APPLICATION OF SEISMIC REFLECTION DATA TO DISCRIMINATE SUBSURFACE LITHOSTRATIGRAPHY

A THESIS

Submitted in fulfilment of the
requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

APPLIED GEOPHYSICS

By
Miss Amita Agarwal
Now
Mrs. AMITA SINVHAL

Forwarded

M. Lihal
22.10.79

PROFESSOR & HEAD
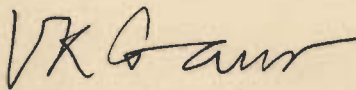Department of Earth Sciences,
University of Roorkee,
ROORKEE-247672

DEPARTMENT OF EARTH SCIENCES
UNIVERSITY OF ROORKEE
ROORKEE-247672, (INDIA)

October, 1979

CERTIFICATE

Certified that the thesis entitled 'APPLICATION OF SEISMIC REFLECTION DATA TO DISCRIMINATE SUBSURFACE LITHO-STRATIGRAPHY' which is being submitted by Mrs.AMITA SINVHAL in fulfilment of the requirements for the award of the Degree of DOCTOR OF PHILOSOPHY IN APPLIED GEOPHYSICS of the University of Roorkee, is a record of the student's own work carried out by her under our supervision and guidance. The matter embodied in this thesis has not been submitted for the award of any other Degree or Diploma.

This is further to certify that the candidate has worked for a period of three years from October 1976 to October 1979 for preparing this thesis at this University.
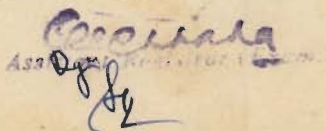
(Dr.V.K. Gaur)
Professor of Geophysics,
Department of Earth Sciences,
University of Roorkee,
Roorkee-247 672 (India).

(Dr.K.N.Khattri)
Professor of Geophysics,
Department of Earth Sciences,
University of Roorkee,
Roorkee-247 672 (India).

University of Roorkee, Roorkee
Certified that the attached Thesis/
Dissertation has been accepted for the
award of Degree of Doctor of
Philosophy / Master of Engineering
Earth Science
No. Ex/150 E191

A C K N O W L E D G E M E N T S

(Amita Sinvhal)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure No.	Title	Page No.

A B S T R A C T

The ever increasing demand for energy has necessitated the exploration of hydrocarbons in stratigraphic traps. The seismic technique can be effectively used to elucidate subsurface stratigraphy and lithology. To interpret seismic responses of geologic sections in terms of subsurface stratigraphic and lithologic information it is necessary to establish a correlation between lithology and suitable parameters abstracted from the seismic response. The present work deals with

(a)   simulating mathematical models for sedimentation processes and calculating their response with the above objective and

(b)   applying the above concept and methodology developed on synthetic data to real seismograms to infer lithostratigraphic information.

Depositional situations may be modeled by using Markov chains. These involve the concept of memory where the nature of successor lithologies are predetermined by preceding lithologies according to certain probabilities. Markov chains with one step memory are therefore applied to model two different depositional conditions of a formation

in a sedimentary basin in India[+]. Accordingly, two Areas X
and Y are considered for the purpose of this study.

Area X corresponds to a dominantly sandy (sand =
53 percent) part of the basin together with coal (26 percent)
and shale (21 percent) constituents. Geophysical well logs
of this area have been used to calculate the probability of
upward transition from one lithology to another at a four
metre sampling interval for a particular formation. These
were used to generate 255 different synthetic stratigraphic
sequences which are collectively designated as Model E.
Area Y corresponds to a dominantly shaly (shale = 60 percent)
part of the same basin, with sand (37 percent) and coal (3
percent) constituents. Another 253 synthetic sequences
generated for this area and designated as Model F were syn-
thesized on the basis of the probabilities of transitions
from one lithology to another as calculated from well log
data. The five hundred and eight sedimentation sequences
thus generated represent sedimentary sequences deposited in
changing environments. Seismic response in time and frequency
domain for these models have been calculated.

[+]Part of the work embodied in the thesis is based on
real field data, courtsey, Oil and Natural Gas Commission,
Dehradun, India. The locations and name of basin have
been suppressed. The sedimentary basin is referred to as
Basin Z, two areas within the basin as Areas X and Y and
the hydrocarbon bearing formation as Formation K.

The models used in this study are composed of homogeneous, isotropic and perfectly elastic layers. The acoustic impedances of these layers were calculated from the velocity and density logs available for the area. The impulse response was calculated and then convolved with a source wavelet to yield conventional looking seismograms. The autocorrelation function, and the power spectrum using maximum entropy methods were computed.

Seventeen variables were picked from the auto-correlation function (ACF) and are A_1/A_0, A_2/A_0, A_3/A_0, where A denotes the ACF at the subscripted lag, A_{min}/A_0, where A_{min} denotes the minimum value of the ACF, T_1, T_2, T_3, where T denotes time of the subscripted zero crossing in the ACF and T_{amin}, the time at which first minima occurs. Nine variables were picked from the power spectrum and are, average power weighted frequency, frequency at which maximum power occurs, frequency at 25th, 50th and 75th percentile values of frequency weighted power, frequencies of 25th, 50th and 75th percentile of power, and frequency at which logarithm of power decreases to zero.

The above mentioned seventeen variables were calculated for all the simulated responses of synthetic stratigraphic sequences. Discriminant analysis which was employed showed that a combination of all the variables can maximally seperate, in the variable space, the two different Models E and F. The discriminating seismic attributes characterize

the two sedimentation sequences and may aid the interpretation of field records in terms of subsurface stratigraphy.

The success achieved in discriminating different depositional situations in computer simulations has led to the test of the method with real seismic data. The formation on which the transition matrices were based for simulating Models E and F was marked on the seismic sections of Areas X and Y and the 387 seismic traces when subjected to the discriminant analysis allowed to distinguish between lithostratigraphic units of Areas X and Y, thereby endorsing the validity of this approach. Contributions of the seventeen variables towards effective discrimination shows that only seven variables, viz., f_8, A_2/A_0, f_1, T_{amin}, f_M, f_3 and A_1/A_0, common to both synthetic and field seismic data, make positive contributions. The variables designated as seismic discriminators of subsurface lithostratigraphy may ultimately help discriminate an oil bearing stratigraphic trap from its barren surroundings in a sedimentary basin.

The statistical method presented here has been shown to be a potential tool for the determination of subsurface lithostratigraphy from seismic data. This constitutes on important additional tool in the exploration for hydrocarbons.

CHAPTER - I

INTRODUCTION

Over a third of the world's power comes from oil.
Its rate of consumption has far exceeded the rate of produc-
tion and discovery of new reserves. This imbalance has
prompted the search for new reserves of oil locked in strati-
graphic, structural and combination traps. Whereas most
structural traps have already been discovered and are being
exploited, stratigraphic traps which have not yet received
their due share of attention hold promise of containing large
reserves of the yet undiscovered oil and gas.

Seismic methods have played an important role in
exploration of oil, especially in locating structural traps.
Recent advances in exploration geophysics have considerably
improved the seismic resolving power thereby enhancing the
chances of locating stratigraphic traps. This has led to the
development of new interpretative modeling techniques which
can help in solving stratigraphic problems, in predicting
lithologies and their inter-relationship which at times yield
information regarding conditions favourable for accumulation
of hydrocarbons.

1.1 SEISMIC STRATIGRAPHIC EXPLORATION

The interpretation of subsurface stratigraphy from
seismic data is possible by studying the nature of reflection
cycles and their termination with respect to adjacent reflection

events (Vail, Mitchum, Todd, Widmier, Thompson, Sangree, Bubb and Hatlelid, 1977). These terminations help to locate boundaries between zones corresponding to specific types of depositional units, each associated with characteristic reflection patterns. The degree of convergence or divergence of reflection is also diagnostic of environmental conditions prevailing during deposition of a sedimentary sequence. Such mechanisms as delta formations, transgression and regression of sea level and tilting of strata are associated with patterns on record sections that make it possible for the interpreter to reconstruct the geologic history of sedimentary areas (Dobrin, 1977). The study of patterns can narrow the search areas in which the depositional environment appears favourable for stratigraphic accumulation of hydrocarbons.

Lyons and Dobrin (1972) have stated that more than half the oil and gas that will be eventually found will be designated as occuring in stratigraphic traps. They have cited the case of a 'mature' exploration province in Oklahoma, where in 1942, 49 percent of the oil and gas pools were stratigraphic, it rose to 62 percent in 1967 with the discovery of four times as many pools. The great size of some of these stratigraphic traps and their greater number will bring them ahead of the structural traps in ultimate reserves.

Definition of stratigraphic features ideally requires the seismic delineation of lithologic boundaries or the resolution of thin beds which depends on seismic wavelengths. Source wavelets of small wavelengths can resolve thin beds compared

to longer wavelengths. If two reflection boundaries are close together in terms of seismic wavelengths they will not be easily recognized on seismic records. This sets a limit on the resolution of thin beds and pinchouts. Widess (1973) has shown that the thinnest bed that can be resolved by seismic method should have a thickness of 5/8th of a wavelength. However, high frequency energy is difficult to record because of its rapid attenuation, which sets a limit on the depth of penetration of seismic energy and therefore high frequency reflections from deep beds may not be obtained.

Lyons and Dobrin (1972) have suggested the following improvements for increasing seismic resolution : use of high frequency source pulse, detonation of shots in consolidated material, high frequency recording, filtering programmes to increase the signal to noise ratio of high frequency reflection and the use of vertical arrays of pressure phones in boreholes.

Pioneering work in interpretation of stratigraphy from seismic data has been carried out by several workers using different approaches. Cook and Taner (1969) and Taner and Koehler (1969) have used interval seismic velocity for identifying lithology. This approach is used as a regular exploration tool for determining the sand shale ratio although it suffers from the limitation that different interval velocities emerge from different choice of intervals. Hilterman (1970); Harms and Tackenberg (1972); Gir (1974); Dedman, Lindsey and Schramm (1975); Khattri and Gir (1975, 1976);

Nath (1975); Neidell (1975); Brown and Fisher (1977); Clement (1977); Galloway, Yancey, and Wipple (1977); Meckel and Nath (1977); Neidell and Poggiagliolmi (1977); Schramm, Dedman and Lindsey (1977); Sieck and Self (1977); Stuart and Caughey(1977); Taner and Sheriff (1977); Vail, Mitchum, Todd, Widmier, Thompson, Sangree, Bubb, Hatlelid (1977); Wiemer and Davis (1977); Sangree and Widmier (1979) are some of the workers who have given significant examples of the use of seismic data to model horizontal and vertical facies changes characterising stratigraphic variablity. The above workers have tackled stratigraphic problems related to modeling using the deterministic approach. The seismic response computed is due to a particular stratigraphic situation. To take into account a large number of vertical variations in lithology Sinvhal (1976) and Khattri, Sinvhal and Awasthi (1979) have introduced a statistical approach in characterizing different stratigraphic situations.

Mathematical modeling of sedimentary sequences is essentially a computational procedure. The model is given in terms of interval properties such as velocity, density and bed thickness. The seismic response that would be generated from the assumed geologic situation gives the synthetic seismogram. The objective of this kind of modeling is to identify lithologic changes by interpreting synthetic data. These lithologic changes offer possibilities to locate stratigraphic traps that can hold hydrocarbons.

Certain complex layering patterns may be associated with certain seismic characteristics that can be identified by statistical techniques. Mathieu and Rice (1969); Avasthi and Vema (1973); Waters and Rice 1975) and Sinvhal, Gaur, Khattri, Moharir and Chander (1979) have identified typical reflection patterns from varying lithologic sequences. Mathieu and Rice (1969) and Waters and Rice (1975) have chosen a part of the Pennsylvanian Morrow Formation in wells along certain seismic lines. Synthetic seismograms were made from the velocity logs obtained from the wells. Pattern recognition techniques involving the use of factor analysis were applied to records along lines between wells and the various kinds of lithology at each shot point were identified and mapped. Mathieu and Rice (1969) applied discriminatory analysis to differentiate between sandstone and shale within a specified stratigraphic interval, using the time domain variables. Although techniques of this type were successful in some cases, the authors note that there were instances where this approach did not predict lithology reliably.

Sinvhal (1976) and Khattri, Sinvhal and Awasthi (1979) have used the impulse response of models of subsurface formations and statistically analyzed them for abstracting seismic parameters which could be characteristic of the stratigraphy and lithology of the formations. Two types of formations have been considered, consisting either of sand-shale sequences of coal-shale sequences. Models of the formations are generated using the Monte Carlo method. It is found that three

features in the power spectrum of the impulse response, namely, the frequency f_e at which the spectrum can be divided into a zone of high energy from a zone of low energy, the lowest frequency, f_p, where there is a significant energy peak and the frequency f_m at which there is maximum energy, can be used statistically to distinguish between the formations consisting of sand-shale sequences and the formations made up of coal-shale sequences. Three additional parameters A_2/A_1, A_2/A_0 and A_1/A_0, where A denotes the autocorrelation function of the impulse response at the subscripted lag are also statistically significant discriminators between the sand-shale formations and the coal-shale formations. The discrimination between the two subgroups of each model consisting of more (or less) than 50 percent of one lithology is also feasible, although there are fewer discriminants available.

1.2 SCOPE AND APPROACH

The most remarkable aspect of the study of Sinvhal (1976) and Khattri, Sinvhal and Awasthi (1979) was that a sand-shale sequence could be distinguished from a coal-shale sequence depending on the content of sand, shale and coal. This was the basic promoting feature behind the present endeavour. However, stratigraphic sequences are not random stacks of lithologies, but each unit bears some relation with the lithounit deposited previously. Any mathematical procedure which takes this into account will give a more realistic model. Markov chains offer ample scope for this and have been used in

the present study. Moreover, the seismic response of a
layered medium is more appropriately estimated by convolving
the impulse response with a source wavelet, which will give a
band limited spectrum. A formation with sand-shale-coal
alterations in a sedimentary basin in India has been takenup
for the present study. The following approach has been
adopted :

(a)   Sedimentation models with sand-shale-coal alterations
        have been constructed using one step Markov Chains.
        The probability of upward transition from one litho-
        logy to another required in generating Markov chains
        is calculated from borehole data, from a sedimentary
        basin in India.

(b)   The seismic impulse response of the above models is
        calculated by using velocity and density information
        from well log data. The response is convolved with
        a source wavelet to give conventional type of
        seismograms.

(c)   The autocorrelation function of the synthetic seismo-
        grams has been calculated.

(d)   Power spectrum of the synthetic seismograms is comput-
        ed using the Fourier transform and maximum entropy
        methods.

(e)   Variables which will be used to characterize lithology
        are searched and picked from (c) and (d). These are
        subjected to the statistical linear discriminant
        analysis.

(f)   Parts of seismic sections corresponding to a
       hydrocarbon bearing formation in India[+] have
       been subjected to the analysis as given in (c),
       (d) and (e).

On the basis of several variables selected from
the autocorrelation functions and the power spectra of
seismograms it has been possible to distinguish between
dominantly sandy and shaly lithologies both in the case
of synthetic and real data.

[+]Part of the work embodied in this thesis is based on
real field data, courtsey O.N.G.C., Dehradun. The loca-
tions have been suppressed. The sedimentary basin is
referred to as Basin Z, two areas within the basin as
Areas X and Y and the hydrocarbon bearing formation as
Formation K.

CHAPTER - II

SIMULATION OF MODELS

The nature, cause, effect and any other aspect of a complex natural phenomenon may be studied by physical, conceptual or mathematical models and may often lead to new ideas and discoveries. In mathematical models the essential aspect of the phenomenon are represented by mathematical relationships based on relatively few parameters. A good model would predict essential features of the phenomenon sufficiently accurately. However, the construction of good models may be restricted by an inadequate understanding of the natural phenomenon and its consequent mathematical formulation, or by the number of parameters to describe it, a restriction imposed by the cost of simulation.

In the present investigations geologic depositional situations have been simulated on a digital computer by a mathematical formulation. The depositional process and environmental condition to represent any geologic situation is first assumed and a model for it is visualized. The process is modeled by choosing certain relevant parameters such as velocities, densities and interface geometry which is used as input data to the computer to simulate the geologic model.

2.1 MONTE CARLO MODELS

Sinvhal (1976), Khattri, Sinvhal and Awasthi (1979) and Sinvhal, Gaur, Khattri, Moharir and Chander (1979) have modeled cyclic repetitions of lithologies and have simulated simple binary systems with alternating sand and shale or shale and coal beds. They employed Monte Carlo technique to select thickness of the successive layers, which took into account the variability of the depositional accumulations. After fixing the model thickness at about 200 m for reasons discussed in section 2.5, individual layers with a thickness having a two way vertical travel time of 6 ms or its multiples were considered for the model. With these constraints the number of layers in each model varied between 10 and 25. Each of these approximately 200 m thick models were overlain by a 200 m thick homogeneous layer and underlain by a homogeneous half-space so that the model could be studied in isolation. The velocities and densities assumed for and assigned to the constituent lithologies are similar to those often met in the field conditions and are given in Table 2.1.

Table 2.1 - Velocities and Densities used for Monte Carlo models (After Sinvhal, 1976).

Constituents of the model	Velocity m/sec	Density g /c.c.
Overburden	1400	2.40
Sandstone	2150	2.05
Shale	2000	1.95
Coal	1500	1.50
Lower Half space	2400	2.20

The sand shale models depict sedimentary environments ranging from oscillating marine to continental environments. The coal shale model depicts environmental transitions from continental to marine conditions.

One hundred and ten sand-shale sequences and an equal number of coal-shale sequences were simulated on a computer. These simulations were grouped into four categories :

(i)   Model A : sand shale ratio ranging between .2 and .5

(ii)   Model B : sand shale ratio ranging between .5 and .8

(iii)   Model C : coal shale ratio ranging between .2 and .5

(iv)   Model D : coal shale ratio ranging between .5 and .8.

The study and analysis based on the seismic response of the above simulations helped in differentiating gross lithologies in the four models discussed above. It is pertinent to note that only random distribution of lithologies and thicknesses generated by the Monte Carlo techniques were considered for the study. The lithounits and their thicknesses in a sedimentary series of beds deposited conformably often have more than two component lithologies and the individual lithounits may be inter-related by a certain probability of transition. Thus the Monte Carlo Models discussed above are rather simplistic and do not define natural depositional sequences met in nature sufficiently accurately. Therefore, it was desirable to extend the study by taking into account a more realistic representation. Markov Chains make it possible to model lithostratigraphy in terms of transition probabilities in which lithounits display

partial interdependence on each other, and have been used in the present study.

2.2 MARKOV CHAINS AND TRANSITION MATRICES

A stratigraphic sequence of conformable beds can be considered as a series of partially interdependent finite number of lithologies. Such a situation can be ideally modeled using Markov Chains which involve the concept of conditional probability. Markov Chains may be regarded as a sequence or a chain of discrete states - in this case lithologies in space (or time) in which the probability of transition from one state to another in the next step in the sequence depends upon the previous state. If the system at a certain point, x_r (or time t_r) depends upon the state at point x_{r-1} (or time t_{r-1}) according to certain probabilities, then it is known as a first order Markov Chain.

Markov Chains can be conveniently used to model complex processes which are subjected to influences that cannot be exactly evaluated. The changes of state can be rigorously examined in terms of their relative probabilities of occurrence. This is evident in cyclical sedimentary sequences, where an underlying pattern of lithologic succession can be discerned, but in which the actual sequence of rock types can only be predicted in terms of their relative probabilities. Carr, Horowitz, Harbar, Ridge, Rooney, Straw, Webb and Potter (1966); Krumbein (1967,68); Potter and Blakely(1967) and Vistelius (1967) have used Markov Chains in modeling

stratigraphic sequences.

Let a Markov Chain consist of three states s_1, s_2 and s_3. Let p_{ij} be the probability of transition from i^{th} state $(i=1,2,3)$ to the jth state $(j=1,2,3)$ in which the ith state underlies the jth state. These probabilities are based on frequency distribution of various transitions amongst different states. Let these probabilities, p_{ij}, be expressed in the form of a transition probability matrix, P, given in the matrix 2.1. In this matrix i and j correspond to rows and /columns respectively, i.e., p_{ij} is the probability of vertical transition from state s_i to state s_j, or p_{12} is the probability of upward transition from state 1 to state 2.

$$P = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{ccc} s_1 & s_2 & s_3 \\ \left[\begin{array}{ccc} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{array} \right] \end{array} \qquad \ldots (2.1)$$

if b_i are the total number of transitions possible from state i to any other state and a_{ij} are the number of transitions from state i to state j, then $P_{ij} = a_{ij}/b_i$, $(i=1,2,3)$ and $(j=1,2,3)$. Several upward transition matrices may be converted into one average matrix by using the formulation

$$\left(\sum_{k=1}^{n} a_{ijk}\, b_{ik} \right) \Big/ \sum_{k=1}^{n} b_{ik} \qquad \ldots (2.2)$$

This gives the element in ith row and jth column of the new matrix, where n matrices have been used in the averaging process.

2.3 TRANSITION MATRICES OF FORMATION K

Two adjoining areas X and Y representing part of a sedimentary basin in India have been considered for the present study. Drilling operations in the sedimentary basin have revealed the presence of thick sedimentary rocks of Tertiary-Quarternary age. The Formation K, in the above basin is usually 200 to 300 metres thick and has a lithological composition of sandstones, shales and coals. This formation is widespread in this basin and is sandwiched between two thick sections of shale and has been modeled and studied in the present investigations.

In Area X, this Formation is represented by thick sand (53 percent), shale (26 percent) and coal (21 percent) sequences. In Area Y, the Formation is dominantly shaly (60 percent) with minor sandstones (37 percent) and a number of coal streaks (3 percent). The Formation was deposited under alternating regressive and transgressive marine environments (O.N.G.C., unpublished reports). The shale underlying the Formation K are thought to be the source rocks for oil found in the reservoir rocks of this Formation. In the present study the sequences of lithology in the Area X represents depositional environments near the basin margins and Area Y represents relatively deeper water environments.

Data from sixteen boreholes in Area X and three from
Area Y have been used in this study. Care was taken to select
those boreholes which are situated near or on the seismic
lines to give better correlation. Since the Area Y exhibits
less lateral variations in lithology, fewer wells were consider-
ed adequate to represent this area. The Formation K of these
stratigraphic sections have been used to obtain upward transi-
tion probability matrices by using transitions from one rock
type to another at four metre sampling interval. One represen-
tative case for each of the Areas X and Y is illustrated in
Figures 2.1 and 2.2, respectively.

To set up a transition probability matrix for transi-
tions from one lithology to its successor, observations of the
upward changes are first recorded in the transition frequency
matrix of Figure 2.1. Each box in the matrix gives the total
number of upward transitions from the state denoted by the
row, to the succeeding state, denoted by the column. A total
of seventy transitions measured at fixed vertical intervals
have been observed for the Formation K in this well. The
transition frequency matrix has been converted into a transi-
tion probability matrix, shown in the same Figures 2.1 and 2.2.

The transition matrices calculated are sensitive to
the sampling interval chosen. If relative frequencies are
taken as a measure or probability then a large number of
transitions should be considered, i.e., the sampling interval
should be very small. The sampling interval of 4 metres

STRATIGRAPHIC SECTION
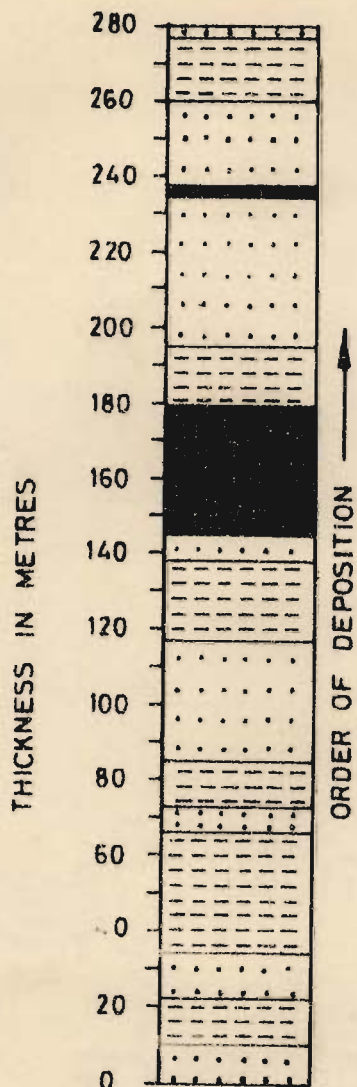
TRANSITION FREQUENCY MATRIX

	S_1	S_2	S_3
S_1	28	5	1
S_2	5	22	0
S_3	0	1	8

TRANSITION PROBABILITY MATRIX

	S_1	S_2	S_3
S_1	.82	.15	.03
S_2	.19	.81	.00
S_3	.00	.11	.89

SANDSTONE S_1

SHALE S_2

COAL S_3

FIG. 2.1_ STRATIGRAPHIC SECTION IN A WELL IN AREA X , WITH ITS UPWARD TRANSITION MATRICES. (SAMPLING INTERVAL IS 4 m.)

STRATIGRAPHIC SECTION
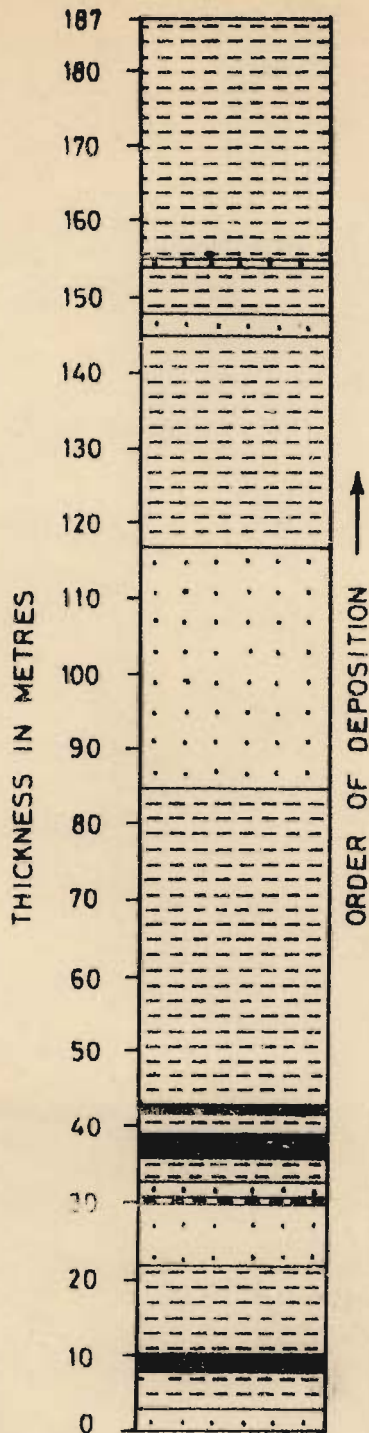
TRANSITION FREQUENCY MATRIX

	S_1	S_2	S_3
S_1	8	5	0
S_2	4	23	2
S_3	0	2	1

TRANSITION PROBABILITY MATRIX

	S_1	S_2	S_3
S_1	.62	.38	.00
S_2	.14	.79	.07
S_3	.00	.67	.33

```
: : :   SANDSTONE   S1

=-=-=   SHALE       S2

█████   COAL        S3
```

FIG.2.2_ STRATIGRAPHIC SECTION IN A WELL IN AREA Y, WITH ITS UPWARD TRANSITION MATRICES. (SAMPLING INTERVAL IS 4 m.)

thickness chosen in the present study was guided by the facts
that (i) the results are to be used for seismic stratigraphic
studies and (ii) the resolution of present day seismic methods
is much less than the above thickness. It may be mentioned
that in the present analysis beds with thickness of less than
4 metres are sometimes either missed or read as 4 metres thick,
because of the restrictions imposed by a fixed sampling inter-
val of 4 metres. Transition matrices thus calculated for 16
wells of Area X and 3 of Area Y are given in Tables 2.2 and
2.3.

Each fractional element in the matrix gives the
probability of upward transition from state s_i to state s_j
where i = 1,2,3 and j=1,2,3 (Table 2.2). The first row in
each of the matrices indicates the transition probabilities
from sandstone to sandstone, shale or coal. For Well E-1
the fraction 28/34 indicates that a total of 34 transitions
have been observed from sandstone, 28 of which are to sand-
stone itself. From the same row it can be observed that 5
transitions are to shale and 1 is to coal. The total probabi-
lity of all transitions possible from sandstone therefore adds
upto 1. This is true for every row in all the matrices. The
zero element indicates that transition from that particular
state, denoted by the row number, to the succeeding state,
denoted by the column number, is not possible. Interesting
cases are Wells E-2, E-7 and E-14, which indicate a total
absence of coal in the Formation K.

Table 2.2 - Upward Transition matrices for the 16 wells
considered for the study of Area X. s_1, s_2 and s_3
are the three lithological states sandstone, shale
and coal in this order.

Well E - 1

	s_1	s_2	s_3
s_1	28/34	5/34	1/34
s_2	5/27	22/27	0
s_3	0	1/9	8/9

Well E - 2

	s_1	s_2	s_3
s_1	11/15	4/15	0
s_2	3/6	3/6	0
s_3	0	0	0

Well E - 3

	s_1	s_2	s_3
s_1	29/40	7/40	4/40
s_2	6/19	12/19	1/19
s_3	5/11	1/11	5/11

Well E - 4

	s_1	s_2	s_3
s_1	17/24	5/24	2/24
s_2	4/23	18/23	1/23
s_3	2/4	1/4	1/4

Well E - 5

	s_1	s_2	s_3
s_1	18/26	7/26	1/26
s_2	5/12	5/12	2/12
s_3	2/6	1/6	3/6

Well E - 6

	s_1	s_2	s_3
s_1	24/31	2/31	5/31
s_2	1/2	1/2	0
s_3	4/9	0	5/9

Well E - 7

	s_1	s_2	s_3
s_1	3/5	2/5	0
s_2	2/7	5/7	0
s_3	0	0	0

Well E - 8

	s_1	s_2	s_3
s_1	42/51	5/51	4/51
s_2	5/13	8/13	0
s_3	3/8	1/8	4/8

Well E - 9

	s_1	s_2	s_3
s_1	37/45	3/45	5/45
s_2	2/8	3/8	3/8
s_3	6/23	2/23	15/23

Well E - 10

	s_1	s_2	s_3
s_1	9/17	7/17	1/17
s_2	7/38	30/38	1/38
s_3	1/5	1/5	3/5

Well E - 11

	s_1	s_2	s_3
s_1	19/29	8/29	2/29
s_2	5/26	20/26	1/26
s_3	3/11	●	8/11

Well E - 12

	s_1	s_2	s_3
s_1	21/29	5/29	3/29
s_2	4/15	10/15	1/15
s_3	3/6	1/6	2/6

Well E - 13

$$
\begin{array}{c@{\qquad}ccc}
 & s_1 & s_2 & s_3 \\
\begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} &
\left[\begin{array}{ccc}
27/35 & 5/35 & 3/35 \\
5/8 & 1/8 & 2/8 \\
3/9 & 2/9 & 4/9
\end{array}\right]
\end{array}
$$

Well E - 14

$$
\begin{array}{c@{\qquad}ccc}
 & s_1 & s_2 & s_3 \\
\begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} &
\left[\begin{array}{ccc}
13/18 & 5/18 & 0 \\
5/7 & 2/7 & 0 \\
0 & 0 & 0
\end{array}\right]
\end{array}
$$

Well E - 15

$$
\begin{array}{c@{\qquad}ccc}
 & s_1 & s_2 & s_3 \\
\begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} &
\left[\begin{array}{ccc}
26/38 & 8/38 & 4/38 \\
7/14 & 5/14 & 2/14 \\
4/9 & 2/9 & 3/9
\end{array}\right]
\end{array}
$$

Well E - 16

$$
\begin{array}{c@{\qquad}ccc}
 & s_1 & s_2 & s_3 \\
\begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} &
\left[\begin{array}{ccc}
20/32 & 11/32 & 1/32 \\
10/36 & 23/36 & 3/36 \\
1/10 & 3/10 & 6/10
\end{array}\right]
\end{array}
$$

Table 2.3 - Upward transition matrices for 3 wells of Area Y.
s_1, s_2 and s_3 are the three lithological states
sandstone, shale and coal, in this order.

Well F - 1

	s_1	s_2	s_3
s_1	9/12	3/12	0
s_2	2/23	20/23	1/23
s_3	0	1/1	0

Well F - 2

	s_1	s_2	s_3
s_1	8/13	5/13	0
s_2	4/29	23/29	2/29
s_3	0	2/3	1/3

Well F - 3

	s_1	s_2	s_3
s_1	15/22	7/22	0
s_2	6/22	15/22	1/22
s_3	0	1/1	0

Table 2.4 – <u>The average matrix characterizing area X</u>

	s_1	s_2	s_3
s_1	0.75	0.17	0.08
s_2	0.25	0.70	0.05
s_3	0.29	0.12	0.59

Table 2.5 – <u>The average matrix characterizing area Y</u>

	s_1	s_2	s_3
s_1	0.68	0.32	0.00
s_2	0.16	0.78	0.06
s_3	0.00	0.73	0.27

Sixteen wells of Area X and three of Area Y have been used to give two average matrices each characterizing a different environment, these are given in Tables 2.4 and 2.5, respectively. The matrix in Table 2.4 characterizes depositional environments near the basin margins while the matrix in Table 2.5 represents relatively deep water environments. It has been already mentioned that Area Y contains very small quantities of coal in the form of thin streaks. These may be missed depending on the sampling interval. Tables 2.3 and 2.5 clearly illustrate the paucity of coal in Area Y as coal has 0.00 probability of succeeding sandstone, 0.06 of succeeding shale and only 0.27 of succeeding itself. It is mostly followed by shale.

2.4 TESTING FOR THE MARKOV PROPERTY

While studying Markov models it is relevant to check whether the process under study actually has the Markov property. For this, a test to distinguish between the two alternative hypotheses, that either the successive events are independent of each other (the null hypothesis) or the events are not independent, is performed. If not independent, they could form a first-order Markov chain. The test statistic λ is given by

$$\lambda = \prod_{i,j}^{m} \left(\frac{p_j}{p_{ij}} \right)^{n_{ij}}.$$

Now,

$$-2 \log_e \lambda = 2 \sum_{i,j}^{m} n_{ij} \log_e (p_{ij}/p_j);$$

which is distributed as χ^2 with $(m-1)^2$ degrees of freedom (Anderson and Goodman, 1957),

where, p_{ij} = probability in box i,j of the transition probability matrix,

p_j = marginal probabilities for the jth column

$$(= \sum_{i}^{m} n_{ij} / \sum_{i,j}^{m} n_{ij}),$$

n_{ij} = transition frequency total in box i,j of the original frequency matrix of observed transitions,

m = total number of states.

This test is illustrated for the average matrix of Area Y. The average tally matrix of Area Y is given by

	s_1	s_2	s_3	Totals
s_1	542	255	0	797
s_2	294	1457	103	1854
s_3	0	8	3	11
Totals	836	1720	106	2662

The values of p_j are calculated by taking the marginal total for the jth column in the tally matrix and dividing it by the overall total. For the three columns,

$p_1 = 836/2662 \quad = 0.31$

$p_2 = 1720/2662 \quad = 0.65$

$p_3 = 106/2662 \quad = 0.04$

$$-2 \log_e \lambda = 2 \left[542 \log_e \frac{0.68}{0.31} + 255 \log_e \frac{0.32}{0.65} + 0 + \right.$$

$$+ 294 \log_e \frac{0.16}{0.31} + 1457 \log_e \frac{0.79}{0.65} + 103 \log_e \frac{0.05}{0.04}$$

$$\left. + 0 + 1720 \log_e \frac{0.73}{0.65} + 3 \log_e \frac{0.23}{0.04} \right]$$

$$= 2 \left[542 \log_e 2.19 + 255 \log_e 0.49 + 294 \log_e 0.52 + \right.$$

$$+ 1457 \log_e 1.21 + 103 \log_e 1.25 + 1720 \log_e 1.12$$

$$\left. + 3 \log_e 5.75 \right]$$

$$= 2 \left[424.87 - 181.90 - 192.25 + 277.73 + 22.98 \right.$$

$$\left. + 19.49 + 4.75 \right]$$

$$= 2 \times 375.67 = 751.34$$

The number of degrees of freedom, $(m-1)^2$ is $(3-1)^2 = 4$. If the level of significance $\alpha = 0.05$ is considered, then the table of values of the χ^2 distribution (Fisher and Yates, 1963) gives the corresponding value of χ^2 as 9.49. The calculated value of $-2 \log_e \lambda$ is 751.34 which is much greater than the tabulated value of χ^2. Therefore the null hypothesis that these transitions are from an independent events process can be rejected and the alternative hypothesis that the transitions have the Markov property can be accepted.

2.5 MARKOV MODELS E AND F

Average matrices characterizing areas X and Y have been used to generate synthetic sequences comparable to the original sequences of these areas. The models generated from these

average matrices for areas X and Y are designated as Markov Models E and F, respectively. A computer programme to generate such sequences from transition probability matrices as given by Harbaugh and Carter (1970) has been used in the present study. Cumulative transition matrices have been computed by adding successively each element of the row to the next element on the right so that the extreme right element attains a value 1.0. The transition probability matrix values of the average matrix for areas X and Y are shown in cumulative form in Tables 2.6 and 2.7 respectively. Using the transition probability matrix the programme generates a sequence of stratigraphic states. The initial state is generated at random, giving each of the states an equal probability of being selected. Pseudo-random numbers are generated in the range 0.0 to 1.0. Following Harbaugh and Carter (1970), the first number is transformed so that it lies within a range of integers extending from 1 to 3, which is the total number of possible states considered for this study. The resulting integer selected in this range provides the starting state. From then on the programme generates subsequent states by sampling the cumulative probability matrix. To select the state at a certain instant the row of probability values pertaining to the state chosen at the immediately preceeding instant is sampled. This is accomplished by generating a random number between 0.00 and 1.00 and progressively comparing it with each element of the appropriate row of the matrix, starting with the lowest value, in the left most column.

Table 2.6 - <u>Cumulative Transition Matrix for area X</u>

$$
\begin{array}{c c c c}
 & s_1 & s_2 & s_3 \\
s_1 & \begin{bmatrix} 0.75 & 0.92 & 1.00 \\ 0.25 & 0.95 & 1.00 \\ 0.29 & 0.41 & 1.00 \end{bmatrix} \\
s_2 & & \\
s_3 & &
\end{array}
$$

Table 2.7 - <u>Cumulative Transition Matrix for area Y</u>

$$
\begin{array}{c c c c}
 & s_1 & s_2 & s_3 \\
s_1 & \begin{bmatrix} 0.68 & 1.00 & 1.00 \\ 0.16 & 0.94 & 1.00 \\ 0.00 & 0.73 & 1.00 \end{bmatrix} \\
s_2 & & \\
s_3 & &
\end{array}
$$

Comparison of the numbers continues until it is found to be equal to or greater than the random number. The column containing that number identifies the next state. The process is illustrated in Figure 2.3.

The synthetic stratigraphic column shown in Figure 2.4 is based on the transition probability values of Table 2.6. Since the transition matrix was based on a 4 metre interval sampling of the well data, each state generated in the synthetic sequence also corresponds to 4 metre thickness. However, for reasons explained later (in Chapter III) these thicknesses have been slightly modified to fit the equal travel time criterion required for generating the synthetic seismogram. The seismic velocity in sandstone is taken as 2362 metres/second and a two way travel time of 4 milliseconds required that the sandstone thickness should be 4.72 metres, therefore each sandstone layer has 0.72 metres added to it. An appropriate addition or subtraction is made in all other lithounits according to the velocities given in Table 2.8, which are calculated from geophysical well log data.

The thickness of the model is governed by two factors, the thickness of the K Formation which can at places be as thick as 300 m, and the wavelength of the source pulse. Since the velocities of the lithounits are approximately 2000 m/sec and the duration of the source pulse is 44 msec, therefore the wavelength of the pulse is around 88 m. If a pair of

First draw   Second draw

Uniform random numbers in
the range 0.0 to 1.0

0.0   0.25   0.50   0.75   1.0

Sandstone to

Cumulative probability vector
for sandstone in the range
0.0 to 1.0

0.0   0.75   0.92 1.0

SAND STONE   SHALE   COAL
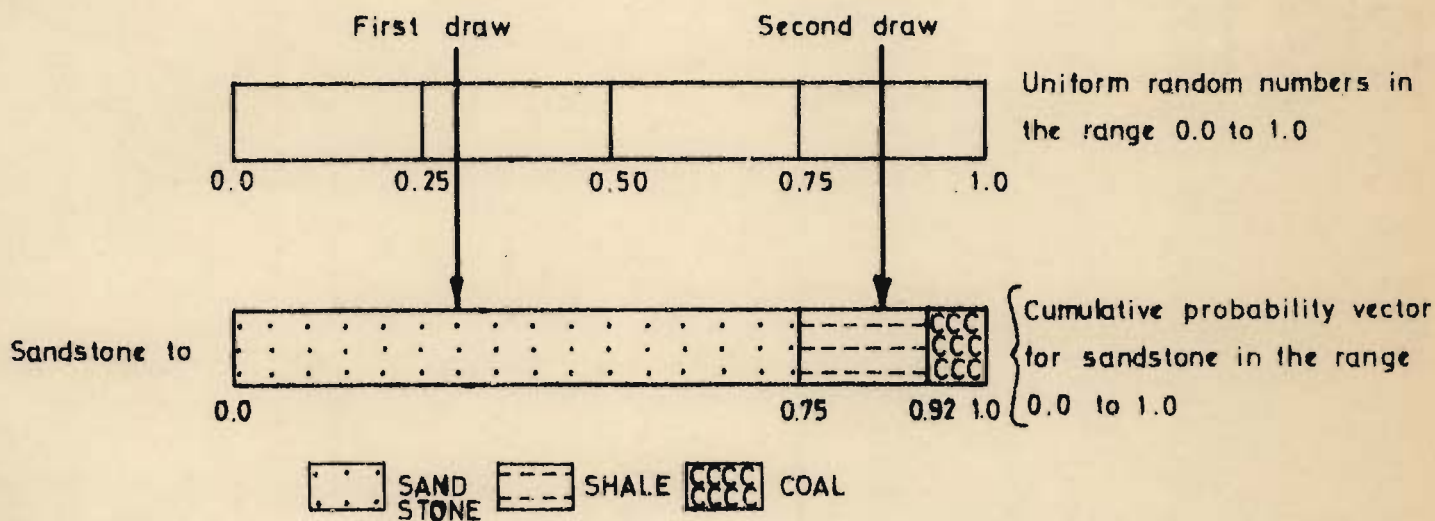
FIG.2.3_ ILLUSTRATION OF THE USE OF UNIFORM RANDOM NUMBER SOURCE
TO SAMPLE FROM CUMULATIVE PROBABILITY VECTOR. FOR EXAMPLE,
ON FIRST DRAW RANDOM NUMBER IS 0.29, WHICH IS LESS THAN
0.75, RESULTING IN THE SELECTION OF SANDSTONE TO SUCCEED
SANDSTONE. ON SECOND DRAW RANDOM NUMBER IS 0.86, WHICH IS
LESS THAN 0.92 BUT GREATER THAN 0.75, RESULTING IN THE
SELECTION OF SHALE TO SUCCEED SANDSTONE.

FIG.2.4_ GRAPHIC DISPLAY OF A SYNTHETIC STRATIGRAPHIC
SEQUENCE EMPLOYING TRANSITION PROBABILITY
VALUES IN TABLE 2.6 (THICK LINES ADDED)

Table 2.8 - <u>Velocities and Densities used for Markov Models</u>
<u>E and F</u>

	Constituents of the models	Velocity m/sec	Density g /c.c.
Markov Model E for Area X	Overburden	1400	2.00
	Sandstone	2875	2.21
	Shale	2629	2.46
	Coal	1998	1.41
	Lower Half Space	2400	2.50
Markov Model F for Area Y	Overburden	1400	2.00
	Sandstone	2362	2.29
	Shale	2192	2.39
	Coal	1929	1.46
	Lower Half Space	2400	2.50

reflecting surfaces are at this or greater distance apart they can be easily resolved on a reflection record. If on the other hand, the surfaces are seperated by less than a wavelength, the resolution becomes difficult as the seperation decreases (Widess, 1973). Therefore the model thickness is kept at about 200 m and it is sandwiched between homogeneous strata to study its effect in isolation. Such situations are often met in sedimentary basins.

Two hundred and fifty five synthetic stratigraphic sequences characterizing Area X, have been simulated by using the cumulative transition matrix given in Table 2.6. Three of these are shown in Figure 2.5. Another two hundred and fifty three stratigraphic sequences characterizing Area Y are generated by using the matrix in Table 2.7 and three of these are shown in Figure 2.6.
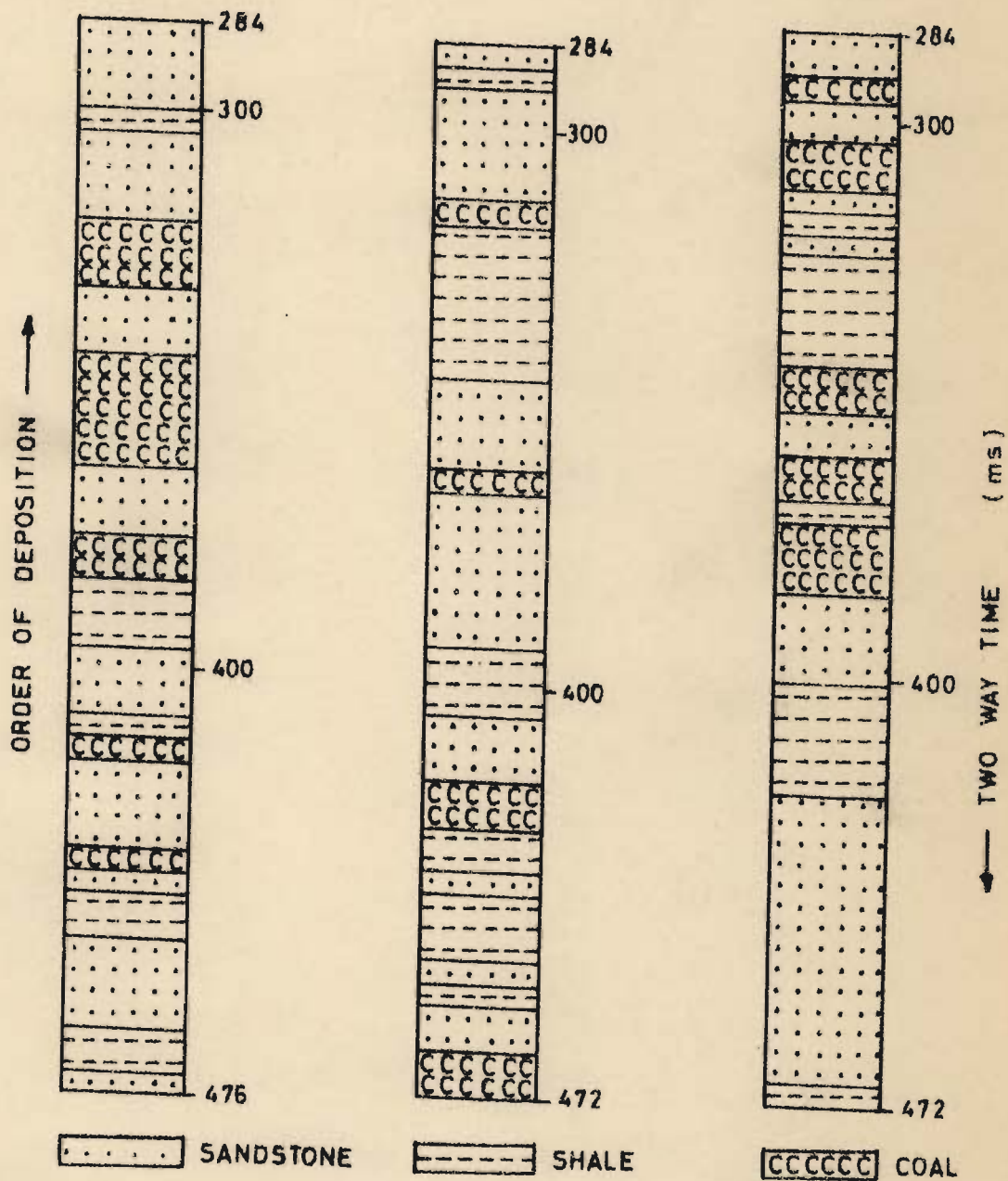
FIG.2.5_ SOME SYNTHETIC STRATIGRAPHIC SECTIONS FOR
MODEL E. THICKNESS IS SHOWN IN TERMS OF
TWO WAY TRAVEL TIME.
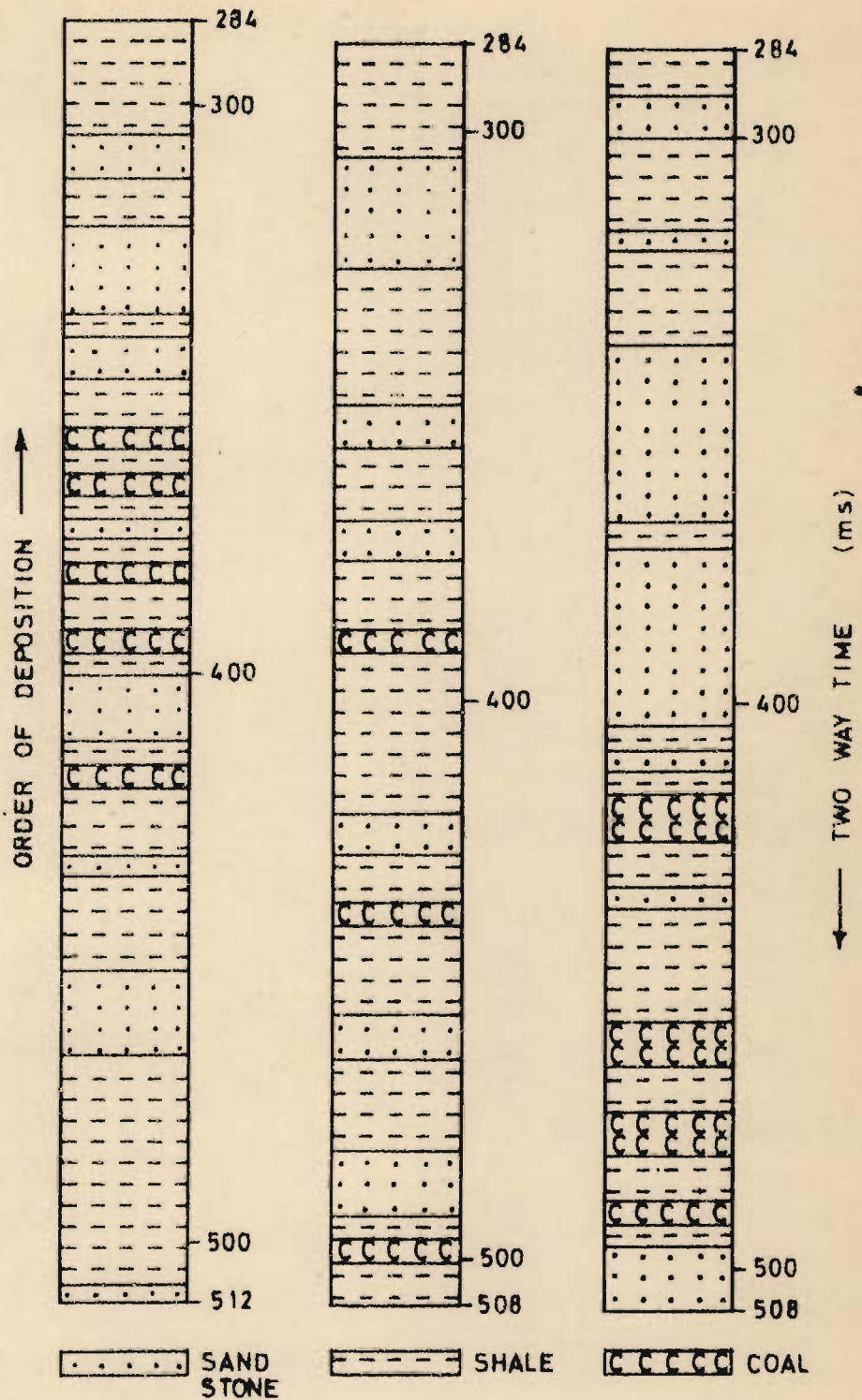
FIG. 2.6 _ SOME SYNTHETIC STRATIGRAPHIC
SECTIONS FOR MODEL F. THICKNESS IS
SHOWN IN TERMS OF TWO WAY TRAVEL
TIME .

CHAPTER - III

SYNTHETIC SEISMOGRAMS AND THEIR
ANALYSIS

Synthetic seismograms were first described by Peterson, Filloppone and Coker (1955). They used artificial reflection records made from velocity logs. These logs were converted in depth to a reflectivity function in time, which was convolved with a presumed source wavelet. Synthetic seismograms have since then assumed considerable importance in seismic exploration.

Wuenschel (1960) introduced a frequency domain approach for calculating synthetic seismograms for normal incidence. Treitel and Robinson (1966) & Claerbout (1968, 1976) have calculated in the time domain the impulse response generated by a source at the surface of a horizontally layered earth, assuming plane waves at normal incidence.

3.1 SYNTHESIS OF A LAYERED MEDIUM FROM ITS ACOUSTIC TRANSMISSION RESPONSE

The seismic response of synthetic stratigraphic columns described in Chapter II can be calculated if the reflection and transmission coefficients at each interface are known. The normal incidence reflection and transmission coefficients r_k and t_k at the interface between the kth and $(k+1)^{th}$ layer (see Figure 3.1) are given by
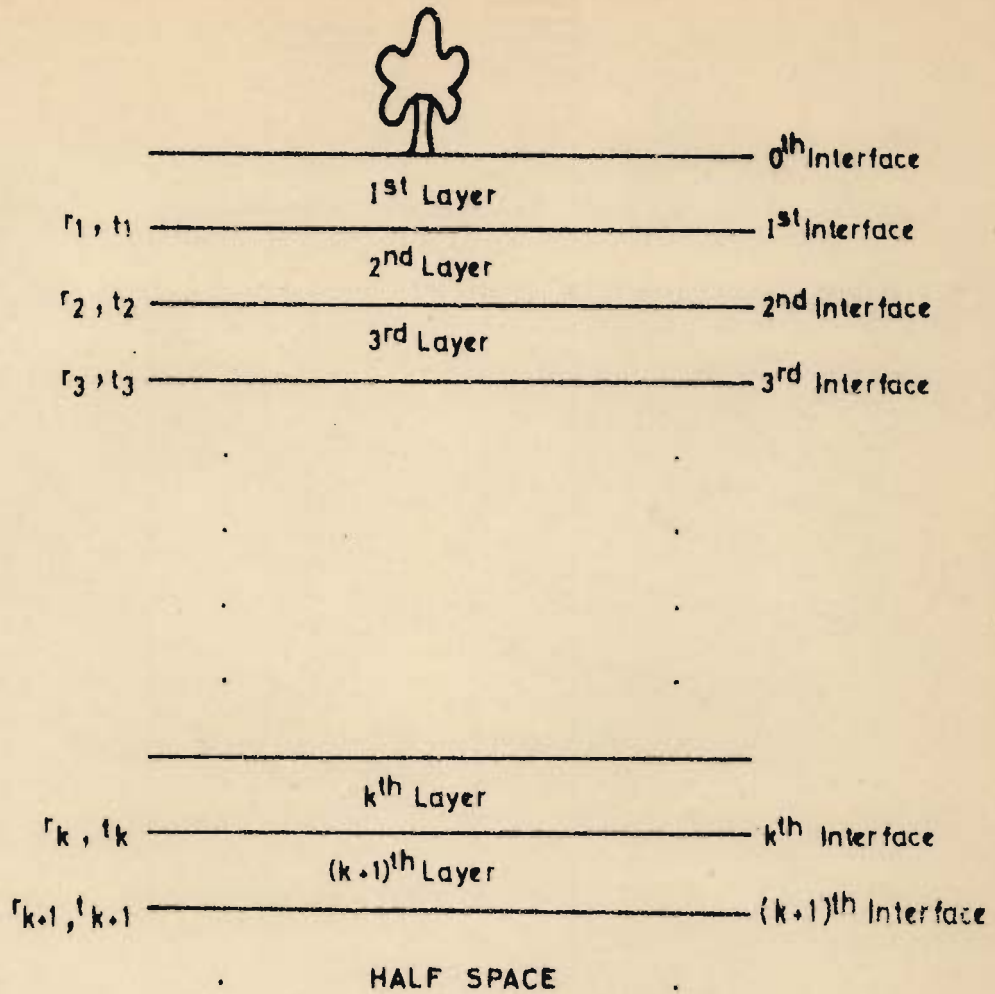
FIG.3.1_ LAYER AND INTERFACE GEOMETRY WITH
REFLECTION AND TRANSMISSION COEFFICIENTS.

$$r_k = \frac{P_k V_k - P_{k+1} V_{k+1}}{P_k V_k + P_{k+1} V_{k+1}}$$

and

$$t_k = \frac{2 P_k V_k}{P_{k+1} V_{k+1} + P_k V_k}$$

Where P_k and P_{k+1} are the densities of the kth and $(k+1)^{th}$ layer respectively, and V_k and V_{k+1} are the seismic velocities of the kth and $(k+1)^{th}$ layer.

Assuming equally spaced interfaces in time the reflection and transmission coefficients are calculated for each interface. Their impulse response can then be calculated using the following method given by Claerbout (1968, 1976).

3.1.1 Impulse Response

Consider a horizontally layered medium in which each layer is homogeneous, isotropic and perfectly elastic. The half-space underlying the layered medium is taken as homogeneous and the top is a free surface. Let the stratification be excited by a downgoing impulsive source at time $t = 0$, which produces plane waves normal to the stratification. Let the reflection and transmission coefficients for the interface between the kth and $(k+1)^{th}$ layer be r_k and t_k respectively. When the ray is travelling from k^{th} layer to $(k+1)^{th}$ layer the transmission and reflection coefficients are denoted by t'_k and r'_k and when the ray is travelling from the $(k+1)^{th}$

layer to k^{th} layer by t_k and r_k respectively. It is hence implied that :

$$t_k = 1 + r_k$$

and

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ldots(3.1)$$

$$r_k = -r'_k$$

In Figure 3.2 the rays are drawn with time displacement along the horizontal axis to make them appear as at an incident and reflected angle of 45°. Lines sloping downward or upward, as indicated by arrows represent downgoing and upcoming waves respectively. When the downgoing ray D' is incident on the interface it is partially reflected as U' and the remaining is transmitted in the next layer as D. The same is true for the upcoming ray U, which is reflected and transmitted as D and U' respectively. In Figure 3.2 the primed and the un-primed layer refer to the kth and $(k+1)^{th}$ layer respectively. The waves U and D' can be extrapolated into the future to get the waves U' and D as given below :

$$U' = t_k U + r'_k D'$$

$$D = r_k U + t'_k D'$$

These equations can therefore be put in the following matrix form :

$$\begin{bmatrix} U' \\ \\ D \end{bmatrix} = \begin{bmatrix} t_k & r'_k \\ \\ r_k & t'_k \end{bmatrix} \begin{bmatrix} U \\ \\ D' \end{bmatrix} \qquad \ldots (3.2)$$
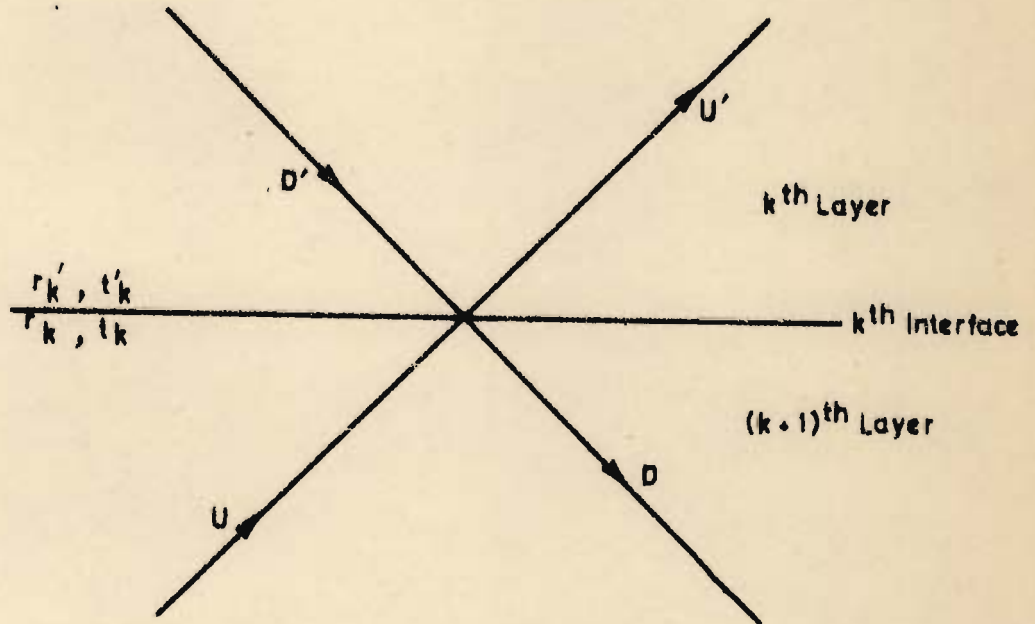
FIG. 3. 2 _ RAYS INCIDENT AND REFLECTED AT A
POINT ON THE kth INTERFACE.

Equations (3.1) and (3.2) can be combined to get the following relation between the primed k^{th} and the unprimed $(k+1)^{th}$ layer

$$\begin{bmatrix} U \\ \\ D \end{bmatrix}' = \frac{1}{t'_k} \begin{bmatrix} 1 & r'_k \\ \\ r'_k & 1 \end{bmatrix} \begin{bmatrix} U \\ \\ D \end{bmatrix} \qquad \dots (3.3)$$

Let $z = W^2 = e^{i\omega T}$, where T, the two way travel time equals the sampling interval of the seismogram. Therefore, multiplication by z^k is equivalent to delaying the function by kT.

The downgoing ray in the upper part of the kth layer is denoted by D and it reaches the kth interface after a time W (Figure 3.3), therefore :

$$D_k = D'_k W^{-1}$$

and

$$U_k = U'_k W.$$

Combining these two for the kth layer the following matrix equation is obtained :

$$\begin{bmatrix} U \\ \\ D \end{bmatrix}_k = \begin{bmatrix} W & 0 \\ \\ 0 & 1/W \end{bmatrix} \begin{bmatrix} U \\ \\ D \end{bmatrix}'_k \qquad \dots (3.4)$$
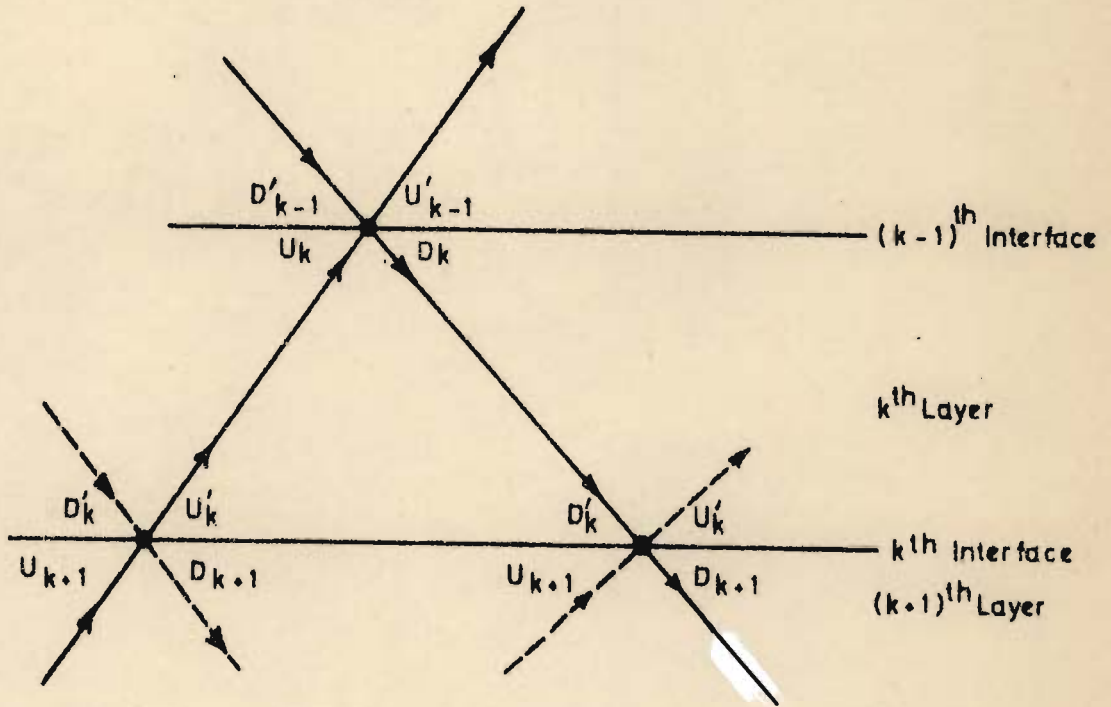
FIG.3.3_ PRIMED AND UNPRIMED NATURE OF RAYS
FOR THE kth LAYER.

where suffix k stands for the kth layer. Combining equations (3.3) and (3.4), relation between displacement at top of the $(k+1)^{th}$ layer with the displacement at top of the kth layer can be obtained:

$$\begin{bmatrix} U \\ \\ D \end{bmatrix}_k = \frac{1}{W\,t_k} \begin{bmatrix} z & zr_k \\ \\ r_k & 1 \end{bmatrix} \begin{bmatrix} U \\ \\ D \end{bmatrix}_{k+1} \quad \dots (3.5)$$

The determinant of the coefficient matrix in equation (3.5) is

$$\begin{vmatrix} \dfrac{z}{t_k W} & \dfrac{zr_k}{t_k W} \\ \\ \dfrac{r_k}{t_k W} & \dfrac{1}{t_k W} \end{vmatrix} = (1-r^2_k)/t^2_k = (1-r_k)/(1+r_k) \quad \dots (3.6)$$

Considering the case of three layers the relationship linking the displacements at top of the first layer with that at top of the half space works out to be

$$\begin{bmatrix} U \\ \\ D \end{bmatrix}_1 = \frac{1}{t_1 t_2 W^2} \begin{bmatrix} z & z\,r_1 \\ \\ r_1 & 1 \end{bmatrix} \begin{bmatrix} z & z\,r_2 \\ \\ r_2 & 1 \end{bmatrix} \begin{bmatrix} U \\ \\ D \end{bmatrix}_3$$

$$\begin{bmatrix} U \\ \\ D \end{bmatrix}_1 = \frac{1}{t_1 t_2 z} \begin{bmatrix} z^2\{1+ r_1 r_2(1/z)\} & z^2\{r_2 + r_1(1/z)\} \\ \\ (r_2 + z.r_1) & (1+z\,r_1 r_2) \end{bmatrix} \begin{bmatrix} U \\ \\ D \end{bmatrix}_3$$

The relation for $(k+1)$ layered medium is given by

$$
\begin{bmatrix} U \\ \\ D \end{bmatrix}_1 = \frac{1}{t_1 t_2 \cdots t_k \ W^k} \begin{bmatrix} z & z\, r_1 \\ \\ r_1 & 1 \end{bmatrix} \begin{bmatrix} z & z r_2 \\ \\ r_2 & 1 \end{bmatrix} \begin{bmatrix} z & z r_k \\ \\ r_k & 1 \end{bmatrix} \begin{bmatrix} U \\ \\ D \end{bmatrix}_{k+1} \qquad (3.6a)
$$

The product of the k matrices is given by

$$
\frac{1}{W^k} \begin{bmatrix} z^k F(1/z) & z^k G(1/z) \\ \\ G(z) & F(z) \end{bmatrix} \qquad \cdots (3.7)
$$

where, $F(z) \overset{k}{\underset{i=1}{\pi}} t_i = 1 + F_1 z + F_2 z^2 + \cdots + F_{k-1} z^{k-1} \qquad \cdots (3.8)$

and $G(z) \overset{k}{\underset{i=1}{\pi}} t_i = r_k + G_1 z + G_2 z^2 + \cdots + G_{k-1} z^{k-1} \qquad \cdots (3.9)$

For a $(k+1)$ layer model the determinant corresponding to (3.6) is given by

$$
\pi \ \det = \overset{k}{\underset{i=1}{\pi}} \ \frac{1 - r_i}{1 + r_i}
$$

$$
= F(z) \ F(1/z) - G(z) \ G(1/z) \qquad \cdots (3.10)
$$

Formula (3.10) says that there are two spectra $F(z) F(1/z)$ and $G(z)G(1/z)$ whose difference is positive as well as frequency independent.

The boundary condition at the surface is that the upcoming wave is the reflection seismogram

$$R(z) = z R_1 + z^2 R_2 + \cdots$$

i.e., $U = R(z)$

Where R_1, R_2, $\cdots$, are the impulses representing the reflected energy. The downgoing wave is the impulsive source $R_0 = 1$ plus the reflection $R(z)$ from the free surface,

$$D = 1 + R(z).$$

Therefore, the boundary conditions at the surface can be represented by

$$\begin{bmatrix} U \\ D \end{bmatrix}_1 = \begin{bmatrix} R \\ 1+R \end{bmatrix} \qquad \cdots (3.10a)$$

Boundary conditions at the half-space underlying the layers is that the upcoming wave is zero and the downgoing wave is some unknown function $T(z)$, i.e.,

$$\begin{bmatrix} U \\ D \end{bmatrix}_{k+1} = \begin{bmatrix} 0 \\ T \end{bmatrix} \qquad \cdots (3.10b)$$

Using equations (3.10a) , (3.10b) and (3.7), equation (3.6a) transforms to:

$$\begin{bmatrix} R \\ 1+R \end{bmatrix} = \frac{1}{W^k} \begin{bmatrix} z^k F(1/z) & z^k G(1/z) \\ G(z) & F(z) \end{bmatrix} \begin{bmatrix} 0 \\ T \end{bmatrix} \qquad \cdots (3.11)$$

On subtracting the first equation in (3.11) from the second, we get :

$$\begin{bmatrix} R \\ 1 \end{bmatrix} = \frac{1}{W^k} \begin{bmatrix} z^k F(1/z) & z^k G(1/z) \\ G(z)-z^k F(1/z) & F(z)-z^k G(1/z) \end{bmatrix} \begin{bmatrix} 0 \\ T \end{bmatrix} \qquad \ldots(3.12)$$

From this the transmitted wave is given by

$$T(z) = W^k / \left[F(z) - z^k G(1/z) \right] \qquad \ldots(3.13)$$

Physically, $T(z)$ must have finite energy and is a delayed minimum phase function.

Defining a new quantity $M(z)$ as $M(z) = F(z)-z^k G(1/z)$, equations (3.11) and (3.12) give the following relations -

$$T(z) = W^k/M(z)$$

$$R(z) = W^{-k} z^k G(1/z) T(z)$$

$$= z^k G(1/z)/ M(z)$$

$$1 + R(z) = F(z)/ M(z)$$

$$R(z) M(z) = z^k G(1/z)$$

$$R(1/z)M(1/z) = z^{-k} G(z)$$

Combining some of these yields

$$\left[1 + R(z) + R(1/z) \right] M(1/z)$$

$$= \left[F(z) M(1/z) + z^{-k} G(z) M(z) \right]/ M(z)$$

$$= \left[F(z) \left\{ F(1/z) - z^{-k}G(z) \right\} + \right.$$

$$\left. z^{-k}G(z) \left\{ F(z) - z^{k} G(1/z) \right\} \right] / M(z)$$

$$= \left[F(z) F(1/z) - G(z) G(1/z) \right] / M(z)$$

$$= (\pi \; det) / M(z)$$

$$= (\pi \; det) \; T(z) / W^{k}$$

For a two layer system

$$\begin{bmatrix} R_0 & R_1 & R_2 \\ R_1 & R_0 & R_1 \\ R_2 & R_1 & R_0 \\ R_3 & R_2 & R_1 \\ \cdot \\ \cdot \\ \cdot \\ R_\infty \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} \pi \; det/M_0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad \ldots (3.14)$$

which on dividing by M_0, where $M_0 = 1 / \prod\limits_{i=1}^{k} t_i$

$$= 1 / \prod\limits_{i=1}^{k} (1 + r_i)$$

becomes,

$$\begin{bmatrix} R_0 & R_1 & R_2 \\ R_1 & R_0 & R_1 \\ R_2 & R_1 & R_0 \\ \cdot \\ \cdot \\ \cdot \\ R_\infty \end{bmatrix} \begin{bmatrix} 1 \\ M_1/M_0 \\ -r_2 \end{bmatrix} = \begin{bmatrix} \pi(1-r^2) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad \ldots (3.15)$$

If the system is a two layer system, the fourth and subsequent equations in (3.15) do not over determine the system, rather they show how to calculate the rest of the reflection seismogram (R_3, R_4, ...) once M has been determined from R_0, R_1 and R_2.

Equation (3.15) may be generalised to the case of many layers by making use of the Levinson recursion algorithm (Levinson, 1949). This yields a reflection seismogram for a sequence of layered rocks. Figure 3.4 shows the same stratigraphic sequence as in Figure 2.4, with the reflection coefficient series and the impulse response as calculated by the method described above.

3.1.2. Source Wavelet and Convolution

It is desirable that the impulse response obtained in section 3.1.1 should be made to appear like a conventional seismogram. This is achieved by convolving the impulse response with a source wavelet of 44 ms duration. Convolution in time domain for sampled functions is given by

$$c(k) = \frac{1}{N} \sum_{\tau=0}^{N-1} g(\tau)\, h(k-\tau)$$

where c is the convolved output, the synthetic seismogram, N is the number of samples in impulse response g, τ is the lag and h is the source wavelet, or as used in this study the Ricker wavelet (Figure 3.6).
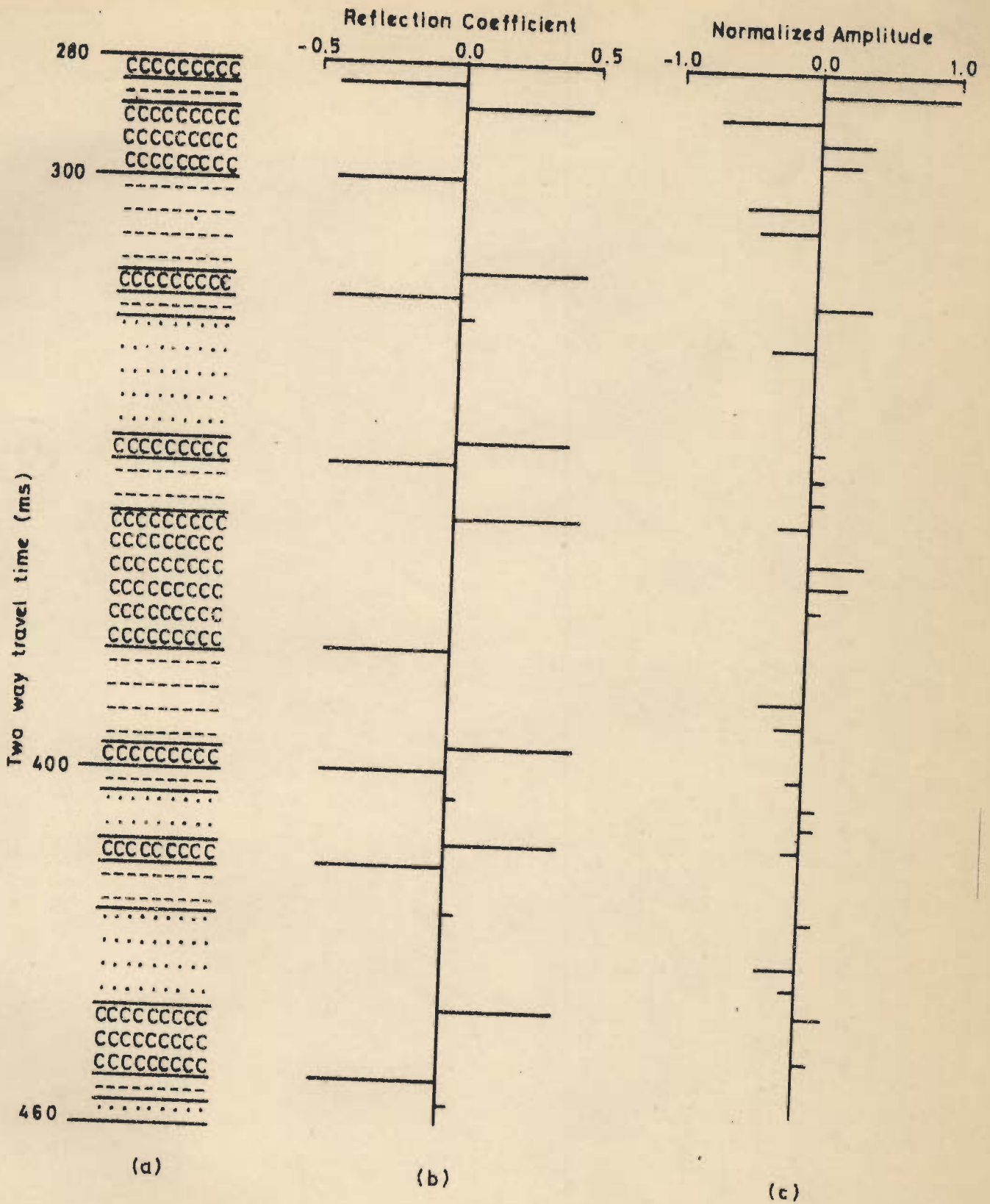
FIG.3.4 _(a)_ SYNTHETIC STRATIGRAPHIC SEQUENCE,

(b)_ REFLECTION COEFFICIENT AT EACH INTERFACE AND

(c)_ IMPULSE RESPONSE.

The source wavelet used in making the synthetic seis-
mograms is usually the Ricker pulse given by Ricker (1953,1978).
Widess (1973) and Dobrin (1976) have shown how reflection
events are modified depending on the relation between the bed
thickness and the wavelength of pulse used. Figure 3.5 shows
how a simple down travelling source pulse, with a waveform
comparable to that of a Ricker wavelet, is altered when it is
reflected from a sequence of boundaries closely spaced in
comparison with the wavelength of the pulse. Each interface
returns a pulse having the same waveform as the source pulse,
but the amplitude and phase (whether or not reversed by a lower
velocity below the boundary) are governed by the reflection
coefficient across it. The resultant of all the individual
reflections is recorded by the geophones placed at the surface.
The difference between the source pulse and the resultant
reflected signal is quite pronounced. However, it is not
possible to isolate the contribution made by any of the indivi-
dual boundaries. It is hence implied that a typical seismic
reflection should be looked upon as an interference pattern
madeup of impulses from many interfaces spread vertically over
hundreds of feet rather than as a simple event originating from
a single lithological interface.

The broad band spectrum of the impulse response would
be modified by the amplitude spectrum of the source wavelet
which shows a maximum at 60 Hz and a bandwidth of 90 Hz
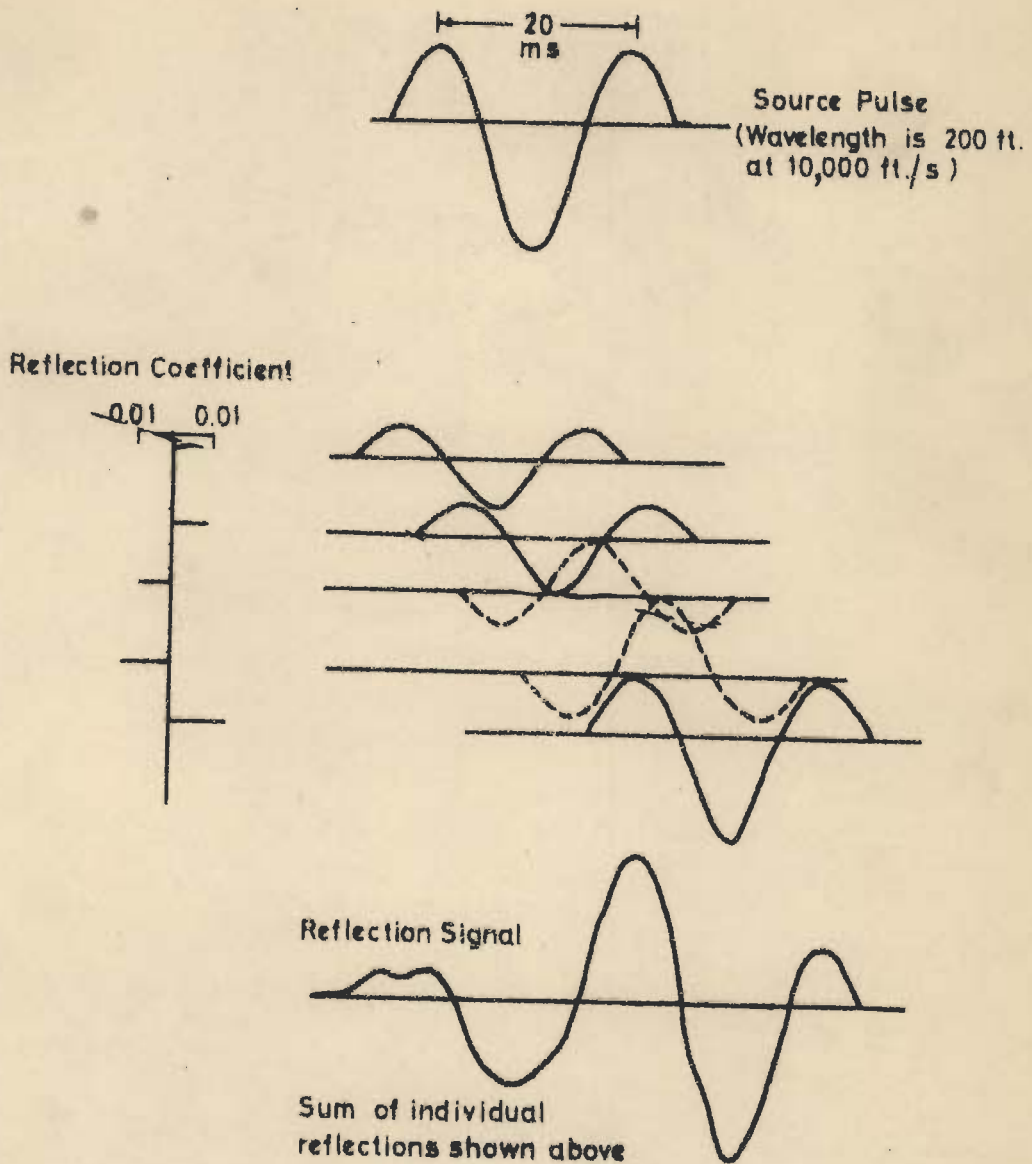
FIG.3.5_ COMPOSITION OF REFLECTIONS FROM SERIES OF FIVE INTERFACES WITH SEPARATIONS SMALL IN COMPARISON TO WAVELENGTH. NOTE CHANGE IN WAVEFORM CAUSED BY REFLECTION PROCESS. WAVEFORMS INDICATED BY DASHED LINES REPRESENT REFLECTIONS WITH PHASE REVERSAL (MODIFIED AFTER DOBRIN , 1976).

(Figure 4.8f). In the present study a total of 508 (255 for Model E and 253 for Model F) impulse responses have been computed, 10 percent noise is superimposed on each, and these are convolved with a 44 ms duration wavelet to produce synthetic seismograms. The 10 percent random noise is taken to account for uncorrelated noise due to wind, microseisms and system components, which are generally encountered in field seismograms. Random numbers in the range -0.5 to + 0.5 were generated and the variance of these numbers was computed to estimate the energy content in the noise. The ratio of signal energy to the noise energy (μ) was taken and each random number generated was subsequently multiplied by 10 percent of $\sqrt{\mu}$ to obtain the desired signal to noise ratio and this was added to the seismogram. Figure 3.6 shows the reflection co-efficient series, the source pulse and the synthetic seismogram constructed for one simulation of Model F. Some more synthetic seismograms for Models E and F are shown in Figures 3.7 and 3.8.

3.2 ESTIMATION OF AUTOCORRELATION FUNCTION

Any signal correlates perfectly with itself. However, if a signal is correlated with a replica of itself displaced by a time-shift τ along the time axis then the amount of correlation will be less. The dependence of correlation on this shift is an important characteristic of the signal (Robinson, 1967). Specifically, the autocorrelation function, A_τ, of a signal is defined as

REFLECTION COEFFICIENT   SYNTHETIC SEISMOGRAM

Normalized amplitude

- 0.35   0.0   0.35

-1.0   1.0

284 ms

284 ms

Time

44 ms

SOURCE WAVELET

480 ms

508 ms
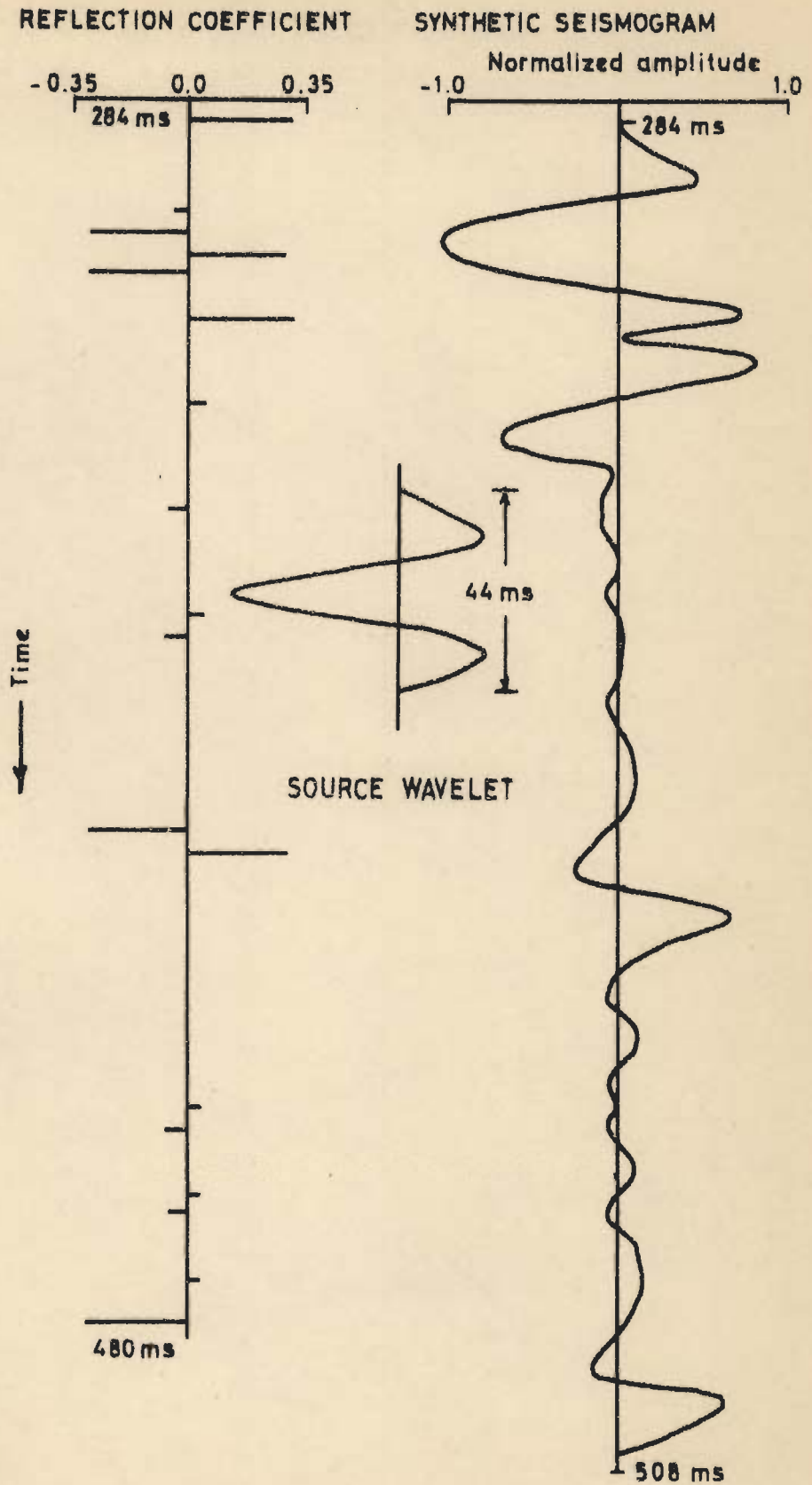
FIG. 3.6   _ REFLECTION COEFFICIENT SERIES AND

ITS SYNTHETIC SEISMOGRAM FOR A

REPRESENTATIVE SIMULATION OF MODEL F.
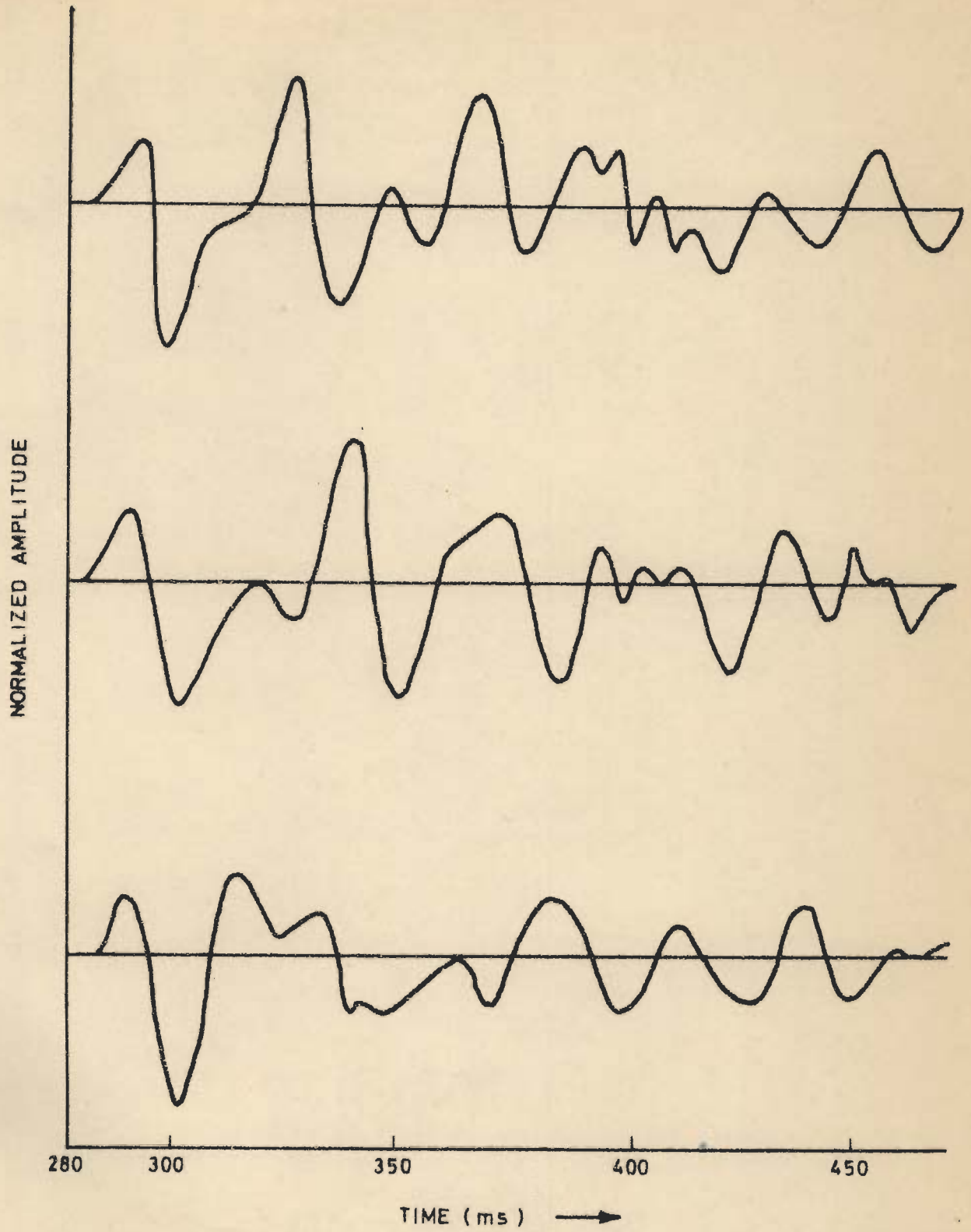
FIG. 3.7 _ SOME SYNTHETIC SEISMOGRAMS FOR MODEL E.
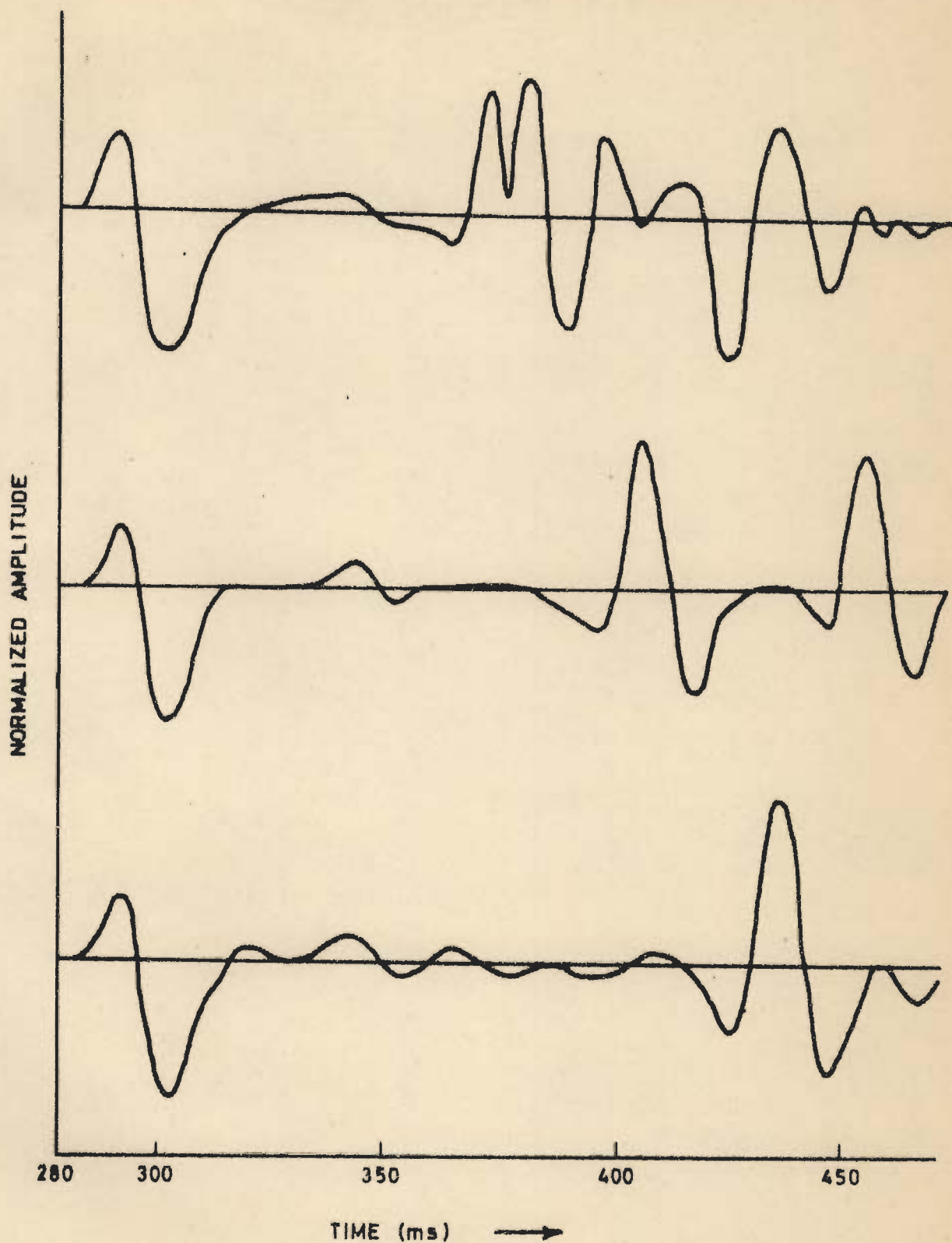
FIG.3.8_SOME SYNTHETIC SEISMOGRAMS FOR MODEL F.

$$A_\tau = \frac{1}{T} \sum_{i=1}^{T} X_i \, X_{i+\tau} \qquad (\tau = 0,1,2, \quad \ldots, \; T-1)$$

The signal $X_{i+\tau}$ represents a replica of the signal X_i advanced by the amount τ. The autocorrelation function is symmetric, i.e., if the replica is shifted to the right or to the left, the result is the same, therefore, it is sufficient to consider only one of these two portions.

The synthetic seismograms give the character of reflections of those sections which have been traversed by the input impulse, or the source wavelet. The appropriate function to study the characteristics of these seismograms is therefore the autocorrelation function (ACF). For constant lag, 4 ms in the present study , if the ACF gives a sharp peak the reflector series is largely uncorrelated, but if the ACF gives a broad based peak then a repetitive element is anticipated. Its rate of decay provides an indication of the frequency bandwidth. For narrow band signals the autocorrelation decays at a slower rate as a function of shift than for broad band signals. The autocorrelation function is the time domain equivalent of the signal power spectrum and is an important analytical tool for random signals.

The autocorrelation function of each of the 508 synthetic seismograms and 387 real seismograms has been computed and some of the representative autocorrelation functions are shown in Figures 4.4-4.7, 4.8(a) and 6.5-6.8.

3.3 ESTIMATION OF POWER SPECTRUM

The study of any signal if carried out only in the time domain will not give the frequency at which a certain character occurs, which may be repeated in a certain pattern. To study the frequency content, the signal is analysed in the frequency domain. The use of fast Fourier transform technique enables efficient computation of Fourier transforms (Bergland, 1969). This is a powerful tool and serves as a bridge between time and frequency domains. It is possible to go back and forth between waveform and spectrum with speed and economy. It helps in finding the periodic components in complex looking signals and their bandwidths. The power spectrum of the synthetic seismogram obtained in Section 3.1 can be studied with the help of this tool.

Conventional methods of estimating power spectra of short time series have certain drawbacks. The periodogram method shows a shift in spectral peaks for truncated sinusoids when the data length is less than 0.58 times the period of the sinusoid (Toman, 1965) and a decrease of resolution when the data length is comparable to the period of the sinusoid (Ulrych, 1972). This method also assumes a periodic extension of the data. The power spectrum $S_p(f)$ is defined as :

$$S_p(f) = \frac{1}{N} \left[\sum_{n=1}^{N} x_n \, h_n \, e^{-2\pi \, ifn} \right]^2$$

The 'data window' h(n) is used to improve the statistical properties of the estimator.

The power spectrum may also be estimated by taking the Fourier transform of the autocorrelation function. It has the drawback that sometimes negative power is indicated if the data length is inadequate. Moreover, estimation of the autocorrelation function unreasonably assumes a zero extension of the time series.

Let x_1, x_2, ..., x_N be the time series under consideration. For a discrete random process of zero mean, the autocorrelation at lag τ is defined as :

$$A(\tau) = E \left[x_n \, x_{n+\tau} \right]$$

$E \left[\, . \, \right]$ denotes the ensemble average of the quantity within the square brackets. Since the time series can be observed for finite time, only an estimate $C(\tau)$ can be obtained using the standard technique. Blackman and Tukey (1958) proposed a power spectral estimator $S_{B-T}(f)$ to overcome the above difficulty in accurate estimation of the autocorrelation function. The power spectrum is given by

$$S_{B-T}(f) = \sum_{\tau=-L}^{+L} C(\tau) \, h(\tau) \, e^{-2\pi i f \tau}$$

where,

$$C(\tau) = \frac{1}{N} \sum_{n=1}^{N-\tau} x_n \, x_{n+\tau}, \quad |\tau| < N$$

The autocorrelation estimate used here is a biased estimate of $A(\tau)$, but it has the computational advantage of a simple scale factor before the summation. As τ approaches N, the accuracy of the estimate $C(\tau)$ decreases because the summation will contain fewer terms. As a result, the $S_{B-T}(f)$ makes use of $C(\tau)$ for values of τ in the range $(-L,L)$, where L is a small fraction of N. The lag window $h(\tau)$ is introduced to obtain the desired statistical properties of the estimator, e.g., stability (Blackman and Tukey, 1958).

Maximum Entropy Method (MEM) originally suggested by Burg (1967) eliminates the necessity of some of these arbitrary assumptions about the data or its autocorrelation function outside the time window. It is particularly useful for short lengths of data sampled at equal intervals. Numerical results published by Ulrych (1972); Ulrych, Smylie, Jensen and Clarke (1973); Chen and Stegen (1974) and Kumar and Mullick (1979) show that the maximum entropy spectral estimator has a better resolving capability.

Ulrych (1972) has shown the superiority of the MEM over the conventional method by using a 1 Hz sinusoid superimposed with 10 percent white noise truncated with a 1 second window (Figure 3.9). Figure 3.10 shows part of a seismogram for Area X and its power spectrum using square of the modulus of the Fourier transform and MEM.

(a)

Time (sec)

(b)

(c)
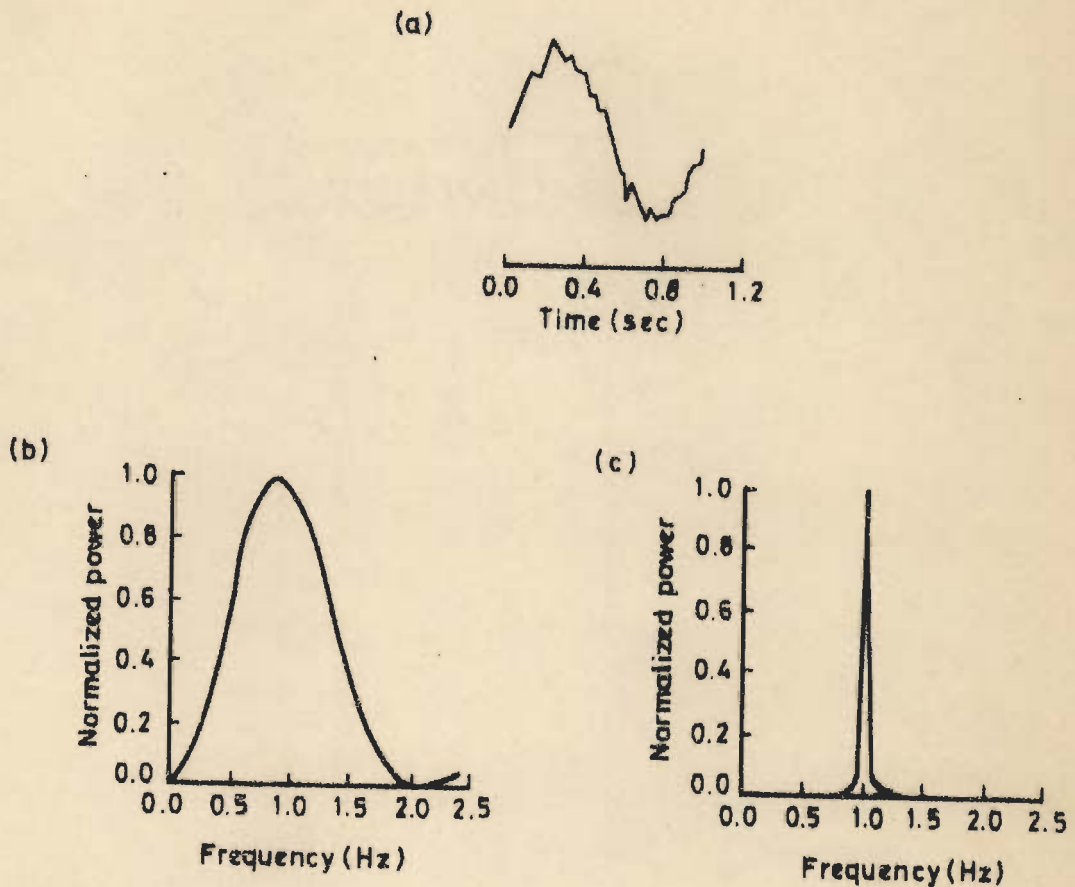
FIG. 3. 9 (a)_ A 1 Hz SINUSOID WITH 10% WHITE NOISE
TRUNCATED WITH A 1 SECOND WINDOW,
(b)_ THE POWER SPECTRUM OF THE SIGNAL IN (a)
COMPUTED BY USING THE SQUARE OF THE
MODULUS OF THE FOURIER TRANSFORM AND
(c)_ THE MAXIMUM ENTROPY POWER SPECTRUM OF
THE SIGNAL IN (a).

(FROM ULRYCH, 1972)
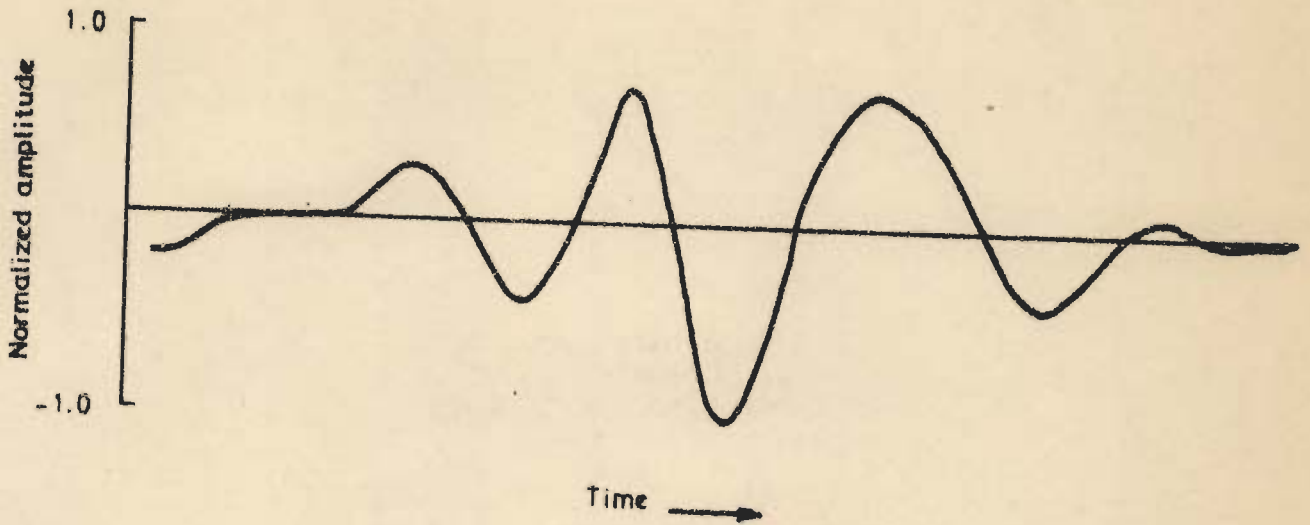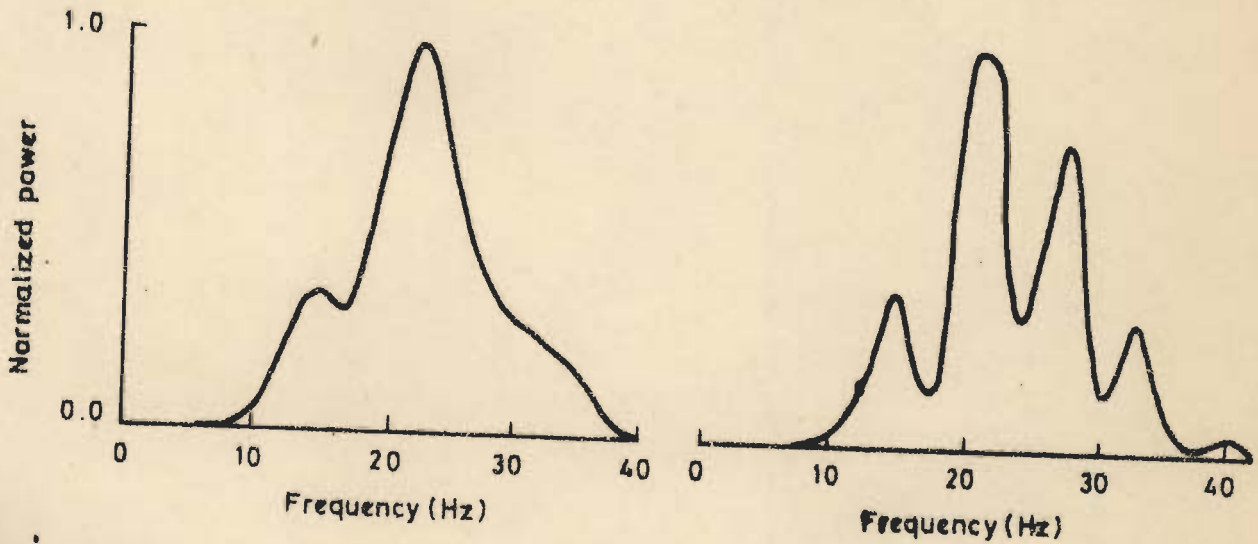
FIG.3.10(a)_ A SEISMIC DATA SERIES FOR AREA **X**,

   (b)_ POWER SPECTRUM OF THE SIGNAL IN (a) COMPUTED BY

        USING SQUARE OF THE MODULUS OF THE FOURIER

        TRANSFORM AND

   (c)_ THE MAXIMUM ENTROPY POWER SPECTRUM OF

        SIGNAL IN (a).

The time series available over (-T, T) can be viewed as the multiplication of one complete sample realization of the process by a rectangular function extending from -T to T. In the frequency domain, this results in the convolution of the true spectrum with a sinc type of function. The width of the major lobe of this sinc function is 1/2 T. Thus any sharp peaks in the spectrum will be broadened leading to loss of resolution. As the record length T increases, the major lobe width decreases and resolution improves. Thus, in conventional spectral estimation methods, the resolution is of the order of 1/T where T is the length of the record. The maximum entropy power spectrum can be estimated as given in Section 3.3.1.

3.3.1. Maximum Entropy Power Spectrum

Consider a random signal as input to a linear system where the infinite sequence of output is given by $(\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots)$ of which the finite segment $(x_0, x_1, \ldots, x_{n-1})$ consisting of n points has been observed.

The autocorrelation function (or the power spectrum) of the above set of observations is to be estimated. The ACF at zero lag is given by

$$\phi_0 = \ldots + x_{-2}x_{-2} + x_{-1}x_{-1} + x_0 x_0 + x_1 x_1 + \ldots + x_{n-1}x_{n-1} + \ldots$$

$$= \frac{1}{T} \int x^2(t)\, dt$$
$$T \to \infty$$

Since only n terms are available for estimating ϕ_0, it may be estimated by

$$\hat{\phi}_0 = \frac{1}{n} \cdot [x_0 x_0 + x_1 x_1 + \ldots + x_{n-1} x_{n-1}]$$

Depending on the sample length n, estimate $\hat{\phi}_0$ of ϕ_0 will have variance, σ_0^2, associated with it. Similarly, the ACF at unit lag is given by

$$\phi_1 = \ldots + x_0 x_1 + x_1 x_2 + \ldots + x_{n-2} x_{n-1} + \ldots$$

which can be estimated as

$$\hat{\phi}_1 = \frac{1}{n-1} [x_0 x_1 + x_1 x_2 + \ldots + x_{n-2} x_{n-1}]$$

with a corresponding variance σ_1^2 .

Similarly,

$$\hat{\phi}_2 = \frac{1}{n-2} [x_0 x_2 + \ldots + x_{n-3} x_{n-1}]$$

$$\vdots$$

$$\hat{\phi}_{n-1} = \frac{1}{n-(n-1)} [x_0 x_{n-1}]$$

where the corresponding variances are $\sigma_2^2, \ldots, \sigma_{n-1}^2$.

In general,

$$\phi_k = E_{n-k} [x_i x_{i+k}]$$

Since the quantities $\hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_{n-1}$ are estimates representing the actual values $\phi_0, \phi_1, \ldots, \phi_{n-1}$; the mean of a large number of such estimates will converge to the true value ϕ_k in the limit as the number of estimates tends to ∞.

The variance associated with each estimate $\hat{\phi}$ is also a function of the number of samples, n, of the time series used in obtaining it, being larger for small n and vice versa. Therefore, $\hat{\phi}_{n-1}$ has the maximum variance of all the estimates. The above concepts are illustrated in Figure 3.11 in which the true values of ϕ_k and the standard deviation ranges on the distribution of $\hat{\phi}_k$ are shown. A particular realization $\hat{\phi}_k$ may be as shown by the heavy line, which may be quite different from the actual value ϕ_k, shown by the dotted line. In general, since more samples have gone into the estimation of small lag ACFs, their values may be expected to be closer to actual values, the error increasing for larger lags. as illustrated by —· —· —·· line. There will be a family of such graphs which will be derived from different data sets obtained by taking sequences at different times. They would, on an average, converge to the true graph of ACF.

The ACF may be estimated more reliably by modeling the random process as an autoregressive process, or as some alternative process in which the observed samples can be used to retrodict and predict past and future samples, respectively i.e.,

$$\underbrace{\ldots, x_{-2}, x_{-1},}_{\substack{\text{retrodiction} \\ \text{estimation} \\ \text{error}}} \quad \underbrace{x_0, x_1, x_2, \ldots, x_{n-1},}_{\text{observed data}} \quad \underbrace{x_n, x_{n+1}, \ldots}_{\substack{\text{predicted estimation} \\ \text{error}}}$$
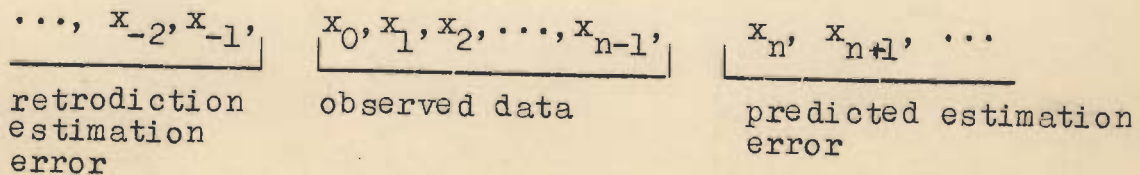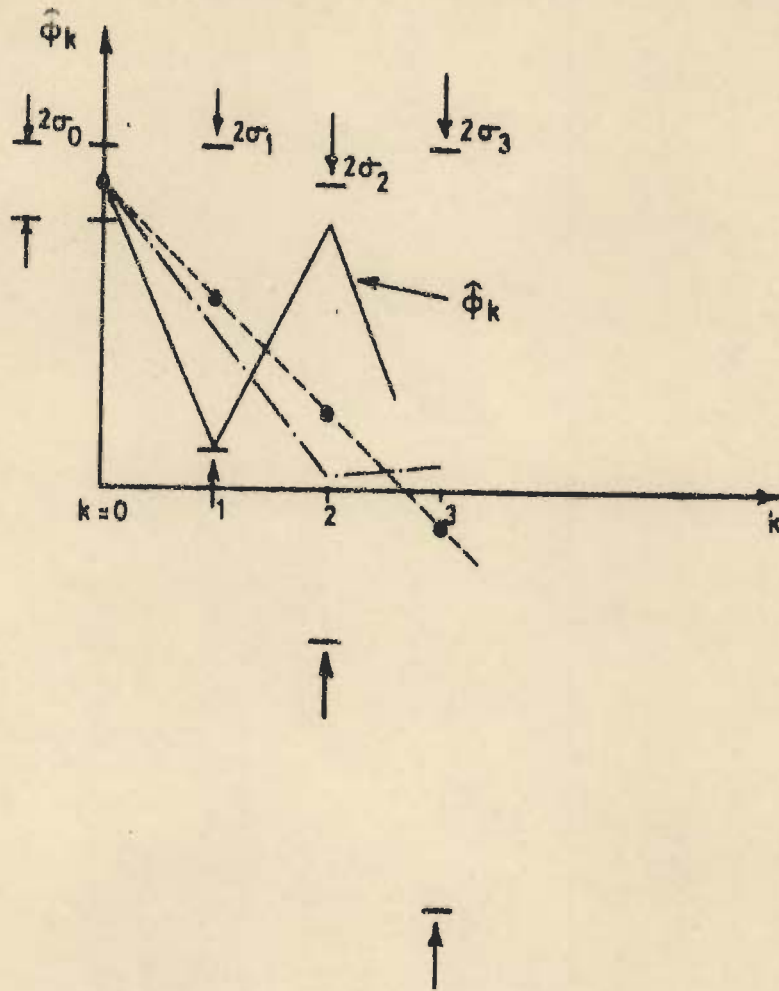
FIG.3.11_THE VARIANCE ASSOCIATED WITH
ESTIMATES $\hat{\phi}_k$, THE AUTOCORRELATION
FUNCTION. ⊕ SHOWS THE TRUE VALUES
ϕ_k FOR THE PROCESS. THE BAR SHOWS
2 STANDARD DEVIATION RANGES ON THE
DISTRIBUTION OF $\hat{\phi}_k$.

An autoregressive process of order N is described by the difference equation :

$$a_0 y(t) + a_1 y(t-1) + \cdots + a_{N-1} y(t-N+1) +$$
$$a_N y(t-N) = x_t$$

This implies that the current value of the output depends upon N previous values of the output $y(t)$ plus the current value of the input x_t. The sequence x_t may be taken to be uncorrelated random noise with,

$$E[x_t] = 0$$

$$E[x_t^2] = \sigma^2$$

The difference equation may be rewritten as :

$$y(t) + \sum_{i=1}^{N} b_i \, y(t-i) = x_t,$$

where $b_i = a_i / a_0$. This equation may be viewed as stating,

$$E\left[\dot{y}(t)\right] = -b_1 y(t-1), \ \ldots, \ -b_N y(t-N) = y_N^p$$

so that x_t is the error in prediction. Then σ^2 can be considered as the mean squared error of prediction, i.e., the error energy of prediction.

An autoregressive model is to be fitted so that σ^2 is a minimum, i.e., the best fit in the least square sense is desired. This corresponds to choosing coefficients b_i, $i = 1, 2, \ldots, N$, in

$$E\left\{x_t^2\right\} = \sigma^2 = E\left\{\left[y(t) + \sum_{i=1}^{N} b_i\, y(t-i)\right]^2\right\}$$

To obtain estimates of b_i in the least square sense or by minimizing the mean square error, σ^2 is differentiated with respect to b_i and equated to zero.

$$\frac{\partial \sigma^2}{\partial b_j} = \frac{\partial}{\partial b_j}\left[E\left\{\left[y(t) + \sum_{i=1}^{N} b_i\, y(t-j)\right]^2\right\}\right] = 0$$

$$E\left\{2\left[\left[y(t) + \sum_{i=1}^{N} b_i\, y(t-i)\right]y(t-j)\right]\right\} = 0$$

$$E\left\{y(t)\, y(t-j) + \sum_{i=1}^{N} b_i\, y(t-i)\, y(t-j)\right\} = 0$$

$$\phi_j + \sum_{i=1}^{N} b_i\, \phi_{i-j} = 0\, ; \quad j = 1, 2, \ldots, N.$$

or,

$$\phi_1 = -\left[b_1\phi_0 + b_2\phi_1 + \cdots + b_n\phi_{N-1}\right]$$
$$\vdots$$
$$\phi_N = -\left[b_1\phi_{N-1} + b_2\phi_{N-2} + \cdots + b_n\phi_0\right];$$

or,

$$\begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_1 & \phi_0 & \cdots & \phi_{N-2} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \phi_{N-1} & & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_N \end{bmatrix} = - \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_N \end{bmatrix}$$

The ACF matrix is of Toeplitz form, and it must be positive definite; it means that all its eigenvectors must be ≥ 0. This gives the coefficients b_1, b_2, $\cdots$, b_N needed for prediction which can be used to extend the original observations for better estimates of ACF. However, the errors increase when longer sequences are predicted, as the predicted values themselves are used for further prediction. The above equation has been manipulated to give (Kanasewich, 1975);

$$
\begin{bmatrix}
\phi_0 & \phi_1 & \cdots & \phi_N \\
\phi_1 & \phi_0 & \cdots & \phi_{N-1} \\
\cdot \\
\cdot \\
\cdot \\
\phi_N & & \cdots & \phi_0
\end{bmatrix}
\begin{bmatrix}
1 \\
\Gamma_2 \\
\cdot \\
\cdot \\
\cdot \\
\Gamma_{N+1}
\end{bmatrix}
=
\begin{bmatrix}
P_{N+1} \\
0 \\
\cdot \\
\cdot \\
\cdot \\
0
\end{bmatrix}
\qquad \cdots (3.16)
$$

where

$$
-\begin{bmatrix}
b_1 \\
b_2 \\
\cdot \\
\cdot \\
\cdot \\
b_{N+1}
\end{bmatrix}
=
\begin{bmatrix}
1 \\
\Gamma_2 \\
\cdot \\
\cdot \\
\cdot \\
\Gamma_{N+1}
\end{bmatrix}
$$

and P_{N+1} is the output power spectrum. For the matrix in equation 3.16 to be positive definite for fixed ϕ_0, $\cdots$, ϕ_{N-1}; ϕ_N will have to lie between a range, the mean of which is taken

to be the estimate of $\not b_N$ (Burg, 1970). The filter $[1, \ulcorner 2, \ldots, \ulcorner N]$ is known as the prediction error filter.

$\ulcorner$ may be considered as a set of prediction filter weights which, when convolved with the input data, will generate a white noise series (Figure 3.12). In the frequency domain the output power spectrum is the product of the input power spectrum and the power response of the filter. The input power spectrum may be obtained by correcting the output power for the response of the filter. In the frequency domain

$$\text{Input power spectrum} = \frac{\text{Output power spectrum}}{\text{Power response of filter}}$$

The input power spectrum is designated by Burg (1967, 1970) as the maximum entropy estimate of power, $P(f)$.

$$P(f) = \frac{P_{N+1} / f_N}{2 \left[1 + \sum_{n=1}^{N} \ulcorner n{+}1 \; e^{-2\pi i f n \wedge t} \right]^2} \qquad \ldots(3.17)$$

where f_N is the Nyquist Frequency and specifies the bandwidth for a sampling interval of $\wedge t$.

The response characteristics of the autoregressive process

$$y(n) + \ulcorner_1 \, y(n-1) + \ldots + \ulcorner_N \, y(n-N) = x_n$$

in z domain are given by :

$$\frac{Y(z)}{X(z)} = \frac{1}{[\ulcorner_1 z + \ulcorner_2 z^2 + \ldots + \ulcorner_N z^N]}$$

which is an all pole system.

WHITENING
FILTER

N TERM INPUT TIME
SERIES WITH SPECTRUM ——————▶ | Γ | ——————▶ OUTPUT TIME SERIES,
$P(f)$ WITH WHITE SPECTRUM,

P_{N+1}
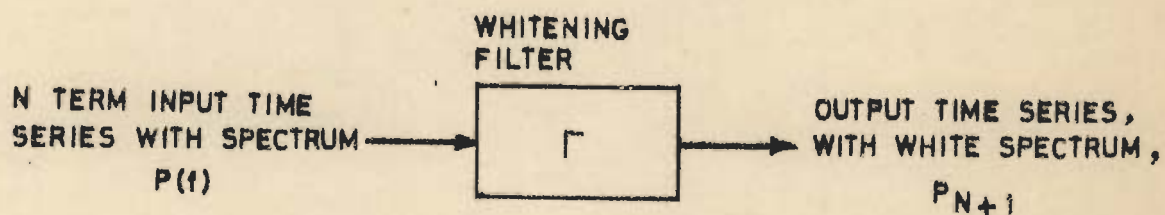
FIG.3.12 _ CONVOLUTION WITH A PREDICTION ERROR FILTER.

The maximum entropy power spectrum $P(f)$ can thus be estimated by using formula (3.17).

The procedure is illustrated for the case of a two term and a three term filter. To estimate the value of the autocorrelation function at the subscripted lag, Burg(1967,1970) first estimated the filter coefficients directly from the data. The ϕ_n and the maximum entropy power spectrum $P(f)$ are then computed from the filter.

Figure 3.13 shows the estimation of a two term prediction error filter $(1, \Gamma\ )$ from an N point long sample of data. It depends on the choice of Γ that minimizes the average power output P_2 of both forward and backward prediction filters. The filter is not run off the ends of the data sample and no assumptions about the time series before and after the data sample are required. The value of Γ which minimizes P_2 is shown in Figure 3.13. $(1, \Gamma\ )$ is a two term minimum phase filter. ϕ_1 and P_2 can be estimated by the matrix equation (3.16), i.e.,

$$\begin{bmatrix} \phi_0 & \phi_1 \\ \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ \Gamma \end{bmatrix} = \begin{bmatrix} P_2 \\ 0 \end{bmatrix}$$

which gives $\phi_1 = -\Gamma \phi_0$

and $\qquad P_2 = \phi_0 (1 - \Gamma^2)$
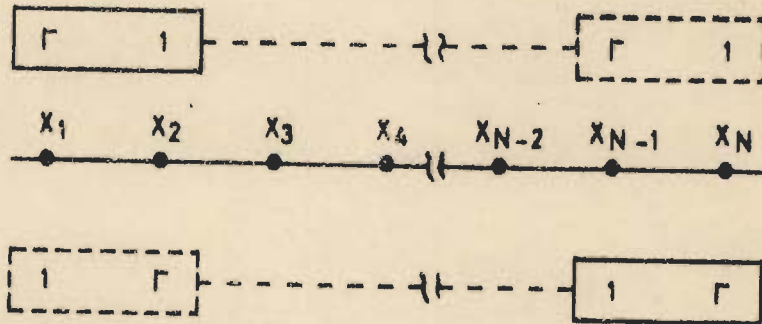
$$P_2 = \frac{1}{2(N-1)} \left\{ \sum_{i=1}^{N-1} (X_{i+1} + \Gamma X_i)^2 + \sum_{i=1}^{N-1} (X_i + \Gamma X_{i+1})^2 \right\}$$

P_2 IS A MINIMUM WHEN

$$\Gamma = -\frac{2(X_1 X_2 + X_2 X_3 + \cdots + X_{N-2} X_{N-1} + X_{N-1} X_N)}{X_1^2 + 2(X_2^2 + X_3^2 + \cdots + X_{N-2}^2 X_{N-1}^2) + X_N^2}$$

FIG.3.13 — ESTIMATION OF A TWO POINT FILTER.

(AFTER BURG, 1970)

ϕ_0 is estimated as the average square value of the data. To obtain higher order prediction error filters, Burg (1967, 1970) has suggested the use of the Levinson algorithm (Levinson, 1949) which is a method of solving simultaneous equations recursively, based on the Toeplitz property and it always yields a minimum phase filter. In the Levinson recursion, a filter of order (n+1) is built from one of order n. For example, for a three term filter

$$
\begin{bmatrix} 1 \\ \Gamma_2 \\ \Gamma_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \Gamma \\ 0 \end{bmatrix} + \Gamma_3 \begin{bmatrix} 0 \\ \Gamma \\ 1 \end{bmatrix}
$$

and thus

$$
P_3 = \frac{1}{2(N-2)} \sum_{i=1}^{N-2} \left\{ \left[x_{i+2} + \Gamma(1+\Gamma_3) x_{i+1} + \Gamma_3 x_i \right]^2 \right.
$$

$$
\left. + \left[x_i + \Gamma(1+\Gamma_3) x_{i+1} + \Gamma_3 x_{i+2} \right]^2 \right\}
$$

Minimizing P_3 with respect to Γ_3 gives (see Figure 3.14)

$$
\Gamma_3 = -2 \sum_{i=1}^{N-2} \frac{(x_{i+2} + \Gamma x_{i+1})(x_i + \Gamma x_{i+1})}{(x_{i+2} + \Gamma x_{i+1})^2 + (x_i + \Gamma x_{i+1})^2}
$$

Again, $|\Gamma_3| \le 1$. The ϕ_2 and P_3 are then estimated by the third order matrix equation,
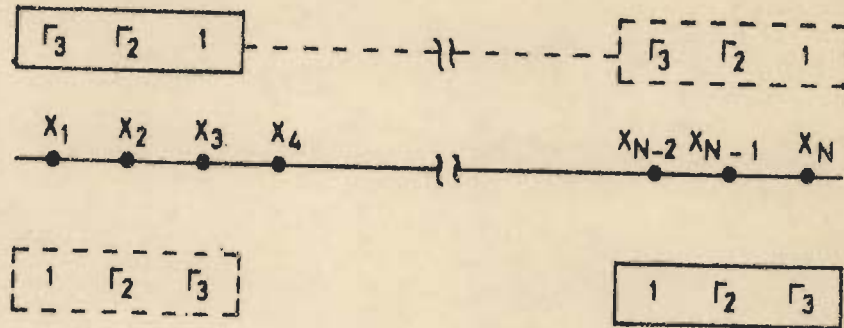
$$P_3 = \frac{1}{2(N-2)} \sum_{i=1}^{N-2} \left\{ (x_{i+2} + x_{i+1}\Gamma_2 + x_i\Gamma_3)^2 + (x_i + x_{i+1}\Gamma_2 + x_{i+2}\Gamma_3)^2 \right\}$$

WHERE $\Gamma_2 = \Gamma(1 + \Gamma_3)$

FIND THE VALUE OF Γ_3 WHICH MINIMIZES P_3
THE ESTIMATE OF THE THREE POINT FILTER $[1, \Gamma(1+\Gamma_3), \Gamma_3]$
WILL BE MINIMUM PHASE AND THUS BE A POSSIBLE
FILTER.

FIG.3.14_ ESTIMATION OF A THREE POINT
FILTER.

(AFTER BURG, 1970)

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_0 & \phi_1 \\ \phi_2 & \phi_1 & \phi_0 \end{bmatrix} \left[\begin{Bmatrix} 1 \\ \Gamma \\ 0 \end{Bmatrix} + \Gamma_3 \begin{Bmatrix} 0 \\ \Gamma \\ 1 \end{Bmatrix} \right] = \left[\begin{Bmatrix} P_2 \\ 0 \\ \Delta \end{Bmatrix} + \Gamma_3 \begin{Bmatrix} \Delta \\ 0 \\ P_2 \end{Bmatrix} \right]$$

$$= \begin{bmatrix} P_3 \\ 0 \\ 0 \end{bmatrix}$$

which yields

$$\phi_2 = -\Gamma_3 \phi_0 - \Gamma(1+\Gamma_3)\phi_1$$

and $P_3 = P_2(1 - \Gamma_3^2)$.

Continuing on, three point filter can be used to form the four point filter through use of the single parameter, Γ_4. Then, applying the filter to the data sample and varying Γ_4 to minimize the output power, the correct four point prediction error filter is estimated. This procedure can be continued on with the assurance that no impossible filter will be obtained. For an M point filter the expression is

$$P_M = P_{M-1}(1 - \Gamma_M^2)$$

The maximum entropy power spectrum P(f) can then be estimated from the filter by using equation (3.17). The final prediction error (FPE) as given by Akaike (1969a, b, 1970) is

$$[FPE]_M = \frac{N + M + 1}{N - M - 1} \cdot P_M^2$$

P_M is calculated by equation (3.17) for successively higher values of m until a minimum is obtained for m = M. This yields an estimate of the mean square error in prediction. As discussed by Ulrych and Bishop (1975), a cutoff of M = N/2 is imposed. The filter that minimizes the FPE is chosen.

CHAPTER - IV

SEISMIC ATTRIBUTES RELATED TO LITHOLOGY

Seismic data interpretation is based on the correlation of events in a seismogram and has till now been used mostly to delineate subsurface structures. Despite the phenomenal success of the correlation of seismic events in a seismic section in the delineation of subsurface structural features, it has not yet proved to be a reliable tool by itself in areas of complicated geology where stratigraphic traps are explored. With the above limitations of the conventional method of interpretation of seismic data, it is not difficult to see the reason why few oil discoveries in stratigraphic traps have been made as compared to those in structural traps, despite the fact that the stratigraphic traps may even outstep structural traps in ultimate reserves (Lyons, 1968).

A number of new approaches to map stratigraphy using seismic data have been proposed recently. One of the most significant parameters that has come into use for identifying lithology is the seismic interval velocity. It is the average velocity of the medium between flat parallel interfaces and is estimated from root mean square (RMS) velocity values for reflection events at the top and bottom of the interval (Smith, 1969; Taner and Koehler, 1969; Taner, Cook and Neidell, 1970). The uncertainty involved in this method becomes extremely large as the interval becomes very thin or when there is significant

departure from horizontal bedding, (Schneider, 1971). Savit
and Matekar (1971) have proposed the use of seismic energy
attenuation as a guide to subsurface lithology.

Although amplitude of a seismic reflection is a
function of many factors other than the reflection coefficients
of the reflecting interface, yet it has been used in seismic
interpretation (Pan and DeBremaecker, 1970; O'Doherty and
Anstey, 1971; Sheriff, 1975; Khattri, Gaur, Mithal and
Tandon, 1978; Khattri, Mithal and Gaur, 1979). Lateral ampli-
tude variations convey information about changes in the
acoustic impedance which may have stratigraphic significance.
Lindseth (1979) has mapped stratigraphic traps by the use of
acoustic impedance log, termed as Seislogs, made from reflec-
tion amplitudes.

Since seismic analysis and display techniques have
become more quantitative, Sheriff (1976) has enumerated
(Table 4.1) the various seismic observations which lead to
seismic interpretation. By combining observations in a
synergetic manner, the reliability of inferences about the
lithology, stratigraphy, fluid content, etc., can be improved
(Marr, 1971).

Table 4.1 - Seismic observations used in geological inter-
pretation (after Sheriff, 1976).

Arrival time	Depth
Differences with location	Dip
Differences with offset	Velocity
Differences in amplitude	Reflectivity
Angular relations	Geologic history
Patterns	Depositional situations
Combinations	Gross lithology
	Stratigraphy
	Fluid content

Synthetic seismograms have been used to interpret
stratigraphic sequences. Harms and Tackenburg (1972) have
suggested the use of lateral changes in the amplitudes, pola-
rity and continuity of reflection in the search for strati-
graphic traps.

Khattri and Gir (1975, 1976) have studied the synthe-
tic seismograms for wave form and spectral characteristics for
four basic sedimentation models : (1) interbedded sand-shale
model representing the seidments of generally fluviatile
origin, (2) interbedded coal-shale model representing deltaic
deposits,(3) sedimentary models representing transgression
and regression of shore lines, and (4) a basal sand model.
Their results have shown that for the first two models
a change in the sand-shale or coal-shale ratio results in

characteristically different seismograms. The nature of the seismogram is also strongly dependent on the arrangement of sand-shale or coal-shale layers, keeping the sand-shale or coal-shale ratios constant. The transgression, regression and basal models also produce characteristically different seismic responses and frequency spectra.

Improvements in seismic data processing techniques make it possible to observe geologically significant information on seismic records. Analysis of a seismic trace permits the transformation to polar coordinates and the measurement of reflection amplitude , instantaneous phase and frequency. These attributes have been coded by colour on seismic sections (Taner and Sheriff, 1977) and this display helps in establishing interrelationships among measurements, and in locating and understanding faults, unconformities, pinchouts, stratigraphic sequences and boundaries and hydrocarbon accumulations.

The aforementioned efforts to correlate some properties of the seismic trace to the subsurface lithological variations can be made qualitatively, as has also been shown by Vail, Mitchum, Todd, Widmier, Thompson, Sangree, Bubb, and Hatlelid (1977). The qualitative approach is not very diagnostic while inferring lithology when either the seismic data quality is not very good or the difference in lithologies are subtle. Even a qualitative approach taking into account a

single or a few measurements from the seismogram may not prove useful. Under such circumstances a multivariate approach may provide the answer.

Mathieu and Rice (1969) and Avasthi and Verma (1973) attempted a discriminant analysis with a linear combination of more than one observed parameter of seismic trace for the determination of variation in stratigraphic conditions. Mathieu and Rice (1969) using this technique have discriminated sandy from shaly sections on the basis of the following parameters extracted from synthetic seismograms : amplitude of a peak, time interval between peaks and changes in wave shape. After working out a linear combination of the measured parameters from synthetic seismograms to distinguish sand from no sand group, the field seismic traces were then classified into one of the above two groups. Their technique was found to be successful in some cases, while in others it failed completely. Similar approach has been adopted by Avasthi and Verma (1973) to infer subsurface stratigraphy from seismic data in Gujarat (India). They choose to study the following parameters : number of cycles of   reflection, predominant frequency of the reflection, time required to reach peak of the envelope of the group reflections (rise time of reflections) and time required to reach average level of trace from peak of envelope of the group of reflections (decay time of envelope). Using the above mentioned approach they have delineated pervious and impervious zones and have determined the thickness of pervious zones.

Auxiliary seismic derived quantities such as amplitude, polarity, frequency, etc., require several different graphical displays for deeper insight into the seismic data (Sheriff, 1977). As such, it is desirable to display at least part of the data embodied in the present study. Since this involves considerable amount of data with 508 synthetic and 387 field seismograms, it is possible to show here only a very limited number of these traces.

Some examples from the 255 traces of synthetic seismograms analysed for Model E are given in Figures 3.7 and 4.1. The duration of the computed seismogram varies between 496 ms and 508 ms two way vertical travel time depending on the velocity, thickness of the lithounits comprising the stratigraphic model and the total thickness of the model which is approximately 200 m (see section 2.5). The seismograms are shown to occur only after 280 ms, as the source pulse travels through a uniform overburden and consequently does not give rise to any reflection from within. The interface between the overburden and the top of the model is a strong reflector as evidenced by the data given in Table 2.8, and this gives a high reflection amplitude at 300 ms in all the seismograms. A coal interface gives high reflection amplitudes and if two coal interfaces occur within 44 ms of each other strong interference is evidenced. Since Model E has on an average 21 percent coal therefore two seismograms are characterized
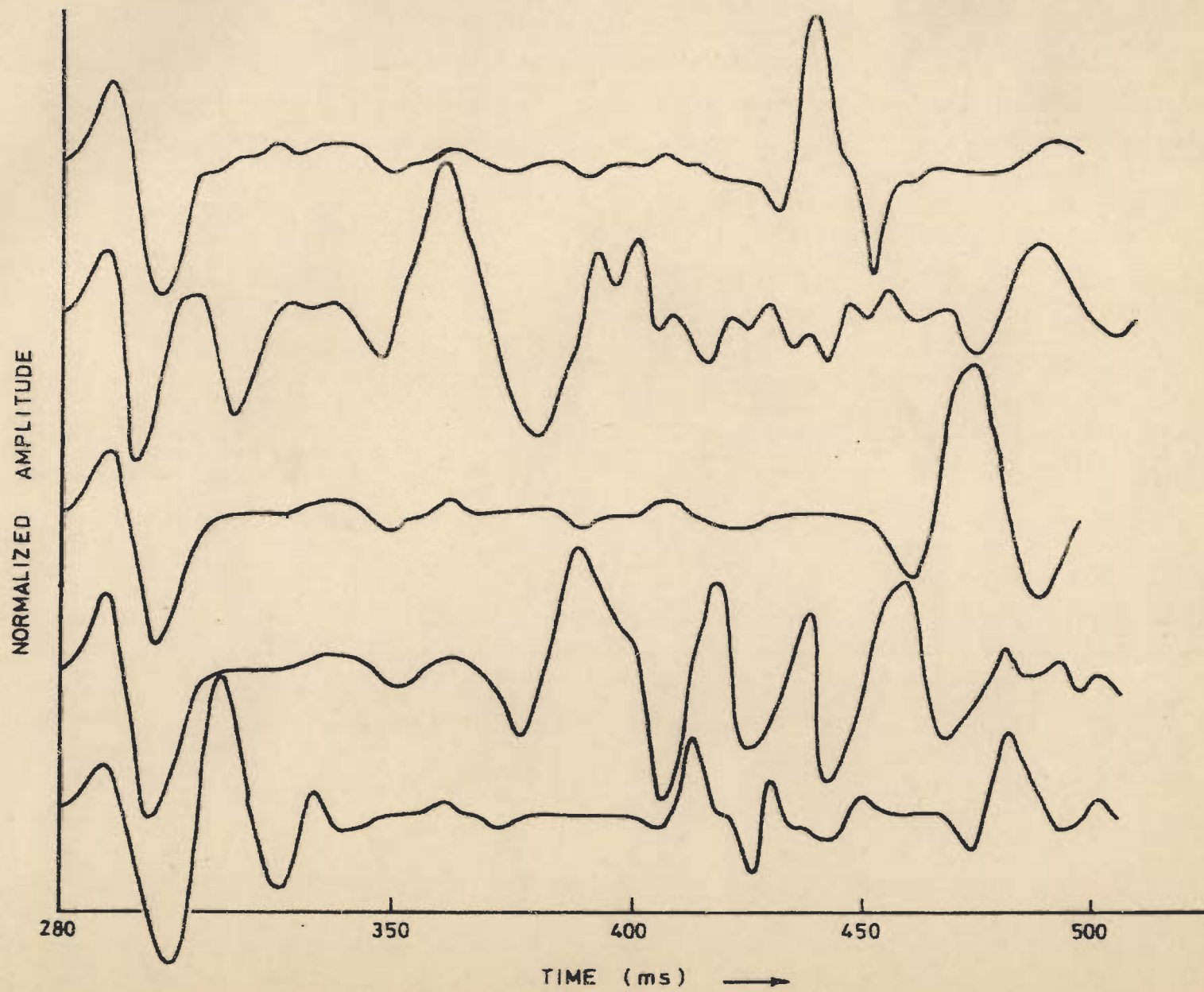
FIG. 4.1_ SOME EXAMPLES FROM THE 255 TRACES OF SEISMOGRAMS ANALYSED

FOR MODEL E .

by high reflection interference patterns. However, Model F which has a relatively small amount of coal, 3 percent, shows seismograms which are relatively smoother, with high reflection amplitudes at coal interfaces and at the interface between the overburden and the model. The seismograms of Model F are of longer duration extending upto 552 ms, as the velocities are lower than that of Model E. Some examples from the 253 traces of reflection seismograms analysed for Model F are given in Figures 3.6, 3.8 and 4.2. These subtle differences between seismograms of the two models are evident, yet these qualitative changes make no contribution towards any knowledge of classifying any seismogram belonging to either of these models. It therefore becomes imperative that a quantitative study is paramount - either of the seismograms or quantities derived from it. With this aim, the information in a seismogram is transferred to the autocorrelation function, power spectrum, cumulative power spectrum, cumulative frequency weighted power spectrum and logarithm of power spectrum; and quantities, hereafter referred to as variables or parameters, are derived from these with a view to quantify each seismogram and thereby to classify each to its relevant model, and assign new seismograms belonging to either of the models to its proper class.

4.1 SEISMIC ATTRIBUTES FROM THE AUTOCORRELATION FUNCTION

Sinvhal (1976) and Khattri, Sinvhal and Awasthi(1979) have used the autocorrelation function of the impulse response
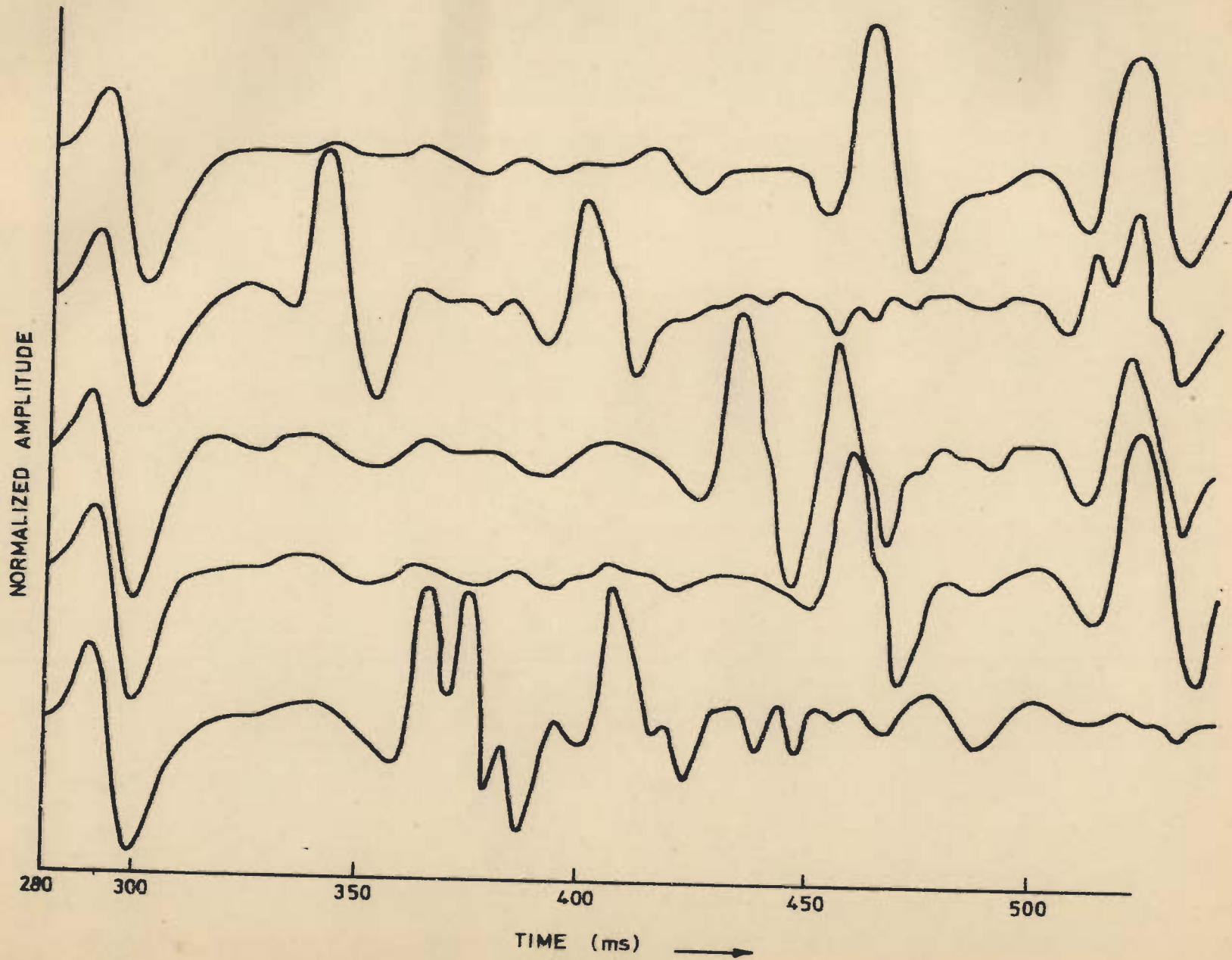
FIG. 4.2 — SOME EXAMPLES FROM THE 253 TRACES OF SEISMOGRAMS
ANALYSED FOR MODEL F.

of models depicting subsurface lithologies, and have abstract-
ed seismic parameters from them to characterize lithostrati-
graphy. The parameters applied to decipher lithology were
A_1/A_0, A_2/A_0, A_2/A_1, where A denotes the autocorrelation
function at the subscripted lag.

These can statistically distinguish between the for-
mations consisting of either sand-shale sequences or coal-
shale sequences, as shown in Table 4.2. The autocorrelation
function is shown in Figure 4.3a, A_0 is not shown as it goes
out of scale but the values of A_1 and A_2 are marked in
Figure 4.3a.

Sinvhal, Gaur, Khattri, Moharir and Chander (1979)
have argued that since A_2/A_1 can be obtained from the other
two variables by the simple relation $(A_2/A_0)/(A_1/A_0)$ and it
bears a deterministic relation to them, therefore, it should
not be used for the discriminatory analysis and it is suffi-
cient to use A_2/A_0 and A_1/A_0 from amongst the three auto-
correlation function variables for any further analysis. For
this reason only, these two variables have been retained for
the present study and some new ones have been searched out
and are listed below. These were used to distinguish different
lithostratigraphic situations.

(1) A_1/A_0 ;

(2) A_2/A_0 ;

Table 4.2 – <u>Parameters from the autocorrelation function to distinguish various pairs of Models based on Kolmogorov-Smirnov statistic (Modified after Sinvhal, 1976)</u>

Model	B	C			D		
A	---	$\frac{A_2}{A_1}$	$\frac{A_2}{A_0}$	$\frac{A_1}{A_0}$	$\frac{A_2}{A_1}$	—	$\frac{A_1}{A_0}$
B		$\frac{A_2}{A_1}$	$\frac{A_2}{A_0}$	$\frac{A_1}{A_0}$	$\frac{A_2}{A_1}$	—	$\frac{A_1}{A_0}$
C					—	—	$\frac{A_1}{A_0}$

Table 4.3 – <u>Parameters from the power spectrum to distinguish various pairs of groups based on the Kolmogorov-Smirnov statistic (Modified after Sinvhal, 1976)</u>

Groups	B	C			D		
A	f_E - -	f_E	f_P	f_M	f_E	f_P	f_M
B		f_E	f_P	f_M	f_E	f_P	f_M
C					-	-	-
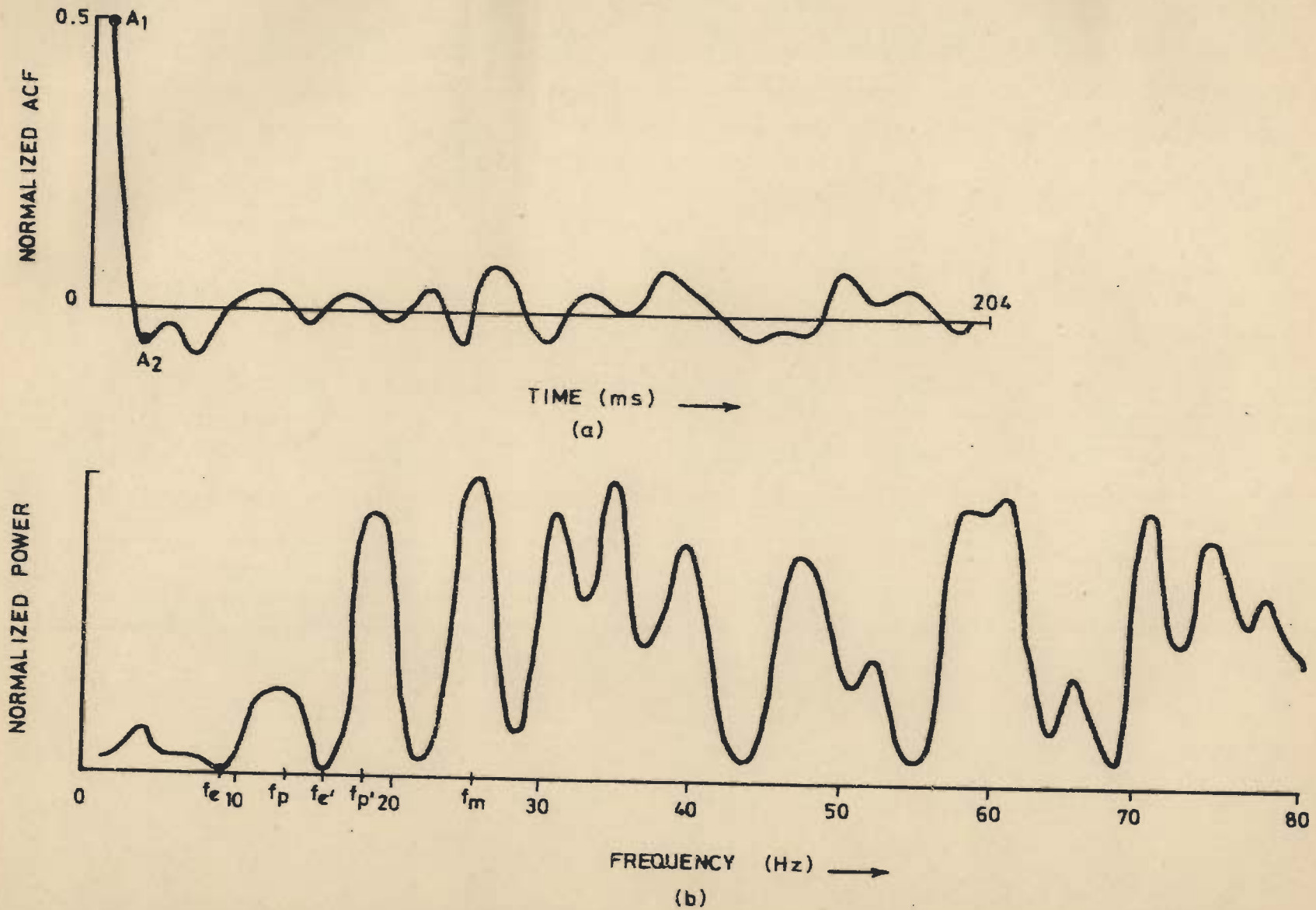
FIG. 4.3 (a)_ AUTOCORRELATION FUNCTION OF IMPULSE RESPONSE FOR MODEL 125 (GROUP D)
A₀ IS NOT SHOWN AND

(b)_ POWER SPECTRUM OF IMPULSE RESPONSE FOR MODEL 125 (GROUP D)

(AFTER SINVHAL, 1976).

$$(3) \quad A_3/A_0 \quad ;$$

$$(4) \quad A_{min}/A_0 \quad ;$$

$$(5) \quad T_1 \quad , \text{ time of the first zero crossing,}$$

$$(6) \quad T_2 \quad , \text{ time of the second zero crossing;}$$

$$(7) \quad T_3 \quad , \text{ time of the third zero crossing and}$$

$$(8) \quad T_{amin} \quad , \text{ time of the first minimum.}$$

These eight variables were picked from the autocorrelation functions of each of the 508 synthetic and 387 real seismograms studied. Some autocorrelation functions with the above mentioned 8 variables marked on it, are shown for synthetic cases in Figures 4.4, 4.5 and 4.8a.

Figures 4.6 and 4.7 show some examples of the 255 and 253 traces of the autocorrelation functions (ACF) analysed for Model E and F respectively. These figures show the highly peaked character of the ACF at zero lag, the amplitudes become negative for all the functions in a very short time of about 12 to 16 ms. The ACFs of Model E, in general, show a highly oscillatory though dissipating character, indicating some oscillations which are also evident in the stratigraphic sections. These oscillations could be due to the dispersed vertical distribution of coal within the model giving rise to cyclic sedimentation sequences. In most cases the ACFs of Model F are relatively flat after the first sharp peak, in keeping with the low coal content of this model.
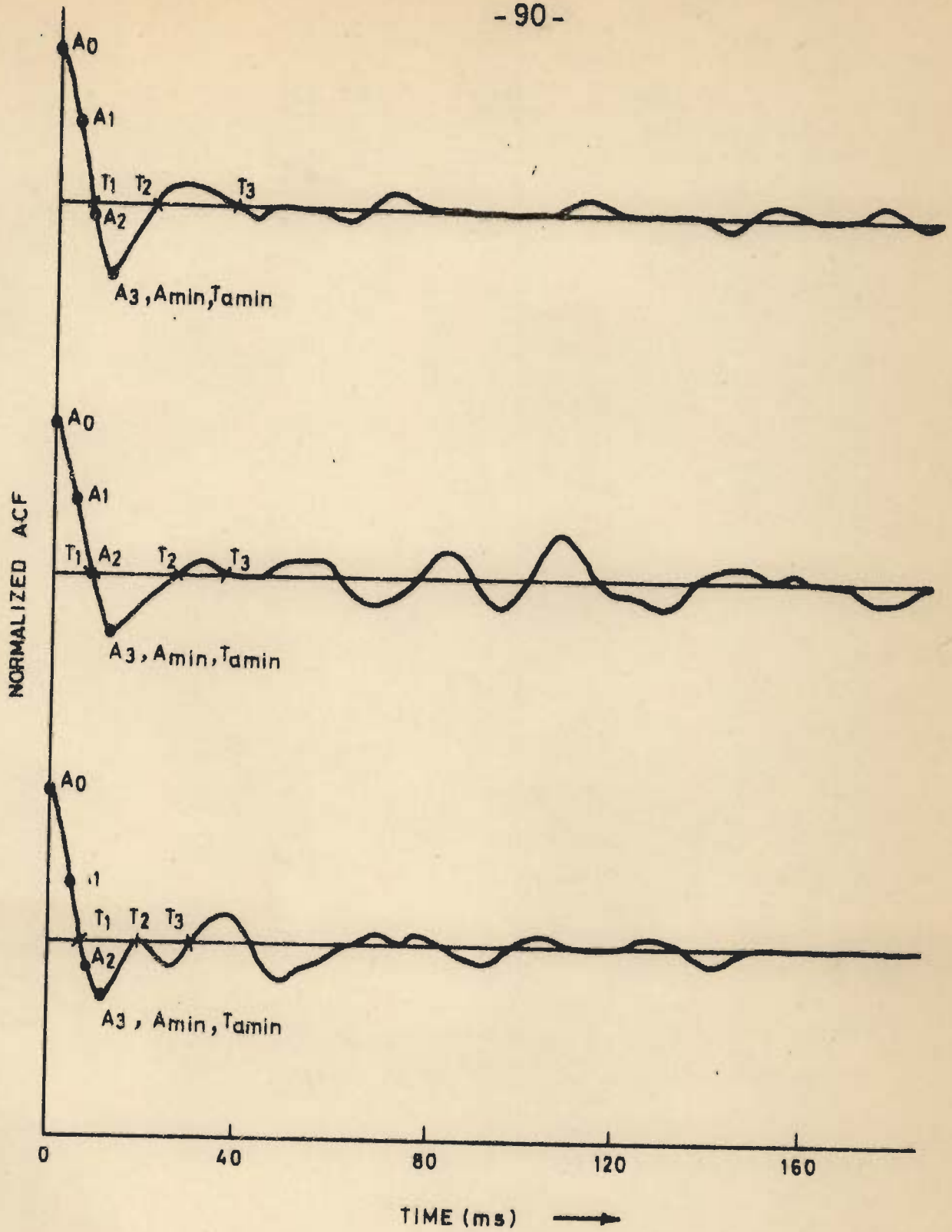
FIG. 4.4_ SOME AUTOCORRELATION FUNCTIONS OF SYNTHETIC

SEISMOGRAMS FOR MODEL E

(ACF = AUTOCORRELATION FUNCTION).
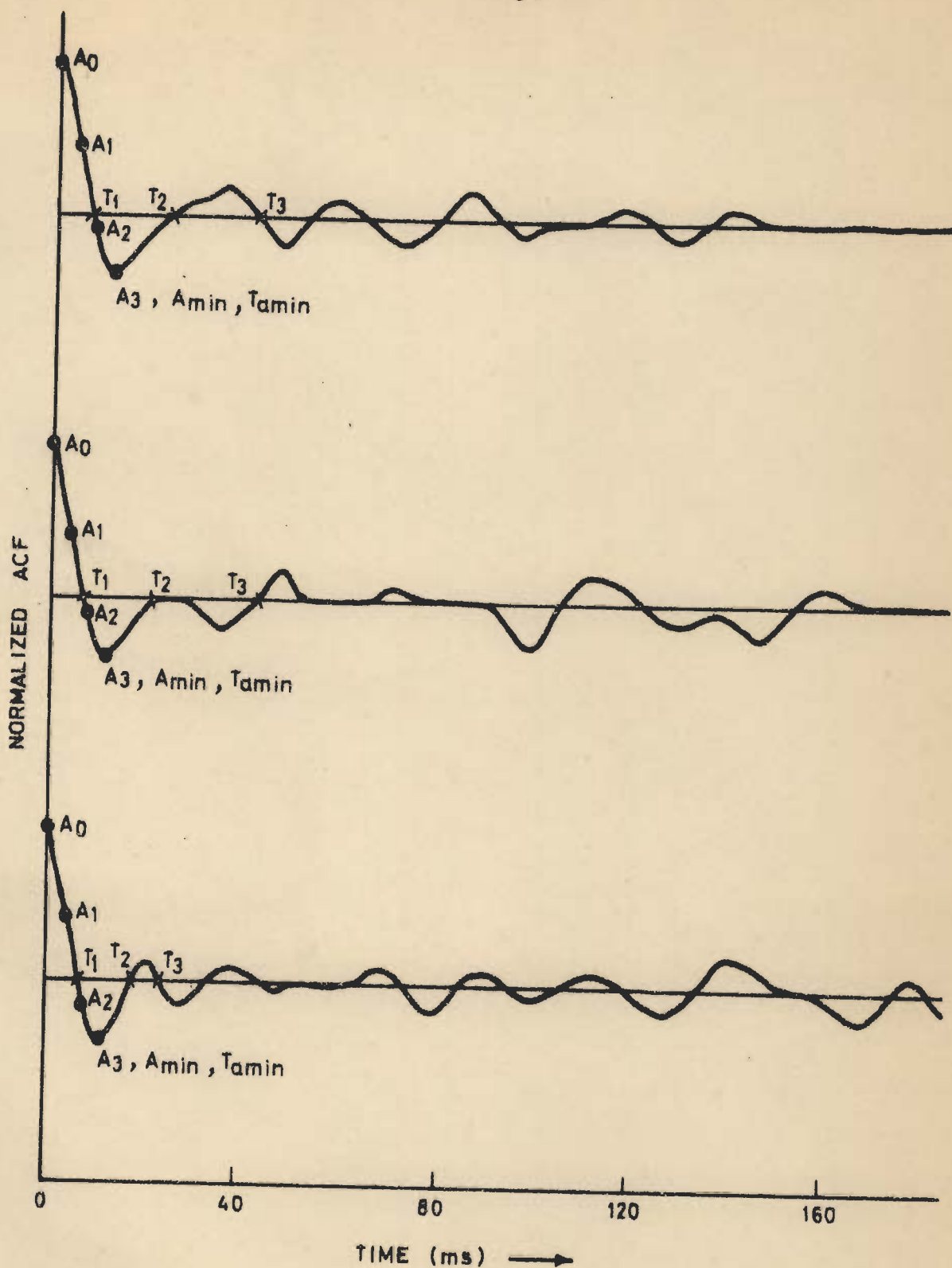
FIG. 4.5 — SOME AUTOCORRELATION FUNCTIONS OF SYNTHETIC SEISMOGRAMS FOR MODEL F (ACF = AUTOCORRELATION FUNCTION).
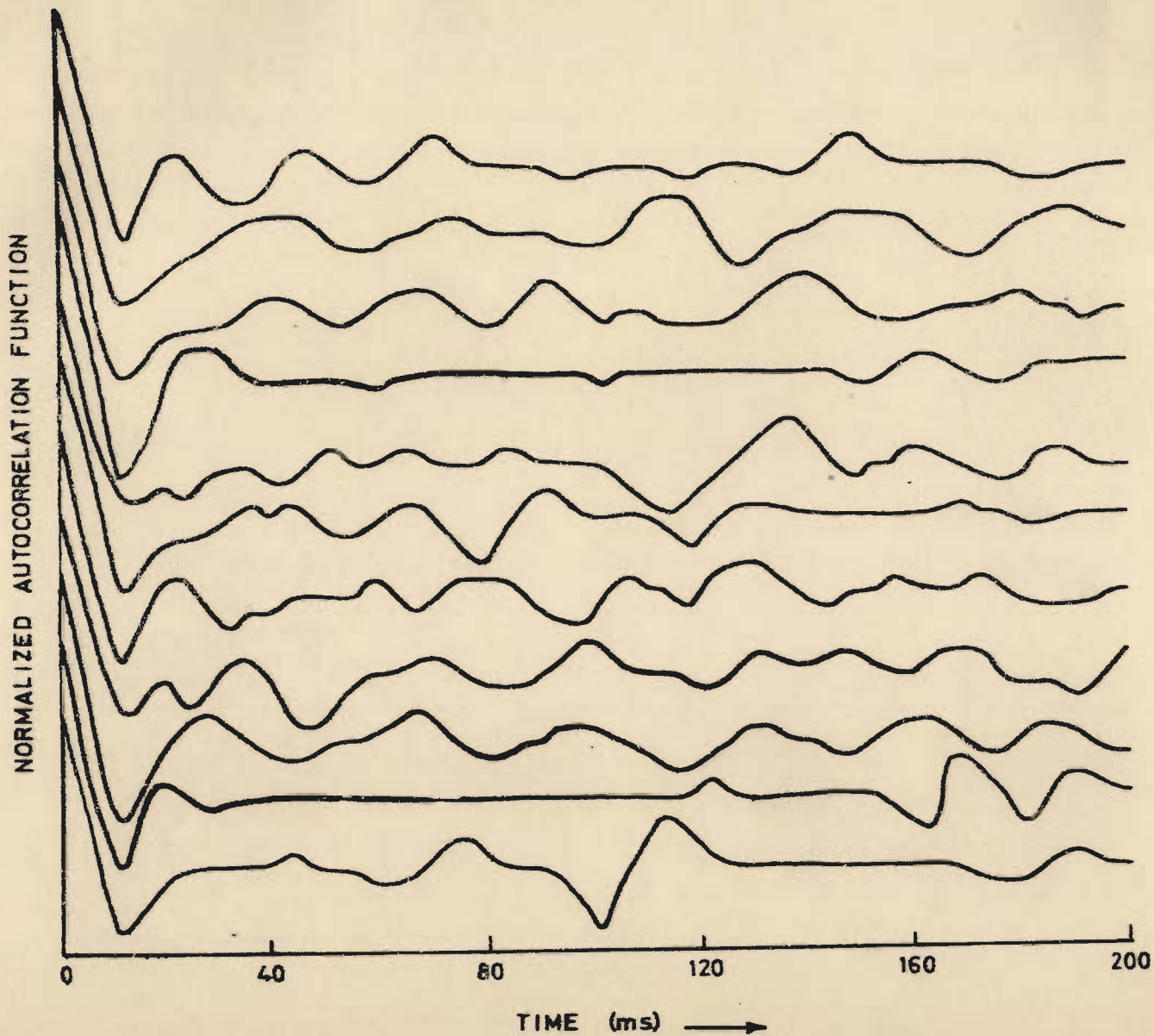
FIG. 4.6_SOME EXAMPLES FROM THE 255 TRACES OF AUTOCORRELATION

FUNCTION ANALYSED FOR MODEL E.
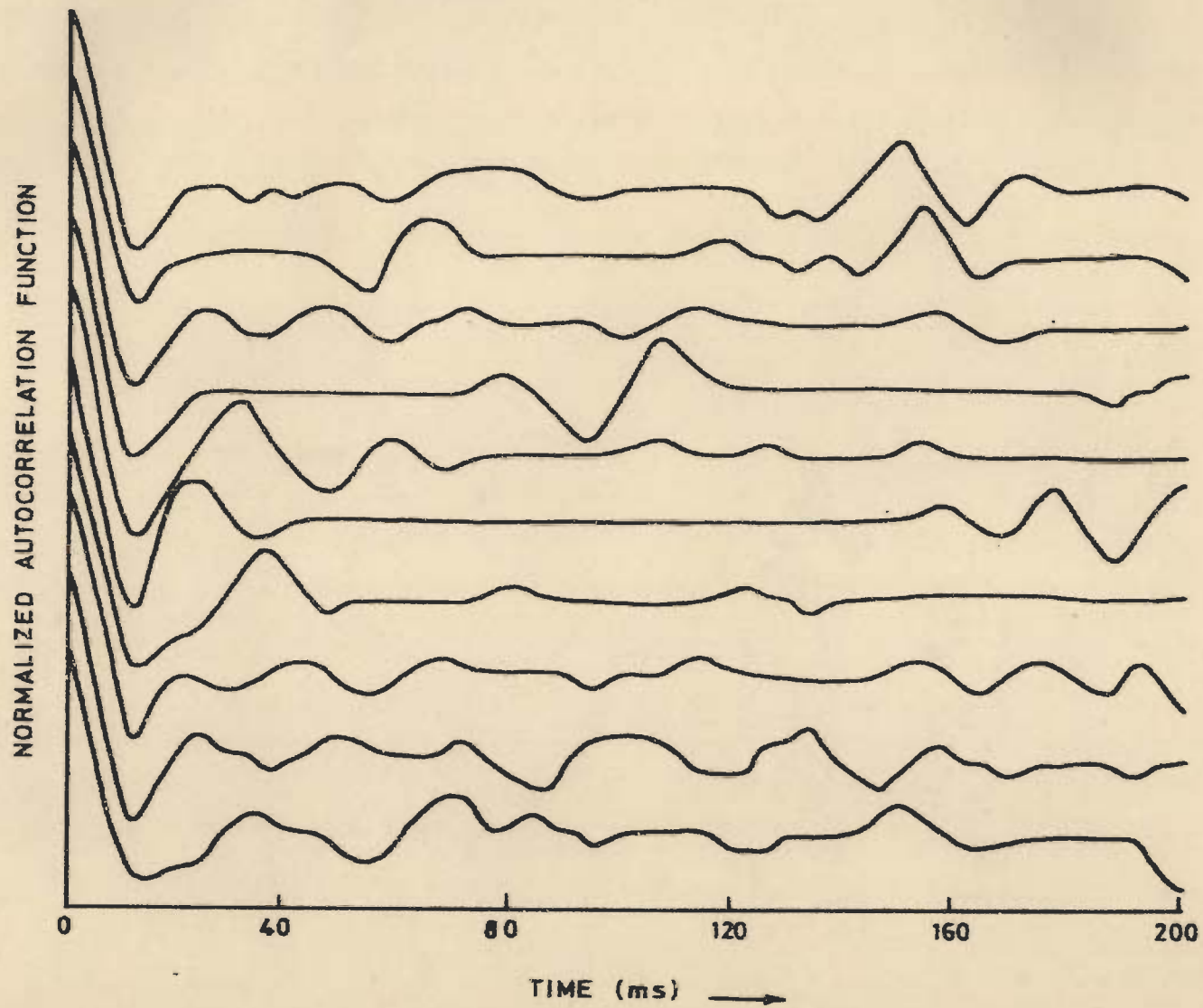
FIG. 4.7_SOME EXAMPLES FROM THE 253 TRACES OF AUTOCORRELATION
FUNCTIONS ANALYSED FOR MODEL F.
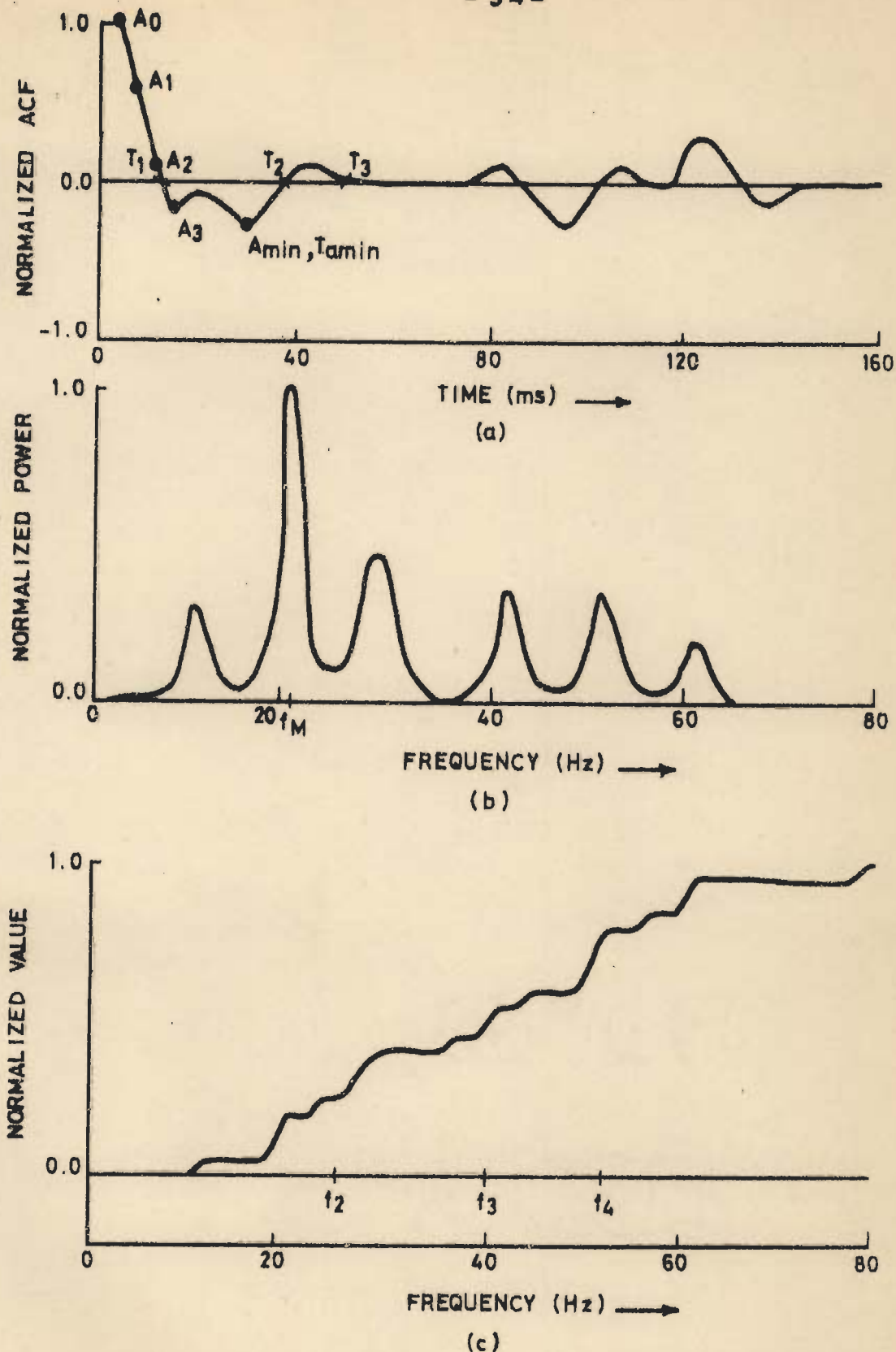
FIG. 4.8 _ ONE CASE OF MODEL E (a) AUTOCORRELATION
FUNCTION,(b) POWER SPECTRUM,(c) CUMULATIVE
POWER WEIGHTED FREQUENCY SPECTRUM,
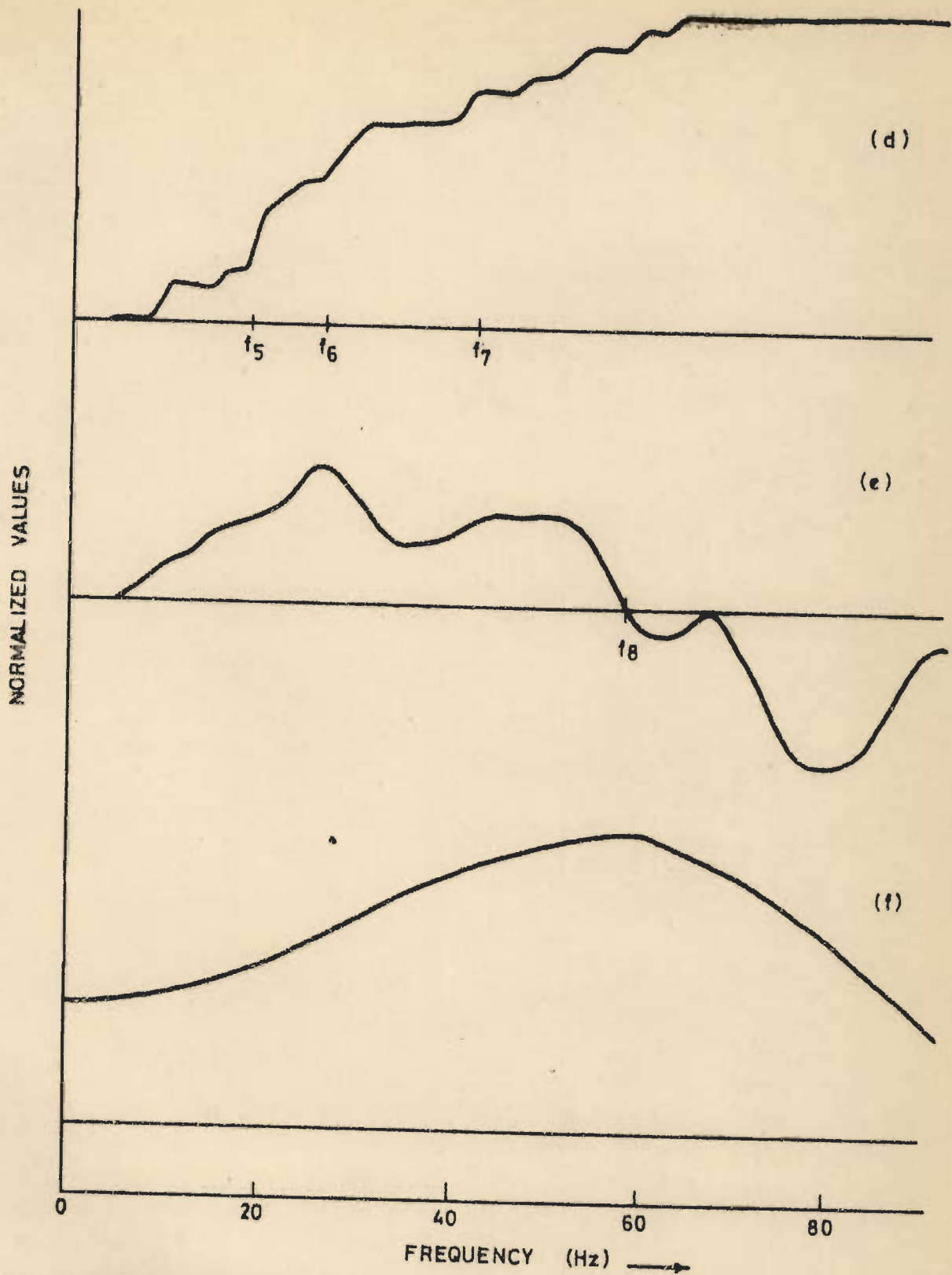
FIG.4.8 (d)_ CUMULATIVE POWER SPECTRUM,

(e)_ LOG POWER SPECTRUM AND

(f)_ AMPLITUDE SPECTRUM OF SOURCE WAVELET

SHOWN IN FIG. 3.6.

The oscillatory character of a seismogram is best studied by taking its power spectrum and its derivatives, and these have been analysed for the present study.

4.2 SEISMIC ATTRIBUTES FROM THE POWER SPECTRUM

Sinvhal (1976) and Khattri, Sinvhal and Awasthi(1979) have used the power spectrum of the impulse response of models depicting subsurface lithologies, and have abstracted the following seismic parameters : f_e, frequency in the power spectrum which divides the band of high and low energy; f_p, frequency of the first significant peak in the power spectrum and f_m, frequency at which the maximum power occurs in power spectrum.

From Figure 4.3(b) it is clear that f_e could also be marked at position shown by f'_e; there is no rigorous criterion to ascertain the frequency f_e; and automatically f_p would shift to a position marked f'_p, because f_p invariably follows f_e. This introduces a certain arbitrariness in picking of these two parameters which are assigned for differentiating lithologies. Despite this drawback f_e and f_p could still differentiate the four groups of lithologies as is evident from Table 4.3 which shows the results of the Kolmogorov-Smirnov test (Miller and Kahn, 1962). The frequency f_m does not suffer from any such drawback, and has been retained for the present study while the frequencies f_e and f_p have been dropped for the above reasons.

The spectral analysis studies by Sinvhal (1976);
Khattri, Sinvhal and Awasthi (1979) and Sinvhal, Gaur, Khattri,
Moharir, and Chander (1979) are for the impulse response of a
stratified medium, which gives a broad band spectrum. In
real cases this spectrum is modified to a band limited spectrum
by that of the source wavelet as well as the attenuating earth.
A Ricker wavelet with maximum amplitude unity and a peaking
frequency of 60 Hz is used for this study (Figure 4.8f) which
is higher than that usually met in field, with the view that
this will give higher seismic resolution (Lyons and Dobrin,
1972).

Nine variables have been identified from the power
spectrum and have been used in distinguishing different kinds
of lithologies. The nine variables picked from the maximum
entropy power spectrum and its derivatives, illustrated in
Figures 4.8b, c, d and e for one simulation of Model E, are
listed below :

(1) f_M,   frequency at which maximum energy occurs
              (Figure 4.8b),

(2) f_1,   the average power weighted frequency of the
              power spectrum,

(3) f_2,   the frequency at which 25th percentile value
              of frequency weighted power occurs,

(4) f_3,   the frequency at which 50th percentile value
              of frequency weighted power occurs,

(5) f_4, the frequency at which 75th percentile value of frequency weighted power occurs(Figure 4.8c);

(6) f_5, the frequency at which 25th percentile of power occurs;

(7) f_6, the frequency at which 50th percentile of power occurs;

(8) f_7, the frequency at which 75th percentile of power occurs (Figure 4.8d) and

(9) f_8, the lowest frequency at which the logarithm of power decreases to zero (Figure 4.8e).

The frequencies f_2, f_3 and f_4 are akin to the notion of pre-emphasis used in communication theory (Panter, 1965). Some power spectra and their derivative spectra, with the above mentioned nine variables marked on some are shown for the synthetic cases in Figures 4.9 - 4.28.

Figures 4.9 and 4.10 show power spectra of some synthetic seismograms of Models E and F respectively, the frequency f_M is marked on them. Figures 4.11 - 4.16 display some examples from the 255 and 253 traces of power spectra analysed for Models E and F. The spectra show considerable variation - some have one or more sharp peaks while the others have several peaks occuring at different frequencies. However, their relationship with the stratigraphic sequences is not explict.
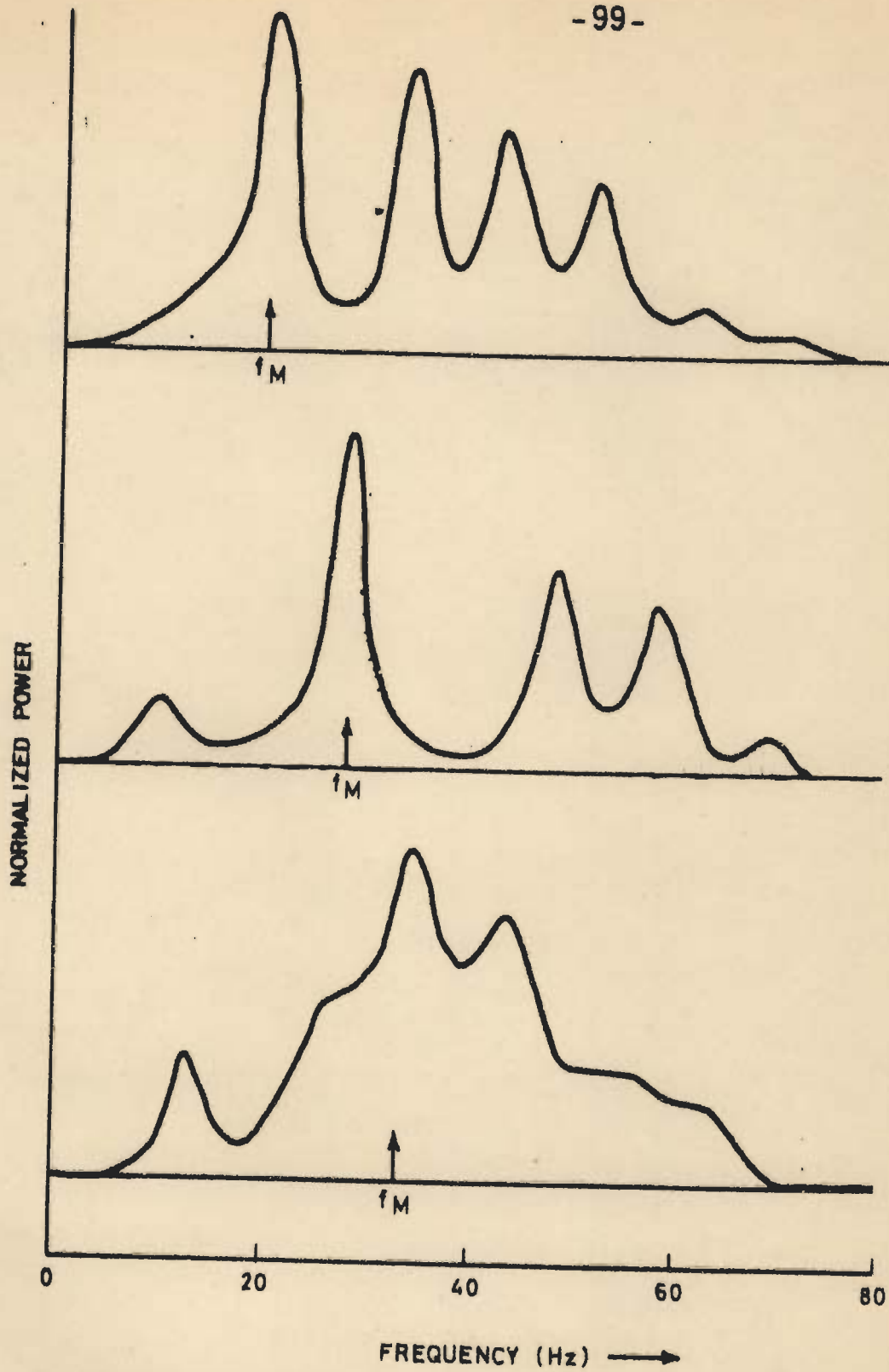
FIG.4.9 _ POWER SPECTRA OF SOME SYNTHETIC
SEISMOGRAMS FOR MODEL E.
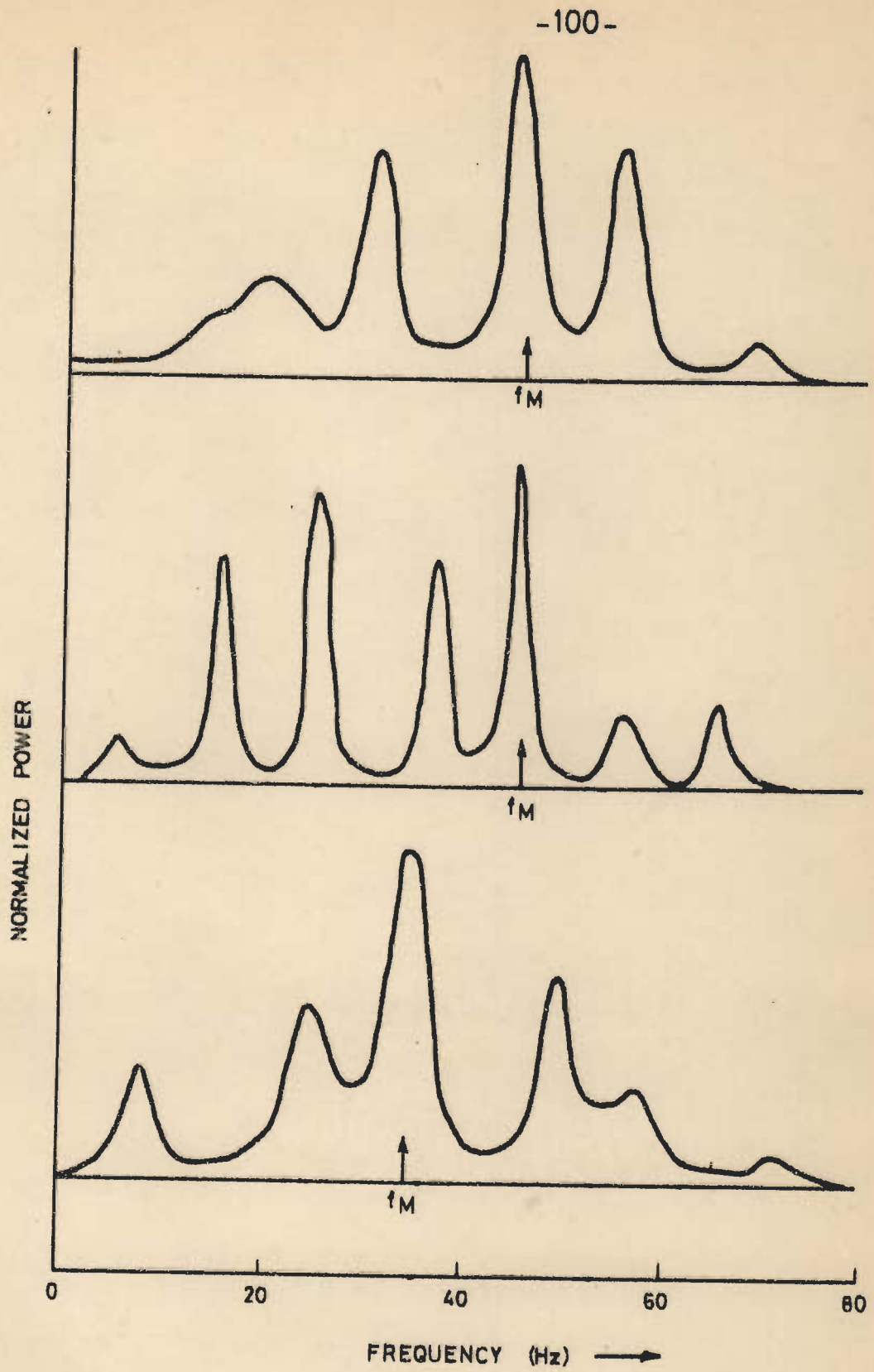
FIG. 4.10_ POWER SPECTRA OF SOME SYNTHETIC
SEISMOGRAMS FOR MODEL F.
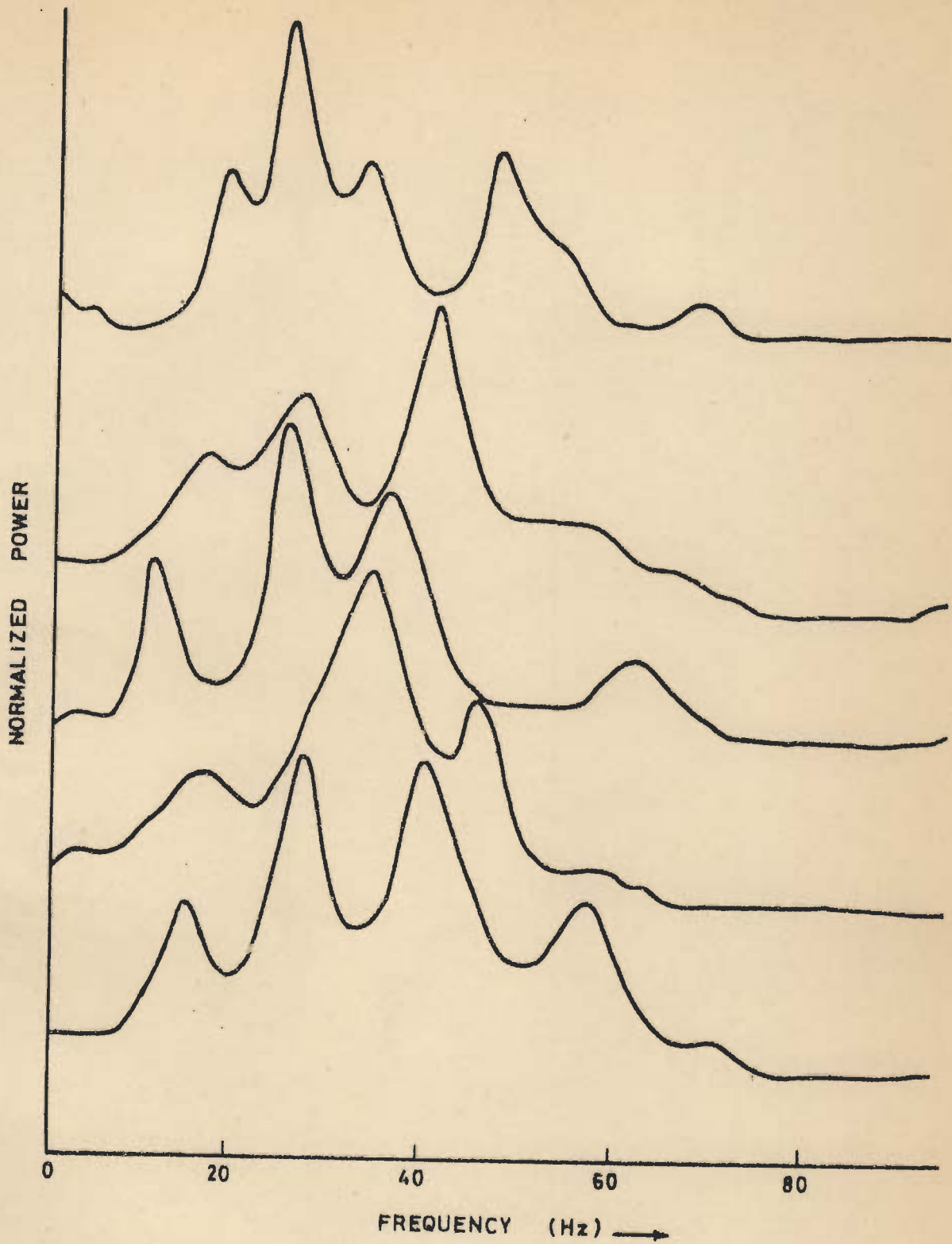
FIG. 4.11 _ SOME EXAMPLES FROM THE 255 TRACES OF POWER

SPECTRA ANALYSED FOR MODEL E.

FIG. 4.12 - SOME EXAMPLES FROM THE 255 TRACES OF POWER SPECTRA ANALYSED FOR MODEL E .
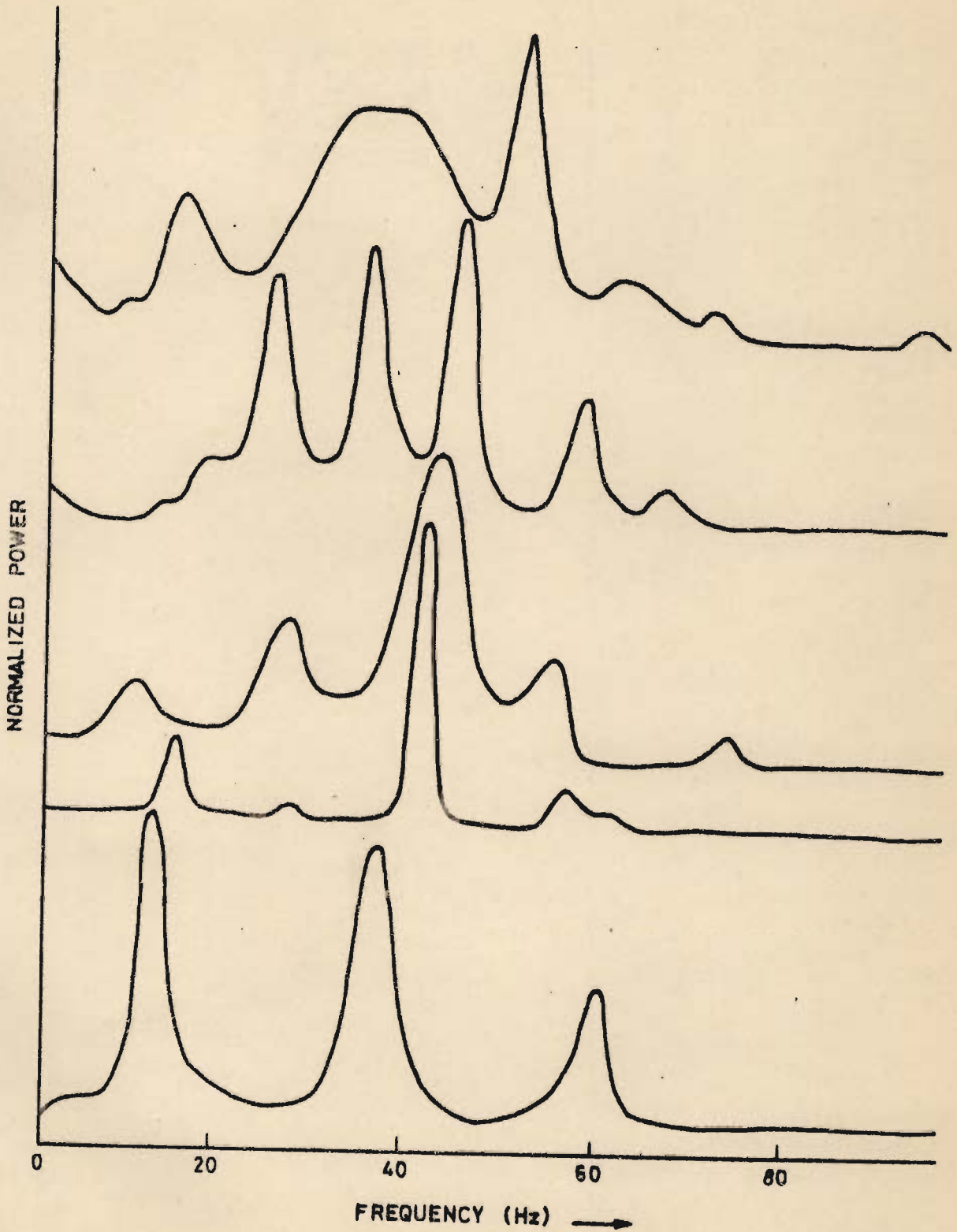
FIG. 4.13 _ SOME EXAMPLES FROM THE 255 TRACES OF POWER

SPECTRA ANALYSED FOR MODEL E .
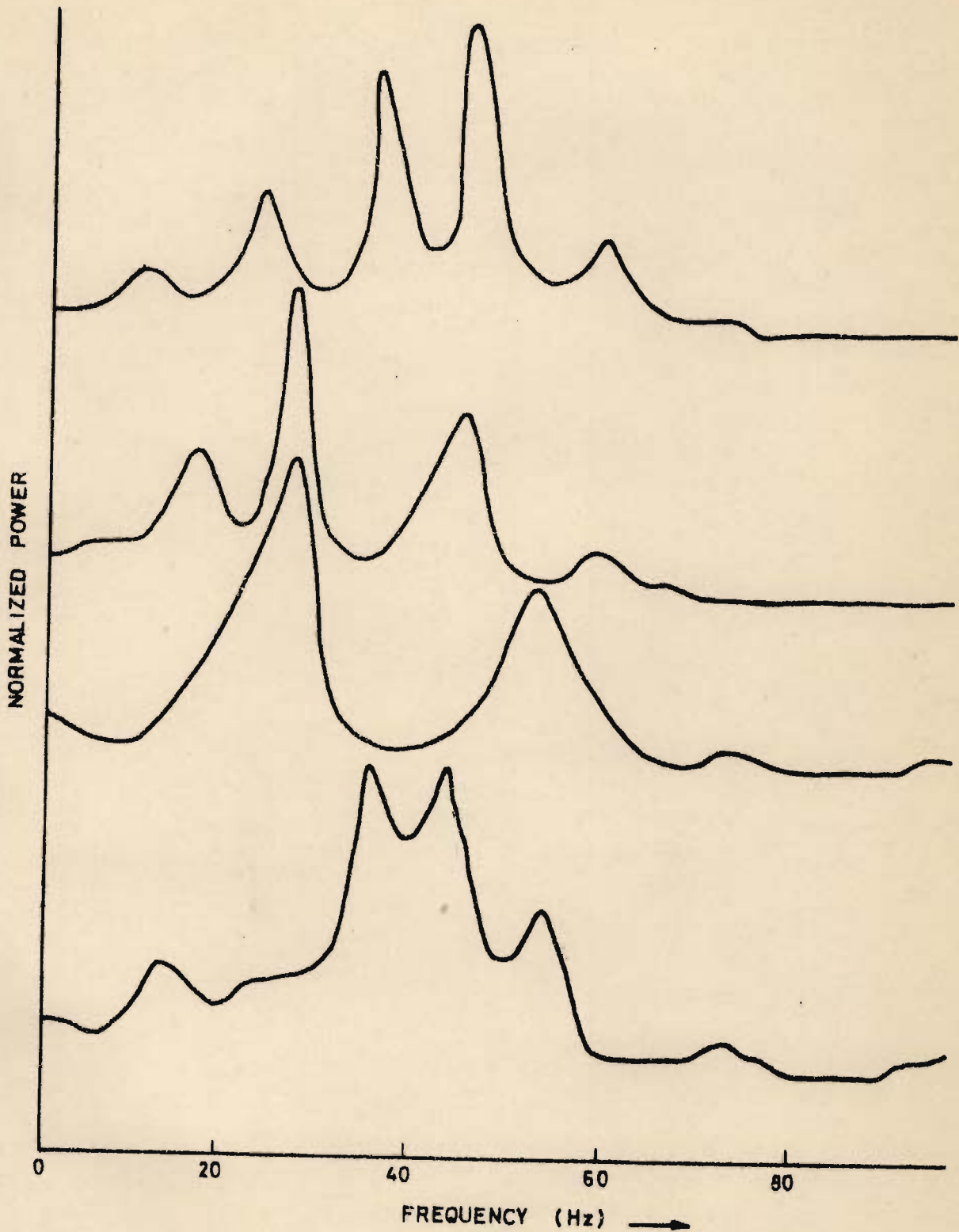
FIG. 4.14 _ SOME EXAMPLES FROM THE 253 TRACES OF POWER
SPECTRA ANALYSED FOR MODEL F.
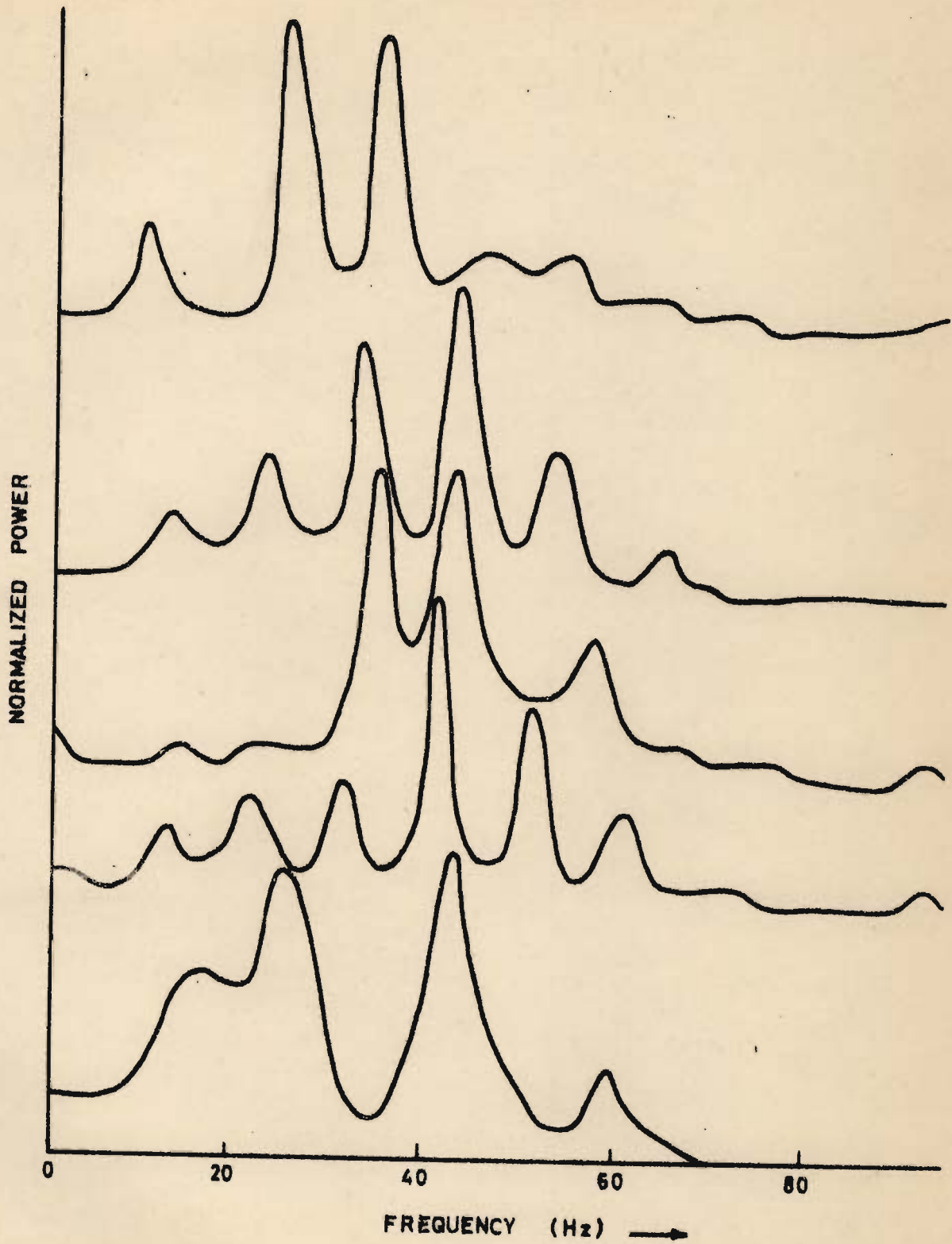
FIG.4.15_SOME EXAMPLES FROM THE 253 TRACES OF POWER SPECTRA ANALYSED FOR MODEL F.
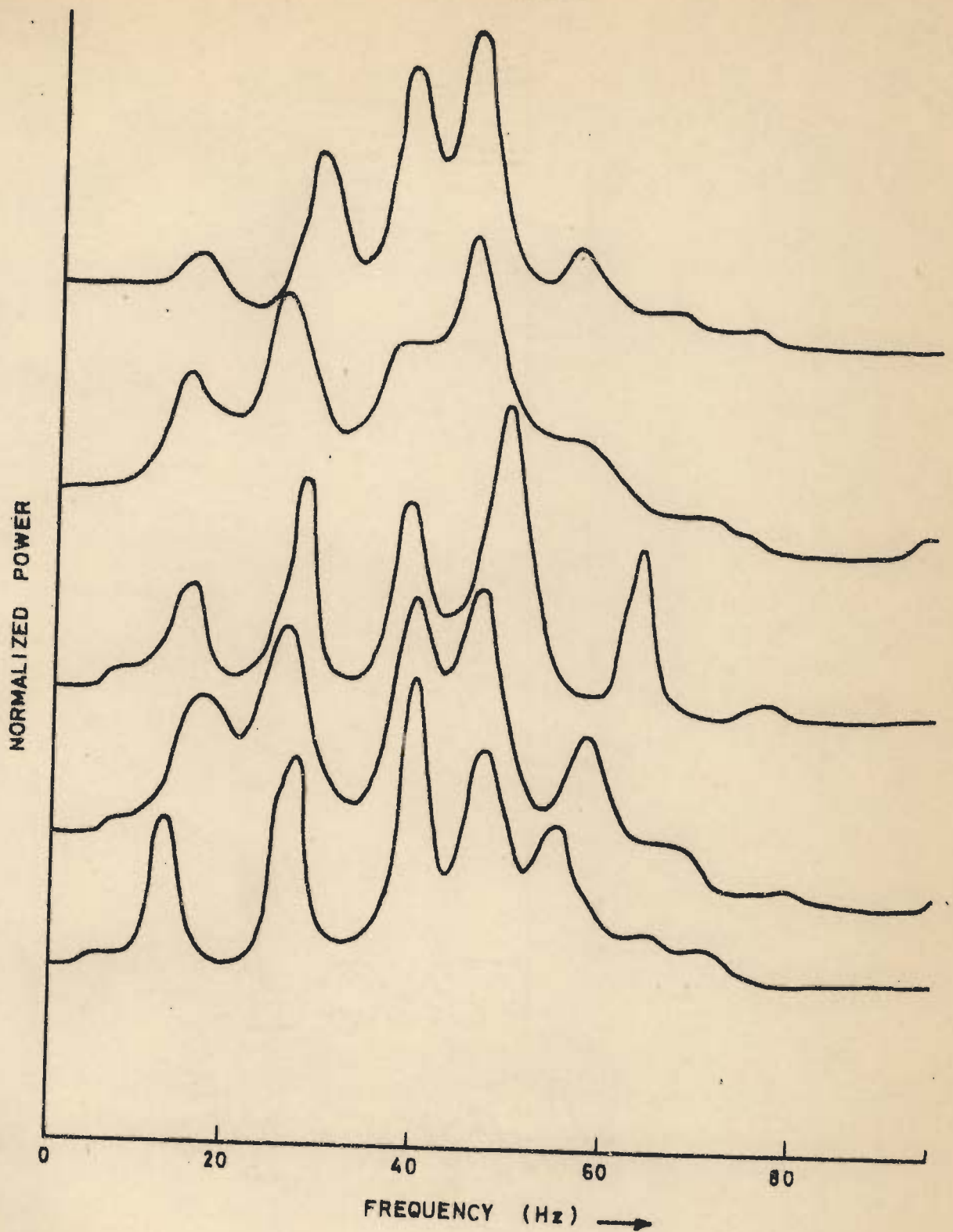
FIG.4.16_SOME EXAMPLES FROM THE 253 TRACES OF POWER SPECTRA ANALYSED FOR MODEL F.

It has till now not been possible to infer lithology directly from the power spectra, though Sinvhal (1976); Khattri, Sinvhal and Awasthi (1979) and Sinvhal, Gaur, Khattri, Moharir and Chander (1979) have been able to distinguish between a dominantly sandy and a coaly model (Table 4.3) by using the attributes of power spectra. It is therefore, plausible that the derivatives of the power spectrum, viz., the cumulative frequency weighted power spectrum, the cumulative power spectrum and the logarithm of power spectrum; and the variables selected from them may be used with the same aspirations.

Each power spectrum can be characterized by an average frequency f_1 which is dependent on both the frequency and the power content of the power spectrum. It is calculated by using the expression $(\sum_{i=1}^{n} P_i f_i)/(\sum_{i=1}^{n} P_i)$, where P is the Power and f is the frequency at the ith point for the n point power spectrum.

Some cumulative frequency weighted power spectra are shown in Figures 4.17 and 4.18 for Models E and F respectively. The variables f_2, f_3 and f_4 which signify the frequencies of 25th, 50th and 75th percentile values of frequency weighted power in that order are also shown in these figures. Because of the definition of these variables f_2 will be the lowest and f_4 the highest frequency amongst these three, with f_3 somewhere in between. Sometimes f_2 and f_3, or f_3 and f_4, and
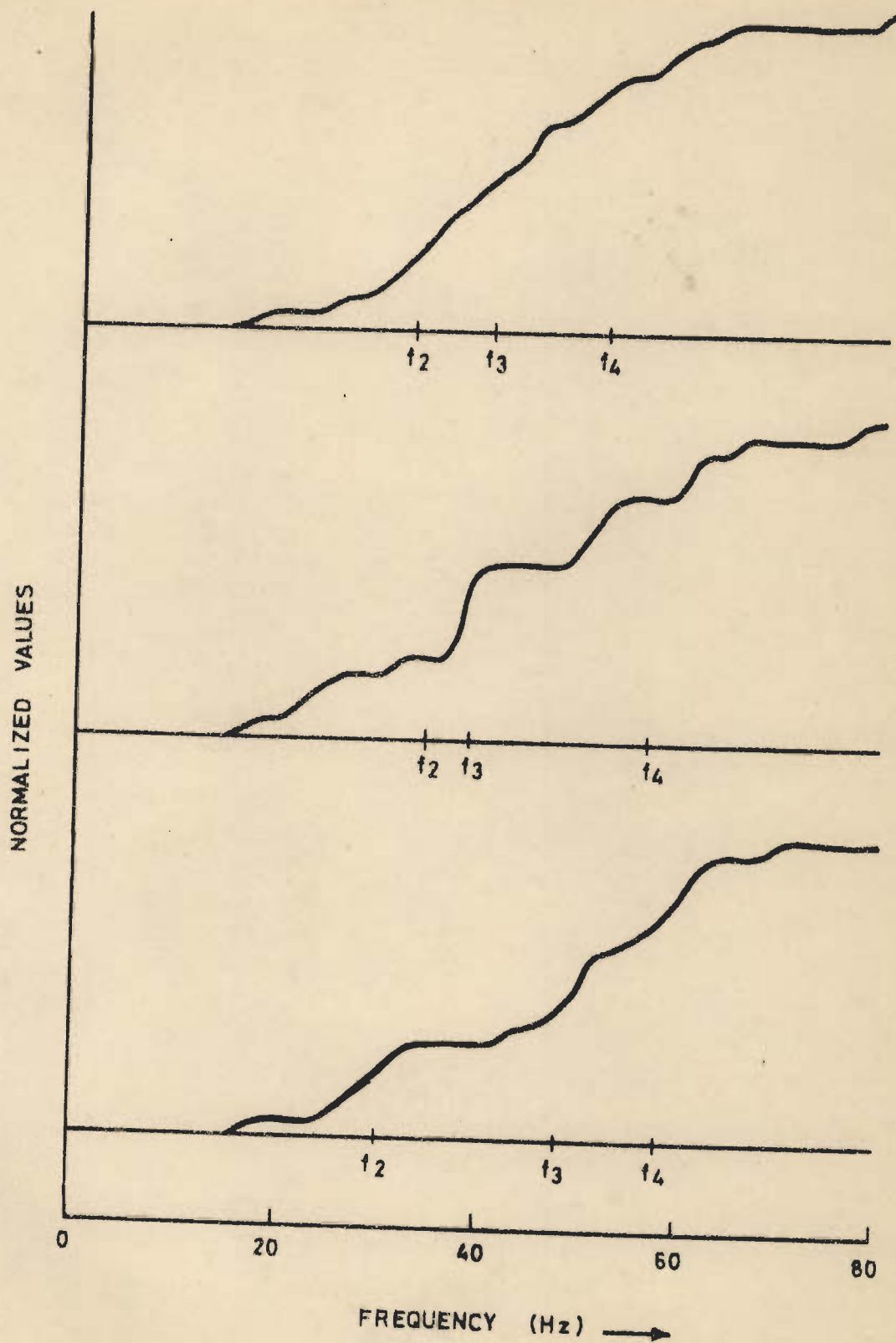
FIG. 4.17 _ SOME CUMULATIVE FREQUENCY WEIGHTED
POWER SPECTRA FOR MODEL E.
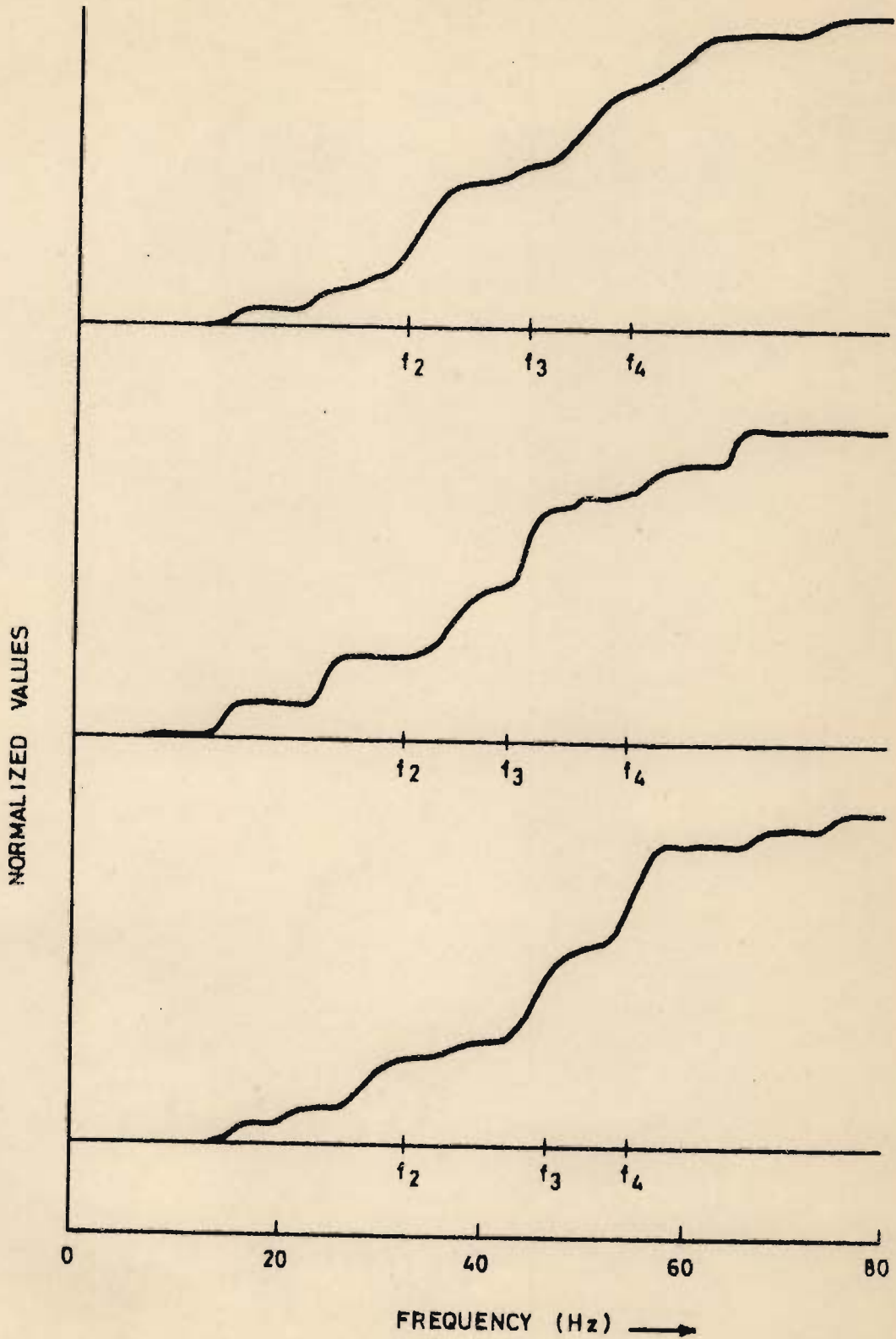
FIG. 4.18 _ SOME CUMULATIVE FREQUENCY WEIGHTED POWER SPECTRA FOR MODEL F.

in some rare cases all the three frequencies may coincide - the latter will indicate one very sharp peak in the spectrum, and consequently one big step in the cumulative frequency weighted power spectrum, which forces all the three frequencies to merge (with respect to the resolution in analysis).

Figures 4.19 and 4.20 show some examples from the 255 and 253 traces of frequency weighted power spectra analysed for Models E and F respectively. All these traces are flat for about the first 15 Hz and then have a gently upward sloping character. This begins to flatten out at about 65 Hz for Model E and at a somewhat higher frequency of about 70 Hz for Model F. The frequency f_4 may give a measure of this condition and may hold the clue to distinguish lithologies depicted by Models E and F. The frequency f_2 replaces the frequency f_e and avoids the arbitrariness involved in picking the latter.

Display of the same data in a different form may sometimes reveal features which were otherwise not obvious. The cumulative power spectra basically have the same information as the power spectra and the frequencies f_5, f_6 and f_7 which represent the 25th, 50th and 75th percentile of power, in that order, may be able to diagnose lithology. These frequencies are marked for 3 cumulative power spectrum traces in Figures 4.21 and 4.22 for Models E and F respectively. Figures 4.23 and 4.24 show some more examples from the 255 and 253 traces of cumulative power spectra analysed for Models E and F
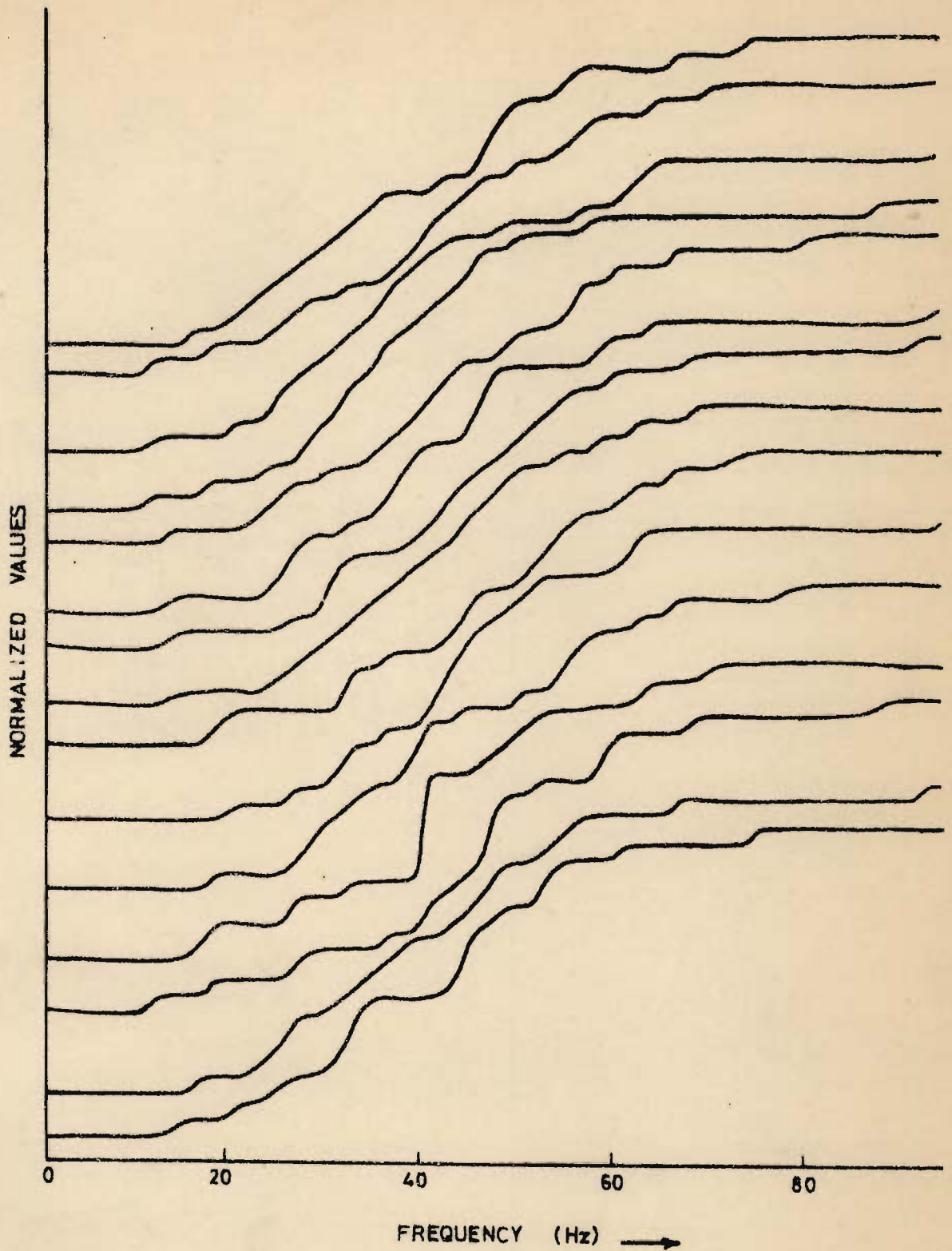
FIG. 4.19 _ SOME EXAMPLES FROM THE 255 TRACES OF FREQUENCY WEIGHTED POWER SPECTRA ANALYSED FOR MODEL E.
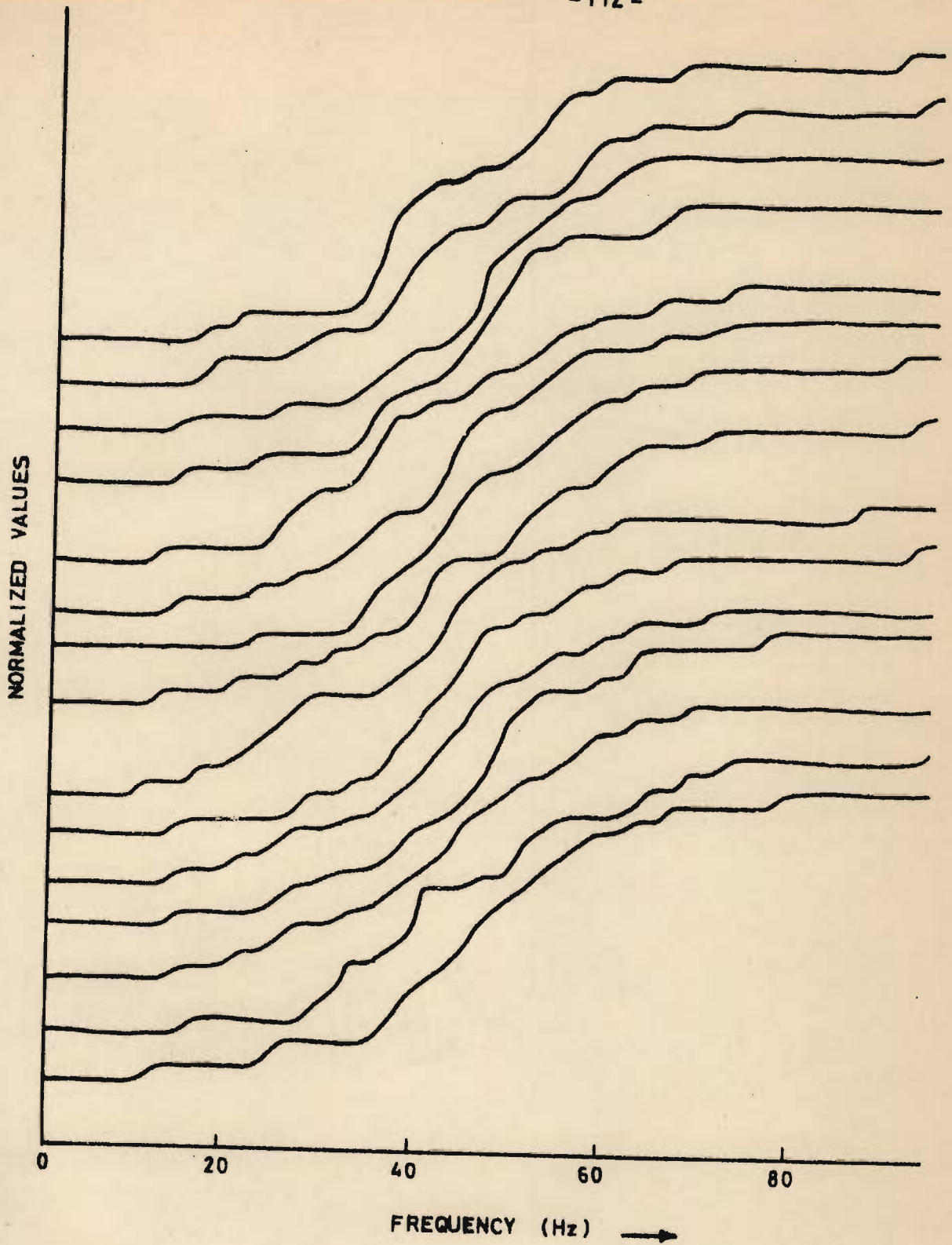
FIG. 4.20_ SOME EXAMPLES FROM THE 253 TRACES OF FREQUENCY WEIGHTED POWER SPECTRA ANALYSED FOR MODEL F.
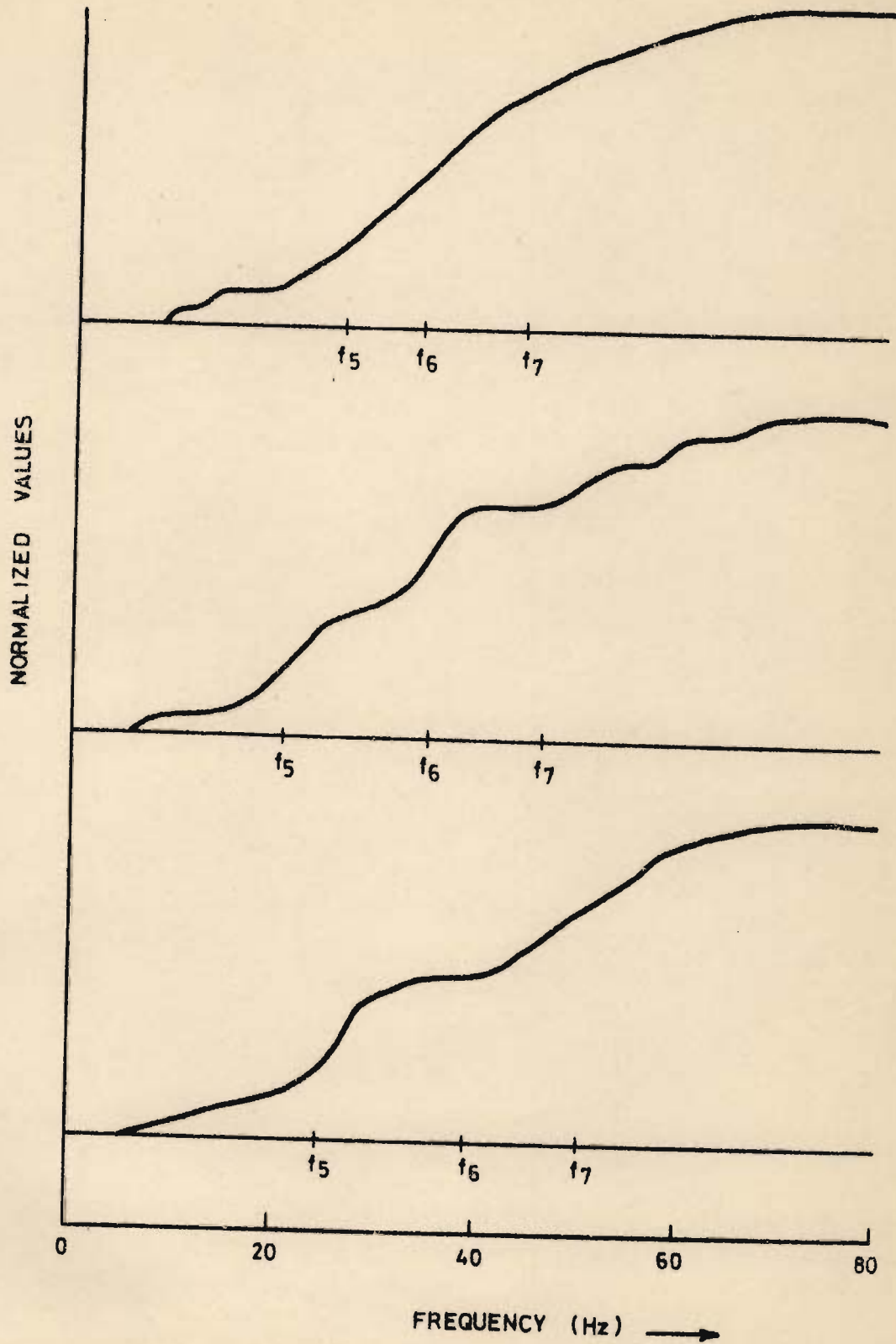
FIG. 4.21 _ SOME CUMULATIVE POWER SPECTRA FOR
MODEL E.
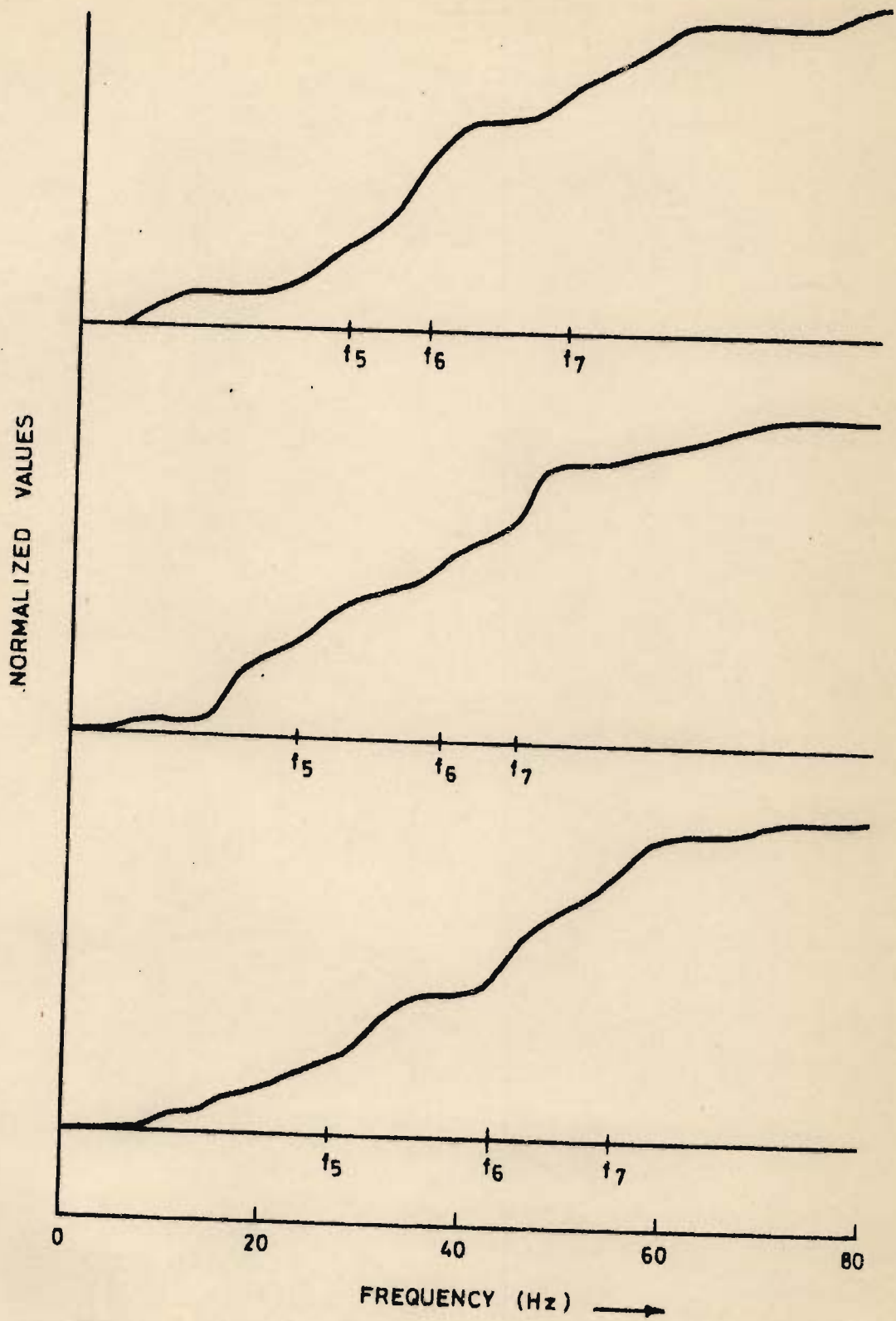
FIG. 4.22 _ SOME CUMULATIVE POWER SPECTRA FOR MODEL F.
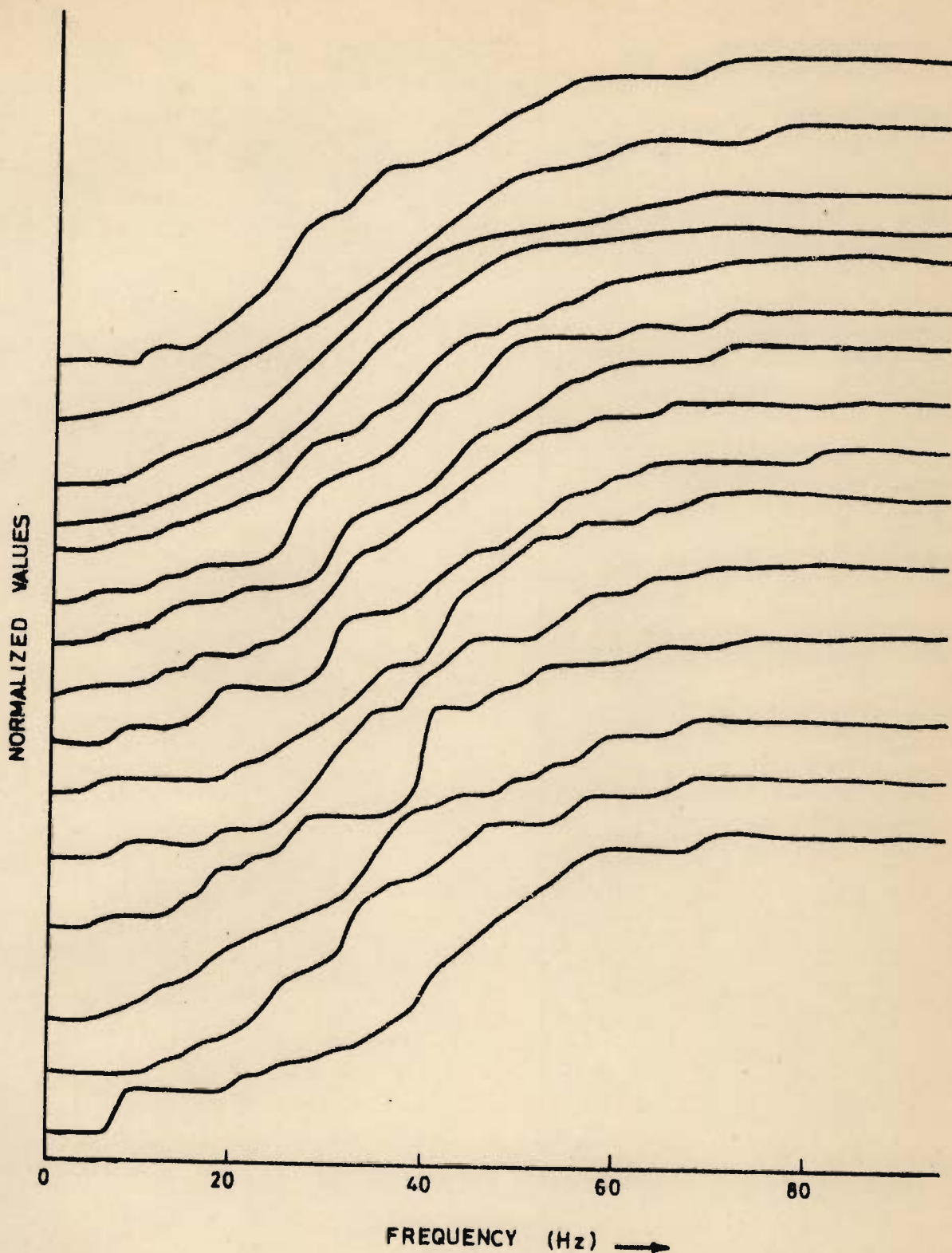
FIG.4.23_SOME EXAMPLES FROM THE 255 TRACES OF CUMULATIVE
POWER SPECTRA ANALYSED FOR MODEL E.
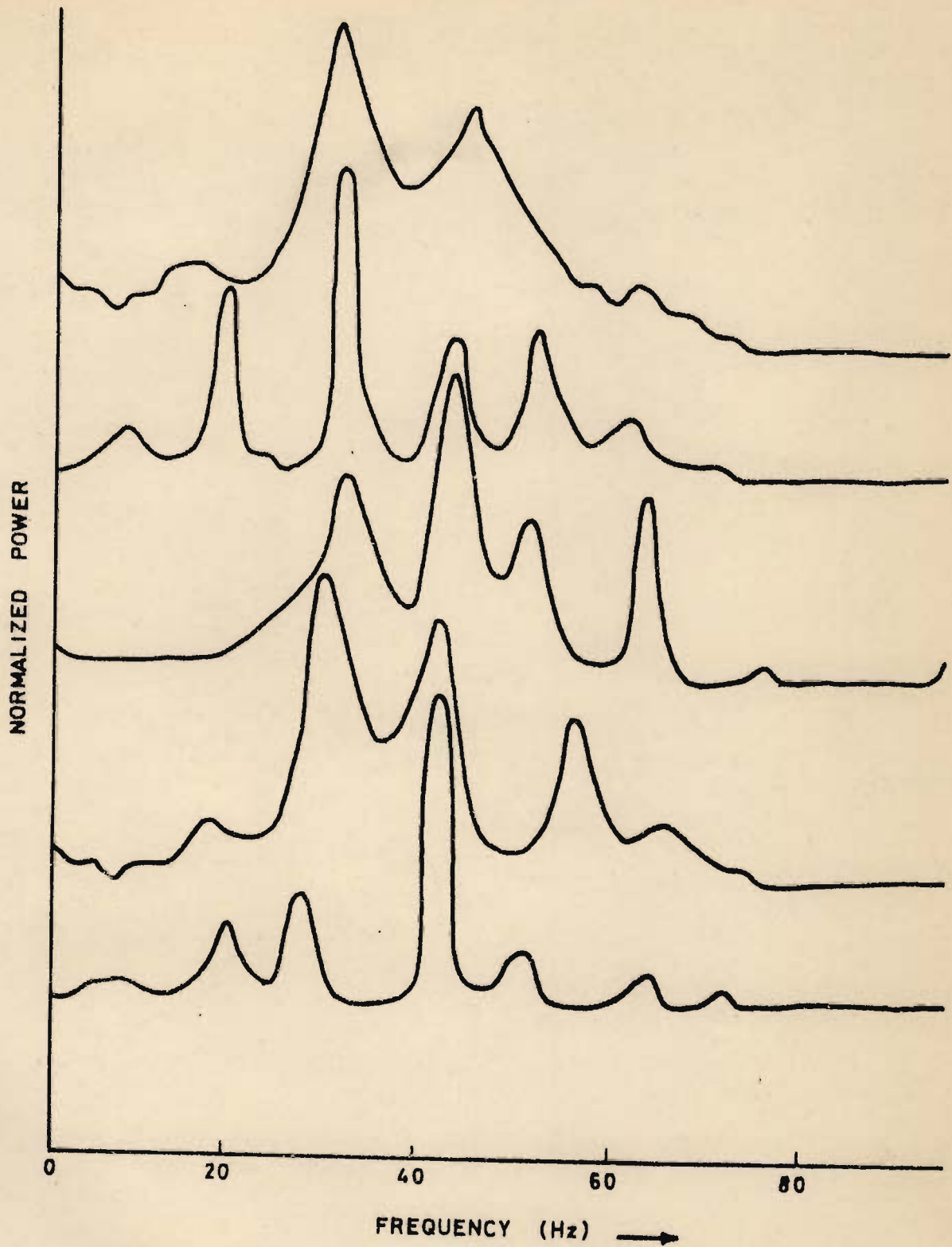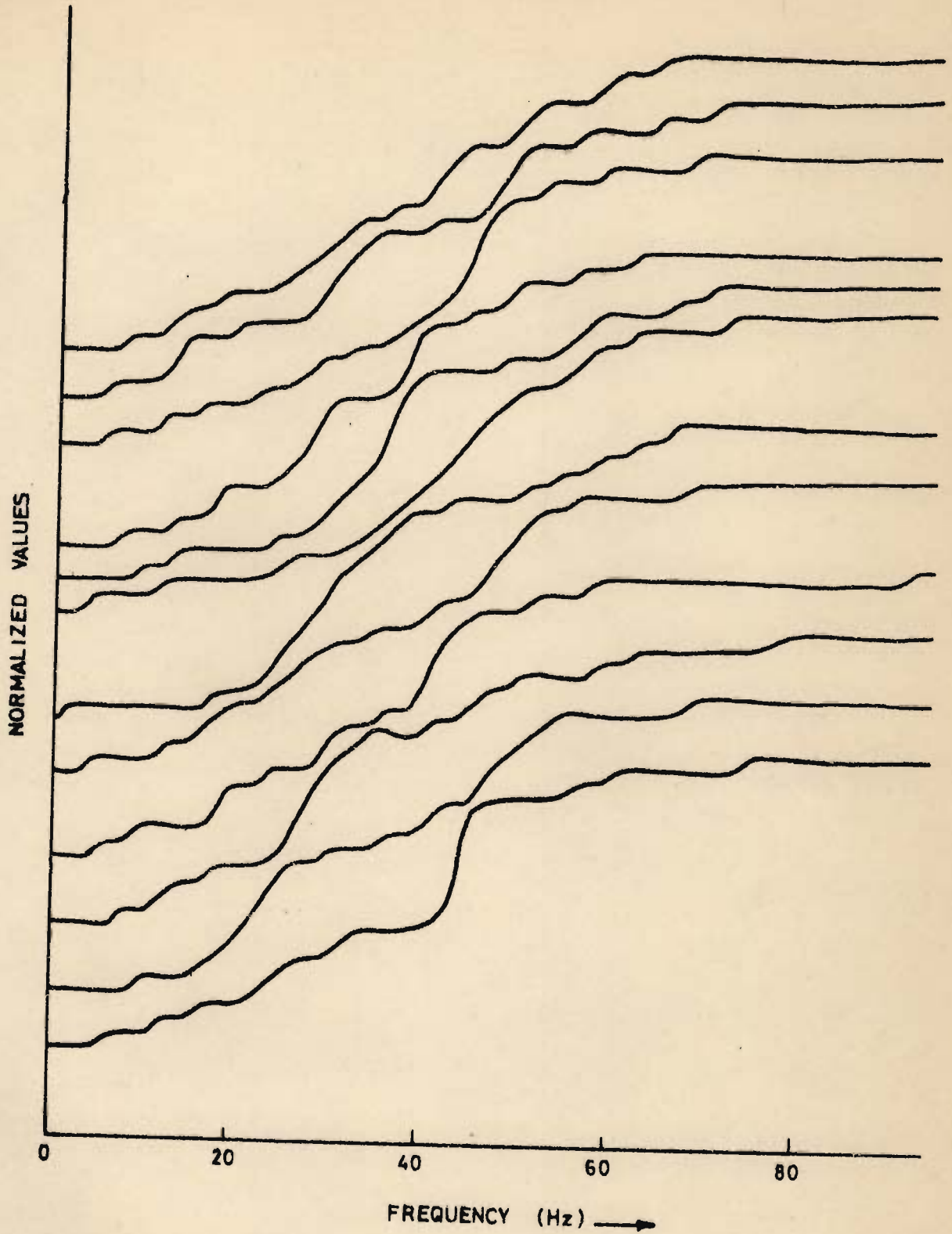
FIG.4.24 _ SOME EXAMPLES FROM THE 253 TRACES OF CUMULATIVE POWER SPECTRA ANALYSED FOR MODEL F.

respectively. The form of these traces is similar to those displayed in Figures 4.19 and 4.20, except that the curves begin to ascend at much lower frequencies.

The logarithm of power is yet another version of the enigmatic power spectrum and the frequency, f_8, at which the logarithm of power decreases to zero at the lowest frequency, is picked, with the oft repeated aim of discriminating lithology. Figures 4.25 and 4.26 show some log power spectra with the frequency f_8 marked on them, for Models E and F respectively. The points marked by $\times$ in Figure 4.25 could have well qualified for this distinction if the definition of f_8 did not include the terms 'decreases' and the lowest frequency. Figures 4.27 and 4.28 show some examples from the traces of logarithm of power spectra analysed for Models E and F respectively.

Seventeen variables have been computed with the aim that they will aid in the interpretation and discrimination of seismograms representing two sets of lithologies, one which is dominantly sandy and has a 53 percent sand, 26 percent shale and 21 percent coal constitution, i.e., Model E, and another which is dominantly shaly and has a 60 percent shale, 37 percent sand and 3 percent coal constitution, i.e., Model F. Discriminant analysis, given in the following Chapter has been carried out with this objective.
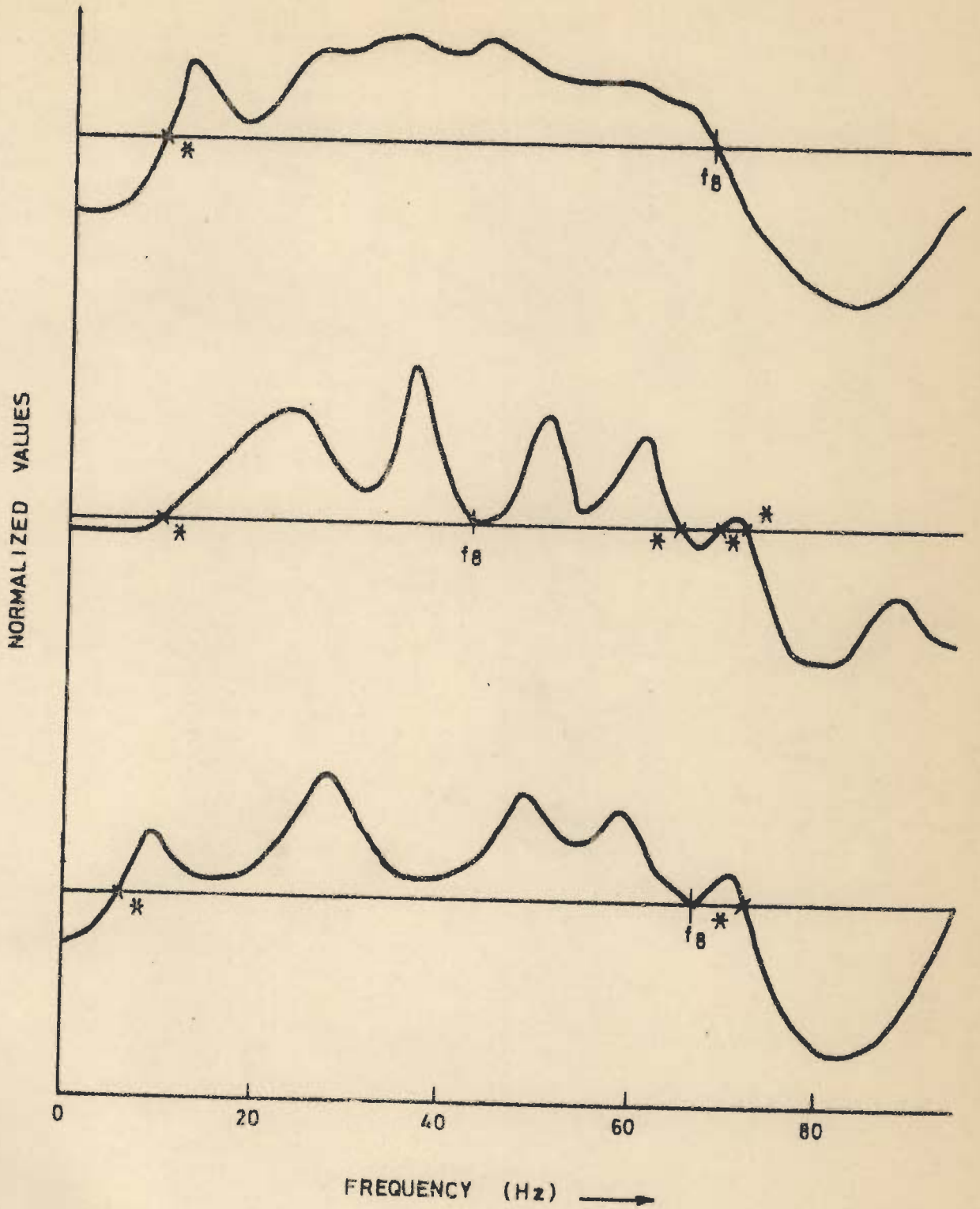
FIG.4.25_ SOME LOG POWER SPECTRA FOR MODEL E.
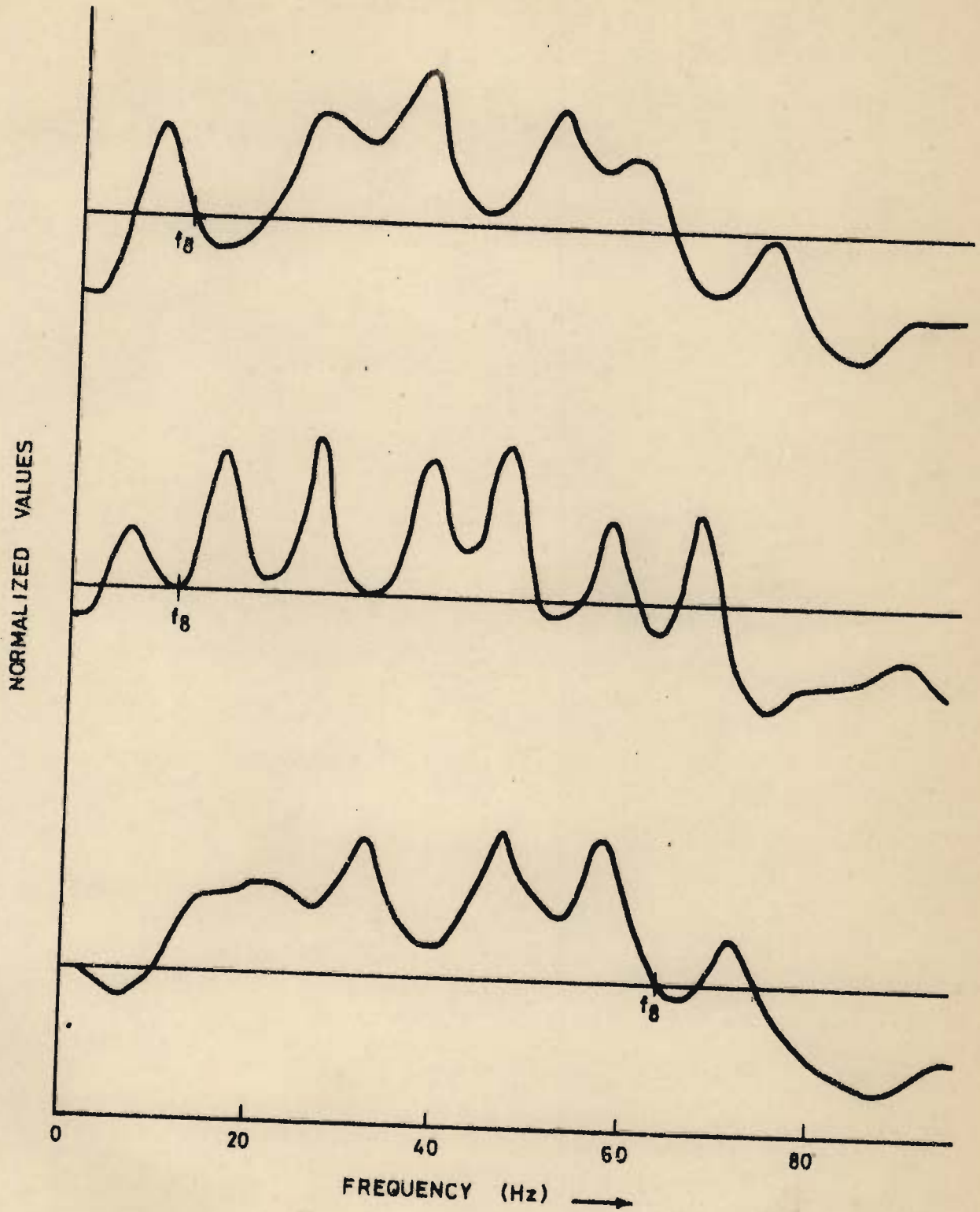
FIG.4.26_SOME LOG POWER SPECTRA FOR MODEL F.
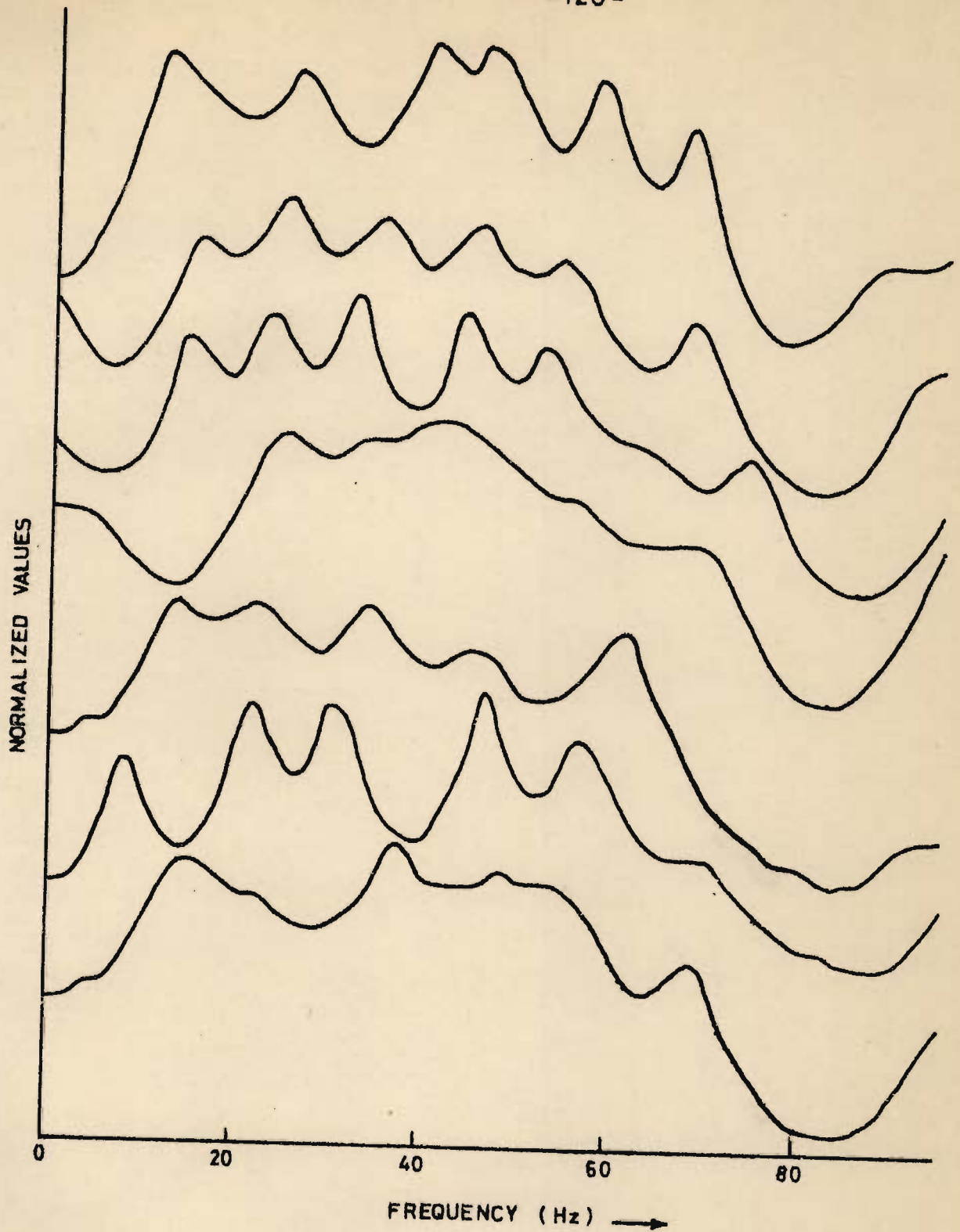
FIG. 4.27 _ SOME EXAMPLES FROM THE 255 TRACES OF LOGARITHM
OF POWER SPECTRA ANALYSED FOR MODEL E.
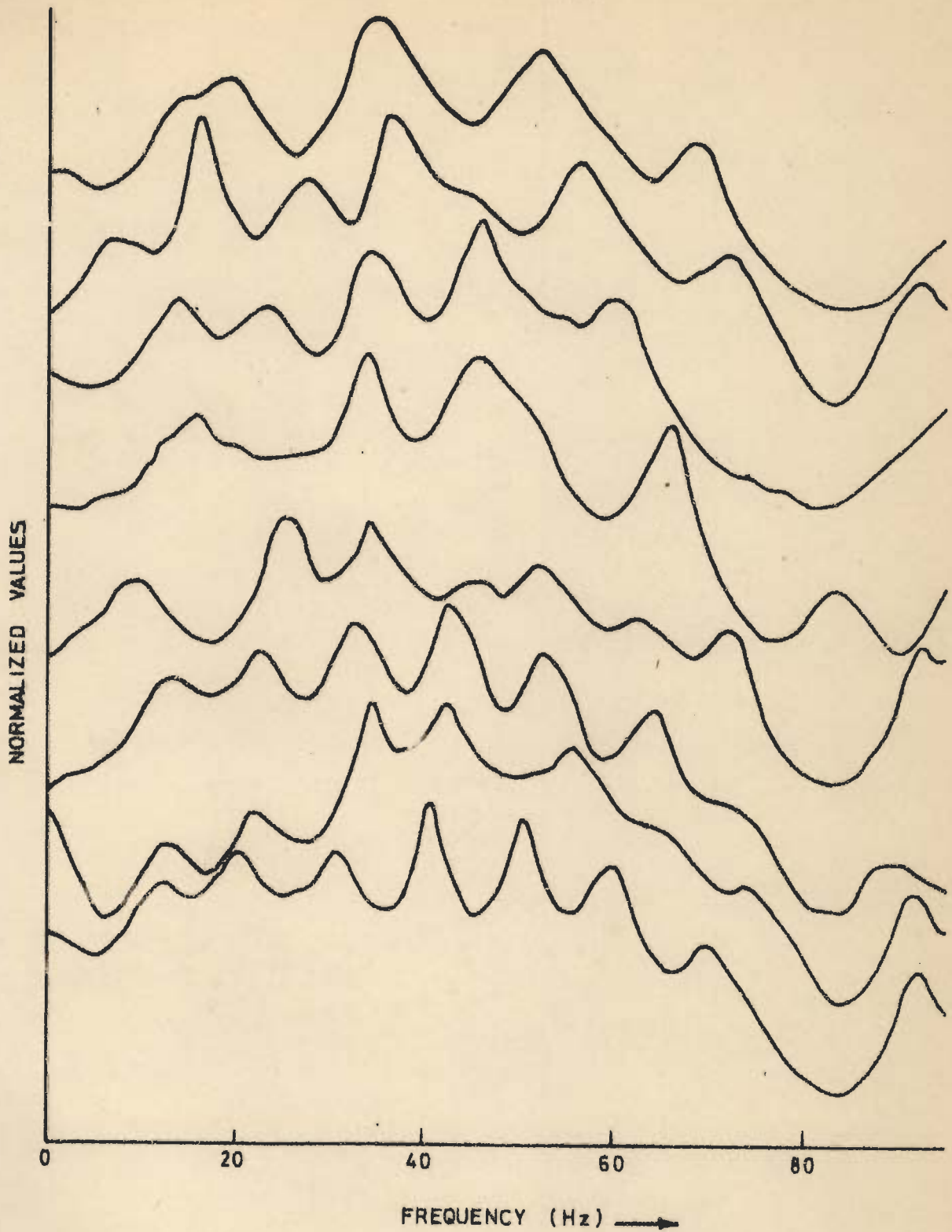
FIG. 4.28 _ SOME EXAMPLES FROM THE 253 TRACES OF LOGARITHM
OF POWER SPECTRA ANALYSED FOR MODEL F.

CHAPTER - V

DISCRIMINANT ANALYSIS

Linear discriminant function analysis is a multi-variate statistical technique of differentiating groups of samples drawn from different populations. This method was originally developed by Fisher (1936) and has been widely used in biometrics and paleobiometries. In geology it has been used to establish setting of sandstones (Middleton, 1962) and volcanics (Chayes, 1964) to classify depositional environments of carbonates (Krumbein and Graybill, 1965) and to distinguish between beach, shelf and fluvial depositional environments (Awasthi, 1979). In Geophysics, it has been used to distinguish dominantly sandy zones from shaly zones (Mathieu and Rice, 1969), pervious zones from impervious zones (Avasthi and Verma, 1973) and between dominantly sand, shale and coal sections (Sinvhal, Gaur, Khattri, Moharir and Chander, 1979).

A population, E, described by m variables may be pictured as a cluster of sample points in m dimensional space. A second population, F, described by the same m variables, consists of a second cluster of points. Discriminant analysis is the computation of a m-dimensional plane that most effectively separates the two clusters. An unknown sample is classified as belonging to one group or the other, depending on which side of the plane it falls. The degree of distinctness of the two groups is measured by the distance between the two

population means.

Five hundred and eight stratigraphic sequences depicting two different kinds of lithologies with sand, shale and coal sequences have been generated as discussed in Chapter II. These are classified as : Model E - which represents 53 percent sand, 26 percent shale and 21 percent coal; and Model F which represents 60 percent shale, 37 percent sand and 3 percent coal. The seismic response of these in time and frequency domain are calculated as discussed in Chapter III, and 17 features have been abstracted from these as discussed in Chapter IV. On the basis of these 17 variables an attempt has been made to distinguish between these two models.

5.1 MATHEMATICAL DEVELOPMENT

For a general case consider m variables that are common to both models, as it is an essential requirement of this formulation. Let there be p and q seismograms of Models E and F, respectively. The first set of assumptions in this approach are that the seismograms in each model are randomly chosen, and it is known for certain that these belong to their respective models. The variables used to discriminate between the two Models E and F should be independent and normally distributed. These variables may be denoted by the following notations :

e_{11}, e_{12}, e_{13}, $\cdots$, e_{1m} as the measured variables of the first seismogram of Model E,

e_{21}, e_{22}, e_{23}, $\cdots$, e_{2m} as the measured variables of the second seismogram of Model E,

$\cdot$
$\cdot$

e_{i1}, e_{i2}, e_{i3}, $\cdots$, e_{im} as the measured variables of the ith seismogram of Model E,

$\cdot$
$\cdot$
$\cdot$

e_{p1}, e_{p2}, e_{p3}, $\cdots$, e_{pm} as the measured variables of the pth seismogram of Model E.

The first subscript denotes the seismogram index and the second subscript is the variable index. e_{ik} denotes the kth variable of the ith seismogram for Model E. Similarly for Model F, f_{ik} denotes the kth variable of the seismogram for Model F. Each seismic realization can be represented as a linear combination, E_i ($i = 1, 2, \cdots, p$), of the m variables as follows :

$$E_1 = \lambda_1 e_{11} + \lambda_2 e_{12} + \cdots + \lambda_m e_{1m}$$
$$E_2 = \lambda_1 e_{21} + \lambda_2 e_{22} + \cdots + \lambda_m e_{2m}$$

$\cdot$
$\cdot$
$\cdot$
$\cdot$

$$E_p = \lambda_1 e_{p1} + \lambda_2 e_{p2} + \cdots + \lambda_m e_{pm}$$

i.e.,

$$E_i = \sum_{k=1}^{m} \lambda_k e_{ik} \qquad \cdots (5.1)$$

or,

$$\sum_{i=1}^{p} E_i = \sum_{i=1}^{p} \left(\sum_{k=1}^{m} \lambda_k e_{ik} \right)$$

Similarly for Model F the linear combination of m variables is given by

$$F_1 = \lambda_1 f_{11} + \lambda_2 f_{12} + \dots + \lambda_m f_{1m}$$

$$F_2 = \lambda_1 f_{21} + \lambda_2 f_{22} + \dots + \lambda_m f_{2m}$$

$$\vdots$$

$$F_q = \lambda_1 f_{q1} + \lambda_2 f_{q2} + \dots + \lambda_m f_{qm}$$

i.e.,

$$F_i = \sum_{k=1}^{m} \lambda_k f_{ik} \qquad \dots (5.2)$$

$$\sum_{i=1}^{q} F_i = \sum_{i=1}^{q} \left(\sum_{k=1}^{m} \lambda_k f_{ik} \right)$$

The λ coefficients in equations (5.1) and (5.2) must be determined such that discrimination between the two Models E and F can be optimized. Fisher's (1936) criterion for this is to find a particular function which maximizes the ratio of the difference between the means of the two models to their standard deviation. This will make all the observations of one model come close together and increase the separation between the two models. Figure 5.1 shows the variables 1 and 2
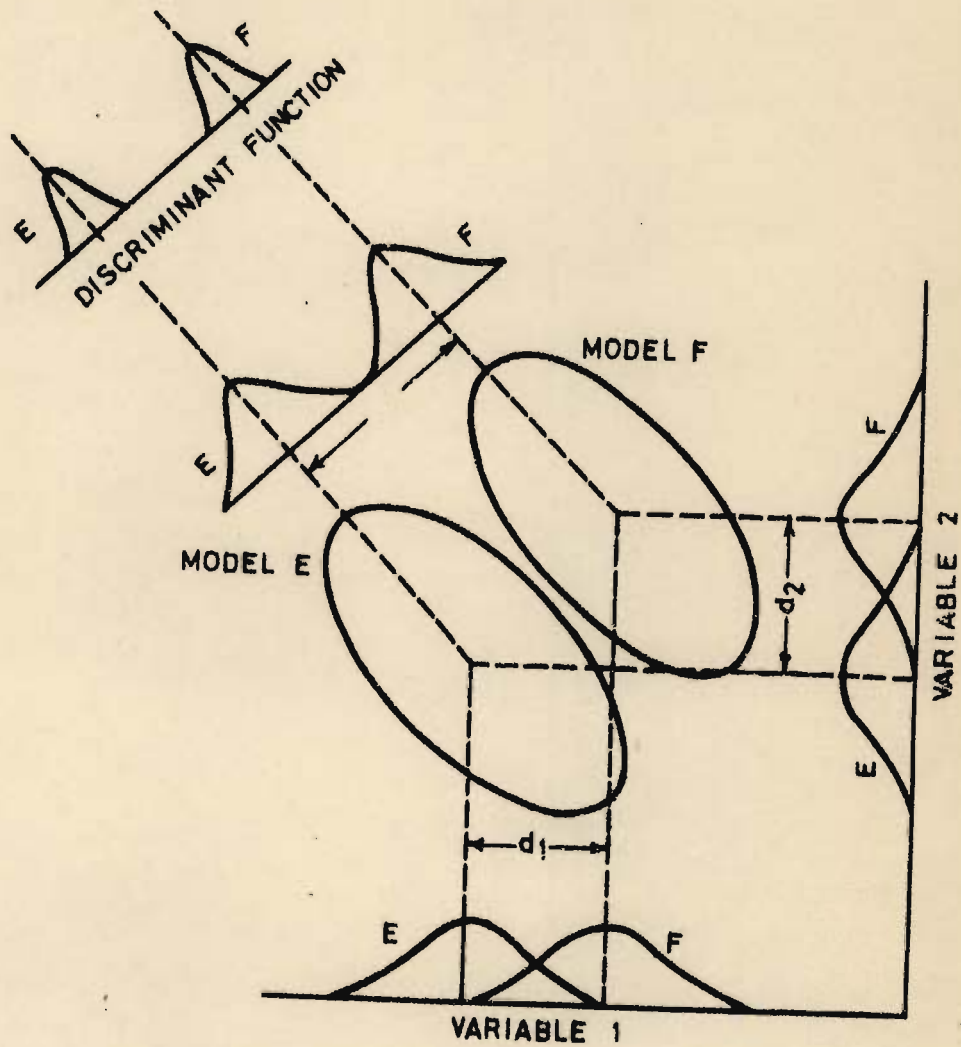
FIG.5.1 _ PLOTS SHOWING OVERLAP BETWEEN MODELS
E AND F ALONG VARIABLES 1 AND 2. MODELS CAN
BE SEPARATED WHEN TWO VARIABLES ARE
CONSIDERED SIMULTANEOUSLY.

(MODIFIED AFTER DAVIS, 1973)

when considered separately overlap and fail to distinguish the
two models E and F, but when considered simultaneously the two
clusters of points belonging to these models are separated by
a distance Δ. Assuming equality of the two variance-covariance
matrices the particular linear function which has to be maximiz-
ed with respect to the λs can be taken as

$$\Delta^2/S^2 = (\overline{E} - \overline{F})^2 / \left[\sum_{i=1}^{p} (E_i - \overline{E})^2 + \sum_{i=1}^{q} (F_i - \overline{F})^2 \right] \quad \ldots (5.3)$$

where S^2 is the pooled sum of squares of deviations from mean
of new variables E_i and F_i of Models E and F, respectively. The
respective means $\overline{E}$ and $\overline{F}$ are defined as follows:

$$\overline{E} = \left(\sum_{i=1}^{p} E_i \right) / p \qquad \qquad \ldots (5.4)$$

$$\overline{F} = \left(\sum_{i=1}^{q} F_i \right) / q \qquad \qquad \ldots (5.5)$$

Δ signifies the difference between the means of the new
variables E_i and F_i for the two models formed by linear combi-
nation of the original variables,

$$\Delta = \overline{E} - \overline{F}$$

$$= \left(\sum_{i=1}^{p} E_i \right) / p - \left(\sum_{i=1}^{q} F_i \right) / q \qquad \ldots (5.6)$$

Substituting the value of E_i and F_i from equation (5.1) and
(5.2) in (5.6) Δ becomes

$$\Delta = (\sum_{i=1}^{p} \sum_{k=1}^{m} \lambda_k e_{ik})/p - (\sum_{i=1}^{q} \sum_{k=1}^{m} \lambda_k f_{ik})/q$$

$$= \sum_{k=1}^{m} \lambda_k \left[(\sum_{i=1}^{p} e_{ik})/p - (\sum_{i=1}^{q} f_{ik})/q \right]$$

$$= \sum_{k=1}^{m} \lambda_k \left[\overline{E}_k - \overline{F}_k \right]$$

$$= \sum_{k=1}^{m} \lambda_k d_k$$

where

$$d_k = \overline{E}_k - \overline{F}_k$$

Therefore,

$$\Delta = \lambda_1 d_1 + \lambda_2 d_2 + \lambda_3 d_3 + \ldots + \lambda_m d_m \qquad \ldots (5.7)$$

Let the variances of variables E_i and F_i corresponding to the Models E and F be given by s^2_E and s^2_F respectively, then,

$$s^2_E = \left[\sum_{i=1}^{p} (E_i - \overline{E})^2 \right]/(p-1)$$

or,

$$\sum_{i=1}^{p} (E_i - \overline{E})^2 = s^2_E(p-1)$$

and similarly

$$\sum_{i=1}^{q} (F_i - \overline{F})^2 = s^2_F(q-1)$$

Thus, the pooled sum of squares s^2 is given by

$$S^2 = (p-1) \cdot s^2_E + (q-1) \, s^2_F$$

$$= \sum_{i=1}^{p} (E_i - \overline{E})^2 + \sum_{i=1}^{q} (F_i - \overline{F})^2 \qquad \ldots (5.8)$$

S^2 is partitioned into two sum of squares which can be put in product form as given below :

From equations (5.1) and (5.4)

$$\sum_{i=1}^{p} (E_i - \overline{E})^2 = \sum_{i=1}^{p} \left[\sum_{k=1}^{m} \lambda_k \, e_{ik} - \frac{\sum_{i=1}^{p} \sum_{k=1}^{m} \lambda_k \, e_{ik}}{p} \right]^2$$

$$= \sum_{i=1}^{p} \left[\sum_{k=1}^{m} \lambda_k \, (e_{ik} - \overline{E}_k) \right]^2$$

$$= \sum_{i=1}^{p} \left[\sum_{k=1}^{m} \sum_{j=1}^{m} \lambda_k \, \lambda_j \, (e_{ik} - \overline{E}_k)(e_{ij} - \overline{E}_j) \right]$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{m} \lambda_k \, \lambda_j \left[\sum_{i=1}^{p} (e_{ik} - \overline{E}_k)(e_{ij} - \overline{E}_j) \right]$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{m} \lambda_k \, \lambda_j \, SE_{kj} \qquad \ldots (5.9)$$

where

$$SE_{kj} = \sum_{i=1}^{p} (e_{ik} - \overline{E}_k) \, (e_{ij} - \overline{E}_j)$$

Similarly,

$$\sum_{i=1}^{q} (F_i - \overline{F})^2 = \sum_{k=1}^{m} \sum_{j=1}^{m} \lambda_k \, \lambda_j \, SF_{kj} \qquad \ldots (5.10)$$

where

$$SF_{kj} = \sum_{i=1}^{q} (f_{ik} - \overline{F}_k)(f_{ij} - \overline{F}_j)$$

Substituting (5.9) and (5.10) in (5.8)

$$S^2 = \sum_{k=1}^{m} \sum_{j=1}^{m} \lambda_k \lambda_j (SE_{kj} + SF_{kj})$$

$$= \sum_{k=1}^{m} \sum_{j=1}^{m} \lambda_k \lambda_j S_{kj} \qquad \qquad \ldots (5.11)$$

Where S_{kj} is defined by

$$S_{kj} = SE_{kj} + SF_{kj}$$

The particular linear function which best discriminates the two models will be one for which the ratio Δ^2/S^2 is maximum. Hence, the function Δ^2/S^2 is maximized with respect to the λ_ks. Therefore Δ^2/S^2 is differentiated with respect to λ_ks and set to zero.

$$\frac{d}{d\lambda_k}(\Delta^2/S^2) = [2\Delta . S^2(d\Delta/d\lambda) - 2S.\Delta^2.(dS/d\lambda_k)] / S^4$$

$$= (2\Delta/S^3). [S.(2\Delta/d\lambda_k) - \Delta(dS/d\lambda_k)]$$

$$= 0 \text{ for maximization.}$$

If Δ/S^3 is zero it means that either Δ, the distance between the two models, would be zero and this defeats the very aim of maximizing, or else S^3, which denotes variability, is infinity, which goes against the philosophy of minimizing the variance. This trivial solution is therefore rejected. Therefore, the acceptable solution is:

$$[S \cdot (d\Delta/d\lambda_k) - \Delta \cdot (dS/d\lambda_k)] = 0$$

which gives $S/\Delta \cdot (d\Delta/d\lambda_k) = (dS/d\lambda_k)$

After maximization the values of λ_ks are fixed, which makes S and Δ individually constants, and therefore S/Δ may be taken as a constant which does not affect the other terms in the above equation. Therefore, the solutions are proportional to this term.

$$(dS/d\lambda_k) = (d\Delta/d\lambda_k) \qquad \ldots (5.12)$$

Using equation (5.11) the left hand side of (5.12) becomes

$$(dS/d\lambda_1) = 2(\lambda_1 S_{11} + \lambda_2 S_{12} + \ldots + \lambda_m S_{1m})$$
$$(dS/d\lambda_2) = 2(\lambda_1 S_{21} + \lambda_2 S_{22} + \ldots + \lambda_m S_{2m})$$

.
.
.

$$(dS/d\lambda_m) = 2 (\lambda_1 S_{m1} + \lambda_2 S_{m2} + \ldots + \lambda_m S_{mm})$$

Similarly, using equation (5.7), the right hand side of (5.12) becomes

$$(d\Delta/d\lambda_k) = d_k$$

for $k = 1, 2, \ldots, m$

Hence, equation (5.12) becomes

$$\lambda_1 S_{11} + \lambda_2 S_{12} + \ldots + \lambda_m S_{1m} = d_1$$
$$\lambda_1 S_{21} + \lambda_2 S_{22} + \ldots + \lambda_m S_{2m} = d_2$$
$$\vdots$$
$$\lambda_1 S_{m1} + \lambda_2 S_{m2} + \ldots + \lambda_m S_{mm} = d_m$$

When k = j, the variance S_{kk} of the kth variable is obtained, and when k ≠ j, the covariance S_{kj} between the variables k and j is obtained. In matrix notation equation (5.12) can be written as

$$
\begin{bmatrix}
S_{11} & S_{12} & S_{13} & \cdots & S_{1m} \\
S_{21} & S_{22} & S_{23} & \cdots & S_{2m} \\
\cdot \\
\cdot \\
\cdot \\
S_{m1} & S_{m2} & S_{m3} & \cdots & S_{mm}
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\cdot \\
\cdot \\
\cdot \\
\lambda_m
\end{bmatrix}
=
\begin{bmatrix}
d_1 \\
d_2 \\
\cdot \\
\cdot \\
\cdot \\
d_m
\end{bmatrix}
\qquad \cdots (5.13)
$$

λ_ks can be obtained from the above set of equations by the Gauss-Jordan elimination method.

5.2 DISCRIMINANT FUNCTION

Following Davis and Sampson (1966) a linear discriminant function R may be defined as

$$R = \sum_{k=1}^{m} \lambda_k \, \psi_k \qquad \cdots (5.14)$$

Where R is known as the discriminant score, and ψ_k are the values e_{ik} or f_{ik} of k variables (in this case seismic responses) of Models E or F. By substituting values of ψ_ks, p and q discriminant scores for Models E and F are obtained, respectively. Let the mean discriminant score for the two models be denoted by R_E and R_F, i.e.,

$$R_E = \left(\sum_{i=1}^{p} R_i \right) / p \qquad \cdots (5.15)$$

and

$$R_F = (\sum_{i=1}^{q} R_i)/q \qquad\qquad \cdots (5.16)$$

Where R_is are taken for the respective groups, then

$$\Delta = R_E \sim R_F \qquad\qquad \cdots (5.17)$$

Where Δ is the maximized distance obtained between the two models, which is equal to Mahalanobis' D^2, a measure of distance between the model means (Davis 1973). To separate the two models a discriminant index R_o is defined as

$$R_o = (R_E + R_F)/2 \qquad\qquad \cdots (5.18)$$

R_o enables to classify a new seismogram to either of the models E and F, provided there is a priori knowledge that it belongs to either of the two models. To test the null hypothesis of equality of multivariate means of the Models E and F, an 'F' test where

$$F_{m, p+q-m-1} = \frac{p \cdot q}{(p+q)(p+q-2)} \cdot \frac{(p+q-m-1)}{m} \cdot D^2 \qquad \cdots (5.19)$$

with m and $(p+q-m-1)$ degrees of freedom is applied. Therefore, when the calculated value is larger than the tabulated value of F, then the multivariate means of the two models are drawn from different populations, i.e., the result of the discriminant analysis is meaningful. If this is not the case, then the multivariate means for the two models are drawn from the same population which renders discriminant analysis meaningless, as

the multivariate means of the variables belong to the same parent population irrespective of the model from which the variables are drawn.

The relative contribution of variable j to the distance between the two model means is measured by a quantity E_j,

$$E_j = \frac{\lambda_j \, d_j}{D^2} \cdot 100 \qquad\qquad \dots (5.20)$$

where d_j is the difference between the jth means of the two groups, and is a measure of the direct contribution of the variable j which does not consider interaction between variables.

The equality of the variance - covariance matrices of the two populations is tested by the following statistic (Seal, 1964):

$$\chi^2 = -2 \left[1 - \left(\sum_{j=1}^{k} \frac{1}{N_j - j} - \frac{1}{N-k} \right) \cdot \frac{2\,m^2 + 3m - 1}{6(m+1)(k-1)} \right] \ln \left[\frac{\prod_{j=1}^{k} \Delta_j^{(N_j-j)/2}}{\Delta^{(N-k)/2}} \right]$$

$$\dots (5.21)$$

where,

   k = number of populations,

   N = total sample size,

   N_j = sample size of the jth population,

   m = number of variables,

   Δ_j = determinant of the jth covariance matrix,

   Δ = pooled covariance matrix.

The above statistic is distributed approximately as chi-squared with $(k-1)$ m $(m+1)/2$ degrees of freedom. Large values reject the null hypothesis which is the equality of covariance matrices of k populations.

5.3 DISCRIMINANT ANALYSIS OF SYNTHETIC DATA

Before embarking on discriminant analysis the assumption of the equality of the variance-covariance matrices is tested by using equation(5.21). The variance-covariance matrices and the pooled variance-covariance matrix for the seventeen variables of Models E and F are given in Tables 5.1, 5.2 and 5.3 respectively. The variables in equation(5.21)will have the following values for the present case :

$$k = 2$$
$$N = 504$$
$$N_1 = 251$$
$$N_2 = 253$$
$$m = 17$$
$$\Lambda_1 = (0.274) \cdot 10^{-8}$$
$$\Delta_2 = (0.569) \cdot 10^{-10}$$
$$\Delta = (0.263) \cdot 10^{-8}$$

for which

$$= -2\left[1-\left(\frac{1}{251-1} + \frac{1}{253-1} - \frac{1}{504-2}\right) \cdot \frac{2 \times 17 \times 17 + 3 \times 17 - 1}{6 \times 18 \times 1}\right]$$

$$\cdot \ln\left[\frac{(0.274 \ 10^{-8})^{125} \cdot (0.569 \times 10^{-10})^{126}}{(0.263 \times 10^{-8})^{251}}\right]$$

Table 5.1 - <u>The variance-covariance matrix for 17 variables of Model E</u>

Variable	T_{amin}	T_1	T_2	T_3	f_5	f_6	f_7	f_8	f_M
T_{amin}	0.120	0.031	0.080	-0.129	-0.112	-0.187	-0.185	-0.051	-0.161
T_1	0.031	0.251	0.396	0.513	-0.449	-0.606	-0.426	0.197	-0.727
T_2	0.080	0.396	2.144	2.337	-1.016	-1.962	-0.889	2.326	-2.347
T_3	0.129	0.513	2.337	6.630	-1.265	-2.507	-1.473	4.637	-3.143
f_5	-0.112	-0.449	-1.016	-1.265	3.280	1.711	0.843	-2.709	1.934
f_6	-0.187	-0.606	-1.962	-2.507	1.711	3.588	1.611	-1.543	4.085
f_7	-0.185	-0.426	-0.889	-1.473	0.843	1.611	3.321	-0.030	1.302
f_8	-0.051	0.197	2.326	4.637	-2.709	-1.543	-0.030	79.830	1.264
f_M	-0.161	-0.727	-2.347	-3.143	1.934	4.085	1.302	1.264	20.915
f_2	-0.141	-0.511	-1.843	-2.535	1.070	2.676	1.378	-2.132	3.615
f_3	-0.190	-0.431	-1.071	-1.624	0.718	1.679	2.357	-0.154	2.375
f_4	-0.023	-0.239	-0.293	-0.696	0.011	0.776	1.677	0.209	0.101
A_{min}/A_0	0.004	0.019	0.043	0.056	-0.091	-0.065	-0.009	0.026	-0.084
A_1/A_0	0.005	0.013	0.045	0.067	-0.045	-0.070	-0.063	0.081	-0.093
A_2/A_0	0.011	0.032	0.070	0.099	-0.104	-0.119	-0.086	0.047	-0.132
A_3/A_0	0.009	0.022	0.047	0.060	-0.108	-0.081	-0.022	0.020	-0.113
f_1	-0.283	-0.698	-2.206	-3.331	2.529	3.700	3.123	-2.937	5.130

Table 5.1 (contd.)

Variable	f_2	f_3	f_4	A_{min}/A_0	A_1/A_0	A_2/A_0	A_3/A_0	f_1
T_{amin}	-0.141	-0.190	-0.023	0.004	0.005	0.011	0.009	-0.283
T_1	-0.511	-0.431	-0.239	0.019	0.013	0.032	0.022	-0.698
T_2	-1.843	-1.071	-0.293	0.043	0.045	0.070	0.047	-2.206
T_3	-2.535	-1.624	-0.696	0.056	0.067	0.099	0.060	-3.331
f_5	1.070	0.718	0.011	-0.091	-0.045	-0.104	-0.108	2.529
f_6	2.676	1.679	0.776	-0.065	-0.070	-0.119	-0.081	3.700
f_7	1.378	2.357	1.677	-0.009	-0.063	-0.086	-0.022	3.123
f_8	-2.132	-0.154	0.209	0.026	0.081	0.047	0.020	-2.937
f_M	3.615	2.375	0.101	-0.084	-0.093	-0.132	-0.113	5.130
f_2	3.115	1.519	0.773	-0.046	-0.063	-0.100	-0.058	3.159
f_3	1.519	2.888	1.449	-0.011	-0.061	-0.091	-0.027	3.028
f_4	0.773	1.449	3.529	0.027	-0.051	-0.046	0.030	2.381
A_{min}/A_0	-0.046	-0.011	0.027	0.005	0.001	0.004	0.005	-0.068
A_1/A_0	-0.063	-0.061	-0.051	0.001	0.003	0.003	0.001	-0.145
A_2/A_0	-0.100	-0.091	-0.046	0.004	0.003	0.008	0.005	-0.154
A_3/A_0	-0.058	-0.027	0.030	0.005	0.001	0.005	0.007	-0.095
f_1	3.159	3.028	2.381	-0.068	-0.145	-0.154	-0.095	7.811

Table 5.2 - The variance-covariance matrix for 17 variables of Model F

Variable	T_{amin}	T_1	T_2	T_3	f_5	f_6	f_7	f_8	f_M
T_{amin}	0.016	0.011	0.030	0.209	-0.012	-0.064	-0.075	0.064	-0.070
T_1	0.011	0.219	0.300	0.698	-0.504	-0.678	-0.290	0.188	-0.717
T_2	0.030	0.300	1.227	2.191	-0.837	-1.928	-0.834	0.976	-2.559
T_3	0.209	0.698	2.191	22.167	-2.197	-3.906	-2.011	4.714	-7.169
f_5	-0.012	-0.504	-0.837	-2.197	3.771	2.264	0.883	-5.973	1.712
f_6	-0.064	-0.678	-1.928	-3.906	2.264	5.115	1.905	-1.729	5.602
f_7	-0.075	-0.290	-0.834	-2.011	0.883	1.905	3.203	-3.287	1.045
f_8	0.064	0.188	0.976	4.714	-5.973	-1.729	-3.287	80.280	3.062
f_M	-0.070	-0.717	-2.559	-7.169	1.712	5.602	1.045	3.062	16.764
f_2	-0.052	-0.596	-1.669	-3.312	1.973	3.988	1.783	-3.102	4.512
f_3	-0.063	-0.276	-0.953	-2.396	0.450	1.712	2.110	-0.399	2.119
f_4	0.011	-0.113	-0.349	-0.514	0.649	1.029	1.698	-4.802	-0.315
A_{min}/A_0	0.001	0.022	0.035	0.085	-0.117	-0.090	-0.005	0.150	-0.099
A_1/A_0	0.001	0.009	0.040	0.100	-0.058	-0.093	-0.076	0.326	-0.072
A_2/A_0	0.002	0.033	0.067	0.147	-0.125	-0.155	-0.078	0.098	-0.166
A_3/A_0	0.002	0.023	0.039	0.105	-0.120	-0.095	-0.010	0.153	-0.110
f_1	-0.058	-0.562	-2.137	-5.300	3.479	4.938	3.884	-2.904	3.705

Table 5.2 (contd.)

Variable	f_2	f_3	f_4	A_{min}/A_0	A_1/A_0	A_2/A_0	A_3/A_0	f_1
T_{amin}	-0.052	-0.063	0.011	0.001	0.001	0.002	0.002	-0.058
T_1	-0.596	-0.276	-0.113	0.022	0.009	0.033	0.023	-0.562
T_2	-1.669	-0.953	-0.349	0.035	0.040	0.067	0.039	-2.137
T_3	-3.312	-2.396	-0.514	0.085	0.100	0.147	0.105	-5.300
f_5	1.973	0.450	0.649	-0.117	-0.058	-0.125	-0.120	3.479
f_6	3.988	1.712	1.029	-0.090	-0.093	-0.155	-0.095	4.938
f_7	1.783	2.110	1.698	-0.005	-0.076	-0.078	-0.010	3.884
f_8	-3.102	-0.399	-4.802	0.150	0.326	0.098	0.153	-2.904
f_M	4.512	2.119	-0.315	-0.099	-0.072	-0.166	-0.110	3.705
f_2	4.091	1.553	1.189	-0.074	-0.087	-0.136	-0.079	4.629
f_3	1.553	2.648	1.181	-0.001	-0.070	-0.065	-0.005	3.517
f_4	1.189	1.181	3.381	0.014	-0.085	-0.044	0.015	4.225
A_{min}/A_0	-0.074	-0.001	0.014	0.006	0.001	0.005	0.006	-0.091
A_1/A_0	-0.087	-0.070	-0.085	0.001	0.005	0.003	0.001	-0.268
A_2/A_0	-0.136	-0.065	-0.044	0.005	0.003	0.008	0.005	-0.154
A_3/A_0	-0.079	-0.005	0.015	0.006	0.001	0.005	0.006	-0.097
f_1	4.629	3.517	4.225	-0.091	-0.268	-0.154	-0.097	14.456

Table 5.3 - Pooled variance-covariance matrix for synthetic data

Variable	T_{amin}	T_1	T_2	T_3	f_5	f_6	f_7	f_8	f_M
T_{amin}	0.068	0.021	0.055	0.169	-0.062	-0.125	-0.130	0.007	-0.115
T_1	0.021	0.235	0.348	0.606	-0.477	-0.642	-0.358	0.193	-0.722
T_2	0.055	0.348	1.684	2.264	-0.926	-1.945	-0.861	1.649	-2.453
T_3	0.169	0.606	2.264	14.429	-1.733	-3.209	-1.743	4.676	-5.164
f_5	-0.062	-0.477	-0.926	-1.733	3.526	1.989	0.863	-4.347	1.823
f_6	-0.125	-0.642	-1.945	-3.209	1.989	4.355	1.759	-1.636	4.846
f_7	-0.130	-0.358	-0.861	-1.743	0.863	1.759	3.262	-1.665	1.173
f_8	0.007	0.193	1.649	4.676	-4.347	-1.636	-1.665	80.062	2.167
f_M	-0.115	-0.722	-2.453	-5.164	1.823	4.846	1.173	2.167	18.831
f_2	-0.096	-0.553	-1.756	-2.925	1.523	3.335	1.581	-2.619	4.066
f_3	-0.126	-0.354	-1.012	-2.012	0.583	1.696	2.233	-0.277	2.247
f_4	-0.006	-0.176	-0.321	-0.605	0.331	0.903	1.687	-2.307	-0.108
A_{min}/A_0	0.003	0.020	0.039	0.070	-0.104	-0.078	-0.007	0.088	-0.092
A_1/A_0	0.003	0.011	0.042	0.083	-0.051	-0.082	-0.070	0.203	-0.082
A_2/A_0	0.006	0.032	0.068	0.123	-0.115	-0.137	-0.082	0.072	-0.149
A_3/A_0	0.005	0.022	0.043	0.082	-0.114	-0.088	-0.016	0.087	-0.112
f_1	-0.177	-0.630	-2.186	-4.327	3.000	4.318	3.501	2.920	4.411

Table 5.3 (contd.)

Variable	f_2	f_3	f_4	A_{min}/A_0	A_1/A_0	A_2/A_0	A_3/A_0	f_1
T_{min}	-0.096	-0.126	-0.006	0.003	0.003	0.006	0.005	-0.177
T_1	-0.553	-0.354	-0.176	0.020	0.011	0.032	0.022	-0.630
T_2	-1.756	-1.012	-0.321	0.039	0.042	0.068	0.043	-2.186
T_3	-2.925	-2.012	-0.605	0.070	0.083	0.123	0.082	-4.327
f_5	1.523	0.583	0.331	-0.104	-0.051	-0.115	-0.114	3.000
f_6	3.335	1.696	0.903	-0.078	-0.082	-0.137	-0.088	4.318
f_7	1.581	2.233	1.687	-0.007	-0.070	-0.082	-0.016	3.501
f_8	-2.619	-0.277	-2.307	0.088	0.203	0.072	0.087	2.920
f_M	4.066	2.247	-0.108	-0.092	-0.082	-0.149	-0.112	4.411
f_2	3.605	1.536	0.982	-0.060	-0.075	-0.118	-0.068	3.892
f_3	1.536	2.768	1.315	-0.006	-0.066	-0.078	-0.015	3.269
f_4	0.982	1.315	3.453	0.021	-0.069	-0.045	0.023	3.304
A_{min}/A_0	-0.060	-0.006	0.021	0.006	0.001	0.005	0.006	-0.079
A_1/A_0	-0.075	-0.066	-0.069	0.001	0.004	0.003	0.001	-0.207
A_2/A_0	-0.118	-0.078	-0.045	0.005	0.003	0.008	0.005	-0.154
A_3/A_0	-0.068	-0.015	0.023	0.006	0.001	0.005	0.006	-0.095
f_1	3.892	3.269	3.304	-0.079	-0.207	-0.154	-0.095	11.145

Therefore, = 125.37

The degrees of freedom (k-1) m (m+1)/2 is 153. The tabulated value of χ^2 are given only till 120 degrees of freedom in Dixon and Massey (1969), but for large values of degrees of freedom the approximate formula given is

$$\chi^2_\alpha = \nu (1 - \frac{2}{9\nu} + z_\alpha \sqrt{\frac{2}{9\nu}})^3$$

where z_α is the normal deviate and ν is the number of degrees of freedom. For the 99th percentile this gives a tabulated value of χ^2_α for 153 degrees of freedom as 196.616, which is much larger than the calculated value, and therefore, the null hypothesis of the equality of the two variance-covariance matrices is accepted.

That each of the variables is normally distributed is tested by plotting the frequency distribution on a probability paper. Most of the variables show normal or near normal distribution and the discriminant function is not seriously affected by limited departures from normality (Davis, 1973). The distribution of some of the variables in real case is shown in Figures 6.30 - 6.33.

The discriminant score, R, is calculated for each of the seismogram and is projected on the discriminant function line (Figure 5.2). To avoid overlapping of points while plotting the data, seismograms with the same value of discriminant function are plotted at different heights. R_E and R_F are the
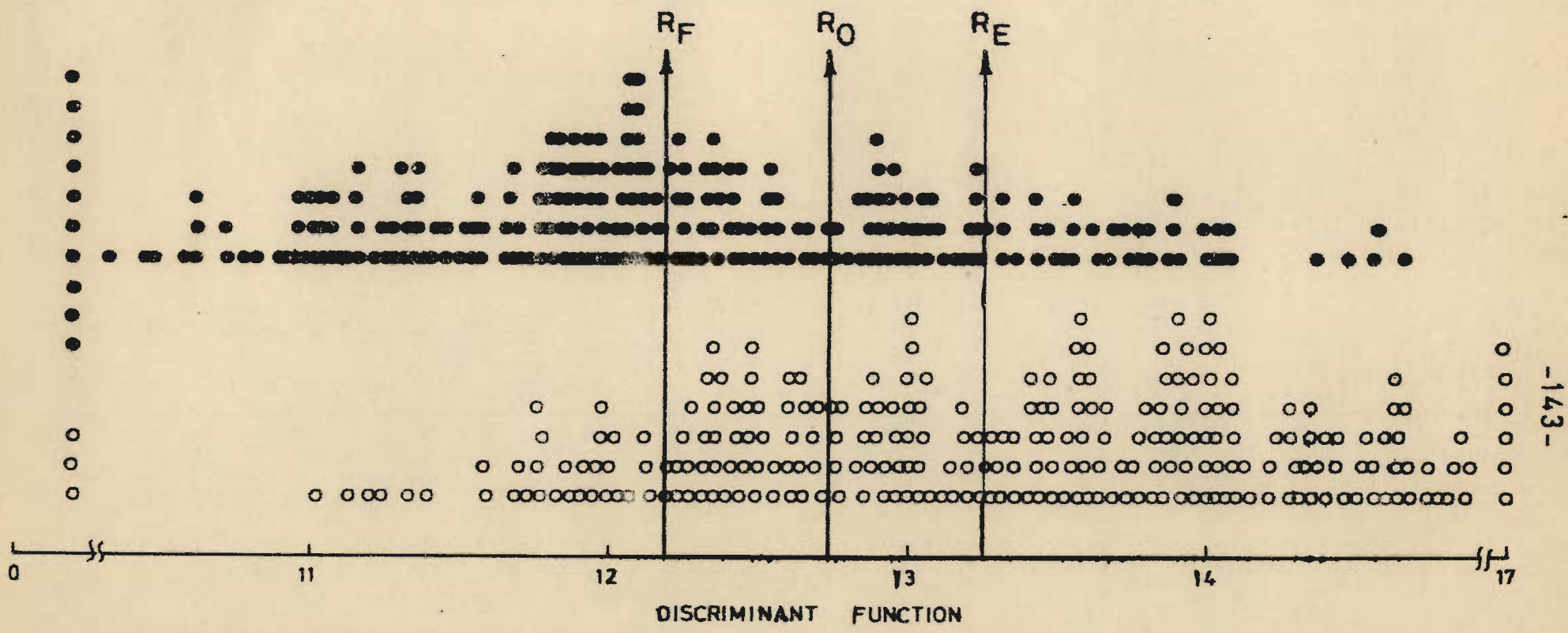
FIG. 5.2 _ DISCRIMINANT ANALYSIS TO DISTINGUISH MODEL E FROM F. 17 VARIABLES ARE TAKEN INTO CONSIDERATION.

multivariate means of seventeen variables of Models E and F, respectively and R_O is the discriminant index. Difference between R_E and R_F is Mahalanobis'distance, D^2. Discriminant scores (equation 5.14) when plotted for seismograms of the two models show some overlap (Figure 5.2). Despite this overlap 70 percent of the total 508 seismograms are correctly classified.

F-test given in equation(5.19)has been applied to test the equality of multivariate means of the Models E and F. The test shows that the two means are significantly different at 95 percent confidence level. The calculated values of R_E and R_F and the percentage contribution of each of the 17 variables are given in Table 5.4.

While most of the variables make positive contributions a few display a negative role. Positive contributions indicate that the variables are meaningful discriminators and the amount of contribution is a measure of the potency of the variable.

Variable A_1/A_O contributes 38.4 percent, 30.5 percent is contributed by f_8, the frequency at which logarithm of power decreases to zero; 24.5 percent is contributed by T_2 - the time of second zero crossing in the ACF and 18.7 percent is contributed by f_M. These variables because of their high percentage of contribution can be considered as powerful

Table 5.4 - <u>Discrimination of Model E from Model F when all</u>
<u>17 variables are considered</u>

Calculated value of F = 7.5483 with 17 and 490 degrees of
freedom.

Tabulated value of F = 1.76 with 17 and ∞ degrees of freedom
at α = 0.05

R_1 = 13.2649

R_0 = 12.7432

R_2 = 12.2215

Sl.No.	Variable	Value of constant	Percentage contributed towards discrimination
1.	T_{amin}	0.2474	1.1
2.	T_1	0.4089	6.5
3.	T_2	0.3334	24.5
4.	T_3	-0.0841	-3.9
5.	A_{min}/A_0	-4.7422	-4.0
6.	A_1/A_0	8.4131	38.4
7.	A_2/A_0	0.7229	2.1
8.	A_3/A_0	1.4399	2.0
9.	f_1	-0.0307	6.9
10.	f_2	0.0422	-3.9
11.	f_3	-0.0433	3.4
12.	f_4	0.0736	-3.8
13.	f_5	0.0254	-1.3
14.	f_6	0.1392	-12.8
15.	f_7	0.0687	-4.4
16.	f_8	0.0490	30.5
17.	f_M	-0.0921	18.7

discriminators of lithologies. Other discriminators are f_1 - the average frequency which contributes 6.9 percent, the time T_1 - which is the time of the first zero crossing in the ACF and contributes 6.5 percent; f_3 - the frequency of 50th percentile value of frequency weighted power contributes 3.4 percent; A_2/A_0 and A_3/A_0 make 2.1 percent and 2.0 percent contributions, respectively and T_{amin} - the time of first minima in the ACF contributes 1.1 percent. These ten variables can be termed as seismic discriminators.

To check the efficacy of this analysis 10 seismograms from each of the Models E and F, which were not part of the aforementioned discriminant analysis, were put to test. The discriminant scores for these 20 seismograms were used to assign a model - either E or F to them. Fifteen of these 20 seismograms could be classified to their correct model, one had the same value as R_0 and four were misclassified. This indicates a 75 percent success in assigning synthetic reflection seismograms to their proper models. This approach, therefore, appears to be successful and as such a dominantly sandy lithology can be distinguished from a dominantly shaly lithology.

CHAPTER - VI

APPLICATION OF DISCRIMINANT ANALYSIS TO
FIELD SEISMIC DATA

The successful discriminant analysis carried out for synthetic data indicates its potentiality for analysing lithostratigraphy using seismograms. In synthetic seismograms the nature of the source pulse and the simulated lithological sequence are the two factors which play a role in shaping the seismogram, the same source pulse was used in generating all the synthetic seismograms. Consequently variables derived from synthetic seismograms are related to the subsurface lithostratigraphy. For field seismograms, however, the source pulse may not uniformly be the same for a suite of seismograms. The observed seismograms may be further modified by the field recording and data processing techniques. The observed seismograms are the total response of the source pulse, the subsurface lithostratigraphy and the recording and processing system. In addition to the above factors degradation due to the presence of source generated noise (e.g., ground roll) as well as ambient noise also occurs. Thus the problem of inferring lithostratigraphy in real cases is relatively more difficult than in the synthetic models. Inspite of the more complicated situations met in nature discriminant analysis has been carried out on real data to find out how successful this concept would be. With this as an aim field seismograms from two different

Areas X and Y characterizing a dominantly sandy and a shaly subsurface lithology, respectively, from a sedimentary Basin Z in India have been subjected to discriminant analysis.

6.1 SEISMIC SECTIONS FOR AREAS X AND Y

Part of the seismic sections of two Areas X and Y, belonging to the Formation K of the sedimentary Basin Z have been subjected to the analysis discussed in Chapters III, IV and V. Four seismic profiles of Area X and three of Area Y were considered for the purpose of this study. It amounts to a total of 70 km of seismic line or 239 traces for Area X and 55 km or 148 traces for Area Y. The details of recording and processing procedures are given in Tables 6.1 - 6.7.

The Formation K was marked on the seismic sections by tying the seismic data with that of the nearby wells by using 5 velocity functions for Area X and 3 for Area Y. The Formation K in Area X could be identified with an accuracy of one reflection cycle, whereas for Area Y the wells were situated at a considerable distance from the seismic lines and therefore the Formation K could be marked with lesser accuracy.

A band of reflections arising from within the Formation K and consisting of a number of cycles is observed on seismic sections. This indicates that the formation consists of a series of thin lithological beds. The strong trough and peak phase alignment of the reflections at places either merge

Table 6.1 - <u>The field recording parameters for Area X</u>

Configuration	Split Spread	CDP
Type of recording	Digital	
SP/VP interval	100 m	
Geophone interval	100 m	
Near Offset	100 m	
Far Offset	1200 m	
Number Traces	24	
Geophones/Trace	12	
Recording filter	10 (2) - 125	
Sampling rate	2 ms	

Table 6.2 - <u>The order in which the seismic sections for</u>
<u>Area X were processed</u>

1.	Record length	4 seconds	
	Sampling interval	4 ms	
2.	Additional Process	True Amplitude Recovery	
3.	Decon before stack	Operator length	= 160 ms
		Window length	= 2000 ms
		Prediction Time	= 2nd zero crossing
4.	Statics	Party supplied	
5.	NMO	SSN No. at position shown V	
7.	Stacking	1200 percent	
8.	Filter		

L.C.	H.C.	Application time in Seconds	Overlap Time
20 - 25	40 - 45	1.0	200 ms
5 - 10	50 - 55	1.5	200 ms
3 - 5	35 - 40	4.0	

9.	Equalization	Two window
10.	Trace Mixing	No.of traces = 3

<u>Note</u> : 6 is residual statics, not applied.

Table 6.3 - <u>Field Parameters used for 3 seismic lines of Area X</u>

1.	Shot depth	29 m	
2.	Shot pattern	Single hole	
3.	Charge size	8.34 kg.	
4.	Gain	Preamplifier	= 36 db
		Initial gain	= 12 db
		Expansion range	= 6 db
		Release rate	= 32 ms
		Final gain	= 84 db
5.	Filters	L.C.	= 10(2) Hz
		H.C.	= 125 Hz
		Notch	= IN

6. Instrument Parameters

   (i) Sampling interval 2 ms

   (ii) Trip delay

     (a) 12th and 13th trace 150 ms

     (b) and 24th trace   700 ms

     (c) In between the rate of increment for other channels is 50 ms

7. Geophone Pattern

   (i) 12 Geophones all in series (1,2,3,3,2,1)

   (ii) Group spacing - 6.5 m

   (iii) Digiphone - 10 Hz were used.

8. Field numbers - 12 fold split spread C.D.P. with 100 m group interval and 100 m in-line off-set.

Table 6.4 - <u>Field Parameters used for the Fourth Seismic</u>
<u>Line of Area X</u>

1.   Shot depth :

From SP 32 to SP 118 average depth 23 m

        SP 120 to SP 278 average depth 29 m

        SP 280 to SP 556 average depth 26 m

2.   Shot Pattern :   Single hole

3.   Shot size :   11.12 kg

4.   Gain :

Pre-amplifier gain - 36 db

Initial gain   - 24 db

Expansion range - 6 db

Release rate   - 32 m sec.

Final gain   - 84 db

5.   Filters :

L.C. 10(2) Hz

H.C. 125 Hz

Notch - IN

6.   Instrument Parameters :

(i) Sampling interval - 2 ms

(ii) Trip delay

    (a) 12th and 13th trace - 200 ms

    (b) 1st and 24th trace - 750 ms

    (c) In between the rate of increment
        for other channels is 50 ms

7.   Geophone pattern :

      (i)   10 geophones all in series (1,2,2,2,2,1)
             upto SP 134

     (ii)   12 geophones all in series (1,2,3,3,2,1)
             upto SP 556

 (iii)   Group spacing - 6.5 m

  (iv)   Digiphones 10 Hz were used.

8.   Fold Numbers :

12 fold split spread CDP with 85 m group
interval and 340 m in-line off-set.

Table 6.5 - <u>The Field recording parameters for Area Y</u>

Digital recording	recorded by SIG - 159
Year recorded	1978-79
SP/VP interval	100 m
Instrument type	SN 328
Geophone Interval	100 m
Near Offset	500 m
Recording filter	10 (2) 125 Hz
Far Offset	2800 m
Sample rate	2 ms
Number Traces	24
Record Length	5.0 seconds
Configuration	End on 12 fold CDP
Geophones/Trace	12

Table 6.6 <u>The order in which the seismic sections for</u>
           <u>Area Y were processed</u>

1. Record length   5.0 seconds

2. Sampling Interval   4 ms

3. Statics   Party supplied

4. NMO   SSN No. at places shown 'V'

5. Residual statics window   1.47 to 1.77 sec

6. Stacking   1200 percent

7. Filter

L.C.	H.C.	Application Time	Overlap Time
50 Hz	Notch	5.0 sec	
10 Hz	45 Hz	2.5 sec	200 ms
5 Hz	35 Hz	5.0 sec	

8. Trace Mixing   No. of traces   = 3

Table 6.7   <u>Field parameters used for Area Y</u>

1.   Shot depth   :   21-24 m

2.   Shot pattern   :   Single hole

3.   Charge size   :   13.9 kg

4.   Gain :

        Preamplifier gain - not supplied

        Initial gain   - 30 db

        Expansion range   - 6 db

        Release rate   - 64 ms

        Final gain   - not supplied

5.   Filters :

        L.C.   10(2) Hz

        H.C.   125 Hz

        Notch -   IN

6.   Instrument parameters :

    (i)   Sampling interval   2 ms

    (ii)   Trip delay

        (a) Channel 1 and 2   -   1400 ms

        (b) Channel 3 and 4   -   1300 ms

        (c) Channel 5 and 6   -   1200 ms

        (d) Channel 7 and 8   -   1100 ms

        (e) Channel 9 and 10   -   1000 ms

        (f) Channel 11 and 12 -   900 ms

        (g) Channel 13 and 14 -   800 ms

(h)	Channel 15 and 16	-	700 ms
(i)	Channel 17 and 18	-	600 ms
(j)	Channel 19 and 20	-	500 ms
(k)	Channel 21 and 22	-	400 ms
(l)	Channel 23	-	300 ms
(m)	Channel 24	-	200 ms

7. Geophone pattern :

    (i)   12 geophones all in series (2,2,2,2,2,2)

    (ii)   Group spacing - 9 m

    (iii) Base length   - 45 m

8. Fold numbers : 12 fold

with each other or diverge and form separate phase alignments. These characteristics of the reflection band may be associated with lateral facies changes, pinch outs, wedging out or thinning and thickening of the lithological beds of small thickness. The exact nature of such features can only be checked by closely spaced well data.

The number of cycles present in a band are generally related to the number of beds in the formation and their thickness. It is noticed that the reflection from within the Formation K, have more cycles in the zones of depression in Area Y. It may be due to an increase in the thickness of these beds in the structurally low zones. The quality of data ranges from fair to good. The two way travel time within the Formation K ranges from 200 ms to 400 ms. This part of the data is retrived from magnetic tapes, and a few of these traces are shown for the relevant time window in Figures 6.1 - 6.4. A comparison of the seismograms from the two areas shows that though it is not easy to pickout any specific differences between them, yet in general, seismograms from Area Y show waveforms broader than those of Area X. Furthermore, the relevant time window chosen for Area Y is at a much longer two way travel time because Formation K is at a greater depth in Area Y.

The autocorrelation functions of the seismic traces are calculated by the method given in Section 3.2 and some of
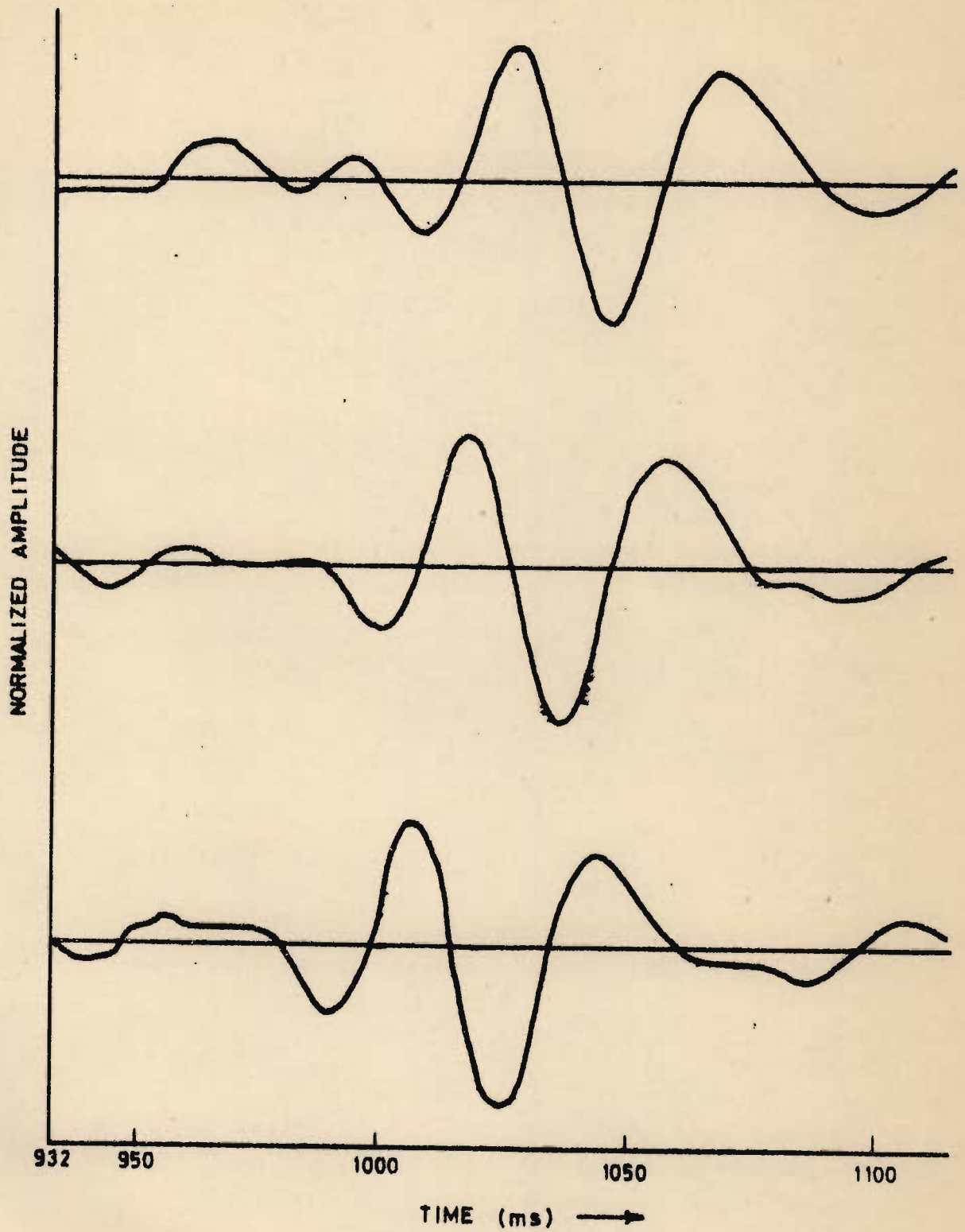
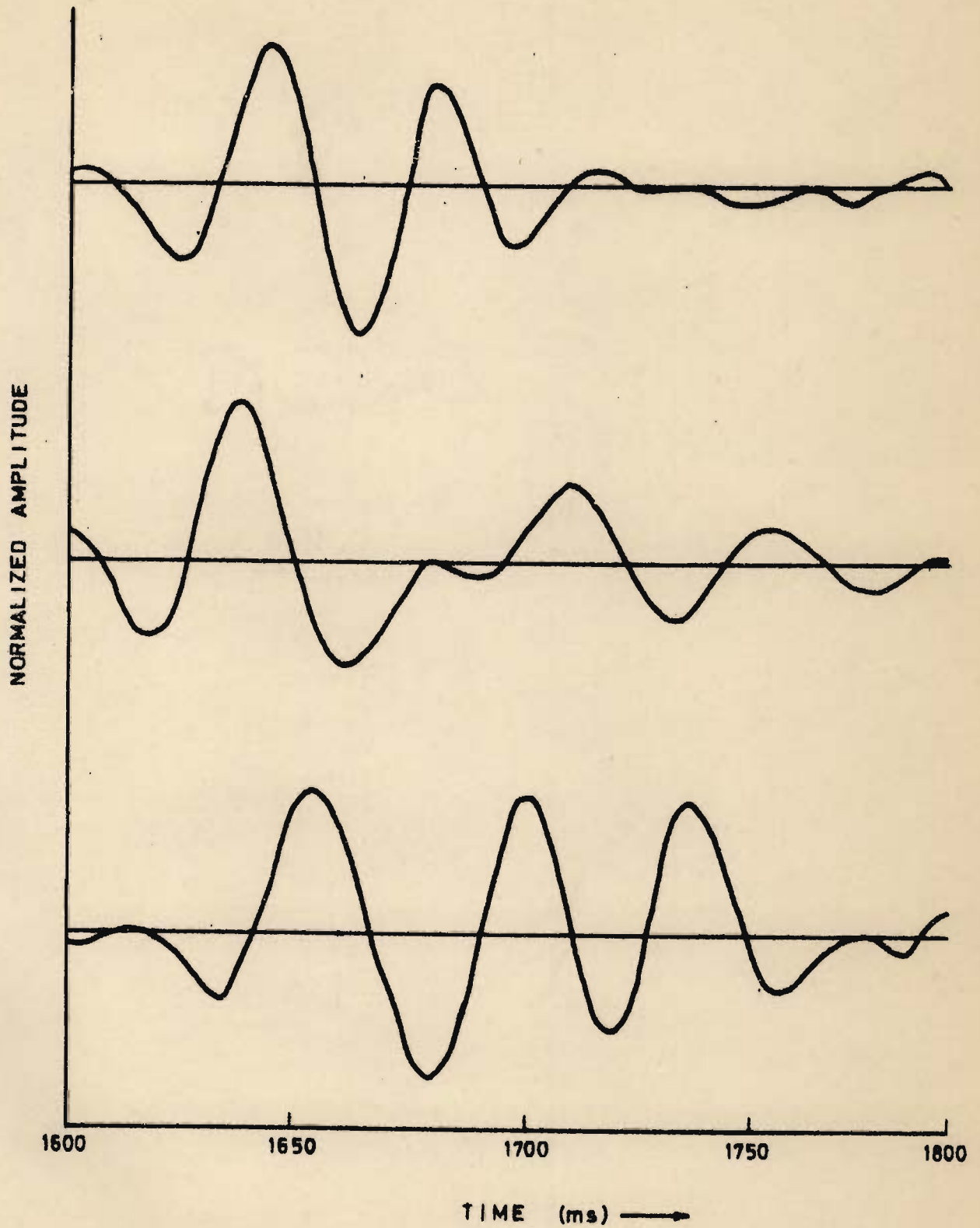FIG.6.1_EXAMPLES OF RELEVANT WINDOW OF SEISMOGRAMS
FOR AREA X.

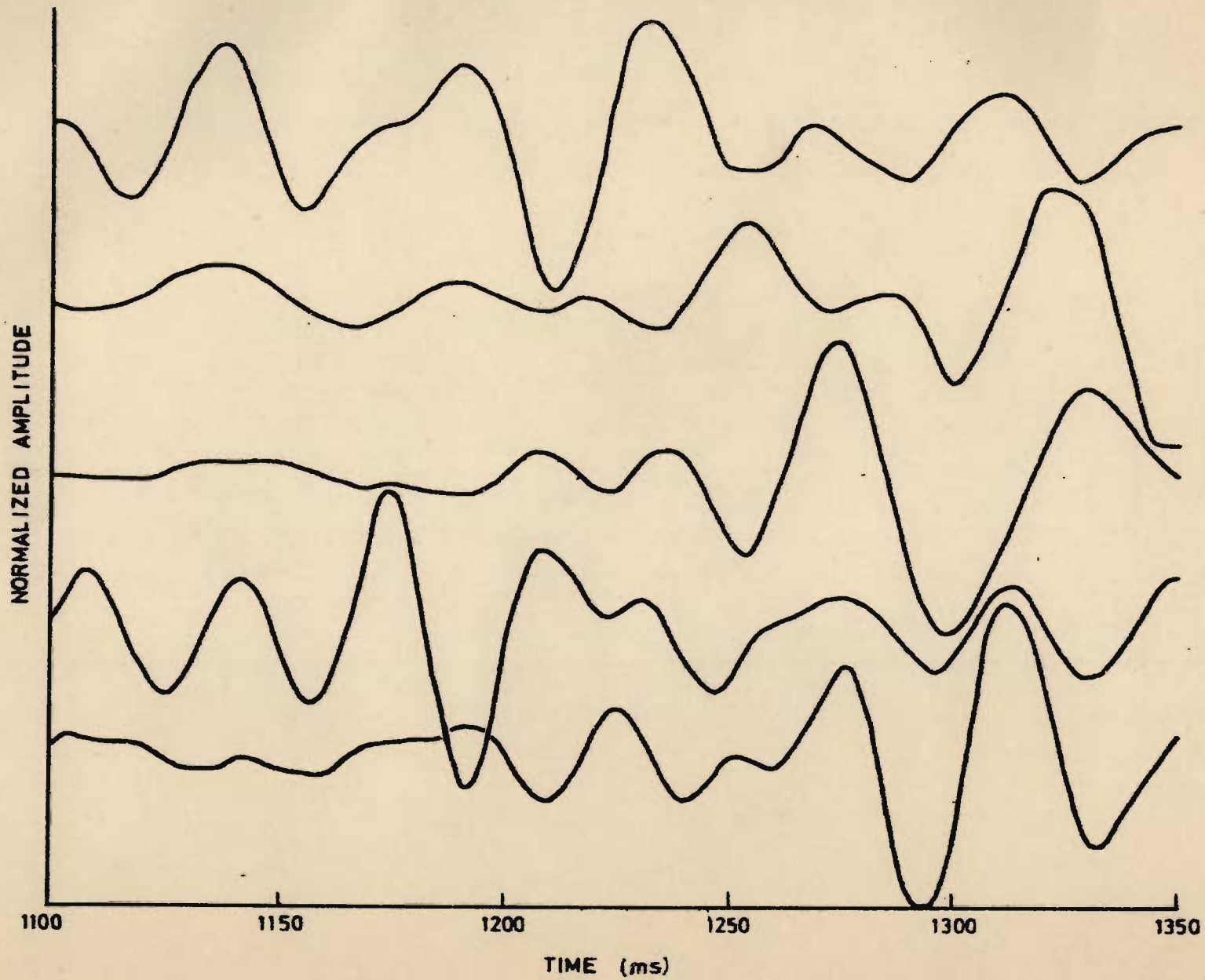FIG.6.2 _ EXAMPLES OF RELEVANT WINDOW OF SEISMOGRAMS FOR AREA Y.

FIG. 6.3 _ SOME EXAMPLES FROM THE 239 TRACES OF SEISMOGRAMS ANALYSED
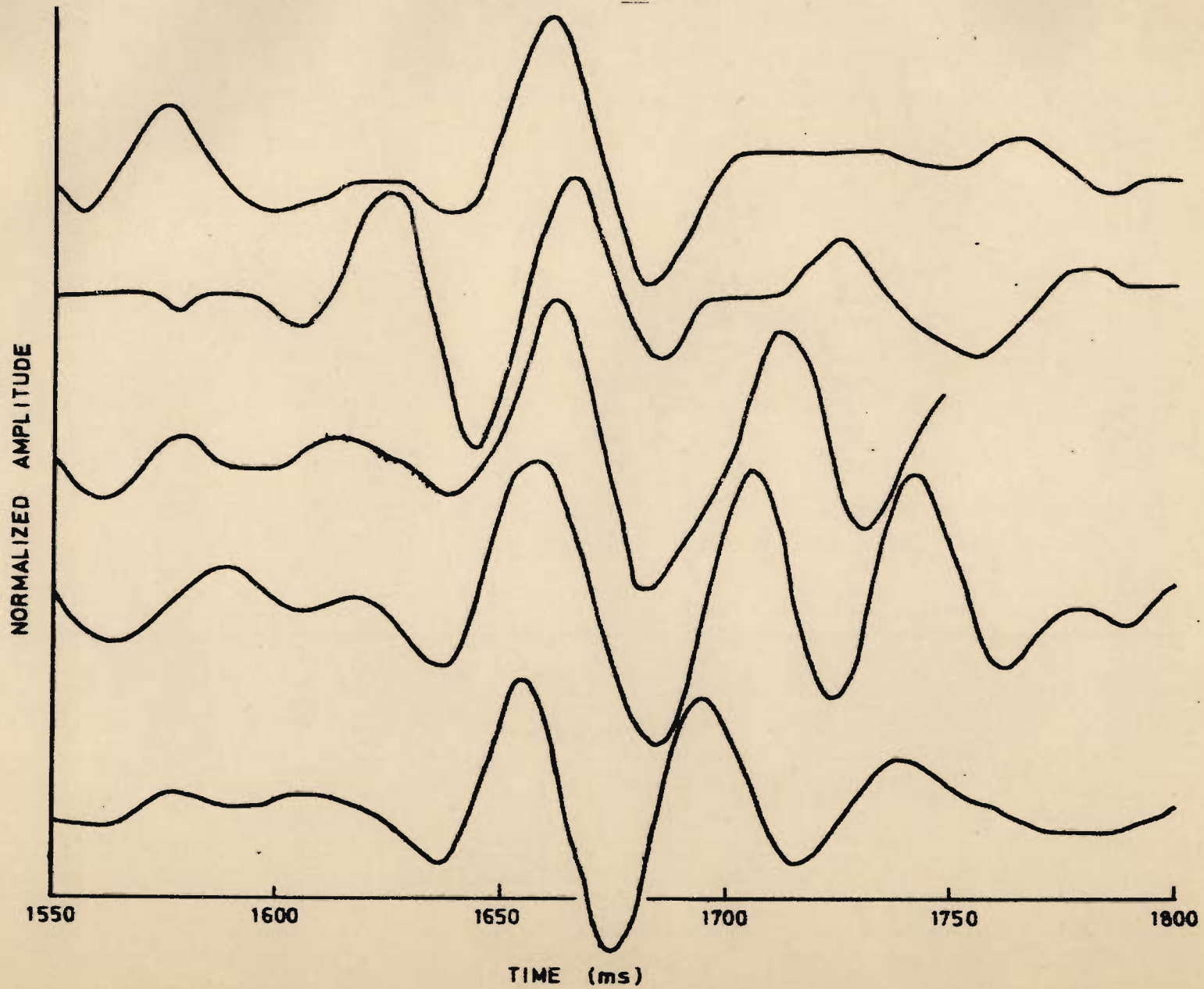
FOR AREA X .

FIG. 6.4 _ SOME EXAMPLES FROM THE 148 TRACES OF SEISMOGRAMS ANALYSED FOR AREA Y.

these are shown in Figures 6.5 - 6.8. The same eight variables that were picked from the ACFs of the synthetic seismograms are also picked from the ACFs of the seismograms of the field data and are shown in Figures 6.5 and 6.6. Figures 6.7 and 6.8 show several ACFs for a comparative study. The autocorrelation functions are broader than what was observed for the synthetic case (Figures 4.6 and 4.7). The ACF traces for Area X are more oscillatory than for Area Y which displays a flatter character. This phenomenon was also observed in the synthetic case where the ACFs of Model E (characterizing Area X) show similar oscillatory nature and the ACFs of Model F (characterizing Area Y) are relatively smooth at larger time lags.

The maximum entropy power spectra for the seismic traces were computed for the present study. Three of these spectra for the Area X and Y with the frequency f_M marked on them are shown in Figures 6.9 and 6.10. Figures 6.11-6.14 and 6.15 - 6.17 show some more examples from the 239 and 148 traces of power spectra analysed for Areas X and Y respectively. The frequency bandwidth of these spectra is subject to the field, recording and processing parameters, and are band limited between 5 and 55 Hz as indicated by the filters in Tables 6.2 and 6.6. The spectra of field seismograms therefore have a frequency band narrower than that for synthetic seismograms (Figures 4.9 - 4.16) but they retain the character of showing one or more peaks.
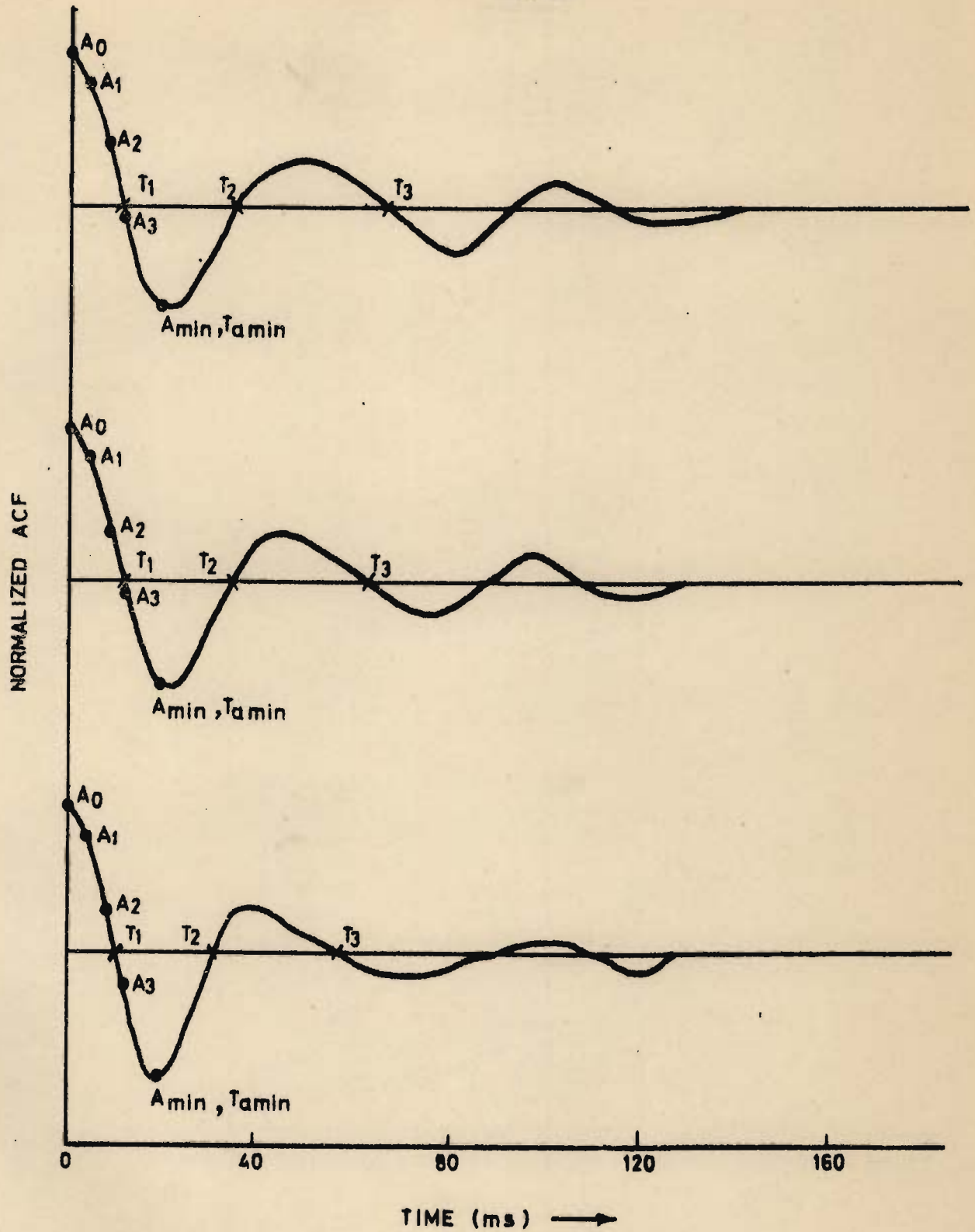
FIG. 6.5 - SOME AUTOCORRELATION FUNCTIONS OF SEISMOGRAMS
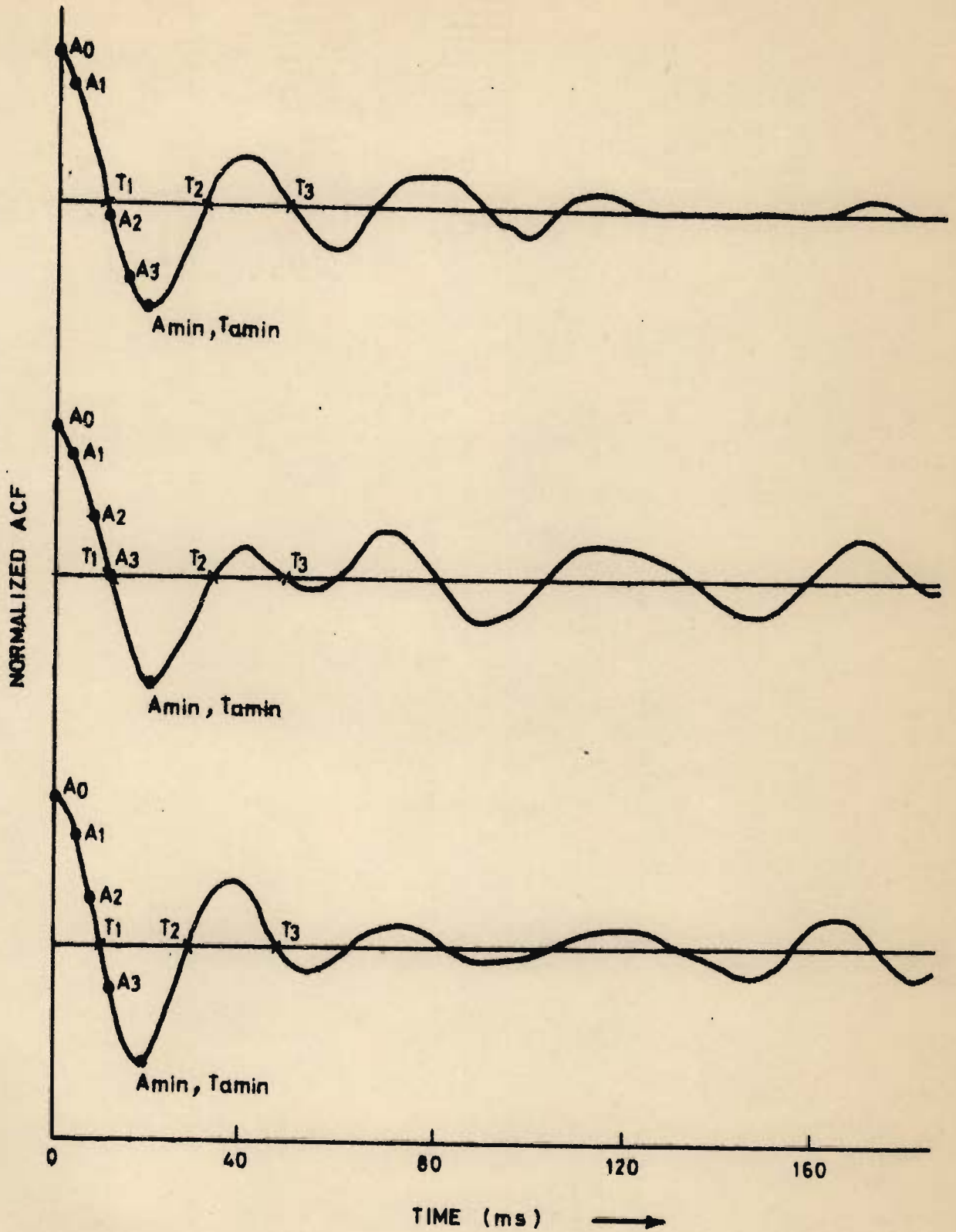
FOR AREA X

(ACF = AUTOCORRELATION FUNCTION).

FIG. 6.6 _ SOME AUTOCORRELATION FUNCTIONS OF SEISMOGRAMS
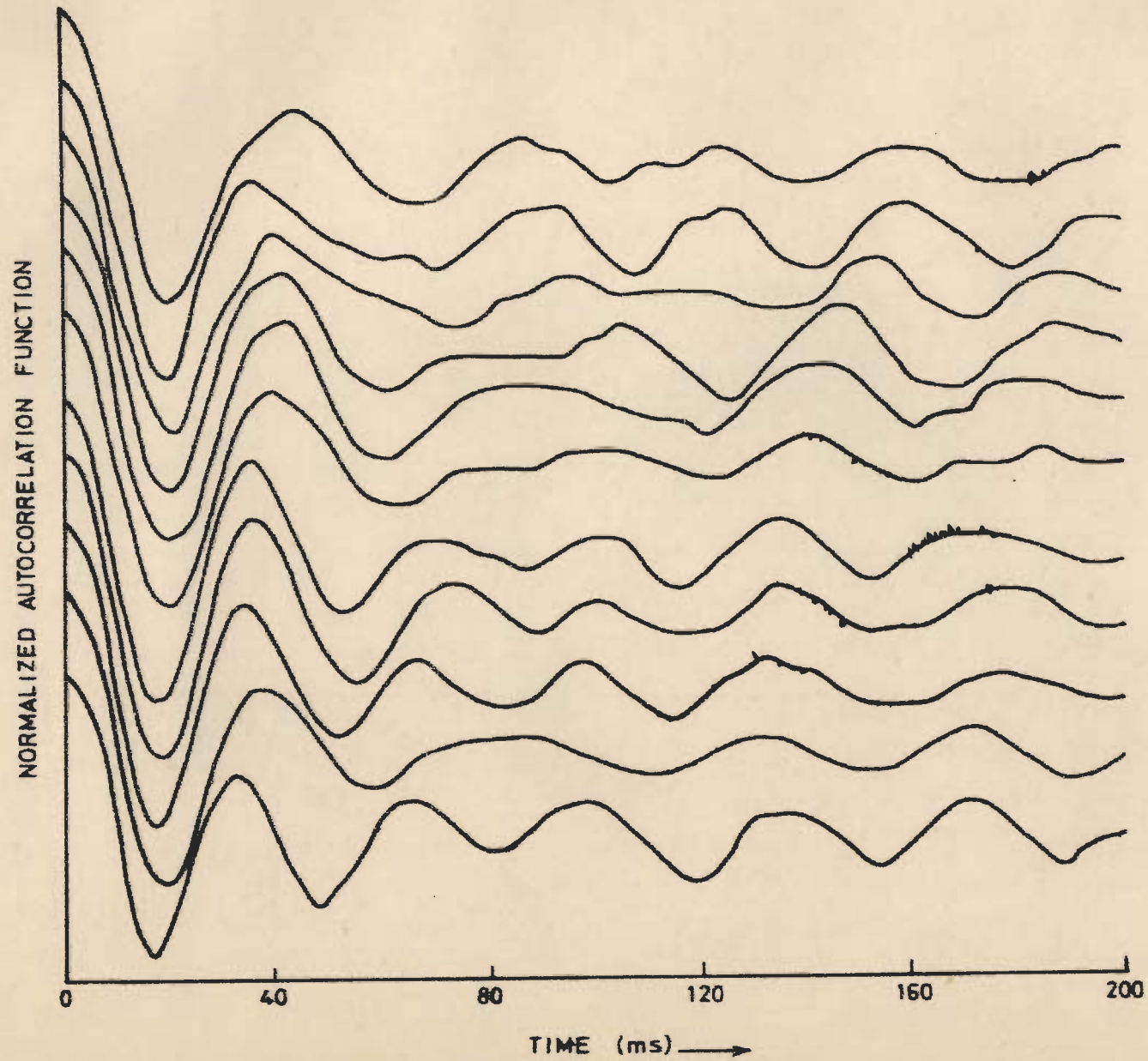FOR AREA Y
(ACF = AUTOCORRELATION FUNCTION).

FIG.6.7_SOME EXAMPLES FROM THE 239 TRACES OF AUTOCORRELATION
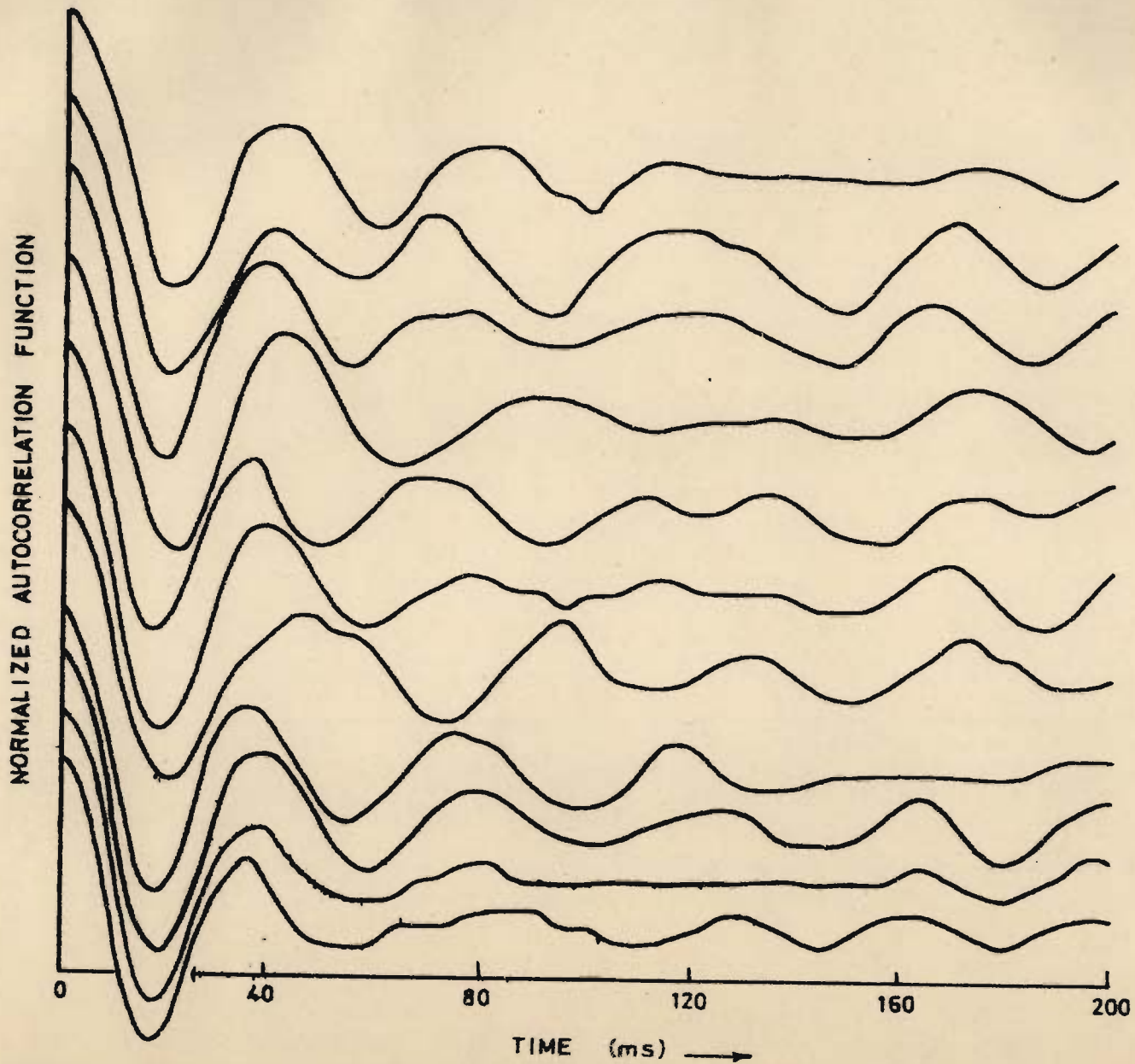
FUNCTIONS ANALYSED FOR AREA X.

FIG.6.8 _ SOME EXAMPLES FROM THE 148 TRACES OF AUTOCORRELATION
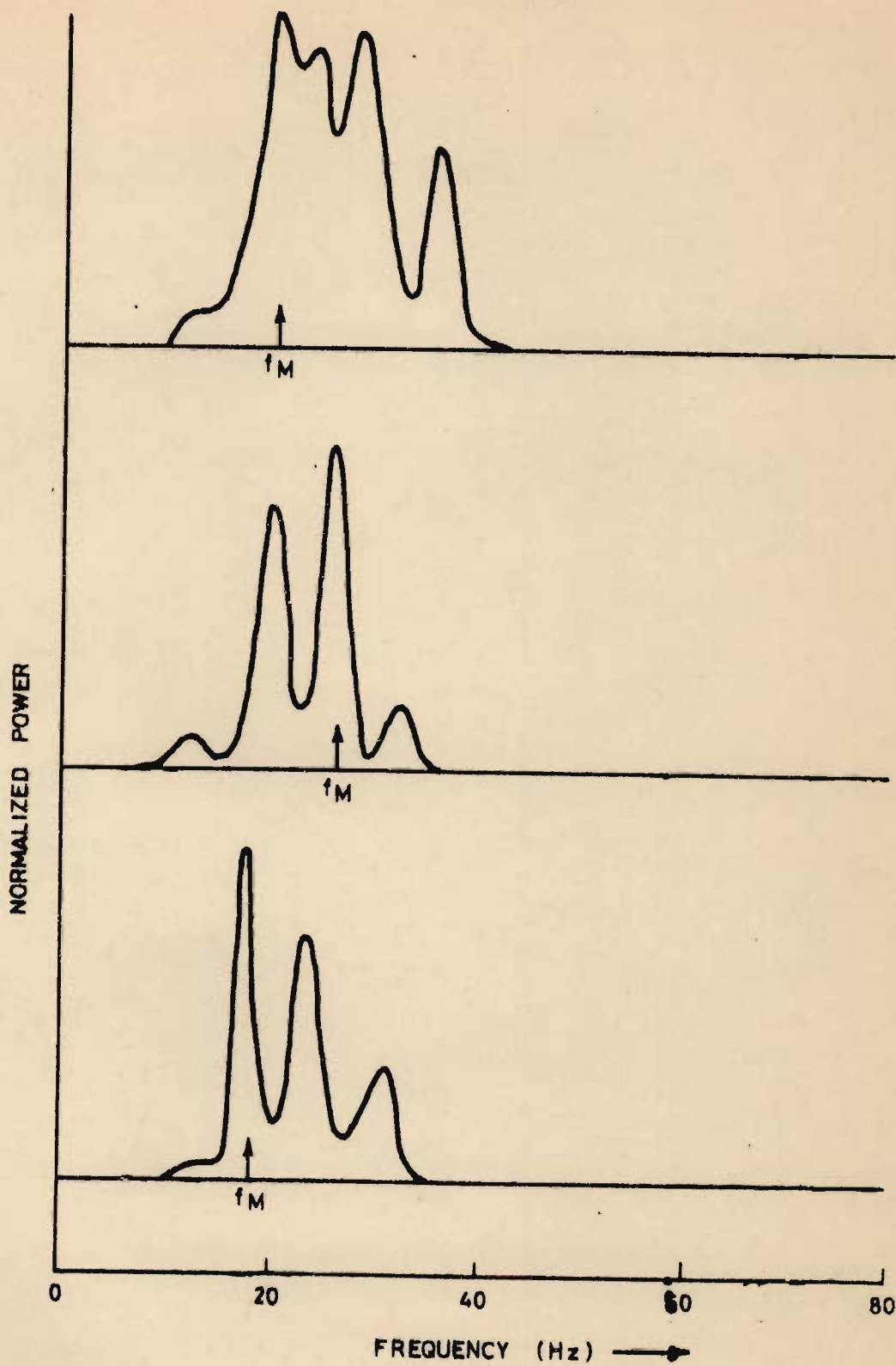
FUNCTIONS ANALYSED FOR AREA Y.

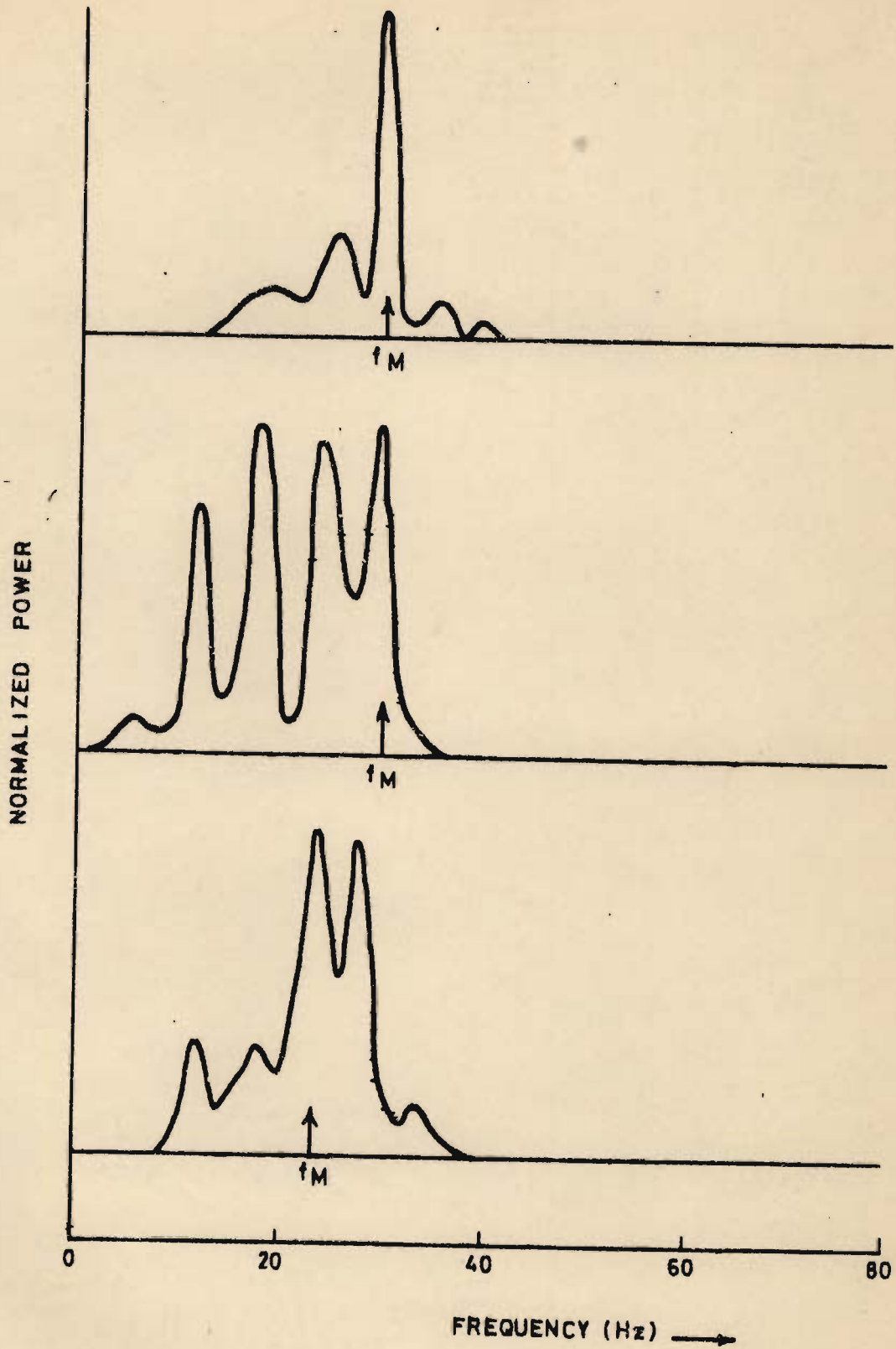FIG. 6.9 - POWER SPECTRA OF SOME SEISMOGRAMS FOR AREA X.

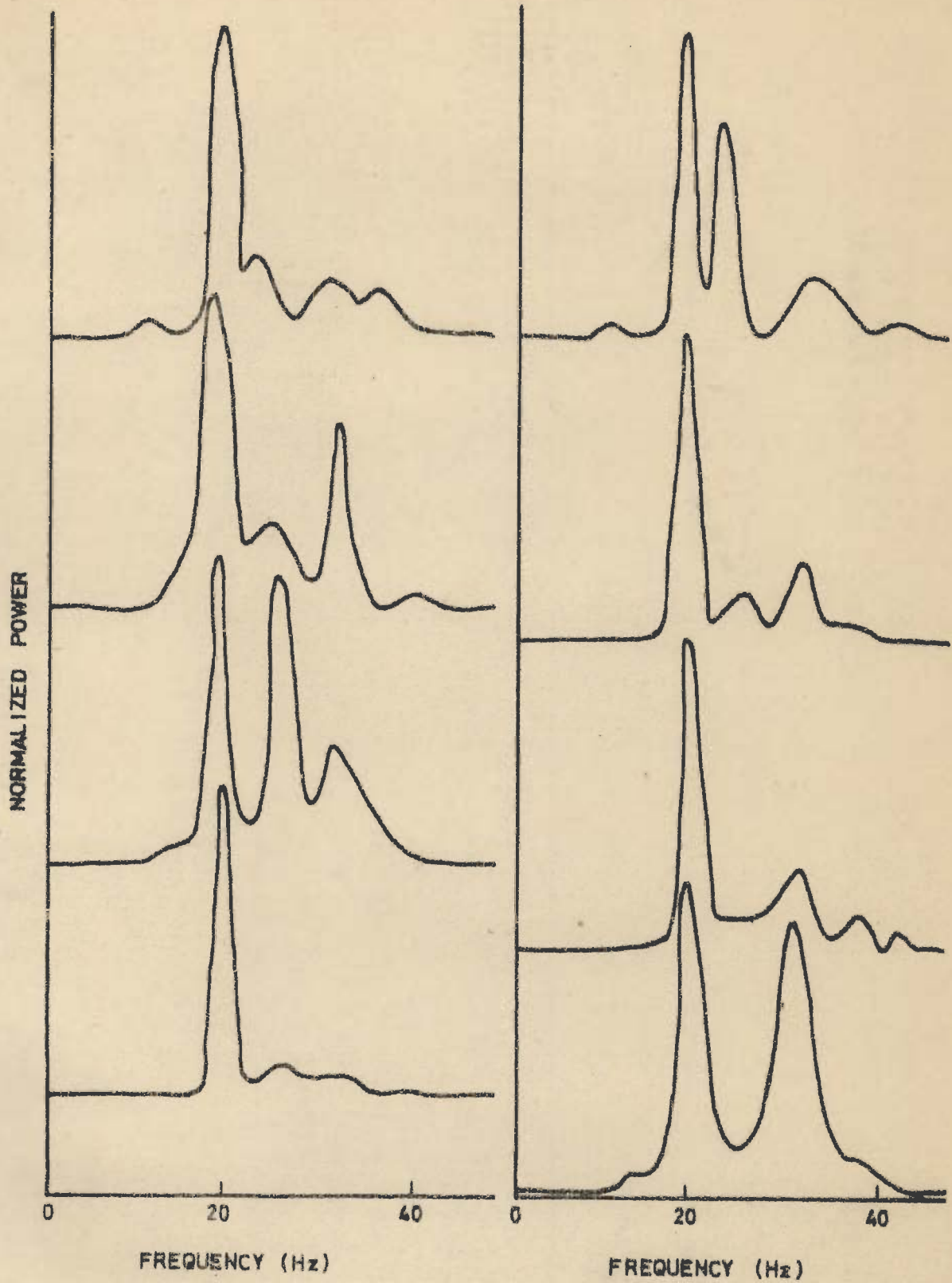FIG.6.10_POWER SPECTRA OF SOME SEISMOGRAMS

FOR AREA Y.

FIG.6.11_ SOME EXAMPLES FROM THE 239 TRACES OF POWER
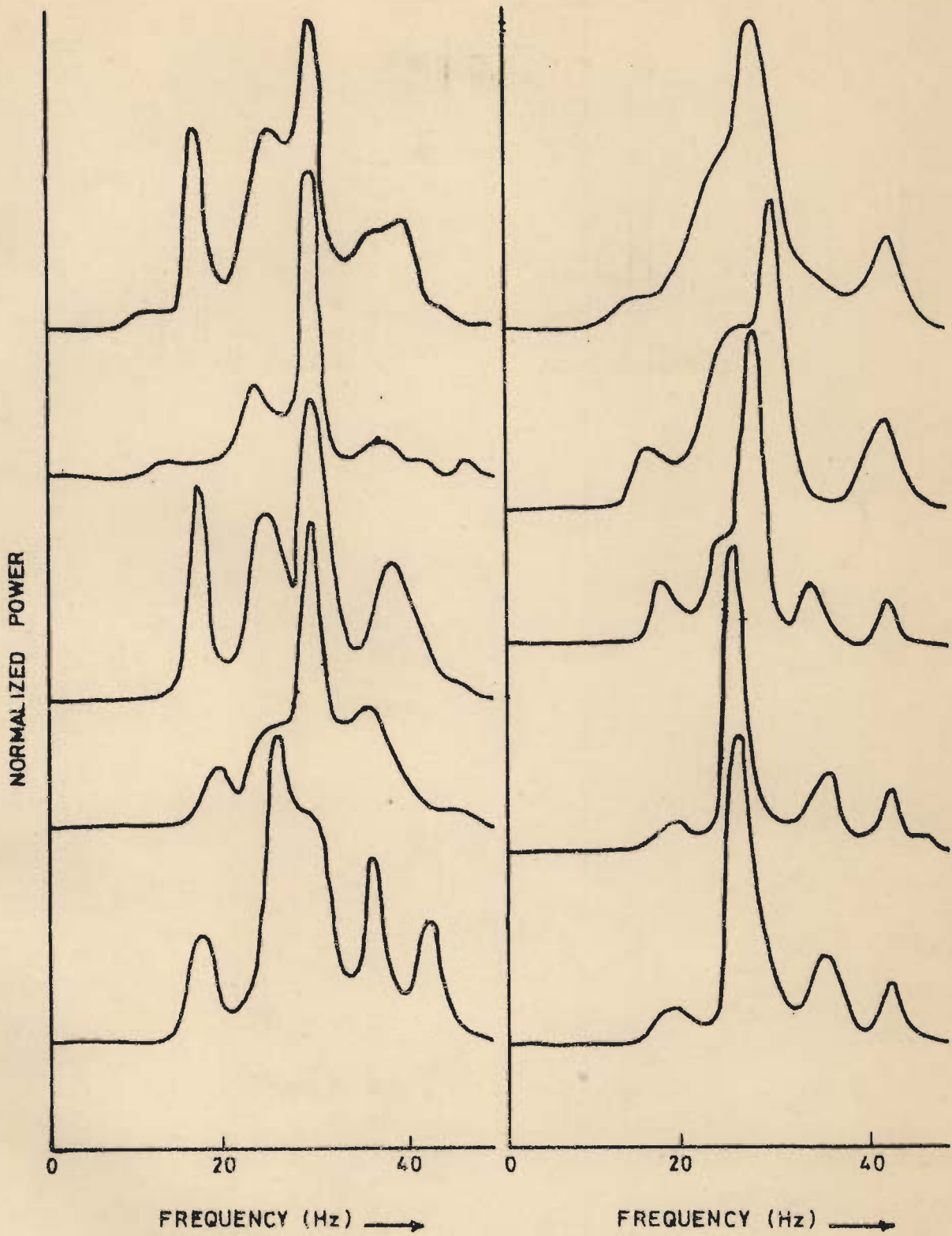
SPECTRA ANALYSED FOR AREA X.

FIG.6.12_SOME EXAMPLES FROM THE 239 TRACES OF POWER

SPECTRA ANALYSED FOR AREA X.

FIG. 6.13 SOME EXAMPLES FROM THE 239 TRACES OF POWER
SPECTRA ANALYSED FOR AREA X.

FIG.6.14_SOME EXAMPLES FROM THE 239 TRACES OF POWER SPECTRA ANALYSED FOR AREA X .

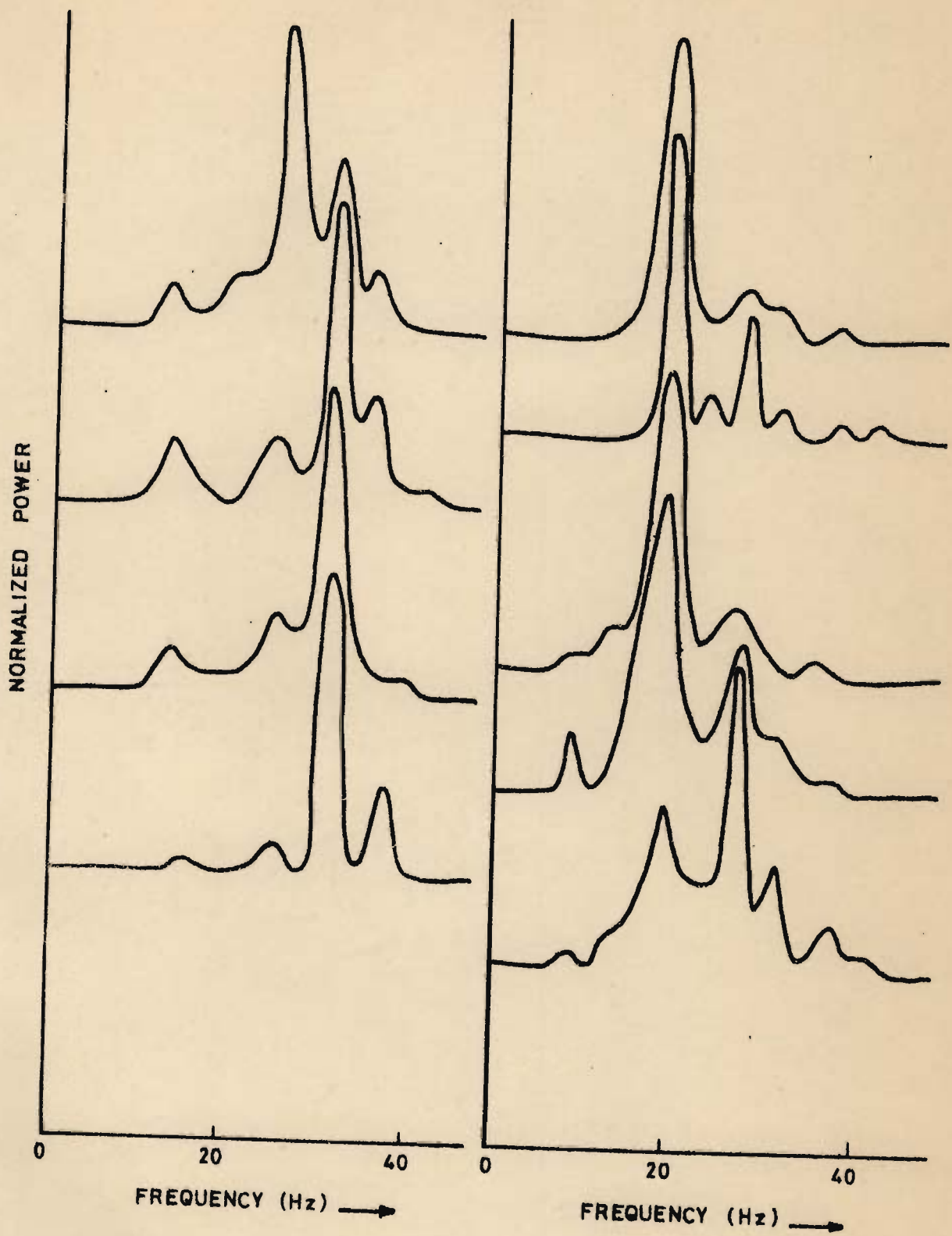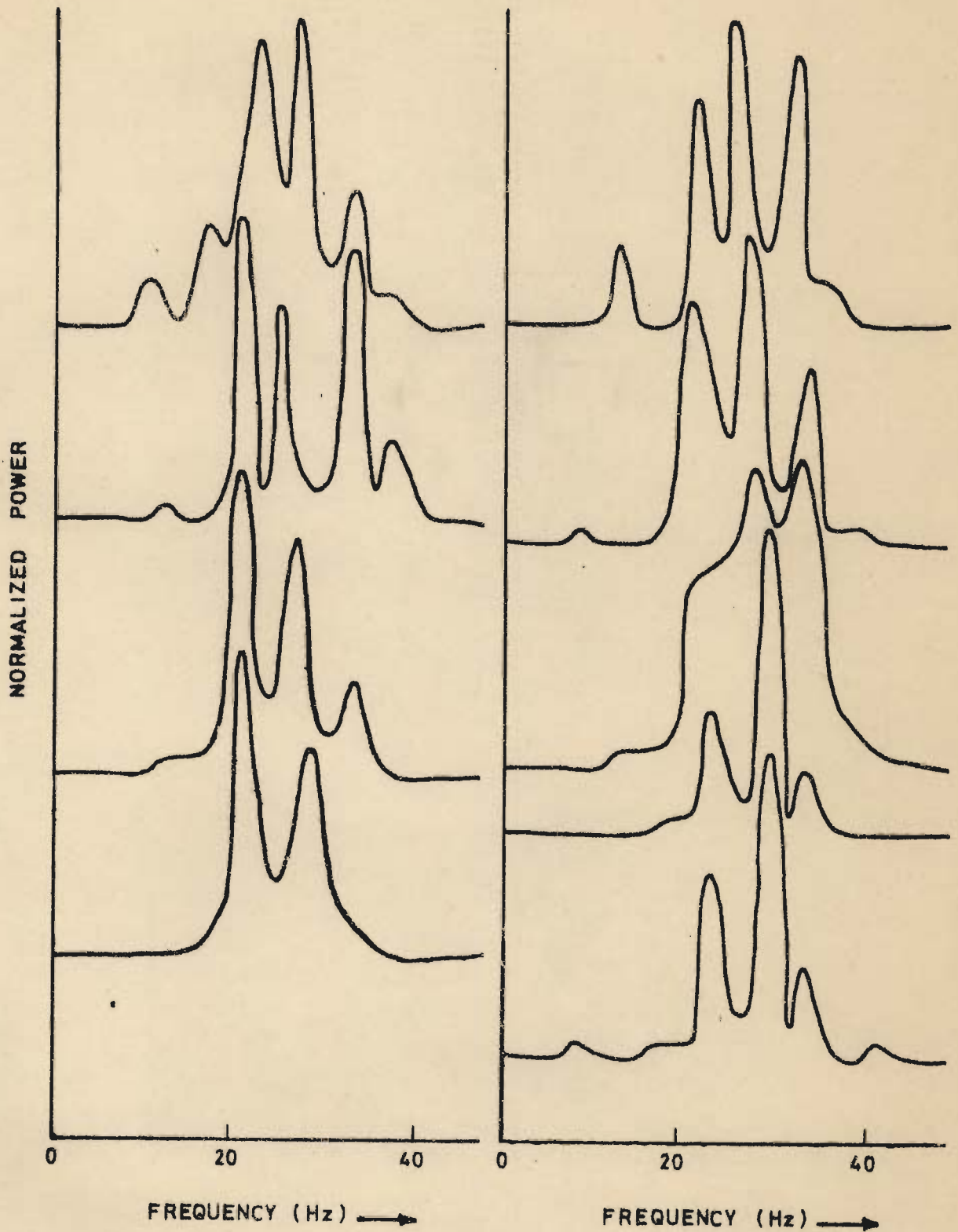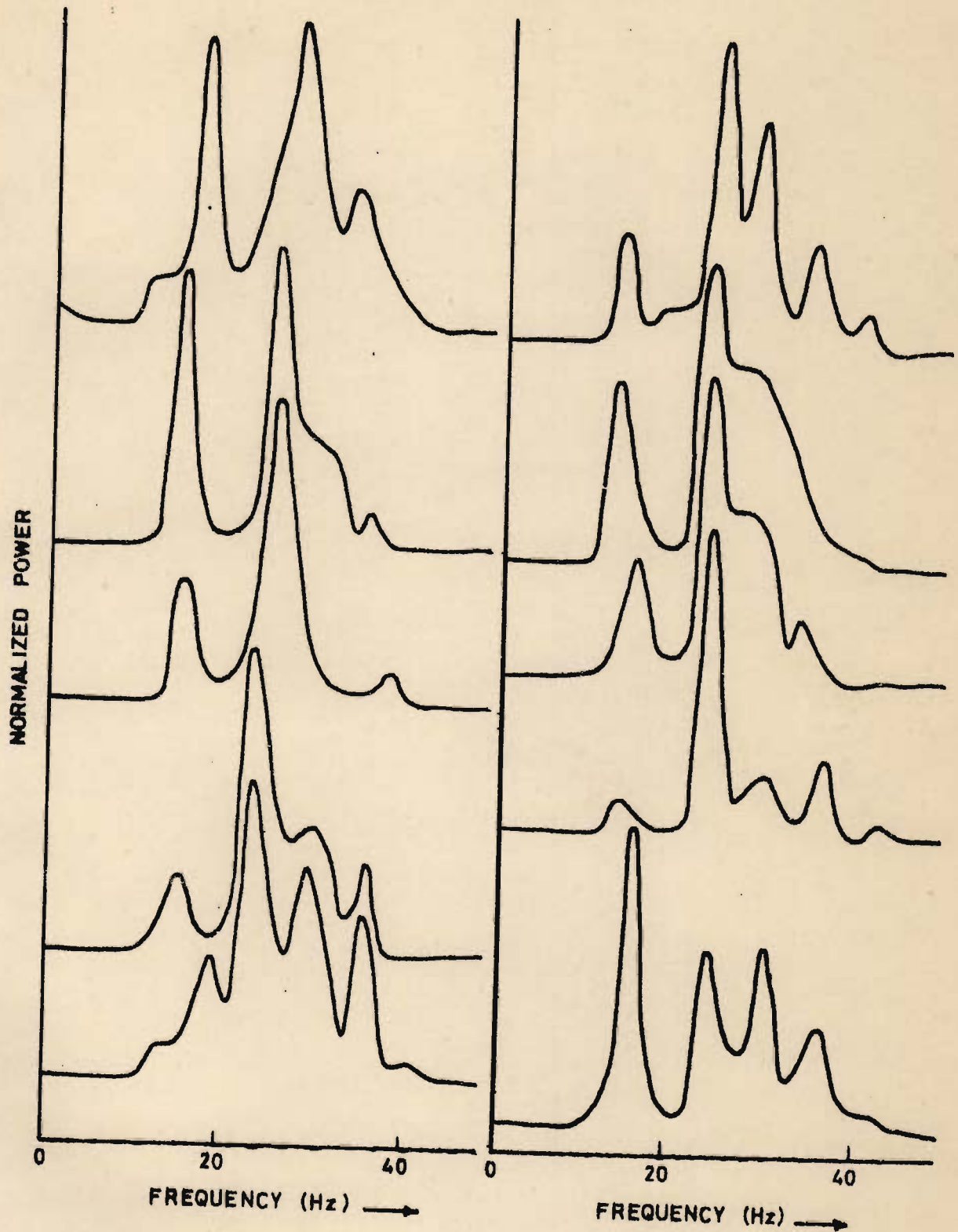FIG.6.15_SOME EXAMPLES FROM THE 148 TRACES OF POWER
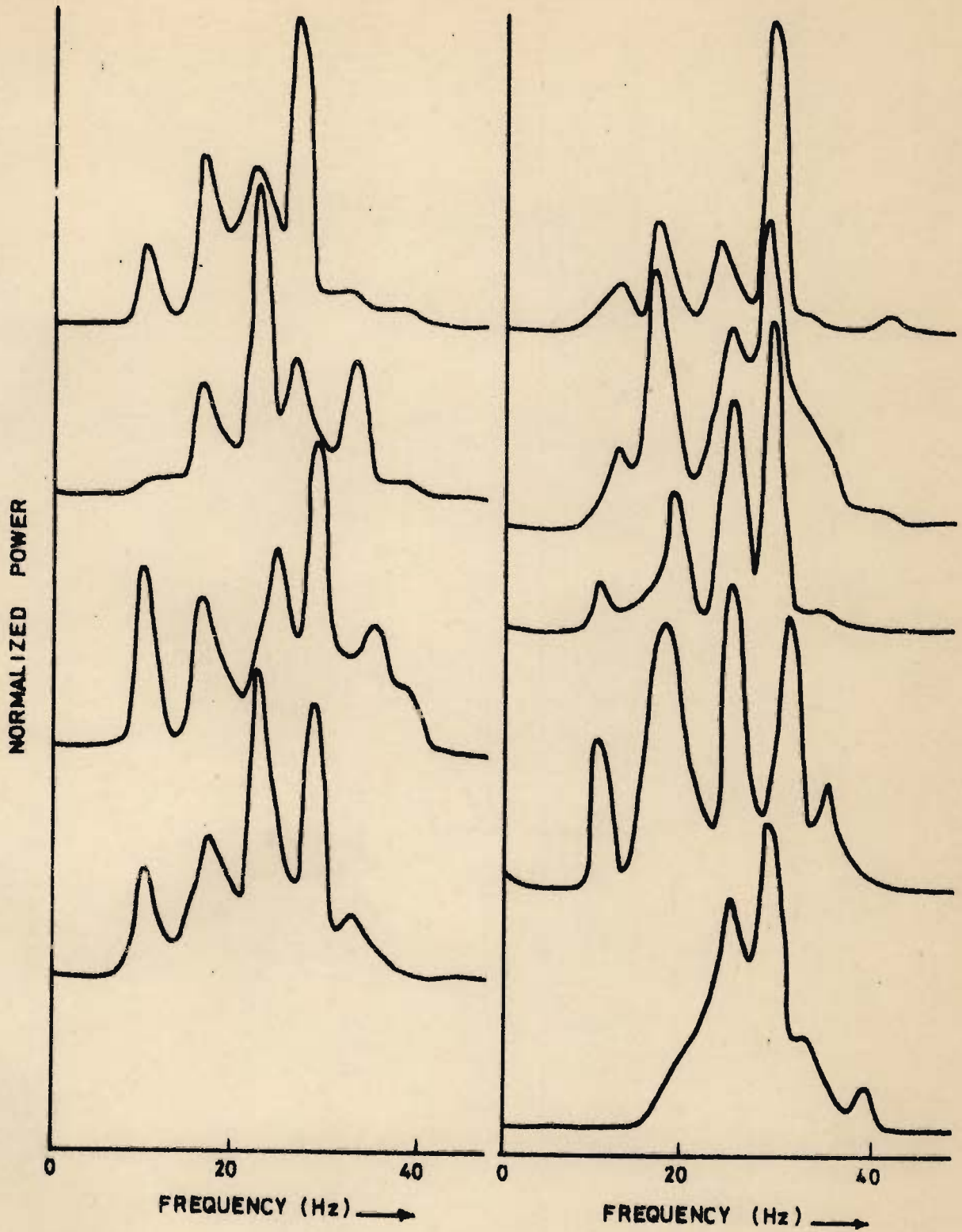SPECTRA ANALYSED FOR AREA Y.

FIG. 6.16_SOME EXAMPLES FROM THE 148 TRACES OF POWER
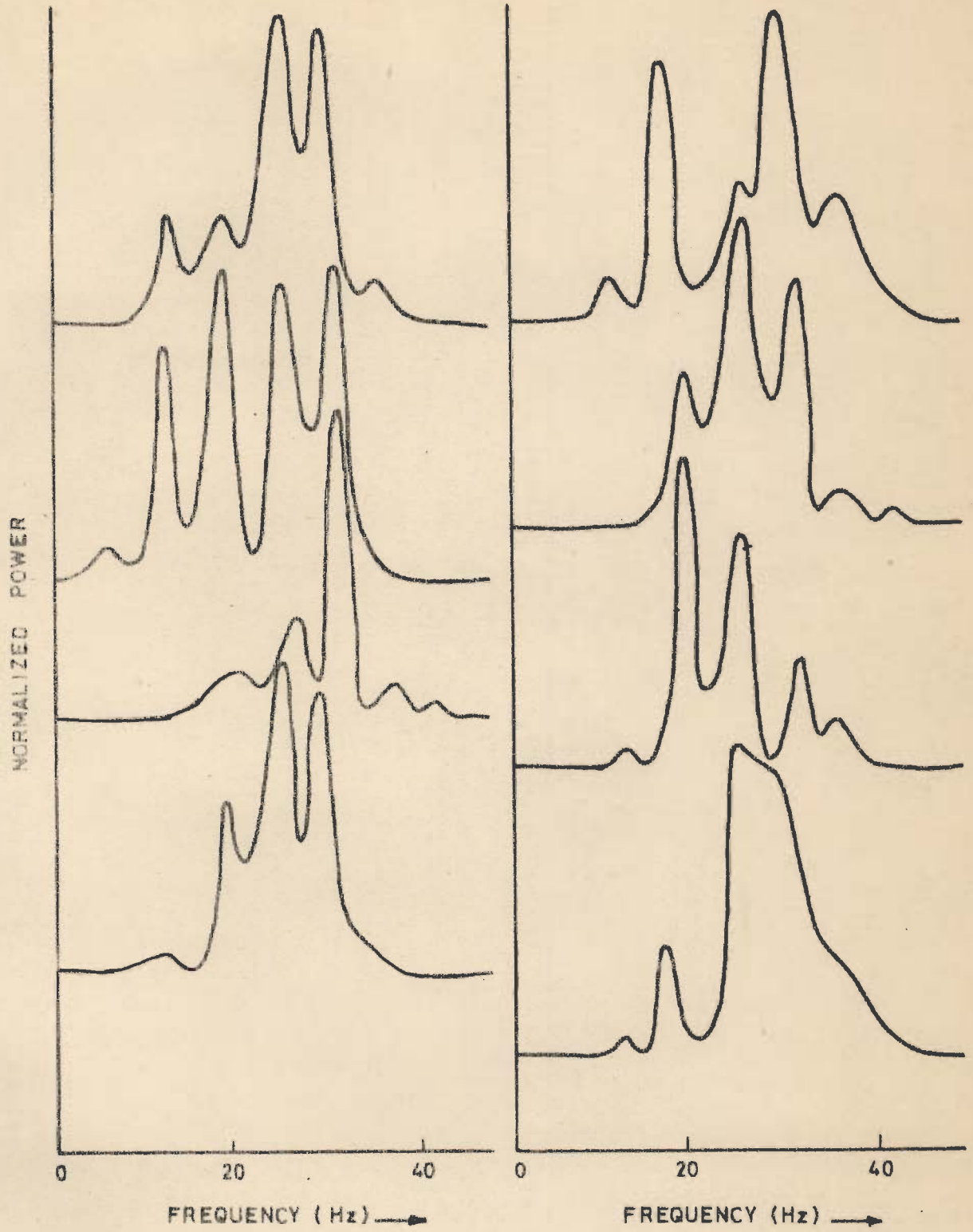SPECTRA ANALYSED FOR AREA Y .

FIG.6.17_SOME EXAMPLES FROM THE 148 TRACES OF POWER

SPECTRA ANALYSED FOR AREA Y.

Each power spectrum for the field seismograms as in the case of synthetic data is characterized by an average frequency, f_1, which takes into account both the frequency and the power content of the spectrum. As the cumulative frequency weighted spectra and the cumulative power spectra are derived from the power spectra they will have the same frequency bandwidth. The frequencies f_2, f_3, f_4, and f_5, f_6 and f_7 therefore lie in a narrow frequency zone as compared to the synthetic case. Some cumulative frequency weighted power spectra for Areas X and Y, with the frequencies f_2, f_3 and f_4 marked on them are shown in Figures 6.18 and 6.19, respectively. Figures 6.20 and 6.21 show some more examples of the frequency weighted power spectra. Figures 6.22 and 6.23 show a few cumulative power spectra for Areas X and Y respectively, with frequencies f_5, f_6 and f_7 marked on them. Figures 6.24 and 6.25 show some further examples of these traces. It is difficult to distinguish visually between the cumulative spectra of the two Areas X and Y. As shall be seen later, this is also evident from the variables derived from the cumulative spectra which show either very small or negative contribution towards discrimination (Table 6.8) of seismograms from the two different areas.

The frequency, f_8, picked from the logarithm of power spectrum is marked on Figures 6.26 and 6.27, and several of these traces are displayed in Figures 6.28 and 6.29. A comparison of Figures 6.26 and 6.27 shows that the frequency f_8 is very different for the two areas, a fact which was not at all

Table 6.8 <u>Discrimination of Seismograms of Area X from Area Y when all 17 variables are considered</u>

Calculated value of $F = 76.3108$, with 17 and 369 degrees of freedom.

Tabulated value of $F = 1.76$, with 17 and ∞ degrees of freedom at $\alpha = 0.05$

Mahalanobis' $D^2 = 14.809$

$R_X = 39.53$

$R_O = 32.12$

$R_Y = 24.71$

Sl.No.	Variable	Constant λ	Percentage Contributed towards discrimination
1.	T_{amin}	-1.08	1.3
2.	T_1	0.27	0.0
3.	T_2	0.42	-0.1
4.	T_3	0.14	0.7
5.	A_{min}/A_0	-3.27	0.0
6.	A_1/A_0	-1.24	0.1
7.	A_2/A_0	-20.63	5.9
8.	A_3/A_0	5.71	-2.1
9.	f_1	0.65	4.2
10.	f_2	-0.37	-0.5
11.	f_3	0.28	0.7
12.	f_4	-0.45	-1.2
13.	f_5	-0.57	-1.9
14.	f_6	-0.71	-1.5
15.	f_7	-0.72	-1.7
16.	f_8	1.85	95.3
17.	f_M	0.26	0.8
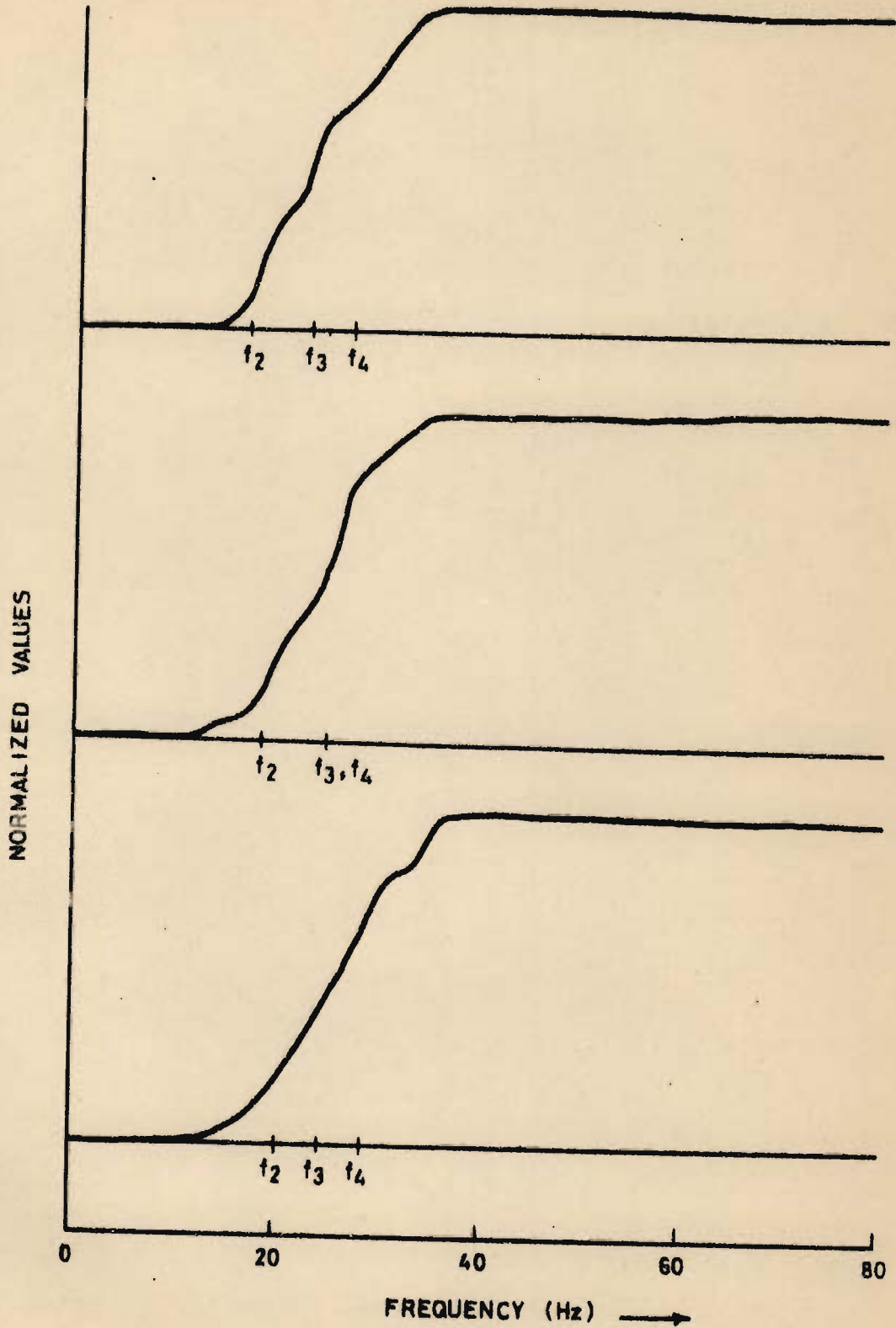
FIG. 6.18 _ SOME CUMULATIVE FREQUENCY WEIGHTED
POWER SPECTRA FOR AREA X .
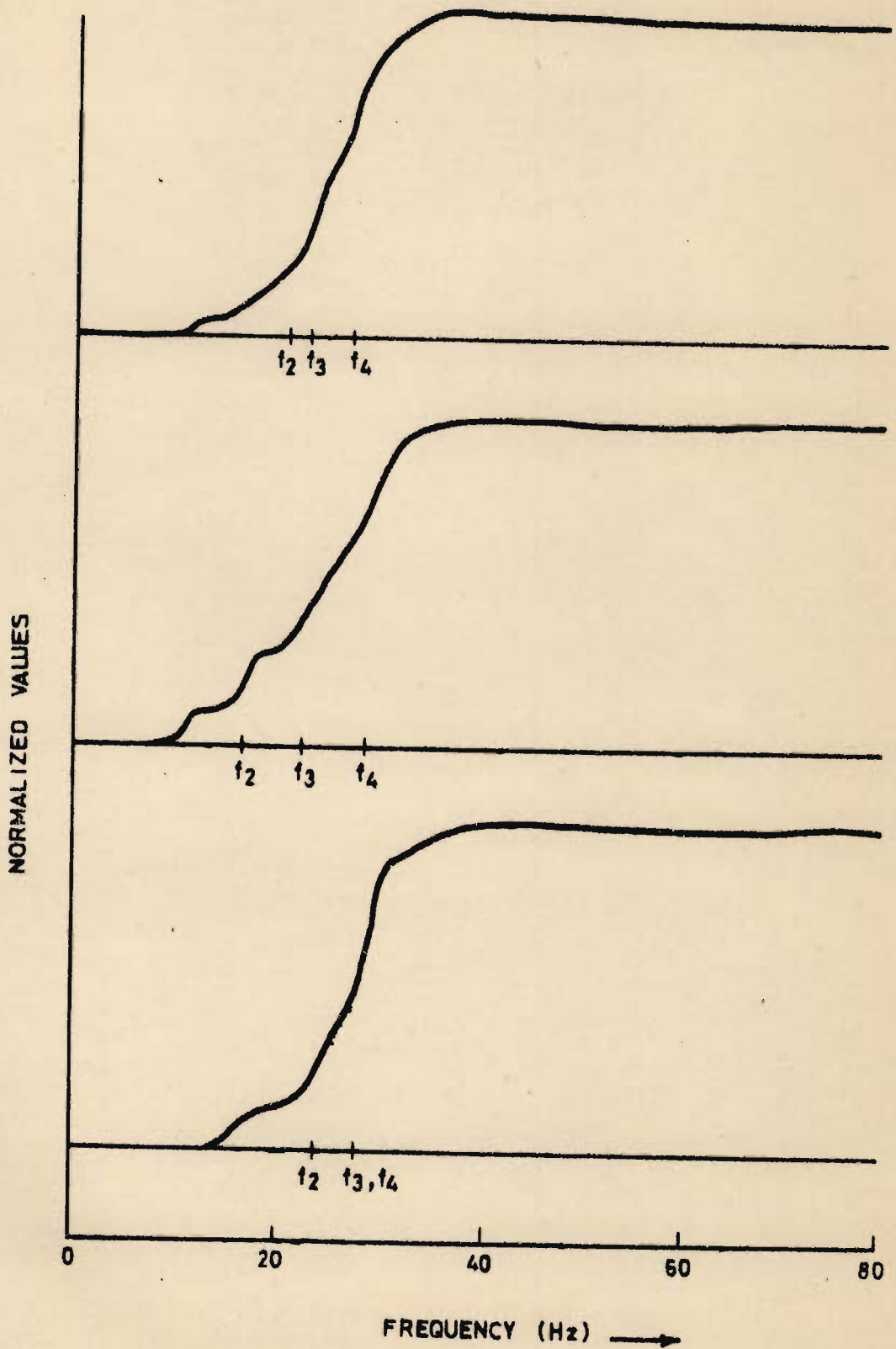
FIG. 6.19 _ SOME CUMULATIVE FREQUENCY WEIGHTED
POWER SPECTRA FOR AREA Y.

FIG.6.20_SOME EXAMPLES FROM THE 239 TRACES OF CUMULATIVE FREQUENCY WEIGHTED POWER SPECTRA ANALYSED FOR AREA X .

FIG.6.21_ SOME EXAMPLES FROM THE 148 TRACES OF CUMULATIVE FREQUENCY WEIGHTED POWER SPECTRA ANALYSED FOR AREA Y.

FIG. 6.22 _ SOME CUMULATIVE POWER SPECTRA FOR AREA X .

FIG. 6.23 _ SOME CUMULATIVE POWER SPECTRA FOR

AREA Y.

FIG.6.24 - SOME EXAMPLES FROM THE 239 TRACES OF CUMULATIVE
POWER SPECTRA ANALYSED FOR AREA X.

FIG. 6.25_ SOME EXAMPLES FROM THE 148 TRACES OF CUMULATIVE

POWER SPECTRA ANALYSED FOR AREA Y.
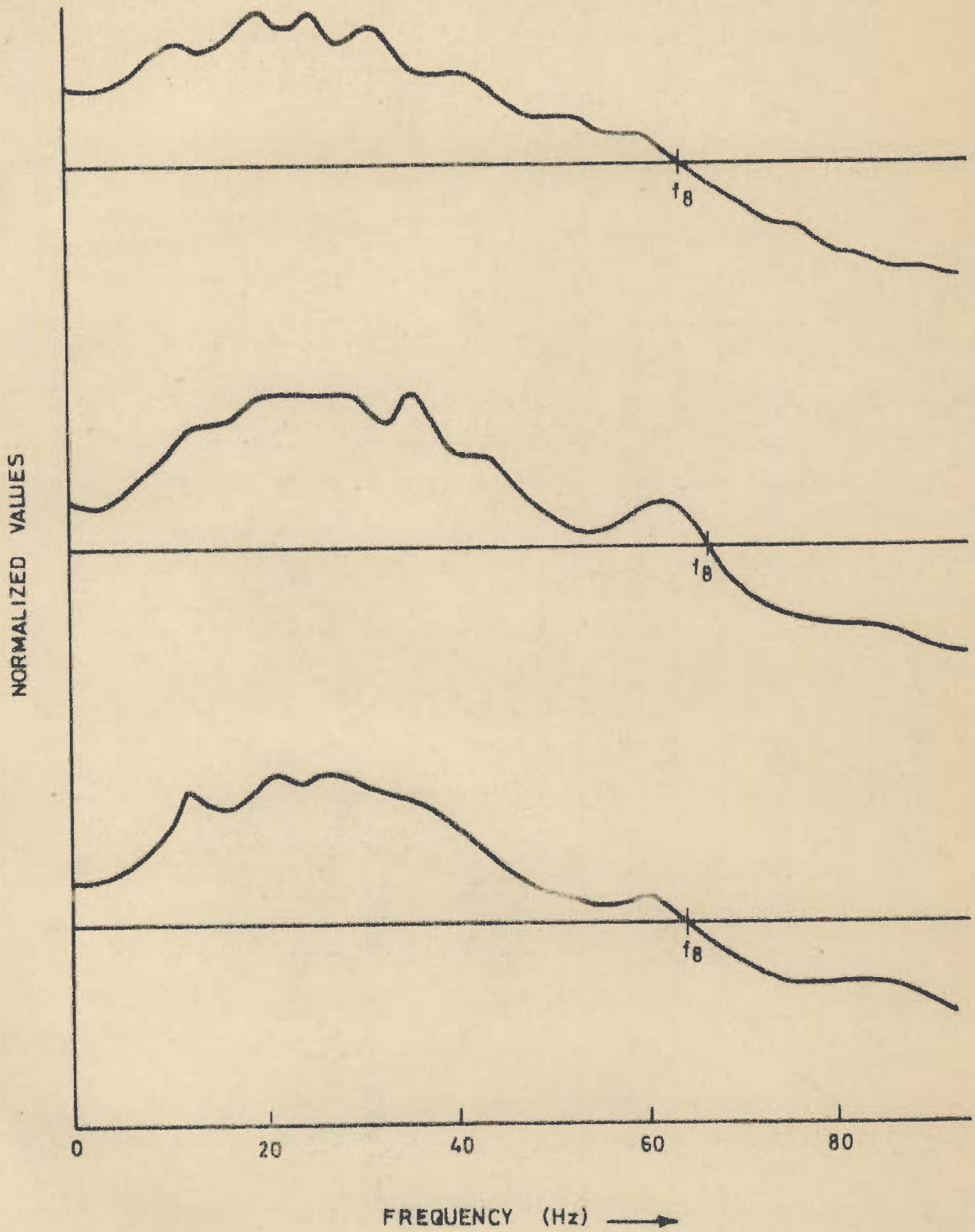
FIG. 6.26_ SOME LOG POWER SPECTRA FOR AREA X .
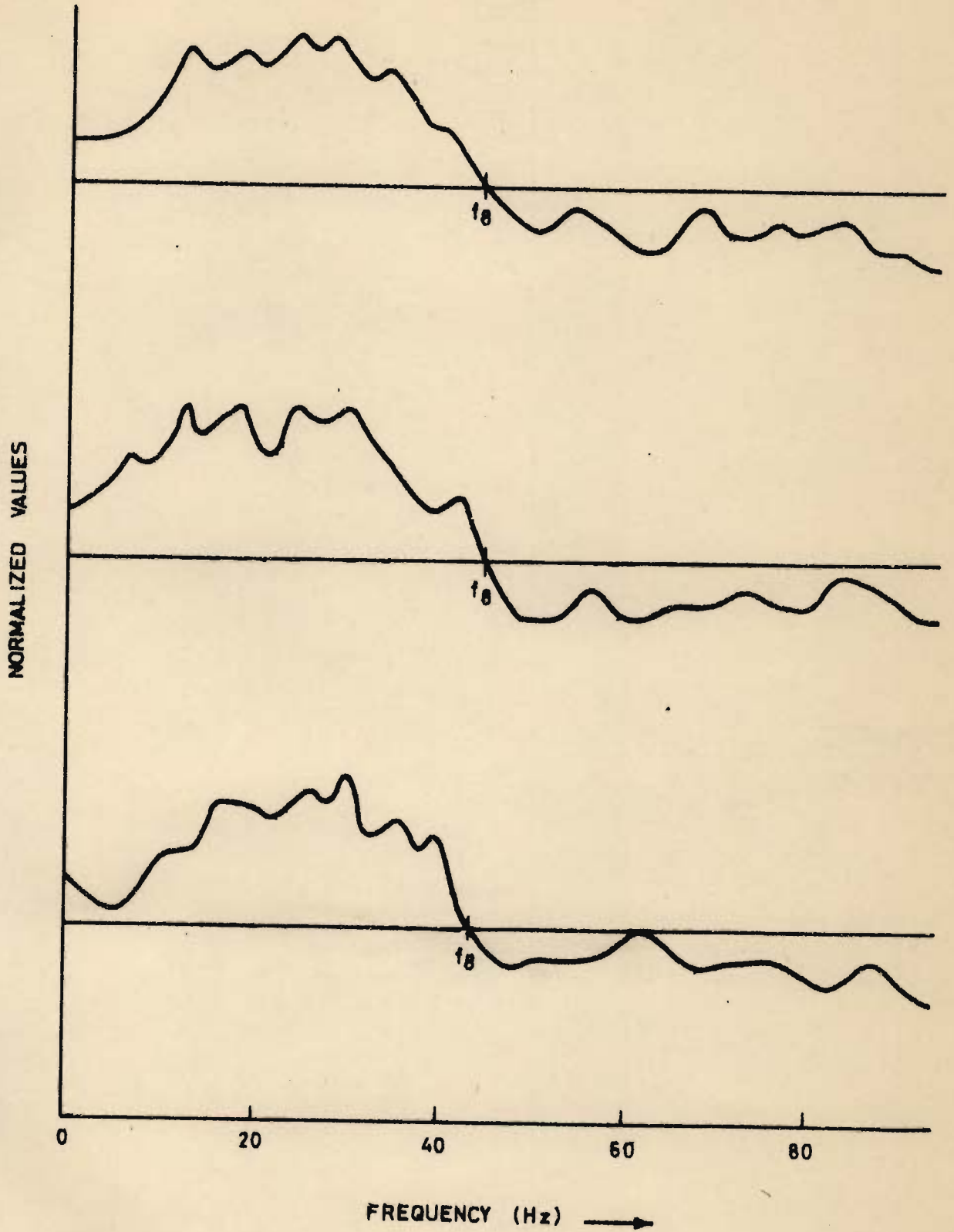
FIG. 6.27 _ SOME LOG POWER SPECTRA FOR AREA Y .
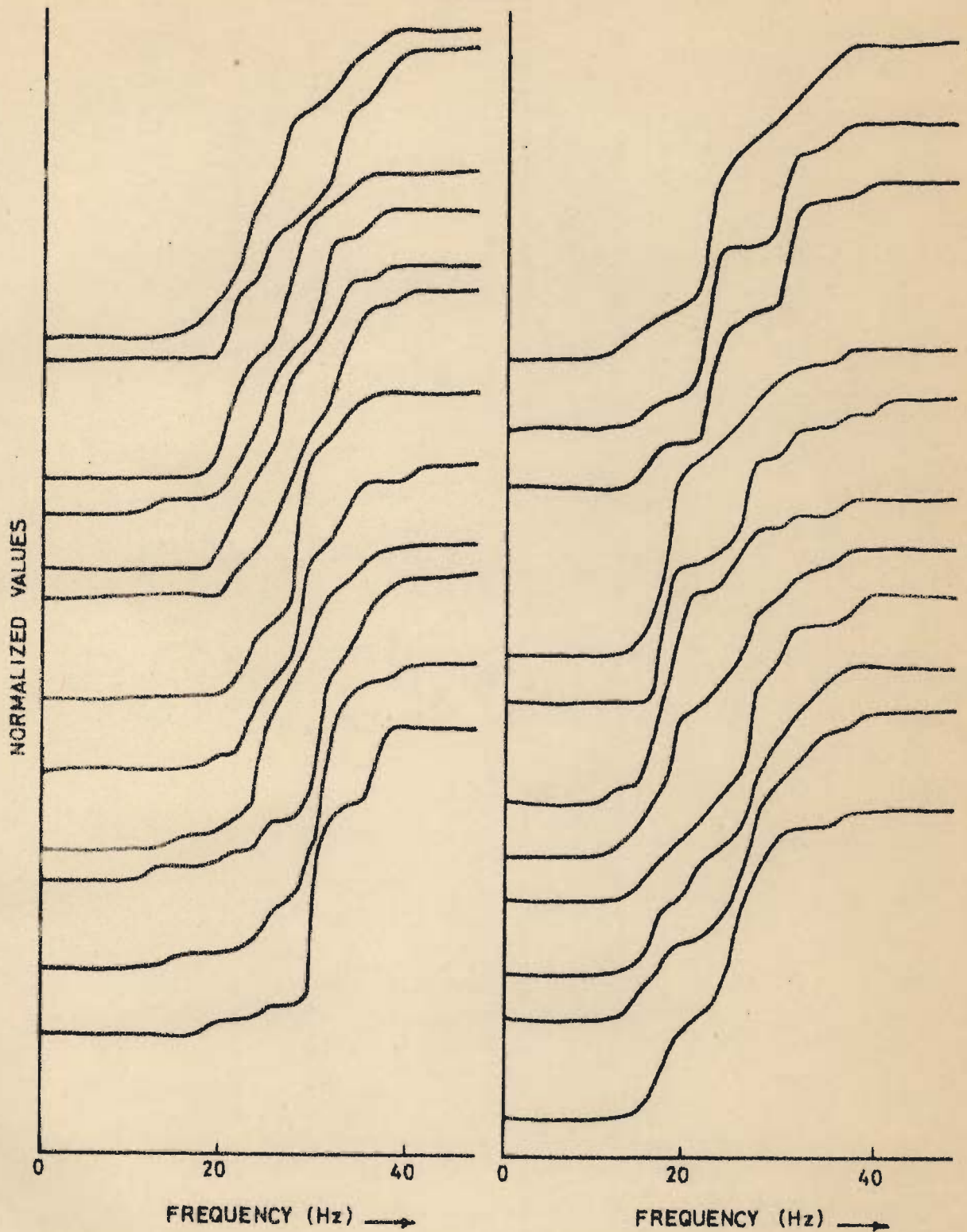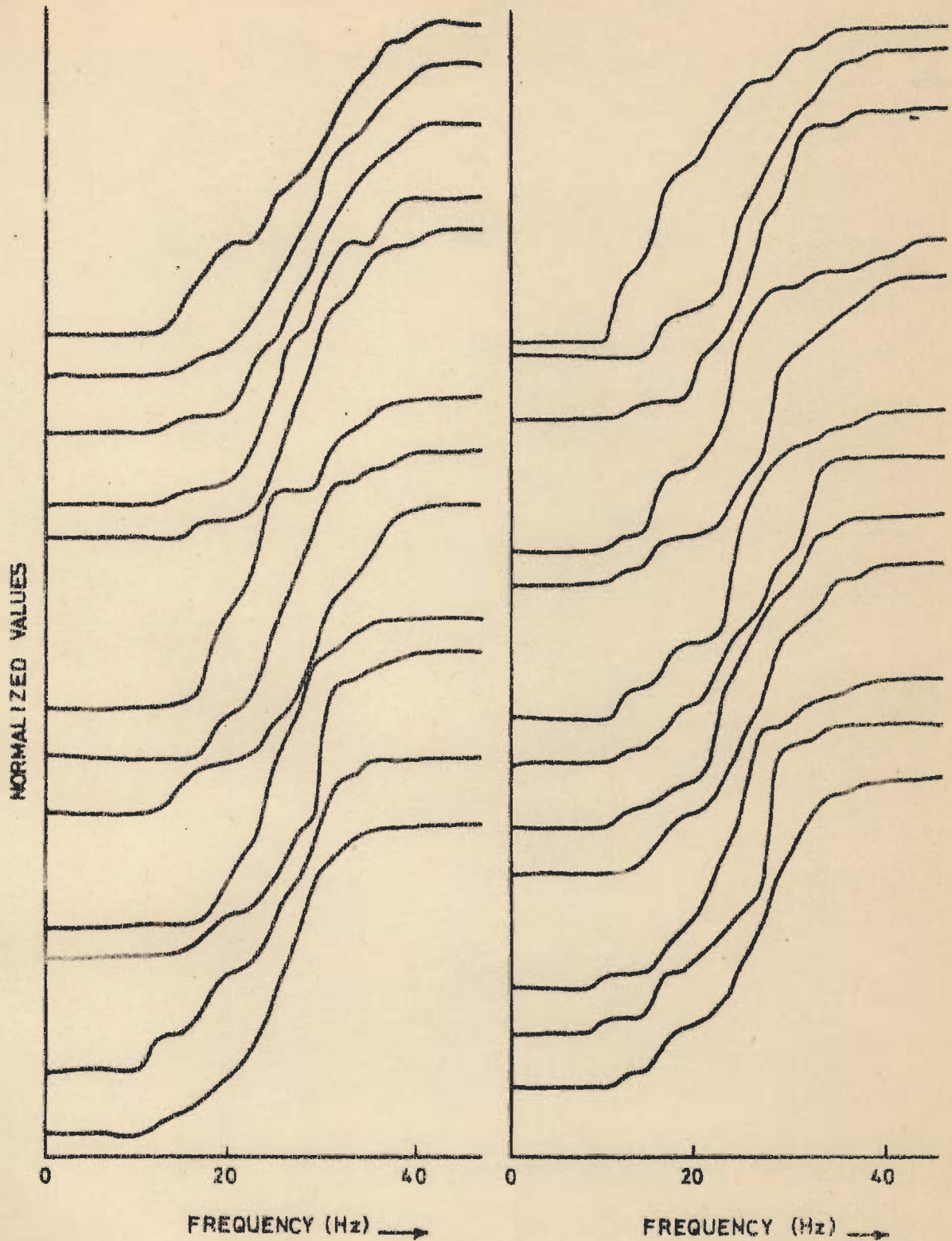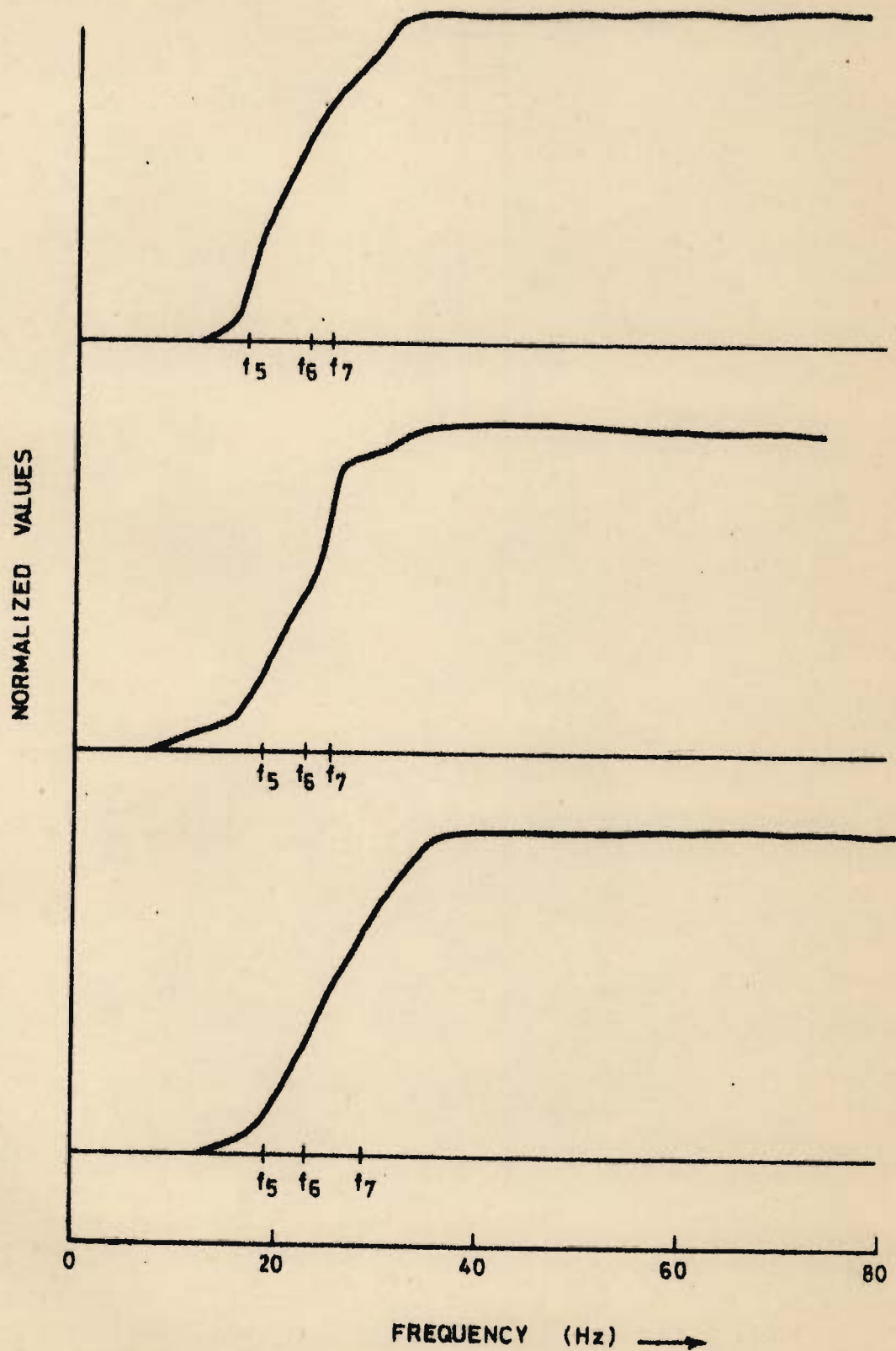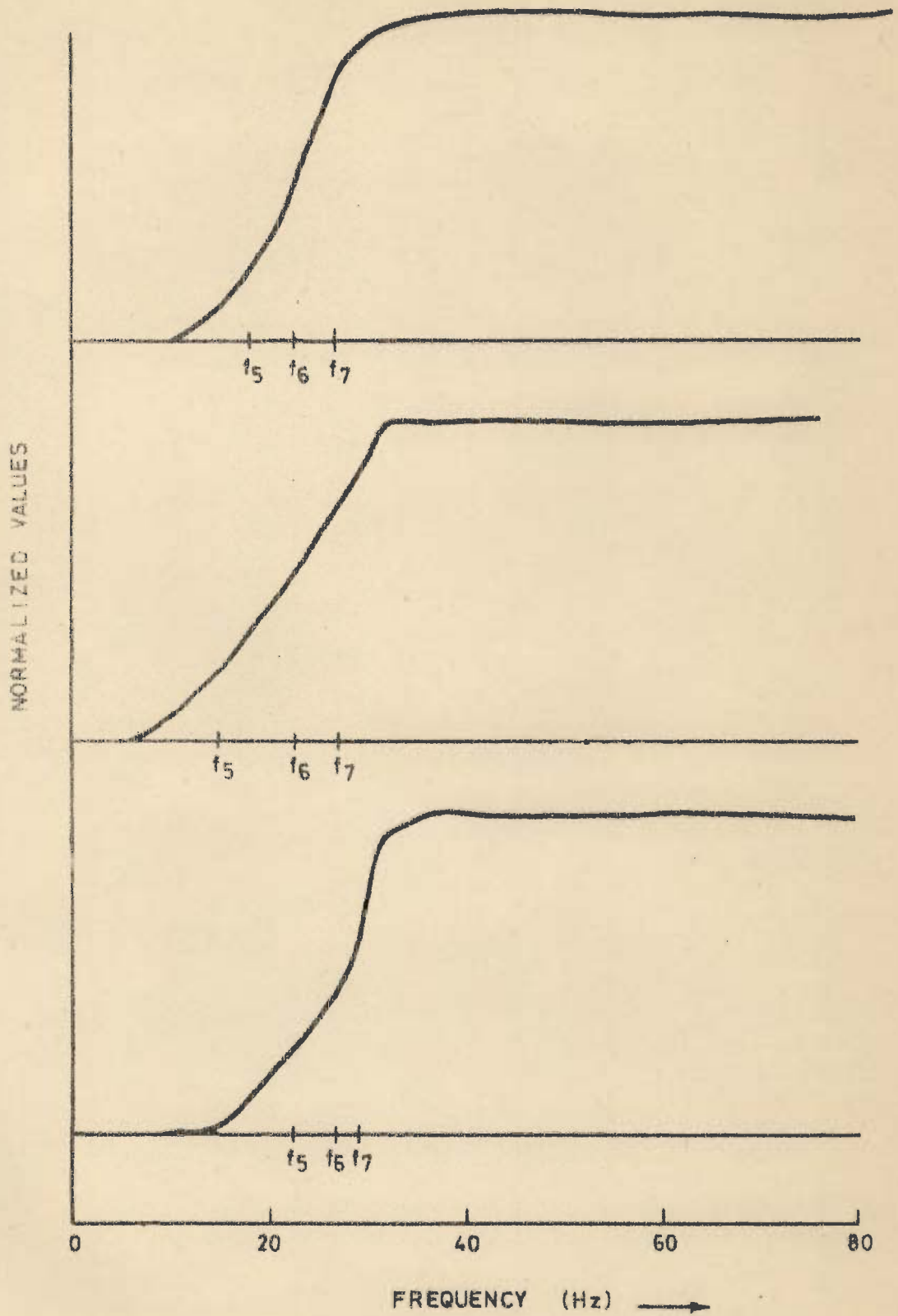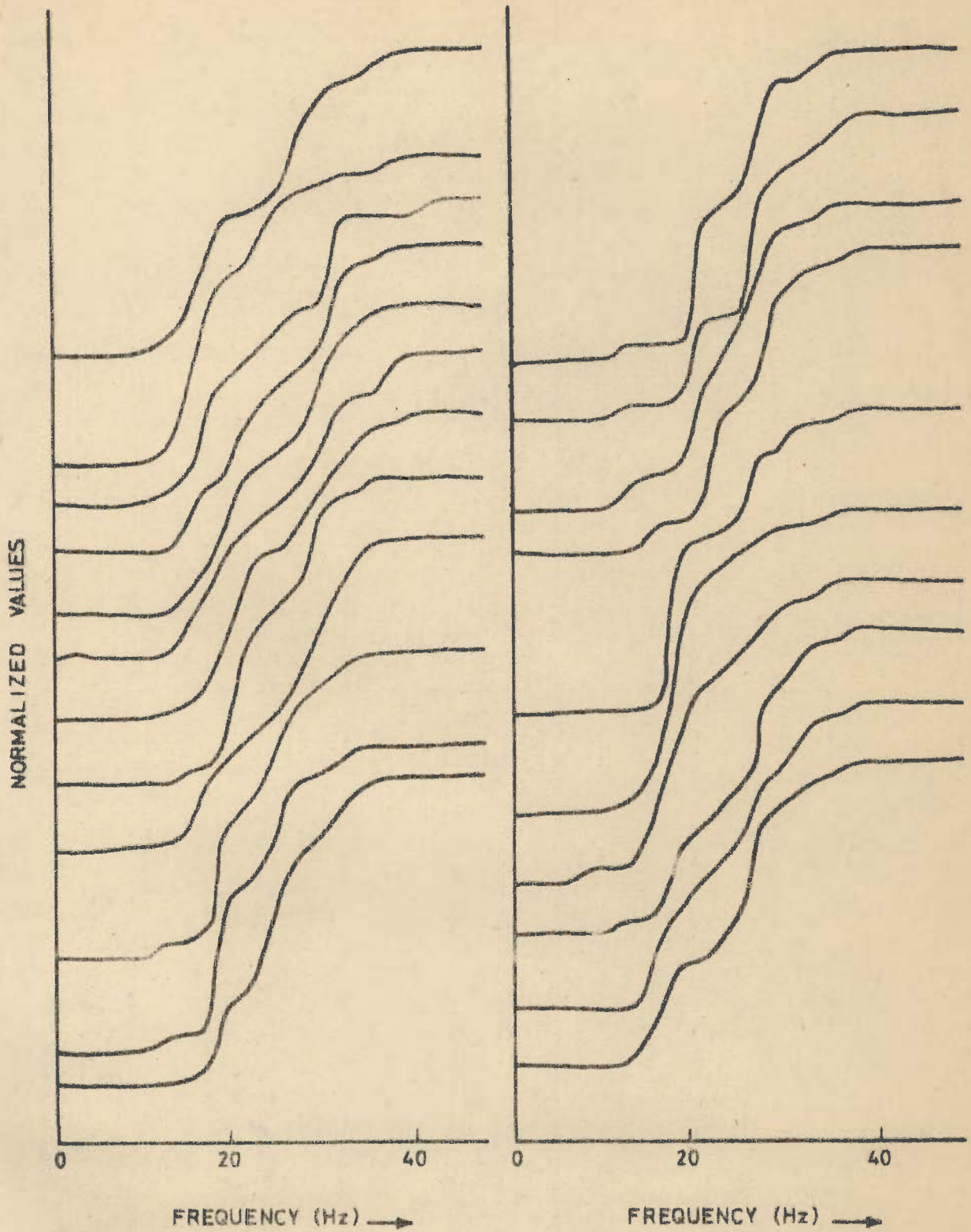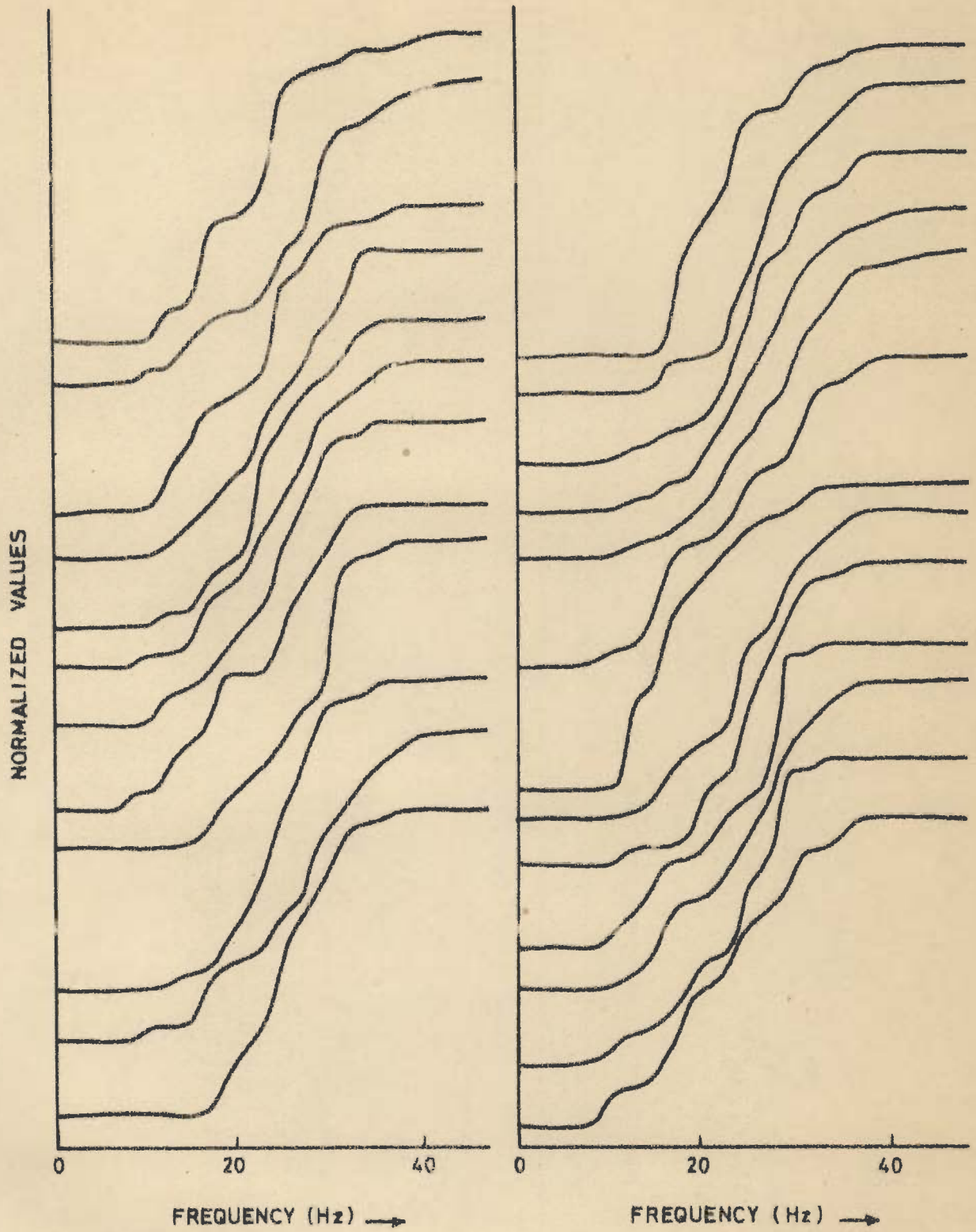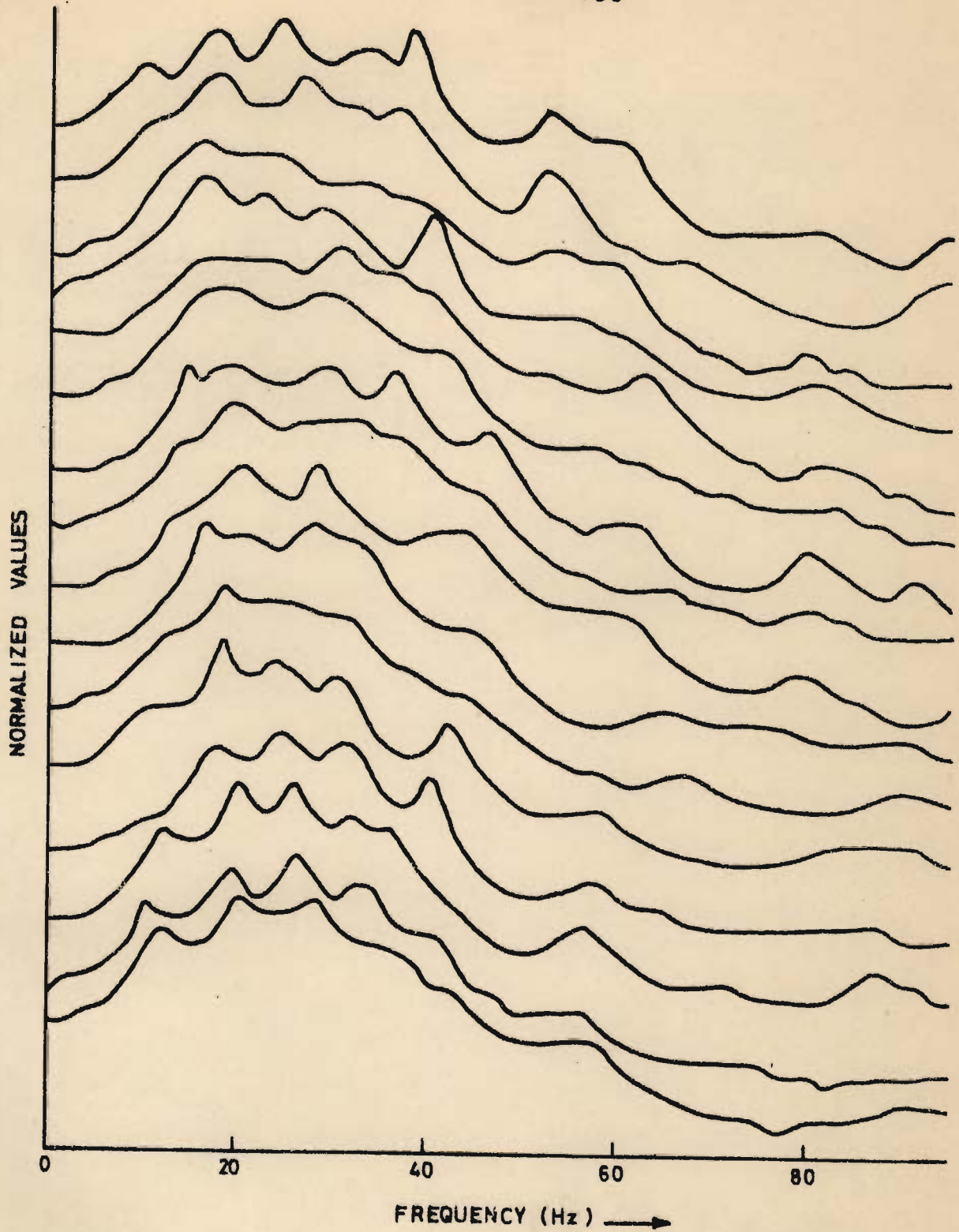
FIG.6.28_SOME EXAMPLES FROM THE 239 TRACES OF LOGARITHM OF POWER SPECTRA ANALYSED FOR AREA X .
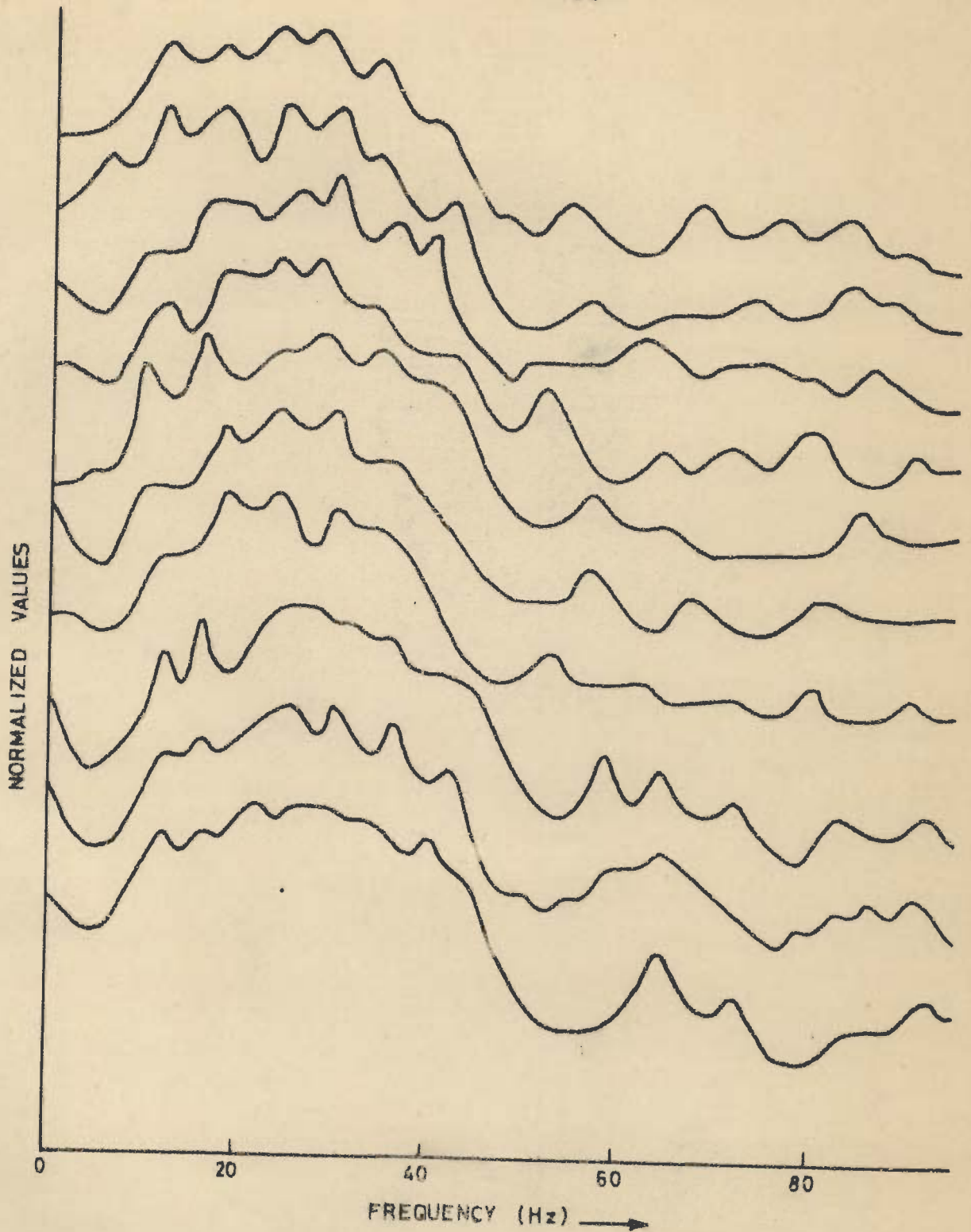
FIG.6.29_SOME EXAMPLES FROM THE 148 TRACES OF LOGARITHM OF POWER SPECTRA ANALYSED FOR AREA Y.

obvious from the power spectra (Figures 6.9 - 6.17), making a very large positive contribution towards discrimination, as is noted from Table 6.8.

6.2 DISCRIMINANT ANALYSIS OF FIELD SEISMOGRAMS

The seventeen variables discussed earlier and shown in Table 6.8, eight of which were from the autocorrelation function and nine from the power spectrum of the seismograms for Areas X and Y were subsequently subjected to discriminant analysis. As a first step towards this analysis, the cumulative distribution of these variables were plotted on the probability paper. These curves shown in Figures 6.30 - 6.33 indicate in general a normal or near normal distribution. However, limited departures from normality do not seriously affect the discriminant function (Davis, 1973).

The discriminant score is calculated from each of the seismograms and is projected on the discriminant function line (Figure 6.34). A perusal of this figure indicates that except for one all the seismograms of Area Y and more than 90 percent of Area X are properly classified. Test for the equality of the variance-covariance matrices (Equation 5.21) gives the calculated values of X^2 as 710.0 as against the tabulated value of 196.6 at 153 degrees of freedom. The null hypothesis of the equality of the two matrices is therefore rejected. However, according to Davis (1973) " In practice, an assumption of equality may be unwarranted".
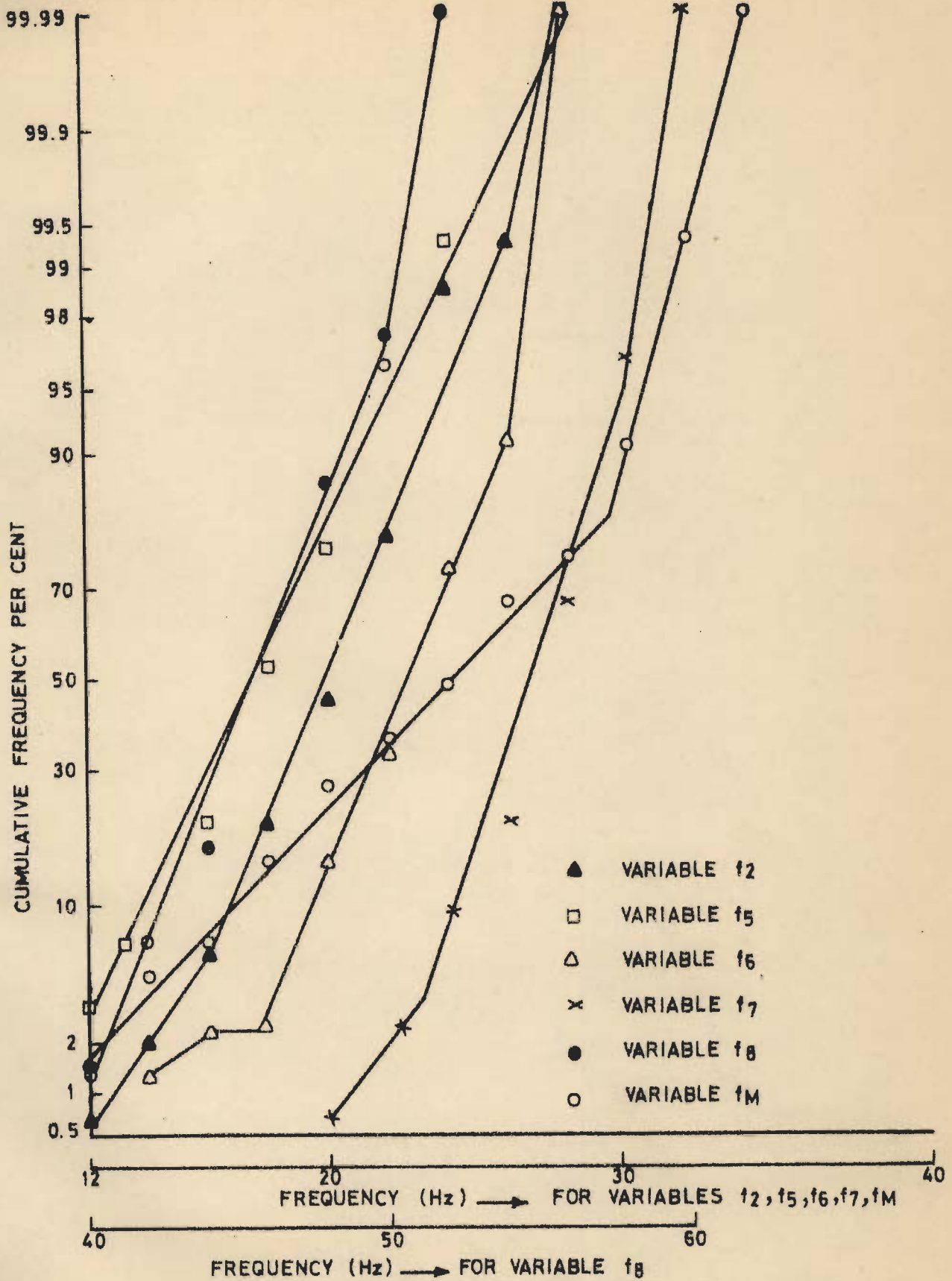
FIG.6.30_PROBABILITY DISTRIBUTION OF VARIABLES $f_2, f_5, f_6, f_7,$ f_8 AND f_M OF AREA Y.
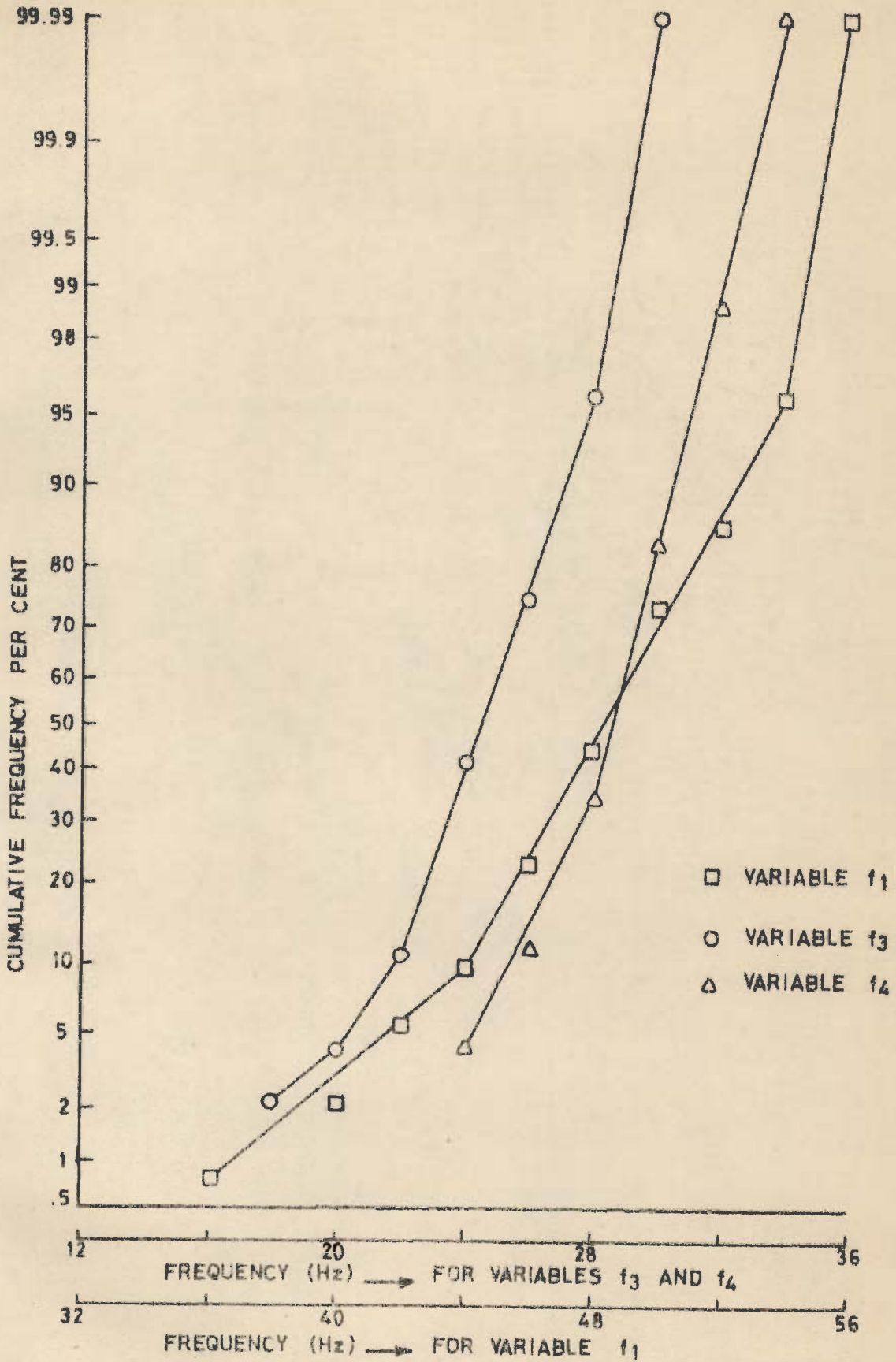
FIG. 6.31_ PROBABILITY DISTRIBUTION OF VARIABLES

f_1, f_3 AND f_4 OF AREA Y.
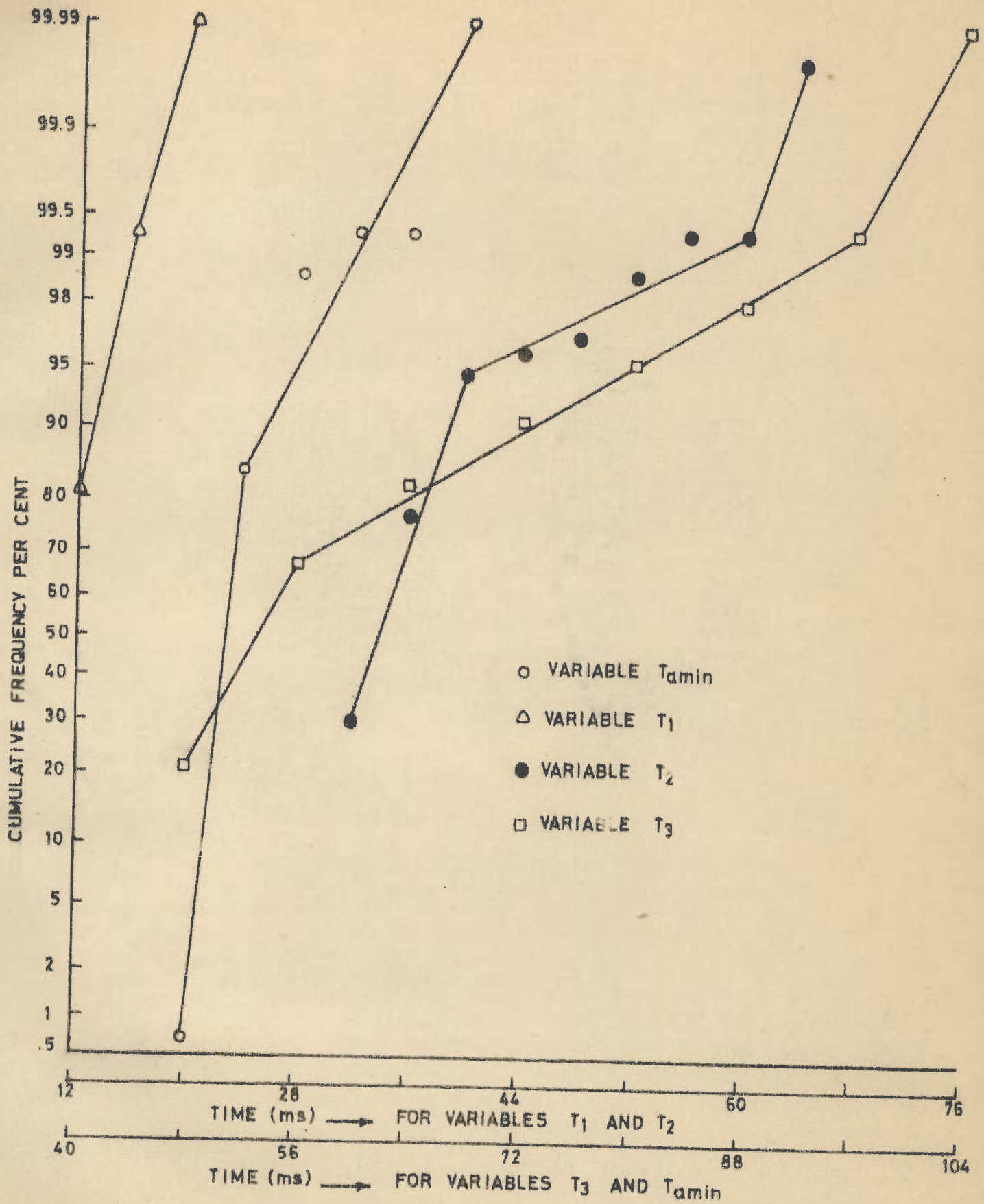
FIG.6.32_PROBABILITY DISTRIBUTION OF VARIABLES T_1, T_2, T_3 AND T_{amin} OF AREA Y.
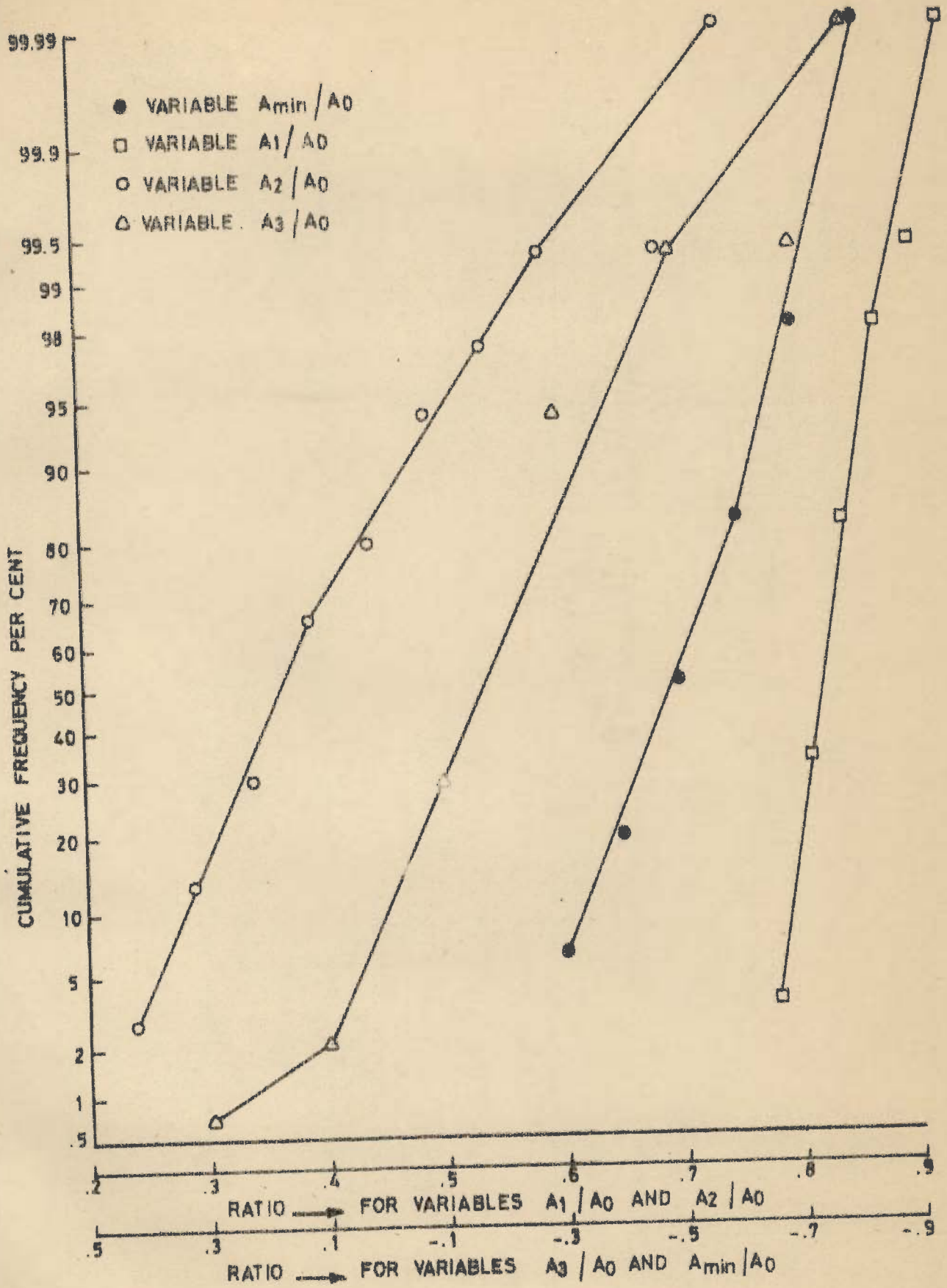
FIG.6.33 _ PROBABILITY DISTRIBUTION OF VARIABLES A_1/A_0, A_2/A_0

A_3/A_0 AND A_{min}/A_0 OF AREA Y.
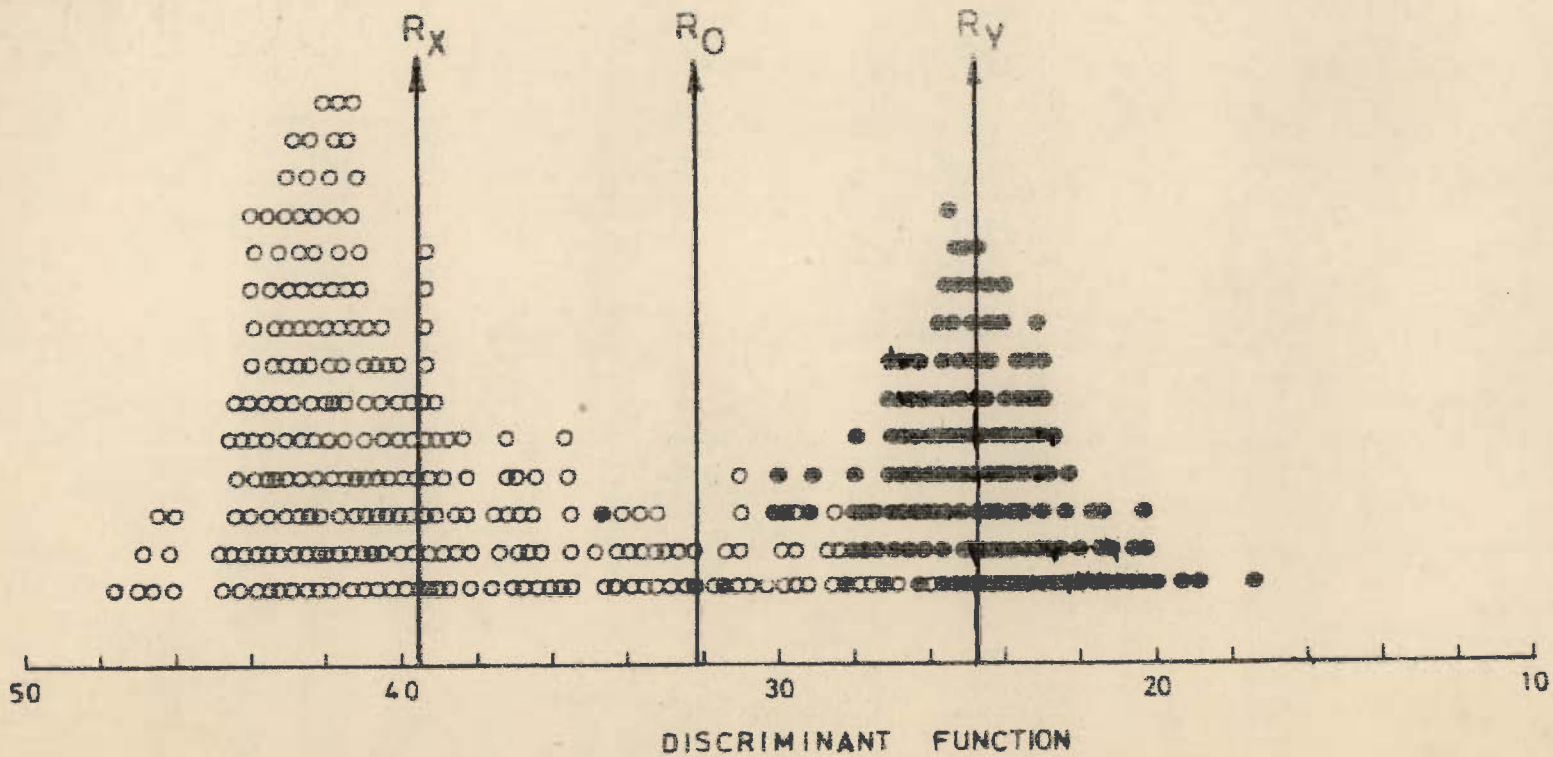
FIG.6.34_ DISCRIMINANT ANALYSIS TO DISTINGUISH AREA X FROM AREA Y.

17 VARIABLES ARE TAKEN INTO CONSIDERATION.

The contribution made by each variable towards discrimination is given in Table 6.8. Frequency variable, f_8, (the frequency at which logarithm of power decreases to zero) gives a very high contribution of 95.3 percent, thereby emerging as a very powerful discriminator. This is in agreement with the discriminant analysis of synthetic data, where the same variable makes a contribution of 30.5 percent. Therefore, the variable f_8 can be used to distinguish between the seismograms from Areas X and Y. However, spectrum of field seismograms is modified by the subsurface lithostratigraphy besides the field recording and processing parameters, such as shot depth and size and recording and processing filters. An increase in the shot depth is expected to give a spectrum richer in higher frequencies, resulting in a corresponding shift of f_8. Parts of one seismic profile of Area X were shot at different depths of 23, 26 and 29 metres, yet no substantial shift in f_8 related to these shot points are observed, thereby infringing, in this case the premise that an increase in shot depth gives rise to higher frequencies. However, f_8 is still subjected to shot size, gain of the recording instrument and the subsequent processing of data. Much of this factor may be taken care of by the trace equalization procedures followed in data processing stage. Therefore inspite of f_8 being somewhat dependent on shot size, gain of the recording instrument etc., it may be used as a useful discriminator of lithologies as found in the present analysis.

Parameter A_2/A_0, the ratio of the autocorrelation
function at lag of two units to zero lag, emerges as the
next best discriminator, with 5.9 percent contribution; in
the synthetic case also this parameter gives a significant
contribution of 2.1 percent. The average frequency, f_1,
contributes 4.15 percent towards discrimination, whereas in
the synthetic case it contributed 6.9 percent, and this also
can be used as a discriminator. The other variables which
make positive contributions in both field and synthetic cases
are T_{amin}, the time of first minima in the autocorrelation
function; A_1/A_0, the ratio of the autocorrelation function at
unit lag to that at zero lag; f_3, the frequency of 50th
percentile value of frequency weighted power; and f_M, the
frequency at which maximum power occurs in the power spectrum.
The values of the constants λ_ks and the percentage contribu-
tion of each variable are given in Table 6.8. This analysis,
therefore, besides discriminating between seismograms of
Area X from those of Area Y, also indicates that out of
seventeen variables only seven, viz., f_8, A_2/A_0, f_1, T_{amin},
A_1/A_0, f_3 and f_M make meaningful contributions towards discri-
mination, in decreasing order of significance.

After discriminating and evolving a classification
criterion a new set of 20 seismograms were collected as test
cases from the known Areas X and Y. As shown in Table 6.8,
if the discriminant score for a particular seismogram is

greater or less than the discriminant index, $R_O = 32.12$, then it can be classified either to the Area X or Y, respectively. The discriminant scores, R_i, (Equation 5.14) were therefore calculated and 90 percent of these twenty seismograms were found to be correctly classified. These test cases therefore demonstrate the validity of this approach and it is therefore concluded that the technique can be used very effectively to solve the problems of discrimination and classification of lithostratigraphy from seismic data.

CHAPTER - VII

DISCUSSIONS AND CONCLUSIONS

A knowledge of subsurface lithostratigraphy of basins without drilling is the oil explorers' dream. The interpretation of subsurface lithostratigraphy directly from seismic data is the solution of this dream, though it is a challenging problem. The present study based on the seismic responses of five hundred and eight synthetic stratigraphic sequences generated using upward transition probability matrices has demonstrated that it is possible to discriminate and identify lithounits using statistical techniques. If several variables can be quantitatively derived from the seismograms and a multivariate strategy be adopted, it becomes possible to decipher lithostratigraphy.

Statistical discriminant analysis has been carried out to distinguish between dominantly sandy, Model E, (sand = 53 percent, shale = 26 percent and coal = 21 percent) and shaly, Model F, (sand = 37 percent, shale = 60 percent, coal = 3 percent) on the basis of seventeen variables. These variables are A_1/A_0, A_2/A_0 and A_3/A_0 where A denotes the autocorrelation function (ACF) at the subscripted lag; A_{min}/A_0, where A_{min} denotes the minimum value of the ACF; T_1, T_2, T_3 where T denotes time of the subscripted zero crossing; T_{amin}, time at which first minima occurs; f_M,

frequency at which maximum power occurs; f_1, the average power weighted frequency; f_2, f_3 and f_4 - frequencies of 25th, 50th and 75th percentile values of frequency weighted power; f_4, f_5 and f_6 - frequencies of 25th, 50th and 75th percentile of power and f_8, the frequency at which logarithm of power decreases to zero. Ten of these, viz., A_1/A_0, f_8, T_2 f_M, f_1, T_1, f_3, A_2/A_0, A_3/A_0 and T_{amin} have been found to contribute positively towards discrimination of gross lithologies corresponding to the Models E and F.

The discriminant analysis has also been applied to field seismic data from two Areas X and Y of a sedimentary basin in India to test the efficacy and demonstrate the applicability of the methodology developed in the present investigations. This analysis, when used as a search technique, indicates that from amongst the eight positive variables : f_8, A_2/A_0, f_1, T_{amin}, f_M, f_3, A_1/A_0 and T_3; identified only seven are common between the synthetic and field seismograms. These are f_8, A_2/A_0, f_1, T_{amin}, f_M, f_3 and A_1/A_0, and have been termed as 'seismic discriminators'.

To generate synthetic seismograms, the same source wavelet and noise character was used in all the cases. Thus the computed responses from the various simulated lithostratigraphies would be mainly characterizing the latter. The deviations on account of noise were observed to be small on account of the large signal to noise ratio.

The variables which give positive contributions towards discrimination are considered to be meaningful discriminators, as these contributions make the Mahalanobis' distance larger. Large distances correspond to strong discrimination. Variables which give negative contribution are detrimental in this analysis. Other workers e.g. Davis(1973) who have used this technique are silent about this negative aspect. However, Sinvhal, Gaur, Khattri, Moharir and Chander, (1979), made an attempt to analyse the effect of eliminating such variables and repeated the entire analysis to find that some of the variables which previously gave positive contributions now made negative contributions. This shows that the sign of contribution of a particular variable also depends upon the set of variables considered in the discriminant analysis. It does not seem that an analytic method to select the subset which will contribute to the distance, D^2 in the positive sense is available, in fact, there may be no such subset of variables.

The modeling of lithostratigraphic situations embodied in the present study and the methodology developed for discrimination are subject to certain limitations. All the lithostratigraphic sequences have been generated with an approximate thickness of 200 m so that a fairly large number of lithologic transitions may be encountered. Therefore each lithounit considered is approximately 4 m thick. This model would restrict the thickness of various component lithounits

to multiples of approximately 4 m.   In nature the thickness
can assume any value, therefore the model would represent
actual thicknesses within a maximum error of ± 2 m. Since the
resolving capability of seismic methods is well below this
order of thickness, therefore this approximation does not
represent any serious limitation of the present work.

The choice of only three distinct lithological
states - sandstone, shale and coal with no provision for
the transition zones also may not seriously affect the
results.   For transition zones which, for example, consist
of silts or carbonaceous shales may not show a significant
variation in impedance contrast and may be grouped into one
of the three states mentioned above.

The assumption of a homogeneous overburden in the
models does not give rise to any reflections from within, and
an accompanying modification of the waveform passing through
it does not occur.   This may constitute a significant devia-
tion from real situations where the layered overburden does
play a significant role in imparting its characteristics
on the reflections arising from layers below it.   The effect
may have serious consequences particularly if the overburden
is highly varying from area to area.   Although the responses
are calculated for normal incidence with no provision for
attenuation or dispersion it is not considered a serious
drawback of the model as in reflection prospecting the angles

of incidence are either close or reduced to normal.
Attenuation is not severe in the frequency band of interest.

The average matrices constructed for the two areas
X and Y are infested with the usual problems of averaging
a data set from diverse areas, but this average matrix still
characterizes broad depositional environments.

The discriminant analysis based on the above model-
ing procedure for simulated data has been found to be
successful showing that the limitations discussed above are
not significant. Furthermore, the same procedure has proved
successful when applied to field seismic data confirming the
suitability of the model and the corresponding method for
actual analysis. Thus as discussed the limitations were
not serious in the field data analysed. The methodology
presented here may therefore be considered as a major step
towards the successful determination of subsurface litho-
stratigraphy.

For discriminant analysis certain assumptions of
normality, equality of variance-covariance matrices and the
independence of variables have to be met. However, in
practice, it is not always possible to satisfy all assump-
tions rigorously in any analysis - some invariably fail.
However, limited departures from such requirements do not
seriously affect the analysis. For example, departures from
equality of the variance-covariance matrices have to be

accepted in practice (Davis, 1973), and similarly limited
departures from normality may also have to be tolerated.
As discussed previously, the analysis concerning field data
does suffer from both of the above mentioned limitations.

For making the discriminant analysis more effective,
one has to account for as many known variables entering the
data (seismogram) as possible. One such variant is the
source pulse. The source pulse may be dependent on shot
size, pattern of shots, shot depth etc. This factor may be
particularly strong on land records. The marine data would
be relatively free from this effect if adequate care is taken
to use similar sources and recording parameters throughout
the area of survey. In land survey, Vibroseis system seems
to be a good method for achieving uniformity of source wave-
form. Similarly uniformity in field and data processing
procedures would be desirable.

The discriminant analysis would be more effective
if the seismograms are as 'wide frequency band' as possible
retaining information in the higher frequencies. The band-
width of field data usually gets narrowed down by the
processing filters. The range of frequencies required for
the use of discriminant parameters based on power spectra
should be extended and this can be achieved by deconvolution
procedures. Besides, the Vibroseis systems permit the input
signal to be known and controlled. A considerable resolution

can be achieved in the reflection seismograms using this technique and applying appropriate pulse compression methods. A limitation that in some cases may exist on the effective use of the frequency variables, is due to the frequency dependent attenuation exhibited by materials. This effect may be severe for deep reflectors, in which case the high frequencies would suffer more and the characteristic features in the spectrum may be attenuated to such an extent that the discriminatory character is lost. If adequate compensation for such losses can be applied, advantage of this discriminatory parameter may be taken in the case of reflections from deeper horizons as well.

The usefulness of the present analysis in exploration programs for hydrocarbons may be outlined as follows :

(i)   An extensive analysis of the known basins would help establish populations of the various seismic parameters, those used here and perhaps newer ones, corresponding to various geological environments of depositions. These could serve as standard reference for classifying virgin or partially explored areas on the basis of their seismic responses.

(ii)   The method can lead to basic information on the variation of lithostratigraphy in various parts of the basin under exploration. By studying reflection

bands at various depths along the profiles to infer lithostratigraphy at various depths and its lateral variation a depositional model for the basin can be formulated. Needless to say that such an information, coupled with the conventional structural picture and the seismic indicators for fluid content, would add another dimension to the exploration concept. It would lead to more cost effective exploration programs.

(iii) A similar analysis may be attempted to establish discriminant functions based on seismograms to predict fluid content. The success in this regard would indeed be a major advance in exploration methods. However, such an investigation remains for the future.

R E F E R E N C E S

Akaike, H., 1969a, Fitting Autoregressive models for Prediction : Ann. Inst. Statist. Math., v.21, p.243-247.

---------- , 1969b, Power Spectrum Estimation through Autoregressive Model Fitting : Ann. Inst. Statist. Math., v. 21, p. 407-419.

---------- , 1970, Statistical Predictor Identification : Ann. Inst. Statist. Math., v.22, p.203-217.

Anderson, T.W., and L.A.Goodman, 1957, Statistical Inference about Markov Chains : Annals of Mathematical Statistics, v. 28, p.89-110.

Avasthi, D.N., and S.K.Verma, 1973, Analysis of the Statistical Structure of Seismic Reflection for Delineation of Stratigraphic Traps for Oil : Symposium on Recent Trends in Exploration of Minerals, Oil and Ground Water, Sponsored by Academy of Sciences USSR and INSA, India, New Delhi, India, Oct. 15-20, 1973.

Awasthi, A.K., 1979, Sedimentological Studies of the Jodhpur Group in the District of Jodhpur and Nagaur, Rajasthan, India : Ph.D. dissertation, University of Roorkee, Roorkee, 314 p.

Bergland, G.D., 1969, A Guided Tour of the Fast Fourier Transform : IEEE spectrum, v.6, p.41-52.

Blackman, R.B., and J.W.Tukey, 1958, The measurement of power spectra : Dover Publication, Inc., New York, 190 p.

Brown, Jr., L.F., and W.L.Fisher, 1977, Seismic - Stratigraphic Interpretation of Depositional Systems : Examples from Brazilian Rift and Pull - Apart Basins : (In AAPG Memoir no.26, Ed. C.E.Payton), 516 p.

Burg, J.P., 1967, Maximum Entropy Spectral Analysis : 37th Annual International SEG Meeting, Oklahoma, Oct. 31,1967, Preprint-Texas Instruments, Dallas.

----------, 1970, A new Analysis Technique for Time Series Analysis : Presented at NATO Advance Study Institute on Signal Processing with Emphasis on Underwater Acoustics.

Carr, D.D., A.Horowitz, S.V.Hrabar, K.F.Ridge, R.Rooney, W.T.Straw, W.Webb, and P.E.Potter, 1966, Stratigraphic Sections, Bedding Sequences and Random Processes: Science, v.154, p.1162-1164.

Chayes, F., 1964, A petrographic distinction between Cenozoic volcanics in and around the open ocean : Jour. Geophysical Res., v.69, p.1573-1588.

Chen, W.Y., and G.R. Stegen, 1974, Experiments with Maximum Entropy Power Spectra of Sinusoids : Jour.Geophysical Res., v.79, p.3019-3022.

Claerbout, J.F., 1968, Synthesis of a Layered Medium from its Acoustic Transmission Response : Geophysics, v.33, p.264-269.

----------, 1976, Fundamentals of Geophysical Data Processing:
with applications to petroleum prospecting : McGraw-
Hill Book Co., New York, 274 p.

Clement, W.A., 1977, A Case History of Geoseismic Modeling of
Basal Morrow-Springer Sandstones, Watonga-Chickasha
Trend : Geary, Oklahoma - T13 N, R 10W : (In AAPG
Memoir no.26, Ed. C.E.Payton), 516 p.

Cook, E.E., and M.T.Taner, 1969, Velocity Spectra and their
use in Stratigraphic and Lithologic Differentiation :
Geophysical Prospecting, v.17, p.433-448.

Davis, J.C., 1973, Statistics and Data Analysis in Geology:
John Wiley and Sons, Inc., New York, 550 p.

----------, and R.J.Sampson, 1966, Fortran II Program for
Multivariate Discrimination Analysis using IBM 1600
Computer : Computer Contribution 4, State Geological
Survey, University of Kansas, Lawrence, Kansas.

Dedman, E.V., J.P.Lindsey, and M.W. Schramm,Jr., 1975, Strati-
graphic Modeling a Step Beyond Bright Spot : World Oil,
v.180, p.61-65.

Dixon, W.J., and F.J.Massey, Jr., 1969, Introduction to
Statistical Analysis : McGraw-Hill Book Co., New York,
638 p.

Dobrin, M.B., 1976, Introduction to Geophysical Prospecting,
3rd ed. : McGraw-Hill Book Co., New York, 630 p.

----------, 1977, Seismic Exploration for Stratigraphic Traps:
(In AAPG Memoir no.26, Ed.C.E.Payton), 516 p.

Fisher, R.A., 1936, The use of Multiple Measurements in
Taxonomic Problems : Ann. Eugenics, v.7,p.179-188
(In Atchley, W.R., and E.H. Bryant (Ed.), Multivariate
Statistical Methods, 1977, Dowen, Hutchinson and Ross,
Inc., Stroudsberg, Pensylvania, 464 p.).

----------, and F.Yates, 1963, Statistical Tables for Biologi-
cal,Agricultural, and Medical Research, 6th edition :
Hafner, New York, 145 p.

Galloway, W.R., M.S.Yancey, and A.P.Wipple, 1977, Seismic
Stratigraphic Model of Depositional Platform Margin,
Eastern Anadarko Basin, Oklahoma : In AAPG Memoir
no.26, Ed. C.E.Payton), 516 p.

Gir, R., 1974, Generation of Synthetic Seismograms to Study
Lithological Variations : M.Tech.Dissertation, Univer-
sity of Roorkee, Roorkee, 64 p.

Harbaugh, J., and G.B.Carter, 1970, Computer Simulation in
Geology : Wiley Interscience, New York, 575 p.

Harms, J.C., and P.Tackenberg, 1972, Seismic Signatures of
Sedimentation Models : Geophysics, v.37, p.45-58.

Hilterman, F.J., 1970, Three Dimensional Seismic Modeling :
Geophysics, v.35, p.1020-1037.

Kanasewich, E.R., 1975, Time Sequence Analysis in Geophysics :
        The University of Alberta Press, Edmonton, Canada,
        364 p.

Khattri, K., and R.Gir, 1975, Reflection Structure : A Seismic
        Guide to Stratigraphic Exploration : The Oil and Gas
        Journal, v.73, p.89-93.

----------, and R.Gir, 1976, A study of the Seismic Signatures
        of Sedimentation Models using Synthetic Seismograms :
        Geophysical Prospecting, v.24, p.454-477.

----------, R.Mithal, and V.Gaur, 1979, Pattern Space of
        Seismic Anomalies Associated with Hydrocarbon Deposits:
        Geophysical Prospecting, v.27, p.339-359.

----------, A.Sinvhal, and A.K.Awasthi, 1979, Seismic Discrimi-
        nants of Stratigraphy derived from Monte Carlo Simula-
        tion of Sedimentary Formations : Geophysical Prospect-
        ing, v.27, p.168-195.

----------, V.Gaur, R.Mithal, and A.K.Tandon, 1978, Seismogram
        synthesis in a multilayered dissipative media : Geo-
        exploration, v.16, p.185-201.

Krumbein, W.C., 1967, Fortran IV Computer Programs for Markov
        Chain Experiments in Geology : Computer Contributions
        Geol. Surv. Kansas, no.13, 38 p.

----------, 1968, Fortran IV Computer Porgramme for Simulation
        of Transgression and Regression with Continuous Time
        Markov Models : Computer Contributions Geol. Surv.
        Kansas, no.26, 38 p.

----------, and F.A.Graybill, 1965, Introduction to Statistical Models in Geology : McGraw-Hill, Inc., New York, 475 p.

Kumar, B.V.K.V., and S.K.Mullick, 1979, Power Spectrum Estimation using Maximum Entropy Method : Jour. Inst. Elect. and Telecom. Engrs., v.25, p.181-194.

Levinson, N., 1949, The Wiener RMS Error Criterion in Filter Design and Prediction, Appendix B : (In Extrapolation, Interpolation and Smoothing of Stationary Time Series by N.Wiener : Technology Press), p.129-148.

Lindseth, R.O., 1979, Synthetic Sonic Logs - a process for Stratigraphic Interpretation : Geophysics, v.44, p.3-26.

Lyons, P.L., 1968, The Potential and the Challenge - Stratigraphic Traps : Unpublished paper presented at SEG 38th Ann. Mtg. Denver, Colorado.

----------, and M.B.Dobrin, 1972, Seismic Exploration for Stratigraphic Traps : AAPG Memoir no.16, Ed.R.E.King, 687 p.

Marr, J.D., 1971, Seismic Stratigraphic Exploration III, Geophysics, v.36, p.676-689.

Mathieu, P.C., and G.W.Rice, 1969, Multivariate Analysis used in the Detection of Stratigraphic Anomalies from Seismic Data : Geophysics, v.34, p.507-515.

Meckel, Jr., L.D., and A.K.Nath, 1977, Geologic Considerations for Stratigraphic Modeling and Interpretation : (In AAPG Memoir no.26, Ed.C.E.Payton), 516 p.

Middleton, G.V., 1962, A multivariate statistical technique
        applied to the study of Sandstone composition : Trans.
        Royal Soc. Canada, v.56, ser.3, sec.3, p.119-126.

Miller, R.L., and J.S.Kahn, 1962, Statistical Analysis in the
        Geological Sciences : John Wiley and Sons, Inc., New
        York, 483 p.

Nath, A.K., 1975, Reflection Amplitudes Modeling can help locate
        Michigan Reefs : Oil and Gas Journal, v.73, p.180-182.

Neidell, N.S., 1975, What are the limits in specifying Seismic
        Models ? : Oil and Gas Journal, v.73, p.144-147.

-----------, and E.Poggigliolmi, 1977, Stratigraphic Modeling
        and Interpretation - Geophysical Principles and Techni-
        ques : (In AAPG Memoir no.26, Ed.C.E.Payton),516 p.

O'Doherty, R.E., and N.A.Anstey, 1971, Reflection on amplitudes:
        Geophysical Prospecting, v.19, p.430-458.

O.N.G.Commission, Unpublished Reports.

Pan, P.H., and J.C.D. De Bremaecker, 1970, Direct location of
        Oil and Gas by the Seismic Reflection Method : Geophysi-
        cal Prospecting, v.18, p.712-727.

Panter, P.F., 1965, Modulation, Noise, and Spectral Analysis :
        McGraw-Hill Book Company, p.759.

Peterson, R.A., W.R. Fillippone, and F.B.Coker, 1955, The
        Synthesis of Seismograms from well Log Data : Geophysics,
        v. 20, p.516-538.

Potter, P.E., and R.F. Blackely, 1967, Generation of a
    Synthetic Vertical Profile of a Fluvial Sandstone
    body : Journal of the Society of Petroleum Engineers
    of AIME, p.243-251.

Ricker, N.H., 1953, Wavelet Contraction, Wavelet Expansion
    and the Control of Seismic Resolution : Geophysics,
    v.18, p.769-792.

----------, 1978, Transient Waves in Visco Elastic Media :
    Elsevier Scientific Publishing Co., Amesterdam, 278 p.

Robinson, E.A., 1967, Statistical Communication and Detection
    with Special Reference to Digital Data Processing of
    Radar and Seismic Signals : Hafner Publishing Co.,
    New York, 362 p.

Sangree, J.B., and J.M.Widmier, 1979, Interpretation of
    Depositional Facies from Seismic Data : Geophysics,
    v.44, p.131-160.

Savit, C.H., and E.J.Matekar, 1971, From 'Where' to 'What' :
    Proceedings of the 8th World Petroleum Conference, 3,
    Applied Science Publishers, London.

Schneider, W.A., 1971, Developments in seismic data processing
    and analysis (1968-1970) : Geophysics, v.36, p.1043-
    1073.

Schramm, Jr., M.W., E.V.Dedman, and J.P.Lindsey, 1977, Practi-
    cal Stratigraphic Modeling and Interpretation : (In
    AAPG Memoir no.26, Ed. C.E.Payton), 516 p.

Seal, H., 1964, Multivariate Statistical Analysis for Biologists : Methuen and Co. Ltd., London, 207 p.

Sheriff, R.E., 1975, Factors affecting seismic amplitudes : Geophysical Prospecting, v.23, p.125-138.

----------, 1976, Inferring Stratigraphy from Seismic Data : AAPG Bull., v.60, p.528-542.

----------, 1977, Limitations on Resolution of Seismic Reflections and Geologic Detail Derivable from Them : (In AAPG Memoir no.26, Ed.C.E.Payton), 516 p.

Sieck, H.C., and G.W. Self, 1977, Analysis of High Resolution Seismic Data : (In AAPG Memoir no.26, Ed. C.E.Payton), 516 p.

Sinvhal, A., 1976, A Stochastic Study of the Properties of Synthetic Seismograms Characterising Sub-surface Lithology : M.Tech. Dissertation, University of Roorkee, Roorkee, 85 p.

----------, V.K.Gaur, K.N.Khattri, P.S.Moharir, and R.Chander, 1979, Mathematical Modeling in Exploration Geophysics : Unpublished Report No.4, Project No.ONGC/NCST/UOR-13.

Smith, M.K., 1969, Development in Seismic Processing for Geologic Interpretation : Paper presented at the 54th Annual meeting of AAPG at Dallas.

Stuart, C.J., and C.A.Caughey, 1977, Seismic Facies and Sedimentology of Terrigenous Pleistocene Deposits in Northwest and Central Gulf of Mexico : (In AAPG Memoir no.26, Ed. C.E.Payton), 516 p.

Taner, M.T., and F.Koehler, 1969, Velocity Spectra-Digital
    Computer Derivation and Applications of Velocity
    Functions : Geophysics, v.34, p.858-881.

----------, and R.E.Sheriff, 1977, Application of Amplitude,
    Frequency, and Other Attributes to Stratigraphic and
    Hydrocarbon Determination : (In AAPG Memoir no.26, Ed.
    C.E. Payton), 516 p.

----------, E.E.Cook, and N.S.Neidell, 1970, Limitations of
    the   Reflection Seismic Method : lessons from
    computer simulations : Geophysics, v.35, p.551-573.

Toman, K., 1965, Spectral Shifts of Truncated Sinusoids :
    Jour. Geophysical Res., v.70, p.1749-1750.

Treitel, S., and E.A.Robinson, 1966, Seismic Wave Propagation
    in Layered Media in terms of Communication Theory :
    Geophysics, v.31, p.17-32.

Ulrych, T.J., 1972, Maximum Entropy Power Spectrum of Truncat-
    ed Sinusoids : Jour. Geophysical Res., v.77, p.1396-
    1400.

----------, and T.N.Bishop, 1975, Maximum Entropy Spectral
    Analysis and Auto Regressive Decomposition : Rev.
    Geophysics and Space Physics, v.13, p.183-200.

----------, D.E.Smylie, O.G.Jensen, and G.K.C.Clarke, 1973,
    Predictive Filtering and Smoothing of Short Records
    using Maximum Entropy : Jour.Geophysical Res., v.78,
    p.4959-4964.

Vail, P.R., R.M.Mitchum, Jr., R.G. Todd, J.M.Widmier,
    S.Thompson, III, J.B.Sangree, J.N.Bubb, and W.G.Hatlelid,
    1977, Seismic Stratigraphy and Global Changes of Sea
    Level : (In AAPG Memoir no.26, Ed.C.E.Payton), 516 p.

Vistelius, A.B., 1967, Studies in Mathematical Geology :
    Consultants Bureau, New York, (Translated from Russian,
    Pt. IV), 294 p.

Waters, K.H., and G.W.Rice, 1975, Some Statistical and Probabi-
    listic Techniques to Optimize the Search for Strati-
    graphic Traps on Seismic Data : Proceedings of the 9th
    World Petroleum Congress, Tokyo, Japan, panel discuss-
    ion 9, paper 1.

Widess, M.B., 1973, How Thin is a Thin Bed ? : Geophysics, v.38,
    v.38, p.1176-1180.

Wiemer, R.J., and T.L.Davis, 1977, Stratigraphic and Seismic
    Evidence for Late Cretaceous Growth Faulting, Denver
    Basin, Colorado : (In AAPG Memoir no.26, Ed.C.E.Payton),
    516 p.

Wuenschel, P.C., 1960, Seismogram Synthesis including Multiples
    and Transmission Coefficients : Geophysics, v.25,
    p.106-129.