Geol. Soc. Malaysia Bulletin 21, December, 1987; pp. 103 – 119

Computer–assisted interpretation of depositional palaeoenvironments based on foraminifera

PHILIP LESSLAR
Sarawak Shell Berhad
Lutong,Sarawak

Abstract: In Sarawak Shell's Geological Laboratory, well samples are analysed in part for their foraminiferal content and this information is used for interpreting the depositional environment and geological age of the section penetrated. This paper addresses the former usage of foraminifera data.

With the aim of minimising the subjectivity involved and of attaining a consistent basis for interpretations, cluster analysis, environmental range charts, identification matrices and a set of interactive programs have been worked into a scheme which enables probabilistic computer-assisted interpretation to be carried out on samples utilising the presence or absence of species. Results are listed with their corresponding probability values and aid the investigator in making consistent environmental interpretations.

INTRODUCTION

The environmental scheme for the Tertiary of NW Borneo developed by the Geological Laboratory of SSB/SSPC is based on a two-fold subdivision:

(i) Bathymetry and

(ii) Holomarine versus fluviomarine environments (Fig. 1).

Interpretation of the palaeoenvironments in samples of well sections is based on sedimentological as well as palaeontological criteria. The composition of foraminiferal assemblages especially, is considered to reflect the depth and nature of their living environment.

The large number of species found in this area (~1500) and the uniqueness that each sample assemblage possesses in terms of species content, frequencies, diversity and preservation make objective and consistent interpretations of depositional palaeoenvironments a difficult task. The problem is accentuated when interpretations are made by several investigators since personal concepts and criteria tend to be developed in addition to established ones depending on the knowledge and experience of particular investigators.

Since these criteria are largely qualitative, and it was felt desirable to develop a quantitative approach which would put some of the accepted criteria on a firmer basis and to point the way to new and useful criteria.

The study was carried out using sidewall sample data available from wells drilled in the Sarawak area (Fig. 2). The information represented by these wells is not only spread over a wide geographical area but also covers the range of environments shown in Fig. 1.

Presented at GSM Petroleum Geology Seminar 1985

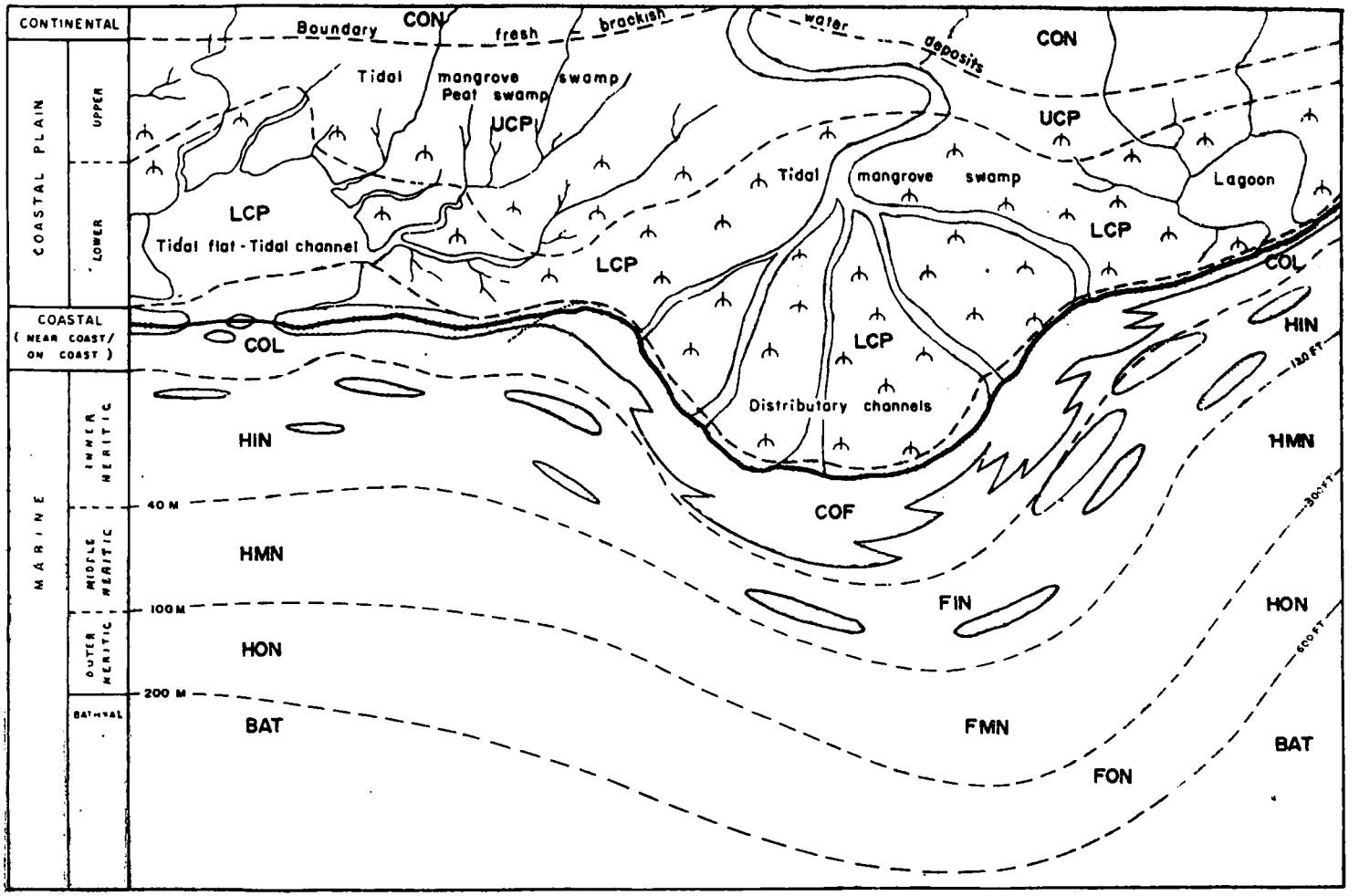


Fig. 1. Schematic outline of NW Borneo Environmental Scheme.

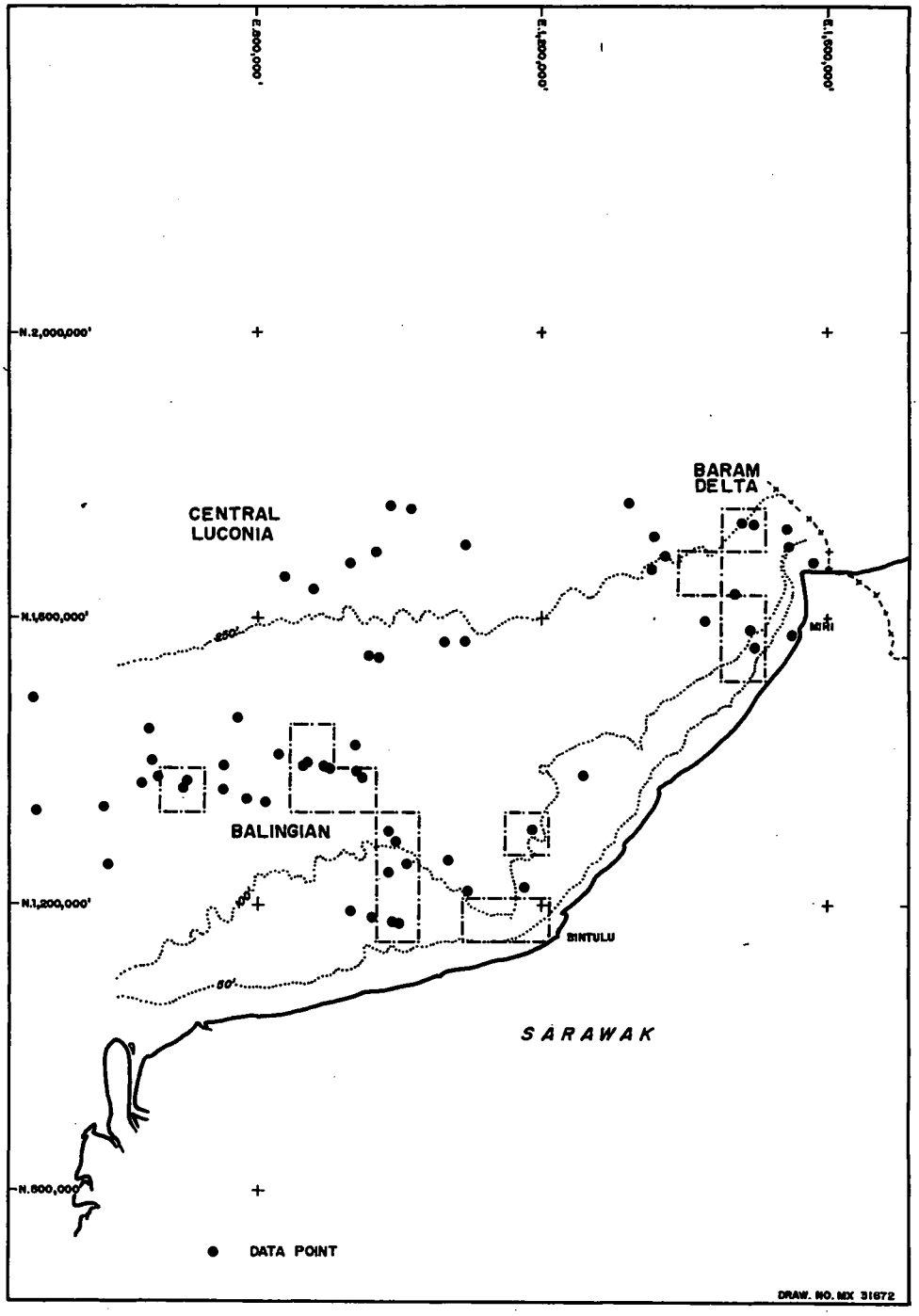


Fig. 2. Data Points Location Map.

METHOD OF INVESTIGATION AND DRAWBACKS IN PREVIOUS METHOD OF INTERPRETATION

The microfossil contents of samples are identified under microscopes by comparing them with a type collection consisting of ~1500 types or against published specimens in literature.

200 foraminifera from each sample are picked and identified. The rest of the sample is scanned for further species which are simply recorded as being present. The result is then studied for foraminiferal markers or assemblages which suggest the depositional environment of this sample.

The identification of the many species plus the interpretation of the environment of deposition are both subjective processes. Consistency can, therefore, be optimised through the use of a type collection. However, the interpretation of the environment of deposition based on fossil assemblages remains the weakest link in the entire process. This becomes especially problematic when experienced staff leave the area.

One solution would be to have a "type-collection" of groups of species that typify the environments of deposition as shown in the environmental scheme of Fig. 1. Such groups would necessarily contain large amounts of species information in order to adequately describe their respective environments. Manual comparison of assemblages against these groups would be extremely difficult. To overcome his difficulty, a computerised method enabling such multivariate comparisons to be rapidly done has been developed and will be discussed below.

STUDY PROCEDURE

The first part of the study can be divided into two main steps (Fig. 3), firstly using cluster analysis to help sort out the sample set into groups which contain samples similar to each other and which are considered to reflect particular depositional environments and secondly to obtain from these groups of samples range charts which show quantitative changes of species percentages across the environment spectrum.

The second part consists of using a set of interactive programs to create identification matrices from range chart data and then to use these matrices in a program which identifies the most likely environmental interpretation of a sample based on its species content in relation to the chosen identification matrix.

CLUSTER ANALYSIS

To objectively compare individual samples and measure their similarities, cluster analysis was used.

Cluster analysis is a multivariate technique which allows comparisons and classification to be done on a set of samples, based on their species content, even when little is known about the structure of the data. It is useful in this case because comparisons take all species in a sample into account and may reveal associations or groupings which were not apparent at first glance. A comprehensive computer package called CLUSTAN (Wishart, 1978) was obtained and implemented for this purpose. In order to enable data

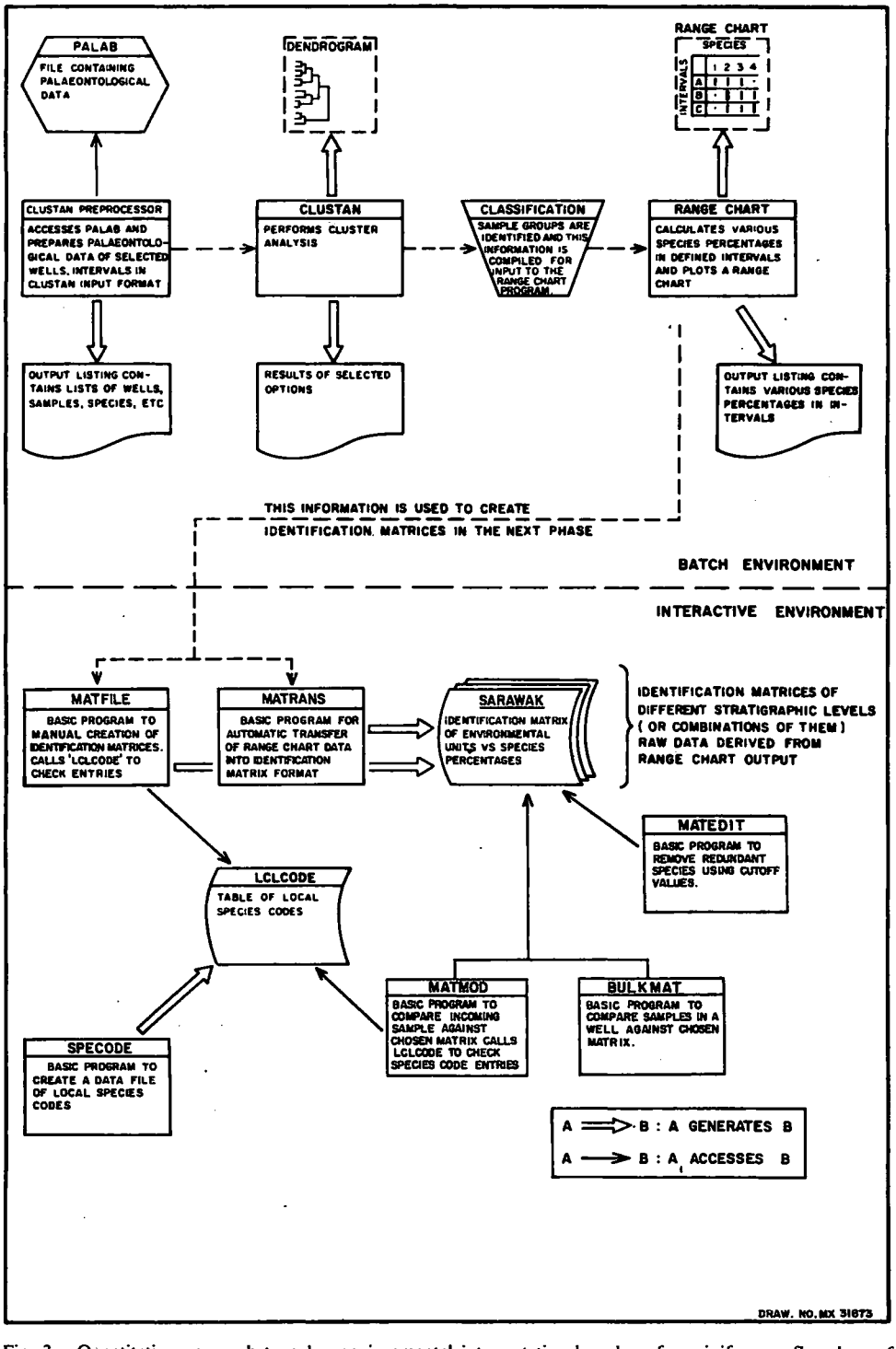


Fig. 3. Quantitative approach to palaeoenvironmental interpretation based on foraminifera — flowchart of working method.

from PALAB (the palaeontological computer data base system in SSB) to be retrieved and arranged in CLUSTAN acceptable format, a program (CLUSTAN preprocessor) was designed which permitted both Q-mode (sample-sample comparisons) and R-mode (species-species comparisons) data to be prepared. Options were built in to allow for well/interval selection, varying cutoff values on species numbers and exclusion of species.

For clustering purposes (Q-mode specifically) each sample can be thought of as a point x in n-dimensional space, where each species represents one dimension. The data of a set of samples can be put in the form of a $p \times n$ matrix where p = number of samples and n = total number of species. This enables the calculation of various coefficients to be done which provide indications of the strength of relationships between samples, one of which arises from the concept of distance (Sneath and Sokal, 1973, p. 124). The stronger the relationship between two sample points in n-dimensional space, the smaller the distance between them.

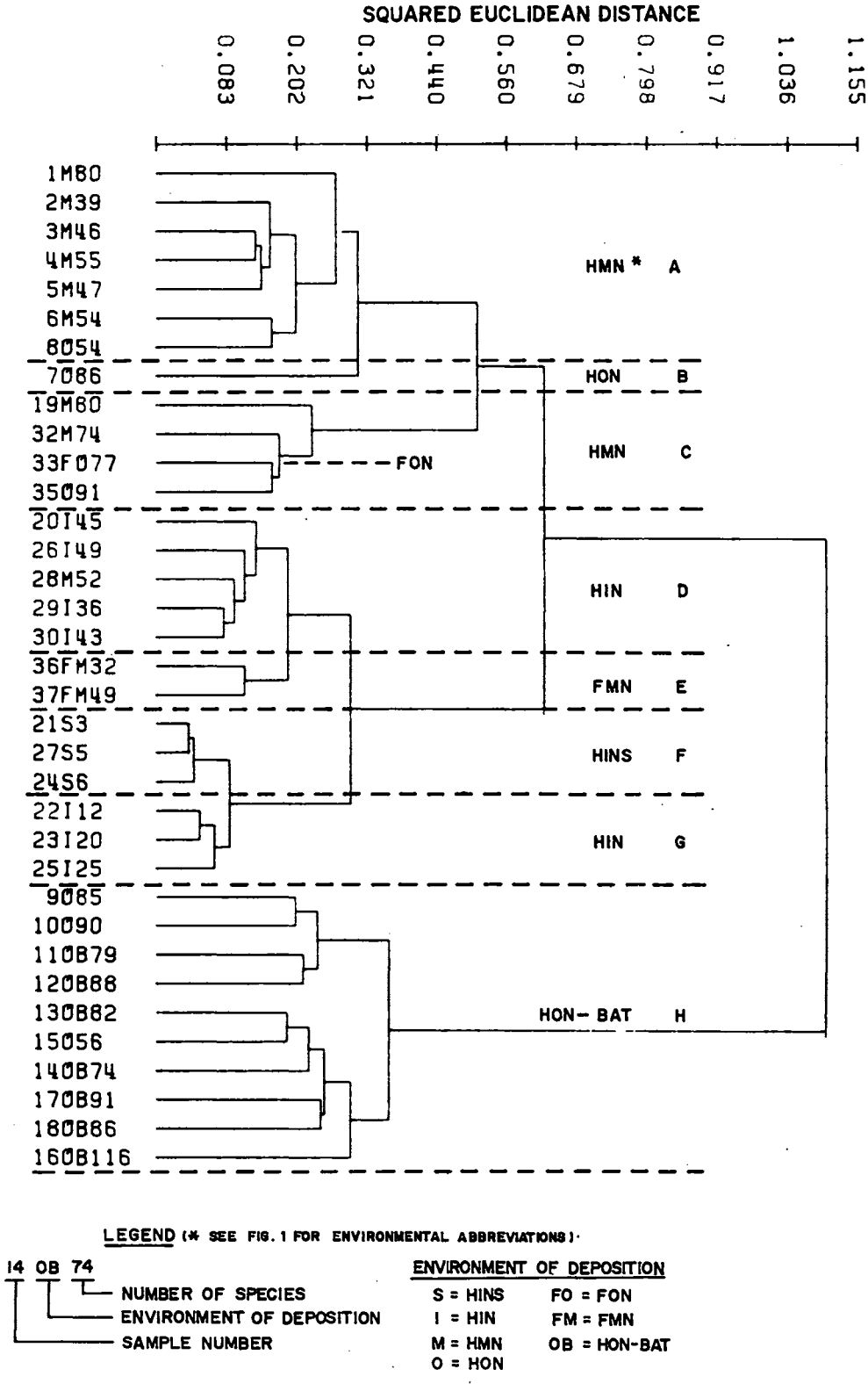
Distances between all combinations of p samples are calculated resulting in a $p \times p$ distance matrix and cluster analysis techniques operate on such a matrix to reveal the interrelationships between the various points.

Cluster analysis is used here to obtain as far as possible, homogenous groups of samples which in general correspond to particular environmental units. Both Q-mode (sample-sample comparison) and R-mode (species-species) cluster analysis were done (using presence-absence data) on each set of data to see if species groups reflected the sample groups found.

The data set consisting of approximately 3000 samples was analysed per well although wells with only a few samples were combined to obtain a total of about 40 samples. Prior to cluster analysis the whole data set was restudied and revised where necessary such that the interpretations used in the present study are assumed to be consistent and that variations due to interpretations made by many different investigators over a long period of time have been reduced to some extent.

In this study, the hierarchical clustering method of Ward was used together with the squared Euclidean distance coefficient. Ward's method results in compact clusters (Fig. 4) and obtains these in a way which minimises the increase in error sum of squares at each point in the clustering process. The effect of chaining (progressive overlap in dendrograms) is not as apparent as with other techniques such as single-linkage and complete-linkage cluster analysis and this simplifies the search for groups. To facilitate comparison and crosschecking of clusters, the same set of data was analysed using average linkage cluster analysis with the Jaccard similarity coefficient (Fig. 5). It can be seen that basically the same clusters are obtained although not necessarily in the same order.

At this point the question of optimally subdividing the dendrogram often arises. Although many methods have been discussed by various workers (Everitt, 1973; Demirmen, 1971), no one method has been universally accepted. In this study the intuitive approach of Demirmen is adopted. Basically he proposes that 'a class is that item or collection of items that, upon visual inspection of the dendrogram, tends to stand out from the neighbouring items or clusters'.



LEGEND (* SEE FIG. 1 FOR ENVIRONMENTAL ABBREVIATIONS)

14 OB 74

- NUMBER OF SPECIES
- ENVIRONMENT OF DEPOSITION
- SAMPLE NUMBER

ENVIRONMENT OF DEPOSITION

S = HINS	FO = FON
I = HIN	FM = FMN
M = HMN	OB = HON-BAT
O = HON	

Fig. 4. Q-mode cluster analysis of well A using Ward's method and the squared euclidean distance coefficient.

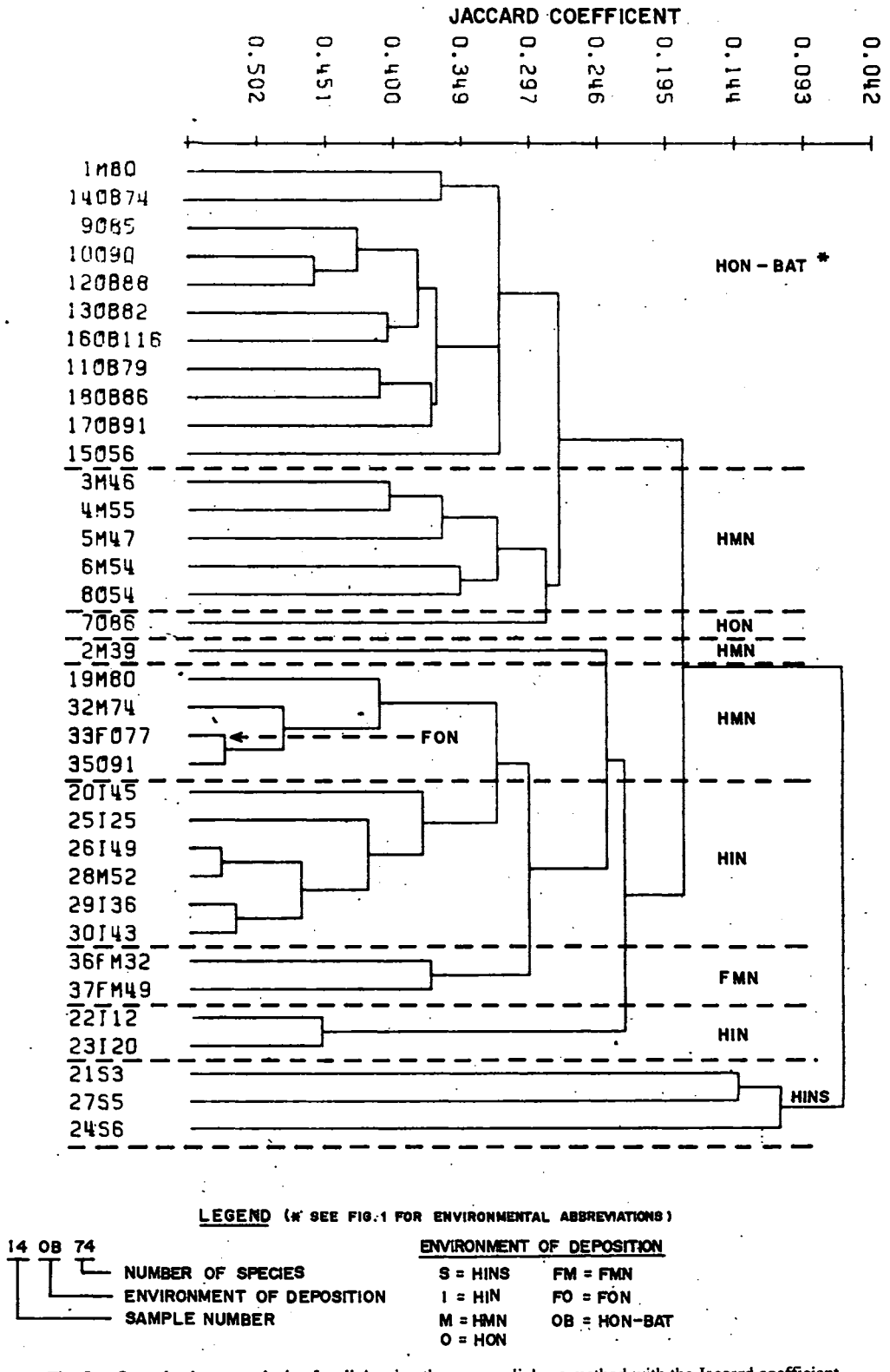


Fig. 5. Q-mode cluster analysis of well A using the average linkage method with the Jaccard coefficient.

Each dendrogram obtained in this study is first subdivided in this way and then studied for inconsistencies within each subgroup. It can be seen in Fig. 4 that the clusters are rather homogeneous in terms of interpreted depositional environment, a reflection of the consistency of the interpretations.

Apparent misfits, e.g. sample 8054, 33F077, and 28M52 were rechecked to see if any reason exists for the misfit. In this case, after rechecking sample 8054 and comparing it with the rest of the samples in Group A which were interpreted as Holomarine Middle Neritic (HMN), no significant differences were found indicating a probable inconsistency in the original interpretation of sample 8054. Sample 33F077, despite being the only Fluviomarine Outer Neritic sample in the set, clustered with the group it most closely resembled i.e. Group C. Sample 1M80 clustered with the Holomarine Middle Neritic group when using Ward's method but in the average linkage dendrogram (Fig. 5), it grouped together with the Outer Neritic to Bathyal group. This is believed to be due to the high number of species rather than to a significant number of deep water elements. The original interpretation was therefore retained. Many clusters were also obtained which consisted of a mixture of environments (Fig. 6). These samples were obviously fairly similar in terms of species content but the difficulty of consistently interpreting them probably meant that this cluster represented an 'intermediate' environment.

As it was the aim to obtain representative groups of samples with as little ambiguity as possible, 'intermediate' clusters such as these were subsequently removed from the study.

In this way it was possible to obtain groups of samples which could be identified with the environmental units in Fig. 1. The foraminiferal data of these groups could now be used to create the environmental range charts discussed in the next section.

ENVIRONMENTAL RANGE CHARTS

One of the utility options in PALAB is a range chart which shows, for all species, their percentage occurrence over the different environmental units. Such quantitative changes provide information regarding the bathymetric distribution of species which are useful criteria for environmental interpretation.

From the cluster analysis part of the study, each of the remaining samples (i.e. those that did not belong to clusters suggesting intermediate environments) were classified into one of the environmental units in Fig. 1. This information was subsequently used as input into the range chart option of the PALAB system.

The range chart which was generated inevitably contained a very large number of species some of which contributed an insignificant amount of information. These species, which occurred sporadically, were eliminated from the chart. In creating this chart only samples with 30 or more specimens were used. Subsequently the program MATEDIT was used to further reduce the noise by eliminating species which did not have an occurrence of 5% or more in at least one environment. This reduced the total number of sidewall samples incorporated in the final matrix to just over 1700 from the original 3000.

The quantitative variations exhibited by various species over the different environmental units are criteria which can be utilised for future interpretations. However, due to

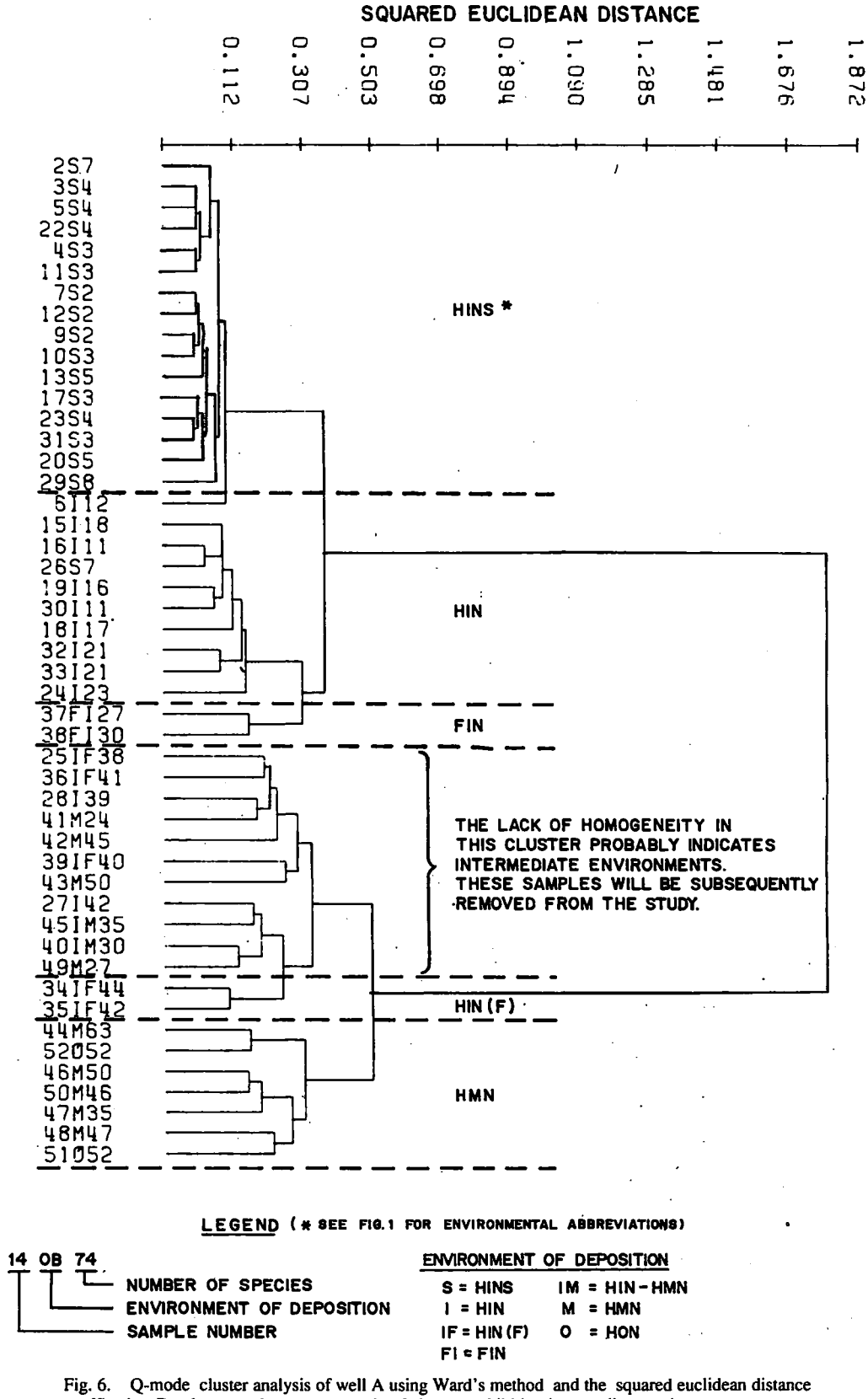


Fig. 6. Q-mode cluster analysis of well A using Ward's method and the squared euclidean distance coefficerit. Dendrogram shows an example of clusters exhibiting intermediate environments.

the size of the identification matrix (13 environments x 411 species), visual comparison of an incoming assemblage against the different assemblages is extremely difficult and automatic techniques are needed to effectively make use of the available information.

COMPUTER-ASSISTED IDENTIFICATION OF DEPOSITIONAL ENVIRONMENTS

One technique for computer-assisted identification compares the assemblage of a sample against a matrix of percent positive character (species) values (Sneath, 1979). The BASIC program presented there was developed initially to aid in the identification of bacteria which were difficult to identify by other methods. The technique is, however, widely applicable and is adapted here to determine (using presence-absence data) the best interpretations of the depositional environment of a sample based on its species content. The program has been modified to facilitate file handling and improve user-friendliness and runs on the VM/CMS (Virtual Machine/Conversational Monitor System), of the IBM 4341 mainframe computer.

Use of this technique requires that a data matrix be available against which an incoming sample can be compared. This identification matrix has the form given in Fig. 7 where each cell in the $q \times n$ matrix contains the percentage of positive occurrence of species in a particular environment. For example, in Fig. 7, species A occurs with a frequency of 70% in the environment 1 and with a frequency of 60% in 2. These figures are obtained in the following way. From the cluster analysis stage of the study, samples are allocated to one of the class of environmental units. Each unit, e.g. Fluviomarine Inner Neritic (FIN), contains its own set of samples which form a spectrum of possible assemblages describing this environment. Given that the set of FIN samples is 100 and that a species, Ammobaculites 1 (Am 1) is presently in 90 of these samples then:

SPECIES / ENVIRONMENT	A	B	. . . . .	n
1	70	80	. . . . .	
2	60	20	. . . . .	
.	.	.		
.	.	.		
.	.	.		
q	.	.		

Fig. 7. Schematic format of the identification matrix.

$$\text{Percent of positive occurrence for Am 1 in FIN} = \frac{90}{100} \times 100$$
$$= 90\%$$

A constraint that has been adopted here is that these percentages are never allowed to attain the values 0 or 100 and range between 1 and 99. One reason for this is that the calculation of the Willcox probability described below involves the multiplication of individual probabilities and the presence of a zero value would result in a likelihood of zero for an environmental unit. In a matrix of 411 species and environments which range from coastal plain to bathyal it is almost certain that there will be at least one zero value in each unit making it impossible to calculate Willcox probabilities. Secondly, from a statistical point of view, a value of 0 or 100% implies certainty and one can never be totally certain that a particular species A does not occur or always occurs in environment B. When creating an identification matrix using MATRANS, values of < 1 are automatically converted to 1 and values > 99 to 99 with negligible effects in practice (Sneath, 1979). Percentages are converted inside the program MATMOD to proportions, P_{ij} for the i^{th} species in environment j.

The range chart program described earlier calculates for each cell of the matrix certain statistics one of which is the percentage of positive occurrence. Using MATRANS, the relevant data from the range chart output can be automatically transferred into a file with the format of Fig. 7.

The identification program MATMOD compares an incoming sample (U) against the set of q environments in the data matrix and lists out in order of merit the best matches with their corresponding probability values. It first calls the selected identification matrix which is stored as a separate file, then the file LCLCODE which contains a list of the 1545 local species codes, e.g. Glm 4 for Glomospira 4, in use in SSB. The investigator is then required to enter one at a time the species codes for U, each entry being verified against LCLCODE before acceptance by the computer.

The Willcox Probability is based on Bayes' Theorem and is the likelihood of the incoming sample U against environment J divided by the sum of the likelihoods of U against all q environments (Willcox *et al.*, 1973). The likelihood L_{UJ} of U against J is:

$$L_{UJ} = \pi \left| U_i + P_{ij} - 1 \right|$$

Where U_i represents the i^{th} species in the identification matrix which if present in U is assigned the value 1 otherwise it has the value zero, P_{ij} is the probability of positive occurrence of species i in environment J, and n is the number of species in the identification matrix. When species i in the identification matrix matches up with one in U, then $U_i = 1$ and P_{ij} is used in the calculation. Because the system uses presence-absence species data, the probability of a negative occurrence (species i not present in U) is one minus the probability of a positive occurrence i.e. $(1 - P_{ij})$.

The Willcox Probability of U against J is given by:

$$P_w(UJ) = \frac{L_{UJ}}{\sum_{k=1}^{q} L_{UJ_k}}$$

In Fig. 8, the results for sample 743 m can be seen. Here the faunal assemblage has been identified with the environment LCP (Lower Coastal Plain) with a very high probability 0.999 and leaves little room for doubt. Some diagnostics are generated by the program and these are used as a further aid to access the calculated results. For example, in Fig. 8, there are not species against the result of LCP whereas for FINS (Fluviomarine Inner Neritic Shallow), the percentages of GLM4 and TRO5 of 9.6 and 6 respectively in the identification matrix are rather low and this therefore has a negative effect on the final probability value. 'Value in unknown' represents the presence or absence of a species in the

```
SAMPLE = 743 m            BEST IDENTIFICATION IS .. LCP
================          CURRENT INTERPRETATION ..
NO.SPECIES = 2       NO.POSITIVE MATCHES WITH IDENT.MATRIX= 2
NO. SPECIMENS =     20           P/B RATIO =  0.00
DIVERSITY INDICES.  YULE-SIMPSON =  1.79, FISHER ALPHA =  0.00

 TAXA             WILLCOX PROBABILITY
-------           -------------------
LCP                    0.9998
FINS                   0.0002
HINS                   0.0000

SPECIES AGAINST   ------>                LCP
SPECIES      PERCENT IN TAXON  VALUE IN UNKNOWN
--------     ----------------  ----------------
         ***     (NONE)     ***

SPECIES AGAINST   ------>                FINS
SPECIES      PERCENT IN TAXON  VALUE IN UNKNOWN
--------     ----------------  ----------------
GLM4                   9.6                +
TRO5                   6                  +

SPECIES AGAINST   ------>                HINS
SPECIES      PERCENT IN TAXON  VALUE IN UNKNOWN
--------     ----------------  ----------------
GLM4                   9                  +
RSPP                   99                 -
TRO5                   6.4                +

 SPECIES     AMT.    %            SCIENTIFIC NAME
---------   -----  -----  ----------------------------------------

TRO5          14    70.0  TROCHAMMINA MACRESCENS BRADY
GLM4           6    30.0  MILIAMMINA FUSCA (BRADY)
-------------------------------------------------------------------
```

Fig. 8. Example of well defined results.

sample being analysed. If a negative value is seen in this column, it means that the species referred to occurs with a high percentage in that environment but the fact that it is absent in the sample downgrades the probability value.

Fig. 9 is an example where the results are not straight forward. The best environment identified is HIN(F) (Holomarine Inner Neritic with some fluviomarine influences) but the probability associated with this determination is only 0.405. The next best environment is FIN (Fluviomarine Inner Neritic) with a probability of 0.294 followed by FMN (Fluviomarine Middle Neritic) with 0.269. These results would lead to the possible conclusion that the assembalge came from the deeper part of the Inner Neritic realm on the fringe of a delta.

CONCLUDING REMARKS

This system is a good tool for operational work, training and experimentation, is also easy to use and, being interactive, has a quick response time. A modified version of MATMOD, called BULKMAT, enables a complete well to be analysed at one time, the raw data having been retrieved from PALAB files by a preprocessor program. BULKMAT then accesses this file and processes it sample by sample. At the end of the run, a list of all samples in increasing order of depth together with their best interpretations and corresponding probabilities is printed. This gives an overall view of the sequence, and possible boundaries can be quickly located and checked. This facility is also very useful for revision work as well as for testing the behaviour of new or updated identification matrices and coefficients.

Fig. 10 shows a BULKMAT run on a section of a well. The summarised results shown are obtained from the final phase of BULKMAT analyses where all the calculations on individual samples have been stores in arrays and can be selectively listed. Selection can be made based on either:

(i) number of species in a sample

(ii) probability value of the best interpretation, P(1)

or (iii) both number of species and probability P(1)

To obtain statistically sound results *in relation to the set of environmental units in the identification matrix used*, one can therefore apply cutoff limits on both species number as well as probabilities. However, it should be noted that the construction of the identification matrix requires each environmental unit to contain sample sets which are as similar as possible and which at the same time contain enough variation to adequately describe the environment. Samples tested against this matrix may not relate to one environment much better than to another because its assemblage overlaps the two. The following examples are taken from the well Example A (Fig. 10).

SWS 822 m	:	P(HIN)	= 0.40	P(FMN)	= 0.35, P(HMN)	= 0.26
SWS 842 m	:	P(HIN)	= 0.64	P(HMN)	= 0.36, P(FMN)	= 0.00
SWS 1200 m	:	P(FON)	= 0.66	P(FMN)	= 0.34, P(HMN(F))	= 0.00

Looking at the values for SWS 822 m, it is possible to conclude that this assemblage is not represented very well in the matrix. This may be due to factors such as contamination or reworking. SWS 842 m and 1200 m suggest interpretations of HIN-HMN and FON-FMN

```
SAMPLE = 2018 m '       BEST IDENTIFICATION IS .. HIN(F)
================        CURRENT INTERPRETATION ..
NO.SPECIES = 22     NO.POSITIVE MATCHES WITH IDENT.MATRIX= 22
NO. SPECIMENS =   165          P/B RATIO =  0.01
DIVERSITY INDICES.  YULE-SIMPSON =  5.71, FISHER ALPHA =  7.14

 TAXA            WILLCOX PROBABILITY
-------          -------------------
HIN(F)                0.4055
FIN                   0.2943
FMN                   0.2690

SPECIES AGAINST   ------>              HIN(F)
SPECIES     PERCENT IN TAXON  VALUE IN UNKNOWN
--------    ----------------  ----------------
SGSPP               7.3                +

SPECIES AGAINST   ------>              FIN
SPECIES     PERCENT IN TAXON  VALUE IN UNKNOWN
--------    ----------------  ----------------
AS2                 5.9                +
BO8                 6.6                +
R18                 4.3                +
R6                  6.6                +
R8                  4.7                +
SGSPP               1                  +
TRI1                3.1                +

SPECIES AGAINST   ------>              FMN
SPECIES     PERCENT IN TAXON  VALUE IN UNKNOWN
--------    ----------------  ----------------
SGSPP               1                  +


 SPECIES    AMT.   %          SCIENTIFIC NAME
---------   -----  -----  ----------------------------------------

PLANKTOT      1           <----TOTAL NUMBER OF PLANKTONICS.
GSPP          0     0.0   GLOBIGERINA SP
H11          11     6.7   HAPLOPHRAGMOIDES NARIVAENSIS (BRONNIMANN)
AMSPP         1     0.6   AMMOBACULITES SP.
AM1           2     1.2   AMMOBACULITES EXIGUUS CUSHMAN & BRONNIMANN
BO8           9     5.5   BOLIVINITA SUBANGULARIS (BRADY)
BOSPP        12     7.3
U1/1A         2     1.2   UVIGERINA PROBOSCIDEA (SCHWAGER)
SGSPP         2     1.2
TRI1          1     0.6   TRIFARINA BRADYI CUSHMAN
AS2           2     1.2
RSPP         56    33.9
R2V1          9     5.5   AMMONIA KETIENZIENSIS (ISHIZAKI)
R6            1     0.6   AMMONIA ANNECTENS (PARKER & JONES)
R26V1/62      2     1.2
R8            5     3.0   PSEUDOROTALIA SCHROETERIANA (PARKER & JONES)
R18           1     0.6   PSEUDOROTALIA FIJIENSIS (CUSHMAN)
ELPHSPP       1     0.6
ELPH1         1     0.6   CELLANTHUS KOEBOEENSE (LEROY)
CISPP        35    21.2
NONSPP        1     0.6
CI2           5     3.0   HETEROLEPA DUTEMPLEI (D'ORBIGNY)
OPSPP         6     3.6
-------------------------------------------------------------------
```

Fig. 9. Example of results that do not point clearly to a single environment but rather suggests one and shows a tendency in a certain direction. In this case, the dominant environment being HIN(F) with a possible inclination towards the deeper part of FIN.

WELL : EXAMPLE A (NS = NUMBER OF SPECIES)

DEPTH (M)	NS	PRESENT INTERP.	BEST THREE ENVIRONMENTS & PROBABILITIES ENV(1)	P(1)	ENV(2)	P(2)	ENV(3)	P(3)
648	24	HIN	HIN	1.00	FMN	.00	FIN	.00
654	19	HIN	HIN	1.00	FMN	.00	HIN(F)	.00
689	53	HIN-HMN	HMN	1.00	HMN(F)	.00	HIN	.00
708	69	HIN-HMN	HMN	1.00	HON	.00	HON-BAT	.00
732	53	HIN	HMN	1.00	HIN	.00	HMN(F)	.00
750	7	HINS	HINS	.98	FINS	.01	FIN	.01
805	22	HIN	HIN	.98	FMN	.02	HMN	.00
822	35	HIN	HIN	.40	FMN	.35	HMN	.26
842	36	HIN	HIN	.64	HMN	.36	FMN	.00
858	14	HIN	HIN	1.00	FIN	.00	HINS	.00
875	25	HIN	HIN	1.00	FMN	.00	HMN	.00
888	50	HMN	HMN	1.00	HMN(F)	.00	FMN	.00
904	61	HMN	HMN	1.00	HON	.00	HON-BAT	.00
940	7	HINS(F)	FINS	.59	HINS	.39	FIN	.02
954	31	HIN	HIN	.98	FMN	.02	HMN(F)	.00
971	52	HIN	HMN	1.00	HMN(F)	.00	HIN	.00
981	34	HIN(F)	HIN	.82	HMN(F)	.08	HMN	.05
991	19	HIN(F)	HIN	.92	FIN	.05	FMN	.03
1015	21	HIN(F)	FMN	.99	HMN(F)	.01	HIN	.00
1027	40	HMN(F)	HMN	1.00	HMN(F)	.00	FMN	.00
1047	64	HMN(F)	HMN	1.00	HON	.00	HMN(F)	.00
1067	72	HMN	HMN	1.00	HON	.00	HON(F)	.00
1079	4	HINS	FINS	1.00	FIN	.00	LCP	.00
1109	68	HMN	HMN	1.00	HON(F)	.00	HON	.00
1121	12	HIN(F)	FIN	.52	FINS	.46	HINS	.02
1136	48	HIN(F)	HMN	.96	FMN	.04	HMN(F)	.00
1151	48	FMN	HMN	.96	HMN(F)	.03	FMN	.00
1171	50	HMN	HMN	1.00	FMN	.00	HMN(F)	.00
1185	84	HMN	HON(F)	1.00	HMN	.00	HON	.00
1200	32	HMN	FON	.66	FMN	.34	HMN(F)	.00
1213	61	HMN	HMN	1.00	HON(F)	.00	HON-BAT	.00
1224	57	HMN	HMN	1.00	HON	.00	HMN(F)	.00

Fig. 10. Summary of results for part of a well showing the manual and computer generated interpretations with their respective probabilities.

respectively. A point worth mentioning is that a result is always calculated regardless of the amount of input data so that before accepting a result, it is important to know how much information has been used in its calculation. For example, the calculation of diversity on samples with only one species has little or no meaning and should be ignored.

The system is also a general tool in the sense that it can be used with different identification matrices for different purposes. For example, if the fauna exhibited a large variation between different geographic areas, separate matrices could be constructed for these areas. Likewise, one could construct a matrix of pollen versus time. These matrices, stores as separate files, can be called from within the main program (see text Fig. 3) whenever required.

Although the system at present uses only presence/absence of data in its analyses, it provides not only a foundation from which future work can develop but also serves as a base for consistent interpretations. For such a computer-assisted system to be measurably improved, one needs to take into account many other factors that investigators routinely utilise, some of which are the relative amounts of and subjective weighting assigned to particular species, reworking, preservation, contamination and sizes of specimens, the species composition in an assemblage, the relationships between different assemblages in the sequence as well as a background of geological knowledge of the area. Some of these criteria are extremely difficult to quantify and investigators develop (often personally unique) 'rules of thumb' based on years of practical experience.

Expert systems or knowledge-based systems (Feigenbaum and McCorduck, 1984; Hayes-Roth, Waterman and Lenat, 1983) are computer programs that combine such 'rules of thumb' or heuristics with a knowledge-base and an inference procedure to enable them to produce an interpretation or analysis of a problem at a level similar to that of an expert. It would appear then that a logical future development of the present system would revolve around the concept of an expert system; one capable of carrying out a dialogue with its human operator as well as being able to explain the line of reasoning for arriving at particular conclusions.

REFERENCES

DEMIRMEN, F., 1971. Mathematical Search Procedures in Facies Modeling in Sedimentary Rocks. *In* Merriam, D.F. (ed.) *Mathematical Models of Sedimentary Processes*. New York, Plenum Press, p. 81-114.

EVERITT, B., 1977. *Cluster Analysis*. London, Heinemann Educational Books, 122 p.

FEIGENBAUM, E.A. and MCCORDUCK, Pamela, 1984. *The Fifth Generation - Artificial Intelligence and Japan's Computer Challenge to the World*. London, Michael Joseph Ltd., 275 p.

HAYES-ROTH, F., WATERMAN, D.A. and LENAT, D.B., 1983. *Building Expert Systems*. Addison Wesley Pub. Co. Inc., 444 p.

SNEATH, P.H. A. and SOPKAL, R.R., 1973. *Numerical Taxonomy*. W.H. Freeman & Co., 573 p.

SNEATH, P.H.A., 1979. BASIC Program for Identification of an Unknown with Presence-Absence Data Against an Identification Matrix of Percent Positive Characters. *Computers and Geosciences*, V. 5, no. 2, p. 195-213.

WISHART, D., 1978. *CLUSTAN User Manual*. Edinburgh University.

Manuscript received 23 March 1987