CLUSTER ANALYSIS : AN AID TO ENVIRONMENTAL INTERPRETATION?

D.W.Perry
Department of Geography
University of Birmingham

Inter-disciplinary exchanges of analytical techniques can lead to problems if the demands they pose,especially with regard to data, are not fully appreciated.Before such a technique is adopted it's use,and possible abuse,should ideally be investigated to determine whether or not it is of significant value.This paper examines the use of cluster analysis in this context,investigating some of the difficulties facing workers in the fields of Palaeoecology and Environmental Archaeology.The views expressed have been reached as a consequence of research carried out in the Leverhulme Research Project on Viking Settlement,Climate and Environmental Change, which is examining patterns of Viking expansion and settlement, and consequent floral and faunal changes,around the North Atlantic.

The questions asked by an investigator obviously determine the design of sampling programme to be used,and hence nature of the analyses that are employed.Sampling requirements are defined to ensure the supply of sufficient data,enabling some form of analysis and interpretation to be performed.The important qualifier in this statement is *sufficient*.This not only implies quantity but quality.It is pointless collecting vast amounts of data if it turns out to be invalid and incapable of meaningful analysis.This point is of fundamental importance in ecological studies,where great emphasis is placed on the careful planning of sample programmes in order to minimise bias and error,thereby facilitating mathematical analysis of the data.In this respect the Institute of Terrestrial Ecology publish a checklist of

questions(Jeffers,1979) within the framework of which a valid sampling programme can be planned.Four of the most important questions are given below.

Defining the population about which inferences are to be made.

1. If there is no logical and practical way of finding samples which are representative of your defined population, is it worth continuing with the investigation at all?

Sample units

2. If sample units are not naturally defined how are they to be defined and limited in space and time?
3. Are you satisfied that there is a sufficiently logical definition of sample units to justify proceeding with the investigation?

Size of samples

4. Is the size of the sample you propose to take adequate i.e. neither too small or too large?

It is the failure of palaeoecological studies to meet any or all of requirements such as these that might at first sight be seen as the *Achilles Heel* of the discipline.However,it is crucial that these requirements derived from the practice of ecology are seen in the correct palaeoecological perspective.Whereas such strictures are arguably vital to the effective practice of ecology,they are by contrast only desirable in respect of

palaeoecology,which is by its very nature concerned with changes in a much broader sense than the typically detailed,smaller scale analyses common to ecology.The basic difference between the disciplines is that the palaeoecologist is far removed in time from the populations about which he would wish to draw inferences. Unlike the ecologist,he has to deal with death assemblages (*thanatocoenoses*) which are the only source of information about the life assemblages (*biocoenoses*) that are of interest.In effect a sample of a sample of the population has to be dealt with; obviously this is a far from desirable situation.Also,there is no way of determining whether the samples are representative,or actual sample sizes equal.To take a modern analogy the quadrat sizes are unequal and the insects recovered are most likely representative of a greater area still (see Kenward,1975).When the dimension of time is added to this equation the data do seem rather inadequate by ecological standards,but as the questions being asked are usually quite different this is not a major methodological difficulty.The study of the mode of formation of palaeoecological samples and associated problems is known as *taphonomy*.It is all too easy to view the problems posed by taphonomy as being insurmountable,indeed in the opinion of Rollins and Donahue (1975),"many palaeoecologists are unwilling to examine seriously fossil assemblages from the point of view that such assemblages might reflect once living communities".

Considerations such as these are only a problem when seen in relation to the question of scale of interest.To workers investigating climatic change the presence in a thanatocoenosis of a transported background element,the *allochthonous* component, is a bonus,enabling conclusions to be drawn concerning a comparatively wide area.In the examination of smaller scale problems the difficulty of determining whether or not the fossils

were members of the same ecological community arises.In studies involving micro-scale archaeological interpretations the allochthonous component may cause an over-representation of certain habitats,as well as a general blurring of ecological information (Kenward,1977).The problems posed by the mixed origins of many archaeological assemblages,and methods for separating the in situ from the transported components have been discussed in detail by Kenward (1978).

A problem pertinent to the work of the Leverhulme Project results from the low diversity of the Icelandic coleopteran fauna.With the exception of most archaeological samples, numbers of individuals recovered are low,and this obviously poses problems from a statistical viewpoint;one answer to which is to consider data only in a presence-absence form.In any event Johnson (1962) provides a cautionary note,stating that for taphonomic reasons,"the number of individuals of a particular species is difficult to interpret".

Given these considerations it would be natural to suggest that anything other than the simplest treatment of the data is useless.Therefore,the question,"Why use cluster analysis?",needs to be examined.Bullock (1971),in considering the investigation of sample data identified two distinct phases of activity: sample description and sample comparison.If sample description, and hence environmental interpretation,is the aim then numerical techniques need not be necessary,sceptics would add *or desirable*. Detailed interpretations can be achieved without ever reaching for a pocket calculator,let alone a mainframe computer.The work of Coope and Osborne amongst others provides an excellent example of what has been labelled this intuitive natural history approach.Kenward has tried to develop this numerically by using indices of diversity as aids to interpretation,especially as

a means of resolving the autochthonous/allochthonous dilemma. However,the real benefit of quantification becomes apparent when the second step of sample comparison,and consideration of environmental variability,is taken.The accurate identification of environmental change can only be achieved by rigorous examination of the data.Whether or not this procedure is effective is largely dependent upon the *mental dexterity* of the investigator,especially when considering the typically large numbers of samples and unwieldy species listings that have to be dealt with.Any method that reduces the initial complexity of this task is useful,and it is at this juncture that computer based analyses can usefully be employed.

The presentation of an easily comprehended summary of the relationships within a data set provides a firmer footing for further investigation than a preliminary *by eye* approach. Experience has shown it to be profitable to re-examine the data with the benefit of a mathematically based interpretation, as this provides an invaluable framework around which the intuitive approach can be used to best effect.This echoes Hodsons (1969) view that "the most fruitful approach seems to be to look at the data in as many different ways as possible, since each new viewpoint may bring out hidden details of real significance ...The intention is to achieve an objectively controlled representation of the data so that interpretation becomes more sraightforward".

The CLUSTAN analysis package,being widely used and readily available,was chosen as the most appropriate means of providing a framework as outlined.It's use resulted from a deliberate decision to make the best use of the data despite it's shortcomings.The availability of a wide range of techniques within the same package comply with the suggestion of Everitt

(1980) that it is useful to apply more than one technique, each based on different assumptions,in order to avoid mis-interpretations.Hierarchical,divisive and iterative relocation procedures have been tried and found to be useful.Clearly,the uppermost question concerning any technique is *does it work* ?

Early on in the present project the techniques were tested to determine their validity.An interpretation produced by Coope and Brophy (1972),based upon a Late Glacial coleopteran succession from Glanllynau in N.Wales was chosen as the benchmark. A CLUSTAN based analysis of the data was found to be in close agreement with the description of environmental variability proposed by the authors.An archaeological test was provided when the coleopteran data from the Brigg 'raft' was analysed (Buckland,1981).The results concurred with those of other workers and also provided interesting information about the complex pattern of environmental change.More recently results obtained from data from an excavation at the farmsite of Storaborg in S.Iceland have proved useful by providing otherwise unobtainable insights into the data,and have stimulated further intuitive investigation.Examples of the graphical results of this investigation are given in figures 1 and 2.Subtle differences in the synanthropic elements of the sample faunas have been identified,and these may well suggest an explanation for the observed pattern of variation in terms of usage of the rooms. Supplementary archaeological and ecological information is being examined in order to investigate the observed pattern of variation more fully,this activity being a direct consequence of the CLUSTAN analysis.

From present experience it is clear that CLUSTAN is an aid to environmental interpretation; but it is simply an aid and

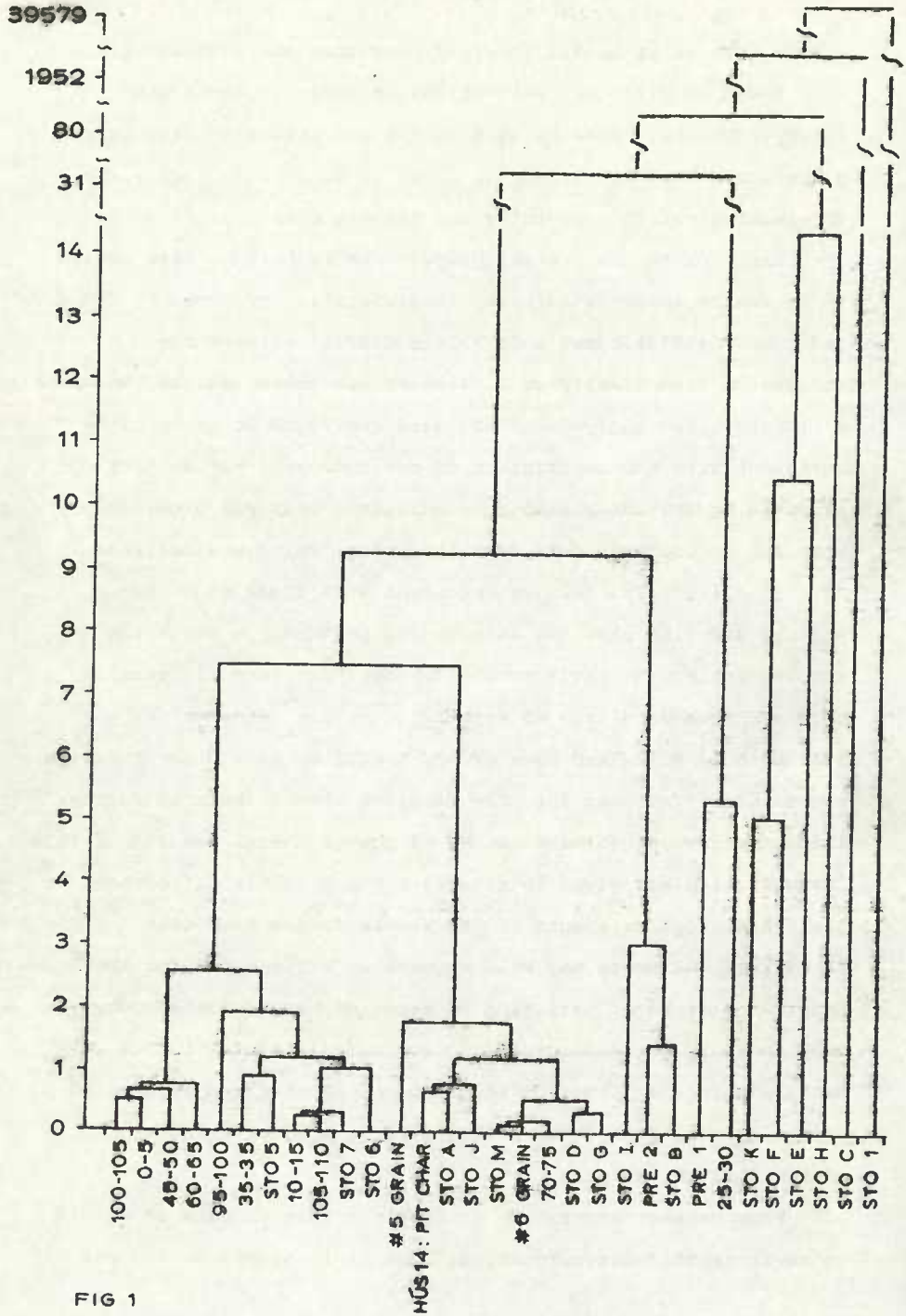
39579
1952
80
31
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
100-105
0-5
45-50
60-65
95-100
35-35
STO 5
10-15
105-110
STO 7
STO 6
#5 GRAIN
HÜS14: PIT CHAR
STO A
STO J
STO M
#6 GRAIN
70-75
STO D
STO G
STO I
PRE 2
STO B
PRE 1
25-30
STO K
STO F
STO E
STO H
STO C
STO 1

FIG 1

nothing more.CLUSTAN is used as a begining,not as an end,and as such takes it's place alongside the more traditional approaches in providing insight into the data.CLUSTAN cannot do otherwise than provide mathematical classifications,taking no account of ecological considerations,viewing each species as being equally distinct from the others.This is a major drawback as the techniques effectively ignore any relationships above the specific level;generic and family similarities cannot be introduced without gross oversimplification.CLUSTAN therfore provides a rather simplistic treatment of the data,and the palaeoecologist must recognise and allow for this effect,using practical experience to determine useful,significant aspects of the computer classifications.In doing this the problems connected with the data,and those of the procedures used must be appreciated. Orloci (1975) appreciated the need for caution,arguing against overestimating the limits to which statistical and other mathematical techniques can be put to useful service.

The real value of cluster analysis,as typified by CLUSTAN, seems to be in producing further questions and hypotheses to be developed along more traditional lines.Greig-Smith (1980) states that"we can only judge by results,not by whether they reproduce our preconceptions,but by whether they are useful in practice or fruitful of hypotheses".The ability of the computer to sift data in no way relieves the palaeoecologist of his responsibility for careful judgement of the results (Raup and Crick,1979).These sentiments are succinctly expressed by Gauch and Whittaker (1981),who state that "classification remains partly an art to which the ecologists experience and understanding may contribute much".Bearing this in mind the most appropriate answer to the question posed in the title of this paper is perhaps that *beauty is in the eye of the beholder.*

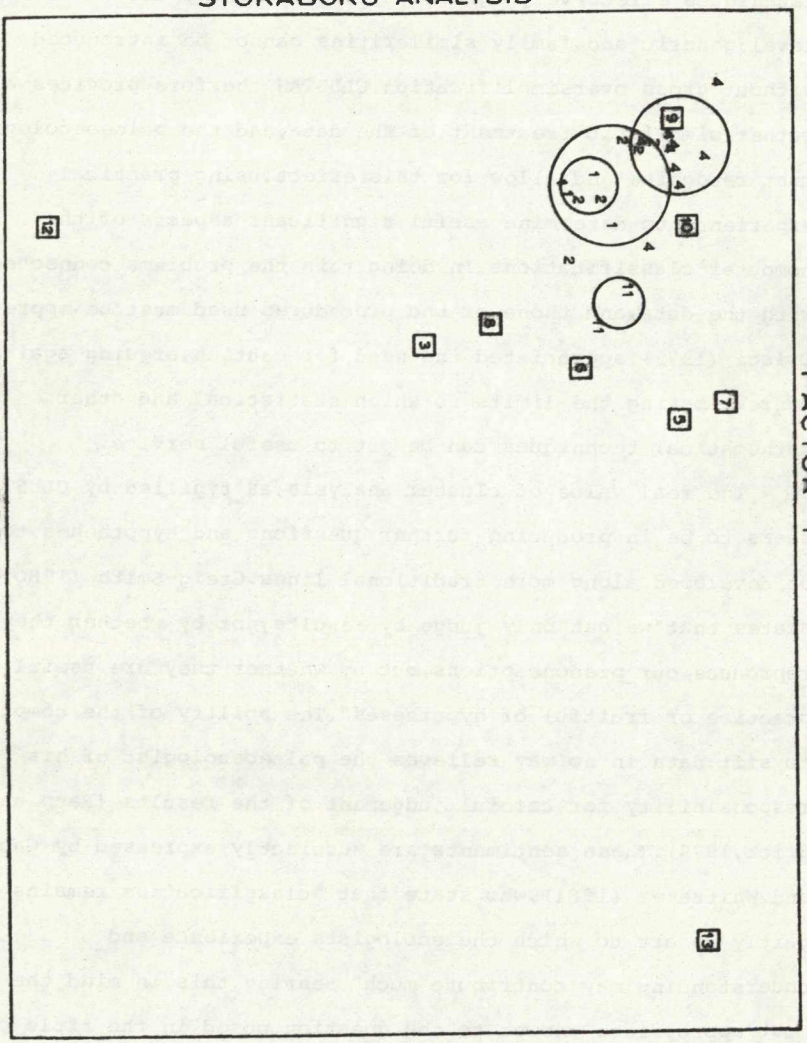

FIG 2

Acknowledgements

Thanks are due to the many people who have provided help and encouragement during my exploration into cluster analysis.

In particular thanks are extended to Paul Buckland for his guidance and enthusiasm over the last three years, and also to Harry Kenward for a particularly productive afternoon in York. Lastly thanks are due to Russell Coope for his valued criticisms of an earlier draft of this paper.

REFERENCES

Buckland,P.C. 'The Brigg'Raft'and her Prehistoric Environment'
1981 ed S.McGrail BAR BRITISH SERIES 89 pp 155-187

Bullock,J.A. 'The Investigation of samples containing many
1971 species I.Sample description'
BIOL.J.LINN.SOC. 3 pp 1-21

Coope,G.R. & 'Late Glacial environmental changes indicated by
Brophy,J.A. a coleopteran succession from N.Wales'
1972 BOREAS I :2 pp 97-142

Everitt,B. CLUSTER ANALYSIS
1980 Halstead Press

Gauch,H.G. & 'Hierarchical classification of Community Data'
Whittaker,R.H. J.ECOLOGY 69 pp 537-557
1981

Greig-Smith,P.'The development of numerical classification and
1980 ordination' VEGETATIO 42:3

Hodson,F.R 'Searching for structure within multivariate
1969 archaeological data'
WORLD ARCHAEOLOGY pp 90-105

Jeffers,J.N.R 'Statistical checklist..Sampling'
1979 INSTITUTE OF TERRESTRIAL ECOLOGY

Johnson,R.G 'Interspecific associations in Pennsylvanian
1962 Fossil Assemblages'
J.GEOLOGY 70

Kenward,H.K 'Pitfalls in the Interpretation of Insect Death
1975 Assemblages'
J.ARCH.SCI 2 pp 85-94

Kenward,H.K 'The value of insect remains as evidence of
1977 ecological conditions on Archaeological sites'

RESEARCH PROBLEMS IN ZOOARCHAEOLOGY
ed D.R.Brothwell,K.D.Thomas & J.Clutton-Brock
Occ.Pub.No 3 Institute of Archaeology

Kenward,H.K 1978 'The Analysis of Archaeological Insect Assemblages:a New Approach'
THE ARCHAEOLOGY OF YORK 19 C.B.A.

Orloci,L.A 1975 MULTIVARIATE ANALYSIS IN VEGETATION RESEARCH
Junk

Raup,D.M.& Crick,R.E 'Measurement of Faunal Similarity in Palaeontology'
J.PALEONTOLGY 53:5

Rollins,H.B.& Donahue,J 1975 'Towards a theoretical basis of palaeoecology: concepts of community dynamics'
LETHAIA 4