DETECTION OF WEAK SEISMIC ARRIVALS IN THE PRESENCE OF MICROSEISM NOISE

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN GEOLOGY AND GEOPHYSICS

AUGUST 1973

By

Gerard J. Fryer

Thesis Committee:

George H. Sutton, Chairman
Will Gersch
Antares Parvulescu
William M. Adams
Rockne H. Johnson

We certify that we have read this thesis and that in our opinion it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Geology and Geophysics.

THESIS COMMITTEE

Chairman

ABSTRACT

Deconvolution of microseism noise from seismic time series can be used to improve the detection of arrivals with very low signal-to-noise ratio. If deconvolution is carried out by inverse filtering (i.e. one step ahead prediction-error filtering) the problem of determining the best deconvolution operator length may be resolved by monitoring the final prediction error statistic of Akaike. This statistic is an estimate of the prediction error variance and is a minimum for the optimum length operator. The operator so obtained is well suited for estimation of the microseism spectra.

This technique has been used successfully for analysis of simultaneous three-component seismometer and hydrophone data for the detection of refracted arrivals at the Columbia ocean bottom seismometer from explosions up to 1,350 km distant, and for the estimation of microseism spectra at the time of each shot.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

INTRODUCTION

The detection of signals in noise is a recurrent problem in seismology and is particularly difficult when the frequency content of the signal is unknown. Enhancement of the signal-to-noise ratio of digital time series may be achieved by prediction-error or inverse filtering, provided the noise is stationary. If the optimum least-squares predictor for the noise is convolved with a time series containing both noise and signals, the error in prediction will be an uncorrelated (white noise) sequence except when a signal is present. When a signal is encountered errors will tend to be large since the signal cannot be predicted from the noise. The superiority of this approach over simple band-pass filtering has been illustrated by Claerbout (1964).

One of the greatest difficulties in prediction-error filtering is determining the optimum operator length for a given finite length of data. Obviously the prediction-error operator must be sufficiently long to reflect all predictable qualities of the noise process. However, since operator coefficients must be estimated from the finite length of data, there will be large uncertainties in the estimates of coefficients for a very long operator. Usually the operator length is increased until the mean square error in prediction (the prediction error variance) falls below some tolerable level. This is equivalent to extending the operator until a reasonable fraction of the energy of the process has been explained (Galbraith, 1971). Such an approach implies that the longer the operator the better its performance, an approach which ignores the limitations imposed by the finite data length.

In the case of one step ahead prediction-error filtering, the

optimum operator length can be found by monitoring the final prediction error (FPE) statistic of Akaike (1969a). This is an estimate of the prediction-error variance expected when a predictor, calculated from one observation of a process, is applied to an independent observation of the same process. Choosing the operator length for which the FPE is a minimum gives the best mean-square compromise between bias and variance errors in the operator.

The same compromise between bias and variance is required in any spectral estimation procedure. Since the one step ahead prediction-error operator is identical to an autoregressive model for the data, autoregressive spectral estimates may be obtained directly from the operator with minimum FPE (for which the optimization problem has already been solved). In geophysical applications autoregressive spectral estimation has become known as maximum entropy estimation (Lacoss, 1971; Ulrych, 1972), but apparently such estimates have not used any optimization procedure to determine the best model size.

The purpose of this thesis is to demonstrate the application of the minimum FPE criterion to seismology by applying the technique to the problem of signal detection and the estimation of noise spectra. Such an application of the FPE procedure has not been suggested in the literature before.

The signal detection problem was experienced during a seismic refraction survey of the crust and upper mantle of the northwest Pacific carried out in 1969 with the Columbia ocean bottom seismometer (OBS) as reciever. Details of OBS, which is located in 3.9 km of water about 200 km west of San Francisco, has been given by Sutton et al (1965) and

Auld et al (1969). During the survey a long refraction line was run from the instrument to a point 2,400 km to the west, using charges of up to 2,000 lb. Storm conditions caused microseism noise to be some 12 to 20 db higher than expected. Few refracted arrivals could be detected beyond 800 km range, even after a considerable variety of band pass filters had been used to analyse the analogue recordings. This range was insufficient to delineate the anticipated low velocity zone in the upper mantle, one of the principle aims of the study, so the data were digitized to apply more sophisticated filtering techniques. The work described here is the pilot analysis of records from three shots to determine the feasibility of routine deconvolution of microseism noise from any record with poor signal-to-noise ratio.

THEORY

Univariate Time Series

If a seismic trace y_t is second order stationary it may be generated by passing a purely random sequence x_t through a linear system with impulsive response b_t. The process has the moving average representation

$$y_t = \sum_{s=0}^{\infty} b_s x_{t-s} \quad , \quad b_o = 1 \qquad (1).$$

The sequence x_t is white noise with zero mean and variance σ^2,

i.e. $$E(x_t) = 0 \quad ; \quad E(x_t x_s) = \sigma^2 \delta_{ts} \qquad (2)$$

where $E(.)$ denotes the expected value of the quantity in parenthesis.

Reduction of the time series to a white noise sequence is termed deconvolution and may be achieved by use of an inverse or whitening filter. If a whitening filter for microseisms is used for deconvolution of a seismic trace y_t, the deconvolved trace will be white except in the region of an event unrelated to the microseism process.

The Z transform (Robinson, 1967) of (1) is

$$Y(z) = \left(\sum_{s=0}^{\infty} b_s z^s \right) . X(z)$$

where $z = e^{-i\omega}$. If δt is the sampling interval then $\omega = 2\pi f \delta t$, the maximum frequency of representation being $1/2\delta t$.

The transfer function between the time series y_t and x_t is

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{s=0}^{\infty} b_s z^s = B(z) \quad .$$

If b_t is invertible (i.e. if b_t is minimum delay) the inverse process has a transfer function between x_t and y_t of $H^{-1}(z)$ where

$$H^{-1}(z) = \frac{X(z)}{Y(z)} = \sum_{r=0}^{\infty} a_r z^r = A(z) \qquad (3).$$

Hence

$$B(z) = \sum_{s=0}^{\infty} b_s z^s = \frac{1}{\sum_{r=0}^{\infty} a_r z^r} \qquad (4).$$

where a_r is the impulsive response of the inverse filter. Either of the two sums in equation 4 may have a finite number of terms, but not both.

From equation 3,

$$X(z) = \left(\sum_{r=0}^{\infty} a_r z^r\right).Y(z) \quad .$$

The inverse Z transform yields the autoregressive representation

$$x_t = \sum_{r=0}^{\infty} a_r y_{t-r} \quad , \quad a_0 = 1 \qquad (5).$$

The relationship between the moving average and autoregressive parameters, b_t and a_t, is

$$\sum_{r=0}^{t} a_r b_{t-r} = \delta_t \qquad \begin{cases} = 1 \text{ if } t = 0 \\ = 0 \text{ if } t \neq 0 \end{cases}$$

(Robinson, 1967).

An interpretation of the moving average and autoregressive representations is given in figures 1 and 2. In figure 1 the time series y_t is represented as the convolution of a "characteristic wavelet" with white noise. In figure 2 the inverse process is illustrated with y_t being decomposed by the "inverse wavelet" into white noise.

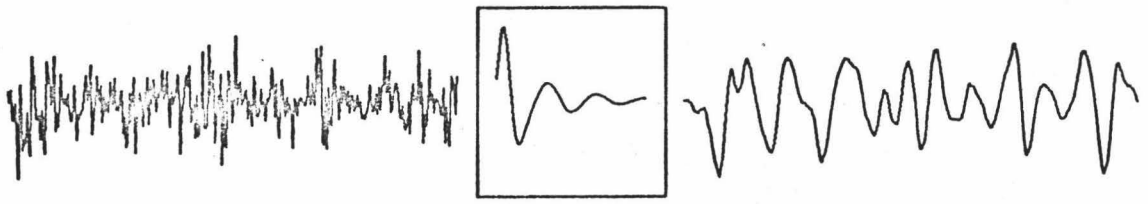

Figure 1. Representation of a moving average process.

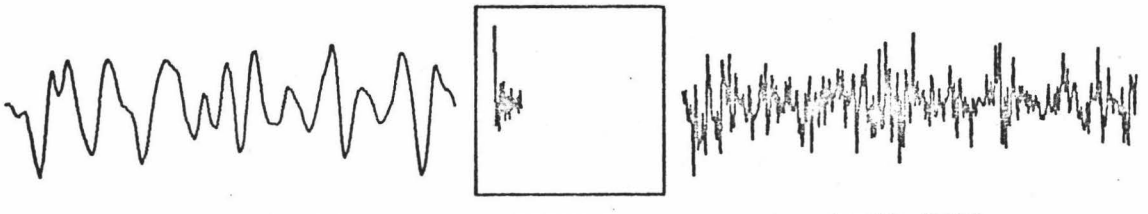

Figure 2. Representation of an autoregressive process.

Equation 5 may be rewritten

$$y_t = -\sum_{r=1}^{\infty} a_r y_{t-r} + x_t$$

For deconvolution we wish to calculate the parameters a_r. For a finite length N of data, linear regression analysis is applied to the finite model

$$y_t = -\sum_{r=1}^{p} a_r y_{t-r} + x_t^{(p)}$$

where $x_t^{(p)}$ is the unpredictable part of y_t when it is assumed that $a_r = 0$ for $r > p$. This gives estimates

$$\hat{y}_t = -\sum_{r=1}^{p} \hat{a}_r y_{t-r}$$

Here $\hat{a}_1, \hat{a}_2, \hat{a}_3, \ldots, \hat{a}_p$ are estimates of $a_1, a_2, a_3, \ldots, a_p$ for which the sum of squares of errors $\sum_{t=p+1}^{N} (y_t - \hat{y}_t)^2$ is a minimum.

The matrix equation to be solved is

$$-\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & & r_0 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_p \end{bmatrix} \qquad (6).$$

(Jones, 1964), where r_k is the empirical covariance for lag k;

$$r_k = E(y_{t+k}y_t) = \frac{1}{N}\sum_{t=1}^{N-k} y_{t+k}y_t$$

(assuming $E(y_t) = 0$). Equation 6 can be rewritten

$$-R\hat{a} = r$$

where R is the covariance matrix, r is the vector of covariances and $\hat{a}$ is a vector of values of $\hat{a}_s$. The solution for $\hat{a}$ is

$$\hat{a} = -R^{-1}r \qquad (7).$$

Note that $\hat{a}$ is calculated from the covariance which contains no phase information. There is an inherent assumption here that the time series is minimum phase in nature (i.e. that a and b are minimum delay) (Robinson, 1967). However, a solution of the type shown in equation 7 may be found regardless of the phase nature of the time series.

With a solution for $\hat{a}$ the deconvolved trace e_t (an approximation of x_t) can be found from

$$e_t = y_t - \hat{y}_t$$

or

$$e_t = \sum_{r=o}^{p} \hat{a}_r y_{t-r} \quad .$$

The most difficult problem in deconvolution is determining the best value for the model length p for a given sample size N. If p is too small some predictable qualities of the time series will have been ignored. If p is too large, the finite sample will result in inaccurate estimates of a and will give inflated prediction-error variance when the model is applied to data outside the sample region. The final prediction error (FPE) statistic of Akaike (1969a) is an index of the performance of the model which can be used to find the best model length.

If the model $\hat{a}$ has been calculated from one observation of the process y_t, the FPE is an estimate of the one step ahead prediction error variance $E(e_t^2)$ of that model when it is applied to an independent observation of y_t. Obviously the optimum model length is the one for which the FPE is a minimum.

The FPE is defined by

$$FPE(p,n) = \sigma^2(1 + \frac{p}{N}) \tag{8}$$

where σ^2 is as given in equation 2. σ^2 is unknown but it can be estimated from

$$\sigma^2 \cong \frac{S_p}{N-p} \tag{9}$$

where S_p is the sum of squares of errors in prediction found in calculating the model. Note that

$$S_p = \sum_{i=p}^{N} e_i^2 = e'e$$

where e is the column vector of values of e_t and e' is the transpose of e (i.e. e' is a row vector).

From equations 8 and 9,

$$FPE(p,N) = \frac{S_p}{N} \left(\frac{N+p}{N-p} \right) .$$

p is the number of parameters fitted; if the mean has been removed from the time series before the analysis, then

$$FPE(p,N) = \frac{S_p}{N} \left(\frac{N+p+1}{N-p-1} \right) \tag{10}.$$

A complete derivation of the FPE scheme has been given by Akaike (1970b).

To apply the FPE scheme in practice, models are fitted recursively for successively larger values of p until a definite minimum in the curve of FPE against p is found. The model with minimum FPE is the optimum (least squares sense) deconvolution operator.

Multivariate Time Series

The generalizations in notation to accomodate multivariate time series are relatively simple. If there are d simultaneous time series the basic equation for an autoregression representation (equation 5) becomes

$$x_t = \sum_{r=o}^{\infty} A_r y_{t-r} \quad , \quad A_o = I \tag{11}$$

where x_t and y_t are d by 1 vectors

$$x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{dt} \end{bmatrix} ; \quad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{dt} \end{bmatrix}$$

and each A_r is a d by d matrix.

The "white" properties of x_t may be represented by

$$E(x_t) = 0 \text{ (zero vector)} ; \quad E(x_r x_s) = \delta_{rs} V \tag{12}$$

where V is a d by d matrix equivalent to σ^2 in the univariate case.

The estimates $\hat{A}$ of A are given by

$$\hat{A} = -R^{-1} r \tag{13}$$

as in the univariate case (equation 7) but each element in the matrices is a d by d sub-matrix.

An efficient recursive scheme for solution of systems of equations such as (13) has been derived independently by Robinson (1963) and Whittle (1963). This technique avoids the inversion of R for each model fitted; models of successively larger orders are fitted with inversion of smaller d by d matrices only.

The final prediction error (equation 8) is modified, for multivariate models to

$$FPE(p,n) = |V| \left(1 + \frac{pd}{N}\right)^d$$

(Akaike, 1971) where $|V|$ is the determinant of V. V is not known but it can be estimated from

$$V = \frac{1}{N-pd} S_p \qquad (14)$$

(Jones, 1964) where S_p is a d by d matrix of residuals (= e'e) equivalent to the univariate S_p of equation 9. The FPE becomes

$$FPE(p,N) = \frac{|S_p|}{(N-pd)^d} \left(1 + \frac{pd}{N}\right)^d$$

$$= \frac{|S_p|}{N^d} \left(\frac{N+pd}{N-pd}\right)^d .$$

If means are removed from the time series before the analysis, this becomes

$$FPE(p,N) = \frac{|S_p|}{N^d}\left[\frac{N + (p+1)d}{N - (p+1)d}\right]^d \qquad (15).$$

Note that if d = 1, equations 15 and 10 are identical, as expected.

The FPE procedure may be used to optimise prediction of one or any subset of the total number of simultaneous time series. If one wishes to predict r of the d time series, the FPE of equation 15 is replaced by the 'FPE of controlled variables', FPEC

$$FPEC(r,p,N) = \frac{S_{r,p}}{N^r} \left[\frac{N + (p+1)d}{N - (p+1)d}\right]^r \qquad (16)$$

(Akaike, 1971) where $S_{r,p}$ is the submatrix of S_p with rows and columns corresponding to the r desired time series.

TECHNIQUE

Analogue recordings of four simultaneous data channels from OBS were digitized at a nominal rate of 50 hertz. The data were from a short period three-component array and a coil hydrophone. These data are referred to in the diagrams as SPZ (vertical), SPH1 (first horizontal), SPH2 (second horizontal) and COIL respectively. For the analogue to digital conversion the data were band-pass filtered with a pass band from 0.83 to 12.5 hz with 6 and 12 db per octave rolloff at the low and high frequency ends respectively. To eliminate unwanted high frequencies the digital data were low-pass filtered using the trapezoidal filter of Gersch (1973). The filter used had a corner at 10 hz and gain was down by 50 db at 12 hz. This allowed a reduction of data by a factor of two (taking every second data point) with a resultant saving in storage and computation time.

All analyses were carried out on an IBM 360/65 computer; FORTRAN subroutines covering all aspects of the theory have been developed and are available on request.

An autoregressive model was fitted to a sample of the microseism noise immediately preceding the expected first arrival for each shot and used for deconvolution of the subsequent data. Typically the microseism sample was thirty seconds long and the model obtained was used for deconvolution of the following eight seconds. Using the model for deconvolution this far ahead of the original data may not seem statistically sound, but there was little improvement in using models derived from longer microseism samples. In no case was the total length of data (sample plus arrival region) more than 45 seconds long. Microseisms

appeared to be reasonably stationary over periods of less than 50 seconds, but samples taken 60 or more seconds apart had spectra that differed significantly.

RESULTS

The results of deconvolution are shown in figures 3 to 5. The figures show the original (low-pass filtered) data and the results of matrix deconvolution over the total data length (sample plus arrival region). Each trace has been normalized relative to its maximum amplitude. Marks on the time axis are 1 second apart. The bottom trace on each figure is a simple binary plot of the deconvolved SPZ data, with amplitude one when the deconvolved SPZ value is outside the 99% confidence range and zero amplitude elswhere.

The processing of shot 181 (Fig. 3) was carried out as a test of the procedure as the arrival was clear on the original data. Shot 181 was a 10 lb charge detonated 48 km from OBS. The signal-to-noise ratio for the first arrivals has been considerably improved by deconvolution, but later arrivals are less well resolved. This is to be expected since the operator is designed for deconvolution of microseism noise rather than for separation of overlapping arrivals. The first arrival on SPZ is a P phase. This does not appear on the original horizontal traces since the incoming disturbance is at near vertical incidence. The appearance of this first arrival on the deconvolved horizontal traces is entirely due to leakage through cross terms in the operator. The first strong arrival on the horizontal traces is S_p, a shear arrival converted from the P phase at the base of the sediment column below OBS (Hussong, 1972).

The original traces of shot 196, a 400 lb charge 335 km from OBS, show no arrival while the deconvolved traces show a clear arrival towards the end of the SPZ, SPH1 and COIL traces (Fig. 4). This arrival

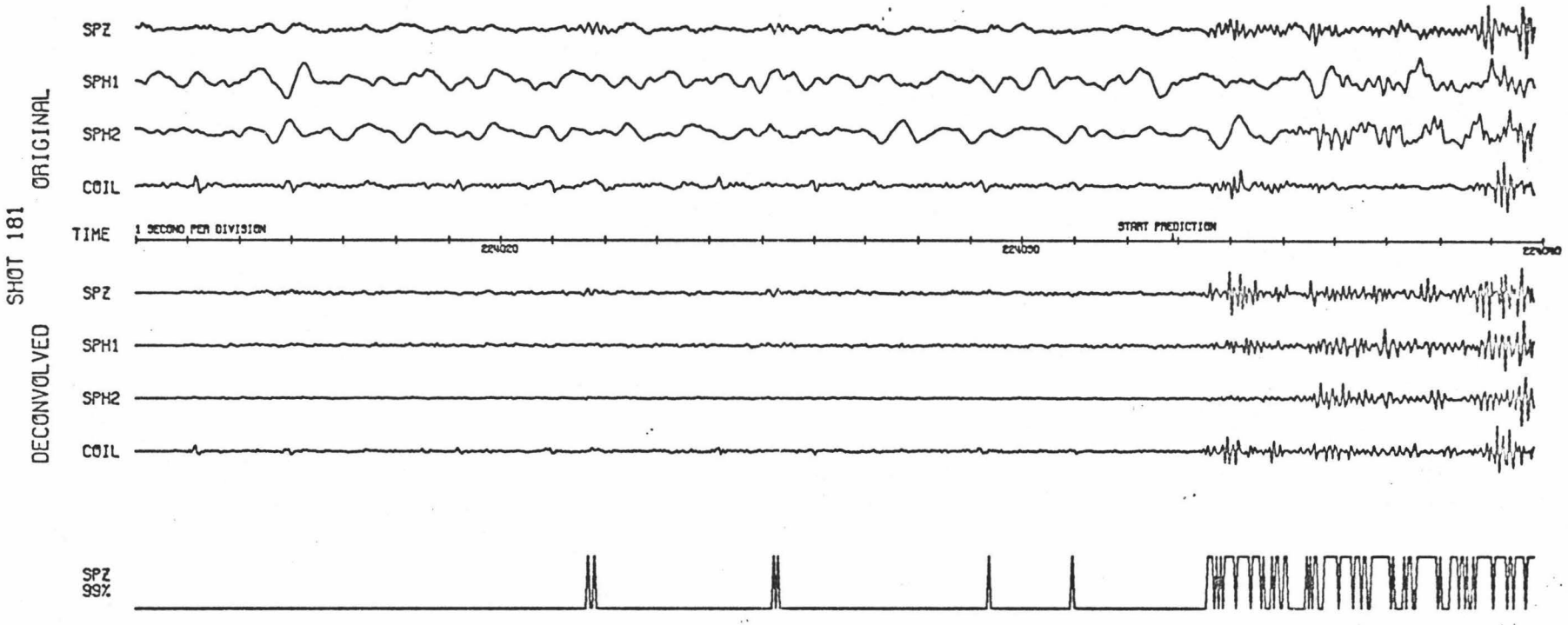

Figure 3. Deconvolution of shot 181 (10 lb charge, 48 km from OBS).

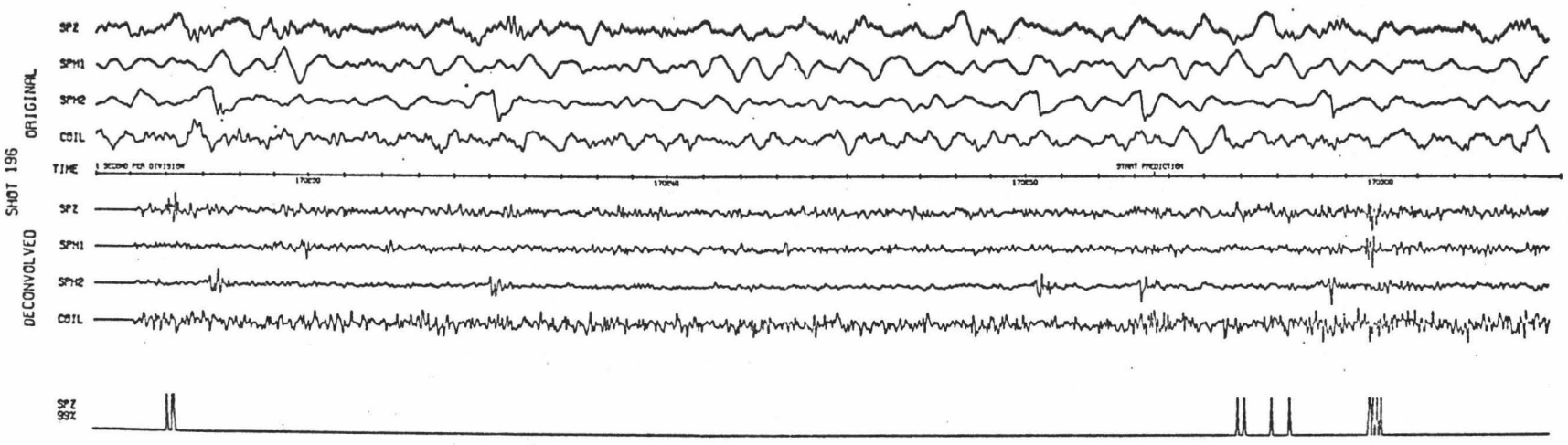

Figure 4. Deconvolution of data for shot 196 (400 lb charge, 335 km from OBS).

appears as a burst of energy on the SPZ 99% trace. Again, appearance of the arrival on SPH1 is a result of leakage through cross terms. SPH2 is almost uncorrelated with the other traces because of apparently random steps in the original data which appear as bursts of energy on the deconvolved SPH2 trace. Because of this poor correlation with other traces, cross terms for SPH2 are small and the arrival does not appear on the deconvolved SPH2 trace. The arrival would probably have been clearer on the other traces if SPH2 had been omitted from the analysis. The four discrete points outside the 99% level for SPZ before the arrival are not significant; the SPZ trace was clipped in this region and several points had to be interpolated. No large errors occur on the other traces at these points.

Shot 207 was a 2,000 lb charge detonated 1,350 km from OBS when microseism noise had become less severe. Two arrivals (arrowed) can be seen on the original SPZ trace (Fig. 5) but they are rather indistinct. Deconvolution sharpens both arrivals and increases the signal-to-noise ratio appreciably, giving better determination of arrival times. The first arrival is seen readily on the deconvolved traces as a change from white noise to a correlated sequence within each trace with associated correlation between the traces. The second arrival is even more obvious because of its large amplitude. Leakage from SPZ and COIL is again responsible for the appearance of the arrivals on the horizontal traces. The apparent arrival times on the horizontal traces are slightly later than on the other two traces; this is probably a result of the predictor reproducing the phase lag between horizontal and vertical motion of the microseisms. Microseisms at OBS generally propagate as Rayleigh or

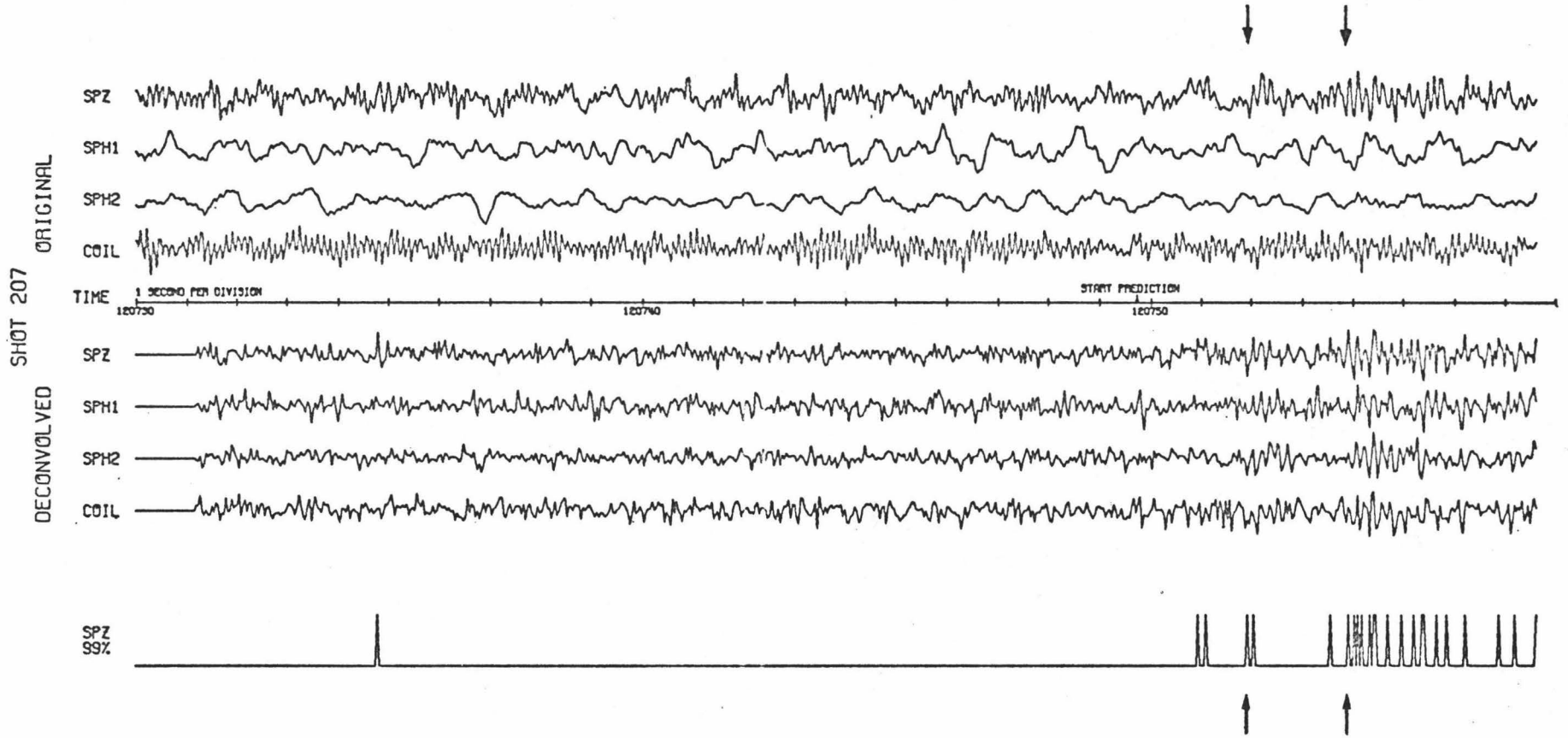

Figure 5. Deconvolution of data for shot 207 (2,000 lb charge, 1,350 km from OBS). Two arrivals are indicated by arrows.

interface waves (Latham and Nowroozi, 1968) so this lag is expected. Why this lag does not appear on the deconvolved trace of shot 196 is not clear. The first double peak on the SPZ 99% trace for shot 207 indicates a possible arrival before the other two, however this is not substantiated by the data from the other channels.

It is important to note that in all the above examples the inclusion of horizontal seismometer and hydrophone data has helped considerably in the detection of arrivals (with the possible exception of SPH2 for shot 196). The simultaneous occurrence of large prediction errors on several traces, even if they are the result of cross-talk, implies that some event very unlike the normal microseism process has occurred. This is much greater justification for picking an arrival than if the data for only one trace were available. All the arrivals suggested here were picked from large prediction error on at least three channels.

For these analyses the autoregressive models were optimized for predictive deconvolution of all channels although the first arrivals were expected to be most apparent on the SPZ traces. Optimizing prediction of SPZ alone at the expense of the other traces (using the FPEC of equation 16) did not improve the detection capability of the system.

As a by-product of the deconvolution procedure, spectral estimates of the microseisms were obtained, following the theory given in the appendix. Power spectra and coherency between each channel were estimated from each model obtained, but only the spectra for the short period vertical traces are given here. Figure 6 shows the microseism spectra for SPZ at the times of shots 181, 196 and 207. The spectra have been

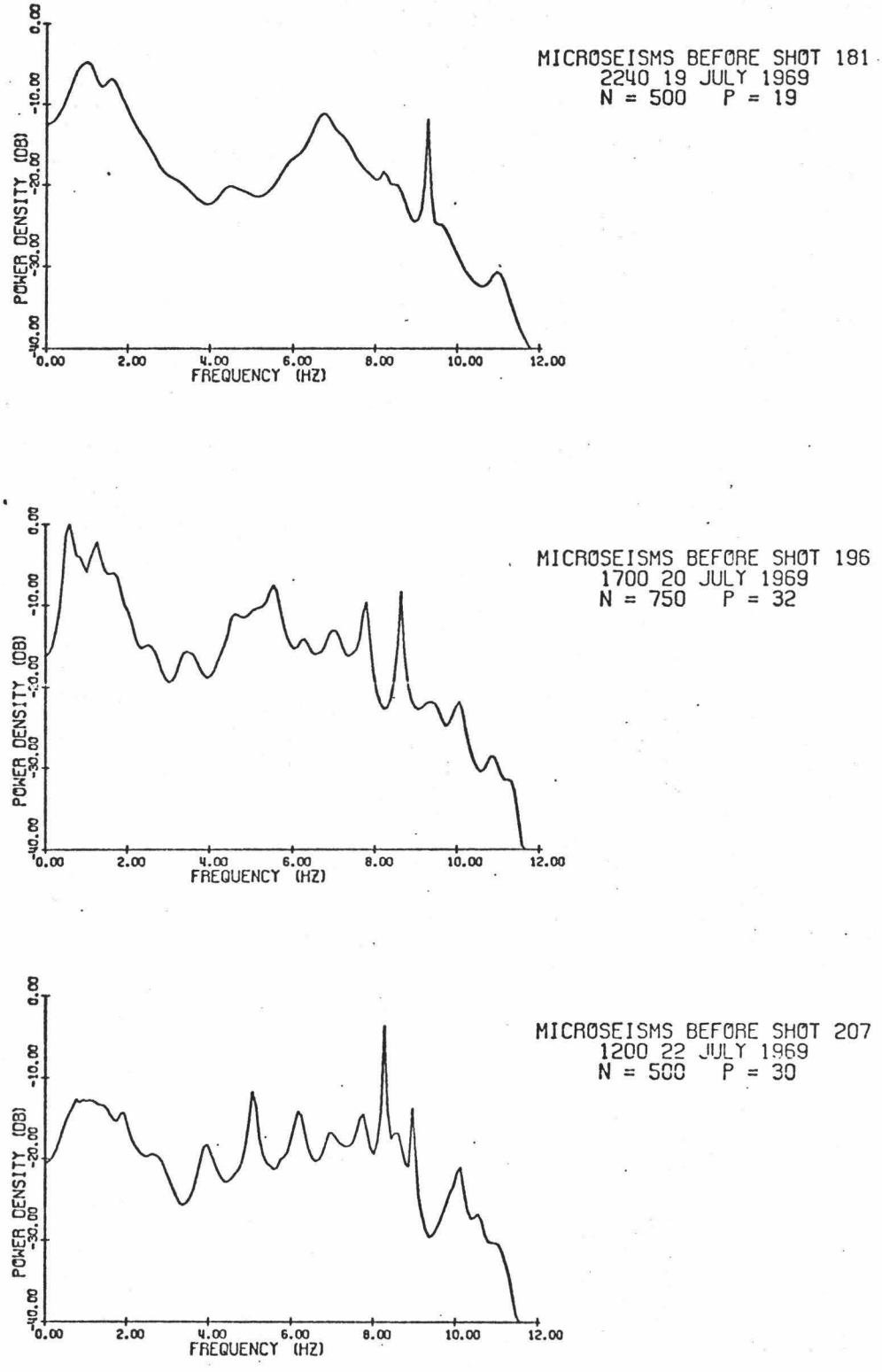

Figure 6. Microseism spectra from short period vertical traces.

normalized with respect to the maximum of 196. The data reflect the passage of an atmospheric storm center over OBS. Shot 181 was detonated when storm conditions were beginning to be hazardous to the shooting ship. The storm was centered over OBS at the time of shot 196 (which shows the highest noise level) and started to die out at shot 207. The very sharp spectral peaks at about 8 hz on all spectra are possibly caused by resonances in the sediment layers (Sutton et al, 1969). The reason for the variation in frequency of these peaks is unclear. It would be worthwhile to follow the variation of the microseism spectra with the path of the storm center, but study of microseisms was not the principle aim of this work.

DISCUSSION AND CONCLUSIONS

The final prediction error scheme provides a much needed simplification to the inverse filtering technique. The procedure can be used to detect arrivals with unknown frequency content which are extremely difficult to isolate by band-pass filtering. The results here suggest that it would be worthwhile to carry out such analyses for all the shots of the OBS long refraction experiment. Results from these analyses will be reported in a later paper (Odegard, 1973).

One aspect of the work that is worthy of more attention is the effect of the cross terms in the autoregressive model. A complete study of the subject of multichannel deconvolution should include single channel deconvolution of all channels as well. In the initial stages of this work it was found that the multichannel approach was significantly superior to single channel deconvolution of the vertical data alone for the detection of arrivals. The single channel deconvolution of the other channels was not investigated.

At present the FPE procedure can only be applied to predictive deconvolution with a unit prediction distance (i.e. one step ahead prediction-error filtering). In many instances a prediction distance of greater than unity yields superior results (Peacock and Treitel, 1969). It would be of great value to extend the definition of FPE so that the optimum operator length could be found for any prediction distance. Work on this problem is proceeding.

ACKNOWLEDGEMENTS

The author wishes to express his thanks to M. E. Odegard for his posing the original signal detection problem and his continued interest in this work, to R. H. Jones for use of his subroutine for solution of the multichannel normal equations, and to H. Akaike for his many helpful comments.

This research was supported by the Office of Naval Research under contract N00014-70-A-0016-0001.

APPENDIX

AUTOREGRESSIVE SPECTRAL ESTIMATION

Traditional spectral analyses (Jenkins and Watts, 1968) have associated with them window functions which are chosen subjectively and are independent of the data; different windows result in different spectral estimates. Spectral estimation from autoregressive (AR) models has a number of advantages over the traditional approach (Gersch, 1970; Lacoss, 1971), the most important of which is that such windowing is not involved. If the FPE procedure is used to choose the optimum AR model this technique becomes completely automatic without the need for any subjectivity.

The energy spectral density of any process y_t is the Fourier transform of its covariance function. From equation 1 the covariance for lag k is

$$r_k = E\,(y_t y_{t+k}) = E\left(\sum_{s=0}^{\infty}\sum_{u=0}^{\infty} b_s x_{t-s} b_u x_{t+k-u}\right)$$

i.e.
$$r_k = \sum_{s=0}^{\infty}\sum_{u=0}^{\infty} b_s b_r \sigma^2 \delta_{k+s-u}$$

or
$$r_k = \sigma^2 \sum_{s=0}^{\infty} b_s b_{s+k} .$$

The Z transform yields the spectral density S

$$S(z) = \sum_{k=-\infty}^{+\infty} r_k z^k = \sigma^2 \sum_{s=0}^{\infty}\sum_{k=-\infty}^{+\infty} b_s b_{s+k}\, z^k .$$

Substituting $u = s+k$,

$$S(z) = \sigma^2 \sum_{s=0}^{\infty} \sum_{u=0}^{\infty} b_s b_u z^{u-s}$$

since $b_u = 0$ for u less than zero.

Hence

$$S(z) = \sigma^2 \sum_{s=0}^{\infty} b_s z^{-s} \sum_{u=0}^{\infty} b_u z^u$$

or $$S(z) = \sigma^2 B(z^{-1}) B(z) \quad .$$

From equation 4, $B(z) = A^{-1}(z)$, so

$$S(z) = \sigma^2 A^{-1}(z^{-1}) A^{-1}(z)$$

or $$S(z) = \frac{\sigma^2}{\sum_{s=0}^{\infty} a_s z^{-s} \sum_{u=0}^{\infty} a_u z^u} \quad \frac{\sigma^2}{\left| \sum_{s=0}^{\infty} a_s z^s \right|^2} \quad .$$

With the autoregressive (AR) model $\hat{a}$ obtained for minimum FPE we use

$$S(z) = \frac{\sigma^2}{\left| \sum_{s=0}^{p} \hat{a}_s z^s \right|^2} \qquad \text{(A-1)}$$

as a spectral estimate where σ^2 is given by equation 9.

The treatment for multivariate spectral estimation follows the above univariate case. The spectrum of a multivariate time series is

$$S(z) = B(z^{-1}) V B(z)' \qquad \text{(A-2)}$$

where V is the variance-covariance matrix of errors in prediction defined in equation 12, and the prime indicates matrix transposition. In terms of the AR parameters

$$S(z) = A^{-1}(z^{-1})\, V\left[A^{-1}(z)\right]' \quad .$$

From the fitted multivariate model with minimum FPE we use

$$S(z) = \hat{A}^{-1}(z^{-1})\, V\left[\hat{A}^{-1}(z)\right]' \qquad \text{(A-3)}$$

as an estimate of the spectrum, where V is given by equation 14. The computation is simplified by the fact that $\left[\hat{A}^{-1}(z)\right]'$ is the complex conjugate transpose of $\hat{A}^{-1}(z^{-1})$, so only one inverse need be calculated at each frequency (Gersch, 1970).

Each value of $S(z)$ is a d by d matrix of auto and cross-spectral estimates. From $S(z)$ it is simple to calculate transfer functions, coherency and phase relationships between the d simultaneous time series (Gersch, 1970).

Spectral estimation from AR models is identical to the maximum entropy technique reported in the geophysical literature (Lacoss, 1971; Ulrych, 1972), however none of the reported maximum entropy estimates use any objective criterion (such as minimum FPE) for choosing the model length p. The importance of the FPE is illustrated in figure 7 where AR spectral estimates are compared with Parzen windowed periodograms. 512 points of a short period vertical record of microseism noise were used with a digitization rate of 30 hz. The FPE curve (Fig. 8) shows a minimum at a model length of 12. The spectral estimates from models of length 6 and 50 show high bias and variance respectively. Such

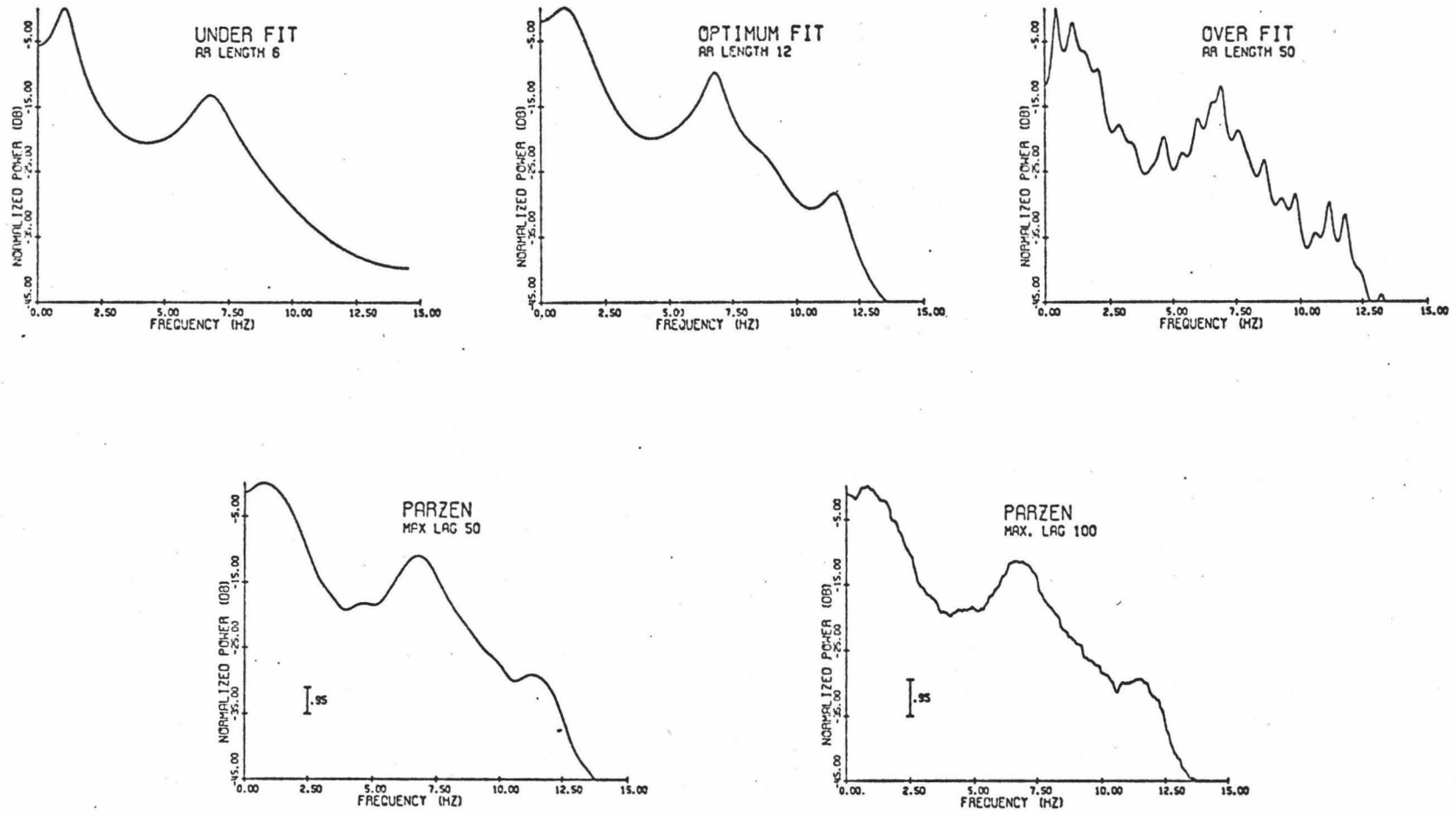

Figure 7. Comparison of autoregressive spectral estimates (above) with Parzen windowed periodograms (below).

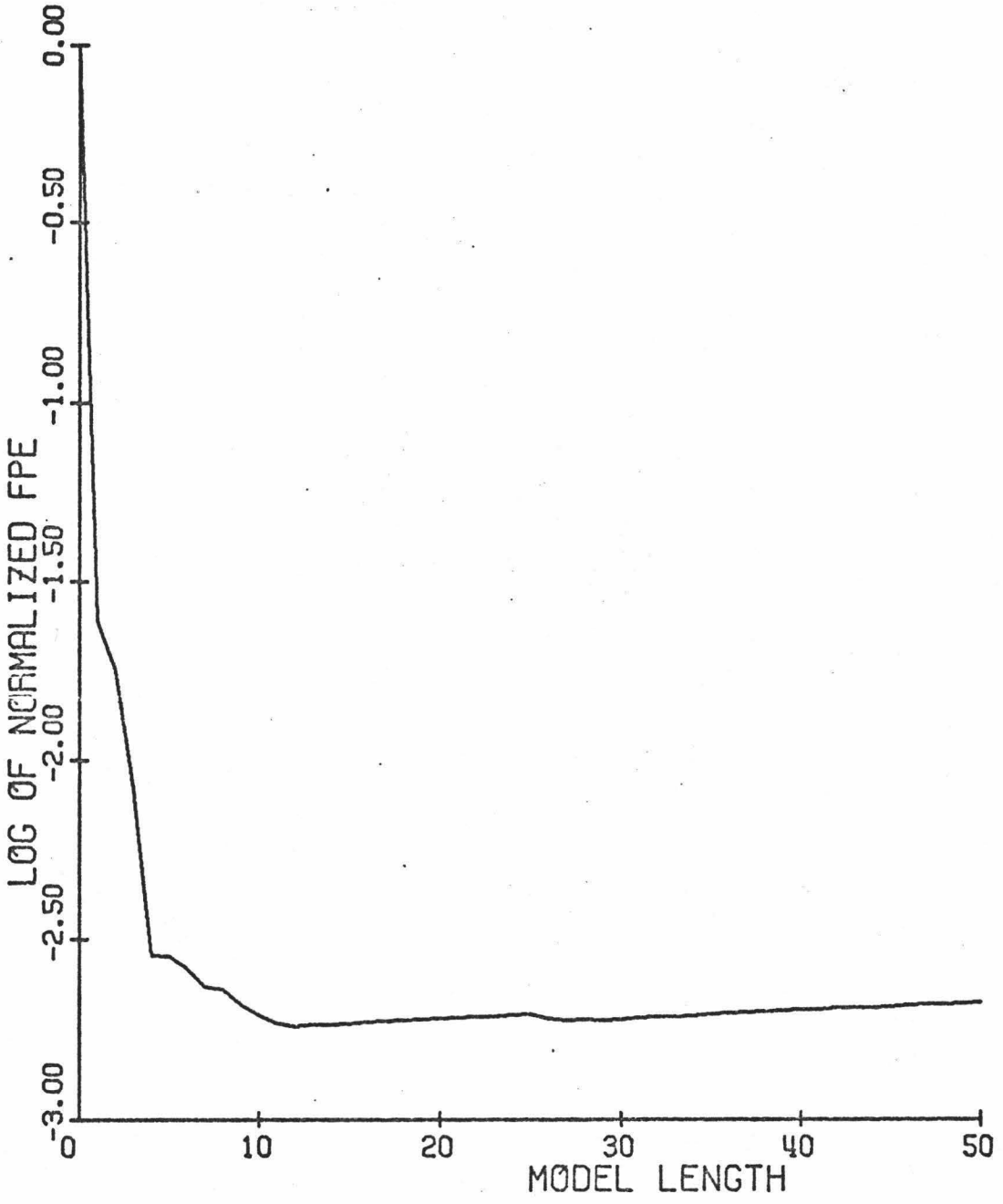

Figure 8. Final prediction error curve for the data in figure 7. FPE values have been normalized with respect to FPE(0). The logarithm of the normalized values is plotted to accentuate the minimum.

variation with the model length p is to be expected. With the inverse waveform, a, limited to a short time duration (p small), its transform A (and hence the spectrum) will vary slowly with frequency. As p is increased the spectrum will vary more rapidly with frequency as shown in figure 7. Note that the optimum AR estimate lies completely within the 95% confidence range of the two Parzen estimates. The optimum AR estimate is probably more reliable than the Parzen estimates since it is not subject to the vagarities of windowing.

A rigorous treatment of AR spectral estimation using the FPE has been given by Akaike (1969b). Further discussion and examples are given by Akaike (1970b) and Gersch and Sharpe (1973).

REFERENCES

Akaike, H., 1969a, Fitting autoregressive models for prediction: Ann. Inst. Stat. Math., v. 21, p. 261-265.

_____, 1969b, Power spectrum estimation through autoregressive model fitting: Ann. Inst. Stat. Math., v. 21, p. 407-419.

_____, 1970a, On a semi-automatic power spectrum estimation procedure: Proc. Third Hawaii International Conference on Systems Science, North Hollywood, Western Periodicals, p. 974-977.

_____, 1970b, Statistical predictor identification: Ann. Inst. Stat. Math., v. 22, p. 203-217.

_____, 1971, Autoregressive model fitting for control: Ann. Inst. Stat. Math., v. 23, p. 163-180.

Auld, B., Latham, G., Nowroozi, A. and Seeber, L., 1969, Seismicity off the coast of northern California determined from ocean bottom seismic measurements: Bull. Seism. Soc. Am., v. 59, p. 2001-2015.

Claerbout, J. F., 1964, Detection of P-waves from weak sources at great distances: Geophysics, v. 29, p. 197-211.

Galbraith, J. N., 1971, Prediction error as a criterion for operator length: Geophysics, v. 36, p. 261-265.

Gersch, W., 1970, Spectral analysis of EEG's by autoregressive decomposition of time series: Math. Bioscience., v. 7, p. 205-222.

_____, 1973, A note on trapezoidal digital filter design: Computers and Biomed. Res., in print.

_____ and Sharpe, D. R., 1973, Estimation of power spectra with finite order autoregressive models: IEEE Trans. Automatic Control, in print.

Hussong, D. M., 1972, Detailed structural interpretations of the Pacific oceanic crust using Asper and ocean-bottom seismometer methods: unpublished doctoral dissertation, University of Hawaii.

Jenkins, G. M. and Watts, D. G., 1968, Spectral analysis and its applications: San Francisco, Holden Day.

Jones, R. H., 1964, Prediction of multivariate time series: Jour. App. Meteorology, v. 3, p. 285-289.

Lacoss, R. T., 1971, Data adaptive spectral analysis methods: Geophysics, v. 36, p. 661-675.

Latham, G. V. and Nowroozi, A. A., 1968, Waves, weather and ocean bottom microseisms: Jour. Geophys. Res., v. 73, p. 3945-3956.

Odegard, M. E., 1973, Structure of the North Pacific upper mantle: doctoral dissertation, University of Hawaii, in preparation.

Peacock, K. L. and Treitel, S., 1969, Predictive deconvolution: theory and practice: Geophysics, v. 34, p. 155-169.

Robinson, E. A., 1963, Mathematical development of discrete filters for the detection of nuclear explosions: Jour. Geophys. Res., v. 68, p. 5557-5567.

_____, 1967, Predictive decomposition of time series with application to seismic exploration: Geophysics, v. 32, p. 418-484.

Sutton, G. H., McDonald, W. G. Prentiss, D. D. and Thanos, S. N., 1965, Ocean-bottom seismic observatories: Proc. IEEE, v. 53, p. 1909-1921.

_____, Odegard, M. E., Mark, N. and Letourneau, N. J., 1969, Research in seismology related to the Columbia ocean-bottom seismograph: Hawaii Inst. Geophysics report HIG-70-12, 66 p.

Ulrych, T. J., 1972, Maximum entropy power spectrum of truncated sinusoids: Jour. Geophys. Res., v. 77, p. 1396-1400.

Whittle, P., 1963, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix: Biometrika, v. 50, p. 129-134.

Except as otherwise noted, copies of all publications listed here are distributed free of charge, as long as available.

Address all requests for publications to:

Publications, Room 262
Hawaii Institute of Geophysics
University of Hawaii
2525 Correa Road
Honolulu, Hawaii 96822

**Out of print*