ANALYSIS OF MAN-INDUCED AND NATURAL RESOURCES OF AN ARID REGION IN CALIFORNIA

GEORGE MAY AND MICHAEL CRAIG

RESEARCH DIVISION
STATISTICAL REPORTING SERVICE
UNITED STATES DEPARTMENT OF AGRICULTURE
WASHINGTON DC

JANUARY 1982

Analysis of Man-Induced and Natural Resources of an Arid Region in California

George May and Michael Craig

Research Division
Statistical Reporting Service
United States Department of Agriculture
Washington DC

January 1982

Analysis of Man-Induced and Natural Resources of an Arid Region in California. George May and Michael Craig, Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington D.C. 20250, SRS Staff Report, January 1982.

ABSTRACT

A study site in California was analyzed with respect to both its man-induced and natural resources. For man-induced resources, estimates were made for major crops and for irrigated cropland based on regressing LANDSAT digital information onto ground enumerated, probability sampled areas. A manual interpretation of LANDSAT images (both from raw and computer classified data) was made to determine the feasibility of mapping soils and land use in an arid climate for the region not under intensive irrigation. Results indicate that a general soils and land use map, over such an environment, can be produced from LANDSAT photointerpretation.

KEY WORDS:LANDSAT, photointerpretation, erosion, fertility, probability sampling, regression, clustering, classification, stratification

This paper was presented at the First Thematic Conference on Remote Sensing of Arid and Semi-Arid Lands, January 1982, held in Cairo, Egypt.

TABLE OF CONTENTS

I. INTRODUCTION

Serghini and Hance [1] discussed the utilization of an area sampling frame in Morocco to collect ground data for quantifying desertification processes. This paper will extend the application of the sampling technique by discussing its role in processing LANDSAT data for purposes of crop and natural resource inventories.

The study site for this analysis, the Imperial Valley, is an arid region located in the southeast corner of California just north of the boundary between the United States and Mexico. The climate in this region is characterized by extreme aridity and high summer temperatures. Average rainfall is approximately three inches per year with half of it falling in intense summer showers. Most of the central part of the Valley is heavily irrigated cropland.

Ground and LANDSAT digital data were collected, combined, and analyzed based on an area sampling frame of this arid environment. The area frame stratification used in California by the Statistical Reporting Service (SRS) is discussed. Planted area estimates for alfalfa, wheat, cotton, and sugar beets were obtained by directly expanding the ground data. LANDSAT digital data were computer analyzed and the resultant classification was combined, using regression techniques, with the ground data to obtain new crop area estimates. Unitemporal, multitemporal, and transformed multitemporal digital data were analyzed and estimates produced from these data sets were compared.

An important element in stratification work and utilization of an area frame is soils information. Many factors such as location of crops, crop intensity, crop yields, wind and water erosion, and salinization are linked to various properties of soils. During this study an attempt was made to extract soils information from interpreting LANDSAT images and analyzing the digital data. The application of LANDSAT derived soils data to area frame studies was demonstrated.

II. THE CALIFORNIA AREA SAMPLING FRAME

Area Sampling Frame Methodology in General

Basically, any sampling frame is a listing or specification of the units from which to sample. In area sampling frame methodology, the units to be sampled are areas of land. Area sampling frames are complete with respect to the target population and are easily updatable. The procedure is to divide the total land area of a country into small areas of defined size using natural boundaries. These small geographic areas must be located, listed , and measured. This listing, along with its size measurements, specifies the Area Sampling Frame (ASF). Aerial photography, topographic maps, other map type products, and LANDSAT satellite images are used as control data for breaking out these areas and for classifying them into groups with similar characteristics. A random sample of these small areas, referred to as the selected segments, are then visited by enumerators who interview the farmers and landowners operating land in the segment. Data obtained by these interviews are used to make statistical estimates about agricultural commodities. Data from these interviews may also be used as ground data for remote sensing analysis.

The samples selected are designed to minimize the 'variance', or error due to sampling, in the final estimates taking into account the costs of the samples. The process of classifying geographic areas into similar groups, as mentioned above, is called stratification. Units are listed by their similar groups, called strata, and each stratum is treated as a separate population. This permits the focusing of the resources for collection of data on those geographic areas where the gain in terms of improvement in the overall estimate will be the greatest.

In the United States, small agricultural areas have been used as sampling units as far back as the late 1940's. By the early 1960's, an area sampling frame had been developed for all states. For many years, SRS has been using sample survey data to estimate area of crops, yield, production, livestock numbers, prices, agricultural wage rates, farm numbers, and other related items in US agriculture. This information is mainly based on a probability sample of segments enumerated in late May and again in early December of each year. Area frames for each state, which provide the basis for this sample, are based on standardized land use definitions in each stratum. Strata definitions used refer to the percentage of land cultivated, the number of dwellings per unit area, land in open range or woodland, and land in non-agricultural uses. Another paper at this conference [1] addresses the use of this methodology to measure indicators of desertification in arid and semi-arid regions.

The California Area Sampling Frame

The area frame currently used in California was constructed in 1978 and early 1979 for use in the June Enumerative Survey (JES) conducted annually in late May. Appendix I gives a description of the stratification and sample allocation used for the entire state of California. Also included in Appendix I is a summary of the cost for construction of this area frame at the state level. It should be noted that the allocation of sample segments to strata was done to minimize the sampling error for crop and livestock estimation without regard to factors concerning desertification, such as soil types, presence of erosion, etc. The construction and use of this frame is covered in a paper by Fecso and Johnson [2]. This report will discuss only those strata pertinent to the area of interest, which is the Imperial Valley in southern California.

Three counties make up the Imperial Valley: San Diego, Riverside, and Imperial. Of these, most of the analysis for this remote sensing study was done in Imperial County. In the study region four strata contain most of the land area. Strata 13 and 19 consist of areas with 50 percent or more cultivation, with the difference between them being that stratum 19 contains vegetable crops mixed with other crops. Stratum 43 (called desert range) consists of barren areas with less than 15 percent cultivation and virtually no crops or livestock. Stratum 50 is made up of nonagricultural areas such as State and National Forests, wildlife refuges, and military reservations. Table 1 gives the numbers of population units and sample segments found in the study region of Imperial County. Note that the sample was drawn at the state level, so that the subset of segments found in a given county may not have segments selected in a given stratum (see 31 and 50 in Table 1). Representative (non-statistical) areas were picked in strata 43 and 50 for later use in manual interpretation and remote sensing analysis giving a total of sixty-six areas and segments for study. Figure 1 shows the location of the major strata boundaries for this area.

Table 1 - JES Sample Allocation in Imperial County Study Region

Strata	Segment Size (hectares)	Population Number	Sample Number
13	259.2	347	13
19	259.2	550	16
31	64.8	144	0
32	25.9	57	1
43	1036.8	448	3
50	518.4	398	0

The use of probability sampling in strata 13 and 19 leads to statistical estimates of areas planted to various crops based only on the data collected by the field enumerators. These estimates are found by calculating the average planted area of a specific crop in the sampled segments and multiplying this average by the total number of population units from which the sample was drawn. Properties of these estimates, called direct expansion estimates, are given in a paper by Hanuschak [3]. The next section will discuss the use of LANDSAT digital data to reduce the sampling error found in these estimates.

III. DIGITAL ANALYSIS OF LANDSAT FOR CROPS

Background and Data Sources

The Statistical Reporting Service uses LANDSAT digital data as the auxiliary variable in a regression estimator to improve the accuracy of ground gathered survey data for planted area estimates of major crops. The SRS operational program in the U.S. is covered in a paper by Kleweno and Miller [4]. Currently, this program produces timely estimates for three crops in selected states of the United States midwest: corn, soybeans, and winter wheat. One objective of the research presented here is an application of this approach to major crop and land use estimation in an arid region. Output products of this digital analysis include statistically based crops estimates and a land cover map showing both vegetation and soils of the Imperial Valley.

Major crops in the Imperial Valley area include alfalfa, durum wheat, winter wheat, upland cotton, sugar beets, and vegetables. These crops are mainly located in the central part of the Valley, bordered on the east, west, and south by irrigation canals and bordered on the north by the Salton Sea. The planting and harvesting seasons vary widely for these crops. Sugar beets and winter wheat are fall and winter planted in this region. Durum wheat is seeded in late winter and early spring, while cotton seeding was most active in April and early May of 1980. Alfalfa is a crop that may be green year round, depending on irrigation practices. Multiple harvests and use for seed make this crop extremely variable with respect to remote sensing analysis. Sugar beet harvest was very active in April and early May of 1980 and mostly complete in early June. Cotton planting in 1980 was complete by late May with very good stands in most areas, while squaring of cotton was not present until mid-June of 1980 because of cool weather. Wheat harvest was very active in late May in this region.

SRS segment data was enumerated in the sampled segments in late May. Enumerators drew field and operator boundaries on black and white, non-current photographs of the segment areas. Farmers with land inside the segment were interviewed and a questionnaire completed giving information on each field and on the entire farming operation. The questionnaires and corresponding photographs made up the majority of the ground 'truth' data set. Since no cultivated land was found in the non-agriculture strata segments (strata 43 and 50), the twenty-nine segments in strata 13 and 19 contained all the cropland information. Because the operational survey is concerned with cultivated land and pasture only, all other land inside a segment is labelled 'wasteland' by the enumerators.

More specific information was needed for land use and soil characteristics in non-agriculture strata. Areas were selected from general soil maps of Imperial and Riverside counties as ground 'truth' to represent soil types and various erosion characteristics in these strata. Thirty-seven areas were selected in the study area and added to the twenty-nine segments. The selection of these areas in non-agriculture strata is discussed in Section V.

Due to satellite malfunctions and cloud cover, only two dates of LANDSAT imagery were available by October of 1980 for this crop season in the Imperial Valley. Specific scene information is given in Table 2 for these acquisitions. The two dates , March 29 and May 31, cover the small grain, alfalfa , and sugar beets crops very well but may have been a little early in the season for upland cotton. Computer compatible tapes (CCT's) plus black and white image products were ordered for both scene dates. Cloud cover and MSS line start problems in these two scenes reduced the study area from the original three county area to Imperial County.

Table 2 - LANDSAT Digital Data - Path 42 Row 37

Scene Id#	Satellite	Data Quality	Cloud Cover	Date MM/DD/YY
30755-17331	LANDSAT-3	Poor	0%	3/29/80
21956-17372	LANDSAT-2	Fair	30%	5/31/80

Processing of Digital and Ground Gathered Data

All information concerning the study region had to be converted to computer accessible form in order to combine it for the analysis stages. This information includes ground truth questionaires, segment and field boundaries, strata boundaries, and LANDSAT MSS digital data. Strata, segment, and field boundaries were digitized to a latitude-longitude coordinate base and the questionnaires data edited and entered into computer disk storage. CCT's were reformatted in the Washington, D.C. office and mailed to the Bolt, Barenek, and Newman computer system in

Boston. The information in each CCT was then registered to a latitude-longitude base using a third-order polynomial transformation computed from corresponding points on maps and LANDSAT image products of the same scale. A multitemporal data set containing eight channels of information for each pixel was obtained by overlaying the two unitemporal dates.

This registration process allowed locating of segment and field information in the LANDSAT digital data. This resulted in a set of eight-channel pixels labelled by crop or land cover type for use in training a computer classifier. Crop types were clustered using the CLASSY algorithm into a varying number of categories depending on the spectral information found in the pixels. CLASSY is a maximum likelihood clustering procedure developed at NASA/JSC [5]. As a general rule, soil and land cover types were given one category each and not clustered. Data from this process were combined into computer form in 'statistics' files containing mean, variance, and covariance statistics for each category.

In addition to the multitemporal (eight-channel) procedure described above, two more approaches to defining statistics files were used. The first, called the Kauth approach, used the 'greenness' and 'brightness' channels of the Kauth-Thomas Transformation [14] to reduce the data from eight to four channels for each labelled pixel. The final approach was to analyze the four channel information from one of the unitemporal dates (March 29). Clustering was done in a similar manner for each of these approaches. Note that only 'pure' pixels (non-boundary pixels from fields with less than 20% wasteland and no double cropping) were used for training. Table 3 summarizes the training pixels and number of categories found for each cover type.

Table 3 - Digital Analysis Training Information

Cover type	Total Pixels	Training Pixels	--Number of Categories--- Multitemp	Kauth	Unitemp
Alfalfa	5825	1724	7	5	3
Upl. Cotton	3034	987	6	5	3
Durum Wheat	4189	2455	10	12	7
Winter Wheat	1350	270	1	4	2
Sugar Beets	1736	764	2	3	2
Sorghum	252	208	1	1	1
Pasture/Hay	1043	188	2	2	1
Vegetables	767	349	4	1	2
Non-crop	15818	14365	30	31	28
Total			63	64	49

Digital Classification and Estimation

The statistics obtained from the approaches described above were used to generate three sets of Gaussian maximum likelihood discriminant functions for pixel classification. Pixel classifications using these functions were performed at two levels, 'small scale' and 'large scale or full frame'. In the three small scale classifications, all pixels within segment boundaries (including both pure and mixed pixels) were classified into categories using the discriminant functions. Thus, each pixel labelled by ground enumeration also had a corresponding category or cover type as estimated by maximum likelihood classification. The large scale procedure consists of classification of the full frame of LANDSAT data covering the entire population from which sampled segments were drawn. A short discussion of these classifications for crop estimation follows.

Totals of enumerated and classified pixels were calculated at the segment level for each cover type. These totals were used to estimate regression coefficients for estimation. The statistical methodology used is covered by Hanuschak [3] and will not be presented here in detail. The best measure of performance of each classifier for SRS purposes is the reduction of sampling error obtained when the classified data is regressed onto the ground enumerated data. For stratum h, this reduction is shown by the coefficient of determination, or r-square, as seen in the following formula:

$$v(\hat{Y}_r) = \sum_h \frac{n_h - 1}{n_h - 2} (1 - R_h^2) v_h(\hat{Y}_{de})$$

where

R_h^2 = R-squared for stratum h,

$v(\hat{Y}_r)$ = regression estimate variance,

$v_h(\hat{Y}_{de})$ = direct expansion variance for stratum h,

and

n_h = number of sample segments in stratum h.

R-squares for the three classifiers are compared in Table 4 for the major crops in the Imperial County study region. As seen in the formula, as the r-square approaches one the sampling error of the regression estimate approaches zero. The largest r-square values for the three analysis approaches are given by the multitemporal approach (using all eight channels of information).

Table 4 - R-square Comparisons* by Stratum for Major Crops

Crop Type	Multitemp		Unitemp		Kauth	
	13	19	13	19	13	19
Alfalfa	.62	.92	.75	.83	.36	.90
Upland cotton	.83	.85	.54	.57	.88	.88
All Wheat	.91	.92	.27	.72	.68	.87
Sugar Beets	.92	.92	.70	.82	.87	.93
All Cropland	.90	.92	.80	.93	.88	.89

* Based on n=29 segments.

Applying the regression relationship obtained from small scale to large scale classification results will yield crop area estimates. Table 5 gives the multitemporal regression and direct expansion estimates for major crops in the study region, along with their corresponding sampling errors. The large scale, multitemporal classification was also used to make a land cover map for this area (described in more detail in Section V). As indicated in the table, the application of regression methodology significantly reduces sampling errors in the estimates.

Table 5 - Area Estimates of Major Crops

Crop Type	Stratum	---Direct---- Expansion		---LANDSAT--- Regression	
		Estimate (ha)	C V (%)	Estimate (ha)	C V (%)
Alfalfa	13	23,500	26.1	15,400	25.8
'	19	36,000	26.2	42,800	6.4
'	Sum	59,500	18.9	58,200	8.3
Upland Cotton	13	14,400	33.5	13,800	14.8
' '	19	14,600	29.5	13,900	12.5
' '	Sum	29,000	22.3	27,700	9.7
All Wheat	13	7,600	46.5	23,700	4.6
' '	19	47,600	18.3	35,600	7.0
' '	Sum	55,100	17.0	59,300	4.6
Sugar Beets	13	5,200	60.8	6,000	15.3
' '	19	10,700	41.1	6,400	20.8
' '	Sum	16,000	34.0	12,300	13.0
All Cropland	13	56,700	14.7	63,700	4.4
' '	19	113,700	10.8	105,000	3.5
' '	Sum	170,400	8.7	168,600	2.7

IV. MANUAL ANALYSIS OF LANDSAT FOR IDENTIFICATION OF SOILS

Area Frame Utilization

In Section II the use of an area sampling frame for crop estimation was discussed. Strata are developed by dividing a geographical area into similar types and intensity of agriculture. In a paper given at this symposium by Serghini and Hance [1] the use of area sampling frame methodology for desertification analysis is discussed. The thrust of using the frame was to quantify desertification processes. Soils information is a very important factor in quantifying crop production and desert encroachment. SRS does not operationally use soils data for building area sampling frames, but indirectly some soils information is included because the better agricultural soils have the higher intensity of cropland. From a desertification standpoint the erodability of soils and their ability to sustain plant growth are important elements.

Soils also affect computer classification results obtained from LANDSAT data. Several researchers [6,7,8] have shown how soil color, texture, and other parameters affect the spectral response received from crops and other vegetative land cover types. Just as a geographical area is stratified by the amount of land cultivated to reduce the sampling errors in crops estimation, perhaps soils data can be used to stratify LANDSAT products to reduce the error in both desertification estimation and classification of digital data. An improvement in the classification of LANDSAT data based on soils stratification could possibly lead to an improvement in the regression estimate. An SRS report by Cook [9] discussed the mixed results that were obtained in conducting research in this area. Significant improvements were obtained in the estimates for sunflowers and rangeland but no improvement was obtained for corn and wheat. Research is continuing in an effort to better quantify soil effects on LANDSAT crop classification and regression estimates.

Rationale for Photointerpretation Work

In the above section a case was built for the importance of soils in stratification work for agriculture and desertification analyses. In many parts of the world soils data is very limited and in some cases non-existent. As part of the Imperial County study an effort was made to develop a soils map for this arid environment. This county offered a timely advantage in that the detailed soils survey report produced by the United States

Department of Agriculture was to be released by spring of 1981. This soils map would be used for verification in determining the quality of the map made in this study.

The objective was to use photointerpretation techniques to delineate soil unit boundaries using a LANDSAT false color print and a geology map of comparable scale. Each soil delineation is described with respect to its origin, parent material, surface color, and surface texture. The rational for this part of the study was that LANDSAT is available over all the arid and semi-arid regions of the world. In most cases some type of geologic map or information would be available over these same areas. Food and Agriculture Organization (FAO) soil maps of the world are also obtainable but at a scale of 1:5,000,000 which limits their usefulness. The soils map produced from this study, even though not linked to a soil taxonomy or classification system, could have potential for use in stratification work for crop and desertification area frame development and also analysis of LANDSAT digital data. From a spectral or scene characterization standpoint this type of map could be useful because it should delineate spectrally homogeneous groups, even though the factors for causing this homogeneity can not be fully determined through photointerpretation.

A 1:250,000 scale, edge-enhanced false color print of the May 31,1980 LANDSAT image was the basis for the photointerpretation . Black and white prints of bands 5 and 7 were also used. A 1:250,000 scale geologic map containing Imperial County was purchased from the California Division of Mines and Geology [10].

Landforms and Geologic Setting

The western two-thirds of Imperial County is in a topographic and structural region called the Salton Trough. It is a landward extension of the depression filled by the Gulf of California but is separated from the Gulf by the broad fan shaped delta of the Colorado River. This trough contains sediments from the Colorado and also local alluvium from the Chocolate Mountains to the east, and California coastal range to the west. The Imperial Valley, which contains the irrigated cropland, lies in the center of this trough. The land surface of this valley slopes northwestward to the Salton Sea which is approximately 232 feet below sea level [11].

The eastern third of the county contains the Chocolate Mountains which are mainly composed of volcanic and metamorphic rock. Alluvial fans from these mountains form distinct patterns as they extend to the old Colorado River flood plain. The Chocolate Mountains reach an altitude of 2500 feet [12].

Development of Soils Map

The first step was to delineate and identify rock types and other geologic features. This was done by transferring data from the geologic map onto clear acetate overlaying the false color LANDSAT image (called FCI). The overlay is shown in Figure 2. In some cases the geology map was used as a guide but the actual delineation was done from photointerpreting the FCI. An example of this is the boundary of the western shoreline of prehistoric Lake Coahuila. Old beach ridges are not as obvious here as they are along the eastern shoreline, but slight tonal differences as seen from the LANDSAT FCI and black and white images did suggest the delineation of parts of the western shoreline.

The next step was to determine direction of slope and drainage patterns. The arrows on Figure 2 indicate slope direction. Topographic maps were available for Imperial County but were not used in this study because the soil interpretation process is aimed toward helping countries where soils and topographic maps do not exist.

After completing the above overlay the delineation of surface soil boundaries began. The main interpretation keys were color, tone, texture, and location/association on the landscape. Soil interpretations were not made for the intensive cropland areas of Imperial Valley because the checkerboard pattern of the lush vegetation made it impossible to obtain bare soils information. In irrigated arid and semi-arid lands where double and triple cropping are practiced it will be very difficult, if not impossible, to obtain imagery when fields are bare. In non-irrigated areas soils will be bare during part of the year or at least spectrally prominent within a scene , even if some vegetation does exist. For example, if the land sustains rangeland then an image can be obtained when the range species are dormant, or at least minimal, to reduce the spectral response from the vegetation.

Several of the soil boundary delineations were easily definable in the eastern half of the county. Alluvium and colluvium deposits derived from upslope material formed distinct lines. Soils boundary delineations within the Lake Coahuila and delta deposits of the Colorado were not as obvious. Associated with each delineation is a description or rational which describes the parent material, origin, surface color, and surface texture. Some delineations are also described with respect to soil erodability. These descriptions are given in Appendix III.

The USDA soil classification textural triangle [13] shown in Figure 3 was used to classify soil texture. The triangle was divided into four groups of soils having high sand, high silt, high clay contents or soils of near equal proportions. The following techniques were used to determine the texture of a soil delineation:(a)drainage pattern analysis, (b)erosion or dune development , (c)knowing the mineral composition of parent material and how these minerals weather to form soils, (d)texture and tonal differences, (e)vegetation patterns and, (f)association of a soil delineation with a specific landform or mode of deposition. Soil color was determined by interpreting and converting the colors seen on the FCI print to normal color.

Mapping Results

The soils map resulting from the photointerpretation is shown in Figure 4. The detailed soil survey map that was to be published earlier in the year and be used for verification has not been released. As a backup, the 'Report and General Soil Map of Imperial County', produced by the USDA and Imperial Irrigation District was used [11]. This soil association map is shown in Figure 5. Appendix IV gives characteristics, taken from the soil survey report, of each soil association. As can be seen on Figure 5, a large area in the northeast corner of Imperial county is a naval reservation and therefore not mapped.

It is very difficult to quantify the accuracy of the interpreted soil map. Figures 4 and 5 are at the same scale so a comparison of any point can be made by the reader. Nine reference points have been denoted on these two figures. For example, there is a high correlation between the maps for the delineation that contains point 3. The interpreted soil boundary is very close to the soil line shown in the soil survey map. The soil texture was also interpreted correctly (refer to the appropriate map symbols in Appendixes III and IV). This high correlation does not hold up in other parts of the map such as the area around point 1.

An attempt was made to quantify accuracy of the interpreted soil textures and soil colors. A grid was positioned onto the soil maps shown in figures 4 and 5 and 50 points were randomly selected. The stated soil texture was obtained from both maps for each point and a comparison was made based on the four divisions of the textural triangle. A 65 percent agreement was obtained between the interpreted texture and survey texture.

Quantifying soil color was more difficult because nine colors were interpreted from the FCI print , whereas only five colors were reported in the survey. A soil color matrix for the 50 points is given in Table 6. This matrix suggests some strong correlations between the interpreted and reported colors. For example, 75 percent of the points falling in the pale brown or light brownish gray soils were interpreted as various shades of gray.

The interpreted soils map shown in Figure 4 has several uses in area sampling frame studies. One application is the identification and estimation of potential cropland. All the soils lying within the Lake Coahuila boundary have similar

Table 6 - Comparison of Interpreted and Surveyed Soil Color for Fifty Random Points

Interpreted Color	Soil Survey Color: Light Brown	Pale Brown-Brown/Gray	Pink/Gray-Red/Brown	Light Brown-Light Gray	Pink
Red/Brown	1	0	0	0	0
Tan	0	2	1	0	0
Gray/Tan	1	2	1	1	0
Red/Tan	5	0	1	0	0
Gray	2	0	1	2	0
Tan/Gray	2	3	0	0	0
Red/Gray	9	2	3	1	1
Yellow/Gray	0	5	0	0	0
Green/Gray	0	2	0	0	2

properties and could support crops. The Imperial Valley is heavily cropped because of the irrigation water obtained from the Colorado River. If additional sources of water became available then more of the lacustrine soils could be put into agriculture use. Figure 6 is an example of this expansion. The top scene was extracted from a March 1975 LANDSAT image and the bottom scene from the May 1980 image. The cropland near point 3 has increased 500 percent over the five year period. This expansion is occurring on the lacustrine soils.

The soil map can also be used to group soil delineations into various textural and color classes for stratification. Delineations could also be described according to their erosion hazard. A rating system could be established to determine the degree of wind and water erosion, and then these ratings grouped accordingly. Section V describes an attempt to use erosion characteristics and inherent fertility descriptions for various soils in the non-agricultural lands in Imperial County for digital LANDSAT analysis and map production.

V. DIGITAL ANALYSIS AND MAPPING IN NON-AGRICULTURAL AREAS

As mentioned earlier in Section III, the normal SRS approach to digital analysis is to lump all non-cropland training areas together (usually called 'wasteland' in the ground truth) and cluster the data. No attempt is made to identify the resulting categories with any land cover or soil grouping, with the exception of water which separates readily. In this analysis, training sites in the non-agricultural strata were selected to represent various soils prevalent in Imperial and Riverside Counties.

Training sites were chosen to reflect various ratings of erosion hazard and inherent fertility as discussed in Section IV. These sites were delineated on soil and topographic maps based on the manual interpretation of the FCI and the corresponding soil association found in the general soils map for the same region. After delineation, these areas were digitized and treated as area frame segments similar to those in the intensive agriculture strata.

Training statistics were generated for high and low rated erosion soils. These were also labelled as to their potential for crop production based on the general soils map description. For the multitemporal (eight channel) data set, the twenty-nine signatures for non-cropland areas included statistics for: a)urban, b)deep water, c)feedlots, d)sand dunes, e)salt deposits around the Salton Sea, and f)various specific soil associations. Another signature was developed for the bad MSS scanner lines in the May 30 image.

For the twenty-four soil specific signatures, ratings by erosion and fertility characteristics led to the following four groupings. The first group, called <u>Soils-1</u>, is characterized as erodable soils with potential for intensive crop use and had three signatures. The <u>Soils-2</u> group, also with three signatures, is described as areas with moderate potential for cultivation and prevalent erosion hazard. The last group with erosion hazard, <u>Soils-3</u>, had little or no potential for crop use and includes badlands, rubble, and rockland areas. This group was represented by ten signatures. The final group of signatures, <u>Soils-4</u>, shows little potential for crop use and only slight chance of erosion.

Cropland and non-cropland signatures were combined into one file and used to classify the full frame multitemporal data set pixel by pixel. The resultant classification was used for both the regression estimation for major crops as described in Section III and to produce a map-type picture product called a Dicomed. The Dicomed print, shown in Figure 7, is generated by assigning a specific color to each category found in the classification. Appendix II gives color groupings used for this print, along with corresponding classification accuracies based on the set of training pixels. In general, color assignments were: green/white for crops, navy for water, red/dark orange for Soils-1, purple for Soils-2, brown/orange/aqua for Soils-3, and yellow/grey for Soils-4. The resultant color patterns seem to correspond well with the photointerpreted and surveyed soils map from Section IV.

No area estimates for the soil categories or groupings were attempted using this classification. However, had the training segments been randomly selected as was done in the cropland strata, both direct expansion and regression estimates could be made using the area frame statistical methodology described in this paper and in the presentation by Hance.

VI. SUMMARY

An inventory of the crop and natural resources was conducted for an arid region in southern California. Ground data enumerated from randomly selected sample segments were combined with LANDSAT data obtained from two dates. LANDSAT photographic and digital data were analyzed by manual interpretation and automated data processing techniques. Sample selection, ground data collection, area estimation, and analyst procedures were based on area sampling frame methodology.

The following three data sets were used in digital analyses: multitemporal (eight channel), multitemporal Kauth-Thomas transformed (four channel), and unitemporal (four channel). Training statistics were developed for alfalfa, cotton, durum wheat, winter wheat, sugar beets, sorghum, vegetables, pasture, various soils, and other non-agriculture land covers. A classification map using these statistics was produced for the study region.

The sampled ground data were expanded and planted area estimates obtained for alfalfa, cotton, wheat, sugar beets, and all cropland. The classified LANDSAT data from each of the three data sets were regressed onto the ground data and three new estimates derived for each of the above categories. The largest r-square values of the regression estimates were obtained using the multitemporal (eight channel) data set. A comparison of the multitemporal (eight channel) regression and direct expansion estimates indicated that the regression methodology significantly reduces the sampling errors of the estimates.

The value of soils data in area sampling frame studies was demonstrated. Soils information was obtained by interpreting false color plus black and white LANDSAT prints in association with a geology map of the same scale. Soils were delineated and described with respect to origin, parent material, soil color, soil texture, and erodability. The interpreted map was compared to a soil association map obtained by survey techniques. It is difficult to quantify the results, but the LANDSAT interpreted soils map can give usable data in areas where soils information is limited or non-existent.

Any future work in interpreting soils from LANDSAT should consider using the World Soil Maps that have an original scale of 1:1,000,000. These maps were produced by the U.S. Soil Conservation Service and were derived using various data inputs such as geology, climate, topography, and, when available, soil maps. These maps were produced for the whole world except the United States and Australia, and therefore were not available for this study. The International Soil Science (ISS) Museum is currently collecting and filing the maps. Inquiries for obtaining these soils data should be directed to the ISS Museum in Wageningen, Netherlands.

VII. References

1) Hance, W. G. and H. Serghini. 1981. 'Quantifying Agricultural Indicators of Desert Encroachment'. Soil Conservation Service, U. S. Department of Agriculture. Washington, D. C.

2) Fecso, R. S. and V. B. Johnson. 1981. 'The New California Area Frame: A Statistical Study'. Document SRS-22. Statistical Reporting Service, U. S. Department of Agriculture. Washington, D. C.

3) Hanuschak, G. A. 1979. 'Precision of Crop Area Estimates'. Proceedings of the Thirteenth International Symposium of Remote Sensing of Environment. ERIM. Ann Arbor, Michigan.

4) Kleweno, D. D. and C. E. Miller. 1981. '1980 AgRISTARS DC/LC Project Summary - Crop Area Estimates for Kansas and Iowa'. Publication AGESS-810414. Statistical Reporting Service, U. S. Department of Agriculture. Washington, D. C.

5) Lennington, R. K. and M. E. Rassbach. 1979. 'CLASSY - An Adaptive Maximum Likelihood Clustering Algorithm'. Volume II. Proceedings of Technical Sessions 1978 LACIE Symposium. JSC-16015. Houston, Texas.

6) Stoner, E. R. and M. F. Baumgardner. 1980. 'Physiochemical, Site, and Bidirectional Reflectance Factor Characteristics of Uniformly Moist Soils'. LARS Technical Report 111679. Laboratory for Applications of Remote Sensing. Purdue University. West Lafayette, Indiana.

7) Westin, F. C. and G. D. Lemme. 1978. 'LANDSAT Spectral Signatures: Studies With Soil Associations and Vegetation'. Photogrammetric Engineering and Remote Sensing. 44:315-325.

8) Imhoff, M. L. and G. W. Petersen. 1980. 'The Role of LANDSAT Data Products in Soil Surveys'. Interim Report, Institute for Research on Land and Water Resources. The Pennsylvania State University. University Park, Pennsylvania.

9) Cook, P. W. 1981. 'A Test of the Effects of Soil Stratification on LANDSAT Crop Classification Accuracy'. Statistical Reporting Service, U. S. Department of Agriculture. Washington, D. C., in progress.

10) Unknown. 1967. 'Geologic Map of California. Salton Sea, El Centro, San Diego Sheets'. California Division of Mines and Geology. Sacramento, California.

11) Soil Conservation Service-U. S. Department of Agriculture. 1967. 'Report and General Soil Map Imperial County, California'. Imperial Irrigation District. El Centro, California.

12) Loeltz, O. J., B. Irelan, J. H. Robison, and F. H. Olmsted. 1975. 'Geohydrologic Reconnaissance of the Imperial Valley, California'. Geological Survey Professional Paper 486-K. U. S. Government Printing Office, Washington, D. C.

13) Soil Survey Staff. 1960. 'Soil Classification, A Comprehensive System, 7th Approximation'. Soil Conservation Service, U. S. Department of Agriculture. Washington, D. C.

14) Kauth, R. J., P. F. Lambeck, W. Richardson, G. S. Thomas, and A. P. Pentland. 1978. 'Feature Extraction Applied to Agricultural Crops as Seen by LANDSAT'. Proceedings of LACIE Symposium. Johnson Space Center. Houston, Texas. II:705-721.

15) Gardner, W. 1979. 'Time and Costs for the New California Area Frame and the New JES Sample'. Official Memo to N. D. Beller dated July 6, 1979. Economics, Statistics, and Cooperatives Service, U. S. Department of Agriculture. Washington, D. C.

APPENDIX I. The California Area Sampling Frame

Early experience with area frame surveys indicated that stratification according to land use was essential for reduction of sampling errors. The sample can be reallocated between strata not only to reduce sampling error but also to alter survey costs. The area frames used by the SRS in the United States are usually stratified into six general land use strata based on the amount of land cultivated. In general, the land use strata are intensive agriculture, extensive agriculture, cities and towns, range, nonagriculture, and water.

The current California frame [2] was not limited to these six land use strata. Information from other sources was used to develop crop-specific strata within the general land use framework. A summary of the strata definitions for the entire state's frame are:

Strata	Percent Cultivated	Description
13	50 - 100	General crops, <10% fruits or vegetables
17	50 - 100	Mostly fruits, tree nuts, or grapes
19	50 - 100	Vegetables mixed with general crops
20	15 - 50	Extensive cropland and hay
31	0 - 15	Residential mixed with agriculture
32	0 - 15	Heavily residential / commercial
41	0 - 15	Private range
43	0 - 15	Desert range
44	0 - 15	Public grazing lands in allotments
45	0 - 15	Public land not in grazing
50	0 - 15	Nonagricultural land
62	0 - 15	Known water greater than 1 sq. mile
07		Large farm operations

The first step in constructing the California frame was classifying all land in the state into one of the defined strata. Transparent copies of county maps were made at the same scale as color LANDSAT image products, and overlaid on them. Overlays showing locations and crop contents of segments sampled during prior surveys were then located and oriented to the scene. Other information gathered for these areas included county estimates of various crops, crop calendars, field sizes, and the date of the LANDSAT images.

Using all available information about an area, the LANDSAT data were photointerpreted into the land use strata defined for California. Strata boundaries were then delineated on a set of county maps, with a photo index(PI) mosaic of aerial photography used to identify final permanent boundaries. Within each strata, areas were further subdivided into units of similar size (called count units). Count unit sizes were measured by digitizing the area on the map base. Each count unit was assigned a number of segments based on dividing the count unit size by the target segment size for that strata. Count units in each stratum were arranged on paper by county to group count units which were agriculturally similar. This leads to a geographic stratification within each stratum, called paper stratification. Count units were broken down further into their assigned number of segments only when necessary as defined by the sample allocation.

Within a land use stratum, simple random samples of equal size were drawn in each paper stratum. These samples were then assigned to replicates to give more flexibility for sample allocation and design changes. All strata were constructed and sampled in this manner with the exception of stratum 44. In this stratum, sample units were based on boundaries of grazing allotments and sampled probability proportional to size. Another stratum (stratum 07), consisted of eleven very large farm operations which were completely enumerated and deleted from the area samples where found.

Current population sizes and sample allocation for the California area frame are shown below:

Strata Label	Target Segment Size*	Population Number Segments	Sample Segments
07	n/a	11	11
13	1.00	6947	240
17	0.50	11228	240
19	1.00	3174	100
20	1.00	7752	120
31	0.25	14760	40
32	0.10	23150	10
41	4.00	10444	100
43	4.00	3984	20
44	n/a	1303	25
45	4.00	4003	8
50	2.00	7086	8
62	n/a	1706*	0
total			922

* square miles

The state of California covers approximately 100 million acres (40.5 million hectares), of which some 15 million acres (6.1 million hectares) is cultivated. For frame construction alone

[15], the cost incurred was $68,194. Selection of the sample for the annual JES survey and provision of field enumeration material was $35,046. Included in these figures was the following breakdown: $67,513 in salaries, $24,738 in materials and travel, and $10,989 in computer costs. The frame construction in California was a research effort for the application of LANDSAT imagery for stratification and the dollar figures quoted above reflect this. Subsequent construction in other states based on this technology has cost substantially less.

APPENDIX II. Dicomed Colors and Accuracies by Cover

Land Cover	Dicomed Colors	Pixels Correct* (%)	Commission Error* (%)
Alfalfa	Dark Green	77.64	4.62
Upland Cotton	White	86.16	26.17
All Wheat	Mint Green	97.79	5.42
Sugar Beets	Moss Green	70.81	15.20
Other Crops	Shades Light Green	-	-
Feedlot	Grey	85.15	10.88
Salt Deposits	Royal Blue	69.02	9.76
Urban	Dark Brown	75.82	15.98
Bad MSS Data	Black	-	-
Deep Water	Navy Blue	-	-
Soils-1	Reds, Dark Orange	80.96	25.62
Soils-2	Purples, Pink	77.37	44.27
Soils-3	Browns, orange, Yellow/Brown, Aqua, Light Blue	85.72	31.18
Dune land	Sand	98.96	0.62
Soils-4	Yellow, Grey	66.67	16.01
All Cropland	Greens, White	98.95	8.35

* Based on training data.

APPENDIX III. Photointerpreted Soil Category Descriptions

(Numbers -N- shown refer to Textural Triangle)

(A) Wind blown deposits overlaying alluvial materials. Sandy surface, tan in color, and susceptible to wind erosion. -1-

(B) Grayish tan alluvial deposits of sand, silt, and clay from Colorado River with some wind blown deposits(striations). High sand content in soil texture. -1-

(C) similar to (B) except no aeolian deposition. -4-

(D) Sandy alluvial deposits from Colorado River with some inherent fertility and moisture capacity to support natural vegetation. -1-

(E) Similar to (D) except decreased sand content and increased fertility/moisture as shown by denser vegetation. Less wind deposition. -4-

(F) Reddish gray alluvial sand, silt and clay(loam) from Colorado and up-slope material. Water erosion is apparent. -4-

(G) Reddish gray alluvial fan of clay, gravel, and sand. Desert pavement and varnish have formed. -1-

(h) Similar to (G) except has a lighter surface and a higher sand content due to the volcanic parent material. -1-

(I) Reddish gray alluvial sand, silt, and clay from up-slope material. -4-

(J) Reddish gray alluvial deposition having its source area from the up-slope (H) and (G) materials. -4-

(K) Coarse textured(sandy), light colored(tans and grays), alluvial soils developed from igneous rock high in quartz content and very little ferromagnesium minerals. -1-

(L) Coarse textured alluvial fan from weathered pyroclastic bedrock high in quartz. Desert pavement and varnish have formed. Surface appears tannish gray. Steep slope supports water erosion. -1-

(M) Mixed area consisting of exposed bedrock and alluvial fans, some forming desert pavement. Conglomeration of schist, gneiss, and volcanic material. Surface appears dark tan. -1-

(N) Alluvial fan developed from parent material high in feldspar, therefore these soils will be fine textured(high clay content). Surface appears to be greenish gray. Drainage patterns of these fans also suggests high clay content. High susceptible to water erosion. -2-

(O) Sandy alluvial deposits from the Colorado River which support some natural vegetation. Soil is highly eroded and dissected. Agriculture production occurring on these soils.

(P) Soil resulting from lacustrine deposits intermixed/overlaid with high clay content alluvial fan material. Agricultural production occurring on this soil. Reddish gray in color. -2-

(Q) Lacustrine deposits(loam) from Lake Coahuilla consisting of sand, silt, and clay. Light gray in color. -4-

(R) Similar soil as (Q) except more inherent moisture/fertility as shown by natural vegetation. -4-

(S) Reddish tan lacustrine deposits that are high in clay content due to the shale parent material. -2-

(T) Reddish gray alluvial fan of mixed composition. Desert pavement and varnish have formed. -4-

(U) Lacustrine deposits of Lake Coahuilla intermixed with alluvial material from the up-slope fan. Tannish gray surface. -4-

(V) Similar to (U) except some uninterpretable difference as suggest by surface patterns.

(W) Residual soils high in clay and silt with a reddish gray surface color. -3-

(X) Lacustrine deposits from Lake Coahuilla with a reddish brown surface color. Smoother texture appearance and lack of dissected drainage suggests a soil texture high in sand. -1-

(Y) Alluvial deposits with a greenish gray surface color derived from the shale and sandstone parent material. -1-

(Z) Lacustrine sand, silt, and clay deposits(loam) from Lake Coahuilla. This delineation appears to be a watershed having its own drainage system which forms a major channel near the bottom of the watershed. Scattered natural vegetation appears to be showing up throughout the watershed. Reddish gray uo gray color. -4-

(AA) Reddish tan lacustrine deposits having a mottled appearance and drainage patterns which suggests higher clay content than soil in (X). -2-

(BB) Reddish gray or tan sand, silt, and clay deposits from Lake Coahuilla with a higher amount of sand and silt than clay as is suggested by the lack of numerous drainage patterns and smooth appearance of this delineation. Susceptible to wind erosion. -1-

(CC) Reddish gray alluvial material that has a predominance of sand in the soil texture due to the parent material being largely granite. -1-

(DD) Alluvial material similar to (CC) except lighter in color and the soil texture has a predominance of clay instead of sand. -2-

(EE) Yellowish gray alluvial deposits that appear to be susceptible to wind and water erosion. Difficult to interpret but appears to be some dune development occurring within this delineation. Also has some pockets of natural vegetation. -4-

(FF) Mainly composed of lacustrine deposits from Lake Coahuilla but has more inherent fertility or moisture due to the natural vegetation response. Reddish brown surface color. -4-

(GG) Terrace type deposits high in clay content with reddish gray surface. -2-

(HH) Colluvium and alluvium deposits high in sand size particles and susceptible to water erosion. -1-

(II) Sandy alluvial deposits that are tan in color and showing water erosion. -1-

(JJ) Alluvium of various mixtures of sand, silt, and clay depending on proximity to source area. This delineation is a collection of intermixed materials and could be subdivided into smaller units. Color varies but mainly reddish gray. -4-

((KK) Light gray alluvium deposits with higher proportions of silt and clay than sand. -3-

(LL) Similar to (KK) except higher in sand and less erosion. -1-

(MM) Gray outwash material of mixed composition. -4-

(NN) Similar to (F) except have sand being deposited by the wind. -1-

APPENDIX IV. Characteristics of the Surveyed Soils

slope symbol	erosion symbol	
A-0 to 2%	no symbol-none/slight	w1-irrigated soils
B-2 to 5%	2-moderate	
C-5 to 9%	3-severe	

Map Symbol	Soil Association	Surface Color	Surface Texture	Erosion Hazard	Inherent Fertility
HT-Aw1	Holtville	Light Brown	Silty clay	None	High
Ga-Aw1	Gadsden	Light brown	Silty clay	None	High
Go-Im-Aw1	Glendale-Imperial	Light Brown	Clay loam	None	High
Go-Im-AB2	Glendale-Imperial	Light Brown	Clay loam	Mod.	High
Im-Aw1	Imperial	Light Brown	Silty clay	None	High
Gi-Vo-Aw1	Gila-Vinton	Light brown	Loam	Slight	Mod.
Gi-Vo-A	Gila-Vinton	Light brown	Loam	Mod.	Mod.
Mx-Gi-Aw1	Meloland-Gila	Light brown	Loam	Slight	Mod.
Mx-Gi-AB	Meloland-Gila	Light brown	Loam	Slight	Mod.
BM-Vo-A	Brazito-Vinton	Light brown-Light Gray	Sandy loam	Mod.	Low
Vo-BM-AB1	Vinton-Brazito	Light brown-Light Gray	Sandy loam	Mod.	Low
Cr-Cd-A	Carrizo-Cajon	Pale Brown Lt. brn. gray	Sand	Slight -mod.	Low
Cr-Cd-BC	Carrizo-Cajon	Pale brown Lt. brn. gray	Sand	Slight -mod.	Low
NI-Im-ABw1	Niland-Imperial	Light brown	Loamy sand	Slight	Low-High
NI-Im-AB2	Niland-Imperial	Light brown	Loamy sand	Mod.	Low-High
ss-Ab-A	Superstition Acolita	Pinkish gray Reddish brn.	Sand	Slight-Mod.	Low
By-HN-AB	Bitterspring Harqua	Pink	Loam	Slight	Low-Mod.
Rz-Os-AC	Rillito-Orita	Light brown	Sandy loam	Slight-Mod.	Low-Mod.
AZ	Alluvial materials	----	----	None	----
BZ	Badlands	----	----	High	----
DL	Dunes	----	----	High	Low
LP	Lava rock	----	----	None	----
PY	Playas	----	----	None	----
RB	Rough broken land	----	----	High	----
RL	Rock land	----	----	High	----

FIGURES

FIGURE 1

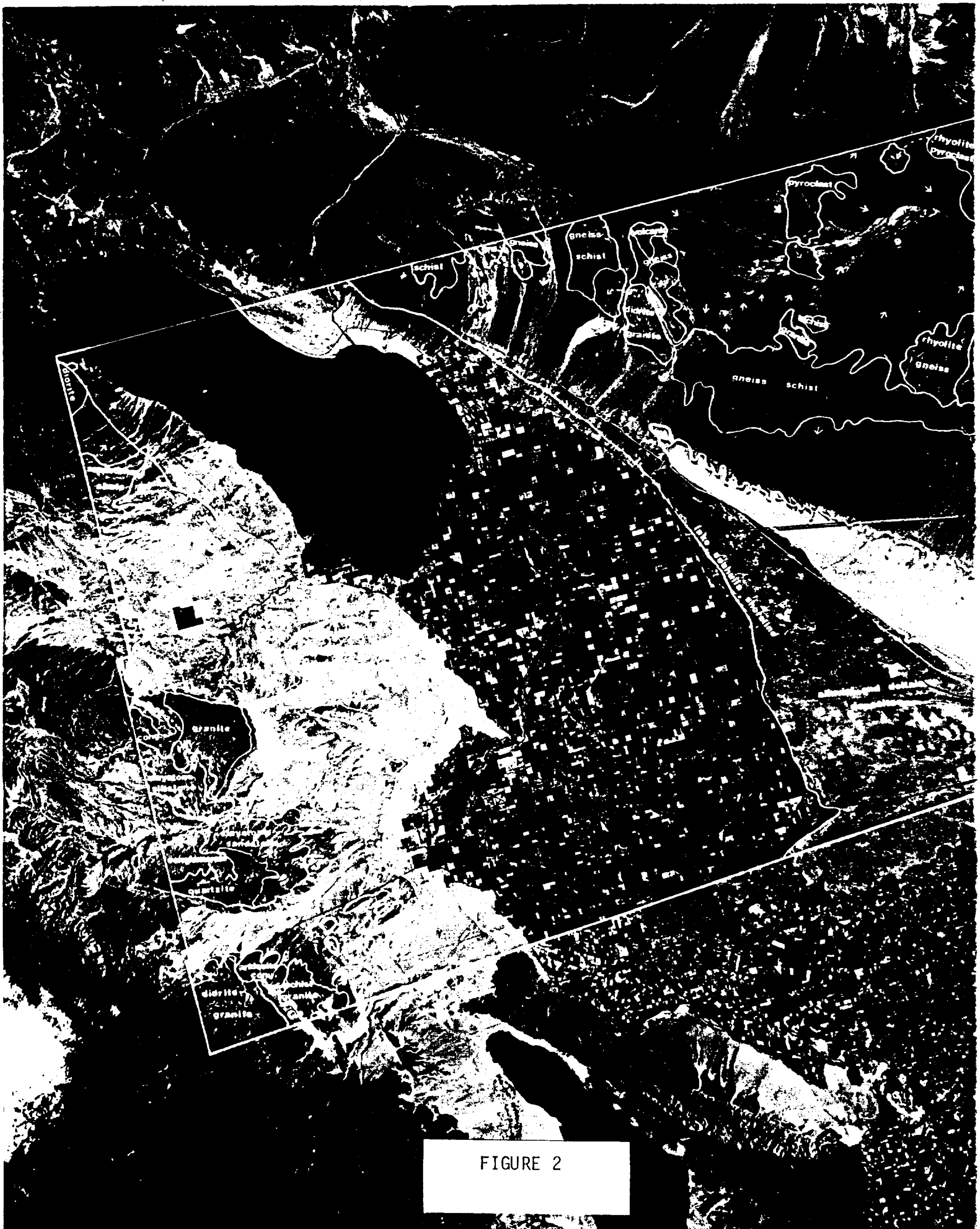
rhyolite
pyroclast
pyroclast
gneiss
schist
schist
gneiss
volcanic
diorite
granite
rhyolite
gneiss
gneiss
schist
diorite
granite
quartzite
diorite
granite
granite

FIGURE 2

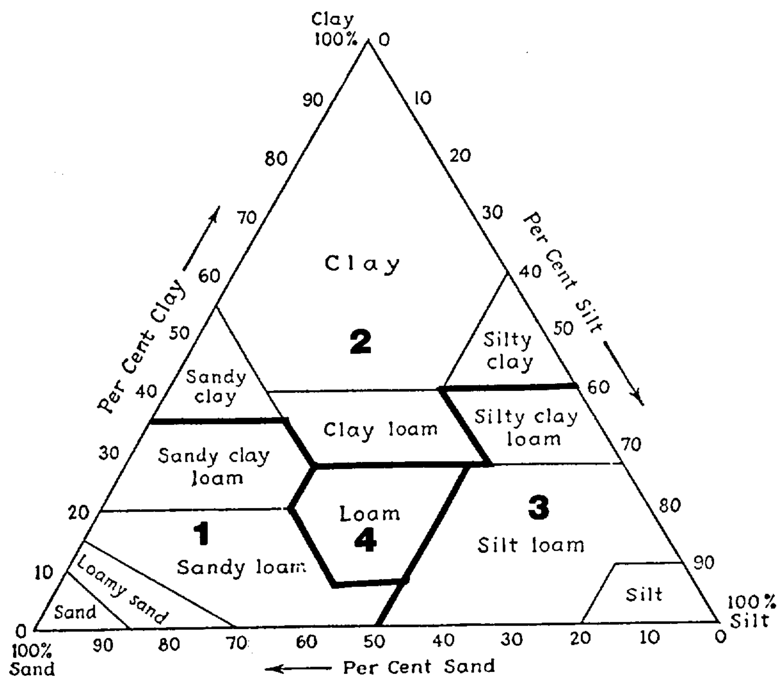
Clay
100%
Per Cent Clay
Per Cent Silt
Per Cent Sand
100%
Sand
100%
Silt
Clay
2
Sandy
clay
Silty
clay
Clay loam
Silty clay
loam
Sandy clay
loam
Loam
4
3
Silt loam
1
Sandy loam
Loamy sand
Sand
Silt

FIGURE 3

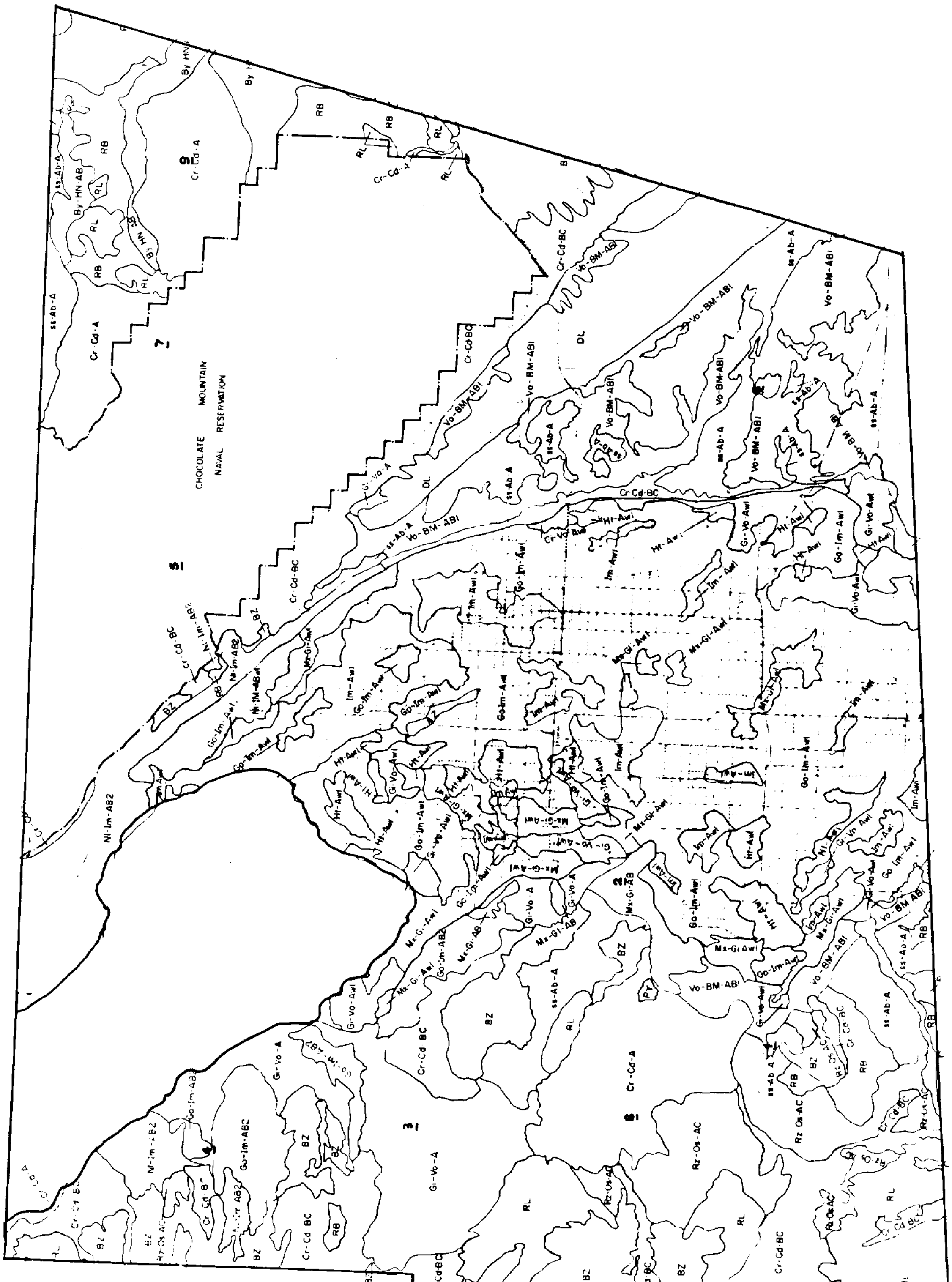

FIGURE 5

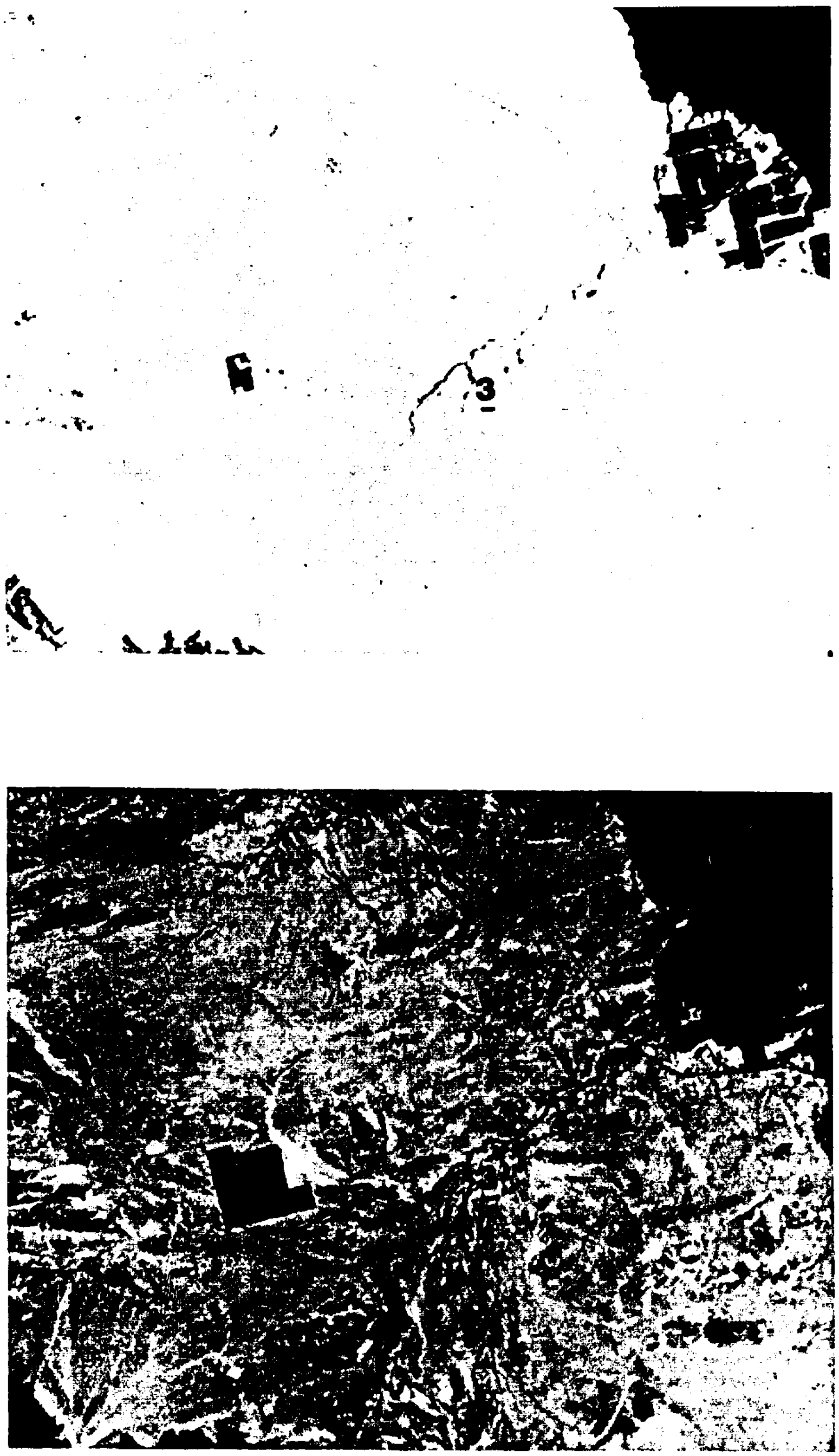

FIGURE 6