Morphometric Methods

in Biostratigraphy

RICHARD A. REYMENT

Preface

This book is mainly about the quantitative analysis of borehole sequences of micropalaeontological data. Other fossil material can also be utilized, provided that the fossils occur in continuous and contiguous sequences. Although wide use is made of a variety of statistical methods, some of them not yet available in statistical textbooks, the book is not, by any means, a text on statistics.

Industrial micropalaeontology marked an important breakthrough in applied geology. At a time when electrical logging methods were only just beginning to be developed, the possibility of using the microfossil content of borehole samples, not only for dating the beds penetrated, but also for correlating between adjacent boreholes, was clearly an innovation of considerable economic importance.

Of recent years, there has been a certain tendency to downgrade the role of micropalaeontology in the petroleum industry. The impressive success of seismic methods in many situations, along with improvements in the standard "Schlumberger techniques", confront an undeniable lack of initiative on the part of micropalaeontologists in refining their methods and this has worked against us.

Nonetheless, microfossils contain a great deal of biological information that can be put to service in a wide variety of palaeontological and, more generally, geological situations. From time to time, isolated attempts have been made at improving the value of fossil data used in biostratigraphical work, usually by means of some kind of statistical procedure. Some 50

years ago, Brinkmann (1929) used quantitative analyses of English Middle Jurassic ammonites in a detailed biostratigraphical study and more recently Shaw (1964) employed much the same philosophy in producing a somewhat more general approach to basically the same problem.

To any palaeontologist who has followed recent developments in ecology and genetics, it is clear that there ought to be reasonable chances of extracting considerable biological information from fossil organisms. I hope to be able to convince the reader of this book that palaeontologists have not yet reached the end of the road by any means and that the biological information residing in fossils can be exploited for purposes of producing a more useful and nuanced stratigraphical analysis in many situations.

An introduction to the ideas expounded here was given in Chapter 5 of my book "Introduction to Quantitative Paleoecology" (1971) in which the preparation of time-ordered variational plots for single variables was considered. In "Multivariate Morphometrics" Blackith and Reyment (1971) took some of these concepts several steps further in that the multivariate generalization of chronoclines was discussed as well as the biological aspects of the methods.

The statistical analysis of sequential variation in fossil organisms is still in its infancy, despite the long history of isolated applications dating back some 70 years. A broad frontal approach to this fundamental subject has yet to appear.

I wish, therefore, to underline the experimental nature of the text I put before you. I should like you to accept it more as a challenge to your own ingenuity, rather than a final report upon which little betterment can be contemplated. I believe I have proved that the approach used here is valid and that the components of ecology, genetic variability, size and shape must be fully recognized if a useful expression of secular changes in fossils is to be achieved. It is, in fact, here that I think the most rewarding advances will be made in the future development of the subject.

Also important is the need for tailoring multivariate statistical methods to fit particular biological problems. Impressive steps in this direction have already been made by Burnaby (1966), Campbell (1979) and Gower (1976), but much remains to be done.

Colleagues and students to whom the ideas of this book have been presented are inclined to be suspicious about the time likely to be involved in making a biolog. The daunting spectre of endlessly measuring and calculating is often enough to quench any ardour felt at the first confrontation with the concepts. Undoubtedly, there is more work involved in doing a fully integrated quantitative study of fossils occurring in a sequence than in the usual run of palaeontological studies. The extra

labour is, however, slight in relation to the preparation of the samples for routine study and the taxonomical work. The problem for industrial laboratories becomes trivial, the costs being small compared with those for exploration and drilling. The advantages accruing from a quantitative biostratigraphical analysis of a section far outweigh the supplementary work, for a wealth of often unsuspected information about palaeoecological relationships can be obtained as a bonus in addition to the variational plots.

Chapter 1 presents the basic concepts of biological logging and topics of a general statistical nature likely to influence the outcome of a quantitative analysis; in this connection, atypical observations are given particular attention. Chapter 2 is concerned with exploring the biological background of quantitative biostratigraphy from the aspects of secular variation in morphological characters, speciation theory, palaeogenetics, ecology and ecophenotypic variation.

A return to questions of statistical methodology is made in Chapter 3; in particular, analyses by the methods of canonical variates and principal components are discussed, together with some newly developed improvements of biostatistical importance. Chapter 4 develops the concept of biologging, passing from simple univariate representations to multivariate statistical solutions. Consideration is given to the problem of establishing reference levels in boreholes and to correlation between boreholes by Gordon's slotting technique.

Chapter 5 is devoted to examples of biologs based on continuous morphometric characters. Case histories for Cretaceous and Tertiary ostracods and foraminifers are discussed and the subjects of growth invariance and stability of canonical vectors illustrated. The presentation of micropalaeontological information in relation to palaeoecological factors as an ecolog is taken up. The palaeoecological significance of electric borehole logs is reviewed in conjunction with some of the examples of this section.

In Chapter 7, the slotting technique is applied to a problem of correlation of physical logs in the face of constraints. A completely worked example for a Nigerian benthic species of foraminifers is given in Chapter 8. Finally, in Chapter 9, points of practical significance for quantitative biostratigraphical studies are considered.

Contents

Preface

1 Introduction

Presentation of the subject—Selection of characters—Frequencies—Problems of sampling—Consequences of multivariate deviations from normality—Normalizing transformation—The bug-bear of outliers—Summarizing comments

2 Biological Foundations

Models of speciation—Allopatric speciation—Selection landscape—Phenotypic substitution—Morphometrics of sequences of fossils—Morphometrical variables used for logging boreholes—Subspecies in palaeontology—Ecological factors—Ecophenotypic variation—Experimental studies in palaeoecology—Can the palaeoenvironment be measured physically?—Reliability of ecological observations on borehole samples—Palaeogenetics and borehole sequences—Spectral analysis of fossil data—Example of the spectral decomposition of variation

1
Introduction

Presentation of the subject

The methods presented in this little book are primarily suited for the special needs of borehole analysis in petroleum exploration. Borehole cores of marine sediments are often rich in microfossils and they provide a potential source of excellent material for the quantitative analysis of palaeontological problems at a level approaching that enjoyed by people working on living material. For several decades now, it has been common practice to identify the micro-organisms contained in borehole samples and to analyse the sediments of cores. Originally, foraminifers, and then ostracods, were the groups of interest but the recognition of the potential biostratigraphical usefulness of palynology heralded a major breakthrough in the early fifties. Of recent years, palaeoecological studies have been undertaken, at the descriptive level, by uniting the observations made on the microfossils with the sedimentary information. This is a natural and logical approach, for many of the facts of importance for understanding the distribution of the microfossils are available in the host-rock. For example, traces and tracks preserved in the sediments can often provide valuable data on the environment of deposition of the sediments, thence the world in which the micro-organisms once lived.

Detailed studies of the sedimentology of a core can aid in elucidating whether the microfauna is *in situ*, or whether it has arrived at its present location by post-mortem transport. Further information can be obtained from the scanning electron microscope (SEM) study of the specimens, as even minute evidence of reworking can usually be identified at high magnifications.

Despite the detail with which boreholes are studied in petroleum laboratories, it is surprising how little of the potential information contained in them, and obtained at great cost, is extracted. The analysis of evolutionary changes and morphological variation caused by ecological influences remains at a descriptive level; the usefulness of morphological changes for aiding the logging of boreholes is seldom realized and there do not seem to have been many attempts to exploit this source of additional information. The determination of the quantitative effects of environmental factors is the only certain manner of initiating and following through an evolutionary and ecological study.

Electronic computers are today available at all petroleum companies and most universities and there is, therefore, no reason why even the most complicated of the computations in this text should hinder the application of quantitative methods in the routine analysis of borehole samples.

The main theme of the present text is then to show how one can extract more biological information from palaeontological borehole samples. As is well known, long-ranging fossils are difficult to use in biostratigraphical and palaeoecological analyses. The solution of the problem can in part be seen as being biostatistical in that such material can often yield valuable stratigraphical results if treated by suitable statistical methods. Let us now briefly consider the main theme of this text in its simplest representation.

If one plots the means of some character or other of the shell of a fossil organism against time, it will be found that the points oscillate back and forth about the grand mean. If the means for each of the samples in chronological order are graphed against location in the borehole (hence time), a type of log of the average morphological variation will be produced, analogous to physical logs.

The construction of a biolog requires the following bases:

1. a biologically realistic background;
2. suitable biometrical methods;
3. geological assessment.

Biological considerations are naturally of primary importance as the variational patterns over time displayed by a shelled organism may be

of ecological origin, they may be genetic or, again, due to both or random. As indicated by several of the examples used here, sustained trends in the morphological variation of a species in a borehole sequence often seem to be relatable to the reaction of the organism to unswaying environmental pressure, although not always (cf. Ozawa, 1975). Cook and O'Donald (1971 p. 93) observed that changes in the average of some size dimension over very short distances may indicate changes in otherwise undetected environmental variables that affect fitness through shell size.

Borehole logs are a natural source of time series. This is a subject that can become very complex, but for most correlations between boreholes, it is sufficient to use simple statistical methods in order to obtain a useful reduction of the data.

The reasons for wanting to make a graphical record of the biological variation of some long-ranging species may not be clear, particularly as the zonal concept, as applied in micropalaeontology, centres about the identification of marker fossils, zonal indices and faunizones and using them for correlation.

The biolog is not proposed as an alternative to classical biostratigraphy. The methods of this book would normally only be attractive when the usual approaches have failed or been deemed ineffective. Such a situation could occur where there is a thick and homogeneous marine sedimentary sequence and just a few species which range throughout these beds. Normally, one would hardly think it worthwhile to carry out a micropalaeontological treatment of such material and correlation would have to be attempted on the physical logs unaided, even these being unlikely to be especially diagnostic.

Logs based on morphological variations of organisms cannot be used for correlations over great distances. Ideally, I should only feel safe in recommending correlation by biologs for boreholes drilled within a single basin. Exceptionally, however, major geological events may leave a pronounced imprint on the palaeoecology of an area, thence the morphology of some shell-bearing organisms.

A more difficult kind of problem must also be considered. If we have a sequence of samples in a borehole and wish to relate the time-dependent fluctuations in the means of the organisms occurring in these samples to determinable factors in the palaeoenvironment (that is, we wish to correlate between a set of "ecovariables" and a set of "morphovariables"), a suitable statistical method of analysis would be one that examined the strength of the association between sets. For example, the variation in size and shape of an organism is controlled by such factors as salinity and the availability of nourishment. Another way of extracting ecological information from fossil material is by analysing relative frequencies of organisms in relation to physicochemical factors of the environment.

Another type of question concerns the need to ascertain if a biological population is morphometrically homogeneous. This necessitates the application of suitable statistical methods of analysis at the outset of an investigation in order to uncover the eventual presence of groups or clusters in the data.

This book is in no wise a text in applied biometry. It is necessary to presuppose that the user has a fair grounding in basic statistical methodology and the interpretation of multivariate morphometric analyses. Suitable introductions are given in Simpson *et al.* (1960) for introductory biometrical analysis, Reyment (1971) for quantitative palaeo-ecological analysis, Morrison (1976) for an elementary introduction to the most common methods of multivariate analysis, and Blackith and Reyment (1971), for the interpretation of multivariate statistical analyses of biological data, with emphasis on taxonometrical problems.

The construction of time-dependent curves for the frequencies of species as a quantitative expression of stratigraphical relationships seems to have been conceived by the Swedish Quaternary geologist, Gerard de Geer. The main force behind the development of pollen-analytical graphs for studying the palaeoclimatology and stratigraphy of the Swedish Pleistocene was provided by a student of de Geer's, Lennart von Post.

Pollen-analytical curves are made from counts on frequencies of species of pollen in samples and are, consequently, a kind of semi-quantitative approach to several palaeontological problems. Pollen analysis has remained one of the principal tools of the palynologist, having achieved its most elaborate treatment in the hands of John Birks (see, for example, Gordon and Birks, 1972).

The use of variability in the dimensions of shelled organisms for producing series in time has received much less attention than frequency diagrams and this subject has been largely ignored by pollen-analysts.

Reyment (1963) used simple multivariate plots for making graphs for use in stratigraphical correlation, later expanding the concept to encompass ecological components (Reyment, 1970). The latter paper introduced a more general approach to the topic in that multivariate methods were utilized.

The idea of using fluctuations in the morphological characters of a shelled species to produce a stratigraphically applicable plot seems to have been presented for the first time by Brinkmann (1929) in his study of British Jurassic ammonites. Shaw (1964) made use of much the same idea in his book on quantitative methods in biostratigraphy. Neither of these applications is directed towards extracting biological information from the data, nor do they take account of differences in means resulting solely from arbitrary mixtures of growth stages and size variations.

I have in the ensuing text made frequent reference to what I have dubbed "biologs" (Reyment, 1978), as well as "curves" based on the morphological variation of organisms as a function of time. These plots are, strictly speaking, not continuous curves, as are gamma ray logs, resistivity logs, etc., since they are made from discrete sampling points in a rock column. Although this point might not seem very significant, it should be kept well in mind when comparisons are being made between a biolog, which is discontinuous in origin, and continuously recorded observations.

The importance of graphically studying any data set cannot be overestimated (Reyment, 1971), and this will become apparent in connection with the examples. Everitt (1978) has published a useful summary of some of the more important graphical methods for multivariate data.

Selection of characters

The selection of characters for use in a morphometric analysis (for a definition of morphometry, see Blackith and Reyment, 1971) should be given careful thought before embarking on a study. Foraminifers and ostracods are often measured in quantitative taxonomical studies, but without due regard for the relevance of the characters chosen. A redundancy analysis of the variables provisionally selected for describing the variation of a species is therefore a desirable starting point.

Confined as we are in palaeontology to the hard parts of organisms for our measurements, it may be difficult to find characters of functional morphological significance, as would be of interest for ostracods, or morphologically well integrated attributes in foraminifers. Consequently, in addition to the task of producing a biolog, there will often be a need for an introductory round of multivariate analyses directed towards mapping the variability of the material and singling out diagnostic variables, in particular, those that are sensitive to modifications in environmental conditions.

Basically, there are two classes of characters, *continuous* and *discontinuous* variables. The most commonly used characters in the present connection are continuous variables, that is, variables which can take any value within a certain range. For example, length, height, breadth of the test of a foraminifer. These characters are measured in millimeters or microns and the values observed for a sample will be found to range without break, in a sufficiently large homogeneous sample, between upper and lower limits.

I use the terms character and variable synonymously. The word morphocharacter is used if the morphological aspect of a character is to be emphasized as opposed to, say, an ecological variable. Customarily, the measurements will be obtained in terms of some arbitrary scale. For practical studies, it may not be necessary to convert these to micrometres (microns).

Discontinuous characters are of the kind: presence or absence of some ornamental features, such as spines in ostracods, number of spines along the margin of the carapace, presence or absence of a posterior spine, ornamental categories, number of lateral ridges on a foraminifer, etc. Discontinuous (or discrete) characters are, as implied by their name, not gradational—there are either five or six spines, but nothing in between.

Now, it is possible to consider a kind of intermediate situation in which one measures lengths, say, of spines. This kind of measure can be rewarding in the study of living organisms, but is seldom useful in palaeontological work, owing to the ease with which ornamental characters are damaged. This is not to say that suitable material cannot occur, and where this is so I recommend its study. In this book, discontinuous characters have been used in a particular kind of analysis involving secular shifts in frequencies. The results of this study are interesting and there is little doubt that much more use should be made of such material in quantitative palaeontological work.

Frequencies

Most people working on the quantification of biostratigraphy make use of a variety of continuous variables, namely, proportions calculated on frequencies of species occurring in a rock column. Although some use of frequencies is made in this book, most of the methods are based on continuously distributed morphometrical values. That is, attention has been directed towards studying variations in the organism rather than the relative numbers of the organisms in the sampled material. The subject of the statistical analysis of frequency data of fossils has been well studied by Gordon and Birks (1972, 1974) with particular reference to palynological data.

I have found that it is possible to find clear relationships between ecological factors of the ancient environment and the organisms that lived therein with respect to variation in their hard parts. It is likewise clear that variations in numbers of organisms also reflect factors in the palaeo-environment and there is no doubt that the study of connections

between environmental factors and relative frequencies of organisms will eventually prove to be a useful tool.

In studying discontinuous variations in fossil ostracods, one has the possibility of utilizing the results obtained by geneticists for living material. For example, the polymorphism shown by the frequencies of anterior spines of *Cyprideis* are paralleled by the discrete banding variations in the snail *Cepaea*. We know very little about the genetics of ostracods, sad to say, but geneticists have put in a great deal of work on *Cepaea* and much of this can be used in unravelling the complicated variational patterns we find in fossil and recent ostracods. This topic is taken up in the next chapter.

Problems of sampling

It is often found that when palaeontological samples are taken from the same immediate area, they may seem to possess different statistical properties; i.e. there may be marked differences in means of a character and in the conclusions to be drawn from the statistical tests, even where relatively large samples are used. Statistical studies made on collections of a fossil species along the strike of a bed may yield sample means and variances that differ statistically significantly from each other. This kind of observation has sometimes been used by opponents to the application of quantitative principles in palaeontology as an argument for the futility of analysing measurements on fossils. The reasons for these differences in means, hence seemingly excessive variability, are, however, usually interpretable as being due to geological causes, and a cautious approach to the study of such data can yield acceptable and reliable results. Particular attention must be paid to the post-mortem sorting of shells (Kilenyi, 1971), reworking, distortion resulting from sedimentary compaction etc., as such factors can easily lead to considerable differences in means and variances in comparison with undistorted material.

Particular attention must be given the identification of *outliers* in the data, as even one divergent specimen can greatly influence variances, covariances and correlations. Gnanadesikan (1977) has devoted a great deal of time to studying and developing a set of methods for picking up outlying observations. There are many techniques available, some based on direct calculation and others on graphical procedures. Barnett and Lewis (1978) have also given the subject detailed treatment.

In most morphometric studies, it is usually assumed that the data are normally distributed. There are univariate tests for normality which are

described in most textbooks on statistics (useful procedures are given in Simpson *et al.*, 1960). The first step in any study is therefore to check the data for *normality*. This can be done conveniently by testing each of the variables, if the study is multivariate, for univariate normality by the usual methods for skewness and kurtosis. This step is then followed by a graphical appraisal of the variables pairwise as bivariate scatter plots. Often, outlying observations can be identified at this stage. The final stage is to carry out a collective multivariate analysis of the samples. Probability plots of the ordered observations can often be very helpful as well.

Caution should be exercised in interpreting the results of multivariate studies of normality as the implications are still not fully understood. The univariate concepts of skewness and kurtosis are well known and for much palaeontological work, it may be sufficient to have obtained an opinion of their importance in a sample in order to be able to make reasonably well founded assumptions about the multivariate normality of the variables being considered.

Mild skewness is not necessarily a non-biological complication and many variables display a natural right skew in their distributions (see Simpson *et al.*, 1960). The coefficient of kurtosis is a measure of peakedness in a sample. Peakedness can often be related to geological causes, such as the effects of size and, or, shape-sorting. Belemnites are very susceptible to shape-sorting as they are easily rolled about by currents. At the microscopic level, ostracod valves may be sorted by bottom currents (cf. Reyment, 1960).

Another kind of complication is mixing of samples (for example, through reworking) which causes a particular variable to have a mixed distribution.

Mixing will also bring about a significant deviation in kurtosis from the properties of the normal distribution as such a mixed distribution will tend to be flattened, bimodal, long-tailed, or polymodal. In the bimodal case, it is easier to identify the mixed nature of the sample and there are now special statistical methods that may be used for extracting a greater amount of useful information from such material than was possible a few years ago. Kurtosis complications can also derive from taxonomic mis-identifications and post-mortem deformation (compaction by sediment load).

The identification of mixed distributions is often an important part of a palaeo-ecological study, as factors of biological significance may lie at the root of the heterogeneity. Fairly suitable methods for treating mixed distributions are available, such as so-called Pearsonian curve analysis for reconstructing the probable shapes of the component distributions.

It is not advisable to attempt to unravel mixed data in the above

manner if one has multiple measurements on the specimens and it may be more valuable to proceed by means of a suitable multivariate method for finding clusters in data sets.

Consequences of multivariate deviations from normality

Very many of the standard multivariate methods in wide use have been specifically designed for normally distributed data. This is the outcome of the fact that the mathematics for treating problems involving multivariate, normally distributed variables are much simpler than for variables that deviate from the normal distribution. If one attempts to pass non-normal data through computer programs for multivariate procedures based on the multivariate normal distribution, one runs the risk of arriving at misleading conclusions.

The risk of doing something seriously wrong in this connection is not equally great for all methods, as some of them are *robust* to deviations from the theoretical requirements. Such a method is that known as the generalized statistical distance. It is well to remember that robustness was not built into the method when it was developed originally; things just turned out to be so. This suggests that it is often worthwhile carrying out empirical tests of methods to see how well they behave under increasingly serious deviations from normality, as such information is not well documented in the statistical literature. Where data are markedly skewed, the generalized distance may be appreciably distorted.

You should be aware of the fact that much work is going on at the present time on the subject of non-parametric multivariate procedures. I do not make use of such methods here, but there seems to be every chance that non-parametric multivariate methods will occupy an important position in the future.

Normalizing transformations

It has become an almost standard applied biometrical procedure to take the logarithms of morphocharacters in multivariate work in order to diminish the effects of differential growth (i.e. allometric growth). Jolicoeur (1963) proposed that the first eigenvector derived from a covariance matrix of logarithmically transformed variables could be used as a multivariate generalization of Huxley's (1932) concept of bivariate

allometry. The same idea has also been taken up by Hopkins (1966). This approach turns out to be a practically useful representation of a multivariate differential growth relationship in cases where almost all of the variance is attached to the first eigenvector of the logarithmic covariance matrix.

Empirical studies of the effects of the logarithmic transformation on morphovariables (Reyment, 1971) have shown that even if the univariate values of skewness and kurtosis are bettered by the transformation, the multivariate analogues may not benefit to the same degree or may, even, have their normality worsened. It should perhaps be mentioned here that in an analysis of morphometrical variation, it is not permissible to transform logarithmically only a few of the variables to improve their conformity with the normal distribution, and to leave untransformed the ones found to be satisfactory. Such a step would influence the expression of growth patterns in the material. This is a biological interpretation and a statistician might well advocate the logarithmic transformation of only some of the variables in order to improve their agreement with the properties of the normal distribution.

The bugbear of outliers

Experience teaches that the most important single source of distorted results is that caused by outliers in the data. As is illustrated in Table I, the inclusion of a wayward specimen in an analysis can wreak havoc with correlations, principal components, factor analysis, etc. It is therefore essential that the possibility of contamination from this source be kept well in mind during an analysis. Outliers can be of many sources, not all of them necessarily unnatural or non-biological. In morphometric studies, they can derive from juvenile individuals—not an infrequent problem with some species of ostracods, deformed specimens, or from misidentified material. Another source of outliers is of geological origin, namely, specimens deformed by compaction of the host sediment. The most common cause of outlying observations is, however, of purely human origin, to wit, mistakes in measurements or punching errors. Gnanadesikan (1977) has made a detailed study of various ways of detecting outliers in a sample. Many of the cases, he considers, are not applicable to palaeontological data; nevertheless, his book is worth detailed study and can be highly recommended to all people working with multivariate statistical methods.

For most purposes, a preliminary graphical appraisal of the data will usually be sufficient to disclose the presence of rogue observations. You

TABLE I

Effect on the correlation matrix of exchanging a normal specimen of *Afrobolivina afra* for one deformed by crushing (N = 22) Example of the deforming influence of a single outlier

1A   Correlation matrix for sample of normal specimens from the borehole Araromi I at a depth of 1550 ft.

	1	2	3	4	5	6	7	8	9
1	1.0000	0.6051	0.7648	0.6891	0.6070	-0.4087	0.6953	0.3981	0.4126
2	0.6051	1.0000	0.8570	0.7225	0.6430	0.0783	0.5695	0.4809	0.4791
3	0.7648	0.8570	1.0000	0.6845	0.6780	-0.2131	0.6643	0.3889	0.5756
4	0.6891	0.7225	0.6845	1.0000	0.6677	-0.1553	0.7303	0.6269	0.4513
5	0.6070	0.6430	0.6780	0.6677	1.0000	-0.2386	0.7273	0.4270	0.4605
6	-0.4087	0.0783	-0.2131	-0.1553	-0.2386	1.0000	-0.3700	0.6490	-0.4598
7	0.6953	0.5695	0.6643	0.7303	0.7273	-0.3700	1.0000	0.6490	0.6124
8	0.3981	0.4809	0.3889	0.6269	0.4270	0.6490	0.6490	1.0000	0.2564
9	0.4126	0.4791	0.5756	0.4513	0.4605	-0.4598	0.6124	0.2564	1.0000

1B   Effect on the above correlation matrix of exchanging an outlier (specimen deformed by crushing so that the values for variables 1, 2 and 7 are distorted) for a normal specimen.

	1	2	3	4	5	6	7	8	9
1	1.0000	0.8605	0.7041	0.5308	0.5873	-0.3641	-0.2577	0.4415	0.4884
2	0.8605	1.0000	0.8045	0.5974	0.6418	-0.0831	-0.2345	0.5104	0.5456
3	0.7041	0.8045	1.0000	0.6845	0.6780	-0.2131	0.1647	0.3889	0.5756
4	0.5308	0.5974	0.6845	1.0000	0.6677	-0.1553	0.4155	0.6269	0.4513
5	0.5873	0.6418	0.6780	0.6677	1.0000	-0.2386	0.2653	0.4270	0.4605
6	-0.3641	-0.0831	-0.2131	-0.1553	-0.2386	1.0000	-0.1189	0.0418	-0.4598
7	-0.2577	-0.2345	0.1647	0.4155	0.2653	-0.1189	1.0000	0.2432	0.1541
8	0.4415	0.5104	0.3889	0.6269	0.4270	0.0418	0.2432	1.0000	0.2564
9	0.4884	0.5456	0.5756	0.4513	0.4605	-0.4598	0.1541	0.2564	1.0000

can usually begin by looking at the histograms of the variables and then continue by checking the pairwise plots of them. These operations are easily programmed and are therefore no more time-consuming than scanning any kind of graphical output. Everitt (1978) has given a summary of some of the more common techniques available for the graphical appraisal of multivariate data.

Summarizing comments

It is now appropriate to summarize some of the remarks made in this section.

First, where samples have been taken along the strike of a bed, it can be expected that there will be certain statistically significant differences in means and variances owing to non-biological variation imposed by any of a number of geological causes.

Secondly, chronologically separated samples usually do (and should) show differences in means, variances and even covariances.

Thirdly, atypical observations, or outliers, can greatly distort the results of statistical calculations. Even a single rogue observation can cause serious trouble.

Fourthly, the effect of logarithmically transforming non-normal data may be a mixed blessing, at least in the case of multivariate samples. Apart from the fact that the investigator is removed one step from the biological relationships, the betterment in multivariate normality may be slight, or even the reverse may take place. Notwithstanding these reservations, biological considerations may demand that data be logarithmically transformed. However, in such cases, this is the prime concern, not the question of improving normality. For example, in analyses of differential growth, particularly those directed towards studying allometric relationships, it is usual to proceed by means of the logarithms of the measurements.

2
Biological Foundations

Models of speciation

Before embarking on the subject of logging palaeobiological variation in sequentially ordered samples, it is necessary that we consider briefly current thought on speciation. This subject is of great interest to palaeontologists at the present time, because an essential part of the proof for modes of speciation will ultimately have to come from palaeontological sources.

This brief review must be made, as it is necessary that we understand, as far as possible, the nature of the oscillatory sequences formed by morphological variation in the hard parts of, for example, ostracods and foraminifers. Should such a series display an evolutionary unidirectional trend? Should the series fluctuate randomly over its stratigraphical range and, eventually, be replaced abruptly by a related (descendant) species? How important are ecological influences on the fluctuations in morphology? Does random variation over-ride evolutionary and ecological causes?

Some idea of the relative importance of these various considerations is

clearly required if the variational histories of fossil species are to be used for making a borehole log.

Among recent publications concerned with this subject, the following can be mentioned: Dodson and Hallam (1977), Eldredge and Gould (1972), Hayami (1973), Hayami and Ozawa (1975), Gould and Eldredge (1977) and Gould (1977). A classical reference is Mayr (1963). The works cited here are not the only references concerned with detailing aspects of the question, although they are noteworthy for the broad coverage of the literature they contain. The journals *Evolution*, *Paleobiology* and *American Naturalist* often contain articles contributing to modern thinking on evolutionary topics.

One might say that the classical concept of evolutionary change as often conceived by palaeontologists is one of gradual and even transition from one species to another. This is sometimes referred to as the gradualistic concept of evolution and it postulates that new species arise by the slow and regular transformation of a large ancestral population into its descendants, this shift taking place over all or most of the range of the ancestral species. This is the concept of sympatric speciation.

In his monumental work on evolution, Heberer (1974, p. 408) used long-term unidirectional trending in foraminifers in illustrating a gradualistic origin of species. Work I have done on a species of benthic foraminifers, and described in Chapter 8, indicates that a gradual size increase in the diameter of the megalospheric proloculus occurs. This observation is akin to, but not exactly equatable with, Ozawa's (1975) conclusion for fusulinids.

Allopatric speciation

As well discussed by Mayr (1963) and White (1978), for example, an alternative model to sympatric speciation and phyletic gradualism can be based on the ethology and distribution of living species and the interplay between the organism and the environment in which it lived. Naturally, the two theories are not diametrically opposed to each other and both processes seem to occur in nature (Mayr, 1963, pp. 414, 415).

Current thought on the subject tends to postulate two variants of allopatric speciation. One of these results from isolation, ethological, ecological or geographical; this is the kind usually considered in an analysis of speciation (Mayr, 1963; 1969, p. 194). The second kind is more recent and can be modelled in terms of the French mathematician, René Thom's, so-called theory of catastrophes, Wright's (1968) concept of the selection landscape, and quantum evolution (Simpson, 1953).

The customary concept of allopatric speciation (cf. Eldredge and Gould, 1972) is that new species arise when a small sub-population of a species becomes locally isolated at the periphery of its geographical range (Mayr, 1963, p. 392), thus forming a so-called "peripheral isolate". A peripheral isolate develops into a new species if isolating mechanisms evolve which block the renewal of gene flow should the new form come into contact with the parent species. This model requires that the new species does not originate in the same geographical location as the parent species, although there is nothing to stop the two co-existing, even competing, if the derived form should invade the area of distribution of the parent species. Eldredge and Gould (1972) suggest that such an event would show up in the fossil record as one species suddenly being replaced by a related one with or without a short period of co-existence, but without the presence of individuals showing transitional morphologies.

The possibility of the remains of individuals deriving from the process of splitting by isolation ever being found is small and so this aspect of the so-called theory of splitting by punctuated equilibria will probably remain conjectural for a long time.

Selection works towards maintaining an equilibrium between the population and its environment and it is therefore to be expected that the morphological features that distinguish the descendant species from its ancestor appear just before or after genetic isolation. Such differences are thought to become accentuated if the two species subsequently co-exist. This phenomenon has been termed character displacement (cf. Pielou, 1974, p. 332). A consequence of this trait, a variety of polymorphism, is that where the species do not occur together, their diagnostic morphological characters may become less unlike.

After having become established in its environment, the morphology of a descendant species will tend to become stabilized.

The model of allopatric speciation by "punctuated equilibria", as presented by Eldredge and Gould (1972), depends on the following requirements: (1) new species arise by the splitting of lineages; (2) the new species develop rapidly; (3) a small sub-population of the ancestral population gives rise to the new species; (4) the new species originates in a very small part of the total geographical extent of the ancestral form; (5) in any local section in which the ancestral species occurs, the fossil record for the origin of the descendant should consist of a sharp morphological break between the two taxa. This break reflects the arrival of the new species into the main area of distribution of the ancestor.

Breaks in evolutionary continuity are not uncommonly observed in the stratigraphical column and these may often be due to geological causes, for example, hiatuses in the rock succession. There may, however, be many examples of allopatric events masquerading as non-successions.

Geneticists tell us that the reason why reproductive isolation does not arise in every local population may be sought within the framework of the concept of homeostatic systems. Such systems are well buffered to resist change and to maintain stability despite disturbances. The importance of peripheral isolates lies in their small size and the usually divergent environment, outside of the normal range of the parent species, which they inhabit. It is in this isolated habitat that selective forces are strong enough and the number of individuals sufficiently reduced to produce the necessary genetic pressure that overcomes homeostasis.

The usual model of allopatric speciation requires some kind of physical isolation to exist. For land animals, it is not difficult to envisage geographical barriers providing the necessary block to gene flow. There are also other kinds of isolation such as ethological characteristics, ecological differentiation, etc. Mayr (1963) gives numerous examples. In the relatively homogeneous marine environment inhabited by ostracods and foraminifers, it is less easy to find convincing evidence of such isolation. This is not to say that physical isolation cannot occur at the microenvironmental level; as an example, we can take the largely different environments lived in by endobionts and epibionts, although these have an interface.

Even the most eager proponents of allopatric speciation do not go so far as to deny the occurrence of sympatric speciation, although detailed documentation is scarce. The time-correlated shift in size of the proloculus of the verbeekinoid foraminifer *Lepidolina multiseptata* seems to be an example of "phyletic gradualism". This is a well-documented study (Ozawa, 1975) in which the size change can be related to a change in the habitat from shallow-water carbonates to clastics deposited in deeper waters. Ozawa (1975) concluded that the gradual drift in the ecological conditions in which this stock lived was sufficient to supply the selection pressure required for the genetic change. A similar case will be analysed in a later chapter.

Selection landscape

This concept is actually due to Wright (1968) but it has been most closely associated with the name of Simpson (1953), who has used it in the practical analysis of horses. Wright developed the theory in 1932 (cf. Wright, 1968). The terrain of possible structural variation is conceived of as a landscape of hills and dales. The extent and direction of variation in a population can be represented by outlining an area and a shape on the field. The direction of positive selection is uphill, that of negative selection, downhill. Selective forces cause the population to move (evolve)

along the surface to the nearest hill, or adaptive peak, until a local maximum on the landscape has been reached. Each population is considered to evolve in each environment on an adaptive surface determined by ecological conditions. These ideas are inherent in the so-called fold-catastrophic model of allopatric speciation.

In the selection landscape concept of Wright and Simpson, the model takes account of all individuals in the sub-population of interest. This is clearly not an easy construction to handle mathematically and the treatment of the subject presented by Dodson (1976) and Dodson and Hallam (1977) employs the mean phenotype instead. The sub-population is assumed to be in equilibrium with the environment it inhabits, except at thresholds, so that the point representing the phenotype will be located at a peak, except where the peak is annihilated by a saddle.

A topological theorem due to Thom (1969) for certain general kinds of discontinuous processes has shown itself to be useful for describing some biological and geological phenomena in which sudden jumps occur. The classical calculus is applicable to processes showing regular changes but it is unable to provide an easy description of systems having sudden discontinuities, such as those occurring at the transition between states in magmas from the liquid phase to the gas phase, the failure of sedimentary rocks under the gradual build-up of stress, cell division, and allopatric speciation under environmental pressure and in the absence of geographical barriers.

The theory is most directly applicable to systems whose mathematical description may be couched in terms of the Aristotelean relationship of velocity being proportional to friction. The palaeontological significance of the model is examined by Dodson and Hallam (1977).

Thom's theorem states that allowing for certain exceptions of no consequence in practical connections, it is always possible to effect a smooth, reversible change of coordinates in such a way that, in the neighbourhood of a given point, the system shows one of seven kinds of behaviour; these are the seven elementary catastrophes. Here the word "catastrophe" is meant to imply sudden change.

The first and simplest of these categories, the "fold catastrophe", is the one of potential significance in biological studies involving speciation.

The behaviour of the system is governed by an "energy function", which may be denoted E. The system is envisaged as tending rapidly to a steady state, that is to a state of equilibrium. The energy function of the fold catastrophe is

$$E = \tfrac{1}{3}x^3 + ax \tag{1}$$

where x denotes some morphological character, normally a single variable but also a multivariate transformed variable. The fitness function $V(u, x)$

can replace E in (1). In this function, the environment is represented by u, and the phenotype by x, as before.

In Fig. 1, it will be seen that for $u < 0$, $V(u, x)$ has a local maximum (or unidimensional peak) and a local minimum (or unidimensional saddle). These cancel out each other at $u = 0$. It is likewise obvious from this figure that $V(u, x)$ has no critical points for $u > 0$. The point $u = 0$ corresponds to an environmental threshold, that is, a point at which the ecological forces operating on the organism trigger off a sudden structural change.

Speciation taking place by such a jump is sometimes referred to as "quantum evolution". The actual quantal phase is therefore very sudden and it is unlikely that the remains of organisms that took part in this phase will be found (cf. Mayr, 1969). After having crossed the threshold, the new species will begin to adjust to the prevailing environmental conditions.

The set of critical points of V, the set of points (u, x) for which $\partial V/\partial x = 0$, is yielded by all points which lie on the parabola in Fig. 2. Here, the lower part of the parabola corresponds to local maxima of V while the upper, dotted part corresponds to the local minima.

On approaching the threshold, there should be a parabolic morphological trend (t^* denotes the threshold time and p^* the mean phenotype at that time)

$$p - p^* = K(t^* - t)^{1/2}$$

where K is a constant. It is this expression that gives rise to the claim that a sub-population approaching a fold-catastrophic threshold follows a
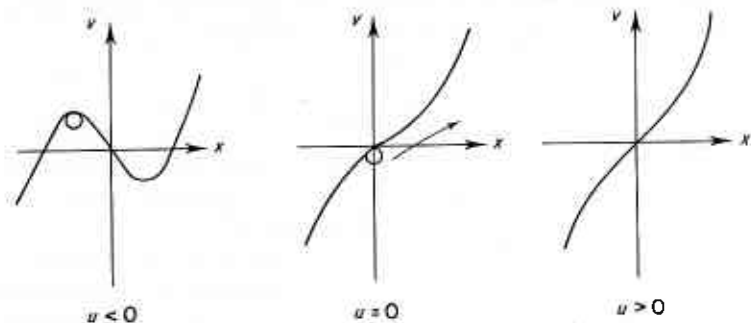

Fig. 1 The effect of varying the environmental parameter u on the function $V(u, x) = \frac{1}{3}x^3 + ux$. The catastrophic element in the depicted sequence of events can be demonstrated by the behaviour of the "bubble of air" shown in the two left-hand figures of the set. For $u < 0$, it is held trapped but at $u = 0$, the trap disintegrates and the bubble rises rapidly to the surface. This homely analogy illustrates the "catastrophic" or sudden change at $u = 0$ in relation to the stability pertaining for $u < 0$.
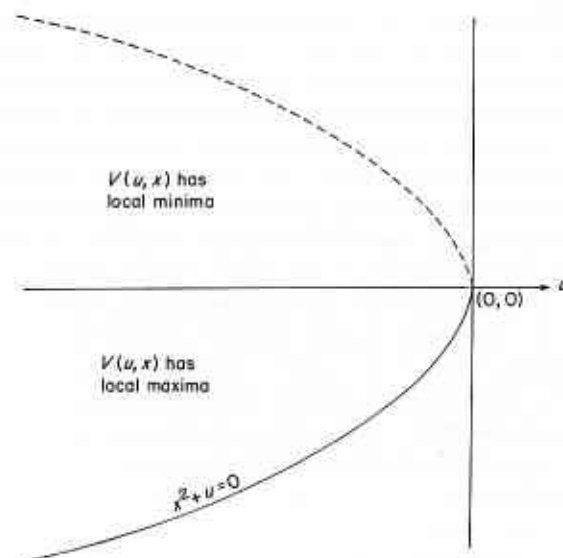
Fig. 2 The "catastrophe curve" is represented by the parabola $\partial V/\partial X = x^2 + u = 0$. A point (u, x) on the solid part of the parabola maximizes $V(u, x) = \frac{1}{3}x^3 + ux$ locally, while a point on the dotted part minimizes $V(u, x)$ locally. The point $u = 0$ is a threshold.

"square-root law". It may be difficult to fit a parabola to a set of points that may be few in number, an outcome of the fact that the catastrophic event is supposed to occur rapidly and so there will be little chance of obtaining satisfactory samples, even under the most favourable of conditions.

Phenotypic substitution

Hayami and Ozawa (1975) have considered morphological changes in terms of genetic variability and selection pressure in the population. The occurrence of morphological breaks in a sequence is thought by these authors to be able to arise through phenotypic substitution. Overlapping of two phena can be explained by the fold-catastrophic model of allopatric speciation; some cases can also be interpreted within the framework of the model of phenotypic substitution. Hayami (1973) has examined this question using rib frequencies in the pelecypod *Cryptopecten vesiculosus*, a Japanese Neogene species. Drawing out the conclusions presented by this model, one would then surmise that a continuous morphological shift ought to be expected where natural selection acts on a continuously varying population.

As you will see in one of the examples later on (Chapter 5), some ostracods lend support to this model, particularly in connection with variation in polymorphic characters. When animals showing polymorphism, apart from sexual dimorphism, are analysed in a quantitative biostratigraphical study, the model of phenotypic substitution should be kept in mind. Ostracods come under this heading for they often show polymorphism in discontinuously varying ornamental characters linked to continuously varying size variables (Reyment and Van Valen, 1969).

To complete this discussion, a few notes on the concept of sympatric speciation are in place. By this model, one species transmutes gradually into another, its successor, by the gradual accumulation of genetically different properties. Thus, over a certain period of time, a closely sampled sequence of the lineage would display infinitesimally slight differences in characters, these differences being less and less marked, the smaller the distance between samples. Hallam (1978) has reviewed possible examples of sympatric evolution.

Morphometrics of sequences of fossils

The main theme of this book is to show how series in time of oscillations in morphometrical variables of fossil species can be utilized practically in quantified biostratigraphy. We have just briefly considered models of speciation, although speciation will seldom be a factor of importance in biological well logging. It is nonetheless necessary to have a reasonably clear concept of what happens in an evolving fossil population with respect to changes in the morphology of the organisms forming it. Clearly, the usual model of allopatric speciation would be accompanied by random fluctuations over the range of the species. The fold-catastrophic model would not be expected to display the same kind of random variation in morphology and there would be a parabolic short-term trend just prior to the speciation event. In sympatric speciation, there would be a long-term trend in the morphological variation. In all cases, ecological effects would bring about shorter or longer periods of trending in morphological variation.

My opinion at the outset of the work leading to the writing of this book was that morphological changes were random, apart from the effects of the environment. There seem to be many cases of this in the material I have studied. However, the species I have analysed in closest detail, *Afrobolivina afra*, does show trending in some of its characters.

Morphometrical variables used for logging boreholes

The interpretation of the biological aspects of the statistical logging or borehole sequences of fossils by using the variations they display from sampling level to sampling level can often be conveniently made in terms of an allopatric model, at least, over short time intervals. In almost all analyses, the question of interpreting a speciation event will not arise. Nevertheless, an attempt should always be made in relatively well-documented sequences to identify trends if they occur.

Palaeogenetics is an insufficiently explored area and we are still much in the dark when it comes to interpreting speciation processes in the fossil record and documented studies are very rare. Gould's (1969) work on the evolution of the landsnail *Poecilozonites bermudensis* is a good example of allopatric speciation. Less well-documented is Reyment's (1975) analysis of the evolutionary relationship between the Late Turonian (Cretaceous) ammonite genera *Subprionocyclus* and *Reesidites* in which an allopatric model was thought to be the best, though not only, way of interpreting the data. In this study, it is important to note that there is no question of a gradient being involved; the essential evolutionary event is reflected in a reversal of loadings in canonical vectors, a subtle quantitative difference.

Allopatric speciation theory as applied to biostratigraphical analysis is diametrically opposed to the approaches of Brinkmann (1929) and Shaw (1964). An interesting, and lengthy, account of phylogeny and ontogeny is to be found in Gould (1977).

Subspecies in palaeontology

As has been noted by biologists on many occasions (cf. Mayr, 1969, p. 37), the application of the concept of subspecies to fossil taxa is not infrequently confused. Species that contain two or more subspecies (an arbitrary taxonomic unit) are said to be polytypic. The frequency of polytypic species varies from group to group and they are most common where species tend to form geographical isolates. Mayr (1969, p. 41) defines a subspecies as an aggregate of phenotypically similar populations which inhabit geographical subdivisions of the range of a species and differing taxonomically from other populations of the species. Consequently, the report of one or more subspecies of the one species from a single locality indicates a misunderstanding of what constitutes a subspecies. In palaeontological connections, it is likewise a wrong use of this taxonomic concept if a number of subspecies are recorded from the same bed and place.

Ecological factors

We have just considered models of speciation in relation to the expression of variational patterns over time. These theories recognize the importance of ecological factors in contributing to the development of populations and isolated sub-populations. Again, this is a field which has not been given nearly enough attention by zoologists and palaeontologists, in spite of the seemingly enormous literature on the subject. There has been an impressive amount of work on the general aspects of the subject but little on such questions as the relationships between microorganisms and their immediate environment. Many of these questions are so specific to a particular organism that it is not possible to extrapolate, for example, from what we know about the effects of temperature on mice to the reaction of marine ostracods to temperature.

I have here summarized some of the more comprehensive literature of interest to the micropalaeontologist but I make no claims for its completeness nor, indeed, for its general validity. In spite of our advances in the study of palaeoecology and the good work being done on the ecology of ostracods and foraminifers with special attention given to features of palaeontological interest, we are no more than at the outset of a task, vast in its scope, bountiful in its ultimate rewards.

For making biologically based logs of borehole sequences, it seems to me to be more useful to use benthic organisms. Although the stratigraphical appeal of planktonic organisms may suggest to the would-be user of logging techniques that such groups ought to possess useful properties, this is not actually so, as they are very difficult to relate to environmental factors recorded in the sediment in which they occur. I know that studies of Pleistocene climatological events have been greatly aided by quantitative analyses of planktonic foraminifers, but there we are dealing with a different scale of values and, moreover, the sediments normally sampled are deep-sea sediments which may be expected to reflect something of the conditions under which the foraminifers lived (cf. Imbrie and Kipp, 1971).

According to marine ecologists, ecological factors of major importance for marine benthic organisms are (cf. Kinne, 1970):

1. salinity;
2. temperature;
3. sediment, including the interstitial environment;
4. interaction with other organisms, including predation;
5. trace elements.

There are doubtless other factors of importance but their effects are not well understood. Still others are of doubtful palaeoecological significance and determinability. A valuable source of information on numerous aspects of marine ecology is the text by Moore (1958). Not all morphological characters will be equally susceptible to environmental effects and the multivariate statistical analysis will need careful study in order that the ecologically stable characters and those that are reactive to the environment can be recognized.

Ecophenotypic variation

Populations of a single species occurring in different habitats in the same region are often visibly different. Such variants are sometimes referred to by ecologists as micro-subspecies or ecological races where a genetic background is suspected. Alternatively, they may be interpreted as nongenetic ecophenotypes, this type being particularly common in plastic species, such as some molluscs.

Ecophenotypic variation is common among euryhaline ostracods, and a case is discussed in the next section.

Mayr (1969, p. 415) has reviewed the importance of habitat characteristics. Species living in a relatively uniform and stable environment are selected primarily for stability and adaptation to this environment. This is often a reasonable classification for many organisms inhabiting a normal marine environment (such an environment can be considered ecologically buffered). Species inhabiting an unstable environment will be strongly selected for genetic flexibility which, in turn, will be less the greater the phenotypic flexibility. An example of the latter case seems to be represented by the Moroccan Cretaceous ostracods discussed in Chapter 5.

Salinity

As an introduction to the topic of ecological factors, we shall consider size and shape shifts, due to salinity differences, in the length of carapace of the living species of ostracods, *Loxoconcha impressa*, observed on material sampled from the River Tamar in southern England by Barker (1963) and restudied by Reyment (1971). Barker's material demonstrates that there is clear association between both of the variables salinity and distance from the mouth of the river, on the one hand, and length of the carapace, on the other. The results show that length of the carapace is

positively correlated with salinity such that the longest shells occur, on the average, near the mouth of the river. Incidentally, this contrasts with the findings of Keen (1971, p. 534) that size and salinity are not correlated in material of *Hemicyprideis montosa* (Jones and Sherbon) studied by him.

The changes in length of the carapace also affect the shape as this change is usually relatively greater than that for height. Barker (1963) also studied Jurassic material of *Fabanella* in the same connection and concluded that his sequence showed evidence of salinity-dependent size changes.

Kilenyi (1969, 1971) found salinity to be an important ecological factor in his study of the ostracods of the Thames estuary. Kinne (1971) pointed out that salinity tolerance is sometimes tied to genetic tolerance and that evolutionary conservatism may go hand-in-hand with this relationship. He also noted (Kinne, 1971, p. 975) that there may be definite structural responses to salinity, Firstly, size can be inhibited by both sub- and supranormal salinities. This reduction in terminal body size is sometimes referred to as "pauperization", a condition well known to micropalaeontologists.

External structures are also subject to salinity-controlled variations; for example, the degree of calcification of the hard parts, meristic characters and the shape of the body. In some brackish water crustaceans, the number of spines may increase as salinity decreases. This is a feature of potential importance in ostracods.

As an example of the relationship between the shape of the body of a crustacean and salinity we can consider the brine shrimp, *Artemia salina*. The canonical variate analysis of shape variation in the first and second canonical variate means for 14 samples of this branchiopod in relation to salinities of 35‰ and 140‰ is illustrated in Fig. 3. This analysis, based on various publications by Gilchrist, was presented in detail in Reyment (1966b). It will be seen that there is both a geographical (thence genetical) and ecological trend in this data. The pronounced effects of salinity on shape variation are apparent along the second canonical axis (cf. Chapter 3).

Temperature

It is generally conceded by marine ecologists that temperature is the most important single ecological variable (Moore, 1958, p. 17). Just as in the case of salinity, some animals (for example, *Ostrea*) have an optimum temperature, above or below which growth decreases. Kinne (1970) observed that temperature is well known to affect the growth rates of small marine invertebrates. Another well-known observation is that many
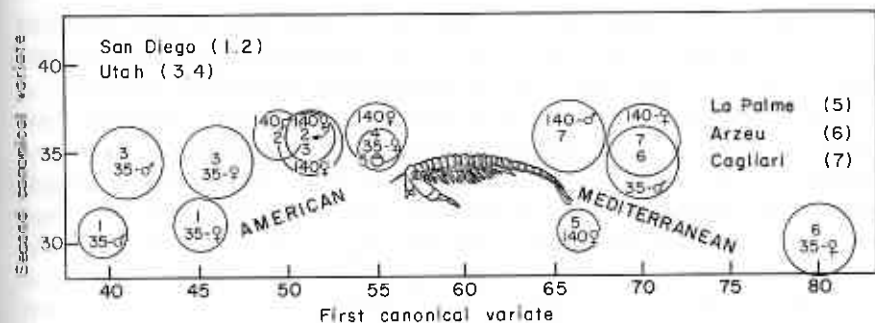
Fig. 3 Example of the relationship between the shape of the body of the brine shrimp *Artemia salina* and salinity. Redrawn from Reyment (1966). The salinities are marked in the confidence circles (35‰, respectively, 140‰). The first canonical variate reflects geographical variation in morphology, the second canonical variate indicates shape variation. The numbers in brackets after the localities identify the samples.

organisms attain a larger terminal size in colder seas than in warmer waters, a result of the foregoing (this may be confounded by the attainment of sexual maturity as in planktonic foraminifers). This also applies to the sizes reached in vertically temperature-differentiated waters. Kinne (1970) also recorded that autumn and winter generations of some protozoans have shorter spines than summer generations. Moore (1958, p. 61) pointed out that there are many exceptions to the correlation of larger terminal size and colder water. The same writer (Moore, 1958, p. 122) drew attention to the fact that salinity tolerance of some marine organisms can vary from place to place.

On a larger scale, it is well to remember that a greater deal of the material studied by micropalaeontologists derives from sediments of near-shore origin. It is therefore important to bear in mind the effects of intense rainy seasons and the salinity fluctuations caused by rain in estuaries, bays and shallow epicontinental seas and connected temperature fluctuations. Moore (1958, pp. 44, 405) noted the multiplicative effect of temperature and salinity together; a lower salinity may be tolerated by some species of crustaceans where the temperature is higher than normal. The question of interaction of two or more ecological factors on an organism has not been investigated to any extent, but it is a field worthy of much research.

Another property of the temperature variable concerns the availability of oxygen, as the relationship between the solubility of a gas and temperature is a critical one. Salinity has a much less important effect on gases in sea water. It is well known that organisms vary in their ability to

survive in oxygen-depleted environments (Moore, 1958, p. 84). Reyment and Brännström (1962) and Reyment and Hallberg (1967) have analysed the effects of oxygen depletion on a species of fresh-water ostracods.

Schopf and Dutton (1976) took up several matters of interest in their study of variation in a species of bryozoans. They noted (Schopf and Dutton, 1976, p. 257) that the gene frequency at a given locality is significantly correlated with temperature at that locality at the warmest time of the year. In the same paper, they recorded that the length of the avicularium changes systematically with temperature or, possibly, with geographical location in some manner or other. They observed, however, that the cline in the length of the avicularium could mainly reflect genetic differentiation and is not ecophenetic. Malmgren and Kennett (1978) have confirmed that the size of the test of *Globigerina bulloides* is correlated with variations in temperature such that lesser mean diameters develop during warmer episodes. They suggest that this relationship can be used as a quantitative palaeoclimatological indicator for the Quaternary.

Properties of the sediment

Intuitively, the properties of the sediment inhabited by benthic organisms can be expected to be of decisive importance in determining the variability of these organisms. Again, we are faced with a troublesome lack of detailed information on the subject for animals of micropalaeontological interest. Kilenyi (1969, 1971) has reported on aspects of the relationships between estuarine ostracods of the Thames and some properties of the substrate. He found fluid mud to be a prohibitive environment as also was black mud. There is a dearth of ostracods in grain sizes ranging from 0.25 to 0.5 mm, although foraminifers are often abundant. Moore (1958) underlined the importance of the interstitial environment and the properties of the sediment (see also Reyment, 1969). Fine sediment may be rich in organic matter, thus favouring detritus-feeders, but also eventually leading to a low value of the redox potential. The interstitial water of fine, organic-rich sediments is less readily renewed and more rapidly depleted in oxygen, than other sedimentary types. Hallberg (1973) has made a detailed study of these questions, although with particular emphasis on the microbiological implications.

Pirson (1977) devoted a large part of his book on log interpretation to demonstrating how physical logs ("Schlumberger logs") can be used for extracting information on the palaeoenvironment from sediments.

As far as I am aware, the most detailed study of the relationships

between sedimentary texture and ostracods has been done by Williams (1969, p. 326). He studied "shell sand" in particular and found that many ostracods are able to find abundant living space in the interstices of this sediment. One of the problems he brought to light was whether ostracods require a continuous system of interstitial spaces to be able to utilize the system or whether they are able to bustle their way from place to place, forcing local barriers. The evolutionary significance of the interstitial environment has been reviewed by Gould (1977, p. 334).

Other factors

I shall only make brief mention of a few other factors of interest in the present connection. Firstly, there is *depth*, undoubtedly of exceptional significance, as amply shown by modern studies on living foraminifers and ostracods. For recent accounts of such research I can do no better than refer you to the pages of the journal *Micropaleontology* and to the publications deriving from the international meetings of the ostracodologists (for example, Neale, 1969; Oertli, 1971; Swain, 1975).

As regards the role of trace elements, this is a further example of an area needing much research. It is of importance to micropalaeontologists as microelements are one of the few environmental components on which we are able to obtain reasonably consistent and accurate measurements. Moore (1958, p. 93) presented such evidence as he could muster for showing that copper, vanadium and beryllium are important components for some organisms. To this we can add zinc for foraminifers. Valentine (1973, p. 214) summarized observations on the relationships between magnesium and temperature and between strontium, temperature and salinity. More examples can be culled from articles dispersed throughout the pages of the journal *Geochimica et Cosmochimica Acta*. A concerted attack on the problem is still outstanding. An example of how the relationship between the chemical properties of the host sediment and the abundance of organisms in a borehole sequence can be used for making a log is given as one of the case histories of this text.

Interactions between organisms are an important component of living ecosystems. For the most part, palaeoecological studies are unable to come to grips with this vital factor with the exception of one kind of interaction—fortunately, the most important of them as regards the effect this has on population densities. This is the case of predation of one species on another. A careful study of ostracod samples can usually yield accurate information on naticid and muricid predation, particular if the work is done with the aid of the scanning electron microscope (SEM).

Sessile organisms are often adversely affected by crowding, a factor that may be significant for foraminifers. The pre-washing examination of cores might be a way of obtaining diagnostic information in this respect. The pH of the original environment was almost certainly not a firstline factor, as has been indicated by chemical oceanographers, who have proved the strong buffering effect of aluminosilicates (Sillén, 1963).

Although not an ecological factor, but rather palaeoecological, Kilenyi (1969) and Reyment (1960) have pointed out the importance of post-mortem transport of microfossils. Again, with the aid of the SEM, a careful watch should be kept on micropalaeontological material being used for logging work for evidence of transport and re-working. Statistical methods are useful for disclosing unnatural relationships between juveniles and adults such as could result from re-working.

There is no reason to expect that a cohort of related species must react in the same phenotypical manner to changes in some environmental component. Work on European terrestrial snails and their genetic reactions to climatic factors has shown that a rise in temperature will cause some species of a genus in an area to become larger, on the average, while others may become smaller under the same conditions, and still others may remain unaffected (see p. 33). The phenotype of an organism is necessarily a result of the interaction of a genotype with the environment. (This can be highly species dependent.) This fact is widely exploited in animal husbandry in experiments devised to test the relative influence of genotypic and environmental factors on domestic animals. In agricultural research, a commonly employed technique is to make replicated plantings that reduce environmental perturbances.

Experimental studies in palaeoecology

To illustrate this section, I shall consider a study by Reyment and Brännström (1962) on the reaction of the freshwater ostracod species *Cypridopsis vidua* to environmental stresses. This is a common species in the northern hemisphere. For ostracods, it is possible to gain valuable empirical information from laboratory studies of clones of parthenogenetic species, such as the one considered here. For such experiments to be successful and to have palaeoecological significance, the environment must be rigidly controlled and measured. Among the many results of interest yielded by the study of *Cypridopsis vidua* may be mentioned those concerning variations in size and shape. It was found that for the environments marked by severe oxygen deficiency and by an excess of

calcium carbonate, there were measurable, though small, differences in the volume of the carapace. The positive allometric relationships between length and height and breadth of the carapace could be shown to be influenced by the environment.

It has often been demonstrated for the plant kingdom that different "races" of a species of plants readily arise from a clone, individuals of which have been subjected to different environmental conditions. Thus, plants with a uniform heredity can look very different when grown in different environments. The range of phenotypes deriving from the interplay between a given genotype and a set of environments constitutes the "norm" or "range of reaction" of that genotype. The results for *Cypridopsis* are probably analogous to those for the plants. Some ostracods give a more spectacular reaction than the species referred to here in that they may develop strong ornamental variations in answer to variations in some ecological component, often salinity.

Can the palaeoenvironment be measured physically?

I have suggested recently (Reyment, 1978b) that the self potential, *SP*, and normal resistivity logs of physical borehole logging methods ("Schlumberger logs") can be used as an indirect quantification of the fossil sedimentary environment (see p. 158). This conclusion was arrived at without knowledge of the results based on very extensive practical experience of Pirson (1977). Pirson (*op. cit.*), in his textbook on well-logging methods, has developed the palaeoenvironmental use of physical logs in an interesting and well-documented manner.

He also shows that the redox potential is an important component of the self-potential curve. Unfortunately, redox logs do not seem to be run very often in most exploratory work and the self-potential curve is the nearest one can come to a reasonably exact measure of the reduction and oxidation properties of the sediment. In the examples presented in the present book, the resistivity curve is mostly used. It closely agrees with the *SP* curve in the material concerned.

Pirson (1977, p. 12) demonstrates how regressive and transgressive sequences can, under favourable circumstances, be charted by the *SP* log, these environments being characterized in part by particular sets of reduction and oxidation potentials. My experience tends to show that as long as the sediment remains reasonably homogeneous, the electrical logs may be useful palaeoecological indicators. When heterogeneities occur, however, these logs do not seem to be relatable to environmental conditions.

The *SP* log has proved to be a suitable tool for mapping the environment of sedimentation as it expresses the geophysicochemical differences in sediments laid down in different environments. As noted, the more direct measure is the redox log but it is not generally available. Under certain conditions, the *SP* curve acts as a subdued redox potential curve. Its role as an ecological measure derives from its reaction to the properties of various minerals, thence the sedimentational environment.

If *Es* denotes the diffusional potential of porous formations and *Eh* the redox potential within formations, the self potential *SP* can be written as a simple additive relationship (Pirson, 1977, p. 18)

$$SP = Es + Eh$$

This indicates that the self potential "contains" the redox potential but that the correspondence between the two will be influenced by the magnitude of *Es*. In a homogeneous rock sequence, *Es* will vary but slightly and the correspondence between *SP* and *Eh* will tend to be uncomplicated.

Sedimentational characteristics, such as cyclicity, can be recognized by curve shapes taken by *SP*, short normal, induction and conductivity curves. The palaeoecological applicability of such properties remains to be fully exploited.

Reliability of ecological observations on borehole samples

Some of the discussions presented in this book (as well as "in the literature") presuppose, often tacitly, that chemical and other observations made on borehole cores may be interpreted in a manner such that the analytical data reflect more or less exactly the environmental conditions that prevailed at the time at which the fossils enclosed within the sediment were living animals. This can be a dangerous assumption to make and before one is tempted to use it, there should be a careful appraisal of the material involved.

Firstly, there is the question of bioturbation and the effect that burrowing organisms have on the primary bedding of sediments, a point which has been well brought out by Schäfer (1962). Recent sediment cores often show a transition from bedded sediment at depth to bioturbated, non-bedded sediment towards the top of the core. Under the microscope, the uppermost layers of sediment can be seen to consist largely of faecal pellets, detritus of various kinds, tracks, etc., but the further down the

core one comes, the less pronounced are the visual effects of the reworking activities of organisms. My observations are based on conditions pertaining in near-shore cores I have taken in the Niger Delta and off the Ivory Coast.

Secondly, there is the question of early diagenetic subsurface chemical reactions, which may cause radical changes in the geochemical properties of fossil sediments. Some interesting aspects of this problem can be obtained from an analysis of data published by Gadel and Mongin (1973). These workers studied a sedimentary core sample taken from the Mediterranean continental shelf of France. They actually interpreted the changes as being due to a gradual increase in salinity of the Mediterranean since the Late Pleistocene, an interpretation that could be supported by the faunal composition of the sediments. Similar changes can, however, take place in sediments as a result of chemical reactions, unrelated to strong fluctuations in salinity. Gadel and Mongin (1973) made 25 observations along the core on the elements Cl, Mg, Ca, Na, K, Fe^{2+} and Mn, listing their determinations in that order. Depth displays highly significant negative correlations with all elements except manganese and ferrous iron. There are strong positive correlations between the elements Mg, Cl, Na and K. The first three eigenvalues and eigenvectors of the correlation matrix for all variables are listed in Table II. The first eigenvector may be interpreted as a "depth vector", in which the negative relationship of five of the variables with depth is brought out, which is the most pronounced feature of the analysis. Ferrous iron dominates the

TABLE II

Eigenvalues and eigenvectors for the French sedimentary core

Eigenvalues as percentages of trace	1 75.3	2 11.6	3 8.6
Variables	Eigenvectors		
	1	2	3
Depth	0.38	0.24	0.22
Cl	−0.40	−0.03	−0.06
Mg	−0.40	−0.07	−0.13
Ca	−0.35	−0.01	0.48
Na	−0.39	0.00	−0.13
K	−0.38	−0.19	−0.33
Fe^{2+}	−0.15	0.93	−0.29
Mn	0.31	−0.21	−0.70

second eigenvector. The plot of the first two sets of transformed observations (not shown) brings out clearly the depth control of the elements, while the plot of the second and third sets of transformed observations shows that the uppermost samples form a distinct cluster (samples from surface to 5 m). The samples from between 5 m and 19 m are distributed without any obvious pattern to their locations.

What can we conclude from this short analysis? It is clear that the chemical properties of the uppermost samples differ from those of the early samples. There is, however, no gradient, at least, as far as can be identified by the rather uncomplicated method used here. This would tend to belie the suggestion that the differentiation is due to an ecological trend and perhaps support the contention that the differences in chemical properties might possibly be ascribable to early diagenetic changes in the sediment.

I have presented this example as a warning that one should keep such possibilities well in mind when interpreting quantitative analyses on borehole logs. In order to give the problem posed here full treatment, it would, however, be necessary to carry out a large-scale statistical analysis in which the various patterns of variation were tracked down in the hope that genuine ecochemical properties of the sediment could be separated from those due to post-depositional chemical reactions.

Palaeogenetics and borehole sequences

Introductory comments

We are now in a position to begin an examination of the possibilities of using fluctuations in the morphology of shells as a means of creating borehole logs, *biologs*, as it were. It should be noted that although the discussion in this book is couched in terms of borehole sequences, the remarks and conclusions apply equally as well to natural outcrop sequences where the time relationships are well understood.

In any sequence of fossil organisms, some of the changes in form witnessed will be environmentally determined and some will be genetically controlled. There can be no clear interface between the two and all changes will, in some manner or other, ultimately have their origins in the genetic constitution of the organism. These are complicated factors and we shall discuss some of them in detail soon.

For the purposes of making first-order models, it is often convenient to

think in terms of splitting the morphological variability into morphological *chronoclines*, operating side by side, one representing changes in size due to ecological factors, the other being due to morphological manifestations of random genetic variation. This might seem to be a very artificial and mechanical approach to the question, but it is one that practical experience has shown to function well in many types of palaeontological problems as a good working hypothesis. Until recently, genetical work had not seemed to find any use for the eigen-analysis of variation in plants and animals. However, Brussard (1975) has successfully applied what is essentially a palaeontological technique to the study of shell polymorphism in European land snails and Leamy (1977) investigated phenotypic correlations in mice by principal components analysis. This latter study is of immediate palaeogenetical interest as he compared and contrasted the results obtained by eigen-analysis with those found from the usual methods of identifying sources of variation in genetical experiments by the analysis of variance. He found that the phenotypic correlation is influenced by both genetic and environmental sources of variation, as postulated in the palaeogenetical first-order model (see Chapter 4). Cook and O'Donald (1971, p. 93) have reported significant results of a similar nature for *Cepaea*. Cervantes *et al.* (1978) obtained similar results to those of Leamy (1977), using principal components analysis on maize.

Interpreting vertical oscillations

We have already concerned ourselves with the role of allopatric speciation in the formation of species and in the vertical variation of morphological characters. The allopatric model seems to provide a means of expressing the variability we observe in some micropalaeontological samples below the species level. We need also to take a further look at the more classical palaeontological view of gradual evolutionary change, particularly from the aspect of vertical variations in discrete characters. As noted earlier, a well-documented study of this topic is that of Hayami (1973) on evolution in the Neogene to Recent pelecypod, *Cryptopecten vesiculosus*. In this study, 17 samples of fossil and living individuals of the species were analysed. It was found that in each sample, most quantitative characters, including the number of radial ribs, are normally distributed. Two discrete phena were found in the late Pleistocene to Recent material, these being readily distinguishable by their separate ornamental patterns. These two variants were designated as phenotypes R and Q by Hayami; the oldest material was found to belong entirely to phenotype Q, the youngest to phenotype R. We have, therefore, before us what appears to be a clear unidirectional shift in a morphological trait.

Viewing the variation over time as genetic, non-sexual and discontinuous, the trend in the morphology of the character under study is interpretable in terms of the change in the relative abundance of the phenotypes. The relative increase in the abundance of phenotype R is considered to be a significant trend in the evolutionary history of the pectinid. The shift in frequencies in this feature could be shown to be correlated with a shift in the overall size of the shell such that the dimensions of the shell display a tendency to decrease, seemingly independently of the environment. Hayami (1973) thinks that this shift is due to natural selection. The trend in frequencies towards an increasing abundance of phenotype R could be taken as indicating the higher adaptive value of this phenotype over phenotype Q. Hayami inferred phenotype R to have the genotypes AA and Aa while phenotype Q represents the genotype aa.

Taking the adaptive value of AA arbitrarily as one and presuming complete dominance selection against aa, and the adaptive value of aa as $1-s$, where s is the coefficient of natural selection, Hayami investigated the effects of shifts in frequencies. Using a late Pleistocene sample and a Recent sample of the pelecypod, the frequencies of the two phenotypes were calculated by means of the usual formula

$$\Delta q = \frac{sq(1-q)^2}{1-s(1-q)^2} \tag{2}$$

Here, q denotes the frequency of the gene A and s was obtained from eqns (3) and (4) below

$$s = \frac{z_t - z_0}{t} \tag{3}$$

with t, time, estimated by some suitable means (Hayami used radiometric dates). The required value of z is obtained from the equation

$$z = \ln\left(\frac{q}{1-q}\right) + \frac{1}{1-q} \tag{4}$$

The order of doing the calculations is that one begins by finding a value of z_t for the upper time-point and a value of z_0 for the lower time-point by means of eqn (4). An assumption has to be made here, namely that s is constant. One then works back through eqns (3) and (2) to produce an estimate of Δq.

Hayami's example was in terms of a geologically short period of time, estimated by him to be about 70 000 years. He concluded that the unidirectional shift in frequencies of the genotypes was due to the

accumulation of a mutant gene by natural selection acting differentially on the genotype.

In micropalaeontological work, the study of variations of the kind expressed in *Cryptopecten* is of great significance for ostracods. Ostracods are prone to show various kinds of polymorphisms and we shall discuss some examples of these in the next section.

Polymorphism in ostracods

Reyment (1963, 1966) described many examples of discrete variation in Paleocene ostracods from Western Nigeria. Some of these are simple, with only one shell character being visibily affected. Others are more complicated with two or more ornamental characters simultaneously involved. These examples seem to be genuine cases of pleiotropism in ostracods (see section on pleiotropism in Sinnott et al., 1958).

In two of the cases studied, both involving Paleocene species of the buntoniid subgenus *Protobuntonia*, there seems to have been a tendency for a mutation to occur which caused the appearance of a posterior spine during advanced ontogeny. In both of the species of this subgenus, a unidirectional shift in the frequencies of the phenotypes took place which led to the early loss of the spinose variant, possibly by selection acting differentially on the genotype, as in the *Cryptopecten* example. Using formulae (2), (3) and (4), and some approximations, a change of gene frequency of $\Delta q = 1.05 \times 10^{-5}$ was obtained, a high value, and much higher than for *Cryptopecten*. The *Protobuntonia* come from a borehole, and the problem of allowing for geographical variation does not arise (a weakness of the *Cryptopecten* material).

The same borehole also yielded material of *Buntonia sensu strictu*, a contenporary of the *Protobuntonia*. The posterior spinosity is not lost in this genus and the evolutionary significnce of the loss of the posterior spine in *Protobuntonia* is not obvious.

Reyment and Van Valen (1969) studied polymorphism in Nigerian fossil and living ostracods. They found, as for Hayami's *Cryptopecten*, that all polymorphisms seem to have equal frequencies in the two sexes. In the Paleocene species *Leguminocythereis lagaghiroboensis* Apostolescu there is polymorphism for a pleiotropic character. Individuals with irregular lateral ornament (manifested largely as breached reticulations) have a broadly rounded posterior, while those with a regularly ornamented lateral surface have a mucronate, or very sharply rounded to bluntly pointed posterior.

A long-term fluctuation in the frequency of a phenotype (mainly marked by a posterior spine) was found for the buntoniid ostracod

Buntonia (*Quasibuntonia*) *livida* Apostolescu. In this species, there seems to have been a tendency for the spinose phenotype to become relatively less common; as to be expected, the frequencies show stochastic fluctuations superimposed on the main trend.

Polymorphism in the lateral ornament seems to be of common occurrence in many groups of ostracods. In addition to the two genera discussed above, such ornamental variations occur in Nigerian Paleocene and Eocene *Ruggieria* and *Cytherella*. Among the more usual pairs of variants, the following pairs are worth mentioning: pitted–unpitted, reticulated regularly–irregularly reticulated, smooth–reticulated. A comparative study on a living Nigerian *Buntonia* not only disclosed the existence of polymorphisms analogous to those found among the Paleocene ostracods, but also much more complex variational patterns in the anterior denticulation (Reyment and Van Valen, 1969).

When it comes to obtaining guidance from the recent literature on the significance of polymorphisms, it is hard to avoid a feeling of disappointment, as the subject has not been given the attention one would like to see for marine organisms. The most useful work is being done on the land snail *Cepaea*, and it is quite possible that some of the results obtained for this gastropod may be applicable to micropalaeontological situations.

Ford (1964, p. 150) has summarized earlier work on *Cepaea*. The five colour bands showing polymorphism seem to be analogous to the anteromarginal spines of ostracods. Even the earlier publications record the opinion that the types of banding developed are correlated with environmental factors. Frequencies are also greatly affected by the predation of birds. Bantock and Price (1975) thought variations in temperature could be important in influencing the frequencies of morphs. Jones and Parkin (1977) noted the connection between the shell polymorphism of *Cepaea vindobonensis* and geographical location in a mountainous terrain. They related this to climatic variation, hence temperature fluctuations. This takes us back to an earlier part of this chapter in which evidence was summarized that points to temperature being perhaps the most important single ecological factor.

Clarke and Murray (1971) did not find frequencies in shell-colour morphs in a Polynesian land snail to be bound to any of a large number of environmental factors. It should perhaps be noted that their data were not analysed by statistical methods relevant to the material. Cook and O'Donald (1971) observed significant size differences to occur between the morphs in one colony of Irish *Cepaea*. Quite small differences in mean breadth were thought to lead to appreciable variation in the capacity to survive.

The influences operating upon the phenotype through the developmental system are complex. Rendel (1967) put forward a method of studying the problem which allows one to measure the resultant and observe changes in it without having to concern oneself with more than the phenotype. He defined his concept of Make by recognizing as an entity the resultant of an incompletely known complex of influences, genetic and environmental, contributing to the formation of a phenotype.

Discrete variations, such as are displayed by ostracods with respect to marginal spines, may be conveniently analysed in terms of Rendel's concept, provided that good ontogenetic series are available.

Demic variation

There are several so-called rules deriving from the early days of biogeographical work of zoologists. I mention these here as they sometimes occur in palaeobiogeographical discussions involving shifts in size. Clines may involve adaptive changes. Bergmann's rule states that in warm-blooded animals, the body size increases with decrease in average temperature. Allen's rule states that exposed portions of the body of a warm-blooded animal decrease in size with decrease of average temperature (cf. Moody, 1964).

A *deme* is defined as the smallest unit of population. Groups of genetically similar individuals bearing an intimate temporal and spatial relationship to each other form a deme. Each deme is isolated to some extent but there is gene-flow between demes. A well-known example of demic variation showing correlation with the environment is provided by the Pacific herring, *Clupea pullasi*, demes of which differ in the number of vertebrae and the number of fin-rays along a "character cline" from north to south along the North American coast. This variation appears to be correlated with temperature. Demic variation can be a source of confounding in an evolutionary study as geographical variations may cloud the picture.

Lewontin (1974, p. 249) reports on several such examples. This book is a mine of useful information on the genetical concepts underlying variability and there is also an interesting set of observations on *Cepaea* and its polymorphisms (*op. cit.*, pp. 234–235). The same author has also edited an important symposium volume (Lewontin, 1968) on population biology and evolution which forms a useful starting point for further study of questions of significance for the analysis of sequences of fossils. The article by Rendel on p. 47 of that book is particularly stimulating.

Spectral analysis of fossil data

In the foregoing pages, several statistical and biological ideas have been put before you which, when viewed as a whole, provide a reasonably broad basis for the development of the main methods used in this book for charting and exploiting vertical fluctuations in the morphology of sequences of fossil species. One of the central concepts for the analysis of meristic characters, i.e. developmental thresholds, is due to Rendel (1967). Another important concept is that of the spectral decomposition of variability by means of eigen-analysis (cf. Chapter 3). This is to be understood as an approximation of only the first order. Nevertheless, the work of geneticists has of late shown that the principal components solution used by zoologists and palaeontologists does appear to fit in with the theory (Brussard, 1975; Leamy, 1977). Statistically, there are still many aspects of this method requiring elucidation; some of these were considered by Jöreskog *et al.* (1976).

A general model for the partition of morphological variance in p variables may be expressed as a vector equation, to wit,

$$x = t + u + e \tag{5}$$

where x is an observational vector containing p components, t is a vector representing size variation of the phenotype resulting from the main interaction between the phenotype and the environment, u is the supposed variation in size of the elements of x ascribable to the genotype alone and only secondarily under the influence of ecological factors and e is the residual variation due to non-biological causes, such as errors in measurement, incorrect taxonomic identifications of some specimens, inaccuracies in the model, deformation of specimens and the like.

As already noted, a useful, though approximate, way of approaching the problem is offered by the method or principal components, as was originally suggested by Jolicoeur and Mosimann (1960). Jöreskog *et al.* (1976) have considered at length questions surrounding this model and it is suggested that you might like to read this discussion (*op. cit.*, Chapters 3 and 4) if you are uncertain about the details of the method and how it is distinguished from similar statistical procedures.

Example of the spectral decomposition of variation

The best way to introduce the idea of the breakdown of variation is by a simple example. Note that this is not a self-contained study and the results presented in Table III are not to be given far-reaching significance.

TABLE III
Eigenvalues and eigenvectors for *Cytherella*

IIIA. Matrix of sums-of-squares and cross-products (covariance matrix multiplied by 17 degrees of freedom) for three measures on the carapace.

	Length	Height	Breadth
Length	0.0114	0.0089	0.0110
Height		0.0183	0.0119
Breadth			0.0146

IIIB. Eigenvalues and eigenvectors of the covariance matrix

	Eigenvalue	Percentage of variance	Eigenvector length	height	breadth
1	0.0021368	82.00	0.49	0.64	0.60
2	0.0003679	14.12	0.55	−0.76	0.36
3	0.0001011	3.88	0.68	0.15	−0.72

N.B. The calculations were made on the raw data.

The method of multivariate statistical analysis used in the present connection is briefly noted in Chapter 3. Jöreskog *et al.* (1976) have devoted an entire book to the subject of the multivariate analysis of single samples, with particular emphasis on applications in the earth sciences. We shall consider the analysis of a trivariate covariance matrix for the variables length, height and breadth of the carapace of a Paleocene ostracod, *Cytherella sylvesterbradleyi* Reyment from western Nigeria (Ogun State). The essential calculations are listed in Table III. No transformations have been applied to the data.

The first eigenvalue accounts for a very large portion of the total variance in the material, as much as 82%. The second eigenvalue amounts to 14.12%, while the third eigenvalue only comes to 3.88% of the total variance (Table III). All the elements of the first eigenvector are positive (Table III) and they are roughly in the proportion of $5:6:6$, indicating almost equal participation of all three variables. The conclusion one might be led to here is that the variation in the three variables resides mostly in fluctuations in the size of the carapace or, more specifically, in its volume. Thus if a carapace were to increase in length by an amount of 0.01 mm, there would be, on the average, a simultaneous increase in the dimensions of height and breadth of about 0.012 mm.

Let us now examine the second eigenvector (Table III). This is approximately composed of elements in the proportion of $6 : -8 : 4$, the height of the carapace being negatively loaded in relation to length and breadth. This second kind of "bipolar" association is often interpreted as an indicator of shape variation in which some elements grow larger as others grow smaller at the same time. The reason for this can be demonstrated here by seeing what would happen if the length variable were to be increased by 0.01 mm. This would lead to an increase of 0.007 mm in the breadth and a decrease of 0.013 mm in height. This kind of shape variation differs from that inherent in the first eigenvector where there is also a shape component in that when the carapace grows larger, all variables participate in the increment, but there will be a progressively more important distortion owing to the greater increases for height and breadth in relation to the length. For the second eigenvalue, a convincing case may be made for the interpretation that some 14% of the total variability resides in individuals that are somewhat longer and more inflated than the majority but which are, at the same time, more depressed. This could be a manifestation of sexual dimorphism in the carapace. The variability residing in the three original variables has been partitioned into a spectrum of new variables, each a composite of the three carapace dimensions. The concept just illustrated derives originally, in part at least, from the zoologist Teissier, but the main elaboration of the concept is due to Jolicoeur (1963). Jolicoeur's interpretation is intuitively attractive and it seems to be supported by the recent work of geneticists. By using the logarithms of the original observations, Jolicoeur (1963) tried to interpret the first eigenvector as a generalized equation of allometry. Hopkins (1966) capitalized on this idea by putting forward a plausible factor-analytical model, whereby the first vector of factor loadings of the covariance matrix of logarithmically transformed variables was suggested as forming generalized coefficients of allometry.

Mosimann (1970) has attempted to approach the size–shape problem by means of ratios (Penrose ratios). Although these results are of considerable theoretical interest, the approximation given by the original Jolicoeur–Mosimann (1960) method is sufficient for our present needs. Bookstein (1978) has given the subject a monographic treatment.

3
Outline of Some Multivariate Statistical Methods

In this chapter, I give short introductions to some of the methods used in the subsequent text. This is not a full coverage of the subject nor does it pretend to tell you how to do the analyses. To learn this, if you do not already have the knowledge, I refer you to standard textbooks in the field. Other methods will be described as they appear in the text. For example, questions of redundancy and stability in multivariate work and the special problem of growth invariance, which is an outgrowth of canonical variate analysis and generalized distances and is best put before you in connexion with an actual example. The multivariate statistical methods of most use in the present connection are those concerned with the simultaneous analysis of several samples.

The basic matrices

The *data matrix* is the matrix formed by N observations (measures of N specimens) on the p variables. These are the N observational vectors of the sample. In this text, the methods are discussed with respect to the

observational vector, x. Each row of X represents a specimen on which p characteristics have been measured.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix}$$

The *mean vector*, denoted here as $\bar{x}$, is the p-variate vector of means of X. Thus, if one sums each column of the data matrix and divides the vector of sums by N, the mean vector results.

The *covariance matrix* is the $p \times p$ symmetric matrix of variances and covariances with the elements

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

The *correlation matrix* is defined as the correlations between the p variables, with each element constructed from S by calculating

$$r_{ij} = \frac{s_{ij}}{\sqrt{(s_{ii} \times s_{jj})}}$$

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

In this very brief presentation of the most common fundamental matrices used here, I have used the notation normally reserved for samples. It has become accepted practice in statistics to employ the Greek letters corresponding to the Roman letters used for sample statistics for denoting the counterparts in the statistical population. In a practical text, such as the present one, discussion mostly centres around the sample quantities and there is seldom need to introduce the population statistics into a particular analysis.

In univariate statistical analysis, the variability of a variate is reflected by the *range* of the observations in the sample and it is measured by the *variance*. The sample mean and the sample variance estimate the population quantities; the larger the sample, the nearer one comes to being able to estimate the true population values by those obtained from the sample. There is, therefore, a certain support for wanting to use large samples in a
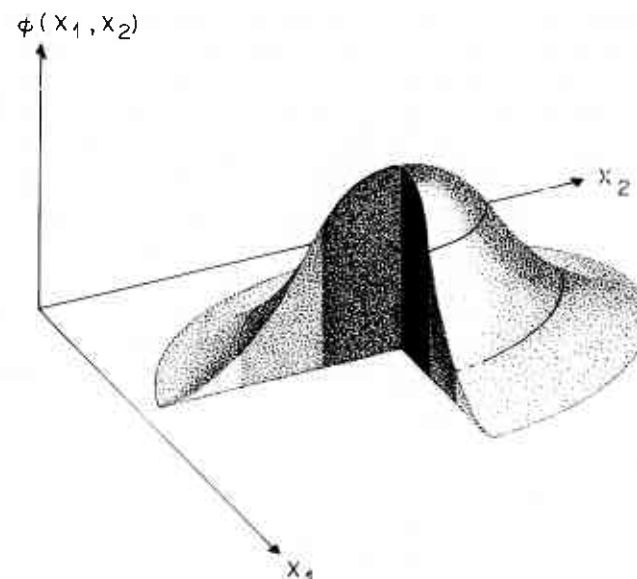
Fig. 4 The bivariate normal distribution. This gives a highly idealized view of the situation for in most cases, the surface will appear more like a maturely dissected volcanic cone.

palaeobiological study. Repeated sampling from a normally distributed population will yield a set of means and variances that cluster around the theoretical values and which themselves are normally distributed.

Passing from one dimension to two or more dimensions brings in the added factor of the covariation between the variables. For two variables, the points plotted for a sample of bivariate normally distributed variables fall within an ellipse, as I have illustrated schematically in Fig. 4. The greatest concentration of points lies in the middle of the ellipse from which they thin out towards its edge. For more than two variables, the shape of the field containing the points is ellipsoidal. For more than three dimensions, one speaks of an hyperellipsoid of scatter.

The scatter ellipsoid

In order to determine the statistical properties of an ellipsoid of scatter, it is sufficient to begin by computing the principal axes of it by some standard method for eigenvalues and eigenvectors. The statistical method for carrying out these operations is known as *principal components*

analysis, a procedure not solely reserved for multivariate normally distributed data, although in this book, the assumption of multivariate normality will always be made. For an account of the concept of principal components, the reader is referred to Jöreskog *et al.* (1976, Chapters 3 and 4). Briefly, and in terms of the sample quantities, the procedure of principal components may be described as follows.

Consider a p-dimensional observational vector $x = (x_1, x_2, \ldots, x_p)$, multivariate normally distributed. The covariance matrix of a sample of N such vectors is S. The eigenvalues of S are $d_1, d_2, \ldots, d_p$ and associated with them are the normalized eigenvectors $f_1, f_2, \ldots, f_p$. The principal components are defined as the linear combinations of these vectors with the x's. Thus, for eigenvector f_i, we should have the relationship

$$y_i = f_{i1}x_1 + f_{i2}x_2 + \ldots + f_{ip}x_p \qquad (6)$$

We note that

$$f_i f'_i = 1 \quad \text{if} \quad i = j$$
$$f_i f'_j = 0 \quad \text{if} \quad i \neq j$$

This is the property of orthogonality for the eigenvectors and it also expresses the fact that these vectors are normalized so that their sums of squares will be one.

The total variation in the material is expressed as the sum of the eigenvalues which is the same as the sum of the original variances. An introductory account of the background of principal components analysis is given by Marriott (1974).

Special morphometrical significance may be attached to the elements of the eigenvectors of principal components analysis (Jolicoeur and Mosimann, 1960; Blackith and Reyment, 1971). Mosimann (1970) has investigated the limitations of this approach and suggested a more general solution.

Two or more scatter ellipsoids

A very useful method for studying the properties of several multivariate samples is that known as canonical variate analysis, biological applications of which are given in Blackith and Reyment (1971) and a data-analytic presentation by Gnanadesikan (1977) as "discriminant coordinates". In a way, this method can be regarded as a generalization of principal components, whereby the considerations applying for one ellipsoid of scatter are

generalized to several ellipsoids. In fact, canonical variate analysis can be interpreted as a two-stage principal component analysis.

In another way, canonical variate analysis can be looked at as being a generalization of the concept of the generalized statistical distance, to which it reduces for two samples, insofar as the canonical root in this case is proportional to the generalized distance and the associated canonical vector is the linear discriminant function.

It is convenient at this juncture to mention briefly the generalized distance, introduced in 1932 by P. C. Mahanalobis. This was one of the first multivariate taxonomical methods developed and one which has been used in a wide variety of problems.

For two populations, one considers the pooled sample covariance matrix S and the sample mean vectors $\bar{x}_1$ and $\bar{x}_2$. The generalized statistical distance is defined as

$$D = \{(\bar{x}_1 - \bar{x}_2)' S^{-1}(\bar{x}_1 - \bar{x}_2)\}^{1/2} \qquad (7)$$

Sometimes it will be found that the component covariance matrices of S are not homogeneous, i.e. they could not have been sampled from the same statistical population. This is not a serious defect for modest differences but it may transpire that the divergencies are so great that some special technique will need to be utilized which takes into account heterogeneity in covariance matrices in producing a generalized statistical distance.

There may be a wide variety of reasons for heterogeneity in covariance matrices and it is recommended that you give the problem careful thought if you encounter it in your research (see Reyment, 1969a; Blackith and Reyment, 1971). The most common cause for serious heterogeneity derives from outliers in one or both of the samples (see p. 11). These may, however, be biologically caused, an example of which is differences in the covariance matrices of carapace measurements on ostracods arising from sexual dimorphism. Reyment (1969b) analysed several examples of this kind of heterogeneity in variances and covariances.

The coefficients of the *linear discriminant function* of R. A. Fisher are obtainable as a half-way step in the computation of the quadratic form in eqn (7). If the vector of linear discriminant function coefficients be denoted as t, the computation is simply

$$t = S^{-1}(\bar{x}_1 - \bar{x}_2) \qquad (8)$$

where $\bar{x}_1$, $\bar{x}_2$ and S have the same meaning as before.

Sometimes *quadratic discriminant functions* are used where the covariance matrices are greatly different. However, as Marriott (1974, p. 39) rightly observes, this is often an unnecessary step, and there are other

ways around the problem (references in Blackith and Reyment, 1971). A disadvantage of quadratic discriminants is that they do not yield interpretable coefficients (the elements of vector t in eqn (8)) in the same way as the linear analogue and the use of this method is mainly justifiable in identification studies.

I shall now take up the question of discriminant functions when there are more than two samples. We have now to consider k samples drawn from k multivariate normal populations, the covariance matrices of which are $S_1, S_2, \ldots, S_k$ statistically equal, and with sample mean vectors $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$.

One does not normally use these matrices directly in the usual suite of computations but rather the sum of the matrices of sums of squares and cross products for each sample (from which the corresponding covariance matrices can be obtained by simple division with $N_i - 1$, $i = 1, \ldots, k$). This matrix, known as the within-groups matrix of sums of squares and cross products, will be designated W. Also required is the matrix B of sums of squares and cross-products of deviations of the means, or between-groups matrix of sums of squares and cross-products.

The required canonical variates are found by solving the equation

$$(B - gW)c = 0 \tag{9}$$

Details of the solution can be obtained from Gnanadesikan (1977) and Marriott (1974). The canonical roots g are yielded by the determinantal equation

$$|B - gW| = 0 \tag{10}$$

Canonical variates is a well-known and well-proven procedure, for which tests of significance are available, and it has enjoyed justifiable popularity in taxonomic work ever since its introduction by Bartlett some 40 years ago. It is basically a method for the graphical display of the relative locations of the sample vectors for the canonical variates, thence Gnanadesikan's term discriminant coordinates.

Although canonical variates is a well-known multivariate statistical method, it is not always successful in obtaining biologically meaningful results in a taxonomical study as regards the interpretability of the coefficients of the canonical vectors. The reason for this may lie with the instability often found in the canonical vectors which, in its turn, is connected with redundant or near-redundant directions of variation. This problem has recently been studied by Campbell (unpublished thesis, Imperial College, London 1979). Owing to the importance of the subject, and the fact that it is not yet available in textbooks, the summary presented in Campbell and Reyment (1978) is reproduced below. (The

canonical roots and the actual ordination are, however, little affected by instability in the coefficients of the canonical vectors.)

Canonical variate analysis of living and fossil organisms on which morphological characters have been measured can sometimes be greatly distorted through overweighting by redundant directions of variation. Instability in the canonical vectors (those of $W^{-1}B$) is often associated with the smallest eigenvalues, particularly if these do not greatly differ from zero.

Instead of working with the within-groups matrix W on n_W degrees of freedom, it is preferable to standardize it to correlational form—i.e., a correlation matrix is made out of it. The between-groups matrix B is scaled similarly. Thus,

$$W^* = S^{-1}WS^{-1} \tag{11}$$

where S is a diagonal matrix whose elements are the square roots of the diagonal elements of W. Consequently,

$$B^* = S^{-1}BS^{-1} \tag{12}$$

The eigenvalues e_i and eigenvectors u_i of W^* are then computed. With the diagonal matrix E formed from the eigenvalues e_i and U made up of the corresponding eigenvectors, $W^* = UEU'$.

The eigenvectors are conveniently scaled by the square root of the corresponding eigenvalue (which produces sphericity). In order to establish the extent of eventual instability in elements of the eigenvectors of $W^{-1}B$, a suitable approach is to proceed by means of shrunken estimators, using a ridge-type technique (see Marquardt and Snee, 1975).

Skrunken estimators are formed by adding shrinkage constants k_i to the eigenvalue e_i before scaling the eigenvectors. Writing $K = \text{diag}(k_1, \ldots, k_p)$, where p is the number of variables, define

$$U^* = U(E + K)^{-1/2} = U^*_{(k_1, \ldots, k_p)}$$

The between-groups matrix in the space of the within-groups principal components is then required. This is computed as

$$G_{(k_1, \ldots, k_p)} = U^{*'}_{(k_1, \ldots, k_p)}B^*U^*_{(k_1, \ldots, k_p)} \tag{13}$$

and set h equal to the ith diagonal element term of G. The ith diagonal element h_i is simply the between-groups sum of squares for the ith principal component. The eigenvalues r of $G_{(0, \ldots, 0)}$ are the usual canonical roots, while the eigenvectors $a^{U'}$ are the standardized canonical vectors for the principal components. The usual canonical vectors c^U are given by

$$c^U = U^*_{(0, \ldots, 0)}a^U$$

Generalized shrunken estimators follow directly from the eigenvectors a^s of $G_{(k_1,\ldots,k_p)}$ using $c^s = U^*_{(k_1,\ldots,k_p)}a^s$. Marked instability in the canonical vectors is associated with small e_p and small diagonal elements of G in (13).

In identifying instability in canonical coefficients, a simple rule to use is to examine the contribution of h_p to the total group separation trace $(W^{-1}B)$, which is just trace $(G_{(0,\ldots,0)}$ or $\sum_{i=1}^{p} h_i$. In situations where p is very small in relation to $|\sum h_i|$, or the corresponding ratio of canonical roots from the usual and shrunken estimator analyses is small, then little loss of discriminatory power will be incurred by excluding the smallest eigenvalue and its eigenvector.

The stability analysis also points out those variables that are contributing little or nothing to the discrimination and these can be eliminated. It transpires that very marked improvements in the interpretability can be achieved by Campbell's method of analysis. This will be demonstrated later on in an example (p. 108).

Relationships between sets of variables

One of the methods I have found useful in preparing variational diagrams for organisms interacting with environmental factors is that of canonical correlation analysis. This has its main value in borehole analysis in allowing a set of morpho-variables to be correlated with a set of environmental variables. It is a well-known procedure and I shall not list the steps here in detail. Explicit accounts are provided by Cooley and Lohnes (1971) and Reyment (1971), the latter paper containing a geological example.

Geometrical relationships between scatter ellipsoids

The relationships between some of the statistical methods briefly reviewed in this chapter are illustrated in Fig. 5. Firstly, we shall take a look at the geometrical interpretation of principal components. Consider the ellipse denoted E_1. The length of the first principal axis of this ellipse AB is proportional to the square root of the first eigenvalue of the bivariate dispersion of variables x and y. The orientation of the line AB is given by its direction cosines. These correspond to the elements of the first
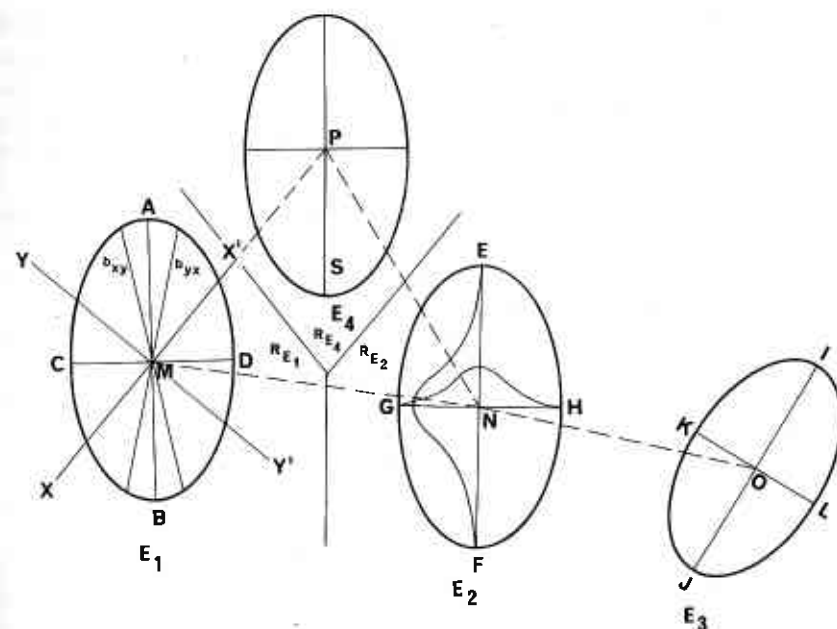
Fig. 5 Diagram illustrating the relationships between some of the multivariate statistical methods reviewed in this chapter. The figure is interpreted on p. 49.

eigenvector of S, the covariance matrix for a set of observations on variables x and y. The second principal axis of E_1, which is the line CD, is likewise interpretable as corresponding to the second eigenvalue and eigenvector of covariance matrix S. Obviously, AB and CD are orthogonal (at right angles to each other).

On ellipse E_1, the original axes XX' and YY' are shown relative to the transformed axes AB and CD in order to demonstrate that the "eigen-operation" produces a rotation of the axes of reference of the data points encompassed by E_1 to new axes, the principal axes of this ellipse. In statistical terms, this rotation involves a transformation from an original set of correlated variables to a new set of uncorrelated variables in the new frame of reference. The relationship between the regression lines to the two variables, b_{xy} and b_{yx}, and the principal axes is also illustrated on ellipse E_1.

The three ellipses E_1, E_2 and E_4 are parallel to each other and of equal size. This indicates that the covariance matrices for each of the samples represented by these ellipses are equal. The lines MP, PN and MN give a proportionate representation of the generalized distances between the

centroids (the means) of the three ellipses. If $MP = PN = MN = 0$, then all three ellipses will coincide and they will have the same centroid. This is the basic problem of the multivariate analysis of variance, which, in the one-way model, tests whether the centroids coincide or not. Ellipse E_2 displays the distributions of the principal component variables along the principal axes EF and GH.

One of the aims of canonical variate analysis is to determine the positions of the centroids, relative to each other, and to display the transformed observational vectors, the canonical variates, on a set of two-dimensional plots, corresponding to the various combinations of the canonical vectors. Additionally, the coefficients of the canonical vectors may be interpreted in biological problems, for example, as indicators of various kinds of variability in morphology. Not all workers are decided on this point. Most canonical variate analyses are made using the covariances and variances directly. A good case can be made, however, for standardizing the covariance matrices to the corresponding correlation matrices. The steps for doing this are supplied on p. 47.

The parallel and equal ellipses E_1 and E_2, with centroids M and N, are separated by a distance MN. This distance is proportional to the generalized statistical distance between centroids M and N. The tangent to the ellipse E_1 at the point of intersection with MN gives the separating line of discriminant analysis (more generally, this will be a hyperplane). The discriminant function is usually taken as the normal or perpendicular line to this.

Specimens deriving from population Π_{E_1}, represented by ellipse E_1, should be identified correctly by the linear discriminant function and vice versa for individuals from population Π_{E_2}. In practice, there is often some degree of overlap and individuals that really belong to one population may be incorrectly assigned to the other. One speaks of the probability of misidentification and the efficiency of the discriminator. The principal reason for difficulties in identification lies with overlapping distributions. That is, the distributions of the two populations of interest are not completely separated from each other. Some idea of how serious this is likely to be can be obtained by testing the significance of the generalized distance by a method known as the Hotelling T^2. This statistic differs from D^2 by a constant obtainable directly from the respective sample sizes. If the T^2-value is highly significant, then there is a good chance that the empirical distributions, and even the theoretical, are clearly separated from each other.

The ellipses E_2 and E_3 differ from each other not only in shape but also in orientation and their first and second principal axes are not parallel. Thus, the calculation of the generalized distance and associated linear

discriminant function will need special methods if the heterogeneity is significantly great.

Consider now the ellipses E_1, E_2, E_4 representing populations having homogeneous covariance matrices. There are three regions of identification for each of the populations Π_{E_1}, Π_{E_2}, and Π_{E_4}, which are denoted as R_{E_1}, R_{E_2} and R_{E_4} in Fig. 5. One may construct linear discriminant functions between all three ellipses for identifying individuals from any of the populations.

If there is no interest in studying the shapes of the sample scatter ellipsoids, a useful simplification of canonical variate analysis can be made by working with concentration hyperspheres.

4

Making a Biolog

In this chapter, we shall concern ourselves with the practical details of making a biolog from sequentially sampled fossils. Usually, the samples will come from boreholes, but there is nothing to preclude the use of good outcrop material. There is also no reason why macrofossils cannot be used in the latter case as long as they are autochthonous. The content of Chapters 2 and 3 provides the necessary basis for the development of this chapter, but certain additional questions are also considered. An important subject is that of the breakdown of variation in samples by the method of principal components and how the results are to be interpreted. The application of the same multivariate statistical method to morphological measurements on foraminifers may disclose components of the life cycle. There is also the problem posed by the need for adjusting for differences due to age-ranges in samples, thence sizes of individuals. This involves the concept of growth-invariant multivariate analysis and it is clearly of importance in the case of the biolog. Another question taken up concerns instability in canonical variate analysis and the influence this may have on the interpretation of a multivariate morphometric analysis.

Somewhat more complicated from the purely practical aspect is the union of observations on the variability of benthic organisms (alternatively, frequencies of benthic organisms) and chemical components of the

palaeo environment. The multivariate method known as canonical correlation analysis is a suitable procedure for treating such data. The practical difficulty is that it may be uncertain whether the organisms involved are *in situ.*

It is naturally very important to be able to correlate between logs from neighbouring boreholes. This topic is considered in the light of a new method for "slotting" borehole observations. Generally, a biolog is most likely to yield superior results in cases where the same few fossil species are found throughout a thick series of lithologically undifferentiated shales. We shall begin by looking at a simple multivariate way of logging on fossil data.

Biolog using a sequence of multivariate means

The simplest kind of biolog is one made by plotting fluctuations in univariate means of a bioseries from sampling level to sampling level in a borehole. This is a rather well-known way of producing a variational curve, although only occasional use seems to have been made of it. The ideas involved are most easily presented in terms of variation in ostracods as this group does not suffer from age–size differences to the same extent as do, for example, foraminifers.

For the purposes of a biostratigraphical logging procedure, a workable first level of approximation is that we assume that major changes in the dimensions of the ostracod carapace (adults only considered) are mainly due to environmental factors. Support for this assumption is available from work reviewed in Chapter 2.

As a consequence of this working hypothesis, the plots of means of the same character measured on several species, and sampled at different stratigraphical levels in the same borehole, can be expected to show the same general pattern of oscillations over time. That is, the plot will display the same overall trend of right and left swings, although the amplitudes will not be very informative, unless consolidated by proper statistical controls. A set of variational curves for a number of species which shows harmony in the pattern of oscillatory swings suggests that all of these species have reacted similarly to the environment.

The essential point here is not that the patterns of the oscillations are near replicas of each other, although this condition does occur. It is that the directions taken by the successive oscillations agree. The reason for this is biological and not statistical. Some organisms react more strongly morphologically than others (even closely related ones) to certain en-

vironmental factors (for example, salinity and temperature) and this may lead to amplitudes of markedly different size. Cytherellid ostracods are less reactive than trachyleberidids, which is evidenced by the biologs shown in Fig. 24.

On the other hand, randomly trending chronological sequences are possibly interpretable as unstressed stochastic variation of genetic origin, largely uncorrelated with changes in the environment. Real environments are uncertain and highly stochastic (May, 1973, p. 109). The input components vary randomly and equilibrium is the average around which the biological system fluctuates. Experimental studies are needed to elucidate this problem. The genetical basis of evolutionary oscillations with respect to gene arrangements has been considered by Dobzhansky (1971). Studying a species of *Drosophila*, he found some changes to be regularly cyclic, following the march of the seasons, and he thought the morphological fluctuations could result from the effects of natural selection.

Biologs from two Nigerian boreholes

I have chosen to illustrate this section with data from two well-studied boreholes in Western Nigeria (Gbekebo I and Araromi I), drilled in Ogun State by the Shell-B.P. Petroleum Development Company of Nigeria Limited. Most of the ostracods occurring in these boreholes are cytherids, the length dimension of which is particularly susceptible to sex-linked variation. For this reason, the analysis recounted here was made on the height dimension, which is less sensitive to sexual dimorphism.

An intuitively satisfying way of studying fluctuations in means is by graphing them against depth (thence time) in the borehole from which the samples were obtained. The reliability of the analysis can be greatly improved by drawing in the confidence intervals for these means. Where thought necessary, tests of significance between critical means can be made and this is often enlightening, particularly where there is some doubt about the reality of right to left, or left to right, swings.

The variations in the means for the ostracods from Araromi are shown in Fig. 6, along with the confidence intervals of the means. Even a cursory glance at the figure shows that there is close agreement in the patterns of variation. To be sure, there are several missing values, but this kind of situation is common for much borehole data. The insertion of the confidence intervals for the means is a valuable accessory as it shows how much reliance can be put on the patterns. In the case of Fig. 6, it will be
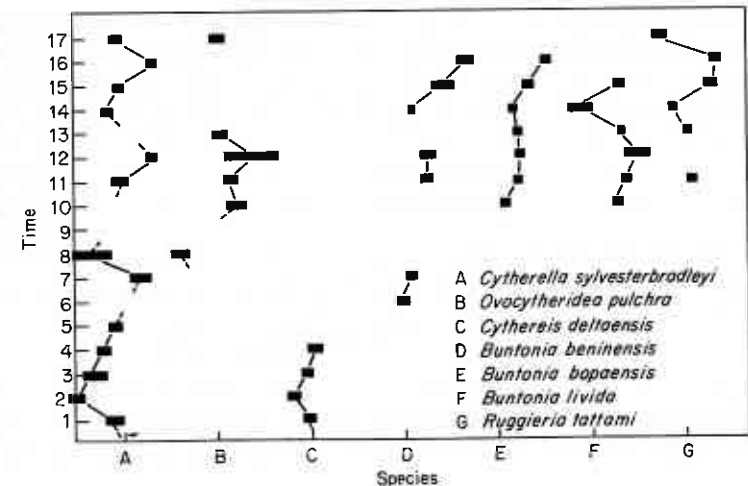
Fig. 6 Variations in the means of seven species of Nigerian ostracods of Late Cretaceous to Early Paleocene age from the Araromi borehole. The lengths of the bars represent the 95% confidence intervals of the means for the height of the carapace. Redrawn from Reyment (1970, Fig. 1).

seen that few of the confidence intervals are broad enough to upset the general picture.

For levels 1 to 4 the variational trend is the same for the species coded as A and C. Levels 10 to 15 have the same morphological directions for the species coded as E and F. Levels 15 to 16 show the same trend for species A, E and G and levels 14 to 17 are analogous for species A and G.

A not unreasonable interpretation of these results would be that all of the carapaces reacted in harmony to one or more environmental factors over the period of time represented and that all the series in time formed by the variation in the mean of carapace height can be expected to be closely correlated.

Principal component logs

We shall now apply the concepts developed earlier in the application of principal component analysis to the preparation of biologs. As a first approximation, the first principal component means will be interpreted as relating to variation in size due to ecological influences (see p. 39).

This time, the Nigerian borehole Gbekebo I was used as a source of data. The eigenvalues and eigenvectors for the three variables length,

height and breadth of carapace, were computed for the separate covariance matrices of seven species of Paleogene ostracods. The first two transformed principal component means (see eqn (14) below) were plotted against depth of sample, as shown in Fig. 7. The equations are

$$\bar{y}_1 = b'_1\bar{x}_1 + b'_2\bar{x}_2 + b'_3\bar{x}_3$$
$$\bar{y}_2 = b''_1\bar{x}_1 + b''_2\bar{x}_2 + b''_3\bar{x}_3 \qquad (14)$$

where the means are denoted as $\bar{x}_i$, b' is the first eigenvector of the covariance (or correlation) matrix and b'' is the second eigenvector of this matrix. In eqn (14), it is the elements of these trivariate vectors that are given.

The principal components mean values are found by substituting the original mean vectors for each species into the first and second equations formed by the first and second principal components, as indicated in eqn (14).

The upper half of Fig. 7 is particularly informative. There is close
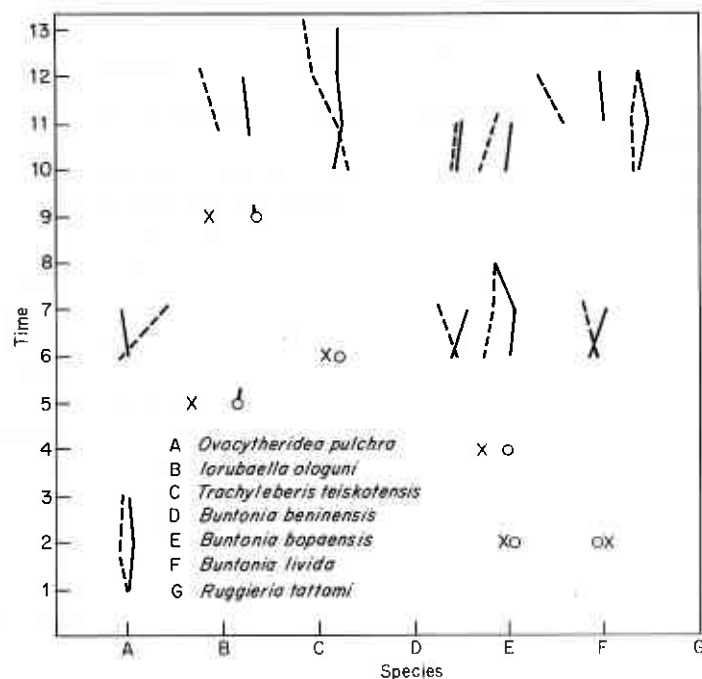
Fig. 7 Plot of the transformed observations for the first two principal component means for seven species of Paleocene ostracods from 13 levels in the borehole Gbekebo I. Redrawn from Reyment (1970, Fig. 2). ——, first transformed means; ---- ×, second transformed means.

agreement in the directions taken by the oscillations in $\bar{y}_1$ for the set of species considered. Comparison of the directions for the plots of $\bar{y}_2$ shows less pronounced concordance.

For levels 10 to 12, the oscillations of the first principal component means agree in sign. For levels 6 and 7, three of the four directions agree. The second principal component means show some agreement in the direction of oscillations, but this is markedly less than for the first component means.

Practical applications of spectral analysis

One may refer to this kind of partition of variability in a multivariate sample as spectral analysis of variation, by an incomplete analogy with the breakdown of light into its component wavelengths.

The value associated with using a principal components log in biostratigraphical work is clear from the point of view that the transformed means of principal components represent a multivariate synthesis, or condensation, of all the variables into a single, new variable, providing that there is a significant correlation structure. This is an obvious improvement over a set of univariate means for logging purposes.

The first suite of principal component means may be interpreted as representing the reaction of the organisms, on the average, to the "gross ecological effect", in the terminology of Reyment (1970).

To summarize the foregoing, we note that as a first-order approximation, it is possible to use the first transformed variable of principal components, based on either the covariance or correlation matrix, for charting the simultaneous reactions of a set of morphovariables to the total contingent of factors of the environment.

Example of the graphical use of principal component analysis

This example illustrates how principal component analysis may be used for the identification of morphological variants associated with the phases of the foraminiferal life cycle. It is included to show some of the capabilities of principal components in displaying morphological variation.

As has been demonstrated by Nyholm (1962), the morphology and

composition of the test of some benthic foraminifers may differ markedly, according to the stage of the life cycle at which the test was constructed. Principal component analysis offers a means of identifying some of the morphological categories of the life cycle in fossil material.

An example of the graphical display of principal components results for the Nigerian Paleocene foraminifer *Afrobolivina africana* (Graham, de Klasz, Rérat) is illustrated in Fig. 8. This figure shows ellipses constructed around the plots of 350 transformed scores for the first two axes of principal components for six measures on the shell (four size dimensions, proloculus diameter and aperture diameter). The ellipses were fitted to the points identified as belonging to the three phases by inspection of the material, the plots having been made without *a priori* decisions on the
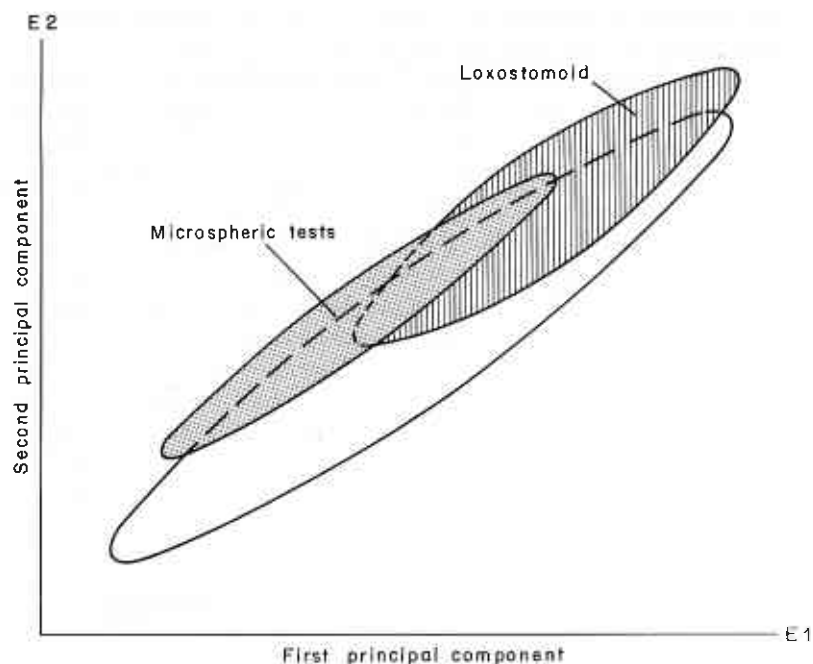
Fig. 8   Principal component analysis of morphs in the life cycle of the Paleocene foraminifer *Afrobolivina africana* from the Akisinde borehole in Western Nigeria. E_1 and E_2 denote the first and second principal components. After Reyment (1966c). The partly covered white ellipse represents megalospheric tests. Microspheric tests are concentrated in a subgroup. Loxostomoid tests (all megalospheres) group in an ellipse which is slightly displaced with respect to normal megalospheres.

morphologies. The results suggest that the material contains at least three stages in the life cycle of this species of *Afrobolivina* (Reyment, 1966c). The existence of a loxostomoid phase seems to be a feature of the genus, as *Afrobolivina afra* may also produce a loxostomoid variant.

The foregoing example gives us food for thought. If a method based on the interpretation of fluctuations in means is to be of any real use, these means must be free of the kinds of effects just found to occur in *Afrobolivina*. Ideally, an organism should be fairly stable in its variational pattern, a prerequisite seldom met by foraminifers, which tend to be highly variable in many respects. Nyholm's (1962) *Cibicides* is certainly an extreme case with its numerous shell variants, but there would seem to be ample evidence for analogous variability in other foraminifers. There is also the question of compositional differences in the tests of the phases.

The vexed problem of growth- and size-confounded data will be given special consideration shortly, but before taking up this vital subject, there is a practical point about principal component analysis I need to bring up.

Principal components can recover information

It is natural to ask whether the method of principal components really can recover information hidden in a set of multivariate measurements of some kind or other. This is a very important question, as the spectral method of analysis depends largely on the ability of principal component analysis to reduce the dimensionality of a problem.

An experiment was made using generated data for three dimensions to illustrate the concept of recoverability, the results of which are depicted in Fig. 9. Three hypothetical environments were constructed so as to differ from each other so slightly that bivariate plots of the raw observations do not show obvious clustering. These plots produce diffuse, homogeneous fields. The data were then subjected to a standard principal component analysis of the covariance matrix, after having been made into a single sample, and the scores of the transformed observations plotted. As is shown by the specimen plot given in Fig. 9, the recovery of the input information is complete, with each of the environmental categories falling into one of the three widely separated clusters. Fuller details are given in Reyment (1966b).

In nature, things seldom work out as impressively as in this constructed example, but the object of the experiment, namely, to illustrate the power of the principal components technique, is here well achieved.
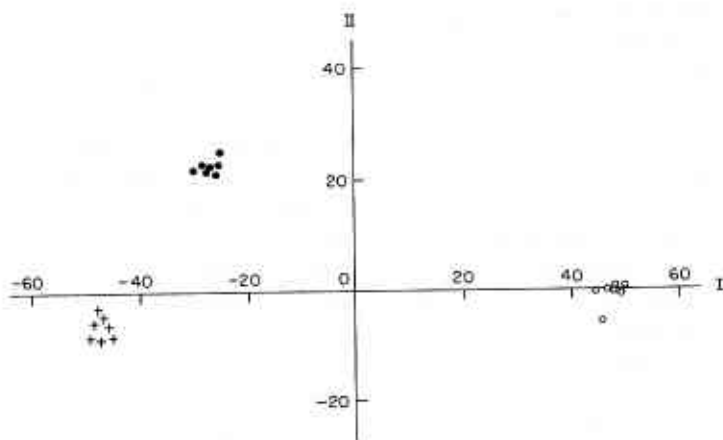
Fig. 9 Diagram illustrating how a principal component analysis plot fully recovers all input information in a set of generated trivariate data. After Reyment (1966b). ○, first hypothetical environment; ●, second hypothetical environment; +, third hypothetical environment. I and II denote the first and second axes of principal components.

Growth invariant biolog

We have just seen how a sample of the tests of a species of foraminifers can differ greatly from morphological homogeneity, a circumstance that can adversely influence the comparison of means.

With ostracods, with their discontinuous growth stages, it is usually possible to identify adult carapaces correctly and one can therefore be certain of applying statistical analyses and tests to material homogeneous with respect to stage of growth. Organisms lacking a terminal growth stage, such as molluscs, corals, foraminifers, confront us with rather special problems.

If we have a sample composed of individuals at different stages of growth, any attempt at comparing their means must fall short of what is required by the statistical prerequisites of such a test. The mean computed for such a sample will obviously be a function of the composition of the mixture of growth stages and genetically and ecologically controlled size differences of organisms making up the sample. Only where we had two samples with exactly the same mixture of growth stages could we expect to obtain compatible means. The idea of circumventing this difficulty for multivariate studies by applying a transformation for achieving growth invariance derives from Burnaby (1966). Gower (1976) has made the

concept more general and Reyment and Banfield (1976) applied Gower's treatment of the problem to the analysis of planktonic foraminifers.

The mathematical problem can be summarized in the following manner. It is desired to find canonical variates among several populations confounded by growth and size effects, which can be represented as gradients.

With k size-difference components and p variables, the effects to be eliminated may be represented by a $p \times k$ matrix K, the rth column of which consists of elements proportionate to the direction cosines of the rth component. The idempotent symmetric matrix

$$Q = I - K(K'K)^{-1}K' \tag{15}$$

projects every sample value onto a space orthogonal to K where they are free from the growth and size effects reflected in K. When there are v populations, all with the same K, canonical variates, l, can be obtained by solving the equation

$$Q(G'G - \lambda W)Ql = 0 \tag{16}$$

where G is the $v \times p$ matrix of sample means and W is the pooled within-populations dispersion matrix (N.B. here, $G'G$ replaces the between-populations dispersion matrix of classical canonical variate analysis). For simplicity, sample sizes are here taken to be equal.

The estimation of K can be difficult. Gower (1976) has considered several ways of doing this, both by internal as well as external estimation. In palaeontology, internal estimation by principal components can often provide a satisfactory solution and biological evidence in support of this is now beginning to appear (see Chapter 2). For many foraminifers (biserial, triserial and quadriserial forms in particular), most of the size variation resides in the length variable of the test and the first principal component does really extract most of the variation due to differences in growth and size. The matrix K can thus be estimated by the first k eigenvectors of W, the calculations being made on the logarithmically transformed observations. (N.B., for $k > 1$, differences additional to size variations are included.)

The solution of eqn (16) requires the computation of a generalized inverse for QWQ, to wit,

$$C = Q(QWQ)^- Q \tag{17}$$

The required solution is then

$$(CG'G - \lambda I)l = 0 \tag{18}$$

The squares of the generalized distances between the means of groups i and j when projected onto the Q-space are given by

$$D^2 = (g_i - g_j)C(g_i - g_j)' \qquad (19)$$

Thus, by using GCG', it is possible to obtain the coordinates of the group means in the Q-space by the method of principal coordinates (Gower, 1966; Jöreskog et al., 1976).

Equation (15), can be written as

$$Q = I - M$$

where $M = K(K'K)^{-1}K'$. A p-variate vector y is projected by M onto the space spanned by K, and Q projects it onto a space orthogonal to K. Thus y can be resolved into components My, confounded with growth- and size-factors and components Qy free from these factors. The Mahanalobis' generalized D^2 can correspondingly be subdivided into two parts, D_M^2 and D_Q^2, related by the equation $D^2 = D_M^2 + D_Q^2$.

A casual reading of Burnaby's paper may suggest that D^2 has been resolved into the orthogonal components D_M^2 and D_Q^2, but this is not so (see Burnaby, 1966). The additivity merely reflects the definition of D_M^2, which is interpreted as the distance lost through working only in the Q-space (Gower, 1976).

The relationships between the partitioned generalized distance can be seen in the photograph of a model based on one of Burnaby's diagrams (Burnaby, 1966, Fig. 2) and shown here as Fig. 10. The M-space is
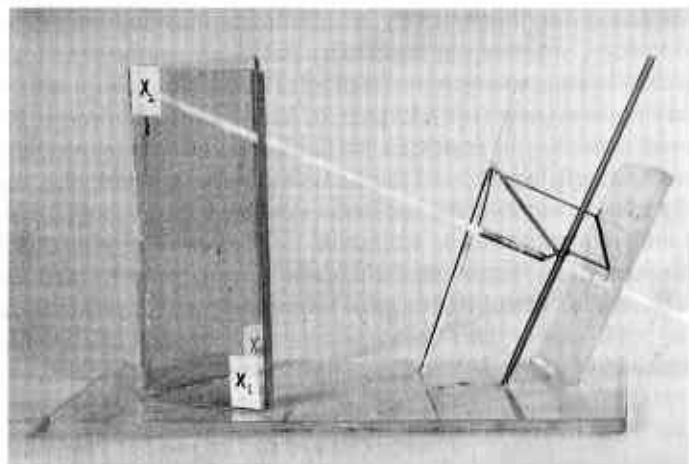
Fig. 10 Photograph of a model for showing the relationships between partitioned generalized distances in connection with the concept of growth invariance. Based on Fig. 2 in Burnaby (1966).

the square to the right, pierced by the white line, the Q-space. The black and grey lines rising from the plane of the x_1 and x_2 axes are two directions of individual growth, parallel to subspace M. The black line on the surface of space M corresponds to D_M^2 and the black portion of space Q corresponds to D_Q^2. These two are together equivalent to the distance between the growth vectors, which corresponds to D^2. (Here, corresponds implies proportionality only.) The reference points for the two individuals involved are taken where the line proportional to D^2 cuts the two growth directions.

It was mentioned earlier on in this section that the coordinates for making the plots can be obtained by the method of principal coordinates. If matrix P contains the eigenvectors of $GCG' = T$, with L denoting the corresponding eigenvalues, then T is a $v \times v$ matrix, P is $v \times p$, and L is a diagonal $p \times p$ matrix. Scaling the vectors so that $P'P = L$ ensures that the rows of P are the coordinates of points referred to principal axes and $PP' = I$. The squared distances between the ith and jth points are (Gower, 1976)

$$t_{ii} + t_{jj} - 2t_{ij}$$

The plot of the first coordinate against stratigraphical position forms a "biolog".

Gower (1976) has also considered external methods of estimating the growth vectors. In the analysis of living organisms, this is often likely to provide an excellent alternative. An external method of estimation of growth vectors is one that utilizes some character which is correlated with the variables of interest and which gives an estimatation of the age of each specimen. For example, the degree of wear on the teeth of rodents.

Even in the case of living animals it can be difficult to find a suitable concomitant variable. Delany and Healy (1964), in their study of the long-tailed field mouse, used one concomitant variable, a measure of tooth-wear, thought to be a good indicator of age. Estimating K by the first k eigenvectors of W may be associated with the difficulty that the principal components so obtained are affected by the scales of measurement of the different variables. This difficulty disappears when all scales are the same or if normalized variates are used. Jolicoeur (1963) pointed out that in biological problems, the logarithms of the variables are usually related linearly, and that taking logarithms is one way in which the principal components can be made scale-free.

Internal estimation by principal components is intuitively reasonable provided the growth effects are the major source of variation within each population. In this case, the elliptical cloud of points representing within-population samples would be spread out around an elongated major axis, the direction of the first principal component.

Growth invariance and a planktonic foraminifer

As an example of the use of the method of growth invariance, I submit the data for *Subbotina pseudobulloides* (Plummer) analysed by Reyment and Banfield (1976) and extracted from Malmgren (1974). The material comes from the Danian of southern Sweden (Scania) from a borehole drilled in connection with the appraisal of limestone deposits. Six characters for size and shape were measured on each specimen by Malmgren; these are illustrated in Fig. 11.

The growth-invariant analysis showed that a size effect could be removed from *S. pseudobulloides* by the first eigenvector of the logarithmic covariance matrix. The effects of extracting one eigenvector (the first) and then two eigenvectors on the values of the generalized distances for the six borehole levels considered are illustrated in Table IV. As noted above, the greatest reduction is brought about for $k = 1$ and for some of the distances, this is very substantial.
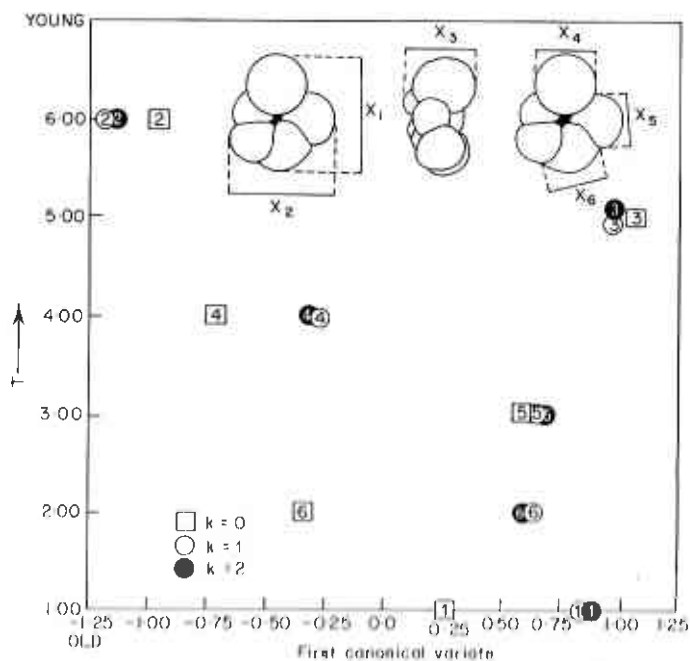
Fig. 11 Plot of the growth-invariant canonical variate means against stratigraphical position for *Subbotina pseudobulloides* for $k = 0$, 1 and 2. After Reyment and Banfield (1976).

TABLE IV
Squared generalized distances for *Subbotina pseudobulloides* for $k = 0, 1, 2$

$k = 0$					
1	0.0000				
2	2.8677	0.0000			
3	1.9539	4.4016	0.0000		
4	1.5322	0.8623	3.6189	0.0000	
5	0.7524	2.9727	0.8738	1.9316	0.0000	
6	1.3623	2.6748	3.1345	0.7847	2.2441	0.0000

$k = 1$					
1	0.0000				
2	2.8647	0.0000			
3	0.7654	3.0867	0.0000		
4	1.3496	0.7246	1.3157	0.0000	
5	0.2759	2.4150	0.7146	0.6821	0.0000	
6	1.0088	2.3848	0.2947	0.7563	0.5923	0.0000

$k = 2$					
1	0.0000				
2	2.8430	0.0000			
3	0.7495	3.0120	0.0000		
4	1.3096	0.6042	1.3101	0.0000	
5	0.2202	2.2688	0.7021	0.6807	0.0000	
6	0.9927	2.3103	0.2947	0.7515	0.5801	0.0000

The canonical variates without the extraction of size differences ($k = 0$) and for one eigenvector of principal components extracted ($k = 1$) are listed in Table V. As only to be expected from the values of the generalized distances, the effects are quite strong when the size-variation is extracted. The coordinates of the means for the first canonical variate with the three combinations for $k = 0$, 1 and 2 were plotted against relative stratigraphical position (Fig. 11). This forms, in effect, a biolog (note, that the ordering of the samples in time is 1-6-5-4-3-2). The change in the shape of the biolog caused by the extraction of size-differences is readily apparent.

Reyment and Banfield (1976) compared the results obtained for *Subbotina pseudobulloides* (Plummer) with those found for *Globoconusa daubjergensis* (Brönniman) from the same sampling levels. They found that the plots of the first canonical means for both species ($k = 1$) gave concordant patterns. This is an interesting result and it could reflect agreement in the reaction of the two species to some environmental feature, possibly of a palaeoclimatological nature.

TABLE V
Canonical variate analyses of *Subbotina pseudobulloides*

	k = 0					k = 1				
	1	2	3	4	5	1	2	3	4	5
Eigenvalues										
	3.1008	1.3326	0.6374	0.2336	0.0235	2.2298	0.6467	0.2667	0.0619	0.0001
Canonical variate loadings										
	-1.6494	-2.0139	-17.4327	-2.5077	15.2510	1.5126	17.9109	0.4724	15.5306	-249.7511
	4.7737	0.3278	-8.0457	-2.6102	-16.2890	-4.0535	7.2746	5.9664	-16.3598	-98.3832
	-8.4805	9.7136	0.9643	3.7573	-0.0948	12.9584	-1.6057	-2.6835	-1.4948	19.6511
	5.6673	-3.3072	7.2132	8.2769	-6.8715	-6.6390	-6.8857	-7.5410	-7.8315	52.2846
	2.4804	0.9962	7.2153	-7.5067	2.5701	-1.6893	-7.5132	6.7275	4.1577	79.9098
	2.4981	-1.1284	7.9632	2.5453	5.4067	-2.4882	-7.6466	-3.7076	4.7446	162.8214
Coordinates of means										
	-0.2762	0.3301	-0.5245	0.2498	0.0258	0.4676	0.5034	-0.1826	-0.0667	0.0043
	0.9390	-0.7096	0.0179	0.1451	-0.0435	-1.1577	0.0329	-0.1904	-0.0278	-0.0033
	-1.0794	-0.3072	0.4137	0.1008	0.0420	0.5503	-0.3420	-0.2395	0.1348	-0.0010
	0.7064	0.1001	0.0143	-0.2162	0.1046	-0.4440	-0.0502	0.2503	0.0903	0.0069
	-0.6122	-0.1777	-0.2669	-0.3043	-0.0652	0.3066	0.2902	0.2569	0.0392	-0.0071
	0.3224	0.7643	0.3455	0.0249	-0.0636	0.2772	-0.4342	0.1052	-0.1698	0.0001

On interpreting biologs

A caution concerning the use of the biolog is now in order. Borehole samples are taken at separated intervals in a borehole and the points making up the log are therefore discrete, not continuous. The latter impression can inadvertently be conveyed through the practice of joining together plots deriving from borehole samples (usually raw frequencies). It is particularly in situations where the discontinuous biolog is being compared with continuous logs of the kind well known to petroleum geologists that confusion can occur, as the geologist will most likely look for agreements between the biological and physical logs in the uncharted areas between the sampling points and only find straight lines. Elementary as this point may seem to you, I have found it to be a common cause of misunderstanding in relating the biological information to the physical and one which can confuse log-analysts confronted for the first time with biological data.

In the light of current developments in what is sometimes known as "pattern analysis", statisticians may look for absolute agreements in the biologs, rather than the significant feature of agreement in the directions taken by the oscillations. This is, however, to misunderstand the nature of the variations.

Before leaving this subject for the moment, I wish to point out that Flessa and Bray (1977) propose using the second eigenvalue of a bivariate morphometric analysis as a measure of size-independent variation. They use the logarithmically transformed variables in their correlation matrix (*op. cit.* p. 352). They plot the eigenvalues against stratigraphical location to produce a kind of log. This procedure seems valid in principle, but only for two variables.

Biolog using stabilized canonical variates

In Chapter 3, I took up the subject of stability in canonical vectors in canonical variate analysis (p. 46). How will the log-analyst know when to select this approach? No hard and fast rule can be given and I can only offer a few practical suggestions.

At the outset of a logging project, I make a point of checking the variables measured on the species selected for constructing the log for redundancy. This can be most easily done, from the aspect of canonical variates analysis, by means of Campbell's method for studying the stability of the coefficients of the canonical discriminants. If the checking is

done early enough, it can result in a considerable saving of time as redundant variables need not be measured. Note, however, that this analysis of redundancy is limited in its validity and it is not good policy to extrapolate from one particular problem to seemingly analogous ones. This means that each project should be initiated by a full appraisal of as many variables as seems reasonable for a given situation. A worked example of this kind of a biolog is given in Chapter 8.

Relating the organisms to the palaeo-environment

Mention has already been made of the possibility of introducing expressions of some ecological factors more directly into quantitative biostratigraphical analysis. The readiest way of doing this is by canonical correlation analysis (see p. 48). For biological and sedimentological applications of this multivariate statistical method see Blackith and Reyment (1971) and Reyment (1972, 1975).

Briefly, this multivariate method seeks out linear combinations of two sets of variates such that the ordinary correlation coefficients between the two new transformed (univariate) variates is a maximum. It finds thus successive maximum interset correlations in order of size. One of these sets may contain observations on frequencies of a number of species, or, observations on morphological variables (as means) and the other set may comprise ecological variables. The magnitudes of the standardized coefficients for the variables of each pair of canonical vectors tends to indicate the important variables in a particular correlation.

The observations on the variables of each of the two sets can be reduced to a series of bivariate relationships whereby each of the new canonical variables represents a linear combination of the respective multivariate sets. The scatter diagram of the canonical correlational scores provides a useful graphical means of illustrating the interplay between certain factors of the palaeoenvironment and abundances of diagnostic organisms. The plot of the two sets of transformed canonical correlation scores against stratigraphical location may be used to produce a pair of logs which relate variations in the environmental factors considered to the frequencies of the organisms.

As is often the case in multivariate applications to geological data, a word of warning must be sounded. A basic premise of canonical correlation analysis is that the data are homogeneous, ideally from a multivariate normal population and with two sub-sets of variables. A basic symmetry in the underlying distributions is desirable, otherwise the influence of
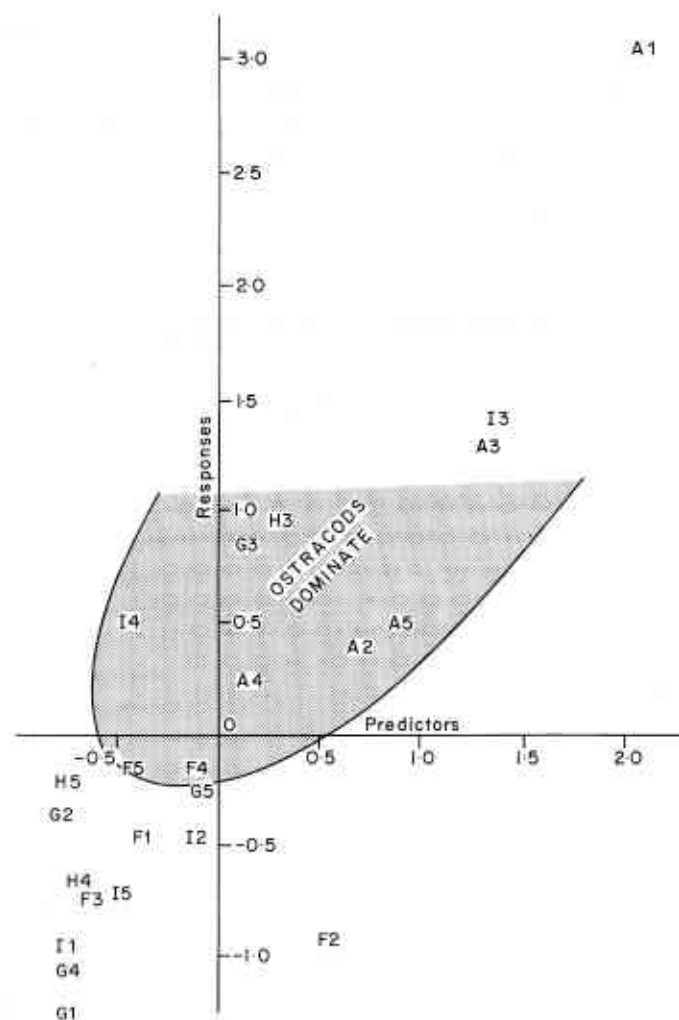
Fig. 12 Plot of canonical correlation "scores" (actually transformed partitioned observational vectors) for an ecological study of Recent ostracods from the Niger Delta. Redrawn from Reyment (1975). The letter–number combinations refer to transects from shallower to deeper water. Thus A1 to A5 are five progressively deeper sampling stations along transect A.

"atypical" observations on the correlation coefficients may be large. Figure 12 illustrates the type of result that can be obtained using canonical correlation analysis of environmental factors (here, pH, Eh, total organic matter of the sediment, $CaCO_3$, distance of the sampling site from the shore, P, and S) on the one hand, against ostracod frequencies

on the other. Figure 12 is the graph of the predictor (environmental) variables against response (ostracod frequencies) variables for sampling sites in the Niger Delta. The samples lying in the dotted field are rich in ostracods while those lying outside this field, and particularly the samples in the lower half of the graph, tend to be poor in ostracods, or to lack them. In the light of the remarks made in the foregoing section, the data analysed here are homogeneous in all respects.

Comparing patterns of logs

An important use of biologs lies in employing them for correlating between adjacent borehole sequences. An example of this concept is shown in Fig. 13. Here, the arrows mark sampling levels in boreholes A and B and the crosses denote plots of values of some kind or other obtained from the morphometric analysis of a fossil organism.

DEVIATIONS ———   ——— DEVIATIONS ———

Borehole A   Borehole B

➡ sampling levels
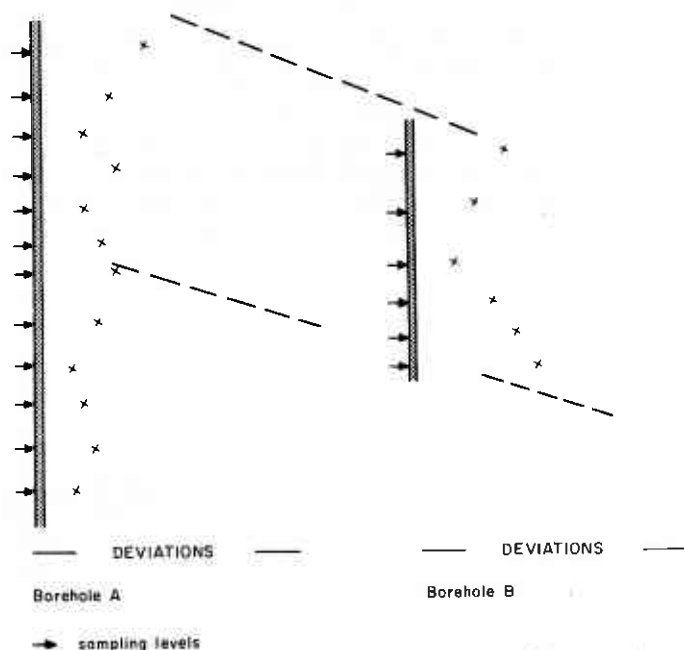
Fig. 13 Schematic representation of the concept of correlating between boreholes A and B by using fluctuations in the means of a species occurring in both boreholes. The fluctuations are plotted as "deviations" from the grand mean. The arrows denote sampling levels in the boreholes, the dotted lines the correlations.

A possible drawback

There is one possibly serious drawback to the preparation of biologs, and to the comparison of adjacent patterns of variation, and it is of a practical nature. Micropalaeontological samples are not always delivered to the laboratory in a form enabling them to be connected directly to a point in the rock sequence penetrated by the drill. Often, bagged composite samples are used, not to speak of mud samples and the like. Bagged samples have the disadvantage that one must work with mixed material. The situation for bagged core pieces can be slightly improved by confining the analysis to the fossils obtained in one suitably sized piece of core rather than washing the entire contents of the bag. The material can be treated as a random sample selected from a certain interval in the rock sequence.

Methods of comparison of logs

Borehole logs form oscillatory curves with, usually, well-defined peaks and troughs occurring in some parts of the sequence. The biolog seldom contains the spikes so typical of some physical logs of rock sequences. Spikes of exceptionally wide amplitude are not infrequently associated with features of the "noise" category and thus deviate from the main geological picture reflected in the log. Nonetheless, in the case of biologs, some important palaeoecological event may be the cause of a marked spike.

Once a sequence of variations has been established for a suite of samples, one will be interested in comparing this log with others obtained from neighbouring boreholes. The comparison can normally be enhanced by the recognition of "datum" lines. There may, however, be complications of various kinds, such as differences in the rate of sedimentation in the two places, a non-sequence, or redeposition of allochthonous sediments, with the result that episodes may be condensed in the one sequence and expanded in the other.

Cross-correlation

The well-known correlation coefficient was adapted many years ago for use in the analysis of serial data. In this version, it is known as the serial correlation coefficient. This correlation coefficient is sometimes useful in comparative work but, for the most part, the cross-correlation coefficient

is more to the point. For cross-correlating between series in time (i.e. borehole sequences, borehole logs) the samples must be taken at the same intervals apart and it must be assumed that the rates of sediment-ation in both sequences were the same. It is also necessary that the sequences compared be free of stratigraphical breaks. In biologging studies, the cross-correlation technique is mainly of use for checking the general agreement in different kinds of logs from the same borehole.

The method for calculating cross-correlation coefficients is essentially the same as that for finding the ordinary correlation coefficient. The so-called "match positions" denote the sites at which observations have been made in the two time series. Davis (1973) gives a useful present-ation of the technique applied to geological data.

Slotting of stratigraphic sequences

It is standard procedure in petroleum exploration to measure properties of the bedrock pierced by a borehole by means of physical methods, such as neutron logging, gamma radiation logging, the electrical self potential (SP), the electrical resistivity of the bedrock, redox logs, and sonic logs. Physical logging of boreholes has as one of its aims, the facilitation of correlations between boreholes.

For introducing the concepts of the following sections, I present the basic notation and concepts of the slotting of sequences at this point. An observational point in a borehole will be referred to as an "object". That is, each depth at which observations have been made in the borehole will be called an object.

In the univariate case, an object will be a single measurement (a gamma-log reading, for example), but more generally, each object will be described by a set of measurements, all of them referring to the same depth in the borehole.

Physical logs are taken continuously, whereas biologs are made from samples taken at discrete intervals. For the purpose of statistical analyses, convenient sampling points are selected for physical logs and this is therefore a justifiable means of putting them on a par with the biolog.

The slotting approach (Gordon, 1973; Gordon and Reyment, 1979) assumes that all relevant distances between each pair of objects may be summarized by a single measure of pairwise dissimilarity. The dissimilar-ity between the jth object in the first sequence, A_j, and the kth object in the second sequence, B_k, will be denoted by $d(A_j, B_k)$. This dissimilarity will be influenced by the differences in the measurements (e.g. gamma ray, self potential, sonic, etc.) recorded for the two objects.

Let $(x_{ij}(A), x_{ik}(B))$ denote the values taken by the ith variable at A_j, B_k respectively $(i = 1, \ldots, p; j = 1, \ldots, m; k = 1, \ldots, n)$. Thus, for example, $i = 1$ could refer to readings of the intensity of gamma radiation, $i = 2$ to readings of the short normal resistivity log, and $i = 3$ to computed values of the biolog.

Then,

$$d(A_j, B_k) = \sum_{i=1}^{p} w_i f(|x_{ij}(A) - x_{ik}(B)|) \tag{20}$$

where w_i denotes the weight to be given to the ith variable and $f(|x_{ij}(A) - x_{ik}(B)|)$ is some monotonic-increasing (= invariably increasing) function of the absolute value of the difference of measurements of $x_{ij}(A)$ and $x_{ik}(B)$. The $[w_i(i = 1, \ldots, p)]$ could be used

1. to standardize each variable so that the readings were all in terms of standard deviations from the mean, or,
2. to give extra weight to variables which experience had indicated to the geologist to be more relevant for purposes of slotting, or,
3. to allow for some combination of (1) and (2).

The second criterion is often of interest as one log might be much more diagnostic than others in the same sequence, under a particular set of conditions, and it might, therefore, be thought a good idea to weight its contribution to the slotting.

The monotonic function f could take many forms, yielding measures of dissimilarity of varying degrees of relevance. For borehole slotting, Gor-don and Reyment (1979) have found the absolute difference measure given by $f(z) = z$ to be very useful.

Gordon (1973, p. 198) has described some properties of measures of dissimilarity. The calculation of the $m \times n$ matrix of dissimilarities be-tween the objects in the two sequences can be made quickly. Once the initial computation has been done, the slotting procedure operates on the matrix of dissimilarities. This means that the amount of work needed to consider simultaneously a number of different kinds of logs when compar-ing two wells is little more than what is required for comparing a single log from one well with the same type of log from the other well.

The measure of discordance $\delta(S_1, S_2)$ between sequences S_1 and S_2 is defined as the sum (over all objects in each sequence) of the dis-similarities between an object and the two objects in the other sequence, which bracket it.

In order to permit comparison of a sequence with several other sequences, it is often convenient to standardize this measure of discor

dance. Thus, if

$$\mu(S_1, S_2) = \sum_{i=1}^{m-1} d(A_i, A_{i+1}) + \sum_{k=1}^{n-1} d(B_k, B_{k+1}) \qquad (21)$$

then a standardized measure of discordance is

$$\psi(S_1, S_2) = \frac{\sigma(S_1, S_2) - \mu(S_1, S_2)}{\mu(S_1, S_2)} \qquad (22)$$

I should mention that this method will always produce a slotting of the sequences, no matter how slight the resemblance between them. The value assumed by the measure of discordance (eqn 21) will give some indication of the degree of similarity of the sequences. Lower values would suggest better agreements, but it is difficult to formulate realistic null models leading to tractable tests of statistical hypotheses. In practical applications, the user of the method will therefore be concerned with exercising his judgement regarding the significance to be attached to the value of the measure of discordance. Generally, a basis for this can be arrived at when the slotting behaviour of a particular type of material has become known as the result of several trials.

Finding reference points in boreholes

Comparisons between boreholes can be greatly improved if marker levels can be located and tied to each other. For the best application of the slotting technique, markers provide an extremely valuable aid. I have already noted this in the foregoing. Unfortunately, in the kinds of situations in which one would be interested in making use of a biolog for correlational purposes, lithological markers will seldom be available (i.e. monotonous shale sequences). In fact, where lithological markers are present in a sequence, the use of the resistivity log for representing environmental fluctuations may be invalidated.

Palaeontological markers are sometimes available but, usually, they tend to be insufficiently concentrated in monotonous, unbroken sedimentary sequences to be useful in the detailed kind of work under consideration. This is not to say that good marker fossils do not occur, nor does it imply that they are never found in the type of sedimentary sequence envisaged here.

Predation in borehole logging

A type of ecological event that may be of great use derives from episodes of heavy predation, such as those caused by drilling gastropods. Ostracods, juvenile pelecypods and juvenile gastropods of the Late Cretaceous and Cenozoic not infrequently display the marks of predation caused by drills of the families Naticidae and Muricidae. Borehole samples sometimes contain eloquent testimony of the activities of, particularly, naticids and such evidence can be useful as a means of establishing a marker level.

As an aside, we may note that in reconstructing the conditions of deposition of a sediment, information on gastropod predation is often valuable, for example, with respect to the ecological factor of bathymetry. Although it is possible to rear the living naticid *Lunatia nitida* in the laboratory without much difficulty, this species only occurs in nature below a depth of 25 m, as shown by observations in the North Sea.

Another point of palaeoecological applicability relates to the niches inhabited by naticids and muricids. Naticids dwell within the sediment and attack their prey by burrowing to it from beneath, if it is an epibiont, and towards it, if it is an endobiont. Muricid drills are epibiontic forms and move over the surface of the sediment towards their prey. A sediment containing pre-dated shells drilled solely by naticids offers a source of information for deducing the nature of the environment inhabited by these organisms. Thus, shells drilled by naticids belong to endobionts, while those drilled by muricids belong to epibionts.

Drills will attack pelecypods, gastropods, even of their own kind, cirripedes and ostracods, the last being most often selected by juveniles. On odd occasions, even foraminifers may be drilled, although such events would seem to be the outcome of an ethological error made by newly hatched individuals.

Drills do not attack all shelled organisms, nor even all genera of a particular family. They tend to be highly selective in what they choose (Moore, 1958; Reyment, 1966a). In the Niger Delta, it seems that species of *Ostrea* are the most sought-after prey, thereafter species of *Cerastoderma* (despite the strength of the lateral ornament). Ostracods are not high on the list of desired prey, but they are preferred to many species of pelecypods and gastropods, both juveniles and adults.

An analysis of ostracod predation by naticids in the Araromi and Gbekebo boreholes of Western Nigeria (Reyment, 1963a) showed that species of *Brachycythere* are the most frequently drilled forms and that individuals of species with unornamented lateral surfaces tend to be more
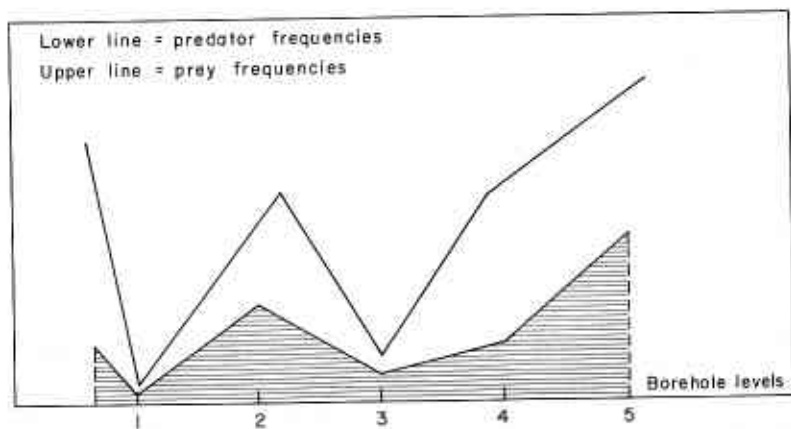
Fig. 14   Reconstruction of a predator (naticid gastropods)–prey (ostracods) relationship for the Nigerian Paleocene. After Reyment (1966b). The frequencies of predators are illustrated at ten times those of the prey.

commonly attacked than costate ones. Almost all of the predation is due to naticids.

Naticids and muricids drill quite different kinds of holes (Reyment, 1966a) Those made by naticids are roughly conical and they are usually countersunk. Those drilled by muricids are smaller and they are parallel-sided without being countersunk. Both kinds of holes bear scratch-marks from the radula, but these are usually more frequent around naticid holes. A predation profile for the Nigerian borehole Araromi I is shown in Fig. 14.

The predation profiles for various levels in a borehole should be established with respect to the following criteria:

1. the species selected for attack—ostracods, juvenile pelecypods, or gastropods;
2. the type of hole drilled—whether made by a naticid or a muricid.

Preliminary sorting of data

It is often difficult to know exactly where to start an analysis of biological borehole data. A useful point at which to commence can be yielded by some kind of ordinating procedure which looks for heterogeneities in a suite of observations. This is a type of problem which has long been of interest to quantitative biologists. A useful approach has been found to be

by the use of some kind of Q-mode factor analysis (cf. Chapter 5 Jöreskog et al., 1976). In such an analysis, a matrix of similarity coefficients or associations is generally used in the stead of the correlation matrix of R-mode analysis. Gower's (1966) method of principal coordinates analysis gives useful results. A basic property of this multivariate statistical method is that it represents graphically "distances" between the individuals of a sample. Principal coordinates analysis has the added advantage over the majority of other Q-mode analytical procedures that it permits the user the option of uniting continuous, discontinuous and qualitative variables in the same analysis. For Euclidean distances, principal coordinates is equivalent to principal components.

A sustained multivariate trend in a set of borehole observations will show up clearly in a principal coordinates analysis. If the analysis be repeated for several species from the same set of samples, time-related patterns may be discovered. The eigenvectors of principal coordinates are not interpretable in the same manner as those of principal components. It is, however, possible to see how efficient a coordinates analysis is by inspecting the eigenvalues. If these fall off very slowly, this indicates that there is much random variation in the material, which is usually an undesirable feature. Biological examples of principal coordinates analysis are featured in Blackith and Reyment (1971).

As an example of the application of principal coordinates to borehole samples, I have illustrated the graphical display for mean vectors for the Paleocene ostracod species Cytherella sylvesterbradleyi from the Nigerian boreholes Araromi I and Gbekebo I. Here, Gower's similarity coefficient was used. There are 20 samples consisting of mean vectors (length, height and breadth of carapace) from the penultimate and adult growth stages of the species. If there had not been any chronological differentiation in the species over time, the plots for the penultimate growth stage (growth stage 7) should form a close cluster, likewise for the adults (growth stage 8). As shown by Fig. 15, there are distinct clusters, but these are elongated and, moreover, not ordered according to location in the boreholes. This implies that we are dealing with markedly variable sequences. A further interesting result of the coordinates plot is that the material for Gbekebo segregates above that for Araromi (there is one exception only).

Canonical variates in biostratigraphy

Canonical variate analysis offers an excellent means of displaying a sequence of variations in a fossil species in that it can be used to produce
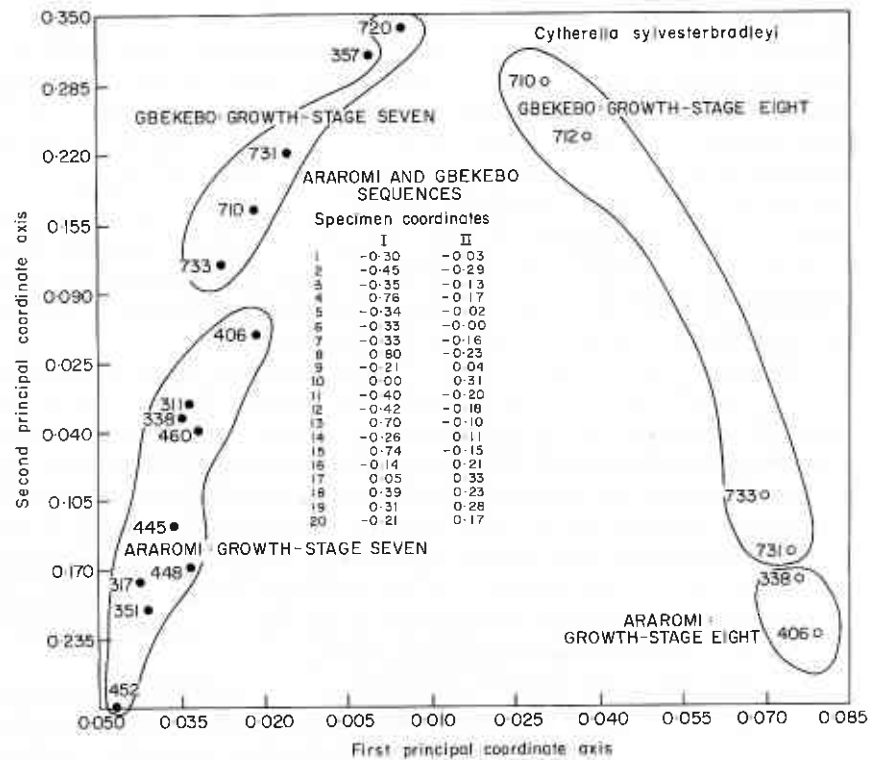
Fig. 15. Graphical display of a principal coordinates analysis of *Cytherella* from the Nigerian Paleocene, from the boreholes at Araromi and Gbekebo. The numbers (e.g. 452, 351) refer to borehole depths in metres.

informative graphical illustrations of the chronological relationships. It allows us to ascertain whether the fluctuations in a sequence of means are significantly different. Finally, it permits a useful representation of variational patterns for comparisons with newly acquired data sets.

Bearing in mind the results of Campbell (1979, unpublished thesis), the standard means of doing canonical variate analysis may not be the best way of studying data of the above-mentioned type and it will probably be found necessary to make a thorough search for redundancy in the material.

As an example of the application of canonical variate analysis in a biostratigraphical connection, I have taken some data on the ostracod species *Cytherella sylvesterbradleyi* Reyment from the boreholes at Araromi and Gbekebo in western Nigeria. The material is Paleocene in age. There are ten samples from five levels in each of the boreholes on

which the variables length, height and breadth of carapace were measured. All samples accord with multivariate normality and the covariance matrices used in the calculations are statistically homogeneous. Araromi spans 16 m and Gbekebo 73 m. There is a size effect in the first canonical vector.

The first two eigenvalues for the ten samples are highly significant and, expectedly, the multivariate analysis of variance for the differences in mean vectors is also significant, thus confirming the strong differentiation in the samples. The plot of the transformed mean vectors for the first and second canonical variates (Fig. 16) does not display unidirectional trend and it may be concluded that the observed differences in multivariate means are due to strong fluctuations in the nature of the environment (cf. Malmgren and Kennett, 1978). This is especially noticeable for sample AR 351.
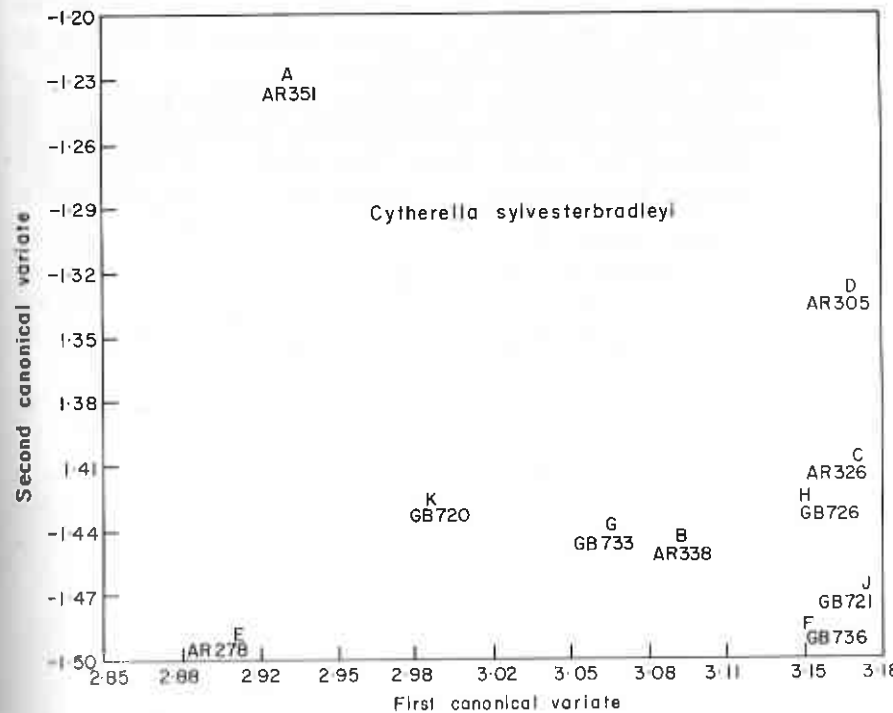
Fig. 16 Transformed means from the canonical variate analysis of Nigerian Paleocene *Cytherella*. The figure shows the effects of what may be ecological size differentiation over time. AR = Araromi, GB = Gbekebo. The Araromi samples are from metre levels 351, 338, 326, 305, 278. The Gbekebo samples are from metre levels 736, 733, 726, 721, 720.

Series in time of scatter ellipsoids

The scatter ellipsoid was discussed in Chapter 3 (p. 48). Figures 17 and 18 illustrate variation in the length (Fig. 18) and orientation (Figs 17, 18) of the first ellipsoid axis for trivariate scatter ellipsoids (length, height and breadth of carapace) for two species of Nigerian Paleocene ostracods from the Araromi and Gbekebo boreholes. These figures are arranged so as to display the relative orientations of the first axes, from sampling level to sampling level. The information on the axes was obtained by the usual method of principal component analysis on each of the sample covariance matrices. The angles shown in Fig. 17 were computed between the first principal axes of the sample covariance matrices for the pairs of samples noted below the individual figures. The second and third axes, all very small in relation to the length of the first, are drawn in. For *Ovocytheridea pulchra*, the species involved in Fig. 17, the first principal axis gives the impression of having been rotated to varying degrees in the samples, as indicated by the angles marked on the figures.

The situation illustrated in Fig. 18 is for the orientation and length of the first axes, relative to a rectangular coordinate system, of samples of *Buntonia bopaensis*. All samples come from consecutive levels in the Araromi borehole. Here it will be seen that the first principal axis shows a

| (a) | (b) | (c) | (d) | (e) | (f) |
| Gb.733/731 | Gb.733/720 | Gb.731/720 | Ar.338/Gb | Ar.338/Gb731 | Ar.338/720 |
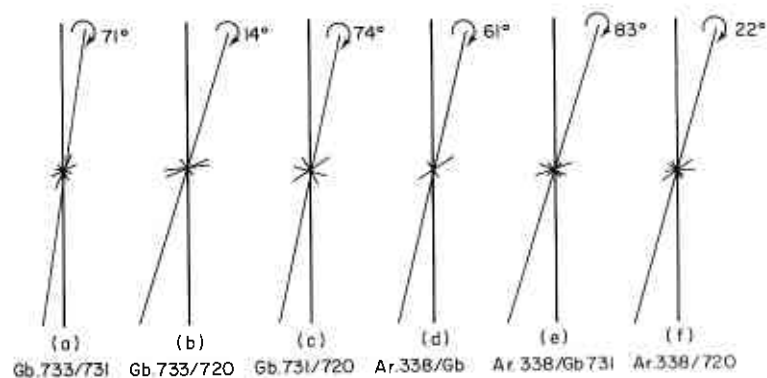
Fig. 17   Lengths of the first axis of the scatter ellipsoid for *Ovocytheridea pulchra* Reyment, a Nigerian Paleocene ostracod. The rotations of this axis relative to each other for the levels considered are marked on the figure. Redrawn after Reyment (1963b). The samples are from the boreholes at Araromi and Gbekebo.
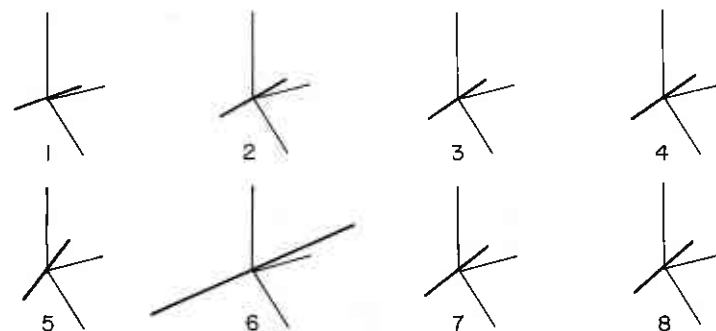
Fig. 18   Lengths of the first principal axes of the scatter ellipsoids for *Buntonia bopaensis*, a Nigerian Paleocene ostracod species (Araromi borehole). Redrawn from Reyment (1963b).

higher degree of variation in orientation than found for the *Ovocytheridea* of Fig. 17. It is also very variable with respect to length which prompts the suggestion that the environmental conditions pertaining might have been less stable than those existing in the environment inhabited by the *Ovocytheridea*.

The foregoing cases are interesting in that they provide a demonstration of how possible minor environmental fluctuations may influence the shape of the scatter ellipsoid, where based on morphological variables, thence the covariance structure of the material. Although there is no direct practical use for such information in the preparation of a biolog, it can be valuable for purposes of interpreting the vectors of principal component, respectively, canonical variate analysis.

5

Secular Variation in Discrete Characters

In this chapter, we shall be studying the feasibility of using meristic shell characters for quantitative biostratigraphical purposes. Although some species of foraminifers display meristic variation, the present discussion has been centred around ostracods.

Two case histories are taken up here. The first of these deals with a species from the Miocene of France, the second of them concerns two species from the Cretaceous of southern Morocco. In neither case are variations in discrete characters suitable for establishing quantified bioseries with sufficient detail to make a biolog. Large-scale trending occurs, however, in both. The Moroccan ostracods provide an example of how the slotting technique can be used for correlating between boreholes.

Variation through time of such characters as the marginal spines of the carapace of some ostracods (as well as other arthropods), longitudinal ribs of some foraminifers, ribbing of pelecypods and gastropods may be useful in certain situations for producing a graphical biostratigraphical display. This is the least studied of the areas treated in this book and the case histories presented below are to be seen as being in the nature of probes.

The French Miocene *Cytheridea*

Reyment *et al.* (1977) analysed variation in discrete characters of the carapace of the Miocene ostracod *Cytheridea acuminata caumontensis* Carbonnel with the end in view of seeing whether meristic characters are suitable for constructing a biolog. Compared with analyses based on continuously varying characters of the ostracod carapace, scant attention has been paid to variation in discontinuous attributes, despite the fact that the taxonomy of fossil ostracods is largely based on ornamental features.

The species of *Cytheridea* analysed here displays three kinds of ornamental variation:

1. variation in six anterior spines;
2. presence or absence of a posterior spine;
3. regular or irregular lateral reticulation.

Not all ostracods show such well-developed meristic variation as the *Cytheridea* now considered. Many factors seem to decide the issue, and in some cases, temperature and salinity may influence the development of the anterior spines (see discussion in Chapter 2, p. 24). Variation in the number of anterior spines in species of the genus *Haplocytheridea*, an important element in the French Miocene, may be an example of this kind of variation. Such variation may be conveniently treated by methods for analysing continuous variation.

The mode of reaction of the ostracod carapace to factors of the environment and the resultant affect of interactions on the ornament is another area of micropalaeontology on which we have little primary information. What we need here are well-designed experiments to test the reaction of a wide range of marine ostracod species to measured variations in the environment. Reyment *et al.* (1977) showed that time-related shifts occur in the frequencies of spines and in the type of ornament. Although we shall be concerned with precise means of exploiting biological variation in later sections, using continuously varying attributes, much can be learned from the analysis of series in time of discontinuous variables. The present case history may be an example of the so-called "Ludwig effect" in which a population contains several morphs which are adapted to particular sub-niches (Mayr, 1963, p. 415).

The living species of *Buntonia* studied by Reyment and Van Valen (1969) seems to be analogous to *Haplocytheridea* with respect to the type of variation shown by its anterior spines. As already mentioned above, this is not the same as the variation found in the anterior spines of *Cytheridea acuminata. Buntonia* and *Haplocytheridea* could possibly have

some lower threshold at which spine production begins, whereas *Cytheridea* is more likely to have a more complicated mechanism. The following questions were considered to be of interest in the *Cytheridea* study:

1. Are the anterior spines significantly correlated with each other?
2. Are the anterior spines significantly associated with the posterior spine?
3. Does the lateral ornament show systematic variation through time?
4. Do any or all of the meristic characters show time-related trending?

Most of the analyses of the data on the spines were made by means of simple methods of statistical analysis (chi-squared and probits). The most complicated procedure used is that of canonical correlation.

Results

The variation in the anterior spines of *Cytheridea* appears at first sight to be polymorphic in nature. Were this so, there would be 2^6 phenotypes to take into account, representing the presence or absence of each spine at each of the six loci and the presence or absence of combinations of spines. Although it is unlikely that the development of each spine is controlled by independent genetical factors, it is interesting to record that some 60 of the possible combinations were actually observed in the material investigated by Reyment *et al.* (1977). It was assumed that there was a "potential to produce spines". This potential can be assumed to be roughly normally distributed in ostracods showing the same type of variation in anterior spines as *Cytheridea acuminata*. Reyment *et al.* (*op. cit.*) thought the potential to have been controlled by multi-factors such as polygenes. To a certain extent, this approach is in line with Rendel's (1967) concept of Make (see p. 37).

The association analysis, presented in Reyment *et al.* (*op. cit.*), showed that the development of anterior spines is closely and positively connected between neighbouring loci. It was also established that the middle spines of the anterior margin tend to be better developed than the external ones. The average potential for making spines can be expressed quantitatively as the sum of the frequencies for each spine. These sums are listed in Table VI. It was found that the spine-making potential tends to increase gradually over the first five of the six levels sampled, only to fall off strongly at the uppermost level. It was concluded that there was a drop in spine-making potential at the top of the sequence considered.

The plots of the shifts in frequencies for all eight characters are closely similar. Multivariate analysis confirmed the inferences based on the

TABLE VI
Spine-making potentials for *Cytheridea acuminata caumontensis*

Stratigraphical level[a]	Potential
1	3.00
2	3.33
3	3.23
4	4.46
5	4.64
6	2.19

[a] 1–6 = oldest level to youngest.

analysis of association in that both point towards a tendency for an increase in the strength of spinosity over time of the first five samples.

The variation in the lateral ornament was also found to show time-related trends. The ratios of the two ornamental categories "regularly pitted shells" and "irregularly pitted shells" were analysed for all levels; they display a significant shift in their relative frequencies over time. The frequency of regularly ornamented individuals tends to increase with time.

It seems possible that the ornamental variation seen here may be of the type known as "quasicontinuous variation" (Falconer, 1970, p. 301). In such organisms, the phenotypic values are discontinuous while the mode of inheritance is continuous. The threshold connects the underlying and continuous variation (genetic and environmental) with the visible and discontinuous variation at the point of discontinuity.

It is convenient to analyse thresholds by the probit transformation of the points of discontinuity (Finney, 1971; Rendel, 1967). This permits us to express the mean phenotype in terms of its standard deviation. The probits for the six levels of *Cytheridea* are listed in Table VII.

TABLE VII
Proportions and probits for the lateral ornamental types

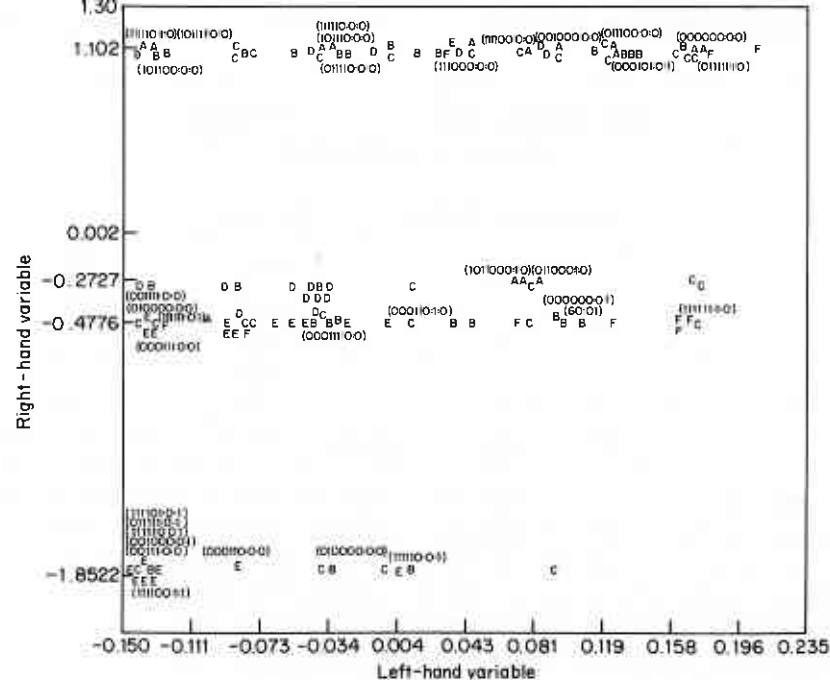| Stratigraphical level	Proportion of irregular phenotype	Probit
1	0.929	6.468
2	0.647	5.377
3	0.625	5.319
4	0.919	6.398
5	0.121	3.830
6	0.500	5.000

Fig. 19 Plot of the canonical correlation "scores" (i.e. transformed partitioned observational vectors) for eight dichotomous measures on the spinosities and ornament of the carapace of *Cytheridea acuminata* from the Miocene of France. Here, the transformed observational vectors for the six anterior spines are plotted against the transformed observational vectors for ornamental category and posterior spinosity. Levels 1 to 6 are here denoted as A to F to avoid confusion with the codings of the anterior spines. Redrawn from Reyment *et al.* (1976).

It is here worth noting that the suggestion presented by the analysis of the anterior spines to the effect that the morphology of the samples from levels one and six is close is not upheld by the lateral ornament. As can be seen from Table VII, the probits for levels one and six are quite different.

Combinations of spines

The four most popular combinations of anterior spines are:

1 1 1 1 1 1 in a proportion of 0.186
1 1 1 1 1 0 in a proportion of 0.149
0 0 0 0 0 0 in a proportion of 0.086
0 1 1 1 1 1 in a proportion of 0.077

These combinations dominate in the material and it may be concluded that the strongest tendency is for all the spines to develop, closely followed by a tendency for the development of five contiguous spines, with the dorsalmost lacking. Less frequent is the third phenotype in which no spines develop. The fourth most common phenotype is the reverse of the second most common and it is here the ventralmost spine that does not appear. The four phenotypes mentioned amount to about one half of the total material from the six sampling levels. The most unpopular phenotypes are those in which the two middle spines are not developed.

Total multivariate analysis

The complete analysis of all the ornamental features showed that there is a shift in them over time for all variables. The variables were also analysed group-wise by means of the method of canonical correlation. The rationale for using this kind of multivariate analysis with discrete data is given by Goldstein and Dillon (1978). The anterior spines constitute a homogenous variational entity. It was demonstrated by Reyment *et al.* (1977) that the posterior spine is not correlated with the anterior spines and that the variations in the lateral ornament are not associated with any of the spines. The canonical correlations between the anterior spines as one set and the posterior spine with lateral ornament as the other set were computed. The plot shown in Fig. 19 derives from this analysis. As to be expected, it reflects the discontinuous structure of the data. It presents the scores for the anterior spines plotted against those for the second set. The formula for the anterior spines (i i i i i i; $i = 1$ or 0) is given for each of the points. There are four discrete classes arranged symmetrically from the stratigraphical point of view. The topmost group contains most of the specimens of level 1, a few from levels 2, 3 and 4 and one ostracod from level 5, and two from level 6. The second group lacks specimens from level 6. It has a few specimens from level 1, the majority of the sample for level 4, a few from level 3 and none from level 5. The third group is devoid of specimens from level 1, but has individuals from all other levels. The fourth, and lowest, group of Fig. 19 lacks specimens from level 1, 4 and 6. It contains mainly ostracods from level 5 and there are a few individuals from levels 2 and 3. This breakdown reflects, of course, the frequencies of the major phenotypes in the material and the complex shift in their relative proportions that has taken place during the slice of Miocene time involved.

Why are the plots spread along lines in the remarkable manner manifested in Fig. 19? This is due to the fact that most of the variability in what is technically a bivariate plot lies with the anterior spines, whereas

the other ornamental categories lack the variational spread of that set. The linear relationships for the two sets are thus selecting four dominant morphological groupings in the material.

Conclusions

Although the analysis of these eight discrete variables does not lead to the desirable situation of our being able to locate a single specimen accurately in time, it does end up with the result that a representative sample can be placed stratigraphically. This is a useful quantitative biostratigraphical result, although insufficient for logging purposes.

The samples of ostracods were not collected from a single stratigraphical sequence nor even from the same immediate area. This is an obvious weak link in the interpretational chain and there is always the possibility that borehole samples from a single well drilled through the part of the Miocene concerned might have yielded more exact results. From the evolutionary point of view, there is evidence of a directed meristic morphological trend in the subspecies over most of the sequence available. The uppermost sample deviates from this tendency. It seems unlikely on our present knowledge that discontinuous variation of the type represented in *Cytheridea acuminata caumontensis* will prove to be practically useful in logging exercises.

Correlations by Moroccan Cenomanian–Turonian ostracods

In the previous example, we considered time-correlated trends in characters of the margin of the carapace. Other properties of the ostracod shell may show discrete variation; for example, in the character of the centrolateral ornament and antero-marginal reticulations, where such occur.

In the Moroccan material, the lateral reticulations and the ridge-structure are dichotomous characters. Although the secular variation in the discrete characters is not useful for biologging (on current indications), it gives a solution, in the present context, to a difficult biostratigraphical problem, namely, that of extrapolating a stratigraphical boundary in a situation of incomplete palaeontological coverage.

In a study of the variability shown by some species of ostracods from the Moroccan province of Tarfaya (Reyment, 1978), I analysed several discontinuous carapace characters. These turned out to be useful for correlating between boreholes and, also, for locating the Cenomanian to Turonian transition in the area.

The drilling was supervised by the Geological Survey of Morocco in 1975 in connection with an evaluation programme for the economic

potential of the shales of the Tarfaya Province. Some of the shales have high contents of organic matter.

The palaeontological problems posed by the study of the material were considerable as the variations found in the species of *Veenia* (*Nigeria*) and *Oertliella*(?) are of the kind described by Liebau (1971, unpublished dissertation). This type of ornamental variation embraces differences in the lateral pattern of reticulations. Liebau (*op. cit.*) refers to macroreticulation, which is defined as a system of meshes with constant location. Macroreticulation can be replaced by microreticulation, which results when one of the calcitic layers of the shell has been reduced. This leads to the formation of meshes of varying diameter, the width of micromeshes always being less than that for macromeshes in the same species. The third and final phase in the reductive process is seen as the development of fine pitting or a smooth lateral surface.

The material

Two species are common in the Tarfaya boreholes, to wit, *Oertliella*(?) *tarfayaensis* Reyment and *Veenia* (*Nigeria*) *rotunda* Reyment. The full analysis (Reyment, 1978) was concerned with variations in the lateral reticulation and also with dichotomous characters of the kind already studied for the French Miocene ostracod (p. 83). For present purposes we shall be mainly concerned with the characters: (1) presence or absence of thick-walled reticulations; (2) whether the carapace is spined or unspined; (3) the presence or absence of a latero-ventral ridge.

The distribution of the two species of interest in the first Tarfayan borehole is as set out below:

Level (m)	Veenia rotunda	Oertliella(?) tarfayaensis
47		×
50		×
138		×
142	×	×
147		
148	×	
150	×	×
152	×	
155	×	×

Oertliella(?) *tarfayaensis* extends throughout the entire sampled interval. For the second borehole, only *O.*(?) *tarfayaensis* was found; it occurs at all sampling levels, namely, 284 m, 290 m, 299 m, 309 m, and 328 m. There is a gap of 88 m over which ostracods are rare or absent.

For the third borehole, both of the species occur in abundance and have the following stratigraphical relationship to each other:

Level (m)	Veenia rotunda	Oertliella(?) tarfayaensis
15		×
19		×
194	×	
197	×	
219	×	
288		×

There is a jump of 175 m over which ostracods occur in such small numbers as to rule out a statistical analysis.

Variational trends

The French Miocene *Cytheridea* could be demonstrated to display trending in some of the discrete characters studied. Analogous variations were also observed for *Veenia* (*Nigeria*) *rotunda*, which are here briefly reviewed. Visual inspection of the material of *Veenia rotunda* suggests that this species may display a unidirectional shift in the frequencies of its ornamental characters over time (Reyment, 1978a). This would be analogous to the shift in frequencies observed for the French *Cytheridea*. The features that seem to show the above-noted shift in frequencies are as follows:

1. the relationship between coarse pitting, fine pitting, and complete smoothness of the lateral surface of the shell;
2. the frequencies of the two antero-marginal rows of reticulations (the relationship being whether two rows or one row develop, or, the antero-marginal area is smooth);
3. in cases where the antero-marginal reticulations are not developed, the proportion between smooth and pitted zones;
4. strength of the sub-central tubercle.

The foregoing variations represent a considerable range in the phenotype of the species, the possible ecological significance of which is taken up later.

I should mention that the morphometrical analysis of the continuous variability in the carapace dimensions did not show obvious trending in either species for length, height and breadth.

Slotting on *Oertliella*(?)

In the Moroccan example, slotting is used to "interweave" the sequences as well as to match up variational patterns. The slotting technique (p. 72) is here applied to frequencies observed for the dichotomous variables, starting with *Oertliella*?

Lateral ornament

Using solely the frequencies for the two morphs for lateral ornament, regularly "reticulated" and "unreticulated", the following slotting result was obtained for boreholes 1 and 2.

Borehole 1	Borehole 2
47 m	
	284 m
	290 m
	299 m
	305 m
50 m	
------Turonian–Cenomanian-------	
transition	
	309 m
138 m	
150 m	
155 m	
	328 m

Here, $\psi = 0.8$ (cf. p. 73). This ordering of the samples condenses levels 284–305 m of borehole 2 into the same time interval as occupied by levels 47–50 m of borehole 1. The stratigraphical boundary between the Cenomanian and Turonian was deduced from the information available from the ammonites and planktonic foraminifers in outcrops and extrapolated from boreholes.

The latero-ventral ridge

The frequencies for the latero-ventral ridge (i.e. whether or not a sub-

alate process is developed) yielded the following slotting result for a value of $\psi = 1.07$.

Borehole 1	Borehole 2
	284 m
	290 m
47 m	
	299 m
	305 m
50 m	
---Cenomanian–Turonian transition---	
138 m	
	309 m
----------------------------	
150 m	
	328 m
155 m	

Thus, the result yielded by slotting the two boreholes on the frequencies of the latero-ventral process is the same with respect to the allocation of the boundary between stages as that given by the frequencies in the lateral ornament.

Lateral ornament and lateroventral ridge

The frequencies for the latero-ventral ridge and those for the lateral ornament were united into a single slotting trial to yield the following ordering. Here, $\psi = 0.99$.

Here again, the Cenomanian–Turonian passage is located in the same stratigraphical position as before.

The slotting results would seem to be a reflection of the observation made earlier on, namely, that there is a secular trend in the frequencies of the discrete characters.

Borehole 1	Borehole 2
	284 m
	290 m
	299 m
47 m	
	305 m
50 m	
---Cenomanian–Turonian transition---	
	309 m
138 m	
----------------------------	
	328 m
150 m	
155 m	

Slotting on *Veenia*

In order to slot borehole 1 with borehole 3, we need to use *Veenia rotunda*. In this case, all four discontinuous characters were employed (see p. 90). Here, for $\psi = 1.15$ (which compares favourably with the results for *Oertliella*(?), the following slotting was obtained:

Borehole 1	Borehole 3
	175 m
	194 m
	197 m
47 m	
50 m	
---Cenomanian–Turonian transition---	
	219 m
138 m	
----------------------------	
150 m	
155 m	

This slotting trial suggests that all sampled levels of the third borehole lie within the Turonian with the exception of the deepest sample, which is located in the passage beds between Cenomanian and Turonian.

Correlating the three boreholes

There is no possibility of bringing about a three-way slotting between the boreholes simultaneously, owing to the disparity in stratigraphical distributions of the two species. However, borehole 1 can be slotted with borehole 2 and also with borehole 3, although boreholes 2 and 3 cannot be compared directly as they lack mutual species at the critical levels. It is, nevertheless, possible to extrapolate from pairwise comparisons so that all three boreholes can be placed in the same stratigraphical context.

First, it was found that three samples of borehole 2 usually assemble above the topmost one of borehole 1, followed by a further sample from borehole 2, followed, in turn, by the second one of borehole 1. Four samples of borehole 2 are located in the Turonian and two of them in the Upper Cenomanian. Using the different ornamental characters selected as being diagnostic for the *Veenia*, three of the samples from borehole 3 fall above the top two of borehole 1.

Summarizing the results yielded by the various slottings, it is possible to establish the following correlation scheme for the three boreholes. (Note, that the levels are not to be thought of as being correlated each to each.) The following chart displays the order of the levels in the three boreholes such as if they were ranged along a single line.

Borehole 1	Borehole 2	Borehole 3
	284 m	175 m
	290 m	194 m
	299 m	197 m
47 m		
	305 m	
50 m		
·········Cenomanian–Turonian transition·········		
	309 m	
138 m		219 m
---		
	323 m	
150 m		
155 m		

Comments on the results

As intimated earlier on, it was not possible to bring about a meaningful correlation between the boreholes on the continuous characters of the carapace alone.

The Moroccan ostracods seem to show habitat characteristics such as occur in selection for phenotypic flexibility in a heterogeneous environment of the type represented by the "*schistes bitumineux*" of the Tarfayan Upper Cretaceous. Both the *Veenia* and *Oertliella*(?) display polymorphism with several morphs possibly adapted to particular subniches (cf. Mayr, 1963, p. 415). This polymorphism seems to be related to that of the *Cytheridea* example but to involve different characters.

Ecophenotypic or habitat variation (Mayr, 1963, p. 142, 1969, p. 152) may be shown by populations of a single species living in different habitats in the same region. Often, these variants may be non-genetic phenotypes. Mayr (1963, p. 415) pointed out that there is little chance that ecological polymorphism will lead to speciation in sexually reproducing diploid organisms; the genes or chromosomes responsible for the polymorphism cannot be removed without spatial segregation of the populations.

In the case of the Moroccan ostracods, at least, there is a possibility that the observed drift in frequencies of characters is a reflection of habitat variation in a varying environment. In that case, the phenotypes would be non-genetic and the trend would lack immediate evolutionary significance. The situation for the French Miocene species is more difficult to assess as the time interval involved is very much greater than for the Cenomanian–Turonian transition.
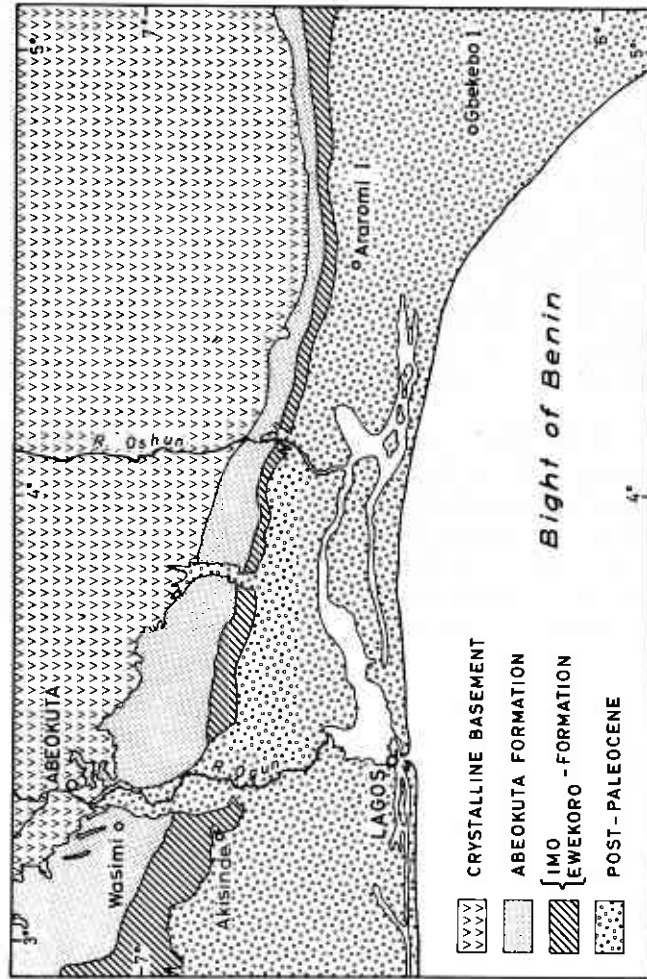
Fig. 20 Geological sketch map showing the locations of the Nigerian boreholes Akisinde, Araromi and Gbekebo in relation to the surface geology.

6
Examples of Biologs

In this chapter, some examples of biologs made using data from the Nigerian boreholes Araromi and Gbekebo are discussed, beginning with the simple method of plotting variations in mean vectors to give an overall size trend. The locations of the boreholes are shown in Fig. 20. The use of discriminant functions for making a biolog from the same data is illustrated. The result is similar to that for the mean vectors on their own and, for straight applications of multivariate analysis, the plots produced by transformed mean vectors will seldom differ markedly from the plots of means.

Where growth and size effects are not of decisive importance, as is so for ostracods, with their discontinuous ontogeny (comparisons based on sexed adults), there may be little advantage attaching to a multivariate statistically produced biolog over a simple plot of the means. Where growth and size differences assume major proportions, such as is the case for organisms with continuous growth (for example, foraminifers), the means require adjusting to make them growth invariant. That is, the effects of differences in size in a sample due to the specimens having different ages are subtracted.

Electrical borehole logs seem to be useful for reconstructing some

properties of the palaeoenvironment (Pirson, 1977). It seems that the properties measured by such logs may have exerted an influence on the morphology of some shell-secreting micro-organisms. Geochemical factors of the host sediment seem to have an influence on the relative abundance of micro-organisms. Relationships between sedimentary chemistry and organisms may be utilized for the production of what can be dubbed an ecolog.

An example using ostracods

Introduction to the problem

Two species of Paleocene ostracods have a wide vertical distribution in the two boreholes Gbekebo and Araromi in Western Nigeria (Ondo State). These are *Cytherella sylvesterbradleyi* Reyment and *Trachyleberis teiskotensis* (Apostolescu). As is typical of the microfossils of the Nigerian Tertiary, the state of preservation of the material is excellent.

Examination of the plots of the secular variation in means of carapace dimensions of samples from the two boreholes showed that the stretch from 369 m to 342 m in the Araromi hole and that running from 736 m to 710 m in the Gbekebo hole, display close similarities in their patterns of oscillation. The question arises as to whether these patterns can be used quantitatively for stratigraphical purposes.

Logging by mean vectors

One way of treating all mean values of a set of samples simultaneously, without the use of multivariate analysis, is by computing the length of the mean vector for each sample, to wit, $(\bar{x}_1^2 + \bar{x}_2^2 + \ldots + \bar{x}_p^2)^{1/2}$, for p variables. These values can be plotted in relation to the positions sampled in the boreholes, as illustrated in Fig. 21. This is a simple way of doing the graphical display, but it does manage to convey a reasonably good impression of the major size trends in the material. Alternatively, the sum of each observational vector can be taken, which also yields an overall size measure. The advantage of such a sum is that a sample variance and confidence intervals can be found.

The vector lengths for the characters length, height and breadth of the carapace are listed in Table VIII. Examination of Fig. 21 shows clearly that the directions of size variation for the two borehole sequences are close and that a change in the direction of variation of the mean vector
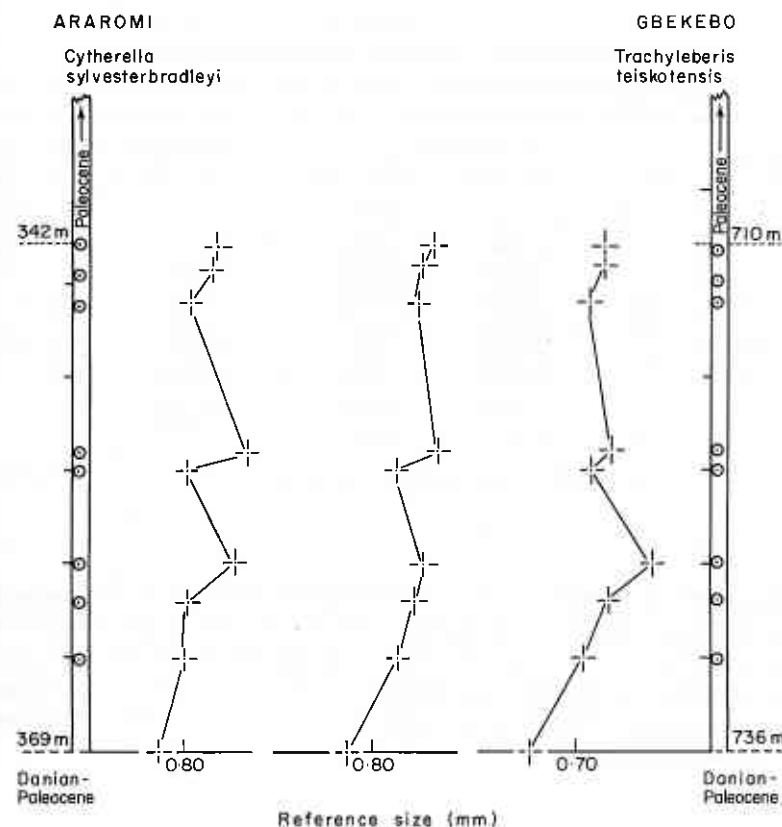
Fig. 21 Fluctuations in vector lengths for two species of Nigerian Palaeocene ostracods plotted against location in the Araromi and Gbekebo boreholes. The horizontal axis indicates deviations from the grand means for each species. The two left-hand plots are for *Cytherella sylvesterbradleyi* Reyment, the right-hand plot for *Trachyleberis teiskotensis* Apostolescu.

length of the Araromi *Cytherella* is mirrored in the Gbekebo *Cytherella*.

In order to broaden the basis for comparison, a species of ostracods of another genus was selected for comparison. The trachyleberidid *Trachyleberis teiskotensis* (Apostolescu) proved to have a pattern closely comparable to that displayed by the sequences for *Cytherella* (Fig. 21).

The initial lining-up of the two boreholes was done in the following manner. Firstly, the passage from Danian to Paleocene was identified by means of the microfauna. This passage is rather rapid in the Gbekebo borehole being identifiable over only a few decimetres of its depth. The passage in the Araromi borehole is seemingly less abrupt. Once the initial

TABLE VIII

Vector lengths for *Cytherella* and *Trachyleberis* from the Nigerian boreholes

Level[a]	Cytherella		Trachyleberis Gbekebo
	Araromi	Gbekebo	
1	0.777	0.777	0.649
2	0.798	0.824	0.707
3	0.806	0.844	0.726
4	0.858	0.853	0.780
5	0.802	0.843	0.718
6	0.870	0.867	0.739
7	0.808	0.853	0.718
8	0.828	0.858	0.728
9	0.831	0.862	0.729

[a] Level 1 denotes the oldest occurrence

arrangement had been made, the variational patterns of the ostracods were slid back and forth alongside each other until a good match was obtained. For just two species, such a match could be spurious. For a large number of species, the likelihood of an incorrect matching of patterns becomes progressively less as the number of concordantly varying species increases.

Angles between mean vectors

A further useful subsidiary tool for logging studies is provided by the angles between mean vectors. You can easily calculate these angles by means of the following cosine formula of vector analysis, namely,

$$\cos \theta = lm/[(l^1l)^{1/2}(m^1m)^{1/2}]$$

where l and m are the two vectors between which the angle is to be calculated. Practical details of doing the calculation are given in Jöreskog *et al.* (1976). The results for the present material of *Cytherella* are the following set of angles: 0, 2.60, 1.40, 1.05, 2.97, 3.67, 2.88, 3.30, 3.56. It will be seen that all are very small indeed, ranging between 0° and 3.6°, which offers support for the hypothesis that the various levels in the boreholes appear to contain microfossils representing organisms that lived at approximately the same time and under the same ecological conditions.

Analysis by canonical variates

The entire data (length, height and breadth of carapace) for the two sequences of *Cytherella* (151 and 93 specimens respectively) and the one sequence of *Trachyleberis* (92 specimens) were analysed by the method of canonical variates (cf. Blackith and Reyment, 1971 p. 88).

The plot of the transformed mean vectors for the first two canonical variates (not scaled to unit variance), shown in Fig. 22, indicates that the first axis discriminates firstly with respect to taxonomical identity and, secondly, chronological order, so that time-dependent shifts in the data show up as clustering of the intercepts along this axis. The second axis also reflects this taxonomical trend to a certain extent, but it tends to
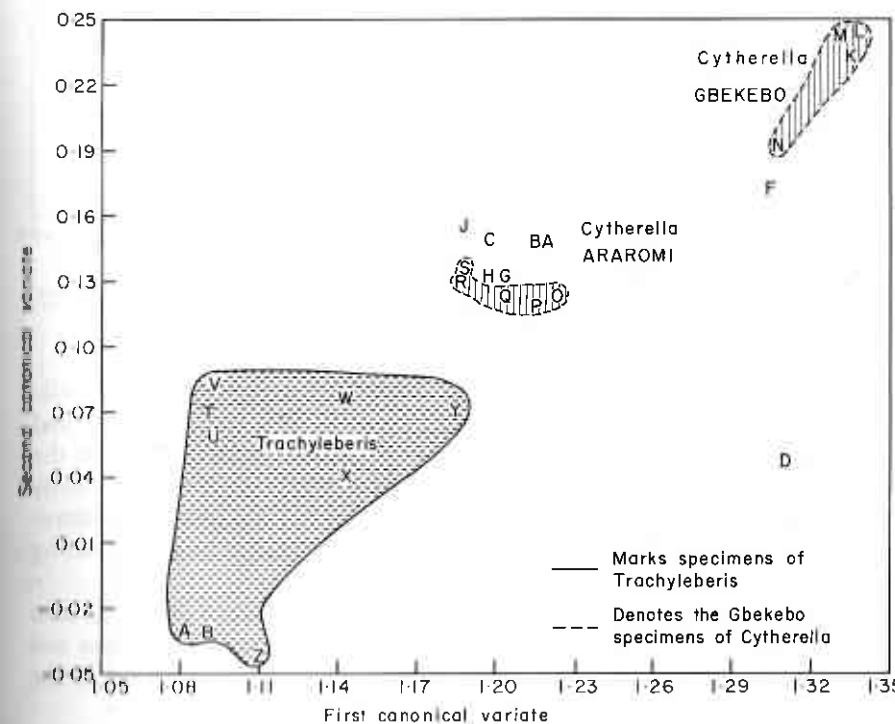
Fig. 22 Plot of the first and second canonical variate means for *Trachyleberis* and *Cytherella* from the boreholes at Araromi and Gbekebo. The letters A to Y denote the ordering of the samples time-wise. Thus, the *Cytherella* from Araromi are marked by the letters A–H, *Cytherella* from Gbekebo by I–Q, while *Trachyleberis* from Gbekebo are denoted by the letters R–Z.
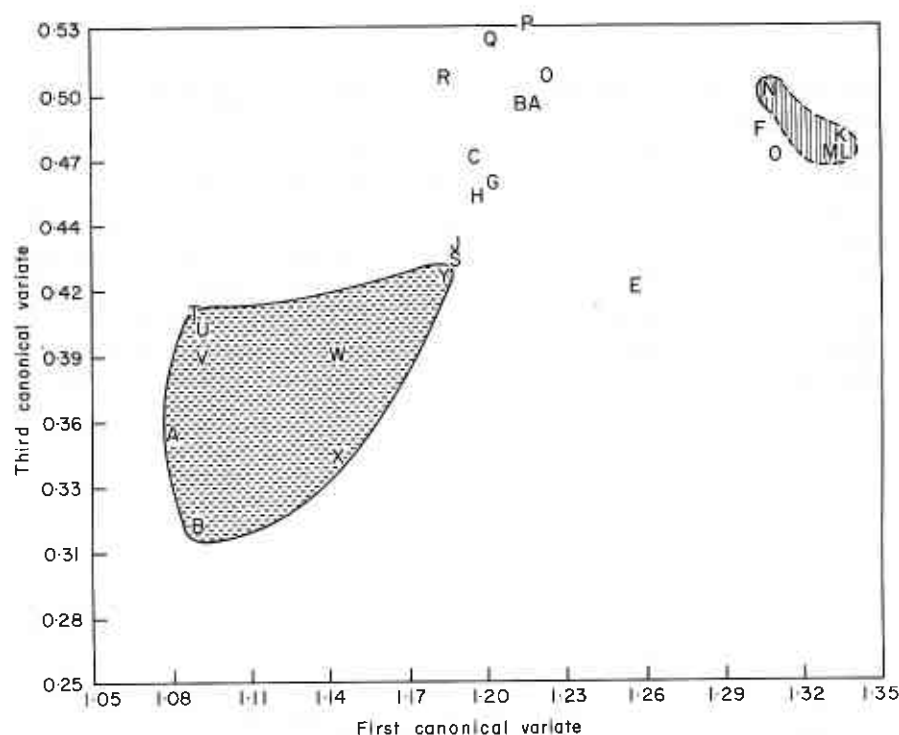
Fig. 23   Plot of the first and third canonical variate means for the data of Fig. 22.
The lettering has the same significance as in Fig. 22.

underscore this differentiation more clearly. The individuals of *Cytherella*
from Gbekebo form two clusters, separated widely from each other and
with the specimens from Araromi lying in between. This suggests that the
Gbekebo *Cytherella* are registering the influence of some factor, possibly
ecological, that affected them during the interval of time under considera-
tion. The Araromi borehole may not have been noticeably influenced by
this incident.

The plot of the first and third canonical means, displayed in Fig. 23 also
presents the taxonomical differentiation observed above, but it does not
bring out the ecological factor. This strengthens the interpretation of the
second canonical vector as a palaeoecological indicator.

Using a simple application of discriminant functions outlined in Black-
ith and Reyment (1971), biologs were made for each of the three
sequences by substituting mean vectors in the discriminant function
constructed between the youngest and oldest samples. The substituted
mean values for the two sequences of *Cytherella* as well as the

Trachyleberis are shown in Fig. 24. The swings or directions taken by the
oscillations for all three curves are identical, although there are differ-
ences in the amplitudes. These are more violent for the *Trachlyeberis*
than for the *Cytherella* as can be demonstrated quantitatively by means of
the generalized statistical distance. For the maximum amplitudinal differ-
ence for the *Cytherella*, $D^2 = 2.12$ and for the *Trachyleberis*, the corres-
ponding generalized distance is $D^2 = 20.72$. We note also that despite the
oscillatory nature of the three sequences, all three display a trend to the
right, though this is not pronounced for the *Cytherella*.

Remarks

The present example is a very simple case of the way in which a biolog
can be made and how it can be used to correlate between two boreholes
within the same sedimentary basin. The disturbed nature of the oscilla-
tions for the *Trachyleberis* suggests that this species may have reacted
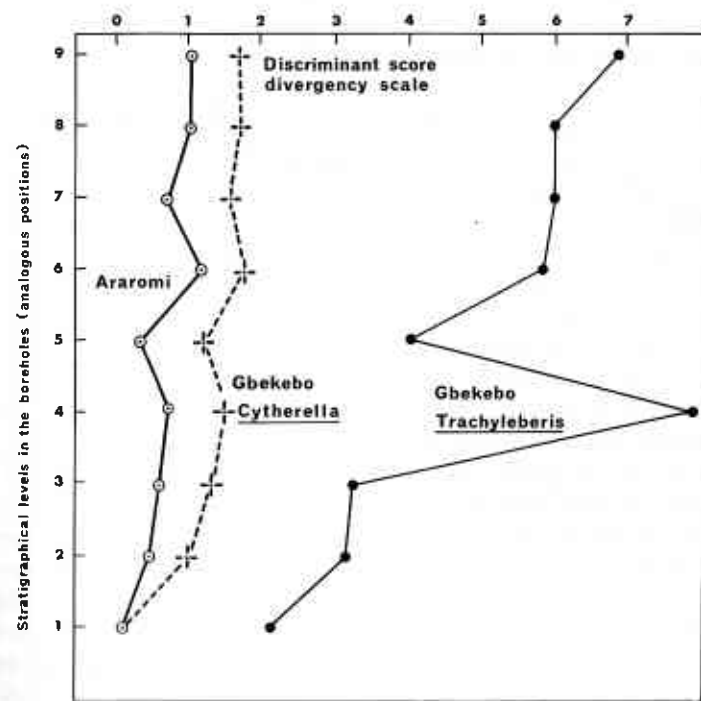
Fig. 24   Biologs constructed from discriminant function scores for ostracods from
the boreholes at Araromi and Gbekebo. The scale of deviations is arbitrary; it is
based on unstandardized deviations from the discriminant mean.

strongly to fluctuations in the environment. The same factors caused morphological reactions in the *Cytherella* but these were less pronounced. A final point needs making here. It is that the three sets of plots were made using three separate discriminant functions and so the amplitudes cannot normally be directly compared. It would be better to use canonical variates for all samples for all species. In the present case, this did not lead to important differences in the plots. The discriminant scores for each plot could also be standardized. Providing there are sufficiently pronounced environmental shifts, it is a relatively easy matter to make a useful biolog. If these environmental variations are not merely of local significance, it is possible to identify their effects in borehole sequences within a not too-restricted area and thus bring about a means of correlating between them by the biologs.

The growth-invariant biolog and foraminifers

Introduction to the problem

The previous example shows how adult ostracod carapaces can be used for making a meaningful biolog. For ostracods, growth and size differences are slight within a certain growth stage and, as already noted, multivariate plots do not differ markedly from those yielded by the means alone. In the present example, we shall be concerned with a far more difficult problem, notably, the preparation of a biolog for a species of bolivinid foraminifers. As you will have gathered from the discussions in Chapters 2 and 3, a major difficulty in interpreting series of morphometrical oscillations, univariate or multivariate, comes from the need for being able to identify genuine morphological changes in an organism as distinct from random variation and spurious shifts in averages resulting from varying mixtures of growth stages, and the like, in a set of samples. Thus, means and variances will be a function of the spectrum of age distributions, the proportions of tests from phases of the life cycle for foraminifers, and analogous complications, in a particular sample. One may also have to contend with morphological manifestations of damage to the test during life, post-mortem damage, the cessation of growth after reproduction and, reworking. All of these factors must be taken into account in a study of the kind reviewed here.

Thus, the statistical situation for continuously growing organisms such as foraminifers is fraught with difficulty, as any given sample will be confounded by size differences due to a variety of causes. A sample of a

species of foraminifers is not statistically homogeneous in that it will be made up of a mixture of growth stages and growth-inhibited morphologies, even though it may be biologically homogeneous, in the broad sense. The method of growth-invariant canonical variates used in the case study treated here was introduced in Chapter 4 (p. 60). The complete analysis of the material is given in Reyment (1978b).

The material

The data derive from 36 borehole sampling levels of the Maastrichtian benthonic foraminifer *Afrobolivina afra* Reyment from the Nigerian well Gbekebo I. The nine characters, illustrated schematically in Fig. 25 were
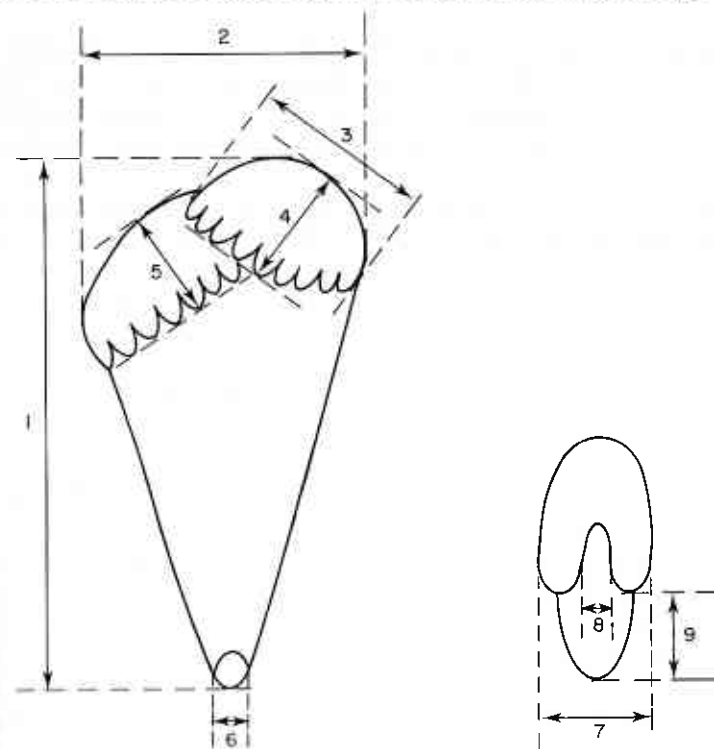
Fig. 25 Sketch showing the locations of the measurements made on *Afrobolivina afra* Reyment. Var 1 = length of test; var 2 = maximum width of test; var 3 = width of final chamber; x_4 = height of final chamber; x_5 = height of second last chamber; x_6 = diameter of proloculus; x_7 = width of test at right angles to the plane of $x_1 - x_2$; x_8 = width of aperture; x_9 = measure of cameral overlap (for last and second last chambers).

measured on each of 778 specimens. The reason for selecting these nine variables is that a preliminary statistical analysis showed that they have reasonably compatible variances, apart from the length measure, and that they give a reasonable representation of the interlocking growth patterns in the species.

The variation in the proloculus is strongly bimodal, but owing to the relative rarity of microspheres (often not more than 5% of a sample) this may not show up. This topic is given detailed treatment in the example analysed fully in Chapter 8.

The growth-invariant curve

With one growth-vector extracted by internal principal component estimation (see p. 61), the first coordinate (see p. 63) was used to construct a size-standardized chronocline for the 36 samples by plotting the values of the coordinates against their stratigraphical positions (i.e. the sampling sites in the borehole sequence).

The details of the steps are given on p. 62. The biolog for the material considered here is illustrated in Fig. 26. In this figure, I have also drawn
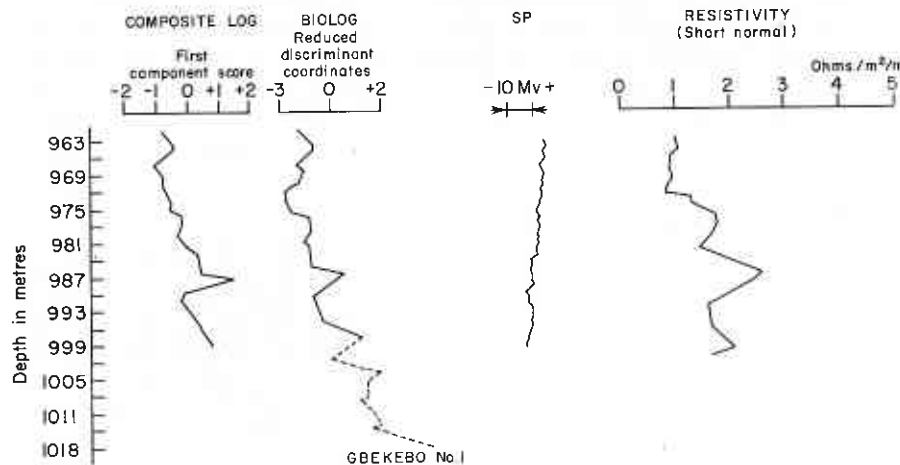
Fig. 26 Growth-reduced biolog, composite log, and the electrical logs for self potential and short-normal resistivity for 36 sampling levels for the biolog and 27 levels for the other three logs. The resistivity log is here represented in a discontinuous state, being made up of the observations at the sampled levels in the borehole at Gbekebo. (Note that the slope of an SP log is opposite to that of the corresponding resistivity logs.)

in the corresponding sampling points for the short normal resistivity curve. As pointed out in Reyment (1978b), there seems to be good agreement in the oscillations shown by the short normal observations and the growth-reduced biolog. There is also an unmistakable similarity with the self potential log, also drawn in in Fig. 26 (remembering that it has an inverse gradient in relation to resistivity logs). All curves display an overall monotonic trend (see below for analysis) in the statistical sense. The biolog for *Afrobolivina afra* extends below the two physical logs, for which observations were not made to the bottom of the borehole. It is important to bear in mind when comparing the biolog with the physical logs that the latter represent records of continuous observations, whereas the biolog derives from discrete sampling points. Direct comparison by overlay, comparing amplitudes of oscillations, etc. can, therefore not be made, although directions of trend can be compared at points on the physical logs that correspond with the levels at which the samples of sediment were taken. It is therefore not permissible to compare the curves by looking for pattern agreements when they are placed over each other. Such a test of agreement can only be made by examining the correlations between analogous points on the curves. This is demonstrated more clearly in Figs. 28 and 29. I am belabouring the question of comparisons of logs because this is a common cause of misunderstanding. In some cases, a smoothed continuous log can be used for a valid comparison with a biolog, but there may be need for care, as is shown in Chapter 8 (p. 49). Quenouille (1959, pp. 24–26) gives some useful tests of monotonic association which are often valuable for analysing biologs.

Conclusions

Residual variation expressed by the growth-reduced discriminant coordinates seems to be due to the response of the foraminiferal organism to environmental factors. Physical logs of the sediments encountered in a borehole might reflect some properties of the ecological background in which these were deposited.

Perhaps the best way of contrasting the biolog and physical logs is by means of comparing the directions of oscillation of the logs at consecutive sampling points. If such a comparison is made here for the biolog and the short-normal resistivity log, there is less than a 10% divergence in gradients of the oscillations. A further refinement worth considering involves standardization of the logs, an approach given further thought in the next chapter.

The curve denoted "composite" log in Fig. 26 was obtained from the first principal component scores for the first growth-reduced discriminant coordinate and the corresponding readings for the two electrical logs. It provides a useful means of summarizing the biological and physical information for the sediments pierced by the borehole. It is also a multivariate condensation of the logs (amounting to more than 80% of the total variation).

There is without doubt a lot of hard work involved in preparing a biolog of the kind presented in this case history, at least, if the whole task falls to the lot of one person. In industrial laboratories, where several persons can be assigned to a task, the effort involved is not prohibitive.

As regards the characteristics of the environment inhabited by the individuals of *Afrobolivina* studied here, the palaeoecological analysis indicates that it was almost certainly normal marine and largely homogeneous. The sediment is a homogeneous clay–shale without sandier or calcareous intercalations over the stretch sampled. The Late Campanian through Early Maastrichtian was a time of very widespread transgression in West and North Africa and the time interval involved for the 36 samples analysed here occurred at the maximum of the transgression, the most extensive of the Cretaceous epicontinental floodings.

The seeming usefulness of the short-normal resistivity log as a kind of general measure of long-term environmental fluctuations is limited to sedimentary sequences in which variations in lithology do not occur. In the treatment of the Nigerian boreholes Araromi and Gbekebo, in which the sediments are remarkably homogeneous, the short-normal log has proved itself to be an invaluable adjunct. Further remarks on the palaeoecological interpretation of the pertinent part of the Gbekebo borehole are given in Reyment (1966b). The potential usefulness of the electric borehole logs as palaeoecological indicators is discussed in Reyment (1978b). Chapter 8 contains an example of the effects of a sandy intercalation on the resistivity log in relation to the biolog (p. 150).

Afrobolivina afra is the dominant faunal component in almost all of the samples considered. It is accompanied by *Gabonella*, *Valvulineria*, and species of *Veenia* and *Cythereis*, some individuals of which show evidence of predation.

Biolog from stabilized canonical vectors

Introduction to the problem

The importance of carrying out a proper introductory analysis of multivariate data was emphasized in the statistical section in Chapter 3. Such

an analysis should comprise

1. A full-scale graphical study of each variable (histograms) and of pairs of variables plotted as scatter diagrams.
2. A careful appraisal of the variables selected for a particular analysis with the end in view of separating out such variables as are redundant.

Often, the two lines of analysis will not only disclose the presence of multivariate outliers in the data, which can sometimes be a serious source of instability in a multivariate analysis, but may also lead to a more parsimonious selection of variables. It has been my observation that the elimination of a few redundant variables can result in a greatly improved canonical variate analysis in that the eigenvectors of $W^{-1}B$ will be stable, thus leading to more reliable results.

The problem we shall consider now concerns the same set of data on *Afrobolivina afra* studied in the foregoing example and the same set of variables (see Fig. 25). The present account is based on the analysis of Campbell and Reyment (1978).

The eigenvalues and last three eigenvectors for all nine variables are listed in Table IX. The smallest eigenvalue accounts for only 1.8% of the variation within groups. The eigenvector corresponding to the smallest eigenvalue (which I shall refer to as the smallest eigenvector, although this is not a strictly logical terminology) reflects a contrast between variables 2 and 3, that is to say, between the width of the test and the width of the last chamber. These loadings are quite large (italics in Table IX) and, bearing in mind the remarks of Chapter 3 (p. 47), it may be suspected that if the corresponding between-groups sum of squares is small, as is the case in the present example (see Table IX), instability in the corresponding canonical variate coefficients may result.

The between-groups sums of squares for all principal components (Table IX) shows that about 40% of the between-groups variation is associated with principal component VII, and 24% with principal component I. The variation for principal component VII results mainly from variable 7 (italics in Table X), with much lesser contributions from the other variables.

The canonical variate analysis can be carried out in terms of the principal components. The coefficients for the first canonical variate (denoted as "CV I Principal Components" in Table X) highlight the contribution from principal component VII. The coefficients for the original variables are determined explicitly from the principal component canonical variate coefficients. Any inflation in these coefficients is exposed in inflated coefficients for those original variables which contribute to the eigenvector from which the principal component is derived.

TABLE IX

Eigenvalues and last three eigenvectors of the within-groups correlation matrix for all nine variables for *Afrobolivina afra*; between-groups sums of squares for each principal component.

		1	2	3	4	5	6	7	8	9
Eigenvalues		4.32	1.24	1.00	0.69	0.48	0.44	0.38	0.29	0.16
		var 1	var 2	var 3	var 4	var 5	var 6	var 7	var 8	var 9
Eigenvectors	VII	0.28	0.26	0.24	-0.21	-0.25	0.03	-0.74	0.33	0.20
	VIII	0.63	-0.49	-0.44	0.14	0.20	0.29	-0.11	0.08	0.10
	IX	0.10	0.69	-0.70	0.05	-0.08	-0.06	0.04	0.05	-0.10
		1	2	3	4	5	6	7	8	9
Between-groups sums of squares for principal components		1.00	0.25	0.24	0.27	0.26	0.07	1.56	0.18	0.21

Sum = 4.05

TABLE X

Canonical variate coefficients for nine variables, including shrunken estimates

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Canonical roots
CV I' principal components	0.54	-0.11	-0.01	0.11	0.05	-0.05	0.79	-0.18	-0.13	2.28 }75%
CV II principal components	-0.64	-0.32	0.15	-0.15	-0.43	-0.10	0.39	0.04	-0.32	0.76
CV I (Standardized original variables)	0.01	-0.59	-0.09	0.43	0.44	0.07	0.99	-0.51	-0.10	
CV II	0.65	0.82	-0.28	-0.44	-0.40	0.14	0.05	0.28	0.26	
CV I $k_9 = \infty$	0.05	-0.37	-0.31	0.44	0.41	0.06	1.01	-0.51	-0.12	2.25
CV II $k_9 = \infty$	0.57	0.30	0.31	-0.53	-0.35	0.24	-0.03	0.37	0.33	0.69

[a] CV = canonical variate.

110

The first canonical variate accounts for 56% of the between-groups variation. The first two canonical variates taken together account for 75% of the variation between groups. The coefficients for the standardized original variables for the first two canonical variates are displayed in Table X.

We shall now consider the effects brought about by shrinkage. The effect of shrinking the contribution of the smallest eigenvector, and associated eigenvalue, namely, principal component IX, is shown in Table X (here, $k_9 = \infty$ denotes that principal component IX has been deleted from the analysis). It will be seen that the two sets of coefficients for the original variables and $k_9 = \infty$ are similar, except for variables 2 and 3. There is a decrease in the coefficient for the second variable and a change in sign of the coefficient for the second variable and a change in sign of the coefficient for variable 3 in CV II.

Principal component VIII was also eliminated. This was found to have little effect beyond that already achieved by the suppression of the ninth principal component. Referring back to Table IX, we see that the eigenvalue corresponding to principal component VIII is about twice that for principal component IX and it may be concluded that its greater contribution to the total variation within groups is above the level which usually causes instability.

The relatively small standardized coefficients for variables 1, 3, 6 and 9, and the instability of particularly variable 3, suggests that an analysis of the five variables 2, 4, 5, 7 and 8 might be worth consideration. As will be seen from Table XI, 34% of the remaining between-groups variation is associated with the smallest eigenvalue and eigenvector; hence, little or no instability in the coefficients can be expected. In fact, elimination of

TABLE XI
Canonical variates analysis for five variables

		1	2	3	4	5	
Eigenvalues		2.82	0.99	0.46	0.43	0.30	
Eigenvectors		Var 2	Var 4	Var 5	Var 7	Var 8	
	IV	-0.10	0.73	-0.67	0.11	-0.08	
	V	0.82	-0.25	-0.43	-0.27	0.07	
Between-groups sums of squares for principal components		1.05	0.34	0.67	0.11	1.12	
		Sum = 3.28					Canonical roots
CV I[a]		-0.64	0.42	0.43	0.95	-0.52	2.25
CV II		0.99	-0.42	-0.33	-0.21	0.46	0.63
CV I $k_4 = \infty$		-0.62	0.26	0.58	0.94	-0.50	2.21

[a] CV denotes canonical variate.

principal component IV has little effect on the coefficients. This result supports the remark on p. 48 concerning stability and the deletion of redundant variables.

From the aspect of redundancy and parsimony in the number of variables utilized in the analysis, it would appear to be sufficient to measure only the five variables: maximum breadth of the test, the heights of the last two chambers, the maximum width of the test, and the width of the aperture.

Making the stabilized biolog

The stabilized coefficients for all nine variables were used for producing the biolog in the present example. The points for each sampling level were obtained by substituting the original sample mean vectors in the first equation of the stabilized canonical variates. The canonical means for the first canonical variate were plotted against position in the borehole Gbekebo I, as illustrated in Fig. 27. This plot also displays the corresponding observations for the short-normal resistivity curve and an "oscillation log", that is, a record of right and left swings for the biolog and the short-normal plot. A simple chi-squared test of the hypothesis that the oscillationary sequences are significantly different was rejected.

The biolog plotted in Fig. 27 shows a general trend to the left. The oscillatory record of the biolog is matched closely by the short-normal resistivity log, an observation supported by the afore-mentioned chi-squared test.

Testing for trend

Although the trend displayed in Fig. 27 seems to be quite clear it is useful to test it for significance. The existence of a time-dependent shift in morphometrical measures of a sequence of a fossil species is of some consequence for the interpretation of a biolog. If significant trend can be shown to occur in both the suite of morphometrical characters and some physical measure of the environment inhabited by the animals, this would provide further support for the belief that the physical log, in the present case, the short-normal resistivity log, reflects some dominant feature of the environment, possibly related to the characteristics of the sediment.
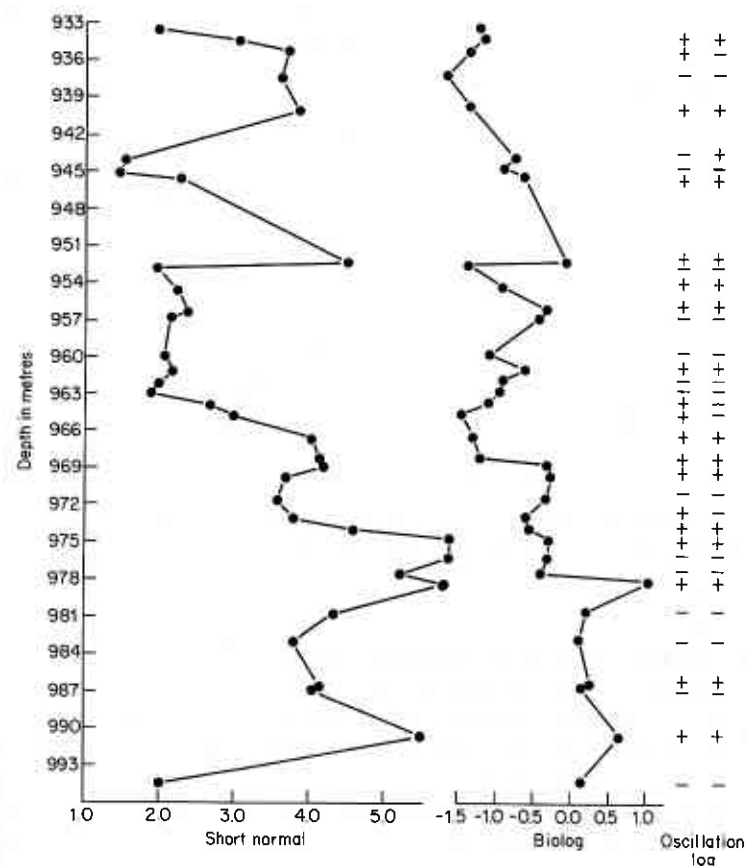
Fig. 27 The biolog formed from the first stabilized canonical variate means plotted against position in the borehole Gbekebo 1, together with the corresponding values of the short-normal resistivity curve and a log of oscillations of these two curves. The left column of the oscillation log denotes the swings to the right or the left for the short-normal curve; the right column does the same thing for the shrunken canonical variate means. The data consist of megalospheres and microspheres. The oscillation log was arbitrarily constructed from level 933 m to level 994 m.

A suitable, albeit large-sample, statistic for testing for trend in a sequence of observations is

$$U = \frac{\sum_{i=1}^{N} (X_i/N - \frac{1}{2}X)}{X\sqrt{(1/12N)}} \tag{23}$$

where the series is observed for the interval $(0, X)$, $X_1, \ldots, X_N$ are the
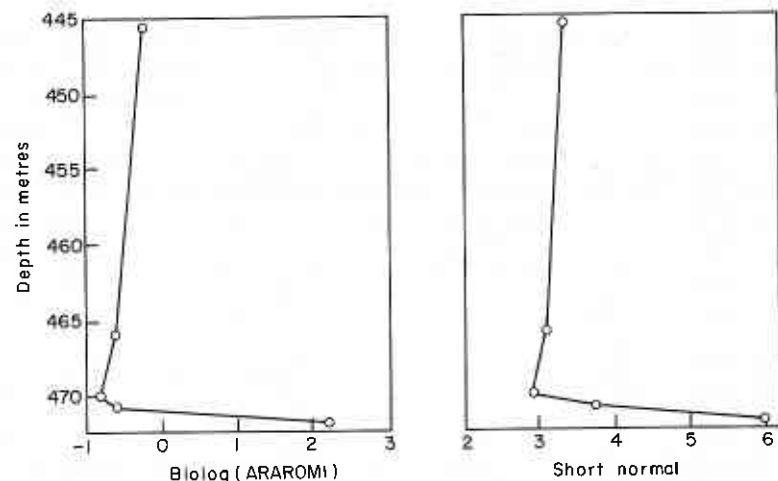
Fig. 28 Comparison between the biolog (canonical variate means) and the short-normal resistivity log (ohms $m^2 m^{-1}$) for the uppermost sampling levels of the Araromi borehole, Nigeria.

cumulative, time-ordered observations, and N is the number of observations. (Cramér, 1946; Cox and Lewis, 1966).

The value U is a standardized normal deviate. If its value exceeds 1.96, this is an indication, at the 95% level of significance, that trend occurs in the sequence of observations.

Test (23) was applied to the 46 consecutive levels in Gbekebo for the set of stabilized first canonical means, and to the 35 consecutive readings for the short-normal resistivity log. For the electrical log, $U = 1.76$, which falls short of statistical significance. For the biolog, $U = 2.05$, which is statistically significant, thus suggesting a possible trend in these data.

The sample sizes are small and this can have a deleterious effect on what is essentially a large sample test.

Another way of examining the structure of sequential data is by means of the serial correlation coefficient. The serial coefficient of correlation can be represented as

$$\rho_j = \frac{\text{Cov}(Y_i, Y_{i+j})}{\text{Var}(Y)} \quad (j = \ldots, -1, 0, 1, \ldots) \tag{24}$$

for the individual time-ordered observations $(Y_1, Y_2, \ldots, Y_N)$.

The first serial correlations computed for both logs display an interesting type of structure, the essential feature of which is that the serial correlation coefficients decrease successively and evenly from highly

TABLE XII

Serial correlation coefficients for the biolog and short-normal resistivity log

	Short normal readings		Biolog	
Lag j	ρ_i	$\sqrt{(N-j)}\rho_i$	ρ_i	$\sqrt{(N-j)}\rho_i$
1	0.84	5.06[a]	0.72	4.96[a]
2	0.71	4.24[a]	0.60	4.13[a]
3	0.61	3.61[a]	0.53	3.64[a]
4	0.55	3.24[a]	0.40	2.76[a]
5	0.58	3.32[a]	0.35	2.35[a]
6	0.52	2.97[a]	0.35	2.36[a]
7	0.48	2.70[a]	0.30	2.00[a]
8	0.43	2.37[a]	0.25	1.69
9	0.24	1.37	0.22	1.50
10	0.04	0.36	0.23	1.52

[a] = significant on at least the 95% level of significance.

significant values for small lags to non-significant values for large lags (Table XII). This pattern suggests that there is a strong dependency between adjacent observations and that this lessens gradually the further apart two observations are located. The sequences are thus far from being random in nature and the non-significant value of U for the short-normal readings is certainly too low, probably being a reflection of the small sample size.

The foregoing analysis is indicative of parallel oscillations in both sets of logging data. This result leads us to the thought that the directed variation in *Afrobolivina* could be environmentally controlled and that the environmental trend is measured by the physical log (see Pirson, 1977).

Morphometric variation and resistivity logs

Two details of comparisons between biologs and short-normal resistivity logs, shown in Figs 28 and 29, will now be described in order to clarify further the discussion of the foregoing section.

Figure 28 compares the two logs for the five uppermost sampling levels of the Araromi borehole. The biolog scale is in terms of values of the canonical variate means and the short-normal resistivity scale is in ohms $m^2 m^{-1}$. Both sets of values fall sharply from level 472 m to 471 m

and then slightly less steeply from levels 471 m to 470 m. Both rise to 466 m and then less pronouncedly to 446 m.

The Gbekebo comparison shown in Fig. 29 is less straightforward than the preceding example. From 891 m to 890 m the resistivity observations remain unchanged while the biolog increases. Both logs then increase sharply from 890 m to 887 m and then fall off to 885 m. Thereafter, the two logs zig-zag identically through 884, 883, 882, 880.5 and 880 m. The directions diverge for the step from 880 to 879 m.

In Fig. 28, the two logs can be compared by overlay and it will be seen that they differ very little from each other. The second example cannot be compared by overlay in a meaningful manner, although eight of the nine steps of the two "walks" take the same directions. This can be seen to be due to differences in amplitude leading to a right-hand or left-hand displacement of the logs, respectively.
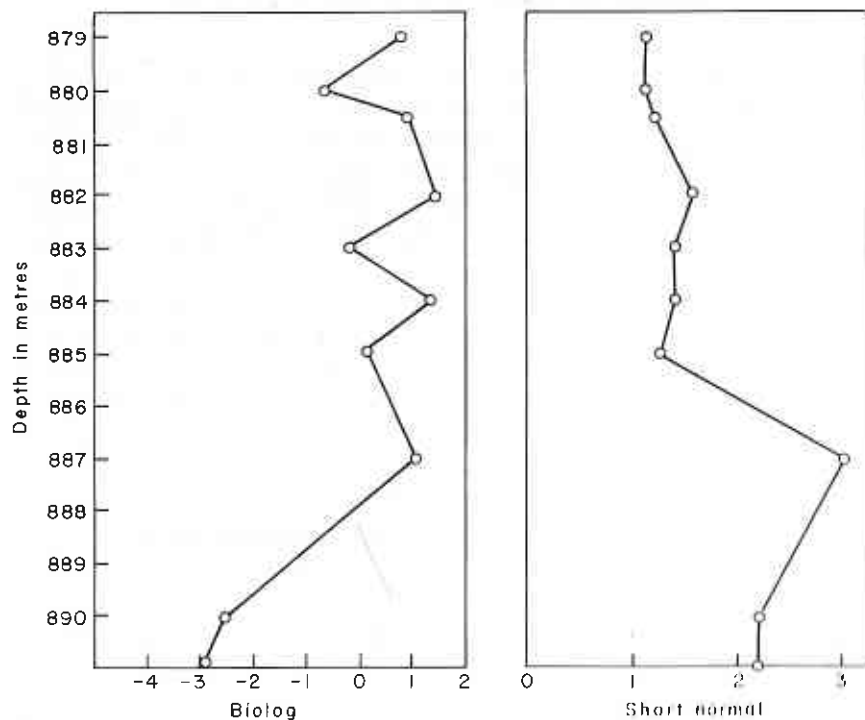
Fig. 29 Comparison between the biolog (values of canonical variate means) and the short-normal resistivity curve (ohms m² m⁻¹) for 10 levels in the Gbekebo borehole.

Quantitative comparison of logs

Making quantitative comparisons of stratigraphical logs such as borehole logs is not as simple as one might think on first consideration, particularly if there is trending with a cyclic component. The reason for this is that there may be sharp lateral displacements in physical logs due to causes beyond statistical control and these generate a clumping effect and a proliferation of atypical values. A measure of association, such as the product–moment correlation coefficient, performs poorly in situations of this kind.

Non-parametric statistical methods seem to offer the most suitable means of studying borehole sequences, various forms of rank correlation coefficients coming first to mind. Useful presentations of rank correlation techniques are given by Bradley (1968, Chapter 5 and pp. 284–287) and Maxwell (1967, Chapter 8). Unfortunately, these methods also tend to be adversely affected by non-geological lateral displacements of logs.

Dichotomizing the observations can sometimes provide a help, so that right-directed swings are coded as $+$, and left-directed swings as $-$; although this loses information on the trending component, this information is often inaccessible for the reason outlined above. It should be noted that dichotomization of a log leads to a fundamental change in the statistical properties of a sequence of observations as the value taken at each point in the time series depends on the value adopted by the previous observation. Such a series forms a two-state Markov chain.

For dichotomized data, a 2×2 contingency table can be used to test an hypothesis of identical populations with respect to directions taken by serially ordered fluctuations for two logs (Bradley, 1968, section 8:2). In the case of a set of the data for *Afrobolivina* from Gbekebo from 80 levels, there are 41 left swings and 39 right swings for the 9-variate biolog and 38 left swings and 42 right swings for the short-normal curve. The resulting χ^2 is not inconsistent with both sets of observations having been drawn from the "same" population.

Variants of the "runs test" can be utilized for testing for randomness in a series of dichotomized logs (Bradley, 1968, sections 11:3 and 12:2). Maxwell (1967) shows how a contingency table of naturally ordered observations can be analysed for trend.

A further useful approach for studying distributions of the type of interest here is given by Cox (1970), who shows how the logistic transform can be applied to binary data to test a 2×2 contingency table for significant differences. Define $Z = \log[(R + \frac{1}{2})/(N - R + \frac{1}{2})]$ where R is, say, the number of right-hand swings on a sequence (sample) of size N. The variance for this expression is $V = [(N + 1)(N + 2)]/[N(R + 1)(N - R + 1)]$.

For two logs, the logistic difference $\Delta = Z_1 - Z_2$ with standard error $S_E = \sqrt{(V_1 + V_2)}$, when expressed as the ratio Δ/S_E, corresponds to a standardized normal deviate, with a two-sided significance test, for testing $\Delta = 0$. When Cox's test was applied to the above-mentioned data from Gbekebo for the two logs mentioned a non-significant value of Δ was obtained, as to be expected, noting that the proportion of right-directed swings for the biolog is 0.51 and for the short-normal curve, 0.48.

For the purposes of testing agreement in patterns, interest attaches to the agreements in directions as a single characteristic and non-agreements as the other state. For the 9-variate biolog and the short-normal log, agreement in directions amounts to a proportion of 0.86 of the total number of observations. As to be expected, this leads to an overwhelmingly significant chi-square test-result of association. The *SP* log, although it shows more divergences (the proportion of agreements is 0.74), also yields a highly significant chi-square for simple association.

It should perhaps be noted that χ^2 is not useful as a measure of the degree of association between two characteristics, although it is a good indicator of the significance of the association. It can, however, be transformed to a measure of association which is independent of sample size, the so-called phi coefficient of the behavourial sciences. This is defined as $\phi = \sqrt{(\chi^2/N)}$, where χ^2 is computed between characteristics A and B, and N is the total sample size (Fleiss, 1973, p. 42).

In discussing non-parametric measures, mention should also be made of the likelihood ratio test known as the "G-test" (Sokal and Rohlf, 1969); G offers some advantages over chi-square.

The runs test, referred to above, can also be applied to the oscillation patterns for the logs. The 80 observations for the biolog compared with the short normal curve contain 19 runs, the corresponding suite for the *SP* log being 26 runs.

In a series of N observations, of which N_1 denotes the number of agreements "a" in oscillatory directions and N_2 the lacks of agreement "b", the obtained a's and b's will occur in random order if each position in the series had equal likelihood of being occupied by an a or b event.

For the short-normal comparison, $N_1 = 69$ and $N_2 = 11$, with 19 runs. A table of the total-number-of-runs test shows that the hypothesis of randomness in the sequence is strongly rejected. For the *SP* log, $N_1 = 59$ and $N_2 = 21$, and 27 runs, the hypothesis of randomness is also rejected. (For references to tables of runs, see Bradley, 1968, pp. 362–363.)

In summary, I think it will be appreciated that the question of demonstrating agreement in patterns of oscillations for comparisons of different types of logs may be relatively easy to do graphically, but rather rounda-

bout when it comes to analysing the actual measurements. Other possibilities are also available but the foregoing serves to present the main ideas involved.

Foraminiferal life cycle and size differences

Nyholm (1962), for living foraminifers, and Reyment (1966c), for fossil foraminifers, have demonstrated how morphological differences can be expected to be linked to the phase in the foraminiferal life cycle from which a particular test (i.e. shell) derives. As noted earlier (p. 104), it is necessary to make a preliminary statistical analysis of a foraminiferal species being chosen for preparing a biolog in order to obtain an opinion on the extent to which size differences can be expected to influence the results. The present material of *Afrobolivina afra* has shown itself to be deficient in individuals belonging to the microspheric phase. The plot of the first and third principal coordinates shown in Fig. 30 for 99 individuals from the four uppermost levels of the borehole Araromi I, Ondo State, Western Nigeria, only contained eight microspheric tests (note, that several points are superimposed in Fig. 30), which is less than 10% of the total sample. A direct count in 35 samples from Gbekebo showed the material to be made up of 16% of microspheric tests. I have offered this figure as a representative example of the entire material of *Afrobolivina* examined in my quantitative analysis of this species. The situation pertaining in other samples is mostly analogous.

It is of interest to note, briefly, the few points occurring in the bottom right-hand corner of Fig. 30. These are exceptionally short tests (cf. Reyment, 1959) with megalospheric proloculi.

Ecolog

Introduction to the problem

We shall now consider an example in which the correlations between frequencies of organisms, on the one hand, and geochemical components of the host sediment, on the other, are used to produce what can be termed an *ecolog*, that is, a log in which chemical elements are related to fluctuations in the frequencies of species of interest over time. The
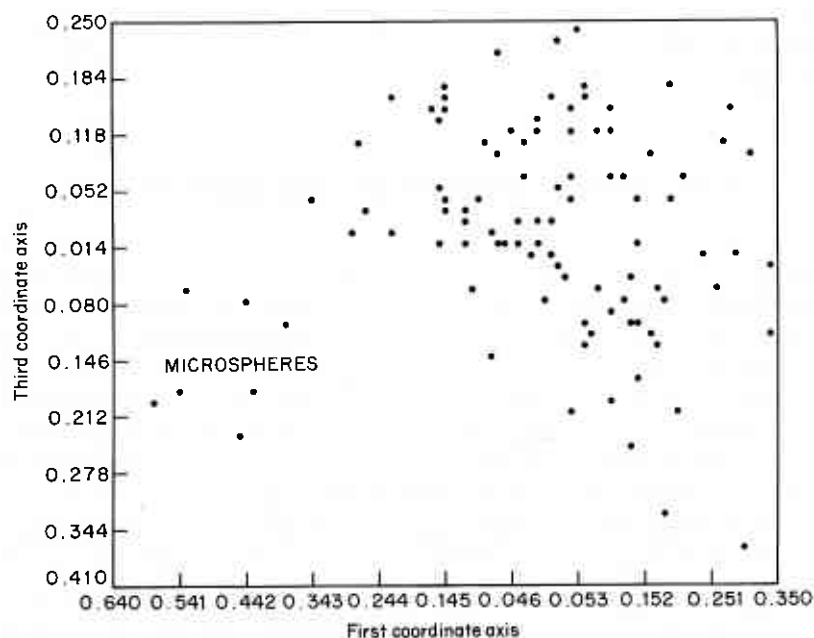
Fig. 30 Plot of the first and second principal coordinates for a representative sample of *Afrobolivina afra* from the Araromi borehole. Note the relative scarcity of microspheric tests in relation to megalospheres.

example is taken from Reyment (1976). In the present connection, it is presumed that some aspects of the interstitial ecology of the palaeo-environment will be recorded in the concentrations of trace elements in the sediment (cf. Waszowiak, 1962). Twenty-six levels in the borehole Gbekebo I were analysed with respect to the 14 elements Si, Fe, Mg, Ca, Na, K, Ti, P, Mn, V, Mo, Sr, Pb and Zn. The frequencies of the foraminifers *Afrobolivina afra* Reyment, *Gabonella elongata* de Klasz and Meijer, and *Valvulineria* sp. nov. were recorded for these levels. The ostracods, being relatively rare, were pooled together for the purposes of the analysis. The most common species is *Ovocytheridea nuda* Grekoff, followed by *Veenia (Nigeria) nigeriensis* Reyment, *Veenia (Veenia) warriensis* Reyment and *Brachycythere armata* Reyment.

The aim of the study was to facilitate the graphical expression of a difficult palaeoecological and biostratigraphical problem. In one direction, it was thought to be of interest to show how all variables considered in the same connection vary over time. In another direction, interest was concentrated on tracing the covariation in frequencies and geochemical indicators.

The material

The correlations between the geochemical variables are shown in Table XIII, with the statistically significant values printed in italics. It will be observed that Si is significantly and positively correlated with Fe, Mn and V, and negatively and significantly with Mg, Ca, Na, P and Sr. The variable Fe is significantly positively correlated with Mn, V and Mo, and significantly negatively correlated with Na, P and Sr, while Na is significantly positively correlated with K, P and Sr. Further significant correlations are as follows. Ti is positively correlated with V, Mo, Pb and Zn, and P is positively correlated with Sr. Manganese is positively correlated with V and Mo. In its turn, Mo is positively correlated with Pb and Zn and negatively with Sr.

For the frequencies of microfossils, the following significant relationships between sets occur (Table XIV). There is a negative correlation between ostracods and Zn, while *Afrobolivina afra* is not correlated significantly with any of the geochemical variables. The frequencies for the *Valvulineria* are correlated positively with Mn and Fe and negatively with Ca, while *Gabonella elongata* is positively correlated with Fe and Mo, and negatively with Ca and Sr.

Results of the analysis

The analysis considered here aims at establishing the correlations between sets, here chemical variables, on the one hand, and species frequencies, on the other. In the present case, there are 14 chemical variables in one set and four species frequencies in the other.

The method of canonical correlations was only briefly noted in Chapter 3 (p. 48). Some of the essential details of the computational steps will now be presented (see Cooley and Lohnes, 1971). In the following, R_{11}, R_{12}, R_{21} and R_{22} denote the partitioning of the correlation matrix R.

The vector variable z_1 contains the chemovariables and the vector variable z_2 the frequencies of the microfossils. The roots of the determinantal equation for the two sets

$$|R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} - \lambda_i I| = 0 \qquad (25)$$

are $\lambda_1 = 0.849$, $\lambda_2 = 0.752$, $\lambda_3 = 0.414$ and $\lambda_4 = 0.182$. The first two of these roots are statistically significant as shown by a standard chi-squared test. These roots are the squares of the canonical correlations, to wit, $R_{c_1} = 0.922$ and $R_{c_2} = 0.867$. These are the maximum correlations between two linear functions of linear combinations "orthogonal" to previous linear combinations.

TABLE XIII

Correlations between 14 geochemical variables. Values in italics are significant at the 5% level (only marked in lower triangle)

	Si	Fe	Mg	Ca	Na	K	Ti	P	Mn	V	Mo	Sr	Pb	Zn
Si	1.00	0.66	-0.43	-0.66	-0.46	-0.13	-0.02	-0.65	0.59	0.57	0.34	-0.42	0.26	-0.09
Fe	0.66	1.00	-0.25	-0.54	-0.30	0.17	0.24	-0.29	0.63	0.58	0.59	-0.52	0.27	0.14
Mg	-0.43	-0.25	1.00	0.65	0.49	0.15	-0.27	0.56	-0.32	-0.28	-0.17	0.35	-0.13	-0.16
Ca	-0.66	-0.54	0.65	1.00	0.77	0.26	0.02	0.61	-0.37	-0.25	-0.28	0.75	-0.14	0.05
Na	-0.46	-0.30	0.49	0.77	1.00	0.45	0.09	0.55	-0.19	-0.15	-0.10	0.62	-0.05	0.07
K	-0.13	0.17	0.15	0.26	0.45	1.00	-0.03	0.37	0.16	0.00	0.09	0.17	0.22	-0.04
Ti	-0.02	0.24	-0.27	0.02	0.09	-0.03	1.00	0.07	0.16	0.41	0.42	0.03	0.47	0.42
P	-0.65	-0.29	0.56	0.61	0.55	0.37	0.07	1.00	-0.29	-0.37	0.02	0.48	0.15	0.20
Mn	0.59	0.63	-0.32	-0.37	-0.19	0.16	0.16	-0.29	1.00	0.58	0.42	-0.27	0.35	0.13
V	0.57	0.58	-0.28	-0.25	-0.15	0.00	0.41	-0.37	0.58	1.00	0.60	-0.21	0.58	0.34
Mo	0.34	0.59	-0.17	-0.28	-0.10	0.09	0.42	0.02	0.42	0.60	1.00	-0.39	0.57	0.56
Sr	-0.42	-0.52	0.35	0.75	0.62	0.17	0.03	0.48	-0.27	-0.21	-0.39	1.00	-0.11	-0.23
Pb	0.26	0.27	-0.13	-0.14	-0.05	0.22	0.47	0.15	0.35	0.58	0.57	-0.11	1.00	0.36
Zn	-0.09	0.14	-0.16	0.05	0.07	-0.04	0.42	0.20	0.13	0.34	0.56	-0.23	0.36	1.00

TABLE XIV

Inter-set correlations for geochemical variables and frequencies of organisms. Values in italics are significant at the 5% level

	Ostracods	Afrobolivina	Valvulineria	Gabonella
Si	0.0695	-0.0038	0.3317	0.2732
Fe	0.1358	-0.0088	0.3844	0.3909
Mg	0.2100	0.1716	-0.0834	-0.1195
Ca	-0.1499	0.2206	-0.3944	-0.4818
Na	-0.2250	0.2465	-0.2430	-0.2414
K	-0.1471	0.0123	-0.0159	0.1510
Ti	-0.2682	0.0907	-0.1760	-0.0080
P	-0.2041	0.0781	-0.2015	-0.0074
Mn	-0.0833	0.1132	0.4338	0.2635
V	-0.3266	0.2511	-0.0458	-0.0110
Mo	-0.3122	0.0774	0.2691	0.4166
Sr	-0.2504	0.2039	-0.3276	-0.4726
Pb	-0.3535	0.1019	-0.2256	0.2169
Zn	-0.4221	0.2242	0.0819	0.1605

The structure coefficients for two canonical factors for all 18 variables are given in Table XV. These coefficients are not commonly used and you will not find them in most texts on multivariate statistical analysis. The main steps involved are as follows (extracted from Cooley and Lohnes, 1971). Having found the roots of eqn (25), the vector d is obtained from the equation

$$(R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} - r_{c_i}^2 I)d_i = 0 \qquad (26)$$

with the constraint that $d_i' R_{22} d_i = 1$. The d_i are the weights for the jth canonical factor of z_2. The corresponding weights for the jth canonical factor of z_i are obtained from the relationship

$$c_i = \frac{(R_{11}^{-1}R_{12}d_i)}{\sqrt{\lambda_i}}$$

Up to now, these steps are the normal ones of canonical correlation. The expansion of the method into a "redundancy analysis" is done by finding the variance extracted by the canonical variables, $s_i' s_i / p_1$, where p_1 denotes the number of variables in vector z_1 (here, this comprises the 14 chemical elements) and $s_2' s_2 / p_2$, where p_2 denotes the number of variables in vector z_2 (in the present example, this is four). We have also that $s_1 = R_{11}c$ and $s_2 = R_{22}d$. The redundancy of set 1 in the presence of set 2 (set 1 contains the chemovariables, set 2 contains the frequencies of the

TABLE XV

Structure coefficients for two canonical factors of the geochemical and species-frequency data

	Geochemical data			Species-frequency data	
Variable	Factor 1	Factor 2	Variable	Factor 1	Factor 2
Si	0.305	−0.215	Ostracods	−0.389	−0.664
Fe	0.360	−0.231	Afrobolivina	0.122	0.141
Mg	−0.257	−0.083	Valvulineria	0.666	−0.740
Ca	−0.381	0.244	Gabonella	0.622	−0.196
Na	−0.080	0.308			
K	0.205	0.228			
Ti	0.119	0.390			
P	0.054	0.374			
Mn	0.489	−0.205			
V	0.257	0.325			
Mo	0.670	0.231			
Sr	−0.265	0.236			
Pb	0.306	0.671			
Zn	0.507	0.363			
Factor redundancy	0.1016	0.0775	Factor redundancy	0.2116	0.197
Total redundancy	0.218		Total redundancy	0.565	

organisms) is defined as

$$R_{d_x} = s_1' s_1 R_{c_x}^2 / p_1$$

where the subscript x labels the canonical factor x. The reverse relationship is expressed by the formula

$$R_{d_y} = s_2' s_2 R_{c_x}^2 / p_2$$

The first canonical factors are $x_1 = c_1' z_1$ and $y_1 = d_1' z_2$. Likewise, the second canonical factors are $x_2 = c_2' z_1$ and $y_2 = d_2' z_2$.

The first canonical factor (Table XV) for the chemical variables comprises significant loadings for most of them. Only Na, Ti and P are so low as to suggest a non-important correlation. The first canonical factor for the species frequencies contains significant loadings for all frequencies except that of *Afrobolivina afra*. This is expected as this species is poorly correlated with the chemovariables. The chemical canonical variate is positively correlated with Si, Fe, K, Mn, V, Mo, Pb and Zn, and negatively correlated with Mg, Ca and Sr. This canonical variate may represent a dipolar relationship between sediment richer in carbonates and clastic sedimentary components. The right-hand canonical variate is positively correlated with the frequencies of *Valvulineria* sp. nov., and *Gabonella elongata*, and negatively with ostracods.

Ecolog from the canonical correlation

The plot of the scores obtained by substituting the partitioned mean vectors into the first pair of linear relationships can be used to produce a palaeo-ecological log in which the fluctuations in the frequencies of the organisms are poised against variations in the chemical components of the host sediment. The log for the 26 levels analysed here is shown in Fig. 31.
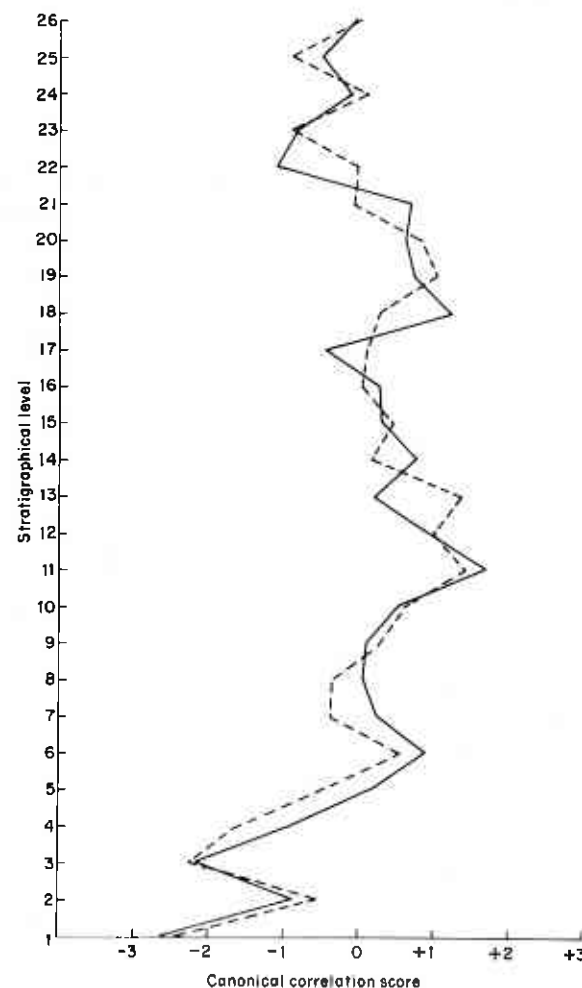
Fig. 31 Ecolog produced from "scores" for the first canonical correlation (geochemical variables balanced against fossil frequencies) plotted against the ordering of 26 samples in the borehole Gbekebo I.

As to be expected from the rather high corresponding canonical correlation, both curves follow the same general trends, although there are numerous deviations in the middle and upper thirds of the plots. These deviations are small but doubtlessly mark periods during which the chemical influences on the organisms were over-ridden by other factors. The lower third of the figure might be an indication of a phase in development during which the chemical components of the environment played a dominant roll. Such a situation could be expected to arise during a period of pronouncedly chemical sedimentation, such as during the early phases in the formation of a marl.

Ecolog by principal coordinates

Using Pythagorean distances between individuals (Gower, 1966), all 18 variables were collected into a single principal coordinates analysis. The plot of the first set of coordinates against location in the borehole, illustrated in Fig. 32, shows the existence of trending in the points in that the youngest samples are displaced in relation to the oldest ones. This could indicate that there was a largely unidirectional ecological trend in the palaeoenvironment over the time covered by the samples. A more subtle and fascinating property of the ecolog is the fact that the youngest samples appear to be in a state of ecological equilibrium (levels 1–9 inclusive), while the older samples seem to reflect conditions that developed in an ecologically perturbed system. This provokes the thought that the ecological system could have been in the process of becoming stabilized in some manner or other, not necessarily optimal for the proliferation of benthic micro-organisms. In fact, the youngest samples are particularly characterized by the predominance of *Afrobolivina afra* over the other three categories.

The principal coordinates log offers a useful way for representing the relationships between organic components of the environment and chemical factors in a single curve. For its meaningful application, the sedimentary sequence should be relatively homogeneous and there should not be shifts in salinity such as occur in a deltaic sequence. A drawback is, of course, the time and expense occasioned by the chemical analyses. Nonetheless, there will be many situations in which the palaeo-ecological aspects of the problem will be such as to justify the preparation of an ecolog.

In the present example, I have used chemical variables for establishing the ecological relationships. Any other kind of factor of the environment can, of course, be used, providing that it can be measured.
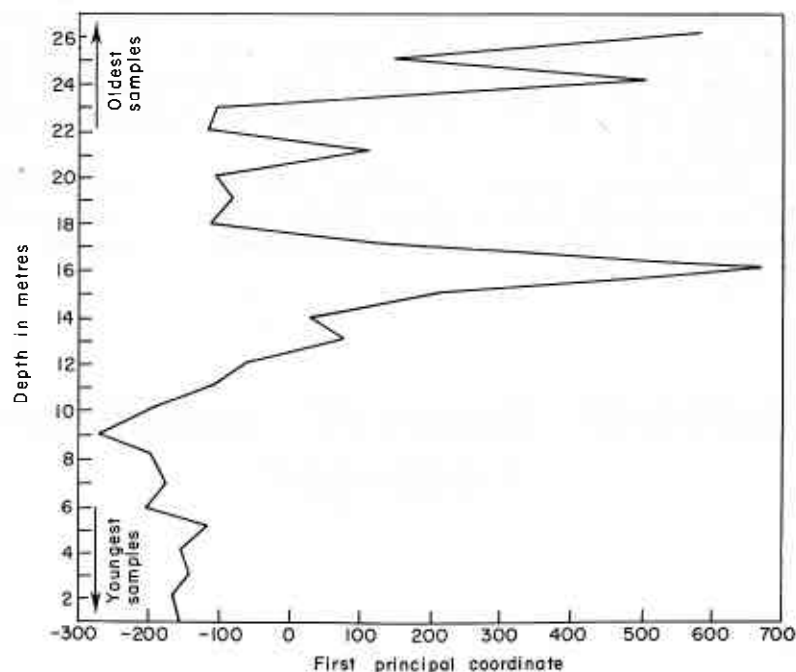
Fig. 32 Principal coordinates ecolog for all variables of the 26 samples from the Gbekebo borehole. Here the first set of coordinates is plotted against the location of the sample in the borehole Gbekebo I. There is a general trend from the youngest to the oldest samples, apart from the strongly expressed fluctuation at 16 m, which may be a reflection of sharp change in environmental conditions.

As noted earlier (p. 63), the methods of principal coordinates and principal components are equivalent procedures if applied to the same measure of association (for example, the correlation coefficient), formed from continuous variables. This is a consequence of the singular value decomposition theorem (the Eckhart–Young theorem, see Jöreskog et al., Chapter 2, section 12). In the principal components treatment, the points are plotted in the space of the variables, a so-called R-technique (R being the correlation matrix, from which the R-concept was derived). The principal coordinates method gives a so-called Q-mode solution (Q being the next letter in the alphabet after R) with the plot in the space of the observations.

If there are more observations than variables, the R-technique is computationally more efficient and, conversely, the Q-method is required if there are more variables than observations.

Note that this duality only applies to continuous variables. For a mixture of variables (e.g. continuous and discontinuous, quantitative and qualitative, dichotomous variables alone or in a combination), principal coordinates, with a suitably constructed association measure, provides a correct approach.

In the foregoing example, I sacrificed optimality in computational speed for practicality in the graphical display, as the computer program used has a very full range of plots for doing principal coordinates.

7

Correlating Between Borehole Sequences

Introduction

The final stage in a borehole analysis is usually an attempt at using the sets of logs established for two or more boreholes within an area for correlating between boreholes. Not infrequently, it will be found possible to make reasonably satisfactory comparisons by the visual comparison of the patterns of fluctuations alone, and if this can be done successfully, then there is no point in using computer techniques.

As I noted in my text on quantitative palaeoecology (Reyment, 1971), the first step in *any* quantitative investigation should be a graphical analysis of the unprocessed observations. This will often show up structures present in the data and thus aid in the planning of the quantitative work.

Gordon and Reyment (1979) discussed the question of comparing logs from adjacent boreholes, both singly, as well as taken simultaneously.

It has long been a major interest of petroleum companies, and satellite organizations, to develop quantitative methods for correlating between boreholes, using physical logs.

A logical and easily accessible starting point is the statistical method of cross-correlation. Rudman and Lankston (1973) have employed this technique, with good results, for data with a constant time interval between successive measurements. For most geological data, however, a more general approach than that of cross-correlation is required, as the thickness of sediment involved in any particular situation cannot, invariably, be translated into the time dimension. Rudman and Lankston (1973) have given quite some thought to "stretching" one borehole relative to the other and interpolating measurements to compensate for differing thicknesses of sediment in the lateral distribution of formations. This imposes a uniformity not found in nature. The procedure proposed by Gordon (1973) automatically deals with the problem of differing thicknesses of strata in sedimentary sequences.

In this chapter, I shall present an example of the application of the slotting technique to a correlation problem. In addition to the presentation of the technique made on p. 72, some further theoretical questions will be taken up in conjunction with the example.

Constrained slotting of sequences

In geological correlation of boreholes, useful stratigraphical information is sometimes provided by the presence of marker beds of wide geographical extent, biological characteristics, peculiarities in logging curves, etc. It is desirable that such information be incorporated into any automated process designed for producing the most meaningful juxtaposition of the stratigraphies of the strata drilled. Some common examples of reference levels are:

1. The persistent marker bed, such as a coal bed, a bone bed, a bed characterized by some other geographical or lithological peculiarity.
2. The case of an easily recognizable zonal element with a short range in time.
3. There may be no clear correspondence in space, but only proximity involved, so that the two parts in adjacent boreholes are near-neighbours, for example, a sedimentary facies. These examples may be defined in terms of constraints.

Type 1 constraint

Where two objects are likely to appear close together in the joint slotting (near-neighbours), we may refer to the situation as a type 1 constraint. As

an example of this kind of constraint, we can take a portion of a borehole section in which a particular faunal assemblage is found, in relation to the location of the same faunal assemblage in an adjacent borehole. The segments of the two boreholes containing this faunal assemblage cannot be expected to be exactly equivalent, but they will be near neighbours.

Type 2 constraint

The type 1 constraint cannot guarantee that two objects of interest will be neighbours in the joint slotting, i.e., that no other object from either sequence occurs between them in the joint slotting. A constraint requiring two specified objects to be neighbours in the joint slotting is called a type 2 constraint. Examples of this are the persistent marker bed (a coal bed, a limestone bed, etc.) or some analogous feature from borehole logging techniques.

Type 3 constraint

A type 2 constraint implies detailed geological knowledge of the nature of the sedimentary sequences. Often, the information available is much less precise. For example, one might have reason to believe that in the joint slotting, there should be some overlap between the sections containing a suite of objects in the one borehole with an analogous suite in the other borehole. Here, one is not insisting on complete correspondence between the two sections, a situation which could be handled by a pair of type 2 constraints linking the tops and bottoms of the two sections. One is, instead, stating that the precise form of the overlap is unknown, but that it is believed that the two sections are not completely separate. This may be termed a "type 3 constraint". Geological examples of this kind of constraint abound. For example, overlapping sedimentary formations and overlapping physical log patterns, and sedimentary facies.

A set of constraints of mixed type can be imposed on the slotting of two sequences, the only requirement being that they be mutually consistent. Owing to the additive form of the statistic $\sigma(S_1, S_2)$ (p. 73), an optimal slotting subject to these constraints can be found by splitting the slotting into parts, and piecing together the optimal slotting of pairs of sub-sequences.

The discordance of the optimal constrained slotting will in general be greater than the discordance of the unconstrained slotting. The amount by which they differ gives an indication of how readily the constraints may be satisfied.

The constrained slotting method is available in a FORTRAN IV program SLOTSEO. Owing to the storage requirements necessary for tracing back the slotting route, the size of the problem which can be analysed by this program is limited. An alternative FORTRAN IV program, BIGSLOT builds up and prints out the calculations sequentially, thus requiring the user himself to trace back the route taken. This greatly extends the size of the problem which can be analysed, but also the work involved.

Example: two boreholes in the Lansing Group of Kansas, USA

The data used here are analysed in detail in a more comprehensive study by Gordon and Reyment (1979). I shall here restrict myself to considering only a part of the results of that study. The data employed consist of the sonic and induction logs for two boreholes drilled by the Skelly Oil Company in western Kansas, namely, No. 1 Bartasovsky well and No. 5 Kisling well, both located in the Cahoj oilfield. Hereinafter, these wells will be referred to as Bartasovsky and Kisling.

The wells span Upper Pennsylvanian rocks included in the Lansing and Kansas City Groups; they are located at about 1.5 km from each other. The sequence lies below the contact between the Douglas Group and the Lansing Group.

There are three marker limestones in the borehole sequences analysed, here denoted, levels A, D and E. As a means of illustrating the slotting technique, the Kansan data allows us to check the success of a "raw slotting" by seeing how well it identified the marker levels. Moreover, the effect of using the marker levels as type 2 constraints can also be tested by these data.

Blocking and filtering

A certain amount of "blocking" may occur in the final slotting, i.e., several objects from the same sequence occur consecutively, with no intervening objects from the other sequence. Blocking occasionally points to interesting differences between the sequences, such as differential compaction, regional dip, etc. On other occasions, blocking occurs in parts of the sequence where there is little variability in the readings.

Providing that this property of the method is kept in mind, coupled with an understanding of how the data arose, there is no objection against splitting such blocks by hand afterwards.

Filtering is often advocated as a means of improving the interpretability of borehole logs. Gordon and Reyment (1979) investigated the effects of filtering and found that the resulting smoothing of the curves tended to promote blocking. Filtering is therefore of limited value in conjunction with slotting, although it may be useful in other connections.

Blocking also occurs in parts of the sequences which differ markedly, but such a situation can generally be distinguished by the higher values taken by ψ.

Standardizing and weighting of variables

I shall mention two different kinds of standardization here. First, if you wish to compare two boreholes on the basis of several different kinds of log recorded in each, it is always good policy to check the effects of standardizing the variables. This is because logs are perforce recorded in different units with different amounts of variability in the recorded scales. One way of standardization is included in formula (20) on page 73. This formulation allows us to give extra weight to variables which experience shows to be more useful in comparative studies.

Secondly, there is the possibility that a variable may have a similar shape in its log at two different sites, but that the scale of the curves may differ. The mean and standard deviation of the readings in each of the four logs of the present example are listed in Table XVI. From this Table, you will see that there could be a case for standardization, although this is not definite. In the full-scale investigation of Gordon and Reyment

TABLE XVI

Summary statistics for logs from the two Kansan boreholes

| Log | Bartasovsky No. 1 | | Kisling No. 5 | |
	Mean	Standard deviation	Mean	Standard deviation
Sonic ($\mu s\,ft^{-1}$)	75.9	13.2	73.8	12.8
Induction (mmhos m^{-1})	226.5	112.7	231.2	117.7

(1979), involving a third borehole, we found ample evidence for the desirability of standardization of the second kind. The question of the standardization of borehole logs is a very general one. A case for standardization could, for example, be made for the examples of Chapter 6.

Slotting on individual logs

Sonic log

In all cases, the slottings were made using 100 observations from each borehole. Here, $\psi = 0.43$, a very low value. In the raw slotting, the A marker was not identified exactly (the true correspondences are given in Table XVII). The result is B5—K8—B6—K9K10 (here, and subsequently, B = Bartasovsky and K = Kisling). The D-marker was not found exactly, owing to the 60-foot level being caught in a block. The E-marker was identified exactly. Standardization by the second method gave improved results, with the D-marker now being located correctly.

Induction log

Here, $\psi = 0.85$, which is still a low value, but higher than for the sonic log, indicating a slightly decreased efficiency in slotting. The A-marker was found exactly and a "close miss" for the D-marker, to wit, K60—B58B59—K61. The E-level was not located exactly, due to blocking. Standardization according to the second criterion given above improved the D and E identifications.

TABLE XVII
Positions of tops of marker limestones in two Kansan boreholes (in feet below the Lansing–Douglas contact)

Limestone marker	Bartasovsky No. 1 $N = 100$	Kisling No. 5 $N = 100$
A	5	10
D	59	60
E	81	84

According to widely accepted practice in the Mid-Western United States, limestone beds in Kansan Carboniferous stratigraphy are labelled alphabetically; in this Table, the top three markers A, D and E are recorded, while limestones B and C, not being considered marker levels, have been left out.
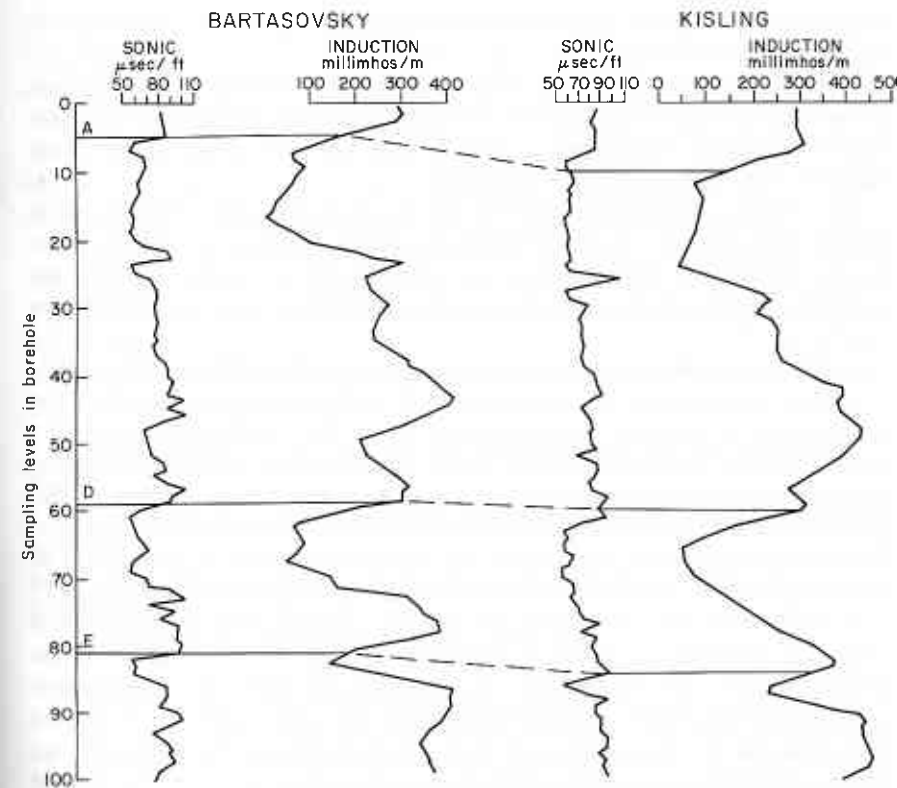
Fig. 33   The sonic and induction logs for the first 100 observational levels in the Kansan boreholes Bartasovsky 1 and Kisling 5. The positions of the three marker limestones (A, D, E) of Pennsylvanian age are shown on the figure.

Both logs together

This gave an overall improved result, with $\psi = 1.57$. The two sets of logs are illustrated in Fig. 33. The A-marker was not exactly identified, and this is a decrease in efficiency in relation to the result achieved by the induction log on its own. Both the D and E levels were found exactly. The results for the raw data must be considered fully acceptable and there would be no need to slot on standardized values.

Constraints

The effects of imposing constraints in order to force the slotting procedure to recognize the three marker levels had the following effects on the

result. First, there was a tendency to increase the value of psi owing to a slight increase in blocking.

Marker A being the most difficult one to identify for the slotting technique, it was found beneficial to make it the only level subject to the type-2 constraint. These slotting results were the most exact as regards an optimal juxtaposition of the sequences.

The slotting technique offers a means of expressing, through ψ, a quantitative estimate of the efficiency of a log in correlating between wells in an exploratory investigation as well as, where marker levels are available, a visual estimate of the success of the log in correlating between boreholes.

I can also mention that the technique has proved useful for other kinds of curve comparisons of geological and palaeontological interest, for example, as in studying in quantitative terms, the evolutionary relationships between ammonite suture lines. One obtains not only a useful element-by-element comparison, but also a quantitative appraisal of the goodness of fit, through the value of psi; this can be extended to apply as a means of comparing the degree of sutural likeness in a genus and in assessing relative likenesses in sets of species belonging to related genera.

A final word here might be of interest. The slotting technique was originally devised in answer to a palynological problem in which it was desired to compare frequencies of pollen species over a fairly extensive area. I have not shown a frequency analysis here. The steps are, however, the same and there is nothing to stop you from comparing borehole logs based on species frequencies. This is a good method and one that should interest micropalaeontologists.

A further use of Gordon's method

Although the main purpose of Gordon's slotting procedure is to compare between boreholes, it may also be applied to comparisons within boreholes. As an example of a situation where such a comparison would be well motivated, we can take the short-normal log and the biolog produced by Campbell's method of stabilized canonical variates (cf. Campbell and Reyment, 1978) for 34 levels in the Nigerian borehole Gbekebo 1.

One of the more interesting results achieved by my project for utilizing the biological information residing in microfossils for making biologs is that, in certain situations which turn out to be optimal for the aims of the method, the electrical resistivity log may be used as an indirect measure

of fluctuations in the palaeo-environment with respect to ecological conditions that existed in the host sediment at the time of life of the organisms (see p. 29). As noted previously, certain basic requirements must be fulfilled if the biolog and concomitant physical logs are to be employed in the way proposed in this book. The most important of these is that the material be free from significant reworking and the sedimentary sequence homogeneous (see p. 150, the sandstone heterogeneity).

In the present example, the data consist of scores on trends for the 34 observational levels in the borehole. That is, I have not used the observations themselves but the record of the directions of the oscillations shown by the logs.

The record of the slotting for the short-normal resistivity curve (R) and the biolog (B) is as follows:

B1B2B3 R1R2R3 B4 R4 B5 R5 B6B7B8 R6 B9 R7 B10B11B12 R8R9

B13B14B15 R10-R13 B16 R14 B17 R15R16R17 B18B19 R18R19 B20

R20 B21 R21 B22B26 R22 B27 R23R24R25 B28 R26 B29 R27R28R29

B30 R30 B31B32B33 R31 B34 R32R33R34

This shows the close agreement in the patterns of the two logs, to which further witnesses the very low value of $\psi = 0.027$. The above slotting was obtained without standardization and without recourse to constraints. From the aspect of the palaeoecological interpretation of the resistivity log in relation to the biolog, the Gbekebo material offers convincing proof of the viability of the postulation that the former is, under specified conditions, a measure of the palaeoenvironment.

The analysis accounted for amounts to a non-parametric application of the slotting technique, a possibility not covered by the existing theory. It works in the present situation probably because the two series are so alike.

8
A Worked Example

This chapter is entirely devoted to the application of the methods outlined in the preceding chapters to a specific problem, some aspects of which have been touched upon in connection with the illustration of details. Ninety-one sampling levels in the Gbekebo borehole are used in which the abundantly occurring foraminifer *Afrobolivina afra* has been measured with respect to the nine characters illustrated in Fig. 25.

Megalospheric tests were selected for the calculations in the interests of homogeneity, as the variation of the microspheric and megalospheric tests shows important and fundamental differences. The resulting biolog agrees well in its record of oscillations with the short-normal resistivity log, except where heterogeneities in the lithology occur. The diameter of the megalospheric proloculus displays, however, a completely different mode, seemingly following a gradual evolutionary trend towards greater size. The microspheric proloculus seems to be invariant over time. The extinction of *Afrobolivina afra* seems, possibly, to follow a pattern compatible with a Thomian catastrophe model.

The biologs yielded by the three alternatives—growth-reduced canonical variates, canonical variates for raw data, and canonical variates for logarithmically transformed data—differ only in minor details, despite the

relatively great variation in the size composition of the samples of the foraminifer.

For the purposes of this illustrative example, I shall use the stabilization method of canonical variates analysis of Campbell (Campbell and Reyment, 1978; Campbell, 1979. Growth-reduced canonical variates gave approximately the same result, with one or two minor exceptions.

The material of *Afrobolivina*

The material analysed consists of samples from 91 levels in the Gbekebo borehole. The species studied is *Afrobolivina afra* which, as already noted, occurs in great abundance in the uppermost Cretaceous beds of coastal Nigeria. The analysis briefly reviewed on p. 109 and presented by Campbell and Reyment (1978) was made on 46 samples comprising mixtures of megalospheric and microspheric tests. This introduces undesirable heterogeneity into the statistical analysis and distortion in the variances and covariances as a result of the disjunctive nature of the distributions of the diameters of the proloculi of the two generations.

The megalospheres were selected for preparing the biolog of my example in order to eliminate the effect noted above. In the present case, this is not as great as one might expect. Some of the results from Campbell and Reyment (1978) are noted on pp. 110 to 111. Comparison of the correlation matrices (Table XIX, this Volume, and Campbell and Reyment, 1978, Table 2) shows the main difference to lie with a marked increase in the correlations of the proloculus with the other variables. The values for the eigenvalues and eigenvectors of $\mathbf{W}^*$, the within-groups correlation matrix, are only slightly affected.

The between-groups sums of squares for all principal components (Table XX) shows 32.6% of the between-groups variation to be associated with principal component VII and 21.8% with principal component I. The variation for principal component VII results from a contrast between variable 7 and most of the other variables. Principal component I is a typical so-called size component. This result does not differ pronouncedly from that obtained for the mixture of megalospheres and microspheres (see Campbell and Reyment, 1978, Table 3).

There is an element of trend in the data and, for this reason, the analysis was made in two ways (cf. Campbell and Reyment, 1978, Fig. 2). There are also differences in the covariances of the younger samples in relation to the older. For these reasons, a first analysis comprising two groups, to wit, 46 older samples and 45 younger ones, was made. The second analysis is for all 91 samples simultaneously.

TABLE XVIII

Means for all samples of megalospheric *Afrobolivina afra* for the first 46 sampling levels in Gbekebo

Levels	N	1	2	3	4	5	6	7	8	9
1	19	73.9105	37.2631	26.0105	15.4000	14.1789	9.7789	19.7421	4.6526	12.4842
2	12	75.2999	32.4583	23.3167	14.4083	13.6333	9.1083	19.0417	4.0333	9.8583
3	16	48.2062	24.6687	18.6562	12.9562	11.8937	8.3687	17.2437	3.5812	8.0625
4	24	87.4374	33.5541	26.6166	15.9000	15.2333	9.4458	20.7083	3.9958	10.4333
5	49	65.1143	31.1102	24.5102	15.9285	14.2775	9.1694	20.0245	3.7469	10.2306
6	29	76.6207	35.2517	27.8827	18.6655	17.6448	12.0828	22.8965	4.2069	11.0138
7	14	61.6428	31.3428	24.4714	15.4857	15.3571	10.5571	20.2357	4.1143	10.4286
8	14	77.4643	30.7071	24.4857	15.8357	14.1429	10.6643	20.3214	3.9429	9.9643
9	16	73.5437	31.7188	25.2562	16.0000	15.4187	11.3812	21.6250	4.2625	9.8750
10	15	75.2533	35.3600	29.6200	19.9133	18.8533	11.1000	27.6266	5.9000	11.8000
11	15	81.6800	35.0866	28.6600	20.4133	19.7066	11.6733	26.3000	5.1333	10.2667
12	15	64.1666	32.5733	25.3000	17.2066	16.2000	10.1667	24.4200	4.3333	10.8000
13	16	62.3937	29.0625	23.0812	18.1437	14.7188	9.7188	20.1875	3.6250	8.5625
14	13	70.5231	28.9538	23.8077	16.6154	15.7615	9.6154	22.6923	3.9000	8.4231
15	16	69.7500	27.0375	23.2500	16.6187	15.1062	9.3375	20.5938	3.7875	7.9063
16	13	68.6077	27.7154	23.5077	16.5461	15.8231	9.3308	21.5385	3.8846	9.5538
17	13	81.0153	34.7230	27.9692	20.3923	17.0154	11.8615	23.6538	4.5231	10.2692
18	25	113.7600	39.9640	33.6280	22.8040	21.2720	14.0680	30.8520	4.3280	11.3400
19	12	73.0083	34.9833	28.7500	20.7083	19.9083	13.0417	25.0833	4.0417	10.5417
20	26	72.8038	31.5692	26.0115	20.2769	18.7692	12.3769	22.6154	3.7500	9.5385
21	13	89.1154	39.7000	31.5692	22.3461	20.9615	14.1538	29.5769	3.7923	11.9615
22	37	73.7784	32.0675	27.4513	20.0513	18.2432	11.7568	24.7432	3.8811	10.2432
23	15	85.53..	33.7000	28.3077	20.2615	18.5385	10.6692	25.8846	4.1846	10.2308
24	18	70.3278	32.7000	26.4222	20.4667	18.5166	11.3056	23.3889	3.8944	9.3056
25	30	68.1933	34.1266	28.0433	22.0067	19.3833	12.3900	23.3167	3.8500	9.7167
26	10	73.4500	34.9200	27.9600	20.6000	19.7000	11.0000	25.4000	3.4900	10.2800
27	12	71.0500	31.1083	26.1583	19.2333	18.4583	10.6667	25.1667	3.4167	9.4167
28	13	81.0615	32.1231	28.1692	21.2077	19.5615	11.0769	25.2308	3.4692	9.3462
29	34	72.0941	30.7059	26.4970	20.7853	17.9676	11.0559	24.6470	3.5382	9.3382
30	29	73.9931	30.6724	26.9965	20.8586	19.0310	10.5345	23.9483	3.6448	9.4483
31	27	60.5629	28.5074	24.2074	18.6852	17.3037	11.0000	22.1963	3.5370	9.2963
32	27	92.2592	36.7185	30.4000	22.2333	20.8259	13.3370	28.8407	3.7296	11.2037
33	18	64.8500	29.6000	24.7666	17.8278	17.4444	10.2222	22.6389	3.7222	8.7500
34	17	81.0823	33.5000	28.2117	20.8353	20.6941	11.8235	28.1765	4.2353	11.1176
35	12	70.4666	32.4250	25.0417	18.1667	18.0833	10.5417	23.9167	3.8750	10.0833
36	11	65.1364	29.8909	25.4818	19.0909	16.8182	10.9273	23.0000	4.0455	9.3182
37	19	79.0368	32.7000	28.3421	20.5737	19.2895	9.6316	25.3947	3.8421	9.6842
38	12	78.5916	30.5666	26.0250	20.5417	18.9500	10.0417	23.4583	3.3167	8.6250
39	14	50.4285	28.0214	22.5714	17.3071	15.7500	9.9071	19.7357	3.6143	8.2714
40	12	58.8583	28.9417	22.1000	18.3083	16.1333	9.0417	20.5000	3.3750	8.5833
41	10	64.7900	29.5900	24.2000	19.3000	17.6000	9.7500	22.7500	3.7500	9.5300
42	14	67.6714	30.6071	24.9000	20.4928	17.7643	10.0143	24.2143	3.6929	10.6786
43	14	111.2571	40.0571	34.1928	25.7143	20.8643	14.7786	31.0714	3.7000	13.1429
44	25	114.6960	40.6040	33.7800	26.0600	23.7520	11.4480	33.2000	4.0400	13.0640
45	19	96.0158	37.2684	30.5000	22.5526	21.1000	12.6842	29.5474	3.9789	11.6842
46	15	74.7600	32.2066	26.6333	20.7200	19.0667	12.5000	26.9066	3.9667	10.0667

TABLE XIX

Correlation matrix computed from matrix **W**, and pooled standard deviations

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
v_1	1.000	0.638	0.685	0.495	0.518	0.245	0.475	0.168	0.230
v_2		1.000	0.830	0.633	0.661	0.509	0.540	0.264	0.341
v_3			1.000	0.639	0.622	0.453	0.543	0.254	0.274
v_4				1.000	0.545	0.384	0.435	0.217	0.230
v_5					1.000	0.366	0.422	0.200	0.278
v_6						1.000	0.318	0.134	0.153
v_7							1.000	0.468	0.478
v_8								1.000	0.332
v_9									1.000
Standard deviations	17.870	4.849	4.354	3.441	3.050	1.697	3.388	0.621	2.074

(Some minor differences in tabulated values are due to the computational precision level (single- or double-) used.)

Variation in the diameter of the proloculus

As can be seen from Fig. 34, the mode of the diameter of the megalospheric proloculus shows a tendency to increase with time. There are here 10 arbitrary grouping of data for all samples. It is significant that the modes of the microspheres, displayed on the same diagram, remain invariant. The topmost histogram in Fig. 34 is for the related species *A. africana*. The modes for the microspheres of this species remains unchanged in relation to *A. afra*, whereas the mode for the diameter of the megalospheric proloculi is much less.

The shift towards greater size of the megalospheric proloculus seems to be a gradual change (cf. Chapter 2, p. 19), of the kind usually referred to as phyletic gradualism.

Steps in the analysis of 46 borehole levels

The means for the 46 samples are listed in Table XVIII. The variables have the same meaning as in Fig. 25 on p. 105. Table XIX contains the pooled correlation matrix for the samples and the pooled standard deviations. The eigenvalues and eigenvectors of the within-groups correlation matrix **W*** for all nine variables are listed in Table XX, together with the diagonal elements of the between-groups matrix in the within-groups principal component space (cf. p. 109).
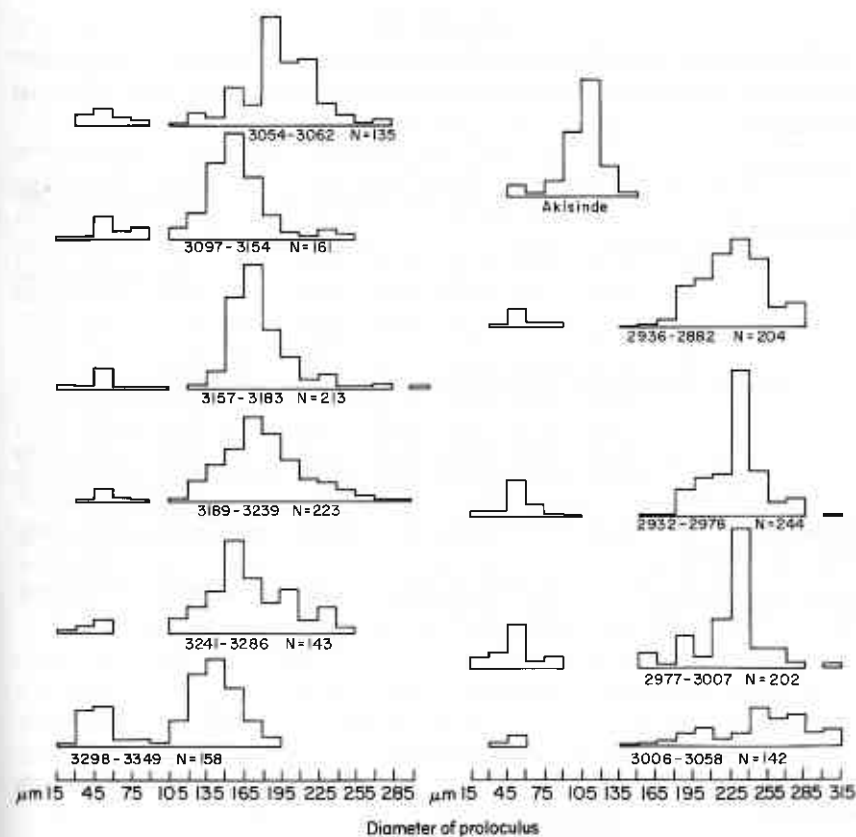
Fig. 34 Histograms for the diameter of microspheric and megalospheric proloculi for ten arbitrary groupings of 91 samples of *Afrobolivina afra* in the Gbekebo borehole and one sample of *A. africana* from the Akisinde borehole.

The standardized canonical vectors for nine variables, including shrunken estimates, are listed in Table XXI. In the present case, only the ninth principal component has been shrunken, as it is small and also related to a small diagonal value of the above-mentioned between-groups matrix (Table XX). These results do not differ markedly from those obtained for the mixture of microspheres and megalospheres (Campbell and Reyment, 1978).

The principal components of **W*** (Table XIX) show an interesting clarification in relation to the mixed analysis in that the diameter of the proloculus of megalospheres occurs in the third component in a negative relationship with length. This can be verified qualitatively as short tests often have unusually large proloculi. This tendency can be seen in the mixed data, but it is clouded by other relationships.

TABLE XX

Eigenvalues and eigenvectors for the within-groups correlation matrix $\mathbf{W}^*$ for all nine variables: between-groups sums of squares for each principal component

Eigenvalues	(e_i)	1	2	3	4	5	6	7	8	9
		4.501	1.237	0.783	0.670	0.508	0.457	0.387	0.300	0.159
Eigenvectors										
	u_1	0.35	0.42	0.42	0.36	0.36	0.26	0.34	0.20	0.22
	u_2	0.16	0.14	0.17	0.17	0.20	0.19	-0.37	-0.61	-0.56
	u_3	0.47	-0.01	0.10	0.01	0.09	-0.86	0.02	-0.12	0.02
	u_4	-0.01	-0.05	0.04	0.11	0.06	-0.10	0.00	0.67	-0.72
	u_5	-0.49	-0.03	-0.15	0.54	0.49	-0.24	-0.32	0.06	0.21
	u_6	0.01	-0.08	0.05	0.71	-0.69	-0.04	0.09	-0.08	-0.01
	u_7	-0.21	-0.24	-0.22	0.09	0.25	-0.04	0.78	-0.32	-0.25
	u_8	0.58	-0.48	-0.49	0.20	0.18	0.28	-0.13	0.10	0.09
	u_9	-0.07	-0.71	0.69	-0.26	0.09	0.06	-0.05	0.00	0.08
		1	2	3	4	5	6	7	8	9
$\mathrm{diag}\{\mathbf{G}_{(0,\ldots,0)}\}$		0.87	0.27	0.35	0.27	0.32	0.09	1.30	0.25	0.26

$\mathrm{tr}\{\mathbf{G}_{(0,\ldots,0)}\} = 3.985$

The pertinent formulae are (11) and (13) on p. 47.

TABLE XXI

Standardized canonical vectors for nine variables including shrunken estimates

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	canonical roots f
a_1^u	0.47	-0.18	0.10	0.12	-0.06	-0.09	-0.77	-0.27	-0.24	—
a_2^u	0.70	0.28	0.08	0.13	0.48	0.08	0.29	-0.08	0.27	
c_1^u	0.00	-0.85	0.01	0.42	0.49	0.29	0.87	-0.49	-0.13	2.021
c_2^u	0.60	0.66	-0.26	-0.39	-0.35	0.32	0.05	0.26	0.26	0.790
$c_{1(0,\ldots,\infty)}^{GI}$	0.07	-0.42	-0.41	0.43	0.44	0.27	0.93	-0.51	-0.17	1.914
$c_{2(0,\ldots,\infty)}^{GI}$	0.56	0.22	0.23	-0.47	-0.33	0.39	-0.08	0.37	0.32	0.737

The relevant formulae are given on p. 110.

The biolog for the first 36 borehole levels

The plots for the stabilized canonical variate means for the megalospheres alone do not differ greatly from those obtained for the mixed sample of

microspheres and megalospheres, illustrated in Fig. 27. There are, however, more agreements in oscillatory directions between megalospheres and the resistivity log than for comparisons between that log and the biolog formed from mixed data. In fact, 34 of 35 matches for the former pair agree in direction, whereas the figure for the latter pairing is 29 agreements out of 35. The removal of the microspheric tests from the analysis has improved the biolog somewhat. (N.B. Observations on the electrical properties of the sediment were only available for 36 of the 46 sampling levels.)

Steps in the analysis of upper 45 borehole levels

Table XXII contains the means for the top 45 levels from the borehole at Gbekebo. The correlation coefficients and standard deviations for $\mathbf{W}$ are given in Table XXIII and the eigenvalues and eigenvectors of that matrix in Table XXIV. Table XXV contains the results of the canonical variate analysis for the first two canonical vectors as well as shrunken estimates.

There are slight differences in the eigenvalues and eigenvectors of $\mathbf{W}^*$, compared with the older set of samples, the most important of these occurring in u_6 (Table XXIV) where variable 7 dominates, whereas it is variable 4 that is highly loaded in Table XX. Also, in u_7, variables 4 and 5 are strongly, and oppositely, loaded, while variable 7 dominates u_7 of Table XX.

The above-noted differences in the structure of some eigenvectors strongly influence the elements of a_1^u and a_2^u (Table XXV) in relation to the results given in Table XXI for the lower 46 levels.

The shrunken estimators of Table 25 are little altered from the original vector elements, such changes as occur being confined to the second canonical vector.

Biolog for the upper 45 sampling borehole levels

The plot of the biolog for the canonical variate means, alongside the short normal resistivity log, is shown in Fig. 35. On the whole, there is less close agreement here than for Fig. 27, which covers the earlier set of observations. The main reason for this is that there is a sandy intercalation between 896 and 900 m. Here, the biolog is the inverse of the resistivity log. However, apart from the sandy level, which doubtlessly

TABLE XXII

Means for all samples of megalospheric *Afrobolivina afra* for the upper set of 45 samples from Gbekebo

Levels	N	1	2	3	4	5	6	7	8	9
1	15	90.1533	39.9800	32.4400	20.9333	19.6000	16.5333	30.0667	3.8333	11.4000
2	27	83.6592	37.4148	29.4148	20.4074	19.7407	16.5555	27.9926	3.7593	10.1296
3	59	76.2389	36.7017	28.8118	20.0729	18.8136	17.4407	27.6017	3.6102	9.6780
4	30	54.6000	28.8200	23.7167	17.7500	16.7333	12.6000	21.0300	3.4200	8.3500
5	24	80.3167	37.0417	29.6667	19.4125	18.3083	14.9583	29.0833	3.3875	9.3333
6	12	70.5417	34.1666	27.5833	18.5417	18.5833	13.5000	25.1000	3.0250	9.4167
7	21	77.8905	36.6524	28.1762	20.0000	20.0000	13.4286	27.0476	3.1571	9.4762
8	10	87.7500	39.3500	30.8100	22.0600	19.8000	13.3000	30.2000	3.4000	9.2000
9	16	64.2000	38.0667	29.2267	19.8667	19.3333	15.2000	29.0000	3.0667	10.3333
10	12	73.9333	36.0167	29.0167	20.5417	20.1667	13.8333	28.4167	3.3750	9.5833
11	16	99.3187	42.2875	34.3875	23.6875	22.1875	15.6563	31.0625	3.8125	11.1875
12	11	82.3000	42.4818	32.0000	22.0000	21.0000	15.0909	31.1818	3.5000	11.4545
13	9	79.4222	36.1555	29.8889	21.2667	20.5111	14.5556	27.8889	3.0556	8.8889
14	12	62.5333	33.4583	27.1667	20.6167	19.4167	14.3333	26.7500	3.4167	8.5417
15	17	99.1615	41.6307	34.9923	25.5461	24.9231	14.9769	31.6538	3.6000	11.6154
16	15	78.3833	38.3833	31.2000	22.8333	20.9583	14.5833	31.6667	3.4167	10.5833
17	15	78.6000	42.1467	34.1866	23.9267	22.0667	15.5333	33.0000	3.4867	9.8000
18	11	84.0454	39.0727	32.4091	22.6364	21.8182	14.5455	31.0000	3.3636	9.0909
19	12	69.7750	33.0583	26.7083	18.0000	17.6250	14.2917	28.5833	3.0833	8.9583
20	14	86.8714	42.4857	33.1857	22.9286	23.1429	15.1429	32.7143	3.4643	11.1429
21	12	73.3333	38.2833	29.6000	22.0000	20.4167	14.7500	31.5000	3.3167	9.4167
22	14	65.6928	37.2214	30.2786	21.0714	19.2500	14.5000	30.2857	3.4357	10.2857
23	14	84.8571	46.1429	37.9286	24.0071	23.4643	16.1429	35.0714	3.4286	10.9286
24	9	106.6111	46.1777	36.2222	26.5555	24.0000	16.3333	37.7778	3.6111	10.5556
25	7	92.3571	34.8571	28.5714	19.6143	19.2857	13.1429	24.5714	2.9143	8.7143
26	12	85.5417	36.3750	30.2500	19.4167	18.6667	14.5417	33.9167	2.7833	9.3333
27	16	85.0437	42.4625	33.0437	23.6875	22.9375	14.5938	33.0625	3.3750	10.4375
28	26	89.1884	44.0192	35.7731	25.7808	23.9231	14.1731	33.6923	3.2808	10.3846
29	29	89.0827	42.4724	35.2483	24.4793	23.2517	14.3793	31.6552	3.1310	9.2759
30	30	90.6200	42.3200	33.1200	23.1566	21.5867	14.2167	32.9667	3.1967	11.4500
31	16	79.4688	39.6563	31.0250	23.5938	22.2188	14.8438	30.1875	3.2062	11.4125
32	24	107.2833	46.4042	37.4541	26.8125	25.3750	15.6875	33.7292	3.3542	10.6042
33	15	75.2200	37.0466	30.1866	23.5667	21.7133	13.2667	30.6667	3.1600	10.5333
34	12	97.5000	42.1583	35.3667	26.4167	24.6667	14.2500	33.7917	3.0667	12.3333
35	16	79.2878	39.4848	31.4272	22.4303	20.7303	15.2303	32.9576	3.3939	10.9212
36	16	86.6000	37.2187	29.9000	20.5562	19.2500	12.5625	28.6563	2.9875	9.7500
37	16	88.9312	37.8375	30.7875	21.1250	19.8062	12.9687	28.5625	3.0187	10.1563
38	15	84.9466	37.2066	28.6266	20.8333	19.7600	12.4533	28.0933	2.9267	9.9533
39	28	85.7857	36.0714	29.5893	20.5857	19.6821	12.6607	28.9714	3.0000	10.0000
40	25	93.9400	38.0320	32.1160	21.1400	19.6680	13.2840	29.0800	3.1000	9.1200
41	15	102.9133	37.3866	31.0133	21.6200	19.6000	14.7333	30.2667	3.1133	10.9000
42	41	97.4317	36.1488	29.2634	20.1195	19.7536	12.8512	28.1707	2.9829	9.2683
43	25	88.5160	35.5840	29.4480	20.3680	19.5360	13.1760	28.4800	3.3120	10.3400
44	68	106.0998	41.5059	34.5206	23.2088	21.4823	12.0088	32.5000	3.3426	10.4912
45	23	120.7739	45.4000	36.7652	24.6130	22.0956	13.1522	34.1522	3.6043	11.3435

TABLE XXIII

Correlation matrix computed from matrix **W**, and pooled standard deviations

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
v_1	1.000	0.705	0.709	0.493	0.513	0.179	0.567	0.236	0.187
v_2		1.000	0.785	0.555	0.573	0.289	0.668	0.306	0.382
v_3			1.000	0.575	0.583	0.271	0.596	0.277	0.210
v_4				1.000	0.628	0.236	0.424	0.228	0.213
v_5					1.000	0.222	0.431	0.171	0.190
v_6						1.000	0.174	0.083	0.069
v_7							1.000	0.405	0.389
v_8								1.000	0.301
v_9									1.000
Standard deviations	18.350	4.466	3.996	2.990	2.548	1.657	3.255	0.405	1.869

TABLE XXIV

Eigenvalues and eigenvectors for the within-groups correlation matrix **W***
for all nine variables; between-groups sums of squares for each principal component

Eigenvalues	(e_i)	1	2	3	4	5	6	7	8	9
		4.342	1.153	0.891	0.717	0.666	0.385	0.364	0.292	0.190
Eigenvectors										
	u_1	0.38	0.43	0.41	0.35	0.35	0.17	0.37	0.21	0.20
	u_2	0.13	0.01	0.16	0.21	0.27	0.29	-0.25	-0.57	-0.61
	u_3	0.23	0.03	0.10	0.06	0.12	-0.93	0.05	-0.16	-0.17
	u_4	-0.17	0.05	-0.13	0.16	0.21	-0.04	-0.10	-0.63	0.69
	u_5	0.34	0.26	0.19	-0.57	-0.45	0.10	0.28	-0.41	0.05
	u_6	-0.37	-0.12	-0.21	-0.18	0.37	0.02	0.76	-0.14	-0.21
	u_7	0.12	0.05	0.04	-0.66	0.63	0.01	-0.34	0.15	0.09
	u_8	-0.70	0.36	0.58	-0.08	-0.07	-0.11	-0.12	0.0	-0.06
	u_9	0.06	-0.77	0.60	-0.04	0.01	0.04	0.10	-0.03	0.17
diag$\{G_{(0,\dots,0)}\}$		0.76	0.16	1.07	0.29	0.26	0.56	0.17	0.41	0.12
tr$\{G_{(0,\dots,0)}\} = 3.798$										

See formulae (11) and (13) on p. 47.

TABLE XXV

Standardized canonical vectors for nine variables including shrunken estimates

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	canonical roots, f
a_1^u	0.57	-0.09	-0.52	-0.30	-0.58	-0.44	0.25	-0.23	0.01	—
a_2^u	-0.27	0.01	-0.75	-0.08	-0.20	0.39	-0.06	0.37	-0.16	—
c_1^u	-0.40	0.24	0.22	0.27	0.20	-0.49	0.71	-0.49	-0.06	1.403
c_2^u	-0.95	0.42	-0.02	0.04	0.15	0.66	0.34	0.23	-0.15	1.165
$c_{1(0,\dots,\infty)}^{GI}$	-0.39	0.22	0.23	0.27	0.19	-0.49	0.71	-0.49	-0.05	1.403
$c_{2(0,\dots,\infty)}^{GI}$	-0.93	0.14	0.02	0.03	0.16	0.68	0.39	0.22	-0.09	1.138

The steps for doing these calculations are given on p. 47.

represents a less favourable environment for *A. afra*, the directions
followed by the oscillations agree well (to the extent of 75% for all 45
levels).

Source of inaccuracy

As demonstrated in the beginning of this chapter (p. 143), there is a
gradual morphological shift in some characters of *A. afra* with time,
manifested most clearly in the diameter of the megalospheric proloculus.
There is also a glide in the covariance structure, as reflected in the
canonical vectors and the eigenvectors of **W***. As is to be expected, this
seems to have an influence on the biolog. For example, the topmost
canonical means of Fig. 35 show a positive trending tendency. However,
the plot of the entire set of observations (Fig. 36) does not show this
trend so distinctly. If only variables showing distinct sensitivity to en-
vironmental effects are included (here, variables 1, 3, 4, 5 and 9), the
trending of the topmost observations is accentuated, becoming paraboli-
cally curvilinear. The suggestion we can make here is that for the analysis
of long sequences of samples in which a "chronocline" in one or more
characters exists, it is advisable to prepare the biolog from subsets of the
data and to join together the curves produced, after standardization of
the canonical variate means. Alternatively, one could delete the variables
showing gradual, directed shifts, basing then the biolog on the variables
that are free from phyletic changes, should such occur.
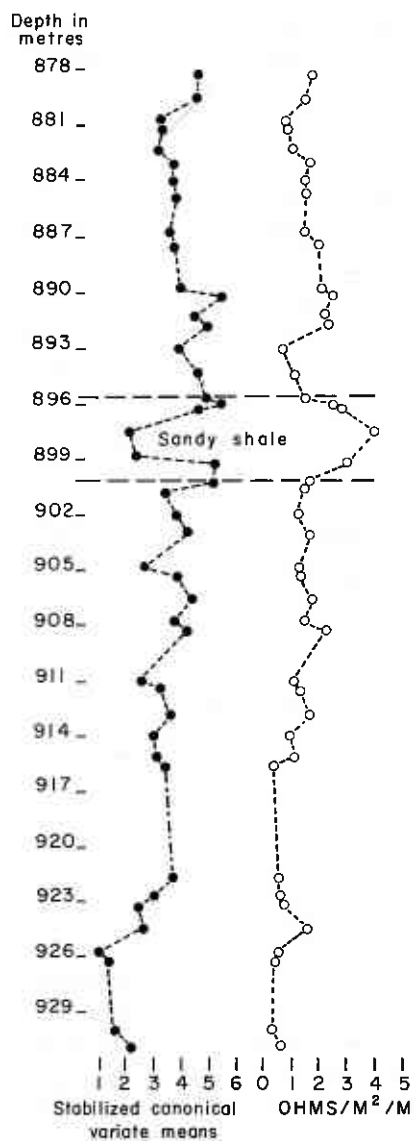
Fig. 35 Biolog formed from the first stabilized canonical variate means for megalospheres for the upper set of 45 samples from the Gbekebo borehole, together with the corresponding values of the short-normal resistivity curve. Sandy shale occurs between 896 and 900 m.
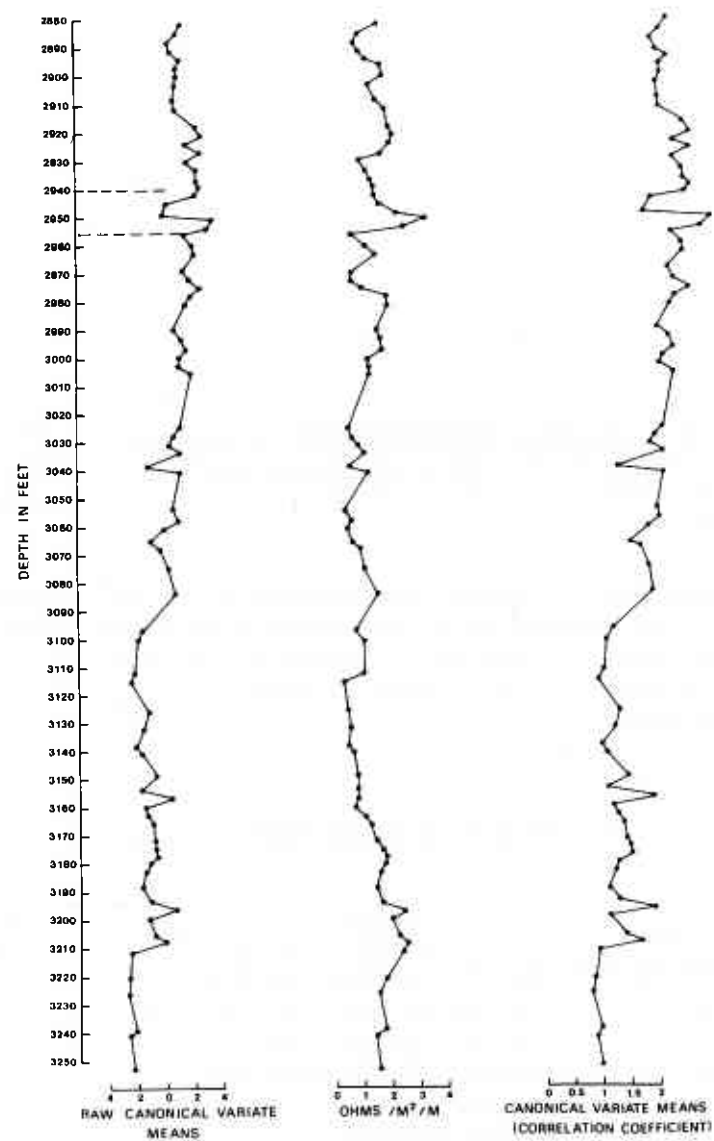
Fig. 36 Canonical variate biologs (logarithmically transformed data and covariance matrices, on the one hand, and canonical variate means from correlation matrices on the other) and the short-normal resistivity log for 81 levels in the Gbekebo borehole. The dotted field denotes a sandy level in the borehole. Feet are used here to denote sampling depth for ease of graphical representation (the original sampling was actually done in feet).
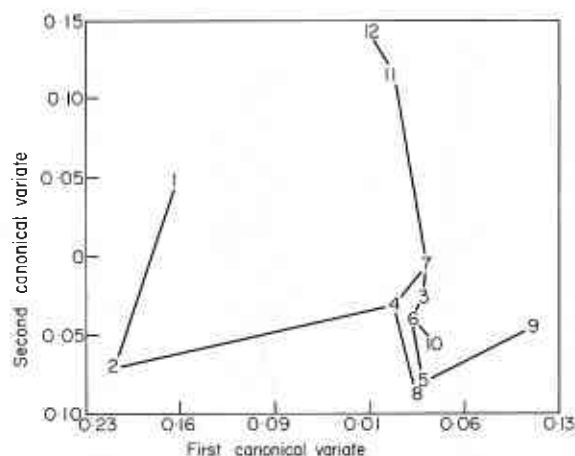
Fig. 37. Plot of the first and second canonical variates for 12 levels in the borehole Gbekebo I. The lines connecting the points are the Prim minimum spanning tree.

It should here be mentioned that the shift in size of the diameter seems to be largely uninfluenced by the changes in size shown by the other variables, which I interpret as being mainly of ecological origin. It seems as though the megalospheric proloculi could have undergone a genuine phyletic change.

All samples analysed simultaneously

Figure 36 shows two biologs (unstabilized, raw data and stabilized canonical variates) for 81 levels in the Gbekebo borehole. The short-normal resistivity log is also shown. A point of interest is that there is an overall trend to the right in the biologs. This is slight admittedly but the shift is unmistakable. The electrical log does not display a sustained trend, although the proportion of coincident directions of oscillation is very high, being 65 out of a total of 80 matches.

A byproduct of a biolog, suggested by its agreement in variational directions with the resistivity and *SP* logs in homogeneous shale, is that in situations where observations on the electrical properties of the sediment are lacking, for some reason or other, the biolog may serve as a first-level approximation to those curves.

9
Summary of Methodology

In this chapter, I shall be concerned with mapping out a general strategy for analysing borehole data. It will be assumed that the data will consist of

(a) pieces of core, sampled at known levels in the borehole,
(b) a suite of physical logs of the usual "Schlumberger" variety,

that is, gamma ray, neutron, electrical logs, sonic, etc. In particular, it is recommended that the short normal resistivity log should figure among these.

Preparation of the samples

There are a wide variety of techniques used in micropalaeontological laboratories for extracting microfossils from sediments. Particular attention should be paid to methods that do not damage the specimens, as some of the forms of biologging that turn out to be useful require accurate observations on discrete characters of the shell, such as spines. Strong heating is seldom a good method to use if you want to make detailed anatomical studies of your fossils.

The preliminary search

The actual scientific work will start with a preliminary search of the washed residues for those long-ranging species which occur in sufficient numbers throughout the boreholes. This is the opposite philosophy of the usual approach adopted in micropalaeontological work in which the search is for short-ranging marker species. For the best results, in conjunction with the physical logs, these will need to be benthic species of foraminifers and ostracods. Planktonic species are difficult to relate to the ecological conditions as reflected in the host sediment; thence the stress on endobionts, in the first instance, and epibionts as a less suitable alternative.

This phase in the analysis is probably its most brittle part. Clearly, there is no way of knowing how rich a suite of samples is going to turn out—some levels may prove to be satisfactory, but there may be many large gaps in the sequence. Of course, you are not obliged to use animal microfossils and in many situations it might be a good idea to advance on a broad front by using both animal and plant fossils.

As an example of the kind of result one can expect, I can take the Nigerian material forming an important part of this text. Only one of the multitude of species, the by now familiar *Afrobolivina afra*, has the vertical and horizontal extensions, as well as the numbers, to be really suitable for the preparation of a fully developed biolog. We have also seen how it played an important part in the preparation of an ecolog. At this point, it is also to be recommended that a sketch map of the geology of the area and the locations of the boreholes be prepared and kept at hand as the various techniques are applied to the data. This might seem obvious but it is worth repeating. Nonetheless, I have frequently noted that we micropalaeontologists do not always take the geographical factor into account. The locations of the Nigerian boreholes referred to in this book, to wit, Akisinde, Araromi and Gbekebo, are shown in Fig. 20.

The measurements

Assuming now that a suitable species, or a set of species, has been selected, the next step in the work is to make the measurements. This will, of necessity, involve a certain amount of guess-work in the beginning as to which of the many possible variables are diagnostic. My suggestion is that you take one or two large samples and make a detailed multivariate analysis of them using, for example, "factor analysis" in the

manner expounded in Jöreskog *et al.* (1976). In this manner, it will usually be found possible to identify redundant variables and, as a consequence, the continued measuring can be restricted to the variables likely to contain the most biological information of relevance to the problem at hand. Note that the term factor analysis is here used generically for a class of techniques for analysing single samples. If you intend using ratios of some kind, take particular care (cf. Simpson *et al.*, 1960, pp. 13–19). You are most likely to run into trouble in situations where there are very pronounced allometric differences between variables.

Regarding the actual technique of making the measurements, I prefer to use scanning electron microscope (SEM) pictures (carefully calibrated). This is not to say that measuring under the light microscope is banned in the present connection. However, the SEM pictures form a permanent record of the objects measured and if you are having trouble with what could be an outlier, you can quickly go back to the specimen in doubt and check your observations. Perhaps a look at Table I (p. 11) again will refresh your memory concerning the havoc that an outlier can wreak with your analysis.

In a normal investigation of a borehole, the number of photographs to be stored may become great and it is therefore advisable to work within the framework of a carefully constructed book-keeping system. In Uppsala we have one set of books for the washing of samples, preliminary micropalaeontological determinations and the necessary information concerning pieces of pre-washed core sampled for geochemical analysis. A second set of notes contains the details of the SEM photography, magnification, calibration, and any special observations on specimens requiring detailed structural study, preservation, evidence of reworking, etc. The third volume contains the measurements, ordered according to depth in the borehole. All three books are cross-referenced so that, for example, the geochemical determinations belonging to a particular level, or the photograph underlying a set of measurements can be identified without trouble. Finally, a fourth set of books is kept for the statistical analyses of the measurements.

Processing of the data

Over the years, I have built up a very comprehensive library of statistical routines, particularly programs for carrying out special multivariate statistical computations. A good deal of this material has been published (see, for example, Reyment, 1971; Blackith and Reyment, 1971). Statistics is,

however, a rapidly evolving science and it is necessary to keep a constant watch on the statistical literature in order to keep the program library fully modern and, in consequence thereof, of progressively increasing value for analytical work. I have therefore considered it pointless to publish computer programs as these would have been out of date even before the book had left the press. The journal *Computers in Geology* is a valuable source of geologically orientated programs.

The analyst will have decided how much time and effort he wishes to put into obtaining a biolog. In industrial connections, time may be of such over-riding significance in a particular project that it may have been concluded that a biolog constructed from fluctuations in means, with or without confidence intervals, must suffice as a first measure. As I have demonstrated in Chapter 4, quite a lot can be won from a simple study of this extent, particularly if it be based on ostracods (probably, also, pollen-grain dimensions).

A word on organization and personnel. There have to be at least two persons on the job if you intend getting something done within a reasonable time. The work can be divided so that one keeps to the SEM photography and measuring while the other can do the statistical work and see that geochemical analyses are done, if required. The SEM work requires a particular kind of person, who will not balk at the thought of five to six hours a day on the instrument and who is able to do the maintainance necessary to keep it in fully satisfactory order. We have found the latter requirement very important in our work as no matter how finicky you are about cleanliness, constant running of an SEM is attended by a deterioration in picture quality. It is good practice to clean the column at far more regular intervals than the manual specifies. Remember, fuzzy pictures make fuzzy measurements.

A logical first step for producing a biolog is the first transformed mean of canonical variates for your samples. My program for canonical variates CANRED (updated and revised as of May, 1979) can take up to 60 stratigraphical levels at a time and 30 variables. Normally, I would go in with this program if dealing with ostracods and expect it to give good results. It has, in its original version, had ten years of international testing (dating from Reyment and Ramdén, 1969) and it has been carefully kept up-to-date. More often than not, you will find yourself dealing with benthic foraminifers and then you should consider, seriously, whether size differences (see p. 67) are going to give you spurious fluctuations in means. More often than not, they will. In that case, you will need to use more complicated tools. There are two avenues open now.

In my opinion, it is best to start with an appraisal of how stable the results for the standard canonical variate analysis are. For this, the

procedure devised by Campbell (see Campbell and Reyment, 1978) is the most suitable method. If the estimates yielded by the standard procedure are stable to the extent of only differing slightly from shrunken estimates, then I should think that you can accept these and go ahead to produce your biolog from the set of first canonical means, as described in the foregoing. If your values turn out to be very unstable, Campbell's method has the built-in advantage that you can see which of the variables are causing the trouble.

I also carry out a study of the effects of removing growth vectors from the set of samples. Often, the biolog deriving from this analysis approximates that obtained from the shrunken estimates of canonical variates. If you are using a foraminiferal species for making your biolog it is always advisable to see how much of the variation you are finding is due to mere size differences due to the growth stage attained by individual specimens. This can be carried out by means of the Burnaby–Gower technique, presented in an earlier chapter.

A further point concerning canonical variates should be mentioned. The interpretation of plots can sometimes pose a special problem and, at the worst, lead to misleading results. To illustrate what I mean, I have included a figure showing the plot of the first and second canonical variates for 12 levels in the Gbekebo borehole (Fig. 37). On first sight, it appears that the canonical variate means of the 12,9-variate samples of *Afrobolivina afra* are differentiated along the first canonical variate axis such that the first two samples lie at one extreme, then there is a cluster of the other ten samples with the last two being displaced by the second canonical axis, but not the first. This is generally an accurate representation of the plot; however, when it comes to interpreting relationships within the principal cluster of the plot, lying between -0.01 and 0.04 on axis 1 and -0.07 and 0.00 on axis 2, incorrect assumptions can easily be made. It seems obvious that samples 5 and 8 and 4 and 6 are very close to each other.

The affinities disclosed by drawing in the minimum spanning tree are, however, different in many respects (Prim, 1957; Everitt, 1978; Ross, 1969). A "tree" spanning n points in multivariate space is said to be a minimum spanning tree if the set of straight-line segments joining pairs of points is such that no closed loops occur, each point is connected by at least one line and, there are paths between any pair of points. Plotting a minimum spanning tree (or Prim network) is a very useful way of bringing out distortions produced by an ordinating procedure such as the reduced dimensional representation of multivariate data by canonical variate analysis. In the present example, it will be seen that the structure of the minimum spanning tree in Fig. 37 discloses that samples 5 and 8 are

really not close to each other and that groups 4 and 6 are closer to 8 and 5, respectively, than 5 and 8 are to each other; likewise for the relationships between samples 4, 7, 3 and 6. Samples 9 and 10 seem to lie reasonably close together but this is also an illusion.

The Prim networks are here computed from the generalized distances between the ten samples and the same conclusions as indicated in the foregoing paragraph can also be arrived at from an inspection of the table of D^2. For ten samples, this is a straightforward procedure but for the 92 samples of the entire *Afrobolivina* material, it is quite out of the question to attempt to rationalize the distance relationships by inspection of a table of D^2. It is in such situations that the minimum spanning tree can provide invaluable information about the relationships in a plot, particularly plots of the kind produced in the present connection in which interested is directed towards identifying secular variations in morphological characters.

The rationale behind the use of the minimum spanning tree for unravelling relationships in two-dimensional plots of multivariate data can be understood if we consider our view of the stars on a clear night. To us, they seem to be spread about on a two-dimensional background and it is not possible for the observer to assess true distance relationships. Stars that seem to lie close together often are separated by vast distances.

Comparing the biolog with the physical logs

The next step is to compare your biolog with the physical logs. The most diagnostic one of these to start with is the short-normal resistivity log although the *SP* log and redox log give good results (Pirson, 1977). This may agree closely with the biolog, as is the case for the Nigerian boreholes discussed in this book and, if so, it may be taken to reflect ecological fluctuations in the environment (cf. Pirson, 1977). In the case of the Nigerian boreholes, the reason for this is not difficult to find, as the sediments containing the microfossils were deposited during a major transgressional–regressional event in latest Cretaceous time, which left a clear imprint on the sedimentary sequence. This kind of situation will not always have pertained and it will probably be found in many cases that the physical and biological logs have little in common. This is likely to arise where the sedimentary sequence is heterogeneous in nature.

If, however, several of the logs show the same trends, it may be useful for advanced interpretational work to produce a composite log (p. 106) from the first principal component, or equivalent statistical procedure, of the logs. This composite log may be suitable for presenting the quantitative information on the borehole as a single curve.

Frequency logging

Side by side with the morphometrical biolog, a frequency log, or logs, for diagnostic long-ranging species can be made. Frequency data were used in the example in which the ecolog was introduced in Chapter 6, and Gordon and Birks (1974) discuss several case histories in detail. This is a well-known approach and one that has been greatly popularized by Pleistocene geologists.

Comparing boreholes

The final phase of the analysis is concerned with comparing logs of adjacent boreholes with the end in view of correlating as closely as possible between them. As has been shown in Chapter 7, the slotting technique is a general procedure which can effectively compare between boreholes on which "curves" are available. If marker beds, or the like, occur, information on the location of these in the boreholes can be used to increase the effectivity of the slotting.

Slotting is an attempt at providing a means of automatic borehole correlation. Not infrequently, the direct (visual) comparison of curves can yield a perfectly satisfactory result. Slotting should be kept at hand as a means of helping to resolve knotty situations. It should also be mentioned that the slotting result is not intended to provide the only possible comparison. It is a tool, and should be understood as such; its main usefulness lies with the possibilities it offers of obtaining an initial, hopefully fairly accurate, juxtaposition of two or more boreholes.

When not to make biologs

The biolog is meant to be used as a means of extricating oneself from a troublesome situation such as can be expected to arise with a thick, monotonous and homogeneous sequence of sediments containing a few long-ranging fossils, none of which is a zone fossil. It cannot be recommended for sedimentary sequences marked by frequent and sharp changes in lithology and with clearly demarcated fossil associations. The traditional methods of palaeontology are perfectly adequate for like sequences and it would be ludicrous to go ahead with the work involved in producing a biological log in such cases.

This piece of advice may seem magisterial, but I fear that once a laboratory has gone to the expense of setting up the analytical procedures necessary for producing a biolog, etc., the temptation may be overpowering to treat all data in this fashion.

Adequate slotting records

Assuming that most investigations will include correlations between boreholes, or correlations between logs of the same borehole (often a useful approach), some attention should be given to the presentation of adequate records. The slotting result ought to be accompanied by information concerning the kinds of logs used, whether or not they have been standardized before processing and, if weighting has been introduced, what the proportions between weights are. As an illustration of how the records can be presented, I have used the data for Fig. 24, the biologs for *Cytherella sylvesterbradleyi* from analogous levels in the Paleocene portions of the Araromi and Gbekebo boreholes. The material is presented in Table XXVI. It is not usually necessary to depict the entire slotting result, although this may be advisable if there are no marker levels. If physical logs enter into the study, their properties should be recorded in the same manner. The results for a slotting based on composite logs may be treated in the same way.

TABLE XXVI

Example of adequate slotting records for the material illustrated in Fig. 24.

	Species = *Cytherella sylvesterbradleyi*	
	Araromi (=A)	Gbekebo (=G)
Mean	0.739	1.399
Standard deviation	0.364	0.474
N	9	9

Slotting results
$\psi = 0.265$
The slotting ordering is

G   G   GGGG   G   G   G
        A   A   AAAA   AA   A
Top   Bottom

No constraints were used, but the readings were standardized.

References

Ayala, F. J. (Ed.) (1976). "Molecular Evolution". Sinauer Associates, Sunderland, MA.

Bantock, C. R. and Price, D. J. (1975). Marginal populations of *Cepaea nemoralis* (L.) on the Brendon Hills, England. I, Ecology and ecogenetics. *Evolution* **29**, 267–277.

Barker, D. (1963). Size in relation to salinity in fossil and recent euryhaline ostracods. *J. Marine Biol. Ass. U.K.* **43**, 785–795.

Barnett, V. and Lewis, T. (1978). "Outliers in Statistical Data". Wiley, New York.

Blackith, R. E. and Reyment, R. A. (1971). "Multivariate Morphometrics". Academic Press, London and New York.

Bliss, C. (1957). "Statistics in Biology". Vol. 1, McGraw-Hill, New York.

Bookstein, F. L. (1978). "The Measurement of Biological Shape and Shape Change". Lecture Notes in Biomathematics, 24, Springer-Verlag, Berlin.

Borowsky, R. (1978). The tailspot polymorphism of *Xiphophorus* (Pisces: Poeciliidae), *Evolution* **32**, 886–893.

Bradley, J. V. (1968). "Distribution-free Statistical Tests". Prentice-Hall, N.J.

Brinkmann, R. (1929). Statistisch–biostratigraphische Untersuchungen an mitteljurassischen Ammoniten: über Artbegriff und Stammesentwicklung. *Abh. Gesell. Wiss. Göttingen, N.F.* **13** 1–249.

An early attempt at using the variation in fossils to establish a type of quantitative stratigraphical analysis.

Brussard, P. F. (1975). Geographic variation in North American colonies of *Cepaea nemoralis*. *Evolution* **29**, 402–410.

Burnaby, T. P. (1966). Growth-invariant discriminant functions and generalized distances. *Biometrics* **22**, 96–110.

In my opinion, this is one of the most important papers ever written in biometry. It is fundamental to the concept of the biolog, in particular, and all multivariate taxonomical work in general.

Campbell, N. C. and Reyment, R. A. (1978). Discriminant analysis of a Cretaceous foraminifer using shrunken estimators. *Math. Geol.* **10**, 347–359.

Campbell, N. C. (1979). Canonical Variate Analysis: Some Practical Aspects. Ph.D. Thesis, Imperial College, University of London, unpublished.

Cervantes, S. T., Goodman, M. M., Casas, E. D. and Rawlings, J. O. (1978). Use of genetic effects and genotype by environmental interactions for the classification of Mexican races of maize. *Genetics* **90**, 339–348.

Clarke, B. and Murray, J. (1971). Polymorphism in a Polynesian land snail *Partula suturalis vexillum*. In "Ecological Genetics and Evolution", (Ed. R. Creed). 51–64. Blackwell Scientific Publications, Oxford and Edinburgh.

Cook, L. M. and O'Donald, P. (1971). Shell size and natural selection in *Cepaea nemoralis*. In "Ecological Genetics and Evolution", (Ed. R. Creed). 93. Blackwell Scientific Publications, Oxford and Edinburgh.

Cooley, W. W. and Lohnes, R. R. (1971). "Multivariate Data Analysis", Wiley, New York.

Cox, D. R. and Lewis, P. A. W. (1966). "The Statistical Analysis of Series of Events", Methuen, London.

Cox, D. R. (1970). "Analysis of Binary Data". Monographs on applied probability and statistics. Methuen, London.

Cramér, H. (1946). "Mathematical Methods of Statistics", Princeton University Press, Princeton, N.J.

Creed, R. (Ed.). (1971). "Ecological Genetics and Evolution", Blackwell Scientific Publications, Oxford and Edinburgh.

Davis, J. C. (1973). "Statistics and Data Analysis in Geology", Wiley, New York.

A well-presented introduction to statistical methods applied to geological problems. Special attention is given to comparing borehole sequences. Computer programs.

Delany, M. J. and Healy, M. J. R. (1964). Variation in the long-tailed fieldmouse (*Apodemus sylvaticus* (L.)) in north-west Scotland. II. Simultaneous examination of all characters. *Proc. R. Soc. B* **161**, 200–207.

Dobzhansky, T. (1971). Evolutionary oscillations in *Drosophila pseudoobscura*. In "Ecological Genetics and Evolution", (Ed. R. Creed). Blackwell Scientific Publications, Oxford and Edinburgh.

Dodson, M. M. (1976). Darwin's law of natural selection and Thom's theory of catastrophes. *Math. Biosci.* **28**, 243–274.

Dodson, M. M. and Hallam, A. (1977). Allopatric speciation and the fold catastrophe. *American Naturalist* **111**, 415–433.

Eldredge, N. and Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. "Models in Paleobiology", (Ed. Schopf), 82–115. Freeman Cooper, San Francisco.

Emigh, T. H. (1977). Partition of phenotypic variance under unknown dependent association of genotypes and environments. *Biometrics* **33**, 505–514.

Everitt, B. (1978). "Graphical Techniques for Multivariate Data", Heinemann, London.

A useful presentation of graphical accessories in multivariate analysis. Can be profitably used in conjunction with Cooley and Lohnes (1971) and Gnanadesikan (1977).

Falconer, D. S. (1970). "Introduction to Quantitative Genetics", Oliver and Boyd, Edinburgh.

Finney, D. J. (1971). "Probit Analysis", Cambridge University Press, London.

Fleiss, J. L. (1973). "Statistical Methods for Rates and Proportions", Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Flessa, K. W. and Bray, R. G. (1977). On the measurement of size-independent morphological variability: an example using successive populations of a Devonian brachiopod. *Paleobiology* **3**, 350–359.

Ford, E. B. (1964). "Ecological Genetics", Methuen, London.

Gadel, F. and Mongin, D. (1973). Analyses sédimentologiques, géochimiques et malacologiques d'une carotte prélevée dans le Golfe du Lion (Méditerranée française). *Palaeogeogr. Palaeoclimatol Palaeoecol.* **13**, 49–64.

Glaessner, M. F. (1945). "Principles of Micropalaeontology", Melbourne University Press, Melbourne.

Gnanadesikan, R. (1977). "Methods for Statistical Data Analysis of Multivariate Observations", Wiley, New York.

Goldstein, M. and Dillon, W. R. (1978). "Discrete Discriminant Analysis", Wiley, New York.

Gordon, A. D. (1973). A sequence-comparison statistic and algorithm. *Biometrika* **60**, 197–200.

Gordon A. D. and Birks, H. J. B. (1972). Numerical methods in Quaternary palaeoecology. I. Zonation of pollen diagrams. *New Phytol.* **71**, 961–979.

Gordon, A. D. and Birks, H. J. B. (1974). Numerical methods in Quaternary Palaeoecology. II. Comparisons of pollen diagrams. *New Phytol.* **73**, 221–249.

This, and Gordon and Birks (1972), are the basic references for the slotting technique. Gordon and Birks (1974) also give an excellent account of how frequency data on several species can be used for quantitative palaeo-ecological reconstructions.

Gordon, A. D. and Reyment, R. A. (1979). Slotting of borehole sequences. *Math. Geol.* **11**, 309–327.

Gould, S. J. (1969). An evolutionary microcosm: Pleistocene and Recent history of the land snail P. (*Poecilizonites*) in Bermuda. *Bull. Mus. Comp. Zool.*, **138**, 407–531.

Gould, S. J. (1977). "Ontogeny and Phylogeny", Belknap Press, Harvard. University, Cambridge, MA.

A valuable reference for allometry as well as the history of thought on phylogeny and ontogeny.

Gould, S. J. and Eldredge, N. (1977). Punctuated equilibria and mode of evolution reconsidered. *Paleobiology* **3**, 115–151.

Gower, J. C. 1966. Some distance properties of latent roots and vectors used in multivariate analysis. *Biometrika* **53**, 325–338.

Principal coordinate analysis introduced and the duality between this Q-technique and principal component analysis discussed.

Gower, J. C. (1976). Growth-free canonical variates and generalized inverses. *Bull. Geol. Instn. Univ. Uppsala N.S.*, **7**, 1–10.

Development of the important concept of Burnaby (1966) for growth invariance in palaeontological and other work.

Hallam, A. (1978). How rare is phyletic gradualism and what is its evolutionary

significance. Evidence from Jurassic Bivalves. *Paleobiology* **4**, 16–25.

Hallberg, R. (1973). The microbiological C–N–S cycles in sediments and their effect on the ecology of the sediment–water interface. *Oikos Suppl.* **15**, 51–62.

Hayami, I. (1973). Discontinuous variation in an evolutionary species, *Cryptopecten vesiculosus*, from Japan. *J. Paleontol.* **47**, 401–420.
   Of fundamental significance in the palaeogenetical analysis of discrete shell characters.

Hayami, I. and Ozawa, T. (1975). Evolutionary models of lineage zones. *Lethaia* **8**, 1–14.

Heberer, G. (1974). "Die Evolution der Organismen", Vol. II (1), 408. Fischer Verlag, Stuttgart.
   A valuable reference text for examples of evolutionary analysis.

Hopkins, J. W. (1966). Some considerations in multivariate allometry. *Biometrics* **22**, 747–760.

Huxley, J. S. (1932). "Problems of Relative Growth", Methuen, London.

Imbrie, J. and Kipp, N. J. (1971). A new micropaleontological method for quantitative micropaleontology: application to a Late Pleistocene Carribean core. *In* "The Late Cenozoic Glacial Ages", (Ed. K. Turekian), 71–181. Yale University Press, New Haven.

Jolicoeur, J. (1963). The degree of generality of robustness in *Martes americana*. *Growth* **27**, 1–27.

Jolicoeur, P. and Mosimann, J. E. (1960). Size and shape variation in the painted turtle. *Growth* **24**, 335–354.
   The concept of Teissier's size factor and shape interpretations of principal components is presented in this paper, probably one of the most significant contributions yet made in applied biometry.

Jones, J. S. and Parkin, D. T. (1977). Experimental manipulation of some snail populations subject to climatic selection. *American Naturalist* **111**, 1014–1017.

Jöreskog, K. G. Klovan, J. E. and Reyment, R. A. (1976). "Geological Factor Analysis", Elsevier, Amsterdam.

Keen, M. C. (1971). A palaeoecological study of the ostracod *Hemicyprideis montosa* (Jones and Sherborn) from the Sannoisian of NW Europe. *In* "Paléoécologie des Ostracodes", (Ed. H. Oertli), *Bull. Centre Rech. Pau SNPA Suppl.* **5**, 523–543.

Kilenyi, T. I. (1969). The problems of ostracod ecology in the Thames estuary. *In* "The Taxonomy, Morphology and Ecology of Recent Ostracods", (Ed. J. Neale), 251–267. Oliver and Boyd, Edinburgh.

Kilenyi, T. I. (1971). Some basic questions in the palaeoecology of ostracods. *In* "Paléoécologie des Ostracodes", (Ed. H. Oertli), *Bull. Centre. Rech. Pau SNPA. Suppl.* **5**, 31–44.

Kinne, O. (Ed.). (1970). "Marine Ecology", Vol. 1, part 1. Environmental Factors, part 1. Wiley Interscience, New York.

Kinne, O. (Ed.). (1971). "Marine Ecology", Vol. 1 part 2. Environmental Factors. 683–1244. Wiley Interscience, New York.
   These two volumes constitute a monumental store of information on the roles of temperature and salinity in marine ecology. They are essential for anybody wishing to unravel palaeo-ecological relationships.

Lande, R. (1975). Natural selection and random genetic drift in phenotype evolution. *Evolution* **30**, 314–334.

Leamy, L. (1977). Genetic and environmental correlations of morphometric traits in randombred house mice. *Evolution* **31**, 357–369.
   Important reference for relating principal components analysis to standard genetical-statistical analyses by the analysis of variance on experimental trials.

Lewontin, R. C. (Ed.). (1978). "Population Biology and Evolution", Syracuse University Press, Syracuse, N.Y.
   This contains many useful case histories which are directly applicable to evolutionary studies in micropalaeontology.

Lewontin, R. C. (1974). "The Genetic Bases of Evolutionary Change", Columbia University Press, New York.

Liebau, A. (1971). Homologe Skulpturmuster bei Trachyleberididen und verwandten Ostrakoden. Dissertation, University of Berlin.

Lynch, E. J. (1964). "Formation Evaluation", Harper and Row, New York; Evanston, London.

Malmgren, B. (1974). Morphometric studies of planktonic foraminifers from the type Danian of Southern Scandinavia. *Stockh. Contr. Geol.* **29**, 1–126.

Malmgren, B. A. and Kennett, J. P. (1978). Test size variation in *Globigerina bulloides* in response to Quaternary paleooceanographic changes. *Nature (London)* **275**, 123–124.

Marquardt, D. W. and Snee, R. D. (1975). Ridge regression in practice. *Am. Statist.* **29**, 3–19.

Marriott, F. H. C. (1974). "The Interpretation of Multiple Observations", Academic Press, London and New York.

Maxwell, A. E. (1967). "Analysing qualitative data", Monographs on applied probability and statistics. Methuen, London.

May, R. M. (1973). "Stability and Complexity in Model Ecosystems", Monographs in Population Biology, No. 6, Princeton University Press, Princeton, N.J.

May, R. M. and Ostler, G. F. (1976). Bifurcations and dynamic complexity in simple ecological models *American Naturalist* **110**, 573–599.

Mayr, E. (1963). "Animal Species and Evolution", Belknap Press, Harvard University, Cambridge, MA.
   Still the basic reference for taxonomically orientated studies with a genetical flavour.

Mayr, E. (1969). "Principles of Systematic Zoology", McGraw-Hill, New York.

Moody, P. A. (1964). "Introduction to Evolution", Harper and Row, New York.

Moore, H. (1958). "Marine Ecology", Wiley, New York.
   A valuable reference for marine ecological observations on invertebrates.

Morrison, D. F. (1976). "Multivariate Statistical Methods", 2nd Ed, McGraw-Hill, New York.

Mosimann, J. E. (1970). Size allometry: size and shape variables with characterizations of the log-normal and generalized gamma distributions. *J. Am. Statist. Assoc.* **65**, 930–945.

Neale, J. E. (Ed.). (1969). "The Taxonomy, Morphology and Ecology of Recent Ostracoda", Oliver and Boyd, Edinburgh.
   A collection of papers on ostracod research, a few of which are of interest to the palaeoecologist.

Nyholm, K. G. (1962). Foraminiferer som fossil och som levande djur. *Svensk Naturvetenskap* **15**, 202–218.
   Results of fundamental importance for the worker on foraminifers presented in

a simple manner. The paper is provided with an English summary and English text captions.

Oertli, H. (Ed.) (1971). Colloque sur la Paléoécologie des Ostracodes. Bull. Centr. Rech. Pau SNPA. Suppl. 5, 953 pp.
Valuable source of ecological information for the worker on fossil and living ostracods.

Ozawa, T. (1975). Evolution of Lepidolina multiseptata (Permian foraminifer) in East Asia. Mem. Fac. Sci. Kyushu Univ. D, 23, 117–164.

Pielou, E. C. (1974). "Population and Community Ecology", 332. Gordon and Breach, New York.

Pielou, E. C. (1977). "Mathematical Ecology", Wiley Interscience, New York.
Two fundamental texts for anyone wishing to penetrate statistical ecology with the end in view of applying useful methods to palaeo-ecology.

Pirson, S. J. (1977). "Geologic Well Log Analysis", Gulf Publishing Co., Houston, Texas.
Electric logs applied to environmental interpretations.

Quenouille, M. H. (1972). "Rapid Statistical Calculations", Griffin London.

Prim, R. C. (1957). Shortest connection matrix network and some generalizations. J. Bell System Techn. 36, 1389–1401.

Rendel, J. M. (1967). "Canalization and Gene Control", Academic Press, London and New York.

Rensch, B. (1954). "Neuere Probleme der Abstammungslehre", Ferdinand Enke Verlag, Stuttgart.

Reyment, R. A. (1959). The foraminiferal genera Afrobolivina gen. nov. and Bolivina in the Upper Cretaceous and Lower Tertiary of West Africa. Stockh. Contr. Geol. 3, 1–57.

Reyment, R. A. (1960). Studies on Nigerian Upper Cretaceous and Lower Tertiary Ostracoda. Part 1. Stockh. Contr. Geol. 7, 238.

Reyment, R. A. (1963). Bohrlöcher bei Ostracoden. Pal. Zeitschr. 37, 283–291.

Reyment, R. A. (1963–66). Studies in Nigerian Upper Cretaceous and Lower Tertiary Ostracoda. Part 2 (1963), Stockh. Contr. Geol. 10, 1–286. Part 3 (1966), Stockh. Contr. Geol. 14, 1–151.
Taxonomical study of Nigerian Palaeogene ostracods and a biometrical and palaeo-ecological analysis for Cretaceous and Tertiary species.

Reyment, R. A. (1966a). Preliminary observations on gastropod predation in the western Niger Delta. Palaeogeogr. Palaeoclimatol. Palaeoecol. 2, 81–102.

Reyment, R. A. (1966b). See Reyment (1963–66 above).

Reyment, R. A. (1966c). Afrobolivina africana (Graham, de Klasz, Rérat); quantitative Untersuchung der Variabilität einer paleozänen Foraminifere. Eclog. Geol. Helvet. 59, 319–337.
Identification of stages in the life cycle of a benthic foraminifer identified by multivariate statistical analysis.

Reyment, R. A. (1969a). Covariance structure and morphometric analysis—a contribution to paleogenetics. Math. Geol. 1, 185–197.

Reyment, R. A. (1969b). Some case studies of the statistical analysis of sexual

dimorphism. Bull. Geol. Instn Univ. Uppsala NS 1, 97–119.

Reyment, R. A. (1969c). A multivariate paleontological growth problem. Biometrics 22, 1–8.

Reyment, R. A. (1969d). Interstitial ecology of the Niger Delta. Bull. Geol. Instn Univ. Uppsala NS, 1, 121–159.

Reyment, R. A. (1970). Spectral breakdown of morphometric chronoclines. Math. Geol. 2, 365–376.

Reyment, R. A. (1971). "Introduction to Quantitative Paleoecology", Elsevier, Amsterdam.
This presents a collection of case histories drawn from quantitative palaeo-ecological studies in which methods of statistical ecology have been used, as well as methods peculiar to palaeo-ecology.

Reyment, R. A. (1972). Models for studying the occurrence of lead and zinc in a deltaic environment. In "Mathematical Models of Sedimentary Processes", (Ed. D. F. Merriam), 233–245. Plenun Press, New York.

Reyment, R. A. (1975). Canonical correlation analysis of hemicytherinid and trachyleberinid ostracodes in the Niger Delta. Bull. Am. Paleontol 65 (282), 141–145.

Reyment, R. A. (1975b). Analysis of a generic level transition in Cretaceous ammonites. Evolution 28, 665–676.

Reyment, R. A. (1976). Chemical components of the environment and Late Campanian microfossil frequencies. Geol. Fören. Stockh. Förh. 98, 322–328.
The ecolog is introduced here for frequencies of microfossils related to the geochemistry of shales.

Reyment, R. A. (1978a). Quantitative biostratigraphical analysis exemplified by Moroccan Cretaceous ostracods. Micropaleontology 24, 24–43.
Analysis of extreme cases of ornamental variation by statistical analysis and the use of the results for correlating the Cenomanian–Turonian boundary.

Reyment, R. A. (1978b). Graphical display of growth-free variation in the Cretaceous benthonic foraminifer Afrobolivina afra. Palaeogeogr. Palaeoclimatol. Palaeoecol. 25, 267–276.

Reyment, R. A. and Banfield, C. (1976). Growth-free canonical variates applied to fossil foraminifers. Bull. Geol. Instn Univ. Uppsala, NS 7, 11–21.

Reyment, R. A. and Brännström, B. (1962). Certain aspects of the physiology of Cypridopsis (Ostracoda, Crustacea). Stockholm Contr. Geol. 9, 207–242.

Reyment, R. A. and Hallberg, R. O. (1967). Notes on free oxygen in connection with negative redox potentials in relation to ostracods. Crustaceana 13, 124–126.

Reyment, R. A., Hayami, I. and Carbonnel, G. (1977). Variation of discrete morphological characters in Cytheridea (Crustacea: Ostracoda). Bull. Geol. Instn Univ. Uppsala. NS 7, 23–36.
Detailed analysis of variation in marginal spines and lateral ornament of a Miocene ostracod.

Reyment, R. A. and Ramdén, H.-Å. (1969). FORTRAN IV program for canonical variates analysis for the CDC 3600 computer. Computer Contr. 47, 39. State Geol. Survey, Lawrence, Kansas.

Reyment, R. A. and Van Valen, L. (1969). Buntonia olokundudui sp. nov. (Ostracoda, Crustacea). A study of meristic variation in Paleocene and recent ostracods. Bull. Geol. Instn Univ. Uppsala, NS 1, 83–94.

Roger, J. (1976). "Paléontologie Evolutive", Masson, Paris.

Excellent presentation of modern ideas in palaeontological research.

Ross, G. J. S. (1969). Minimum spanning trees. Algorithm AS 13. *Appl. Stat.* **18**, 103–104.

Rudman, A. J. and Lankston, R. W. (1973). Stratigraphic correlation of well logs by computer technique. *Bull. Amer. Ass. Petrol. Geol.* **57**, 577–588.

Schäfer, W. (1962). "Aktuo-Paläontologie", Kramer, Frankfurt-am-Main.
   Monographic treatment of palaeontological questions viewed in the light of the study of Recent organisms in tidal flats of the North Sea.

Schindel, D. E. and Gould, S. J. (1977). Biological interaction between fossil species: character displacement in Bermudian land snails. *Paleobiology*, **3**, 259–269.

Schopf, T. J. M. and Dutton, A. R. (1976). Parallel clines in morphologic and genetic differentiation in a coastal zone marine invertebrate: the bryozoan *Schizoparella errata. Paleobiology*, **2**, 255–264.

Shaw, A. B. (1964). "Time in Stratigraphy". McGraw-Hill, New York.
   An early attempt at using statistical methods for stratigraphical purposes.

Sillén, L. G. (1963). How has seawater got its present composition? *Svensk Kemisk Tidskrift* **75**, 161–177.

Simpson, G. G. (1953). "The Major Features of Evolution", Columbia University Press, New York.
   A basic reference for concepts of evolution with an emphasis on palaeontological points of view.

Simpson, G. G., Rowe, A. and Lewontin, R. (1960). "Quantitative Zoology", McGraw-Hill, New York.

Sinnott, E. W., Dunn, L. C. and Dobzhansky, T. (1958). "Principles of Genetics", 5th edn, McGraw-Hill, New York.

Sokal, R. R. and Rohlf, F. J. (1969). "Biometry", Freeman, San Francisco.

Stewart, I. (1975). The seven elementary catastrophes. *New Scientist* **68** (976), 447–448.

Swain, F. M. (Ed.). (1975). *Biology and Paleobiology of Ostracoda. Bull. Am. Paleontol.* **65**, 687.

Thom, R. (1969). Topological models in biology. *Topology* **8**, 313–335, 89–116. *In* "Towards a Theoretical Biology", Vol. 3. (Ed. C. H. Waddington) Edinburgh University Press, Edinburgh.

Thom, R. (1975). "Structural Stability and Morphogenesis", Benjamen, Reading, MA.
   English language version of Thom's principal treatise on catastrophe theory.

Valentine, J. W. (1973). "Evolutionary Paleoecology of the Marine Biosphere", Prentice-Hall, Eaglewood Cliffs.

Waszowiak, R. (1962). Geochemische Untersuchungen an rezenten Molluskenschalen mariner Herkunft. *Freiberger Forschungshefte* **C136**, 155.

White, M. J. D. (1978). "Modes of Speciation". Freeman, San Francisco.

Williams, R. (1969). Ecology of the Ostracoda from selected marine intertidal localities on the coast of Anglesey. *In* "The Taxonomy, Morphology and Ecology of Recent Ostracoda", 299–327. Oliver and Boyd, Edinburgh.

Wright, S. (1968). Evolution and the Genetics of Populations. I. "Genetic and Biometric Foundations". University of Chicago Press, Chicago.

Zeeman, E. C. (1975). "Catastrophe Theory", Addison Wesley, Reading MA.

Index