Reprinted from

Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

A LANDSAT-BASED INVENTORY PROCEDURE FOR AGRICULTURE IN CALIFORNIA

S.L. WALL, R.W. THOMAS, C.E. BROWN

Remote Sensing Research Program
Berkeley, California

E.H. BAUER

National Aeronautics and Space
Administration/Ames Research Center
Moffett Field, California

I. ABSTRACT

California's agriculture is one of the most diversified in the world. Successful production of the 200 commercially grown crops is based on the irrigation of 9.9 million acres of land. Since 1975 the California Department of Water Resources has been cooperating with the National Aeronautics and Space Administration and the University of California in addressing the applicability of Landsat imagery and digital data to aid in the mapping and estimation of irrigated land and specific crop types. An inventory of irrigated acreage over the entire state was conducted in 1979. The inventory system utilized stratified regression estimation based on multitemporal Landsat and single date ground measurements of irrigated proportion. Approximately 9.86 million acres were irrigated statewide, with a relative standard error of ± 1.17% at the 99% confidence level (absolute S.E. = ± 0.89%). Development of a digital version of the inventory system described above is nearing completion. Two major sub-tasks were addressed: (1) Landsat multitemporal registration; and, (2) classification/labeling procedures. To identify irrigated land, a simple vegetation indicator, the ratio of Landsat MSS band 7 to MSS band 5 was used. Ground sample units were linked with the Landsat-derived measurement of irrigated land to produce estimates over a 1.98 million acre test site. The third phase of the project is the development of a baseline, computer-assisted mapping and area-estimation system for crop type. Work is centering on: (1) the development of efficient multicrop Landsat classification procedures; and (2) development of simulation techniques that will allow identification of crop area estimation procedures using registered Landsat and ancillary data.[1]

II. INTRODUCTION

Endowed with large areas of level terrain, productive soils, a well developed irrigation system and a long growing season, California has developed one of the most productive and varied agricultural industries in the world. Sixty-nine major crops are grown on a large commercial scale in a state that recognizes the cultivation of more than 200 crops. Despite the well known concentrations of population in the Los Angeles Basin and San Francisco Bay Area, most counties in California are still agriculturally oriented. Based on this orientation, California continues as the leading farm state grossing $13.7 billion in agricultural marketings in 1980.[2]

Considering the importance of agriculture to the economy of California as well as its contribution to the national food supply and foreign export, J.B. Kendrick, Jr., University of California Vice President for Agriculture and University Services authorized a task force to study agricultural policy challenges for California in the 1980's. One of the task force's major topics dealt with agriculture as a competitor for resources. They state, "During the 1980's, California farmers can expect increasing competition, both within agriculture and between agriculture and nonfarm sectors, for basic resources--including water, land, energy and labor. Of these essential inputs, water will be most scarce during the next decade (if not longer) and is most likely to limit food and fiber production in the state.[3] Currently the California Department of Water Resources (DWR) estimates that 9.9 million acres are irrigated at least once during the growing season. This water requirement is met primarily from surface sources supplemented by groundwater extraction, and by the use of large-scale water transport projects. Agriculture is the prime recipient of the available water, utilizing about 85% of the supply.[4]

Given the myriad crops grown, their variable need for water, and agriculture's overwhelming total demand for water, the Department of Water Resources has long recognized the need for specific land use data as an input to state water planning. Since the late 1940's DWR has been performing a continuing survey to monitor land use changes over the state. Because of manpower and budgetary constraints, approximately one-seventh of the state is surveyed during a given year. The Land Use Survey is compiled through the interpretation of current 35mm color aerial photography supplemented with field inspections. From the data collected, DWR is able to generate 7-½ minute quadrangle maps showing the land use classification to cover type, including crop identification, and the area of irrigated land. Since each land use is associated with a specific water demand, total water consumption forecasts can be made for DWR planning purposes.

Given the limitations of an area-limited, single-date survey system, the Department of Water Resources has been cooperating with the National Aeronautics and Space Administration (NASA) and the University of California in addressing the applicability of Landsat imagery and digital data to aid water management decision-making. These decisions are based partly on information that a Landsat-aided inventory system should be able to provide: (1) the proportion of the total area of the State that is irrigated in a given year, and (2) the distribution and area of specific crop types. Emphasis through the Spring of 1982 has been on the development of an inventory system to accurately estimate the proportion of land irrigated at least once during the growing season. Development has included testing of both human analysis of Landsat color composite imagery and computer-assisted analysis of Landsat digital data. Currently, the development of techniques for estimating and mapping specific crop types is receiving increasing effort.

III. STATEWIDE INVENTORY OF IRRIGATED LAND

Prior to the statewide inventory of irrigated land undertaken in 1979, two pilot studies had been conducted on limited test sites within the state. Based on the findings of these tests, and upon consultation with DWR, inventory objectives of the statewide test were outlined. A survey design was defined for the purposes of: (1) estimating the proportion of land irrigated at least once during the growing season to within ± 5% relative error, 95 times out of 100 for each of ten hydrologic basins within California, and to within ± 3%, 99 times out of 100 for the state; (2) summarizing the irrigated acreage at three levels - hydrologic basin, county (58 counties) and statewide; and (3) operationally producing estimates within one year at a cost of one to two cents ($1979) per agricultural acre. In addition, the survey system was to utilize only techniques that could be implemented with inhouse capabilities by DWR.

A. DESIGN AND SAMPLE ALLOCATION

To meet the low error goals specified by DWR, a stratified regression sample design was developed for use with irrigated proportion measurements from Landsat imagery and ground survey. This design utilized relatively inexpensive manual interpretation of Landsat imagery to obtain irrigated proportion observations for all Phase I sample units. A sample of these units was then selected for more expensive, though more accurate, ground field delineation and labeling as to irrigation practice. Random selection of ground units (Phase II) and subsequent estimation of irrigated land proceeded on a stratified basis to minimize basin-wide sampling error.

To provide a list of potential ground sample segments, the agricultural area within each hydrologic basin was divided into a grid of rectangular segments and assigned to a sampling stratum. These strata, which also assisted in controlling Landsat measurement error, were produced by interpreting multi-year 1:1,000,000 Landsat color composite transparencies for general agricultural land use and field size. Seven strata were identified: (1) dryland farming; (2) field crops with fields generally smaller in area than 40 acres and proportion of irrigation unknown; (3) field crops with fields generally smaller in area than 40 acres with known high proportion irrigated; (4) field crops with fields greater than or equal to 40 acres in area; (5) orchard and vineyard areas with fields generally smaller than 40 acres; (6) orchard and vineyard areas with fields 40 acres or larger; and (7) unusual agricultural areas (for example, dispersed agriculture in foothill areas). When this stratification was intersected with the sample grid, a total of 6004 sample segments were created statewide.

Precision control and associated ground sample allocation was performed independently by hydrologic basin. Within each basin, sample segments were allocated to each land use stratum according to Neyman optimal rules through the use of a non-linear programming algorithm. This algorithm required stratum-specific estimates for Landsat-to-ground irrigated proportion measurement correlation, irrigated proportion variance, and ground measurement cost in order to compute the sample size for each stratum. From the total population, 606 or 10.1% of the possible segments were selected with equal probability and without replacement for ground data collection by DWR (Figure 1). Each selected segment (average size = 2151 acres), was delineated on a U.S. Geological Survey 7.5 minute quadrangle for ground survey.

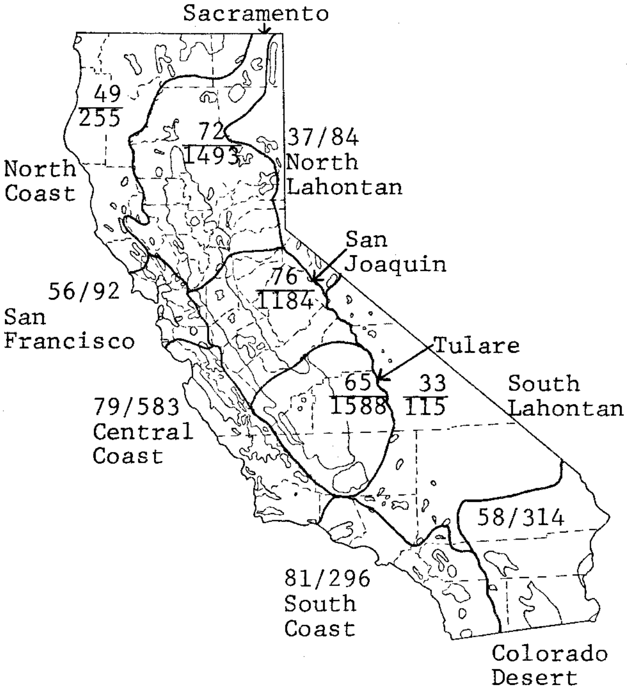

Figure 1. Population and Sample Allocation by Hydrologic Basin. The agricultural area of California is shown in gray. Hydrologic Basin boundaries are indicated by the heavy black lines; county boundaries are dashed lines. The numbers indicate the number of ground sample segments allocated/the total population for each basin.

B. LANDSAT MEASUREMENT

California's long growing season (most of the agricultural area exceeds 280 days without killing frost) and varied crop production necessitates utilizing multiple dates of Landsat to insure the identification of irrigated land. Fortuitiously, the Mediterranean and desert climates that dominate the growing areas provide a nearly cloud-free summer. A simple interpretation procedure was developed requiring the use of three seasonally-distributed Landsat acquisitions. The cornerstone of the system is: if a crop is growing in mid-summer in California or other semi-arid to arid areas it must be irrigated. Using this fact, a summer Landsat acquisition was used as the base date to establish the initial extent of irrigation. In addition a late spring/early summer date was used to determine the acreage of crops such as small grains (mostly irrigated in California) that are harvested early. The final interpretation was done on a fall scene to locate any other areas that were missed previously (this is especially important in areas of multiple cropping where the growth, harvesting and replanting of crops can be very rapid and may occur entirely between the first and second dates used). Interpretation was done on 1:150,000 scale prints enlarged from color composite transparencies. Enlargement and interpretation were done on a county basis.

Considering the low error goals required by DWR, maximizing Landsat-to-ground correlation at the Landsat phase was critical. Ancillary information consisting of county agricultural statistics, current farm newspaper and magazine reports, weekly crop-weather reports and antecedent DWR land use maps and statistics were provided to the analyst as source material upon which he could develop reasonable expectations as to the probability of irrigation. Using all or a subset of the available data, the analyst built a mental model of what he expected to see on the dates of Landsat imagery provided for each county.

Image characteristics traditionally used in manual photographic interpretation were used in the analyses of the Landsat imagery. For the majority of the interpretation the most critical characteristics were: (1) pattern (is this an area of agricultural fields?); and (2) color (is this field a color expected for an irrigated field on the date being analyzed?). Other characteristics the analyst relied on were texture shape of fields and the location of fields. These last three characteristics were particularly important when analyzing mountain areas, riparian/agriculture interfaces, the fringes of well developed agriculture and areas of dispersed cultivation.

C. RESULTS

The Landsat and ground measurements were reduced to proportion data and linked, using regression estimators, to compute the estimates of irrigated land and their associated errors. This procedure produced a statewide estimate of 9.86 million acres irrigated at least once during the 1979 growing season. Acreage tabulations of irrigated land provided by DWR showed that the Landsat-based estimate differed by less than 0.4% at the state level. Table 1 shows the stratified estimates, absolute standard errors, and relative standard errors for each hydrologic basin and the state.

Viewing the error estimates as a whole, the original goal of ± 5% 95 times out of 100 was met in all ten basins on the basis of absolute standard error and in six out of ten basins based on relative error (relative error was within 6.3 % 95 times out of 100 in three of the remaining four basins). Given the fact that this was a first-time inventory over most of the state, the estimated error performance obtained in 1979 should be considered quite good. Information gained on sample variance, correlation and strata performance by basin will allow improved error performance in subsequent inventories and reduce the cost to achieve given levels of performance.

IV. COMPUTER-ASSISTED ESTIMATION AND MAPPING OF IRRIGATED LAND

A logical outgrowth of the manual analysis system just described is the development of an estimation and mapping technique that utilizes computer assisted analysis of multitemporal Landsat digital data. Ideally, the accuracy goals outlined in Section III would be met using a computer-based system and the Department would gain the additional benefit of having the irrigated area data in a form suitable for geographic information system manipulation. At present three major sub-tasks are being studied: (1) the precise registration of multitemporal Landsat date-to-date and Landsat-to-map; (2) the selection of the most appropriate classification/labeling procedure; and (3) the evaluation of potential sampling designs.

A. REGISTRATION OF MULTITEMPORAL LANDSAT DIGITAL DATA

Precise registration of Landsat digital data for a number of dates, as well as mosaicking adjoining path and row acquisitions, is necessary for optimizing classification and outputting results. Figure 2 presents the procedure used to register entire Landsat scenes of differing dates one to another. Here all "secondary" scenes of a common Path/Row location are registered to one "primary" scene. That is, all pixels of the primary scene are kept in the original raw data location, and pixels from differing or secondary overpass dates are manipulated to overlay those of the primary scene.

Table 1. Results of the 1979 Statewide Inventory of Irrigated Land

Basin	Stratified Estimate (% of area irrigated in the sample frame)	Absolute Standard Error at 95% C.L.	Relative Standard Error at 95% C.L.
Central Coast	31.91	1.84	5.77
Colorado Desert	82.15	1.40	1.70
North Coast	53.52	2.04	3.81
North Lahontan	58.73	2.68	4.56
Sacramento	65.38	1.80	2.75
San Francisco	21.85	1.22	5.56
San Joaquin	74.78	2.55	3.41
South Coast	45.79	2.88	6.28
South Lahontan	27.38*	3.81*	13.91*
Tulare	82.04	2.00	2.44
STATE	67.09	95: 0.89 99: 1.17	95: 1.32 99: 1.74

*Unstratified estimate

Any computer compatible tape (CCT) may be chosen as the primary scene.

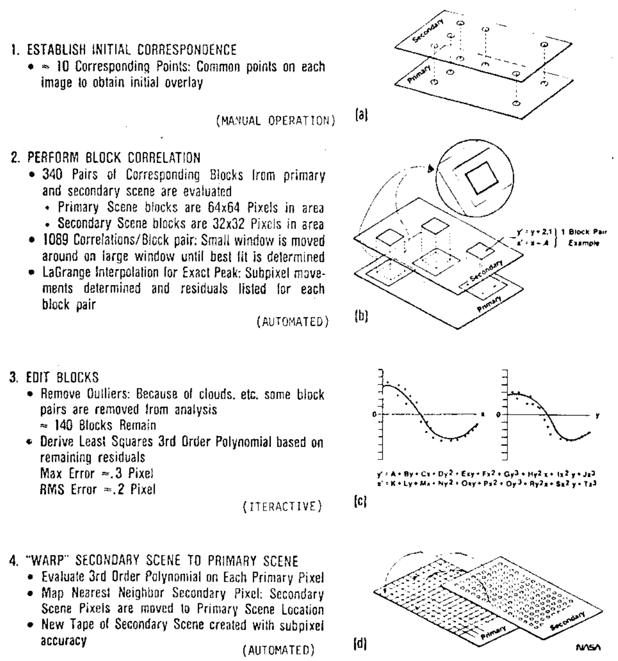

Figure 2. Scene-to-Scene Registration

In Step 1, initial correspondence between a primary and secondary scene is obtained by digitizing the corner tick marks and about ten identifiable matching points scattered over 1:1,000,000 scale Landsat images. Using this initial correspondence, 340 pairs of approximately overlaid blocks spaced uniformly throughout the scene are extracted (primary scene blocks are 32 x 32). Block correlation as depicted in Step 2 is accomplished by performing the same operation on each of the 1089 possible overlay positions on the larger primary block, calculating the gradient correlation at each position, and performing a LaGrange interpolation at the maxima of this correlation surface. The X and Y shifts giving maximum correlation represent the secondary to primary scene correspondence for this block pari to subpixel accuracy. Block editing in Step 3 takes the resulting 340 corresponding point pairs and calculates the best third degree polynomial (in a least squares sense) fitting these pairs. Because block pairs do not always correlate well due to cloud cover, land use changes, snow cover or data errors, editing to remove outliers is performed until a maximum root mean square error of 0.2 pixels is obtained (in a post-1979 scene experience, usually at least 140 block pairs remained after this editing process). In Step 4 the secondary image is warped to the primary image by evaluating the least-squares polynomial obtained in Step 3 at each primary pixel. This gives the corresponding secondary scene location, which is converted to actual secondary scene row and column addresses via a nearest neighbor mapping. A new CCT is thus created with sub-pixel accuracy. This new CCT is in Landsat format and not yet registered to a map base; however, when the primary scene is registered to the map base, all the overlaid secondary scenes also become registered.

B. CLASSIFICATION

One of the test sites selected for developing and testing digital classification of irrigated land is a 1.98 million acre area in the Sacramento Valley. In addition to three dates of registered Landsat, a multi-layered data base was created that contained county boundaries for summarization, the land use stratification described in Section III-A, sample segment boundaries, and sample segment ground-checked field information.

A simple vegetation indicator, the ratio of MSS Band 7 to MSS Band 5, was used to identify irrigated land. Since actively growing vegetation generally has a higher 7/5 ratio than other cover classes, agricultural land with its healthy vegetation and high percent canopy cover, should have a higher 7/5 ratio than native vegetation or fallow land. Three dates of Landsat 7/5 ratioed data mimicking the acquisitions used in the manual procedure, were selected to identify irrigated land. For each date a threshold 7/5 value was determined to separate irrigated agricultural land from non-irrigated land.

Two methods of determining the threshold value are being tested: (1) setting the value by 30-minute block; and (2) setting the value by reference to the seven land use strata described earlier. To set the value, each date is displayed on an interactive system and pixels with values below a specified value are deleted from the data set. This 7/5 value is adjusted visually by reference to Landsat color composite imagery and a sample of ground data until only actively growing cropland remains. Each pixel is then labeled either irrigated or not irrigated on each of the dates. By combining the three dates of labeled Landsat, a map is produced showing land irrigated at least once during the growing season as well as the sequence of irrigation.

C. RESULTS OF ESTIMATION AND MAPPING

A test of the first method for setting the threshold value has been completed and is shown in Table 2. Using the data base with its matched ground and Landsat-classified sample segments, estimates were calculated using a weighted regression estimator. For purposes of evaluation, the results of the block-setting procedure are presented by stratum. Estimates were calculated for each stratum by comparing the proportion mapped as irrigated on the ground sample segments with the proportion labeled as irrigated using the 7/5 ratioed digital Landsat data. A regression equation for each stratum was then computed and was used as a "correction" factor for the entire population within each stratum.

Table 2. Proportion Irrigated by Land Use Stratum Using a Weighted Regression Estimator

Stratum	Proportion Irrigated	Standard Error	Population Size	Sample Size	r^2
1	.35145	.11178	139	7	.47740
2	.59923	.00113	45	4	.95692
4	.79651	.03391	587	29	.85920
5	.87183	.04997	102	4	.99121
6	.75831	.00590	9	4	.99905
7	.24812	.04460	17	4	.69372

By combining the stratum estimates, summary statistics for the entire basin area were computed. This system estimated 72.4% of the population to be irrigated with a relative standard error of ± 7.98% at the 95% confidence level. As expected digital analysis works well in areas that are mostly agriculture. In dryland areas (Stratum 1) and in areas of dispersed agriculture (Stratum 7), the simple 7/5 discriminant when set by 30-minute block was not as effective. Currently we are testing the second method suggested and setting the irrigation threshold value by stratum. Preliminary results from this effort are very encouraging.

The Department of Water Resources has specific needs for accurate in-place mapping of irrigated fields in addition to acreage estimation. Map accuracy depends on minimizing: (1) errors of omission (missing land that was actually irrigated); and (2) errors of commission (classifying land as irrigated when it really was not). Thirty-two 7.5-minute quadrangles were selected in a checkerboard pattern for error analysis. On each quad, 24 points were chosen and summarized by irrigation label, ground label, crop type, and field size. Overall map accuracy was quite good: Percent Correct = 94.0; Percent Omission = 7.4; and Percent Commission = 6.3.

V. CROP TYPE ANALYSIS USING LANDSAT DIGITAL DATA

As part of evolution of techniques to enable DWR to use Landsat effectively, we are addressing two major multicrop issues: (1) the development of an efficient classification procedure; and (2) the development of simulation techniques to allow identification of cost-effective crop area estimation procedures using registered Landsat and ancillary data. Initial work is being done on sixty-four 7.5-minute quadrangles in the Sacramento Valley. This area was selected because the crop mix is diverse, the mix is representative of much of the agriculture in northern California; and DWR had collected detailed ground data over the entire site.

The crops selected for analysis are beans, corn, orchard, pasture, rice, small grains, sorghum, sugar beets, and tomatoes. General and year-specific crop calendars were generated for each of these crops using Weekly Crop and Weather Reports. These calendars show crop phenology and pertinent cultivation practices and were useful in selecting appropriate Landsat acquistions. The initial set of acquisitions selected for analysis were: May 4, May 30, June 26, August 28, and October 3. Each of these scenes was registered to the 7.5-minute quadrangle base.

A. CLASSIFICATION PROCEDURE

For purposes of classification we wanted to test: (1) the utility of simply and inexpensively computed bands to identify crop type; and (2) the validity of using a stratification scheme based on the sequence of irrigation to reduce potential classification confusion among crops.

Ratioed spectral bands were created for each of the five dates. MSS Band 7 to

MSS Band 5 was created to measure the ratio of reflected infrared energy to reflected red energy. A ratio band of MSS Band 5 to MSS Band 4 was created as an approximate measure of vegetation senescence indicating the end of a plant's growth cycle. Finally a Euclidean albedo band,

$$EB = ((MSS\ 4)^2 + (MSS\ 5)^2 + (MSS\ 6)^2 + (MSS\ 7)^2)^{\frac{1}{2}} ,$$

was created to measure the brightness of the vegetation.

Spectral statistics were obtained within the test area for the selected crops. The mean value of the three bands, together with the standard deviation and value range by date were determined on a field basis. The resulting data were then plotted against time for each crop and were compared for crop separability. A crop matrix was then prepared showing the best bands and dates for crop differentiation.

B. STRATIFICATION

The test area was stratified into general crop groupings based on irrigation timing. Using the 7/5 ratio technique for identifying irrigated land described in Section IV-B, a map was created where each pixel was labeled as irrigated or not irrigated on each of the three dates used. This map has eight irrigation classes ranging from irrigated on none of the dates to irrigated on all three dates. Because of crop phenology and other crop calendar events, these irrigation classes tend to separate general crop groups (Table 3).

Table 3. Irrigation Patterns and Associated Crop Groups

Irrigation Pattern	Crop Groups
not irrigated	non-vegetated, native vegetation, non-irrigated small grains
May 4 only	small grains, truck crops
August 28 only	rice, vegetable crops
October 3 only	orchard
May 4 & August 28	rice, field crops, orchard
May 4 & October 3	pasture, orchard
August 28 & Oct 3	field crops, rice
All 3 dates	pasture, orchard

Using the crop separability matrix, bands were chosen for each irrigation stratum to separate the crop types found within that stratum. For a 30-minute block within the test site, unsupervised clustering and classification was performed within each stratum using these input bands. This clustering and classification used all of the digital values from the selected Landsat bands and combined them into spectrally similar groups which could then be labeled as to crop type.

Ground data maps for the 30-minute block were digitized and registered to the Landsat data. The classification output and the ground data were intersected to provide the crop mix for each spectral class. Each spectral class was given the label of the ground data crop group with the highest proportion within that class.

Preliminary evaluation indicates that certain crops and crop groups are discernable. Small grains and rice, with their highly individualized temporal and spectral patterns make them easily identifiable (percent correct = 95% and 90%, respectively). Orchard as a crop group is quite easily distinguishable (81% correct); however, separating the various fruit and nut varieties is not feasible with these techniques. Separating individual field crops requires additional research (percent correct = 5% - 56%). The crop calendars of these crops are quite similar, with planting and harvesting times overlapping. Also, many of these crops are quite similar spectrally causing them to be easily confused.

VII. REFERENCES

1. Wall, S.L., R.W.Thomas, C.E.Brown, M.Eriksson, J.Baggett, E.H.Bauer and C.A.Grigg. 1981. Irrigated Lands Assessment for Water Management - Applications Pilot Test. Annual Report. Space Sciences Laboratory Series 22, Issue 23. University of California, Berkeley.
2. Department of Food and Agriculture, State of California. 1981. California Agriculture. U.S. Department of Agriculture, Sacramento, CA.
3. University of California. 1978. Agricultural Policy Challenges for California in the 1980's. Agricultural Sciences Publications, Special Publication No. 3250, Richmond, CA.
4. Department of Water Resources, State of California. 1974. The California Water Plan - Outlook in 1974. Department of Water Resources Bulletin No. 160-74, Sacramento, CA.

Ethel H. Bauer. Ethel Bauer is presently with NASA/Ames Research Center as Ames Project Manager for the California Irrigated Landsat Assessment for Water Management Applications Pilot Test project. Ethel has been with NASA for the past 20 years, working first at the George C. Marshall Space Flight Center in Huntsville, Alabama. After graduation from Huntingdon College and Vanderbilt University with a B.A. in Business Administration and a B.S. in Mathematics, she worked in the areas of guidance, performance and dynamics for the Apollo and Sylab missions, and in system studies for the Shuttle Program. She transferred to Ames Research Center in California in 1974 and has been working in the remote sensing field since that time, primarily in the areas of agriculture and water related projects.

Catherine E. Brown. Ms. Brown received a B.A. in Geography from the University of California, Berkeley and an M.A. in Geography from California State University, Hayward. She is a member of the American Society of Photogrammetry, the Association of Pacific Coast Geographers, the Association of American Geographers, and is currently Secretary-Treasurer of the California Remote Sensing Council. Ms. Brown has extensive experience in the analysis of remotely sensed data for the assessment of agricultural resources. Her research activities have dealt mainly with the use of Landsat data for the estimation and mapping of agricultural resources and have included both manual and computer-assisted analysis.

Randall W. Thomas. Mr. Thomas is currently a staff scientist with the Remote Sensing Research Program, University of California at Berkeley, and is also a Ph.D. candidate in the University's Department of Forestry and Conservation. Since joining the Program in 1973, Mr. Thomas has participated in a variety of wildland and agricultural inventory development studies. Recent work has emphasized the design and evaluation aspects of irrigated lands inventory in California.

Sharon L. Wall. Ms. Wall received a B.A. in Geography from the University of California, Davis. Since that time she has been employed at the Remote Sensing Research Program, University of California at Berkeley involved in research projects utilizing satellite and aerial data in agriculture, forestry, range and recreation studies. Currently, she is Project Manager of the Irrigated Lands Assessment for Water Management Applications Pilot Test. She is a member of the Association of Pacific Coast Geographers, California Remote Sensing Council and the American Society of Photogrammetry where she is on the Board of Directors for the Northern California Chapter.