P.R.L.
TECHNICAL NOTE

TN-88-61

A SOFTWARE PACKAGE FOR CROSS SPECTRAL ANALYSIS EMPLOYING THE DIFFERENCE METHOD

S.K. SHAH

December, 1988

PHYSICAL RESEARCH LABORATORY
AHMEDABAD-9

DOCUMENT CONTROL AND DATA SHEET

1. REPORT NO. : PRL : TN-88-61

2. TITLE AND SUBTITLE :

"A Software Package for Cross Spectral Analysis Employing the difference Method"

3. REPORT DATE: Dec., 1988

4. TYPE OF REPORT: Technical Note

5. PAGES: 23

6. PRICE: Unpriced

7. NO. OF REFERENCES: 2

8. AUTHORS: S.K. Shah

9. a) PURPOSE: Cross Spectral Analysis

b) USEFUL FOR: Any user for analysing the time series data

10. ORIGINATING UNITS/DIVISIONS AND ADDRESS:

Library
Physical Research Laboratory
Navrangpura
Ahmedabad 380 009

11. SPONSORING AGENCY: Nil

12. ABSTRACT: Enclosed

13. KEYWORDS: Difference Method, Quadrature Spectra, Co-spectra Squared Cross Amplitudes, Coherency, Phase Spectra, Bandwidth of a Spectral Window, 95% Confidence Limits.

14. DISTRIBUTION STATEMENT: Distribution unlimited

15. SECURITY CLASSIFICATION: Unclassified

CONTENTS

A SOFTWARE PACKAGE FOR CROSS SPECTRAL ANALYSIS EMPLOYING THE DIFFERENCE METHOD

S.K. SHAH

PHYSICAL RESEARCH LABORATORY

AHMEDABAD

ABSTRACT

The package gives the power spectral density using the algorithm for solution of a difference method given by Jenkins and Watts (1968). In Fast Fourier Transform (FFT), Cooley-Tukey algorithm is based on radix 2 so that either the number of data points or the lag size for calculating auto- and cross-covariances are expressible as a power of 2. The solution of the difference method algorithm is free from the above limitation. The package calculates the co-spectra, qudrature spectra, the squared cross amplitudes, phase, the coherency, the gain and the residual spectra. It also calculates 95% confidence limits and the bandwidth of a spectral window. The user has a choice of selecting a proper lag window viz: Bartlett, Tukey or Parzen. The advantage of the package is that it does not require any complex variables. This results in a very substantial saving of computer time and resources.

INTRODUCTION

Time series analysis is now widely used in many branches of engineering, physical sciences and economics. One important aspect of time series analysis is spectrum analysis which is concerned with the splitting up of time series into different frequency components.

Power spectrum analysis is derived from principles first enunciated by Wiener (1930, 1949). It is based on the premise that time series are not necessarily composed of a finite number of oscillations, each with a discrete wavelength. The spectrum, therefore, yields a measure of the distribution of variance in a time series over a continuous domain of all possible wavelengths, each arbitrarily close to the next, ranging from an infinite wavelength to the shortest wavelength that can be resolved by any scheme of harmonic analysis.

The most important assumptions made about a time series are that the corresponding stochastic process is stationary and that a stationary stochastic process may be adequately described by the lower order moments of its probability distribution function. The lower order moments include mean, variance and co-variance. Since the statistical properties of stationary series do not change with time, these properties can be conveniently summarised by computing certain functions from the data. The functions which are used for this purpose are the auto-covariance and

cross-covariance functions. The Fourier transform of the covariance function namely the sample spectrum has the undesirable property that its variance does not decrease as the length of the time series increases. The improved estimation of the spectrum can be obtained by smoothing the sample spectrum. The more one smooths, the smaller is the variance but the larger is the bias or distortion. Hence a compromise has to be achieved between bias and variance. That is the variance of the smoothed spectral estimator can be reduced by making the value of the truncation point M small. However, reducing the value of M increases the bias and causes more distortion of the theoretical spectrum. Hence the exact choice of M is a vital matter.

DESCRIPTION

The package first reads two time series, $X_1(t)$ and $X_2(t)$, of time t. The data are then smoothed using moving average method. The different lag windows are incorporated in the package namely, Bartlett, Tukey and Parzen. The two auto-covariances of $X_1(t)$ and $X_2(t)$ and cross-covariance between $X_1(t)$ and $X_2(t)$ as well as between $X_2(t)$ and $X_1(t)$ are calculated using the standard equations ,

$$C_{xx}(K) = [\sum_{t=1}^{N-K}(X_{it}-\bar{X}_i)(X_{jt+k}-\bar{X}_j)]/N \quad ;k=0,1\ldots N-1$$

where $\bar{X}_i = [\sum_{t=1}^{N} X_{it}]/N$; $i,j=1,2$

Since power spectrum is an even function of frequency, it is only necessary to calculate it over the range of $0 \leq f \leq 1/2\Delta$. Δ is the sampling interval. However, to preserve the Fourier transform relationship between the sample spectrum and the sample autocovariance function, it is necessary to double the power associated with each frequency in the above range.

$$\hat{C}_{xx}(i) = 2\Delta[C_{xx}(0) + 2\sum_{K=1}^{L-1} C_{xx}(k)W(k)\cos(\pi ki/f)], \quad 0 \leq i \leq f$$

where W(k) is the lag windowing function.

The package allows three different lag windows.

(1) Tukey : $W(k) = 0.5(1 + \cos \pi K/M)$, $K \leq M$

(2) Bartlett: $W(K) = 1 - K/M$, $K \leq M$

(3) Parzen : $W(K) = 1 - 6(K/M)^2 + 6(K/M)^3$, $K \leq M/2$

$= 2(1-K/M)^3$, $M/2 < K \leq M$

The above equation is solved using the solution of a difference method algorithm given by Jenkins and Watts (1968).

A linear difference equation is of the form

$$Y_r = \alpha_1 Y_{r-1} + \alpha_2 Y_{r-2} + \cdots + \alpha_m Y_{r-m} + \beta_0 X_r + \cdots + \beta_n X_{r-n}$$

The quantities Y_r, Y_{r-1}, ---- Y_{r-m} and X_r, X_{r-1} ---- X_{r-n} could be values of continuous signals Y(t) and X(t) at the instance $t = r\Delta$, $(r-1)\Delta$, ---- $(r-m)\Delta$, $(r-n)\Delta$ respectively.

This has the Z transform

$$H(z) = \frac{\beta_0 + \beta_1 Z^{-1} + \text{----} + \beta_n Z^{-n}}{1 - \alpha_1 Z^{-1} - \text{-----} - \alpha_m Z^{-m}}$$

Factoring Z^{-m}, substituting p=Z and equating the denominator to zero gives characteristic equation of the discrete system.

$p^m - \alpha_1 p^{m-1}$ ---- $- \alpha_m = 0$ has the roots $\pi_1, \pi_2,$ ----π_m lie inside the unit circle.

The system is stable provided $|\pi_i| < 1$

i.e., all α_i lie in the triangular region.

i.e., $\sum \alpha_i < 1$.

This algorithm has the advantages of relatively high speed, high accuracy and computes only one cosine function for each frequency point. In Fast Fourier Transform, Cooley-Tukey algorithm is based on radix 2 so that either the number of data points or the lag size for calculating the covariances are expressible as a power of 2. The solution of the difference method has an advantage over this restriction. As this method needs to compute only one cosine function, computer storage area can be reduced substantially.

Now the sample cross spectrum is the Fourier transform of the sample cross covariance function. In fact, this sample cross spectrum is a complex quantity and may be

written as the product of a real function called the sample cross amplitude spectrum and a complex function called the sample phase spectrum. The cross amplitude spectrum shows whether frequency components in one series are associated with large or small amplitudes at the same frequency in the other series. Similarly, the phase spectrum shows whether frequency components in one series lag or lead the components at the same frequency in the other series.

Since the sample cross spectrum is a complex quantity, the alternative expressions are used as a sum of a real and imaginary part known as the sample co-spectra and the quadrature spectra using the equation set

$$L_{12}(i)=2[l_{12}(0)+2\sum_{k=1}^{L-1} l_{12}(k)W(k)\cos(\pi ik/f)] \quad , \quad 0\leqslant i\leqslant f$$

$$Q_{12}(i)=4\sum_{k=1}^{L-1} q_{12}(k)W(k)\sin(\pi ik/f) \quad , \quad 1\leqslant i\leqslant f-1$$

where $l_{12}(k)$ and $q_{12}(k)$ are the even and odd cross covariance function estimates.

The statistical properties of the cross correlation between the two time series can be described by the squared coherency. It provides a non-dimentional measure of the correlation between two time series as a function of frequency. In fact, the coherency plays the role of a correlation coefficient defined at each frequency of f. Thus it shows that when the noise spectrum is identical to the output spectrum, the squared coherency is zero. In other words, the squared coherency is zero when the output

consists entirely of noise. The squared coherency is unity when the output spectrum is simply the input spectrum multiplied by the square of the gain of the system. The squared coherency is small when the output signal to noise ratio is small and large when the ratio is large.

The package calculates the squared coherency using the equation

$$K_{12}^2(i)=\sqrt{L_{12}^2(i)+Q_{12}^2(i)}/C_{11}\hat{}(i)C_{22}\hat{}(i) \quad , \quad 0\leq i\leq f$$

as the basis for a frequency-domain test for correlation between the two time series.

The covariance property shows that the phase and coherency estimations are uncorrelated and hence it is permissible to derive confidence intervals for these spectra separately. Apart from the effect of smoothing, the varianace of absolute coherency is identical to the variance of an ordinary correlation coefficient. Hence R.A. Fisher's Z-transformation can be applied. Using the equations

$$Y_{12}(i)=\text{arctanh}(|K_{12}(i)|)$$

$$Y_{12}(i) \pm 1.96\sqrt{\text{var}(Y_{12}(i))} \quad , \quad 0\leq i\leq f$$

the upper and the lower confidence limits of coherency are constructed. The ratio of the cross spectral estimator to the input spectral estimator is known as the estimator of gain. Using the equations

$$G(i)=\sqrt{L_{12}^2(i)+Q_{12}^2(i)}/\hat{C}_{11}(i)$$

$$F_{12}(i)=\arctan(-Q_{12}(i)/L_{12}(i)) \quad , \quad 0 \leqslant i \leqslant f$$

the gain spectra and phase spectra are calculated.

A measure of the reduction in variance of the estimator due to smoothing by the spectral window is known as the bandwidth of a spectral window. In order to obtain a good estimate of a peak in a spectrum, the width of a spectral window must be the same order as the width of the peak. According to the users' choice of a lag window, the appropriate bandwidth is calculated. The variance of the spectral estimator is inversely proportional to the bandwidth of the spectral window.

PACKAGE TESTING

We have tested the package using the data given in reference (1).

Realisations of two bivariate autoregressive processes are taken as an example. The two processes are as below:

$$X_{1t} = 0.6\ X_{1t-1} - 0.5\ X_{2t-1} + Z_1t$$

$$X_{2t} = 0.4\ X_{1t-1} + 0.5\ X_{2t-1} + Z_2t$$

Table 1 shows the exact data points of above two processes for the calculations. The Tukey lag window is applied to the above mentioned data. Lag size for calculating the coherency spectra and the phase spectra is taken as 16. The bandwidth (B.W.) of a spectral window is superimposed on the

coherency spectra. 95% confidence limits are also plotted in Fig. 5. The package took 25.98 sec. computer time to calculate the above results.

PROBLEM STATEMENT

The package has been tested on a geological data set in the context of following problem.

In the loess deposits of Kashmir, several palaeosol layers are seen. These layers exhibit relatively high values of magnetic susceptibility (χ) and its frequency dependent component (χ_{fd}) as seen in figure 8. The age assignment of some of the palaeosol layers is based on interpretation of thermoluminescence dates obtained independently. It is believed that the palaeosol layers were found as a result of pedogenesis during climatically warm humid phases. Oxygen-18 ($\delta^{18}O$) variations from sea bottom cores indicate the global climatic variations, being the response of variable land ice water storage during glacial and interglacial periods. To see if the palaeosols in the loess deposits in Kashmir are in fact in some way related to the global palaeoclimatic variations, we have used the present cross correlation program to study the periodicities of coherency between the two time series.

STRUCTURE OF THE PACKAGE

The package is developed in FORTRAN 77 language and extensively tested on DEC-10 computer.

The main program COHER calls the following different modules.

(1) SMOOTH - Calculates the moving average for the given two series, supplied by the number of points and the number of order to be smoothed.

(2) AUCS - Calculates the auto and cross covariances with the different number of time lags for the given data.

(3) AMAX - Finds the maximum and minimum values and its position from the given array.

(4) ATSPEC - Calculates the auto spectrum using the auto covariance for the different number of frequency points.

(5) CRSPEC - Calculates the co-spectrum, Quadrature spectrum and squared cross amplitudes for the given two data arrays.

(6) SIGMA - Calculates the mean and the standard deviations for the given array.

(7) SLOG10 - Converts the spectrum points into logarithmic values.

(8) PEAK - Finds the peak values from the auto-spectrum; the coherency spectrum and the phase spectrum

to get the different periodicities.

This package contains 385 lines. COHER.EXE occupies 68 blocks on the disk.

HOW TO USE THE PACKAGE

The package requires two input data files and five output files. User has to create the input data files 1 and 2 for the data array 1 and 2 respectively as below:

The first record of each data file should contain total number of data points in I4 format and 40 characters long title for the data identification.

The second and onward records should contain the data points each separated by commas.

RUN COHER.EXE [302,107]

RUN INPUT DATA FILE 1	=	give file name for input file 1
SUPPLY INPUT DATA FILE 2	=	give file name for input file 2
SUPPLY OUTPUT FILE (AUTO-CORR 1 & 2)	=	give file name for auto correlation values for array 1 and 2 to be stored.
SUPPLY OUTPUT FILE (CROSS-CORR)	=	give file name for cross-correlation values between array 1 & 2 to be stored
SUPPLY OUTPUT FILE (SPECTRUM 1 & 2)	=	give file name for auto spectrum values for array 1 & 2 to be stored.

SUPPLY OUTPUT FILE (COHERENCY)	=	give file name for coherency spectrum values between array 1 & 2 to be stored
SUPPLY OUTPUT FILE (PHASE)	=	give the file name for phase spectra to be stored.
SAMPLING INTERVAL	=	give sampling interval between two adjacent data points.
LAG SIZE	=	give number of lags for calculating coherency and phase spectra.
SMOOTHING REQUIRED? (1 = YES / 0 = NO)	=	give appropriate value
NO.OF POINTS,NO.OF ORDERS	=	if smoothing is required, give the no. of points and no. of orders to be smoothed.
LAG WINDOWS (1=TUKEY, 2=BARTLETT, 3=PARZEN)	=	give the appropriate lag window.

The package executes the data, given in input file (1) and (2) and creates five output files mentioned as above for plotting purpose.

Output file 1	contains auto correlation values for array (1) and auto correlation values for array (2).
Output file 2	Contains cross correlation values between the array (2) and (1) as well as

between the array (1) and (2).

Output file 3 — Contains auto spectra values for array (1) and auto spectra values for array (2).

Output file 4 — Contains coherency spectra values, the upper 95% confidence limit values, the lower 95% confidence limit values for coherency spectra and the bandwidth values of a spectral window.

Output file 5 — Contains phase spectra values, the upper 95% confidence limit values, the lower 95% confidence limit values for phase spectra.

The package will also give the individual periodicities for array (1), array (2) and the coherency as mentioned in Fig. 7.

FIGURE CAPTIONS

Fig. 1: Auto correlation of process 1.

Fig. 2: Auto correlation of Process 2.

Fig. 3: Cross correlation between Process 1 & 2.

Fig. 4: Power spectra of Process 1 & 2.

Fig. 5: Coherency spectra between Process 1 & 2.

Fig. 6: Phase spectra.

Fig. 7: The sample output for the individual periodicities.

Fig. 8: Field and mineral magnetic stratigraphies

of selected loess-palaeosol section in Kashmir.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. S.K.Gupta and Dr. P.Sharma for their keen interest and valuable suggessions. Thanks are also due to Mr. P.Pillai for typing of the manuscript.

REFERENCES

(1) Jenkins and Watts (1968), Spectral analysis

(2) Reid J.S. (1979),Confidence limits and Maximum Entropy Spectra and its applications.

TOTLE NO. OF DATA POINTS : 100 (ROW WISE)

LINEAR PROCESS 1

-0.9	-0.2	-1.9	-1.1	1.4	2.1	2.8	0.6	-0.7	-1.8
-3.8	-2.4	1.0	0.7	-0.2	1.0	0.1	-0.4	-0.7	0.9
-1.6	-0.4	-1.4	-1.7	-1.2	-2.0	-0.2	0.4	1.3	0.7
0.3	0.5	-1.9	-0.9	-1.5	-0.1	1.0	0.0	-0.8	0.1
-0.6	-0.5	-0.1	1.2	1.5	0.6	0.4	2.2	3.2	2.1
0.4	-0.2	0.6	-0.5	2.4	1.0	-0.5	-2.5	-2.1	-1.0
-0.1	-1.9	-1.5	1.5	3.3	3.1	1.7	0.8	1.5	0.9
-0.9	-1.2	0.9	1.7	3.1	0.2	-1.0	0.1	0.1	0.1
-2.6	-1.3	1.1	3.2	1.9	0.5	-1.1	0.5	-0.6	0.2
1.2	-1.0	-1.6	1.1	-0.7	-0.9	-0.1	0.2	0.5	-2.2

LINEAR PROCESS 2

0.8	1.1	-1.1	-2.4	-1.8	-0.8	-0.4	1.3	1.8	2.4
0.4	-2.1	-1.9	-1.3	-1.8	-0.3	0.7	0.5	0.7	0.7
0.1	0.6	1.5	0.1	0.5	-1.4	-2.6	-1.7	-0.4	0.6
2.2	-0.2	0.6	-0.2	-1.5	-0.6	-1.1	0.9	-0.7	-0.4
0.5	0.5	0.0	-0.7	0.2	0.6	-1.3	-0.3	0.3	2.2
3.0	1.6	-0.4	-0.6	-0.1	3.0	2.1	0.8	0.9	-1.5
-0.4	-0.4	-1.4	-4.2	-0.7	-1.0	0.4	0.1	-1.9	-0.1
0.2	1.0	0.0	0.4	0.2	2.7	0.6	0.3	1.1	0.5
-1.1	-2.3	-2.0	-0.8	1.0	1.7	0.6	2.0	1.0	1.9
2.2	3.1	0.6	0.5	1.3	0.6	0.1	0.0	2.3	1.9

TABLE - 1

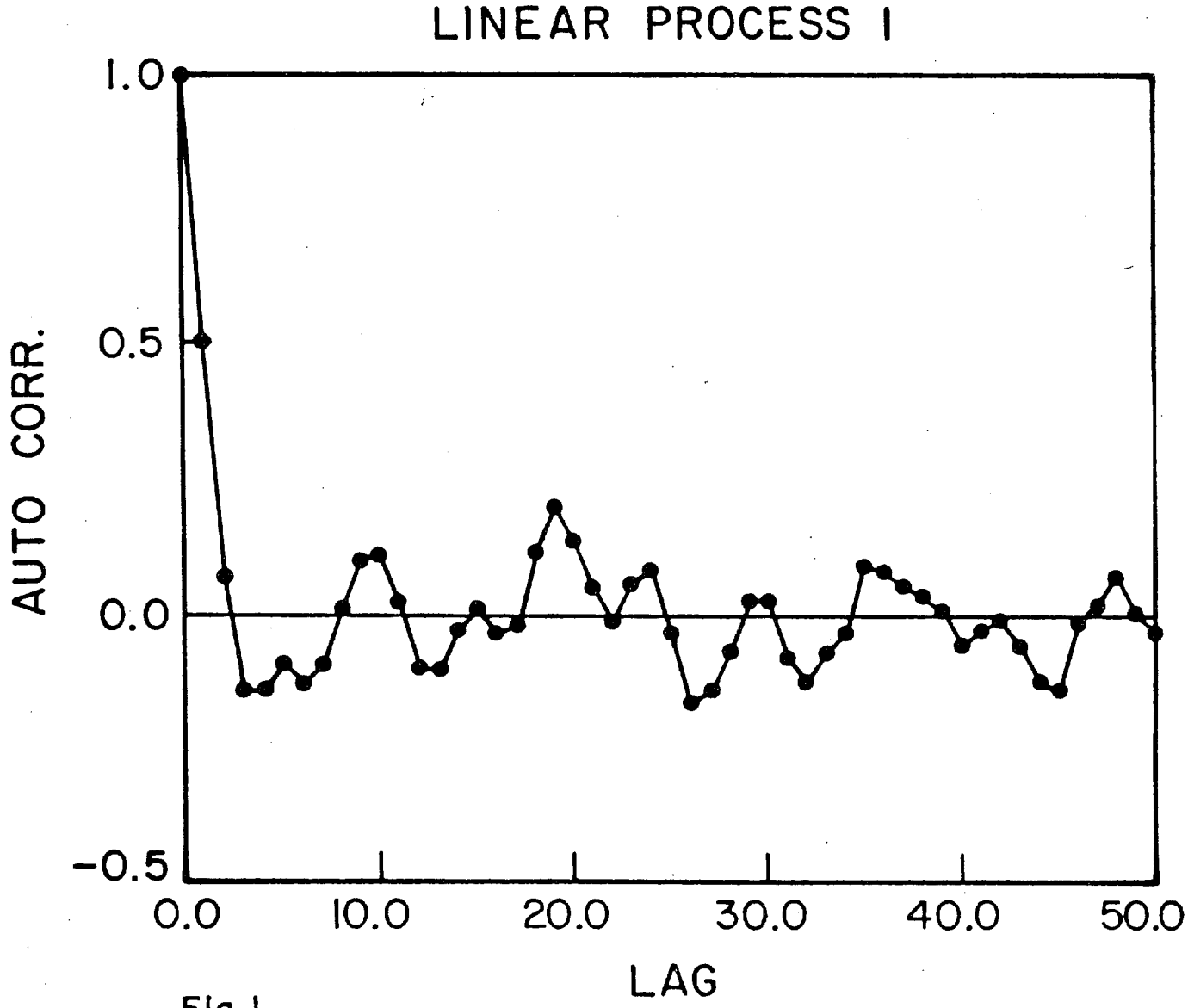
LINEAR PROCESS I
1.0
0.5
0.0
-0.5
AUTO CORR.
0.0
10.0
20.0
30.0
40.0
50.0
LAG

Fig. 1

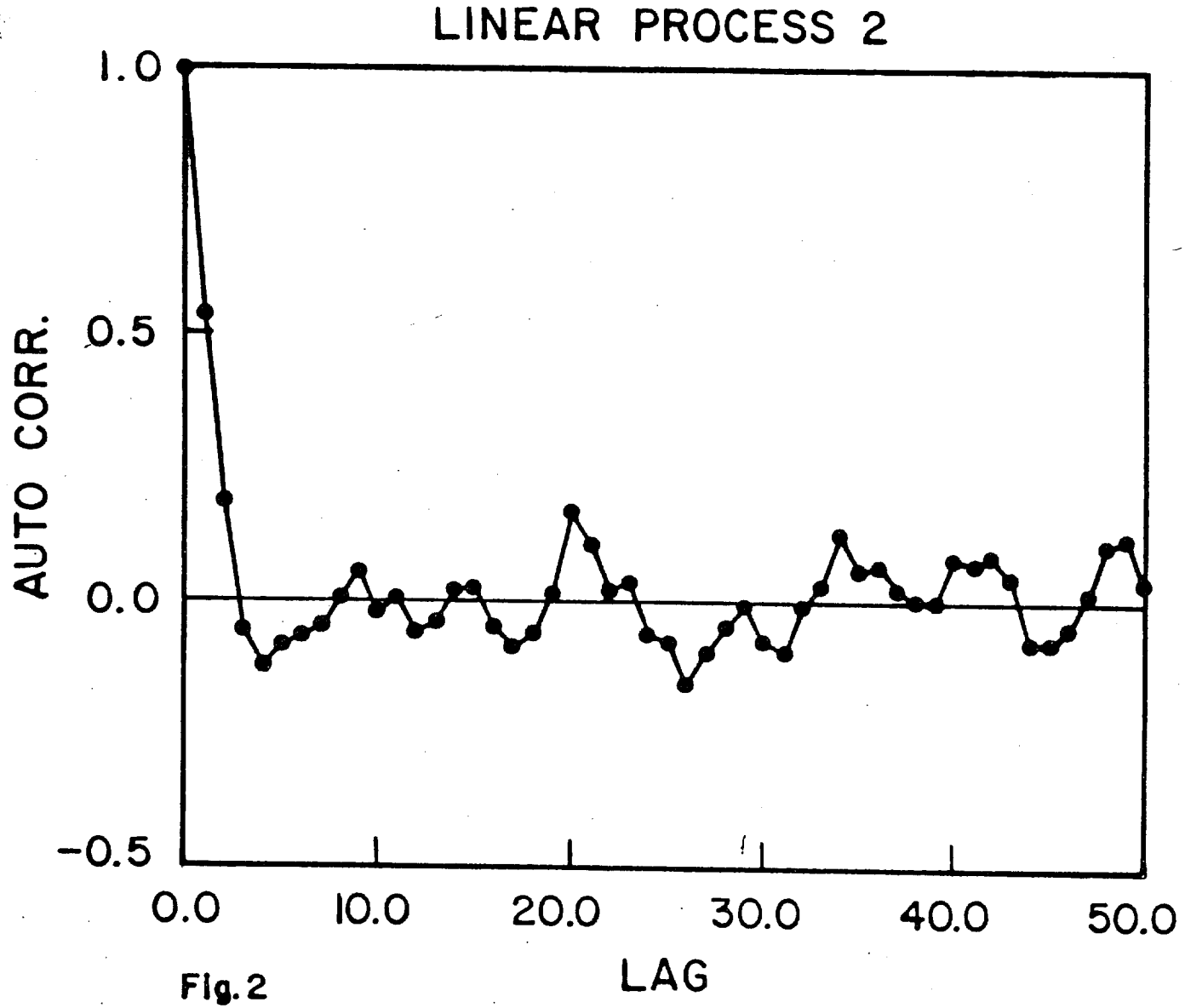
LINEAR PROCESS 2
1.0
0.5
0.0
-0.5
AUTO CORR.
0.0
10.0
20.0
30.0
40.0
50.0
LAG

Fig. 2

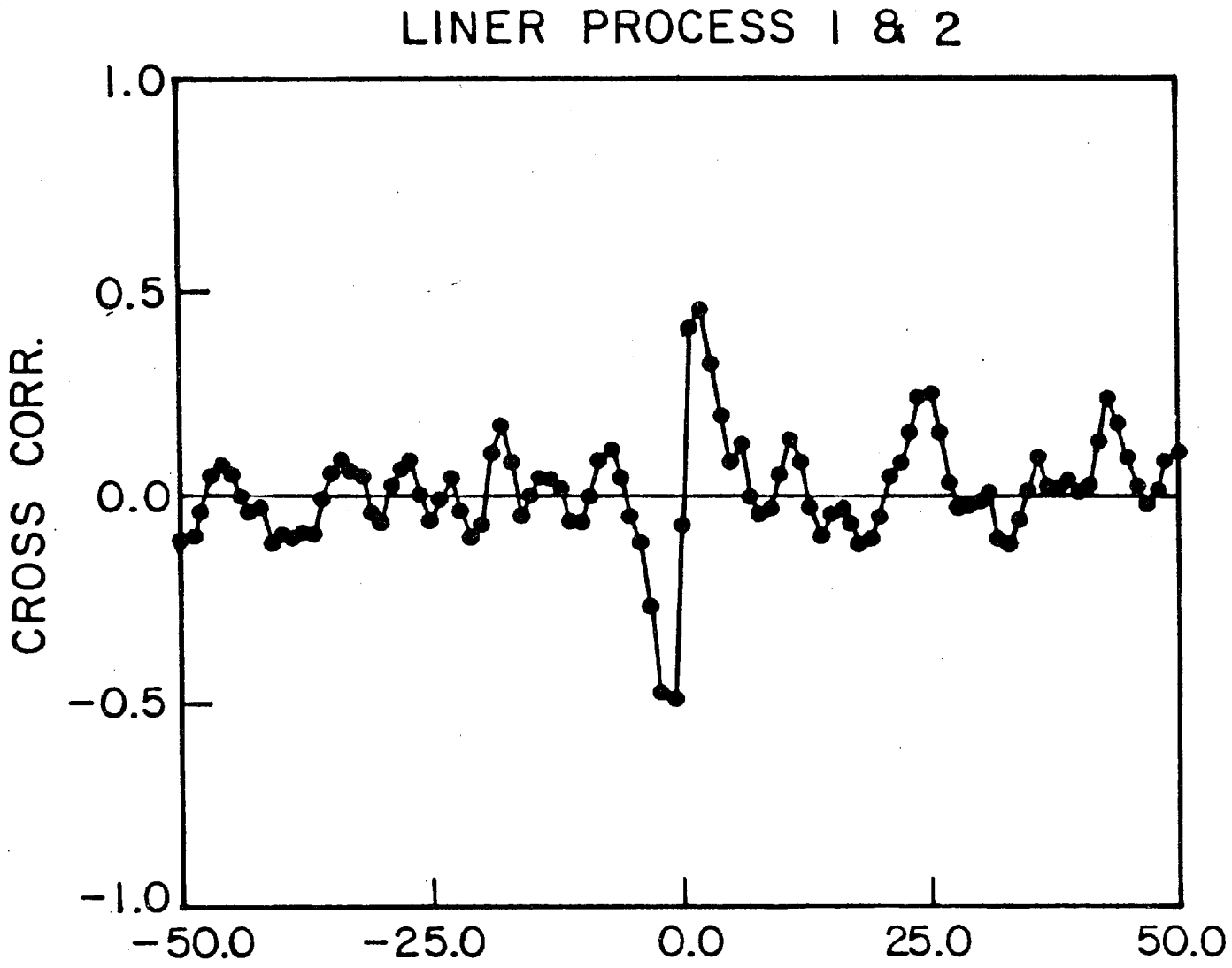
LINER PROCESS 1 & 2
1.0
0.5
0.0
-0.5
-1.0
CROSS CORR.
-50.0
-25.0
0.0
25.0
50.0
LAG

Fig.3

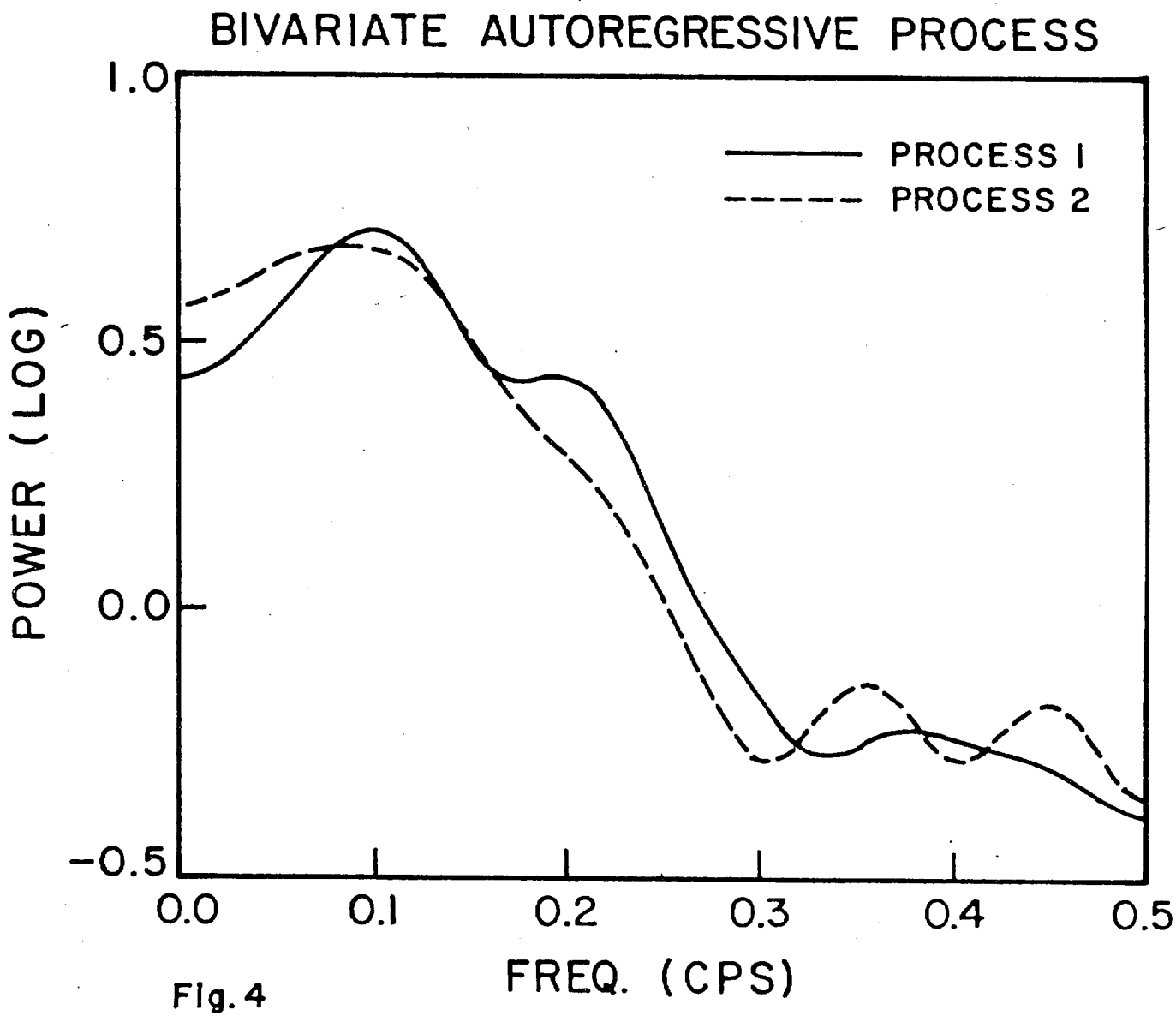
BIVARIATE AUTOREGRESSIVE PROCESS
PROCESS 1
PROCESS 2
POWER (LOG)
1.0
0.5
0.0
-0.5
0.0
0.1
0.2
0.3
0.4
0.5
FREQ. (CPS)

Fig. 4

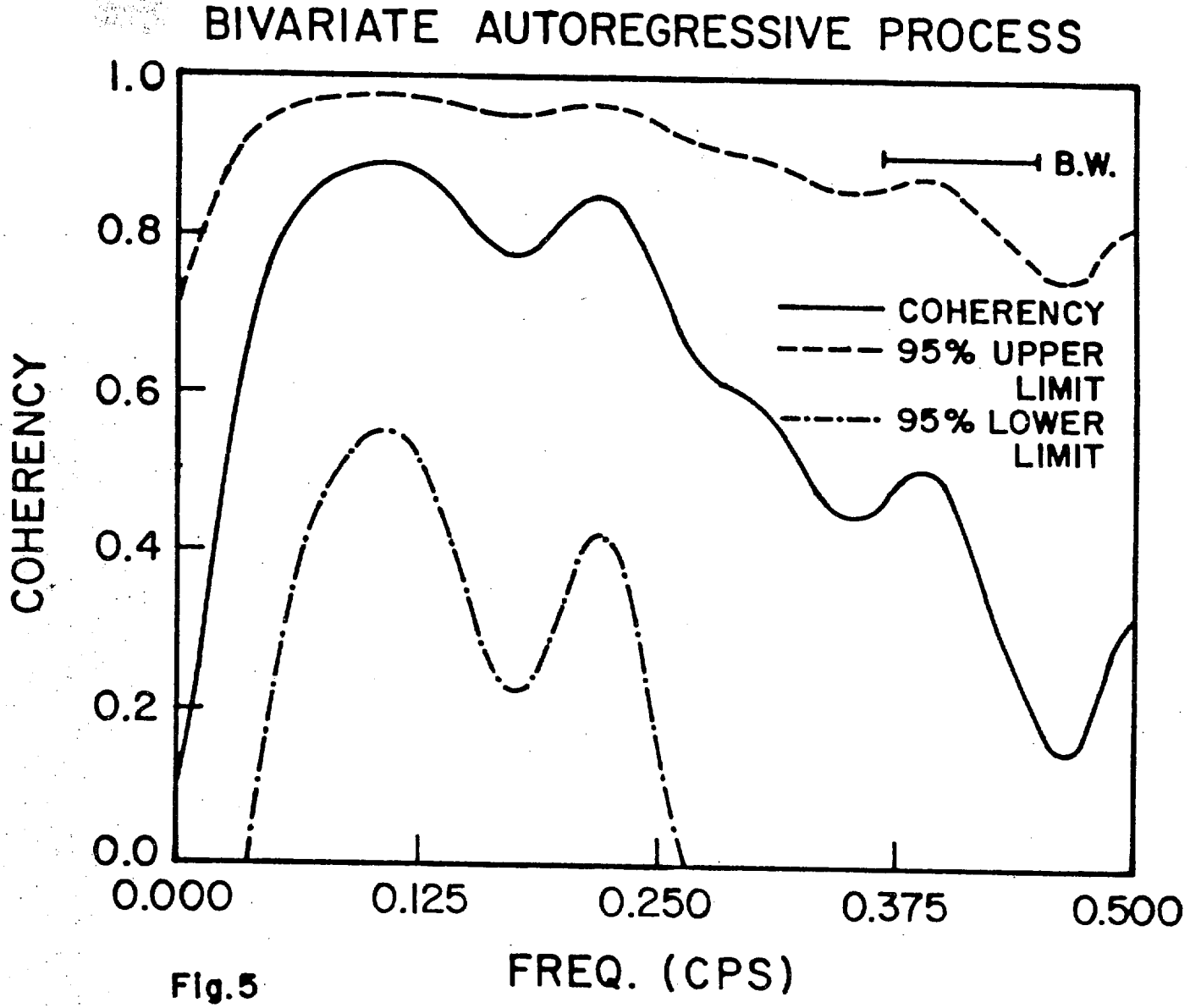
BIVARIATE AUTOREGRESSIVE PROCESS
1.0
0.8
0.6
0.4
0.2
0.0
COHERENCY
B.W.
COHERENCY
95% UPPER LIMIT
95% LOWER LIMIT
0.000
0.125
0.250
0.375
0.500
FREQ. (CPS)

Fig.5

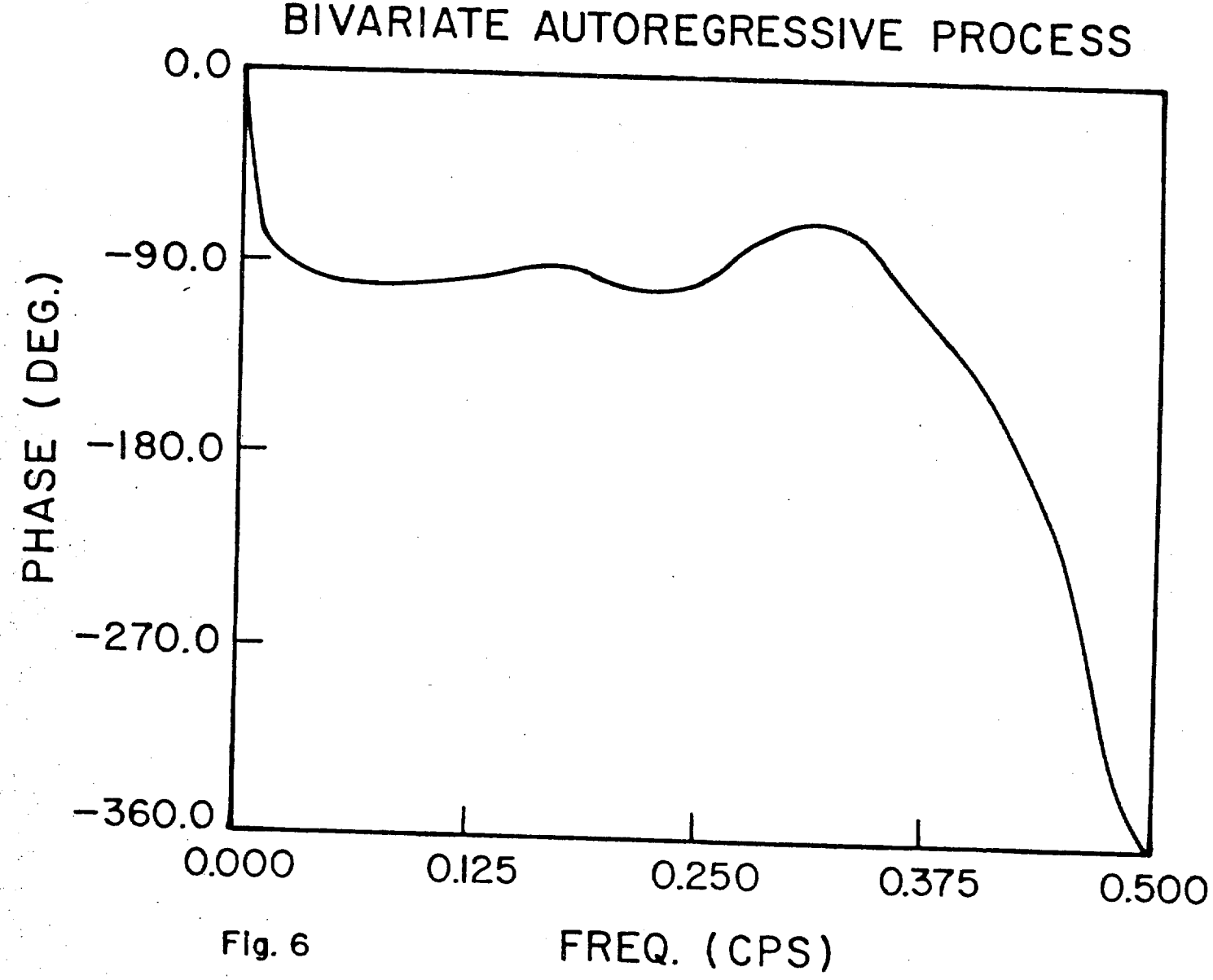
BIVARIATE AUTOREGRESSIVE PROCESS
0.0
-90.0
-180.0
-270.0
-360.0
PHASE (DEG.)
0.000
0.125
0.250
0.375
0.500
FREQ. (CPS)

Fig. 6

NO. OF DATA POINTS = 100
LAG SIZE = 16
DT = 1.0000

SOLUTION OF DIFFERENCE EQU. ALGORITHM

LINEAR PROCESS 1

PEAK---FREQ.	0.104167	POWER:	0.711487	PERIODICITY:	10
PEAK---FREQ.	0.197917	POWER:	0.433242	PERIODICITY:	5
PEAK---FREQ.	0.375000	POWER:	-0.225033	PERIODICITY:	3

LINEAR PROCESS 2

PEAK---FREQ.	0.093750	POWER:	0.685465	PERIODICITY:	11
PEAK---FREQ.	0.354167	POWER:	-0.140102	PERIODICITY:	3
PEAK---FREQ.	0.447917	POWER:	-0.179045	PERIODICITY:	2

COHERENCY SPECTRUM

PEAK---FREQ.	0.104167	POWER:	0.891181	PERIODICITY:	10
PEAK---FREQ.	0.218750	POWER:	0.850378	PERIODICITY:	5
PEAK---FREQ.	0.385417	POWER:	0.505646	PERIODICITY:	3

FIG - 7

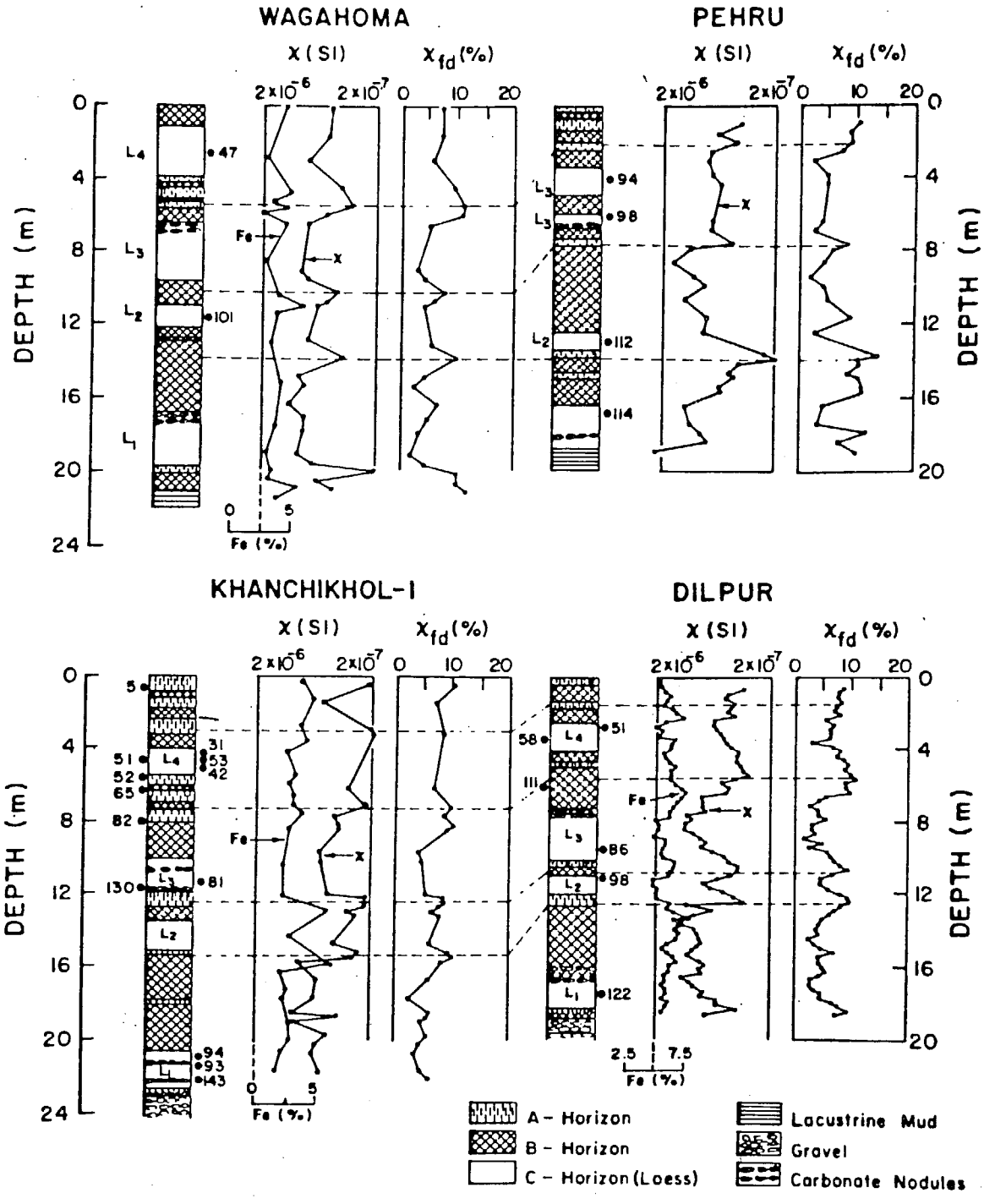
WAGAHOMA
PEHRU
KHANCHIKHOL-I
DILPUR
χ (SI)
χfd (%)
DEPTH (m)
Fe
Fe (%)
A – Horizon
B – Horizon
C – Horizon (Loess)
Lacustrine Mud
Gravel
Carbonate Nodules

Fig. 8