THE DENDROGRAPH

By

RICHARD B. McCAMMON

and

GUENTHER WENNINGER

University of Illinois at Chicago

COMPUTER CONTRIBUTION 48
State Geological Survey
The University of Kansas, Lawrence
1970
EDITORIAL STAFF
D.F. Merriam* Editor
Technical Editors
John C. Davis* Owen T. Spitz
Associate Editors
Frederic P. Agterberg John H. Hefner* Sidney N. Hackens*
Richard W. Fetzner* J. Edward Klovani William C. Krumbein*
James M. Forgetson, Jr.* R.H. Lippert
John C. Griffiths
John W. Harbaugh*

Editor's Remarks

COMPUTER CONTRIBUTION 48, "The dendrograph" by R.B. McCammon, is another program for graphic display. Geologists are increasingly using quantified classification techniques and will find this computer program of value in their work. For a limited time the Geological Survey will make available the program on magnetic tape for $10.00 (US) or $20.00 (US) on punched cards.

For a complete list of COMPUTER CONTRIBUTIONS, write Editor, COMPUTER CONTRIBUTIONS, Kansas Geological Survey, The University of Kansas, Lawrence, Kansas 66044.

Computer Contribution

1. Mathematical simulation of marine sedimentation with IBM 7090/7094 computers, by J.W. Harbaugh, 1966 .... (out of print) $0.50
2. A generalized two-dimensional regression procedure, by J.R. Dempsey, 1966 .... $0.75
3. FORTRAN IV and MAP program for computation and plotting of trend surfaces for degrees 1 through 6, by Mont O'Leary, R.H. Lippert, and O.T. Spitz, 1966 .... $0.75
4. FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer, by J.C. Davis and R.J. Sampson, 1966 .... $0.50
5. FORTRAN IV program using double Fourier series for surface fitting of irregularly spaced data, by W.R. James, 1966 .... $0.75
6. FORTRAN IV program for estimation of cladistic relationships using the IBM 7040, by R.L. Bortcher, 1966 .... $1.00
7. Computer applications in the earth sciences: Colloquium on classification procedures, edited by D.F. Merriam, 1966 .... $1.00
8. Prediction of the performance of a solution gas drive reservoir by Muskat's equation, by Apolonio Baca, 1967 .... $0.50
9. FORTRAN IV program for mathematical simulation of marine sedimentation with IBM 7040 or 7094 computers, by J.W. Harbaugh and W.J. Wahlstedt, 1967 .... $1.00
10. Three-dimensional response surface program in FORTRAN II for the IBM 1620 computer, by R.J. Sampson and J.C. Davis, 1967 .... $0.75
11. FORTRAN IV program for vector trend analyses of directional data, by W.T. Fox, 1967 .... $1.00
12. Computer applications in the earth sciences: Colloquium on trend analysis, edited by D.F. Merriam and N.C. Cocke, 1967 .... $1.00
13. FORTRAN IV computer programs for Markov chain experiments in geology, by W.C. Krumbein, 1967 .... $1.00
14. FORTRAN IV programs to determine surface roughness in topography for the CDC 3400 computer, by R.D. Hobson, 1967 .... $1.00

* Active Member, ° Associate Member, † Junior Member, American Association of Petroleum Geologists
THE DENDROGRAPH

by

Richard B. McCammon and Guenther Wenninger

ABSTRACT

The dendrogram has been developed as a two-dimensional diagram for depicting the mutual relationships among a group of objects whose pairwise similarities are given. The construction of the dendrogram is based on results of clustering using the unweighted pair-group method. The resultant hierarchical arrangement reflects both the within-group and the between-group similarity. In order to accentuate the hierarchical group structure, a rule of ordering is used to impart a pyramid shape to the dendrogram. To allow greater flexibility in the computer program for constructing dendrograms, an option is provided for the user to enter either correlation coefficients or distance functions as the measure of pairwise similarity between objects. Two examples are given to illustrate the usefulness of dendrograms in extracting meaning from multivariate data arrays.

INTRODUCTION

The search for pattern characterizes much of what geologists do at present. For the most part, the search is concentrated on the elucidation of multivariate data arrays. In many instances, this has been facilitated through the use of graphic forms of data display. One notable graphic form which has found application in a number of geological problems is the dendrogram. Although devised originally by systematic zoologists to illustrate the hierarchical relationships among organisms based on observed similarities (Mayr, Linsley, and Usinger, 1953, p. 58), the dendrogram is used increasingly by the geologist to depict the mutual relationships for (1) different samples with multiple attributes, and (2) different variables measured for a set of samples. In either situation, what is desired and usually obtained is data compression without significant loss of information. Because the dendrogram has consistently had this effect on the expression of multidimensional data, it has been widely in geology (Purdy, 1963; Behrens, 1965; Bonham-Carter, 1968; Parks, 1966; Merriam and Sneath, 1966; Valentine and Peddicord, 1967; Rowell, 1967; McCammon, 1968a).

Recently, the unidimensionality of the dendrogram has been extended to two dimensions and the new graphic form has been termed the dendrogram (McCammon, 1968b). The dendrogram has the advantage of displaying similarities within groups as well as between groups of ordered objects. This tends to lessen the distortional effect inherent in any portrayal of data in fewer than the original number of dimensions.

The purpose here is to describe a computer program that produces dendrograms. A program is provided which makes it possible for anyone having access to a computer and a plotter system to produce his own dendrograms. If a user does not have access to a plotter, it is possible to modify the plotting portion of the program so that the dendrogram can be produced on a printer. The user also can construct dendrograms manually using a straightedge and ruler with results computed by the program. The reason for offering the program with plotter is that with the time-sharing concept coming into vogue with computers, it is likely that most computer centers will soon have plotter capabilities.

To make the program as usable as possible, the assumption was made that the user previously would have calculated or somehow estimated the pairwise similarities of the items which he wished to arrange in hierarchical order. The reason, as numerical taxonomists have found (Sokal and Sneath, 1963, p. 166), is that the problem of deciding the best generalized measure of pairwise similarity for a given set of objects having multiple attributes is unresolved. Furthermore, if the items to be ordered represent different variables measured for a set of samples, a problem as to which measure to choose arises whenever continuous variables are to be correlated with discrete variables. Also, there is a problem if measurements for certain samples are missing. The attitude here is that the user of the program must treat problems as unique to his particular set of data, and the necessary steps are taken beforehand so that the input is a set of pairwise similarity coefficients.

One main option, however, exists in the program. In general, there are two broad classes of measures of pairwise similarity. On one hand, there are values that increase as the two objects compared more resemble one another. The notable example is the correlation coefficient defined for paired observations for two continuous variables. On the other hand, there are values that decrease the more the two objects resemble one another. The notable example here is the Euclidean distance measure defined
for two objects located as points in some coordinate space. Other measures which are analogous to the Euclidean measure but do not possess necessarily full metric properties may be used. Most of these, however, satisfy all but the triangular inequality and are called semimetrics. Both measures are accepted by the program.

During execution, the program operates on the similarity measures as if they were either metric or semimetric quantities. This requires that if correlation-type measures are entered as input, it is necessary to transform them into metric quantities. The transformation used is the arc cosine transformation. In the situation of the correlation coefficient, it transforms the correlation measure into one possessed with all properties of a metric (Blumenthal, 1953, p. 16). The transformation has a similar effect on Jaccard's coefficient. The transformation is diagrammed in Figure 1. In instances where correlation-type coefficients are used, the arc cosine transformation will have the desired effect of converting coefficients to distance-type measures. To ease the subsequent interpretation of the resultant dendrograms generated by correlation type measures, the within-group similarity measures are transformed by taking the cosine and scaling the within-group similarity axis in correlation units.

The method used for ordering the objects to form the dendrograph has been described previously (McCann, 1968b). Essentially, the method is the unweighted pair-group method of clustering (UWPGM) modified to provide, in addition, intergroup similarity. The literature on cluster analysis has grown enormously over the past few years and it would be impossible here to review its critical aspects. Also it is not possible to justify the preference of UWPGM over the dozens of other algorithms that have been proposed. Perhaps because it is one of the simplest to implement and because it treats objects equally, UWPGM has gained favor among cluster analysts. The assumptions underlying the various clustering schemes now in circulation however need a closer examination. An important step in this direction has been made recently by Gower (1967). For the present, UWPGM appears as versatile as any of the other methods.

Acknowledgments. The algorithm used to cluster objects and the computer program for constructing dendrograms on a plotter were developed originally at Gulf Research & Development Company. Special thanks go to Robert O'Hearn and Alexandra Tavlarides for assistance in development of the plotter program. Since that time the program has been rewritten for general use by the junior author. Special thanks go to Michael Goldman, David Rappaport, and Dennis Barley of the Computer Center staff at the University of Illinois at Chicago Circle for making available the plotting subroutines for the 1627 Calcomp plotter.

PROGRAM OPERATION

Program Dimensions

The program is dimensioned currently to cluster and construct a dendrogram for 200 objects. This number may be increased simply by increasing the dimensions of the appropriate program variables provided the necessary core storage is available at execution. No change in the logic of the program is necessary. To change the dimensions of the program variables, it is necessary only to change their dimensions in the main program. This has been made possible through the use of vector arrays in the calculations. Although the object name array is a matrix, the row dimension is fixed regardless of the number of objects to be clustered.

Order of Input Cards
1. Program control and title card
2. Plotter control card
   Object name array cards
3. Variable format input data card
   Data cards
Program usage

Card 1:

Cols.

1-4  M = number of objects to be clustered (negative M should be used if correlation measure coefficients are given; positive M should be used if distance measure coefficients are given)

5-8  KLAM = frequency with which intermediate stages of clustering are printed (if KLAM=0, only the final stage of clustering is printed)

9-80 A = problem identification which will appear as the title for the dendrograph

Card 2:

1-10  AENG = length of dendrograph in inches (between-group scale axis) (AENG \leq 120.)

11-20  AHI = height of dendrograph in inches (within-group scale axis) (AHI \leq 8.)

21-25  FNCR = units that tic marks along the within-group scale axis are apart (if FNCR=0., then, for +M, FNCR=1.0; for -M, FNCR=0.1)

Object name array cards

Columns 1-72 are used to read object names. Sixteen columns are reserved for each object name and there are five names to a card. The order in which object names are given must be the same as the ordering of the pairwise similarity coefficients.

Card 3:

Columns 1-72 are used to read the format for the pairwise similarity coefficients. The order in which the data are read is given as follows.

Data Cards:

The pairwise similarity coefficients are read as a lower triangular matrix with one row of the matrix read at a time. Thus,

\[
\begin{align*}
c_{11} & \\
c_{21} & c_{22} \\
c_{31} & c_{32} & c_{33}
\end{align*}
\]

is read as:

first card: \( c_{11} \)

second card: \( c_{21}, c_{22} \)

third card: \( c_{31}, c_{32}, c_{33} \)

If more than one card is needed to read any row, the data are continued on additional cards until all of the row is read.

The sequence of cards is repeated if a new problem is to be read.

Subprograms

The listed subprograms are provided along with the main program. The user need only check that the card decks are present before program execution.

IPOS   integer compare function
SEARCH  element search function
DENDRO  dendrograph plot subroutine

The following subroutines refer to plotter instructions and must be supplied by the user because plotter arrangements differ from one computer installation to the next. A more complete description follows this listing:

LIMIT   set maximum length
BGNPLT  begin plot
ENDPLT  end plot
SYMBOL  symbol plot
NUMBER  number plot
PLOT    location plot with pen control

The following are excerpts quoted from part of a user's manual prepared by the staff of the Computer Center at the University of Illinois at Chicago. The information contained in the excerpts should be sufficient for the user to prepare his own subroutines compatible with the plotter at his computer installation. The subroutines described here were designed to provide capability for plotting on a
Calcomp 1627 Plotter via the System 360 channel adapter.

Introduction

This is a user-oriented program designed to produce an X-Y plot of digital data. These plotting subroutines produce a plot file on the disk which, upon completion, is plotted via the 360-1800 channel adapter on the 1627 Plotter. The plotter converts tabulated digital information into graphic form. Bar charts, flow charts, organization charts, engineering drawings, and maps are among the many graphic forms of data which can be plotted on the 1627 Plotter.

The actual recording is produced by the incremental movement of the pen on the paper surface (y-axis) and/or the paper under the pen (x-axis). The pen is mounted on a carriage that travels horizontally across the paper as viewed from the front of the plotter. The vertical plotting motion is achieved by rotation of the pen feed drum, which also acts as a platen.

The drum and the pen carriage are bi-directional; i.e., the pen moves forward or backward, and the pen moves left or right. Control is also provided to lower or raise the pen to or from the paper surface. The pen remains in the "up" or "down" position until directed to change to the opposite status.

The drum and pen carriage movements and the pen status are controlled by digits transferred to the 1627 Plotter from the S/360 file. Each digit is decoded into a directional signal and relayed to the 1627 Plotter. Each signal to the Plotter causes a 1/100 inch incremental movement of the pen carriage and/or paper, or pen up or a pen down movement.

The maximum size of the plot in the y-direction is limited by the length of the drum; on the plotter at this installation, this is 11 inches. The initial origin is established .54 inches above the bottom of the page and the maximum size of the plot in the x-direction is 24 inches; thus the range of values available is:

\[-0.54 \leq Y \leq 10.46 \quad \text{and} \quad -4.0 \leq X \leq 20.0\]

All actual plotting is done in Background Partitions after the user has ended execution; because it is in background, both the S/360 and the 1800 can proceed in their normal facilities. FORTRAN is the only language required for the use of this program with the addition of the following JCL card:

//GO.PLOT18 DD DSN=SYST1.PLOTIN, DISP=(MOD,KEEP)

1. Subroutine BGNPLT

Calling Sequence:

CALL BGNPLT (I,EBCDCIC)

where:

1 is the number of characters in the EBCDCIC String

EBCDCIC is the Plot Identification Literal String. It is to be in quotes or H format. IT MUST BE THE PROGRAMMER'S NAME.

purpose: This subroutine must be called before any plotting is to be done. It opens the plot file and gives the user exclusive use of it. The EBCDCIC string will be drawn as a identifier .5 inches high along the y axis. The pen is then moved 4 inches away from the identifier and placed .54 inches away from the edge of the paper.

ALL LIMITS ARE RESET.

example:

CALL BGNPLT (9, 'NIXON R M')

Note: 1 \leq 9

2. Subroutine ENDPLT

Calling Sequence:

CALL BGNPLT (I,EBCDCIC)

purpose: This subroutine closes and releases the user's exclusive control of the plot file. Then it positions the pen for the next plot.

Note: ENDPLT must be called before another call to BGNPLT can be made.

3. Subroutine LIMIT

Calling Sequence:

CALL LIMIT (XL)

where:

XL: the maximum length of the plot in inches

purpose: This subroutine allows the user to specify the maximum length of the desired plot.

example:

XL = 24,

CALL LIMIT (XL)

Note: XL \leq 120.

4. Subroutine NUMBER

Calling Sequence:

CALL NUMBER (XA,YA,HEIGHT,(ANUM),THETA, N)

where:

XA and YA: the x and y coordinates in inches of the bottom left-hand corner of where you want the number to appear.

HEIGHT: the height in inches of the number or symbol to be plotted. For legibility it should be at least .07.

ANUM: floating point number or a floating point variable name where the number is located.

THETA: degrees from the horizontal, which you want the number to be drawn at.

N: \geq 0, the N places to the right of the decimal point will be drawn; N=0, the integral portion will be drawn with a decimal point; N=-1, no decimal point will be drawn and N-1 digits will be dropped from the number.

purpose: To supply additional labeling above and beyond that supplied by AXIS.

example:

CALL NUMBER (5.0, 10.0, 14, SUM, 0, 0, 0)

For example if SUM = 17.395, then

N = 3 WILL GENERATE 17.395

N = 0 WILL GENERATE 17.

N = -1 WILL GENERATE 17

5. Subroutine SYMBOL

Calling Sequence:

CALL SYMBOL (XA, YA, HEIGHT,EBCDCIC,THETA,

where:

XA and YA: coordinates of lower left corner of first symbol to be plotted, except in the case of a negative NS, where they are the coordinates of the center of the symbol.

HEIGHT: same as in NUMBER.

EBCDCIC: a Hollerith string or an address of a Hollerith string to be plotted or a number specifying a centering symbol depending on the value of NS.

THETA: same as in NUMBER.

NS: If NS > 0, NS is the number of EBCDCIC characters to be plotted from the EBCDCIC string specified; if NS < 0, SYMBOL will expect EBCDCIC an integer specifying a centering symbol (see symbol list). If NS < -1, the pen will be raised before moving to the new location; if NS = -9, the pen will be lowered.

purpose: To provide labeling of alphanumeric information on the plotter.
CALL SYMBOL (1.0, 8.0, 0.25, 5) LABEL, 0.0, 5
This will plot LABEL with the left-hand corner of
the L starting at (1.0, 8.0) relative to the origin.

CALL SYMBOL (1.0, 7.0, 0.14, 4, 0.0, 1)
This will draw an X centered at (1.0, 7.0) relative
to the origin.

6. Subroutine PLOT

This subroutine has the entry point PLOT. Its primary purpose is the trans-
ation of data into plotter format and the writing of the necessary plot
file. It also monitors the pen position and will terminate the computer
run if the pen goes out of the range of these limits. The limits are:

+X = XL inches; see LIMIT for XL

- X = 4 inches

+Y = 10.46 inches

- Y = .54 inch

Entry PLOT

Calling Sequence:

CALL PLOT (X, Y, IND)

where:

X = x-coordinate in inches of new pen position
relative to origin

Y = y-coordinate in inches of new pen position
relative to origin

IND = 0 or ±1 no change in vertical position of
pen
± 2 put pen down
± 3 put pen up
999 dump the current buffer

If the value of IND is negative, after the pen has
moved to the new position, zeros are stored as its
old position. This effectively establishes the new
position as a new reference point (origin) for sub-
sequent plotting. If IND is not equal to any of the
above, it is automatically set equal to 3.

purpose:

To move the pen from current position to a new po-
sition. It is the basic entry for plotting.

example:

CALL PLOT (10.0, 10.0, 2) This command will
draw a line from where the pen currently is to the
coordinates (10.0, 10.0) with the pen down.

EXAMPLES

To see how dendrographs are produced and
how to interpret output from the program, let us
consider two examples. The first example is based
on data reported by Valentine (1966) in a study of
latitude changes in the molluscan shelf fauna of
western North America between Baja California,
and Point Barrow, Alaska. The data on which the
study was based consisted of faunal lists compiled
over many years for shelved benthonic Mollusca of
the classes Pelecypoda and Gastropoda recorded
from depths less than 183 m. Range data were com-
piled for some 2077 forms. The purpose of the
study was to define major marine provinces based
on the fauna and to explain the observed latitude
changes. Valentine employed numerical methods
to analyze the data. For each degree of latitude,
a faunal list was prepared. The latitude ranged
from 23° N to 72° N, and for each pair, a measure
of faunal resemblance was calculated. The measure
used was Jaccard's coefficient defined for each pair
of latitudes as the number of common species divided
by the total number of different species reported for
the two latitudes. The matrix formed from the set of
pairwise latitude similarity coefficients constitutes
the input to the dendrogram program. The input
is shown in Table 1 which is the first part of the out-
put from the program. Note that the data are entered
as a lower triangular matrix. Because Jaccard's co-
efficient is a correlation-type measure, it is neces-
sary to transform the coefficients by taking the arc
cosine which transforms the coefficients into distance
measures. The transformed input is shown next. The
data now are ready for clustering. In this situation,
we are not interested in the intermediate stages of
clustering and thus the next output step gives the
Final Results of clustering (Table 1). It is from the
Final Results that we construct the dendrogram. The
results contain five columns which we can interpret
(starting from the left). The first column labelled
NAME contains the name given to each object. In
our example, the objects refer to latitudes, and they
are represented by numbers ranging from 72 to 23.
The numbers appear displaced to the right because
t here are sixteen spaces allowed for each object
name and the numbers have been right adjusted. The
next column labelled NP lists the objects by number
in the order they appear in the dendrogram. The
objects are numbered in sequence in the order they
appear in the input data. For this example, the
latitudes were entered starting from latitude 72N to
23N. Thus, corresponding to number 1 in column
NP is latitude 72 in column NAME, corresponding
to 2 is 71, 3 is 70, etc., to 50 which is latitude 23.
The objects are arranged in this order in the dendo-
graph. The next column labelled ORDER gives the
order in which the objects are joined, in this instance,
latitudes, are latitudes 47 and 46. The next two are
latitudes 68 and 69. The next combines latitude 67
with the previously joined latitudes 68 and 69. This
is continued until all latitudes have been combined
into a single group. To construct the dendrogram,
the connecting lines are drawn to form groupings in
the order indicated in column ORDER. The last two
columns contain the distances involved in connecting
the objects to form the hierarchical groupings. The
first is the within-group distance and the second the
between-group distance. The latter states the dis-
tance apart between two adjacent objects. The
former gives the distance from the origin where two
adjacent groups are connected. Because all con-
nections are made between two objects and only
one connection is possible between objects, the
distances are given between those two objects for
which the connecting line is drawn. In this manner,
the dendrogram is constructed step-by-step in the
order that the groups are formed in the clustering
process. Figure 2 contains the dendrogram produced
on the plotter for the data described. The latitude
Table 1. - Results of cluster analysis for Pacific Coast molluscan data taken from Valentine (1966).

<table>
<thead>
<tr>
<th>LATITUDE</th>
<th>PACIFIC COAST MOLLUSCAN RANGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF GROUPS IS 50</td>
<td></td>
</tr>
</tbody>
</table>

**ORIGINAL INPUT MATRIX (LOWER HALF BY ROWS)**

1. 1.0000
2. 0.85CC 1.0000
3. 0.82CC 0.95CC 1.0000
4. 0.82CC 0.95CC 0.9600 1.0000
5. 0.81CC 0.95CC 0.96CC 0.9900 1.0000
6. 0.8000 0.93CC 0.94CC 0.9800 0.9700 1.0000
7. 0.75CC 0.87CC 0.88CC 0.9200 0.9100 0.9400 1.0000
8. 0.71CC 0.82CC 0.83CC 0.8600 0.85CC 0.8800 0.9200 1.0000
9. 0.64CC 0.72CC 0.74CC 0.77CC 0.7600 0.7900 0.8700 0.8900 1.0000
10. 0.6200 0.72CC 0.73CC 0.7500 0.7500 0.7700 0.8000 0.8700 0.9300 1.0000
11. 0.62CC 0.70CC 0.71CC 0.7400 0.7300 0.7500 0.7900 0.8600 0.9100 0.9700 1.0000
12. 0.57CC 0.64CC 0.65CC 0.67CC 0.67CC 0.6800 0.7100 0.7400 0.8400 0.9000 1.0000
13. 0.45CC 0.50CC 0.50CC 0.5200 0.5200 0.5300 0.5600 0.6000 0.6500 0.6900 0.7000 0.7600 1.0000
14. 0.40CC 0.46CC 0.46CC 0.4700 0.4700 0.4800 0.5100 0.5500 0.6000 0.6400 0.6500 0.7100 0.8500 1.0000
15. 0.35CC 0.39CC 0.40CC 0.40CC 0.40CC 0.4100 0.4400 0.4700 0.5100 0.54CC 0.5500 0.60CC 0.7200 0.8300 1.0000
16. 0.26CC 0.32CC 0.33CC 0.34CC 0.34CC 0.3600 0.3900 0.4300 0.4500 0.4500 0.5000 0.60CC 0.6800 0.7600 1.0000
17. 0.24CC 0.27CC 0.27CC 0.28CC 0.28CC 0.2900 0.3000 0.3200 0.3500 0.37CC 0.3800 0.4100 0.50CC 0.5700 0.62CC 0.7500 1.0000
18. 0.1700 0.19CC 0.20CC 0.2000 0.2000 0.2000 0.2200 0.2300 0.2500 0.26CC 0.2700 0.2900 0.3700 0.4100 0.4600 0.5600 1.0000
19. 0.1200 0.14CC 0.15CC 0.1500 0.1500 0.1500 0.1600 0.17CC 0.1800 0.1900 0.2000 0.22CC 0.2800 0.3200 0.3600 0.4500 1.0000
20. 0.12CC 0.14CC 0.14CC 0.14CC 0.14CC 0.14CC 0.1500 0.1600 0.17CC 0.18CC 0.18CC 0.21CC 0.27CC 0.3100 0.3500 0.4300 1.0000
21. 0.12CC 0.13CC 0.14CC 0.14CC 0.14CC 0.14CC 0.1500 0.1600 0.17CC 0.18CC 0.18CC 0.2000 0.26CC 0.3000 0.3400 0.4200 1.0000
22. 0.11CC 0.13CC 0.13CC 0.14CC 0.14CC 0.14CC 0.1500 0.1600 0.16CC 0.17CC 0.17CC 0.20CC 0.25CC 0.2900 0.3200 0.4100 1.0000
23. 0.11CC 0.13CC 0.13CC 0.13CC 0.13CC 0.14CC 0.1400 0.1500 0.1500 0.16CC 0.16CC 0.17CC 0.17CC 0.19CC 0.2500 0.2800 0.3200 0.4000 1.0000
24. 0.10CC 0.12CC 0.12CC 0.12CC 0.12CC 0.12CC 0.13CC 0.1300 0.1400 0.15CC 0.15CC 0.16CC 0.16CC 0.20CC 0.23CC 0.26CC 0.3000 0.3700 1.0000
25. 0.45CC 0.55CC 0.77CC 0.83CC 0.84CC 0.90CC 0.9100 1.0000
<table>
<thead>
<tr>
<th>NAME</th>
<th>NP</th>
<th>ORDER</th>
<th>WGR</th>
<th>BGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>26</td>
<td>1</td>
<td>0.1415</td>
<td>0.1415</td>
</tr>
<tr>
<td>46</td>
<td>27</td>
<td>9</td>
<td>0.2391</td>
<td>0.2731</td>
</tr>
<tr>
<td>45</td>
<td>28</td>
<td>8</td>
<td>0.2003</td>
<td>0.2003</td>
</tr>
<tr>
<td>44</td>
<td>25</td>
<td>12</td>
<td>0.2750</td>
<td>0.3290</td>
</tr>
<tr>
<td>43</td>
<td>30</td>
<td>26</td>
<td>0.4301</td>
<td>0.7403</td>
</tr>
<tr>
<td>48</td>
<td>25</td>
<td>32</td>
<td>0.5099</td>
<td>0.5907</td>
</tr>
<tr>
<td>28</td>
<td>35</td>
<td>21</td>
<td>0.3871</td>
<td>0.5333</td>
</tr>
<tr>
<td>25</td>
<td>34</td>
<td>6</td>
<td>0.2003</td>
<td>0.2003</td>
</tr>
<tr>
<td>40</td>
<td>33</td>
<td>13</td>
<td>0.2763</td>
<td>0.3143</td>
</tr>
<tr>
<td>41</td>
<td>32</td>
<td>7</td>
<td>0.2003</td>
<td>0.2003</td>
</tr>
<tr>
<td>42</td>
<td>31</td>
<td>38</td>
<td>0.6338</td>
<td>1.2535</td>
</tr>
<tr>
<td>24</td>
<td>29</td>
<td>44</td>
<td>0.7661</td>
<td>1.0336</td>
</tr>
<tr>
<td>27</td>
<td>36</td>
<td>29</td>
<td>0.4643</td>
<td>0.5545</td>
</tr>
<tr>
<td>35</td>
<td>38</td>
<td>15</td>
<td>0.2838</td>
<td>0.2838</td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>47</td>
<td>0.9197</td>
<td>1.1077</td>
</tr>
<tr>
<td>56</td>
<td>17</td>
<td>42</td>
<td>0.6902</td>
<td>1.0164</td>
</tr>
<tr>
<td>55</td>
<td>18</td>
<td>35</td>
<td>0.5815</td>
<td>0.8529</td>
</tr>
<tr>
<td>49</td>
<td>24</td>
<td>30</td>
<td>0.4729</td>
<td>0.5559</td>
</tr>
<tr>
<td>54</td>
<td>15</td>
<td>27</td>
<td>0.4314</td>
<td>0.5217</td>
</tr>
<tr>
<td>50</td>
<td>23</td>
<td>16</td>
<td>0.2838</td>
<td>0.2838</td>
</tr>
<tr>
<td>51</td>
<td>22</td>
<td>20</td>
<td>0.3711</td>
<td>0.4064</td>
</tr>
<tr>
<td>52</td>
<td>21</td>
<td>17</td>
<td>0.3176</td>
<td>0.3176</td>
</tr>
<tr>
<td>53</td>
<td>20</td>
<td>48</td>
<td>1.1305</td>
<td>1.3947</td>
</tr>
<tr>
<td>47</td>
<td>45</td>
<td>4</td>
<td>0.2003</td>
<td>0.2003</td>
</tr>
<tr>
<td>23</td>
<td>50</td>
<td>23</td>
<td>0.3965</td>
<td>0.4945</td>
</tr>
<tr>
<td>25</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORDER</th>
<th>WGR</th>
<th>BGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>0.5719</td>
<td>0.7474</td>
</tr>
<tr>
<td>47</td>
<td>0.6633</td>
<td>0.8005</td>
</tr>
<tr>
<td>40</td>
<td>0.8851</td>
<td>1.0465</td>
</tr>
<tr>
<td>34</td>
<td>0.7102</td>
<td>0.9495</td>
</tr>
<tr>
<td>45</td>
<td>0.5906</td>
<td>0.7121</td>
</tr>
<tr>
<td>32</td>
<td>0.5096</td>
<td>0.6156</td>
</tr>
<tr>
<td>36</td>
<td>0.4036</td>
<td>0.5093</td>
</tr>
<tr>
<td>31</td>
<td>0.2003</td>
<td>0.2003</td>
</tr>
<tr>
<td>42</td>
<td>1.2482</td>
<td>1.4719</td>
</tr>
<tr>
<td>49</td>
<td>0.5548</td>
<td>0.5548</td>
</tr>
<tr>
<td>53</td>
<td>0.6378</td>
<td>0.6793</td>
</tr>
<tr>
<td>45</td>
<td>0.7844</td>
<td>1.0001</td>
</tr>
<tr>
<td>50</td>
<td>0.6821</td>
<td>1.1818</td>
</tr>
<tr>
<td>57</td>
<td>0.2456</td>
<td>0.2456</td>
</tr>
<tr>
<td>59</td>
<td>0.3498</td>
<td>0.4019</td>
</tr>
<tr>
<td>62</td>
<td>0.4207</td>
<td>0.4915</td>
</tr>
<tr>
<td>61</td>
<td>0.3910</td>
<td>0.5315</td>
</tr>
<tr>
<td>67</td>
<td>0.3347</td>
<td>0.4378</td>
</tr>
<tr>
<td>71</td>
<td>0.2832</td>
<td>0.3323</td>
</tr>
<tr>
<td>68</td>
<td>0.2505</td>
<td>0.3053</td>
</tr>
<tr>
<td>67</td>
<td>0.1958</td>
<td>0.2230</td>
</tr>
<tr>
<td>68</td>
<td>0.1415</td>
<td>0.1415</td>
</tr>
</tbody>
</table>
Figure 2. - Dendrogram depicting hierarchical pattern of Pacific Coast molluscan ranges.
pattern that emerges is distinct and shows clearly how a system of marine provinces could be proposed. Two points should be made. The pyramid shape of the dendrogram is not by accident. As described in the original paper (McCammon, 1968b, p. 1665), a rule given for ordering the objects as they are joined into groups results in the pyramid structure. As seen in Figure 2, it greatly enhances the interpretive quality of the dendrogram. Because the input for this problem was a set of correlation-type coefficients, the within-group distance measures were transformed by taking the cosine, and the transformed values are shown plotted in Figure 2. This strategy has been employed to preserve the sense of the original data.

As a second example, some unpublished data from a pattern recognition study are considered. The example will serve to show how distance-type similarity measures are used to produce dendrograms. Table 2 is a listing of the card input for this example. The data represent interpoint distances between gravity centers of environments based on a random sample of 500 points taken from a map of the Mississippi River deltaic plain of southeastern Louisiana prepared by Kolb and others (1958). The environments represented are natural levee, point bar, marsh, swamp, beach, lacustrine, and bay-sound. The purpose of the sampling was to obtain statistical measures of spatial order in a deltaic complex. The matrix of interpoint distances between gravity centers was constructed from the sample. It will suffice here to consider the graphic portrayal of spatial ordering of environments using the dendrogram.

Table 3 lists the results of clustering the interpoint distances for the seven environments. Because distance-type measures are involved, no transformation of the coefficients is necessary. To show more of the working of the program, an intermediate stage of clustering is listed. This was done by assigning a value of 5 to the KLAM control variable which was interpreted as a request to have the intermediate stages of clustering printed out at every fifth stage of clustering. For this example, it amounted to the listing of one intermediate stage. The form of the intermediate clustering stage is similar to the final clustering with the exception that zeroes appear where ordering has not been determined. The intermediate groups are indexed by the lowest ranking element within each group. In the example, there remains group 1 and group 4 at the fifth stage of clustering; the former having 4 elements, the latter, 3 elements. This information is contained in the two lines below the intermediate stage of clustering. The final results given next form the basis for constructing the dendrogram which is shown in Figure 3. It reveals clearly the dichotomous nature of marine versus nonmarine environments and the transitional character of the lacustrine environment. The environments can be taken as representative elements for pattern recognition in deltaic-type deposits situated elsewhere in the geologic column.

Table 2. Card input for environment center of gravity data based on random sample of 500 points taken from map prepared by Kolb and others (1958).

<table>
<thead>
<tr>
<th>7</th>
<th>MISSISSIPPI DELTA REGION ENVIRONMENTAL GRAVITY CENTERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>7.</td>
</tr>
<tr>
<td>NATURAL LEVEE</td>
<td>POINT BAR</td>
</tr>
<tr>
<td>LACUSTRINE</td>
<td>BAY-SOUND</td>
</tr>
</tbody>
</table>

(7F10.1)

<table>
<thead>
<tr>
<th></th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.4</td>
<td>0.0</td>
</tr>
<tr>
<td>9.8</td>
<td>12.5</td>
</tr>
<tr>
<td>27.8</td>
<td>36.8</td>
</tr>
<tr>
<td>55.2</td>
<td>64.2</td>
</tr>
<tr>
<td>2.4</td>
<td>27.8</td>
</tr>
<tr>
<td>42.8</td>
<td>53.9</td>
</tr>
</tbody>
</table>
Table 3. Results of cluster analysis of environment center of gravity data. Intermediate results given for fifth stage of clustering.

MISSISSIPPI DELTA REGION ENVIRONMENTAL GRAVITY CENTERS

NUMBER OF GROUPS IS 7

ORIGINAL INPUT MATRIX (LCWER HALF BY ROWS)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5.4000</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>9.8000</td>
<td>12.5000</td>
</tr>
<tr>
<td>4</td>
<td>27.8000</td>
<td>36.8000</td>
</tr>
<tr>
<td>5</td>
<td>55.2000</td>
<td>64.2000</td>
</tr>
<tr>
<td>6</td>
<td>22.4000</td>
<td>27.8000</td>
</tr>
<tr>
<td>7</td>
<td>42.8000</td>
<td>50.9000</td>
</tr>
</tbody>
</table>

CLUSTERING STAGE NO. 5

<table>
<thead>
<tr>
<th>NP</th>
<th>NGR</th>
<th>ORD</th>
<th>WGR</th>
<th>BGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5.4000</td>
<td>9.4000</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>10.5667</td>
<td>11.1500</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>16.2167</td>
<td>21.8667</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>17.6000</td>
<td>17.6000</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>5</td>
<td>21.0333</td>
<td>22.7500</td>
</tr>
</tbody>
</table>

FINAL RESULTS

<table>
<thead>
<tr>
<th>NAME</th>
<th>NP</th>
<th>ORDER</th>
<th>WGR</th>
<th>BGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATURAL LEVEE</td>
<td>1</td>
<td>1</td>
<td>9.4000</td>
<td>9.4000</td>
</tr>
<tr>
<td>POINT BAR</td>
<td>2</td>
<td>2</td>
<td>10.5667</td>
<td>11.1500</td>
</tr>
<tr>
<td>SWAMP</td>
<td>3</td>
<td>3</td>
<td>16.2167</td>
<td>21.8667</td>
</tr>
<tr>
<td>LACUSTRINE</td>
<td>6</td>
<td>6</td>
<td>30.9809</td>
<td>40.8500</td>
</tr>
<tr>
<td>BEACH</td>
<td>5</td>
<td>5</td>
<td>21.0333</td>
<td>22.7500</td>
</tr>
<tr>
<td>BAY-SOUND</td>
<td>7</td>
<td>4</td>
<td>17.6000</td>
<td>17.6000</td>
</tr>
<tr>
<td>MARSH</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. - Dendrograph depicting hierarchical pattern of spatial arrangement of environments in the Mississippi River deltaic plain of southeastern Louisiana.
REFERENCES


DENDROGRAPH COMPUTER PROGRAM

A COMPUTER PROGRAM TO ARRANGE A SET OF OBJECTS WHOSE
PAIRWISE SIMILARITY COEFFICIENTS ARE GIVEN INTO MUTUALLY EXCLUSIVE
HOMOGENEOUS SUBGROUPS AND TO DISPLAY THE RESULTS IN THE FORM OF A
DENDROGRAPH. THE METHOD USED FOR CLUSTERING IS THE UNWEIGHTED PAIR
GROUP METHOD. THE PRESENT PROGRAM IS DIMENSIONED TO CLUSTER UP TO
200 OBJECTS. THIS CAN BE EXTENDED BY INCREASING THE DIMENSIONS OF
THE AppROPRIATE PROGRAM VARIABLES. THE PRESENT PROGRAM USES THE
1627 CALCOMP PLOTTING VIA A SYSTEM/360--1800 CHANNEL ADAPTER TO DRAW
THE DENDROGRAPH. FEW MODIFICATIONS WOULD BE NECESSARY TO DRAW
DENDROGRAPHS ON A DIFFERENT PLOTTING SYSTEM.

ORDER OF INPUT CARDS
1. PROGRAM CONTROL AND TITLE CARD
2. PLOTTING CONTROL CARD
   ** OBJECT NAME ARRAY CARDS **
3. VARIABLE FORMAT INPUT DATA CARD
   ** DATA CARDS **

PROGRAM CONTROL VARIABLES

M - NUMBER OF OBJECTS TO BE CLUSTERED
   POSITIVE M DISTANCE MEASURE COEFFICIENTS
   NEGATIVE M CORRELATION MEASURE COEFFICIENTS

KLAM - FREQUENCY WITH WHICH INTERMEDIATE STAGES OF CLUSTERING
   ARE PRINTED
   DEFAULT OPTION - ONLY THE FINAL STAGE OF CLUSTERING IS
   PRINTED

PLOTTING CONTROL VARIABLES

AENG - DENDROGRAPH LENGTH IN INCHES (BETWEEN-GROUP SCALE AXIS ,
   AENG.LE.120.)

AH1 - DENDROGRAPH HEIGHT IN INCHES (WITHIN-GROUP SCALE AXIS ,
   AH1.LE.8.)

FNCR - UNITS THAT TIC MARKS ALONG THE WITHIN-GROUP SCALE AXIS
   ARE APART
   DEFAULT OPTIONS : +M FNCR=1.0
                   -M FNCR=0.1

PROGRAM VARIABLES

MATR - INPUT DATA VECTOR OF LENGTH M*(M+1)/2
NAME - OBJECT NAME MATRIX WITH DIMENSION 4 BY M
WIGR, BEGR, NP, ICRRD, NGR, NN, NCM, NAR, ALIN - WORKING VECTORS OF
   LENGTH M
SUBROUTINES SUPPLIED ALONG WITH MAIN PROGRAM

IPOS INTEGER CCOMPARE
SEARCH ELEMENT SEARCH
DENDRO DENDROGRAPH PLCT

SUBROUTINES TO BE SUPPLIED BY USER

LIMIT SET MAXIMUM LENGTH
BGNPLT BEGIN PLCT
ENDPLT END PLCT
SYMBOL SYMBOL PLCT
NUMBER NUMBER PLCT
PLCT LOCATION PLOT WITH PEN CONTROL

**************
MAIN PROGRAM
**************

REAL*4 MATR, NAME
INTEGER*4 SEARCH
DIMENSION MATR(20100), NGR(200), NN(200), NOM(200), NG(200), LGR(200),
NCPR(200), P(200), IORD(200), NAME(4, 200), ALIN(200), A(18), FMT(18)

READ(5, 10, END=100) M, KLA, (A(I), I=1, 18)
MM=M
MT=M
M=IBS(M)
READ(5, 40) AENG, AHI, FNCR
READ(5, 30) ((NAME(I, J), I=1, 4, J=1, M)
READ(5, 20) (FMT(J), J=1, 18)
WRITE(6, 11) (A(I), I=1, 18), M
DO 15 J=1, M
JJ=(J*(J-1))/2
NG(J)=J
NCPR(J)=J
NCPR(J)=J
NAME(J)=J
IORD(J)=J

LOWER HALF OF MATRIX IS READ ROW WISE FROM LEFT TO RIGHT

DO 15 I=1, M
JJ=(I*(I-1))/2
WRITE(6, 21) I, (MATR(JJ+J), J=1, I)
107 IF(MM.GT.0) GO TO 107
MM=(M*(M+1))/2
DO 105 I=1, MM
IF(MATR(I).GT.1.) MATR(I)=1.
105 MATR(I)=ARCS(MATR(I))
WRITE(6, 101)
GO TO 102
107 MH=M+1
ML=M-1
*************** MAIN LOOP FOR CALCULATING CLUSTERS ***************

DO 95 MAIN=1,ML
LCT=MH-MAIN
LCM=LCT-1
AMIN=100000.

MATRIX IS SEARCHED FOR SMALLEST WITHIN GROUP MEASURE

DO 35 I=1,LCM
K=I+1
IMA1=NN(I)+(NN(I)*(NN(I)-1))/2
DO 35 J=K,LCT
IT=(NN(J)+(NN(J)-1))/2
IMA2=NN(I)+IT
IMA3=NN(J)+IT
LA=NOM(I)+NOM(J)
ANOM=(LA*(LA-1))/2
TEMP=(MATR(IMA1)*MATR(IMA2)+MATR(IMA3))/ANOM
IF(TEMP.GT.AMIN) GO TO 35
ITI=I
ITJ=J
AMIN=TEMP
35 CONTINUE
LA=NOM(ITI)+NOM(ITJ)

BEGINNING AND ENDS OF GROUPS TO BE JOINED ARE FOUND BY SEARCH

JKL=NN(ITI)
IBEG1=SEARCH(NGR,JKL)
JKL=NN(ITJ)
IBEG2=SEARCH(NGR,JKL)
IEND1=IBEG1+NOM(ITI)-1
IEND2=IBEG2+NOM(ITJ)-1
IT=(NN(ITJ)+(NN(ITJ)-1))/2
IMA1=NN(ITI)+(NN(ITI)*(NN(ITI)-1))/2
IMA2=NN(ITI)+IT
IMA3=NN(ITJ)+IT
JKL=NOM(ITI)+NOM(ITJ)
BEGR(IBEG2)=MATR(IMA2)/FLCAT(JKL)
WGR(IBEG2)=AMIN
IORD(IBEG2)=MAIN
TEMP=MATR(IMA1)+MATR(IMA2)+MATR(IMA3)

OUTPUT ARRAYS ARE SHIFTED INTO THE NEW ORDER
THE RIGHTMOST GROUP IS SHIFTED LEFT

LDIF=IBEG2-IEND1-1
KLIM=NOM(ITJ)
IF(LDIF) 77,17,18
18 DO 45 I=1,LDIF
DO 45 J=1,LDIF
KS=IBEG2+I-J
MST=NGR(KS)
NGR(KS)=NGR(KS-1)
NGR(KS-1)=MST
MST=NPI(KS)
NPI(KS)=NPI(KS-1)
NPI(KS-1)=MST
45 CONTINUE
MST=IORD(KS)
IORD(KS)=ICRD(KS-1)
IORD(KS-1)=MST
GW=WIGR(KS)
WIGR(KS)=WIGR(KS-1)
WIGR(KS-1)=GW
GW=BEGR(KS)
BEGR(KS)=BEGR(KS-1)

45 BEGR(KS-1)=GW
17 IF(NOM(ITI)-NOM(ITJ)) 131,141,141
141 KNO2=NOM(ITJ)
KS=IEND1+1
GO TO 78
131 KNO1=NOM(ITI)
KNO2=NOM(ITJ)
DO 125 I=1,KNO1
DO 125 J=1,KNO2
KS=IEND1+J-I
MST=NP(KS+1)
NP(KS+1)=NP(KS)
NP(KS)=MST
MST=IORD(KS+1)
IORD(KS+1)=IORD(KS)
IORD(KS)=MST
GW=WIGR(KS+1)
WIGR(KS+1)=WIGR(KS)
WIGR(KS)=GW
GW=BEGR(KS+1)
BEGR(KS+1)=BEGR(KS)
125 BEGR(KS)=GW
KS=IBEG1+KNO2
IORD(KS)=ICRD(IBEG1)
WIGR(KS)=WIGR(IBEG1)
BEGR(KS)=BEGR(IBEG1)
IORD(IBEG1)=I
WIGR(IBEG1)=0.0
BEGR(IBEG1)=0.0
KNO2=NOM(ITI)

C     IF THE NEW RIGHT GROUP HAS MORE THAN TWO ELEMENTS
C     ITS ORDER IS REVERSED
C
78 IF(KNO2.LE.2) GO TO 111
KNO1=KNO2/2
KS=KS-1
IEND2=IBEG1+NOM(ITI)+NOM(ITJ)
DO 135 I=1,KNO1
MST=NP(IEND2-I)
NP(IEND2-I)=NP(KS+I)
NP(KS+I)=MST
MST=IORD(IEND2-I)
IORD(IEND2-I)=ICRD(KS+I+1)
IORD(KS+I+1)=MST
GW=BEGR(IEND2-I)
BEGR(IEND2-I)=BEGR(KS+I+1)
BEGR(KS+I+1)=GW
GW=WIGR(IEND2-I)
WIGR(IEND2-I)=WIGR(KS+I+1)
WIGR(KS+I+1)=GW
135 CONTINUE
LATEST GROUP IS PARTIALLY RELAbeLED IN 'NGR'

DO 55 I=1,KLIM
55 NGR(IEND1+I)=NN(ITI)

ROW AND COLUMN NN(ITJ) ARE ADDED ON TO ROW AND COLUMN NN(ITI)

DO 75 I=1,M
   KLA=IPOS(NN(ITI),I)
   KLE=IPOS(NN(ITJ),I)
75 MATR(KLA)=MATR(KLA)+MATR(KLE)
   MATR(IMA1)=TEMP
   LDIF=LCT-ITJ
   IF(LDIF) 77,27,28
28 CONTINUE

DO 85 I=1,LDIF
   NN(ITJ+I-1)=NN(ITJ+I)
85 CONTINUE

NOM(ITJ+I-1)=NOM(ITJ+I)

NOM(ITI)=LA
   IF(KLAM.EQ.0) GO TO 95
   KS=MAIN-(MAIN/KLAM)*KLA
   IF(KS.NE.0) GO TO 95
   WRITE(6,31) MAIN
   WRITE(6,61) NP(I),NGR(I)
   DO 65 J=2,M
55 WRITE(6,71) IORD(J),WIGR(J),BEGR(J)
   WRITE(6,61) NP(J),NGR(J)
   WRITE(6,41)(NN(I),I=1,LCM)
   WRITE(6,41)(NOM(J),J=1,LCM)
95 CONTINUE

WRITE(6,121)
KS=NP(I)
WRITE(6,81) NAME(L,KS),L=1,4,KS
DO 205 I=2,M
   WRITE(6,181) IORD(I),WIGR(I),BEGR(I)
   KS=NP(I)
205 WRITE(6,81) NAME(L,KS),L=1,4,KS
   WRITE(6,61)
   CALL DENDRO(WIGR,BEGR,NS,IORD,NAME,ALIN,M,FNCR,AENG,AHI,MT,A)
   GO TO 999
77 WRITE(6,51) LDIF
   GO TO 999

10 FORMAT('1',I1)
10 FORMAT(2I4,18A4)
11 FORMAT('1',30X,18A4,/' NUMBER OF GROUPS IS ',I3)
20 FORMAT(18A4)
21 FORMAT('0',I3,(1X,16F8.4))
30 FORMAT(20A4)
31 FORMAT('1 CLUSTERING STAGE NO. ',I3,/' NP',5X,'NGR ORD',4X,'WGR 1 3GR')
40 FORMAT(2F10.3,F5.2)
41 FORMAT('0',3314)
51 FORMAT('0',I4)
61 FORMAT(' ',I3,5X,I3)
71 FORMAT(' ',12X,I3,2F8.4)
81 FORMAT(' ',4A4,I4)
91 FORMAT('0',20X,'ORIGINAL INPUT MATRIX (LOWER HALF BY ROWS)')
101 FORMAT('1 TRANSFORMED INPUT MATRIX')
121 FORMAT('1',30X,'FINAL RESULTS/// NAME',14X,'NP',5X,'ORDER',5X,
1 'WGR',5X,'BGR')
181 FORMAT(' ',27X,I3,2X,2F8.4)
END

C
****
INTEGER FUNCTION IPCS(IA,IB)
****
C
IF(IA.GT.IB) GO TO 1
IPOS=IA+(IB*(IB-1))/2
GO TO 2
1 IPOS=IB+(IA*(IA-1))/2
2 RETURN
END

C
****
INTEGER FUNCTION SEARCH(NAR,NINT)
****
C
DIMENSION NAR(1)
DO 5 J=1,200
   IF(NINT.EQ.NAR(J)) GC TO 10
5 CONTINUE
   J=-1
10 SEARCH=J
RETURN
END
SUBROUTINE DENDRO(WIGR,BEGR,NP,ICRD,NAME,ALIN,M,FNCR,AENG,AHI,MT,1A)

REAL NAME
INTEGER SEARCH
DIMENSION WIGR(1),BEGR(1),NP(1),ICRD(1),NAME(4,1),ALIN(1),A(1)
MM=M-1
ICRD(1)=M+1
ICRD(M+1)=M+1
AENG=AENG-1.

SCALE X-AXIS IN INCHES

DO 15 J=3,M
15 BEGR(J)=BEGR(J)+BEGR(J-1)
SCAX=AENG/BEGR(M)
UPI=BEGR(M)/AENG
DO 25 J=1,M
25 BEGR(J)=BEGR(J)+UPI/2.0
TOTX=BEGR(M)+UPI/2.0
DO 35 J=1,M
35 BEGR(J)=BEGR(J)*SCAX

SCALE Y-AXIS IN INCHES

IH=SEARCH(ICRD,MM)
TOTY=WIGR(IH)
IF(MT.LT.0) TOTY=1.0-COS(WIGR(IH))
SCAY=AHI/TOTY
DO 45 I=1,M
IF(MT.LT.0) GO TO 44
WIGR(I)=WIGR(I)*SCAY
GO TO 45
44 WIGR(I)=(1.0-COS(WIGR(I)))*SCAY
45 CONTINUE
CALL BGNPLT(8,'MCCAMMCN')

SET X-AXIS LIMIT

XLEN=48.
CALL LIMIT(XLEN)

SET ORIGIN

XORG=0.
YORG=2.

PRINT TITLE

X=XORG-2.
Y=.5
CALL SYMBOL(X,Y,.125,A(1),90.,72)

DRAW IN BOUNDARY LINES

CALL PLCT(XORG,YORG,3)
X=XORG+AENG+1.
CALL PLOT(X,YORG,2)
   Y=YORG+AH1+.4
CALL PLOT(X,Y,2)
CALL PLOT(XORG,Y,2)
CALL PLCT(XORG,YORG,2)

C   PEN NOW LOCATED AT ORIGIN
C
C   PLACE TIC MARKS ALONG X-AXIS
C
   IAX=TOTX
   AX=XORG
   DO 145 I=1,IAX
      AX=AX+SCAX
      Y1=YORG+.06
      Y2=YORG-.07
      CALL PLOT(AX,Y1,3)
   145 CALL PLOT(AX,Y2,2)
C
C   LABEL AND PLACE TIC MARKS ALONG Y-AXIS
C
   IF(MT.LT.0) GO TO 36
   IF(FNCR.EQ.0.) FNCR=1.
   NTCT=TOTY/FNCR
   FNMB=0.
   FDIS=FNCR*SCAY
   GO TO 37
36 IF(FNCR.EQ.0.) FNCR=.1
   NTOT=TOTY/FNCR
   FDIS=FNCR*SCAY
   FNMB=1.
   FNCR=-FNCR
37 X=XORG-.425
   Y=YORG-.125
   CALL NUMBER(X,Y,.125,FNMB,0.,1)
   AI=FNMB
   AY=YORG
   IF(NTOT.EQ.0) GO TO 27
   DO 85 I=1,NTCT
      AI=AI+FNCR
      AY=AY+FDIS
      X1=XORG+.125
      X2=XORG-.125
      CALL PLOT(X1,AY,3)
      CALL PLOT(X2,AY,2)
      AYY=AY-.125
      X=XORG-.425
   85 CALL NUMBER(X,AYY,.125,AI,0.,1)
C
C   PRINT OBJECT NAMES
C
   27 DO 55 I=1,M
      AX=XORG+BEGR(I)
      IL=NPI(I)
      Y=YORG-.254
   55 CALL SYMBOL(AX,Y,.125,NAME(I,IL),90.,16)
   DO 65 I=1,M
   65 ALIN(I)=0.0
GENERATE DENDROGRAPH

DO 75 I=1, MM
IH=SEARCH(IORD, I)
BAR=YORG+ALIN(IH)
BAL=YORG+ALIN(IH-1)
TOP=YORG+WGR(IH)
SR=XORG+BEGR(IH)
SL=XORG+BEGR(IH-1)
CALL PLOT(SR, BAR, 3)
CALL PLOT(SR, TOP, 2)
CALL PLOT(SL, TOP, 1)
CALL PLOT(SL, BAL, 1)
K=IORD(IH)
J=IH
67 J=J+1
IF(K-IORD(J)) 68, 67, 67
68 IEND=J-1
J=IH
77 J=J-1
IF(K-IORD(J)) 78, 77, 77
78 IBEG=J
ALIN(IBEG)=WGR(IH)
ALIN(IEND)=WGR(IH)
CENT=(BEGR(IH)+BEGR(IH-1))/2.0
BEGR(IBEG)=CENT
BEGR(IEND)=CENT
CALL ENDPLT
RETURN
END
CHARACTERS AVAILABLE IN SYMBOL ROUTINE (IBM 360)*

* Please note that all identifying numbers are base 16
PROGRAM ABSTRACT

Title (If subroutine state in title): THE DENDROGRAPH

Date: February 1970

Author, organization: Richard B. McCammon, Department of Geological Sciences, University of Illinois at Chicago, Chicago, Illinois 60680

Direct inquiries to: same as above

Name: Address:

Purpose/description: To produce a two dimensional diagram called the dendrograph in which is displayed the mutual relationships among a group of objects whose pairwise similarities are given.

Mathematical method: Unweighted pair group method of clustering (UWPGM)

Restrictions, range: The program is currently dimensioned to cluster up to 200 objects. Plotting subroutines must be supplied by the user.

Computer manufacturer: IBM Model: 360/50

Programming language: FORTRAN IV

Memory required: 27 K Approximate running time: 2 - 5 minutes

Special peripheral equipment required: Plotter unit

Remarks (special compilers or operating systems, required word lengths, number of successful runs, other machine versions, additional information useful for operation or modification of program)
16. FORTRAN IV program for the GE 625 to compute the power spectrum of geological surfaces, by J.E. Esler and F.W. Preston, 1967  
17. FORTRAN IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers, by G.F. Bonham-Carter, 1967  
19. FORTRAN II time-trend package for the IBM 1620 computer, by J.C. Davis and R.J. Sampson, 1967  
21. FORTRAN IV program for computation and display of principal components, by W.J. Wahlstedt and J.C. Davis, 1968  
25. FORTRAN IV CDC 6400 computer program to analyze subsurface fold geometry, by E.H.T. Whitten, 1968  
26. FORTRAN IV computer program for simulation of transgression and regression with continuous-time Markov models, by W.C. Krumbein, 1968  
28. KWIK88 a FORTRAN IV program for multiple regression and geologic trend analysis, by J.E. Esler, P.F. Smith, and J.C. Davis, 1968  
29. FORTRAN IV program for harmonic trend analysis using double Fourier series and regularly gridded data for the GE 625 computer, by J.W. Harbaugh and M.J. Sackin, 1968  
30. Sampling a geological population, by J.C. Griffiths and C.W. Ondrick, 1968  
31. Multivariate procedures and FORTRAN IV program for evaluation and improvement of classifications, by Ferruh Demirsen, 1969  
32. FORTRAN IV programs for canonical correlation and canonical trend-surface analysis, by P.J. Lee, 1969  
33. FORTRAN IV program for construction of Pi diagrams with the Univac 1108 computer, by Jeffrey Warner, 1969  
34. FORTRAN IV program for nonlinear estimation, by R.B. McCammon, 1969  
35. FORTRAN IV computer program for fitting observed count data to discrete distribution models of binomial, Poisson and negative binomial, by C.W. Ondrick and J.C. Griffiths, 1969  
36. GRAFPAK, graphic output subroutines for the GE 635 computer, by F.J. Rohlf, 1969  
38. FORTRAN II programs for 8 methods of cluster analysis (CLUSTAN I), by David Wishart, 1969  
39. FORTRAN IV program for the generalized statistical distance and analysis of covariance matrices for the CDC 3600 computer, by R.A. Reymert, Hans-Ake Ramden, and W.J. Wahlstedt, 1969  
40. Symposium on computer applications in petroleum exploration, edited by D.F. Merriam, 1969  
41. FORTRAN IV program for sample normality tests, by D.A. Preston, 1970  
42. CORFAN—FORTRAN IV computer program for correlation, factor analysis (R- and Q-mode) and varimax rotation, by C.W. Ondrick and G.S. Srivastava, 1970  
43. Minimum entropy criterion for analytic rotation, by R.B. McCammon, 1970  
44. FORTRAN IV CDC 6400 computer program for constructing isometric diagrams, by W.B. Wray, Jr., 1970  
45. An APL language computer program for use in electron microprobe analysis, by D.G.W. Smith and M.C. Tomlinson, 1970  
46. FORTRAN IV program for Q-mode cluster analysis on distance function with printed dendrogram, by J.M. Parks, 1970  
47. FORTRAN IV program for canonical variates analysis for the CDC 3600 computer, by R.A. Reymert and H. Ramden, 1970  