To the Graduate Council:

I am submitting herewith a thesis written by Dennis Lee Campbell entitled "The Mapping of Desert Alluvial Soils through the Use of Remote Sensing and GIS Techniques." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree Masters of Science, with a major in Engineering Science.

T. Paludan

T. Paludan, Major Professor

We have read this thesis
and recommend its acceptance:

Jack H. Hanson

Alfonso Peyral

Accepted for the Council:

Lew Minkel

Associate Vice Chancellor
and Dean of the Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at The University of Tennessee, Knoxville, I agree that the Library shall make it available to borrowers under the rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature Dennis L. Campbell

Date 8/18/93

THE MAPPING OF DESERT ALLUVIAL SOILS THROUGH THE USE OF REMOTE SENSING AND GIS TECHNIQUES

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Dennis Lee Campbell

December 1993

DEDICATION

This thesis is dedicated to my parents

Ms. Violet Campbell and the late Mr. Booker T. Campbell

and

to those who perished during the African famines of 1985 through 1987. Their suffering compelled me to pursue this research undertaking.

ACKNOWLEDGEMENTS

I would like to extend my appreciation to my advisor and committee chairman, Dr. Ted Paludan, and other committee members, Dr. Jack Hansen and Dr. Al Pujol for their assistance and suggestions. I would be remiss if I did not express a very special thank you to Dr. Jack Hansen. His level of excellent in research is exemplary and I am honored to have had the privilege to have worked with him. As a first-rate engineering scientist (as well as human being) he is, without question, an asset to the University.

I am particularly grateful for the assistance and encouragement I have received from Mr. Cliff Whitehead, Planning Director, Tennessee Wildlife Resources Agency and other employees: Mr. Bill Turner, Mr. David Tirpak and Mr. Bill Christie.

I am also grateful to Mr. Charles Finchum, Executive Director of the Boys and Girls Clubs of Middle Tennessee for time granted away from my work with the youth to finish this research project.

I would like to thank Dr. Neil Weber, Director of the Remote Sensing Center at Murray State University and the Laboratory Director, Ms. Patricia Bumba for the usage of their facilities and equipment in the early stages of this research endeavor.

I would also like to thank the entire London family for their love, support, and friendship that I have cherished over the years. God has blessed me with the opportunity to meet a family of winners.

Finally, I would like to thank my mother, Ms. Violet Campbell, for her never ending encouragement, support and love; my father, the late Booker T. Campbell, for his inspiration, love, and legacy of commitment; and God Almighty, for giving me the strength to persevere.

ABSTRACT

Originating from the uplands of semi-arid regions, streams and rivers transport sediments downslope. Steep changes in the gradient of the land, it is widely believed, forced the deposition of sediments over vast regions of the desert plain through time. Various studies indicate that these sediments, or alluvium, may occupy as much as 70 percent of the surrounding desert basins (Merifield, 1987). Interestingly, since antiquity a number of soils in these areas have been known to be agriculturally productive (Encyclopedia Britannica, 1978).

Many of the potientially fertile soils in semi-arid regions, particularly in underdeveloped regions, remain unmapped. Less expensively, improved digital classification of satellite imagery could expedite the process of generating vital soil data. Methods of enhancing the digital classification of soils in a semi-arid environment are investigated. Specifically, image classification techniques are combined with the use of ancillary data and then discussed based of its ability to discriminate semi-arid soils. Results may be observed in map format.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	The Importance of Mapping Desert Alluvial Soils	1
	Research Objective	4
II.	STUDY AREA	6
	Selection Criteria	6
	Geography and Climate	7
	Geomorphological Description	9
	Soils and Agriculture of the Big Smokey Valley Survey Area	10
III.	SOIL CLASSIFICATION TECHNIQUES	15
	Conventional Methods of Producing a Soil Survey	15
	Spectral Classification Techniques	16
	Principal Component Analysis	21
	Geographical Information Systems	25
IV.	DISCRIMINATION OF SEMI-ARID SOILS: INTEGRATING SATELLITE IMAGERY WITH ANCILLARY DATA	27
	Previous Work	27
	Study Procedures	30
	Description of Study Sites	34
V.	ANALYSIS RESULTS	36

VI. CONCLUSION..... 48

LIST OF REFERENCES..... 49

APPENDICES..... 52

APPENDIX A. TABLES..... 53

APPENDIX B. FIGURES..... 61

APPENDIX C. SOFTWARE TERMINOLOGY..... 76

APPENDIX D. TERMINOLOGY..... 79

VITA..... 82

LIST OF TABLES

TABLE		PAGE
1.	Format of variance-covariance matrix	24
2.	Variance-covariance matrix for all seven TM bands for image site ET2	54
3.	Eigenvalues computed for ET2	55
4.	Error matrix, percent correct, commission error, and omission error for ET2MAX (non-transformed, singular signature)	56
5.	Error matrix, percent correct, commission error, and omission error for ET2MAX6 (non-transformed, merged signature)	57
6.	Error matrix, percent correct, commission error, and omission error for ET2MAXSEP (transformed, separate signature)	58
7.	Error matrix, percent correct, commission error, and omission error for ET2CLAS1 (transformed, merged signature)	59
8.	Soil type classification comparison	60

LISTS OF FIGURES

FIGURE		PAGE
1.	Illustration of an alluvial fan.....	62
2.	Areal perspective of Study Area coinciding with soil survey.....	63
3.	Location of Study Site ET2.....	64
4.	Soil Conservation Survey soil areas.....	65
5.	Discarded classification distribution of a sample signature.....	66
6.	Acceptable classification distribution of a sample signature.....	67
7.	ET2MAX (non-tranformed signature input, singular).....	68
8.	ET2MAX large scale view with superimposed soil areas defined.....	69
9.	ET2MAX6 (non-transformed, signature input merged).....	70
10.	ET2MAX6 large scale view with superimposed soil areas defined.....	71
11.	ET2MAXSEP (transformed signature input singular).....	72
12.	ET2MAXSEP large scale view with superimposed soil areas defined.....	73
13.	ET2CLAS1 (transformed, signature input merged).....	74
14.	ET2CLAS1 large scale view with superimposed soil areas.....	75

CHAPTER I

INTRODUCTION

The Importance of Mapping Arable Desert Alluvial Soils

Covering more than a third of the earth's surface area, arid and semi-arid lands range from sparsely vegetated desert fringe areas that support a variety of nomadic inhabitants to vast areas of low crop productivity in Africa, Latin America, and Asia that sustain large populations of people and livestock. Due to climatic factors, less than 50 percent of these regions are suitable for agricultural uses while human and climatological effects continue to reduce the percentage of arable desert lands. Overgrazing and the elimination of forests on mountain slopes combined with radical climatic changes, as observed in the Sub-Sahara, have a devastating effect on desert populations during times of severe drought (Hills, 1966).

Hence, increasing environmental pressures on the fragile ecosystems of semi-arid developing regions necessitate more accurate and rapid methods of delineating, characterizing, and mapping the remaining soils most suitable for agricultural development and management. Soil surveys, as will be discussed, produced through conventional means require extensive field and laboratory investigations. Consequently,

such surveys demand sizeable capital resources and a considerable amount of time to complete depending on the size of the area (Gerrard, 1981).

The cultivation of massive layers of accumulated deposits from mountain hillslopes, commonly termed alluvial fan soils, dates back to antiquity (Encyclopedia Britannica, 1979). Located in semi-arid regions, where cloud shading and vegetation cover are for the most part sparse, these soils are particularly discernable on images produced from earth resources surveillance satellites (Merifield, 1987).

For more than a decade, soil investigation studies have progressively employed Landsat image data to produce generalized soil surveys at the reconnaissance level. Usually mapped at scales of 1:50,000 to 1:250,000, Landsat imagery has been successfully used as a base for the production of semi-detailed maps. Of particular value, has been the demonstrated ability of using Landsat imagery to provide information for underdeveloped arid regions in the form of semi-detailed maps never before obtainable through conventional methods. In Syria, for example, Landsat imagery combined with ground observations of soils and plant communities, has been used to produce a landcover evaluation of the entire country at a scale of 1:100,000. In Kenya, Landsat data has also been employed in a similar manner in the mapping of the Narok region (Colwell, 1983).

Although the use of remote sensing for soil type

delineation is extremely complex because of the inherent dynamic and complex biological nature of soils, the approach offers advantages over traditional methods of conducting resource surveys. Some of the many advantages include (Colwell, 1983):

(1) The ability to achieve a synoptic view and to rapidly acquire a suitable soil database.

(2) The use of multispectral data, thus broadening informational data.

(3) The capability to detect change in landscape cover by multiple coverage in the time domain.

(4) The reduction of expenses involved in the completion of a soil survey.

More recently, the digital integration of ancillary data into the data gathering image analysis process presents an opportunity previously unachieved: image differentiation of semi-arid soils beyond the generalized reconnaissance level to the more detailed soil type classification level.

Research Objective

Through the integration of image and ancillary information, this study is an attempt to assess the use of remote sensing combined with GIS (geographical information system) techniques to digitally classify semi-arid alluvial soils from satellite imagery at the soil type level of classification. Using specific image classification algorithms, soils exhibited on a satellite image of a soil survey area in Big Smokey Valley, Nevada are mapped. Image classification techniques are performed based on selecting supervised training sets in known soil types as defined by the superimposition of U.S. Soil Conservation Service (SCS) soil classifications on areas of concern having multispectral satellite data. Classification acceptability is assessed by a comparison of semi-automated results with the actual soil survey areas (ancillary data) stored as vector data in a geographical information system. The results, displayed in map format, may be used to assist in the areal expansion of a soil survey investigations .

As was stated, the primary focus of this thesis is to investigate the usefulness of remote sensing and geographical information system (GIS) techniques in the mapping of desert alluvial soils at the soil type level of differientation. The investigation begins with a discussion of why the pursuit of such a study is paramount. Chapter Two discusses the criteria

used for the selection of the study area, as well as its overall geography, climate, geomorphology, agriculture, and soils. Chapter Three reviews conventional approaches in the generation of a soil survey and discusses image classification techniques and geographical information systems. A review of the direct applications of image classification and GIS approaches in scene differientation is given in Chapter Four, along with a discussion of the research procedures applied to this study. A description of the specific study site examined is also presented. Chapter Five reveals the results produced from the application of research procedures used in Chapter Four. The final chapter, Chapter Six, examines the practicality of the results and suggests areas for further research.

CHAPTER II

STUDY AREA

Selection Criteria

Alluvial fan development in semi-arid regions exist in numerous desert areas globally. However, to facilitate the acquisition of necessary satellite imagery and soil survey data, a site in the United States was sought. In the western half of the country, within the semi-arid areas of the Basin and Range physiographic province, any alluvial fan region present exhibiting some degree of agriculture development was considered a potential candidate for analysis. Satellite imagery costs and the availability of soil survey reports, however, dramatically decreased the number of sites from which to chose. With the privatization of Landsat, one Thematic Mapper scene can cost thousands of dollars. Consequently, the acquisition of non-proprietary imagery (inexpensive imagery produced before the commercialization of Landsat) was sought.

Shortly after securing access to the use of needed image processing and geographical information systems, the author contacted the Remote Sensing Center at Murray State University concerning the possibility of obtaining non-proprietary Thematic Mapper data. Fortunately, out of the four scenes available, the one which encompassed the Big Smokey Valley

Region in central Nevada proved to be the most useful and consequently was chosen for the purposes of this study (coincidently, the author had already considered this area as one of the candidate areas and obtained a soil survey report of the area before having any prior knowledge of the existence of non-proprietary imagery of the region!). The seven band Thematic Mapper (TM) scene was recorded on June 30, 1984 and has 30 meter resolution.

In brief, the Big Smokey Valley Area was chosen for the following reasons: (1) significant alluvial fan development present in a semi-arid region; (2) some degree of agricultural development existing in the fan region; (3) the existence of a soil survey report; and (4) the obtainability of non-proprietary data.

Geography and Climate

Specifically defined, alluvial fans are layers of sediments deposited at a mountain base primarily by streams emerging downslope and various other fluvial processes. Where fluvial paths are free of lateral constraints, fan deposits may extend miles beyond the base into adjacent plain areas. Typically fan-shaped in areal view, these deposits commonly comprise 70 percent of flat neighboring areas and up to 30 percent of the total surface area in a desert region (Blissenbauch, 1954). Thus, because of their prominence in

arid and semi-arid zones, most are considered a characteristic desert landform (Encyclopedia Britannica, 1979).

Geographically located in the Great Basin section of the Basin and Range physiographic province in northwest Nye County, Nevada, the study area is an internally drained valley enclosed by mountains, low foothills, and broad alluvial fans. Extending from the Lauder and Nye County border to the north, to the town of Tonopah to the south, the area as surveyed by the U.S. Soil Conservation Service is elongated in shape and comprises 567,680 acres in total area (Soil Conservation Service, 1972).

The survey area comprises two valleys. To the north, one valley consisting of alluvial fans and a large playa known as Lake Tab, is flanked by the Toiybe and Toquima Mountain ranges on its eastern and western side, respectively. In the southern portion of the survey area, the other valley extends west from the upper section of the San Antonio mountains across broad alluvial fans and the valley floor known as Tonopah flat to the hills of the Esmeralda-Nye County line. Elevations of the landforms are as follows. The mountain ranges have a maximum elevation of 8,500 feet, the valley floor has an elevation of about 4,800 feet, the elevations of the alluvial fans in the northern valley are 5,600 to 6,500 feet and 4,500 to 6,000 feet in the southern valley (Soil Conservation Service, 1972).

Climatologically, because these valleys are on a plateau between major mountain ranges, a semi-arid climate is

produced. Situated to the west, the Sierra Nevada Mountain range captures much of the moisture content of air moving inland from the Pacific during fall, winter, and spring seasons. Moist air, forced to rise over each successive mountain range, loses much of its moisture on the western slopes. Air traveling down the eastern slope undergoes adiabatic warming and is relatively dry by the time it reaches the interior valleys. Consequently, precipitation in the survey area is light and averages only five inches annually. The primary source of moisture in the survey area that arrives during the summer is the northward flow of warm air from the Gulf of Mexico (Soil Conservation Service, 1972).

Geomorphological Description

Geometrically, if not deformed by local tectonism or either coalescing adjacent fans, the shape of alluvial fans approximate that of a half-buried cone. The profile shape of a fan from the apex to its margin is convex and slightly concave upward, respectively. From a radial profile, the general range of slope is two to five degrees. However, because gradients are dependent on fan size, dominant rock type of source area, and depositional and tectonic histories, slope angles may differ greatly. Fans of recently tectonically uplifted areas generally have steeper slopes. The highest point of elevation on an alluvial fan is termed the apex of

the fan and develops where water laden materials emerge from mountain areas. The fanhead is the upper section of the fan encompassing the apex. The midfan is the area of the fan encompassing the apex and the lower margins of the fan or the base. The downcutting or entrenchment of stream channels into mountain areas and the resulting deposition of sediments are the primary processes associated with the formation of alluvial fans (Blissenbach, 1954 and see Figure 1).

The interplay of climatological and geographical factors contributes significantly to the universal human occupation and cultivation of many of these mountain border areas. The presence of mountains in a desert area contributes to decreased aridity, particularly on the windward side of the mountain exposed to moist air. Land adjoining these areas benefit from the presence of groundwater in the alluvial fans received from the moister mountain area (Hills, 1966).

Soils and Agriculture of the Big Smokey Valley Survey Area

Soil, a natural medium for the growth of land plants, is a mixture of rock fragments, minerals, organic matter, air, water and living organisms (Gerrard, 1981). Soil typing in part is done by digging holes to expose the soil profiles. A soil profile is a sequence of layers or horizons in the soil which extend from the surface to unaltered bedrock. Various profiles are examined and compared with other profiles in the

region. Similar profiles are classified and named in accordance with the nationally accepted taxonomy. The soil classification system adopted for the study site was accepted by the National Cooperative Survey in 1965 (Soil Conservation Service, 1972).

Soils with major horizons that are similar in thickness, arrangement, or in other major characteristics, make up a soil series. The name of a soil series is generally coined or named after a geographic feature or town located within the proximity of the soils of interest. Because of differences in the surface layer, slope, stoniness, and other characteristics that affect soil usefulness, soil series are divided into phases. Names associated with soil phases indicate features that affect its management. For example, Mazuma coarse sandy loam, is one of several phases within the Mazama series (Soil Conservation Service, 1972).

Because of the impracticality of representing on a map small, scattered portions of soil of another type within an area that is predominantly of a recognized soil phase, mapping units are used. Mapping units of two types are shown on the soil map of the Big Smokey Valley Area: soil complexes and soil associations. Soil complexes consists of areas of two or more soils so intricately mixed in areal extent that it is not practical to represent them separately on a soil map. Joined by a hyphen, the names of a soil complex is chosen to coincide with those predominantly cited in the field. Soil

associations consists of neighboring soils of extensive size represented as one unit on a soil map because of lack of field work required to map them separately or a need to generalize similar soil types into categories of larger areal extent. Association names are also separated by a hyphen, such as the Tomel-Laxel association. As exhibited on the soil map (see Figure 4), the symbols identify the names of each mapping unit. Since the soils were mapped at high and low intensity levels, they are distinguished on the map by the use of different symbols. Soils mapped at high or intense amount of field work are labeled with a capital letter followed by a small letter. Soils mapped at a low intensity are designated by two capital letters (Soil Conservation Service, 1972).

For broad interpretive purposes, soil associations in the Big Smokey Valley Survey Area can be grouped into four general kinds of landscapes. Soils of the uplands and mountains form in the residuum and colluvium are derived from basalt, andesite, rhyolite, and volcanic ash and are primarily comprised of the Old Camp-Pinwater-Rock Outcrop Associations. These soils comprise 14 percent of the survey area. Soils of this type are also found in the southwestern part of the survey region (Soil Conservation Service, 1972).

Soils surveyed on valley fill, outwash plains, and alluvial fans were formed from alluvium derived from basalt, rhyolite, tuffs, limestone, and shale. The surface layers of the soils are gravelly and coarse, moderately coarse, or

medium in texture. The four associations: Dobel-Bluewig-Lyda, Mazuma-Quima-Broyles, Wardenot-Laxal, Vigius-Unsel-Ardivey that comprise this group of soils make up 49 percent of the survey area (Soil Conservation Service, 1972).

Formed in alluvium, soils on alluvial fans and aprons are derived from basalt, rhyolite, andesite, tuffs, limestone, and granite rocks. Stumble-Tipperary and Yomba-Temper-Koyen are the two soil associations predominately sited in these areas and comprise 20 percent of the survey area (Soil Conservation Service, 1972).

Playas and soils on flats and basins are formed from silty lacustrine sediment derived from a variety of rocks. The two soil associations, Playa and Umberland-Yobe-Orizaba are present within these landforms and comprise 17 percent of the survey area (Soil Conservation Service, 1972).

The agricultural uses of soils in the Big Smokey Valley survey area are predominantly for raising beef cattle. Consequently, soils are irrigated for the production of native grasses, grass-legume mixtures, and alfalfa agricultural crops. Irrigation of these soils is needed to supplement the sparse rainfall as streams provide very small amounts of water after the middle of June. Soils of alluvial fans and valley plains support sparse vegetation and limited livestock grazing. Sparse vegetation is also located in the uplands and lower mountains for grazing, however, due to the steepness of the area cattle grazing is legally prohibited.

All totalled, 2,000 acres have been irrigated. Public lands in the survey area are administered by the Bureau of Land Management (Soil Conservation Service, 1972).

CHAPTER III

SOIL CLASSIFICATION TECHNIQUES

Conventional Methods of Producing a Soil Survey

Soil surveys in the form of maps and reports provide assessments of the properties of soils which may be used to create a database. In part, such a database forms the basis for making informed decisions in rural and urban planning operations. Homogeneous units, referred to as soil series, describe the land use potential and their responses to changes in a manner useful to management (Soil Conservation Service, 1972). Because the properties of soils change gradually in vertical and horizontal directions, the process of producing a soil survey is complex. Soil boundaries, although depicted as discrete parcels of land, are transitional and intergrades are common (Gerrard, 1981).

Aerial photo interpretation combined with soil field sampling forms the basis for soil series mapping in which land areas of similar characteristics are mapped into units or series, defined by specific types of soil horizons. Mapping units are defined according to a national soil classification system so as to maintain consistency in nomenclature, and are evaluated with respect to important local agricultural uses. Major boundaries between mapping units are delineated (Soil

Conservation Service, 1972).

Field samples for each prospective mapping unit are collected and subjected to physical, chemical and mineralogical laboratory analysis. Soil laboratory results form the basis for the classification of a mapping unit. The classification system used in this study, is defined in Soil Taxonomy, a publication issued by the Soil Conservation Service, U. S. Department of Agriculture. In the final product, maps depicting patterns of soils classes within the survey region are produced within each mapping unit (Soil Conservation Service, 1972).

Spectral Classification Techniques

Through the process of multispectral classification, Landsat Thematic Mapper (TM) images composed of brightness values for pixels recorded from multiple regions of the electromagnetic spectrum can be transformed into informational classes of interest. Combinations of the electromagnetic radiation bands characterize surface features at varying levels of complexity. In general, digital image processing methods attempt to statistically cluster brightness or spectral values that characterize terrain classes with a minimum degree of error (as will be discussed). Two basic methods implement this process: supervised and unsupervised classification.

Supervised Classifiers

Generally supervised classification techniques are preferred when informational classes that comprise a scene are known thus allowing the analyst to select and sample spectral values within each class. The method employs pixel averages from areas of known identity to classify individual pixels whose spectral values have not yet been associated with a unique identity. Training sets (also called training samples) are sets of pixels chosen that correspond to specific classes. Pixels with values that correspond to the statistics of the training sets are thus identified and may be displayed (Campbell, 1987).

Generally, the process is initiated by the selection of classification schemes based on categories of interest such as soil type, vegetation cover, etc. Training sets are selected using field collected data, map, or other surrogate data (i.e., soil survey maps) which correspond to areas of interest identified on the image. Statistics computed from sample pixel data sets of each category of interest are then used to create training set signature values, that define a category, for example a the terrain class (i.e., soil type). Once created, signature data may be tested (as will be discussed) to determine if they truly represent each class.

After a set of reliable signatures is created and evaluated, image classification is performed. Analyzing each

pixel independently, the relationship of each pixel to each training set signature is defined according to criteria established in the classification algorithm or decision rule chosen. Pixel data values that pass the criteria contained within the algorithm, are assigned to a class determined by its associated signature. Three types of supervised classification algorithms that use signatures to classify the unknown pixels in a scene will be discussed.

The parallelepiped classification algorithm defines the range of values within the chosen training data sets taken from each spectral band. Unidentified pixels whose values fall within the range defined by the training sets are assigned to the designated category. Parallelepiped algorithms may also use standard deviations within training areas as range or threshold boundaries. Although the procedure assigns a majority of pixels to the designated classes, pixels of unknown classes that satisfy the criteria for more than one class maybe assigned to a class erroneously (Campbell, 1987).

The minimum distance algorithm defines classes based on the minimum distances between the means of each chosen training area and the pixels of unknown classes. Each pixel is assigned to the closest class mean or centroid based on distance. The calculation of the distance from the mean of the training data to each unknown pixel is performed using the Euclidean distance formula. However, because the algorithm

does not account for variability contained within and between classes, classes may overlap. If pixel values lie equally in the same range of values or are equi-distant from more than one class mean, then large classification errors may result using the previously described methods (Campbell, 1987; Jensen, 1986).

The third type of supervised classification method, the maximum likelihood algorithm, accounts for the variability in the brightness values of the class before individual pixels are assigned to a designated terrain classes. The approach is based on the assumption that pixels in both the training data and the specified classes exhibit normal or Gaussian frequency distribution in each band. One type of maximum likelihood approach based on the Bayes theorem and used in this study, does not necessarily that the occurrence of the classes are equally probable (Campbell, 1987).

Unsupervised Classifiers

If the location of the classes or the number of classes present in the scene is unknown, unsupervised classification methods are commonly used. Requiring minimal input from the analyst, the unsupervised classification algorithms automatically cluster pixels by various numerical operations. One method is similar to the minimum distance to means procedure previously described. Clusters of spectral values

are displayed and can be compared with the location of categories of interest. Training data generated by the analyst is not required (Jensen, 1986). Most operate in a two-pass mode. The first pass groups pixel values as it reads through the data set and a mean value is computed with each cluster generated. In the second iteration, the entire data set, pixel by pixel, is assigned to the mean vectors. The clusters produced are then matched to terrain classes specified by the analyst (Erdas, Inc., 1991; Campbell, 1987).

Advantages, Disadvantages, and Errors

In the application of these two broad image classification techniques, two types of errors are produced. Errors of commission assign pixels to classes erroneously. In omission errors, pixels are not assigned to their appropriate class (Campbell, 1987).

Supervised classification allows an estimate of accuracy of training sets through the examination of training data prior to scene classification. However, the selection of training data imposes a classification structure on the data which may not be consistent with the natural spectral classes in the image. Consequently, it may be difficult to differentiate classes using training data if the training set selector incorrectly picks training sets that do not correspond to the spectral data of a classification group.

Unsupervised classification does not require extensive prior knowledge of the region considered and because input is minimal the opportunity for human error is minimized. On the other hand, correspondence between the natural grouping of data and associated informational categories may be inadequate (Erdas, Inc., 1991; Campbell, 1987; Jensen, 1986;). Supervised classification is the chosen method of analysis used in this study.

Principal Component Analysis

The analysis of original bands of remotely sensed digital data, extracted from computer-compatible tapes and displayed, can be extremely difficult. Interpretation problems are primarily attributed to the significant amount of data present in the multispectral image. Data redundancy or correlation between bands arises from the following phenomena (Schowengerdt, 1983):

(1) Natural spectral correlation. For example, vegetation records low levels of reflectance in bands 4 and 5, and registers high reflectance values in 6 and 7.

(2) Topographic shading. Because topographic slope and aspect remains a constant throughout each band of the imaged area, the degree of topographic shading is identical for all bands even though the effect on each band varies.

(3) Electromagnetic spectrum spectral sensitivities overlap between neighboring spectral bands.

Additionally, another problem arises in the choice of bands to be used to create a color composite image that displays an optimum amount of information contained in the several band data set. Hence, the choice of the best bands for analysis is not straightforward (Cahvez et al., 1989).

The application of a principle components transformation (also referred to as principle component analysis, PCA, or Karhunen-Loeve analysis) to raw remote sensor data often has resulted in a more interpretable image with a minimum number of bands needed as data input. Consequently, computer processing time and storage space are also decreased. Before the application of the technique, a measure of the mutual change in pixel values between any two bands of data is required (Jensen, 1986).

Covariance is a measure of the joint variation of two variables about their common mean. The covariance (Cov_{kj})

between pixel values in bands k and j is equal to (Jensen, 1986):

$$Cov_{kj} = \frac{n\sum_{i=1}^{n}(BV_{ik} \times BV_{ij}) - \sum_{i=1}^{n} BV_{ik} \sum_{i=1}^{n} BV_{ij}}{n(n-1)}$$

where BV_{ik} is the i^{th} pixel value of band k, and BV_{ij} is the i^{th} measurement of band j with n pixels in the area of interest. Results of the data computed from the covariance equation above can be arranged in a 7 X 7 variance-covariance symmetric matrix (see Table 1, below).

Table 1. Format of variance-covariance matrix.

	Band 1	Band 2	Band 3...	Band 7
Band 1	Variance	$Cov_{1,2}$	$Cov_{1,3}$...	$Cov_{1,7}$
Band 2	$Cov_{2,1}$	Variance	$Cov_{2,3}$...	$Cov_{2,7}$
Band 3	$Cov_{3,1}$	$Cov_{3,2}$	Variance	$Cov_{3,7}$
.				.
.				.
.				.
Band 7	$Cov_{7,1}$	$Cov_{7,2}$	$Cov_{7,3}$	Variance

The principal component analysis procedure may then be performed on the 7 X 7 matrix of correlated multispectral data. The method involves essentially the calculation of eigenvectors for the matrix and in so doing, the transformation translates and rotates the original coordinate axes of the data into axes which correspond to the maximum and minimum directions of variance in the data. The new axes, termed eigenvalues or principle components, represent the transformation of redundant, correlated sensor data to non-redundant, uncorrelated multispectral data whose variance

diminishes and therefore the importance of its inclusion as data input, increases with each component computed (Jensen, 1986; Siegal and Gillespie, 1980).

Generally, the first principal component generated accounts for the largest amount of total variance in the data set. The second axis accounts for much of the remaining variance and is uncorrelated with the first axis. This process continues until the remaining axes generated contribute negligible amounts of variance. As a rule, the first three components make up 99.9 percent of the total variance contained within seven bands of data. Thus these three components, for the most part, represent the maximum amount of non-redundant information that can be extracted from the set (Campbell, 1987).

Geographical Information Systems

Demonstrably, geographical information systems (GIS), have proven to be powerful tools for the analysis and management of geographical data. A GIS is a spatial database designed to accommodate extensive volumes of geographical data derived from a variety of sources. The system allows the user to store, retrieve, manipulate, display, and analyze the data according to user specifications. Results generated may then serve as a test bed for the study of environmental processes, planning decisions, trends, and similar types of analyses

(Environmental Systems Research Institute, Inc., 1991).

Geographical data input is achieved by the representation of data by locational (x,y) and non-locational (i.e., attributes) characteristics. A surface, for example, may exist at an x,y location and have z attributes which may be qualitative (i.e., soil type at a location) or quantitative (i.e., the elevation at the same location). Four fundamental types of topological data are stored in a GIS: points, lines, polygons, and surfaces. Vector and raster topological representation are the two most commonly used formats for the storage of these data types (Burrough, 1986).

Map features are implicitly represented by vector or polygon encoding which registers their boundaries by a listing of coordinates. Links or line segments are numbered and associated with nodes, as well as left and right polygon identification. Coordinates are encoded with each node. Greater complexity in data structure increases processing time for more sophisticated spatial analyses (Burrough, 1986). Raster or cellular structures represent topological explicitness; small data cells comprise points, lines, and polygons. Grid cells encoded by superimposing a matrix over an area of interest with attribute information assigned to each cell are called pixels (Burrough, 1986). A vector based geographical information system (Arc/Info) was employed in this study.

CHAPTER IV

DISCRIMINATION OF SEMI-ARID SOILS: INTEGRATING SATELLITE IMAGERY WITH ANCILLARY DATA

Previous Work

As previously discussed, multispectral classification of semi-arid soils is based on the belief that different soil types have distinguishable electromagnetic radiation reflections which may be defined by the use of various statistical classifiers. However, the accuracy of multispectral classification using statistical classifiers alone may be highly variable (Hutchinson et al., 1982).

Therefore, a number of methods other than the use of new classification algorithms, have been employed to improve classification accuracy. One such method has been to increase the number of spectral representative samples used in defining classes. Increasing the number of spectral observations, however, does not necessarily improve classification results. Another approach is to consider class attributes, such as size, shape, pattern, and the spatial association of other objects. Because many attributes are not easily derived from images, particularly in the case of soil discrimination, this method has also proven ineffective (Hutchinson et al., 1982).

For more than two decades, many have suggested that the

optimum utility of remotely sensed data is ultimately determined by the ability to integrate the data with other ancillary spatial information (Colwell, 1983). Typically, the incorporation of ancillary data for the purposes of classification improvement has followed one of three approaches: integrating data either before, during, or after classification using classifier operations, postclassification sorting techniques, or stratification methods (Hucthinson et al., 1982).

In the merging of ancillary data with classifier operators, several approaches have been followed. One method has been to add to the number of spectral bands used for classification by creating an ancillary data channel or band. Such a method, termed "the logical channel" approach, without the appropriate modification of spectral sampling algorithms improves classification only marginally and presents new problems in the development of class statistics. For example, a common problem is that new observations (extracted from ancillary data) require samples to describe signatures as well as their association (covariance) with other spectral data (Hutchinson et al., 1982).

Other attempts have primarily involved the modification of the maximum likelihood classifier discussed previously. Because in many classification operators, prior probabilities are not considered or are assumed to be equal for all classes, the classifier may be modified through the use of prior

probabilities before classification based on the estimated spatial composition of classes that comprise a scene or known occurrences between classes and ancillary data. Guassian distributions are assumed in the application of maximum likelihood classifiers, an assumption, however, which may not be valid when applied to ancillary data (Hutchinson et al., 1982).

In a multispectral scene an object class is seldom represented by a single spectral class. Consequently, spectral classes are merged after classification into clusters which represent object classes. However, as is often the case, a single subset of a spectral class may represent subsets of more than one class. Postclassification sorting handles these as problem spectral classes, and with the employment of ancillary data, incorrectly assigned pixels are assigned their appropriate class (Hutchinson et al., 1982).

The use of ancillary data with stratification techniques involves the division of the scene of interest into smaller homogenerous areas or strata based on a selection or criterion rule. Each stratum may then be processed separately and then merged in the final product (Campbell, 1987). Because the technique is performed prior to the execution of image classification, the technique is deterministic in nature and gradations in the scene are not easily accommodated. Additionally, incorrect stratification selection rules can erroneously affect classification results produced (Hutchinson

et al., 1982). Thus, care must be taken in its administration. The application of such an approach is employed in this study.

Study Procedures

As has been stated, the research purpose is to differentiate semi-arid alluvial soil classes by the integration of satellite imagery and ancillary data. The test site area is the alluvial soils of the Big Smokey Valley (BSV) survey area in central Nevada. The use of two computer systems, a satellite image processing system (Erdas) and a geographical information system (Arc/Info), made the attainment of the objective possible. Various image processing modules in Erdas were executed once the image data, approximately one-eighth the size of the original Landsat TM scene, was acquired on magnetic tape (reel to reel) media and loaded onto a computer tape drive.

Using the Erdas READ module, image data was displayed and through an extensive comparison of topographic features used to locate road network intersections which formed georeference points, the image site precisely corresponding to the same area as observed on BSV soil survey maps was located on the eastern- most edge of the scene. The image file of the area, named "NEVALL", contained all seven bands of data for the area of interest. The Erdas CUTTER module was used to extract only the image data which directly corresponded to the

area observed on the corresponding soil maps (Figure 2).

The image file NEVALL was then subjected to the geo-rectification process. To obtain the needed ground control points, twenty-five intersections from the image were matched with twenty-five georeferenced road intersections from manually digitized soil survey maps stored in the geographical informational system, Arc/Info. Using state plane map projection, the coordinates of the two sets of twenty-five road intersections (TM image and map coordinates) were entered into the Erdas module called COORDN used to compute the transformation matrices. The matrix represents a set of coefficients from polynomial equations computed from ground control points and is used to transform the coordinates into a chosen map projection or convert from one map projection to another. The size and accuracy of the matrix depends on the order of the transformation (see Appendix D, for definition).

The final step of the geo-referencing process involves image rectification or the process of making image data conform to a chosen map projection within a given range of error tolerance. The acceptable root mean square error of tolerance or the distance between the actual ground control and the same transformed point displayed on the image NEVALL was plus or minus 0.8 of a pixel. The RECTIFY module in Erdas performed this transformation with a choice of three resampling methods: bilinear interpolation, cubic convolution, and nearest neighbor resampling techniques. Because of its

ability to maintain the original values of the scene, nearest neighbor resampling was the chosen method (see Appendix D, for further definition).

Once geo-referenced the Erdas module CUTTER was again employed to extract forests, cultivated agricultural lands, urban areas, and mountainous regions from the scene leaving only the land cover type of interest for analysis, namely, bare and sparsely vegetated alluvial soils. All subsequent image processing results produced from the scene, as well as subset scenes were prefixed by the letters "ET." A site located in the southern portion of the scene, comprising 2,598 acres, was chosen for analysis and labelled ET2 (see Figures 2 and 3, Appendix B).

The Erdas principal component analysis module, PRINCE operates on the original ET2 data set to produce a transformed ET2 data set. Classification results of both data sets could later be assessed. Additionally, as will be discussed, PRINCE was also used to optimize band selection for the classification of the non-transformed ET2 spectral data.

For the purposes of this study, soil areas delineated by vector lines are superimposed and georeferenced with both sets of data through the use of Arc/Info LIVE-LINK. Such a configuration, allows the analyst to interactively select sample spectral values or training samples (displayed in the form a digitized polygons) contained within each soil type area as overlayed on the image. Using a documented

methodology for the selection of training samples (Campbell, 1987), a primary objective was to locate sets of pixels that represent accurately the spectral variation exhibited within each soil type. All totalled, one hundred and forty-two training samples were chosen from both data sets (71 for non-transformed and 71 for transformed).

Training sets are collected from the separate non-transformed and transformed ET2 data. The Erdas module SIGEXT is then used to compute a set of statistical data that defines the training samples in terms of the number of bands, the mean, standard deviation, and variance of the data in each band for every sample.

Once training samples, now termed signatures, are created, each are evaluated on the basis of its ability to represent spectrally distinct soil types as chosen by the analyst. Using programs ALARM and HISTOGRAM in the Erdas module SIGMAN, signatures are evaluated. ALARM allows the analyst to view the estimated classification area that the signatures represent compared to the original soil type areas as exhibited on the display device. Signatures were saved or discarded on the basis of their ability to represent only their associated soil types (see Figures 5 and 6). If a signature displayed an accurate representation, relative to its associated soil area borders, the second test was performed. HISTOGRAM computes the frequency of occurrence of brightness values represented by a signature, and displays

this information graphically in the form of a histogram. A signature exhibiting a histogram with bimodal frequency distribution is indicative of more than one spectral soil class and was discarded. Those possessing a unimodal distribution (indicative of a signature representing one class) were accepted. Twenty-five signatures were accepted for the supervised classification of the non-transformed ET2 data. Thirty-one signatures were accepted for the transformed ET2 data.

In this study, the training set data was either handled as a single separate set for each sample area within a soil type or all samples within a soil type were averaged to give a single mean value describing the overall soil type. The Erdas MAXCLAS module was used to perform maximum likelihood classification for both the non-transformed and transformed data with separate classification calculations for averaging or non-averaging of training sets within each soil type.

Description of Study Site ET2

The study site ET2 represents soils found on valley fills, outwash plains, and alluvial fans in the region. The site located in the southern central portion of the scene, is approximately 2,958 acres in size and comprises six different soil types, each described as follows (Soil Conservation Service, 1972):

Deerlodge stony loam soils (DEC) are moderately sloping brownish gray soils found in small, irregularly shaped regions on convex, dissected alluvial fans.

Laxel gravelly fine sandy loam (LMA) are nearly level light gray soils found in irregularly shaped areas on the toeslopes of alluvial fans.

The Tomel-Laxal association (TN) are light gray, gently to moderately sloping soils found in large irregularly shaped areas on dissected alluvial fans. Wardenot sandy loam soils (WA) are light brownish gray soils found in very large irregularly shaped regions on alluvial fans and toeslopes.

Monte Cristo-Playas complex (MO) are nearly level soils found in small, medium, and long narrow bands on flood plains. The soils are light gray in color. Mazuma fine sandy loam soils (MCA) are light gray in color and are found in medium sized irregular regions on board alluvial fans and toeslopes.

CHAPTER V

ANALYSIS RESULTS

Supervised classification techniques were applied to study the non-transformed and transformed ET2 study site data employing two training set signature input methods. Classification accuracy assessments were made for the following generated classification maps: non-transformed with the signature input singularly, non-transformed with the signature input merged, transformed with the signature input singularly, and transformed with the signature input merged. Essentially, two types of data are exhibited on each of the four maps (Figures 8, 10, 12 and 14). The colored regions represent the supervised classification of pixels; each color is representative of the spectral classification pattern of each designated soil area. The superimposed blue vector lines represent the actual SCS soil areas.

Classification accuracy may be assessed through the computation of overall accuracy, percent correct, error of commission, and error of omission, each of which may be expressed as follows (Jensen, 1986; University of California; 1973):

Overall accuracy = total number of correct interpretations/ total number of interpretations (incorrect + correct) x 100

Percent correct = number correct interpretations (pixels) of a soil type /total number (pixels) of a soil type present x 100

Error of Commission = number of incorrect interpretations (pixels) of a soil type /total number of a soil types (pixels) selected by the classifier x 100

Error of Omission = 100 - Percent correct

The classification results are presented in Figures 7 - 14 and Tables 4 - 8 (Appendices B and A, respectively). The number and class of each pixel contained with within each SCS soil type are presented in the form of error matrices (see Tables 4 through 7). The estimates of classification accuracy are based on these matrix tables. Each matrix reveals the number of pixels classified for each soil type as well as the total number of pixels each soil type comprises. Manual digitization of each SCS soil area superimposed over the

classification maps made the acquisition of such data possible.

Using the Erdas modules DIGSCRN and CUTTER, each superimposed SCS area was digitized and stored as a separate file. Erdas BSTATS was then used to display the number and the class of the pixels contained with a given SCS area. A discussion of results is given for the following soil four classification maps (Figures 8, 10, 12, and 14) and Tables 4 through 8).

Image Site ET2: Non-transformed Data

Each set of signature statistical data may be entered into the classification algorithm singularly (one signature at a time), or as a merged (average) statistical value representing each soil group. A comparison of the output products produced based on each of the two signature input approaches is presented for the non-transformed data case, and the PCA transformed case.

Examination of soil classification map ET2MAX (signatures input singularly, see Figures 7 and 8) in comparison with ET2MAX6 (using the identical signature set, but the data sets for each soil type were averaged before input, see Figures 9 and 10) reveals differences between the two non-transformed data classifications. The most noticeable difference is the Monte Cristo (MO) classification shown in green. The Monte

Cristo soil classification, shown in Figures 8 and 9, encroaches on the Laxel-Tomal (LMA), the Mazuma (MCA), and the Tomel(TN) soil series mainly at the middle-right side of the image. Although it exhibits a high percent accuracy and a relatively low omission error (71.0 percent and 29.0 percent, respectively), the high commission error (81.9 percent) shown in Table 4 primarily collaborates this observation.

In comparison, for ET2MAX6 (non-transformed, merged), MO exhibits noticeably less encroachment. The lower MO commission error (71.5 percent) seen in Table 5 confirms this observation and appears to have been produced at the expense of a lower percent correct and a higher error of omission (63.1 and 36.9 percent, respectively) versus those seen in ET2MAX (71.0 and 29.0 percent, respectively). As a result of less encroachment of the MO soils in ET2MAX6, the classification accuracy of the LMA appears to have improved (see Figure 10). The higher percent accuracy and lower error of omission of LMA (78.5 percent and 21.5 percent, respectively) in Table 5 compared to lower percent accuracy and higher error of omission of LMA (68.0 and 32.0 percent, respectively) in Table 4, agrees with this observation. Again, along with these improvements, an increase in the commission error occurred (20.8 percent in Table 5 versus 15.0 percent in Table 4).

The classification accuracy of the TN soils in ET2MAX6 (Figure 10) appears noticeably less than those classified in ET2MAX (Figure 8). The percent accuracies and error of

omission agree with what is presented in the images. In Table 5, TN has a lower percent correct and a higher omission error (44.2 and 55.8 percent, respectively) than in Table 4 (55.7 and 44.4 percent). Accordingly, the commission error is less (11.8 percent in Table 5 compared to 17.8 percent in Table 4).

With the exception of the TN soils, the soil classification of ET2MAX6 is superior to that of the ET2MAX soils. Improved classification appears to be primarily attributable to the reduction in the encroachment of the MO soils in ET2MAX6 relative to those seen in ET2MAX.

Image Site ET2 - Transformed Data

Identical classification techniques were used for the transformed case as were applied to the non-transformed image data. As shown in Figures 11 through 14, the soil classification borders for ET2MAXSEP (singular) and for ET2CLAS1 (merged) more closely agree with those defined by the soil survey than for the non-transformed cases previously discussed. Because of spectral signature inseparability between of the Mazuma and Monte-Cristo soils in the transformed case, the soils were considered one soil type and labelled MCA-MO.

Table 3 gives the variances (or eigenvalues) of the seven principal components for the ET2 site. These signatures used in a maximum likelihood classification produced the results

exhibited in ET2MAXSEP and ET2CLAS1. As shown, 97.88 percent of the scene variance of non redundant, uncorrelated data was compressed into the first three principal components. Compared to the non-transformed output, overall classification results exhibit a noticeable improvement in accuracy, which apparently is attributable to the classification optimization of data produced through the execution of principal component analysis.

In observing classification of the MCA-MO soils in ET2MAXSEP (transformed, signature input singular), the encroachment of these soils is significantly reduced relative to the same soils encroachment exhibited in the non-transformed classifications. A comparison with the non-transformed, singular and merged results is made by averaging the MCA and MO values for each accuracy measure, namely, percent correct, commission and omission errors given in Tables 4 and 5. Importantly, a lower commission error is achieved for the transformed singular (41.3 percent, Table 6) as compared to non-transformed, singular calculated 60.0 percent average. Even though this lowers the percent accuracy to 49.1 percent and increases the omission error to 50.9 percent for MCA-MO compared to the non-transformed singular case calculated averages of 61.1 and 36.7 percent respectively, the beneficial effect is to improve overall classification accuracy.

The LMA classification appears noticeably improved in

ET2MAXSEP compared to the non-transformed classifications of the same soil. Exhibiting a higher percent accuracy and a lower omission error (83.6 and 16.4 percent) compared to those in Table 4 (68.0 and 32.0 percent) and Table 5 (78.5 and 21.5 percent), the commission error is not as improved in comparison of Table 6 values (19.2 percent) with those in Table 4 (15.0 percent) and Table 5 (20.2 percent).

Improved accuracy is also observed in the classification of the TN soils in ET2MAXSEP. Relative to the non-transformed percent accuracies and omission errors for TN presented in Table 4 (55.7 and 44.4 percent) and in Table 5 (44.2 and 55.8 percent), an improvement in these categories is seen in Table 6 (69.7 and 30.3 percent, respectively). The commission error of these soils, however, has increased as exhibited in Table 6 (24.1 percent) in comparison to Table 4 (15.8 percent) and Table 5 (20.8 percent).

A lesser degree of encroachment of the MCA-MO soils is seen in ET2CLAS1 (transformed, input merged) than the non-transformed examples previously presented. However even less encroachment is observed in ET2MAXSEP (transformed, singular). Examination of Tables 4 through 7 confirm this observation. MCA-MO, as shown in Table 7 has a higher percent error and a lower omission error (65.0 and 35 percent, respectively) relative to the average computed for MCA-MO in Tables 4 (with averages of 63.3 and 36.7 percent) and Table 5 (with averages of 64.0 and 35.9 percent). The lower commission error of

MCA-MO (54.2 percent) in Table 7 relative to higher commission errors of the same soils seen in Tables 4 and 5 (with averages of 60.8 and 60.0 percent), however, is primarily brought about by the reduced encroachment of MCA-MO on other soil types as shown in ET2CLAS1. Because MCA-MO in the transformed singular classification has the lowest commission error (41.3 percent, Table 6), encroachment of the soil classes on other soils is considerably less for ET2MAXSEP.

The TN soils appear to have a higher classification accuracy than the same soils observed in the non-transformed data cases and the transformed, singular data. Again, higher percent correct and lower omission errors are exhibited (76.0 and 24.0, percent, respectively) in Table 7 compared to those in Table 4 (55.7 and 44.4 percent, respectively) and in Table 5 (44.2 and 55.8 percent, respectively). These improved accuracies are apparently achieved at the expense of a higher commission error (34.8 percent) versus the non-transformed cases (17.7 percent and 11.8 percent).

Classification accuracy differences between the LMA in ET2CLAS1 and the same soil in the non-transformed data appear marginal. A noticeable difference in classification is observed in a comparison with the transformed, singular data, ET2MAXSEP. A comparison of the percent correct, commission and omission errors of LMA shown in Table 7, indicate that all accuracies are within 5 percent for the same accuracy tests computed for LMA in Tables 4 and 5.

Overall comparison of the two non-transformed and transformed images and data might best be expressed in the overall accuracy given at the bottom of Tables 4 through 7. The lowest overall accuracy is for the non-transformed singular signature classification (65.1 percent); followed by the non-transformed merged signature (68.2 percent); followed by the transformed merged signature (70.5 percent). The transformed singular signature exhibited the highest overall accuracy with 75.9 percent. The high overall accuracy exhibited by the transformed, singular signature input was primarily brought about by the more contained spectral signature used for the classification of MCA-MO which produced the lowest commission error (41.3 percent) and thus reduced the encroachment of this category on other categories.

An analysis of the results considering SCS soil type acreages as compared to classified acreages was performed to determine the percent area correct (Table 8). In this test the number of acres classified in each of the soil types was compared with the true acres given by the SCS soil types. For each of the four classification tests, a percent difference, the weighted mean percentage area incorrect and correct was calculated.

The highest weighted mean percentage area correct was the transformed singular signature (91.4 percent) followed by the transformed merged signature (77. 3 percent); followed by the non-transformed merged signature (73.5 percent); and

lastly the non-transformed merged signature (71.3 percent). These figures confirm the accuracy statements made on the basis of the images and the figures for percent correct, commission errors and omission errors formerly discussed.

A discussion of these accuracies relative to three other published studies reveals an interesting comparison. Using Landsat Multispectral Scanner (MSS) image data at 90 meter resolution, Westin and Lemme (1978) conducted a similar study to classify three unvegetated soils at the soil association level. Spectral signatures were extracted through the use of overlays of soil association boundaries manually superimposed over the scene representing combinations of four MSS bands (4, 5, 6, and 7). Representative of each classification category, merged signatures on the basis of their group means and variances were input into a minimum distance classifier. On the basis of spectral data point count, a 43 percent overall accuracy was achieved. This method was similar to the non-transformed merged signature (Table 5) of this study which had an overall accuracy of 68.2 percent.

In an attempt to discriminate three soil types, May and Petersen (1975) employed an airborne multispectral scanner with a 13 channel wavelength interval of 0.4 to 2.6 micrometers. Signatures were generated from training areas located in the field that exhibited a uniform spectral response. Signatures that were input into a supervised classification algorithm, singularly and merged, produced

different classification results. On the basis of the percentage of area mapped, merged signatures obtained classification accuracies for the three soils of 85 percent, 75 percent, and 27 percent. When input singularly, classification accuracy improved and were "closely related to the field survey" (May and Petersen, 1975). In substantiation of this claim, however, no corresponding data was presented for the singular classification. Petersen's results of 85 percent, 75 percent, and 27 percent, compare to the non-transformed results of this study for six soil types of 86.0, 78.5, 44.2, 66.4, 63.1, and 65.0 percent correct (Table 5).

Perhaps in a manner that more closely parallels the investigations of this study relative to the others reviewed thus far, Kornblau and Cipra (1983) used Landsat MSS data to map semi-arid soils at the soil type level of discrimination. Using a supervised methodology, signatures were selected through the superimposition of soil maps overlaid on image data recorded in four bands. Comparable to merged signatures, each signature representative of each of the eleven soil units was input into a Gaussian maximum likelihood algorithm. Results were presented in the form of a weighted mean agreement with soil maps produced a weighted accuracy of 32.8 percent on the basis of 877 pixels. This compares to the weighted mean percentage correct based on area for the non-transformed merged signatures of 73.5 percent for this study (Table 8). The lesser accuracy obtained in these studies is

probably attributable to one or a combination of the following reasons: (1) the studies reviewed employed the use of Landsat MSS data possessing lesser spatial and spectral resolution compared to Landsat TM data used in this study; and (2) an inadequate classification algorithm was chosen by Westin and Lemme 1978; the minimum distance classifier they used did not account for data variability.

CHAPTER VI

CONCLUSION

On the basis of results exhibited in the classification maps and the accuracy analyses presented, desert alluvial soils may be differentiated at the soil type level. To this end, the execution of principal component analysis used in supervised classification proved superior in the classification of alluvial soils relative to identical methods applied to the same, non-transformed image data. Even greater soil classification accuracies was achieved when non-averaged transformed signature data was used with the maximum likelihood classifier.

It may be concluded from these results that a soil spectral classification database may be produced and extensive mapping of alluvial soils may be pursued. In conclusion, Thematic Mapper (TM) satellite imagery can be used to delineate and map desert alluvial soils located in a semi-arid region and thus contribute to locating those soils generally most suitable for agricultural uses.

LIST OF REFERENCES

LIST OF REFERENCES

1. Blissenbach, Erich. Geology of Alluvial Fans in Semi-arid Regions. Bulletin of the Geological Society of America, Vol. 65, pp. 175 - 190, 1954.

2. Burrough, P.A. Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford, England, pp. 13-37, 1986.

3. Campbell, James B. Introduction to Remote Sensing. The Guildford Press, New York, New York, pp. 269-472.

4. Cahvez, Pat S., and Kwarteng, Andrew Y. Extracting Spectral Constrast in Landsat Thematic Mapper Image Data Using Selective Principal Component Analysis. Photogrametric Engineering and Remote Sensing, Vol. 55, pp. 339 - 348, 1989.

5. Colwell, Robert N., ed. The Manual of Remote Sensing. American Society of Photogrammetry, Second Edition, Vol. 2, Falls Church, Virginia, pp. 2111-2228, 1983.

6. Encyclopedia Britannica, Inc. Alluvial Fans. Encyclopedia Britannica, Inc., Vol. 1, pp. 611-617, 1978.

7. Environmental System Research Institute, Inc. Understanding GIS: The Arc/Info Method. Environmental System Research Institute, Inc., Redlands, California, pp. 1-27, 1991.

8. Erdas, Inc. Erdas Field Guide. Second Edition, Atlanta, Georgia, pp. 31-167, 1991.

9. Gerrard, A.J. Soils and Landforms: an Intregration of Geomorphology and Pedology. George Allen and Unwin, London, England, pp. 10-26, 1981.

10. Hills, E.S. Arid Lands: A Geographical Appraisal. Methuen and Co Ltd, London, England, pp. 219-220, 1966.

11. Hutchinson, Charles F. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement. Photogrammetric Engineering and Remote Sensing, Vol. 48, pp. 123-130, 1982.

12. Jensen, John R. Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice-Hall, Englewood Cliffs, New Jersey, pp. 63-226, 1986.

13. Kornblau, M.L. and J. E. Cipra. Investigation of Digital Landsat Data for Mapping Soils Under Range Vegetation. Remote Sensing of Environment, Vol. 13, pp. 103 - 112, 1983.

14. May, G. A. and G. W. Petersen. Spectral Signature Selection for Mapping Unvegetated Soils. Remote Sensing of Environment, Vol. 4, pp. 211 - 220, 1975.

15. Merifield, P.M. Mapping Desert Alluvial Deposits from Satellite Imagery. Progress in Desert Research, Rowan and Littlefield, Totowa, New Jersey, pp. 198-214, 1987.

16. Schowengerdt, Robert A. Techniques for Image Processing and Classification in Remote Sensing. Academic Press, New York, New York, pp. 159-161, 1983.

17. Siegal, Barry S., and Alan R. Gillespie. Remote Sensing in Geology. Wiley and Sons, New York, pp. 409-411, 1980.

18. United States Department of Agriculture, Soil Conservation Service. Soil Survey of Big Smokey Valley, Nevada: Part of Nye County. United States Government Printing Office, Washington, D.C., pp. 1-13, 1972.

19. University of California, Berkeley. An Integrated Study of Earth Resources in the State of California Based on ERTS-1 and Supporting Aircraft Data. University of California, Berkeley, Berkeley, California, pp. 48-52, 1973.

20. Westin, F. C. and G. D. Lemme. Landsat Spectral Signatures: Studies with Soil Associations and Vegetation. Photogrammetric Engineering and Remote Sensing, Vol. 44, pp. 315 - 325, 1978.

APPENDICES

APPENDIX A

Table 2. Variance-covariance matrix for all seven TM bands for image site ET2.

		$Band_k$						
		1	2	3	4	5	6	7
$Band_x$	1	19.63	12.39	19.88	17.71	19.78	3.41	19.78
	2	12.39	8.90	13.83	12.60	14.11	1.97	14.11
	3	19.88	13.83	24.10	22.59	25.05	1.79	25.05
	4	17.71	12.60	22.59	24.03	25.30	-1.11	25.05
	5	19.78	14.11	25.05	25.30	32.05	-0.60	32.05
	6	3.41	1.97	1.79	-1.11	-0.60	7.63	-0.60
	7	19.78	14.11	25.05	25.30	32.05	-0.60	32.05

Table 3. Eigenvalues computed for ET2.

	Eigenvalues	Variance %	Total %	Angle (degrees)	Scale
1	128.61	86.73	86.67	44.40	1
2	11.88	8.01	94.68	51.25	1
3	4.65	3.14	97.81	75.99	1
4	1.98	1.33	99.15	97.00	1
5	0.76	0.51	99.66	81.09	1
6	0.51	0.34	100.00	89.01	1
7	0.00	0.00	100.00	90.00	1

Table 4. Error matrix, percent correct, commission error, and omission error for ET2MAX (non-transformed, singular signature.

	Interpreted Soil Classes								
		Number of Pixels							
		DEC	LMA	TN	WA	MO	MCA	Prt. Ct. (%)	C.E. (%)	O.E (%)
Actual SCS Soil Classes	DEC	670	101	77	0	1	0	78.9	26.4	21.1
	LMA	55	3983	232	4	1465	118	68.0	15.8	32.0
	TN	186	330	1579	24	647	71	55.6	17.7	44.4
	WA	0	0	24	432	165	41	65.4	8.5	34.6
	MO	0	216	0	2	556	9	71.0	81.9	29.0
	MCA	0	55	6	10	239	386	55.6	38.2	44.4
	Overall Accuracy = 65.1 %									

Table 5. Error matrix, percent correct, commission error, and omission error for ET2MAX6 (non-transformed, merged signature).

	Interpreted Soil Classes								
		Number of Pixels							
		DEC	LMA	TN	WA	MO	MCA	Prt. Ct. (%)	C.E. (%)	O.E (%)
Actual SCS Soil Classes	DEC	730	68	51	0	0	0	86.0	44.7	14.0
	LMA	151	4598	105	7	704	292	78.5	20.8	21.5
	TN	413	744	1251	29	303	97	44.2	11.8	55.8
	WA	17	6	7	439	136	57	66.4	9.2	33.6
	MO	1	265	0	4	494	19	63.1	71.5	36.9
	MCA	10	127	5	5	97	462	65.0	50.1	35.0
	Overall Accuracy = 68.2 %									

Table 6. Error matrix for percent correct, commission error, and omission error for ET2MAXEP (transformed, separate signature).

	Interpreted Soil Classes							
		Number of Pixels						
		DEC	LMA	TN	WA	MCA-MO	Prt. Ct. (%)	C.E. (%)	O.E. (%)
Actual SCS Soil Classes	DEC	753	70	3	0	23	88.3	32.5	11.7
	LMA	121	470	12	360	4894	83.6	19.2	16.4
	TN	242	1979	24	76	516	69.7	24.1	30.3
	WA	0	58	523	75	6	79.0	21.3	21.0
	MCA-MO	0	33	103	725	618	49.1	41.3	50.9
	Overall Accuracy = 75.9 %								

Table 7. Error matrix, percent correct, commission error, and omission error for ET2CLAS1 (transformed, merged signature).

	Interpreted Soil Classes							
		Number of Pixels						
		DEC	LMA	TN	WA	MCA-MO	Prt. Ct. (%)	C.E. (%)	O.E (%)
Actual SCS Soil Classes	DEC	89	0	14	773	4	87.9	34.4	12.1
	LMA	970	16	3862	201	772	66.4	14.5	33.6
	TN	2198	79	225	205	186	76.0	34.8	24.0
	WA	45	481	0	0	137	72.4	22.5	27.6
	MCA-MO	71	45	382	0	929	65.0	24.6	35.0
	Overall Accuracy = 70.5 %								

Table 8. Soil Type Classification Comparison

Number of Acres and Percent Area by Soil Types for Each Classification Test

Soil Series	SCS Soil Series Map	ET2MAX Non-Transformed Singular Signature		ET2 MAX6 Non-Transformed Merged Signature		ET2MAXSEP Transformed Singular Signature		ET2CLAS1 Transformed Merged Signature	
	Acres	Acres	%Diff	Acres	%Diff	Acres	%Diff	Acres*	%Diff
DEC	189	205	+8.5	294	+55.6	248	+31.2	255	+34.9
LMA	1301	1043	-19.8	1293	-0.6	1341	+3.0	1013	-22.1
TN	632	436	-31.0	322	-49.1	585	-7.4	744	+17.7
WA	147	118	-19.7	121	-17.7	159	+8.2	140	-4.8
MO+MCA	329	796	+141.9	568	+72.6	265	-19.5	446	+35.6
Weighted Mean Percentage Area Incorrect		28.7		26.5		8.6		22.7	
Weighted Mean Percentage Area Correct		71.3		73.5		91.4		77.3	

* Acreage normalized for a total of 2598 acres

APPENDIX B

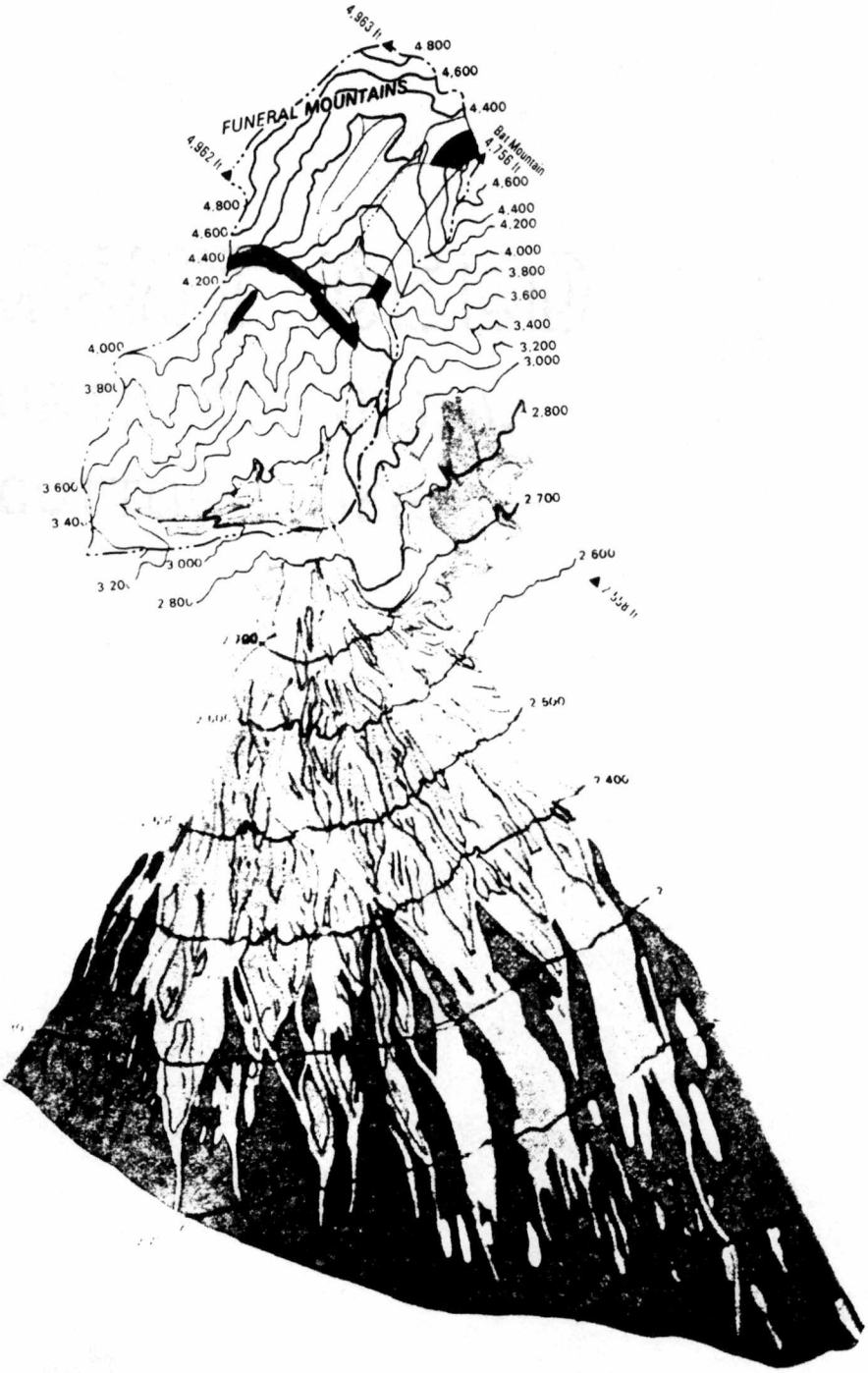


Figure 1. Illustration of an alluvial fan (Encyclopedia Britannica, 1978).

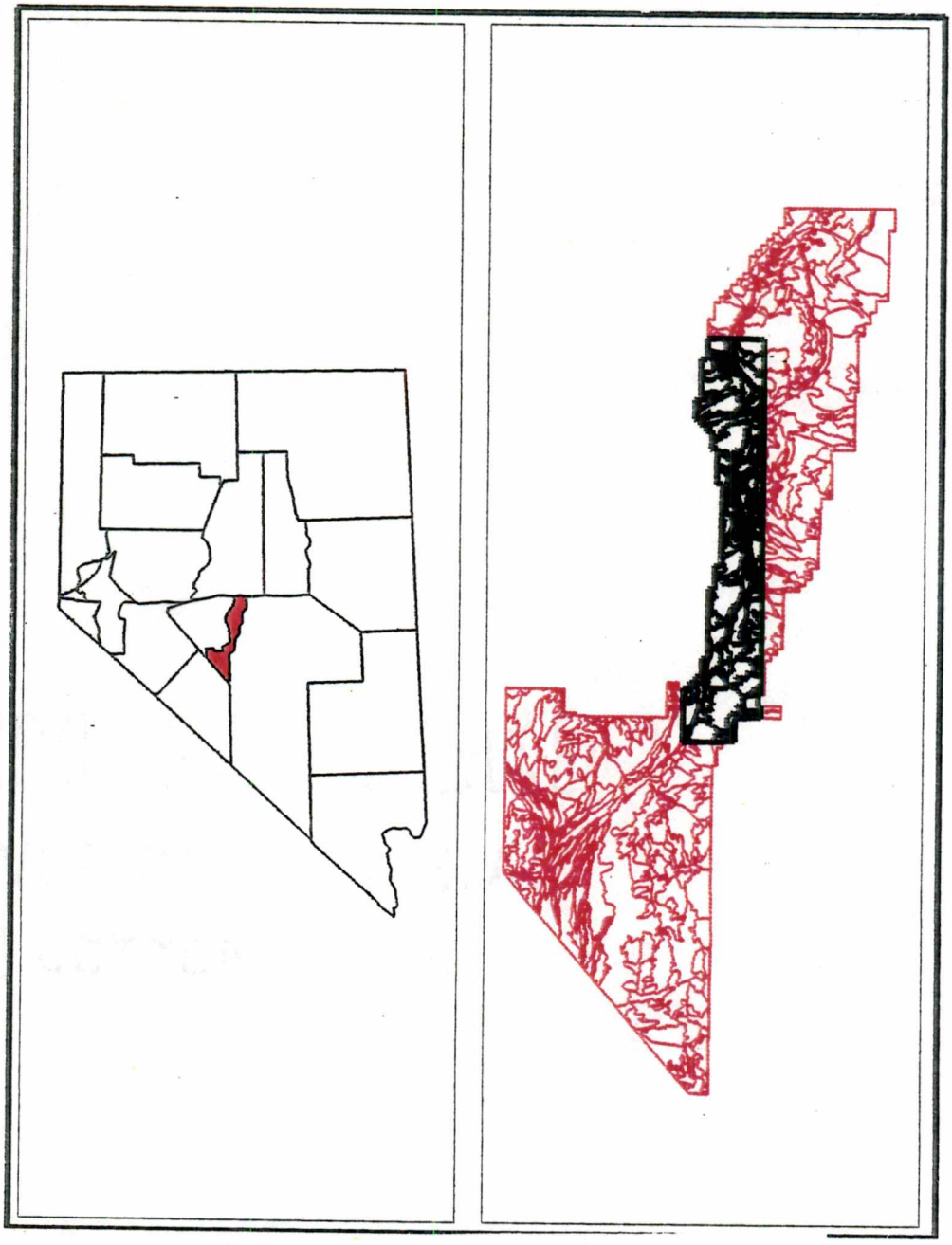


Figure 2. Areal perspective of Study Area coinciding with soil survey.

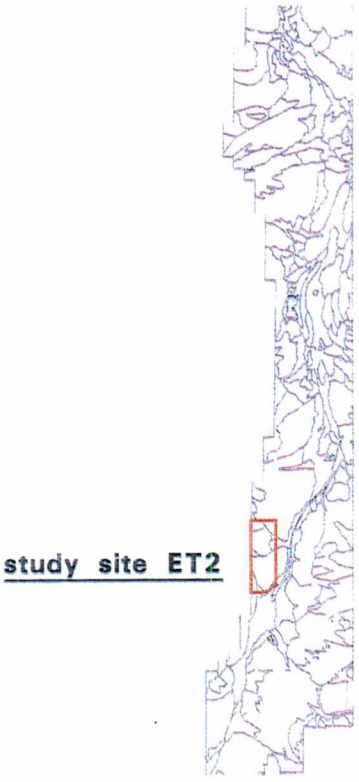


Figure 3. Location of Study Site ET2.

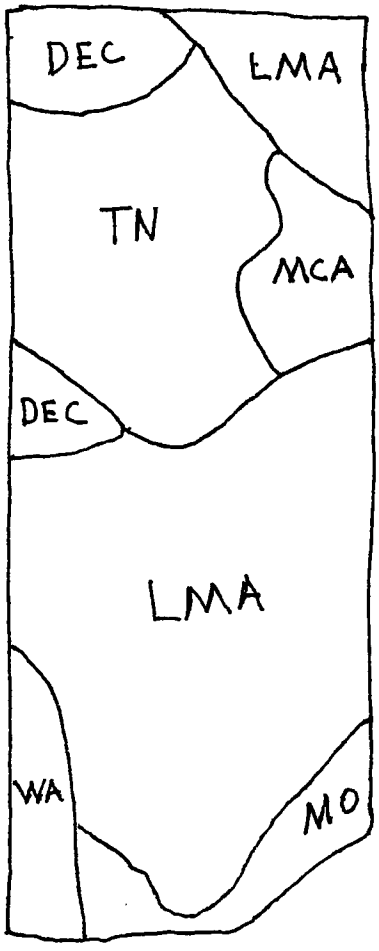


Figure 4. Soil Conservation Survey soil areas.

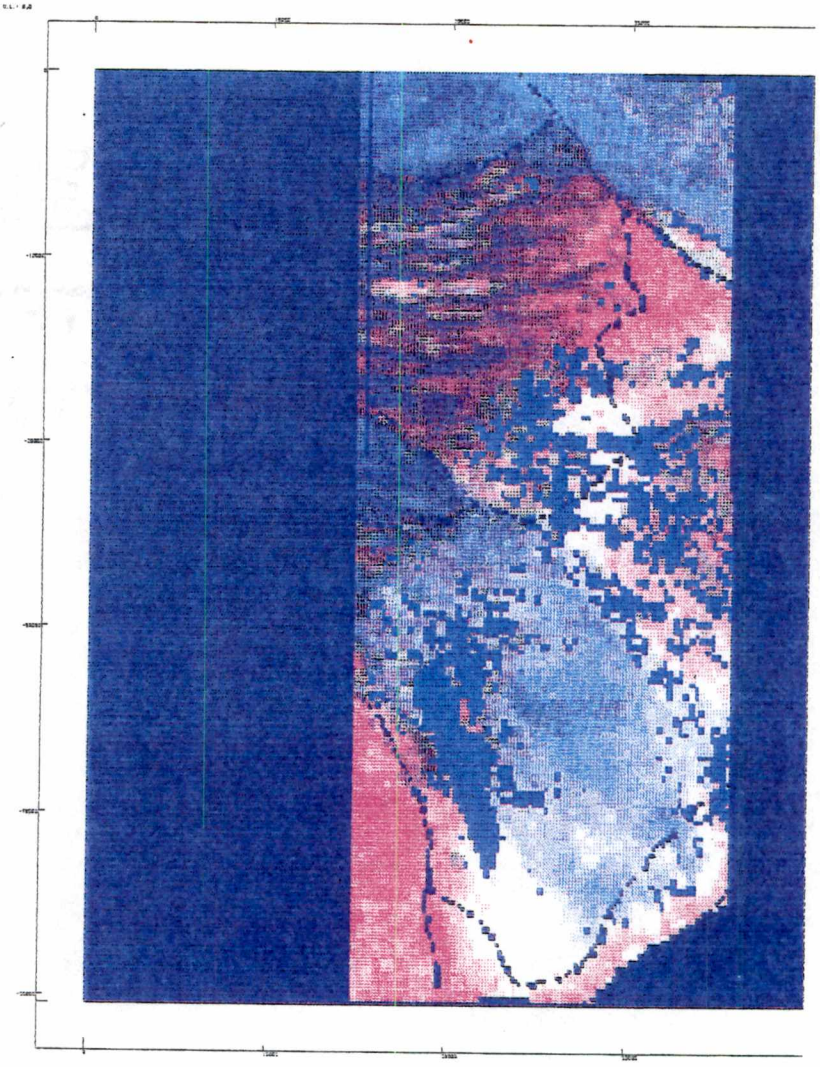


Figure 5. Discarded classification distribution of a sample signature.

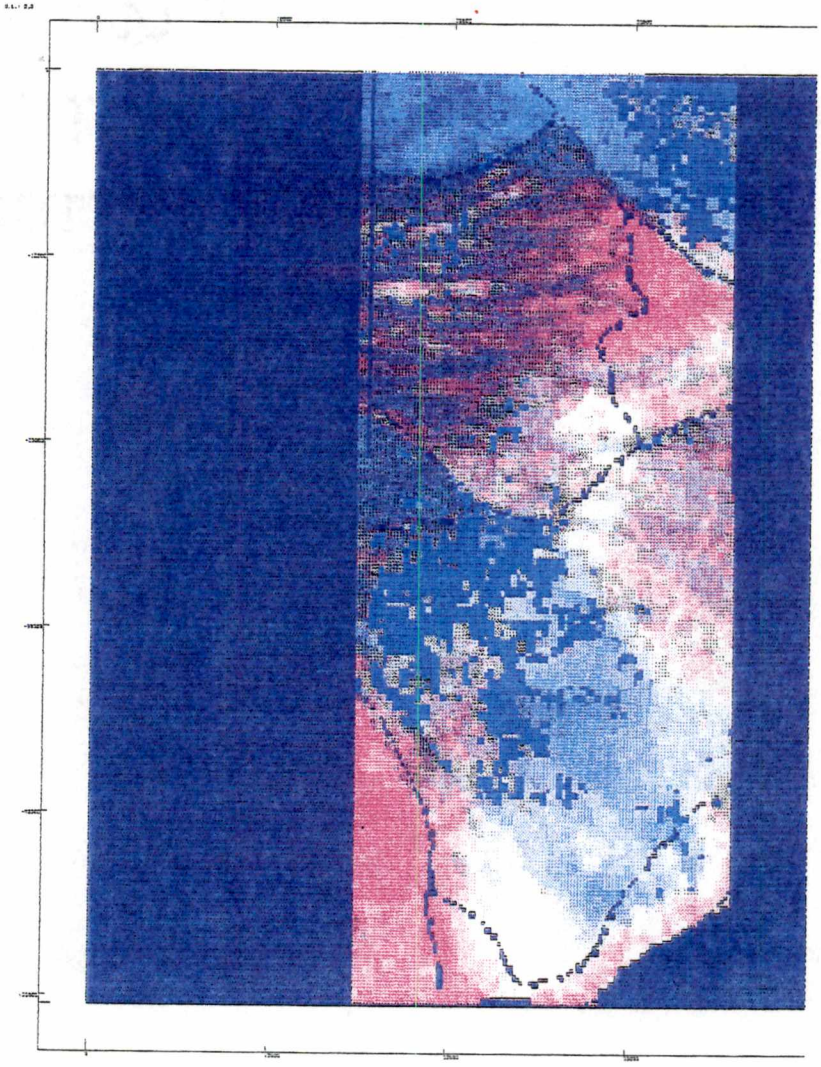
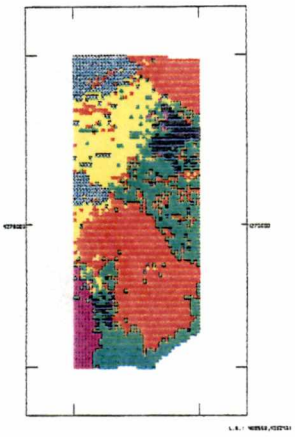
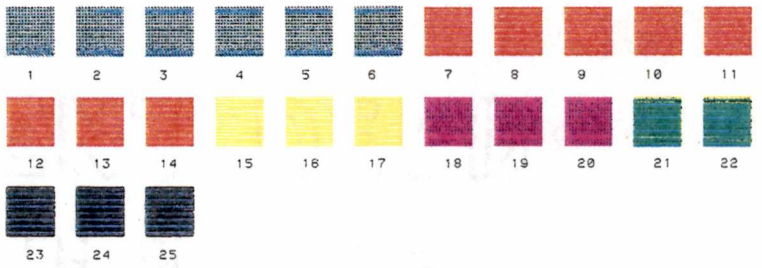


Figure 6. Acceptable classification distribution of a sample signature.





The variable name is :

VALUE	CLASS NAME	NO. OF POINTS	%	NO. OF ACRES
1	DEC22SIG	126.	1.08	28.019
2	DEC42SIG	46.	0.39	10.229
3	DEC62SIG	141.	1.21	31.354
4	DEC72SIG	151.	1.29	33.578
5	DEC162SIG	254.	2.17	56.482
6	DEC182SIG	205.	1.75	45.586
7	LMA12SIG	296.	2.53	65.822
8	LMA32SIG	449.	3.84	99.845
9	LMA62SIG	233.	1.99	51.813
10	LMA152SIG	731.	6.26	162.553
11	LMA172SIG	1165.	9.97	259.063
12	LMA202SIG	165.	1.41	36.691
13	LMA212SIG	127.	1.09	28.241
14	LMA242SIG	1525.	13.05	339.116
15	TN92SIG	232.	1.99	51.590
16	TN262SIG	505.	4.32	112.298
17	TN282SIG	1222.	10.46	271.738
18	WA72SIG	136.	1.16	30.243
19	WA142SIG	159.	1.35	35.135
20	WA152SIG	238.	2.04	52.924
21	MO72SIG	136.	1.16	30.243
22	MO122SIG	2832.	24.24	629.756
23	MCA72SIG	209.	1.79	46.476
24	MCA122SIG	356.	3.05	79.164
25	MCA142SIG	46.	0.39	10.229
TOTALS:		11684.		2598.187

Figure 7. ET2MAX (non-transformed signature input, singular)

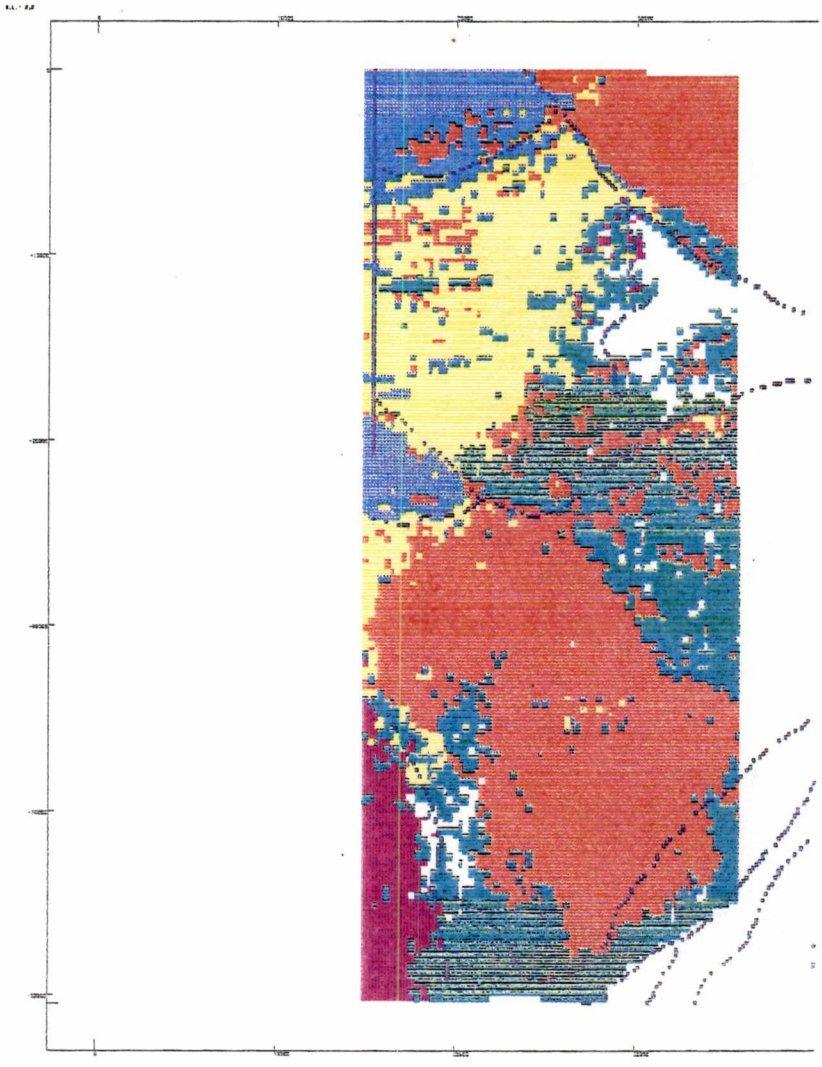
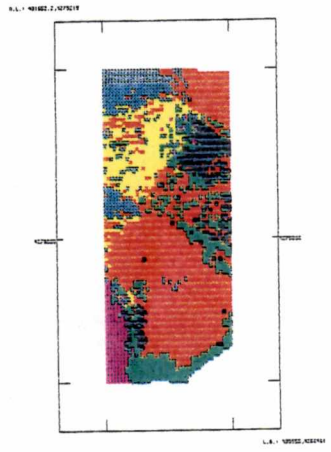
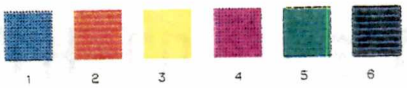


Figure 8. ET2MAX large scale view with superimposed soil areas defined.





The variable name is :

VALUE	CLASS NAME	NO. OF POINTS	%	NO. OF ACRES
1	DEC	1320.	11.30	293.530
2	LMA	5817.	49.79	1293.534
3	TN	1447.	12.38	321.771
4	WA	544.	4.66	120.970
5	MO	1661.	14.22	369.359
6	MCA	895.	7.66	199.022
TOTALS:		11684.		2598.187

Figure 9. ET2MAX6 (non-transformed, signature input merged).

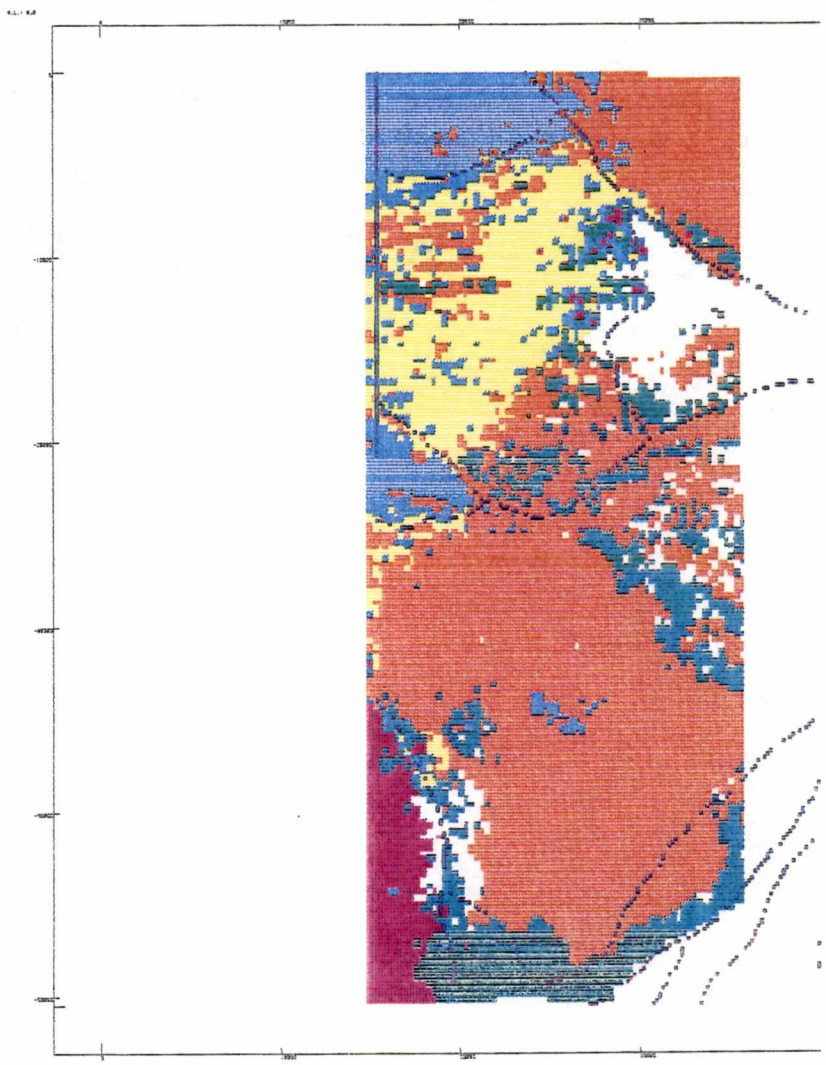
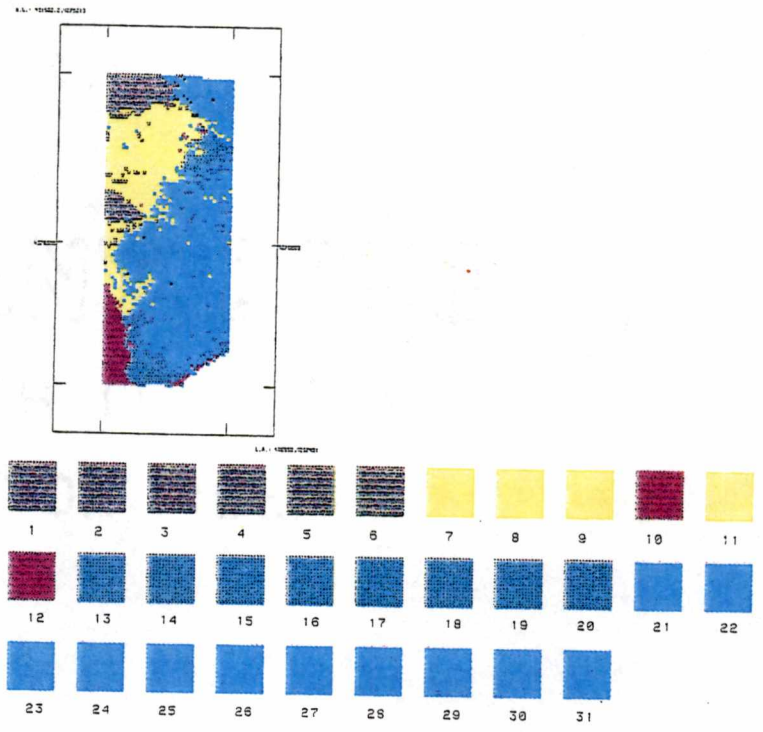


Figure 10. ET2MAX6 large scale view with superimposed soil areas defined.



The variable name is :

VALUE	CLASS NAME	NO. OF POINTS	%	NO. OF ACRES
1	DEC12SIG	66.	0.56	14.677
2	DEC52SIG	423.	3.62	94.063
3	DEC62SIG	114.	0.98	25.350
4	DEC72SIG	257.	2.20	57.149
5	DEC32SIG	238.	2.04	52.924
6	DEC22SIG	19.	0.16	4.225
7	TN12SIG	588.	5.03	130.754
8	TN22SIG	139.	1.19	30.910
9	TN42SIG	821.	7.03	182.567
10	WA12SIG	347.	2.97	77.163
11	TN32SIG	1081.	9.25	240.383
12	WA22SIG	370.	3.17	82.277
13	MCA-M012	143.	1.22	31.799
14	MCA-M022	210.	1.80	46.698
15	MCA-M032	289.	2.47	64.265
16	MCA-M042	57.	0.49	12.675
17	MCA-M052	46.	0.39	10.229
18	MCA-M062	35.	0.30	7.783
19	MCA-M072	169.	1.45	37.581
20	MCA-M082	240.	2.05	53.369
21	LMA12SIG	273.	2.34	60.707
22	LMA32SIG	162.	1.39	36.024
23	LMA22SIG	534.	4.57	118.746
24	LMA42SIG	83.	0.71	18.457
25	LMA52SIG	77.	0.66	17.123
26	LMA62SIG	3091.	26.45	687.350
27	LMA72SIG	191.	1.63	42.473
28	LMA92SIG	347.	2.97	77.163
29	LMA102SIG	399.	2.64	88.713
30	LMA112SIG	650.	5.56	144.541
31	LMA82SIG	315.	2.70	70.047
TOTALS:		11684.		2598.187

Figure 11. ET2MAXSEP (transformed signature input singular).

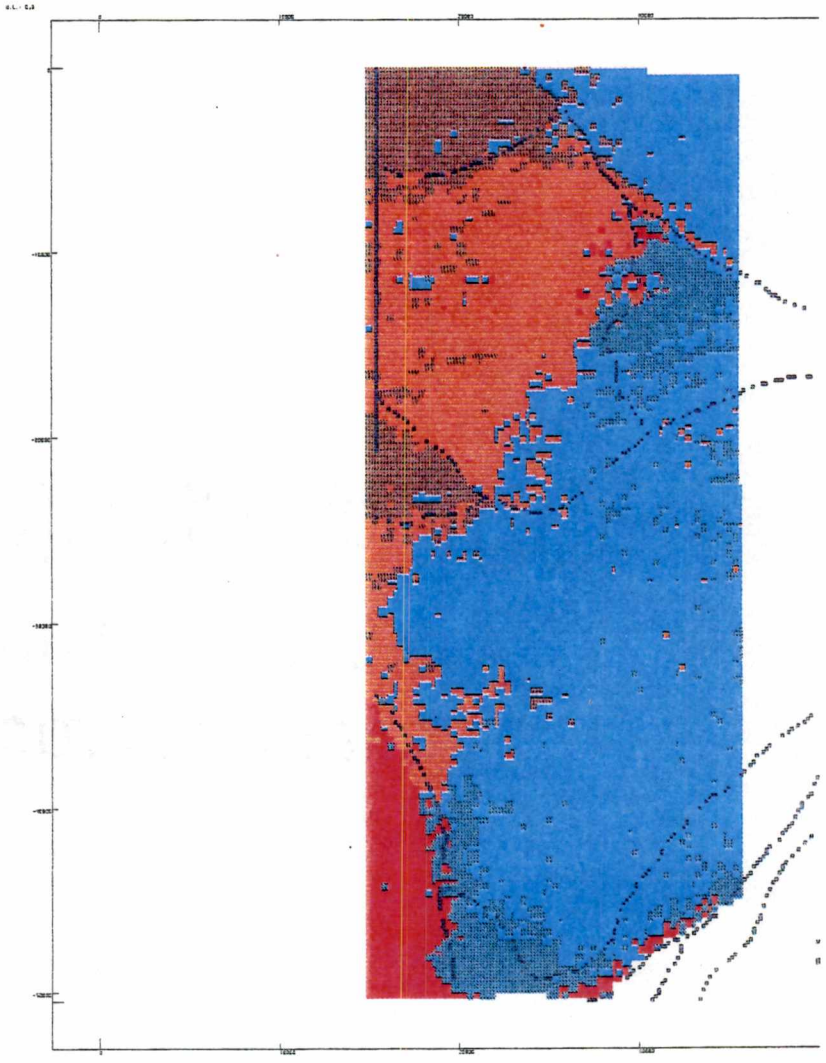
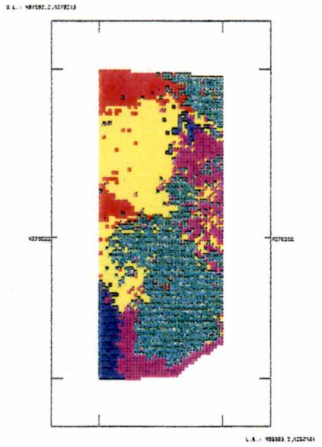


Figure 12. ET2MAXSEP large scale view with superimposed soil areas defined.





The variable name is : STUDY AREA #2

VALUE	CLASS NAME	NO. OF POINTS	%	NO. OF ACRES
1	Tomel-Laxal association (TN)	3227.	28.63	717.593
2	Wardenot sandy loam (WA)	608.	5.39	135.202
3	Laxel sandy loam (LMA)	4392.	38.97	976.655
4	Deerlodge stony loam (DEC)	1109.	9.84	246.610
5	Mazuma-Monte Cristo (MCA-MO)	1935.	17.17	430.289
TOTALS:		11271.		2506.348

Figure 13. ET2CLAS1 (transformed, signature input merged).

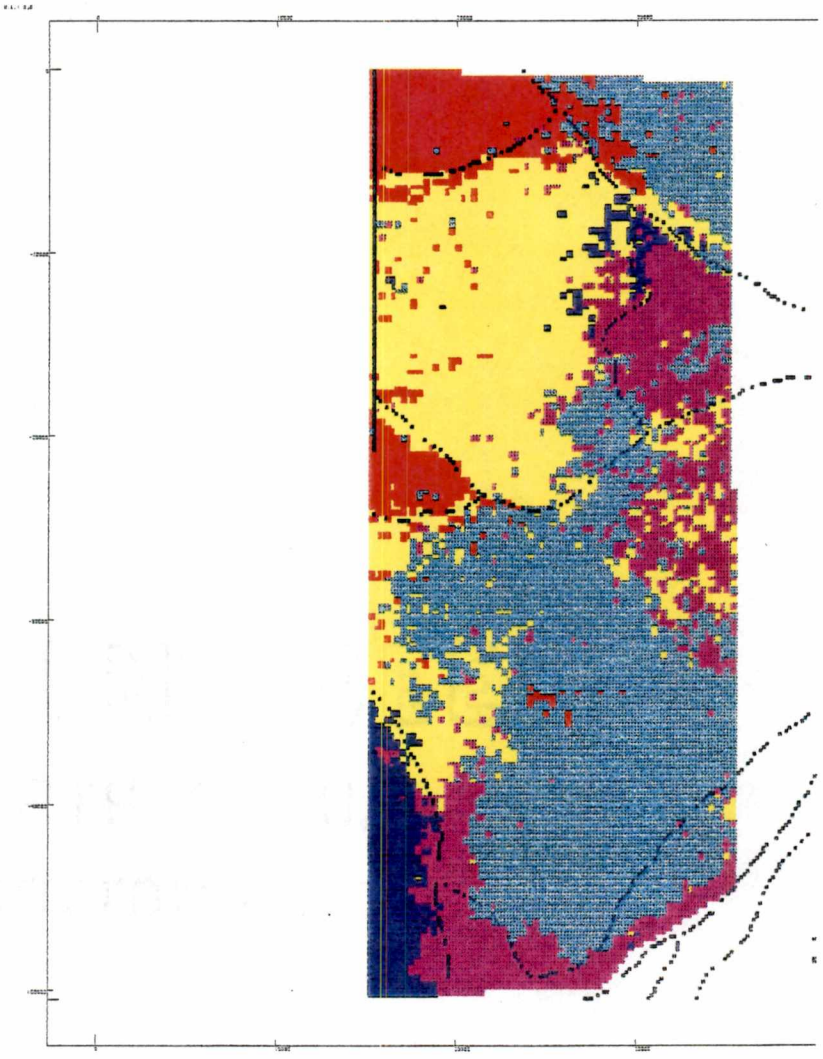


Figure 14. ET2CLAS1 large scale view with superimposed soil areas.

APPENDIX C

ALARM is a command in the ERDAS module SIGMAN which displays or highlights pixels classified by signature data.

ARC is the main programming environment of ARC/INFO. It contains commands which activate each of the other subsystems (ie., ARCPLOT, ARCEDIT, INFO, etc.,...).

ARCEDIT serves as the graphics and database editor for ARC/INFO. The subsystem combines all the facilities for digitizing map coverages with a comprehensive set of editing commands.

ARC/INFO is a geographical information system used to automate manipulate, analyze, and display geographic data in digital form.

ARC/INFO LIVE-LINK interactively allows the user to superimpose ARC/INFO vector coverages over corresponding image data.

ARC MACRO LANGUAGE (AML) allows the user to group commands into a string of executions.

ARCPLOT is an interactive cartographic and mapping subsystem of ARC/INFO. The module enables the user to perform an extensive array of map graphics.

BSTATS is a module in ERDAS that allows the user to access any statistics associated with image or geographical data.

COORDN is a module in Erdas that creates a transformation matrix of the 1st through 10th order, by the input of ground control points (GCPs). Once stored, the transformation matrix is used as data input into the rectification module, RECTIFY.

CUTTER in the Erdas module user to extract polygonal areas from an image.

DIGSCRN in ERDAS allows the digitzation of vector data superimposed over an image.

INFO is a complete relational database manager for tabular (attribute) data associated with geographic map coverages. ARC/INFO keeps account of and updates map feature attribute tables stored as INFO data files. The subsystem allows ARC/INFO to maintain the relationship between map features and corresponding tabular records.

LDDATA (load data) in the Erdas module allows the user to load image data from a reel to reel magnetic tape or from a disk file.
MAXCLAS module permits the execution in ERDAS of the maximum likeihood classification algorithm.

READ allows the user in Erdas to display an image stored on a disk, to a display device.

RECTIFY is an Erdas module which transforms coordinates of the pixels in the input image file (GCPs) to correspond to the chosen map coordinate system in the output file.

SIGEXT permits a user in ERDAS to compute the signature statistics from training samples and create accompanying signature files.

SIGMAN in ERDAS, allows the user to view (using the command ALARM) and merge signature statistics.

APPENDIX D

Bilinear Interpolation computes a value for each output pixel based on the weighted average of four input pixels of closest proximity. The weight or influence of one input pixel value is determined by proximity; the nearer pixels are given greater influence in the output values than the more distant pixels.

Cubic Convolution resampling assigns values to output pixels in a manner similar to bilinear interpolation with the exception that each estimated value in the output matrix is found by assessing the values of a neighborhood of 16 pixels in the reference grid.

Least Squares Regression is a statistical procedure for finding the "best fitting" line for a set of points that minimizes the sum of the squares of derivations of the observed data (ground control data) from those predicted (reference coordinate system).

Nearest Neighbor resampling assigns the values of each geometrically incorrect pixel to the closest geometrically correct pixel. Because method does not alter pixel brightness valves during resampling, subtle changes in brightness values in the image are maintained.

Transformation Matrix is an array of values that establishes the geometry of the output image and its relationship to the input image. The matrix consists of coefficients used in polynomial equations to convert coordinates. Based on the degree of distortion in the imagery, the number of GCP's (ground control points) and their locations relative to each other, higher order polynomials may be required to express the need transformation. The order of a polynomial is simply the highest exponent used in the polynomial expression. Second order or higher transformations are called nonlinear transformations and are used to correct nonlinear distortions. This process is also known as rubber sheeting.

Root Mean Square Error is determined in the computation of the coefficients for the transformation matrix, using the least square regression method, the objective is to find the polynomial equations which produce the least amount of error when used to transform GCP's into georeferenced coordinates. As show in figure , the distance between the GCP georeferenced coordinates and the polynomial curve is called the Root Mean Square (RMS) error. If the sum of the RMS errors for all ground control points are greater than the RMS tolerance defined by the user, the algorithm will delete ground control points that have the largest RMS error. The program, COORDN, then recomputes the coefficients using the

remaining GCP's. If too few points remain to calculate a least square regression or if the error tolerance remains high the program will automatically discontinue.

VITA

Dennis Lee Campbell was born in Memphis, Tennessee, the son of Booker T. and Violet Campbell. He attended elementary and secondary schools in the Memphis City Schools District and graduated from Melrose High School. He matriculated at Washington University at St.Louis where he received a Bachelor's degree in Earth and Planetary Sciences. Continuing his education, he accepted a graduate research assistantship at the University of Tennessee Space Institute and pursued a Master's degree in Engineering Science.