
The California Cooperative Remote Sensing Project: Final Report

Christine A. Hlavka, Ames Research Center, Moffett Field, California
Edwin J. Sheffner, TGS Technology, Inc., Moffett Field, California

July 1988

NASA

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

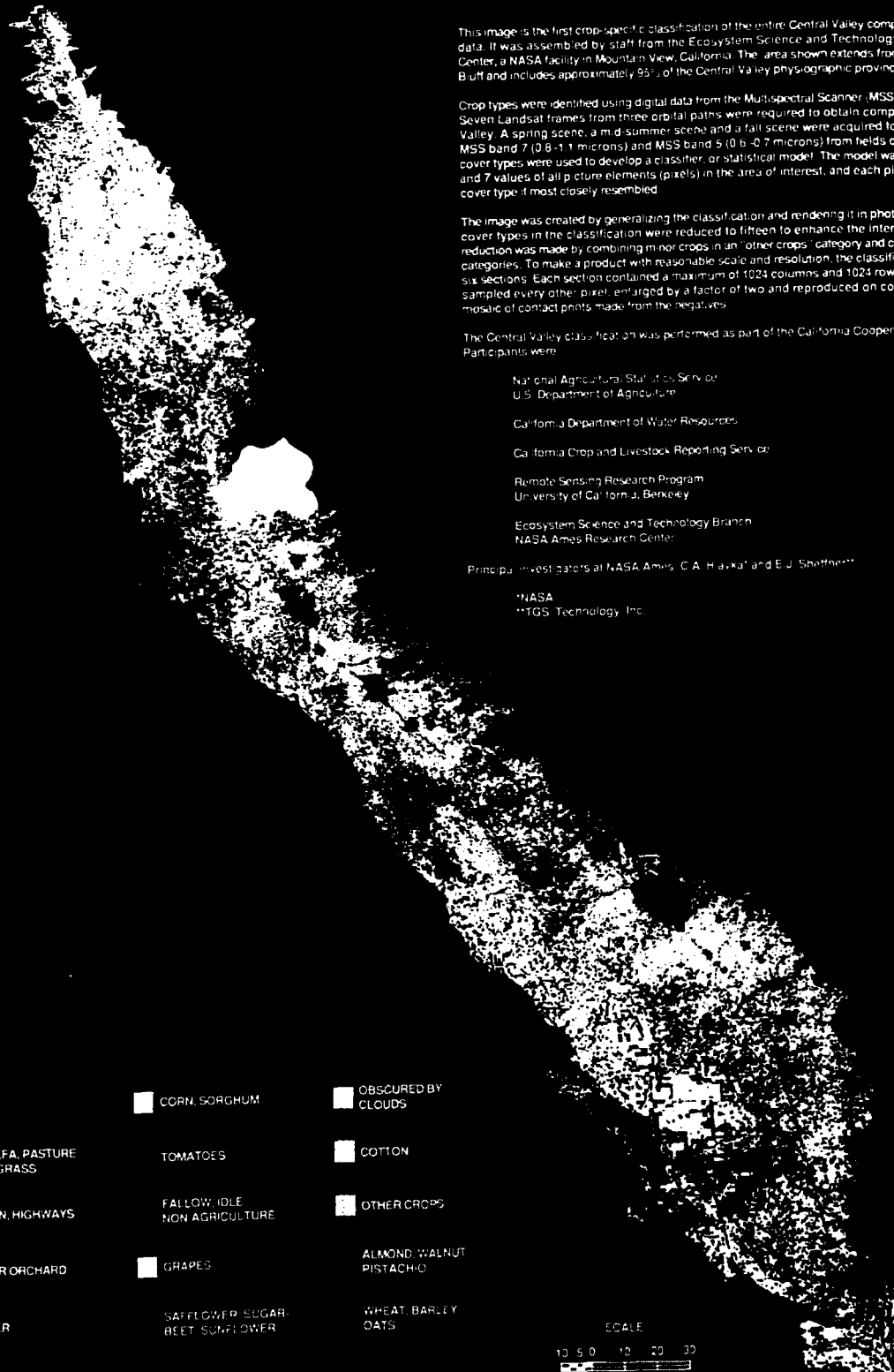
	<u>Page</u>
SUMMARY.....	1
1. INTRODUCTION.....	3
2. BACKGROUND.....	5
2.1 Landsat Data.....	5
2.2 The USDA Use of Landsat Data.....	6
2.2.1 The June Enumerative Survey.....	6
2.2.2 EDITOR Data Processing.....	6
2.2.2.1 JES data encodement.....	7
2.2.2.2 Landsat data processing.....	8
2.2.2.3 Estimation.....	9
2.2.3 EDITOR History.....	11
2.3 CDWR Use of Landsat Data.....	12
3. THE COOPERATIVE AGREEMENT.....	15
3.1 Project Goals and Participating Organizations.....	15
3.1.1 Planning Sessions.....	16
3.1.2 Ground Surveys.....	16
3.1.3 Landsat Data.....	17
3.1.4 Landsat Data Processing.....	17
3.1.5 Software Support.....	17
3.1.6 Data Communications.....	19
3.1.7 Landsat Data Products.....	21
3.2 Early Research Tasks.....	21
3.2.1 The Small Grains Task.....	21
3.2.2 The Multi-Crop Task.....	23
3.3 Development of MIDAS.....	24
3.3.1 Workstation Configuration.....	25
3.3.2 Workstation Communications.....	25
3.3.3 Workstation Software.....	27
3.3.3.1 CIE and ELAS.....	27
3.3.3.2 PEDITOR.....	27
3.4 Plan for the 1985 Inventory.....	28
3.4.1 Technical Approach.....	28
3.4.2 Role of Participating Organizations.....	30
3.4.3 Preparations.....	31
3.4.3.1 PEDITOR testing.....	31
3.4.3.2 Code written for CCRSP.....	32
3.4.3.3 Current year photography.....	32
4. THE 1985 INVENTORY.....	33
4.1 The Test Site.....	33
4.2 Ground Data Acquisition.....	36

4.2.1	June Enumerative Survey.....	36
4.2.2	Transect Survey.....	36
4.2.3	Preparation of the Survey Data.....	38
4.3	Landsat Data Preparation.....	39
4.3.1	Landsat Data Acquisition.....	39
4.3.2	Reformatting of Landsat Scenes.....	40
4.3.3	Formatting Landsat Coverage of Segment Data.....	42
4.4	Landsat Data Processing and Interpretation.....	42
4.4.1	Training for Classification.....	42
4.4.2	Statistics File Editing.....	43
4.4.3	Classification and Aggregation.....	44
4.5	Estimation.....	45
4.5.1	The Original Landsat Estimates.....	45
4.5.2	Revised Estimates.....	49
4.5.3	Experimental Estimates.....	52
4.6	Landsat Map Products.....	63
4.7	Data System Performance.....	65
4.7.1	PEDITOR Software.....	66
4.7.2	Other MIDAS Software.....	66
4.7.3	MIDAS Hardware.....	67
4.7.4	BBN/EDITOR.....	67
4.7.5	Networking.....	67
5.	ANALYSIS OF INVENTORY DATA.....	69
5.1	Data System Performance.....	69
5.2	Accuracy Assessment.....	70
6.	CONCLUSIONS AND RECOMMENDATIONS.....	73
6.1	Data Processing.....	73
6.1.1	Implementation of PEDITOR.....	73
6.1.2	Processing Environment.....	74
6.1.3	Data System Recommendations.....	74
6.2	Use of Landsat Data.....	75
6.2.1	Acreage Estimates.....	75
6.2.2	Landsat Map Products.....	77
	APPENDIX A - PEDITOR MODULES AND FUNCTIONS AS OF 2/20/87.....	79
	APPENDIX B - RECOMMENDATIONS FOR THE 1985 INVENTORY.....	81
	APPENDIX C - ROBUST REGRESSION ESTIMATION.....	87
	REFERENCES.....	92

LIST OF TABLES AND FIGURES

	<u>Page</u>
Table 1: CCRSP Hardware and Software Systems.....	18
Table 2: 1985 Inventory--Data Processing of ARC.....	34
Table 3: California Area Frame Strata Definitions.....	37
Table 4: Landsat Acquisitions for 1985 Central Valley Inventory.....	41
Table 5: 1985 California JES Area Frame Population and Sample Sizes.....	46
Table 6: Regression Parameters for Acreage Estimates (a) AD42	
Table 6: Continued (b) AD43	
Table 6: Concluded (c) AD44	
Table 7: Comparison of Several Types of Landsat Estimates for Rice and Walnuts with CCLRS Estimates	
Table 8A: Landsat Acreage Estimates: Alfalfa.....	53
Table 8B: Landsat and CCLRS Acreage Estimates: Almonds.....	54
Table 8C: Landsat and CCLRS Acreage Estimates: Corn.....	55
Table 8D: Landsat and CCLRS Acreage Estimates: Cotton.....	56
Table 8E: Landsat and CCLRS Acreage Estimates: Grains (Wheat and Barley).....	57
Table 8F: Landsat and CCLRS Acreage Estimates: Grapes.....	58
Table 8G: Landsat and CCLRS Acreage Estimates: Rice.....	59
Table 8H: Landsat and CCLRS Acreage Estimates: Tomatoes.....	60
Table 8I: Landsat and CCLRS Acreage Estimates: Treefruit.....	61
Table 8J: Landsat and CCLRS Acreage Estimates: Walnuts.....	62
Table 9: Robust Estimates of Regression Parameters	
Figure 1: CCRSP Network.....	20
Figure 2: MIDAS System Configuration.....	26
Figure 3: Central Valley Inventory - Data Processing Steps.....	29
Figure 4: 1985 Central Valley Inventory - Study Site and Landsat Frame Locations.....	35

CENTRAL VALLEY OF CALIFORNIA DISTRIBUTION OF MAJOR CROPS—1985



This image is the first crop-specific classification of the entire Central Valley completed utilizing Landsat digital data. It was assembled by staff from the Ecosystem Science and Technology Branch at Ames Research Center, a NASA facility in Mountain View, California. The area shown extends from south of Bakersfield to Red Bluff and includes approximately 95% of the Central Valley physiographic province.

Crop types were identified using digital data from the Multispectral Scanner (MSS) on Landsat 4 and Landsat 5. Seven Landsat frames from three orbital paths were required to obtain complete coverage of the Central Valley. A spring scene, a mid-summer scene and a fall scene were acquired for each frame. Data values in MSS band 7 (0.8-1.1 microns) and MSS band 5 (0.6-0.7 microns) from fields of known crop and other land cover types were used to develop a classifier, or statistical model. The model was applied to the MSS bands 5 and 7 values of all picture elements (pixels) in the area of interest, and each pixel was assigned to the land cover type it most closely resembled.

The image was created by generalizing the classification and rendering it in photographic prints. Twenty-eight cover types in the classification were reduced to fifteen to enhance the interpretability of the image. The reduction was made by combining minor crops in an "other crops" category and combining all orchards into two categories. To make a product with reasonable scale and resolution, the classification was divided into thirty-six sections. Each section contained a maximum of 1024 columns and 1024 rows of pixels. The sections were sampled every other pixel, enlarged by a factor of two and reproduced on color negatives. The image is a mosaic of contact prints made from the negatives.

The Central Valley classification was performed as part of the California Cooperative Remote Sensing Project. Participants were:

National Agricultural Statistics Service
U.S. Department of Agriculture

California Department of Water Resources

California Crop and Livestock Reporting Service

Remote Sensing Research Program
University of California, Berkeley

Ecosystem Science and Technology Branch
NASA Ames Research Center

Principal investigators at NASA Ames: C.A. Haykal and E.J. Shefferson

NASA
MTGS Technology, Inc.

- | | | |
|--------------------------------|-------------------------------------|-----------------------------|
| RICE | CORN, SORGHUM | OBSCURED BY CLOUDS |
| ALFALFA, PASTURE
HAY, GRASS | TOMATOES | COTTON |
| URBAN, HIGHWAYS | FALLOW, IDLE
NON AGRICULTURE | OTHER CROPS |
| OTHER ORCHARD | GRAPES | ALMOND, WALNUT
PISTACHIO |
| WATER | SAFFLOWER, SUGAR
BEET, SUNFLOWER | WHEAT, BARLEY
OATS |



THE CALIFORNIA COOPERATIVE REMOTE SENSING PROJECT:

FINAL REPORT

Christine A. Hlavka
NASA Ames Research Center

and

Edwin J. Sheffner
TGS Technology, Inc.

SUMMARY

The U.S. Department of Agriculture (USDA), the California Department of Water Resources (CDWR), the Remote Sensing Research Program of the University of California (UCB) and the National Aeronautics and Space Administration (NASA) completed a 4-yr cooperative project on the use of remote sensing in monitoring California agriculture. This report is a summary of the project and the final report of NASA's contribution to it. The cooperators developed procedures that combined the use of Landsat Multispectral Scanner imagery and digital data with ground survey data for area estimation and mapping of the major crops in California. An inventory of the Central Valley was conducted as an operational test of the procedures. The satellite and survey data were acquired by USDA and UCB and processed by CDWR and NASA. The inventory was completed on schedule--demonstrating the plausibility of the approach, although further development of the data processing system is necessary before it can be used efficiently in an operational environment.

(Photograph at left shows crop-specific classification of the entire California Central Valley completed using Landsat digital data. A 35mm slide of this photograph is included in this paper and is located in the envelope attached to the inside back cover.)

1. INTRODUCTION

If California were a separate nation, it would have the fifth largest economy in the world. The foundation of the California economy is agriculture. Exploiting the advantages of a virtual year long growing season, massive irrigation projects, abundant tillable land and a variety of soil types and microclimates, Californians grow commercially over 200 different crops.

The agricultural resource is monitored closely. The responsibility lies with several state and federal agencies including the California Department of Water Resources (CDWR) and the National Agricultural Statistics Service (NASS), formerly the Statistical Reporting Service (SRS), of the U.S. Department of Agriculture (USDA). A tally of irrigated lands and estimates of water use is annually compiled by CWR. Because water demand varies with crop type, CDWR conducts crop inventories as well. Annual crop inventories are conducted by NASS as part of its mandate from Congress to collect and distribute state and national agricultural statistics. Both agencies support research on methods to improve data collection and processing procedures so that the required information can be obtained more efficiently and with greater accuracy.

In 1982 a cooperative agreement was signed by CDWR, NASS, the Remote Sensing Research Program of the University of California, Berkeley (UCB) and the National Aeronautics and Space Administration (NASA) - Ames Research Center (ARC). The long-range (4-yr) goal of the joint research project was to develop procedures for area estimation and mapping of the major crops in California using Landsat digital data as the primary data source. The principal funding agency was NASS.

The joint research project was conducted in four stages, each stage corresponding, generally, to a fiscal year (FY):

- FY83 - Evaluation of inventory techniques
- FY84 - Design of inventory experiment
- FY85 - Perform operational test of inventory procedure
- FY86 - Evaluate procedure performance

The following report is a summary of the work done under the auspices of the cooperative agreement. The 1985 inventory experiment and the work performed at ARC in support of it are emphasized. The joint research project was truly cooperative. Participants met regularly, worked together closely, and shared responsibility. Although a joint report on results would have been appropriate, at the request of NASS, separate reports on the 1985 inventory are being submitted by ARC, UCB, and CDWR. This report focuses on the contributions and responsibilities of the staff at ARC--specifically, the Ecosystem Science and Technology Branch (ECOSAT: NASA code SLE).

In the course of preparing for the 1985 inventory, many specific research tasks were completed. Some tasks had significance beyond the context of the cooperative agreement and have been reported on separately. Those tasks are referred to in the

report, and the results are summarized. The reader should consult the references for a more detailed report on specific accomplishments.

The body of the report is divided into five sections. The "Background" provides a description of Landsat and a summary of how NASS and CDWR processed and applied Landsat data prior to the cooperative agreement. "The Cooperative Agreement" describes how the agencies worked together, the goals of the project, the tasks assigned to ARC and the evolution of research within the project. Section 4, "The 1985 Inventory," is a review of the design, implementation and evaluation of the 1985 inventory experiment. The report ends with "Conclusions and Recommendations," as seen from the perspective of ARC.

The California Cooperative Remote Sensing Project involved a great many people. The project was conceived and supported by Bill Caudill (NASS), Bob MacGregor (CCLRS), Glen Sawyer (CDWR) and Ethel H. Bauer (ARC). Management and technical assistance was provided by Bill Pratt (NASS), Richard Sigman (NASS), Randall W. Thomas (UCB), Ron Radenz (CCLRS), Dave Kleweno (CCLRS), George May (CCLRS) and James G. Lawless (ARC). The core programming staff included Martin Ozga (NASS), Martin Holko (NASS), Anthony Travlos (UCB), Paul Ritter (UCB), Robert Slye (NASA), and Gary Angelici (Sterling Software). The primary responsibility for data collection, processing and analysis for the 1985 inventory fell to Jay Baggett (CDWR). He was aided by Catherine Brown (UCB), Louisa Beck (UCB), Charles Ferchaud (CDWR) and others.

Assistance with the preparation of the manuscript was provided by Honoris Ocasio (TGS Technology).

The authors wish to express their appreciation for the efforts of all those who contributed to this undertaking.

2. BACKGROUND

The efforts of the four organizations involved in the CCRSP were not joined by chance but were based on a recent history of common interests and cooperative work in remote sensing and application of satellite data.

2.1 Landsat Data

Landsat is the name of a series of earth observing satellites developed by NASA to monitor renewable and nonrenewable resources. All Landsat satellites are polar orbiting and provide repeat, daytime observations of any area on Earth every 16 days. The multispectral scanner (MSS) aboard Landsat collects imagery from reflected light from the earth's surface in four ranges of wavelengths in the electromagnetic spectrum. These spectral bands are: MSS4: 0.4 μm to 0.5 μm (visual green); MSS5: 0.6 μm to 0.7 μm (visual red); MSS6: 0.7 μm to 0.8 μm (near infrared); and MSS7: 0.8 μm to 1.1 μm (near infrared). Bands MSS5, MSS6, and MSS7 are particularly useful in observations of vegetation because chlorophyll absorbs red light and the mesophyll tissue in plant leaves reflects near infrared radiation. Approximately 10 million picture elements, or pixels, make up a Landsat scene. Each pixel represents the reflectance from 0.8 acre on the ground, and the full scene roughly covers a square area of about 10,000 square miles.

The location of a scene is specified by path and row numbers. A path is traced out by the north to south orbital motion of the satellite during daylight hours on a given day within the 16-day cycle. These paths, which overlap slightly and cover the globe, are cut into rows of scenes, so that each row corresponds to an interval in latitude.

Landsat scenes of the United States (U.S.) are distributed through the Eros Data Center, and may be obtained as photo products or in digital form on computer-compatible tapes (CCTs). Scenes on tape are encoded in four brightness levels, corresponding to the four MSS bands, for each pixel. The tapes are formatted in such a way that the locations of the brightness levels for each pixel, in terms of file number on the tape, record number and byte, are a function of the Landsat scene coordinates (Space Oblique Mercator, or SOM). These coordinates are essentially the scan number within the scene, and position within the scan line. The SOM coordinates can be calibrated to latitude and longitude, or to Universal Transverse Mercator (UTM) coordinates by using information about the location and attitude of Landsat relative to the earth contained on the tape (ref. 1). Greater precision is achieved from calibrations based on regression analysis of sample points whose location are known in both SOM and ground coordinates. The calibration-control-point information is usually obtained by the user of the data by measurements on the Landsat imagery and on high-quality maps, such as the 7.5-min quadrangles at 1:24,000 scale available from U.S. Geological Survey (USGS). Landsat tapes of some areas also contain some control-point information (refs. 1,2). Because the orbit and attitude of Landsat are not perfectly stable, the calibration may differ

slightly between acquisitions of a scene. If more than one Landsat acquisition is used in a study of an area, the SOM coordinates of one acquisition are chosen as the standard coordinates for the study and other Landsat acquisitions are "registered" to the standard. This means that the coordinates of the other scenes are calibrated to the standard, and the file(s) containing brightness data are then reformatted, or "resampled," so that the coordinate systems of all Landsat acquisitions are now the same. Sometimes other geographical data used in the study is also registered to the standard SOM coordinate system.

The features of the Landsat system: the spatial resolution, the repeat coverage, the spectral resolution, the coverage per scene, the reasonable cost, and the availability of the data in both photographic and digital formats, make the data potentially useful in an agricultural inventory.

2.2 The USDA Use of Landsat Data

The USDA, through NASS (SRS), began using Landsat data in the mid 1970's. Landsat imagery is used routinely now as an aid in development of sampling frames for crop and livestock inventories, and Landsat digital data have been used to improve the precision of crop-acreage estimates. Both activities have been in support of the June Enumerative Survey (JES)--the primary mechanism for obtaining large area crop estimates in the U.S.

2.2.1 The June Enumerative Survey

The JES is a survey conducted annually by state (ref. 3). Plots are selected for survey by a stratified random sample (refs. 3,4). Each state is divided into strata based on land use. Strata boundaries are first drawn on enlargements of Landsat imagery and/or aerial photography then transferred to medium-scale maps, usually county highway maps. The area in square miles of each stratum within a county is tabulated.

A random sample of segments, parcels of land usually one square mile in area, is drawn from those strata containing a significant amount of agriculture. Segment boundaries are located on large-scale aerial photography. The photographs are given to enumerators who visit the sites during the JES. The enumerators draw in the field boundaries on the aerial photography and identify the contents of the field primarily through interviews with farmers, and secondarily through windshield surveys. The crop/land-use type may be a crop (e.g., wheat, sorghum, tree fruit, etc.), natural feature (e.g., grass, pasture, etc.), or nonagricultural land use (e.g., commercial, industrial, urban, farmstead). The survey data is used to develop acreage estimates for major crops, by proration by area, i.e. direct expansion (ref. 4).

2.2.2 EDITOR Data Processing

In the latter half of the 1970's, USDA developed a procedure for generating improved area estimates using Landsat digital data in conjunction with the JES data

(ref. 5). Landsat imagery is interpreted by computer using a statistical classification algorithm to label pixels according to crop type or land use. The interpreted imagery is then integrated with digitized geographic boundaries, i.e., JES stratum and segment boundaries, to create tables of pixel counts by crop type/land use. The tables are then correlated and integrated with the JES data to form acreage estimates. The agency uses the Landsat estimates to supplement the proration estimates.

Because of the volume of data in a Landsat scene (about 40 million pieces of information) and the combination of data sources involved in the procedure, automated data processing was a prerequisite for area estimation with Landsat. EDITOR is a software system developed by USDA to perform the data processing required for the Landsat acreage-estimation procedure. EDITOR is based on procedures developed at Purdue University and implemented with an image-processing system called LARSYS. LARSYS techniques were adapted for use with JES survey data to create EDITOR.

EDITOR is a modular system originally written in Sail, Fortran, Rational Fortran (Ratfor), and Macro programming languages (refs. 6,7). Data are passed between modules by writing and reading files. A feature of EDITOR, possibly unique at the time of its creation, is the ability to process a variety of types of data coded in text or binary format.

Three categories of data are manipulated in EDITOR - ground data, Landsat data and statistical data. Ground data consists of information on the location, size, contents and condition of specific fields, and the number of ground sample segments by county, stratum and analysis district. The ground data is maintained in formats suitable for data processing. Landsat digital data is stored on tape as full or half scenes, or is stored on disk in files containing all the data for the segments being analyzed, the data for specified crop types only, or files in which the data has been classified. Statistical data is generated by operations on the ground data and Landsat data.

EDITOR was used by NASS with some technical support provided by ARC. Portions of EDITOR were also used at ARC for research on applications of remote sensing.

The data flow within EDITOR is summarized below. The EDITOR approach to data processing and a version of the EDITOR software were used by CCRSP (refs. 5-8).

2.2.2.1 JES data encodement- Much of the manipulation of ground data files is completed prior to integration with Landsat data. The data collected by the JES enumerators, i.e. the per field information collected from the ground sample segments, are encoded in ground truth files. These files are created in a binary format on a system outside of EDITOR, and are read by EDITOR modules when the acreage estimates are calculated.

Additional files required for the integration of Landsat data with the JES survey are created and used within EDITOR. The boundaries of the JES strata within each county are digitized, a process that converts the information marked on a map

to a digital format. The boundaries of each stratum are treated as polygons. A latitude and longitude for each vertex of a polygon is recorded along with a label associating the polygon with a stratum. Files containing polygon data are referred to as "network files." Longitude and latitude coordinates are calibrated to Landsat SOM coordinates so that the pixels in the scene can be associated with strata. The network files are reformatted to form EDITOR "mask files" so that counts of Landsat pixels within boundaries can be made. In a similar manner, the boundaries and crop type/land use for each field in the JES segments are digitized and encoded in mask files.

2.2.2.2 Landsat data processing- Landsat data is processed to generate pixel counts for each crop type to be included in the acreage report. The estimation technique requires pixel counts both by segment and by stratum. The computationally intensive steps required to generate pixel counts by stratum are performed on a supercomputer.

For the sake of computational efficiency, the Landsat data is prepared in two formats for processing in the EDITOR system. The computer is then "trained" to recognize crop/land-use type on the JES segments. The Landsat imagery is interpreted by the computer and classified imagery is generated. Finally, pixels are counted on the classified imagery with reference to the mask files. These steps in Landsat data processing are described in the following text.

The first type of reformatting is for processing steps associated with JES strata and is performed with software outside of EDITOR. The information in Landsat computer compatible tapes is rearranged so that the brightness values associated with pixels on each scan line on a scene are contained in a single record. Sometimes data from two Landsat observations of a scene are used. As mentioned in section 2.1, the Landsat coordinates vary between two observations. This problem is corrected for by a process called "registration," in which the coordinates of one observation are calibrated to the coordinates of the other. The USDA procedure involves location of control points on both observations of the scene. The first few points are located manually, and then a hundred or so are located with an automated technique on a supercomputer. The brightness for both dates is then interleaved so that eight numbers are associated with each pixel.

After the Landsat scene has been reformatted, the data are extracted for the segments located in the scene. The segment(s) specific digital data is the input for the the second type of reformatting, termed "packing." Packing is one of the unique features of EDITOR. A packed file contains a compressed version of a multi-dimensional histogram, i.e., the number of pixels for each vector of brightness values by segment or by crop/land-use type (as identified by JES enumerators).

The computer is "trained" to recognize crop/land-use type by a process called "clustering" performed on Landsat data packed by crop/land-use type. A cluster can be thought of as a subtype of the crop/land-use type. Each cluster is determined by a combination of factors that affect the appearance of a patch of ground on Landsat imagery. These factors include agricultural practices and soil color. The probabilistic distribution of brightnesses for each crop/land-use type is modeled as a

mixture of multivariate normal (Gaussian) distributions (ref. 9) in which each normal distribution represents a cluster. In EDITOR, an algorithm called CLASSY (ref. 10) divides the pixels into groups, or clusters, so that the shape of the multidimensional histogram, or scattergram, of brightness values for the cluster conform to the that expected for a normal distribution. The number of pixels, band means, and covariance matrix of each cluster are evaluated and assembled, with the crop/land-use type and a number label for the cluster, in a cluster statistics file. A separate cluster statistics file is developed for each Landsat path, because the acquisition dates differ among paths. Each Landsat path is considered a separate "analysis district."

Classification of Landsat imagery is performed by maximum likelihood discrimination (refs. 7,9,11). Each pixel is labeled with the number of the cluster it most closely resembles. Resemblance to a cluster is defined as the likelihood of observation of the brightness values of the pixel if it belonged to the cluster, that is, if the combination of crop/land-use type and other conditions associated with the cluster were true for that pixel. The likelihood is highest near the cluster mean. In subsequent steps, to derive acreage estimates and map products, pixels are associated with a crop/land-use type, i.e, the type associated with the cluster number in the cluster statistics file.

"Aggregation" is the tabulation by cluster number of all the pixels in the area defined by a EDITOR mask file. Aggregation is performed to get pixel counts on strata within each county of a survey, and within each JES segment. The aggregations are used to compute acreage estimates.

2.2.2.3 Estimation- Regression estimates (ref. 4, Chapter 7) make use of two sources of information about the geographical distribution of crops: sampled crop acreages collected as part of JES, and counts of Landsat pixels labeled by crop on classified imagery. Estimation on a regional scale is performed in three steps in EDITOR. A linear relationship between Landsat pixel counts and ground acreage is developed by regression analysis of the classified segment data and JES statistics. The relationship is then applied to pixel counts on classified, full frame, Landsat imagery. In the final step, the estimates for all analysis districts are combined to create a state level estimate for each crop of interest (ref. 12). Estimation on the county scale is performed by a single module in EDITOR. The estimates are described in the following text.

If the land-use map were perfectly accurate, then crops' acreages could be calculated by multiplying the pixels counts by the pixel area (0.8 acres/pixel). Because there is significant errors in the classification, regression is used to estimate the relationship between pixels and acres. Regional estimates are derived by least squares estimates of mean acreage Y_h for a given crop per segment (square mile) on each land use stratum h within each Landsat analysis district, or path, as follows:

$$\text{Est}(Y_h) = y_h + b_h \times (X_h - x_h) \quad (1a)$$

or, equivalently as:

$$\text{Est}(Y_h) = b_0 + b_1 \times X_h \quad (1b)$$

with:

$$b_0 = y_h - b_h \times x_h$$

and:

$$b_1 = b_h$$

where x_h and X_h are JES sample and population (entire strata within the analysis district) pixel counts per segment, y_h is the mean JES sample acreage, and b_h is derived from least squares estimation. The estimate of total acreage is computed as $N \times \text{Est}(Y_h)$ where N is the area of stratum h in square miles, that is, the number of segments required to cover the stratum.

This type of estimate is generally more accurate than direct expansion wherein total acreage is estimated by $(N/n) \times y_h$, because Landsat pixels counts are used to correct for the difference in crop prevalence between the sample and the population (stratum/path) as a whole.

The improvement in accuracy depends on the correlation r between pixels and acreage, that is, the variance of $\text{Est}(Y_h)$ is approximately $[(1 - n/N)/N][1 - r^2]\text{Var}_y$ which can be compared to a variance of $[(1 - n/N)/N][\text{Var}_y]$ for a direct expansion estimate.

These estimates are then summed over strata and analysis districts to form the regional crop estimates. The standard error for each estimate is computed using the standard formula (ref. 4, Chapter 7). Each estimate is statistically independent of the others, therefore the standard error for the regional estimates are computed by root mean squares of sums of standard errors for the district/stratum estimates.

Equation (1) defines a linear relationship between the $\text{Est}(Y_h)$ and X_h . A low value (less than 0.8) of the slope term b_h compensates for a tendency for other crops or types of land use to be identified as the crop of interest in the Landsat classification. Conversely, a tendency for the Landsat classification to under-identify the crop is corrected by a high value for b_h . Usually, b_h is computed on each stratum/path, thus allowing for possibly different patterns of confusion among crops and land use types.

County estimates are derived using a modification of standard least squares regression developed by Battese and Fuller (refs. 13-15), with NASS support and collaboration. The intercept in equation (1) is altered on a per county basis, in the estimation of parameters for a linear model of the relationship between acreage Y and pixels X includes a "county effect." In ordinary least squares regression, the regression line goes through the means point x, y , as in equation (1). The

Battese-Fuller (B-F) line, lies between the line in equation (1) and a parallel straight line going through county means point:

$$Y = y_{h,c} + b_h \times (X_{h,c} - x_{h,x,c}) \quad (2)$$

The position of the B-F line is determined by a factor d as follows:

$$\text{Est}(Y_{h,c}) = y_h + b_h \times (X_{h,c} - x_{h,c}) + d \times D \quad (3)$$

where D is the vertical distance between lines (1) and (2). The factor d is computed to minimize the standard error of the estimate. This value of d turns out to be the proportion of the variance(VAR) in the residuals of equation (1) due to "county effect":

$$d = \text{VAR}(\text{between counties})/\text{VAR}(\text{total}) \quad (4)$$

The value of $\text{Est}(Y_h)$ is computed using equation (3) and then adjusted so that the estimates of total acreage for the counties in an analysis district add up to the regional estimate. County estimates are formed by summing district/stratum B-F estimates in the same way the regional estimates are computed.

2.2.3 EDITOR History

The first Landsat satellite was launched in 1972. The following year, NASS began development of EDITOR. The system was completed in 1978. EDITOR was used first, and has been most successful, for crop area estimates in the Midwestern states. Large field size, relatively short growing seasons, and the small number of crops grown make the Midwest particularly suitable for inventories with Landsat data. One acquisition per Landsat scene is sometimes sufficient to be able to identify the crop(s) being surveyed. The NASS program with EDITOR expanded so that by 1983 Landsat based estimates for seven states were being generated annually.

EDITOR was written at the Center for Advanced Computation at the University of Illinois in association with NASS and ARC. It has undergone amendment and modification since 1978 at NASS and ARC, but the basic processing steps have not changed. In the early 1980's EDITOR was operated by NASS on a PDP10 computer at Bolt, Berenek, and Newman (BBN) in Cambridge, Massachusetts. The more computationally intensive procedures were performed on the Cray computer at ARC. At that time, the agency made a decision to rewrite the software resident on the BBN system so that it could be operated on a number of different machines. Because ARC was familiar with EDITOR, NASS contracted with NASA to undertake the bulk of the recoding. Work on the new code, called PEDITOR for portable EDITOR, began in 1983.

PEDITOR was completed to the satisfaction of NASS in 1985. It was installed on an IBM system for agency use with a link to a commercially operated Cray computer. All, or part, of PEDITOR has been implemented on a VAX 11/780 (VMS), a SUN2 workstation (UNIX), and the MIDAS workstations (XENIX) at ARC, UCB, and CDWR. PEDITOR was used by CCRSP for the Central Valley inventory in 1985. Much is written about

PEDITOR and the MIDAS workstation in the following pages. The PEDITOR and MIDAS projects occurred concurrently with CCRSP, and several staff members from NASS, ARC and UCB worked on more than one of the projects. However, the three projects were administratively and managerially separate. The decisions taken by CCRSP, discussed below, to use PEDITOR as the primary data processing package for the 1985 inventory and to attempt to perform most of the data processing on MIDAS meant that the fate of the three projects became intertwined. PEDITOR and MIDAS are discussed in some detail in this report because it is impossible to evaluate the results of the 1985 inventory without knowledge of the history and operational characteristics of the hardware and software systems used.

2.3 CDWR Use of Landsat Data

Among its many achievements, California is the most populous state, the most expensive state in which to live, and the state with the most comprehensive program of water management. Water management is mandated by the needs of agriculture and the peculiar propensity of Californians to settle where the water isn't-- approximately 75% of the state's population lives south of the Tehachapi Mountains in a region that receives only 10% of the state's annual precipitation. Since 1957, the state has operated under a master plan for the development and allocation of its water resources. The CDWR was assigned the task by the State Legislature of periodically updating and supplementing the plan.

CDWR operates an ongoing inventory program to meet its information needs. The department generates land-use maps at 1:24,000 scale that include crop-coverage information in agricultural areas. The size of the state and cost of gathering information preclude compilation of new land use maps every year. In fact, the state is covered on a 7-yr cycle, wherein several counties are mapped each year.

CDWR began work in the late 1970's with the Remote Sensing Research Program at UCB, the Remote Sensing Unit of the Department of Geography at the University of California, Santa Barbara, and ARC to develop crop survey methods using Landsat imagery. The project, known as the Irrigated Lands Project (ILP), consisted of four tasks directed toward development of procedures for:

1. Estimation of irrigated land area using manual interpretation of Landsat photoproducts,
2. Estimation of irrigated lands using automated classification of Landsat digital data,
3. Crop-type mapping by manual interpretation of Landsat data, and
4. Crop-type mapping by automated classification of Landsat digital data.

The four tasks were reported on in the fall of 1982 (refs. 16,17).

The procedure for estimation of irrigated lands with Landsat digital data met accuracy specifications set by CDWR. The manual technique was adopted first for departmental use, and the automated technique is becoming operational.

The multi-crop identification procedure using Landsat digital data was an extension of the procedure for identifying irrigated land. However, the irrigated-lands inventory was carried out for the entire state, and the multi-crop research was limited to a pilot test in a localized area.

The test site for the automated multicrop classification procedure was a 30-min block in the Sacramento Valley. The area was stratified according to the prevalence of irrigation as observed on a series of dates covering the growing season. Landsat digital data within each stratum was classified in order to identify the crops grown in each field within the test site. The results of the test indicated that the procedure worked well for some crops and crop groups (e.g., rice, small grains, and orchards), and that additional work might prove fruitful.

3. THE COOPERATIVE AGREEMENT

3.1 Project Goals and Participating Organizations

In early 1982, the converging interests of NASS and CDWR stimulated the growth of a joint research project. Both agencies wished to continue to pursue the use of Landsat digital data and imagery for crop identification and area estimation in California. The interests of NASS focused on how much was grown at state and county levels, while CDWR was concerned with the local and statewide distribution of crops as well. The California office of NASS, the California Crop and Livestock Reporting Service (CCLRS), was familiar with, and supported, the research goals of the national office and recognized the potential for sharing information with CDWR. The possibility that a single procedure could generate products satisfying the needs of both agencies enhanced interest in the project.

The California Cooperative Remote Sensing Project (CCRSF) was administered under a "memorandum of understanding" (MOU) or cooperative agreement. The goal of the program was to, "... determine the extent to which agricultural remote sensing data can be used in the various State and Federal information programs in California, and to explore the possibility of sharing this technology in continuing State and Federal programs." The inclusion of UCB and ARC was because of their expertise in applications of remote sensing and their history of collaboration with NASS and CDWR. The MOU for a 4-yr project was signed in the spring of 1982. The bulk of the funding to support the work was to be provided by NASS.

The obligations of the four participants were specified in the MOU. Ames Research Center agreed to:

1. Cooperate and consult with other organizations at all stages of the project.
2. Participate in research and development of remote sensing techniques applicable to California agriculture.
3. Perform Landsat MSS full-frame classifications.
4. Provide software support for CDWR digitizing equipment.
5. Provide software support for putting CDWR files into suitable format and transferring them to BBN for processing.
6. Provide high-altitude flight data.
7. Provide photo and map products.
8. Test the EDITOR code as developed within the cooperative project.

The major components involved in the project are described in the following subsections. The roles of the coop members are outlined, with elaboration on ARC's participation.

3.1.1 Planning Sessions

All decisions about the operation of CCRSP were made by representatives from the participating organizations during regularly scheduled meetings. Project meetings were held approximately every other month during the first three stages of the project. The meeting schedule varied in the 6 mo proceeding the 1985 inventory and during the analysis phase. During periods of intense effort, meetings were held as often as every other week.

The meetings were chaired by the representative from NASS. David Kleweno filled that role from the start of CCRSP until the summer of 1984. He was replaced by George May who worked as project coordinator until January 1986. No representative from NASS attended the meetings after January 1986. During the final year of CCRSP, project meetings were chaired by Randall Thomas of UCB, but no individual was designated as project coordinator.

The CCRSP meetings were used for presentation of progress reports, discussion of issues, assignment of tasks, and planning sessions. Perhaps the greatest benefit derived from the meetings was the opportunity they gave the sponsoring agencies, particularly NASS, to maintain the focus on their priorities. It was of value to the CCRSP research staff to receive ongoing evaluation and direction from the ultimate users of the research. These benefits were lost the final year of CCRSP because NASS was unable to send a representative to the meetings.

In addition to the regularly scheduled meetings of CCRSP, project reviews were held semi-annually, usually around the first of the year, and early summer. However, no review was held between September 1984 and the final review in October 1986.

3.1.2 Ground Surveys

Ground surveys were required at various times during the course of the project. The surveys were conducted by CDWR, UCB, and CCLRS.

The CDWR provided ground data to the program from surveys conducted prior to CCRSP and from surveys designed for CCRSP. The small grains task undertaken early in CCRSP (see 3.2.1) used ground data from CDWR provided on 7.5' quadrangle maps. The data were collected as part of the on-going, field-level data-collection effort of the agency.

Ames Research Center assisted ground survey efforts by providing high altitude photography. The photography came from the High Altitude Missions Branch at ARC which acquires aerial photography and other airborne sensor data for research projects. The data are collected by U-2 and ER-2 aircraft operated by NASA. One of the products generated by the branch, high altitude, color-infrared photographic

transparencies at a scale of 1:126,000, is particularly useful for field location and crop identification. High altitude photography was used by CDWR and UCB for work done within CCRSP prior to the 1985 inventory, and it was used by USDA enumerators during the 1985 JES (see 3.4.3).

3.1.3 Landsat Data

Landsat digital data and hard copy imagery was required during all phases of CCRSP. NASS ordered the Landsat data from the EROS Data Center.

Digital data for research prior to the 1985 inventory, and the 1985 inventory data, were sent to ARC where they were entered into the CCT library of ECOSAT. Photoproducts of the imagery prior to the 1985 inventory were also sent to ARC. Photoproducts for the 1985 inventory were sent to CDWR where they were used for scene-to-scene registration and for general reference. The photoproducts were 1:1,000,000 scale black and white transparencies or prints of individual Landsat bands, usually MSS bands 5 and 7, for each scene of interest.

3.1.4 Landsat Data Processing

ARC is the site of one of the most advanced computational facilities in the U.S. The capabilities of the systems available to the ARC staff far exceeded those available to the other CCRSP participants. Although network links made it possible for personnel from CDWR or UCB to operate the machines at ARC from off-site locations, there were numerous instances during the project when processing was done by ARC personnel to complete the processing more efficiently. CCRSP data processing needs were assigned the highest priority by ECOSAT staff.

The most computationally intensive computer processing required by CCRSP was full-frame classification. When performing a maximum likelihood classification of a pixel with EDITOR, the discriminant function for each class (cluster), a quadratic function in the Landsat reflectance values, is computed. The total number of arithmetic operations is approximately proportional to:

$$PB^2C$$

where: P = number of pixels in the scene(about 10 million)
B = number of bands in the Landsat data set(four or more)
C = number of classes(clusters, as many as 255 for CCRSP)

Because of the billions of arithmetic operations required, full-frame classification could be done efficiently only on a supercomputer; for CCRSP, the Cray X-MP at ARC.

3.1.5 Software Support

The CCRSP required sophisticated data handling for preparation, operation and evaluation of the inventory. Table 1 is a summary of the sites, systems, and software used frequently during the course of the project.

TABLE 1.- CCRSP HARDWARE AND SOFTWARE SYSTEMS

Site	System	Operating system	Software		
BBN ^a	PDP10/20	Tenex	EDITOR PEDITOR		
ARC ^b	Cray X-MP	COS	CLASSY CLUSTER WARP BCORR COMPILE AGGR AMERGE		
			VAX 11/780	VMS	PEDITOR(partial) REFORM
			MIDAS	XENIX	PEDITOR ELAS CIE
			SUN	UNIX	PEDITOR(partial)
RSRP ^c	NOVA		DIANA		
	MIDAS	XENIX	PEDITOR ELAS CIE		
CDWR ^d	MIDAS	XENIX	PEDITOR ELAS CIE		

^aBolt, Berenek, and Newman, Cambridge, Massachusetts

^bAmes Research Center, Mountain View, California

^cRemote Sensing Research Program, Berkeley, California

^dCalifornia Department of Water Resources,
Sacramento, California

No analyst or research group was familiar with all the systems when CCRSP began. Indeed, some of the systems, such as MIDAS, didn't exist. Analyst training occurred concurrently with the development of the program. In general, the flow of training information descended the hierarchy of experience within CCRSP, particularly experience with EDITOR/PEDITOR software, and passed from NASS to ARC to CDWR and UCB.

EDITOR, now PEDITOR, emerged as the primary software system for the inventory. It is a complex package that contains a large number of somewhat inflexible, operationally independent programs. The system performs all functions needed to create an area estimate.

EDITOR training for ARC analysts began in the spring of 1982 and continued through 1984. It was aided by a short training program conducted by NASS in Washington and an EDITOR operations manual compiled by Martin Holko of NASS (ref. 18). Ames Research Center's experience with EDITOR was also aided by participation in an agricultural inventory of the Snake River Plain in Idaho. The inventory was performed in 1983-84, and EDITOR was used for data processing (ref. 19).

Ames Research Center assisted other participants in CCRSP with their data processing requirements as needed. The assistance included consultation on EDITOR/PEDITOR processing, ELAS and CIE training on MIDAS, system operations on the VAX 11/780 in ECOSAT, and Cray job setup and submittal. The bulk of the assistance provided by ARC occurred during the first stage of the project, when much of the data processing was done at BBN, and during the 1985 inventory, when ARC was the site for all of the large-scale data processing.

3.1.6 Data Communications

Data communication links were crucial to the operation of CCRSP. ARC was the hub of a network linking all CCRSP participants. The network was provided to transfer data for processing, maintain an electronic mail service, and to update PEDITOR software. The CCRSP network is illustrated in figure 1.

Data communications within CCRSP were maintained jointly by ARC and UCB. The primary network software was Kermit, supplemented by Decnet, Arpanet, UUCP, and Telenet when and where appropriate.

As CCRSP began, it was assumed that much of the data would be processed at BBN. Software was needed by CDWR to generate files in, and convert files to, EDITOR format. Additionally, CDWR needed file transfer and communication capabilities with BBN for data processing. Ames Research Center provided CDWR with two network links to BBN. Both links required connection over public access telephone lines. One link, Telenet, was accessible directly from Sacramento; the other link, Arpanet, required access to ARC via a telephone line and a subsequent connection to a Arpanet node.

LANDSAT DATA PREPARATION

LANDSAT DATA ACQUISITION
↓
DATA REFORMATTING
↓
SCENE-TO-SCENE REGISTRATION
↓
SIX-BAND TAPE COMPILATION

GROUND DATA PREPARATION

GROUND DATA COLLECTION
JES DATA-TRANSECT DATA
↓
GROUND DATA FILES
↓
SEGMENT DIGITIZATION
↓
SEGMENT REGISTRATION
↓
SEGMENT MASK GENERATION

CLUSTERING

DATA PACKING

CLUSTERING

STATISTICS FILE EDITING

CLASSIFICATION AND AGGREGATION

STRATA MASK GENERATION

STRATA MASK EDITING

SAMPLE CLASSIFICATION

FULL FRAME CLASSIFICATION

AGGREGATION

ESTIMATION

REGRESSION WITH JES SAMPLE

REGIONAL AND COUNTY ESTIMATION

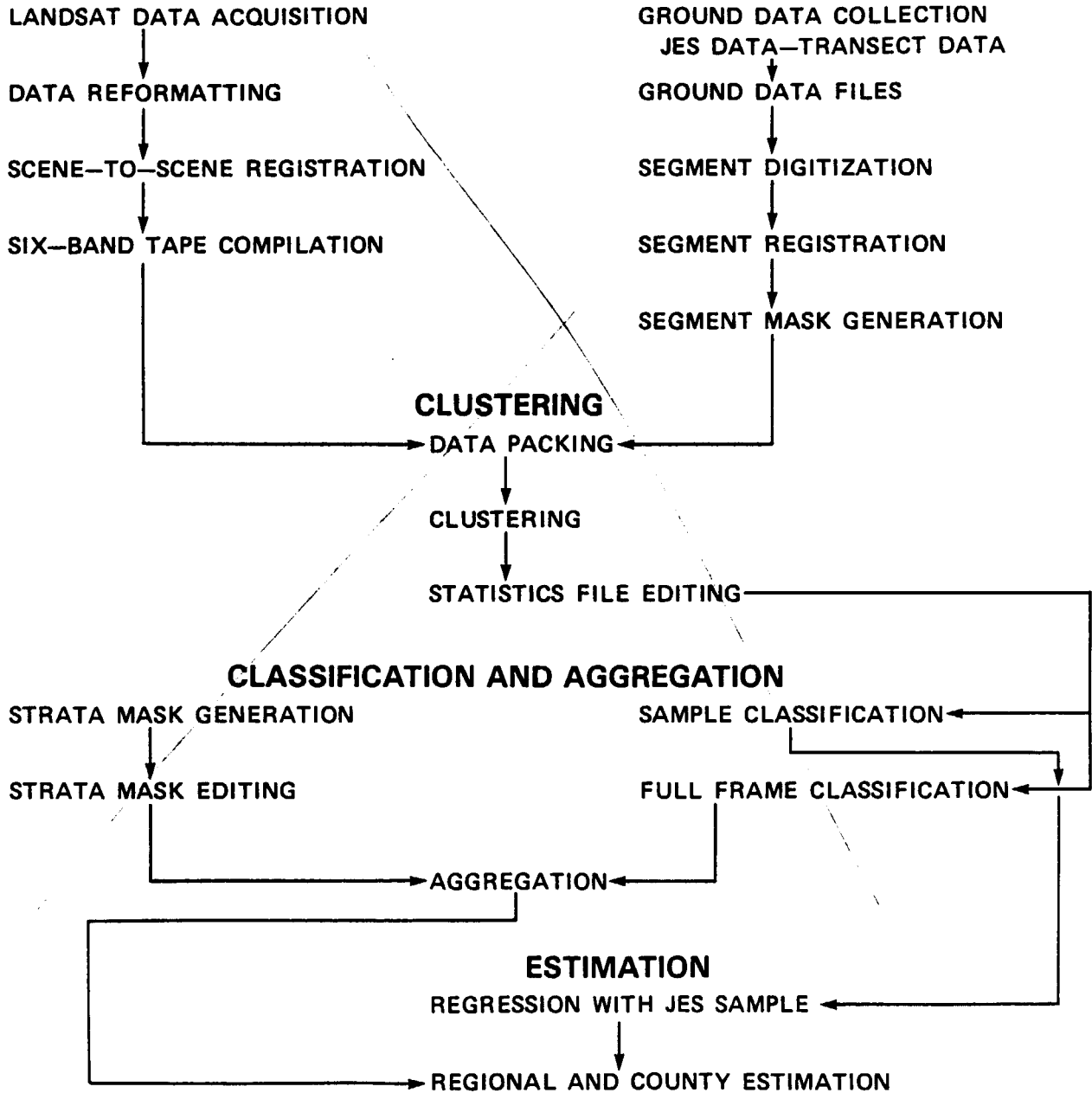


Figure 1: CCRSP Network

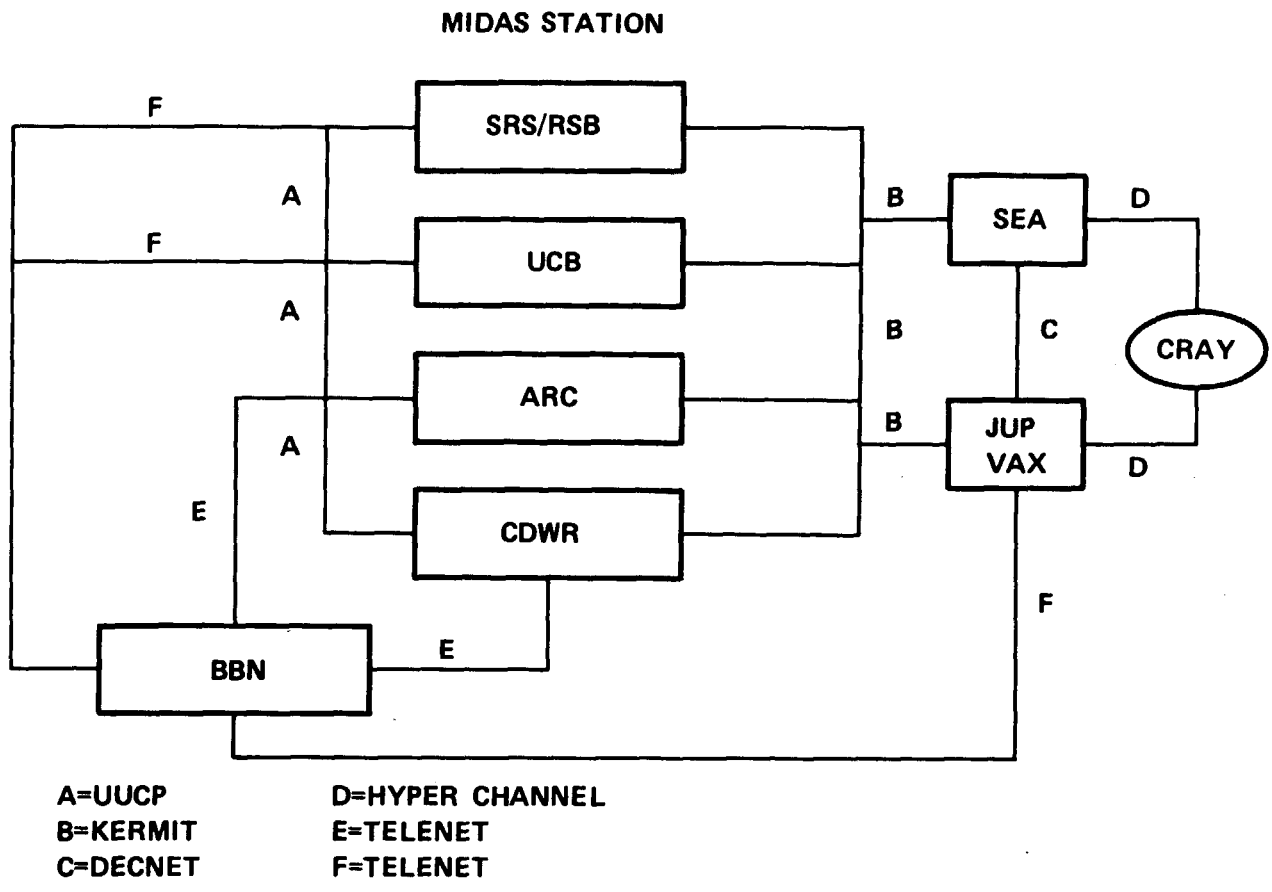


Figure 1: CCRSP Network

3.1.7 Landsat Data Products

Photo and map products was supplied by ARC to the CCRSP participants at various times during the project. These included 1:24,000 scale quadrangle maps of small grains generated by the two small-grains classification techniques described in section 3.2.1, aerial photography enlarged for use by field enumerators, and a mosaic of the Central Valley classification.

3.2 Early Research Tasks

The first phase of CCRSP was an evaluation of existing inventory techniques. The evaluation was considered a prerequisite for design of the 1985 inventory. When CCRSP began, the only large-scale, multicrop inventories in the U.S. based on Landsat digital data were in the Midwest. The California environment and California agriculture differ substantially from the Midwest (e.g., greater variety of crops, longer growing season, greater variety in topography and soils). There was no basis of assuming that inventory procedures developed and tested in the Midwest would be appropriate in California. The 1985 inventory was intended to produce both acreage estimates and map products; no previous large-scale inventories had attempted both from a single procedure. The early research tasks also provided the participants, other than NASS, with an opportunity to become acquainted with the algorithms and approach to data processing of EDITOR.

Two early research tasks were the development of a procedure for identification and mapping of small grains, and an evaluation of techniques for multi-crop labelling.

3.2.1 The Small Grains Task

CDWR had experimented with a manual technique for mapping small grains (wheat, oats, barley) with Landsat data. The technique was based on the phenology of small grains and the appearance of the phenological stages in Landsat imagery.

The phenology of grain is distinctive because it is an early crop. In California, grain is prepared and planted in late fall. The field remains bare until the grain emerges in winter. It grows to full height in early spring, then matures and is harvested in late spring or early summer. The CDWR technique involved labeling a field on Landsat photoproducts according to whether or not it appeared covered with green vegetation on three observation dates. If a field was labeled as bare on a fall observation date, green in early spring, and bare or stubble covered in early summer, the field was labeled grain.

Because the results of the CDWR technique were promising, an early research task for CCRSP was to test methods for automated identification of small grains in California using logic similar to the manual technique. The research on identification of grains was considered useful because it addressed two issues related to identification of multiple crops, i.e., what labelling techniques work well in

California, and how many dates of Landsat imagery are required for successful crop identification in the California environment.

The number of Landsat observations that would be required for a multicrop inventory was vital information because of the cost of acquiring the imagery and the adjustments that would have to be made in analysis procedures if more than two observations were needed. The EDITOR procedure, for example, was not equipped to process more than eight bands (four bands from two Landsat acquisitions) of data. It was postulated that the long growing season in California would mandate the use of three or more Landsat observations for accurate crop identification. The CDWR experience with manual labelling of grains supported that assumption.

The small grains research was accomplished with Landsat data from five observations taken during the 1981 growing season. The earliest Landsat acquisition was 14 November 1980, the last acquisition was 6 July 1981. The test site was Yolo County in the southern part of the Sacramento Valley. The JES segments were selected as training areas for the classifiers. The crop/land-use identifications for the fields within the segments came from current year CDWR inventory data. Classifications were generated for all two, three and four date combinations, and for the five dates taken together. The classification techniques mimiced the logic for the manual CDWR approach in that initial classifications were made on Landsat data from the individual observations, and final class (grain/nongrain) assignments were a function of the combination of single-date classes.

Classification accuracy was measured using the percent of pixels in the JES segments identified correctly. EDITOR contained software to generate the statistics. The classifications were also evaluated for per-field accuracy by visually comparing Landsat map products with CDWR land-use maps. Grain acreage estimates were developed and compared to CDWR figures from its comprehensive land use maps of the test site.

Two of the grain identification methods were developed and tested at ARC--"layered classification" and "band ratio thresholding" (ref. 20). In the layered classification approach, a separate maximum likelihood classification was generated for each date. All pixels were labelled grain or nongrain. The single-date classifications were combined, i.e, layered, to produce a composite classification in which each pixel was given a unique number depending on which dates it was labeled grain. With each combination of dates, pixels labeled grain on all dates were labeled as grain; pixels labeled grain on no dates were labeled "nongrain." The labels for "mixed" classes were assigned at the discretion of the analyst.

The band-ratio thresholding technique used an adaptation of the technique for automated mapping of irrigated lands developed in the ILP (section 2.2) to take the place of the manual interpretation involved in the small grains procedure developed by CDWR. It has been shown that the ratio of a near infrared band (MSS7) to the visual red band (MSS5) is well correlated with the amount of green biomass (ref. 16). The ILP technique (task 2, section 2.3) labeled all pixels with band-ratio values above a cut-off value of 1.0 as covered with green vegetation on the date of Landsat observation. The band ratio thresholding technique was a modified

version of the ILP technique, wherein a threshold value was selected for each date in such a way as to minimize errors of omission in the identification of grain. Layered classification and band-ratio thresholding were compared to an approach developed by UCB. The UCB approach used Kauth transformed data in the analysis (ref. 21). The three techniques produced similar results in terms of acreage estimates and measures of map accuracy, but the band-ratio threshold approach produced more visually pleasing maps and better definition of field patterns. Three dates of Landsat observations produced better results than one or two dates, but no important improvements were achieved with four or five observations.

The experiments at ARC were completed in 1982 and were reported by Sheffner et al. (ref. 20). The experiments on small grain conducted by UCB continued. The technique UCB developed, called polygon vector analysis, was reported on during the CCRSP semi-annual review in Berkeley in February 1984.

3.2.2 The Multi-Crop Task

The results of the small grains experiment indicated that classification technique was probably not critical to the accuracy of Landsat map products or estimates. The EDITOR approach, maximum likelihood classification on combined imagery from all Landsat acquisitions, was, therefore, chosen as the method for multi-crop survey and mapping within the CCRSP. Given the schedule of the project, it was prudent to choose a technique which was fully implemented unless another technique was clearly superior.

Two key issues remained to be addressed prior to completion of the design of the 1985 inventory experiment. Although the small grains work indicated that three Landsat acquisitions were optimum for grains classification, the number of acquisitions needed for multiple crops remained unresolved. It was also of interest to determine whether a transformation of Landsat data, the brightness values in the four MSS bands, would lead to better classification accuracies. In 1983-84, a series of experiments were conducted to settle these and other issues. The experiments were designed by UCB and were carried out in conjunction with ARC.

An expanded version of the small-grains data set was used for testing (section 3.2.1). Approximately 60 JES segments were used in the analysis. Classification of the data within the segments was done using all two-, three-, and four-date combinations and all five dates. For each classification, the correlation with CDWR ground data was determined. All data processing was done with EDITOR.

The combination of an early spring date and two summer dates produced the most accurate classifications. No significant improvement was achieved with an additional acquisitions (one to two) either earlier or later in the growing season.

Tests on acquisitions were run concurrently with tests on data compression options. The data-compression tests were necessary because of the eight-channel limit in EDITOR/PEDITOR processing. Extending the channel limit would have resulted in costs for software development and data processing. Three data-compression options were tested. The options were:

1. Four MSS bands from two acquisitions (no compression)
2. MSS bands 5 and 7 only (two, three, and four dates)
3. Linear combinations of MSS bands designed to measure vegetation greenness and scene brightness (Kauth transformation)

Option 1 was investigated by NASS (ref. 22). Options 2 and 3 were tested by UCB and ARC. For the latter two options, all MSS5 and MSS7 classification and estimation tasks were performed by ARC personnel. The Kauth transformations were applied to the Landsat data at ARC using the Video Image Communications and Retrieval (VICAR) software package, developed at the Jet Propulsion Laboratory in Pasadena, on an IBM-360. Ames Research Center also assisted UCB with data processing on the Kauth transformed data set.

The tests showed that MSS bands 5 and 7 generated results comparable to the other data compressions indicating that transformations or extension of the eight channel limit were not necessary.

The use of JES samples to train the Landsat classifier and to develop the regression lines used for estimates with the classified Landsat data tends to bias estimates of classification accuracy and derived acreage estimates. This bias is due to the fact that accuracy, and correlation with ground "truth," is generally higher on areas used to train the classifier than on the image as a whole. The two-date study by NASS (ref. 22) included an investigation of the magnitude of this bias. The JES segments used were split into two non-overlapping sets, "set A" and "set B." Two separate classifications were made (one used set A for training and the other used set B). Correlation to ground "truth" was measured with each set, a total of four correlations (two for each classification). Correlations were substantially higher when the same set of segments was used for training and correlation than when one set was used for training and the other set was used for correlation. The result may have been due to the small sample sizes involved.

As a result of the NASS test, the plan for the 1985 inventory specified separate ground-sample units for development of the classifier and accuracy assessment.

3.3 Development of MIDAS

Microprocessor Image Display and Analysis (MIDAS) is a prototype, microprocessor-based workstation developed at ARC under the sponsorship of NASS and the U.S. Geological Survey. The sponsoring agencies wished to determine if a workstation could perform most of their Landsat-related data processing, including both computation and interactive display of imagery.

MIDAS was designed to take advantage of the then new technology in 16-bit microprocessors. The workstation was built with "off-the-shelf" components. MIDAS was one of the first attempts to assemble a workstation that was reasonably priced and that would provide an analyst access to software tools and machine memory

capacity available, previously, only in larger, multi-user devices. The first MIDAS workstation was operational in 1983 (ref. 23). Within a year, seven MIDAS systems were assembled and distributed to CCRSP participants.

3.3.1 Workstation Configuration

The MIDAS system configuration is shown in figure 2. Four workstations were assembled at ARC. The ARC systems contained a MC68000 CPU board, a 1024 x 1024 x 1 graphics board, 512K error-correcting multibus RAM, a disk controller board, an ETHERNET controller board, and a 1024 x 800 black and green monitor. Each system was equipped with an 80 MB Winchester-type disk drive except for one workstation which has a 160 MB disk. These components allowed the workstation to function as a microcomputer with a large amount of data storage, as required for processing geographical information. Two of the four ARC systems contained components for the interactive display of Landsat imagery, i.e., a color frame-buffer interface board linked to a 512 x 512 x 24 color frame buffer with pan and zoom, color lookup tables, two graphics overlay planes, high-speed hardware vector generator, a pixel arithmetic unit, a hardware character generator, an 11" x 11" graphics tablet and a 19" high-resolution red/green/blue color monitor (ref. 23).

Three other MIDAS workstations were assembled by A. Travlos (UCB). One each was installed at UCB, CDWR and the Survey Research Branch of NASS in Washington. All three were equipped with a display device, as described above, and a 1600 bpi tape drive. The workstation at UCB has a 160 MB disk.

The seven MIDAS workstations were in place by the end of 1984.

3.3.2 Workstation Communications

Communication among the MIDAS workstations is accomplished in two ways. The MIDAS workstations at ARC are linked by ETHERNET, a high-speed, direct cable linkage. One of the ARC workstations, designated "FOO," has access to a modem for communication with off-site systems. All off-site MIDAS workstations have a similar capability. The workstations at CDWR, NASS, UCB and ARC (FOO) "talk" to each other over public phone lines using either the UUCP utility in XENIX, for electronic mail, or Kermit, a public domain software developed at Columbia University, for file transfer and communication, to conduct the communication.

Prior to, and during, the 1985 inventory, the MIDAS stations needed access to BBN. Access was required for file transfer and data processing. The electronic linkages comprising the CCRSP network are illustrated in figure 1. Kermit was used for most communications among MIDAS stations at different CCRSP sites. Arpanet, a system maintained by the Defense Advanced Research Projects Agency (DARPA) for communications among government and university research centers was used for most of the communications between MIDAS and BBN (ref. 24). Arpanet supported communications between the VAX network at Ames, which includes the VAX in ECOSAT, or the VAX network at UCB, and the BBN system in Boston. A MIDAS station at ARC or UCB could communicate with BBN by connecting to a VAX using Kermit and then linking the VAX to BBN using Arpanet. Some backup methods of communications, involving Telenet and

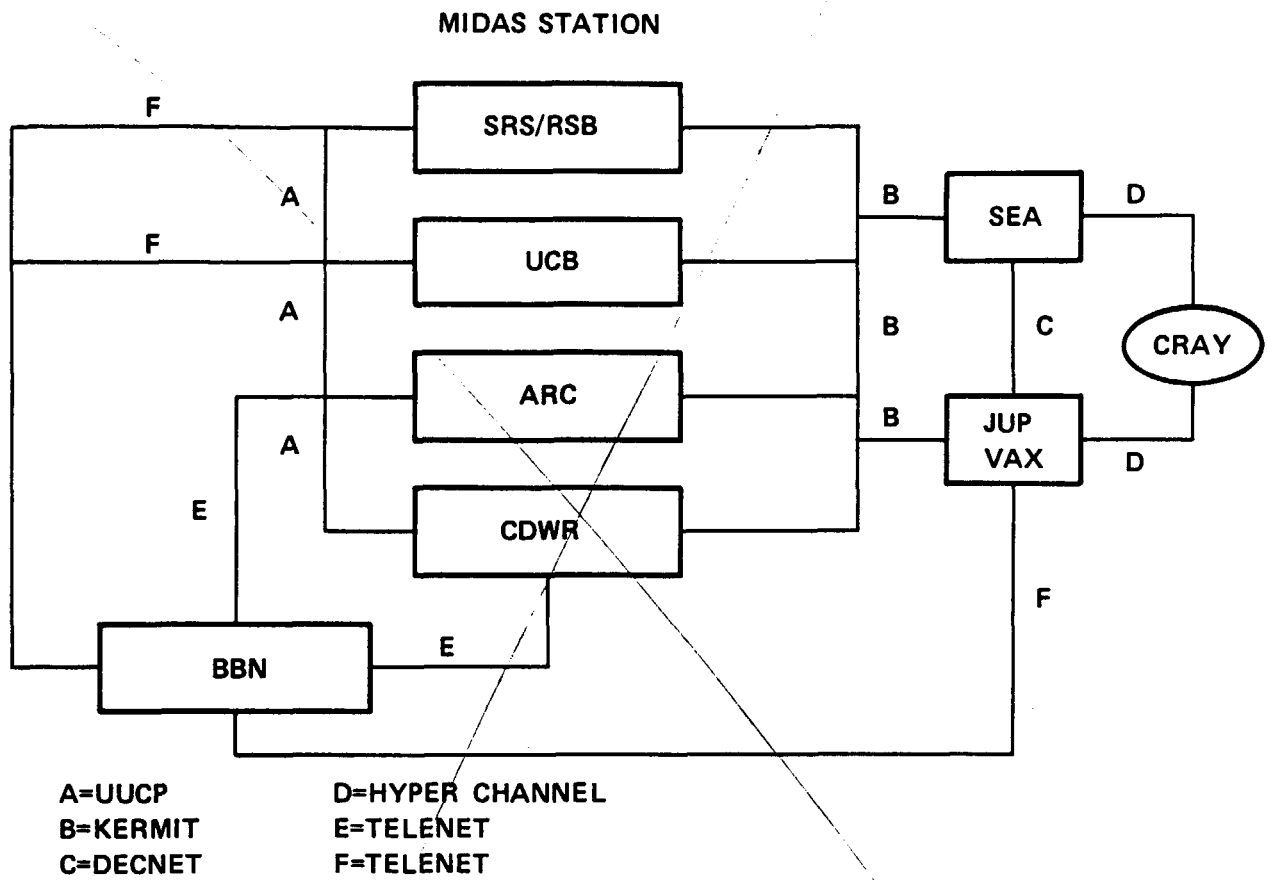


Figure 2: MIDAS System Configuration

public telephone lines, were included in the system of linkages illustrated in figure 1 to provide backup access to BBN.

3.3.3 Workstation Software

MIDAS has a XENIX operating system and is equipped with three software packages for digital data manipulation: Classified Image Editor (CIE), Earth Resources Laboratory Applications Software (ELAS), and PEDITOR.

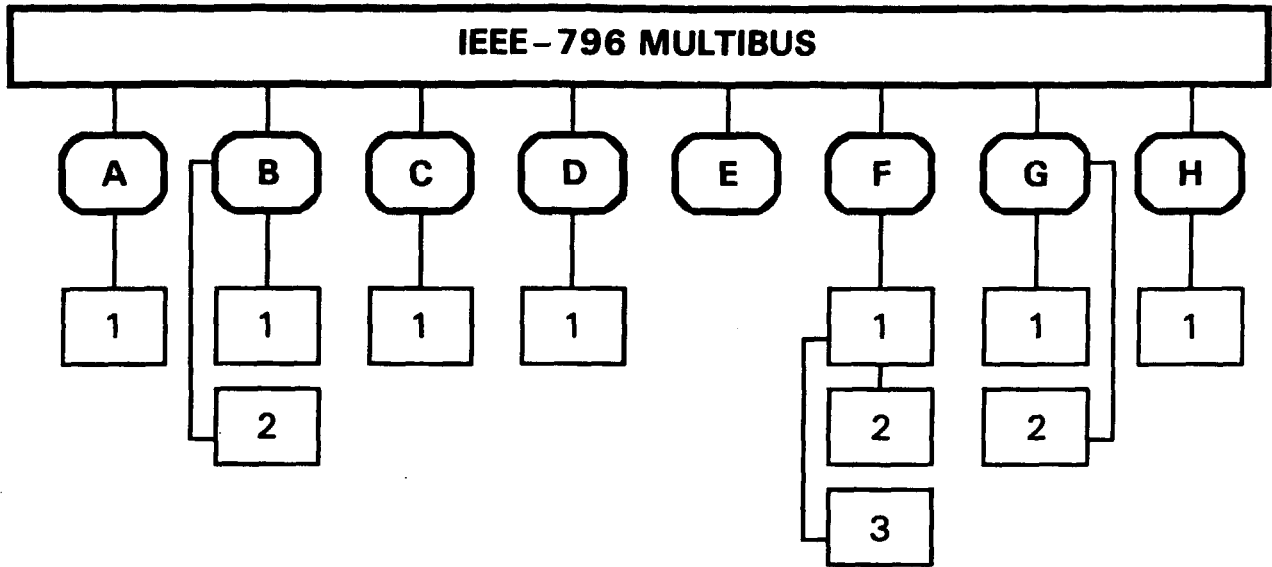
3.3.3.1 CIE and ELAS- CIE was written at ARC by Walt Donovan. It is a special purpose package designed for the display and editing of single band images, especially classified images. The classified image appears as a map on a color graphics terminal. The image may be displayed in shades of grey or in color. Color assignments are made by associating a color name with a class number or numbers, or a range of grey levels. Usually all clusters corresponding to a crop type or land use are displayed in one color. A color key is displayed along side of the map and is updated as color assignments are made. CIE was used by CCRSP to edit classifications before hard copies of the data were generated.

ELAS is a general purpose image-processing system developed at the National Space Technology Laboratory. When MIDAS was brought on line, ELAS was implemented on the new workstations by William Erickson, which was the first implementation of ELAS on a UNIX-like operating system. ELAS includes modules for simultaneous display of up to three bands of imagery. ELAS was used by CCRSP to display Landsat MSS bands 4, 5, and 7 so that the imagery would look similar to a high altitude, color infrared photograph.

3.3.3.2 PEDITOR- The rationale behind the development of PEDITOR is described in section 2.2.3. The conversion of EDITOR code to PEDITOR began in 1983 and was completed in the fall of 1985. Most of the EDITOR code operational on the BBN system was rewritten in Pascal. The format of the new code was chosen to make the code as transportable as possible.

Appendix A lists the PEDITOR modules and includes a brief description of each module's function. Approximately 80% of PEDITOR code was written, i.e. converted from the EDITOR system, at ARC. Some modules and libraries were written at UCB and some, such as the modules to "pack" data and perform the estimation calculations, were written at NASS. The code was tested by NASS and ARC prior to, and during, the inventory. The tests performed are described in section 3.3.5.

The MIDAS stations at ARC designated "FOO" was the depository for the official version of PEDITOR. As modules, libraries, and standard reference files were completed or updated at UCB, NASS, or ARC they were transferred to FOO. The ease of communications among the workstations made it possible to distribute PEDITOR code electronically. In 1984, UCB assumed the responsibility to distribute PEDITOR updates to all workstations and BBN. Upgrades or reloads involving more than one or two modules or other files were sometimes accomplished by writing the files containing the code to magnetic tape and reading the tape at the remote sites.



A-ETHERNET CONTROLLER
 B-SERIAL I/O INTERFACE
 C-TAPE DRIVE CONTROL
 D-2181 DISK CONTROL
 E-512K RAM
 F-COLOR GRAPHICS INTF

 G-MC68000, 256K RAM
 SERIAL I/O
 H-1024 x 800 GRAPHICS
 CONTROL

A1-ETHERNET TRANSCEIVER
 B1-LINE PRINTER
 C1-TAPE DRIVE
 D1-80 MB DISK

 F1-COLOR GRAPHICS CONTROL
 F3-GRAPHICS TABLET
 G1-KEYBOARD

 H1-GRAPHICS DISPLAY

B2-MODEM

 F2-COLOR MONITOR

 G2-OPERATOR
 CONSOLE

Figure 2: MIDAS System Configuration

3.4 Plan for the 1985 Inventory

A list of recommendations for the 1985 inventory was compiled by UCB, based on the findings of the CCRSP research. These were reviewed at one of the regular CCRSP meetings and subsequently presented to management of the CCRSP organizations and CCLRS at the semi-annual review of CCRSP in Berkeley in September 1984. The list is reproduced in Appendix B. A preliminary list of crops to be reported on and a prioritization of study sites were made based on the recommendations and the interests of NASS, CCLRS and CDWR. The UCB recommendations and preliminary decisions made at the September meeting were then reviewed in Washington by NASS. Most of the technical recommendations made by UCB for the inventory were approved, and NASS wrote an implementation plan for the inventory. The plan included a revised list of crops, choice of study site, technical methodology, pre-inventory preparations, and a work schedule.

The primary goal of the inventory was an operational test of the use of Landsat data to develop estimates and to map major crops in California. The study site was the Central Valley, specifically, 19 counties within the Central Valley. Acreages estimates were to be reported for 10 major crops in Central California: alfalfa, almonds, corn, cotton, grain (wheat and barley), grapes, rice, deciduous tree-fruit(citrus, olives, kiwi, etc.), tomatoes, and walnuts. These acreages were to be reported at the regional level by January 1986, and at the county level by March 1986. The schedule was designed to test the feasibility of obtaining Landsat-based estimates in a timely manner, i.e., in time to have an impact on the annual acreage estimates issued by CCLRS. Map products showing the distribution of the crops and major land-use types would be produced from the classification of the Landsat imagery and evaluated in terms of accuracy and utility in support of CDWR land-use inventories.

The secondary objective of the inventory was a test of MIDAS. The procedures for the inventory were a modified version of standard EDITOR processing. A significant difference was that most of the processing be done on MIDAS with PEDITOR. All CCRSP participants were equipped with MIDAS stations by 1984. CDWR and NASS, especially CCLRS, appeared interested in developing the operational potential of the workstation. In response to the presence of MIDAS and the then imminent completion of PEDITOR, MIDAS was selected as the system of choice for the inventory. The decision to use MIDAS was made with the understanding that BBN would be available to assume the data processing burden should MIDAS prove inadequate for the job.

3.4.1 Technical Approach

The data processing steps involved in the Central Valley inventory are summarized in figure 3. The inventory design differed from typical NASS processing in five ways:

1. Use of three Landsat observations over the study site, rather than one or two,
2. Use of Landsat bands 5 and 7 only from each acquisition,

LANDSAT DATA PREPARATION

LANDSAT DATA ACQUISITION
↓
DATA REFORMATTING
↓
SCENE-TO-SCENE REGISTRATION
↓
SIX-BAND TAPE COMPILATION

GROUND DATA PREPARATION

GROUND DATA COLLECTION
JES DATA-TRANSECT DATA
↓
GROUND DATA FILES
↓
SEGMENT DIGITIZATION
↓
SEGMENT REGISTRATION
↓
SEGMENT MASK GENERATION

CLUSTERING

DATA PACKING

CLUSTERING

STATISTICS FILE EDITING

CLASSIFICATION AND AGGREGATION

STRATA MASK GENERATION

STRATA MASK EDITING

SAMPLE CLASSIFICATION

FULL FRAME CLASSIFICATION

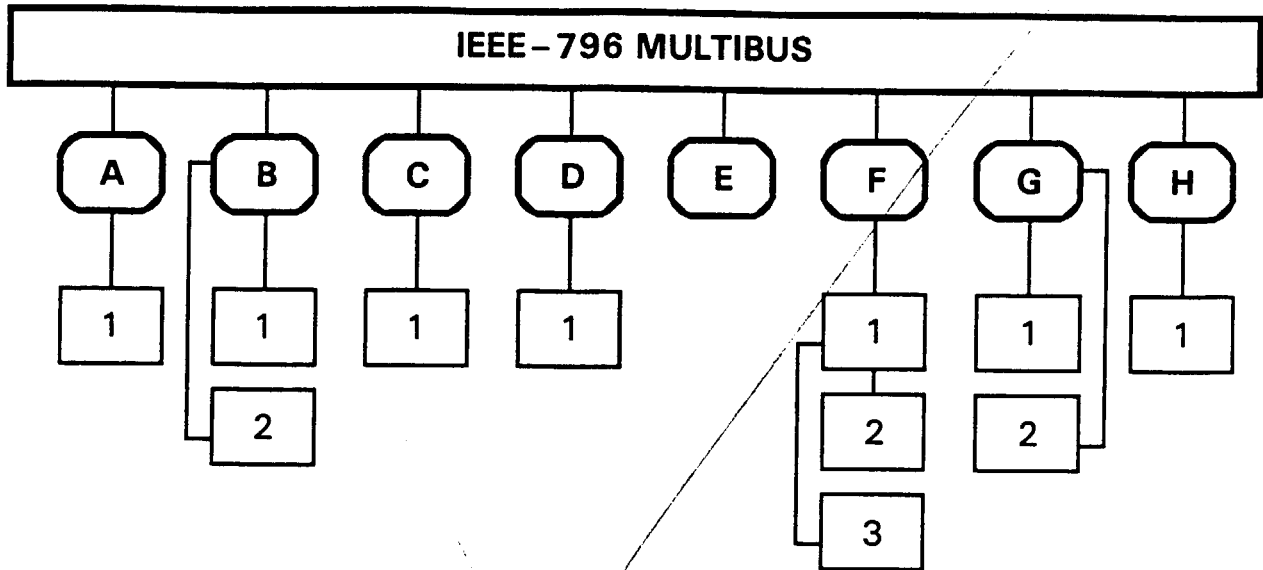
AGGREGATION

ESTIMATION

REGRESSION WITH JES SAMPLE

REGIONAL AND COUNTY ESTIMATION

Figure 3: Central Valley Inventory - Data Processing Steps



A-ETHERNET CONTROLLER
 B-SERIAL I/O INTERFACE
 C-TAPE DRIVE CONTROL
 D-2181 DISK CONTROL
 E-512K RAM
 F-COLOR GRAPHICS INTF
 G-MC68000, 256K RAM
 SERIAL I/O
 H-1024 x 800 GRAPHICS
 CONTROL

A1-ETHERNET TRANSCEIVER
 B1-LINE PRINTER
 C1-TAPE DRIVE
 D1-80 MB DISK
 F1-COLOR GRAPHICS CONTROL
 F3-GRAPHICS TABLET
 G1-KEYBOARD
 H1-GRAPHICS DISPLAY

B2-MODEM
 F2-COLOR MONITOR
 G2-OPERATOR
 CONSOLE

Figure 3: Central Valley Inventory - Data Processing Steps

3. Transect ground data collection - typical processing used JES data only,
4. Map product generation - map product capabilities were included at the request of CDWR, including detailed accuracy assessment on of the JES segments, and
5. Evaluation of the accuracy of JES survey data.

Three dates of Landsat data were to be used because the results of preparatory research by CCRSP (section 3.2.2) indicated that three dates would yield the best results for an inventory of California agriculture. One spring and two summer Landsat observations would be acquired for each of the seven frames required to cover the study site. The analysis would be done using MSS bands 5 and 7 from each date.

Two issues related to sample allocation were addressed in the design of the inventory. There were indications from research by NASS that bias was introduced into the estimates when the same ground sample segments were used to train the classifier and develop the estimates (section 3.2.2). There was concern by UCB that the number of ground sample segments in the JES might not be adequate to sample the spectral variation in California agriculture and was, therefore, inadequate to train the classifier. Both issues were resolved by drawing upon the resources of CDWR to conduct an additional survey.

In the 1983-84 growing season, CDWR conducted a systematic sample of a section of the Sacramento Valley to determine if such a technique would sample adequately the variation of crop types present. The survey was conducted as an independent test, and the data collected was not used in any other task. Details of the sampling scheme are described in section 4.2.2. The survey results indicated that the systematic sample provided sufficient data on the crop types of interest to train the computer to recognize them in the Landsat imagery. It was decided by CCRSP to use the CDWR survey technique to gather data for training the classifier (cluster statistics file) and reserve the JES data for acreage estimation and accuracy assessment.

Map products were to be generated from classified Landsat imagery because CDWR typically delineates survey information on maps. The accuracy of Landsat classification would be evaluated within that agency by comparison of the Landsat crop map with other crop maps.

3.4.2 Roles of the Participating Organizations

The inventory plans called for most tasks to be completed by NASS and CDWR. NASS, through CCLRS, was responsible for the JES data set. Preparation of the data set included collection, tabulation and digitization of the 600 JES ground sample segments in the study site. George May was transferred by NASS to Sacramento to take charge of the NASS/CCLRS tasks in CCRSP. May also chaired the meetings of the CCRSP working group and was the unofficial manager of the inventory project until his resignation from USDA in December, 1985.

The primary data processing role in the inventory was taken on by CDWR. Most of the data processing was intended for the MIDAS workstation at CDWR. That site was selected because CDWR and CCLRS wanted an operational demonstration of its capability, and CDWR was committed to continuing MIDAS operations after the completion of the inventory. In addition to the data processing load assigned to CDWR, the agency also took on the planning, coordination, collection, tabulation, and digitization of the transect data--the training data for the Landsat classification. The ground data collected for the inventory by CDWR was an extension of work done by the agency for the preliminary studies reported in section 3.3.2.

The ground data collection effort for the transect data, software support, data processing assistance, and technical advice was provided by UCB. Upon completion of the inventory, UCB was to review the inventory procedures and assist with the assessment of the quality of the crop maps and acreage estimates generated.

Ames Research Center was required to continue support activities, as described in section 3.1, in particular, assistance with software, provision of U-2 photography, and execution of processing steps on the Cray X-MP. In addition, ARC was responsible for reformatting Landsat data tapes for processing within the EDITOR/PEDITOR system and the development of acreage estimates following tabulation of the ground survey and Landsat data.

3.4.3 Preparations

The inventory implementation plan included schedules and assignments for tasks that needed to be addressed prior to the actual inventory. These included completion and testing of PEDITOR, completion of non-PEDITOR software related to the inventory such as software to generate six-band Landsat data files (MSS5 and MSS7 from each of three dates), acquisition of current year photography of the JES segments, and preparation for transect data collection.

3.4.3.1 PEDITOR testing- PEDITOR was tested by NASS and ARC. Testing by NASS was performed at BBN with a four-channel Landsat data set. By the summer of 1985, NASS confirmed, to the agency's satisfaction, the operation of all modules completed at that time and declared PEDITOR operational. Additional tests were performed by CCRSP to confirm proper function with a Landsat data set containing more than four bands and to confirm the proper function of PEDITOR on MIDAS.

Tests of PEDITOR for CCRSP were begun by ARC in the spring of 1985. The test data set was the 1982 Yolo County data set that had been used for the multi-crop research described in section 3.2. The data set included an eight-channel tape of Landsat data (MSS bands 5 and 7 from four dates) and approximately 60 JES segments. A copy of the data set was available at BBN.

The tests were performed in two stages. In the first stage, identical data was processed through a PEDITOR module at BBN and on a MIDAS workstation at ARC. If no run-time errors were encountered on either system, and the results were identical, the proper operation of the module on MIDAS was confirmed. An error in the logic of a module would produce erroneous results even though the module appeared to operate

correctly. Second stage testing was done to verify the computations. Verification was achieved by comparing the results obtained with PEDITOR on MIDAS with results obtained using functionally similar ELAS modules or other software, and by cross-checking the output from one PEDITOR modules.

The manner of testing described was used to verify the correct operation of the PEDITOR modules and MIDAS for the inventory data processing steps up to and including classification. Testing of the estimation modules was delayed until the fall of 1985 when they were completed by NASS. By that time, the inventory was well underway, and the inventory data set served as the test data set.

3.4.3.2 Code written for CCRSP- A limited amount of new computer code was developed at ARC for the 1985 inventory experiment. The program COMPILE was the most significant piece of code added. It was written to generate a 6-band tape of MSS bands 5 and 7 from multiple registered Landsat acquisitions (section 4.1.4).

Modifications to some PEDITOR code was also required for the inventory. The most common change was to enlarge an array because of the large number of segments used in the survey for the training data and the large number of spectral classes resulting from clustering three dates of Landsat imagery.

The PEDITOR modules involved in classification were installed on the ECOSAT VAX at ARC to take advantage of the greater speed of the VAX, compared to the MIDAS workstation, and the facilities available to submit files directly to the front-end computer for the Cray.

3.4.3.3 Current year photography- An ARC U-2 flight in October, 1984 acquired color infrared photography of the entire Central Valley. The photography was used in both the JES and the transect survey. The best time to acquire photography for a current year survey of the Central Valley is in the early spring of the survey year. By that time, virtually all field boundaries for spring and summer plantings have been defined. However, because of the large number of prints needed and the need to have the prints marked with segment boundaries prior to the June survey, the October 1984 flight data was used. It was assumed that minimal field boundary changes would occur after the October date.

The U-2 photography was acquired as 9"x9" color transparencies. Samples from the transparencies were enlarged, converted to color prints and submitted to CDWR and CCRLS in Sacramento for comment. The scale and resolution of the product were acceptable to both agencies, but CCLRS preferred the prints in black and white so that tract and field boundaries and other enumerator marks added to the the photography in color would not cause confusion.

Enlargements were made at ARC for approximately 300 segments. The prints were delivered to CCLRS in early spring of 1985. They were annotated with segment boundaries in Washington D.C. and were used by the enumerators in the JES.

4. THE 1985 INVENTORY

The following account of the 1985 presents a general description of events and details of the steps in which ARC was a substantial contributor (Table 2). The account begins with a description of the study site because geography effected much of the work and the results obtained. Data collection, data processing, acreage estimation and map products are discussed in turn. The account concludes with a summary of system performance.

This section of the report is intended to be a comprehensive guide to the data processing for the 1985 inventory. The processing was documented in the CCRSP PEDITOR Procedural Manual written by CCRSP participants and available through UCB.

4.1 The Test Site

The test site and location of the Landsat frames which cover it are shown in figure 4. The Central Valley is the heartland of California agriculture. It is an elongated basin that stretches from the foothills of the Tehachapi Mountains south of Bakersfield north north-west approximately 400 miles to the southern extension of the Cascade Mountains north of Redding. The width of the Valley varies between 40 and 80 miles. It is bounded on the east by the Sierra Nevada Mountains and on the west by the North Coast and Central Coast Ranges. The flatness of the Valley floor is broken only by the Sutter Buttes north of Sacramento and the low-lying Dunnigan and Montezuma Hills on the western edge of the Sacramento Valley. Water is the key to the formation of the Central Valley and to its current economic health. The floor of the Central Valley is underlain by sediments deposited in the basin by the drainage of the San Joaquin River and its tributaries in the south and the Sacramento River and its tributaries in the north. Because of the summer drought in the Valley, typical of a Mediterranean climate, most agriculture is irrigated.

The Central Valley was selected as the test site because a large proportion of the major crops grown in California are farmed there. Over 90% of the corn, cotton, grain sorghum, and nut crops, and 100% of the rice harvested in California come from the Central Valley. The Central Valley was an appropriate size for the test. It is small enough for implementation of a transect survey, and large enough to represent a good operational test of an inventory design involving Landsat.

The diversity of crops grown in the test site affected the inventory in two ways. First, a larger than normal sample size was required to garner adequate training for the classifier. Second, a given crop tended to occur in a minority of segments with many crops concentrated in a subregion of a land-use stratum within a Landsat analysis district resulting in imprecision in expansion estimates based on JES segments. Regression on Landsat pixels to correct the estimate of the mean acreage per square mile could potentially lead to great improvement in the accuracy of acreage estimates provided there was a high correlation between pixels and acreage and sufficient data in the JES survey to develop a regression line.

TABLE 2.- 1985 INVENTORY--DATA PROCESSING AT ARC

Processing stage	Function	Job setup	Job execution
Landsat data preparation	Reformat tapes	M/P	C/S
	Scene registration	M/P	C/M
	Confirm registration	M/E	M/E
	Map calibration	M/P	M/P
	Digitized segments	M/P	M/P
	Register segments	M/E	M/P
Clustering	Reformat training data	M/P	M/P
	Cluster	M/P	C/Ed
	Stat file edit	M/P/E	M/P/E
Classification/ aggregation	Classify segments	M/P	C/Ed
	Full frame classify	M/P	C/Ed
	Aggregate	M/P	C/M/P
Estimation	Regression	M/P	M/P
	Large scale	M/P	M/P
Map product generation	Recode data	M/E	M/E
	Generate maps	M/E	M/E

Key: M = MIDAS
Ed = EDITOR

C = CRAY
S = stand alone programs

P = PEDITOR
E = ELAS



Figure 4: 1985 Central Valley Inventory - Study Site and Landsat Frame Locations

The large counties in the Central Valley were both an advantage and a drawback for crop surveys using Landsat data. Large counties hold more JES segments thus improving the estimate of the "county effect" in the relationship between acreage and pixel counts and increasing the potential for good estimates of acreage at the county level. The complexity of data processing was directly proportional to county size; the larger the county, the greater the difficulty getting the county digitized properly and the more likely the county will cross a Landsat frame boundary. Three major agricultural counties in the Central Valley, Fresno, San Joaquin, and Sacramento, fell across Landsat frame boundaries in the inventory.

4.2 Ground Data Acquisition

Two ground data surveys were conducted as part of the inventory. The June Enumerative Survey collected data for area estimation and accuracy assesment. The Transect Survey collected data for classifier training. The data from both surveys was encoded in computer files for processing with EDITOR/PEDITOR software.

4.2.1 June Enumerative Survey

The JES data was collected in early June as part of the standard survey of California crops. The sample segments were selected by NASS using stratified random sampling. A standard set of land-use strata have been defined by NASS for the U.S. This stratification was used in California amended with a tree fruit/grapes stratum (ref. 25). The strata definitions for California are given in Table 3.

The standard JES survey procedure was changed for the inventory by the acquisition of previous year photography. NASS supplies its enumerators with panchromatic, medium scale aerial photography for each ground sample segment to be surveyed. The photograph of a segment is annotated with the outline of the segment. Field boundaries present at the time the photograph was taken are usually apparent. During the survey, the enumerator draws in the tract and field boundary lines on the photograph using the boundaries in the photograph as a guide for locating where the lines should be correctly drawn. The photography is usually updated about every 7 yr but may be older. Field boundaries can change significantly in 7 yr, and it is often difficult for enumerators to draw boundaries accurately on old photography. By supplying enumerators with recent photography, it was hoped that errors in field size and boundary location would be reduced.

There were also a few minor differences in definitions of crop/land-use categories in response to requests from CCRSP. A few new crop types were defined, for example over-wintered sugar beets were differentiated from sugar beets planted in the current year to provide information to be used in accuaracy assesment.

4.2.2 Transect Survey

Perhaps the most significant divergence in standard EDITOR processing during the 1985 Central Valley inventory was the use of an independent data set to train the classifier. The JES data, normally used for training and testing, was reserved

TABLE 3.- CALIFORNIA AREA FRAME STRATA DEFINITIONS^a

Stratum 13 ^b	Fifty percent or more cultivated, mostly general crops with less than 10 percent fruit or vegetables.
Stratum 17 ^b	Fifty percent or more cultivated, mostly fruit, tree nuts, or grapes mixed with general crops.
Stratum 19 ^b	Fifty percent or more cultivated, mostly vegetables mixed with general crops.
Stratum 20 ^b	Fifteen to fifty percent cultivated, extensive cropland and hay.
Stratum 31	Agri-urban, more than 20 dwellings per square mile, residential mixed with agriculture.
Stratum 32	City, more than 20 dwellings per square mile, heavily residential/commercial, virtually no agriculture.
Stratum 41	Privately owned range, less than 15% cultivated.
Stratum 43	Desert range, barren areas with less than 15% cultivated, virtually no crops or livestock.
Stratum 44	Public grazing lands, Bureau of Land Management or Forest Service grazing allotments.
Stratum 45	Public land not in grazing.
Stratum 50	Nonagricultural, includes state and national forests, wildlife refuges, military reservations, and similarly designated land.
Stratum 62	Known water (not sampled), larger than 1 sq. mile.

^aFrom M. L. Holko, "1982 Results of the California Cooperative Remote Sensing Project," SRS staff report No. AGES840305, March 1984.

^bMajor crop strata--other strata not used for crop estimation in this study.

for estimation and accuracy assessment. A separate set of sample ground data was collected to train the classifier. The new ground data was collected by driving transects through the agricultural areas of interest. The data set is referred to as the Transect Survey.

The Transect Survey was operated by CDWR, NASS and UCB personnel. Approximately 2500 segments were selected and visited in mid-spring and mid-summer. Segment selection was designed to achieve a comprehensive representation of crops and crop conditions affecting crop appearance, as well as representation of types of land use present in the agricultural land use strata in the California. The ground sample units were picked through a systematic sample as follows:

1. CDWR land use maps were used to quantify crop mix on a 2.5 min block basis.
2. Areas characterized by homogenous soil color and other factors affecting the appearance of agricultural fields on imagery were identified on 1984 Landsat photoproducts.
3. The above information, supplemented with 1985 U-2 photography, was used to locate transects that would maximize contact with crops of interest and sample areas with different appearance factors.
4. The transects were drawn on road maps and assignments for the gathering of field observations were made by county.
5. Along the transects, stops were made every 2 miles and all fields adjacent to the stop and above a minimum size were included in the survey.

Field enumerators drew field boundaries on a map and recorded field contents, i.e. crop type or land use. High altitude photography acquired by ARC in 1985 was used to check the data before it was encoded.

4.2.3 Preparation of the Survey Data

The tabulated data collected by JES enumerators required key punching and entry on magnetic tape. The task was completed by NASS. The tapes were returned to CDWR. The ground data were loaded on MIDAS and distributed to other CCRSP sites. The field observations collected by CDWR and UCB staff for the transect survey were entered into computer files at CDWR interactively.

The strata network files delineating the California JES sampling frame for 16 of the 19 counties in the Central Valley test site were supplied by NASS. Strata maps were sent to ARC for Kern, Kings and Tulare Counties. The maps were digitized on a tablet at ARC using PEDITOR software at BBN. All 19 strata network files were converted to mask files using PEDITOR on a MIDAS workstation at ARC.

Field boundaries were digitized into segment files at CDWR and CCLRS. The JES segments were digitized at CCLRS. The task was made easier by the need to digitize interior field boundaries for only the 10% sample of the JES segments that were to be used for accuracy assessment. The transect segments were digitized at CDWR using

the Osborne I. As the ground data was digitized and segment network files created, the files were transferred to MIDAS, and copies were sent to BBN.

Segment registration is a two-step process. The first step is the generation of a calibration file for each Landsat scene relating latitude and longitude to Landsat SOM coordinates. The calibration file is created on a digitizing tablet by locating control points on 1:250,000 maps and on 1:1,000,000 Landsat photoproducts obtained from EROS Data Center. About 20-30 control points are needed. After the points are located, least squares analysis is performed on the points, and a calibration file is generated. The calibration files for CCRSP were all created at CDWR.

The latitude/longitude coordinates in the segment network file are inexact due to small errors in calibration when the segments are digitized. The next step in registration corrects for the small error introduced. The calibration file is used to identify a block of data in the Landsat scene containing the segment. A grey-scale plot of one band of Landsat data from the block containing the segment is generated. A plot that shows the field boundaries of the segment, drawn at the same scale as the grey-scale plot, is also produced. The second plot, a vector plot, contains tic marks that allow it to be placed correctly on the grey scale plot. The vector plot is overlaid on the grey-scale plot and shifted until the field boundaries in the segment plot appear to lie in the proper location in the Landsat data gray scale. The shift required for each segment is the number of pixels the segment plot was moved, in the x and y directions, from the location predicted by the calibration file to its proper location. The x and y shifts are recorded and entered into a text file, with segment number, to be used for mask generation.

The intention of CCRSP was to generate all plots on MIDAS, however, the vector plots were generated by NASS in Washington because of the unexpected length of time required to create the plots on MIDAS. Most of the segment shifting was done at CDWR with help from UCB and ARC. Segment registration proceeded smoothly even though, for 90% of the JES segments, internal fields boundaries were not available to match the vector and grey scale plots. Upon completion of segment shifting, mask files were created using the appropriate calibration file, segment network file and file of segment shifts. All segment mask generation for the inventory was done at CDWR.

4.3 Landsat Data Preparation

Acquisition and preparation of the Landsat data started at about the same time in the spring as ground data collection but wasn't completed until the end of October 1985. Because of the quantity of data that had to be manipulated, and the need to store the data on tape, all Landsat data preparation was done at ARC.

4.3.1 Landsat Data Acquisition

The Landsat data was ordered by NASS. The EROS catalogue containing the scene information was checked regularly as the time window for acquisitions was entered.

The window extended from March through September. The intent was to obtain a spring acquisition, an early summer acquisition, and a late summer acquisition for each of the seven frames required for the inventory. Within that window, acquisitions were selected to maximize data quality and minimize percent cloud cover. Data was ordered as computer compatible tape and 1:1,000,000 black and white prints. The tapes were the primary data source. The prints were needed for the preliminary work in scene-to-scene registration and for general reference.

The frames came from three Landsat paths. The acquisitions selected are listed in Table 4. The worst cloud problem was in frame path 44, row 33 where part of Sacramento County between the cities of Sacramento and Lodi was obscured by cloud or cloud shadow on the 5 May 85 acquisition. Some patchy clouds were also unavoidable on path 42.

The data tapes were sent to CCLRS first and forwarded to ARC for processing. The first two acquisitions for each of the seven frames was received at ARC by early August. The last acquisition was received by early October.

4.3.2 Reformatting of Landsat Scenes

The Landsat digital data arrived from CCLRS in EROS band interleaved (BIL) format on 6250 bpi tapes. The data left ARC reformatted into six-band data sets in the two formats described in section 2.2.2.1 and suitable for processing at BBN or on a MIDAS workstation.

The first step in preparation of the data was to reformat the EROS raw data tapes so that the data would be compatible with the registration software on the Cray. The second step was to register the data. Before the scenes could be registered, a base date had to be selected.

The second acquisition in each frame was chosen as the base (primary) date, i.e., the Landsat SOM coordinates on the second date were chosen as the coordinates for the six-band data sets.

The reformatted data was registered using the block correlation technique (ref. 7). The initial overlay was performed at CDWR. ARC took advantage of the automated block editing in the latest version of BCORR to eliminate manual editing of the correlation blocks. Sample sections were extracted and displayed from each pair of registered scenes to verify that the registration was correct.

The third acquisition of all frames was received at ARC in late September. By early October, the scenes were registered to the primary dates, and all frames were ready for the final processing step, creation of the six-band data set.

The program COMPILER created the new tapes. The frames were split into an eastern and western half so that no tape file contained more than 1950 columns. There was a 100-column overlap between the halves. Splitting the data was necessary because BBN could not read records longer than 20,000 bytes, and MIDAS could not read a tape file that extended beyond a single reel of tape. The COMPILER program

TABLE 4.- LANDSAT ACQUISITIONS FOR 1985
CENTRAL VALLEY INVENTORY

Path/row (frame)	Acquisition date	Scene ID
42/35	20 MAR 85	50384-18043
	2 JUL 85	41082-17582
	12 SEP 85	50560-18032
42/36	20 MAR 85	50384-18045
	2 JUL 85	41082-17585
	12 SEP 85	50560-18034
43/34	14 MAY 85	50439-18100
	17 JUL 85	50503-18094
	18 AUG 85	50535-18093
43/35	14 MAY 85	50439-18102
	17 JUL 85	50503-18101
	18 AUG 85	50535-18095
44/32	5 MAY 85	50430-18152
	8 JUL 85	50494-18151
	25 AUG 85	50542-18144
44/33	5 MAY 85	50430-18154
	8 JUL 85	50494-18153
	25 AUG 85	50542-18151
44/34	5 MAY 85	50430-18161
	8 JUL 85	50494-18160
	25 AUG 85	50542-18153

read the three registered scenes for a frame, extracted MSS bands 5 and 7 from each half scene, and wrote a new six-band tape in BBN format at 1600 bpi. BBN format (pixel interleaved, 664 byte header) was chosen so that the data set would be compatible with BBN and MIDAS.

4.3.3 Formatting Landsat Coverage of Segment Data

The Landsat coverage of the JES and Transect Survey segments was processed to form compact data files by packing portions of data from the half scene tapes. In the standard EDITOR procedure (section 2.2.2.1), one packed file is created for each crop/land-use type within a Landsat analysis district (Landsat path.) The inventory, however, generated multiple packed files for each crop/land-use type because of the large number of training segments and a 300 segment limit on the number of segments in a packed file. Two packed files were required for each crop/land-type use in two of the analysis districts, and three packed files were required in the third.

CDWR intended to create all the packed files in Sacramento but was precluded from doing so by the slow speed of the MIDAS station and the propensity for the system to crash if more than five crop/land-use types were packed at one time. Because of the interest in completing the estimates on schedule, some of the packed files were created at UCB and ARC.

4.4 Landsat Data Processing and Interpretation

Maximum likelihood classification was performed to label each pixel in the six-band data sets with a crop type or land-use category. The data was clustered first to estimate the distribution parameters for the crop/land-use types. Following classification, pixel counts on JES segments were tabulated, and the tabulations were used to compute crop acreage estimates and to gauge the accuracy of the classification.

4.4.1 Training for Classification

The classifier for each analysis district was trained to recognize crops and land-use types by clustering the brightness data contained in the sample segments of the Transect Survey. Virtually all clustering for the inventory was done with the CLASSY algorithm on the Cray X-MP at ARC. The intent of the inventory was to set up and submit the CLASSY jobs from CDWR. As in packing, however, the press of time forced a change in plans. The files packed at UCB and CDWR were sent to ARC where the clustering jobs were set up and submitted.

Job set up for clustering was accomplished using a PEDITOR module on the SEA VAX. The module, CRAY, automatically formats Cray job control language when the user identifies the type of job to run and other parameters required for the particular job type. The output from the CRAY module is a text file that can be submitted directly to the Cray, via DECNET, from the SEA VAX. Multiple-packed files were submitted in a single job, but each file was clustered individually.

The output from the CLASSY program was a text file containing the statistical information on the data clustered. The output file was returned automatically to the SEA VAX. The CRAY module was used, again, to reformat the data in the output file. After reformatting, every crop/land-use type clustered had a separate statistics file in PEDITOR format. The statistics files were written to tape and mailed back to CDWR or UCB as appropriate, for the next step in processing, editing the statistics files.

4.4.2 Statistics File Editing

The statistics files for each Landsat analysis district were combined and edited. Editing was required because, for all analysis districts, the number of clusters in the combined statistics file exceeded 255, the maximum number of clusters the classifier could process. Editing was also advisable because of the following:

1. Some clusters were associated with a small number of data points. The significance of these clusters and the stability of the statistics, particularly the covariance matrices, were thus in doubt.

2. Improvements in the quality of the classifications might result from correcting for imperfections in the training data. The transect data might have contained a few errors, undetected mistakes in the deliniation of fields or in recording of crop/land-use type. Some fields could be atypical or highly variable in appearance.

3. In a few cases, the training data was drawn from areas near an edge of the Landsat path, with no valid reflectance data on one or more observation dates. These points had values of 255 on some channels.

The strategy for editing clusters was agreed upon after several discussions at CCRSP meetings. Responsibility for this work was split among members. Path 42 was assigned to CDWR and USDA, Path 43 to ARC, and Path 44 to UCB. The task required some discretion, but the following criteria were used by all cluster editors to remove clusters:

1. Clusters with less than 100 points,

2. Clusters with three or more channels with very high variance (17 or more grey levels in the standard deviation in brightness detected by the Landsat scanner),

3. Clusters with a mean of 255 in one or more channels, indicating invalid Landsat data,

4. Clusters that were similar to a large number of other clusters.

The clusters were removed in the order shown above. In the last step, the Swain-Fu distance (refs. 11,26) was calculated for all cluster pairs and was the criterion for removing clusters. The Swain-Fu distance measures spectral

similarity, or degree of overlap, between two clusters using a formula which normalizes the ordinary Euclidean distance between cluster means by a factor related to the cluster shapes and volumes, as indicated by the cluster statistics. Clusters that were less than 0.6 in Swain-Fu distance from more than 20 other clusters were scrutinized carefully. Some clusters were similar to 50 or 60 other clusters. Most of similar cluster pairs were of the same crop/land-use type. Clusters that were similar to clusters associated with dissimilar crop/land-use types, that is, areas in the transect segments that should have looked different on the Landsat data on one or more of the three observation dates, were the first to be eliminated. The final statistics files for paths 42, 43, and 44 contained 223, 169, and 231 clusters respectively.

4.4.3 Classification and Aggregation

The classification job control files were created on the ECOSAT Vax with PEDITOR software, and the Landsat data were classified on the Cray at ARC. Two classifications were performed per analysis district. The same classifier, i.e. the same cluster statistics file, was used for both classifications. A "small-scale" classification was performed first on packed JES segments and "large-scale" classification on the half-scene tapes. "Large-scale" classification and aggregation, tabulation of pixel counts by crop/land-use label and stratum, were performed sequentially as two parts of the same Cray job.

Aggregation requires mask files of the strata. The program will abort if the strata mask window extends beyond any edge of the classification. Because most strata mask files extend across Landsat image boundaries, the mask files must be edited, i.e., split along frame boundaries before use in aggregation. That type of mask splitting is a standard EDITOR processing step, and the strata mask files for the 1985 inventory were split in that manner. However, additional editing of the strata mask files was required for the 1985 inventory, because several of the frames were split into two halves as described in section 4.3.2.

Strata mask-file editing for CCRSP was accomplished at ARC the week of 9 December 1985. Personnel from CDWR came to ARC to complete the processing. The masks were edited to exclude the parts of each frame that were outside the three-date overlap zone, and split to accommodate masks that crossed analysis district boundaries. Some of the processing was done on MIDAS. Errors in the software, and the slowness of the system compelled the analysts to perform some operations at BBN.

Following completion of the strata mask editing, large-scale classification and aggregation proceeded. Each half frame was classified with the appropriate statistics file and the labelled pixels were aggregated with the strata mask files. The data processing was done on the Cray with job set-up on the SEA VAX. The outputs from the job were a tape of the half-frame classification and a text file of the pixel counts by crop/land-use type and stratum aggregations. The tape was stored for later copying and distribution to CCRSP participants. The text file was reformatted using the CRAY module in PEDITOR. The text file was split, by county, into individual aggregation files and each file was written to disk in PEDITOR format. Each aggregation file contained a tabulation, by class and stratum, of the pixels

) classified within the area delineated by the mask, a county or portion of a county within a half scene. The aggregation files were used in estimation.

4.5 Estimation

Several techniques were used to include Landsat pixels in the computation of acreage estimates. The primary techniques were those developed in EDITOR and described in section 2.2.2.3. Those techniques were used to generate the estimates reported for deadlines established in inventory plan (section 3.4) and for revisions made at a later date. Additionally, two experimental procedures were tested - ratio estimation (ref. 4, Chapter 6) and robust regression (ref. 18, Chapter 5).

Most of the data processing for estimation was performed on a MIDAS workstation at ARC using PEDITOR software. The county estimates were generated using the EDITOR programs at BBN because they had not been included in PEDITOR, and therefore were not available on MIDAS. Files were transferred to and from BBN using Kermit between SEA VAX and MIDAS stations at ARC, and FTP between the SEA VAX and BBN.

Personnel at ARC relied upon NASS staff, in particular Martin Ozga and Maryanne Cummins, for guidance during the data processing for estimation because of ARC's lack of experience with some of the modules, and the tendency for features and performance of the software to change from year to year. The procedures were documented as they were used in the work reported here.

) Experimental estimates were generated on the ECOSAT VAX. Programs in the Biomedical Data Programs (BMDP Statistical Software, Department of Biomathematics, University of California, Los Angeles) were used to examine the frequency of occurrence of selected crops in the agricultural strata and for computation for ratio estimates. A Fortran program, using a subroutine from the IMSL Library (International Mathematical and Statistical Libraries, Inc., Houston, Texas) was used for development of robust regression estimates with inventory data and for testing the performance of the robust procedure with simulated data, as described in Appendix C.

4.5.1 The Original Landsat Estimates

) Standard EDITOR/PEDITOR procedures were followed to develop estimates described in section 2.2.2.3. Two decisions were made by the analyst. The first decision was selection of land-use strata to be included in the estimates. As indicated in Table 5, the most important land-use strata were 13, 17, and 19. Stratum 20 was included for grain acreage. The second decision was the choice of type of estimator for each stratum (within an analysis district)--regression with Landsat data or proration with JES data only. Proration was used for strata without sufficient JES data for development of a regression line, or in cases where the regression line seemed "unreasonable." An ideal regression line would have a zero Y-axis intercept and a slope of 0.8. As can be seen in Tables 6A-6C, many of the regression lines were very different from the ideal. It was not clear whether the variance in regression line parameters was the result of consistent patterns of omission or commission in the classifications of Landsat or the result of other factors such as

TABLE 5.- 1985 CALIFORNIA JES AREA FRAME POPULATION (N)^a
AND SAMPLE (m)^b SIZES

Stratum	AD42 N/m	AD43 N/m	AD44 N/m	ADDE ^c N/m
13	2095/67	956/25	2336/73	385/23
17	1637/61	1898/80	917/40	0/13
19	49/1	1510/37	606/7	99/13
20	604/15	332/8	607/13	64/5
31	0/1	0/2	0/2	0/2
32	0/0	0/0	0/2	0/1
41	0/3	0/2	0/2	0/14
43	0/0	0/0	0/0	0/0
44	0/0	0/0	0/0	0/5
45	0/0	0/0	45/0	0/1
50	0/0	0/0	0/0	0/1
62	0/0	0/0	0/0	0/0

^aSize in # segments as represented in the frame unit file developed by CCRSP. Strata with little acreage in agriculture were not included.

^bSize in # segments in the JES sample used in development of regression lines, except for analysis district ADDE.

^cAn artificial district. In the frame unit file, N represents size of regions in paths 43 and 44 covered by clouds or smoke on the date of one or more Landsat passes. In the segment catalogue file, m represents segments listed as located in ADDE because of cloud or smoke cover, data processing problems, or outside of Landsat coverage (Kern County).

TABLE 6.- REGRESSION PARAMETERS FOR ACREAGE ESTIMATES

(a) AD42

Crop	Stratum 13			Stratum 17			Stratum 20		
	b ₀	b ₁	r ²	b ₀	b ₁	r ²	b ₀	b ₁	r ²
Alfalfa	-7.4	0.57	0.60	-0.8	0.11	0.23	-14.1	0.76	0.82
Almonds	-16.9	1.63	0.56	-7.1	0.86	0.56	1.1	0.00	0.00
Corn	-3.0	1.32	0.76	2.3	0.17	0.07	-12.2	1.71	0.93
Cotton	42.6	0.78	0.75	-1.8	1.15	0.76	-24.3	1.25	0.93
Grain	-9.8	0.77	0.69	-0.5	0.44	0.39	-43.8	0.94	0.66
Grapes	-1.0	0.11	0.05	-20.4	0.81	0.78	-15.6	0.59	0.50
Tomatoes	+	+	+	0.1	-0.09	0.00	+	+	+
Tree fruit	0.0	2.5	0.14	-6.6	1.09	0.80	-2.2	0.53	0.19
Walnuts	1.5	1.91	0.25	1.7	0.92	0.06	4.3	-1.37	0.02

+0.0 acres in JES survey--no estimation performed

(b) AD43

Crop	Stratum 13			Stratum 17			Stratum 19			Stratum 20		
	b ₀	b ₁	r ²	b ₀	b ₁	r ²	b ₀	b ₁	r ²	b ₀	b ₁	r ²
Alfalfa	46.8	0.67	0.11	-13.3	1.1	0.55	-4.8	0.89	0.62	-38.8	0.94	0.76
Almonds	2.00	0.18	0.06	-14.1	1.13	0.59	2.6	0.00	0.00	2.7	0.96	0.07
Corn	8.9	0.72	0.33	-0.8	0.94	0.62	20.00	0.31	0.45	-30.5	2.51	0.8
Cotton	15.00	0.77	0.84	-8.2	0.76	0.71	-10.6	0.89	0.81	-55.4	1.21	0.84
Grain	-10.3	0.87	0.9	-2.1	0.38	0.37	15.5	0.3	0.47	-8.9	0.91	0.55
Grapes	2.2	0.01	0.00	-24.5	1.05	0.88	-0.2	0.12	0.07			
Rice	-3.3 ^a	0.73 ^a	0.87	-0.5	1.96	0.99	-4.5	0.73	0.87	-101	4.16	0.84
Tomatoes	-12.7	1.51	0.58	0.00	0.00	0.00	11.8	0.07	0.00	+	+	+
Tree fruit	0.00	0.00	0.01	-5.7	0.67	0.34	4.00	0.05	0.00	+	+	+
Walnuts	-0.5	0.85	0.21	-2.2	2.87	0.5	0.2	1.23	0.75	+	+	+

^aStrata 13 and 19 combined because 0.0 acres in JES survey for stratum 13 alone.

+0.0 acres in JES survey

(c) AD44

Crop	Stratum 13			Stratum 17			Stratum 19			Stratum 20		
	b ₀	b ₁	r ²	b ₀	b ₁	r ²	b ₀	b ₁	r ²	b ₀	b ₁	r ²
Alfalfa	-10.00	1.14	0.79	-3.7	0.83	0.61	+	+	+	+	+	+
Almonds	-16.7	1.03	0.46	-8.01	0.99	0.52	+	+	+	-0.1	0	0.20
Corn	-19.3	1.31	0.64	3.9	0.17	0.08	-35.6	2.21	0.83	+	+	+
Grain	-2.7	0.77	0.61	-2.7	0.83	0.92	7.8	0.83	0.91	-1.6	0.45	0.54
Grapes	+	+	+	1.1	0.33	0.05	+	+	+			
Rice	11.9	0.97	0.81	1.8	0.95	0.45	-3.00	0.91	0.99	+	+	+
Tomatoes	0.4	0.2	0.08	-10.3	0.76	0.68	-48.2	1.04	0.89	+	+	+
Tree fruit	-10.5	0.27	0.27	-38.3	1.15	0.55	-5.2	0.15	0.4	-0.4	0.02	0.25
Walnuts	1.2	0.32	0.21	16.00	0.16	0.03	-3.4	3.59	0.97	+	+	+

+0.0 acres in JES survey -- no estimation performed

Note: Less than 5 JES segments in stratum 19; no estimation performed

insufficient sampling or outliers, i.e., atypical segments. Large values of R^2 in some cases were an indication that the former hypothesis was sometimes correct. Further analysis and investigation was not possible due to time constraints, therefore the rather conservative decision was made to use proration (direct expansion) if the regression line slope was less than 0.5 or greater than 1.5. Estimates for tomatoes, walnuts, and, in some analysis districts, corn, rice, and almonds were largely based on proration.

Acreages in cloud- or smoke-obscured areas in Fresno and Sacramento Counties were estimated by proration on JES data alone. The cloud-obscured areas were identified by defining a cloud stratum and using it in addition to the land-use strata defined for the JES.

Estimates of the crops by analysis district were developed using both MIDAS/PEDITOR and BBN/USDA-EDITOR. All modules for the computations involved had been installed and tested on MIDAS/PEDITOR, but some modules could be executed on the BBN system more rapidly. The BBN system was therefore used for some steps in the procedure in order to save time. The analysis district estimates were reported in January 1986. As mentioned earlier, county estimates were developed at BBN, and these were reported in early March.

All estimates were reviewed with Ron Radenz of the CCLRS, and by other CCLRS staff members. The table of Landsat estimates for a crop included, for each of the 19 counties in the test site, the estimate of total acreage and of root mean square error (RMSE) and a breakdown of the numbers by method of estimation, regression and proration. The Landsat estimates of acreages by county were compared with 1985 preliminary planted and preliminary harvested acreages developed at the CCLRS for corn, wheat, rice, and cotton, and with acreage estimates for grapes, almonds, and walnuts listed in 1984 California Fruit and Nut Acreage (ref. 27). The quality of the Landsat estimates was also judged by examining the RMSE included in the tables created by PEDITOR. It was noted that there was good agreement between the Landsat estimates and the CCLRS numbers in many cases, but considerable disagreement in others. Several major cases of disagreement occurred for crop estimates for Kern and Tulare Counties in the southern part of the Central Valley, where the differences between the Landsat acreages and the CCLRS acreages were several RMSE's. The Landsat estimate of rice in Merced County was only about a tenth that reported by the CCLRS and several RMSEs below the CCLRS number.

Other cases of disagreement occurred when the Landsat estimate was primarily based on proration, that is, when the prorated part of the acreage estimate was larger than the regression part. Estimates involving a large component from proration had large estimates of RMSE. This was because the difference between crop prevalence in a county/stratum and crop prevalence in the stratum as a whole was not accounted for in the prorated acreage estimate, because there usually only a few and sometimes no JES segments in a county/stratum for estimating this difference. There was, however, sufficient data to estimate the variation in crop prevalence among counties, and this was included in the estimate of RMSE. The Landsat acreages therefore tended to be inaccurate as shown by comparison with CCLRS acreages, but in general, estimates of precision for these estimates were accurate, i.e., the

magnitude of the difference between the Landsat estimate and the CCLRS acreage was the same or smaller than the RMSE.

The regression part of the total county estimate was negative in a few cases, due to a negative intercept in the regression line and a small number of pixels for the crop. This anomaly was a possible contributor to underestimated acreages such as that for Merced rice and lead to a negative estimate for the rice acreage in Solano County (compared to a report of zero acres in the CCRLS report).

The conclusion of the meeting with Radenz was that while some of the estimates looked good and the inclusion of RMSE's for the estimates was a potentially useful feature of the Landsat acreage report, the cases of inaccuracy were a problem. It was agreed that these problems would be investigated so that they could be understood and perhaps rectified. The emphasis was to be on cotton, rice, and grapes, because CCRLS had confidence in its acreage reports for these crops and the accuracy of Landsat estimates could be assessed. The CCLRS estimates for grapes were particularly accurate because a special survey had been made the previous year (ref. 28).

4.5.2 Revised Estimates

The problems noted above with some of the estimates were addressed at ARC by reworking the county estimates using the original input data, because it was judged that some of the problems were due to ineffective estimation techniques. The Landsat classification in Merced County was viewed on the MIDAS system color monitor using ELAS software. Pixels which had been labelled rice were concentrated in areas known to be the primary rice-growing region in the county. There were 6547 rice pixels, equivalent to over 5,000 acres in terms of the area represented by these pixels, which was much closer to the CCLRS figure (10,100 acres) than the original Landsat estimate (1063 acres). The classification was therefore judged to be generally accurate in delineation of rice fields, and that the problem with the estimate was not due to bad data. As noted above, other estimates derived primarily by proration or involving regression with a negative intercept tended to be inaccurate, suggesting that changes in estimation technique might lead to better results.

Some revisions to the original county estimates reported in early March were made. Experimental estimation techniques were tested and will be described below. These were restricted to Paths 43 and 44, because of concern about the quality of data, in particular potential problems in the classified Landsat imagery caused by procedural errors in Path 42. Much of the Landsat data analysis had been reviewed by CDWR and UCB as part of the accuracy assessment task summarized in section 6. A review of the cluster statistics file for path 42 indicated that some clusters might have been mislabeled, that is, associated with crop different than that reported in the JES survey for fields with the pixels used for development of cluster statistics. Some problems were noted with data for Paths 43 and 44, but these were judged, at least by ARC personnel, to be minor in effect, and the all the new estimates described in this report were based on the same data as the original estimates.

A few experiments were conducted in order to discover the best rules for decisions in construction of a revised set of county estimates. New estimates were created only for counties in Paths 43 and 44, because of concern over the effect of analyst error in classification of Path 42, and only for crops for which CCLRS acreages were available, so that the quality of the results could be judged. New estimates were created, wherein pixel counts for the major agricultural areas, strata 13, 17, and 19, were always included in computations. To ensure that there were a sufficient number of JES segments to develop the regression equations, strata were grouped. The grouping was based on CDWR's judgement from familiarity with the geography of the Central Valley and with the JES data, and on measurements of crop frequencies in segments within strata. Strata 13 and 19 were similar to one another in terms of physical geography and crop mixes and were therefore grouped to form a new combined stratum. Stratum 17 was always included in the estimates separately because the crop mix, dominated by vines and orchards, was distinctive. Stratum 20, when included in an estimate, was also kept separate, as it was primarily rangeland.

Ratio estimates (see section 4.5.3) were developed for walnuts and rice to explore both the effect of including Landsat data from additional strata and of eliminating the possibility of a negative intercept in the regression line. Regression estimates for these crops were also made on the grouped strata. The results are shown in Table 7. The ratio estimates were similar to regression estimates. The new rice estimates were somewhat better than the original estimates, except that the new estimates for rice in San Joaquin County, where the quality of the classification might have been effected by a few thin clouds on the July acquisition of Landsat imagery, were much higher than the CCRLS estimate. The new walnut estimates were much closer to those in the 1984 Fruit and Nut Report than the estimates completed earlier were, because stratum 17 was prorated in the original estimates making the estimates depended on the prevalence of walnuts in the JES segments. Ratio or regression estimates made a major difference because the prevalence of walnuts, as indicated by pixel counts, was much higher in stratum 17 as a whole than in the JES segments.

The results of the experiments with walnuts and rice indicated that better estimates for other crops might be obtained by using regression instead of proration wherever possible, that is, on grouped strata with sufficient numbers of segments. New regression estimates were made rather than ratio estimates because of more developed software and because a good estimator for the variance of ratio estimates at the county level had not been developed. If the estimate of acreage within a stratum was negative, it was replaced with an estimate of zero.

In keeping with the philosophy of maximal use of Landsat pixel counts, the use of stratum 20 was reexamined. Although stratum 20 contains mostly native vegetation, pasture, and grain fields in hilly areas in the Central Valley, other crops were grown there in the 1985 growing season. Personnel involved in the CCRSP had noted almond orchards in some upland areas in the Central Valley during an observational tour of crops in the area. The statewide estimates of corn, grapes, almonds, and walnuts in the JES report [George May, personal communication] indicated that more than 5% of these crops were grown in stratum 20. Estimates were generated for

TABLE 7.- COMPARISON OF SEVERAL TYPES OF LANDSAT ESTIMATES
FOR RICE AND WALNUTS WITH CCLRS ESTIMATES

	Landsat Estimates			CCLRS(4)
	1st est. (1)	2nd est. (2)	ratio est. (3)	PHV
Rice Acreage Estimates for Counties in Landsat Pathes 43 and 44				
Butte	73897	73868	76954	72000
Colusa	114465	113599	113262	97000
Contra Costa	--	1645	654	--
Glenn	67883	67823	69209	68000
Madera	1400	0	393	200
Merced	1063	4963	5528	11000
Placer	5129	4511	3350	4000
San Joaquin	3089	10034	9205	4000
Solano	0	0	1888	--
Stanislaus	1797	0	957	2500
Sutter	77922	77920	60262	72000
Tehama	1435	1420	1333	1600
Yolo	17261	20081	28683	25000
Yuba	26155	24531	23523	27000
14 County Total	391496	400395	395201	384300
Walnut Acreage Estimates for Counties in Landsat Pathes 43 and 44				
Butte	3184	5352	12987	14879
Colusa	2683	5240	2607	4593
Contra Costa	874	1616	494	4552
Glenn	3146	3175	6856	5140
Madera	4102	2514	2354	1822
Merced	6737	8496	12740	8662
Placer	613	242	332	944
San Joaquin	11810	31038	22834	28568
Solano	1557	1472	1614	3102
Stanislaus	7534	35658	26419	24770
Sutter	2628	5593	6285	13957
Tehama	2300	3326	10488	11242
Yolo	2854	3411	2974	6714
Yuba	1360	1877	4532	5744
14 County Total	51382	109010	1133516	134689

- (1) The preliminary estimate using Landsat data, reported in March 1986, Fuller-Battese regression on Landsat on some strata, proration on other strata.
- (2) Using Landsat, Fuller Battese regression on all land use strata with agriculture achieved by defining the regression line slope and intercepts on grouped strata; one group was general crops strata (13, 19), the other group -- the tree crop stratum (17).
- (3) Using ratio estimation (as in Cochran, Sampling Techniques), with the ratio multiplier defined for the grouped strata used in the 2nd estimate.
- (4) Preliminary Planted Acreage from CCLRS, see Note 1 for Table 8C.

these crops both without and, in cases where there were at least two JES pixels with acreage in the crop of interest, with stratum 20. Inclusion of stratum 20 improved estimates of almonds and corn. There were not sufficient JES data to include stratum 20 for walnuts and grapes. Visual examination and review of pixel counts in stratum 20 revealed a very significant problem with errors of commission in the discrimination of grapes on the Landsat imagery so that a good estimate of grape acreage in this stratum with Landsat data might not have been possible even if more JES data were available.

For most crops, the new regression estimates were very similar to the original estimates. The original estimates were, therefore, reported at the final CCRSP review and are shown in Table 8. As expected, the estimates for walnut acreage were much higher than in the original estimates. The newer estimates for grapes in analysis district 44, rice in stratum 43, and tomatoes in analysis district 44 were closer to CCLRS estimates. These estimates are shown in Table 8.

The best estimates for a given crop were the estimates which included the use of Landsat data to estimate acreage in the most important stratum for the crop, and the estimates selected for Tables 8A-8H conform to this rule. Estimates for grapes and tree crops included regression on Landsat pixels in stratum 17. The best estimates for field and row crops included regression on stratum 13, or on strata 13 and 19 combined.

4.5.3 Experimental Estimates

The ratio estimate was tested on two crops, rice and walnuts, because the small number of fields in the JES containing these crops led to regression lines with negative intercepts, and sometimes negative county estimates. The estimate of the county mean was of the form:

$$\text{Est}(Y_{c,h}) = R \times X_{c,h} \quad (5)$$

with:

$$R = y_h/x_h \quad (6)$$

In cases where there was insufficient data for regression, or development of a ratio estimate, strata were pooled as stated in the previous section. The results, shown in Table 7, were similar to those achieved by regression on the same stratum groups, as shown by examination of Tables 8A-8H and discussed in the section 4.5.2. The sum of acreage estimates for all the counties within analysis districts 43 and 44 were closer to the sum of CCLRS estimates when the ratio estimator was used.

Some experiments with a robust regression estimator were carried out. The ordinary least squares (OLS) regression line is often strongly affected by a few outlier points. In agricultural inventories that use Landsat, outliers may have occurred due to errors in JES information or to some condition in a field, such as infestation with weeds or disease, that leads to an atypical spectral response in

TABLE 8A.- LANDSAT ACREAGE ESTIMATES:
ALFALFA

Estimation technique by Landsat path:

P42--s/re^a for strata 12,20; pr^b for stratum 17

P43--s/re for strata 13, 17, 19, 20

P44--s/re for strata 13, 17; pr for strata 19, 20

County	Estimate	Standard error
Butte	3246	1944
Colusa	7908	2922
Contra Costa	---	---
Fresno	82386	13795
Glenn	21401	2374
Kern	65320	10448
Kings	36281	8820
Madera	33797	9723
Merced	79036	10422
Placer	9639	5059
Sacramento	10247	1568
San Joaquin	58566	6964
Solano	8048	1282
Stanislaus	60092	8892
Sutter	3704	1661
Tehama	4905	607
Tulare	52760	6114
Yolo	16144	2636
Yuba	2229	813

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

TABLE 8B.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
ALMONDS

Estimation technique by Landsat path:

P42--s/re^a for stratum 17; pr^b for strata 13, 19, 20

P43--s/re for strata 17, 20; pr for strata 13, 19

P44--s/re for strata 13, 17; pr for strata 19, 20

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Butte	41050	4471	38820
Colusa	6981	3659	14055
Contra Costa	---	---	2267
Fresno	36023	7828	31204
Glenn	13899	3440	12333
Kern	34003	8453	83926
Kings	16670	11132	4922
Madera	30360	15821	33174
Merced	65829	12259	65854
Placer	630	12272	152
Sacramento	5881	3430	23
San Joaquin	38671	4824	37631
Solano	1565	1746	2900
Stanislaus	62996	17508	64545
Sutter	6599	2812	4973
Tehama	10404	2003	7627
Tulare	37135	7601	11187
Yolo	7586	3734	10184
Yuba	5839	1613	1823

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

^cEstimates from L. O. Larson, L. S. Williams, and S. Severson, California Fruit and Nut Acreage, California Crop and Livestock Reporting Service, July, 1985.

TABLE 8C.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
CORN

Estimation technique by Landsat path:

P42--s/re^a for stratum 13; pr^b for strata 17, 19, 20

P43--s/re for strata 13, 17; pr for strata 19, 20

P44--s/re for strata 13; pr for strata 17, 19, 20

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Butte	5988	3659	3000
Colusa	9228	5546	14000
Contra Costa	---	---	---
Fresno	78535	37503	20000
Glenn	14018	4518	10000
Kern	14346	7786	9000
Kings	23350	7632	27000
Madera	13470	3529	18000
Merced	43269	19680	51000
Placer	1923	1896	300
Sacramento	38751	9454	62000
San Joaquin	52173	31501	96000
Solano	14878	4047	53000
Stanislaus	53613	6518	60000
Sutter	17705	10213	9000
Tehama	1826	1082	2000
Tulare	47050	5763	50000
Yolo	40308	12513	44000
Yuba	2369	1689	3000

TABLE 8D.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
COTTON

Estimation technique by Landsat path:

P42--s/re^a for strata 13, 17, 20; pr^b for stratum 19

P43--s/re for strata 13, 17, 19, 20

P44--no estimates, very little cotton grown in P44

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Kern	286236	10292	310000
Kings	238973	11951	270000
Madera	43025	11469	45000
Merced	75741	10566	65000
Tulare	129420	8241	160000

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

^cPreliminary estimates of planted acreage developed by the California Crop and Livestock Reporting Service for the 1985 growing season--obtained through private communication with Ron Radenz.

TABLE 8E.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
GRAINS (WHEAT AND BARLEY)

Estimation technique by Landsat path:

P42--s/re^a for strata 13, 17, 20; pr^b for stratum 19

P43--s/re for strata 13, 17, 19, 20

P44--s/re for strata 13, 17, 19, 20

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Butte	21935	3412	19800
Colusa	45537	2946	27500
Contra Costa	---	---	---
Fresno	99243	24059	88500
Glenn	31239	4116	34800
Kern	65917	14752	55000
Kings	74420	14903	63000
Madera	35911	13273	31700
Merced	47991	7042	30800
Placer	4526	2587	800
Sacramento	29034	21179	28900
San Joaquin	45777	6104	45200
Solano	42484	3492	35000
Stanislaus	24865	4204	8100
Sutter	37613	3528	66500
Tehama	8899	1522	8000
Tulare	42714	9940	45700
Yolo	77954	5198	77300
Yuba	3313	2933	2200

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

^cPreliminary estimates of planted acreage (wheat only) developed by the California Crop and Livestock Reporting Service for the 1985 growing season. Obtained from Ron Radenz.

TABLE 8F.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
GRAPES

Estimation technique by Landsat path:

P42--s/re^a for stratum 17, 20; pr^b for strata 13, 19

P43--s/re for stratum 17; pr^b for strata 13, 19, 20

P44--s/re for stratum 17

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Butte	2011	884	246
Colusa	335	33	147
Contra Costa	1203	508	962
Fresno	231333	24790	214097
Glenn	528	747	1456
Kern	45937	4911	93236
Kings	5623	1372	4085
Madera	77459	5487	87225
Merced	32019	13535	18541
Placer	0	0	126
Sacramento	3859	2242	3705
San Joaquin	73491	21207	55355
Solano	352	289	1233
Stanislaus	30548	4383	20574
Sutter	1316	949	12
Tehama	515	768	162
Tulare	67066	4318	84538
Yolo	219	385	1272
Yuba	565	473	359

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

^cAcreege summarized in the 1984 California Fruit and Nut Acreage, report from a special survey undertaken at industry request and supported by the Winegrowers of California, the California Raisin Advisory Board, and the California Table Grape Commission, with matching funds from USDA. The complete report is California Grape Acreage 1985, May 1986, by J. Tippet, R. Radenz, D. Kleweno, and K. Hintzman.

TABLE 8G.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
RICE

Estimation technique by Landsat path:

P42--no estimates, very little rice grown in P42

P43--s/re^a for strata 13/19, 17

P44--s/re for strata 13, 17, 19; pr^b for stratum 20

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Butte	73879	6186	72000
Colusa	114465	8704	97000
Contra Costa	---	---	---
Glenn	67888	7269	68000
Madera	178	363	200
Merced	4789	1572	11000
Placer	5129	1543	4000
Sacramento	10883	5508	9400
San Joaquin	3089	3320	4000
Solano	0	4266	0
Stanislaus	547	945	2500
Sutter	77922	6524	72000
Tehama	1435	770	1600
Yolo	17261	9800	25000
Yuba	26155	3429	27000

TABLE 8H.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
TOMATOES

Estimation technique by Landsat path:

P42,43--no estimates using Landsat because no significant correlation between pixels and JES acreage

P44--s/re for strata 13/19, 17

County	Landsat		CCLRS ^d
	Estimate	Standard error	
Butte	1887	2870	---
Colusa	6531	3804	10100
Contra Costa	3809	1876	5150
Glenn	3083	3273	---
Placer	93	709	---
Sacramento	3844	4897	3900
San Joaquin	21451	13167	29100
Solano	13263	2324	11500
Sutter	15988	4318	15600
Tehama	1540	1010	---
Yolo	29983	5973	43500
Yuba	2268	1611	---

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

^cSee note 1, Table 8C.

^dSee note 1. Processing tomatoes only except for San Joaquin County which includes 5700 acres of fresh tomatoes.

TABLE 81.- LANDSAT ACREAGE ESTIMATES:
TREEFRUIT

Estimation technique by Landsat path:

P42--s/re^a for strata 17, 20; pr^b for strata 13, 19

P43--s/re for stratum 17; pr for strata 13, 19, 20

P44--s/re for stratum 17; pr for strata 13, 19, 20

County	Landsat	
	Estimate	Standard error
Butte	37645	2491
Colusa	10724	1222
Contra Costa	---	---
Fresno	76365	8385
Glenn	21833	1612
Kern	41009	10237
Kings	7150	3118
Madera	10739	1953
Merced	25563	4914
Placer	1423	1442
Sacramento	5097	1149
San Joaquin	26890	7205
Solano	9683	1459
Stanislaus	24402	2877
Sutter	37346	3257
Tehama	24860	1749
Tulare	155654	8314
Yolo	11031	2614
Yuba	22281	1478

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

TABLE 8J.- LANDSAT AND CCLRS ACREAGE ESTIMATES:
WALNUTS

Estimation technique by Landsat path:

P42--s/re^a for stratum 17; pr^b for strata 13, 19, 20

P43--s/re for stratum 17; pr for strata 13/19, 17

P44--s/re for stratum 17; pr for strata 13/19, 17

County	Landsat		CCLRS ^c
	Estimate	Standard error	
Butte	5353	2075	14897
Colusa	5241	1604	4593
Contra Costa	1616	667	4552
Fresno	11296	6667	3285
Glenn	3176	1785	5140
Kern	5028	2289	1367
Kings	3789	1136	4794
Madera	2514	1465	1822
Merced	8496	2301	8662
Placer	243	227	944
Sacramento	1834	549	205
San Joaquin	11810	6379	28568
Solano	1473	946	3102
Stanislaus	35658	3059	24770
Sutter	5594	1881	13957
Tehama	3326	1601	11242
Tulare	15297	2226	26163
Yolo	3412	2148	6714
Yuba	1877	916	5744

^aLeast squares regression on a single variable

^bProration on JES survey data (Landsat not used)

^cEstimates from L. O. Larson, L. S. Williams, and S. Severson, California Fruit and Nut Acreage, California Crop and Livestock Reporting Service, July, 1985.

the Landsat pixels contained in the field. When outliers are present, LS regression sometimes yields poor estimates. An alternative regression line, where outliers are down-weighted, might lead to improved estimates. Such a robust estimator, described in Appendix C, was developed following Huber (ref. 18).

The Huber robust estimator was tested on simulated data, created using random number generators, and on selected data from the 1985 inventory. The simulations showed that if the distribution of deviations from a straight-line relationship between pixels and acreage was a mixture of two normal (Gaussian) distributions with different variances, the small variance corresponding to the more common segments and the larger variance to outlier segments, the robust estimator was more precise than the least square regression estimate.

Table 9 shows regression parameters derived by robust estimation and by least squares for selected crops in AD43 and AD44. The behavior of the robust estimator was dominated by statistics for JES segments that did not contain the crop to be estimated, according to JES data, but often contained a substantial number of pixels assigned to the crop. Each crop was contained in a minority of the sampled segments in the survey. That minority contained most of the segments flagged as outliers and, subsequently, down-weighted in the computation of the crop-specific regression parameters. In most cases, the down-weighting of these outliers led to a decreased y-axis intercept b_0 , a higher slope b_1 , and a small increase in R^2 . In all cases, the differences between the robust parameters and the LS parameters were within one standard deviation and, therefore, not significant. In a few cases, the R^2 value was smaller for the robust regression line. This occurred in cases where the LS R^2 was already low--less than 0.50. The decrease was due to an effective decrease in the range of the data caused by down-weighting some of the data points.

4.6 Landsat Map Products

The agency in CCRSP most interested in map products was CDWR. Copies of the registered data set, and the classified images from the seven frames were sent to CDWR from ARC as they became available; CDWR has experimented with generating map products from those materials.

A mosaic of the seven frames of classified imagery was assembled at ARC and distributed to the CCRSP participants. The classifications were edited before the mosaic was compiled. Urban areas, major north/south highways and major rivers (following the California definition i.e., any creek, run, stream, or gulch with moving water year-round is a major river), and locations with cloud cover were redrawn on the classification.

The editing was done to increase the accuracy and interpretability of the classification. The features redrawn were generally mis-classified, because they were not agricultural and there was insufficient training data for them.

TABLE 9.- ROBUST ESTIMATES OF REGRESSION PARAMETERS

Stratum 13/19					
	b_0	b_1	r^2	no ¹	nnz ²
AD43:					
Almonds	+	+	+	+	+
Corn	23.2, 18.2 ³	0.32, 0.35 ³	0.40, 0.48 ³	7/62	6/32
Cotton	-0.7, 0.8	0.84, 0.85	0.81, 0.86	5/62	4/24
Grapes	+	+	+	+	+
Rice	+	+	+	+	+
Walnuts	-0.6, -1.8	1.18, 1.25	0.63, 0.86	8/62	4/8
AD44:					
Almonds	-15.8, -8.9	1.02, 0.57	0.46, 0.33	6/80	5/5
Corn	-23.2, -19.5	1.44, 1.27	0.65, 0.66	8/80	5/27
Grapes	+	+	+	+	+
Rice	10.3, 2.1	0.97, 0.98	0.82, 0.89	10/80	7/32
Tomatoes	-8.9, -4.3	0.39, 0.27	0.23, 0.15	9/80	9/15
Walnuts	1.6, 0.2	0.33, 0.06	0.19, 0.04	11/80	10/10
Stratum 17					
	b_0	b_1	r^2	no ¹	nnz ²
AD43:					
Almonds	-14.1, -15.0 ³	1.14, 1.23 ³	0.59, 0.67 ³	6/62	4/18
Corn	-0.8, -3.0	0.94, 0.95	0.62, 0.71	11/80	9/17
Cotton	-8.2, -7.7	0.77, 0.75	0.71, 0.75	11/80	6/20
Grapes	-24.4, 24.6	1.06, 1.05	0.88, 0.90	9/80	6/38
Rice	+	+	+	+	+
Walnuts	-2.2, 0.61	2.87, 0.35	0.50, 0.04	12/80	10/16
AD44:					
Almonds	-8.0, -10.2	0.99, 0.92	0.52, 0.55	5/40	4/12
Corn	3.9, 1.2	0.17, 0.07	0.08, 0.04	5/40	5/5
Grapes	+	+	+	+	+
Rice	+	+	+	+	+
Tomatoes	-10.3, -10.8	0.76, 0.77	0.68, 0.73	5/40	4/5
Walnuts	16.0, 15.3	0.16, 0.15	0.03, 0.03	3/40	3/19

+Insufficient data for robust estimation; <5 segments with Crop or <8 segments in sample

¹Number of outliers as a fraction of sample

²Number of non-zero outliers as a fraction of sample non-zero segments

³First number is LS parameter, second number is robust parameter

The images were edited as follows: A band of Landsat imagery (usually MSS5) from the primary scene in each frame was loaded on MIDAS and displayed using CIE software (section 3.3.3). Viewing the displayed image, and using a cursor and bit pad, polygons were inscribed around the features of interest. A roll of high altitude photography and a roadmap of the Central Valley were consulted to locate the boundaries of the features as accurately as possible. The polygons outlined were stored on disk. The classified image was then displayed and the polygons redisplayed with it. CIE was used to change the digital values of the pixels within the polygons to a new value representing the feature in that location. For example, all pixels in the areas identified as urban were assigned the digital value 250. The pixels values were altered in the image first, and, after confirmation that the new value was correct, the pixel values on the disk file were changed similarly.

After all the frames were edited, hard copy imagery was produced. The classified images were split into 512×512 pixel pieces, each piece was sampled from a 1024×1024 block of imagery. The pieces were enlarged by a factor of two and written to photographic negatives using a Dicomed film writer. Thirty-six prints were required for coverage of the entire Central Valley. The final product was generated by assembling a mosaic of the prints, cutting off the area outside the Central Valley physiographic province, and photographing the mosaic.

4.7 Data System Performance

The hardware and software applied to the 1985 data constituted what is referred to below as the data system. The software included the PEDITOR, ELAS and CIE software packages, special purpose, single function programs, and communications software (Kermit and Arpanet). Data processing was performed on MIDAS workstations, the VAX 11/780 in the ECOSAT branch at ARC, the Cray X-MP at ARC and a PDP20 at BBN.

The demands placed on the data system differed during the inventory and evaluation phases of the 1985 test. The inventory phase requirements were primarily operational. The functional requirements of the software were known in advance. The machines of preference were selected, and emphasis was on processing the data as quickly as possible. Flexibility and experimentation became guiding principles for the evaluation phase. An outline for the evaluation was prepared before the inventory, but changes in the availability of data for the evaluation forced the analysts to design new tests and make unanticipated demands on the data system.

The data system was adequate to meet the goals of the 1985 inventory, but the data processing was accomplished with difficulty. At a meeting of the CCRSP participants after the conclusion of the inventory phase, it was generally agreed that the data system employed for the inventory was not operational. Problems encountered with the data system fell into six categories. Some problems were directly related to the structure and operation of the data system, others developed from the manner in which the data system was used. The categories were:

1. Uniformity of PEDITOR code
2. Flexibility of PEDITOR code
3. Analyst training
4. System speed
5. Disk space
6. Software/hardware errors (bugs)

Data system performance during the inventory phase is described in sections 4.8.1 to 4.8.5. The performance during the evaluation phase is described in section 5.

4.7.1 PEDITOR Software

A protocol for distribution of PEDITOR code to all CCRSP participants was established by UCB and presented at the project review in October 1984. The protocol was accepted, and UCB was assigned the responsibility to implement it. A distribution protocol was necessary because new PEDITOR modules were being completed at a rapid pace by programmers at NASS, ARC, and UCB between the fall of 1984 and the start of the inventory. In addition, as the completed code was tested and bugs were uncovered, code fixes were broadcast to all users, and the code was recompiled as necessary.

The distribution of changes to PEDITOR became erratic during the inventory. The breakdown was caused by a number of factors. PEDITOR was completed, to the satisfaction of NASS, in the summer of 1985. The version that resided at BBN was transferred to Washington and became the mainframe version operated by NASS on the Martin-Marietta system in Florida. With that event, and the cessation of USDA operations at BBN, an on-going need for the distribution of changes in the code was no longer clear. Bug reports were not regularly available to all users, and each programmer fixed any bugs and distributed new codes as he or she deemed appropriate.

The effect of the breakdown in distribution of the code was exactly what the protocol was implemented to avoid i.e., a different PEDITOR at each node of CCRSP. The differences were often significant. PEDITOR modules that did not function correctly at CDWR, for example, worked correctly at ARC. Fixing bugs became much more difficult as the programmer had to determine which versions of the module and related libraries were being accessed before making a correction. The lack of a uniform PEDITOR code contributed to nagging delays in data processing during the inventory and became a more significant problem later.

4.7.2 Other MIDAS Software

ELAS and CIE were complimentary to the 1985 inventory. Both software packages were used for tasks during the evaluation phase of CCRSP. Problems encountered with ELAS were few and were generally due to inherent weaknesses or limitations in the modules. CIE was used marginally but without difficulty.

4.7.3 MIDAS Hardware

The lack of a test data set similar in size and complexity to the inventory data set proved to be unfortunate. Although MIDAS performed well with test data, system performance degraded substantially during the inventory. The degradation was noted particularly in system speed, but system operations were also adversely affected by disk size and undefined system bugs. It is likely that, had the operational characteristics of MIDAS been better understood prior to the inventory, a recommendation to perform the data processing at BBN would have been made.

MIDAS is slow. The system was designed to perform the functions typically encountered during digital image processing. However, it was agreed during the planning for the inventory that certain functions, e.g., classification of large areas, while possible on MIDAS, would have to be completed on a more powerful machine in order to meet the reporting schedule of the inventory.

The size of the 1985 data set slowed, substantially, the processing of virtually all PEDITOR functions on MIDAS. Generally, operations intended for MIDAS were completed on MIDAS, but the speed of the system forced the sharing of some operations with other MIDAS workstations and completing a few operations on another system. For example, data packing, intended for the CDWR MIDAS, was split between the MIDAS workstations at CDWR and ARC. The output from CLASSY was reformatted into crop-specific statistics files, and the output from Aggregation was reformatted into county-specific aggregation files, on the ECOSAT VAX rather than a MIDAS workstation.

Disk storage capacity on MIDAS adversely affected the operations of the system. About 30 MB of storage was available for inventory data. The storage capacity was sufficient, at best, for the data from one analysis district. Consequently, substantial offloading and loading of data was required. The time required for data management had a significant impact on the efficiency of operations during the inventory and evaluation phases.

4.7.4 BBN/EDITOR

Data processing on BBN/EDITOR was confined to estimation and strata mask editing. No serious problems were encountered with the BBN operations. At certain times of the day, delays in response time, probably caused by congestion in the network, slowed processing. NASS phased out its BBN account in 1986, therefore PEDITOR at BBN ceased to be available by the summer of 1986 and was not an alternative for data processing during the evaluation phase of CCRSP.

4.7.5 Networking

A MIDAS-VAX-Cray network was established to facilitate data processing during the inventory. The crucial link was the connection between the CDWR MIDAS and ARC. The link was accomplished with Kermit software. Kermit worked as expected; it was accurate but slow. While planning the inventory, it was felt that jobs for the Cray could be set up at CDWR, transferred to the SLE VAX, submitted to the Cray and

the output file transferred back to CDWR. Most of the Cray processing was actually set up at ARC either by ARC personnel or by off-site personnel logging on to the SLE VAX. Most files were transferred by tape. Only relatively small text files were moved routinely among systems by Kermit.

5. ANALYSIS OF INVENTORY DATA

Evaluation of the inventory results began in March 1986 after delivery of the county level estimates to CCLRS. Ames Research Center supported the evaluation tasks with data processing and software assistance as required.

5.1 Data System Performance

Significant problems encountered during the data evaluation phase precluded as thorough an analysis of the data as desired or planned. Some problems were related directly to the operation of the data system.

Many of the problems with the data system described in section 4.7 had a greater negative impact on the evaluation than on the inventory. For example, the inflexibility of PEDITOR software was a minor problem for the inventory but caused dismay during the evaluation when attempts were made to process data in an unorthodox manner.

The cornerstone of the evaluation design was the 10% sample of JES segments in which internal field boundaries had been digitized. The sample was compiled by NASS personnel in Sacramento. Examination of the test data set in the spring of 1986 uncovered inconsistencies, omissions and apparent errors in many of the segments. Attempts to confirm the accuracy of all the segments and to reconstruct the flawed segment data proved to be impossible, because the field data entered by the JES enumerators on the aerial photographs had already been erased. The questionable validity of the 10% sample set off a search for alternate test data. Using either the transect data or the remaining JES data required manipulating PEDITOR modules in a manner beyond their original intent. Such manipulations proved to be difficult, if not impossible, and sometimes generated data of no value.

The inflexibility of PEDITOR software was compounded by inadequate analyst training. An analyst can be trained to perform standard PEDITOR processing with a modicum of effort. To understand the intricacies of the code and to be able to manipulate the code to its fullest extent requires interest, skill, and substantial experience. Most of the data processing burden for the inventory fell to CDWR. Prior to the inventory, no CDWR staff member had performed any area estimation with PEDITOR. With the assistance of CCLRS, ARC, and UCB, all of whom had personnel with some EDITOR/PEDITOR experience, CDWR was able to complete the inventory. The data processing burden for the evaluation phase fell to CDWR and UCB. The experience with PEDITOR that both agencies gained during the inventory proved to be inadequate when faced with the data processing needs of the evaluation. Lack of a thorough understanding of the code contributed to errors that delayed and limited the effectiveness of the analysis.

The evaluation process uncovered a number of software/hardware bugs or unexpected features. The errors encountered were often difficult to resolve, because it was uncertain if the cause of the error was analyst inexperience, a

genuine bug in the software, a peculiarity of MIDAS hardware, or use of an outdated version of a module or related library. Correcting an error was often an intensive undertaking that was usually organized in an informal manner, often required a significant amount of programmer and analyst time, and was not always successful. The greatest impedence to a quick resolution of several of the data system problems was the lack of uniform PEDITOR code among the workstations.

Data system problems were not confined to the workstations. Early in the evaluation phase, some inconsistencies were noted in the output from the CLASSY clustering algorithm implemented on the Cray X-MP at ARC. An error was traced to a bug in the preprocessor accessed by CLASSY. Data sets read from disks were interpreted incorrectly if larger than one Cray block. The error had escaped detection because it occurred only in relatively large files and only with six-channel input data.

Because the error in the preprocessor affected the output from clustering and, consequently, the cluster statistics files, it cast doubt on the validity of the final classification and area estimates. New classifications and estimates were generated. The original and corrected results were presented at the project review in October 1986.

5.2 Accuracy Assessment

The University of California, Berkeley and CDWR worked together to assess the accuracy of the ground survey data and classified Landsat imagery. The analysis of the classified imagery led to an examination of some of the data processing algorithms and procedural steps in EDITOR. Ames Research Center provided data processing support for accuracy assessment as needed. The results from accuracy assessment are summarized here, because they affected the acreage estimation work at ARC. The estimates are described in section 4.5. Cathy Travlos (UCB) and Jay Baggett (CDWR) conducted the accuracy assessment and reviewed the EDITOR processing. Visual examination of the classified Landsat data on a MIDAS color graphics monitor, and comparison with recent CDWR land-use maps showed that, in most areas, the quality of the classified imagery was good. Field labels were accurate and most fields were well defined. The classification appeared better in the Sacramento Valley than in the San Joaquin Valley. Many errors in the classification were explained by similarities in appearance and phenology among crop/land-use categories as "seen" by Landsat. For example, confusion in the classification between wild grasses and grains was attributed to the similar appearance of the two land-cover types and their concurrent growth stages. Similarly, native riparian vegetation sometimes was confused with tree crops, because one species in the riparian vegetation, a native walnut, was similar in appearance and phenology to the commercially grown English walnut.

Other errors in the classification were more difficult to explain. Confusion between grapes and other crops, including cotton, was noted with concern, because it was unexpected. The confusion may have been caused by the presence in the vineyard of some understory with a phenology similar to that of cotton or by some condition,

) such as excessive salinity in cotton fields, that caused deficient plant development. No explanation could be verified.

Procedural errors and software errors were also uncovered. There appeared to be inaccurate crop reporting in the JES in at least one county, Colusa County. For certain segments, the Landsat classification and the CDWR survey work were in agreement, but the JES labelled the fields differently. Numerous bugs in the software were encountered during the course of the inventory. Most of the bugs were due to the size of the data set and did not affect the estimates.

6. CONCLUSIONS AND RECOMMENDATIONS

Two objectives of CCRSP were to conduct a current year, Landsat-based, inventory of the Central Valley and to supply acreage estimates, from the inventory, to CCLRS in a timely manner. Those objectives were met. A secondary objective of CCRSP was to perform the data processing on a MIDAS-VAX-Cray network and evaluate the operational characteristics of the system. The program demonstrated the feasibility of doing large-scale (multiple Landsat frame) area estimation when most data processing functions were performed on a microprocessor. The system that was used, however, is not operational, and will not be operational until a number of problems are resolved.

6.1 Data Processing

With the exception of the county estimation program (ESTCO), all data processing for the 1985 inventory could be performed on a MIDAS workstation and Cray XMP.

6.1.1 Implementation of PEDITOR

In the midst of the 1985 inventory, PEDITOR began to evolve into two different systems. The workstation version of PEDITOR resided, officially, on a MIDAS workstation at ARC. A mainframe PEDITOR was implemented by NASS on an IBM system at Martin-Marietta in Florida. Both systems evolved from a common source but changed in response to different operating environments and analyst needs. The mainframe system was optimized for operations, but the workstation system was nurtured in an experimental environment. Most differences between the systems are minor; the basic data flow remains the same. It is likely, however, that some bugs, recognized and corrected on one system, have not been changed on the other. Currently, there is no ongoing communication between the programmers developing the workstation version and the programmers working the mainframe version, nor is there managerial direction on how the two systems should evolve.

Because of the workstation/mainframe divergence, no standard version of PEDITOR exists. However, it may be appropriate to maintain two PEDITORS. As long as there is a need for an operational, mainframe system, it is reasonable to optimize the operational efficiency of the software for that purpose. But if PEDITOR is wanted as an experimental tool, the software should have greater flexibility and be more interactive than is practical in an operational system. If two systems are maintained, there must be centralized oversight so that successful modifications to the experimental version will be incorporated into the operational version.

Whatever the outcome of the workstation/mainframe divergence in PEDITOR, a standard version of PEDITOR must be distributed and maintained on the workstations. The existing differences among the PEDITORS at CDWR, ARC, UCB, and NASS make operation of the system exceedingly difficult. The differences became crucial during the evaluation phase of CCRSP and were particularly troublesome at CDWR. A number of "quick fixes" were introduced into the CDWR PEDITOR. As a result, the

CDWR system is probably different from any of the others. Correcting the problem will require two tasks: distribution and implementation of a common code, and retesting the code to assure that all bugs have been removed from the latest version. After a common PEDITOR is established, a protocol for distribution of changes must be reinstated and strictly adhered to, or the local system manager must take responsibility for maintaining the system.

6.1.2 Processing Environment

The experience of the 1985 inventory suggests that the PEDITOR software cannot be evaluated properly without considering the hardware on which it operates.

The MIDAS workstation was not adequate for an inventory of the size completed. It is likely, however, that the existing MIDAS configuration could process a frame or analysis district efficiently. It is also likely that the new generation of workstations, e.g., Sun3/4, or Apollo, could take on a task as large as the Central Valley inventory. Preliminary tests of PEDITOR code performed on a SUN2 workstation at Ames were completed with greater speed and fewer system errors than on MIDAS.

All data processing completed on the Cray worked well. Because there was no software, prior to the inventory, to create a six-channel data set, the compilation of those tapes required a significant amount of analyst interaction in the data processing. However, analyst interaction was limited to job set-up and confirmation of results. There is no inherent need for the analyst to manipulate or view the data from the acquisition of the the raw data through registration of the scenes, to generation of the data set. Combining the processing stages into a single job would improve the efficiency of the process. Furthermore, it is possible that the improved stability of the newer Landsat platforms (refs. 2,29,30) and the control-point information on the Landsat CCT's (refs. 1,31) might make simplified versions of processing multirate Landsat imagery possible.

The workstation-mainframe network worked well for communication and small-scale data transfers but was not adequate, as expected, for transferring large data sets. Unless a high-speed interface is established between the workstation and the mainframe, there will be a need to transfer data via magnetic tape. Without on-site personnel dedicated to monitoring and assisting the flow of data tapes, the need to use tape transfers will reduce the efficiency of the processing and increase the time required to generate estimates and other products.

6.1.3 Data System Recommendations

The following recommendations are made based on the preceding discourse:

1. A determination should be made on the functional future of PEDITOR. Maintaining operational and experimental PEDITORs is recommended. The operational PEDITOR should be optimized to perform an established procedure. The Experimental PEDITOR should be highly interactive with the accompanying flexibility.

2. A uniform workstation version of PEDITOR should be completed, tested and distributed to all interested users. The code should be tested on simple (single scene) and complex (multiple scene/frame) data sets before it is certified. A commitment on which machines will be supported and a statement on the extent of the support are needed.

3. Present users of MIDAS should consider alternate machines. Assuming proper operation of MIDAS, its limited disk space, uncertain support, and "old technology" weigh against relying on it as an operational tool.

4. The Cray software for preparation of Landsat data should be modified so that all processing stages, from AMERGE through COMPILE, are combined in a single job.

5. An operational program for area estimation using a workstation/mainframe network should not be implemented unless a high-speed communications link is available for large-scale data transfers.

6.2 Use of Landsat Data

Conclusions and recommendations about the utility of Landsat for crop surveys in California are listed below for the two modes of usage tested in CCRSP work: improvement of JES estimates of major crop acreages and crop mapping.

6.2.1 Acreage Estimates

1. The use of Landsat pixel counts for estimating acreages for the land-use strata where most of the crop in question is to be found usually improves the accuracy of regional and county estimates. The improvement can be quite substantial for crops that are concentrated in a few areas within the lands use stratum, such as walnuts and rice.

2. The quality of the estimates is not sensitive to details in technique. As was noted in early CCRSP research, correlation between pixels counts and acreage varied considerably among crops and localities and to a lesser extent among classifications involving different methods of discrimination. The results reported for the 1985 inventory indicate that regression estimates using standard OLS formulae result in acreage numbers which are similar to ratio estimates and to robust regression estimates.

3. The estimates might be improved by additional analysis to locate crops by land use stratum in each county. Some crops in California are being cultivated in the "rangeland stratum" (stratum 20). Others may be missing one or more of the "agricultural strata"(strata 13, 17, 19) in some counties. An analyst familiar with the geography of Californian agriculture, such as a member of CDWR or the CCLRS, could locate crops by visual analysis of the Landsat crop map with overlays delineating county and stratum boundaries. Such an analysis would be useful in cases where there are few JES segments with the crop of interest in order to choose to

either estimate crop acreage with zero, a proration estimate, or a Landsat estimate using calibration (regression) parameters developed with data from another stratum.

4. The estimates reported for the 1985 survey were effected by the small (JES) sample size. The data for some segments were lost due to procedural problems, therefore the quality of results was less, by some unknown amount, than was potentially achievable with the methods used.

5. Generation of acreage estimates with processed Landsat data for 10 crops was slow and difficult with the computer hardware and software used for the 1985 inventory. The following changes are recommended:

a) Either change hardware to increase speed of execution of programs or change text files containing user input for a program, and these were used extensively when the input was in the form of lists of items of the same type (lists of segments, lists of aggregation files, etc.). The programs required input of many types however, and the some programs had to be run once for every crop and analysis district.

b) Put more information (location of segments by county) in the "segment totals file" used to develop the regression line parameters for an analysis district, so that it can be used for development of the B-F county estimates.

c) Increase the use of tabular forms for storage of data in files and in program output. A standard form could be the following: line 1 is title (for example, "estimates by county"), line 2 is headings for columns of data, line 3 is the Fortran format for each of the succeeding lines, and the remaining lines contain data in rows (one line=one row) and columns. As in the example, the title would always identify what the rows correspond to. Such a tabular form is compact and easy to read.

d) Files containing data by segment should contain missing value code to make it easier to handle cases of missing data properly and flexibly. In the 1985 inventory, some experimentation was required to ensure that missing data was dropped rather than treated as zero valued. Much of the missing data in the 1985 inventory was due to segments under cloud or smoke cover on the Landsat data, and some of these segments were treated as missing data even for proration.

e) The code for regression estimates should include an option for replacement of negative stratum estimates with zeros.

f) Consider use of a statistics program package such as SAS for estimation work. Some of the suggestions for improvement in the operation of EDITOR/PEDITOR above lead in this direction, as tabular files with missing data codes are supported by most packages. Statistical program packages would support continuing research and changes in procedures. Use of a package of development of estimates would also accomodate interagency work such as the CCRSP. Once files with pixels

) counts were developed, estimates could be made by an analyst with no knowledge of image processing or EDITOR.

6.2.2 Landsat Map Products

1. The software on the MIDAS system, including ELAS and CIE, supports map product development.

2. Regional Landsat map products, such as the one included in this report, show the major areas of cultivation of important crops.

3. Landsat map products show field by field distribution of crops in most areas in the major crop strata.

APPENDIX A

PEDITOR MODULES AND FUNCTIONS AS OF 2/20/87

Module Name	Function
accum	Accumulate Estimates with Proration
addagg	Add or Subtract Aggregation Files
aggr	Aggregation Functions
asma	Automatic Segment Matching
badpix	Range Check of Pixel Values
button	Button to Menu File Creation
calcor	Calculate Coordinates
cated	Segment Catalog File Editing
clas	Maximum Likelihood Classification
clust	Cluster Window Data
cmaskp	Pixel Count of Mask Fields
compak	Combine Packed Files
compar	Compare Categorized Files
correc	Percent Correct Calculation
cpedit	Control Point Editor
cracon	Translate Cray Aggregation Output
crtape	Cray Tape Read
cvstat	Convert a Statistics File
cvwin	Convert Window to Cie/Elas Format
dlgscn	Scan DLG Tape
dspdlg	Digital Line Graph File Display
dspmsk	Display a Mask File
dspwin	Display Window File
editcr	Examine Correlation Output
edunit	Frame Unit File Editing
epwin	Elas to Peditor File Conversion
eraspl	Clear Display
estl	Large Scale Estimation
ests	Sample Estimation
extent	Determine Segment Window Extents
gmfdip	DIP File Generation
gmfras	Raster File Generation
group	Group Categories in a File
gtruth	Ground Truth File Editing
ident	Identify a File
imgen	Generate an Image from a Mask
lltape	Line-by-Line Tape Read
mapima	Mapping Functions
mctynm	County Check Between Files
medit	Edit a Mask File
modcm	Categorized Color Mapping

Module Name	Function
mshgen	Segment Mask Generation
mshplit	Split Masks by Frame
ncray	Cray Job Creation
nwstcl	Reformat Classy Output
pack	Field Selection for Analysis
pedit	Peditor Driver
peeker	Binary File Dump Routine
poly	Polygon Functions
prmenu	Button Assignments
rdcorr	Reformat Cray Correlations
reclas	Reclassify a Categorized Image
refo	Reformat Window File
regdlg	Register DLG to Image
rtdisp	Load Image to Display
rtinit	Initialize Display Device
runsys	Execute System Commands
scat	Scattergram of Pixel Values
segdsp	Segment/Polygon File Display
seged	Segment Network Editing
segplt	Segment Plotting
setdst	Alter Display Status
showds	Show Display Status
stated	Statistics File Editing
stot	Totals File Editor
stplot	Statistics File Plotting
subwin	Subwindow Window File
svcal	Save Segment Calibrations
tapdlg	DLG Tape Read
tapwin	Tape Read to Create Window File
tdcopy	Read Binary Tape to Disk
wrtape	Write Binary Tape from Disk

APPENDIX B

RECOMMENDATIONS FOR THE 1985 INVENTORY

(Presented by R.W. Thomas, Remote Sensing Research Program, during the CCRSP review, October, 1984.)

I. Information Product Objectives

A. Acreage Estimates

1. Analysis district, land-use stratum, and county estimates by
 - a. crop type - cotton, barley, wheat, rice, tomatoes, permanent pasture, corn, and alfalfa highest priority
 - b. broader land-use category
 - c. irrigated vs. nonirrigated land
2. Regional and statewide estimates for above categories
3. Sample frame-count unit totals for selected categories

B. Map products

1. 7 1/2' thematic maps
2. Regional class maps with survey features shown.

II. Recommended Inventory System

A. Sample frame

1. Use the current USDA frame for acreage estimation
2. Use an independently constructed "frame" for estimation of crop/land-use type spectral means and covariances

B. Sample allocation

1. Acreage estimation: use set of 1985 JES segments
2. Estimation of spectral parameters: Obtain a systematic sample of fields along a county transect
 - a. use CDWR land-use maps to quantify crop presence on a 2 1/2' block basis
 - b. locate areas of "homogenous" spectral mix using Landsat imagery from previous years
 - c. use a and b together with high-flight photography and road maps to locate road transect by county expected to adequately sample range of target crops and confusor/spectral distributions
 - d. draw a systematic sample of stop points (one every 2 miles) along each transect
 - e. all fields (satisfying minimum size criteria) touching these stop points will be selected for estimating spectral means and covariances

C. Measurements

1. Ground

a. JES

- 1) June Survey: standard with these additions:
 - use of more current photography
 - comment section to flag questionable labels or unusual field conditions
 - presentations to enumerators in May
 - possible assignment of most experienced enumerators to survey segments later used for accuracy assessment
 - capture of field boundaries on acetate copy for later use in evaluation
- 2) Followup survey
 - check questionable labels or unusual conditions
 - check intension fields

b. Transect

- 1) Windshield survey
- 2) Record for each field selected: stop number, field number, crop/crop/land-use, irrigated vs. not, note bad field, comments, date
- 3) Visit twice Mid-Spring Mid-Summer

2. Landsat

a. Assumptions

- 1) MIDAS, ARC VAX, ARC Cray will be the primary network used for processing
- 2) BBN will be used as backup
- 3) DWR MIDAS will be the primary MIDAS used for processing data for the main 1985 test
- 4) DWR and SSO will have the primary responsibility for actual data processing for the main 1985 test
- 5) UCB MIDAS and personnel will be available to process overload as necessary
- 6) NASA-ARC MIDAS/personnel as final backup

b. Initial processing

- 1) Acquire MSS data
 - three date goal, May through early August
- 2) Reformat digital data
- 3) Register scene-to-scene
- 4) Perform Tasselled Cap transformation if necessary
- 5) Generate six-channel data tapes

c. Spectral training

- 1) transect field digitization
 - Osborne/MIDAS interface
- 2) creation of "segment" catalog file (index)
 - fields at each stop will be considered to form a segment
- 3) create segment mask file and register to Landsat
 - image display to check

- 4) cluster by crop/land-use type
 - CLASSY, ISOCLAS
 - 5) classify transect fields
 - 6) edit clusters
 - analysts aids:
 - standard statistics and plots
 - multi-segment display
 - eliminate anomalous signatures, possibly also field edge clusters
- d. Landsat classification
- 1) Initial stratification
 - a) obtain digitized USDA land-use stratification (use to remove areas not subject to Landsat-aided estimation)
 - b) digitize out major "blobs" of urban and residential, possibly also riparian
 - use CDWR land-use maps and recent aerial photography
 - c) create mask for each stratification, register to Landsat, and check registration
 - d) merge both masks into one mask to be used in classification
 - 2) Classify sample segments
 - a) use edited statistics file excluding nonagricultural classes from training
 - b) apply CLASSIFY (maximum likelihood classifier)
 - threshold to a special fill category of those pixels with posterior probability less than a threshold established during cluster editing
 - c) perform an error analysis
 - generate tabulation and percent correct files on accuracy assessment segments
 - run regression, obtain plots and x,y table files
 - display multiple segment block files
 - identify outlier segments using regression plots and tables
 - examine outlier segments using USDA field maps, tabulations, display, and raw data statistics to determine cause of error
 - drop segments (or fields?) for which strong evidence exists that ground data is inaccurate
 - 3) Classify full frame
 - a) classify as with sample segments
 - b) summarize counts by class by count unit, county, stratum, and analysis district
 - c) if summary for some other region is desired then digitize, create mask, register to Landsat, and apply to class map for count summary

- D. Acreage estimation
 - 1. Form of estimator
 - a. Analysis district level (by stratum and crop type)
 - use standard USDA single variate regression estimator
 - b. County level
 - use standard USDA regression procedure
 - 2. Aggregate by
 - a. stratum
 - b. analysis district
 - c. county
 - d. statewide
 - e. regionally if implemented
- E. Map products
 - 1. 7 1/2' quadrangle maps
 - a. a sample of these will be produced on electrostatic plotter
 - b. map to contain crop/crop/land-use symbols, state plane lines, stratum boundaries, and masked area
 - 2. Color photographs
 - a. crop group and type shown for 1/4 Landsat scene
 - b. coordinate system and other cultural feature separates from USGS photo-overlaid into image

III. Associated Experiments in 1985

- A. Use of masking to improve classification performance
 - examine impact of various levels of detail in mask on one or two study sites
- B. Use of second pass classification
 - use of an additional classification step to remove confusion between selected crops
- C. Development of a procedure for complete area cluster definition
 - to determine how well transect fields and JES segments actually sample the range of spectral variability so as to improve sample allocation
- D. Development of improved estimation procedures
 - for more robust estimation in the presence of outliers
 - to take advantage of omission and commission error
 - to evaluate alternative county estimation procedures
- E. Further development of map-product capability

IV. Test site recommendation

- Central Valley
 - a. large proportion of major crops
 - b. doable in terms of implementation of transect training for next year
 - c. doable in terms of processing load
 - d. appropriate next level for efficient "large area" learning
 - e. high probability of success for stepping into operational implementation

- Lower priority
 - a. Central coast
 - for grains only
 - Salinas Valley complex, requires experiment of it's own
 - b. Imperial Valley
 - could add cotton, sugar beets, alfalfa, wheat
 - logical step for 1986 or shortly after

APPENDIX C

ROBUST REGRESSION ESTIMATION

The Fortran program ROBREG was developed to estimate robust estimates of regression parameters relating pixel counts to JES acreage. The macro TESTRR was used to test the performance of ROBREG with simulated data.

Model: Acreage in the i -th sample is

$$y_i = a + b \times x_i + r_i$$

where $x(i)$ is the number of pixels in the i -th segment. The $\{r_i\}$ are independently and identically distributed with mean 0 and standard deviation s .

Robust estimation: The mean acreage is estimated by

$$y = a + b \times X$$

where X is the mean segment acreage. By the ordinary least squares technique, a and b are chosen to minimize the sum of D_i^2 over all i , where D_i is the distance of the i -th point from the regression line:

$$D_i = y_i - a - b \times x_i.$$

The resulting estimates are sensitive to outliers, points with unusually large D_i . Robust estimates of a type called M -estimates replace the squaring function with a function which increases more slowly as D_i increases. An M estimate of a and b are the values of a and b that minimize the sum of

$$\text{RHO}[(y_i - a - b \times x_i)/S]$$

over all i . S is a scale factor which is included so that estimates will be scale invariant, i.e. yield the same estimates for y with different measurement units for y . Forms for RHO, algorithms for computation of a and b and formulae for estimating these estimates are presented in Chapter 5, Section 14 of Robust Statistical Procedures (Huber, 1977). Huber's "proposal 2" RHO is defined as:

$$\text{RHO}_z = z^2 \text{ if } z \leq C.$$

$$\text{RHO}_z = z^2 \text{ if } z > C.$$

For ROBREG, S was chosen to be s , the standard deviation of the r_i , and C was chosen to be 2.0 so that identification of outliers would be similar to the technique of evaluation of studentized residuals in outlier analysis following Belsey, Kuh, and Welsh and implemented in the EDITOR/PEDITOR system.

Algorithm: ROBREG implements Huber's "W algorithm" to compute estimates a and b by an iterative least squares procedure. Weighted least square estimates are computed by formulae given by Draper and Smith (Draper and Smith, 1966) with the IMSL subroutine RLFOR. Initial estimates are determined by ordinary least squares. The weights on subsequent iterations are determined by the RHO function and the residuals $\{r_i\}$. For the choice of RHO above, the weights are 1.0 if $r_i/s < 2.0$ and less than 1.0 if $r_i/s > 2.0$, specifically $1.0/(r_i/s)$. The estimates is updated on every iteration. The program stops when changes in values of a and b are small compared to estimates of their variances, which are computed using s. The ROBREG program reports on a, b and y, estimated standard errors for these estimates, and the number of "outliers" in the last iteration, and also the number of iterations involved in the computations.

Performance Tests: A test version of ROBREG included subroutines for generating simulated data, using IMSL subroutines for simulating numbers from uniform or normal distributions. A "population" of N pairs of x,y numbers were generated. The independent variable (pixel counts) were from $U(0,1)$, the uniform (rectangular) distribution with a range of values between 0 and 1. The y_i were determined by the formula:

$$y_i = (B0 + B1 \times x_i + r_i) \times 200.$$

with r_i from $N(0,S1^2)$ or $N(0,S2^2)$, normal distributions with two different standard deviations. The y_i thus simulated acreages which fit the regression model. The occurrence of outliers was simulated by the mixture of distributions for the residuals $\{r_i\}$. Values of y_i computed with the second normal distribution tended to be outliers because $S2$ was always specified as much larger than $S1$. The frequency of outliers was determined by P, the proportion of the r_i that were simulated samples from $N(0,S2^2)$.

The program generated 100 samples with NR pairs of numbers drawn randomly from the population. Estimates were computed from each sample. The initial estimates for a, b, and s were either ordinary least squares estimates or weighted least squares, with weights determined by the deviation of y_i from $0.8 \times x_i$.

The macro TESTRR ran ROBREG with a particular set of parameters for the simulated test data sets and initial definition of outliers:

- N,NR - population size, size for each of the 100 samples
- B0,B1 - true regression intercept(/200.), slope
- S1,S2,P - parameters of the distribution of the residuals; standard deviations and proportions of mixed normal distributions $N(0,S1^2)$ and $N(0,S2^2)$ with proportions (1-P) and P.
- NONEG - set equal to 1 to force minimum y value to 0.0
- CSTART - initial definition of outliers is $y(i) - 0.8 \times x(i) > CSTART$, "infinite" means ordinary least squares initialization.

The output from ROBREG was analyzed in TESTRR by using the Minitab statistical program package. The mean and variance of y were computed over all samples to determine the accuracy and precision of the estimates. The mean of the sample estimate of the standard deviation of y was computed so that it could be compared to that of y computed over all samples.

A few test runs were made in July, 1985 (Table C1). The input parameters were set as follows:

N,NR (population size, sample size) - 200,20
 B0,B1 (intercept,slope) - (0.,.8) or (.1,.8)
 S1,S2,P (residuals) - (.1,.4,0.0),(.01,.4,.2),or(.1,.4,.2)
 NONEG (minimum values of $y(i)$ set to zero) - 0(no) or 1(yes)
 CSTART - 40.0, 200.0, or infinite(least squares)

TABLE C1.- TESTRR RUNS IN JULY, 1985

Test name	B0	B1	S1	S1	P	NONEG	CSTART
TEST01	0.0	0.8	0.1	0.4	0.0	0	200.
TEST02	"	"	"	"	"	1	200.
TEST03	"	"	"	"	"	0	40.
TEST04	"	"	"	"	"	1	40.
TEST05	0.0	0.8	0.1	0.4	0.0	0	200.
TEST06	"	"	"	"	"	1	200.
TEST07	"	"	"	"	"	0	40.
TEST08	"	"	"	"	"	1	40.
TEST2	0.1	0.8	0.1	0.4	0.2	0	infinite
TEST3	"	"	"	"	"	1	"
TEST4	"	"	"	"	"	1	"

The performance results are shown in Table C2. For each test, the following statistics are presented:

- Y - population mean value of y_i over all N values.
- INITPRED(m/s) - mean/standard deviation of initial(first iteration) y computed over 100 samples.
- RRPRED(m/s) - mean/standard deviation of robust (last iteration) y computed over 100 samples.

Comparison of statistics for RRPRED for test runs that were identical except for CSTART, for example TEST05 and TEST08, show that robust estimates are not effected by the initial iteration. Comparison of the statistics for INITPRED and RRPRED show that mean values for initial and final estimates are similar. Results for TEST2, TEST3, and TEST4 show that robust estimates were slightly less variable overall.

TABLE C2.- ACCURACY OF TESTRR ESTIMATES

Test name	Y	INITPRED(m/s)	RRPRED(m/s)
TEST01	82.57	82.81/5.00	82.86/5.04
TEST02	82.99	83.07/4.83	83.12/4.86
TEST03	82.57	82.79/4.97	82.86/5.04
TEST04	82.99	83.07/4.80	83.12/4.86
TEST05	82.97	82.99/7.85	83.15/5.64
TEST06	84.83	84.73/6.57	83.19/4.92
TEST07	84.83	83.11/4.23	83.18/4.92
TEST08	82.97	83.06/4.46	83.18/5.65
TEST2	-	104.00/9.16	104.63/8.24
TEST3	-	104.97/8.33	104.69/7.89
TEST4	-	111.52/13.44	110.83/13.21

Table C3 shows results for testing the accuracy of the sample estimate for the standard deviation of y and also displays the mean number of outliers per sample.

TABLE C3.- ACCURACY OF ESTIMATED STANDARD DEVIATIONS

Test name	s(INITPRED) m(est)/s	s(RRPRED) m(est)/s	#outliers m
TEST01	4.52/5.00	4.65/5.03	0.35
TEST02	4.33/4.83	4.46/4.86	0.40
TEST03	4.77/4.97	4.58/5.04	0.34
TEST04	4.55/4.80	4.40/4.86	0.38
TEST05	7.29/7.85	3.92/5.64	2.16
TEST06	5.94/6.57	3.13/4.91	2.23
TEST07	5.22/4.23	3.12/4.92	2.25
TEST08	6.07/4.46	3.91/5.66	2.15
TEST2	8.48/9.16	7.44/8.24	-
TEST3	7.65/8.33	7.04/7.89	-
TEST4	12.2/13.4	12.00/13.21	-

Estimated standard deviations of robust estimates were lower than sample standard deviations, especially in test runs with more frequent outliers. Estimated standard deviations of weighted least squares estimates tended to be lower than sample standard deviations when CSTART was 40.

These preliminary tests indicated that robust estimates tend to downweight outlier values while maintaining overall precision and accuracy of estimates. Some further work is needed to develop accurate estimates of the standard deviation of the robust estimate.

REFERENCES

- Belsey, D.A., E. Kuh, and R.E. Welsh, Regression Diagnostics, John Wiley and Sons, 1980.
- Draper, N.R. and H. Smith, Applied Regression Analysis, John Wiley and Sons, 1966.
- Huber, P.J., Robust Statistical Procedures, Society for Industrial and Applied Mathematics, 1977.
- International Mathematics and Statistics Library, Reference Manual, IMSL, Inc., 1982.
- Ryan, T.A., B.L. Joiner, and B.F. Ryan, MINITAB Student Handbook, Duxbury Press, 1976.

REFERENCES

1. Landsat 4 Data Users Handbook, U.S. Geological Survey, National Oceanic and Atmospheric Administration, 1984. (For sale by the Distribution Branch, Text Products Section, U.S. Geological Survey, 604 S. Pickett St., Alexandria, VA 22304.)
2. Thermodsgard, J. M.; and DeVries, D. J.: Geodetic Accuracy of Landsat-4 Multiscanner (MSS) Data. Landsat-4 Science Characterization Early Results, vol. I, Feb. 1983, pp. 133-142.
3. "Scope and Methods of the Statistical Reporting Service." U.S. Dept. of Agr., SRS, Misc. Publication No. 1308, Sept. 1983.
4. Cochran, W. G.: Sampling Techniques, Third Edition. John Wiley and Sons, Inc., New York, 1977.
5. Hanuschak, G. A.; Craig, M.; Ozga, M.; Luebbe, R.; Cook, P.; Kleweno, D.; and Miller, C.: Obtaining Timely Crop Area Estimates Using Ground-Gathered and Landsat Data. U.S. Dept. of Agr., ESCS Technical Bulletin No. 1609, Aug. 1979.
6. Ozga, M. L.: Experience with the Use of Supercomputers to Process Landsat Data. International Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Indiana, June 1984.
7. Ozga, M.; Donovan, W. E.; and Gleason, C.: An Interactive System for Agricultural Estimates Using Landsat Data. International Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Indiana, June 1977.
8. Ray, R. M. III; Thomas, J. D.; Donovan, W. E.; and Swain, P. H.: Implementation of ILLIAC IV Algorithms for Multispectral Image Interpretation, Final Report, CAC Document No. 112, Center for Advanced Computation, University of Illinois at Urbana-Champaign, Illinois, June 1974.
9. Anderson, T. W.: An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, Inc., New York, 1958.
10. Lennington, R. K.; and M. E. Rossback: CLASSY - An Adaptive Maximum Likelihood Clustering Algorithm. Ninth Annual Meeting of the Classification Society, Clemson University, South Carolina, May 1978.
11. Swain, P. H.: Pattern Recognition: A Basis for Remote Sensing Analysis. LARS Information Note 111572, Sept. 1973.

12. Holko, M. J.; and Signman, R. S.: The Role of Landsat Data in Improving U.S. Crop Statistics. Eighteenth International Symposium on Remote Sensing of Environment, Paris, France, Oct. 1984.
13. Battese, G. E.; and Fuller, W. A.: Predictions of County Crop Areas Using Survey and Satellite Data. Survey Section Proceedings, American Statistical Association Annual Meeting, Detroit, Michigan, 1981.
14. Chikara, R. S.: Discussion. Survey Section Proceedings, American Statistical Association Annual Meeting, Detroit, Michigan, 1981.
15. Sigman, R.; and Walker, G.: The Use of Landsat for County Estimates of Crop Areas: Evaluation of the Huddleston-Ray and Battese-Fuller Estimators. U.S. Dept. of Agr., SRS, Sept. 1982.
16. Wall, S. L.; Brown, C. E.; Erickson, M.; Grigg, C.A.; and Thomas, R. W.: Irrigated Lands Assessment for Water Management. Annual Report, NASA Cooperative Agreement NCC2-54, Space Sciences laboratory, University of California, Berkeley, Apr. 1981.
17. Wall, S. L.; Thomas, R. W.; Brown, C. E.; and Bauer, E. H.: Landsat-Based Inventory System for Agriculture in California. Remote Sensing of Environment, vol. 14, 1984, pp. 267-278.
18. Huber, P. J.: Robust Statistical Procedures. Society for Industrial and Applied Mathematics, Philadelphia, 1977.
19. Sheffner, E. J.: Crop Area Estimation in Idaho using EDITOR. NASA CR 177327, Dec. 1984.
20. Sheffner, E. J.; Hlavka, C. A.; and Bauer, E. M.: Two Techniques for Mapping and Area Estimation of Small Grains in California Using Landsat Digital Data. Seventeenth International Symposium on Remote Sensing of Environment, Michigan, May 1983.
21. Kauth, R. J.; and Thomas, G. S.: The Tasseled Capp--A Graphic Description of the Spectoral-Temporal Development of Agricultural Crops as seen by Landsat. Proceedings of Machine Processing of Remotely Sensed Data, 1976.
22. Holko, Martin J.: 1982 Results for the California Cooperative Remote Sensing Project. U.S. Dept. of Agr., SRS, Mar. 1984.
23. Erickson, W. K.; Hofman, L. B.; and Donovan, W. E.: The Microprocessor Workstation: An Alternate Hardware Architecture For Remotely Sensed Image Analysis. Eighth William T. Pecora Symposium, South Dakota, 1983.
24. Quarterman, J. S.; and Hoskins, J. C.: Notable Computer Networks. Communications of the ACM, vol. 29, no. 10, Oct. 1986, pp. 932-971.

25. Fesco, R. V. J.: The New California Area Frame. U.S. Dept. of Agr., SRS, SRS-22, Sept. 1981.
26. Swain, P. A.; and Davis, S. M.: Remote Sensing: The Quantitative Approach. McGraw-Hill, Inc., New York, 1978.
27. Larson, L. O.; and Williams, L. S.: 1984 California Fruit and Nut Acreage. California Crop and Livestock Reporting Service, Sacramento, California, July 1985.
28. Tippett, J.; Randenz, R.; Kleweno, D.; and Hintzman, K.: California Grape Acreage 1985, California Crop and Livestock Reporting Service, Sacramento, California, May 1986.
29. Imhoff, M. L.; and Alford, W. L.: Geometric Accuracy Assessment of Landsat-4 Multispectral Scanner(MSS) Data. Landsat-4 Science Characterizations Early Results, vol. I, Feb. 1983, pp. 143-158.
30. Welch, R.; and Usery, E. L.: Geometric Accuracy of Landsat-4 MSS Image Data. Landsat-4 Science Characterization Early Results, vol. I, Feb. 1983, pp. 123-132.
31. Wrigley, R. C.: Personal communication.



Report Documentation Page

1. Report No. NASA TM 100073		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle The California Cooperative Remote Sensing Project: Final Report			5. Report Date July 1988		
			6. Performing Organization Code		
7. Author(s) Christine A. Hlavka and Edwin J. Sheffner (TGS Technology, Inc., Moffett Field, CA)			8. Performing Organization Report No. A 88105		
			10. Work Unit No. 146-66-01		
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035			11. Contract or Grant No.		
			13. Type of Report and Period Covered Technical Memorandum		
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001			14. Sponsoring Agency Code		
			15. Supplementary Notes Point of Contact: Christine A. Hlavka, Ames Research Center, MS 242-4 Moffett Field, CA 94035 (415) 694-6060 or FTS 464-6060 35mm slide in envelope attached to inside back cover of this report.		
16. Abstract The U.S. Department of Agriculture (USDA), the California Department of Water Resources (CDWR), the Remote Sensing Research Program of the University of California (UCB) and the National Aeronautics and Space Administration (NASA) completed a 4-yr cooperative project on the use of remote sensing in monitoring California agriculture. This report is a summary of the project and the final report of NASA's contribution to it. The cooperators developed procedures that combined the use of Landsat Multispectral Scanner imagery and digital data with ground survey data for area estimation and mapping of the major crops in California. An inventory of the Central Valley was conducted as an operational test of the procedures. The satellite and survey data were acquired by USDA and UCB and processed by CDWR and NASA. The inventory was completed on schedule--demonstrating the plausibility of the approach, although further development of the data processing system is necessary before it can be used efficiently in an operational environment.					
17. Key Words (Suggested by Author(s)) Remote sensing Landsat Digital data processing June enumerative survey EDITOR			18. Distribution Statement Unclassified-Unlimited Subject Category - 43		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of pages 98	22. Price A05