AN APPLICATION OF LANDSAT DIGITAL DATA TO AIR QUALITY PLANNING IN THE TUCSON URBAN AREA

by

Charles Hayward Sauerwein, Jr.

A Thesis Submitted to the Faculty of the

DEPARTMENT OF GEOGRAPHY, REGIONAL DEVELOPMENT AND URBAN PLANNING

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF ARTS
WITH A MAJOR IN GEOGRAPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 0

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the intersts of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Charles H. Sauerwein Jr.

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

David A. Mouat
DAVID A. MOUAT
Assistant Professor of Geography
and Arid Lands

15 August 1980
Date

ACKNOWLEDGMENTS

The research for and writing of this report were done at the Applied Remote Sensing Program, Office of Arid Lands Studies, University of Arizona, under the direction of Drs. David A. Mouat and Charles F. Hutchinson. Support was provided by a National Aeronautics and Space Administration grant number NGL 03-002-313.

I would like to express my gratitude to my committee, Drs. Mouat, Hutchinson, and Richard W. Reeves, for their advice, patience, and comments on the thesis draft. Particular appreciation is expressed to Dr. Robert A. Schowengerdt for his encouragement, helpful suggestions, and tutoring on the use of the computer software used in this report. I am also grateful to Ms. Karen Heidel of the Pima Association of Governments, Environmental Planning Division, for her help in familiarizing me with the air quality planning process in Tucson. Additional thanks go to Ms. Billie Jo Lobley for her graphics work and to numerous people at the Applied Remote Sensing Program for their helpful suggestions and encouragement.

Finally, I would like to convey special thanks to my friends and family; especially my parents, whose moral support was instrumental in the completion of this project.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS vi

LIST OF TABLES vii

ABSTRACT viii

CHAPTER

1. INTRODUCTION 1

Statement of Problem 1
Purpose 1
The Air Quality Planning Process 3
The Study Area 6

2. TOTAL SUSPENDED PARTICULATES IN THE TUCSON URBAN AREA 14

Meteorological Variables and TSP Patterns 14
Total Suspended Particulates 20

3. THE CLASSIFICATION OF LANDSAT DIGITAL DATA 28

Applicability to Dust Source Monitoring 28
Spectral Data and Observation Space 32
Logic and Purpose of Classification 37
Approaches to Classification of Landsat Digital Data . 38

4. SURFACE COVER CLASSES IN CLASSIFICATIONS OF LANDSAT DIGITAL DATA 47

Previous Work on Classification of Landsat Digital Data in Urban Areas 48
Class Descriptions 54
Built-Up Area 54
Outlying Area 60

5. METHODOLOGY AND RESULTS 64

Preprocessing 65
Digital Image Processing 65
Classification 68
Results 72

TABLE OF CONTENTS--*Continued*

Page

6. SUMMARY AND CONCLUSIONS 82

REFERENCES 85

LIST OF ILLUSTRATIONS

Figure		Page
1.	Tucson urban and air planning areas	7
2.	Tucson urban area extracted from Landsat scene	8
3.	Five year monthly arithmetic means of TSP levels (Health and Welfare Building sampler)	17
4.	TSP concentrations in Tucson (annual geometric means in micrograms/m^3)	18
5.	General air flow in Tucson area during temperature inversion	19
6.	TSP sampler sites in the Tucson urban area (1980)	24
7.	Annual TSP sampler data for sites in the Tucson urban area (geometric means in micrograms/m^3)	26
8.	Landsat coverage of Arizona	30
9.	Spectral responses of water, soil, and vegetation	34
10.	Spectral responses of water, soil, and vegetation in two-dimensional observation space	36
11.	Clusters in observation space	40
12.	Parallelepiped classifier	43
13.	Decision rules for two classifiers	46
14.	Tucson sample area	67
15.	Unsmoothed image data	73
16.	Smoothed image data	74
17.	Classification example: South Tucson and Tucson Central Business District	79
18.	Classification example: Residential area in Tucson	80
19.	Classification example: Outlying area	81

LIST OF TABLES

Table		Page
1.	Population Growth in Tucson	12
2.	Air Quality Standards for TSP	22
3.	TSP Sampler Sites in the Tucson Urban Area	25
4.	Landsat System Information	33
5.	Land Use and Land Cover Classification System for Use with Remote Sensor Data	49
6.	Mean Reflectance Values of Training Data for All Classes	56
7.	Classification Summary	75
8.	Summary of Classification Accuracy by Test Sites	77

ABSTRACT

In urban centers in arid and semi-arid areas of the American West and Southwest, including Tucson, dust resulting from human disturbance of the ground and deflation has caused ambient concentrations above federal standards. In Tucson, the location and extent of dust sources are highly variable because of the continuing population growth, accompanying development, and resulting disturbance of the desert environment. The variable location of dust sources in Tucson is not monitored with current annual inventorying methods.

The Landsat satellite system provides repetitive coverage of the earth in image and digital data formats. Landsat, therefore, can provide air quality planners with an opportunity to monitor dust sources and related potential dust emissions. A classification of general surface cover was performed to determine the efficiency of Landsat digital data in monitoring disturbed areas. Overall accuracy averaged 90%, although inspection of the classification map revealed confusion among certain classes. Nevertheless, the ability to identify high dust emission potential areas was demonstrated. It is proposed that annual monitoring of surface cover conditions with Landsat digital data will aid air quality planners in lowering ambient dust levels to comply with federal air quality standards.

CHAPTER 1

INTRODUCTION

Statement of Problem

The Tucson urban area has been classified as part of a non-attainment area for total suspended particulates (TSP) by the Environmental Protection Agency (EPA) because of violations of the National Ambient Air Quality Standards (NAAQS) (Heidel, Hyde, et al., 1978). In Tucson, the TSP problem is essentially a dust problem; primarily the result of disturbance of the desert environment by the urban populace (Heidel, Dixon, and Young, 1978; Moyers, 1977). As part of the required Nonattainment Area Plan (NAP) a detailed TSP emission inventory for Pima County was completed in 1978 (Heidel, Hyde, et al., 1978). This inventory identified all sources on a small grid basis for base year 1975. Annual emission updates estimate the total emissions by pollutant for the entire county without regard to the location of these sources. The location of dust sources is highly variable because of the substantial population growth and associated ground disturbance in Tucson. Thus, air quality planners in Tucson are at a disadvantage in meeting TSP standards without yearly updates of the changing location of dust sources.

Purpose

The purpose of this thesis is to examine the potential of Landsat digital data for inventorying areas of potential dust emissions. Landsat satellites have been orbiting the earth since 1972, providing essentially complete coverage of the earth every 18 days. Landsat senses the earth's reflectance in four wavelength intervals (or "bands") of the visible and near infrared spectrum and records the data in digital format. In this format the data can be manipulated by computers to perform image processing to enhance features for interpretation purposes and image analysis operations, such as classification, to extract information.

A classification of surface cover was performed on the Tucson urban area using Landsat digital data. Classes were selected on the basis of their potential dust emissions. It is proposed that annual monitoring of general surface cover changes resulting from activities of the urban populace be done using Landsat data. In this way air quality planners would have a current picture of the locations and extent of areas that have high dust emission potential. Air quality analysts could then determine if a relationship exists between the change in location and the extent of those areas and dust concentrations.

Four potential benefits could derive from the use of Landsat data in the air quality planning process.

1. Failure to meet EPA air quality standards could result in a loss of federal funds from the Departments of Transportation and Housing and Urban Development amounting to $140 million through 1982. If the proposed updates using Landsat data are implemented in the NAP, the additional information provided could be instrumental in averting federal funding losses.
2. Due to the high cost of preparing detailed, gridded emission inventories, the conduction of such inventories is totally

dependent on EPA funding. Updates using Landsat data could be used to identify source areas responsible for specific TSP violations at a fraction of the cost of obtaining this information from traditional methods.

3. Monitoring the variable locations and extent of dust sources could be useful for local planning agencies by identifying dust problem areas. As an efficient cost effective dust control strategy, abatement measures (road paving, land use regulation) could be focused on areas of high potential dust emissions located with the Landsat updates. Concentrating abatement efforts might substantially reduce the $20 million expenditures estimated by air quality planners needed to implement the NAP strategy and achieve compliance with the air quality standards.
4. The information obtained through annual Landsat updates could be used in the evaluation of sites in the Tucson urban area for new TSP samplers and for the relocation of existing samplers now designated as site specific and thus of little value for area-wide air quality planning.

The Air Quality Planning Process

Although the first federal air pollution legislation in the United States was enacted a quarter of a century ago, it was not until the Clean Air Act Amendments of 1970 were passed that the current goals and strategies of national air pollution control were initiated. In 1970, the Environmental Protection Agency (EPA) was established, and, as one of its initial acts, air quality standards were set for the entire nation. Heretofore, each state was to establish its own air quality standards and draft plans to meet those standards by setting specific emission levels by source and a timetable for achieving compliance. According to the new legislation, each state was to develop plans to implement EPA's national ambient air quality standards (NAAQS).[1]

In 1971, standards were established for particulate matter, carbon monoxide, hydrocarbons, oxides of nitrogen and sulfur, and photochemical oxidants, to serve as threshold levels below which no adverse effects would occur. Amendments to the Clean Air Act in 1977 required the classification of areas that had not attained ambient standards as "nonattainment areas" and required states to draw and implement plans for attainment by December 31, 1982 (Easton and O'Donnell, 1977).

As a result of the Clean Air Act Amendments of 1977, Pima County, located in southern Arizona, was designated as a nonattainment area for total suspended particulate matter (TSP). As an initial step in developing the Nonattainment Area Plan (NAP), the Pima County Air Quality Control District (PCAQCD) of the Arizona Department of Health Services submitted a detailed gridded particulate emissions inventory to EPA in February, 1978 (Heidel, Dixon, and Young, 1978). The purpose of the inventory was to identify, quantify, and characterize particulate emissions from all sources on a small grid basis for all of Pima County. In addition, the report contained projections of emission levels at the end of each five year period from 1980 through the year 2000.

1. For a detailed review of federal air pollution legislation in the United States, see Berry and Horton (1974) and Bibbero and Young (1974).

Due to the extensive requirements of the NAP, the Pima Association of Governments (PAG) was designated as the lead agency for NAP development in cooperation with the Pima County Air Quality Control District (PCAQCD) (Heidel, Hyde, et al., 1978). The PCAQCD is responsible for monitoring pollutants, analysis of pollutant data, and technical support for planning efforts; e.g., the technical analysis of the TSP NAP in 1978. PAG on the other hand is responsible for planning strategies to meet the NAAQS (Bridson, 1980).

Emissions inventories have been compiled by the PCAQCD since 1973 for all pollutants regulated by EPA. Total emissions for the entire county are estimated without regard to the location of the sources. In contrast, in the 1978 Emission Inventory for TSP (base year data for 1975), emissions were defined by source and calculated within a grid network of areas ranging from 1 km^2 to 10 km^2 depending on population density.

Modifications were proposed to the NAP in September, 1978, based on the results of the gridded emission inventory. Most relevant to this thesis was the proposal to limit planning to the Tucson urban area. Two justifications were offered. First, the primary concern of air quality control is the protection of public health and approximately 90% of the total population of Pima County resides in the Tucson urban area. Second, outside of the Tucson urban area, the only known violations of the NAAQS for TSP in Pima County have occurred in the communities of Ajo and Rillito, where mining and agricultural activities, respectively, are the main sources of TSP and the reasons for NAAQS violations (Heidel, Hyde, et al., 1978). According to NAP projections, none of the non-urban areas of the county except Ajo and Rillito would exceed NAAQS through the year 2000.

The proposal to limit the planning area to the urban area was rejected by EPA and a new Tucson Air Planning Area (TAPA) (Figure 1), based on the physical airshed boundaries affecting Tucson, was officially established (Environmental Protection Agency [EPA], 1979). Fifteen per cent of this newly redefined nonattainment area for TSP falls outside Pima County. This 15% is considered to be important enough to include in the planning area, but is not considered in the NAP by PAG because it is outside its jurisdiction.

The Study Area

Tucson occupies a sediment-filled basin at an elevation of 2,500 feet and is surrounded by mountains ranging from 5,000 to 9,000 feet (Figure 2). The Santa Cruz River flows intermittently through the basin in an entrenched arroyo. The main tributary washes to the Santa Cruz, Cañada del Oro and Rillito Creek, are usually dry but may overflow their banks during times of heavy rainfall. Rainfall averages less than 11 inches annually in the city, with a bi-seasonal regime of approximately 60% in summer and 40% in the winter months. In July, daily high and low temperatures average 98.3° and 74.2°F, respectively, and January's high and low average 63.5° and 38.2°F (National Oceanic and Atmospheric Administration [NOAA], 1978).

Vegetation in and around Tucson is typical of the Lower Sonoran Lifezone (Lowe, 1972). Creosotebush (<u>Larrea divaricata</u>) dominates the basin or bottomland areas, characterized by shrubs, dwarf shrubs, and

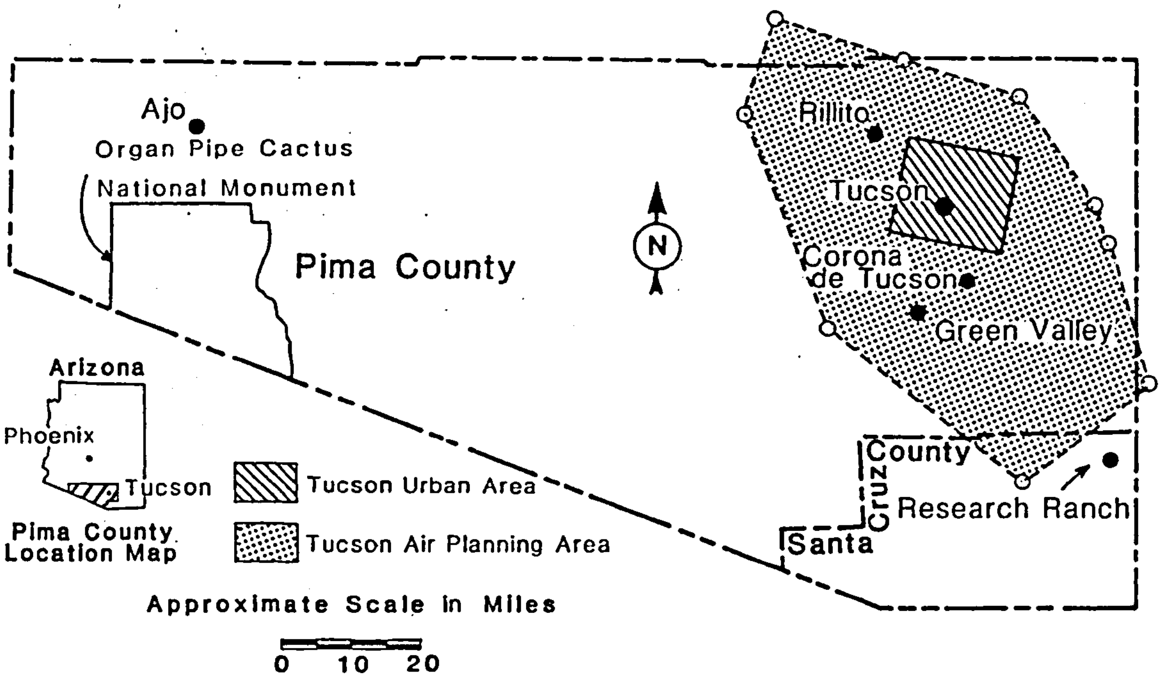


Figure 1. Tucson urban and air planning areas.

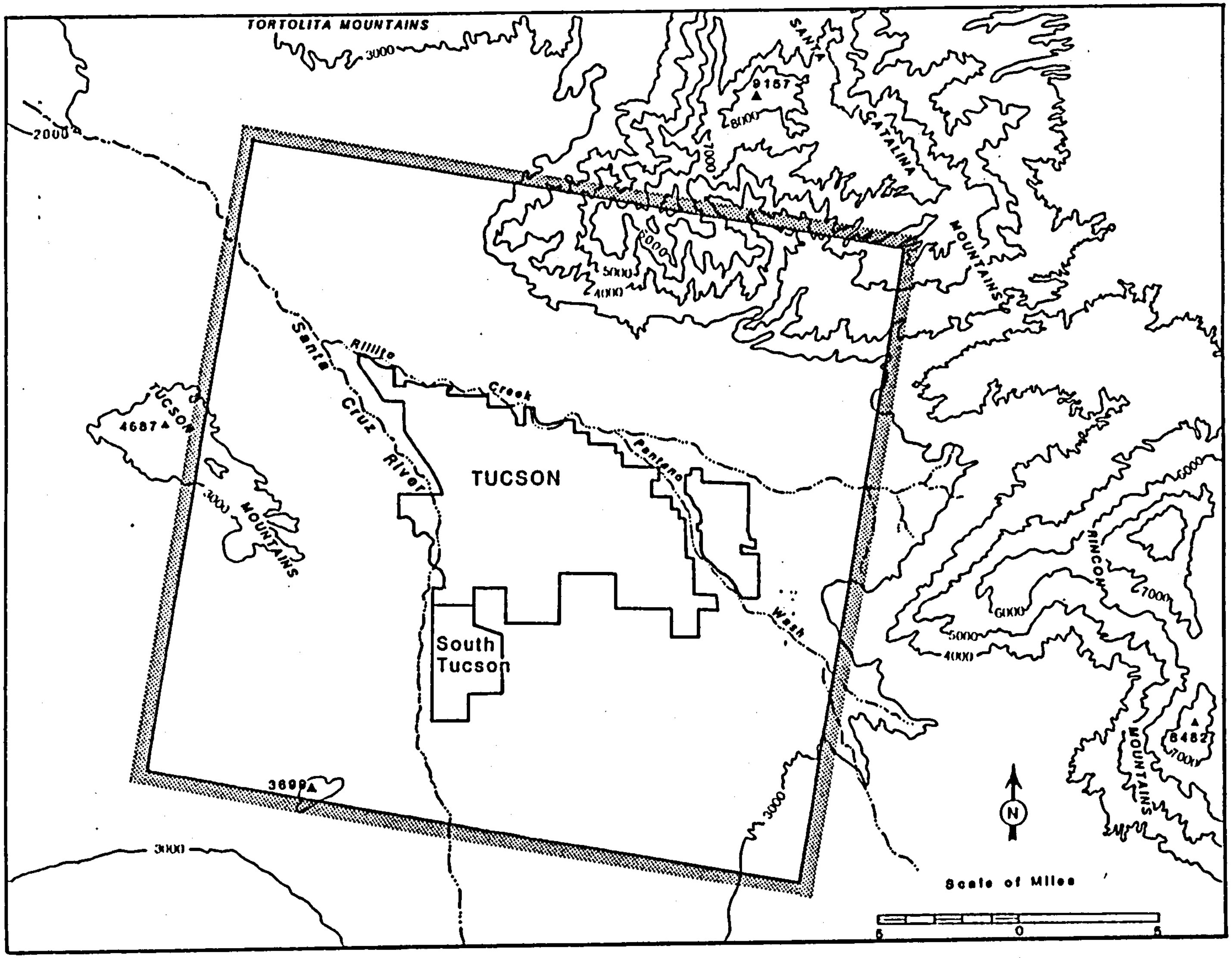


Figure 2. Tucson urban area extracted from Landsat scene.

a general lack of trees. In the surrounding rocky foothills and on the coarse soil of the bajadas the paloverde-saguaro (Cercidium-Cereus) community is dominant. This community consists predominantly of small-leaved desert trees as well as shrubs and cacti. Riparian vegetation in this area generally consists of mesquite (Prosopis juliflora), blue paloverde (Cercidium floridum), and catclaw (Acacia greggii), along with other trees and shrubs. Steenburgh and Warren (1977) showed that vegetative cover in similar and proximate areas ranges from 8-15% for creosotebush-dominated communities and approximately 22-32% for paloverde-saguaro communities.

Thus, Tucson is situated in a hot and dry environment, where vegetative cover is sparse. There are extensive areas of bare ground that readily emit dust into Tucson's air when disturbed and continue to do so through deflation. In addition, the revegetation of the disturbed areas is a slow process, which results in a prolonged period of high dust emissions.

In the Tucson urban area the TSP problem is essentially a dust problem rather than one of emissions from industrial activities and fuel combustion (Heidel, Dixon, and Young, 1978; Moyers, 1977). Results of the February, 1978, TSP Emissions Inventory for Pima County revealed that over 90% of the total emissions were dust, mostly from unpaved roads and road shoulders, and construction and agricultural activities. Only 8.5% of TSP were attributed to stationary sources, mainly residential and industrial fuel combustion and less than 1% to mobile sources including motor vehicles, aircraft, and trains. In the Tucson urban area, dust sources were estimated to account for over 93% of TSP emissions, stationary sources 6%, and mobile sources 1% (Heidel, Hyde, et al., 1978). In a study done at The University of Arizona Analytical Center on the chemical properties of TSP in the Tucson urban area, the soil component was found to comprise from 50-82% of the total mass (Moyers, 1977). In a TSP emissions inventory prepared for EPA in the Phoenix urban area, 112 miles to the northwest of Tucson and situated in a similar desert environment, dust emissions were estimated to account for 98% of all TSP emissions (Richard, Tan, and Avery, 1977).

The desert setting of Tucson is not by itself responsible for the NAAQS violations (Heidel, Dixon, and Young, 1978; Moyers, 1977; Hartmann, 1976; Caldwell, 1971; Pima County et al., 1977). TSP levels at Organ Pipe Cactus National Monument and at Corona de Tucson, both relatively undisturbed areas of the Sonoran Desert, have remained well below air quality standard levels in the time that TSP samplers have been in operation at these locations (Heidel, Hyde, et al., 1978). The Research Ranch, Inc., located 70 miles southeast of Tucson, was used as a desert "background" station in The University of Arizona Analytical Center study on TSP. This choice was based on the supposed absence of local disruptive human activities and the resulting native desert conditions maintained on the land. Total mass concentrations at this site were well below NAAQS levels while only one urban sampler out of eight in Tucson met the standard levels for TSP. These data strongly suggest a low background of desert TSP concentrations, and that violations of air quality standards in Tucson are primarily the result of activities of the urban populace. The "natural" desert background concentration has been estimated to comprise only one-third of the typical urban levels in Tucson (Pima County Air Quality Control District [PCAQCD], 1980).

Tucson is an area of rapid growth. The population of the Tucson urbanized area grew 29.4% from 1960 to 1970 (U. S. Bureau of the Census, 1973). The Bureau of the Census defined "urbanized areas" as consisting of a central city or cities with a combined population of 50,000 or more and the urban fringe or "surrounding closely settled territory." While the population within Tucson's incorporated limits grew 23.5% between 1960 and 1970 the most dramatic increase occurred in the urban fringe where the population grew almost 115%. The per cent of the Tucson urbanized area population residing in the urban fringe also rose from 6.4% in 1960 to 10.6% in 1970.

In 1975 a special population census was completed for Pima County, but no statistics were generated for the Tucson urbanized area. This special census does provide tabulations for the Tucson Census County Division (CCD), which is considered to represent the local trade or service area of Tucson. The statistics of the Tucson CCD are much more closely aligned with the urbanized area statistics than to those of the Tucson SMSA (equivalent to all of Pima County), and show a 14.6% increase in population between 1970 and 1975 (Table 1).

With such growth, the location and extent of potential dust sources are highly variable from year to year especially in the urban fringe. In the annual emission inventories compiled by the PCAQCD, the variable location and extent of many dust sources are not accounted for. For those inventories, estimations of area dust source emission are projected from the 1975 base year data using a variety of secondary data

Table 1. Population Growth in Tucson

	1960	1970	% Increase 1960-70	1975	% Increase 1970-75
Tucson Urbanized Area	227,433	294,184	29.4%	--	--
Central City[a]	212,892	262,933	23.5%	298,683	13.6%
Urban Fringe	14,541	31,251	114.9%	--	--
Tucson Census County Division[a] (Local Trade & Service Area)	219,896	290,661	32.2%	333,137	14.6%
Tucson SMSA (Pima County)	265,660	351,667	32.4%	449,544	27.8%

Sources: U. S. Bureau of the Census (1973, 1976); Gennaro (1979).

[a]Changes in boundaries have occurred since 1960.

(Heidel, Dixon, and Young, 1978). Projected changes in population are used to estimate emissions from road shoulders, sand and gravel pits, and reentrained dust from paved streets. Projected changes in population density are used to estimate emissions from construction activities. Emissions from agricultural and mining activities are estimated from changes in acreages obtained through questionnaires. Updated estimated emissions from unpaved roads are made by adding the mileage of new dirt and gravel roads to existing mileage estimates and by monitoring changes in VMT (vehicle miles traveled) on these roads.

If Landsat digital data can be used successfully to identify high dust emission potential areas, this information could be used by air quality planners to understand direct relationships between source emissions and dust concentrations in Tucson's air. This information could provide a basis for planning realistic dust abatement measures. Without a comprehensive picture of the changing spatial distribution of potential dust emission areas, air quality planners may be seriously hampered in their efforts to reduce dust levels to comply with federal standards.

CHAPTER 2

TOTAL SUSPENDED PARTICULATES IN THE TUCSON URBAN AREA

Based on annual averages from 1973 to 1978, only 9.4% of the labor force in Pima County is employed in manufacturing (Gennaro, 1979). Most of the manufacturing activities are not heavy polluting industries and as a result the pollution in Tucson is not highly toxic. Dust or soil particles comprise the majority of the TSP mass in Tucson's air. Most of the dust is attributed to the disturbance of the ground by activities of the urban populace. In arid and semi-arid areas, revegetation of disturbed areas can be a much slower process than in a humid environment. As a result, disturbed areas are potential sources of dust for longer periods of time in arid and semi-arid areas than they would be in more humid areas.

Meteorological Variables and TSP Patterns

Urban centers in arid and semi-arid regions of the American West and Southwest have had difficulty in meeting NAAQS for TSP. It has been found that dust resulting from human disturbance of the ground and deflation cause ambient concentrations above federal standards in these urban centers (EPA, 1977). Data for 1976 (PCAQCD; 1980) show that TSP levels in five other large urban areas in the dry region of the U. S. (Phoenix, Albuquerque, Las Vegas, El Paso, and Salt Lake City) exceed the NAAQS. Surface disturbance and the enforcement of dust abatement measures influence TSP levels in these urban areas, but meteorological variables play a large role in controlling temporal and spatial patterns of TSP.

In Tucson, daytime heating during summer months produces intense vertical and horizontal mixing in the atmosphere. The injection and residence time of dust in the atmosphere is increased by this turbulence. Small particles, mostly secondary particulates formed by complex chemical reactions between gases, are mixed and diluted by this activity. Thus, concentrations of the larger dust particles are high in the hot, dry summer months while concentrations of the smaller secondary particles are relatively low.

In mid-summer, usually in early July, upper level moist tropical air is drawn into Arizona around a high pressure cell protruding into the central United States from the Atlantic Ocean. This moist air mass is conditionally unstable and subjected to orographic uplift and intense thermal heating in the Tucson area, which trigger strong vertical currents and heavy but brief rainshowers. In the mountainous southeast section of Arizona, the ratio of summer to winter precipitation is higher than in flatter southwestern Arizona because of the greater mechanical uplifting of the moist tropical air (Sellers and Hill, 1974). Tucson receives 60% of its precipitation in the summer. Downdrafts cooled by the condensation of moisture in cumulonimbus clouds sometimes cause severe dust storms in and near Tucson, especially when the subsequent rain evaporates before reaching the ground. Usually, though, the ensuing rainfall removes the dust from the air through washout. As a result, TSP levels are at a minimum in Tucson during the peak of the summer rainy season (July-September).

In the winter months, most of Arizona's precipitation is associated with large scale low pressure cells and fronts that pass eastward across the state. These cyclonic storms occur intermittently in constrast to the summer rains which occur almost daily, and thus TSP concentrations are not affected as much by washout in the winter. Of more importance to TSP concentrations in the winter months are the radiational inversions that often persist until mid-day and may last for several days without breaking (Mees, 1965). Though some dust particles are removed by sedimentation, the overall effect of the persistent winter radiational inversions in Tucson is higher concentrations of TSP (Figure 3).

Whereas air turbulence and stability and rainfall affect the seasonal pattern of TSP concentrations in Tucson, the spatial distribution of TSP levels is strongly influenced by relative elevation and the orientation of surface winds. The Tucson Basin slopes gently downward toward the northwest. Low-lying areas in this direction tend to have higher TSP concentrations than other parts of the airshed (Figure 4), in response to nighttime and early morning air drainage produced by radiational cooling of the air near the ground (Figure 5). In the spring and fall, the nighttime and early morning winds are from the southeast more than 80% of the time (Sellers and Hill, 1974). Early morning commuter traffic also generates abundant dust, especially from heavily traveled roads with unpaved shoulders (PCAQCD, 1980), and much of this is transported to the northwest by the prevailing southeasterly

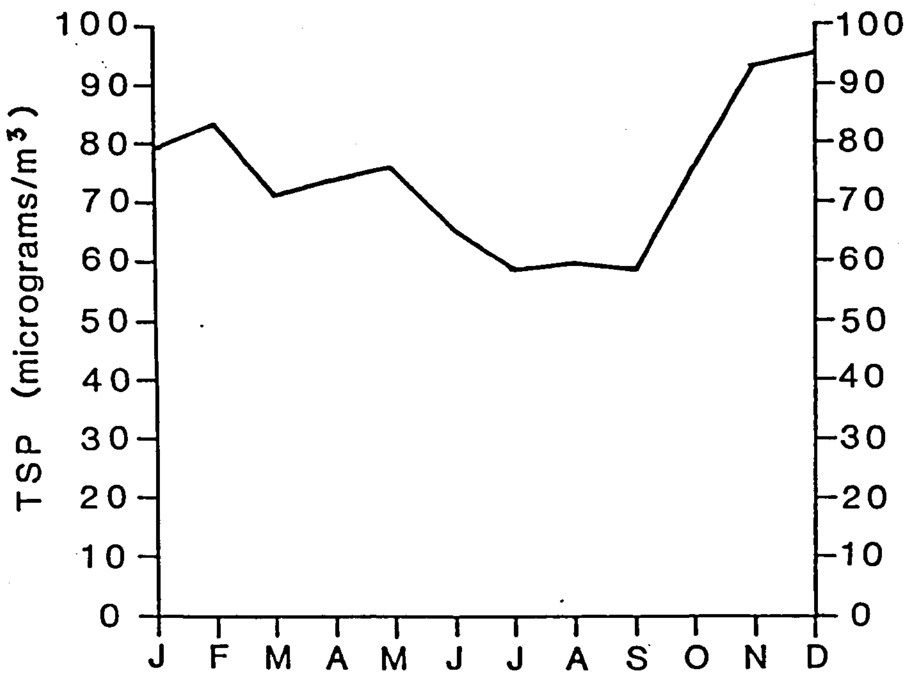


Figure 3. Five year monthly arithmetic means of TSP levels (Health and Welfare Building sampler).

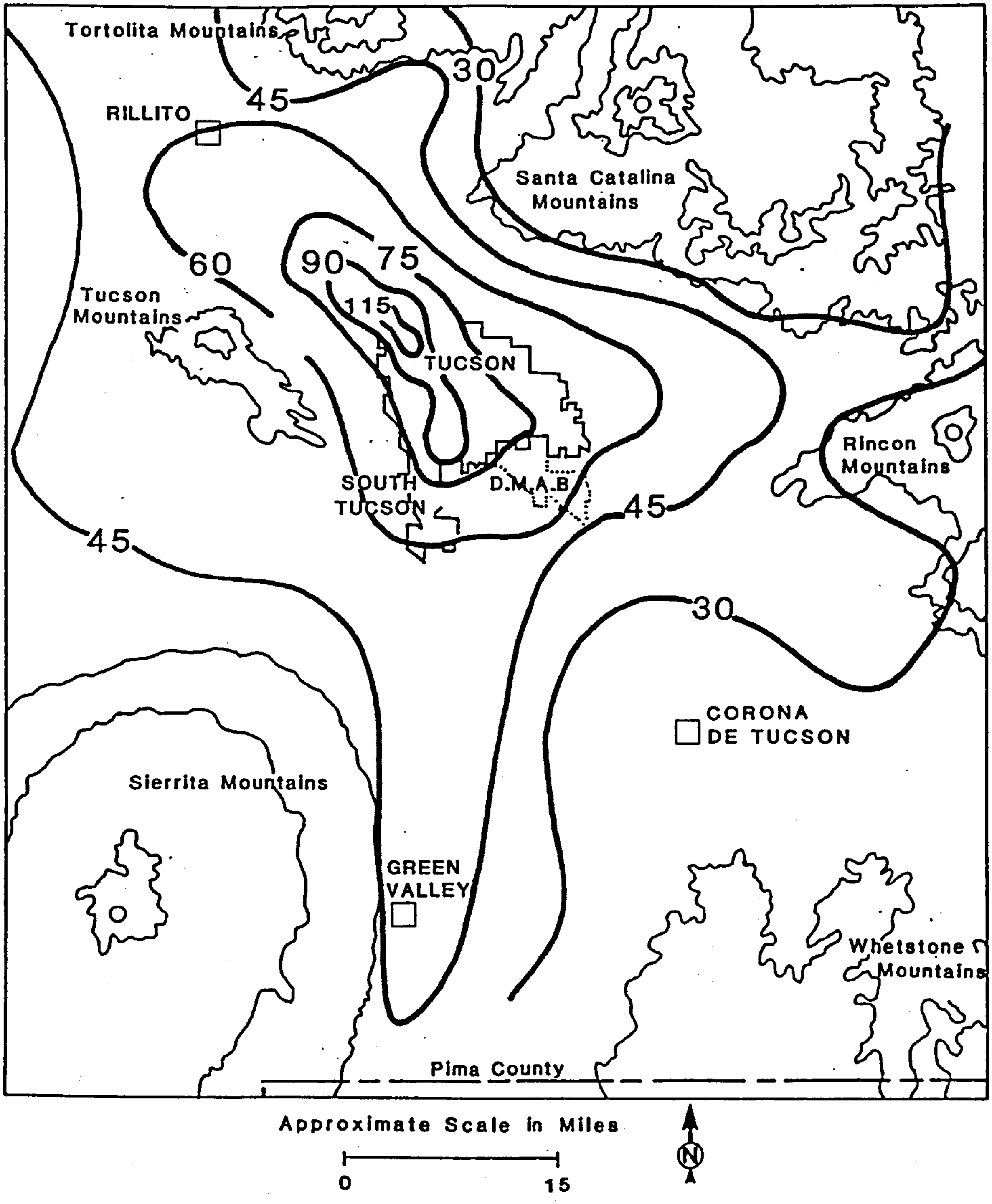


Figure 4. TSP concentrations in Tucson (annual geometric means in micrograms/m^3).

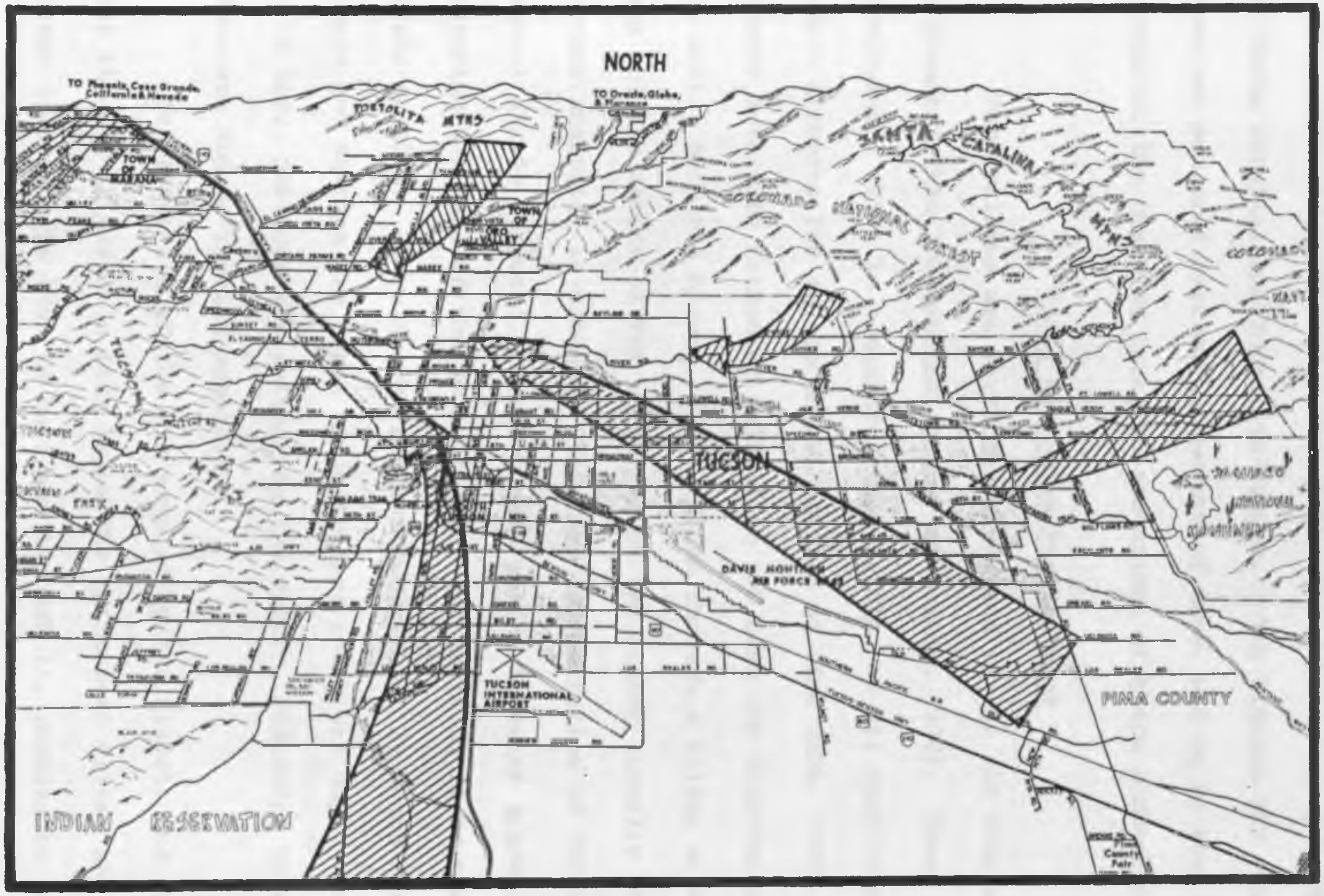


Figure 5. General air flow in Tucson area during temperature inversion -- Adapted, with permission, from a map drawn by R. A. Drafting Service (1979).

winds. There is also a tendency for these downslope winds to stagnate in the north and northwest parts of the area where they are impeded by the Santa Catalina and Tortolita ranges. The highest TSP concentrations are measured near major sources of dust, and it is thus necessary to monitor the changing location and extent of these sources frequently.

Total Suspended Particulates

Particulates are particles dispersed in the air that are smaller in diameter than 500 microns (Berry and Horton, 1974). Total suspended particulates (TSP) is a measure, used by EPA, of all particles that remain airborne for an appreciable period of time (EPA, 1977). The Federal Reference Method of collecting TSP uses the high-volume sampler (hi-vol). With the hi-vol, air is drawn through a filter at a constant rate and the amount of TSP collected on the filter (usually adequate for further analysis) is weighed. The mass concentration of TSP in the ambient air is computed, in micrograms per cubic meter ($\mu g/m^3$), by measuring the mass of the collected TSP and dividing it by the volume of air sampled. Hi-vols are run continuously for 24 hours for each sample and samples are taken at least once every six days. On rainy or humid days, moisture may collect on the filter, resulting in a slower flow-rate and an invalid sample.

The flow-rate of the air sampled by the hi-vol, and the fact that the air is drawn in from below the horizontal filter and must change direction by 180° to be drawn through it, combine to produce a filtering effect whereby those particles larger than 60-70 microns in diameter are separated as a result of their mass inertia and do not reach the filter (Birkle, 1978). This, in effect, sets the maximum size of particulates which are included in the TSP measurement to about 60-70 microns. Previous analysis of TSP collected by the hi-vol indicated that the threshold may be as low as 30 microns (Cowherd et al., 1974).

The mass of TSP, from samples collected in fourteen urban areas in the U. S., shows a bimodal distribution. The distinction between fine and coarse particles is made at approximately 2 microns. The finer particles make up most of the TSP in terms of particle numbers and surface area, whereas the larger particles comprise most of the mass. The size range for the finer particles is called the "accumulation" mode since these particles are formed by coagulation of even smaller particles or by chemical reactions between gases and previously existing particles. The larger particles, in the "coarse" mode, are dominated by compounds of soil and mineral derived elements and are formed mainly from mechanical processes such as some industrial processes, soil disturbance, reentrained street dust, and rubber tire wear (EPA, 1977).

National ambient air quality standards (NAAQS) for TSP were established by EPA to denote the legal limit on the level of atmospheric contamination necessary to protect against adverse effects on public health and welfare. National primary standards are levels judged necessary, within an adequate margin of safety, to protect the public health. National secondary standards are levels judged necessary to protect the public welfare from any known or anticipated adverse effects of a pollutant (EPA, 1971). Specifically, welfare standards are for the protection of plants and building materials. Primary and secondary standards were established for two types of TSP measures. First, an annual geometric mean[2] is computed for the data from each sampler. A geometric mean is used, instead of an arithmetic mean, because it reduces the noticeable effect of a few large values on the annual value. A second measure used, which focuses on single occurrences of poor air quality, is a maximum 24-hour concentration that is not to be exceeded more than once per year. The state of Arizona has established primary standards that are more stringent than the national primary standards (Table 2).

Table 2. Air Quality Standards for TSP

Air Quality Standards ($\mu g/m^3$)	Arizona	Federal Primary	Federal Secondary
24 hour	150	260	150
Annual Geometric Mean	75	75	60

2. A geometric mean is computed by the following equation: $\sqrt[N]{S_1 \cdot S_2 \cdot S_3 \cdot \ldots S_n}$ where N is the number of valid samples (S) taken for a given year.

TSP Sampling Network and Data

Figure 6 and Table 3 show the locations and addresses, respectively, of TSP sampling sites in the Tucson urban area. Some samplers, notably those in the downtown area, South Tucson, The University of Arizona, and at Davis-Monthan Air Force Base, were sited to monitor TSP levels in areas of high activity. Most of the other samplers were located along the boundary of the city-urban fringe to monitor changes in TSP levels in areas undergoing rapid development. From the spatial distribution of these samplers it seems that most of the urban area is being represented and thus a good overall picture of TSP concentrations in Tucson's air is being provided by the existing monitoring network.

The five samplers in operation since 1972 represent varied locations within the study area. While three of the samplers (Health and Welfare Building, University of Arizona, and South Tucson) are closely situated in relatively high density urban areas, the other two (Prince Road and Magnetic Observatory) are situated (in 1972) in areas of rapid surface cover change. If a current classification update is performed, areas around all existing samplers can be monitored and compared to surface conditions in 1972. Ideally, another date midway between 1972 and the present would be used in addition to a current update, since most of the current samplers would have been in operation and more comparisons of surface cover changes to changing TSP levels could be done.

Annual data for all samplers in the study area are provided in Figure 7. No clear-cut trends exist in general for TSP levels in

Figure 6. TSP sampler sites in the Tucson urban area (1980) -- Adapted, with permission, from a map drawn by R. A. Drafting Service (1979).

Table 3. TSP Sampler Sites in the Tucson Urban Area

Station No.	Name	Address
1	Health & Welfare Bldg.	151 West Congress Street
4	Florence Highway & Magee Road	8211 North Oracle Road
6	South Tucson	1810 South 6th Avenue
7	Prince Road Station	1016 West Prince Road
8	Magnetic Observatory	7920 East Tanqúe Verde
10	Davis-Monthan Air Force Base	Civil Engineering Building
12	Border Patrol	1970 West Ajo Way
13	Hughes-Nogales	8100 South Nogales Highway
14	Orange Grove	3401 West Orange Grove
15	Alvernon-Fort Lowell	3915 East Fort Lowell
16	Pima College	2202 West Anklam
17	Golf Links-Harrison	2181 South Harrison Road
19	University of Arizona	Civil Engineering Building

	Health & Welfare Bldg.	University of Arizona	South Tucson	Prince Road	Magnetic Observatory	Davis-Monthan	Border Patrol	Hughes-Nogales	Orange Grove	Alvernon-Ft. Lowell	Pima College	Golf Links	Florence Hwy. - Magee Rd.
1966	106	80	142	177	100								
67	102	89	121	167	88								
68	94*	70	115	132	72								
69	98	78	147	160	92								
70	88	96	149	184	93								
71	107	88	140	229	115								
72	88	98	95	110*	55*								
73	78	67	116	109	47	88	73						44
74	77	69	107	81	47	80	60		76				65
75	67	71	96	81	60	73	79	54	87	174		62	66
76	71	73	118	94*	59	72	76	51	85	168		57	54
77	68	71	119	122	67	76	72	50	85	155	50	63	73
78	56	61	117	110	56	69	66	43	87	126	42	55	50
79	70	89	101	129	61	69	69	54	109	120	47	65	60

*Samplers relocated.

Number of Samplers

Health & Welfare Bldg. 1966--every day
1967-73--weekdays
1974 to present--every other day

Other 1966-1974--every two weeks
1975 to present--every six days

Figure 7. Annual TSP sampler data for sites in the Tucson urban area (geometric means in micrograms/m^3).

Tucson. Most samplers show yearly variations which neither increase nor decrease consistently. Variations, most probably, are tied to changing local conditions around samplers, some too small to be detected by Landsat. It may be shown subsequently that TSP levels at some samplers correlate well with proximate general surface cover changes.

CHAPTER 3

THE CLASSIFICATION OF LANDSAT DIGITAL DATA

Landsat, originally named the Earth Resources Technology Satellite (ERTS), is an unmanned satellite system designed to monitor resource and environmental conditions. Landsat 1 was launched in July 1972 and remained operational until January 1978, and Landsat 2 was operational from January 1975 to January 1980. Landsat 3 was launched in March 1978 and continues in operation to date. As of this date the next Landsat is scheduled to be launched in early 1982. Congressional legislation has assured the continuation of data in Landsat's format well into the 1980's by follow-on satellite systems (Doyle, 1978).

Applicability to Dust Source Monitoring

Certain characteristics of the Landsat system are significant in light of the application to air quality planning in Tucson. First, Landsat has provided repetitive coverage of the entire earth, excluding small areas around the poles, every 18 days, since 1972. Thus, a consistent data base is available for historical studies. Landsat senses the reflectance of the earth's surface in wavelengths that do not penetrate cloud cover. Since Tucson is situated in a region of very low cloud cover, good and frequent coverage of the Tucson area is usually available.

Second, Landsat moves in a sun-synchronous orbit so that every scene is sensed at approximately the same local time, between 9:30 and 10:00 a.m. Uniform illumination, a result of the constant sun angle, therefore, is maintained from scene to scene, facilitating comparison between scenes. Corrections for varying sun elevation are easily implemented for comparison between scenes of different seasons. For the purpose of monitoring changes in areas of high potential dust emissions, Landsat scenes should be chosen near the dame date for each year to minimize seasonal changes in vegetative cover which are highly dependent on seasonal rainfall patterns in Tucson.

Third, each Landsat scene covers an area of 185 by 185 km or 100 by 100 nautical miles. The entire Tucson urban area is included in one Landsat scene (Figure 8) in which the scale and the illumination are essentially constant throughout.

Fourth, the multispectral scanner (MSS) aboard Landsat senses the earth's reflectance simultaneously in four different wavelength intervals, or bands, of the electromagnetic spectrum (the green and red bands of the visible spectrum and two near infrared bands). Reflectance values of ground features vary with wavelength. Thus, the Landsat MSS offers four observations with which to differentiate between ground features.

Fifth, Landsat data are available in digital format. Energy reflected from the ground is sensed and converted into an electrical signal for recording and transmission as digital data. The digital data for all four spectral bands are recorded on computer compatible tapes (CCTs). The digital data can then be manipulated, using

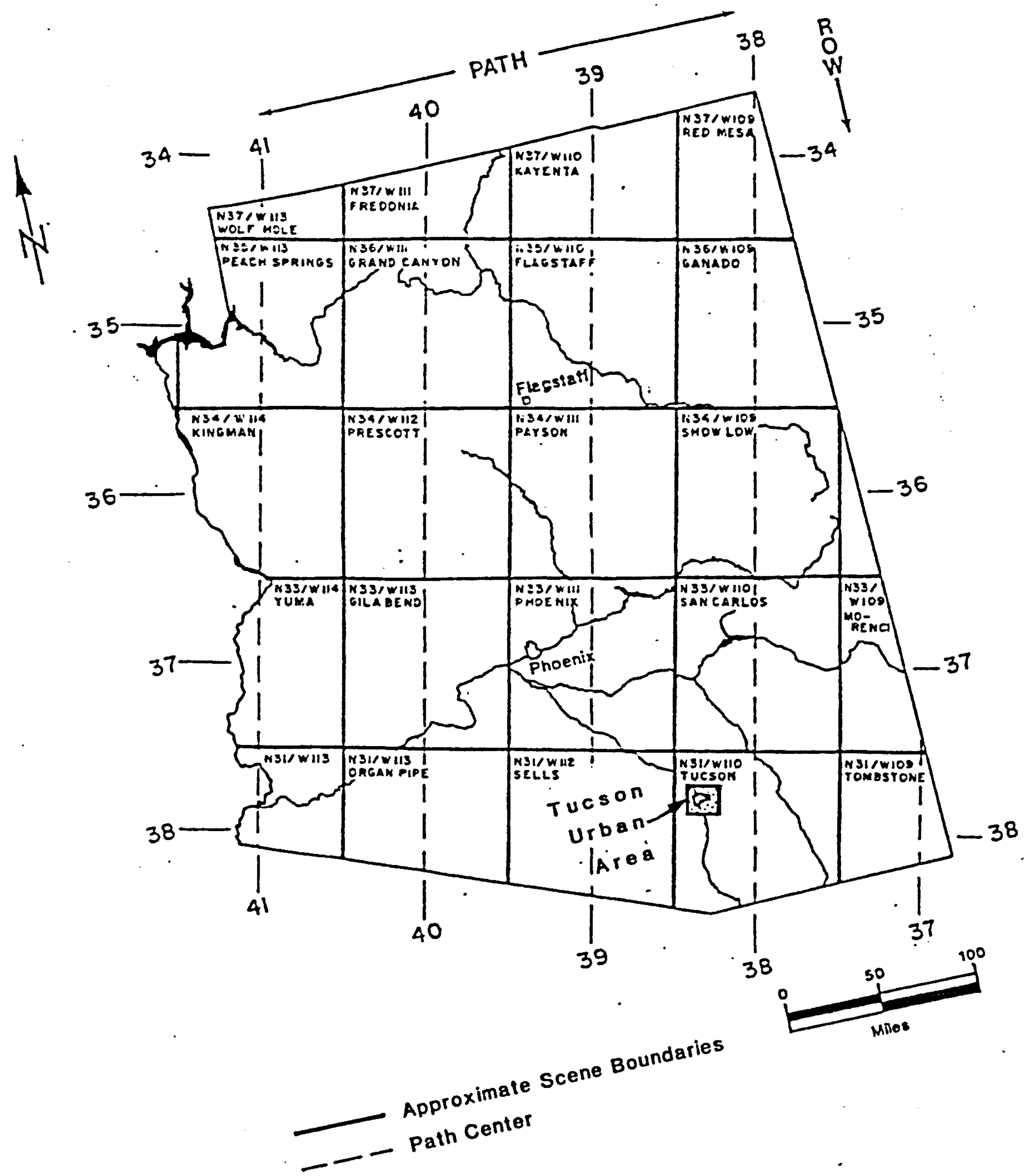


Figure 8. Landsat coverage of Arizona.

computers, for enhancement or classification purposes. Landsat data are also available in an image format. The digital data can be used to produce black and white scenes of each spectral band at scales ranging from less than 1:1,000,000 to 1:250,000. The digital data can also be processed to enhance imagery for various purposes and to extract information not readily detectable from the raw data per se. Color composite images, the equivalents of infrared color photographs, can be produced from the black and white originals.

Sixth, Landsat CCTs and imagery are readily available to the public from the EROS (Earth Resources Observation System) Data Center in Sioux Falls, South Dakota. Computer searches of available CCTs and imagery for a specified area are provided free of charge.

A disadvantage of the Landsat system for this investigation is the 76 by 76 meter, or 1.43 acre, instantaneous field of view of the multispectral scanner. Each data point or pixel (picture element) of the digital image is an average reflectance value from an area roughly equivalent to the instantaneous field of view.[3] Urban areas usually have more heterogeneous surface cover than non-urban areas. Thus, a classification of an urban area using Landsat digital data will involve averaging, or integration, of more diverse surface features within a pixel than would occur, for instance, in a rangeland scene. As a result, lower classification accuracies can be expected within urban areas.

---

3. See Slater (1979) for a discussion of the instantaneous field of view of the Landsat multispectral scanner.

Table 4 provides some additional information on the Landsat system. For a detailed description of orbital parameters, payloads, data processing, data products, applications, and user services see National Aeronautics and Space Administration (NASA) (1976).

Spectral Data and Observation Space

The spectral response of each pixel or picture element of a Landsat scene is sensed and measured in four different portions of the electromagnetic spectrum. Incident radiation may be absorbed, emitted, reflected, scattered, or transmitted through a ground feature. The area on the ground corresponding to a pixel in a Landsat scene may be identified by the way in which the features within that area interact with incident radiation. The spectral response of a feature, then, is a product of this interaction and is quantified by Landsat into brightness levels, since radiant energy in the portion of the spectrum that Landsat senses is primarily reflected. These brightness levels within the four Landsat spectral bands may be used to identify a feature and are collectively called its spectral signature.

General spectral curves are shown in Figure 9 for vigorous vegetation, bare soil, and water based on their spectral signatures in the four Landsat bands. Vigorous vegetation has a very high reflectance in the infrared bands, usually on the order of three times its reflectance in the visible spectrum. Bare soil that does not have a distinctive color, but does have a uniform level of dryness, usually exhibits a flat spectral curve with only a slightly higher response in the infrared bands. Water, in contrast, has its highest response in

Table 4. Landsat System Information

Orbit (near polar, sun synchronous)

Altitude--570 miles (917 km) nearly circular

Inclination--81° (relative to a plane passing through the equator)

Orbit Period--103 minutes

Orbits/Day--14 (crossing the equator at 9:30 a.m. local time)

Successive Orbits--Shift westward 1,785 miles (2,875 km) at the equator

Adjacent Orbits--1st and 15th (Orbit 1--Day 1, Orbit 1--Day 2), 2nd and 16th, etc.

--orbit on one day displaced westward 99 miles (59 km) from corresponding orbit on previous day

--sidelap--15% (16 miles--25 km) at equator
57% at 60° latitudes
85% near the poles

Scene width--115 miles (185 km, 100 nautical miles)

Coverage Cycle--18 days (entire earth, excluding areas near the poles)

Multispectral Scanner Data

Pixels (Picture Elements)--a digital image data point whose reflectance value is an average value for a 76 by 76 m (1.43 acre) area on the ground

Spectral Bands

Landsat Band	Spectral Region	Color	Brightness Levels
4	.5-.6 μm	green	0-127
5	.6-.7 μm	red	0-127
6	.7-.8 μm	infrared	0-127
7	.8-1.1 μm	infrared	0-63

Adapted from Short et al. (1976).

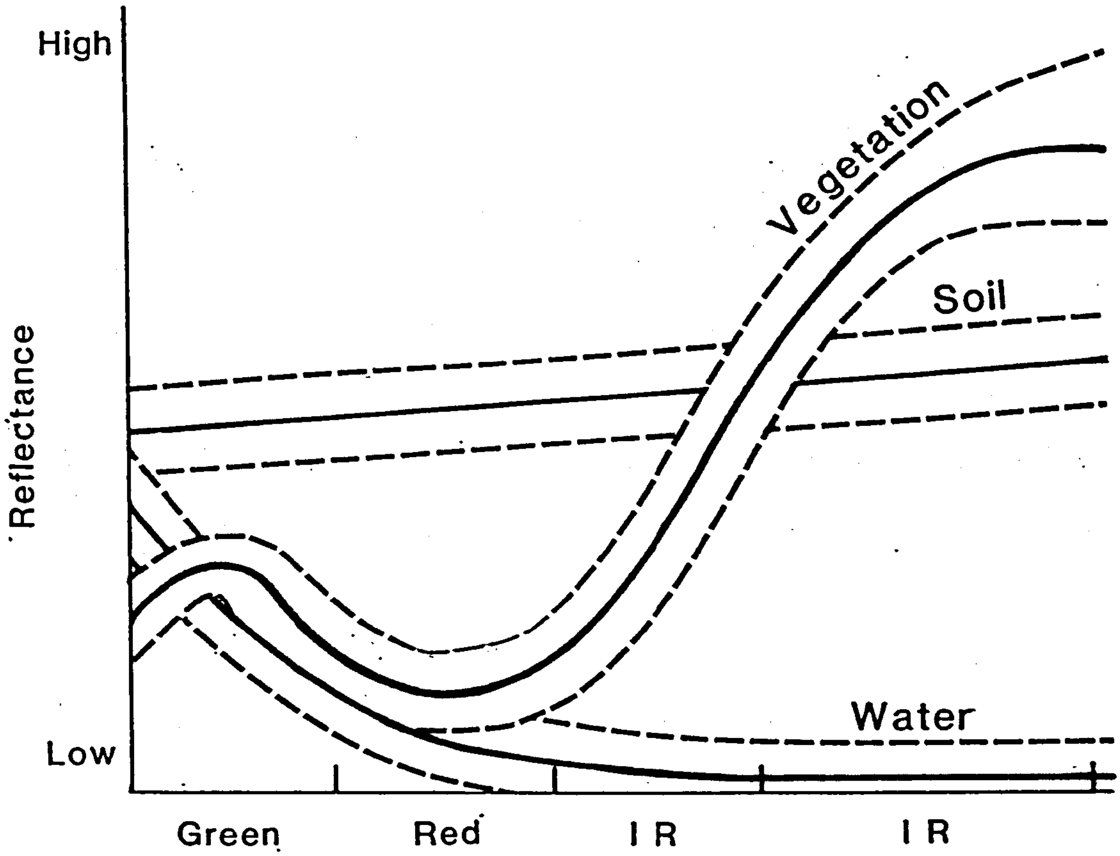


Figure 9. Spectral responses of water, soil, and vegetation.

the shortest wavelengths and an increasingly lower response corresponding with increasingly longer wavelengths toward the infrared bands.

If the spectral responses for vigorous vegetation, bare soil, and water in two bands are plotted against each other, the spectral signatures may be represented in 2-dimensional observation space as shown in Figure 10. Because of natural variation in the spectral response of any class of features, the spectral curve of each is actually a fluctuating variable within some range, denoted by the dotted lines in Figure 9. Due to this variability, the spectral response of like ground features will be represented in observation space by a cluster of points, or pattern, rather than a single point.

The advantage of having the spectral responses of features in more than one part of the spectrum is shown in Figure 10. If ground features were to be grouped into either vigorous vegetation, bare soil, or water classes on the basis of their spectral response in only one band, confusion might exist since the ranges of their responses may overlap. With information from two bands the spectral separation of features into these classes in 2-dimensional observation space is greatly facilitated. With Landsat data the observation space is 4-dimensional providing even a greater possibility of class separation. In real world situations, however, the spectral response of most classes of features overlap in even the 4-dimensional observation space of Landsat data. Thus the task of class separation is not an easy one and requires statistical techniques which optimize the spectral differences between classes.

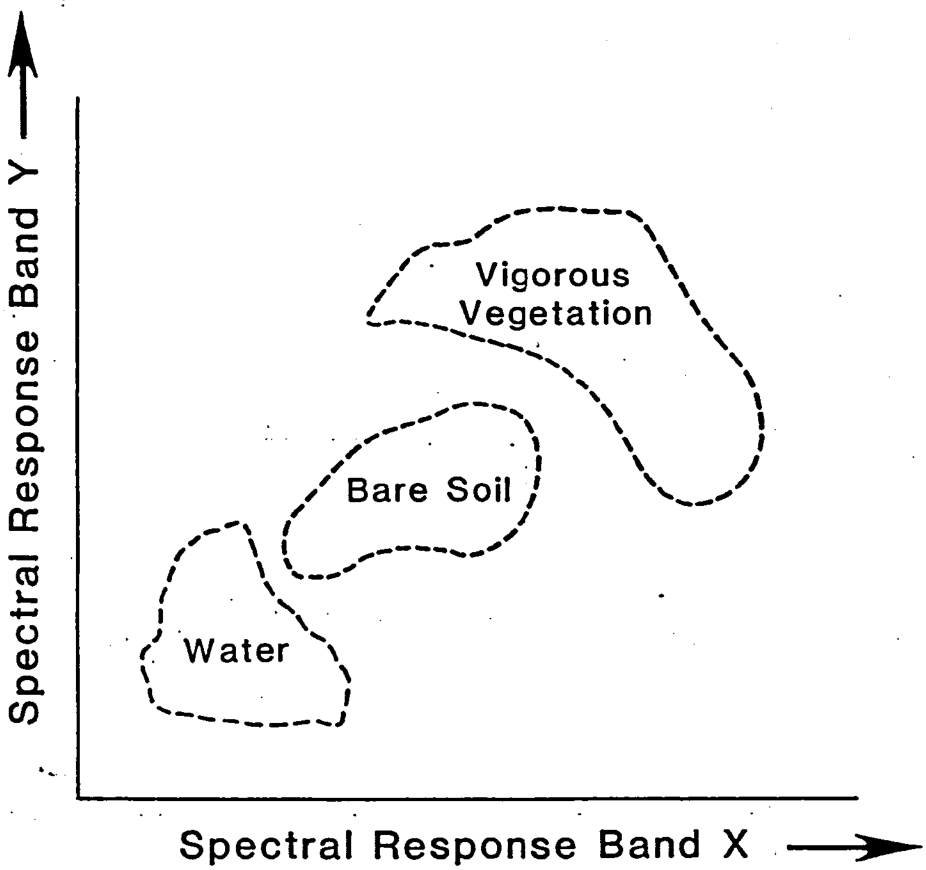


Figure 10. Spectral responses of water, soil, and vegetation in two-dimensional observation space.

Logic and Purpose of Classification

Classification is the process by which objects are grouped into classes on the basis of properties or relationships they have in common. The resulting system of classes can be derived by either inductively classifying individuals into groups based on differentiating characteristics or deductively by logical division, in which the universe of individuals is divided according to some principle (Grigg, 1965). As Grigg (1965) points out, the latter method is valid only if an a priori understanding of the cause of differences, i.e., the principle, is full and complete. No such assumption is made when individuals are judged to be similar and grouped together a posteriori. Still, an advantage of logical division is that the detail of the system can be defined to meet the needs of the immediate problem (Nunnally and Witmer, 1970). In most practical applications both methods are used to reach the final classification system (Grigg, 1965; Sokal, 1974; Witmer, 1977).

The purpose of classification is to give order to the objects studied (Grigg, 1965). A terminology must be developed so that these objects can be identified and placed into classes. The most important result of classification is that inductive generalizations can be made about the objects studied (Grigg, 1965). As Sokal (1974, p. 1116) states,

> The paramount purpose of a classification is to describe the structure and relationship of the constituent objects, and to simplify these relationships in such a way that general statements can be made about classes of objects.

Thus, classification is an integral step in the scientific method. Order must be established so that inductive generalizations can be proposed about the objects classified. In this case the objects grouped into classes of varying dust emission potential are areas of different surface cover. The differentiating characteristic used to separate different surface cover is spectral response in the four Landsat bands. The hypotheses to be tested subsequent to this project will involve the relationship between dust emission potential from the classes of surface cover defined here to dust levels measured by samplers throughout the study area. The subsequent hypotheses must necessarily depend on the accuracy of the classification and the validity of the methods used to obtain the classification.

Approaches to Classification of Landsat Digital Data

The role of the computer in modern classification has been paramount. New techniques, designed to solve problems that at one time seemed to be analytically intractable, have been developed as a result of the computer's great analytical capacity. Solutions that are exceedingly tedious to carry out by hand are easily handled by modern computers. A most important by-product of the application of computers to the problem of classification is the development of new algorithms, which in turn has led to increased consistency and optimization of the classification process (Sokal, 1974).

The mass of digital data contained in a Landsat scene (seven million pixels per band or twenty-eight million pixels total) necessitates the use of a computer in the classification process. An algorithm must be chosen which establishes the rules by which the objects are classified. The chosen algorithm, known as the "classifier," must be "trained" so that the differentiating characteristics, the criteria by which the objects are grouped, can be established.

A Landsat classification algorithm may be trained by either a supervised or unsupervised approach or by a modification thereof. In the supervised approach, the user locates through ground checking, map data, or aerial photography, sites within the study area that are representative of each class to be included in the classification. These sites must be as homogeneous as possible while still being sufficiently typical of the whole class in question. Homogeneity is important so that each class is as spectrally distinct as possible. The entire spectral range of each class must be included in its training data to ensure that all features within the training data of a class can be included in that class. These training sites are then located within the scene so that the classifier can use the spectral information of these sites to group pixels into the appropriate classes.

In the unsupervised approach, the "training" is achieved through statistical methods; commonly through cluster analysis. In cluster analysis, a sample of pixels is examined to see where natural clusters of data occur (Figure 11). With very small study areas, the clustering process may be performed on the entire area, cost permitting. In effect, then, the training and classification is combined into one step (Swain, 1978). The probability density function of the sampled spectral data is used to distinguish the modes or clusters of data. The number of clusters, the minimum number of pixels per cluster, the maximum

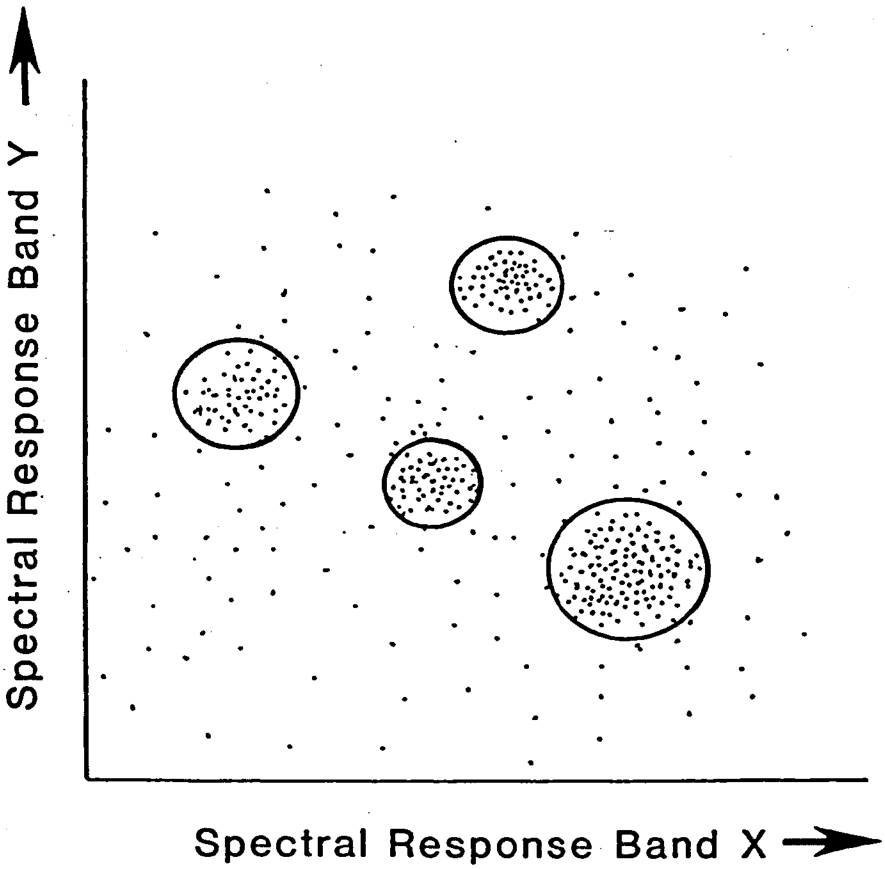


Figure 11. Clusters in observation space.

allowable variance within a cluster, and the distance (in observation space) between clusters can all be manipulated by the user. Certain clusters of data may be combined, divided, or deleted until the number of clusters equals a predetermined number set by the user. The user then examines the classification map to see how the clusters in observation space correspond to classes of objects on the ground. The user may then decide to do more iterations of the clustering process until the clusters in observation space correspond to meaningful classes on the ground.

Each approach has its advantages. The supervised approach allows the user to specify only those classes that are important for the purpose of the classification. The unsupervised approach is more objective, relying on statistical analysis to group the data into classes. The unsupervised approach can still be used when good ground data, necessary to determine training sites in the supervised approach, is not available. However, in the unsupervised approach there is sometimes so much variability within a scene that the task of finding the correct number and size of clusters to correspond to meaningful classes on the ground is confusing. Also when ground classes are only marginally separable in observation space, the unsupervised approach is generally not as effective as the supervised approach in separating meaningful, ground classes because of the analyst's reduced control (Swain, 1978).

A "modified" unsupervised, or "guided" cluster analysis approach has also been used in Landsat classifications. This approach combines the advantages of allowing some control in the selection of training sites for specified classes, characteristic of the supervised approach, with the advantage of a significant amount of statistical objectivity, inherent in the unsupervised approach (Gaydos and Newland, 1978; Hutchinson, 1978). In this approach a section of the study area containing all of the variability of the desired classes is delineated and cluster analysis is performed only on this training area. Unwanted spectral variability within the scene is not used by the classifier in this case, and as a result confusion between classes should be decreased. As in the other training approaches, these data are then supplied to a classifier to determine how the rest of the study area is to be grouped.

Classifiers use either probabilistic or deterministic rules to divide observation space (Steiner and Salerno, 1975). Probabilistic classifiers use rules based on the probability distribution, commonly assumed to be a gaussian distribution. Deterministic methods, conversely, do not use probability concepts and thus produce a unique answer in terms of categorization.

An example of a deterministic method is the parallelepiped classifier, in which observation space is divided solely on the limits of the ranges of reflectance values in each band (Figure 12). With this approach all pixels are not classified and, if there is much overlap between classes, an excessive number of classes may be generated which may have little meaning on the ground (Schlien and Goodenough, 1974).

Another slightly more sophisticated deterministic approach is the minimum-distance-to-means (or centroids in observation space)

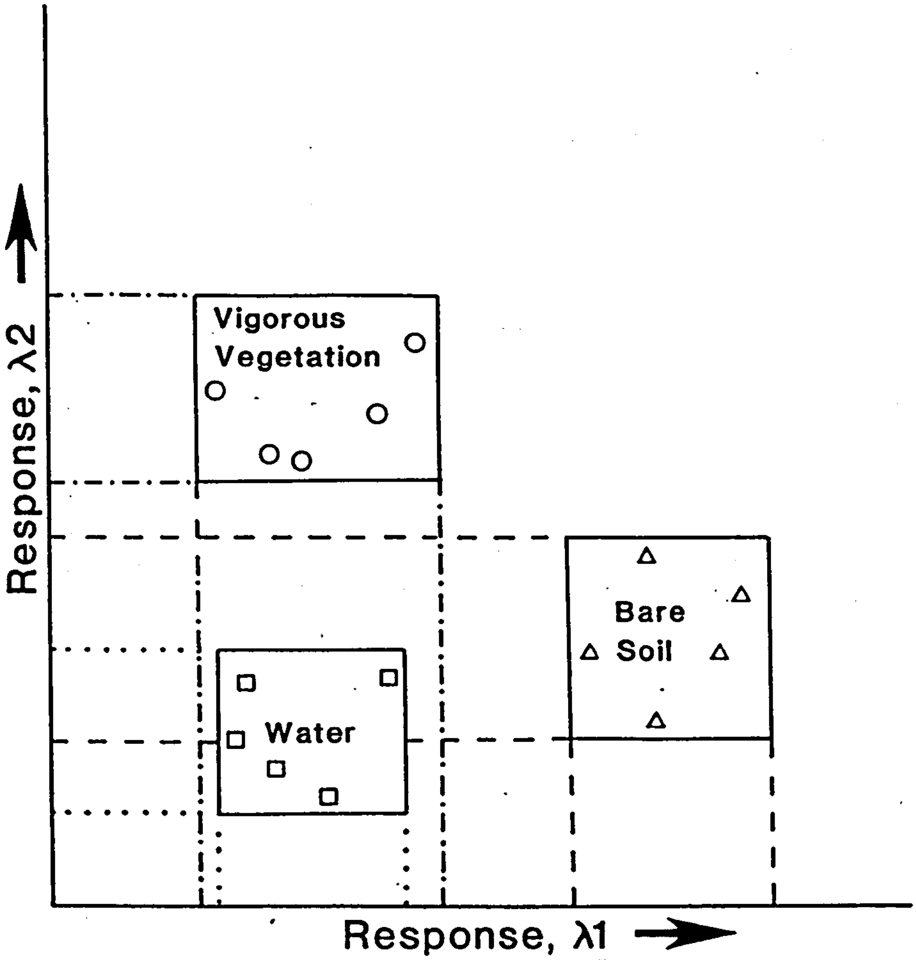


Figure 12. Parallelepiped classifier.

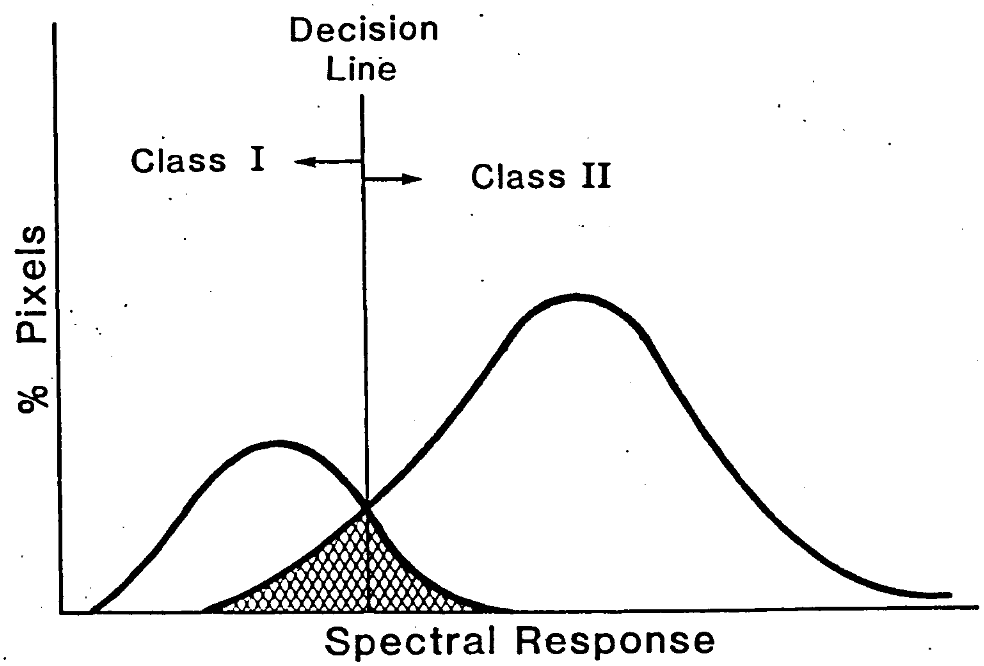
classifier. With this classifier, all pixels can be classified and greater accuracy can be expected than with a parallelepiped classifier (Wacker and Landgrebe, 1972).

A probabilistic approach which is widely used and accepted is the maximum-likelihood classifier. Because it makes use of probability theory, the parameters of each class must be estimated from the training data. Assuming gaussian distributions for the class populations, the means and covariance matrices are estimated from the training data. For each pixel in the study area the probability of membership is calculated for each class. According to the Bayes optimal decision rule (Swain, 1978; Steiner and Salerno, 1975), each pixel is then assigned to the class for which it has the highest probability of belonging. The gaussian assumption for class populations may be moderately violated without severely affecting classification accuracy, except in situations where the training data are clearly bimodal (Swain, 1978; Schlien and Goodenough, 1974).

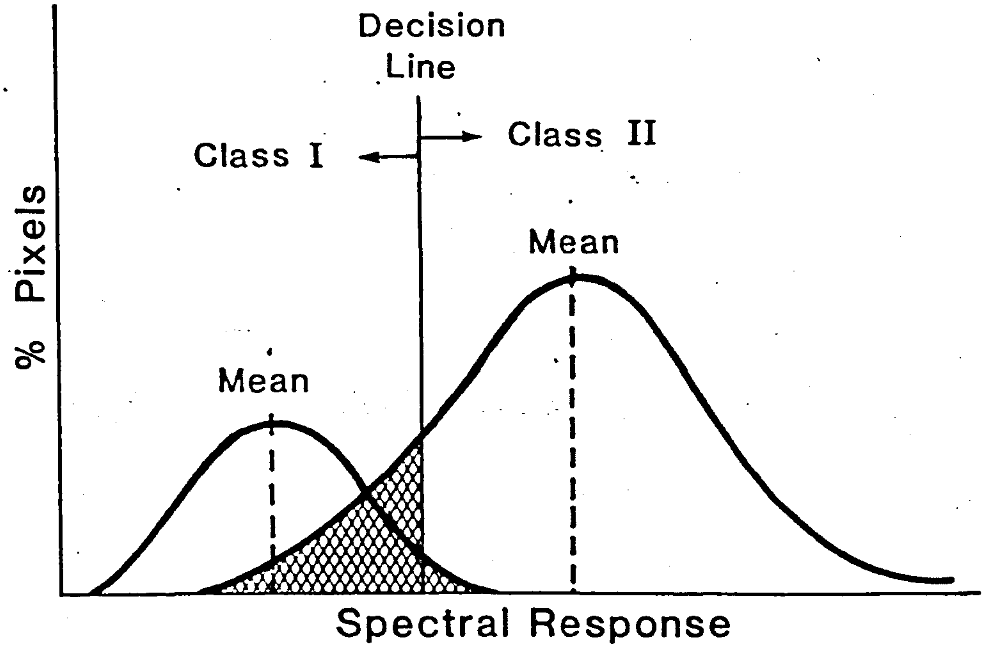
Probabilistic classifiers may be preferred over deterministic methods for three reasons (Swain, 1978). First, statistical analysis helps to account for the natural variations of spectral responses that obscure differences between classes that could otherwise reduce classification accuracy. Second, statistical methods are tolerant of errors resulting from uncertainty as to the true identity of training data, as long as the frequency of error is relatively low. For example, a training site for corn may include small areas of weeds and bare soil that would confuse deterministic classifiers. Third, when spectral responses overlap in observation space and parts of two classes are indistinguishable, statistical methods allow for class assignments that are most probably correct.

Figure 13 illustrates where the decision is made to assign a pixel to a class (in one-dimensional observation space) and what the error will be with the maximum-likelihood and minimum-distance-to-means classifiers. In the former case (Figure 13a), the Bayes optimal decision rule is used, which theoretically minimizes the amount of error resulting from confusion when two ranges of spectral response overlap. The decision to assign pixels to either Class I or II is made at the point of intersection between the two probability density curves. The amount of error resulting from this maximum-likelihood decision is illustrated by the cross-hatched area.

In Figure 13b, the decision to assign pixels to either Class I or II, with the minimum-distance-to-means classifier, is made midway between the means of each class distribution. Again, the amount of resulting error is illustrated by the cross-hatched area. As one can see, the amount of additional error is a function of the distance of the decision line from the intersection of the probability density curves, or the distance from where the Bayes optimal decision is made. The Bayes optimal decision therefore minimizes the error of misclassification.



(a) Maximum-likelihood classifier



(b) Minimum-distance-to-means classifier

Figure 13. Decision rules for two classifiers.

CHAPTER 4

SURFACE COVER CLASSES IN CLASSIFICATIONS OF LANDSAT DIGITAL DATA

For the classes used in the Landsat classification of the Tucson urban area, two criteria had to be met. First, each class had to be spectrally separable from the other classes in order for the classification to work. Second, for the purpose of the classification, each class had to represent areas that were fairly uniform throughout in dust emission potential so that they could be ranked for further analysis. Since there is no way of assuring the uniformity of potential dust emissions within general surface cover classes, the validity of their ranking may be open to question. Nevertheless, through interaction with local air quality planners, the classes chosen and their theoretical potential for dust emissions were seen to offer promise of utility for long term air quality planning purposes in Tucson (Heidel, 1979; Buchholz, 1980). If some classes described below are grouped together, valid and accurate distinctions could be then made between the dust mission potential of the resulting classes.

In a study done by Bach (1971) in 1969 atmospheric turbidity was measured over different land uses in the Cincinnati urban area. Bach used a Volz sun photometer to measure the depletion of solar intensity, the turbidity coefficient, which can be related to the amount of particulate matter present in the atmosphere. According to Bach's measurements, TSP mass concentrations were worst over industrial areas.

They improved progressively from the city center to residential areas and suburbs and were best over city parks. He suggested that "green spaces" (vegetated open space, parks, or gardens) should be placed around industrial and residential areas because of their filtering effect on dust, soot, and fly ash. "Green spaces" of this kind would have a very low potential for dust emissions both because they do not emit pollutants and because they act as air filters.

Previous Work on Classification of Landsat Digital Data in Urban Areas

A land use and land cover classification system has been designed for use with remote sensing data (Anderson et al., 1976). This system was created in response to the need for standardization of land use and land cover classes throughout the United States, in order to reduce duplication of effort and to make possible the aggregation of available data. Classes in Levels I and II (Table 5) were devised as standard classes that could be delineated using Landsat data (the principal source for Level I classes) and color infrared photography at a scale of 1:80,000 or less (the principal source for Level II classes). There is a built-in flexibility with these general classes in that classes of Levels III and IV can be designed to meet more specific needs.

The Central Atlantic Regional Ecological Test Site (CARETS) Project was one of several developed to test and evaluate this classification system. In one part of the CARETS Project, the utility of land use and land cover information derived from this system was analyzed with respect to air quality planning needs (Reed and Lewis,

Table 5. Land Use and Land Cover Classification System for Use with Remote Sensor Data

Level I	Level II
1. Urban or Built-up Land	11. Residential
	12. Commercial and Services.
	13. Industrial
	14. Transportation, Communications, and Utilities
	15. Industrial and Commercial Complexes
	16. Mixed Urban or Built-up Land
	17. Other Urban or Built-up Land
2. Agricultural Land	21. Cropland and Pasture
	22. Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas
	23. Confined Feeding Operations
	24. Other Agricultural Land
3. Rangeland	31. Herbaceous Rangeland
	32. Shrub and Brush Rangeland
	33. Mixed Rangeland
4. Forest Land	41. Deciduous Forest Land
	42. Evergreen Forest Land
	43. Mixed Forest Land
5. Water	51. Streams and Canals
	52. Lakes
	53. Reservoirs
6. Wetland	61. Forested Wetland
	62. Nonforested Wetland
7. Barren Land	71. Dry Salt Flats
	72. Beaches
	73. Sandy Areas other than Beaches
	74. Bare Exposed Rock
	75. Strip Mines, Quarries, and Gravel Pits
	76. Transitional Areas
	77. Mixed Barren Land
8. Tundra	81. Shrub and Brush Tundra
	82. Herbaceous Tundra
	83. Bare Ground Tundra
	84. Wet Tundra
	85. Mixed Tundra

Table 5.--Continued

Level I	Level II
9. Perennial Snow or Ice	91. Perennial Snowfields
	92. Glaciers

From Anderson et al. (1976).

1975). It was proposed that EPA use information from this system to monitor the location and changes in major point and area emissions sources for various pollutants. Available land use information for the study area had differing accuracy, scales, and dates. Considerable time and cost was required to compile this information at a uniform scale and in categories suitable for air quality planning. Using the classification system developed by Anderson et al. (1976), high levels of TSP were found to correspond to high concentrations of industrial, commercial, and old, dense residential land uses. Lower emissions were related to low density residential, commercial, transportation, and industrial land uses and areas of water and agricultural activities. However these specific classes would be of little utility to air quality planning in arid and semi-arid urban areas where dust resulting from ground disturbance is the main source of TSP (EPA, 1977).

Most work on Landsat classifications in urban areas is an outgrowth of work performed by the Geography Program of the U. S. Geological Survey on research and development of the "Anderson" classification system. Digital Landset classifications are preferred over image interpretation, in most cases, because they allow the classification of surface features whose spectral differences are not discernible in the imagery and because objects to be grouped can theoretically be as small as individual pixels. Unsupervised training, together with a maximum-likelihood classifier, is used in many cases in attempts to distinguish both Level I and Level II classes. Dornbach and McKain (1973) had poor results in a supervised approach to an urban classification using Landsat digital data because of the heterogeneity of their classes. In addition, the integration of the reflectance values of many diverse surface features within a pixel area, the "mixed picture element" problem, made for poor training sites. In general, the unsupervised approach is judged optimal in specialized classifications of small selected areas, i.e., urban areas, and the supervised approach is preferred in large areas that are predominantly rural (Joyce, 1974).

One reason for suggesting alternate approaches for urban and rural area classifications is that within a class spectral signatures of natural features vary more than cultural features. This user control (supervised approach) would be to greater advantage in urban area classifications where spectral signatures are more homogeneous within a class of features (Hutchinson, 1978). Of course, the variability in natural feature classes may justify using the unsupervised approach in non-urban areas but the integration of diverse surface features within a pixel in urban areas adds variability to class signatures that can make supervised training difficult.

Problems may be encountered in separating urban from non-urban areas with Landsat digital data, because of the propensity for highly vegetated urban areas to be spectrally confused with cropland and pasture land and other vegetated non-woody areas. For improved results in these cases, a stratification of the study area into built-up and rural areas and separate classifications for each are necessary (Dornbach and McKain, 1973; Ellefsen, Swain, and Wray, 1973; Rohde, 1978). Some Level II classes (residential) specified by Anderson et al. (1976) are classified regularly with high accuracy in urban areas, but others (12--commercial, 13--industry, and sometimes 14--transportation; see Table 5) are spectrally indistinguishable (Dornbach and McKain, 1973; Gaydos and Newland, 1978; Mausel, Todd, and Baumgardner, 1974).

In certain studies Level III classes, specifically new and old residential or suburban and inner city, are said to be spectrally separable using Landsat digital data (Dornbach and McKain, 1973; Mausel et al., 1974). But this is more a function of the amount of vigorous vegetation present (lawns, trees, and shrubs) rather than of unchanging characteristics of these areas. In the Pacific Northwest Land Resource Inventory Demonstration Project (LRIDP) (1979), distinctions were successfully made between high, medium, and low land use intensity, the equivalents to Level II classes. As Ellefsen (1974) points out, reflectance is the raw material used by the classifier in computer-aided classifications of Landsat digital data, and there are none of the common references used by the photointerpreter (size, shape, shadow, tone, texture, pattern, position) with which to infer functional distinctions such as retail activity, education, wholesale trade, or transportation.

Observations have been made on urban classes that are more easily separated in arid and semi-arid areas. Ellefsen et al. (1974) reported that in work on classification of the Phoenix area, non-vegetated vacant lots were correctly identified in an area thought to be a built-up string of commercial land use. This identification was probably due to the high reflectance of these lots in contrast to their surroundings, which is a result of a lack of weeds and grasses that would quickly revegetate vacant lots in more humid areas. Gaydos and

Newland (1978) point out that parks and golf courses are easier to separate from pasture and grassland in arid and semi-arid areas where vegetation in non-agricultural rural areas usually has a lower per cent cover than in humid areas.

Class Descriptions

In initial classifications, confusion was found to exist between certain classes that are essentially limited to the urban built-up areas and classes of similar reflectance found solely in the outlying, virtually undeveloped areas. The study area was stratified into built-up and outlying areas so that separate classifications could be performed thereby eliminating this unnecessary spectral confusion and reduced classification accuracy. The following descriptions of classes from both subareas are intended to summarize the nature of the training sites selected for each class and those areas that were grouped into each class by the maximum-likelihood classifier. The descriptions of these ground features were derived from photointerpretation of color infrared photography at a scale of 1:120,000 and orthophotoquads at a scale of 1:24,000. The color infrared photography is from the same month as the Landsat CCT (August 1972) and the orthophotoquads are derived from high-altitude photography of Tucson taken in May and November 1972.

Built-Up Area

Disturbed-Vacant Land. Areas in this class have little or no vegetative cover and are disturbed often enough so that little surficial crust is present. The reflectance of the training sites for this class is high throughout the four Landsat bands.[4] When the ground is disturbed the resultant decrease in particle size generally results in an increased reflectance. This is due to greater light scattering, with more surface area exposed, and the reduction in microshadows occurring between particles under oblique illumination. Conversely, undisturbed areas, where a surface crust or desert pavement occurs, will have lower reflectance values (Orr, n.d.). Dust emission potential for these areas is very high, a result of a surface readily subject to deflation and frequently subject to mechanical disturbance and the reentrainment of dust.

Although dust emission potential is related to the silt content of soil (Cowherd et al., 1974) for this class, differences in dust emission potential resulting from the extent to which an area is disturbed tend to override soil texture factors. An area is determined to be disturbed solely on the basis of its reflectance on the Landsat CCT and photography used. Many areas included in this class are actively disturbed throughout the year, but others such as construction sites may have high dust emission potential only for part of the year. Since the extent to which an area has been disturbed cannot be determined from the CCT and photography, the variation in dust emission potential from these source areas is considered to be greater than variations attributed to differing silt content. Finally, soil data of the

---

4. See the first four columns in Table 6 to compare the mean reflectance values of the training sites for all classes.

Table 6. Mean Reflectance Values of Training Data for All Classes[a]

	Band 4	Band 5	Band 6	Band 7	7/5	5/4
<u>Built-Up Areas</u>						
Disturbed-Vacant	70	86	85	72	34	106
Developed-Disturbed	57	63	64	56	37	95
Native Vegetation (Low Cover)	54	67	67	60	38	106
Paved	37	38	31	27	28	87
Developed	48	50	58	56	49	88
Vigorous Vegetation	35	29	69	80	130	70
<u>Outlying Areas</u>						
Disturbed-Vacant	65	82	78	67	33	108
Developed-Disturbed	58	65	68	60	38	97
Native Vegetation (Low Cover)	52	63	63	57	38	103
Native Vegetation (Medium Cover)	46	50	53	49	42	92
Native Vegetation (High Cover)	40	42	52	55	59	89
Basalt	34	33	34	33	44	82
Vigorous Vegetation	34	28	68	79	134	67

[a]Band 7 values were multiplied by two so the range of brightness values corresponds to that of the other bands. Values for both ratios were linearly stretched to allow increased differentiation between classes.

necessary detail are not available. Thus, soil texture is not considered in developing dust emission potential classes.

Examples of land parcels comprising this class include vacant lots disturbed by human activity, including playgrounds, areas disturbed by off road vehicles, areas cleared for construction and areas disturbed by automobiles, e.g., unpaved parking lots.

Developed-Disturbed Land. This is a heterogeneous class of diverse cultural features interspersed with small areas of disturbed-vacant land too small to be included in Class 1 because of the "mixed picture element" problem. Reflectance values of training sites for this class are medium to medium high and are a product of the integration of the reflectance of a wide range of ground features that are apparently dominated by high reflectance from the disturbed-vacant land. Potential dust emissions for areas in this class are very high for the same reasons given for Class I. Also areas of unpaved roads with high dust emissions (Heidel, Dixon, and Young, 1978; McCaldin, 1977; Cowherd et al., 1974) are included in this class. Examples of areas classified as developed-disturbed land are industrial and established residential areas with a considerable number of small disturbed-vacant lots and unpaved roads. Also included in this class are areas of new construction consisting of a mixture of disturbed land and buildings, roads, vegetation, etc.

This class is not ideal to use because of the heterogeneity of its surface cover. Through consultation with local air quality planners (Heidel, 1979) it was decided that an attempt should be made to include areas like these in a class because of their high dust potential. The determination of training sites for this class took considerable time and effort before an area was found that had a fairly regular pattern of disturbed land interspersed with cultural features. The best training data came from new construction sites where the streets were paved and where structures had not yet been erected, so that the high reflectance of the cleared areas was regularly integrated with the streets' low reflectance.

Because the surface cover of the class is comprised of surface cover from several other classes, there is the potential for interclass spectral confusion. However, from inspection and testing of the final classification minimal misclassification of this class was observed. The most recurring confusion was with the native vegetation--low per cent cover class. Because the dust emission potential for these classes are similar they could be combined into one class with the second highest dust emission potential, for further analysis.

Native Vegetation--Low Per Cent Cover. These areas are basically undisturbed lots with a low per cent vegetative cover; usually creosotebush dominated vegetation. The reflectance of training sites for this class is medium high since the spectral response in such areas of low cover, 8-15% (Steenbergh and Warren, 1977), is largely from the light desert soil. A high spectral response was not obtained in the near infrared Landsat bands because of the low vegetative cover. Dust emission potential for areas in this class is less than for Classes 1 and 2 because these are not extensively disturbed and a protective vegetative cover is present to inhibit deflation. From close examination of the verification photography, some areas assigned to this class include either unpaved roads or small disturbed areas. Therefore this class is ranked above subsequently described classes in dust emission potential. Examples of areas classified as native vegetation--low cover are small lots in the city that have small amounts of disturbed areas but that generally have a continuous cover of vegetation.

Paved Areas. Areas in this class are dark asphalt streets and parking areas. The reflectance for the paved training sites is very low in all four Landsat bands. Dust emission potential for this class of surface cover is lower than for Classes 1, 2, and 3 yet should be higher than either developed land or vigorous vegetation. The source of emissions in paved areas is reentrained dust. Paved areas are high activity areas and any dust, dirt, or gravel found in these areas may be injected into the air by mechanical means such as reentrainment from cars. Examples of areas classified as paved are major transportation routes with adjacent parking lots and large parking lots for shopping malls.

Developed Land. Areas included in this class are generally newer established residential areas with little disturbed-vacant land. In addition to areas of medium to high density housing, including grass lawns, trees, and paved streets, some industrial and commercial areas lacking disturbed vacant land are also included in this class. Reflectance values of training data for this class are medium to low for all four Landsat bands. Dust emission potential should not be very high for this class. However, small portions of areas included in this class may have a high potential for dust emissions because of the high level of human activity, the presence of small but undetected disturbed areas and the reentrainment of dust from streets in residential areas. Care was taken to limit this class to areas where no disturbance could be detected on the verification photography and assign all other developed areas to the developed-disturbed class.

Vigorous Vegetation. This class consists of highly vegetated areas of grasses, shrubs, and trees not native to the Tucson area. Reflectance values for the vigorous vegetation training data are, as expected, low in the visible bands and high in the near infrared bands. Dust emission potential for this class is very low. Even the disturbed areas within this class, e.g., baseball diamonds, sand traps, and gravel parking lots in parks, are small and adjacent trees and grasses act to filter dust from the air (Bach, 1971). Areas included in this class are mostly parks, golf courses, and playing fields adjoining educational facilities.

Outlying Area

Disturbed-Vacant Land. This class is essentially identical to Class 1 in the built-up area classification. Almost all places assigned to this class in classification of the outlying area are cleared construction sites and are generally larger than similar sites in the built-up area.

Developed-Disturbed Land. This class is virtually the same as Class 2 in the built-up area classification. Examples of areas that are assigned to this class in the outlying area classification are almost solely construction sites where some building or paving of streets has occurred. Reflectance values for this class are only slightly higher than for Class 2 in the built-up area. This is due to integration of the reflectance values of medium cover native vegetation in outlying area rather than integration of the reflectance values of paved streets, buildings, and non-native vegetation in the built-up area.

Native Vegetation--Low Per Cent Cover. This class is identical to Class 3 in the built-up area classification. Mean reflectance values of the training data for this class in both classifications are very similar. Examples of areas included in this class are mainly flat low-lying areas to the south of the built-up area.

Native Vegetation--Medium Per Cent Cover. Areas included in this class are basically undisturbed areas of medium density cover located in the foothills surrounding the built-up area. Reflectance of the training data used for this class is uniformly medium for all four Landsat bands. The vegetative cover, approximately 22-32% (Steenbergh and Warren, 1977), is not great enough to give a high response in the near infrared bands. Dust emission potential for this class is lower than in Class 3 as a result of the higher vegetative cover and coarser soils characteristic of foothill areas. Small disturbed areas may be included in this class if not large enough to be assigned to Class 1, but they are of minimal area and of little consequence.

Native Vegetation--High Per Cent Cover. Areas of riparian vegetation are the main component of this class. Training data reflectance is uniformly medium-low in all four bands for this class and is not very different from the training data reflectance for Class 4. The medium response in the infrared bands for this class shows a lack of vegetation detectable by Landsat. The reason for the low infrared response may be that much of the riparian vegetation or vegetation in drainageways is tall trees and shrubs that cast shadows on the surrounding ground and vegetation, resulting in lower reflectance values in all bands. Dust emission potential for areas assigned to this class is fairly low. Whether or not it is lower for this class than the native vegetation medium cover class (Class 4) is difficult to say. Since riparian vegetation is situated in and around dry washes, the dust potential for these areas may be higher than for Class 4. In any case this class could be aggregated with Class 4 and possibly Class 6--Basalt to form a more general medium to low potential dust emission class. For the most part native vegetation--high per cent cover areas are restricted to drainage areas.

Basalt. This class was created after a classification of the entire study area was performed in which areas of volcanic material, mainly to the west of the city in the Tucson Mountains, tended to be confused with paved areas in the city. Reflectance of training data for this class is uniformly low, only slightly higher than for the paved class in the built-up area classification. If a ranking by single class is desired in further analysis of this classification data it is recommended that this class be combined with Class 4 (Native vegetation--medium cover) and perhaps with Class 5. The important point to stress is that all three classes have low potential for dust emission.

Vigorous Vegetation. This class is identical to Class 6 in the built-up area classification. Areas assigned to this class are virtually all golf courses, since few schools and irrigated parks are situated in the outlying areas.

Cropped fields could be assigned to any one of a number of the above mentioned classes in the classification, depending on the stage of growth on the date of the Landsat CCT. In order to avoid problems of constant and extreme changes in dust emission potential inherent in agricultural production, farmed fields were delineated on the classification map through interpretation of color infrared photography and orthophotoquads.

CHAPTER 5

METHODOLOGY AND RESULTS

The computer compatible tape (CCT) used for this project was generated from data acquired by Landsat on August 22, 1972 for the Tucson area (Scene ID 8103017271500). This CCT was chosen for several reasons. First, it was available from the Applied Remote Sensing Program (ARSP) of the Office of Arid Lands Studies at The University of Arizona. As a demonstration project, it was felt that the acquisition of a new CCT was not warranted since TSP data are available for 1972. In addition, the five samplers in operation at that time were located in areas representative of both built-up and undeveloped areas. Second, in anticipation of further analysis of these data and subsequent updates of the classification, the 1972 CCT was the oldest image with which to compare changes in surface cover and related potential dust emissions. Third, good ground data are available for this date. Color infrared photography of the study area is available at an approximate scale of 1:120,0-0 for August 1, 1972 (NASA U-2 mission 72-129). In addition 1:24,000 scale orthophotoquads of the study area derived from NASA U-2 photography taken in May and November 1972 are also available.

Preprocessing

The Landsat CCT, originally obtained from EROS Data Center in Sioux Falls, South Dakota, was preprocessed using program PREPROC[5] to make necessary geometric adjustments and detector striping corrections to upgrade the quality of the image data. The study area was then extracted from the preprocessed image using program EXTRACT.

Digital Image Processing

Subsequent processing of the extracted data was done using the <u>SADIE (System at Arizona for Digital Image Experimentation) version 2.2 image processing software users manual</u> (1979). Specifically, a filtering subroutine in SADIE, MEDIAN, was run to remove some of the severe striping that was evident in band 6. A sliding 3 by 3 pixel window was scanned over the band 6 input image data. The median value within that window was calculated and assigned to the central pixel in the window. In this way, the noise, usually extreme values unrelated to the reflectance of the ground, was removed from the image data and replaced with a value characteristic of that area of the image. Windows of different sizes were tested, but by visual inspection of the results the 3 by 3 pixel window was found to remove most of the noise with minimum alteration of the rest of the data.

Because of the large size of the study area, 500 by 480 pixels, a small part of the study area covering portions of both the built-up and outlying areas was extracted using the SADIE subroutine SUBSAM

---

5. Descriptions of this program and all other software used in this project are available from the Applied Remote Sensing Program, Office of Arid Lands Studies, The University of Arizona.

(Figure 14). This allowed preliminary work to be performed at a reduced cost. Specifically, this small area, the Tucson Sample Area (75 by 175 pixels) was used to test class feasibility for the final classification and to examine features other than the Landsat bands that might be used to help separate these classes in observation space.

A major purpose of image processing is to change digital images so that information may be extracted from them that is not apparent in the original data. A common information extraction process is band ratioing, a technique whereby the reflectance values of pixels in one band are divided by the values of corresponding pixels in another band. As can be seen by examining Figure 9, if band 7 values are divided by band 5 values, vigorous vegetation will have values greater than one, bare soil will have values of near one, and water will generally have values of less than one. In this way the differences between certain types of ground features can be enhanced to aid, for instance, in the classification process.

Various ratios of Landsat bands 5 and 7 and bands 5 and 6 have been used to detect and measure biomass (Richardson and Weigand, 1977; Tucker, 1979). The information derived from the ratio between bands 5 and 7 has also been shown to provide the best basis for separation of abiotic and biotic environments and also to quantify desert vegetation density. The ratio of band 5 to band 4 has been shown to be effective in separating abiotic spectral classes (Heilkema, 1978).

Several ratios were computed with SADIE subroutines and were used together with the four Landsat bands in a classification of the

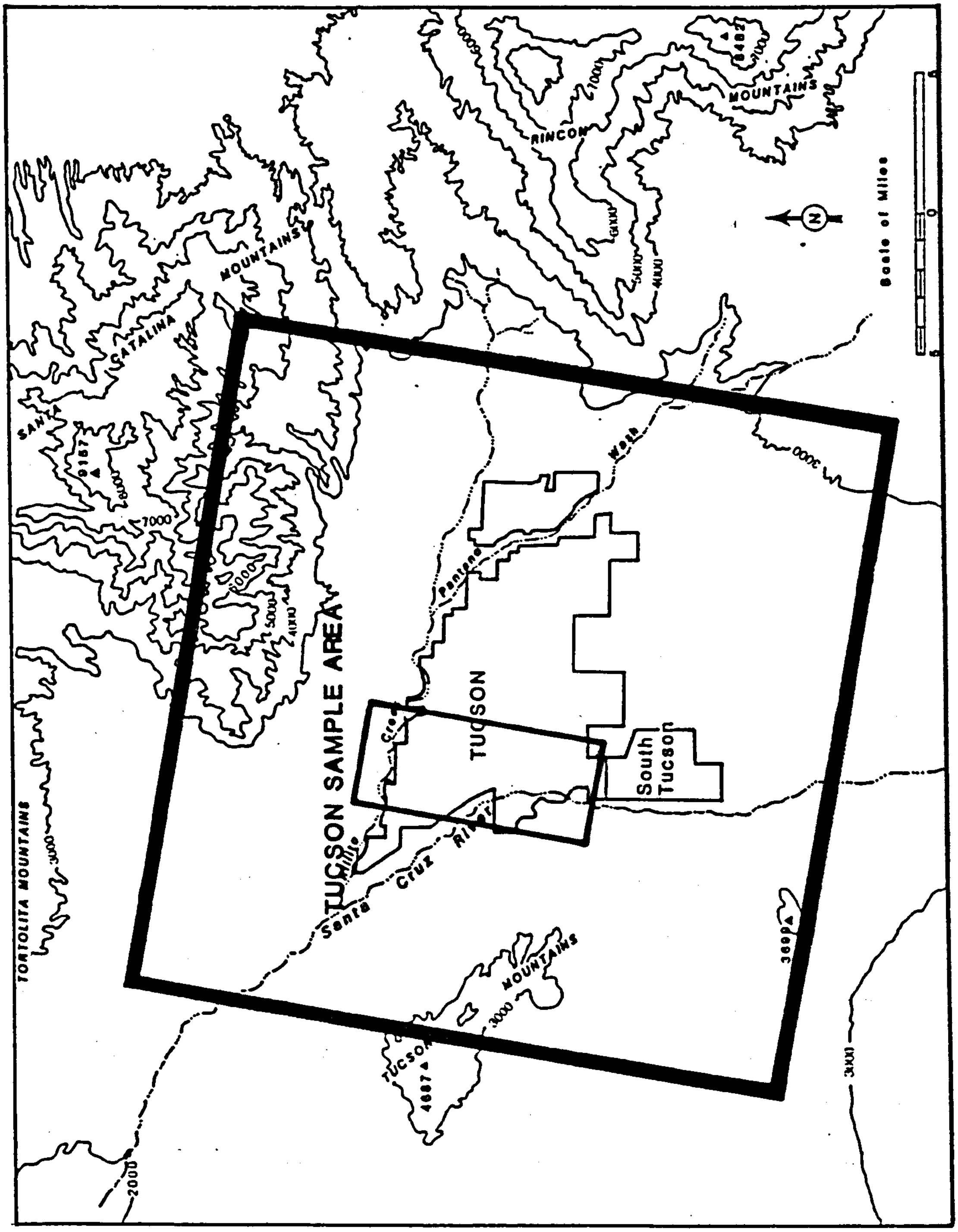


Figure 14. Tucson sample area.

Tucson Sample Area.[6] These features were combined into one image file at a scale of 1:24,000 with the SADIE subroutine CALOUT to be used by the classification software. This scale was chosen because it is large enough so that many urban features can be detected on gray scale maps of the different bands and also because 7.5 minute U. S. Geological Survey topographic maps are available at the same scale, to be used for identifying the exact locations of ground features. This scale has also been recommended for use in urban area Landsat classification in order to record land use on cover delineations in sufficient detail (Joyce, 1974).

Classification

The CALSCAN image classification program (Schowengerdt, 1979) which utilizes the maximum-likelihood algorithm was used to perform the classifications. CALSCAN consists of four modules: STAT, SELECT, CLASSIFY, and DISPLAY. Since no clustering training algorithms (unsupervised approach) were available, a supervised classification was performed. Data from the user-supplied training sites for each class were statistically analyzed by the STAT module to estimate the nature of class populations. Probabilities of pixel membership in each class are derived from these statistics. The classes used are (1) disturbed-vacant land, (2) developed-disturbed land, (3) native vegetation--low per cent cover, (4) native vegetation--medium per cent

---

6. Ratios of Band 6/Band 4, Band 5/Band 4, Band 7/Band 5, Band 6/Band 5, $\sqrt{\frac{\text{Band 7 - Band 5}}{\text{Band 7 + Band 5}} + .05}$, $\sqrt{\frac{\text{Band 6 - Band 5}}{\text{Band 6 + Band 5}} + .05}$

cover, (b) developed land, and (6) vigorous vegetation (see Chapter 4 for class descriptions). From the STAT data, SELECT was used to determine which combinations of features (spectral bands and ratios) optimized the interclass divergence, or the separation between classes. The band 7 to band 5 and band 5 to band 4 ratios were found to be the most effective of those tested in increasing the interclass divergence. Classifications were then performed with CLASSIFY and DISPLAY, using the four Landsat bands and the two ratios. Accuracy of approximately 90% in the training site classifications of the Tucson sample area warranted the use of the classes defined above in the classifications of the entire study area.

Classifications were performed on the entire study area. Two classes, paved areas and native vegetation--high per cent cover, were added because more extensive areas of each were found in the study area outside the Tucson Sample Area. While the training data accuracies were over 90% for all classes, visual comparison of the classification map results and the ground data showed confusion between the paved class and the native vegetation classes and between the native vegetation--medium per cent cover class and the developed land class.

To avoid this confusion, the study area was stratified into built-up and outlying areas, with appropriate classes in each area. This is a common procedure used in other Landsat digital classifications with similar problems (Dornbach and McKain, 1973; Ellefsen et al., 1973; Rohde, 1978). The boundary between the built-up area and outlying areas was drawn using the color infrared photography and the orthophotoquads based on the location of consolidated developed land and undisturbed areas of medium per cent cover native vegetation. Virtually all developed land was included in the built-up area and all areas of medium per cent cover native vegetation were included in the outlying area. This boundary was then transferred to a gray scale map of the study area so the vertices of the polygon boundary could be identified by pixel location.

Program SIPS (Spatial Information Processing System) was used to outline this boundary on the data for the entire study area.[7] All pixels within the built-up area were given a value of one and all pixels outside it were given a value of zero. The data for each band and ratios were then multiplied by this outlying area "mask" using SADIE subroutines. All areas within the built-up area would have their original values and all outside it would have values of zero. A built-up area mask was also created by applying a reverse linear stretch to the outlying mask. In this stretch, performed with SADIE subroutines, all zero values were changed to one and all values of one were changed to zero. The built-up area mask was then multiplied by the original feature data so that all pixels within the built-up area would have a value of zero and all pixels in the outlying area would have their original reflectance values.

Classifications were then performed separately on the built-up and outlying areas using CALSCAN, as was done in the Tucson Sample Area. A thresholding operation was then performed on each classification. Thresholding is a technique, used with the maximum-likelihood classifier,

---

7. Program SIPS was adapted from software developed at the University of California, Riverside (Nichols, 1979).

which allows the user to set the level of probability that must be attained before a pixel is assigned to a class. Thresholding, in effect, establishes another class of "not classified" for all those pixels what are dissimilar to any of the classes according to the threshold specified by the user. When a low threshold (a slight lowering of the acceptable probability) was applied to the built-up area classification, all pixels in the outlying area with values of zero in all of the features were thresholded out as not classified. The same technique was used with the outlying area classification.

A low threshold value was applied to both classifications so that only those pixels with a very low probability of belonging to any of the classes would not be classified. It was felt that most of the pixels in both areas could be reasonably assigned to one of the classes because of the general surface cover classes specified. In the outlying area, pixels within shadows in the mountainous areas were thresholded out, while in the built-up area pixels corresponding to light colored aircraft runways and taxi areas at Davis Monthan Air Force Base and some smaller areas in the city, were not classified.

Upon examining the classification maps for both areas, it was found that an inordinate amount of detail, especially in the built-up area, made the maps extremely difficult to read. It was originally thought that this detail would be desirable since it would be a true representation of the diversity of surface cover in the city. The purpose of a map, however, is to convey information. In order to have a product that would communicate results in an acceptable way, the RECLASS option in the DISPLAY module in CALSCAN was used to "smooth" the data on the classification maps. With this option each pixel is reclassified on the basis of its own value and that of its eight neighbors. If a majority of these neighbors are members of one class the pixel is reclassified into that "majority" class. If no majority exists, the pixel remains as originally classified. A comparison of a detailed portion of the classification map, which includes parts of South Tucson and the Tucson central business district, is shown in Figures 15 and 16. The data are less cluttered and less confusing in the "smoothed" map. It more readily conveys essentially the same information and thus will probably be of more utility than the "unsmoothed map."

Results

A classification summary for both areas is shown in Table 7. According to these results, developed land comprises over 42% and native vegetation--low per cent cover over 15% of the built-up area. Disturbed-vacant land and developed-disturbed land together comprise almost 20% of the built-up area. The portion of this area unclassified at the thresholds adopted is almost 8%. Paved areas and areas of vigorous vegetation account for less than 4% and 2% of the built-up area, respectively.

In the outlying area, native vegetation--medium cover, --high cover, and --low cover account for 36%, 17%, and 16% of the surface area, respectively. The developed-disturbed land class comprises the next largest portion, 12% of the outlying area, while disturbed-vacant land accounts for another 5%. Basalt areas cover almost 10% of the outlying

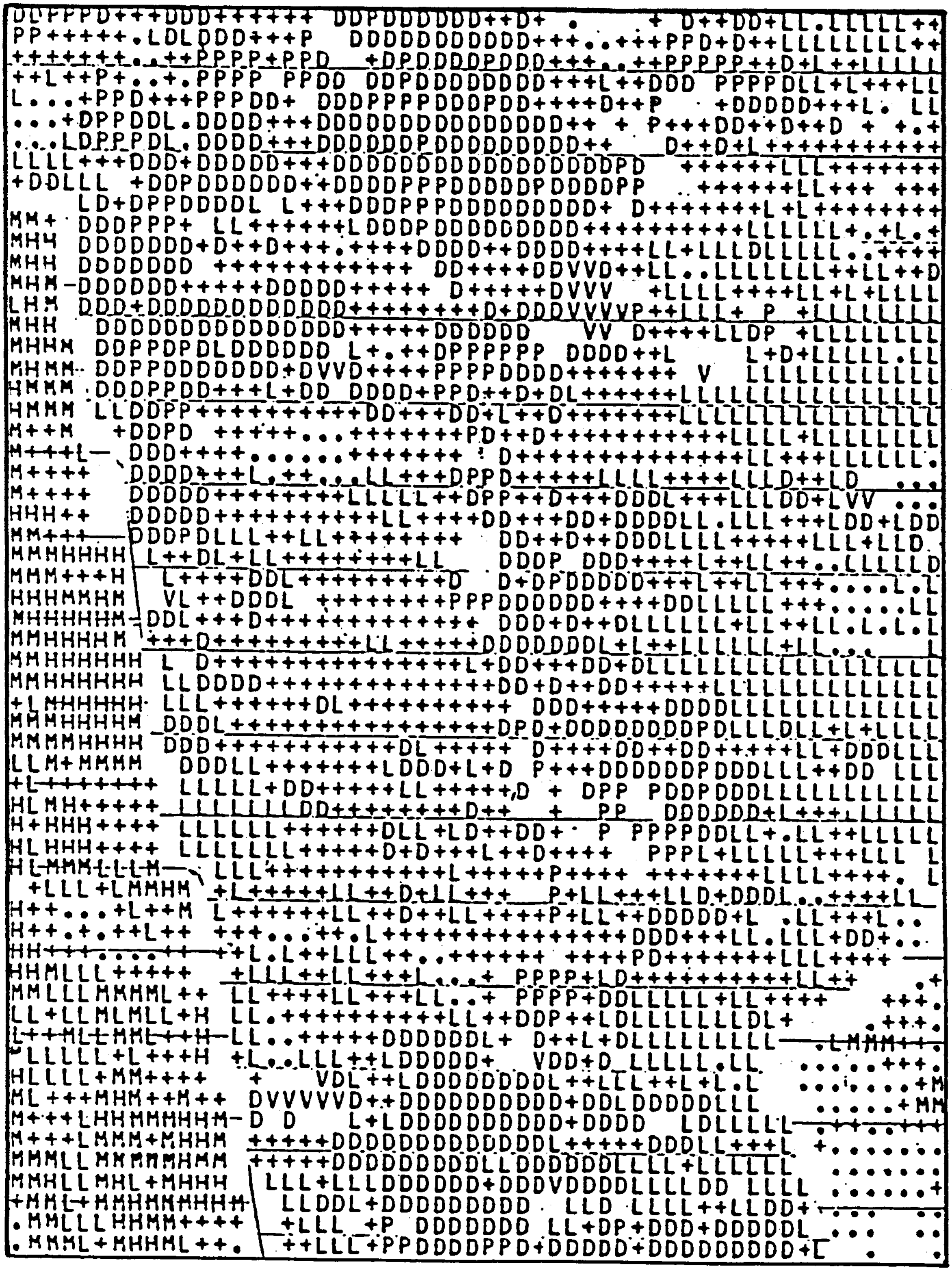


Figure 15. Unsmoothed image data.

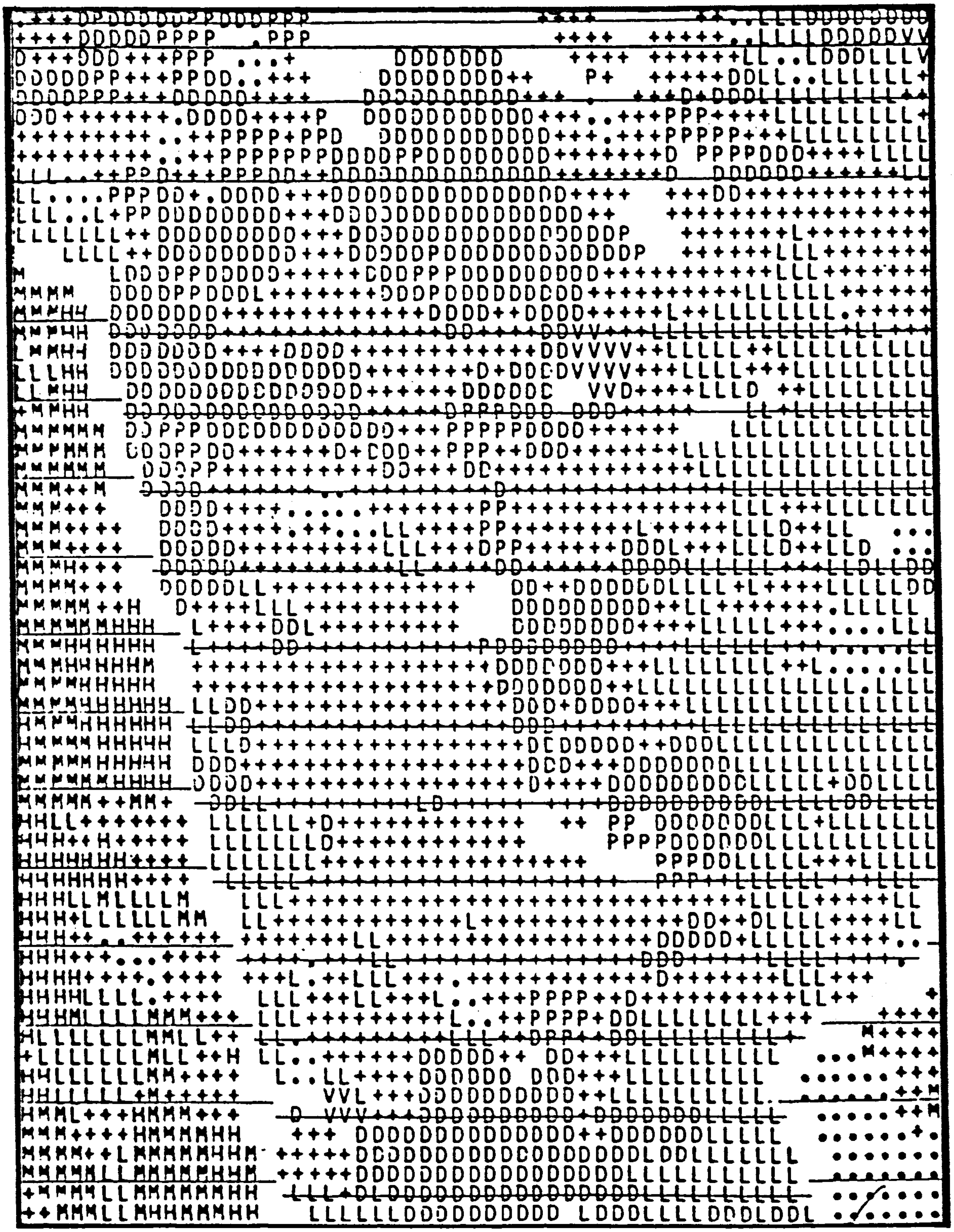


Figure 16. Smoothed image data.

Table 7. Classification Summary

Built Up Area	% of Outlying Area	% of Built Up Area	% of Tucson Urban Area
Disturbed--Vacant Land		4.23	.63
Developed--Disturged Land		15.49	2.31
Native Vegetation--Low Cover		24.82	3.70
Paved Areas		3.22	.48
Developed Land		42.59	6.35
Vigorous Vegetation		1.74	.26
Thresholded (unclassified)		7.91	1.18
		Total	14.91
Outlying Areas			
Disturbed-Vacant Land	4.76		4.05
Developed-Disturbed Land	12.08		10.28
Native Vegetation (Low Cover)	16.05		13.66
Native Vegetation (Medium Cover)	35.73		30.40
Native Vegetation (High Cover)	17.18		14.62
Basalt	9.52		8.10
Vigorous Vegetation	.22		.19
Thresholded (unclassified)	4.45		3.79
		Total	85.09
Entire Study Area			
Native Vegetation (Medium Cover)			30.40
Native Vegetation (Low Cover)			17.36
Native Vegetation (High Cover)			14.62
Developed-Disturbed			12.59
Basalt			8.10
Developed Land			6.35
Thresholded (unclassified)			4.97
Disturbed-Vacant			4.68
Paved			.48
Vigorous Vegetation			.45

area, whereas vigorous vegetation accounts for only a fraction of one per cent of this area. Just over 4% of this area is "thresholded out" and remains unclassified.

Table 7 also shows what the entire study area results are when the subarea classifications are combined. Altogether, undisturbed native vegetation with low, medium, and high cover and basalt areas account for a total of 70% of the Tucson urban area. Both disturbed classes combine for over 17% and developed and paved areas comprise 7% of the total study area. Vigorous vegetation accounts for less than one-half of one per cent of the study area. Only 5% of it is thresholded out as unclassified.

The accuracy of the classification was calculated using a common technique (Swain, 1978; Bauer et al., 1979) of designating test sites and estimating the overall error from the accuracy of the data in these sites. Test sites and training sites should be selected at the same time and should be interchangeable. As in the case with training sites, test sites should be typical of the whole class in question. Estimates of error are also made from the training sites but are only accurate if the training samples are truly representative of the overall data set to be classified. In practice, this is rarely the case and thus estimations of error from training sites are usually optimistic (Swain, 1978). The accuracy of test site estimates is a much more realistic measure of the overall classification accuracy.

The test site accuracies are shown in Table 8. Overall accuracies are approximately 90% for both areas. These results are

Table 8. Summary of Classification Accuracy by Test Sites

	Per Cent Correct	Number of Samples
Built-Up Area		
Disturbed-Vacant Land	83.0	47
Developed-Disturbed Land	100.0	70
Native Vegetation--Low Cover	81.5	54
Paved Areas	91.7	24
Developed Land	100.0	72
Vigorous Vegetation	96.3	27
Overall Accuracy 92.9%		
Outlying Area		
Disturbed-Vacant Land	80.3	66
Developed-Disturbed Land	63.3	60
Native Vegetation--Low Cover	97.6	42
Native Vegetation--Medium Cover	100.0	110
Native Vegetation--High Cover	81.1	37
Basalt	100.0	56
Vigorous Vegetation	83.3	30
Overall Accuracy 88.0%		

promising, but because these sites were chosen a priori as most typical and representative of each class, they seem slightly optimistic and biased. Testing the accuracy of the entire study area with a stratified random sample was contemplated, using techniques devised by Ginevan (1979). Because one of the classes used, developed-disturbed land, is heterogeneous and composed of a mixture of several other classes, interpretation of correct individual pixel classification could not be done in an objective and consistent manner. As a result, evaluation of classification accuracy was limited to the test site method.

Figures 17, 18, and 19 show examples of the classification maps overlayed on corresponding 7.5 minute U. S. Geological Survey topographic sheets. From these examples, a subjective idea of the accuracy of the classification may be derived, and some notion of potential utility to air quality planning can be obtained.

Figure 17. Classification example: South Tucson and Tucson Central Business District.

EXPLANATION

Orange Overlay:

- Paved Areas
- Native Vegetation--High Percent Cover
- Developed-Disturbed Land
- Native Vegetation--Medium Percent Cover

Black Overlay:

- Vigorous Vegetation
- Developed Land
- Native Vegetation--Low Percent Cover
- Disturbed-Vacant Land
- Thresholded (unclassified)

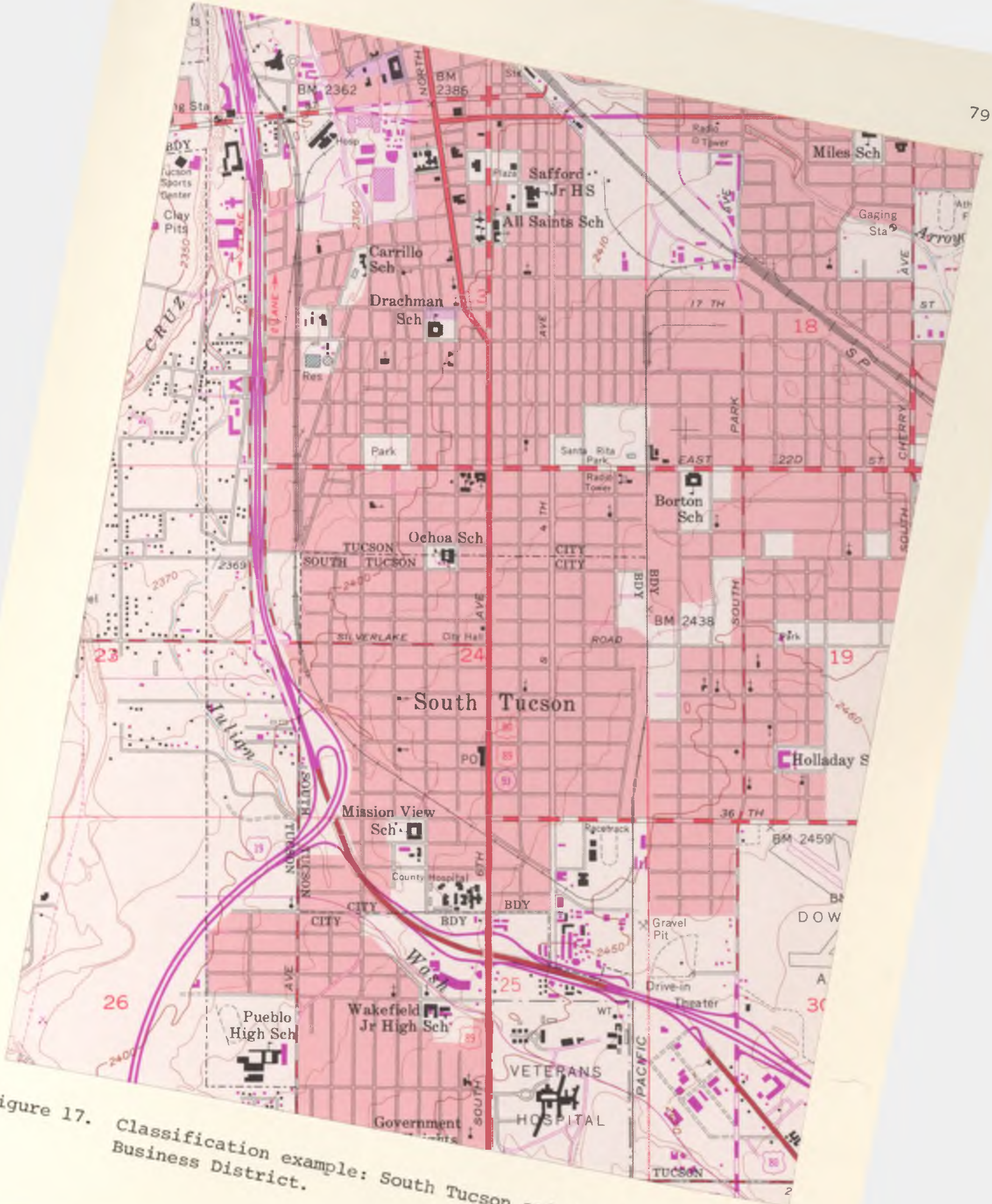


Figure 17. Classification example: South Tucson and Tucson Central Business District.

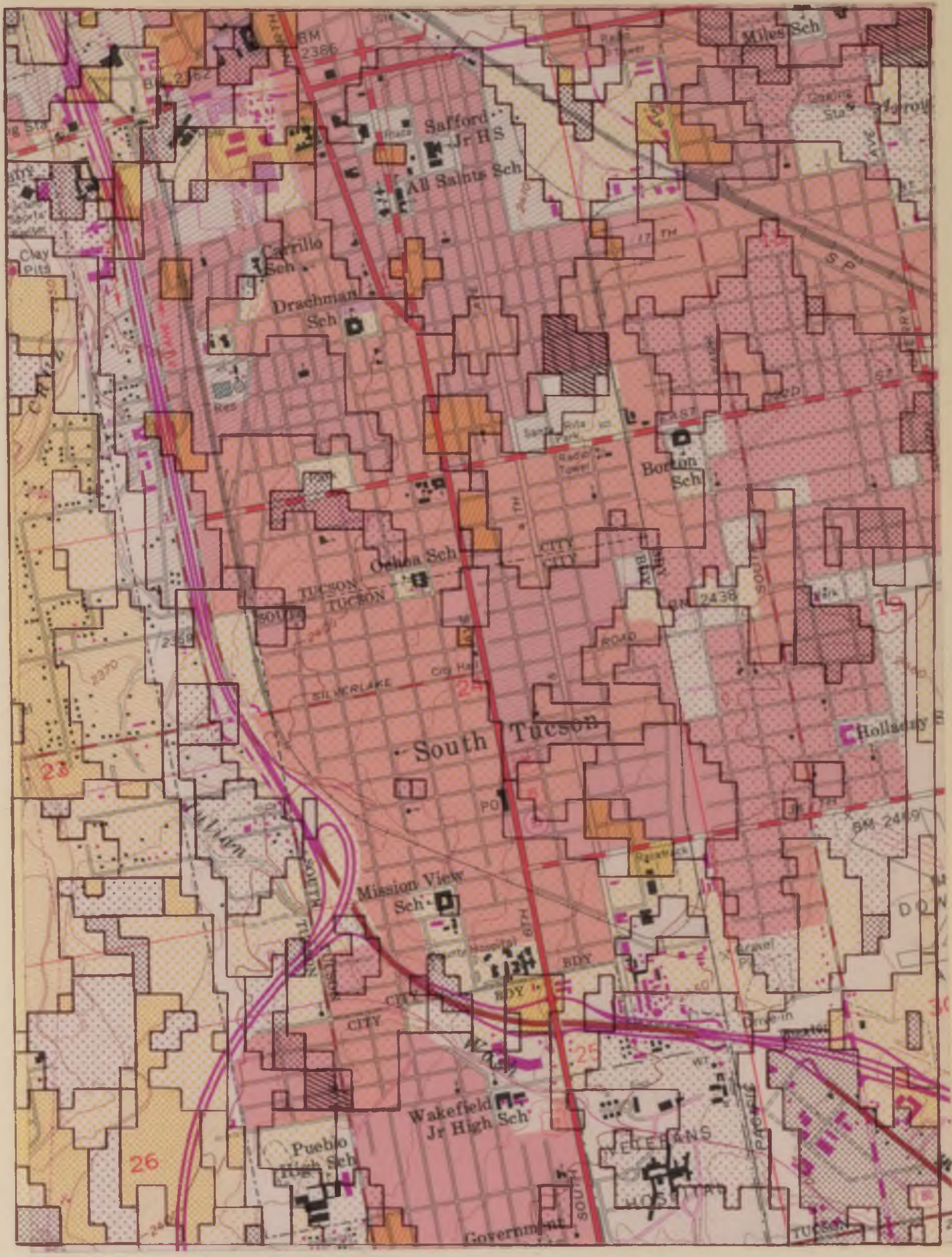


Figure 17. Classification example: South Tucson and Tucson Central Business District.

Figure 18. Classification example: Residential area in Tucson.

EXPLANATION

Orange Overlay:

Paved Areas

Developed–Disturbed Land

Black Overlay:

Vigorous Vegetation

Developed Land

Native Vegetation--
Low Percent Cover

Disturbed–Vacant

U Thresholded (Unclassified)

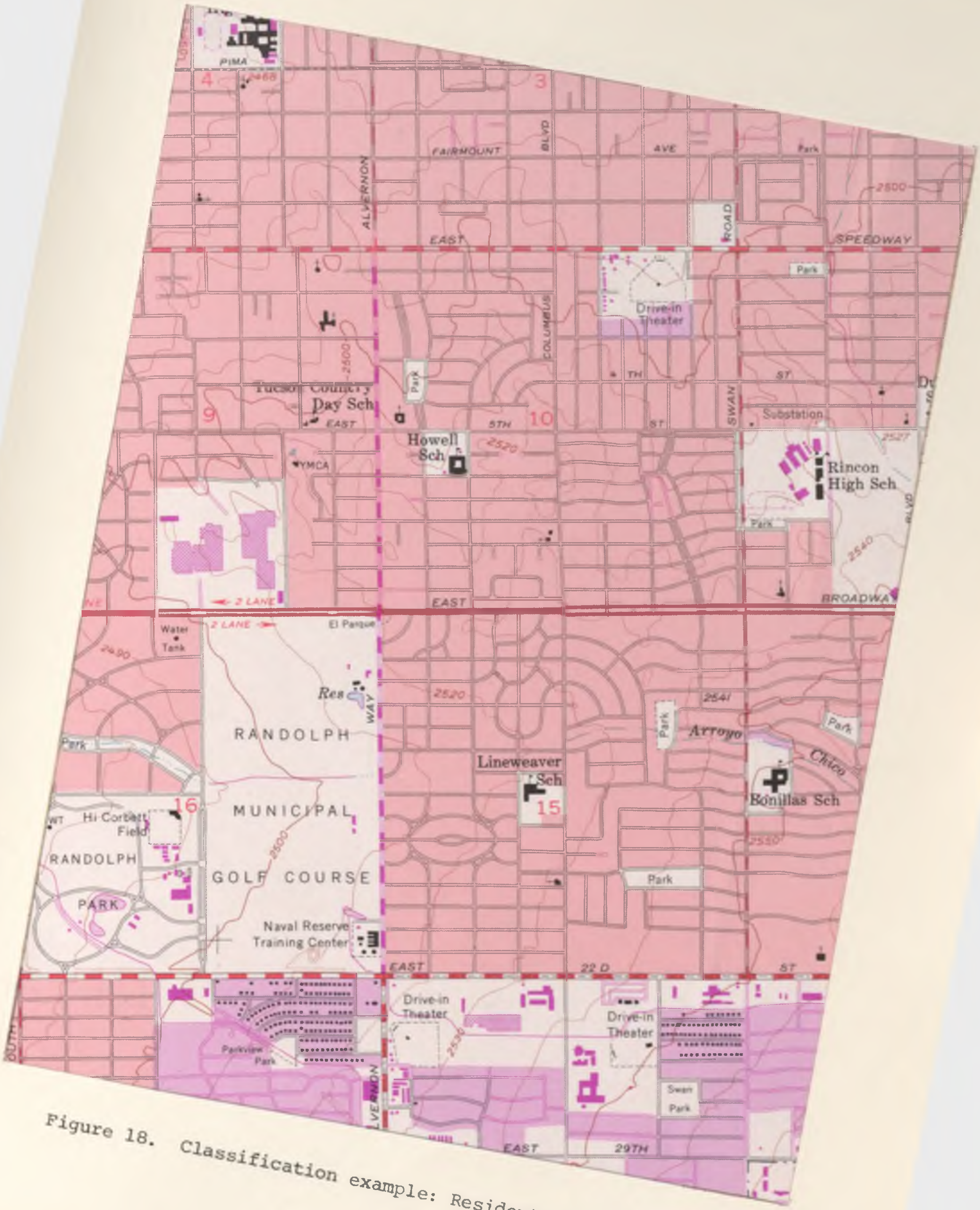


Figure 18. Classification example: Residential area in Tucson.

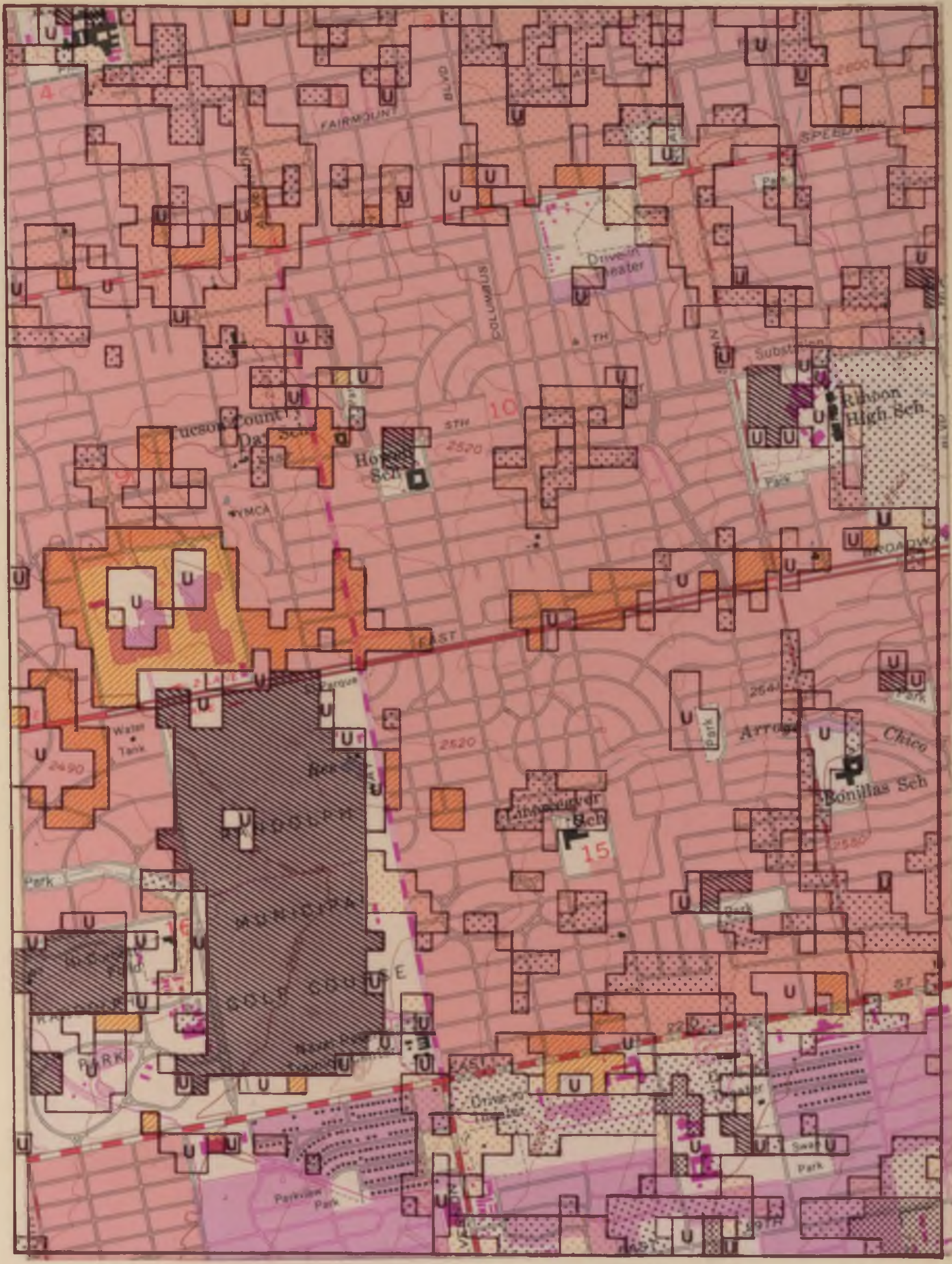


Figure 18. Classification example: Residential area in Tucson.

Figure 19. Classification example: Outlying area.

EXPLANATION

Orange Overlay:

- Basalt
- Native Vegetation--
  High Percent Cover
- Developed-Disturbed Land
- Native Vegetation--
  Medium Percent Cover

Black Overlay:

- Native Vegetation--
  Low Percent Cover
- Disturbed-Vacant
- U Thresholded (unclassified)

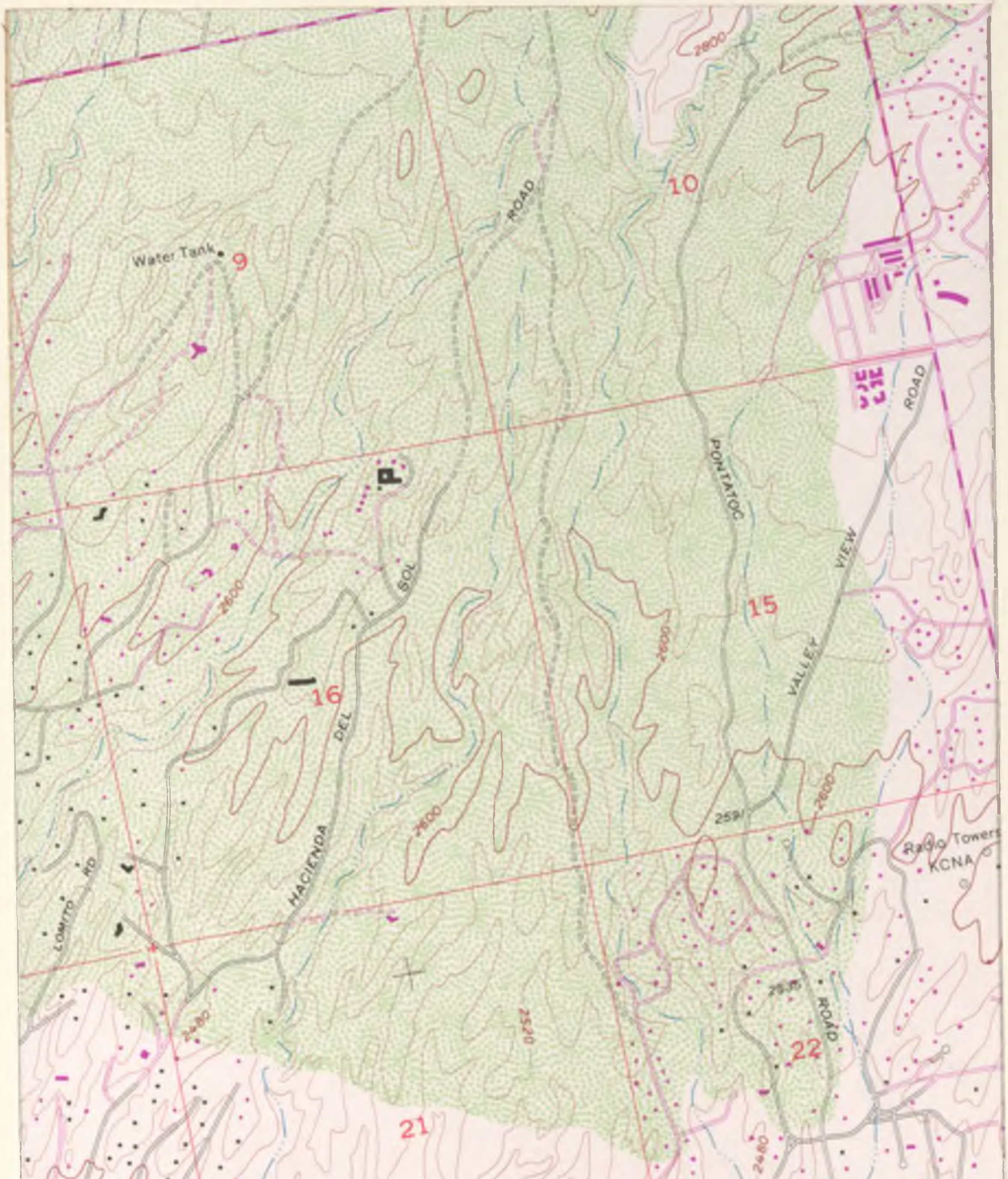


Figure 19. Classification example: Outlying area.

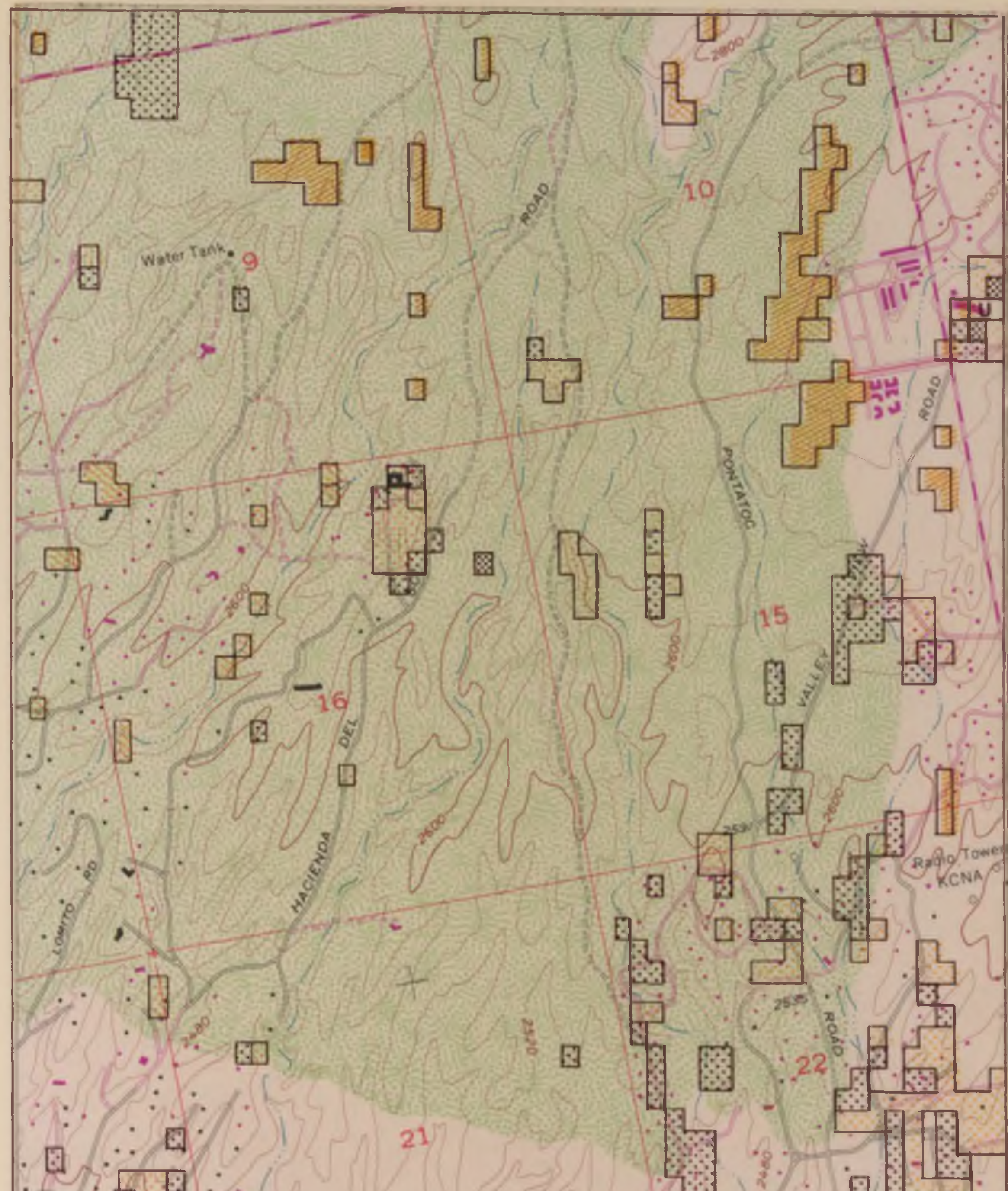


Figure 19. Classification example: Outlying area.

CHAPTER 6

SUMMARY AND CONCLUSIONS

This project was designed to test the utility of Landsat digital data for monitoring dust sources in the Tucson urban area. The Tucson urban area is part of the Tucson Air Planning nonattainment area for TSP, where the main reason for NAAQS violations is dust resulting from human disturbance of the ground (Heidel, Dixon, and Young, 1978). Investigation was confined to the urban area, where the vast majority of people within the nonattainment area reside. As a result, ground disturbance and dust emissions are greatest there.

A supervised classification using Landsat digital data was performed on the study area to determine the location and extent of general surface cover classes that can be related to varying potential dust emissions. These classes were used, as opposed to classes of specific sources, in light of the 76 by 76 meter pixel size of the Landsat multispectral scanner. The study area was stratified into built-up and outlying subareas in order to minimize confusion between classes that are spectrally similar but yet are located in the different subareas.

The accuracy of the classification for the entire study area, estimated from test sites, is approximately 90%. In the outlying area the accuracy is 88%. The disturbed-vacant land and developed-disturbed land classes are classified correctly 80.3% and 63.3% of the time, respectively. Most of the error in test samples of these two classes results from confusion between each other. If these classes are combined into one class of disturbed areas, the accuracy of this combined class is increased to 92.1% and the overall accuracy of the outlying area classification increases to 94.3%. In the built-up area, classification accuracy is 92.9%. The accurcay of the two disturbed classes is 83.0% and 100.0%, respectively. No confusion exists between these classes in the test samples of the built-up area classification.

From close inspection of the verification photography it is suggested that in some cases the test site accuracies are misleading. Disturbed-vacant land, the class with the highest potential dust emission, is classified with a high degree of accuracy. In a few cases, developed areas with closely situated buildings having white roofs (particularly mobile home and trailer parks) are classified in the disturbed-vacant class. While these misclassifications are minimal, it was found that confusion exists between the developed-disturbed and native vegetation--low cover classes in many cases. As a result, it is recommended that these classes be grouped in one class, for subsequent analysis. This class would then represent areas with the second highest potential for dust emissions.

The location and extent of this combined class and the disturbed-vacant land class, detected from Landsat digital data, could provide extremely useful information for air quality planners. Areas included in these classes vary from year to year in size and location. By using Landsat classifications to monitor these changes over a few years air quality analysts could determine if relationships exist between these source areas and TSP concentrations. Microinventory techniques could be used to monitor areas in close proximity to TSP samplers (see Pace, Axetell, and Zimmer, 1978; Pace, 1979), while the Landsat derived data could be used to provide a broad picture of dust-related surface conditions.

One drawback to annual monitoring of high potential dust emission areas with Landsat digital data is that many of the transitory sources may be missed, depending on whether or not these source areas are being actively disturbed at the date of the chosen scene. This problem cannot be resolved unless more updates are performed. But it is maintained that this technique is best suited to long term planning efforts and that annual updates will enable planners to determine where their efforts should be focused. Dust abatement measures could be focused on persistent source areas, such as industrial areas with unpaved parking and storage areas. The more transitory sources, such as construction areas, could be dealt with as they occur with stricter enforcement of air quality statutes.

The utility of Landsat digital data in monitoring dust sources has been demonstrated. In arid and semi-arid urban areas with dust problems related to surface cover disturbance, this tool provides air quality planners with an increased perspective with which to view the problem and to develop proficient dust abatement measures.

REFERENCES

Anderson, James R., Ernest E. Hardy, John T. Roach, and Richard E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data (U. S. Geological Survey Professional Paper 964). U. S. Geological Survey, Arlington, Virginia.

Bach, Wilfrid. 1971. Atmospheric turbidity and air pollution in greater Cincinnati. The Geographical Review 61(4):573-593.

Bauer, Marvin E., Jan E. Cipra, Paúl E. Anuta, and Jeanne B. Etheridge. 1979. Identification and area estimation of agricultural crops by computer classification of Landsat MSS data. Remote Sensing of Environment 8:77-92.

Berry, Brian J. L., and Frank E. Horton. 1974. Managing the air resource. Pages 124-166 in Brian J. L. Berry and Frank E. Horton, eds., Urban environmental management. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Bibbero, Robert J., and Irving G. Young. 1974. Legislative basis of air-pollution control systems. Pages 139-186 in Robert J. Bibbero and Irving G. Young, eds., Systems approach to air pollution control. John Wiley and Sons, Inc., New York.

Birkle, Michael. 1978. Monitoring. Pages 263-282 in Hanns Malissa, ed., Analysis of airborne particles by physical methods. CRC Press, West Palm Beach, Florida.

Bridson, Lynne. 1980. Air quality analyst, Pima County Health Department Air Quality Control District, Tucson, personal communication.

Buchholz, Nick. 1980. Air quality planner, Pima Association of Governments, Tucson, personal communication.

Caldwell, Roger L., ed. 1971. Air pollution in Tucson (a report to the Mayor and Council of the City of Tucson). Tucson Advisory Committee on air pollution.

Cowherd, Chatten, Jr., Kenneth Axetell, Jr., Christine M. Guenther, and George A. Jutze. 1974. Development of emission factors for fugitive dust sources (EPA-450/3-74-037). U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Dornbach, John E., and Gerald E. McKain. 1973. The utility of ERTS-1 data for applications in land use classification. Pages 439-456 in Stanley C. Freden, Enrico P. Mercanti, and Margaret A. Becker, eds., Third Earth Resources Technology Satellite-1 Symposium, volume 1: technical presentations, section A, December 10-14, 1973 (NASA SP-351). National Aeronautics and Space Administration, Washington, D. C.

Doyle, Frederick J. 1978. The next decade of satellite remote sensing. Photogrammetric Engineering and Remote Sensing 44(2): 155-164.

Easton, Eric B., and Francis J. O'Donnell. 1977. The Clean Air Act Amendments of 1977: refining the national air pollution control strategy. Journal of the Air Pollution Control Association 27(10):943-947.

Ellefsen, Richard. 1974. Methods and problems in computer-aided land-use classification of four U. S. urban areas: an experiment using ERTS-1 multispectral data and high altitude color IR air photography. Association of American Geographers, Proceedings 6:69-73.

Ellefsen, Richard, Leonard Gaydos, Philip Swain, and James R. Wray. 1974. New techniques in mapping urban land use and monitoring change for selected U. S. metropolitan areas: an experiment employing computer-assisted analysis of ERTS-1 MSS data. Pages 51-63 in Symposium on Remote Sensing and Photointerpretation, Proceedings, volume 1. Ottawa, Canada.

Ellefsen, Richard, Philip Swain, and James Wray. 1973. Urban land-use mapping by machine processing of ERTS-1 multispectral data: a San Francisco bay area example. Pages 2a-7-22 in Symposium on Machine Processing of Remotely Sensed Data, October 16-18 (IEEE 73 CHO 834-2GE). Institute of Electrical and Electronics Engineers.

Environmental Protection Agency. 1971. National primary and secondary ambient air quality standards. Federal Register 36(84):8186-8187.

__________. 1977. Control strategy preparation manual for particulate matter (EPA-450/2-77-023). Standard Implementation Branch, Control Programs Development Division. Research Triangle Park, North Carolina.

__________. 1979. Air quality control regions, criteria and control techniques; attainment status designations. Federal Register 44(70):21264.

Gaydos, Leonard, and Willard L. Newland. 1978. Inventory of land use and land cover of the Puget Sound region using Landsat digital data. Journal of Research U. S. Geological Survey 6(6):807-814.

Gennaro, Nat de, ed. 1979. Arizona statistical abstract: a 1979 data handbook. Northland Press in cooperation with the Division of Economic and Business Research, College of Business and Public Administration, University of Arizona, Tucson.

Ginevan, Michael E. 1979. Testing land-use map accuracy: another look. Photogrammetric Engineering and Remote Sensing 45(10):1371-1377.

Grigg, David. 1965. The logic of regional systems. Association of American Geographers, Annals 55(3):465-491.

Hartmann, William K. 1976. Air quality: pollution forecast and application to planning (a report to City of Tucson Department of Planning). Planetary Science Institute Science Applications, Inc., Tucson.

Heidel, Karen. 1979. Senior Engineer, Pima County Health Department Air Quality Control District, Tucson, personal communication.

Heidel, Karen, Sharon Dixon, and Mary Young. 1978. Particulate matter emission inventory and emission projections: Pima County, Arizona. Pima County Health Department Air Quality Control District, Tucson.

Heidel, Karen, Peter Hyde, William Montgomery, Bob Ohlfest, and Mary Young. 1978. Technical analysis for the nonattainment area plan for total suspended particulates Pima County, Arizona. Pima County Health Department Air Quality Control District, Tucson.

Hielkema, Jelle U. 1978. Advanced training and research on satellite remote sensing techniques and applications in the United Kingdom and the United States, 1 October 1977-30 September 1978. Food and Agriculture Organization of the United Nations, Rome.

Hutchinson, Charles F. 1978. The digital use of Landsat data for integrated land resource survey: a study in the eastern Mojave Desert, California. Ph.D. dissertation, University of California, Riverside.

Joyce, Armond T. 1974. Land use and mapping. Pages 15-32 in Stanley C. Freden, Enrico P. Mercanti, and David B. Friedman, eds., Third ERTS Symposium, volume 3, discipline summary reports, December 10-14, 1973 (NASA SP-357). National Aeronautics and Space Administration, Washington, D. C.

Lowe, Charles H. 1972. The vertebrates of Arizona. University of Arizona Press, Tucson.

Mausel, Paul W., William J. Todd, and M. F. Baumgardner. 1974. An analysis of metropolitan land-use machine processing of Earth Resources Technology Satellite data. Association of American Geographers, Proceedings 6:54-57.

McCaldin, Roy O. 1977. Fugitive dust study for Pima County Air Quality Control District Tucson, Arizona. Unpublished.

Mees, Quentin M. 1965. Tucson air pollution: 1959-1964 (Engineering Experiment Station Report Number 7). University of Arizona, Tucson.

Moyers, Jarvis L. 1977. The identification and analysis of urban air quality patterns. University Analytical Center, University of Arizona, Tucson.

National Aeronautics and Space Administration. 1976. Landsat data users handbook (Document 76SD4249). National Aeronautics and Space Administration Goddard Space Flight Center. Greenbelt, Maryland.

National Oceanic and Atmospheric Administration. 1978. Local climatological data, annual summary with comparative data: Tucson, Arizona. Environmental Data and Information Service, National Climatic Center, Asheville, North Carolina.

Nichols, David A. 1979. Spatial information processing system. Department of Earth Sciences, University of California, Riverside.

Nunnally, Nelson A., and Richard E. Witmer. 1970. Remote sensing for land-use studies. Photogrammetric Engineering 36(5):449-454.

Orr, Donald G. n.d. Imate forming process. Pages I-1-29 in Donald G. Orr, Introduction to digital image analysis of Landsat data for the Bureau of Land Management. Earth Resources Observation System Data Center, Sioux Falls, South Dakota.

Pace, Thompson G. 1979. An empirical approach for relating annual TSP concentrations to particulate microinventory emissions data and monitor siting characteristics. U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Pace, Thompson G., Kenneth Axetell, and Robert Zimmer. 1978. Microinventories for TSP, in Air Pollution Control Association Specialty Conference on Emission Factors and Inventories, Proceedings. Anaheim, California.

Pacific Northwest Land Resources Inventory Demonstration Project. 1979. Urban application of Landsat information, Tacoma. Pixel Facts 20:1-4.

Pima County Air Quality Control District. 1980. Air quality in Tucson, Arizona, 1979. Pima County Health Department, Tucson.

Pima County, City of Tucson, City of South Tucson, and Town of Oro Valley. 1977. The comprehensive plan: findings and policies.

R. A. Drafting Service. 1979. Tucson 1980's Mjaor Streets and Area Guide Map. Tucson, Arizona.

Reed, Wallace E., and John E. Lewis. 1975. Land use information and air quality planning: an example of environmental analysis using a pilot national land use information system (Final Report--volume 7), Central Atlantic Regional Ecological Test Site Project). National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland and U. S. Geological Survey, Reston, Virginia.

Richard, George, Ron Tan, and Jim Avery. 1977. An implementation plan for suspended particulate matter in the Phoenix area, volume 2--emission inventory (EPA-450/3-77-0021b). U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Richardson, A. J., and C. L. Weigand. 1977. Distinguishing vegetation from soil background information. Photogrammetric Engineering and Remote Sensing 43(12):1541-1552.

Rohde, Wayne G. 1978. Improving land cover classification by image stratification of Landsat data. Pages 729-742 in Twelfth International Symposium on Remote Sensing of the Environment, Proceedings, volume 1, April 20-26, 1978. Center for Remote Sensing Information and Analysis, Environmental Research Institute of Michigan, Ann Arbor.

SADIE (System at Arizona for Digital Image Experimentation) version 2.2 image processing software users manual. 1979. Applied Remote Sensing Program, Office of Arid Lands Studies, University of Arizona, Tucson.

Schlien, S., and D. Goodenough. 1974. Quantitative methods of processing the information content of ERTS imagery for terrain classification. Pages 237-265 in Second Canadian Symposium on Remote Sensing, volume 1.

Schowengerdt, Robert A. 1979. CALSCAN users manual. Applied Remote Sensing Program, Office of Arid Lands Studies, University of Arizona, Tucson.

Sellers, William D., and Richard H. Hill, eds. 1974. Arizona climate: 1931-1972. Revised second edition. University of Arizona Press, Tucson.

Short, Nicholas M., Paul D. Lowman, Jr., Stanley C. Freden, and William A. Finch, Jr. 1976. Mission to earth: Landsat views the world (NASA SP-360). National Aeronautics and Space Administration, Washington, D. C.

Slater, Philip N. 1979. A re-examination of the Landsat MSS. Photogrammetric Engineering and Remote Sensing 45(11):1479-1485.

Sokal, Robert R. 1974. Classification: purposes, principles, progress, prospects. Science 185(4157):1115-1123.

Steenbergh, Warren F., and Peter L. Warren. 1977. Ecological investigation of natural community status Organ Pipe Cactus National Monument. Cooperative National Park Resources Study Unit, University of Arizona, Tucson.

Steiner, Dieter, and Anthony E. Salerno, eds. 1975. Remote sensing data systems, processing and management. Pages 611-803 in Robert G. Reeves, ed., Manual of Remote Sensing. American Society of Photogrammetry, Falls Church, Virginia.

Swain, Philip H. 1978. Fundamentals of pattern recognition in remote sensing. Pages 136-186 in Philip H. Swain and Shirley M. Davis, eds., Remote sensing: the quantitative approach. McGraw-Hill, Inc., New York.

Tucker, Compton. 1979. Red and photographic infrared linear combinative for monitoring vegetation. Remote Sensing of the Environment 8(2):127-150.

U. S. Bureau of the Census. 1973. Census of population 1970, volume 1 characteristics of the population, part 4 Arizona. U. S. Government Printing Office, Washington, D. C.

__________. 1976. Special census of Pima County, Arizona, October 20, 1975 (current population reports, series p-28, special censuses, number 1541). U. S. Government Printing Office, Washington, D. C.

Wacker, A. G., and D. A. Landgrebe. 1972. Minimum distance classification in remote sensing. Pages 577-599 in First Canadian Symposium on Remote Sensing, Proceedings, volume 2. Department of Energy, Mines and Resources, Ottawa.

Witmer, Richard E. 1977. The USGS land use and land cover classification system. Pages 10-19 in James R. Anderson, ed., Land use and land cover maps and statistics from remotely sensed data (Remote Sensing of the Electro-magnetic Spectrum 4(4). University of Nebraska, Omaha.