A COMPUTER-ASSISTED METHOD FOR DERIVING SOIL MAPS OF VIRGINIA COUNTIES

by

Jerry Wayne Ziewitz

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science

in

Fisheries and Wildlife Sciences

Approved:

R. H. Giles, Jr., Chairman

E. W. Carson

L. A. Nielsen

October, 1982

Blacksburg, Virginia

A COMPUTER-ASSISTED METHOD FOR DERIVING SOIL MAPS OF VIRGINIA COUNTIES

by

Jerry Wayne Ziewitz

(ABSTRACT)

Procedures were developed for a computer-based geographic information system to map the probable soils of a county in southwestern Virginia having no soil map. To facilitate the association of known soil characteristics and interpretations with derived soil units, the procedures were based on a soil map available for a similar, mapped county in the region. Using a grid-cell size of 1/81 km^2 (1.2 ha or 3.0 acres), data were collected for geology, slope, aspect, topographic shape, distance to ridge, and distance to stream in both counties, and for soils in the county having a soil map.

A hierarchical, monothetic, divisive classification program was written to combine these data and produce a dichotomous key to soil types. About 36,000 cells belonging to 26 soil types were classified with 55 percent accuracy using data available for the entire study area. Twenty-

three of the 26 types could not be distinguished in the classification. Using more detailed data available for an 11 km^2 (4.3 sq mi) subarea, 900 cells belonging to 19 soil types were classified with 69 percent accuracy. Eleven soil types in the subarea could not be distinguished in the classification. High variability within some of the larger soil types mapped in the study area and inaccuracies in the topographic data were the most probable causes of the apparent inability to classify many of the soil types, and the results were not applied to the mapped county.

ACKNOWLEDGEMENTS

I thank Dr. Robert H. Giles, Jr., for his guidance and counsel. He set me on a path I shall probably follow the rest of my professional career. I thank Drs. L. A. Nielsen and E. W. Carson for serving as committee members and for reviewing this manuscript.

Many people assisted in the collection of data. I am gratefull to Bill Burks, Steve Cherry, Dave Graham, Betsy McGraf, and Leslie Mehring. I thank Charles Stanley of the Divsion of Mineral Resources for providing his personal copy of a geologic survey.

I was fortunate to have two mentors in computer programming and geographic data handling: Dr. Gregory T. Koeln and A. Blair Jones, III. Without their patient, day-to-day advice and assistance, my work would have been immensely more difficult. They taught me well, and I am grateful.

I owe a special thanks to Kevin Jordan and Greg Kroll for their help and commiseration. I thank John Keith and Lisie Kitchel for their generosity and hospitality during my last month in Blacksburg. Finally, I thank Kathryn, my wife, for her assistance in the preparation of this manuscript and for her support of my work in general. She has been an inspiration to me, without which, my work would not have been quite as rewarding.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

LIST OF APPENDICES x

INTRODUCTION 1

LITERATURE REVIEW 6

Geographic Information Systems 6
Applications of Soil Data in Geographic Information Systems 8
Soil Data for Unsurveyed Areas 10

METHODS 14

Overview 14
The Study Area 21
Partitioning the Study Area 23
Data Collection 27
Soil 28
Source of Data 28
Method of Digitizing 30
Geology 33
Source of Data 33
Method of Digitizing 38
Elevation 38
Attributes Derived from Elevation 43
Streams 46
The Classification Program DIVIDE 54
General Design of DIVIDE 55
Input 58
The Classification Process 58
Derivation of _H 60
Special Data Requirements and Considerations . . 62
Division of Clusters 63
Output 64
Other Programs 65
KEYOUT 65
DVMTRX 65
CLASSACC 67

RESULTS 68

The Classification Generated for the Construction Data Set 68
The Classification Generated for Map-043 77
Predictive Ability of the Classification 87

DISCUSSION AND CONCLUSIONS 94

LITERATURE CITED 104

APPENDIX A 111

APPENDIX B 133

VITA 175

LIST OF TABLES

Table page

1. Summary of status of Virginia soil survey (as of January, 1982), most recent surveys listed. 3

2. Soil types occurring in the Dickenson-like part of Wise County and in Map-043. 29

3. Geologic formations occurring in the Dickenson-like part of Wise County as mapped by Eby (1965). . . 37

4. Geologic formations occurring in Map-043 as mapped by Miller (1965). 39

5. Slope classes used in program DIVIDE. 45

6. Categories defined for the relative degree of influence of a nearby large or small stream as a soil-forming factor. 52

7. Attributes selected for divisions of the construction data set. 71

8. Diversity within each soil type with respect to the clusters derived for the construction data set. . . . 76

9. Accuracy per soil type in the classification generated for the construction data set. 78

10. Attributes selected for divisions of the Map-043 data set. 82

11. Diversity within each soil type with respect to the clusters derived for the construction data set. 88

12. Accuracy per soil type in the classification generated for the Map-043 data set. 89

13. Diversity within each soil type with respect to the clusters derived for the validation data set. . . 92

14. Accuracy per soil type in the classification derived for the validation data set. 93

LIST OF FIGURES

Figure page

1. Counties of southwestern Virginia. 22

2. POWER maps comprising Wise County, Virginia. 26

3a. Map of digitized soil types for four POWER maps in Wise County. 34

3b. Legend to map of digitized soil types in Fig. 3a. . . 35

4. Map of digitized geologic formations for four POWER maps in Wise County. 40

5. Map of slope classes for four POWER maps in Wise County. 47

6. Map of aspect classes for four POWER maps in Wise County. 48

7. Map of topographic shapes for four POWER maps in Wise County. 49

8. Map of distance-to-nearest-ridge for four POWER maps in Wise County. 50

9. Map of categories defined for distance-to-nearest-stream for four POWER maps in Wise County. . 53

10. Warnier-Orr diagram depicting the general processes performed by program DIVIDE. 56

11. Matrices showing how program DIVIDE restructures geographic data in the POWER system. 59

12. Dendrogram showing the sequence of divisions for the construction data set. 69

13a. Map of clusters for two POWER maps in the construction data set. 73

13b. Legend for the map of DIVIDE clusters given in Fig. 13a. 74

14. Map of DIVIDE clusters assigned to predominate soil types in two maps of the construction data set . 79

15. Dendrogram showing the sequence of divisions for the Map-043 data set. 80

16a. Map of digitized soil types in Map-043. 83

16b. Legend to map of digitized soil types in Fig. 16a. . 84

17a. Map of DIVIDE clusters in the Map-043 data set. . . . 85

17b. Legend to map of DIVIDE clusters in Fig. 17a. 86

18. Map of DIVIDE clusters assigned to predominate soil types in the Map-043 data set. 90

LIST OF APPENDICES

Appendix page

A. FORTRAN IV source code for program DIVIDE 111

B. Output from program DIVIDE 133

INTRODUCTION

Modern land-use planning requires the collection and analysis of large amounts of natural resources data. This includes data for topography; bedrock geology; mineral deposits; soils; surface and ground water resources; flooding hazards; climate; forest, farm and wetlands resources; fish and wildlife resources; and areas having scenic, historic, scientific and recreational value (Bauer 1973:2). To deal with both the volume and complexity of such data, computer-based geographic information systems (GIS) have been developed. The GIS stores, processes, and displays land-resources data referenced by geographic coordinates. It is a powerful tool for manipulating such data rapidly to make timely land use decisions. Past applications of the GIS to land use planning are well documented (Allen 1973, Fabos and Caswell 1976,Ferris and Fabos 1974, Lyle and von Wodtke 1974, Schneider and Amanullah 1979), and the frequency of applications in the future is likely to increase.

Because soil is such an important element of the land resource, when available, soil data have been used extensively in land use planning. Several authors have reported the value of soil surveys in regional and community planning projects. Examples may be found in the volumes edited by Bartelli et al. (1966), and by Beatty et al.

(1979). In the last 10 to 15 years, computers have been used to store, manipulate, and display soil maps (Butler et al. 1977, Jansen and Fenton 1978, McCann 1976, Meyers et al. 1974, Nichols 1975, Nichols and Bartelli 1974, Ragg 1977, Rogoff et al. 1980, Steinitz 1970). The demand for computer-generated interpretive maps and computer-assisted analyses of soils data is increasing. The opportunities for integrating soil data with other land resource data in geographic information systems are numerous.

Making a soil survey is a slow and methodical process requiring the skills of highly-trained soil scientists, field and lab technicians, and cartographers. It often requires over 10 years to map the soils of a county, complete the necessary correlation and interpretation studies, and publish the survey. Many localities are without soil surveys and planning efforts in these areas are therefore seriously handicapped, since the soil influences nearly all uses of land. In Virginia, 21 counties and several independent cities have never been soil surveyed. The current status of the soil survey in Virginia is summarized in Table 1.

Hamm (1978) addressed the problem of obtaining preliminary soil information for the unsurveyed areas in Virginia. Using an algorithm developed for the POWER geographic information system (Giles et al. 1976, Koeln et

Table 1. Summary of status of Virginia soil survey (as of January 1982), most recent surveys listed.

Mapping Completed Pre-1941	Mapping Completed 1941-1981 (published)	Mapping Completed 1941-1981 (unpublished)	Mapping in Progress	Unmapped
Appomatix	Augusta	Albemarle	Accomack	Alleghany
Arlington	Campbell	Clarke	Bedford	Amelia
Bland	Carrol	Frederick	Botetourt	Amherst
Grayson	Charlotte	Giles	Dinwiddie	Bath
Halifax	*Chesapeake	*Hopewell	Essex	Brunswick
Lee	Chesterfield	Ilse of Wight	Greene	Buchanan
Rockbridge	Culpeper	James City	Greensville	Buckingham
Russell	Fairfax	Middlessex	King William	Caroline
Scott	Fauquier	Montgomery	New Kent	Craig
Smyth	Fluvanna	Prince George	Northampton	Cumberland
Southampton	Gloucester	Prince William	Pittsylvania	Dickenson
Tazewell	Goochland	Pulaski	Powhatan	Essex
Wise	Hanover	Richmond	Rockbridge	Floyd
	Henrico	Roanoke	Shenandoah	Henry
	King George	Rockingham	Washington	Highland
	Lancaster	Spotsylvania	Wythe	King and Queen
	Loudon	*Virginia Beach		Nelson
	Louisa	Warren		Page
	Lunenburg	*Williamsburg		Patrick
	*Lynchburg	York		Surry
	Madison			Sussex
	Matthews			
	Mecklenburg			
	Northumberland			
	Nottoway			
	Orange			
	Prince Edward			
	Rappahannock			
	Stafford			
	*Suffolk			
	Westmoreland			

* - Independent city. Those not listed have either never been surveyed or were surveyed prior to 1941.

al. 1981), Hamm combined data on geologic formations, slope, aspect, and topographic shape to produce a map of likely soil units, since these land attributes are important factors in soil formation. It was proposed that the units derived might serve as a form of soil information at a level of detail between county soil association maps and more detailed county soil surveys. Hamm's algorithm, however, did not provide for correlating the units it derived with known soil units, e.g., series or associations, or with known soil properties. A method having this capability would be preferable.

Predicting a geographic attribute, such as soil type, based on other related attributes, is a problem particularly suited to analysis assisted by a GIS. Within the often broad constraints of available computing resources and the capabilities of the GIS itself, it is possible to overlay automatically dozens of different maps, examine the relationships therein, and produce predictive equations or classifications. A general method for using a GIS to predict land attributes based on other attributes would have many applications. Preliminary planning maps might be produced for areas lacking soil surveys, forest cover type maps, wildlife habitat maps, and other maps, where appropriate ancillary mapping is available. In addition, such maps could potentially serve as research tools and as

aids to survey operations in unmapped areas.

This thesis is designed to further Hamm's earlier work (1978) with the POWER system. The objective is to develop and test procedures for using a geographic information system to provide soils information for an unsurveyed county based on geology, topography, hydrography, and knowledge of soils in a similar county. The approach is to address the objective as a general problem in the use of a GIS so that the procedures developed may be easily applied to other areas and to other objectives.

LITERATURE REVIEW

GEOGRAPHIC INFORMATION SYSTEMS

> The complexity of Earth resources analysis and land use planning has prompted tecnological advances in automated data collection, storage, retrieval, manipulation, and display of spatial or geographically referenced data. Automation continues to increase as decision makers in both the public and private sectors require current natural resources and land use information, as well as detailed analysis of their interrelationships (Knapp and Rider 1979:57).

The GIS combines advances in natural resources data collection, remote sensing, cartography, and computer science into a versatile tool for planning and research. Eighty-five full GIS, having a wide range of capabilities and applications, are described in the International Geographic Union's review of computer software for handling spatial data (Calkins and Marble 1980). Rogoff (1978) provided a detailed annotated bibilography of six statewide systems. Salmen et al. (1977), Schnieder and Amanullah (1979), and Wilcott and Gates (1977) reviewed the capabilities and accomplishments of several additional systems.

The basic difference between a GIS and other computer-based information systems is the spatial or geographic nature of the data (Honeycutt al. 1980:1). The data of a GIS are referenced by geographic coordinates (e.g., geodetic latitude and longitude, universal transverse mercator (UTM),

state plane, or U.S. township and range coordinates). Data are entered as points (e.g., benchmarks, weather stations), lines (e.g., stream courses, highways), or areas (e.g., forest cover types, soil mapping units). Calkins and Tomlinson (1977) and Honeycutt et al. (1980) gave detailed accounts of data types and data structures in GIS. Davis and MuCullagh (1975) have edited a volume that addresses the theoretical and technical aspects of computer-assisted cartography, an important component of most GIS.

Kennedy and Guinn (1975:76) described the GIS as a data-driven system:

> "...while models which use the data are important to support the decision-making activities of those who use the system, a large portion of the investment [in a GIS] will be in obtaining, converting, and storing new data."

Most of the data in a GIS relate to areas, i.e., to land use types, land ownerships, census districts, etc. Areal data are usually stored and processed in either a grid or polygon data structure. Because the software written for these two structures is very different, most GIS can be described as grid systems, polygon systems, or systems able to accept both kinds of data.

Geographic registration of areal data in grid systems is accomplished by the orientation and dimensions of the grid itself, whereas polygon systems specify geodetic (X,Y) coordinates to describe the boundaries of individual map

units. Polygon systems are able to store a more accurate representation of areal data than grid systems, but grid systems generally require less computer storage and are easier to program. The advantages and disadvantages of grid *vs.* polygon data structures have been documented (Pueker 1972, Durfee 1974, Honeycutt et al. 1980).

The POWER GIS was used in this research. POWER was developed at Virginia Polytechnic Institute and State University during the mid-70's as an aid to the selection of minimum-impact powerline corridors (Giles et al. 1976, Jones 1976, Smart 1976). It is a grid system based on UTM coordinates. Continued research and development with POWER have made it the heart of a large and versatile statewide GIS (Koeln et al. 1981). Data have been collected for elevation, streams, lakes, wetlands, soils, land use, recreational sites, urban density, roads, utilities, and historic sites. POWER has been applied to evaluating the environmental impacts of large physical projects (Cason 1980), mapping the probable ranges of endangered terrestrial mammals (Hoar 1980), and siting general aviation airports (Koeln 1980).

APPLICATIONS OF SOIL DATA IN GEOGRAPHIC INFORMATION SYSTEMS

Soil data are included in the data bases of several statewide GIS (Calkins and Marble 1980), and since 1970, the

literature contains frequent references to computer-assisted applications of soil data. The major use has been the generation of interpretive maps showing soil suitability for various land uses. Hunter et al. (1966) contended that the greatest restriction in the use of soil surveys was the difficulty in obtaining interpretive maps. With computer-assisted processing and display of soil data, maps showing soil features or soil interpretations can be produced on demand to satisfy specific user requests, generally in a short amount of time and at a small cost. Nichols and Bartelli (1974) discussed the characteristics and uses of computer-generated interpretive soil maps. Rogoff et al. (1980) developed procedures by which ratings of soil potentials for five urban land uses were generated using a computer-based natural resource information system and displayed in computer-drawn interpretive maps.

Soil data can be readily combined with other data in a GIS. Butler et al. (1977) used data for soil, land use, slope, wetlands, ground water, and surface water to map areas of probable suitability for two on-site waste disposal technologies. McCann (1976) produced maps to help guide urban development in one Arizona county using a computer mapping system with data for soil, topography, geology, hydrology, and wildlife. In Massachusets, another effort to plan urban growth made extensive use of a GIS containing

soil and other natural resources data (Fabos and Caswell 1977, Ferris and Fabos 1974).

SOIL DATA FOR UNSURVEYED AREAS

When soil mapping is not available, the possibilities for land use planning and prudent management of the land are greatly diminished. Voelker (1977) noted the difficulty encountered in regional planning efforts when an adequate soil data base was lacking. Because soils are related to other characteristics of the land for which data bases often already exist, it may be possible to produce maps that approximate soil conditions in areas without soil surveys.

Soils are conceived by soil scientists as independent natural bodies that have properties reflecting the effects of local and zonal soil-forming factors (Soil Survey Staff 1975:1-5). Buol et al. (1980:105) defined a factor of soil formation as "an agent, force, condition, or relationship, or combination of these, which influenced, or may influence a parent material of a soil with the potential of changing it." The many possible factors of soil formation are usually generalized as five major factors: climate, living organisms (especially vegetation), parent material, relief, and time (Jenny 1941). Boundaries between soils occur where one or more of the soil-forming factors have been different at some time during the development of the soil.

Working with the POWER GIS in two Virginia counties, Hamm (1978) combined maps of geologic formations, slope, aspect, and topographic shapes to produce a map of soil units for use in regional planning. These soil units were intended to fulfill a need for information at a level of detail between generalized soil association maps and more detailed county soil surveys. Hamm digitized geologic formations from a 1:250,000 scale regional map and elevation from 1:24,000 scale topographic maps. The cell size was 1/9 km^2 (11 ha or 27 acres). Maps of slope, aspect, and topographic shape were generated from the cellular elevation data. Different types within each of these land attributes were aggregated, i.e., certain geologic formations were grouped, broad slope and aspect classes were formed, and topographic shapes were combined. These maps were overlayed yielding 92 of 96 possible soil units. Eleven infrequent units were combined with other similar units.

Hamm did not report a comparison of the derived soil map with the soil association map available for one of the two counties. If a derived soil map is to be used, such a comparison is necessary so that soil features and interpretations can be associated with its units. Inaccuracies in the regional geologic map were cited as possible sources of error in units of the derived map, and it was recommended that future applications of the

methodology employ more detailed geologic mapping.

Hamm's methodology was deficient in that it did not employ actual soil data in determining how to aggregate and overlay other data. Associations of sufficient strength to delineate soil boundaries were hypothesized, not disclosed by analysis. By Hamm's methodology, only trial and error comparisons of actual and derived soil maps can demonstrate whether good decisions are made with respect to the choice of attributes and the manner in which attributes are handled.

Croswell (1981) used the Kentucky Natural Resources Information System to delineate landform-based soil units for 1/225-km² (0.4-ha or 1.1-acre) cells. The approach was similar to Hamm's work with the POWER system except that geologic data were not used and more attention was given to associating derived with actual soil units. Slope and aspect were synthesized from cellular elevation data and seven distinct landforms (comparable to the topographic shapes that Hamm employed) were manually interpreted from 1:24,000 scale aerial photography. Croswell defined a small number of slope classes and treated aspect as either north- or south-facing. Decisions governing how these data were combined to produce a composite map of soil units were based upon a study of soil maps available for an area near the unsurveyed area of interest. Combinations which appeared to

frequently delineate soil boundaries were determined by a manual overlay of the landform, slope, aspect, and soil maps. After field checking, rules for combining the data were established which created seven soil units.

Croswell did not report how well the derived soil units corresponded with actual soil mapping units in the surveyed area, and without this information it is difficult to evaluate fully the methodology. Nevertheless, it represented an improvement over Hamm's methodology in that the manner by which data was combined to produce soil units was guided by available soil data.

Croswell's methodology, however, involved a cumbersome process of manual overlays and visual analysis. Smart (1976:22) outlined the deficiencies of procedures based on manual map overlays. Such overlays are generally expensive and slow to produce, and relatively few maps can be included before the resulting image becomes either too complex or too blurred to interpret. Given an appropriate algorithm for analyzing composite map units, the automated overlay capabilities of a GIS would prove far more efficient and probably more accurate. The development of such an algorithm is central to the objective of this research.

METHODS

OVERVIEW

Methods have been sought for deriving soil units in unsurveyed areas that meet three basic requirements: 1) the methods should be generalizable; 2) the methods should be as free from subjective manipulations of the data and subjective evaluations of the results as possible; 3) the methods should produce results that are readily verifiable.

In keeping with the first requirement, all procedures have been designed to be as flexible and as data independent as possible, and have been documented using a general terminology. Although the specific objective dealt with providing soil information for an unsurveyed county of Virginia, the opportunity existed to develop methods that would apply to this objective and to other objectives as well. Given appropriate data, the techniques and programs described in the following sections could also be used to map the probable occurrence of plant associations that provide food and cover for various wildlife species. For the second requirement, weighting and other subjective manipulations of data in the analysis have been kept to a minimum. To avoid the problem of deriving soil units that cannot be characterized without extensive field work (a shortage of field data is the problem that this project is

intended to relieve), the methods are designed to extend existing knowledge of soils, not to generate hypotheses about soils that will not be tested. Addressing the third requirement, a reasonable approach is outlined for testing the predictive ability of the procedures with existing data by the technique of cross validation.

The basic problem that this research addresses is predicting one geographic attribute based on other related geographic attributes. The statistical method most commonly applied to this kind of problem is multiple linear regression. Regression, and other related statistical methods, are intended for the analysis of continuous random variates sampled from a large population. The data for many geographic attributes meet these requirements, but many do not. Non-quantitative attributes, such as soil and geologic types, are difficult to use in regression, especially when this kind of attribute becomes the dependent variable in the analysis.

It would be possible in some cases to replace non-quantitative attributes with quantitative surrogates, for example, to replace soil types with corresponding values for permeability, depth to bedrock, shrink-swell potential, etc., but such data are also not always available in a form suitable for statistical analysis. Such data are often given as a range of values, or as a series of highly

detailed measurements for one sample from one site.

More robust methods were sought that could be applied to the analysis of soil types, surficial geologic types, vegetative cover types, etc., as non-quantitative attributes and that could be applied to the analysis of quantitative attributes as well. In addition, methods were sought that would require as few assumptions as possible about the method of data collection.

When predicting a geographic attribute is regarded as a problem in the classification of individuals, the data previously discussed become much more tractable, especially when the attribute of interest is a non-quantitative attribute. In a grid-based geographic information system, such as the POWER system, the individuals to be classified are grid cells, the spatial units with which all data are associated. The attributes associated with these individuals are characteristics of the land such as bedrock type, slope clas, topographic shape, etc. The problem, then, becomes one of grouping cells on the basis of selected attributes to disclose the nature of another particular attribute of interest.

Williams (1971) presented a dichotomous key to classification methodologies. An early decision in the key is between intrinsic and extrinsic classifications. In an intrinsic classification, individuals are grouped based on

attributes that are all considered equivalent. The purpose is simply to group similar individuals. An intrinsic classification may then be examined to determine whether the groups derived reflect discontinuities in other attributes, but no reference is made to these attributes in constructing the classification. Attributes which do not serve specifically as criteria for the grouping of individuals are termed extrinsic attributes. For example, sample sites were classified in a study of a marine benthic community based on the occurrence of different species, (species were used as the intrinsic attributes) and the resulting groups of sites were later analyzed with respect to extrinsic attributes such as depth, salinity, and substratum characteristics (Stephenson et al. 1972). Williams (1971:307-308) outlined the difference between this kind of analysis and a classification that directly addresses relationships between attributes:

> In an extrinsic classification, however, an external attribute is declared in advance, and the resulting groups, though based entirely on the internal attributes, are required to reflect discontinuities in the external [extrinsic] attribute as closely as possible, the intrinsic pattern being overridden if need be.

The terms extrinsic and intrinsic are used to distinguish the manner in which attributes are handled in a classification, not necessarily to distinguish the nature of the attributes themselves. An extrinsic atribute is simply

one that is singled out as the attribute of special interest in an extrinsic classification.

Tracey (1968) applied an extrinsic classification strategy to group a series of pastures based on the occurrence of different plant species. Tracey's aim was to create groups that reflected various pasture treatments: burning, mowing, etc. Macnaughton-Smith (1965) applied similar methods to sociological analyses. In a grid-based geographic information system, exploring the relationships between soil types and other cell attributes may be considered as an extrinsic classification problem, and this was the approach taken. Soil type was the extrinsic attribute in the classification, and bedrock type, slope class, topographic shape, etc., were the intrinsic attributes. If the objective had been, instead, to classify cells in a way that would reflect vegetative cover type, then vegetative cover type would have become the extrinsic attribute and soil type, slope class, etc., might have been used as intrinsic attributes. Because the terms extrinsic and intrinsic are used frequently in the following description of my methods, it is important to realize that these terms serve only to distinguish the role that a particular attribute assumes in the classification procedure.

Any classification scheme must employ some means of grouping or separating individuals in a population, and this implies some measure of similarity or dissimilarity. Clifford and Stevenson (1975:49-82) discussed several measures in detail and noted that some are better suited to certain types of data than others. Many measures apply only to attributes measured on a continuous quantitative scale or to simple binary attributes, and are inappropriate for dealing with non-quantitative attributes.

An information theory measure of similarity/dissimilarity was selected to derive soil units based on other attributes because of its ability to handle a variety of data types, including quantitative and non-quantitative data (Lance and Williams 1968a and 1968b). The measure, called _H, is basically a measure of diversity within a group. It is useful for generating classifications because it can serve as either a criterion for fusing similar groups or for separating dissimilar groups. Its derivation and application in the method developed will be explained later.

Classifications may be hierarchical or nonhierarchical, but _H is most readily incorporated in a hierarchical classification. A hierarchical classification may be divisive, proceeding by a sequence of divisions from the population to the individual, or it may be agglomerative,

proceeding by a sequence of fusions in the opposite direction. A divisive classification is better suited to a data set composed of many individuals described by few attributes (Clifford and Stevenson 1975:101). This is characteristic of data sets belonging to grid-based GIS, in which large data sets containing tens of thousands of cells are described by fewer than 50 attributes. Although many advantages lie with agglomerative methods, such methods would prove too costly for large data sets, therefore a divisive method was developed.

Divisive classifications may be monothetic, dividing the population based on a single attribute at a time, or polythetic, dividing by several attributes at a time. Few polythetic divisive programs have been developed, and I do not know of any that may be considered an extrinsic classification. Monothetic methods are far simpler than their polythetic counterparts, and because the former were readily adaptable to an extrinsic classification problem, a monothetic divisive strategy was followed.

In summary, the method developed is an extrinsic, hierarchical, monothetic divisive classification for predicting one geographic attribute based on other related geographic attributes. The program written to generate the classification, called DIVIDE, and the programs written to evaluate and apply the classification, are designed as parts

of the POWER GIS. (Giles et al. 1976, Koeln et al. 1981). The program listing for DIVIDE is given in Appendix A.

THE STUDY AREA

Recent data collection and analysis for the POWER system have been focused on the counties of southwestern Virginia to meet the information needs of land use planning in these areas. The soils of several of these counties were surveyed during 1935-1940 at a scale of 1:24,000 (Fig. 1). One of the counties not surveyed, Dickenson County, is adjacent to two of the counties that were surveyed, Wise and Russell Counties. Nearly all of Dickenson county shares common geomorphology with a large portion of Wise County (Eby 1923, Giles 1921).

The general soil map of the state (USDA-SCS and VPI & SU 1979) shows both counties composed entirely of two major soil associations, except for the Powell Valley portion of Wise County, which is shown as a third association. The larger of the two associations is characterized as "Shallow to deep, steep to very steep soils formed mainly in residuum from acid sandstone, shale and phyllites; on mountains." The smaller is characterized as, "Moderately deep, gently sloping to steep soils formed in residuum from acid shale; on uplands of dissected valleys."



Fig. 1. Counties of southwestern Virginia. The soils of all counties shown, except Dickenson and Buchanan counties, were surveyed during 1935-1940.

This map is far too generalized to serve as a local or even a regional planning tool, but it and the geologic surveys of the region suggest that knowledge of the soils of Wise County, provided by its more detailed soil survey, might be applied to Dickenson County to produce a preliminary source of soil information. The methods developed in this thesis, however, may apply not only to Wise and Dickenson Counties, but to other regions where information available for one area is pertinent to another similiar area.

PARTITIONING THE STUDY AREA

Tracey (1968) applied a monothetic divisive procedure to classifying pastures which had been subjected to a variety of treatments: burning, mowing, etc. Tracey's aim was to predict the treatment history of sites by the occurrence of different plant species. Half of the sites sampled were randomly assigned to a data set used to construct the classification, and the other half to a second data set used to test the classification. The extrinsic attribute, pasture treatment, was used in generating the classification with the first data set, but no reference was made to this attribute in applying the classification to the second data set. The accuracy of the classification in the second data set was a measure of its predictive ability.

Following Tracey's example, data for the Wise County study area were initially partitioned into two data sets: one to construct a classification of soil types and another to test the predictive ability of the classification. The two are hereinafter referred to as the construction data set and the validation data set, respectively.

In order to explain how the study area was partitioned into two data sets, it is necessary to describe the data structure of POWER, the geographic information system used in this research. In POWER, different types of data , elevation, land use, etc., are stored in separate computer files called mapfiles. POWER mapfiles are composed of blocked data. Each block contains 900 cells in a 30 X 30 square matrix and a label designating its geographic location. A block of 900 cells with its associated label is called a POWER map (a map is one logical record in computer storage). Most of the data for POWER has been collected for 1/9-km^2 (11-ha or 27-acre) cells, and some for 1/81-km^2 (1.2-ha or 3.0-acre) cells. A 1/9-km^2 cell contains nine 1/81-km^2 cells. The smaller cell size was selected in order to work with more precise data.

The partitioning of the study area was acomplished by randomly selecting a POWER map for the construction data set from the 130 POWER maps that cover Wise County, and then by assigning every other map in the county to one or the other

data set. This method of partitioning the study area was better than one which either randomly or systematically assigned individual cells because: 1) it would most likely yield data sets containing similar proportions of all attributes; 2) it would simulate the eventual application of the classification to Dickenson county by keeping adjacent map cells between the two data sets at a low number; 3) it would require a small amount of bookeeping on the two data sets since maps are more easily manipulated in the POWER system than are individual map cells.

Maps located in and around the Powell Valley region of Wise County were excluded from both construction and validation data sets since this area is unlike the potential application area, Dickenson County. However, one map in the Powell Valley, referred to as Map-043, was singled out for a separate detailed analysis due to the availability of 1:24,000 scale geologic mapping (Miller 1965) for this area. Fig. 2 shows the partitioning of the Wise County study area into the two data sets and the location of Map-043. Fig. 2 also shows the location of Map-129, a map randomly selected for comparisons of different sources of elevation data, and the location of four maps chosen for the display of data in subsequent figures. The size of the study area prohibits the meaningfull display of entire data sets in one-page figures. Fifty-four maps comprised the construction data

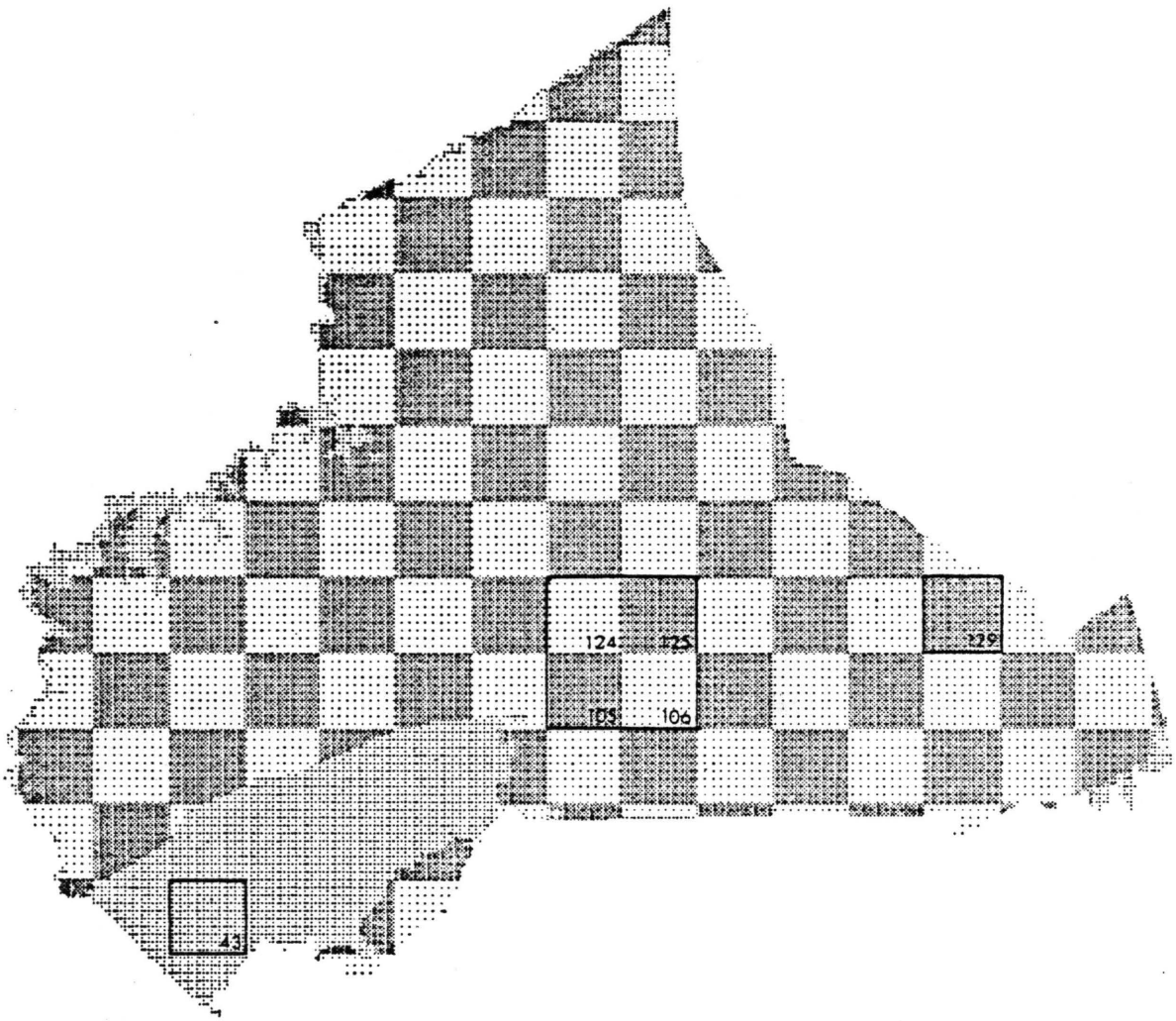


Fig. 2. POWER maps comprising Wise County, Virgina. The darkest areas were assigned to the construction data set and the lighter areas to the validation data set. The areas having and intermediate tone are geologically unlike neighboring Dickenson County. Map-043 was selected for a separate analysis and Map-129 for comparisons of different sources of elevation data. Maps 105, 106, 124, and 125 were selected as an area for the display of data in several figures that follow. Each map contains 900 1/81-km^2 (1.2-ha or 3.0-acre) cells.

set and 55 maps comprised the validation data set.

DATA COLLECTION

The selection of attributes by which to predict the attribute of interest in this study, soil, was guided by the following considerations: 1) attributes should be related to recognized factors of soil formation; 2) attributes should be available from published maps, either directly as digitized data or indirectly as data synthesized from the former; 3) attributes should be meaningful in the context of a grid-based geographic information system using a cell size of 1/81 km^2 (1.2 ha or 3.0 acres).

The soil survey for Wise County reported that, "Since climate and vegetation have no doubt been relatively constant, they are not to be considered the major factors that have brought about differences in the morphology of soils," (Perry et al. 1954:95). Attributes for predicting soil in Wise County were therefore limited to those relating to parent material, relief, and age of landform. Following is a description of the methods of data collection (or synthesis) for the soil extrinsic attribute and for each intrinsic attribute used to predict soil in program DIVIDE.

Soil

Source of Data

Mapping of the soils of Wise County was completed in 1940 and the final soil survey was published in 1954. The mapping scale was 1:24,000 and the minimum mapping unit was about 2 ha (5 to 10 acres). Units as small as 1 ha. (about 3 acres) were delineated, however, if an area contrasted highly with the surrounding soils. The 3.0-acre cell size is therefore an appropriate unit for data collection.

Thirty-seven soil types were identified in the Wise County survey. The soil type is the lowest category in the system of soil taxonomy used in this survey, which was made before implementing the system currently in use (Soil Survey Staff 1975). Concepts of soil classification have greatly changed in the new system and the reader is cautioned that soils named in the Wise County survey and referenced in this thesis are not necessarily the same soils identified by similar names in more recent surveys of other areas.

Seventeen of the 37 soil types were mapped as two or more soil phases, based on differences in slope or relief, to give a total of 56 different mapping units. To reduce the number of units to code for computer storage and processing, and because data for slope and relief were available from other sources, I worked with soil types and

Table 2. Soil types occurring in the Dickenson-like part of Wise County and in Map-043.

Soil Type	Dickenson-like part of Wise Co. # cells	%	Map-043 #cells	%
Atkins loam	1468	2.1	2	0.2
Clarksville-Fullerton stony loams			3	0.3
Coeburn silt loam	4699	6.5		
Coeburn-Muskingum complex	7491	10.5		
Dekalb loam	110	0.2		
Enders silt loam	197	0.3		
Fullerton loam			4	0.4
Greendale fine sandy loam	10	*		
Hagerstown silt loam	7	*		
Hagerstown stony silt loam	19	*		
Hartsells fine sandy loam	1208	1.7	16	1.8
Hayter loam			3	0.3
Hayter stony loam			22	2.4
Holston fine sandy loam	166	0.2		
Jefferson loam	108	0.2	9	1.0
Jefferson stony loam	1780	2.5	102	11.3
Leadvale loam	889	1.2	3	0.3
Made Land	258	0.4		
Mine Dump	49	0.1		
Monogahela fine sandy loam	9	*		
Montevallo shaly silt loam			24	2.7
Muskingum stony loam	42063	58.8		
Muskingum fine sandy loam	831	1.2	25	2.8
Muskingum stony fine sandy loam	1858	2.6	551	61.2
Muskingum-Litz silt loam			82	9.1
Philo loam	287	0.4	7	0.8
Pope fine sandy loam	329	0.5	8	0.9
Pope loamy fine sand	427	0.6		
Rolling stony land	1174	1.6	1	0.1
Rough stony land	5407	7.6	10	1.1
Sequatchie fine sandy loam	131	0.2		
Stony alluvium	592	0.8		
Teas-Dandridge silt loam	5	*	8	0.9
Westmoreland silt loam	4	*	20	2.2
Total cells	71576		900	

* Less than 0.1 percent

not with soil phases. Table 2 gives the approximate areal and proportional extent of each soil type in the Dickenson-like part of Wise County, and in the POWER map (Map-043) selected for a separate analysis.

The Wise soil survey, along with other surveys of the time, was an agricultural soil survey, and little attention was given to non-agricultural uses of the land in interpreting the soils mapped. The publication, "Interpretive Guide to the Soils of Southwest Virginia TVA [Tennessee Valley Authority] Counties" updates the agricultural interpretations given by the original survey and provides additional characterization data and non-agricultural interpretations that are most important for the area (Lietzke and Porter 1978).

Method of Digitizing

Digitizing is the process of translating data in map form to a computer-compatible form. The methods of digitizing for the grid-based POWER geographic information system are documented in previous work with the system (Jones 1976, Koeln 1980). Digitizing soil data is similar to digitizing land use, vegetative cover and other types of areal data. Briefly, the process as currently practiced in the POWER system involves the following steps:

1) Derivation of a coding scheme. Each soil type is assigned

a numeric code.

2) Geographic registration. A soil map is placed on an electronic graphics tablet and a scaled 30 X 30 grid is then oriented on the map with respect to the proper UTM coordinates. One 30 X 30 grid, called a POWER map, is coded at a time.

3) Data entry. The soil type codes are assigned to grid locations and placed in computer storage. This is aided by a minicomputer linked to a storage device and the graphics tablet mentioned earlier. The program PWR30X30, developed by A. Blair Jones, III, of the Department of Fisheries and Wildlife Sciences, VPI and SU, Blacksburg, Virginia, manages the computer hardware involved.

Initially, cells were assigned soil type values according to a strict majority rule coding scheme, i.e., the soil type occupying the most area in a cell was coded. This tended to exclude many of the smaller and narrower mapping units thereby seriously reducing variability in the digitized version of the soil map. Since an objective in the study was to disclose patterns of soil variability, the coding scheme was changed to a modified majority rule. Units which occupied an area the size of one cell or larger, but did not occupy the majority of any of the cells in which they occurred, were assigned to as many cells as approximated their size. The cells chosen were those in

which the unit appeared to be concentrated. This coding scheme preserved units occurring as islands or narrow strips within larger units. This coding scheme introduced an additional complexity to the coding process, in that coding decisions had to be made in the context of groups of cells, rather than on a cell by cell basis, but it achieved the desired effect of producing a digitized map that more closely reflected soil variability and the areal extent of all units.

Once a 30 X 30 map was digitized, it was transferred from the microcomputer system to a POWER mapfile on the university computer system using the POWER program TEKCELLS, also developed by Mr. Jones. A mapfile contains all the 30 X 30 maps of a single attribute for the entire study area.

The digitized and transfered map was then plotted on an electrostatic plotting device at the same scale as the original soil map using programs BRDIME and SYMPLOT, developed with the assistance of Dr. G. T. Koeln of the Department of Fisheries and Wildlife Sciences, VPI and SU, Blacksburg, Virginia. BRDIME plots boundaries between adjacent cells having different values, and SYMPLOT plots the actual cell values. These plots were used to check the digitized map for coding errors by aligning the plots over the soil map on a light table. Errors were corrected by editing the original minicomputer file and then transferring

the edited POWER map to the university computer again. Fig. 3 shows the digitized soils of four maps in the central part of Wise County. Fig. 3 was produced on an electrostatic plotting device using program RGRAY, developed by Dr. Koeln.

Geology

The parent materials of soil may be classified as:
1) residuum, material derived from the weathering of consolidated rock in place;
2) colluvium, deposits of rock fragments and soil material accumulated by the action of gravity; or,
3) alluvium, material deposited by the action of streams.

Most of the soils of Wise County are considered to be formed from residuum (Perry et al. 1954:13-15). The geologic surveys for Wise County (Eby 1923:72) and Dickenson County (Giles 1921:15) both report that, in general, the character of soil in the region is largely dependent upon the nature of the underlying rock formations. The pattern of surficial geologic formations of the area was therefore considered a good attribute for predicting soils.

Source of Data

The Virginia Geologic Survey Bulletin, "The Geology and Mineral Resources of Wise County and the Coal-bearing Portion of Scott County, Virginia" (Eby 1923) provided

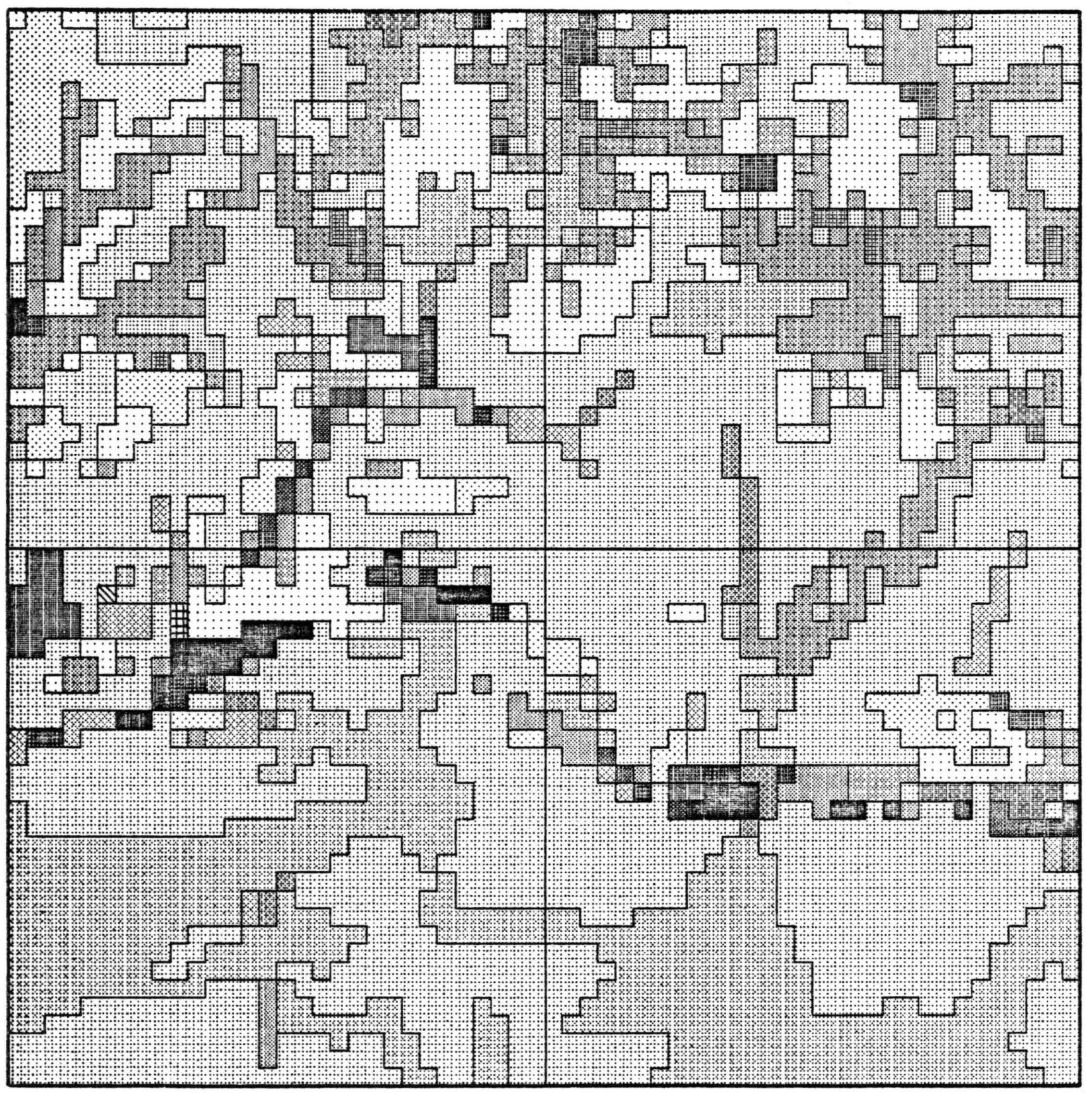


Fig. 3a. Map of digitized soil types for four POWER maps in Wise County. The scale is 1:48,000. A legend is given in Fig. 3b.

Soil Type	No. cells	Tone
Atkins loam	123	
Coeburn silt loam	338	
Coeburn-Muskingum complex	151	
Enders silt loam	24	
Hartsells fine sandy loam	83	
Holston fine sandy loam	29	
Jefferson stony loam	66	
Leadvale loam	72	
Made Land	5	
Mine Dump	1	
Monogahela fine sandy loam	3	
Muskingum stony loam	1662	
Muskingum fine sandy loam	16	
Muskingum stony fine sandy loam	32	
Philo loam	21	
Pope fine sandy loam	2	
Pope loamy fine sand	20	
Rolling stony land	332	
Rough stony land	574	
Sequatchie fine sandy loam	15	
Stony alluvium	31	

Fig. 3b. Legend to map of digitized soil types in Fig. 3a.

geologic data for Wise County. In this survey, 23 different bedrock formations were identified and mapped at a scale of 1:62,500. Only five of these formations occurred in the construction and validation data sets. The other 18 occurred in the Powell Valley region of the county considered unlike neighboring Dickenson county.

A similar survey, "The Geology and Coal Resources of Dickenson Co. Virginia" (Giles 1921) provided the geologic data for the potential application area. Six different bedrock formations were identified and mapped at a scale of 1:62,500. Four of these six are the same formations mapped in the Wise County survey, the other two are subdivisions of a formation also mapped in Wise County. The Dickenson-like part of Wise County, from which the construction and validation data sets were derived, was defined as the area having the same geologic formations as found in Dickenson County. Table 3 lists the formations occurring in Dickenson County and the Dickenson-like part of Wise County and gives the approximate areal and proportional extent of each in the latter.

The source of geologic data for the Map-043 data set was a 1:24,000 scale map (Miller 1965) rather than the 1:62,500 scale map (Eby 1923) used for the construction data set. Although three different parts of Wise county have been the subject of 1:24,000 scale geologic surveys, Map-043

Table 3. Geologic formations occurring in the Dickenson-like part of Wise County as mapped by Eby (1965).

Geologic Formation	# cells	%
Wise formation	38445	53.5
Gladeville sandstone	4377	6.1
Norton formation	18897	26.3
Lee formation	9980	13.9
Undifferentiated Mississippian	99	0.1
Total Cells	71798	

was the only full POWER map in the county for which this type of data was available. Table 4 lists the formations and the approximate areal and proportional extent of each formation occurring in Map-043.

Method of Digitizing

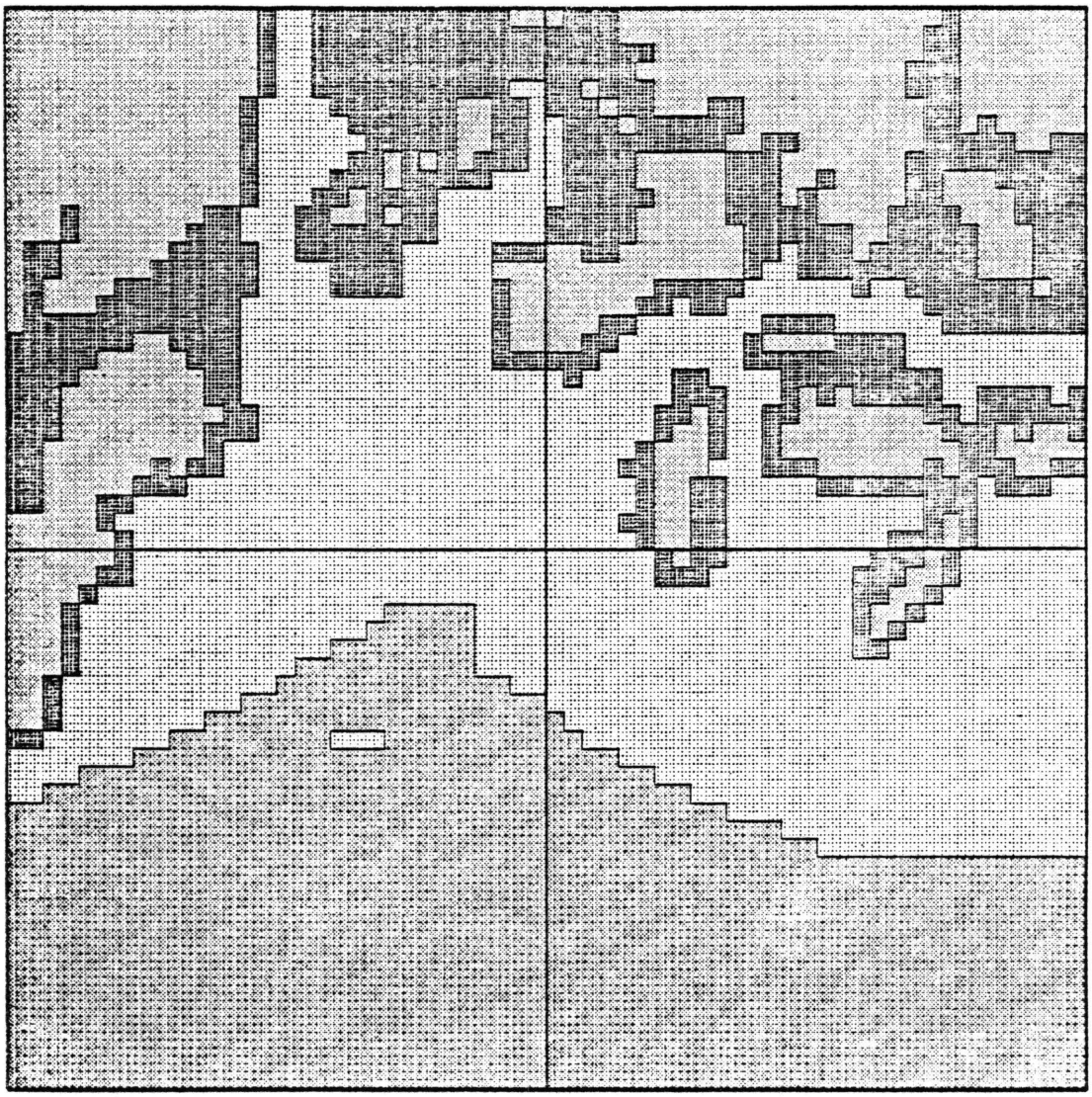
The two geologic maps were digitized and checked for errors following the same methods described for the soil data, except that the geographic registration of the geologic data was handled differently. Whereas UTM coordinates were readily transfered to the 1:24,000 soil quadrangles from the corresponding USGS 1:24,000 topographic quadrangles, the different scale and nonstandard coverage of the geologic maps required a more involved procedure. Latitude-longitude control points were taken from the geologic maps and were converted to UTM coordinates. These control points were then plotted by computer along with a UTM grid at a scale of 1:62,500. The plots were then used to transfer UTM registration to the geologic data. Fig. 4 shows the formations of four maps in the central part of Wise County.

Elevation

Elevation is an important attribute, because it provides the raw data for modeling a variety of useful

Table 4. Geologic formations occurring in Map-043 as mapped by Miller (1965).

Geologic Formation	# cells	%
Bluefield formation	6	0.7
Greenbrier limestone	9	1.0
Maccrady and Price formations	7	0.8
Chattanooga shale	70	7.8
Wildcat Valley sandstone	17	1.9
Hancock limestone	30	3.3
Clinton formation	313	34.8
Clinch sandstone	390	43.3
Sequatchie formation	46	5.1
Reedsville shale	12	1.3
Total Cells	900	



Geologic Formation	# Cells	Tone	Geologic Formation	# Cells	Tone
Wise Formation	673		Norton Formation	1177	
Gladeville Sandstone	648		Lee Formation	1102	

Fig. 4. Map of digitized geologic formations for four POWER maps in Wise County. The scale is 1:48,000.

terrain features. Initally, 3-acre cell size elevation data were taken from a USGS Digitial Terrain Tape of the region (USGS 1980). The tape available for the Wise-Dickenson study area was produced by electronically scanning small scale (1:250,000) topographic maps of the area. Discrepencies between this data and other especially reliable data, however, raised serious concerns about the accuracy of the tape. It was frequently observed in plots of the data that cells of minimum and maximum elevation did not agree with vector-digitized data for stream courses and ridge lines. This inaccuracy is probably due to the small scale of the maps used to produce the tape. The best alternative to the Digital Terrain Tape, coding elevations for the centroid of each 3-acre cell in the study area from 1:24,000 topographic maps, was beyond the resources of this project.

The next best alternative was to use elevation data interpolated from existing 27-acre-cell-size data. The 27-acre cell (1/9 km^2 or approximately 11.1 ha) is the more commonly used cell size in the POWER system, and elevation data has been coded for the entire state of Virgina at this cell size. A 3 X 3 block of 3-acre cells comprise one 27-acre cell in the POWER system. To convert from 27-acre data to 3-acre data, a 27-acre cell value becomes the value of the 3-acre cell in the center of the block. The other

eight 3-acre cell values in the block are interpolated from neighboring 27-acre cell values. Admittedly, interpolation extends the 27-acre data beyond their intended use in regional planning, but the possibility was worth investigating that interpolation might yield a better elevation data set than otherwise available.

With the assistance of Dr. G. T. Koeln, four different interpolations were compared with 3-acre elevation data coded from a 1:24,000 topographic quadrangle for one POWER map randomly selected from the study area. The root mean square difference was calculated between each interpolation (including also data from the Digital Terrain Tape) and the 3-acre map. The root mean square difference is a simple measure of the average absolute difference between a series of paired numbers. The paired numbers in this case are the elevation values of cells from the interpolations and cells from the more accurate 3-acre map. The best interpolation was the one having the smallest root mean square difference with the 3-acre map.

The best interpolation was a bicubic spline interpolation generated by a program developed by Dr. Koeln. It had a root mean square difference of 27.1 m (88.9 ft) when compared to the map coded at 3-acre-cell-size resolution. Further investigation revealed that much of the difference between the 3-acre map and the interpolated maps

was due to coding errors in the original 27-acre data from which the interpolated maps were derived. The root mean square difference between the 27-acre cell values and the corresponding 3-acre cell values was 21.6 m (70.9 ft). Elevation coding is a slow and tedious process and some error was to be expected, but the level of error indicated in this one map exceeded what I considered a level acceptable for my analysis. Still, these data were nearly an order of magnitude better than the only feasible alternative, data from the Digital Terrain Tape, which had a root mean square difference of 49.8 m (163.3 ft) when compared to the 3-acre data.

The small size of the Map-043 study area made it possible to use elevation data other than that interpolated from 27-acre data. Elevations were coded for 3-acre cells directly from the most recent USGS 1:24,000 topographic map of the area, which had a contour interval of 12.2 m (40 feet).

Attributes Derived from Elevation

Relief, the elevations or inequalities of a land surface, is considered a major soil forming factor:

> Within specific geographic regions, the following soil properties are commonly found to be relief related: 1) depth of the solum; 2) thickness and organic matter content of the A horizon; 3) relative wetness of the profile; 4) color of the profile; 5) degree of horizon differentiation; 6) soil reaction; 7) soluble salt

> content; 8) kind and degree of "pan" development; 9) temperature; 10) character of the initial material. (Buol et al. 1980:123).

Relief has been characterized in many different ways in previous studies with soil. Elevation, gradient of slope, and aspect (slope direction) were found statistically significant predictors of selected soil properties in Iowa (Walker et al. 1968). In Saskatchewan, Canada, Acton (1965) found that the type of soil occurring on a particular slope segment was related to the gradient of the segment and the position of the segment on the entire slope.

Using the POWER system program TOPO (Lawrence 1976, Jones et al. 1976), three relief attributes were derived from the interpolated elevation data: slope, aspect, and topographic shape. TOPO is a version of Grender's (1976) program, TOPO III, modified to accommodate the POWER data structure. The predominant slope of a cell is calculated based on the greatest change in elevation between a cell and its neighboring cells. Aspect is determined as one of eight principle compass directions that the predominant slope faces. The topographic shape of each cell is classified as either 1) slope, 2) ridge, 3) valley, 4) knob, 5) depression, or 6) saddle.

Because the classification program DIVIDE requires that continuous attributes be converted to an interval scale, the POWER program CATEGORY (Jones et al. 1976) was used to

Table 5. Slope classes used in program DIVIDE.

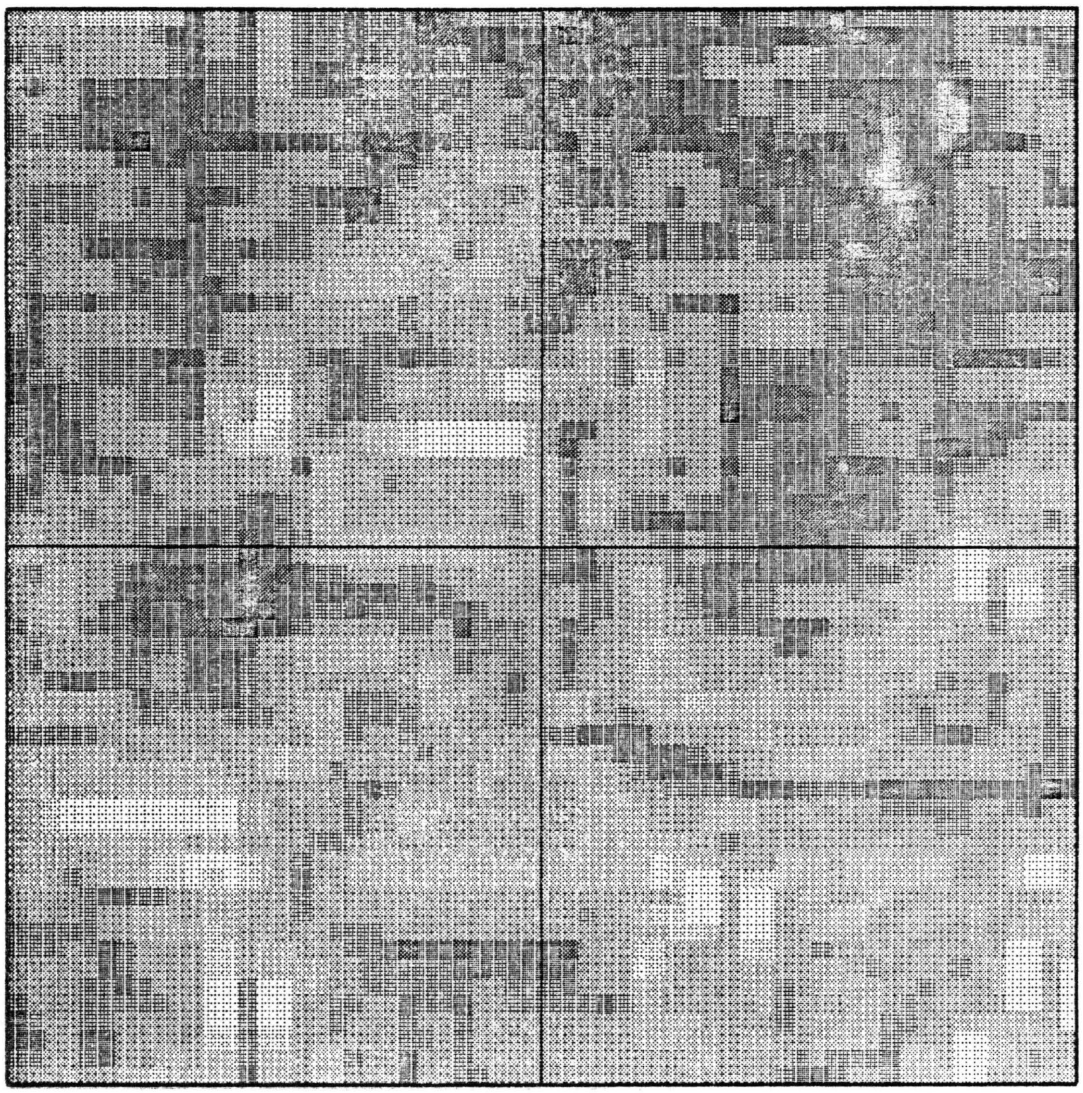
Slope Class	Range in Gradient of Slope (percent)
1	0 - ≥ 3
2	> 3 - ≥ 5
3	> 5 - ≥ 10
4	> 10 - ≥ 15
5	> 15 - ≥ 30
6	> 30 - ≥ 50
7	> 50 - ≥ 60
8	> 60 - ≥ 100

convert slope in percent to slope classes (Table 5). The classes chosen were those most frequently used in the Wise County soil survey to characterize soil types and phases. Figs. 5, 6, and 7 show slope, aspect, and topographic shape, respectively, for four maps in central Wise County.

Another relief attribute, distance to nearest ridge, was derived from the topographic shape attribute by the program DIST4, developed by Dr. Koeln. DIST4 computed for every cell in the study area the distance (in units equal to the width of a map cell) to the nearest cell classified as a ridge by program TOPO. This attribute was considered an indicator of the relative position of a cell on a slope. Fig. 8 shows distance to nearest ridge for four maps in central Wise County.

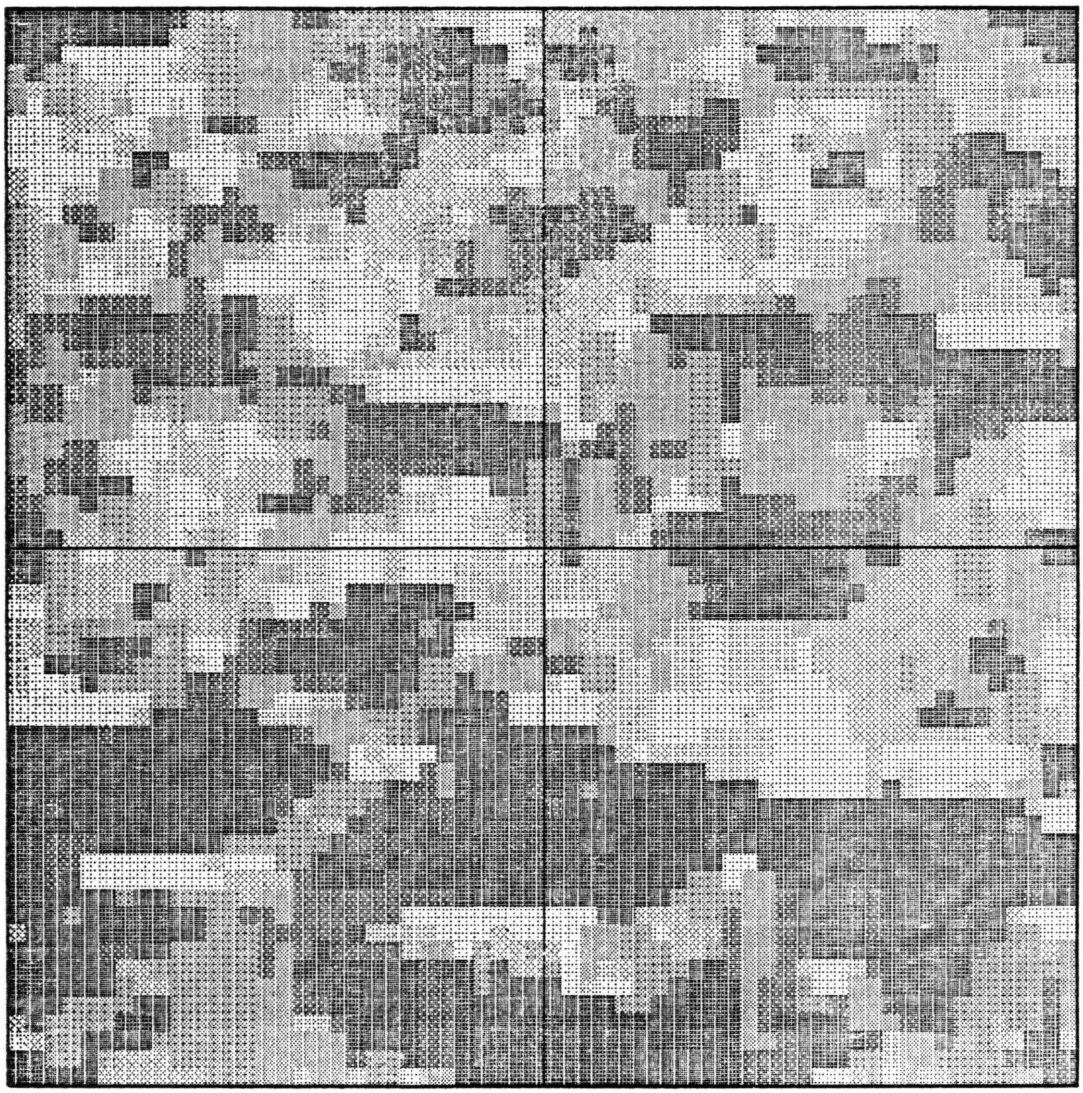
Streams

Small deposits of colluvium and alluvium occur along ephemeral and perennial streams in many parts of Wise County (Perry et al. 1954); therefore, information about the location of streams was used as an attribute for predicting soil. The stream courses of Wise and Dickenson Counties were digitized as lines, not cells, from USGS 1:24,000 topographic maps using the SEGMENT system developed by A. Blair Jones, III. Because ephemeral streams were mapped inconsistently in the study area, appearing fregently in



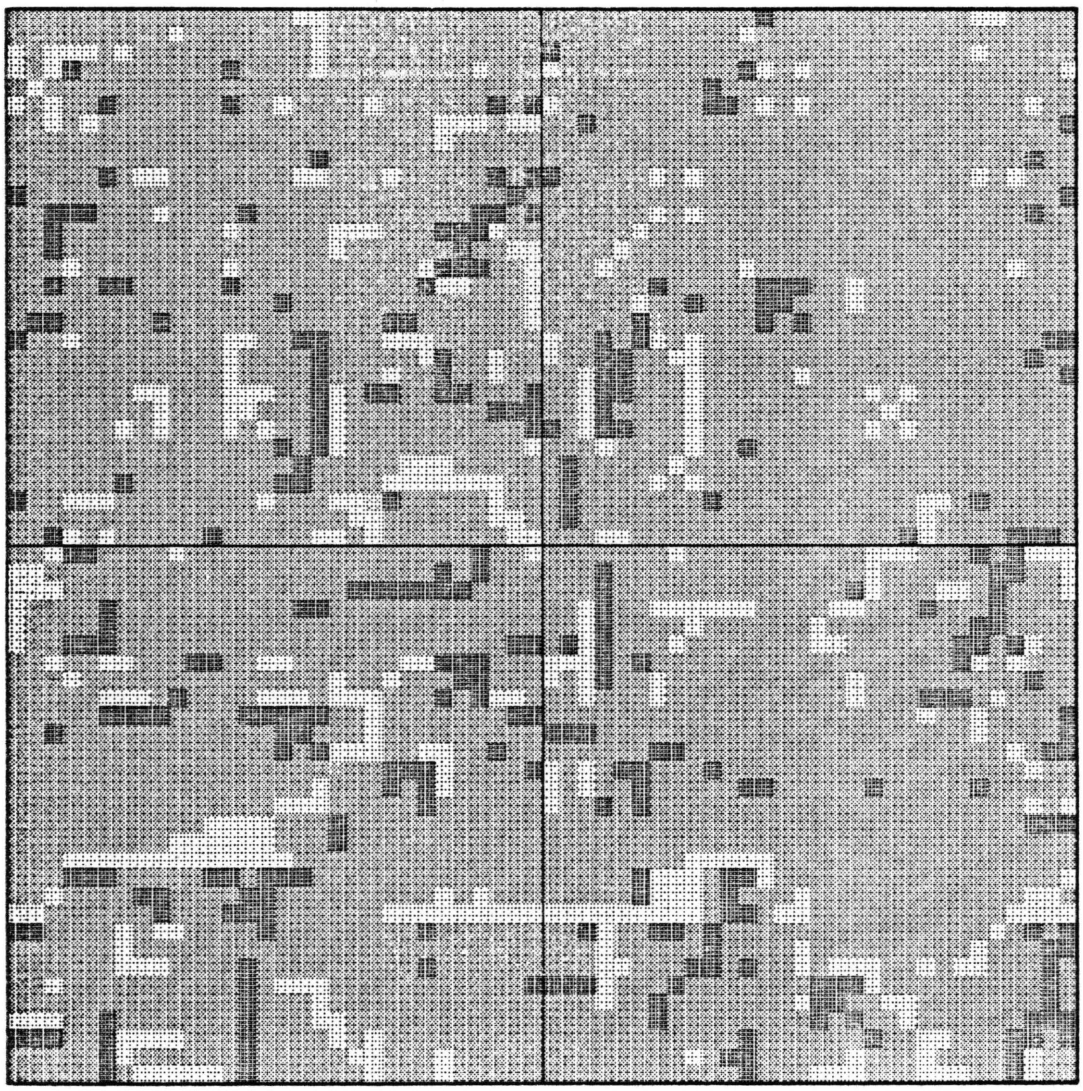
Slope Class (%)	# Cells	Tone	Slope Class (%)	# Cells	Tone
0-3	31		15-30	1399	
3-5	164		30-50	630	
5-10	589		50-60	67	
10-15	618		60-100	102	

Fig. 5. Map of slope classes for four POWER maps in Wise County. The scale is 1:48,000.



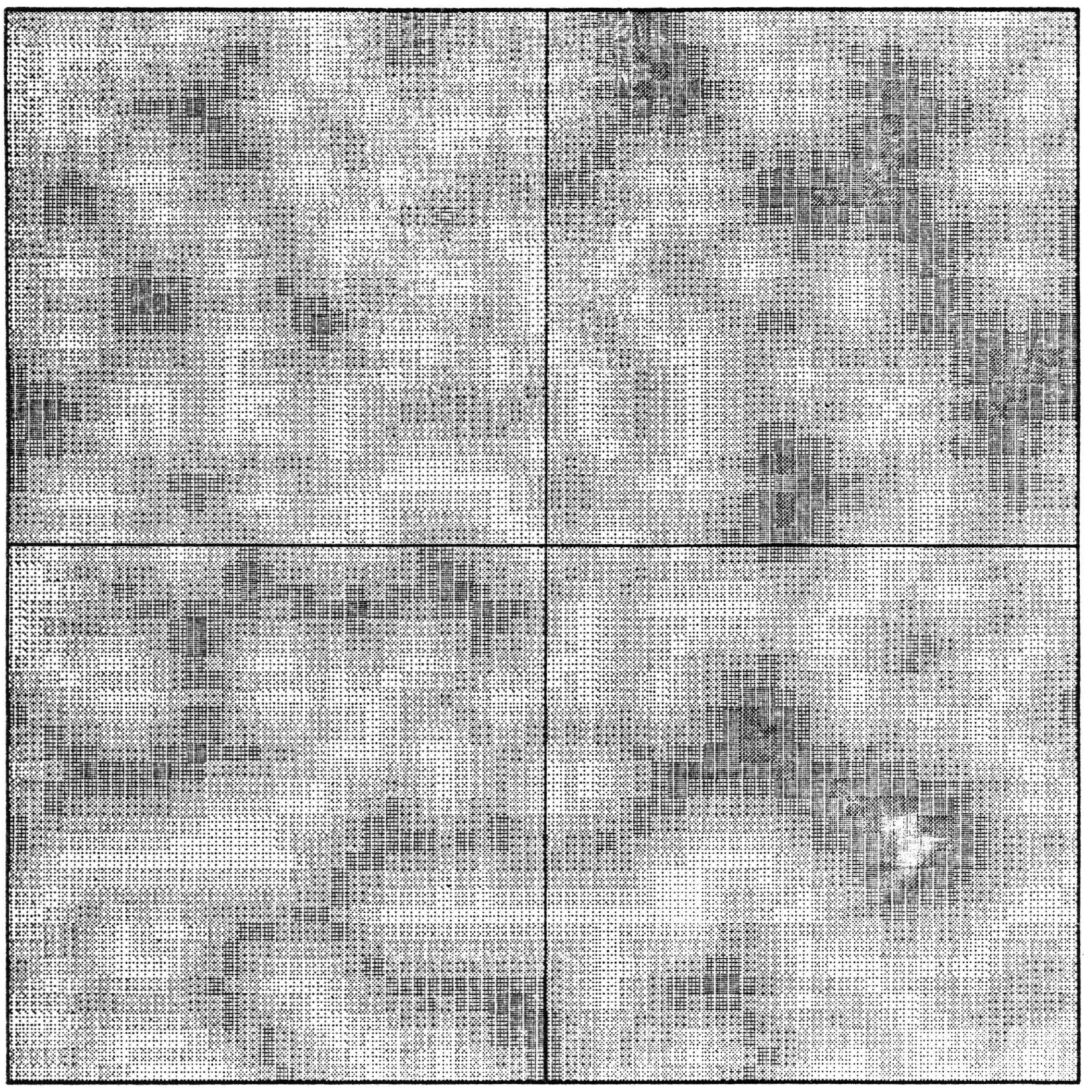
Aspect	# Cells	Tone	Aspect	# Cells	Tone
NE	280		SW	318	
E	426		W	506	
SE	366		NW	359	
S	529		N	816	

Fig. 6. Map of aspect classes for four POWER maps in Wise County. The scale is 1:48,000.



Topographic Shape	# Cells	Tone	Topographic Shape	# Cells	Tone
Slope	2889		Knob	1	
Ridge	400		Saddle	3	
Valley	307				

Fig. 7. Map of topographic shapes for four POWER maps in Wise County. The scale is 1:48,000.



Distance to Ridge (Cells)	# Cells	Tone	Distance to Ridge (Cells)	# Cells	Tone
0	400		4	379	
1	858		5	174	
2	939		6	45	
3	800		7 and greater	5	

Fig. 8. Map of distance to nearest ridge for four POWER maps in Wise County. The scale is 1:48,000.

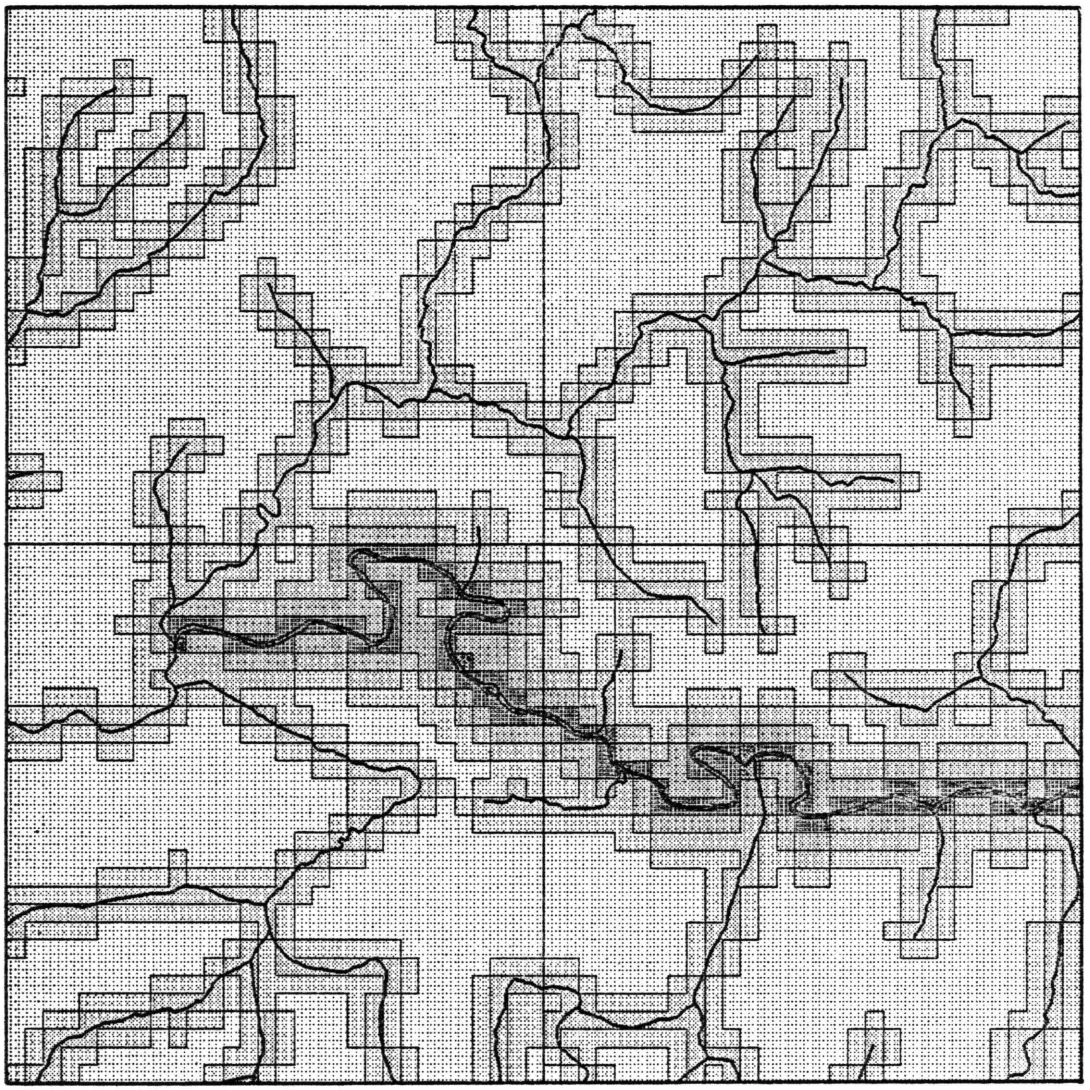
some quadrangles and not at all in others, only large and small perennial streams were digitized. Using Mr. Jones' program LINES, these data were converted to the grid-based POWER format.

The Wise County soil survey shows the soils of terraces and bottom lands generally occurring within 300 to 400 m of large streams and within 100 to 200 m of small streams. Distance to nearest stream was therefore considered a better attribute for predicting soil than mere presence or absence of streams within a cell. Program DIST4, developed by Dr. Koeln, was used to create separate mapfiles of distance-to-nearest-large-stream and distance-to-nearest-small-stream. To account for a nearby large stream having a greater influence on a cell than a nearby small stream, and for a large stream extending its influence farther than a small stream, values in the two distance mapfiles were converted to a scale expressing the relative degree of influence of a nearby stream as a soil-forming factor. This conversion, accomplished by program CATEGORY (Jones et al. 1976) is summarized in Table 6. A third mapfile was then created containing the larger cell values from the two overlaid mapfiles of categorized distances. Fig. 9 shows four maps from the final distance-to-streams mapfile.

In summary, data were collected or synthesized for the following map-cell attributes to use in program DIVIDE: 1)

Table 6. Categories defined for the relative degree of influence of a nearby large or small stream as a soil-forming factor.

Type Stream	Distance (in number of cells) 0	1	2	3	4 or more
Small	3	2	1	1	1
Large	5	4	3	2	1



Distance to Stream Category	# Cells	Tone	Distance to Stream Category	# Cells	Tone
1	1790		4	133	
2	919		5	103	
3	655				

Fig. 9. Map of the categories defined for distance to nearest stream for four POWER maps in Wise County. The scale is 1:48,000.

soil types, 2) surficial geologic formations, 3) slope, 4) aspect, 5) topographic shape, 6) distance to nearest ridge, and 7) distance to nearest stream.

THE CLASSIFICATION PROGRAM DIVIDE

As outlined earlier, a hierarchical, monothetic divisive classification strategy was followed. Because a reasonable approach to inferring soil conditions in an unsurveyed area is to use soil information available in a similar, surveyed area, the strategy then becomes further defined as an extrinsic classification. The program developed as a part of the POWER geographic information system, called DIVIDE, is to my knowledge the only example of such a strategy applied to geographic data.

DIVIDE is patterned, in part, after the hierarchical, divisive classification program MULTDIV, developed by Lance and Williams (1971). Both programs iteratively divide a population of individuals (in my case, a population of grid cells in the POWER system) into two dissimilar clusters until a predetermined number of clusters are derived. In DIVIDE, the clusters are groups of cells. Both programs select attributes by which to accomplish the divisions using a measure of dissimilarity based on the Shannon diversity index. Beyond these most basic considerations, however, the two programs are quite different. DIVIDE is designed as an

extrinsic classification in which the clusters derived are intended to reflect the nature of an external attribute declared in advance, whereas MULTDIV is an intrinsic classification in which all attributes are considered equivalent. Perhaps the most significant difference between the two programs is the fact that the input and output of DIVIDE are linked to a GIS. The FORTRAN IV source code for DIVIDE is given in Appendix A.

General Design of DIVIDE

Before explaining some the important details of DIVIDE, an overview of the program is presented. The design of program DIVIDE was greatly facilitated by the use of a structured programming technique called Warnier-Orr diagramming (Higgins 1979). The technique is a convenient shorthand for representing the basic features of program design: hierarchy, sequence, iteration, and decision. A Warnier-Orr diagram depicting the general processes performed by DIVIDE is given in Fig. 10. In this form of diagramming, hierarchy is symbolized by bracketing in a fashion similar to conventional set notation, except that right hand brackets are left off. Reading across the diagram from left to right shows how a complex process is broken into successively simpler component processes. Reading from top to bottom on the diagram shows the sequence

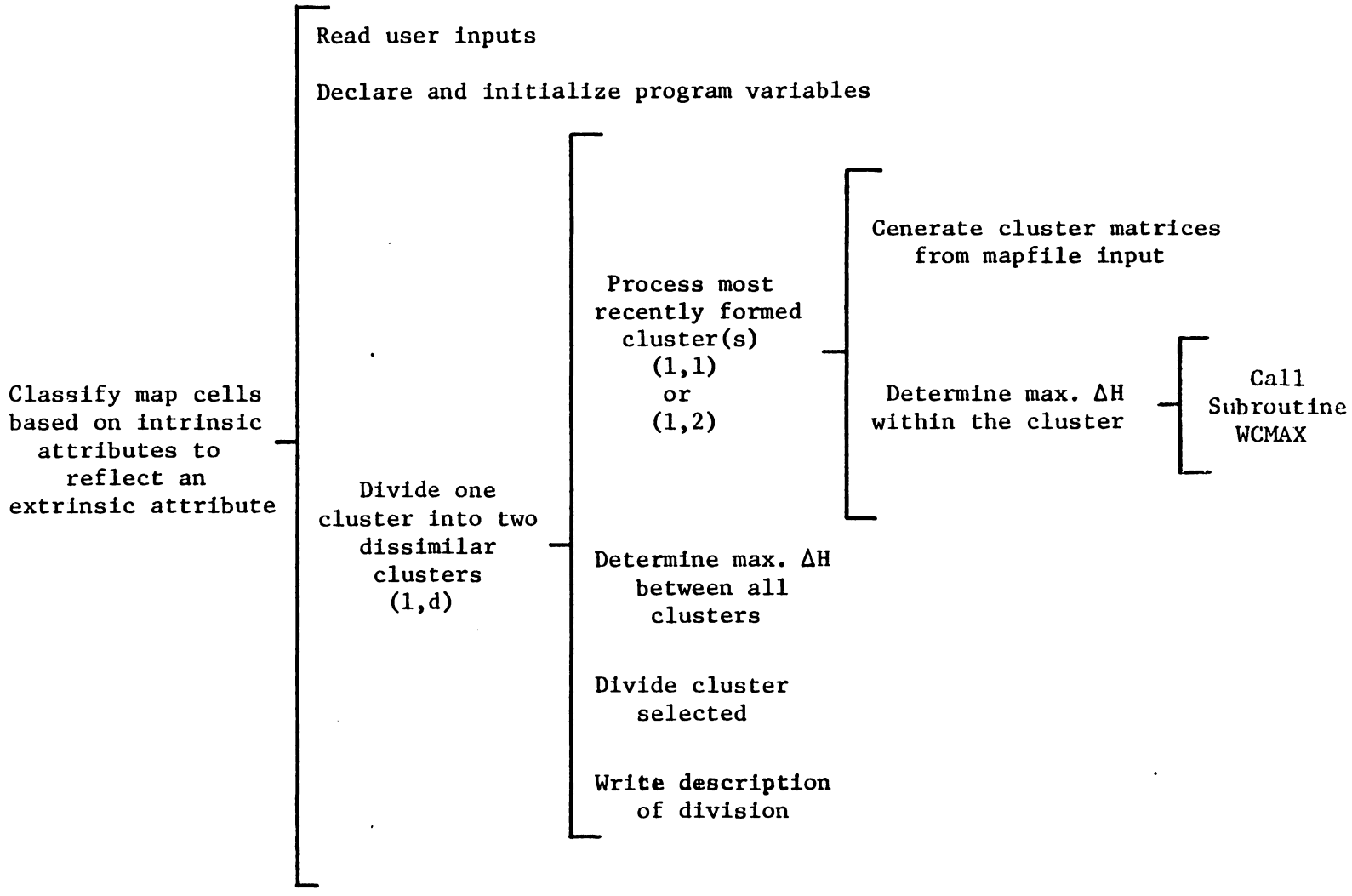


Fig.10. Warnier-Orr diagram depicting the general processes performed by program DIVIDE.

in which processes occur.

Iteration is symbolized by the notation "(1,d)". This is interpreted as "Repeat the process to the right from 1 to d times." In Fig. 10, d stands for the number of extrinsic attribute states to be "keyed out" minus 1. For example, to derive 26 clusters corresponding to 26 soil types in the study area, program DIVIDE would be required to "Divide one cluster into two dissimilar clusters" 25 times. This is the central process that DIVIDE performs.

Decision is symbolized simply by the word "or." The only example of decision in the general structure of DIVIDE given in Fig. 10 is that shown beneath the statement, "Process the most recently formed cluster(s)." The placement of the "or" here indicates that this process is to be performed (1,1) or (1,2) times. This is because during the first iteration of the higher order step, "Divide one cluster into two dissimilar clusters," only one cluster exists, i.e., the entire population of map cells. During all subsequent iterations of this step, however, two newly formed clusters are available for processing. Other features of Warnier-Orr diagramming are not shown in this example, but were used often in more detailed design diagrams. The interested reader is referred to Higgins (1979). Further reference will be made to Fig. 10 as some of the processes are explained.

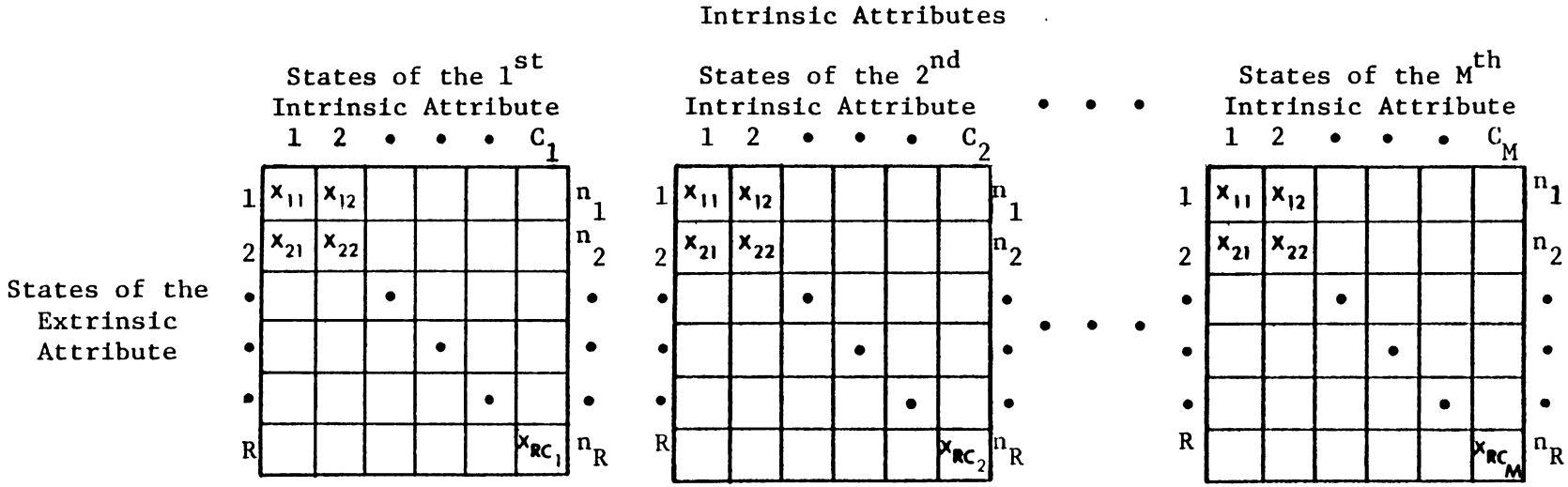
Input

DIVIDE reads and then restructures POWER system mapfile data into a series of matrices. This process is described as "Generate cluster matrices from mapfile input," in Fig. 10. Each of these matrices is a two-way table that opposes the extrinsic attribute, soil, and a particular intrinsic attribute, such as geology, topographic shape, or slope. Fig. 11 is a representation of these matrices. The geographic location of map cells is not incorporated in this restructuring, therefore a "label" mapfile separately stores the location and cluster affiliation of all cells.

Each entry in an extrinsic vs. intrinsic matrix represents the number of map cells that share the extrinsic attribute state (e.g., the soil type Muskingum stony loam) given by the matrix row and the intrinsic attribute state (e.g., the geologic formation Gladeville sandstone) given by the matrix column. These matrices generated from mapfile data are essentially contingency tables for pairs of attributes.

The Classification Process

Once all the matrices describing a cluster have been generated from mapfile input, the subroutine WCMAX (an acronym for "within cluster maximum") determines which particular combination of attribute states would produce the



M = the number of attributes (matrices)

R = the number of states (rows) of the extrinsic attribute

C_k = the number of states (columns) of the k^{th} intrinsic attribute; $k = 1, 2, \ldots, M$

X_{ij} = in a given matrix(k), the number of map cells having the i^{th} extrinsic attribute state and the j^{th} intrinsic attribute state; $i = 1, 2, \ldots, R$ and $j = 1, 2, \ldots, C_k$

n_i = the number of map cells having the i^{th} extrinsic attribute state; $i = 1, 2, \ldots, R$

Fig.11. Matrices showing how program DIVIDE restructures geographic data in the POWER system.

best possible division of that cluster. When this is determined for each of the two clusters formed as a result of the previous division, the best possible division must be selected from all existing clusters. The measure of "best" in both cases is an information theory measure of dissimilarity based on the Shannon diversity index, called _H. These processes are noted as "Determine max. _H within the cluster," and "Determine max. _H between all clusters," in Fig. 10

Derivation of _H

_H is derived from the Shannon diversity index, H. In program DIVIDE, the diversity of an extrinsic attribute vs. intrinsic attribute matrix is computed as:

$$H = N \log N - \sum_{i=1}^{R} n_i \log n_i$$

where: N = the total number of map cells in the matrix
R = the number of extrinsic attribute states (rows)
n(i) = the number of map cells per extrinsic attribute state

The magnitude of the index depends on the base of the logarithms used and on N, the number of individuals in the matrix. Throughout this work, logarithms to the base 10 have been used.

Small values of H indicate that most cells occur in a few extrinsic attribute states (rows) and large values

indicate that cells are evenly distributed among the different states. A matrix containing all rows (extrinsic attribute states) and columns (intrinsic attribute states) that describe a cluster is referred to as matrix C. The value of H computed for this matrix is denoted by H(c).

Subroutine WCMAX searches for the combination of attribute states that provides the best possible division. To accomplish this function, WCMAX systematically splits each C-matrix into two submatrices, A and B. The split is made by columns (intrinsic attribute states), e.g., geologic formations 2 and 3, vs. geologic formations 1, 4, and 5. The Shannon index is then computed for the two submatrices to give H(a) and H(b). The measure computed to compare this division with all other possible divisions is _H:

$$_H = H(c) - (H(a) + H(b))$$

_H is the loss in diversity resulting from the division of matrix C into the submatrices A and B. _H is larger for divisions that yield highly dissimilar A and B submatrices, submatrices in which soil types have been effectively separated. The particular A-B split which maximizes the value of _H is, for the intrinsic attribute under consideration, the best split for "keying out" the soil types. _H is the sole discriminating function in DIVIDE. It is analogous to the Chi-square test statistic as it is used in contingency-table analysis.

Special Data Requirements and Considerations

For intrinsic attributes possessing three or more equivalent states (e.g., geology and topographic position), DIVIDE computes _H for all possible A-B splits. Such attributes are termed non-ordered attributes, since any ordering of states is inappropriate. Attributes which are measurable on some discrete or continuous scale, on the other hand, are termed "ordered" attributes. Because a non-ordered attribute having k states can be divided into two clusters in $2^{k-1} - 1$ different ways, a reasonable limit had to be placed on k. I judged that 10 was a reasonable limit, but since the program structure for 10 will also accomodate 11, 11 became the actual limit to the number of states for non-ordered attributes in DIVIDE. Eleven states can be divided into two clusters in 1023 different ways.

For ordered attributes (e.g., slope, distance from nearest stream, elevation below nearest ridge), the range in values must be converted to an interval scale which, for convenience, is also limited to 11 states. This means that a continuous attribute, such as slope, must be converted to a discrete attribute, i.e., to slope classes. This process of categorization was discussed under the methods for data collection. Division of a cluster by an ordered attribute is considered only for single breaks in the range of intervals for that attribute. For example, slope classes 1

and 2 *vs.* slope classes 3, 4, 5, ...etc., is a valid split, but slope classes 1 and 3 *vs.* 2, 4, 5, ...etc., is not. This convention was adopted to prevent the illogical grouping of states at opposite ends of a continuum and to reduce computing costs.

Division of Clusters

That cluster which yields the maximum value of _H compared to all other clusters is always the next cluster divided in program DIVIDE. Fig. 10 notes this process as "Divide cluster selected." The division of clusters continues until a predetermined number of clusters has been derived. A division is implemented through the label mapfile by relabeling those cells belonging to the cluster selected for division. Those cells having the characteristics of the A submatrix are given one label and those having the characteristics of the B submatrix are given another label. The labeling scheme used is designed to show cluster relationships in the sequence of divisions. A cluster labeled X begets clusters labeled 2X and 2X + 1 upon division. From these labels, the pathway taken by any cluster through the hierarchical tree of divisions can be recontructed as a dendrogram. Dendrograms were used to display graphically the sequence of divisions.

Output

DIVIDE generates three forms of output:

1) Mapfile output. The label mapfile contains the cluster identity of all cells processed.

2) Division descriptions and summary of divisions. After each division of a cluster, DIVIDE reports the identity of the cluster, the division attribute, the value of _H, and the composition by soil types of the two clusters created by the division. Following all divisions, DIVIDE reports the labels of the final set of clusters derived.

3) Input for program KEYOUT. This is a condensed form of the second category of output that can be read by KEYOUT to apply the classification to the validation data set and to an unsurveyed area. KEYOUT is explained in more detail later. The output function of DIVIDE is noted in Fig. 10 as "Write description of division."

DIVIDE describes the extrinsic attribute composition of new clusters in three ways to facilitate a quick interpretation of the progress of the classification. For each extrinsic type occurring in a new cluster, each type identified by a numeric code and a four-letter acronym, DIVIDE reports the number of map cells, the percentage of the cluster, and the percentage of the total number of map cells of each type. Ideally, all final clusters would be composed of 100 percent of the map cells of one extrinsic

type, indicating a perfect classification.

OTHER PROGRAMS

KEYOUT

DIVIDE works by succesively dividing the population of map cells of an area in two clusters. Therefore the output of DIVIDE is essentially a dichotomous key. Program KEYOUT follows this dichotomous key to classify map cells in another area without reference to the extrinsic attribute used to construct the key. This other area should be described by the same compliment of attributes that describe the construction data set. KEYOUT creates a mapfile having cell values corresponding to the labels of the final clusters formed by DIVIDE. A cell of a given value in the KEYOUT mapfile is characterized by the same intrinsic attribute states as the cluster having that value as its label in the mapfile output of DIVIDE.

DVMTRX

Program DVMTRX (an acronym for "diversity matrix") was developed to assist in the preliminary analysis of intrinsic-extrinsic attribute relationships and in the evaluatation of results from program DIVIDE. DVMTRX provides an indication of the interdependence of two

geographic attributes in the POWER system based on diversity measures similar to those employed by program DIVIDE.

DVMTRX overlays the mapfiles for two attributes, X and Y, from the same study area and produces a matrix of attribute X vs. attribute Y. Each block in the matrix contains the number of map cells that share the X attribute state given by the matrix row and the Y attribute state given by the matrix column. As with program DIVIDE, continuous attributes should be transformed to an interval scale.

For each row and column of the matrix, DVMTRX computes diversity per map cell, H(i), which is the Shannon diversity index divided by the total number of map cells in the row or column (Clifford and Stevenson 1975:71). This measure approaches 0 as the association increases between a particular row or column and the opposing attribute. DVMTRX also computes an index, S.I., of overall similarity between the two attributes proposed by Estabrook (1967). The index S.I. assumes values near 1 when the two attributes are highly dependent, and values near 0 when the two attributes are virtually independent. For a derivation of this index, one that is consistent with the terminology used to discuss the diversity indices employed by program DIVIDE, the reader is referred to Clifford and Stevenson (1975:206-208).

CLASSACC

Program CLASSACC (an acronym for classification accuracy) was written specifically to report the percent agreement between the digitized extrinsic attribute mapfile and the predicted extrinsic attribute mapfile. Each cluster derived by program DIVIDE was assigned the soil type that dominated each cluster using program CATEGORY (Jones et al. 1976). This was done for both the construction data set mapfile created by DIVIDE and for the validation data set mapfile created by KEYOUT. CLASSACC was applied to both data sets to determine overall classification accuracy and accuracy per soil type. Overall accuracy is calculated as the total number of cells correctly classified divided by the total number of cells. Classification accuracy per soil type is calculated as the number of cells correctly classified per soil type divided by the number of cells per soil type.

RESULTS

Program DIVIDE was applied to two data sets, to the 54 maps of the construction data set, and to the one map, Map-043, in the Powell Valley area that was selected for a separate analysis using more detailed sources of data. The intrinsic attributes comprising both data sets were geology, slope, aspect, topographic shape, distance to nearest ridge, and distance to nearest stream.

THE CLASSIFICATION GENERATED FOR THE CONSTRUCTION DATA SET

Twenty-six soil types occurred in the construction data set. DIVIDE derived 26 clusters by successively dividing the population of map cells based on the intrinsic attributes. The dendrogram in Fig. 12 shows the sequence of divisions for the construction data set. The scale given along the left margin of the dendrogram is the value of _H (the decrease in diversity) associated with each division. Divisions are noted by horizontal lines in the dendrogram. The numbers appearing at the bottom margin of the dendrogram are the labels of terminal clusters, clusters that were not further divided.

Attributes selected for dividing clusters earlier in the sequence of divisions are more strongly associated with soil types than those selected later in the sequence of divisions. An attribute assuming more than two states can

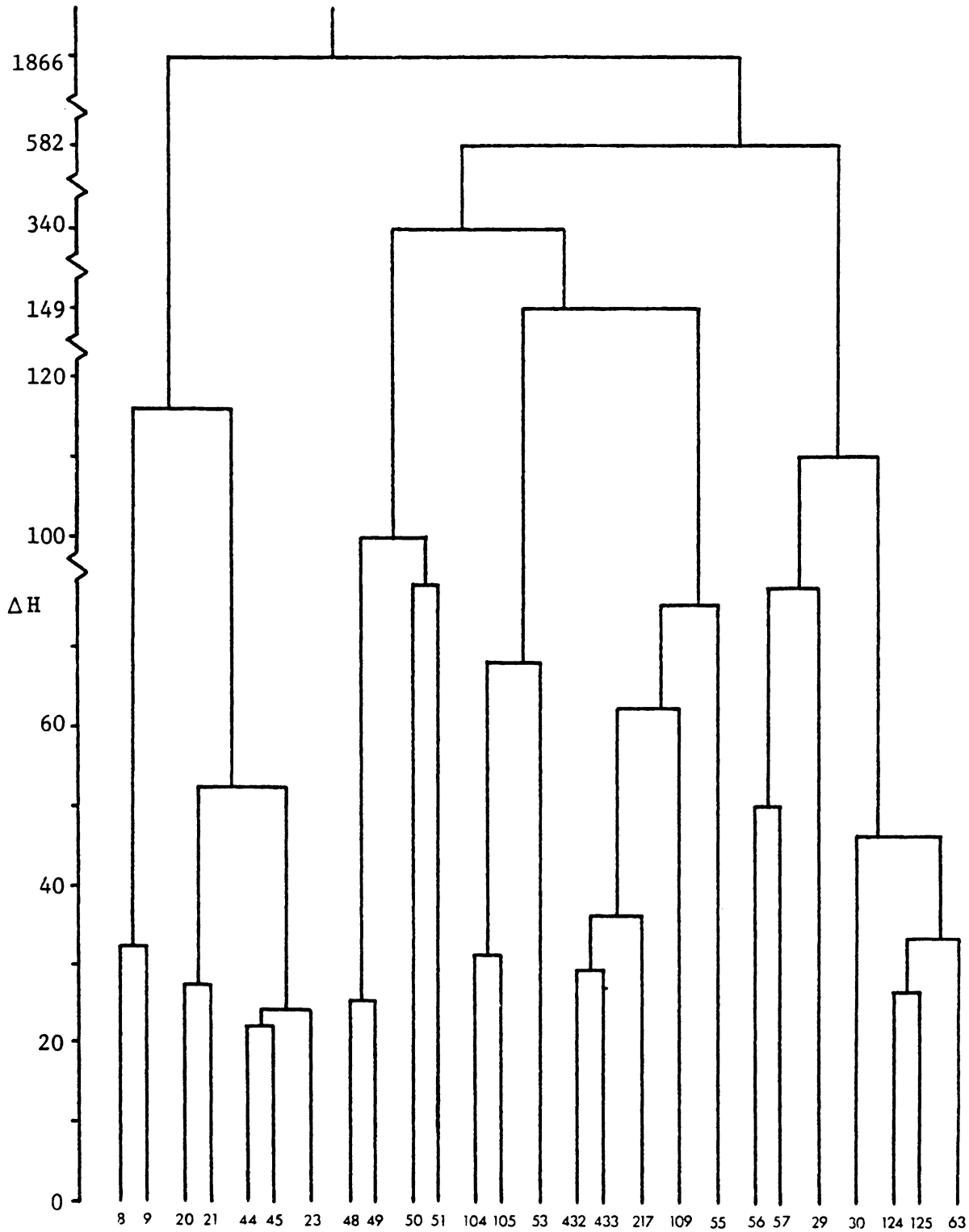


Fig. 12. Dendrogram showing the sequence of divisions for the construction data set. The numbers along the lower margin identify clusters created by the divisions.

be selected for more than one division. For example, the attribute, slope, which assumed eight different slope classes (Table 5), was selected for six divisions. In the third division (see Fig. 12), map cells having slope classes 1 though 4 were separated from those having slope classes 5 through 8. In the twelfth division, some of these map cells were further separated into clusters having slope classes 5 and 6 and those having slope classes 7 and 8. Geology proved the attribute most strongly associated with soil types in the construction data set. It was selected for the first division and for seven divisions thereafter. Distance to nearest stream was selected for the second division and for six subsequent divisions. Table 7 lists the attributes selected for all divisions.

Descriptions of the terminal clusters are based on constancy and fidelity of soil types. Complete constancy means that a cluster is composed entirely of map cells from one soil type. Complete fidelity means that all maps cells having a particular soil type occur in one cluster and in one cluster only. Ideally, if the intrinsic attributes are related strongly enough to soil types, the terminal clusters would be composed mostly of map cells from a single soil type, i.e., the clusters would exhibit both high constancy and high fidelity. Both properties are important to this classification. High constancy permits an easy association

Table 7. Attributes selected for divisions of the construction data set.

Division Number	Division Attribute	Division Number	Division Attribute
1	Geology	15	Geology
2	Distance to Stream	16	Distance to Stream
3	Slope	17	Distance to Stream
4	Geology	18	Distance to Ridge
5	Slope	19	Slope
6	Geology	20	Distance to Stream
7	Geology	21	Distance to Stream
8	Geology	22	Slope
9	Distance to Stream	23	Distance to Stream
10	Geology	24	Geology
11	Geology	25	Aspect
12	Slope		
13	Aspect		
14	Slope		

between a cluster and a soil type, whereas high fidelity demonstrates an effective discrimination between different soil types.

The clusters derived for the construction data set show an intermediate level of constancy. Although none of the terminal clusters derived for the construction data set were composed entirely of cells from one soil type, in 15 of the 26 clusters, greater than 50 percent of the cells comprising a cluster belonged to one soil type. The largest type in the county, Muskingum stony loam (hilly and steep phases combined), was the dominant type in 13 of these 15 clusters. The other two clusters consisted mostly of cells characterized as rough stony land (limestone material and Muskingum soil material units combined). The remaining clusters were of mixed composition. The composition and defining characteristics of all 26 clusters derived for the construction data set are given in Appendix B.

The clusters derived for the construction data set show low fidelity. All soil types in the construction data set occurred in more than one cluster, i.e., none of the types were fully characterized by the intrinsic attributes. Fig. 13 is a plot of the clusters for two of the four POWER maps shown in Fig. 3, which is a plot of digitized soils. Only two maps are shown because only two of the four were a part of the construction data set for which DIVIDE produced

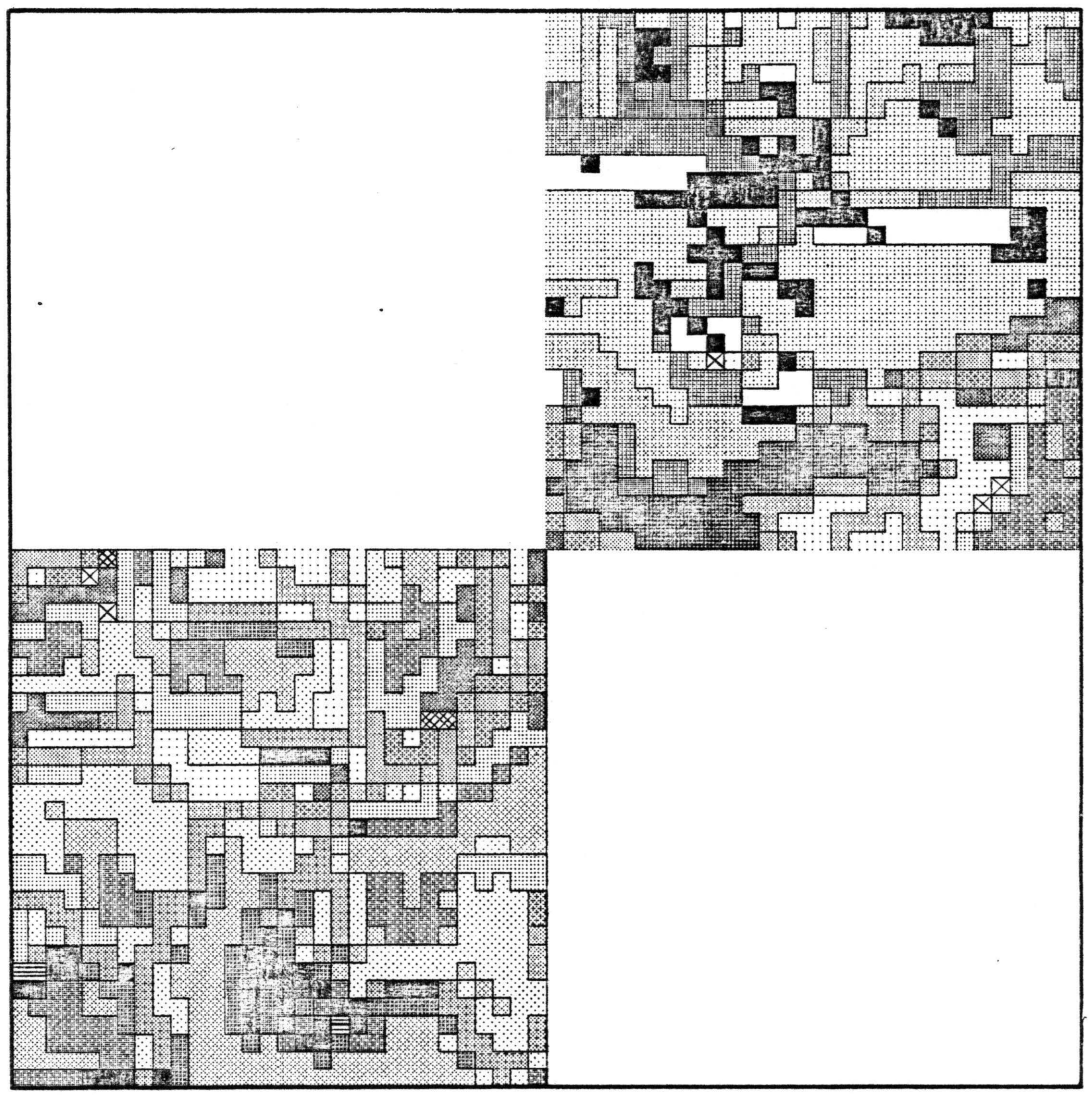


Fig. 13a. Map of DIVIDE clusters for two POWER maps in the construction data set. The scale is 1:48,000. A legend is given in Fig. 13b.

Cluster Label	No. Cells	Tone	Cluster Label	No. Cells	Tone
50	102		9	118	
51	171		104	5	
29	72		105	51	
55	108		432	284	
53	23		433	99	
109	3		20	21	
56	73		21	3	
57	52		124	158	
30	80		125	50	
217	57		48	81	
63	120		49	52	
8	17				

Fig. 13b. Legend for the map of DIVIDE clusters given in Fig. 13a.

clusters. Program DVMTRX was used to compare the digitized-soils mapfile with the mapfile of clusters generated by DIVIDE. The similarity index, S.I., calculated for this comparison was 0.34 on a scale of 0 to 1, where 1 is complete similarity. Low similarity, indicated by this low value of S.I., is evident when comparing Fig. 3 and Fig. 13. The areas in which the largest soil types occur in Fig. 3, Muskingum stony loam and Rough stony land, contain several different clusters in Fig. 13. This indicates that these two soil types are variable with respect to the intrinsic attributes, geology, slope, aspect, etc., since each cluster is defined by a unique set of intrinsic attribute states. In general, the map of clusters for the construction data set appeared more detailed than the map of digitized soils.

The diversity per map cell within each soil type, H(i), also computed by DVMTRX, is given in Table 8. H(i) approaches 0 as the association between a soil type and the DIVIDE clusters increases, i.e., as fidelity increases. For example, if all map cells having the soil type Philo loam occurred in the cluster labeled 15, the value of H(i) for i=Philo loam, would be 0. As an aid to interpreting Table 8, those soil types that occurred mostly in a single cluster (i.e., over 50 percent of the cells comprising a type) have values of H(i) less than 0.5.

Table 8. Diversity within each soil type with respect to the clusters derived for the construction data set.

Soil Type	H(i)
Atkins loam	1.21
Coeburn silt loam	1.08
Coeburn-Muskingum complex	1.13
Dekalb loam	0.47
Enders silt loam	1.00
Greendale fine sandy loam	0.46
Hagerstown silt loam	0.42
Hagerstown stony silt loam	0.54
Hartsells fine sandy loam	1.10
Holston fine sandy loam	0.92
Jefferson loam	0.99
Jefferson stony loam	1.15
Leadvale loam	1.20
Monogahela fine sandy loam	0.26
Muskingum stony loam	1.16
Muskingum fine sandy loam	1.13
Muskingum stony fine sandy loam	1.07
Philo loam	1.01
Pope fine sandy loam	0.99
Pope loamy fine sand	0.97
Rolling stony land	1.06
Rough stony land	1.08
Sequatchie fine sandy loam	0.84
Stony alluvium	0.96
Teas-Dandridge silt loam	0.46
Westmoreland silt loam	0.60

To compare further the mapfile of DIVIDE clusters with the mapfile of digitized soils, each of the DIVIDE clusters was assigned the soil type that dominated each cluster. Due to the mixed composition of the clusters, this resulted in all clusters being assigned to one of only three soil types that dominated the clusters. By this procedure, overall classification accuracy was 55.3 percent. The classification accuracy per soil type is given in Table 9. Fig. 14 shows the clusters interpreted as the dominant soil type of each cluster for the two maps shown in Fig. 13.

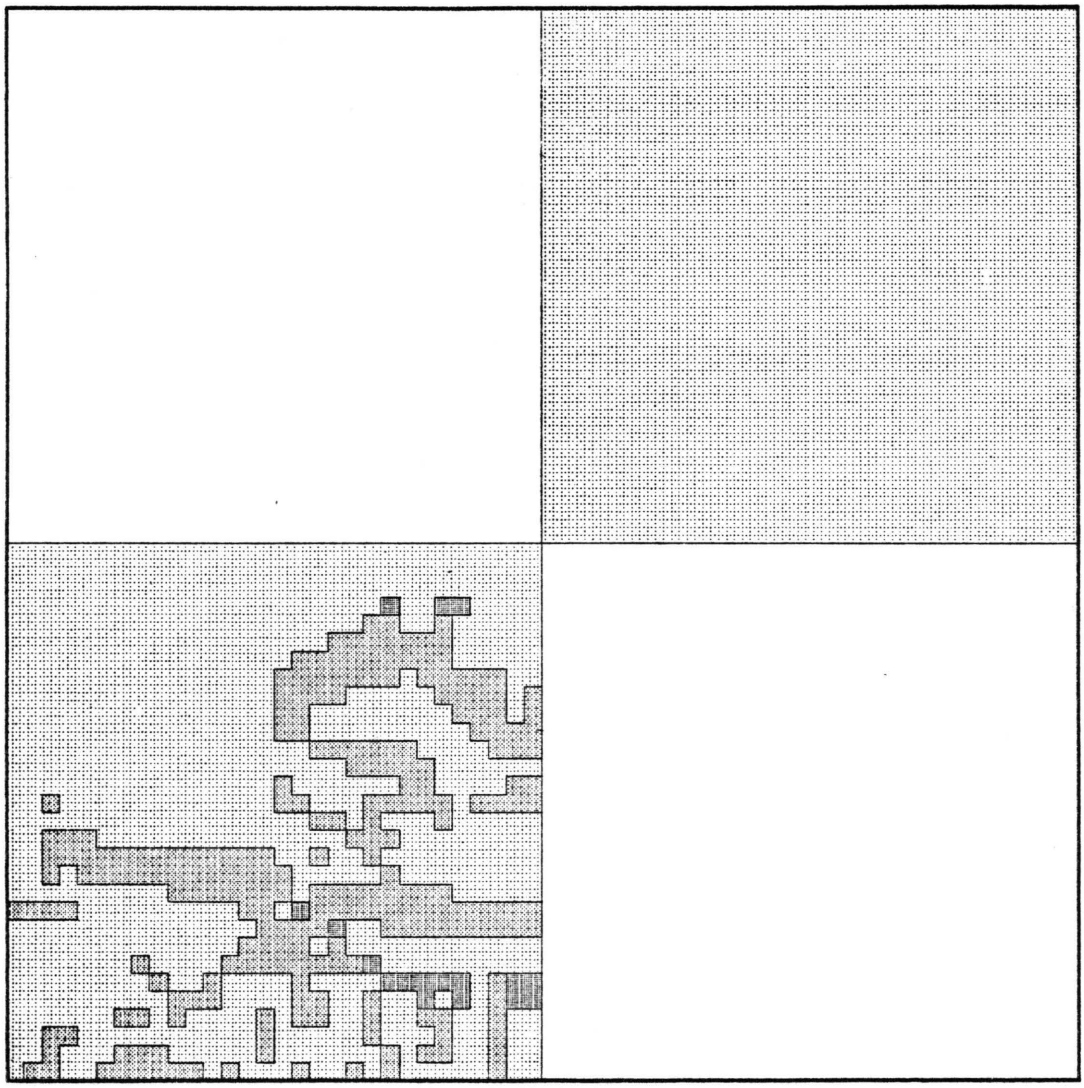
THE CLASSIFICATION GENERATED FOR MAP-043

Nineteen soil types occurred in Map-043. The sequence of divisions by which DIVIDE derived 19 clusters for this area is shown as a dendrogram in Fig. 15. Because Map-043 is a study area containing far fewer map cells than the construction data set, the values of _H given in Fig. 15 are much smaller than the values given in Fig. 12, the previous dendrogram.

Geology was the attribute selected for the first division of the Map-043 data set, but distance to nearest stream and aspect were each selected more frequently than geology. Distance to nearest ridge, which accounted for one division in the construction data set, was not selected as a

Table 9. Accuracy per soil type in the classification generated for the construction data set.

Soil Type	Accuracy %
Atkins loam	0.0
Coeburn silt loam	0.0
Coeburn-Muskingum complex	0.0
Dekalb loam	0.0
Enders silt loam	0.0
Greendale fine sandy loam	0.0
Hagerstown silt loam	0.0
Hagerstown stony silt loam	0.0
Hartsells fine sandy loam	0.0
Holston fine sandy loam	0.0
Jefferson loam	0.0
Jefferson stony loam	0.0
Leadvale loam	0.0
Monogahela fine sandy loam	0.0
Muskingum stony loam	90.72
Muskingum fine sandy loam	0.0
Muskingum stony fine sandy loam	10.01
Philo loam	0.0
Pope fine sandy loam	0.0
Pope loamy fine sand	0.0
Rolling stony land	0.0
Rough stony land	34.09
Sequatchie fine sandy loam	0.0
Stony alluvium	0.0
Teas-Dandridge silt loam	0.0
Westmoreland silt loam	0.0
Overall Classification Accuracy	55.3



DIVIDE Soil Type	# Cells	Tone
Muskingum stony loam	1551	
Muskingum stony fine sandy loam	17	
Rough stony land	232	

Fig. 14. Map of DIVIDE clusters assigned to predominate soil types in two maps of the construction data set. The scale is 1:48,000.

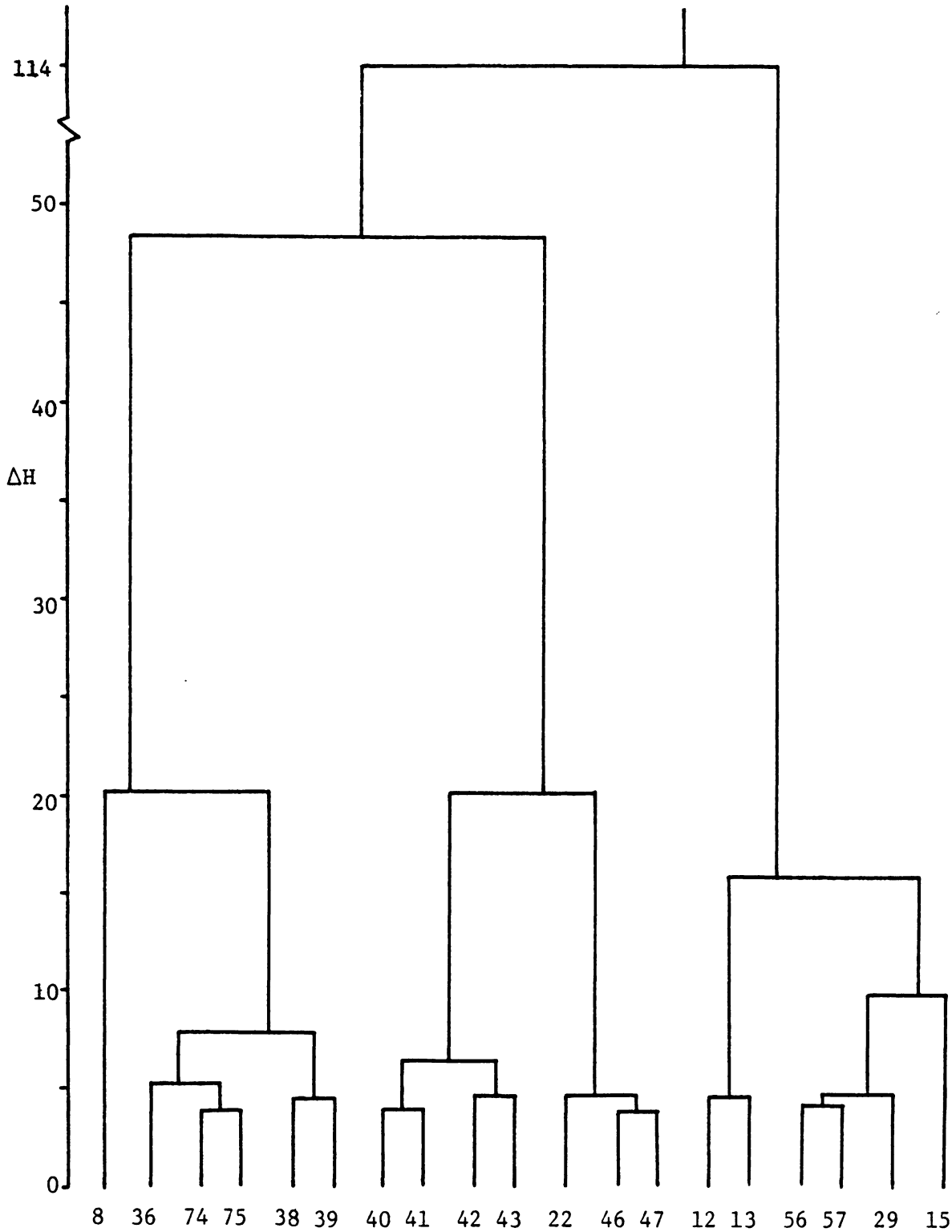


Fig. 15. Dendrogram showing the sequence of divisions for the Map-043 data set. The numbers along the lower margin identify clusters created by the divisions.

division attribute in the Map-043 data set. Table 10 lists the attributes selected for all divisions.

As in the classification generated for the construction data set, the classification generated for the Map-043 data set did not achieve complete constancy in an cluster. None of the terminal clusters were composed entirely of one soil type. In 12 of the 19 clusters, however, greater than 50 percent of the cells comprising a cluster belonged to one soil type. The composition and defining characteristics of the clusters derived for Map-043 are given in Appendix B.

DIVIDE produced a classification of higher fidelity in the Map-043 data set than in the construction data set. Four of the types occurring in Map-043 were fully characterized by the intrinsic attributes, i.e., four types showed complete fidelity. The similiarity index, S.I., calculated for the digitized soil types *vs.* the DIVIDE clusters was 0.63. This value is nearly twice the value computed for the comparison between digitized soil types and clusters in the construction data set. The greater similarity is evident when comparing Figs. 16 and 17. Fig. 16 is a plot of the digitized soil types for Map-043 and Fig. 17 is a plot of the clusters derived by DIVIDE for this area. As in the construction data set, though, the map of clusters for Map-043 appeared more detailed than the digitized soil map. Muskingum stony fine sandy loam, which

Table 10. Attributes selected for divisions of the Map-043 data set.

Division Number	Division Attribute	Division Number	Division Attribute
1	Geology	11	Distance to Stream
2	Geology	12	Distance to Stream
3	Distance to Stream	13	Distance to Stream
4	Geology	14	Aspect
5	Geology	15	Aspect
6	Slope	16	Topographic Shape
7	Distance to Stream	17	Topographic Shape
8	Slope	18	Aspect
9	Aspect		
10	Aspect		

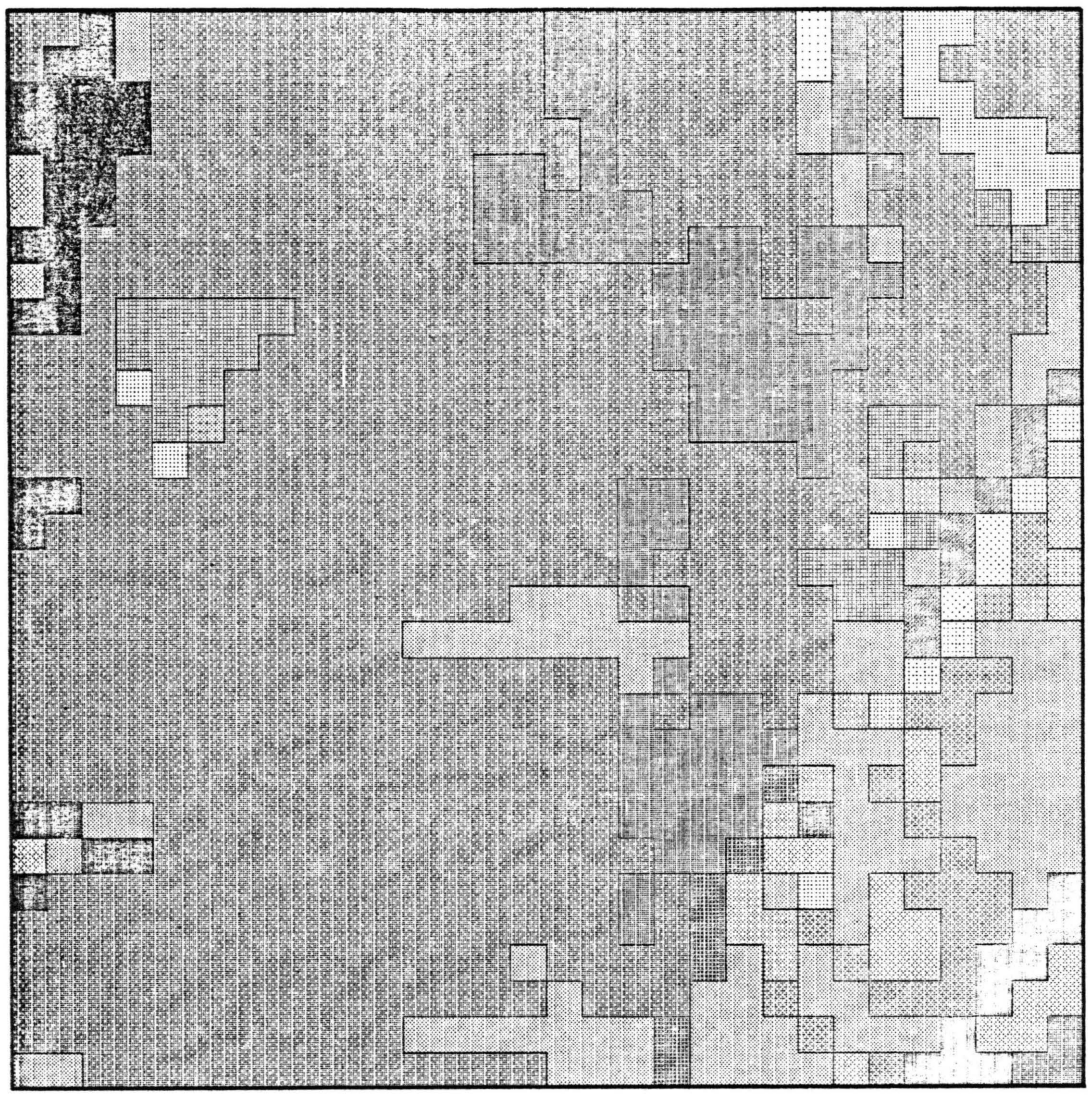


Fig. 16a. Map of digitized soil types in Map-043 of Wise County. The scale is 1:24,000. A legend is given in Fig. 16b.

Soil Type	No. Cells	Tone
Atkins loam	2	
Clarksville-Fullerton stony loam	3	
Fullerton loam	4	
Hartsells fine sandy loam	16	
Hayter loam	3	
Hayter stony loam	22	
Jefferson loam	9	
Jefferson stony loam	102	
Leadvale loam	3	
Montevallo shaly silt loam	24	
Muskingum fine sandy loam	25	
Muskingum stony fine sandy loam	551	
Muskingum-Litz silt loam	82	
Philo loam	7	
Pope fine sandy loam	8	
Rolling stony land	1	
Rough stony land	10	
Teas-Dandridge silt loam	8	
Westmoreland silt loam	20	

Fig. 16b. Legend to map of digitized soil types in Fig. 16a.

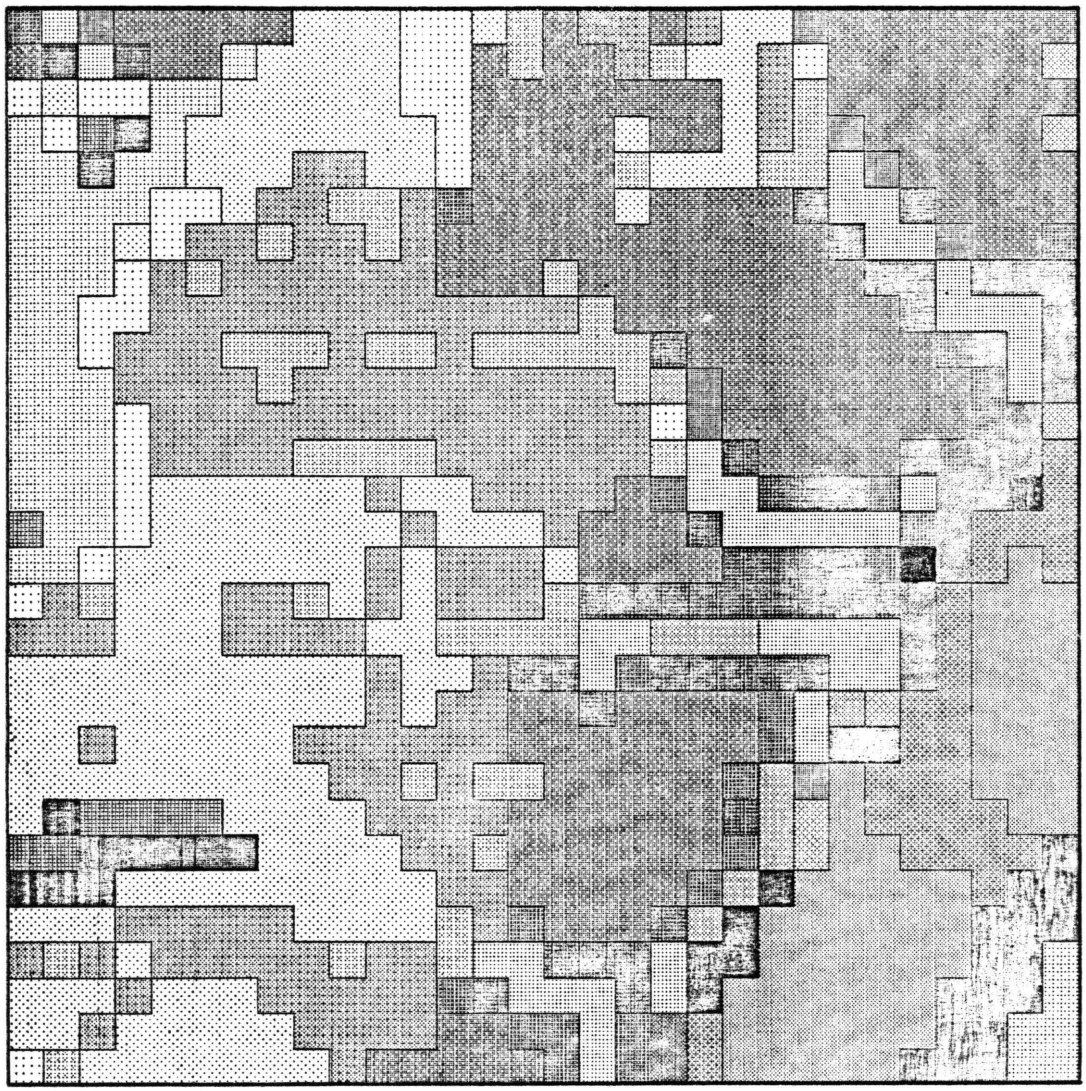


Fig. 17a. Map of DIVIDE clusters in the Map-043 data set. The scale is 1:24,000. A legend is given in Fig. 17b.

Cluster Label	No. Cells	Tone	Cluster Label	No. Cells	Tone
8	182		39	42	
15	16		56	10	
36	8		57	6	
42	28		40	153	
43	161		41	48	
29	14		46	8	
12	59		47	11	
13	28		74	24	
22	45		75	47	
38	10				

Fig. 17b. Legend to map of DIVIDE clusters in Fig. 17a.

occurs in slightly over half of the cells of Map-043, is represented by several clusters. For most other soil types, though, fidelity was generally high, as indicated by the near-zero values of diversity per soil type, H(i), given in Table 11.

By calling each cluster derived by DIVIDE the soil type that dominated each cluster, clusters were assigned to eight soil types out of the 19 occurring in Map-043. By this procedure, overall classification accuracy was 69.3 percent. The classification accuracy per soil soil type is given in Table 12. Fig. 18 shows the clusters for the Map-043 data set interpreted as the dominant soil type of each cluster.

PREDICTIVE ABILITY OF THE THE CLASSIFICATION

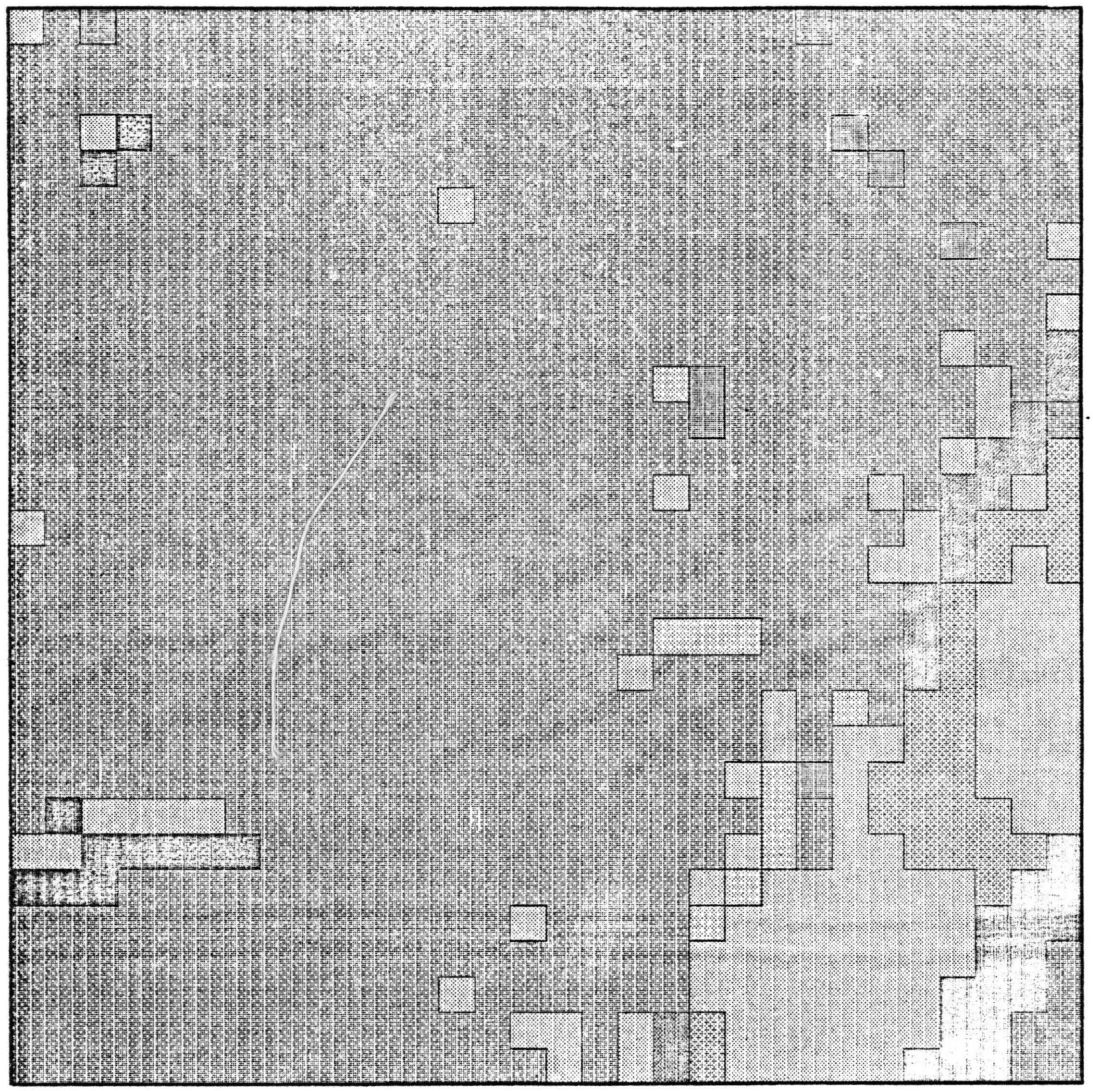
The classification generated by program DIVIDE for the construction data set was applied to the validation data set using program KEYOUT. KEYOUT determined the cluster to which each cell in the validation data set belonged, based on the same intrinsic attributes used in DIVIDE: geology, slope, aspect, topographic shape, distance to nearest ridge, and distance to nearest stream. The same measures used to compare clusters and soil types in the construction data set were then applied to clusters and soil types in the validation data set.

Table 11. Diversity within each soil type with respect to the clusters derived for the construction data set.

Soil Type	H(i)
Atkins loam	0.30
Clarksville-Fullerton stony loams	0.0
Fullerton loam	0.60
Hartsells fine sandy loam	0.36
Hayter loam	0.28
Hayter stony loam	0.72
Jefferson loam	0.64
Jefferson stony loam	0.94
Leadvale loam	0.48
Montevallo shaly silt loam	0.39
Muskingum fine sandy loam	0.66
Muskingum stony fine sandy loam	0.85
Muskingum-Litz silt loam	0.32
Philo loam	0.59
Pope fine sandy loam	0.0
Rolling stony land	0.0
Rough stony land	0.0
Teas-Dandridge silt loam	0.32
Westmoreland silt loam	0.65
Total cells	900

Table 12. Accuracy per soil type in the classification generated for the Map-043 data set.

Soil Type	Accuracy %
Atkins loam	0.0
Clarksville-Fullerton stony loams	0.0
Fullerton loam	0.0
Hartsells fine sandy loam	0.0
Hayter loam	0.0
Hayter stony loam	0.0
Jefferson loam	33.33
Jefferson stony loam	50.98
Leadvale loam	0.0
Montevallo shaly silt loam	37.50
Muskingum fine sandy loam	0.0
Muskingum stony fine sandy loam	96.55
Muskingum-Litz silt loam	4.88
Philo loam	0.0
Pope fine sandy loam	100.00
Rolling stony land	0.0
Rough stony land	100.00
Teas-Dandridge silt loam	75.00
Westmoreland silt loam	0.0
Overall classification accuracy	69.3



DIVIDE Soil Type	# Cells	Tone	DIVIDE Soil Type	# Cells	Tone
Jefferson loam	10		Muskingum-Litz silt loam	8	
Jefferson stony loam	101		Pope fine sandy loam	14	
Montevallo shaly silt loam	28		Rough stony land	16	
Muskingum stony fine sandy loam	712		Teas-Dandridge silt loam	11	

Fig. 18. Map of DIVIDE clusters assigned to predominate soil types in the Map-043 data set. The scale is 1:24000

Ten clusters defined by DIVIDE did not occur in the validation data set. Clusters are groups of map cells having a particular compliment of attribute states. The fact that ten particular compliments of attribute states were unique to the construction data set may be evidence that the checkerboard-like partitioning of the county into construction and validation data sets did not yield two highly similar data sets.

The similarity index, S.I., calculated for the digitized soil types *vs.* the clusters in the validation set was 0.22, indicating low similarity. The diversity per map cell within each soil type, H(i), is given in Table 13.

Clusters were assigned the same soil types in the validation data set as in the construction data set. The absence of ten clusters in the validation data set contributed to a reduction in the overall classification accuracy from 55.3 to 44.9 percent. The classification accuracy per soil type is given in Table 14.

Table 13. Diversity with each soil type with respect to the clusters derived for the validation data set.

Soil Type	H(i)
Atkins loam	0.97
Coeburn silt loam	0.81
Coeburn-Muskingum complex	0.80
Dekalb loam	0.11
Enders silt loam	0.60
Greendale fine sandy loam	0.0
Hagerstown silt loam	0.14
Hagerstown stony silt loam	0.33
Hartsells fine sandy loam	0.75
Holston fine sandy loam	0.76
Jefferson loam	1.01
Jefferson stony loam	0.97
Leadvale loam	0.99
Monogahela fine sandy loam	0.0
Muskingum stony loam	0.83
Muskingum fine sandy loam	0.86
Muskingum stony fine sandy loam	0.82
Philo loam	0.88
Pope fine sandy loam	0.92
Pope loamy fine sand	0.86
Rolling stony land	0.87
Rough stony land	0.75
Sequatchie fine sandy loam	0.86
Stony alluvium	0.86
Teas-Dandridge silt loam	0.28
Westmoreland silt loam	0.0

Table 14. Accuracy per soil type in the classification derived for the validation data set.

Soil Type	Accuracy %
Atkins loam	0.0
Coeburn silt loam	0.0
Coeburn-Muskingum complex	0.0
Dekalb loam	0.0
Enders silt loam	0.0
Greendale fine sandy loam	0.0
Hagerstown silt loam	0.0
Hagerstown stony silt loam	0.0
Hartsells fine sandy loam	0.0
Holston fine sandy loam	0.0
Jefferson loam	0.0
Jefferson stony loam	0.0
Leadvale loam	0.0
Monogahela fine sandy loam	0.0
Muskingum stony loam	100.00
Muskingum fine sandy loam	0.0
Muskingum stony fine sandy loam	0.0
Philo loam	0.0
Pope fine sandy loam	0.0
Pope loamy fine sand	0.0
Rolling stony land	0.0
Rough stony land	0.04
Sequatchie fine sandy loam	0.0
Stony alluvium	0.0
Teas-Dandridge silt loam	0.0
Westmoreland silt loam	0.0
Overall Classification Accuracy	44.9

DISCUSSION AND CONCLUSIONS

The classification program DIVIDE produced 26 clusters corresponding to 26 soil types for map cells comprising half of the Dickenson-like part of Wise County called the construction data set. The clusters were groups of cells sharing the same compliment of attribute states, i.e., the same geologic types, the same slope classes, etc. The clusters were derived by a series of steps designed to reflect the soil type of the greatest number of cells in the data set. If geologic types, slope classes, etc., were strongly associated with soil types, the clusters would have been composed mostly of cells having unique soil types. The set of attribute states defining such clusters could then be said to characterize that soil type and could be used to predict the occurrence of that soil type in the other half of the Dickenson-like part of Wise County, the validation data set. Given acceptable accuracy in the validation data set, the classification could then be applied in Dickenson County to produce a map of expected soil types, based on the similarity of this county with Wise County. This map could be checked in the field at a few randomly selected locations for further verification.

The classification generated for the construction data set was dominated by three of the 26 soil types that

occurred in the Dickenson-like part of Wise County: Muskingum stony loam, Muskingum stony fine sandy loam, and Rough stony land. These three types comprise over half of the mountainous terrain of Wise County and probably of Dickenson County as well. This is why DIVIDE achieved 55.3 percent classification accuracy while classifying only three of the 26 soil types. High variability within these large soil types with respect to geologic types, slope classes, etc., prevented an effective discrimination of other soil types by the classification.

Unfortunately, the greatest utility for this project in Dickenson County would have been an ability to identify the other 23 soil types, which are types exhibiting a fairly broad range of opportunities and restrictions for the uses and management of land. The three types which dominated the classification are typically unsuited for most important land uses due to excessive stoniness or slope.

The absence of ten clusters defined for the construction data set in the validation data set was evidence of unexpected disimilarity between the two data sets. Given that the two data sets contained nearly equal proportions of all attribute states considered singly and that both were contained within the same region, it seems unlikely that several combinations of attribute states would have occurred in one and not the other. The absence of

several clusters may constitute further evidence of inaccuracies in the data, or it may suggest that it is difficult to apply information derived from one area to another area believed similar. Given these findings for the validation data set and the composition of clusters in both data sets, I judged it inappropriate to apply the classification to Dickenson County.

There are three directions to take as possible explanations of the results of the classification:
1) The inability of program DIVIDE to effectively differentiate soil types may be an artifact of the methods of data collection or of the classification strategy itself;
2) Inaccuracies in the sources of data could be responsible; or
3) perhaps the results are simply a reality of the attributes, i.e., soil types as presently described and mapped are not well associated with geologic types, slope classes, etc.

The coding rule established for digitizing the soil and geologic maps may not have been the rule best suited to the problem of classifying soils. As outlined in the Methods, a modified majority rule was adopted to avoid excluding many of the soil and geologic mapping units which occurred as small islands or narrow strips throughout the study area. Although this coding scheme produced digitized maps that

more closely resembled the source maps, it may have made it more difficult to classify cells based on attributes which were not digitized in this manner. Most of the soil-related attributes were derived from elevation data. Elevation is coded for the centroid of a cell. In order to achieve a more consistent data set, one in which all attributes pertain precisely to the same feature of a cell, it might have been better to code soil and geologic mapping units for the centroid of a cell as well. Point digitizing of areal data is not common practice. Since the 3-acre cell size was comparable to the minimum mapping unit of the source map, though, it would not have been inappropriate, and it may have produced a better set.

The most obvious deficiency in the data concerns the attributes derived from elevation. Elevation data for 3-acre cells was interpolated from existing 27-acre cells data. There were demonstrated inaccuracies in these data which provided the basis for modeling slope, aspect, topographic shape, and distance to nearest ridge. These inaccuracies may have been of significant magnitude and frequency to effect serious errors in the terrain models. Collecting more accurate 3-acre elevation data for the entire study area was beyond the means of this project.

In an attempt to separate the possible effects of inaccurate data from the methodology itself, a small study

area called Map-043 was established in the southwestern part of Wise County. The source of geologic data for Wise County as a whole was a 1:62500 scale map, but a more detailed 1:24000 scale map was available for Map-043. In addition, elevation data digitized directly for 3-acre cells was used rather than data interpolated from 27-acre cells. Program DIVIDE achieved greater success in differentiating soil types with the Map-043 data set than with the construction data set. It was possible to map eight of 19 soil types in Map-043 as opposed to three of 26 in the construction data set, and classification accuracy was increased from 55.3 percent to 69.3 percent. This could have been due to stronger associations between soils and the other attributes in Map-043, to more accurate and more precise data, both, or to other unknown causes.

The classification for Map-043, although an improvement over the classification for the construction data set, was also charcterized by the dominance of a large soil type in several clusters. Consequently, the map of clusters appeared more detailed than the corresponding map of soil types. Differences between cells with respect to topographic shape, distance from stream, etc., did not necessarily correspond to differences in the soil. Part of this phenomenon may have been a function of the coding scheme established for digitizing the soil maps. Soil types

were digitized instead of soil phases, which are defined by differences in slope and relief. This coding scheme was elected when I believed that elevation data from the Digital Terrain Tapes would provide adequate 3-acre topographic data. It is possible, however, that there are differences between between soil phases belonging to the same soil type beyond acknowledged differences in slope and relief. The occurrence of a soil type in more than one cluster may reflect such differences.

Lietzke and Porter (1975:2) note that many of the soil taxa described in the Wise County soil survey, and other surveys of the time, have since been split into two or more taxa based on greater knowledge of soil properties. Although soil mapping units are seldom, if ever, delineated containing 100 percent pure taxonomic units, the maping units of the Wise survey may be too heterogeneous (in terms of soil individuals) to characterize by other attributes. In addition, the 1:24000 mapping scale of the survey may have been too small to map sufficiently homogeneous mapping units. Nevertheless, the Wise survey constitutes the best soil data available for Wise County and probably the best model available for the soils of Dickenson County. In order to facilitate the association of derived soil units with known soil properties and interpretations, methods were developed to disclose relationships that might accurately

reproduce the Wise survey for extension to Dickenson County. This approach was necessary to avoid having to characterize derived soil units by extensive field sampling, which is the work of a soil survey.

Although problems with the sources of data certainly account for part of the inability to adequately classify soil types, the results also suggest a weakness in the methodology itself. DIVIDE is a hierarchical monothetic divisive classification program. By a series of divisions, based on one attribute at a time selected from several attributes, DIVIDE attempts to "key-out" cells in the POWER system to reflect another attribute, soil, set apart from the rest. Because resemblence between cells was based on only one attribute per division, DIVIDE was prone to separate cells belonging to large, variable soil types. Cells belonging to a particular soil type, but lacking a division attribute common to other cells of that type, frequently became members of different clusters. This monothetic aspect of the classification would not have presented a problem if soil types had proven more homogeneous with respect to more attributes. Whether due to a deficiency in the quality of the data or simply a reality of the attributes, most soil types proved too heterogeneous, in terms of the attributes geology, slope, etc., to be accurately classified by a monothetic strategy.

A polythetic strategy would probably have been better suited to the problem, but computer algorithms for polythetic classifications are much more complex than for monothetic classifications, and have been developed only recently (Clifford and Stevenson 1975:100). The adaptation or development of a polythetic algorithm for use with geographic data of the POWER system was beyond the means of this project. Program DIVIDE could be modified to reallocate apparently missclassified cells in a polythetic fashion, but this would require much further work with no guarantee of success.

It would have been possible to allow program DIVIDE to continue the division of clusters in an attempt to create smaller clusters containing fewer soil types. Similar clusters of less mixed composition could then have been combined to yield a better classification. The dendrograms for the two classifications generated by DIVIDE, however, show how the final divisions were accomplishing only small decreases in ϕH, the measure of disimilarity. These divisions were "splitting hairs" and further divisions would not likely have improved the differentiation of soil types. It would slso have been possible to exclude the larger, more heterogeneous soil types from the classification and perhaps then achieve greater accuracy with the remaining types. The results of such a classification, however, would be

difficult to apply in an unsurveyed area, since the characteristics of the excluded soil types cannot be readily distinguished from those not excluded.

Although unsuccessful in its application with the soils of southwestern Virginia, the methods developed in this thesis nevertheless overcame the major limitation of previous approaches: soil data from a surveyed area is incorporated directly into an automated process for producing expected soil units in a similar unsurveyed area. The program which accomplishes this process, DIVIDE, holds promise for other applications with geographic data. DIVIDE provides an objective and efficient means for determining how various geographic attributes can be combined to disclose information about a particular attribute of interest. DIVIDE and its sister programs KEYOUT, DVMTRX, and CLASSACC, are general-purpose tools, i.e., they are suited for use with other POWER data sets. At the time of this writing, another researcher working with the POWER system is preparing to use DIVIDE to predict forest cover types and wildlife habitat suitability based on data collected for soils, topography, and climatic parameters (M. L. Fies, pers. comm.). Program DIVIDE is capable of processing large data sets containing both quantitative and nonquantitative geographic attributes. It is not necessary to weight or assign priorities to different attributes. The

results can be objectively evaluated in terms of classification accuracy and are readily tested in a validation data set.

LITERATURE CITED

Acton, D.F. 1965. Relationships of pattern and gradient of slope to soil type. Canadian J. Soil Sci. 45:96-103.

Allen, N. 1973. Computer mapping for land use planning, COMLUP. U.S. Forest Service, Intermountain Region, 140 pp.

Bartelli, L.J., A.A. Klingebiel, J.V. Baird, and M.R. Heddleson. 1966. Soil surveys and land use planning. Soil Sci. Soc. of Am. and Am. Soc. of Agronomy. Madison, Wisconsin. 196 pp.

Bauer, K.W. 1973. The use of soils data in regional planning. Geoderma, 10:1-26.

Beatty, M.T., G.W. Peterson and L.D. Swindale (eds.). 1979. Planning the uses and management of land. Am. Soc. Agronomy, Crop Sci. Soc. of Am., Soil Sci. Soc. of Am., publishers. 1028 pp.

Buol, S.W., F.D. Hole and J.R. McCracken. 1980. Soil genesis and classification (2nd ed.). Iowa State Univ. Press, Ames, Iowa. 404 pp.

Butler, K.S., B.H. McCown, and W.A. Gates. 1977. Use of a computer-based resource data system to evaluate on-site waste disposal alternatives. J. Soil and Water Cons. 32(5):214-219.

Cason, T.W.,Jr. 1980. PHAEDRUS: a system for the preliminary evaluation of large physical projects. M.S. Thesis, Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 306pp.

Calkins, H. and D.F. Marble. 1980. Computer software for spatial data handling: vol. 1, full geographic informatoin systems. International Geographical Union, Commission on Geographical Data Sensing and Processing. Ottawa, Ontario. 272 pp.

Calkins, H. and R.F. Tomlinson. 1977. Geographic information systems, methods and equipment for land use planning. U.S. Geologic Survey. Wash., D.C. 88 pp.

Clifford, H.T., and W. Stephenson. 1975. An introduction to numerical classification. Academic Press, New York. 229 pp.

Croswell, P. 1981. Automated soil modeling with the Kentucky Natural Resources Information System. Paper presented at annual meeting of the Assoc. Am. Geographers. 9 pp.

Davis, J.C. and M. McCullagh (eds.). 1975. Display and analysis of spatial data. John Wiley and Sons, Ltd. London. 287 pp.

Durfee, R.C. 1974. ORRMIS: Oak Ridge regional modeling information system. Oak Ridge National Laboratory. Oak Ridge, Tennessee. 98 pp.

Eby, J.B. 1923. The geology and mineral resources of Wise County and the coal-bearing portion of Scott County, Virginia. Va. Geologic Survey Bull. No. 24. Charlottesville, Va. 617 pp.

Estabrook, G.F. 1967. An information theory model for character analysis. Taxonomy 16:8697.

Fabos, J.G. and S.J. Caswell. 1976. Composite landscape Assessment. Research Bull. No. 637, Massachusetts Ag. Exper. Station. Amherst, Mass. 323 pp.

Ferris, K.H. and J.G. Fabos. 1974. The utility of computers in landscape planning: the selection and application of a computer mapping and assessment system for the metropolitan landscape planning model (METLAND). Research Bull. No. 617, Massachusetts Ag. Exper. Sta. Amherst, Mass. 116 pp.

Giles, A.W. 1921. The geology and coal resources of Dickenson County, Virginia. Va. Geologic Survey Bull. No. 21. Charlottesville, Va. 224 pp.

Giles, R.H., Jr., A.B. Jones III, and C.W. Smart. 1976. POWER: a computer system for corridor location. Office of Biological Services, U.S. Fish and Wildl. Service, Wash., D.C. 30 pp.

Grender, G.C. 1976. TOPO III. A FORTRAN program for terrain analysis. Computers and Geosciences. 2:195-209.

Hamm, C.P. 1978. A soil information algorithm for the Appalachian Ridge and Valley Province of Virginia. M.S. Thesis. Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 77 pp.

Higgins, D.A. 1979. Program design and construction. Prentice-Hall, Inc. New Jersey. 189 pp.

Hoar, A.R. 1980. A methodology for mapping probable ranges of endangered terrestrial mammals within selected areas of Virginia. M.S. Thesis, Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 191 pp.

Honeycutt, D.M., M. Brooks, and A.J. Kimerling. 1980. Geographic information systems: a review of selected operational and functional capabilities. Oregon State Univ. Corvalis, Ore. 72 pp.

Hunter, W.R., C.W. Tipps, and J.R. Coover. 1966. Use of soil maps by city officials for operational planning. Pages 31-36 In L.J. Bartelli (eds.), Soil surveys and land-use planning. Am. Soc. of Agronomists. Madison, Wisconsin. 196 pp.

Jansen, I.J. and T.E. Fenton. 1978. Computer processing of soil survey information. J. Soil and Water Conservation. 33(4): 188-190.

Jenny, H.J. 1941. Factors of soil formation. McGraw-Hill, New York. 281 pp.

Jones, A.B., III. 1976. POWER: a computer information system for land-use decisions. M.S. Thesis. Virginia Polytechnic Inst. and State Univ., Blacksburg, Va. 194 pp.

Jones, A.B. III, C.W. Smart, and R.H. Giles, Jr. 1976. Operating manual for POWER: a computer system for corridor location and land-use decisions. Dept. of Fisheries and Wildlife Sciences, Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 406 pp.

Kennedy, M. and C. Guinn. 1975. Automated spatial information systems: avoiding failure. Urban Studies Center, Louisville, Kentucky. 83 pp.

Knapp, E.M., and D. Rider. 1979. Automated geographic information systems and Landsat data: a survey. Pages 57-68 In Computer mapping in natural resources and the environment. Laboratory for Computer Graphics and Spatial Data Analysis, Harvard Univ. Cambridge, Massachusetts.

Koeln, G.T. 1980. A computer-assisted general aviation airport location and evaluation system for Virginia. PhD Dissertation. Virginia Polytechnic Inst. and State Univ., Blacksburg, Va. 235 pp.

Koeln, G.T., A.B. Jones, and R.H. Giles, Jr. 1981. Virginia's geographic information systems. Unpublished manuscript, Commonwealth Data Base, Dept. of Taxation, Richmond, Va. 187 pp.

Lance, G.N. and W.T. Williams. 1967. Mixed-data classificatory programs, I. Agglomeration systems. Australian Computer J. 1:15-20.

Lance, G.N. and W.T. Williams. 1968a. Mixed-data classificatory programs, II. Divisive systems. Australian Computer J. 1:82-85.

Lance, G.N. and W.T. Williams. 1968b. Note on a new information-statistic classificatory program. Computer J. 11:195.

Lance, G.N. and W.T. Williams. 1971. A note on a new divisive classificatory program for mixed data. Computer J. 14:154-155.

Lawrence, G.E., Jr. 1976. A computer-based insolation mapping algorithm for large areas. M.S. Thesis, Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 129 pp.

Lietzke, D.A. and H.C. Porter. 1978. Interpretive guide to the soils of southwest Virginia TVA counties. Extension Division, Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 129 pp.

Lyle, J. and M. von Wodtke. 1974. An information system for environmental planning. J. Am. Assoc. Planners. 40(6):394-413.

Macnaughton-Smith, P. 1965. Some statistical and other numerical techniques for classifying individuals. Home Office. Studies in the causes of delinquency and the treatment of offenders. No. 6. HM Stationery Office, London. 33 pp.

McCann, M.R. 1976. The application of computer mapping to soils and land-use planning. M.S. Thesis, Univ. of Arizona. 55 pp.

Meyers, C.R., R.C. Durfee and T.C. Tucker. 1974. Computer augmentation of soil survey interpretation for regional planning applications. Proc. Annual Meeting, Soil Conservation Soc. of Am., 29:129-138.

Miller, R.L. 1965. Geologic map of the Big Stone Gap quadrangle, Virginia. U.S. Geological Survey. Wash., D.C.

Nichols, J.D. 1975. Characteristics of computerized soil maps. Soil Sci. Am. J. 39:927-932.

Nichols, J.D. and L.J. Bartelli. 1972. Computer generated interpretive maps. The earth around us. Proc. Annual Meeting. Soil Conservation Soc. of Am. 27:20-24.

Nichols, J.D. and L.J. Bartelli. 1974. Computer-generated interpretive soil maps. J. Soil Water Conservation 29:232-235.

Perry, H.H., P.C. Connor, A.M. Baisden, C.S. Coleman, E.F. Henry, and A.W. Sinclair. 1954. Soil Survey of Wise County, Virginia. U.S.D.A. Soil Conservation Service, Va. Ag. Exper. Station and Tennessee Valley Authority. 114 pp.

Peuker, T.K. 1972. Computer cartography. Am. Assoc. of Geographers, Commission on College Geography. Resource Paper No. 17. 75 pp.

Ragg, J.M. 1977. The recording and organization of soil field data for computer areal mapping. Geoderma. 19:81-89.

Rogoff, M.J. 1978. Statewide computer-based land information systems: an annotated bibliography of an emergent field. Council of Planning Librarians, Exchange Bibliography No. 1490. 20 pp.

Rogoff, M.J. E. Dersch, D.L. Mokma, and E.P. Whiteside. 1980. Computer-assisted ratings of soil potentials for urban land uses. J. of Soil and Water Conservation. 35(5):237-241.

Salmen, C., J. Gropper, J. Hammill, and C. Reed. 1977. Comparison of selected operational capabilities of fifty-four geographic information systems. Western Govenors Policy Office. Ft. Collins, Colorado. 50 pp.

Schneider, D.M. and S. Amanullah. 1979. Comptuter-assisted land resources planning. Am. Planning Association, Planning Advisory Service Report No. 339.

Smart, C.W. 1976. A computer-assisted technique for planning minimum impact transmission right of way routes. PhD. Dissertation, Virginia Polytechnic Inst. and State Univ. Blacksburg, Va. 192 pp.

Soil Survey Staff. 1975. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. Ag. Handbook No. 436. Soil Conservation Service, Wash., D.C. 754 pp.

Steinitz, C. 1970. Landscape resource analysis. Landscape Architecture Quart. 60(2):101-105.

Stephenson, W., W.T. Williams, and S. Cook. 1972. Computer analyses of Petersen's original data on bottom communities. Ecological Monogr. 42:387-415.

Tracey, J.G. 1968. Investigations of changes in pasture composition by some classificatory methods. J. Appl. Ecological 5:639-648.

U.S. Geologic Survey. 1980. Digital terrain tapes: user guide, 2nd ed. National Cartographic Information Center. Reston, Virginia. 12 pp.

Voelker, A.H. 1976. Indices: a technige for using large spatial data bases. Oak Ridge National Laboratory, ORNL/RUS-15. Oak Ridge, Tennessee. 84 pp.

Walker, P.H. , G.F. Hall, R. Proty. 1968. Relation between landform parameters and soil properties. Soil Sci. Soc. Am. Proc. 32:101-104.

Wilcott, J.C., and W.A. Gates. 1977. A review of existing geographic information systems and some recommendations for future systems. Pages 553-568 In Classification, inventory, and analysis of fish and wildlife habiat. Office of Biological Services, U.S. Fish and Wildlife Service. Wash., D.C. 604 pp.

Williams, W.T. 1971. Principles of clustering. Annual Review Ecology and Systematics. 2:303-326.

APPENDIX A

FORTRAN IV SOURCE CODE FOR PROGRAM DIVIDE

```
C*************************************************************
C                                                            *
C                          DIVIDE                            *
C                                                            *
C*************************************************************
C                                                            *
C PURPOSE                                                    *
C                                                            *
C   DIVIDE OVERLAYS POWER MAPFILES AND PRODUCES A CLASS-     *
C IFICATION OF CELLS IN ONE MAPFILE BASED ON OTHER           *
C MAPFILES.  THE MAPFILE WHICH IS THE OBJECT OF THE CLASS-   *
C IFCATION IS CALLED THE EXTRINSIC ATTRIBUTE.  THE MAP-      *
C FILES UPON WHICH THE CLASSIFICATION IS BASED ARE CALLED    *
C INTRINSIC ATTRIBUTES.  DIVIDE ITERATIVELY SPLITS AN INI-   *
C TIAL POPULATION OF CELLS INTO DISSIMILAR CLUSTERS, RE-     *
C FLECTING DISCONTINUITIES IN THE EXTRINSIC ATTRIBUTE.       *
C                                                            *
C   CELLS WHICH ARE TO COMPRISE THE INITIAL POPULATION OF    *
C CELLS SHOULD BE ASSIGNED A VALUE OF 1.0 IN AN ADDITIONAL   *
C MAPFILE CALLED THE LABEL MAPFILE.  DIVIDE READS THE        *
C LABEL MAPFILE TO DETERMINE WHICH CELLS TO PROCESS AND      *
C SUBSEQENTLY WRITES TO THIS MAPFILE THE VALUES WHICH        *
C IDENTIFY DIFFERENT CLUSTERS.  THE LABEL MAPFILE IS A       *
C SPATIAL REPRESENTATION OF THE RESULTS OF THE CLASSIFI-     *
C CATION.  THE MORE THIS MAPFILE RESEMBLES THE MAPFILE OF    *
C THE EXTRINSIC ATTRIBUTE, THE BETTER THE CLASSIFCATION.     *
C                                                            *
C   DIVIDE PROVIDES A PRINTED OUTPUT SUMMARIZING EACH STEP   *
C IN THE CLASSIFICATION.  A CODED FORM OF THIS SUMMARY IS    *
C ALSO GENERATED FOR USE AS INPUT TO ANOTHER PROGRAM         *
C CALLED KEYOUT.  KEYOUT APPLIES THE CLASSIFCATION TO        *
C OTHER AREAS WITHOUT REFERENCE TO THE EXTRINSIC ATTRI-      *
C BUTE.                                                      *
C                                                            *
C DATA SET REQUIREMENTS                                      *
C                                                            *
C   DIVIDE IS PARTICULARLY SUITED TO PROCESSING WHAT MAY     *
C BE TERMED NON-QUANTITATIVE OR NON-ORDERED ATTRIBUTES       *
C WHICH ASSUME NO MORE THAN 11 STATES, E.G., A GEOLOGY       *
C MAPFILE CONTAINING 11 GEOLOGIC FORMATIONS.  DIVIDE WILL    *
C PROCESS QUANTITATIVE OR ORDERED ATTRIBUTES, E.G., AN       *
C ELEVATION MAPFILE, BUT SUCH ATTRIBUTES MUST FIRST BE       *
C CONVERTED TO A DISCRETE SCALE ALSO ASSUMING NO MORE THAN   *
C 11 STATES.  CELL VALUES FOR BOTH KINDS OF ATTRIBUTES       *
C SHOULD NOT EQUAL OR EXECEED 12.0.                          *
C                                                            *
C   DIVIDE IS LIMITED TO A STUDY AREA CONTAINING LESS THAN   *
C 5000 POWER MAPS DESCRIBED BY LESS THAN 15 INTRINSIC        *
C ATTRIBUTES.                                                *
C                                                            *
```

```
C I/O UNITS                                                          *
C                                                                    *
C   05 - CARD IMAGES (FORMAT FREE ON CARDS 1 AND 2)                  *
C     CARD 1) NO. OF MAPFILES                                        *
C             NO. OF NONORDERED ATTRIBUTES                           *
C             NO. OF DIVISIONS TO PERFORM                            *
C             NO. OF ATTRIBUTE STATES OF THE EXTRINSIC               *
C               ATTRIBUTE                                            *
C     CARD 2) NO. OF ATTRIBUTE STATES OF ALL THE INTRINSIC           *
C               ATTRIBUTES (IN THE ORDER GIVEN BY THE UNIT           *
C               NUMBERS ASSIGNED TO EACH)                            *
C     CARD 3) INTRINSIC ATTRIBUTE NAMES (A4,1X,A4,1X,...)            *
C     CARD 4) EXTRINSIC ATTRIBUTE STATE NAMES (A4)                   *
C     CARD 5)     .         .         .     .    .                   *
C     CARD N)     .         .         .     .    .                   *
C                                                                    *
C   08 - POWER DEFINE FILE                                           *
C   09 - FIRST NON-ORDERED INTRINSIC ATTRIBUTE MAPFILE               *
C   10 - SECOND NON-ORDERED INTRINSIC ATTRIBUTE MAPFILE              *
C      -   .          .          .          .         .               *
C   (UNIT NUMBERS INCREASE BY 1'S)                                   *
C      - LAST NON-ORDERED INTRINSIC ATTRIBUTE MAPFILE                *
C      - FIRST ORDERED INTRINSIC ATTRIBUTE MAPFILE                   *
C      -   .          .          .          .         .               *
C      - LAST ORDERED MAPFILE                                        *
C      - EXTRINSIC ATTRIBUTE MAPFILE                                 *
C      - LABEL MAPFILE                                               *
C   26 - OUTPUT FOR PROGRAM KEYOUT                                   *
C                                                                    *
C SUBROUTINES CALLED                                                 *
C                                                                    *
C   GETDEF - A POWER SYSTEM UTILITY THAT READS THE POWER             *
C            DEFINE FILE.                                            *
C   GETMAP - A POWER SYSTEM UTILITY THAT READS A POWER               *
C            MAP FROM ONE OR SEVERAL MAPFILES.                       *
C   PUTMAP - A POWER SYSTEM UTILITY THAT WRITES A POWER              *
C            MAP TO ONE OR SEVERAL MAPFILES                          *
C   WCMAX  - DETERMINES WHICH INTRINSIC ATTRIBUTE AND                *
C            WHICH COMBINATION OF STATES WILL YIELD THE              *
C            WITHIN-CLUSTER-MAXIMUM VALUE OF A MEASURE OF            *
C            BETWEEN-GROUP DISSIMILARITY.                            *
C                                                                    *
C MAJOR PROGRAM VARIABLES                                            *
C                                                                    *
C        ABBREVIATIONS USED:                                         *
C             E-A = EXTRINSIC ATTRIBUTE                              *
C             I-A = INTRINISIC ATTRIBUTE                             *
C             E-A-S = EXTRINSIC ATTRIBUTE STATE                      *
C             I-A-S = INTRINSIC ATTRISTATE                           *
C                                                                    *
```

```
C BCMXDI - BETWEEN-CLUSTER-MAXIMUM DELTA-I.                          *
C COUNT(I,J,K) - ARRARY WHICH COUNTS CO-OCCURRENCES OF               *
C  E-A-S(I) AND I-A-S(J) IN A CLUSTER, FOR ALL I-A(K).              *
C LABEL(I) - CELLS BELONGING TO A CLUSTER(I) ARE ASSIGNED            *
C  A LABEL TO IDENTIFY THE CELLS. THE LABEL ALSO DES-               *
C  CRIBES THE CLUSTER'S POSITION IN THE DIVISIVE HIER-               *
C  ARCHY: CLUSTER X BEGETS CLUSTERS 2X AND 2X+1.                     *
C LFTU - THE JCL UNIT NO. OF THE LABEL MAPFILE.                      *
C LUNIT - THE PROGRAM UNIT NO. OF THE LABEL MAPFILE.                 *
C MAPNUM - POWER MAP NUMBER OF MAP CURRENTLY BEING PRO-              *
C  CESSED.                                                           *
C MAPS(I,J,K) - ARRAY THAT CONTAINS DATA FOR ONE POWER               *
C  MAP(I,J) FROM ALL MAPFILES(K).                                    *
C NATT - NO. OF I-A MAPFILES.                                        *
C NCELLS - TOTAL NO. CELLS PROCESSED.                                *
C NCLUST - MAX. NO. OF CLUSTERS THAT CAN BE GENERATED:               *
C  NCLUST=(NDIV*2)-1                                                 *
C NCSKPD - NO. CELLS SKIPPED.                                        *
C NDIV - NO. DIVISIONS TO BE PERFORMED.                              *
C NIST - NO. OF STATES OF THE E-A.                                   *
C NJST(I) - NO. OF STATES OF THE I'TH I-A.                           *
C NNOATT - NO. NON-ORDERED ATTRIBUTES.                               *
C NSUBC - NO. CELLS IN CLUSTER(KUNIT).                               *
C NUNIT - NO. MAPFILES, INCLUDING THE E-A                            *
C  AND LABEL MAPFILES.                                               *
C PCNTCL(I) - PERCENT CELLS IN A CLUSTER BELONGING TO THE            *
C  I'TH E-A STATE. PROVIDES A MEASURE OF CONSTANCY.                  *
C PCNTOT(I) - PERCENT OF THE TOTAL NO. OF CELLS OF AN                *
C  E-A-S- BELONGING TO A CLUSTER(I). PROVIDES A MEASURE              *
C  OF FIDELITY.                                                      *
C STNAME - FOUR-LETTER ACRONYMNS GIVEN TO THE E-A-S.                 *
C STTC - TOTAL NO. CELLS BELONGING TO A E-A-S.                       *
C SUNIT - PROGRAM UNIT NO. OF THE E-A MAPFILE.                       *
C TERM(I) - LOGICAL VARIABLE INDICATING WHETHER A                    *
C  CLUSTER(I) HAS BEEN DIVIDED YET.                                  *
C WCMXDI(I) - WITHIN-CLUSTER(I)-MAX. DELTA-I.                        *
C XA(I,J) - J'TH I-A-S BELONGING TO THE "A" GROUP OF THE             *
C  CLUSTER(J) SELECTED FOR DIVISION.                                 *
C XATT(I) - THE I-A SELECTED FOR DIVISION OF CLUSTER(I).             *
C XB(I,J) - J'TH I-A-S BELONGING TO THE "B" GROUP OF THE             *
C  CLUSTER(J) SELECTED FOR DIVISION.                                 *
C XCLUST - THE CLUSTER SELECTED FOR DIVISION.                        *
C XFTU - THE JCL UNIT NO. OF THE I-A MAPFILE SELECTED FOR            *
C  DIVISION OF CLUSTER(XCLUST).                                      *
C XNA(I) - NO. OF I-A-S BELONGING TO THE "A" GROUP OF                *
C  CLUSTER(I) SELECTED FOR DIVISION.                                 *
C XNB(I) - NO. OF I-A-S BELONGING TO THE "B" GROUP OF                *
C  CLUSTER(I) SELECTED FOR DIVISION.                                 *
C XNSA(I) - NO. OF E-A-S BELONGING TO GROUP "A" OF                   *
C  CLUSTER(I) SELECTED FOR DIVISION.                                 *
```

```
C XNSB(I) - NO. OF E-A-S BELONGING TO GROUP "B" OF          *
C  CLUSTER(I) SELECTED FOR DIVISION.                        *
C XNSUBA(I) - NO. OF CELLS BELONGING TO GROUP "A" OF        *
C  CLUSTER(I) SELECTED FOR DIVISION.                        *
C XNSUBB(I) - NO. OF CELLS BELONGING TO GROUP "B" OF        *
C  CLUSTER(I) SELECTED FOR DIVISION.                        *
C XROWSA(I,J) - NO. OF CELLS IN E-A-S(J) BELONGING TO       *
C  GROUP A OF CLUSTER(I) SELECTED FOR DIVISION              *
C XROWSB(I,J) - NO. OF CELLS IN E-A-S(J) BELONGING TO       *
C  GROUP B OF CLUSTER(I) SELECTED FOR DIVISION              *
C XSA(I,J) - J'TH E-A-S BELONGING TO GROUP "A" OF           *
C  CLUSTER(I) SELECTED FOR DIVISION.                        *
C XSB(I,J) - J'TH E-A-S BELONGING TO GROUP "B" OF           *
C  CLUSTER(I) SELECTED FOR DIVISION.                        *
C                                                           *
C***********************************************************
C
C DECLARE MAIN PROGRAM VARIABLES
C
      DEFINE FILE 09(5000,930,U,M1),10(5000,930,U,M2),
     &  11(5000,930,U,M3),12(5000,930,U,M4),
     &  13(5000,930,U,M5),14(5000,930,U,M6),
     &  15(5000,930,U,M7),16(5000,930,U,M8),
     &  17(5000,930,U,M9),18(5000,930,U,M10),
     &  19(5000,930,U,M11),20(5000,930,U,M12),
     &  21(5000,930,U,M13),22(5000,930,U,M14),
     &  23(5000,930,U,M15),24(5000,930,U,M16),
     &  25(5000,930,U,M17)
C
      INTEGER NUNIT,NNOATT,NIST,NJST(15),ISKIP(6),
     &  IANAME(15),STNAME(35),
     &  SUNIT,FTUNIT,LUNIT,LABEL(69),
     &  XCLUST,XATT(69),XFTU,
     &  XNA(69),XNB(69),XNSA(69),XNSB(69),
     &  XA(69,10),XB(69,10),XSA(69,35),XSB(69,35),
     &  XROWSA(69,35),XROWSB(69,35)
C
      REAL COUNT(35,11,15),MAPS(30,31,17),WCMXDI(69),
     &  NSUBC(69),FXA(10),STTC(35),XNSUBA(69),XNSUBB(69)
C
      LOGICAL TERM(69)
C
      COMMON/AREA1/COUNT,XATT,XA,XB,XSA,XSB,XNA,XNB,
     &  XNSA,XNSB,XROWSA,XROWSB,NIST,NJST,NSUBC,WCMXDI,
     &  NNOATT,ICLUST,NATT,TERM,STTC,IDIV,XNSUBA,XNSUBB
C-----------------------------------------------------------
C
C READ UNIT 05 INPUT
C
      READ(5,*) NUNIT,NNOATT,NDIV,NIST
```

```
      NATT=NUNIT-2
C
      READ(5,*) (NJST(IATT),IATT=1,NATT)
C
      READ(5,501) (IANAME(IATT), IATT=1,NATT)
  501 FORMAT(15(A4,1X))
C
      DO 50 IST=1,NIST
        READ(5,502) STNAME(IST)
  502   FORMAT(A4)
   50 CONTINUE
C-----------------------------------------------------------
C
C CHECK UNIT 05 INPUT - IF OUT OF BOUNDS, WRITE ERROR
C MESSAGE, THEN STOP.
      IF(NUNIT.GT.17.OR.NIST.GT.35.OR.NDIV.GT.34) GO TO 99
      DO 16 I=1,NATT
        IF(NJST(I).LE.11) GO TO 16
        GO TO 99
   16 CONTINUE
C
      GO TO 55
C
   99 CONTINUE
      WRITE(6,6999)
 6999 FORMAT('1','PROGRAM DIVIDE CANNOT ACCOMODATE:'/
     & 3X,'MORE THAN 17 MAPFILES'/
     & 3X,'           35 EXTRINSIC ATTRIBUTE STATES'/
     & 3X,'           34 DIVISIONS'/
     & 3X,'           11 STATES IN ANY INTRINSIC ATTRIBUTE.'/
     & '0','PROCESSING TERMINATED')
      STOP
C-----------------------------------------------------------
C
   55 CONTINUE
C
C WRITE HEADER INFORMATION
      WRITE(6,6001) NIST, (STNAME(IST),IST=1,NIST)
 6001 FORMAT('1',5X,'PROGRAM DIVIDE'/
     & '0','NUMBER OF CLUSTERS TO DERIVE:',I4/
     & '0','EXTRINSIC ATTRIBUTE-STATE ACRONYMNS:'/
     & 3X,10(A4,1X)/
     & 3X,10(A4,1X)/
     & 3X,10(A4,1X)/
     & 3X,5(A4,1X))
C
      WRITE(6,6002) NATT, (IANAME(I),I=1,NATT)
 6002 FORMAT('-','NUMBER OF INTRINSIC ATTRIBUTES:',I4/
     & '0','INTRINSIC ATTRIBUTES ACRONYMNS:'/
     & 3X,10(A4,1X)/
```

```
     & 3X,5(A4,1X))
C
      WRITE(6,6003) (NJST(J),J=1,NATT)
 6003 FORMAT('0','NUMBER OF ATTRIBUTE-STATES ASSOCIATED '
     & 'WITH EACH:'/
     & 2X,10I5/
     & 2X,5I5)
C-----------------------------------------------------------
C
C INITIALIZE MAIN PROGRAM VARIABLES
C
      SUNIT=NUNIT-1
      LUNIT=NUNIT
      LFTU=LUNIT+8
      NCLUST=(2*NDIV)-1
      NCSKPD=0
C
      DO 10 ICLUST=1,NCLUST
        TERM(ICLUST)=.TRUE.
        LABEL(ICLUST)=0
   10 CONTINUE
C
      ICLUST=0
      LABEL(1)=1
C-----------------------------------------------------------
C
      DO 100 IDIV=1,NDIV
C
C DIVIDE ONE CLUSTER INTO TWO DISSIMILAR CLUSTERS
C
C-----------------------------------------------------------
C
C PROCESS THE TWO MOST RECENTLY FORMED CLUSTERS
      NNEW=2
      IF(IDIV.EQ.1) NNEW=1
      DO 200 INEW=1,NNEW
C-----------------------------------------------------------
C
C INITIALZE NEW CLUSTER (THE DO 200-LOOP) VARIABLES
C
      ICLUST=ICLUST+1
      NSUBC(ICLUST)=0.0
C
      DO 300 KUNIT=1,15
C
        DO 400 JST=1,11
C
          DO 500 IST=1,35
            COUNT(IST,JST,KUNIT)=0.0
  500     CONTINUE
```

```
  400   CONTINUE
  300 CONTINUE
C
C-----------------------------------------------------------
C
C GENERATE COUNT ARRAY FOR THE CLUSTER TO BE PROCESSED
C
      REWIND 8
      IF(IDIV.EQ.1)CALL GETDEF(8)
      IF(IDIV.GT.1)READ(8,8100)ISKIP
C
  600 CONTINUE
      READ(8,8100,END=800)MAPNUM
 8100 FORMAT(I4)
C
C READ MAP VALUES FOR ALL MAPFILES
C
        DO 900 KUNIT=1,LUNIT
          FTUNIT=KUNIT+8
          CALL GETMAP(FTUNIT,MAPNUM,MAPS(1,1,KUNIT))
  900   CONTINUE
C
C DETERMINE WHETHER CELL BELONGS TO CURRENT CLUSTER BY
C CHECKING VALUE OF LABEL MAPFILE.
C IF NOT, SKIP TO THE NEXT CELL.
C
C DETERMINE WHETHER ANY MAPFILE VALUE FOR THE CELL IS OUT
C OF BOUNDS. IF SO, SKIP TO THE NEXT CELL.
C
C COUNT OCCURANCES OF ALL MAPS ARRAY VALUES.  ASSIGN
C TOTALS TO COUNT ARRAY.
C
        DO 1000 IY=1,30
C
          DO 1100 IX=1,30
            IF(MAPS(IX,IY,LUNIT).NE.LABEL(ICLUST))
     &       GO TO 1100
            IF(MAPS(IX,IY,SUNIT).GE.1.0 .AND.
     &       MAPS(IX,IY,SUNIT).LE.NIST) GO TO 1110
            IF(IDIV.EQ.1) NCSKPD=NCSKPD + 1
            GO TO 1100
C
 1110       CONTINUE
C
            IST=MAPS(IX,IY,SUNIT)
            NSUBC(ICLUST)=NSUBC(ICLUST)+1.0
C
            DO 1200 KUNIT=1,NATT
              IF(MAPS(IX,IY,KUNIT).GE.1.0 .AND.
     &         MAPS(IX,IY,KUNIT).LE.NJST(KUNIT)) GO TO 1200
```

```
              IF(IDIV.EQ.1) NCSKPD=NCSKPD + 1
              GO TO 1100
 1200       CONTINUE
C
            DO 1350 KUNIT=1,NATT
              JST=MAPS(IX,IY,KUNIT)
              COUNT(IST,JST,KUNIT)=COUNT(IST,JST,KUNIT)+1.0
 1350       CONTINUE
 1100     CONTINUE
 1000   CONTINUE
C
        IF(IDIV.EQ.1) NCELLS = NSUBC(ICLUST)
C
C RETURN TO READ ANOTHER MAP
C
        GO TO 600
  800 CONTINUE
C-------------------------------------------------------------
C
C DETERMINE MAX. WITHIN-CLUSTER DELTA-I (WCMXDI(ICLUST))
C VARIABLES SHARED BETWEEN MAIN AND WCMAX ARE COMMON/AREA1/
C
      CALL WCMAX
C-------------------------------------------------------------
C
  200 CONTINUE
C
C-------------------------------------------------------------
C
C DETERMINE MAX. DELTA-I BETWEEN TERMINAL CLUSTERS
C
      BCMXDI=0.0
C
      DO 1500 ICOMP=1,ICLUST
        IF(.NOT.TERM(ICOMP)) GO TO 1500
        IF(WCMXDI(ICOMP).LT.BCMXDI) GO TO 1500
        IF(WCMXDI(ICOMP).EQ.BCMXDI) GO TO 1600
        BCMXDI=WCMXDI(ICOMP)
        XCLUST=ICOMP
        GO TO 1500
C
 1600   CONTINUE
        IF(XNA(ICOMP).LE.XNA(XCLUST)) GO TO 1500
        XCLUST=ICOMP
 1500 CONTINUE
C
C ASSIGN NEW GROUP IDENTIFIERS TO THE LABEL ARRAY.
C
      TERM(XCLUST)=.FALSE.
      ILEFT=ICLUST+1
```

```
      IRIGHT=ICLUST+2
      LABEL(ILEFT)=2*LABEL(XCLUST)
      LABEL(IRIGHT)=LABEL(ILEFT)+1
C-----------------------------------------------------------
C
C RE-LABEL MAP CELLS BELONGING TO THE CLUSTER SELECTED FOR
C DIVISION.
C
      REWIND 8
      READ(8,8100)ISKIP
 1700 CONTINUE
      READ(8,8100,END=1800)MAPNUM
C
      KUNIT=XATT(XCLUST)
      XFTU=KUNIT+8
      CALL GETMAP(XFTU,MAPNUM,MAPS(1,1,KUNIT))
      CALL GETMAP(LFTU,MAPNUM,MAPS(1,1,LUNIT))
C
      NA=XNA(XCLUST)
C
C CONVERT INTEGER VALUES DESIGNATING ATTRIBUTE STATES
C ASSOCIATED WITH THE 'A' CLUSTER (XA) TO REAL VALUES FOR
C COMPARISON WITH POWER MAPFILE VALUES.
C
      DO 1850 IA=1,NA
        FXA(IA)=FLOAT(XA(XCLUST,IA))
 1850 CONTINUE
C
      DO 1900 IY=1,30
C
        DO 2000 IX=1,30
          IF(MAPS(IX,IY,LUNIT).NE.LABEL(XCLUST)) GO TO 2000
C
          DO 2100 IA=1,NA
            IF(MAPS(IX,IY,KUNIT).NE.FXA(IA)) GO TO 2100
            MAPS(IX,IY,LUNIT)=LABEL(ILEFT)
            GO TO 2000
 2100     CONTINUE
          MAPS(IX,IY,LUNIT)=LABEL(IRIGHT)
 2000   CONTINUE
 1900 CONTINUE
C
      CALL PUTMAP(LFTU,MAPNUM,MAPS(1,1,LUNIT))
C
      GO TO 1700
C
 1800 CONTINUE
C-----------------------------------------------------------
C
C WRITE DESCRIPTION OF DIVISION SELECTED
```

```
C
      I=XATT(XCLUST)
      WRITE(6,6100)IDIV,XCLUST,LABEL(XCLUST),XATT(XCLUST),
     &  IANAME(I),BCMXDI
 6100 FORMAT('1','DIVISION NUMBER ',I3/
     & 1X,'CLUSTER NUMBER ',I3/
     & 1X,'CLUSTER LABEL ',I3/
     & 1X,'DIVISION ATTRIBUTE ',I3,', ',A4/
     & 1X,'INFORMATION LOSS RESULTING FROM DIVISION OF '
     & 'CLUSTER, ',E14.6)
C
      N=XNA(XCLUST)
      WRITE(6,6200) LABEL(ILEFT),(XA(XCLUST,IXA),IXA=1,N)
 6200 FORMAT('-','ATTRIBUTE STATES ASSOCIATED WITH NEW'
     & 'CLUSTER LABELED',I4,':'/1X,10I3)
      WRITE(6,6250)
 6250 FORMAT('0',8X,'                                % OF TYPE'/
     & ' ','EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA')
C
      N=XNSA(XCLUST)
C
      DO 2300 IXSA=1,N
        PCNTCL=(XROWSA(XCLUST,IXSA) / XNSUBA(XCLUST)) * 100.
        I=XSA(XCLUST,IXSA)
        PCNTOT=(XROWSA(XCLUST,IXSA) / STTC(I)) * 100.
        WRITE(6,6300)XSA(XCLUST,IXSA),STNAME(I),
     &   XROWSA(XCLUST,IXSA),PCNTCL,PCNTOT
 6300   FORMAT(' ',2X,I2,1X,A4,4X,I6,6X,F5.1,8X,F5.1)
 2300 CONTINUE
C
      N=XNB(XCLUST)
      WRITE(6,6200) LABEL(IRIGHT),(XB(XCLUST,IXB),IXB=1,N)
      WRITE(6,6250)
C
      N=XNSB(XCLUST)
C
      DO 2400 IXSB=1,N
        PCNTCL=(XROWSB(XCLUST,IXSB) / XNSUBB(XCLUST)) * 100.
        I=XSB(XCLUST,IXSB)
        PCNTOT=(XROWSB(XCLUST,IXSB) / STTC(I)) * 100.
        WRITE(6,6300)XSB(XCLUST,IXSB),STNAME(I),
     &   XROWSB(XCLUST,IXSB),PCNTCL,PCNTOT
 2400 CONTINUE
C
C-----------------------------------------------------------
  100 CONTINUE
C-----------------------------------------------------------
C
C WRITE SUMMARY OF DIVISIONS, I.E., WHICH CLUSTERS
C REPRESENT THE FINAL GROUPINGS OF MAP CELLS.
```

```
C
C
      WRITE(6,6525)
 6525 FORMAT('1','TERMINAL CLUSTERS ARE LABELED:'/)
C
      DO 999 I=1,IRIGHT
        IF(.NOT.TERM(I)) GO TO 999
        WRITE(6,6550) LABEL(I)
 6550   FORMAT(' ',I4)
  999 CONTINUE
C
      WRITE(6,6600) NCELLS,NCSKPD
 6600 FORMAT('0','NUMBER OF CELLS PROCESSED:',I10/
     & '0','NO. OF CELLS SKIPPED BECAUSE ATTRIBUTE STATE'/
     & ' ','VALUES WERE OUT OF RANGE:',I10)
C-----------------------------------------------------------
C
C WRITE INPUT TO PROGRAM KEYOUT
C
      KNUNIT = NUNIT - 1
      WRITE(26,2601) KNUNIT, NIST
 2601 FORMAT('1',2I3)
C
      DO 643 I=1,NCLUST
        WRITE(26,2602) LABEL(I), TERM(I), XATT(I), XNA(I)
 2602   FORMAT(' ',I6,L2,I3,I2)
  643 CONTINUE
C
      DO 644 I=1,NCLUST
        K=XNA(I)
        WRITE(26,2603) (XA(I,J), J=1,K)
 2603   FORMAT(' ',10I3)
  644 CONTINUE
C-----------------------------------------------------------
      STOP
      END
```

```
C*************************************************************
C                                                            *
C                      SUBROUTINE WCMAX                      *
C                                                            *
C*************************************************************
C                                                            *
C PURPOSE                                                    *
C                                                            *
C   TO DETERMINE HOW BEST TO DIVIDE A CLUSTER, C, INTO TWO   *
C GROUPS, A AND B.  FOR EACH INTRINSIC ATTRIBUTE (I-A) IN    *
C A MAPFILE, WCMAX DETERMINES THE COMBINATION OF ATTRI-      *
C BUTE STATES (I-A-S) THAT YIELDS THE MAXIMUM VALUE OF       *
C DELTA-I.  DELTA-I IS A MEASURE OF DISSIMILARITY BETWEEN    *
C GROUPS.  FOR NON-ORDERED ATTRIBUTES, THIS INVOLVES THE     *
C THE CALCULATION OF EVERY POSSIBLE DIVISION OF THE SET OF   *
C ATTRIBUTE STATES INTO TWO GROUPS.  FOR ORDERED ATTRI-      *
C BUTES, IT INVOLVES ONLY CONSIDERING DIVISIONS BY SIMPLE    *
C BREAKS IN THE RANGE OF STATES.  THE ATTRIBUTE THAT         *
C YEILDS THE MAXIMUM VALUE OF DELTA-I COMPARED TO ALL        *
C OTHER ATTRIBUTES BECOMES A CANDIDATE FOR AN ACTUAL DIV-    *
C ISION OF A CLUSTER IN THE MAIN PROGRAM.                    *
C                                                            *
C MAJOR SUBROUTINE VARIABLES                                 *
C  (THOSE SHARED IN A COMMONBLOCK WITH THE MAIN PROGRAM      *
C  ARE DESCRIBED IN THE MAIN PROGRAM)                        *
C                                                            *
C A(I) - I'TH I-A-S OF THE "A" GROUP.                        *
C B(I) - I'TH I-A-S OF THE "B" GROUP.                        *
C BFLAG(I) - LOGICAL VARIABLE INDICATING WHETHER THE I'TH    *
C  I-A-S BELONGS TO THE B GROUP.                             *
C C(I) - I'TH I-A-S OF THE OF CLUSTER (THE "A" AND "B"       *
C  GROUPS ARE DERIVED FROM C, THE CLUSTER).                  *
C COUNT(I,J,K) - SEE MAIN PROGRAM DOCUMENTATION              *
C DI - DELTA-I, THE MEASURE OF DISSIMILARITY BETWEEN THE     *
C  A AND B GROUPS.                                          *
C INFOA - INFO. CONTENT OR DIVERSITY OF THE "A" GROUP,       *
C  WITH RESPECT TO ONE I-A.                                  *
C INFOB - INFO. CONTENT OR DIVERSITY OF THE "B" GROUP,       *
C  WITH RESPECT TO ONE I-A.                                  *
C INFOC - INFO. CONTENT OR DIVERSITY OF THE CLUSTER          *
C  AS A WHOLE, WITH RESPECT TO ONE I-A.                      *
C NA(I) - NO. I-A-S BELONGING TO THE "A" GROUP OF            *
C  CLUSTER(I).                                               *
C NATT - SEE MAIN PROGRAM DOCUMENTATION                      *
C NAW1S - NO. OF ATTRIBUTES HAVING ONE STATE ONLY.           *
C NB(I) - NO. I-A-S BELONGING TO THE "B" GROUP OF            *
C  CLUSTER(I).                                               *
C NC(I) - NO. I-A-S BELONGING TO THE I'TH CLUSTER.           *
C NIST - SEE MAIN PROGRAM DOCUMENTATION                      *
C NJST - SEE MAIN PROGRAM DOCUMENTATION                      *
```

```
C NNOATT - SEE MAIN PROGRAM DOCUMENTATION                              *
C NS - NO. E-A-S IN THE CLUSTER.                                       *
C NSA - NO. E-A-S IN THE "A" GROUP OF THE CLUSTER                      *
C NSB - NO. E-A-S IN THE "B" GROUP OF THE CLUSTER                      *
C NSUBA - NO. CELLS BELONGING TO THE "A" GROUP                         *
C NSUBB - NO. CELLS BELONGING TO THE "B" GROUP                         *
C NSUBC - NO. CELLS BELONGING TO THE CLUSTER                           *
C S(I) - I'TH E-A-S OF THE CLUSTER.                                    *
C SA(I) - I'TH E-A-S OF THE "A" GROUP.                                 *
C SB(I) - I'TH E-A-S OF THE "B" GROUP.                                 *
C SCOL(I) - SUM OF COLUMNS, I.E., THE NO. CELLS IN THE                 *
C  I'TH I-A-S.                                                         *
C SPLSRA - SUM-PRODUCT-LOG-SUM-ROWS-"A" GROUP, A TERM IN               *
C  THE CALCULATION OF INFOA.                                           *
C SPLSRB - SUM-PRODUCT-LOG-SUM-ROWS-"B" GROUP, A TERM IN               *
C  THE CALCULATION OF INFOB.                                           *
C SPLSRC - SUM-PRODUCT-LOG-SUM-ROWS CLUSTER, A TERM IN                 *
C  THE CALCULATION OF INFOA.                                           *
C SRA(I) - SUM ROWS A, THE NO. CELLS OF THE I'TH I-A-S                 *
C  BELONGING TO THE "A" GROUP.                                         *
C SRB(I) - SUM ROWS B, THE NO. CELLS OF THE I'TH I-A-S                 *
C  BELONGING TO THE "B" GROUP.                                         *
C SRA(I) - SUM ROWS C, THE NO. CELLS OF THE I'TH I-A-S                 *
C  BELONGING TO THE CLUSTER.                                           *
C STTC(I) - NO. CELLS OF THE I'TH E-A-S IN THE ENTIRE                  *
C  POPULATION OF CELLS BEING PROCESSED.                                *
C TERM(I) -                                                            *
C WCMXDI-                                                              *
C XA-                                                                  *
C XATT-                                                                *
C XB-                                                                  *
C XNA-                                                                 *
C XNB-                                                                 *
C XNSA-          SEE MAIN PROGRAM DOCUMENTATION.                       *
C XNSB-                                                                *
C XNSUBA-                                                              *
C XNSUBB-                                                              *
C XROWSA-                                                              *
C XROWSB-                                                              *
C XSA-                                                                 *
C XSB-                                                                 *
C***********************************************************************
C
      SUBROUTINE WCMAX
C
C DECLARE AND INITIALIZE SUBROUTINE VARIABLES
C
      INTEGER S(35),C(11),NJST(15),SCOL(11),
     &  A(5),B(10),SA(35),SB(35),
     &  XATT(69),XNA(69),XNSA(69),XNB(69),XNSB(69),
```

```
     &   XA(69,10),XB(69,10),XSA(69,35),XSB(69,35),
     &   XROWSA(69,35),XROWSB(69,35),ISRA(35),ISRB(35)
C
      REAL COUNT(35,11,15),NSUBC(69),NSUBA,NSUBB,
     &   WCMXDI(69),SRC(35),SRA(35),SRB(35),
     &   INFOA,INFOB,INFOC,
     &   STTC(35),XNSUBA(69),XNSUBB(69)
C
      LOGICAL BFLAG(11),TERM(69)
C
      COMMON/AREA1/COUNT,XATT,XA,XB,XSA,XSB,XNA,XNB,
     &   XNSA,XNSB,XROWSA,XROWSB,NIST,NJST,
     &   NSUBC,WCMXDI,NNOATT,ICLUST,NATT,TERM,
     &   STTC,IDIV,XNSUBA,XNSUBB
C
C-----------------------------------------------------------
C
C INITIALIZE THOSE VARIABLES REQUIRING RESETTING
C FOR EACH CALL TO SUBROUTINE WCMAX
C
      WCMXDI(ICLUST)=0.0
      NS=0
      SPLSRC=0.0
      NAW1S=0
C-----------------------------------------------------------
C
C DETERMINE WHICH EXTRINSIC ATTRIBUTE-STATES ARE ASSOCIATED
C WITH THE CLUSTER AND AND CALCULATE THE INFORMATION
C CONTENT OF THE FULL C MATRIX, INFOC.
C
C
      DO 400 IST=1,NIST
        SRC(IST)=0.0
C
        N=NJST(1)
C
        DO 500 JST=1,N
          SRC(IST)=SRC(IST)+COUNT(IST,JST,1)
  500   CONTINUE
C
        IF(SRC(IST).LT.1.0) GO TO 400
        NS=NS+1
        S(NS)=IST
        SPLSRC=SPLSRC+SRC(IST)*ALOG10(SRC(IST))
  400 CONTINUE
C
      IF(IDIV.GT.1) GO TO 525
C
      DO 530 IST=1,NIST
        STTC(IST)=SRC(IST)
```

```
  530 CONTINUE
C
C IF THE CLUSTER CONTAINS MAP CELLS BELONGING TO ONLY
C ONE EXTRINSIC ATTRIBUTE-STATE, DIVISION OF THE CLUSTER
C IS NOT NECESSARY. RETURN TO MAIN PROGRAM.
C
  525 CONTINUE
C
      IF(NS.GT.1)GO TO 450
        TERM(ICLUST)=.FALSE.
        WCMXDI(ICLUST)=0.0
        XNA(ICLUST)=0
        RETURN
  450 CONTINUE
C
      INFOC=NSUBC(ICLUST)*ALOG10(NSUBC(ICLUST))-SPLSRC
C-----------------------------------------------------------
C
      DO 100 KUNIT=1,NATT
C
C PROCESS EACH INTRINSIC ATTRIBUTE
C
C-----------------------------------------------------------
C
C INITIALIZE LOOP VARIABLES, THOSE REQUIRING RESETTING FOR
C EACH ATTRIBUTE.
C
      NC=0
      N=NJST(KUNIT)
C
      DO 150 IC=1,11
        BFLAG(IC)=.TRUE.
  150 CONTINUE
C-----------------------------------------------------------
C
C DETERMINE WHICH INTRINSIC ATTRIBUTE-STATES ARE ASSOCIATED
C WITH THE CLUSTER.
C
      DO 200 JST=1,N
        SCOL(JST)=0
C
        DO 300 IS=1,NS
          IST=S(IS)
          SCOL(JST)=SCOL(JST)+COUNT(IST,JST,KUNIT)
  300   CONTINUE
C
        IF(SCOL(JST).LT.1) GO TO 200
        NC=NC+1
        C(NC)=JST
  200 CONTINUE
```

```
C
C IF AN ATTRIBUTE IN THE CLUSTER IS REPRESENTED BY ONLY ONE
C STATE, PROCESS NEXT ATTRIBUTE.  IF ALL ATTRIBUTES HAVE
C ONLY ONE STATE, SET CLUSTER VARIABLES AND RETURN TO MAIN.
C
      IF(NC.GT.1)GO TO 250
        NAW1S=NAW1S+1
        IF(NAW1S.NE.NATT) GO TO 100
          TERM(ICLUST)=.FALSE.
          WCMXDI(ICLUST)=0.0
          XNA(ICLUST)=0
          GO TO 100
  250 CONTINUE
C-------------------------------------------------------------
C
C FOR ORDERED ATTRIBUTES BYPASS THE FOLLOWING STEP
C
      IF(KUNIT.GT.NNOATT) GO TO 600
C-------------------------------------------------------------
C
C GENERATE ALL COMBINATIONS FOR NON-ORDERED ATTRIBUTE
C STATES.
C
      NTC=NC/2
C
      DO 700 IR=1,NTC
        I1=1
        J1=1
        K1=1
        L1=1
        M1=1
C
        I2=NC-4
        J2=NC-3
        K2=NC-2
        L2=NC-1
        M2=NC
C
        IF(IR.LT.5)I2=1
        IF(IR.LT.4)J2=1
        IF(IR.LT.3)K2=1
        IF(IR.LT.2)L2=1
C
        IF(IR.EQ.5 .AND. NC.EQ.10) I2=1
        IF(IR.EQ.4 .AND. NC.EQ. 8) J2=1
        IF(IR.EQ.3 .AND. NC.EQ. 6) K2=1
        IF(IR.EQ.2 .AND. NC.EQ. 4) L2=1
C
    5   DO 15 I=I1,I2
          IF(IR.LT.5)GO TO 4
```

```
            J1=I+1
            A(1)=C(I)
            BFLAG(I)=.FALSE.
C
    4       CONTINUE
C
            DO 25 J=J1,J2
              IF(IR.LT.4)GO TO 3
              K1=J+1
              IA=IR-3
              A(IA)=C(J)
              BFLAG(J)=.FALSE.
C
   3          CONTINUE
C
              DO 35 K=K1,K2
                IF(IR.LT.3)GO TO 2
                L1=K+1
                IA=IR-2
                A(IA)=C(K)
                BFLAG(K)=.FALSE.
C
    2           CONTINUE
C
                DO 45 L=L1,L2
                  IF(IR.LT.2)GO TO 1
                  M1=L+1
                  IA=IR-1
                  A(IA)=C(L)
                  BFLAG(L)=.FALSE.
C
    1             CONTINUE
C
                  DO 55 M=M1,M2
                    A(IR)=C(M)
                    BFLAG(M)=.FALSE.
C-----------------------------------------------------------
C
C CALCULATE DI (FOR DELTA-I), THE INFORMATION LOSS RESULTING
C FROM A DIVISION OF THE CLUSTER BY THE CURRENT GROUPING OF
C ATTRIBUTE STATES.
C
                    NSUBA=0.0
                    NSUBB=0.0
                    SPLSRA=0.0
                    SPLSRB=0.0
                    NA=IR
                    NB=NC-IR
                    NSA=0
                    NSB=0
```

```
                    IB=0
C
                    DO 750 IC=1,NC
                      IF(.NOT.BFLAG(IC))GO TO 750
                      IB=IB+1
                      B(IB)=C(IC)
  750               CONTINUE
C
                    DO 800 IS=1,NS
                      SRA(IS)=0.0
                      SRB(IS)=0.0
                      IST=S(IS)
C
                      DO 900 IA=1,NA
                        JST=A(IA)
                        SRA(IS)=SRA(IS)+COUNT(IST,JST,KUNIT)
  900                 CONTINUE
C
                      NSUBA=NSUBA+SRA(IS)
                      SRB(IS)=SRC(IST)-SRA(IS)
C
                      IF(SRA(IS).LT.1.0) GO TO 850
                      NSA=NSA+1
                      SA(NSA)=IST
                      ISRA(NSA)=IFIX(SRA(IS))
                      SPLSRA=SPLSRA+SRA(IS)*ALOG10(SRA(IS))
C
  850                 CONTINUE
C
                      IF(SRB(IS).LT.1.0) GO TO 800
                      NSB=NSB+1
                      SB(NSB)=IST
                      ISRB(NSB)=IFIX(SRB(IS))
                      SPLSRB=SPLSRB+SRB(IS)*ALOG10(SRB(IS))
  800               CONTINUE
C
                    NSUBB=NSUBC(ICLUST)-NSUBA
                    IF(NSUBA.LT.1.0.OR.NSUBB.LT.1.0)GO TO 44
C
                    INFOA=NSUBA*ALOG10(NSUBA)-SPLSRA
                    INFOB=NSUBB*ALOG10(NSUBB)-SPLSRB
                    DI=INFOC-INFOA-INFOB
C-------------------------------------------------------------
C
C COMPARE DI WITH WCMXDI(ICLUST), THE CURRENT WITHIN-
C CLUSTER-MAXIMUM INFORMATION LOSS IF THE CLUSTER IS
C DIVIDED.
C
                    IF(DI.LT.WCMXDI(ICLUST)) GO TO 44
C
```

```
                    WCMXDI(ICLUST)=DI
                    XATT(ICLUST)=KUNIT
                    XNA(ICLUST)=NA
                    XNSA(ICLUST)=NSA
                    XNB(ICLUST)=NB
                    XNSB(ICLUST)=NSB
                    XNSUBA(ICLUST)=NSUBA
                    XNSUBB(ICLUST)=NSUBB
C
                    DO 1000 IA=1,NA
                      XA(ICLUST,IA)=A(IA)
 1000               CONTINUE
C
                    DO 1100 ISA=1,NSA
                      XSA(ICLUST,ISA)=SA(ISA)
                      XROWSA(ICLUST,ISA)=ISRA(ISA)
 1100                 CONTINUE
C
                    DO 1150 IB=1,NB
                      XB(ICLUST,IB)=B(IB)
 1150               CONTINUE
C
                    DO 1175 ISB=1,NSB
                      XSB(ICLUST,ISB)=SB(ISB)
                      XROWSB(ICLUST,ISB)=ISRB(ISB)
 1175               CONTINUE
C
   44                BFLAG(M)=.TRUE.
C
   55              CONTINUE
C
                   BFLAG(L)=.TRUE.
   45            CONTINUE
C
                 BFLAG(K)=.TRUE.
   35          CONTINUE
C
              BFLAG(J)=.TRUE.
   25       CONTINUE
C
            BFLAG(I)=.TRUE.
   15    CONTINUE
C
  700 CONTINUE
C
      GO TO 100
C
C-------------------------------------------------------------
C
  600   CONTINUE
```

```
C
C GENERATE ALL COMBINATIONS OF ORDERED ATTRIBUTE STATES.
C
      NR=NC-1
C
      DO 1200 IR=1,NR
        NSUBA=0.0
        NSUBB=0.0
        SPLSRA=0.0
        SPLSRB=0.0
        NA=IR
        NB=NC-IR
        NSA=0
        NSB=0
C
        DO 1300 IS=1,NS
          SRA(IS)=0.0
          SRB(IS)=0.0
          IST=S(IS)
C
          DO 1400 IA=1,NA
            JST=C(IA)
            SRA(IS)=SRA(IS)+COUNT(IST,JST,KUNIT)
 1400     CONTINUE
C
          NSUBA=NSUBA+SRA(IS)
          SRB(IS)=SRC(IST)-SRA(IS)
C
          IF(SRA(IS).LT.1.0) GO TO 1350
          NSA=NSA+1
          SA(NSA)=IST
          ISRA(NSA)=IFIX(SRA(IS))
          SPLSRA=SPLSRA+SRA(IS)*ALOG10(SRA(IS))
C
 1350     CONTINUE
C
          IF(SRB(IS).LT.1.0) GO TO 1300
          NSB=NSB+1
          SB(NSB)=IST
          ISRB(NSB)=IFIX(SRB(IS))
          SPLSRB=SPLSRB+SRB(IS)*ALOG10(SRB(IS))
 1300   CONTINUE
C-------------------------------------------------------------
C
C CALCULATE DI, THE INFORMATION LOSS RESULTING FROM A
C DIVISION OF THE CLUSTER BY THE CURRENT GROUPING OF
C ATTRIBUTE STATES.
C
C
        NSUBB=NSUBC(ICLUST)-NSUBA
```

```
        IF(NSUBA.LT.1.0.OR.NSUBB.LT.1.0)GO TO 1200
C
        INFOA=NSUBA*ALOG10(NSUBA)-SPLSRA
        INFOB=NSUBB*ALOG10(NSUBB)-SPLSRB
        DI=INFOC-INFOA-INFOB
C-------------------------------------------------------------
C COMPARE DI WITH WCMXDI(ICLUST), THE CURRENT WITHIN-
C CLUSTER-MAXIMUM INFORMATION LOSS IF THE CLUSTER IS
C DIVIDED.
C
        IF(DI.LT.WCMXDI(ICLUST)) GO TO 1200
C
        WCMXDI(ICLUST)=DI
        XATT(ICLUST)=KUNIT
        XNA(ICLUST)=NA
        XNSA(ICLUST)=NSA
        XNB(ICLUST)=NB
        XNSB(ICLUST)=NSB
        XNSUBA(ICLUST)=NSUBA
        XNSUBB(ICLUST)=NSUBB
C
        DO 1500 IA=1,NA
          XA(ICLUST,IA)=C(IA)
 1500   CONTINUE
C
        DO 1600 ISA=1,NSA
          XSA(ICLUST,ISA)=SA(ISA)
          XROWSA(ICLUST,ISA)=ISRA(ISA)
 1600   CONTINUE
C
        DO 1700 IB=1,NB
          IC=IB+NA
          XB(ICLUST,IB)=C(IC)
 1700   CONTINUE
C
        DO 1800 ISB=1,NSB
          XSB(ICLUST,ISB)=SB(ISB)
          XROWSB(ICLUST,ISB)=ISRB(ISB)
 1800   CONTINUE
 1200 CONTINUE
C-------------------------------------------------------------
C
  100 CONTINUE
      RETURN
      END
```

APPENDIX B

OUTPUT FROM PROGRAM DIVIDE

ABBREVIATIONS USED TO DENOTE SOIL TYPES

AA	Atkins loam
CHKL	Coeburn silt loam
CEFG	Coeburn-Muskingum complex
DA	Dekalb loam
EA	Enders silt loam
GA	Greendale fine sandy loam
HA	Hagerstown silt loam
HB	Hagerstown stony silt loam
HD	Hartsells fine sandy loam
HGH	Holston fine sandy loam
JAB	Jefferson loam
JCD	Jefferson stony loam
LAB	Leadvale loam
MC	Monogahela fine sandy loam
MMN	Muskingum stony loam
MFG	Muskingum fine sandy loam
MLK	Muskingum stony fine sandy loam
PA	Philo loam
PB	Pope fine sandy loam
PC	Pope loamy fine sand
RAB	Rolling stony land
RCD	Rough stony land
SA	Sequatchie fine sandy loam
SB	Stony alluvium
TAB	Teas-Dandridge silt loam
WAB	Westmoreland silt loam

ABBREVIATIONS USED TO DENOTE INTRINSIC ATTRIBUTES AND STATES

Attribute Abbreviation	Attribute	State Code	State
ASPE	Aspect	1	Northeast
		2	East
		3	Southeast
		4	South
		5	Southwest
		6	West
		7	Northwest
		8	North
DRDG	Distance to Ridge	1	Cell is a ridge
		2	Ridge is 1 cell away
		3	Ridge is 2 cells away
		etc.	etc.
		8	Ridge is 8 cells away
DSTR	Distance to Stream	1	
		2	
		3	See Table 6, pg. 52
		4	
		5	
GEOL	Geologic Formation	1	Wise Formation
		2	Gladeville sandstone
		3	Norton Formation
		4	Lee Formation
		5	Undifferentiated Mississippian
SHAP	Topographic Shape	1	Slope
		2	Ridge
		3	Valley
		4	Knob
		5	Depression
		6	Saddle
SLOP	Slope	1	
		2	
		...	See Table 5, pg. 45

OUTPUT FOR THE CONSTRUCTION DATA SET

```
     PROGRAM DIVIDE

NUMBER OF CLUSTERS TO DERIVE:  26

EXTRINSIC ATTRIBUTE-STATE ACRONYMNS:
  AA    CHKL CEFG DA    EA    GA    HA    HB    HD    HGH
  JAB   JCD  LAB  MC    MMN   MFG   MLK   PA    PB    PC
  RAB   RCD  SA   SB    TAB   WAB

NUMBER OF INTRINSIC ATTRIBUTES:   6

INTRINSIC ATTRIBUTES ACRONYMNS:
  GEOL ASPE SHAP SLOP DSTR DRDG
NUMBER OF ATTRIBUTE-STATES ASSOCIATED WITH EACH:
     5    8    6    8    5    8

 NUMBER OF MAPS PER ROW      =    19
 REFERENCE Y MAP COORDINATE  =     1
 REFERENCE X MAP COORDINATE  =     1
 MINIMUM MAP NUMBER          =     1
 MAXIMUM MAP NUMBER          =   285
 ADJUSTMENT FACTOR           =     0
 MINIMUM Y MAP COORDINATE    =     1
 MINIMUM X MAP COORDINATE    =     1
 MAXIMUM Y MAP COORDINATE    =    15
 MAXIMUM X MAP COORDINATE    =    19
```

****

DIVISION NUMBER 1
CLUSTER NUMBER 1
CLUSTER LABEL 1
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.186638E+0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 2:
4 5

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	91	1.7	12.2
2	CHKL	55	1.0	2.2
3	CEFG	193	3.6	4.8
4	DA	67	1.3	100.0
5	EA	1	0.0	1.0
6	GA	9	0.2	100.0
7	HA	7	0.1	100.0
8	HB	15	0.3	100.0
9	HD	139	2.6	25.9
10	HGH	8	0.2	8.4
12	JCD	43	0.8	5.0
13	LAB	31	0.6	6.4
15	MMN	1774	33.4	8.6
16	MFG	118	2.2	31.1
17	MLK	687	12.9	65.1
18	PA	5	0.1	3.0
19	PB	3	0.1	2.1
21	RAB	6	0.1	1.1
22	RCD	2021	38.1	71.8
23	SA	11	0.2	15.5
24	SB	16	0.3	5.3
25	TAB	5	0.1	100.0
26	WAB	4	0.1	100.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 3:
1 2 3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	652	2.1	87.8
2	CHKL	2490	8.1	97.8
3	CEFG	3813	12.5	95.2
5	EA	99	0.3	99.0
9	HD	397	1.3	74.1
10	HGH	87	0.3	91.6
11	JAB	42	0.1	100.0
12	JCD	824	2.7	95.0
13	LAB	456	1.5	93.6
14	MC	7	0.0	100.0
15	MMN	18858	61.6	91.4
16	MFG	261	0.9	68.9

17	MLK	369	1.2	34.9
18	PA	163	0.5	97.0
19	PB	142	0.5	97.9
20	PC	240	0.8	100.0
21	RAB	563	1.8	98.9
22	RCD	795	2.6	28.2
23	SA	60	0.2	84.5
24	SB	288	0.9	94.7

****

DIVISION NUMBER 2
CLUSTER NUMBER 3
CLUSTER LABEL 3
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.581949E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 6:
1 2

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	286	1.2	38.5
2	CHKL	2208	9.0	86.8
3	CEFG	3293	13.4	82.2
5	EA	92	0.4	92.0
9	HD	365	1.5	68.1
10	HGH	11	0.0	11.6
11	JAB	22	0.1	52.4
12	JCD	423	1.7	48.8
13	LAB	312	1.3	64.1
15	MMN	15665	63.9	75.9
16	MFG	201	0.8	53.0
17	MLK	329	1.3	31.2
18	PA	45	0.2	26.8
19	PB	44	0.2	30.3
20	PC	67	0.3	27.9
21	RAB	427	1.7	75.0
22	RCD	663	2.7	23.5
23	SA	8	0.0	11.3
24	SB	66	0.3	21.7

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 7:
3 4 5

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	366	6.0	49.3
2	CHKL	282	4.6	11.1
3	CEFG	520	8.6	13.0
5	EA	7	0.1	7.0
9	HD	32	0.5	6.0
10	HGH	76	1.3	80.0
11	JAB	20	0.3	47.6
12	JCD	401	6.6	46.3

13	LAB	144	2.4	29.6
14	MC	7	0.1	100.0
15	MMN	3193	52.5	15.5
16	MFG	60	1.0	15.8
17	MLK	40	0.7	3.8
18	PA	118	1.9	70.2
19	PB	98	1.6	67.6
20	PC	173	2.8	72.1
21	RAB	136	2.2	23.9
22	RCD	132	2.2	4.7
23	SA	52	0.9	73.2
24	SB	222	3.7	73.0

****

DIVISION NUMBER 3
CLUSTER NUMBER 4
CLUSTER LABEL 6
DIVISION ATTRIBUTE 4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.340812E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 12:
1 2 3 4

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	153	2.6	20.6
2 CHKL	846	14.6	33.2
3 CEFG	882	15.3	22.0
5 EA	57	1.0	57.0
9 HD	233	4.0	43.5
10 HGH	8	0.1	8.4
11 JAB	5	0.1	11.9
12 JCD	82	1.4	9.5
13 LAB	167	2.9	34.3
15 MMN	2726	47.2	13.2
16 MFG	116	2.0	30.6
17 MLK	107	1.9	10.1
18 PA	28	0.5	16.7
19 PB	3	0.1	2.1
20 PC	24	0.4	10.0
21 RAB	225	3.9	39.5
22 RCD	99	1.7	3.5
23 SA	4	0.1	5.6
24 SB	16	0.3	5.3

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 13:
5 6 7 8

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	133	0.7	17.9
2 CHKL	1362	7.3	53.5
3 CEFG	2411	12.9	60.2
5 EA	35	0.2	35.0

9	HD	132	0.7	24.6
10	HGH	3	0.0	3.2
11	JAB	17	0.1	40.5
12	JCD	341	1.8	39.3
13	LAB	145	0.8	29.8
15	MMN	12939	69.0	62.7
16	MFG	85	0.5	22.4
17	MLK	222	1.2	21.0
18	PA	17	0.1	10.1
19	PB	41	0.2	28.3
20	PC	43	0.2	17.9
21	RAB	202	1.1	35.5
22	RCD	564	3.0	20.0
23	SA	4	0.0	5.6
24	SB	50	0.3	16.4

****

DIVISION NUMBER 4
CLUSTER NUMBER 7
CLUSTER LABEL 13
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.149145E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 26:
1

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	96	0.8	12.9
2	CHKL	901	7.4	35.4
3	CEFG	1588	13.0	39.6
5	EA	19	0.2	19.0
9	HD	46	0.4	8.6
10	HGH	2	0.0	2.1
11	JAB	10	0.1	23.8
12	JCD	238	1.9	27.5
13	LAB	95	0.8	19.5
15	MMN	8775	71.8	42.5
16	MFG	16	0.1	4.2
17	MLK	33	0.3	3.1
18	PA	8	0.1	4.8
19	PB	23	0.2	15.9
20	PC	28	0.2	11.7
21	RAB	141	1.2	24.8
22	RCD	188	1.5	6.7
23	SA	4	0.0	5.6
24	SB	16	0.1	5.3

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 27:
2 3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	37	0.6	5.0

2 CHKL	461	7.1	18.1
3 CEFG	823	12.6	20.5
5 EA	16	0.2	16.0
9 HD	86	1.3	16.0
10 HGH	1	0.0	1.1
11 JAB	7	0.1	16.7
12 JCD	103	1.6	11.9
13 LAB	50	0.8	10.3
15 MMN	4164	63.9	20.2
16 MFG	69	1.1	18.2
17 MLK	189	2.9	17.9
18 PA	9	0.1	5.4
19 PB	18	0.3	12.4
20 PC	15	0.2	6.3
21 RAB	61	0.9	10.7
22 RCD	376	5.8	13.4
24 SB	34	0.5	11.2

****

DIVISION NUMBER 5
CLUSTER NUMBER 2
CLUSTER LABEL 2
DIVISION ATTRIBUTE 4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.115631E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 4:
1 2 3 4

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	37	2.2	5.0
2 CHKL	28	1.7	1.1
3 CEFG	76	4.5	1.9
4 DA	14	0.8	20.9
6 GA	5	0.3	55.6
8 HB	10	0.6	66.7
9 HD	114	6.8	21.3
10 HGH	7	0.4	7.4
12 JCD	13	0.8	1.5
13 LAB	19	1.1	3.9
15 MMN	583	34.8	2.8
16 MFG	78	4.7	20.6
17 MLK	310	18.5	29.4
18 PA	4	0.2	2.4
21 RAB	4	0.2	0.7
22 RCD	364	21.7	12.9
23 SA	4	0.2	5.6
24 SB	2	0.1	0.7
25 TAB	3	0.2	60.0
26 WAB	2	0.1	50.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 5:
5 6 7 8

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	54	1.5	7.3
2	CHKL	27	0.7	1.1
3	CEFG	117	3.2	2.9
4	DA	53	1.5	79.1
5	EA	1	0.0	1.0
6	GA	4	0.1	44.4
7	HA	7	0.2	100.0
8	HB	5	0.1	33.3
9	HD	25	0.7	4.7
10	HGH	1	0.0	1.1
12	JCD	30	0.8	3.5
13	LAB	12	0.3	2.5
15	MMN	1191	32.8	5.8
16	MFG	40	1.1	10.6
17	MLK	377	10.4	35.7
18	PA	1	0.0	0.6
19	PB	3	0.1	2.1
21	RAB	2	0.1	0.4
22	RCD	1657	45.6	58.8
23	SA	7	0.2	9.9
24	SB	14	0.4	4.6
25	TAB	2	0.1	40.0
26	WAB	2	0.1	50.0

****

DIVISION NUMBER 6
CLUSTER NUMBER 5
CLUSTER LABEL 7
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.109602E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 14:
3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	173	7.4	23.3
2	CHKL	110	4.7	4.3
3	CEFG	197	8.5	4.9
5	EA	4	0.2	4.0
9	HD	13	0.6	2.4
10	HGH	49	2.1	51.6
11	JAB	12	0.5	28.6
12	JCD	91	3.9	10.5
13	LAB	57	2.5	11.7
14	MC	5	0.2	71.4
15	MMN	1113	47.9	5.4
16	MFG	56	2.4	14.8
17	MLK	40	1.7	3.8
18	PA	74	3.2	44.0

19	PB	52	2.2	35.9
20	PC	31	1.3	12.9
21	RAB	9	0.4	1.6
22	RCD	101	4.3	3.6
23	SA	30	1.3	42.3
24	SB	109	4.7	35.9

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 15:
1 2

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	193	5.1	26.0
2	CHKL	172	4.6	6.8
3	CEFG	323	8.6	8.1
5	EA	3	0.1	3.0
9	HD	19	0.5	3.5
10	HGH	27	0.7	28.4
11	JAB	8	0.2	19.0
12	JCD	310	8.3	35.8
13	LAB	87	2.3	17.9
14	MC	2	0.1	28.6
15	MMN	2080	55.4	10.1
16	MFG	4	0.1	1.1
18	PA	44	1.2	26.2
19	PB	46	1.2	31.7
20	PC	142	3.8	59.2
21	RAB	127	3.4	22.3
22	RCD	31	0.8	1.1
23	SA	22	0.6	31.0
24	SB	113	3.0	37.2

****

DIVISION NUMBER 7
CLUSTER NUMBER 6
CLUSTER LABEL 12
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.100770E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 24:
1

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	94	2.8	12.7
2	CHKL	537	15.9	21.1
3	CEFG	510	15.1	12.7
5	EA	29	0.9	29.0
9	HD	64	1.9	11.9
10	HGH	2	0.1	2.1
12	JCD	48	1.4	5.5
13	LAB	75	2.2	15.4
15	MMN	1804	53.5	8.7
16	MFG	13	0.4	3.4

17	MLK	11	0.3	1.0
18	PA	8	0.2	4.8
19	PB	2	0.1	1.4
20	PC	17	0.5	7.1
21	RAB	114	3.4	20.0
22	RCD	33	1.0	1.2
23	SA	1	0.0	1.4
24	SB	8	0.2	2.6

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 25:
2 3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	59	2.4	7.9
2	CHKL	309	12.8	12.1
3	CEFG	372	15.4	9.3
5	EA	28	1.2	28.0
9	HD	169	7.0	31.5
10	HGH	6	0.2	6.3
11	JAB	5	0.2	11.9
12	JCD	34	1.4	3.9
13	LAB	92	3.8	18.9
15	MMN	922	38.2	4.5
16	MFG	103	4.3	27.2
17	MLK	96	4.0	9.1
18	PA	20	0.8	11.9
19	PB	1	0.0	0.7
20	PC	7	0.3	2.9
21	RAB	111	4.6	19.5
22	RCD	66	2.7	2.3
23	SA	3	0.1	4.2
24	SB	8	0.3	2.6

****

DIVISION NUMBER 8
CLUSTER NUMBER 15
CLUSTER LABEL 25
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.768743E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 50:
3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	42	2.4	5.7
2	CHKL	232	13.4	9.1
3	CEFG	259	14.9	6.5
5	EA	15	0.9	15.0
9	HD	87	5.0	16.2
10	HGH	5	0.3	5.3
11	JAB	4	0.2	9.5
12	JCD	28	1.6	3.2

13 LAB	60	3.5	12.3
15 MMN	756	43.6	3.7
16 MFG	52	3.0	13.7
17 MLK	96	5.5	9.1
18 PA	10	0.6	6.0
19 PB	1	0.1	0.7
20 PC	4	0.2	1.7
21 RAB	18	1.0	3.2
22 RCD	57	3.3	2.0
24 SB	8	0.5	2.6

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 51:
2

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	17	2.5	2.3
2 CHKL	77	11.4	3.0
3 CEFG	113	16.7	2.8
5 EA	13	1.9	13.0
9 HD	82	12.1	15.3
10 HGH	1	0.1	1.1
11 JAB	1	0.1	2.4
12 JCD	6	0.9	0.7
13 LAB	32	4.7	6.6
15 MMN	166	24.5	0.8
16 MFG	51	7.5	13.5
18 PA	10	1.5	6.0
20 PC	3	0.4	1.2
21 RAB	93	13.7	16.3
22 RCD	9	1.3	0.3
23 SA	3	0.4	4.2

****

DIVISION NUMBER 9
CLUSTER NUMBER 12
CLUSTER LABEL 14
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.765305E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 28:
3

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	160	8.0	21.5
2 CHKL	88	4.4	3.5
3 CEFG	157	7.8	3.9
5 EA	4	0.2	4.0
9 HD	5	0.2	0.9
10 HGH	12	0.6	12.6
11 JAB	12	0.6	28.6
12 JCD	86	4.3	9.9
13 LAB	51	2.5	10.5

15	MMN	1023	51.0	5.0
16	MFG	37	1.8	9.8
17	MLK	36	1.8	3.4
18	PA	73	3.6	43.5
19	PB	37	1.8	25.5
20	PC	23	1.1	9.6
21	RAB	9	0.4	1.6
22	RCD	83	4.1	2.9
23	SA	3	0.1	4.2
24	SB	108	5.4	35.5

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 29:
4 5

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	13	4.1	1.7
2	CHKL	22	6.9	0.9
3	CEFG	40	12.5	1.0
9	HD	8	2.5	1.5
10	HGH	37	11.6	38.9
12	JCD	5	1.6	0.6
13	LAB	6	1.9	1.2
14	MC	5	1.6	71.4
15	MMN	90	28.2	0.4
16	MFG	19	6.0	5.0
17	MLK	4	1.3	0.4
18	PA	1	0.3	0.6
19	PB	15	4.7	10.3
20	PC	8	2.5	3.3
22	RCD	18	5.6	0.6
23	SA	27	8.5	38.0
24	SB	1	0.3	0.3

****

DIVISION NUMBER 10
CLUSTER NUMBER 9
CLUSTER LABEL 27
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.753025E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 54:
3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	30	0.6	4.0
2	CHKL	303	5.7	11.9
3	CEFG	605	11.4	15.1
5	EA	9	0.2	9.0
9	HD	44	0.8	8.2
11	JAB	6	0.1	14.3
12	JCD	81	1.5	9.3
13	LAB	32	0.6	6.6

15	MMN	3549	66.9	17.2
16	MFG	51	1.0	13.5
17	MLK	180	3.4	17.0
18	PA	8	0.2	4.8
19	PB	15	0.3	10.3
20	PC	7	0.1	2.9
21	RAB	16	0.3	2.8
22	RCD	334	6.3	11.9
24	SB	32	0.6	10.5

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 55:
2

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	7	0.6	0.9
2	CHKL	158	13.0	6.2
3	CEFG	218	17.9	5.4
5	EA	7	0.6	7.0
9	HD	42	3.5	7.8
10	HGH	1	0.1	1.1
11	JAB	1	0.1	2.4
12	JCD	22	1.8	2.5
13	LAB	18	1.5	3.7
15	MMN	615	50.5	3.0
16	MFG	18	1.5	4.7
17	MLK	9	0.7	0.9
18	PA	1	0.1	0.6
19	PB	3	0.2	2.1
20	PC	8	0.7	3.3
21	RAB	45	3.7	7.9
22	RCD	42	3.5	1.5
24	SB	2	0.2	0.7

****

DIVISION NUMBER 11
CLUSTER NUMBER 8
CLUSTER LABEL 26
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.679180E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 52:
1

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	47	0.6	6.3
2	CHKL	717	8.5	28.2
3	CEFG	1181	13.9	29.5
5	EA	15	0.2	15.0
9	HD	37	0.4	6.9
11	JAB	4	0.0	9.5
12	JCD	112	1.3	12.9
13	LAB	39	0.5	8.0

15	MMN	6008	70.9	29.1
16	MFG	15	0.2	4.0
17	MLK	32	0.4	3.0
18	PA	2	0.0	1.2
19	PB	6	0.1	4.1
20	PC	8	0.1	3.3
21	RAB	92	1.1	16.2
22	RCD	153	1.8	5.4

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 53:
2

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	49	1.3	6.6
2	CHKL	184	4.9	7.2
3	CEFG	407	10.8	10.2
5	EA	4	0.1	4.0
9	HD	9	0.2	1.7
10	HGH	2	0.1	2.1
11	JAB	6	0.2	14.3
12	JCD	126	3.4	14.5
13	LAB	56	1.5	11.5
15	MMN	2767	73.6	13.4
16	MFG	1	0.0	0.3
17	MLK	1	0.0	0.1
18	PA	6	0.2	3.6
19	PB	17	0.5	11.7
20	PC	20	0.5	8.3
21	RAB	49	1.3	8.6
22	RCD	35	0.9	1.2
23	SA	4	0.1	5.6
24	SB	16	0.4	5.3

****

DIVISION NUMBER 12
CLUSTER NUMBER 20
CLUSTER LABEL 54
DIVISION ATTRIBUTE 4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.618560E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 108:
5 6

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	22	0.5	3.0
2	CHKL	296	6.6	11.6
3	CEFG	581	13.0	14.5
5	EA	8	0.2	8.0
9	HD	44	1.0	8.2
11	JAB	6	0.1	14.3
12	JCD	74	1.7	8.5
13	LAB	30	0.7	6.2

15	MMN	2846	63.6	13.8
16	MFG	51	1.1	13.5
17	MLK	177	4.0	16.8
18	PA	7	0.2	4.2
19	PB	13	0.3	9.0
20	PC	4	0.1	1.7
21	RAB	16	0.4	2.8
22	RCD	276	6.2	9.8
24	SB	23	0.5	7.6

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 109:
7 8

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	8	1.0	1.1
2	CHKL	7	0.8	0.3
3	CEFG	24	2.9	0.6
5	EA	1	0.1	1.0
12	JCD	7	0.8	0.8
13	LAB	2	0.2	0.4
15	MMN	703	84.9	3.4
17	MLK	3	0.4	0.3
18	PA	1	0.1	0.6
19	PB	2	0.2	1.4
20	PC	3	0.4	1.2
22	RCD	58	7.0	2.1
24	SB	9	1.1	3.0

****

DIVISION NUMBER 13
CLUSTER NUMBER 11
CLUSTER LABEL 5
DIVISION ATTRIBUTE 2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.515410E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 10:
1 2 6 8

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	21	1.0	2.8
2	CHKL	21	1.0	0.8
3	CEFG	78	3.8	1.9
4	DA	39	1.9	58.2
5	EA	1	0.0	1.0
7	HA	3	0.1	42.9
8	HB	4	0.2	26.7
9	HD	18	0.9	3.4
12	JCD	19	0.9	2.2
13	LAB	5	0.2	1.0
15	MMN	834	40.2	4.0
16	MFG	33	1.6	8.7
17	MLK	229	11.0	21.7

18 PA	1	0.0	0.6
19 PB	3	0.1	2.1
21 RAB	1	0.0	0.2
22 RCD	752	36.2	26.7
23 SA	2	0.1	2.8
24 SB	8	0.4	2.6
25 TAB	2	0.1	40.0
26 WAB	1	0.0	25.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 11:
3 4 5 7

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	33	2.1	4.4
2 CHKL	6	0.4	0.2
3 CEFG	39	2.5	1.0
4 DA	14	0.9	20.9
6 GA	4	0.3	44.4
7 HA	4	0.3	57.1
8 HB	1	0.1	6.7
9 HD	7	0.4	1.3
10 HGH	1	0.1	1.1
12 JCD	11	0.7	1.3
13 LAB	7	0.4	1.4
15 MMN	357	22.9	1.7
16 MFG	7	0.4	1.8
17 MLK	148	9.5	14.0
21 RAB	1	0.1	0.2
22 RCD	905	58.1	32.1
23 SA	5	0.3	7.0
24 SB	6	0.4	2.0
26 WAB	1	0.1	25.0

****

DIVISION NUMBER 14
CLUSTER NUMBER 18
CLUSTER LABEL 28
DIVISION ATTRIBUTE 4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.497493E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 56:
1 2 3 4

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	94	15.1	12.7
2 CHKL	41	6.6	1.6
3 CEFG	62	10.0	1.5
5 EA	2	0.3	2.0
9 HD	3	0.5	0.6
10 HGH	9	1.4	9.5
11 JAB	5	0.8	11.9
12 JCD	17	2.7	2.0

13 LAB	28	4.5	5.7
15 MMN	213	34.2	1.0
16 MFG	22	3.5	5.8
17 MLK	16	2.6	1.5
18 PA	42	6.7	25.0
19 PB	7	1.1	4.8
20 PC	16	2.6	6.7
21 RAB	7	1.1	1.2
22 RCD	18	2.9	0.6
23 SA	2	0.3	2.8
24 SB	19	3.0	6.3

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 57:
5 6 7 8

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	66	4.8	8.9
2 CHKL	47	3.4	1.8
3 CEFG	95	6.9	2.4
5 EA	2	0.1	2.0
9 HD	2	0.1	0.4
10 HGH	3	0.2	3.2
11 JAB	7	0.5	16.7
12 JCD	69	5.0	8.0
13 LAB	23	1.7	4.7
15 MMN	810	58.5	3.9
16 MFG	15	1.1	4.0
17 MLK	20	1.4	1.9
18 PA	31	2.2	18.5
19 PB	30	2.2	20.7
20 PC	7	0.5	2.9
21 RAB	2	0.1	0.4
22 RCD	65	4.7	2.3
23 SA	1	0.1	1.4
24 SB	89	6.4	29.3

****

DIVISION NUMBER 15
CLUSTER NUMBER 13
CLUSTER LABEL 15
DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.463394E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 30:
2

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	25	5.5	3.4
2 CHKL	36	8.0	1.4
3 CEFG	77	17.0	1.9
5 EA	1	0.2	1.0
9 HD	7	1.5	1.3

10	HGH	5	1.1	5.3
11	JAB	1	0.2	2.4
12	JCD	49	10.8	5.7
13	LAB	14	3.1	2.9
15	MMN	153	33.8	0.7
16	MFG	4	0.9	1.1
18	PA	13	2.9	7.7
19	PB	4	0.9	2.8
20	PC	8	1.8	3.3
21	RAB	48	10.6	8.4
22	RCD	3	0.7	0.1
23	SA	1	0.2	1.4
24	SB	3	0.7	1.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 31:
1

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	168	5.1	22.6
2	CHKL	136	4.1	5.3
3	CEFG	246	7.5	6.1
5	EA	2	0.1	2.0
9	HD	12	0.4	2.2
10	HGH	22	0.7	23.2
11	JAB	7	0.2	16.7
12	JCD	261	7.9	30.1
13	LAB	73	2.2	15.0
14	MC	2	0.1	28.6
15	MMN	1927	58.4	9.3
18	PA	31	0.9	18.5
19	PB	42	1.3	29.0
20	PC	134	4.1	55.8
21	RAB	79	2.4	13.9
22	RCD	28	0.8	1.0
23	SA	21	0.6	29.6
24	SB	110	3.3	36.2

****

DIVISION NUMBER 16
CLUSTER NUMBER 24
CLUSTER LABEL 108
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.359219E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 216:
1

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	5	0.2	0.7
2	CHKL	198	7.1	7.8
3	CEFG	382	13.6	9.5
5	EA	7	0.2	7.0

9	HD	39	1.4	7.3
12	JCD	28	1.0	3.2
13	LAB	11	0.4	2.3
15	MMN	1761	62.8	8.5
16	MFG	32	1.1	8.4
17	MLK	146	5.2	13.8
21	RAB	12	0.4	2.1
22	RCD	174	6.2	6.2
24	SB	8	0.3	2.6

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 217:
2

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	17	1.0	2.3
2	CHKL	98	5.9	3.9
3	CEFG	199	11.9	5.0
5	EA	1	0.1	1.0
9	HD	5	0.3	0.9
11	JAB	6	0.4	14.3
12	JCD	46	2.8	5.3
13	LAB	19	1.1	3.9
15	MMN	1085	64.9	5.3
16	MFG	19	1.1	5.0
17	MLK	31	1.9	2.9
18	PA	7	0.4	4.2
19	PB	13	0.8	9.0
20	PC	4	0.2	1.7
21	RAB	4	0.2	0.7
22	RCD	102	6.1	3.6
24	SB	15	0.9	4.9

****

DIVISION NUMBER 17
CLUSTER NUMBER 31
CLUSTER LABEL 31
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.332400E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 62:
3

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	167	5.3	22.5
2	CHKL	136	4.3	5.3
3	CEFG	242	7.7	6.0
5	EA	2	0.1	2.0
9	HD	12	0.4	2.2
10	HGH	10	0.3	10.5
11	JAB	7	0.2	16.7
12	JCD	244	7.8	28.1
13	LAB	73	2.3	15.0

15 MMN	1839	58.6	8.9
18 PA	31	1.0	18.5
19 PB	42	1.3	29.0
20 PC	109	3.5	45.4
21 RAB	79	2.5	13.9
22 RCD	24	0.8	0.9
23 SA	21	0.7	29.6
24 SB	102	3.2	33.6

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 63:
4 5

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	0.6	0.1
3 CEFG	4	2.5	0.1
10 HGH	12	7.5	12.6
12 JCD	17	10.6	2.0
14 MC	2	1.2	28.6
15 MMN	88	54.7	0.4
20 PC	25	15.5	10.4
22 RCD	4	2.5	0.1
24 SB	8	5.0	2.6

****

DIVISION NUMBER 18
CLUSTER NUMBER 10
CLUSTER LABEL 4
DIVISION ATTRIBUTE 6, DRDG
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.321111E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 8:
1 2 3 4

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	27	2.0	3.6
2 CHKL	27	2.0	1.1
3 CEFG	70	5.3	1.7
4 DA	14	1.1	20.9
6 GA	2	0.2	22.2
8 HB	4	0.3	26.7
9 HD	79	6.0	14.7
10 HGH	5	0.4	5.3
12 JCD	13	1.0	1.5
13 LAB	11	0.8	2.3
15 MMN	473	35.8	2.3
16 MFG	59	4.5	15.6
17 MLK	195	14.8	18.5
18 PA	3	0.2	1.8
21 RAB	4	0.3	0.7
22 RCD	330	25.0	11.7
24 SB	2	0.2	0.7
25 TAB	2	0.2	40.0

```
  26 WAB           1         0.1          25.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED   9:
  5  6  7  8
                                          % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   1 AA          10         2.8          1.3
   2 CHKL         1         0.3          0.0
   3 CEFG         6         1.7          0.1
   6 GA           3         0.8         33.3
   8 HB           6         1.7         40.0
   9 HD          35         9.8          6.5
  10 HGH          2         0.6          2.1
  13 LAB          8         2.2          1.6
  15 MMN        110        30.9          0.5
  16 MFG         19         5.3          5.0
  17 MLK        115        32.3         10.9
  18 PA           1         0.3          0.6
  22 RCD         34         9.6          1.2
  23 SA           4         1.1          5.6
  25 TAB          1         0.3         20.0
  26 WAB          1         0.3         25.0
**************************************************************
DIVISION NUMBER  19
CLUSTER NUMBER  22
CLUSTER LABEL  52
DIVISION ATTRIBUTE   4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.308916E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 104:
  5  6
                                          % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   1 AA          47         0.7          6.3
   2 CHKL       646         9.1         25.4
   3 CEFG      1044        14.8         26.1
   5 EA          14         0.2         14.0
   9 HD          36         0.5          6.7
  11 JAB          3         0.0          7.1
  12 JCD         98         1.4         11.3
  13 LAB         37         0.5          7.6
  15 MMN       4889        69.2         23.7
  16 MFG         13         0.2          3.4
  17 MLK         32         0.5          3.0
  18 PA           2         0.0          1.2
  19 PB           6         0.1          4.1
  20 PC           8         0.1          3.3
  21 RAB         86         1.2         15.1
  22 RCD        107         1.5          3.8
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 105:
  7  8
```

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
2 CHKL	71	5.1	2.8
3 CEFG	137	9.8	3.4
5 EA	1	0.1	1.0
9 HD	1	0.1	0.2
11 JAB	1	0.1	2.4
12 JCD	14	1.0	1.6
13 LAB	2	0.1	0.4
15 MMN	1119	79.9	5.4
16 MFG	2	0.1	0.5
21 RAB	6	0.4	1.1
22 RCD	46	3.3	1.6

****

DIVISION NUMBER 20
CLUSTER NUMBER 32
CLUSTER LABEL 216
DIVISION ATTRIBUTE 2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.294983E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 432:
3

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
2 CHKL	13	4.0	0.5
3 CEFG	22	6.7	0.5
9 HD	4	1.2	0.7
12 JCD	1	0.3	0.1
13 LAB	1	0.3	0.2
15 MMN	194	59.0	0.9
16 MFG	2	0.6	0.5
17 MLK	66	20.1	6.3
22 RCD	24	7.3	0.9
24 SB	2	0.6	0.7

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 433:
1 2 4 5 6 7 8

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	5	0.2	0.7
2 CHKL	185	7.5	7.3
3 CEFG	360	14.6	9.0
5 EA	7	0.3	7.0
9 HD	35	1.4	6.5
12 JCD	27	1.1	3.1
13 LAB	10	0.4	2.1
15 MMN	1567	63.3	7.6
16 MFG	30	1.2	7.9
17 MLK	80	3.2	7.6
21 RAB	12	0.5	2.1
22 RCD	150	6.1	5.3

24 SB	6	0.2	2.0

****

DIVISION NUMBER 21
CLUSTER NUMBER 26
CLUSTER LABEL 10
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.272244E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 20:
1 2

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	6	0.4	0.8
2	CHKL	20	1.2	0.8
3	CEFG	69	4.2	1.7
4	DA	38	2.3	56.7
7	HA	2	0.1	28.6
8	HB	4	0.2	26.7
9	HD	14	0.9	2.6
12	JCD	9	0.6	1.0
13	LAB	4	0.2	0.8
15	MMN	679	41.8	3.3
16	MFG	30	1.8	7.9
17	MLK	191	11.7	18.1
21	RAB	1	0.1	0.2
22	RCD	556	34.2	19.7
25	TAB	2	0.1	40.0
26	WAB	1	0.1	25.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 21:
3 4 5

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	15	3.3	2.0
2	CHKL	1	0.2	0.0
3	CEFG	9	2.0	0.2
4	DA	1	0.2	1.5
5	EA	1	0.2	1.0
7	HA	1	0.2	14.3
9	HD	4	0.9	0.7
12	JCD	10	2.2	1.2
13	LAB	1	0.2	0.2
15	MMN	155	34.5	0.8
16	MFG	3	0.7	0.8
17	MLK	38	8.5	3.6
18	PA	1	0.2	0.6
19	PB	3	0.7	2.1
22	RCD	196	43.7	7.0
23	SA	2	0.4	2.8
24	SB	8	1.8	2.6

****

DIVISION NUMBER 22
CLUSTER NUMBER 34
CLUSTER LABEL 62
DIVISION ATTRIBUTE 4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.263450E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 124:
1 2 3 4

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	65	7.5	8.7
2 CHKL	67	7.7	2.6
3 CEFG	90	10.3	2.2
5 EA	1	0.1	1.0
9 HD	9	1.0	1.7
10 HGH	6	0.7	6.3
11 JAB	4	0.5	9.5
12 JCD	70	8.0	8.1
13 LAB	25	2.9	5.1
15 MMN	410	47.1	2.0
18 PA	10	1.1	6.0
19 PB	17	2.0	11.7
20 PC	21	2.4	8.7
21 RAB	38	4.4	6.7
22 RCD	7	0.8	0.2
23 SA	7	0.8	9.9
24 SB	23	2.6	7.6

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 125:
5 6 7 8

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	102	4.5	13.7
2 CHKL	69	3.0	2.7
3 CEFG	152	6.7	3.8
5 EA	1	0.0	1.0
9 HD	3	0.1	0.6
10 HGH	4	0.2	4.2
11 JAB	3	0.1	7.1
12 JCD	174	7.7	20.1
13 LAB	48	2.1	9.9
15 MMN	1429	63.0	6.9
18 PA	21	0.9	12.5
19 PB	25	1.1	17.2
20 PC	88	3.9	36.7
21 RAB	41	1.8	7.2
22 RCD	17	0.7	0.6
23 SA	14	0.6	19.7
24 SB	79	3.5	26.0

****

DIVISION NUMBER 23
CLUSTER NUMBER 14
CLUSTER LABEL 24
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.246316E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 48:
1

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	56	2.6	7.5
2 CHKL	401	18.4	15.8
3 CEFG	341	15.7	8.5
5 EA	24	1.1	24.0
9 HD	53	2.4	9.9
12 JCD	21	1.0	2.4
13 LAB	54	2.5	11.1
15 MMN	1117	51.3	5.4
16 MFG	11	0.5	2.9
17 MLK	9	0.4	0.9
18 PA	3	0.1	1.8
20 PC	3	0.1	1.2
21 RAB	62	2.8	10.9
22 RCD	23	1.1	0.8

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 49:
2

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	38	3.2	5.1
2 CHKL	136	11.4	5.3
3 CEFG	169	14.2	4.2
5 EA	5	0.4	5.0
9 HD	11	0.9	2.1
10 HGH	2	0.2	2.1
12 JCD	27	2.3	3.1
13 LAB	21	1.8	4.3
15 MMN	687	57.6	3.3
16 MFG	2	0.2	0.5
17 MLK	2	0.2	0.2
18 PA	5	0.4	3.0
19 PB	2	0.2	1.4
20 PC	14	1.2	5.8
21 RAB	52	4.4	9.1
22 RCD	10	0.8	0.4
23 SA	1	0.1	1.4
24 SB	8	0.7	2.6

DIVISION NUMBER 24
CLUSTER NUMBER 27
CLUSTER LABEL 11
****

DIVISION ATTRIBUTE 1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.239232E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 22:
4

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	21	1.4	2.8
2	CHKL	6	0.4	0.2
3	CEFG	39	2.6	1.0
4	DA	14	0.9	20.9
6	GA	4	0.3	44.4
7	HA	4	0.3	57.1
8	HB	1	0.1	6.7
9	HD	4	0.3	0.7
10	HGH	1	0.1	1.1
12	JCD	11	0.7	1.3
13	LAB	6	0.4	1.2
15	MMN	357	23.6	1.7
16	MFG	7	0.5	1.8
17	MLK	132	8.7	12.5
21	RAB	1	0.1	0.2
22	RCD	891	59.0	31.6
23	SA	5	0.3	7.0
24	SB	6	0.4	2.0
26	WAB	1	0.1	25.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 23:
5

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	12	26.1	1.6
9	HD	3	6.5	0.6
13	LAB	1	2.2	0.2
17	MLK	16	34.8	1.5
22	RCD	14	30.4	0.5

****
DIVISION NUMBER 25
CLUSTER NUMBER 48
CLUSTER LABEL 22
DIVISION ATTRIBUTE 2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.263074E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 44:
3 7

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	3	0.4	0.4
2	CHKL	2	0.2	0.1
3	CEFG	4	0.5	0.1
4	DA	1	0.1	1.5

8	HB	1	0.1	6.7
12	JCD	7	0.8	0.8
15	MMN	206	24.3	1.0
16	MFG	4	0.5	1.1
17	MLK	72	8.5	6.8
21	RAB	1	0.1	0.2
22	RCD	542	63.9	19.2
24	SB	4	0.5	1.3
26	WAB	1	0.1	25.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 45:
4 5

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1	AA	18	2.7	2.4
2	CHKL	4	0.6	0.2
3	CEFG	35	5.3	0.9
4	DA	13	2.0	19.4
6	GA	4	0.6	44.4
7	HA	4	0.6	57.1
9	HD	4	0.6	0.7
10	HGH	1	0.2	1.1
12	JCD	4	0.6	0.5
13	LAB	6	0.9	1.2
15	MMN	151	22.8	0.7
16	MFG	3	0.5	0.8
17	MLK	60	9.0	5.7
22	RCD	349	52.6	12.4
23	SA	5	0.8	7.0
24	SB	2	0.3	0.7

TERMINAL CLUSTERS ARE LABELED:

50
51
29
55
53
109
56
57
30
217
63
8
9
104
105
432
433
20

```
 21
124
125
 48
 49
 23
 44
 45
NUMBER OF CELLS PROCESSED:     35915
NUMBER OF CELLS SKIPPED BECAUSE ATTRIBUTE STATE
VALUES WERE OUT OF RANGE:        236
```

OUTPUT FOR THE MAP-043 DATA SET

```
     PROGRAM DIVIDE

NUMBER OF CLUSTERS TO DERIVE:  19

EXTRINSIC ATTRIBUTE-STATE ACRONYMNS:
  AA   CB   FBC  HCD  HE   HF   JAB  JCD  LAB  MDE
  MFG  MKL  MH   PA   PB   RAB  RCD  TAB  WAB

NUMBER OF INTRINSIC ATTRIBUTES:   6
INTRINSIC ATTRIBUTES ACRONYMNS:
  GEOL ASPE SHAP SLOP DSTR DRDG
NUMBER OF ATTRIBUTE-STATES ASSOCIATED WITH EACH:
    10    8    6    8    5    8

 NUMBER OF MAPS PER ROW      =     1
 REFERENCE Y MAP COORDINATE  =     3
 REFERENCE X MAP COORDINATE  =     5
 MINIMUM MAP NUMBER          =     1
 MAXIMUM MAP NUMBER          =     1
 ADJUSTMENT FACTOR           =     6
 MINIMUM Y MAP COORDINATE    =     3
 MINIMUM X MAP COORDINATE    =     5
 MAXIMUM Y MAP COORDINATE    =     3
 MAXIMUM X MAP COORDINATE    =     5
```

```
**************************************************************
DIVISION NUMBER    1
CLUSTER NUMBER    1
CLUSTER LABEL    1
DIVISION ATTRIBUTE    1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.113915E+03
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED    2:
  1  7  8  9 10
                                          % OF TYPE
EXTR. TYPE   NO. CELLS   % CLUSTER   IN STUDY AREA
   1 AA            2         0.3          100.0
   4 HCD          15         2.0           93.8
   6 HF            8         1.0           36.4
   7 JAB           5         0.7           55.6
   8 JCD          47         6.1           46.1
   9 LAB           2         0.3           66.7
  11 MFG          25         3.3          100.0
  12 MKL         549        71.6           99.6
  13 MH           82        10.7          100.0
  14 PA            4         0.5           57.1
  18 TAB           8         1.0          100.0
  19 WAB          20         2.6          100.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED    3:
  2  3  4  5  6
                                          % OF TYPE
EXTR. TYPE   NO. CELLS   % CLUSTER   IN STUDY AREA
   2 CB            3         2.3          100.0
   3 FBC           4         3.0          100.0
   4 HCD           1         0.8            6.3
   5 HE            3         2.3          100.0
   6 HF           14        10.5           63.6
   7 JAB           4         3.0           44.4
   8 JCD          55        41.4           53.9
   9 LAB           1         0.8           33.3
  10 MDE          24        18.0          100.0
  12 MKL           2         1.5            0.4
  14 PA            3         2.3           42.9
  15 PB            8         6.0          100.0
  16 RAB           1         0.8          100.0
  17 RCD          10         7.5          100.0
**************************************************************
DIVISION NUMBER    2
CLUSTER NUMBER    2
CLUSTER LABEL    2
DIVISION ATTRIBUTE    1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.482271E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED    4:
  7
```

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	0.3	50.0
4 HCD	13	4.2	81.3
6 HF	1	0.3	4.5
7 JAB	5	1.6	55.6
8 JCD	32	10.2	31.4
9 LAB	1	0.3	33.3
11 MFG	13	4.2	52.0
12 MKL	160	51.1	29.0
13 MH	80	25.6	97.6
14 PA	4	1.3	57.1
19 WAB	3	1.0	15.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 5:
1 8 9 10

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	0.2	50.0
4 HCD	2	0.4	12.5
6 HF	7	1.5	31.8
8 JCD	15	3.3	14.7
9 LAB	1	0.2	33.3
11 MFG	12	2.6	48.0
12 MKL	389	85.7	70.6
13 MH	2	0.4	2.4
18 TAB	8	1.8	100.0
19 WAB	17	3.7	85.0

****

DIVISION NUMBER 3
CLUSTER NUMBER 4
CLUSTER LABEL 4
DIVISION ATTRIBUTE 5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.201443E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 8:
1

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
4 HCD	12	6.6	75.0
8 JCD	3	1.6	2.9
11 MFG	6	3.3	24.0
12 MKL	94	51.6	17.1
13 MH	67	36.8	81.7

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 9:
2 3

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	0.8	50.0
4 HCD	1	0.8	6.3
6 HF	1	0.8	4.5

```
   7 JAB          5        3.8          55.6
   8 JCD         29       22.1          28.4
   9 LAB          1        0.8          33.3
  11 MFG          7        5.3          28.0
  12 MKL         66       50.4          12.0
  13 MH          13        9.9          15.9
  14 PA           4        3.1          57.1
  19 WAB          3        2.3          15.0
**********************************************************************
DIVISION NUMBER   4
CLUSTER NUMBER   5
CLUSTER LABEL   5
DIVISION ATTRIBUTE   1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.200501E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  10:
  8
                                          % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   1 AA           1        0.3          50.0
   4 HCD          2        0.5          12.5
   8 JCD         12        3.1          11.8
   9 LAB          1        0.3          33.3
  11 MFG         12        3.1          48.0
  12 MKL        354       90.8          64.2
  13 MH           2        0.5           2.4
  18 TAB          1        0.3          12.5
  19 WAB          5        1.3          25.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  11:
  1  9 10
                                          % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   6 HF           7       10.9          31.8
   8 JCD          3        4.7           2.9
  12 MKL         35       54.7           6.4
  18 TAB          7       10.9          87.5
  19 WAB         12       18.8          60.0
**********************************************************************
DIVISION NUMBER   5
CLUSTER NUMBER   3
CLUSTER LABEL   3
DIVISION ATTRIBUTE   1, GEOL
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.157847E+02
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED   6:
  4  5
                                          % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   2 CB           3        3.4         100.0
   3 FBC          1        1.1          25.0
```

4	HCD	1	1.1	6.3
5	HE	2	2.3	66.7
6	HF	11	12.6	50.0
7	JAB	3	3.4	33.3
8	JCD	43	49.4	42.2
9	LAB	1	1.1	33.3
10	MDE	22	25.3	91.7

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 7:
2 3 6

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
3	FBC	3	6.5	75.0
5	HE	1	2.2	33.3
6	HF	3	6.5	13.6
7	JAB	1	2.2	11.1
8	JCD	12	26.1	11.8
10	MDE	2	4.3	8.3
12	MKL	2	4.3	0.4
14	PA	3	6.5	42.9
15	PB	8	17.4	100.0
16	RAB	1	2.2	100.0
17	RCD	10	21.7	100.0

****

DIVISION NUMBER 6
CLUSTER NUMBER 11
CLUSTER LABEL 7
DIVISION ATTRIBUTE 4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.972990E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 14:
3 4 5 6

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
3	FBC	3	10.0	75.0
5	HE	1	3.3	33.3
6	HF	1	3.3	4.5
7	JAB	1	3.3	11.1
8	JCD	10	33.3	9.8
12	MKL	2	6.7	0.4
14	PA	3	10.0	42.9
15	PB	8	26.7	100.0
16	RAB	1	3.3	100.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 15:
8

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
6	HF	2	12.5	9.1
8	JCD	2	12.5	2.0
10	MDE	2	12.5	8.3
17	RCD	10	62.5	100.0

```
**************************************************************
DIVISION NUMBER   7
CLUSTER NUMBER   7
CLUSTER LABEL   9
DIVISION ATTRIBUTE   5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.772127E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  18:
  2
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   1 AA           1        1.3         50.0
   7 JAB          1        1.3         11.1
   8 JCD         11       13.9         10.8
   9 LAB          1        1.3         33.3
  11 MFG          7        8.9         28.0
  12 MKL         41       51.9          7.4
  13 MH          13       16.5         15.9
  14 PA           2        2.5         28.6
  19 WAB          2        2.5         10.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  19:
  3
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   4 HCD          1        1.9          6.3
   6 HF           1        1.9          4.5
   7 JAB          4        7.7         44.4
   8 JCD         18       34.6         17.6
  12 MKL         25       48.1          4.5
  14 PA           2        3.8         28.6
  19 WAB          1        1.9          5.0
**************************************************************
DIVISION NUMBER   8
CLUSTER NUMBER   8
CLUSTER LABEL  10
DIVISION ATTRIBUTE   4, SLOP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.631421E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  20:
  4  5
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   4 HCD          2        1.0         12.5
   8 JCD         10        5.0          9.8
   9 LAB          1        0.5         33.3
  11 MFG         11        5.5         44.0
  12 MKL        176       87.6         31.9
  13 MH           1        0.5          1.2
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  21:
  6  7  8
```

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	0.5	50.0
8 JCD	2	1.1	2.0
11 MFG	1	0.5	4.0
12 MKL	178	94.2	32.3
13 MH	1	0.5	1.2
18 TAB	1	0.5	12.5
19 WAB	5	2.6	25.0

****

DIVISION NUMBER 9
CLUSTER NUMBER 14
CLUSTER LABEL 18
DIVISION ATTRIBUTE 2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.511984E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 36:
5 6

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	12.5	50.0
7 JAB	1	12.5	11.1
12 MKL	1	12.5	0.2
13 MH	4	50.0	4.9
19 WAB	1	12.5	5.0

ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 37:
1 2 3 4 7 8

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
8 JCD	11	15.5	10.8
9 LAB	1	1.4	33.3
11 MFG	7	9.9	28.0
12 MKL	40	56.3	7.3
13 MH	9	12.7	11.0
14 PA	2	2.8	28.6
19 WAB	1	1.4	5.0

****

DIVISION NUMBER 10
CLUSTER NUMBER 17
CLUSTER LABEL 21
DIVISION ATTRIBUTE 2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
0.469176E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED 42:
5 6 7

EXTR. TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
1 AA	1	3.6	50.0
8 JCD	1	3.6	1.0
11 MFG	1	3.6	4.0

```
  12 MKL          21        75.0          3.8
  19 WAB           4        14.3         20.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  43:
  1  2  3  4  8
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   8 JCD          1        0.6          1.0
  12 MKL        157       97.5         28.5
  13 MH           1        0.6          1.2
  18 TAB          1        0.6         12.5
  19 WAB          1        0.6          5.0
**************************************************************
DIVISION NUMBER  11
CLUSTER NUMBER  12
CLUSTER LABEL  14
DIVISION ATTRIBUTE   5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.469038E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  28:
  1  2
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   3 FBC          2       12.5         50.0
   5 HE           1        6.3         33.3
   6 HF           1        6.3          4.5
   7 JAB          1        6.3         11.1
   8 JCD          7       43.8          6.9
  12 MKL          2       12.5          0.4
  14 PA           1        6.3         14.3
  16 RAB          1        6.3        100.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  29:
  3
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   3 FBC          1        7.1         25.0
   8 JCD          3       21.4          2.9
  14 PA           2       14.3         28.6
  15 PB           8       57.1        100.0
**************************************************************
DIVISION NUMBER  12
CLUSTER NUMBER  10
CLUSTER LABEL   6
DIVISION ATTRIBUTE   5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.456157E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  12:
  1
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   4 HCD          1        1.7          6.3
```

```
   6 HF            7        11.9          31.8
   7 JAB           3         5.1          33.3
   8 JCD          34        57.6          33.3
   9 LAB           1         1.7          33.3
  10 MDE          13        22.0          54.2
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  13:
  2  3
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   2 CB            3        10.7         100.0
   3 FBC           1         3.6          25.0
   5 HE            2         7.1          66.7
   6 HF            4        14.3          18.2
   8 JCD           9        32.1           8.8
  10 MDE           9        32.1          37.5
**************************************************************
DIVISION NUMBER  13
CLUSTER NUMBER   9
CLUSTER LABEL  11
DIVISION ATTRIBUTE   5, DSTR
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.445671E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  22:
  1
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   6 HF            6        13.3          27.3
  12 MKL          28        62.2           5.1
  18 TAB           1         2.2          12.5
  19 WAB          10        22.2          50.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  23:
  2  3
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   6 HF            1         5.3           4.5
   8 JCD           3        15.8           2.9
  12 MKL           7        36.8           1.3
  18 TAB           6        31.6          75.0
  19 WAB           2        10.5          10.0
**************************************************************
DIVISION NUMBER  14
CLUSTER NUMBER  15
CLUSTER LABEL  19
DIVISION ATTRIBUTE   2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.441769E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  38:
  6  8
                                        % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
```

```
   6 HF            1        10.0          4.5
   7 JAB           3        30.0         33.3
   8 JCD           1        10.0          1.0
  12 MKL           3        30.0          0.5
  14 PA            2        20.0         28.6
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  39:
  1  2  3  4  5  7
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   4 HCD           1         2.4          6.3
   7 JAB           1         2.4         11.1
   8 JCD          17        40.5         16.7
  12 MKL          22        52.4          4.0
  19 WAB           1         2.4          5.0
**************************************************************
DIVISION NUMBER  15
CLUSTER NUMBER  22
CLUSTER LABEL  28
DIVISION ATTRIBUTE    2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.399495E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  56:
  7  8
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   3 FBC           1        10.0         25.0
   7 JAB           1        10.0         11.1
   8 JCD           7        70.0          6.9
  14 PA            1        10.0         14.3
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  57:
  2  3  6
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   3 FBC           1        16.7         25.0
   5 HE            1        16.7         33.3
   6 HF            1        16.7          4.5
  12 MKL           2        33.3          0.4
  16 RAB           1        16.7        100.0
**************************************************************
DIVISION NUMBER  16
CLUSTER NUMBER  16
CLUSTER LABEL  20
DIVISION ATTRIBUTE    3, SHAP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.392896E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  40:
  1  2
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   4 HCD           2         1.3         12.5
```

```
   8 JCD          2        1.3          2.0
   9 LAB          1        0.7         33.3
  11 MFG         10        6.5         40.0
  12 MKL        137       89.5         24.9
  13 MH           1        0.7          1.2
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  41:
  3  6
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   8 JCD          8       16.7          7.8
  11 MFG          1        2.1          4.0
  12 MKL         39       81.3          7.1
**************************************************************
DIVISION NUMBER  17
CLUSTER NUMBER  27
CLUSTER LABEL  23
DIVISION ATTRIBUTE   3, SHAP
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.379750E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  46:
  3
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   6 HF           1       12.5          4.5
   8 JCD          3       37.5          2.9
  12 MKL          2       25.0          0.4
  19 WAB          2       25.0         10.0
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  47:
  1  2
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
  12 MKL          5       45.5          0.9
  18 TAB          6       54.5         75.0
**************************************************************
DIVISION NUMBER  18
CLUSTER NUMBER  19
CLUSTER LABEL  37
DIVISION ATTRIBUTE   2, ASPE
INFORMATION LOSS RESULTING FROM DIVISION OF CLUSTER,
  0.366962E+01
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  74:
  3  7
                                      % OF TYPE
EXTR. TYPE  NO. CELLS  % CLUSTER  IN STUDY AREA
   8 JCD          8       33.3          7.8
  11 MFG          3       12.5         12.0
  12 MKL          8       33.3          1.5
  13 MH           3       12.5          3.7
  14 PA           2        8.3         28.6
ATTRIBUTE STATES ASSOCIATED WITH NEW CLUSTER LABELED  75:
```

1 2 4 8

EXTR.	TYPE	NO. CELLS	% CLUSTER	% OF TYPE IN STUDY AREA
8	JCD	3	6.4	2.9
9	LAB	1	2.1	33.3
11	MFG	4	8.5	16.0
12	MKL	32	68.1	5.8
13	MH	6	12.8	7.3
19	WAB	1	2.1	5.0

TERMINAL CLUSTERS ARE LABELED:

8
15
36
42
43
29
12
13
22
38
39
56
57
40
41
46
47
74
75

NUMBER OF CELLS PROCESSED: 900

NUMBER OF CELLS SKIPPED BECAUSE ATTRIBUTE STATE VALUES WERE OUT OF RANGE: 0

The vita has been removed from the scanned document