Geoscience Canada

PALEOSCENE #8. Biostatistics in Paleontology

Brian Jones

Volume 15, Number 1, March 1988

URI: https://id.erudit.org/iderudit/geocan15_1art01

See table of contents

Publisher(s)

The Geological Association of Canada

ISSN

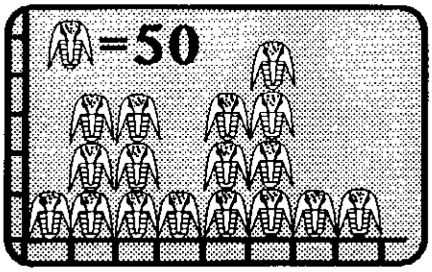
0315-0941 (print)
1911-4850 (digital)

Explore this journal

Cite this article

Jones, B. (1988). PALEOSCENE #8. Biostatistics in Paleontology. *Geoscience Canada*, *15*(1), 3–22.

Articles



PALEOSCENE #8. Biostatistics in Paleontology

Brian Jones
Paleontological Collections
Department of Geology
University of Alberta
Edmonton, Alberta T6G 2E3

Introduction

Paleontology is a multi-faceted subject that involves taxonomy and community analysis as two of its cornerstones. Both qualitative and quantitative approaches have been incorporated into the various methods employed in such analyses. Each paleontologist obviously has his or her own preference that is commonly expressed in terms of the use of qualitative or quantitative methods.

The use of statistics has gained favour among some paleontologists since it offers hope for the precise mathematical delineation of species or communities. With respect to species this is based on the assumption that most species in paleontology are essentially defined on morphology. Thus it should, at least in theory, be possible to accurately define species on the basis of measured parameters. Attempts to do this (*e.g.*, Sokal, 1966; Sokal and Sneath, 1963, Jardine and Sibson, 1971) have met with varying degrees of success. This is essentially because of the problems (1) inherent with deriving the measured parameters, and (2) inherent to the statistical methods used. In addition, a certain proportion of the problem must be laid at the feet of paleontologists themselves who refuse to use any statistics in their taxonomy or establishment of communities, or conversely, use statistical procedures without any consideration of the qualitative observational data. Obviously, a more central approach should be taken because the combination of qualitative and quantitative data can yield very sound paleontological conclusions.

The science of statistics has been applied to virtually every facet of paleontology. It is impossible in the context of this paper to review all of these methods and their application in paleontology. Consequently, attention has been focussed on (1) univariate and bivariate statistics that can be used in virtually any taxonomic study, and (2) cluster analysis, which is a powerful, but underused, tool in the delineation of communities or fossil associations. In each case, the format involves: (1) an outline of the basic statistics involved, as well as the formulae needed for their derivation; (2) a discussion of the problems that may be encountered in the application of these statistics; and (3) a listing of selected studies that have utilized the statistics being discussed; these should serve to give the interested reader an insight into the various ways in which the statistics can be used. Since most of the author's own studies have involved the Silurian and Devonian brachiopods of Arctic Canada, examples have been drawn from those in order to demonstrate the points being made. In addition, two complete sets of data are provided in Appendices 1 and 2 so that interested students can utilize them in their own derivation of the statistics used in this paper.

The processing of statistical data can pose a problem in itself. In years past this problem was overcome simply by using various statistical programs that are associated with mainframe computers. In some cases, however, it is necessary to write programs that are specifically tailored to the problem at hand. The latter course of action also has the advantage of leading to a greater understanding of the mechanics behind the statistics being used. The last two years or so have seen the advent of microcomputers with large memories and powerful programs. Indeed, virtually any of the univariate or bivariate statistics outlined in this paper can be derived from standard programs associated with most microcomputers. A note of caution must be added in this respect, however, for many of these programs do not document the actual methods used in the derivation of the statistics. In some cases this is not critical because it involves statistics that can only be calculated in one way (*e.g.*, mean). Conversely, many graphical programs allow the computation of the equation relating the x and y variable without specifying the method used. Given that there are many different methods of deriving such equations, each with its specific set of problems, it is imperative that the method of calculation be known. In some respects, lack of knowledge of the methodology negates the equation that is calculated. Nevertheless, the usefulness of microcomputers cannot be underestimated. In addition to the calculation of the statistics, many of the present microcomputers offer excellent graphic packages that can be easily used to present the statistical information obtained from any analysis. To demonstrate this point, all of the figures in this paper, with the exception of Figures 4 and 13, have been generated directly from a Macintosh Plus® computer with the aid of programs such as MacDraw®, Macpaint® and Superpaint®.

Acquisition of Data

The starting point of any statistical analysis is obviously the acquisition of the data base. Great care must be taken in this respect since poor data will invariably lead to even poorer statistical results. If the statistics are to be used for taxonomic purposes, such problems can be reduced to a minimum by (1) carefully selecting the parameters that are to be used, and (2) exercising great care in the actual measurement of the parameters.

Selection of parameters. In most cases the selected parameters will be (1) a straight line measurement (*e.g.*, length of shell - Figure 1A); (2) an angle (*e.g.*, apical angle - Figure 1A); or (3) the numbers of a particular feature that are present (*e.g.*, number of ribs on a brachiopod shell). Further parameters can be generated by combining any of the above parameters (*e.g.*, length/width ratio). Such derived parameters are commonly very useful although some care must be exercised if they are used in the plotting of bivariate graphs. For example, a plot of length *versus* length/width ratio will produce a graph and a correlation coefficient which is, at least in part, a function of the fact that length is involved in both parameters.

An inherent problem with some of these parameters stems from the fact that they are straight line measurements. For example, the length of a shell as defined in Figures 1A and 1B is actually the shortest distance between the anterior and posterior ends of the shell — it is not a measurement of the actual length of shell material between those two points. Thus, shells of vastly different convexities will have the same length if the standard measurement of shell length is used. In all probability the shell length should be measured along the shell surface that stretches from the posterior to the anterior margins (Figure 1B). While this may well be the best measurement, it does leave the paleontologist with the problem of trying to measure the distance on these curved surfaces.

The computation of ratios such as the length/width ratio is commonly done on the premise that it will give an indication of the shape of the shell being considered. This must, however, be treated with caution since a ratio such as the length/width ratio is in reality only a crude measure of the shell shape. This is easily demonstrated by the fact that an ellipse and a rectangle, while being substantially different shapes, may have the same length/width ratio (Figure 1C).

Measurement of parameters. Once the parameters have been selected for study it is important to take every care in deriving their values from each fossil being considered. There will always be some operator error related to the use of the measuring instruments or in some cases to the measuring instrument itself. The aim is therefore one of reducing that error to a minimum. The best possible test for determining the accuracy of measurement is to have at least one set of fossils measured three times: (1) once by the primary investigator; (2) a second time by the primary investigator some days after the initial measuring; and (3) by somebody other than the primary investigator, preferably with a different measuring instrument. Each set of measurements should then be analysed and tested to see if significantly different statistical measures are being derived from the different data sets. If this should be the case, then the causes of such differences must be determined before further measuring is done.

Collection of data for comparison of faunas. The comparison of faunas from different localities is obviously dependent on the quality of the collections made from those localities. Poor or biased collections will lead to poor statistical results. Although the collection of such data might appear to be straightforward, it is actually fraught with problems that can seriously affect the final results. If one of the objects of a study is to delineate communities or faunal associations, there are a number of ways in which such collections can be made in order to minimize the problems typically associated with this stage of the data acquisition, namely: (1) collect all the fossils that can be found at a given outcrop with particular care being taken to ensure that all sizes of fossils are collected; (2) select a given area and collect all the fossils in that area — for example, if bedding plane exposures are available then one square metre quadrants can be used as the standard sampling area; or (3) collect as much material from an outcrop as possible within a given time interval — for example, allocate one hour per outcrop. Each of these techniques, as well as any other techniques that are used, has its own set of associated problems. The paleontologist must select a collecting method which is deemed best suited to the material and outcrops being considered. At all times the paleontologist should be aware of such problems and select the techniques which minimize the possible sources of error. It should be remembered, however, that these types of collecting problems are common to all aspects of paleontology and not just those involving statistical procedures.

Micropaleontologists commonly follow a different collecting technique since the success of their collecting is uncertain until they have processed their bulk samples on return to their laboratories. Many would claim that such techniques are without bias for the sampling commonly involves the collection of a certain amount of sample from fixed intervals throughout a section (*e.g.*, 5 kg of sample at every 5 m throughout the section). Even this technique, however, has biases because the micropaleontologists (1) commonly sample the rocks which they know from experience have a higher probability of yielding the microfossils that they are interested in; (2) must assume that there is no contamination during their processing; (3) must assume that no fossils are destroyed during processing; and (4) must assume that all specimens are actually picked from the residue that is left after processing. As with the megafossils, these potential problems can be minimized by careful sampling and processing.

Univariate Statistics

As the name implies these are statistics that refer to a single variable of a given data set. Such basic statistics are important because they allow detailed description and, ultimately, comparison of a single variable from different data sets. The following statistics are important:

Minimum = Lowest value determined for variable

Maximum = Highest value determined for variable

Range = Maximum - minimum

Mean ($\bar{x}$) = $\frac{\sum x}{n}$

Standard Deviation (**s**) = $\frac{\sqrt{\sum (x - \bar{x})^2}}{(n - 1)}$

Coefficient of variation (**V**) = $\frac{100\,(sx)}{\bar{x}}$

Standard error of the mean ($s_{\bar{x}}$) = $\frac{sx}{\sqrt{n}}$

where **n** = sample size
x = variable being considered.

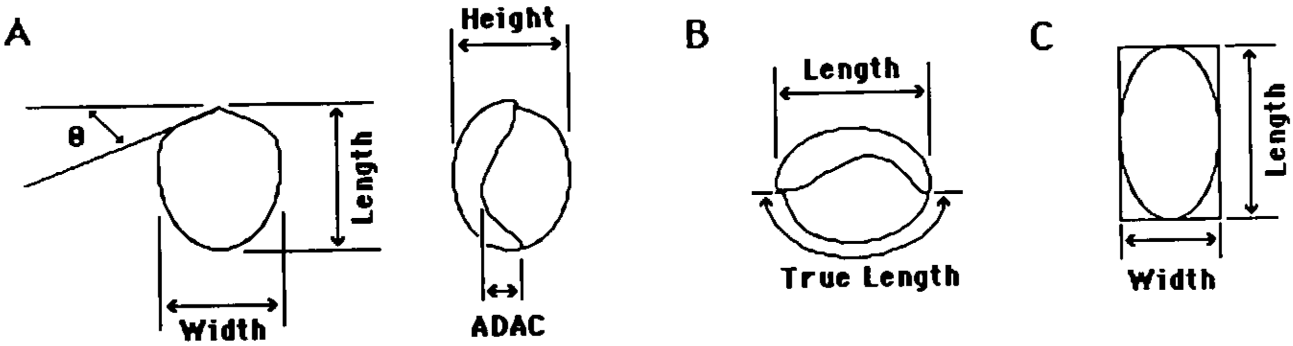


Figure 1 **(A)** *Definition of parameters that have been used in the statistical analysis of the brachiopod* Atrypoidea. *(After Jones, 1974).*
(B) *Comparison of straight line measurement of shell length* **(L)** *as opposed to true length measured along shell surface* **(TL)**.
(C) *Diagrammatic illustration of a rectangle and an ellipse that have the same length/width ratios.*

Table 1 gives examples of these statistics derived from data for assemblages W1 (*A. phoca*) and X14 (*A. foxi* f. A) that are given in Appendices 1 and 2.

Portrayal of data. Univariate data are typically portrayed as a histogram which gives a very good and clear visual impression of the distribution of the recorded values throughout the total recorded range of the variable (Figure 2). Attention has been focussed on the form of histograms since they can yield information about population dynamics. Before such studies can be commenced, however, it is important to be satisfied that the form of the histogram is not due to (1) poor sampling, or (2) closure. Poor sampling, usually caused by small sample size or bias to large specimens can produce histograms that have more than one mode, badly skewed histograms, or histograms with very little spread. The easier way of avoiding these potential problems is to collect large samples and avoid collecting bias. If this is impossible, as is the case in certain instances, then the use of histograms for the interpretation of population dynamics should be avoided. The problem of closure, which is common to virtually all paleontological data, occurs when a particular parameter has an upper limit. For example, a brachiopod may have a well developed sulcus with 1 to 5 ribs in it, depending upon the age of the individual shell. If a particular population of that brachiopod contains mostly adults then the resultant histogram of the number of ribs in the sulcus will be skewed to the right with the mode in the 5 rib category. While there will be a tail developed to the left (1 to 4 ribs) there will be no tail to the right for the simple reason that the brachiopod never has more than 5 ribs in its sulcus. In cases such as this, the skewed distribution is correct — it is not an artifact of poor collecting or poor data collection. If sampling was good and there is no evidence of closure then the shape of the histogram can yield valuable information about the population dynamics. Craig and Hallam (1963) clearly demonstrated that the shape of a histogram is a function of growth rates, recruitment rates and mortality rates. Interested readers should refer to Craig and Hallam's (1963) classic work for more information on how these factors control the form of histograms.

Table 1 **Univariant statistics for length, width, and apical angle for assemblages X14 (*Atrypoidea foxi* f. A) and W1 (*Atrypoidea phoca*). *SEM* = standard error of the mean; *Std. Dev.* = standard deviation. (Original data given in Appendices 1 and 2.)**

	X14			W1		
Statistic	**Length**	**Width**	**AA**	**Length**	**Width**	**AA**
Minimum	1.67	1.47	35.00	0.96	0.93	40.00
Maximum	3.94	3.03	60.00	2.19	2.15	76.00
Range	2.27	1.56	25.00	1.23	1.22	36.00
Mean	2.59	2.27	47.94	1.69	1.64	55.32
SEM	0.05	0.04	0.61	0.04	0.03	0.89
Std. Dev.	0.44	0.33	5.06	0.27	0.26	6.82
Coeff. Variation	16.87	14.58	10.55	15.97	16.05	12.33

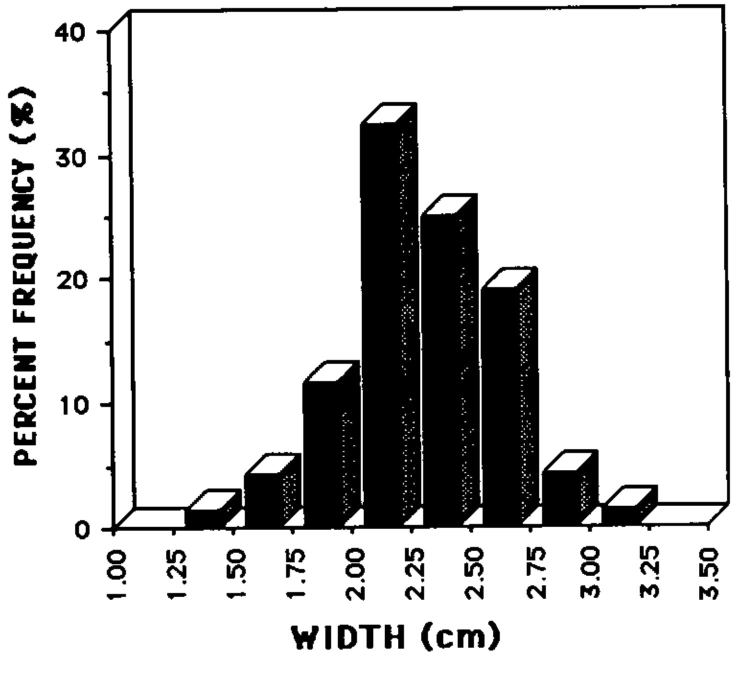


Figure 2 *Example of a histogram for shell width of* Atrypoidea foxi *f. A (assemblage X14 - Appendix 2).*

The impact of histograms can be accentuated by indicating the position of the mean value, median value and mode on that diagram. Histograms, however, have a number of disadvantages which seriously hinder their use, namely: (1) statistical parameters other than mean, median and mode cannot be shown on them, and (2) their drawing requires a certain vertical height; if a number of samples is being shown on the same diagram, a space problem may result. In certain cases, the latter problem may be more serious than the former problem.

A very useful alternative involves the use of Dice-diagrams which are relatively easy to construct (Figure 3). The Dice-diagrams are particularly useful because (1) they allow clear and accurate portrayal of the minimum, maximum (and hence the range), mean, standard error of the mean and the standard deviation (Figure 3), and (2) the Dice-diagrams are relatively small in size and are thus more convenient to use in situations where numerous samples are being compared (Figure 4).

Comparison of data. Both histograms and Dice-diagrams can be easily used to examine the similarity between two variables. If only two or three samples are involved then histograms can be used and a good visual impression of the similarity of two samples can be obtained (Figure 5). It must be emphasized, however, that this is strictly a visual comparison and, apart from a comparison of the means, there can be no statistical measure of the similarity or dissimilarity between the various data sets being compared. In such a situation, judgement as to similarities and/or dissimilarities can be severely impaired by the wish or need to see differences. If this is the case, the whole reason for using statistics is negated.

Dice-diagrams can be used for the comparison of numerous samples with great effect (Figure 4). This is especially true in situations where stratigraphically ordered samples are involved (Figure 4; see also Table 2). Like histograms, comparison using Dice-diagrams gives a good visual impression of differences and similarities. Dice-diagrams, however, have a distinct advantage in that they can be used to test whether the mean values of two data sets are statistically the same or different. Since the central box of the Dice-diagram (defined by the mean ± 2 SEM) is a very good approximation of the 95% confidence limits about the mean, it can be used for comparing the mean of one data set with the mean of another. If the central boxes of the two samples overlap, then the means cannot be considered statistically different (Figure 6). Conversely, if the central boxes do not overlap, the mean values can be considered statistically different (Figure 6).

When only two samples are involved the comparison of the mean values in this manner is straightforward since the result is either (1) the mean values are different, or (2) the mean values are not different (Figure 6). In situations involving 3 or more samples, however, some care must be exercised since unusual results can be obtained in certain circumstances. This problem can be demonstrated by considering a contrived example involving samples A, B and C (Figure 7). Comparison of the mean values of A, B and C using the method just described yields the result that:

(1) A is not significantly different from B,
(2) B is not significantly different from C,
(3) C is significantly different from A.

Combining results 1 and 2 would lead to the conclusion that the mean value of C is not significantly different from the mean value of A — a conclusion that is in direct contradiction to result 3. This enigma is not unique to comparisons generated from the use of Dice-diagrams since similar results can be obtained using other statistical tests. This type of result commonly occurs in situations in which there is a distinct increase or decrease in the mean value of a parameter from a set of samples arranged in stratigraphic order (Figures 4 and 7). In such situations, emphasis should be placed on the trend rather than the comparison between successive pairs of samples.

Studies with univariate comparisons involving histograms. *Brachiopods:* Amsden (1974), Copper (1986), Harper (1969), Hewitt and Hurst (1977), Imbrie (1956), Jones (1974, 1977, 1978, 1979a, 1981), Watkins (1975); *Bivalves:* Craig and Hallam (1963), Hallam (1967), Sørensen (1984); *Microfossils:* Nyberg and Schopf (1984).

Studies with univariate comparisons involving Dice-Diagrams. *Methodology:* Jones (1974); *Brachiopods:* Alvarez (1982, 1984), Jones (1977, 1978, 1979a, 1981); *Gastropods:* Graus (1974 - variation of Dice-diagram); *Radiolarians:* Lazarus *et al.* (1985 - variation of Dice-Diagram); *Microcrinoids:* Lane and Sevastopulo (1985 - variation of Dice-Diagram).

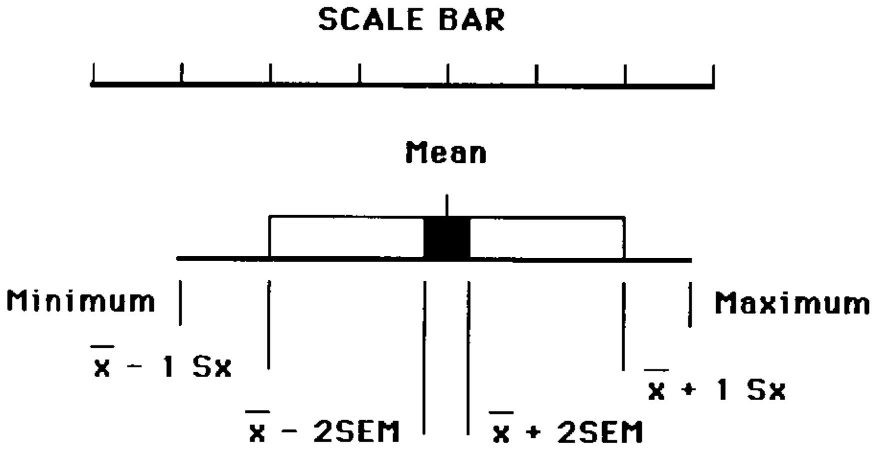


Figure 3 *Construction of Dice-diagram. **(1)** The base line stretches from the minimum to the maximum value of the parameter; **(2)** the open box is defined by the mean plus and minus 1 standard deviation; **(3)** the filled box is defined by the mean plus and minus 2 standard errors of the mean (SEM); and **(4)** the small vertical tick at the top of the filled box is defined by the mean value of the parameter.*

Table 2 **Example of correlation table for parameters derived from *Atrypoidea phoca* from locality GF-F (Goose Fjord, Ellesmere Island). (From Jones, 1981, Table 3).**
L **= length;** ***W*** **= width;** ***H*** **= height;** ***ADAC*** **= absolute deflection of anterior commissure;** ***RDAC*** **= relative deflection of anterior commissure;** ***AA*** **= apical angle;** ***PCI*** **= plan circularity index;** ***SCI*** **= side circularity index.**

	L	W	H	ADAC	RDAC	AA	PCI	SCI
L	1.00	0.86	0.86	0.75	0.52	0.25	0.48	0.55
W		1.00	0.77	0.65	0.43	0.10	-0.02	0.54
H			1.00	0.88	0.59	0.34	0.37	0.89
ADAC				1.00	0.89	0.35	0.36	0.77
RDAC					1.00	0.29	0.28	0.53
AA						1.00	0.34	0.33
PCI							1.00	0.18
SCI								1.00

$$\mathrm{RDAC} = \frac{\mathrm{ADAC} \times 100}{\mathrm{H}} \qquad \mathrm{PCI} = \frac{\mathrm{L} \times 100}{\mathrm{L} + \mathrm{W}} \qquad \mathrm{SCI} = \frac{\mathrm{H} \times 100}{\mathrm{L} + \mathrm{H}}$$

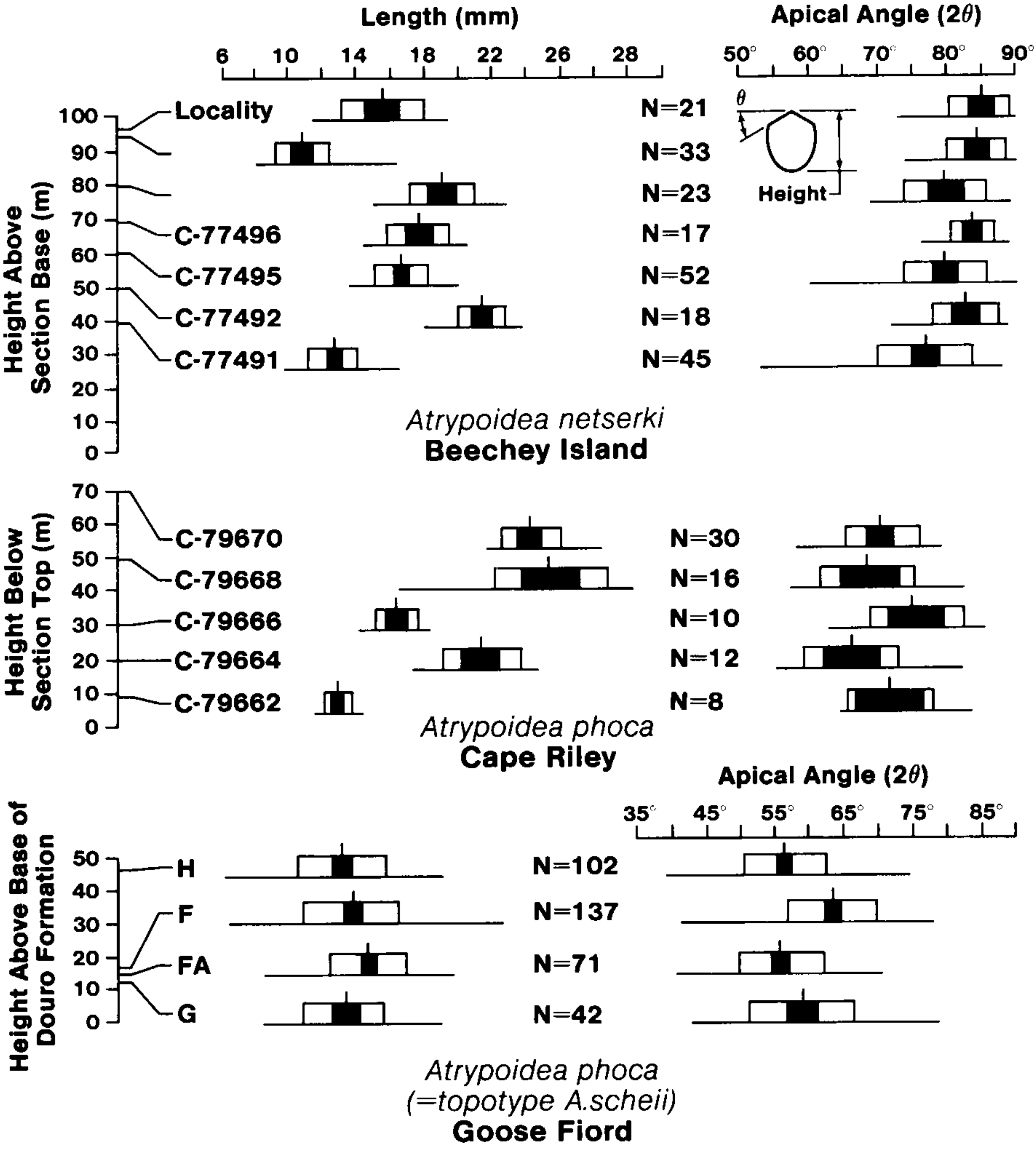


Figure 4 *Example of use of Dice-diagrams for the comparison of the length and the apical angle of* Atrypoidea phoca *and* Atrypoidea netserki *from Silurian strata of Arctic Canada. (From Jones, 1981, fig. 2).*

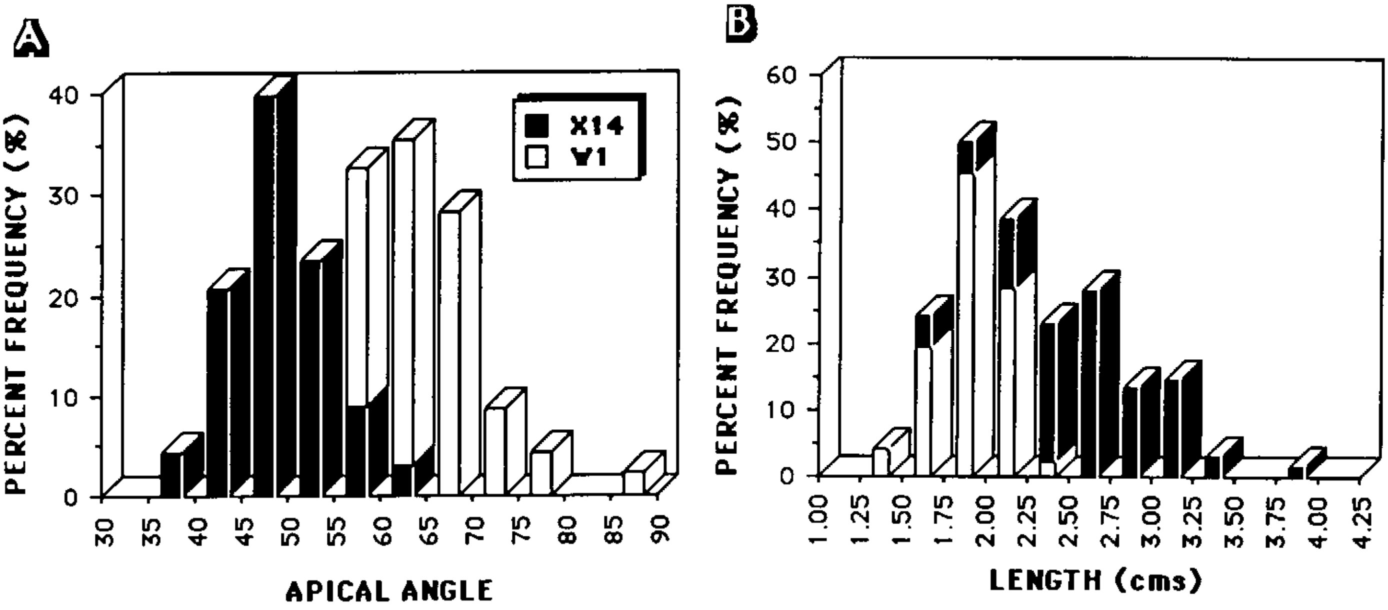
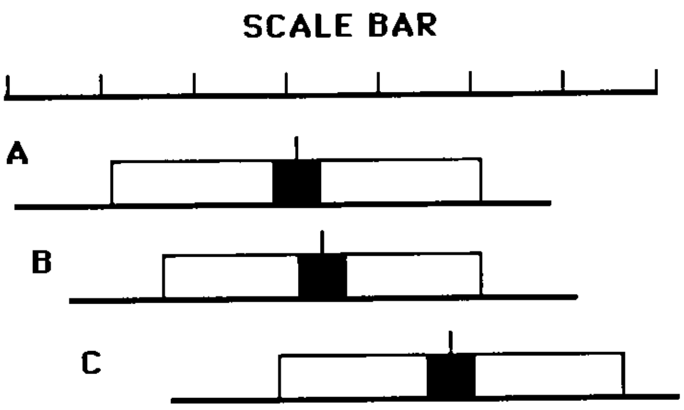


Figure 5 *Comparison of apical angle (**A**) and length (**B**) for* Atrypoidea foxi f. *A (assemblage X14) and* Atrypoidea phoca *(assemblage W1). See Appendices 1 and 2 for data.*



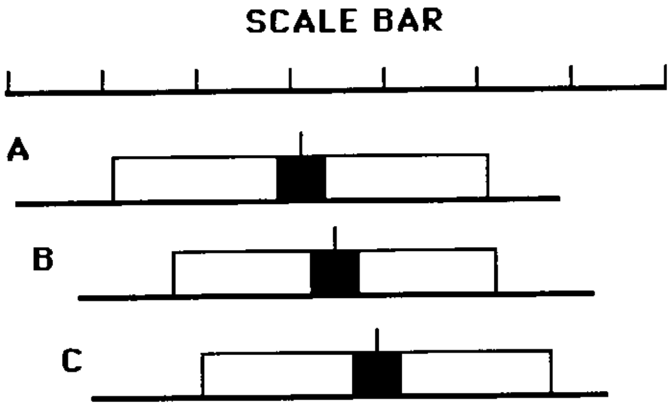


Figure 6 *Comparison of mean values using Dice-diagrams. Samples A and B are not significantly different; sample C is significantly different from both samples A and B.*

Figure 7 *Dice-diagrams in which comparison of the mean values yield the results that A = B, B = C but C ≠ A. See text for further discussion.*

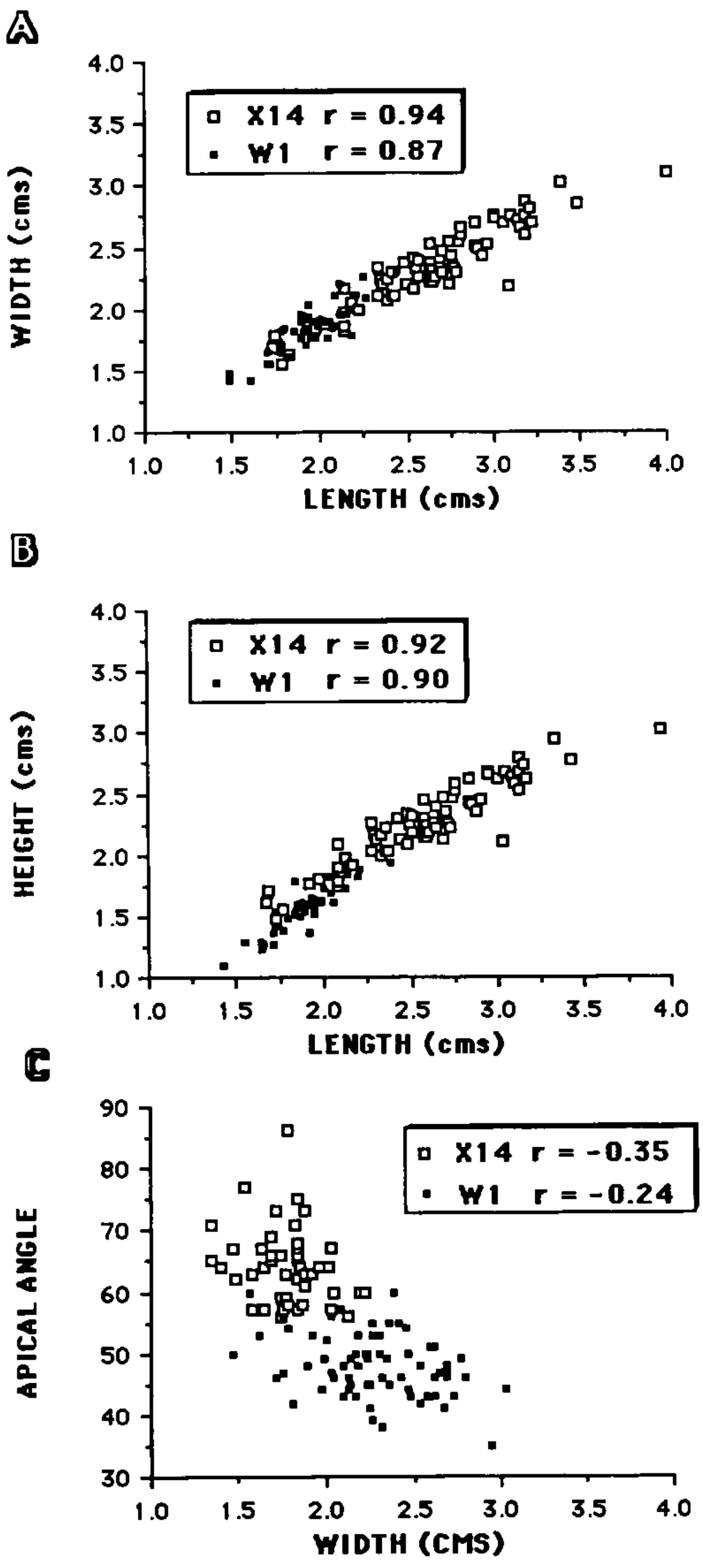


Figure 8 *Examples of bivariate graphs that compare length versus width (A), length versus height (B), and width versus apical angle (C) for* Atrypoidea foxi f. A *(assemblage X14) and* Atrypoidea phoca *(assemblage W1). Note high correlation coefficients (r) associated with each plot.*

Bivariate Statistics

As the name implies, these types of statistics investigate the statistical relationship between two variable parameters that are usually termed the X and Y variables. The simplest way of portraying the relationship between the two variables is a bivariate graph (*e.g.*, Figure 8). Once such a graph has been constructed it is necessary to derive statistical measures of the relationship between the two variables. These include correlation coefficient and determination of growth equations.

Correlation coefficient. One of the basic statistics applicable to bivariate analysis is the correlation coefficient (r) which is calculated by:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

The value of **r** will range from 0.0 (no correlation) to plus or minus 1.0 (perfect correlation). Values between 0 and 1 give a measure of the correlation between variables X and Y. Thus, a value of 0.50 or less is generally indicative of a poor correlation (*e.g.*, Figure 9) while a value of 0.80 or above is generally indicative of a good correlation (*e.g.*, Figure 8). It is important to note, however, that a value of 0.80 does not mean that the correlation is twice as good as a correlation coefficient of 0.40.

Relative growth. Bivariate statistics provide a method of studying relative growth (*i.e.*, how one variable changes relative to another variable). This is important in many taxonomic studies because the paleontologist is commonly interested in how one parameter (*e.g.*, width of shell) behaves relative to another parameter (*e.g.*, shell length) during ontogeny. The bivariate graph, which provides a good visual impression of this relationship, is of paramount importance in this type of study. Visual inspection can determine whether allometric (points follow a curved line — Figure 10) or isometric (points follow a straight line) growth relationships are present. This basic assessment is necessary because different statistics are involved in the determination of growth lines (= line of relative growth) for allometric and isometric growth.

Allometric Growth. The growth line for allometric growth is given by:

$$y = b x^{\mathbf{a}} \quad \text{, where}$$

x and y = variables being considered, and,
b and **a** = mathematical parameters which vary according to the nature of the curve.

The fact that the growth line is a curve indicates that the two variables are increasing at different rates and that the x/y ratio changes as the animal grows (Imbrie, 1956, p. 228). Although the specific growth rates (*i.e.*, rate at which the logarithm of a variable is changing) are unequal, they maintain a ratio that is constant and equal to **a** (Imbrie, 1956, p. 228). The value of **a**, which is the growth ratio, is the slope of the growth line on double logarithmic graph paper while **b** (initial growth index) is the absolute value of **y** when **x** = 1 (Imbrie, 1956, p. 228). These values can be obtained by plotting data on double logarithmic graph paper and solving the equation:

$Y = aX + B$, where

$Y = \log y \quad X = \log x \quad B = \log b.$

Isometric growth. In many cases, bivariate plots of paleontological data do not show any hint of a logarithmic relationship between the x and y variables. In some cases this indicates that a real linear trend exists (*i.e.*, isometric growth). In many cases, however, the apparent linear trend results because (1) the curvature of the true growth line is very slight and becomes disguised because of the scatter of points around it (Figure 10 - curve B); and/or (2) the sample being studied is somewhat restricted and only plots around the linear portion of the true curved growth line (Figure 10 - curve C). Irrespective of the underlying cause, this growth can be treated as isometric growth (= allometric growth with **a** = 1) and much simpler mathematical treatments of data can be used. Essentially, the growth line will have the form $y = ax + b$, where **a** = gradient of line and **b** = y-intercept of line (*i.e.*, value of y when x = 0).

Determination of growth lines for isometric growth. If the two variables have a correlation coefficient of 1.0, it means that there is no scatter of points around the linear growth line. In such cases, the equation of the growth line can be easily and accurately given by an equation of the form $y = ax + b$. Unfortunately, most paleontological data result in a scatter of points on a bivariate graph and methods for deriving the best estimate of the growth line must be examined. Essentially, there are three ways of estimating the equations of growth lines, namely: (1) Regression Equations, (2) Major Axis, and (3) Reduced Major Axis.

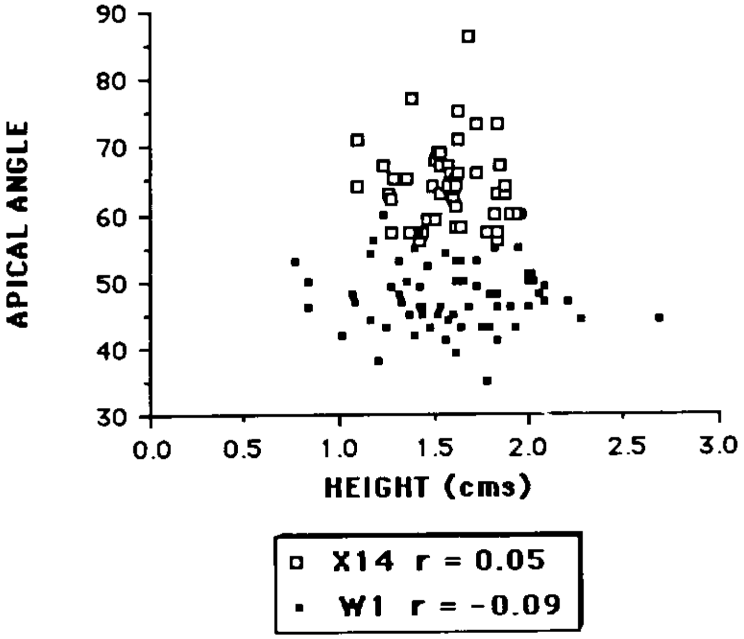


Figure 9 (upper) *Example of bivariate graph with very low correlation between the height and apical angle of* Atrypoidea foxi *f. A (assemblage X14) and* Atrypoidea phoca *(assemblage W1).*

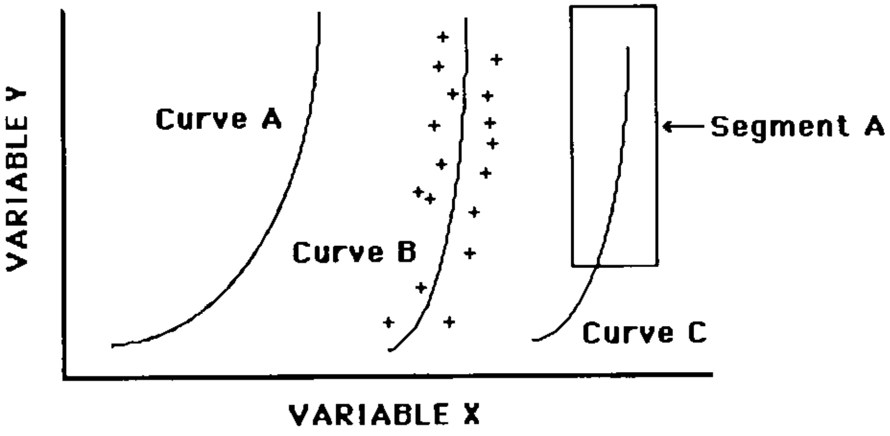


Figure 10 (middle) *Examples of allometric growth curves. Curve A is well defined and thus easy to recognize as allometric growth. Curve B is less pronounced and if coupled with a wide scatter of plotted points, as shown, would be difficult to recognize as allometric growth. If the plotted points only occurred in segment A of curve C a more or less linear relationship between variables x and y would be suggested.*

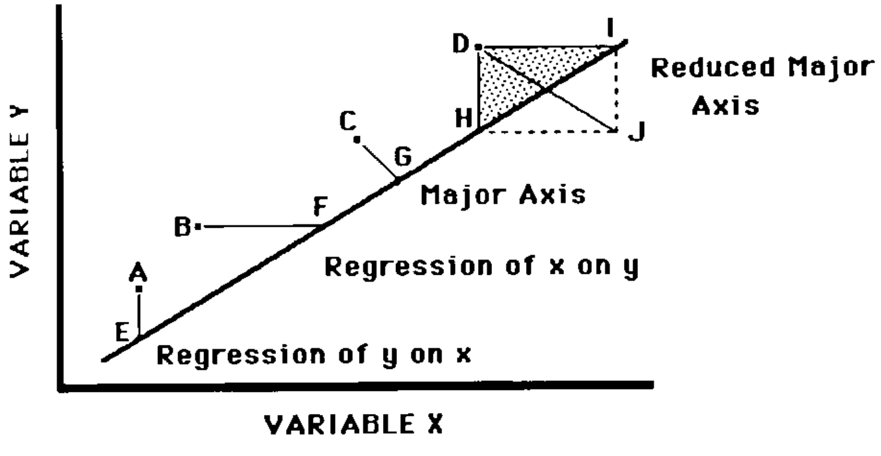


Figure 11 (lower) *Diagram showing the different distances used in the computation of regression equations (x on y and y on x), a Major Axis, and a Reduced Major Axis. (Based on Imbrie, 1956, fig. 2). See text for meaning of letters A to J.*

Regression equations. This procedure actually involves the calculation of two equations, one for the *regression of y on x* and one for the *regression of x on y*. The regression of y on x is defined as the line that minimizes the sum of the squares of the deviations from that line (line AE on Figure 11), the deviations being measured perpendicular to the x-axis (Imbrie, 1956, p. 230). The regression of x on y is similar except that the deviation is measured in a direction perpendicular to the y-axis (line BF on Figure 11). In both cases, there is the underlying assumption that one variable is independent and the other variable dependent. From a paleontological point of view this means that the procedure cannot be used since biological data always involve variability (both real and observed) of both the x and y variates (Kermack and Haldane, 1950; Kermack, 1954; Imbrie, 1956).

Major axis. "A line that minimizes the sum of the squares of the perpendicular distances from each point to the desired line is called the major axis" (Imbrie, 1956, p. 230; line CG on Figure 11). Although seemingly a reasonable way of treating variable paleontological data, it has been shown that the slope of the major axis changes with the unit of measurement (Kermack and Haldane, 1950). It is thus unsuitable for use in taxonomy (Imbrie, 1956, p. 230).

Reduced major axis. "This line minimizes the sum of the areas of the triangles formed by lines drawn from each point to the desired line and parallel with the **x** and **y** axes" (Imbrie, 1956, p. 230; HDI on Figure 11). This line is particularly useful since (1) it makes no assumptions of independence, (2) it is invariant under change of scale, (3) it is simple to compute, and (4) results obtained from its use are intuitively more reasonable than the corresponding results obtained from regression analysis (Imbrie, 1956, p. 231). Furthermore, the use of a reduced major axis allows the computation of further statistics for the comparison of two or more such axes.

Table 3 **Example of calculated or z statistics for the comparisons of two growth axes for the width (x-axis) *versus* length (y-axis) from assemblages BJ8D *(Atrypoidea phoca)* and BJ33 *(Atrypoidea foxi)*. Both assemblages from the Douro Formation, Goodsir Creek, Cornwallis Island, NWT.**

Assemblage	Equation	Error of slope	Dispersion	
			Absolute	Relative
BJ8D	$y = 1.01x + 0.13$	0.05	0.11	6.01
BJ33	$y = 1.23x - 0.13$	0.07	0.09	4.63

For comparison of gradients with Null Hypothesis that both equations have the same slope:

$$z = \frac{(a_1 - a_2)}{\sqrt{s_{a_1}^2 + s_{a_2}^2}} = \frac{(1.23 - 1.01)}{\sqrt{0.07^2 + 0.05^2}} = \frac{0.22}{\sqrt{0.0074}} = 2.56$$

Since $z > 1.96$, $P < 0.05$ and Null Hypothesis of equal slopes is rejected.

For comparison of constants with Null Hypothesis that both equations have same constant (*i.e.*, initial growth indices):

$$z = \frac{x_o(a_1 - a_2) + (b_1 - b_2)}{\sqrt{s_{a_1}^2(x_o - \bar{x}_1) + s_{a_1}^2(x_o - \bar{x}_2)}}$$

Given that $\bar{x}_1 = 1.40$ and $\bar{x}_2 = 1.28$,
Let $x_o = \bar{x}_1$ then,

$$z = \frac{1.40\,(1.40 - 1.28) + (-0.13 - 0.13)}{\sqrt{0.05\,(1.40 - 1.01)}} = \frac{0.17 - 0.26}{\sqrt{0.02}} = \frac{0.09}{0.14} = 0.64$$

Since $z < 1.96$, $P > 0.05$ and Null Hypothesis of equal y-intercepts cannot be rejected.

Calculation of Reduced Major Axis and associated statistics. For a straight line with the equation of the form $y = ax + b$,

$$a = \frac{s_y}{s_x} \qquad s_a = a\sqrt{\frac{1 - r^2}{n}}$$

$$b = \bar{y} - \bar{x}a \quad , \qquad \text{where}$$

a = growth ratio
$\mathbf{s_x}$ = standard deviation of *x*
$\mathbf{s_y}$ = standard deviation of *y*
$\mathbf{s_a}$ = standard error of *a*
r = correlation coefficient
n = sample size
b = initial growth index.

Comparison of Reduced Major Axes. The simplest way of comparing two or more growth lines (= Reduced Major Axis) is to plot them on a graph and visually inspect them. In many cases, the axes involved may be so different that further statistical tests are not needed. With some sets of growth axes it may be difficult to decide confidently if the growth lines are the same or different. In this situation a number of very useful statistical tests can be applied to the data (Table 3). In effect, the comparison of two reduced major axes can be treated as (1) a comparison of the gradients of the two lines, and (2) the position of the y-intercepts of the two lines (*i.e.*, the initial growth indices).

Comparison of gradients of growth lines. This test, conducted under the hypothesis that the two growth lines being considered have the same slope, is done by calculating the statistic **z**, where:

$$z = \frac{a_1 - a_2}{\sqrt{s_{a_1}^2 + s_{a_2}^2}} \quad .$$

If the value of **z** < 1.96, the probability *(P)* that the observed difference $(a_1 - a_2)$ arose by chance is greater than 0.05 (Imbrie, 1956, p. 237). Conversely, if $z > 1.96$, then the probability that so great a difference arose by chance is 0.05 or less (Imbrie, 1956, p. 237). If a **z** value of 2.58 is used, then the probability can be expressed at the 0.01 level. If $P < 0.05$ $(z > 1.96)$, then the hypothesis of equal slopes can be rejected and the observed differences are considered significant (Imbrie, 1956, p. 237). If this is true then no further statistical tests are needed.

Comparison of initial growth indices. This comparison is only used in situations where the gradients of the growth line are not considered significantly different. The **z** statistic is calculated by:

$$z = \frac{x_o(a_1 - a_2) + (b_1 - b_2)}{\sqrt{s_{a_1}^2(x_o - \bar{x}_1) + s_{a_1}^2(x_o - \bar{x}_2)}} \quad , \quad \text{where}$$

$\mathbf{x_o}$ = some biologically significant value.

Once the **z** value has been determined, it is treated in the same manner as the **z** value was in the comparison of the gradients of the growth lines.

One of the main problems with this test is the choice of x_0 since it is difficult to define a "biologically significant value" without bias. A simple solution is to make x_0 equal to $\bar{x}_1$ since this reduces the above formulae to:

$$z = \frac{x_1 (a_1 - a_2) + (b_1 - b_2)}{s_{a_2} (x_1 - \bar{x}_2)}$$

Other possible values for x_0 include the **x** value for which the vertical distance between the growth lines is (1) at a minimum — if the wish is to demonstrate that a significant difference exists between the initial growth indices, or (2) at a maximum — if the wish is to demonstrate that there is not a significant difference between the growth lines (Imbrie, 1956, p. 237). If a means of choosing x_0 without bias can be derived, then the test can be used with some success.

Other methods for comparing Reduced Major Axes. If only two or three growth axes are being compared, then the method of comparing pairs of axes in terms of their gradient and initial growth indices can be used without difficulty. If, however, the task is to compare many growth axes (*e.g.*, 10 or more), then various practical problems are created by using these test procedures (Jones, 1983). For example, a situation in which 10 growth axes are all being compared, one with another, would require (n(n-1))/2 or 45 comparisons. This is compounded by the fact that each comparison of two reduced major axes usually involves two separate tests. Thus, a total of 90 **z** values may be obtained from an investigation of 10 growth lines. As noted by Jones (1983), the problem then becomes one of deciding how to deal with this amount of data, and, more importantly, how to decipher the information. This problem can be overcome quite simply by plotting a bivariate graph which plots ***a*** (*i.e.*, the growth ratio) against ***b*** (*i.e.*, the initial growth index). Inspection of such a graph can give a very clear visual impression of the relationships between the various growth lines being considered (Figure 12; Table 4). If needed, additional statistical procedures can be applied to this type of bivariate graph (see Jones, 1983).

Scatter of points around a growth axis. In some cases, it is desirable to have a measure of the spread or dispersion of the plotted points around the reduced major axis. This can be obtained by calculating either the Absolute Dispersion or the Relative Dispersion about the Reduced Major Axis (Table 3). Both of these calculations hinge on the fact that each point deviates from the axis by the horizontal distance d_x and the vertical distance d_y (Imbrie, 1956, p. 239). A measure of the dispersion of the points can therefore be derived by calculating the vector sum of d_x and d_y. The total dispersion, which can be expressed as the standard deviation (s_d) of those diagonal distances (Imbrie, 1956, p. 239), is given by:

$$s_d = \sqrt{2(1 - r)(s_x^2 + s_y^2)}$$

Imbrie (1956, p. 241) has pointed out, however, that the absolute dispersion is a measure of shape variability since deviation in either **x** or **y** will affect s_d. To overcome this problem relative dispersion is used by calculating:

$$z = \frac{100\, s_d}{\sqrt{\bar{x}^2 + \bar{y}^2}}$$

Imbrie (1956, p. 241), who termed this the Coefficient of Relative Dispersion about the reduced major axis, noted that it expresses the amount of shape variation as a proportion of the average shape attained by the sample.

A visual measure of the dispersion of the plotted points on a bivariate graph can be obtained by contouring the graph according to the density of points per unit area on the graph (Figure 13). This technique allows a clear visual impression of the spread of the points (*e.g.*, Trewin, 1976) as well as the identification of modal groups (Figure 13).

Caution in use of bivariate graphs. If two assemblages of brachiopods are being compared through the use of bivariate graphs, then some care must be exercised in the manner in which that comparison is made. In many cases, it is tempting to make such a comparison on the basis of a single graph, for example a length *versus* width graph. Such a graph may suggest that there is no difference between the two assemblages of brachiopods being considered. It is important to emphasize, however, that this only means that the two assemblages are alike in terms of this aspect of their growth. It may well be that they are significantly different in other aspects of their growth. Thus, it is important in this type of study to investigate as many bivariate graphs as possible.

An insight to this problem is provided by the comparison of assemblage X14 (*Atrypoidea foxi* f. A) and W1 (*Atrypoidea phoca*) through the use of length *versus* width, length *versus* height and width *versus* apical angle (Figure 8). If the length *versus*

Table 4 **Data used for plotting graph in Figure 12. Data from Jones (1983). Gradient (a) and constant (b) from the reduced major axes relating width (x-axis) to length (y-axis).**

Species	Assemblage	Gradient (a)	Constant (b)
A. foxi f. A	X14	1.32	-0.41
	Y18	1.22	-0.21
	T43	1.41	-0.51
A. foxi f. B	FA60A	1.20	-0.15
	X2	1.25	-0.26
	T22A	1.23	-0.08
	BJ33	1.13	-0.13
	M113	1.10	-0.01
	GC5	1.09	-0.02
A. phoca	Z18	0.96	0.10
	W1	0.99	0.14
	T31	1.07	-0.01
	FA22	1.01	0.05

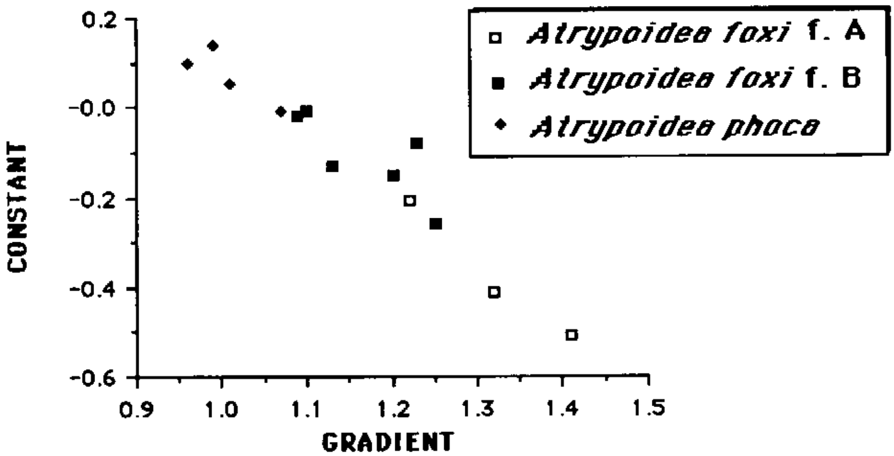


Figure 12 *Example of bivariate graph in which the gradient (**a**) is plotted against the constant (**b**) of reduced major axes calculated for the width (x-axis)* versus *length (y-axis) of various assemblages of* Atrypoidea foxi *f. A,* Atrypoidea foxi *f. B and* Atrypoidea phoca*. Original data given in Table 4.*

width (Figure 8A) and the length *versus* height (Figure 8B) graphs were considered alone, it would be difficult to argue that two different species were involved in the comparison. Conversely, a plot of width *versus* apical angle (Figure 8C) suggests that the two assemblages are quite different in this aspect of their growth.

Examination of the same data using three-dimensional graphs also illustrates the problems that can be encountered if attention were focussed on a single bivariate graph (*e.g.*, Figure 14). If such a graph is viewed in a direction parallel to the Z axis (height), then the view is in effect a bivariate graph of length *versus* apical angle (Figure 14A) and there is little apparent difference between the two assemblages of brachiopods being considered (Figure 14A). By rotating the graph in three-dimensional space (using MacSpin® program), it is possible to obtain an oblique view of the graph and the clusters of points for the two assemblages (Figure 14B). An exploded view of such a plot clearly shows the difference between the two assemblages of brachiopods (Figure 14C).

Studies with bivariate graphs.

Methodology: Imbrie (1956), Kermack (1954), Kermack and Haldane (1950), Jones (1983); *Contoured bivariate graphs:* Jones (1978), Trewin (1976); *Brachiopods:* Alvarez (1982, 1984), Amsden (1974, 1978), Amsden and Boucot (1958); Cooper (1972), Copper (1986), Hewitt and Hurst (1977), Imbrie (1956), Jones (1974, 1977, 1978, 1979b, 1981, 1983), Jones and Rong (1982), Mergl (1985), McGugan and May (1965), Mills and Langenheim (1987), Peck *et al.* (1987), Pocock (1966), Watkins (1975); *Ostracodes:* Berdan (1972, 1986); *Crinoids:* Pabian and Strimple (1979); *Corals:* Dixon (1979, 1986), Elias (1985), Hodges and Roth (1986); *Conodonts:* McHargue (1982), Murphy and Cebecioglu (1986); *Eurypterids:* Kues and Kietzke (1981); *Sponges:* Laghi *et al.* (1984); *Blastoids:* Waters *et al.* (1985).

Comparison of Faunas from Different Localities

A common result of paleontological research are data pertaining to the number of specimens of each species at a particular locality. This invariably leads to the question of how the fauna at one locality compares to the fauna at another locality. Although there are a number of different ways in which this question can be approached from a statistical point of view, only two will be considered here, namely (1) simple graphical methods, and (2) cluster analysis.

Graphical comparison of faunal data. One possible way of comparing the composition of faunas from different localities is by using various graphical presentations of the data. This can be in the form of graphs that show the number of specimens of a given species for all the localities being considered (*e.g.*, Jones and Smith, 1985, fig. 9). This method is quite useful in certain circumstances and can be used to determine trends in the distribution of species in a given area (*e.g.*, Jones and Smith, 1985, fig. 9). Alternatively, the data can be presented in the form of a histogram with the different components of the fauna illustrated on the individual columns (Figure 15). In many cases, such plots give a very clear picture of any major trends that may exist in the data. While these methods give a good visual impression of the data, they do not allow any statistical evaluation of those trends and they must therefore be described in qualitative terms.

Cluster analysis. Cluster analysis is a powerful statistical procedure that is extremely useful in the evaluation of paleontological data. In the particular case being considered here, the data comprise a number of localities for which the number of specimens of each species is known. Before cluster analysis can be performed, however, a number of important decisions must be made, namely (1) choice of data type, (2) choice of similarity index, and (3) choice of clustering method.

Choice of data. With most paleontological information there is a choice between binary data and variable data (= numerical data). This is especially true if the data had been assembled with the knowledge that statistical procedures were to be used in their manipulation.

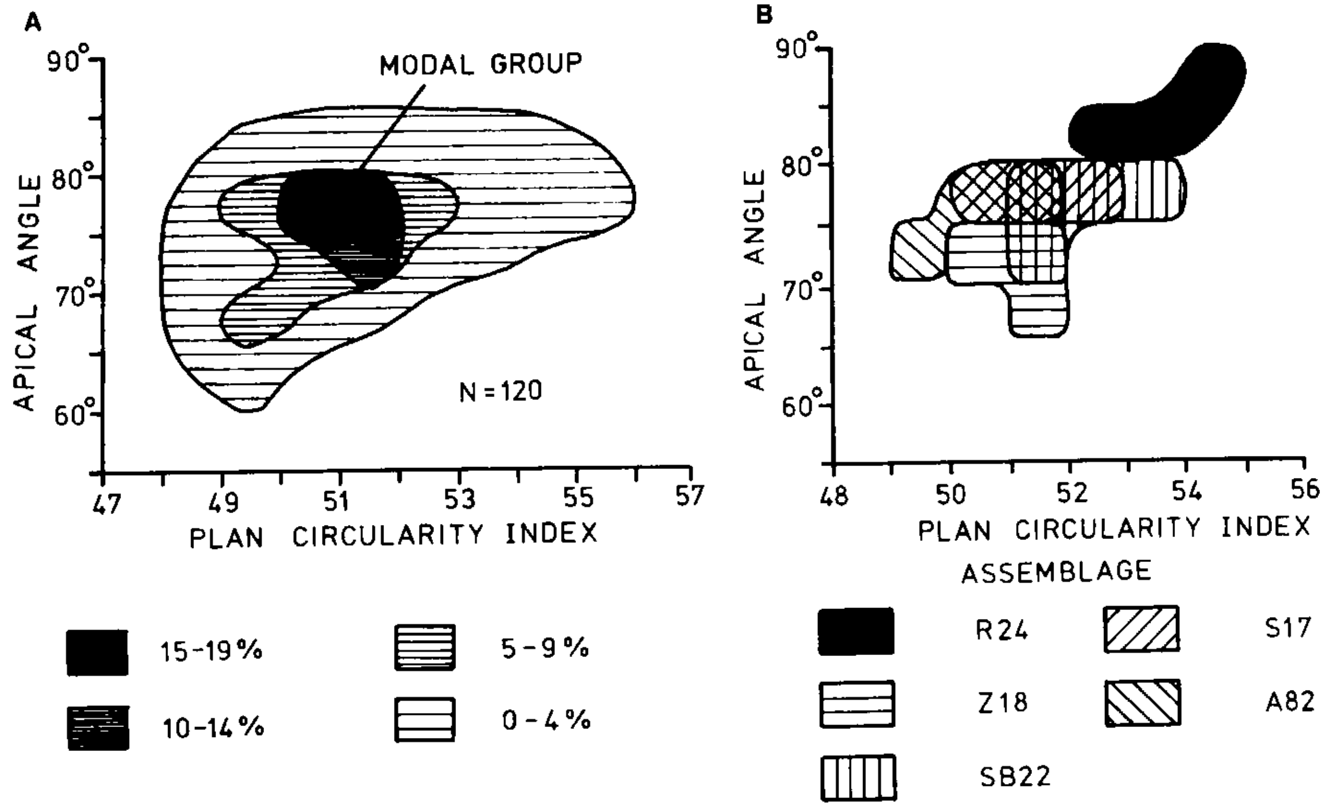


Figure 13 (A) *Contoured bivariate graph of apical angle* versus *Plan Circularity Index for* Protathyris praecursor *which allows the identification of a modal group.* (B) *Comparison of modal groups for different assemblages of* Protathyris praecursor. *(From Jones, 1978, fig. 9).*

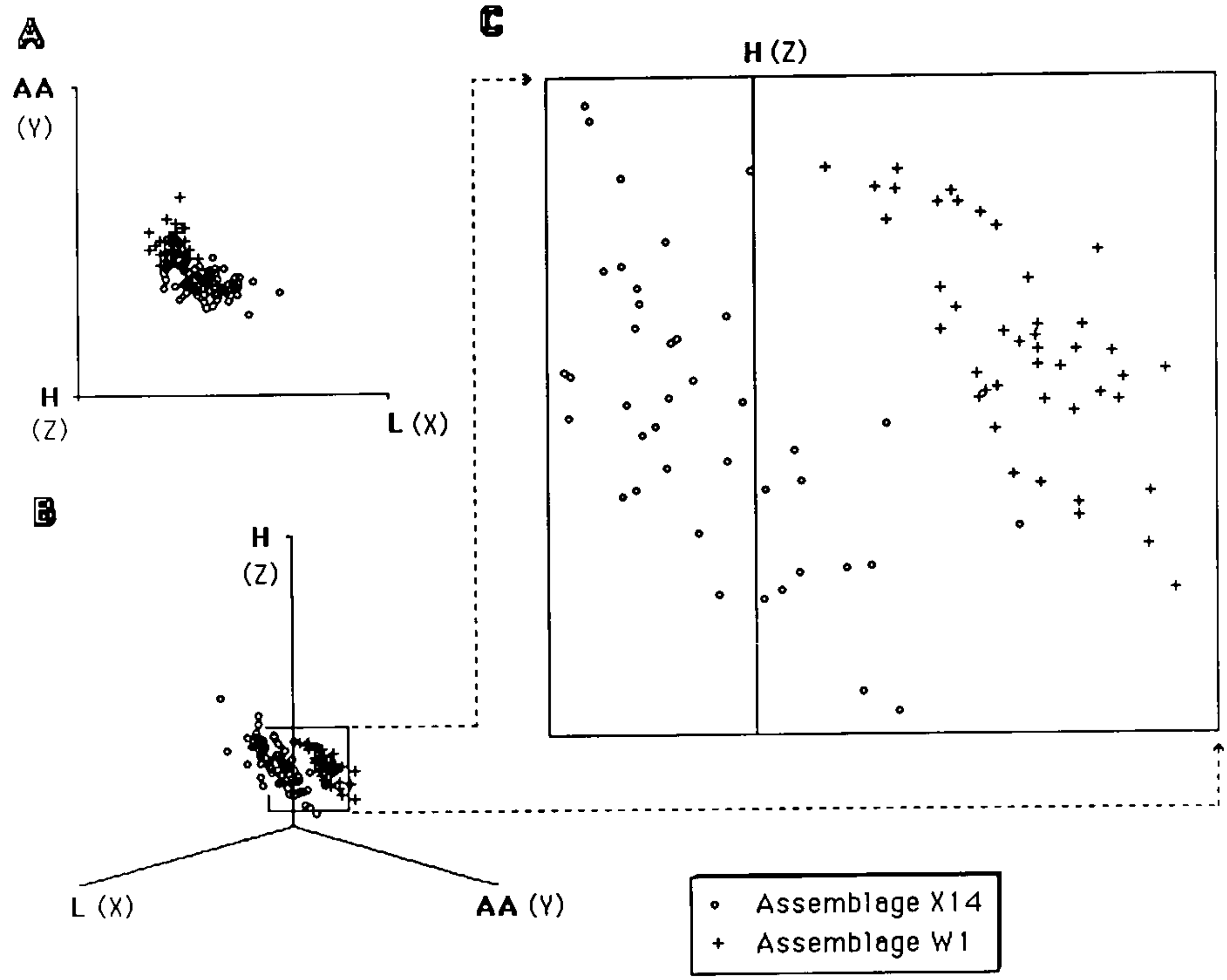


Figure 14 (upper)

- **(A)** *Three-dimensional plot of length **(L)**, apical angle **(AA)** and height **(H)** oriented so that the z-axis is coming straight out of the paper. This gives the effect of a bivariate graph involving only the length and the apical angle.*
- **(B)** *Same graph as in Figure 14A except that it has been rotated around the origin in order to obtain a better view of the plotted points.*
- **(C)** *Exploded view of plot shown in Figure 14B in order to illustrate the difference between assemblages X14 and W1. The scale has been omitted from the axes in order to prevent crowding and confusion on the diagrams. Plots derived using MacSpin™ program on a Macintosh Plus® microcomputer.*

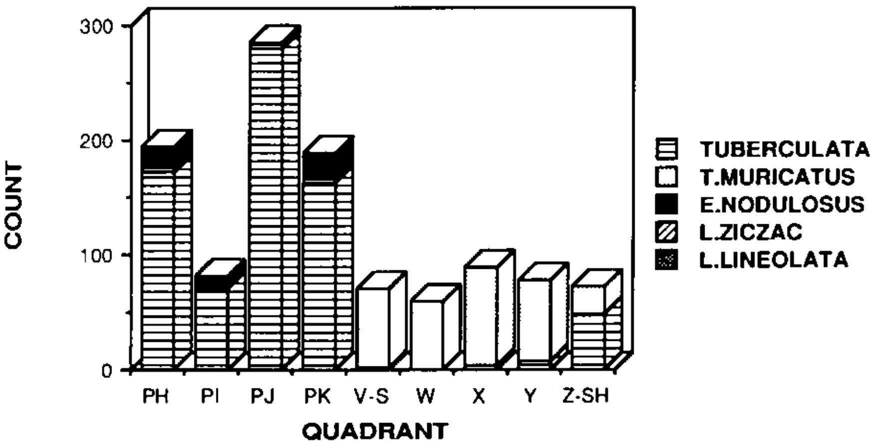


Figure 15 (lower right) *Example of stacked histograms showing variation and hence the trends in faunas from different localities. This particular example shows the variation and trends in gastropod faunas that occur on a rocky shoreline on the northwest coast of Grand Cayman Island. Samples are from 1 m^2 quadrants measured successively from the shoreline (sample **PH**) inland. Based on data collected by B. Jones and S.G. Pemberton.*

Table 5 Listing of similarity indices showing the type (*dis* = distance; *sim* = similarity), the type of data it can be used with (*num* = numerical or variable; *bin* = binary) and the name. (Modified from CLUSTAN: Wishart, 1978, p. 112).

Type	Data	Name or originator or reference
dis	num	Squared euclidean distance
dis	bin	Squared euclidean distance
sim	num	Product-moment correlation
sim	bin	Simple matching coefficient
sim	bin	Similarity ration (Jaccard)
sim	bin	Czekanowski-Dice
sim	bin	Sokal-Sneath
sim	bin	Sokal-Sneath
sim	bin	Rogers-Tanimoto
sim	bin	Kulczynski
sim	bin	Sokal-Sneath
sim	bin	Hamann
sim	bin	Dot product (Russell-Rao)
sim	bin	Kulczynski
sim	bin	Sokal-Sneath
sim	bin	Cosine (Ochiai)
sim	bin	Sokal-Sneath
sim	bin	Product-moment correlation
sim	bin	Yule
dis	bin	Size difference
dis	bin	Pattern difference
dis	num	Average distance
dis	bin	Average distance
dis	num	Error sum of squares
dis	bin	Error sum of squares
sim	num	Dot product
sim	num	Cosine
sim	num	Similarity ratio
dis	num	Size difference
dis	num	Shape difference
dis	bin	Shape difference
sim	num	Dispersion
sim	bin	Dispersion
dis	num	Variance
dis	bin	Variance
dis	num	Nonmetric (Bray-Curtis)
dis	bin	Nonmetric (Bray-Curtis)
sim	num	USER coefficient
sim	bin	USER coefficient
dis	bin	Information statistic

Binary data is based on the very simple premise that a species is either present or absent at a particular locality. In many respects, such data are very useful because no emphasis is placed on the abundance of a particular species. Thus, a species is regarded as being present irrespective of whether it is represented by one specimen or by 1,000 specimens. It is this aspect of the data that can cause potential problems. In the case of brachiopod faunas, one or two species commonly dominate while the remaining species are present in minor numbers. The dominant species change as environmental or stratigraphic positions change. Thus, a species that is dominant at one locality may only be a minor element of the fauna at another locality. This aspect of dominance, which is of extreme importance in the comparison of brachiopod faunas, cannot be accommodated by the use of binary data.

Given that binary data cannot accommodate the aspect of dominance, it would seem logical to suspect that variable data — *i.e.*, data which record the actual number of each species collected — could be used to overcome the problem. From both theoretical and statistical points of view, this is the case. Unfortunately, there are practical problems which can cause some serious difficulties with such data; these focus on the problem of sampling, since there can be no assurance that the collection at hand is truly representative of the actual fauna. Biases in the data can be introduced by collecting problems (*e.g.*, bias toward collecting the larger shells at the expense of smaller shells), preservation problems (*e.g.*, larger shells badly crushed), or simply the time available for collecting at each locality. These factors, as well as many others, all have the potential to affect the number of specimens actually collected. This, in turn, will affect the statistical results.The problems with the variable data can be minimized to a certain extent by careful collecting techniques that reduce the known sources of error. Even with such care, problems may exist, since one collection may contain 2,000 specimens while another collection may only contain 200 specimens. Since the computation of the similarity indices may involve actual numbers of specimens, the very fact that one collection is so much bigger than the other will affect the results. The problem of sample size can be overcome to a certain extent by reducing all the data to a percentage basis. In effect, each sample is treated as being of a standard size (*i.e.*, 100 shells). Where large samples are involved this procedure seems to be quite effective.

Choice of similarity coefficients. The choice of a similarity index is not as straightforward as might be expected since there are numerous indices that can be used. For example, the computer package CLUSTAN (Wishart, 1978) lists 40 different similarity/dissimilarity coefficients (Table 5) that can be used for binary data (27) and variable data (13). With binary data, most coefficients revolve around the assessment of the frequency counts A, B, C, and D where, with reference to samples I and K:

A = number of species common to both I and K; **B** = number of species present in I and absent in K; **C** = number of species absent in I and present in K; and, **D** = number of species absent from both I and K.

Once these frequency counts have been established, a similarity or dissimilarity coefficient can be computed. Of the 27 possible coefficients listed in CLUSTAN (Wishart, 1978), some are more useful than others. This can be demonstrated by considering the Jaccard Coefficient, the Rogers and Tanimoto Coefficient, and the Simple Matching Coefficient.

Jaccard coefficient – a similarity coefficient that ranges from 0 to 1 which is calculated by:

$$\frac{A}{A + B + C}.$$

Thus, emphasis is placed on those species which are common to both samples being considered. At the same time, those species which are absent from both samples (D) do not enter into the computation.

Rogers and Tanimoto coefficient – a similarity coefficient that ranges from 0 to 1 which is calculated by:

$$\frac{(A + D)}{(A + D) + 2(B + C)}.$$

This coefficient places emphasis on the species that are present in both samples (A) as well as the species that are absent from both samples (D). The mismatches (B and C) are given double weight. This coefficient tends to give poor results, because (1) emphasis is placed on joint absences (which can be increased simply by adding more and more species that are known to be absent from both samples), and (2) in the denominator double weight is given to the mismatches (B and C).

Simple matching coefficient – a similarity coefficient that ranges from 0 to 1 which is calculated by:

$$\frac{A + D}{A + B + C + D}.$$

This is a straightforward computation which takes into account both the joint presences (A) and the joint absences (D) of species relative to the total number of species present (A + B + C + D).

Three coefficients based on binary data have been presented in order to demonstrate the various approaches that have been taken in the computation of similarity and dissimilarity coefficients. It is apparent from these, however, that great care must be taken in the selection of a suitable coefficient (Table 6). While it is possible to evaluate the merits of each coefficient simply by considering the equation by which it is derived, the ultimate test involves the application of all the coefficients to a single data set (Table 6). It is advisable when dealing with any data set for the first time to calculate a number of different coefficients and then thoroughly evaluate the results prior to choosing a particular one.

There are also numerous similarity/dissimilarity coefficients that can be used when dealing with variable data (Table 3). Many of these coefficients involve the calculation of the distance between points plotted in Euclidean space. The concept behind this can best be understood by considering two points plotted on a simple bivariate graph. The distance between such points can be easily calculated by applying the Pythagoras Theorem; thus,

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

where **d** = shortest distance between the two points that are defined by the co-ordinates (x_1, y_1) and (x_2, y_2).

If the data were plotted on a three-dimensional graph, then the Pythagoras Theorem can be extended so that:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2},$$

where **d** = shortest distance between the two points that are defined by the co-ordinates (x_1, y_1, z_1) and (x_2, y_2, z_2).

In the case of paleontological data, there are commonly more than three variables involved and it is sometimes necessary to compute similarity coefficients using all the variables. This can be done through the use of Euclidean space which is an n-dimensional extension of the three-dimensional graph with the axes X, Y and Z. In the case of Euclidean space, the graph is considered to have n axes (where n = number of variables being considered), each axis being at right angles to all other axes. It is obvious that such a graph would be impossible to construct. Nevertheless, the distances between pairs of points can be calculated by extending the Pythagoras Theorem such that:

$$d = \sqrt{\sum_{i=1}^{n} (x_{i1} - x_{i2})^2},$$

where

i = number of variables being considered

$\mathbf{x_{i,1}}$ = first variable of group 1

$\mathbf{x_{i,1}}$ = first variable of group 2.

In many situations, it is more useful to express the similarity scale on a scale of 0.0 to 1.0; thus the formula becomes:

$$d = 1 - \sqrt{\sum_{i=1}^{n} \frac{(x_{i1} - x_{i2})^2}{N}}.$$

The above formulae will only work if x_i is positive and less than 1.0. Thus, x_i must be normalized on a scale of 0 to 1.0. Maximum use of the scaling can be achieved by setting the minimum value of x_i to 0 and the maximum value of x_i to 1.0 by:

$$\text{Normalized data} = \frac{(\text{True value - Minimum value})}{(\text{Maximum value - Minimum value})}.$$

If five samples are being compared using this technique, then a 5 × 5 similarity matrix will be obtained in which all possible pairs of samples have been compared. Cluster analysis then proceeds as follows: (1) matrix is searched for the highest similarity index; (2) the data for the two samples having the highest similarity index is combined and averaged using a weighting factor of 1 for each sample (*i.e.*, if the new data comprise two samples, the weighting factor is 2); (3) one of the two samples is deleted from further consideration while the other one is retained; (4) a new similarity matrix is calculated using the averaged data obtained in step 2 and ignoring the sample deleted in step 3; (5) the procedure starting at step 1 is repeated.

The calculation of similarity indices and the following cluster analysis is relatively straightforward. It is, however, a long process which involves the repeated computation of similarity matrices. For this reason, it is best to either write the necessary computer program or make use of some of the readily available computer packages. CLUSTAN (Wishart, 1978) is excellent in this respect since it allows the use of many different similarity indices (Table 5) and clustering techniques (Table 7).

Choice of clustering procedure. Once the similarity matrix has been derived, it is necessary to analyse it further in order to determine the clusters that are present. One of the more straightforward techniques has been described in order to demonstrate the general principles involved in the derivation of clusters. There are, however, numerous other statistical procedures that can be used to derive such clusters (Table 7) and it is thus necessary to decide which technique should be used for the particular set of data being considered. Unfortunately, there are no rules governing the choice of such procedures and it essentially becomes a question of personal choice. In most cases, it is advisable to run the same similarity matrix through a number of different clustering procedures in order that an assessment of the relative merits of each can be obtained. Any procedure producing a dendrogram that exhibits chaining (*e.g.*, Figure 16B) should not be used. This results from successive samples being successively combined with the previously defined cluster and should be discarded because it does not provide a good means of subdividing the data being considered. Apart from this, the choice of method is open to the investigator. Of the procedures listed in Table 7, Ward's method (error sum of squares) seems to give the best results.

Q-mode and R-mode Cluster Analysis. Data pertaining to the distribution of species are typically arranged in a table with species on the horizontal axis and localities on the vertical axis. In terms of cluster analysis, such data can be investigated by performing both (1) Q-mode, and (2) R-mode cluster analyses.

Table 6 **Computation and comparison of the Jaccard Coefficient (1), Rogers and Tanimoto's Coefficient (2) and the Simple Matching Coefficient (3) using the artificial samples I and K with A = 7; B = 3; C = 4; D = 5.**

$$(1) = \frac{A}{A+B+C} = \frac{7}{7+3+4} = \frac{7}{14} = 0.50$$

$$(2) = \frac{(A+D)}{(A+D)+2(B+C)} = \frac{7+5}{(7+5)+2(3+4)} = \frac{12}{26} = 0.46$$

$$(3) = \frac{A+D}{A+B+C+D} = \frac{7+5}{7+5+3+4} = \frac{12}{19} = 0.63$$

Table 7 **Listing of methods that can be used to derive clustering from similarity matrices. (Modified from CLUSTAN: Wishart, 1978, p. 32).**

Method	Usual coefficients
Nearest neighbour (single linkage)	all
Furthest neighbour (complete linkage)	all
Group average (average linkage)	all
Centroid	distance size difference shape difference
Median (Gower's method)	distance
Ward's method (error sum of squares)	distance
Lance-Williams flexible BETA	distance
McQuitty's similarity analysis	all

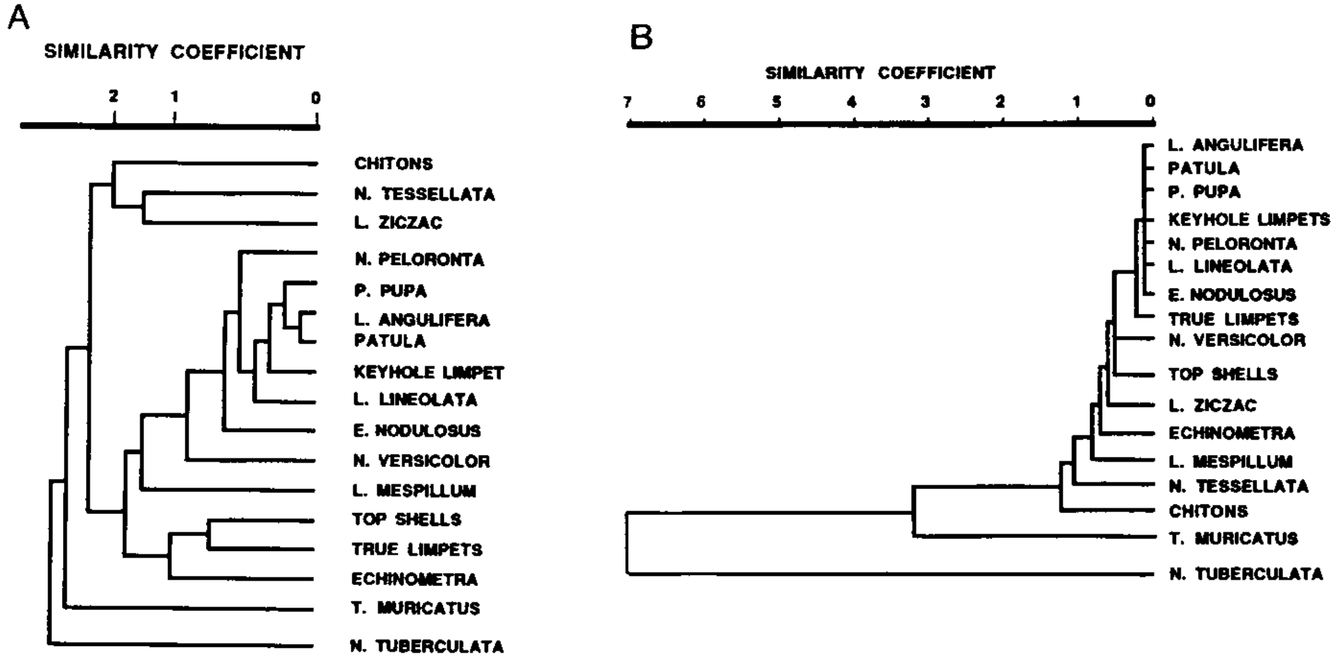


Figure 16 *R-mode cluster analysis based on variable data using 39 samples of rocky shoreline faunas from the coasts around Grand Cayman Island. All samples based on collections from 1 m² quadrants made by B. Jones and S.G. Pemberton.*
(A) *Similarity index based on Squared Euclidean Distance (Table 5) and clustering by Ward's Method (Table 7).*
(B) *Similarity index based on Squared Euclidean Distance and clustering by the Nearest Neighbour (single linkage) method (Table 7).*

Q-mode analysis investigates the relationship between localities; in other words, which localities are most similar in terms of the species present at each locality. Conversely, R-mode analysis investigates the relationship between the species; in other words, which species typically occur together.

With any data set both Q-mode and R-mode analyses should be used since both will yield valuable information.

Presentation and interpretation of cluster analysis. The results of cluster analysis are typically displayed in the form of dendrograms (tree diagrams) since these provide the best visual impression of the relationships present (*e.g.*, Figures 16 and 17). Such diagrams usually permit the establishment of distinct "clusters" of samples (for Q-mode analysis) or species (for R-mode analysis) by simple visual inspection. A more rigorous division of the dendrograms can be achieved by selecting a certain similarity index and extending it as a line across the dendrogram (this is termed a phenon line). The phenon line serves to segment the dendrogram into distinct clusters (Sokal, 1966; Jardine and Sibson, 1971). There are no set rules governing the position of the phenon line although it should obviously be selected with some care. Once such clusters have been delineated

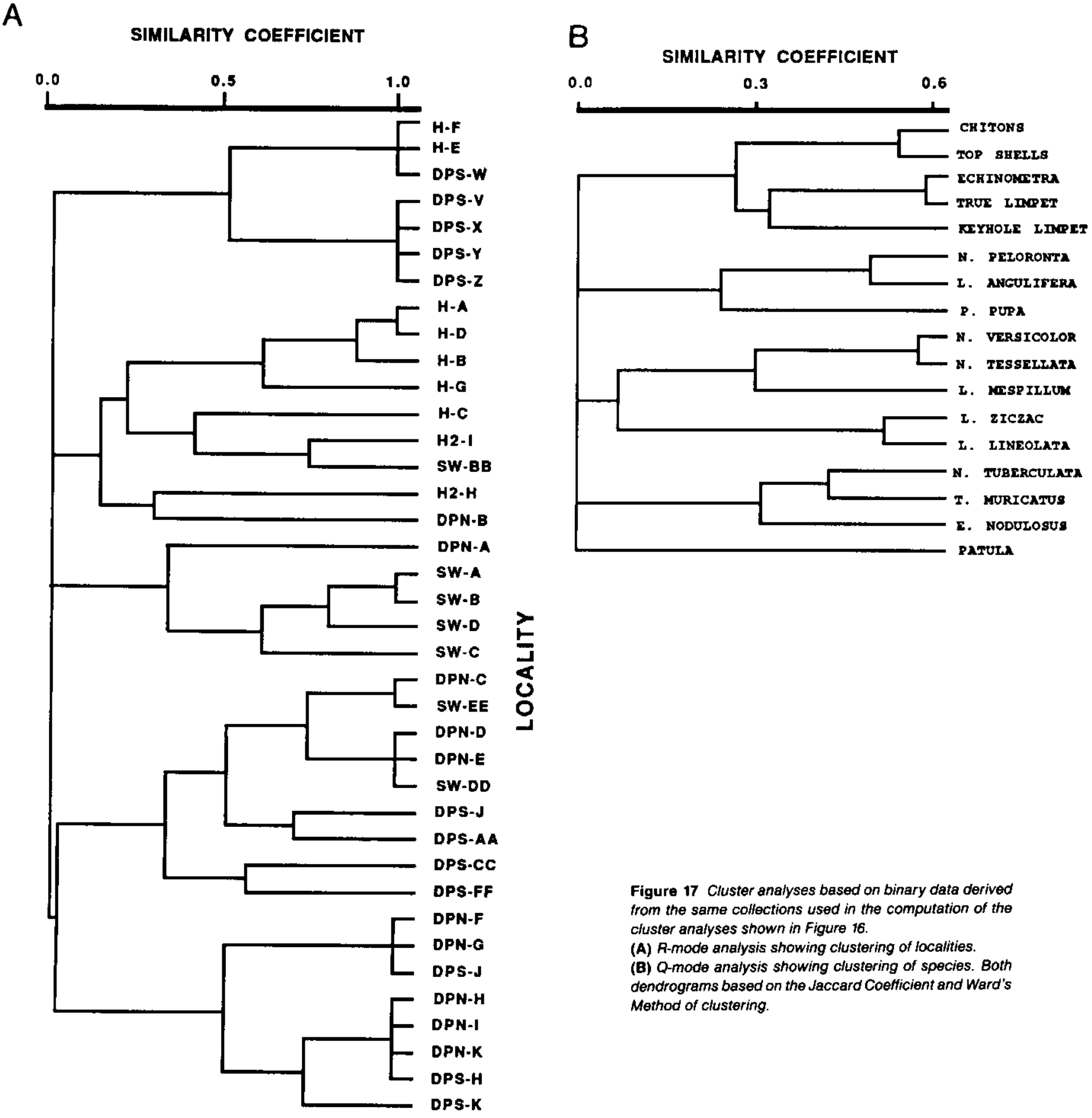


Figure 17 *Cluster analyses based on binary data derived from the same collections used in the computation of the cluster analyses shown in Figure 16.*

(A) *R-mode analysis showing clustering of localities.*

(B) *Q-mode analysis showing clustering of species. Both dendrograms based on the Jaccard Coefficient and Ward's Method of clustering.*

they can be labelled and then incorporated into any subsequent discussion concerning the distribution of various faunas.

If a large number of samples and/or species is involved in a study, deciphering the patterns obtained from the Q-mode and R-mode analyses can be awkward and cumbersome. This can be overcome by constructing a table in which (1) the species are ordered according to the sequence defined by the R-mode analysis, and (2) the localities are ordered according to the sequence defined by the Q-mode analysis (Figure 18). This creates a grid which is, in effect, a cross-plot of the Q-mode cluster dendrogram against the R-mode cluster dendrogram (Figure 18). This diagram is completed by cycling through each locality and marking or shading the appropriate squares to indicate the species that are present at the particular locality being considered. This basic cross-plot can be modified by:

(1) in the case of binary data, the aspect of dominance by one or more species can be incorporated into the plot by showing the numbers of specimens (expressed as actual number of specimens collected or percentage of the fauna) of each species that were collected; while this gives an accurate assessment of totals, it is difficult to identify visually any patterns that may be present;

(2) ranking the species according to the numbers of individuals collected at each locality. In many cases, the faunas can be broken down in 1 to 4 fairly common species and numerous other relatively rare species. The corresponding squares in the grid created by the cross-plot can then be shaded according to the ranking of the species (*e.g.*, black for the most common species and successively lighter shades of grey for the 2nd, 3rd and 4th ranked species). The rarer elements of the fauna are recorded as being present without any ranking. This is a powerful technique because it permits easy division of the grid by simple visual inspection (Figure 18);

(3) an alternative to the procedure described in (2) is to construct a cumulative bar graph on the right hand side of the cross-plot which shows the actual numbers (or percentage) or each species present at the locality being considered (Figure 19). Again, the rarer elements of the fauna (for example, all those species that constitute less than 10% of the fauna) need not be incorporated in the graph. This method, like method 2, also provides an excellent visual impression of the data;

(4) in addition to either method 2 or 3, a second bar graph or scatter graph can be constructed which shows (a) the number of species present (species diversity) at each locality, and (b) the number of specimens collected at each locality (Figure 18). These plots can be extremely useful for they provide an excellent visual impression of the overall nature of the data base. In many cases, it becomes apparent that there is a strong correlation between the defined clusters, the species diversity and the number of specimens collected (*e.g.*, Figure 18). The problem then becomes one of deciding if the correlation is a real one or one created by the collecting techniques that were employed.

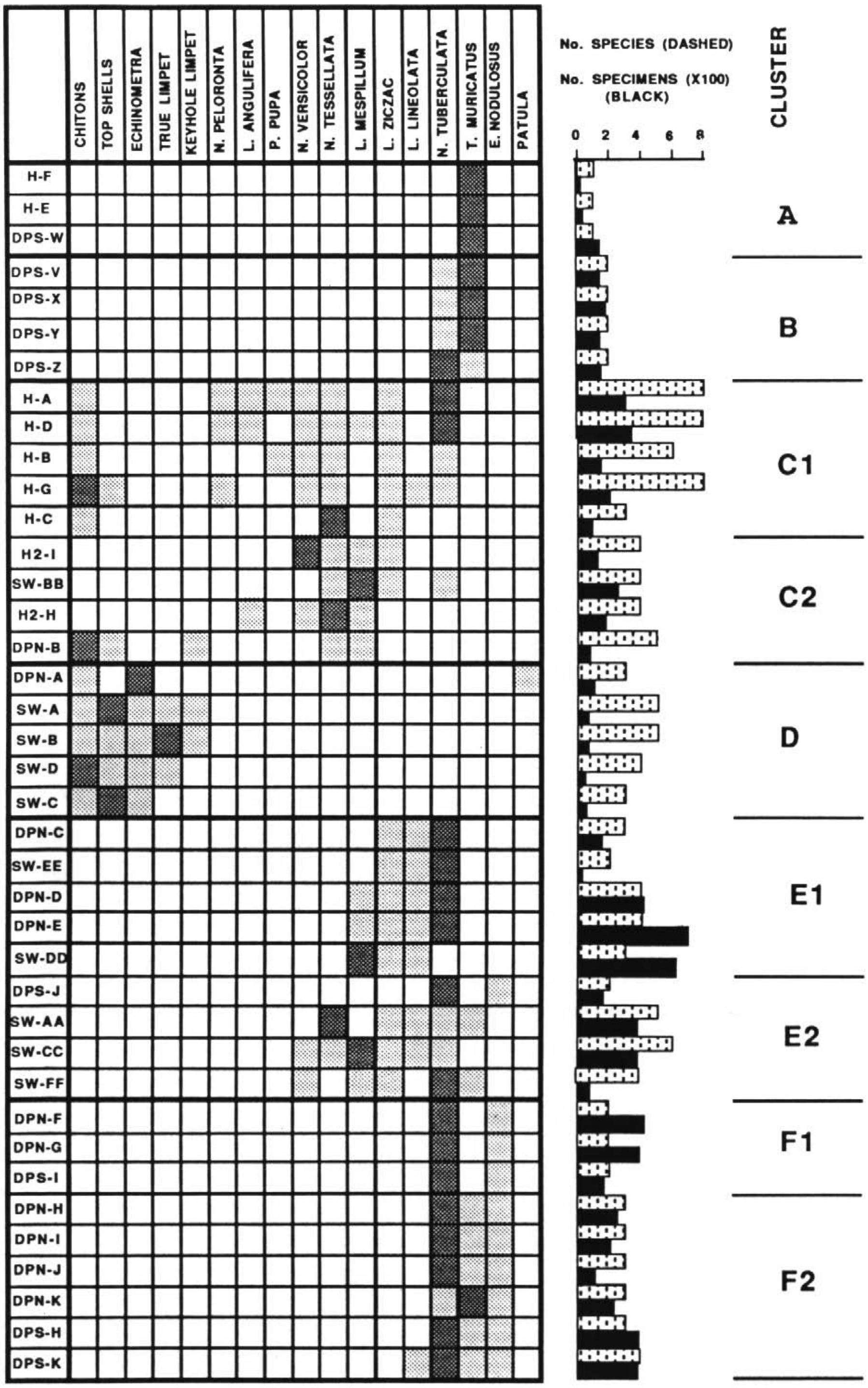


Figure 18 *Cross-plot of localities (ordering based on dendrogram in Figure 17A) versus species (ordering based on dendrogram in Figure 17B). Dark shading indicates that the species is the most common element of the fauna in a sample while the light shading indicates which species are present. The bar graphs on the right side of the diagram indicate the sample size and the species diversity for each locality. This cross-plot is used to define clusters **A** to **F**.*

Cluster analysis is a powerful, but probably underused, tool which can be of considerable use in the analysis of faunal associations. As with most statistical procedures, it is important to have an understanding of the basic mathematical concepts involved since this can have a radical affect on the interpretation of the results obtained.

Studies involving cluster analysis.
Methodology: Harbaugh and Merriam (1968), Harper (1978), Hazel (1970), Jardine and Sibson (1971), Ludvigsen *et al.* (1986), Parks (1966), Sokal (1966); *Total fauna:* Baarli (1987), Beus (1984), Fox (1968), McDonald (1975), McGhee and Sutton (1981), Scott (1974), Sepkoski and Sheehan (1983), Springer and Bambach (1985), Thayer (1974); *Brachiopods:* Jones and Smith (1985), Noble *et al.* (1976); *Gastropods and Bivalves:* Colbath (1985), Valentine and Peddicord (1967); *Bryozoans:* Anstey (1986), Pachut and Anstey (1984); *Crinoids:* Webster (1981); *Echinoids:* McKinney and Zachos (1986); *Ostracodes:* Benson (1975), Izuka and Kaesler (1986); *Traces:* Kitchell and Clark (1979); *Trilobites:* Ludvigsen and Westrop (1983); *Foraminifera:* Mello and Buzas (1968), Rogers (1976), Hallock and Glenn (1985); *Conodonts:* von Bitter (1972), Nowlan and Barnes (1981).

Concluding Remarks

There are many statistical procedures that can be of considerable use to the paleontologist. Although some of the more common ones have been outlined in detail in this paper, it must be realized that there are numerous other statistical procedures that can be of considerable use to the paleontologist. Virtually any paleontological study could be improved by the addition of statistical information even if it is of the basic univariate type.

The advent of image analysis will have a serious impact on the application of statistics in paleontology. Such systems permit (1) the easy and very accurate acquisition of parameters from fossils, and (2) the easy and very accurate derivation of parameters such as area of shell. Although the first has considerable practical advantages, it is really the second aspect that will substantially advance the use of statistics in paleontology. This is because the image analysis systems will allow the easy computation of parameters that have been very difficult to obtain in the past. With the availability of parameters such as area of shell or the center of gravity of an area being considered, more meaningful statistical analyses should be possible. Wilson (1983) and Klapper and Foster (1986) have already demonstrated the advantage of this type of analysis with shark teeth and conodonts, respectively. Similar studies should become more common in the future.

Acknowledgements

I am indebted to the Natural Sciences and Engineering Research Council of Canada (grant No A6090) which has supported the research on which much of this paper is based. I am also grateful to Dr. O.A. Dixon (University of Ottawa) who encouraged me to pursue this line of research while I was a graduate student under his supervision, and Drs. G.S. Nowlan (Geological Survey of Canada) and R. Hall (University of Calgary) for reviewing an earlier version of this manuscript.

References

Alvarez, F., 1982, Variacion morfologica de *Anathyris alejensis* Comte 1938 en su localidad y estrato Tipo: Trabajos de Geologia, Universidad de Oviedo, v. 12, p 195-204.

Alvarez, F., 1984, Cambio temporal y variabilidad morfologica algunos braquipodos Devonicos de la cordillera Cantabrica (No de Espana) pertenecientes al genero *Anathyris*: Trabajos de Geologia, Universidad de Oviedo, v. 14, p. 121-130.

Amsden, T.W., 1974, Late Ordovician and Early Silurian articulate brachiopods from Oklahoma southwestern Illinois, and eastern Missouri: Oklahoma Geological Survey, Bulletin 119, 154 p.

Amsden, T.W., 1978, Articulate brachiopods of the Quarry Mountain Formation (Silurian), eastern Oklahoma: Oklahoma Geological Survey, Bulletin 125, 75 p.

Amsden, T.W. and Boucot, A.J. 1958, Stratigraphy and paleontology of the Hunton Group in the Arbuckle Mountain Region: Oklahoma Geological Survey, Bulletin 78, 199 p.

Anstey, R.L., 1986, Bryozoan provinces and patterns of generic evolution and extinction in the Late Ordovician of North America: Lethaia, v. 19, p. 33-51.

Baarli, B.G., 1987: Benthic faunal associations in the Lower Silurian Solvik Formation of the Oslo-Asker Districts, Norway: Lethaia, v. 20, p. 75-90.

Benson, R.H., 1975, The origin of the psychrosphere as recorded in changes of deep-sea ostracode assemblages: Lethaia, v. 8, p. 69-83.

Berdan, J.M., 1972, Brachiopoda and Ostracoda of the Cobleskill Limestone (Upper Silurian) of central New York: United States Geological Survey, Professional Paper 730, p. 16-19.

Berdan, J M., 1986, New ostracode genera from the Lower Devonian McMonnigal Limestone of central Nevada: Journal of Paleontology, v. 60, p. 361-378.

Beus, S.S., 1984, Fossil associations in the High Tor Limestone (Lower Carboniferous) of south Wales: Journal of Paleontology, v. 58, p. 651-667.

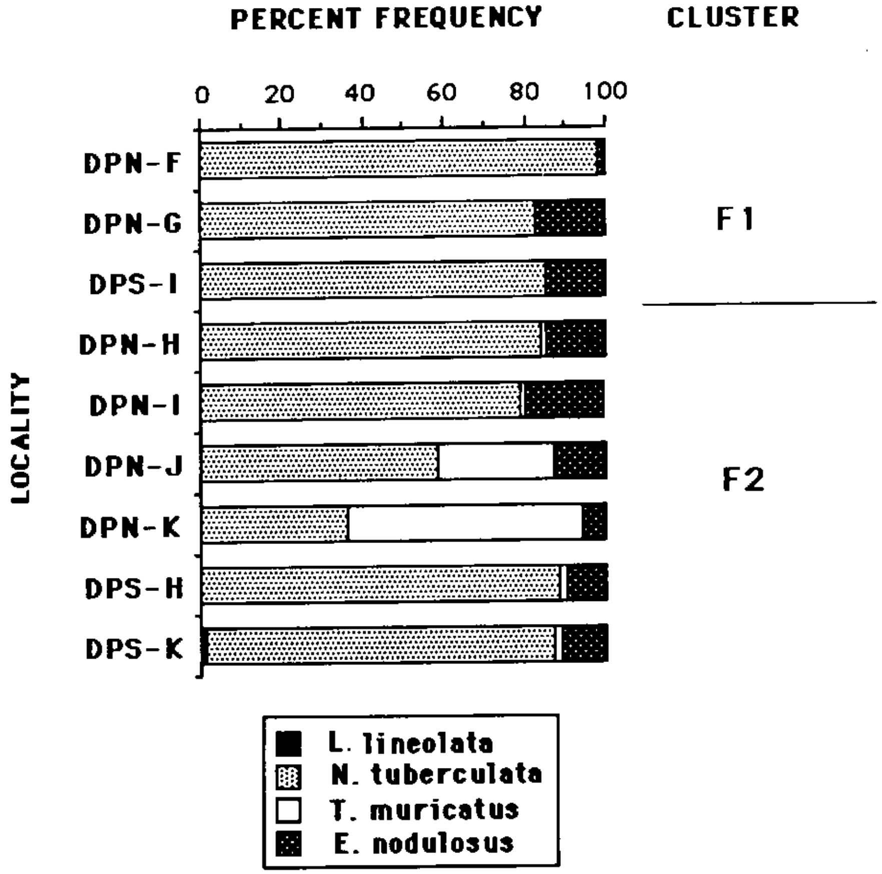


Figure 19 *Stacked bar histograms for faunas that occur in clusters F1 and F2 defined on Figure 18. Such bar histograms give a good visual impression of the composition of the constituent faunas and provide added information for the establishment of clusters and sub-clusters.*

Colbath, S.L., 1985, Gastropod predation and depositional environments of two molluscan communities from the Miocene Astoria Formation at Beverly Beach State Park, Oregon: Journal of Paleontology, v. 59, p. 849-869.

Cooper, G.A., 1972, Homeomorphy in Recent deep-sea brachiopods: Smithsonian Contributions to Paleobiology, Number 11, 25 p.

Copper, P., 1986, Evolution of the earliest smooth spire-bearing Atrypoids (Brachiopoda: Lissatrypidae, Ordovician-Silurian): Palaeontology, v. 29, p. 827-866.

Craig, G.Y. and Hallam, A., 1963, Size-frequency and growth-ring analyses of *Mytilus edulis* and *Cardium edule*, and their paleoecological significance: Palaeontology, v. 6, p. 731-750.

Dixon, O.A., 1979, Late Silurian Plasmoporid and Stelliporellid corals (Heliolitidae) from the Canadian Arctic: Journal of Paleontology, v. 53, p. 642-656.

Dixon, O.A., 1986, The Heliolitid coral *Acidolites* in Ordovician-Silurian rocks of eastern Canada: Journal of Paleontology, v. 60, p. 26-52.

Elias, R.J., 1985, Solitary Rugose corals of the Upper Ordovician Montoya Group, southern New Mexico and westernmost Texas: The Paleontological Society, Memoir 16, 58 p.

Fox, W.T., 1968, Quantitative paleoecologic analysis of fossil communities in the Richmond Group: Journal of Geology, v. 76, p. 613-640.

Graus, R.R., 1974, Latitudinal trends in the shell characteristics of marine gastropods: Lethaia, v. 7, p. 303-314.

Hallam, A., 1967, The interpretation of size-frequency distributions in molluscan death assemblages: Palaeontology, v. 10, p. 25-42.

Hallock, P. and Glenn, E.C., 1985, Numerical analysis of foraminiferal assemblages: a tool for recognizing depositional facies in Lower Miocene reef complexes: Journal of Paleontology, v. 59, p. 1382-1394.

Harbaugh, J.W. and Merriam, D.F., 1968, Computer applications in stratigraphic analysis: John Wiley and Sons Inc., New York, 282 p.

Harper, C.W., 1969, Rib branching patterns in the brachiopod *Atrypa reticularis* from the Silurian of Gotland, Sweden: Journal of Paleontology, v. 43, p. 183-188.

Harper, C.W., 1978, Groupings by locality in community ecology and paleoecology: tests of significance: Lethaia, v. 11, p. 251-257.

Hazel, J.E., 1970, Binary coefficients and clustering in biostratigraphy: Geological Society of America, Bulletin, v. 81, p. 3237-3252.

Hewitt, R.A. and Hurst, J.M., 1977, Size changes in Jurassic liparoceratid ammonites and their stratigraphical and ecological significance: Lethaia, v. 10, p. 287-301.

Hodges, L.T. and Roth, A.A., 1986, Orientation of corals and stromatoporoids in some Pleistocene, Devonian, and Silurian reef facies: Journal of Paleontology, v. 60, p. 1147-1158.

Imbrie, J., 1956, Biometrical methods in the study of invertebrate fossils: American Museum of Natural History, Bulletin 102, p. 217-252.

Izuka, S.K. and Kaesler, R.L., 1986, Biostratinomy of ostracode assemblages from a small reef flat in Maunalua Bay, Oahu, Hawaii: Journal of Paleontology, v. 57, p. 662-685.

Jardine, N. and Sibson, R., 1971, Mathematical Taxonomy: Wiley Series in Probability and Mathematical Studies, John Wiley and Sons, Inc., New York, 286 p.

Jones, B., 1974, A biometrical analysis of *Atrypella foxi* n. sp. from the Canadian Arctic: Journal of Paleontology, v. 48, p. 963-977.

Jones, B., 1977, Variation in the Upper Silurian brachiopod *Atrypella phoca* (Salter) from Somerset and Prince of Wales islands, Arctic Canada: Journal of Paleontology, v. 60, p. 347-360.

Jones, B., 1978, Taxonomy and intraspecific variation of *Protathyris praecursor* Kozlowski, 1929 from the Upper Silurian of Somerset Island, Arctic Canada: Journal of Paleontology, v. 52, p. 8-27.

Jones, B., 1979a, *Atrypoidea erebus* n. sp. from the Late Silurian of Arctic Canada: Journal of Paleontology, v. 53, p. 187-196.

Jones, B., 1979b, *Atrypoidea* zonation of the Upper Silurian Read Bay Formation of Somerset and Cornwallis Islands, Arctic Canada: Canadian Journal of Earth Sciences, v. 16, p. 2204-2218.

Jones, B., 1981, *Atrypoidea* species from the Canadian Arctic Islands: Canadian Journal of Earth Sciences, v. 18, p. 1539-1561.

Jones, B., 1983, Analysis of growth patterns in brachiopods and its potential use for the identification of species: Journal of Paleontology, v. 57, p. 662-685.

Jones, B. and Rong, Jia-Yu, 1982, Comparison of the Upper Silurian *Atrypoidea* faunas of Arctic Canada and southern China: Journal of Paleontology, v. 56, p. 924-937.

Jones, B. and Smith, G. P., 1985: Paleoecology of the brachiopod faunas in the Lower Devonian Eids Formation of southwest Ellesmere Island, Arctic Canada: Journal of Paleontology, v. 59, p. 957-974.

Kermack, K.A., 1954, A biometrical study of *Micraster coranguinum* and *M. (Isomicraster) senonensis:* Royal Society of London, Philosophical Transactions, v. 237, p. B375-B428.

Kermack, K.A. and Haldane, J.B.S., 1950, Organic correlation and allometry: Biometrika, v. 37, p. 30-41.

Kitchell, J.A. and Clark, D.L., 1979, A multivariate approach to biofacies analysis of deep-sea traces from central Arctic: Journal of Paleontology, v. 53, p. 1045-1067.

Klapper, G. and Foster, C.T. 1986, Quantification of outlines in Frasnian (Upper Devonian) platform conodonts: Canadian Journal of Earth Sciences, v. 23, p. 1214-1222.

Kues, B.S. and Kietzke, K.K., 1981, A large assemblage of new eurypterid from the Red Tanks Member, Madeira Formation (Late Pennsylvanian - Early Permian) of New Mexico: Journal of Paleontology, v. 55, p. 709-729.

Laghi, G.F., Martinelli, G. and Russo, F., 1984, Localization of minor elements by EDS microanalysis in aragonitic sponges from the St. Cassian Beds, Italian Dolomites: Lethaia, v. 17, p. 133-138.

Lane, N.G. and Sevastopulo, G.D., 1985, Redescription of *Allagecrinus americanus* Rowley, 1895, a Late Devonian microcrinoid: Journal of Paleontology, v. 59, p. 438-445.

Lazarus, D., Scherer, R.P. and Prothero, D.R., 1985, Evolution of the radiolarian species-complex *Pterocanium:* a preliminary survey: Journal of Paleontology, v. 59, p. 183-220.

Ludvigsen, R. and Westrop, S.R., 1983, Trilobite biofacies of the Cambrian-Ordovician boundary interval in northern North America: Alcheringa, v. 7, p. 301-319.

Ludvigsen, R., Westrop, S.R., Pratt, B.R., Tuffnell, P.A. and Young, G.A., 1986, Dual biostratigraphy: Zones and biofacies: Geoscience Canada, v. 13, p. 139-154.

Mello, J.F. and Buzas, M.A., 1968, An application of cluster analysis as a method of determining biofacies: Journal of Paleontology, v. 42, p. 747-758.

Mergl, M., 1985, Cryptonella (Brachiopoda, Cryptonellacea) in the Lower Devonian of the Prague Basin (Central Bohemia). Vestnik Ústredniho ustavu geologickelo, v 60, p. 355-360.

Mills, P.C. and Langenheim, R.L., 1987, Wolfcampian brachiopods from the Bird Spring Group, Wamp Spring area, Las Vegas Range, Clark County, Nevada: Journal of Paleontology, v. 61, p. 32-55.

Murphy, M.A. and Cebecioglu, M.K., 1986, Statistical study of *Ozarkodina excavata* (Branson and Mehl) and *O. tuma* Murphy and Matti (Lower Devonian, *Delta* Zone, Conodonts, Nevada): Journal of Paleontology, v. 60, p. 865-869.

McDonald, K.B., 1975, Quantitative community analysis: recurrent group and cluster techniques applied to the fauna of the Upper Devonian Sonyea Group, New York: Journal of Geology, v. 82, p. 473-499.

McGhee, G.R. and Sutton, R.G., 1981, Late Devonian marine ecology and zoogeography of the Central Appalachians and New York: Lethaia, v. 14, p. 27-43.

McGugan, A. and May, R., 1965, Biometry of *Anthracospirifer curvilateralis* (Easton) from the Carboniferous of south-east British Columbia: Journal of Paleontology, v. 39, p. 31-40.

McHargue, T.R., 1982, Ontogeny, phylogeny, and apparatus reconstruction of the conodont genus *Histiodella*, Joins Fm., Arbuckle Mountains, Oklahoma: Journal of Paleontology, v. 56, p. 1410-1433.

McKinney, M.L. and Zachos, L.G., 1986, Echinoids in biostratigraphy and paleoenvironmental reconstruction: a cluster analysis from the Eocene Gulf Coast (Ocala Limestone): Palaios, v. 1, p. 420-423.

Noble, J.P.A., Logan, A. and Webb, G.R., 1976, The Recent *Terebratulina* Community in the rocky subtidal zone of the Bay of Fundy: Lethaia, v. 9, p. 1-17.

Nowlan, G.S. and Barnes, C.R. 1981, Late Ordovician conodonts from the Vauréal Formation, Anticosti Island, Quebec: Geological Survey of Canada, Bulletin 329, Part 1, p. 1-49.

Nyberg, A.V. and Schopf, J.W., 1984, Microfossils in stromatolitic cherts from the Upper Proterozoic Min'yar Formation, southern Ural Mountains, USSR: Journal of Paleontology, v. 58, p. 738-772.

Pabian, R.K., and Strimple, H.L. 1979, Notes on biometrics, paleoecology and biostratigraphy of *Cibolocrinus conicus* Strimple from Oklahoma, Kansas, and Nebraska: Journal of Paleontology, v. 53, p. 421-437.

Pachut, J.F. and Anstey, R.L., 1984, The relative information content of Fourier-structural Binary (presence-absence) and combined data sets: a test using the H.A. Nicholson collection of Paleozoic Stenolaemate bryozoans: Journal of Paleontology, v. 58, p. 1296-1311.

Parks, J.M., 1966, Cluster analysis applied to multivariate geologic problems: Journal of Geology, v. 74, p. 703-715.

Peck, L.S., Clarke, A. and Holmes, L.J., 1987, Size, shape and the distribution of organic matter in the Recent Antarctic brachiopod *Liothyrella uva*: Lethaia, v. 20, p. 33-40.

Pocock, Y.P., 1966, Devonian schizophoriid brachiopods from western Europe: Palaeontology, v. 9, p. 381-412.

Rogers, M.J., 1976, An evaluation of an index of affinity for comparing assemblages in particular of foraminifera: Palaeontology, v. 19, p. 503-515.

Scott, R.W., 1974, Bay and shoreface benthic communities in the Lower Cretaceous: Lethaia, v. 7, p. 315-330.

Sepkoski, J.J. and Sheehan, P.M., 1983, Diversification, faunal change, and community replacement during the Ordovician radiations, *in* Tevesz, M.J.S., and McCall, P.L., eds., Biotic Interactions in Recent and Fossil Benthic Communities: Plenum Press, New York, p. 673-717.

Sokal, R.R., 1966, Numerical Taxonomy: Scientific American, v. 215, p. 106-116

Sokal, R.R. and Sneath, P.H.A., 1963, Principles of Numerical Taxonomy: W.H. Freeman and Co., San Francisco, 359 p.

Sørenson, M., 1984, Growth and mortality in two Pleistocene bathyal micromorphic bivalves: Lethaia, v 17, p. 197-210.

Springer, D.A. and Bambach, R.K., 1985, Gradient *versus* cluster analysis of fossil assemblages: a comparison from the Ordovician of southwestern Virginia: Lethaia, v. 18, p. 181-198.

Thayer, C.W., 1974, Marine paleoecology in the Upper Devonian of New York: Lethaia, v 7, p. 121-155.

Trewin, N H , 1976, *Isopodichnus* in a trace fossil assemblage from the Old Red Sandstone: Lethaia, v. 9, p. 29-37.

Valentine, J.W. and Peddicord, R.G., 1967, Evaluation of fossil assemblages by cluster analysis: Journal of Paleontology, v. 41, p. 502-507.

von Bitter, P.H. 1972, Environmental control of conodont distribution in the Shawnee Group (Upper Pennsylvanian) of eastern Kansas: University of Kansas Paleontological Contributions, Article 59, 105 p.

Waters, J.A., Horowitz, A.S. and Macurda, D.B., 1985, Ontogeny and phylogeny of the Carboniferous blastoid *Pentremites* : Journal of Paleontology, v. 59, p. 701-712.

Watkins, R., 1975, Silurian brachiopods in a stromatoporoid bioherm: Lethaia, v 8, p. 53-61.

Webster, G.D., 1981, New crinoids from the Naco Formation (Middle Pennsylvanian) of Arizona and a revision of the family Cromyocrinidae: Journal of Paleontology, v. 55, p. 1176-1199.

Wilson, M.V.H. 1983, Digitizing outlines of fossil specimens using video frame capture: Canadian Journal of Earth Sciences, v. 20, p. 1747-1751.

Wishart, D., 1978, CLUSTAN (User Manual - Third Edition): Inter-University/Research Councils Series, Report No. 47, 175 p.

Worsley, D. and Broadhurst, F.M., 1975, An environmental study of Silurian atrypid communities from southern Norway: Lethaia, v. 8, p. 271-286.

Accepted 21 September 1987.

Appendix 1 **Data for assemblage W1 of *Atrypoidea phoca* from the Douro Formation, Pressure Point, Somerset Island. *L* = Length; *W* = Width; *H* = Height; *AA* = Apical angle. All measurements in centimetres.**

L	W	H	AA	L	W	H	AA
2.07	1.88	1.84	73	2.39	2.22	1.94	60
1.94	1.82	1.63	71	2.06	1.88	1.62	61
2.13	1.71	1.72	73	1.98	1.83	1.61	65
2.15	2.04	1.90	60	1.86	1.64	1.49	64
2.05	2.12	1.83	56	1.89	1.68	1.54	69
2.02	1.77	1.83	63	2.00	1.83	1.73	66
2.11	1.92	1.88	63	1.98	1.69	1.63	66
0.94	1.84	1.58	67	1.79	1.74	1.47	59
1.93	1.78	1.65	58	1.72	1.74	1.42	56
1.83	1.83	1.78	57	1.94	1.84	1.51	68
1.84	1.87	1.54	63	2.04	1.78	1.68	86
1.90	1.83	1.60	62	2.19	2.19	1.82	60
1.65	1.47	1.23	67	1.65	1.58	1.28	57
1.88	1.85	1.58	64	1.88	1.96	1.62	64
1.88	1.86	1.61	58	2.14	2.03	1.85	67
1.92	1.68	1.36	65	2.03	2.03	1.83	57
1.66	1.48	1.27	62	1.77	1.53	1.39	77
1.94	1.83	1.63	75	1.42	1.34	1.09	71
1.74	1.76	1.44	57	1.85	1.74	1.59	66
1.71	1.65	1.37	57	2.20	2.01	1.88	64
1.43	1.40	1.10	64	1.73	1.63	1.53	67
1.71	1.57	1.26	63	1.83	1.69	1.52	69
1.55	1.34	1.29	65	1.86	1.77	1.50	59

Appendix 2 **Data for assemblage X14 of *Atrypoidea foxi* f. A from the Douro Formation, Pressure Point, Somerset Island. *L* = Length; *W* = Width; *H* = Height; *AA* = Apical angle. All measurements in centimetres.**

L	W	H	AA	L	W	H	AA
3.94	3.03	2.68	44	2.75	2.58	1.75	43
2.85	2.41	1.95	55	2.68	2.12	1.58	44
2.58	2.30	1.62	50	2.17	1.92	1.32	53
3.12	2.80	1.90	46	2.58	2.14	1.72	49
3.16	2.62	2.00	51	2.48	2.34	1.43	49
3.13	2.53	2.05	48	2.42	2.30	1.64	53
2.94	2.68	1.84	48	2.29	2.16	1.35	50
2.69	2.48	1.79	43	2.08	2.08	1.43	57
3.13	2.69	2.00	46	2.48	2.09	1.48	43
3.10	2.59	2.01	51	2.28	2.26	1.40	55
3.08	2.64	2.08	47	2.59	2.14	1.52	45
2.71	2.26	1.73	53	2.27	2.04	1.53	46
2.75	2.53	1.40	42	1.92	1.76	1.08	47
3.15	2.73	1.64	43	2.44	2.13	1.44	46
3.43	2.77	2.08	49	2.13	1.97	1.16	44
2.83	2.43	1.84	46	2.30	2.13	1.43	46
2.70	2.35	1.82	55	2.35	2.23	1.37	45
2.49	2.24	1.56	41	2.63	2.26	1.61	39
2.08	1.75	1.18	56	2.64	2.23	1.72	49
2.33	2.17	1.25	43	2.33	2.00	1.47	52
2.37	2.03	1.18	56	2.50	2.32	1.20	38
2.74	2.47	2.03	50	1.67	1.61	0.77	53
3.04	2.68	2.21	47	2.08	1.89	1.07	48
2.88	2.36	1.60	45	2.51	2.18	1.32	48
3.00	2.62	1.93	43	1.73	1.47	0.84	50
2.83	2.62	1.69	46	2.36	2.03	1.33	47
1.97	1.81	1.02	42	2.12	1.98	1.27	49
2.94	2.67	1.83	41	3.03	2.10	1.79	48
2.73	2.22	2.00	50	1.68	1.71	0.84	46
2.60	2.18	1.62	53	1.77	1.56	1.23	60
2.68	2.32	1.53	46	2.08	1.78	1.16	54
2.64	2.39	1.97	60	2.57	2.23	1.66	50
2.59	2.24	1.44	45	2.58	2.45	1.56	54
2.90	2.46	2.28	44	3.33	2.95	1.78	35