

## SHE ANALYSIS FOR BIOFACIES IDENTIFICATION

MARTIN A. BUZAS<sup>1</sup> AND LEE-ANN C. HAYEK<sup>2</sup>

## ABSTRACT

Most quantitative methodologies for designating biofacies do so by ordering, in some way, all the possible similarities between stations along some gradient. The procedure is considered successful when the ordering produces mappable units (biofacies or biotopes). In contrast, SHE analysis for biofacies identification (SHEBI) serially accumulates contiguous stations and designates biofacies as groups of contiguous stations that exhibit the properties of known statistical distributions for biological communities. SHEBI identifies biological communities and always insures mappability.

The number of species,  $S$ , the information function,  $H$ , and the measure of evenness,  $E$ , are related by the decomposition equation  $H = \ln S + \ln E$ . Each of these quantities is calculated as  $N$ , the number of individuals, is accumulated by adding successive samples (stations). In all cases, for a multispecies population (community) these variables form linear trends on a log scale. As  $N$  accumulates with each sample,  $S$  usually increases and the decomposition equation establishes constraints on  $H$  and  $\ln E$ . If  $H$  remains constant, then  $\ln E$  must decrease precisely as  $\ln S$  increases. If  $\ln E$  remains constant, then  $H$  must increase precisely as  $\ln S$  increases. If  $H$  increases and  $\ln E$  decreases as  $\ln S$  increases with accumulation, then the ratio  $\ln E/\ln S$  may remain constant. Sometimes,  $\ln S$  and  $H$  increase while  $\ln E$  and  $\ln E/\ln S$  decrease. Departures from linear trends indicate a mixture of communities and, hence, a new biofacies.

The total population from 35 stations ranging in depth from 20m to 2696m in traverse 6 of Parker (1954) in the northeastern Gulf of Mexico was analyzed using the SHEBI technique. Seven biofacies were recognized. These biofacies compare favorably with the six recognized by Parker. However, the SHEBI procedure recognized three biofacies at depths shallower than 100m while Parker recognized one. At depths greater than 200m SHEBI recognized two biofacies while Parker recognized three.

SHEBI establishes boundaries by examining departures from the linear trends expected for the statistical distribution describing a multispecies population or community. Thus, SHEBI provides us not only with a means of designating biofacies, but also with a quantitative definition of a community.

## INTRODUCTION

Wherever modern benthic foraminifera are surveyed along a depth gradient, the species examined exhibit a zonation with depth. While the distribution of each species is

unique (Culver and Buzas, 1981), groups of species can be recognized as principally inhabiting particular areas such as inner shelf, outer shelf, and so on. These groups of species are often called biofacies (Gressly, 1838) even though the original term referred to the biological aspect of rocks or sediments. Initially, investigators examined their data and proposed depth zonations based on the distribution and abundance of the taxa contained therein (e.g., Natland, 1933; Phleger, 1960). When computers and multivariate methods became available, researchers took advantage of the many numerical techniques to quantify biofacies more objectively (e.g., Kaesler, 1966; Buzas, 1967, 1969; Culver, 1988). Today, the use of a myriad of numerical methodologies for the designation of biofacies is commonplace.

As might be expected, biofacies, regardless of how they are recognized, do not all contain the same number of species. Traverses with numerous stations distributed from shallow to deeper water usually show an increase in species richness ( $S$ ) per station to at least the outer shelf (Sen Gupta and Kilbourne, 1974). At greater depths, species richness per station exhibits one of three patterns: increase, decrease or relatively constant (Gibson and Buzas, 1973). The information function,  $H$  (Shannon, 1948), shows a similar pattern while,  $E$ , a measure of evenness derived from it (Buzas and Gibson, 1969), often shows no apparent trend (Gibson and Buzas, 1973; Sen Gupta and Kilbourne, 1974; Gibson and Hill, 1992). While biodiversity patterns are used to characterize areas such as nearshore vs offshore and so on, they are not used to define biofacies.

In this paper we show how patterns of biodiversity can be used to recognize biofacies through analysis of  $S$ ,  $H$ , and  $E$ . This is now possible because a decomposition equation for the relationship of  $S$ ,  $H$ , and  $E$  has been derived (Buzas and Hayek, 1996; Hayek and Buzas, 1997). A new methodology is involved. The new approach differs from the usual procedure used by researchers in an important aspect. Instead of examining the data station by station or taking the mean of several contiguous stations, we accumulate the data so that  $S$ ,  $H$ , and  $E$  are examined as a function of  $N$  (the number of individuals). That is, we compute the quantities  $S$ ,  $H$  and  $E$ , related through the decomposition equation, on each set of contiguous stations as the samples (stations) are accumulated. The expected behavior of  $S$ ,  $H$ , and  $E$  with the accumulation of  $N$  for statistical distributions of communities is known uniquely. Thus, departures from known expected patterns define boundaries. We examine these boundaries for a traverse in the Gulf of Mexico (Parker, 1954).

## THE SHE RELATIONSHIP

The familiar Shannon information function (Shannon, 1948) is

$$H = -\sum_{i=1}^S p_i \ln p_i \quad (1)$$

<sup>1</sup> Department of Paleobiology, NHB MRC-121, Smithsonian Institution, Washington, D.C. 20560, U.S.A.

<sup>2</sup> Mathematics & Statistics, NHB MRC-136, Smithsonian Institution, Washington, D.C. 20560, U.S.A.

where  $H$  is the information function,  $p_i$  is the proportion of the  $i$ th species and  $S$  the number of species. The maximum value of  $H = \ln S$  or  $e^H = S$  where  $e$  is the base of the natural logarithms. This maximum value occurs when all the  $p_i$ 's are equal in which case  $e^H/S = 1$ . When the values of the  $p_i$ 's are not equal this ratio will be less than one and will indicate the degree of evenness of the  $p_i$ 's. Consequently, Buzas and Gibson (1969) defined a measure of evenness as

$$E = e^H/S \quad (2)$$

where  $E$  is the measure of evenness or equitability. Hill (1973) showed that this measure of evenness,  $E$ , complies with his mathematical continuum of diversity measures. Now (2) can also be written as  $e^H = SE$ . Taking the natural log ( $\ln$ ) of both sides we obtain

$$H = \ln S + \ln E \quad (3)$$

This decomposition formula (3) uses  $H$  not in the usual way as a diversity measure, but as a vehicle to partition  $H$  into components of species richness,  $S$ , and evenness,  $E$  (Buzas and Hayek, 1996; Hayek and Buzas, 1997). Equation (3) is incisive because, for the first time, researchers have a simple way to examine evenness separately from richness within a single multispecies system.

First, consider that it is always true that  $0 < E \leq 1$ . Therefore, the value of  $\ln E$  is always negative and (3) indicates the  $H$  is made up of  $\ln S$  minus the amount of evenness exhibited by the population. Because the maximum value of  $H = \ln S$ , (3) could also be written as  $H = H_{\max} + \ln E$ .

When a natural population is repeatedly sampled without replacement, the number of individuals observed increases along with  $S$  until all the species inhabiting the area are accounted for or, more likely, the researchers' resources are exhausted. The number of individuals,  $N$ , is, of course, correlated with the area,  $A$ , sampled. Consequently,  $S$  is a function of, or can be related to  $N$  and  $A$ ; that is, the statements  $S = f(N)$  and  $S = f(A)$  are both true. Both of these relationships can be fit with power curves which can be written as  $S = aN^b$  or  $S = cA^d$ , where  $a$ ,  $b$ ,  $c$  and  $d$  are constants. Often these equations are expressed in a linear form as

$$\ln S = \ln a + b \ln N \quad (4)$$

or

$$\ln S = \ln c + d \ln A$$

(e.g. Hanski and Gyllenberg, 1997; Hayek and Buzas, 1997). Power curves when plotted on an arithmetic scale form a "hollow curve". That is, the curve rises sharply at the outset and then levels off, sometimes asymptotically, at some value of  $N$  or  $A$ . When plotted on a log-log scale, such curves form a straight line. While the relationship between  $S$  and  $N$  or  $A$  has been long established (e.g. Preston, 1948, 1962; MacArthur and Wilson, 1967), the consequences for  $H$  and  $\ln E$  through consideration of (3) have not been fully explored.

If we form a total for the number of species in two samples, say  $S_1$  and  $S_2$ , we expect, in their accumulation, that usually  $S_1 < S_{(1+2)}$  where  $S_{(1+2)}$  is the total in both. Consideration of (3) indicates that if  $\ln S$  increases during accumulation, then to satisfy equation (3) either  $H$  or  $\ln E$  or both must also change. Let us examine the possibilities.

If two samples are accumulated so that  $S_1 < S_{(1+2)}$ , then  $\ln S_1 < \ln S_{(1+2)}$ . In this situation if either  $H$  or  $\ln E$  remains constant, then equation (3) requires the non-constant member to vary exactly as  $\ln S$ . For example, if  $H_1 = H_{(1+2)}$  and  $H$ , therefore, remains constant with accumulation, then to satisfy (3)  $\ln E$  must become more negative by exactly the same amount that  $\ln S$  becomes more positive. If a number of samples are accumulated and  $H$  remains constant, then  $\ln E$  must become increasingly negative as  $\ln S$  becomes increasingly positive. At each step in the accumulation procedure equation (3) must be satisfied.

Much of the increase in  $S$  is by the addition of rare species. In a natural population, more species will be represented by one individual, then by two, three, and so forth. The inclusion of these rare species is necessary for a meaningful analysis and does not present any problem, as it would if the analysis depended on *particular* rare species. If a SHE analysis were conducted again over the same area, we would probably not encounter the same rare species represented by, say, one individual, but would instead encounter others represented by one individual. In like fashion, as  $N$  accumulates some individuals originally represented by one individual move into a category represented by two, and so on. However, in the process of accumulating  $N$ , previously undiscovered species appear which are represented by one individual.

One way a constant  $H$  over accumulated samples can occur is with an increase in the proportion of the abundant species and/or a decrease in the proportions of rare species, either of which will result in a lower value of  $E$  (more dominance). A special case conforming to this situation is Fisher's log series (Fisher, 1943). Bulmer (1974) showed that the expected value of  $H$  for the log series is  $H = \ln \alpha + 0.58$  where  $\alpha$  is Fisher's measure of diversity and 0.58 is Euler's constant. Now Fisher's  $\alpha$ , the parameter of the log series, is also a constant for any particular log series and, consequently, for any log series  $H$  is a constant. Hayek and Buzas (1997) found a constant value of  $H$ , but not with the value expected for a log series, for a hectare of trees in Bolivia.

Another possibility as samples are accumulated with a concomitant increase in  $S$  is that  $\ln E$  can remain constant. In this circumstance (3) tells us that  $H$  must increase precisely as  $\ln S$ . A way of keeping  $E$  constant is for a decrease to occur in the proportions of the more abundant species as samples accumulate. A special case is MacArthur's (1957) broken stick model. May (1975) indicated that for the broken stick model the expected value of  $H$  is  $H = \ln S - 0.42$ . Note the resemblance of this equation to (3), indicating that  $\ln E = -0.42$ , which gives an  $E$  value of 0.66. Buzas and Hayek (1996) found a constant  $\ln E$  for a one hectare plot of trees in Guyana, but the value of  $\ln E$  was not the expected one for a broken stick model.

Earlier we indicated that the maximum possible value of  $H$  is  $H = \ln S$ . When this is true, then  $H/\ln S = 1$ . When this is not the case the value of the ratio will be less than one, and defines another measure of evenness,  $J$ , introduced by Pielou (1966) as  $J = H/\ln S$ . This can also be written as  $H = J \ln S$ . May (1975) indicated that for a log normal distribution (Preston, 1948) the expected value of  $H$  is  $H = (1 - \gamma^2) \ln S$  where  $\gamma$  is a parameter (a constant) of the log

normal. A comparison of these results for  $H$  shows that  $J = (1 - \gamma^2)$ ; consequently for the log normal,  $J$  is a constant. With a little algebraic manipulation, in terms of (3),  $H = (1 + \ln E/\ln S)\ln S$ . Consequently, this shows that for a log normal, the ratio  $\ln E/\ln S$  is a constant. A way to achieve a constant  $\ln E/\ln S$  in a multispecies population as species accumulate is to have the proportions of the more abundant species get smaller. However, this reduction in species proportions will be less than when  $\ln E$  remains constant. Thus, the accumulation when  $\ln E/\ln S$  stays constant will show a slight decrease in the evenness as measured by  $E$ . A constant  $\ln E/\ln S$  was found for a one hectare plot of trees in Front Royal, Virginia (Hayek and Buzas, unpublished). In the situation for which  $\ln E$  or  $E$  remains constant (as indicated in the previous paragraph), the ratio  $\ln E/\ln S$  or  $J$  will show a slight increase.

We have also encountered data sets where none of the above quantities are constant. Sometimes as  $S$  increases with accumulation,  $H$  increases slightly, and  $\ln E$  and  $\ln E/\ln S$  decrease. One way this can be accomplished in nature is having the abundant species remain relatively constant while adding species represented by one individual (singletons).

As mentioned previously, we have observed instances where  $H$  or  $\ln E$  remain constant, but the values are not those expected for a particular distribution. This is because in ecological studies only three statistical distributions are most often used in data fitting. Hayek and Buzas (1997) discussed the interrelationships of these three (broken stick, log series, log normal). For more complete identification further mathematical derivation is required to identify entropy ( $H$ ) uniquely for additional members of this family of distributions.

In all of the situations described above, we have been referring to observations obtained from a relatively uniform habitat that is inhabited by a multispecies population, namely, a biofacies. As we accumulate  $N$  and  $S$ , the value of  $E$  or  $\ln E$  remains constant or decreases. However, as we accumulate across biofacies this pattern will show disruption as we pass from one biofacies to another. It is this disruption in pattern that we look for when establishing biofacies through SHEBI.

#### NORTHEASTERN GULF OF MEXICO

In her classic paper, Parker (1954) enumerated the foraminifera from about 200 samples from 11 traverses in the northeastern Gulf of Mexico. We arbitrarily selected traverse 6 for SHE analyses. Although Parker (1954) enumerated live as well as total (live + dead) populations, we selected the total populations because of the much larger numbers observed. Strictly speaking, a biofacies can be considered as the biological component of a formation and is a geological term. The living population represents a community in the ecological sense. However, it can also be argued that the total population gives a better representation of a fauna over ecological time and often resembles the live population anyway, taphonomic effects notwithstanding (Buzas, 1965; Scott and Medioli, 1980; Loubere and Gary, 1990; Loubere et al, 1993). Consequently, we will refer to our analysis of the total population as an analysis of biofacies. For our pur-

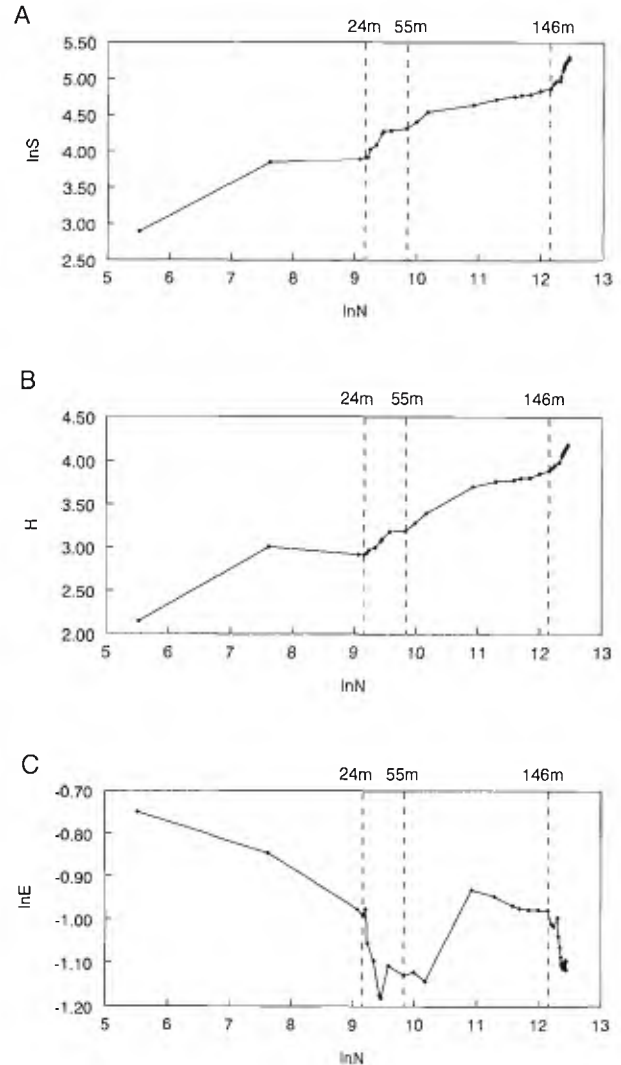


FIGURE 1. Plots of SHE analysis of Parker's (1954) traverse 6. Dashed lines show depths for particular values of accumulated  $\ln N$ . (A)  $\ln S$  vs  $\ln N$ . (B)  $H$  vs  $\ln N$ . (C)  $\ln E$  vs  $\ln N$ .

poses the terms community and biofacies are regarded as synonymous (Schopf, 1996).

In traverse 6, Parker (1954) occupied 35 stations ranging in depth from 20 to 2697 m. She presented her data as the percent of each species observed at each station. In all, 201 taxa were encountered in traverse 6. For the procedure followed here we converted the percent values back to the number of individuals observed. At the first station, 20m, the values are  $\ln S = 2.89$ ,  $H = 2.14$ , and  $\ln E = -0.75$  ( $E = 0.47$ ). After these values are calculated for the initial sample in the traverse, a stepwise accumulation procedure is used for the remaining samples. That is, the number of individuals for each unique species from the first and second station are summed and  $\ln S$ ,  $H$ , and  $\ln E$  are recalculated. The number of individuals in the first, second, and third are then summed and  $\ln S$ ,  $H$ , and  $\ln E$  are recalculated, and so on. Although tedious, the accumulation procedure is straightforward and can be carried out easily on any spread sheet package. We used QUATTRO (1994) for the calculations. The results are shown in Fig. 1A, 1B, 1C. Notice

TABLE 1. SHEBI analysis for Parker (1954) traverse 6. Depth is in meters, N is the number of individuals in the sample,  $\ln N_{cum}$  is the ln of cumulative N,  $S_{obs}$  is the number of species in the sample, S+ is the number of species not observed in the previous sample,  $S_{cum}$  the cumulative number of species,  $\ln S_{cum}$  the ln of  $S_{cum}$ , H is the Shannon information function for cumulative samples,  $\ln E$  the ln of E, the Buzas and Gibson measure of evenness, and  $(1 + \ln E / \ln S_{cum})$  is equal to J, Pielou's measure of evenness.

Biofacies	Depth	N	$\ln N_{cum}$	$S_{obs}$	S+	$S_{cum}$	$\ln S_{cum}$	H	$\ln E$	E	$(1 + \ln E / \ln S_{cum})$
1	20	250	5.52	18	18	18	2.89	2.14	-0.75	0.47	0.74
	22	1798	7.62	46	29	47	3.85	3.00	-0.84	0.43	0.78
	24	6700	9.08	39	2	49	3.89	2.92	-0.98	0.38	0.75
	24	848	9.17	38	1	50	3.91	2.92	-0.99	0.37	0.75
2	29	375	5.93	31	31	31	3.43	2.87	-0.56	0.57	0.84
	27	401	6.65	35	13	44	3.78	3.00	-0.79	0.45	0.79
	24	1092	7.53	47	11	55	4.01	3.15	-0.86	0.42	0.78
	39	1096	7.99	67	15	70	4.25	3.27	-0.98	0.38	0.77
	43	273	8.08	32	2	72	4.28	3.28	-1.00	0.37	0.77
	49	1693	8.50	44	1	73	4.29	3.16	-1.13	0.32	0.74
	55	4191	9.12	36	2	75	4.32	2.88	-1.44	0.24	0.67
3	64	3307	8.10	58	58	58	4.06	3.16	-0.90	0.41	0.78
	91	4291	8.94	70	21	79	4.37	3.37	-1.00	0.37	0.77
4	100	28955	10.27	79	79	79	4.37	3.59	-0.78	0.46	0.82
	106	24349	10.88	73	12	91	4.51	3.63	-0.88	0.42	0.80
	113	27972	11.30	71	8	99	4.60	3.66	-0.94	0.39	0.80
	119	12112	11.44	65	4	103	4.63	3.68	-0.96	0.38	0.79
	128	18625	11.63	68	1	104	4.64	3.68	-0.96	0.38	0.79
	139	23529	11.82	76	6	110	4.70	3.74	-0.96	0.38	0.80
	146	27473	12.00	79	5	115	4.74	3.78	-0.97	0.38	0.80
5	155	9542	9.16	72	72	72	4.28	3.64	-0.64	0.53	0.85
	165	8991	9.83	68	17	89	4.49	3.67	-0.81	0.44	0.82
	183	13772	10.38	68	5	94	4.54	3.59	-0.96	0.38	0.79
6	223	2083	7.64	65	65	65	4.17	3.41	-0.76	0.47	0.82
	446	7816	9.20	74	30	95	4.55	3.55	-1.00	0.37	0.78
	555	3106	9.47	66	10	105	4.65	3.61	-1.05	0.35	0.77
	631	2207	9.63	73	7	112	4.72	3.67	-1.05	0.35	0.78
	650	1624	9.73	63	3	115	4.75	3.68	-1.06	0.35	0.78
	677	1600	9.82	66	3	118	4.77	3.70	-1.06	0.34	0.78
	823	2502	9.95	67	6	124	4.82	3.74	-1.08	0.34	0.78
	860	2104	10.04	62	1	125	4.83	3.75	-1.08	0.34	0.78
	960	3603	10.19	70	3	128	4.85	3.73	-1.12	0.33	0.77
	1144	5100	10.36	64	3	131	4.88	3.72	-1.15	0.32	0.76
7	1573	1301	7.17	56	56	56	4.02	3.15	-0.87	0.42	0.78
	2697	595	7.55	48	20	76	4.33	3.40	-0.93	0.39	0.78

that  $\ln E$  shows considerably more variation than  $\ln S$  or H even though they are all related by (3). From our outline of the SHE relationship given above, we recall that for a multispecies population, the value of  $\ln E$  should either be constant or decrease. This is clearly not the case. An increase in the value of  $\ln E$  indicates the addition of new species with enough relative abundance to increase the evenness. Consequently, the graph confirms what we already knew, namely, that more than one foraminiferal biofacies exists between the depths of 20 to 2697 m. Fig. 1A, 1B, and 1C show that for  $\ln S$ , H, and  $\ln E$  changes in slope occur at  $\ln N = 9.17$  (24m),  $\ln N = 9.83$  (55m), and  $\ln N = 12.15$  (146m). In addition,  $\ln E$  (Fig. 1C) indicates the presence of at least three other places of demarcation. In general,  $\ln E$  is the most sensitive of the trio and we will use it extensively in what follows. However, the accumulations after many thousands of individuals may distort the appearance of the later biofacies because of the influence in the accumulation procedure of species encountered earlier in the traverse which are no longer present at later depths. Consequently, we will break up the traverse and examine it piece by piece. Looking at Fig. 1C notice that the first four stations have decreasing

values of  $\ln E$ . We designate these four stations ranging in depth from 20 to 24 m as the first biofacies (first group in Table 1). The data from these stations are now eliminated from consideration. The fifth station (29 m) now becomes the first for the next SHEBI procedure. The results are shown in Fig. 2A which indicates that the first seven stations show a steadily decreasing trend for  $\ln E$ . These stations ranging in depth from 29 to 55 m are designated as biofacies 2. These stations are now deleted from the data set as we proceed with further calculations. The remaining stations beginning with Parker's 12th station (station 62 at 64 m, Parker, 1954) are now used to continue the SHEBI. Fig. 2B illustrates that two stations at 64 and 91 m display a set of decreasing values of  $\ln E$ . These are designated as biofacies 3. Deleting these stations, we re-accumulate and produce Fig. 2C. Here three stations show a decreasing trend followed by four more with a constant  $\ln E$ , and then a sharp decline after 146m. These seven stations ranging in depth from 100 to 146 m are designated as biofacies 4. Deleting these, we produce Fig. 2D which shows three stations before a sharp break in the slope. These stations, ranging in depth from 155 to 183 m, are designated as biofacies 5. Proceed-

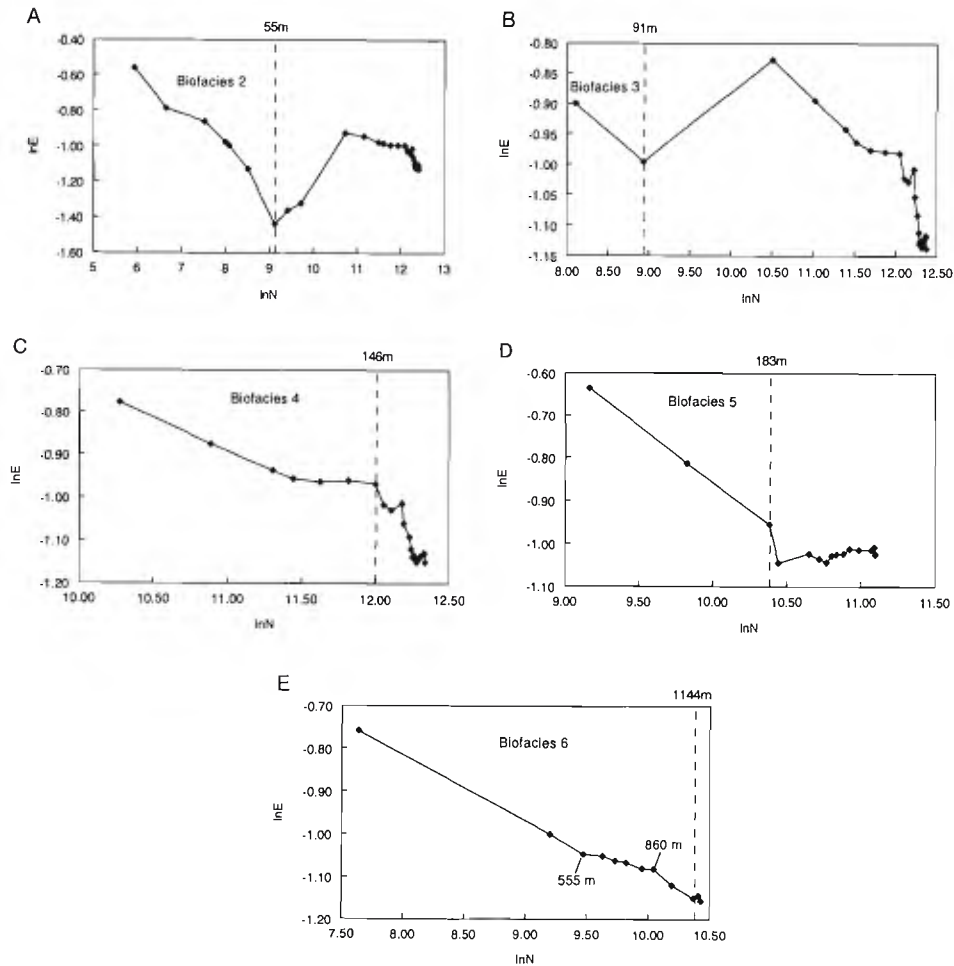


FIGURE 2. Plots of  $\ln E$  vs  $\ln N$  with serial deletions for the accumulation procedure. Dashed lines indicate depths at which biofacies are identified. (A) first 4 stations from Fig. 1 are deleted. (B) first 7 stations from Fig. 2a are deleted. (C) first two stations from Fig. 2B are deleted. (D) First 7 stations from Fig. 2C are deleted. (E) First 3 stations from Fig. 2D are deleted.

ing in like fashion, we produce Fig. 2E, which shows three more stations before a slight change in slope at  $\ln N = 9.20$  (555m). Unlike the previous breaks, this one is more subtle. Another possible break is at  $\ln N = 10.04$  (860m). However, neither of these are like the sharp delineations experienced with biofacies 1 through 5. Additionally, the samples are quite widely spaced over a considerable depth range. Consequently, we took the conservative viewpoint and chose neither. Instead, we chose the upturn in  $\ln E$  at the second last station ( $\ln N = 10.40$ , 1573m). Biofacies 6 then constitutes the 10 stations from 223m to 1144m. Finally, the last two stations are designated as biofacies 7 (1573 and 2697 m). We have then by this simple procedure designated 7 biofacies for Parker's traverse 6. Values are shown in Table 1. When referring to Table 1, the reader should keep mind that the accumulation procedure begins each time with the first entry in each biofacies.

A comparison of biofacies obtained from SHEBI and Parker's generalized biofacies based on 11 traverses is shown in Fig. 3. The results are surprisingly similar considering the difference in criteria for discrimination. Parker chose biofacies based on the changes in the relative abundance of individual taxa while the SHEBI procedure examined

changes in the multispecies structure of the total population. In shallower water, SHEBI indicates the presence of three biofacies while Parker only designated one. However, Parker did recognize a subfacies at 30–50 m and indicated considerable variability of relative abundance in shallow water. Parker recognized a boundary between 555 and 631m and while we noted a possible change in slope (Fig. 2E) at this location, we did not designate a biofacies.

## DISCUSSION

Statistical distributions of naturally occurring multispecies populations (communities) exhibit a non-increasing linear trend for  $\ln E$  vs  $\ln N$ . The criteria for biofacies identification using SHEBI is to identify a group of contiguous stations that exhibit such a linear trend. Departures from a non-increasing linear trend define the boundaries of the biofacies. This criteria provides us with a quantitative definition of what constitutes a community.

Unlike most ordination procedures that compare all possible pairs of samples and sometimes produce uninterpretable results, the SHEBI analysis assures continuity and mappability because all the areas recognized by SHEBI are

Depth (m)	SHEBI	Parker
20		
22		
24	1	
24		
29		
27		
24		1
39	2	
43		
49		
55		
64		
91	3	
100		
106		
113		
119	4	2
128		
139		
146		
155		
165	5	3
183		
223		
446		4
555		
631		
650	6	
677		
823		5
860		
960		
1144		
1573		6
2697	7	

FIGURE 3. Comparison of biofacies obtained by SHEBI and Parker (1954) for traverse 6.

contiguous. This contiguity of the samples does not imply that we must observe a trend with depth, only that SHEBI will uncover a series of biofacies with depth, if it occurs. We believe this is a most desirable property for biofacies analysis. On the other hand if the same biofacies repeats itself along a traverse, the SHEBI procedure would recognize only another additional biofacies. However, in all biofacies analyses, regardless of the methodology used, the researcher should always examine the data to determine the constituency.

The SHEBI procedure could not unequivocally discriminate among the 10 stations distributed from 223m to 1144m. Parker (1954) recognized a boundary between 555m and 631m in her scheme for the entire northeastern Gulf (Fig. 3). However, unlike the shallower stations, the stations deeper than 183m are distributed over a very much wider geographic area. The lack of replication and the wide geographic range of the stations in this interval leaves the question of further subdivision unresolved.

When enough samples are available within a biofacies, a SHE analysis will allow us to identify the properties or structure of the underlying statistical distribution. In other

words, SHE analysis can be used for community structure identification (SHECSI).

As an example of SHECSI, we will examine the three (2, 4, 6) biofacies with the most samples identified by SHEBI in traverse 6. Biofacies 2 (29 to 55 m,  $n = 7$ , Table 1) is defined by a pattern of increasing  $\ln S$  and an accompanying decrease in values of both  $\ln E$  and  $(1 + \ln E/\ln S)$ . The value of  $H$  increases and then decreases so that the final value at 75 species is nearly identical to the initial value at 31 species (Table 1). The regression coefficient  $b$  for the equation  $H = \ln \alpha + b \ln N$  is 0.05 indicating that the predicted values for  $H$  are nearly constant. Ideally, for a log series the observed values of  $H$  should remain constant. For a log series equation (3) requires that the values for the regression coefficients for  $\ln S$  and  $\ln E$  should be nearly equal and of opposite sign. In our example, the former has a value of 0.29 and the latter  $-0.24$ , the difference of 0.05 is the regression coefficient for  $H$ . The predicted value of  $H$  for a log series is  $H = \ln \alpha + 0.58$ . With  $\ln N = 9.12$  and  $S = 75$ , the value for  $\alpha = 11.2$  (see, Hayek and Buzas, 1997, for calculation procedure) which results in a predicted  $H = 2.99$ . The observed value at  $S = 75$  is 2.88 (Table 1) and the average of all observed values in the biofacies is  $H = 3.09$ . These observations are consistent with a log series distribution for biofacies 2.

Biofacies 4 with  $n = 7$  is at depths of 100 to 146 m. The values of  $\ln S$  and  $H$  increase while the values of  $\ln E$  decrease at the outset and then become constant (Table 1). The regression coefficient for  $H$  vs  $\ln N$  is 0.10 and for  $\ln E$  vs  $\ln N$ ,  $-0.11$ . The values of  $(1 + \ln E/\ln S)$  are nearly constant throughout (Table 1). The regression coefficient for  $\ln E/\ln S$  vs  $\ln N$  is  $-0.02$ . This constancy of  $\ln E/\ln S$  is expected with a log normal distribution.

The third example is biofacies 6 with  $n = 10$  at depths of 223 to 1,144 m. As  $\ln S$  increases,  $\ln E$  decreases while  $H$  increases (Table 1). The regression coefficient for  $H$  vs  $\ln N$  is 0.13 and for  $\ln E$  vs  $\ln N$ ,  $-0.14$ . The values for  $(1 + \ln E/\ln S)$  shown in Table 1 are nearly constant and the regression coefficient for  $\ln E/\ln S$  vs  $\ln N$  is  $-0.02$ . These results are also most consistent with a log normal distribution.

The SHECSI analysis, then, allows the researcher to go beyond the designation of biofacies procedure of SHEBI and examine the underlying distribution. At the present time, however, the data are too scanty to speculate on an overall pattern or meaning. The main purpose of this paper is simply to outline the procedure for designating biofacies using the technique of SHEBI.

#### SUMMARY

1. SHE analysis (Buzas and Hayek, 1996; Hayek and Buzas, 1997) is a generalized methodological set of procedures for the study of multispecies population structure. SHE analysis can be used for community structure identification (SHECSI) and biofacies identification (SHEBI).
2. A naturally occurring community consists of  $N$  individuals which are distributed among  $S$  species. The calculation of the information function,  $H$ , uses both the number of species and their relative abundances.  $H$  is decomposed by the equation  $H = \ln S + \ln E$  where  $E = e^H/S$  is a measure of evenness.

3. As samples are accumulated within a single community, N and S increase. The decomposition equation places constraints on H and lnE. Communities have known statistical distributions that exhibit a non-increasing linear trend for lnE vs lnN.

4. The underlying premise when using the SHE approach for the identification of biofacies (SHEBI) is to identify demarcations or departures from non-increasing linear trends.

5. Using SHEBI for the analysis of biofacies along some environmental gradient such as depth is a simple procedure. The procedure involves identifying a break and re-examining the remainder of the data. We begin by subjecting the entire transect or traverse to a SHE analysis. A plot of lnE vs lnN is examined for changes in a non-increasing linear trend. The first change is designated as a biofacies boundary. The station(s) occurring before the break are designated as composing the first biofacies and are then deleted from further consideration. The data set now consists of the remaining stations. The analysis is repeated and the first change in slope is recognized as a second biofacies boundary. The station(s) occurring before the break are designated as biofacies two and are then deleted from the data set. The procedure continues until we run out of stations and the entire traverse is demarcated into ecological zones or biofacies.

6. The SHEBI procedure differs from all other approaches because it identifies the zones or biofacies by examining the structure of multispecies populations.

#### ACKNOWLEDGMENTS

We thank J. Jett for her assistance with computation and graphics. J. A. Comiskey for the QUATTRO program and R. S. Carney, S. J. Culver and P. J. D. Lamshead for helpful comments on the manuscript. This is contribution number 440 from the Smithsonian Marine Station at Link Port.

#### REFERENCES

- BULMER, M. G., 1974, On fitting the Poisson lognormal distribution to species-abundance data: *Biometrics*, v. 9, p. 101-110.
- BUZAS, M. A., 1965, The distribution and abundance of Foraminifera in Long Island Sound: *Smithsonian Miscellaneous Collections*, v. 149, p. 1-89.
- , 1967, An application of canonical analysis as a method for comparing faunal areas: *Journal of Animal Ecology*, v. 36, p. 563-577.
- , 1969, On the quantification of biofacies. *Proceeding of the North American Paleontological Convention*, September 1969, Part B, p. 101-116.
- , and GIBSON, T. G., 1969, Species diversity: benthonic Foraminifera in Western North Atlantic: *Science*, v. 163, p. 72-75.
- , and HAYEK, L. C., 1996, Biodiversity resolution: An integrated approach: *Biodiversity Letters*, v. 3, p. 40-43.
- CULVER, S. J., 1988, New foraminiferal depth zonation of the Northwestern Gulf of Mexico: *Palaios*, v. 3, p. 69-85.
- , and BUZAS, M. A., 1981, Distribution of Recent benthic foraminifera in the Gulf of Mexico: *Smithsonian Contributions to the Marine Sciences*, No. 8, 698p.
- FISHER, R. A., 1943, The relation between the number of species and the number of individuals in a random sample of an animal population. Part 3. A theoretical distribution for the apparent abundance of different species: *Journal of Animal Ecology*, v. 12, p. 54-58.
- GIBSON, T. G., and BUZAS, M. A., 1973, Species diversity: Patterns in modern and Miocene foraminifera of the Eastern Margin of North America: *Geological Society of America Bulletin*, v. 84, p. 217-238.
- , and HILL, E. E., 1992, Species dominance and equitability: Patterns in Cenozoic foraminifera of eastern North America: *Journal of Foraminiferal Research*, v. 22, p. 34-51.
- GRESSLY, A., 1838, *Observations géologiques sur le Jura Soleurois*: Schweizer Gesell. Gesamten Naturwiss., Neue Denkschr. v. 2, p. 1-112.
- HANSKI, I., and GYLLENBERG, M., 1997, Uniting two general patterns in the distribution of species: *Science*, v. 275, p. 397-400.
- HAYEK, L. C., and BUZAS, M. A., 1997, *Surveying natural populations*: Columbia University Press, New York, 563p.
- HILL, M. O., 1973, Diversity and evenness: A unifying notation and its consequences: *Ecology*, v. 54, p. 427-432.
- KAESLER, R. L., 1966, Quantitative re-evaluation of ecology and distribution of Recent foraminifera and ostracoda of Todos Santos Bay, Baja California, Mexico: *The University of Kansas Paleontological Contributions*, Paper 10, p. 1-50.
- LOUBERE, P., and GARY, A., 1990, Taphonomic process and species microhabitats in the living to fossil assemblage transition of deeper water benthic Foraminifera: *Palaios*, v. 5, p. 375-381.
- , GARY, A., and LAGOE, M., 1993, Generation of the benthic foraminiferal assemblage: Theory and preliminary data: *Marine Micropaleontology*, v. 20, p. 165-181.
- MACARTHUR, R. H., 1957, On the relative abundance of bird species: *Proceedings of The National Academy of Sciences*, v. 43, p. 293-295.
- , and WILSON, E. O., 1967, *The theory of island biogeography*: Princeton University Press, Princeton, 203p.
- MAY, R. M., 1975, Patterns of species abundance and diversity: in Cody, M. L., and Diamond, J. M., (eds.), *Ecology and evolution of communities*, Belknap Press of Harvard University Press, Cambridge, p. 81-120.
- NATLAND, M. L., 1933, The temperature and depth distribution for some Recent and fossil Foraminifera in the southern California region: *Bulletin of the Scripps Institute of Oceanography, Technical Series*, v. 3, p. 225-230.
- PARKER, F. L., 1954, Distribution of the foraminifera in the Northeastern Gulf of Mexico: *Bulletin of the Museum of Comparative Zoology*, v. 111, p. 452-588.
- PIELOU, E. C., 1966, The measurement of diversity in different types of biological collections: *Journal of Theoretical Biology*, v. 13, p. 131-144.
- PHLEGER, F. B., 1960, *Ecology and distribution of Recent foraminifera*: The Johns Hopkins Press, Baltimore, 297p.
- PRESTON, F. W., 1948, The commonness and rarity, of species: *Ecology*, v. 29, p. 254-283.
- , 1962, The canonical distribution of commonness and rarity (two parts): *Ecology*, v. 43, p. 185-215, 410-432.
- QUATTRO PRO, 1994, *WordPerfect*. Novell Applications Group, Novell, Inc., Version 6.0, Orem, Utah, 752p.
- SCHOPF, K. M., 1996, Coordinated stasis: biofacies revisited and the conceptual modeling of whole-fauna dynamics: *Palaeogeography, Palaeoclimatology, Palaeoecology*, v. 127, p. 157-175.
- SCOTT, D. B., and MEDIOLI, F. S., 1980, Living vs. Total foraminiferal populations: their relative usefulness in paleoecology: *Journal of Paleontology*, v. 54, p. 814-831.
- SEN GUPTA, B. K., and KILBOURNE, R. T., 1974, Diversity of benthic foraminifera on the Georgia Continental Shelf: *Geological Society of America Bulletin*, v. 85, p. 969-972.
- SHANNON, C. E., 1948, A mathematical theory of communication: *Bell System Technical Journal*, v. 27, p. 379-423, v. 27, p. 623-656.

Received 9 September 1997

Accepted 26 January 1998

