N69-32541

HSER 5006

CASE FILE
COPY

Analysis of Patterns
of n-Alkane Distributions

by

DIAN R. HITCHCOCK
BIOMEDICAL SYSTEMS DEPARTMENT

FINAL REPORT
PHASE IV NASw871

STATISTICAL DECISION PROBLEMS
IN LARGE SCALE BIOLOGICAL EXPERIMENTS
TO
OFFICE OF SPACE SCIENCES
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D. C.

JULY 15, 1969

Hamilton
Standard DIVISION OF UNITED AIRCRAFT CORPORATION U
A®

ANALYSIS OF PATTERNS OF N-ALKANE DISTRIBUTIONS

by

Dian R. Hitchcock

Final Report Phase IV
NASw 871

STATISTICAL DECISION PROBLEMS IN LARGE SCALE
BIOLOGICAL EXPERIMENTS

TO

Office of Space Sciences
National Aeronautics and Space Administration
Washington, D.C.

July 15, 1969

Submitted by _____
Dian R. Hitchcock, Principal Investigator

Biomedical Systems Department
Hamilton Standard
Division of United Aircraft Corporation
Farmington, Connecticut 06032

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

INTRODUCTION

At the time this project was initiated in 1965, it was part of a general inquiry regarding the applicability of concepts of information content to the problem of distinguishing material of biotic origin from material of non-biotic origin. The basic assumptions were and are that the fundamental distinction between biological and non-biological materials is that the former differ quantitatively from the latter by representing local accumulations of "negative entropy" whose occurrence is highly unlikely in an enviornment not infected with living systems, e.g., on the basis of physical principles. In particular, living systems establish and maintain a chemical free energy gradient and utilize the energy store thus established to power biotic processes. After death, the chemical free energy gradient is gradually dissipated by the action of non-living environmental processes and through the action of biological agents, until the material once comprising the living system is no longer recognizable as an area of relatively low entropy. (15)

Implicit in this position is the hypothesis that planetary biologies are self-organizing systems and the course of evolution is characterized by an increase in the order or complexity of the total planetary biological system. Such increase in complexity is also paralleled by an evolutionary increase in complexity of individual biological units such as individual organisms, or communities of interacting organisms, but whether contemporary examples of units of phylogentically ancient origin are significantly less ordered than contemporary examples of phylogentically more recent units remains to be determined.

It is in keeping with these assumptions to suppose that the negative entropy characteristics of living systems are expressed in a variety of ways in addition to the observable chemical free energy gradient represented by the

fact that living systems are composed of reduced carbon compounds with a potential to react with atmospheric oxygen.

For example, the highly organized morphological properties of living systems correspond to intricate visual patterns which differ greatly from random visual noise and from the visual patterns of geological structures, cloud formations, etc. An examination of the chemical substances making up terrestrial living systems reveals that they represent a very small subset of the possible macro-molecules which could be assembled from the same elemental components. The very sensitively balanced ecological relationships between members of a community of living systems similarly reveal a high degree of order.

It is tempting to identify these and many other widely recognized aspects of biological specificity and control as examples of the high "information content" of biological systems.

There are, however, a number of very serious problems which must be solved before one can equate apparent order and complexity with negative entropy in the information theoretic or thermodynamic sense of this term. First one needs to define the kind of "order" in question very carefully and to specify in detail what observable features of biological systems are hypothesized to reflect this order. Secondly, there must be a procedure for objectively and reliably measuring the degree of order present in the biological systems in question. (Preferably the measurement techniques employed should be analogous to techniques used to estimate entropy as that term is more rigorously employed in other disciplines.)

If these requirements can be satisfied, one will have defined a concept of biological order in operational terms, but the utility of this concept will still have to be verified by demonstrating a relationship between biological order and other properties of interest.

We have postulated that living systems represent areas of low entropy (in the thermodynamic or information theoretic sense of that term) and that after death, the order stored in the living system is gradually dissipated by living and non-living environmental processes until the once living material is no longer recognizably an area of low entropy. (15)

If this is the case, then it should be possible to observe a measurable decline in order associated with the biological and non-biological degradation of once living material. This clearly occurs when biological agents - consumers and decomposers - utilize the products of photosynthesis, increasing the entropy of these food materials in the process of extracting available free energy from them. Energy content is a useful index to order in examining this kind of degradation. It is clear however that a criterion defined in terms of free energy does not serve to aid in distinguishing biological processes as an origin of order from non-biological processes. Free energy gradients certainly occur on planets devoid of life; furthermore, the fact that carbonaceous chondrites contain hydrocarbons with a potential to undergo energy releasing reactions with the oxygen of the terrestrial atmosphere does not support the conclusion that they are the products of an extraterrestrial biology.

On earth the synthesis of complex hydrocarbons from very simple precursors occurs routinely in laboratories. Although these materials are biological artifacts in the sense that they are the products of advanced human technology, they may still be characterized as materials on non-biological origin, and their existence further emphasizes the inadequacy of the free energy criterion as an index to biological origin and degree of degradation.

The studies reported here constitute an attempt to define a criterion of order which would serve as the basis for the following:
    a.   Classifying hydrocarbons of biological origin in terms of known
          or presumed degree of degradation by non-biological and biological
          agents in the terrestrial environment.

3

b. Distinguishing between terrestrial hydrocarbons of recent
   biological origin and similar hydrocarbons synthesized in a
   laboratory.

c. Testing the hypothesis that biological order varies as a function
   of phylogenetic age.

The distributions of relative abundances of n-alkanes in materials of
biological origin have been observed by many workers to exhibit characteristic
differences in shape that are correlated with known or presumed
origin or with estimated age. (1) (9) (10) (11). Many have emphasized
the fact that many recent hydrocarbons show a pronounced predominance of
high molecular weight alkanes with odd numbers of carbon atoms over alkanes
with even numbers of carbon atoms. A more careful examination of representative
n-alkane distributions reveals that many materials of recent origin show a
spiky uneven pattern throughout their whole range, whereas distributions of
n-alkanes from ancient crude oils and shale oils tend to be unimodal, and
relatively smooth. (Figure 1)

It seemed to us that in a naive but very fundamental sense, distributions
of relative abundances of n-alkanes of recent origin can be termed more
complex than those of ancient origin, and that this visual complexity appears
to decline with increasing age. If this is the case, and if these differences
in complexity in fact reflect differences in the "information content" or
biological order of the corresponding materials, two hypotheses regarding
the cause of this decline in order with increasing age are possible:

a. The order incorporated in living materials is degraded by
   biological and non-biological agents in the local environment.

b. The order in living materials increased with evolutionary time
   so that at the time of deposition the most ancient materials
   initially reflected a lower degree of order than less ancient
   ones.

Both mechanisms would act in the same direction with respect to time. If we
suppose that complexity of individual organisms (if it has increased at all)
has probably changed only slightly since Cambrian times, such differences

may well have been obliterated by the effects of degradation.

Reasoning that the effects on order of both of these postulated mechanisms are approximately indexed by the age of the once living material -- at least for materials more than 10 million years old - we concluded that the "order" we postulated might be indexed by the visual complexity of the n-alkane distribution, and that if we could devise a suitable measure of unevenness it might serve as a basis for the desired classifications.

The results of these studies may be summarized as follows:

I.   A simple measure of visual complexity was developed and shown to correspond well to (a) subjectively perceived unevenness of a plot of the relative abundance of n-alkanes and to (b) differences in unevenness which are difficult to discriminate visually. This measure is independent of the range of the molecular weights of the n-alkanes and permits the comparison of distributions containing different numbers of components.

II. Complexity so defined correlates very strongly with estimated age in a sample of hydrocarbons of diverse geographical and geological origin varying in age from 50 to 2700 million years.

III. Seventy-two n-alkane distributions of recent origin from a diversity of materials, including carbonaceous chondrites, land plants and animals, bacteria, marine algae, and kelp and two samples of synthetic n-alkanes produced by the Fischer-Tropsch method, were examined and their complexity compared. All these distributions were taken from the literature and consequently the observations are somewhat difficult to interpret because we cannot assume consistent controls for error.   The observations support the following hypotheses:

a.   Complexity varies in the expected direction with phylogenetic age in "fresh" biological materials, but these results could also be explained by contamination of the sample materials.

b.   Passage through the digestive track of animals reduces the complexity

5
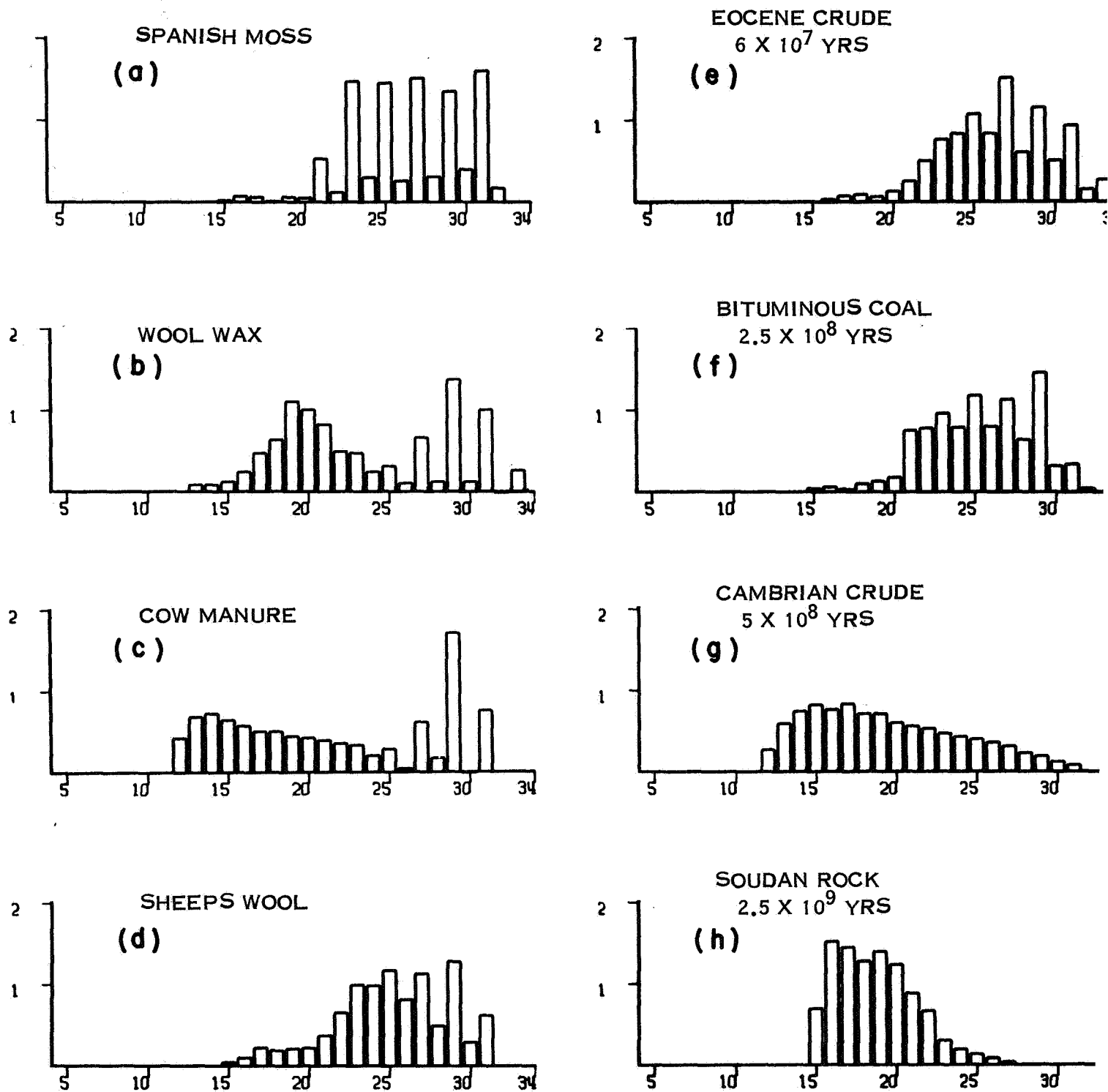
Figure 1. Representative distributions of n-alkanes from modern and ancient biological materials.

of plant materials, and biological degradation is in this respect analogous to the non-biological degradation effected by low-temperature pyrolysis.

c. Synthetic hydrocarbons are much less complex than biological hydrocarbons.

d. The carbonaceous chondrites differ significantly in complexity from synthetic hydrocarbons, and from modern biological hydrocarbons, but do not differ from ancient geological terrestrial hydrocarbons. These observations could have arisen if synthetically produced hydrocarbons were contaminated with terrestrial biological material.

e. N-alkanes from terrestrial geological deposits range in complexity from values as high as those of some modern land plant materials in young deposits to low values as low as synthetic hydrocarbons in very ancient deposits. If environmental degradation processes are largely responsible for this smoothing, then we may conclude that these processes can alter biological materials until they are indistinguishable in complexity from materials of non-biological origin.

IV. Simulated random degradation experiments were conducted to explore the effects on pattern complexity of random fragmentation of carbon-carbon bonds in n-alkanes, under the assumption that all such fragments are transformed into n-alkanes. The different models studied varied in the following respects:

A. Given that a molecule experienced fragmentation, the position of the break was equally likely for all c-c bonds in each molecule, or was Gaussian normal with the mean at the middle of the molecule, and standard distribution equal to 1/6 its length.

B. The initial conditions of the experiment define the number of molecules which experience fragmentation. This may decline exponentially with time, in which case the initial conditions may be thought of as defining a liklihood of fragmentation for each molecular species present in the starting mixture or produced

during the course of the simulated experiment (so that the production of fragments varies as e^{-kt}). Alternatively the initial conditions may be thought of as specifying some constant number of fragmentations per unit time, so that the number of fragmentations is a linear function of time.

C.   The relative fragility of molecules is either independent of their molecular weight, or a linear function of it so that longer molecules are proportionately more fragile.

D.   The reaction vessel may be thought of as porous and thereby permitting the migration of some molecules at a rate proportional to their relative vapor pressures during the degradation process. Alternatively the vessel may be thought of as completely closed, so that no fragments escape.

Using these various programmed options, we simulated the degradation on n-alkane mixtures analogous to modern materials, and drew the following conclusions:

1.   All degradation models resulted in an eventual decline in measured complexity.

2.   Patterns analogous to those of ancient hydrocarbons can be produced only if low molecular weight components are removed at rates proportional to their vapor pressures.

3.   When combined with migration all degradation models transformed a highly complex mixture of n-alkanes to mixtures similar in shape to those found in ancient terrestrial deposits. The models incorporating the linear fracture patterns were capable of mimicing virtually all the shapes observed in the series of n-alkanes from ancient hydrocarbons. Models incorporating Gaussian fracture patterns did not succeed in mimicing as many.

4.   A comparison of the results of simulated degradation of n-C_{28} with those observed empirically suggests that while none of the

7

modeled experiments exactly reproduces reported laboratory results, these are better approximated by the linear fracture pattern combined with proportional fragility and evaporation of low molecular weight components than by any of the other models.

5. Although all the degradation models incorporating migration and non-Gaussian fracture patterns produced sequences of n-alkane patterns which are visually similar to those observed in naturally occurring terrestrial deposits, the measured change in complexity with simulated time does not correspond to that observed in the former as a function of real time. Specifically, if total fragmentation is a linear function of time, the complexity measure is erratic and declines very rapidly. If fragmentation varies with e^{-kt}, and if fragility increases with the length of the molecule, then the plot of log complexity against modeled time resembles that of log complexity against log time in the set of age dated naturally occurring hydrocarbons.

We may speculate that if actual degradation (e.g., random fragmentation of n-alkanes accompanied by the transformation of some or all fragments to new n-alkanes) resembles modeled degradation then one of the following can be the case:

a. If the deposited material was approximately equal in complexity initially, (so that degradation is responsible for the observed linear correlation of complexity with age) then degradation (cleavage) rate must (1) have been relatively independent of local environmental history and (2) have declined exponentially with time by about 2 orders of magnitude, during the interval from 50 to 2700 million years before present.

b. If degradation rate is sharply dependent on thermal and/or radiation properties of the local geological environment, then the degradation can have played only a very minor role in transforming the shapes of the deposited mixtures. We must then conclude from the evidence of the linear correlation between age and complexity that mean complexity of deposited materials increased linearly with time during the first 2 billion years of life on earth.

8

Part I of the following describes the measure of complexity developed and its validation.

Part II presents the results of studies of the relationship between age and complexity.

Part III reports the studies of complexity in modern biological materials, synthetic n-alkanes, carbonaceous chondrites and the biological and thermal degradation of n-alkanes from modern biological materials.

Part IV reports the results of the computer simulation of alkane degradation.

PART I MEASURES OF PATTERN COMPLEXITY

A. POWER DENSITY FUNCTIONS AND DERIVED MEASURES

The measure employed to obtain a description of the complexity of n-alkane distributions is a function of the relative power density function of the envelope of the distribution of n-alkanes represented as percentage chromatograph peak height or peak area, or as percent by weight of the total n-alkanes.

Computing the Power Density Spectrum

We assume that the envelope of the n-alkane distribution is a one-period segment of a continuous signal, and that the observed peak heights are digitized points obtained by sampling this signal at N equally spaced intervals, where N is the number of different n-alkanes in the mixture analyzed *.

Thus, if C_{20} is the lowest molecular weight alkane and $n-C_{31}$ the highest, then n = 12. The total distribution is assumed to represent one period and consequently the sampling rate in units of angular frequency is $360^{\circ}/N$ where N is even, or $360^{\circ}/N-1$ where N is odd.

We first transform this signal by subtracting the mean amplitude of all the N components from each of them to obtain a representation in terms of positive and negative deviations from this mean. Let $y_0, y_1, \ldots, y_{N-1}$ be the peak heights of the original signal. Then $y'_0, y'_1, \ldots, y'_{N-1}$ is the transformed representation where

$$y'_j = y_j - \frac{1}{N} \sum_{i=1}^{N} y_i \qquad (1.1)$$

* The validity of this assumption for the purposes of obtaining a unique description of n-alkane distribution patterns is discussed in Note 1.

Any signal of this nature can be represented as the sum of cosine and sine functions of period $360°$, $360°/2$, .., $360°/k$ (where $k = N/2$ for N even, or $(N-1)/2$ for N odd) where each cosine and sine is multiplied by an appropriate coefficient A_i or B_i. A_i is the coefficient of the cosine component of frequency $360°/i$ and B_i the corresponding sine coefficient.

The values of these coefficients are determined from the transformed distribution as follows:

$$A_i = \frac{2}{N} \sum_{j=0}^{N-1} y'_j \cos\left(\frac{2\pi}{N} ij\right) \qquad (1.2)$$

$$B_i = \frac{2}{N} \sum_{j=0}^{N-1} y'_j \sin\left(\frac{2\pi}{N} ij\right) \qquad (1.3)$$

For N even:

$$A_k = \frac{1}{N} \sum_{j=0}^{N-1} y'_j \cos\left(\frac{2\pi}{N} ij\right) \qquad (1.4)$$

Thus, the A_1 and B_1 coefficients define cosine and sine functions whose period is the same as that of the original signal and whose amplitude is determined by the values of these coefficients. A_2 and B_2 define functions of twice this frequency, A_3 and B_3 of three times the frequency and so forth. We term the cosine and sine functions thus defined the frequency components of the signal (18).

The original signal can be reconstructed from these coefficients A_i and B_i since

$$y'_j = \sum_{i=1}^{k} A_i \cos\left(\frac{2\pi}{N} ij\right) + B_i \sin\left(\frac{2\pi}{N} ij\right) \qquad (1.5)$$

We will refer to these terms as the <u>Fourier components</u> of the signal.

The variability of a set of numbers is a function of the differences
between each value and the mean value, and is defined as the <u>variance v</u>:

$$\underline{v} = \frac{\sum (y_i - \bar{y})^2}{N} \qquad\qquad (1.6)$$

where $\bar{y}$ = mean y

In communications terminology, the variance of a periodic signal per
period is referred to as its <u>power</u> or <u>energy</u>. When the cosine and
sine coefficients are computed according to equations (1.2) through (1.4),
the total variance $\underline{v}$ is equal to one-half the sum of the squares
of these coefficients (19):

$$\underline{v} = \frac{1}{2} \sum_{i=1}^{k} (A_i^2 + B_i^2) \qquad\qquad (1.7)$$

In communications theory, it is useful to analyze signals in terms of
the amount of total signal variance or "power" contributed by components at
each observed frequency, since the ability to decode properly a signal used
for communications purposes is partly a function of the signal power at
frequencies employed for communications.

The <u>frequency spectrum</u> of a periodic signal is the set of frequency
components which may be present in it. The <u>power density function</u> is a
plot of power against frequency in this spectral interval. Since frequency
analysis is usually applied to continuous time varying signals, the convention
is to express the components of the frequency spectrum in terms of angular frequer

relative to some unit time T, which corresponds to the sampling period during which N sample points are collected. Thus, if N sample points are collected at N equal intervals during a period of duration T, the frequencies present will extend from a minimum near 0 to a maximum of k cycles per unit time T, and each frequency may be expressed in terms of its rate (in angular frequency) relative to the sampling interval N. Thus, if the sampled period is one second and the sampling interval (corresponding to the N sample points) is $\underline{1}$ seconds, then the angular frequency per sampling interval of a component of $\overline{\overline{N}}$ cycles per second will be $360^\circ j/N$. (If N is odd, the angular frequency is $360^\circ j/N-1$). It will be seen that the frequency spectrum expressed in this manner will contain frequencies only in the interval 0 to 180°, since no matter how large N becomes, $360^\circ j/N$ cannot fall outside this range, if j N/2.*

In plotting a power density function one plots the quantity $A_j{}^2 + B_j{}^2$ on the Y axis against the angular frequency $360^\circ j/N$ (or $360^\circ j/N-1$) on the X axis. Since the highest frequency is always 180°, differences in the number of n-alkanes represented in hydrocarbon samples will correspond to differences in the number of Fourier components which can be evaluated and hence to the number and spacing of points in the power density spectrum. The more n-alkanes in the hydrocarbon sample, the greater the resolution of the power density spectrum.

We assume that the hypothetical signal corresponding to the envelope of the plot of n-alkanes is a signal containing some power at every frequency in the interval from 0 to 180°.**   This assumption permits us to view the power density function as a continuous function. It follows that the power contribution computed from each of the k Fourier components corresponds to the power in each of k frequency <u>bands.</u> Thus, each point on the measured

---

*   In fact, the frequency spectrum extends from -180° to $+180^\circ$, but is symmetrical about the y axis, so we consider only the positive portion of this spectrum.

**   See Note 2 for a discussion of the validity of this assumption.
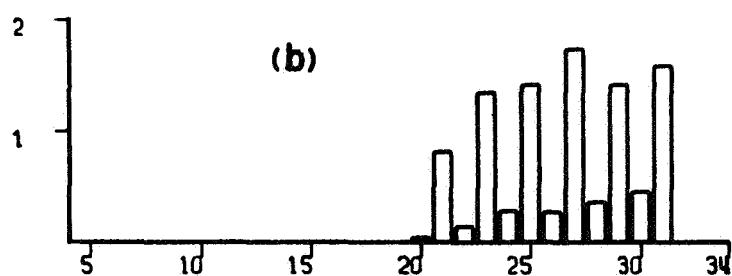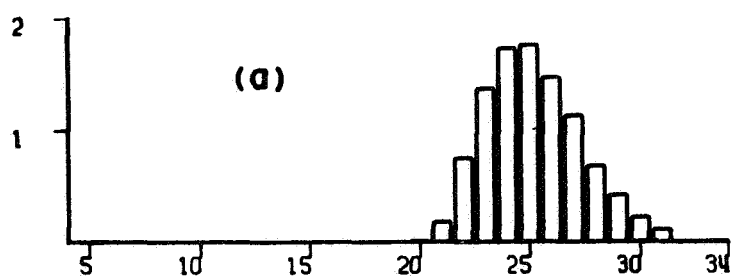
Figure 2. N-alkane distribtuions used as examples in computation of complexity measures and in mixing experiments. 2a is Shell wax; 2b is Spanish moss (only those components in range C_{20} to C_{21} are used).

power density plot is a measure of the power in a <u>band</u> of frequencies in the frequency spectrum, rather than a single frequency. For example, if K = 6, then the first point on the power density plot is 360° /12 or 30° , and the band of frequencies represented extends from 0 to 30° . The next point is plotted at 2 x 360° /12 or 60° , and this is assumed to represent the power present in the band from 30° to 60° , and so forth. Thus, each point in such a plot should be viewed as representing a band of frequencies corresponding to 1/k of the total spectrum from 0 to 180° .

Figures 2, 3, and 4 provide graphic illustrations of the manner in which cosine and sine components may be summed to reconstruct two distributions of n-alkanes.

Figure 2A is a bar plot of peak heights of n-alkanes in the range C_{20} to C_{31} extracted from a sample of commercial wax; 2B shows n-alkanes in the same range extracted from a fresh sample of Spanish moss. Figures 3 and 4 show plots of the cosine and sine components and the summed Fourier components computed from the distributions shown in 2A and 2B. The Fourier components of each of these distributions are tabulated in Table I.

The data plotted in Figure 3 are taken from the analysis of the Shell wax distribution pictured in Figure 2A. The first column of Figure 3 shows cosine and sine functions multiplied by the appropriate Fourier terms A_i and B_i. Thus, the first plot labeled C 30 in column 1 is a cosine wave whose frequency is equal to the length of the original distribution multiplied by the appropriate Fourier coefficient A_1, which in this case is -39.49, thus resulting in an inverted sine wave whose maximum amplitude is 39.49. The second plot in this column is a sine wave of the same frequency multiplied by 23.71, which is the value of B_1. The sum of these two functions is shown in the first row of column 2. In column three this sum is plotted over the original complete Shell wax distribution, to show the manner in which this original signal can be reconstructed by summing its frequency components.

TABLE IA   SHELL WAX

i	Fourier Components A_i	B_i	Power Density Function $\frac{1}{2}(A_i^2 + B_i^2)$	Relative Power $(A_i^2 + B_i^2)/\sum(A_j^2 + B_j^2)$
1	-39.49	23.71	1060.81	.96019
2	-3.60	-8.28	40.76	.03681
3	-0.2	-2.32	2.71	.00246
4	0.09	0.25	0.04	.00003
5	0.62	0.52	0.33	.00030
6	-0.5	-----	0.25	.00022

TABLE IB   SPANISH MOSS

i	Fourier Components A_i	B_i	Power Density Function $\frac{1}{2}(A_i^2 + B_i^2)$	Relative Power $(A_i^2 + B_i^2)/\sum(A_j^2 + B_j^2)$
1	-7.96	-8.57	68.40	.06593
2	-2.56	-5.24	17.01	.01636
3	-1.62	-9.0	41.81	.04025
4	-2.89	-1.69	5.60	.00540
5	3.58	-2.43	9.36	.00901
6	-29.94		896.40	.86306
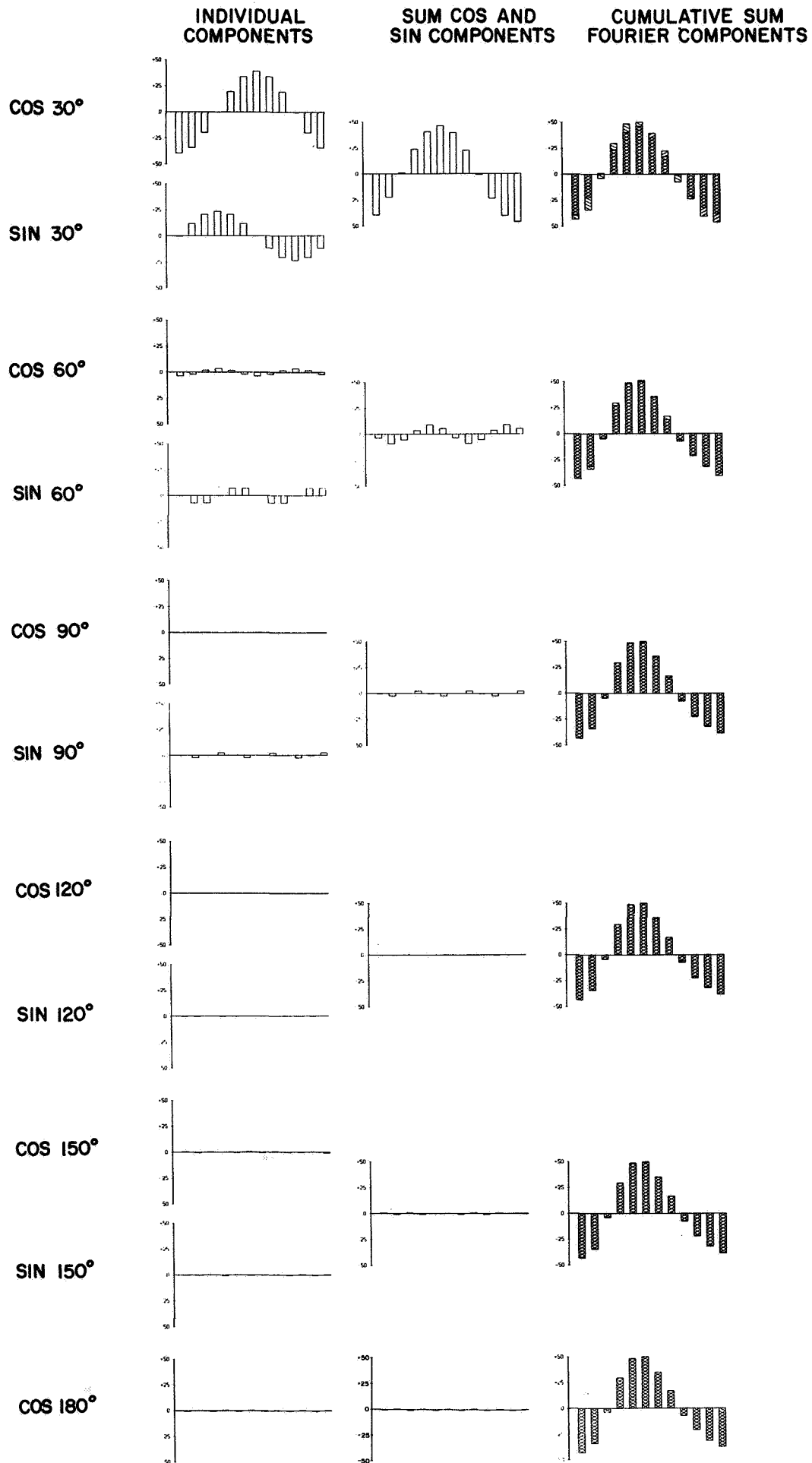
The amount of mismatch is indicated by the area of single hatching on this superimposed plot.

The third and fourth plots in column 1 show cosine and sine functions whose period is half that of the original distribution, multiplied by the appropriate Fourier coefficients $A_2 = -3.60$ and $B_2 = -8.28$. The adjacent column 2 plot is the sum of these sine and cosine functions. The column 3 plot in this line shows the cumulative sum of the first two frequency functions (e.g., row 1 and 2 in column 2) plotted against the original distribution.

The Fourier components for successive values of i are shown in the remaining rows. Sine B_6 is = 0, so no sine function is shown in row 6.

Visual comparison of the column 2 plots in various rows reveals that the lowest frequency component in row one makes the largest contribution to the original signal, which is another way of saying that most of the total signal variance is contributed by this lowest frequency component (see column ', Table 1A).

Figure 4 shows the same information for the very uneven Spanish moss distribution plotted in Figure 2B. Here the lowest five frequency components provide only a very poor approximation of the original distribution, the corresponding column 3 plots showing large areas of mismatch. Most of the signal variance is contributed by the highest frequency component.

It is clear that differences between the shapes of the two signals are reflected in the differences in the magnitudes of their cosine and sine coefficients at each frequency. In general, smooth distributions have relatively large low frequency components while the reverse is true of irregular uneven distributions. The differences in smoothness can be expressed in terms of the total amount of the variability in the original signal which is accounted for by each Fourier component.

16

SPANISH MOSS FOURIER COMPONENTS

INDIVIDUAL COMPONENTS	SUM COS AND SIN COMPONENTS	CUMULATIVE SUM FOURIER COMPONENTS

COS 30°

SIN 30°

COS 60°

SIN 60°

COS 90°

SIN 90°

COS 120°

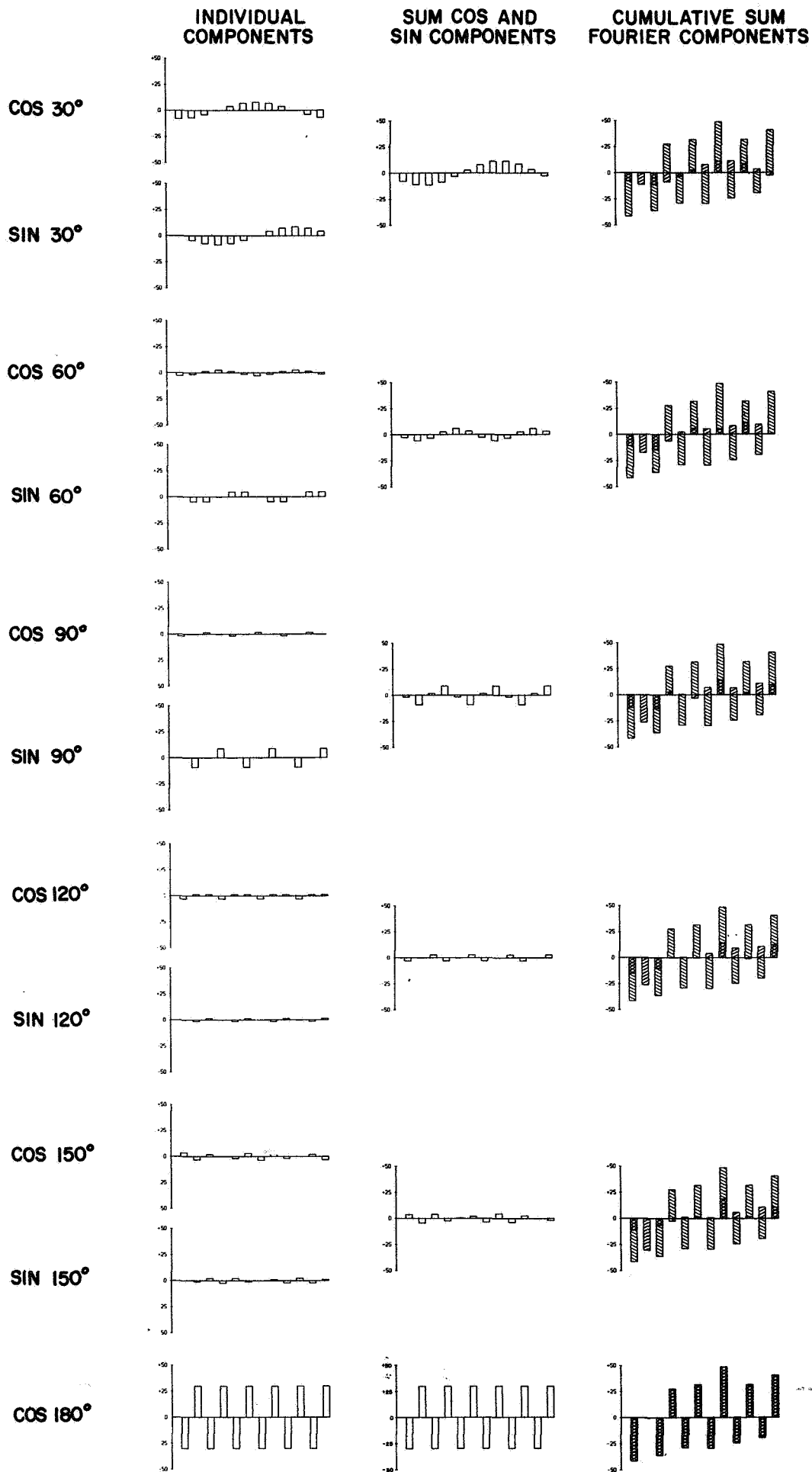SIN 120°

COS 150°

SIN 150°

COS 180°

Figure 4. Fourier components of Spanish moss distribution.

In the two examples provided above, the total spectrum is divided into six equal segments corresponding to the six Fourier components which can be evaluated when N= 12. The total spectrum from 0 to 180° is thus divided into six segments. The power density function of each of these two distributions is tabulated in Table I. The differences in smoothness between the two original distributions is clearly exhibited in their different power density functions, the smoother of the two having virtually all its power concentrated in the low frequency bands, while the uneven n-alkane distribution has its power concentrated in the highest frequency band.

Relative Power Density - The H Function

The examples cited above demonstrate that uneven curves tend to have more power in higher frequencies than smooth curves. This is always the case if the variance of the two curves is equal or nearly equal, as is the case in this example. But plots of relative amplitudes of chromatograph peaks rarely have the same variance, and two Gaussian normal plots which are equally "smooth" will differ in variance if their standard deviations differ. Thus, in order to compare different distributions with respect to unevenness, we must normalize them with respect to variance. This is achieved by expressing the variance in each frequency band as a proportion or percent of total variance, which is effected by dividing the variance contributed by each frequency component by the total variance. The result is the relative power density function $H_1, H_2, \ldots, H_k$ where

$$H_i = \frac{1}{2} (A_i^2 + B_i^2)/\underline{v} \qquad\qquad (1.8)$$

These relative power density functions are tabulated in Table I and pictured in the form of bar plots in Figure 5, where the width of the bar corresponds to the width of the frequency band.

17
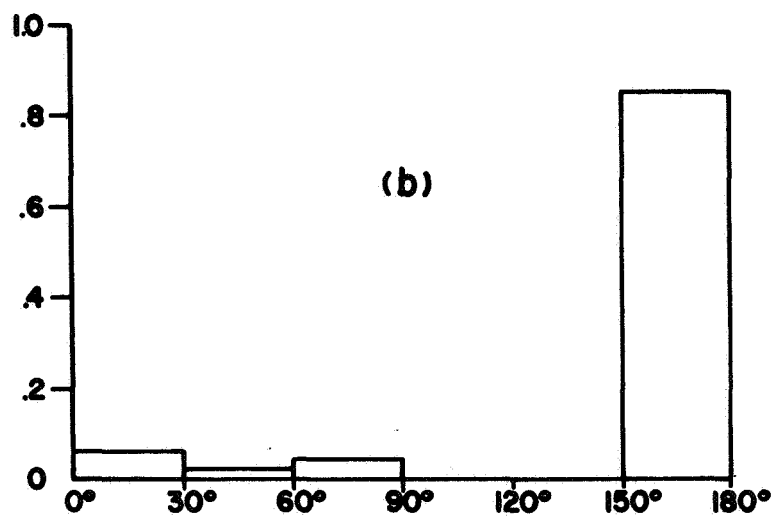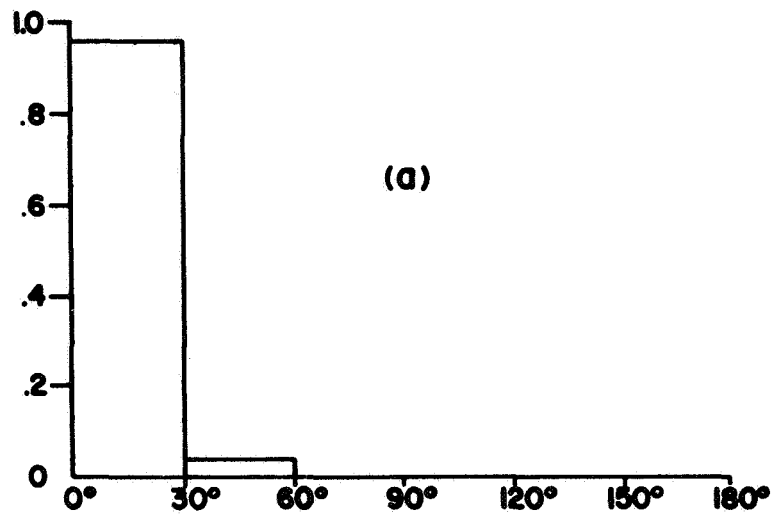
RELATIVE POWER DENSITY DISTRIBUTIONS

Figure 5. Relative power distributions of shell wax (5a) and Spanish moss (5b).

The number N of n-alkanes in hydrocarbon samples is not constant but varies
with the source of the hydrocarbons.

Because we assume that the observed n-alkane pattern corresponds to exactly
one period, the number of frequency bands will be $N/2$ or $(N-1)/2$ and hence
will vary with the number of n-alkanes present in the mixture analyzed.
Figure 6 shows plots of two n-alkane distributions containing different
numbers of n-alkanes. In order to compare n-alkane distributions such as
these we require a technique for transforming their power density distribu-
tions so that the power in corresponding segments of the frequency spectrum
can be directly compared.

Figure 7a is a plot of relative power against frequency for the distribution
containing 24 n-alkanes shown in 6b, where $k = 12$. The length of the Y
axis is the same as in Figure 5, but because there are 12 components rather
than 6, the width of the bars is correspondingly narrowed. Such a power
density function cannot be readily compared with the distributions in Figure
5 because each bar corresponds to the power in a band comprising $1/12$ of the
total frequency spectrum. However, by summing adjacent components, we can
transform this plot into a plot of power against frequency where each frequency
band consists of $1/6$ of the total spectrum. The result is shown in Figure 7b.
This transformation introduces no error since the power in the band of frequencies
from 0 to 30° is clearly eqaul to the power in the band from 0 to 15° plus that
in the band from 15° to 30° and so on.

We may extend this transformation technique to generate new plots of relative
power against $k' < k$ frequency bands by appropriately integrating the areas in
the original power density plot. Figure 7c is a plot of the relative power
density function of the n-alkane distribution shown in 6a where $n = 16$ and
$k = 8$. Figure 7d shows this replotted for $k' = 6$. Thus, in 7d the first
bar represents the sum of the power in the band from 0 to 22.5° of the original

18
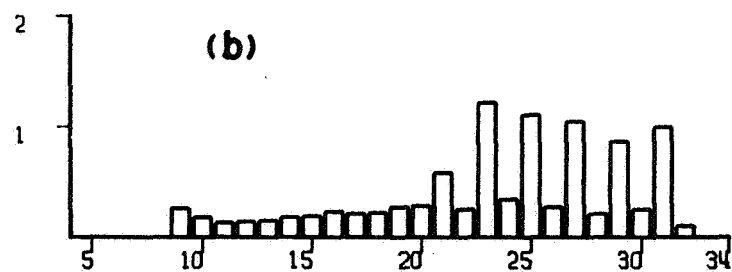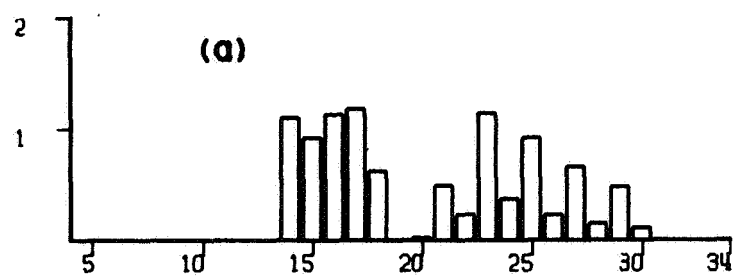
Figure 6. N-alkane distributions containing different numbers of components. 6a is a contaminated Spanish moss containing 17 components; 6b is a sample of cow manure containing 24 components.
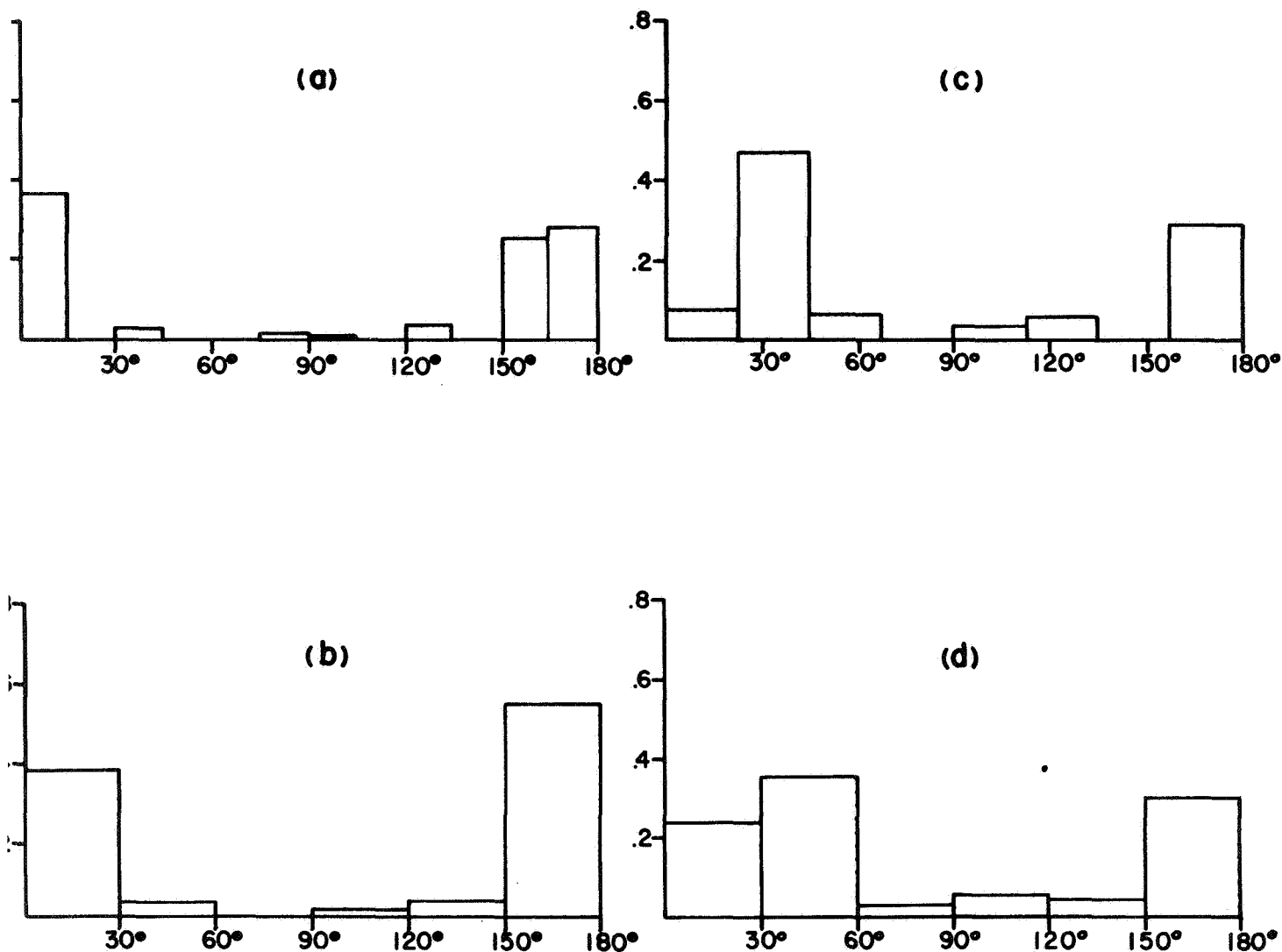
APPROXIMATE POWER DENSITY DISTRIBUTIONS

Figure 7. Bar plots of relative power density and approximate power density functions of the cow manure n-alkanes shown in Figure 6b. 7a shows the distribution of power in the 12 frequency bands estimated from the 24 components of the cow manure n-alkanes. Each band is 15° wide. 7b is an approximation of the same distribution for k' = 6, obtained by summing adjacent bands.

7c shows the 8 bands estimated from the 17 components of the Spanish moss n-alkanes. Each band is 22.5° wide. 7d is an approximation of the same distribution for k' = 6, obtained by summing corresponding segments of the frequency spectrum in 7a. Thus, the first band of 8b is equal to the content of the first band of 8a (0 to 22.5°) plus one-third the power in the second band of 7b.

power density distribution plus 1/3 the power in the band from 22.5° to 45° and so forth. While this approximation procedure always results in a loss of information (in that the original power density function cannot be reconstructed from the approximation) it can be argued that it introduces no error in that it is fundamentally equivalent to sampling some interpolated envelope at an appropriate rate. If such an envelope pattern analysis yields insights into differences between samples of hydrocarbons, then the assumptions implicit in the analysis technique will be justified.

Perceived Uneveness and Complexity

These approximation procedures permit us to contrast relative power in each of six bands of the total frequency spectrum for these four hydrocarbon samples shown in Figures 2 and 6. The more uneven a distribution is, the greater the percentage power represented in its high frequency components. This fact immediately suggests that unevenness may be objectively measured as a function of percentage power in high frequencies. Table 2 shows the relative power in each sixth of the spectrum for all four of these samples. Inspection reveals that subjectively perceived unevenness does, in fact, correlate with percent power in the high frequency bands.

Percentage power in high frequencies of the spectrum is the complexity measure we employ in the work reported here. Use of the relative power density function permits the estimation of percentage power in the top 1/k' portion of the spectrum. Table 3 shows these measures for the eight n-alkane distributions plotted in Figure 1. H'_6 is the percentage power in the top 1/6 of the spectrum, H'_8 the percentage power in the top 1/8 and so forth. The ranking imposed by these complexity measures corresponds well to that which is estimated visually. Some of these are difficult to rank on the basis of subjectively perceived unevenness: Figures 1b and 1c, 1d and 1e and 1g and 1h. These same pairs have complexity values which are close to one another. This implies that complexity measures presented here correspond well to subjective estimates, although additional factors probably also influence judegments of perceived unevenness.*

* See note 3.

19

TABLE 2

		1	2	3	4	5	6
Spanish Moss	(Figure 2B)	.066	.016	.040	.005	.009	.863
Spanish Moss	(Figure 6a)*	.375	.035	.015	.014	.026	.534
Cow Manure	(Figure 6b)*	.239	.365	.025	.046	.035	.289
Shell Wax	(Figure 2a)	.957	.040	.003	.0002	.0002	.0004

Table 2   Distribution of power in six frequency bands of materials shown in Figures 2 and 6 ranked in order of decreasing uneveness. Asterisks indicate approximate power density functions.

20

TABLE 3

Figure	H'_6	H'_8	H'_9	H'_{10}
1a	.434	.371	.350	
1b	.372	.320	.303	.290
1c	.370	.250	.210	.178
1d	.194	.175		
1e	.146	.108	.095	
1f	.110	.088	.081	
1g	.0037	.0027	.0024	
1h	.0007			

Table 3.   Measured complexity values of materials shown in
Figure 1.

B.   VALIDATION OF THE MEASURE

Discriminibility Requirement

We have demonstrated that the percentage variance accounted for by high-frequency components of an n-alkane distribution correlates well with perceived smoothness of the original bar plot distribution. But, to be useful, a measure of smoothness also ought to discriminate differences in smoothness which are so slight as to be imperceptible to the unaided eye.

We may test discriminability of the smoothness measure by generating a series of synthetic n-alkane distributions which differ in smoothness, but so slightly that the difference is not readily perceived.   Such a series can be generated by adding very small quanitites of a hydrocarbon sample which is highly uneven to a sample which is very smooth. The larger the percentage of uneven sample in the resulting mixture, the more uneven the mixture should be. Consequently, if the frequency measures do index uneveness, they should correlate positively with the percentage of uneven sample in the mixture.

Simulated Mixing Experiment

Let the percentage abundance of alkanes C_{20}, C_{21}, C_{22},..C_{31} in the very smooth shell wax distribution pictured in Figure 2a be denoted by x_0, x_1,..x_{11} and similarly let y_0, y_1,...y_{11} be the percentage abundance of the corresponding n-alkanes in the uneven Spanish moss distribution shown in Figure 2b.

We may generate a new hydrocarbon mixture by adding measured quantities of these two samples together so that the resulting mixture will contain a known proportion α of Spanish moss and 1- of shell wax. We denote the resulting mixture z_0, z_1,..z_{11} where
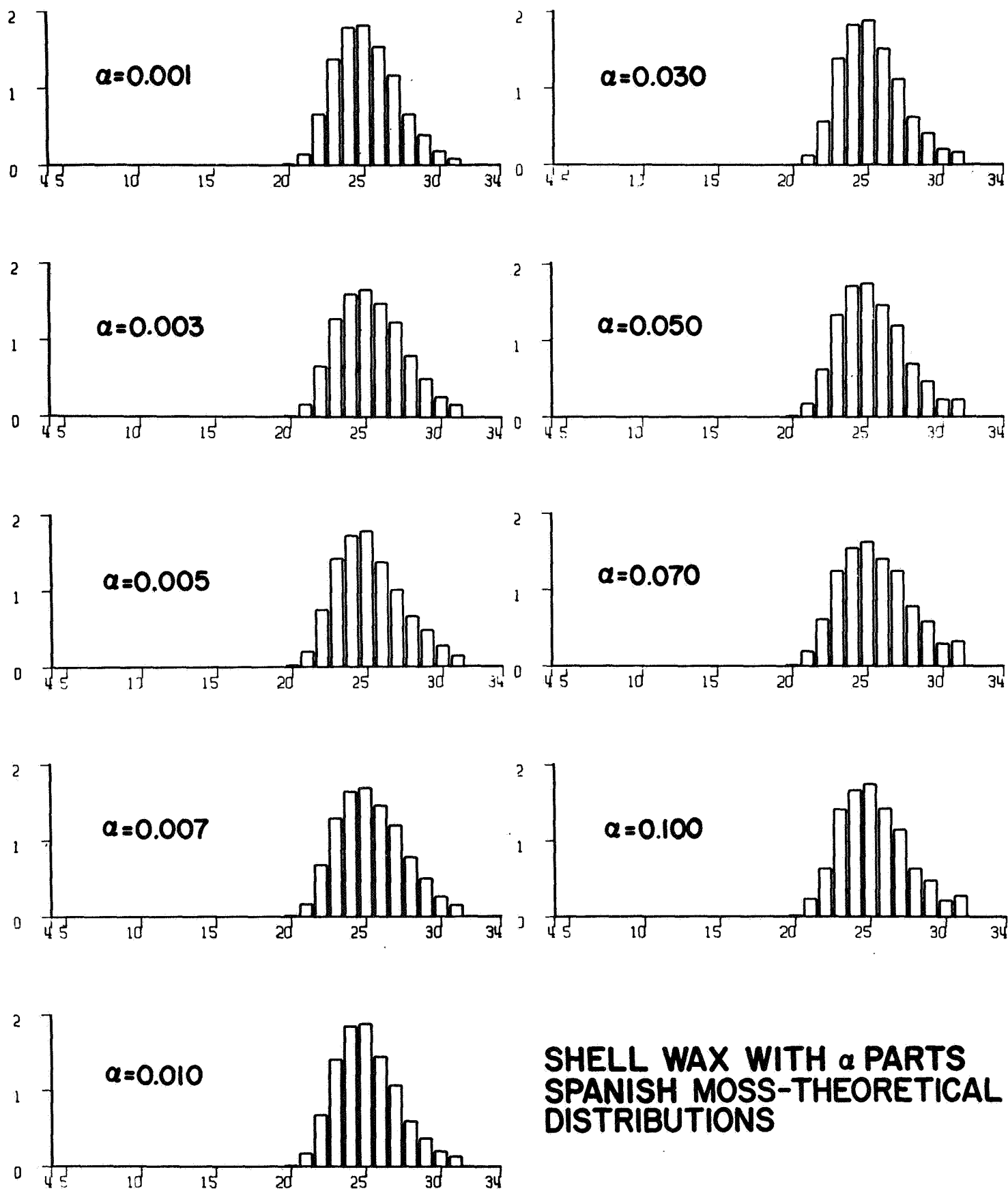
$$z_i = (1 - \alpha) x_i + \alpha y_i \qquad (1.10)$$

Figure 9. Nine theoretical mixtures of n-alkanes consisting of α parts of the Spanish moss n-alkanes shown in Figure 2B with 1-α parts of Shell wax n-alkanes shown in Fig. 2a.

Substituting known values of x_i and y_i obtained from the Shell wax and Spanish moss chromatograms in equation 1.10, we computed values of z_i for nine theoretical mixtures of the two, where α, the proportion of Spanish moss alkanes varied from 0.001 to 0.1. Figure 9 shows these mixtures. Mixtures containing less than 5% Spanish moss are visually indistinguishable from one another, and those containing more can be distinguished only with difficulty.

When these theoretical mixtures were subjected to frequency analysis, the differences in smoothness were shown to be reliably discriminated by the frequency measure: the percent of total power in the highest frequency band (the top 1/6 of the spectrum) increases with the proportion of Spanish moss n-alkanes in the theoretical mixtures.

In Figure 11, H_6 is plotted against α, and this proposed measure of uneveness is seen to correlate positively with α. (This function appears to vary exponentially, but this is not true for larger values of α.)

Physical Mixing Experiment

The simulated mixing experiment may be viewed as providing a prediction of the frequency distributions which would be observed if Spanish moss alkanes were physically mixed in appropriate proportions with the shell wax alkanes. Such a physical experiment was conducted by Peter Simmonds, who chromatographed the resulting nine mixtures. The proportion of each n-alkane in each mixture was estimated by measuring the corresponding peak height. The resulting mixtures are shown in Figure 10. Only the 7% and 10% mixtures are clearly more uneven than the others. The complexity measure H_6, the percentage power in the top 1/6 of the spectrum, is plotted against α in Figure 11. The fit is fairly good, although there are some discrepancies. In Figure 12 these observed H_6 values are plotted against the predicted values obtained in the simulated mixing experiments.
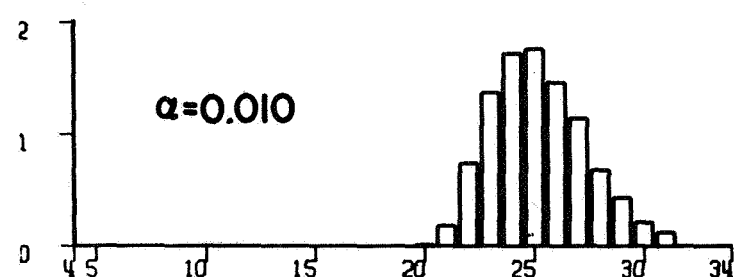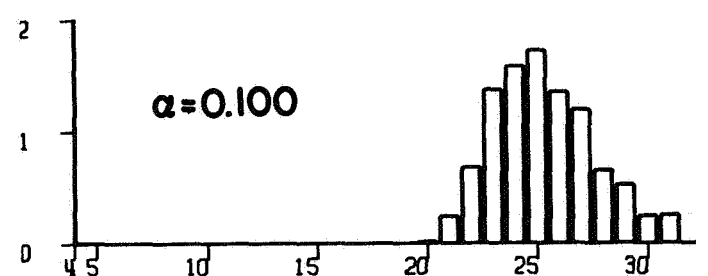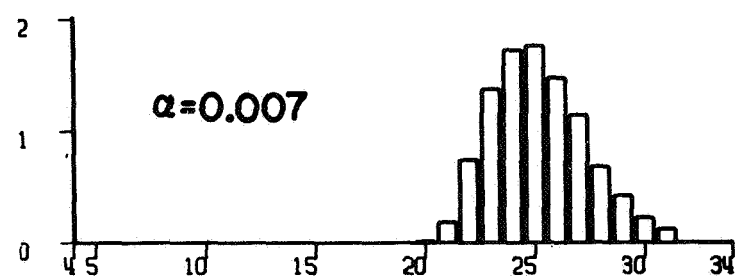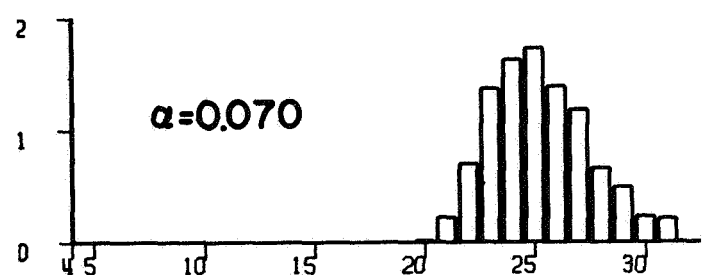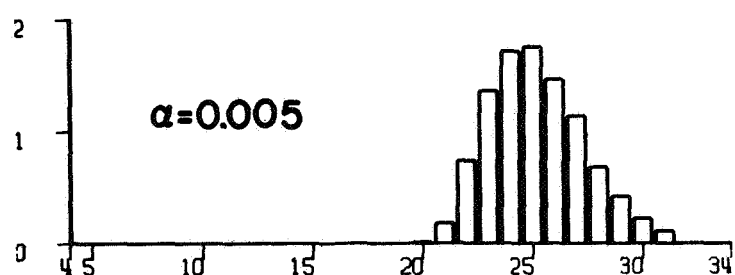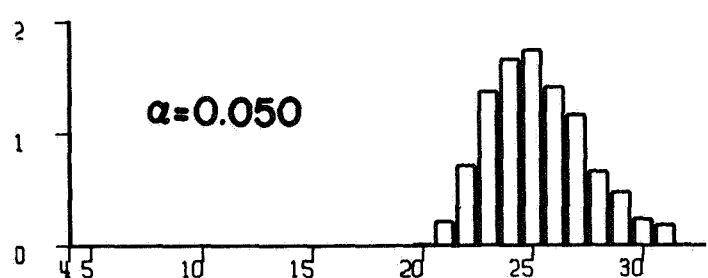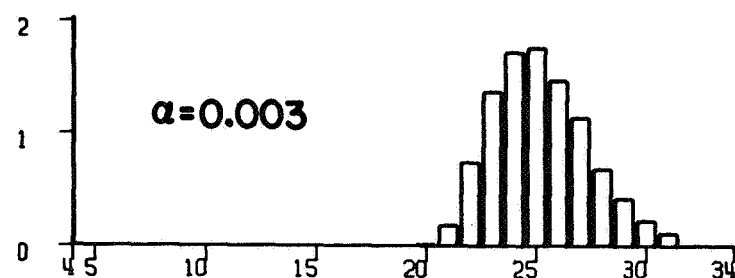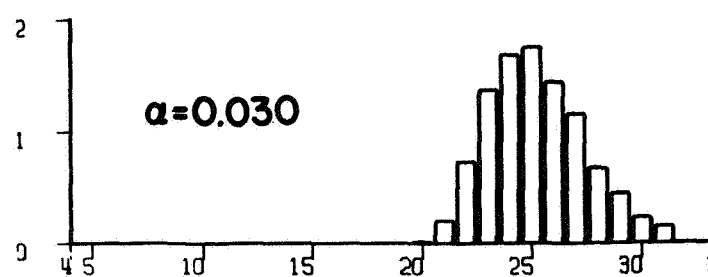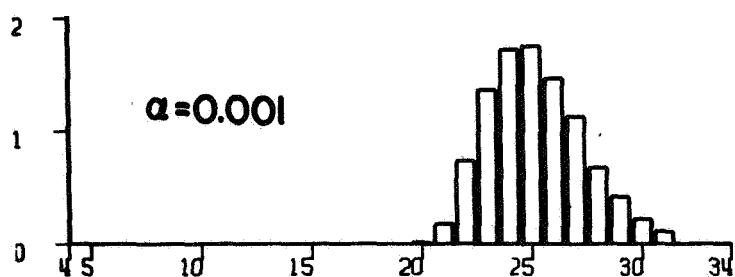
Figure 10. Distributions obtained by physically mixing appropriate quantities of Shell wax and spanish moss n-alkanes.
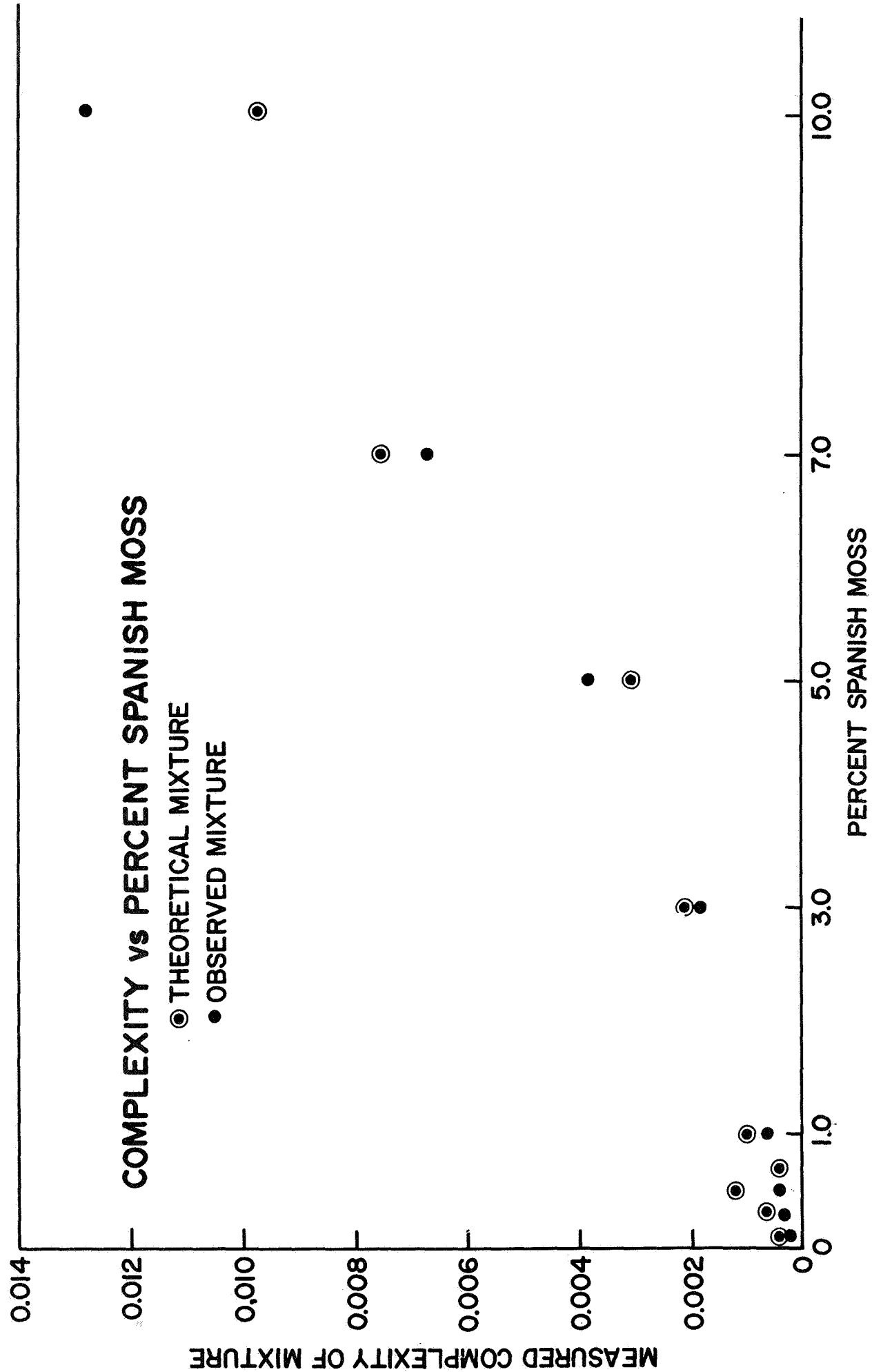
Figure 11. Plot of complexity (H_6) against percent Spanish moss for the n-alkane mixtures shown in Figures 9 and 10.

Analysis of the variance of the linear regression of the observed on the theoretical H_6 functions reveals a very good fit: $F = 142.8$ ($P_{H_0} < 3 \times 10^{-5}$ for 7 degrees of freedom). The calculated value of the intercept is $A = 5.51 \times 10^{-5} \pm .42$, which is not significant. The slope fitted with $A = 0$ is $0.848 \pm .058$. This slope, and the 95% confidence limits around the regression line are plotted in Figure 12. This slope is clearly significantly different from 1, which implies one or more of the following:

a)   An error in the prediction model, which could occur if the values of x_i and y_i substituted in expression 1.10 to obtain the predicted z_i values are wrong;

b)   A systematic error in the experimental procedure;

c)   A random error in the experimental procedure affecting one observed point which depresses the calculated value of the slope.

We are unable to decide which among these three alternatives may in fact be the case. However, if the point corresponding to $\alpha = .1$ is removed, the re-calculated value of the slope is not significantly different from unity. We are inclined to feel that this point is in error. It should be noted that the errors discussed here are, in absolute magnitude, less than 0.2% of total power. Although these may seem small according to conventional standards, such differences should not be neglected, as subsequent work reported below has shown even smaller differences to be significantly correlated with age of ancient hydrocarbon deposits.

Error Sources

It is instructive to consider the sources of errors that could lead to systematic or random deviations from the predicted values:

1.   Errors in measuring or mixing the samples. When alkanes from a smooth and an uneven sample are mixed like this, the resulting power estimate is heavily dependent on α^2, so small errors in measurement of the samples
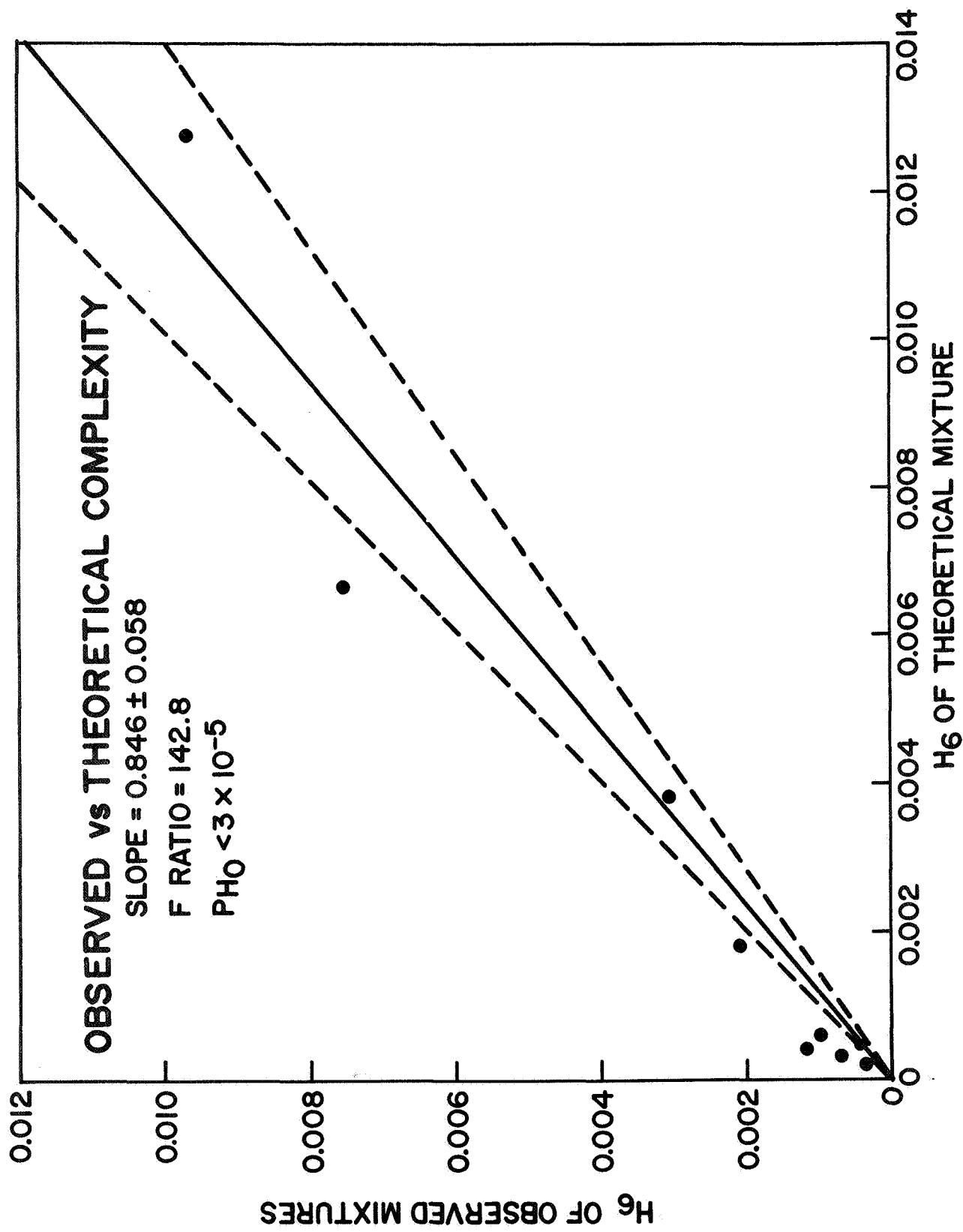
Figure 12. Observed and theoretical complexity measures. Straight line is the calculated linear regression. Dashed lines shown 95% confidence limits on the regression line. The slope would be 1 if the last point were omitted.

would have a relatively large effect on the resulting power functions.

2. Chromatograph errors. The output of the chromatograph is a peak whose area is assumed to vary linearly with the quantity of the corresponding n-alkane present in the mixture. However, in practice, detector output is not always linear, nor is the amplification of the signal from the detector. Non-linearities such as these could introduce consistent errors in the power estimate.

3. Data reduction errors. For simplicity of data reduction we have used peak amplitude rather than peak area to obtain an estimate of the quantity of each n-alkane in each mixture. Such a procedure yields correct estimates only if all peaks are of constant width. In the chromatograms used in this experiment, the peak widths of some components differed slightly from those of others in all the samples run, so some components were probably incorrectly measured on all runs. A bias in the observed power estimate could result.

The process of measuring the peak heights themselves is one that is subject to small random errors. The greatest difficulty in measuring peak height derives from the difficulty of deciding where the base line is since some base line noise is almost always present. No single rule for fixing base line location works in all situations and different rules give different results in at least some peaks. These small random errors are independent of peak height or molecular weight, and they have a negligible effect on the power estimate.

Considering the variety and ubiquity of error sources, the observed agreement between predicted and observed power estimates is excellent.

These examples demonstrate that frequency measures constitute a sensitive discriminator of pattern smoothness when the experimental data are reasonably good. However, this mixing experiment and the discussion of error sources suggest that considerable caution is required in interpreting results: contamination of samples with crude oil, insecticides, floor wax, or any other material with a distinctive n-alkane pattern will introduce a shift in the smoothness of the resulting distribution in a fashion readily discriminated by the complexity measure.

C. SUMMARY

1. The percentage of total variance accounted for by the highest frequency components of a distribution of n-alkanes is proposed as a measure of the "complexity" of the n-alkane mixture. This measure is computed in the conventional way, except that the period of the signal is assumed to be N the number of n-alkanes in the range of C_n represented in the mixture, where each n-alkane is represented as percentage abundance of the total mixture of n-alkanes. If 2 or more mixtures contain the same number of components, they may be directly compared with respect to percentage power in one or more of the high frequency bands.

2. When two or more n-alkane mixtures do not contain like numbers of components, estimates of the relative power content of like portions of the frequency spectrum can be obtained by appropriately integrating the relative power density function.

3. Complexity measures computed in this fashion are independent of the range of carbon numbers represented in the n-alkane mixture.

4. Complexity so defined corresponds well to subjective estimates of perceived unevenness and to known differences in unevenness which are difficult or impossible to discriminate visually.

26

PART II   AGE AND COMPLEXITY

Our original hypothesis was that materials of recent biological origin are characterized by a high but unknown degree of "order" and that after the death of the organism this order is degraded by environmental agents. If we equate biological order with visual complexity as measured by the frequency functions described in the previous section, the hypothesis under consideration is that the complexity of the n-alkane distributions of ancient hydrocarbons declines with their age.

A.   PROBLEMS OF EXPERIMENTAL DESIGN

The statistical and methodological problems associated with an attempt to determine whether age correlates inversely with measured complexity are numerous and should be reviewed.

A rigorous test of this hypothesis should attempt to distinguish differences in complexity which are due solely to differences in age from those which may be independent of age. We suppose that any reliable age dependent differences in complexity are attributable to one or more physical processes acting through time in the environment of the sample, or to initial differences in complexity which are a function of type of material and time of origin. Consequently, a good experimental test of the hypothesis should control for such obvious physical differences as the nature of the material (e.g., coal, crude oil, shale oil) and such differences in the geological environment as would suggest differences in the geologic history of the samples. Control could in principle be achieved either by randomizing on these factors, or by suitable replication. Variations in complexity can be caused by differences in the treatment of the sample: the extraction techniques used to isolate the n-alkanes and the chromatographic and data reduction procedures used to measure their relative abundance are possible sources of variation and these can be controlled by replication. Obviously, data from different laboratories using different analytical procedures should not be combined unless a careful collaborative program insures

appropriate statistical control of between-laboratory variations. The samples
employed should represent an appropriate distribution of ages: desirably the
age dimension should be sampled at approximately equal intervals and errors in
the age estimate should be eliminated, or be controlled (by employing similar
dating procedures and obtaining reliable estimates of the inherent dating
error).

B. PROCEDURE

Clearly these requirements for a rigorous experimental test of the hypothesis
are very difficult if not impossible to satisfy. At the time we began our
study an elaborate and expensive program of data collection could not be justified.
We chose therefore to conduct a very preliminary exploration of the relationship
between complexity and age by using n-alkane distribution data available to us in
the literature. The only criteria we employed in selecting distributions from
the literature were:

(1) An age in years could be assigned to the sample. In some cases the author
    of the source article specified an age in years; in other cases only a
    geological period was specified. In the former case we employed the
    author's estimate, in the latter, we assigned an age in years based on Kulp's
    time scale. (6)

(2) Estimates of the relative abundances of all extractable n-alkanes were
    provided. Many authors are interested only in the high molecular weight
    components conventionally employed in computing carbon preference indices,
    and so do not publish the entire distribution. Such data cannot be used
    for our purposes. Therefore, we used distributions which appeared to be
    "complete" in the sense that they declined to low values at both ends.

(3) Good peak amplitude measurements are possible. In practise this meant
    that when primary literary sources were available we made our own measure-
    ments on reproductions of the published original chromatograms, or on good
    bar plots. Most published line plots proved to be too small in scale to
    measure accurately. Three samples were obtained from Clark (3) who gives

digital values for the peak heights. It was, of course, impossible to confirm the reliability of the amplitude measurements provided by Clark, but we assume them to be adequate for our purpose.

The computed complexity measures of these distributions were then analyzed to determine whether they exhibit a significant regression on estimated age of the source material, with no attempt to control for error. This strategy may be criticized on many grounds. Not only does it represent a deliberate though common violation of cannons of experimental design mentioned above, but it introduces a vexing problem of distinguishing what we may term "duplicate" samples from "replicate" samples.

It is customary for research workers to exchange samples of hydrocarbons, and to publish n-alkane distributions which have been obtained by different laboratory procedures from the same sample. Sometimes an author will publish one such chromatogram in one paper and another in another paper. It is obvious that such distributions provide control of errors due to different sample handling procedures, but would not reflect differences due to different physical processes acting on material found in geologically diverse kinds of formations or in the same formation at geographically diverse locations or in different strata.

We may term all n-alkane distributions obtained from the same physical sample duplicates, and reserve the term replicates for samples of the same estimated age which in fact were derived from geologically or geographically diverse deposits. It is clear that the difference can be statistically important, for a large number of duplicates which yield very close complexity measures can either introduce a spurious degree of significance if they fall near a regression line, or falsely reduce the calculated significance if they result in a greatly increased scatter, since all observations are given equal weights. Replicates, if available, serve a valuable purpose even in this preliminary exploration by contributing to the variance about the regression line on which one bases an estimate of the percentage of variance which is accounted for by the independent variable--in this case, age. Unfortunately literature sources

do not always permit one to distinguish between duplicates and replicates. Consequently, in collecting our data, we attempted to make this distinction on the basis of internal evidence of the shape of the distribution itself, or on the basis of authorship, or on the basis of descriptions of the source of the material or the analytical procedures employed. The resulting samples include some distributions we judged to be duplicates and some we judged to be replicates. Bar plots of the relative abundances of n-alkanes in these samples are shown in Figure 13. The samples are:

(a)   Non-marine Eocene Shale, Welte as reported by Clark (3). We assigned the age 4.7×10^{7} years.

(b)   Green River Shale 1. Eglinton and Calvin (1) Age estimate provided by the authors.

(c)   Green River Shale 2. Calvin (1) Age estimate provided by the author.

(d)   Eocene Crude. Simmonds (13)

(e)   -(h) Shale oils from lower Cretaceous sediments. Kenvolden (7). 1 and 2 are from the Mowry formation, 3 and 4 from the Thermopoliss formation. Our assigned age of 1.2×10^{8} years is the mid-point of the lower half of the Cretaceous as given by Kulp (6).

(i)   Bituminous Coal. Simmonds. (13) Our assigned age is the midpoint of the Permian.

(j)   Antrim Shale 1. Calvin (1) Age estimated by author.

(k)   Antrim Shale 2. Eglinton and Calvin (2). Age assigned by authors. This sample and the preceding one are very similar, though a different age estimate is given by the same author. For this reason, and because the C_{12} and C_{13} measurements differed, we assumed they represent independent sampl replicates.

(l) &
(m)   Devonian Sediments 1 and 2. From Welte (4) as reported by Clark (3). Assigned age corresponds to the mean of the Devonian. We assume these to be replicates.

(n)   Cambrian crude oil, Simmonds (13). Assigned age is the end of the Cambrian.
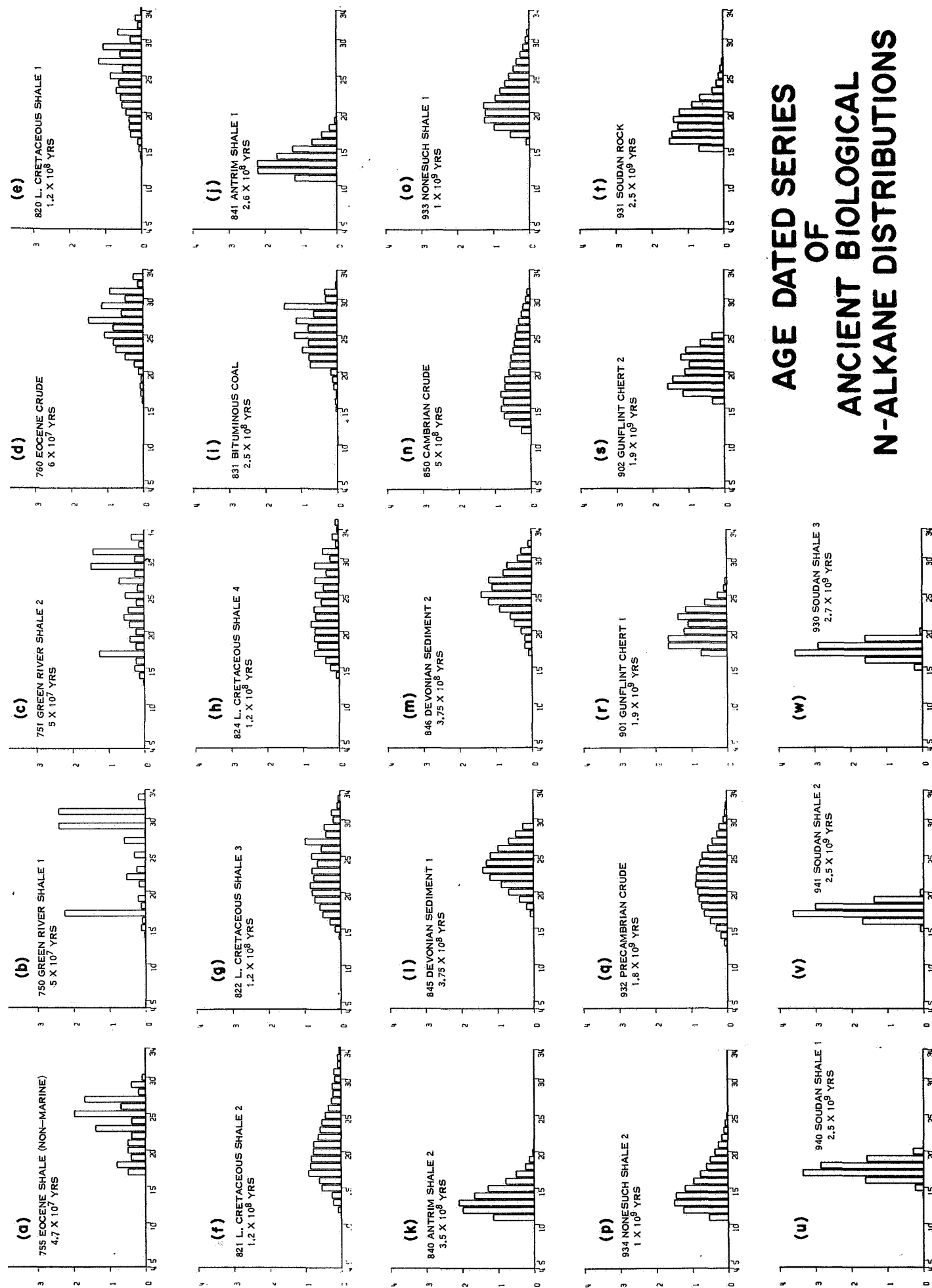
AGE DATED SERIES
OF
ANCIENT BIOLOGICAL
N-ALKANE DISTRIBUTIONS

Figure 13. Twenty-three n-alkane distributions from geological hydrocarbon deposits. Sources cited in text.

(o)   Nonsuch Shale 1. Barghoorn, Meinschein, and Schopf (8). Age assigned by authors.

(p)   Nonsuch Shale 2. Calvin (1). Both this sample and the preceding come from the same formation, but we assumed they are replicates rather than duplicates because of the difference in authorship, and the great differences in the distributions themselves.

(q)   Precambrian Crude. Simmonds (13). Our assigned age is the mean of Precambrian, assuming it began 3×10^9 years ago.

(r) & Gunflint Chert 1 and 2. Oro, et. al. (12) Age assigned by authors. These
(s)   two distributions are from the same sample and, hence, we assume them to be duplicates rather than replicates.

(t)   Soudan Rock, Oro et. al. (17) Age assigned by the authors.

(u)   Soudan Shale 1. Calvin (1), age assigned by the author.

(v)   Soudan Shale 2. Douglas and Eglinton (9). Age assigned by the authors.

(w)   Soudan Shale 3. Eglinton and Calvin. (2) Age assigned by authors. All three Soudan Shale samples have similar distributions, but they differ in detail. However, because all three are reported by essentially the same group of authors, we have assumed that 1 and 2 are duplicates because the same age assigned is assigned to both, while sample 3 is from a different and older stratum.

C.   RESULTS

The Duplicate-Replicate Problem

The seriousness of the problem represented by the difficulty of distinguishing between duplicates and replicates can be evaluated by determining whether their inclusion as separate points yields significantly different estimates of the regression line or of the variance around the line. One may test the error contribution of the duplicates and replicates by contrasting the statistical estimates obtained on three groups of data: One group includes all measurements as separate data points; one contains all replicates as separate points but carries suspected duplicates as single points corresponding to the means of the measured frequency estimates of the duplicated items; one contains only one point for each age value--duplicates or replicates of like age having been averaged together. Such tests for all of the analyses presented below revelaed no significant differences between the estimates calculated for each of the

TABLE 4:   PARTIAL REGRESSIONS OF COMPLEXITY ON AGE

A H'_3 and Age for Long and Short Distributions

Group	N	A	B	F Ratio	P_{H_0}	r^2
I	23	9.188 ± 1.586	$-1.277 \pm .184$	48.16	1.4×10^{-6}	.696
II	20	9.106 ± 1.794	$-1.272 \pm .209$	37.06	1.4×10^{-5}	.673
III	14	9.040 ± 2.064	$-1.261 \pm .240$	27.61	2.3×10^{-4}	.697

B H'_5 and Age, Long distributions only

Group	N	A	B	F Ratio	P_{H_0}	r^2
I	20	12.70 ± 1.912	$-1.727 \pm .225$	58.83	$.86 \times 10^{-6}$	.766
II	18	12.51 ± 2.23	$-1.708 \pm .262$	42.4	1.1×10^{-5}	.726
III	13	12.0 ± 2.07	$-1.645 \pm .243$	45.8	$.42 \times 10^{-4}$	.806

C H'_3 and age, Long distributions only

Group	N	A	B	F Ratio	P_{H_0}	r^2
I	20	10.96 ± 1.671	$-1.495 \pm .197$	57.73	$.97 \times 10^{-6}$	.762
II	18	11.05 ± 1.819	$-1.51 \pm .214$	49.65	$.44 \times 10^{-5}$	.756
III	13	10.79 ± 1.858	$-1.476 \pm .281$	45.95	$.41 \times 10^{-5}$	.807

three groups except for increases in the standard errors of the values of the regression coefficients, and decreases in the significance of the regression to be expected from the reduction in the degrees of freedom. (See Table 4). This means, simply, that it makes essentially no difference whether the controversial points of like age are duplicates or replicates because the scatter they introduce is approximately the same as that contributed by points which correspond to different ages. We may conclude that the results presented here are not in fact biased by the unlucky selection of duplicates.

Complexity and Age

The n-alkane distributions in Figure 13 contain from 6 to 22 components. The entire set can be compared only with respect to the percentage power in the top third of the spectrum, (the H'_3 function) since the power in narrower bands cannot be estimated for the three short distributions containing only 6 or 7 points. All the remainder contain at least 10 points, and for these it is possible to estimate the percentage power in the top 1/5 of the total spectrum (H'_5).

In examining the relationship between complexity and age, we employed both H'_3 and H'_5 as complexity measures and calculated the significance of the regression.

In determining the significance of an observed F ratio, it is conventional to distinguish only the levels provided in standard tables: namely, whether the liklihood under the null hypothesis of the observed ratio is less than .05, .01, .005 or .001. Because our data sources are so varied, and because we wished to contrast the effectiveness of the H'_3 measure with that of the H'_5 measure we have calculated a function P_z which is the exact probability under the null hypothesis of observing an F ratio equal to the calculated one (21). The computed P_z is accurate to $\pm 2.5 \times 10^{-8}$.

Tables 4A and B contrast the statistical parameters calculated from the six groupings of the data. In all cases the observed regression is highly significant.

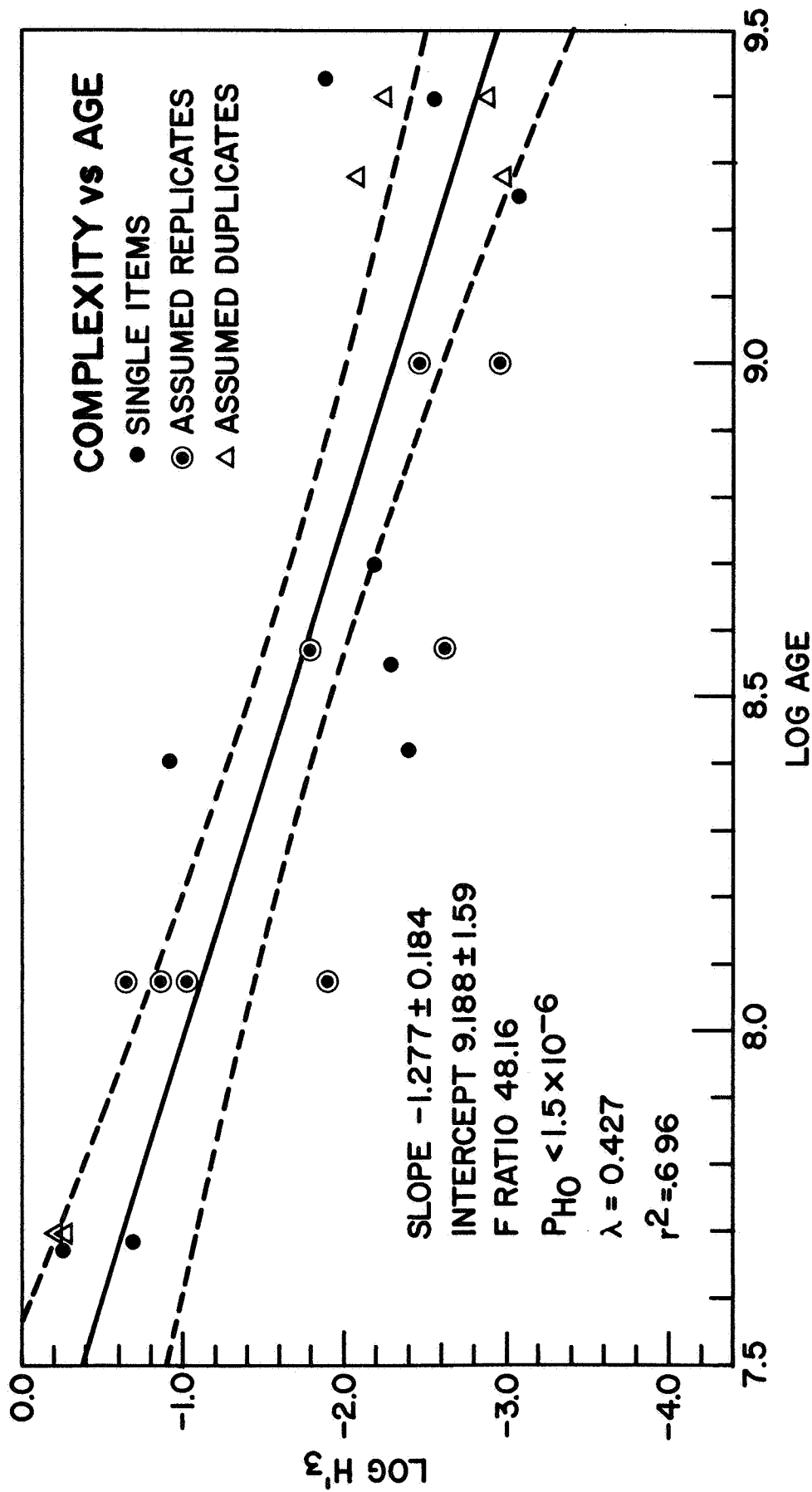Figure 14. Regression of complexity on age for the twenty-three materials shown in Figure 13. Complexity is measured by H'$_3$ the percentage power in the top 1/3 of the frequency spectrum. P$_{HO}$ is the liklihood that these results are due to chance. Dashed lines indicate the 95% confidence limits. The rightmost two triangles and the rightmost solid circle are the three short distributions containing only 6 or 7 points.

Figures 14 and 15 are plots respectively of the log* of the H'_3 and H'_5 complexity measures against log age for the entire set of points---e.g., including replicates and duplicates.

It will be noted that the fitted line in both cases rises to a value greater than 1 at about 7.3. Since by definition the H' function cannot exceed 1, its log cannot exceed 0, the relationship cannot be extrapolated beyond the range shown.

Lambda is the standard deviation of an x value estimated for a known y by means of the regression equation $y = A + Bx$. That is, given an observed y_0 we can estimate the corresponding x_0 by the formula:

$$\frac{y_0 - A}{B} = x_0 \mp \lambda \qquad (2.1)$$

The quantity lambda is the standard deviation of x_0 and consequently it constitutes a convenient index of accuracy of the regression equation.

In these groups of data the complexity measure allows prediction of the age accurate to about $\pm.33$ log units for the H'_5 function and $\pm.43$ log units for the H'_3 function.

The square of r, the correlation coefficient, corresponds to the percentage of total variance accounted for by the independent variable age (22). Between 70 and 80% of variance is accounted for by the regression on age.

For the reasons given above, these data cannot be said to demonstrate unequivocally that complexity of the n-alkane distribution declines with age. Such a conclusion requires an improved experimental design which insures at least a balanced sampling of materials of diverse geological, geographical and temporal origin, and control of sample handling procedures. But in view of the very high levels
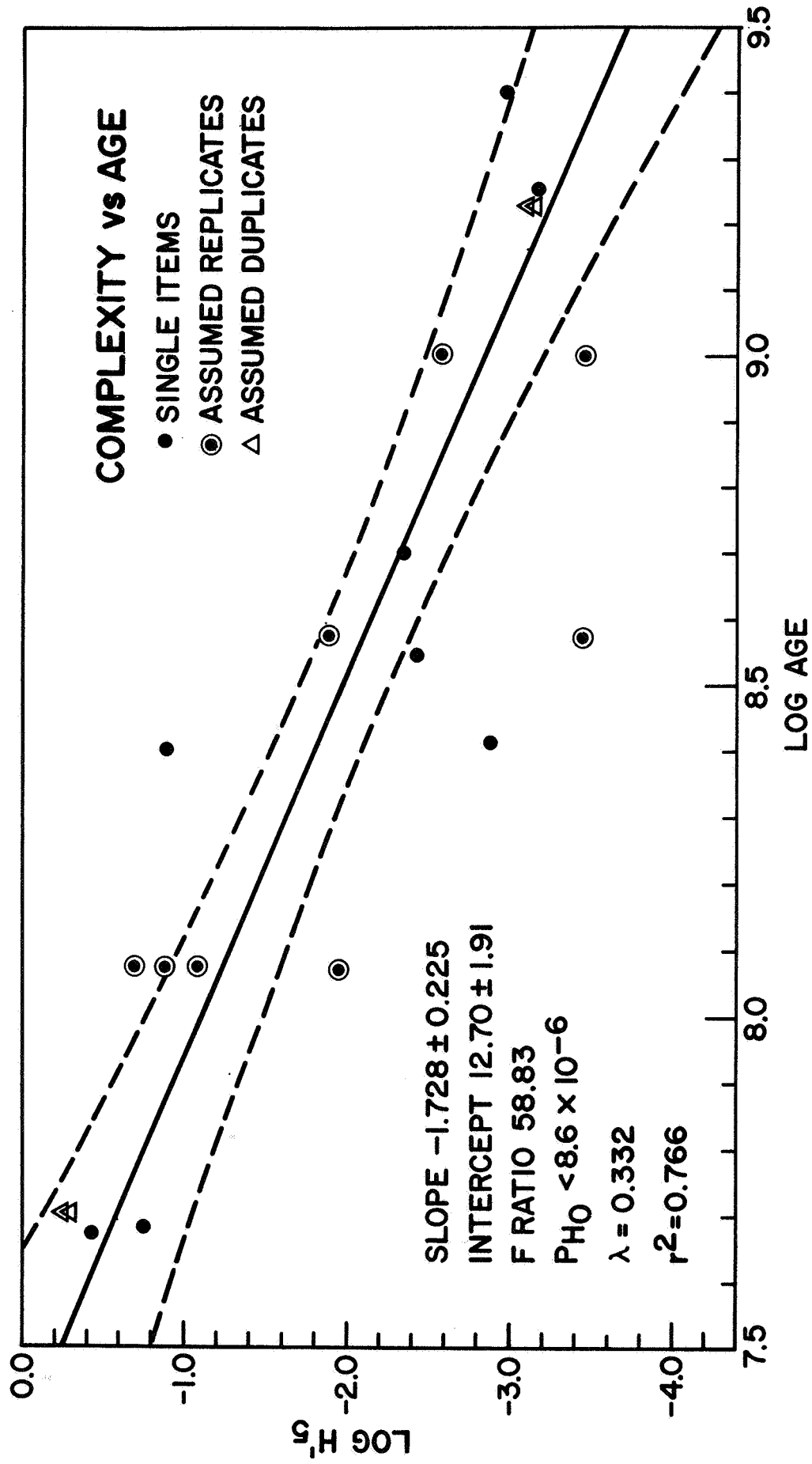
* Common logarithms are used throughout.

Figure 15. Regression of complexity on age for the twenty long distributions containing at least 10 components shown in Figure 13. Complexity is measured by H'_5, the percentage power in the top 1/5 of the frequency spectrum.

of significance reported here, it seems reasonable to accept the conclusion
that complexity is strongly influenced by one or more time dependent factors
until a more rigorous test of the hypothesis justifies its rejection.

Comparison of H'_3 And H'_5 Complexity Measures

One of the objectives of this project was to develop a measure of complexity
which permits the comparison of distributions which differ both in respect to
the range of the molecular weights represented and the total number of components
present. Carbon preference indices (CPI) which are the only measures presently
used to contrast aspects of the pattern of relative distributions of n-alkanes
are used primarily to contrast the distributions in respect to a common range
of components--conventionally C_{24} to C_{33} or even C_{35}. The complexity measures
presented here ignore the range represented, and the use of the integrated
power functions (the H' functions) permits the comparison of distributions
with different numbers of components.

In the preceding section we have used the H'_3 function to measure complexity
in a group of distributions including some short ones for which the H'_5 measure
cannot be estimated. The H'_5 function was used to contrast complexity in the
same group after the short distributions were removed. We may determine
whether the two complexity measures differ in their ability to discriminate
pattern properties that correlate with age by comparing their predictive effect-
iveness in the same group of long distributions for which both measures can be
computed. The results of an analysis of the partial regression of H'_3 on age
for this group are shown in Table 4C. Comparing these results with the data
of Table 4B, we see that both complexity measures perform about equally well,
there being no significant difference between the resulting values of lambda,
r^2 or P_{H_0}.

H'_3 and H'_5 are of course very strongly correlated, since the top 1/3 of the
frequency spectrum necessarily includes the top 1/5. Analysis of the regression

TABLE 5 COMPARISONS OF H'$_3$ AND H'$_5$ COMPLEXITY MEASURES

A (H'$_5$/H'$_3$) and age for long distributions only

Group	N	A	B	F ratio	P_{H_0}		r^2
I	20	1.742 $\pm$ .849*	-.232 $\pm$.10	5.38	.031	1.1	.230
II	18	1.451 $\pm$ .802*	-.196 $\pm$.10	4.29	.052	1.11	.212
III	13	1.211 $\pm$ .688*	-.169 $\pm$.08	4.39	.058	1.01	.285

B H'$_3$ and H'$_5$ for long distributions only

Group	N	A	B	F ratio	P_{H_0}		r^2
I	20	8.2×10^{-2} *	.868 $\pm$.037	545	10^{-10}	.037	.968
II	18	1.5×10^{-3} *	.823 $\pm$.041	405	10^{-10}	.037	.962
III	13	-2.2×10^{-4} *	.818 $\pm$.050	266	10^{-7}	.043	.960

* A not significantly different from 0.

of H'_3 on H'_5 shows that the former accounts for about 97% of the variance
of the latter (Table 5B). Consequently, it is possible that the observed
similarity in discrimination is due entirely to this correlation, and that
the H'_3 measure contributes nothing that is not included in the H'_5 measure.
If this is the case, the differences between the slopes of the two measures
cannot be significant. The question is important because we assume that the
decline in complexity, if real, may be due to one or more age-dependent phy-
sical process acting on the materials deposited; clearly the more information
we can obtain regarding the way in which complexity changes with age, the more
likely it is that one can identify processes that may be responsible. Thus,
if the difference in the slopes is in fact significant, then we should look
for a processes which results in the more rapid decline of higher frequencies.

This hypothesis may be directly tested by analyzing the relationship between
age and the ratio of the two measures, since if H'_5 does decline more rapidly than
H'_3 we would expect the ratio to show a corresponding decline. Table 5A shows
the results of an anlaysis of the relationship between age and the ratio of
H'_5 to H'_3. The change is clearly significant only for group 1 where N = 20,
but P_{H_0} is so near the 5% level for groups 2 and 3 that we may accept these
results as indicating that the ratios do decline with age, and consequently
that the H'_5 measure does in fact decline more rapidly than H'_3.

Comparison of Complexity and Odd Carbon Preference

An odd carbon preference in the high molecular weight region of the n-alkane
distribution is considered an indication of biological origin (9), (11). Odd
carbon preference is clearly related to pattern complexity in the range of C_n
over which it is defined. Consequently, we may inquire whether odd carbon
preference also correlates significantly with age.

The range over which the CPI is measured is arbitrary. Although C_{25} to C_{33} or
C_{35} is a common range employed in the comparison of relatively long distributions,
some authors may use a much broader range where some preference is observed.
The odd carbon preference index is usually defined as the mean of two ratios:

$$\text{Odd CPI } C_L \text{ to } C_H = 1/2 \left[\frac{\sum \text{Odd } C_L \text{ to } C_H}{\sum \text{Even } C_{L-1} \text{ to } C_{H-1}} + \frac{\sum \text{Odd } C_L \text{ to } C_H}{\sum \text{Even } C_{L+1} \text{ to } C_{H+1}} \right]$$

Where L and H specify the range of the odd carbon numbers used. Conventionally one uses CPI's to compare distributions which contain detectable quantities of all the components in the range over which the measure is defined (although one often sees in the literature generalizations regarding CPI's computed for different ranges). This is partly due to the fact that the measure is unstable under shifts in range.

Because of the diversity of ranges represented in our sample, we computed a number of different indices and compared them all to age. In general, a CPI is used to compare preference among the components of highest molecular weight. Consequently, we first defined carbon preference in terms of some C_n at the beginning of the range, and included all higher odd carbon number components shorter than the highest even component present.

Thus, if C_m is the highest molecular weight component present in the mixture, the range was C_n to C_{m-1} if m were even, or C_n to C_{m-2} if m were odd. A further requirement is that there be some minimum number of components in the range from C_n to C_m. After devising a number of indices defined in this fashion, we found the highest correlations between odd carbon preference - termed CPI_{21} - in the range C_{21} and up for "long" distributions containing at least 10 components in that range. In contrast, in the same group of distributions, carbon preference in the range C_{25} and up - CPI_{25} - shows a much less significant regression on age. (See Table 6A). We found this somewhat surprising in view of the fact the higher molecular weight components are generally thought to be more diagnostic of origin.

CPI_{21} corresponds to a larger portion of the total distribution than CPI_{25}, and

38

TABLE 6   COMPARISONS OF CPI AND COMPLEXITY

A   Long Distributions Only

Group	N	A	B	F Ratio		Pz	r^2
CPI_{25}	14	4.780 ± 1.751	-.540 ± .212	6.512	.024	.722	.352
CPI_{21}	14	3.139 ± .922	-.353 ± .112	10.02	.008	.581	.455
CPI_A	14	2.972 ± .933	-3.350 ± .113	8.81	.011	.620	.423
H'_5	14	13.54 ± 2.526	-1.818 ± .306	35.32	.0009	.310	.746

B   Distributions Containing at Least 6 Components Above $n\text{-}C_{21}$

Group	N	A	B	F Ratio		Pz	r^2
CPI_{21}	17	2.650 ± .663	-.292 ± .079	13.84	.002	.648	.480
H'_5	17	12.953 ± 3.506	-1.748 ± .229	58.21	2.6×10^{-6}	.316	.795

C   All Distributions

Group	N	A	B	F Ratio		Pz	r^2
CPI_A	23	1.985 ± .511	-.214 ± .059	13.07	.002	.820	.384
H'_3	23	9.188 ± 1.586	-1.277 ± .184	48.16	1.4×10^{-6}	.427	.696

consequently is a somewhat better indicator of total pattern. We therefore
decided to explore the relationship between a CPI defined for the whole
distribution and age, and defined CPI_A as the preference of all odd carbons
which are neither the first nor the last components in the distribution.
CPI_A proved to be a better discriminator than CPI_{25} and only slightly
inferior to CPI_{21}. Although all these preference indices show significant
partial regressions on age, none approach the complexity measures in
predictive effectiveness. Thus, Table 6 shows H'_5 a better predictor
of age than all CPI's for the 14 long distributions containing at least
10 components in the range $n-C_{21}$ to $n-C_m$. Table 6B contrasts H'_5 to
CPI_{21} for 17 distributions containing at least 6 components in this
range, and Table 6C contrasts CPI_A with H'_3 for all 23 distributions.
In all cases the complexity measure is a far better discriminator of
age than the preference index, as judged either by the value of lambda or the
level at which the null hypothesis is rejected.

Figures 16 and 17 are plots of Log CPI against Log age. Examination of
these plots shows that in more recent material there is considerable
scatter while the odler materials are approximately equal in CPI. In
short, although a straight line can be fitted to these data, the CPI
values are not as useful indices to age as the lambda values of Table
6 might suggest.

D. SUMMARY OF CONCLUSIONS

The primary conclusions that may be drawn from the results reported in
this section are:

1.   Complexity as measured by percentage power in the top 1/3 and
top 1/5 of the frequency spectrum is observed to correlate strongly
with estimated age in an inhomogeneous group of hydrocarbons of diverse
geological, temporal, and geographical origin, percentage power in the
top segment of the frequency spectrum declining approximately linearly
with age in the range from about 5×10^7 to 2.7×10^9 years.

2.   The rate of decline of percentage power with age is slower fro the top
1/3 of the frequency spectrum than for the top 1/5, and this difference is
significant.
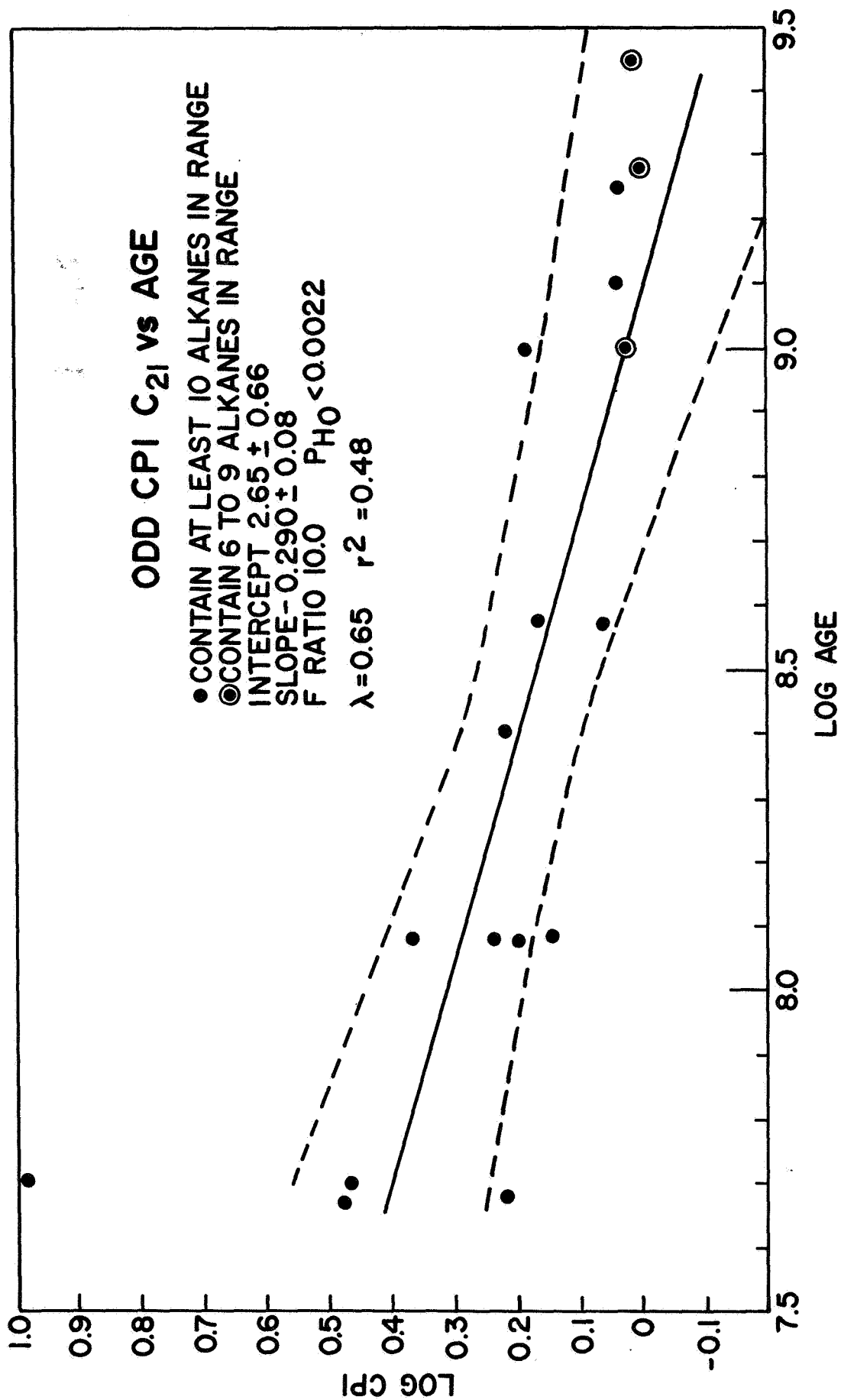
Figure 16. Regression of CPI_{21} on age for 17 of the distributions in Figure 13 containing at least 6 components above C_{21}. The scatter for the leftmost four points may be contrasted with the complexity values for these distributions in Figures 14 and 15.
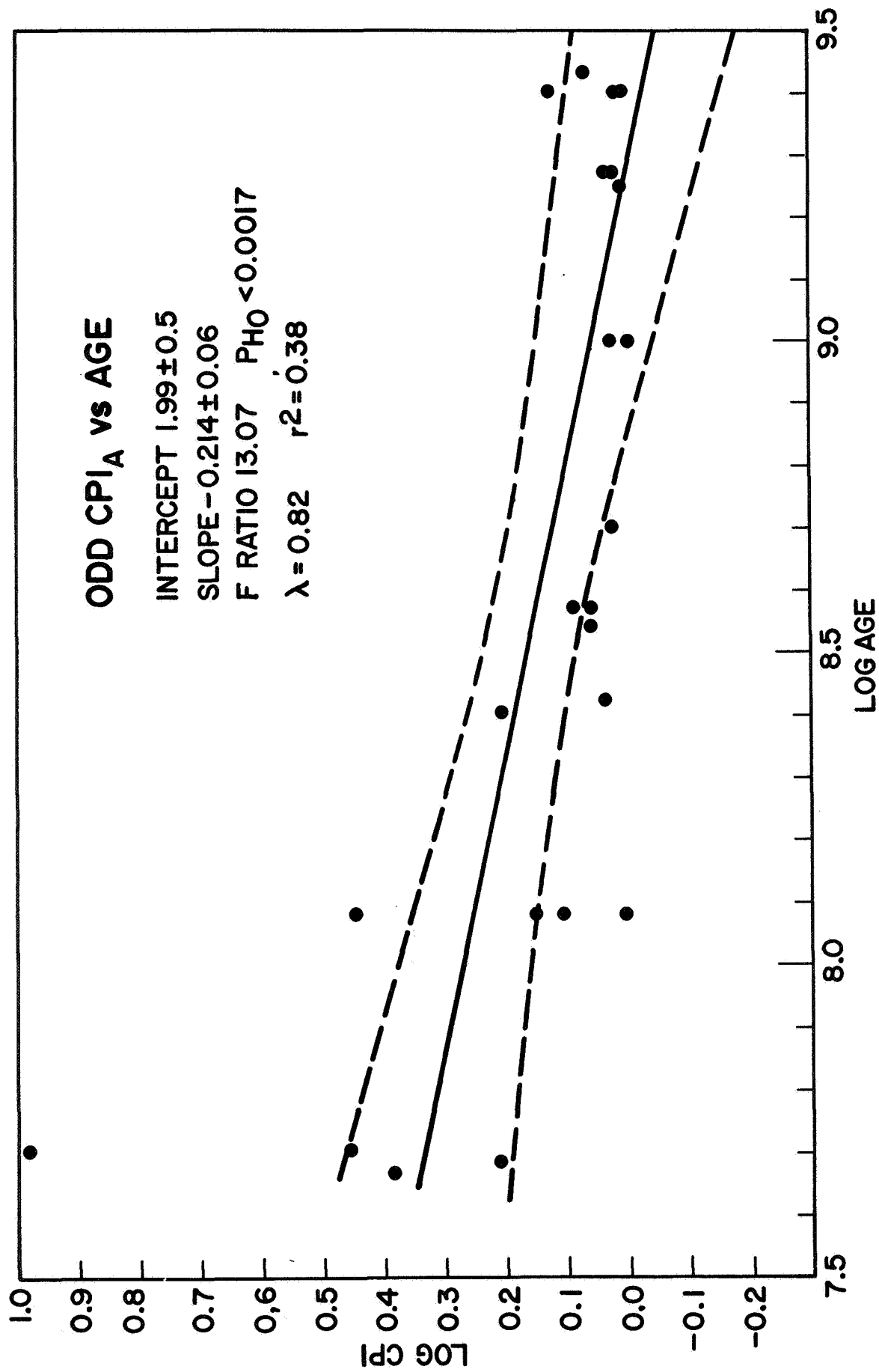
Figure 17. Regression of Odd CPI$_A$ computed for the whole distribution on age for all 23 distributions shown in Figure 13.

3. Odd carbon preference indices are functions of pattern complexity, and are also observed to show a significant linear regression on age in the set of samples examined. The strongest correlation is observed for indices where the range of odd carbons considered represents a large proportion of the total distribution, the best fit being observed when the preference index is computed for all odd carbons in the range C_{21} and up. An odd carbon preference defined over the whole distribution is almost as good a discriminator of age. However, no matter how defined, odd carbon preference is not nearly as good a discriminator of age as the complexity functions presented.

4. These results support the conclusion that one or more age-dependent factors affect the overall shape of the distribution of relative abundance of n-alkanes. While these shape changes are partially indexed by odd carbon preference, they are much more sensitively discriminated by the complexity measures presented.

PART III   COMPLEXITY AND ORIGIN OF n-ALKANES IN NATURALLY-OCCURRING AND
              SYNTHETIC HYDROCARBONS

Our original motivation in attempting to quantify the differences in perceived

smoothness of n-alkane distributions lay in the common observation that the

degree of unevenness in these distributions tends to be associated with their

origin, namely modern biological materials have more uneven distributions than

ancient biological (geological) distributions which are in turn more uneven

than synthetic materials (n-alkanes synthesized by the Fischer-Tropsch method).

Therefore, it should come as no surprize that a measure of smoothness should

exhibit the same association in the same set on which the subjective conclusion

is based. A more important question is whether measured complexity indexes more

subtle classifications than these.

We have shown this to be the case in ancient biological materials where the

complexity varies very significantly with estimated age. Does complexity serve

similarly to discriminate among modern materials of diverse origin? Does it

rank modern materials on any meaningful continuum, and, in particular, does

it yield a ranking that could be construed as corresponding to "biological

order"?

                         A.   PROCEDURE

As in the work described in the preceding section, we were limited in this inquiry

to a consideration of n-alkane distributions published in the literature, which

introduces many methodological problems. Although these materials can be classi-

fied in various subgroupings, and thus yield a broad range of categories within

which we may look for significant associations with measured complexity, we

42

cannot assume that they are, statistically speaking, good random samples from the populations in these categories. (For most of these categories, it would be difficult to decide what a random sample should consist of.)

There is no control on interlaboratory variations in analysis technique, and not all samples can be assumed to be free from contamination. In addition, in this study we made extensive use of Clark's catalogue of n-alkane distributions (3), and it is likely that in many instances he does not report the amplitudes of all extractable n-alkanes present in the material analyzed, but follows the original authors in reporting only those in a specific range of interest. Furthermore, since we were limited to the literature, we were unable to obtain sufficient examples from all interesting categories to draw reliable inferences from all our observations.

A total of 95 different n-alkane distributions taken from ten distinct categories were examined and the H'_3 complexity measure computed for each. The categories and materials are described below. Six of the ten categories are materials of biolgocial origin, containing a total of 64 different n-alkane distributions of which all but 13 were obtained from Clark (3). Eight of the remaining were analyzed by Simmonds (13) and 5 were obtained from primary literature sources. The categories and the codes assigned them are described below:

Code 1   Land animals. Nine materials comprised this category: 2 samples each of cow manure, bat guano, and beeswax and one sample each of adult housefly, wool wax, and sheep's wool. The range in complexity was from .239 for sheep's wool to a high of .708 for beeswax. Of all

43

large amounts of partially digested algae and/or petroleum contamination.

Code 6   Kelp. One sample with $H'_3 = .131$.

Code 9   Synthetic n-alkanes. Two samples of n-alkanes synthesized by the Fischer Tropsch method, with H'_3 measurements of .00233 and .00257. See Figure 18.

Code 10   Floor wax and laboratory floor dust from Oro, et. al. (17). One sample of each is included, primarily because they represent the smoothest distributions observed, and because they may be argued to represent very old biological materials subjected to fairly extensive chemical and thermal degradation during the industrial petroleum refining process of which they constitute a residue (under the reasonable assumption that the floor dust alkanes are in fact floor wax alkanes). The H'_3 values are .00042 and .00052.

Code 11   Carbonaceous Chondrites. A total of four samples - two each from Orgueil and Murray meteorites which range in H'_3 value from .0021 to .0776 from Oro, et. al. (17). We may class these as samples of unknown origin, keeping in mind the fact that they may have been contaminated with material of recent biological origin. These are pictured in Figure 18.

Code 12   Aged biological samples. The twenty-three samples in this category are the twenty-three used in the study of complexity and age. The H'_3 range is .00081 for Precambrian crude oil (1.8×10^9 years) to .582 for Green River Shale 2 (5×10^7 years).
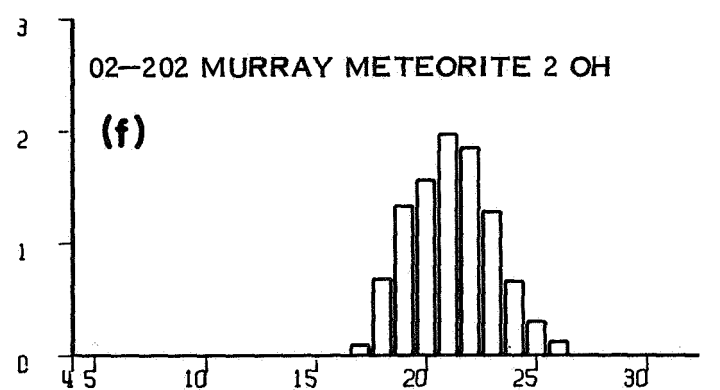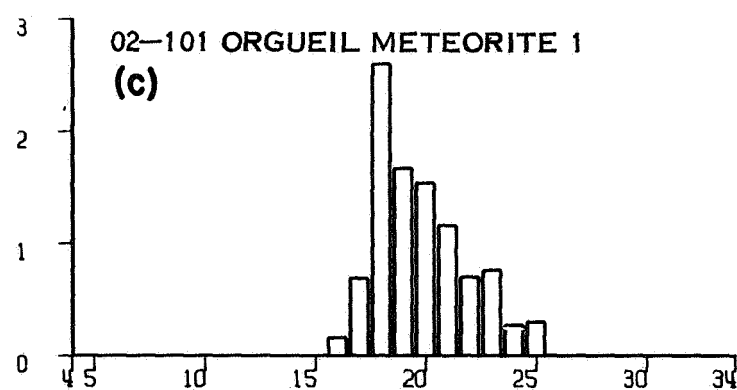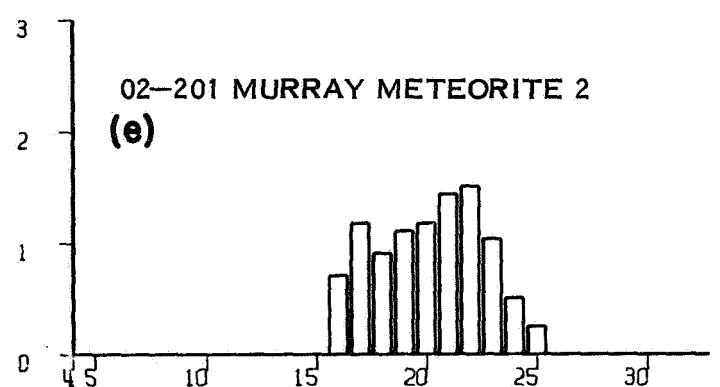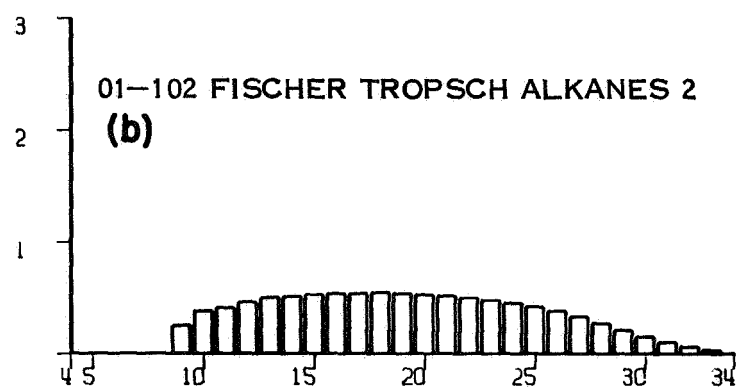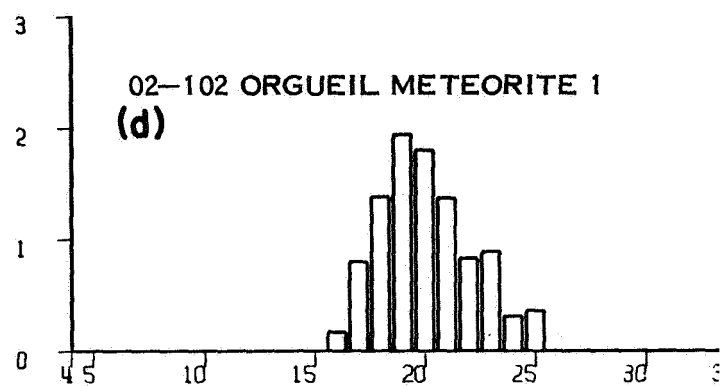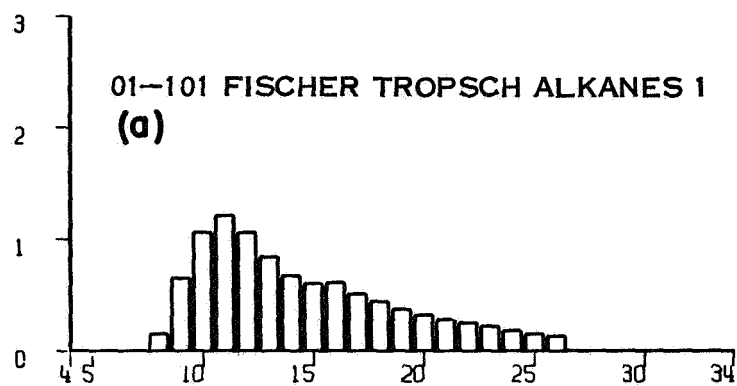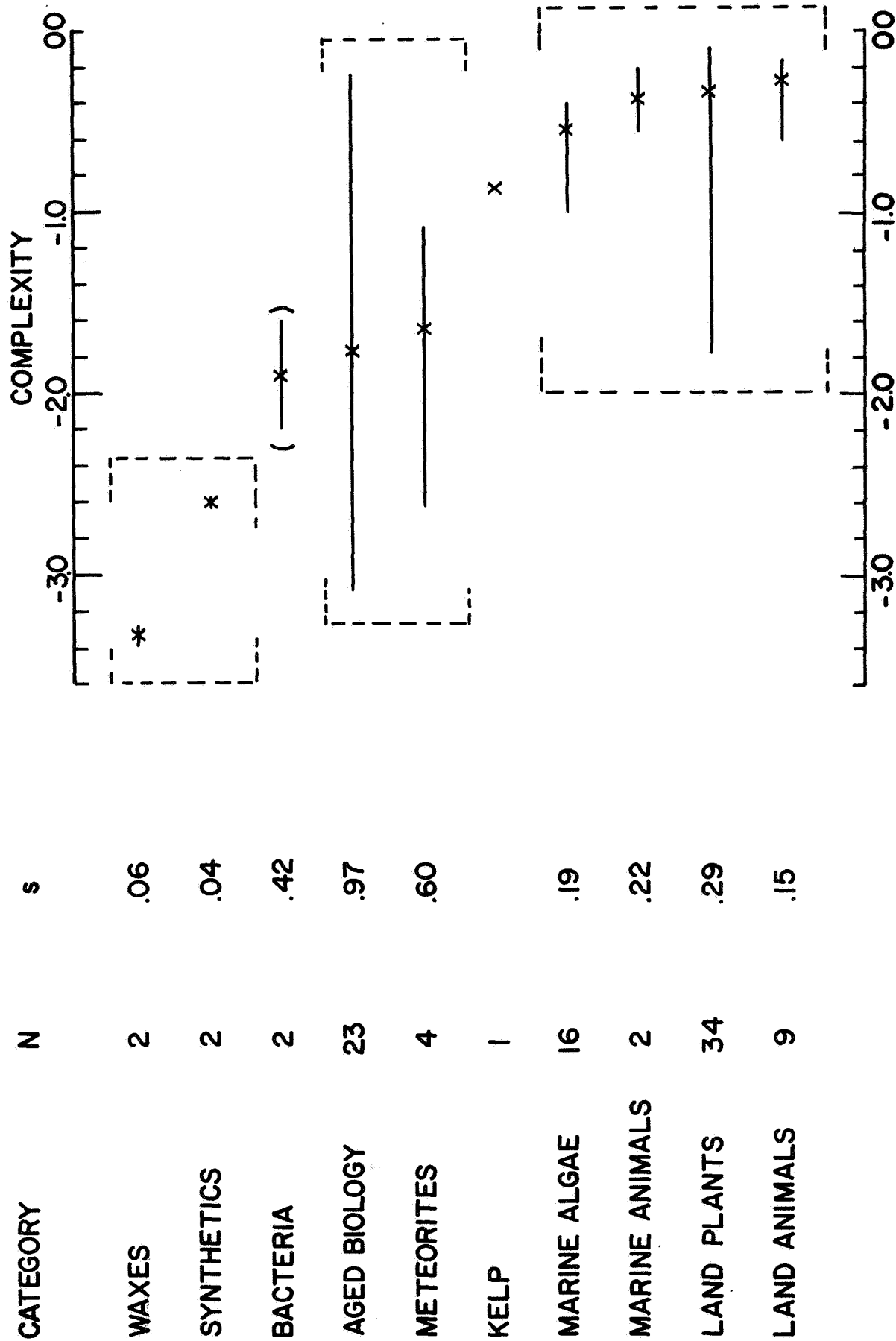
44

Figure 18. N-alkanes from Fischer Tropsch synthesis and carbonaceous chondrites. These categories differ significantly in mean complexity.

these samples only adult housefly represents tissue. The use of cow manure and bat guano which represent partially digested plant material may bias this category.

Code 2 Bacteria. Two samples: B. cereus from Tomabene and Oro (16), with an H'_3 of .0064 and Sarcina lutea albro from Clark (3) with an H'_3 of .025. Both are extremely smooth distributions (see Figure 18).

Code 3 Land Plants: 34 distributions mostly of leaves, leaf waxes, grasses and cereals. The H'_3 range is from a low of .017 for wood extract of Manilkara bidentata to .783 for oats. Although several subgroupings of plant types are represented, these subgroups often correspond to a single experimenter. Consequently, we did not group by plant type lest this reflect variations due to laboratory technique, selectivity in reporting the n-alkanes present, or control of contamination rather than variability due to plant type.

Code 4 Marine algae. Sixteen samples of marine algae reported by a single author (Clark (3)) comprise this category. No controls on the possibility of contamination by petroleum pollution are reported. The range of H'_3 is from .097 for Chryptophycae to .4 for Corallina Officinalis.

Code 5 Other marine animals. This category contains only two samples: mixed plankton with $H'_3 = .292$ and herring oil with $H'_3 = .6$. This group is too small and too heterogeneous for meaningful comparisons but we retained it because we could not justify putting either of these examples in other groups. The mixed plankton measurement seems low in contrast to land animals but this may be due to the presence of

CATEGORY	N	s
WAXES	2	.06
SYNTHETICS	2	.04
BACTERIA	2	.42
AGED BIOLOGY	23	.97
METEORITES	4	.60
KELP	1	
MARINE ALGAE	16	.19
MARINE ANIMALS	2	.22
LAND PLANTS	34	.29
LAND ANIMALS	9	.15

RANGE AND MEAN COMPLEXITY BY CATEGORY

Figure 19. Mean and range of log complexity values of 10 categories of materials. N is the number of items in each category. Dashed brackets group categories whose means do not differ significantly from one another but differ significantly from all other categories. Parenthesis indicate pivotal categories which differ significantly from all but adjacent categories.

Figure 19 shows the range in complexity values and the geometric mean complexity

value for the 10 categories considered, ranked in order of increasing complexity.

Although there is considerable overlap in complexity between categories, this

presentation suggests that some groups may differ significantly from one another.

The statistical significance of the complexity differences may be examined by

means of Fischer's test of least significant difference between the means (23).

This test assumes that the means of the sampled items exhibit only a random

deviation from the true population mean. The log transformation is used to

reduce skew and variance.

This test results in the groupings indicated by the dashed brackets in Figure 19.

Thus, the waxes and synthetic materials form a group in that their means do

not differ significantly from one another and do differ significantly from all

other groups enclosed by brackets. Meteoritic n-alkanes and alkanes from ancient

biological materials form a similar group. The bacteria are a pivotal category

between these two groups in that its mean does not differ significantly from

that of either category ranked next to it, but does differ from all other

categories. Since the kelp category contains only one item, its groupings

are not significant and it is included only for comparison purposes. By this

test all material of modern biological origin except the bacteria also form a

group whose means do not differ significantly from one another.

For categories containing more than five items, the Kruskall-Wallace one-way

analysis of variance test may be used to test the significance of rank ordering

imposed by the complexity measure (24). When applied to the marine algae, land
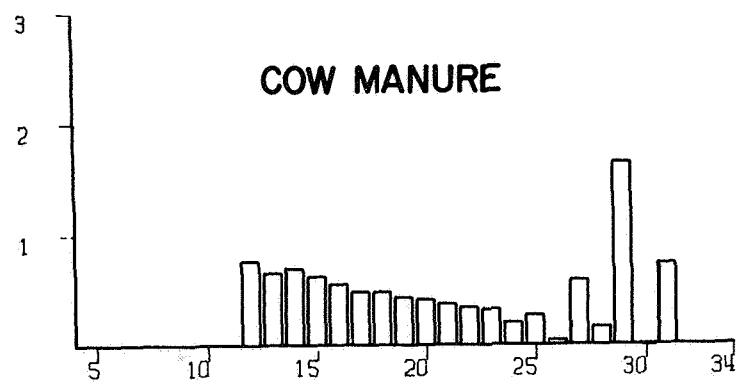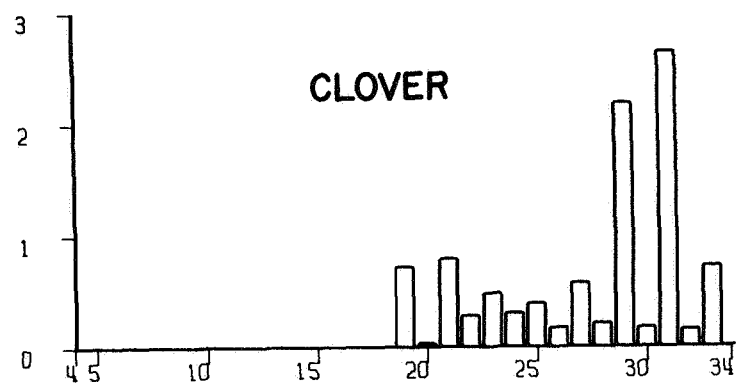
Figure 20. Biological degradation of n-alkanes. Figure 20A is n-alkanes from spotted burr clover; 20B is n-alkanes from manure of cows fed on the clover. From Oro, et. al. (11).

plant and land animal categories this test shows that the marine algae are ranked significantly lower than the members of the other two categories, but that these categories do not differ significantly in ranking from one another.

Before discussing the possible interpretations of these findings, we present additional evidence regarding the effects of biological and nonbiological degradation on the complexity of n-alkane distributions.

C.   BIOLOGICAL AND NONBIOLOGICAL DEGRADATION OF N-ALKANE MIXTURES

We have proposed that both biological and nonbiological processes acting on n-alkanes of biological origin are capable of removing biological order. If complexity reflects order, we would expect these degradation processes to reduce complexity. We present one case of biological degradation and one of nonbiological degradation which appear to support this hypothesis.

Biological Degradation

Figure 20 shows n-alkanes extracted from spotted burr clover (whole plant) and those from the manure of cows fed on the clover: H'_3 of the clover is .688, that of the manure is .545.

If the n-alkanes in the manure in fact were derived from those of the whole plant by a process of biological degradation, then we may conclude that in this case at least biological degradation does in fact reduce measured complexity. In (11) Oro, et. al., points out that n-alkanes extracted from leaves of clover have a distribution which resembles that of cow manure in the range C_{23} to C_{33} more

47

strongly than do the n-alkanes from whole plants. Unfortunately these authors did not publish the original chromatogram of the leaf alkanes, and the published graph was too incomplete to be usefully analyzed by us.

It will be observed that the occurrence of biological degradation of n-alkanes introduces a further problem into the analysis of the significance of the pattern of the n-alkane distribution. The highest concentrations of naturally-occurring n-alkanes are usually to be found in the materials of the outer surfaces - waxes, oils and tough horny material. All of these areas tend to be heavily colonized by microflora, the density per unit area often approximating that in a very heavily populated garden soil, and it may be supposed that these microflora degrade the n-alkanes present. It follows that contamination with biologically degraded n-alkanes may contribute significantly to the wide spread in complexity values observed for such materials. We may, for example, wonder whether the very smooth distributions of the n-alkanes of the two bacteria samples discussed above reflect the bacterial cell structures or the presence of materials degraded by them.

Non-Biological Degradation

After deposit, biological hydrocarbons may be subjected to thermal degradation. Although the mean temperatures of the environments of such materials may not approach those necessary for pyrolosis in a laboratory environment, we might expect that exposure to even moderate temperatures for long time periods would have an effect on the n-alkanes initially deposited. According to our hypothesis, such degradation should result in a reduction of biological order and may

48
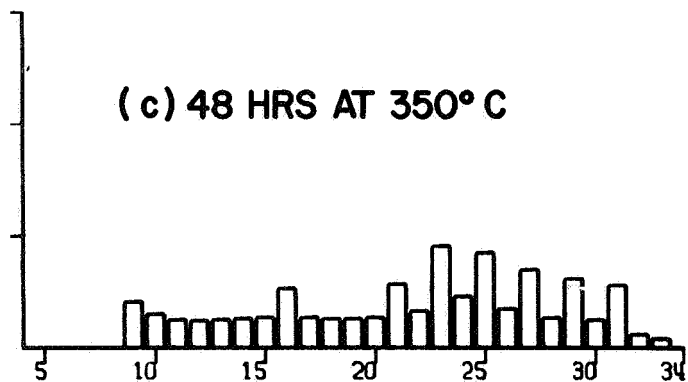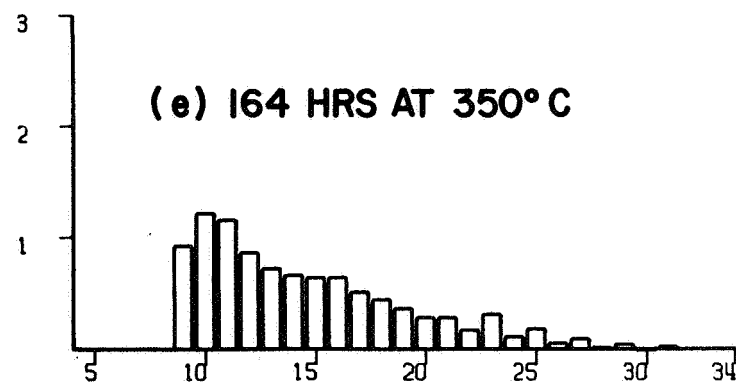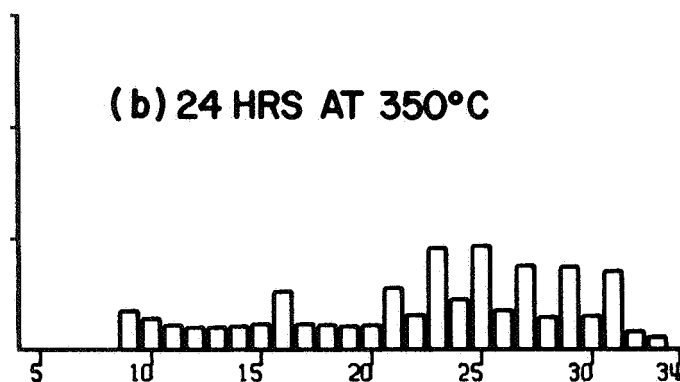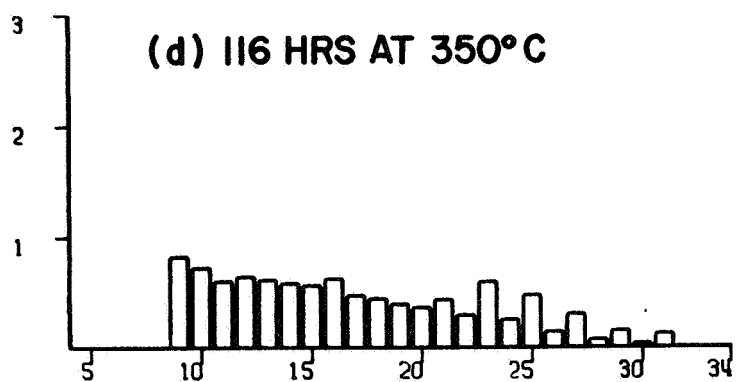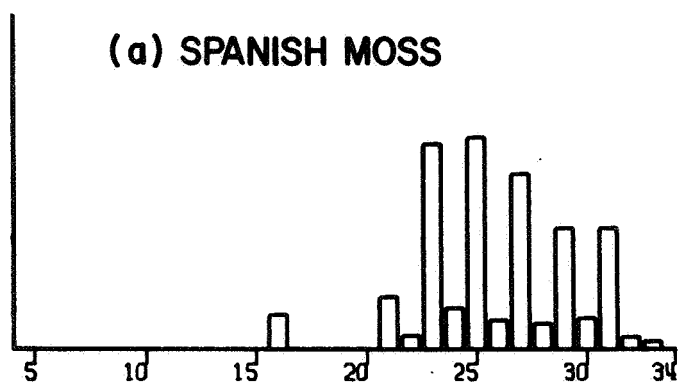
Figure 21. Thermal degradation of biological n-alkanes. Alkanes were extracted from Spanish moss and subjected to pyrolysis in a sealed vessel.

therefore reduce complexity as well. The effects of thermal degradation were explored by Peter Simmonds (13) who subjected the n-alkanes extracted from samples of plant materials to thermal degradation.

Figure 21 shows one such experiment. Figure 21a is the starting material, b, c, d and e are taken from chromatograms of the same mixture after it had been pryolozed at 350° C for 24, 48, 116 and 164 hours. The bar plots correspond to relative amplitudes of peaks which could correspond to n-alkanes or n-alkanes and n-alkenes. We do not know what percentage of the starting material was transformed to other pyroloysis products. Some attempt was made to prevent loss of the volatile low molecular weight components. In spite of these uncertainties, it is clear that pyrolysis results in the production of low molecular weight products (n-alkanes or n-alkenes) and reduces the visual complexity of the initial distribution, resulting in a progressive smoothing with time. Thus, Figure 21b shows that after only 24 hours the relative abundance of the high molecular weight components with odd numbers of carbon atoms is greatly reduced, while the intervening even numbered molecules have increased in relative abundance. This fact in combination with the appearance of components in the range C_9 to C_{15} not present in the starting material results in a great reduction of the variance of the 24-hour distribution relative to that of the starting material. Because our measure of complexity is the ratio of power in a given band or set of bands to variance, the complexity does not show a smooth decline during this interval. Figure 22 is a plot of complexity against time for the whole distribution. It will be noted that H'_3 and H'_5 drop sharply at first, then "plateau" or even increase slightly in the next interval and thereafter decrease rapidly. H'_9 actually increases during the first twenty-four
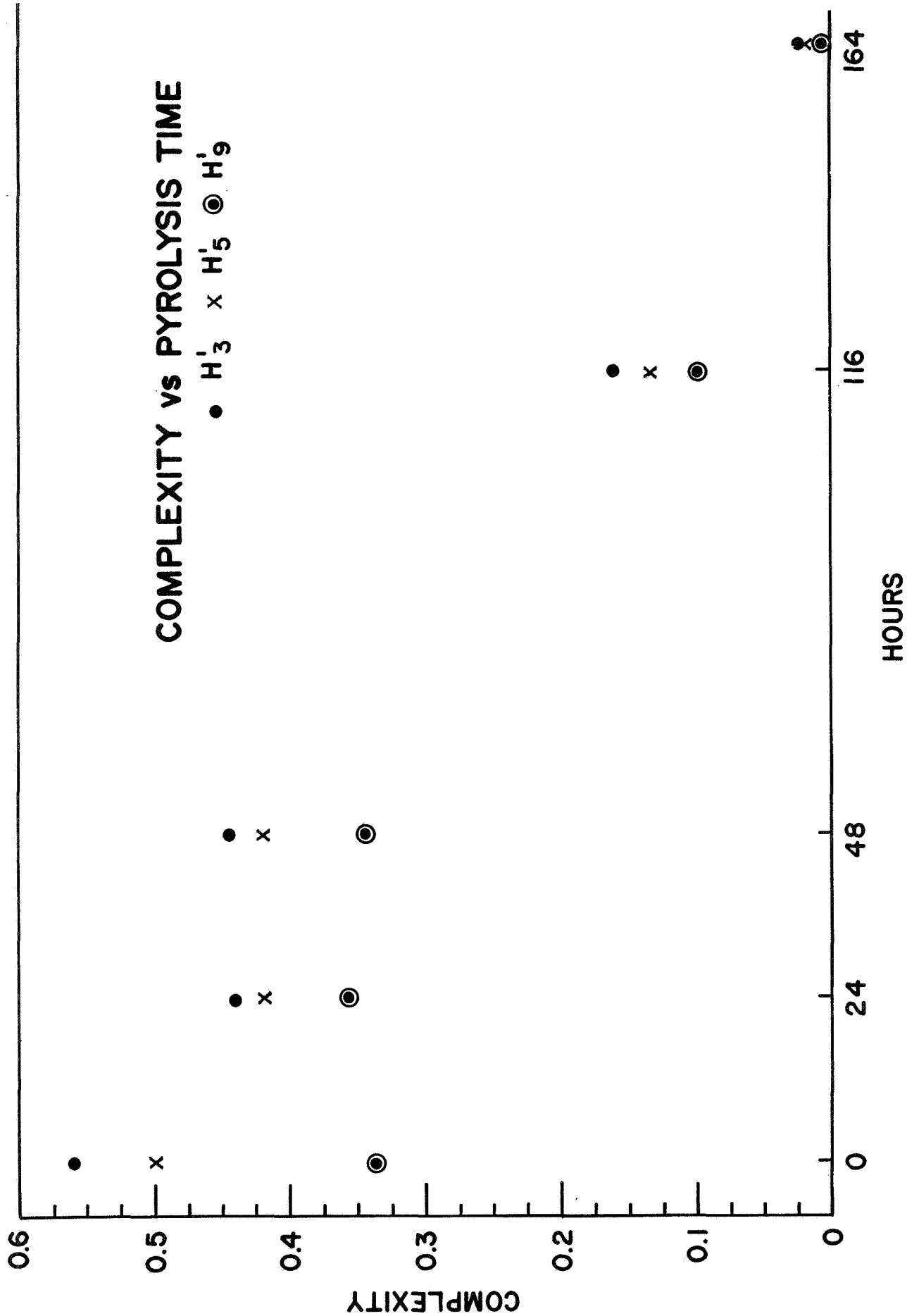
49

Figure 22. Complexity changes with time during pyrolysis. Degraded mixtures contain more components than the starting mixture.

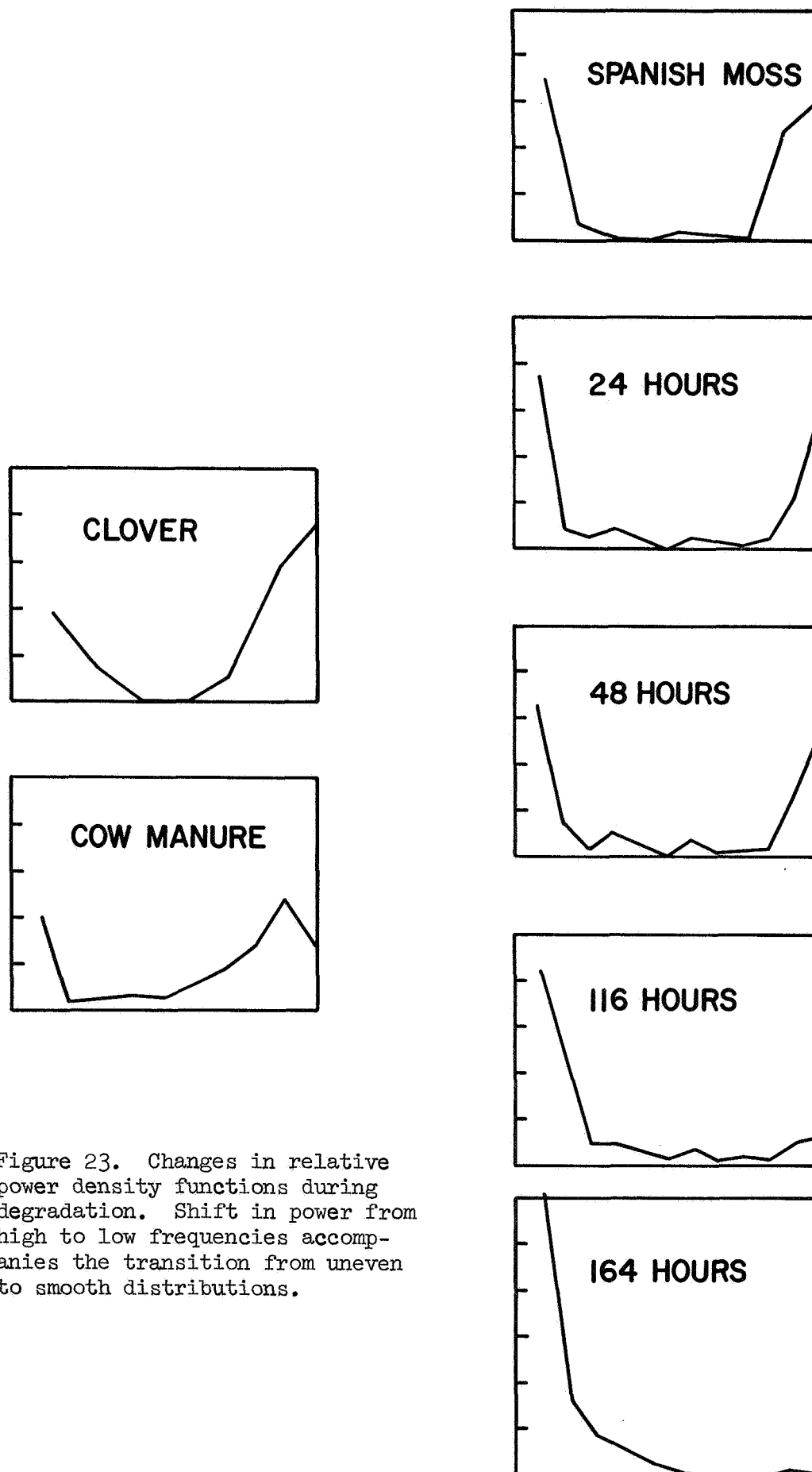# BIOLOGICAL AND THERMAL DEGRADATION

Figure 23. Changes in relative power density functions during degradation. Shift in power from high to low frequencies accompanies the transition from uneven to smooth distributions.

hours, and declines thereafter. Studies of simulated degradation processes reported below have shown that plateauing or increase in complexity measures during the first stages of degradation are greatly influenced by relative vulnerability to fragmentation of molecules of different weight, increases or plateauing being more pronounced when all molecular species are equally likely to experience fragmentation.

Additional information regarding the changes occurring during degradation can be obtained by examining the entire relative power density functions plotted in Figure 23.

D. DISCUSSION

The data on thermal degradation of the Spanish moss hydrocarbons and biological degradation of clover imply that both of these mechanisms act to reduce the complexity of n-alkane distributions of biological origin. Taken together with the change in complexity with age observed in the age-dated series, we may conclude that some of the age-related differences in the patterns of n-alkanes may be due to degradation. The question remains whether any of these differences are due to inherent differences in the complexity of the original starting material. A clear difference in complexity of modern biological materials as a function of phylogenetic age would answer this question, but the data we have presented do not support such a conclusion, although they do not rule it out. While mean complexity roughly ranks recent biological materials in approximate order of phylogenetic origin, these differences are either not significant statistically, or could have arisen as a result of contamination. Thus, the low position of bacteria could be due to the fact that these organisms are very

efficient degraders of n-alkanes of plant and animal origin, or it might even be due to contamination with petroleum products. Similarly, there are good grounds for questioning the marine algae distributions in view of the widespread oil pollution of the marine environment. The fact that the land plants exhibit a broad range in complexity values which actually overlaps that of the marine algae suggests that these differences should not be interpreted as reflecting differences in inherent structure, in spite of the significance of the difference in ranking. However, it might be that the land plants are also contaminated with insecticides containing petroleum products.

The complexity of the synthetic hydrocarbons also introduces interesting questions regarding the use of complexity measures for distinguishing between biological and nonbiological materials. If we assume that these materials are representative of n-alkanes of nonbiological origin, and that the age-dated series are of biological origin, (as demonstrated by the presence of pristane and phytane in the same series (Calvin (1)) but differ primarily in degree of degradation, then we may conclude that degradation is capable of removing all evidence of biological complexity from n-alkanes. (This conclusion is supported by simulation studies presented below which show that random fragmentation of n-alkanes can produce comparable changes in pattern complexity if the fragments thus formed are transformed into alkanes.) The observed range of values for the n-alkanes from carbonaceous chondrites is consistent with the hypothesis that they consist of synthetic hydrocarbons contaminated by material of biological origin.

It appears that these data, while equivocal, strongly suggest that n-alkane distribution data obtained under conditions in which contamination is controlled could serve to test these hypotheses regarding relationship between the phylogenetic age and complexity, and to further confirm the complexity-reducing effects of degradation.

PART IV   SIMULATED DEGRADATION OF N-ALKANES

In preceeding sections of this report, we have shown that the complexity

of the pattern of relative abundance of n-alkanes declines approximately

linearly with age in a sample of hydrocarbons of diverse geological and

geographical origin varying in age from 50 to 2700 million years.

We have also proposed that these differences may be due to one or both

of the following mechanisms:

a.   Random degradation of the pattern by physical and/or biological

processes acting in the environment of the material after depo-

sition, and

b.   Increase with time of the complexity of the material at the time

of deposition.

In Part III we have shown that biological degradation and thermal degradation

act to reduce the complexity of the n-alkane distributions of modern materials

of biological origin. In these latter studies, it appears that high molecular

weight n-alkanes are reduced in relative abundance while low molecular weight

n-alkanes not present in the starting material are produced. This suggests

that degradation consists of the cleavage of carbon-carbon bonds in n-alkanes

accompanied by the production of new n-alkanes from fragments thus produced.

Would such degradation processes, acting over a long period of time and

accompanied presumably by the loss of more volatile and hence more mobile

low molecular weight components cause n-alkane mixtures characteristic of

modern land plants to resemble visually the patterns of relative abundances

of n-alkanes observed in the age-dated series presented above? Would such

degradation processes reduce the complexity of n-alkane mixtures in a manner

analogous to that observed in the age-dated series?

The simulated degradation studies presented here were undertaken to answer

these questions. Specifically, we wished to explore the effects on complexity

and on the overall shape of the distribution of the simplest conceivable kinds

52

of "random" degradation and migration. The purpose of these studies is not to validate any particular detailed model of n-alkane breakdown, but rather to determine whether in general the consequences of repeated fragmentation would be to modify the shapes and the measured complexity of n-alkane distributions in such a fashion as to mimic those observed in the age-dated series or those produced during pyrolysis.

A. DEGRADATION MODELS

We assume that a specified starting mixture of n-alkanes is changed only by the following two mechanisms:

1. The loss through fragmentation of n-alkanes, which consists of the destruction of one carbon-carbon bond and the consequent creation of two daughter molecules whose combined length equals that of the parent molecule. In other words, all fragmented molecules are assumed to be immediately converted into n-alkanes of shorter length than the parent molecule.

2. The loss through migration of low molecular weight components in quantities proportional to their relative vapor pressures away from the mixture.

Fragmentation Models

The fragmentation models we studied varied in the following respects:

1. The number of fragmentations per unit simulated time. This was either a constant fixed at the beginning of the simulated experiment, or it was specified as a fixed percentage of the number of carbon-carbon bonds present in the mixture subjected to degradation, so that the number of fragmentations per unit time declined at least exponentially during the duration of the simulated degradation experiment.

53

2. The relative fragility of n-alkanes of different lengths. If fragmentations are randomly distributed among all the molecules in the mixture, then all molecules are equally fragile, so that the number of molecules of any species actually experiencing fragmentation during a given time period depends on its concentration in the mixture and not on its molecular length. Alternatively, if fragmentations are randomly distributed among all the carbon-carbon bonds in the mixture, then the number of fragmentations per unit time will be proportional to n-1 as well as to the concentration of a given C_n in the mixture.

3. The distribution of fractures among the bonds in each molecule subjected to fragmentation. Given that a molecule experiences destruction of one of its bonds, the length of the two daughter molecules formed is determined by the location of the fracture, and consequently the distribution of lengths of daughter molecules depends upon the relative liklihood of fracture of each of the bonds in the parent molecules. Two models of bond fracture were employed in the simulated experiments reported here. The first model provided that each of the carbon-carbon bonds were equally likely to be destroyed, so that equal numbers of all possible daughter molecules are produced from fragmentation of a given parent. We term this the linear or non-Gaussian fracture model. The other model explored provided that bond fractures are distributed among the bonds in a Gaussian normal fashion, where the mean of the distribution was the midpoint of the molecule and the standard deviation was approximately 1/6 the length of the molecule. This is referred to as the Gaussian fracture model. The differences between the Gaussian and non-Gaussian fracture pattern are shown in Figure 24 where the product yield for each pattern from n-C_{28} is plotted.

54

WEIGHT DISTRIBUTION OF n-C$_{28}$ FRAGMENTS
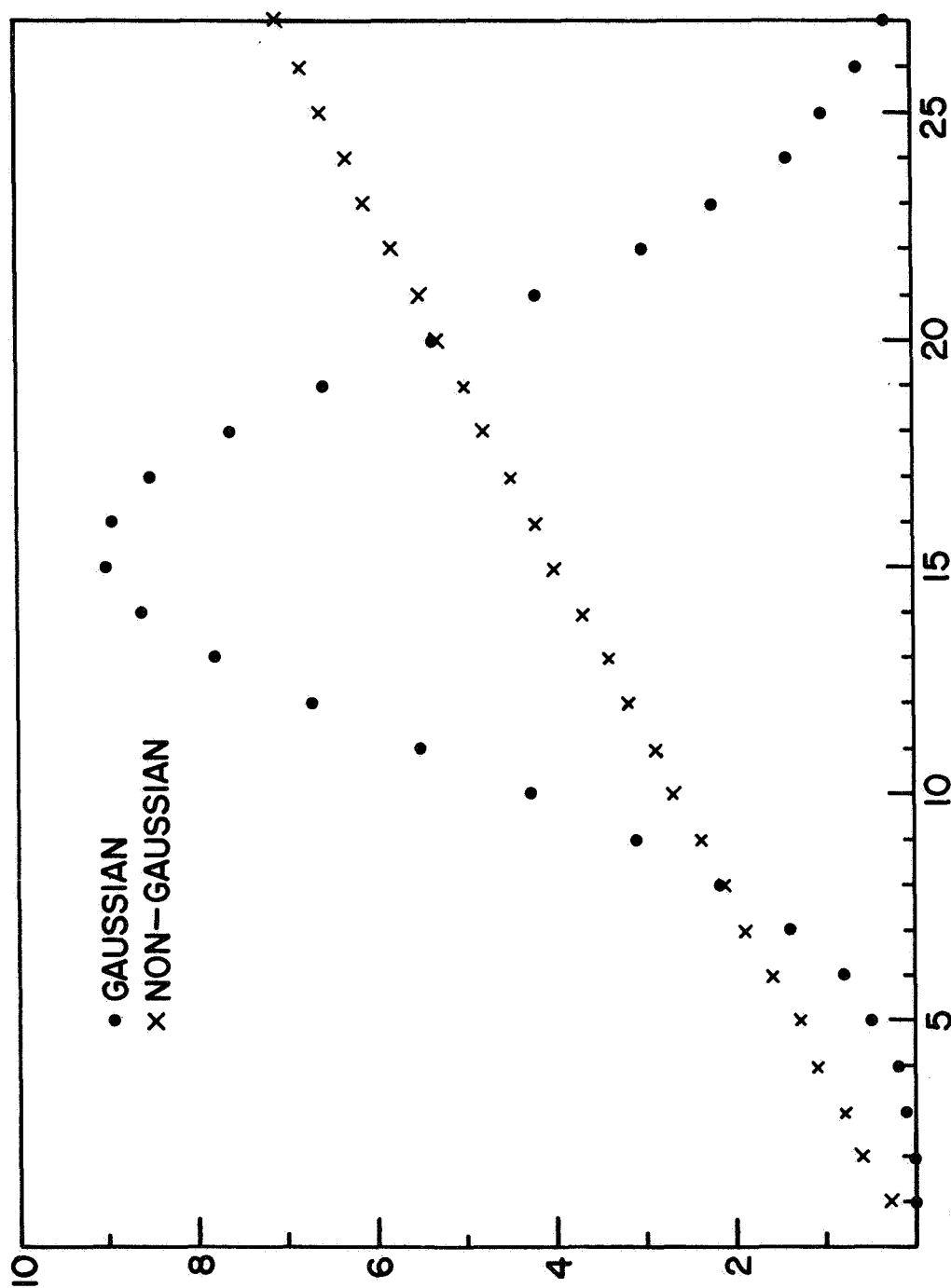
• GAUSSIAN
× NON-GAUSSIAN

Figure 24. Distribution by weight of fragments created from n-C$_{28}$ by Gaussian and non-Gaussian bond fracture patterns.

These three pairs of alternative models define highly simplified modes of
degradation which may be thought of in terms of the following idealized
interpretations of the effects of either temperature or background radiation
as the causitive degradation mechanism.

Thermal interpretation. If the thermal environment causes the degradation, we
may think of it as specifying a fixed heat input of so many units of energy
per unit time which must be dissipated in the destruction of c-c bonds so
that under constant thermal conditions a fixed number of bonds are destroyed
per unit time. Alternatively, we may think of the thermal environment as
determining the temperature of the mixture, which in turn defines a liklihood
of fragmentation per unit interval over the bonds in the mixture corresponding
to the percentage fragmented. Under either circumstance, it seems reasonable
to suppose some difference in relative fragility of molecules of different
lengths as a linear function of molecular length. If the dependence is
slight, then the true difference might be well approximated by the equal
fragility model; alternatively, the proportional fragility model indicates
the general kinds of effects to be expected when fragility is strongly
dependent on molecular weight.

The thermal interpretation suggests that actual fracture is due to some
kind of vibration or bending action imposed on the molecules. If so, we
might suppose that molecules are more likely to break in the middle than
at the ends. The Gaussian fracture model was designed to indicate the
effects of this kind of fragmentation.

Radiation interpretation. If we choose to consider high energy particle
bombardment as the causitive agent in fragmentation, then we may consider
the initial conditions specified in terms of a fixed flux density per unit
time bombarding a "target" mixture of n-alkanes, where the number of hits on
carbon-carbon bonds depends on their density in the mixture and where only
carbon-carbon bonds are subject to destruction. It follows that if the
volume of the mixture is approximately a constant function of the number of carbon

55

atoms in it, the number of hits will decline with time as bonds are removed. We cannot think of any interpretation of the effects of radiation which would yield a constant number of fragmentations per unit time. Similarly, radiation would seem to be consistent with the proportional fragility model whereby larger molecules represent proportionally larger targets, and inconsistent with the equal vulnerability model. Radiation would also presumably yield equal liklihood of fracture among bond positions in each molecule, and could not be construed as yielding a Gaussian fracture pattern.

Computer Program

The computer program employed to simulate these degradation processes represents the continuous process of degradation as a large number of discrete time units. At the beginning of the program, the starting mixture of n-alkanes is specified in terms of the relative abundance by weight of each of the n-alkanes present in the mixture, and the mixture is assumed to weigh exactly 100 grams. The program computes the number of molecules of each species present in the mixture and all computations are subsequently performed on these numerical quantities. During each time unit the number of fragmentations occurring in each species is computed, the number of daughter products of each length is calculated, and records are appropriately adjusted. After these fragmentation effects are calculated, the program computes the effects of "migration".

Migration Model

The migration model assumes that the local conditions of temperature, pressure and "pourosity" of the surrounding structures are such as to permit the loss of low molecular weight components in ratios determined by their relative vapor pressures. We assume arbitrarily that the duration of the cycle is sufficiently long to permit the loss of all constituents present up to some C_n, and that the loss of constituents of greater molecular weight
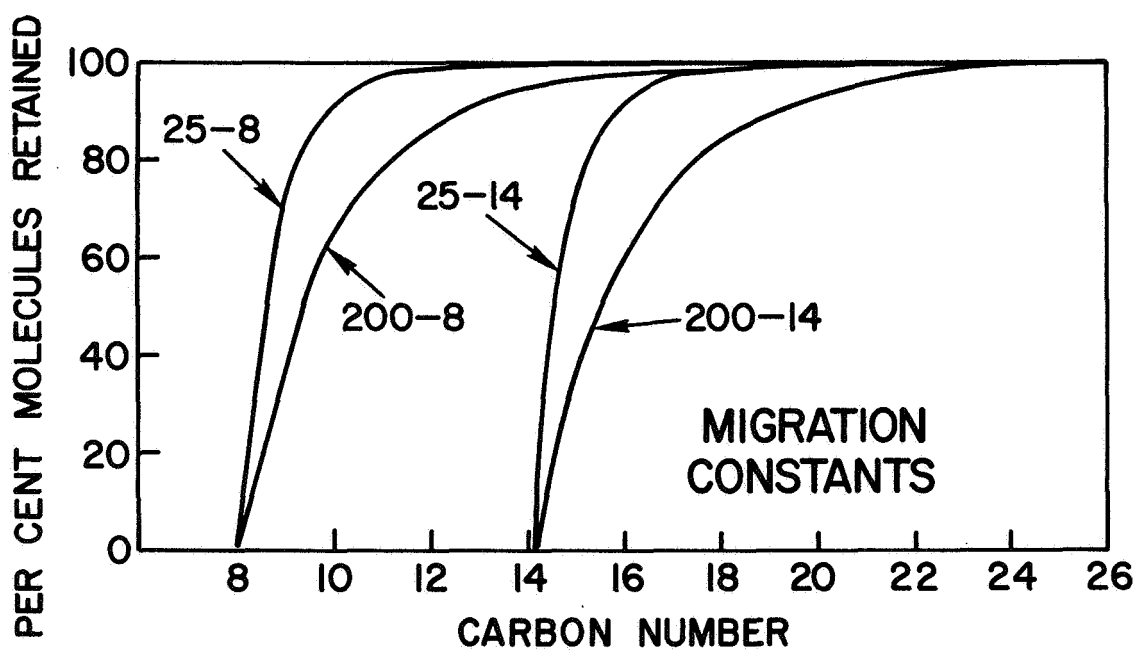
Figure 25. Migration constants indicate the amounts of each species retained after simulated migration.

will depend upon the ratio of their vapor pressures to that of C_n at some arbitrarily selected hypothetical temperature. Thus, the migration constants denoted by the expression "25-8" provide that all n-alkanes up to n-C_8 are lost through migration and that 30.54% of the n-C_9 is lost--since that is the ratio of the vapor pressure of n-C_9 to n-C_8 at $25^{o}C$--9.3% of n-C_{10}, and so forth. The migration constants employed in the runs reported here are plotted in figure 25 . It will be noted that the effect of migration on the pattern of n-alkanes is dependent on the distribution and number of daughte molecules produced in the range in question, which in turn is dependent upon the number and distribution subjected to fragmentation, and the rate and mode of fragmentation.

Sampling

Periodically the mixture subjected to degradation is "sampled" and "chromatographed". This entails plotting the relative abundance by weight of all "recoverable" n-alkanes, where we assume that all components above C_8 are recovered. The weight of the recovered sample and the H' complexity measures are calculated, together with other useful values.

B.   EFFECTS OF DEGRADATION ON THE SHAPES
       OF COMPLEX n-ALKANE MIXTURES

The simulated degradation experiments are defined by four major pairs of options:

    a.   Molecular fragility: molecules of all lengths are equally fragile or fragility is proportional to n-1, the number of carbon-carbon bonds in the molecule.

    b.   Bond fracture patterns: cleavages are equally distributed among all bond positions, so that daughter molecules of all lengths from C_1 to C_{n-1} are produced from species C_n in equal quantities, or breaks are distributed in a Gaussian normal fashion along the length of the molecule.

DEGRADATION WITHOUT MIGRATION

NON-GAUSSIAN PROPORTIONAL FRAGILITY

NON-GAUSSIAN EQUAL FRAGILITY

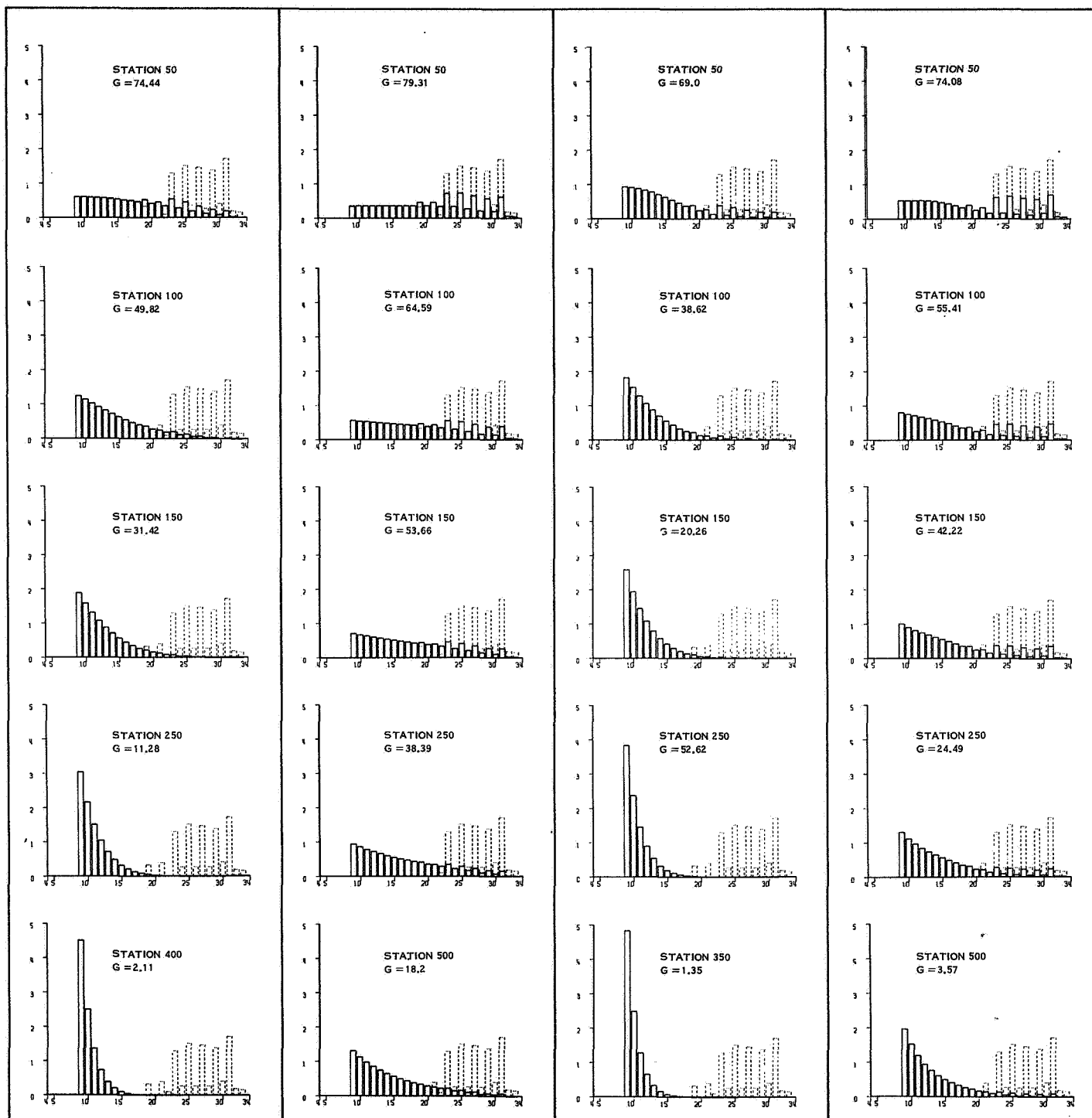GAUSSIAN PROPORTIONAL FRAGILITY

GAUSSIAN EQUAL FRAGILITY

Figure 26. Simulated degradation of Spanish moss n-alkanes without migration. Dashed lines indicate the shape of the starting mixture.

c.   Migration: no migration occurs, or one of four sets of migration
      constants is used (25-8, 25-14, 200-8, 200-14).

d.   Fragmentation rate: the number of fragmentations per unit time is
      a linear function of the number of bonds remaining in the mixture
      so that the yield declines at least exponentially, or a constant
      number are destroyed during each time unit.

The first three options determine the shape transformations that occur
during simulated degradation, the last determines the rate of change of
shape with simulated time.

Degradation without Migration

The differential effects of the first two options on the shape of the dis-
tribution are shown in Figure 26 . In this figure, and all others showing
the simulation results, the dashed lines describe the original distribution
employed, G is the weight of the recovered sample in grams, and the station
number is the number of degradation cycles up to that point in the program.
The plots in Figure 26 show interim degradation mixtures obtained with four
combinations of the first two options. All four experiments employed the
exponential rate model and none employed migration. The starting mixture
is a distribution Spanish moss n-alkanes.

In all four experiments the spikiness concentrated among the high molecular
weight components gradaully erodes with time, the rate of erosion being
faster with the proportional fragility model. The primary difference between
the Gaussian and non-Gaussian bond fracture patterns is that in the former
cases, a greater proportion of the daughter products contain fewer than 9
carbon atoms, and hence are not "recoverable". As a result the weight declines
more rapidly when the Gaussian model is used and the remaining high molecular
weight components represent a slightly larger proportion of the recovered
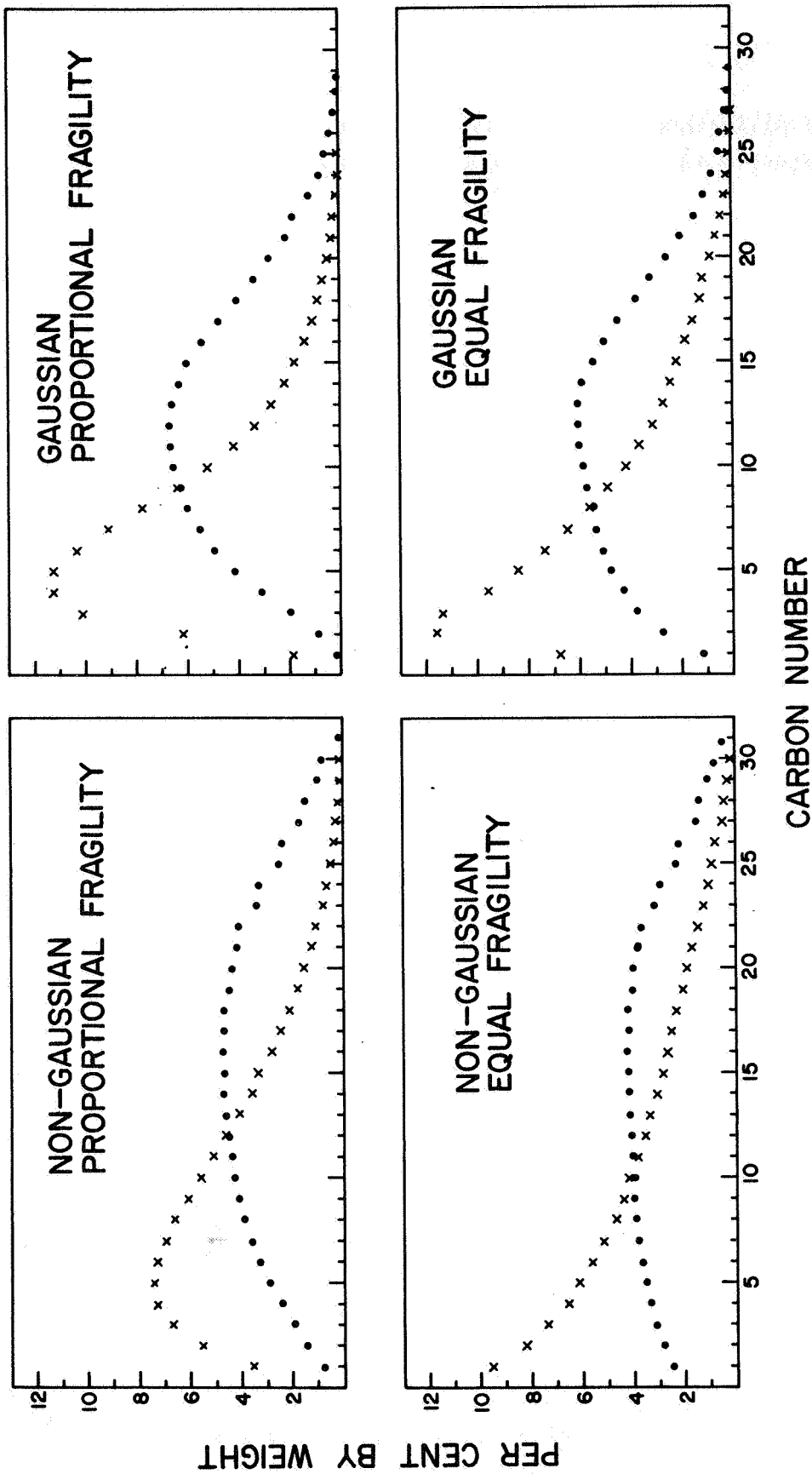sample. Consequently, erosion of the spikiness proceeds slightly less rapidly.

58

Figure 27. Product yields during the first 50 cycles and during the next 100 cycles for all four simulated experiments shown in Figure 26.

A slight hump in the fragments is also observable in the early stages of both these Gaussian experiments.

These differences in the patterns of fragment distribution are more clearly illustrated in Figure 27 where the relative weights of the fragments produced during the first 50 stations and the next 100 stations are plotted for each of the four experiments. Thus, with the non-Gaussian fracture pattern, fewer low molecular weight products are produced during the first 50 stations with the proportional fragility model than with the equal fragility model because more of the fragments are derived from fragmentation of high molecular weight products. During later stages of degradation the peak product yield shifts toward the left, reflecting the fact that fewer of the high molecular weight parent molecules remain. Most of the fragments produced by the Gaussian fracture pattern are concentrated in the range below n-C_9; this concentration increases with time.

Effects of Migration

Migration as modeled in these experiments profoundly affects the shapes of the degraded mixtures. By removing low molecular weight components the migration option smooths the left end of the distribution and reduces the total weight of the recovered mixture. This tends to concentrate fragmentations among the high molecular weight components not subject to migration, but this effect is partially balanced by the reduction of the total weight of the mixture so that these components tend to represent a relatively large percentage of the total recovered sample.

Non-Gaussian Models. Figure 28 shows distributions obtained during degradation of Spanish moss using the straight radiation model (proportional fragility of molecules, non-Gaussian fracture pattern) in conjunction with the four sets of migration constants. The balancing effect is clearly illustrated here. The greater the quantity of material removed by migration, the slower the apparent rate of decline of spikiness. This results from the fact that even though the

DEGRADATION WITH MIGRATION
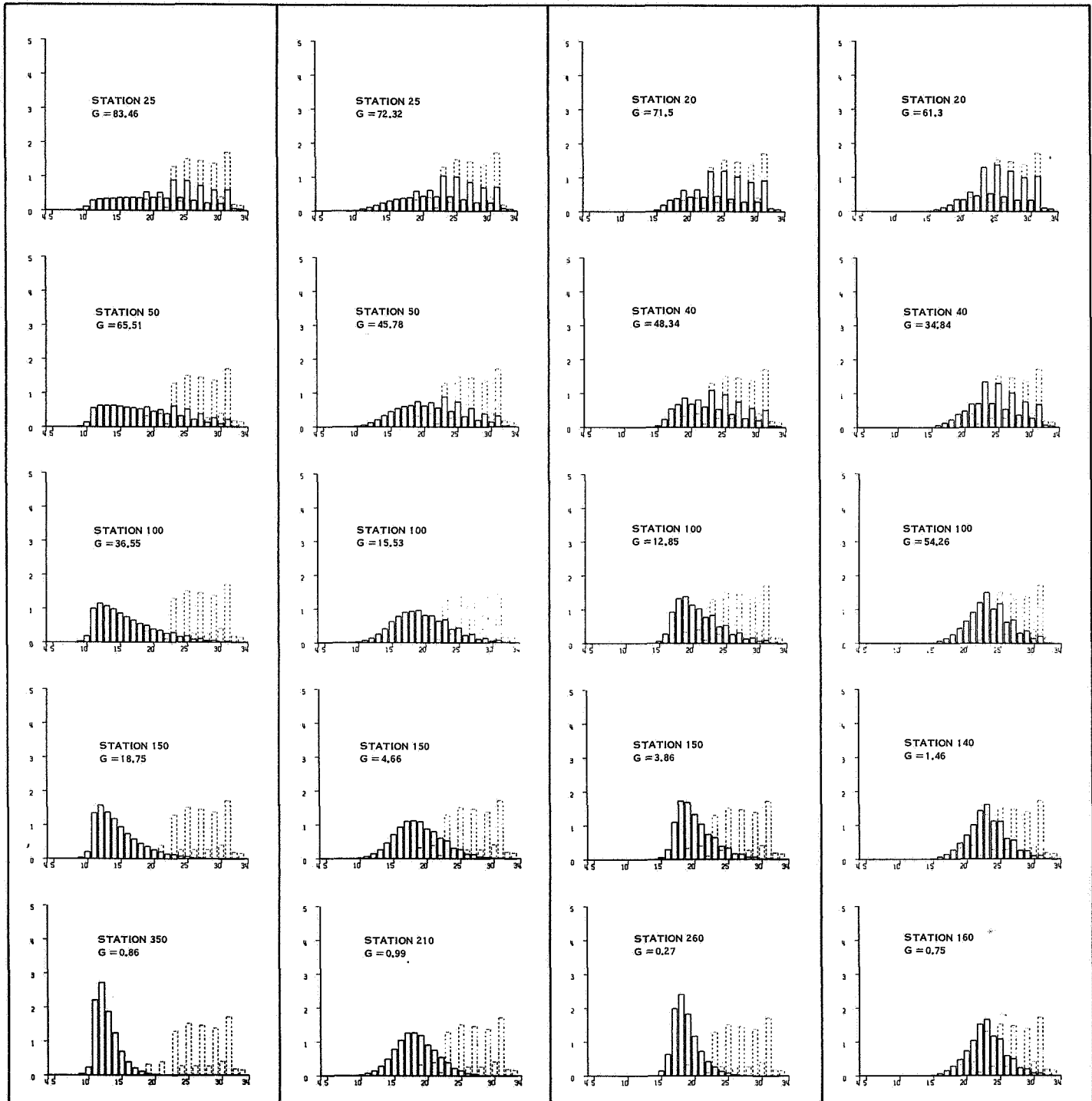
NON - GAUSSIAN, PROPORTIONAL FRAGILITY

Figure 28. Simulated degradation of Spanish moss with migration and non-Gaussian proportional fragility models.

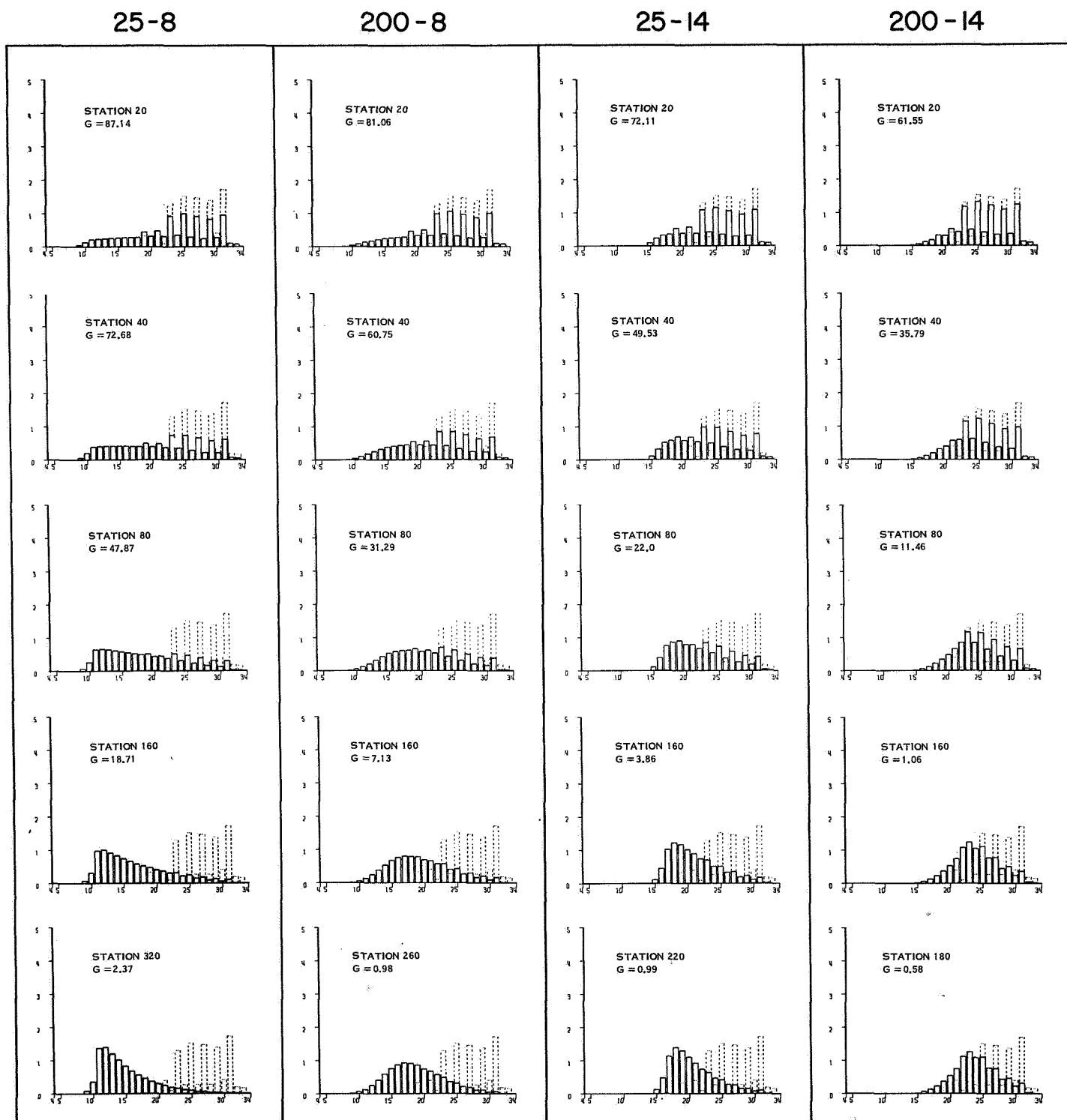# DEGRADATION WITH MIGRATION

NON-GAUSSIAN, EQUAL FRAGILITY

Figure 29. Simulated degradation of Spanish moss with migration and non-Gaussian equal fragility models.

high molecular weight components account for a greater percentage of the mixture at any given time, the total number of bonds present and hence the number of fragmentations declines with the weight of the sample.

The first four rows of Figure 28 show the interim patterns at approximately equal points in the degradation sequence. Normally we stopped degradation when the mixture had declined to 1 gram (1% by weight of the starting mixture) but three of these experiments were continued much longer to see whether some of the very short smooth distributions shown in the age-dated series of Figure 13 of this report could be obtained. Inspection of the first three figures of row five suggests that this is the case. In particular, the migration constants 25-14 (column 3) might have resulted in distributions analogous to the three very old Soudan shale distributions had the degradation program continued for a very much longer time.

Comparison of Figure 28 with Figure 13 shows that this degradation model can transform a very spiky and complex Spanish moss distribution into patterns strongly reminiscent of many of those found in the age-dated series.

Very similar results occur when the non-Gaussian fracture pattern is combined with the equal fragility option and the four sets of migration constants. These are shown in Figure 29. As might be expected, employment of the equal fragility option whereby molecules of all lengths are equally vulnerable to fragmentation, (e.g., equal percentages of each species are fragmented during any time unit) results in a slightly slower decline in spikiness per unit time. Otherwise, the results are much the same as those shown in Figure 28.

Gaussian Models. The effects of migration on the Gaussian models are shown in Figures 30 and 31. The overall shape transformations are very similar to those obtained with the non-Gaussian fracture pattern, but the erosion of spikiness does not progress as rapidly or as far as in the former case because this pattern produces relatively few products in the range $n\text{-}C_{20}$ to $n\text{-}C_{30}$ where the starting material exhibits a pronounced unevenness. With this model, unevenness is eroded primarily by the loss through fragmentation of the components with odd numbers of carbon atoms. Consequently, the erosion is effected only when the

DEGRADATION WITH MIGRATION

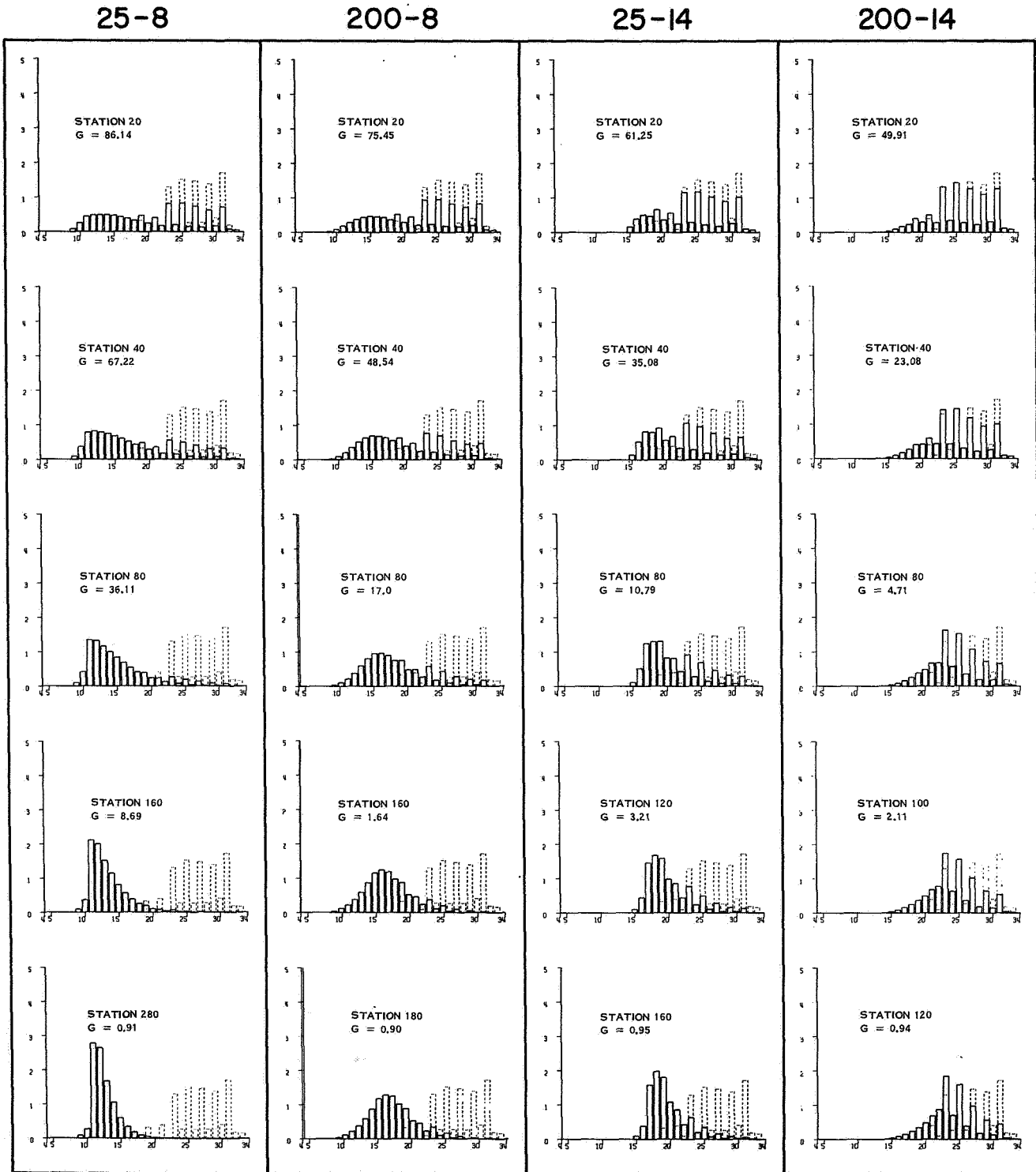## GAUSSIAN, PROPORTIONAL FRAGILITY

Figure 30. Simulated degradation of Spanish moss with migration and Gaussian proportional fragility models.

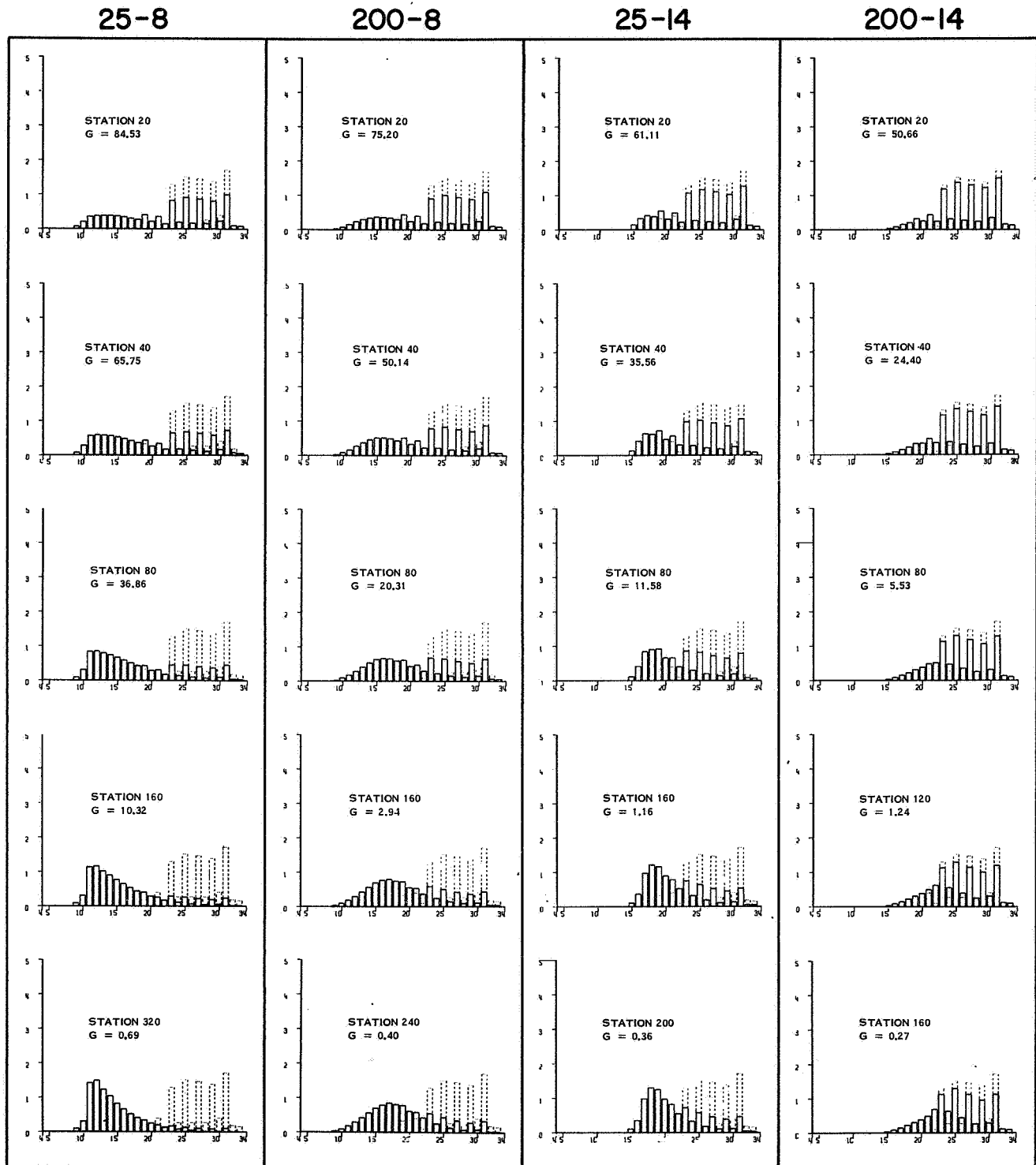# DEGRADATION WITH MIGRATION

GAUSSIAN, EQUAL FRAGILITY

Figure 31. Simulated degradation of Spanish moss with migration and
Gaussian equal fragility models.

simulated experiment lasts long enough for the appropriate number of cleavages to occur. This results when weight loss through migration is minimal, as for example, when the migration constants 25-8 and 200-8 are used.

These four figures suggest that almost any mode of random fragmentation can eventually transform a complex n-alkane mixture into mixtures that mimic the visual features of distributions found in terrestrial deposits, provided migration is permitted.

Degradation of Other Complex Mixtures

In order to see whether less complex starting mixtures could be transformed into shapes characteristic of the most ancient terrestrial n-alkane distributions, we subjected three additional starting mixtures to degradation using the non-Gaussian fracture models. This also permitted us to study rate of change of complexity as a function of starting mixtures.

The starting mixtures employed were Eocene Crude (Figure 13d) Cambrian Crude (Figure 13n) and the n-alkanes from a recent marine sediment (Kenvolden (7)). Figure 32 shows some of the results. Column one shows one experiment run with the equal fragility model. The experiments illustrated in the remaining columns used the proportional fragility model. Visual differences in the starting materials are almost entirely obliterated, but the very short distributions were not obtained.

Degradation of $n-C_{28}$

Henderson, et. al. (14) subjected $n-C_{28}$ to pyrolysis at $375^{\circ}C$. Column one of Figure 33 contains plots of the relative abundance of n-alkanes observed after 75 and 150 hours of pyrolysis. After 75 hours, 89% of the mixture consists of $n-C_{28}$ and 11% of components in the range $n-C_{13}$ to $n-C_{27}$. Seventy-five hours later, $n-C_{28}$ represents only about 8% of the total mixture and the products in the range $n-C_{13}$ to $n-C_{27}$ have assumed a definite smooth unimodal distribution reminiscent of n-alkanes from ancient deposits.

Column 2 shows simulated degradation of $n-C_{28}$ when all products were assumed to remain in the reaction vessel during the reaction (i.e., no migration) and no
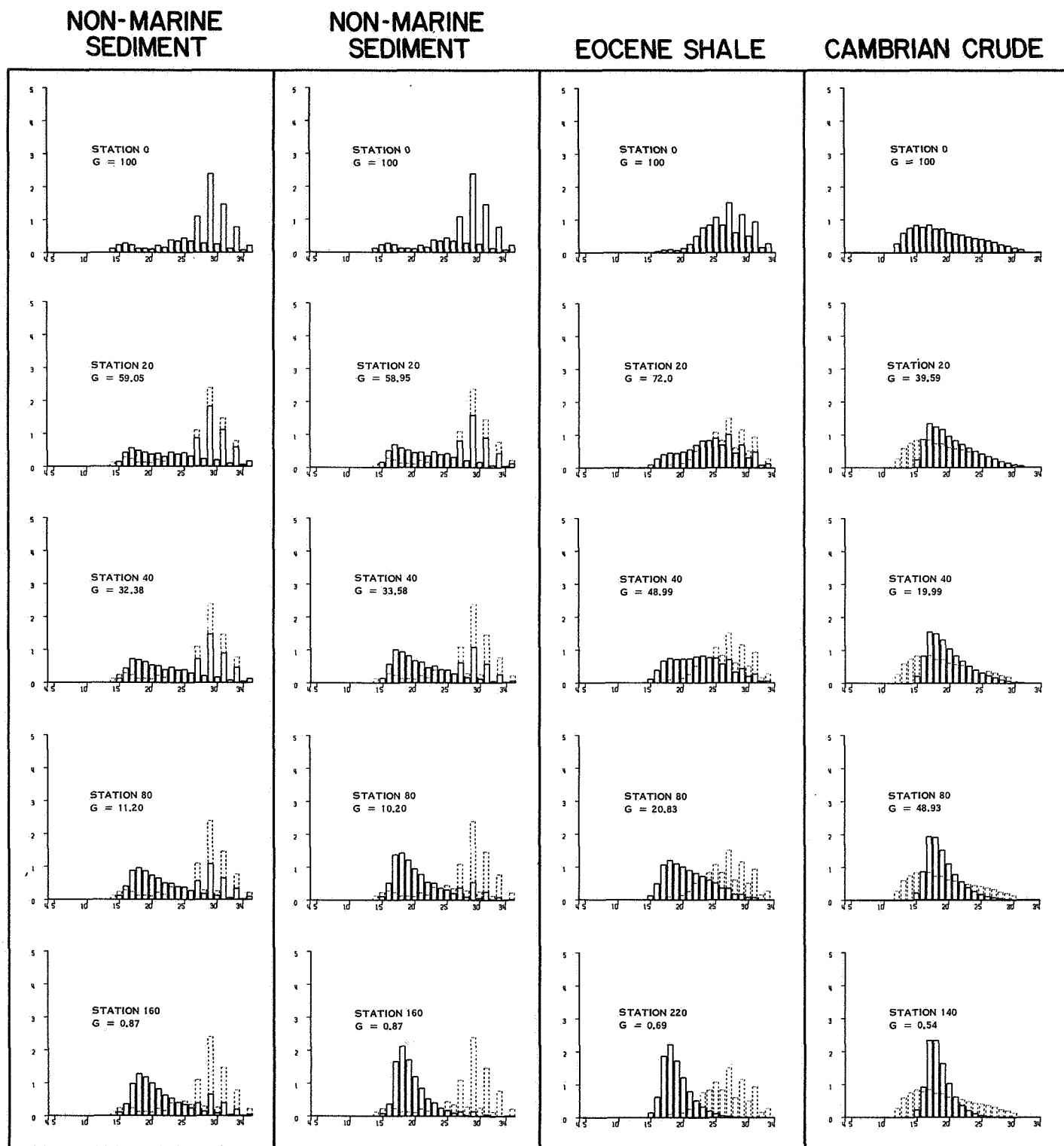
DEGRADATION OF DIVERSE MIXTURES

Figure 32. Degradation of other complex mixtures with non-Gaussian equal fragility (column one) and proportional fragility models (columns two, three and four).

components were "lost" during work up. All runs employ the proportional fragility option.

The first three plots of column two show interim stages of a simulated experiment employing the Gaussian fracture pattern; the last two plots use the non-Gaussian fracture pattern. In both cases, when the n-C_{28} has declined to a low abundance, the product distribution shows the characteristic increase in relative abundance of lower molecular weight products due to secondary degradation of the high molecular weight products.

It is not known how many low molecular weight products were lost by Henderson during work up (14), but examination of column 2 suggests that evaporation of products in the range n-C_{10} to n-C_{16} could transform the simulated mixtures into patterns similar to that observed after 150 hours of pyrolysis.

Our simulation program permits us to simulate "evaporation" of products during "work up" by eliminating from the mix specified percentages of low molecular weight products. We employ the migration constants 25-14 to indicate the pattern transformations which would result if all products up to n-C_{14} were completely lost and every higher molecular weight component is reduced by an amount proportional to the ratio of its vapour pressure at 25°C to that of n-C_{14}. That is, we use the migration constants 25-14 plotted in Figure 25.

The results of these simulated evaporation experiments are shown in columns three (Gaussian and four (non-Gaussian). It will be observed that while these evaporation constants do not yield a result strongly similar to Henderson's, evaporation at a higher assumed temperature would be very likely to yield simulated degradated mixtures very similar to the 150 hour mixture.

The discrepancies between the observed and simulated n-C_{28} pyrolysis experiment are suggestive. It would seem that with different evaporation constants, the non-Gaussian model would yield patterns that strongly mimic the observed results if fragility were more strongly dependent on n (if, for example, it varied with 2n), and, perhaps, if n-C_{27} and n-C_{26} yields different slightly from the yields of the remaining daughter products. In any case, it would seem that simulation provides a useful tool for exploring testable hypotheses regarding the production of n-alkanes from other n-alkanes during pyrolysis.
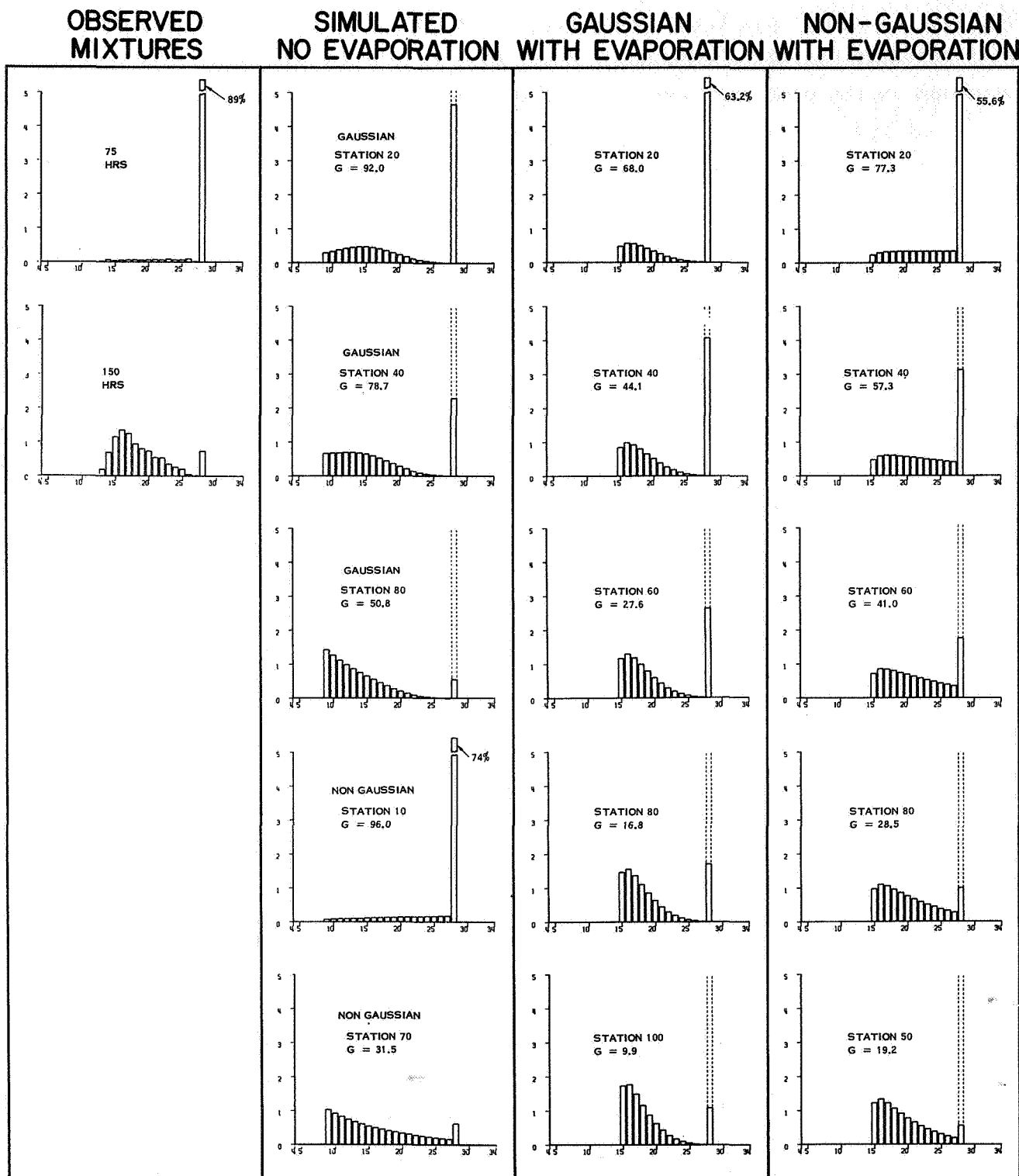
PYROLYSIS OF N-C$_{28}$

Figure 33. Degradation of n-C$_{28}$: column one, observed by Henderson, et. al. (14); columns two through four, simulated.

C.   COMPLEXITY CHANGES DURING SIMULATED DEGRADATION

We have shown that all the degradation models employed can, when combined
with suitable migration constants, mimic most of the n-alkane distributions
observed in the geological series in Figure 13. Although these models
represent only a few of the many "random" degradation models that could be
explored, the results presented above tend to suggest that almost any
pattern of n-alkane destruction that is accompanied by transformation
of at least some of the fragments produced into new n-alkanes will produce
patterns visually similar to those seen in progressively older terrestrial
deposits from almost any starting distribution provided suitable migration
constants are employed. It is therefore tempting to suppose that comparable
forms of degradation have played a major role in producing the n-alkane
mixtures present in terrestrial deposits from precursor n-alkanes.

If this is the case, then we should expect that the simulated degradation
experiments produce a decline in measured complexity that is consistent
with the relationship between complexity and age demonstrated in Part II
above.

Observed Complexity Changes

Figures 34 through 36 show the complexity changes and the decline in the
weight of recorded sample observed in the runs illustrated in Figures 28
through 32. All of these illustrations are for runs employing the declining
fragmentation rate rather than the constant rate, for the simple reason that
the constant rate resulted in a rapid and dramatic accelerating decline in
complexity with model time, completely inconstent with the results observed.

Figures 34 shows the results observed with the non-Gaussian fracture model.
A comparison of these plots shows that the overall decline in complexity is
approximately linear for most of the simulated time for all 8 runs, the
primary differences being that for equivalent migration constants, the model

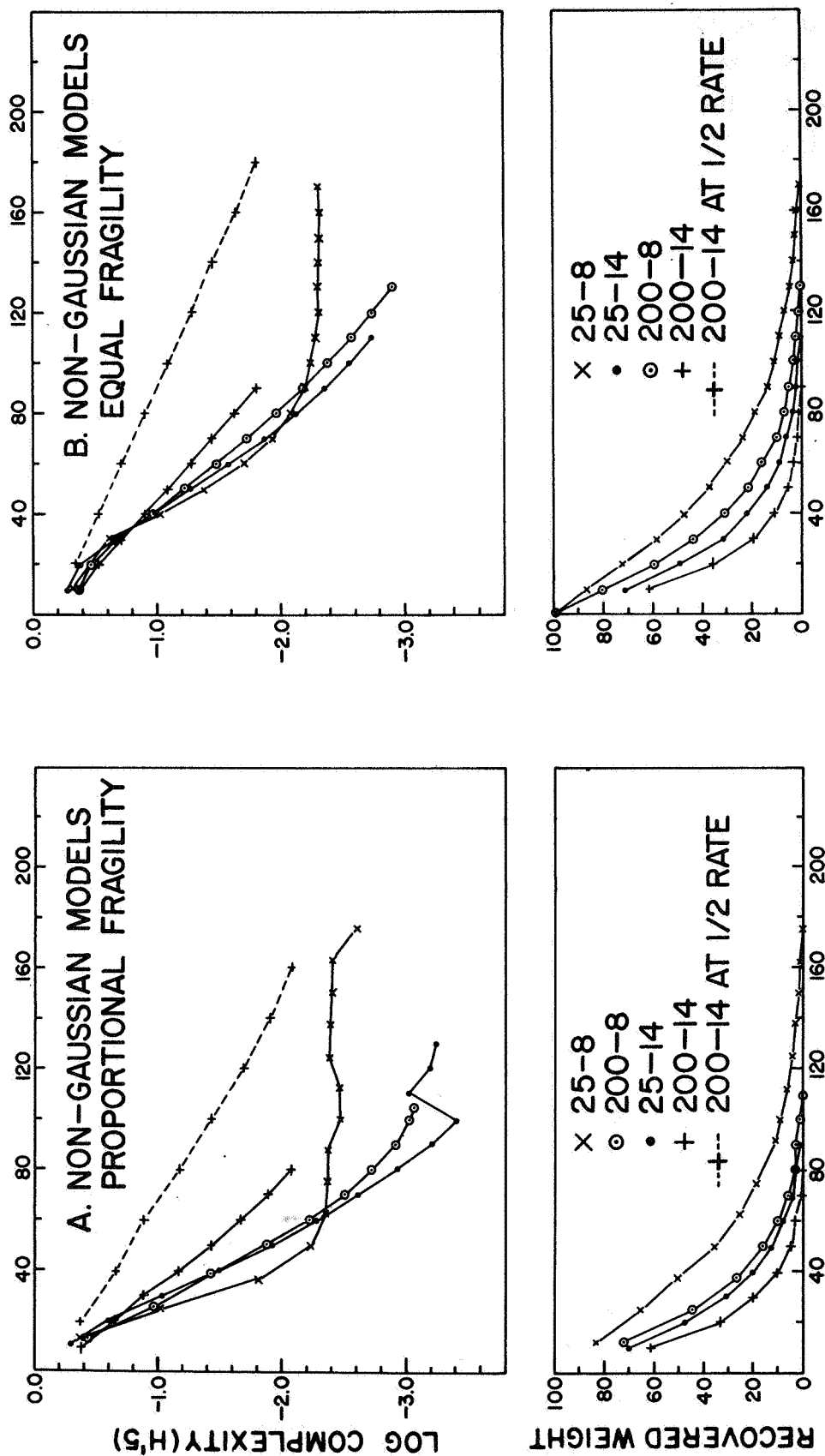# COMPLEXITY CHANGES DURING DEGRADATION WITH MIGRATION

Figure 34. Changes in complexity and weight of recovered sample with simulated time during non-Gaussian degradation.

with proportional fragility tends to result in somewhat lower complexity values than the equal fragility model. The complexity values on each plot are quite similar, which shows that the migration constants do not greatly affect the complexity values, although they do have major effects on the time required for the mixture to decline to 1 gram or less.

In all these experiments the fragmentation rate (number of fragmentations per unit time) was set at 8% of the bonds present in the mixture. Changes in the rate merely result in changing the X axis. Thus, in Figure 32 the dotted line indicates the complexity curve which would result if the experiment using migration constants 200-14 had been run with a fragmentation rate of 4%.

All these runs used the same starting material -- the distribution of n-alkanes observed in a sample of Spanish moss with a high degree of complexity $(H'_5 = .518)$ and not all of them achieved the range of log complexity values observed in the age-dated series (-0.2 to -3.48). Figure 34a has two curves which show a somewhat erratic behaviour after station 100. This is due to a reduction in the number of components in the distribution which occurred when the highest molecular weight components simply dissappeared entirely. In the experiment using 25-8 this occurred when the mixture had declined to about 10 grams. In 25-14 it occurred when the mixture declined to slightly less than 1 gram. This run was continued in order to attempt to get the mixture to lose a large number of components to see if the shapes characteristic of the very old distributions could be obtained. The experiment was stopped at station 130 when the weight was .26 grams. The shape changed only very slightly during this interval.

The complexity changes observed with the Gaussian models are shown in Figure 35. Here we find that the differences between the molecular fragility options are of the same kind but greater in magnitude than was the case for the

64

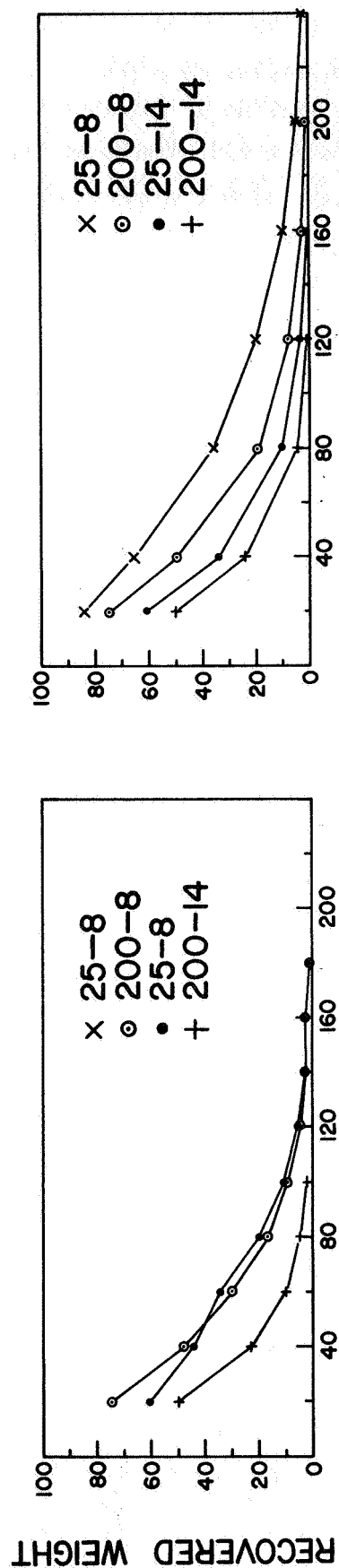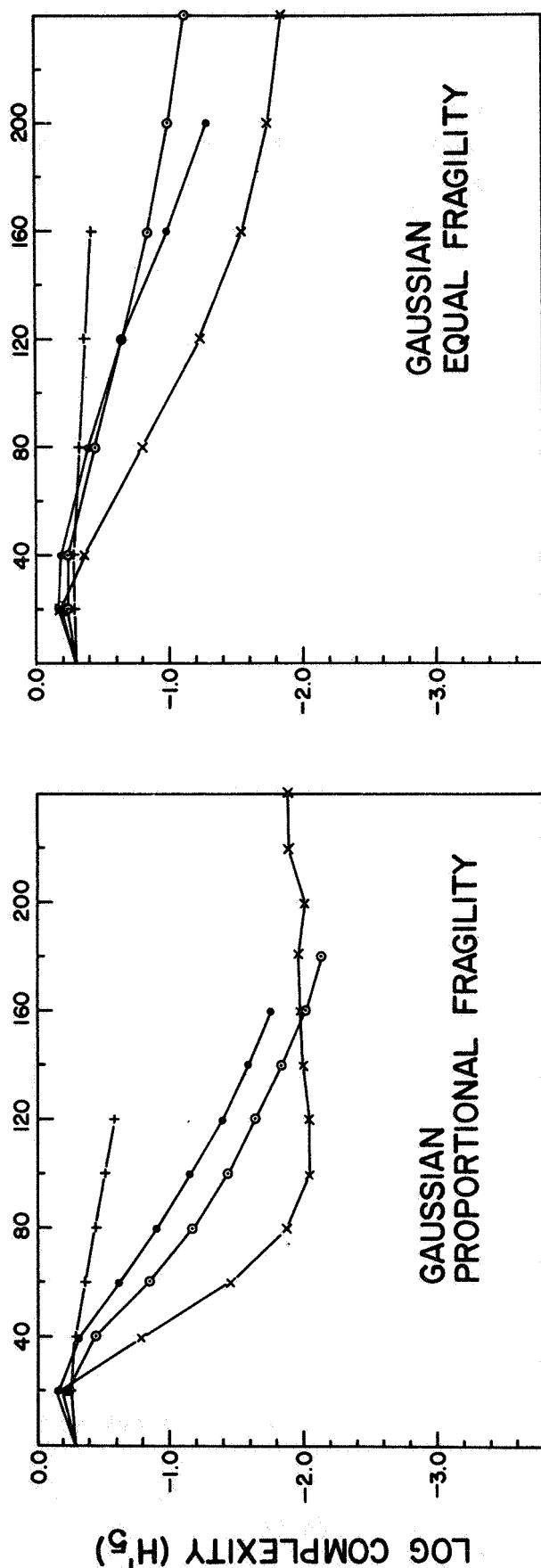# COMPLEXITY CHANGES DURING DEGRADATION

GAUSSIAN MODELS

Figure 35. Changes in complexity and weight of recovered sample with simulated time during Gaussian degradation.

non-Guassian models. Thus, if we compare the 25-14 and 25-8 experiments in both plots, we find the proportional fragility model resulting in a faster decline in complexity to lower values than the corresponding experiments with the equal fragility model. In both cases the migration constants have a greater effect on the spread of the complexity values than is the case for the non-Gaussian fracture pattern. The complexity value did not decline to as low values in these runs as in the runs using the non-Gaussian models.

The Gaussian model used here employed a standard deviation about 1/6 the length of the molecule, resulting in a relatively narrow spread of product molecules. If a much larger standard deviation had been used, the result would be greater smoothness in the distributions, in a slower decline of weight.

Figure 36 shows the decline in complexity observed with different starting materials using the non-Gaussian models and migration constants 25-14. Both experiments using Cambrian Crude as a starting mixture quickly degenerated to a very short distribution whose complexity measure behaved somewhat eratically.* When the proportional fragility option is employed the rate of change of log complexity with time is nearly independent of the complexity or shape of the starting mixture. When the equal fragility option is employed, rate of change is also influenced by the shape of the starting mixture. Thus, the sediment exhibits a very slow complexity decline with time while the Spanish moss declines relatively rapidly, though these two starting mixtures are of approximately equal complexity.

Discussion

The random degradation models presented here are not proposed as models of

---

* This is partly an artifact of the complexity measure due to the need to represent the n-alkane mixture as percent abundance. Consequently, the variance of smooth unimodal distribution can become very large when the number of frequency components declines.

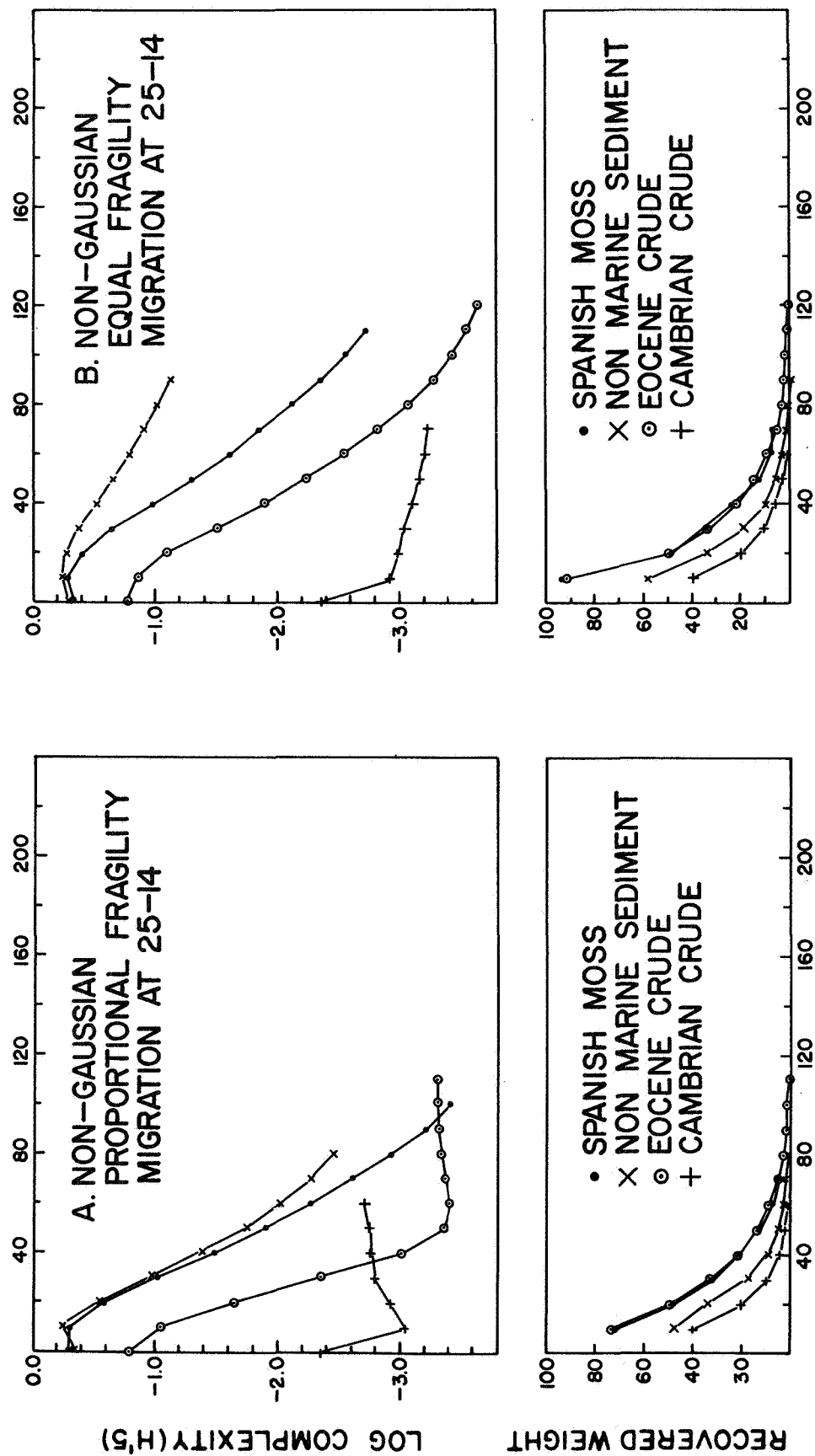# COMPLEXITY CHANGES OF DIVERSE STARTING MIXTURES

Figure 36. Changes in complexity and weight of recovered sample using different starting materials and non-Gaussian fracture pattern with migration constants 25-14.

the effects of physical degradation processes acting on biogenic material deposited in the terrestrial environment. However, the fact that they produce patterns that mimic the patterns of these n-alkane distributions with striking success naturally leads to speculations regarding the possibility that all the age-related differences observed in the former are due to degradation processes similar in their effects to the kinds we have simulated. In short, it is tempting to interpret these data as implying that all of these ancient biogenic hydrocarbons, when initially deposited, consisted of materials comparable in complexity to modern land plants and animals, and that all of the age-related differences now observed among them are due to differences in the duration of the degradation processes.

Several features of the simulation results suggest that this is not likely.

The first is the relationship between measured complexity and modeled time. In the age-dated series, we found that complexity declines approximately linearly with time, whereas in the simulated studies it declines exponentially with time at least for most of the period observed. Accomodating these results to the observed complexity-age relationship would require the assumption that the physical processes responsible for degradation have changed in such a fashion as to result in a decline in degradation rate by about 2 orders of magnitude in the interval in question (50 to 2700 million years before present), throughout al the environments from which these materials were obtained.

The second somewhat inconsistent feature is that rate of decline in complexity is strongly dependent on assumed fragmentation rate. Slight differences in fragmentation rate could be expected to result in exceedingly large differences in the observed complexity after a short time, if complexity declines exponentially with time. Alternatively, if complexity declines linearly with time, differences in degradation rate due to long duration differences in the local environments would tend to result in an increase with age in the spread of the observed complexity values.

66

There is some suggestion in Figure 14 that this occurs up to the midpoint of the
Paleozoic. (It may be, of course, that the long-term average fragmentation
rate does not differ greatly among the geologic environments in which hydro-
carbons are now found--that, for example the extreme differences that
occurred were of short duration interspered by long periods during which
very little degradation occurred.)

It will be noted that in many cases complexity declines to a low value and
remains at that value for a relatively long time, although sometimes the
minimum is followed by a period of eratic increase in complexity.

In view of these inconsistencies in the complexity-time relationships, we
are inclined to conclude that while degradation probably played some role in
modifying the shapes of n-alkane distributions, it does not seem likely
that it is entirely responsible for all the time related differences.

Thus, even if we assume that alternative forms of degradation (fragmentation)
can yield a decline in complexity which is linear with time rather than ex-
ponential, it would seem that differences in the geological environments,
starting materials, and degree of biological degradation would be so great
as to result in a very large variance about the regression line - indeed,
we would expect to see little if any linear correlation through the "noise"
of the environmentally induced variation.

In other words, if we take the plot of Figure 15 as representing a random
sampling of environments as well as ages, and if we assume that (1) complexity
declines linearly with age at a rate which varies with the environmental con-
ditions and (2) starting materials showed a complexity range characteristic
of that of modern land plants, then we must conclude that environmental
conditions were virtually identical for all localities represented and the
degradation induced rate of change of complexity was very slight. Alternatively,
if we suppose that part of the variance observed in this plot is due to differ-
ences in local environments than we must conclude that the total degradation

effect is small, and consequently that the shape differences are partly due to differences in the complexity of the initial starting material.

Further studies are needed to determine the probable role of degradation as a factor influencing shape change. These should include (a) identification of degradation modes which result in a linear decline in complexity with time (b) studies correlating complexity of ancient materials with known differences in thermal and/or radiation environment and (c) further studies of the relationship between phylogenetic age and complexity, and of the role of biological agents as degraders of pattern complexity.

D.   SUMMARY OF CONCLUSIONS

1.   Patterns strikingly similar to those observed in the age-dated series can be produced from a highly complex starting mixture by all the degradation models simulated, provided migration is permitted.

2.   Patterns reminiscent of those observed during pyrolysis of Spanish moss n-alkanes can be produced from all degradation models without migration.

3.   Some patterns similar to those observed during pyrolysis of $n-C_{28}$ can be produced by the non-Gaussian models combined with evaporation of low molecular weight products.

4.   Complexity declines exponentially with simulated time when migration is permitted for most of the degradation models studied. The migration constants affected the rate of decline only slightly with the linear fracture models; with the Gaussian fracture models, the rate of decline was greatly influenced by the migration constants.

5.   It seems unlikely that all the age-related shape differences observed are the result of degradation of starting material similar in complexity

to modern materials by processes with results analogous to those modeled. If degradation processes do have similar effects, we must conclude from the relatively small variance of the plot of complexity versus age that degradation effects on complexity are relatively small and consequently that complexity of biogenic material has increased with evolutionary time.

NOTES

1. We do not in fact invoke these assumptions in any essential fashion
    because we interpret the results of the analyses performed as applying
    only to the pattern of the envelope of the n-alkanes actually observed.
    In particular, we do not use these results to "predict" the concentrations
    of very low or very high molecular weight alkanes that may be present in
    the mixture but not observed on account of the limitations of the analy-
    tical techniques employed. Similarly, while we make use of the concept
    of the continuous envelope of the n-alkane distribution, we do not suppose
    that an interpolated envelope bears any necessary relationship to the
    distribution of branched chains or other hydrocarbons whose peaks frequentl
    occur on a chromatogram between those corresponding to adjacent n-alkanes.

    This descriptive use of Fourier techniques contrasts with more conventiona
    applications in which the validity of the assumptions is very critical in
    that the characteristics observed in a short sample of a continuous signal
    are interpreted as describing unobserved segments of that signal. In
    contrast, our "assumptions" are merely metaphors which simplify the
    transition to the convenient terminology employed in such predictive
    applications.

2. By definition an aperiodic signal possesses a continuous power density
    spectrum and is to be distinguished from a strictly periodic signal,
    which has a power density function which is a line spectrum. The differ-
    ence is important only when one is considering continuous time varying
    functions, and in this case it is usually possible to distinguish the two
    in practise. In this descriptive application we are not in fact making
    any assumptions at all, but merely describing a transformation used to
    compare features of patterns of n-alkane distributions containing differing
    numbers of n-alkanes. The transformation is the same one that could be
    validly applied to predict the spectral distributions of real time-varying

signals which are aperiodic and continuous and whose distribution of
power within each of the original k segments of the frequency spectrum
is in fact flat, as suggested by the bar plot presentation of the power
density spectrum. If such a signal exists, then sampling it at $2K'$ (where $K' \langle K$)
equally spaced intervals would reveal a spectral distribution identical
with the one obtained by the transformation technique we employ. In
both cases, the resulting spectral approximation is a predictable
function of the original N sample points, and it represents a loss
of information in that the values of the original N points is not
a unique function of the value of the transform.

3.   Very uneven distributions containing one or two very large components
constituting fifty percent or more of the total are perceived as more
uneven than the H' measure reflects. Similarly, short unimodal dis-
tributions with a few components tend to be perceived as more smooth
than they are ranked by their measure.

We believe that further development of the measure and exploration of
factors involved in subjective judgements would result in a more useful
measure of complexity than the H' functions presented here.

LITERATURE CITED

(1)   Calvin, M. "Chemical evolution", Proceedings of the Royal Scoiety,
       Series A, Vol. 288 pp. 441 to 446, November 30, 1965.

(2)   Eglington, G., and M. Calvin. "Chemical fossils", Scientific American,
       January 1967, pp. 32-43.

(3)   Clark, C., Occurrence of Normal Paraffin Hydorcarbons in Nature,
       Technical Report, Reference Number 66-34, Woods Hole Oceanographic
       Institution, Woods Hole, Massachusetts, 1966.   UNPUBLISHED MANUSCRIPT

(4)   Welte, D. H., "Nichtfluchtige Kohlenwasserstaffe in Kernprohen des
       Devons und Karbons der Bohrung Munsterland 1", Fortshr. Geol.
       Rheinld. W. Westf. 12, 559-568.

(5)   Belsky, T., R. B. Johns, E. K. McCarthy, A. L. Burlingame, W. Richter
       and M. Calvin, "Evidence of life processes in a sediment two and a
       billion years old", Nature, Vol. 206, May 1, 1965, pp. 446-447.

(6)   Kulp, Lawrence, J. "Geologic time scale", Science, Vol. 133, Ap. 14,
       1961, pp. 1105-1114.

(7)   Kenvolden, Keith A., "Molecular distributions of normal fatty acids
       and paraffins in some lower Cretaceous sediments", Nature, Vol. 209,
       Number 5023, Feb.2, 1966, pp. 573-577.

(8)   Barghoorn, E. A., W. G. Meinschein, J. S. Schopf., "Paleobiology of a
       precambrian shale", Science, Vol. 148, April 23, 1965, pp. 461-472.

(9)   Douglas, A. G. and G. Eglington, "The distribution of alkanes", Chapter
       4 of Comparative Phytochemistry, T. Swain, ed. Academic Press New York,
       pp. 57-77.

(10)   Cooper, J. E. and E. E. Bray, "A postulated role of fatty acids in
       petroleum formation",   Geochimica et Cosmochimica Acta, Vol 27,
       pp. 1113 to 1127, 1963.

(11)   Oro, J., D. W. Nooner, and S. A. Wilkstrom, "Parafinic hydrocarbons
         in pasture plants", Science Vol. 147, pp. 870-873, February 1965

(12)   Oro, J., D. W. Nooner, A. Zlatski, S. A. Wikstrom, and E. S. Barghoorn
         "Hydrocarbons of biological origin in sediments about two
         billion years old", Science Vol. 148, April 2, 1965, pp. 77-79.

(13)   Simmonds, P. G., Ph.D. Thesis, University of Houston, June 1966.

(14)   Henderson, W., J. Eglington, P. Simmonds, and J. E. Lovelock,
         "Thermal Alteration as a Contributory Process to the Genesis
         of Petroleum", Nature Vol. 219, Sept. 7, 1968, pp. 1012-1016.

(15)   Hitchcock, D. R. and Lovelock, J. E., "Life detection by atmospheric
         analysis", Icarus, Vol. 7, No. 2, Sept. 1967, pp. 149-159.

(16)   Tomabene, T., and J. Oro, Science Vol. 150 (1965), p. 1048.

(17)   Oro, J., D. W. Nooner, A. Zlatkis, and S. A. Wilkstrom, Life Sciences
         and Space Research, 4 63 (1966).

(18)   Korn, Granio A. and T. A. Korn, Mathematical Handbook for Scientists
         and Engineers, McGraw Hill, Inc., New York 1961, pp. 659-662.

(19)   Ibid. p. 135

(20)   Snedecor, G. W., and W. G. Cochran, Statistical Methods, Iowa State
         University Press. Ames Iowa, 1967, pp. 135-158.

(21)   National Bureau of Standards, Handbook of Mathematical Functions,
         U. S. G. P. O. Washington, D. C., June 1964, p. 947.

(22)   Snedecor and Cochran, op. cit. pp. 172-190.

(23)   <u>Ibid</u>. p. 272.

(24)   Siegal, Sidney, "Nonparametric Statistics for the Behavioral Sciences",
         McGraw-Hill, Inc., New York, 1956, pp. 185-193.

ACKNOWLEDGEMENT