The New Characteristic Analysis (NCHARAN) Program
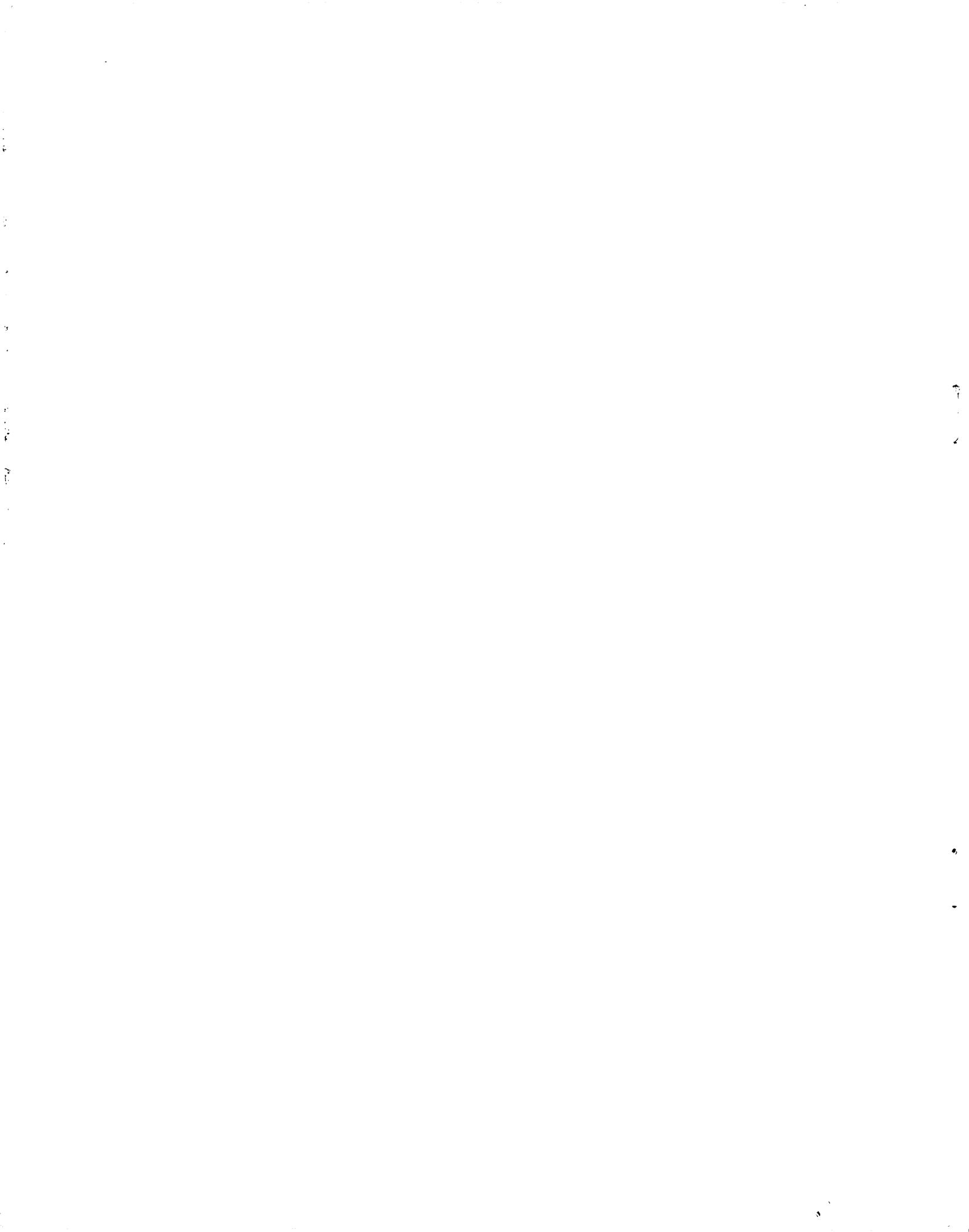
U.S. GEOLOGICAL SURVEY BULLETIN 1621

The New Characteristic Analysis (NCHARAN) Program

By Richard B. McCammon, Joseph Moses Botbol,
Richard Sinding-Larsen, and Roger W. Bowen

*A computer program for integrating geologic,
geochemical, and geophysical data with respect to
mineral deposit models*

GEOLOGICAL SURVEY BULLETIN 1621

DEPARTMENT OF THE INTERIOR
WILLIAM P. CLARK, Secretary

U.S. GEOLOGICAL SURVEY
Dallas L. Peck, Director

CONTENTS

FIGURES

TABLES

The New Characteristic Analysis (NCHARAN) Program

By Richard B. McCammon, Joseph Moses Botbol, Richard Sinding-Larsen,[1]
and Roger W. Bowen

Abstract

The new characteristic analysis (NCHARAN) program is
a method for integrating geologic, geochemical, and geophys-
ical data in mineral exploration and regional resource ap-
praisal. Data are integrated by relating measurements, obser-
vations, and inferences to deposit models. The NCHARAN
program allows the user to interactively select region cells for
defining deposit models and variables or logical combinations
of variables that best represent these models.

INTRODUCTION

Characteristic analysis was first proposed as a
method for classifying base-metal mining districts on the
basis of the presence of characteristic mineralogic,
lithologic, and structural associations (Botbol, 1968,
1969, 1971). The idea was to represent the attributes pres-
ent in each district as a binary occurrence model in which
the number "1" denoted the presence of the attribute
and the number "0" denoted that the attribute had not
been recognized. By combining districts, it was possible
to represent the contributions of the various attributes.
The attributes of such a model could be compared to
those of other districts, thereby making it possible to
classify new districts. From these applications came the
recognition that characteristic analysis offered a multi-
variate technique for combining geologic, geochemical,
and geophysical data for target selection in exploration.
Since then, it has been used to identify exploration targets
for a wide variety of deposit types (Botbol and others,
1978; Sinding-Larsen and others, 1979; McCammon and
others, 1979; Sinding-Larsen and Strand, 1981; McCam-
mon and others, 1983a,b). More recently, characteristic
analysis has been used in uranium resource assessment
(McCammon and others, in press).

The purpose of this paper is to describe the new or
latest version of the characteristic analysis (NCHARAN)
computer program. The current version includes options
for the selection of region cells for defining deposit
models and the selection of variables and combinations
of variables that constitute the models. Moreover, the
program provides for the display of results and thus
allows review, reselection, and refinement of a model.
Most important, the performance of the above-
mentioned operations in an interactive computing mode
can result in a timely and meaningful interpretation of
the data available to the geologist.

The current version of the program is written in
FORTRAN IV and has been implemented on the Honey-
well MULTICS system.[2] The Tektronix TCS (terminal
control software) (Tektronix, Inc., 1977) package pro-
vides graphics support. The dependency on operating
system and graphics support are constraints particular to
the U.S. Geological Survey (USGS) implementation of
the code. Users who wish to use the program on other
systems must first disaggregate the code and rewrite those
parts of the program that are system or software depend-
ent. The system dependencies are described in more detail
in the Appendix. Operational flow charts are given which
provide the information necessary for recoding. If
recoding is necessary, we recommend that the code for
mathematical algorithms be retained and used as needed
in the new procedures. A magnetic tape of the source
program is available for a nominal fee from the authors.

DATA REPRESENTATION

In characteristic analysis, variables take on the
values $+1$, -1, and 0. These values represent the possi-
ble states in which a given measurement, observation, or
inference about a variable is classified with respect to a
particular deposit model. The values $+1$, -1, and 0 refer
to favorable, unfavorable, and unevaluated states,
respectively. For deposit models defined by the presence
or absence of attributes, it is usual to assign a $+1$ value
to presence and a -1 value to absence. Where presence
or absence cannot be determined, a 0 value is assigned.
For most models it is necessary to transform ranges of
values of variables into ternary form, that is, to construct
a transformation that classifies a given range into one of
the three possible states. The manner in which this trans-
formation is performed, however, is not necessarily part
of characteristic analysis.

The ternary transformation performed on a vari-
able depends upon the nature of the model and the nature
of the data. For models based on geochemical data,
favorable $(+1)$ states are indicated by local anomalies,

---

[1]Professor Sinding-Larsen is at the Geologisk Institutt, Univ. i Trond-
heim, Trondheim, Norway.

[2]The use of trade names is for identification only and does not imply
endorsement by the U.S. Geological Survey.

which can be found by taking second derivatives (Botbol and others, 1978, p. 535) as shown in figure 1. In this instance, values are classified as favorable or unfavorable depending upon their position above or below an inflection point. For data in two-dimensional space, the magnitude of measurement is classified according to the position above or below the loci of inflection points defined by a mathematical surface representing the spatial distribution of the measured values.

For deposit models based on geophysical data, favorable (+1) states are assigned on the basis of regional gradients, local increases or decreases in measured values, and special features seen on regional contour maps. Figure 2 shows an example in which the measured values of magnetic intensity are classified according to geometric shapes of mathematical surfaces represented by a set of values. In this case, the isogams are examined to see whether a valley head or side slope, a shoulder, a peak on a plateau or ridge, or a depression is present in the contour pattern.

Examples of ternary transformations performed on data for sandstone uranium deposit models are given in table 1. For each variable, the transformation applied is based on a range of values, or a set of states, that best distinguishes areas with deposits from areas without deposits. To a large extent, the success or failure of characteristic analysis for a given set of data rests on the ability of the interpreter to determine what constitutes favorability for each variable with respect to a model.

To facilitate handling many variables for many cells, the transformed values of variables are stored in the computer in a binary format in a file called a "matarray". Currently, the program provides for up to 85 variables for up to 6,000 cells. The data in the matarray are stored in a compressed mode (Bowen, 1977) in which 17 ternary-transformed values are contained in each 36-bit word. The matarray is a matrix of rows and columns, each row of the matrix representing the value of the variables in each cell. Each cell is represented by the ternary-transformed values of the variables at a specified location. Blanks are used for missing data.

To assist in the preparation of data used in characteristic analysis, a preprocessor (PREPRO) computer program is available (Bridges and others, in press) that enables the geologist to transform the data into ternary form. PREPRO's options include variable-input formats, cursor input, ordering among a set of input variables, and selectable ternary transformations. PREPRO has made it possible for the geologist to preprocess the data rapidly and efficiently before characteristic analysis.

GENERAL DESCRIPTION OF METHOD

Characteristic analysis is best described as a method for integrating exploration data with respect to deposit models. A deposit model is defined as a weighted linear combination of ternary-transformed variables. In most instances, the variables are evaluated at locations specified on a grid. Hence, it is convenient to identify each value of a variable by a number that refers to a cell within the grid. Grids usually are rectangular arrays of square cells.

As indicated earlier, a ternary-transformed variable takes on one of three values, $+1$, -1, and 0, with respect to a particular deposit model. The basic assumption in
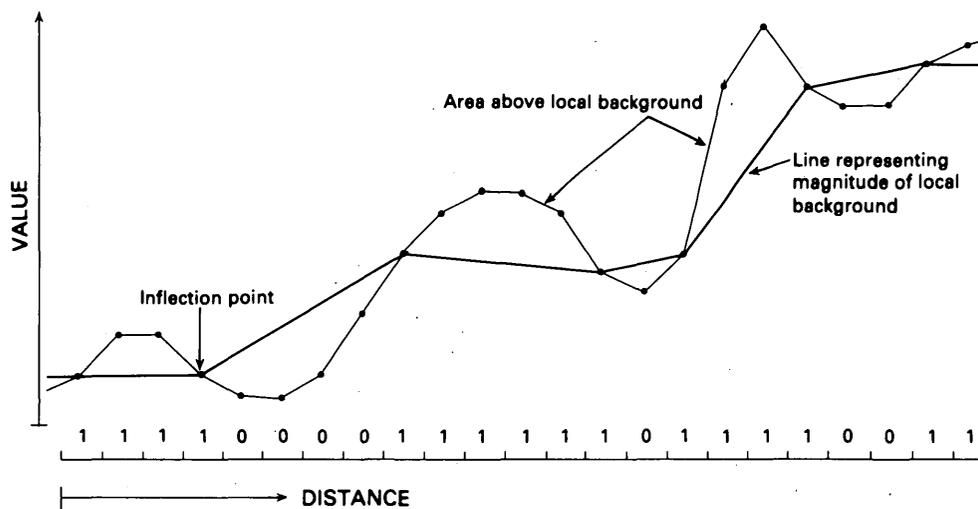

Figure 1. Hypothetical geochemical profile; areas above local inflection points (negative second derivative), labeled 1, are considered favorable, and other locations, labeled 0, are considered unevaluated or unfavorable.
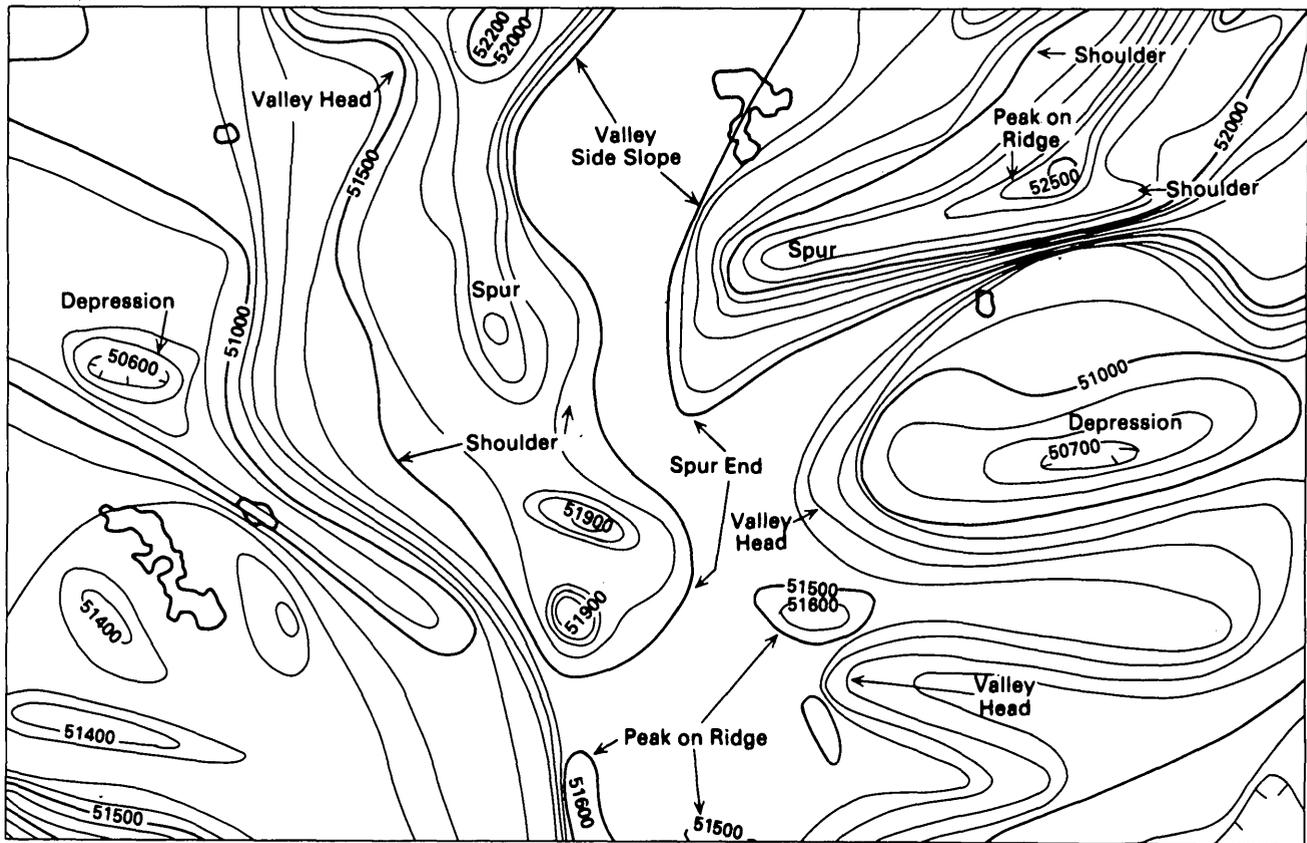
Figure 2. Magnetic contour map indicating regional and local morphological features. Valley side slope is regarded as a regional magnetic gradient, and a local decrease can be regarded as a shoulder. Favorable states (+1) are coded if the presence of a morphological feature is judged favorable for the presence of mineralization. For example, a steep valley side slope indicates a possible lithological boundary and is therefore favorable for massive sulfides due to their presence at the boundaries of volcaniclastic units.

characteristic analysis is that cells containing deposits (discovered and undiscovered) of the type represented by a particular model have similar attributes when described by a set of appropriately selected variables.

On this basis, the favorability, f, of a given cell with respect to a model composed of n variables is defined by

$$f = a_1 x_1 + a_2 x_2 \ldots + a_n x_n,\qquad (1)$$

where a_i ($i = 1,\ldots,n$) represents weight and x_i ($i = 1,2,\ldots,n$) represents the variable. The determination of the weights is based on the observed values of the n variables in cells selected to represent the model.

Suppose that m cells are chosen to represent a deposit model based on n variables. This defines a $m \times n$ matrix, $\mathbf{X}$, that contains ternary-transformed values for the n variables. The weights, a_i, in equation 1 are determined by solving the matrix equation

$$(\mathbf{X'X})\, a = \lambda a \qquad (2)$$

where λ is the largest eigenvalue of the product matrix $\mathbf{X'X}$. $\mathbf{X'}$ is the transpose of $\mathbf{X}$. A more complete description of the mathematical development of equation 2 and the method of solution can be found in an earlier paper (McCammon and others, 1983b).

The weights, a_i, are scaled so that values of f in equation 1 lie between -1 and $+1$. High (positive) values, that is, values near $+1$, indicate a close match with a model and are interpreted as being highly favorable for occurrence of a deposit represented by the selected variables describing the model. Low (negative) values, that is, values near -1, indicate a mismatch with a model and are interpreted as being highly unfavorable for occurrence of a deposit represented by the model. Values around 0 indicate neither match nor a mismatch and so give no information about the occurrence of a deposit represented by the model.

Table 1. Example of ternary transformations performed on data that convert the data to a favorable (+1) state with respect to area-related uranium sandstone models. Data from McCammon and others (in press)

[dim, dimensionless; —, variable not used; ss, sandstone; ms, mudstone; altn, number of ss-ms alternations; nmaltn, altn per unit thickness; magthk, thickness of magnetite destruction interval]

| | | | | | Area-related uranium deposit model | | | | |
| | | | | | Trend | | | Remnant | Roll |
Geologic factor	Variable No.	Name	Data type	Units	1	2	3		
Favorable host-rock deposition.	1	thickness	real	feet	—	115–280	—	—	—
	2	net ss thickness	real	feet	200–280	—	≥ 130	160–280	≥ 200
	3	ss/ms	real	dim.	2.5–7.0	2.5–7.0	—	3.0–8.5	·
	4	altn	integer	dim.	—	2.0–7.0	—	—	—
	5	nmaltn	real	dim.	—	—	0.5–2.5	—	—
Favorable alteration preparation.	6	sscolor	character	dim.	gray	—	—	—	—
	7	mscolor	character	dim.	gray	—	—	—	—
	8	magthk	real	feet	>0.	>0.	>0.	>0.	—
Favorable preservation	9	oxidation	character	dim.	reduced	reduced	reduced	oxidized	transitional

COMBINATIONS OF VARIABLES

In many cases, a single variable is inadequate for describing a particular aspect of a model. This is particularly true for genetic-geologic models (Finch and others, 1980). In these models, combinations of variables are of interest. For example, in the genetic-geologic model for tabular humate uranium deposits, a favorable host-rock-sedimentary-source factor is indicated by evidence for both an epiclastic organic source and either an epiclastic granitic or a pyroclastic source (McCammon and others, in press).

Our experience has been that, in most cases, deposit models are best constructed by using logical combinations of variables, that is, AND, OR, and NOT combinations. Consequently, the x_i's in equation 1 have been extended to include logical combinations of ternary-transformed variables. The truth table (table 2) was devised to find such combinations. The truth table defines the values of all logical combinations of two ternary-transformed variables, consistent with the rules of logical operations. Complex logical expressions involving more than two variables can be evaluated by successive application of the rules for two variables.

MODEL GENERALIZATION

A problem that arises in model definition is cell selection. Generally, the selected cells are those that contain deposits of the type represented by the model.

Table 2. Truth table for ternary-valued variables A and B
[The numbers 1, −1, and 0 represent favorable, unfavorable, and unevaluated states, respectively, with respect to a given deposit model]

A	B	A OR B	A AND B	NOT A
1	1	1	1	−1
1	0	1	0	−1
1	−1	1	−1	−1
0	1	1	0	0
0	0	0	0	0
0	−1	0	−1	0
−1	1	1	−1	1
−1	0	0	−1	1
−1	−1	−1	−1	1

Because any single deposit is unique, the data used to define a model often are overly restrictive; data from other areas that may contain undiscovered deposits of the type represented by the model may not match the model perfectly. However, what is desired in regional exploration is a way to identify favorable geologic settings within which undiscovered deposits may occur. With this in mind, it is a matter of practical importance to consider how models of existing deposits can be expanded into generalized models of cells, most (not all) of which will contain deposits.

The following procedure describes how an existing deposit model can be generalized. Consider an existing deposit model, based on n variables, with k cells that contain deposits. The weights for such a model can be determined using equation 2. Suppose now that we apply

equation 1, using the weights derived from equation 2 to other cells around the cells that contain deposits and determine that l of these cells are equally or more favorable than the k cells in the model. We must now decide whether this result is due to chance or to some relation between those cells and the cells in the model. If a relation does exist, including the larger number of cells in the model can lead to identification of similar geologic settings outside the model area. The question can be answered by considering the matches among the n ternary-transformed variables for the $k+l$ cells.

For each of the n variables, there is an ordered array of $k + l$ 1's, 0's, and -1's, for example (1, 1, 0, -1,..., 1). For any two variables expressed as vectors $\mathbf{u}$ and $\mathbf{v}$, the number of non-zero matches, m, can be expressed as:

$$m = \mathbf{u}'\mathbf{v} \tag{3}$$

where $\mathbf{u}'$ is the transpose of $\mathbf{u}$. If it is assumed that the observed sequences of values in $\mathbf{u}$ and $\mathbf{v}$ could have occurred in any order, then the probability that the observed number of matches would not be exceeded by chance can be calculated (the formula for calculating this probability can be found in McCammon and others, 1983a). If this probability is high, it is concluded that the two variables are correlated. If the probability is low, it is concluded that the two variables are not correlated. For n variables, an $n \times n$ symmetric matrix, $\mathbf{P}$, can be formed that contains the n^2 pairwise probabilities. $\mathbf{P}$ can be substituted for $\mathbf{X}'\mathbf{X}$ in equation 2 and a different set of weights, a_j, calculated. Using these weights in equation 1, we can calculate the favorability, f, for the $k + l$ cells, and those p cells having the highest values are then selected as the cells that represent the generalized model. This procedure can lead to better recognition of similar geologic settings outside a model area, as shown by McCammon and others (1983a).

USER GUIDE

The NCHARAN program consists of nine basic operations. These are (1) database selection, (2) model-cell selection, (3) selection of variables, (4) construction of logic variables, (5) computation of weights, (6) computation of degree of association, (7) display of selected variables or degree of association, (8) saving the degree of association, and (9) quitting the program. Each operation consists of optional steps. Following any operation, the user can branch to any other operation. Following each step of any operation, the user can branch to any other step in that operation. A question mark (?) or the word "help" can be entered for any prompt during a session, at which point the program identifies the user's present

position and the next logical step to be performed. All operation prompts are labelled with three question marks (???), and all operational step prompts are identified by the words "next move??" In most cases if the user responds to a prompt with an impossible or incorrectly formatted response, the program will signal a discrepancy and repeat the prompt.

NCHARAN uses the principle of graphics image storing. This means that any graphic generated by the program can be stored as a graphic file. Recall of the graphic file will display the graphic on the screen, so that in Operation 2, for instance, the cell locations of digital maps can be referenced by cursor. Note that it is only the map image that is displayed. The data contained in the image are not accessible.

The following sections describe each operation of the program. A flow chart accompanies each description. The flow charts are not intended to show all options for each step; rather, they indicate the logical flow as perceived by the user.

Operation 1. Data entry

Data entry, the first operation, is invoked automatically upon initiation of a session. The program prompts the user for the name of a matarray. If there is any access problem with the file name or format, the user is informed of the discrepancy, control passes to the beginning of the operation, and the user is once again prompted for a data file name. No other operation can be performed until a matarray has been entered. After correct input, the program displays the following five data matrix parameters:
1. Name of matarray
2. Total number of cells in the study area
3. Total number of variables
4. Number of rows of cells in the map area
5. Number of columns of cells in the map area

Once a database is entered, subsequent access to Operation 1 displays the attributes of the current database or prompts for a new database and displays its attributes.

Figure 3 shows the flow diagram for Operation 1.

Operation 2. Model-cell selection

The model-cell selection operation is used whenever the user wishes to initialize, modify, or cancel the list of model cells. The cell numbering convention for this procedure requires that cell number 1 be in the lower left corner of the region, and the cells are numbered increasing from left to right, one row at a time. The seven steps of this operation are
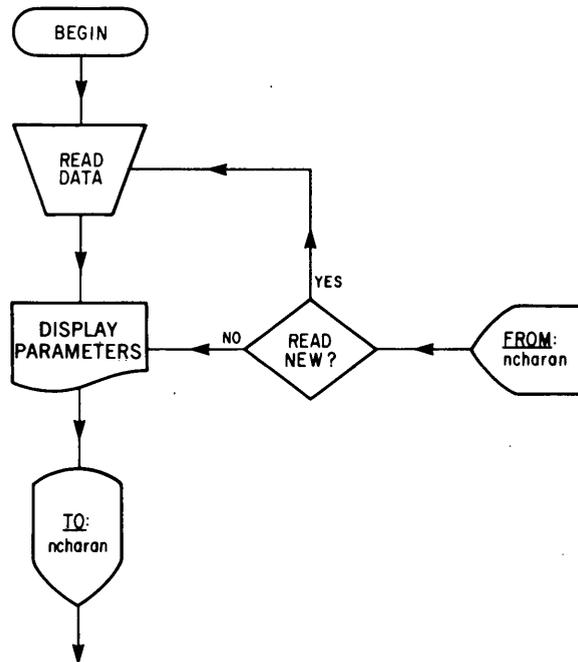
Figure 3. Flow diagram for Operation 1, data entry.

1. Model-cell file input
2. Model-cell keyboard entry
3. Cursor cell entry/deletion
4. Display of model-cell list
5. Save the model-cell list as a file
6. Consistency check and return
7. Initialize current model-cell list to zero

Step 1. Model-cell file input

In this step the user can enter previously created files of model-cell numbers. Model-cell files are created in Step 5 of this operation. After entry of any file of model-cell numbers, the user can choose to repeat the file entry procedure and concatenate other files to the existing array, or to display the region with outlines of the present aggregate of model cells, or to do both. Errors occurring during file input are reported to the user.

Step 2. Keyboard model-cell number entry

This step permits the direct keyboard entry of model-cell numbers. This step can be iterated for concatenation of cell numbers or display of model cell outlines or both.

Step 3. Cursor model-cell number entry or deletion

This step uses the screen cursor to locate and add or delete model cells from the current list of model-cell numbers. For a working base map, the user can either

display a plot of the current model cells or display any graphics file that was previously plotted on the screen and stored. Any plot of the variables, logic variables, or degree of association for the region can be saved as a graphic file and used later to select model cells.

After the display of whatever graphics image is selected, the program displays the cursor and prompts the user. The five available options are

1 add a cell
2 delete a cell
3 polygon vertex (except the final vertex)
4 final polygon vertex
. quit the cursor

To add a cell to the current list, the user must position the cursor anywhere within the desired cell, enter the number 1, and depress the carriage return [CR]. The program responds by inscribing the outline of the selected model cell within the regional grid boundary of that cell. If that cell has already been selected, the program so informs the user and the user is asked to try again. The program also notifies the user if the cursor is positioned outside the region boundary.

Model-cell deletion is performed in essentially the same way used to add single cells, except that in this case the user enters the number 2 instead of 1. The program responds with the letter X inscribed within the cell boundary. If the cell has not as yet been selected (and thus cannot be deleted), the user is notified and asked to try again. The program also notifies the user if the cursor is positioned outside the boundary of the region.

A polygon is a rapid and convenient way to select a large number of model cells in a single step. In this case

the user positions the cursor at each successive vertex of a polygon and enters the number 3. The program responds by inscribing an asterisk (*) within the cell at that vertex and, except for the first vertex, drawing a line from the current vertex to the previous vertex.

At the final polygon vertex, the user must enter the number 4, and the program responds with an asterisk at the vertex, a line drawn to the previous vertex, a closure line drawn to the first vertex, and the inscription of model cell outlines in all cells enclosed by the polygon. The program automatically removes cell number redundancies without notifying the user.

By entering a period (.) the cursor step is terminated; the user can repeat the step, display the resultant current model-cell outlines, or continue with any step in the operation.

Step 4. Model-cell number display

In this step both the total number of selected cells in the model and a list of the cell numbers is displayed. This option is often used if the user wishes to identify previously selected cells.

Step 5. Save the model cells

Step 5 provides the capability to save the current list of cells as a file for future use. The program prompts the user for the name of the file to be generated and informs the user of any errors occurring during generation of the file.

Step 6. Consistency check and return

Step 6 invokes an automatic model consistency check and a return to the operation level. In the model consistency check, the program examines the selected model variables in all model cells. If no variables have been selected, the user is notified and the program returns to the operation level. If model variables have been selected, the user is notified which cells lack data and the data that they lack. The user can then either delete all cells that lack the necessary data or continue with an inconsistent model. For example, the data for only one variable might be missing from a majority of cells. It might then be advisable to retain the inconsistent cells in the model and delete the particular variable in Operation 3.

If all the variables are required, the inconsistent cells may be deleted. Caution must be exercised when deleting inconsistent model cells, particularly if the list of model cells had already been saved. Failure to delete the inconsistent cells can cause the program to abort a session when trying to calculate the degree of match for these cells.

Step 7. Initialize current list to zero

Execution of Step 7 initializes the number of selected model cells and the array of model cell numbers to zero.

Figure 4 shows the flow diagram for Operation 2.

Operation 3. Selection of model variables

The selection of model variables is made whenever the user wishes to initialize, modify, or cancel the list of model variables. The six steps in this operation are as follows:
1. Display the list of available variables
2. Add or delete variables from current list
3. Save the current list of variables
4. Display the current list of variables
5. Check model consistency and return to operation level
6. Initialize the number of selected variables and the list of selected model variables to zero

Step 1. Display the list of available variables

In this optional step the user can display the entire ordered list of variables in the current data array. Names of variables may be up to four characters; the display gives both names of variables and their corresponding numbers. All reference to variables in the characteristic analysis procedure is by their assigned numbers, not their names. It is a good practice to make a hard copy of this list for use throughout a session.

Step 2. Add or delete variables from the current list

This step provides the option of initializing a list or adding or deleting variables from the current list of selected model variables. To initialize a list or to add variables, the user may read the list of model variable numbers from a previously created file or enter the numbers directly from the keyboard. The user is notified of any errors occurring when the program is accessing an input file. To delete variables, the user must enter the number of the selected variable on the keyboard.

Step 3. Save the current list of selected model variables

In this step, the user can save the current list of selected model variables in a file for future use. The user is notified of any errors occurring in the file generation. Care should be taken when saving a list that includes logic variables generated in Operation 4. Although the number of the logic variable will be saved, the variable must be redefined for each session and each change of database.
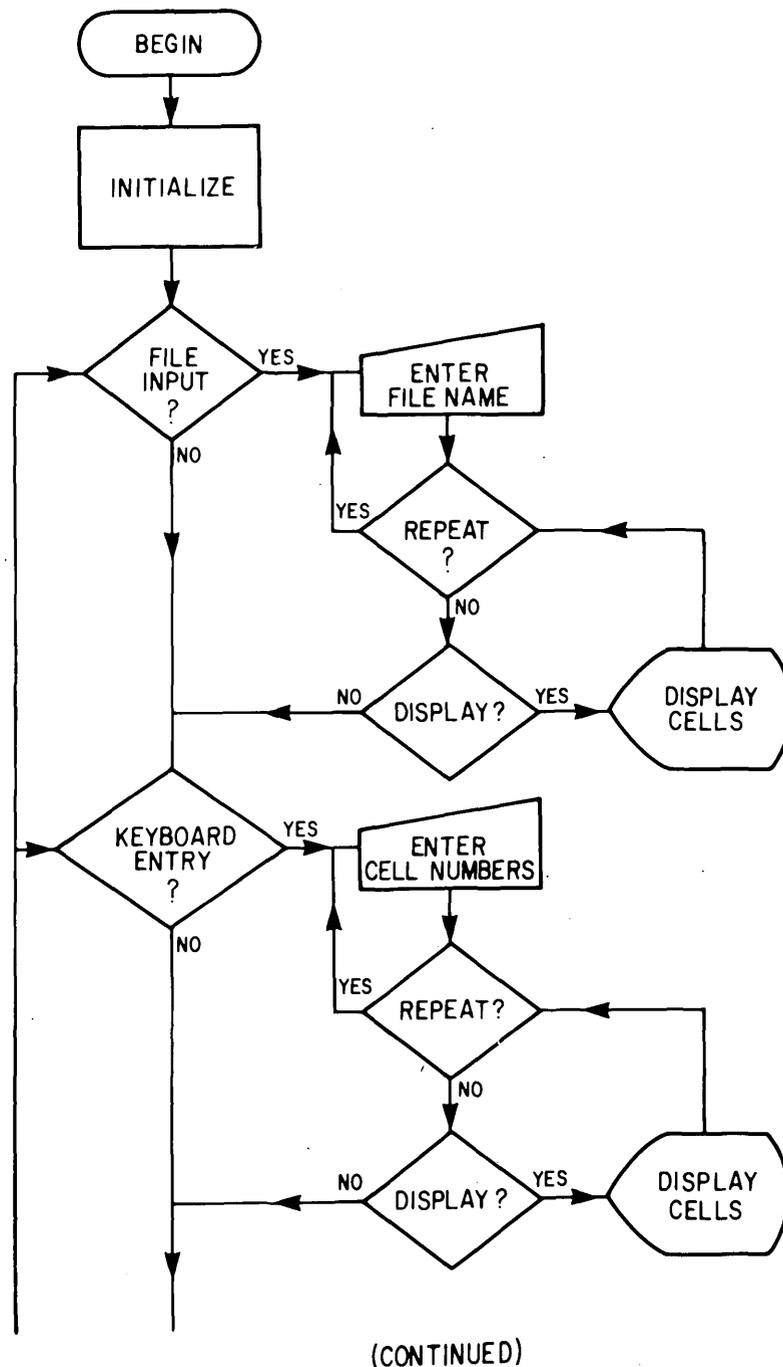
Figure 4. Flow diagram for Operation 2, model-cell selection.

Step 4. Display the list of selected model variables

In this step the program displays the names and corresponding numbers of the current list of selected model variables.

Step 5. Check model consistency and return to operation level

This step invokes an automatic model consistency check and return to operation level. Each model cell is checked for missing data for each model variable selected, and the user is notified of the cells that have

missing data and the corresponding missing variable or variables. As in Step 6 of Operation 2, the user may continue with an inconsistent model or delete the inconsistent cells from the model.

Step 6. Initialize the number of selected variables and the list of selected model variables to zero

This step initializes the list and number of selected model variables to zero.

Figure 5 shows the flow diagram for Operation 3.
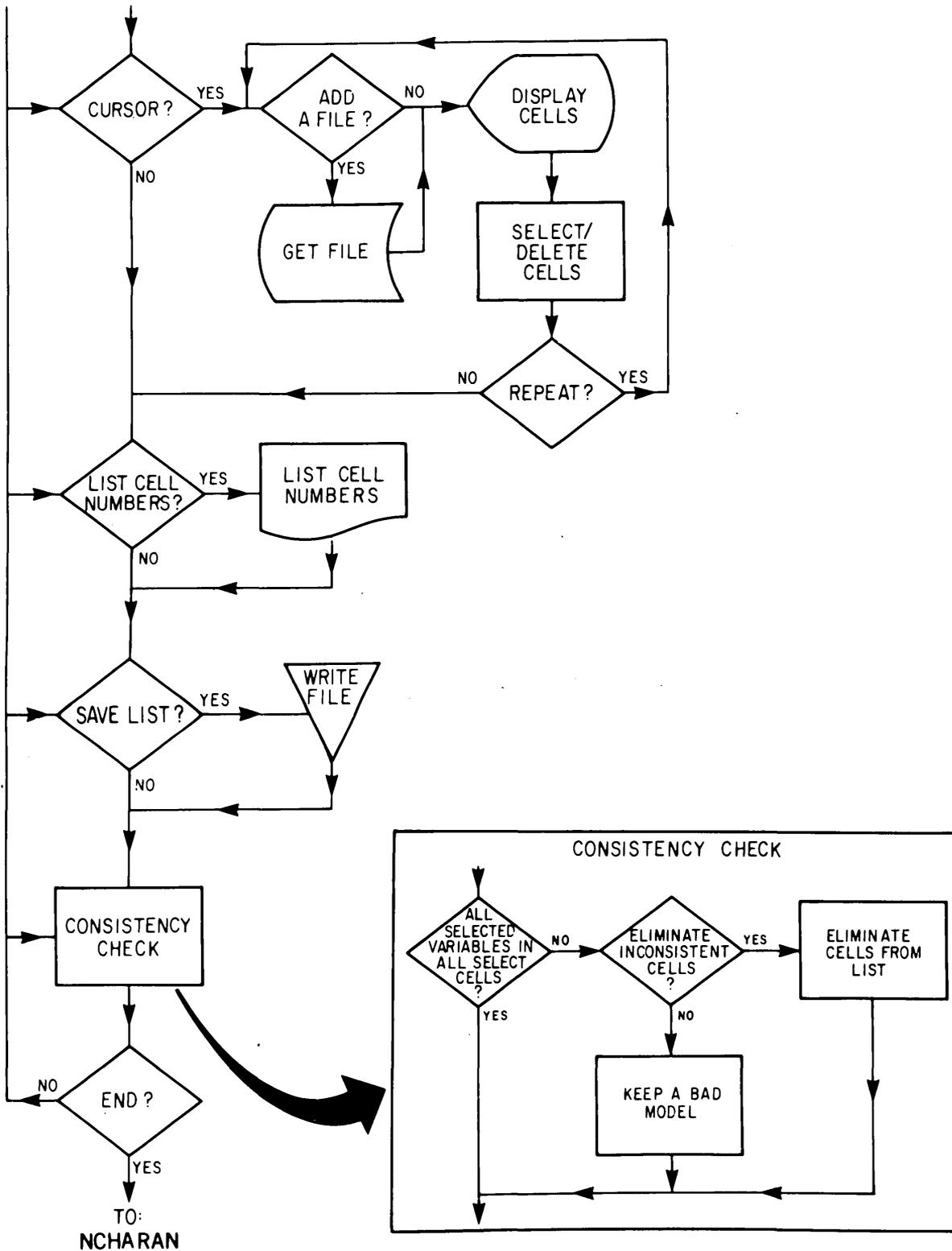
Figure 4. Continued.

Operation 4. Construction of logic variables

This operation enables the user to create logic variables. In characteristic analysis, a logic variable is a variable that is composed of logical combinations of other variables (including previously defined logic variables). These variables are stored only during the time when the corresponding database is open and are deleted
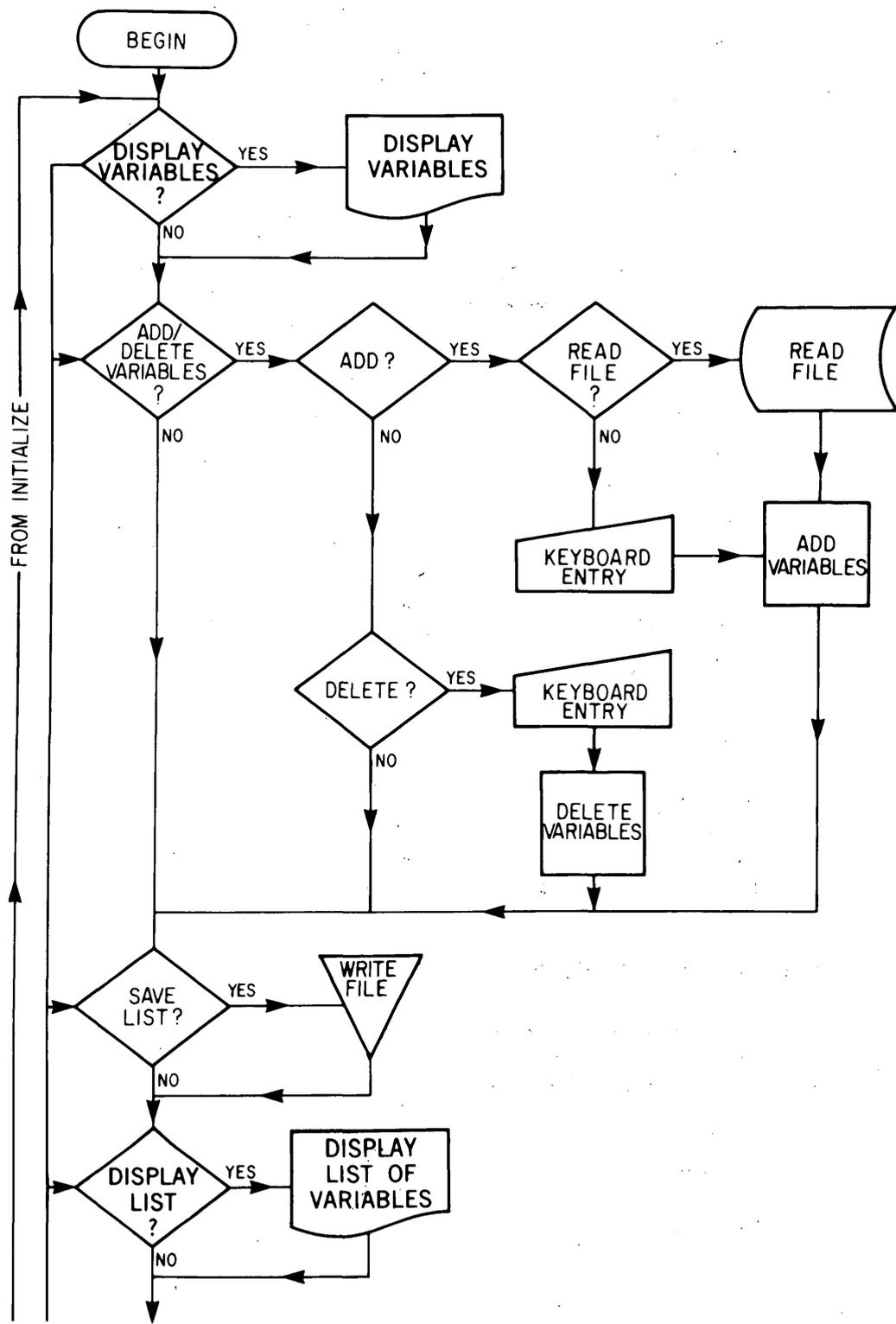
(CONTINUED)

Figure 5. Flow diagram for Operation 3, variable selection.

Figure 5. Continued.

whenever a session is terminated or the database is changed. The user may also read a logic variable from a file created external to the program.

Figure 6 shows the flow diagram for Operation 4.

Operation 5. Computation of model weights

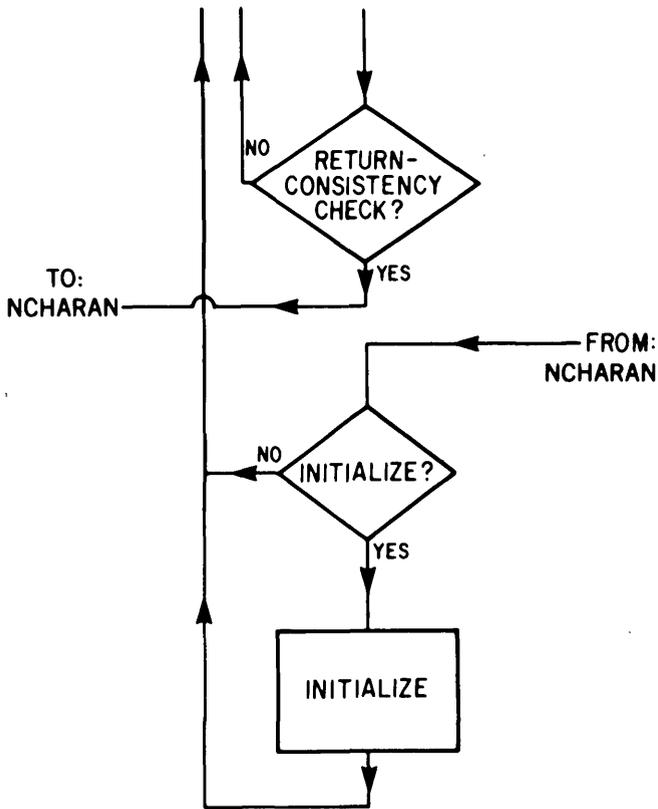This operation enables the user to enter previously computed variable weights or to compute the weights of selected model variables. The seven steps in this operation are as follows:

1. Enter previously computed weights
2. Compute product and probability matrices
3. Display the matrices
4. Select weight-computation method
5. Display the weights
6. Save the weights
7. Return to operation level

Step 1. Enter previously computed weights

In this step the user can enter previously computed weights either directly from the keyboard or from a previously written file of weights. In either case, the list of selected model variables must be arranged in the same

order as the list of variable weights being entered. If not, the user is notified. If weights are entered by reading a file, the file must have been created in Step 6 of this operation.

Step 2. Compute product and probability matrices

In this step the program automatically computes the product matrix for the current model. The user can also compute the probability matrix using sampling with or without replacement. (Sampling without replacement is more accurate but requires more computer time.) These matrices are described in the next step.

Step 3. Display the matrices

The user has the option to display the product matrix, the tally matrix, and the probability matrix. The output consists of an annotation declaring the type of matrix (either "product," "tally," or "probability"), row and column labels, page numbers, and the values. Screen output is confined to 30 variables (120 characters) across the page and 59 lines with one variable per line. Before each matrix is displayed, the program pauses for the user to copy and clear the screen. Page numbers and matrix labels are useful when joining separate printout pages of large matrices.

For the product matrix, the values represent the number of positive-positive plus negative-negative matches minus the positive-negative mismatches between pairs of variables. For the tally matrix, the main diagonal represents the number of positive values, the upper-right triangular portion represents the number of positive-positive matches, and the lower-left triangular portion contains the number of negative-negative matches for pairs of variables. For the probability matrix, the upper-right triangular portion represents the number of positive-positive plus negative-negative matches for pairs of variables, and the lower-left triangular portion represents the probability that these matches are not due to chance.

Step 4. Select weight-computation method

There are three options for computing model weights: (1) taking the square root of the sum of squares for each row of the product matrix, (2) taking the first principal component of the product matrix, and (3) taking the first principal component of the probability matrix. Only one method can be executed at a time, and only the most recently computed weights are current in the program (unless precomputed weights have been entered to override the most recently computed weights).
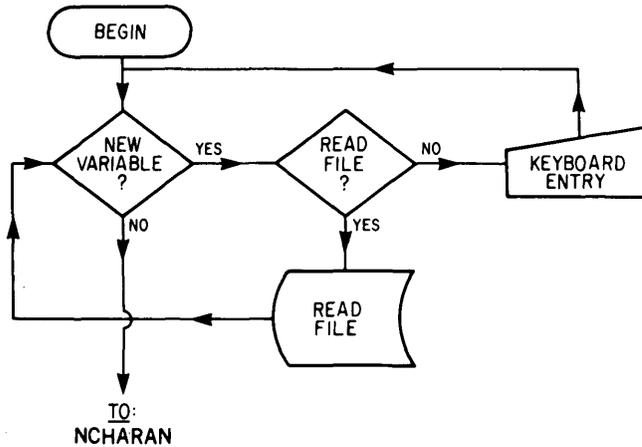
Figure 6. Flow diagram for Operation 4, construction of logic variables.

Step 5. Display the weights

This step displays the variable number, the variable name, and the computed weights. The user may sort the model weights in decreasing order before printing.

Step 6. Save the weights

In this step the user can save the current list of weights for future use in Step 1 of this operation. The user is prompted for an output file name and is informed of any errors occurring in the file generation.

Step 7. Return to operation level

Step 7 permits return to the operational stream in the program.

Figure 7 shows the flow diagram for Operation 5.

Operation 6. Computation of degrees of association

In this operation, the current model weights are used to compute the degree of association (match) for each region cell with the model. The user also has the option to specify a subset of cells stored previously as a cell file. The user may display a frequency distribution of the degree of association within both model cells and region cells. These distributions can then be used to define natural groupings of degrees of association. Figure 8 shows the flow diagram for Operation 6.

Operation 7. Display of variables or degree of association

Operation 7 provides the capability of displaying either variables (including logic variables) in the current

database or the degrees of association from any previous cycle through the program. The displays can be supplemented by including either the grid or the model cell outlines or both. All displays have at least one dimension plotted full screen. Thus, databases with different dimensions will be displayed at different scales. An advantage of the one-dimension full-screen concept is that the displays exhibit maximum legibility. A disadvantage is that data bases with different cell dimensions, regardless of their map scale, will be displayed at different scales. Because the screen is used mostly as a device to interactively display data relationships, the choice between maximum legibility and uniform fixed scale was made in favor of maximum legibility.

The following are the three steps in this operation:
1. Plot a variable from the current data array
2. Plot the degree of association
3. Return to operation level

Step 1. Plot a variable from the current data array

In this step the user can plot any variable, including logic variables, of the current data base. Data values are symbolically plotted as follows:

Plotted symbol	Data value
+	+1
0	0
−	−1
[blank]	−2 (outside area of interest)

In addition to the plotted symbols, the user can elect to plot either the region grid or the model-cell outlines or both. The plot can be saved as a graphic file for future use in model-cell selection. In this case the user is prompted for the name of the graphic file to be created. The program informs the user of any errors that occur during file generation. This file is not a file of data
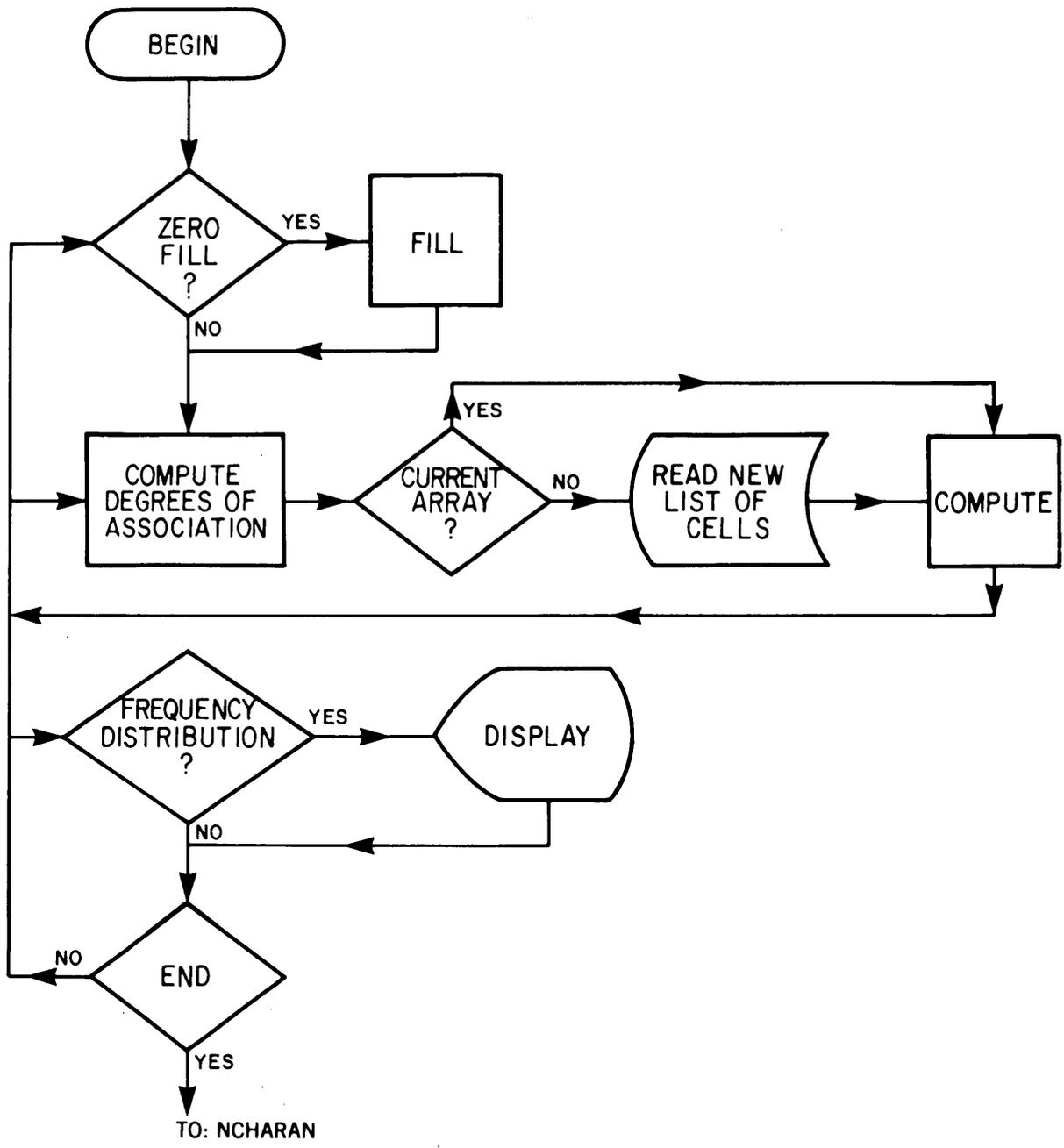
Figure 7. Flow diagram for Operation 5, computation of model weights.

Figure 8. Flow diagram for Operation 6, computation of degrees of association.

values from which the map was drawn; it is a picture that corresponds exactly to the gridded regional outline displayed for cell selection in Operation 2.

Step 2. Plot the degrees of association

In this step the user may plot either the current array of degrees of association or a previously computed file of degrees of association saved in Operation 8. The user must specify the internal class boundaries needed to group the degrees of association into classes. The maximum and minimum values define the outermost class boundaries. The program allows at most eight internal boundaries, which define as many as nine classes. These class boundaries must be specified in ascending order, separated by commas, and entered all on one line. Starting with the class containing the lowest degree of association, the program numbers the classes from 1 through n, where n equals the number of the highest class, and plots the class numbers at their appropriate cell positions. The user may save the plot as a graphics file for future use and is notified of any errors occurring during file generation.

Step 3. Return to operation level

This step permits return to the operation level. Figure 9 shows the flow diagram for Operation 7.

Operation 8. Save the degrees of association

In Operation 8, the user can save the degrees of association as a file for future use. Only the currently
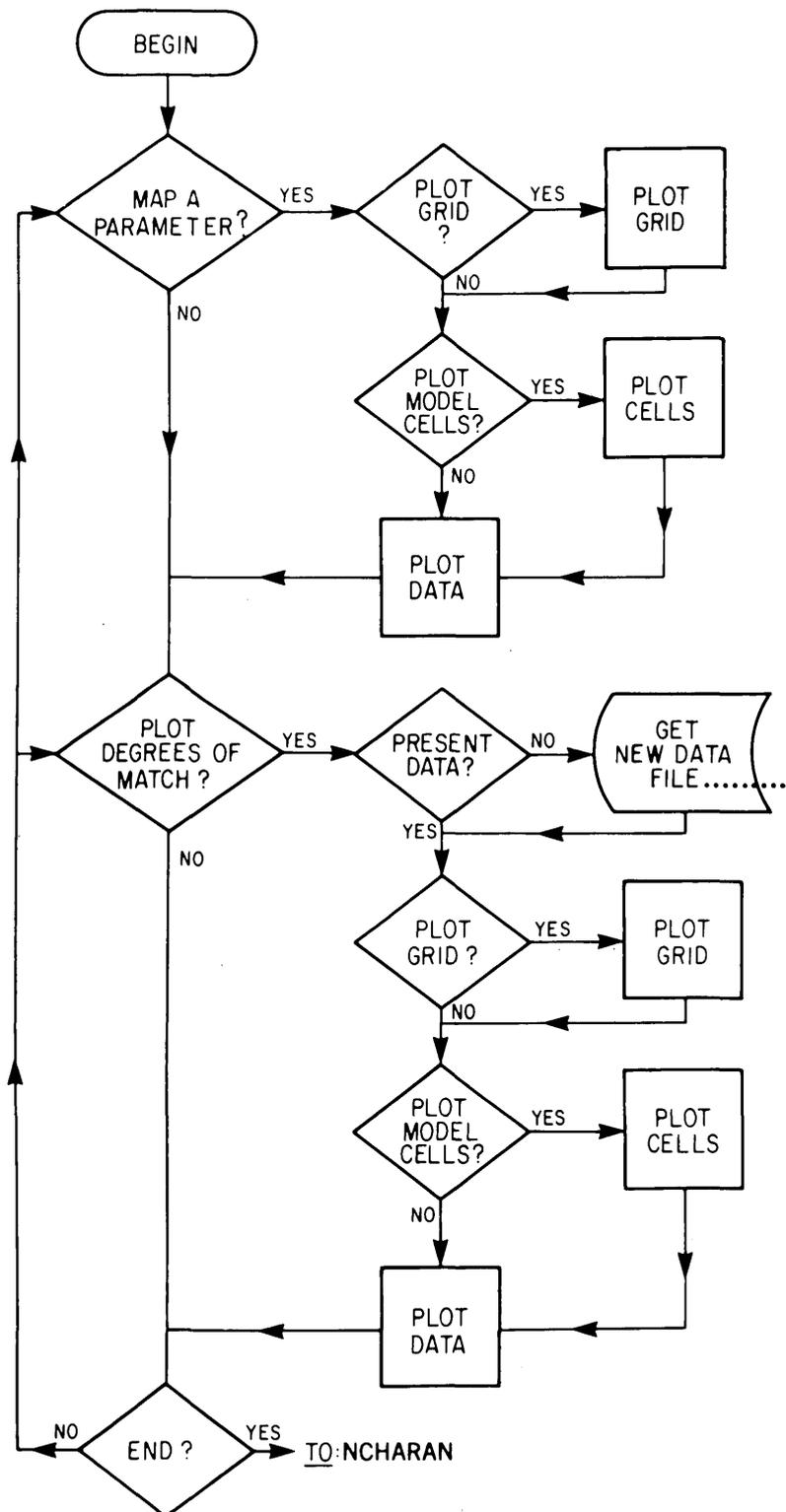

Figure 9. Flow diagram for Operation 7, display of variables or degrees of association.

computed degrees of association generated in Operation 6 can be saved, and the file that is generated can only be used in the program as input to Step 2 of Operation 7. The user is notified of any errors occurring during file generation. A hard copy of the frequency distribution of the degrees of association should be made for each file generated in this operation.

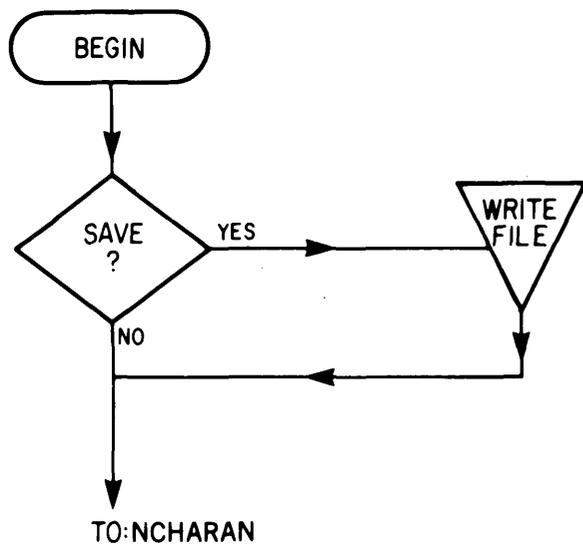Figure 10 shows the flow diagram for Operation 8.

Figure 10. Flow diagram for Operation 8, save the degrees of association.

Operation 9. Quit NCHARAN

This operation permits normal termination of the program. The user can branch to any other operation or quit the program.

SAMPLE SESSION

The sample session presented below demonstrates the use of the NCHARAN program. Because of the many steps within each operation, it is not practical to demonstrate here all the options in the program. The sample session should provide the reader, however, with a basic understanding of the operations performed in characteristic analysis. The data used in the example are from Bridges and others (1983) and relate to the study by McCammon and others (in press) on estimating the undiscovered uranium endowment in the Westwater Canyon Member of the Morrison Formation in the San Juan Basin, N. Mex. The matarray "sjbdat" consists of values of some 80 ternary-transformed variables for 5,292 cells defined in the study area. To assist the reader, the variables used in the example are displayed at the end of the session. The two cell files that are read in the example were created in an earlier session and consist of cells chosen to represent the model and cells chosen for evaluation with respect to the model, respectively. As further aids to the reader, the responses of the user are in bold-face type, and marginal comments are provided that give a more complete explanation. [CR] means that the user has pressed the [RETURN] key.

Session	Comments
ncharan	Call NCHARAN
Enter transmission rate in characters/second **960** [CR]	
[Screen clears]	
Would you like a brief statement of operating instructions?? **yes** [CR]	
To obtain an explanation of any step in this procedure type "help" or "?" at any pause.	Ask for instructions.

The main operational sequence of the characteristic analysis program is composed of the following sequential operations:

1. Data file input and matrix parameter review.
2. Model cell selection.
3. Selection/Deletion of variables.
4. Construction of logic variables.
5. Computation of weights.
6. Computation of degrees of association.
7. Display of computed values.
8. Save the degrees of association.
9. Terminate the session.

The prompt ??? is given before each major operation is executed, and 'CR' automatically enters the next operation in the sequence. Entry of any number from 1 to 8 will cause the flow to branch to that operation.

For each operation there are specific prompts, and "help" or "?" will cause explanations of these to be printed during the prompting process within any operation.
"ncharan" operation 1. Data file entry.

Enter the name of the ternary (or binary) encoded data file. **sjbdat** [CR]   Call matarray
The data array is sjbdat
Number of cells 5292
Number of variables   81
Map area is 84 cells N-S by   63 cells E-W.
???   **2** [CR]   Call for operation 2

"ncharan" operation 2. Model cell selection.
Next move?? **1** [CR]

You are about to begin....
Step 1. Model cell FILE ENTRY.
Enter FILE NAME with cell number list or CR for next local step. **rm1** [CR]   Read in file that
Enter r to repeat, d to display, CR to continue. **d** [CR]   consists of model cells
                              [Screen clears]
rm1
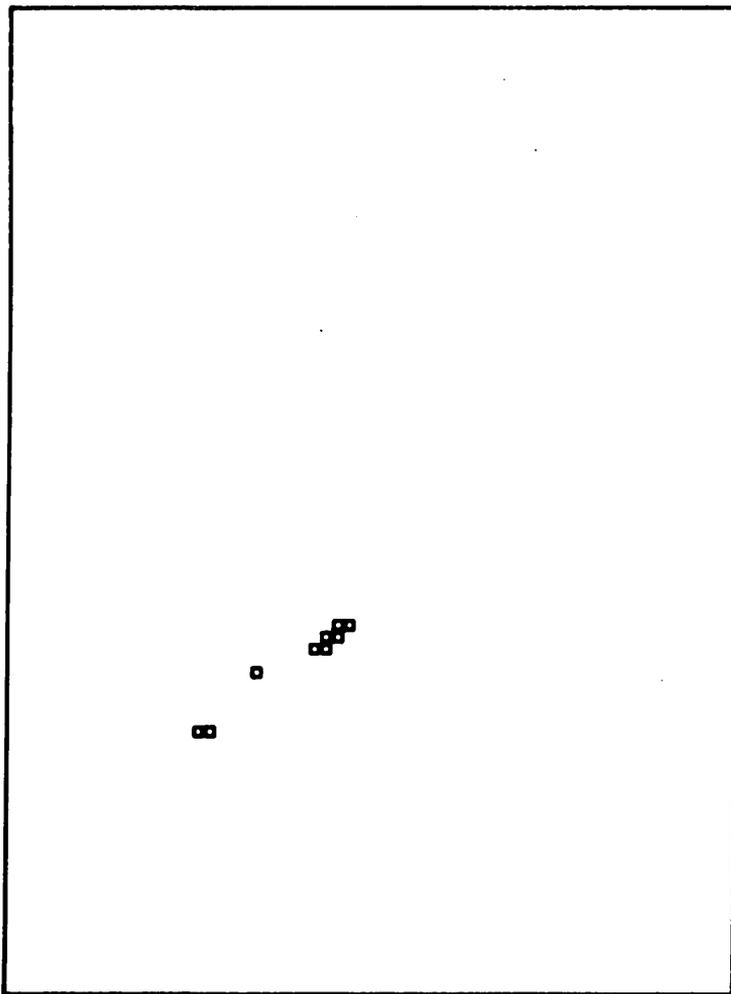
Display of cells in
model

Next move?? **?** [CR]   Request possible
You are selecting model cells. By entering the appropriate step number   moves
you can branch to any of the following steps:

1. Read the model cell list from a file.
2. Enter model cell numbers from the keyboard.
3. Add or delete cells by cursor.
4. Display the current list of model cells.
5. Save the model cells as a file.
6. Check consistency of the model and return to NCHARAN.
7. Wipe out the current selected list of cells and start from scratch.

Enter the number of a desired STEP, or
CR to begin STEP 2

Next move?? **6 [CR]**

you are about to begin....
Step 6. Check consistency of model and return to NCHARAN.
Model cells have been chosen, but NO VARIABLES have as yet been selected.
??? [CR]

"ncharan" operation 3. Selection/Deletion of variables.
Next move?? **1 [CR]**

The following are all available variables:

rm1e	1	rm2e	2	rm3e	3	rm5e	4	rm6e	5	rm7e	6	rm9e	7,
rmad	8		9	rmae	10	expl	11	expt	12	exin	13		14,
	15	dph1	16	dph2	17	dph3	18	dph4	19	mask	20	rmn2	21,
rm2t	22	rm3t	23	rm5t	24	rm6t	25	rm7t	26	rm9t	27	rm1n	28,
rm2n	29	rm3n	30	rm5n	31	rm6n	32	rm7n	33	rm9n	34	rm1s	35,
	36	rm3s	37		38	rm6s	39	rm7s	40		41		42,
	43		44	rm5a	45	rm6a	46	rm7a	47	rm9a	48	rm1c	49,
rm2c	50		51		52	rm6c	53	rm7c	54	rm9c	55	ct01	56,
ct21	57	ct02	58	ct22	59		60	scdt	61	scdm	62	scdr	63,
scor	64		65	mcdt	66	mcdm	67	mcdr	68	mcor	69	mgth	70,
mgpt	71	kaol	72	woxt	73	woxm	74	woxr	75	fm1f	76	rm2f	77,
rm3f	78	rm4f	79	rm5f	80	rm6f	81								

Display of variables (Note that not all the numbers are associated with variables)

Next move?? **5 [CR]**

You are about to begin....
Step 5. Return to "ncharan",
Model cells have been chosen, but NO VARIABLES have as yet been selected.
Am returning to NCHARAN.
??? **4 [CR]**

"ncharan" operation 4. Logic variable construction.
Enter constructs in the form nv = F(v1, v2,....,vn)
    where nv is less than or equal to the prompt number
    and F is an expression containing operators and variable nos.
Enter CR only to exit this step.
next new variable = 82
82 = 21 + (28*35) [CR]
next new variable = 83
83 = (61 + 66)*70 [CR]
next new variable = 84
[CR]
??? **3 [CR]**

nv stands for new variable
Define logical combinations of variables to form new variables

----

"ncharan" operation 3. Selection/Deletion of variables.
Next move?? **2 [CR]**

You are about to begin...
Step 2. Addition/Deletion of variables to the list of selected ones.

Enter:
    1 to ADD variables,
    2 to DELETE variables,
    CR to return to the variable selection stream.
1 [CR]
Enter r to read vars. from a file, d for direct entry, CR for local stream.
d [CR]
Enter the variable numbers. **82,83,73,*** [CR]

Next move?? ? [CR]
You are selecting variables. By entering the appropriate step number, you can
branch to any of the following steps:
    1. Display a list of all available variables.
    2. Add(Delete) variables to(from) the current list.
    3. Save the current list as a file.
    4. Display a list of current selected variables.
    5. Return to NCHARAN.
    6. Wipe out the current list of selected variables and start from scratch.
Next move?? **3** [CR]
You are about to save a file of selected variables.
Enter:
    The name of the file to be written, or CR to return to stream
[CR]
Next move?? **4** [CR]
You are about to begin...
Step 4. Display a list of current selected variables.
The following are the SELECTED variables:
    v82   82   v83   83   woxt 73
Next move?? [CR]
You are about to begin...
Step 5. Return to "ncharan".
CONGRATULATIONS!!! You have a consistent model.
??? **5** [CR]

"ncharan" operation 5. Computation of weights.
Weight calculation steps are as follows:
    step   function
    1   read user-supplied weights from a file or the keyboard
    2   compute product, tally, and probability matrices
    3   display the matrices
    4   select a method and compute weights
    5   display the weights
    6   save the weights in a file
    7   return to NCHARAN
Enter CR to execute step 2
Next move?? [CR]
Probability calculation:
Enter CR to bypass, 1 to use sampling-without-replacement model
    or 2 to use sampling-with-replacement model
[CR]
Enter CR to execute step 3
Next move?? [CR]
Do you wish to display the product matrix?? **Y** [CR]
This is a pause for copying. Enter CR to continue. [CR]
[screen clears]
product   matrix, page 1 of 1
            82   83   73
            v82 v83 woxt
82 v82   7   7   7
83 v83   7   9   5
73 woxt   7   5   9

Selection of variables
for model; asterisk
signifies end of list

Product matrix

Do you wish to display the tally matrix?? **Y**

This is a pause for copying. Enter CR to continue. [CR]
[screen clears]
tally   matrix, page 1 of 1
               82   83   73
               v82   v83   woxt
82   v82   7   7   7
83   v83   0   9   7
73   woxt   0   0   7
Enter CR to execute step 4
Next move?? [CR]
Select a method of weight calculation
     1   :   centroid method
     2   :   1st principal component of product matrix
     3   :   1st principal component of probability matrix
or CR to bypass weight calculation

Enter 1, 2, 3, or CR to bypass **2** [CR]
Enter CR to execute step 5
Next move?? [CR]
Do you wish to display the weights?? **Y** [CR]
Do you want the weights displayed in sorted order? **Y** [CR]
     variable   name   weight
     82   v82   0.577
     83   v83   0.577
     73   woxt   0.577
Enter CR to execute step 6
Next move?? [CR]
Enter CR or the file name on which weights are to be saved
[CR]
Enter CR to execute step 7
Next move?? [CR]
??? **6** [CR]

"ncharan" operation 6. Computation of degrees of association.

Do you wish to compute the degs of assoc for all cells?
no [CR]

You are about to begin....
Step 1. Model cell FILE ENTRY.

Enter FILE NAME with cell number list or CR for next local step. **ora__mask** [CR]

Do you wish to substitute zero (0) for blank (missing) data items? (y or n)
n [CR]
Degrees of association have been computed.
Enter:
     CR to clear screen and display a frequency distribution of scores, or
     1   to return to NCHARAN.
[CR]
     [screen clears]

Tally matrix

Read in file of cells
to be evaluated

frequency distribution

	class interval		model freq	cells pct	non-model freq	cells pct
1	−1.00	−0.90	0	0.00	0	0.00
2	−0.90	−0.80	0	0.00	0	0.00
3	−0.80	−0.70	0	0.00	0	0.00
4	−0.70	−0.60	0	0.00	224	12.27
5	−0.60	−0.50	0	0.00	0	0.00
6	−0.50	−0.40	0	0.00	0	0.00
7	−0.40	−0.30	0	0.00	386	21.14
8	−0.30	−0.20	0	0.00	0	0.00
9	−0.20	−0.10	0	0.00	0	0.00
10	−0.10	−0.00	0	0.00	0	0.00
11	−0.00	0.10	2	22.22	542	29.68
12	0.10	0.20	0	0.00	0	0.00
13	0.20	0.30	0	0.00	0	0.00
14	0.30	0.40	0	0.00	461	25.25
15	0.40	0.50	0	0.00	0	0.00
16	0.50	0.60	0	0.00	0	0.00
17	0.60	0.70	0	0.00	25	4.65
18	0.70	0.80	0	0.00	0	0.00
19	0.80	0.90	0	0.00	0	0.00
20	0.90	1.00	7	77.78	128	7.01
totals			9		1826	

Frequency distribution of degree of association for cells in model and cells being evaluated

```
??? 7 [CR]
"ncharan" operation 7. Display of computed values and/or raw data.
Next move?? [CR]
Enter:
     the NUMBER OF THE VARIABLE that you wish to plot, or
     CR to return to the local stream.
[CR]
Next move?? [CR]
You are about to plot the degrees of association.
Enter:
     CR to proceed with a plot of the degrees of association, or
     the number of the step to which you wish to branch.
[CR]
Enter CR to plot current data, or, enter name of file to be plotted.

Do you wish to plot the grid (y or n)? n [CR]
Do you wish to plot training cell outlines (y or n)? n [CR]
Enter the desired map type. 1 [CR]
Enter the boundaries in ascending order, separated by commas, and all on one line.
     Exclude the upmost and lowest boundaries. .3,.6,.9 [CR]
Enter 1 to save graphics IN ADDITION TO SCREEN output, CR for SCREEN ONLY.
[CR]
```
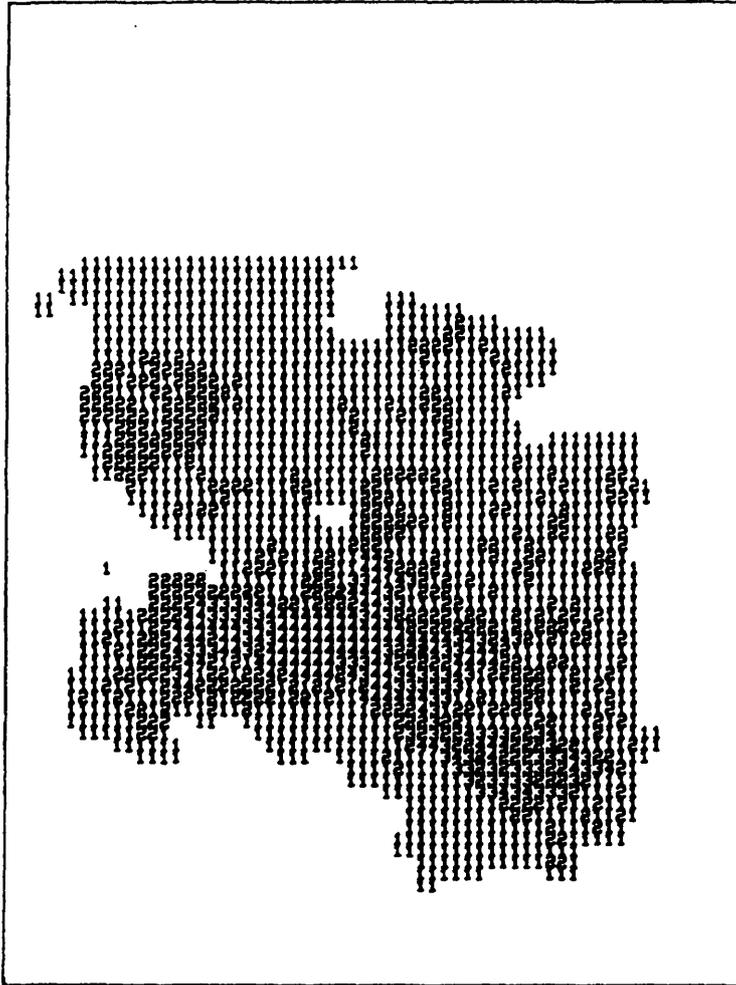
Partitioning of degree of association into 4 favorability classes ranging from class 1, least favorable, to class 4, most favorable

ora__mask

Next move?? [CR]

??? **8** [CR]

"ncharan" operation 8. Save the degrees of association.
Enter:
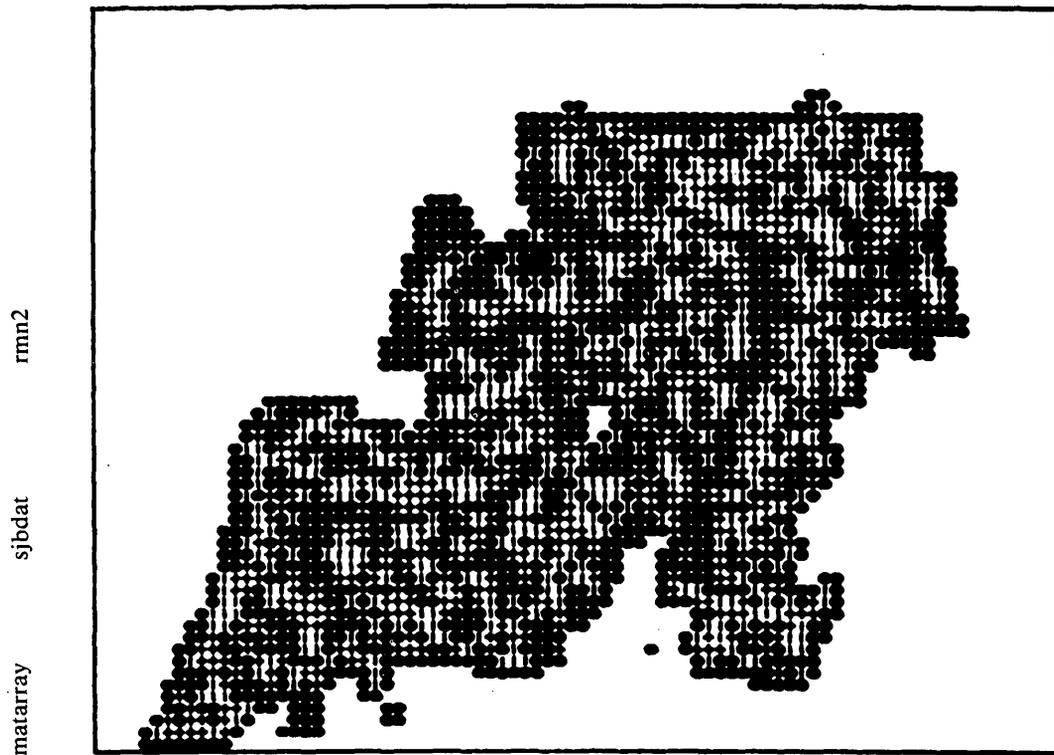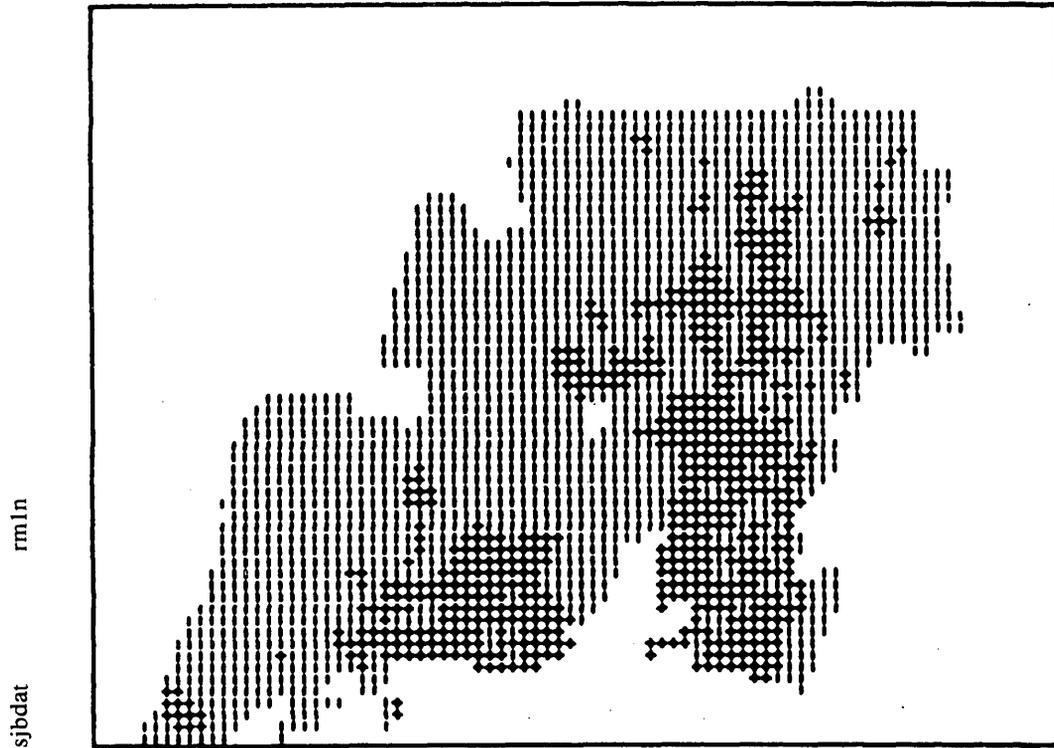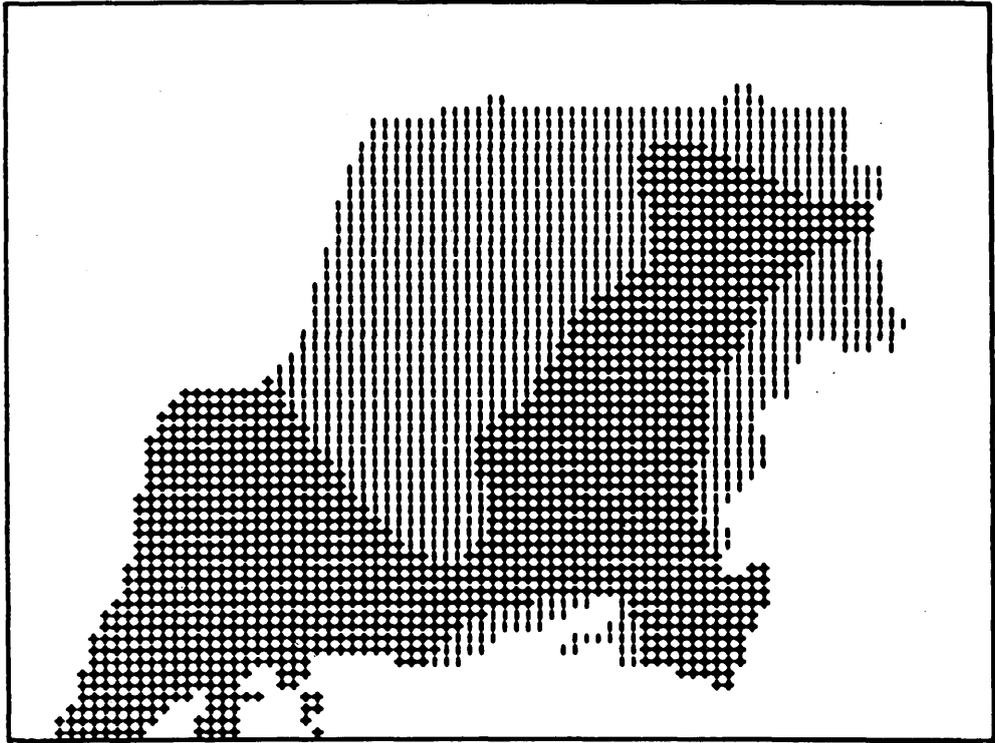   The name of the file to be created, or CR for next local step.

Save degree of
association for cells
evaluated with
respect to model

rm1fav
???[CR]

You are at the end of an analysis. Enter CR to QUIT, or SOME OTHER OPERATION NUMBER.
[CR]

The following are the displays of the variables used in the example above
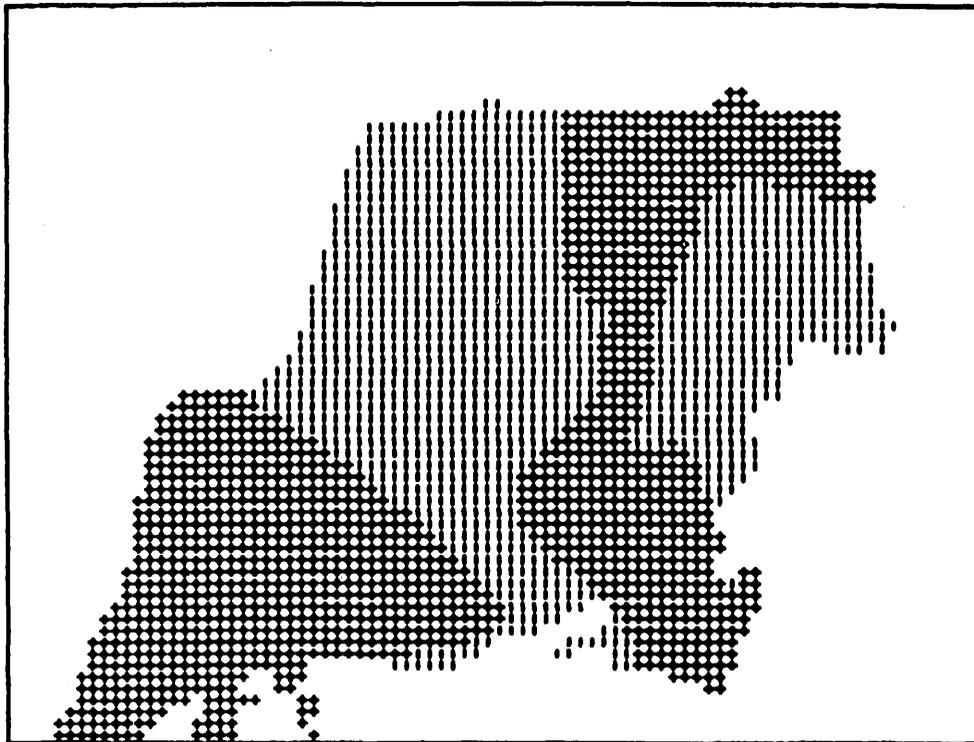

rm1n

sjbdat

rmn2

sjbdat

matarray

scdt

sjbdat

rm1s
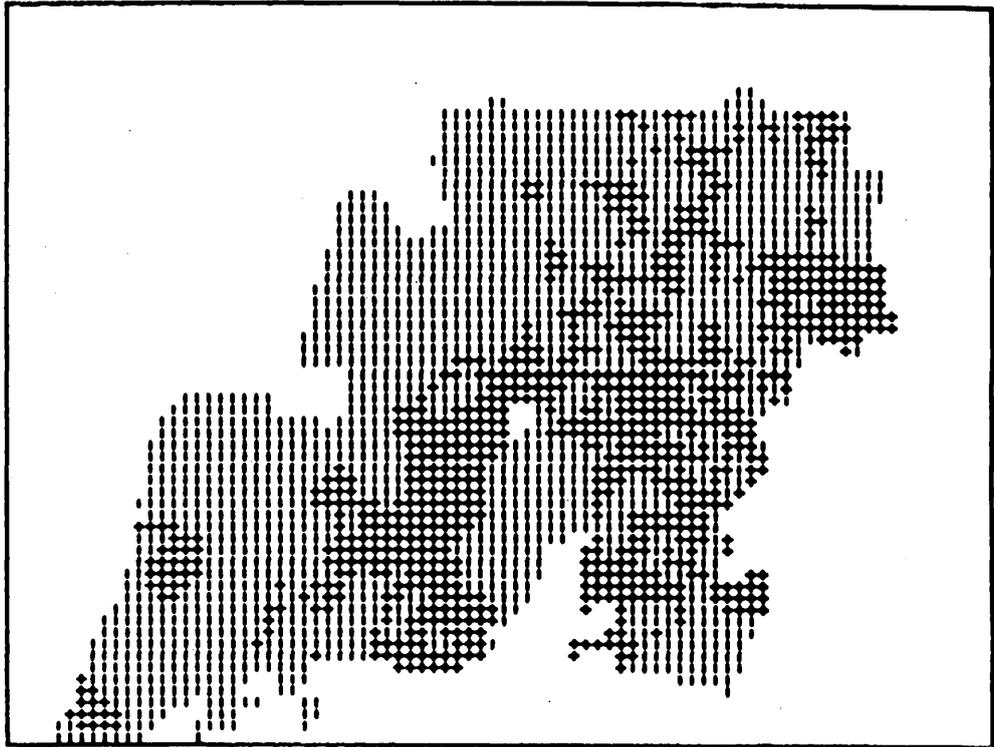
sjbdat

mgth

sjbdat

mcdt
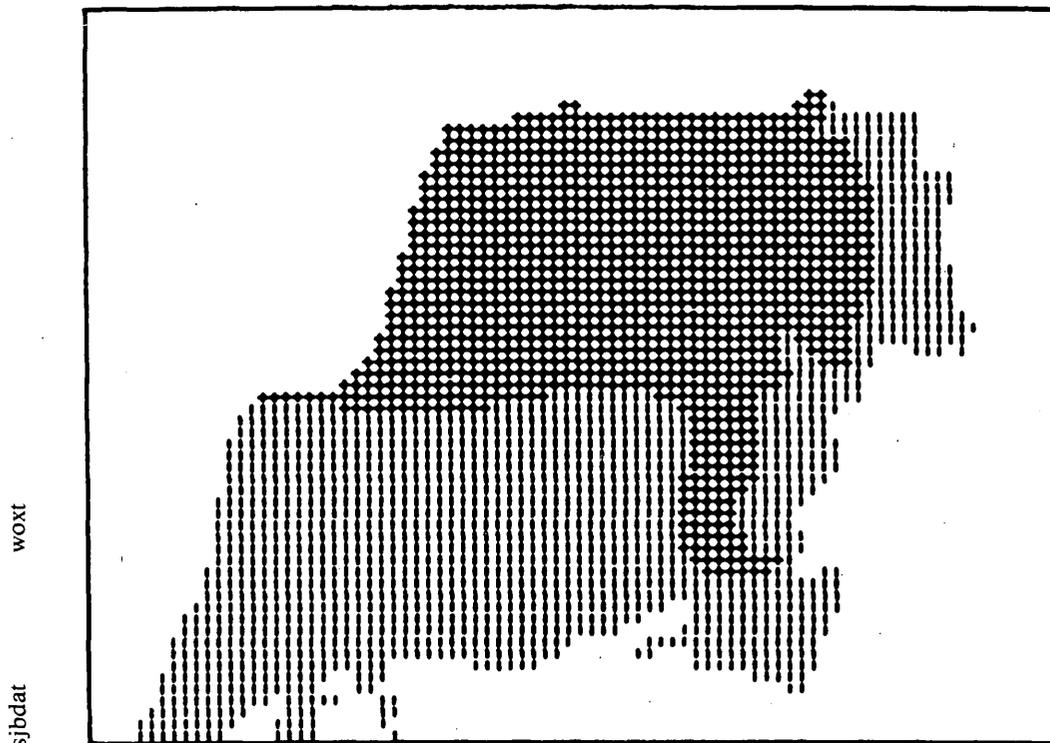
sjbdat

woxt

sjbdat

REFERENCES CITED

Botbol, J. M., 1968, Characteristic analysis of base metal mining districts in the continental United States: Unpublished Ph.D. thesis, University of Utah, Salt Lake City, Utah.

_____1969, A model way to analyze the mineralogy of base metal mining districts: Mining Engineering, v. 22, no. 3, p. 56-59.

_____1971, An application of characteristic analysis to mineral exploration, in Decision-making in the mineral industry: Canadian Institute of Mining and Metallurgy, Special Volume 12, p. 92-99.

Botbol, J. M., Sinding-Larsen, R., McCammon, R. B., and Gott, G. B., 1978, A regionalized multivariate approach to target selection in geochemical exploration: Economic Geology, v. 73, no. 4, p. 534-546.

Bowen, R. W., 1977, Compression of ASCII files on MULTICS: MULTIPS, U.S. Geological Survey, Computer Center Division, Reston, Va., v. 2, no. 16, p. 7-8.

Bridges, N. J., Hanley, J. T., and McCammon, R. B., in press, PREPRO — A computer program for encoding regional exploration data for use in characteristic analysis: Computers & Geosciences.

Bridges, N. J., Kork, J. O., and McCammon, R. B., 1983, Lithologic, petrographic, and resource data to accompany report on estimation of uranium endowment of the Westwater Canyon Member, Morrison Formation, San Juan Basin, New Mexico: USGS-GD-83-007, NTIS PB 83-225042.

Finch, W. I., Granger, H. C., Lupe, Robert, and McCammon, R. B., 1980, Genetic-geologic models — a systematic approach to evaluate geologic favorability for undiscovered uranium resources. Part I, Research on uranium resource models, a progress report: U.S. Geological Survey Open-File Report 80-2018-A, 55 p.

McCammon, R. B., Botbol, J. M., McCarthy, J. H., Jr., and Gott, G. B., 1979, Drill-site favorability for concealed porphyry copper prospect, Rowe Canyon, Nevada based on characteristic analysis of geochemical anomalies: Society of Mining Engineers Fall Meeting, Preprint No. 79-311, 7 p.

_____1983b, Characteristic analysis applied to multiple geochemical anomalies over a concealed porphyry copper prospect, Rowe Canyon, Nevada: American Institute of Mining, Metallurgical and Petroleum Engineers Transactions, v. 272, p. 1998-2002.

McCammon, R. B., Botbol, J. M., Sinding-Larsen, R., and Bowen, R. W., 1983a, Characteristic analysis — 1981. Final program and a possible discovery: Mathematical Geology, v. 15, no. 1, p. 59-83.

McCammon, R. B., Finch, W. I., Kork, J. O., and Bridges, N. J., in press, Estimation of undiscovered uranium endowment in the Westwater Canyon Member, Morrison Formation, San Juan Basin, New Mexico, using a data-directed numerical method: U.S. Geological Survey Professional Paper 1311-T.

Sinding-Larsen, R., Botbol, J. M., and McCammon, R. B., 1979, Use of weighted characteristic analysis as a tool in resource assessment: Evaluation of Uranium Resources Proceedings Advisory Group Meeting, International Atomic Energy Agency, Vienna, p. 275-285.

Sinding-Larsen, R. and Strand, G., 1981, Quantitative integration of mineral exploration data. A case study of the Grong mining region of Norway: Episodes, v. 1981, no. 1, p. 9-12.

Tektronix, Inc., 1977, 4010 A01 PLOT 10 Terminal Control System: Beaverton, Oregon, Tektronix, Inc.

APPENDIX. SYSTEM DEPENDENCIES

The system dependencies fall into the categories of hardware, language, operating system, and proprietary graphics software.

Hardware dependencies are primarily related to word size. To reduce the amount of memory required and yet accomodate a data matrix consisting of up to 85 variables for 6,000 cells, a compression technique was devised. Since a variable may assume only the values zero through three, this value may be represented using only two bits; hence, 17 values may be stored in each 36-bit word and an entire cell of 85 variables may be stored in 5 36-bit words. When subroutine NCHINIT is called, the array "npower" is initialized to successive powers of four. This array is used by the subroutines PUTVAL and IVAL, which perform the insertion and extraction of values from the data matrix.

The language dependencies are related to the MULTICS MR68 Fortran, which is commonly called "old_fortran." It is assumed that single precision variables can contain four characters and that double precision variables can contain eight characters; ENCODE/DECODE statements are used as well as the "free-field" format item "v." A number of subroutines use the MULTICS FORTRAN AUTOMATIC statement to control storage allocation. This statement is used only to reduce disk storage requirements for the compiled program and can be removed on those systems not supporting dynamic allocation of data arrays.

The operating system dependencies relate to calling various MULTICS commands and routines from within NCHARAN. In order to input screen coordinates using the graphics cursor in subroutine DRWCEL, the MULTICS command STTY must be called to permit the input of coordinate values that coincide with system control characters. Subroutine OPENF calls the PL/1 routine OPENER to attach and open files. This routine is used to trap invalid file names, thus avoiding abnormal program termination. Subroutine OPENF also calls the MULTICS routine IO_CALL to close and detach files. Subroutine PASSER calls the MULTICS command FO to redirect output from the console to a file and MULTICS command CO to redirect output back to the console. Subroutine TKINIT calls the MULTICS routines SETUP.TEKTRONIX.TCS and SETUP. TEKTRONIX.AG2 to initialize the Tektronix Plot-10 software. Subroutine TKOUT uses the MULTICS command STTY to permit output of data which represents a graphic image. That image is then printed using the MULTICS PRINT command.

The proprietary graphics software used is the Tektronix Plot-10 Terminal Control System and the Tektronix Plot-10 Advanced Graphics System.