon_{ik}}{\partial \theta_j} \Delta\theta_j = -\sum_{i=1}^{2} \sum_{k}^{n-m} \epsilon_{ik}^0 \frac{\partial \epsilon_{ik}}{\partial \theta_\alpha} \tag{9}$$

$$\alpha = 1, 2, \cdots, m$$

Equation (9) is now linear in the set of corrections $\Delta\theta_j$ and can therefore be solved by standard matrix methods. The corrections are substituted into (6), and the process is repeated until the corrections $\Delta\theta_j$ become sufficiently small.

In order to find the derivative $(\partial \epsilon_{ik} / \partial \theta_j)$ we differentiate (1) as follows:

$$\frac{\partial \epsilon_{1k}}{\partial \theta_j} = \sum_{i=1}^{m} x_{k+m-i} \frac{\partial g_{mi}}{\partial \theta_j}$$

$$\frac{\partial \epsilon_{2k}}{\partial \theta_j} = \sum_{i=1}^{m} x_{k+i} \frac{\partial g_{mi}}{\partial \theta_j} \tag{10}$$

where the summations now begin at $i = 1$ because $g_{m0} = 1$ and its derivative vanishes.

Since $g_{jj} = U \sin \theta_j$,

$$\frac{\partial}{\partial \theta_j} = \frac{\partial g_{jj}}{\partial \theta_j} \frac{\partial}{\partial g_{jj}} = U \cos \theta_j \frac{\partial}{\partial g_{jj}} \tag{11}$$

By using the Levinson algorithm we can find all necessary derivatives:

$$\frac{\partial g_{ik}}{\partial g_{jj}} = \frac{\partial g_{i-1,k}}{\partial g_{jj}} + g_{ii} \frac{\partial g_{i-1,i-k}}{\partial g_{jj}} + g_{i-1,i-k} \delta_{ij} \tag{12}$$

For example, $\partial g_{21} / \partial g_{11} = 1 + g_{22} \partial g_{21} / \partial g_{22} = g_{11}$, $\partial g_{22} / \partial g_{11} = 0$, and $\partial g_{mm} / \partial g_{mm} = 1$. Note also that $\partial g_{ik} / \partial g_{jj} = 0$ if $j > i$.

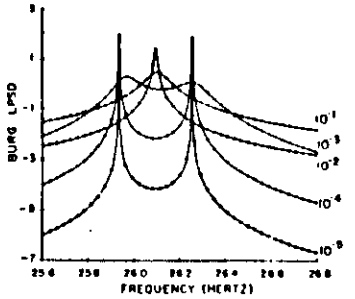


Fig. 1. Burg log power spectral density (LPSD) versus frequency in hertz for signals consisting of 101 samples, 0.01 s apart, of a unit amplitude sine wave of 26.25-Hz frequency and 45° initial phase in additive Gaussian white noise with amplitudes 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. Five filter weights are used.

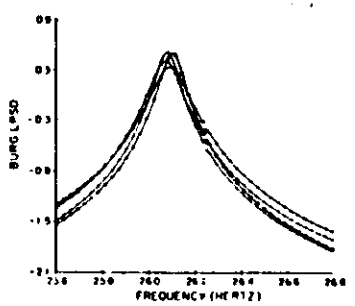


Fig. 3. Burg log spectra of signals similar to those in Figure 1 except that the noise level is 0.1 and four different noise realizations have been used.

If we need the gradient of P_m with respect to the independent variables θ_n, we can write

$$\nabla_\theta P_m = \sum_{n=1}^{m} \frac{\partial P_m}{\partial \theta_n} \hat{\theta}_n \tag{13}$$

where $\hat{\theta}_n$ is a unit vector and

$$\frac{\partial P_m}{\partial \theta_n} = (n-m)^{-1} \sum_{j=1}^{n} \sum_{k=1}^{m} \epsilon_{nk} \frac{\partial \epsilon_{jk}}{\partial \theta_n}$$

$$= -(n-m)^{-1} \times \text{(right side of (9))} \tag{14}$$

i.e., we have already computed $(\partial P_m/\partial \theta_n)$.

Since we can calculate P_m and its gradient $\nabla_\theta P_m$, we can also minimize P_m by using standard methods if difficulty is encountered in using (9). In practice, both methods are used in alternation; by using this technique, all test cases tried so far have converged satisfactorily. Satisfactory convergence is obtained if the minimum so derived is really smaller than the value of the function when each independent variable is varied in a small increment on both sides of the minimum and if, in addition, the magnitude of the gradient at the minimum is smaller than that on either side in any direction.

Results on Test Cases

According to the results of paper 1, the worst splitting-shifting occurs for portions of sine waves an odd number of quarter cycles long with initial phase of 45°. Therefore the test case to be studied here is a unit amplitude sine wave with a frequency of 26.25 Hz and an initial phase of 45° sampled 101 times in 1 s (the Nyquist frequency is 50 Hz). In paper 1 this case was studied by using 25 filter weights and a noise level of 0.0001. In order to understand the effect of changing noise level we will use the same signal with Gaussian noise of amplitudes 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} but only 5 filter weights.

Figure 1 shows the Burg results. With noise levels of 10^{-5} and 10^{-4}, sharp doublets are observed. With a noise level of 10^{-3} the doublet is quite flat, and with noise levels of 10^{-2} and 10^{-1} the two peaks have coalesced into one. Except for the high degree of symmetry displayed in Figure 1, these results are typical: with low noise the spectrum displays a sharp doublet which becomes flatter and disappears as the noise level is raised. Note that the same noise realization was used in each case so that the effect of noise level could be studied.

The new method was then applied to all five of these cases with the results given in Figure 2. All splitting was cured in the low-noise cases, and the frequency shift was greatly reduced in the other two cases. Quantitatively, with the highest noise level the frequency shift amounts to 0.150 Hz when the Burg method is used and only 0.022 Hz, about 7 times smaller, when the new method is used.

To see if the results with the highest noise level are typical, the experiments were repeated using four different noise realizations and a noise amplitude of 0.1. Figure 3 gives the Burg results, and Figure 4 gives the results from the new method.

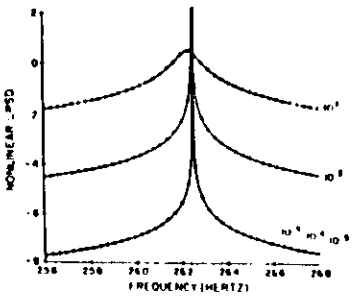


Fig. 2. Nonlinear log spectra of the same signals as in Figure 1.

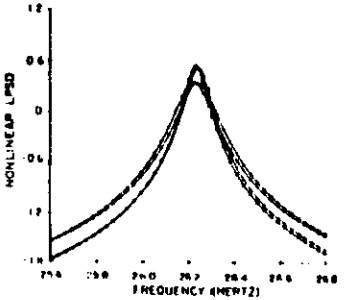


Fig. 4. Nonlinear log spectra of the same signals as in Figure 3.

The picture is entirely consistent, the frequency shift is about 0.6% when the Burg method is used and about 0.08% when the new method is used.

Summary and Conclusions

The problem, using Burg's maximum entropy method, of spectral line splitting in the presence of low noise levels and frequency shifting in the presence of moderate noise levels has been solved. By setting each reflection coefficient equal to a constant times the sine of a real angle and then varying all the angles independently, we gain more freedom for a minimization of the mean forward and backward error power. This increased freedom with its consequent reduction in error power is precisely what is required to heal splitting and to correct shifting in the case of sinusoidal signals in the presence of Gaussian white noise.

The splitting-shifting problem is not caused by too high a noise level or by too large a number of prediction error coefficients, and it is not related to the problem of line shifting sometimes observed in fast Fourier transform spectra.

A complete Fortran package which finds Burg spectra and then optionally applies the new nonlinear technique has been prepared. Seriously interested scientists are invited to write to the author for a copy of the program and its documentation.

Acknowledgments. The author thanks Henry Radoski and Edward Zawalick for many stimulating discussions during the evolution of this work. This solution was first presented at the AGU Fall Annual Meeting, December 1975.

The Editor thanks T. J. Ulrych and S. Wernecke for their assistance in evaluating this paper.

References

Burg, J. P., Maximum entropy spectral analysis, Ph.D. thesis, Stanford Univ., Stanford, Calif., 1975.

Cain, J. C., S. J. Hendricks, R. A. Langel, and W. V. Hudson, A proposed model for the international geomagnetic reference field—1965, *J. Geomagn. Geoelec.*, *19*, 335–355, 1967.

Chen, W. Y., and G. R. Stegen, Experiments with maximum entropy power spectra of sinusoids, *J. Geophys. Res.*, *79*, 3019–3022, 1974.

Currie, R. G., Period and Q_w of the Chandler wobble, *Geophys. J. Roy. Astron. Soc.*, *20*, 179–185, 1974.

Fougere, P. F., A solution to a problem of spontaneous line splitting in maximum entropy power spectrum analysis (abstract), *Eos Trans. AGU*, *56*, 1054, 1975.

Fougere, P. F., E. J. Zawalick, and H. R. Radoski, Spontaneous line splitting in maximum entropy power spectrum analysis, *Phys. Earth Planet. Interiors*, *12*, 201–207, 1976.

Jackson, P. L., Truncation and phase relationships of sinusoids, *J. Geophys. Res.*, *72*, 1400–1402, 1967.

Levinson, N., The Wiener RMS (root mean square) error criterion in filter design and prediction, *J. Math. Phys. Cambridge Mass.*, *25*, 261, 1947.

Radoski, H. R., P. F. Fougere, and E. J. Zawalick, A comparison of power spectral estimates and application of the maximum entropy method, *J. Geophys. Res.*, *80*, 619–625, 1975.

Smylie, D. E., G. K. C. Clarke, and T. J. Ulrych, Analysis of irregularities in the earth's rotation, *Methods Comput. Phys.*, *13*, 391–430, 1973.

Ulrych, T. J., and T. N. Bishop, Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys. Space Phys.*, *13*, 183–200, 1975.

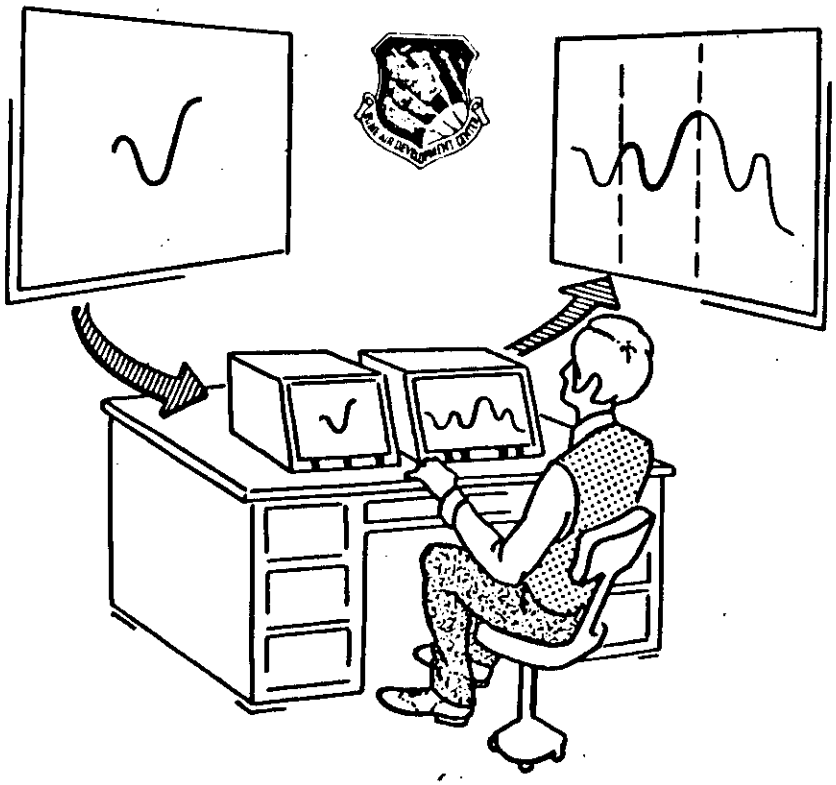
Ulrych, T. J., and R. W. Clayton, Time series modelling and maximum entropy, *Phys. Earth Planet. Interiors*, *12*, 188–200, 1976.

(Received September 7, 1976;
accepted December 2, 1976.)

PROCEEDINGS OF THE

RADC *Spectrum Estimation* WORKSHOP

24, 25 & 26 MAY, 1978



A SOLUTION TO THE PROBLEM OF SPONTANEOUS LINE SPLITTING IN MAXIMUM ENTROPY POWER SPECTRUM ANALYSIS OF COMPLEX SIGNALS

PAUL F. FOUGERE

Air Force Geophysics Laboratory
Hanscom AFB, Bedford, MA 01731

Abstract

Under certain conditions, Burg's maximum entropy spectra of real signals in the presence of additive noise show either spontaneous line splitting (low noise levels) or appreciable frequency shifting (at moderate noise levels). This difficulty arises because an unnecessary constraint is imposed during the minimization of the mean error power. A similar problem arises when the Burg technique is applied to complex signals. This paper presents a solution to this latter problem.

Introduction

In a recent paper, Fougere et al [1], have shown that under certain conditions the maximum entropy method using Burg's prediction error coefficients produces power spectra which display spurious line splitting in the presence of very low noise. In a second paper hereinafter called Paper II, Fougere [2] showed that this type of splitting occurred only if the noise level were sufficiently low; when the noise level is gradually increased the spectrum is broadened and the multiple peaks coalesce into a single peak shifted substantially away from the correct value. In Paper II, Fougere presented a solution to that problem and showed that using that solution, splitting was cured in the low noise case and shifting was reduced considerably in the moderate noise cases. These two papers treated only the case of real input time series, and therefore real prediction error coefficients.

For general information on the maximum entropy technique see the PhD thesis by Burg [3] and papers by Smylie et al [4], Ulrych and Bishop [5], Ulrych and Clayton [6], and Radoski et al [7].

The present paper extends these results to the case of a complex time series requiring complex prediction error coefficients. The first observations of spontaneous splitting with

a complex signal in the presence of very low noise were by R.W. Herring (private communication). This paper will follow closely the structure given in the "Detailed Mathematics" section of paper II. Each of the equations in paper II will be written in the appropriate complex form. The equation numbers will be the same.

Detailed Mathematics

We are given an n-point sample $(x_1, x_2, \ldots, x_n)$ of complex numbers x_i, measured at equally spaced values of a single real independent variable, usually considered to be time. Define an (m+1) point prediction error filter (PEF) $(1, g_{m1}, g_{m2}, \ldots, g_{mm})$ where each g_{ij} is a complex variable, such that the k'th prediction errors are:

$$\epsilon_{1k} = \sum_{i=0}^{m} x_{k+m-i}\, g_{mi}$$

$$\epsilon_{2k} = \sum_{i=0}^{m} x_{k+i}\, g^{*}_{mi} \qquad (1)$$

$$k = 1, 2, 3, \ldots, i-m$$

where g^{*}_{mi} is the complex conjugate of g_{mi}, $g_{mo} \equiv 1$, and ϵ_{1k} and ϵ_{2k} are the forward and backward prediction errors, respectively.

Now the mean square prediction error, or mean error power, in both time directions is:

$$P_m = 0.5\,(n-m)^{-1} \sum_{s=1}^{2} \sum_{k=1}^{n-m} \epsilon_{sk}\epsilon^{*}_{sk} \qquad (2)$$

If the PEF's (with leading "1" suppressed) of all orders 1, 2, ..., m are gathered in one complex matrix G_m, we may write:

$$G_m = \begin{pmatrix} g_{11} & & & \\ g_{21} & g_{22} & & \\ \vdots & & & \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix} \qquad (3)$$

The generalization of the Levinson Algorithm is given by

$$g_{jk} = g_{j-1,k} + g_{jj}\, g^{*}_{j-1,j-k} \qquad (4)$$

This simple, two term formula allows the off diagonal elements of the jth row of G_m to be determined wherever the diagonal elements $(g_{jj},\ j=1,m)$ are known.

Burg has shown that if these diagonal elements (also called reflection coefficients) all lie in the range $|g_{jj}|<1$, then the PEF is minimum phase, that is its Z transform has all its zeroes outside the unit circle.

In order to enforce this condition we set

$$g_{jj} = U \sin\theta_j\, e^{i\phi j} \qquad (5)$$

where θ_j and ϕ_j are any real angles and U is a positive constant slightly less than unity. The discussion in paper II on the significance of U, can be carried over unchanged to the present case of complex input data. Briefly, U is adjusted so that all of the roots of the Z transform of the PEF all lie <u>outside</u> the unit circle and none lie <u>on</u> it.

We now follow a method used by Cain et al [8]; variations in θ_j and ϕ_j are written

$$\theta_j = \theta^{o}_j + \Delta\theta_j$$

$$\phi_j = \phi^{o}_j + \Delta\phi_j \qquad (6)$$

Now expand the prediction errors ϵ_{sk} in a Taylor series about θ^{o}_j

and ϕ_j^o and retain only the first two orders:

$$\epsilon_{sk} = \epsilon_{sk}^o + \sum_{j=1}^{m}\left(\frac{\partial \epsilon_{sk}}{\partial \theta_j}\Delta\theta_j + \frac{\partial \epsilon_{sk}}{\partial \phi_j}\Delta\phi_j\right) \qquad (7)$$

Substitute (7) into (2) to get:

$$P_m = 0.5(n-m)^{-1}\sum_{s=1}^{2}\sum_{k=1}^{n-m}\left|\epsilon_{sk}^o + \sum_{j=1}^{m}\left(\frac{\partial \epsilon_{sk}}{\partial \theta_j}\Delta\theta_j + \frac{\partial \epsilon_{sk}}{\partial \phi_j}\Delta\phi_j\right)\right|^2 \qquad (8)$$

Set $\partial P_m/\partial\Delta\theta_\alpha = 0$ to find the minimum error power and then rearrange the resulting equations to get:

$$\sum_{j=1}^{m}\sum_{s=1}^{2}\sum_{k=1}^{n-m}\left[\left(\frac{\partial \epsilon_{sk}}{\partial \theta_j}\frac{\partial \epsilon_{sk}^*}{\partial \theta_\alpha} + \frac{\partial \epsilon_{sk}^*}{\partial \theta_j}\frac{\partial \epsilon_{sk}}{\partial \theta_\alpha}\right)\Delta\theta_j + \left(\frac{\partial \epsilon_{sk}}{\partial \phi_j}\frac{\partial \epsilon_{sk}^*}{\partial \theta_\alpha} + \frac{\partial \epsilon_{sk}^*}{\partial \phi_j}\frac{\partial \epsilon_{sk}}{\partial \theta_\alpha}\right)\Delta\phi_j\right] = -\sum_{s=1}^{2}\sum_{k=1}^{n-m}\left(\epsilon_{sk}^o\frac{\partial \epsilon_{sk}^*}{\partial \theta_\alpha} + \epsilon_{sk}^{o*}\frac{\partial \epsilon_{sk}}{\partial \theta_\alpha}\right)$$

There are three expression in parentheses. Note that each has the form $A + A^* = 2\,\mathrm{Re}\{A\}$. Thus the equation becomes

$$\sum_{j=1}^{m}\sum_{s=1}^{2}\sum_{k=1}^{n-m}\left[\mathrm{Re}\left(\frac{\partial \epsilon_{sk}}{\partial \theta_j}\frac{\partial \epsilon_{sk}^*}{\partial \theta_\alpha}\right)\Delta\theta_j + \mathrm{Re}\left(\frac{\partial \epsilon_{sk}}{\partial \phi_j}\frac{\partial \epsilon_{sk}^*}{\partial \theta_\alpha}\right)\Delta\phi_j\right] = -\sum_{s=1}^{2}\sum_{k=1}^{n-m}\mathrm{Re}\left(\epsilon_{sk}^o\frac{\partial \epsilon_{sk}^*}{\partial \theta_\alpha}\right) \qquad (9a)$$

Similarly when we set $\partial P_m/\partial\Delta\phi_\alpha = 0$ we get

$$\sum_{j=1}^{m}\sum_{s=1}^{2}\sum_{k=1}^{n-m}\left[\mathrm{Re}\left(\frac{\partial \epsilon_{sk}}{\partial \theta_j}\frac{\partial \epsilon_{sk}^*}{\partial \phi_\alpha}\right)\Delta\theta_j + \mathrm{Re}\left(\frac{\partial \epsilon_{sk}}{\partial \phi_j}\frac{\partial \epsilon_{sk}^*}{\partial \phi_\alpha}\right)\Delta\phi_j\right] = -\sum_{s=1}^{2}\sum_{k=1}^{n-m}\mathrm{Re}\left(\epsilon_{sk}^o\frac{\partial \epsilon_{sk}^*}{\partial \phi_\alpha}\right) \qquad (9b)$$

where, in both (9a) and (9b): $\alpha = 1, 2, \ldots, m$.

The 2m equations (9) are now <u>linear</u> in the corrections $\Delta\theta_j$ and $\Delta\phi_j$ and can therefore be solved by standard matrix methods. The corrections are then substituted into (6) and the process is repeated until the corrections $\Delta\theta_j$ and $\Delta\phi_j$ become sufficiently small.

In order to find the derivatives: $\partial \epsilon_{sk}/\partial\theta_j$, $\partial \epsilon_{sk}^*/\partial\theta_j$, $\partial \epsilon_{sk}/\partial\phi_j$ and $\partial \epsilon_{sk}^*/\partial\phi_j$, we differentiate (1) as follows:

$$\frac{\partial \epsilon_{1k}}{\partial \theta_j} = \sum_{i=1}^{m} x_{k+m-i}\frac{\partial g_{mi}}{\partial \theta_j}\ ;\ \frac{\partial \epsilon_{1k}}{\partial \phi_j} = \sum_{i=1}^{m} x_{k+m-i}\frac{\partial g_{mi}}{\partial \phi_j}$$

$$\frac{\partial \epsilon_{2k}}{\partial \theta_j} = \sum_{i=1}^{m} x_{k+i}\frac{\partial g_{mi}^*}{\partial \theta_j}\ ;\ \frac{\partial \epsilon_{2k}}{\partial \phi_j} = \sum_{i=1}^{m} x_{k+i}\frac{\partial g_{mi}^*}{\partial \phi_j} \qquad (10)$$

and four equations resulting from (10) by complex conjugation, for example

$$\frac{\partial \epsilon_{1k}^*}{\partial \theta_j} = \sum_{i=1}^{m} x_{k+m-i}^*\frac{\partial g_{mi}^*}{\partial \theta_j}$$

Next we rewrite eq. (5) and its complex conjugate

$$g = U\sin\theta e^{i\phi}$$
$$g^* = U\sin\theta e^{-i\phi} \qquad (A)$$

where we have temporarily dropped the subscripts for convenience. Equations (A) may be inverted to yield:

$$\theta = \theta(g, g^*)$$
$$\phi = \phi(g, g^*) \qquad (B)$$

If we treat g and g^* as independent variables, we can express the partial derivatives with respect to θ and ϕ as follows:

$$\frac{\partial}{\partial\theta} = \frac{\partial g}{\partial\theta}\frac{\partial}{\partial g} + \frac{\partial g^*}{\partial\theta}\frac{\partial}{\partial g^*}$$
$$\frac{\partial}{\partial\phi} = \frac{\partial g}{\partial\phi}\frac{\partial}{\partial g} + \frac{\partial g^*}{\partial\phi}\frac{\partial}{\partial g^*} \qquad (C)$$

Substituting $e^{i\phi} = \cos\phi + i\sin\phi$ in (A) and performing the differentiation indicated in (C) we arrive at:

$$\frac{\partial}{\partial\theta} = U\cos\theta\left[\cos\phi\left(\frac{\partial}{\partial g} + \frac{\partial}{\partial g^*}\right) + i\sin\phi\left(\frac{\partial}{\partial g} - \frac{\partial}{\partial g^*}\right)\right]$$
$$\frac{\partial}{\partial\phi} = U\sin\theta\left[-\sin\phi\left(\frac{\partial}{\partial g} + \frac{\partial}{\partial g^*}\right) + i\cos\phi\left(\frac{\partial}{\partial g} - \frac{\partial}{\partial g^*}\right)\right] \qquad (D)$$

Write

$$D_j^{\pm} \equiv \frac{\partial}{\partial g_{jj}} \pm \frac{\partial}{\partial g_{jj}^*} \qquad (E)$$

Then

$$\frac{\partial}{\partial\theta_j} = U\cos\theta_j(\cos\phi_j\, D_j^+ + i\sin\phi_j\, D_j^-)$$
$$\frac{\partial}{\partial\phi_j} = U\sin\theta_j(-\sin\phi_j\, D_j^+ + i\cos\phi_j\, D_j^-) \qquad (11)$$

Note that because g_{jj} and g_{ii}^* are independent variables, $\partial g_{jj}/\partial g_{ii}^* = 0$ and $\partial g_{jj}^*/\partial g_{ii} = 0$, for any values of i and j.

We now apply (E) to the Levinson Algorithm (4). The result is

$$D_j^{\pm} g_{ik} = D_j^{\pm} g_{i-1,k} + g_{i-1,i-k}^*\,\delta_{ij} + g_{ii}\, D_j^{\pm} g_{i-1,i-k}^* \qquad (12)$$

Where the middle term arises because

$$D_j^{\pm} g_{ii} = \left(\frac{\partial}{\partial g_{jj}} \pm \frac{\partial}{\partial g_{jj}^*}\right) g_{ii} = \delta_{ij} \qquad (F)$$

If we take the complex conjugate of equation (E) we get:

$$(D_j^{\pm})^* = \frac{\partial}{\partial g_{jj}^*} \pm \frac{\partial}{\partial g_{jj}} = \pm D_j^{\pm} \qquad (G)$$

Therefore, anytime we need derivatives of any terms in g^* we can write

$$D_j^{\pm} g_{km}^* = \pm(D_j^{\pm} g_{km})^* \qquad (H)$$

Finally, the derivatives with respect to θ_j and ϕ_j are determined by substituting (12) into (11).

The gradient of P_m with respect to the independent variables θ_K and ϕ_K is written:

$$\nabla_{\theta,\phi} P_m = \sum_{\alpha=1}^{m}\left(\frac{\partial P_m}{\partial\theta_\alpha}\hat{\theta}_\alpha + \frac{\partial P_m}{\partial\phi_\alpha}\hat{\phi}_\alpha\right) \qquad (13)$$

where $\hat{\theta}_\alpha$ and $\hat{\phi}_\alpha$ are unit vectors.
Applying (13) to (2) yields:

$$\frac{\partial P_m}{\partial\theta_\alpha} = 0.5(n-m)^{-1}\sum_{s=1}^{2}\sum_{k=1}^{n-m}\epsilon_{sk}\frac{\partial\epsilon_{sk}^*}{\partial\theta_\alpha} + \epsilon_{sk}^*\frac{\partial\epsilon_{sk}}{\partial\theta_\alpha}$$
$$= (n-m)^{-1}\sum_{s=1}^{2}\sum_{k=1}^{n-m}\mathrm{Re}\left(\epsilon_{sk}\frac{\partial\epsilon_{sk}^*}{\partial\theta_\alpha}\right)$$

Thus,

$$\frac{\partial P_m}{\partial\theta_\alpha} = -(n-m)^{-1} \times (\text{right side of (9a)}) \qquad (14a)$$

milarly,

$$\frac{\partial P_m}{\partial \theta_\alpha} = -(n-m)^{-1} \times \text{(right side of (9b))} \qquad (14b)$$

This completes the formal derivation. Clearly these sults must be programmed for a computer before we can test the thod. When such a program has been written and checked out, will be made available on request to seriously interested lentists.

References

Fougere, P.F., E.J. Zawalick, and H.R. Radoski, 1976, "Spontaneous line splitting in maximum entropy power spectrum analysis", Phys. Earth Planet. Interiors, 12, P. 201-207.

Fougere, P.F., 1977, "A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis", J. Geophys. Res., 82, P. 1051-1054.

Burg, J.P., 1975, "Maximum entropy spectral analysis", Ph.D. thesis, Stanford Univ., Stanford, CA.

Smylie, D.E., G.K.C. Clarke and T.J. Ulrych, 1973, "Analysis of irregularities in the earth's rotation", Methods Comput. Phys., 13, P. 391-430.

Ulrych, T.J. and T.N. Bishop, 1975, "Maximum entropy spectral analysis and autoregressive decomposition", Rev. Geophys. Space Phys., 13, P. 183-200.

Ulrych, T.J. and R.W. Clayton, 1976, "Time series modelling and maximum entropy", Phys. Earth Planet. Interiors, 12, P. 188-200.

Radoski, H.R., P.F. Fougere and E.J. Zawalick, 1975, A comparison of power spectral estimates and application of the maximum entropy method", J. Geophys. Res., 80, P. 619-625.

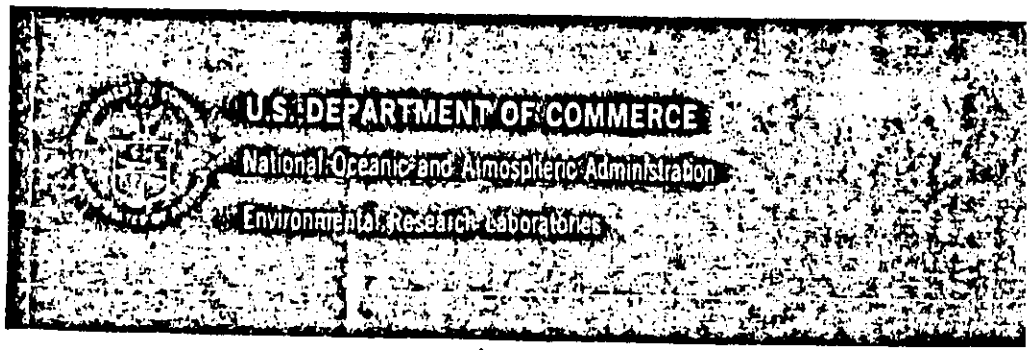
Cain, J.C., S.J. Hendricks, R.A. Langel, and W.V. Hudson, 1967, "A proposed model for the international geomagnetic reference field - 1965", J. Geomagn. Geoelec., 19, P. 335-355.

Solar-Terrestrial Predictions Proceedings

Volume 3: Solar Activity Predictions

Richard F. Donnelly, Editor

Space Environment Laboratory
Boulder, Colorado



SUNSPOTS: POWER SPECTRA AND A FORECAST

Paul F. Fougere
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731, U.S.A.

High resolution power spectra of the Zürich sunspot numbers for the period 1849 to 1978 are obtained using the maximum entropy technique of Burg. The monthly means are first smoothed using a digital, least-squares, band-pass filter with 193 weights, and then decimated to obtain a more manageable series. An overall power spectrum, which displays multiple structure and harmonic series is then obtained.

In order to explain the complications in the spectrum, a dynamic spectrum is displayed. This is obtained by finding the spectra of data samples 66 years long, repeatedly slipped by 2 years for a total of 78 spectra. The spectra are then much simpler and vary smoothly with epoch. The nominal 11-year line varies in period from about 10 to 12½ years.

In the process of obtaining a power spectrum a prediction error filter is derived. This linear filter may then be used to make predictions as follows: using sets of data containing 5 solar cycles of unsmoothed monthly values beginning with cycles 9 through 15, predictions of the next 12 months of the time series are made and compared to the observed values. RMS errors vary between 5 and 33 and lead to an expectation that the error of prediction for cycle 21 will be about 20. The entire cycle 21 is then predicted using 50, 100, 150 prediction error coefficients. The maximum of cycle 21 is predicted to be 130 ± 20 at 1980.1 ± 0.2.

Introduction

The subject of this paper is the time series of Zürich sunspot numbers, R_z. The basic data set consists of monthly mean values of R_z beginning in 1749 and running without gaps to October 1978, the latest available value. See Chernosky and Hagan (1958) for the earlier data. There is some smoothing built into the way in which sunspot counts are obtained, because each daily count uses the entire visible hemisphere, which is rotating with a period of 27 days, or a rate of 13 1/3 degrees per day. Thus the daily count is really a 13 1/3-day running mean of those spots and groups which could be observed on a 13 1/3 degree lune centered at central meridian. Despite the low-pass filter, one of the striking features of the series of monthly means of R_z is the large amplitude, high frequency noise which rides

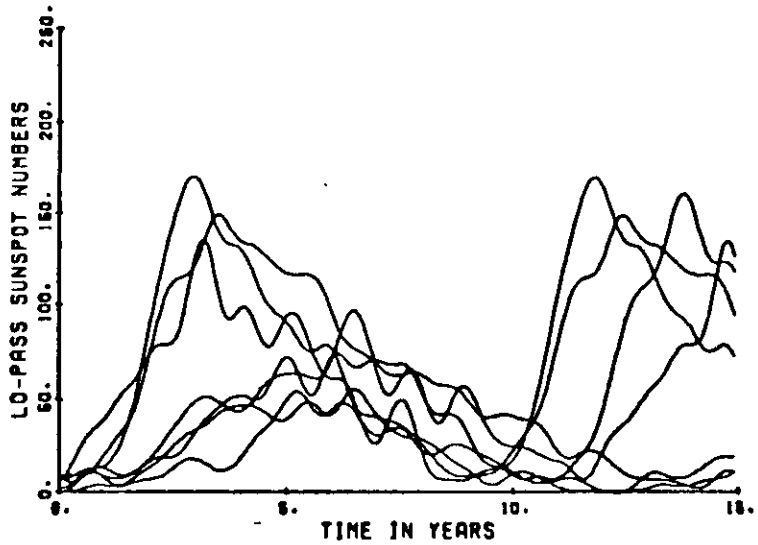


FIGURE 1. SOLAR CYCLES 1-7. DIGITAL LOW-PASS FILTER WITH 13 WEIGHTS HAS BEEN USED. EACH CYCLE BEGINS AT TIME OF SMOOTHED MINIMUM.

on top of the other striking feature, the 11-year period. Accordingly, it has become customary, for example in the monthly Solar Geophysical Data bulletin, to further smooth the monthly averages by using a running mean of 13 monthly values with the 2 extreme months weighted at ½.

So dominant is the appearance of the 11-year cycle that the individual cycles have been assigned numbers beginning with cycle 1 in 1755 to cycle 21, the current cycle, which started in mid-1976. Since the time of minimum can be located with reasonable accuracy, the cycles begin and end at a minimum. Figure 1 shows the first 7 cycles with the time origin set at the time of minimum. These data have been smoothed using a 13-point digital least-squares, low-pass filter following the method of Behannon and Ness (1966). This simple filter has a more nearly accurate low-pass response than that of the 13-point running mean.

Despite the use of the filter, the extreme variability of the data is evident. All of the cycles are bumpy; cycle length, from minimum to minimum, varies between 9 and 15 years; time from minimum to maximum varies from about 2½ to 6 years, and the size of the maximum varies by a factor of 4. These early 7 cycles have been criticized in the classical paper on sunspot prediction by McNish and Lincoln (1949). According to that paper the early data are considerably less reliable and even belong to a different statistical population from the modern data beginning in 1834.

Cycles 8 through 20, the modern data, are shown in Figure 2. For these data, the length of the cycle varies less, between 10 and 12½ years. The

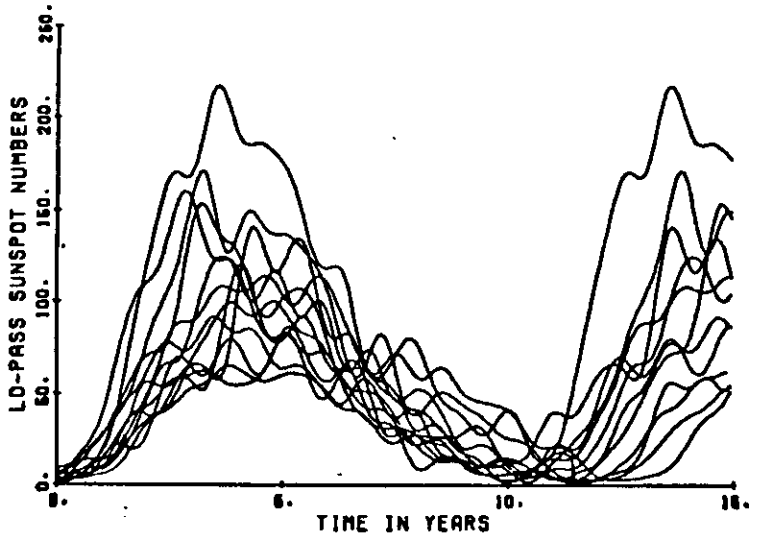


FIGURE 2. SOLAR CYCLES 8-20. DIGITAL LOW-PASS FILTER WITH 13 WEIGHTS HAS BEEN USED. EACH CYCLE BEGINS AT TIME OF SMOOTHED MINIMUM.

time to maximum is now about 2½ to 4½ years, but the height of the maximum still varies by a factor of 4: from about 50 to more than 200.

The extreme variability of these data is further illustrated in Figure 3, showing a 3-dimensional representation of all 20 cycles. Cycle 19, the largest ever, which began in 1954, is so large that it completely hides cycle 20.

Power Spectra

Now we would like to determine the power spectrum of this time series using the maximum entropy method (Burg, 1975) which works admirably on rather short time series. The original 2758 monthly means constitute a rather long series. We can easily remedy this by decimation but only after smoothing to prevent aliasing. Accordingly, a low-pass filter with a cutoff period of 2½ years was designed. At the same time, a high-pass filter was used to remove the mean and very long time trends. The response of the resulting band-pass filter is shown in Figure 4. The digital least-squares filter used 193 weights to achieve this response. Notice the filter beginning to cut off at about 2½ years, as designed, so that the response at 2 years and shorter times (high frequencies) is close to zero. Thus we can sample the output safely once a year and introduce no aliasing because all frequencies higher than the new Nyquist frequency of 0.5 cycles per year have been removed.

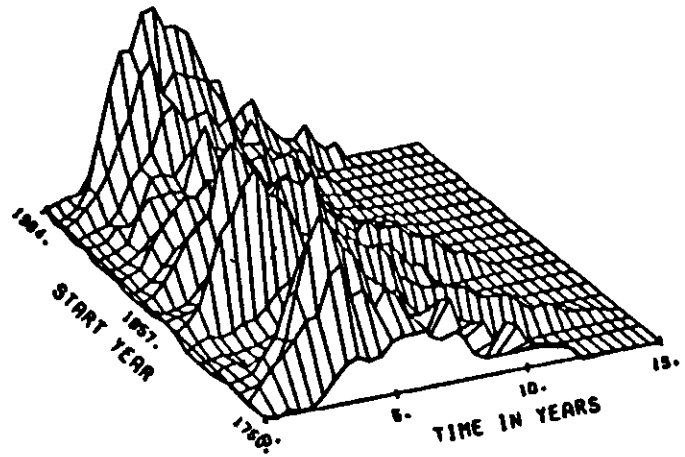


FIGURE 3. PERSPECTIVE PLOT OF CYCLES 1-20.

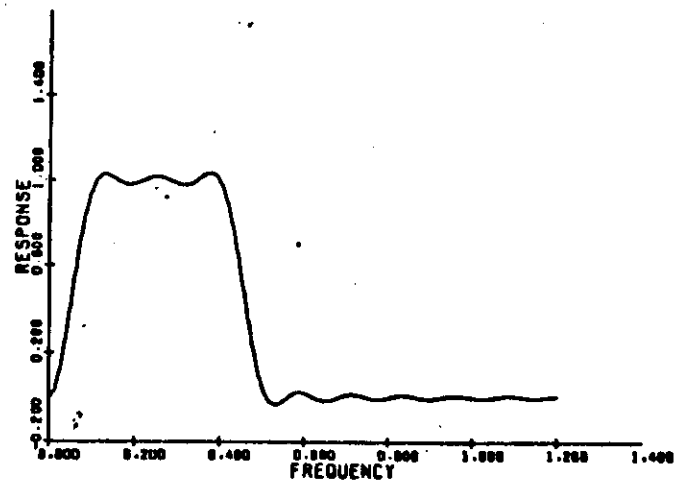


FIGURE 4. RESPONSE OF DIGITAL LEAST-SQUARES, BAND-PASS FILTER. 193 WEIGHTS ARE USED. FREQUENCY IS IN UNITS OF CYCLES PER YEAR.

The results are shown in Figure 5. The original unsmoothed, monthly means are given in the bottom panel with values ranging from 0 to 250 and the dates ranging from the year 1750 to the year 2000. The band-pass filtered output is shown in the top panel, and the differences between the original and band-pass data, the residuals, are given in the middle panel. Thus the original data set in the bottom panel is the sum of the top 2 panels, month by month. The top panel shows a smooth, zero-mean time series which is still not stationary because the amplitudes are quite variable. The residual series, in the central panel, contains essentially all of the high-frequency noise plus the slowly-varying mean or DC level. It is now quite safe to decimate the band-pass filtered series by 12 to obtain 211 yearly values.

The maximum entropy method of Burg (1975) was applied to these 211 numbers and the number of prediction error filter weights, analogous to the number of lags in a Blackman-Tukey spectrum, was varied from 4 to 130. The 3-dimensional representation of these spectra is shown in Figure 6. The nominal 11-year line, which carries most of the power, splits into two very stable lines with periods of about 11.0 and 10.0 years. There is also clear evidence of power peaking at about 12 years and again at about 8½ years. The second harmonic of the 11-year doublet is also a doublet at about 5.5 and 4.8 years. The richness of structure in the spectrum is simply another manifestation of the extreme variability of the time series.

The complicated spectrum also can be explained by taking small overlapping segments of data, performing spectral analysis on each and watching the spectrum change with time. Such a dynamical spectrum is shown in Figure 7. Each spectrum is based on a segment of data 66 years long.

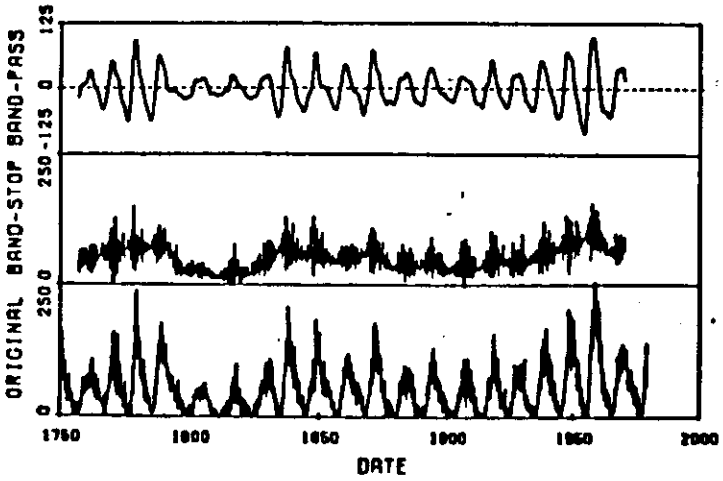


FIGURE 5. SUNSPOT NUMBERS VERSUS DATA FROM 1750 TO PRESENT (OCT. 1978).
TOP PANEL: BAND-PASS FILTERED DATA (193 WEIGHTS).
MIDDLE PANEL: ORIGINAL MINUS BAND-PASS.
BOTTOM PANEL: ORIGINAL, UNSMOOTHED MONTHLY MEANS.

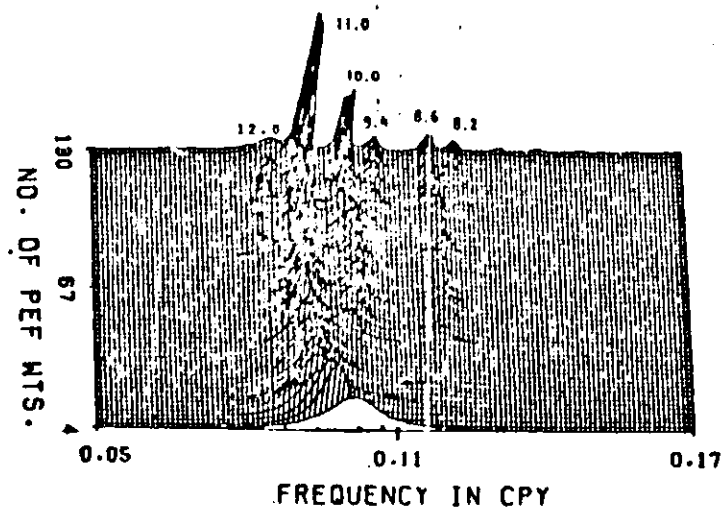


FIGURE 6. PERSPECTIVE PLOT OF MAXIMUM ENTROPY SPECTRA OF BAND-PASS FILTERED DATA SAMPLED ONCE PER YEAR. THE NUMBER OF PREDICTION ERROR FILTER (PEF) WEIGHTS RANGES BETWEEN 4 AND 130 IN STEPS OF 2. PERIODS IN YEARS ARE GIVEN AT APPROPRIATE PEAKS.

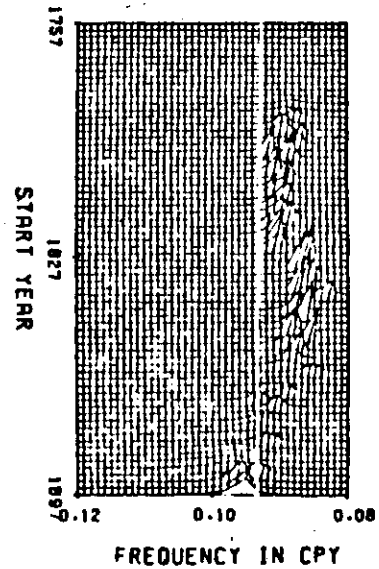


FIGURE 7. DYNAMIC MAXIMUM ENTROPY SPECTRUM OF 66 YEARS (APPROXIMATELY 6 CYCLES) OF BAND-PASS FILTERED DATA SAMPLED ONCE PER YEAR. THE START YEAR ADVANCES BY 2 YEARS FOR EACH SPECTRUM SHOWN.

The segments are overlapped by 64 points, so that the start year moves by 2 years for each new segment. We are now concentrating on a narrow band of periods between 8 and 12½ years. The peaks in the earliest data are too small to be seen in these linear power spectral density plots. The nominal 11-year line shows an instantaneous period varying between 10 and 12½ years while the power spectral density varies by over two orders of magnitude.

Predictions

At the heart of the maximum entropy method is the determination of a prediction error filter. This is a linear filter whose scalar product with a segment of data yields the error in a one-step-ahead prediction. We find the filter essentially by minimizing the sum of squares of such prediction errors over the set of observations. Once we have such a filter we can use it to make predictions off both ends of the series as far as we wish to go simply by applying the filter and moving over the data and the newly acquired predictions.

The predictions of the sunspot numbers were tested as follows. Beginning in 1843, 5 cycles of original unsmoothed monthly values, containing 708 monthly values, constitute a data set for which spectra are obtained for a number of prediction-error filter-weights ranging from 50 to 300 in steps of 50. For each filter, predictions were made for the following 12 months and the RMS prediction error was determined. This is possible because predictions and observed values which had not been used to obtain the predictions are available. The entire procedure was repeated seven times, using 5 cycles of monthly data beginning with cycles 9, 10, 11 through cycle 15, and making predictions into cycles 14 through 20 respectively.

The mean square prediction errors, all quite reasonable numbers from a low of 5 to a high of 33, are shown in Figure 8. Here are plotted the cycle numbers from 14 through 21 and the RMS prediction errors for 12 predictions ranging from 0 through 40. Curiously, the RMS error seems to alternate in magnitude, being low for even cycles, and appreciably higher for odd-numbered cycles. Accordingly a guess is made that the RMS prediction error for the real predictions for cycle 21 will be around 20 for predictions up to 12 months ahead.

Predictions were made using 50 to 300 weights for the entire cycle 21 and plotted. The most consistent behavior for the entire cycle comes from sets with 100, 150, and 200 weights and these are shown in Figure 9. Smoothing these results, the predictions of the maximum and its date become:

$$\text{MAX } (R_z) = 130 \pm 20$$

$$\text{DATE} = 1980.1 \pm 0.2$$

The entire cycle seems quite reasonable with a minimum around 1986.2.

Of the 12 predictions tested for each of the cycles 14 thru 21, the prediction errors were essentially oscillatory. This reflects the noise in the data set rather than any monotonic degradation in the prediction as the prediction time increases. The number 12 was chosen because of the prior

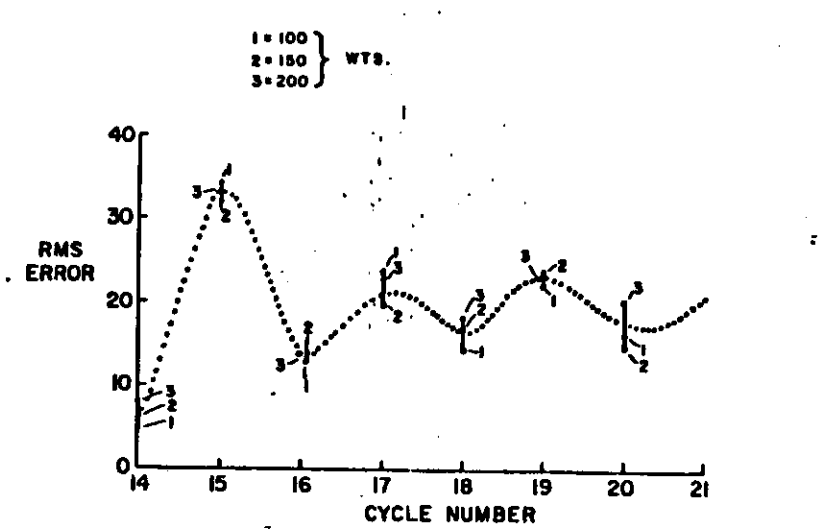


FIGURE 8. RMS ERRORS, FOR PREDICTION OF MONTHLY MEANS 1,2,3,.....,12 MONTHS AHEAD FOR 100, 150 AND 200 WEIGHTS. FIVE SOLAR CYCLES OF ORIGINAL UNSMOOTHED MONTHLY MEAN ARE USED AS INPUT TO THE MAXIMUM ENTROPY POWER SPECTRAL ANALYSIS PROGRAM. NUMBER OF OBSERVATIONS WAS 708, 708, 684, 684, 660, 648 AND 624 RESPECTIVELY.

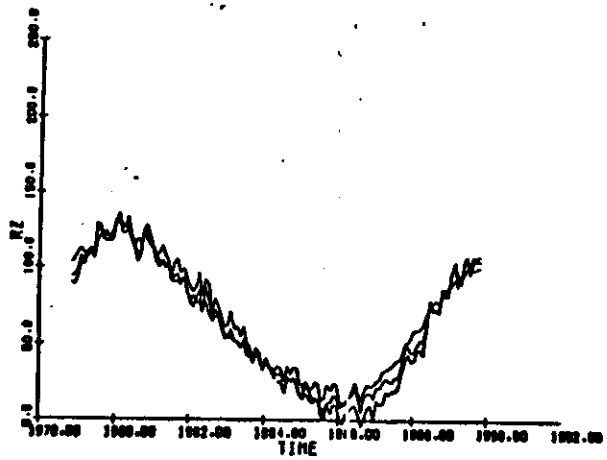


FIGURE 9. PREDICTION OF UNSMOOTHED SUNSPOT NUMBERS MADE USING THREE DIFFERENT SETS OF PREDICTION ERROR FILTERS WITH 100, 150 AND 200 WEIGHTS. DATA USED ARE 670 ORIGINAL UNSMOOTHED MONTHLY MEANS STARTING IN JANUARY, 1923.

expectation that the maximum of cycle 21 was about 12 months away from the date of the latest available observed monthly mean sunspot number (Oct 1978). Thus the expected error should be quite close to 20 even for predictions 14 to 16 months ahead.

In summary, the sunspot data have been displayed and their extreme variability discussed. The data have been filtered with a digital band-pass filter and the resulting series has been decimated. Maximum entropy power spectra vary in a stable manner as the number of weights is varied. A dynamic spectrum showed that the nominal 11-year line varies in period between 10 and 12½ years. A prediction technique has been tested using seven separate data sets each five cycles long. For predictions of monthly means 1,2,3,.....12 months ahead, the RMS prediction error varied between 5 and 35 and lead to the expectation that the RMS prediction error should be about ±20 for the prediction of the maximum of cycle 21. A forecast for the entire cycle has been given, with the maximum expected to be 130±20 at 1980.1 ± 0.2.

The programs used here, the filter design program, the three-dimensional plotting program, and the maximum entropy spectral analysis program, are all available on request to seriously interested scientists. Also available is a small card deck containing all of the sunspot monthly means.

References

Behannon, K.W. and N.F. Ness, (1966), The Design of Numerical Filters for Geomagnetic Data Analyses, NASA TN D-3341.

Burg, J.P., (1975), Maximum Entropy Spectral Analysis, Ph.D. Thesis, Stanford University.

Chernosky, E.J. and M.P. Hagan, (1958), The Zürich Sunspot Number and Its Variations for 1700-1957, J. Geophys. Res., 63, (4).

McNish, A.G. and J.V. Lincoln, (1949), Prediction of Sunspot Numbers, Trans, AGU, 30, 673-685.

Spectral Resolution of Fougere's Maximum Entropy Spectral Analysis

C. H. CHEN

Abstract—Computer results are presented to illustrate the spectral resolution of the nonlinear maximum entropy spectral analysis developed by Fougere. Comparison is also made with the recently published results in this PROCEEDINGS.

I. Introduction

Accurate spectral analysis is required in many applications such as geophysics, radar, sonar, etc. The Burg's maximum entropy spectral analysis [1] is one of several major high resolution spectral analysis methods for short length data records. In the Burg's method a prediction error filter is obtained by minimizing the error power output from the filter in both time directions. For sinusoidal waveforms, the method has the drawbacks of line splitting in low noise case and frequency shifting in high noise case. In the method proposed by Fougere [2], [3], the filter coefficients are redetermined iteratively by starting with the Burg's filter coefficients. The iterative procedure involves nonlinear optimization with respect to the filter coefficients with the constraint that the filter must be stable by requiring the reflection coefficients to have magnitudes less than 1. The constraint is necessary and sufficient to produce a true prediction error filter which yields the lowest possible error power. The Fougere's method not only removes the drawbacks of the Burg's method described above but also provides a much better spectral resolution [4], [5].

II. Computer Results

Fig. 1 shows the spectral analysis of the Zurich Sunspot Number (Fig. 1(a)). With only 20 data points, the Fougere's method can provide as accurate a spectrum (Fig. 1(c)) as that obtained from the Burg's method using a 200 data points (Fig. 1(b)). The nominal 11 year periodicity as reported by Fougere [6] is determined here accurately from both Burg and Fougere spectra.

Manuscript received December 29, 1980; revised February 4, 1981. This work was supported by the Air Force Office of Scientific Research under Grant AFOSR 80-0143 and Contract N00014-79-C-0494.

The author is with the Electrical Engineering Department, Southeastern Massachusetts University, North Dartmouth, MA 02747.

0018-9219/81/0700-0839$00.75

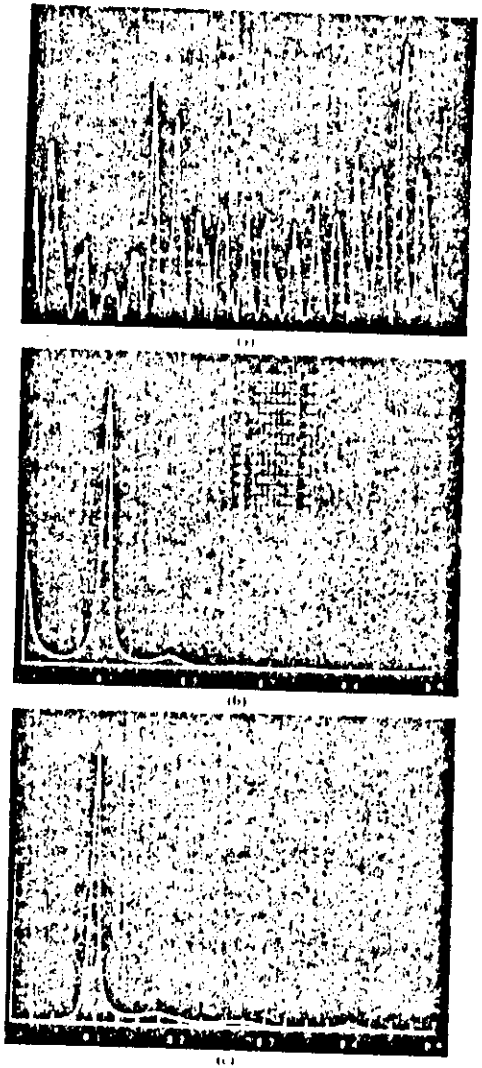


Fig. 1. (a) Plot of Zurich Sunspot Number from 1700 to 1979 at one point per year. Sampling rate 1. (b) Plot of Burg's spectrum for 200 data points and 15 filter weights with linear scale. (c) Plot of Fougere's spectrum (nonlinear method) for 20 data points from 1960 to 1979 at one point per year, with 8 filter weights, also in linear scale.

Now consider the discrete signal with two-frequency sinusoids plus noise given by

$$y(n) = e^{j(2\pi f_1 n + \theta)} + e^{j(2\pi f_2 n)} + b(n), \quad n = 1, 2, \ldots, N$$

where $\theta = 45°$, $f_1 = 0.5$ Hz, $f_2 = 0.52$ Hz, $N = 25$, $b(n)$ is a complex-valued white Gaussian noise sequence and the sampling rate is 1 Hz. The spectral plots (Fig. 2(a)) of the signal with 10 filter weights clearly demonstrate that the Fougere's method (solid curve) can resolve the two frequency components while the Burg's method cannot at 5-dB signal-to-noise ratio[1]. There is no frequency shifting in the Fougere method. Further computer study indicates that the 0.02-Hz separation between the two frequency components is about the minimum allowable. If we consider the real part of the signal only then the frequency separation must be increased to 0.04 Hz with the two frequencies being 0.5 Hz and

[1]Here signal-to-noise ratio is the ratio of signal power to the variance of complex noise sequence.

SPONTANEOUS LINE SPLITTING IN MULTICHANNEL MAXIMUM ENTROPY POWER SPECTRA

PAUL F. FOUGERE

U.S. AIR FORCE GEOPHYSICS LABORATORY
HANSCOM AFB, MA, 01731, U.S.A.

ABSTRACT

The twin problems which plague Burg Maximum Entropy Power Spectra are spectral line shifting of sinusoidal signals in the presence of moderate noise levels, and spontaneous line splitting the presence of low noise levels. The problems and their solutions and some incorrect attempts at solution are discussed for the case of single channel real and complex signals. The multichannel generalizations of the Burg technique have been studied numerically and will be shown to produce serious line shifting for two-channel sinusoidal signals. Thus far no certifiable line splitting has been found.

INTRODUCTION

The Burg Technique[1,2], which determines a maximum entropy power spectrum directly from time series data using forward and backward residuals, has been generalized to apply to multichannel signals. Generalizations were accomplished independently and nearly simultaneously by four different authors: Jones[3], Nuttall[4], Strand[5] and Morf et al.[6]; All these authors except Morf et al. also present FORTRAN programs. The program used in the present research was put together by Currie[7] using parts of Jones' program and parts of Strand's program modified by using some of Morf's ideas. The present author added a set of subroutines which find, for each channel, the accurate location and height of all spectral peaks and the integral under the spectrum. The integral is used in conjunction with Parseval's theorem to check the accuracy of the computed spectrum. The specific results to be presented in this paper employ two real data channels but the program is applicable to any number of channels and has been tested and works correctly for more than two channels.

The single channel splitting/shifting problem and its solution.

In 1974, Chen and Stegen[8] ran some computer experiments, finding Burg maximum entropy power spectra for sinusoids of varying length and initial phase with varying amounts of gaussian white noise. They showed that real problems exist even in the simple case of a single sine-wave plus noise. An excessive shift of the frequency at the peak of the spectrum, away from the true input frequency might occur or occasionally the peak would split into one or more comparably sized components. They showed correctly that the splitting/shifting problem is worst for an initial phase equal to an odd multiple of 45° and a signal length equal to an odd number of quarter cycles. However, they incorrectly attributed splitting to noise amplification using an overly large number of prediction-error-filter weights. They also claimed that frequency shifting was similar to that observed by Jackson[9] for FFT spectra. In fact, Jackson's worst cases were sinewaves an even number of cycles long with either 0° or 90° initial phase. In these cases Burg spectra are neither split not shifted but are extremely accurate with very sharp, correctly located peaks.

In 1976, Fougere, Zawalick and Radoski[10] correctly developed the systematics of spontaneous line splitting. They showed that:

1. Splitting is not caused by amplification of noise. In fact splitting is sharpest and most pronounced when the noise level is lowest.
2. Splitting is not caused by using too many prediction-error filter weights.
3. Splitting is worst (as correctly pointed out by Chen and Stegen) with a combination of a signal length equal to an odd number of quarter cycles and an initial phase equal to an odd multiple of 45°. In this worst case, split peaks become broader and less distinct as the noise level is raised and eventually coalesce to a shifted frequency at high enough noise levels. In this worst case also, lengthening the signal does not help. Even for a signal as long as 49 1/4 cycles, splitting is still easy to detect.

In order to understand the attempts that have been made to cure the splitting/shifting problem of Burg Spectra it is necessary to discuss the Burg Technique itself. In his original 1968 paper Burg[2] stated the expression for the prediction error variance as the mean square of forward and backward prediction errors. This variance is a function of the prediction error coefficients. Burg then used the modern Levinson Algorithm to replace all coefficients except the last one (the reflection coefficient) by functions of the reflection coefficient and of previously deter-

ained reflection coefficients. Burg assumed that the previous reflection coefficients are held fixed, and minimized the error variance with respect to the current reflection coefficient. In this manner, all reflection coefficients are determined recursively using a single minimization at each stage. This procedure guarantees that all reflection coefficients are bounded by unity in magnitude and that the resulting filter is a prediction error filter. The corresponding Toeplitz matrix of estimated autocorrelations is nonnegative definite and the basic autocorrelation function theorem is satisfied.

In an attempt to ameliorate the splitting shifting problem, Ulrych and Clayton[11] in 1976 dropped all constraints from the Burg Technique and solved the least squares problem of minimizing the unconstrained sum of squares of the forward and backward prediction errors, a procedure which Burg[2] had specifically rejected as incorrect! In this method the reflection coefficients lose their special significance and the Levinson Recursion is not used at all. Since there are no constraints the reflection coefficients are not bounded by unity magnitude, and thus the prediction error filter can become unstable and the power spectrum can become negative. This method, originally rejected by Burg because of its latent instability, has achieved widespread popularity under many different guises. Some of them will be discussed later.

In 1977, Fougere[12] presented a solution to the splitting problem. Recall from the discussion on the Burg Technique that Burg assumes previously determined reflection coefficients are held fixed in order to determine the present reflection coefficient. It is this constraint which makes the Burg Technique so simple to apply and yet it is an unnecessary constraint. The correct necessary and sufficient condition for a set of numbers to constitute a prediction error filter is that the set is determined from a set of reflection coefficients each of which is less than unity in magnitude and that the Levinson recursion holds at each stage. In his non-linear technique, Fougere sets each reflection coefficient equal to the sine of a real angle times a fixed constant slightly less than one (typically $(1-10^{-6})$ and minimizes the mean forward and backward error power with respect to these unconstrained real angles. The resulting set of reflection coefficients is used to generate the prediction error filter by means of repeated applications of the Levinson recursion. This is not an additional constraint. The Levinson recursion is simply a transformation between two basis sets which span the same vector space. The non-linear procedure is so called because the error power is a function of the sines of the independent variables. The normal equations are solved by standard nonlinear minimization techniques. At any stage in the iteration the reflection coefficients are always less than unity in magnitude and the prediction error filter is always guaranteed to be a stable minimum phase filter with all the roots of its Z transform (using positive powers of Z) lying outside of the unit circle. The power spectrum is always positive, and the autocorrelation function, which can be determined easily, always satisfies the basic autocorrelation function theorem. The maximum number of prediction error filter coefficients m can be shown to be $m = 2/3\, n$ where n is the number of observations. Using this value of m, it should be possible to achieve a perfect fit to any set of input data -- even random noise. Using the non-linear technique this condition has been met as follows: A set of n values of gaussian white noise of variance σ^2 is used as input. The non-linear procedure is run for $m = 2/3\, n$ and iterated until the error variance is many orders of magnitude lower than σ^2; in fact, in many tests the error variance was effectively zero. Thus, a perfect fit had been achieved. This shows that the number of degrees of freedom (df) of the prediction error filter is exactly equal to the df of the data and that therefore the constraint is exactly necessary and sufficient.

The multichannel splitting/shifting problem.

In order to study the multichannel splitting/shifting problem we began by trying a signal which exemplified the problem for single channel signals. This is a single sine wave of frequency 1 Hz sampled every 0.05 seconds, with additive noise:

$$f_1(n) = \sin\left((n-1)\pi/10 + \phi_1\right) + A\varepsilon_1(n)$$

$$f_2(n) = \sin\left((n-1)\pi/10 + \phi_2\right) + A\varepsilon_2(n),$$

$$n = 1, 2, \ldots, N$$

where ϕ_1 and ϕ_2 are the initial phase in each channel and $\varepsilon_1(n)$ and $\varepsilon_2(n)$ are independent realizations of random numbers in the range ± 0.5 and A is the noise amplitude.

The first cases studied are $\phi_1 = 45°$, $\phi_2 = 135°$; A = 0.0005, .001, .002, .004; and N = 16, 17, ...25. Figure 1 shows the location of the principal peaks for A = 0.0005. Note that the spectra peaks are in essentially identical location for each channel even though we have different noise realization in the two channels. Figure 1 also shows the apparent frequency in the cases $\phi_1 = -45°$; $\phi_2 = -135°$. As expected there is a node at N = 21 points, one full cycle, where there is excellent agreement between true and apparent frequencies.

Figure 2 shows three more cases with results not anticipated by analogy to the single channel results. All three cases use an initial phase of 0° in channel 1. In channel 2 we use: 0°, 90° and no signal and all — only random noise, respectively. All three signals show similar behavior, with true nodes at 21 and 31 points, ie 1 and 1 1/2 cycles respectively and approximate nodes at odd quarter cycles. Note that even in the case of no signal at all in channel 2 the channel 2 spectrum has peaks almost identical to those shown in Figure 2 for channel 1. Of course, the power spectral density in channel 2 is down many orders of magnitude below that in channel 1, but the "cross talk" between channels is disturbing.

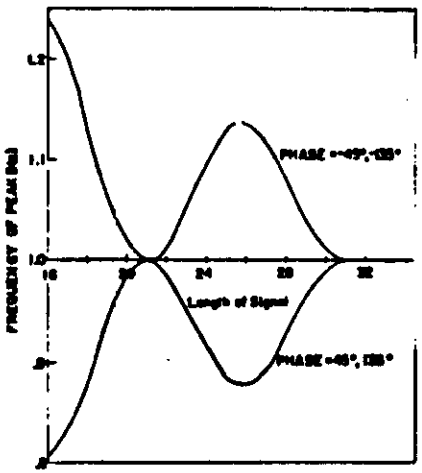


Figure 1. Frequency of peak power spectral density of two-channel signal with 0.0005 noise in each channel versus length of signal for two different combination of initial phase. Peak frequencies were essentially identical in both channels.

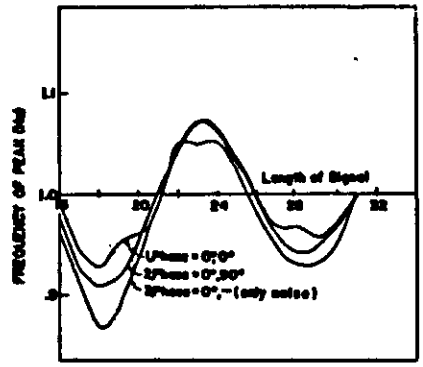


Figure 2. Same as Figure 1 for three more combinations of initial phase. Even for signal number 3 with only noise in channel 2, both channels showed the same peak frequencies.

References

1. Burg, J.P., "Maximum Entropy Spectral Analysis", Ph.D. Thesis. Stanford University, Palo Alto, CA 123 pp. (1975).

2. Burg, J.P., "A New Analysis Technique for Time Series Data" NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics," (1968).

3. Jones R.H., "Multivariate Autoregression Estimation Using Residuals", in Applied Time Series Analysis (Findley, ed.), Academic Press, Inc. 1978.

4. Nuttall, A.H. "Multivariate Linear Predictive Spectral Analysis Employing Weighted Forward and Backward Averaging: A Generalization of Burg's Algorithm" Naval Underwater Systems Center Tech Report #5501 (1976).

5. Strand, O.N., "Multichannel Complex Maximum Entropy (autoregression) Spectral Analysis" IEEE Trans. Inform. Theory, IT-17, 493, (1971).

6. Morf, M., Vieira, A., Lee, D.T.L., and Kailath, T., "Recursive Multichannel Maximum Entropy Spectral Estimation", IEEE Transaction on Geoscience Electronics, GE-16, 85 (1978).

7. Currie, R.G., Private Communication (1980)

8. Chen, W.Y. and Stegen, G.R., "Experiments with Maximum Entropy Power Spectra of Sinusoids", J. Geophys. Res., 79, 3019-3022, (1974).

9. Jackson, P.L., "Truncations and Phase Relationships of Sinusoids", J. Geophys. Res., 72, 1400-1403, (1967).

10. Fougere, P.F. Zawalick, E.J. and Radoski, H.R., "Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", Phys. Earth and Plan. Int., 12, 201-207, (1976).

11. Ulrych, T.J. and Clayton, R.W., "Time Series Modelling and Maximum Entropy", Phys. Earth and Plan. Int., 12, 188-200, (1976).

12. Fougere, P.F., "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", J. Geophys. Res., 82, 1051-1054, (1977).

Reprinted from PROCEEDINGS OF THE 2nd INTERNATIONAL SYMPOSIUM
ON COMPUTER AIDED SEISMIC ANALYSIS AND DISCRIMINATION, August 1981

SPONTANEOUS LINE SPLITTING IN BURG MAXIMUM ENTROPY POWER SPECTRA: A REVIEW

PAUL F. FOUGERE

U.S. AIR FORCE GEOPHYSICS LABORATORY, HANSCOM AFB, MA, 01731, U.S.A.

Abstract

The paper will discuss the problems of spontaneous line splitting of Burg[1,2] Maximum Entropy Power Spectra in the presence of low noise levels and excessive line shifting in the presence of high noise levels. It will critically review various attempts which have been made to ameliorate these problems. Numerical examples using simple data sets will illustrate the basic problems and some of the pitfalls which await the unwary user of some of the unconstrained methods which have become popular more because of their speed than because of their correctness. The non-linear procedure which has cured all cases of spontaneous line splitting so far encountered and which always produces a stable prediction error filter and a meaningful positive spectrum will be described. The extension of the methods, both non-linear and unconstrained, from real to complex data sets will be discussed briefly. The further extension, to multichannel data will be introduced and some new, preliminary results will be presented.

Introduction

In 1974, Chen and Stegen[3] ran some computer experiments, finding Burg maximum entropy power spectra for sinusoids of varying length and initial phase with varying amounts of gaussian white noise. They showed that real problems exist even in the simple case of a single sinewave plus noise. An excessive shift of the frequency at the peak of the spectrum, away from the true input frequency might occur or occasionally the peak would split into one or more comparably sized components. They showed correctly that the splitting/shifting problem is worst for an initial phase equal to an odd multiple of 45° and a signal length equal to an odd number of quarter cycles. However, they incorrectly attributed splitting to noise amplification using an overly large number of prediction-error-filter weights. They also claimed that frequency shifting was similar to that observed by Jackson[4] for FFT spectra. In fact, Jackson's worst cases were sinewaves an even number of cycle long with either 0° or 90° initial phase. In these cases Burg spectra are neither split not shifted but are extremely accurate with very sharp, correctly located peaks.

In 1976, Fougere, Zawalick and Radoski[5] correctly developed the systematics of spontaneous line splitting. They showed that:

1. Splitting is not caused by amplification of noise. In fact splitting is sharpest and most pronounced when the noise level is lowest.
2. Splitting is not caused by using too many prediction-error filter weights.
3. Splitting is worst (as correctly pointed out by Chen and Stegen) with a combination of a signal length equal to an odd number of quarter cycles and an initial phase equal to an odd multiple of 45°. In this worst case, split peaks become broader and less distinct as the noise level is raised and eventually coalesce at a shifted frequency at high enough noise levels. In this worst case also, lengthening the signal does not help. Even for a signal as long as 49 1/4 cycles, splitting is still easy to detect.

In order to understand the attempts that have been made to cure the splitting/shifting problem of Burg Spectra it is necessary to discuss the Burg Technique itself. In his original 1968 paper Burg[2] stated the expression for the prediction error variance as the mean square of forward and backward prediction errors. This variance is a function of the prediction error coefficients. Burg then used the modern Levinson Algorithm to replace all coefficients except the last one (the reflection coefficient) by functions of the reflection coefficient and of previously determined reflection coefficients. Burg assumes that the previous reflection coefficients are held fixed, and minimizes the error variance with respect to the current reflection coefficient. In this manner, all reflection coefficients are determined recursively using a single minimization at each stage. This procedure guarantees that all reflection coefficients are bounded by unity in magnitude and that the resulting filter is a prediction error filter. The corresponding Toeplitz matrix of estimated autocorrelations is nonnegative definite and the basic autocorrelation function theorem is satisfied.

In an attempt to ameliorate the splitting shifting problem, Ulrych and Clayton[6] in 1976 dropped all constraints from the Burg Technique and solved the least squares problem of minimizing the unconstrained sum of squares of the forward and backward prediction errors, a procedure which Burg[5] had specifically rejected as incorrect! In this method the reflection coefficients lose their special significance and the Levinson Recursion is not used at all. Since there are no constraints the reflection coefficients are not bounded by unity magnitude, and thus the prediction error filter can become unstable and the power spectrum can become negative. This method, originally rejected by Burg because of its latent instability, has achieved widespread popularity under many different guises. Some of them will be discussed later.

In 1977, Fougere[7] presented a solution to the splitting problem. Recall from the discussion on the Burg Technique that Burg assumes previously determined reflection coefficients are held fixed in order to determine the present reflection coefficient. It is this constraint which makes the Burg Technique so simple to apply and yet it is an unnecessary constraint. The correct necessary and sufficient condition for a set of numbers to constitute a prediction error filter is that the set is determined from a set of reflection coefficients each of which is less than unity in magnitude and that the Levinson recursion holds at each stage. In his non-linear technique, Fougere sets each reflection coefficient equal to the sine of a real angle times a fixed constant slightly less than one (typically $(1-10^{-7})$ and minimizes the mean forward and backward error power with respect to these unconstrained real angles. The resulting set of reflection coefficients is used to generate the prediction error filter by means of repeated applications of the Levinson recursion. This is not an additional constraint. The Levinson recursion is simply a transformation between two basis sets which span the same vector space. The non-linear procedure is so called because the error power is a function of the sines of the independent variables. The normal equations are solved by standard nonlinear minimization techniques. At any stage in the iteration the reflection coefficients are always less the unity in magnitude and the prediction error filter is always guaranteed to be a stable minimum phase filter with all the roots of its Z transform (using positive powers of Z) lying outside of the unit circle. The power spectrum is always positive, and the autocorrelation function, which can be determined easily, always satisfies the basic autocorrelation function theorem.

The maximum number of prediction error filter coefficients m can be shown to be $m = 2/3\ n$ where n is the number of observations. Using this value of m, it should be possible to achieve a perfect fit to any set of input data -- even random noise. Using the non-linear technique this condition has been met as follows: A set of n values of gaussian white noise of variance ϵ^2 is used as input. The non-linear procedure is run for $m = 2/3\ n$ and iterated until the error variance is many orders of magnitude lower than ϵ^2; in fact, in many tests the error variance was effectively zero. Thus, a perfect fit had been achieved. This shows that the number of degrees of freedom (df) of the prediction error filter is exactly equal to the df of the data and that therefore the constraint is exactly necessary and sufficient.

The paper will present examples of splitting and its cure using the non linear method and examples of instabilities induced by using some of the many different appearing but really identical unconstrained methods.

References

1. Burg, J.P., "Maximum Entropy Spectral Analysis", Ph.D. Thesis, Stanford University, Palo Alto, CA 123 pp. (1975).

2. Burg, J.P., "A New Analysis Technique for Time Series Data" NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics," (1968).

3. Chen, W.Y. and Stegen, G.R., "Experiments with Maximum Entropy Power Spectra of Sinusoids", J. Geophys. Res., 79, 3019-3022, (1974).

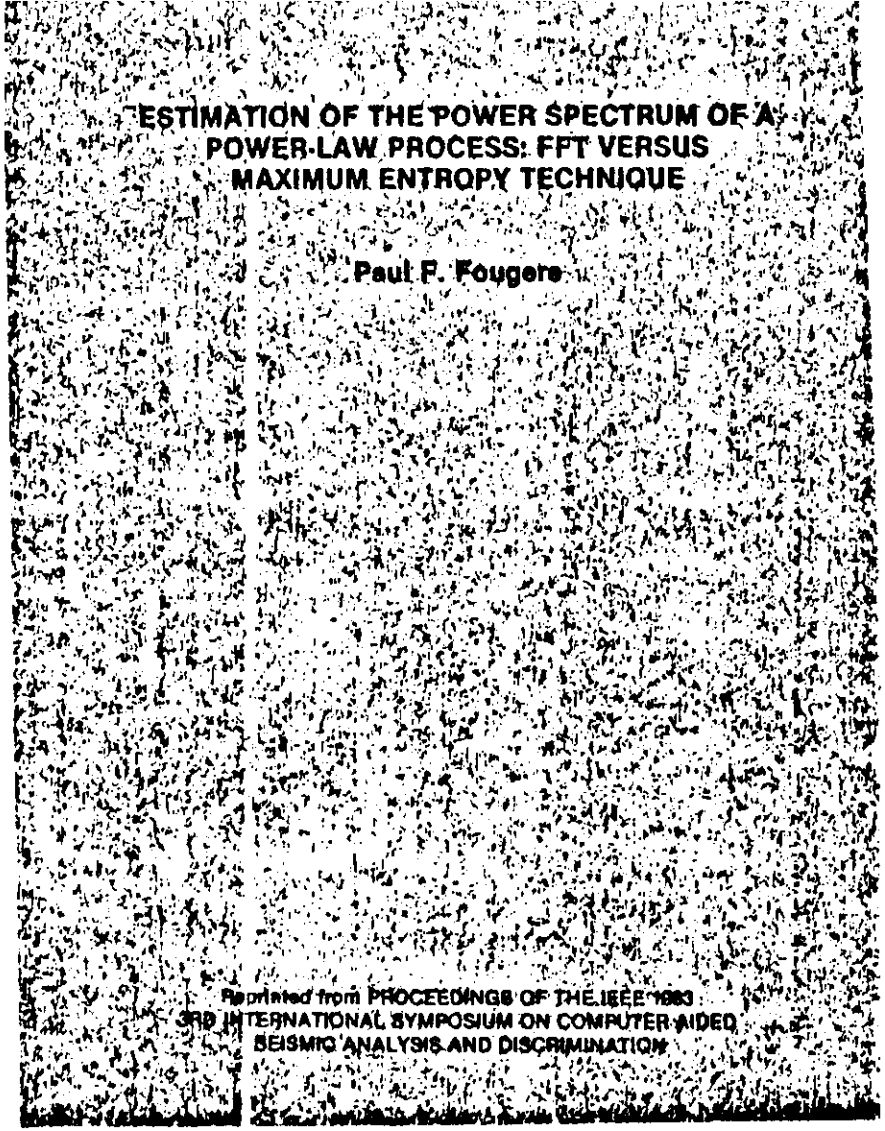
4. Jackson, L.B., "Truncations and Phase Relationships of Sinusoids", J. Geophys. Res., 72, 1400-1401, (1967).

5. Fougere, P.F. Zawalick, E.J. and Radoski, H.R., "Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", Phys. Earth and Plan. Int., 12, 201-207, (1976).

6. Ulrych, T.J. and Clayton, R.W., "Time Series Modelling and Maximum Entropy", Phys. Earth and Plan. Int., 12, 188-200, (1976).

7. Fougere, P.F., "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", J. Geophys. Res., 82, 1051-1054, (1977).

28 34



80

ESTIMATION OF THE POWER SPECTRUM OF A POWER-LAW PROCESS: FFT VERSUS MAXIMUM ENTROPY TECHNIQUE

by Paul F. Fougere

Air Force Geophysics Laboratory
Hanscom Air Force Base, Bedford, MA 01731

Abstract

Power spectra, estimated by the maximum entropy method and by an FFT based periodogram method, are compared using simulated time series. The time series are computer generated by passing Gaussian white noise through low pass filters with precisely defined magnitude response curves such that the output time series have power-law spectra: $P(f)=Af^{-p}$. Ten different values of p between 0.5 and 5.0 are used. Using 4000 independent realizations of these simulated time series, it is shown that maximum entropy results are superior (usually greatly superior) to the periodogram results even when endmatching or windowing or both are used before the power spectra are estimated. Without the use of endmatching or windowing or both, the FFT results are useless at best and very misleading at worst.

Introduction

A power-law process is one whose power spectrum can be written: $P(f) = Af^{-p}$. If such a spectrum is plotted as log P(f) versus log f, then the slope of the resulting straight line is -p, where p is the so-called spectral index. Many geophysical processes, such as scintillation of radio waves passing through the ionosphere, are modelled as power-law processes and the spectral index then yields important information on the scale sizes of the ionospheric irregularities which produce the scintillation[1].

Most published reports dealing with power-law processes estimate the power spectrum by using a periodogram - calculated from an FFT algorithm. The periodogram has the disturbing property that the variance of the estimates increases as the number of original time series observations increases. Statistically the periodogram provides what is called an inconsistent estimate of the power spectrum. It is this inconsistency which has led to a number of very clever "remedies" including windowing of the data; smoothing of the periodogram; and stacking followed by averaging of the estimated power spectrum of many independent realizations of the original noise process.

About fifteen years ago, John Burg[2,3] presented papers at two international meetings in which he introduced his radical new technique based upon the work of Shannon[4] and Jaynes[5] on Information Theory. The two Burg papers as well as many others on modern spectrum estimation are reprinted in Childers[6]. See also Burg[7]. The method, called the Maximum Entropy Method or MEM for short, has revolutionized the subject of power spectral estimation. MEM is still not yet widely accepted and practiced throughout the geophysical community despite the fact that MEM has been shown to produce a smoother spectrum with higher resolution than the FFT based techniques, including the Cooley-Tukey and the slightly older Blackman-Tukey methods. See, for example, Radoski et al[8,9]. For an excellent tutorial review on all of the modern spectral analysis techniques, including MEM and FFT, see Kay and Marple[10].

It is the object of this paper to compare the FFT based and MEM techniques for power-law processes using spectral indices varying between 0.5 and 5.0.

Simulation of the Power Law Process

For any linear, time invariant filter, the output power spectrum is equal to the input power spectrum times the square of the magnitude of the frequency response of the filter. See, for example, Rabiner and Gold[11], page 414. If the input is Gaussian white noise, its power spectrum is a constant, the variance of the noise. Thus the output power spectrum is simply a constant times the magnitude-squared frequency response of the filter. If a filter can be designed to have a frequency response in the form: f^{-p} then its output power spectrum, when excited by white noise with unit variance will be:

$$P(f) = f^{-2p}$$

A program written by McClellan et al[12] and available on tape from IEEE can be used to design a finite input response (FIR) filter with any desired magnitude frequency response. It is only necessary to write a simple subroutine to define the desired frequency response.

38

The Computer Experiments

Using this program, filters were designed whose squared magnitude had a power-law response, in the frequency range 0.01 to 0.5, with indices of 0.5 to 5 in steps of 0.5; 10 filters in all. In order to compare the two methods systematically, each filter was used to produce 100 independent red-noise realizations for each power-law index. Four sets of experiments were run, characterized by differing treatments of the red-noise before the spectral estimates.

EXPERIMENT A: Raw Data

The 1000 experiments of case A are summarized in Figure 1A. Here we see the spectral indices for FFT versus those for MEM. The MEM indices are always reasonable with relatively little scatter for the entire range of index 0.5 to 5.0. The FFT indices always show greater scatter but are reasonable if the true index lies between 0.5 and 2.0. As the true index becomes greater than 2, the FFT results become worse - that is an increase in true index results in a decrease in the index approximated by FFT. Note that for each point plotted, the same red-noise realization was used for both MEM and FFT.

An explanation of the failure of the FFT results is that the FFT technique produces a Fourier analysis of the data sample. An inverse Fourier analysis, back into the time domain, produces a periodic function of time, whose period is equal to the duration of the original data. If the original data set does not look periodic, that is if the first and last points are not equal, then the resulting discontinuity in value of the periodic function produces a distorted power spectrum estimate. If the discontinuity is large, the resulting spectrum distortion can be fatal.

The practitioners of the FFT technique come to the rescue at this point with a remedy called "end matching". They find the straight line which connects the first and last observations and then subtract this straight line from the data. We are thus led to:

EXPERIMENT B: Endmatching

Now the same 10 filters as in experiment A are used but each rednoise realization (time series) is endmatched before spectral estimation. Once again, the same data set is used as input to both MEM and FFT. Figure 1B shows the results with a dramatic improvement in the FFT indices up to a true index of 4.0. The results at 4.5 and 5.0 are again unacceptable. Note also that the MEM results are slightly but definitely worse. There is more scatter, and at an index of 0.5 the MEM results are biased upward. This is further evidence that end-matching really changes the spectrum of the data set whose spectrum we are trying to estimate.

Clearly then, end-matching is required for FFT spectra - without it most of the spectrum estimates are completely unreliable. Except in the low index range, for true indices between 0.5 and 2.0, the FFT results are useless, because an FFT result of 2.0 could have arisen from a true index anywhere between 2.0 and 5.0.

Just as clearly, endmatching should be avoided for MEM calculations, where it does not help but degrades the spectral estimates

EXPERIMENT C: Windowing

The failure of the FFT results at true indices of 4.5 and 5.0 may perhaps be explained by discontinuities in slope of the original data set at the beginning and end. The cure suggested and used by the FFT community is tapering - looking at the data through a window.

The actual window used was recommended in Welch [13] and is:

$$W_j = 1 - (2j/(L+1) - 1)^2 \qquad ; j = 1, 2 \ldots L.$$

For those interested in "window carpentry", no less than 44 distinct windows are discussed in great detail in Harris [14]. The same 1C filters used in experiment A and B were again used in experiment C. Here the above window was applied routinely to all red-noise realizations before the spectra were estimated. The results are given in Figure 1C. The FFT results are now usable for indices between 0.5 and 4.5 but the results at 5.0 are severely biased. Once again, as with the use of endmatching, windowing degrades the MEM spectra. The MEM scatter is even larger here then it was for endmatching.

EXPERIMENT D: Both Endmatching and Windowing

Here the original red noise realizations are first endmatched and then windowed before the spectral estimation. Figure 1D shows the results. Now at last the FFT results are usable for the entire range of spectral indices from 0.5 to 5.0. The MEM results show that endmatching and windowing, taken either singly or together, degrade the MEM spectra. Note well that in all cases, for all experiments, the MEM scatter is smaller than the FFT scatter.

Discussion

The above 4000 runs, collected in four experiments of 100 runs each on ten distinct spectral indices between 0.5 and 5.0, show convincingly that the Maximum Entropy Method, applied directly to the raw data, yields results superior to those produced by the FFT method on several counts.

1. The MEM spectral shape is always smooth and nearly linear. The FFT shape is highly variable and noisy.

2. When straight lines are fit to the spectra the resulting spectral indices always are more highly variable with FFT than with MEM.

3. The average MEM index for 100 realizations is far superior to the average FFT index unless the data are arbitrarily subjected to either endmatching or windowing or both.

4. Only when both endmatching and windowing are simultaneously applied to the data are the results of FFT spectral estimations usable over the entire range of spectral indices from 0.5 to 5.0. Even when both are applied to the data and FFT used, better results could have been obtained by using MEM on the raw data.

Note

A greatly expanded version of the SUMMARY presented here was submitted for publication in the Journal of Geophysical Research in April 1983.

References

1. Crane, R.K., "Spectra of Ionospheric Scintillation", J. Geophys. Res., 81, 2041, (1976).

2. Burg, J.P., Maximum entropy spectral analysis. Paper presented at 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma (1967).

3. Burg, J.P., "A New Analysis Technique for Time Series Data", NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics (1967).

4. Shannon, C.E. and W. Weaver, "The Mathematical Theory of Communication", University of Illinois Press Urbana, 1949.

5. Jaynes, E.T., "Information Theory and Statistical Mechanics I", Phys. Rev., 106, 1957, 620-630, (1957).

6. Childers, D.G., "Modern Spectrum Analysis, New York IEEE Press, 1978.

7. Burg, J.P., "Maximum Entropy Spectral Analysis", Ph.D. Thesis. Stanford University, Palo Alto, CA, 123 pp (1975).

8. Radoski, H.R., P.F. Fougere and E.J. Zawalick, "A comparison of power spectral estimates and applications of the maximum entropy method", J. Geophys. Res., 80, 619-625, 1975.

9. Radoski, H.R., E.J. Zawalick and P.F. Fougere, "The superiority of maximum entropy power spectrum techniques applied to geomagnetic micropulsations", Phys. Earth Planetary Interiors, 12, 208-216, 1976.

10. Kay, S.M. and S.L. Marple Jr., "Spectrum Analysis - A Modern Perspective", Proc. IEEE, 69, 1380-1418, (1981).

11. Rabiner, L.R. and B. Gold., "Theory and Application of Digital Signal Processing, Prentice Hall Inc. (1975).

12. McClellan, J.H., T.W. Parks, and L.R. Rabiner, "FIR Linear Phase Filter Design Program" in Programs for Digital Signal Processing, IEEE Press, (1979).

13. Welch, P.D., "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Trans Audio Electroacoustics AU-15, 70-73, (1967).

14. Harris, F.J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", Proc. IEEE, 66, 51, (1978).

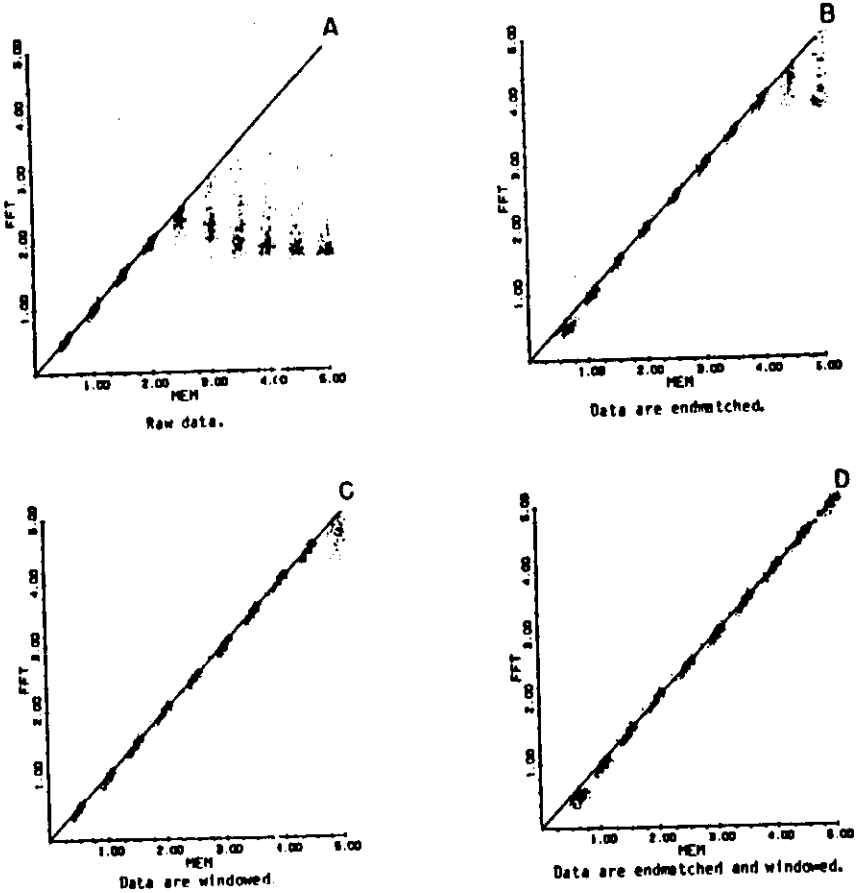


Figure 1 All Panels: Each point is the observed FFT slope versus the MEM slope. There are 100 independent realisations of the power-law process for each of the ten slopes -0.5, -1.0...-5.0. In every case the same time series was used as input to MEM and to FFT.

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 89, NO. A5, PAGES 3023-3027, MAY 1, 1984

Eighty-Eight Year Periodicity in Solar-Terrestrial Phenomena Confirmed

J. FEYNMAN

Department of Physics, Boston College

P. F. FOUGERE

Air Force Geophysics Laboratory, Hanscom Air Force Base

The existence of a 60-100 year periodic variation in solar and/or solar-terrestrial phenomena has been a matter of dispute for many years. A wide variety of data sets previously have been analyzed, and the results of the analyses have been interpreted as showing evidence either for or against such a variation. However, all data sets that are proxy for solar wind in the ecliptic at 1 AU show variations consistent with a period of about 88 years. Here we report that a maximum entropy spectral analysis of the number of aurora reported per decade in Europe and the Orient from 450 A.D. to 1450 A.D. shows a strong stable line at a 88.4 ± 0.7 years. Since the data set contains 11 cycles, this analysis establishes the reality of the "long cycle" for 1000 years. The mean amplitude and phase are then estimated from a superposed epoch analysis. The mean amplitude was 2.2 auroral reports per decade and the last minimum phase in these data occurred between 1403 A.D. and 1413 A.D.

INTRODUCTION

Firmly establishing whether or not long-period variations exist in solar output and in solar terrestrial relations is of fundamental importance in understanding the sun and solar wind and in our ability to predict variations in interplanetary and magnetospheric environments. However, there has been a great deal of interest and controversy concerning the reality of systematic changes in solar output in the frequency range of one or two cycles per century. This type of possible variation has been called, among other things, the Gleissberg variation, the "long cycle," the "87-year cycle," and the secular variation. The frequency range is particularly intriguing because on the one hand there are hints of such changes in the historical record and on the other hand the period is so long that it has been difficult to accumulate data adequately to test for such a variation. Furthermore, tests made on different data sets have resulted in conflicting conclusions.

STATUS OF THE PROBLEM

Recently, *Feynman* [1983] reviewed six data sets which had been used previously as proxy data for solar outputs to test for periodic changes in the 60–100 year period range. These data sets were chosen because in each case the relationship between the observed quantities and the solar wind was understood at least in principle. For this reason, data pertaining to solar-weather relationships were not included. The six data sets were the sunspot numbers since 1720 as an expression of the solar cycle, the *aa* index of geomagnetic activity, a post-Maunder minimum auroral data set, a pre-Maunder minimum auroral data set, and two sets of ^{14}C data. The different data sets are, of course, proxy data for different aspects of solar variability. The sunspot number is related to the photospheric magnetic field. The ^{14}C is produced indirectly by cosmic rays that have propagated through broad reaches of the heliosphere and so refers to properties of the solar wi throughout the heliosphere. Geomagnetic variations and r roral counts refer to solar output in a very restricted region space, i.e., the ecliptic plane at 1 AU. Since the data sets r proxy for at least three different aspects of solar activity a solar wind, a periodicity in any one of these quantities do not necessarily imply a periodicity in the others, and disagre ments exist among the results only when two analyses proxy data for the same aspect of the sun and/or solar wi yield conflicting results.

The post-1720 11-year sunspot number cycles had relative small amplitudes at the beginning of the 19th and of the 20 centuries, and these small amplitude cycles have been wide interpreted as being due to minimums in an 87-year period variation. *Sonett* [1982] has recently run a maximum entro power spectrum on these data and, as expected, found a line an 87-year period. However, there are only 2½ cycles of da at that period, and in our opinion that is not a long enoug time series to establish a variation as periodic. The post-17 frequency of auroral sightings in Sweden [*Rubenson*, 188 and the geomagnetic variations as measured by *aa* [*Mayau* 1973] are also consistent with an 87-year cycle. Although ne ther of these latter two sets of solar terrestrial data shown yearly average value that is proportional to the yearly averag sunspot number [*Feynman and Crooker*, 1978; *Legrand an Simon*, 1981; *Feynman*, 1983], they both have minimum valu at about the same time as the sunspot cycle amplitude mini mums, i.e., Swedish auroral reports minimized in the earl 19th and 20th centuries [*Silverman and Feynman*, 1980] and a minimized in the early 20th century.

The situation is different with the two ^{14}C data sets. *Stuiv* [1980] carried out a power spectral analysis of the post-70 A.D. rates of production of ^{14}C and did not find any increas in power in the 60- to 100-year frequency range. However, i an independent study, *Lin et al.* [1975] did a covariance func tion analysis of 8000 years of ^{14}C anomaly data and found broad increase in power in the region of 80 years, which the interpreted as evidence for the Gleissberg cycle. This conflic of results is not yet resolved.

The final data set discussed by *Feynman* [1983] is the me

Paper number 4A0311.
0148-0227/84/004A-0311$02.00

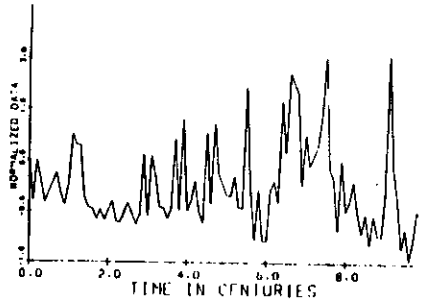


Fig. 1. The normalized and detrended number of auroral observations per decade. Zero corresponds to 450 A.D. The data before normalization and detrending is from *Siscoe* [1980].

lieval auroral set which was reviewed by *Siscoe* [1980] and which will be further analyzed here. The data set is derived rom reports of auroras seen in Europe and in the Orient from 150 A.D. to 1450 A.D. Siscoe, following *Keimatsu* [1976], investigated the accuracy of the data by comparing the number of auroras per century reported from China and Europe separately. Not only were the envelopes of the two frequency distributions remarkably close, but the actual number of reports from the two areas were surprisingly similar. In addition, features corresponding to the 7th century medieval auroral minimum and the 12th century medieval auroral maximum are seen in the ^{14}C data. Siscoe also presented the number of auroras seen per decade in the combined European-Oriental data set. In Figure 1, we show the data to be used here, i.e., the data taken from *Siscoe* [1980] and then normalized and detrended. As Siscoe pointed out and as had been often suggested before (see review by *Siscoe* [1980]), the data set appeared to show a periodic variation with a mean period of about 87 years. In order to test the validity of this observation in more objective way, *Feynman* [1983] carried out a modified superposed epoch analysis for the 87-year signal and found what appeared to be a statistically significant result. However, this method of analysis is unsatisfactory because the period must be chosen ahead of time and so cannot be accurately determined and because the evaluation of the statistical significance of the results is subjective.

In this paper we carry the objectivity of the analysis of the medieval auroras further by using a maximum entropy method (MEM) spectral analysis. This method has been shown to be superior to other methods in certain cases in that the period of the variation can be determined very accurately *Radoski et al.*, 1975]. Once the period is known from MEM, can be used in a superposed epoch analysis to determine the amplitude and phase, as shown below.

Determination of the Period

The maximum entropy method (MEM) of power spectral analysis consists of two independent steps, invented by John Burg in 1967 and 1968. See Burg's Ph.D. thesis [*Burg*, 1975], as well as papers by *Radoski et al.* [1975], *Smylie et al.* [1973], and *Ulrych and Bishop* [1975] for detailed reviews and discussions of the method. See also the excellent review of modern spectrum estimation by *Kay and Marple* [1981], as well as a set of important reprints, including the original Burg papers, edited by *Childers* [1978].

The first of these two steps is the so-called Burg technique which determines a prediction error filter with m weights which, when run in both time directions over the M data points, minimizes the mean square prediction error. Since the data are being used to predict later (or earlier) values of the data, the method is also called autoregressive.

The second step is called MEM and uses the m prediction error coefficients in an expression derived by Burg to estimate the power spectral density by maximizing entropy subject to constraints. This two step method will be referred to as Burg-MEM. This technique has been shown to produce smooth, accurate spectra with enhanced resolution relative to the older Blackman-Tukey [*Blackman and Tukey*, 1959] or Cooley-Tukey [*Cooley and Tukey*, 1965] methods. *Fougere et al.* [1976] have shown that in certain cases, however, Burg-MEM has some undesirable properties, namely, line splitting in the low-noise cases and line shifting in moderate- and high-noise cases more characteristic of real data sets. These problems were solved by *Fougere* [1977] by using a nonlinear generalization of the Burg technique. This new two-step procedure, which we will call nonlinear MEM, also has been applied to our 100-point data set. The results, although in general agreement with Burg-MEM results, should be considered to be quantitatively superior.

In order to test the stability of the spectral lines, both Burg and nonlinear MEM spectrums were calculated using a variety of weights. The result of the nonlinear analysis for 35 filter weights is shown in Figure 2*a* and the Burg result for 50 filter weights is shown in Figure 2*b*. In both cases the spectrums show a dominant narrow line at a frequency corresponding to

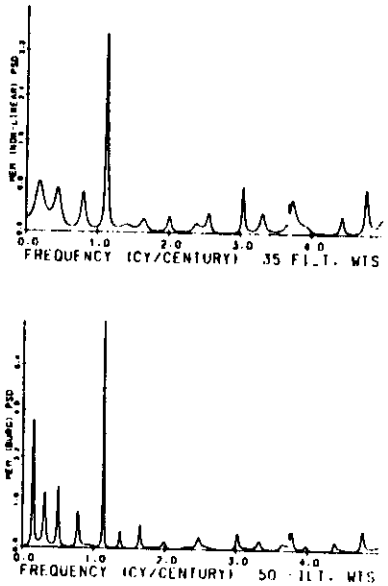


Fig. 2. Typical maximum entropy spectrums of the data shown in Figure 1. The top panel shows a nonlinear spectrum calculated using 35 filter weights, and the lower spectrum uses the Burg method with 50 filter weights. The lines at 1.13 cycles/century (~88 years) are stable and dominant.

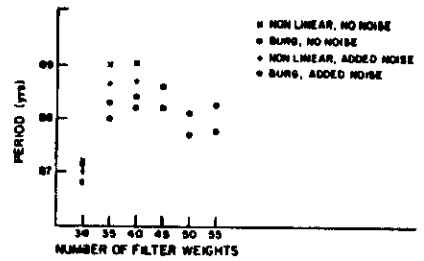


Fig. 3. Estimates of the period of the long cycle from 16 spectrums of the data in Figure 1. Our best estimate of the period is 88.4 ±0.7 years.

one cycle per 88 years. The effect of uncertainty in the data base was tested by adding noise of amplitude one to the data. This corresponds to an uncertainty of about three major mid-latitude aurora per decade, since the range of values in the original data was from 0 to 15 auroral observations. The power spectrums were almost indistinguishable from those run on the data without noise.

Although the spectrums are very stable, there are slight differences in the periods of the peak. In Figure 3 we have plotted the peak periods for 19 spectrums. Except for the four spectrums with 30 filter weights, all of the periods are longer than 87.5 years. Since the frequency determination is probably less accurate for 30 weights than for the higher weights, these four spectrums have not been included in the determination of the period. All of the 15 other spectrums show peaks, the periods of which lie in the range 88.4 ± 0.7 years, and that value is adopted as our best estimate. There can be no doubt that the line found in our spectrums is real, and since there are 1000 years of data covering 11 cycles, the reality of the periodicity during these 1,000 years is firmly established.

Amplitude and Phase

As stated above, MEM power spectral analysis cannot determine the amplitude and phase of a variation but, given the period, these quantities can be found from a superposed epoch analysis of the data. The zero times of the epochs were chosen to make the average interval length 88 years. Since the data consists of the number of auroras per decade, it is not possible to use intervals of exactly 88 years each. Instead, each interval consists of 9 decades of data, but the zero time is adjusted so that the last decades of two intervals (750 A.D. and 1190 A.D.) are also used as the first decades of the next intervals. The results are shown in Figure 4, where the closed circles represent the individual data points, the open circles the bin averages, and the open triangles the fit given by

$$H(X) = 4.9 + 2.2 \cos\left[\frac{2\pi}{8.8}(X - 1) - 22.6^\circ\right]$$

where X is the bin number. This function was determined by a least squares fit to the bin averages and gives a good estimate to the amplitude and phase of the long cycle variation. The amplitude is given in recorded auroral sightings per decade.

Because of the importance of these results, it is worthwhile to review the superposed epoch method briefly to evaluate their validity. Any function can be decomposed into the sum of periodic terms,

$$f(t) = \Sigma_i A_i \cos(\omega_i t + \alpha_i)$$

If we assume that the period of one of the terms is known, then we can write

$$f(t) = A \cos(\omega t + \alpha) + \Sigma_i B_i \cos(\omega_i t + \beta_i)$$

where ω is known but A, α, and the B_i, ω_i, and β_i are unknown. The purpose of the superposed epoch analysis is to find A and α. We construct a time series by sampling $f(t)$ at intervals Δt. The gth term of the series is

$$f(g\,\Delta t) = A \cos(\omega g\,\Delta t + \alpha) + \Sigma_i B_i \cos(\omega_i g\,\Delta t + \beta_i) \qquad (1)$$

However, since we know ω, we can choose Δt so that $\omega L\,\Delta t = 2\pi$, where L is an integer. Using this choice of Δt we perform a superposed epoch analysis of the time series by binning the data in L bins. That is, we group the data samples so that in the first bin we have

$$f(0), f[(L)\,\Delta t], \cdots, f[(sL)\Delta t], \cdots, f[(m-1)L\,\Delta t]$$

where the data set is mL periods long. In the $(n+1)$th bin we have

$$f[(n)\,\Delta t], f[(n+L)\,\Delta t], \cdots, f[(n+sL)\,\Delta t], \cdots, f[(n+(m-1)]L\,\Delta t\}$$

where s is an integer.

The average value of the members of the $(n+1)$th bin is

$$G_{n+1} = \frac{1}{m}\sum_{p=0}^{m-1} f[(n+pL)\,\Delta t] \qquad (2)$$

or, more explicitly,

$$G_{n+1} = \frac{1}{m}\sum_{p=0}^{m-1} A \cos[\omega(n+pL)\,\Delta t + \alpha] + \frac{1}{m}\sum_{p=0}^{m-1}\Sigma_i B_i \cos[\omega_i(n+pL)\,\Delta t + \beta_i] \qquad (3)$$

but we have chosen Δt so that

$$\cos[\omega(n)\,\Delta t + \alpha] = \cos[\omega(n+1)\,\Delta t + \alpha] = \cdots = \cos[\omega n + sL)\,\Delta t + \alpha] = \cdots$$

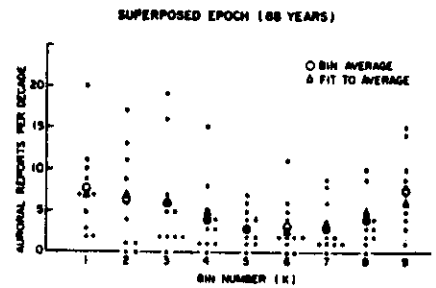


Fig. 4. Superposed epoch analysis of data from *Siscoe* [1980]. Solid dots show data points, open circles give bin averages, and open triangles are calculated from the best fit estimates. If the same number of aurora was reported in the same bin in more than one cycle, the points are plotted next to one another. For example, the figure shows that two auroras were reported in the third decade during four cycles. The average length of the period used in the superposition was 88 years.

and the $m + 1$ terms in the first sum are all equal, and the sum is given by

$$mA \cos [\omega(n)\,\Delta t + \alpha]$$

However, for a given ω_i the terms in the second summation in (3) will not have a predetermined phase relationship to one another. If $\omega_i \propto \omega$ and not commensurate, this quasirandom sampling will cause the term to vanish (provided that B_i is not too large). Even if $\omega_i \gg \omega$ and/or $B_i \gg A$, the process of summation will decrease the relative importance of the second summation. The G_{n+1} will then be of the form

$$G_{n+1} = A \cos [\omega(n)\,\Delta t + \alpha] + E_{n+1}$$

where E_{n+1} is the contribution to the bin average of the summations over p and i in the second term of expression (3).

The size of the error in determining A and α due to the n dependence of E_n can be estimated by fitting the bin average with a cosine function (thus determining the best values of A and α) and comparing the fitted values H_{n+1} with the actual values of G_{n+1} from the data.

In Figure 3 the G_{n+1} are given by open circles and H_{n+1} by triangles. It is clear that the contribution of the n dependence of E_n is small. The errors in A and α are therefore small, although we do not here attempt to put a formal estimate of limits on them.

We conclude that the mean amplitude of the long cycle was 2.2 recorded auroral sightings per decade and the phase was such that the last minimum in these data occurred in the decade between 1403 A.D. and 1413 A.D. A long extrapolation of the results to modern times would produce a minimum circa 1850, whereas the observed minimum in auroral frequency occurred circa 1815. Such a long extrapolation may be unwarranted because the phase is a rather poorly determined function of quite noisy data. The entire phase problem requires further study.

Discussion and Conclusion

We have shown that the number of recorded auroral reports/decade in midlatitude Europe and Asia was periodic with a period of 88.4 $\pm$0.7 years from 450 A.D. to 1450 A.D. Although it appears to be difficult, if not impossible, to avoid concluding that solar output must also have displayed periodic behavior, it is not possible to identify the parameters that changed. The problems that occur in such an identification have been discussed elsewhere [*Feynman*, 1982] and will not be treated at length here. Suffice it to say that almost no studies have been done relating midlatitude auroral observations to solar wind parameters. We have only the general observation that high levels of geomagnetic activity are related to low-latitude auroral appearances. We also know, of course, that geomagnetic activity is associated with solar wind velocity and with southward interplanetary magnetic field, but the empirical relationship is still a matter of dispute [*Holzer and Slavin*, 1983; *Kamide*, 1983].

There is evidence that the changes in solar outputs that take place in connection with the long cycle are quite different from those taking place because of the 11-year cycle. This was inferred from a study of auroral frequency from 1890 to 1935 at high geomagnetic latitude in Sweden [*Silverman and Feynman*, 1980]. The data from Karesuando (geomagnetic latitude 65.2°) shows both a well-marked variation with the long cycle and an equally clear 11-year variation. But whereas the number of auroras reported reaches relative minimums at each 11-year cycle minimum, it reaches a relative maximum at the minimum of the long cycle. Since the frequency, position, and intensity of auroras is determined by solar outputs such as interplanetary wind velocity and magnetic field intensity, it follows that the long cycle variation cannot be dominated by changes in the same parameters as those determining the 11-year cycle.

Although we cannot definitively identify which solar outputs change during the long-cycle variation, we have shown that the long cycle in solar terrestrial relations is real and periodic, that it is present in 1000 years of auroral data, and that the period is 88.4 $\pm$0.7 years.

Acknowledgments. We thank Xiao-Yue Gu for her help in fitting the curve in Figure 4. One of us (J.F.) was supported by National Science Foundation grant ATM 811753.

The Editor thanks two referees for their assistance in evaluating this paper.

References

Blackman, R. B., and J. W. Tukey, *The Measurement of Power Spectra from the Point of View of Communication Engineering*, Dover, New York, 1959.

Burg, J. P., Maximum entropy spectral analysis, Ph.D. thesis 123 pp., Stanford Univ., Palo Alto, Calif., 1975.

Childers, D. G. (Ed.), *Modern Spectrum Analysis*, pp. 34–48, IEEE Press, New York, 1978.

Cooley, J. W., and J. W. Tukey, An algorithm for the machine computation of complex Fourier series, *Math. Comput.*, *19*, 297–301, 1965.

Feynman, J., Geomagnetic and solar wind cycles, 1900–1975, *J. Geophys. Res.*, *87*, 6153, 1982.

Feynman, J., Solar wind variations in the 60–100 year period range: A review, *NASA Conf. Publ. 2280*, 333, 1983.

Feynman, J., and N. U. Crooker, The solar wind at the turn of the century, *Nature*, *275*, 626, 1978.

Fougere, P. F., A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis, *J. Geophys Res.*, *82*, 1051, 1977.

Fougere, P. F., E. J. Zawalick, and H. R. Radoski, Spontaneous line splitting in maximum entropy power spectrum analysis, *Phys. Earth Planet. Inter.*, *12*, 201, 1976.

Holzer, R. E., and J. A. Slavin, Reply, *Geophys. Res. Lett.*, *88*, 4955, 1983.

Kamide, Y., Comment on "An evaluation of three predictors of geomagnetic activity" by R. E. Holzer and J. A. Slavin, *Geophys. Res. Lett.*, *88*, 4953, 1983.

Kay, S. M., and S. L. Marple, Jr., Spectrum analysis: A modern perspective, *Proc. IEEE*, *69*, 1380–1418, 1981.

Keimatsu, M. A., A chronology of aurorae and sunspots observed in China, Korea, and Japan, *Ann. Sci.*, *13*, 1, 1976.

Legrand, J. P., and P. A. Simon, Ten cycles of solar and geomagnetic activity, *Sol. Phys.*, *70*, 173, 1981.

Lin, Y. C., C. Y. Fan, P. E. Damon, and E. I. Wallick, Long term modulation of cosmic ray intensity and solar activity cycle, *Conf. Pap. Int. Cosmic Ray Conf.* 995–999, 1975.

Mayaud, P. N., A hundred year series of geomagnetic data 1868–1967, Indices aa, Storm sudden commencements, *IAGA Bull. 33*, 1973.

Radoski, H. R., P. F. Fougere, and E. J. Zawalick, A comparison of power spectral estimates and applications of the maximum entropy method, *J. Geophys. Res.*, *80*, 619, 1975.

Rubenson, R., Catalogue des Aurores Boreales observees en Suede, *K. Sven. Vetensk. Akad. Handl.*, *18*, 1, 216, 1882.

Silverman, S. M., and J. Feynman, The changing aurora of the past three centuries, p. 407, in *Exploration of the Polar Upper Atmosphere*, edited by C. S. Deehr and J. A. Holtet, D. Reidel, Hingham, Mass., 1980.

Siscoe, G. L., Evidence in the auroral record for secular solar variability, *Rev. Geophys. Space Phys.*, *18*, 647, 1980.

Smylie, D. E., G. K. C. Clarke, and T. J. Ulrych, Analysis of irregularities in the earth's rotation, *Methods Comput. Phys.*, *13*, 391, 1973.

Sonett, C. P., Sunspot time series: Spectrum found from square law modulation of the Hale cycle, *Geophys. Res. Lett.*, *9*, 1313, 1982.

Stuiver, M., Solar variability and climate change during the current millennium, *Nature*, *286*, 868, 1980.

Ulrych, T. J., and T. N. Bishop, Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys. Space. Phys.*, *13*, 183–200, 1975.

J. Feynman, Department of Physics, Boston College, Chestnut Hill, MA 02167.

P. F. Fougere, Air Force Geophysics Laboratory, Hanscom Air Force Base, MA 01731.

(Received December 12, 1983;
revised February 2, 1984;
accepted February 22, 1984.)

On the Accuracy of Spectrum Analysis of Red Noise Processes Using Maximum Entropy and Periodogram Methods: Simulation Studies and Application to Geophysical Data

PAUL F. FOUGERE

Journal of Geophysical Research, vol. 90, no. A5, May 1, 1985

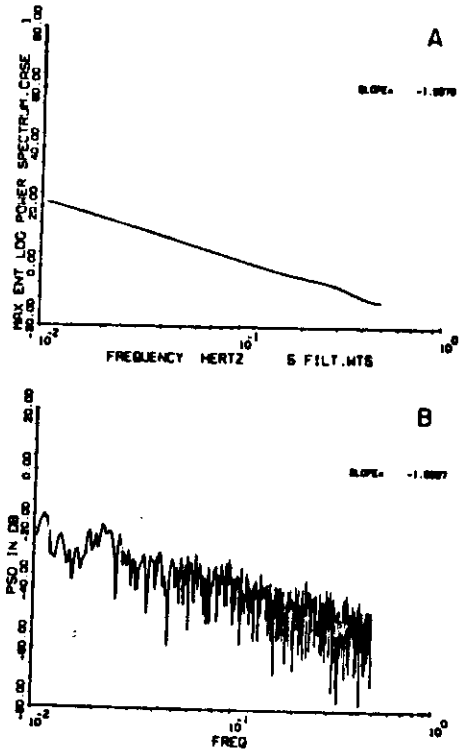


Fig. 2. (*a*) Maximum entropy spectrum of the signal from Figure 1*d*. The ordinate is $10 \log_{10}$ (PSD). (*b*) Periodogram of the same signal.

periodogram. But note that the MEM spectrum is a true power spectral density, a continuous function of frequency which may be evaluated over an arbitrary grid of any density; the periodogram is based upon the finite Fourier transform, and hence it is essentially a sampled or discrete estimate of the power spectrum.

It is the object of Part 2 of this paper to compare the unaveraged periodogram and Burg-MEM techniques for power law processes using spectral indices varying between 0.5 and 5.0.

2. Simulation of the Power Law Process

For any linear, time invariant filter, the output power spectrum is equal to the input power spectrum times the square of the magnitude of the frequency response of the filter (see, for example, *Rabiner and Gold* [1975, p. 414]). If the input is Gaussian white noise, its power spectrum is a constant, the variance of the noise; the output power spectrum is simply a constant times the magnitude-squared frequency response of the filter. If a filter can be designed to have a frequency response in the form $f^{-\beta}$, then its output power spectrum when excited by white noise with unit variance will be

$$P(f) = f^{-2\beta}$$

A program written by *McClellan et al.* [1979] and available on tape from the Institute of Electrical and Electronics Engineers (see also *Digital Signal Processing Committee* [1979]) can be used to design a finite input response (FIR) filter with any desired magnitude frequency response. It is only necessary to write a simple subroutine to define the desired frequency response. Since the frequency response of a power law filter with negative slope is infinite at zero frequency, this one point must be excluded from the calculations. In practice, a band is chosen such that the frequency response follows a power law from f_0 to f_1, where $f_0 > 0$ and $f_1 \leq 0.5$, where f is a frequency normalized by the sampling frequency and runs from 0 to 0.5. The sampled output from the FIR filter is a discrete time moving average (MA) process. An infinite impulse response (IIR) filter (all pole) which produces a power law spectrum over a limited frequency range can also be designed. Such a filter would represent an autoregressive process. The coefficients of the AR model for the filter are estimated as an intermediate step in the Burg-MEM algorithm.

Since the random number generator used to create the Gaussian white noise input can produce a virtually unlimited supply of independent random numbers, many different realizations of the colored noise (power law) process can be produced easily.

3. The Computer Experiments

Using the *McClellan et al.* [1979] program, filters were designed whose squared magnitude had a power law response in the frequency range 0.01 to 0.5, with indices of 0.5 to 5 in steps of 0.5 (10 filters in all). The impulse response of the filter with index 2.0 is shown in Figure 1*a*. The impulse response is symmetric about the $t = 0$ point, and thus the frequency response is zero phase; that is, it is purely real. The frequency response is the Fourier transform of the impulse response, and it may be approximated as closely as desired by using a FFT of the impulse response augmented by a sufficiently large number of zeroes. The process of augmenting a discrete function with a large number of zeroes before performing a finite Fourier transform is called zero padding and results in a closer spacing of the transformed values. Effectively, zero padding in the input produces interpolation in the output. Since the frequency response of a digital filter is the discrete Fourier transform of the (essentially discrete) impulse response, with zero padding the discrete Fourier transform can be approximated as closely as desired by the finite Fourier transform. Figure 1*b* shows the squared frequency response of the filter with index 2, approximated using a 2048-point FFT. The actual slope, obtained by fitting a straight line to the computed points using least squares, is -2.0000.

Using the filter described in Figures 1*a* and 1*b*, the sample of Gaussian white noise, shown in Figure 1*c*, produced the realization of a power law process with index 2 (the FIR filter output) shown in Figure 1*d*. The maximum entropy spectrum using five prediction error filter weights is given in Figure 2*a*, and the periodogram (unaveraged, point by point square of magnitude of FFT) is shown in Figure 2*b*. The MEM spectrum is much smoother than the FFT spectrum: there is much less variance from one frequency estimate to the next, and the shape itself is nearly linear, reflecting the linearity of the true spectrum. For each spectrum the slope is found by fitting a straight line to the power spectral density (PSD) estimates using least squares. In this procedure, all of the computed points shown in Figure 2*a* or Figure 2*b* are used to fit a straight line in the form

$$\log y = -m \log x + A$$

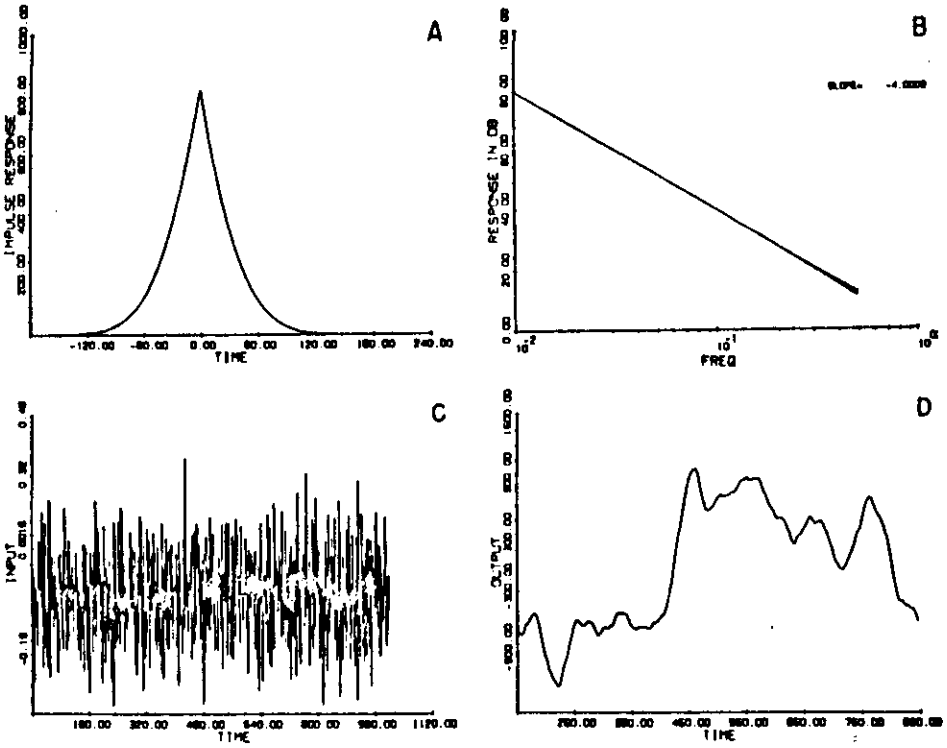


Fig. 3. (a) Impulse response of a filter designed to have a squared power law frequency response with a slope of −4. A total of 301 weights are used. (b) Frequency response of the filter. The slope is −4.0009. (c) The 1000-point sample of Gaussian white noise used as input to the filter. (d) The 700-point output. At each end, 150 points are lost because the symmetrical filter is 301 weights long.

where m is the desired spectral index, using the method of least squares. The MEM slope is −1.9578, and the periodogram slope is −1.9397. Thus the smoothed behavior of the periodogram in this case is acceptable.

When the spectral index is increased to 4, the impulse response is that shown in Figure 3*a*, the squared frequency response is that shown in Figure 3*b*, the white noise input is that given in Figure 3*c*, and the red noise realization is that given in Figure 3*d*. The MEM spectrum, this time using 10 weights, is given in Figure 4*a*, and the periodogram in Figure 4*b*.

In this case the MEM spectrum is still quite acceptable, but the periodogram is not! There is essentially no relationship between the periodogram and the true spectrum. The true slope is −4, and the periodogram slope is only −1.9580.

An explanation of the periodogram difficulty derives from the fact that a finite data set is equivalent to an infinite data set multiplied by a rectangular window which is unity inside the measurement interval and zero outside. Multiplication of the actual time series by a window function implies that the overall transform is a convolution of the desired transform with the transform of the window function. The Fourier transform of the rectangular window has the form $\sin(\pi f l)/\pi f$; this has a relatively narrow central lobe and very high side lobes. If the window were infinitely wide, the side lobes would disappear, and the central lobe would become a Dirac delta function; convolution with such a function would simply reproduce the correct spectrum. But for a finite-sized window the convolution produces some deleterious effects. The first is that the true spectrum is smeared out or defocused; spectral resolution is limited to the width of the main lobe $(NAT)^{-1}$ Hz. The second, and more important for our discussion, is that the high side lobes increase the apparent power at points away from the central lobe. In effect, the decrease of power spectral density with increasing frequency, instead of characterizing the physical process, depends upon the window alone. Thus spectral indices greater than about 2 cannot be obtained using a rectangular window. Many other windows have been designed to offset this high side lobe behavior. The rectangular window gives the narrowest main lobe; other windows designed to reduce side lobes do so at the expense of increasing main lobe width. Nevertheless, some such window is essential: spectral estimation of red noise processes using periodograms requires the use of some nonrectangular window.

In order to compare the two methods systematically, each filter was used to produce 100 independent red noise realizations for each power law index. Four sets of experiments were run, characterized by differing treatments of the red noise before the spectral estimates. In all cases the order of the prediction error filter in MEM was set to 6.

Experiment A: Raw Data

The 1000 experiments of case A are summarized in Figure 5*a*. Here we see the spectral indices for periodograms versus those for MEM. The MEM indices are always reasonable with relatively little scatter for the entire range of index 0.5 to 5.0. The periodogram indices always show greater scatter but are reasonable if the true index lies between 0.5 and 2.0, because of the side lobe behavior discussed above. As the true index becomes greater than 2, the periodogram results become worse; that is, an increase in the true index results in a decrease in the index approximated by the periodogram. Note that for each point plotted, the same red noise realization was used for both MEM and the periodogram.

Another explanation of the difficulty with the periodogram results is that the periodogram technique produces a Fourier analysis of the data sample. An inverse Fourier analysis, back into the time domain, produces a periodic function of time, whose period is equal to the duration of the original data. If the original data set does not look periodic, that is, if the first and last points are not equal, then the resulting discontinuity in value of the periodic function produces a distorted power spectrum estimate. If the discontinuity is large, the resulting spectrum distortion can be fatal. For an illuminating discussion of this problem, which is called "spectral leakage," see the paper by *Harris* [1978] on the use of windows.

Note that this phenomenon does not have any effect on the MEM spectrum, which is specifically designed to operate on the given data and only the given data. No periodic extension of the data is required for MEM as it is for the periodogram technique.

It is in the treatment of missing data that the sharpest and most obvious difference between the periodogram technique and MEM occurs. *Jaynes'* [1982] maximum entropy principle says that in inductive reasoning our general result should make use of all available information and be maximally noncommittal about missing information by maximizing the entropy of an underlying probability distribution. The maximum entropy method, developed by Burg in 1967, begins by assuming that the first few lags of an autocorrelation function are given exactly. Writing the power spectral density as the Fourier transform of the entire infinite autocorrelation function (ACF) and using an expression for the entropy of a Gaussian process in terms of its PSD, Burg solved the constrained maximization problem: Find the PSD whose entropy is maximum and whose inverse Fourier transform yields the given ACF. Maximization is with respect to the missing (infinite in number) ACF values. The result of all this is an extrapolation formula for the ACF. The MEM power spectrum is the exact Fourier transform of the infinitely extended ACF.

The discontinuity in value of the periodogram can be removed by "end-matching," in which a straight line is fit to the first and last data points and then this straight line is subtracted from the data. End-matching may be used to remove the low-frequency components, whose side lobes cause spectral distortion.

Experiment B: End-Matching

Now the same 10 filters as in experiment A are used, but each red noise realization (time series) is end-matched before spectral estimation. Once again, the same data set is input to both MEM and the periodograms. Figure 5 the results with a dramatic improvement in the peric indices up to a true index of 4.0. The results at 4.5 and again unacceptable. Note also that the MEM res slightly but definitely worse. There is more scatter, a index of 0.5 the MEM results are biased upward.

Clearly then, end-matching is required for spect mation using periodograms: without it most of the s estimates are unreliable. Except in the low index ra true indices between 0.5 and 2.0, the periodograms ar ted. A spectral index of 2.0 could have arisen from a tr anywhere between 2.0 and 5.0.

Just as clearly, end-matching should be avoided MEM calculations, where it does not help but degra spectral estimates.

Experiment C: Windowing

The difficulty with the periodogram results at true in 4.5 and 5.0 may perhaps be explained by discontin slope of the original data set at the beginning and e suggested cure is "tapering," or looking at the data th window which deemphasizes the data at both ends b plying the data by a function that is near zero at

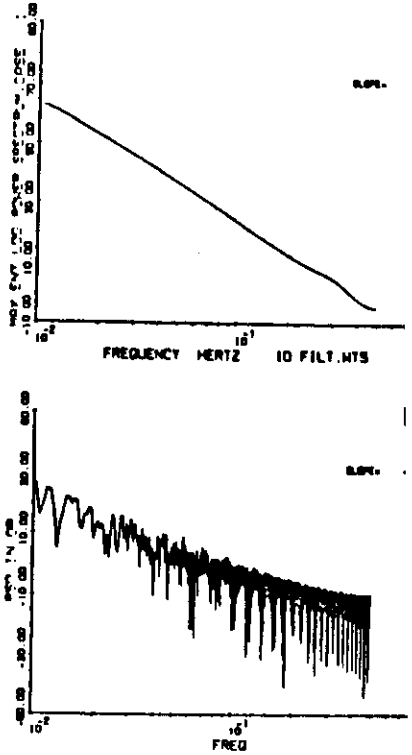


Fig. 4. (a) Maximum entropy spectrum of the signal from Fig The ordinate is $10\log_{10}$ (PSD). (b) Periodogram of the same si

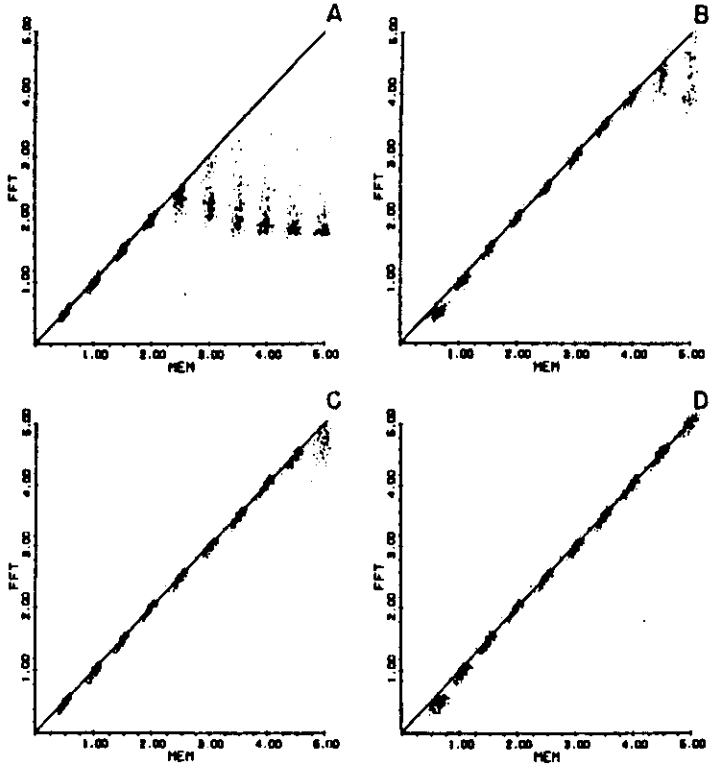


Fig. 5. Observed FFT (periodogram) index versus the MEM index (index is the negative value of the slope). There are 100 independent realizations of the power law process for each of the 10 indices 0.5, 1.0, ···, 5.0. In every case the same time series was used as input to MEM and to FFT (periodogram). (*a*) Raw data. (*b*) End-matched data. (*c*) Windowed data. (*d*) End-matched and windowed data.

points and higher in the center. This tapering does indeed reduce the tendency to discontinuity in slope.

The actual window used was recommended by *Welch* [1967] and is

$$W_j = 1 - [2j/(L+1) - 1]^2 \qquad j = 1, 2, \cdots, L$$

For those interested in "window carpentry," no fewer than 44 distinct windows are discussed in great detail in the review paper by *Harris* [1978]. The same 10 filters used in experiments A and B were again used in experiment C. Here the above window was applied routinely to all red noise realizations before the spectra were estimated, but end-matching was not used. The results are given in Figure 5*c*. The periodogram results are now usable for indices between 0.5 and 4.5, but the results at 5.0 are still biased. Once again, as with the use of end-matching, windowing degrades the MEM spectral estimate. The MEM scatter is even larger here then it was with end-matching.

Experiment D: Both End-Matching and Windowing

Here the original red noise realizations are first end-matched and then windowed before the spectral estimation. Figure 5*d* shows the results. Now the periodogram results are usable for the entire range of spectral indices from 0.5 to 5.0. The MEM results show that end-matching and windowing, taken either singly or together, degrade the MEM spectral estimates. Note that in all cases, for all experiments, the MEM scatter is smaller than the periodogram scatter. The quantitative results for all four experiments are collected in Table 1, where the mean, standard deviation, and maximum and minimum values are all given.

The results of the simulation are summarized in Table 2. The several experiments have shown that the Burg-MEM technique should be used on the raw data and that some form of windowing and detrending is required before the periodogram method can be employed for the estimation of the spectral index of a power law power spectrum. In Table 2 the standard deviation values were used to calculate the expected 90% confidence limits for the spectral index estimates. It is seen that the FFT or single-periodogram method produces an uncertainty which is almost a factor of 2 larger than the uncertainty for the Burg-MEM technique (± 0.015 versus ± 0.009). The 90% confidence intervals (CI in the table) are very small, and either method must be judged acceptable.

A difficulty in the use of the Burg-MEM technique can also

TABLE 1. Statistical Summary of the 4000 Experiments

True Index	Mean Index		Standard Deviation		Minimum		Maximum	
	MEM	FFT	MEM	FFT	MEM	FFT	MEM	FFT
				Experiment A: Raw Data				
0.5	0.5155	0.4991	0.0541	0.0678	0.3488	0.3547	0.6160	0.6546
1.0	1.0308	0.9975	0.0534	0.0705	0.8905	0.8263	1.1400	1.1786
1.5	1.5108	1.4956	0.0530	0.0698	1.3712	1.3380	1.6316	1.6646
2.0	1.9938	1.9749	0.0546	0.0711	1.8833	1.8246	2.1248	2.1588
2.5	2.5063	2.3468	0.0550	0.1280	2.3802	1.9569	2.6248	2.6553
3.0	3.0241	2.3374	0.0548	0.2784	2.8773	1.7457	3.1507	2.9979
3.5	3.5119	2.1354	0.0579	0.3060	3.3660	1.7264	3.6810	3.2385
4.0	3.9855	2.0058	0.0612	0.2506	3.8402	1.7499	4.1375	3.3308
4.5	4.4798	1.9134	0.0627	0.1897	4.3133	1.6931	4.6044	3.0970
5.0	4.9813	1.8694	0.0637	0.1855	4.8104	1.7218	5.1130	3.2826
				Experiment B: End-Matched Data				
0.5	0.6607	0.5096	0.0767	0.0699	0.4456	0.3480	0.8378	0.6743
1.0	1.0693	1.0003	0.0572	0.0661	0.9463	0.8597	1.2004	1.1852
1.5	1.5136	1.4964	0.0547	0.0695	1.3855	1.3518	1.6295	1.6906
2.0	1.9899	1.9928	0.0545	0.0708	1.8771	1.8062	2.1206	2.2038
2.5	2.5030	2.4940	0.0538	0.0697	2.3741	2.3082	2.6250	2.6812
3.0	3.0229	2.9923	0.0553	0.0770	2.8756	2.7881	3.1229	3.1696
3.5	3.5112	3.4860	0.0581	0.0751	3.3601	3.2890	3.6501	3.6359
4.0	3.9850	3.9550	0.0623	0.0775	3.8309	3.7450	4.1316	4.1355
4.5	4.4950	4.2794	0.0627	0.1629	4.3497	3.8593	4.6417	4.7137
5.0	4.9869	4.3143	0.0635	0.3632	4.8330	3.6926	5.1223	5.1244
				Experiment C: Windowed Data				
0.5	0.5102	0.4875	0.0622	0.0794	0.3019	0.2930	0.6249	0.6516
1.0	1.0243	0.9861	0.0625	0.0809	0.8033	0.7873	1.1311	1.1298
1.5	1.5041	1.4865	0.0650	0.0837	1.2805	1.2798	1.6271	1.6369
2.0	1.9871	1.9865	0.0685	0.0869	1.7483	1.7680	2.1192	2.1505
2.5	2.5007	2.4871	0.0698	0.0861	2.2533	2.2679	2.6351	2.6547
3.0	3.0217	2.9878	0.0697	0.0890	2.7594	2.7572	3.1627	3.1685
3.5	3.5104	3.4891	0.0696	0.0915	3.2532	3.2527	3.6536	3.6713
4.0	3.9824	3.9872	0.0732	0.0905	3.7381	3.7532	4.1281	4.1648
4.5	4.4811	4.4415	0.0793	0.0983	4.2217	4.1652	4.6250	4.6105
5.0	4.9830	4.7165	0.0819	0.1989	4.7177	4.0513	5.1505	5.1166
				Experiment D: End-Matched and Windowed Data				
0.5	0.6383	0.4877	0.0850	0.0792	0.4739	0.2928	0.8190	0.6519
1.0	1.0593	0.9858	0.0664	0.0811	0.8462	0.7873	1.2291	1.1301
1.5	1.5074	1.4864	0.0641	0.0837	1.2938	1.2796	1.6394	1.6370
2.0	1.9841	1.9865	0.0688	0.0867	1.7376	1.7695	2.1122	2.1509
2.5	2.4983	2.4871	0.0700	0.0861	2.2422	2.2622	2.6288	2.6556
3.0	3.0195	2.9878	0.0694	0.0894	2.7542	2.7558	3.1576	3.1696
3.5	3.5086	3.4886	0.0695	0.0920	3.2496	3.2525	3.6539	3.6630
4.0	3.9823	3.9861	0.0748	0.0923	3.7293	3.7641	4.1169	4.1504
4.5	4.4863	4.4926	0.0792	0.0955	4.2226	4.2289	4.6336	4.6677
5.0	4.9890	4.9884	0.0811	0.0952	4.7283	4.7529	5.1608	5.1671

Each entry is based upon a set of 100 realizations of the red noise process. FFT stands for FFT-based periodogram.

be seen in Table 2. The MEM estimates are biased. That is, the confidence interval does not include the true value. This small bias value is of the order of the uncertainty in the periodogram method. The bias can be changed by using a different order for the AR process. Recall that the order is set at 6; no attempt was made to optimize it.

But notice also that for MEM in six cases the error is negative and in four cases it is positive. For FFT the error is always positive, indicating a systematic bias. The periodogram systematically underestimates the spectral index. Most of the errors are quite close to the 10% confidence interval. For both MEM and FFT, however, the bias is quite small and should not prove troublesome.

It is noted that by averaging the spectral index estimates obtained from four periodograms the confidence bounds can be reduced to less than the bounds for a single Burg-MEM estimate. Using overlapping spectra as recommended [*Nuttall and Carter*, 1982], equivalent results could be obtained from averaged periodograms by using a data set 3 times the length of that required for the Burg-MEM analysis. The trade-off between the use of the two techniques is evident. The nonparametric method employing windowed and averaged periodograms requires more data to produce the same result as can be obtained from the Burg-MEM algorithm. The parametric Burg-MEM algorithm, however, requires the selection of the correct order to produce unbiased results, but the order is not known a priori. When faced with a time series from an unknown process, both techniques should be applied, and the parameters of the models (such as order of the process) adjusted to provide consistent results [*Jenkins and Watts*, 1968].

9. 45

TABLE 2. Summary of Simulation Experiments

True Index	MEM (Experiment A) Mean	Error[1]	CI	FFT (Experiment D) Mean	Error[1]	CI
0.500	0.516	−0.016	0.009	0.488	0.012	0.013
1.000	1.031	−0.031	0.009	0.986	0.014	0.013
1.500	1.511	−0.011	0.009	1.486	0.014	0.014
2.000	1.994	0.006	0.009	1.987	0.013	0.014
2.500	2.506	−0.006	0.009	2.487	0.013	0.014
3.000	3.024	−0.024	0.009	2.988	0.012	0.015
3.500	3.512	−0.012	0.010	3.489	0.011	0.015
4.000	3.986	0.014	0.010	3.986	0.014	0.015
4.500	4.480	0.020	0.010	4.493	0.007	0.017
5.000	4.981	0.019	0.011	4.988	0.012	0.016

[1]Error = true index − mean index.

CI is the expected 90% confidence interval $t\sigma/\sqrt{n}$, where t is the 5% critical value of Student's distribution using 100 degrees of freedom, σ is the standard deviation from Table 1, and $n = 100$. For a single observation, CI = $t\sigma$, i.e., 10 times as large. Note that if |error| > CI, a bias exists.

Discussion

The above 4000 runs, collected in four experiments of 100 runs each on 10 distinct spectral indices between 0.5 and 5.0, show that correctly applied, the maximum entropy method and periodogram techniques yield results which may be thought of as complementary.

1. The MEM spectral shape is always smooth and nearly linear. The unaveraged periodogram shape is highly variable and noisy.
2. When straight lines are fit to the spectra, the resulting spectral indices are more variable with periodograms than with MEM.
3. Slight biases can result from the use of MEM unless care is taken in determining the order of the process to be analyzed.

Explanation

Since the difficulties with the periodogram-based techniques have been explained briefly, it seems in order to present an intuitive explanation of the success of the MEM technique. The MEM spectrum is based upon the determination, from the data sample, of a prediction error filter, which finds the error in a one-step-ahead prediction as a linear combination of m previous sample values. The same filter is used to make predictions in both time directions (it is merely reversed to make predictions from the future into the past). The mean square prediction error in both time directions is minimized by varying the prediction error coefficients. Because the filter makes predictions of the time series based upon previous values of the time series itself, it is also called an autoregressive filter.

Now the red noise process whose spectrum we are trying to estimate is also an autoregressive process, and thus MEM is ideally suited to the estimation of the autoregressive parameters and indeed produces spectra which are close to ideal.

4. Application to Geophysical Data

Ionospheric scintillation occurs when a radio wave, transmitted by a satellite toward a receiver on the earth, passes through a disturbed ionosphere. Both the phase and the amplitude of the wave suffer low-frequency (~0.001 to 100 Hz) perturbations known as phase and amplitude scintillation, respectively. If the wave is detected in a suitable receiver, the scintillation can be separated from the carrier and can then yield important information on the nature of the ionospheric irregularities which are the source of the scintillation [*Yeh and Liu*, 1982].

The data to be analyzed here are amplitude scintillation data sampled at 36 Hz for 5 min from the MARISAT satellite in January 1981. Figure 6*a* displays the data set, which was chosen especially because it contains a quiet segment from 0 to about 2½ min, a moderately noisy section from 2½ to 3½ min, and a highly scintillating section from 3½ to 5 min. The changes in character of the noise record are quite abrupt and easy to see in the time record of Figure 6*a*.

It may legitimately be asked whether power spectrum analysis could be employed to monitor the development of such a process. A dynamic spectrum was constructed from the ~5-min data sample as follows. There were 10,981 observations in this data set, which was divided into 60 batches consisting of 361 points (10 s each) and overlapped by 181 points (5 s).

Some elementary statistics for each batch are shown in Figure 6*b*, which gives the maximum (top curve), standard deviation (middle curve), and minimum (bottom curve). The input power to each power spectrum is simply the square of the standard deviation. Figures 6*a* and 6*b* show that the signal is approximately stationary, that is, the mean and standard deviation are approximately constant, during the three separate time intervals 0–2½ min, 2½–3½ min, and 3½–5 min. These

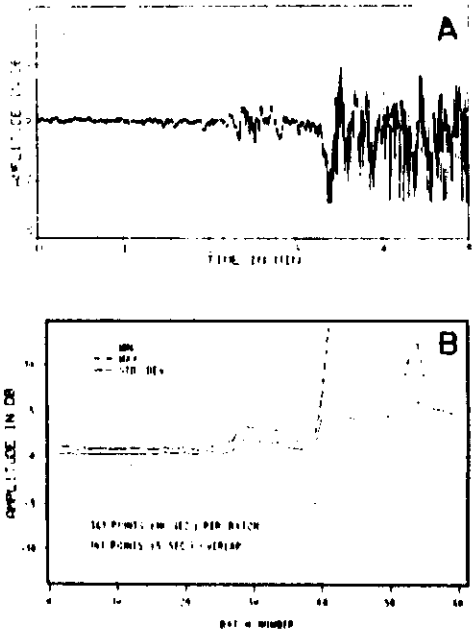


Fig. 6. (*a*) Amplitude scintillation data taken from the MARISAT satellite in January 1981. The sampling rate is 36 Hz. There is background noise from 0 to 2½ min, moderate scintillation from 2½ to 3½ min, and fully saturated scintillation from 3½ to 5 min. (*b*) Minimum, maximum, and standard deviation of each batch of 361 observations of data from Figure 6*a*. The batches are overlapped by 181 points.

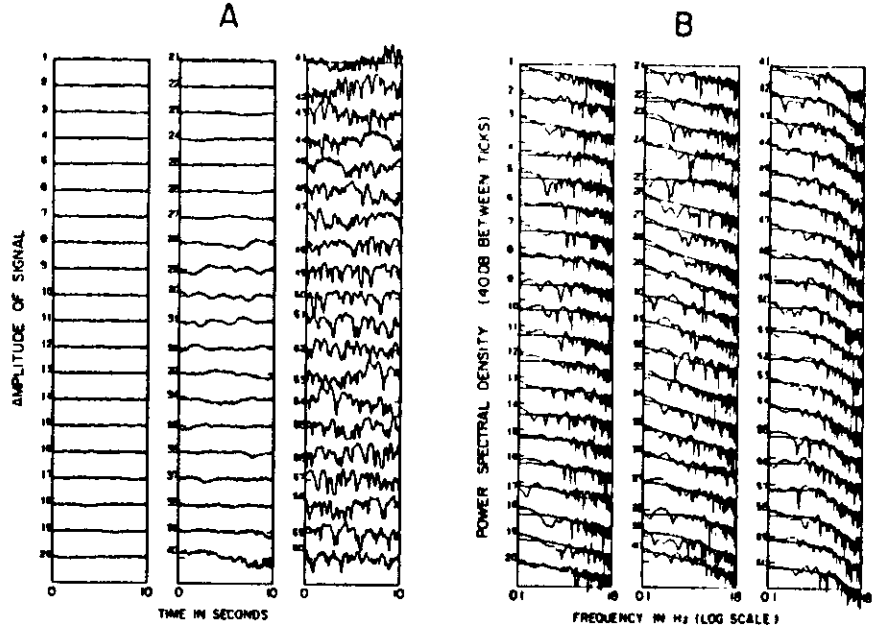


Fig. 7. (*a*) The 60 overlapped data batches. (*b*) Smooth curves: Burg-MEM spectra of order 5 of the time series shown in Figure 7*a*. Noisy curves: periodograms of the same data after end-matching and windowing.

times correspond to batches numbered 1–27, 28–39, and 40–60. A power spectrum was then obtained for each batch. The 60 overlapped data sections are shown in Figure 7*a*, and the MEM and unaveraged periodogram spectra are shown in an identical format in Figure 7*b*. The smooth curve is the MEM result based upon five prediction error filter weights applied directly to the raw data. The periodograms which appear as the noisy curves in Figure 7*b* were obtained on the same data sets after first end-matching and windowing using a three-term Blackmar-Harris window. This is one of the "best" windows described by *Harris* [1978]. It can be seen that in every case the Burg-MEM results are an "ideally" smoothed version of the periodogram results or, more important, that the periodogram results show the statistical fluctuations to be expected when no form of averaging is used. If one had only the periodogram results, one might be tempted to "see" significant spectral peaks, and these peaks in any case would tend to obscure real changes in slope in the underlying process. Such important changes in slope (and indeed changes in the character of the spectrum) are easy to identify in the MEM spectra and can be seen to correspond to changes in the character of the time series as seen in Figure 6. The smoothness of the MEM spectrum is affected by the choice of the order of the spectrum. Since the process is nearly stationary over a number of spectra, additional information can be obtained by averaging the spectral estimates obtained from the periodograms. Notice that the spectra numbered 27–37 inclusive are all extremely close to pure power law spectra with a single spectral index over the entire frequency band of interest, 0.1–18 Hz. This fact should serve to validate and motivate the simulation results presented earlier. Real geophysical data sets sometimes can be accurately represented as realizations of pure power law processes.

Notice also that the character of the individual data se changes quite sharply from batch 25 to batch 29. The Bur MEM spectra similarly show changes of character from bat 25 to batch 29. Once again the character of the signal and the associated spectra changes abruptly from pure power la to composite power law, approximated by two or more ap proximately linear sections at batch 38. Fully saturated scinti lation is evident by batches 40 and 41 in the time record. Th spectra here are composite power law, piecewise linear, wit three different spectral indices, small, medium, and large, an a spectral minimum near the Nyquist frequency (18 Hz). Thes results, the spectra of batches 38–41, are expanded and show in detail in Figure 8*a*. The sharp change of character fror batches 38 and 39 to batches 40 and 41 is dramatic and easil visible. Figure 8*b* shows the periodogram results. In order t distinguish the four curves at all, a bias of 20, 40, and 60 dB was added to curves 39, 40, and 41, respectively. The chang of character from curve 38 to curve 41 is obscured by th statistical fluctuation.

At the risk of belaboring the point, these qualitative change in character of the spectrum are clearly and easily visible i the Burg-MEM results but are at least partially obscured i the fluctuations which are the constant companion of una veraged periodogram results.

Note finally that certain features clearly visible in the Burg MEM spectra could not even be imagined by examining the time series record. For example, the fully saturated three-component spectra evident in batches 41–47 change to predominantly two-component (white plus red) spectra in batches

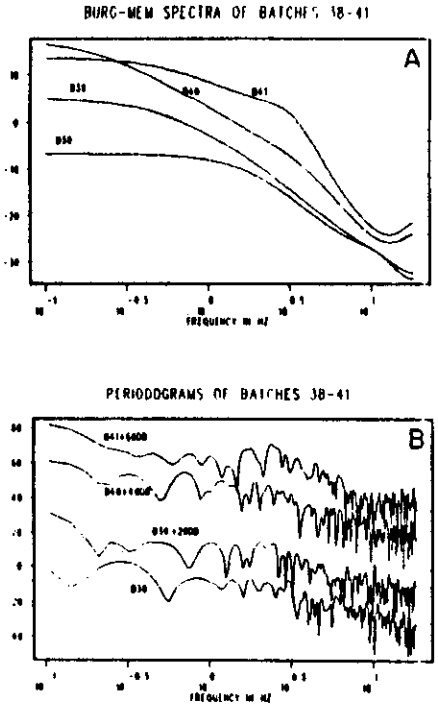


Fig. 8. Burg-MEM spectra of order 5 of data batches 38–41. (*b*) Periodograms of the same data after end-matching and windowing; 40, and 60 dB have been added to batches 39, 40, and 41, respectively.

54. Of course, it is the desire to see features not obvious in time record that motivates the use of the power spectrum he first place.

Another way of visualizing a dynamic power spectrum is a ee-dimensional surface representation with hidden lines re-ved. Figure 9*a* shows such a representation of the 60 over-ped spectra using Burg-MEM. The qualitative picture is of power spectral density changing smoothly with time h two regions of rapid but smooth increase. The com-ion picture from the overlapped periodograms, Figure 9*b*, wise shows two regions of rapid increases, but the high-plitude noise obscures all other features.

5. Conclusions

. The Burg maximum entropy method applied to time ies realizations of red noise processes produces consistently ooth power spectra.

2. Without averaging, the periodogram method applied to same data sets produces power spectra with large statis-al fluctuations which may obscure the true spectral vari-ons. End-matching or windowing or preferably both are olutely essential if meaningful periodogram results are to obtained.

At this point we will return briefly to the issue raised in the roduction, that of statistical consistency. It was mentioned t the periodogram produces a statistically inconsistent spectral estimate: the variability of the spectrum does not decrease as the data sample increases in size. This is a stochastic result which has nothing to do with the deterministic problems of "spectral leakage," which can be greatly reduced using end-matching or windowing or both.

This stochastic result has been well known for a long time. Subsequent statistical analysis has shown that there are two methods of reducing the variance in the periodogram spectral estimate. Both involve a smoothing procedure, one in the frequency domain, and the other in the time domain. In the frequency domain the variance can be reduced by applying various smoothing formulas to a set of adjacent spectral estimates in a periodogram. The simplest of the smoothing formulas is the running mean. The greater the number of ordinates which are smoothed, the greater the reduction in variance. But, of course, at the same time the frequency resolution of the smoothed spectrum is similarly reduced. Thus a trade-off must be made between decreased variance, which is desirable, and decreased resolution, which is undesirable.

In the time domain the smoothing procedure which is useful for reducing spectral variance is that of averaging successive independent or overlapped periodograms. The price paid here is a corresponding reduction in time resolution. Thus if, as in the MARISAT data, the spectrum changes abruptly in time, because the underlying time series changes abruptly, the process of averaging n consecutive periodograms would blur any sharp changes in character of the spectrum.

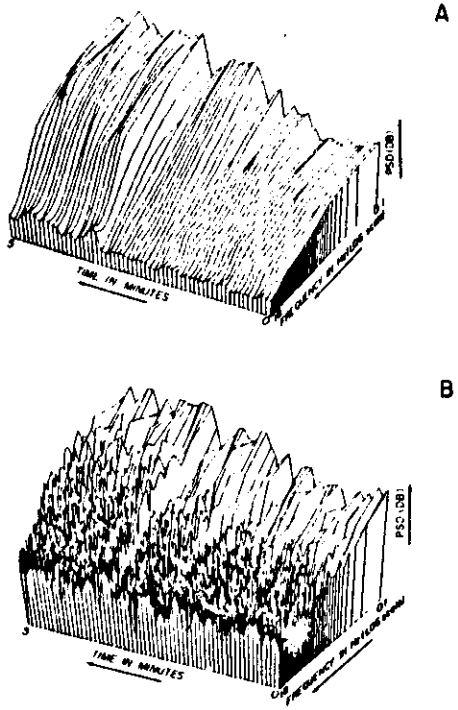


Fig. 9. (*a*) Order 5, Burg-MEM dynamic spectra of all 60 data batches. (*b*) Periodogram dynamic spectra of the same data after end-matching and windowing.

If, however, the process under scrutiny were stationary and sufficient data were available, the method of averaging periodograms would produce a reasonable, smooth spectrum. To illustrate this point, we return briefly to our simulation results and show in Figure 10 successively stacked and averaged periodograms using 1, 2, 3, 4, 5, 10, 25, 50, 75, and 100 independent periodograms and, on the bottom, a single maximum entropy spectrum for comparison. The original spectral index was 2.0 (slope of -2.0). The number given next to each spectrum is the rms deviation of the spectrum estimate from the power law spectrum determined by least squares. With 100 averaged periodograms the rms deviation is still a little larger than that of one maximum entropy spectrum. An adequate estimate of the spectral index, however, requires the use of only four periodograms, as indicated in the discussion of Table 2.

By way of a final resolution which may provide an intuitive understanding of the differences between maximum entropy and periodogram spectral analysis for red noise processes, Figure 11 shows maximum entropy power spectra of a single realization of a simulated power law process with index 2.5. Here we see the number of filter weights increasing from top to bottom and, at the very bottom, the periodogram (after end-matching and windowing). As the number of filter weights increases, the MEM spectral appearance becomes more and more jagged until at 512 weights the MEM spectrum resembles quite closely the raw periodogram. Note that there is very little change in the spectrum shown in the top four curves, with 2, 4, 8, and 16 weights, but that below that, for 32, 64, $\cdots$ weights, more and more meaningless detail is displayed.

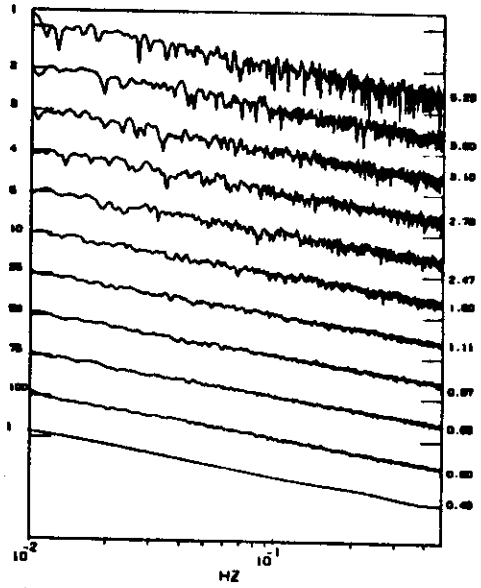


Fig. 10. Stacked and averaged periodograms (top 10 curves) of a simulated power law process with index 2. The number of independent spectra stacked is shown on the left (1, 2, 4, 8, $\cdots$), and the rms deviations of the displayed spectrum from the power law spectrum determined by least squares is shown on the right. The bottom spectrum is a single MEM spectrum. The vertical scale is 30 dB between ticks.

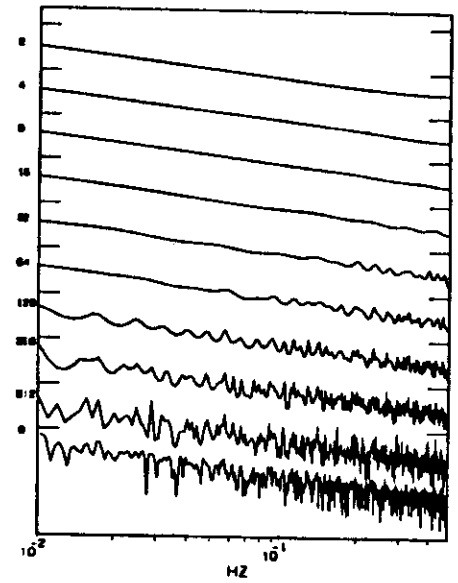


Fig. 11. Maximum entropy spectra (top nine curves) of a simulated power law process with index 2.5. The number of prediction error filter weights is shown at the left (2, 4, 8, 16, $\cdots$, 512). The periodogram is shown at the bottom for comparison. The vertical scale is 30 dB between ticks.

A complete FORTRAN package which finds Burg-MEM spectra has been prepared. Seriously interested scientists are invited to write to the author for a copy of the program and its documentation. Please indicate preferred tape density (800 or 1600 BPI) and code (EBCDIC or ASCII).

Acknowledgments. It is a pleasure to thank the reviewers, whose criticism and suggestions have helped me to produce a much better paper. I also want to thank Santimay and Sunanda Basu and Herbert C. Carlson for the use of the MARISAT data and for many very useful discussions. Finally, I am grateful to Celeste Gannon and Elizabeth Galligan for patiently and expertly typing the many, many draft versions as well as the final version of the manuscript.

The Editor thanks the referees for their assistance in evaluating this paper.

References

Armstrong, J. W., and W. A. Coles, Analysis of three-station interplanetary scintillation, *J. Geophys. Res.*, *77*, 4602–4610, 1972.

Burg, J. P., Maximum entropy spectral analysis, Ph.D. thesis, 123 pp., Stanford Univ., Stanford, Calif., 1975.

Burg, J. P., Maximum entropy spectral analysis, in *Modern Spectrum Analysis*, edited by D. G. Childers, IEEE Press, New York, 1981*a*.

Burg, J. P., A new analysis technique for time series data, in *Modern Spectrum Analysis*, edited by D. G. Childers, IEEE Press, New York, 1981*b*.

Childers, D. G. (Ed.), *Modern Spectrum Analysis*, IEEE Press, New York, 1981.

Crane, R. K., Spectra of ionospheric scintillation, *J. Geophys. Res.*, *81*, 2041–2050, 1976.

Digital Signal Processing Committee, IEEE Acoustics, Speech, and Signal Processing Society, *Programs for Digital Signal Processing*, IEEE Press, New York, 1979.

Harris, F. J., On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE*, *66*, 51–83, 1978.

Jaynes, E. T., On the rationale of maximum-entropy methods, *Proc. IEEE*, *70*, 939–952, 1982.

Jenkins, G. M., and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, Calif., 1968.
Kay, S. M., and S. L. Marple, Jr., Spectrum analysis—A modern perspective, *Proc. IEEE*, *69*, 1380–1418, 1981.
Larsen, M. F., M. C. Kelley, and K. S. Gage, Turbulence spectra in the upper troposphere and lower stratosphere at periods between 2 hours and 40 days, *J. Atmos. Sci.*, *39*, 1035–1041, 1982.
McClellan, J. H., T. W. Parks, and L. R. Rabiner, FIR linear phase filter design program, in *Programs for Digital Signal Processing*, IEEE Press, New York, 1979.
McKenzie, J. F., Similarity solution for non-linear damping of Alfvén waves, *J. Plasma Phys.*, *28*, 317–323, 1982.
Nuttall, A. H., and G. C. Carter, Spectral estimation using combined time and lag weighting, *Proc. IEEE*, *70*, 1115–1125, 1982.
Rabiner, L. R., and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N. J., 1975.
Radoski, H. R., P. F. Fougere, and E. J. Zawalick, A comparison of power spectral estimates and applications of the maximum entropy method, *J. Geophys. Res.*, *80*, 619–625, 1975.
Radoski, H. R., E. J. Zawalick, and P. F. Fougere, The superiority of maximum entropy power spectrum techniques applied to geomagnetic micropulsations, *Phys. Earth Planet. Inter.*, *12*, 208–216, 1976.
Schuster, A., On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, *J. Geophys. Res.*, *3*, 13–41, 1898.
Shannon, C. E., and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.
Welch, P. D., The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoust.*, *AU-15*, 70–73, 1967.
Woo, R., and J. W. Armstrong, Spacecraft radio scattering observations of the power spectrum of electron density fluctuations in the solar wind, *J. Geophys. Res.*, *84*, 7288–7296, 1979.
Yeh, K. C., and C. H. Liu, Radio wave scintillation in the ionosphere, *Proc. IEEE*, *70*, 324–360, 1982.

P. F. Fougere, Air Force Geophysics Laboratory/LIS, Hanscom Air Force Base, MA 01731.

(Received March 25, 1983;
revised December 18, 1984;
accepted December 19, 1984.)

Radio Science, Volume 20, Number 3, Pages 463–476, May–June 1985

Large-amplitude electric field fluctuations near the Harang discontinuity

Nicolaos A. Saflekos[1]

Physics Department, Boston College, Chestnut Hill, Massachusetts

William J. Burke and Paul F. Fougere

Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts

(Received August 6, 1984; revised December 7, 1984; accepted February 4, 1985.)

Recently, much attention has focused on physical mechanisms responsible for the formation and evolution of the plasma density irregularities in the *F* region ionosphere that disrupt communications at high latitudes. The polar orbiting S3-2 satellite with its complement of high resolution of plasma, particle and field detectors provides useful information on the dynamics of some irregularities. This study focuses on plasma density and electric field measurements taken as the satellite flew poleward along the Harang discontinuity during a substorm. The position of the Harang discontinuity is established from simultaneous ground magnetometer measurements and DMSP imagery. The plasma density showed a distinct maximum near the poleward boundary of auroral electron precipitation that was two orders of magnitude higher than in the trough and in the (winter) polar cap. Very intense (> 100 mV/m) electric field fluctuations were detected along the steep, equatorward, plasma density gradient near the poleward boundary of the oval. Power spectral densities of these fluctuations did not obey expected power laws but had strong spectral lines between 1 and 16 Hz that do not seem to map to the peak of the *F*-layer. This suggests a magnetospheric source for at least some high-latitude electric field irregularities. Simultaneously measured spectra for plasma densities consistently showed k^{-2} power laws over the 0- to 8-Hz range. This is consistent with the predictions of nonlinear, gradient drift theory.

INTRODUCTION

Results from the DNA Wideband satellite have sparked a renewed interest in the morphology and aetiology of ionospheric irregularities at high latitudes. The Wideband scintillation measurements indicate that 1- to 10-km-scale irregularities in the midnight sector of the auroral oval are *L* shell-aligned sheets [*Rino and Matthews*, 1980]. They are found in regions of large-scale field aligned currents and TEC gradients [*Fremouw et al.*, 1977; *Rino et al.*, 1978]. Because convection in the midnight sector has a strong equatorward component, regions of poleward (equatorward) density gradients should be stable (unstable) against the gradient drift (E × B) instability [*Simon*, 1963; *Ossakow and Chaturvedi*, 1979] showed that even in regions of poleward density gradients the plasma could be linearly unstable to the current convective instability [*Kadomtsev*, 1965]. This m requires the presence of field-aligned currents car by thermal plasma with strengths sufficient to o come the effects of stabilizing E × B drifts. Howe the linearly unstable modes have wave vectors pendicular to the density gradients and thus can explain *L* shell-aligned sheets. Subsequent anal [*Chaturvedi and Ossakow*, 1979; *Keskinen et* 1980] show that linearly unstable modes tran energy to harmonics that are nonlinearly unst and whose wave vectors are along the density gr ent. Numerical modelling of the current-convec instability predicts that at long wavelengths the n linearly saturated waves should have power spec densities that follow power laws with spectral ind in the 2–2.5 range [*Keskinen and Ossakow*, 1982].

The Wideband observations and their plausible planation in terms of current-convective instabili eventually led to the launch of the HILAT satellit July 1983. In the meanwhile experimental effo were extended to exploit other sources of informat about density structures in the ionospheric *F* regi Based on ISIS 1 thermal plasma measureme [*Phelps and Sagalyn* [1976] found that in the auro

[1] Now at Southwest Research Institute, Department of Space Sciences, San Antonio, Texas.

Paper number 5S0097.
0048-6604/85/005S-0097$08.00

48

asma density zone irregularities obeyed a power ectral law with an index of ~2. Rocket measure-nts in conjunction with Wideband and Chatanika oherent radar backscatter operations detected aller spectral indices [*Kelley et al.*, 1980]. The dif-ences between the satellite and rocket measure-nts can be reconciled if one considers that struc-es, with scales of several tens of kilometers, could t be detected in a rocket experiment. *Kelley et al.* 982] cite evidence of such tens of kilometers struc-e in measurements by the Chatanika radar in the dnight sector [cf. *Kelley et al.*, 1982, Figure 4]. ey suggest that the tens of kilometers scale vari-ons reflect structure in low-energy (< 1 keV) elec-n precipitation. This conjecture has been verified the polar cap [*Weber and Buchau*, 1981] and in auroral oval [*Basu et al.*, 1983] where several ndred electron volts electron structures are found coincide with topside density enhancements.

The purpose of this paper is to present a case dy of topside ionospheric irregularities observed S3-2 near the Harang discontinuity during a sub-rm. The irregularities were detected by both the ctric field and plasma density sensors, embedded an extensive, morningside, region 2 [*Iijima and temra*, 1978] current system. These field-aligned rents are carried mostly by energetic, precipitating ctrons, not by thermal particles. In all but one tance the irregularities were found in regions of uatorward density gradients. Despite our inability measure the westward component of the dc elec-field, this aspect of the data is consistent with the dient drift hypothesis. In this case, however, the ctrum of the electric field fluctuations showed dis-ct spectral lines in the 1- to 16-Hz (7.5- to 0.5-km) ge. It is not understood how such irregularity ctra are explained by presently existing theory.

The paper is made of three parts describing the 2 instrumentation, irregularity measurements and liscussion of their relationship to our present un-standing of irregularity formation.

INSTRUMENTATION

The S3-2 satellite was launched in December 1975 o a polar orbit with an initial apogee, perigee and lination of 1557 km, 240 km and 96.3°, respec-ly. It was spin stabilized with a nominal spin iod of 20 s and a spin axis perpendicular to the ital plane. There were two independent scientific kages to measure aeronomic and mag-osphere/ionosphere coupling parameters. During ew brief periods of ~10 min duration, while the satellite was within view of a tracking station, data from both systems were collected simultaneously. The aeronomy experiments included an ion mass spectrometer. The magnetosphere/ionosphere coupling package consisted of detectors to measure quasi-dc electric and magnetic fields, the flux of auroral electrons, the densities and temperatures of thermal electrons and the drift motion of thermal ions. The instrumental packages and modes for data reduction have been described by *Phibrick* [1976] and *Burke et al.* [1980].

OBSERVATIONS

During the period 2018–2028 UT on January 6, 1976, a Defense Meteorological Satellite Program (DMSP) satellite and the S3-2 satellite crossed the nightside auroral region. Optical imagery of the auroral scene is given in Figure 1. Two grids superposed on the image represent geographic and geomagnetic coordinates projected to an altitude of 100 km. The projections of the subsatellite tracks show that DMSP was moving westward, morning to evening, across the midnight sector of the oval and S3-2 poleward. Ephemeris data found at the bottom of Figure 2, show that the S3-2 trajectory was slightly to the east of the MLT midnight meridian. The separation in time of the trajectories crossing was approximately 2 min. This can be seen by comparing the position of DMSP as a function of universal time, given by the scale to the left of the image, and that of S3-2 explicitly marked along its trajectory. For the purpose of understanding overall geomagnetic conditions we have also given the AKR power measured by the Hawkeye satellite at an altitude of 20 R_E and normalized to an emission source at 7 R_E. The horizontal bars to the left of Figure 1 represent 3-min averaged values of AKR plotted at the midpoint of the interval. Geomagnetic conditions at the time of the S3-2 pass have been established through an examination of standard geomagnetic indices and auroral magnetometer records January 6, 1976, was a moderately disturbed day with the $\Sigma Kp = 20$ and $Kp = 4$ in the 1800–2100 UT interval. The pass took place during an expansion phase of a multiple substorm period that began at 1530 UT and lasted through the remainder of the day. At 2026 UT the *AE* and *AL* indices were 385 and 340 nT, respectively. The AKR power at 2025 UT was 5×10^{-4} W m^{-2} Hz^{-1}, the highest value achieved during the ongoing substorm. The relative strengths of the eastward and westward electrojets determined from *AE* and *AL* and the DMSP imagery suggest that most of

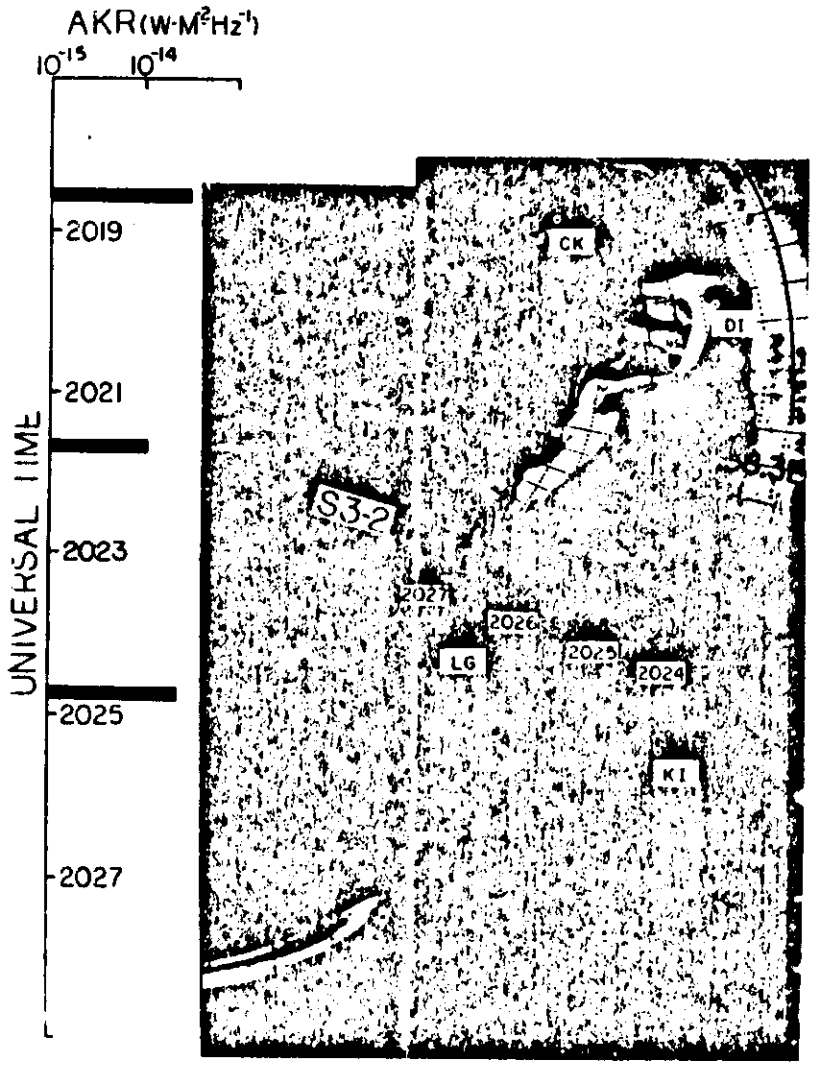


Fig. 1. Imagery of northern polar region from DMSP between 2019 and 2029 UT on January 6, 1976. The subsatellite trajectory of S3-2 and the AKR power flux are provided on and to the left of the image, respectively.

the AKR came from the postmidnight sector of the oval.

The locations of magnetic observatories at Leirvogur (LG), Kiruna (KI), Cape Chelyushkin (CK) and Dixon Island (DI) are also given in Figure 1. An examination of horizontal and vertical magnetic deflections (not shown) indicate that at the time of the S3-2 overpass both Leirvogur and Kiruna were

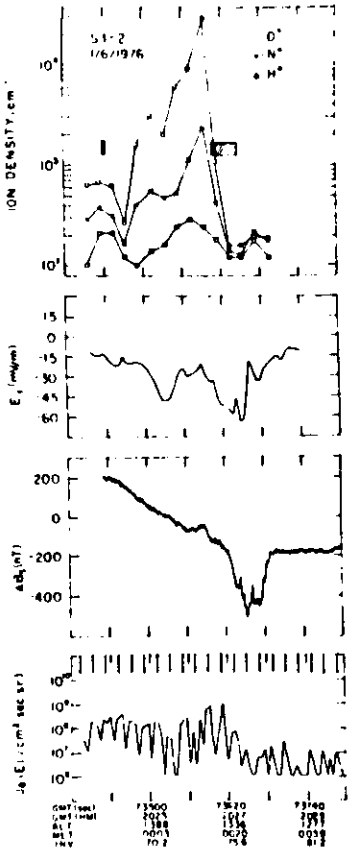


Fig. 2. Measurements from S3-2 of (1) the 20-s averaged densities of O^+, N^+, and H^+, (2) the 5-s averages of the electric field component along the satellite trajectory, (3) the magnetic perturbation transverse to the satellite trajectory, and (4) the directional flux of electrons with energies between 80 eV and 17 keV. The arrows in the top panel mark the two regions where large amplitude electric field fluctuations were detected.

nder the influence of an eastward electrojet located lose to the latitude of Kiruna. At approximately 032 UT the direction of the electrojet changed from astward to westward. Both Cape Chelyushkin and ixon Island were continuously under the influence f a westward electrojet located at intermediate latiudes. As discussed below, the electric fields and field-ligned currents measured by S3-2 are characteristic of the morning sector. From its proximity to both Leirvogur and Kiruna we conclude that the S3-2 trajectory was just to the morning side of the Harang discontinuity.

Figure 2 summarizes measurements from four instruments taken during the auroral zone passage of S3-2. The top panel gives 20-s average densities of H^+, N^+, and O^+ ions. The second panel gives the 5-s average values of the electric field component along the satellite trajectory ($E_\perp$). Negative values of $E_\perp$, antiparallel to the satellite velocity, mean that the convective field has an equatorward component. The third panel is a plot of magnetic perturbations transverse to the satellite trajectory. In our satellite coordinate system, regions of negative (positive) slope in plots of ΔB_z versus time are consistent with field-aligned current sheets directed out of (into) the ionosphere. The bottom panel gives the directional flux of electrons with energies between 80 eV and 17 keV. Intermittently placed arrows at the top of this panel indicate times when the detector sampled downward moving electrons. Electron fluxes of $\sim 10^8$/cm^2 s sr isotropically distributed over the downcoming hemisphere indicate that at the time of instrument turnon the satellite was in the region of diffuse auroral precipitation. The presence of field-aligned currents in the diffuse aurora (region 2) that are directed out of the ionosphere and of equatorward electric fields that drive westward Hall currents supports our contention that the S3-2 satellite was just on the morning side of the Harang discontinuity. In this region one also expects that the electric field should have a westward component driving an equatorward component of convection. Unfortunately, owing to uncertainties in surface contact potentials of the axial dipole, the dc part of the westward electric field component could not be reliably measured.

Returning to the top panel of Figure 2, we note that at S3-2 altitudes between 1400 and 1300 km, the cold plasma was predominantly made up of heavy ions O^+ and N^+ and had large-scale north-south gradients. The plasma density rose from $\sim 10^3$ cm^{-3} at invariant latitude 67° to $\sim 3 \times 10^4$ at 75°, then decreased rapidly to 3×10^2 at 77°. We have examined the raw electric field measurements and found two regions of rapid electric field variations (discussed below) marked by bars in the top panel of Figure 2. Outside of these two regions, variations in the measured electric fields were smooth and orderly. The equatorward region of electrostatic fluctuations, detected at 73440 UT s occurred in a relative plateau of cold plasma density. The poleward bursts of electric fluctuations extend across the steep density gradient between 75° and 77° invariant latitude. These bursts occur in the most poleward portion of the region 2 [*Iijima and Potemra*, 1978] currents and across the transition from central plasma sheet to boundary plasma precipitation [*Winningham et al.*, 1975]. In this region the average energy of precipitating electrons typically changes from a few keV to several hundred electron volts. Electrons with lower energies produce ionization at F rather than E layer altitudes [*Roble and Rees*, 1977].

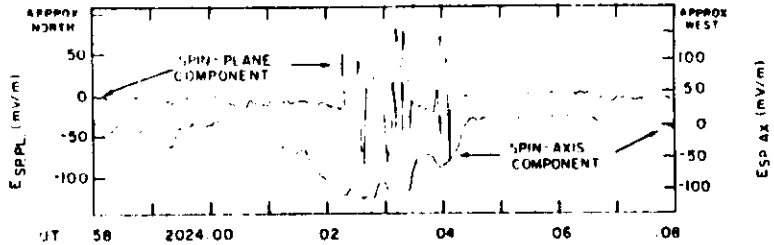


Fig. 3. Electric field measurements from the spinning and axial dipoles in the equatorward region of fluctuations marked in the top panel of Figure 2.

Figures 3 and 4 provide detailed measurements of the spin-plane and spin-axis dipoles in the vicinity of the equatorward and poleward electric field fluctuations, respectively. In the case of the equatorward event, rapid electric field fluctuations with amplitudes of up to ±50 mV/m are observed in the north-south component between 2024:02 and 2024:04 (73442-44 s) UT, but not in the east-west component. Rather, over this 2-s period there is a strong (~100 mV/m), smooth, eastward turning of the electric field. An eastward directed electric field component is not expected in the midnight sector. Its narrow latitudinal extent of ~0.1° suggests that it is derived from a region of positive (negative) space charge located to the west (east) of S3-2's trajectory. This is exactly the space charge distribution expected where the polarity of the magnetospheric shielding layer changes from positive before midnight to negative after midnight [*Harel et al.*, 1981; *Rich et al.*, 1980].

Figures 4a and 4b give the potential differences measured between the ends of the spin plane and axial dipoles, respectively, in the vicinity of the poleward electric field fluctuations. The data are arranged to span four consecutive 10-s intervals. The end points of the four periods correspond to times when the spinning dipole was most closely aligned with the earth's magnetic field. In preparing the spin-plane data for presentation the potentials due to dc biases and cross-field motion of the satellite have been subtracted from raw measurements. Since the dc bias is not known, only the cross-field induction potential has been subtracted from raw, axial dipole measurements. To convert potential differences into electric fields it is necessary to divide by the dipole lengths of 27.72 m (spin plane) and 11.0 m (axial). In the case of the spinning dipole it is also necessary to divide by sin Ψ, where Ψ is the angle between B and the dipole [*Burke et al.*, 1980].

Both Figures 4a and 4b show that in the first 10-s interval the two electric field components were slowly varying in amplitude and direction. Rapid electric field fluctuations are observed throughout the last three intervals. The most intense fluctuations were detected during the second interval in which amplitudes of ±4 V and ±2 V were typical of the spinning and axial dipole measurements, respectively. These correspond to electric field amplitudes of ±140 mV/m. There are three points that we wish to make regarding the electric field variations. First, these fluctuations differ significantly in frequency, amplitude and polarization from those shown in Figure 3. Second, there is a high degree of correlation between the fluctuations measured by the two dipoles. Third, the fluctuations are largely electrostatic rather than electromagnetic. The last point is easily established. If the waves were electromagnetic there would be magnetic fluctuations δB related to the electric fluctuations δE.

$$\delta B = \delta E / V_A \qquad (1)$$

where V_A is the local Alfven speed. At an altitude of 1300 km and magnetic latitude of 75° the earth's magnetic field's intensity is $\sim 3.5 \times 10^{-5}$ T. With an ambient plasma density, consisting mostly of O^+, in the $10^4 - 10^5$ cm^{-3} range, we calculate that V_A

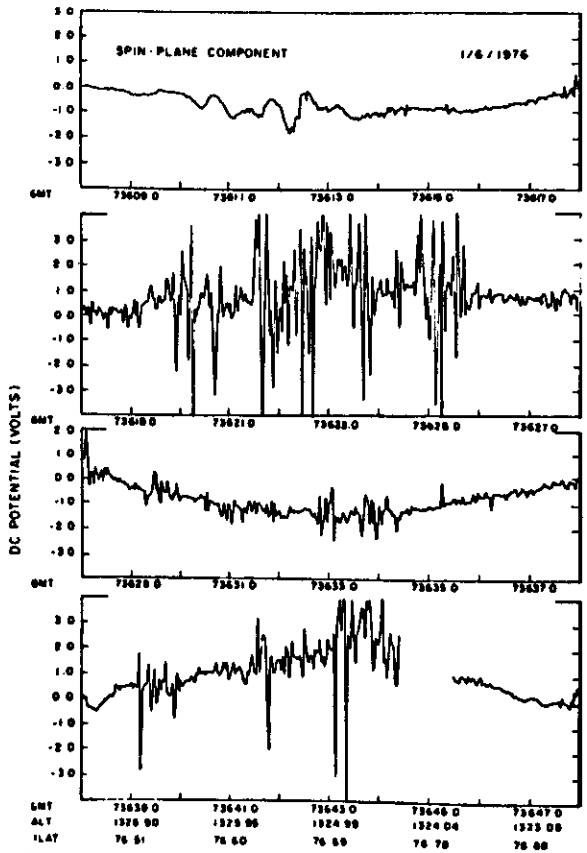


Fig. 4a. Potential differences measured across the spinning dipole in the poleward region of electric fluctuations marked in the top panel Figure 2.

should be between 2 and 6 × 10^3 km/s. For electric field fluctuations of ±140 mV/m, equation (1) predicts δB variations of ±(25 to 75) nT. With a 5 nT/bit resolution sampled 32 times per second, such magnetic fluctuations should be easily detectable. We have examined our magnetic field measurements at maximum resolution and have not found variations corresponding to the observed electric variations. In fact, since all data points are reproduced in the ΔB_z panel of Figure 2, such Alfven waves would appear as a burst of noise near 73620 s UT.

To facilitate comparisons with previously reported irregularity measurements and theoretical expectations, we have performed power spectral analyses on the δE fluctuations shown in Figures 4a and 4b by using the method of maximum entropy [*Ulrych and Bishop*, 1975; *Radoski et al.*, 1975; *Fougere*, 1977]. Figures 5a and 5b give power spectral densities (PSD) as functions of frequency for the spinning and axial dipole measurements by using 10 and 30 filter weights, respectively. The higher filter weight analysis more efficiently resolves individual spectral line contributions. The scale lengths cited at the top of the figures assume that the irregularities are stationary in the plasma's frame of reference [*Baker et al.*, 1983] and that the satellite speed is 7.5 km/s.

The analysis was performed by first subtracting the mean and any linear trend from the data. Figure 5 contains results from five consecutive 10-s intervals. Analyses were performed upon 7 s of data in the first four cases. Because of an instrument calibration beginning at 73644.5 s UT, only 5 s of data were used in the last cases. The first spectrum, which is from the 10-s interval prior to encountering the δE fluctuations is provided as a background reference. Also for reference we have superimposed lines with slopes characteristic of k^{-2} and k^{-3} power laws, where k represents wave number (2π times frequency/satellite velocity). Finally, above the high-frequency end of each spectrum in Figure 5a we have noted the total power under the curve normalized to the power contained in spectrum 3 for the spinning dipole. Note

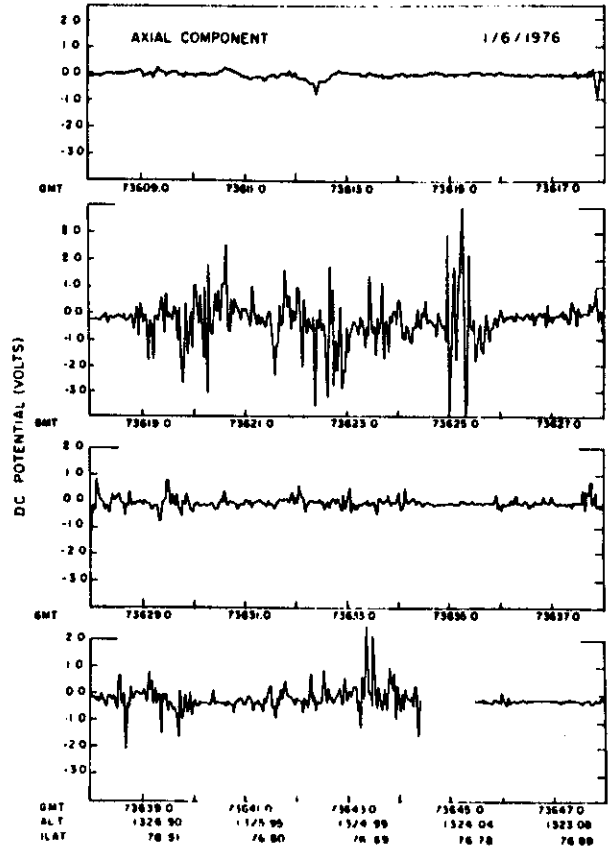


Fig. 4b. Potential differences measured across the axial dipole in the poleward region of electric fluctuations marked in the top panel of Figure 2.

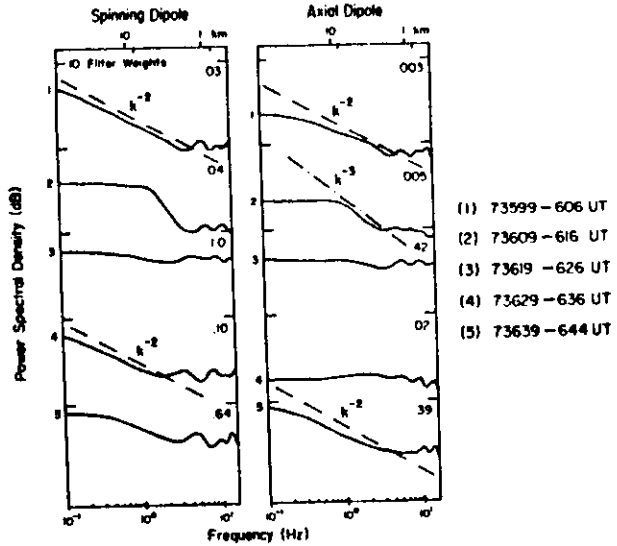


Fig. 5a. Maximum entropy power spectral densities versus frequency using 10 filter weights, prior to and during the interval of electric field fluctuations shown in Figure 4. Tick marks are separated by 40 dB. The integrated powers normalized to case 3 for the spinning dipole are noted in the upper right corners.

that the power contained in spectrum 3 is about 2 orders of magnitude above background.

There are five aspects of the PSD's shown in Figure 5 that we wish to emphasize:

1. Except in the background case, none of the spectra follows a power law over the entire frequency range.

2. In the portions of individual spectra showing power law dependencies the spectral index is in the 2–3 range. However, in the cases of greatest intensity, 3 for both dipoles, the spectra are more characteristic of white noise rather than power laws.

3. The power and frequencies of spectral lines vary from dipole to dipole and from interval to interval. However, in the region of strong E fluctuations, cases 2–5, there are consistently spectral lines near 1.5, 2.5 and 10 Hz.

4. There exists no simple harmonic relationship between spectral lines.

5. The satellite spin (0.05 Hz) has no influence on the spectral analysis.

One particularly intriguing aspect of the S3-2 measurements is the presence of discrete lines in the E PSD's between 73619 and 73645 s UT. Signals with frequencies near 10 Hz are evident in the raw data shown in Figures 4a and 4b. To further study these signals we passed the raw electric field measurements through the 9- to 10.8-Hz filter shown in Figure 6. The results of this analysis for the interval 73642–73643 s UT show a very narrow spectral line with a central frequency of 9.5 Hz. In the right panels of Figure 6 we have plotted the 9.5-Hz component of the E fluctuations detected by the spinning and axial dipoles. Over this interval the electrostatic waves are linearly polarized at an angle of ~45° to the satellite trajectory. At other subintervals the polarization was not so simple.

Representative measurements from the Langmuir probe and the ion drift meters are given in the top and bottom panels of Figure 7. The electron density measurements were taken with the sensor grid in the constant voltage mode. Data taken during periods when the grid voltage was being swept have been suppressed. The ion density measurements are normalized to times when the sensor in the spin plane faced in the ram direction. A cursory glance at these measurements reveals both large- and small-scale variations in the plasma density. Large-scale variations measured by these thermal plasma sensors are consistent with those from the ion mass spectrometer. For example, at 73513 and 73594 s UT the ion mass spectrometer showed local density maxima of 3×10^3 and 3×10^4 cm^{-3}, respectively. At these times the Langmuir probe measurements were $\times 10^3$ and 4×10^4 cm^{-3}. Drift meter sensors 5 measured a density maximum of 4×10^4 cm^{-3} 73593 s UT. Note that where the plasma sens measurements were sampled 16 times per second, t mass spectrometer data plotted in the top panel Figure 2 were sampled once every 20 s when t detector aperture faced the ram direction.

In examining small-scale variations care must exercised to distinguish between current fluctuatio due to variations in the ambient density and satelli potential. Satellite potentials change in response varying fluxes of auroral particles. Auroral electro tend to charge satellites negatively, thus suppressi thermal electron currents and enhancing thermal ic currents. In crossing regions of thermal density flu tuations, the ion and electron detector measuremen should correlate positively. In regions of satellite p tential fluctuations, variations of the thermal senso should anticorrelate. Comparisons of the tv measurements are valid only when one of the i sensors faces close to the ram direction. With t limitation, we have compared the three sets measurements and are only able to identify o "electron density" decrease as clearly due to an i creasingly negative satellite potential. This occurr at 73565 s UT and is marked by the symbol ϕ_s in t top panel of Figure 7. Where the electron curre decreased by a factor of 4, the current to drift me sensors 4–8 showed no significant modification.

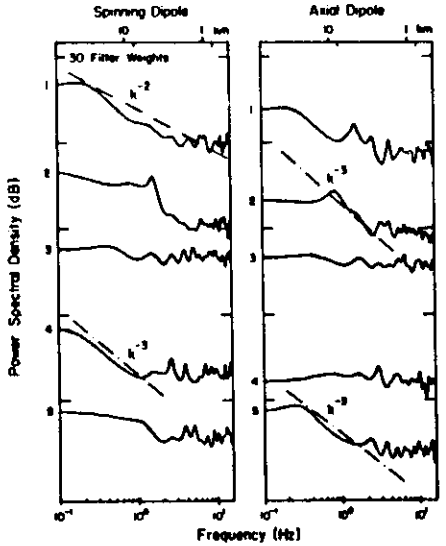


Fig. 5b. Maximum entropy power spectral densities for electric field fluctuations versus frequency using 30 filter weight for same intervals as Figure 5a.

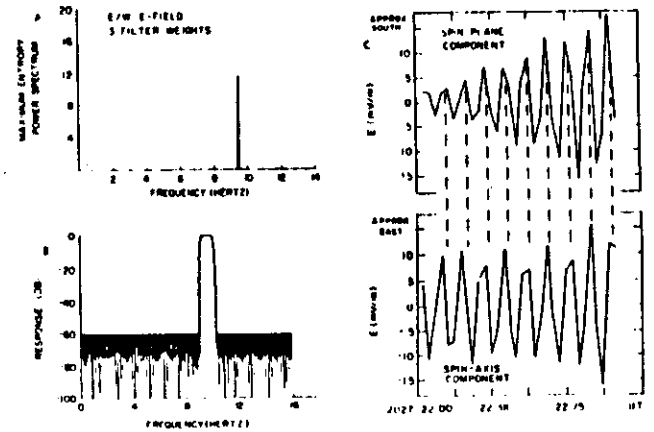


Fig. 6. Highly peaked PSD at 9.5 Hz (Figure 6a) obtained from the axial dipole for the 1-s interval following 20:17:22 (73642 s) UT using the 9- to 10.8-Hz filter shown in Figure 6b. Results of passing signals from both dipoles through this filter are shown in Figure 6c. The dashed lines in Figure 6c are aids to the eye for perceiving phase coherence between the two signals.

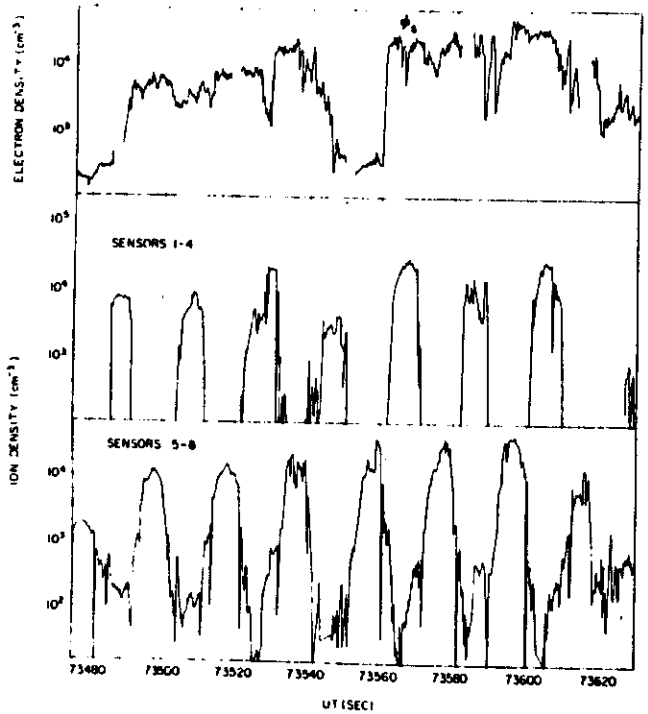


Fig. 7. Densities of thermal electrons and ions measured by the Langmuir probe and the two ion drift meters on S3-2.

Turning attention to the electron density variations, we note three features of the measurements: (1) electron density decreases can be greater than an order of magnitude (e.g., 73540–73560 s UT), (2) poleward density gradients tends to be very steep and devoid of smaller scale variations (e.g., 73490, 73530 and 73560 s UT), and (3) equatorward density gradients are relatively weak and have smaller-scale irregularities embedded in them (e.g., 73525, 73540, 73610 s UT). It should be noted that the scales of small-scale irregularities embedded in regions of equatorward density gradients are much longer than those of electric field variations shown in Figure 4.

To quantify the spectral characteristics of the density fluctuations, we have performed maximum entropy analyses of Langmuir probe measurements from the 73599–73544 s UT interval. The first four PSD's for $\delta n/n$ shown in Figure 8 were calculated by using data from 10 s (160 samples) of data. Otherwise, they are from the same intervals as the δE spectra of Figure 5. The fifth spectrum was compiled from 5 s (80 samples) of data. The dashed lines represent k^{-2} PSD's. There are four aspects of Figure 8 worth noting:

1. All of the $\delta n/n$ spectra obey power laws over most of the 0.1- to 8-Hz frequency range.
2. Spectra 1, 2 and 3 show a distinct spectral line centered at 5.5 Hz. No similar line appears in corresponding δE spectra.
3. The line doublet centered at 4.5 and 6.5 Hz in spectrum 5 has no analogate in the δE spectra.
4. Because of the different sampling rates, $\delta n/n$ and δE spectra cannot be compared to the 8- to 16-Hz range.

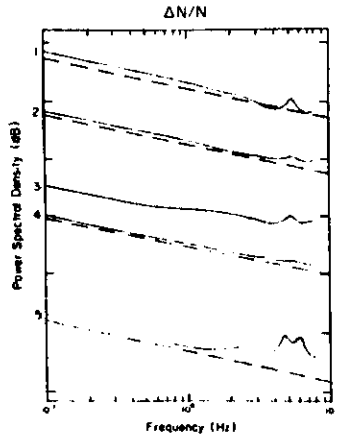


Fig. 8. Power spectral densities for $\delta n/n$ for same five intervals as δE fluctuations in Figure 5. Dashed lines represent k^{-2} spectra.

SUMMARY AND DISCUSSION

In the previous section we have described a set of observations taken during a substorm period on January 6, 1976, in which DMSP and S3-2 crossed the midnight sector of the auroral oval. The S3-2 trajectory carried it just to the morningside of the Harang discontinuity where convection electric fields have relatively strong westward components. Throughout the pass O^+ and N^+ were the dominant ion species. Two bursts of strong electric field fluctuations were observed by the S3-2 electric field detector. The equatorward burst was detected in a region of no perceptible density gradient. It appears to be associated with an ionospheric projection of the magnetospheric shielding layer rather than being a manifestation of a gradient driven plasma instability. The poleward region of electric field fluctuations spanned the location of an equatorward density gradient in a region of upward field-aligned currents and in the transition between central and boundary plasma sheet precipitation. The absence of δB fluctuations suggests that the fluctuations are electrostatic. Spectral analysis showed significant power in the 0.7- to 1-km-scale range (8–10 Hz). Analysis of the Langmuir probe and drift meter measurements revealed δn variations with scale sizes >2 km. The smaller scale structures appeared in regions of equatorward gradients with tens of kilometers scale size density fluctuations. In the region of strong δE fluctuations the $\delta n/n$ PSD obeyed k^{-2} power laws over most of the 0.1- to 8-Hz range. Individual lines at 5.5 Hz have no analogates in the δE PSD's.

There are three aspects of the S3-2 measurements that relate to our current understanding of irregularity formation. These are (1) the relationship between electron fluxes and topside density enhancements, (2) the role of E layer conductivity in the maintenance of density irregularity, and (3) the relationship between electric fields and density irregularities.

Electron precipitation and topside densities

Dyson and Winningham [1974] first noted a coincidence of severe topside irregularities and the onset of 300 eV electron precipitation near the dayside cleft. *Kelley et al.* [1982] suggested that topside density variations with scale sizes of tens of kilometers are initiated by similarly structured soft electron precipitation, then are distributed throughout the high latitude region by convection. Electrons with energies <1 keV tend to create ionization pairs at F rather than E layer altitudes. In the F layer ionization loss rates due to recombination are relatively slow. Thus cold plasma has a long lifetime to ambipolar diffuse to topside altitudes. This hypothesis is consistent with recently reported satellite measurements using rather different techniques [*Basu et al.*, 1983, 1984; *Rodriguez and Szuszczewicz*, 1984]. It is also consistent with the general features of the topside density and auroral electron flux distributions shown in the top and bottom panels of Figure 2, respectively. The equatorward density gradient between 75° and 77° spatially corresponds to a similar gradient in precipitating electron flux. The thermal electron density at 73540 s UT (Figure 7) increased by a full order of magnitude as the electron flux increased by a factor of 5 (Figure 2). The relatively low density-maximum of 8×10^2 cm^{-3} at 73440 s UT corresponding to energetic electron fluxes $>10^8$/cm^2 s sr, also confirms the hypothesis of Kelley and co-workers. Observe that at 73440 s UT as S3-2 encountered central plasma sheet (>1 keV) precipitation the electron density maximum was 8×10^2 cm^{-3}, whereas comparable fluxes of boundary plasma sheet electrons (<1 keV) produced an ionization maximum of 3×10^4 cm^{-3} at 73600 s UT.

'ole of E region conductivity

Based on theoretical considerations *Vickrey and 'elley* [1982] have postulated that the lifetimes of *F* :gion irregularities are strongly influenced by the onductivity of the ionospheric *E* region. In the pres-nce of density gradients particles tends to diffuse cross magnetic field lines to weaken the gradients. he rate of diffusion is given by $\rho_\alpha \nu_\alpha$ where ρ_α is the yroradius of the α species and ν_α is the collision rate f that species. O^+ ions have gyroradii of ~5 m and iffuse at greater rates than electrons with gyroradii ~3 cm. In the absence of any *E* region conduc-vity charge separations develop and ion diffusion is :tarded by ambipolar electric fields. Diffusion pro-eds at slow electron rates and density gradients ssociated with irregularities tend to be long lived. If, n the other hand, the *E* region is highly conducting, obile electrons can move along magnetic field lines neutralize the ambipolar fields. Topside diffusion roceeds at relatively fast ion rates, and the lifetimes f irregularities are reduced.

In the top panel of Figure 7 we see irregularities in ie plasma density from scales of 100 km down to ie 2-km resolution of the detector. The very steep radients observed throughout the data stream, e.g., ear 73590 s UT, appear to be consistent with the nall *E* region conductivity model of Vickrey and elley. We note that between 2026:30 and 2027 UT ery little auroral luminosity appears on the S3-2 ajectory in Figure 1, consistent with a low *E* region onductivity.

lectric field and density variations

Relating S3-2 observations of δE and $\delta n/n$ fluctu-ions near the Harang discontinuity to present eory is a perplexing exercise. On the one hand, the uctuations shown in Figures 4 and 6 are associated ith equatorward gradients in the plasma density. heoretical considerations indicate that such regions hould be subject to gradient drift (E × B) instabil-ies. Near the Harang discontinuity the convective lectric field has a strong westward component E_w. he linear growth rate for the gradient drift insta-ility is

$$\gamma = \frac{\nabla n}{n}\frac{E_w}{B} \tag{2}$$

a the poleward region of electric field fluctuations ie density decreased by 2 orders of magnitude, 4.6 -fold, over a latitudinal distance 2°. Thus the gradi-nt scale length $L = (\nabla n/n)^{-1}$ is ~43 km. If we assume that the equatorward component of the plasma convective drift was 1 km/s (E_w = 35 mV/m), then the growth time γ^{-1} = 23 s. A flux tube would thus take approximately 10 growth times to drift across the density gradient. The large amplitudes of the δE fluctuations would then suggest that the insta-bility had evolved into the nonlinear stage.

There are however, two aspects of the data that do not fit simply into the plasma instability picture. These concern the amplitude relationship between δn and δE and the power spectral distributions required by the gradient drift theory.

Amplitude relationships between δn and δE can be obtained from consideration of the continuity and momentum equations, which for thermal electrons in the *F* region can be well approximated by

$$\frac{\partial}{\partial t}\delta n + \delta V \nabla n = 0 \tag{3}$$

and

$$k_B T_e \nabla \delta n + e n_0 \delta E = 0 \tag{4}$$

where $\delta V = \delta E/B$ represents plasma drift fluctu-ations, k_B is the Boltzmann constant and T_e is the electron temperature. Fourier decomposition of equations (3) and (4) gives

$$\delta n/n_0 = (-E_w/BL i\omega)(\delta E/E_w) = (\delta E/E_w) \tag{5}$$

since $i\omega = \gamma = E_w/BL$ (equation (2)) for a purely growing wave. Also, from the electron momentum equation it can be shown that

$$\delta E = ik(k_B T_e/e)(\delta n/n_0) \tag{6}$$

Thus if the spectrum of δE fluctuations could be rep-resented by *N* Gaussian-shaped lines superimposed on a power law spectrum with spectral index *m*,

$$\delta E = \delta E_0\left[(k/k_0)^{-m} + \sum_{i=1}^{N} a_i e^{-[k_i - k]^2/\Delta_i^2}\right]$$

then $\delta n/n$ spectrum would have similar lines superim-posed on a $k^{-(m+1)}$ power law. The symbols a_i, k_i and Δ_i represent the amplitude, wave number and half width of the *i*th spectral line, respectively. Returning to equation (5), we see that for δE fluctuations with amplitudes near 140 mV/m with frequencies between 7 and 10 Hz, then $\delta E/E_w \sim \delta n/n > 1$. As we have stated above it is not possible to compare the S3-2 δn and δE measurements at frequencies greater than 8 Hz. However, other satellites, flying in the same alti-tude/local time sector, as S3-2, with much higher sampling rates have usually reported power law spec-

tra with $\delta n/n$ fluctuations of just a few percent near 10 Hz [*Phelps and Sagalyn*, 1976].

In order to predict the spectral characteristics of waves generated in the gradient drift instability proc-ess, *Keskinen and Ossakow* [1981] have studied the evolution of the E × B instability into the nonlinear regime and found that for scale sizes between 100 km and 0.5 km, the spectra follow power laws with spect-ral indices in the 2–2.5 range. Indeed, the PSD's for $(\delta n/n)^2$ shown in Figure 8 are consistent with the nonlinear, gradient drift theory.

Equation (6) shows that the PSD for the electric field $(\delta E)^2$ should be 2 powers of *k* less than $(\delta n/n)^2$. Thus for $(\delta n/n)^2 \propto k^{-2}$, $(\delta E)^2$ should have a white noise spectrum. We see from Figures 5*a* and 5*b* the electric field power spectral density obeys a white noise power law only in the cases of spectra 3 for both the spinning and axial dipoles over the entire spectral range. However, even here as the number of filter weights is increased to 30 more detailed spectral lines in the 1- to 16-Hz range. As far as we can determine, the presence of strong discrete lines in the δE PSD's is not a prediction of the E × B instability models.

Finally, there are two further observational elements that suggest an alternative to a gradient drift explanation for the large-amplitude electric field fluctu-ations reported here. First, preliminary data from the HILAT satellite at 840 km show frequent occur-rences of large amplitude plasma drift fluctuations coincident with rapidly varying low-energy electron fluxes [*Rich et al.*, 1984]. In a region of field aligned potential drop the PSD's plasma drift and the energy flux of precipitating electrons, in the energy channels closest to the potential drop were identical even in discrete spectral lines [*Basinska and Burke*, 1984]. Second, P. F. Bythrow (personal communication, 1984) has not found large-amplitude drift fluctu-ations of these kinds in data retrieved from a similar drift meter on AE-C at 300- to 400-km altitude. A survey of all S3-2 electric field measurements from January, February, and June 1976 [*Denig and Rich*, 1983] indicated that δE fluctuations similar to those reported here only occur at altitudes greater than 700 km.

The observations from S3-2, HILAT and AE-C can be reconciled most easily if the δE fluctuations do not map to altitudes near the peak of the *F* layer. If this is true, there are two significant consequences: (1) the magnetosphere is the most likely source of the fluctuations, and (2) structured, field-aligned potential drops must occur between 400 and 800 km. These potential drops are of the order of 100 V and would produce structured modulations of field aligned, low-energy electron fluxes. These electrons in turn would structure $\delta n/n$ in the lower *F* layer where plasma processes cause cascading to typical power law spectra. However, the nonlinearly evolved $\delta n/$ processes do map to the altitude of S3/2 and HILAT

Acknowledgments. The authors wish to thank Frederick Rich of AFGL for his assistance in plotting S3-2 data an Michael C. Kelley of Cornell University for his comments on a early draft of this paper. The work was supported in part by A contract F19628-82-K-0011 with Boston College.

REFERENCES

Baker, K. D., R. A. Greenwald, and R. T. Tsunoda, Very hig latitude *F* region irregularities observed by HF radar bac scatter, *Geophys. Res. Lett.*, *9*, 904, 1983.

Basinska, E. M., and W. J. Burke, HILAT observations of plasm irregularities in the cusp, *Eos Trans. AGU*, *65*, 1040, 1984.

Basu, S., S. Basu, E. MacKenzie, S. Basu, H. C. Carlson, D. Hardy, F. J. Rich, and R. C. Livingston, Coordinated measur ments of low energy electron precipitation and scinti lation/TEC in the auroral oval, *Radio Sci.*, *18*, 1151, 1983.

Basu, S., S. Basu, E. MacKenzie, W. R. Coley, W. B. Hanson, an C. S. Lin, *F* region electron density irregularity spectra ne auroral acceleration and shear regions, *J. Geophys. Res.*, *8* 5554, 1984.

Burke, W. J., D. A. Hardy, F. J. Rich, M. C. Kelley, M. Smiddy, Shuman, R. C. Sagalyn, P. J. L. Wildman, and S. T. Lai, He trodynamic structure of the late evening sector of the auror zone, *J. Geophys. Res.*, *85*, 1179, 1980.

Chaturvedi, P. K., and S. L. Ossakow, Nonlinear stabilization the current convective instability in the diffuse aurora, *Geophy Res. Lett.*, *6*, 957, 1979.

Denig, W. F., and F. J. Rich, Measurements of narrow latitudin region of intense antisunward convective flow in the late morn ing sector, *Eos Trans. AGU*, *64*, 801, 1983.

Dyson, P. L., and J. D. Winningham, Topside ionospheric spect *F* and particle precipitation in the dayside magnetospher clefts, *J. Geophys. Res.*, *79*, 5219, 1974.

Fremouw, E. J., C. L. Rino, R. C. Livingston, and M. C. Cousin A persistent subauroral scintillation enhancement observed Alaska, *Geophys. Res. Lett.*, *4*, 539, 1977.

Fougere, P. F., A solution to the problem of spontaneous lin splitting in maximum entropy power spectrum analysis, *J. Geo physics. Res.*, *82*, 1051, 1977.

Harel, M., R. A. Wolf, P. H. Reiff, R. W. Spiro, W. J. Burke, F. Rich, and M. Smiddy, Quantitative simulation of a mag netospheric substorm, I, Model logic and overview, *J. Geophy Res.*, *86*, 2217, 1981.

Iijima, T., and T. A. Potemra, Large-scale characteristics of fiel aligned currents associated with substorms, *J. Geophys. Res.*, *8* 599, 1978.

Kadomtsev, B. B., *Plasma Turbulence*, pp. 11–15, Academic, Ne York, 1965.

Kelley, M. C., K. D. Baker, J. C. Ulwick, C. L. Rino, and M. Baron, Simultaneous rocket probe, scintillation, and incoherent scatter radar observations of irregularities in the auroral zon ionosphere, *Radio Sci.*, *15*, 491, 1980.

elley, M. C., J. F. Vickrey, C. W. Carlson, and R. Torbert, On the origin and spatial extent of high-latitude *F* region irregularities, *J. Geophys. Res.*, *87*, 4469, 1982.
eskinen, M. J., and S. L. Ossakow, Nonlinear evolution of plasma enhancements in the auroral ionosphere, I, Long wavelength irregularities, *J. Geophys. Res.*, *87*, 144, 1982.
eskinen, M. J., S. L. Ossakow, and B. E. McDonald, Nonlinear evolution of diffuse auroral *F* region ionospheric irregularities, *Geophys. Res. Lett.*, *7*, 573, 1980.
ssakow, S. L., and P. K. Chaturvedi, Current convective instability in the diffuse aurora, *Geophys. Res. Lett.*, *6*, 332, 1979.
helps, A. D., and R. C. Sagalyn, Plasma density irregularities in the high-latitude topside ionosphere, *J. Geophys. Res.*, *81*, 515, 1976.
ilbrick, C. R., Recent satellite measurements of upper atmospheric composition, *Space Res.*, *16*, 289, 1976.
sdoski, H. R., P. F. Fougere, and E. J. Zawalick, A comparison of power spectral estimates and applications of maximum entropy method, *J. Geophys. Res.*, *80*, 619, 1975.
ch, F. J., W. J. Burke, M. C. Kelley, and M. Smiddy, Observations of field-aligned currents in association with strong convection electric fields at subauroral latitudes, *J. Geophys. Res.*, *85*, 2335, 1980.
ch, F. J., R. A. Heelis, W. B. Hanson, P. Anderson, B. Holt, L. Harmon, D. R. Zucarro, C. R. Lippincott, D. Girouard, and W. Sullivan, Cold plasma measurements on HILAT, *Johns Hopkins Appl. Phys. Lab. Tech. Dig.*, *5*, 114, 1984.
no, C. L., and S. J. Matthews, On the morphology of auroral zone radio wave scintillation, *J. Geophys. Res.*, *85*, 4139, 1980.
Rino, C. L., R. C. Livingstone, and S. J. Matthew, Evidence for sheet-like auroral ionospheric irregularities, *Geophys. Res. Lett.*, *5*, 1039, 1978.
Rodriguez, R., and E. P. Suzuszczewicz, High-latitude irregularities in the lower *F* region: Intensity and scale size distributions, *J. Geophys. Res.*, *89*, 5575, 1984.
Roble, R. G., and M. H. Rees, Time dependent studies of the aurora: Effects of particle precipitation on the dynamic morphology of ionospheric and atmospheric properties, *Planet. Space Sci.*, *25*, 991, 1977.
Simon, A., Instability of a partially ionized plasma in crossed electric and magnetic fields, *Phys. Fluids*, *6*, 382, 1963.
Ulrych, T. J., and T. N. Bishop, Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys.*, *13*, 183, 1975.
Vickrey, J. F., and M. C. Kelley, The effects of a conducting *E* layer on classical *F* region cross-field plasma diffusion, *J. Geophys. Res.*, *87*, 4461, 1982.
Weber, E. J., and J. Buchau, Polar cap *F* layer auroras, *Geophys. Res. Lett.*, *8*, 125, 1981.
Winningham, J. D., F. Yasuhara, S. I. Akasofu, and W. J. Heikkila, The latitudinal morphology of 10-eV to 10-keV electron fluxes during magnetically quiet and disturbed times in the 2100-0300 MLT sector, *J. Geophys. Res.*, *80*, 3148, 1975.

W. J. Burke and P. F. Fougere, Air Force Geophysics Laboratory, Hanscom Air Force Base, MA 01731.
N. A. Saflekos, Southwest Research Institute, Department of Space Sciences, San Antonio, TX 78284.

A REVIEW OF THE PROBLEM OF SPONTANEOUS LINE SPLITTING IN MAXIMUM ENTROPY POWER SPECTRAL ANALYSIS

Paul F. Fougere

Air Force Geophysics Laboratory, Hanscom AFB, Massachusetts 01731

C. Ray Smith and W. T. Grandy, Jr. (eds.),
Maximum-Entropy and Bayesian Methods in Inverse Problems, 303–315.

1. Introduction

The maximum entropy method of power spectral estimation was first discovered in papers presented in 1967 and 1968 by John Burg. These two seminal papers have revolutionized the subject of power spectrum estimation. Hundreds of papers have been written on the subject of high resolution spectral estimation since that time. Burg's two papers are available in the book Modern Spectrum Analysis edited by Don Childers and published by IEEE Press (1978). Another excellent source of information on the method is Burg's thesis of 1975.

Very briefly the Maximum Entropy Method (MEM) consists in maximizing the entropy rate of a process, subject to constraints that express whatever is known about the process. If the first few lags of the autocorrelation function are known, then an expression for entropy is written in terms of the power spectral density and the expression is then maximized with respect to the unknown autocorrelation lags. The constraint is that the autocorrelation function is the Fourier transform of the power spectral density.

In the so-called Burg technique (the 1968 paper) the power spectrum is determined directly from the data without any need to estimate autocorrelations and without any artificially introduced windows to corrupt the given information. The autocorrelation (or autocovariance) function is never needed, but it can be produced as a simple by-product if desired.

The topic of this paper is a review of an annoying problem that arises as a result of a very subtle constraint involving the Burg technique: the problem of spontaneous spectral line splitting and shifting, and of some methods that have been designed to ameliorate this problem.

The fascination with the Burg technique is due at least in part to its success in handling very short signals. Figure 1a shows a rather extreme case: one-half cycle of a 3-1/3-Hz sinusoid sampled four times at a Nyquist frequency of 10 Hz. The plotter program has simply connected the points with straight lines.

Figure 1b shows the Burg power spectrum obtained using four filter weights—a beautiful sharp line at 3-1/3 Hz. (The number of filter weights is analogous to the number of lags in a Blackman-Tukey spectrum.) This kind of result was first presented by Ulrych (1972).

2. The Splitting Problem

Now, however, if one more point is added, to give the picture shown in Fig. 2a, we have two-thirds of a cycle of the same signal, and its power spectrum, shown in Fig. 2b, splits into two components that roughly straddle the correct frequency. This difficulty was first pointed out by Chen and Stegen (1974), but some of their systematics were wrong and they incorrectly ascribed the unsatisfactory spectrum as due to amplification of noise peaks, caused by the use of too many prediction error filter coefficients. They also claimed incorrectly that this problem was similar to that observed by Jackson (1967) for FFT spectra. Shortly after the work of Chen and

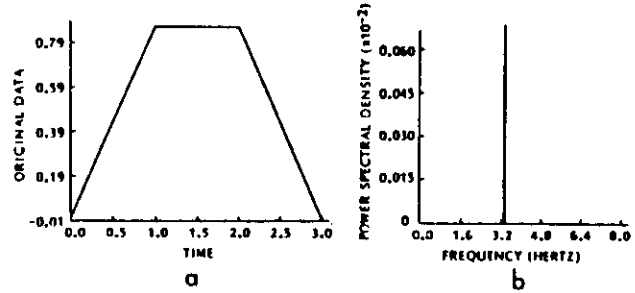


Fig. 1. a. One-half cycle of a 3-1/3-Hz sine wave sampled four times. The four points are simply connected by straight lines. b. The Burg spectrum of the signal.

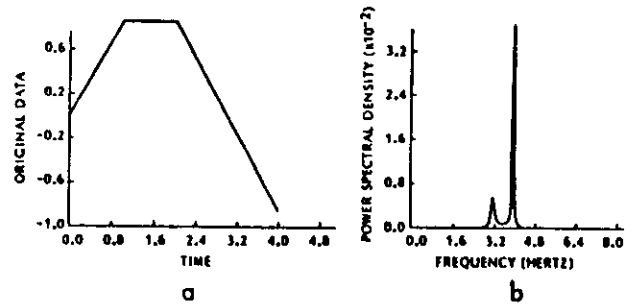


Fig. 2. a. Two-thirds cycle of the same signal as in Fig. 1. b. The Burg spectrum of the signal.

Stegen, Fougere, Zawalick, and Radoski (1976) systematically explored the problem as follows.

Figure 3 shows spectra of the following signals: unit-amplitude, 1-Hz sine waves sampled 21 times at a Nyquist frequency of 10 Hz; 10^{-4} Gaussian white noise added and 20 filter weights. This is one full cycle: the variable is initial phase. There is no splitting—every spectrum is a very sharp accurate needle—and we have used 20 of a possible 21 filter weights. Note that

the classical FFT-based or Blackman-Tukey spectra are never as sharp as this!

Next, in Fig. 4, we use 1.25 cycles with varying initial phase. Every spectrum splits into a sharp doublet unless the initial phase is 0°, 90°, or 180°, that is, a multiple of π/2. But this signal is terribly short: only 1.25 cycles.

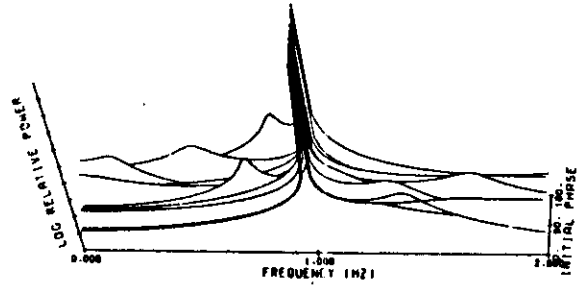


Fig. 3. Nineteen Burg spectra of one cycle of a unit-amplitude, 1-Hz sinusoidal signal, to which Gaussian noise of amplitude 10^{-4} has been added, sampled 21 times. The initial phase varies from 0° to 180° in steps of 20°. Twenty filter weights were used.

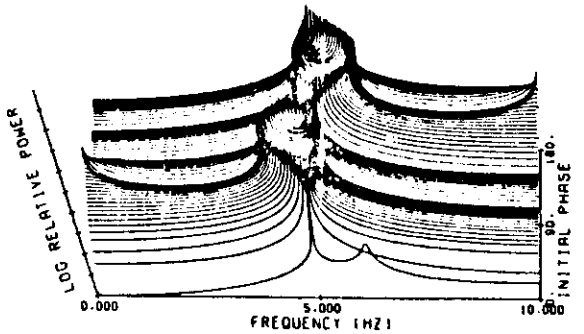


Fig. 4. Ninety-one Burg spectra of 1.25 cycles of a unit-amplitude, 5-Hz sine wave (with 10^{-4} amplitude of additive Gaussian noise), sampled six times using six filter weights. Initial phase varies from 0° to 180° in steps of 2°. Note that there is no splitting at 0°, 90°, and 180° and maximum splitting at 45° and 135°.

The next question arises: What happens when the signal is lengthened? Figure 5a shows the result. The signals are 1.25 Hz, 3.25 Hz, and so on up to 49.25 Hz with a Nyquist frequency of 50 Hz. Every one splits, even with 49.25 cycles of signal, and we have used only 25 of a possible 101 weights—not too high, by anyone's standards. Figure 5b is an enlarged view of the central region.

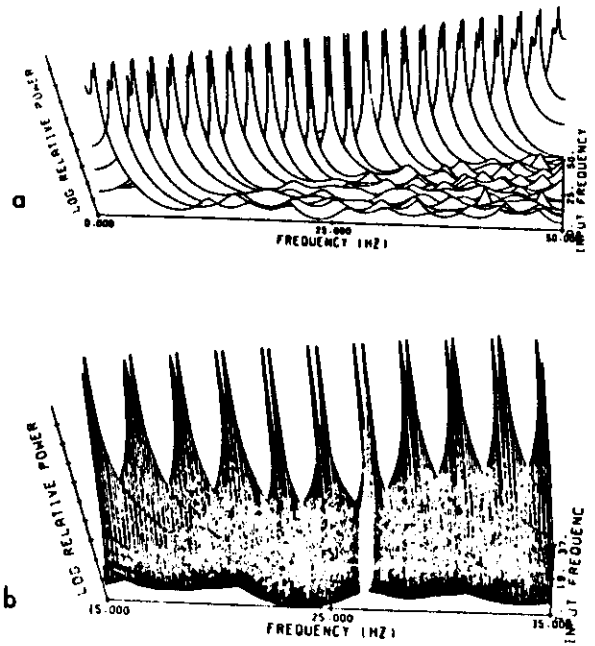


Fig. 5. a. Burg spectra of unit amplitude sinusoidal signals at 45° initial phase, sampled 101 times in 1 second. The frequencies of the signals are 1.25, 3.25, 5.25, ..., 49.25 Hz. Gaussian white noise of amplitude 10^{-4} has been added to each signal. b. Expanded view of the central region.

3. The Solution to the Splitting Problem

Now I will present some elementary mathematics to set the stage for a discussion of various attempts to solve this problem. Linear prediction is at the very heart of the Burg technique. Let the data set be

$$X_1, X_2, X_3, \ldots, X_n ,$$

and define a prediction filter $G_1, G_2, \ldots$ such that a prediction at X_3, $\hat{X}_3$, is given by the simple convolution

$$\hat{X}_3 = G_1X_2 + G_2X_1 . \tag{1}$$

Define the prediction error at X_3 as ϵ_3:

$$\epsilon_3 = X_3 - \hat{X}_3 = X_3 - (G_1X_2 + G_2X_1) . \tag{2}$$

Define prediction error coefficients:

$$g_0 = 1; \quad g_1 = -G_1; \quad g_2 = -G_2 ; \tag{3}$$

then

$$\epsilon_3 = g_0X_3 + g_1X_2 + g_2X_1 \tag{4}$$

and ϵ_3 is called a PEF for Prediction Error in the Forward direction. If the same filter is reversed and run backward, we define δ_1, the prediction error in the backward direction, or PER (Prediction Error Reverse)

$$\delta_1 = g_0X_1 + g_1X_2 + g_2X_3 . \tag{5}$$

We can picture the filters moving over the data to generate prediction errors at all of our observed times in both time directions. But for a three-point filter we begin the forward prediction at X_3; we never run the filter off the data—we use all of the given data and only the given data. There is never any need to make assumptions about what the data are like outside of our given data set. There is never a periodic extension of data as in the Cooley-Tukey FFT method or a zero extension as in the Blackman-Tukey method. Also there is never any tapering—which really ought to be called tampering—of the data with resulting loss of information! The idea of running the filter in both time directions is extremely important. It explicitly takes care of the fundamental symmetry principle that the power spectrum of a stationary process is invariant under a time reversal. Note that we are limiting ourselves to stationary signals. Some authors criticize the Burg technique because it does not work well with exponentially growing or decaying signals. This is nonsense! The Burg technique was not designed for such signals. In fact, all techniques designed to estimate power spectral density require at least approximate stationarity.

Now the best linear prediction error filter minimizes some function of the forward and backward errors. Wiener (1949) chose to minimize the mean square forward error. On the other hand, the Burg norm is given as the arithmetic mean of forward and backward squared error

$$P_m = \frac{1}{2(n-m)} \sum (\overline{PEF^2} + \overline{PER^2}) , \tag{6}$$

where n is the number of observations = sample size and m is the number of g's = filter length. Let the sequence $\phi_0, \phi_1, \phi_2, \ldots$ define the autocorrelation function. Then the Wiener (1949) equation [also called the Yule-Walker equation (Yule, 1927; Walker, 1931)] for real data can be written

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_m \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \phi_{m-1} \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \phi_{m-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \phi_m & \phi_{m-1} & \phi_{m-2} & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ g_{m1} \\ g_{m2} \\ \cdot \\ \cdot \\ \cdot \\ g_{mm} \end{bmatrix} = \begin{bmatrix} P_0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} . \tag{7}$$

The matrix of autocorrelations is a highly symmetric form, called a Toeplitz matrix, with all of the elements on any diagonal equal to each other. This symmetric form greatly simplifies the solution of Eq. (7).

The matrix of prediction error filters, stripped of their leading unity elements, is

$$G = \begin{bmatrix} g_{11} & & & & \\ g_{21} & g_{22} & & & \\ g_{31} & g_{32} & g_{33} & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ g_{m1} & g_{m2} & g_{m3} & \cdots & g_{mm} \end{bmatrix} . \tag{8}$$

The diagonal elements,

$$g_{11} = g_{11} + g_{21}g_{11}$$
$$g_{21} = g_{21} + g_{11}g_{22}$$
$$g_{22} = g_{22} + g_{21}g_{21}$$

called reflection coefficients, are the critically important ones because of the Levinson (1947) recursion given in Eq. (9). In general,

$$g_{nk} = \varepsilon_{n-1,k} + g_{nn}g_{n-1,n-k}. \quad (9)$$

Once we know g_{11} and g_{12} we can find g_{21}. Once we know g_{11}, g_{12}, and g_{13}, we can find g_{31} and g_{32} to complete the third-order prediction error filter. In general, given the set of reflection coefficients, the set of all prediction error filters is easily determined by using this simple and elegant algorithm repeatedly. Given only the first m diagonal elements of Eq. (8), the entire m×m matrix is determined using the Levinson recursion. Conversely, given the prediction error filter, the backward Levinson recursion yields all of the reflection coefficients. That is, given the mth row of Eq. (8), the entire m×m matrix can be found.

Now in the Burg technique the Levinson recursion is used to write the expression for the mean forward and backward error power in terms of the reflection coefficients. If previously determined reflection coefficients are kept fixed, the current one is found by minimizing the error power with respect to the one current reflection coefficient. That procedure guarantees that all of the reflection coefficients are bounded in magnitude by unity and that the resulting filter is a prediction error filter. The corresponding Toeplitz matrix of estimated autocorrelations is nonnegative definite and the basic autocorrelation function theorem is satisfied (Burg, 1975).

Unfortunately, the constraint of fixing previously determined reflection coefficients is not necessary to achieve these ends. It is necessary only that all reflection coefficients at any stage be bounded by unity in magnitude. The independent variables, g_{ii}, $i = 1,\ldots,m$, must be constrained such that $|g_{ii}| < 1$. Such a constraint can be satisfied identically, that is, for any choice of the variables θ_i, $i = 1,\ldots,m$, by setting $g_{ii} = U \sin\theta_i$ with U slightly less than 1; typically $U = 1 - 10^{-7}$. With the independent variables chosen to be the set of real angles θ_i, the minimization problem is now nonlinear. This is the solution to the problem that I discovered for real signals (Fougere, 1977) and for complex signals (Fougere, 1978). Minimizing the mean error power with respect to the set of θ's completely solves the line-splitting and shifting problem. It is true that the objective function P_n is a nonlinear function of the angles θ_i and that the nonlinear minimization is nontrivial, but the 1977 and 1978 papers show how to find the gradient of the objective function, and so standard nonlinear minimization techniques, which require the value of the function and its gradient, can be used to solve the problem quite effectively.

Figure 6 shows Burg spectra of the worst case: a signal consisting of a sine wave of frequency 26.25 Hz with 45° initial phase sampled 101 times at a Nyquist frequency of 50 Hz. The noise levels are shown on the right. Note that the splitting is strongest when the noise level is lowest. As the noise level is raised, the peaks broaden and finally coalesce into a single broad peak located at a shifted frequency. When the nonlinear method just described is applied to the same signals, no splitting occurs, as shown in

Fig. 7. The method is called nonlinear because the independent variables are not the reflection coefficients but arc-sines of the reflection coefficients. The independent variables enter the expression for the mean error power nonlinearly.

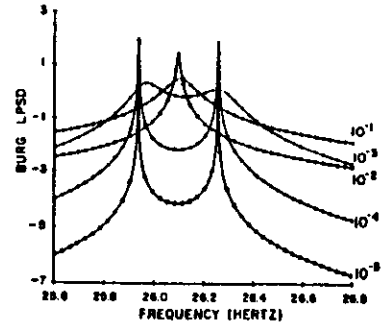


Fig. 6. Burg log power spectral density (LPSD) versus frequency for signals consisting of 101 samples, 0.01 second apart, of a unit-amplitude sine wave of 26.25-Hz frequency and 45° initial phase in additive Gaussian white noise with amplitudes 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. Five filter weights were used.

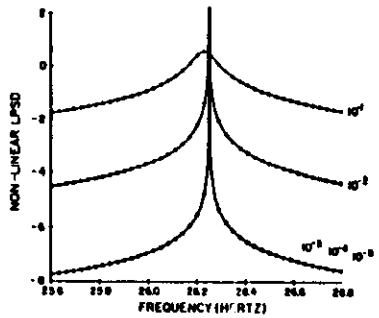


Fig. 7. Nonlinear log spectra of the same signals as in Fig. 6.

One further simple bit of arithmetic shows that if we use m = 2/3n we should be able to achieve a perfect fit to any set of data. If we have m filter weights and n observations, there are 2(n-m) independent prediction errors (in the forward and backward time directions). If m = 2(n-m), then a perfect fit can be achieved because there are as many independent variables as there are independent prediction errors. Solving the equation yields m = 2/3n for a perfect fit. This is not a desirable thing to do—it is merely a test of the method. In practice, a perfect fit can be achieved under either of two conditions. If the constraints are not tight, that is, if no reflection coefficient (g_{ii}) is near unity in magnitude, then it is easy to achieve a perfect fit at m = 2/3n. I have observed a perfect fit when the input data set consists of white noise samples. This condition can be observed on the computer when the mean square prediction error becomes so small as to be computationally indistinguishable from zero. If the computer word length contains k significant decimal figures, then the ratio of mean square prediction error to the mean square observation is less than 10^{-k}.

If the constraints are tight, as they tend to be with sine waves in noise, it is sometimes impossible to achieve a perfect fit. In these cases, when the constraints are dropped so that some reflection coefficient is allowed to become greater than 1 in magnitude, a perfect fit can again be achieved. Of course the resulting filter is nonphysical. These remarks are not intended to suggest that achieving a perfect fit is desirable; it never is, except as a very powerful check on the correct operation of the program.

In summary, the nonlinear method employs the lightest possible constraint to the reflection coefficients—that they all are bounded by unity in magnitude; and it guarantees that the prediction error filters will always be stable minimum-phase filters and the resulting spectrum will always be positive. All of the roots of the Z-transform of the prediction error filter lie outside the unit circle.

4. The Unconstrained Least Squares Method

It is simpler, of course, to remove all constraints entirely and to solve a minimization problem in which the normal equations of least squares are linear in the prediction error coefficients.

The following authors all suggest and recommend such a method: Nuttall (1976), Ulrych and Clayton (1976), Kay and Marple (1979), Swingler (1979), Barrodale and Erickson (1980), Marple (1980), Lang and McClellan (1980). The list of names should be headed by that of John Burg, who first described it as erroneous and rejected it in his 1968 paper. When the unconstrained method actually produces a prediction error filter and the backward Levinson recursion shows that in fact the reflection coefficients are all bounded by unity in magnitude, then the results must be algebraically identical to those of the nonlinear method. But there is absolutely no guarantee that this situation will obtain on any given data set. It is especially unlikely to work correctly in the low noise situations.

Some of these very authors also claim that the cause of line splitting/shifting is the constraint of using the Levinson recursion. This is totally absurd. The Levinson recursion is in no sense a constraint—but it is quite magical! Given any set of numbers whatever, determined in any way imaginable, which are alleged to be prediction error coefficients, we can apply the backward Levinson recursion. If the resulting reflection coefficients are all of magnitude less than unity, then the original numbers are indeed valid prediction error coefficients; otherwise they are not! Conversely, given any set of numbers all of which are less than unity in magnitude, we can use the direct Levinson algorithm to find a valid prediction error filter.

5. Summary

Following is a complete summary of the nonlinear technique, putting together four equations we have just seen:

$$P_m = \frac{1}{2(n-m)} \sum_{k=1}^{n-m} (\epsilon_k^2 + \delta_k^2) , \tag{10}$$

$$\epsilon_k = \sum_{i=0}^{m} X_{k+m-i}\, a_{mi}$$

$$\delta_k = \sum_{i=0}^{m} X_{k+i}\, a_{mi} \tag{11}$$

$$a_{nk} = a_{n-1,k} + a_{nn} a_{n-1,n-k} \tag{12}$$

$$a_{ii} = U \sin \theta_i \; ; \; i = 1,2,\ldots,m, \; U < 1 . \tag{13}$$

The mean error power P_m is given in Eq. (10). This is what I previously called the Burg norm. The prediction errors PEF = ϵ_k and PER = δ_k are given in Eqs. (11). The Levinson recursion is written out in Eq. (12), and the simple necessary and sufficient constraints to produce a stable prediction error filter are given in Eq. (13). We substitute Eq. (13) into Eq. (12); Eq. (12) into Eqs. (11); Eqs. (11) into Eq. (10) and then minimize the mean error power P_n with respect to the true independent variables, the angles θ_j. This produces the optimal correct solution to the problem first posed by Burg, and it completely solves the line-splitting and shifting problem. It is

true that the objective function P_n is a nonlinear function of the angles θ_j and that the nonlinear minimization is nontrivial; but the original papers (Fougere 1977, 1978) show how to find the gradient of the objective function, and so standard nonlinear minimization techniques that require the value of the function and its gradient are used to solve the problem quite effectively.

References

Burg, J. P. (1967) Maximum entropy spectral analysis. Paper presented at 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma.

Burg, J. P. (1968) A new analysis technique for time series data, NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics.

Burg, J. P. (1975) Maximum entropy spectral analysis, Ph.D. thesis, Stanford University, Palo Alto, Calif., 123 pp.

Chen, W. Y., and G. R. Stegen (1974) Experiments with maximum entropy power spectra of sinusoids, J. Geophys. Res. 79, 3019.

Childers, D. G., ed. (1978) Modern Spectrum Analysis (New York: IEEE Press).

Fougere, P. F. (1977) A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis, J. Geophys. Res. 82, 1051.

Fougere, P. F. (1978) A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis of complex signals, Proceedings of the RADC Spectrum Estimation Workshop, May 1978.

Fougere, P. F., E. J. Zawalick, and H. R. Radoski (1976) Spontaneous line splitting in maximum entropy power spectrum analysis, Phys. Earth Planet. Inter. 12, 201.

Jackson, P. L. (1967) Truncations and phase relationships of sinusoids, J. Geophys. Res. 72, 1400.

Kay, S. M., and S. L. Marple, Jr. (1979) Sources of and remedies for spectral line splitting in autoregressive spectrum analysis, p. 151 in Records, 1979 International Conference on Acoustics, Speech, and Signal Processing.

Lang, S. W., and J. H. McClellan (1980) Frequency estimation with maximum entropy spectral estimators, IEEE Trans. Acoust. Speech Signal Process. ASSP-28, 716.

Levinson, N. (1947) The Wiener RMS (root mean square) error criterion in filter design and prediction, J. Math. Phys. (Cambridge, Mass.) 25, 261.

Marple, S. L., Jr. (1980) A new autoregressive spectrum analysis algorithm, IEEE Trans. Acoust. Speech Signal Process. ASSP-28, 441.

Nuttall, A. H. (1976) Spectral analysis of a univariate process with bad data points, via maximum entropy, and linear predictive techniques, Naval Underwater Systems Center, Tech. Rept. 5303, New London, Conn.

Swingler, D. N. (1979) A comparison between Burg's maximum entropy method and a nonrecursive technique for the spectral analysis of deterministic signals, J. Geophys. Res. 84, 679.

Ulrych, T. J. (1972) Maximum entropy power spectrum of truncated sinusoids, J. Geophys. Res. 77, 1396.

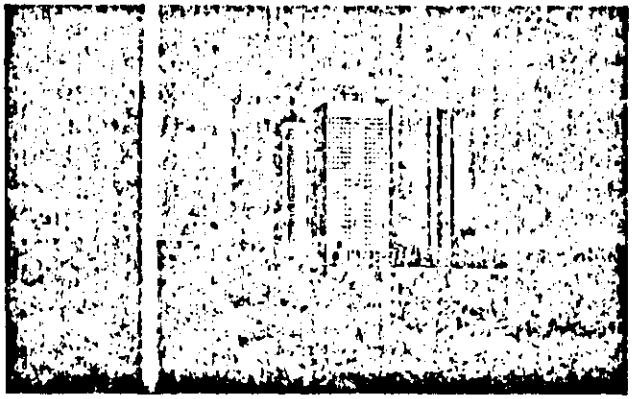
Ulrych, T. J., and R. W. Clayton (1976) Time series modelling and maximum entropy, Phys. Earth Planet. Inter. 12, 188.

Walker, G. (1931) On periodicity in series of related terms, Proc. R. Soc. London, Ser. A, 131, 518.

Wiener, N. (1949) Extrapolation, Interpolation and Smoothing of Stationary Time Series (New York: Wiley and Sons).

Yule, G. U. (1927) On a method of investigating periodicities in distributed series, with special reference to Wolfer's sunspot numbers, Phil. Trans. R. Soc. London, Ser. A, 226, 267.

61



ICASSP 86
PROCEEDINGS

APRIL 7, 8, 9, 10, 11, 1986
KEIO PLAZA INTER-CONTINENTAL HOTEL
TOKYO, JAPAN

Sponsored by

THE IEEE ACOUSTICS, SPEECH, AND SIGNAL PROCESSING SOCIETY,
THE INSTITUTE OF ELECTRONICS AND COMMUNICATION ENGINEERS OF JAPAN,
and
THE ACOUSTICAL SOCIETY OF JAPAN

IEEE-IECEJ-ASJ INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING

Reprinted from PROCEEDINGS OF IEEE-IECEJ-ASJ INTERNATIONAL CONFERENCE
ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, April 7 - 11, 1986

MAXIMUM-ENTROPY SPECTRUM FROM A NON-EXTENDABLE AUTOCORRELATION FUNCTION

PAUL F. FOUGERE

AFGL/LIS
HANSCOM AFB, MA, 01731, USA

INTRODUCTION

In the first "published" reference on maximum entropy spectra, an enormously influential and seminal symposium reprint, Burg [1] announced his new method based upon exactly known, error free autocorrelation samples. Burg showed that if the first n samples were indeed the beginning of a legitimate autocorrelation function (ACF) that the next sample (n+1) was restricted to lie in a very small range. If that (n+1) sample were chosen to be in the center of the allowed range, then the sample number (n+2) would have the greatest freedom to be chosen, again in a small range. In the same paper, Burg also showed that the extrapolation to the center point of the permissable range corresponds to a maximum entropy situation in which the available data were fully utilized, while no unwarranted assumptions were made about unavailable data. In fact unmeasured data were to be as random as possible subject to the constraint that the power spectral density produce ACF values in agreement with the known, exact ACF.

Some time later it was recognized that the realization of exactly known ACF values rarely if ever occurs in practice and that the ACF is usually estimated from a few samples of the time series or from some other experimental arrangement and is, therefore, subject to measurement error. Thus the concept of exact matching of given ACF values was weakened to approximate matching: up to the error variance. Since then there have been several extensions of the Maximum Entropy Method (MEM) to include error prone ACF estimates. Ables [2] suggested an extension but gave no practical method or results. The earliest practical extension was due to Newman [3] who showed that with a slightly generalized definition, one could obtain maximum entropy spectra given noisy ACF estimates.

The Newman method appeared to work well - at least for small problems. I am very much indebted to Will Newman for sending me a copy of his FORTRAN program which I compiled to run on the AFGL CYBER - 750 computer. The problems that Newman had written-up as test cases worked very nicely. However, when I attempted to use the method on larger problems involving noisy ACF measurements, obtained from an interferometer (an interferogram), I had difficulties. For example, if we ignore errors in the ACF simply solve the Yule-Walker equations using simple program of Ulrych and Bishop (1975), order 20 the Yule-Walker equations failed, I produced a reflection coefficient greater than and a resultant negative error power.

Now applying Newman's program, I was able extend the valid range to 30 or so but with difficulty. The program allows an initial adjustment the zero-order correlation (ρ_0) upward until solution is obtained and then tries to lower t value of ρ_0 as much as possible. In order obtain convergence for Newman's iterative procedure, increasingly large ρ_0 values were requi and even so convergence was painfully slow. then decided that it is really improper to v the ρ_0 lag, because the entropy will increase without bound as ρ_0 is increased, and thus there will be no maximum entropy solution.

Other methods have appeared in the literatu notably Schot and McLellan [4]. This paper wa written in such a general fashion, to allow application to multichannel and/or multidimensi data, that I have been unable to utilize it prop ly for a test. Also the program was written "C" programming language (not spoken by the Cyber): an additional roadblock; even if the authors had made the program available it wo have been a great labor to convert it to FORTR

It was then decided to search for some via and simple alternative to the above mentio methods.

THE NEW METHOD

At the very heart of the maximum entro method of BURG is the Fundamental Autocorrelat Theorem. This theorem,announced in the 1967 pa and proved by BURG in his Ph.D thesis [5], can stated "the first n numbers $\rho_0,\rho_1,\ldots,\rho_n$ constitute the beginning of an Autocorrelation Fu tion (ACF) if, and only if, the Toeplitz autoco relation matrix is nonnegative definite. If now write down the modern Yule-Walker equatio as:

130

62

$$\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_n \\ \rho_1 & \rho_0 & & \\ \vdots & & \ddots & \\ \rho_n & & & \rho_0 \end{pmatrix} \begin{pmatrix} 1 \\ g_{n1} \\ \vdots \\ g_{nn} \end{pmatrix} = \begin{pmatrix} P_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1)$$

square matrix is the Toeplitz autocorrelation
ix where the prediction error filter is written
n1 ,gn2 ,··· gnn and the error power at stage n
n+1. The numbers g11,g22,···· gnn are the so-
ed reflection coefficients (due to a geophysical
ogy involving seismic reflection at interfaces).
fundamental ACF theorem is completely equivalent
he statement that the first (n+1) numbers,
$\rho_1,\cdots\rho_n$, are the beginning of an ACF if,
only if, the reflection coefficients, (g_{jj},
) are all of magnitude less than or equal to
; that is, $|g_{jj}| < 1$. This extremely power-
nd simple condition was the basis for the
sseful and optimal non-linear methods of
ire [6,7] which solve the line splitting and
shifting problem associated with the Burg
ique. We will now proceed to derive the new
ique using a method which parallels the non-
r method exactly. For if we have been given
numbers $\rho_0,\rho_1\cdots\rho_n$ and we use the Yule- Walker
ions (see for example the extremely simple
am in Ulrych and Bishop [8]) and find any
, we know that there are errors in some or
f the ρ's. The given numbers simply cannot
e beginning of an ACF! But if we look at our
Walker equations again, we can see that using
ottom row, a very simple recursion for the
an be derived.

$$\rho_k = -\sum_{j=1}^{k} \rho_{k-j}\, g_{kj} \quad ; k=1,2 \ldots j-1 \qquad (2)$$

example:

$$\rho_1 = -\rho_0 g_{11} \quad ; \quad \rho_2 = -(\rho_1 g_{21} + \rho_0 g_{22})$$

Given the prediction error filter (PEF) and
the simple Levinson Recursion:

$$g_{jk} = g_{j-1,k} - g_{jj}\, g_{j-1,j-k} \quad ; \quad \begin{matrix} j=2,3,\ldots,n \\ k=1,2\ldots j-1 \end{matrix} \qquad (3)$$

n obtain the PEF directly from the set of re-
ion coefficients. Thus we now see that as
was we begin with reflection coefficients,
f which lie in the range $-1<g_{jj}<1$ we will
s get an ACF. This condition is trivial to
ce if we simply set $g_{jj} = \text{Using}_j$ where U
constant, very slightly less than 1, and
any real angle. This is the nub of the
mely simple new method. We start off by
ng all $g_{jj} = \text{Using}_j$ and then find the
iven by these numbers. We then minimize the
distance, R, between our new acceptable ACF, ($\hat{\rho}_k$), and the given unacceptable ACF (ρ_k), where:

$$R^2 = \sum_{k=1}^{n} (\hat{\rho}_k - \rho_k)^2 \qquad (4)$$

The result will always be a legitimate ACF and the extension (from n lags to ∞ lags) via the maximum entropy method will always produce a Maximum Entropy Spectrum.

It might now be argued that we have never written down an expression for entropy and maximized it. This is true but the extension of the allowable ACF will always be a "Maximum Entropy Extension". If our original given ACF were badly in error then our new ACF will fit the old ACF, but not very well. Nevertheless, the fit will always be as good as is allowed by the given ACF when the new method is allowed to converge. The starting guess for the iterative solution is obtained by extrapolating the last acceptable ACF of order j up to the full length of the original given ACF, using equation (5) also to be found in Ulrych and Bishop [8]:

$$\hat{\rho}_k = \begin{cases} \rho_k & ; k=1,2..m \\ \displaystyle\sum_{j=1}^{m} \rho_{k-j}\, a_j & ; k=m+1,\ldots n \end{cases} \qquad (5)$$

The method, which has been programmed in FORTRAN, utilizes the IMSL subroutine ZXSSQ, which is a non-linear least squares routine. The required subroutine ZXSSQ can be obtained from IMSL; note that many computer libraries subscribe to IMSL and have copies of ZXSSQ already.

The program is available on 9 track tape on request by seriously interested scientists. Please do not send a blank tape but do include the required type density, either 800 or 1600 BPI and the required code, either EBCDIC or ASCII.

ACKNOWLEDGEMENT

I am very grateful to Bill Neuman for letting me use his program and to Elizabeth Galligan, Mrs. Celeste Gannon, and Mrs. Pauline Beardsley for their excellent typing of this manuscript.

REFERENCES

1. Burg, J.P., (1967), "Maximum Entropy Spectral Analysis", Reprinted in "Modern Spectrum Analysis", (1978), Ed D.C. Childers, IEEE Press, N.Y.

2. Ables, J.G., (1974), "Maximum Entropy Spectral Analysis" Astron. Astrophys. Suppl. Series 15, 383-393 (1972), Proc. Symp. on the Collection and Analyses of Astrophysical Data.

3. Neuman, W.I., (1977) "Extension to the Maximum Entropy Method", IEEE Trans. Inform. Theory, IT-23, 89-93.

4. Schott, J.P. and J.H. McClellan, (1984) "Maximum Entropy Power Spectrum Estimation with Uncertainty in Correlation Measurements", IEEE Trans. ASSP, 32, 410-418.

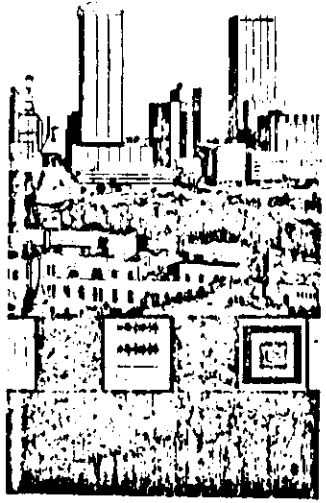
5. Burg, J.P., (1975), "Maximum Entropy Spectral Analysis" Ph.d. Thesis, Stanford University, Palo Alto, CA, 123pp

6. Fougere,P.F., (1977), "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", J. Geophys. Res. 82, 1051,1056.

7. Fougere,P.F., (1978), "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis of Complex Signals", Proc. RADC Spectrum Estimation Workshop, Rome, N.Y.

8. Ulrych,T.J. and T.N. Bishop, (1975), "Maximum Entropy Spectral Analysis and Autoregressive Decomposition", Rev. Geophys. and Space Phys. 13, 183-200.

REPRINTED FROM



ICASSP 85

PROCEEDINGS

MARCH 26-29, 1985
HYATT REGENCY HOTEL
TAMPA, FLORIDA

SPONSORED BY
THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS,
ACOUSTICS, SPEECH AND SIGNAL
PROCESSING SOCIETY

IEEE INTERNATIONAL CONFERENCE ON
ACOUSTICS, SPEECH, AND SIGNAL PROCESSING

SPECTRUM MODEL-ORDER DETERMINATION VIA SIGNIFICANT REFLECTION COEFFICIENTS

Paul F. Fougere
Air Force Geophysics Laboratory
Hanscom AFB, MA 01731-5000
(617) 861-3221

ABSTRACT

A new model-order determination method for maximum entropy or autoregressive power spectra is presented. The method is based upon a simple F-(variance ratio) test for the reflection coefficients. The method is applicable to Burg-MEM, non-linear-MEM and unrestricted least squares methods and their many variants as well as to spectra derived via maximum entropy from autocorrelation estimates or measurements. The method will be tested on simulated signals - both wide-band: power-law processes and narrow band: sinusoids in noise.

INTRODUCTION

For maximum-entropy or autoregressive power spectra, the determination of optimum model order - that is the best number of prediction-error-filter coefficients, is of crucial concern. For if the model order is chosen to be too small, the power spectrum will be too smooth and will omit spectral features which are really present in the data. If the model order is too high - the spectrum will then show spurious detail - exhibiting spectral peaks for which the evidence in the data is weak or non-existent.

At this time, there are about ten distinct procedures designed to aid in the solution of this problem. Some of these procedures work best for broad-band noise signals and underestimate the order required for accurate representation of narrow band processes - typically and surprisingly important - sine waves in noise.

The present paper presents another such procedure; this one is based upon a significance test for the reflection coefficients. Note that even spectral analysis techniques, such as the unrestricted least-squares technique, which do not determine reflection coefficients directly, can use this suggested test as long as they produce a stable prediction error filter, because the inverse Levinson algorithm yields the reflection coefficients. Note also that it does not matter whether we start with time-series data or with a set of autocorrelation estimates. If the starting point is time series data and the Burg-technique [1] or the non-linear technique of Fougere [2] is employed, the reflection coefficients are estimated directly. If the starting point is any estimate or measurement of the ACF, then the Yule-Walker equations which can be solved via the modern Levinson recursion, the reflection coefficients.

FORMULATION OF THE PROBLEM

The Yule-Walker equations [1] may be s
for real time series data and therefore fo
ACF:

$$\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_n \\ \rho_1 & \rho_0 & & \\ & & & \\ \rho_n & & & \rho_0 \end{pmatrix} \begin{pmatrix} 1 \\ g_{n1} \\ \\ g_{nn} \end{pmatrix} = \begin{pmatrix} P_n \\ 0 \\ \\ 0 \end{pmatrix}$$

where $(\rho_0, \rho_1, \ldots, \rho_n)$ is a vector containi
zeroth and first n ACF estimates; $(1, g_{n1}, g_{n2}$
is the n'th order prediction error filter;
the n'th reflection coefficient and P_n i
prediction error variance.

From the modern Levinson recursion as a
by Burg [1], we have

$$P_{j-1} = P_{j-2}\,(1-g^2_{j-1,j-1})$$

This is the sample variance (of the backwar
forward prediction errors) on $2(n-j+1)$ degre
freedom (df).

Similarly

$$P_j = P_{j-1}\,(1-g_{jj}^2)$$

is a sample variance on $2(n-j)$ df.

Now the sum of squares Q_j is equal to a
ance times its df; this is a χ^2 variable.

$$Q_{j-1} = 2(n-j+1)\,P_{j-1}$$

$$Q_j = (2(n-j))\,P_j$$

Since Q_j and Q_{j-1} are each χ^2 variables, t
difference ΔQ_j is also a χ^2 variable on 2

$$\Delta Q_j = Q_{j-1} - Q_j = 2(n-j+1)P_{j-1} - 2(n-j)P_{j-1}(1-g_{jj}^2) \quad (5)$$

where use is made of Eq. 3.

A slight simplification of (5) gives

$$\Delta Q_j = 2 P_{j-1}(1+(n-j)g_{jj}^2) \quad (6)$$

Since ΔQ_j is a χ^2 variable on 2 df, its variance is $\Delta Q_j/2$ and there is a variance ratio

$$VR = \frac{\Delta Q_j/2}{P_j} = \frac{1+(n-j)g_{jj}^2}{1-g_{jj}^2} \quad (7)$$

This is a variance ratio on 2 (numerator) and 2(n-j) (denominator) degrees of freedom. Note the important features of equation (7):

1. The variance ratio increases linearly with n. This could have been expected since, given more input data, smaller spectrum features can be measured with greater confidence.

2. As $g_{jj} \to o$, $VR \to 1$. Thus the numerator and denominator variances are estimating the same quantity, namely the variance of the backward and forward errors, i.e. the prediction error power.

Now if we have 2 independent, unbiassed estimators of the same variance, their ratio, a variance ratio, is distributed according to the F distribution on 2 and 2(n-j)degrees of freedom. Then the Hypothesis, that $g_{jj}=o$ can be tested against the alternative, that $g_{jj} \neq o$ by using a simple F test on 2 (numerator) and 2(n-j) denominator df.

3. As $j \to n$, $VR \to 1/(1-g_{jj}^2)$. Thus for very large model order, the reflection coefficient would have to be very large to be significant. Once again this seems reasonable. It should rarely, if ever be necessary to use as many prediction error coefficients as observations to obtain good spectrum estimates. Indeed as an aside we will show now that if each reflection coefficient really possesses one degree of freedom, then a perfect fit can be attained on any data set if the model order m = 2/3 n. For if m, the model order = number of independent reflection coefficients = 2(n-m) = number of independent prediction errors in both time directions then: m = 2(n-m); m = 2/3 n for a perfect fit. This condition has been observed under the following condition: a) Input data is white noise. Using my own Non-Linear technique [2], I have observed a perfect fit to random-number input. The way that a "perfect fit" is observed is that the mean square prediction error, P_j, suddenly drops by many orders of magnitude when the threshhold 2/3n is passed. If there are g significant figures in a computer word then at 2/3n the error drops to 10^{-g} or smaller. The effect is quite dramatic and quite easy to recognize. This situation will not occur for Burg spectra where each reflection coefficient has less than one degree of freedom owing to the that the prediction error filter goes unstable before this stage is reached. With random-number input, 2/3n may be reached, at which time the situation of "zero" prediction error should be realized.

To give an intuitive feeling for significant reflection coefficients, Table 1 lists a few critical values of the F test [3] for an assumed model order of 10 with 28-, 40-,70-,2 0- and 410-point data samples.

SIGNIFICANCE LEVEL	NUMBER OF OBSERVATIONS 28	40	70	210	410
0.0001	12.06	10.80	9.97	9.43	9.32
	0.61	0.49	0.36	0.20	0.14
0.001	8.43	7.77	7.32	7.03	6.97
	0.53	0.42	0.31	0.17	0.12
0.01	5.25	4.98	4.78	4.66	4.63
	0.43	0.34	0.24	0.13	0.09

Table 1. Upper entry = critical variance ratio, lower entry = magnitude of smallest significant reflection coefficient. Assumed model order (m)=10.

We see, as expected, that with only 28 observations a given reflection coefficient has to be as large as 0.61 to be significant at 0.01% level. This is really 0.01% or 1/10,000. Thus we are saying that a reflection coefficient as large as 0.61 could have arisen in random sampling from a reflection coefficient of zero only once in 10,000 trails. On the other hand, with 410 observations, a reflection coefficient with magnitude as large as 0.09 will be significant at the 0 01 or 1% level.

SIMULATED TESTS OF THE NEW SIGNIFICANCE TEST

In this paper, only verification that the new method is sensible will be given. To compare the new method with all 10-12 existing methods in any kind of fair comparison may well be desirable but is certainly outside of the scope of the present paper.

The first set of tests will be based upon broad-band noise, specifically red-noise with power law spectra $P(f) = Af^{-g}$. In a previous paper [4] I showed that simple BURG-MEM spectra were superior to periodograms for the class of signals called red-noise in which there is appreciably more power at low frequencies than at high frequencies. A convenient special case of a red-noise spectrum is a power-law spectrum; this is a spectrum with only one important parameter, the spectral index, the negative of the spectral slope on a log-log plot. Since we know in advance the "correct spectrum", the method of model order selection can be judged fairly on whether it predicts the correct spectrum.

The second kind of test to which the new method will be subjected will be narrow band: specifically a set of sinusoids in a white noise background. The test signals will be of the form:

$$x_k = \sum_{i=1}^{n} A_i \sin(\omega_i t_k) + N_k$$

where N_k is white noise and ω_i will be chosen to be equally spaced frequencies in the Nyquist band, e.g. one peak would be at $\nu_n/2$; 2 peaks at $\nu_n/3$ and 2 $\nu_n/3$ etc.

REFERENCES

[1] Burg, J.P., "Maximum entropy spectral analysis", Ph.D. thesis, Stanford Univ., Stanford, CA., 1975.

[2] Fougere, P.F., "A solution to the problem of spontaneous line splitting, in maximum entropy power spectrum analysis", J. Geophys. Res., 82, P. 1051-1054, 1977.

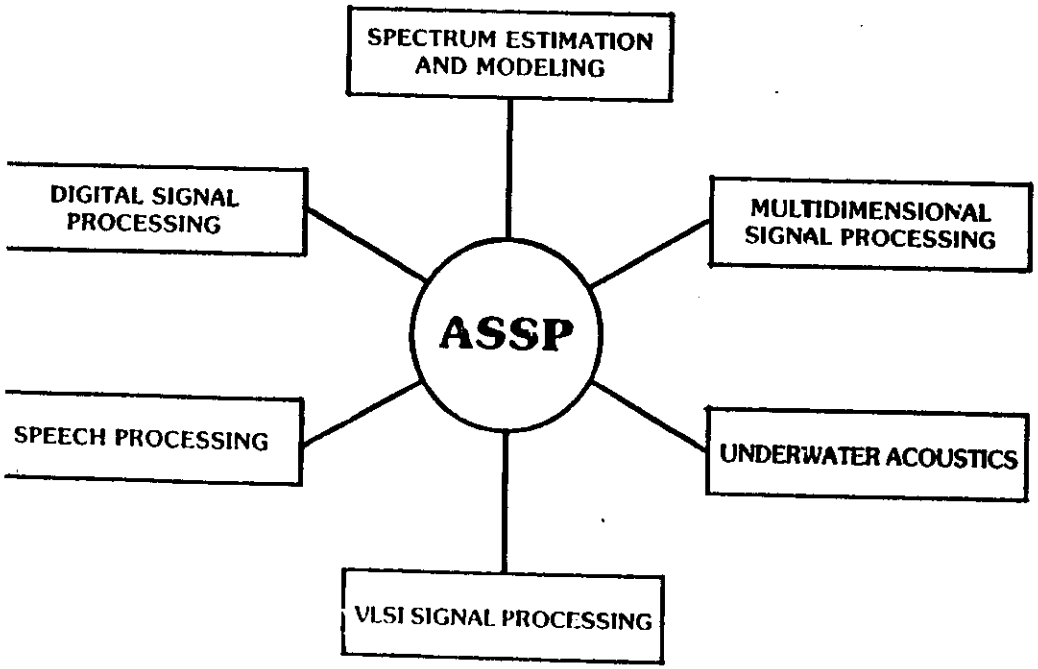
[3] Mardia, K.V. and P. J. Zemroch, "Tables of the F- and related distributions with algorithms" Academic Press, 1978.

[4] Fougere, P.F., "On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods: Simulation studies and application to geophysical data", J. Geophys. Res. (in press).

THE INSTITUTE OF
ELECTRICAL AND
ELECTRONICS
ENGINEERS, INC.

THIRD ASSP WORKSHOP
ON
SPECTRUM ESTIMATION AND MODELING

NOVEMBER 17-18, 1986
NORTHEASTERN UNIVERSITY
BOSTON, MASSACHUSETTS



86TH0167-7

Paul F. Fougere

Air Force Geophysics Laboratory
Hanscom AFB MA 01731

ABSTRACT

This paper reviews applications of the Maximum Entropy Method (MEM) of spectrum estimation to eight distinct problems of interest to the Air Force at the Air Force Geophysics Laboratory (AFGL): (1) geomagnetic pulsations; (2) sunspot data; (3) auroral counts, (3) satellite electric field data; (5) ionospheric phase and amplitude scintillation; (6) neutral air density measured by satellite; (7) "Fourier transform" spectroscopy of atmospheric molecules; and finally (8) Fourier transform mass spectrometry.

INTRODUCTION

Since 1975, AFGL has been applying MEM to Air Force problems [1]. In each case, for each different type of data set, MEM has been compared to the "classical" methods, either Blackman-Tukey [BT] or the Cooley-Tukey periodogram method using the Discrete Fourier Transform [DFT]. In every single case, MEM has been shown to be demonstrably superior to the classical methods. The superiority lies in both increased resolution and increased smoothness in the frequency domain and for slowly varying, nearly stationary signals increased resolution in the time domain. For slowly varying, non-stationary signals a time window can be moved through the data set in such a way that each snapshot portrays a very nearly stationary process. Using MEM, a shorter time signal is required than with classical methods and further MEM requires no external smoothing. Thus spectral features which change with time can be more readily identified using MEM.

Furthermore, the MEM spectrum is invariably much smoother than the periodogram even when many periodograms are averaged to reduce the noisiness of the individual periodogram. Examples of AFGL applications which portray these features will now be presented.

1. GEOMAGNETIC PULSATIONS [2]

Geomagnetic pulsations are low-frequency signals superimposed upon the earth's magnetic field. Their period range is defined somewhat arbitrarily to lie between 0.2 and 600 sec. Their amplitude may be a few hundred gammas to milligammas, while the earth's field is of the order of $5 \cdot 10^4$ γ at the earth's surface ($1\ \gamma = 1nT$).

Radoski et al., [2] applied BT, DFT and MEM to a sample of magnetic pulsations, x and y (horizontal) components observed at Sudbury, MA on 1 September 1970. They found good agreement on the location of spectral peaks among the three methods, but the MEM results showed higher resolution than either BT or DFT and were smoother (less noisy) than DFT. Figure 1 shows this comparison. Figure 2 shows a dynamic MEM spectrum obtained by moving an overlapped time window through the data set. Two times of peak activity are seen at about 0700 and 0730 with spectral energy concentrated near 0.01 Hz a secondary peak near 0.015 Hz. The smoothness the spectra is very evident and is required detailed physical analysis of the pulsation portrayed.

2. SUNSPOT DATA [3]

Sunspot activity has been observed for hun of years and carefully recorded since about 176 Many important geophysical parameters are well correlated with sunspot activity, notably magnetic activity and auroral activity. Of co a very active sun invariably causes severe com cations disruptions.

The Zurich sunspot index, Rz, takes into count the number of sunspot groups as well as number of individual sunspots which appear in light on the visible diec. Much of the power series of sunspot counts appears at a perio eleven years - the well known "sunspot cyc

In order to study this long cycle and even longer cycles, the monthly mean sunspot is used. The range of these monthly numbers essentially zero to about 250.

Figure 3 displays a run of 228 years of mon mean Rz data. The bottom panel shows the orig data, while the upper panel shows band-pass and middle panel shows band-stop data. The filter is a 193 weight, symmetrical FIR filter desi using the method of Behannon and Mems [5].

Figure 4 shows a dynamical MEM spectrum co ing the start years 1757 to 1897. Each spectru based upon 66 years of data. The non-stationa of the sunspot cycle is plainly evident in Figures 3 and 4.

The purpose of the study was to make pre tions for the next sunspot cycle given all avail data. Using the prediction error filter of Burg technique from 670 monthly values beginni January 1923 and using 100, 150 and 200 weig the next cycle was predicted in Figure 5. maximum value was predicted to be 130 ± 20 1980.1 ± 0.2. The actual maximum occurred in number 1979 and reached a smoothed value of Thus, the MEM prediction was rather too low occurred at very nearly the right time.

3. AURORAL SIGHTINGS [6]

The Aurora Borealis has been observed fo far longer time than the sunspots. A remarka time series gleaned from auroras reported in Ch and Europe separately during the years 450 to 1 AD was published by Simcox [7]. This time ser consists of the number of reported auroral sighti during each decade and thus the 100-point t series covers 1000 years.

Figure 6 shows this normalized time ser after the mean and a linear trend were remov Feynman and Fougere [6] performed MEM analysis

this signal using both the Burg procedure [8] and the non-linear technique of Fougere [9]. Figure 7 shows the results. Both spectra show a stable line at 1.13 cycle per century with a period of 88.4 years. Due to the close connection between auroras and other solar-geophysical phenomena, the existence of this stable line shows that the so-called long cycle or Gleissberg cycle in solar-terrestrial relations is real and quasi-periodic and present in 1000 years of auroral data.

4. SATELLITE ELECTRIC FIELD DATA [10]

Saflekos, Burke and Fougere [10] have studied the electric field as measured on the polar orbiting S3-2 satellite on a poleward pass during a substorm. Burg-MEM spectra of electric field fluctuations did not obey expected power laws but had strong spectral lines between 1 and 16 Hz. Simultaneously measured spectra for plasma densities consistently showed f^{-2} power laws over the 0-8 Hz range.

Discrete spectral lines with frequencies near 10 Hz are believed to be due to linearly polarized electrostatic waves. Figure 8 covers approximately one second of data. Panel A shows the very sharp 9.5 Hz spectral peak obtained after filtering with the sharp cutoff band pass filter, (symmetric FIR designed using the McClellan, et. al., program [11]) shown on Panel B. The C Panel shows the data on two simultaneously measured components of the E-field after filtering. The phase coherence between these two components is quite extraordinary.

5. RED-NOISE [12]

A noise process is "red" if it exhibits appreciably more power at low frequencies than at high frequencies. If, in addition, the power spectrum of the process is linear over a certain frequency band on a log-log plot, it is called a "power-law" process and the negative of the slope is called the spectral index. Many naturally occurring red processes are also approximately power-law over an appreciable frequency band.

For these kinds of signals, MEM produces a dramatic improvement over DFT-type spectra. In a systematic series of simulations of power-law processes with indices ranging in steps of 0.5 from 0.5 to 5.0 Fougere [12] showed that the MEM spectrum is always smooth, very nearly linear and possesses an index quite close to the true index. For the same signals, DFT-type spectra required spectral windows and smoothing to even approximate the correct indices. The DFT spectral shape was also excessively noisy so that averaging even 100 periodograms did not produce a spectrum as smooth as one MEM spectrum.

An example of naturally occurring red, power-law like noise is given by ionospheric scintillation which occurs when a radio wave, transmitted by a satellite toward a receiver on the earth, passes through a disturbed ionosphere. Both the phase and the amplitude of the wave suffer low-frequency (~ 0.001 to 100 Hz) perturbations known as phase and amplitude scintillation, respectively. If the wave is detected in a suitable receiver, the scintillation can be separated from the carrier and then can yield important information on the nature of the ionospheric irregularities which are the source of the scintillation.

The data analyzed here were amplitude scintillation data sampled at 30 Hz for 5 min from the MARISAT satellite in January 1981. Figure 9 displays the data set, which was chosen especially because it contains a quiet segment from 0 to about 1 1/4 min, a moderately noisy section from 2 1/4 to 3 1/4 min, and a highly scintillating section from 3 1/4 to 5 min. The changes in character of the noise record are quite abrupt and easy to see in the time record of Figure 9.

It may legitimately be asked whether power spectrum analysis could be employed to monitor the development of such a process. A dynamic spectrum was constructed from the ~ 5-min data sample as follows. There were 10,981 observations in this data set, which was divided into 60 batches consisting of 361 points (10s each) and overlapped by 181 points (5s). Spectra were estimated for each batch.

Figure 10 shows MEM and DFT spectra of four of the batches: numbers 38, 39, 40, 41 which span a time of appreciable change in the time series. Corresponding changes in character of the spectrum which are clearly evident in the MEM spectra are almost totally obscurred in the noisy DFT spectra.

6. NEUTRAL AIR DENSITY [13]

Neutral air density in the 200 km altitude range is of importance in determining satellite drag. It has been measured on board an Air Force satellite in a low altitude, low eccentricity, high inclination orbit using a Satellite Electrostatic Triaxial Accelerometer (SETA). The results have been normalized to 200 km altitude using the MSIS-83 atmospheric density model. Samples are taken every 4 seconds; thus the Nyquist frequency is 0.125 Hz.

Figure 11 shows orbits collected on a day of low magnetic activity: Day 201 of 1983. Band A shows the raw data (with mean removed) and Band B shows the Burg-MEM spectra of order 50. The sharp falloff at about 6×10^{-2} Hz shows the operation of the anti-aliasing filter applied before sampling. The day was so quiet, with respect to variations in air density, that all five spectra, taken from five successive full orbits, are remarkably similar with the spectral index printed above each spectrum varying only very slightly.

7. FOURIER TRANSFORM SPECTROSCOPY [14]

In this study of the infrared spectra of atmospheric molecules, e.g., CO_2 and OH, an experimental arrangement is established in which an "interferogram" is measured. After certain corrections are applied and the interferogram is symmetrized with respect to zero lag, the result is operationally identical to a measured autocorrelation function (ACF).

In order to compare the accuracy of the maximum entropy method (MEM) with the conventional Fourier Transform Method (DFT) computer simulations were run. Note that since we are given an ACF and not a time series we do not use Burg-MEM, but simply MEM. We solve the modern Yule-Walker equations (see Ulrych and Bishop [15] for a simple program) for the reflection coefficients and then use the Levinson Recursion to find the prediction error filter and hence the power spectrum.

In the computer simulations an ACF, R, with N+1 lags, is created by assuming known intensities, A_i, and frequencies, f_i, of n lines in the ν_3 branch of the CO_2 spectrum.

$$R(k) = \sum_{i=1}^{n} A_i \cos(2\pi(k-1)\, dt\, f_i); k=0,1,\ldots N \quad (1)$$

where we used n = 63 lines and N = 250 lags.

The MEM results are given in Figure 12, which shows the MEM spectrum, the given input spectrum and the relative errors in amplitude for the MEM spectral lines. The MEM amplitude was estimated by fitting each spectral peak to a Lorentzian line shape and then integrating analytically under the Lorentzian. The amplitude errors are usually much smaller than 1% of the input amplitude, while the largest error, which occurs at the smallest input amplitude, is still only 4%.

The MEM frequency determination was extraordinarily accurate. The lowest 53 peaks reproduced the input frequencies exactly (to six significant figures, while the largest error in the remaining 10 occurred at line 63: input = 89.2814 cm^{-1}; MEM = 89.2934 cm^{-1}. This peak is the smallest peak with an amplitude only 2×10^{-4} of the maximum amplitude and yet its frequency is recovered to an accuracy of part in 9000.

In contrast, when the same input ACF was apodized using a triangular window and then a Fourier transform was taken, the resulting spectrum was determined much less accurately than the MEM spectrum. The frequency errors are about 100 times larger than those of MEM and the amplitude errors are also much larger than corresponding MEM errors. For the lines with the smallest amplitude, not even one significant figure is correct.

8. FOURIER TRANSFORM MASS SPECTROMETRY [16]

The Fourier Transform Mass Spectrometer (FTMS) was invented in 1978 by Comisarow and Marshall in 1974 (17). The principle employed in the operation of the FTMS is excitation of all ions in a cell (in the presence of a magnetic field B) to cyclotron orbits by a fast rf chirp pulse containing the necessary frequencies. The FTMS measures the image current induced in a pair of electrodes by the oscillating ions.

The image current is a composite of harmonic currents with frequencies that are related to the mass to charge ratios of the ions present in the ion trap. The resulting time domain signal is digitized and analyzed by FFT to estimate a power spectrum which can be converted into a mass spectrum by appropriate calibration.

The resolution of the FTMS is proportional to BT/m where T is the observation time and m is the ionic mass. Long observation times which are necessary for good resolution, cause a decay of the signal with time due to collisions with background gases thus causing degradation of the resolution.

It was not until eleven years later (1975) that Rahbee [16] first applied Burg-MEM to the time signals generated by his FTMS with remarkable results. Figure 13 shows the time signal which results when two isotope of CO_2 with nominal masses of 44 and 45 are used in the cell. Figure 14a shows the FFT spectrum obtained using a 2048 point FFT while 14b shows the corresponding Burg-MEM spectrum of order 120 using only the first 200 observations. The Burg-MEM results achieve a resolution at least ten times as great as FFT while using only 1/10 of the data. Such increased resolution using a much shorter length of data is extremely important for following chemical reactions-experiments now in the planning stage.

REFERENCES

1. H.R. Radoski, P.F. Fougere, E.J. Zawalick, "A comparison of power spectral estimates and applications of the maximum entropy method," J. Geophys. Res., 80, 619, 1975.

2. H.R. Radoski, E.J. Zawalick, P.F. Fougere, "The superiority of maximum entropy power spectrum techniques applied to geomagnetic micropulsations," Phys. Earth and Planet. Int., 12, 208, 1976.

3. P.F. Fougere, "Sunspots: Power spectra and a forecast," in Solar Terrestrial Predictions Proceedings, NOAA, Boulder, CO, Dec 1978.

4. E.J. Chernosky, M.P. Hagan, "The Zurich sunspot number and its variations for 1700-1957," J. Geophys. Res., 63, (4), 1958.

5. K.W. Behannon, N.F. Ness, "The design of numerical filters for geomagnetic data analyses," NASA TN D-3341, 1966.

6. J. Feynman, P.F. Fougere, "Eight-eight year periodicity in solar-terrestrial phenomena confirmed," J. Geophys. Res., 89, 3023, 1984.

7. G.L. Siscoe, "Evidence in the auroral record for secular solar variability," Rev. Geophys. Space Phys., 18, 647, 1980.

8. J.P. Burg, "A new analysis technique for time series data," NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, 1967.

9. P.F. Fougere, "A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis," J. Geophys. Res., 82, 1051, 1977.

10. N.A. Saflekos, W.J. Burke, P.F. Fougere, "Large amplitude electric field fluctuations near the Harang Discontinuity," Radio Sci., 20, 463, 1985.

11. J.H. McClellan, T.W. Parks, L.R. Rabiner, "FIR linear phase filter design program," in Programs for Digital Signal Processing, IEEE Press, New York, 1979.

12. P.F. Fougere, "On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods: Simulation studies and application to geophysical data," J. Geophys. Res., 90, 4355, 1985.

13. F. Marcos, P.F. Fougere, J. Forbes, Unpublished notes.

14. A.S. Zachor, D.S. Smith, "A study of the maximum entropy method of power spectrum estimation as applied to interferometer data," Final Report (Part 2) on Utah State University Sub-contract No. 85-077, 1986.

15. T.J. Ulrych, T.N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," Rev. Geophys. Space Phys., 13, 183, 1975.

16. Rahbee, A., Application of maximum entropy spectral analysis to Fourier Transfrom Mass Spectrometry, Chem. Phys. Lett., 117, 3523, 1985.

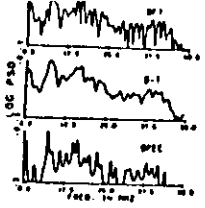


Fig. 1 Pulsation spectra.

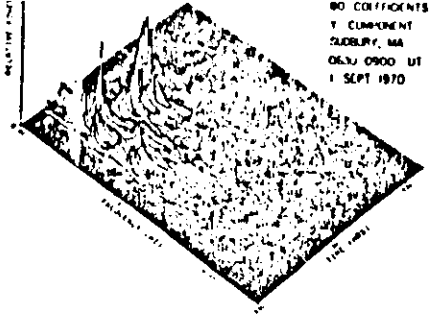


Fig. 2 Dynamic MEM pulsation spectrum.

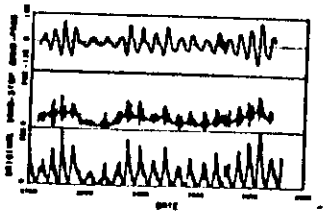


Fig. 3 Sunspot numbers.

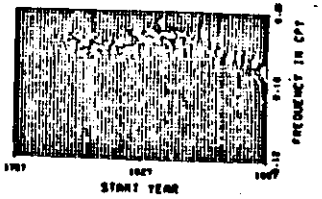


Fig. 4 Dynamic MEM sunspot spectrum.

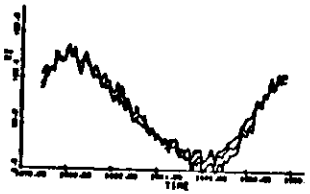


Fig. 5 Sunspot predictions.

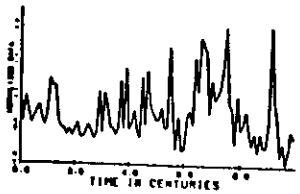


Fig. 6 Auroral observations.

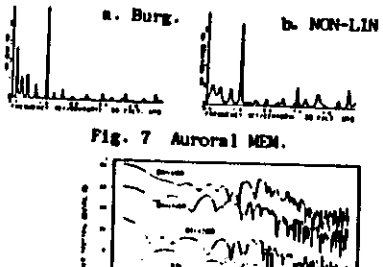


Fig. 7 Auroral MEM.

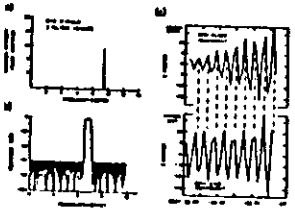


Fig. 8 a. MEM spectrum
b. Filter
c. Filtered data

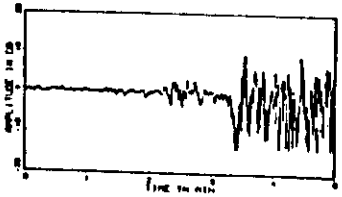


Fig. 9 Scintillation data.

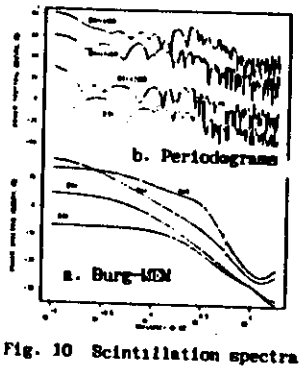


Fig. 10 Scintillation spectra

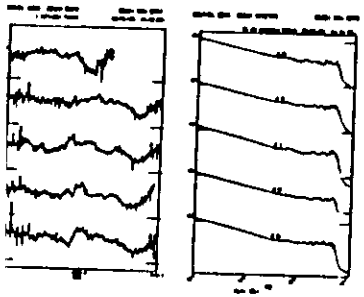


Fig. 11 Neutral density.
a. raw data; b. Burg-MEM spectra.

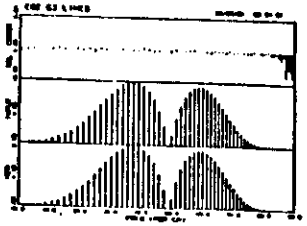


Fig. 12 CO_2 Bottom: MEM
Middle: Input
Top: Amplitude Error.

Fig. 13 CO_2 Isotopes; data.

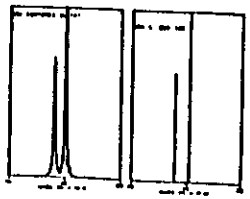


Fig. 14 CO_2 Isotopes;
a. 2K FFT; b. Order 120 MEM.

TO APPEAR IN:

ICASSP 87

PROCEEDINGS

APRIL 6 - 9, 1987
REGISTRY HOTEL
DALLAS, TEXAS

SPONSORED BY
THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS,
ACOUSTICS, SPEECH AND SIGNAL
PROCESSING SOCIETY

IEEE INTERNATIONAL CONFERENCE ON
ACOUSTICS, SPEECH AND SIGNAL PROCESSING

68

ON THE EXTREME ACCURACY OF MAXIMUM ENTROPY SPECTRUM ESTIMATION FROM AN ERROR-FREE AUTOCORRELATION FUNCTION

Paul F. Fougere

AFGL/LIS
Hanscom AFB, MA. 01731-5000

ABSTRACT

The Maximum Entropy Method (MEM) is compared to the periodogram method (DFT) for the estimation of line spectra given an error-free autocorrelation function (ACF). In one computer simulation run, a 250 lag ACF was generated as the sum of 63 cosinusoids with given amplitudes, Ai, and wave numbers, f_i. The wave numbers cover a band from 0 to 89.239 cm^{-1} with a Nyquist of 90 cm^{-1} and are given to an accuracy of six significant figures. The amplitudes range over four orders of magnitude and are given to five significant figures. The resulting normalized ACF was truncated to five significant figures.

The resulting MEM spectrum was remarkably accurate. The lowest 53 spectral peaks reproduced the input frequencies exactly (to six significant figures) while the largest error in the remaining 10 lines occurred at line 63: input = 89.2814; MEM = 89.2934. The relative amplitude of this tiny line is only 2×10^{-4} and yet its frequency was recovered to one part in 9000.

INTRODUCTION

The maximum entropy method (MEM) was introduced in 1967 by John Burg [1]. Given the first few error-free lags of an autocorrelation function (ACF) the basic idea is to find the infinitely long ACF which maximizes the determinant, D, of the ACF matrix. Since the Shannon entropy of a time series is proportional to D, maximizing D is equivalent to maximizing the Shannon entropy. In the same paper, Burg showed that the maximum entropy extension can be obtained simply by solving the modern Yule Walker equations for the reflection coefficients, and using the modern Levinson recursion to find prediction error coefficients g_i. The power spectral density (PSD), P(f) is obtained by evaluating the well known expression:

$$PSD(f) = \frac{P_{n+1}\Delta t}{\left|1+\Sigma g_i \exp(j2\pi f i \Delta t)\right|^2} \qquad (1)$$

The maximum entropy PSD is simply the Fourier Transform of the infinitely extended ACF.

Many studies have illustrated the desirable properties of the MEM spectrum, for example:

a. High frequency resolution [2,3]
b. Smoothness and high time resolution for slowly changing broadband precesses (eg. ionospheric scintillation)[4].

Few studies have addressed the question of accuracy in either frequency estimation or amplitude estimation. Although Lacoss [5] showed that the amplitude of a spectral peak is proportional to the area under the PSD centered at the peak, conventional wisdom says that while MEM frequencies are well determined, the amplitudes are questionable.

SIMULATION STUDY

The present paper attacks the question of accuracy in a rather realistic setting by simulating the infrared spectrum of the ν_2 band of the atmospheric molecule CO_2. The "true" spectrum in this band is quite nicely represented by a "stick spectrum" in which each "line" has a frequency and amplitude but no shape. The spectrum is the sum of a finite number, n, of delta functions; in this case n = 63. Figure 1, middle panel shows the given input spectrum.

The ACF is a sum of cosinusoids whose frequencies and amplitudes are given in Table I. The original wave numbers run from 2300 cm^{-1} to 2390 cm^{-1}, but are "heterodyned" here to the range 0.0 to 90.0 cm^{-1} simply by subtracting 2300 cm^{-1}. The heterodyned wave numbers are given to four decimal places while the ACF using a Nyquist of 90 cm^{-1} is calculated up to lag 250 to five significant figures and is normalized by dividing by the leading term (the zeroth lag). The zeroth lag is then multiplied by 1.00004, effectively adding white gaussian noise of amplitude 0.00004 to the underlying raw data series, because otherwise the Yule Walker equations used to determine the reflection coefficients are singular.

The MEM spectrum is evaluated on a coarse 501 point frequency grid lying betweeen 0 and the Nyquist frequency. This coarse grid is searched for maxima. As soon as 3 points are found such that the middle one has a PSD greater then that of both its neighbors a search is initiated to find the maximum raw PSD and the frequency. The interval containing the peak is repeatedly halved until the maximum and its two neighbors are all within ±1% of the maximum raw PSD. This procedure ordinarily locates the peak frequency to an accuracy of 6 or 7 significant figures. Then a 101 point, fine grid of points in the immediate neighborhood of the peak is fit to a Lorentzian Shape:

$$PSD(f) = A\,(d^2 + (f-fo)^2)^{-1} \qquad (2)$$

where A/d^2 is the PSD at the peak, $f = f_o$. Note that if $f-f_o = d$ then $PSD = A/2d^2$, thus $2d$ is the half width at half maximum. The integral under the Lorentzian from $f_o-\Delta$ to $f_o+\Delta$ is easily found to be:

$$\text{Area} = 2A/d \ \tan^{-1}(\Delta / d) \qquad (3)$$

and as $\Delta \to \infty$, Area $\to \pi A/d$. Thus if d is appreciably less than the separation between adjacent peaks, very little error (typically less than 1%) is incurred by integration between $\pm\infty$. As a check on the integration, Parseval's Theorem is used: the mean input power equals the area under the PSD from zero to the Nyquist.

The resulting spectrum is displayed as a stick spectrum on the lower panel of Figure 1. In the same spectrum, there appeared a few very tiny spurious lines, which were troublesome at first until it was realized that a very simple amplitude test could eliminate every one. The spurious lines are not included in either the figure or the table.

As a further indication that these spurious lines are a "non-problem" a linear least squares analysis was performed to fit a set of cosines having the MEM frequencies to the original 250 lag ACF. All of the spurious frequencies were included in the analysis. All spurious lines have amplitudes 10^{-4} to 10^{-5} times the amplitude of their nearest neighbors, and thus all spurious lines are negligibly small.

To provide a classical comparison with the MEM results, the usual Fourier procedure was applied to the 250 lag ACF after triangular apodization. Instead of using the FFT algorithm, the Fourier transform was calculated from its definition and the same search procedure was used for it as had been used for MEM. A summary of the Fourier results follows:

1. There are 19 spurious lines, resulting from sidelobes. They are smaller by factors of varying only from about 10 to 1000 than their nearest neighbors.

2. Frequency errors range from 0.17E-3 to 0.13 - much larger then corresponding MEM errors

3. Amplitude errors are much larger than corresponding MEM errors. For the smallest lines not even one significant figure is correct.

As a further indication of the remarkable accuracy of the MEM analysis, Figure 1, top panel shows the relative amplitude errors. Most are less than 1% while the largest are less than 4%.

Figure 2 zooms in on the high frequency end of the Spectrum. Here the input and MEM amplitudes vary only bweteen 0 and 0.02 and still the smallest lines are barely discernable. The largest relative amplitude errors are less than 4 and they occur on lines where amplitudes are smaller than the largest line by a factor of abou 6000.

Table I is provided so that other methods ma be tried on the identical 250-lag ACF used here. Can any other method do better than MEM?

REFERENCE

1. Burg, J.P., "Maximum Entropy Spectral Analysis", reprinted in Modern Spectrum Analysis, D.G. Childers, Ed., IEEE Press, (1983).

2. Ulrych, T.J., "Maximum Entropy Power Spectrum of Truncated Sinusoids", J. Geophys. Res. 77, 139 (1972).

3. Radoski, H.R., P.F. Fougere and E.J. Zwalick, "A Comparison of Power Spectral Estimates and Application of the Maximum Entropy Method". J. Geophys Res.80,619,(1975).

4. Fougere, P.F., "On the Accuracy of Spectrum Analysis of Red Noise Processes using Maximum Entropy and Periodogram Methods: Simulation Studies and Application to Geophysical Data". J. Geophys. Res.90, 4355, (1985).

5. Lacoss, R.T., "Data Adaptive Spectral Analysis Methods", Geophysics, 36, 661 (1971).

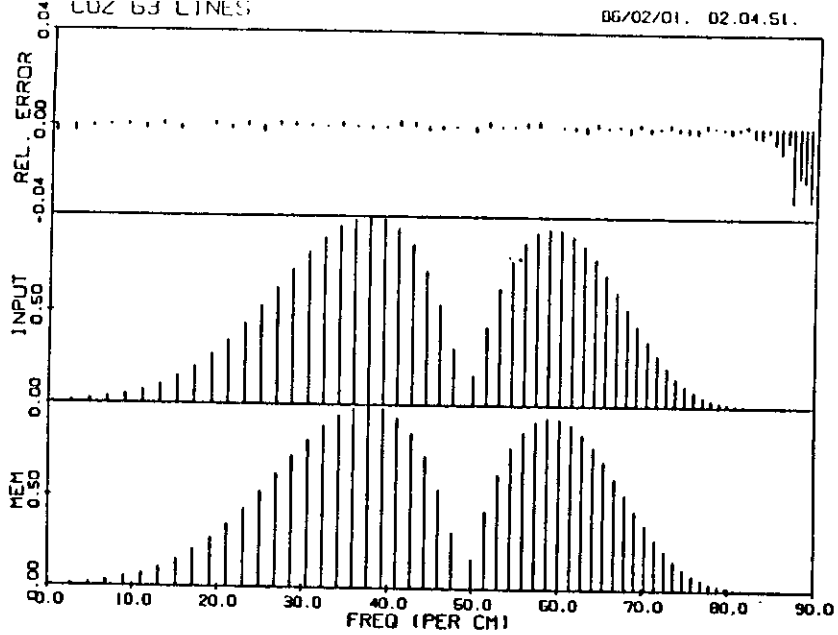


Figure 1. The MEM and input stick spectra. The top panel is the amplitude error divided by the input amplitude.

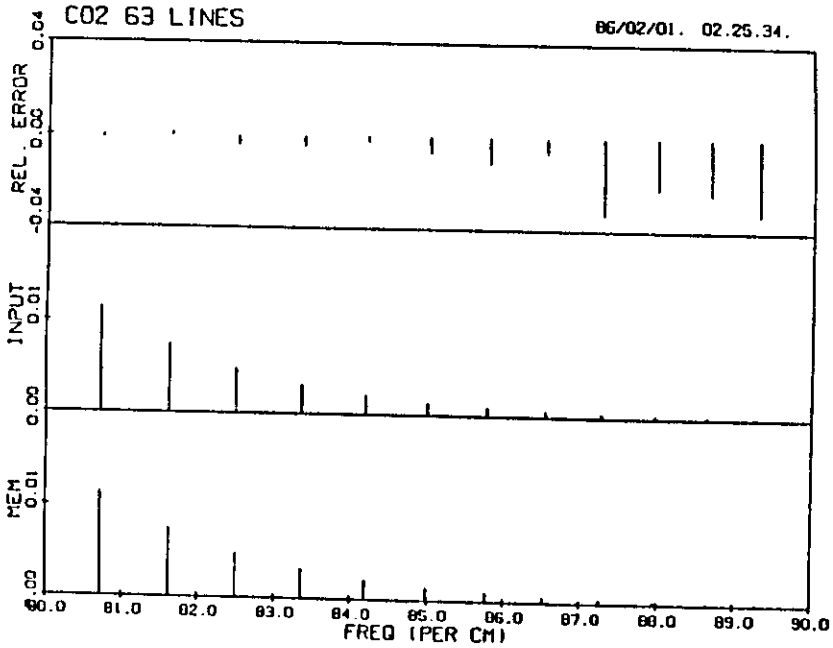


Figure 2. Same as Figure 1 for the high end (80 - 90 cm^{-1}) of the spectrum. The amplitude for MEM and input now run from 0 to 0.02.

Table . Details of the input and MEM spectra.

N	Input Freq.	MEM Freq.	Input Intensity	Area	Total	A	d
1	00.4764	.4764	.12648E-10	.4166E-03	.0004	.329E-10	.711E-(
2	02.6503	2.6503	.19299E-10	.6353E-03	.0011	.127E-10	.108E-(
3	04.8002	4.8003	.29195E-10	.9593E-03	.0020	.631E-10	.592E-(
4	06.9262	6.9262	.43656E-10	.1433E-02	.0034	.228E-10	.144E-(
5	09.0281	9.0281	.64329E-10	.2110E-02	.0056	.232E-10	.989E-(
6	11.1060	11.1060	.93209E-10	.3064E-02	.0086	.415E-10	.122E-(
7	13.1599	13.1599	.13245E-09	.4341E-02	.0130	.394E-10	.816E-(
8	15.1897	15.1897	.18424E-09	.6060E-02	.0190	.172E-10	.255E-(
9	17.1954	17.1954	.25045E-09	.8221E-02	.0272	.480E-10	.525E-(
10	19.1769	19.1769	.33196E-09	.1088E-01	.0381	.229E-10	.189E-(
11	21.1343	21.1343	.42863E-09	.1409E-01	.0522	.314E-10	.200E-(
12	23.0674	23.0674	.53850E-09	.1765E-01	.0699	.376E-10	.192E-(
13	24.9764	24.9764	.65790E-09	.2166E-01	.0915	.460E-10	.191E-(
14	25.8611	26.8611	.78158E-09	.2561E-01	.1171	.513E-10	.180E-(
15	28.7215	28.7215	.90325E-09	.2961E-01	.1467	.564E-10	.171E-(
16	30.5577	30.5577	.10155E-08	.3332E-01	.1801	.408E-10	.110E-(
17	32.3694	32.3694	.11122E-08	.3652E-01	.2166	.481E-10	.118E-(
18	34.1569	34.1569	.11870E-08	.3891E-01	.2555	.233E-10	.539E-(
19	35.9199	35.9199	.12351E-08	.4056E-01	.2960	.425E-11	.944E-(
20	37.6586	37.6586	.12518E-08	.4114E-01	.3372	.460E-12	.101E-(
21	39.3728	39.3728	.12322E-08	.4047E-01	.3777	.804E-12	.179E-(
22	41.0625	41.0625	.11720E-08	.3838E-01	.4160	.223E-11	.522E-(
23	42.7277	42.7277	.10650E-08	.3490E-01	.4509	.459E-11	.118E-(
24	44.3685	44.3685	.90287E-09	.2969E-01	.4806	.440E-10	.133E-(
25	45.9847	45.9847	.67738E-09	.2226E-01	.5029	.121E-10	.490E-(
26	47.5763	47.5763	.37864E-09	.1243E-01	.5153	.918E-11	.664E-(
27	49.9176	49.9176	.19892E-09	.6545E-02	.5219	.263E-10	.361E-(
28	51.4477	51.4477	.53216E-09	.1743E-01	.5393	.305E-10	.157E-(
29	52.9531	52.9531	.78358E-09	.2573E-01	.5650	.395E-10	.138E-(
30	54.4338	54.4338	.96281E-09	.3161E-01	.5966	.530E-10	.151E-(
31	55.8899	55.8899	.10817E-08	.3544E-01	.6321	.300E-10	.763E-(
32	57.3212	57.3212	.11498E-08	.3766E-01	.6697	.743E-10	.178E-(
33	58.7277	58.7277	.11757E-08	.3860E-01	.7083	.541E-10	.126E-(
34	60.1095	60.1095	.11662E-08	.3830E-01	.7466	.451E-10	.106E-(
35	61.4664	61.4664	.11266E-08	.3703E-01	.7837	.664E-10	.161E-(
36	62.7985	62.7985	.10620E-08	.3495E-01	.8186	.404E-10	.104E-(
37	64.1058	64.1058	.97711E-09	.3202E-01	.8506	.725E-10	.204E-(
38	65.3882	65.3882	.87710E-09	.2881E-01	.8795	.393E-10	.123E-(
39	66.6457	66.6457	.76722E-09	.2517E-01	.9046	.769E-10	.275E-(
40	67.8782	67.8782	.65364E-09	.2150E-01	.9261	.493E-10	.206E-(
41	69.0858	69.0858	.54178E-09	.1776E-01	.9439	.404E-10	.205E-(
42	70.2684	70.2684	.43693E-09	.1437E-01	.9583	.425E-10	.266E-(
43	71.4261	71.4261	.34268E-09	.1125E-01	.9695	.567E-10	.454E-(
44	72.5586	72.5586	.26179E-09	.8583E-02	.9781	.426E-10	.447E-(
45	73.6662	73.6662	.19480E-09	.6404E-02	.9845	.654E-10	.920E-(
46	74.7487	74.7487	.14160E-09	.4656E-02	.9892	.376E-10	.727E-(
47	75.8061	75.8061	.10064E-09	.3311E-02	.9925	.174E-10	.472E-(
48	76.8383	76.8383	.70104E-10	.2298E-02	.9948	.317E-10	.124E-(
49	77.8455	77.8455	.47978E-10	.1574E-02	.9963	.378E-10	.216E-(
50	78.8274	78.8274	.32331E-10	.1062E-02	.9974	.186E-10	.158E-(
51	79.7842	79.7842	.21533E-10	.7083E-03	.9981	.177E-10	.225E-(
52	80.7158	80.7158	.14209E-10	.4669E-03	.9986	.215E-10	.415E-(
53	81.6222	81.6224	.93199E-11	.3056E-03	.9989	.192E-10	.565E-(
54	82.5033	82.5032	.61051E-11	.2012E-03	.9991	.153E-10	.684E-(
55	83.3591	83.3592	.40068E-11	.1321E-03	.9992	.143E-10	.973E-(
56	84.1896	84.1900	.26455E-11	.8700E-04	.9993	.162E-10	.168E-(
57	84.9949	84.9941	.17628E-11	.5824E-04	.9994	.212E-10	.328E-(
58	85.7747	85.7746	.11891E-11	.3946E-04	.9994	.293E-10	.669E-(
59	86.5293	86.5268	.81315E-12	.2685E-04	.9994	.400E-10	.134E-(
60	87.2584	87.2605	.56404E-12	.1912E-04	.9994	.497E-10	.234E-(
61	87.9622	87.9656	.39674E-12	.1330E-04	.9995	.516E-10	.349E-(
62	88.6405	88.6573	.28267E-12	.9576E-05	.9995	.375E-10	.352E-(
63	89.2934	89.3026	.20357E-12	.6899E-05	.9995	.124E-10	.162E-(