EXTRA ORIGINAL COPY INSTEAD OF ~~ZEROX~~

420 W. 118th St.
New York, NY 10027

CR-157145

Semi-Annual Progress Report

Sept., 1977

NSG 5080

(E78-10133) APPLICATION OF DIGITAL ANALYSIS OF MSS DATA TO AGROENVIRONMENTAL STUDIES Semiannual Progress Report (Columbia Univ.) 111 p HC A06/MF A01 CSCI 02C G3/43 N78-25499 Unclas 00133

Columbia University in the City of New York | *New York, N. Y. 10027*

DEPARTMENT OF GEOGRAPHY

International Affairs Building
420 West 118th Street

APPLICATION OF DIGITAL ANALYSIS OF MSS DATA TO AGRO-ENVIRONMENTAL STUDIES

Semi-Annual Progress Report
NASA Grant NSG-5080
September 1, 1977

Principal Investigators

Kempton E. Webb
Colin J. High
Jerry C. Coiner

CONTENTS

I. Columbia University, Geography Department, Remote Sensing Project.

Since June 1975, Columbia University, Department of Geography has been conducting research on various aspects of applying LANDSAT and other multispectral data to a broad range of topics related to agriculture, agricultural development, and the environment. The research is carried out in collaboration with the NASA/Goddard Institute for Space Studies (GISS) and is intended to support the Institute's program in Earth Resources.

The primary purpose of the research is to aid in developing a global agricultural monitoring system which would rely on satellite data collection systems as major data sources. This involves the study of three specific topics. 1) agricultural change, 2) traditional agricultural systems and 3) relationships between agriculture and the physical environment.

Primary data sources for these studies are simulated advanced multispectral sensors of various resolutions, 24-Channel multispectral scanner (MSS) data, LANDSATS 1 and 2, as well as other remote sensing data made available through the offices of GISS. MSS, LANDSAT, Skylab and aircraft spectrometer data are analyzed by means of GISS-developed software and computing facilities.

The faculty of the Department of Geography, Columbia works closely with GISS personnel as well as supervises the student research assistants who may undertake individual projects within the framework of this proposal.

Participating Columbia University faculty members from the Department of Geography are:

Professor Kempton E. Webb, Chairman

Professor Colin J. High

Dr. Jerry C. Coiner

In addition to the three principal investigators, the staff includes five funded graduate students, three part-time student technicians and four nonfunded graduate researchers.

The following reports and publications were presented or published by the staff during the reporting period.

1. "Information Systems Concepts for Integrating Landsat and Other Data Sources," Jerry C. Coiner, working paper presented at the Conference/Workshop on The Use of Landsat Data in the Analysis of Non-Western Ecosystems, Wenner-Gren Foundation For Anthropological Research, New York, May, 1977.

2. "Atlas of Selected Crop Spectra; Imperial Valley, California," S. G. Ungar, W. Collins, J. Coiner, et al, June 1977.

3. "A First Interpretation of East African Swiddening via Computer Assisted Analysis of 3 Landsat Tapes," Francis P. Conant & Tina K. Cary. Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, June, 1977.

Copies are included in the Appendix to this report.

II. Experiments in Progress

A. Support to Thematic Mapper Design: Classification Studies

An extensive series of investigations have been conducted at the Goddard Institute for Space Studies to determine appropriate sensor parameters for the Thematic Mapper instrument under development for LANDSAT D, which is scheduled to go into operation some time in the early 1980's. This multispectral scanner will provide data in up to 7 bands at a higher resolution than the scanners now on LANDSATs 1, 2, and C. During the six months ending August 31, 1977 particular emphasis was placed on the role of thermal data in classification accuracy.

The purpose of these investigations has been to assess the effectiveness of alternative combinations of channels and resolutions for agricultural crop analysis. The Thematic Mapper instrument is simulated using selected channels from data acquired by a 24-Channel multispectral scanner flown over selected Large Area Crop Inventory Experiment (LACIE) test sites in Williams County, North Dakota and Finney County, Kansas during the summer of 1975. The data were subsequently degraded from an acquisition resolution of approximately 6 meters to both 30 and 60 meter resolutions.

Of particular interest are the studies designed to test the impact on classification accuracy of the addition of a thermal band (band 7, 10.10-12.00 micrometers) at both 30 meter and 120 meter resolutions to combinations of the 30 meter reflective bands (bands 1, 3, 4, 5, and 6, in the range of from 0.47 to 1.62 micrometers).

These studies were conducted using data for the LACIE Intensive Study Site in Williams County, North Dakota, August 15, 1975. The procedure for this set of studies was as follows:

1. Data for Band 7, at both 30 and 120 meter resolution was added to data for the full set of reflective bands, and classification was performed using each of these two configurations.

2. In order to determine if band 7 would have any significant impact when added to a smaller set of reflective bands, each configuration of the 4 reflective bands (1, 3, 4+5, 6), taken 3 bands at a time, was tested on the training areas. The 3-band reflective configuration that performed most poorly in terms of total classification accuracy was selected as the one to which the thermal band would be added.

3. The 3-band reflective configuration selected as described in the previous step and its associated thermal band configuration were used to classify on the available test sites (3 flight lines for the 120 meter-resolution thermal data and 1 flight line for the 30 meter-resolution thermal data).

4. Steps 2 and 3 above were repeated in order to test sequentially the addition of the thermal band to a 2-band reflective combination and finally to a single reflective band.

The classification results for the experiments described above are shown in Table 1. The crop identification accuracy statistics for the flight line data are plotted in Figure 1.

B. Support to Airborne Spectroradiometer Studies

GISS/Columbia has developed a high resolution spectroradiometer. This instrument acquires line trace spectra over a range from .43 to 1 micron for sequential areas approximately 18 meters square. To aid in interpreting these spectra and relating them to specific ground locations a 35 mm photograph is acquired with every tenth spectrum.

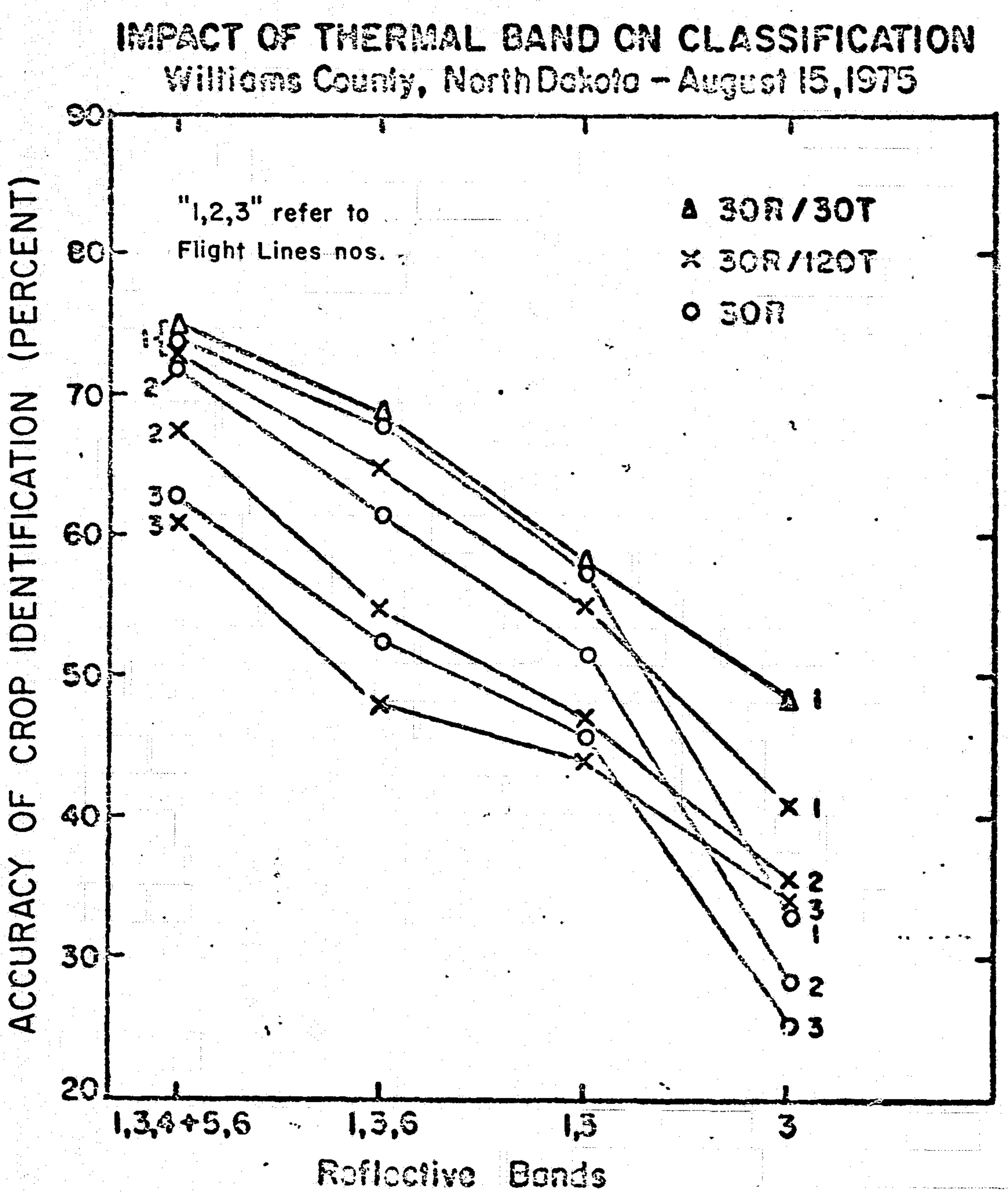
Data acquired by this system over Imperial Valley, California were selected for inclusion in the GISS Crop Spectra Atlas, the design and production of which the Columbia staff provided major support for. (Included as document 2 in the Appendix of this report are Part I and the appendices of the Crop Spectra Atlas.) The Crop Spectra Atlas is designed to make available to interested investigators high resolution spectral data for representative agricultural ground surfaces.

An intensive photointerpretation of the 35 mm aerial photography collected concurrently with the spectral data was performed in order to support and clarify the latter. Additional support in the preparation of this Atlas included preparation of the spectral plots, determination of criteria for homogeneity, and the selection of spectra meeting these criteria to ensure that each spectra set represented a single homogeneous agricultural field state.

TABLE 1 SUMMARY STATISTICS
THERMAL BAND CLASSIFICATION STUDIES
Williams County, North Dakota
August 15, 1975

RESOLUTION	30M REFLECTIVE/120M THERMAL								30M REFLECTIVE 30M THERMAL			
BANDS	1, 3, 4+5, 6	1, 3, 4+5, 6, 7	1, 3, 6	1, 3, 6, 7	1, 3	1, 3, 7	3	3, 7	1, 3, 4+5, 6, 7	1, 3, 6, 7	1, 3, 7	3, 7
DATA SET	ACCURACY OF CROP IDENTIFICATION (PERCENT)											
TRAINING AREAS	83	85	74	76	62	62	43	47	86	81	66	57
FLIGHT LINE 1	74	73	67	65	58	55	33	42	75	69	58	49
FLIGHT LINE 2	72	68	60	53	50	46	28	36				
FLIGHT LINE 3	65	61	53	48	46	44	26	34				
DATA SET	ACCURACY OF CROP ACREAGE ESTIMATION (PERCENT)											
TRAINING AREAS	86	86	75	78	60	60	47	45	86	79	61	51
FLIGHT LINE 1	77	76	63	58	46	42	23	28	78	63	43	33
FLIGHT LINE 2	71	68	48	32	30	19	12	12				
FLIGHT LINE 3	62	48	35	29	25	27	7	17				

FIGUR I


IMPACT OF THERMAL BAND ON CLASSIFICATION
Williams County, North Dakota - August 15, 1975
ACCURACY OF CROP IDENTIFICATION (PERCENT)
"1,2,3" refer to
Flight Lines nos.
30R/30T
30R/120T
30R
Reflective Bands
1,3,4+5,6
1,3,6
1,3
3

Columbia staff prepared the text and graphics for the Atlas.

C. LANDSAT Applications to Agricultural Change Studies

1. Monitoring Agricultural Transition in Southern New Jersey

This research is exploring the possibilities of monitoring agricultural change with LANDSAT data. One such attempt using data from southern New Jersey is described below.

1.1. Background

For this study a 13,675 hectare (approximately 34,000 acres) area in southern New Jersey was chosen, an area which is known to be under considerable pressure to adjust to new agricultural circumstances. The study area is located about 40 kilometers south of Philadelphia, Pennsylvania, and 15 kilometers west of Vineland, New Jersey. The study area lies in both Salem and Cumberland counties, within the Upper Pittsgrove and Alloway townships in Salem County and the Upper Deerfield Township in Cumberland County. The landscape is gently rolling, underlain by geologic formations of loosely consolidated sedimentary deposits. Soils in the parts of these townships which are in the study area belong to a soil association that is silty, well-drained to poorly drained, nearly level to gently sloping.

Agriculture and food processing combine to form New Jersey's largest industry. Over the past few years, the state's total area in vegetables has declined whereas total area in wheat, corn, and soybeans has increased. The area harvested in Cumberland and Salem counties for selected field crops is detailed in Table 2. These latter crops have always played a part in the New Jersey agricultural system, but their recent areal increase is indicative of a statewide trend. Table 3 shows the decline in hectares harvested for commercial vegetables, both fresh market and processing.[1] From this table it appears that the processing vegetables have undergone the more severe drop. Economic and environmental pressures have forced many members of the vegetable processing community in New Jersey to move their operations elsewhere. In recent years, the Del Monte Corporation, P. F. Ritter (Curtis-Burns Corporation), and the Seabrook Foods Corporation have ceased purchasing New Jersey-grown produce. New Jersey farmers have had to find either new markets or new crops; if they choose the latter alternative, field crops are often the most suitable because (1) a farmer can easily convert to this labor-extensive system and (2) wheat, soybeans, and corn have been drawing an increasingly higher price on the market (wheat and corn dropped in 1977).

The approximately 14,000 hectares which comprise the test area for this study lie in and around the area formerly farmed by Seabrook Farms, Incorporated, which closed its

---

[1]Significant discrepancies have been noted with respect to data in this category as reported annually in New Jersey Agricultural Statistics for the years 1971 through 1976. (Each annual report gives statistics in this category for preceding years as well as the current year.) For example the 1972 Statistics reported the total area harvested in 1971 for all commercial vegetables for processing as 17,200 hectares. The 1976 Statistics reported 17,600 hectares in this category for 1971, and in 1977 the figure given was 13,400 hectares, as shown in Table 3. According to the New Jersey Crop Reporting Service, the latest (1977) figures are the most accurate, implying that it took six years to rectify a 4,000-hectare error--an error amounting to nearly one-third the acreage in that category. It appears that LANDSAT acreage estimators would be more efficient in this regard.

TABLE 2

AREA HARVESTED IN HECTARES--SELECTED FIELD CROPS

CUMBERLAND AND SALEM COUNTIES, 1971-1976

County/Year	Corn	Wheat	Soybeans
Cumberland			
1971	1,290	930	1,170
1972	1,150	1,290	1,170
1973	1,320	1,620	1,340
1974	1,340	1,940	1,700
1975	1,460	2,270	1,900
1976	2,020	2,350	3,040
Salem			
1971	5,260	1,340	2,020
1972	4,130	1,660	2,430
1973	4,740	1,990	2,910
1974	5,020	3,360	3,640
1975	5,220	3,160	3,850
1976	6,480	3,320	5,950

SOURCE: New Jersey, Department of Agriculture, New Jersey Crop Reporting Service, New Jersey Agricultural Statistics, August 1972, August 1975, August 1976, August 1977.

TABLE 3

AREA HARVESTED IN HECTARES--ALL COMMERCIAL VEGETABLES

STATE OF NEW JERSEY

1971-1976

Year	For Fresh Market	For Processing	Total
1971	24,140	13,400	37,540
1972	24,120	12,300	36,420
1973	22,610	12,970	35,580
1974	20,850	13,800	34,650
1975	19,280	11,080	30,360
1976	19,850	5,700	25,550

SOURCE: New Jersey, Department of Agriculture, New Jersey Crop Reporting Service, New Jersey Agricultural Statistics, August 1977.

fresh vegetable processing plant in Cumberland County in March 1976. This meant that 1,485 hectares (3,670 acres) of their own and 8,100 hectares (20,000 acres) under contract with 150 farmers would be released from their current land uses. All 1,485 hectares farmed by Seabrook are included in the test area except for a small area surrounding the town of Seabrook itself. Also included are privately owned farms, most of which operated independently of Seabrook; these are included as a comparison to the Seabrook land which had been completely devoted to vegetables.

The departure of one processor does not ordinarily signify a trend; however, in this case the reasons given for such action relate to regional and state problems and are similar in nature to those expressed by Ritter and Del Monte. It appears, then, that the Seabrook action, along with those of Ritter and Del Monte before it, are part of a larger agricultural change occurring in the state.

1.2. Experimental Design

The purpose of this study is to detect agricultural systems change by separating two agricultural systems, corporate vegetable farming and field cropping, in order that the transition from the former to the latter may be detected. This study traces land utilization over time in an area where two agricultural systems exist in close proximity, each imposing a different spatial pattern on the landscape.

With respect to "land utilization in time," crop calendars vary for each type of agricultural system. For example, vegetables are planted mainly during the warmer months, whereas winter grains are planted in the Fall. Table 4 shows the approximate periods of peak growth for a variety of crops grown in this area. It appears from this list that the actively photosynthesizing period for vegetables extends later into the Fall than it does for field crops. Winter grains are seeded in October and do not make an impact on land cover until early November. Thus, vegetables and field crops have somewhat different cyclical patterns, and this study attempts to trace these cycles over time.

In addition to the temporal component of land utilization, the spatial component must be considered. Field crops and vegetables impose different percentages of canopy cover on the soil background. The influence of exposed soil on spectral response is discussed below.

Surface features influence the amount of solar energy reflected into space from the earth to an earth-oriented sensor. Vegetable, soil, exposed rock, and water reflect energy differently in each spectral band. In the agricultural area under examination in this paper, vegetation and soil are the most dominant surface features. Although vegetation types differ in their spectral response curves, they do exhibit general characteristics distinct from other types of land cover. Actively photosynthesizing plants are charaterized by a relatively large amount of near infrared reflectance (Bands 3 and 4 on the MSS) and a relatively small amount of red reflectance (Band 2 on the MSS). Soils exhibit spectral patterns distinct from vegetation while also differing within soil

TABLE 4. PEAK GROWTH PERIODS OF FIELD CROPS AND VEGETABLES

	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Field Crops									
Wheat		X	X						X
Rye		X	X	X					X
Oats			X	X					X
Barley			X						
Soybeans					X	X			
Corn					X	X			
Hay-Alfalfa				X	X	X			
Vegetables									
Cauliflower						X	X		
Carrots				X	X	X	X	X	
Broccoli						X	X		
Spinach--Fall									
Spring	X	X	X						
Squash					X	X	X	X	
Kale--Fall							X		
Spring			X					X	
Baby Limas					X	X	X		
Green Beans				X	X	X	X		

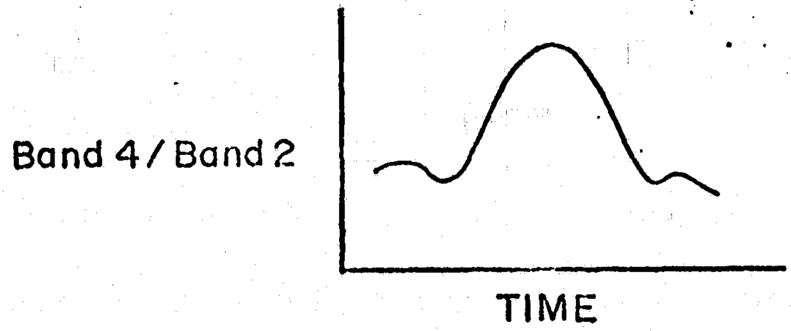
This table is compiled from information received from Seabrook Farms, Inc. and from published agricultural statistics of the New Jersey Crop Reporting Service.

categories. Generally, spectral response curves of soils are rather flat in the visible region with a decrease in reflectance in the near infrared. The reflectivity of soils in the visible is also dependent upon soil color--a lighter soil will be more reflective than a darker one. The coastal plain soils in the test area are rather bright and will therefore be very influential on the visible bands' spectral responses.

The increase in near infrared reflectance which occurs as green plants develop is associated with the maturation of the plant leaf. The reduced amount of reflected radiation in the red band is due to the absorption of light in the .6 to .7 micrometer region by chlorophyll-a, the major pigment involved in photosynthesis. Green light is absorbed as well as red but in lesser amounts. Thus, in the LANDSAT MSS data, a decrease in reflectance will occur in Band 2 with respect to Bands 1, 3, and 4 in pixels dominated by actively photosynthesizing vegetation. As a plant approaches senescence chlorophyll absorption decreases.

Because leaves reflect increasing amounts of near-infrared light as they mature and because the pigmented parts of the mesophyll reflect decreasing amounts of visible light as chlorophyll increases, the amount of reflectance in Bands 2 and 4 should be a good indicator of vigorous vegetative cover. In this study, the ratio of Band 4 to Band 2 is used as an indicator. Since an increase in Band 4 is associated with a decrease in Band 2, a ratio of the two bands is more sensitive to vegetation changes than, for example, one band alone.

Plotting the ratio of Band 4 to Band 2 over time for a specific ground area, yields a curve for each of the four years of the study, similar to the diagram below.



It is hypothesized that a change in the shape of this plot, such as a shift in the peak, from one year to the next represents a change in the agricultural system because of the different temporal growth patterns of vegetable and field crops. The year 1976 is particularly important in the Seabrook area because of the land use change described earlier; a sharp departure from the curves of the three previous years would indicate a rapid adjustment to the Seabrook Farms' decision to cease farming and processing in the area of study.

Although measurements of reflected energy are made at approximately the same time of day, the sun elevation at a specific latitude at a given hour changes throughout the year. To account for this variation, a sun elevation correction factor was introduced which adjusts all measurements to a solar angle of 90^{0}. In addition, the use of the Band 4/Band 2 ratio reduces solar angle differences as well as signal fluctuations and differences attributable to changing atmospheric conditions by using a relative measure, a ratio.

The analysis employs digital LANDSAT data and relies upon, as support materials, LANDSAT imagery, aerial photography, United States Geological Survey maps, Seabrook Farms Corporation maps, and ground-collected information.

Although LANDSAT data are available for passes made every eighteen days from July 1972 to January 1975 and every nine days from January 1975 through most of 1976, not all of it is useable, due to excessive haze or cloud cover or to technical problems in producing computer compatible tapes of the data. Out of 119 possible coverages, 26 acquisitions (or 22 percent) were found to be useable in this study. All 26 scenes are preprocessed for geometric correction and rectification to a Universal Transverse Mercator grid.

Imagery is used initially to choose suitable passes and to geographically identify the corresponding digital data. A LANDSAT image, or "scene" corresponds to a block of data over a geographic area of approximately 34,225 square kilometers (13,225 square miles), and each data tape contains information from an entire scene. (See Table 5 for a listing of LANDSAT data used in this study.) From each scene twenty-four smaller areas of 595 hectares (1400 acres), or "quads", are used as the basic unit of analysis. These areas may encompass up to 40 agricultural fields. It is felt that this unit is suitable for the long-run objective of the study--to simulate global monitoring. In addition, the unit conforms to

TABLE 5 LANDSAT DATA USED IN THIS STUDY

	DATE	SCENE IDENTIFICATION NUMBER
1973	February 13	1205-15141
	June 1	1313-15141
	July 7	1349-15134
	August 12	1385-15131
	August 30	1403-15125
	October 23	1457-15113
1974	April 3	1619-15083
	April 21	1637-15080
	June 14	1691-15063
	July 20	1727-15052
	October 18	1817-15023
1975	May 22	5033-14530
	June 9	5051-14520
	August 2	5105-14485
	October 13	5177-14443
	November 18	5213-14422
1976	March 5	5321-14352
	March 23	5339-14341
	April 10	5357-14330
	April 19	2453-14571
	April 28	5375-14314
	June 12	2507-14553
	July 18	2543-14545
	August 23	2579-14535
	September 28	2615-14525

technical constraints in the computer programs which are used in support of the project. Each set of twenty-four areas is registered spatially from date to date. Therefore the study concerns the change in reflectance of twenty-four identical areas over the course of four years--1973 through 1976.

An unsupervised classification routine developed at GISS is used to obtain a nominal spectral response for each quad. Normally, this routine is used to delimit clusters of pixels which have similar response curves. In this study, however, the object is to delimit one single cluster for each quad into which approximately 95 percent of all the pixels in the quad fall. This percentage was chosen to filter out highly reflective, exotic pixels which might influence the quad's spectral response disproportionately. The classification routine is applied to twenty-four quads from each of twenty-six passes to obtain 624 characteristic 4-band spectral responses. The reflectances for Bands 4 and 2 are ratioed for each of the 624 nominal spectral responses and then plotted against time for each of the 24 quads.

1.3 Preliminary Results

To date, classification has been performed on all of the 24 quads for each of the available 26 LANDSAT passes and the Band 4/Band 2 ratios calculated and plotted. Two sample plots are presented in Figures 2a and b for quads 2 and 23, which are representative of field crops and vegetables respectively, as determined from information collected in the field and maps from Seabrook Farms.

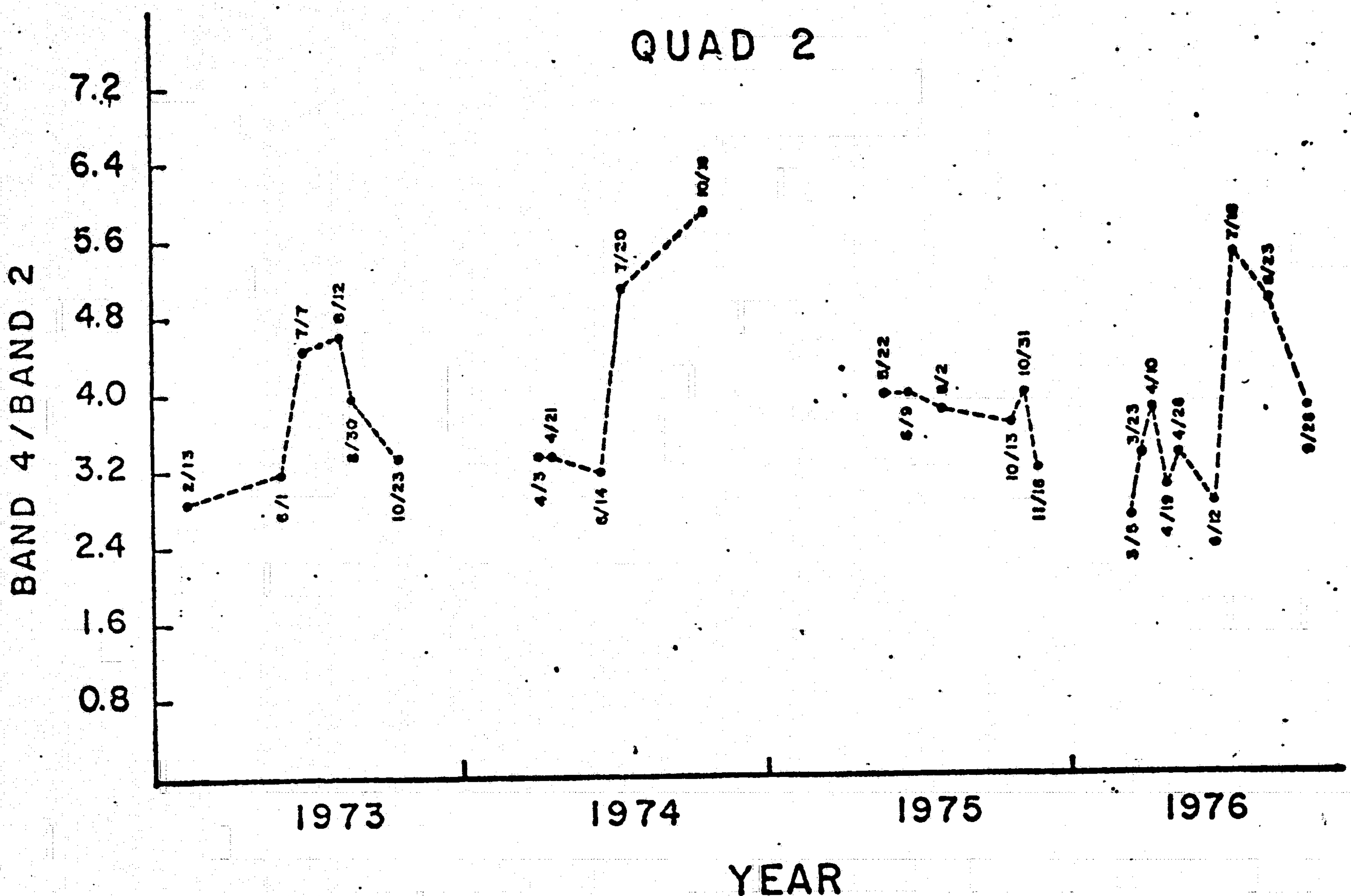


Figure 2a Quad 2. (field crops)

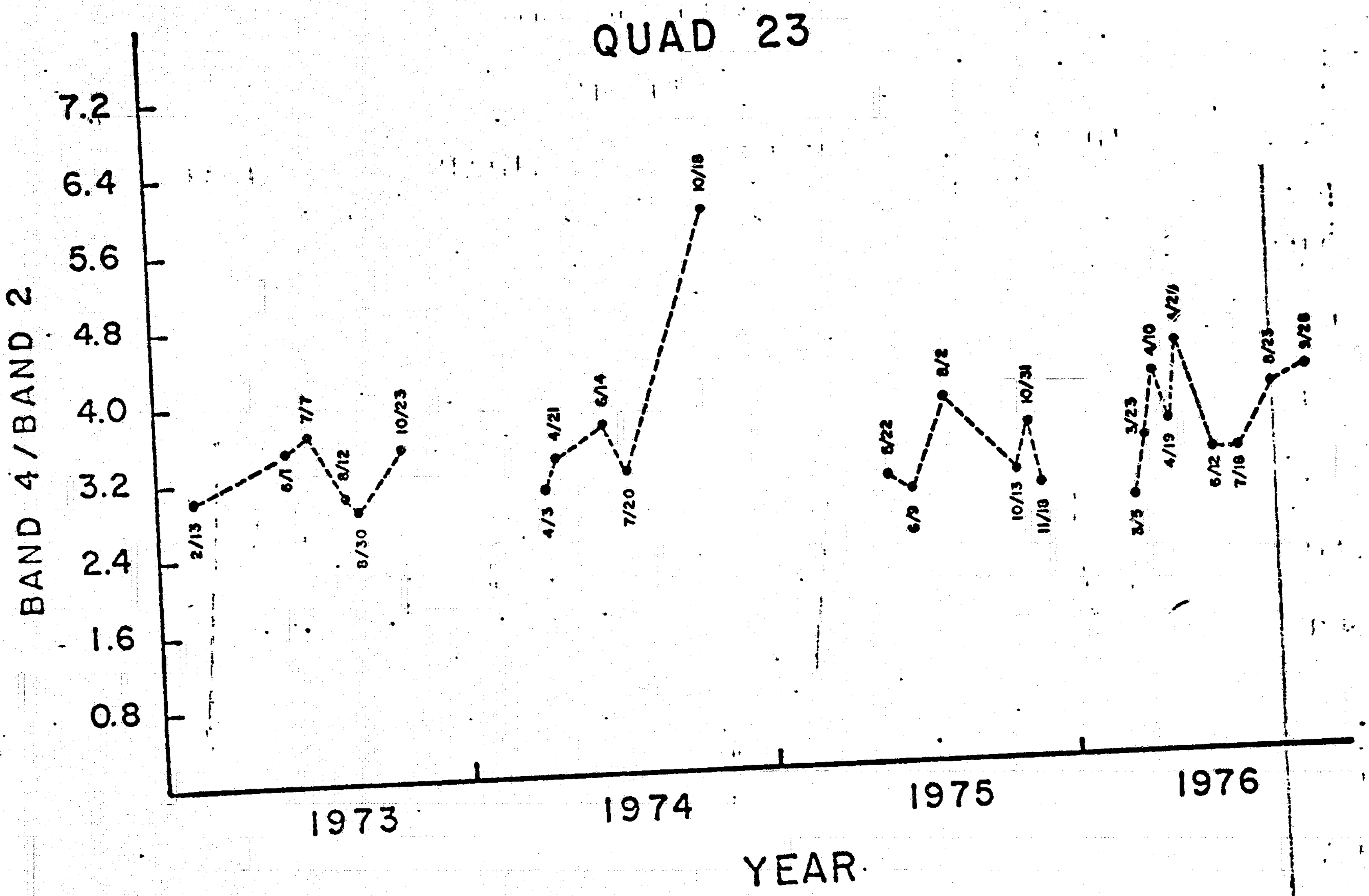


Figure 2b Quad 23. (vegetables)

The shapes of the curves differ from year to year because of a lack of concurrence in satellite acquisition times for usable data. The two quads can be compared to each other within each year, however. The curves diverge within the summer months when both field crops and vegetables are growing. Field crops tend to be in full canopy during July and August whereas vegetables tend to have less canopy cover with rows being separated by strips of soil. The large amount of exposed soil decreases the Band4/Band 2 ratio, and this is evident in the figure. Curve separation may also be related to meteorological events which have immediately preceded data acquisition. For example, the LANDSAT pass during the summer of 1975 occurred ten to fifteen days after a series of severe storms destroyed crops in the southern New Jersey region. In Figure 2, this event is reflected in the similarity of the two plots of 1975 during the summer, when they would be expected to diverge.

Initial analysis of the ratio plots for vegetable quads as compared with field crop quads indicates significant differences between the two agricultural systems. Additional statistical tests are being conducted.

2. East African Swidden Agriculture

An investigation into the feasibility of detecting shifting cultivation in East Africa has encouraging preliminary results, which are reported in detail in document 3 of the Appendix.

The study site is in the West Pokot District of Kenya. LANDSAT computer-compatible tapes for three dates are available:

scene 1067-07221, 9 September 1972; scene 1193-07230, 1 February 1973; and scene 2063-07111, 26 March 1975.

Supplementary data include 1956 black-and-white aerial photography and data from 1961-62 anthropological fieldwork.

Using the supplementary data in conjunction with gray scale printouts of geometrically corrected LANDSAT data, a preliminary study site of 595 hectares was selected. An approximately ten-hectare area in a region locally known as Asar always contains active cultivation and was thus selected as the training area. From this area, the most frequent spectral response in each band was determined, and this became the training data. The preliminary study site was classified, for each of the three available dates, into two categories: each pixel was or was not in the same category as that defined by the training spectral response pattern.

Distribution of points put into the category defined by the training data was well-correlated with the ecozone in which shifting cultivation occurs. The percentage of classified points falling within the ecozone was 77%, 82%, and 80% for the three dates. This led to the inference that the defined training class does represent shifting cultivation. Classification results indicate a significant increase in swiddening activity within the three-year span of the LANDSAT data used in the study.

The same training data were then used to classify a larger area (approximately 19,000 hectares). Results are being interpreted.

D. Effect of Environmental-Management Parameters on Computer Classification of Agricultural Land Use

Current research in computer classification of remote sensing techniques concentrates on acreage estimation of agricultural land use. The techniques used for agreage estimation identify and then separate radiance values of the scene into categories. These values are usually considered discrete elements of the environment, and no attention is given to their composition. Another way of examining the spectral response of any landscape is to view it as a composite of management and physical environmental characteristics, which make varying contributions to the landscape's radiance. It is the purpose of this research to assemble major environmental-management parameters into a model measuring their influence on spectral signatures of an agricultural landscape. The area of study is in a LACIE intensive study site in the northern Great Plains. The remote sensor data was acquired by the Johnson Space Center 24-Channel multispectral scanner, August 15, 1975.

The computer classification method used in this digital analysis consists of two major elements: 1) the ground truth system and 2) the radiance values of the landscape. The first element divides the scene into finite geographic units of varying sizes and shapes (i.e., agricultural fields, soil boundaries), and the second element represents the spectral response of the environment. It is the function of the classification algorithm to distinguish between and assign each pixel to its respective land use/land cover category. A change in the ground truth system results in classification of the same landscape into different categories. This may reveal a greater degree of homogeneity within a category, resulting in higher classification accuracies and a more thorough understanding of the spectral reflectances.

In agricultural studies, two categories of environmental parameters are identified: management and physical environment. Management consists of such farming factors as tillage direction, fertilization, planting date, and irrigation. The physical environment category includes elements such as soil type, soil color, and terrain slope. The relative influence of a set of parameters on land use classification is unique for each landscape.

The study region for this investigation is a 5 x 7.5 kilometer area of Williams County, North Dakota, where severe winters and dry summers predominate. Spring wheat is the major crop in a landscape intermixed with summer fallow and pasture fields. Summer fallow is practiced almost exclusively as an anti-wind erosion measure in strip cropping patterns. Pasture includes harvested and non-harvested oats and other natural grasses which may or may not be grazed. The agricultural practices are influenced to some extent by remnant glacial features, such as potholes and till knobs. Potholes create areas of poor drainage which become marshy during stages of evaporation after the rainy season. Sections of patchy bare soil (till knobs) produce a mottled effect in the appearance of the landscape.

The management practices and physical environmental factors of this area comprise a landscape system of varying spectral responses. Explanation of the spectral variability will be attempted through the testing initially of three environmental/management parameters as separate ground truth systems: tillage direction, planting date, and soil type.

To date, the three ground truth files have been constructed in preparation for computer classification of the multi-spectral scanner data. The ground truth management system used in this study is a computerized geographic information system developed in cooperation with GISS.[1]

Procedures for ground truth file construction differed for management and physical environmental parameters. The expression of management parameters (e.g., tillage direction and planting date) in the landscape are on a field basis. Therefore each management parameter was input using a field number ground truth array. Crop cover designations for each agricultural field were replaced by tillage direction and planting date. Information regarding tillage direction and

[1]See Coiner, J. C. and Ungar, S. G., "Ground Truth Management System to Support Multispectral Scanner (MSS) Digital Analysis." Proceedings of the ASP-ASCM Joint Annual Meeting, Washington, D. C., February 27 - March 5, 1977.

planting date was extracted from LACIE ground truth inventory forms (June 1975), except for tillage direction in summer fallow fields, which was acquired through interpretation of 1:20,000 RC-14 color infrared aerial photography (Aug. 15, 1975).

By contrast, physical environmental parameters are spatially independent of field boundaries. Thus, a new ground truth file was constructed for soil type using a 1:126,720 soil survey map of Williams County from the North Dakota Soil Survey Report (Bulletin No. 473, July 1968).

Continuing work on this experiment involves computer classification using each of the 3 ground truth systems in order to determine the relative influence of each environmental-management parameter on classification accuracy for the land cover classification categories. It is expected that each parameter will have differing magnitudes of influence, both within and between each category; however, the degree of distinguishability is not anticipated.

III. Other Activities

Staff members of the Columbia Remote Sensing Project were invited to participate in a conference - workshop, "The Use of Landsat Data in the Analysis of Non-Western Ecosystems", sponsored by the Wenner-Gren Foundation for Anthropological Research, New York, May 23 - 26, 1977. Each participant was selected to fulfill an integral role within the conference, which was designed to bring together anthropologists, geographers, ecologists, and technological experts to deal with the topic. Among the working papers prepared in advance was a discussion of data (appended as document 1 to this report) prepared by Dr. Jerry Coiner of the Columbia group. Additional participants were: Prof. Colin High, Tina Cary, Connie Cambria, and Helene Wilson.

APPENDIX

Working paper prepared in advance for participants in
Conference/Workshop on

THE USE OF LANDSAT DATA IN THE
ANALYSIS OF NON-WESTERN ECOSYSTEMS

Francis P. Conant, Organizer

May 23 - 26, 1977

Not for Publication

INFORMATION SYSTEMS CONCEPTS FOR INTEGRATING LANDSAT AND OTHER DATA SOURCES

Jerry C. Coiner

WENNER-GREN FOUNDATION FOR ANTHROPOLOGICAL RESEARCH
14 East 71 Street
New York, New York 10021

INFORMATION SYSTEMS CONCEPTS FOR INTEGRATING LANDSAT AND OTHER DATA SOURCES

Jerry C. Coiner
Department of Geography
Columbia University
New York, New York 10027

INTRODUCTION

LANDSAT data availability for nonwestern ecosystem studies is a major addition to traditional data sources, thereby impacting traditional analyses. In the past, most studies of nonwestern ecosystems have been place specific and descriptive, with normative theory derived from localized observation sets. LANDSAT data are multiple-place specific, representing integrated coverage of large areas of the Earth's surface. In addition, they represent physical properties of the Earth's surface and have not been "filtered" through the perceptual mechanism of an in-situ observer. Because of these characteristics, three problems are created for the researcher:

(1) LANDSAT data must be subjected to analysis before they can be considered information directly bearing on specific research questions.

(2) Data from LANDSAT available for analysis are multiplied several orders of magnitude over that obtained conventionally, thereby requiring new data management techniques not normally associated with place-specific research.

(3) LANDSAT data call for a mechanism to allow their association with traditional data so that analysts can conduct fully-integrated, multiple source research.

The second and third problems will provide the focus of this paper, which suggests that both the volume of data and the integration of multiple data sources require development of geographic information systems to conduct research incorporating LANDSAT data.

An information system can be defined as "a group of entities and activities meaningfully connected and satisfactorily bounded, which interact for a common purpose or purposes" (Thomas and Shafer, 1970). More importantly, the function of an information system is to organize data and process them to information in such a way as to facilitate the analysis phase of research. This is accomplished by a series of component processes within the information system: specification, collection, transmission, processing, display, dissemination and use (Steiner and Salerno, 1975). Some of these are self-evident in terms of human ecosystems research. However, certain aspects of the information processing sequence are not self-evident and will be elaborated further.

A geographic or spatial information system is a special type of information system. It is unique in that it retains the spatial context of the data being input to the system, allowing the user to interrogate the system not only in terms of substantive data properties (How large? How many? What kind?), but also in terms of location (How many at place A? What kind at place B?). Spatial data are needed because through the concept of place, it becomes possible to integrate multiple types of data of the kind LANDSAT and traditional anthropologic resources generate. Whether dealing with traditional ethnographic descriptions or sets of radiance values from a LANDSAT pixel, both share two properties. They were acquired

(1) at a specific location, and

(2) at a specific time.

These two universal data characteristics form the cornerstone upon which the information system is built and must be maintained to allow integration of the data from multiple sources.

I. Sources and Types of Data

Information systems are dependent upon data that can be taken from numerous existing sources as well as new data sources. In terms of human ecosystem research, this means that the system would have to deal with classic descriptive literature or qualitative data as well as various types of quantitative (numeric) data. The problems of classical descriptive literature arise from the fact that condition and perception are not well separated. For example, to say that the annual precipitation of a location is 200 mm is descriptive of a condition, whereas to say that the rainfall is sufficient to allow flooded paddy rice to be grown in the lowlands is a perception. The former is easily subjected to information systems constraints because it is quantifiable and free of ambiguity, while the latter poses a difficult problem for integration into the system's data base because the condition inherent in the perception cannot be easily separated from the researcher's individual assessment and unambiguously quantified. Integrating different types of data is further compounded by a basic dichotomy between quantitative and qualitative data. Quantitative and qualitative data may reveal two different aspects of the geographic place by forming two different sets of data--quantitative, which describes the universal aspects of place; and qualitative, which emphasizes uniqueness to place. This makes the integration of traditional qualitative data with universal data sources, in this case LANDSAT, more difficult, because they address entirely different aspects of the same geographic location.

In an information systems context, data are normally thought of as quantitative, or essentially, observations that can be reduced to numbers. This reduction to numbers implies measurement. Measurement can be done on four different scales. In general, the higher the scale of measurement, the more information the observation contains. The simplest level of measurement is classification, or a <u>nominal</u> scale. This is done by establishing classes and assigning each observation to a class. For instance, if in a number of regions the predominant agricultural activity is either swidden or sawah, each region can be assigned to a predominant agricultural system and a count can be made of the number in each class.

The next level of measurement is the rank order, where additional information is available to assign each observation to a rank relative to other observations. Because the relationship of each observation is relative to other observations in the set, the observations are measured on an ordinal scale. If a precise numeric value can be assigned to distances between observations, again the informational content of the measurement has increased, and an interval scale can be employed. This interval scale allows true measurement of observation (as does the ordinal scale), while the nominal scale provides only the ability to count observations. Interval scale measurements can be further refined to ratio scales if the interval scale begins at true zero (Hammond and McCullagh, 1974, pp. xiv-xvi).

Types of quantitative measurements used to form the data base of information systems will determine interrelationships between data. For instance, LANDSAT ratio scaled radiances lose much of their informative power it only compared to nominally scaled ground collected data. Arguments presented here lead to the speculation that types of data collected to support geographic information systems incorporating LANDSAT-based analyses will require ground investigations which collect quantitative data with higher levels of measurement information (scaling).

At present, major scientific disciplines collect, store and analyze data about their areas of specialization, but incongruities arise when multidisciplinary studies are attempted and integration of data is needed from the social, earth, and natural sciences. Incompatibilities between the disciplinary data bases occur because the paradigms of the disciplines adhere to assumptions which tend to create unique and nonoverlapping data bases when each discipline focuses on the same real-world object, including geographic space.

II. Geocoding

Traditional static methods of displaying spatial relationships may be limited as tools to analyze human ecosystems. In dynamic situations of ecosystems, it is obvious that a static data base, of the type represented by a single land cover map, cannot provide all the data necessary for successful research analysis. From this inadequacy emerges the requirement for large scale geographic information systems to support research. One source of data, which encompasses both human infrastructure and numerous aspects of the physical environment while retaining spatial orientation, is aerial photography and more recently, remotely sensed data from satellites such as LANDSAT I and II. However, in its raw form, remotely sensed imagery, whether from aerial cameras or satellite scanners, does not represent an analytical base. It must be subjected to a systematic approach that includes collection, analysis, storage, retrieval and display, i.e., it must be formulated and processed into a geographic information system.

The most commonly used method of storing environmental and spatially extensive socio-economic data, particularly that derived from aerial photographs and LANDSAT, is the cellular or equal area grid system. It consists of overlaying on the landscape a series of parallel lines oriented perpendicular to each other. A Cartesian coordinate system is thereby formed where the intersection of each perpendicular pair defines an x,y coordinate. Tomlinson describes the systems that are dependent on this type of geocoding, of which the most widely known is the New York State Land Use and Natural Resources Inventory (Tomlinson, 1972, pp. 635-1200; Belcher, et al., 1971). Grided data can be presented in small scale maps, as was done for the Minnesota Land Use Inventory, while other systems store grided data for use with computers, such as the University of Wisconsin Institute for Environmental Studies REMAP Systems. Also, the Harvard Computer Graphics Laboratories SYMAP series of programs operate using grid data.

A product similar to one generated by these systems is shown in Figure 1. This land use map was produced by storing for each x,y location the generalized primary land use (most area extensive) found delimited in a 10 acre cell to the lower right of the township/range reference. By using this storage procedure for all the data elements collected, the area analyzed is described in terms of the status of each data element at the same point on the Cartesian coordinate system. Thus, a data vector, the z dimension, is recorded for each x,y location, with the length of the vector dependent on the number of data elements specified and computer hardware available. One or a number of the z components can be made up of classifications derived from LANDSAT data.

Cellular geocoding offers the advantage of being subject to the same type of analytical methods as quadrate approaches widely used by ecologists. Simple non-parametric techniques, such as Chi-square and measures of association can be applied by both natural and social scientists to grided data elements, allowing analysis of change that occurs within the landscape when measured on a cell by cell direct comparison.

Probably the greatest weakness of cellular arrays lies in the fact that they necessarily abstract the landscape at the data acquisition stage. The degree to which data are abstracted varies with the grid system employed and should, in theory, be related to the resolution of detail to be extracted from the landscape. However, grid cell size is usually a design specification based on external constraints, not on the data to be stored. Consequently, under and over-generalization of the landscape for specific utilizations can occur. This is not a serious restriction in the case of undergeneralization, because the cells can be aggregated and data elements factored, but overgeneralization is a serious problem. It would require returning to the basic data source and reinterpreting each data element to correct the deficiency.

One of the strongest features of a cellular array is the locational accuracy and the flexibility which the Cartesian coordinate system permits, i.e., the ability to transfer locational information from geo-coordinate systems (latitude, longitude) to geographic reference systems, such as the Universal Transverse Mercator (UTM). This allows handling of data on large areas, and in certain cases, where geographic abstraction above the cellular level can be tolerated, cells can be combined to reach the appropriate geographic scale for analysis.

A criticism that has been leveled at the equal area grid or cellular approach is that it does not retain the true boundaries between activities occurring on the Earth's surface. If we were looking for diffusion and expansion of certain types of land use, or if we were considering the relationships between vegetation communities which responded to precise environmental controls, e.g., soil types, boundaries would be important.

The Canadian Geographic Information System has been developed which uses digitizers that acquire data about boundaries within the landscape for storage in a land use information system. Figure 2 is a flow diagram for the Canadian system and shows the two components of data recovery, the map-like plot detailing boundaries of the land uses and the data table listing the land uses and their areas. As the table indicates, land use units have different sizes because they are delimited by boundaries recorded as they naturally occur. Boundaries are acquired from the landscape by digitization in terms of a Cartesian coordinate system, while the land uses themselves are interpreted and stored separately.

Although the Canadian system presents a realistic rendition of the boundary conditions as they are perceived in traditional map form or in an aerial photograph, there is no indication that this approach retains a more accurate analytical frame than does the cellular data. Comparison of Figures 1 and 2 shows that the two geocoding systems describe the landscape in disparate ways, since the emphasis in cellular data is on what happens at each place and in boundary data, on the morphology of the phenomena delimited by the boundary. Under these conditions, no comparison between the two systems appears possible. Further research is warranted if place states are to be analyzed in a boundary recovery system.

The major weakness of boundary geocoding is that it does not permit equal area analysis to be conducted. This means that the data will have to be reformed (generalized) into equal area units (cells) prior to the use of analytical techniques in either the earth and natural sciences or in comparative factorial ecology. In general, fewer methods of analysis exist for considering unequal areas than equal areas. Therefore, changes in boundary patterns must be assessed on a relatively qualitative basis at present.

Spatial data can also be stored employing the concepts of topology, the study of geographic configurations. It is possible to approximate the spatial appearance of a universe by defining a series of links and nodes. Human transportation networks are ideally suited for demarcating the landscape as a topologic graph, with routes forming the links and intersections the nodes. Surface transportation networks are the building blocks upon which topologic systems for storing geographic data are constructed.

To establish a spatial unit for topologic geocoding, areas are circumscribed by a series of links, usually the surface transport routes in the study site, with node points where the boundaries of the area change direction. Clearly, units of unequal area may result.

Probably the most commonly used and widely known topologic system is the GBF/DIME File developed by the U.S. Bureau of the Census. The DIME file is based on nodes and links which are essentially defined by the surface transportation network within the area for which the file is developed. Figure 3 is a small portion of a DIME file taken from the Census Use Study Manual. The file, like a boundary file, does not demark equal areas and is essentially a generalized replication of the human transport network within the landscape. DIME files are created for specific U.S. metropolitan areas defined by the standard metropolitan statistical areas (SMSA's), but the files are not contiguous for the entire U.S. and thus, they are not acceptable for studies requiring analysis of urban areas and their hinterland or rural areas alone.

The DIME system's main advantage is that it defines relatively small human population units (approximately 300), which are to a high degree invariant. Its major weakness is that in achieving invariance in population, high variance in area per sampling unit results. From the point of view of acquisition of socio-economic data and its analysis, the GBF/DIME file is an excellent data base, in that it can handle both socio-economic area statistics, such as the population and income of population within a census block, and network and flow concepts related to communication. However, from the point of view of equal area and continuous coverage over large geographic units, the file suffers from its very nature. That is, the block or enumeration districts within the unit are not spatially uniform (equal area) and the data collected in this manner are not subject to analysis via techniques similar to those used by ecologists. Although not generally a fault of the file, data may also vary in areal unit specification from Census to Census, therefore creating aggregation problems for time series analysis (Coiner, 1975, pp. 52-60).

The three types of geocoding described here have all formed the spatial framework of information systems used in conjunction with LANDSAT data. But the question that remains unanswered is what do each of these landscape generalizations discard from the actual landscape and how important is this to an understanding of human ecosystems existing on the landscape?

III. Example of Data Storage, Retrieval and Display

A geographic information system, designed specifically for a Third World site, Marinduque Island, Republic of the Philippines, is presented here to give an example of geographic data sensing in an information systems context. The experimental information system for Marinduque Island allows the collection and maintenance of spatial relationships inherent in the landscape through generalization of a preselected number of data elements into a uniform grid cell (25 hectares or 60.25 acres). Storage is achieved by arraying the data elements as records, based on the grid cell's location within the predefined UTM geographic reference system. Either aggregated data, categorized and combined into tables, or spatially displayed data in the form of computer maps, can be recalled for analysis and planning support. Programming allows presentation of discrete symbology or choropleth categories in the computer generated maps (Figure 4). Display of either symbols or patterns can be accomplished in multiple colors. In the initial tests, an IBM 370/145 was the basic computer hardware and was supplemented by an off-line CALCOMP drum-type plotter for automated cartographic products.

The Marinduque data set consisted of three time series, 1948, 1967, and 1972, taken from aerial photographs. Each of the data series were divided into two parts, 1) a land use classification, and 2) selected point data related to human settlement infrastructure. Data elements were assigned to three categories by spatial type, a point (e.g., house), a line (e.g., road) or area (e.g., mangrove swamp). Each spatial type was then analyzed relative to changes over time within types and between types.

Land use categories interpreted from aerial photographs of Marinduque Island were originally developed as part of an effort to support agrarian reform throughout the Philippines, particularly Central Luzon. Because of its original purpose, the classification required slight modification and supplementary land use categories to describe the Marinduque landscape. Of the land reform project categories, twenty-one were identified as area extensive on Marinduque (Table 1). A numeric code was assigned to aid in storing the land use data on computers. The classification logic used in the three-digit computer code was as follows: hundredths digit--general land use type; tenths and lower digits--specific land cover. To mark areas that had similar land uses, a photo symbol code was applied to the photographs, a necessary step when data takeoff was a separate operation from photo interpretation.

Studies of land use dynamics indicate that area extensive land cover categories, seen in Table 1, can have large cyclic components which do not reflect change wrought by development. These cyclic components may be induced by such environmental factors as long term climatic variations and competition between plant communities. In an effort to retain cyclic components in the data and to overcome intrinsic generalization of land use within a grid system, data were acquired not only on the most area extensive land use but also on the second and third most extensive in each cell. Also, certain types of land uses were not treated as area extensive and were recorded separately. The latter were considered part of the selected development data in Table 2.

Selected data were collected to determine infrastructure changes thought to be associated with development, i.e., changes in the number of dwelling units, schools, churches, industrial and governmental facilities, the condition of transport facilities, and the numbers and location of alternate sources of food, such as fish ponds and fishing boats. These additional data elements were thought important to avoid distortions created by employing land use data alone to measure change.

For planning purposes, the information system was operationalized in the following way. Each of the 4,008 twenty-five hectare cells, which when combined represent a virtual map of Marinduque Island in the data base, were studied for changes in and between various development variables. These changes were measured and tested using nonparametric statistics to determine the effect of specific infrastructure additions or deletions (e.g., roads, school buildings, industrial facilities, etc.), on other infrastructure components and proximate land use. Since the grid was spatially registered through time, change occurring within the cell could be positionally related, allowing assessment of the change's impact.

Studies now underway are intended to develop methods for more continuous updating of the Marinduque data base with LANDSAT data. The primary effort will be to develop for each twenty-five hectare cell a single indicator of change as derived from UTM preregistered and aggregated LANDSAT data. Anomalous or unique changes over time could thus be detected and corroborated either with ground reconnaissance or high resolution aircraft imagery.

Employing an experimentally designed geographic information system that used remotely sensed data from Marinduque Island, Philippines in the form predominantly of aerial photography, it has been shown that changes associated with human settlement can be monitored. The system allowed the identification of change and its location, and through analysis of this change, development was measured in a spatial context (Coiner, 1976).

SUMMARY

This paper has provided background on the concepts of geographic information systems, the types of data they require, and the methods used to implement a specific system now supporting human settlement research.

It has been widely documented that LANDSAT data can be readily structured into an information system context, but it has yet to be demonstrated what aspects of traditional data about human ecosystems can be encompassed in the same information system. The best hope of integration seems to rest on the commonality of place. Provision should be made in future field work to increase the informational content of field measurements, possibly leading to better congruency between LANDSAT and field data.

Epilogue

Scientific hypotheses are commonly held to be data dependent, reflecting the types of data available to the investigator. To date, the new data source (LANDSAT) seems to have been applied to research questions borne from other data sources. Should not effort be expended now in determining what research questions about human ecosystems should be asked in the context of LANDSAT?

REFERENCES

Belcher, D. J., R. L. Shelton, and E. E. Hardy, 1971, New York State Land Use and Natural Resources Inventory, Final Report, Cornell University.

Coiner, J. C., 1975, "Geographic Base File Assumptions for Human Settlement Studies," in Geocoding Applications, University of Colorado, Denver.

Coiner, J. C., 1976, "Geographic Information Systems in Development Planning," paper presented Annual Meeting, Association of American Geographers.

Information Canada, 1973, The Canadian Geographic Information System: An Overview, Ottawa, Canada.

Hammond, R. and P. S. McCullagh, 1974, Quantitative Techniques in Geography, Clarendon Press, Oxford.

Merchant, J. W. and B. H. Waddell, 1974, The Use of High Altitude Aerial Photography of Inventory Wildlife Habitat in Kansas: An Initial Evaluation, Center for Research, Inc., Technical Report No. 2230-14-1.

Steiner, D. and A. E. Salerno, 1975, "Remote Sensor Data Systems, Processing, and Management," in Manual of Remote Sensing, The American Society of Photogrammetry, Falls Church, Virginia.

Tomlinson, R. F., 1972, Geographical Data Handling, Vol. I and II, International Geographical Union, Commission on Geographical Data Sensing and Processing, Ottawa, Canada.

Thomas, E. N. and J. L. Shofer, 1970, Strategies for the Evaluation of Alternative Transportation Plans, National Cooperative Highway Research Program Report No. 96.

U.S. Bureau of the Census, 1970, The DIME Geocoding System, Census Use Study, Report No. 4.

TABLE 1. LAND USE CLASSIFICATION FOR MARINDUQUE ISLAND, PHILIPPINES

Computer Coding No.	Photo Symbol	Land Use
101[a]	Ar	Land intensively used for rice
102	Ac	Corn, beans, forage crops, upland croplands
104	Aom	Mixed orchard (backyard orchards around homesites)
105	Aoc	Coconut orchard
108	Ai	Fallow lands
109	Aq	Aquatic lands (commercial fish ponds, salt beds, moss production, oyster beds)
112[b]	Am	Mangrove swamps (to include nipa palm areas)
201[c]	Fr	Natural forest stands
202	Fb	Forest brushland (go-back, *parang*) and grassland (*cogon*)
203[b]	Fg	Natural bamboo (include all riverine tree vegetation)
301[d]	Rh	High density greater than 10 houses per centimeter on 1:15,000 air photo
302	Rm	Medium density 5 to 10 houses per centimeter on 1:15,000 air photo
303	Rc	Road strip and cross road community
503[be]	Es	Strip mining
605[f]	Pm	Military base and governmental centers
706[g]	Ta	Airport
901[h]	Ns	Sand/beach
902	Nr	Rock, exposed bare earth
903	Wn	Fresh water lakes and ponds greater than 7.5 hectares
904	Wr	Permanent streams and river (to include the entire width of river courses)
906[b]	Wn	Open salt water

[a]100 series, agriculture
[b]Added land use category for Marinduque study
[c]200 series, forest land
[d]300 series, residential land use
[e]500 series, extractive industry land use
[f]600 series, public and semi-public land use
[g]700 series, transportation land use
[h]900 series, non-productive land use

TABLE 2. SELECTED DEVELOPMENT DATA ELEMENTS COLLECTED PER 25 HA. CELL

Number of occupied houses (dwelling units)

Number of schools

Number of churches

Number of government buildings

Number of buildings associated with industrial activity, including oil tanks and storage areas

Number of commercial buildings, including *sari-sari* store front houses where identifiable

Number of market places

Type of roads/trails
- 0 = no road or trail
- 1 = all-weather road
- 2 = seasonal road, usually passable to wheeled vehicles
- 3 = trail, not normally usable by wheeled vehicles

Number of bridges

Number of boats

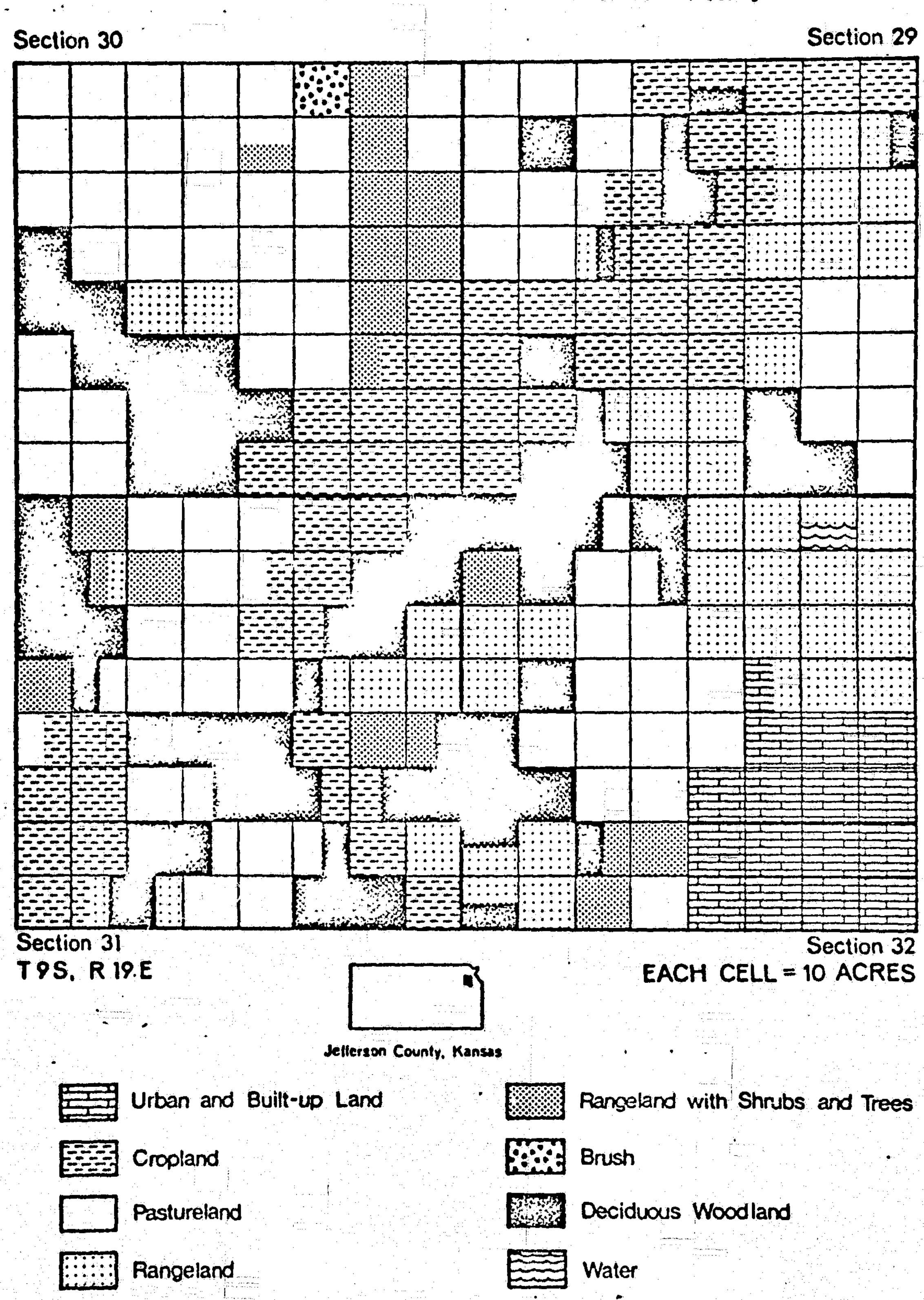
Number of piers or boat landings, beaches

Number of concrete drying platforms or basketball courts

Number of fish ponds

Number of fish traps

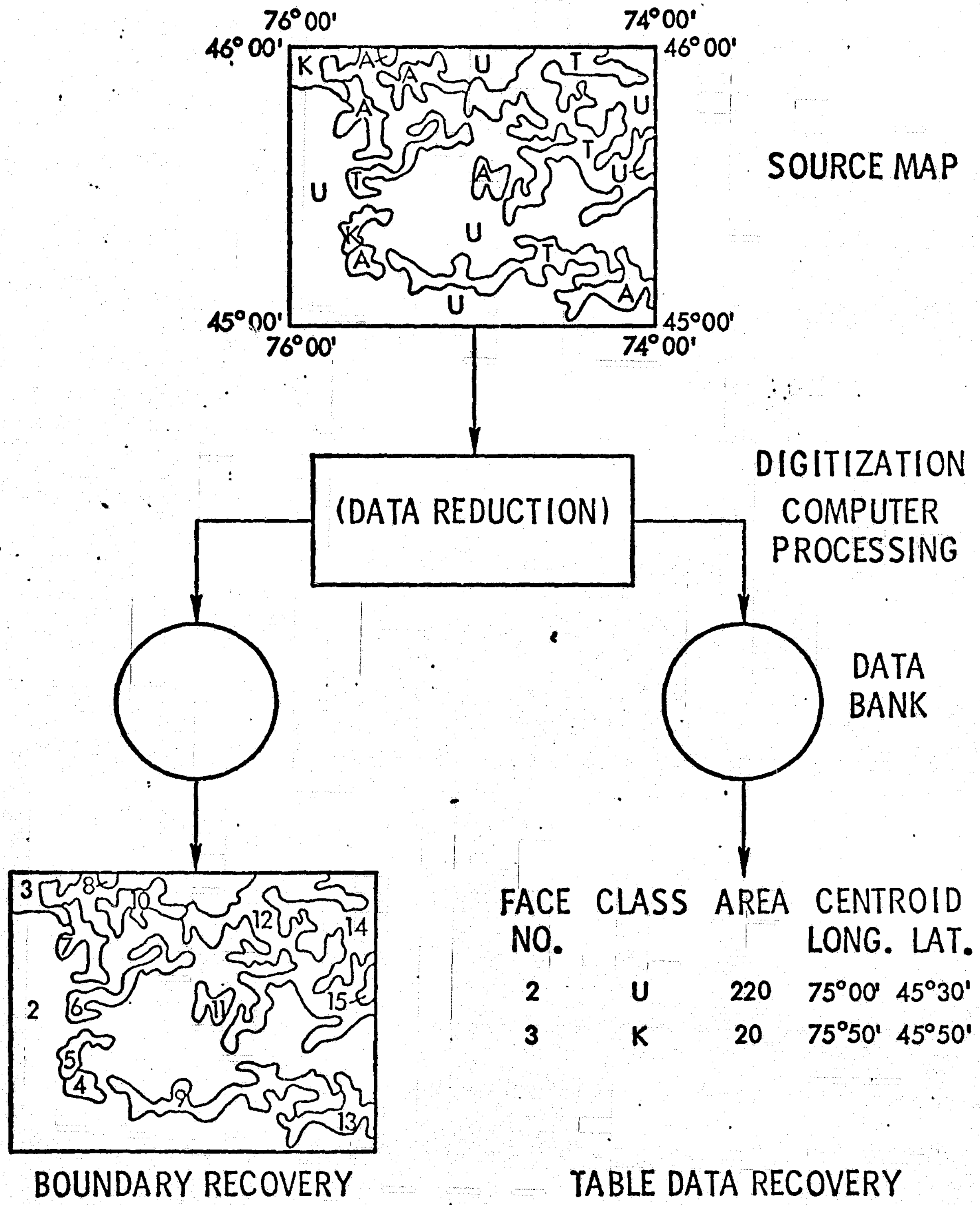
FIGURE 1. SIMULATED DIGITAL MAP: AN OUTPUT PRODUCT OF A GRIDED ARRAY GEOGRAPHIC BASE FILE.



AFTER MERCHANT AND WADDELL, 1974.

ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 2. BOUNDARY STORAGE AND RECOVERY

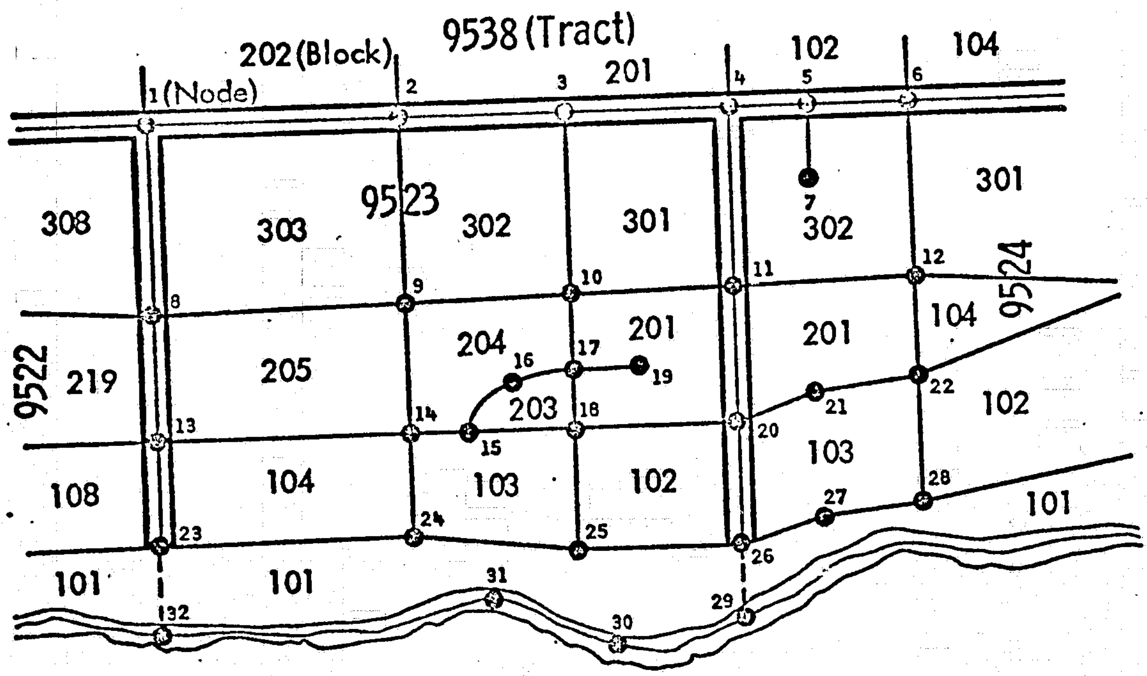


FACE NO.	CLASS	AREA	CENTROID LONG.	CENTROID LAT.
2	U	220	75°00'	45°30'
3	K	20	75°50'	45°50'

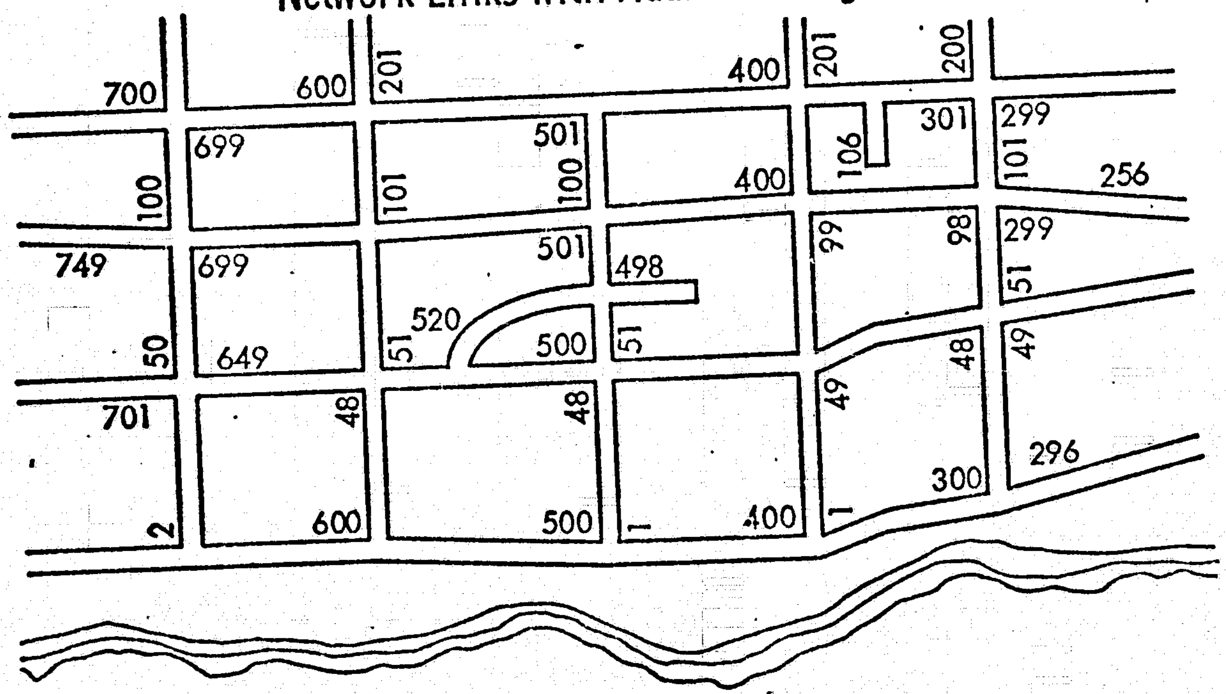
(AFTER, INFORMATION CANADA, 1973)

FIGURE 3. GBF/DIME TOPOLOGY

MONOPOLY HEIGHTS (a)
Tract-Block Structure with Network Nodes

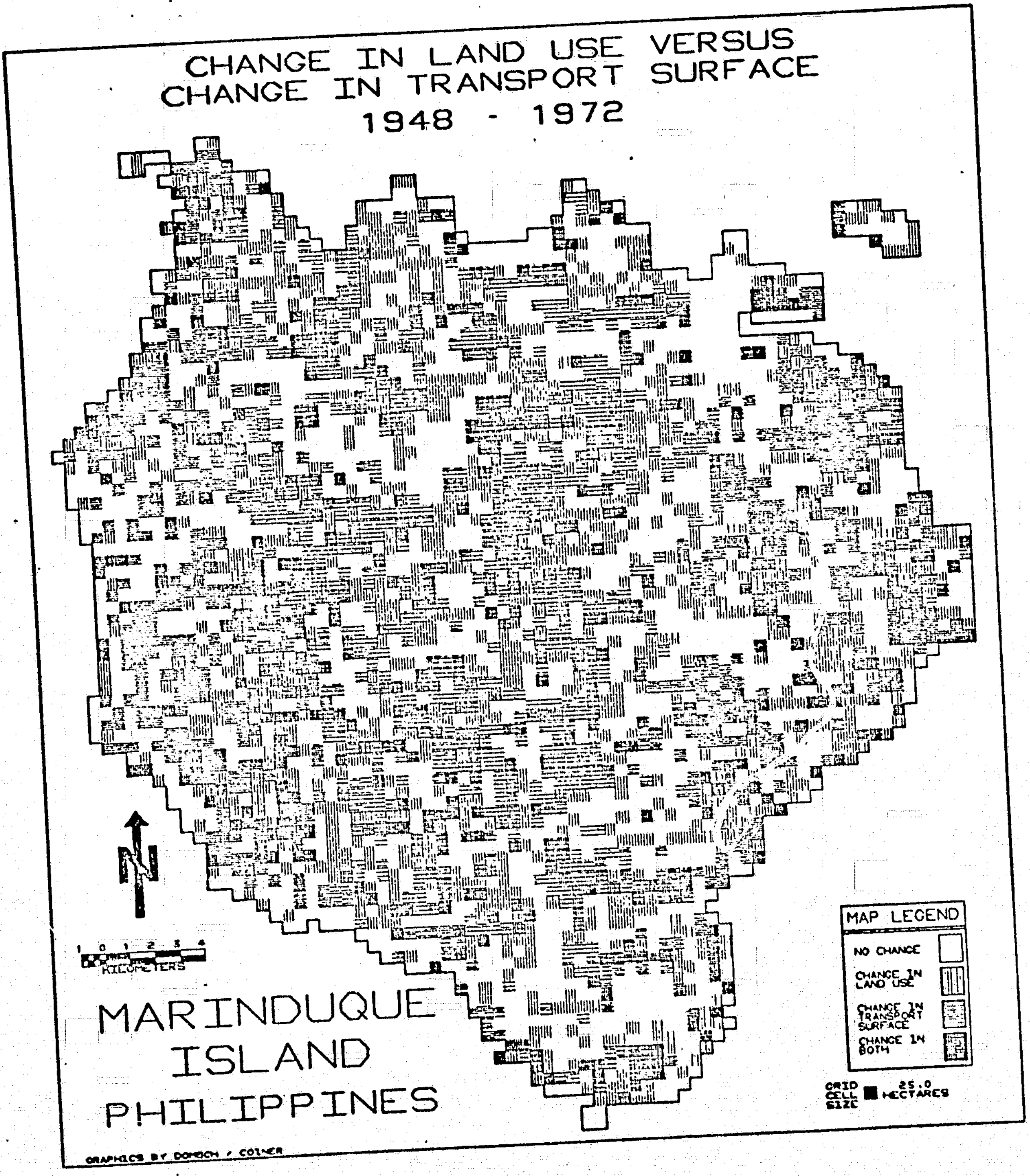


MONOPOLY HEIGHTS (b)
Network Links with Address Ranges



ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 4.


CHANGE IN LAND USE VERSUS
CHANGE IN TRANSPORT SURFACE
1948 - 1972
MAP LEGEND
NO CHANGE
CHANGE IN LAND USE
CHANGE IN TRANSPORT SURFACE
CHANGE IN BOTH
GRID CELL SIZE 25.0 HECTARES
1 0 1 2 3 4
KILOMETERS
MARINDUQUE ISLAND PHILIPPINES
GRAPHICS BY DOMSCH / COINER

ATLAS OF SELECTED CROP SPECTRA

Included here are Part I (Introduction and Key to Atlas) and Appendices A - E.

ATLAS OF SELECTED CROP SPECTRA
IMPERIAL VALLEY, CALIFORNIA

JUNE 1977

Prepared By

Stephen G. Ungar	NASA, Institute for Space Studies
William Collins	Columbia University
Jerry C. Coiner	Columbia University
Dwight Egbert	General Telephone and Electronics
Richard Kiang	General Telephone and Electronics
Tina Cary	Columbia University
Peter Coulter	General Telephone and Electronics
Nurit Landau	General Telephone and Electronics
Elaine Matthews	Columbia University
Stephen Lytle	Columbia University
Katie Prentice	Columbia University
Nancy Lytle	Columbia University
Alex Rodriguez	Columbia University
Jack Flamholz	General Telephone and Electronics
William Beck	Dartmouth College
Nancy Wasserman	Dartmouth College
Duke Angier	General Telephone and Electronics
Scott Lydiard	Columbia University

Prepared in part under NASA Grants NSG-5080, NGR 33-008-191, NSG-5014, and NASA Contract NAS5-20749.

TABLE OF CONTENTS

CHAPTER I.

Introduction and Key to Atlas

The purpose of this publication is to make available to interested investigators high spectral resolution reflected radiance data for representative agricultural landscapes. These spectral data are accompanied by supporting ground observations and interpretation details from simultaneously acquired aerial photographs. Thus, this Atlas provides a single reference source for representative spectral reflectance data covering a variety of precisely described agricultural crops, crop conditions and tillage states.

The spectral reflectance data represent measured solar radiant energy reflected from the crop surface (acquired at nadir and averaged over an 18 by 18 meter area) throughout the visible and near infrared portion of the electromagnetic spectrum (from 430 to 1000 nanometer wavelength in 410 bands of 1.4 nm width). These data are presented graphically, with each set of plots representing a homogeneous ground surface type.

Spectra of agricultural landscapes presented in this Atlas were selected from a larger data set. Each set of spectra describes a specific homogeneous crop, crop condition or tillage state. To meet the criterion of homogeneity, roads, fence lines, homesteads and other such landscape elements were avoided.

Each page of the Atlas presents a set of spectra representing a single homogeneous agricultural landscape. In principle, spectra from different fields can be aggregated to create a single Atlas entry for a specific crop, crop condition or tillage state. However, difficulties in aggregating spectra acquired under different observing conditions (e.g., sun elevation, aircraft attitude) have constrained the agricultural landscapes presented here to parts of fields. In some cases, more than one set of spectra appear in the Atlas for a specific crop state. In these instances, differences are attributable entirely to changes in observing conditions and each spectra-set is equally representative of the agricultural landscape described. In all cases, the measured standard deviation within a set of spectra is less than ten percent of the mean value at all wavelengths.

The spectral plots constitute the bulk of this Atlas. Appendices provide detailed information to the reader interested in specific aspects of data acquisition and analysis procedure. Appendix A describes the airborne spectroradiometer instrument and associated calibration procedures used in acquiring the data. All data contained in the Atlas were acquired during the 1975 growing season (May and September) from ground sites in Imperial Valley, California. These sites and data acquisition flight parameters are discussed in Appendix B. Collection of supporting ground information is the subject of Appendix C. Additional surface condition information was derived from photo interpretation of 35 mm color aerial photography collected simultaneously with the spectroradiometer data. The method of interpretation and definitions of descriptive terms comprise Appendix D. Appendix E discusses availability of the spectral data on computer compatible tape and availability of supporting aerial photography.

The following description provides a brief key to the sample Atlas page shown in Figure 1. For more details about specific aspects of the data acquisition and analysis procedure the reader is referred to the appropriate appendices.

a. The title indicates the crop type and its growth stage as shown in Table 1.

b. The field description incorporates ground observations and soils data. Details of how this information was obtained are given in Appendix C.

c. The results of photo interpretation are presented in the form of a set of keywords as described in Table 2. A complete discussion of the photo interpretation keywords is included in Appendix D.

d. This plot is a representative spectrum of the

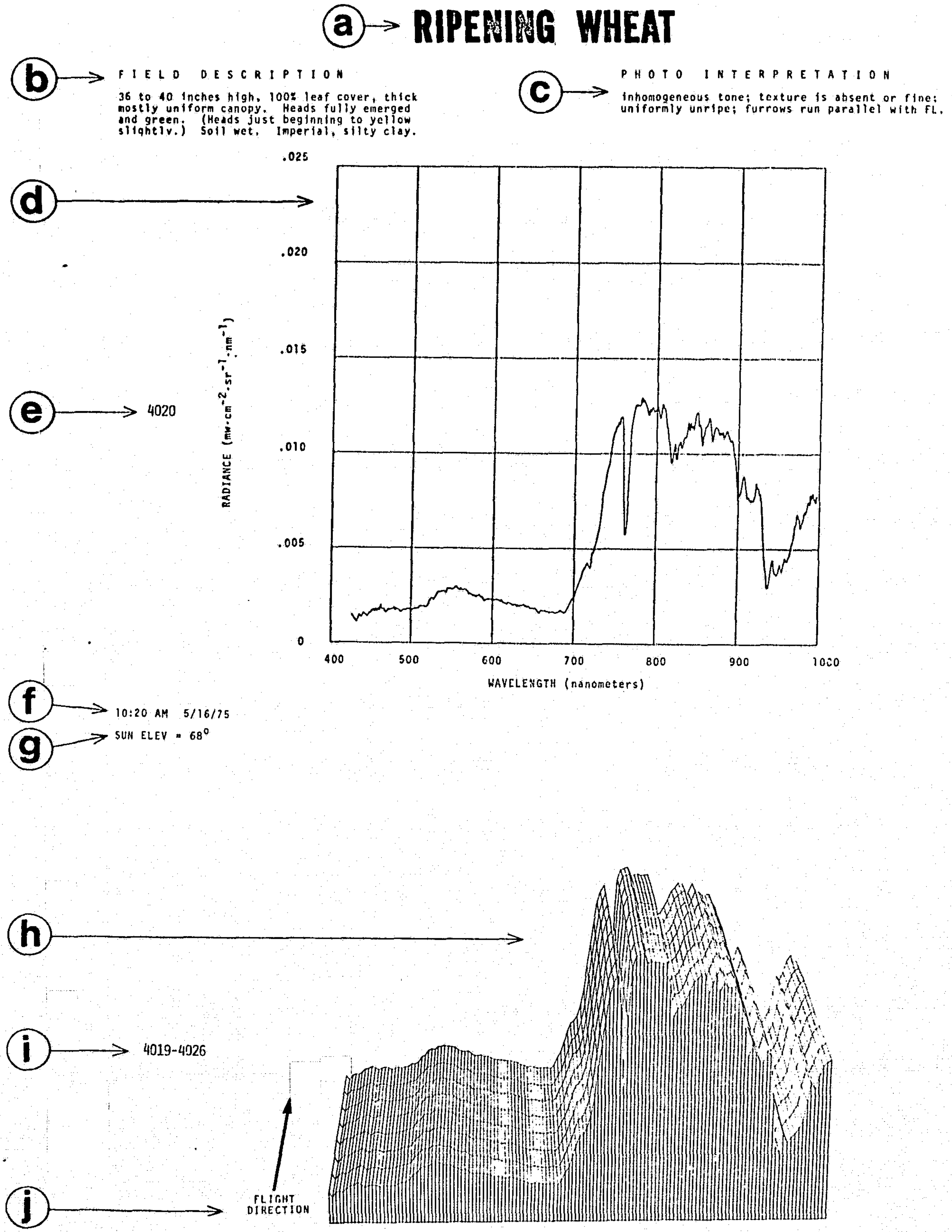


Figure 1. Sample Atlas page

TABLE 1. CROP GROWTH STAGE CHARACTERISTICS

Crop	Growth Stage	Characteristics
Alfalfa	Maturing	6 inches to 20 inches high
	Mature	20 inches high and above
	Harvested	less than 6 inches high
Asparagus		
Bare Soil		
Barley	Ripe	
Carrots	Harvested	
Cotton	Emergent	6 inches high or less
	Mature	blossoms and balls, balls closed
	Ripening	balls opening
Melons		
Sorghum	Emergent	8 inches high or less
	Headed	heads present
Sugar Beets	Emergent	6 inches to 12 inches high
	Thinned	12 inches high
	Mature	24 inches to 36 inches high
	Harvested	plant litter in field
Wheat	Booted	established, no heads
	Headed	heads present and green
	Ripening	fully headed, turning yellow
	Ripe	turned yellow

TABLE 2. PHOTO INTERPRETATION IMAGE CHARACTERISTICS

<u>Tone</u>	homogeneous inhomogeneous
<u>Texture</u>	none or absent fine medium coarse differential--range
<u>Density</u>	bare soil low medium high differential--range
<u>Cover</u>	bare soil little 1/3 2/3 near total total
<u>Crop Status</u>	growth stage, unusual patterns or conditions, and ripening description for wheat and barley fields
<u>Furrow Direction</u>	parallel or perpendicular to Flight Line

agricultural landscape being described. The arithmetic average is calculated for the spectra-set at all wavelengths. The representative spectrum is the spectrum within the set which has the minimum RMS (root mean square) difference from the average when summed over all wavelengths.

e. This number is the catalog spectrum number of the representative spectrum shown in the plot.

f. The spectra acquisition date and time (PDT) are taken from the flight logs and reflect the time at the start of a flight line. No single flight line was longer than five minutes in duration.

g. The sun elevation angle measured from the horizon plane was calculated for the time, date, and location of each set of observations.

h. The second plot on each data sheet is a three-dimensional representation of sequentially acquired spectra which make up the homogeneous unit of crop, crop condition or tillage state. The spectra are plotted one behind the other in the sequence in which they were acquired along the flight line. The horizontal and vertical axes are wavelength and radiance respectively as in the representative spectrum (d). To provide an interpretable display, the number of points plotted on the x-axis was reduced by averaging the radiance values for three wavelength bands or channels at a time.

i. The catalog spectrum numbers for the first and last spectra in each set are given for reference.

j. The flight direction indicates the order of acquisition for each set of spectra.

This Atlas is a product of the research activities of several groups which comprise the Earth Resources Program at the NASA Goddard Institute for Space Studies.

Instrumentation was developed and flown by Dr. William Collins, Columbia Univeristy, Department of Geology. Dr. Hong-Yee Chiu, Goddard Institute for Space Studies assisted in instrument design. Data processing methods were developed by Dr. Stephen G. Ungar, Goddard Institute for Space Studies. Photo interpretation and production support was provided by Columbia University, Department of Geography, under the direction of Dr. Jerry C. Coiner. Preliminary data processing was performed at Dartmouth College by William Beck and Nancy Wasserman.

Appendix A. INSTRUMENTATION AND CALIBRATION

The aircraft instrument system used to collect high resolution spectra employs a Princeton Applied Research "Optical Multichannel Analyzer" (OMA) (Princeton Applied Research Corp., 1975), which drives a silicon vidicon detector coupled to a Jarrel-Ash .33 meter spectrometer, and a computer-compatible tape recorder (Figure A-1). Collecting optics image a ground target on the entrance slit of the Ebert design spectrometer. Radiation through the slit is collimated, dispersed by a grating, and refocused on the vidicon raster. The vidicon samples the spectrum on the detector in 500 wavelength intervals at the rate of 64 microseconds per channel and 32 milliseconds per frame. The detector output is amplified at the sensor, then sent to the A/D converter and digital processing. Digital data are stored in memory 1 where up to 10^4 scans can be summed. Spectral data in memory 1 are read out in real time on the display scope, and summed data are dumped into an interface built to automatically control the system. The interface recycles the OMA and sensor while dumping the summed data into a computer formatted eight track magnetic tape recorder. The interface monitors data going on tape for errors, and it triggers a 35 mm ground truth camera on every tenth spectrum going to tape. Table A-1 correlates

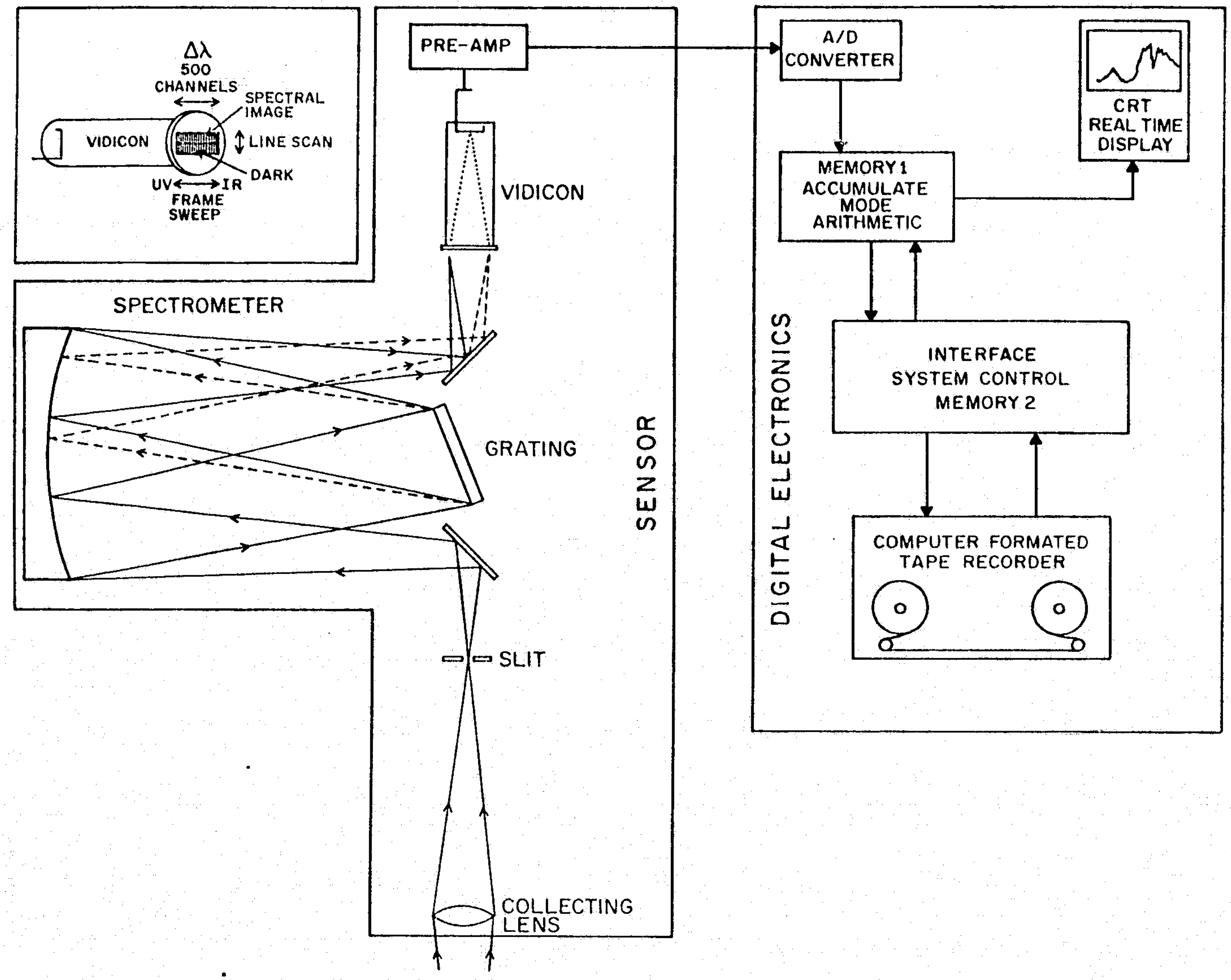


Figure A-1. Schematic diagram of the airborne instrument system. Radiation through the front optics is dispersed by the spectrometer and refocused on the vidicon. The vidicon raster format is shown in the inset diagram. Analog signals from the detectors are sent to the electronics package for processing and recording.

TABLE A-1. FLOW DIAGRAM OF THE AIRBORNE INSTRUMENT INTERNALLY PROGRAMMED OPERATION

Time Domain	Procedure	Data Material
64 microsec.	vertically scan one channel	one 13.96Å wide spectral channel
32 m sec.	sweep one vidicon frame of 500 channels	one spectrum to memory 1
320 m sec.	sum up 10 frames in memory 1	one averaged spectrum
3 x 32 m sec.	memory 1 data IBM formatted and recorded on tape	one averaged spectrum on tape
10 x 416 m sec.	camera triggered every 10th spectrum	one ground truth photograph

The operator gives the start signal at the beginning of a flight line and the sequence of events and timing are carried out with automatic recycling until the operator gives the stop signal.

the time domain procedure with the data format. The entire system functions continuously and automatically along an aircraft flight line.

The silicon target of the vidicon detector is a microscopic array of photodiodes spaced 8μ between centers and read out by an electron beam 20μ in diameter. A spectrum displayed on the vidicon raster is dispersed along the horizontal x-axis, or frame sweep dimension, of the detector (inset in Figure A-1). Each of 500 "monochromatic" channels in a frame is read out by a vertical line scan of 64 microseconds and output serially. The line scan is electronically chopped to remove dark current, which is read on the lower half of the raster where no signal is applied.

In the usual mode, photoelectric detectors, such as those used in LANDSAT, respond to radiant flux as a function $F(x,y,t)$ of the spatial coordinates on the detector surface and time

$$S(t) = \int \int_{\Sigma} F(x,y,t)\, dx\, dy \ [\text{counts}\cdot t^{-1}] \qquad (1)$$

where $S(t)$ is the digitized detector signal at the time t of the readout integrated over the detector surface Σ. The photodiodes in a vidicon detector are storage devices responding to total energy incident between readouts. The time relationships are lost, but the spatial relations can be preserved according to the scan mode used. The output signal of the OMA detector electronics can be written

$$S(x) = \int_0^T \int_{\Sigma_{y,\Delta x}} F(x,y,t)\ dt\ dy\ [\text{counts}\cdot\text{ch}^{-1}] \qquad (2)$$

where T is the readout interval. The line scan integrates the total energy incident between readouts, 32 msec., over the chromatic image of the spectrometer entrance slit focused on a 2.5 mm by 25μ detector element. Spectral information is preserved in 500 intervals in Δx along the frame sweep but spatial variables in each chromatic image are integrated over the line scan area $\Sigma y,\Delta x$. The vidicon system, functioning as a parallel input device, achieves considerable gain in sensitivity and simplicity of design over serial input devices using radiant flux detectors. Parallel optical input is the essential design element in the present application. It permits high spectral resolution measurements in 500 bands to be made in real time, and equally important, the simple design affords reliable operation from a light, inexpensive aircraft.

The instrument package consists of an electronics rack and the sensor, together weighing about 123 kilograms. It is flown in a light airplane at 610 meters above the terrain with the sensor looking vertically at the ground along the flight path. The instantaneous field of view is 18 meters by 1.8 meters (Figure A-2). The shape is determined by the rectangular dimensions of the spectrometer entrance slit. The long axis of the field of view is fixed perpendicular to the flight direction. Ten spectra, or frame sweeps, are summed over .32 seconds corresponding to 18 meters of aircraft motion along the ground track. The sum of ten frame sweeps is recorded on tape and later averaged to obtain the spectral response integrated over an 18 meter square field of view. In the cycle of add ten frames - dump to tape - start new summation, three frames are skipped. Blocks of data on tape in 500 spectral bands correspond to consecutive 18 meters square integrated target areas along the flight path with about 5 meters of spacing between. Ground truth photographs taken every 10th dump to tape have about 60 percent overlap.

Instrument Calibration

Aircraft spectral measurements in 500 channels per spectrum are calibrated to convert recorded digital data from the instrument in counts per channel to apparent radiance per one nanometer interval. Calibration measurements are made in the laboratory before the instrument is installed on the aircraft for a series of survey flights and again after it returns. The absolute calibration relates aircraft measurements to a standard lamp. One calibration is applied to all measurements in a single flight line in

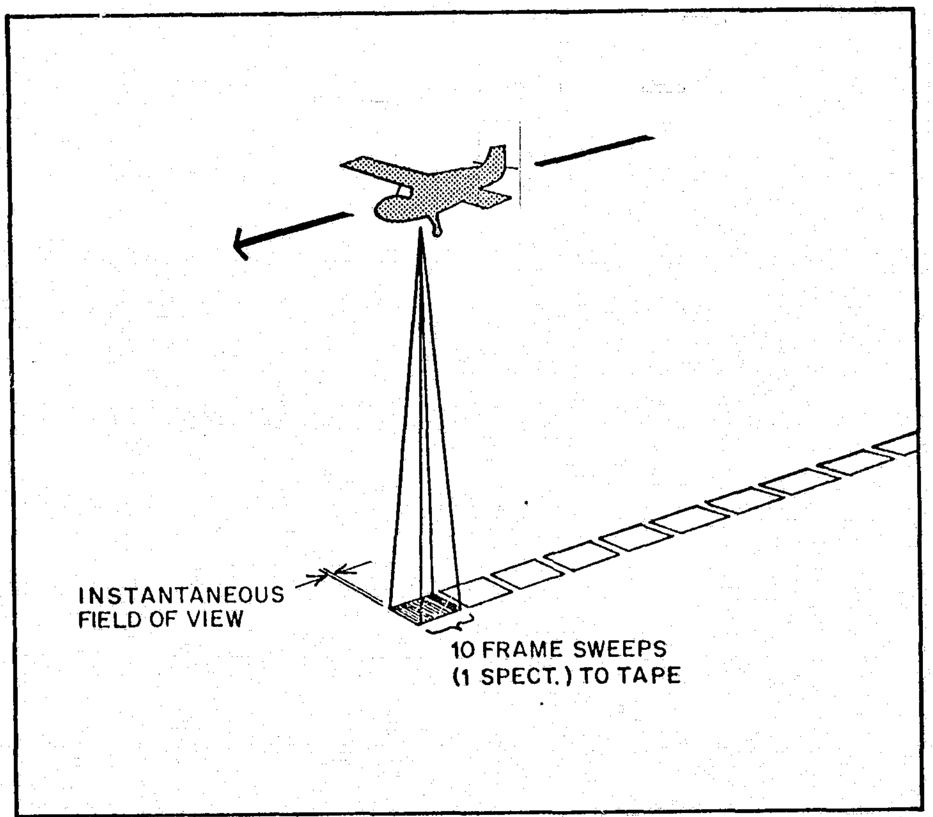


Figure A-2. Survey procedure. Spectral measurements are integrated over consecutive 18-meter-square areas along the ground track and recorded on tape. Flight lines are arranged to transect targets of interest.

order to preserve the separation of inherent inaccuracies of the absolute calibration from the precision of relative measurements along the flight line. Precision among spectral measurements in a flight line is affected by error in wavelength-channel correlation, caused by drift in grating position due to mechanical and temperature induced strain; drift in sensor electronics and detector sensitivity, mainly a function of temperature variation; and microphonics in the optical and vidicon systems. The drift due to temperature variations are small over a flight line of 2 to 3 minutes duration, and optical noise is controlled by careful design and construction. The precision over a flight line will be much better than the accuracy of absolute calibrations. The absolute calibrations will be affected by those factors affecting the precision only over much greater time periods and ranges in temperature. Absolute calibrations are also dependent on the accuracy and precision of the calibration source and laboratory conditions and procedures. The radiation standard used is an ISCO current regulated tungsten lamp (Instrumentation Specialties Company, Inc., n.d.) calibrated against a National Bureau of Standards lamp. Measurements made after the flight are compared with those taken at the same instrument settings before the flight. These are compared to determine drift in responsivity over the flight period. Calibration of the flight data is done in the after flight calibration simulating instrument settings and temperatures at the time of the field measurements for each set of instrumentation circumstances.

The signal output of the instrument can be broken down, disregarding field effects, as a function of the mean spectral radiance of the target and the instrument factors. Radiant energy through the system is affected by the optical properties of each component in the collecting optics and in the dispersing device before it reaches the detector element. At the detector, the efficiency of the photon to electric signal conversion is a function of wavelength, position on the detector surface, and angle from the normal of incident radiation. The analog electronic signal produced at the detector is processed through several stages of amplification in which amplifier gains modify the signal before it is digitized. All of the instrument factors combined determine the radiance responsivity R_N of the total system at the time of the measurement. The amplified and digitized output signal S_I of the instrument can be written

$$S_I = \int_{\Delta\lambda} \overline{N}_\lambda(\lambda) R_N(\lambda) d\lambda \ [\text{counts}] \tag{3}$$

for each channel of a spectral measurement.

The spectral response of each instrument component can be assumed constant over the narrow wavelength $\Delta\lambda$ of an instrument channel. The output signal by channel can be written in simplified form with all instrument parameters included:

$$S_I = \overline{N}_\lambda \; T \; \tau_0 \; \tau_F \; E_G \; R_D \; \Delta\lambda \; [\text{counts}]. \tag{4}$$

The throughput T, as already discussed, includes the geometric factors that determine the field of view Ω' and the effective area A of the entrance aperture. Included in T is the f-stop, or aperture stop, setting of the collecting optics used during the field observation. The transmittance of the optical train τ_0 includes all lenses, mirrors, and the detector window. Spectral transmission properties of all filters used for each observation are included in the term τ_F, and E_G is the spectral efficiency of the grating for non-polarized light. The detector response R_D is strongly dependent on wavelength, position on the light sensitive surface, and angle of incidence. To compensate for this as much as possible, the grating is set with a frequency standard to position the spectrum on the vidicon detector as it was in the field measurement.

An instrument transfer function can be defined for each specific optical configuration:

$$C_\lambda = [T \; \tau_0 \; \tau_F \; E_G \; R_D \; \Delta\lambda]^{-1} \; [\text{counts} \cdot w^{-1} \cdot cm^2 \cdot sr \cdot ch.]. \tag{5}$$

A set of instrument functions for each of the 500 channels is calculated for every combination of f-stop, filters, and grating position used in the field observations. The instrument function can be calculated from Equation (4):

$$C_\lambda = \overline{N}_{\lambda_L} \Big/ S_I \; [\mathrm{counts}\cdot w^{-1}\cdot cm^2\cdot sr\cdot ch.] \tag{6}$$

where $\overline{N}_{\lambda_L}$ is the radiance of a white diffusing plate irradiated by the ISCO lamp of known radiant intensity. Field measurements are calibrated from

$$N'_\lambda = S_T \; C_\lambda \; [w\cdot cm^{-2}\cdot sr^{-1}\cdot ch.^{-1}] \tag{7}$$

where N' is the apparent radiance of a target, S_T is the recorded signal for the target, and C_λ is the channel by channel transfer function for the specific instrument configuration. The spectral curves for the lamp measurement, the transfer functions, and the resultant calibrated lamp spectrum calculated from Equation (7) are shown in Figure A-3. The baseline (amplifier bias) is subtracted from the curves as it is in all the data.

Frequency calibration is accomplished using a krypton emission line spectral lamp (Figure A-4). The peaks of known wavelength can be located within 0.2 channels on the instrument display scope in real time. The channel width

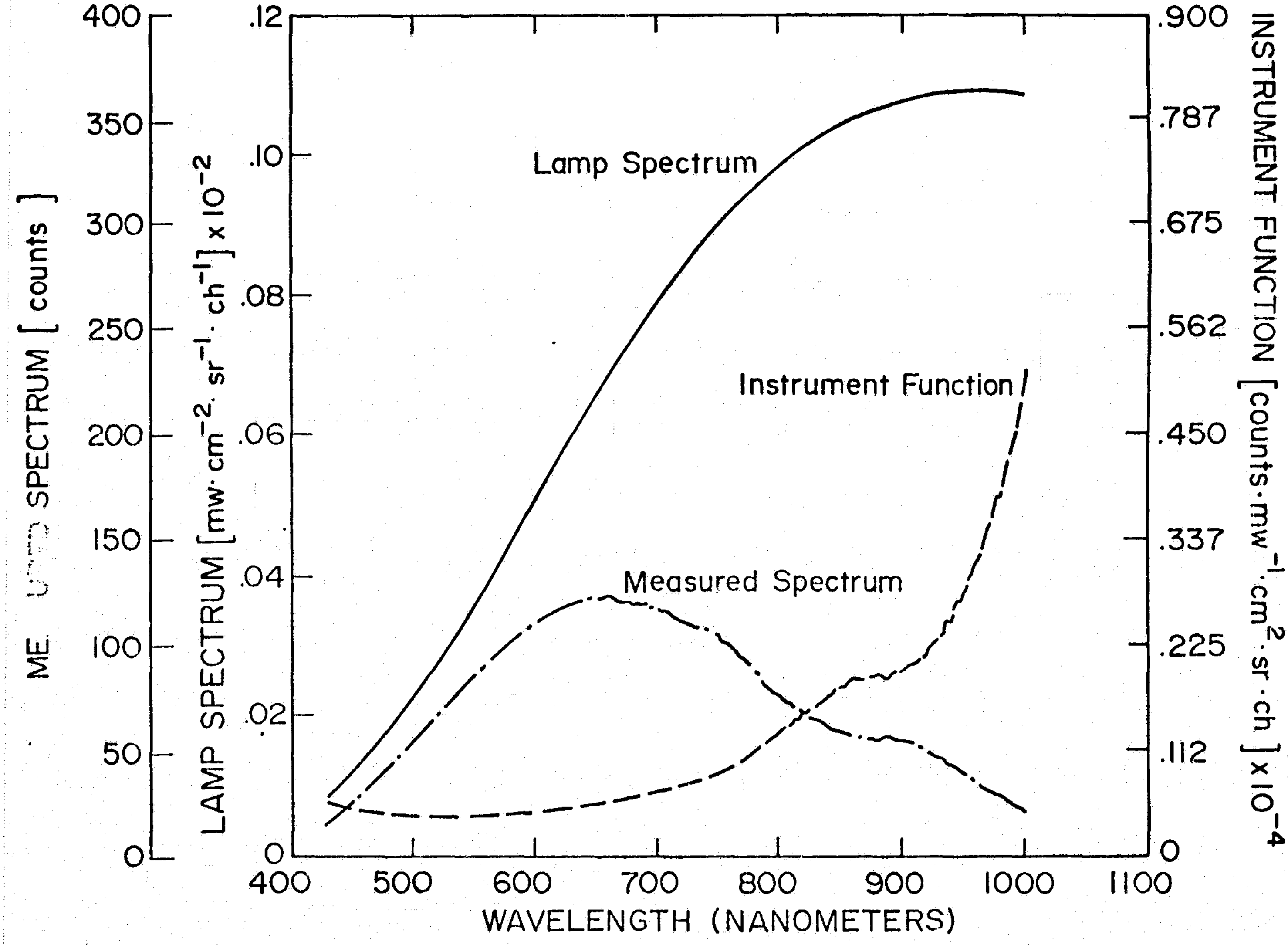


Figure A-3. True spectral curve of the ISCO calibration lamp (solid line) calculated as the product of the measured spectrum and the instrument function curve. The instrument function curve was determined from a similar measured spectrum of the calibration lamp.

ORIGINAL PAGE IS OF POOR QUALITY

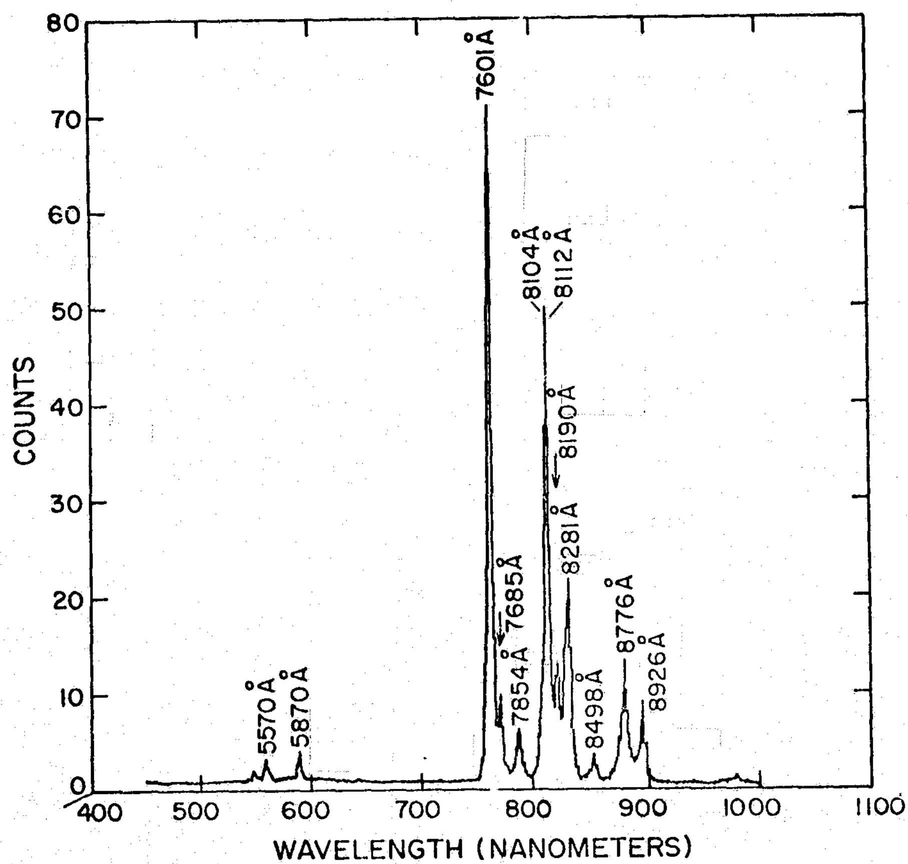


Figure A-4. Emission lines of a krypton gas lamp. Location and spacing of the lines of precisely-known frequencies are used in determining the instrument channel width and wavelength scale.

is calculated from the number of channels between the various peaks of known frequency. The channel width, 13.96Å $\pm$ .025Å is determined by the grating dispersion and the spectrometer focal length. A 72 grooves per millimeter grating disperses the wavelength interval from 4000Å to 11000Å over 12.5 mm in the system spectrometer that has been modified to approximately .1 meter focal length. A split filter is placed in front of the detector to eliminate second order effects between 7000Å and 11000Å. The low dispersion grating gives a very even wavelength distribution across the detector. Error translated to the end channels due to uneven dispersion is 6.25Å, less than half a channel. Frequency alignment is made by positioning the 7601Å krypton peak on channel 250. Field calibration for grating setting is obtained from the position of the atmospheric oxygen absorption band centered at 7605Å. This absorption band is inherent in every field spectrum, allowing accurate monitoring of wavelength setting in the field data.

The field of view was mapped using a light source through a small aperture at 61 meters from the sensor. It is 15 milliradians in the long slit dimension and 1.5 milliradiansin the slit width direction. The system linearity with change in intensity measures $\pm$1 percent using calibrated neutral density filters and the ISCO lamp. The detector and amplifier RMS noise, looking at a series of recorded reference baselines (front lens covered), is ± 1 count in the accumulate ten scans mode. Most target measurement signals are between 100 and 700 counts per scan. The signal to noise ratio is greater than 100:1 in most cases. The maximum dynamic range is 750. The lag in response to signal change on the accumulate 10 scans mode was measured by repeatedly opening and closing the aperture while looking at a strong light source. The signal rises to within 12 percent of the maximum in the first recorded spectrum on opening the lens, and the signal drops to 17 percent of the last signal on closing off the front lens. The slow response time must be considered in planning flight lines and interpreting the data.

There is a maximum spectral standard deviation of 5.5 percent in calibration measurements made before and after the two aircraft flights to Nevada. The apparent drift in sensitivity can be traced mainly to temperature effects and frequency misalignment among measurements. The maximum standard deviation in sensitivity between the normal operating ranges of $15^{\circ}C$ to $27^{\circ}C$ is ± 3.5 percent in the blue spectral region (Figure A-5). There is a possible ± 1 channel absolute error in reproducing the grating settings as they were in the field. The oxygen absorption band is checked in every flight line, and the relative wavelength drift is always less than 1/2 channel. A drift of one channel can cause up to 2.5 percent standard deviation in the spectral

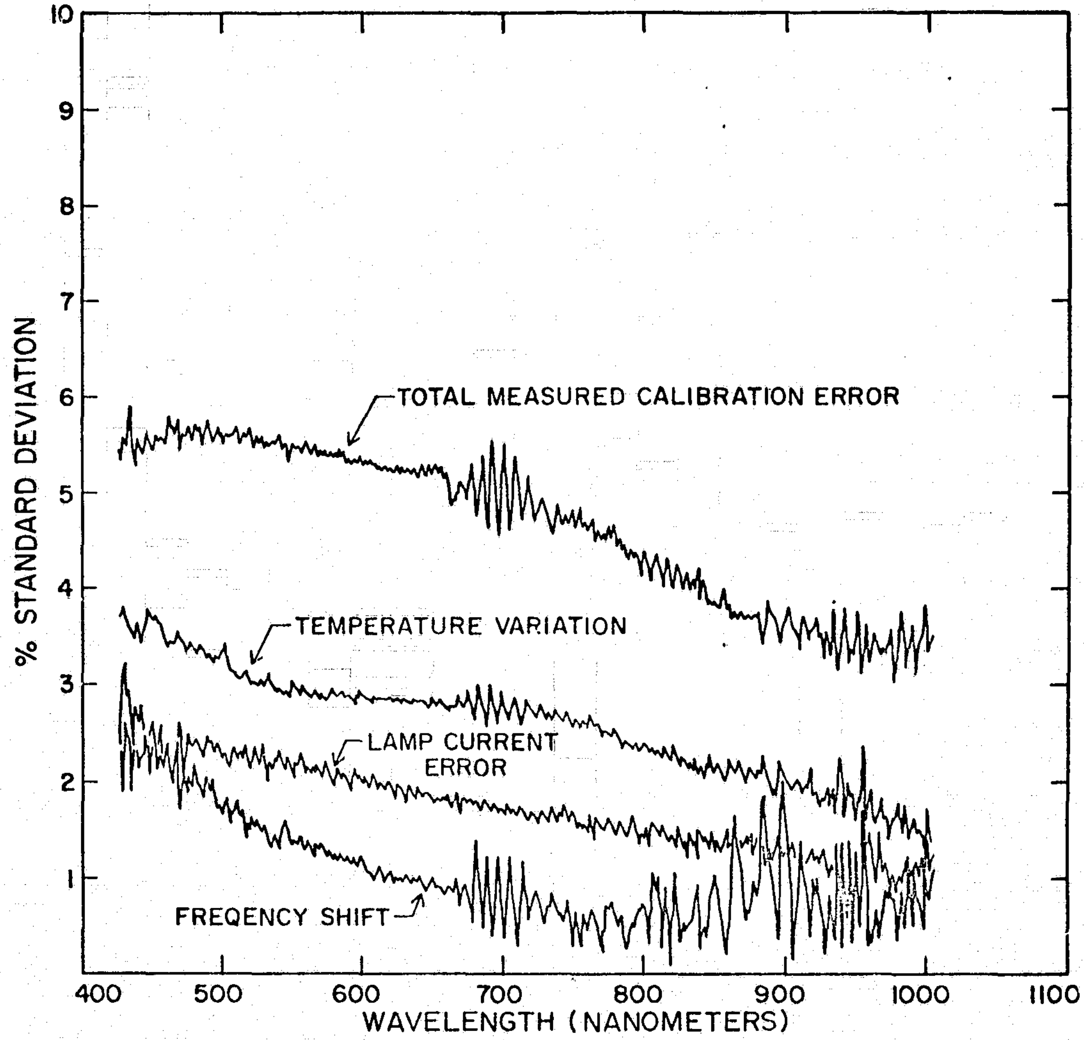


Figure A-5. Percent standard deviation as a function of wavelength among all before- and after-flight calibration lamp measurements. The three lower curves indicate measurement variation as temperature, frequency setting, and lamp current were varied under controlled laboratory conditions. The noise at 7000Å is the edge effect of the split filter for eliminating second-order overlap.

measurements. There is a possible error of the same amount due to lamp current adjustment error. The other sources of error, calibration lamp aging and electronics drift, have not been monitored.

The precision among all data collected will be on the order of the error discussed above. Drift in sensitivity over lamp measurements of 2 to 5 minutes, however, is less than ±1 percent. This should also be the precision of the data within a flight line. Most of the error in the relative calibrations from one flight line to another over long time periods would be removed by on-board calibration. The accuracy, or absolute calibration has the added error of the calibrated light source. The manufacturer's estimate on that is ±5 to 10 percent (Instrumentation Specialties Company, Inc.) relative to the NBS standard.

References

Collins, William. "Spectroradiometric Detection and Mapping of Areas Enriched in Ferric Iron Minerals Using Airborne and Orbiting Instruments." Ph.D. dissertation, Columbia University, 1976.

Instrumentation Specialties Company Inc. "Instrumentation Manual - Model 5RC Spectroradiometer Calibrator." Lincoln, Nebraska: Instrument Specialties Company Inc., n.d.

Princeton Applied Research Corp. "Optical Multichannel Analyzer (OMA), Operating and Service Manual." Princeton, NJ: Princeton Applied Research Corp., 1975.

Appendix B. SITE AND MISSION

Site

The data presented in this document were collected in Imperial County, California (33^{O}N 115^{O}W). This county, which encompasses nearly three million acres (1.5 million hectares), is bordered by Riverside County (north), the Colorado River (east), Mexico (south), and San Diego County (west).

The county can be described as comprised of two parts: the western two-thirds, dominated by a broad structural trough having a NNW-SSE axis; and the eastern one-third, where the broad, gently-sloping alluvial fans from the Chocolate Mountains meet the floodplain of the Colorado.

The trough in the western part of the county contains Colorado River sediments, and alluvium from the Chocolate Mountains and the California Coastal Range. The Salton Sea is in the center of the trough. Old Lake Cahuilla, now dry, is also in the center of the trough, and contains irrigated farmland which comprises the site locations.

Elevations in the Imperial County range from 71 meters below sea level to more than 1200 meters above sea level. Irrigation occurs only at elevations below 75 meters.

The climate of Imperial County is dry with hot summers and cool winters. Average rainfall is approximately 3 inches per year, of which about half falls in high intensity summer showers and about half in gentle winter rains.

Day temperatures rise above 100^{o}F almost every day, May through October, dropping to the low 60's at night. Mean temperatures over a 50-year period at Imperial are 52.9 for January, 91.2 for July, 71.5 for the year.

Average annual humidity is about 30 percent and is often higher in July and August.

Highest evapotranspiration takes place during the period June through September. Consumptive use in this period is about one-third of an inch per day (USDA, 1967).

This study specifically focused on two sites of irrigated agricultural land in the Cahuilla dry lake, Imperial Valley, California. One site is located in the southern part of the county near the town of Calexico, while the other is located in the eastern-central portion of the county (Figure B-1). Together, they encompass approximately 3,840 acres or 1500 hectares of crop land. The crop types include both continuous-cover crops, such as alfalfa, wheat and barley, and row crops, such as sugar beets, melons, asparagus, carrots, sorghum and cotton. Crop distribution for the two sites is presented for each date of data collection in Figures B-2 and B-3.

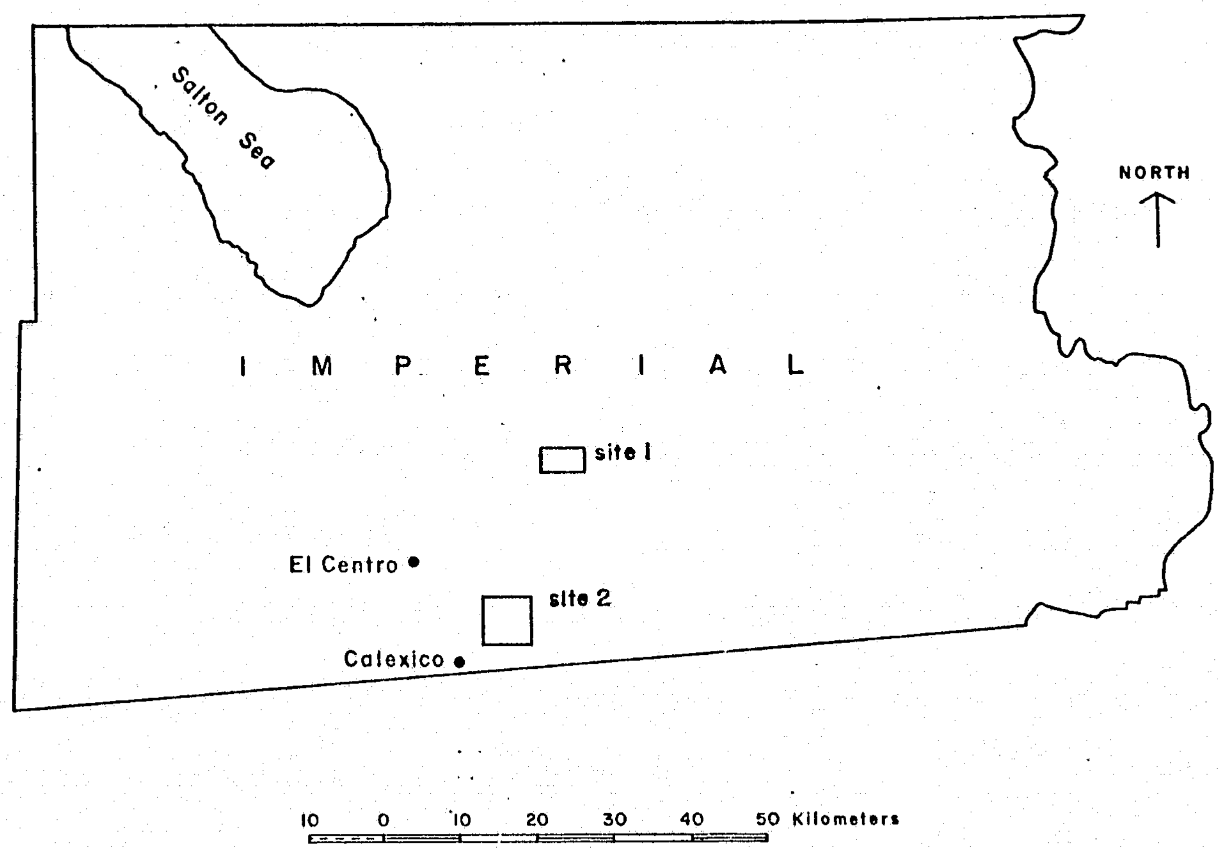


Figure B-1. Location of test sites, Imperial County, California.

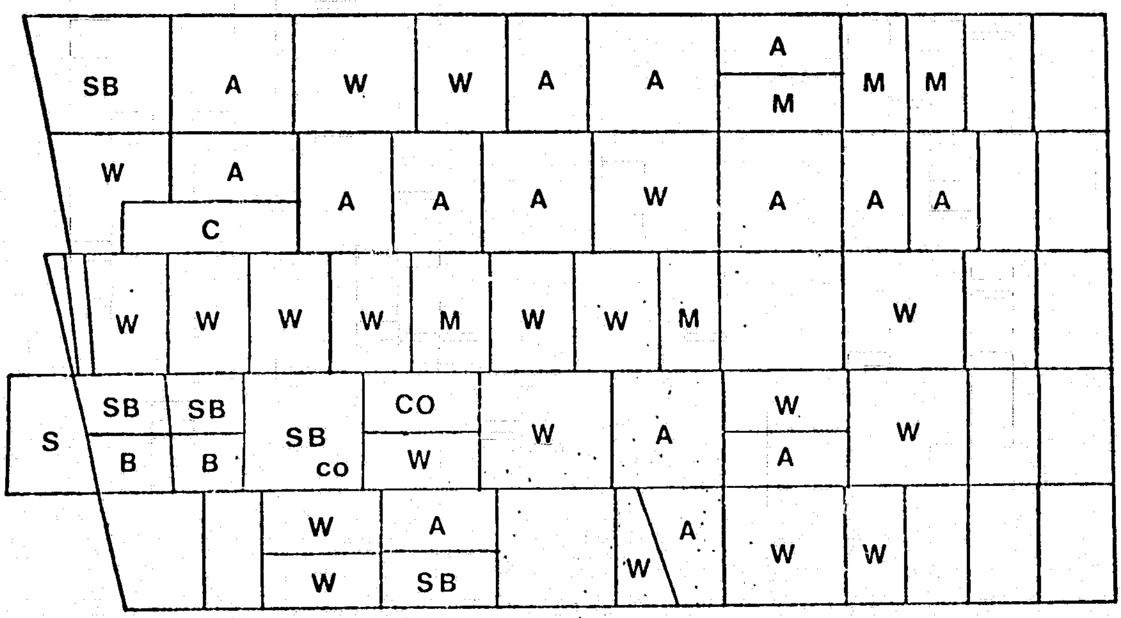


Figure B-2. Fields in site 1 for which spectra are included in the Atlas, labeled by crop type: A=alfalfa, B=bare soil, C=carrots, CO=cotton, M=melons, S=sorghum, SB=sugar beets, W=wheat. Lower case letters are used to denote September crop when it differs from May crop.

ORIGINAL PAGE IS OF POOR QUALITY

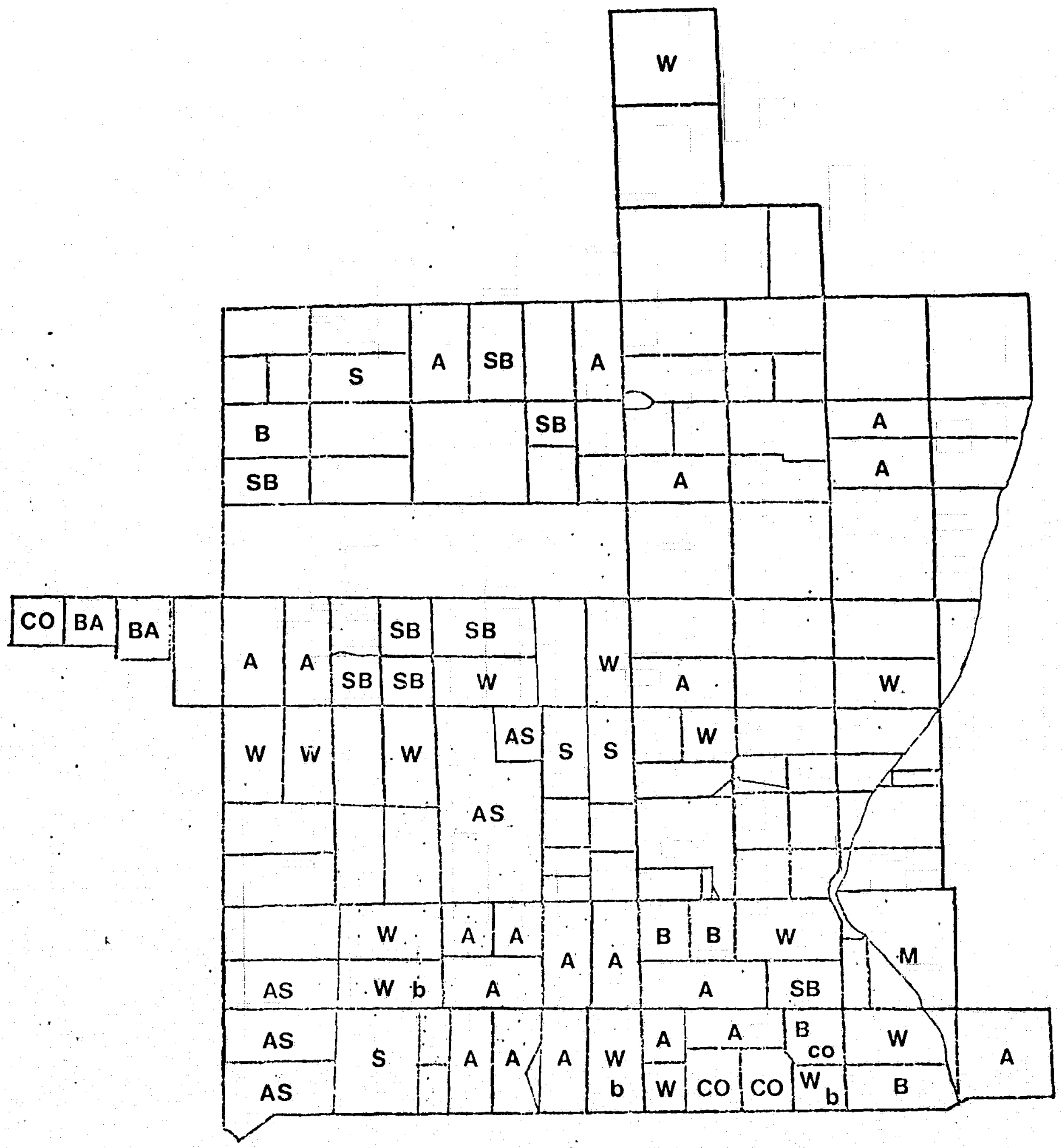


Figure B-3. Fields in site 2 for which spectra are included in the Atlas, labeled by crop type: A=alfalfa, AS=asparagus, B=bare soil, BA=barley, CO=cotton, M=melons, S=sorghum, SB=sugar beets, W=wheat. Lower case letters are used to denote September crop when it differs from May crop.

ORIGINAL PAGE IS
OF POOR QUALITY

Mission

Spectroradiometer data and 35 mm color photography were simultaneously collected from a twin-engine Aero Commander at an altitude of 610 meters above local terrain. These flights were flown east-west, as shown in Figures B-4 and B-5. Site 1 was covered by ten flight lines approximately 5 miles (8km) in length. Site 2 was covered by eleven east-west flight lines, ten approximately 4 miles (6.5 km) long and one approximately 1 mile (1.6 km). Not all flight lines were flown on all missions, and some were flown twice on the same mission (Table B-1).

Figure B-4. Flight lines for site 1. Each flight line is approximately 5 miles (8 km) in length. Not all flight lines were flown on all missions, and some were flown twice on the same mission.

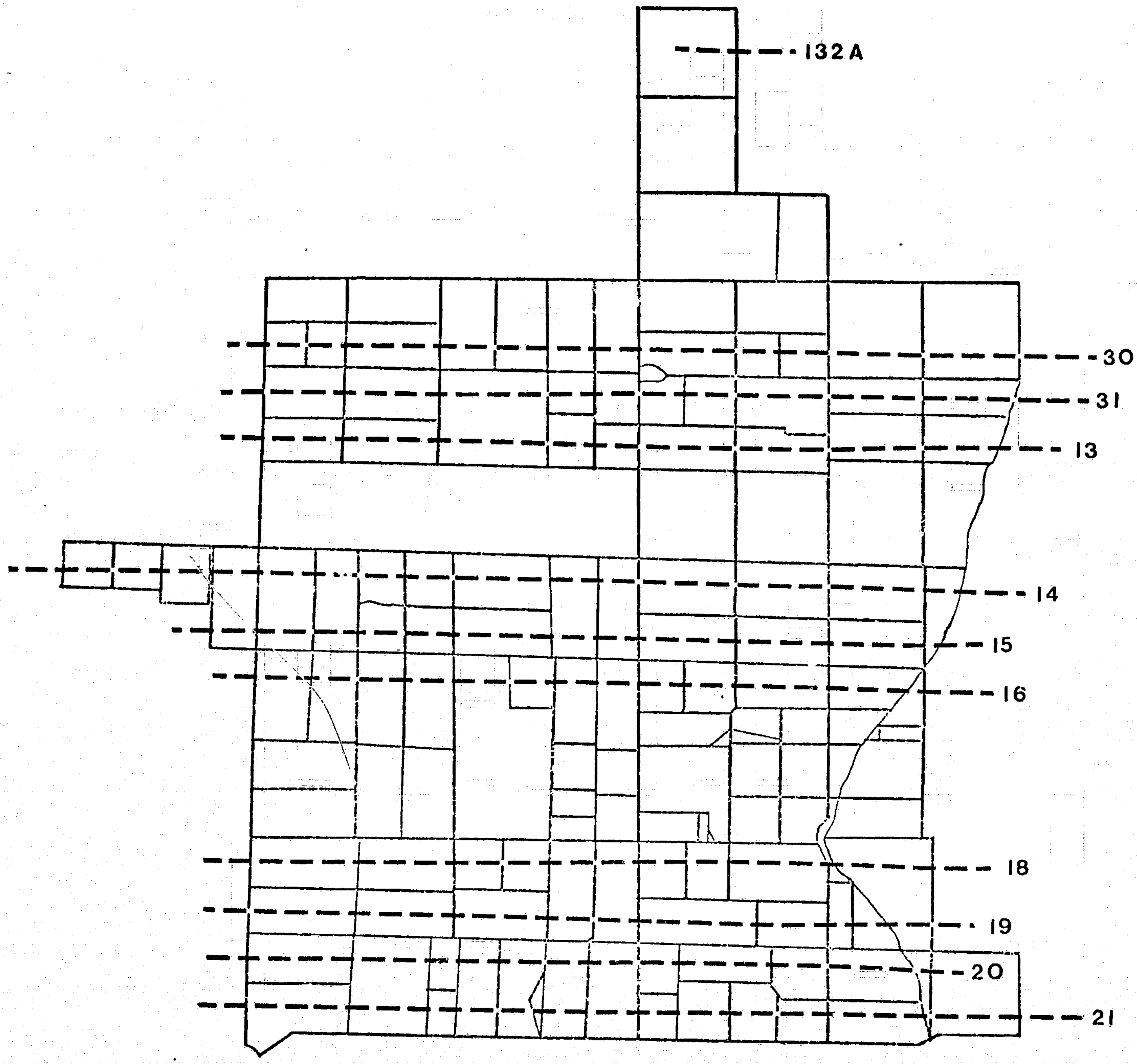


Figure B-5. Flight lines for site 2. Ten of the flight lines are approximately 4 miles (6.5 km) in length; the other one is approximately 1 mile (1.6 km) in length. Not all flight lines were flown on all missions, and some were flown twice on the same mission.

TABLE B-1. MISSIONS OVER IMPERIAL VALLEY TEST SITES, 1975

Site	Date	Time (Duration)	Flight Lines[a]
2	5/15/75	9:28 AM - 10:42 AM	13, 14, 15, 16, 18, 19, 21[b], 30, 31, 132A
1	5/16/75	9:54 AM - 11:03 AM	1, 2, 3[b], 4, 5, 6, 8, 10
1	5/20/75	10:55 AM - 11:10 AM	7, 9
2	5/20/75	11:15 AM - 11:25 AM	14
1	9/23/75	11:19 AM - 11:30 AM	3,4
2	9/23/75	10:26 AM - 10:37 AM	19, 20, 21
2	9/23/75	11:42 AM - 11:50 AM	21

[a]See Figures B-4 and B-5.

[b]Collected twice.

Appendix C. GROUND OBSERVATIONS

Ground truth was collected by GISS, Columbia and Dartmouth personnel on the same day as the overflights. Ground data were taken using standard windshield survey techniques accompanied by ground assessment in each field to determine irrigation state, crop, crop height and density, and overall field condition. These data were augmented by discussion with and participation of local USDA personnel as well as interested farmers who volunteered information and assistance. Also, Imperial Irrigation District provided extensive supporting data on crops, crop calendars and irrigation status.

In addition to these direct observations, baseline soil data have been extracted from the General Soil Map and Report, Imperial County, California (USDA, 1967). These data elements include soil type, soil texture, and soil color for the Ap horizon. Soil color is given in the Munsell notation when the soil is reported dry.

Reference

U.S. Department of Agriculture Soil Conservation Service. Report for General Soil Map, Imperial County, California. El Centro, California, 1967.

Appendix D. PHOTO INTERPRETATION SUPPORT

Photo interpretation of the 35 mm color aerial photography collected concurrently with the spectroradiometer data was performed for the purpose of supporting and clarifying the spectroradiometer data, and also to supplement the ground observations by uncovering broad and extensive patterns which were unrecognizable in ground field work.

An intensive interpretation was performed on individual fields using 35 mm (strip) color transparencies at a scale of 1:20,000. These strips were acquired simultaneously with the spectroradiometric data. The coverage of the fields encompassed 840 frames with approximately 60% overlap. Bausch and Lomb Zoom 240 optical equipment was utilized with varying magnifications, ranging from 4 to 7.5X, according to the image characteristic under examination. Each field was classified using the image characteristics, tone, texture, density, cover, crop status and furrow direction (see Table 2, Chapter 1).

C-2

When examining tone, fields were classified according to their homogeneity. If the tone was interrupted because of contrasts due to ripening, visibility of bare soil areas, soil moisture patterns, or any other crop condition, the field was graded inhomogeneous.

Texture results from differential height in the crop. For example, a coarse texture is present when growth is sporadic, harvesting is in progress, or when a crop, such as melons, is in an early growth phase and height difference between the clusters of plants and bare soil is large. Texture is absent or fine when the crop has smooth and uniform cover, or when the field contains bare soil. In those fields where texture was differential, the range was specified and exceptions were mentioned.

Both density and cover are descriptions of the amount of soil that is visible in the field, however, it is quite possible for a crop to have total cover but medium density, for example. To clarify the difference between terms, density is basically the number of plants per unit area and describes the thickness of the crop and how closely planted it is. Crop cover is more a description of the growth stage and vigor of the crop. If a crop is stressed, the cover can be spotty and discontinuous while a healthy (unstressed) crop will tend to have total or near total cover. Likewise, newly planted crops have very low cover while mature crops tend to have total cover.

The classification of furrow direction was very straightforward. The furrows either ran parallel or perpendicular to the flight line. However, the term "furrow" should be further defined since it is used loosely to refer not only to furrows but also to crop rows and irrigation control ridges in continuous-cover crops. In some fields, the crop density was so high that the furrows were not visible. In these cases, no mention was made of their direction.

Any unusual pattern or conditions fell under the crop status category. These included such things as ripening stages, and irrigation, growth, or soil moisture patterns. The location or size of these patterns was identified in terms of ground distance and proximity to field boundaries. Also, the possible reasons for unusual conditions were often mentioned.

APPENDIX E. ADDITIONAL DATA PRODUCTS

The spectra used in compiling this Atlas are available on computer-compatible digital tape (CCT) for more detailed analysis. The digital tape includes coded information obtained from the ground observations and photo interpretation, in addition to the spectral radiances.

The CCT contains a total of 3458 records. Each record represents one spectral observation and is 844 bytes long (1 byte = 8 bits), consisting of a 24 byte ancillary data field followed by an 820 byte field containing the spectral radiances. The format of each record is shown in Table E-1.

Simultaneously acquired overlapping color aerial photographic coverage of areas where spectra were obtained is available at a scale of 1:20,000 on 35 mm positive transparencies at reproduction cost.

For further information on availability of data, contact Dr. Stephen G. Ungar, NASA, Institute for Space Studies, 2880 Broadway, New York, New York 10025.

TABLE E-1. BINARY RECORD FORMAT

Byte	Information
1-24	Ancillary data*
1-2	No. of first spectrum in set (16 bit integer)
3-4	No. of last spectrum in set (16 bit integer)
5-6	No. of spectrum representing set (16 bit integer)
7	Month
8	Day
9	Hour
10	Minutes
11	Sun elevation (degrees)
12	Crop code: 00=alfalfa; 01=asparagus; 02=bare soil; 03=barley; 04=carrots; 05=cotton; 06=melons; 07=sorghum;08=sugar beets; 09=wheat
13	Growth stage code: 00=emergent; 01=booted; 02=headed; 03=thinned; 04=maturing; 05=mature; 06=ripening; 07=ripe; 08=harvested
14	Height, lower limit (cm)
15	Height, upper limit (cm)
16	Percent crop cover
17	Soil moisture code: 00=wet; 01=dry
18	Tone code: 00=homogeneous; 01=inhomogeneous
19	Texture code: 00=none or absent; 01=fine; 02=medium; 03=coarse; 04=differential
20	Density code: 00=bare soil; 01=low; 02=medium; 03=high; 04=differential
21	Cover code: 00=bare soil; 01=little; 02=one-third; 03=two-thirds; 04=near total; 05=total; 06=nonuniform
22	Furrow direction code: 00=parallel to flightline; 01=perpendicular to flightline
23-24	Fill bytes
25-844	Spectral radiances (2 bytes or 16 bits per channel)**

*This segment of each record is descriptive of the spectra-set to which the recorded spectrum belongs. When information is not available, the byte is set to integer value 255.

**Reflected energy is recorded for 410 channels in uniformly spaced integer levels from 0 to 10000 representing 0.0 to 0.025 $mw \cdot cm^{-2} \cdot sr^{-1} \cdot nm^{-1}$. The first channel is centered at 426.86 nanometers. Channel widths are 1.396 nm.

A FIRST INTERPRETATION OF EAST AFRICAN SWIDDENING VIA COMPUTER-ASSISTED ANALYSIS OF 3 LANDSAT TAPES

FRANCIS P. CONANT
Hunter College, CUNY

TINA K. CARY
Columbia University

I. INTRODUCTION

This paper reports a successful preliminary application of machine processing of LANDSAT data to the identification of swidden farming in East Africa (Figure 1). Analysis of "slash and burn" or shifting cultivation via LANDSAT requires recognition of the characteristics intrinsic to swiddening: by Western standards, field sizes are small, their borders are irregularly shaped and merge with natural features of the terrain; in the area of East Africa with which we are concerned, multiple cropping of as many as 25 cultigens is common, and the small fields in which these crop complexes are grown are interspersed among land at various stages of fallow and regeneration of plant cover.

These characteristics of swidden farming combine to achieve what Geertz[1] has called a "canny imitation" of the natural landscape. This mimetic effect makes swidden fields indistinguishable from surrounding plant cover by visual inspection of standard LANDSAT imagery (Figure 2). In contrast, areas of Western-style agriculture are readily apparent. However, our analysis of the digital LANDSAT data does allow swidden areas to be differentiated.

The special problems which swidden farming pose require "a reorientation of techniques and typologies."[2] Such a reorientation is justified for two reasons. First is the number of people for whom swidden farming is the life-support system: in 1963, Conklin[3] put this number at 200 million persons in Africa, Asia, and the New World. Second, the homogeneous quality of the LANDSAT data, and the capability of making repeated observations, facilitate analysis of spatio-temporal variables intrinsic to swiddening.[4]

II. DESCRIPTION OF THE AREA

Our East African study area is West Pokot District, Rift Valley Province, Kenya. The study area lies midway between Mt. Elgon to the southwest and Lake Turkana (formerly Lake Rudolph) to the northeast. The District is about 5,000 square kilometers and includes a population of 60,000 persons, including about 45,000 swidden farmers. The balance employ a combination of farming and herding as subsistence techniques.[5]

The District's physiographic characteristics include a range of altitude from 800 meters to 4,000 meters above sea level. The probability of rainfall meeting evapo-transpiration requirements varies from .1 at lower altitudes in the north to .9 at higher altitudes in the south.[6] The soils, formed from a variety of substrates, tend to be easily exhausted. Along a cline from high to low altitude, plant cover includes "Alpine" meadow, hardwood forest, *Commiphora* scrub, riverine thickets, scattered *Acacia*, and a grassy cover which ranges from tall and dense to short and patchy (Figure 3). The preponderance of swidden farming is at mid-elevations, in association with *Commiphora* regrowth.

The swidden area we used for training is in a region known locally as *Asar* (see Figures 4 and 5), which is part of a narrow valley oriented north and south. The Asar region in general and the swidden area in particular are part of a subsistence ecozone the Pokot refer to generically as *kamass*, that is, mid-altitude slopes used extensively for swiddening. We use the term ecozone to mean an area sharing similar characteristics and exploited by a particular subsistence activity.[7]

The upper and lower limits of the kamass ecozone in the Asar region are approximately 2,500 meters and 1,500 meters above sea level. In the Asar region, a further characteristic of this ecozone is that the slopes are generally east-facing (B, C, and D in Figure 6). Cultivation takes place on slopes of varying degrees of steepness in plots of varying size, often less than half a hectare. A single plot may contain sorghum, millet, beans, gourds and squashes. Harvesting of different crops occurs at different times in a single plot, and at different times in different parts of the same ecozone. Throughout the ecozone, areas of mixed cultivation are interspersed among plots previously returned to fallow. The fallow period may be as long as 20 years, by which time a mature community of secondary regrowth is established.

The decision to return a fallow area to active cultivation involves a complex set of socio-ecological factors which include population pressure, food reserves, and a need to reestablish a family's claim to the land, as well as the amount of time the land has been in fallow and the degree of regeneration of plant cover. A plot re-opened for cultivation is therefore likely to be surrounded by areas at various stages of the fallow period, from newly returned to fallow, to completely regenerated and mature regrowth. This situation is schematized in Figure 7, in which an area of active cultivation (marked AC) is surrounded by areas at different stages of fallow (marked F-1, F-5, etc., with the numbers referring to years in fallow).

The cycling of plots between active cultivation and fallow in the ecozone we studied tends to involve shorter periods of time at lower elevations than at higher elevations. That is, close to the bottom of the valley where water for cultivation is more readily available the fallow cycle is shorter, and plots are returned to active cultivation from a herbaceous stage rather than a ligneous stage of plant growth; at higher altitudes where water for cultivation is less readily available, and depends in part on the uncertain use of an irrigation technique developed by the Pokot, the fallow period is long enough that trees become established and reach maturity. Despite the many factors involved, a generalization is possible: areas of active cultivation on the upper slopes of the ecozone in the Asar region tend to be surrounded by areas of more mature regrowth, including trees, than areas of active cultivation on the lower slopes of the same ecozone. Regrowth in fallow areas on the lower slopes tends to be limited to grasses and bushes. This differentiation is evident in the 1956 aerial photography, in which there is a clear demarcation of upper and lower slopes by a band of relatively dense regrowth (see Figure 5).

III. DATA DESCRIPTION

Our knowledge of West Pokot District derives largely from data collected during anthropological fieldwork in 1961-62. In addition to these field data, there is complete aerial coverage of the District, collected in 1956. Since even recent maps of the District are relatively incomplete, the two main sources of information are the 1956 aerial photographs and the 1961-62 fieldwork. The time lapse between these data and the LANDSAT data limits the kinds of interpretation the present study can make.

The LANDSAT data consist of three tapes, each representing a different portion of the cycle of dry and wet seasons characteristic of the District, as follows: 1972, mid-dry season; 1973, late dry, and 1975, early wet. Quality ratings for these tapes were, respectively, 8888, 8888, and 5588.

IV. DATA ANALYSIS

The first step in data analysis was a preprocessing step to de-skew, rotate, and rescale the data. This geometric correction process resulted in line printer output at a nominal scale of 1:24,000[8] All three data sets were geometrically corrected, and gray scale printouts were produced for a region approximately equal to 1/25th of a scene. The gray scales had 16 levels or ranges of spectral response values, with one alphanumeric symbol or overprint combination assigned to each level.

To simplify correlating the aerial photography with the LANDSAT data, photographic enlargement was used. The original photography was at a nominal scale of 1:44,000; as enlarged to match the gray scales, the nominal scale was 1:24,000. For final scale adjustment, a Bausch and Lomb Zoom Transfer Scope was used to register the photographic image of the swidden area to its location on the gray scale.

By using the aerial photography to provide a framework for spatial referencing, a 32 x 40 pixel training site centered on the swidden area was selected.

The next step was to obtain spectral response values for training the classification algorithm, developed by Dr. Stephen Ungar at NASA Goddard Institute for Space Studies. The spectral responses in all four bands were tabulated for pixels falling within the previously delimited swidden area. From these tables, two methods of obtaining training data were tried: (1) the average, or mean, and (2) the most frequent, or modal response. Both methods gave similar results. The ease of determining the modal responses led us to use that method for the analysis. For all three dates, the modal response in each band for the swidden area was determined (Table 1). These modal responses were used to train the classifier. For each date, one modal spectral response was used to define the informational class swidden. The classifier then made a binary decision for each pixel: the pixel was or was not within the spectral class defined by the training spectrum and the program parameters. Program parameters were selected empirically to maximize correct classification of data points within the swidden area.

V. RESULTS

Classification results are shown superimposed on the aerial photography (Figure 8). Table 2 displays the distribution of classified points according to the areas demarcated in Figure 5.

The distribution of points is highly correlated with the ecozone comprised of lower slopes marked C, the training area marked D (also lower slope), and upper slopes marked B. The percentage of points falling within this ecozone is, for three classifications, 77%, 82%, and 80%. This leads us to infer that the spectral class defined by our training method does represent the informational class of swiddening.

VI. SUMMARY, DISCUSSION, AND CONCLUSIONS

This paper has reported a first attempt at interpreting swidden farming in East Africa via machine analysis of three LANDSAT scenes. Each scene represents a different portion of the seasonal cycle of alternating wet and dry periods, a characteristic of the area notably affecting food production.[9] The training area was selected because it represents a traditional system of East African swiddening, utilizing mid-altitude slopes of mountains and escarpments.[10] This training area is part of a larger ecozone defined by upper and lower altitudes and the east-facing orientation of its slopes.

Inspection of the 1956 aerial photography (Figure 5) and observations made during the 1961-62 fieldwork show only light use of the upper slopes in the kamass ecozone of the Asar region. Our classification results indicate an increase in swiddening activity over the period 1956-1975: Furthermore, within the 3 year span of our LANDSAT data, results indicate a considerable increase in swiddening activity (Table 2).

Given the necessity in the swiddening system of retiring areas of active cultivation to fallow, and opening up new areas of cultivation elsewhere, and given the shared characteristics of the ecozone, which make relocation elsewhere within the same ecozone likely, we feel this analysis resulted in (1) a successful identification of an area of swiddening in East Africa and (2) that computer analysis makes this possible. In the standard image products, the mimetic effect associated with swiddening tends to blur distinctions between the natural surround, active cultivation, and land in fallow.

We have not presented the results of our classifications in terms of an increase of so many hectares because to do so might give a false impression of accuracy. In view of our available ground information, we claim only the identification of swiddening activity within an ecozone. As for an increase in this activity over earlier levels, we recognize different possible explanations. For example, one season may be better than another for the detection of swiddening, and this alone might account for all or most of the difference in number of points classified in the three data sets. However, there is also the possibility that the changing number of points classified reflects a corresponding change in swidden activities. Only further ground truth can resolve this. The acquisition of additional information is necessary to achieve more definitive results than the preliminary ones we have presented here. The continued importance of swidden farming in East Africa, and elsewhere in the tropical and subtropical world, warrants further investigation of this application of the LANDSAT data.

VII. REFERENCES

1. Geertz, Clifford. Agricultural Involution. Berkeley: University of California Press, 1963.

2. Thaman, R. "Remote Sensing of Agricultural Resources." In John E. Estes and Leslie W. Senger (eds.), Remote Sensing. Santa Barbara: Hamilton Publishing Company, 1974, 189-223.

3. Conklin, Harold. The Study of Shifting Cultivation. Studies and Monographs VI. Washington, D.C.: Department of Social Affairs, Pan American Union, 1963.

4. Conant, Francis P.; Reining, Priscilla; and Lowes, Susan, editors. "Satellite Potentials for Studying Subsistence Productivity and Population Deployment." A Report to the National Science Foundation, Washington, D.C., 1975.

5. Conant, Francis P. "Korok: A Variable Unit of Physical and Social Space among the Pokot of East Africa." American Anthropologist, 67: 429-434.

6. Porter, Philip W. "Environmental Potentials and Economic Opportunities--A Background for Cultural Adaptation." American Anthropologist, 67: 409-420.

7. Harrison, G. A., and Boyce, A. J., editors. The Structure of Human Populations. Oxford: Clarendon Press, 1972.

8. Bernstein, Ralph, and Ferneyhough, Dallam G., Jr. "Digital Image Processing." Photogrammetric Engineering and Remote Sensing, 41 (December 1975), 1465-1476.

9. McLoughlin, Peter F. M., editor. African Food Production Systems. Baltimore: John Hopkins Press, 1970.

10. de Wilde, John C. Experiences with Agricultural Development in Tropical Africa. 2 volumes. Baltimore: Johns Hopkins Press, 1967.

Table 1. Modal Responses Used for Training.

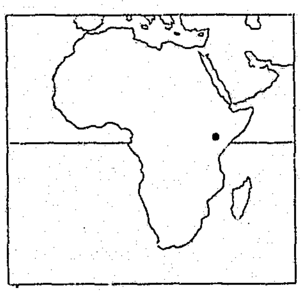
Scene ID	Date		Responses Band 4	Band 5	Band 6	Band 7
1067-07221	9/28/72	count	29	22	46	24
		energy*	.566	.346	.637	1.752
1193-07230	2/1/73	count	28	26	43	24
		energy	.547	.409	.596	1.752
2063-07111	3/26/75	count	30	36	54	27
		energy	.572	.492	.636	1.859

*mw/cm^{2}sr

Table 2. Number and Distribution of Points Classified.* Table entries correspond to areas designated by letter in Figure 5.

Date Scene ID Season	Asar Region Ecozone B	C	D	Riv-erine	West Facing Slopes E	F	Else-where A	G	Total
9/28/72 1067-07221 mid-dry	62	4	19	8	5	9	3	0	110
2/1/73 1193-07230 end-dry	56	13	14	7	5	4	2	0	101
3/26/75 2063-07111 early wet	93	38	22	16	8	1	14	0	192

*The convention was adopted that if a point lay along a demarcation line in Figure 5, the point was assigned to the area to the left of that line.



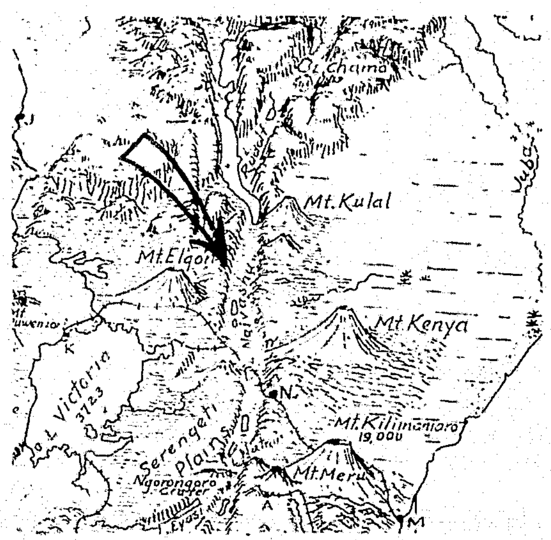


Figure 1. Location of study area in Africa (dot, above) and in East Africa (arrow, right). Based on Lobeck, A. K., "Physiographic Diagram of Africa." New York: Hammond, 1946.

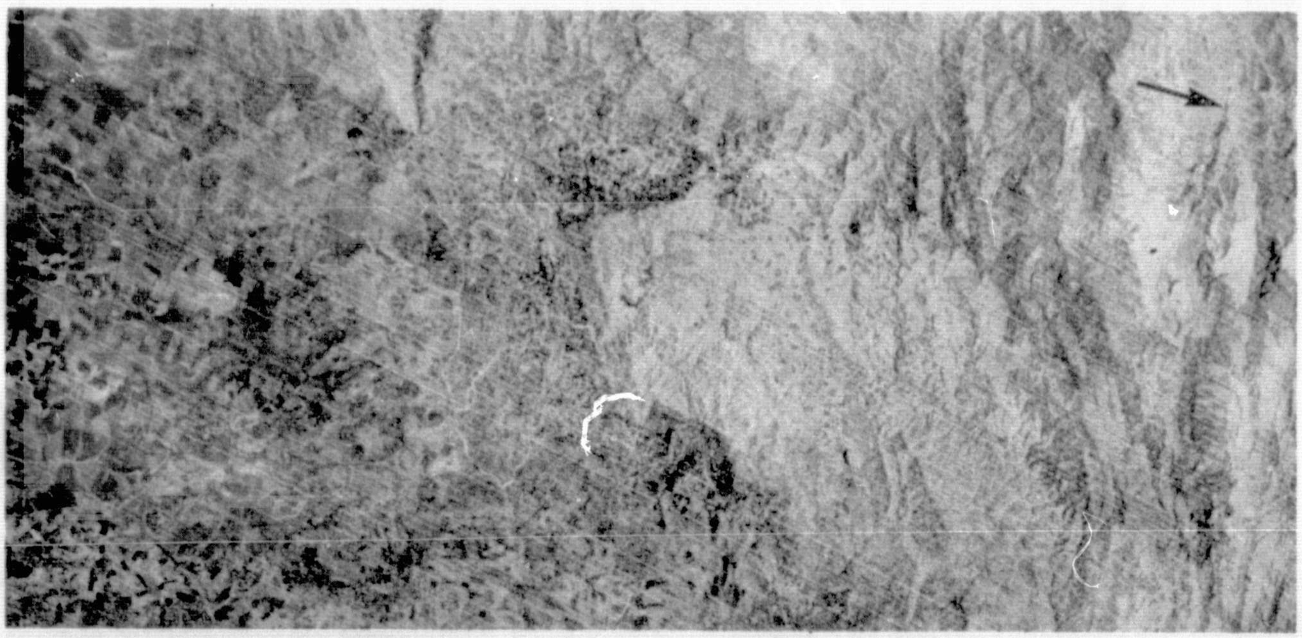


Figure 2. LANDSAT scene 2063-07111, 26 March 1975, Band 7, enlarged to about 1:300,000. Large-area agriculture is visible on the left of the scene, while shifting or swidden cultivation is barely apparent on the right. Arrow indicates study area.

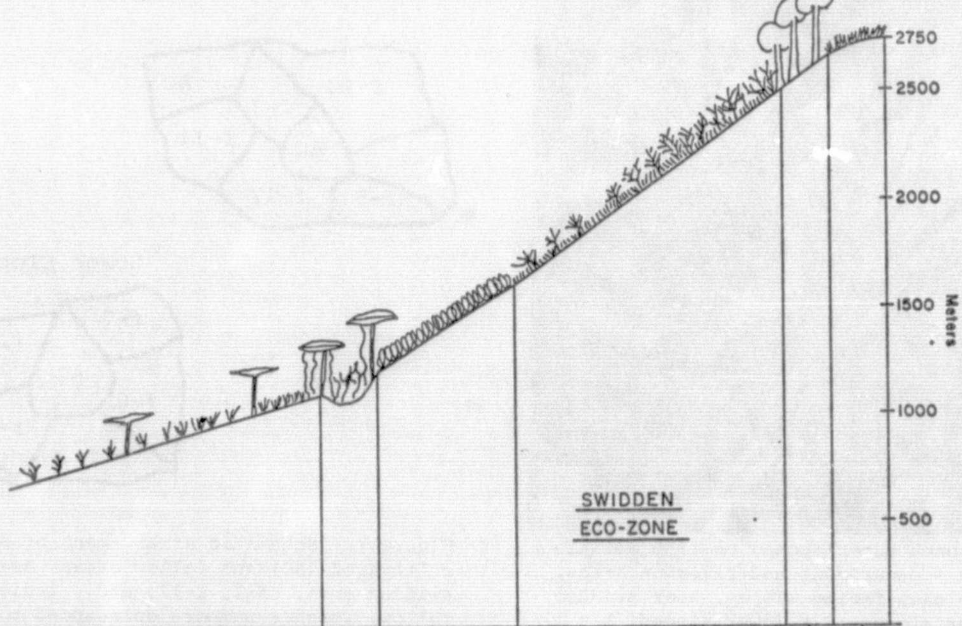


Figure 3. Schematic arrangement of plant cover, West Pokot District, Kenya, according to elevation above sea level. (Schematic courtesy of J. Coiner)

Figure 4. Field Photograph, 1961-62, of the southern portion of the Asar region swidden area used for training. North is to the right. Figure 5 is an aerial photograph of the same location.

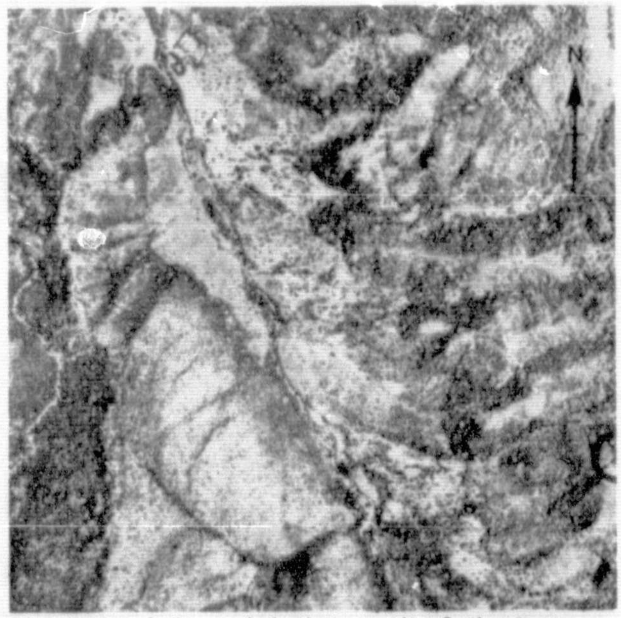


Figure 5. 1956 aerial photograph of the Asar region swidden area used for training, enlarged to a nominal scale of 1:24,000.

Figure 6. Ecozones superimposed on 1956 aerial photography. A = escarpment and riverine areas. B, C, and D are east-facing slopes, Asar swidden area: B = upper slopes, C = lower slopes, D = lower slope training area. E, F, and G are west-facing slopes: E = lower slopes, F = upper slopes, G = high altitude area. The narrow band separating east-facing and west-facing slopes is riverine.

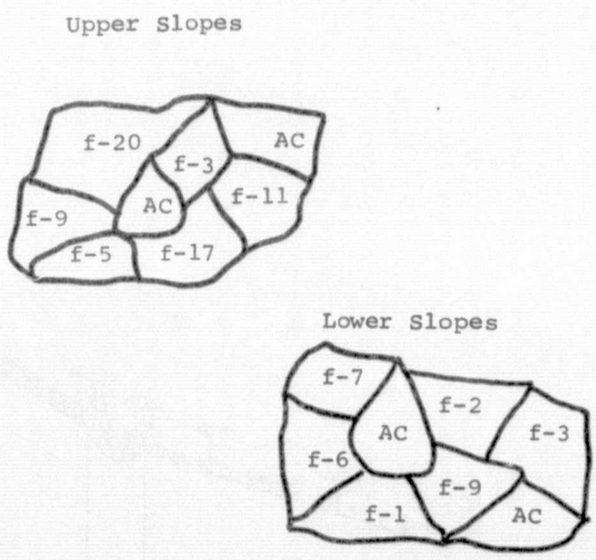
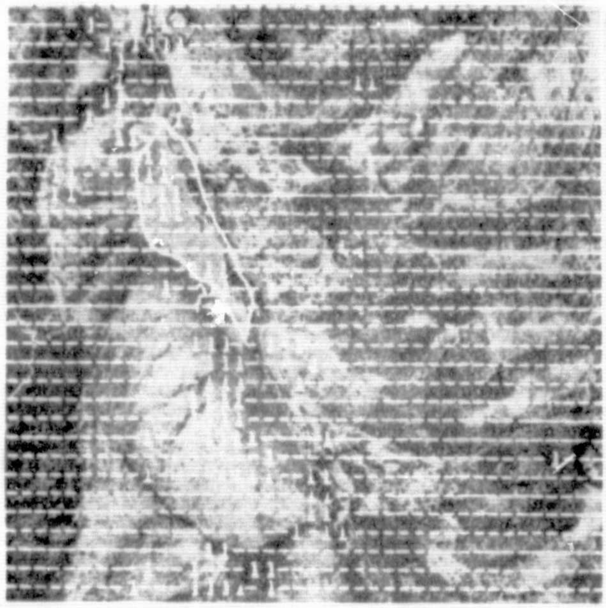


Figure 7. Schematic arrangement of actively cultivated (AC) and fallow areas, Asar region swidden area. f-1, f-17, etc., indicate years in fallow. Higher numbers correspond to ligneous regrowth, and lower numbers to herbaceous.

8a. LANDSAT scene 1067-07221, 28 Sept. 1972.



8b. LANDSAT scene 1193-07230, 1 Feb. 1973.

8c. LANDSAT scene 2063-07111, 26 Mar. 1975.

Figure 8. Classification results generated by computer-assisted analysis of digital LANDSAT data, superimposed on 1956 aerial photography. Training area is outlined; "1" = points classified as swidden, "-" are points not so classified.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to Professor Jerry C. Coiner of the Department of Geography at Columbia University and to Drs. Dwight Egbert and Stephen G. Ungar at the NASA Goddard Institute for Space Studies for their encouragement and advice in the conduct of the present research. The participation of Ms. Tina Cary was made possible through NASA Grant NG 5080. The acquisition of LANDSAT imagery and tapes was made possible through grants from Hunter College (1 June 1976) and the Wenner-Gren Foundation for Anthropological Research (Grant #3177). The 1961-62 fieldwork was part of the NSF/NIMH project "Culture and Ecology in East Africa," organized by Professor Walter Goldschmidt, Department of Anthropology, University of California, Los Angeles.

The authors, however, are solely responsible for the analyses and interpretations presented in this paper.

Francis P. Conant is Professor of Anthropology at Hunter College, City University of New York. He received his BA from Cornell University in 1950 and a PhD in Anthropology from Columbia University in 1960. His main research interests are in ecological anthropology.

Tina Cary is a Graduate Research Assistant in the Department of Geography at Columbia University. She received her BS from Purdue University in 1972. She was a Remote Sensing Data Analyst and Training Specialist at the Laboratory for Applications of Remote Sensing at Purdue University, 1972-1975.

ORIGINAL PAGE IS
OF POOR QUALITY