

191

FINDING THE IDEAL CYCLOTHEM

by

655
WILLIAM C. PEARN

B. S., University of Kansas, 1954
M. S., University of Kansas, 1959
Ph.D., University of Kansas, 1963

6
A MASTER'S REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

Approved by:

Leslie F. Marcus
Major Professor

LD
2668
R4
1964
P 359
c.2

TABLE OF CONTENTS

	page
INTRODUCTION.	1
Acknowledgements	3
SIMPLIFYING THE MOORE IDEAL CYCLOTHEM	4
THE DISCORDANCE INDEX, G.	7
FORMULATING THE QUESTIONS	9
GENERATING THE POPULATION OF SEVEN-UNIT SEQUENCES	9
A POPULATION OF ALTERNATIVE IDEALS.	11
VARIATION OF G IN AN IDEALIZED COMPOSITE SECTION.	15
DISTRIBUTION OF G IN A SAMPLE OF ROCK SEQUENCES	22
PROBLEMS OF CLASSIFICATION.	31
Fusulinid Requirement.	32
Inclusiveness of Unit-3.	33
Inclusiveness of Unit-2.	33
Degree of Clasticity Required for Unit-1	33
Thickness.	34
Generalizations and Directions	34
CONCLUSIONS	36
REFERENCES.	37
APPENDIX A, Programs Used	38
Program 1.	39
Program 2.	44
Program 3.	48
Program 4.	54
APPENDIX B, An Alternative to MAG	59

INTRODUCTION

The general repetitive nature of Pennsylvanian and Permian rock sequences is well known and is particularly striking in parts of the midcontinent region. It seems that the concept of cyclic sedimentation is well established, at least within this region and stratigraphic interval.

Throughout geologic history epicontinental seas have repeatedly inundated the present land masses. In virtually every locality where a sedimentary sequence exists at all, it contains a record of the transgressions and regressions of such seas. It is natural to equate depositional cycles with marine oscillations. However, the large number of oscillations seemingly required for deposition of the Pennsylvanian cyclothems has led to speculation that physical transgression and regression may not have been involved in each individual "cycle". Throughout this paper, the terms transgression and regression should be understood to stand symbolically for whatever mechanism may actually have been operative.

A portion of the investigation reported here was designed to question the existence of underlying mechanisms governing the nature of repetitive sedimentary sequences. This was done only as a logical step in the development of the methods employed. It may be assumed with confidence that such mechanisms exist. Speculation concerning the nature of these mechanisms is interesting, but conclusions are difficult if not impossible to prove.

More practical questions concern the nature of the repetitive record itself. For instance, in a given region and within a given stratigraphic interval, what lithologic units constitute the ideal cyclothem? What sequence, if any, is repeatedly though imperfectly represented by actual rock sequences?

A well-known ideal cyclothem, which seems to be applicable to Pennsylvanian rocks in Kansas, is that proposed by R. C. Moore (1935). Recognition of an ideal cyclothem has been possible only after the study of large numbers of actual rock sequences. The process is inductive. From an essentially infinite number of possible ideals one is selected which seems to fit the observational data at least as well as any other.

If the selected ideal cyclothem implies a reasonable transgressive-regressive mechanism, as is certainly true of the Moore ideal, then the overall concept takes on additional weight as a unifying hypothesis. A recognized ideal cyclothem has the attributes of a "natural law" in the sense that it helps to organize diverse observational data in terms of a single, simple, and reasonable hypothesis. An ideal cyclothem is not only scientifically useful but intellectually satisfying.

The purpose of this investigation was to provide the operational mechanics of an objective procedure for making the necessary inductive step in the recognition of an ideal cyclothem. The methods used were specifically designed to answer certain questions about cyclothem in Kansas. Some of the procedural details, especially of classification and sampling, were dictated by expedience and tailored to make use of available

data. Refinements and improvements will be necessary. It is hoped, however, that the general approach used here will prove useful in future studies and other areas.

Portions of the geologic discipline, e.g. mineralogy and petrology, have long been tolerably quantitative. For many years, however, W. C. Krumbein has stood virtually alone, the outstanding proponent of a quantitative approach to problems in stratigraphy and sedimentation. Grain-size analysis has become popular, and facies maps are a standard procedure. There have been followers but few innovators.

Recently, a number of other capable people have demonstrated their interest in new ways of extending the quantitative influence. Aided and abetted by modern computer capabilities, a movement is afoot. Although in most papers the long range goals of this movement are only implied, the trend is clear and refreshing.

Acknowledgements

The author wishes to thank the members of his committee, headed by Dr. Leslie Marcus of the Department of Statistics, Kansas State University. Thanks are due also to Dr. E. J. Zeller, Geology Department, University of Kansas, for the suggestion that questions about cyclothemms might be handled by considering numerical sequences. The State Geological Survey of Kansas, especially Dr. D. F. Merriam and his staff, was of invaluable service both in supplying the data and in making available the facilities of the IBM 1620 computer center at

the University of Kansas. At Kansas State University, both the 1620 and the 1410 computers were likewise freely utilized.

SIMPLIFYING THE MOORE IDEAL CYCLOTHEM

How well does the Moore ideal cyclothem describe rock sequences within the stratigraphic range for which it was intended? With this general question in mind, the first step was to formulate the Moore ideal as a numerical sequence. Moore's original numerical designations were easily adapted (see Table 1).

The reason for combining the shales .1 and .2 into a single lithologic unit, 2, was merely that the first criterion in classification was to be gross lithology. It was desirable, in so far as possible, to restrict the application of other criteria such as fossil content and the presence or absence of coal. Similarly, the regressive units .6, .7, .8, .9 would have been difficult to distinguish from their transgressive counterparts .1, .2, .3, .4 except on the basis of relatively subtle distinctions. For that reason, corresponding transgressive and regressive units were considered equivalent. With these modifications the Moore ideal cyclothem is expressible as:

. . . 1 2 3 4 5 4 3 2 1 2 3 4 5 4 3 2 1 2 3 4 5 4 3. . .

an infinite sequence consisting of adjacent transgressive (1 2 3 4 5) and regressive (5 4 3 2 1) hemicycles. The units 1 and 5 are at once both transgressive and regressive and will here be called pivotal lithologies.

Table 1. Revised designations for cyclothem units, .
after Moore (1935, p. 24-25).

original description	designation	
	original	revised
Sandstone.	.0	1
Shale (and coal).	.9	2
Shale, typically with molluscan fauna.	.8	
Limestone, algal, molluscan, or with mixed molluscan and molluscoid fauna.	.7	3
Shale, molluscoids dominant.	.6	4
Limestone, contains fusulinids, associated commonly with molluscoids.	.5	5
Shale, molluscoids dominant.	.4	4
Limestone, molluscan, or with mixed molluscan and molluscoid fauna.	.3	3
Shale, typically with molluscan fauna.	.2	2
Shale, (and coal) may contain land plants.	.1	
Sandstone.	.0	1

Moore (1935, p. 26) anticipated the consideration of such an infinite sequence when he remarked:

The entire cyclothem thus records a single marine pulsation....This nearly symmetrical or harmonic sort of rhythm might be expressed numerically by the sequence 0-1-2-3-4-5-4-3-2-1-0.

In order to complete the classification, it was necessary to consider additional criteria. Specifically, it was necessary to distinguish between the shales, 2 and 4, and between the limestones, 3 and 5. The scheme shown in Table 2 was adopted.

Table 2. Criteria for the distinction between non-sandstones

	2 if shale	4	3 if limestone	5
FAUNA	Plant remains (pos. ident.) or mixed fauna. pelecypods, inarticulate brachiopods, gastropods, ostracodes indicative-- especially in the absence of (4) indicators. But both may be unfossiliferous.	Fusulinids (pos. ident.) or a relatively abundant mixed fauna. Crinoids, corals, bryozoans diagnostic; articulate brachiopods indicative.	Unfossiliferous or a variable assemblage, but without fusulinids.	Fusulinids required may be more or less abundant and occur with or without an assemblage like (4).
COLOR, PURITY, TEXTURE	Black fissile shale or coal (pos. ident.) Blk non-fissile reds, greens, maroons, indicative.	Ordinarily gray or buff. Beds with (4) fossils often described as poorly bedded, clayey or highly calcareous.	Impure, highly ferruginous or sandy or may be pure. Thin bed and poor consolidation indicative, but may be massive.	Relatively pure and massive, but these criteria not diagnostic.
POSITION	yellow, argillaceous, (others?) generally non-indicative.			

If and only if other criteria fail, assign the lithologic identification most concordant. On this basis alone, for instance, a shale between limestones of types 3 and 5 may be considered type 4. Nondescript limestones in association with types 1 and 2 are to be considered type 3.

The primary criteria correspond to Moore's original descriptions.

If the chief purpose of this investigation were to establish, once and for all, the "best" ideal cyclothem for the area considered, the classification of Table 2 would have to be considered inadequate. The investigation purports to be objective, yet the classification contains many subjective elements. Still worse, the ultimate appeal to "position" when decision seems hopeless assumes the underlying Moore ideal, and to answer questions about the Moore ideal on this basis is decidedly circular.

Classification, however, may be considered a separate problem. The purpose of this investigation is not to arrive at unshakeable conclusions, but rather to indicate a line of attack which should lead to objective conclusions, given a better classification, more detailed data, and so forth.

THE DISCORDANCE INDEX, G

The second step in the procedure was to define a numerical statistic to serve as a measure of the amount of deviation of any actual rock sequence from the Moore (or some other) ideal sequence. For this purpose, the discordance index G was defined as follows:

- 1) Observe the first lithologic unit, a_1 ($a_1 = 1, . . . , 5$) of the finite sequence of interest.
- 2) Consider a portion of the ideal sequence beginning with a_1 and such that a_1 occurs within a transgressive hemicycle.

- 3) Sum the number of lithologic units which would have to be inserted to convert the observed sequence of (1) to the ideal sequence of (2). Call this sum G_1 .
- 4) Consider a portion of the ideal sequence beginning with a_1 and such that a_1 occurs within a regressive hemicycle.
- 5) Sum as in (3), but comparing the observed sequence of (1) with the ideal sequence of (4). Call this sum G_2 .
- 6) The statistic G , characteristic of the observed sequence of interest and the ideal being considered, is the minimum of G_1 and G_2 .

$$G = \min(G_1, G_2)$$

This definition will, perhaps, be clarified by an example.

Consider a seven-unit actual sequence as follows:

actual sequence	2 3 2 5 3 2 1
ideal (transgressive)	... <u>2</u> <u>3</u> 4 5 4 3 <u>2</u> 1 2 3 4 <u>5</u> 4 <u>3</u> <u>2</u> <u>1</u> ...
sum of omitted units	4 + 4 + 1 = 9 = G_1
ideal (regressive)	... <u>2</u> 1 2 <u>3</u> 4 5 4 3 <u>2</u> 1 2 3 4 <u>5</u> 4 <u>3</u> <u>2</u> <u>1</u> ...
sum of omitted units	2 + 4 + 4 + 1 = 11 = G_2
$G = \min(G_1, G_2) = G_1 = 9.$	

The statistic G is called the discordance index because it represents the number of omissions from the observed sequence if the ideal is really applicable. The larger the value of G , the less likely it seems that the observed sequence actually resulted from the transgressive-regressive repetitions implied by the ideal. Because equivalent lithologies in the transgressive and regressive hemicycles are considered indistinguishable, it is logical to characterize the observed sequence by the choice

of initial transgression or regression which minimizes G. In this way, the ideal sequence being considered is given the "benefit of the doubt".

Clearly, the discordance index so defined was not the only possible choice for a measure of observed deviation from ideal sequences. The investigation reported here rests on the assumption that G was a natural and interesting choice. However, the methods used would be adaptable to other statistics, and this is a possible direction for future investigation.

FORMULATING THE QUESTIONS

The general question which guided the translation of the Moore ideal into a numerical sequence and the definition of the discordance index must be made more specific. It can be rephrased in the following alternative forms:

QUESTION A. 1. How well does the Moore ideal cyclothem describe the rock sequences summarized as the composite section of the Kansas Rock Column?

2. Would some other ideal sequence describe these "facts" better?

QUESTION B. 1. How well does the Moore ideal cyclothem describe actual rock sequences observed in outcrops reported from Kansas localities?

2. Would some other ideal sequence describe actual rock sequences better?

3. Is there adequate reason to believe that actual rock sequences are not random?

GENERATING THE POPULATION OF SEVEN-UNIT SEQUENCES

In order to begin to answer the above questions, it was necessary to restrict the length of the actual sequences which would

serve as data units. In particular, question B3 required that the distribution of G in a population of finite sequences be known. If the population of all possible sequences of length L were generated, the distribution of G in that population could be determined. On the assumption of equal likelihood among the sequences, the probability of occurrence of any particular G-value could also be found. The following information was desired:

- 1) All permutations of L lithologies chosen from the five recognized lithologic types such that identical lithologies do not occur in adjacent positions in sequence. This restriction was necessary because the actual sequence 1223454, for instance, would probably be reported as 123454.
- 2) The values of G which result from comparing each sequence of the population with the Moore ideal.

This was accomplished with the aid of an IBM 1620 computer program (see Appendix A). It can easily be shown that the number of sequences in such a population is given by:

$$N = n(n - 1)^{L-1} \quad \text{where}$$

n is the number of distinct lithologies recognized, five in this case, and L is the length of the sequences to be generated.

To see this, we may visualize the filling of L positions in sequence by n distinct kinds of items. Let a_1, a_2, \dots, a_L be the items to be chosen. There are n choices for a_1 , and for each of these there are n-1 choices for a_2 . The single restriction is $a_1 \neq a_2$. For given a_1, a_2 there are n-1 choices for a_3 , and so forth. In general,

position	1	2	3	L
item	a_1	a_2	a_3	a_L
choices	n	n-1	n-1	n-1

$a_i \neq a_{i+1}$ from which the above result

is clear.

It was desirable to fix L in such a way that the population would be of a manageable size, while the length of actual sequences to be used would be sufficient to test the hypotheses of interest. Intuitively, it would not have been wise to use actual sequences of length 2, for example, to test hypotheses concerning an ideal sequence with hemicycle length 5. After preliminary considerations of this kind, L was chosen as 7 and the population consisting of

$$N = 5(4)^6 = 20,480$$

sequences was generated. At the same time each G was calculated.

Table 3 shows the distribution of G in this population. If the sequences of the population are equally likely to occur in nature, then each possible value of G (0,1,. . . ,18) will have the probability shown in column 3. In other words, Table 3 gives the expected frequencies of occurrence for each possible G-value under the hypothesis of random deviation from the Moore ideal.

A POPULATION OF ALTERNATIVE IDEALS

Would some other ideal sequence describe the facts better? It was necessary to ask in turn, what other ideal sequences are possible? Any sequence which contains each of the recognized lithologies at least once could be taken as an ideal hemicycle. Some sequences such as

1 2 3 2 3 2 3 4 3 2 3 2 3 2 3 2 3 2 3 4 5 4

do not seem reasonable in terms of the complexity of the transgressive-regressive mechanism implied if such a sequence were to be considered a hemicycle. Nevertheless, there is no

limit to the number of sequences which might improve upon the Moore ideal cyclothem. In order to search systematically for the "best" ideal sequence, it was necessary to restrict the universe of ideals in some manner.

Table 3. Distribution of G in the population of seven-unit sequences based on the Moore ideal.

G	No. of	Pr(G) H_0	Cum. Prob. (%)
0	15	.000732	0.073
1	37	.001807	0.254
2	101	.004932	0.747
3	209	.010205	1.768
4	389	.018994	3.667
5	621	.030322	6.699
6	895	.043701	11.069
7	1148	.056055	16.675
8	1638	.079981	24.673
9	1967	.096045	34.277
10	2061	.100635	44.341
11	1833	.089502	53.291
12	2273	.110986	64.390
13	2245	.109619	75.352
14	1770	.086426	83.995
15	904	.044141	88.409
16	1019	.049756	93.385
17	863	.042139	97.599
18	492	.024023	100.001
	<u>20480</u>		

Because the population of seven-unit sequences was already available, it was convenient to consider a set of ideal hemicycles obtained by examining each sequence of the larger population to see whether either the 5th or 6th positions could be considered pivotal. This was also accomplished with the aid of a computer program (see Appendix A).

The procedure was:

- 1) Designate the lithologies in each sequence as a_1, a_2, \dots, a_7 where the subscripts indicate the position in sequence. Then $a_i = k$ ($i=1\dots 7$; and $k=1\dots 5$).
- 2) If the set consisting of a_1, a_2, \dots, a_5 contains each integer (1,2,...,5) exactly once, i.e. the first five lithologies are all different, then the sequence is a potential ideal generator subject to satisfaction of the restriction in (3).

If the set consisting of a_1, a_2, \dots, a_6 contains each integer (1,2,...,5) at least once, i.e. exactly one lithology is repeated among the first six, then the sequence is a potential ideal generator subject to satisfaction of the restriction of (4).

- 3) If $a_4 = a_6$ and $a_3 = a_7$, then the sequence is pivotal around a_5 and the hemicycle length is 5.
- 4) If $a_5 = a_7$, then the sequence is pivotal around a_6 and the hemicycle length is 6.

For example, consider the sequence 3 2 4 1 5 3 5. Among the first five lithologies, all are represented. However, $a_4 \neq a_6$ ($1 \neq 3$), so that the sequence is not pivotal around a_5 . Among the first six lithologies, exactly one lithology (3) is repeated, and $a_5 = a_7 = 5$. The sequence is pivotal around a_6 , and a_1, a_2, \dots, a_6 constitute an ideal hemicycle.

The population of hemicycles obtained in this manner has 1200 members. From these hemicycles 660 distinct ideal sequences can be generated. Consider the original seven-unit sequences 3 2 4 1 5 3 5 and 3 5 1 4 2 3 2. Both will contribute six-unit hemicycles to the 1200-member population, but these will be merely transgressive and regressive, (obverse and reverse) hemicycles of the same sequence:

...3 2 4 1 5 3 5 1 4 2 3 2 4 1 5 3 5 1 4 2 3 2 ...

However, the original sequences 2 3 4 1 5 1 4 and 1 5 1 4 3 2 3 generate distinct ideals even though the first six positions satisfy the obverse-reverse relationship. The first sequence is pivotal around a_5 and yields the ideal $\dots \underline{2\ 3\ 4\ 1\ 5\ 1\ 4}\ 3\ 2\ 3\ 4\dots$ while the second is pivotal around a_6 and yields the ideal $\dots 1\ 5\ 1\ 4\ 3\ \underline{2\ 3\ 4\ 1\ 5\ 1\ 5}\ 1\ 4\ 3\ 2\ 3\dots$

It must be emphasized that the population of 660 ideals generated in this way is by no means exhaustive; however, it is exhaustive of symmetric ideals having five- and six-unit hemicycles. A symmetric ideal is defined as one in which adjacent hemicycles are obverse and reverse, as opposed to sequences like $\dots 1\ 2\ 3\ 4\ 5\ 1\ 2\ 3\ 4\ 5\ 1\ 2\ 3\ 4\ 5\dots$, which might be called simply repetitive ideals. It should be mentioned here that simply repetitive ideals may best describe actual rock sequences in some areas. Moore (1935) and others have noted that the typical Illinois cyclothem is probably of the simply repetitive type.

Any symmetric ideal which might conceivably constitute an improvement upon the Moore ideal either (1) belongs to the 660-member population described above, or (2) has hemicycle length at least seven. The latter possibility is by no means unthinkable. The ideals to be considered were restricted in the particular manner described only because the next larger population, including seven-unit hemicycles, would have been too large to have been exhaustively analysed in the time available. Either faster computers or a continuing program of study could allow for

expanding the present investigation in the direction of a larger population of ideals.

VARIATION OF G IN AN IDEALIZED COMPOSITE SECTION

Answers to questions A1 and A2 involved a comparison between one abstraction, the population of ideal sequences, and another abstraction, the idealized composite section of the Kansas Rock Column. The connection with reality attained later by the use of actual measured sections was here lacking. Accordingly, the answers obtained should be considered relatively non-pertinent. This part of the investigation was designed to illustrate how the necessary restriction in sequence length could be overcome if it became desirable to analyse data pertaining to long sequences of actual rock units (perhaps from continuous coring operations).

The Kansas Rock Column (Moore, et al., 1951) was consulted, and the stratigraphic interval to be used was chosen. The interval conformed to that covered by available measured sections used later; it extended from the Pleasanton Group (Hepler Sandstone, Missourian) below into the Council Grove Group (Roca Shale, Wolfcampian) above. By studying the descriptions of each formation and member, the number and classification of recognized lithologies within the interval were determined. The chief problems of classification at this stage were:

- 1) deciding what sequence of lithologies to use when it happened that a formation or member was described as being differently represented at different Kansas localities, and
- 2) deciding upon the number of distinct lithologies to be included when a formation or member was described as "alternating shales and limestones" or the like.

The unavoidable subjectivity of these decisions was not critical in this phase of the study, since the purpose of the undertaking was primarily illustrative. A total of 278 lithologies were recognized and classified within the interval.

With the aid of a third computer program (see Appendix A) this information was subjected to the following steps in

- analysis:
- 1) the 278 lithologies were considered in seven-unit subsequences from bottom to top. The first seven lithologies constituted the first subsequence, lithologies two through eight constituted the second subsequence, and so forth, making a total of 272 subsequences in all.
 - 2) For the Moore ideal cyclothem, a member of population of ideals, G was computed for each individual subsequence. The values of G were combined in a five-point moving average, and the variation of G through the interval was graphically displayed (see Figure 1).
 - 3) The average value of G over the interval was determined for the Moore ideal.
 - 4) For all distinct remaining members of the population of ideals the average values of G over the interval were also obtained.

A detailed discussion of the features of Figure 1 will not be undertaken because the connection with reality is at best problematic. However, the following feature of Figure 1 is perhaps sufficiently general to be considered "real":

Levels of G are noticeably higher in the Kansas City Group and below as well as in the Admire Group and above. Clearly, the interpretation is that the Moore ideal is more descriptive of the "facts" within the Middle Missourian through Virgilian of Kansas than elsewhere in the interval considered.

Several points demand mention in connection with the analysis described above. First, the sequential consideration

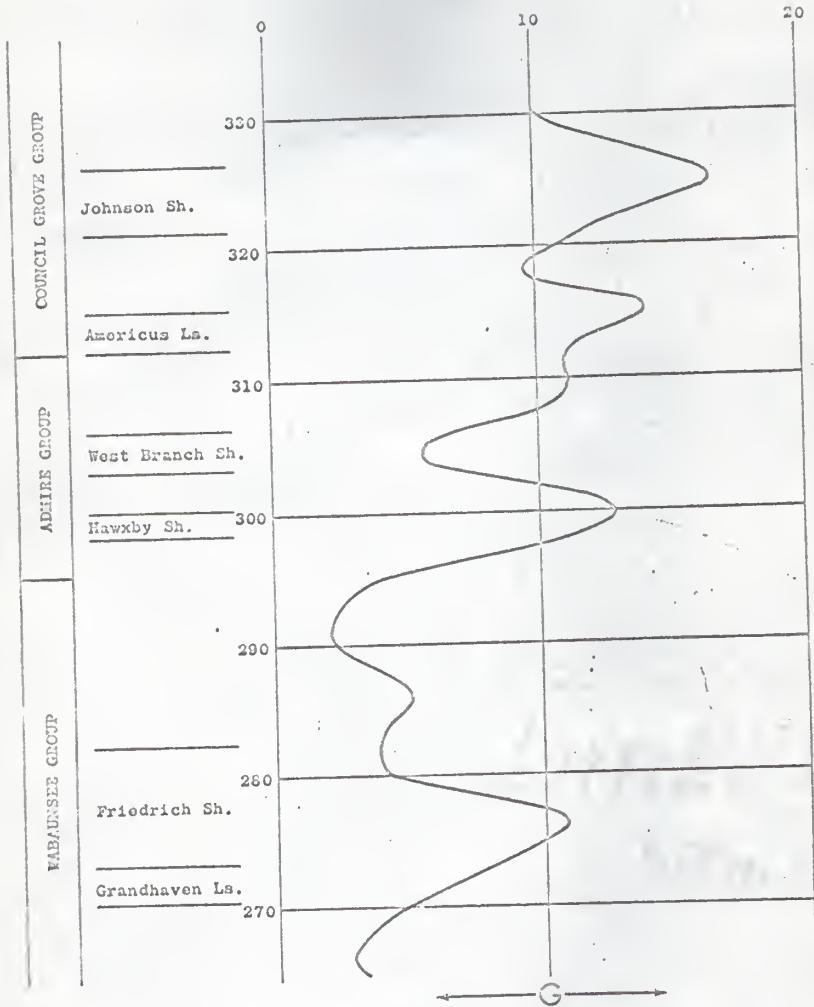


Figure 1. Variation of G through the idealized composite section (five-point moving average).

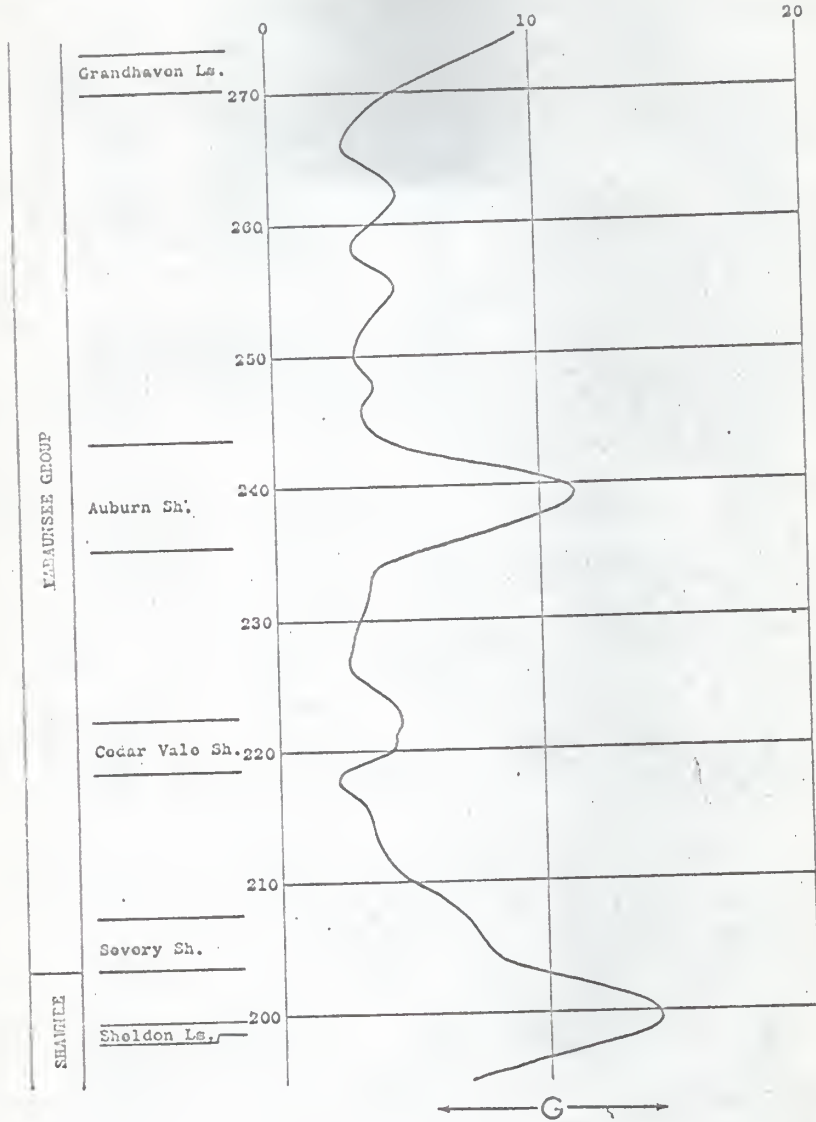


Figure 1. (continued)

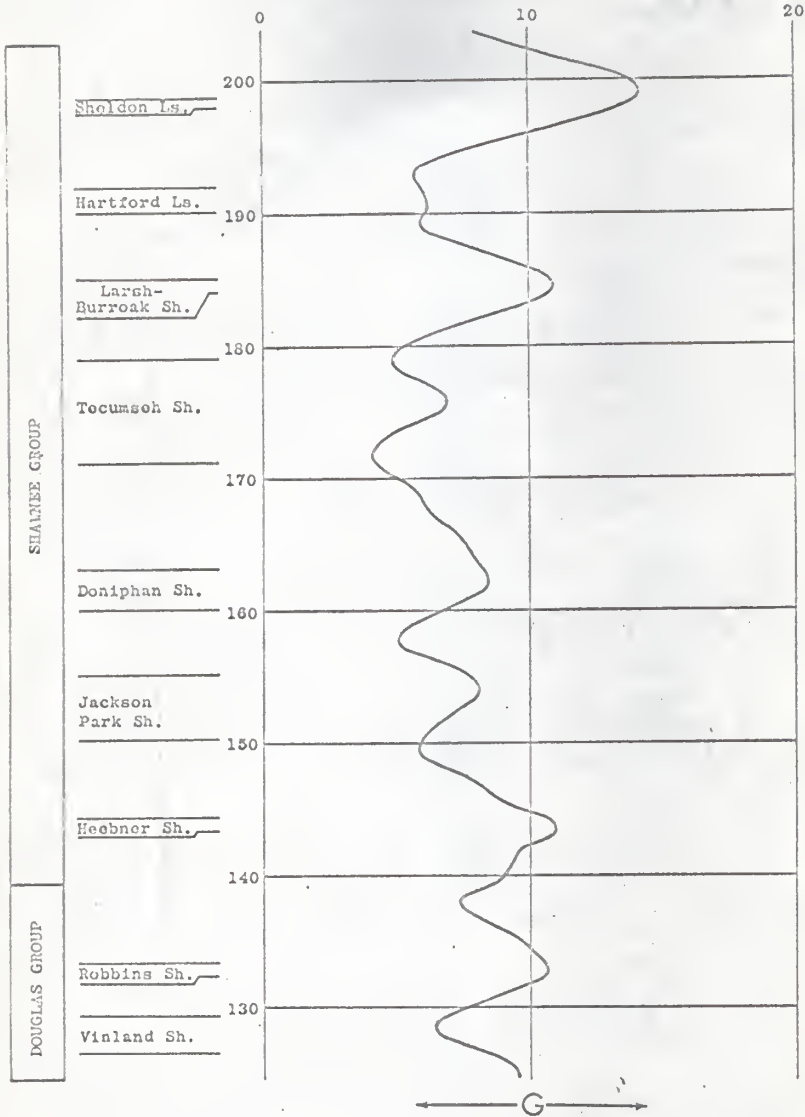


Figure 1. (continued)

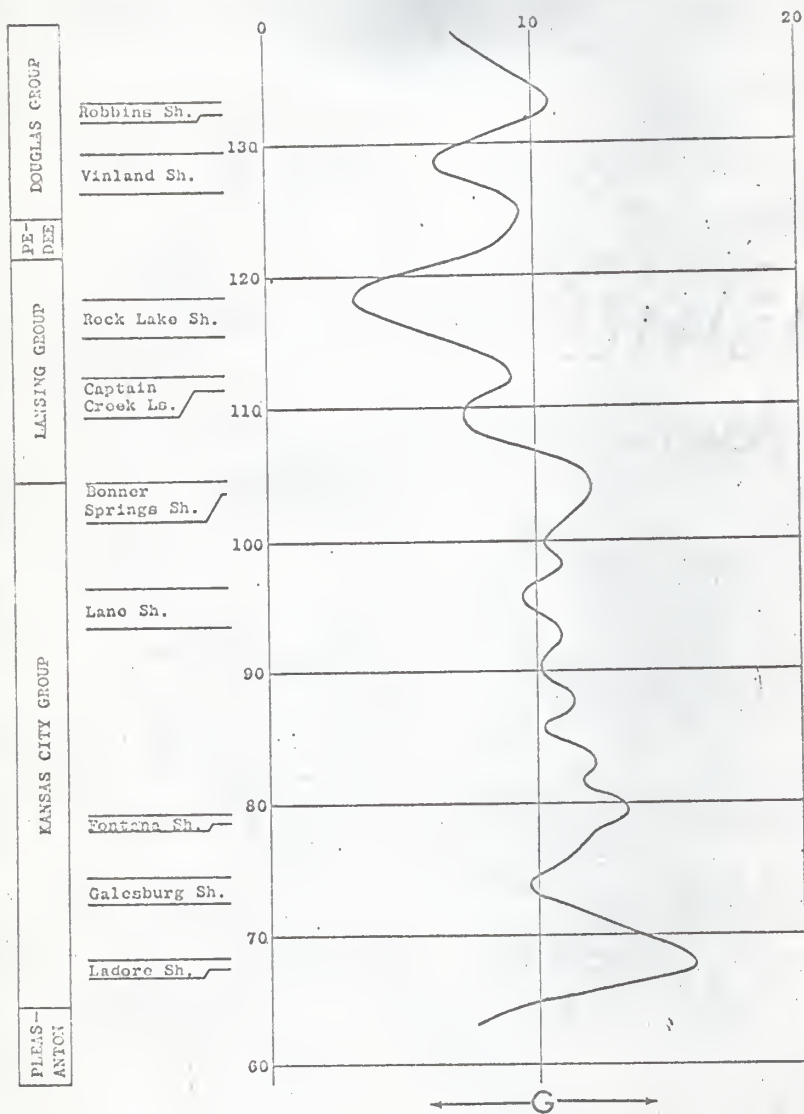


Figure 1. (concluded)

of seven-unit subsequences is only one possible method of obtaining an average G over an extended sequence. A comparison of this method with a reasonable alternative is given in Appendix B.

Secondly, because the input lithologies were visualized as representing a continuous sequence, freedom of choice as to the starting point (transgressive or regressive) could not be allowed for each subsequence. Rather, the distinct values of G_i resulting from different starting points were first accumulated and averaged over the entire interval and then minimized to obtain the final G . Each G_i represented a single initial choice of transgression or regression (for the first subsequence). Compared to the procedure for calculating G in a single seven-unit sequence, the distinction here is summarized in the statement that reported values were minimized average G (hereafter called MAG) values over the interval.

Finally, when ideals with six-unit hemicycles are considered there may be as many as four starting points which will yield distinct G_i , rather than two. In such cases, the reported MAG was the minimum of the averaged G_i , where $i = 1, 2$ or $1, 2, 3$ or $1, 2, 3, 4$ depending on certain characteristics of the ideal under consideration.

Analysis of the complete population of ideals showed that no member had MAG less than that of the Moore ideal. For explaining the composite section of the Kansas Rock Column on the basis of the least- G criterion, the Moore ideal is the best

possible sequence among all symmetric ideals with five- or six-unit hemicycles. However, no basic significance is claimed for this result because, as has been previously mentioned, the data from the Kansas Rock Column was pre-synthesized and correspondingly unreal. In addition, bias may well have been introduced by the writer during translation of descriptions into numerical sequences.

DISTRIBUTION OF G IN A SAMPLE OF ROCK SEQUENCES

Answers to questions B1, B2, and B3 were obtained from a sample of actual seven-unit rock sequences drawn from available measured sections within the region shown in Figure 2.

The State Geological Survey of Kansas kindly made available a file of measured sections and provided a map of locations for an initial selection of about 400 sections. This selection included all available sections which happened to display at least seven lithologic units within the interval from Hepler Sandstone to Roca Shale. The preliminary set of 400 sections was subjected to a sampling procedure as follows:

- 1) A grid was superimposed upon the map showing the location of, and stratigraphic group(s) represented in each available section.
- 2) The number of available sections per group was tabulated for each grid subdivision.
- 3) The total number of sections to be retained was set provisionally at 250, and it was decided that group representation should be proportional to the "size" of the group.

- 4) The percent of the total interval actually occupied by each stratigraphic group had been previously estimated by

$$P_i = \frac{n_i}{N} (100)$$

where P_i = percent of interval represented by the i th group.

n_i = number of recognized lithologies within the i th group, estimated from the Kansas Rock Column.

N = estimated total number of recognized lithologies in the interval studied.

- 5) These considerations dictated that the group representation in the final sample should be as follows:

Group represented	% of sample (= P_i)
Council Grove	8
Admire	6
Wabaunsee	33
Shawnee	23
Douglas	6
Pedee	1
Lansing	6
Kansas City	14
Pleasanton	3

- 6) Where a group was originally represented to excess, sections were discarded from those grid subdivisions containing the most representatives of the group in question. The particular sections to be discarded were randomly chosen. In this way 250 sections were chosen from the available 400.
- 7) Locations of the desired 250 sections were then communicated to Dr. D. F. Merriam, who provided Xerox copies of the measured sections and descriptions. To this point the writer was unaware of the detailed characteristics of the sections to be used.
- 8) On each section containing more than seven lithologic units, according to the classification system of Table 2, a starting point was randomly chosen. Upward from this starting point, seven successive units were defined and classified as 1,2,3,4,5.

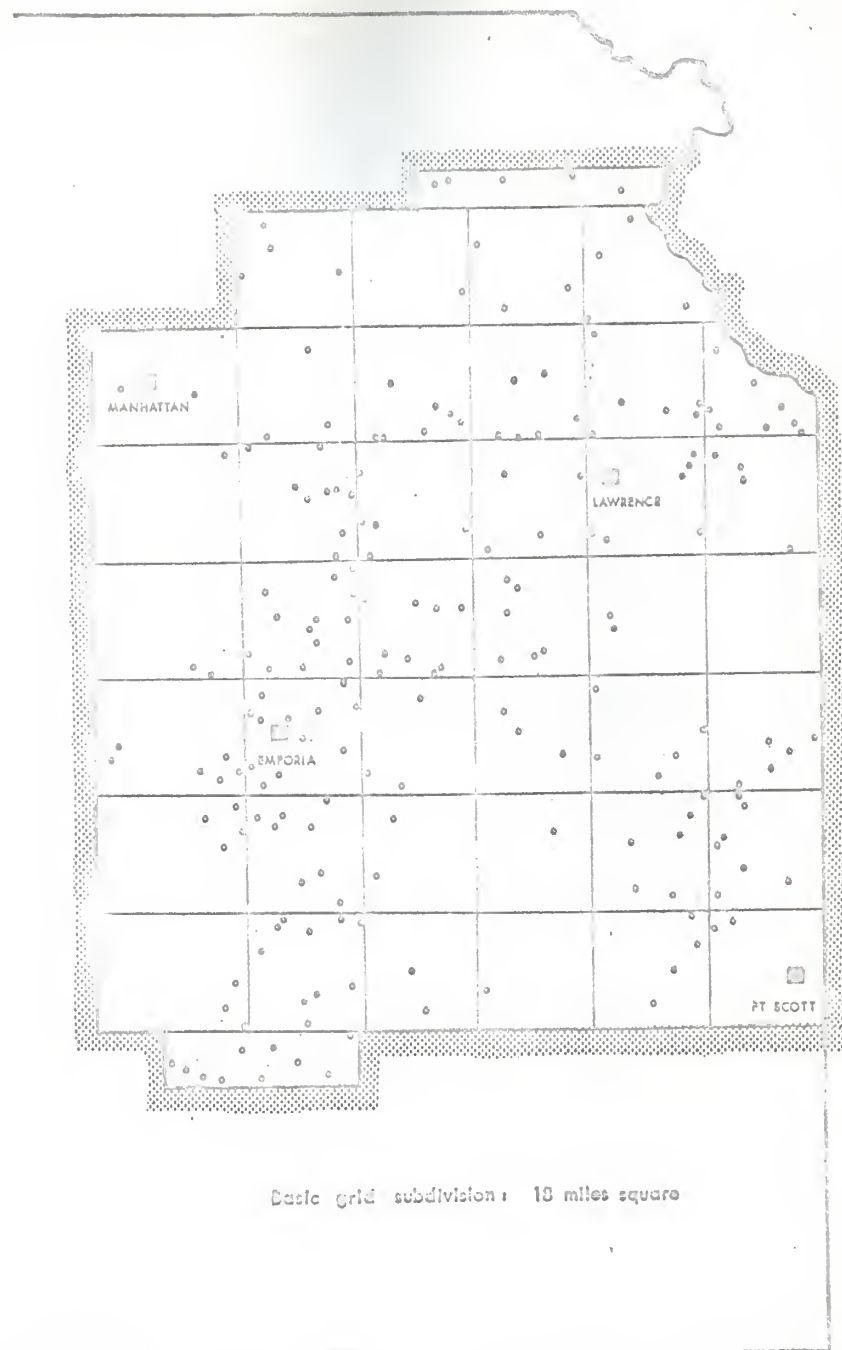


Figure 2. Distribution of sample localities.

- 9) For several different reasons, mainly because of difficulty in interpreting descriptions, 15 sections were considered unsuitable and were discarded. The final sample consisted of 235 seven-unit sequences.
- 10) To check the percentage of group representation each sequence in the final sample was classified to group. In cases of overlap, the section was counted twice. Compare the break-down below with that of (4).

Group represented	% of sample
Council Grove	8.1
Admire	7.0
Wabaunsee	31.4
Shawnee	20.3
Douglas	7.0
Pedee	1.1
Lansing	6.2
Kansas City	16.2
Pleasanton	2.6

The chief purpose of this procedure was to insure that the final sample of seven-unit sequences would be spread over the geographic area and the stratigraphic interval of interest. The writer feels that this kind of "representativeness" is a desirable feature of geologic sampling, in which true randomness is usually not at issue. In the present case, certainly, it was not a matter of choosing between the kind of sample obtained and a truly random sample. Ideally, a random sample would have had both locality and stratigraphic interval (group) randomly predetermined. It would have been necessary to be able to go to any locality and there observe a section within any stratigraphic group. The obvious difficulty is that when one is restricted to surface measurements, he is also restricted by the fact that outcrops are where you find them.

In addition, it was necessary for this study to consider only those sections already measured and recorded. No sample of the available sections could have been considered a random sample of the population to which inference was to be made, i.e. seven-unit sequences in the three-dimensional area-interval of interest. That the sample actually used was a reasonable approximation to that goal is, at this point, simply assumed.

With the aid of a fourth computer program (see Appendix A) the sample was analysed in a manner similar to that already described for the composite section. Differences were as follows:

- 1) The 235 sections were separate entities and the choice of a minimizing starting point was left open for each seven-unit sequence.
- 2) The average values of G were determined after the 235 separate minimumizations, hence were average minimum G (AMG) values, rather than MAG values as before.

The G-values corresponding to each observed sequence, with reference to the Moore ideal, formed the basis of a simple test in answer to the question B3. Consider the null hypothesis, H_0 : The sample of observed sequences was drawn from a population described in Table 3, i.e. every conceivable seven-unit sequence had equal opportunity to appear in the sample because the sequences occur randomly in nature.

For the sake of brevity, call this H_0 the randomness hypothesis. The alternative hypothesis, then, is a nonrandomness or the simple negation of H_0 .

Table 5 shows the observed number of occurrences for the various G-values, the expected number according to the distribution under H_0 (from Table 3), and calculated quantities necessary for a chi-square test of H_0 , where

G = a value of the discordance index.

O_G = the number of times (out of 235) the particular G was observed.

E_G = the number of times the particular G would be expected to have occurred under H_0 (=235 x column three of Table 3).

Table 5. Data for the chi-square answer to B3.

G	O_G	E_G	$O_G - E_G$	$(O_G - E_G)^2 / E_G$
0	2	8.62	14.38	23.99
1	0			
2	3			
3	6			
4	12			
5	12	7.13	4.87	3.33
6	17	10.27	6.73	4.41
7	9	13.17	4.17	1.32
8	29	18.80	10.20	5.53
9	12	22.57	10.57	4.95
10	26	23.65	2.35	0.23
11	15	21.03	6.03	1.73
12	27	26.08	0.92	0.03
13	14	25.76	11.76	5.37
14	26	20.31	5.69	1.59
15	3	10.37	7.37	5.24
16	18	11.69	6.31	3.41
17	2	15.55	11.55	8.58
18	2			

Groupings at the extremes of the observed distribution were made in order to satisfy the chi-square requirement

that $\min(E_G) = 7$. The procedure is to calculate the statistic:

$$\chi^2 = \sum_G \frac{(O_G - E_G)^2}{E_G} = 69.71$$

and to note that in large samples χ^2 is approximately chi-square distributed under H_0 . Reference to tabled chi-square with 13 degrees of freedom ($m-1$, where $m = n_0$ cells used) reveals that under H_0 the probability of observing a χ^2 this large or larger is much less than 0.00001. The randomness hypothesis is most decidedly to be rejected. It may be desirable to emphasize the assumptions under which the above chi-square test is a valid rejection of the randomness hypothesis. We assume:

- 1) that if the recognized lithologies actually occurred in random sequences in nature, then any sequence would be as likely to occur as any other.
- 2) that the population distribution of G under the randomness hypothesis would be the same as the distribution derived by generating all possible sequences and assigning them equal probabilities.
- 3) that we have a random sample from the population of interest, namely the population of all seven-unit sequences, within the defined three-dimensional area-interval.
- 4) that the dependence of the theoretical G -distribution on the ideal sequence of reference (the Moore ideal) does not affect the test of randomness.

Assumptions (1) and (2) would appear to be justified. Assumption (3) as we have already seen, is invalid but should be approximately true. Assumption (4) is reasonable because the alternative to randomness is unspecific. The particular kind of order we visualize in order to be able to calculate G has

no direct bearing on the question, "Does any order exist?". In other words, the test would be expected to reject with any choice of reference sequence.

The AMG values obtained from the analysis of the entire population of 660 distinct ideals will allow no simple interpretation. Of the ideals tested 78 yielded AMG less than that of the Moore ideal. The twenty smallest AMG are listed in Table 6.

Among this surprisongly large number of "improvements" over the Moore ideal, the best is . . . 1 2 3 4 5 2 5 4 3. . . . The chief difference between this and the Moore ideal is the extra unit-2 per hemicycle. Table 7 shows another set of the hemicycles which generate ideals with relatively low AMG. The grouping is intended to illustrate some of the reasons for the results obtained. Note first that all ideals shown have six-unit hemicycles, and the unit repeated in the hemicycle is either 2 or 3. Both of the observations hold for all 78 "improvements".

Table 8 shows the distribution and frequency of the recognized units among the positions (a_i) of the sample sequences. The high proportions of units 2 and 3 would seem to account for the fact that ideals with extra units 2 or 3 have low AMG, other factors remaining constant. Note also the low proportion of unit-1 in the sample. Table 7 shows that the position unit-1 occupies has relatively little effect on the value of AMG. In the first group of four ideal hemicycles, for instance, the change in position of unit-1 from a_1 to a_4 caused the change in AMG rank from 1 to 7.

Table 6. Value and rank of twenty smallest AMG.

ideal hemicycle	AMG	rank
1 2 3 4 5 2	7.4596	1
1 2 5 4 3 2	7.5106	2
2 1 3 4 5 2	7.7702	3
2 1 5 4 3 2	7.8468	4
2 3 1 4 5 2	7.8979	5
1 2 3 4 2 5	8.0043	6
2 3 4 1 5 2	8.0383	7
2 1 3 4 2 5	8.0894	8
2 1 3 2 4 5	8.1021	9
1 2 3 2 4 5	8.1489	10
1 3 2 5 4 3	8.2213	11
2 3 1 4 2 5	8.2596	12
1 3 2 5 3 4	8.2638	13
2 3 1 2 4 5	8.2723	14
3 1 2 5 4 3	8.3191	15
2 1 3 2 5 4	8.3404	16
3 1 2 5 3 4	8.3447	17
1 2 3 2 5 4	8.3872	18
2 3 4 1 2 5	8.3957	19
1 3 2 3 4 5	8.4085	20
1 2 3 4 5 (Moore)	9.9319	79

Table 7. Selected AMG showing relationships: * indicates reverse of a previously listed hemicycle.

ideal hemicycle	AMG	rank
1 2 3 4 5 2	7.4596	1
2 1 3 4 5 2	7.7702	3
2 3 1 4 5 2	7.8979	5
2 3 4 1 5 2	8.0383	7
1 2 5 4 3 2	7.5106	2
2 1 5 4 3 2	7.8468	4
*2 5 1 4 3 2	8.0383	7
*2 5 4 1 3 2	7.8979	5
1 2 3 4 2 5	8.0043	6
2 1 3 4 2 5	8.0894	8
2 3 1 4 2 5	8.2596	12
2 3 4 1 2 5	8.3957	19
1 2 3 2 4 5	8.1489	10
2 1 3 2 4 5	8.1021	9
2 3 1 2 4 5	8.2723	14
2 3 4 1 4 5	9.6468	68

Table 8. Distribution and frequency of recognized lithologic units in the sample of seven-unit sequences.

Position	Unit				
	1	2	3	4	5
1	26	79	71	22	37
2	24	92	62	28	29
3	14	76	81	32	32
4	17	85	68	33	32
5	13	80	82	28	32
6	7	93	57	38	40
7	15	65	86	27	42
TOTAL	116	570	507	208	244
percent	7.05	34.65	30.82	12.65	14.83

In summary, the relative proportions in the sample of the units 1-5 interact with the ordering of these units in the ideal and AMG is a complex function of both. This should have been obvious at the outset. Are all 78 ideals with low AMG to be considered improvements over the Moore ideal? If the classification of lithologies were entirely objective and unambiguous, the answer would be an unqualified yes.

PROBLEMS OF CLASSIFICATION

The classification used here is deficient. It has already been mentioned that the tie-breaking "position" criterion begs the question. In a sense, the use of such a criterion is the most serious deficiency of this study. In another sense, it was largely irrelevant. Given criteria adequate to assign every lithology in the chosen area-interval to one of the recognized a priori classes, the need for tie-breaking

would have been automatically removed. This study has been chiefly concerned with the mechanical procedures whereby useful conclusions would be reached, given, as a point of departure, just such an objective and unambiguous classification of cyclothemic units.

The following brief discussion is intended as the barest food for thought concerning the difficulties to be encountered in any future attack on problems of classification. The discussion is in terms of the specific questions asked here, but the implications are more general.

Fusulinid Requirement

In the Moore ideal, the type-5 unit is pivotal between the hemicycles in such a way that if physical transgression and regression is visualized, then unit 5 represents maximum transgression or the so-called "deep-water" limestone. It may be true that the presence of fusulinids is one of the best criteria for recognizing such a unit. Still, the type-5 unit which contains fusulinids at one locality may be physically continuous with a limestone which is type-3 at another locality because fusulinids are lacking. If a true facies change is so indicated, such a situation need not concern us too much. On the other hand, if other faunal elements remain the same we may legitimately wonder whether fusulinids are all that important. With special regard to the present study, it is probable that fusulinids may be lacking in the descriptions of some measured sections though present at the outcrop.

Inclusiveness of Unit 3

The limestones encountered in the sections used for this investigation are fusulinid-bearing, fossiliferous (no fusulinids), or unfossiliferous; massive to thin-bedded and often wavy-bedded; hard and dense to soft, argillaceous or "punky"; pure to ferruginous or otherwise impure; and so forth. Almost any combination of such adjectives describes some limestone in the interval considered. In what sense can all non-fusulinid limestones be considered equivalent? In particular it seems likely that the many impure and thin-bedded limestones interbedded with shales and not distinguished as members should be separated from other type-3 units.

Inclusiveness of Unit 2

A similar objection can be made concerning the criteria for recognizing unit 2. As a general rule, the shales of the interval considered tend to be less fossiliferous than adjacent limestones. This alone accounts for the scarcity of positively identifiable type-4 units, and the majority of shales became type-2 by default as it were. Unit 2 may be marine or nonmarine, fossiliferous or unfossiliferous, and any color at all.

Degree of Clasticity for Unit 1

An attempt to use the classification of Table 2 on descriptions of measured sections is especially difficult when the terms siltstone, mudstone, and conglomerate are encountered. Is siltstone to be called sandstone or shale? Is mudstone to be considered shale or, if calcareous, impure limestone?

What about conglomeratic limestones? The relative scarcity of type-1 units in the sample is probably "real" regardless of classification difficulties, but we may wonder whether the presence of sandstone is really an environmental measure. The sandstone environment, whatever it may be, could have been present at many points in time which did not happen to coincide with a supply of coarse clastics.

Thickness

Thickness is a criterion whether it should be or not. For this investigation, all lithologic units less than .3 feet thick were ignored. Clearly there must be some such arbitrary cut-off point. Is it then reasonable to assign equal weight to all limestones, for instance, from .3 to 20 feet in thickness?

Generalizations and Directions

We may distinguish at least three types of troublesome questions stated or implied in the above discussion:

- 1) How many lithologies should be recognized?
- 2) What combination of criteria will effect the assignment of actual rock units to the n recognized categories without ambiguity?
- 3) Given an appropriate set of criteria, how should they be weighted, i.e. what is the order of their relative importance?

There exist no set procedures to tell us which criteria may be of importance, but intuitively we may conclude that it will be necessary to consider many types of criteria. Surely an objective synthesis should draw information from many

fields. Paleontology, mineralogy, petrology, sedimentology, geochemistry, all may be called upon to contribute to the store of measurable variables from which a set of criteria appropriate for the purpose at hand may somehow be chosen. A subjective guiding principal for preliminary selection of criteria would include an evaluation, in terms of current geologic thought, of the "amount of information" about ancient environment contained in any particular variable.

Various types of cluster and factor analysis exist which could be applied to such preliminary criterion matrices, and in theory at least useful answers to questions like (1) and (2) would eventually result. For an interesting example of factor analysis applied to a geologic problem see Imbrie and Purdy (1962). Question (3) could then be approached in a relatively straightforward manner through the use of discriminant functions.

Development of a fully objective classification designed specifically for an investigation such as this would be a long and arduous task. By side-stepping the difficult job and anticipating some of the potential returns on such an investment of effort, this study may serve as some small motivation. The even more difficult task of developing a master classification which would be adequate with reference to a broader field of problems is not outside the realm of possibility. The first steps should be taken with such a larger goal already in mind.

CONCLUSIONS

It is easy to see, in retrospect, that the classification used here was such that the preponderance of units 2 and 3 in the sample was inevitable. Any change in the classification which tended to equalize the proportions of the recognized units would probably tend to reduce the number of improvements on the Moore ideal. Of course, this is not to be considered a goal, i.e. justification of the appropriate criteria must be based on independent evidence.

The purpose of this investigation will have been served if any motivation has been provided toward the development of an objective classification based on geochemical and/or lithologic indicators of environment. In addition it is hoped that the distinction is fully grasped between what is reasonable and what is demonstrable. In the opinion of the writer, there is some degree of evidence here that the Moore ideal cyclothem is, after all, the truth behind the complexity of the observable quantities. But opinion is relatively worthless. Refinement of the criteria for classification may ultimately render the truth susceptible to demonstration by methods similar to those developed here.

In the meantime, geology as a scientific discipline needs more and better attempts to demonstrate the truth of its reasonable hypotheses. If nothing else, such attempts will often demonstrate that our basic methods of observation, measurement, and classification are inadequate to deal

systematically with the larger problems. We need to become increasingly aware that the only slightly exaggerated formulation, "How do you feel about cyclothems?", is simply not a scientifically meaningful question. We need to become increasingly willing to focus our attentions on hypotheses at least potentially susceptible to proof and on methods oriented toward the realization of that potential.

REFERENCES

- Imbrie, J., and Purdy, E. G. (1962) Classification of Modern Bahamian Carbonate Sediments: in Classification of carbonate rocks, Am. Assoc. Petroleum Geologists Special Pub.
- Moore, R. C. (1935) Stratigraphic Classification of the Pennsylvanian Rocks of Kansas: The State Geol. Survey of Kansas, Bull. 22, 256 pp.
- Moore, R. C., et al (1951) The Kansas Rock Column: The State Geol. Survey of Kansas, Bull. 89, 132 pp.

APPENDIX A. Programs Used.

The programs reproduced here were written with the specific problems of this study in mind, and most of them are inflexible. In other areas of study, for different classifications of recognized lithologies, or for different lengths of the sequences to be sampled many modifications in the programs would be necessary. In addition, the use of SENSE SWITCHES, PAUSES, etc., shows that personal supervision of the running of these programs was necessary. The forms given are the original IBM 1620 programs, although programs 3 and 4 were later modified for unsupervised use on an IBM 1410-1401 tape oriented system. Finally, the writer does not claim that the forms used here are either elegantly conceived or characterized by optimum running-time. For these reasons, it is not expected that others will desire to use these programs in precisely their present forms.

However, experienced programmers may find it helpful to use bits and pieces of the present forms in programs designed to accomplish similar ends. Accordingly, the following presentation is given:

- I. The general objectives of major steps in the execution of each program.
- II. Definitions or descriptions of the important symbolic designations used in each program.
- III. The FORTRAN source program itself with comments keyed to the general objectives.

PROGRAM 1 -- Population Generator

- ① Read control
 - a. length of sequences to be used (here 7) = LSEQ
 - b. number of recognized lithologies (here 5) = NDIG
 - c. first sequence to be considered (digits entered in reverse order); example -- 1 4 2 5 3 1 3. *
- ② Initialize and enter Moore ideal.
- ③ Depending on (1a) enter nested loops designed to increment each digit of (1c) between original values and maximum values equal to NDIG.

For the example in (1c) the sequence actually under consideration is the reverse: 3 1 3 5 2 4 1. The next sequences to be considered will be 3 1 3 5 2 4 2, 3 1 3 5 2 4 3, 3 1 3 5 2 4 4, and 3 1 3 5 2 4 5. If NDIG = 5, the program will then skip to 3 1 3 5 2 5 1, and so forth.

- ④ For each sequence under consideration, the adjacent digits are examined. If two adjacent digits are equal, the sequence is not to be considered a member of the population and the program increments as in (3) and goes on to the next sequence.
- ⑤ When no two adjacent digits of the sequence considered are equal, the program proceeds to calculate G for that sequence
 - a. by considering the initial digit as belonging to the transgressive hemicycle of the Moore ideal, and
 - b. by considering the initial digit as belonging to the regressive hemicycle of the Moore ideal, whence
 - c. $G = \min(5a, 5b)$.
- ⑥ Punch the digits of each population member together with the corresponding G-value.
- ⑦ Continue until the last candidate for inclusion in the population has been considered. For the example given, the last candidate would be 5 5 5 5 5 5 5. Of course,

*Also entered here is INUM; see definition below.

PROGRAM 1 -- Population Generator

- ⑦ (cont) the last sequence to be accepted as a population member will be 5 4 5 4 5 4 5. Similarly, the first sequence to be entered for consideration will be 1 2 1 2 1 2 1.

Important symbols.

LSEQ = length of sequences to be generated.

NDIG = number of recognized lithologies.

INUM = number of sequences to be considered by the successive enumeration procedure before skipping to a new starting point. Can be used to save time by skipping a long series of trial sequences all of which will be rejected.

MRET = multiplying factor for extending Moore ideal into the M(k) positions.

IN(I) = subscripted variable representing input digits of beginning sequence; maximum I = 12.

IS1 to_ unsubscripted variables set equal to the IN(I)
IS12 and used as the lower index of the nested DO's.

KJ(J), = subscripted variables used for juggling the
LU(K) original (and incremented) digits of sequences.

MST1 = starting point for calculation of G (transgressive).

MST2 = second starting point (regressive).

IG1 = G calculated from MST1.

IG2 = G calculated from MST2.

IGMIN = min(IG1, IG2).

C CYCLOTHEM PROBLEM
 C PROGRAM 1 -- POPULATION GENERATOR
 C

DIMENSION KU(12),LU(12),M(56),IN(12)
 100 FORMAT (2(I3))
 101 FORMAT (12(I3),I5)
 102 FORMAT (12(I3),3(I5))
 104 FORMAT (11THAT IS ALL)

C
 C ① READ CONTROL
 C

99 READ 100, LSEQ,NDIG
 92 READ 101, (IN(J),J=1,12),INUM
 IF (SENSE SWITCH 1) 91,90
 91 PAUSE

C
 C ② INITIALIZE
 C

90 INPUT=LSEQ+1
 DO 98 I=1,LSEQ
 KU(I)=IN(I)
 98 CONTINUE
 DO 97 I=INPUT,12
 KU(I)=0
 97 CONTINUE
 IS1=IN(1)
 IS2=IN(2)
 IS3=IN(3)
 IS4=IN(4)
 IS5=IN(5)
 IS6=IN(6)
 IS7=IN(7)
 IS8=IN(8)
 IS9=IN(9)
 IS10=IN(10)
 IS11=IN(11)
 IS12=IN(12)
 INC=0
 MRET=NDIG-2

C
 C ③ ENTER MOORE IDEAL AND EXPAND INTO THE M(K) POSITIONS
 C

96 J=INC*NDIG
 K=INC*MRET
 DO 95 I=1,NDIG
 L=I+J+K
 M(L)=I
 95 CONTINUE
 J=(INC+1)*NDIG
 DO 94 I=1,MRET
 L=I+J+K
 M(L)=NDIG-I
 94 CONTINUE
 INC=INC+1
 IF (INC-7) 96,93,96

C
C
C

③ ENTER NESTED LOOPS

```

93 GO TO (1,2,3,4,5,6,7,8,9,10,11,12),LSEQ
12 DO 50 I12=IS12,NDIG
   KU(12)=I12
11 DO 50 I11=IS11,NDIG
   KU(11)=I11
10 DO 50 I10=IS10,NDIG
   KU(10)=I10
 9 DO 50 I9=IS9,NDIG
   KU(9)=I9
 8 DO 50 I8=IS8,NDIG
   KU(8)=I8
 7 DO 50 I7=IS7,NDIG
   KU(7)=I7
 6 DO 50 I6=IS6,NDIG
   KU(6)=I6
 5 DO 50 I5=IS5,NDIG
   KU(5)=I5
 4 DO 50 I4=IS4,NDIG
   KU(4)=I4
 3 DO 50 I3=IS3,NDIG
   KU(3)=I3
 2 DO 50 I2=IS2,NDIG
   KU(2)=I2
 1 DO 50 I1=IS1,NDIG
   KU(1)=I1
   J=1

```

C
C
C

④ EXAMINE ADJACENT DIGITS

```

25 IF(KU(J)-KU(J+1)) 30,50,30
30 IF (J-LSEQ+1) 40,45,40
40 J=J+1
   GO TO 25
45 CONTINUE
   KINC=KU(LSEQ)

```

C
C
C

⑤ SET STARTING POINTS FOR IG1,IG2

```

   GO TO (501,502,503,504,505,),KINC
501 MST1=1
   GO TO 700
502 MST1=2
   MST2=8
   GO TO 700
503 MST1=3
   MST2=7
   GO TO 700
504 MST1=4
   MST2=6
   GO TO 700
505 MST1=5

```


C
C
C ⑤ CALCULATE IG1

```

700 J=0
    IG1=0
    DO 701 I=1,LSEQ
    K=LSEQ+1-I
702 MUP=MST1+J
    IF (KU(K)-M(MUP)) 703,704,703
703 IG1=IG1+1
    J=J+1
    GO TO 702
704 J=J+1
701 CONTINUE
    IF (KU(LSEQ)-1) 705,706,705
705 IF (KU(LSEQ)-NDIG) 710,706,710
706 IG2=IG1
    GO TO 900

```

C
C
C ⑤ CALCULATE IG2

```

710 J=0
    IG2=0
    DO 711 I=1,LSEQ
    K=LSEQ+1-I
712 MUP=MST2+J
    IF (KU(K)-M(MUP)) 713,714,713
713 IG2=IG2+1
    J=J+1
    GO TO 712
714 J=J+1
711 CONTINUE

```

C
C
C ⑤ FORM MINIMUM G

```

900 IF (IG1-IG2) 902,902,903
902 IGMIN=IG1
    GO TO 904
903 IGMIN=IG2
904 DO 901 J=1,12
    I=13-J
    LU(I)=KU(J)
901 CONTINUE

```

C
C
C ⑥ OUTPUT

```

    PUNCH 102, (LU(I),I=1,12),IG1,IG2,IGMIN
    INUM=INUM-1
    IF (INUM) 905,92,50
905 IF (SENSE SWITCH-1) 92,50
50 CONTINUE
    TYPE 104
    END

```

PROGRAM 2 -- Eligible Ideals.

- ① Read one card at a time from output of PROGRAM 1, i.e. the individual digits of each member of the total population of LSEQ-unit sequences. Note, however, that this program was written to handle only the case where LSEQ = 7.
- ② Initialize.
- ③ Examine the first five digits of the sequence being considered.
 - a. If each of the integers 1,2,3,4,5 is present exactly once
 - b. examine the sixth digit to see if it is the same as the fourth and likewise the third vs seventh. If both,
 - c. accept the sequence as an ideal generator with hemicycle length 5 and punch digits of original sequence, hemicycle length, and G-value of original sequence.
 - d. If conditions are not satisfied, go to (4).
- ④ Examine the first six digits of the sequence being considered.
 - a. If exactly one of the digits 1,2,3,4,5 is repeated,
 - b. examine the seventh digit to see if it is the same as the fifth. If so,
 - c. accept the sequence as an ideal generator with hemicycle length 6 and punch as in (3c).
- ⑤ Repeat until total input population is exhausted.

Note: this program does not yield the final population of ideals, as the obverse-reverse relationship discussed in the text was revealed and sorted out by inspection of these results.

Important symbols.

IA(J) = subscripted variable carrying digits of the input sequences to be tested; J = 1,7

IGEE = G-value for the sequence considered, along for the ride.

PROGRAM 2 -- Eligible Ideals.

Important symbols. (cont)

IAG01, = set in turn to each IA(J) and used in branching
IAG02 operations.

IB1 = locations used to indicate occurrence or non-
IB2 occurrence of the lithologic classifications
. 1 to 5.

IB5

C PROGRAM 2 -- ELIGIBLE IDEALS

C

DIMENSION IA(7)

100 FORMAT (15X,7(I3),10X,I5)

101 FORMAT (7(I3),5X,I3,36X,I5)

102 FORMAT (13HLOGICAL ERROR)

C

C

① READ ONE DATA CARD

C

200 READ 100, (IA(J),J=1,7),IGEE

C

C

② INITIALIZE

C

IPIV=1

IB1=0

IB2=0

IB3=0

IB4=0

IB5=0

C

C

③ EXAMINE FIRST FIVE DIGITS

C

DO 201 J=1,5

IAG01=IA(J)

GO TO (1,2,3,4,5),IAG01

1 IB1=1

GO TO 201

2 IB2=1

GO TO 201

3 IB3=1

GO TO 201

4 IB4=1

GO TO 201

5 IB5=1

201 CONTINUE

202 IBSUM=IB1+IB2+IB3+IB4+IB5

GO TO (203,204,204,204,204,205),IPIV

203 IF (IBSUM-5) 300,206,300

204 TYPE 102

PAUSE

205 IF (IBSUM-5) 200,400,200

206 IPIV=5

GO TO 400

300 IPIV=6

C

C

④ EXAMINE SIXTH DIGIT

C


```
300 IPIV=6
      IAGO2=IA(6)
      GO TO (11,12,13,14,15),IAGO2
11  IB1=1
      GO TO 202
12  IB2=1
      GO TO 202
13  IB3=1
      GO TO 202
14  IB4=1
      GO TO 202
15  IB5=1
      GO TO 202
```

```
C
C      3 CHECK DIGITS 4,6 AND 3,7
```

```
C
400 IF(IPIV-5) 204,401,405
401 IF (IA(4)-IA(6)) 404,402,404
402 IF (IA(3)-IA(7)) 200,403,200
```

```
C
C      3,4 OUTPUT
```

```
C
403 PUNCH 101, (IA(J),J=1,7),IPIV,IGEE
      GO TO 200
404 IPIV=6
405 IF (IA(5)-IA(7)) 200,403,200
      END
```

PROGRAM 3 -- MAG over an Extended Sequence.

- ① Read Control -- number of lithologies in extended sequence (here 278) = NLITH.
- ② Read Data from cards 1 to NLITH
 - a. integer representing classification of lithology in question,
 - b. identification number.
- ③ Read one at a time from the population of ideal generators (reduced output from PROGRAM 2) and extend the given digits through a sequence of $M(k)$, $k = 1$ to 500.
- ④ Consider seven-unit subsequences consisting of the integers in (2a) for cards 1 through 7, cards 2 through 8, , cards i through $i+6$, etc.
- ⑤ Find starting points which may yield different subsequence G-values.
- ⑥ By comparison with the $M(k)$, calculate the G-value for each of the subsequences of (4) and according to the setting of SENSE SWITCH 1
 - a. punch and accumulate subsequence G-values or
 - b. accumulate only.
 - c. Calculate average G when subsequences are exhausted.
- ⑦ Repeat step (6) for initial starting positions of (5). Minimize the distinct averages so obtained; minimum is MAG.
- ⑧ Punch ideal being considered (first seven digits) and the MAG value.
- ⑨ Repeat from (3) until population of ideals is exhausted.

Important symbols.

NLITH = sample size; number of lithologies in extended sequence, maximum 450.

IA(J) = classification of lithologies present, $J = 1, \text{NLITH}$.

ITHEM(K) = integers in ideal considered.

NPIV = pivotal position of ideal considered.

LOGIC = indicator of transgression or regression used in extending ideal.

PROGRAM 3 -- MAG over an Extended Sequence.

Important symbols. (cont)

MST(J) = starting points, J = 1,4.

NREP = number of subsequences, NLITH-6.

IC = an accumulator showing which starting point is being used; also occurs as subscript in MST(IC), IGEE(IC), etc.

IGEE(K) = G-values per subsequence and starting point.

GEE(K) = equivalent values in floating point.

SUMG(K) = GEE(K) summed over subsequences.

AVGEE(K) = average G over subsequences per starting point.

AVMIN = minimum of AVGEE(K) = MAG.

C CYCLOTHEM PROBLEM
 C PROGRAM 3 -- MAG OVER AN EXTENDED SEQUENCE
 C

DIMENSION IA(450),ID(450),M(500),ITHEM(7),MST(4),IGEE(4)
 DIMENSION GEE(4),SUMG(4),AVGEE(4)

100 FORMAT (7(I3),5X,I3)
 101 FORMAT (I5,60X,I5)
 102 FORMAT (I3)
 103 FORMAT (2(I4),3X,I3,10X,7(I3),5X,I3)
 104 FORMAT (7(I3),F14.8)

C
 C ① READ CONTROL
 C

198 READ 102, NLITH

C
 C ② READ DATA
 C

DO 199 J=1,NLITH
 199 READ 101, IA(J),ID(J)

C
 C ③ READ ONE IDEAL GENERATOR
 C

200 READ 100, (ITHEM(J),J=1,7),NPIV

C
 C INITIALIZE
 C

DO 208 I=1,4
 208 SUMG(I)=0.
 LOGIC=1
 I=2
 J=0
 M(1)=ITHEM(1)
 M(500)=0

C
 C ③ EXTEND IDEAL INTO THE M(K) POSITIONS
 C

201 K=I-J
 M(I)=ITHEM(K)
 IF (M(500)) 202,202,300
 202 GO TO (1,2),LOGIC
 1 IF (NPIV-K) 203,204,203
 2 IF (NPIV-K) 204,203,204
 203 LOGIC=1
 IF (K-1) 206,205,206
 205 J=J+2
 206 I=I+1
 GO TO 201
 204 LOGIC=2
 IF (K-1) 205,207,205
 207 LOGIC=1
 GO TO 206

300 IF (SENSE SWITCH 2) 700,701
700 PAUSE

C
C
C

(5) FIND STARTING POINTS, MST(I) I=1,4

701 IF (NPIV-5) 311,301,311
301 IF (IA(1)-ITHEM(1)) 302,305,302
302 IF (IA(1)-ITHEM(2)) 303,306,303
303 IF (IA(1)-ITHEM(3)) 304,307,304
304 IF (IA(1)-ITHEM(4)) 309,308,309
305 MST(1)=1
GO TO 497
306 MST(1)=2
MST(2)=8
GO TO 498
307 MST(1)=3
MST(2)=7
GO TO 498
308 MST(1)=4
MST(2)=6
GO TO 498
309 MST(1)=5
GO TO 497
311 IF (IA(1)-ITHEM(1)) 312,315,312
312 IF (IA(1)-ITHEM(2)) 313,316,313
313 IF (IA(1)-ITHEM(3)) 314,317,314
314 IF (IA(1)-ITHEM(4)) 319,318,319
315 NP61=1
MST(1)=1
GO TO 320
316 NP61=2
MST(1)=2
GO TO 320
317 NP61=3
MST(1)=3
GO TO 320
318 NP61=4
MST(1)=4
GO TO 320
319 IF (IA(1)-ITHEM(5)) 346,345,346
320 NIAS=0
DO 322 K=1,6
IF (IA(1)-ITHEM(K)) 322,321,322
321 NIAS=NIAS+1
322 CONTINUE
IF (NIAS-2) 330,331,330
330 GO TO (497,342,343,344),NP61
331 GO TO (351,354,356,357),NP61
342 MST(2)=10
GO TO 498
343 MST(2)=9
GO TO 498

```

344 MST(2)=3
    GO TO 498
345 MST(1)=5
    MST(2)=7
    GO TO 498
346 MST(1)=6
    GO TO 497
351 IF (IA(1)-ITHEM(3)) 352,360,352
352 IF (IA(1)-ITHEM(4)) 353,361,353
353 IF (IA(1)-ITHEM(5)) 363,362,363
354 IF (IA(1)-ITHEM(4)) 355,364,355
355 IF (IA(1)-ITHEM(5)) 366,365,366
356 IF (IA(1)-ITHEM(5)) 368,367,368
357 MST(2)=6
    MST(3)=8
    GO TO 499
360 MST(2)=3
    MST(3)=9
    GO TO 499
361 MST(2)=4
    MST(3)=8
    GO TO 499
362 MST(2)=5
    MST(3)=7
    GO TO 499
363 MST(2)=6
    GO TO 498
364 MST(2)=4
    MST(3)=8
    MST(4)=10
    GO TO 500
365 MST(2)=5
    MST(3)=7
    MST(4)=10
    GO TO 500
366 MST(2)=6
    MST(3)=10
    GO TO 499
367 MST(2)=5
    MST(3)=7
    MST(4)=9
    GO TO 500
368 MST(2)=6
    MST(3)=9
    GO TO 499
497 MST(2)=0
498 MST(3)=0
499 MST(4)=0
    IF (SENSE SWITCH 2) 800,500
800 PAUSE

```

```

C
C ⑥ CALCULATE G PER SUBSEQUENCE
C

```

```

500 NREP=NLITH-6

```

```

IC=1
IF (SENSE SWITCH 2) 900,501
900 PAUSE
501 M1ST=MST(IC)-1
DO 599 I=1,NREP
IGEE(IC)=0
J=0
502 K=M1ST+J+I
IF (M(K)-IA(I)) 503,504,503
503 J=J+1
GO TO 502
504 LMIN=I+1
LMAX=I+6
N=K-I
DO 598 L=LMIN,LMAX
506 LUP=N+L
IF (IA(L)-M(LUP)) 505,598,505
505 N=N+1
IGEE(IC)=IGEE(IC)+1
GO TO 506
598 CONTINUE
C
C ⑥A PUNCH SUBSEQUENCE G IF DESIRED
C
IF (SENSE SWITCH 1) 600,601
600 PUNCH 103, ID(I),ID(LMAX),IGEE(IC),(ITHEM(JA),JA=1,7),IC
C
C ⑥B CONVERT TO FLOATING POINT AND ACCUMULATE
C
601 GEE(IC)=IGEE(IC)
SUMG(IC)=SUMG(IC)+GEE(IC)
599 CONTINUE
REP=NREP
AVGEE(IC)=SUMG(IC)/REP
C
C ⑦ CHECK FOR REMAINING STARTING POSITIONS AND REPEAT IF NECESSARY
C
602 IF (IC-4) 603,604,603
603 IC=IC+1
IF (MST(IC)) 501,605,501
605 AVGEE(IC)=999.
GO TO 602
C
C ⑦ MINIMIZE OVER THE STARTING POINTS
C
604 AVMIN=AVGEE(1)
DO 607 I=2,4
IF (AVMIN-AVGEE(I)) 607,607,608
608 AVMIN=AVGEE(I)
607 CONTINUE
C
C ⑧ OUTPUT
C
PUNCH 104, (ITHEM(K),K=1,7),AVMIN
GO TO 200
END

```

PROGRAM 4 -- AMG over a Sample of Actual Sequences.

- ① Read Control -- number of sequences in sample (here 235)
= NSEQ.
- ② Read Data from cards 1 to NSEQ.
 - a. identification number
 - b. seven digits representing classification of the distinct lithologies in the sequence.
- ③ Read one at a time from the population of ideal generators and extend the given digits through a sequence $M(k)$, $k = 1$ to 100.
- ④ Depending on the lithology repeated in the ideal and the first lithology of the sample sequence, locate the starting points (among the $M(k)$) which may yield different values of G per sample sequence.
- ⑤ Calculate G for each of the NSEQ sample sequences, minimize over starting points, and accumulate.
- ⑥ Divide the accumulation of (5) by NSEQ to obtain AMG.
- ⑦ Punch digits of ideal considered, AMG.
- ⑧ Repeat from (3) until population of ideals is exhausted.

Important symbols.

NSEQ = sample size; number of seven-unit sequences used,
maximum 300.

ISEQ(I,J) = classification (J) of the lithologies in each
sample sequence ($I = 1, NSEQ$).

ITHEM(K), M(K), NP1V, all as in PROGRAM 3.
LOGIC, MST(K), IC

SUMG, AVG = as in PROGRAM 3 but no longer subscripted since
minimization is prior to accumulation.


```

C      CYCLOTHERM PROBLEM
C      PROGRAM 4 -- AMG OVER A SAMPLE OF ACTUAL SEQUENCES
C
      DIMENSION ISEQ(300,7),ITHEM(7),M(100),MST(4),IGEE(4),ISCOD(300)
100  FORMAT (I3)
      01  FORMAT (I3,5X,7(I3))
102  FORMAT (7(I3),5X,I3)
103  FORMAT (I3,10X,I3,10X,7(I3))
104  FORMAT (4HFOR 7(I3),3X,21HAVERAGE MINIMUM G IS F12.8)
C
C      ① READ CONTROL
C
198  READ 100, NSEQ
C
C      ② READ DATA
C
      DO 199 I=1,NSEQ
199  READ 101, ISCOD(I),(ISEQ(I,J),J=1,7)
C
C      ③ READ ONE IDEAL GENERATOR
C
200  READ 102, (ITHEM(K),K=1,7),NPIV
C
C      INITIALIZE
C
      M(1)=ITHEM(1)
      M(100)=0
      LOGIC=1
      I=2
      J=0
      ISUMG=0
C
C      ③ EXTEND IDEAL INTO THE M(K) POSITIONS
C
201  K=I-J
      M(I)=ITHEM(K)
      IF (M(100)) 202,202,300
202  GO TO (1,2),LOGIC
      1  IF (NPIV-K) 203,204,203
      2  IF (NPIV-K) 204,203,204
203  LOGIC=1
      IF (K-1) 206,205,206
205  J=J+2
206  I=I+1
      GO TO 201
204  LOGIC=2
      IF (K-1) 205,207,205
207  LOGIC=1
      GO TO 206
300  DO 599 IA=1,NSEQ
      IF (SENSE SWITCH 2) 700,701
700  PAUSE
C
C      ④ FIND STARTING POINTS, MST(I) I=1,4

```

C

```
701 IF (NP1V-5) 311,301,311
301 IF (ISEQ(IA,1)-ITHEM(1)) 302,305,302
302 IF (ISEQ(IA,1)-ITHEM(2)) 303,306,303
303 IF (ISEQ(IA,1)-ITHEM(3)) 304,307,304
304 IF (ISEQ(IA,1)-ITHEM(4)) 309,308,309
305 MST(1)=1
   GO TO 497
306 MST(1)=2
   MST(2)=8
   GO TO 498
307 MST(1)=3
   MST(2)=7
   GO TO 498
308 MST(1)=4
   MST(2)=6
   GO TO 498
309 MST(1)=5
   GO TO 497
311 IF (ISEQ(IA,1)-ITHEM(1)) 312,315,312
312 IF (ISEQ(IA,1)-ITHEM(2)) 313,316,313
312 IF (ISEQ(IA,1)-ITHEM(3)) 314,317,314
314 IF (ISEQ(IA,1)-ITHEM(4)) 319,318,319
315 NP61=1
   MST(1)=1
   GO TO 320
316 NP61=2
   MST(1)=2
   GO TO 320
317 NP61=3
   MST(1)=3
   GO TO 320
318 NP61=4
   MST(1)=4
   GO TO 320
319 IF (ISEQ(IA,1)-ITHEM(5)) 346,345,346
320 NIAS=0
   DO 322 K=1,6
   IF (ISEQ(IA,1)-ITHEM(K)) 322,321,322
321 NIAS=NIAS+1
322 CONTINUE
   IF (NIAS-2) 330,331,330
330 GO TO (497,342,343,344),NP61
331 GO TO (351,354,356,357),NP61
342 MST(2)=10
   GO TO 498
343 MST(2)=9
   GO TO 498
344 MST(2)=8
   GO TO 498
345 MST(1)=5
   MST(2)=7
   GO TO 498
346 MST(1)=6
   GO TO 497
```

```

351 IF (ISEQ(IA,1)-ITHEM(3)) 352,360,352
352 IF (ISEQ(IA,1)-ITHEM(4)) 353,361,353
353 IF (ISEQ(IA,1)-ITHEM(5)) 363,362,363
354 IF (ISEQ(IA,1)-ITHEM(4)) 355,364,355
355 IF (ISEQ(IA,1)-ITHEM(5)) 366,365,366
356 IF (ISEQ(IA,1)-ITHEM(5)) 368,367,368
357 MST(2)=6
    MST(3)=8
    GO TO 499
360 MST(2)=3
    MST(3)=9
    GO TO 499
361 MST(2)=4
    MST(3)=8
    GO TO 499
362 MST(2)=5
    MST(3)=7
    GO TO 499
363 MST(2)=6
    GO TO 498
364 MST(2)=4
    MST(3)=8
360
    GO TO 500
365 MST(2)=5
    MST(3)=7
    MST(4)=10
    GO TO 500
366 MST(2)=6
    MST(3)=10
    GO TO 499
367 MST(2)=5
    MST(3)=7
    MST(4)=9
    GO TO 500
368 MST(2)=6
    MST(3)=9
    GO TO 499
497 MST(2)=0
498 MST(3)=0
499 MST(4)=0
    IF (SENSE SWITCH 2) 800,500
800 PAUSE

```

C
C (5) CALCULATE G PER SAMPLE SEQUENCE AND OVER STARTING POINTS
C

```

500 DO 598 IC=1,4
    IF (MST(IC)) 501,598,501
501 IGEE(IC)=0
    N=MST(IC)-1
    DO 597 L=2,7
506 LUP=N+L
    IF (ISEQ(IA,L)-M(LUP)) 505,597,505
505 N=N+1
    IGEE(IC)=IGEE(IC)+1
    GO TO 506

```

```
597 CONTINUE
   IF (IC-1) 601,600,601
```

```
C
C   (5) MINIMIZE OVER STARTING POINTS
C
```

```
600 IGM=IGEE(1)
     GO TO 598
601 IF (IGM-IGEE(IC)) 598,598,602
602 IGM,IGEE(IC)
598 CONTINUE
```

```
C
C   (5) SUM MINIMUM G OVER THE SAMPLE SET
C
```

```
ISUMG=ISUMG+IGM
IF (SENSE SWITCH 1) 603,599
```

```
C
C   PUNCH MINIMUM G PER SAMPLE SEQUENCE IF DESIRED
C
```

```
603 PUNCH 103, ISCOD(IA),IGM,(ITHEM(J),J=1,7)
599 CONTINUE
     SEQ=NSEQ
     SUMG=ISUMG
```

```
C
C   (6) AVERAGE
C
```

```
AVG=SUMG/SEQ
```

```
C
C   (7) OUTPUT
C
```

```
PUNCH 104, (ITHEM(J),J=1,7),AVG
GO TO 200
END
```


APPENDIX B. An Alternative to MAG.

Let us first express MAG somewhat more generally as:

$$\bar{G}_L = \frac{\sum_{j=1}^{n-L+1} \left(\sum_{i=j}^{j+L-2} g_i \right)}{n-L+1}$$

where n = number of lithologies in an extended sequence,

L = length of subsequence (7 for MAG),

g_i = increment of the total G due to the transition from a_i to a_{i+1} .

Then given the minimizing initial choice of transgression or regression, $\bar{G}_L = \text{MAG}$, as used in the body of this paper.

The reasons for using this kind of measure in Program 3 were:

- a) to make MAG magnitudes comparable to those obtained for 7-unit sequences, and
- b) to make the subsequence G 's easily available from the Program, as for the purposes of Figure 1 subsequence G 's were of interest in their own right.

It was only after this procedure had been adopted and applied that a certain disadvantage became clear.

An alternative procedure for satisfying (a) above would have been:

$$\underline{G}_L = (L-1) \frac{\sum_{i=1}^{n-1} g_i}{n-1}$$

i.e. the number of transitions in a subsequence of length L times the average increment per transition in the extended sequence.

Furthermore, \underline{G}_L is the better measure in the sense that it makes use of all the information in the extended sequence and gives equal weight to each g_i .

On the contrary, \bar{G}_L is "biased" in the sense that g_1 and g_{n-1} are used only once, g_2 and g_{n-2} are used only twice, . . . , g_{L-2} and g_{n-L+2} are used only $L-2$ times. All the remaining g_i (if any) are used $L-1$ times, and \bar{G}_L would be the same as G_L if it were not for this "bias" with reference to the first and last $L-2$ terms.

This observation leads logically to the question of how one might convert from \bar{G}_L to G_L . A little reflection reveals that

$$\sum_{j=1}^{n-L+1} \left(\sum_{i=j}^{j+L-2} g_i \right) + \sum_{i=1}^{L-2} (L-i-1) g_i + \sum_{i=n-L+2}^{n-1} (L-n+i-1) g_i = (L-1) \sum_{i=1}^{n-1} g_i \quad (A)$$

which holds for certain conditions on n, L .

To illustrate take an arbitrary sequence with $n=15$ and $L=7$ and find the g_i with reference to the Moore ideal.

a_i	g_i	first term in (A)	correction terms
1			
2	0	1(0)	+5(0)
5	2	2(2)	+4(2)
3	1	3(1)	+3(1)
2	0	4(0)	+2(0)
1	0	5(0)	+1(0)
2	0	6(0)	
5	2	6(2)	
3	1	6(1)	
4	4	6(4)	
3	2	5(2)	+1(2)
4	4	4(4)	+2(4)
3	2	3(2)	+3(2)
2	0	2(0)	+4(0)
5	4	1(4)	+5(4)
sums	22	85	11 + 36

Whence by (A) $85 + 11 + 36 = 6(22) = 132$.

In the example, the middle g_i for which no correction is necessary are 0,2,1,4 (for $i = 6,7,8,9$). In general, the number of such g_i is: $(n-1) - 2(L-2)$. It is easy to see from the example that (A) will hold whenever the two groups of correction terms do not "overlap", i.e. when

$$(n-1) - 2(L-2) \geq 0$$

$$n \geq 2L - 3$$

For $L=7$, the particular case of interest in this paper, $n > 11$ is sufficient to guarantee that (A) holds. In the section concerning the idealized composite sequence, the value of n is 273.

Expressing (A) in terms of \bar{G}_L and \underline{G}_L we have:

$$(n-L+1) \bar{G}_L + \sum_{i=1}^{L-2} (L-i-1) g_i + \sum_{i=n-L+2}^{n-1} (L-n+i-1) g_i = (n-1) \underline{G}_L \quad (B),$$

and for the case of interest where $L=7$:

$$(n-1) \underline{G}_7 - (n-6) \bar{G}_7 = \sum_{i=1}^5 (6-i) g_i + \sum_{i=n-5}^{n-1} (i-n+6) g_i \quad (C).$$

Let the right-hand side of (C) be called $k(I)$, a positive integer depending on the particular ideal in question and on the first six and last six a_i of the extended sequence. Then

$$(n-1) \underline{G}_7 - (n-6) \bar{G}_7 = k(I).$$

For any practical problems it should be possible to show that some

$$k_1 < k(I) < k_2. \quad (D)$$

Even for the purposes of this paper, however, the procedure is laborious. It involves enumeration of all distinct configurations of ideal hemicycles and comparison of these

with the leading and trailing six a_i of the extended sequence used. Details are omitted, but it can be shown that for the idealized composite sequence described in the body of this paper $10 < k(I) < 134$. In the same part of the study, as previously mentioned, $n = 278$. Finally, the results showed that $7.897 < \bar{G}_7 (=MAG) < 20.717$. So that

$$k_1 < n\underline{G}_7 - \underline{G}_7 - n\bar{G}_7 + 6\bar{G}_7 < k_2$$

$$k_1 - 5\bar{G}_7 < n(\underline{G}_7 - \bar{G}_7) - (\underline{G}_7 - \bar{G}_7) < k_2 - 5\bar{G}_7$$

$$\frac{k_1 - 5\bar{G}_7}{n - 1} < \underline{G}_7 - \bar{G}_7 < \frac{k_2 - 5\bar{G}_7}{n - 1}$$

$$\frac{10 - 5(20.717)}{277} < \underline{G}_7 - \bar{G}_7 < \frac{134 - 5(7.897)}{277}$$

$$-.34 < \underline{G}_7 - \bar{G}_7 < +.38$$

(D)

provides bounds for the difference between the two measures.

A few calculations were carried out using (C) and values of MAG actually obtained from PROGRAM 3. In most cases, the difference $|\underline{G}_7 - \bar{G}_7|$ is much less than the absolute extremes of (D). Values appear to be concentrated around .02. The largest actually calculated value was approximately .17. For sample sizes above 200, then, the difference between the alternative measures of average G over an extended sequence is essentially negligible. For smaller sample, \underline{G}_7 might be preferable.

FINDING THE IDEAL CYCLOTHEM

by

WILLIAM C. PEARN

B. S., University of Kansas, 1954

M. S., University of Kansas, 1959

Ph.D., University of Kansas, 1963

AN ABSTRACT OF
A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

Rock sequences can be translated into numerical sequences by associating with each recognized lithology an integer (1 to 5) according to a fixed classification scheme. Finite numerical sequences, corresponding to actual measured sections, can then be compared to infinite numerical sequences, corresponding to ideal cyclothem repetitions. If an actual sequence is considered to be a fragmentary ideal, with some lithologies missing owing either to non-deposition or subsequent removal, the deviation of the actual from the ideal can be measured by a discordance index defined as the minimum value of the number of missing lithologies. The ideal sequence which best explains the overall characteristics of a sample of finite sequences is the ideal for which the average value of the discordance index is least.

In this manner, the "best" ideal cyclothem for an area and stratigraphic interval of interest can be determined from arithmetic operations on a sample of actual rock sequences derived from measured sections within the area and interval. The method is developed, and application is made to measured sections from the Missourian-Wolfcampian interval of northeast Kansas.

Of particular interest is the ideal sequence which corresponds to the ideal cyclothem proposed by Moore for this region. Results seem to indicate that the best ideal cyclothem for the area-interval considered would be quite similar to that proposed by Moore. However, the classification used was

clearly inadequate. Deficiencies of the classification are discussed briefly.

A short sermon is preached, in which present conclusions are dismissed while the general methods and approach of this study are highly recommended.