CORD STUDY TECHNICAL NOTE 42

GEOGRAPHIC DATA PROCESSING CONCEPTS AND THEIR APPLICATION TO PARK AND OTHER LAND USE PLANNING

M. F. Goodchild

ABSTRACT

The issues surrounding the automated processing of geographic data are subtle and complex. This paper examines the field from the particular stance of the land use planner, on the assumption that the major purpose of automated processing in this context is to answer questions of a geographic nature that cannot be readily answered by manual map analysis. In this context study areas are usually small, and study emphasis is more on the ability to provide rapid answers to complex questions than on the banking of large volumes of data.

Issues are reviewed under the general headings of the nature of geographic data, the basic pros and cons of machine processing, the alternatives in data storage, and the issues in the encoding process. The conclusions reached are consistent with the designs of several operating geographic information systems.

The PLUS system developed by the author is used to provide two sets of illustrations in the final sections of the paper. One is of the encoding and verification of a typical geographic source document to produce an accurate data base for a variety of applications. The other shows the PLUS/2 system in operation as it responds to a set of queries about a study area in Southeastern Manitoba from a fifteen-map data base of physical geographical variables.

PREFACE

Early in 1973 the author was asked to prepare a paper for the Department of Indian and Northern Affairs on the then current status of geographic data processing and its potential as a tool for park planning (see Reference 23). That paper identified the major issues and reviewed the various approaches that have been adapted to them. The paper was discussed at a series of meetings with interested representatives of a number of agencies and led (in the summer of 1973) to a contract with the Lands Directorate of Environment Canada for the development of a limited land-use planning system to satisfy a number of objectives. First, the system was to permit the manipulation of maps in a variety of ways in an interactive mode. This sort of facility is impractical in many geographic information systems but is feasible in a system dedicated solely to land-use planning because the accuracy requirements are less severe. Second, the system was to make the maximum use of existing data banks by providing interfaces to them. (The greatest obstacle to the adoption of geographic data processing has been the problem of data encoding, which requires a large commitment to hardware and staff. Yet this obstacle is slowly being eroded as more data becomes available in agency archives.) Finally, the system was to allow the input of raw data in limited quantities and in the most efficient form possible.

The system became known as PLUS (see Reference 26), organised into two sections, PLUS/1 containing the data bank interfacing and data input, and PLUS/2 the interactive map manipulation. A case study was used to provide test data and to demonstrate the use of the package (see Reference 25). Interfaces were developed with the co-operation of the Canada Geographic Information System (Lands Directorate, 1974) and the Canada Soil Information System (Dumanski and Kloosterman, 1973) and the system has been demonstrated in a variety of contexts (see, for example, Reference 27). Data can be input and output in a variety of other standard forms, allowing direct interfacing with SYMAP, CALFORM and POLYVRT (Harvard University Laboratory for Computer Graphics, 1974) and US Census DIME files (US Bureau of the Census, 1970).

The present paper reviews the concepts and important issues underlying geographic data processing, and provides illustrations and examples drawn from the author's work with PLUS. Development of the system is continuing.

THE NATURE OF GEOGRAPHIC DATA

The principal purpose of this paper is to review the field of geographic data processing in the particular context of land use planning. (For a more general review see Reference 12, 58). It might be as well to begin with an attempt at a definition of the phrase "geographic data", before discussing the problems it presents. Consider the question "How much of the land that lies within five miles of a certain town is under 500 feet above sea level?" If a data base is to be capable of answering such a question, it must contain information on height and on location. It might for example contain the precise locations of contour lines in the area, through the locations of a sequence of points at which the contours change direction. Or, as another alternative, the topography might be represented by accurate heights at every intersection of a very fine grid mesh laid over the area. With either method, an answer could be given through an appropriate series of arithmetic and logical steps.

The two methods mentioned above are often referred to as the polygon and grid systems of data encoding; each is geographic because each contains explicit locational information. As a counterexample, suppose that the topography had been represented by the average height of land in each Census Enumeration Area on the map. It would then have been impossible to answer the original question without reference to the Enumeration Area locations and boundaries. So a geographic data base can be defined as one possessing explicit geographic information.

The traditional method of storing, displaying and communicating geographic information has of course been the map, a stylised two-dimensional representation of reality. All of the methods currently available for processing and communicating information are one-dimensional, working with a single linear stream of data, and geographic information must be represented in this way if any of the current data processing technologies are to be applied to it. This difference in dimensionality is responsible for a fundamental paradox in all efforts at image or picture processing, pattern recognition, or geographic data processing; since the eye is a two-dimensional processor, it is often as or even more effective than the digital computer in certain kinds of operations. The computer is often only a marginally beneficial processor of geographic data.

Geographic data can be classified in a variety of ways; in a discussion of data processing it is particularly useful to distinguish between point, line and area data. A point representation is often used when a map must show the locations of objects whose size is much less than the intervening distances between the objects. Line representations are used for such phenomena as railways or roads; area representations are used when characteristics are to be ascribed to substantial tracts of land, as on soil maps for example.

Several other rather specialised types of map can be classified as variations on this typology. Contour maps can be seen as (a) area maps on which the contours are area boundaries, or (b) line maps on which heights are to be interpolated between neighbouring lines. Points are sometimes used to infer areas, as when the locations of Census districts are given by centroids.

A map can display several types of information at once, but for data processing purposes it is difficult to deal with more than one. The term 'coverage' is used in the paper to refer to a single type of data and a defined geographic area. So a coverage is a one-dimensional classification of a two-dimensional segment of the earth's surface. An area-map coverage consists of a single partitioning of an area into non-overlapping, homogeneous zones.

The relative importance of each type will of course depend on the context. However, the scale of most land use planning is sufficiently detailed that the bulk of requisite information is areal. Much of the relevant data, on soils, land use, vegetation, etc., is collected directly in areal form. In addition, most socio-economic data is collected by predetermined areal units and is therefore also areal. On the other hand, topography, climatic variables and water table depths are usually determined by point sampling and represented through contoured maps. Highway rights of way are sufficiently narrow that they can often be dealt with as line data. In summary, while areal data predominates, all of the standard types can be recognised as relevant to land use planning.

Finally, accuracy is an important issue which can easily be overlooked. The paper touches on the question at several points, as the degrees to which the results of data processing reflect the information on the source maps. There are complex payoffs between the costs of processing and the accuracy of the results. Basically, there is no value in providing accuracy in data that will never be used in analysis. Nor is there value in an elaborate system which can provide precise answers if they are based on inaccurate data. An air photo analyst may subjectively draw a line which encloses land of a roughly homogeneous soil type; yet the data processing system will objectively ascribe the precise soil type to precisely all of the land within the line.

The Case for Machine Processing

Geographic data processing is a relatively new field. Although it has been possible to handle geographic data since the earliest days of electronic computers, it is only through the enormous reduction in costs per operation over the last 20 years, and the simultaneous increases in storage capacity, that processing has become economically feasible on a realistic scale. (The points made in this section are discussed in more detail in Goodchild (1973).)

Three types of geographic data manipulation are considered in this section. The advantages of automatic processing are assessed in the context of mapping, then in relation to geographic information retrieval, and finally in the analysis of geographic relationships and model calibration. All three areas are to some degree relevant to present day land use planning, and are assessed in that context. But, as well, it is important to bear in mind that the development of geographic data processing techniques may open new directions for land use planning in the future.

Automated Cartography

Once geographic information has been coded, it can readily be retrieved in graphic form as a computer-produced map (see Reference 48). Two main methods exist, based on the plotter or computer-driven pen, and the line printer. In Canada, both the Canada Land Inventory, through the Canada Geographic Information System (Lands Directorate, 1974) and the Department of Agriculture, with the Canada Soil Information System (Reference 15) have produced massive data files of coded information readily usable for producing plotted maps. Canada Land Inventory coverages are described in Reference 15. Similar efforts in other areas can be found in the GRDSR program of Statistics Canada (Reference 55), the National Topographic Survey of the Department of Energy, Mines and Resources, and other agencies, both federal and provincial.

In the line printer case, relatively crude maps have usually been made by overprinting symbols to create different shadings (see Figure 4) although more recent variants such as the electrostatic plotter produce more acceptable results by a dense application of small dots.

At the simplest level, automatic cartography is not an efficient way of making a map. It would clearly take at least as long to describe the map to the machine by encoding it as it would to plot it by hand. But there are many sets of circumstances which may justify computerising a map. First, once encoded the map is stored by the computer so that the benefits of encoding extend beyond the cartographic exercise. The data might be used to plot the same map at different scales or to draw only part of it: or The coded map might be used as a source of information in some analytic problem. Second, there are cases in which the same computer-coded base might be used to produce different maps. For example, once the outlines of census areas are coded it is possible to produce innumerable maps of different census variables very cheaply. Furthermore, one may wish to process the map between the coding and drawing stages. It might be necessary to change projection or scale: or computer processing might be used to make corrections and modifications before drawing an updated final copy of a map. And finally, a great deal of work in the printing process (in the preparation of colour separations) can be avoided by effective automation.

But while each of the operations mentioned may be very relevant to the routine production of topographic, soil or census maps by the appropriate agencies, the land use planner is unlikely to justify encoding geographic data for the sole purpose of automated cartography. He is more likely concerned with a unique, limited study of a specific area, with very little likelihood that the data will be useful to other studies in the future, at least without extensive updating and revision.

Information Retrieval

The land use planner must work with answers to a wide variety of geographic questions, ranging from the simple "What is here?" on a single coverage to complex comparisons of several coverages, such as "How much land of types x, y and z on coverages A, B and C lies within N miles of location L?", or problems involving a search over feasible locations to find optimum sites. Some of these questions can be answered by eye, and it is not difficult to construct a case in which the eye can accomplish an operation (evaluation) more rapidly than a third-generation computer, given a clear map and a prior encoding. Even when the capabilities described above may fit directly into planning practice, more advanced operations are likely to do so only in the long term.

Consider the following example. A power line is planned across productive agricultural land and must follow the route of minimum cost, defined as the composite of construction cost, land acquisition, transmission cost and agricultural production loss. Potential agricultural production can be calculated from agricultural capability, which is available in map form, but the actual relationship has yet to be established. Advanced geographic data processing capabilities might be used in this context, first to establish the relationship between potential production and land capability (by regressing past yields against the map of capability) and then to select the optimum site by a minimising search algorithm. Whatever the set of operations finally considered useful for land use planning, it is clear that any system must have the potential for rapid and easy addition of new capabilities as the need for them becomes apparent.

THE ISSUES IN MACHINE STORAGE

In discussing the costs and benefits of various approaches to machine storage of geographic data sets, there are three separate sets of criteria to be considered. First, there are alternative methods for actually encoding data, each with its own advantages and disadvantages. Second, long-term storage presents its own problems of efficiency and reliability. And third, there is a wide variety of eventual applications of the data which must be reflected in the methods of storage and retrieval.

The two methods of automated cartography, plotter and printer, correspond precisely to the two major methods of data storage, polygon and grid. Plotter maps are drawn by driving a pen around each of a number of boundaries; polygon data is stored as a sequence of points which, when connected, indicate polygon boundaries. Similarly, printer maps are created by placing a symbol in each of a finite array of cells; correspondingly, grid data is stored as a collection of discrete values associated with grid cells.

Within each of the two classes there are innumerable variants. The record in a polygon data set can consist of all that land classified by a certain code, or a complete contiguous polygon, or a segment of a polygon boundary between two junctions, or a single boundary point. These may usefully be referred to as levels 1, 2, 3 and 4 files respectively. Certain variants contain more than one level. The SYMAP structure is a pure level 2 form, whereas the CALFORM structure combines level 4 and level 2 subfiles. (See Reference 30.)

The grid data structure may vary according to the convention for ordering cells, although the commonest by far is the basic printer sequence, by rows from the top left corner. More importantly, the structure may be compressed so that extensive runs of the same cell type are replaced by integral multipliers. The literature on data structures is extensive; for reviews see References 00, 00.

The accuracy with which polygon data represents its source document depends on the accuracy with which each boundary line is encoded. If the boundaries are excessively contorted, they may require a high density of points to create an accurate representation, while the largely straight lot lines of a land use map can be coded with a much lower density. But in general a moderate number of boundary points are sufficient to make an adequately accurate polygon data set.

By contrast, a grid data set has a definite and predetermined level of accuracy once the grid cell size is set. Relationships exist between grid cell size and the reliability of any estimate of area, or any other geographic parameter. (See Reference 24, 57.) Yet in spite of the sacrifice of precision, gird data has a very clear advantage in any of the more complex forms of geographic data processing. Maps can be made equally readily from either data type, but any operation which requires the evaluation of land type at a specific location, or the comparison of different maps, is far more readily executed with grid data than with polygon. With grid data, it is a simple matter to calculate the cell in which a point lies, and thus to retrieve its characteristics, but with polygon data it is a substantial problem to associate a specific polygon with a given point.

For this reason most advanced data processing of the type outlined above has been carried out on gridded data sets. The sacrifice of accuracy has been judged to more than compensate for the excessive computational problems of information retrieval from polygon data. Optimally, each coverage is coded by a precisely similar grid, so that the nth cell in the first grid corresponds to the nth in every other.

But while information retrieval and analytic operations are best performed on grid data, there is no need to store all data in that form, as transfer is easy between the two types, particularly from polygon to grid. Grid data sets are only rational once a level of accuracy and a study area have been determined. Thus one finds that the polygon structure has been adopted by most agencies with responsibility for data collection, storage and cartography, while grid data is favoured for sophisticated retrieval and analysis. Ideally, data should be collected and stored in polygon form, and then gridded for each specific study area and level of accuracy. Collecting directly in grid form places too many restrictions on the usefulness of the data, while analysis using polygon data is bound to be inordinately expensive.

THE ISSUES IN DATA ENCODING

The encoding of point or line data presents no particular problems. Co-ordinates can be taken directly from source documents with readily available digitising equipment. Grid coding of areal data is similarly elementary, though unlikely for the reasons discussed in the preceding section. But the encoding of areal data in polygon form has historically been one of the main stumbling blocks of geographic data processing. In the simplest manual system a digitiser operator indicates a sequence of points that he wishes to have encoded, together with the contents of each polygon. There are many variants of the process, depending on the degree to which computer processing is used in the eventual editing. A few possibilities are discussed below.

The SYMAP Method: Complete Polygon Encoding

In this system, the polygon constitutes a file record (a level 2 file according to the above notation). The operator codes a series of points defining each polygon outline in clockwise order, and then codes the contents of each polygon. See Figure 1(a). As a result every boundary is coded twice, as part of the polygon on either side. The two versions of each line will inevitably conflict, so the crude image is often processed to resolve any deviations of less than some allowable amount, say 0.05-inch, to create a clean file.

The CALFORM Method

The file cleaning process mentioned above is potentially dangerous, since any real detail in the map at a scale of less than 0.05-inch will be removed along with the spurious detail. The CALFORM method attempts to resolve this potential ambiguity and to economise in effort by avoiding the double encoding of each line. Every polygon vertex on the map is first numbered and location encoded in a level 4 file. Then the polygons are coded by listing the sequence numbers of their vertices, to create a level 2 file. See Figure 1(b). Although the method becomes unmanageable when large numbers of vertices have to be numbered, it avoids both of the difficulties of the SYMAP system. But a realistic encoding method must be made to deal with thousands of polygons and hundreds of thousands of vertices on a single map.

PC(Segment) Methods

PC methods divide the polygon networks into sections of boundary between junctions (a level 3 file) and so avoid the duplication problem of the level 2 methods. See Figure 1(c). But each junction now creates a problem, since it occurs at the end of each of several records, and may be coded at a slightly different point each time. So a certain amount of processing must be used to resolve discrepancies to within an allowable limit. A variety of methods have been devised to resolve discrepancies in an unambiguous way. In some cases each junction is encoded in a separate digitising phase; in others, each segment is named according to the polygons it bounds.

PC encoding can become a complex and confusing task for a digitizer operator, who must continually switch between point encoding and the identification of polygons. In some variations of the method, polygons are identified in a separate phase of the encoding process after the entire point encoding is complete. This is easier for the operator, but requires a more sophisticated processing stage to resolve more complex ambiguities at junctions. The IC method implemented in PLUS/1 is one example, see Reference 25. The operator first codes the boundaries by identifying vertices, but without necessarily beginning and ending records at each junction. The polygons are then identified by encoding one point within each one and giving its characteristics. Computer software identifies all junctions, breaks the encoded boundaries at appropriate points, and attaches each polygon type to the segments forming each polygon boundary.

Semi-automatic Methods

The encoding methods described above can be automated in various ways. First, the digitising process can be accelerated if points are coded automatically, at given intervals of time or over given distances, rather than by a conscious act of the operator. Second, a great deal of work has gone into making the line-following process semi-automatic, with a limited amount of operator control to resolve ambiguities. The map is usually scanned by a device similar to a TV camera, moving unaer the control of a small "dedicated" computer. Ambiguities are brought to the operator's attention when they cannot be resolved internally. (For example, Reference 8.)

Fully Automatic Coding

The technology necessary to read an entire map and place it in computer storage has existed for a long time. Unfortunately, such a vast amount of data is created if the map is scanned with any degree of precision that the cost of sorting out polygons by computer processing is inordinately high. In addition, most source documents must be redrawn prior to scanning to remove any marks that do not represent polygon boundaries, an operation that often occupies as much human effort as manual digitising. But several trends suggest that such systems will enjoy increasing popularity in the future. Computer processing is becoming cheaper, while the cost of manual operations is increasing. And with better software, it should be possible to scan images that are closer to the source document in quality, without requiring expensive redrafting. But at the present time the

PC (Segment) methods discussed above are clearly the most practical and economical for the creation of polygon data sets except in the largest agencies.

Editing Data Sets

These are two fundamental problems in editing or updating areal data sets and they are largely responsible for the difficulties which all practical systems have experienced. First, there are internal consistencies within any polygon data set, so that changes may have to be made at several points in the file at once. For example, if a boundary point location is changed in SYMAP data set, the change must be made to at least two of the coded points in the file. Or if a polygon characteristic is changed in a PC file, it must be changed in all of the records which make up the polygon boundary. So it is essential that editing be done using specialised software so that internal consistency can be preserved.

Secondly, there is a fundamental difficulty in relating the contents of a file to the appearance of a map. The coordinates of a point cannot be associated with particular lines on a map without the aid of a digitiser or plotting device, so both machines are usually considered essential to the editing process. Some editing systems use graphics terminals to display sections of a data set, so that the user can identify the location of errors on the original source document. Errors can then be corrected with the terminal cursor, which the operator can position to revise locations or to indicate changes. This sort of man/machine interaction appears to be the best solution to the editing of polygon data, but suffers at the moment from the smallness of graphics terminal screens. The ideal system would operate at the same scale as the source document, perhaps through a combined digitiser/plotter system with a head that could be driven or positioned manually, in an interactive mode.

EXAMPLES

The Encoding Process for Polygon Data

The first example illustrates the preparation of an areal data set of the polygon type from the source document, using the PLUS software package. The map chosen for encoding is one showing the pattern of Census Enumeration Areas in the city of London, Ontario; there are 478 separate polygons on the map, mostly with simple rectilinear outlines. The map was coded using a simple manual digitiser, with a cross hair cursor connected by a rigid arm and steel cables to two incremental counters. A digitiser operator pressed a foot switch whenever a point was to be encoded, causing the point's x and y co-ordinates to be recorded on a punched card, to the nearest 1/100-inch. The overall accuracy, expressed as the ability of the operator to find and encode the same point at infrequent intervals, is about 0.05-inch.

The IC method discussed briefly above was used to encode the map. In the first stage the operator coded the entire image as a network of lines, breaking the sequence whenever necessary to move to another section of image. Figure 1(d) shows a typical sequence and should help to clarify the method. In a second stage, a single point is located arbitrarily in each polygon, in this case its Enumeration Area identification number. This primary coding operation occupied the digitiser operator for ten hours, in which time roughly 10,000 image boundary points were coded, and 500 identifiers.

As a first step in correcting the errors in this raw data, the image and centres were plotted out at the same scale as the source document (Figure 2). Several kinds of error can be detected immediately, such as those which result when the operator fails to properly indicate a break in the encoding of the image, or when sections of image or centres are omitted. Another two hours were spent in modifying the data to remove such errors.

The data were then processed by a PLUS/1 programme package (POLYSORT) designed to identify junctions, make appropriate breaks in the line image, and attach the polygon identification to the appropriate side of each PC record. At the same time, any gaps at junctions caused by an undershoot or overshoot during digitising were removed to within a certain tolerance (0.1-inch). Nevertheless, errors could still exist in the file at this stage. Junctions could fail to close within the prescribed tolerance, or there could be problems that were not visually obvious at the earlier error correction. More seriously, the use of an error tolerance means that it is difficult to resolve real detail at scales approaching that tolerance without ambiguity.

Three checks for potential errors were made at this stage, besides those already described. First, the internal consistency of the file was checked by linking together all of the PC's forming each polygon boundary, to ensure that no gaps existed and that each was coded with the correct polygon identification. Any errors were flagged by the PLUS/1 program (EDITONE) so that they could be checked and corrected. As a second check, each polygon identification was verified from the original map. Any changes must maintain the internal consistency of the file and so must be made by an appropriate program package (EDITTWO). Finally, the PC file was plotted at the scale of the source document to ensure that the boundary locations were correct (Figure 3).

A clean polygon data set can be used in a variety of ways. The coded map can be gridded at any scale, with any size of grid cell (GRID). See Figure 4. It can be used to calculate areas, perimeter lengths, centroid locations and other geographic summaries (RPG); it can form the basis for mapping using one of the standard polygon mapping packages (SYMAP or CALFORM) and if necessary the data structure can be modified to any variant of the polygon form (STRUCT); and polygon maps can be overlain (OVER) to produce composites in which each combination of polygons becomes a new, unique figure. (GRID, RPG, STRUCT and OVER are PLUS/1 routines described in Reference 27.)

Information Retrieval with Grid Data

The second example demonstrates the use of geographic data in an information system using PLUS/2. Under this system, answers to a variety of queries are available at a user's terminal (a teletype or cathode ray tube device) in response to simple commands. In the following pages results of an actual run are photographed from a screen in precisely the way a user would see it, with underscoring applied to everything that would be typed by the user. The actual case is taken from a study of the Roseau River basin in southern Manitoba (Reference 27). Data on the area has been gridded, based on 168 columns and 36 rows, so that each map cell corresponds to one quarter section of land. Fifteen coverages of the area were stored in this grid form. The "resource value" recorded for each cell on each map coverage is the predominant value of the appropriate variable in the area covered by that cell.

A computerized "planning" session begins (Figure 5) by the user asking for assistance. The list of commands available in PLUS/2 contains several routines for the analysis of map data, for the preparation of summaries, display of maps, creation of new maps and the combination of existing coverages. The command 'COVERAGES' produces a list of maps in the system; these names, along with the size of grid area and other basic information on the study area are passed from session to session: they can also be entered by a dialogue initiated by the 'RESTART' command.

Incidentally the PLUS.2 system recognises four kinds of coverage, Alpha, Numeric, Lines and Points. Alpha and Numeric maps both consist of cellular arrays, but differ in the allowed list of operations. The values on a Numeric map are assumed to be measurements on some scale, whereas no interval or ordinal relationship is assumed between the alphabetic or numeric symbols on an Alpha map.

After finding out what converages are available the user might ask for a display of one of the maps in the system. 'AGD' is, in reality, the primary agricultural capability class, on a scale of 1 to 7. Because only 72 print positions are available on the terminal being used, the system asks for a part of the map to be specified. It has previously determined that the map must be displayed as six 'pages', two rows and three columns, so that page 1,2 corresponds to the top centre of the map. Page 1,1 is the top left, and page 2,3 the bottom right. (See Figure 6.)

These maps are crude images that can be produced at the terminal in seconds. More permanent maps of published quality might be produced from the same data by directing output to a line printer. A command 'PILOT' is available in

PLUS/2 when permitted by the specific computer system in use.

What Proportion of Land is Class 2?

In displaying a map, the system is merely making a crude reproduction of a source document. But the same data can be used to answer a series of questions. Consider first a question which can be answered through the 'TABULATE' command. Since the AGD file being considered is Alpha, the values on the map are reported in the order in which they are found, and in the actual request, the output shows that 12.3% of the area Class 2, a total of 742 cells (Figure 7).

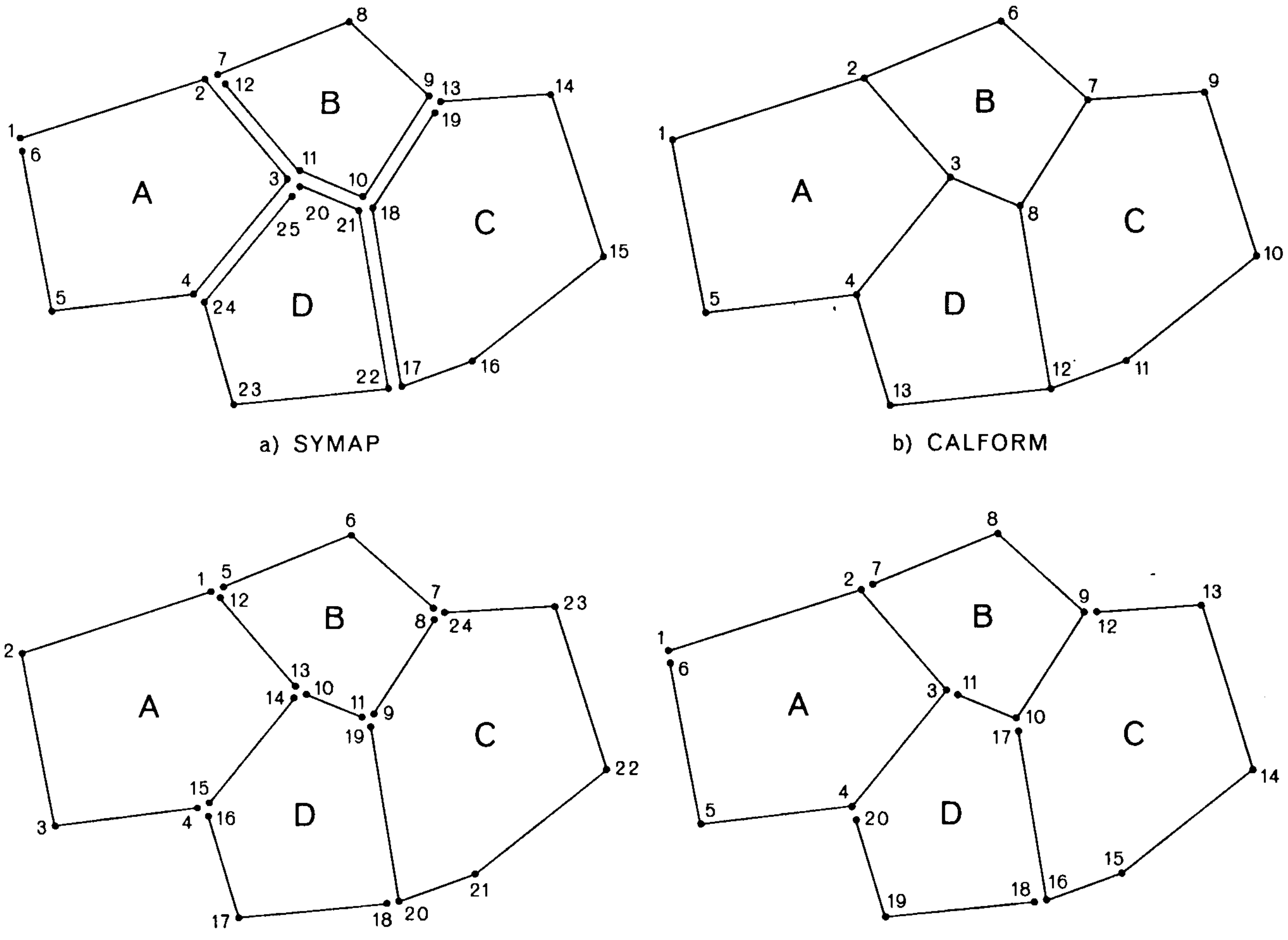
How Much Land is Class 2 With an Average Water Table Depth of 12-30 Inches?

The coverage 'SLG' contains a classification of the depth to water table, using the number 3 to indicate depths in the range 12-30 inches. So the question above requires a comparison of two maps and an evaluation of the area over which two classes coincide. The PLUS/2 'CROSSTAB' command will tabulate all coincidences between two coverages; it shows that of the 742 cells of Class 2 agricultural capability, 126 have the required class of water table depth (Figure 7). This corresponds to a total area of 31.5 square miles.

How Much Class 2 Land Lies Within 10 Miles of Stuartburn?

This query can be answered with the 'DTAB' command, which analyses a map according to distance from a fixed point. The location of Stuartburn is given in terms of the network of cells, as row 18, column 50. The user now indicates how the map is to be analysed. Each cell in the map will be identified in two ways, according to its type, and to its distance from Stuartburn. The user defines ranges of aistance, by indicating the upper limit of each range, and the symbol the range is to be given in the output table. Distances must be given in terms of cells, bearing in mind that each quarter-section cell is one half mile across. By specifying limits of 20 and 200 cells, the user is in effect indicating ranges of less than and greater than 10 miles, since no distances of greater than 200 cells occur on the map. From the output, Figure 7, the user infers that only 9 cells have the required characteristic, the majority of Class 2 land lying outside the ten mile radius.

Figure 1 Illustration of Alternative Digitizing Methods



a) SYMAP

b) CALFORM

c) PC

d) IC

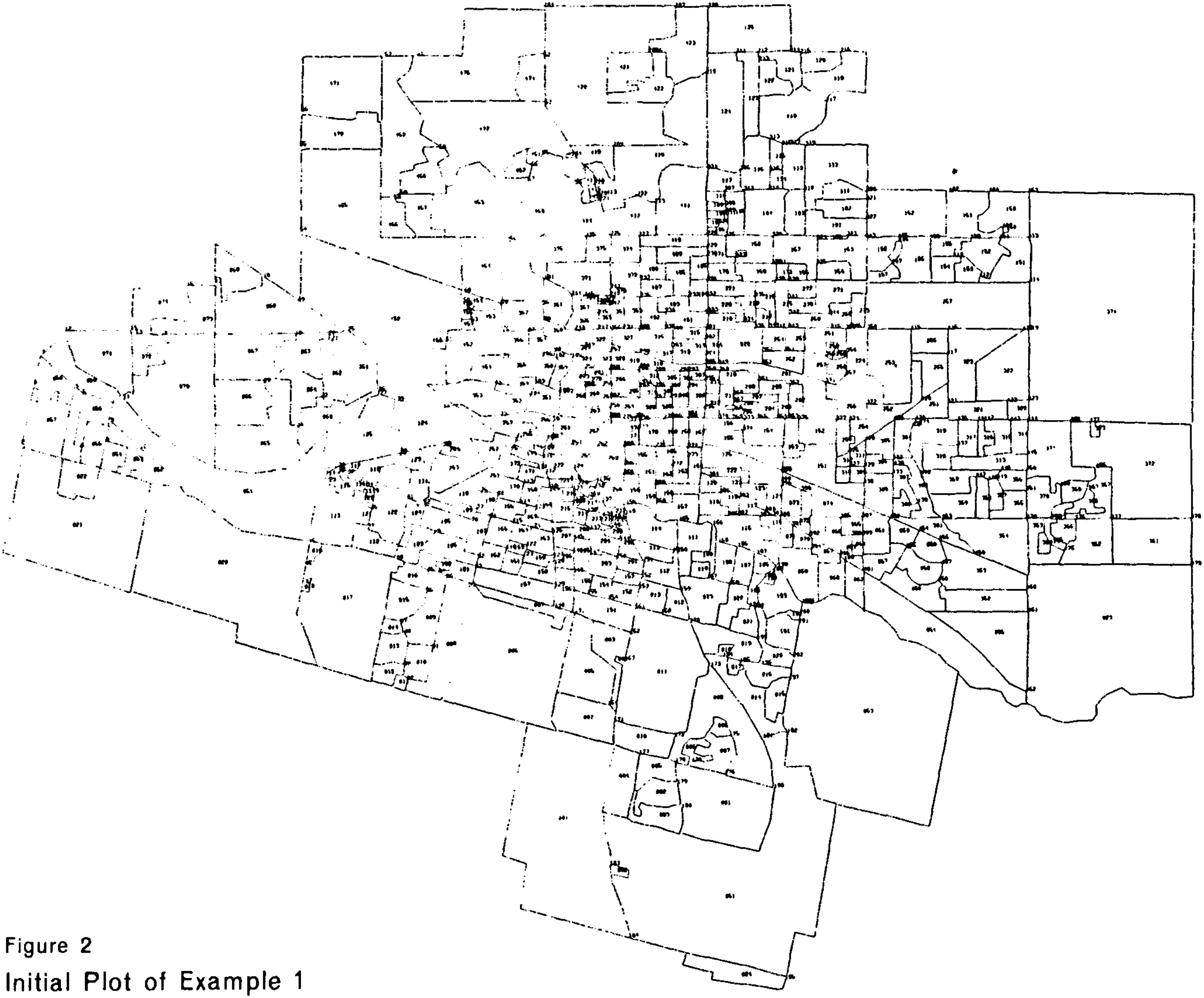


Figure 2
Initial Plot of Example 1

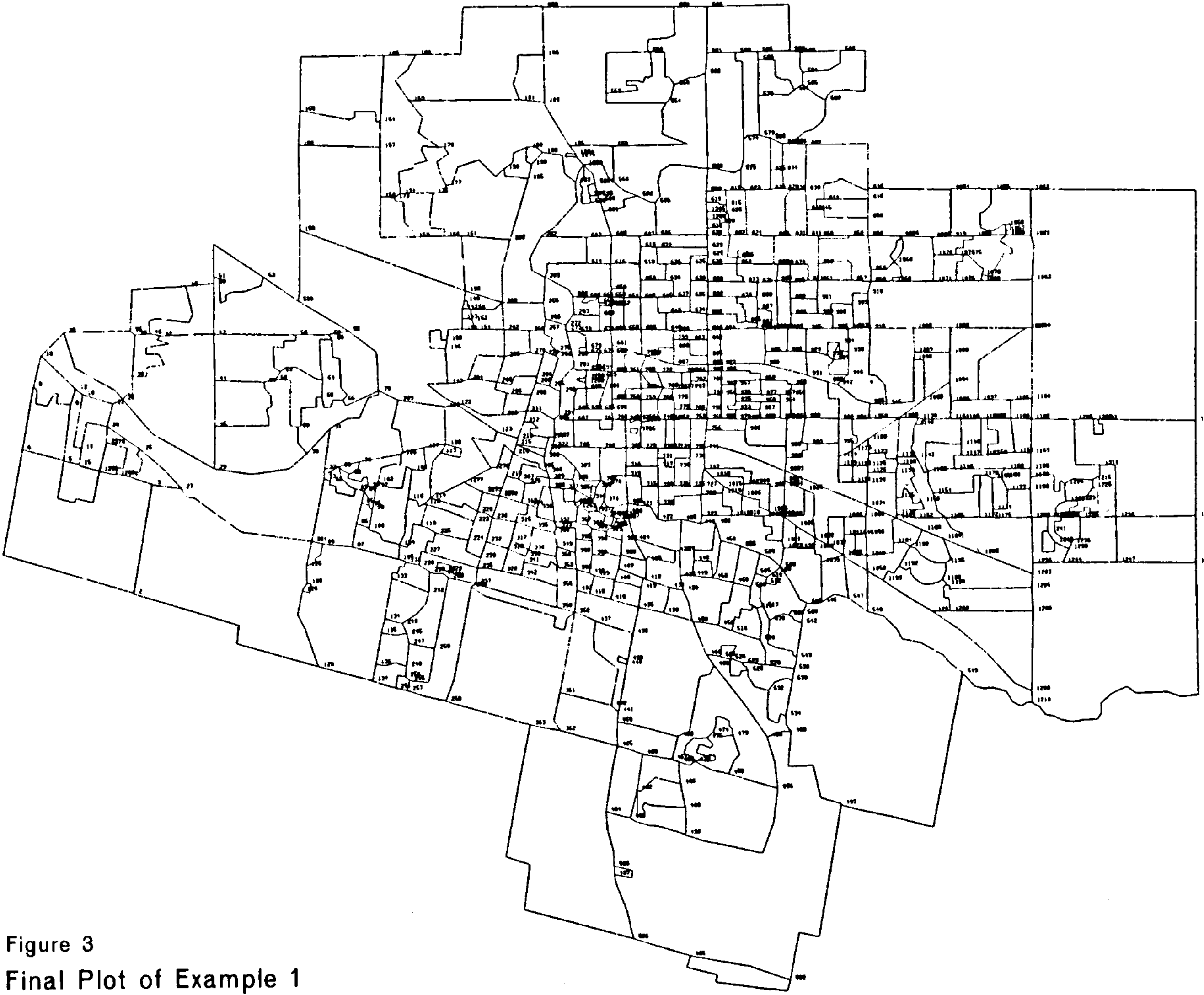


Figure 3

Final Plot of Example 1

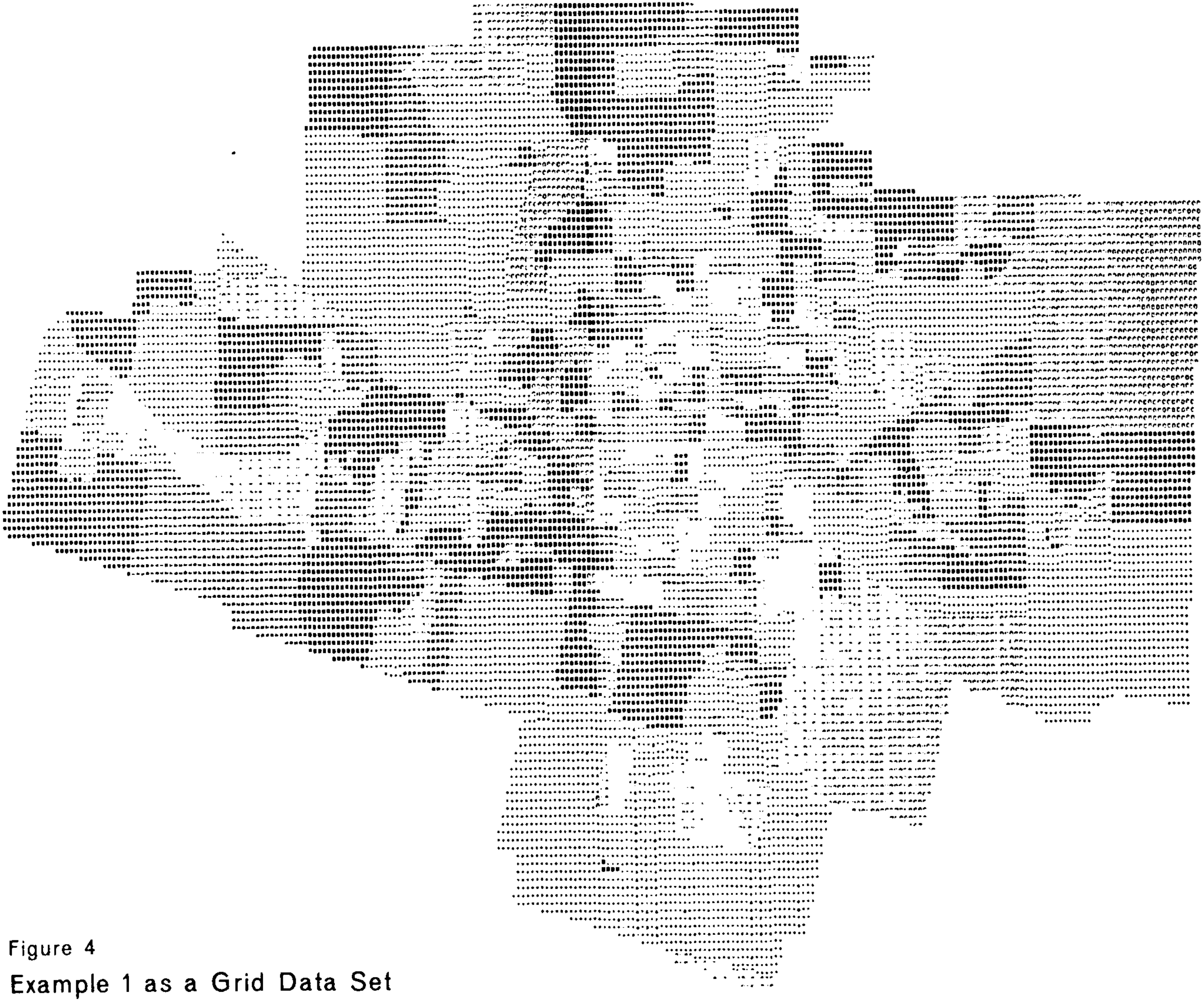
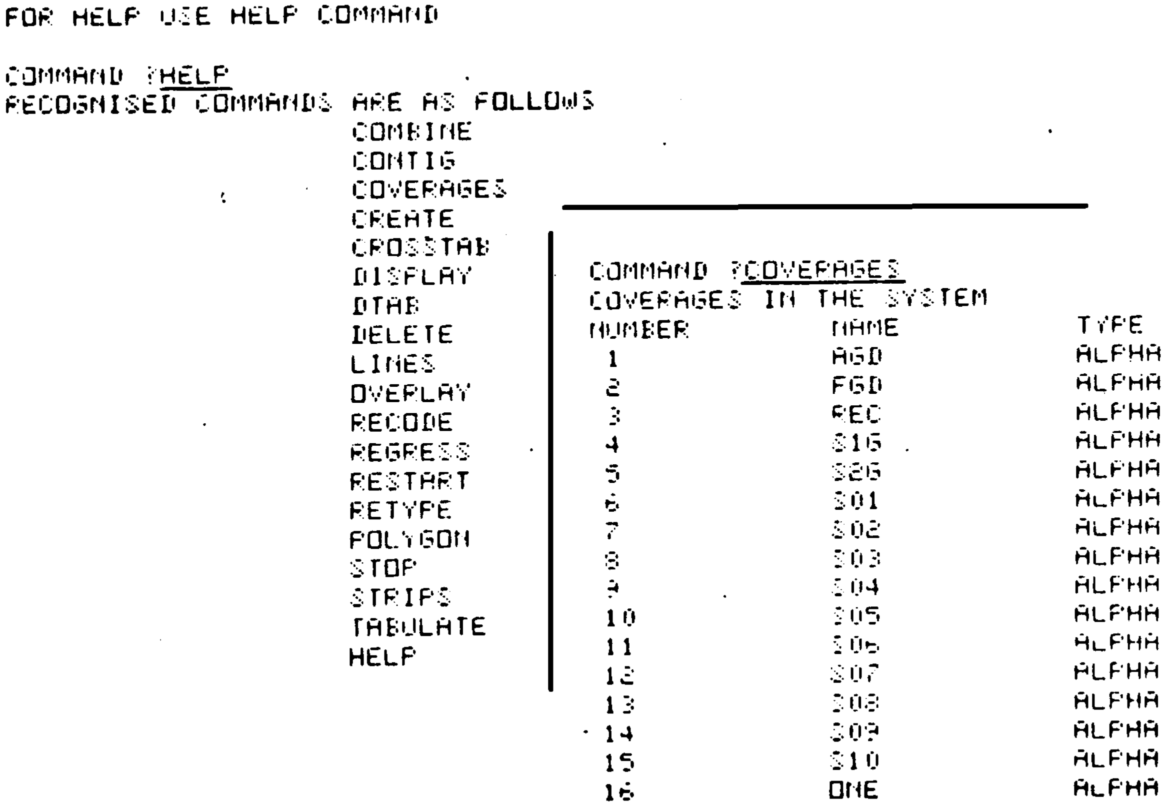


Figure 4

Example 1 as a Grid Data Set

FIGURE 5

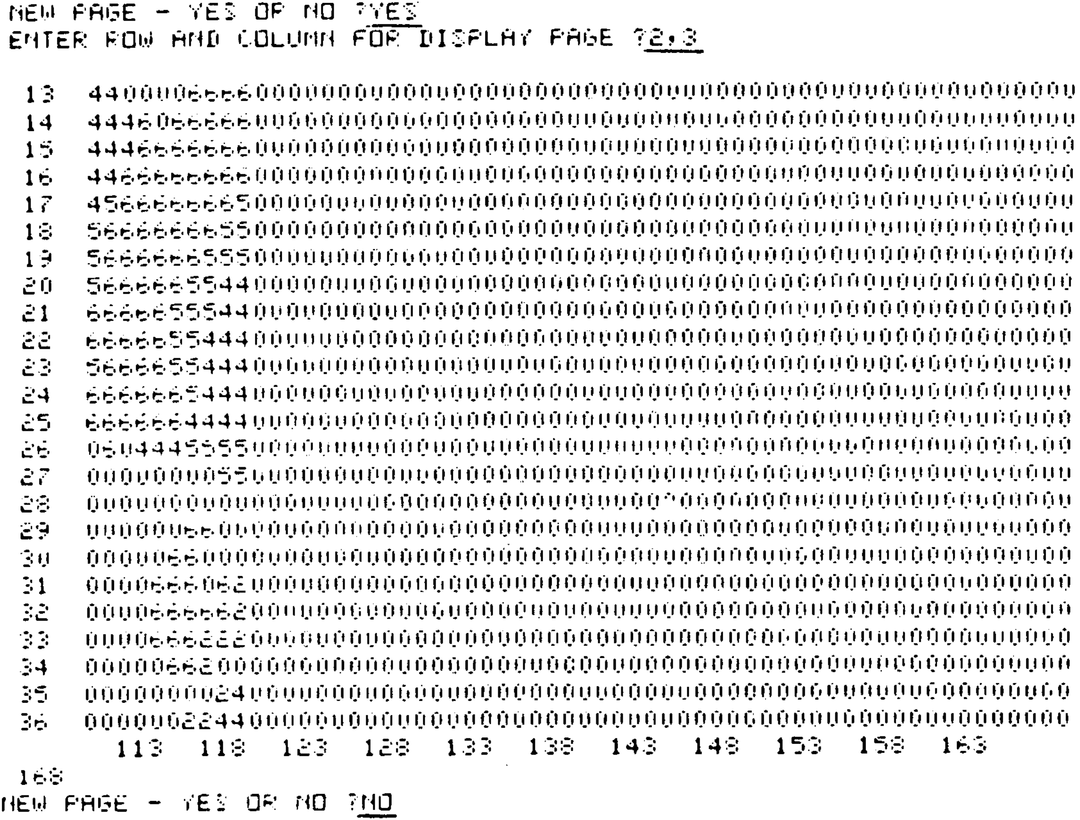
```
      PLUS/2 BASIC VERSION              (12 JUNE 74)
```



```
FOR HELP USE HELP COMMAND

COMMAND ?HELP
RECOGNISED COMMANDS ARE AS FOLLOWS
                   COMBINE
                   CONTIG
                   COVERAGES
                   CREATE
                   CROSSTAB
                   DISPLAY
                   DTAB
                   DELETE
                   LINES
                   OVERLAY
                   RECODE
                   REGRESS
                   RESTART
                   RETYPE
                   POLYGON
                   STOP
                   STRIPS
                   TABULATE
                   HELP
```

```
COMMAND ?COVERAGES
COVERAGES IN THE SYSTEM
NUMBER            NAME             TYPE
  1               AGD              ALPHA
  2               FGD              ALPHA
  3               REC              ALPHA
  4               S16              ALPHA
  5               S26              ALPHA
  6               S01              ALPHA
  7               S02              ALPHA
  8               S03              ALPHA
  9               S04              ALPHA
 10               S05              ALPHA
 11               S06              ALPHA
 12               S07              ALPHA
 13               S08              ALPHA
 14               S09              ALPHA
 15               S10              ALPHA
 16               ONE              ALPHA
```

FIGURE 6

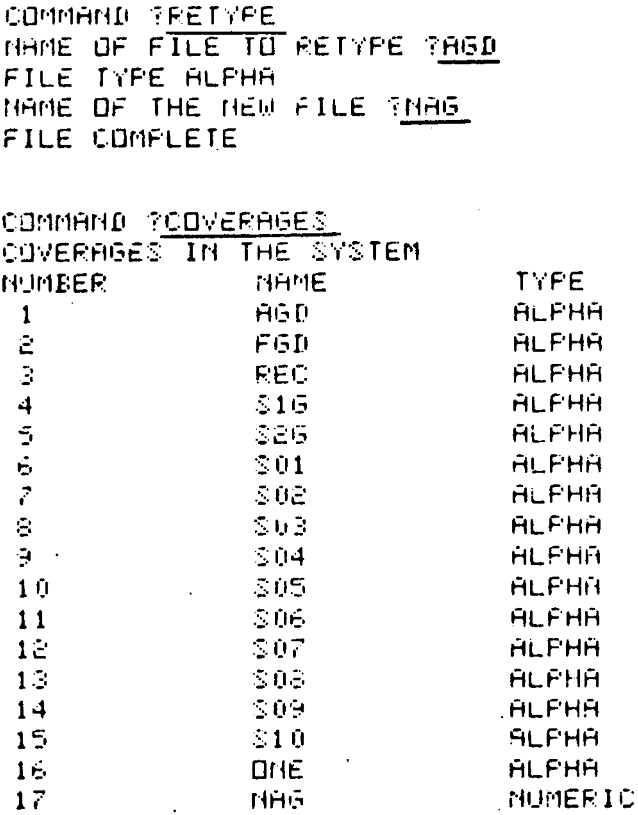


```
NEW PAGE - YES OR NO ?YES
ENTER ROW AND COLUMN FOR DISPLAY PAGE ?2,3

 13   440000666600000000000000000000000000000000000000000000000000
 14   444606666600000000000000000000000000000000000000000000000000
 15   444666666600000000000000000000000000000000000000000000000000
 16   446666666600000000000000000000000000000000000000000000000000
 17   456666666500000000000000000000000000000000000000000000000000
 18   566666665500000000000000000000000000000000000000000000000000
 19   566666655500000000000000000000000000000000000000000000000000
 20   566666554400000000000000000000000000000000000000000000000000
 21   666665554400000000000000000000000000000000000000000000000000
 22   666665544400000000000000000000000000000000000000000000000000
 23   566665544400000000000000000000000000000000000000000000000000
 24   666666544400000000000000000000000000000000000000000000000000
 25   666666444400000000000000000000000000000000000000000000000000
 26   060444555500000000000000000000000000000000000000000000000000
 27   000000005500000000000000000000000000000000000000000000000000
 28   000000000000000000000000000000000000000000000000000000000000
 29   000000066000000000000000000000000000000000000000000000000000
 30   000000660000000000000000000000000000000000000000000000000000
 31   000066606200000000000000000000000000000000000000000000000000
 32   000066666200000000000000000000000000000000000000000000000000
 33   000066622200000000000000000000000000000000000000000000000000
 34   000000662000000000000000000000000000000000000000000000000000
 35   000000002400000000000000000000000000000000000000000000000000
 36   000000224400000000000000000000000000000000000000000000000000
          113  118  123  128  133  138  143  148  153  158  163
 168
NEW PAGE - YES OR NO ?NO
```

FIGURE 7

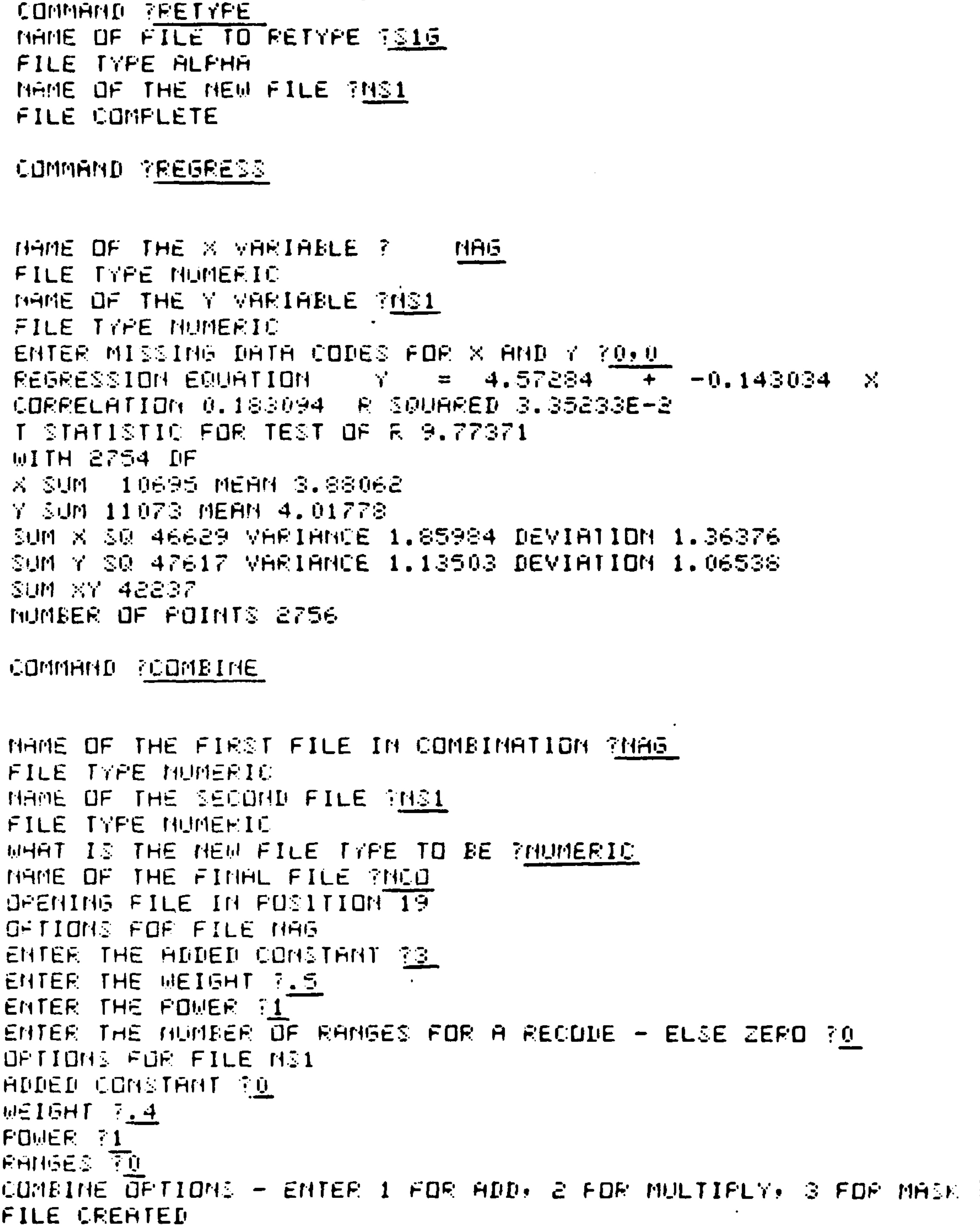
```
COMMAND ?TABULATE
NAME OF FILE TO BE TABULATED ?AGD
FILE TYPE ALPHA
TABULATING FILE TYPE ALPHA
LEVEL            NAME            TOTAL CELLS     PERCENT
  1               2                  742            12.2685
  2               3                  596            9.8545
  3               5                  1158           19.1468
  4               4                  587            9.70569
  5               0                  2609           43.1382
  6               6                  356            5.88624

COMMAND ?CROSSTAB


NAME OF FIRST FILE IN CROSSTAB ?AGD
FILE TYPE ALPHA
NAME OF SECOND FILE ?SIG
FILE TYPE ALPHA
CROSSTABULATION - AGD IN ROWSSIG IN COLUMNS
       3               4       5       1       2
2       215     74      67      372     8       6       742
3       106     126     162     197     1       4       596
5       262     227     48      500     3       118     1158
4       104     173     74      213     3       20      587
0       603     586     377     222     0       821     2609
6       61      83      98      28      0       86      356

        1351    1269    826     1532    15      1055    6048


CHISQUARE STATISTIC 1668.75 WITH 25 DF

COMMAND ?DTAB

NAME OF FILE FOR DISTANCE TABULATION ?AGD
FILE TYPE ALPHA
DISTANCES ARE IN GEOGRAPHIC UNITS. NUMBER OF RANGES ?2
SPECIFY THE UPPER LIMITS OF EACH RANGE, THEN THE
NEW SYMBOL OF VALUE FOR EACH RANGE
LIMIT AND SYMBOL OF RANGE 0   ?20,<10
LIMIT AND SYMBOL OF RANGE 1   ?200,>10
ENTER ORIGIN POINT AT ROW AND COLUMN NUMBER ?18,50
CROSSTABULATION - DISTANCES IN COLUMNS, COVERAGE IN ROWS
 .               <10   >10
2                  9     733   742
3                  142   454   596
5                  483   675   1158
4                  402   185   587
0                  156   2453   2609
6                  10    346   356
                1202  4846
```

FIGURE 8



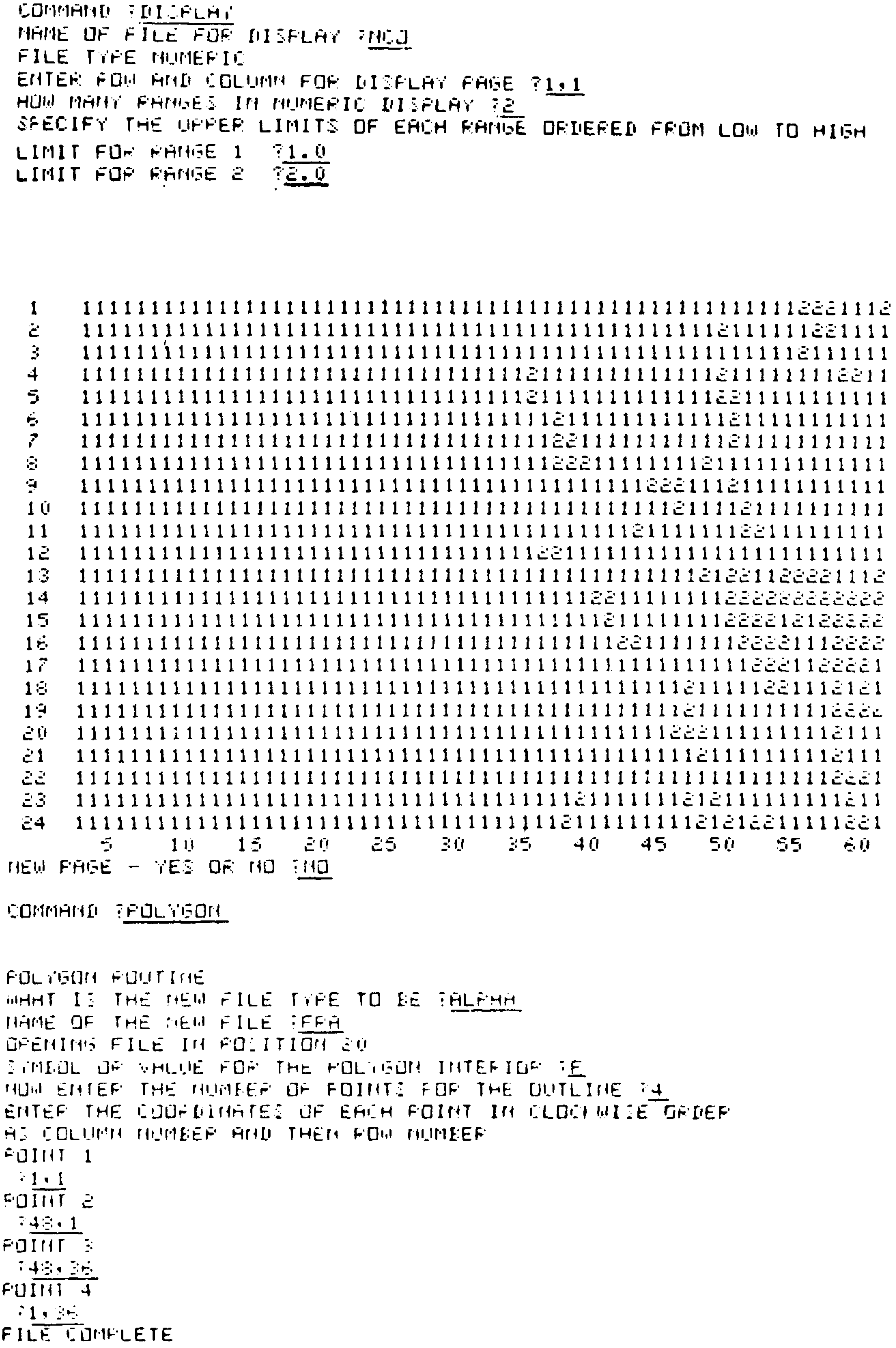
```
COMMAND ?RETYPE
NAME OF FILE TO RETYPE ?AGD
FILE TYPE ALPHA
NAME OF THE NEW FILE ?NAG
FILE COMPLETE


COMMAND ?COVERAGES
COVERAGES IN THE SYSTEM
NUMBER             NAME              TYPE
 1                 AGD               ALPHA
 2                 FGD               ALPHA
 3                 REC               ALPHA
 4                 S16               ALPHA
 5                 S26               ALPHA
 6                 S01               ALPHA
 7                 S02               ALPHA
 8                 S03               ALPHA
 9                 S04               ALPHA
10                 S05               ALPHA
11                 S06               ALPHA
12                 S07               ALPHA
13                 S08               ALPHA
14                 S09               ALPHA
15                 S10               ALPHA
16                 ONE               ALPHA
17                 NAG               NUMERIC
```

FIGURE 9



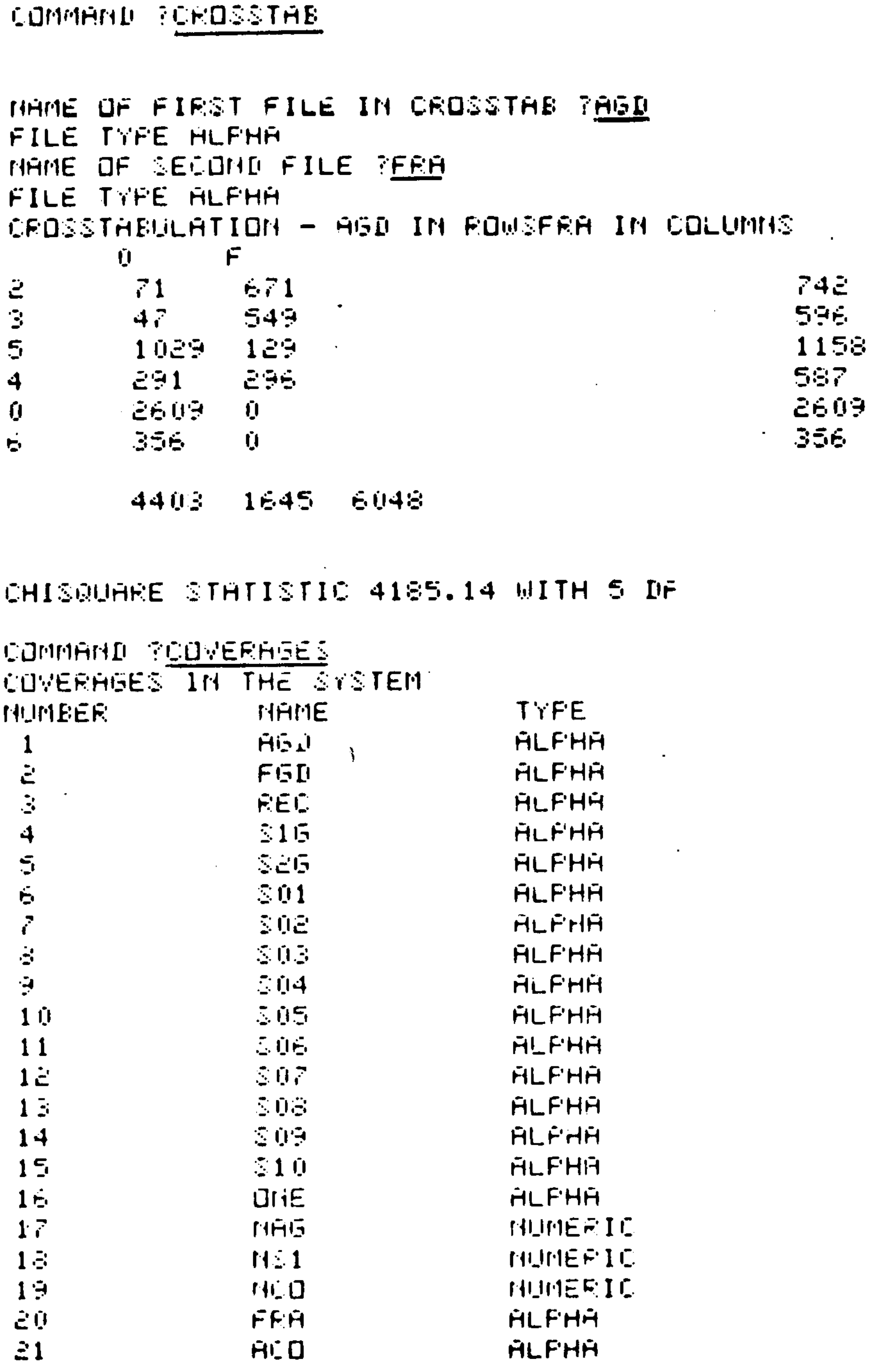
```
COMMAND ?RETYPE
NAME OF FILE TO RETYPE ?S1G
FILE TYPE ALPHA
NAME OF THE NEW FILE ?NS1
FILE COMPLETE

COMMAND ?REGRESS


NAME OF THE X VARIABLE ?    NAG
FILE TYPE NUMERIC
NAME OF THE Y VARIABLE ?NS1
FILE TYPE NUMERIC
ENTER MISSING DATA CODES FOR X AND Y ?0,0
REGRESSION EQUATION    Y   =  4.57284   +  -0.143034  X
CORRELATION 0.183094  R SQUARED 3.35233E-2
T STATISTIC FOR TEST OF R 9.77371
WITH 2754 DF
X SUM  10695 MEAN 3.88062
Y SUM 11073 MEAN 4.01778
SUM X SQ 46629 VARIANCE 1.85984 DEVIATION 1.36376
SUM Y SQ 47617 VARIANCE 1.13503 DEVIATION 1.06538
SUM XY 42237
NUMBER OF POINTS 2756

COMMAND ?COMBINE


NAME OF THE FIRST FILE IN COMBINATION ?NAG
FILE TYPE NUMERIC
NAME OF THE SECOND FILE ?NS1
FILE TYPE NUMERIC
WHAT IS THE NEW FILE TYPE TO BE ?NUMERIC
NAME OF THE FINAL FILE ?NCO
OPENING FILE IN POSITION 19
OPTIONS FOR FILE NAG
ENTER THE ADDED CONSTANT ?3
ENTER THE WEIGHT ?.5
ENTER THE POWER ?1
ENTER THE NUMBER OF RANGES FOR A RECODE - ELSE ZERO ?0
OPTIONS FOR FILE NS1
ADDED CONSTANT ?0
WEIGHT ?.4
POWER ?1
RANGES ?0
COMBINE OPTIONS - ENTER 1 FOR ADD, 2 FOR MULTIPLY, 3 FOR MASK
FILE CREATED
```

FIGURE 10



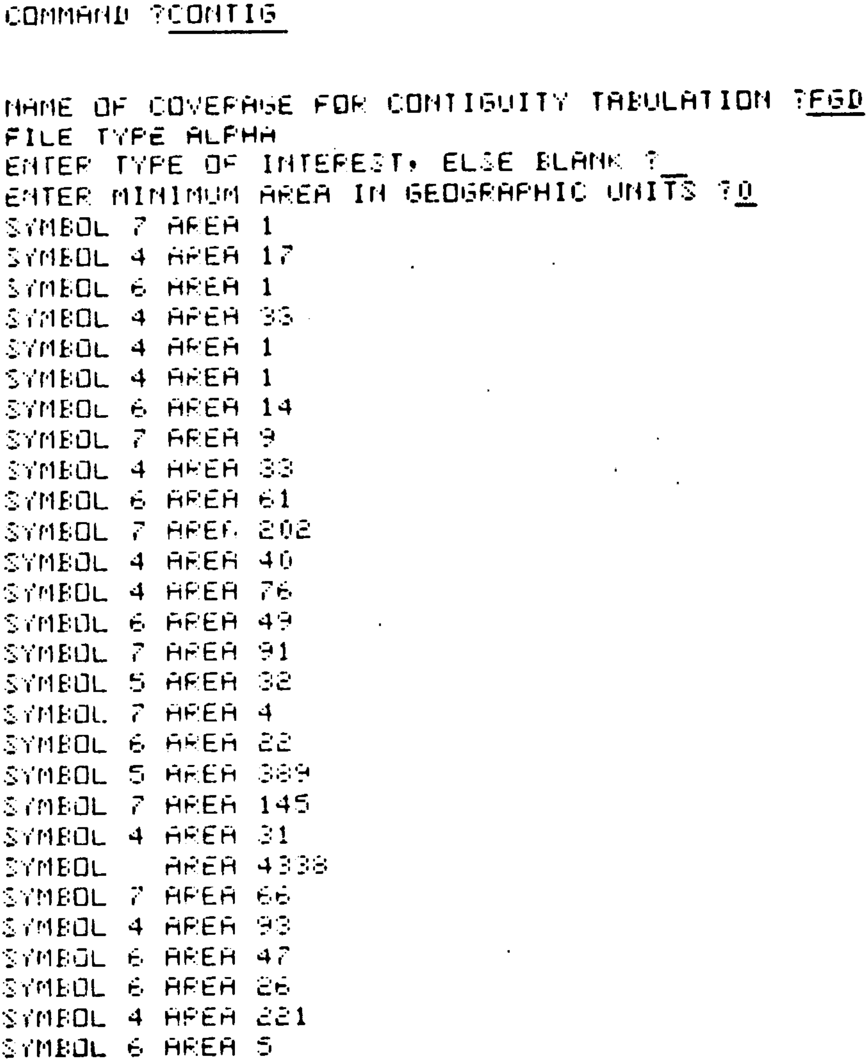
```
COMMAND ?DISPLAY
NAME OF FILE FOR DISPLAY ?NCJ
FILE TYPE NUMERIC
ENTER ROW AND COLUMN FOR DISPLAY PAGE ?1,1
HOW MANY RANGES IN NUMERIC DISPLAY ?2
SPECIFY THE UPPER LIMITS OF EACH RANGE ORDERED FROM LOW TO HIGH
LIMIT FOR RANGE 1  ?1.0
LIMIT FOR RANGE 2  ?2.0



  1   111111111111111111111111111111111111111111111111111112221112
  2   111111111111111111111111111111111111111111111112111111221111
  3   111111111111111111111111111111111111111111111111111112111111
  4   111111111111111111111111111111111211111111111112111111112211
  5   111111111111111111111111111111111211111111111112211111111111
  6   111111111111111111111111111111111112111111111111211111111111
  7   111111111111111111111111111111111112211111111111211111111111
  8   111111111111111111111111111111111112221111111121111111111111
  9   111111111111111111111111111111111111111111222111211111111111
 10   111111111111111111111111111111111111111111112111121111111111
 11   111111111111111111111111111111111111111112111111122111111111
 12   111111111111111111111111111111111122111111111111111111111111
 13   111111111111111111111111111111111111111111111121221122221112
 14   111111111111111111111111111111111111112211111111222222222222
 15   111111111111111111111111111111111111111211111111222212122222
 16   111111111111111111111111111111111111111122111111122221112222
 17   111111111111111111111111111111111111111111111111112221122221
 18   111111111111111111111111111111111111111111111211111221112121
 19   111111111111111111111111111111111111111111111211111111112222
 20   111111111111111111111111111111111111111111112221111111112111
 21   111111111111111111111111111111111111111111111121111111112111
 22   111111111111111111111111111111111111111111111111111111112221
 23   111111111111111111111111111111111111121111111212111111111211
 24   111111111111111111111111111111111111211111111121212211111221
          5    10   15   20   25   30   35   40   45   50   55   60
NEW PAGE - YES OR NO ?NO

COMMAND ?POLYGON


POLYGON ROUTINE
WHAT IS THE NEW FILE TYPE TO BE ?ALPHA
NAME OF THE NEW FILE ?PPA
OPENING FILE IN POSITION 20
SYMBOL OR VALUE FOR THE POLYGON INTERIOR ?E
NOW ENTER THE NUMBER OF POINTS FOR THE OUTLINE ?4
ENTER THE COORDINATES OF EACH POINT IN CLOCKWISE ORDER
AS COLUMN NUMBER AND THEN ROW NUMBER
POINT 1
 ?1,1
POINT 2
 ?48,1
POINT 3
 ?48,36
POINT 4
 ?1,36
FILE COMPLETE
```

FIGURE 11



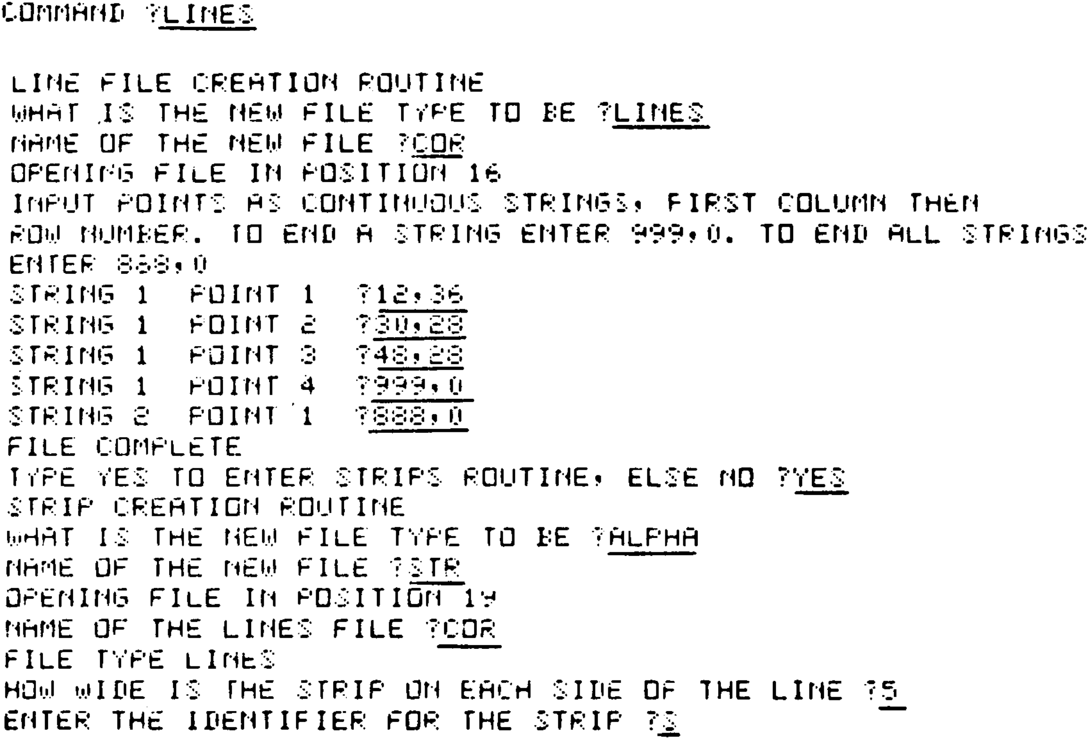
```
COMMAND ?CROSSTAB


NAME OF FIRST FILE IN CROSSTAB ?AGD
FILE TYPE ALPHA
NAME OF SECOND FILE ?FRA
FILE TYPE ALPHA
CROSSTABULATION - AGD IN ROWSFRA IN COLUMNS
        0      F
2        71     671                                         742
3        47     549                                         596
5        1029   129                                         1158
4        291    296                                         587
0        2609   0                                           2609
6        356    0                                           356

         4403   1645   6048


CHISQUARE STATISTIC 4185.14 WITH 5 DF

COMMAND ?COVERAGES
COVERAGES IN THE SYSTEM
NUMBER              NAME               TYPE
 1                  AGD                ALPHA
 2                  FGD                ALPHA
 3                  REC                ALPHA
 4                  S1G                ALPHA
 5                  S2G                ALPHA
 6                  S01                ALPHA
 7                  S02                ALPHA
 8                  S03                ALPHA
 9                  S04                ALPHA
 10                 S05                ALPHA
 11                 S06                ALPHA
 12                 S07                ALPHA
 13                 S08                ALPHA
 14                 S09                ALPHA
 15                 S10                ALPHA
 16                 ONE                ALPHA
 17                 NAG                NUMERIC
 18                 NS1                NUMERIC
 19                 NCO                NUMERIC
 20                 FRA                ALPHA
 21                 ACO                ALPHA
```

FIGURE 12



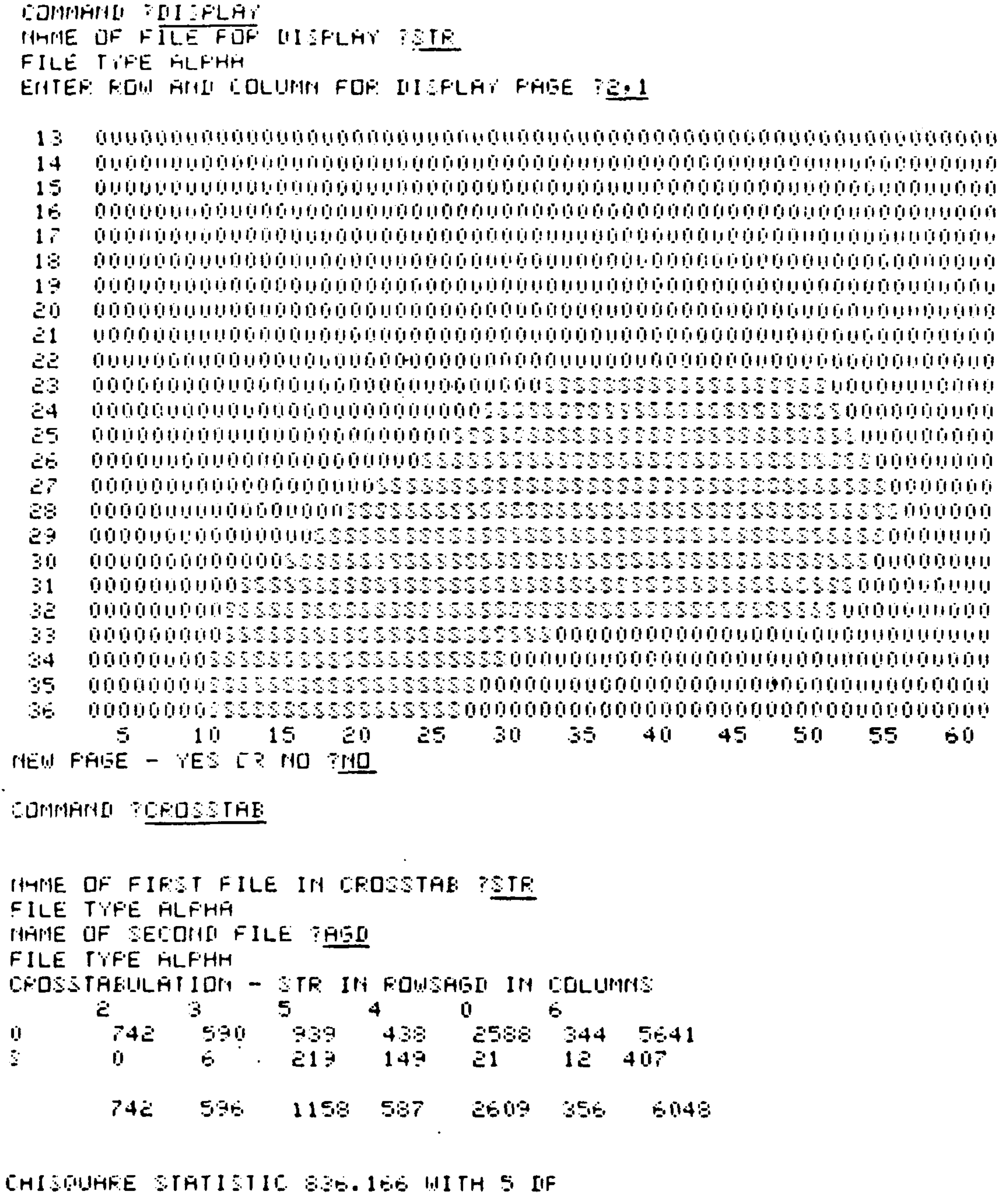
```
COMMAND ?CONTIG


NAME OF COVERAGE FOR CONTIGUITY TABULATION ?FGD
FILE TYPE ALPHA
ENTER TYPE OF INTEREST, ELSE BLANK ?_
ENTER MINIMUM AREA IN GEOGRAPHIC UNITS ?0
SYMBOL 7 AREA 1
SYMBOL 4 AREA 17
SYMBOL 6 AREA 1
SYMBOL 4 AREA 33
SYMBOL 4 AREA 1
SYMBOL 4 AREA 1
SYMBOL 6 AREA 14
SYMBOL 7 AREA 9
SYMBOL 4 AREA 33
SYMBOL 6 AREA 61
SYMBOL 7 AREA 202
SYMBOL 4 AREA 40
SYMBOL 4 AREA 76
SYMBOL 6 AREA 49
SYMBOL 7 AREA 91
SYMBOL 5 AREA 32
SYMBOL 7 AREA 4
SYMBOL 6 AREA 22
SYMBOL 5 AREA 389
SYMBOL 7 AREA 145
SYMBOL 4 AREA 31
SYMBOL   AREA 4338
SYMBOL 7 AREA 66
SYMBOL 4 AREA 93
SYMBOL 6 AREA 47
SYMBOL 6 AREA 26
SYMBOL 4 AREA 221
SYMBOL 6 AREA 5
```

FIGURE 13



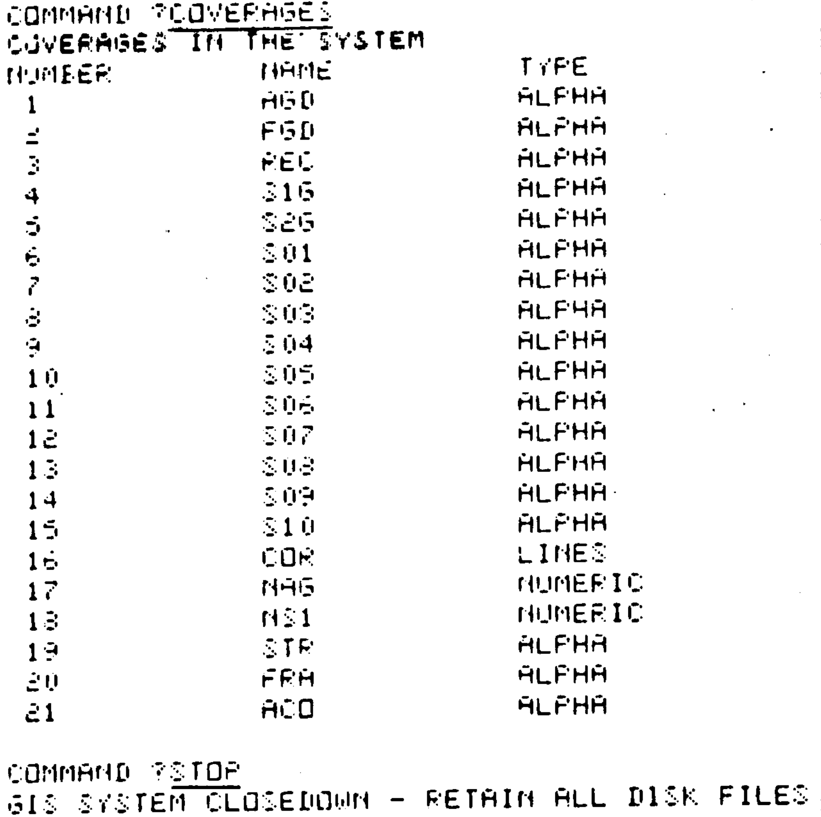
```
COMMAND ?LINES


 LINE FILE CREATION ROUTINE
 WHAT IS THE NEW FILE TYPE TO BE ?LINES
 NAME OF THE NEW FILE ?COR
 OPENING FILE IN POSITION 16
 INPUT POINTS AS CONTINUOUS STRINGS, FIRST COLUMN THEN
 ROW NUMBER. TO END A STRING ENTER 999,0. TO END ALL STRINGS
 ENTER 888,0
 STRING 1  POINT 1  ?12,36
 STRING 1  POINT 2  ?30,28
 STRING 1  POINT 3  ?48,28
 STRING 1  POINT 4  ?999,0
 STRING 2  POINT 1  ?888,0
 FILE COMPLETE
 TYPE YES TO ENTER STRIPS ROUTINE, ELSE NO ?YES
 STRIP CREATION ROUTINE
 WHAT IS THE NEW FILE TYPE TO BE ?ALPHA
 NAME OF THE NEW FILE ?STR
 OPENING FILE IN POSITION 19
 NAME OF THE LINES FILE ?COR
 FILE TYPE LINES
 HOW WIDE IS THE STRIP ON EACH SIDE OF THE LINE ?5
 ENTER THE IDENTIFIER FOR THE STRIP ?S
```

FIGURE 14



```
COMMAND ?DISPLAY
NAME OF FILE FOR DISPLAY ?STR
FILE TYPE ALPHA
ENTER ROW AND COLUMN FOR DISPLAY PAGE ?2,1

 13  000000000000000000000000000000000000000000000000000000000000
 14  000000000000000000000000000000000000000000000000000000000000
 15  000000000000000000000000000000000000000000000000000000000000
 16  000000000000000000000000000000000000000000000000000000000000
 17  000000000000000000000000000000000000000000000000000000000000
 18  000000000000000000000000000000000000000000000000000000000000
 19  000000000000000000000000000000000000000000000000000000000000
 20  000000000000000000000000000000000000000000000000000000000000
 21  000000000000000000000000000000000000000000000000000000000000
 22  000000000000000000000000000000000000000000000000000000000000
 23  000000000000000000000000000000SSSSSSSSSSSSSSSSSSS00000000000
 24  00000000000000000000000000SSSSSSSSSSSSSSSSSSSSSSSS0000000000
 25  000000000000000000000000SSSSSSSSSSSSSSSSSSSSSSSSSSS000000000
 26  0000000000000000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSS00000000
 27  0000000000000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS0000000
 28  00000000000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS000000
 29  000000000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS0000000
 30  0000000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS00000000
 31  0000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS000000000
 32  000000000SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS0000000000
 33  000000000SSSSSSSSSSSSSSSSSSSSSS00000000000000000000000000000
 34  00000000SSSSSSSSSSSSSSSSSSSS00000000000000000000000000000000
 35  00000000SSSSSSSSSSSSSSSSSS0000000000000000000000000000000000
 36  00000000SSSSSSSSSSSSSSSSS00000000000000000000000000000000000
       5   10   15   20   25   30   35   40   45   50   55   60
NEW PAGE - YES OR NO ?NO

COMMAND ?CROSSTAB


NAME OF FIRST FILE IN CROSSTAB ?STR
FILE TYPE ALPHA
NAME OF SECOND FILE ?AGD
FILE TYPE ALPHA
CROSSTABULATION - STR IN ROWSAGD IN COLUMNS
      2      3      5      4      0      6
0       742    590    939    438    2588   344   5641
S       0      6      219    149    21     12   407

        742    596    1158   587    2609   356    6048


CHISQUARE STATISTIC 836.166 WITH 5 DF
```

FIGURE 15



```
COMMAND ?COVERAGES
COVERAGES IN THE SYSTEM
NUMBER            NAME             TYPE
 1                AGD              ALPHA
 2                FGD              ALPHA
 3                REC              ALPHA
 4                S1G              ALPHA
 5                S2G              ALPHA
 6                S01              ALPHA
 7                S02              ALPHA
 8                S03              ALPHA
 9                S04              ALPHA
 10               S05              ALPHA
 11               S06              ALPHA
 12               S07              ALPHA
 13               S08              ALPHA
 14               S09              ALPHA
 15               S10              ALPHA
 16               COR              LINES
 17               NAG              NUMERIC
 18               NS1              NUMERIC
 19               STR              ALPHA
 20               FRA              ALPHA
 21               ACO              ALPHA

COMMAND ?STOP
GIS SYSTEM CLOSEDOWN - RETAIN ALL DISK FILES
```

Can Agricultural Capability Be Predicted From Water Table Depth?

Neither 'AGD' nor 'SlG' were initially declared as Numeric files, although both contain exclusively numeric symbols, which in turn represent values on crude scales of capability and depth respectively. So they both satisfy the requirements of Numeric files, and their types are changed using the 'RETYPE' command to form files 'NAG' and 'NSl'. The 'REGRESS' command can now be used to test for a predictive relationship (Figure 9). Any cells containing the code '0' in either file are omitted, since the code denotes missing data, leaving 2,756 cases to be evaluated by linear regression. The correlation of 0.183 indicates that a very weak relationship is present, with good capability (low values) corresponding to shallow water tables, but that many other factors also affect the capability index.

Yield Prediction

Suppose that past analyses have indicated that hay yields, in thousands of pounds per acre, can be predicted from the equation

$$\text{Yield} = 3 + 0.5 + 0.4\ \text{SlG}$$

The 'COMBINE' function can be used to produce a new coverage in which each cell shows the combined yield prediction from the capability and water depth coverages. The options in the command allow a wide variety of algebraic and logical combinations. See Figure 9.

How Much Class 2 Land Lies in the Municipal District of Franklin?

The outline of the District is used to make a coverage 'FRA' showing the critical area with the symbol F and the rest of the map as O, by invoking the 'POLYGON' command. A crosstabulation of 'FRA' with 'AGD' then shows that 671 cells have the required capability class. This and other new coverages created during the session now appear in the 'COVERAGES' list with appropriate types. (See Figure 11.)

What is the Largest Continuous Block C1 Class 4 Forestry Land?

The tabulations above have paid no attention to continguity, so that a total of 100 cells may exist as a continuous tract, or as 100 small fragments. The system can produce summaries of contiguous areas through the 'CONTIG' command. An analysis of the map of forestry capability, 'FGD', shows that 221 cells, or 54.25 square miles are available in one unit. (See Figure 12.)

How Much Class 2 Land

Lies Within 2.5 Miles of the CNR Right of Way?

This is answered in three steps. First, the location of the right of way is supplied to the system through the 'LINES' command, by giving the locations of points at which the lines in the network change direction. Then the 'STRIPS' command creates a coverage based on the Lines file by distinguishing the cells within a critical distance of the network by a particular symbol. Finally, a crosstabulation of this new coverage with 'AGD' shows that there is no Class 2 land within the critical strip. (See Figure 14.)

CONCLUSION AND SUMMARY

The second example illustrated the use of a geographic information system to answer a set of queries that would be largely impossible by the manual analysis of mapped information. The kind of system exemplified by PLUS/2 is capable of providing answers to complex questions both rapidly and cheaply, using data stored in grid form. While it is possible in principle to perform the same operations on polygon data, the computer processing times and costs are much greater and far outweigh the corresponding increase in accuracy. Furthermore, it is doubtful if a user could cope with the volume of data produced in an analysis of polygon data sets. The size of a grid cell in a grid analysis can be adjusted to provide the level of resolution and generalisation appropriate to a particular study, whereas polygon data is constrained to a constant, high level of precision.

The major expense of the system described here, and for that matter of all information systems, lies in the collection preparation and maintenance of the data base. This paper has been written from the planning point of view, on the assumption that the agency using such a system has no explicit responsibility for the acquisition of any particular kind of data, but rather is concerned with making the best use of geographic data sets maintained by other agencies, such as Statistics Canada or the Canada Land Inventory. As such, the final objective of the system must be the ability to respond to queries such as those in the second example; data storage and cartography are not likely to be major objectives by comparison. The selection of data structures is thus dictated by the need for grid form of levels of accuracy can be decided in advance. But more frequently a planning study will require various levels of accuracy as it moves from general studies of an area to detailed examination of critical zones. In such cases it is appropriate that data be first encoded in polygon form, and then overlaid with various grid cell schemes as necessary.

These arguments can be summarised in a scenario for a typical study. The area to be studied is first delineated, and enquiries made to determine the amount of data already available in various agency data banks. Such data is likely to be of polygon form, since agencies with a responsibility for acquiring data will usually avoid any loss of resolution in the encoding process. Additional data will be needed, besides that available in data banks, and must be encoded from maps. To avoid a premature choice of a level of resolution, such data is best encoded in polygon form, the precise method depending on the equipment available, following the arguments made earlier in the paper.

An initial level of resolution is now determined, and all available data gridded at that level to form a data base for analysis. If the accuracy must be changed later to permit a detailed analysis of part of the study area, the polygon data sets can be regridded with no difficulty. Initially, more information will be needed as the study progresses and should be acquired in grid form if needed for one level of resolution, or on polygon form if required more generally.

Several current trends in the computer hardware industry are likely to affect the area of geographic information systems in the near future, in some cases by altering these conclusions, in others by reinforcing them. First, the cost per operation is likely to continue to drop, along with the cost per unit of central memory. The effect will be to improve the feasibility of automatic polygon data. This should encourage the maintenance of large polygon data sets, which will relieve planning agencies of much of the responsibility for data collection for planning studies. Secondly, the introduction of new forms of solid state circuitry at vastly lower cost and smaller size is leading to the introduction of parallel-processing systems, which can perform many similar operations simultaneously. This trend is of particular importance to geographic problems, in which large arrays must be processed with highly repetitive operations. Finally, although geographic arrays are large, processing is basically sequential in many operations, such as those performed by PLUS/2. Many of the features of large computer systems (such as extensive core memory and direct access disk) tend to be unnecessary, and indeed PLUS/2 can be operated efficiently in a small mini-computer system with a fast central processor and a sequential access disk but very little core.

This paper has identified the major issues in geographic data encoding and processing from the viewpoint of applications in the planning field. Geographic data processing is now entering a rather lengthy phase of demonstration and application. Planners, not computer technicians, must be made more aware of the possibilities it offers, through more efficient responses to geographic questions and through the new kinds of geographic analysis that it permits.