

---

An Application of Cluster Analysis as a Method of Determining Biofacies

Author(s): James F. Mello and Martin A. Buzas

Reviewed work(s):

Source: *Journal of Paleontology*, Vol. 42, No. 3 (May, 1968), pp. 747-758

Published by: [SEPM Society for Sedimentary Geology](#)

Stable URL: <http://www.jstor.org/stable/1302372>

Accessed: 15/04/2012 01:30

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



SEPM Society for Sedimentary Geology is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Paleontology*.

<http://www.jstor.org>

## AN APPLICATION OF CLUSTER ANALYSIS AS A METHOD OF DETERMINING BIOFACIES

JAMES F. MELLO AND MARTIN A. BUZAS

U.S. Geological Survey and Smithsonian Institution, Washington, D.C.

**ABSTRACT**—Data on 44 species of Recent Foraminifera in 99 samples collected from three traverses off the central Texas coast are re-examined using the Jaccard Coefficient for data comparison and the weighted pair group method with simple arithmetic averages for clustering of coefficients. The purpose of this study is to evaluate the utility of the method and to compare the results obtained with other appraisals of the same data. The technique considers only presence or absence and is simple to understand. The analyses show good agreement of station (sample to sample) distribution compared with the results of previous investigations, and interspecies relationships are meaningful. The method, although not statistical, is a useful, rapid, and inexpensive empirical tool for analysis of data.

### INTRODUCTION

ONE of the more difficult problems in the study of faunal and floral populations is the discovery of meaningful patterns. All the various methods by which such patterns are searched for are here referred to as biofacies analysis. The usual procedure in biofacies analysis involves the careful examination of data tables depicting the abundance, or at least the occurrence, of species in samples. The extent of such an examination depends on the researcher's aims and the volume and complexity of the data. The final results, the meaningful patterns revealed by the data, are always partly intuitive; the requisite for success, here called meaningful, is basically an intuitive appreciation of relationships among the data. Except for small bodies of data, the calculation of even simple kinds of potentially "meaningful" relationships was, until recently, far too time consuming. However, the large electronic computer permits the researcher to utilize large amounts of data and to undertake calculations which were impossible only a few years ago. Computers not only have made it possible to handle much larger bodies of data within the traditional intuitive framework but also have opened up broader possibilities which require assessments not based on past experience. In other words, the capacity of the computer to make comparisons forces new thought with regard to many kinds of relationships which could not be assessed before and thus which are without clear "meaning" in the context of past experience.

The purpose of this study is to evaluate the utility of a technique for biofacies analysis in which data comparisons, calculated as Jaccard Coefficients, are clustered by the weighted pair group method with simple arithmetic averages (WPGMA). The method requires only presence or absence data and is very simple in terms of computation and comprehension. The data

chosen for the analyses are from a study by Phleger (1956) on the total and living foraminiferal populations off the central Texas coast.

Kaesler (1966) applied the same method successfully to ostracode and foraminifer data from Todos Santos Bay, and Maddocks (1966) successfully used it on ostracode data from the Nosey Bé area of Madagascar. We are testing the method's utility again here because we are dealing with a different environmental setting, and because we feel that before the method can be used with confidence in paleoecology more than one or two trials should be attempted in the Recent. We chose Phleger's (1956) data because they have already been analyzed by Phleger, who based his interpretations on careful examination, and by Buzas (1967), who used a multivariate statistical method. The three methods of analysis are different, as are the criteria for discrimination. The results of the two earlier analyses provide *a priori* knowledge of the expected outcome, and, at the same time, provide an opportunity to compare the different methods.

### ACKNOWLEDGMENTS

We wish to thank Rosalie Maddocks and Richard Boardman of the Smithsonian Institution and J. Thomas Dutro of the U.S. Geological Survey for critically reviewing the manuscript. Thanks are also due to Louis Purnell who drafted the figures.

The arrangement of our names in the citation is arbitrary, and no seniority is intended. Publication has been approved by the Director, U. S. Geological Survey.

### METHOD

We will discuss the calculations used in this study in terms of the compared entities. In sample by sample comparison—the *Q*-mode described by Sokal & Sneath, (1963, p. 124)—each sample is compared with every other sample on

TABLE 1—Data matrix. +, species present; -, species absent.

		samples			
		A	B	C	D
s	1	+	+	-	-
p	2	+	+	-	+
e	3	+	-	+	+
i	4	+	+	+	-
e	5	-	+	-	+

the basis of species presence or absence in each. In species by species comparison, the R-mode described by Sokal & Sneath (1963, p. 124), each species is compared with every other species on the basis of their presence in the respective samples. An example of a data matrix, with presence denoted by + and absence by -, is shown in table 1.

TABLE 2—Coefficients of similarity.

Ssm	Simple Matching Coefficient
Sj	Jaccard Coefficient
Ss	Simpson's Index 2
WHERE, for sample by sample (Q-mode) analysis:	
NJK	number of species contained in both samples being compared
NjK	number of species present in sample K and absent in sample J
NjK	number of species present in sample J and absent in sample K
Njk	number of species in the total fauna being considered but absent in both samples J and K
N min (J, K)	number of species in the smaller sample, J or K.
Ss	$\frac{NJK}{Nmin (J, K)}$
Ssm	$\frac{NJK + Njk}{NJK + Njk + NjK}$
Sj	$\frac{NJK}{NJK + Njk + NjK}$

We considered three coefficients for use in this study. The first, called the Coefficient of Jaccard by Sokal & Sneath (1963, p. 133), can be used to compare pairs of entities (samples, for example) on the basis of the presence or absence of attributes of those samples (species in this case). Positive matches and mismatches of the attributes are taken into account in the coefficient, but negative matches are not (table 2). The number of occurrences in common in the two entities being compared is divided by the total number of unique occurrences in both entities. An example of the application of this coefficient is shown in table 3. We also considered the use of the Simple Matching Coefficient (Sokal & Sneath, 1963, p. 133) in which the number of negative matches is introduced into both the numerator and denominator (see table 2). Finally, Simpson's Index 2 (1960, p. 302) in which the number of occurrences in common between two entities is divided by the number of occurrences of the smaller entity (see table 2) was considered. Reasons for our choice of the Jaccard Coefficient are discussed later.

Computation and clustering were carried out on an IBM 7094 computer using a program written in Fortran IV. The Program is available from the authors.

PREVIOUS WORK

The data for the present study were taken from a paper by Phleger (1956) on Recent Foraminifera off the central Texas coast. The samples are from the three traverses shown in text-figure 3. Traverse III has 29 stations, Traverse V has

TABLE 3—Calculated examples of similarity coefficients, using the data of text-fig. 1.

		Coefficients		
		Ss	Ssm	Sj
A	compared with B	.75	.80	.60
A	" " C	1.00	.80	.50
A	" " D	.66	.70	.40
B	" " C	.50	.60	.20
B	" " D	.66	.70	.40
C	" " D	.50	.60	.25

Note: For Ssm we assume a total fauna of 10 species

88, and Traverse VII has 31. Because of the limitations of the available computer, every third station had to be deleted in our analysis, so that Traverse III is represented by 19 stations, Traverse V by 58, and Traverse VII by 21. Phleger (1956) recorded 75 species from the area. Buzas (1967) deleted those species which were very rare and used only 44 species. The same 44 species, listed in our text-figures 5 and 6, were used in this study.

Phleger (1956, p. 111) recognized faunal assemblages (biofacies) by means of examination and comparison of species depth ranges, shown in his figures 17–20. Delimitation of each depth-related biofacies, and the assignment of samples to it, was based on:

... consideration of the percent of total population within a single sample, frequency trends between samples, consistency of occurrence in depth, and relative number of samples in which each species was present.

Phleger (1956) recognized biofacies boundaries at 20–30 m, 50–70 m, and 100 m. At the same time he indicated that the living population increases in number of individuals toward the northeast (Traverse VII).

Location of the traverses, and Phleger's suggested faunal boundaries, allowed Buzas (1967) to subdivide arbitrarily the area into five subareas: Traverse III 0–30 m, Traverse V 0–30 m, 30–60 m, 60–110 m, and Traverse VII 0–30 m. In his study Buzas did not use Traverse IV because it contained only 7 stations and did not use the 100 m boundary suggested by Phleger because only 5 stations, from Traverse V, occur deeper than 100 m. The mean vectors of species abundances for the five subareas were then compared using a statistical method called canonical analysis. The results of the canonical analysis showed that for the *total* population all three 0–30 m subareas are similar, whereas the 30–60 m subarea and the 60–110 m subarea are each distinct. These results are completely compatible with Phleger's separation of the foraminiferal fauna into biofacies. Canonical analysis of the *living* population showed that the 0–30 m subareas of Traverse III and Traverse V are similar, whereas subareas 0–30 m of traverse III, 30–60 m of Traverse V, and 60–110 m of Traverse V are each distinct. This analysis indicates that an additional biofacies, occupying Traverse III, 0–30 m, is present in the living population. Phleger noted a larger living population in samples from this interval of Traverse III but did not recognize them as comprising a separate biofacies, whereas in the canonical analysis, which uses species abundances as a criterion, this separation was made. On the whole, agreement between the two studies, differing as they do in

methodology and criteria of discrimination, is surprisingly good.

In any area, such as the Texas Gulf Coast, determination of the total number of species would require examination and species assignment of all individuals in the area. If this task were to be attempted, it would be found that in the early stages of specimen examination the number of species would increase rapidly, but in later stages, as the number of individuals grew large, the addition of a new species would be a rare event. The same relationship pertains, of course for individual samples of an area. This relationship between the number of individuals and the number of species has been the subject of much research and is discussed by Preston (1962).

When a series of samples are taken on a one-mile square grid, say, we assume that each sample represents a square mile. Buzas (1965, 1968) and Lynts (1966) have shown that the spatial distribution of Foraminifera is often heterogeneous. Therefore in taking only one sample, one may, by chance, have sampled an area with few individuals, although nearby, within the same square mile, there are abundant individuals. Clearly, the abundance and the number of species are related and an accurate estimate of either one requires more than one sample. Because of spatial heterogeneity we believe it is too much to assume that areas represented by single samples can be considered adequately sampled. Consequently, we would like to put more emphasis, as one does intuitively, on those samples which contain many individuals and of course at the same time many species. The Jaccard Coefficient does this. We have selected it as a compromise between Simpson's Index 2 which we believe is too insensitive to sample diversity and the Simple Matching Coefficient which can be too sensitive to inadequate sampling.

To our knowledge the first use of the quantitative techniques applied to this study for ecological analysis was by Kaesler (1966). Kaesler (1966, p. 23) notes three assumptions which are made in the kind of analysis attempted by both him and us:

- 1) Biofacies and biotopes exist in the study area;
- 2) A sample adequately represents the population of organisms at a station;
- 3) Biotopes are mappable.

However, Kaesler (1966) used the term biofacies for the species lists which result from clustering by the R-mode. We suggest that this usage is too restrictive and have applied the term biofacies in the more conventional sense throughout this paper. We use the term biofacies when talking about an area which is defined by species or when talking about the species which are con-

tained in it. We also use the term regardless of the criteria by which it is defined, that is presence or absence, abundance, relative abundance, etc. This kind of usage requires, of course, that the biofacies be mappable.

In his biofacies analysis (sample by sample comparisons), Kaesler (1966, p. 31) used the Simple Matching Coefficient, reasoning that the absence of species in any two samples being compared is as important as their presence in that the mutual absence suggests that both samples are from environments inhospitable to that species. For reasons stated above we prefer the Jaccard Coefficient. Kaesler actually carried out his analyses using both coefficients and the results are in agreement. Maddocks (1966, p. 15) who used both Jaccard and Simple Matching Coefficients for sample by sample analysis of live ostracode populations found that the results obtained using the Simple Matching Coefficient were not interpretable ecologically, although results from both coefficients were congruent in dealing with subfossil samples.

Kaesler (1966) used the Jaccard Coefficient for species by species analysis. In support of the exclusion of negative matches he noted that:

Whereas the absence of both species A and species B at station 1 is of ecologic interest, it provides no useful information for clustering species into biofacies.

Perfect similarity between two species caused by negative matches alone would not justify grouping the species in the same biofacies, so negative matches must be ignored.

We concur with this reasoning and have used the Jaccard Coefficient in our species by species analysis.

#### CLUSTERING TECHNIQUE

Computers can compare limitless numbers of samples or species with each other by means of similarity coefficients, but the matrix of values produced when more than a few tens of items are compared is beyond efficient human absorption and interpretation. In order to make the results more intelligible a means for relating the values embodied in the matrix, called clustering, must be used. The storage capacity for the program and computer which we used was a matrix of  $99 \times 99$  items. In view of this limitation it was necessary, as already mentioned, to reduce the number of samples by one-third before they could be considered.

Clustering was carried out using the weighted pair group method with simple arithmetic averages (Sokal & Sneath, 1963, p. 182-185, 189-194, 309-310). In this method the matrix of similarity coefficients is examined, and those samples (or species) which have the highest re-

lationships with each other are grouped together. This initial clustering step relates all samples which have values higher than that value at which a third sample becomes eligible to join an initial cluster, at which point the entire similarity matrix is recalculated. The relationship of each pair to all other pairs and as yet unclustered samples is the arithmetic average of the similarity coefficient values between each member of the pair and each member of all other pairs or individual samples. In the next step new clusters are formed in the same fashion as were the original clusters except that the admission of new clusters or individuals to already existing clusters is based on the arithmetic average value of the potential member with the members of the already existing cluster. It is in this and subsequent clustering steps that weighting takes place in that the arithmetic average of the values between the potential new member and *all* members of the existing cluster which it will join is the factor upon which admission is based. A simple example of the weighting and arithmetic averaging steps in clustering is shown in table 4.

TABLE 4—Clustering steps applied to Jaccard Coefficient matrix generated in text-fig. 3.

		original matrix			
	A				
	B	.60			
	C	.50	.20		
	D	.40	.40	.25	
		A	B	C	D
Step 1: cluster					
	A    B				
	.6				
Step 2: recalculate					
	A+B				
	C .35				
	D .40 .25				
	A+B    C    D				
Step 3: cluster					
	A    B    D				
	.6				
	.4				
Step 4: recalculate					
	A+B+D				
	C .30				
	A+B+D    C				
Step 5: cluster					
	A    B    D    C				
	.6				
	.4				
	.3				

The process of clustering and recalculation is repeated until all samples have been clustered or until the relationship between remaining unclustered samples or groups of samples is zero. After completion of clustering the results are compiled and printed out in a dendrogram reflecting the calculated relationships.

The true relationships between 99 related variables requires 98-dimensional space for accurate representation. To represent them in two-dimensional space is to distort true relationships. Sokal & Sneath (1963, p. 189) report that Sokal & Rohlf (1962) have found that the weighted pair group method with arithmetic averages gave the highest correlation (i.e., the least distortion among the methods tested) with the original correlation coefficients (0.86). Although the weighted pair group method with arithmetic averages was used for clustering sample and species data like ours by Kaesler (1966) and by Maddocks (1966), Sokal & Sneath (1963) discuss it in terms of its use in a strictly hierarchical biologic system. It is possible that some other clustering method is more suitable to our data, but too little is known about the degree and effect of distortion from any of the available methods as applied to these kinds of data to permit complete confidence in the selection of any one of them.

#### DENDROGRAM INTERPRETATION

The selection of a suitable level at which the dendrogram reflects meaningful relations between samples (or species) presupposes that such meaningful relations exist over the sampled area and that the data are good enough to reflect them. Natural discontinuities in the dendrogram more or less objectively define groups of samples or species as belonging together, but other criteria, such as a requisite number of clusters, or the closest match of clusters to other criteria, may be used for the subdivision. Interpretations of clustered data by Kaesler (1966) and Maddocks (1966) have been based on the selection of a single value or level on the dendrograms above which relationships are considered to be significant. Although we followed the same procedure in our study, we are not sure that there is any compelling reason to use a single level of demarcation in nontaxonomic analysis. It might well happen that clusters chosen at several levels within a single dendrogram might more closely approach reasonable sample or species arrangements. Kaesler (1966, p. 33), while not explicitly advocating this procedure, states that:

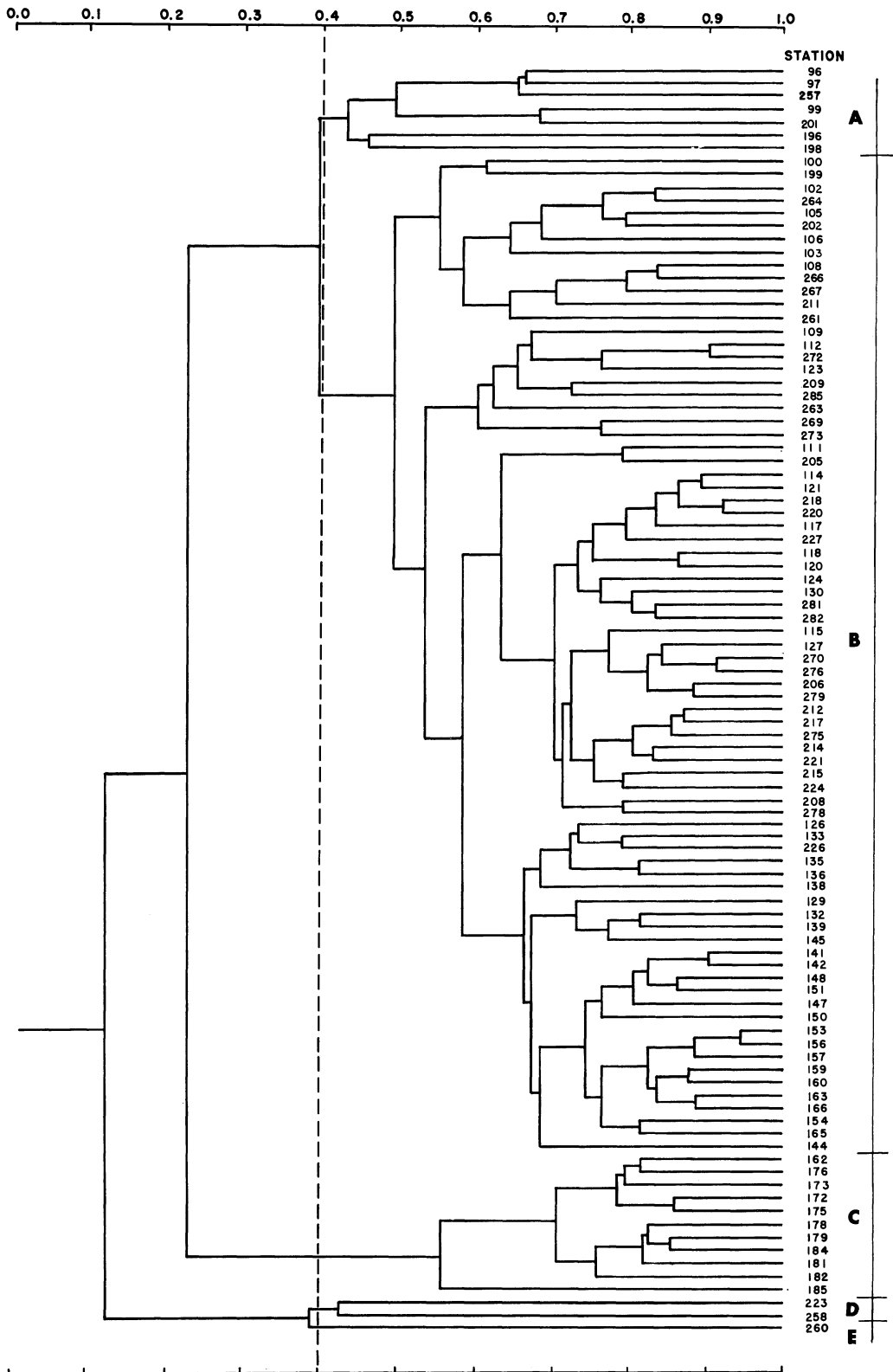
The best procedure in biofacies analysis is probably to avoid drawing lines and to let the dendrograms stand alone as representation of similarity.

We have re-examined a body of carefully accumulated and well-studied data for the purpose of assessing a simple and rapid method of analysis of species and sample relationships. We believe that, when adequate data are used, the method gives reasonable results. The utility of the method as determined in this re-evaluation of Phleger's (1956) data, and the details of the results, are explained below.

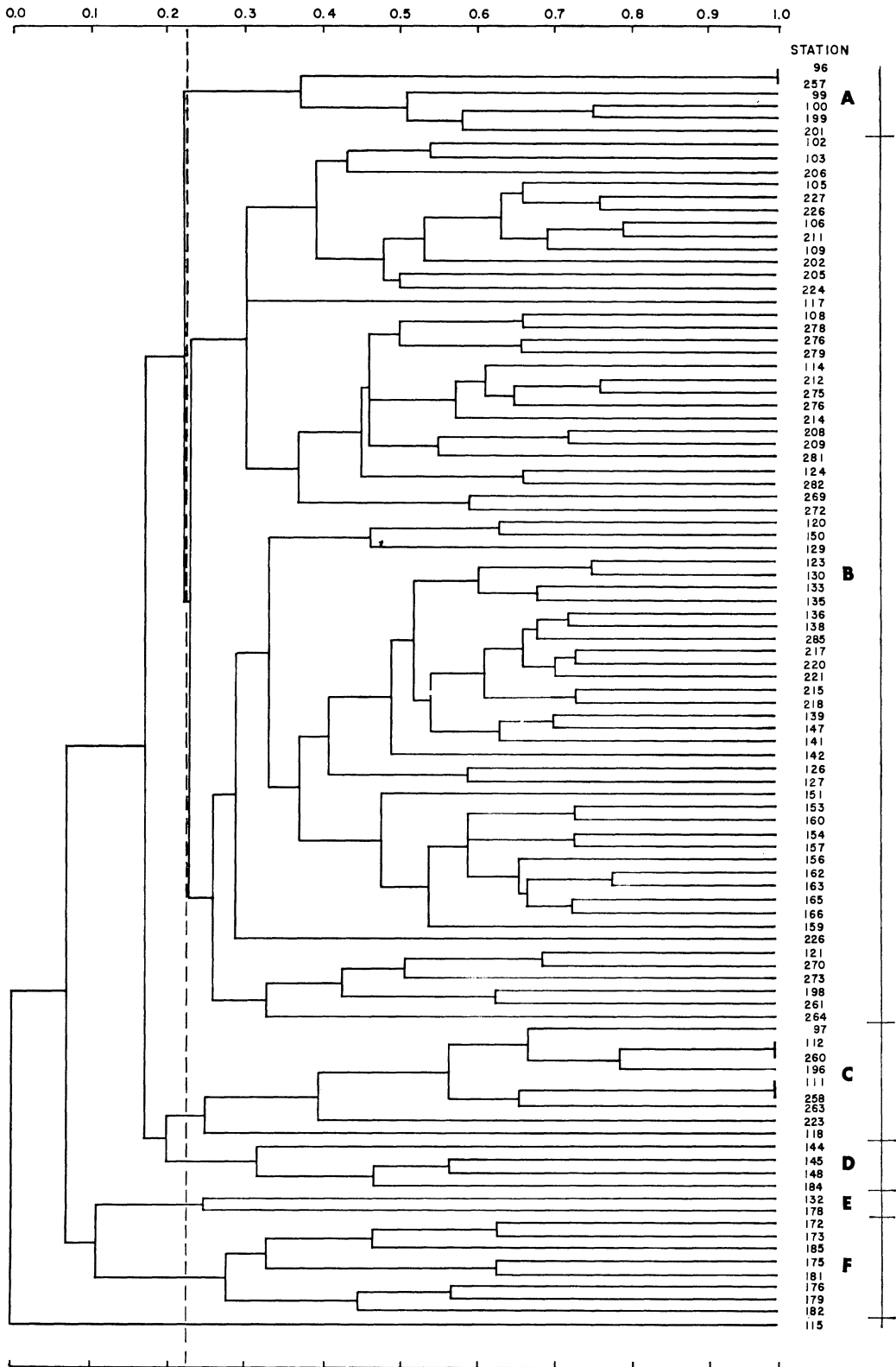
#### RESULTS

*Q-mode—Total population.*—Results of the Q-mode (sample by sample) analysis for the total population are shown in the dendrogram of text-figure 1. The clusters at the 0.40 level are shown areally in text-figure 3. At this level there are three large clusters, a fourth consisting of two stations, and a fifth consisting of one station. The boundary between clusters A and B shown in text-figure 3 is at a depth of about 10 m. The boundary between B and C is at about 77 m. Stations 223 and 258 which make up cluster D have fewer species than the stations in cluster A. Examination of the data matrix shows that station 260, cluster E, is quite similar to stations in cluster D. This similarity is shown in the dendrogram, but because cluster E joins cluster D at a lower level than the chosen 0.40 phenon line, it must be treated separately. Relationships between the major clusters A, B, and C (text-fig. 1) show that clusters A and B are more closely related to each other than they together are related to the deeper water cluster C. Although not shown here, maps were drawn for the 0.50 and 0.55 phenon levels also. The basic pattern remains the same but as the phenon level increases the number of smaller clusters present within larger areas increases.

*Q-mode—Living population.*—Results of the Q-mode analysis for the living population are shown in the dendrogram of text-figure 2. Areal representation of clustering at the 0.239 level is shown in text-figure 4. The three larger areas shown in text-figure 4 agree well with the three larger areas of the total population shown in text-figure 3. The boundary between clusters A and B is at about 12 m. The boundary between clusters B and F is at about 77 m. As can be seen from text-figure 4, there are more representatives of small clusters within the areas defined by the large clusters than there were for the total population. Clustering at levels below 0.239, for example at 0.1, eliminates not only the small clusters C, D and E, but also eliminates the distinction between A and B which we wish to retain for the sake of compatibility with the distribution of total population clusters. As the level of discrimination is increased above 0.239 more small clusters appear within the larger

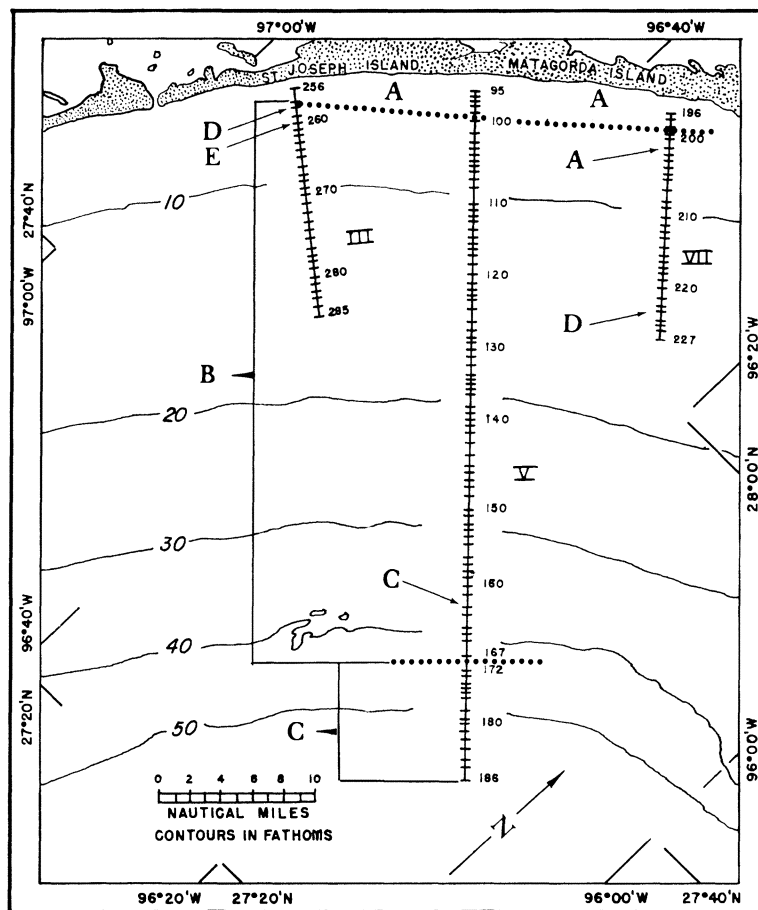


TEXT-FIG. 1—Dendrogram for Q-mode analysis of the total population.



TEXT-FIG. 2—Dendrogram for Q-mode analysis of the living population.





TEXT-FIG. 3—Areal representation of Q-mode clusters for the total population.

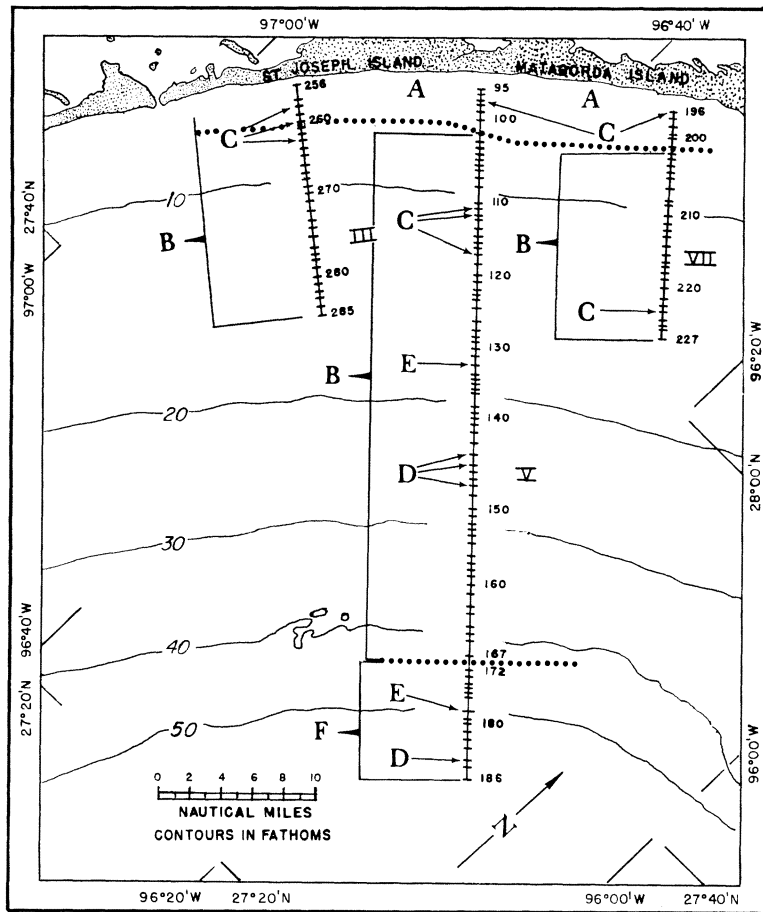
areas and eventually the results become unintelligible.

*R-mode—total population.*—The dendrogram for the R-mode (species by species) analysis of the total population (text-fig. 5) shows that at the 0.200 level there are four clusters, two of which are composed of one species each. The first of these single-species “clusters” contains *Elphidium matagordanum*, which is very rare in occurrence. The second contains *Palmerinella palmerae*, which does not occur at all in the reduced list of stations included in this study. An indication, although not proof, that this species was not present in any of the samples is that it fails to join any other cluster.

Examination of the data matrix shows that the large cluster consisting of *Ammobaculites dilatatus* through “*Rotalia*” *rolshauseni* contains species which either occur throughout the area or are found only in shallow water. The subcluster *Bigenerina irregularis* through *Virgulina*

*spincostata* contains species which occur often and throughout the area. *Ammobaculites dilatatus* and *A. exiguus*, which independently join the cluster at low levels, were found in only a few samples as were the species from *Bolivina lowmani* through “*Rotalia*” *rolshauseni*. The subcluster *Ammoscalaria pseudospiralis* through *Buccella hannai* occurs only in a few samples from deeper water.

The large cluster *Angulogerina bella* through *Eponides umbonatus* contains species which occur more frequently in deep than in shallow water. The subcluster *Angulogerina bella* through *Bulimina marginata* contains frequently occurring species most often found in deeper water samples. The species in the subcluster *Seabrookia earlandi* and *Stetsonia minuta* occur infrequently but most often in deeper water samples. Finally, the subcluster containing *Bolivina fragilis* through *Eponides umbonatus* consists of species very much restricted to the deeper water samples



TEXT-FIG. 4—Areal representation of Q-mode clusters for the living population.

and less frequently occurring than those of the *Angulogerina bella*–*Bulimina marginata* subcluster.

In summary, the major subcluster *Bigenerina irregularis*–*Virgulina spinicostata*, consisting of frequently occurring and widely distributed species, dominates the major cluster *Ammobaculites dilatatus*–“*Rotalia*” *rolshauseni* and has several infrequently occurring species associated with it. The two major subclusters in the major cluster *Angulogerina bella*–*Eponides umbonatus* both contain species found most commonly in deeper waters.

*R-mode—Living population.*—The dendrogram for the R-mode of the living population is shown in text-figure 6. There are six clusters at the 0.020 level. The first cluster consists of species *Ammobaculites dilatatus* through *Seabrookia earlandi*. Examination of the data matrix indicates that the subcluster *Ammobaculites dilatatus* through *Stetsonia minuta* contains species which

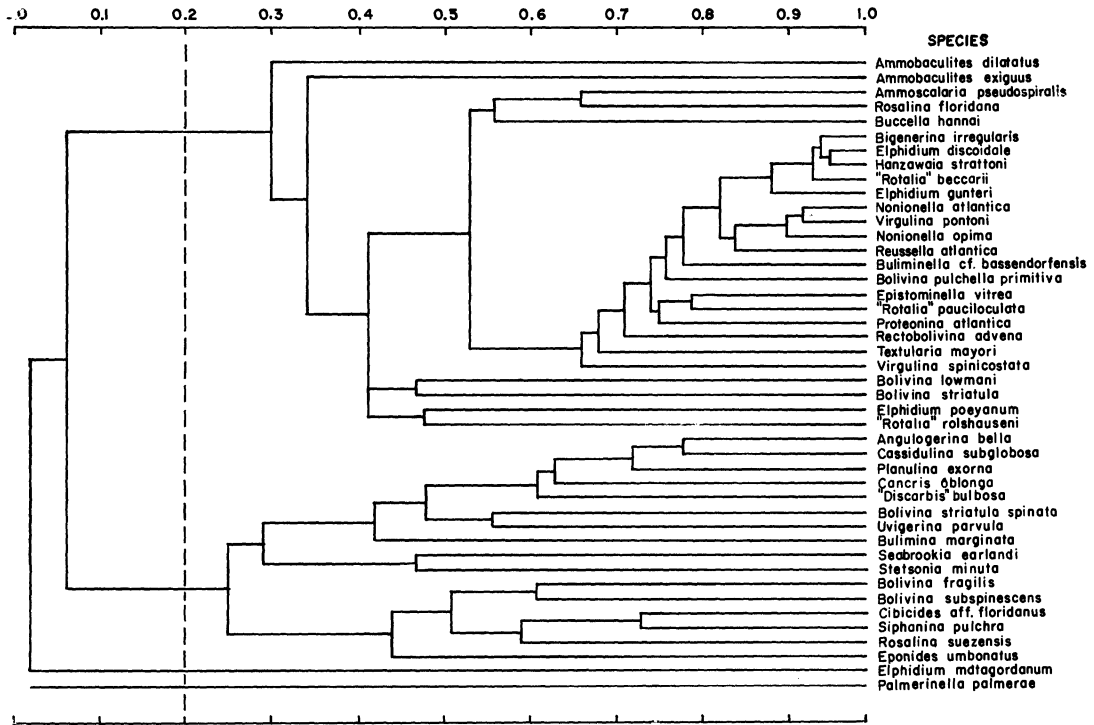
have very few occurrences. The large subcluster consisting of species *Ammoscalaria pseudospiralis* through *Seabrookia earlandi* consists of species which occur frequently and generally are distributed throughout the area.

The second cluster at the 0.021 level consists of the species *Bulimina marginata* through *Rosalina suezensis*. These species are scattered in occurrence, with some tendency toward increasing occurrence in deeper water. The deepest occurrence, however, is less than 60 m.

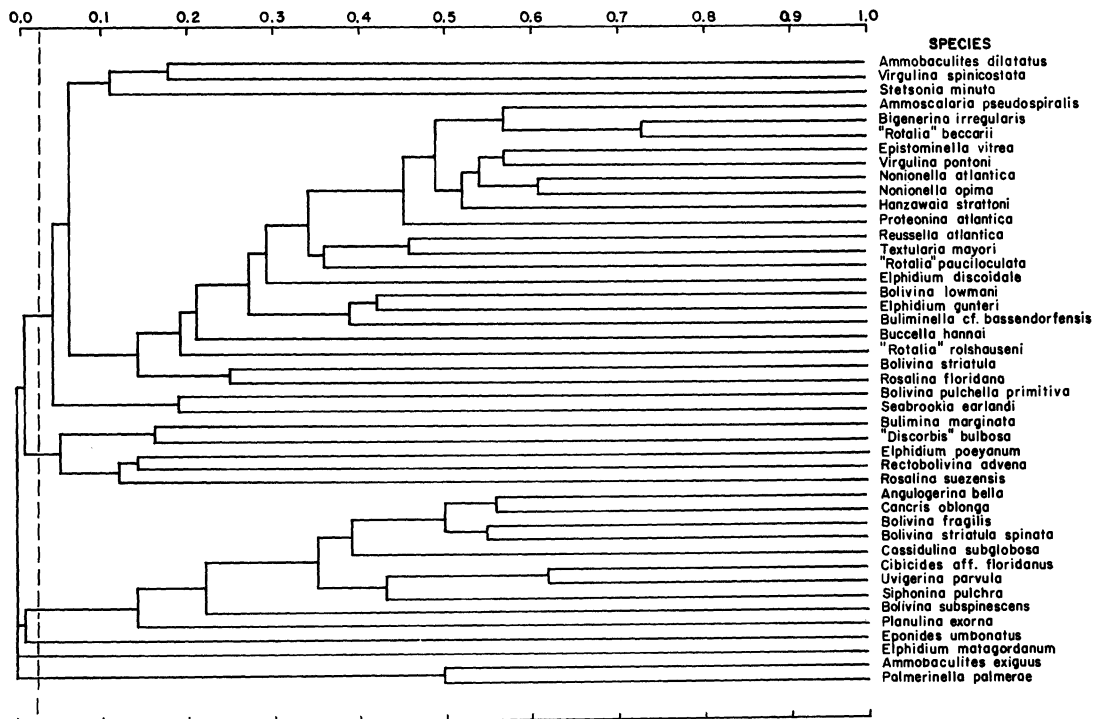
The third cluster consists of the species *Angulogerina bella* through *Planulina exorna*. These species occur most frequently in the deeper area. All ten of them were present in the deep water cluster of 16 species (*Angulogerina bella*–*Eponides umbonatus*) of the total population.

The fourth cluster consists of *Eponides umbonatus* which occurs rarely and only in the deeper area.

The fifth and sixth clusters consist of the



TEXT-FIG. 5—Dendrogram for R-mode analysis of the total population.



TEXT-FIG. 6—Dendrogram for R-mode analysis of the living population.

species *Elphidium matagordanum* through *Palmerinella palmerae*. These species are very rare in occurrence.

The agreement between the dendrograms for the R-mode of the total and living populations is very good. In both cases there is a grouping of species into shallower and/or ubiquitous species and deeper species.

#### DISCUSSION

These analyses indicate that the method described is useful for biofacies analysis. Q-mode analysis divided the stations into three large areally recognizable groups. The depth limits are somewhat different from those chosen by Phleger (1956) and subsequently used for analysis by Buzas (1967), but the agreement is good. The agreement between the depth boundaries determined for the total and the living populations in this study is also quite good. The former has boundaries at 10 m and 77 m, and the latter at 12 m and 77 m.

It should be kept in mind that the clustering method used *must* cluster the data. In the present study there was good *a priori* knowledge of the expected pattern of station relationships. Could meaningful clusters be recognized without *a priori* knowledge? Providing we are willing to stipulate that the clusters must be areally recognizable (that is, meaningful when mapped) the answer is yes. If no areally recognizable units (groups or clusters of contiguous stations) are displayed in the dendrogram, then it may be assumed (if sampling is adequate) that such a pattern does not exist in the study area. This is in agreement with Kaesler's (1966) statement that for biofacies analysis one must assume that "Biotopes are mappable."

One must keep in mind, however, that the pattern is based on presence or absence data. To recognize a pattern in the multispecies populations we may analyze by this method, we require some species to be discrete in their areal distribution. If all species occurred everywhere in the area of study but with different abundances, no areal pattern could be recognized. For example, Phleger's (1956) observations and Buzas' (1967) analyses indicated that the 0-30 m northeasternmost area, Traverse V, can be discriminated from the two other 0-30 m areas in the living population because of the greater species densities found there. In the present study, this difference could not be recognized.

The R-mode analysis successfully clustered species restricted to the deeper area, species which were widespread in their distribution, and species which were very rare. However, these results are more difficult to interpret than Q-mode results. In the Q-mode a cluster of stations can

simply be compared with their geographic position to see if an areally defined unit is formed. In the R-mode the cluster of species must first be compared with the samples it occurs in to see if an areally recognizable unit is present. Hopefully, the R-mode will tell us which species are responsible for the areal units recognized through the Q-mode. The R-mode may, however, also point out interesting associations which are not spatial patterns in the usual sense and, therefore, are not mappable. Clusters of species which continually occur together throughout the study area may reveal an interesting pattern but may not be mappable or even intelligible in the context of present knowledge. Clustering in the R-mode forces each species to belong to one group or another. While such clustering is not objectionable for samples which are geographically located and which we want to be mappable, it can be for species. Serious difficulties arise in the interpretation of the R-mode because we have no clear-cut criterion such as mappability by which to evaluate the results.

We believe the method used in the present study can reveal meaningful patterns based on presence or absence. The method is simple, and the broad mappable patterns it outlines would probably be recognizable to the investigator without using it when only a few species and a few samples are involved. When the number of species and the number of samples becomes large, however, the method greatly simplifies what otherwise would be an impossible task for the investigator. The method is not statistical in the sense that it is not soundly based on probability theory. It is possible to use statistical methods on presence-absence data (see, for example, Macnaughton-Smith, 1965), but this has not been adequately explored to date for biofacies analysis. If abundance data are available, as they were in the present analysis, and interesting areas can be delineated for comparison before analysis, a statistical method such as canonical analysis is much to be preferred. On the other hand, as an exploratory aid or as an aid in interpreting data which have been collected in such a manner that a statistical analysis is not warranted, we believe the method used here is a valuable empirical tool.

#### REFERENCES

- BUZAS, M. A., 1965, The distribution and abundance of Foraminifera in Long Island Sound: Smithsonian Misc. Coll., v. 149, p. 1-89.
- , 1967, An application of canonical analysis as a method for comparing faunal areas: Jour. Animal Ecology v. 36, p. 563-577.
- , 1968, On the spatial distribution of Foraminifera: Cushman Found. Foram. Research Contr. v. 19, pt. 1, p. 1-11, figs. 1-5.

- KAESLER, R. L., 1966, Quantitative re-evaluation of ecology and distribution of Recent Foraminifera and Ostracoda of Todos Santos Bay, Baja California, Mexico: Kansas Univ., Paleont. Contr. Paper 10, p. 1-50, text-figs. 1-23, tables 1-14.
- LYNTS, G. W., 1966, Variation of foraminiferal standing crop over short lateral distances in Buttonwood Sound, Florida Bay: Limnology and Oceanography, v. 11, no. 4, p. 562-566.
- MACNAUGHTON-SMITH, P. N. M., 1965, Some statistical and other numerical techniques for classifying individuals: Her Majesty's Stationary Office, Home Office, London, 33 pages.
- MADDOCKS, R. F., 1966, Distribution patterns of living and sub-fossil podocopid Ostracodes in the Nosy Bé area, Northern Madagascar: Kansas Univ., Paleont. Contr. Paper 12, p. 1-72, text-figs. 1-63, tables 1-12.
- PHLEGER, F. B., 1956, Significance of living foraminiferal populations along the central Texas coast: Cushman Found. Foram Research Contr., v. 7, p. 106-151.
- PRESTON, F. W., 1962, The canonical distribution of commonness and rarity pt. 1: Ecology, v. 43, no. 2, p. 185-215.
- SIMPSON, G. G., 1960, Notes on the measurement of faunal resemblance: Am. Jour. Sci., Bradley Volume, v. 258-A, p. 300-311.
- SOKAL, R. R., & ROHLF, F. J., 1962, The comparison of dendrograms by objective methods: Taxon, v. 11, p. 33-40.
- SOKAL, R. R., & SNEATH, P. H. A., 1963, Principles of numerical taxonomy: San Francisco, W. H. Freeman & Co., 359 p.

MANUSCRIPT RECEIVED AUGUST 21, 1967

The Smithsonian Institution contributed \$250.00 from the Charles D. and Mary Vaux Walcott Research Fund toward publication of this paper.