NASA CR-166,370

NASA CONTRACTOR REPORT 166370

Irrigated Lands Assessment For Water Management: Technique Test

Sharon L. Wall
Catherine E. Brown
Marian Eriksson
Carolyn A. Grigg
Randall W. Thomas
Robert N. Colwell

University of California, Berkeley

John E. Estes
Larry R. Tinney
Jay O. Baggett

University of California, Santa Barbara

Glenn Sawyer
Charles Ferchaud

California Department of Water Resources

NASA Cooperative Agreement NCC 2-54
April 1981

NASA

NASA CONTRACTOR REPORT 166370

Irrigated Lands Assessment for Water Management: Technique Test

Sharon L. Wall
Catherine E. Brown
Marian Eriksson
Carolyn A. Grigg
Randall W. Thomas
Robert N. Colwell

Remote Sensing Research Program
University of California, Berkeley
Berkeley, California 94720

John E. Estes
Larry R. Tinney
Jay O. Baggett

Geography Remote Sensing Unit
University of California, Santa Barbara
Santa Barbara, California 93106

Glenn Sawyer
Charles Ferchaud

California Department of Water Resources
Sacramento, California 95814

Prepared for
Ames Research Center
under Cooperative Agreement NCC 2-54

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

N83-17921#

ABSTRACT

TASK I - A procedure for estimation of irrigated land using full frame Landsat imagery and a sample of ground data was demonstrated statewide in 1979. Relatively inexpensive interpretation of multidate Landsat photographic enlargements was used to produce a map of land irrigated at least once during the calendar year.

Maps of irrigated land based on ground survey were also obtained for a sample of ground sample units allocated for each hydrologic basin in the state. The Landsat and ground maps were then linked by regression equations to enable precise estimation of irrigated land area by county, basin, and statewide.

Land irrigated at least once in California during 1979 was estimated to be 9.86 million acres, with an expected error of less than 1.75 percent at the 99 percent level of confidence. This figure was found to be within one-half of one percent of a corresponding acreage estimate developed by the California Department of Water Resources (DWR) using data obtained from conventional ground mapping and sampling techniques. Most basin acreage estimates of irrigated area were estimated to be within 2 to 6.5 percent of their true values 95 times in 100.

To achieve the same level of error with a ground-only sample would have required a minimum of 3 to 5 times as many ground sample units statewide. The operational cost for a ground-only sytem would therefore, based on preliminary figures, be in the range of 1.5 to 2.5 times that required for a corresponding Landsat-ground system. An additional advantage of the Landsat-aided approach is the availability of a complete area map for irrigated land. This product would not be forthcoming from the ground-only system unless a complete area mapping was done at considerable extra cost.

The Landsat-ground estimation procedure demonstrated in Task I has been designed to complement current DWR mapping programs. On the average, one seventh of the state of California is mapped to crop type and land use each year by DWR. These data are then used by the Department to assist in the planning and management of the State's water delivery system. The Task I Landsat system's role is to enable inexpensive, statewide estimation of land irrigated in any given year, broken down by county and basin. As such, the Landsat procedure represents a new capability for obtaining near-real time data on changes in agricultural water use throughout the State.

TASK II - A procedure for relatively inexpensive computer classification of Landsat digital data to irrigated land categories was developed further during the previous year. This technique is designed to replace the manual Landsat classification employed in Task I where cost-effective. The objectives of a DWR inventory system utilizing this digital technique are to (1) produce regression estimates of irrigated land area as in Task I for counties, basins, and statewide; and to (2) provide a digital data base for easy computer-based production of map products of varying kinds.

Classification results based on the ratio of Landsat band 7 to band 5 (a vegetation greeness indicator) gave good results for several counties in the California Central Valley in 1979. Comparison of 7/5 ratio acreage values to

DWR ground survey figures for the same year or adjusted to the same year showed a difference of only 0.3% in Sacramento county and 1.5% in Kern County. Differences in Tulare and Kings counties were 6.8 and 10.8 percent respectively. None of the Landsat acreage figures just cited were adjusted by regression on ground sample unit data.

Regression of Landsat 7/5 ratio irrigation class acreage data from seven 30 minute by 30 minute blocks covering the Sacramento Valley floor in 1979 gave a ground-calibrated acreage estimate to within 8 percent at the 95 percent confidence level. Correlations between matching Landsat and ground sample unit measurements were found to be somewhat lower than those experienced with the manual technique of Task I. Stratum-specific classification and the use of Landsat brightness measures were proposed to correct this problem.

TASK III - This task has been directed towards development of manual Landsat interpretation techniques for crop type mapping. During this reporting year, effort was focused on (1) obtaining regional crop distribution data and crop phenology data useful in the interpretation process, and on (2) the initial development of a procedure for manually and inexpensively mapping small grains acreage with Landsat color composite imagery. Definition of an efficient small grains mapping technique will be important in providing more accurate identification of small grains fields on current DWR map products.

TASK IV - The objective of this task is to develop a baseline, computer-based mapping and area estimation system capable of meeting many of the California DWR's land use information needs. Work during the previous year has focused on (1) developing an efficient multicrop Landsat classification procedure; and on (2) developing simulation techniques that will allow identification of cost-effective crop area estimation procedures using registered Landsat and ancillary data.

TABLE OF CONTENTS

Page

NASA TECHNICAL OFFICER

Ethel H. Bauer

M/S 242-4
NASA/Ames Research Center
Moffett Field, CA 94035

The research reported on under this Cooperative Agreement (NCC 2-54) was conducted previously under NASA Contract NAS 5-20969 (Goddard Space Flight Center, Greenbelt, MD) and NASA Grant NSG 2207 (Ames Research Center, Moffett Field, CA).

ACKNOWLEDGMENTS

Much of the success of the project is directly attributable to continuing interest and support from the California Department of Water Resources. We would particularly like to acknowledge Mr. James Wardlow (State Office, Sacramento), who has been an active participant on the project. We would also like to acknowledge Mr. J. Kono, Mr. F. Littel and Mr. G. Sata (Central District, Sacramento); Mr. C. Ferchaud, Mr. R. McGill and Mr. G. Pixley (Northern District, Red Bluff); Mr. J. Berthelot, Mr. F. Stumpf and Mr. J. Williams (San Joaquin District, Fresno); and Mr. R. Smith and Mr. R. Yoha (Southern District, Los Angeles) who have been particularly active participants on the project.

In addition, we would like to express our appreciation of our colleagues within the University of California: from the Remote Sensing Research Program (Berkeley Campus), the Director of the Program, Mr. Stephen DeGloria; Engineer, Mr. Anthony Travlos; Photographers, Mr. Stephen Harui and Ms. Dorothy Robinson; Statistician, Mr. Thomas Gossard; and Programmers, Ms. Evelyn Gray, Mr. Agnis Kaugars, Mr. Paul Ritter and Mr. Derek Taylor; and from the staff research personnel, Ms. Hilary Davis and Ms. Nicole Nedeff; from the Geography Remote Sensing Unit (Santa Barbara Campus), Photographer, Mr. Joseph Scepan; and from the staff research personnel, Ms. Tara Torburn.

We have also received valuable assistance from Mr. Bryce Jope and Mr. George Katibah of CALTRANS (Sacramento).

1.0 INTRODUCTION*

The state of California is the location of one of the most complex and productive agricultural industries in the world. California's agriculture is diverse, with no single crop dominating the State's farm economy. With the cultivation of some 200 crops, most crops individually account for less than two percent of the State's total gross farm income.

California's gross cash receipts from farm marketings in 1980 totalled $13.7 billion. With this income, California continues as the leading farm state, with nearly ten percent of the total from only three percent of the nation's farms. The State leads the nation in the production of 49 crop and livestock commodities and is one of the top five producers in another 25. California accounts for half of the nation's cash receipts for fruits and nuts and for about one-third of the vegetables.

This abundance of agricultural production results from the cultivation of the State's 3.8 million hectares (9.8 million acres) of farms. The 1980 production was composed of 25 million metric tons (27.6 million tons) of field crops, 11.2 million metric tons (12.3 million tons) of vegetables and 10.2 million metric tons (11.2 million tons) of fruits and nuts. California's production for 1980 was a record 46.5 million metric tons (51.2 million tons).[1]

Much of the success of this agricultural production is founded on the availability of water for irrigation. The California Department of Water Resources (DWR) estimates that approximately 3.8 million hectares (9.5 million acres) are irrigated at least once during the growing season. This water is derived from surface sources, groundwater extraction and the construction of large-scale water transport projects. Agriculture is the prime recipient of the available water, utilizing about 85% of the supply.

In 1957, California Water Code Section 10005 established the California Water Plan. It is a "comprehensive master plan to guide and coordinate the planning and construction of works required for the control, protection, conservation and distribution of the water of California to meet present and future needs for all beneficial uses and purposes in all areas of the State."[2] The responsibility for updating and supplementing the Plan was assigned to the Department of Water Resources.

"The Department carries out this responsibility through a statewide planning program, which guides the selection of the most favorable pattern for the use of the State's water resources, considering all reasonable alternative courses of action. Such alternatives are evaluated on the basis of technical feasibility and economic, social, and institutional factors. The program comprises:

---

* All principal measurements and calculations were performed using customary units.

[1] Department of Food & Agriculture, State of California, "California Agriculture - 1980"

[2] Department of Water Resources, State of California, "The California Water Plan Outlook in 1974," Bulletin No. 160-74, November 1974

- Periodic reassessment of existing and future demands for water for all uses in the hydrologic study areas of California.

- Periodic reassessment of local water resources, water uses, and the magnitude and timing of the need for additional water supplies that cannot be supplied locally.

- Appraisal of various alternative sources of ground water, surface water, reclaimed waste water, desalting, geothermal resources, etc. - to meet future demands in the areas of water deficiency.

- Determination of the need for protection and preservation of water in keeping with protection and enhancement of the environment.

- Evaluation of water development plans.[3]

A summary status of conditions and expectations is published every four years in the form of a comprehensive bulletin (Bulletin 160) that is used to provide information to aid in guiding and coordinating the use of California's water resources.

To meet these responsibilities, DWR has long recognized the need for specific land use data as an input to state water planning. Since the late 1940's the Department has been performing a continuing survey to monitor land use changes over the State. Because of manpower and budgetary constraints, only a portion of the State (approximately one-seventh) is surveyed during any given year. In DWR's surveys, two types of output are produced, (1) land use surveys which record the nature and extent of present water-related land development, and (2) land classification surveys designed to determine the location and extent of lands with physical characteristics suited to specific kinds of development. The more pertinent of these surveys to the projects discussed in this report, is the land use survey. It is compiled through the interpretation of current 35 mm aerial photography supplemented with field inspections. Tabulations of the acreage of each specific land use class are then summarized by 7-1/2 minute quad sheet, county and other area subdivisions such as water agency or hydrographic area. Figures 1-1 and 1-2 show the land use legend and a completed land use map prepared by DWR.

As seen in Figures 1-1 and 1-2, each parcel of agricultural land has been designated as either irrigated, the prefix "i", or non-irrigated, "n". This condition is determined by the interpretation of aerial photography and the gathering of supplementary field data as mentioned above. From the data collected, DWR is able to generate maps showing the land use classification to cover type, including crop identification, and the acreage of irrigated lands. Since each land use is associated with a specific water demand, total water consumption forecasts can then be made. Due to the limitations of the one date survey, however, the DWR survey is not considered accurate as to the proportion of acreage devoted to small grains or multiple cropping.

---

[3] Ibid

In addition to their normal survey techniques, DWR has been actively participating since 1975 with NASA and the University of California on several projects designed to investigate the feasibility of estimating irrigated acreage and determining cropping practices within the State utilizing a Landsat-based remote sensing system. Based on the results of these studies, information acquired from the analysis of satellite imagery may become a valuable supplement to the land use information presently collected by DWR. The use of the satellite system allows DWR the opportunity to analyze data from several dates during the growing season and the ability to collect data over the entire state in one year.

AGRICULTURE

Each parcel of agricultural land use is labeled with a notation consisting basically of three symbols. The first of these is a lower case "i" or "n" indicating whether the parcel is irrigated or nonirrigated. This is followed by a capital letter and number which denote the use group and specific use as shown below.

C SUBTROPICAL FRUITS
1 Grapefruit
2 Lemons
3 Oranges
4 Dates
5 Avocados
6 Olives
7 Miscellaneous subtropical fruits

D DECIDUOUS FRUITS AND NUTS
1 Apples
2 Apricots
3 Cherries
5 Peaches and Nectarines
6 Pears
7 Plums
8 Prunes
9 Figs
10 Miscellaneous or mixed deciduous
12 Almonds
13 Walnuts

G GRAIN AND HAY CROPS
1 Barley
2 Wheat
3 Oats
6 Miscellaneous and mixed hay and grain

F FIELD CROPS
1 Cotton
2 Safflower
3 Flax
4 Hops
5 Sugar beets
6 Corn (field or sweet)
7 Grain sorghums
8 Sudan
9 Castor beans
10 Beans (dry)
11 Miscellaneous field

T TRUCK AND BERRY CROPS
1 Artichokes
2 Asparagus
3 Beans (green)
4 Cole crops
6 Carrots
7 Celery
8 Lettuce (all types)
9 Melons, squash, and cucumbers (all kinds)
10 Onions and garlic
11 Peas
12 Potatoes
13 Sweet potatoes
14 Spinach
15 Tomatoes
16 Flowers and nursery
18 Miscellaneous truck
19 Bushberries
20 Strawberries
21 Peppers (all types)

P PASTURE
1 Alfalfa and alfalfa mixtures
2 Clover
3 Mixed pasture
4 Native pasture

V VINEYARDS

R RICE

I IDLE
1 Land cropped within the past three years but not tilled at time of survey
2 New lands being prepared for crop production

S SEMIAGRICULTURAL AND INCIDENTAL TO AGRICULTURE
1 Farmsteads
2 Feed lots (livestock and poultry)
3 Dairies
4 Lawn areas

Special conditions are indicated by the following additional symbols and combinations of symbols.

A ABANDONED ORCHARDS AND VINEYARDS
F FALLOW (tilled but not cropped at time of survey)
S SEED CROPS
Y YOUNG ORCHARDS AND VINEYARDS
X PARTIALLY IRRIGATED CROPS

INTERCROPPING (or interplanting) is indicated as follows: $i\,\frac{D13\text{-}y}{T9}$ = a melon crop planted between rows of young walnut trees

URBAN

UC - URBAN COMMERCIAL

UC 1 Miscellaneous establishments (offices and retailers)
UC 2 Hotels
UC 3 Motels
UC 4 Apartments, barracks (three family units and larger)
UC 5 Institutions (hospitals, prisons, reformatories, asylums, etc., having a reasonably/stable 24-resident population)
UC 6 Schools (yards mapped separately if large enough)
UC 7 Municipal auditoriums, theaters, churches, buildings, and stands associated with race tracks, football stadiums, baseball parks, rodeo arenas, etc.
UC 8 Miscellaneous high water use (indicates a high water use not covered above)

UI - URBAN INDUSTRIAL

UI 1 Manufacturing, assembling, and general processing
UI 2 Extractive industries (oil fields, rock quarries, gravel pits, public dumps, rock and gravel processing plants, etc.)
UI 3 Storage and distribution (warehouses, substations, railroad marshalling yards, tank farms, etc.)
UI 6 Saw mills
UI 7 Oil refineries
UI 8 Paper mills
UI 9 Meat packing plants
UI 10 Steel and aluminum mills
UI 11 Fruit and vegetable canneries and general food processing
UI 12 Miscellaneous high water use (indicates a high water use not covered above)

UV - URBAN VACANT

UV 1 Miscellaneous unpaved areas
UV 4 Miscellaneous paved areas

UR - URBAN RESIDENTIAL

One and two family units, including trailer courts

RECREATION

RR RESIDENTIAL
Permanent and summer home tracts within a primarily recreational area. (The estimated number of houses per acre is indicated by a number in the symbol.)

RC COMMERCIAL
Commercial areas within a primarily recreational area (includes motels, resorts, hotels, stores, etc.)

RT CAMP AND TRAILER SITES
Camp and trailer sites in a primarily recreational area

P PARKS

NATIVE

NV NATIVE VEGETATION

NR RIPARIAN VEGETATION
NR 1 Swamps and marshes
NR 2 Meadowland

NW WATER SURFACE

NC NATIVE CLASSES UNSEGREGATED

Figure 1-1. Legend developed by the California Department of Water Resources and used in their land use surveys.

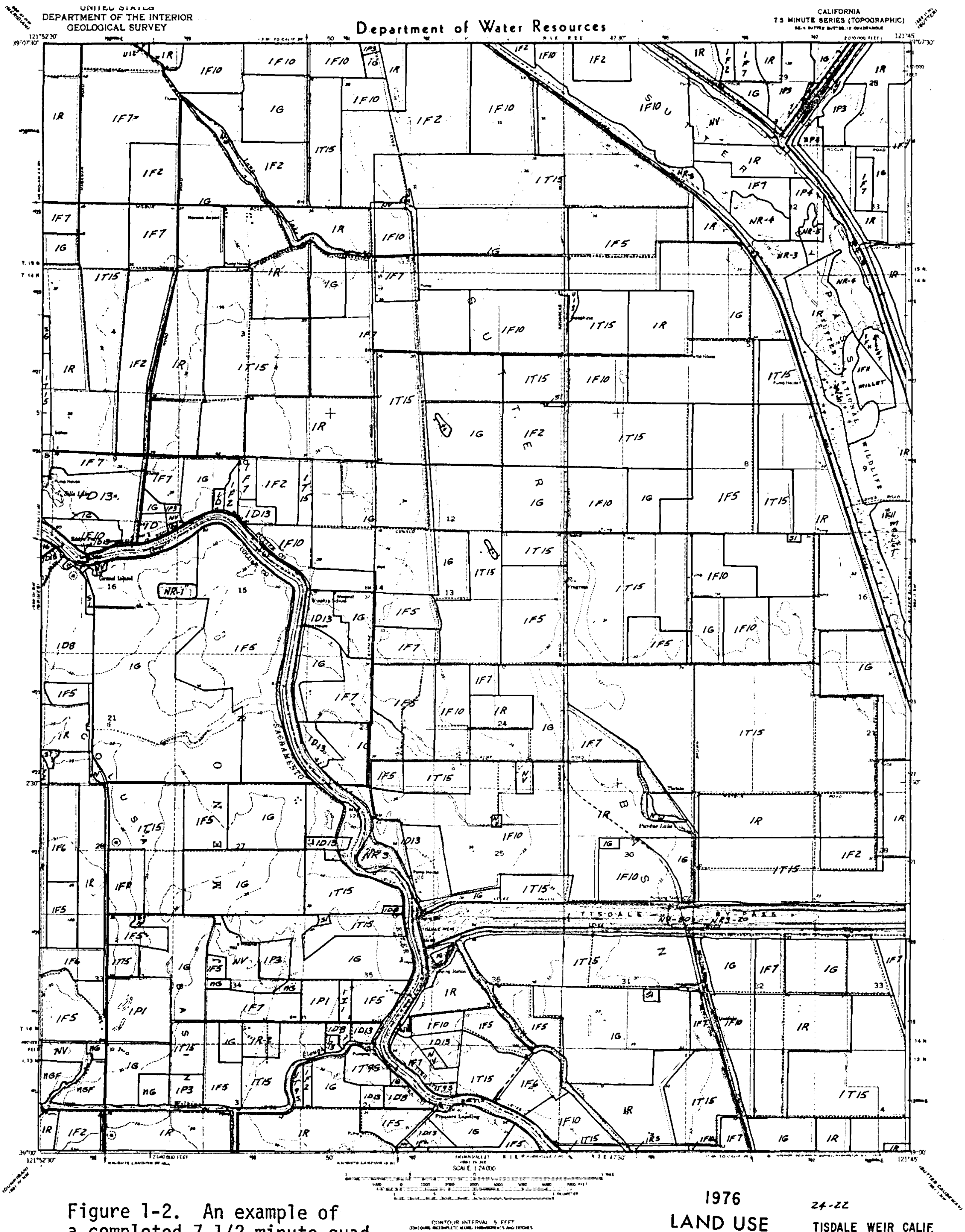


Figure 1-2. An example of a completed 7-1/2 minute quadrangle land use survey map prepared by DWR.

2.0 OBJECTIVES

Throughout the life of this Applications Pilot Test and the projects that directly proceeded it, there have been a number of fundamental questions which have motivated technique development and validation. These questions fall into three major categories; first, can a Landsat-based system deliver acceptable results on both the area of irrigated land and the area of specific crop types? Second, can we develop procedures that make optimal use of manual analysis of Landsat image products and digital analysis of Landsat computer compatible tapes? Third, can inventory systems that meet both the estimation and/or mapping requirements of the Department of Water Resources be developed? To answer these questions, the project work was divided into four major tasks:

- Task I - Estimation of irrigated land using manual analysis techniques
- Task II - Estimation and mapping or irrigated land using digital analysis techniques
- Task III - Estimation/mapping of crop type using manual analysis techniques
- Task IV - Estimation/mapping of crop type using digital analysis techniques

Each of these tasks has specified information requirements associated with it. Tables 2-1 through 2-3 outline the mapping and estimation needs for each of the tasks, the relationship between them, the expanding sophistication of the tasks (from I to IV), as well as the constraints that have directed the procedural developments. The goals detailed in these matrices have guided the operation of the project and provided the standards by which the results were evaluated.

More specifically, for 1980 the following objectives were defined: (1) for Task I, complete the estimation of the total irrigated area of the State of California and perform a detailed evaluation of the procedures and results; (2) for Task II, continue the development of the MSS band 7-to-MSS band 5 ratio techniques for the digital estimation and mapping of irrigated land and demonstrate these techniques on three areas, the Tulare Hydrologic Basin, Sacramento County and the Sacramento Hydrologic Basin; (3) for Task III, continue development of crop phenology diagrams, plan for the development of a 7.5 minute quad-based agricultural information system, and prepare for a limited manual survey of small grains; and (4) for Task IV, re-evaluate the results of crop type classification done in Kern County and continue the development and demonstration of classification and sampling techniques for digital crop type estimation and mapping in the Sacramento Valley. (Figure 2-1) As in previous years, the completion and evaluation of the Task I statewide inventory dominated the project during 1980. Significant progress was also made on the other tasks, especially on the development of techniques for the digital estimation and mapping of irrigated land (Task II).

TABLE 2-1

(1) INFORMATION REQUIREMENTS

A. ESTIMATION

TASK	I	II	III	IV
1. PARAMETERS TO BE ESTIMATED	IRRIGATED AREA (PROPORTION OF AGRICULTURAL AREA)	AS IN I	CROP AREA	CROP AREA * ● CROP AREA CHANGE ● WATER USE
2. LAND USE CLASSES FOR WHICH PARAMETER ESTIMATES DESIRED	AREAS IRRIGATED AT LEAST ONCE	"	SELECTED SET OF CROP CATEGORIES (E.G. SMALL GRAINS)	ALL CROPS FOR WHICH DWR PRESENTLY PROVIDES SUMMARY; GENERALLY 10-20 PER BASIN
3. CLASSES FOR WHICH SAMPLING ERROR CONTROLLED	SAME	"	SAME	SIGNIFICANT CROPS: F (AREA, WATER USE, WATER QUALITY) GENERALLY ≤ 10 PER BASIN
4. REPORTING LEVEL AT WHICH ERROR IS CONTROLLED	HYDROLOGIC BASIN	"	COUNTY	COUNTY
5. SAMPLING ERROR GOAL	± 5% AT 95% C.L.		TBD	VARIABLE BY CROPS 5-20% @ 90-95 C.L.
6. REPORTING LEVELS FOR WHICH ESTIMATES DESIRED	COUNTY, BASIN, STATE	"	DETAILED ANALYSIS UNIT (DAU), COUNTY	DETAILED ANALYSIS UNIT (DAU), COUNTY
7. OTHER AVAILABLE INFORMATION	A. AREA IRRIGATED AND ERROR BY STRATUM B. AREA IRRIGATED BY DATE (IF MATCHING GROUND DATA)	"	TBD	1. AREA BY CLASS BY STRATUM

Table 2-2

B. Mapping

Task	I	II	III	IV
1. Land use classes to be mapped	Vegetated/irrigated land in agricultural areas	A. Irrigated, non-irrigated, exclusion areas B. Irrigated area by date	TBD	All DWR land use classes* • Areas of land use change • Multicropped areas • Water use classes • Other: water quality; drainage
2. Ground registration accuracy				
Maximum error (95%)	150m	x:115m, y:160m		As in II
Median of absolute error (Based on Regression Fit)		x:45m, y:55m		Better if possible for small field areas
3. Spectral class purity[1] "Classification accuracy"	NA	85% +		70% + Highest on significant classes
4. Reporting units for information summary	NA	Any agricultural area		As in II
5. Map products (USGS quadrangle dominant)	None originally planned[2]	1. Tabular summary, 1:125,000 map 2. Hardcopy color/B&W print transparencies 3. Digital data base		7.5' map, Digital data base

[1] Frequency of most commonly occurring land use class in a given spectral class.

[2] Though digitized "irrigated" map available, as in II.

Table 2-3

C. Constraints

TASK	I	II	III	IV
1. Cost per agricultural acre (Jan. 79 dollars)	1 - 2¢	1.5 - 4¢ (EXCLUDING HARDCOPY)	2 - 3¢	TBD
2. Time required for inventory exclusive of planning	1 YEAR	1 YEAR	3 MOS.-1 YR.	VARIABLE
3. Special expertise required				
REMOTE SENSING	YES	YES	YES	YES
BASIC SAMPLING	YES	YES	YES	-
ADVANCED SAMPLING	NO	NO	PREFERRED	YES
CUSTOM PHOTOPROCESSING	YES	NO	PREFERRED	PREFERRED
COMPUTER PROGRAMMING	NO	PREFERRED	PREFERRED	YES
4. Special equipment required				
PHOTO LAB	YES	YES	YES	YES
DIGITIZER	PREFERRED	PREFERRED	PREFERRED	YES
BATCH MAINFRAME	PREFERRED	YES	PREFERRED	YES
INTERACTIVE DIGITAL ANALYSIS SYSTEM	NO	PREFERRED	NO	YES

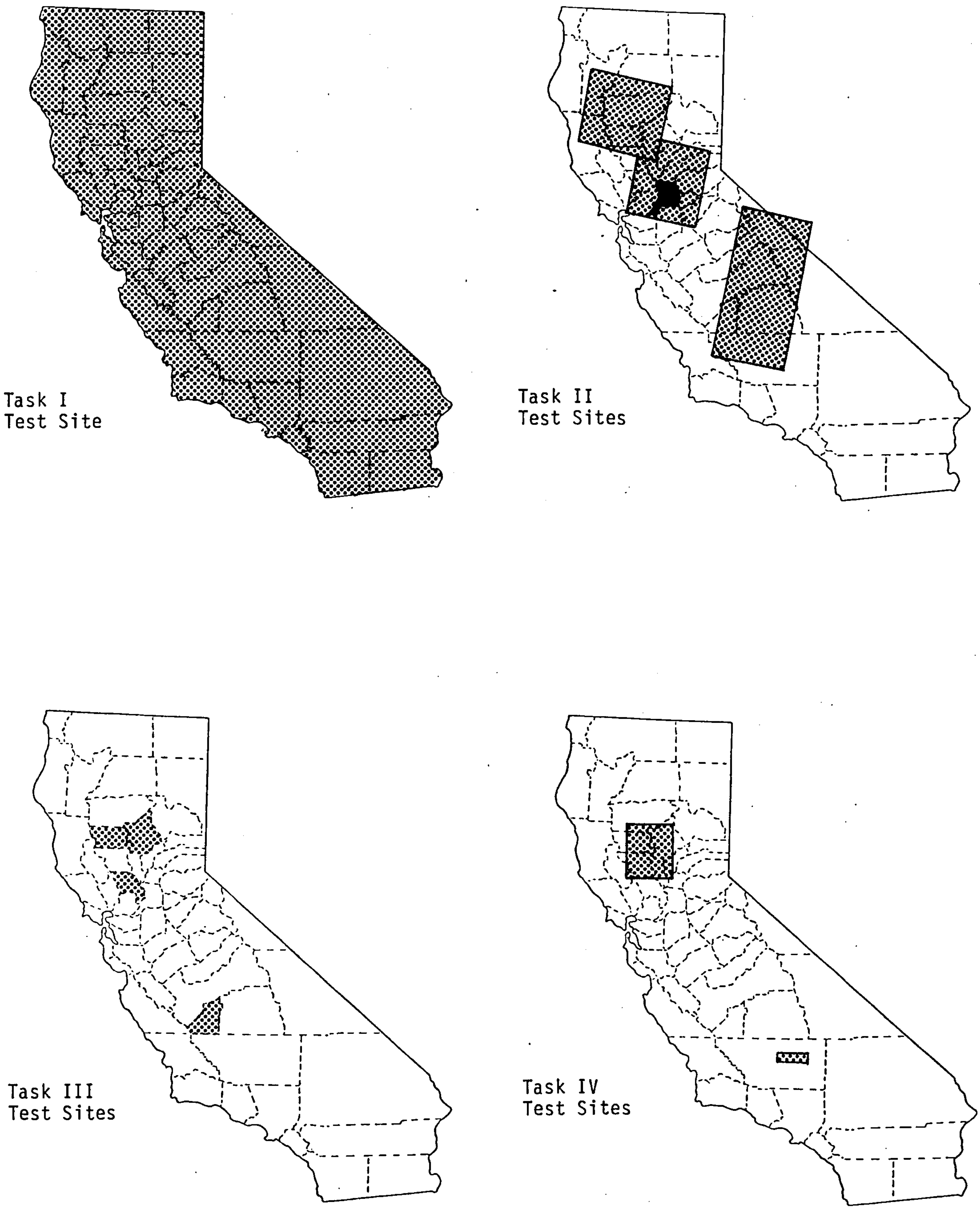


Figure 2-1. Test Site Locations

3.0 ESTIMATION OF IRRIGATED LAND USING MANUAL ANALYSIS TECHNIQUES (TASK I)

When attempting to produce a highly accurate, repeatable estimate of irrigated land over a state as large and complex as California, a detailed analysis flow is an integral part of the design process. Based on our experience on previous projects (see Wall, Baggett, et al, 1980), five major sub-tasks were defined to guide the processing of the data from the initial definition of information requirements to the final production and evaluation of results. Figure 3-1 presents the analysis flow with its five major attributes:

- Design and sample allocation
- Stratification and sample frame construction
- Landsat measurement
- Ground measurement
- Estimation, results and evaluation

In keeping with the flow presented in Figure 3-1, the Task I description that follows will adhere to that organization. The first four sub-tasks are summarizations of work reported on in the 1980 report. The estimation, results and an evaluation of the results, as well as recommendations for modification to the inventory system are presented in detail.

3.1 Design and Sample Allocation

Specifying the inventory design required addressing several key issues: (1) defining the information required by the California Department of Water Resources; (2) generating a data set to be used as a preliminary population model to test and refine the previously used estimation system; (3) applying statistical techniques (Monte Carlo) to the data set to simulate model performance; with the simulation testing various mathematical models, evaluating the stratification scheme and determining expected sample sizes for hydrologic basins; (4) specifying the mathematical model, stratification procedures and sample frame for the 1979 inventory; and, (5) computing the actual sample allocation.

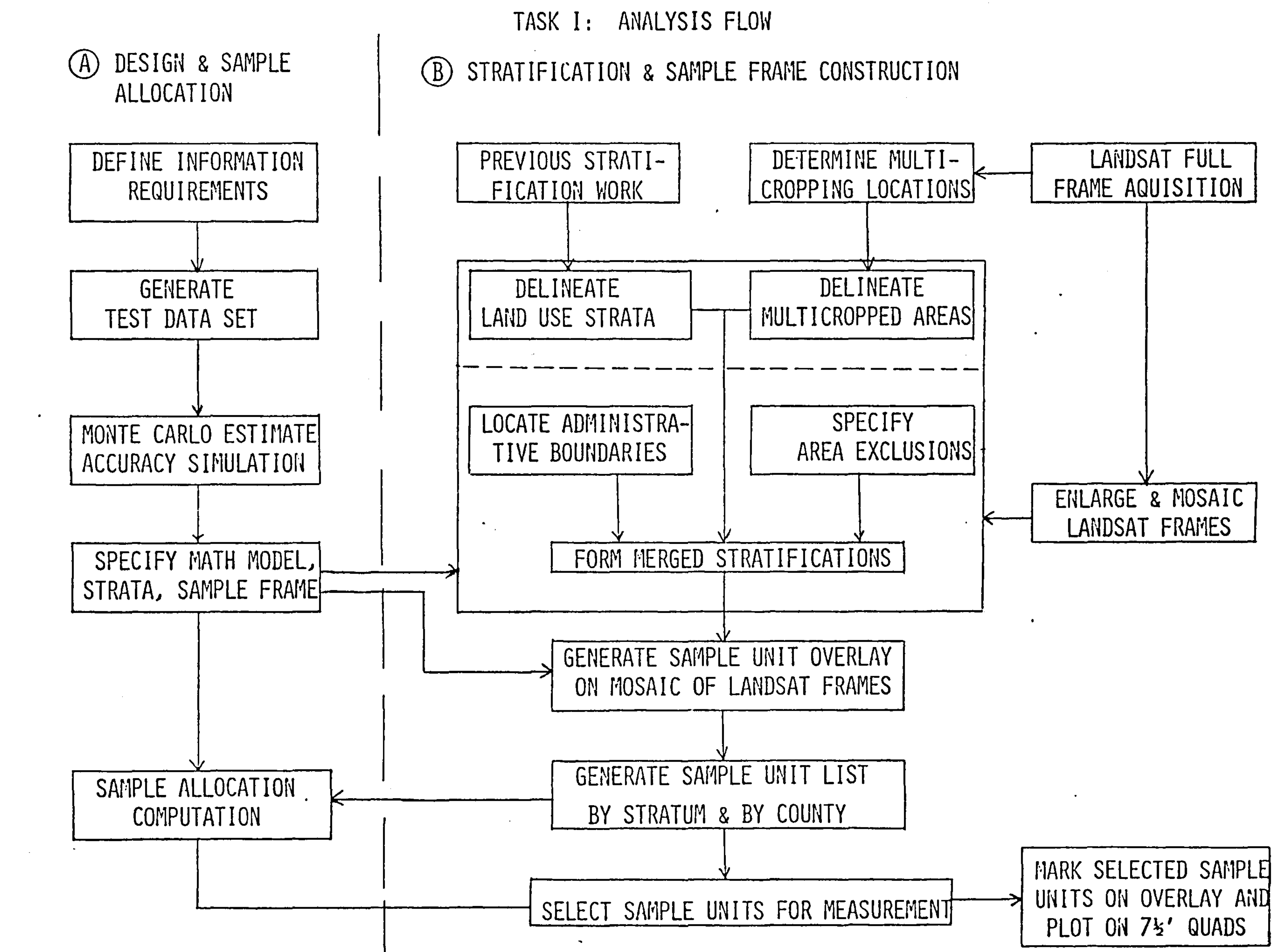


Figure 3-1. Task I analysis flow. Task I was divided into five major organizational sub-tasks: (A)design and sample allocation; (B)stratification and sample frame construction; (C)Landsat measurement; (D)medium scale photography and ground measurement; and (E) estimate summary, evaluation and report.

C LANDSAT MEASUREMENT

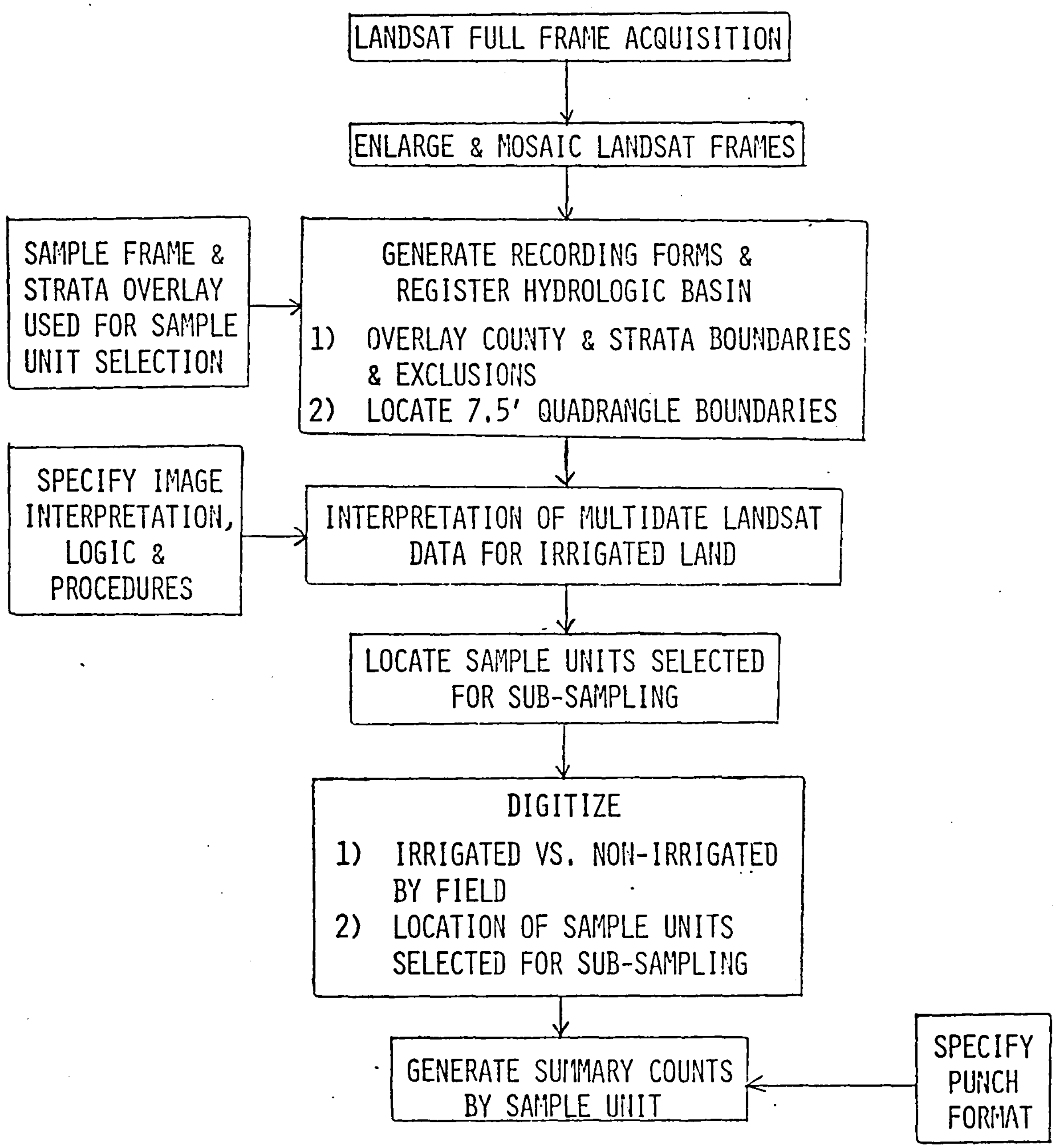


Figure 3-1 (cont'd)

(D) MEDIUM SCALE PHOTOGRAPHY AND GROUND MEASUREMENT

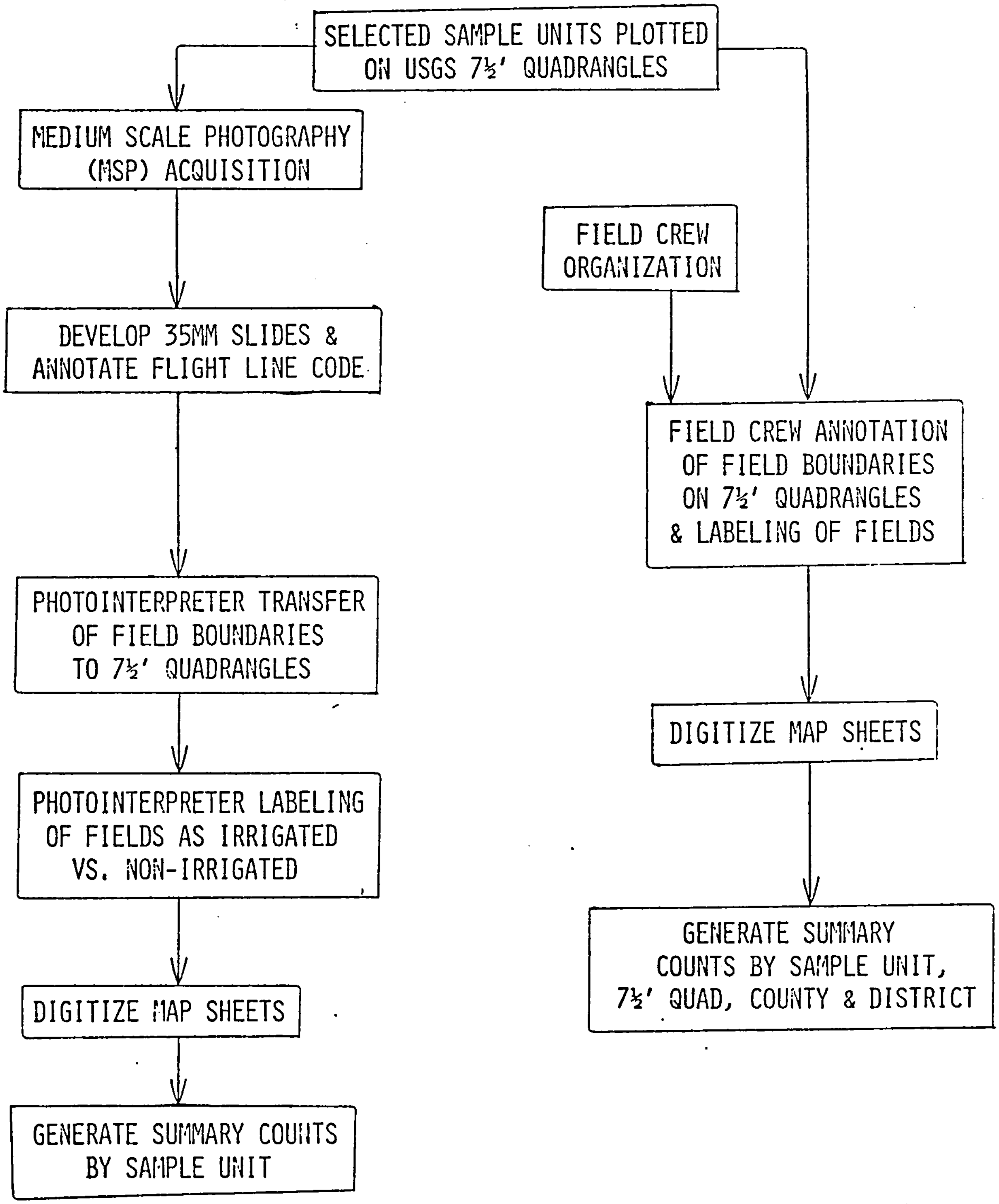


Figure 3-1 (cont'd)

TASK I: ANALYSIS FLOW

Ⓔ ESTIMATE SUMMARY, EVALUATION, REPORT

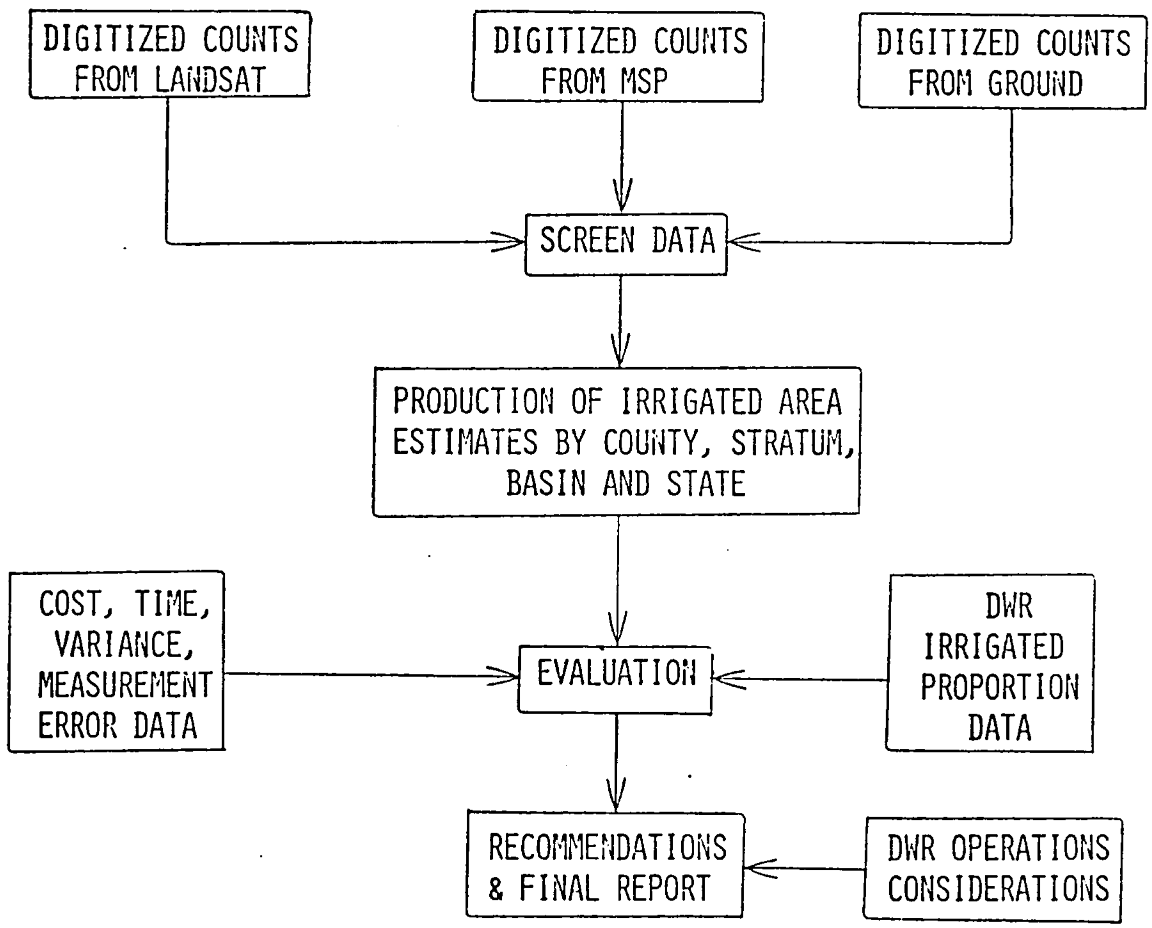


Figure 3-1 (cont'd)

3.1.1 Definition of Information Requirements

A necessity in any project is to strictly and accurately define information requirements. This procedure demands frank appraisal by the user agency as to what is really needed and a straightforward explanation of what can be expected from a particular remote sensing system. Certain fundamental questions designed to carefully define DWR's information needs were posed. These questions, and the responses provided by DWR, formed the base upon which the Task I design was built. Table 3-1 briefly summarizes those questions and responses.

Table 3-1. Design requirements for the statewide estimation of irrigated land.

. Type of information?	. Estimation of the proportion of irrigated land
. Areas of summary?	. Hydrologic Basin (10) . County (58) . State
. Time?	. Inventory data summary within one year, exclusive of planning phase
. Accuracy?	. Estimate precision control at hydrologic basin level
	. True value of proportion irrigated to fall within ± 5% of estimate 95 times out of 100
. Cost?	. Not formally specified, but in the range of 1 to 2 cents per agricultural acre
. Technology constraints?	. Must be implementable by current DWR personnel and processing capabilities

3.1.2 Generation of Test Data Set and Monte Carlo Simulation

Once inventory information needs were established, the sample design phase progressed to the next step; evaluating the previously used estimation procedures and developing an improved system to meet refined and updated inventory objectives. To proceed with this evaluation, data collected and analyzed for a previous study conducted on fourteen counties was used. This data set consisted of 141 locationally matched pairs of Landsat and ground data sample segments (each sample segment ≅ 5 square miles). Using this paired data, Monte Carlo simulations were performed to: (1) test an alternative to the regression estimator used to link estimates made at various phases (Landsat, aerial photo, ground); (2) examine the value of stratification (originally designed to control measurement error) for controlling sampling error; and (3) compute the approximate number of sample segments needed to achieve a sampling error of $\pm$ 5% at the 95% level of confidence within a hydrologic basin.

The Monte Carlo simulation was used to test the relative performance of two estimators: regression and biased ratio. The biased ratio was evaluated as an alternative since this estimator exhibits lower variance under certain conditions. The form of these estimators, including their variance estimators, is given in Appendix 2. The results of this evaluation showed a lack of any significant difference between the two estimators' performance as exhibited by very similar values of average bias and standard error of the estimate at both the 95% and 99% levels of confidence. Except for small sample sizes, the regression estimator exhibited lower bias and variance than did the ratio estimator. Though the regression estimator was judged superior to the ratio over most strata, the Monte Carlo did not clearly indicate which estimator, if either, was "the best" to accomplish the objectives of the statewide inventory of irrigated land. Based on these results, a more in-depth analysis of other mathematical estimators was performed.

Using the same data set for the 14-County Study, two additional estimators, the simple random sample (SRS) and unbiased ratio were evaluated with the biased ratio and regression forms. The performance of all four was evaluated deterministically by predicting group variance (σ_y^2) and correlation (σ) between Landsat and ground. By determining the variance of the estimators using variable sample sizes the relative performance of each estimator could be evaluated. That estimator exhibiting the lower variance for given sample sizes would be preferred for the statewide inventory.

By examining variance plotted against sample size (sample size ranging from 2-25, including an estimate of variance for very large sample sizes, $n \to \infty$), it was shown that the regression estimator was superior to all others for large sample sizes ($n \geq 5$). For small sample sizes both ratio estimators were superior to regression but indistinguishable from each other. Because of the standard error of the biased ratio estimator was at best 13% less than that of the unbiased ratio estimator and given the advantages of using an estimator with no bias, the unbiased ratio estimator was recommended for use with small sample sizes.

The second objective of the Monte Carlo tests was to evaluate the agricultural practice stratification used in the Sacramento Valley fourteen county test. This stratification, based on field size and land use, had been designed to control measurement error associated with the manual interpretation of agricultural land that varies considerably in "ease" of interpretation and accurate line placement. The major purpose of the Monte Carlo test, in this instance, was to evaluate the utility of the stratification in reducing sampling error as well. When the stratification and regression estimators were combined in the Monte Carlo simulations, stratification did little to reduce sampling variance. Since this result would have significant implications for this and future studies, further investigation of regression variance behavior was performed. It was found that stratification would significantly decrease the variance only if the mean square error for the stratified case is significantly less than the mean square error of the unstratified case. (i.e. only if the regression slopes and/or intercepts differed significantly between strata.) After reviewing the Monte Carlo results, the stratification done for the fourteen county test site was redesigned in anticipation of achieving differing regressions. (The recommended stratification scheme is described in Section 3.1.3.)

The final function of the Monte Carlo simulation was to compute the approximate number of sample segments that would be needed to achieve the stated accuracy requirements. As DWR was responsible for the collection of the ground sample data (Phase II), the preliminary computation was to provide a guideline for planning DWR manpower requirements. Computation of the final sample size for the 1979 APT inventory is discussed in Section 3.1.4 . For each sample size (n = 5, 10, 15, 20 ...) used in the Monte Carlo tests, the number of samples (n*) which gave estimates of irrigated proportion that fell within 5% and 10% of the true estimate was determined. This number was converted to a percentage by dividing by the number of cycles (m) and multiplying by 100: (n*/m x 100). These percentages were then graphed against sample size. Preliminary sample sizes necessary to achieve ±5% error 95 times out of 100 were then predicted from these graphs by hydrologic basin. Based on this preliminary analysis a maximum number of 80 segments per hydrologic basin, or 800 units for the entire state, was used by DWR for planning.

3.1.3 Specification of the Mathematical Model, Stratification Scheme and Sample Frame

The Monte Carlo simulations described in Section 3.1.2 provided the information needed to refine the mathematical estimators used to produce the Task I estimate of irrigated acreage. The simulations also indicated that modifications to the stratification scheme would be necessary if the stratification was to be used to reduce sampling as well as measurement error. The sampling frame remained similar to that used in the previous studies (cluster sample units, 1.6 x 8.0 kilometers in size [1 x 5 miles], prior to adjustment for stratum boundaries).

Specification of the Mathematical Model

As in the previous studies, the primary equations (estimators) used to link Landsat and ground area measurements to produce estimates of irrigated area were of the linear regression type. The general form of these equations, as adapted to the irrigated lands problem, was established in previous studies. In the present study, two phases were employed: a census at the Landsat phase (Phase I) and a simple random sample within strata at the ground phase (Phase II). Section 3.5 and Appendix I discuss the estimation procedure and present the equations for estimation of irrigated land and associated error.

Specification of the Stratification Scheme

Based on the results of the Monte Carlo analysis (Section 3.1.2), the stratification scheme used for this year's statewide estimation was modified. The modifications were designed to reduce sampling variance as well as control measurement error. These new strata were composed of areas that, on Landsat 1:1,000,000 color composite transparencies, appear to be:

Table 3-2. Stratification scheme used in the allocation of sample units for the Task I estimation of irrigated land.

Stratum Number	Stratum Description
1	Generally dry farmed
2	Field crop areas dominated by fields less than 16 hectares (40 acres) in size
3	Field crop areas dominated by fields less than 16 hectares (40 acres) in size with known high proportion irrigated
4	Field crops dominated by fields 16 hectares (40 acres) or larger in size
5	Orchards and vineyards less than 16 hectares (40 acres) in size
6	Orchards and vineyards 16 hectares (40 acres) or larger or larger in size
7	Unusual agricultural areas

The procedure used to produce the statewide stratification is described in Section 3.2, Stratification and Sample Frame Construction. This revised stratification scheme was evaluated at the end of the Task I inventory and will be discussed in Section 3.7.

Specification of the Sampling Frame

For the geographic areas, sampling frames usually are constructed as either a point system referenced by coordinates or an arbitrary clustering of areas into some convenient size unit (e.g. rectangular areas). The project objective as well as statistical and implementation considerations all enter into decisions which lead to the "optimum" strategy for sampling the population. Photo-related variables were (and may be in the future) a major part of the system either as a separate phase or as an aid to ground data collection. Therefore, the sampling frame should allow maximum use of the photographic capabilities for a given expenditure of effort. For this reason, point systems are not practical; to photograph a large number of different points with a single or pair of images is very costly. A cluster system is more economical since larger units allow additional information to be obtained at little incremental cost.

Initially, the decisions on sample unit size and configuration were based largely on practical considerations as insufficient data existed to simulate and optimize sample unit dimensions for large area inventories in California. A nominal 1.6 x 8.0 km (1 x 5 mi) sampling unit was used in the preliminary studies because: (1) DWR's standard aerial survey photography covers a one-mile wide strip, (2) a five mile length is easily located and flown over several dates, and (3) the north-south orientation corresponds to DWR's survey techniques.

These same considerations were valid for the present study; thus, the nominal 1.6 x 8.0 km (1 x 5 mi) north-south oriented unit was maintained. Two modifications were made, however. Given the choice during sample frame development of having two small or one large sample unit, the larger unit was favored. This was done to decrease the errors due to possible misregistration of units when they transferred onto maps and Landsat enlargements. The second change was in sample unit orientation. The north-south orientation was maintained in the Central Valley and other agricultural areas where road networks were primarily oriented north-south. The sample units in upland areas and small valleys were oriented along major landforms and/or main thoroughfares. This was done to prevent having a large number of small sample units at the expense of having only a very few large units, and increasing driving efficiency for the ground data collection.

3.1.4 Sample Allocation Computation

As can be seen in the analysis flow (Figure 3-1), the sample allocation computation was based on input from two major sources: (1) the specification of the mathematical model, stratification scheme and sample frame and (2) a sample unit list summarized by stratum and county.

Since the sampling design for Task I included the use of stratification, allocating the sample units required the distribution of sample units among the strata for each hydrologic basin. The distribution of units could have been simple proportional to the relative size of each stratum. Since the 1976 14-County Study gave estimates of within stratum variance (σ^2) and correlation (ρ), the optimum (theoretically giving smallest variance) allocation of sample units to each stratum (n_i) can be accomplished by minimizing variance subject to a cost constraint, as follows:

minimize: $$V(Y) = \sum_{h=1}^{L} W_h^2 \, (1 - \hat{\rho}_h^2) \, \hat{\sigma}^2_{y_h} \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \left(1 + \frac{1}{n_h - 3}\right) \quad (1)$$

subject to: $$C = \sum_{h=1}^{L} n_h c_h \quad (2)$$

where:

$V(Y)$ = estimate of variance of the estimate of basin proportion irrigated

L = number of strata

h = stratum index

W_h = proportion of basin occupied by stratum h

$\hat{\rho}_h$ = sample correlation between Landsat and ground data in stratum h as determined in the 14-County Study

$\hat{\sigma}^2_{y_h}$ = sample variance of proportion irrigated for ground data in stratum h as determined in the 14-County Study

n_h = sample size in stratum h

N_h = population size in stratum h

C = maximum relative cost permitted in basin

c_h = weighted average relative cost of stratum h

The values of L, W_h and N_h came from summary tables for each hydrologic basin. The basin summary tables were compiled from similar tables constructed for each County. The information summarized on the county table was derived from detailed county sample unit lists that described each sample unit in terms of agricultural practice stratum, presence or absence of grain and/or vegetables and relative ease of ground access.

The constraint function (Equation 2) uses the average relative cost of ground checking a sample unit in a particular stratum (c_h). In the 14-County Study all sample units (SUs) were located on the floor of the Sacramento Valley and were considered equally accessible. As sample units were allocated over the entire state for the 1979 inventory, the assumption of equal accessibility was not valid. Therefore, sample units were divided into three accessibility categories. Relative cost weights (c_h) were then determined for each stratum.

After all terms were defined, a computer algorithm, FCDPAK*, was used to minimize Equation 1 subject to Equation 2.

For each hydrologic basin, the total number of sample units was allowed to vary over the range of 30 to 200. FCDPAK determined the optimal allocation of these units to each stratum (n_h). Percent confidence intervals at 95, 98, 99 and 99.9% levels of confidence were also calculated for each allocation. These percent standard confidence intervals were then plotted against the total sample size. Figure 3-2 illustrates a typical plot using the Tulare Basin allocation. From these plots, the total number of sample units required to achieve $\pm 3\frac{1}{2}\%$ at the 95% confidence level was determined by interpolation. As the stated inventory accuracy objective was $\pm 5\%$ at 95% confidence level, this more conservative criteria insured against the possibility that chance alone would cause a failure to meet the stated goal in any hydrologic basin. As seen in Figure 3-2, the total number of sample units required to meet the $\pm 3\frac{1}{2}\%$ at the 95 criterion in the Tulare Basin was interpolated to be 65. This value is then compared to the FCDPAK values bordering this interpolated estimate (i.e. 62 and 81 sample units). The FCDPAK stratum allocation within the Tulare Basin for the 62 and 81 sample units is tabulated in Table 3-3.

To achieve the desired stratum-level allocation of the 65 basin units, a second interpolation was performed using the optimal FCDPAK stratum allocation for 62 and 81 basin units. This procedure was used for all the hydrologic basins. The resulting allocation of sample units by basin and by stratum is given in Table 3-4.

After all the sample units were allocated by stratum for each of the hydrologic basins, the units were physically annotated on map sheets for subsequent ground survey by DWR personnel. Measurement of both the sample units on the ground and the Landsat census is described in the following Sections.

3.1.5 Summary

The design process is a critical element in any inventory activity. It serves to specify the framework for data acquisition, analysis, summary, and storage and retrieval. By specifying this framework, all phases of an inventory are performed in a coordinated fashion, thus increasing the probability of successfully achieving the stated inventory objectives. For the design

---

*FCDPAK (Feasible Conjugate Direction Package for the Solution of Differentiable Mathematical Programs) was developed by Best (1972) to solve the general problem of maximizing a function subject to linear and/or nonlinear constraint functions. The program's only shortcoming is that solutions to n_h are generated in noninteger form. This problem was solved by use of the following contingency table:

$$\text{if } n_h - \text{integer}(n_h) < 0.1, \text{ then } n_h = \text{integer}(n_h)$$

$$\text{if } n_h - \text{integer}(n_h) \geq 0.1, \text{ then } n_h = \text{integer}(n_h) + 1$$

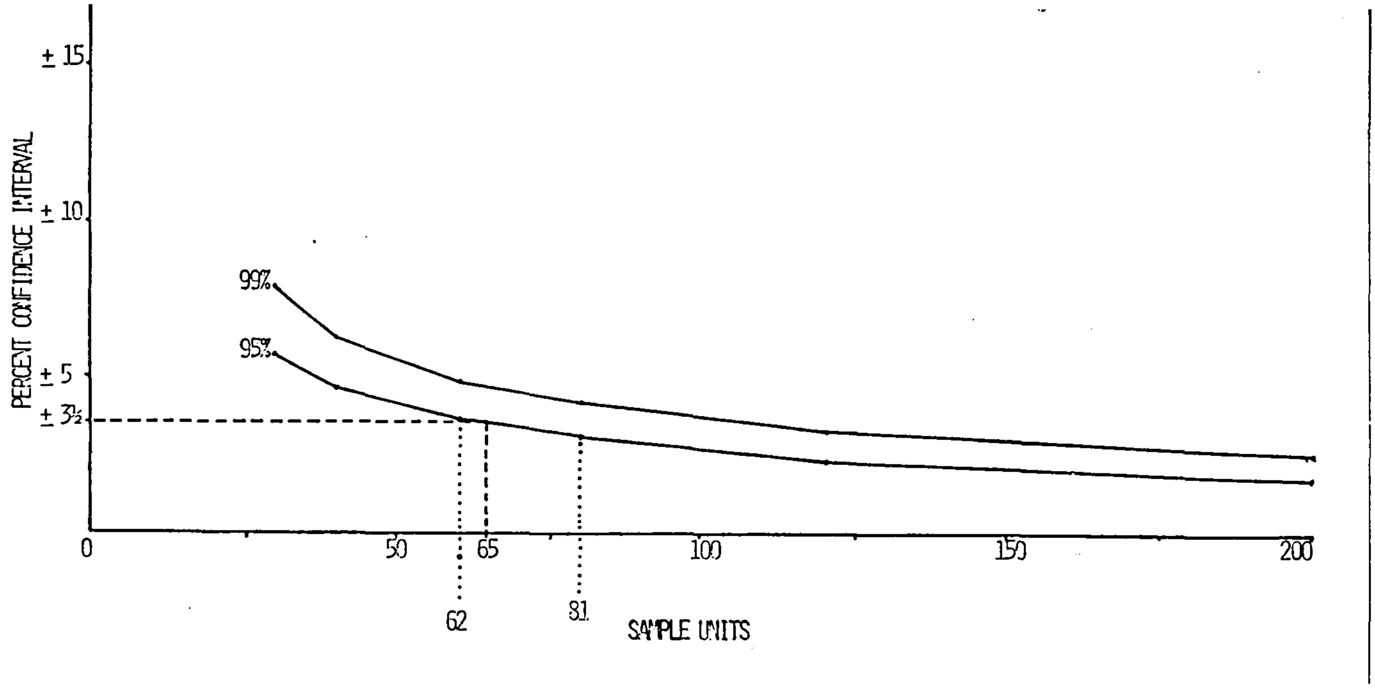


Figure 3-2. Percent standard confidence interval plotted against total sample size for the Tulare Basin.

Table 3-3. Allocation of sample units by stratum. The values were calculated by interpolation from the allocation shown in Figure 3.2.

Stratum	Values from FCDPAK	Values used by interpolation	Values from FCDPAK
1	4	4	4
2	0	0	0
3	0	0	0
4	44	46	61
5	5	5	5
6	9	10	11
7	0	0	0
Total	62	65	81
Accuracy	±3.6 @ 95	±3.5 @ 95	±3.1 @ 95

Table 3-4. Sample unit allocation by stratum for each hydrologic basin.

HYDROLOGIC BASIN	STRATUM 1	2	3	4	5	6	7	TOTAL
Central Coastal	26	7	28	8	-	5	6	80
Colorado Desert	-	-	-	42	4	8	4	58
North Coastal	4	6	-	30	12	-	-	52
North Lahontan	-	12	-	26	-	-	-	38
Sacramento Valley	8	10	-	39	5	4	6	72
San Francisco Bay	19	4	11	7	14	-	-	55
San Joaquin	6	5	-	53	5	14	-	83
South Coastal	12	18	9	11	24	-	8	82
South Lahontan	7	-	-	39	-	-	6	52
Tulare	4	-	-	46	5	10	-	65
California	86	62	48	301	69	41	30	637

process to be successful, the interrelationships between data collection, decision-making, and management must be understood, documented and integrated into the design. Only by understanding these interrelationships in concert with historical inventory practices can realistic assumptions be made, functional relationships documented, and operational systems developed and implemented.

The current design effort for the 1979 inventory had the benefit of close cooperation with the user agency, California Department of Water Resources, who provided invaluable management and decision-making insight and critical information needs on a state-wide basis. Furthermore, DWR conducted the complete ground survey effort providing the sensitive, costly, and compulsory ground data to drive the state-wide estimation process.

Based on the combined DWR input and previous studies conducted by the University of California, important historical data were available for the design process. The experience and the data were used to (1) generate and refine assumptions, (2) evaluate various estimator alternatives, (3) evaluate the effect of stratification on sampling and measurement errors, (4) calculate estimates of variance and data plane (i.e. Landsat-ground) correlations, both critical for the sample size calculations, and (5) generate accessibility/cost constraint functions paramount in the sample allocation process.

After numerous analyses, the inventory design was completed and implemented. The 1979 design may be summarized as follows:

GOAL: Estimate the proportion of irrigated acreage in the state of California to within ±5% allowable error at the 95% level of confidence.

DATA TYPES:

- Multitemporal Landsat color composite imagery enlarged to a scale of 1:150,000 (Phase I)
- Ground data collected by DWR; supplemented with 35mm aerial photography (Phase II)
- USGS maps at scales 1:1,000,000, 1:250,000, 1:62,500 and 1:24,000
- U-2 color infrared aerial photography at a scale of 1:130,000 and 1:24,000

SAMPLING FRAME:

- Sampling frame of area units (clusters)
- 1.6 x 8.0 km rectangular sample unit
- Orientation of sample units predominately north-south; allowed to vary with local topography and road network

STRATIFICATION:

- Hydrologic basin, county
- Agricultural practice/land use
- Small grain and vegetable
- Exclusions

MATHEMATICAL MODEL AND SAMPLE ALLOCATION:

- Two phase design
- Census at Phase I (Landsat)
- Simple Random Sample within strata/basin at Phase II (ground data)
- Phases linked using regression estimator for large sample sizes and an unbiased ratio estimator for small sample sizes

When the statewide inventory is completed, a detailed evaluation of the Task I design process can begin. The evaluation will allow further refinement of assumptions, sampling frame (including size, shape and orientation of SU's), two phase sampling, stratification, sample allocation, and the estimation procedure (i.e. equation used to link phases and predict errors; and, procedures used to aggregate strata estimates into final estimates).

3.2 STRATIFICATION AND SAMPLE FRAME CONSTRUCTION

Stratification is a commonly used technique designed to reduce variance by systematically placing boundaries that separate homogeneous units. For Task I the major purposes of stratification were to: (1) allow summary of data by administrative units (hydrologic basin, county and state), (2) reduce sampling and measurement error, (3) enhance the allocation of sample units, and (4) flag areas for early and/or multiple ground data collection. The production of three stratifications was necessary to address the purposes just described: (1) administrative boundaries were defined by use of a DWR-supplied map delineating hydrologic basins and county boundaries were located from USGS 1:24,000 and 1:250,000 scale topographic maps (Figure 3-3); (2) an agricultural practice stratification was developed to reduce sampling and measurement error and enhance the allocation of sample units; and (3) areas of small- grain and vegetable cultivation were stratified to help optimize ground data collection. The latter two stratifications will be described in Sections 3.2.1 and 3.2.2 As shown on the analysis flow (Figure 3-1), a merged stratification was formed that became the basis for the sample unit list required to compute the sample allocation.

Figure 3.3. Counties and hydrologic basins of California.

3.2.1 Agricultural Practice Stratification

The agricultural practice stratification was based on two general factors that are critical to both manual and digital classification of Landsat land use and field size. When defining land use for the purpose of estimating irrigated agriculture, there are several pertinent factors to be examined: (1) the presence/absence of any agriculture; (2) historically known or topographically defined areas of dryland vs. irrigated agriculture and (3) variations in agricultural cropping practices within a generally irrigated area (i.e. field crops vs. orchards). The problems caused by small field size affect the human analyst where detecting and identifying fields as well as accurately drawing boundaries becomes difficult and tedious and to the computer where the edge effect of mixed pixels and precise registration of acquisitions is critical.

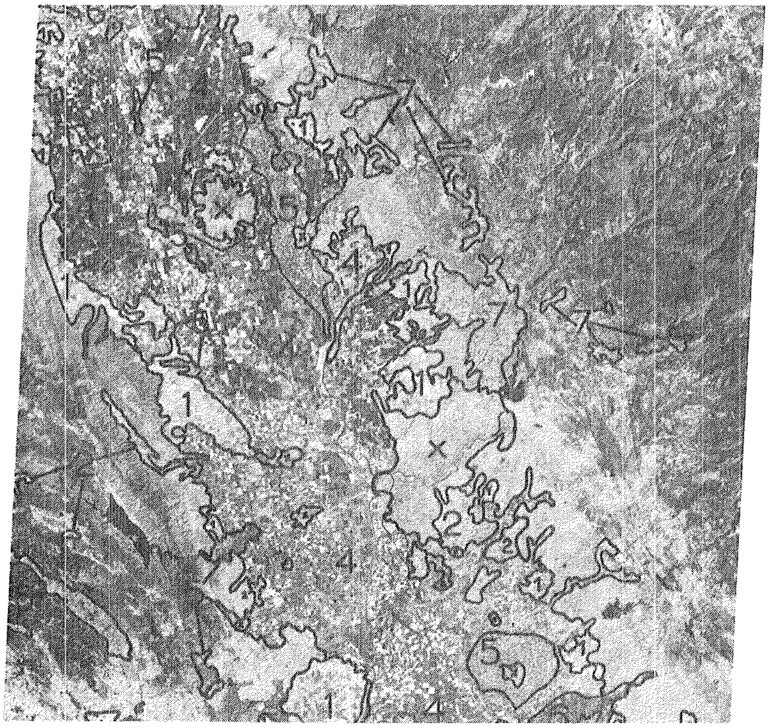
To minimize interpreter variability, the entire state (approximately 30 Landsat frames per date) was stratified by a single analyst into one of the seven strata described above. Since the minimum sample unit size was one square mile, areas less than that were not delineated (areas less than one square mile are subject to measurement for total irrigated acreage on Landsat but were considered too small to act as individual sample units). Stratification was done by overlaying clear acetate on 1:1,000,000 Landsat color composite transparencies and delineating the appropriate stratum. Multitemporal Landsat imagery was used to verify the consistency of the delineation. Since quite different agricultural practices and, therefore, quite different strata may appear similarly on any single date of imagery, it is very important to utilize the multitemporal capability and synoptic coverage of full frame Landsat to obtain an accurate, repeatable stratification. Figure 3-4 shows an example of the agricultural practice stratification used in the Sacramento Valley.

3.2.2 Small Grain and Vegetable Stratification

In order to direct the collection of field data two additional stratifications were necessary. Areas of small grain and vegetable cultivation have historically posed a problem in ground data collection due to: (1) early harvest of grains and subsequent plowdown, and (2) multiple cropping in vegetable areas. To ensure ground data acquisition at the optimum time, areas of grain cultivation and vegetable cultivation were stratified separately on 1:1,000,000 Landsat transparencies for each hydrologic basin.

After examining historical data on vegetable cultivation, boundaries of historical vegetable cultivation areas published by the California Crop and Livestock Reporting Service were transferred to the Landsat imagery. These boundaries were refined by reference to the Landsat imagery to account for land use changes and urban encroachment.

Small grain cultivation areas were delineated through analysis of 1976 through 1978 Landsat imagery. The grain areas were then classified into: (1) dryland grain farming; (2) areas of less than 21 percent grain; (3) 21-40 percent grain; and (4) greater than 40 percent grain.



Stratum Number	Stratum Description
1	Generally dry farmed
2	Field crop areas dominated by fields less than 16 hectares (40 acres) in size
3	Field crop areas dominated by fields less than 16 hectares (40 acres) in size with known high proportion irrigated
4	Field crops dominated by fields 16 hectares (40 acres) or larger in size
5	Orchards and vineyards less than 16 hectares (40 acres) in size
6	Orchards and vineyards 16 hectares (40 acres) or larger in size
7	Unusual agricultural areas

Figure 3-4. Agricultural practice stratification. Stratification similar to this was completed for the entire state and was used for the allocation of sample units that were ground checked. (Sacramento is located slightly southeast of center and marked with an "X").

Areas where multiple cropping occurs were examined through historical data and information from the county farm advisors. Most multiple cropping in California occurs in grain areas, where grain is followed by a field crop such as corn or beans, or in vegetable areas, where one vegetable crop follows another. The previous delineation of the grain and vegetable cultivation areas, therefore, included the majority of the multiple crop areas.

3.2.3 Formation of the Merged Stratification

Merging the agricultural practice, small grain and vegetable stratifications as well as locating administrative and exclusion areas was necessary before the sample unit list could be generated. Locating administrative boundaries, such as counties, exclusion areas (established wildlife refuges, cities), and assigning an access code is facilitated by reference to available maps. Since the agricultural practice and croptype stratifications were based on the spatial and spectral information provided by Landsat, it was felt that an appropriate base for the merging of these functions was a combination of 1:250,000 scale USGS topographic maps and 1:250,000 scale Landsat enlargements. Enlargements were made on a county basis, by reference to the USGS maps. These enlargements and the associated maps provided the base upon which the sample frame of 1.6 x 8.0 km (1 x 5 mile) units was created. The subsequent sample unit lists provided the population from which the ground data units were selected. In addition to providing the sample frame base, the combination of information available from the maps and enlargements was critical for accurate transfer of the sample unit boundaries selected for ground checking to the 1:24,000-scale (7.5') USGS maps used by DWR for field work.

For each of the 58 counties in California, the land use strata, grain cultivation boundaries and vegetable cultivation boundaries were enlarged from the original 1:1,000,000 scale to the 1:250,000 scale Landsat prints. By matching topographic features on both the original transparencies and the enlarged prints, accurate transfers of the boundaries were made. County boundaries were drawn from the 1:250,000 scale maps and overlayed on the merged strata boundaries. At this point all image defined agricultural phenomena were tied to the county base map. Hydrologic basin boundaries, provided by DWR, were transferred onto the overlays for those counties that were split into more than one hydrologic basin. Accurate location of this boundary was particularly important in those areas where the basin boundary crossed agricultural land, since misplacement of the boundary would result in farmland being transferred to the wrong basin. See Figure 3-5.

3.2.4 Generation of Sample Unit List

When the merging of the strata and the location of county, hydrologic and exclusion areas was complete, each county consisted of a set of irregularly shaped polygons defined by some combination of the strata. Each polygon was labelled indicating the appropriate land use stratum, the presence of vegetables, the presence and proportion of grain and the general accessibility of each polygon. The merged and annotated overlay was then placed over a gridded template of 1.6 x 8 km (1 x 5 mile) sample units and

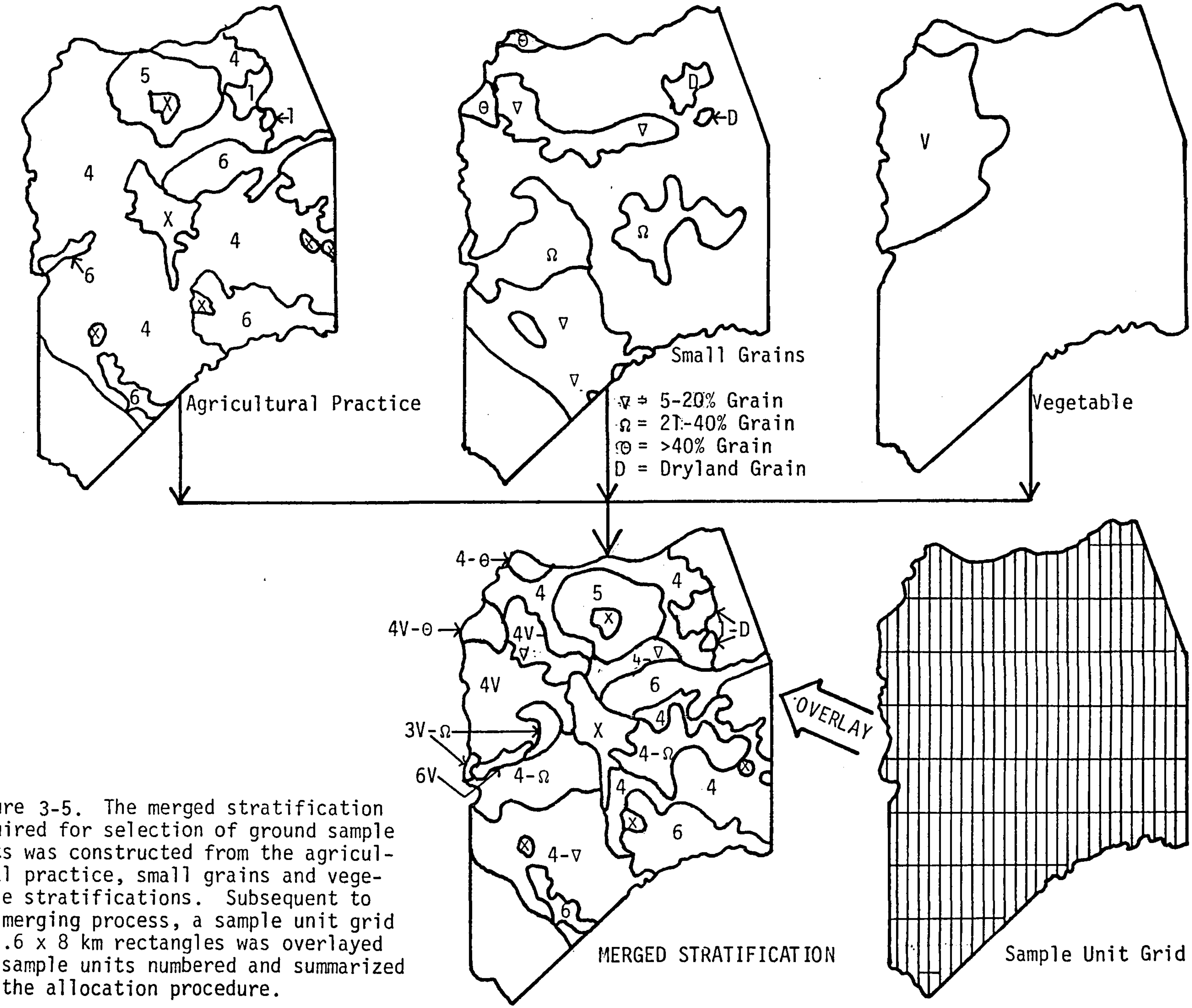


Figure 3-5. The merged stratification required for selection of ground sample units was constructed from the agricultural practice, small grains and vegetable stratifications. Subsequent to the merging process, a sample unit grid of 1.6 x 8 km rectangles was overlayed and sample units numbered and summarized for the allocation procedure.

the 1:250,000 USGS map. In those areas where the predominant field pattern was oriented north-south, the sample unit grid was placed to coincide with section lines. This was done to increase the ease and efficiency of the field data collection effort. In areas where the topography or historical land development caused the dominant field pattern to be oriented in other directions (i.e. Salinas Valley) the sample unit grid was placed so as to conform with the developed road/field pattern system. The sample unit grid was traced onto the county boundary overlay for all areas that fell within the stratified area.

Although a sample unit was nominally defined to be 8 km (5 mi) long, actual length varied from 1.6 to 11.2 kilometers (1 to 7 mi). Editing of the sample units removed those less than 259 hectares (640 acres) in area and those portions of units that were less than .4 kilometer (.25 mile) wide. Each sample unit was then numbered and placed in a sample unit list.

The information from the sample unit list was summarized in a table for each county. Similar summary tables were made for each hydrologic basin. The basin summary sheet was used to calculate average relative access cost within each stratum and the proportion of the basin represented by each stratum. This information, along with the number of sample units in each stratum (the population size) was used to compute the stratum sample sizes as described in Section 3.1.4.

3.2 Preparation of 7.5 Minute Quads for Ground Measurement

After the units to be ground surveyed were randomly selected, the boundaries of these sample units were visually transferred from the 1:250,000 scale overlay to USGS 7.5 minute quadrangles (scale = 1:24,000). Since DWR uses the 7.5 ' map as the base for their land use surveys the quadrangles were a logical, compatible choice for field crew use. After the sample unit boundaries were transferred, each unit was labelled with its sample unit number and its stratum label.

Finally, a recording form for the entire hydrologic basin was prepared. It contained a summary of those sample units that had been selected for each county within the basin including their stratum labels, whether or not the field work was to be done twice in the season (vegetable and grain strata were field checked early), and the name(s) of the 7.5' quadrangles involved. This summary recording form and the frosted overlays were sent to the appropriate DWR district offices having jurisdiction over each of the hydrologic basins. A map showing the approximate location of 637 sample units ground checked by DWR is shown on Figure 3-6.

3.3 LANDSAT MEASUREMENT

Earlier work in the estimation of irrigated acreage as well as in other agricultural projects has relied on the use of multitemporal Landsat data to monitor the dynamic agricultural environment. As in previous projects, the 1979 estimation procedure was based on the complete measurement of agricultural land on Landsat imagery at three critical time periods. The recommended acquisition windows were based on expected crop calendar, county cropping

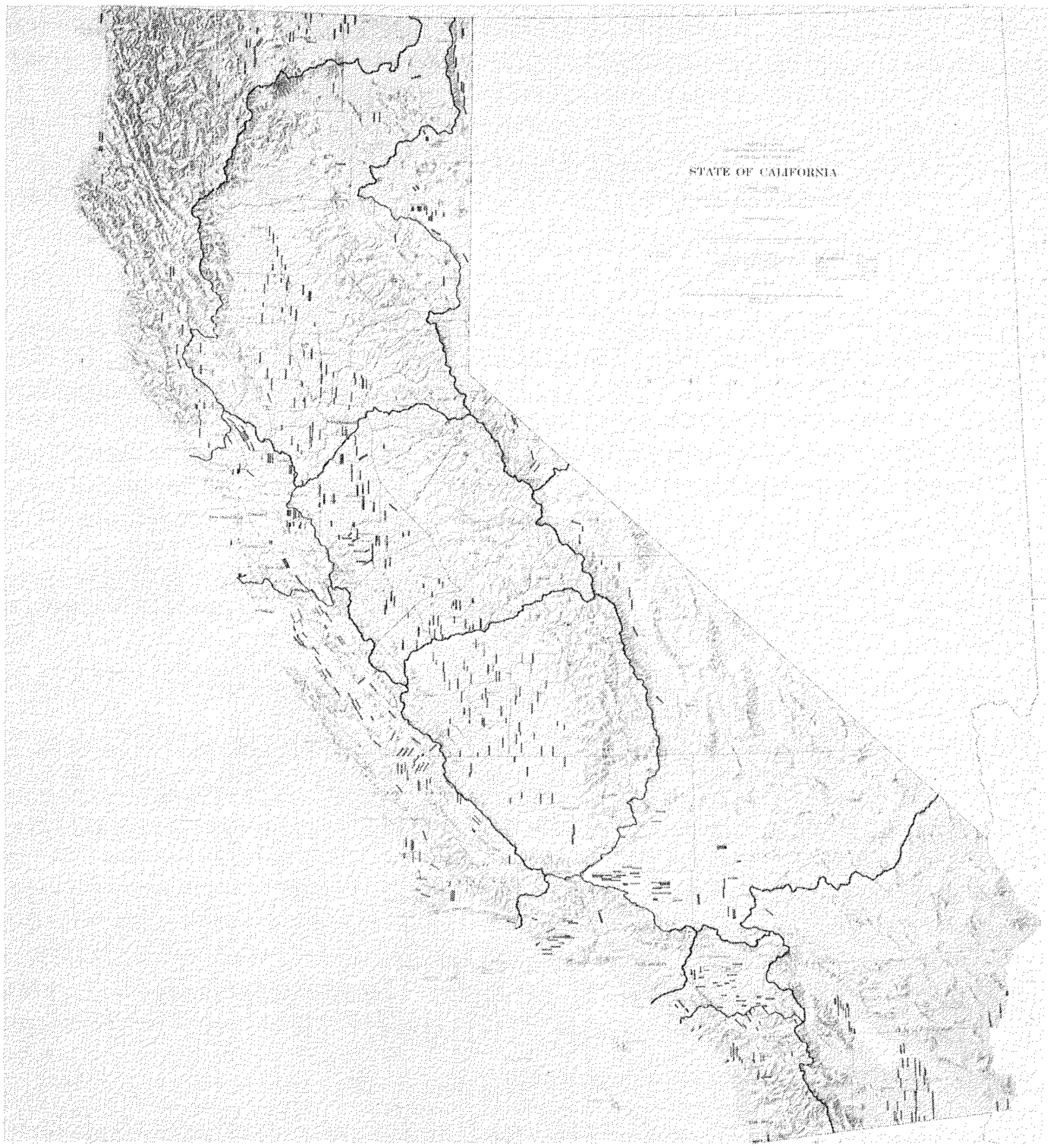


Figure 3-6. Distribution of ground (Phase II) sample units. Each of these 637 units was checked by DWR to determine the location of irrigated fields.

practices, historical cropping trends and consultation with DWR personnel. These three time periods were late July-early August when maximum canopy coverage is expected; May to monitor small grains; and late September-early October to aid in the detection of multiple cropped acreage.

3.3.1 Landsat Acquisition

The full frame Landsat 1:1,000,000 transparencies required for interpretation were ordered by NASA/Ames Research Center through the EROS Data Center (EDC) using normal ordering procedures. To obtain complete coverage of California thirty-three scenes are needed. Normally, California's virtually cloud-free summers over major agricultural areas provide ample opportunity to select from a large variety of acquisitions to obtain the optimal data set. In 1979 the satellite and ground processing problems often combined to severely limit or nullify any choice of acquisitions. Although certainly not the optimal date selection, three time periods of imagery were generally available for each of the counties.

3.3.2 Enlarging and Mosaicing Landsat Frames

In 1979, as in the earlier projects, measurement at the Landsat phase was done on 1:150,000 scale enlargements of each county. On a county basis, each available Landsat frame was evaluated for image quality (i.e. line drop, "smearing"), color balance, exposure and miscellaneous items such as cloud and smoke. Following this evaluation, the best combination of dates and frames was selected for enlargement.

When the enlargement was completed, each county was mosaiced together and mounted on stiff posterboard. Counties that would have a mosaiced size greater than approximately 750 cm x 1 meter were divided and mounted on separate boards. This size limitation facilitated handling and interpretation as well as storage.

3.3.3 Generation of Recording Forms

Forms for recording the interpretation done on the multitemporal Landsat enlargements were created for each county. To produce the form, the 1:150,000 scale county boundaries plotted by Caltrans were located on one of the completed mosaics for each county. The county boundary was then traced onto a second overlay; the originally plotted boundary was archived. The agricultural practice strata, exclusions and hydrologic basin boundaries were transferred from the 1:250,000 scale overlays by interpretation. The agricultural practice strata boundaries were necessary because (1) interpretation responsibilities were divided between analysts based on these strata boundaries, and (2) digitization of interpretation results was needed by stratum. Exclusion areas were also transferred from the 1:250,000 overlays; reference was also made to 1979, U-2, 1:130,000 scale CIR aerial photography to refine boundary placement. The hydrologic basin boundaries were needed for summarization of results and as a logical way to divide work between the Berkeley and Santa

Barbara campuses. Superimposed on the overlay, which was now a composite of county, agricultural practice strata, exclusion and hydrologic basin boundaries was placed a grid that defined the borders of 7.5 minute quadrangles. The grid was used as a mechanism for organizing interpretation, a unit for documenting the time required to perform interpretation and as a potential area for summarization and comparison of results with DWR's land use surveys.

3.3.4 Interpretation of Multitemporal Landsat Imagery

The interpretation logic and procedures for identifying irrigated land in California are basically the same that have been used in the past projects. The analyst is required to make a decision on whether a particular parcel of land is irrigated. To do this the analyst relies on a variety of image characteristics and logical expectations of the presence and appearance of irrigated land.

Providing the analyst with sufficient data to develop reasonable expectations is critical to accurate measurement at the Landsat phase. Prior to interpreting a particular county the analyst was given a variety of ancillary information upon which to build his expectations. These included (1) California Crop-Weather which is published on a weekly basis by the California Crop and Livestock Reporting Service and summarizes weather conditions over the state as well as land preparation, planting, growth condition and harvesting of field crops, fruit and nut crops, vegetable crops and livestock (pasture and range conditions). The information is summarized by region and provides the means for constructing year/regional specific crop calendars; (2) Agricultural Commissioner's crop reports for 1979 which list county acreage by crop type; (3) California-Arizona Farm Press that publishes weekly reports on all facets of agriculture in the West including land preparation planting, irrigation and water problems, pest and disease management, fertilization, plant variety performance, economic marketing and tax issues, legislation and harvesting; (4) California Grower and Rancher - a monthly published magazine on agriculture in California (written and published regionally); (5) 1979, U-2, 1:130,000, color infrared photography of the majority of agricultural land in California, and (6) antecedent DWR land use survey quads and summary statistics. Using all or a subset of the available data, the analyst builds a mental model of what he expects to see on the dates of Landsat imagery provided for each county.

Image characteristics traditionally used in manual photographic interpretation are exploited in the analysis of Landsat imagery. For the majority of the interpretation, the most critical characteristics are (1) pattern (is this an area of agricultural fields?) and (2) color (is this field the color expected for an irrigated field on the date being analyzed?). Other critical characteristics that analyst relies on are texture, shape of fields, and location of fields. These last three characteristics are particularly important when interpreting the mountain areas, along rivers and streams, (intermingled riparian vegetation) on the fringes of well developed agriculture and in areas of dispersed agriculture such as the foothills.

The interpretation procedure called for analysis to be done in a specific manner. The structure of the interpretation system was designed to (1) eliminate variability in the method interpreters use and (2) allow for a detailed evaluation of the separate parts of the analysis system. The procedure called for:

- Within each hydrologic basin assignment of a single interpreter was made to each stratum. An interpreter may analyze more than one stratum per basin, but no stratum should be interpreted by more than one analyst.

- Using the 7.5' grid as a base, interpretation proceeded on a quad-by-quad basis moving left to right and top to bottom.

- Interpretation was done on the mid-summer date first, the spring image second and fall date last. In strata where irrigated agriculture dominates, the analyst delineated areas that were not showing active vegetative growth in July/August. These areas were marked with a single dot. The overlay was then placed over the May image and the blocks marked with a single dot were checked; if these areas were interpreted as irrigated cropland in May, a second dot was added. The analyst then proceeded to the final date and checked the remaining singlely-dotted areas.

- Within each 7.5' quad the analyst recorded the time required to interpret each stratum on each date.

- Re-checked areas as necessary.

3.3.5 Digitization of Measurement Results

Upon completion of the interpretation, the results must be tabulated for input to MPHASE. The first step in this process was to locate the sample units that had been selected for ground checking. Accurate location is absolutely vital to the estimation procedure since the comparison of ground proportion irrigated to Landsat interpreted proportion irrigated "corrects" the estimate and provides the data needed to compute accuracy statements. Location was accomplished by reference to the 7.5' quadrangle maps with an overlay of the ground annotated sample units. By visual comparison of the ground data (field pattern) and map features (i.e. roads, canals, railraods) to the 1:150,000 scale Landsat enlargements, accurate location was possible.

The proportion of irrigated land was then calculated by digitizing the total area of each sample unit and the area that was irrigated. Each sample unit was digitized and recorded separately. The remainder of the interpretation was digitized by stratum within each county.

3.4 GROUND MEASUREMENT

For each of the 637 Phase II sample units, DWR district personnel made a field-by-field inspection to determine the presence of irrigation. Using 7.5' USGS quads with the plotted sample unit outlines as a base, field boundaries were drawn and each field coded. In many cases, detailed ground data including specific crop type mapping was done. At a minimum, the ground crews mapped parcels as irrigated or non-irrigated grain, safflower, field crop, pasture, other agricultural classes or lawn areas; fallow, farmsteads, feedlots, or dairies; native vegetation, water surfaces or unsegregated native vegetation; and six classes of urban. More than one visit was made to many of the units to verify multiple cropping.

When collecting the field data DWR often used their previously mapped land use survey quads (Figures 1-1 and 1-2) and the 35mm color aerial slides from which the maps were derived as aids for defining field boundaries. Color infrared 1:130,000 scale aerial photography flown by the U-2 during the spring and early summer of 1979 was also used extensively as soon as it was available. For a few units where access was particularly difficult, low altitude aerial observation of the unit provided the necessary information.

Each of these sample units was then tabulated by DWR and acreages output in a variety of forms: (1) by hydrologic basin - individual sample units listed by county (Figure 3-7); (2) by 7.5' quadrangles - sample unit(s) and county; (3) by county-cumulative summary of all sample units within the county; and (4) by DWR district-cumulative summary of all sample units mapped by the individual district offices. In total, DWR personnel ground checked (at least once) and tabulated approximately 520,400 hectares (1,286,000 acres) across the State.

3.5 ESTIMATE OF IRRIGATED AREA

Estimates of land irrigated at least once during 1979 were produced by county, by hydrologic basin, and statewide. These numbers, together with their associated estimates of error, would represent the principal product of an operational version of the inventory system demonstrated in this study. Tables 3-5a, 3-6a, 3-6b & 3-7 present the 1979 estimates. The estimates reported in those tables should not be considered 'simple numbers' to be accepted without question, but should in fact be treated as dynamic values that depend on sample frame, measurement system, and estimation procedure among other factors. Consequently a review of the major inventory procedures and assumptions is important in understanding and using the final estimates.

The characteristics of the sample frame, sample allocation, and measurement components of this Landsat-aided irrigated lands inventory system were described previously in Wall _et al_ (1980) and have also been summarized in the previous sections. The focus _here_ will be to explain how the resulting Landsat and ground measurement data were converted into irrigated land estimates.

STATE OF CALIFORNIA
THE RESOURCES AGENCY

79-74

CO. SERVICE AREA SUMMARY — LAND USE — IN ACRES — 1979 — LAND SAT-SOUTHERN DIST

DEPARTMENT OF WATER RESOURCES

REPORT DATE 02/04/80

① SPECIAL CONDITIONS: A ABANDONED, F FALLOW, M MILITARY, B BURNED, PARTIALLY IRRIG., P PARKS

② PLANTING SEASON: A SPRING, S SUMMER, C FALL, D WINTER

③ SOURCE OF WATER & TYPE OF DIVERSION: 1. STREAM-PUMP, 2. DRAIN-PUMP, 3. WELL-PUMP, 4. OTHER-PUMP, 5. STREAM-GRAVITY, 6. DRAIN-GRAVITY, 7. WELL-GRAVITY, 8. OTHER-GRAVITY

QUAD NUMBER	COUNTY CODE	HYDRO UNIT	SERVICE AREA CODE NUMBER	DIVERSION CODE	LAND USE SYMBOL: MAJOR CROP	LAND USE SYMBOL: INTER CROP	① SPECIAL CONDITION	② PLANT SEASON	③ SOURCE OF WATER & TYPE OF DIVERSION	IRRIGATED LANDS: INCLUDING FALLOW	IRRIGATED LANDS: EXCLUDING FALLOW	NONIRRIGATED LANDS: INCLUDING FALLOW	NONIRRIGATED LANDS: EXCLUDING FALLOW	SERVICE AREA AND OR HYDRO UNIT TOTALS	QUAD TOTALS
	13		0002		G					227	227				
										227*	227*				
					R					235	235				
										235*	235*				
					F					44	44				
										44*	44*				
					P					378	378				
					P		T			81	81				
										459*	459*				
										965**	965**				
					NW							1852	1852		
												1852*	1852*		
					NC							772	772		
												772*	772*		
												2624**	2624**		
					U							123	123		
												123*	123*		
												123**	123**		
														3712	
	13		0004		G					167	167				
					G		T			111	111				
										278*	278*				
					F					352	352				
					F		T			317	317				
										669*	669*				
					T		T			33	33				
										33*	33*				
					I					156	156				
										156*	156*				
										1136**	1136**				
					NC							390	390		
												390*	390*		
												390**	390**		
					U							8	8		
												8*	8*		
												8**	8**		
														1534	
	13		0016		G		T			511	511				
										511*	511*				
					R					202	202				
										202*	202*				
					F					1166	1166				
					F		T			509	509				
										1675*	1675*				
					P1					87	87				
										87*	87*				

Figure 3-7. DWR tabulation of individual sample units by county.

TABLE 3-5A: RESULTS OF 1979 STATEWIDE INVENTORY OF IRRIGATED LAND

BASIN	STRATIFIED ESTIMATE (PERCENT OF AREA IRRIGATED IN SAMPLE FRAME)	100 x ABSOLUTE S.E. AT 95% C.L.	RELATIVE S.E. AT 95% C.L.
NORTH COAST	53.52	2.04 *	3.81
SAN FRANCISCO	21.85	1.22 *	5.56
CENTRAL COAST	31.91	1.84 *	5.77
SOUTH COAST	45.79	2.88	6.28
COLORADO DESERT	82.15	1.40 *	1.70
SOUTH LAHONTAN	27.38[A]	3.81[A]	13.91[A]
NORTH LAHONTAN	58.73	2.68	4.56
SACRAMENTO	65.38	1.80 *	2.75
SAN JOAQUIN	74.78	2.55 *	3.41
TULARE	82.04	2.00 *	2.44
STATE	67.09	95: 0.89 99: 1.17	95: 1.32 99: 1.74

TABLE 3-5B: FOR COMPARISON ONLY - UNSTRATIFIED RESULTS OF 1978 STATEWIDE INVENTORY OF IRRIGATED LAND

BASIN	UNSTRATIFIED ESTIMATE (PERCENT OF AREA IRRIGATED IN SAMPLE FRAME)	100 x ABSOLUTE S.E. AT 95% C.L.	RELATIVE S.E. AT 95% C.L.
NORTH COAST	53.18	2.39	4.49
SAN FRANCISCO	21.19	2.20	10.37
CENTRAL COAST	32.58	2.15	6.60
SOUTH COAST	45.25	2.40	5.31
COLORADO DESERT	82.25	1.30	1.58
SOUTH LAHONTAN	27.38	3.81	13.91
NORTH LAHONTAN	58.73	2.45	4.17
SACRAMENTO	65.44	1.63	2.49
SAN JOAQUIN	75.16	2.31	3.08
TULARE	81.46	2.09	2.57
STATE	67.04	95: 0.88 99: 1.15	95: 1.31 99: 1.72

TABLE 3-5c: UNSTRATIFIED TASK I ESTIMATES CORRECTED FOR TYPE OF ALLOCATION TO STRATA

BASIN	100 x ABSOLUTE S.E. AT 95% C.L.
North Coast	3.11
San Francisco	1.96
Central Coast	2.27
South Coast	2.73
Colorado Desert	2.15
South Lahontan	3.81
North Lahontan	2.52
Sacramento	2.30
San Joaquin	2.34
Tulare	2.51
State	95: 1.09 99: 1.43

TABLE 3-6a: Stratified summary statistics for the area within the sample unit frame. Regression with factor 5.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53521	321070	6029	12238
San Francisco	191654	0.21852	41880	1108	2329
Central Coast	1380040	0.31906	440316	12572	25420
South Coast	598866	0.45787	274203	8510	17223
Colorado Desert	818231	0.82147	672152	5670	11447
South Lahontan	235626	0.27383	64522	4402	8977
North Lahontan	175456	0.58726	103038	2297	4695
Sacramento	3388466	0.65381	2215413	30327	60823
San Joaquin	2788914	0.74778	2085494	35419	71145
Tulare	4080305	0.82038	3347400	40721	81769
State	14257457	0.67091	9565489	64477	127020

TABLE 3-6b: Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346785
San Francisco	512	2586376	5623	2778543	47503
Central Coast	13475	5789056	24725	7182571	465041
South Coast	62133	6289499	63810	6950499	338013
Colorado Desert	28421	11852213	10830	12698865	682982
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117981
Sacramento	211744	13452904	37823	17053114	2253236
San Joaquin	542467	6704753	51098	10036134	2136592
Tulare	123300	5977461	42352	10181065	3389752
State	1000375	85067824	294255	100325656	9859744

Table 3-7. County estimates based on the weighted-unstratified model.

County	Inside Acres	Prop Irrig	Acres Irrig	S.E. (acres)	Excl Acres	Outside Acres	Ex & Out Irrig	Total Acres	Total Irrig
Alameda	35808	0.20437	7318	2163	512	482615	5015	518935	12333
Alpine	14071	0.30028	4225	1059	0	451350	688	465421	4913
Amador	10907	0.76512	8345	1055	0	368034	316	378941	8662
Butte	325022	0.71862	233567	22570	14982	720116	2835	1060120	236403
Calaveras	0	0.	0	0	0	655715	1053	655715	1053
Colusa	401541	0.77695	311977	27883	5510	327667	43	734718	312020
Contra Costa	115508	0.52964	61178	9370	0	370500	6034	486008	67212
Del Norte	14919	0.58190	8681	1155	0	627571	0	642490	8681
El Dorado	19616	0.11699	2295	1362	0	1112280	1928	1131896	4223
Fresno	1499786	0.83009	1244957	100606	50435	2057156	7489	3607376	1252446
Glenn	370991	0.67563	250653	25762	10061	461687	860	842740	251512
Humboldt	75657	0.32805	24819	5856	0	2209857	3967	2285514	28786
Imperial	595280	0.85959	511697	22252	9603	2275803	3287	2880686	514984
Inyo	31683	0.34954	11074	3338	0	6439435	2230	6471118	13305
Kern	1208298	0.78631	950097	90248	0	4005342	35980	5213641	986076
Kings	691649	0.80384	555975	53036	35962	157713	486	885324	556461
Lake	44004	0.30721	13518	3056	0	800563	1056	844567	14575
Lassen	118355	0.50783	60104	6080	4314	2878058	11565	3000727	71670
Los Angeles	138798	0.18266	25353	13901	4377	2372842	7565	2516017	32918
Madera	410092	0.65970	270538	39664	38775	913974	8756	1362842	279294
Marin	0	0.	0	0	0	377393	461	377393	461
Mariposa	0	0.	0	0	0	894479	288	894479	288
Mendicino	72478	0.32023	23210	5610	0	2155594	279	2228072	23489
Merced	725187	0.78865	571919	70140	181085	457383	15482	1363655	587401
Modoc	203746	0.74043	150860	8938	6940	2456047	11077	2666732	161937
Mono	60610	0.58197	35273	3816	0	1951702	5544	2012312	40817
Monterey	511204	0.45914	234714	47051	8007	1536931	1125	2056142	235839
Napa	48339	0.31740	15343	3069	0	453412	187	501751	15530
Nevada	30313	0.12329	3737	2105	0	588398	1235	618711	4972
Orange	37611	0.41968	15785	3513	0	470837	2748	508448	18533
Placer	193399	0.33930	65620	13430	406	756248	859	950053	66479
Plumas	62085	0.18388	11416	4311	0	1588495	2333	1650580	13749
Riverside	434233	0.52227	226787	23375	43352	4078911	22781	4556495	249568
Sacramento	356086	0.62629	223013	20144	21945	270650	1557	648681	224570
San Benito	124741	0.37823	47181	9961	0	740074	1428	864815	48609
San Bernadino	95242	0.48689	46372	5343	0	12718372	21759	12813614	68131
San Diego	121421	0.41910	50888	10523	25192	2587831	22068	2734444	72955
San Francisco	0	0.	0	0	0	30443	0	30443	0
San Joaquin	710511	0.75262	534745	68721	121672	72061	5148	904244	539893
San Luis Obispo	534008	0.10701	57144	49150	2135	1578858	9658	2115002	66803
San Mateo	4605	0.48262	2222	306	0	270995	827	275600	3049
Santa Barbara	155176	0.44387	68878	14282	579	1475465	10773	1631220	79651
Santa Clara	55357	0.49159	27213	3377	0	773207	720	828564	27933
Santa Cruz	40236	0.57639	23192	3703	2754	242212	1169	285202	24361
Shasta	112608	0.49302	55518	7819	0	2320549	8426	2433157	63944
Sierra	27840	0.68768	19145	1933	0	581838	634	609678	19780
Siskiyou	299259	0.64113	191864	22884	7006	3902853	16814	4209118	208678
Solano	278318	0.63448	176587	16788	78822	222472	2558	579612	179145
Sonoma	118844	0.21798	25906	7315	0	887805	3642	1006649	29548
Stanislaus	489547	0.82526	404004	47349	191784	277631	5160	958961	409164
Sutter	333264	0.85744	285754	23142	51784	0	0	385048	285754
Tehama	218727	0.49919	109186	15188	8052	1641421	4766	1868200	113952
Trinity	0	0.	0	0	0	2028052	802	2028052	802
Tulare	929336	0.83065	771953	71261	36903	2133106	4163	3099346	776116
Tuolumne	0	0.	0	0	0	1436630	641	1436630	641
Ventura	152542	0.64299	98083	14247	12407	991918	3602	1156868	101685
Yolo	471407	0.71555	337315	32735	19351	155340	146	646099	337462
Yuba	127194	0.71585	91052	8832	5667	273928	2240	406789	93292
State	14257459		9558250		1000375	85067840	294255.	100325656	9852506

3.5.1 Data Screening

Before the sample unit measurement data could be used to produce estimates, they were screened to detect errors in locating sample unit boundaries, interpreting irrigation status, digitizing field boundaries, and recording data. Figure 3-8 illustrates this process.

The first step in this procedure was to convert digitized or tabulated irrigated count data from individual Landsat and ground sample units to proportional values. Resulting irrigated proportion values were then compared from spatially-matched Landsat and ground sample units. An error check was initiated when such spatially-matched units were found to differ by more than a previously specified constant. The size of this constant was arbitrary. A value of 10 percent of sample unit area was chosen as a 'reasonable' constant for the 1979 demonstration under the assumption of high expected Landsat-to-ground correlation. As experience is gained, previous information on error bounds can be used in formal tests for outliers.

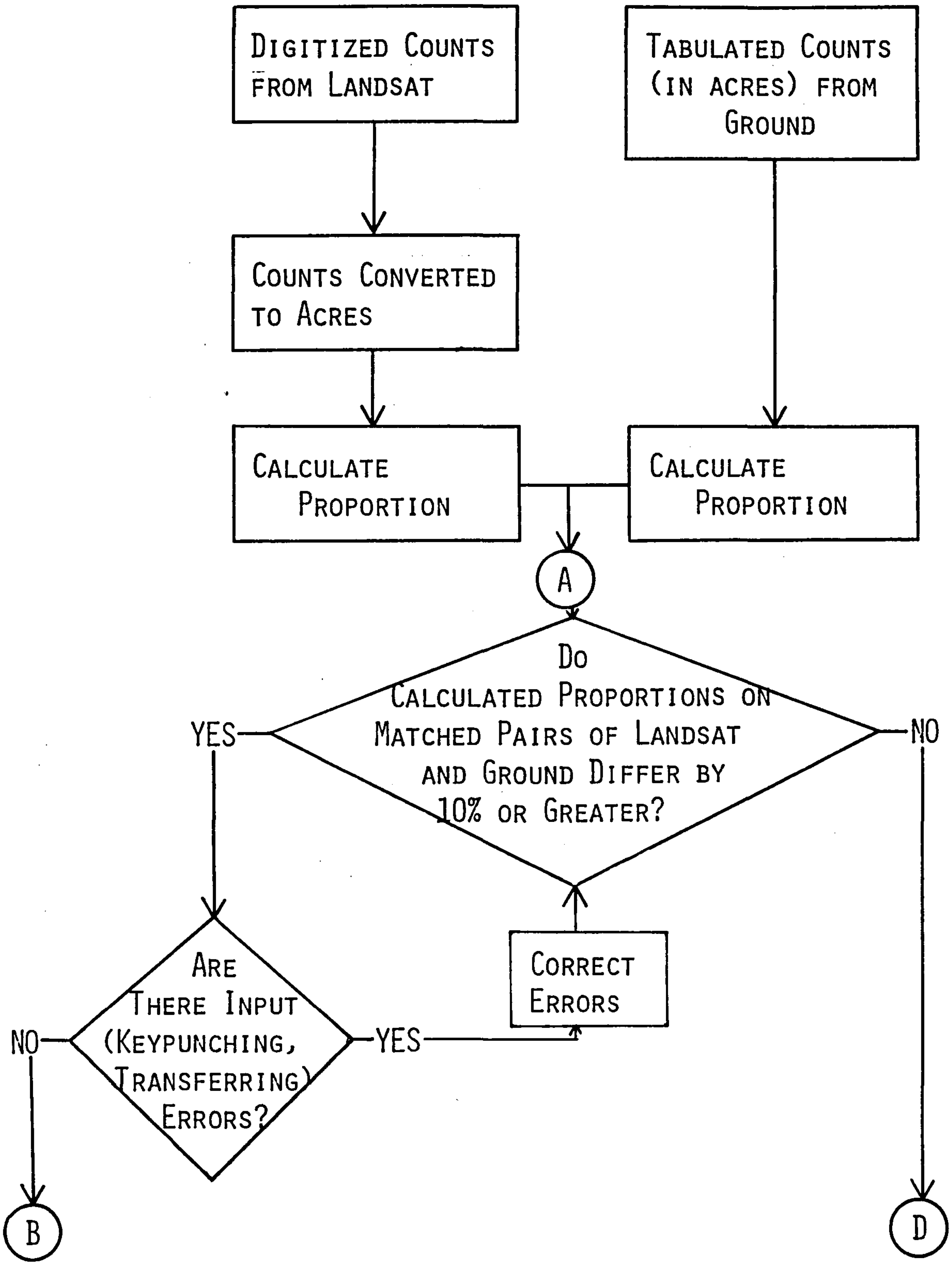
The initial step in an error check consisted of a verification of keypunching or count accumulation errors on data forms.* If detected these errors were corrected and the Landsat and ground sample units were again compared using the 10 percent rule. If no errors of this type were detected for sample units differing by 10 percent or more, then sample unit boundary locations were checked. To facilitate the location check, the boundary of the ground sample unit in question was plotted on color infrared highflight photography. This was accomplished with use of the 1:24,000 acetate overlay for the ground unit. Visual comparison of this boundary as seen on highflight and on a corresponding Landsat image and sample unit overlay enabled detection of Landsat sample unit location errors. When these errors occurred, the Landsat sample unit was moved to its proper position and irrigated/non-irrigated area redigitized. Field boundaries within the ground sample unit were also checked against the highflight aerial photography. If errors were detected, ground data boundaries were redigitized only after consultation with and approval of DWR personnel. On very rare occasions, the ground sample unit itself was found to be mislocated. When this occurred, the ground sample unit location was declared the proper location and the Landsat sample unit moved accordingly.

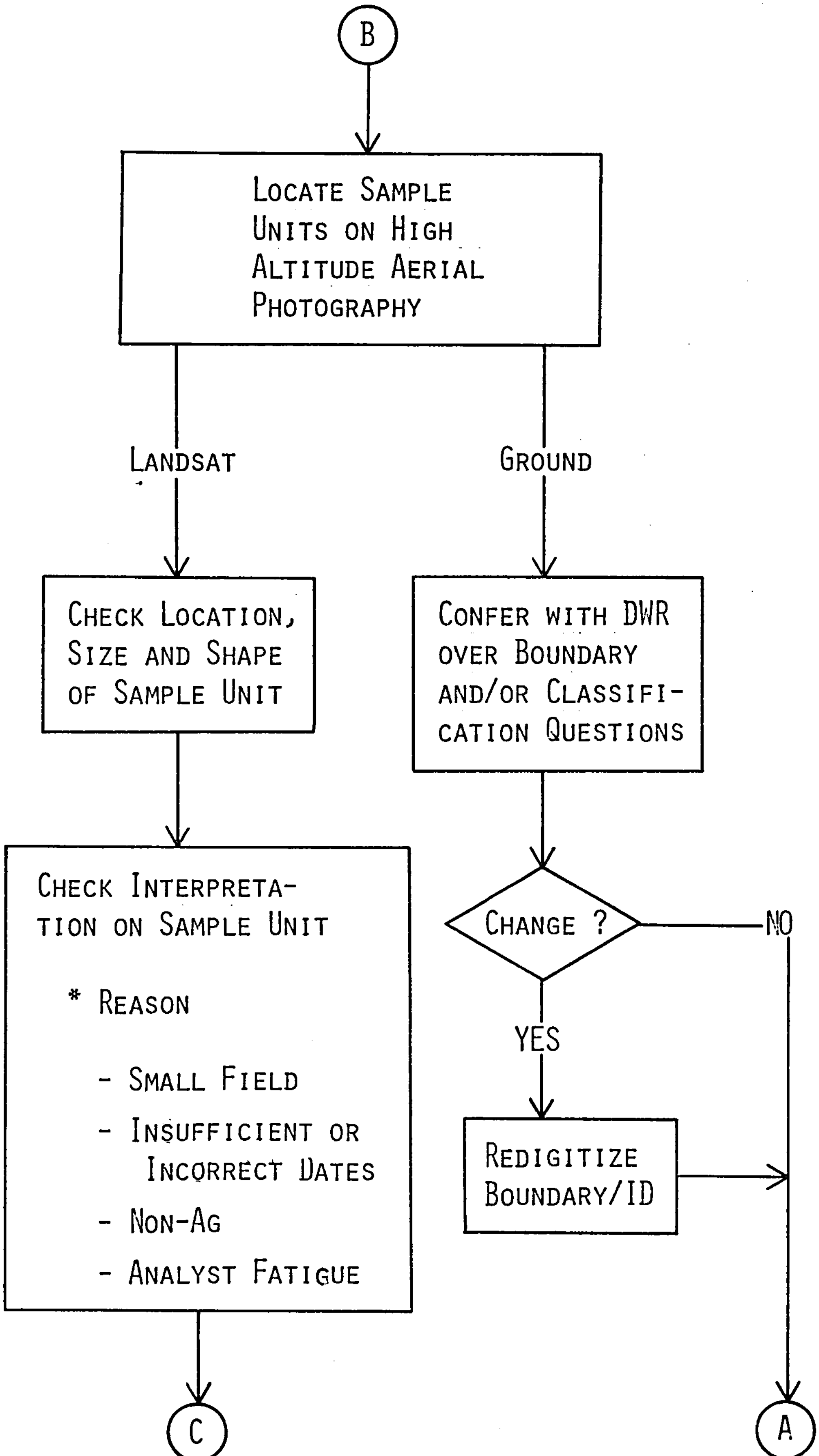
The final step in the sample unit screening process was a check of the irrigation classification on a field-by-field basis. Questionable ground field labelling was detected by reference to the presence of vegetation indicated on the color infrared highflight photography used to check location. Changes in ground field labels were made only after consultation with and the approval of DWR personnel. Landsat sample unit irrigated/non-irrigated labelling was checked by reference to the multidate Landsat imagery itself as well as to the highflight photography. If labelling errors were systematic (i.e. consistently made and of the same type over sample units in a given stratum) then the entire stratum was reinterpreted and redigitized - including the sample unit areas. On the other hand, if labelling errors were found to have a random pattern, then no changes were made to either the sample units or to the stratum.

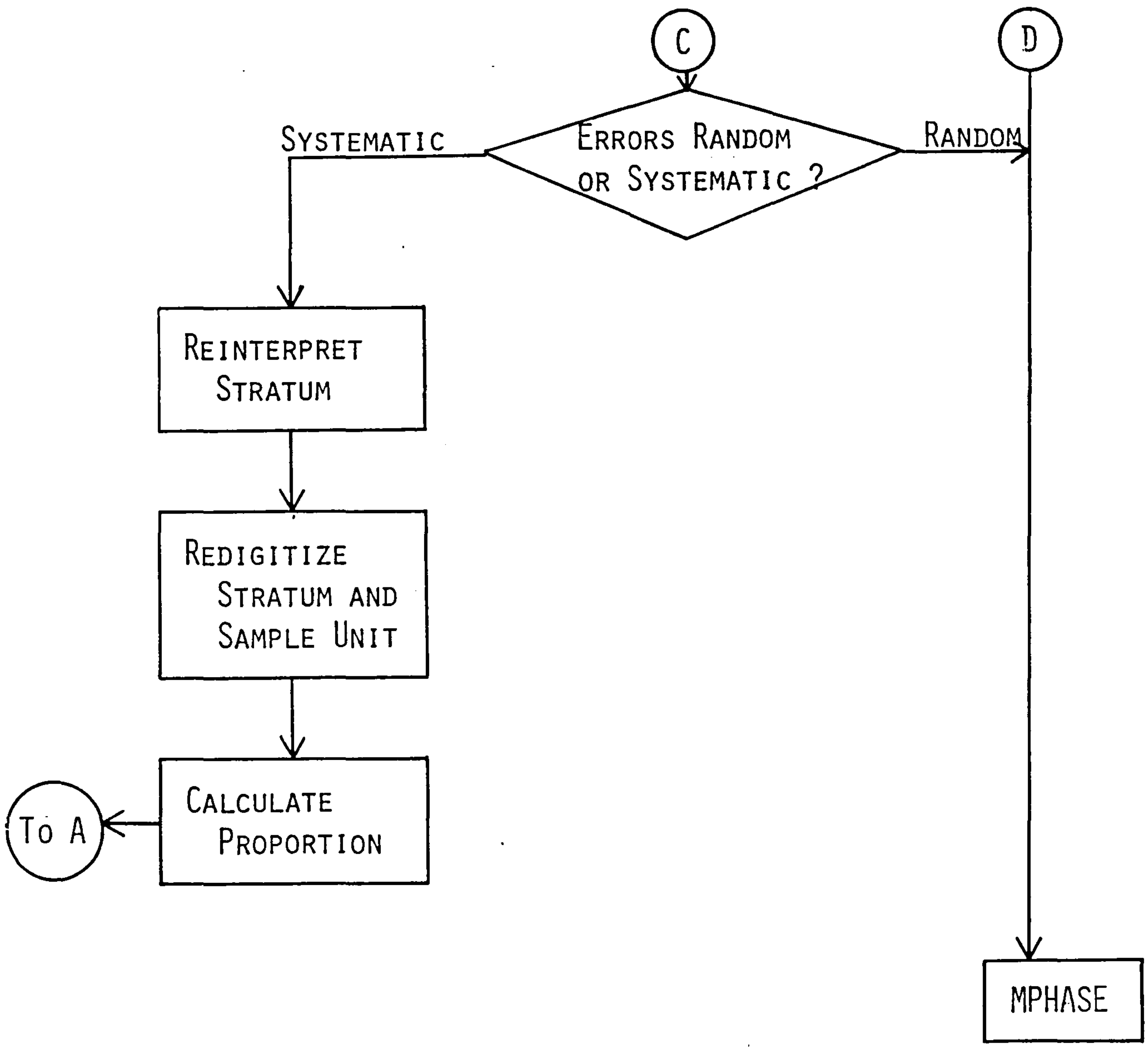
---

* This step should always be included, regardless of whether the error limit is exceeded.

FIGURE 3-8. TASK I - SCREENING MEASURED DATA




B
LOCATE SAMPLE UNITS ON HIGH ALTITUDE AERIAL PHOTOGRAPHY
LANDSAT
GROUND
CHECK LOCATION, SIZE AND SHAPE OF SAMPLE UNIT
CONFER WITH DWR OVER BOUNDARY AND/OR CLASSIFICATION QUESTIONS
CHECK INTERPRETATION ON SAMPLE UNIT
* REASON
- SMALL FIELD
- INSUFFICIENT OR INCORRECT DATES
- NON-AG
- ANALYST FATIGUE
CHANGE ?
NO
YES
REDIGITIZE BOUNDARY/ID
C
A


C
D
ERRORS RANDOM OR SYSTEMATIC ?
SYSTEMATIC
RANDOM
REINTERPRET STRATUM
REDIGITIZE STRATUM AND SAMPLE UNIT
CALCULATE PROPORTION
TO A
MPHASE

After sample unit data were screened and corrected where appropriate, they were entered into a data file for use in producing estimates. Other inventory data were also checked. Digitized count data for stratum-wide irrigated/non-irrigated area, area irrigated outside the sample frame, and area of exclusions were checked for completeness and proper scaling of digitizer output on a county-by-county basis. Summing of these areas over counties within basins was also verified. County and basin boundary locations and digitized total area figures were checked as well.

3.5.2 Basin and Statewide Estimates of Irrigated Area

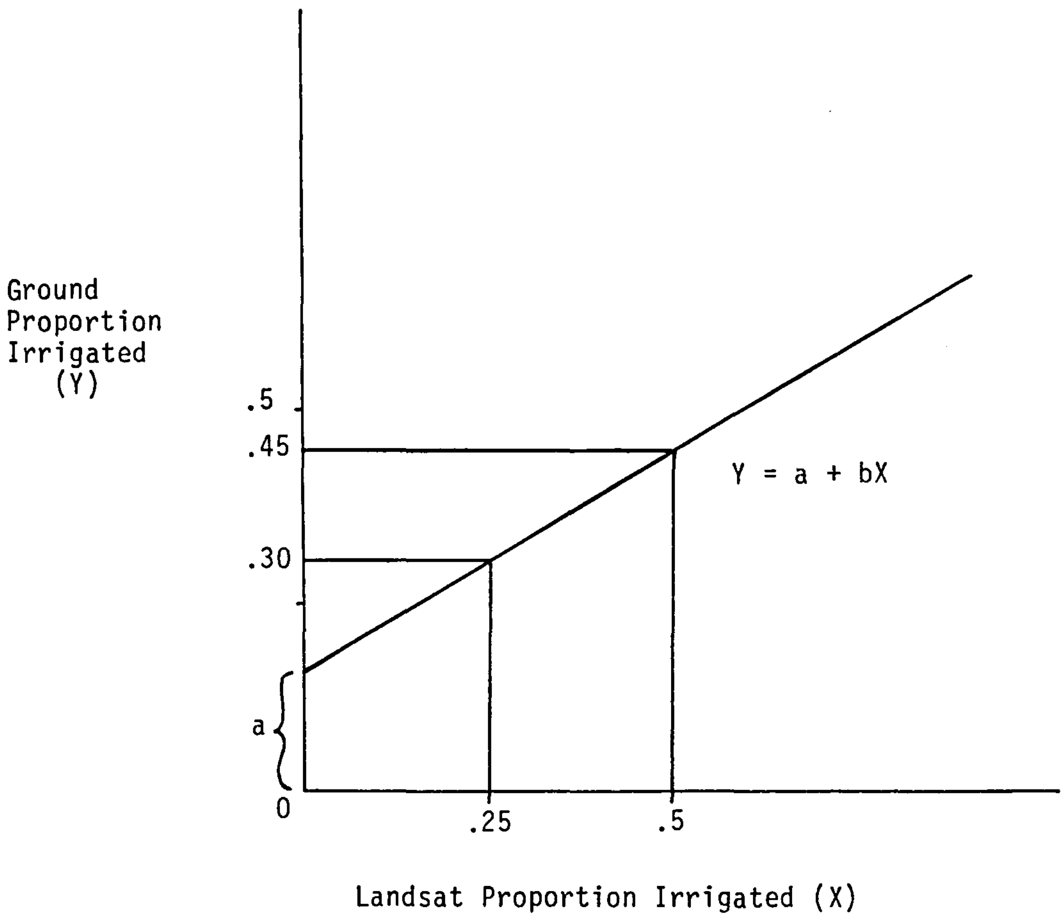
An Overview of the Estimation Approach

Year-specific estimates of irrigated land within a hydrologic basin were produced in the following way. First, for areas within the sample frame, a relationship was established between the screened Landsat measurement of irrigated land by sample unit and the corresponding ground measurement. Then, given the relatively inexpensive basin-wide Landsat measurement of irrigated land, the relationship between Landsat and ground data was used to 'predict' or estimate the basin-wide ground value of irrigated land. This approach is illustrated in Figure 3-9. There the relationship between Landsat and ground observations is represented by a straight line. Thus, for example, if the proportion of total sample frame area irrigated at least once was determined to be 50 percent based on digitization of Landsat interpretation, use of the straight line in Figure 3-9 would give a basin-wide ground estimate of approximately 45 percent. To see this result, project a vertical line from .5 (i.e. 50%) on the horizontal Landsat axis to the solid straight line. Then project a horizontal line from the point just intersected on the straight line to the vertical ground axis. A value of .45 should result.

Application of this conceptually simple method of irrigated land estimation is predicated on the probability sampling approach employed in this study. For the area estimation problem, probability sampling can be described as a procedure whereby 1) a population of area elements (say individual acres) is grouped into sample units; 2) each such unit is assigned some probability of selection; and then 3) a sample of these units are then selected at random according to those probabilities for measurement of irrigated area. Given this objective method for establishing and selecting sample units, estimates of irrigated area having known error properties can be constructed. In particular, the Landsat-to-ground relationship can be used to produce an estimate of basin-wide irrigated land that 1) tends to be centered on the true value of ground-measured irrigated land in the basin, and for which 2) an objective statement of error can be given. Measurement data obtained through non-probability (or subjective) sampling does not enable specification of estimation procedures with these two properties.

The within sample frame estimation procedure just described produced the values for irrigated proportion shown in numerical column 1 (counting left to right) in Table 3-5a, and the values of irrigated acreage shown in column A3 of Table 3-6a. A small percentage (approximately three percent statewide) of irrigated land did, however, lie outside the sample frame. This irrigated acreage was located in areas excluded from the interior of the sample frame (e.g. urban zones and marsh areas) or in small pockets of agricultural land outside the

Figure 3-9: Schematic for Linear Regression Line



a = regression line intercept

b = regression line slope

sample frame (e.g. along stream courses in foothill or mountain areas). Measurements of irrigated land in these areas were made only on Landsat and not calibrated with ground data. Figures for this category of irrigated acreage are reported in column B3 of Table 3-6b. Total basin-wide irrigated acreage was computed by adding the values shown in columns A3 and B3 and is reported in the right-most column of Table 3-6b. Plus or minus values for error, shown in columns 2 and 3 of Table 3-5a and columns A4 and A5 of Table 3-6a, were based on within-sample frame estimates only. No statement for error on irrigated land outside the sample frame was available due to the absence of an established Landsat-to-ground measurement relationship.

The Specific Procedure: Regression Estimation

Given this general overview of the estimation process, we may now proceed to a more formal description of the procedure. Within-sample frame estimates of irrigated land were produced using a regression estimator. The estimator itself was an equation that related Landsat measurements to ground measurements of irrigated land. It was constructed by 'fitting' a straight line through the spatially paired Landsat measurements (random variable X) and ground measurements (random variable Y) of irrigated proportion. This fitting was accomplished by minimizing the squared deviations between the line and the actual paired (x,y) values. Having established this line (e.g. the solid line shown in Figure 3-9), the basin-wide Landsat measurement for irrigated land was substituted into the straight line formula to produce an estimate of basin-wide ground irrigated land. Use of the simple linear regression estimator assumed that 1) the relationship between X and Y was in fact a straight line; 2) the variance of Y about the regression line was constant over the range of X; and 3) the errors (distance between the line and the actual (x,y) observations) were independent of one another. In addition, it is assumed that the X's are fixed when computing the regression slopes and intercepts, and that the expected error for any value of X should be zero.

The expectation that these assumptions would hold was based on results of earlier pilot studies (Wall, et al, 1977, 1978). Selection of the regression estimator over other possible linear estimators was in turn based on results of two studies reported last year (Wall, et al, 1980). The first of these, a Monte Carlo exercise, compared ratio (straight line relationship forced through the origin) and regression estimators over a range of sample sizes, individually by stratum and for strata combined. A second study compared the relative sampling efficiency of simple random sampling, biased and unbiased ratio sampling, and regression sampling over a range of sample sizes under expected inventory conditions. Both studies indicated the general superiority of the regression estimator. The unbiased ratio estimator was found to be superior at very low sample sizes, but was not used in producing the final 1979 inventory estimates due to a problem associated with its use when X and/or Y were close to zero.

Two major regression estimation approaches were used with the 1979 survey data. The first, or primary, procedure for producing irrigated land estimates was that of stratified regression estimation. If the relationship between X and Y is linear and if the slope of the regression or its intercept with the Y (ground) axis differ significantly between land use strata, then a stratified estimator should produce less biased stratum-specific estimates of irrigated land and should (depending on sample size) give a lower basin-wide sampling error. Since a principle assumption (to be tested) in the 1979 inventory was that the land use strata were in fact different in this respect, the use of the stratified regression estimator was appropriate.

Sample unit observations (measurements) of X and Y were reported as the proportion of sample unit irrigated as opposed to area irrigated. This was done in order to eliminate error due to differences in computing total area of spatially-matched Landsat and ground sample units. These observations were in turn weighted by size (area) of the sample unit.* Motivation for weighting was based on the concern that smaller sample units would be subject to proportionally higher measurement error rates due to ground registration, digitizing, and, in some cases, interpretation error. The effect of sample unit weighting was to give each acre sampled equal weight in determining the regression line. If the error or variance about the regression line did in fact decrease with increasing size of sample unit, then weighting should transform the scatter of paired (x,y) data points to a more linear condition and thereby produce a lower sampling error.

A regression line of the form:

$$\hat{y} = a + bx$$

was then fitted through the scatter of matched (x,y) sample unit size - weighted observations in each land use stratum. The resulting estimates for regression line intercept (a) and slope (b) minimized the sum of squared deviations between actually observed y's and the corresponding, estimated values ($\hat{y}$) taken from the regression line. An irrigated proportion estimate for an entire stratum was then obtained by substituting the measurement of Landsat irrigated proportion (X) for that entire stratum into the regression model above. These estimates represented the proportion of the area <u>within the sample frame</u> within the given basin which was irrigated at least once during 1979.

A stratified estimate of irrigated proportion for the area within the sample frame within a <u>given basin</u> was produced by forming a weighted average of stratum estimates, where the weight assigned to each stratum represented the proportion of basin-wide sample frame area included inside the given stratum. An estimate of irrigated proportion <u>within the statewide</u> sample frame was formed in similar fashion. Weights were assigned to the estimates of proportion irrigated for each basin according to the relative sample frame area within each. A weighted sum of proportions was then formed to produce the statewide, stratified estimate.

Estimates of acreage irrigated were obtained by simply (1) multiplying stratum-specific estimates of irrigated proportion times the total area within the sample frame in that stratum; and then (2) summing the resulting acre-

* Sample unit size expressed as the area in the given sample unit relative to the average area for sample units in the same land use stratum and basin.

ages over strata to produce an acreage irrigated figure for each basin. To this basin total was added the irrigated acreage identified through manual, interpretation of Landsat imagery in areas outside of the sample frame.** In most cases this represented a small percent of the total irrigated acreage within a given basin.

An estimate of the total statewide acreage irrigated at least once during 1979 was produced by direct addition of the totals for each basin. Separate figures for statewide within-sample frame and outside-sample frame irrigated acreage were also obtained and reported.

The second major approach used to estimate area of irrigated land was that of unstratified regression estimation. This procedure will be the most efficient of the two approaches, that is give the lowest sampling error for fixed sample size, when the regression line slopes and intercepts do not differ significantly between strata. The unstratified estimate was constructed by pooling the spatially-matched (x,y), weighted observations from each separate stratum within a given basin into one 'grand' stratum. Each observation was weighted according to the size of the given sample unit relative to the average sample unit size in the entire basin. A regression line was then fitted through these paired (x,y) observations and the resulting relationship was used to estimate ground irrigated area in the manner described above in the discussion of Figure 3-9.

The resulting basin estimates of irrigated proportion were multiplied by the area within their corresponding sample frames to produce estimates of within-sample frame irrigated acreage. To this total was added the acreage identified by Landsat interpretation as irrigated outside the sample frame. Statewide estimates of irrigated proportion and acreage were obtained in a manner analogous to that described for the stratified case.

Appendix IA presents the mathematical formulation of the estimation problem and the equations used to produce the basin and statewide estimates of irrigated land. Review of this appendix is highly recommended for a complete understanding of the method and assumptions involved in the estimation process.

3.5.3 Basin and Statewide Estimates of Error

Error associated with the regression estimates of irrigated area described in the last section can be separated into two components. These are known as bias and sampling error. Bias can be defined to be the difference between the true value of irrigated area for the administrative area in question (the area within the sample frame within a given basin in this case) and the expected value of the estimator when averaged over all possible samples of given size. In the general linear model if the relationship between X and Y is truely a straight line over the range of X, then the regression estimator of proportion or area will be unbiased (Cochran 1977). Data from previous pilot studies (Wall et al 1977, 1978) and from Y versus X plots for the 1979 inventory (see the Evaluation section) indicated that the relationship was

---

** Includes wildland, range, and urban areas outside the contiguous boundary of the sample frame, and areas excluded from the interior of the sample frame - e.g. urban areas, urban fringe, wildlife refuges, small areas of rangeland, etc.

very near linear, and therefore the bias was expected to be small relative to the size of the sampling error. Consequently, the regression estimators were assumed to be minimally biased in this inventory demonstration.

Sampling error is the error introduced by measuring only a portion of ground units instead of the whole area as would be done in a standard California Department of Water Resources mapping survey. Estimates of sampling error were obtained for the regression estimators described in the last section. These estimators were based on the formulas for the variance of paired (x,y) observations about the appropriate regression line. Appendix IB presents these formulas and illustrates their use in producing the error values reported in Tables 3-5a and 3-5b.

Three different expressions of sampling error are provided in those tables. The first, absolute error in percent, represents error expressed as a percent of the area in the sampling frame. It was computed by multiplying the estimated standard error (the square root of the regression variance), which is expressed in units of percent of sample frame, by a statistic (Student's t) used to give the desired probability with which the true value of irrigated proportion should be covered by the interval

$$\hat{p} - (\text{std error} \times t) \leq p \leq \hat{p} + (\text{std error} \times t)$$

where p represents the estimate of proportion irrigated, expressed in percent (i.e. x 100) for a given area - e.g. a basin or statewide. The product (std error x $t_{1-\alpha}$) is termed the absolute sampling error at the $1-\alpha$ level of confidence, or alternatively, the $1-\alpha$ percent confidence interval half-width for absolute error expressed in percent. For each basin, α was set to 5 percent so that the confidence interval covering p shown in the equation above was expected to include the true value of p 95 times out of 100. This statement assumes that repeated estimates of p will tend to occur according to a normal (bell shaped) distribution centered on the true value of p. Absolute sample error values are given for stratified sampling in column #2 of Table 3-5a (counting left to right) and in column #2 of Table 3-5b for unstratified sampling. Statewide error is reported at the 99 percent level of confidence as well as at the 95.

If error as a percent of the irrigated proportion estimate is of interest, then the second expression for error will be of value. This expression, termed relative error in percent was defined by dividing the estimated standard error by the estimate ($\hat{p}$) itself, then multiplying the result by the appropriate Student's t - statistic. Thus the $1 - \alpha$ percent confidence interval half-width expressed as a percent of the estimate was (std error $\div \hat{p}$) x $t_{1-\alpha}$ x 100. This expression was also termed the relative sampling error in percent at the $1-\alpha$ level of confidence. Estimates for relative error are shown in column #3 of Tables 3-5a and 3-5b.

A third expression for error was reported in column A5 of Table 3-6a. This error represented absolute error expressed in acres and was computed by multiplying the estimated standard error times the total number of acres in the basin sample frame, and the resulting product times the appropriate Student's t - statistic. Thus column A5 represents the $1-\alpha$ (95) percent confidence interval half-width in acres. The full confidence interval, centered on the estimated acreage irrigated within the sample frame, will be expected to include the true value of irrigated acreage 95 times out of 100 if the estimates are normally distributed about the true value.

It must be emphasized that the errors provided in this report apply only to irrigated proportion or acreage estimates within the sample frame. No error statements are available for the Landsat measurement of irrigated acreage outside the sample frame. The additional error associated with this outside acreage is likely small on a statewide basis, as only three percent of the irrigated acreage fell in this category. Statistically valid statements of error for this acreage will be available if such areas can be economically included within a sample frame in future inventories.

3.5.4 County-Specific Estimates of Irrigated Land and Associated Error

The basin regression estimators were also used to produce estimates of the area irrigated within individual counties. This approach was used because the per county ground sample sizes were generally too small to develop stable county-specific, Landsat-to-ground regression equations.

The primary method for generating county estimates was as follows:

1) irrigated area within a given county within the sample frame of a given basin was determined by digitizing the Landsat interpretation;

2) total area of the sample frame associated with a given basin in the county was also determined by digitizing;

3) these data were then used with basin-specific, unstratified regression equations developed from weighted observations* to predict irrigated area within the county;

4) an estimate of total irrigated area within a given county was formed by adding the predicted value of irrigated area within the sample frame, obtained in the previous step, to the Landsat-measured irrigated area outside the sample frame; and

5) standard error was computed for the county, within-sample frame estimate of irrigated acreage using the formula for the variance of a predicted value.

The equations necessary to implement this procedure were similar in form to those used in the stratified basin estimation problem. These equations are presented in Appendix IC. Errors reported in Table 3-7 represent error primarily due to sampling and were based on the formula for the sample variance of values predicted by regression. That is, since the regression equations used to estimate county irrigated proportions were not developed exclusively on the given county data set, an additional component of variance (for predication) was added to the usual regression expression for variance. Furthermore, errors for county estimates were reported under the assumption that the regression estimates of irrigated area were minimally biased. The validity of this assumption will be examined during the coming year's evaluation.

---

* observations weighted by the area within each sample unit relative to the average area within the sample units in the given basin.

County error values given in Table 3-7 are cited at the one standard error level (i.e. Student's t was set equal to one). Thus the true value of irrigated acreage was expected to fall within plus or minus one standard error (for regression with prediction) of the estimated value 68 times out of 100 - assuming a series of estimates themselves would be distributed normally and centered on the true value. As in the case of the basin estimates, no error statement was available for irrigated land outside the sample frame.

3.6 RESULTS FROM THE 1979 STATEWIDE INVENTORY OF IRRIGATED LAND

3.6.1 Basin and State Estimates

Basin and statewide results from the 1979 California Irrigated Lands APT are shown in Tables 3-5a, 3-6a and 3-6b. Table 3-5a presents results in terms of estimated irrigated proportions and associated absolute and relative errors for the area with the sample frame. Table 3-6a provides corresponding results in acres and Table 3-6b gives irrigated acreage totals for areas outside the sample frame as well as within. All figures reported in these tables were based on stratified regression estimation. This was designated the primary design for this inventory.

Inspection of Table 3-5a shows that, statewide, the percentage of land within the sample irrigated at least once during 1979 was estimated to be 67.1 percent. On a basin basis, there was a tendency for lower proportion irrigated in the cooler, moister hydrologic units (located in the coastal and northern portions of the state). Highest percentages irrigated occurred as expected in desert and mid to southern Central Valley areas.

An absolute measure of irrigated area is provided by the irrigated acreage figures shown in column A3 of Table 3-6a and in the last column on the right in Table 3-6b. Column A3 represents the acreage irrigated in the sample frame, while the column cited in Table 3-6b gives the total acreage over areas within and outside the sample frame. Reference to the last row shows that the within-frame estimate of statewide irrigated acreage was 9.57 million acres. Addition of the 'outside' irrigated acreage to that figure brought the total estimated statewide irrigated acreage in 1979 to 9.86 million acres, a difference of approximately three percent.

Estimates of total basin irrigated acreage given in the last column of Table 3-6b are seen to range from 47.5 thousand acres in the San Francisco hydrologic unit to nearly 3.4 million acres in the Tulare unit. Of the total 9.86 million acres irrigated statewide, 78.9 percent (7.78 million acres) was located in the three basins encompassing the Central Valley. These basins were the Sacramento, San Joaquin, and Tulare hydrologic units. The four coastal basins - North Coast, San Francisco, Central Coast, and South Coast - contained 12.1 percent (1.20 million acres) of the total irrigated acreage in the state in 1979. The remaining irrigated acreage, 9.0 percent (0.88 million acres), was located in the desert/mountain North Lahonton, South Lahonton, and Colorado Desert hydrologic basins.

Inspection of the second numerical column (counting left to right) in Table 3-5a shows that in all basins the confidence interval half-width, expressed as a percent of the sample frame, was within plus or minus five percent, 95 percent of the time. In fact, the highest absolute error was 3.81 percent* in the South Lahonton unit and the next highest was 2.88 percent in the South Coast basin. Statewide absolute error was, under the assumptions presented in the previous section, estimated to be less than or equal to 1.17 percent of the sample frame 99 times out of 100.

---

* only an unstratified estimate was available for the South Lahonton basin.

Relative error, expressed as a confidence interval half-width in numerical column three of Table 3-5a, was found to exceed plus or minus five percent of the estimate at the 95 percent level of confidence in four basins out of ten. In three of these cases - San Francisco, Central Coast, and South Coast - the estimated relative error did not exceed 6.28 percent for stratified regression. Statewide, the 99 percent confidence interval half-width for relative error was estimated to be 1.74 percent for stratified regression.

Column A5 of Table 3-6a shows that this statewide error represented 127 thousand acres at the 95 percent level of confidence. The corresponding value at the 99 percent level of confidence was approximately 166,800 acres. It is important to emphasize that these error figures were estimated for only the irrigated acreage within the sample frame. Error values for the approximately three percent of statewide irrigated land outside the sample frame were not available.

Stratified Versus Unstratified Regression Results

Unstratified regression estimates were prepared and are reported in Table 3-5b. These results were obtained in order to determine the effect of stratification on estimates of irrigated proportion and associated error.

Comparison of Tables 3-5a and 3-5b shows close agreement between irrigated proportion figures, both statewide and for individual basins. The state totals differed by only 0.05 percent of sample frame area from each other. Corresponding basin estimates differed by no more than one percent of sample frame area.

Overall, the stratified regression approach produced smaller basin confidence interval half-widths (both absolute and relative) in four out of the nine basins where both estimates were available. No stratified estimate was available for a tenth basin, the South Lahonton, because all sample unit observations had been classified to a single stratum. All values shown for this basin were based on unstratified regression estimators.

In the four basins where stratified regression was superior, significant differences in regression slopes or intercepts between two or more strata were evident (see Tables 3-11 and 3-12 in the evaluation section). Differences between strata regression coefficients were generally much less pronounced in the five basins where unstratified regression produced smaller confidence interval half-widths. This result is consistent with the commonly accepted conditions under which stratified regression is expected to be superior to unstratified regression.

Two additional factors contributed to the difference between errors reported for the stratified and unstratified cases. The first of these was a relative gain in precision for the unstratified estimate due to a larger number of degrees of freedom than given to the corresponding stratified estimate.*

---

* As explained in Appendix IB, degrees of freedom for the stratified case will (by equation 28) always fall between the smallest of the terms $n_h - 2$ and their sum. In contrast, degrees of freedom for the unstratified case equaled $(\sum_{h=1}^{L} n_h) - 2)$ a value larger than maximum value for stratified sampling. Tables 3-32 and 3-33 in the evaluation section present the computed degrees of freedom for each estimate of error.

For a given size of standard error this meant a smaller Student's t value and therefore a narrower confidence interval half-width for the unstratified estimate.

A second factor favoring narrower unstratified confidence intervals was a more stable sample unit weighting procedure. Recall that the weight assigned to each unstratified observation was determined relative to the average area of all sample units in a given basin, while that of stratified observations was computed with respect to only the average area within sample units occupying a given stratum. Weights based on the latter (stratified) procedure should tend to be somewhat more variable than those based on the former. More variable weights, in turn, contribute to higher sampling variance.

Even allowing for the considerations cited above, the error performance of stratified regression relative to unstratified was considered disappointing. In order to determine if some other factor might have limited precision gains due to stratification, an examination was made of the method for aggregating stratum (x,y) observations to produce unstratified values. The procedure had been to simply treat these observations as being obtained from sample units allocated in an unstratified, random manner. In effect, the observations were defined to represent a single, undifferentiated data set. Unstratified regression formulas outlined in Appendices IA and IB were then applied to these data to produce the values shown in Table 3-5b.

If the number of sample units selected from each land use stratum had been proportional to the relative size (area) of that stratum, the estimates of unstratified sampling error obtained by the procedure described above should not be biased.* However, allocation of ground sample units to strata was optimal and not proportional in the 1979 inventory. Thus the estimates of unstratified error reported in Table 3-5b could be either higher or lower than those obtained by an 'inherently' unstratified sample. Appendix II presents an evaluation of this problem, including a comparison of standard errors computed according to the formulas cited earlier versus standard errors predicted for an 'inherently' unstratified allocation.

Table 3-5c gives the estimated 'inherently' unstratified values for absolute error in percent. Comparison with Table 3-5b shows that the 'inherently' unstratified figures were higher than the original estimates in eight out of the nine basins where observations from separate strata had been aggregated together. Comparison to Table 3-5a shows the stratified absolute error to be less than the corresponding 'inherently' unstratified values in seven of the nine basins where a stratified estimate was available. These basins are identified by asterisks in Table 3-5a. Statewide, the stratified estimate of irrigated proportion gave a 22 percent reduction in sampling error when compared to the 'inherently' unstratified estimate.

---

* Since random selection of units within strata in this manner would tend to mimic the random selection of units from an unstratified sampling frame.

It appears from this analysis that, in addition to providing separate estimates by stratum, the stratified design may give reduction in sampling error in some basins and statewide. However, further operational experience with this design will be necessary to verify this finding.

Summary of Results

Viewing the error estimates as a whole, the original goal of $\pm$ five percent 95 times out of 100 was met in all ten basins on the basis of absolute error and in six out of ten based on relative error.* This conclusion will be valid <u>if</u> the assumptions made for regression estimation in the 1979 were, in fact, true. In addition, the relative size (percent) error for measurements of irrigated acreage outside the sample frame must be no larger than the within frame values. Otherwise, percent errors for the total irrigated land by basin will exceed those reported here.

The same qualifications apply to the statewide estimated errors. Table 3-5a reports these to be less than two percent at the 99 percent level of confidence for both absolute and relative error. Both were less than one percent at the 95 percent level of confidence.

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.

Continuing Evaluation of Results

An accuracy assessment of the Irrigated Lands APT estimates of irrigated land and associated error is presently underway. APT estimates for 1979 are being compared with corresponding estimates produced by the California Department of Water Resources. These DWR figures are based on 1) 'wall-to-wall' irrigated land and crop surveys performed by the Department for several counties in 1979, 2) California Agricultural Commissioner reports for 1978 extrapolated to 1979, and 3) Crop and Livestock Reporting Service 1978 sample estimates extrapolated to 1979. Comparison of acreage estimates is proceeding on a county-by-county basis, seeking to identify sources of error in either source. DWR county estimates will then be aggregated into basin and statewide estimates for comparison to the figures developed by the Landsat-aided inventory.

3.6.2 <u>County Estimates</u>

Table 3-7 displays the county irrigated land estimation results. Proportion of the sample frame irrigated was seen to range from zero in several counties, located primarily in coastal and mountain areas, to a high of 86 percent in Imperial County. Many Central Valley counties had within-frame irrigated proportions of 75 percent or more, and most were greater than 60 percent.

---

* Relative error was within 6.3 percent 95 times out of 100 in three of the remaining four basins in the case of stratified sampling.

Total county irrigated acreage, reported in the column on the right, varied widely. Counties having the most irrigated acreage included Fresno (1.25 million acres), Kern (.99 million acres), and Tulare (.78 million acres). Several others topped 500,000 acres including Imperial, Kings, Merced, and San Joaquin counties. The total statewide estimate of irrigated acreage based on addition of county totals was 9.85 million acres. This figure compares favorably to the statewide estimate of 9.86 million acres reported in Table 3-6b, the actual difference (7,238 acres) representing less than one tenth of one percent of the Table 3-6b value.

Standard errors ran approximately 5 to 15 percent of the estimated irrigated acreage within the sample frames of counties having sizable amounts of irrigated land. Standard errors ranging from 20 to 50 percent of the estimate were not uncommon in counties containing smaller acreages of irrigated land. As expected, these county errors were larger than their basin counterparts. This was due to the fact that 1) the original ground sample was allocated to control error at the basin level, not the county level, and 2) the county estimates of within-frame irrigated acreage represented predictions using the regression equations developed for the basins. When the regression estimator is used in a predictive fashion, an extra term must be added to the variance formula to account for the variation of observations about the regression line. The net result is a larger error interval than would be obtained if samples were allocated for precision control at the county level.

It is recommended that the reported standard errors be used as a guide to expected error only when the sample frame contains most of the irrigated land. In some counties, primarily those with smaller total irrigated area, the proportion of irrigated land found outside the sample frame was large. Since no error statement was available for these areas, the reported standard error may significantly under estimate the true value. Future surveys may be able to include a large portion of these areas within the sample frame, thereby eliminating this source of uncertainty.

Table 3-8 illustrates the points made in the previous paragraph. This table compares the county acreage estimates obtained in the 1979 inventory with those provided by the California DWR. The DWR figures were obtained from their own complete area surveys in 1979 for several counties and from extrapolation of 1978 California Agricultural Commissioner and Crop and Livestock Reporting Service reports for other counties. The fourth numerical column following every county shows the difference between the corresponding irrigated acreage estimates as a percent of the DWR estimate. Inspection of this column indicates that the percent difference between estimates was within or very near to the predicted standard error value for counties having significant amounts of irrigated land. However, larger errors than predicted sometimes occurred in counties with smaller acreages of irrigated land.

The comparison shown in Table 3-8 should not be considered final. An evaluation during this coming year will seek to identify the source of differences between county estimates. Since error may be found in either or both, final conclusions regarding the performance of the Landsat-aided county estimation procedure for irrigated acreage must be withheld at present.

TABLE 3-8. COUNTY ESTIMATES

A Comparison of DWR Estimates with APT
Estimates Based on Weighted-Unstratified Values

COUNTY	ESTIMATES BY COUNTY (IN THOUSANDS OF ACRES) DWR	APT	DIFFERENCE (IN THOUSANDS OF ACRES)	DIFFERENCE AS PERCENT OF DWR'S ESTIMATE
Alameda	14.4	12.3	-2.1	-14.6
Alpine	6.3	4.9	-1.4	-22.2
Amador	4.6	8.7	+4.1	+89.1
Butte	252.6	236.4	-16.2	- 6.4
Calaveras	2.7	1.1	- 1.6	-59.3
Colusa	307.5	312.0	+ 4.5	+ 1.5
Contra Costa	58.3	67.2	+ 8.9	+15.3
Del Norte	5.8	8.7	+ 2.9	+50.0
El Dorado	7.1	4.2	- 2.9	-40.8
Fresno	1310.8	1252.4	-58.4	- 4.5
Glenn	240.0	251.5	+11.5	+ 4.8
Humboldt	24.8	28.8	+ 4.0	+16.1
Imperial	527.4	515.0	-12.4	- 2.4
Inyo	16.5	13.3	- 3.2	-19.4
Kern	991.5	986.0	- 5.5	- .6
Kings	613.7	556.5	-57.2	- 9.3
Lake	16.3	14.6	- 1.7	-10.4
Lassen	80.2	71.7	- 8.5	-10.6
Los Angeles	41.2	32.9	- 8.3	-20.1
Madera	353.1	279.3	-73.8	-20.9
Marin	.6	.5	- .1	-16.7
Mariposa	.8	.3	- .5	-62.5
Mendocino	21.7	23.5	+ 1.8	+ 8.3
Merced	492.4	587.4	+95.0	+19.3
Modoc	172.0	161.9	-10.1	- 5.9
Mono	36.8	40.8	+ 4.0	+10.9
Monterey	184.5	235.8	+51.3	+27.8
Napa	18.0	15.5	- 2.5	-13.9
Nevada	11.1	5.0	- 6.1	-55.0
Orange	21.2	18.5	- 2.7	-12.7
Placer	42.3	66.5	+24.2	+57.2
Plumas	37.7	13.7	-24.0	-63.7
Riverside	250.0	249.6	- .4	- .2
Sacramento	196.7	224.6	+27.9	+14.2
San Benito	54.2	48.6	- 5.6	-10.3
San Bernardino	72.1	68.1	- 4.0	- 5.5
San Diego	85.0	73.0	-12.0	-14.1
San Francisco	0.0	0.0	-	-
San Joaquin	573.2	539.9	-33.3	-5.8
San Luis Obispo	58.4	66.8	+ 8.4	+14.4
San Mateo	4.8	3.0	- 1.8	-37.5
Santa Barbara	90.9	79.7	-11.2	-12.3
Santa Clara	42.0	27.9	-14.1	-33.6
Santa Cruz	24.0	24.4	+ .4	+ 1.7
Shasta	53.0	63.9	+10.9	+20.6
Sierra	16.4	19.8	+ 3.4	+20.7
Siskiyou	196.0	208.7	+12.7	+6.5
Solano	179.6	179.1	- .5	- .3
Sonoma	35.0	29.5	- 5.5	-15.7
Stanislaus	402.0	409.2	+ 7.2	+ 1.8
Sutter	298.6	285.8	-12.8	- 4.3
Tehama	97.2	114.0	+16.8	+17.3
Trinity	1.4	.8	- .6	-42.9
Tulare	710.9	776.1	+65.2	+ 9.2
Tuolumne	2.9	.6	- 2.3	-79.3
Ventura	111.9	101.7	-10.2	- 9.1
Yolo	327.0	337.5	+10.5	+ 3.2
Yuba	97.9	93.3	- 4.6	- 4.7
STATE	9894.2	9852.5	-41.7	- 0.4

3.7 EVALUATION OF THE 1979 IRRIGATED LANDS ESTIMATION PROCEDURE

3.7.1 Evaluation of Differences Among Strata

Analysis of Variance

One of the major test and evaluation objectives of the 1979 irrigated lands inventory was to determine whether the relationship between Landsat measurements (X) and ground measurements (Y) varied significantly between different land use strata. Stratified sampling will give lower basin-wide standard error than unstratified sampling when there are significant differences between strata regression slopes or intercepts. In addition, stratum-specific estimates of irrigated area will be biased if an unstratified estimator is used when strata regression coefficients differ significantly.

An analysis of variance was performed to obtain a measure of statistical difference between regression coefficients. Three hypotheses were tested. These were

$H_1 : b_h = b_k$ for all h,k (i.e. all strata regression slopes are equal within a given basin),

$H_2 : a_h = a_k$ for all h,k (i.e. all strata regression intercepts are equal within a given basin), and

$H_3 : b_h = b_k$ and $a_h = a_k$ for all h,k (i.e. all strata slopes are equal <u>and</u> all strata intercepts are equal in a given basin).

To test these hypotheses, four models relating ground to Landsat measurements were formulated. The first, Model 0, expressed Y as a linear (regression) function of Landsat X plus a term for error about the regression line, where both the intercept and slope were allowed to vary between strata. That is

$$y_{hi} = a_h + b_h x_{hi} + e_{hi} \quad , \qquad \text{(Model 0)}$$

where

y_{hi} = ground measurement (irrigated proportion) in sample unit i of land use stratum h,

x_{hi} = Landsat measurement of irrigated proportion in sample unit i of stratum h,

a_h = regression line intercept in stratum h,

b_h = regression line slope in stratum h, and

e_{hi} = term for error $(=y_{hi} - (a_h + b_h x_{hi}))$ about the regression line in sample unit i of stratum h.

A second model was defined for the case where all strata slopes were assumed to be constant in a given basin. Thus

$$y_{hi} = a_h + b\, x_{hi} + e_{hi} \quad . \qquad \text{(Model 1)}$$

In a similar fashion, a model appropriate to the case where all intercepts were constant was defined by

$$y_{hi} = a + b_h x_{hi} + e_{hi} \quad . \qquad \text{(Model 2)}$$

Finally, a model for the case where both the slope and intercept were assumed to be constant for all strata in a given basin was given by

$$y_{hi} = a + b\, x_{hi} + e_{hi} \quad . \qquad \text{(Model 3)}$$

An additional assumption in each model above was that the e_{hi} were independent of one another and were each distributed according to a normal distribution with a mean of zero and a constant variance of σ^2.

A test of hypotheses 1, 2, and 3 was then performed by comparing, through a one-way analysis of variance, Model 0 with alternative Models 1, 2, and 3 respectively. An F statistic, a measure of the difference in residual* variation about the regression line for one model versus another, was constructed** for each of the three model comparisons. Differences in residual variation between the models were due to the assumption that one or both of the regression coefficients were equal across land use strata in a given basin. The F value calculated in this manner was referred to tabulated percentage points of the F-distribution. The hypothesis in question was rejected if the table showed that the probability of obtaining the calculated F value under the assumption that the hypothesis was true was less than or equal to .05.

Table 3-9 summarizes the calculated F statistics for each basin associated with the tests of hypotheses 1, 2, and 3. The degrees of freedom (df) used to locate the proper tabulated F value are also given there. Two entries for df are given: the first v_1 = no. of strata minus one, and the second v_2 = no. of ground sample units in the basin minus the number of strata.

---

* A residual was defined to be the difference between an observed y_i and the corresponding regression line estimated value $\hat{y}_i$ for a given value of x_i.

**where F =

$$F = \frac{\dfrac{(\text{sum of squares for error for Model 1, 2, or 3}) - (\text{sum of squares for error Model 0})}{\text{Difference in degrees of freedom between models}}}{\text{mean square for error for Model 0}}$$

Inspection of the right-most column in Table 3-9 shows which hypotheses were rejected at the five percent level of significance in the various basins. Thus, differences among slopes were found to be statistically significant at the five percent level in only the North Coast and San Francisco hydrologic basins. Differences among intercepts were seen to be significant in the San Francisco and Central Coast basins. When hypotheses of both constant slopes and constant intercepts were considered, rejection occurred in the three basins mentioned above plus the Tulare unit.

In order to identify which regression coefficients were causing rejection of the hypotheses, contrasts (differences) were formed between pairs of strata slope or pairs of intercept values and these were then evaluated for statistical significance. The Scheffé method for simultaneous confidence intervals was used to establish this significance. This method tests whether the difference between the two slopes (or intercepts) under question is significantly different than zero; if the confidence interval about the difference did not include zero then the hypothesis of no difference between the two strata coefficients was rejected at the given level of statistical significance. The rejection level chosen for these pairwise tests was again five percent. Table 3-10 summarizes the results for tests of difference between regression slopes in the North Coast and San Francisco basins.

The results shown in Table 3-10 can be related to the values of the slope coefficients for each basin (shown in Table 3-11). For example, the small slope reported for stratum 5 in the North Coast was found to be significantly different from both the slope in stratum 2 and the slope in stratum 3, but the latter were not different from each other or that of stratum 1. Similarly, the low slope in stratum 4 of the San Francisco unit relative to that of stratum 3 caused rejection of the hypothesis of equal slopes.

Construction of Scheffé confidence intervals for regression intercepts showed that only those contrasts involving stratum 4 in the San Francisco hydrologic unit and stratum 3 in the Central Coast unit were declared significant at the .05 level. Table 3-12 lists the regression intercepts by basin. Intercepts in the two strata just mentioned are clearly much larger than their counterparts in other strata. Inspection of Tables 3-11 and 3-12 indicates that both the slope and intercept in stratum 1 of the Tulare basins were probably responsible for the rejection of H_3 in that hydrologic unit.

The statistical tests of difference between regression coefficients described above do not, however, provide a complete picture of between stratum differences. They identify only the most significant contrasts. The relative importance of differences in slope and intercept must be judged in the larger context of their overall impact on the basin-wide estimate of irrigated area and the associated estimate of error. Important factors to consider in this regard include 1) the relative area in each stratum, 2) the relative dispersion of observations around the regression line in each stratum (summarized in the form of an (X,Y) correlation coefficient, and their pattern of dispersion*, 3) the proportion of land irrigated in each stratum, and 4) the sample size in each stratum. For example, if a stratum had little area and was only moderately irrigated, it might have little impact on estimated error even if its regression coefficients differed

---

* including 1) the shape of the X,Y distribution as a whole, 2) the resulting pattern of residuals about the regression line, and 3) the distribution of residuals with respect to the size (area) of sample unit

Table 3-9: Resulting F Statistics from ANOVA for Hypothesis H_1, H_2, and H_3

H_1: Model 0 vs Model 1:

North Coast	-	F = 6.485*	d.f. = 3,41
San Francisco	-	F = 3.238*	d.f. = 4,46
Central Coast	-	F = .672	d.f. = 5,67
South Coast	-	F = .357	d.f. = 5,69
Colorado D.	-	F = .168	d.f. = 3,50
South Lahonton	-	-----	= ----
North Lahonton	-	F = .699	d.f. = 1,33
Sacramento	-	F = 1.188	d.f. = 5,60
San Joaquin	-	F = 1.223	d.f. = 5,64
Tulare	-	F = .821	d.f. = 3,57

H_2: Model 0 vs Model 2:

North Coast	-	F = .955	d.f. = 3,41
San Francisco	-	F = 3.464*	d.f. = 4,46
Central Coast	-	F = 2.833*	d.f. = 5,67
South Coast	-	F = .187	d.f. = 5,69
Colorado D.	-	F = .653	d.f. = 3,50
South Lahonton	-	----	----
North Lahonton	-	F = 1.816	d.f. = 1,33
Sacramento	-	F = .767	d.f. = 5,60
San Joaquin	-	F = .576	d.f. = 5,64
Tulare	-	F = .219	d.f. = 3,57

H_3: Model 0 vs Model 3:

North Coast	-	F = 4.148	d.f. = 6,41
San Francisco	-	F = 2.180	d.f. = 8,46
Central Coast	-	F = 3.150	d.f. = 10,67
South Coast	-	F = .678	d.f. = 10,69
Colorado D.	-	F = 1.775	d.f. = 6,50
South Lahonton	-	-----	-----
North Lahonton	-	F = 1.139	d.f. = 2,33
Sacramento	-	F = .927	d.f. = 10,60
San Joaquin	-	F = .856	d.f. = 10,64
Tulare	-	F = 2.585*	d.f. = 6,57

*Hypothesis of no significant difference rejected at $\alpha = .05$ significance level; i.e. if hypothesis in question was true, the observed values for the regression coefficients would only occur less than or equal to 5 times in 100 trials - an event that would not support the hypothesis of no significant difference.

TABLE 3-10. Results for Scheffé Tests of Difference Between Regression Slopes**

North Coast: $S^2 = d \cdot F_{\alpha;d,n-r}$
$= 3F_{.05;3,41}$ so $S \doteq 2.919$

Stratum Pair i,j	Slope Difference $\beta_i-\beta_j$	$var(\beta_i-\beta_j)$	95% Confidence Interval Bounds Lower	Upper
1,2	.02665	14.42131	-11.058	11.111
1,4	.14126	14.40935	-10.939	11.221
1,5	.49147	14.41359	-10.590	11.573
2,4	.11461	.01490	- .242	.471
* 2,5	.46482	.01914	.061	.869
* 4,5	.35021	.00718	.103	.598

San Francisco: $S^2 = 4F_{.05;4,46}$ so $S \doteq 3.225$

1,2	.20339	14.57468	-12.108	12.515
1,3	.06439	14.56028	-12.241	12.370
1,4	.81463	14.58922	-11.503	13.132
1,5	.23452	14.55941	-12.071	12.540
2,3	-.13900	.02940	- .692	.414
2,4	.61124	.05834	- .168	1.390
2,5	.03113	.02853	- .514	.576
* 3,4	.75024	.04394	.074	1.426
3,5	.17013	.01413	- .213	.533
4,5	-.58011	.04307	- 1.249	.039

* 95 percent confidence does not cover origin, so hypothesis of no significant difference between slopes is rejected at the $\alpha = .05$ significance level.

**Scheffé tests for contrasts among β's for cases (North Coast & San Francisco) where F-test rejected ($\alpha = .05$) H : $\beta_i = \beta_j$ for all i,j (Model 0 vs. Model 1).

Table 3-11. Stratum specific regression coefficients (slopes) by basin.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	1.15704	1.12994	------	1.01535	.66512	------	------	------
San Francisco	1.07466	.87125	1.01024	.25999	.84011	------	------	------
Central Coast	1.18087	1.00293	.88619	1.00307	------	.96214	1.08628	------
South Coast	1.66812	1.16774	1.05261	1.08444	1.08301	------	1.24437	------
Colorado Desert	------	------	------	1.02018	1.06329	1.04550	1.05981	------
South Lahontan	------	------	------	------	------	------	------	1.01276
North Lahontan	------	.86718	------	.96226	------	------	------	------
Sacramento	.71822	.77358	------	.92217	.91926	.98465	1.12657	------
San Joaquin	1.19883	.87839	.92879	.87361	.87616	.98050	------	------
Tulare	.82957	------	------	.99478	.85035	.93739	------	------

Table 3-12. Stratum specific regression coefficients (intercept) by basin.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	-.00252	-.00314	------	-.00078	.05955	------	------	------
San Francisco	-.00003	.10773	-.01338	.22019	.07198	------	------	------
Central Coast	-.00211	.01121	.13282	-.03373	------	.02657	.00461	------
South Coast	-.00939	.03526	-.03146	.05243	.02061	------	-.02321	------
Colorado Desert	------	------	------	-.01468	-.00329	.03503	.01557	------
South Lahontan	------	------	------	------	------	------	------	-.00418
North Lahontan	------	.11668	------	.01837	------	------	------	------
Sacramento	.03807	.14329	------	.02478	.03804	-.00073	.00694	------
San Joaquin	-.00802	.05844	.04166	.07735	.15221	.00863	------	------
Tulare	-.06613	-------	-----	.00068	.03058	.03827	------	------

significantly from those of other strata in the same basin. A measure of this impact would be the estimated stratum standard error computed according to Equation 21 or 22 of Appendix IB. Yet even inspection of standard errors must be tempored by consideration of the sample size used to compute both that error and the underlying regression coefficients in the first place. Larger stratum sample size might stabilize regression coefficients at different values as well as reduce estimated sampling error.

Clearly, the problem of determining the real significance of differences in X,Y relationships between strata is not simple. Decisions regarding the 'effective' importance of differences and resulting recommendations regarding which strata to keep must be made relative to the objectives of the inventory. If only basin-wide estimates are of importance, if simplicity in conduct of the inventory is desired, and if the presence of strata do not significantly decrease basin-wide standard error, then an unstratified design is to be preferred. On the other hand, even if difference in basin error is negligable, a requirement for within-stratum estimation would necessitate a stratified design.

Supplementary Analysis of Differences Between Stratum Statistics

A number of statistics were calculated in order to more completely evaluate the difference between the seven major land use strata. These statistics, summarized in the following tables, were developed for the case of stratified regression estimation with weighted sample unit observations described earlier. In addition to the stratum-specific regression coefficients given in Tables 3-11 and 3-12, the list includes:

Table 3-13 - estimate of proportion irrigated in each stratum by equation 8,

Table 3-14 - estimate of standard error in each stratum as a percent of percent of the sample frame by equations 17 and 18,

Table 3-15 - estimate of the square of (X,Y) correlation by stratum,

Table 3-16 - stratum weights, i.e. the relative proportion of total basin area in each stratum,

Table 3-17 - ground sample size in each stratum,

Table 3-18 - stratum sample unit population size, i.e., the total number of sample units in each stratum in a given basin,

Table 3-19 - estimated irrigated acreage in each stratum from the expression inside the summation sign in equation 9,

Table 3-20 - estimated standard error in acres in each stratum from the expression inside the summation sign in equation 19 times the square of the area in the given stratum,

Table 3-21 - estimated coefficient of variation (CV) for each stratum, where CV = 100 times Table 3-14 value divided by Table 3-13 value, and

Table 3-22 - estimated 95 percent confidence interval half-width by stratum (equal to Table 3-21 value times Student's-t with df = n_h - 2).

Examination of regression coefficient values in light of this information, especially stratum weights, irrigated proportions and standard errors, correlation, and sample size, lead to several observations:

1) Stratum 1 slope tended to differ from those of other strata in several basins. This appeared to be due to low proportion irrigated.small sample size, and resulting difficulty in establishing a stable regression line;

2) At least one stratum was found to be significantly different from others in the three northern-most coastal basins:

   a) Stratum 5 in the North Coast unit gave a low slope indicating a tendency for Landsat over-estimation of irrigated land. Examination of the plot in Figure 3-10 shows that this did not occur in all sample units. A preliminary analysis (discussed in the next section) indicated that error here was due to the difficulty of determining from Landsat imagery whether orchards or vineyards were irrigated. As this stratum represented 20 percent of the sample area for the North Coast basin, its maintenance as a distinct stratum seems to be advisable;

   b) Stratum 4 in the San Francisco unit gave both slope and intercept significantly different from other strata. Landsat-to-ground correlation was extremely low (.52), reflected in the fact that the regression line itself was nearly flat. However, the high intercept and low slope were largely due to one outlier observation*. Otherwise, the Landsat measurements in stratum 4 tended to over-estimate, rather than under-estimate ground proportion irrigated. (See Figure 3-11) Since this stratum occupies only seven percent of the sample frame, it would be desirable from a sampling standpoint to combine stratum 4 with other strata. Successful combination would, however, depend upon a careful review of the reasons for error in stratum 4 and an analysis of the reoccurring impact of this stratum on the error associated with a combined stratum estimate. Ideally, this year's evaluation will identify the sources of Landsat measurement error in stratum 4 leading to a minimization of this problem in future inventories. A decision to combine strata must also be weighed against the possibility of increasing bias in stratum-specific estimates when these are

---

* A single observation can have significant impact on the location of a regression line at low sample size. In this case, one observation having a high weighted Landsat proportion irrigated pulled the regression line down and the intercept up. The sample size was seven.

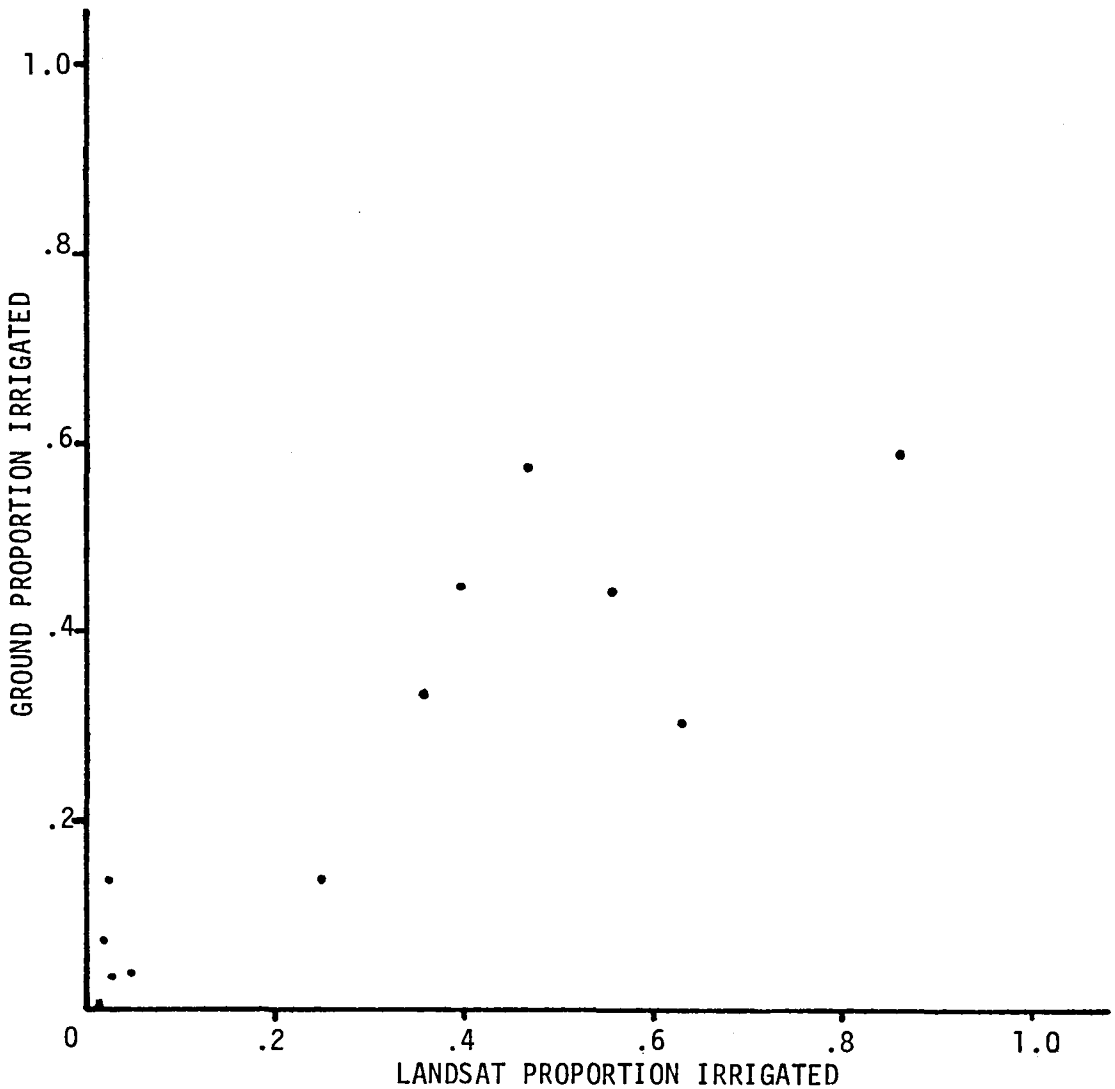


Figure 3-10. North Coast - Stratum 5; Weighted Observations

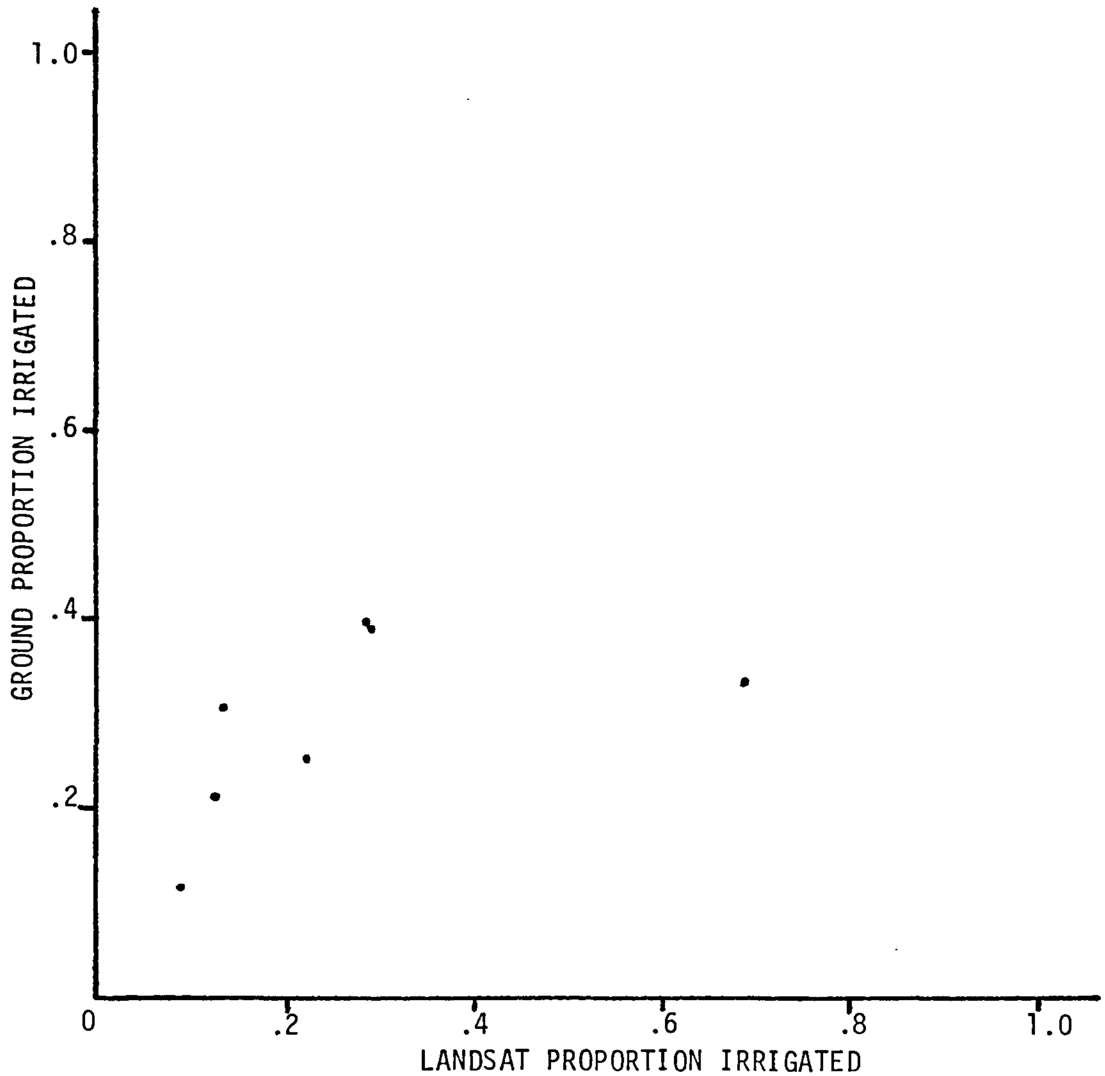


Figure 3-11. San Francisco - Stratum 4; Weighted Observations

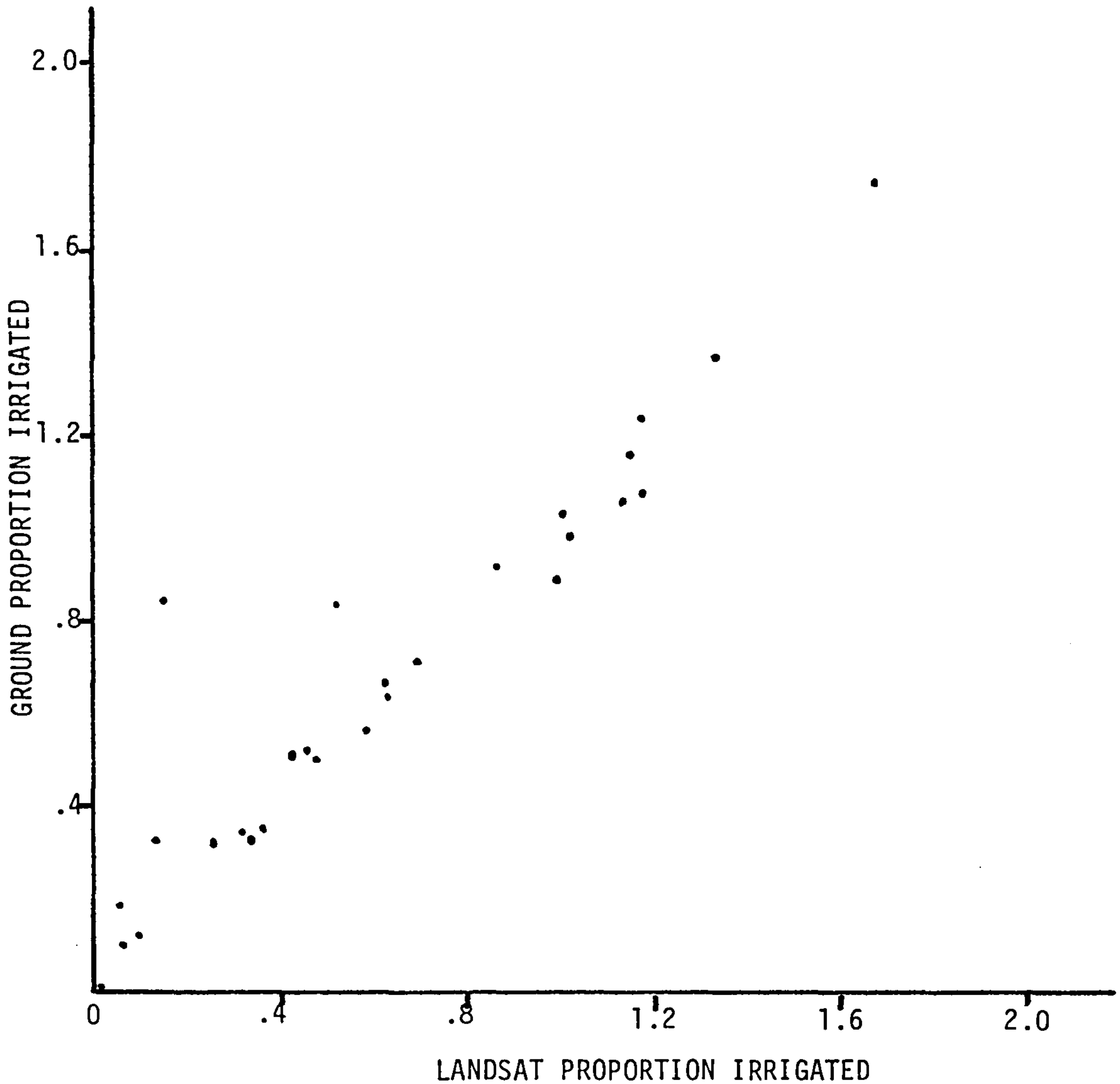


Figure 3-12. Central Coast - Stratum 3; Weighted Observations

desirable (e.g. for county prediction);

c) Stratum 3 in the Central Coast unit had an intercept significantly higher than others in that basin. This appears to have been due to the difficulty in detecting young vineyards on Landsat imagery.

This error manifested itself in several low to medium proportion irrigated sample units (see Figure 3-12), thereby pulling the intercept up and the slope down. Thus, contrary to the impression given by the slope, Landsat over-estimation generally did not occur in this stratum. With this understanding in mind, combination of this stratum with other field crop strata would appear feasible.

3) Apparent, though not statistically significant, differences did occur in other strata. For example, in the South Coast the slope for stratum 2 differed from the slopes for other field crop strata. This was also true for the field crop strata in the Sacramento basin. A relatively high intercept was also obtained for stratum 2 in the Sacramento unit. The proportion of sample unit area in stratum 2 was not large in either basin, being 13 percent in the South Coast and 8.3 percent in the Sacramento basin. Combination with other strata appears to be feasible.

4) Though statistically insignificant, stratum 7 tended to have a higher slope than either the field crop or orchard strata. Inspection of Y versus X plots showed this was often due to one or two outlier observations, though there did seem to be a slight tendency to under-estimate irrigated acreage. This tendency was expected given the diversity of land use patterns occurring in stratum 7 and the attendant difficulty in detecting some irrigated fields.

Combination of Strata

Looking at the strata results as a whole, it appears that combination of strata is possible. The field crop strata (2,3,4), originally distinguished on the basis of field size and irrigated proportion, did not appear to be distinct enough statistically to justify separate sample allocation or estimation. Possible exceptions to this statement were found in the San Francisco and Central Coast basins. However, even in these cases, combination of field crop strata appears feasible due to the relatively small area occupied and the potential for reduction of Landsat interpretation error in future surveys. Combination of the orchard and vineyard strata (5 and 6) also appears to be justified. No significant differences occurred between these strata when they appeared together.

In order to examine the characteristics of a sampling system based on combined strata, sample observations from the 1979 inventory were combined according to the strategy suggested above. The resulting design consisted of four strata. Strata 1 and 4 in this alternative system represented the old strata 1 and 7 respectively. Irrigated proportion and land use pecularities particular to these strata, and the resulting potential for regression line dissimilarity indicated their retention as separate strata. The new stratum 2 included the old field crop strata 2, 3, and 4.

Table 3-13. Stratum specific estimated irrigated proportions for the regression with factor 5 model.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	.00798	.21113	------	.67396	.27509	------	------	------
San Francisco	.00243	.51289	.34070	.28836	.42541	------	------	------
Central Coast	.02720	.34778	.71218	.46554	------	.40857	.27636	------
South Coast	.14877	.56574	.54253	.43472	.58444	------	.42603	------
Colorado Desert	------	------	------	.83780	.45308	.77228	.16097	------
South Lahontan	------	------	------	------	------	------	------	.27383
North Lahontan	------	.49251	------	.60926	------	------	------	------
Sacramento	.11172	.49514	------	.80795	.84907	.80352	.18216	------
San Joaquin	.24434	.69949	.76775	.76142	.91143	.83626	------	------
Tulare	.22866	------	------	.83807	.81983	.79309	------	------

Table 3-14. Stratum specific standard errors by basin (for regression with factor 5).

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	.00386	.00935	------	.01217	.02690	------	------	------
San Francisco	.00082	.03606	.01290	.01260	.01728	------	------	------
Central Coast	.00845	.01736	.02504	.01402	------	.00685	.01638	------
South Coast	.02718	.02940	.01228	.06300	.02470	------	.05035	------
Colorado Desert	------	------	------	.00770	.01835	.01326	.00570	------
South Lahontan	------	------	------	------	------	------	------	.01868
North Lahontan	------	.02590	------	.01496	------	------	------	------
Sacramento	.02094	.04055	------	.01154	.01053	.02308	.02005	------
San Joaquin	.07732	.03301	.07576	.01568	.02687	.02133	------	------
Tulare	.12756	------	------	.01143	.01779	.02152	------	------

Table 3-15. Stratum specific correlations (r-squared) by basin.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	.63539	.99353	------	.97328	.78615	------	------	------
San Francisco	.65395	.93971	.94151	.27521	.77777	------	------	------
Central Coast	.80274	.93563	.89170	.99360	------	.99799	.95985	------
South Coast	.70659	.85953	.99370	.46772	.90461	------	.66180	------
Colorado Desert	------	------	------	.97265	.98967	.99480	.97127	------
South Lahontan	------	------	------	------	------	------	------	.78344
North Lahontan	------	.84469	------	.94865	------	------	------	------
Sacramento	.86876	.80802	------	.97178	.99903	.99757	.76432	------
San Joaquin	.85171	.97235	.92251	.94077	.98507	.96010	------	------
Tulare	.34735	------	------	.95386	.99403	.95887	------	------

Table 3-16. Stratum weights by basin.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	.03454	.07292	------	.68697	.20557	------	------	------
San Francisco	.44406	.03675	.14115	.07532	.30272	------	------	------
Central Coast	.48399	.03898	.32005	.08047	------	.04389	.03262	------
South Coast	.19413	.13041	.14428	.14415	.31175	------	.07528	------
Colorado Desert	------	------	------	.88490	.01453	.09378	.00679	------
South Lahontan	------	------	------	------	------	------	------	1.00000
North Lahontan	------	.18842	------	.81158	------	------	------	------
Sacramento	.13820	.08331	------	.66369	.03693	.02469	.05317	------
San Joaquin	.05663	.01798	.04362	.70035	.03871	.14270	------	------
Tulare	.01868	------	------	.81368	.04626	.12138	------	------

Table 3-17. Stratum specific sample sizes by basin.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	4	6	--	27	12	--	--	49
San Francisco	19	4	11	7	15	--	--	56
Central Coast	26	7	28	7	--	5	--	79
South Coast	10	18	9	11	25	--	6	81
Colorado Desert	--	--	--	42	4	8	4	58
South Lahontan	--	--	--	--	--	--	--	33
North Lahontan	--	12	--	25	--	--	--	37
Sacramento	8	10	--	39	5	4	6	72
San Joaquin	6	4	4	43	5	14	--	76
Tulare	4	--	--	46	5	10	--	65

Table 3-18. Stratum specific population sizes by basin.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	7	17	--	176	55	--	--	255
San Francisco	44	7	14	8	19	--	--	92
Central Coast	287	29	181	47	--	19	20	583
South Coast	63	51	26	42	89	--	25	296
Colorado Desert	--	--	--	263	7	39	5	314
South Lahontan	--	--	--	--	--	--	--	115
North Lahontan	--	20	--	64	--	--	--	84
Sacramento	212	150	--	951	60	47	73	1493
San Joaquin	93	28	51	769	42	201	--	1184
Tulare	45	--	--	1294	58	191	--	1588

Table 3-19. Stratum specific estimated irrigated acres for the regression model.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	165	9235	------	277745	33925	------	------	321070
San Francisco	207	3612	9217	4162	24681	------	------	41880
Central Coast	18167	18709	314555	51700	------	24748	12442	440321
South Coast	17295	44183	46878	37529	109115	------	19205	274205
Colorado Desert	------	------	------	606611	5387	59258	895	672150
South Lahontan	------	------	------	------	------	------	------	64522
North Lahontan	------	16282	------	86757	------	------	------	103039
Sacramento	52317	139783	------	1817003	106254	67215	32821	2215392
San Joaquin	38591	35081	93407	1487225	98387	332819	------	2085511
Tulare	17431	------	------	2782438	154753	392784	------	3347406
State	144173	266886	464056	7151169	532501	876824	65363	9565494

Table 3-20. Standard errors in acres by stratum and basin for th regression withfactor 5 model.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7	All Strata
North Coast	80	409	------	5015	3317	------	------	6028
San Francisco	70	254	349	182	1003	------	------	1109
Central Coast	5644	934	11060	1557	------	415	737	12577
South Coast	3160	2296	1061	5439	4611	------	2270	8508
Colorado Desert	------	------	------	5575	218	1017	32	5672
South Lahontan	------	------	------	------	------	------	------	4402
North Lahontan	------	856	------	2130	------	------	------	2296
Sacramento	9806	11448	------	25952	1318	1931	3612	30319
San Joaquin	12212	1656	9217	30627	2901	8489	------	35430
Tulare	9724	------	------	37948	3358	10658	------	40737
State	19537	11870	14440	56075	7399	13805	4330	64490

Table 3-21. Stratum specific coefficients of variation for the regression with factor 5 model.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7
North Coast	48.37	4.43	----	1.81	9.78	----	----
San Francisco	33.74	7.03	3.79	4.37	4.06	----	----
Central Coast	31.07	4.99	3.52	3.01	----	1.68	----
South Coast	18.27	5.20	2.26	14.49	4.23	----	11.82
Colorado Desert	----	----	----	0.92	4.05	1.72	3.54
South Lahontan	----	----	----	----	----	----	----
North Lahontan	----	5.26	----	2.46	----	----	----
Sacramento	18.74	8.19	----	1.43	1.24	2.87	11.01
San Joaquin	31.64	4.72	9.87	2.06	2.95	2.55	----
Tulare	55.79	----	----	1.36	2.17	2.71	----

Table 3-22. Stratum specific estimated 95 percent confidence interval half-width in acres

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6	Stratum 7
North Coast	208.13	12.30	----	3.72	21.79	----	----
San Francisco	71.19	30.25	8.57	10.96	8.78	----	----
Central Coast	64.12	12.52	7.23	7.55	----	5.34	----
South Coast	42.13	11.02	5.35	32.29	8.74	----	28.92
Colorado Desert	----	----	----	1.86	17.43	4.20	15.24
South Lahontan	----	----	----	----	----	----	----
North Lahontan	----	11.72	----	5.08	----	----	----
Sacramento	45.86	18.89	----	2.89	3.95	12.36	30.56
San Joaquin	87.86	20.31	42.46	4.16	9.38	5.56	----
Tulare	240.03	----	----	2.75	6.91	6.26	----

Orchard and vineyard strata were combined to form the new stratum 3. Statistics corresponding to those presented in Tables 3-11 through 3-18 were generated for the new stratified sampling scheme and are shown in Tables 3-23 through 3-30. Stratum estimates of irrigated proportion and variance were obtained using the formulas for stratified regression presented earlier. Observation weights were again based on the size of the sample unit relative to the average size of sample units over all units in the given stratum.

Review of these tables for the four strata case shows that X,Y correlation in the new strata 2 and 3 was generally higher than average correlation among strata that were combined. However, combined strata correlation was in some cases somewhat lower than that obtained in original strata. Combined strata standard errors tended to be no larger, and in many cases smaller, than those obtained for the original strata. Regression slopes in new strata 2 and 3 tended to stabilize in the range of .9 to 1.1 and intercepts in the range -.02 to .08. In only two cases were stratum sample sizes, after combination, less than 10 units. Problems of small population size in some basins were reduced as well. Low population size remained a problem in stratum 1 of the North Coast and in the new stratum 4 of the Central Coast, South Coast, and Colorado Desert basins.

From the standpoint of minimizing hydrologic basin estimate error, the performance of the four strata design relative to that of the seven strata and unstratified designs is shown in Table 3-31. There it can be seen that differences in standard error between the stratified and unstratified designs tended to be small, with the exception of the three northern-most coastal basins. The same was true with respect to the 95 percent confidence interval half-widths. Of the two stratified designs, the four strata scheme tended to give somewhat lower estimated error.

Conclusions Regarding Stratification

Thus, on the basis of estimated error, no one strategy for stratification appears to be consistently superior. Reductions in estimated variance due to stratification generally occurred when significant differences in the size of regression coefficients occurred between strata. This was especially evident in the coastal basins. Elsewhere, gains in basin-wide precision were not found to be significant.* It would seem attractive to argue that stratification should be applied where the difference in estimated basin-wide error justified it. Otherwise, stratification should not be required, thereby simplifying sample allocation, Landsat interpretation and irrigated land estimation.

---

* A more recent study, however, suggests that gains due to stratification may in fact be possible in the Colorado Desert, Sacramento, and San Joaquin basins. These results, reported in Appendix II, were obtained when unstratified sampling error was adjusted for type of sample allocation used in 1979.

Table 3-23. Stratum specific regression coefficients (slopes) by basin for the combined strata.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	1.15704	1.02000	.66512	------	------
San Francisco	1.07466	.88723	.84011	------	------
Central Coast	1.18087	.92868	.96214	1.08628	------
South Coast	1.66812	1.00234	1.08301	1.24437	------
Colorado Desert	------	1.02018	1.05618	1.05981	------
South Lahontan	------	------	------	------	1.01276
North Lahontan	------	.94382	------	------	------
Sacramento	.71822	.91061	.94781	1.12657	------
San Joaquin	1.19883	.87829	.95794	------	------
Tulare	.82957	.99478	.87537	------	------

Table 3-24. Stratum specific regression coefficients (intercepts) by basin for the combined strata.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	-.00252	.00082	.05955	------	------
San Francisco	-.00003	.04708	.07198	------	------
Central Coast	-.00211	.07961	.02657	.00461	------
South Coast	-.00939	.05421	.02061	-.02321	------
Colorado Desert	------	-.01468	.02007	.01557	------
South Lahontan	------	------	------	------	-.00418
North Lahontan	------	.03981	------	------	------
Sacramento	.03807	.03725	.03034	.00694	------
San Joaquin	-.00802	.07240	.04432	------	------
Tulare	-.06613	.00068	.05704	------	------

Table 3-25. Stratum specific estimated irrigated proportions for the regression with factor 5 model where strata 2,3,and 4 and strata 5 and 6 are combined.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	.00798	.63217	.27509	------	------
San Francisco	.00243	.34997	.42541	------	------
Central Coast	.02720	.63402	.40857	.27636	------
South Coast	.14877	.50589	.58444	.42603	------
Colorado Desert	------	.83780	.72575	.16097	------
South Lahontan	------	------	------	------	.27383
North Lahontan	------	.58725	------	------	------
Sacramento	.11172	.77055	.84170	.18216	------
San Joaquin	.24434	.75897	.85749	------	------
Tulare	.22866	.83807	.79161	------	------

Table 3-26. Stratum specific standard errors by basin for the combined strata. Regression with factor 5 estimator.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	.00386	.01045	.02690	------	------
San Francisco	.00082	.01295	.01728	------	------
Central Coast	.00845	.01848	.00685	.01638	------
South Coast	.02718	.01570	.02470	.05035	------
Colorado Desert	------	.00770	.00162	.00570	------
South Lahontan	------	------	------	------	.01868
North Lahontan	------	.01207	------	------	------
Sacramento	.02094	.01102	.01074	.02005	------
San Joaquin	.07732	.01409	.01871	------	------
Tulare	.12756	.01143	.01699	------	------

Table 3-27. Stratum specific correlations (r squared) for the combined strata.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	.63539	.97664	.78615	------	------
San Francisco	.65395	.84616	.77777	------	------
Central Coast	.80274	.91195	.99799	.95985	------
South Coast	.70659	.95383	.90461	.66180	------
Colorado Desert	------	.97265	.99276	.97127	------
South Lahontan	------	------	------	------	.78344
North Lahontan	------	.93665	------	------	------
Sacramento	.86876	.96463	.99753	.76432	------
San Joaquin	.85171	.93992	.96675	------	------
Tulare	.34735	.95386	.96692	------	------

Table 3-28. Stratum specific weights for the combined strata.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	.03454	.75989	.20557	------	------
San Francisco	.44406	.25322	.30272	------	------
Central Coast	.48399	.43950	.04389	.03262	------
South Coast	.19413	.41884	.31175	.07528	------
Colorado Desert	------	.88490	.10831	.00679	------
South Lahontan	------	------	------	------	1.00000
North Lahontan	------	1.00000	------	------	------
Sacramento	.13820	.74700	.06162	.05317	------
San Joaquin	.05663	.76195	.18141	------	------
Tulare	.01868	.81368	.16764	------	------

Table 3-29. Stratum specific sample sizes for the combined strata.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	4	33	12	--	49
San Francisco	19	22	15	--	56
Central Coast	26	42	5	20	79
South Coast	10	38	25	25	81
Colorado Desert	--	42	12	5	58
South Lahontan	--	--	--	--	33
North Lahontan	--	37	--	--	37
Sacramento	8	49	9	6	72
San Joaquin	6	51	19	--	76
Tulare	4	46	15	--	65

Table 3-30. Stratum specific population sizes for the combined strata.

Basin	Stratum 1	Stratum 2	Stratum 3	Stratum 4	All Strata
North Coast	7	193	55	--	255
San Francisco	44	29	19	--	92
Central Coast	287	257	19	20	583
South Coast	63	119	89	25	296
Colorado Desert	--	263	46	5	314
South Lahontan	--	--	--	--	115
North Lahontan	--	84	--	--	84
Sacramento	212	1101	107	73	1493
San Joaquin	93	848	243	--	1184
Tulare	45	1294	249	--	1588

Table 3-31. Comparison of Stratified and Unstratified Regression Estimates of Standard Error and Confidence Interval Half-Width

	- - - - - Standard Error - - - - -		
Basin	7 Strata	4 Strata*	Unstratified*
North Coast	.00982	.00968	.01187
San Francisco	.00578	.00618	.01906
Central Coast	.00911	.00911	.01080
South Coast	.01421	.01203	.01207
Colorado Desert	.00693	.00693	.00647
South Lahonton	.01868	.01868	.01868
North Lahonton	.01309	.01207	.01207
Sacramento	.00895	.00881	.00817
San Joaquin	.01270	.01208	.01161
Tulare	.00998	.01001	.01047

- - - - -95 Percent Confidence Interval Half-Width- - - - -			
Basin	7 Strata	4 Strata*	Unstratified*
North Coast	.02040	.01958	.02387
San Francisco	.01215	.01297	.02198
Central Coast	.01842	.01825	.02150
South Coast	.02876	.02405	.02402
Colorado Desert	.01399	.01398	.01296
South Lahonton	.03810	.03810	.03810
North Lahonton	.02676	.02451	.02451
Sacramento	.01795	.01766	.01630
San Joaquin	.02551	.02420	.02312
Tulare	.02004	.02009	.02093

* Variance not corrected for original allocation procedure (See Section 3.6.1 and Appendix II; see also Table 3-5c).

This approach to the stratification decision must be weighed against the desirability of producing accurate sub-basin estimates, on either a county or land use stratum basis. In this context, some form of stratification might be appropriate even in those basins not stratified for reduction of basin-wide error.

In order to maintain the greatest flexibility in future surveys, some form of stratification will be desirable. Given the results to date, the four strata design would be the most likely candidate. Actual implementation of this design should also address two other stratification problems identified in this study. The first of these was the occurrence of irrigated land in dispersed agriculture outside the sample frame, occasionally occupying an important segment of irrigated acreage at the county level. In a future inventory, an attempt should be made to incorporate some or all this land in one of the four strata - with the fourth stratum (the original stratum 7) appearing to be the best possibility.

The second problem was the occurrence of irrigated land in areas excluded from the interior of the sample frame. These areas generally constituted an urban/agriculture mix or marsh lands not ordinarily used for agriculture. Grouping exclusion areas with the new stratum 2 or creating a separate sampling stratum should be considered in a future survey design.

Summary of Findings and Conclusions

1) Differences did exist between strata in some basins - some statistically significant and some not;

2) Where strata having significant area differed in slope or intercept, stratified regression gave smaller estimated error than unstratified regression. This occurred in four of the nine basins where a comparison was possible;

3) Analysis of differences between strata suggested a four strata approach. The four strata regression estimate of the 95 percent confidence interval half-width was smaller than its seven stratum counterpart in seven of the nine basins where a comparison was possible;

4) Several instances of low Landsat-to-ground correlation occurred. These are evaluated in the next section;

5) Exclusive of interpretation error per se, the dryland stratum suffered from lack of adequate imagery, low proportion irrigated, and in a number of cases, low sample size. As a result, the proper sample allocation and estimation procedure in this stratum is open to question.

6) Inclusion of a greater proportion of exclusion areas and agricultural land outside the contiguous sample frame should be considered in future surveys; and

7) The four strata design would serve as a starting point for a revised stratification scheme in future inventories.

3.7.2 Comparison of Alternative Estimators

Description of the Estimators

Bias in estimates of within-sample frame irrigated area and the size of the associated estimates of the sampling error can be affected by the type of equation used to link Landsat and ground observations. Different equations, or estimators as they are called, perform best under certain assumptions about the joint distribution of Landsat (X) and ground (Y) observations. Two preliminary, pre-1979 inventory, comparisons of alternative irrigated land estimators were described in Wall et al (1980). These comparisons indicated that the regression estimator should give results with lowest error at moderate to large sample sizes - but conclusions were tentative, depending on statistics derived from the Sacramento basin in 1976. Availability of data obtained over the ten basins in 1979 permitted a comprehensive analysis of relative estimator performance.

The objective of the estimator comparison performed following the 1979 inventory was to (1) determine if several well known linear estimators produced significantly different within-frame estimates of irrigated proportion, and to (2) determine which linear estimator tended to give the lowest estimated sampling error. Estimator behavior was compared under stratified and unstratified conditions, and with sample unit observations weighted by sample unit size and not weighted in any way.

The list of estimators of irrigated proportion and variance is given below:

1) Regression estimator: a straight line relationship between X and Y that may intercept the y-axis at any point; this estimator will not be significantly biased if the relationship between X and Y is linear; it will give the lowest sample variance among the class of parametric linear estimators if the paired (x,y) observations are distributed about the regression line in the same manner over the range of X; i.e., the error variance (or equivalent of the variance of Y) is independent of X.

   Two forms of the variance for this estimator were considered in this evaluation:

   a) a form of regression variance that used the set of Landsat sample unit measurements (x_i) obtained in the 1979 sample to compute the variance of the estimated slope; this was the form of the variance used to estimate errors reported in the results section.

   b) a form of regression variance that used the expected (average) value for the variance of slope assuming the entire population of (x_i) by basin to be normally distributed; this form of variance is most often used in computing sample size, n; but is not sample-specific;

2) Biased ratio estimator: a straight line relationship between X and Y that must pass through the origin; this estimator will be biased on the order of 1/n, i.e. the bias will decrease in direct proportion to the increase in sample size; this ratio estimator will produce the lowest sample variance among the class of simple linear estimators when the variance of Y increases in direct proportion to the size of X;

3) Unbiased ratio estimator: of the several suggested in the literature, the one by Goodman and Hartley (1958) was selected for analysis here; a straight line relationship between X and Y is assumed, but is not constrained to pass through the origin; this estimator should give the lowest estimated variance when the variance of Y increases in direct proportion to the square of X;

4) Difference estimator: a straight line relationship between X and Y not constrained to pass through the origin; the estimate of $\bar{Y}$ is obtained by estimating the difference between $\bar{Y}$ and $k\bar{X}$ and then adding that difference to $k\bar{X}$; this estimator is seen to have the same form as the regression estimator, except that the constant k is pre-assigned (instead of estimated as b) according to previous experience or according to some expectation based on a conceptual model of the relationship between X and Y; the advantage of the difference estimator, in addition to its simplicity, is that more degrees of freedom are available - thereby giving a smaller Student's t-statistic and therefore a smaller confidence interval half-width than a corresponding regression estimate, assuming k has been chosen properly; this potential advantage over regression is also shared by the ratio estimators;

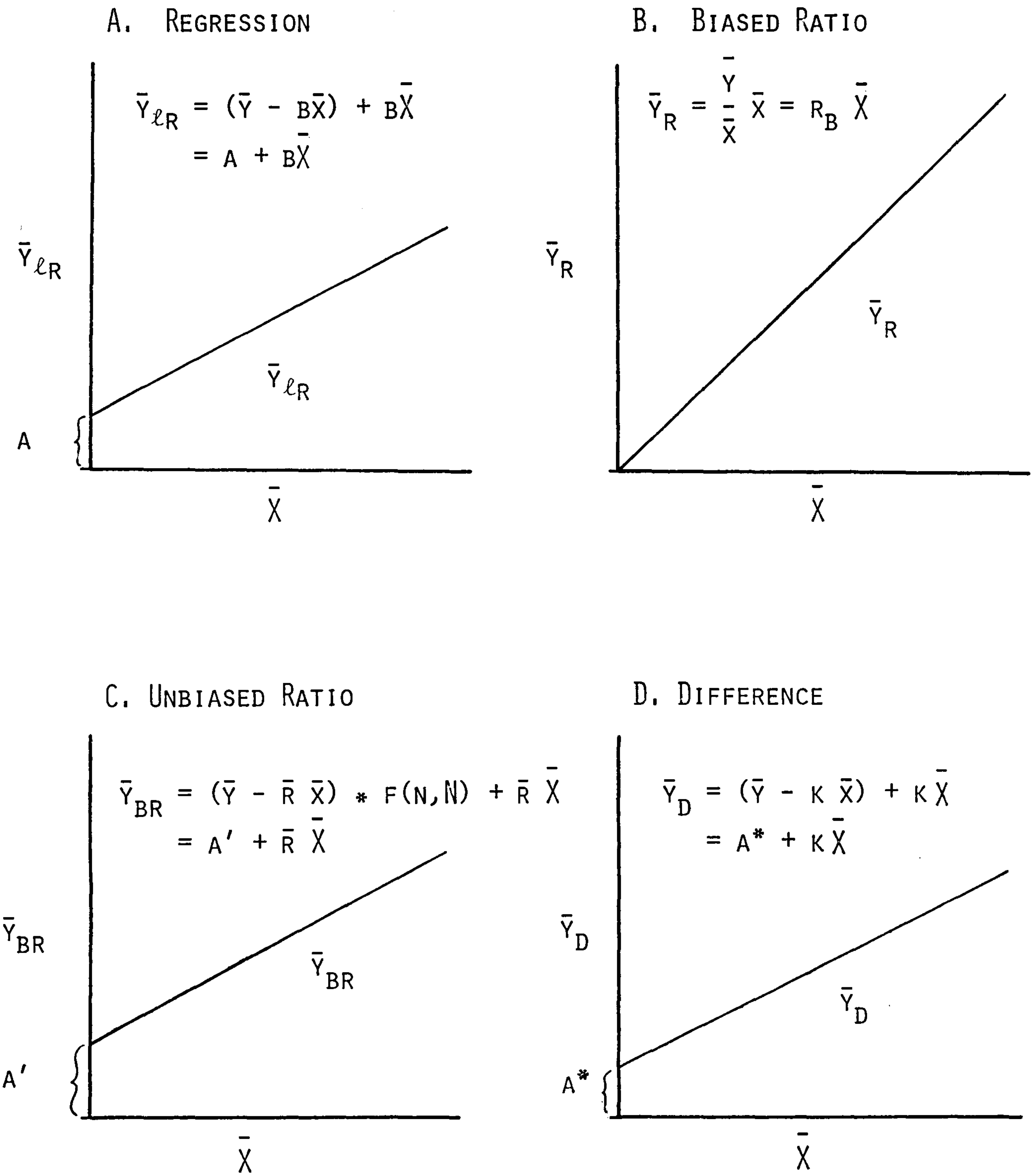
5) Combined regression and unbiased ratio estimator: at low sample size within a stratum the analysis reported last year in Wall et al (1980) indicated that the unbiased ratio estimator would give lower estimated sample variance than its regression counterpart; thus a combined estimator was defined such that if sample size was less than or equal to a predetermined number in a given stratum, then unbiased ratio estimation was used - otherwise regression estimation was employed; combined re-regression/ratio estimates were computed for both forms of the regression estimator variance introduced above.

The regression, biased ratio, unbiased ratio, and difference estimators are shown in Figure 3-13. There it can be seen that each uses a different estimate of slope and intercept for the straight line relating Y to X. A formal presentation of each of the estimators described above is given in Appendix II. Note that the primary regression estimator (#1a) described above and used to produce the estimates reported in the estimation results section, is termed regression with factor 5 (f5) in that Appendix. The second regression estimator (#1b above) is denoted as regression with factor 3 (f3). Detailed estimation results for each basin are presented in that appendix for each of the estimators introduced here.

Comparison of Weighted Estimators

When the estimators described above were applied to observations weighted by size of sample unit, sample unit size-weighted, or simply 'weighted', estimators resulted. These weighted estimators were than compared in terms of their estimated

FIGURE 3-13. TYPES OF LINEAR ESTIMATORS EXAMINED



WHERE $\bar{R} = \Sigma \frac{Y_I}{X_I} \div N$

WHERE K IS SOME CHOSEN CONSTANT, E.G. K=1 error and estimated basin proportion irrigated. The purpose of this comparison was to (1) identify the estimator(s) tending to give the lowest estimated error among the basin populations sampled, (2) determine the relative performance of the estimators under different joint X,Y distributions, and (3) determine the difference between the regression estimate of proportion irrigated and similar estimates produced by the other linear estimators. It was recognized that the exact differences among estimators would depend on the particular sample of ground units drawn in each basin. However, major trends were expected to reflect real differences.

The method of comparison proceeded by first ranking each estimator according to the size of its estimated error within a given basin. Thus, in the case of unstratified sampling, standard errors were ordered from lowest to highest within a given basin - the estimator having the lowest standard error being assigned the rank of 1 and the highest the rank of 5.* When a tie occurred, each estimator was assigned the average of the next two ranks not yet awarded. Standard errors were ranked separately in each basin. A similar procedure was followed in the stratified situation except that the 95 percent confidence interval half-widths were ranked instead of the standard errors. This was done to incorporate the estimator-specific effect of the formula (equation 28 of Appendix IB) used for calculating degrees of freedom in the stratified case. Tables 3-32 and 3-33 present the standard errors, confidence interval half-widths, degrees of freedom, and estimates of irrigated proportion for each estimator by basin in the stratified and unstratified cases, respectively. Numbers in parenthesis represent the assigned ranks.

Ranks were then summed by estimator over the ten basins. These rank sums, presented in Table 3-34, were used to judge the relative overall error performance of the different linear estimators. Use of ranks for this purpose enabled identification of general trends while deemphasizing exact differences between estimated errors resulting from the particular ground sample chosen.

In order to explain the ranking results and thereby qualify the conclusions of the estimator comparison, the distribution of ranks over basins were considered in the light of the joint distribution of X and Y. Each estimator will tend to perform better or worse, depending on the shape of the joint distribution of Landsat (X) and ground (Y) observations. Plots for each basin were prepared showing Y versus X, the residuals of the regression versus X, and the residuals versus sample unit weight.** Sample unit observations represented irrigated proportions weighted by relative sample unit size. Stratified observations were weighted by the ratio of the sample size for the given sample unit (A_{hi}) to the

* The two combined regression/ratio estimators were not employed in the unstratified case, leaving only 5 estimators.

** The residual of the regression of y on x, defined, for a given observation x_i, to be the difference between the value $\hat{y}_i$ predicted from the regression and the value y_i as observed on the ground. It is a measure of how well the regression explains the dependent variable. A residual plot is intended to identify departures from assumptions of independence and specification; there should be no obvious trends in the plots.

Table 3-32. Results for the various estimators using the stratified observations.

BASIN	REGRESSION (f3)	REGRESSION (f5)	UNBIASED RATIO	BIASED RATIO	DIFFER-ENCE	COMBINED (f3)	COMBINED (f5)
			STANDARD ERRORS				
North Coast	.00982	.01005	.01975	.01051	.01059	.01036	.01058
San Francisc o	.00613	.00578	.01367	.00681	.00601	.00953	.00937
Central Coast	.00929	.00911	.01693	.01123	.00944	.00939	.00922
South Coast	.01272	.01421	.01287	.01262	.01226	.01272	.01421
Colorado Desert	.00701	.00693	.00727	.00703	.00688	.00702	.00694
South Lahontan	.01837	.01868	.01829	.01860	.01779	.01837	.01868
North Lahontan	.01302	.01309	.01307	.01315	.01263	.01302	.01309
Sacramento	.00887	.00895	.01007	.01050	.00968	.00908	.00904
San Joaquin	.01335	.01270	.01518	.01398	.01372	.01317	.01295
Tulare	.00993	.00998	.00968	.00950	.00949	.01014	.01019
			95% CONFIDENCE INTERVAL HALF WIDTHS				
North Coast	(1) .01996	(2) .02040	(7) .04188	(4) .02138	(6) .02158	(3) .02095	(5) .02140
San Francisco	(3) .01283	(1) .01215	(7) .02897	(4) .01405	(2) .01250	(5) .02253	(6) .02293
Central Coast	(3) .01877	(1) .01842	(7) .03453	(6) .02267	(5) .01907	(4) .01894	(2) .01861
South Coast	(3.5).02553	(6.5).02876	(5) .02574	(2) .02526	(1) .02452	(3.5).02553	(6.5).02876
Colorado Desert	(5.5).01416	(2) .01399	(7) .01462	(4) .01415	(1) .01388	(5.5).01416	(3) .01401
South Lahontan	(3.5).03746	(6.5).03810	(2) .03725	(5) .03790	(1) .03624	(3.5).03746	(6.5).03810
North Lahontan	(3.5).02672	(5.5).02676	(2) .02670	(7) .02685	(1) .02583	(3.5).02672	(5.5).02676
Sacramento	(2) .01780	(3) .01795	(6) .02021	(7) .02124	(5) .01941	(4) .01819	(1) .01390
San Joaquin	(4) .02693	(1) .02551	(7) .03044	(6) .02800	(5) .02750	(3) .02643	(2) .02598
Tulare	(4) .01995	(5) .02004	(3) .01939	(2) .01905	(1) .01901	(6) .02034	(7) .02044
			DEGREES OF FREEDOM				
North Coast	34.73	35.27	16.41	33.91	32.56	39.15	39.76
San Francisco	19.16	18.51	16.64	24.10	21.18	7.36	6.96
Central Coast	40.51	40.41	31.43	41.62	40.83	42.03	42.11
South Coast	52.86	38.61	60.65	58.68	59.23	52.86	38.61
Colorado Desert	42.43	42.52	47.93	45.57	43.42	42.52	42.73
South Lahontan	31.00	31.00	32.00	32.00	32.00	31.00	31.00
North Lahontan	27.99	29.27	30.82	30.13	29.16	27.99	27.27
Sacramento	52.95	52.82	51.73	39.15	54.61	55.65	55.29
San Joaquin	43.17	50.69	54.97	56.26	55.34	51.80	53.04
Tulare	49.81	51.71	56.76	54.08	55.10	52.60	54.53
			IRRIGATED PROPORTION				
North Coast	.53521	.53521	.53451	.53600	.53530	.53472	.53472
San Francisco	.21852	.21852	.21163	.21785	.21807	.21279	.21279
Central Coast	.31906	.31906	.32208	.32145	.32074	.31949	.31949
South Coast	.45787	.45787	.46696	.46408	.45102	.45787	.45787
Colorado Desert	.82147	.82147	.82376	.82281	.82058	.82144	.82144
South Lahontan	.27383	.27383	.27683	.27291	.27320	.27383	.27383
North Lahontan	.58726	.58726	.58248	.58362	.58559	.58726	.58726
Sacramento	.65381	.65381	.65183	.65214	.65521	.65059	.65059
San Joaquin	.74778	.74778	.75390	.75258	.75432	.74635	.74635
Tulare	.82038	.82038	.82246	.82197	.82316	.82168	.82168

Table 3-33. Results for the various estimators using the unstratified observations.

BASIN	REGRESSION (f3)	REGRESSION (f5)	UNBIASED RATIO	BIASED RATIO	DIFFER-ENCE	STRATA COMBINED
			STANDARD ERRORS			
North Coast	(2) .01082	(4) .01187	(5) .01636	(3) .01089	(1) .01063	.00968
San Francisco	(1) .01051	(5) .01906	(4) .01527	(3) .01128	(2) .01060	.00618
Central Coast	(2) .01072	(3) .01080	(5) .01416	(4) .01164	(1) .01062	.00911
South Coast	(3) .01214	(1.5).01207	(5) .01555	(4) .01322	(1.5).01207	.01203
Colarado Desert	(3) .00631	(4) .00647	(5) .00663	(2) .00630	(1) .00620	.00693
South Lahontan	(3) .01837	(5) .01868	(2) .01829	(4) .01860	(1) .01779	.01868
North Lahontan	(3) .01219	(1) .01207	(5) .01306	(4) .01264	(2) .01216	.01207
Sacramento	(1.5).00817	(1.5).00817	(5) .01001	(3) .00867	(4) .00918	.00881
San Joaquin	(2) .01164	(1) .01161	(5) .01366	(3) .01255	(4) .01259	.01208
Tulare	(4) .01008	(5) .01047	(2.5).00998	(2.5).00998	(1) .00997	.01001
			95% CONFIDENCE INTERVAL HALF WIDTHS			
North Coast	.02177	.02387	.03290	.02189	.02137	.01958
San Francisco	.02106	.02198	.03059	.02260	.02123	.01297
Central Coast	.02135	.02150	.02818	.02317	.02115	.01825
South Coast	.02417	.02402	.03095	.02631	.02401	.02405
Colorado Desert	.01265	.01296	.01328	.01262	.01242	.01398
South Lahontan	.03746	.03810	.03725	.03790	.03624	.03810
North Lahontan	.02475	.02451	.02650	.02564	.02466	.02451
Sacramento	.01629	.01630	.01996	.01728	.01831	.01766
San Joaquin	.02319	.02312	.02720	.02501	.02508	.02420
Tulare	.02015	.02093	.01994	.01994	.01991	.02009
			DEGREES OF FREEDOM			
North Coast	47.00	47.00	48.00	48.00	48.00	39.55
San Francisco	54.00	54.00	55.00	55.00	55.00	23.88
Central Coast	77.00	77.00	78.00	78.00	78.00	57.27
South Coast	79.00	79.00	80.00	80.00	80.00	62.33
Colorado Desert	56.00	56.00	57.00	57.00	57.00	42.58
South Lahontan	31.00	31.00	32.00	32.00	32.00	31.00
North Lahontan	35.00	35.00	36.00	36.00	36.00	35.00
Sacramento	70.00	70.00	71.00	71.00	71.00	55.03
San Joaquin	74.00	74.00	75.00	75.00	75.00	57.45
Tulare	63.00	63.00	64.00	64.00	64.00	52.60
			IRRIGATED PROPORTION			
North Coast	.53182	.53182	.56517	.53344	.53461	.53920
San Francisco	.21192	.21192	.18562	.20481	.20595	.21848
Central Coast	.32579	.32579	.31270	.32014	.32451	.31876
South Coast	.45251	.45251	.44940	.45121	.45289	.45504
Colorado Desert	.82245	.82245	.82626	.82386	.82255	.82107
South Lahontan	.27383	.27383	.27683	.27291	.27320	.27383
North Lahontan	.58725	.58725	.58990	.58925	.58881	.58725
Sacramento	.65443	.65443	.66009	.65745	.65911	.65260
San Joaquin	.75164	.75164	.75685	.75532	.75565	.74769
Tulare	.81458	.81458	.81593	.81445	.81677	.81890

average size of sample unit in that stratum ($\bar{A}_h$). Unstratified weighting was performed by ignoring strata and forming a weight from the ratio of sample unit size (A_i) for sample unit i to the average size over the whole basin ($\bar{A}$). More detail on weighting is given in Appendices I and II. Figures 3-14 to 3-25 show examples of the resulting plots for several basins.

The relative error performance between stratified and unstratified application of the linear estimators was also evaluated. Differences in the relative ordering of the summed ranks for each estimator between stratified and unstratified situations was noted. In addition, a comparison was made of the relative size of standard errors and confidence interval half-widths for corresponding stratified and unstratified estimates produced by the same estimators.

Results for Weighted Estimators

General observations regarding error ranking were as follows:

(1) On the average, the regression and difference estimators produced the narrowest confidence interval half-widths for the stratified case, and the smallest standard errors for the unstratified case;

(2) On the average, the difference estimator produced somewhat smaller standard errors and confidence interval half-widths than the regression estimator;

(3) Stratified factor 3 (assumption of normally-distributed Landsat observations) and factor 5 (no assumption made on the distribution of Landsat observations) regression ranked closely, though factor 3 gave smaller confidence interval half-widths in six out of ten basins;

(4) The biased and unbiased ratio estimators generally gave the largest standard errors or confidence interval half-widths, the latter estimator having the average highest error ranking;

(5) The combined unbiased ratio (at low sample size) and regression estimators had average error rankings midway between the regression and ratio estimators; and

(6) Stratification produced smaller standard errors and confidence interval half-widths in four out of nine basins where a comparison was possible.

Patterns of estimator error ranking relative to assumptions on the joint distribution of paired (x,y) observations were as follows:

(1) The difference estimator did best in basins where the slope in all land use strata was close to unity (one);

Table 3-34. Rank sums for the various estimators

	Regression (f3)	Regression (f5)	Unbiased Ratio	Biased Ratio	Differ-ence	Combined (f3)	Combined (f5)
Stratified*	33	33.5	53	47	28	41	45
Unstratified**	24.5	31	43.5	32.5	18.5	--	--
Unweighted***	--	20	38	27	15	--	--

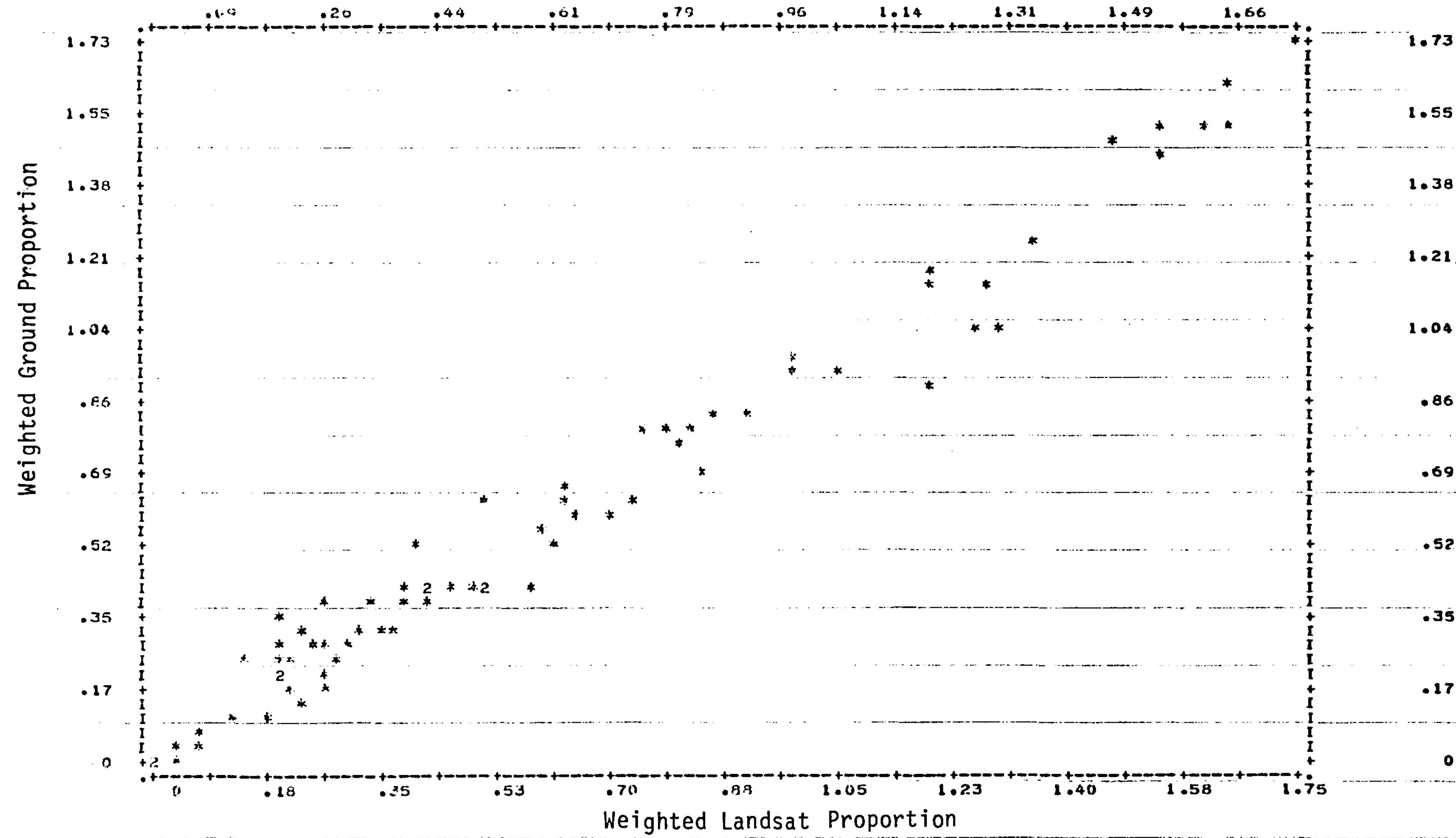
*Difference between one or more rank sums found to be significant at the $\alpha=.10$ level using the non-parametric Friedman statistic for comparison of blocked treatments. Dependence between estimators due to use of the same sample observations must be ignored to use this statistic.

**Difference between rank sums found to be significant at the $\alpha=.01$ level using the Friedman statistic.

***Difference between rank sums found to be significant at the $\alpha=.01$ level using the Friedman statistic.

Figure 3-14. Unstratified Weighted Ground Observations of Irrigated Proportion Versus Matching Weighted Landsat Observations for the Sacramento Basin

SCATTERGRAM OF (DOWN) WTG
(ACROSS) WTL



UNSTRATIFIED PLOTS 03/24/81 20.51.56. PAGE 6

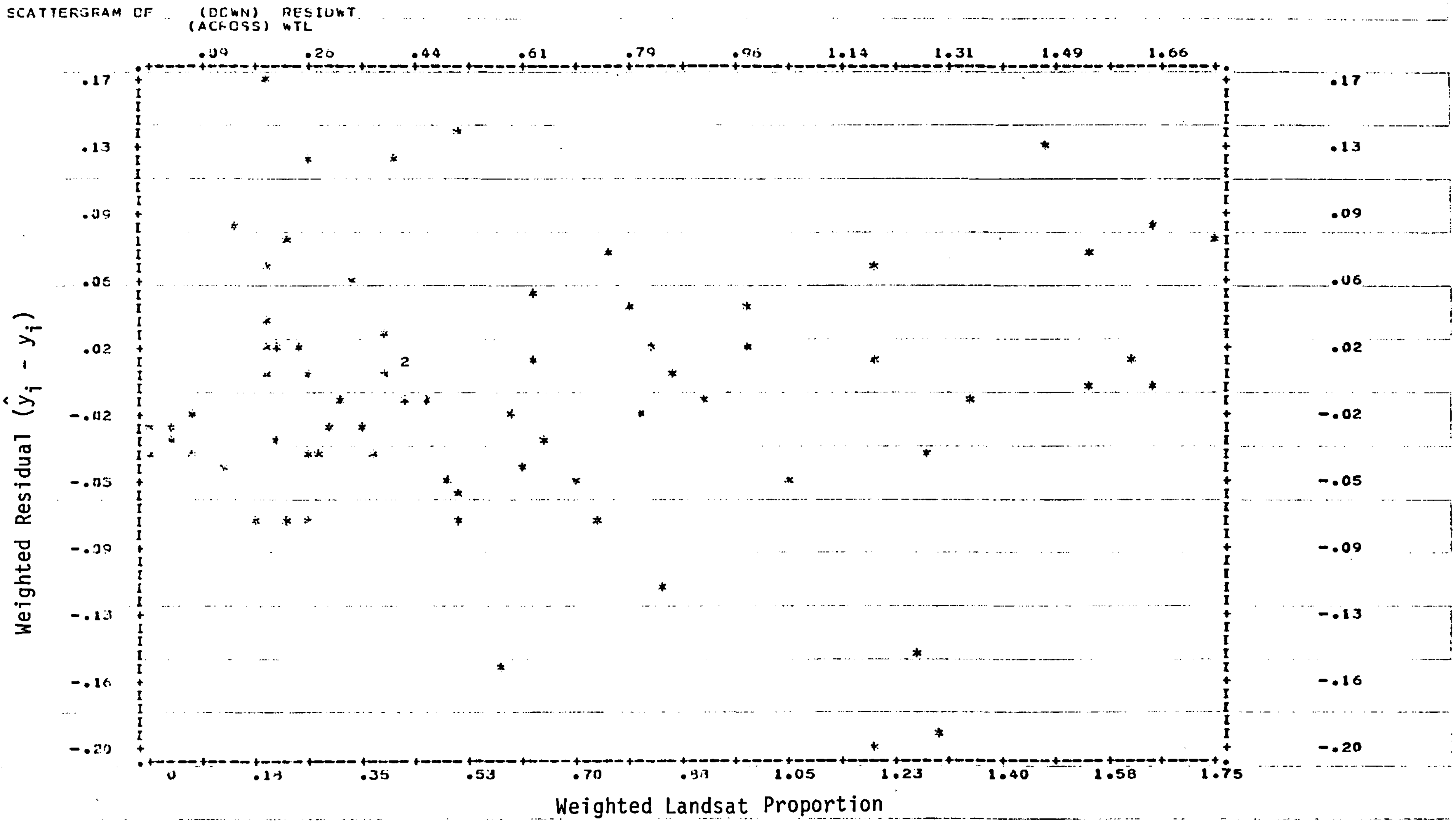
STATISTICS..

CORRELATION (R)-	.98753	R SQUARED -	.97521	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.07051	INTERCEPT (A) -	.03212	STD ERROR OF A -	.01366
SIGNIFICANCE A -	.01075	SLOPE (B) -	.91957	STD ERROR OF B -	.01752
SIGNIFICANCE B -	.00001				

PLOTTED VALUES - 72 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-15. Unstratified Residuals ($\hat{y}_i - y_i$) Versus the Corresponding Weighted Landsat Observations of Irrigated Proportion for the Sacramento Basin



UNSTRATIFIED PLOTS 03/24/81 20.56.36. PAGE 4

STATISTICS..

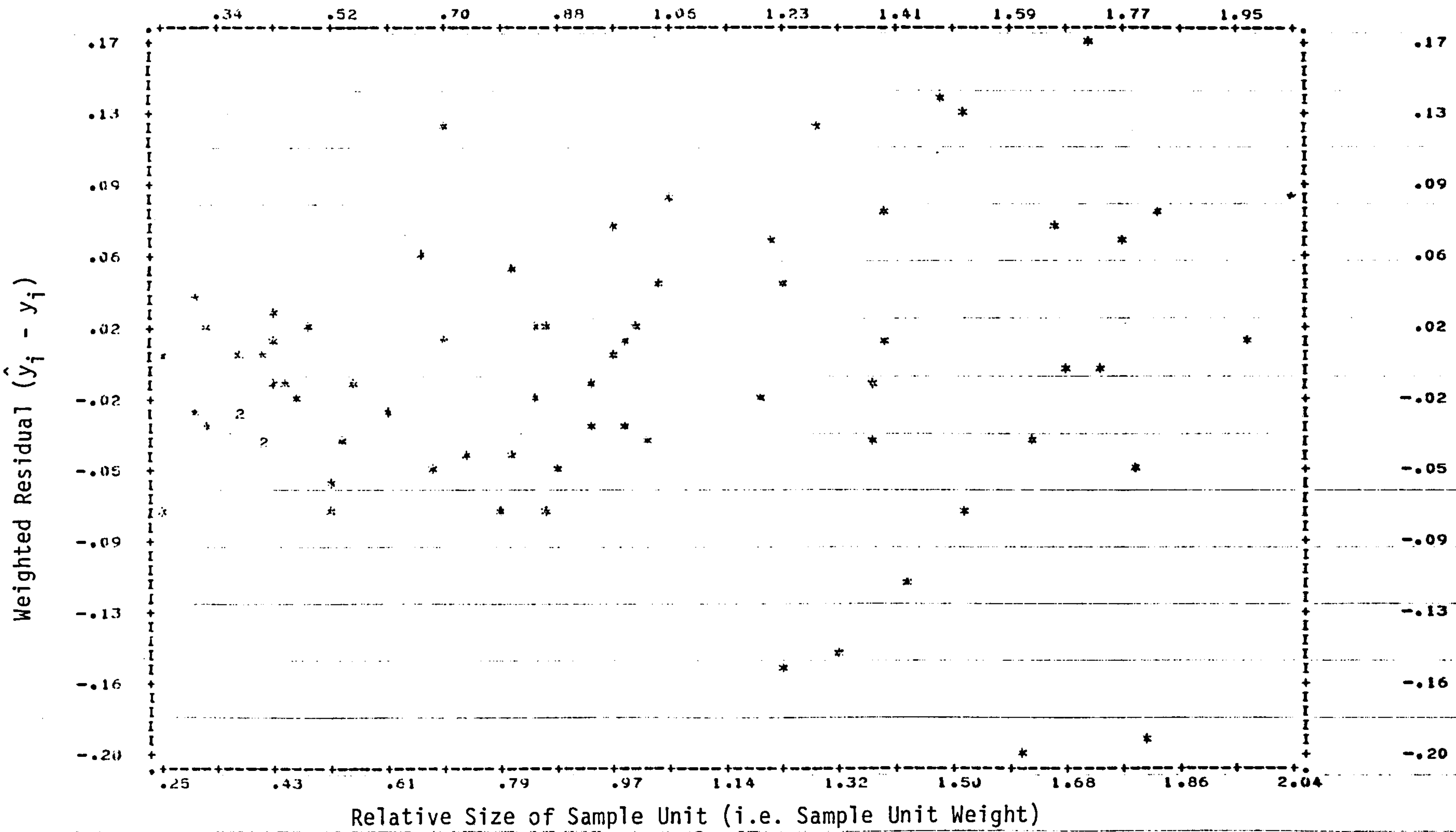
CORRELATION (R)-	.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.07051	INTERCEPT (A) -	-.21766E-08	STD ERROR OF A -	.01366
SIGNIFICANCE A -	.50000	SLOPE (B) -	.32718E-08	STD ERROR OF B -	.01752
SIGNIFICANCE B -	.50000				

PLOTTED VALUES - 72 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-16. Unstratified Residuals Versus Corresponding Relative Size of Sample Unit for the Sacramento Basin

SCATTERGRAM OF (DOWN) RESIDWT
(ACROSS) WT2



-106-

UNSTRATIFIED PLOTS 03/24/81 20.56.36. PAGE 6

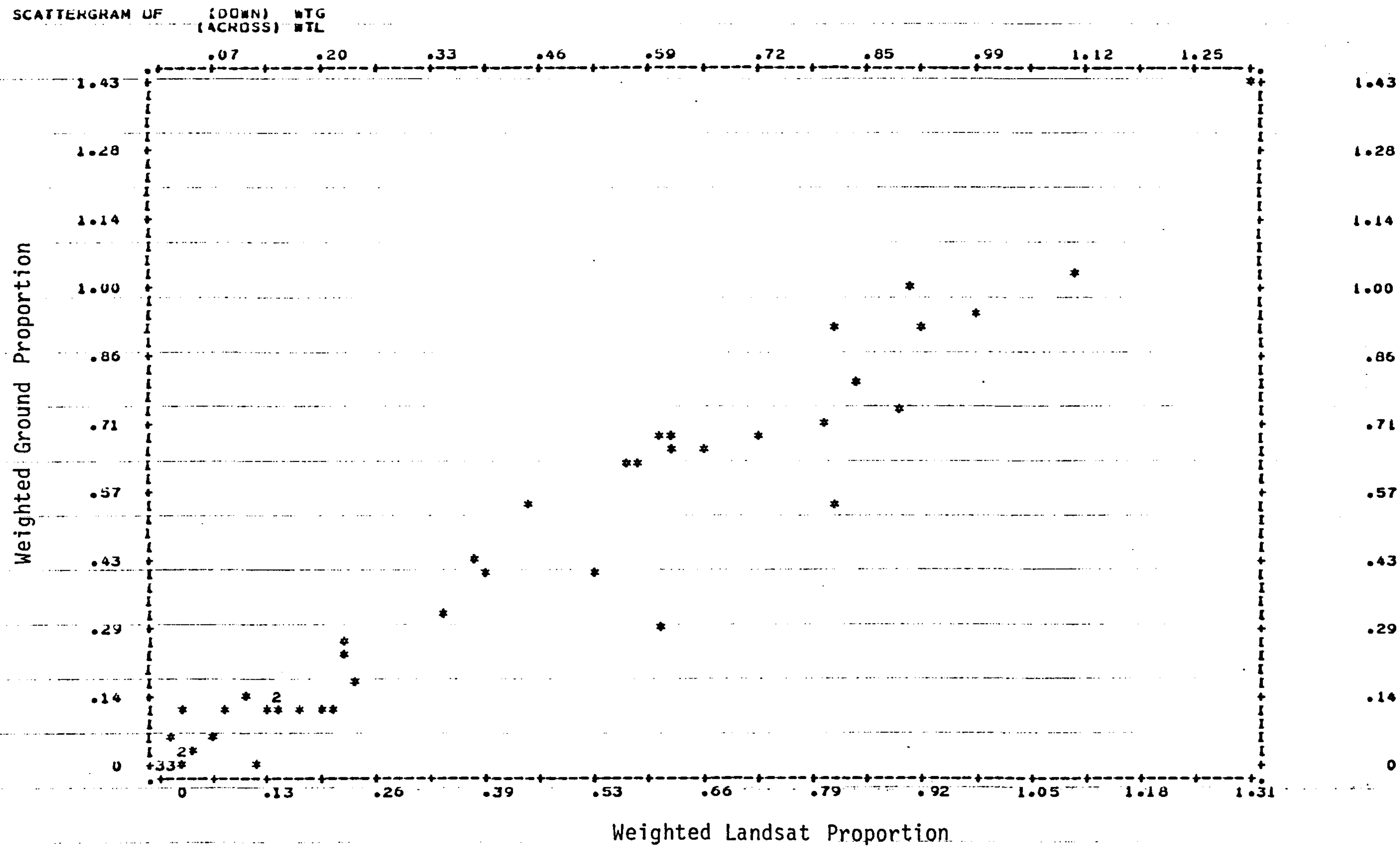
STATISTICS..

CORRELATION (R)-	.10599	R SQUARED -	.01123	SIGNIFICANCE R -	.18777
STD ERR OF EST -	.07011	INTERCEPT (A) -	-.01433	STD ERROR OF A -	.01807
SIGNIFICANCE A -	.21520	SLOPE (B) -	.01478	STD ERROR OF B -	.01657
SIGNIFICANCE B -	.18777				

PLOTTED VALUES - 72 EXCLUDED VALUES - 0 MISSING VALUES - 0

******* IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-17. Unstratified Weighted Ground Observations of Irrigated Proportion Versus Matching Weighted Landsat Observations for the North Coast



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 6

STATISTICS..

CORRELATION (R)-	.97378	R SQUARED -	.94825	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.08338	INTERCEPT (A) -	.00409	STD ERROR OF A -	.01753
SIGNIFICANCE A -	.40824	SLOPE (B) -	.98177	STD ERROR OF B -	.03346
SIGNIFICANCE B -	.00001				

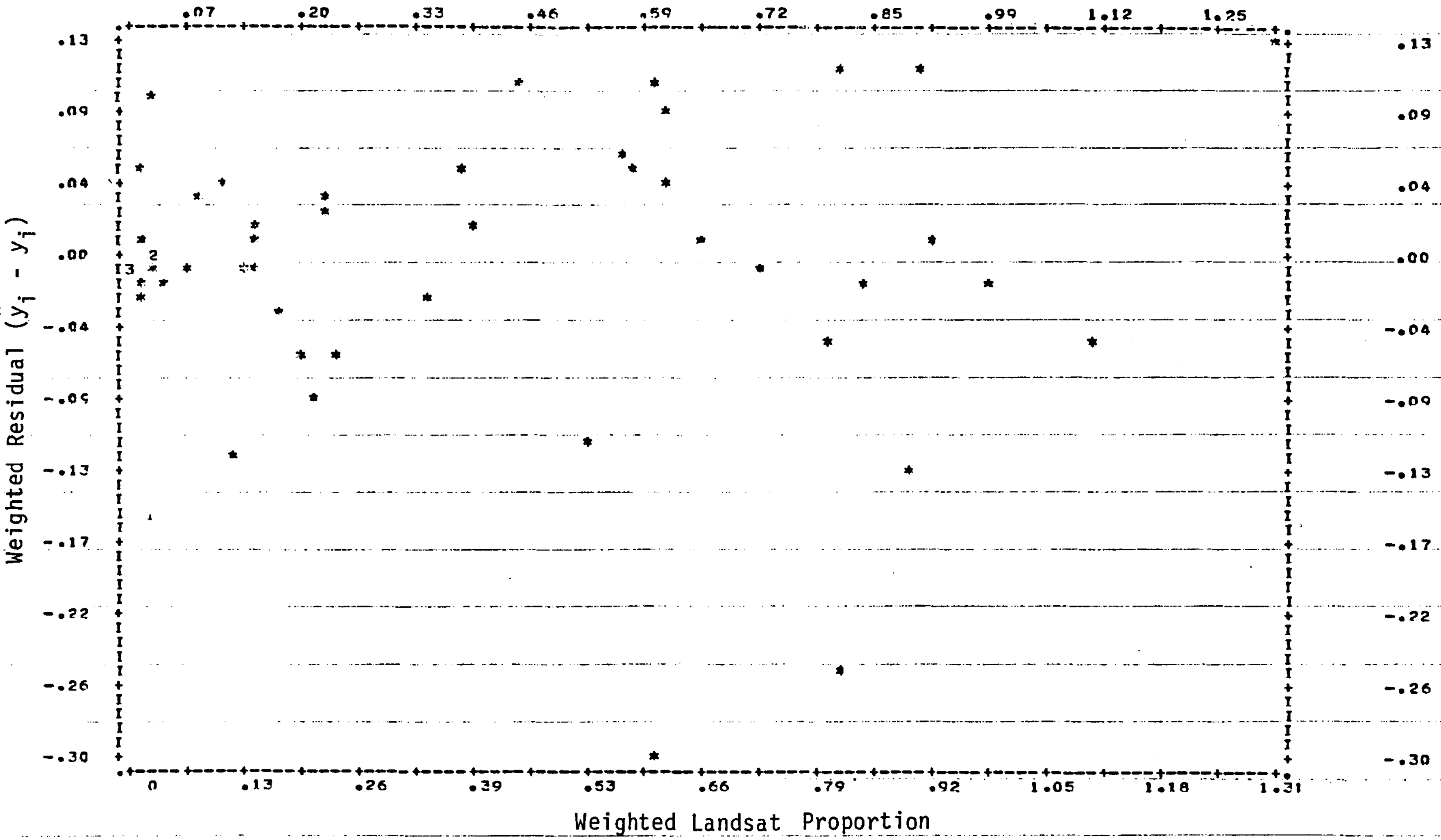
PLOTTED VALUES - 49 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-18. Unstratified Residuals ($\hat{y}_i - y_i$) Versus the Corresponding Weighted Landsat Observations of Irrigated Proportion for the North Coast Basin

FILE FLCTIT (CREATICN CATE = 01/11/82)
SUBFILE NC

SCATTERGRAM OF (DOWN) RESIDWT
(ACROSS) WTL



UNSTRATIFIEC FLCTS 01/11/82 16.45.31. PAGE 4

STATISTICS..

CORRELATICN (R)-	.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.08338	INTERCEPT (A) -	-.23340E-08	STD ERROR OF A -	.01753
SIGNIFICANCE A -	.50000	SLCPE (B) -	.97443E-09	STD ERROR OF B -	.03346
SIGNIFICANCE B -	.50000				

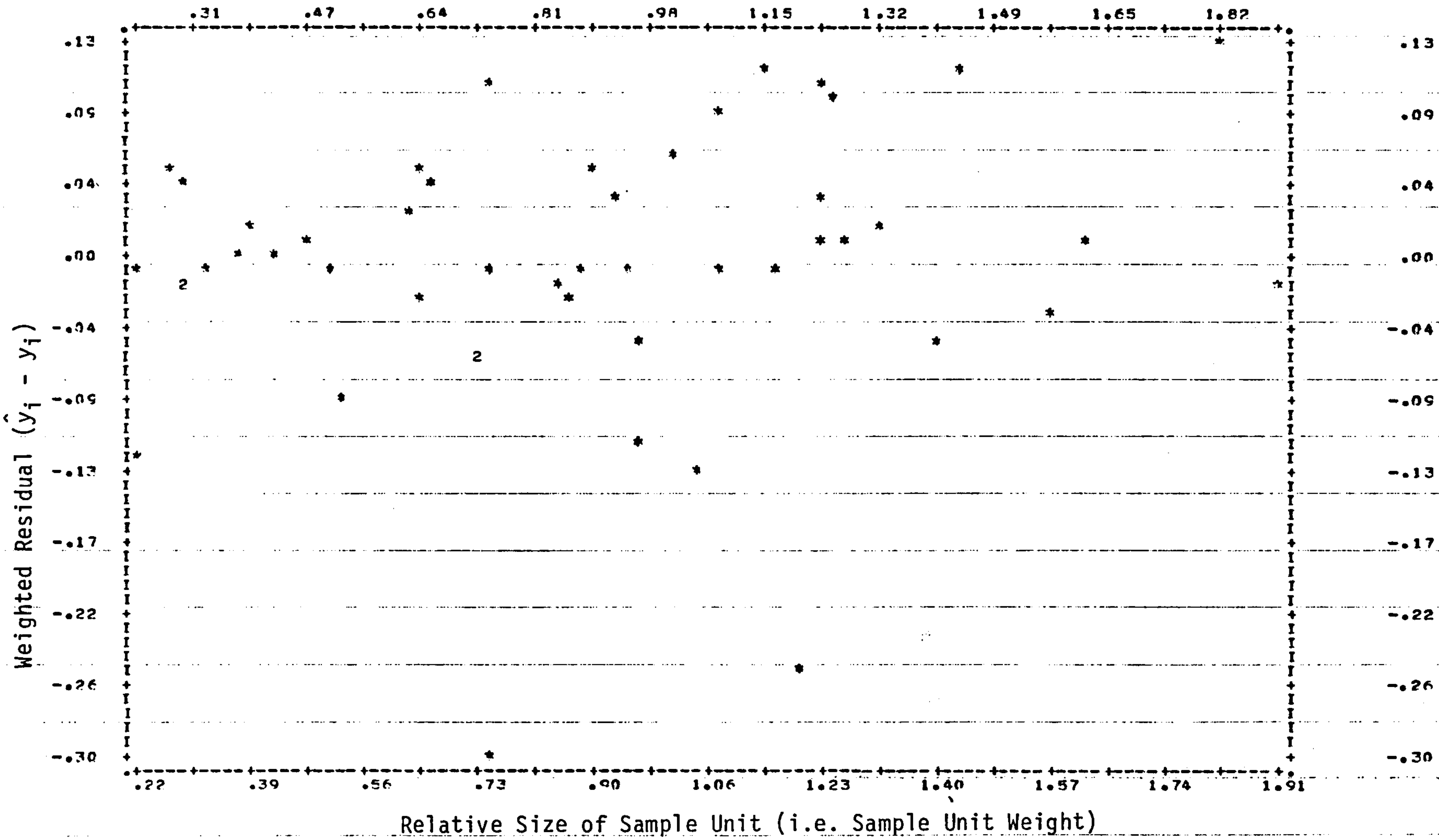
FLCTTED VALUES - 49 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-19. Unstratified Residuals Versus Corresponding Relative Size of Sample Unit for the North Coast Basin

FILE PLOTIT (CREATION DATE = 01/11/82)
SUBFILE NC

SCATTERGRAM OF (DOWN) RESIDWT
(ACROSS) WT2



UNSTRATIFIED PLOTS 01/11/82 16.45.31. PAGE 6

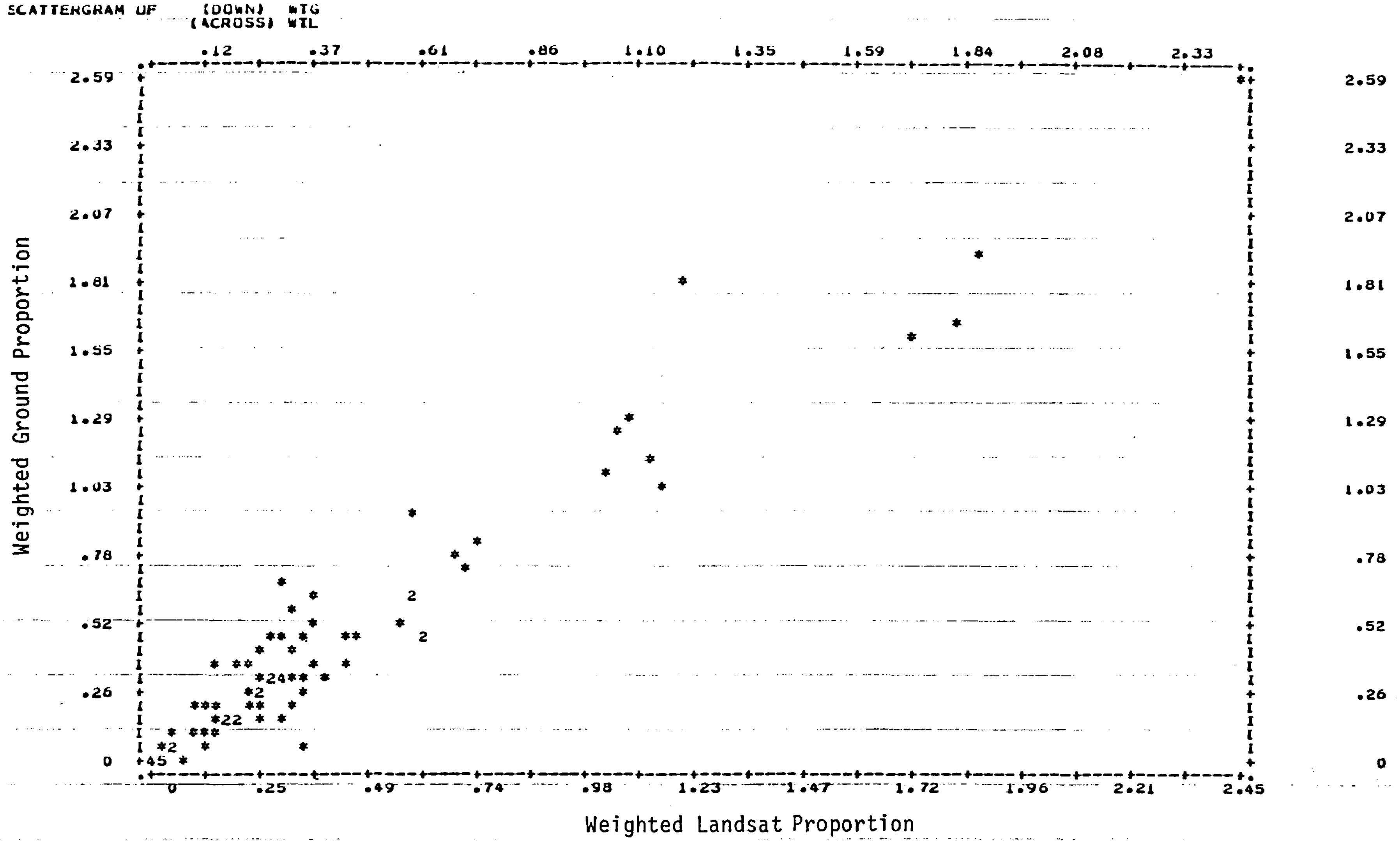
STATISTICS..

CORRELATION (R)-	.15585	R SQUARED -	.02429	SIGNIFICANCE R -	.14246
STD ERR OF EST -	.08237	INTERCEPT (A) -	-.02635	STD ERROR OF A -	.02705
SIGNIFICANCE A -	.16753	SLOPE (B) -	.02991	STD ERROR OF B -	.02765
SIGNIFICANCE B -	.14246				

PLOTTED VALUES - 49 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-20. Unstratified Weighted Ground Observations of Irrigated Proportion Versus Matching Weighted Landsat Observations for the South Coast



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 30

STATISTICS..

CORRELATION (R)-	.96612	R SQUARED -	.93338	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.12739	INTERCEPT (A) -	.04344	STD ERROR OF A -	.01900
SIGNIFICANCE A -	.01246	SLOPE (B) -	1.03137	STD ERROR OF B -	.03100
SIGNIFICANCE B -	.00001				

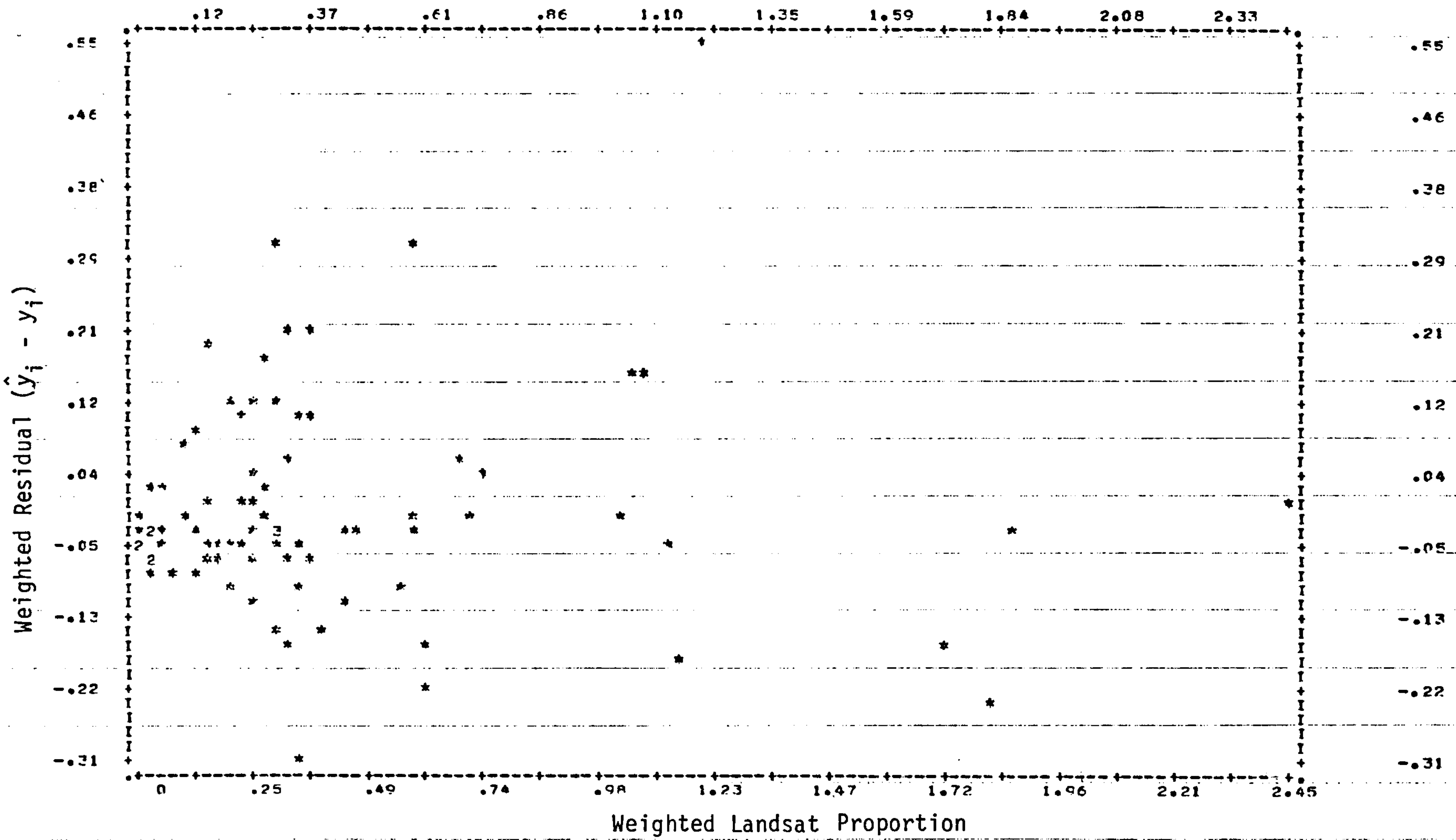
PLOTTED VALUES - 81 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-21. Unstratified Residuals $(y_i - y_i)$ Versus the Corresponding Weighted Landsat Observations of Irrigated Proportion for the South Coast Basin

FILE FLCTIT (CREATION DATE = 01/11/82)
SUBFILE SC

SCATTERGRAM OF (DOWN) RESIDWT
(ACROSS) WTL



UNSTRATIFIED FLCTS 01/11/82 16.45.31. PAGE 16

STATISTICS..

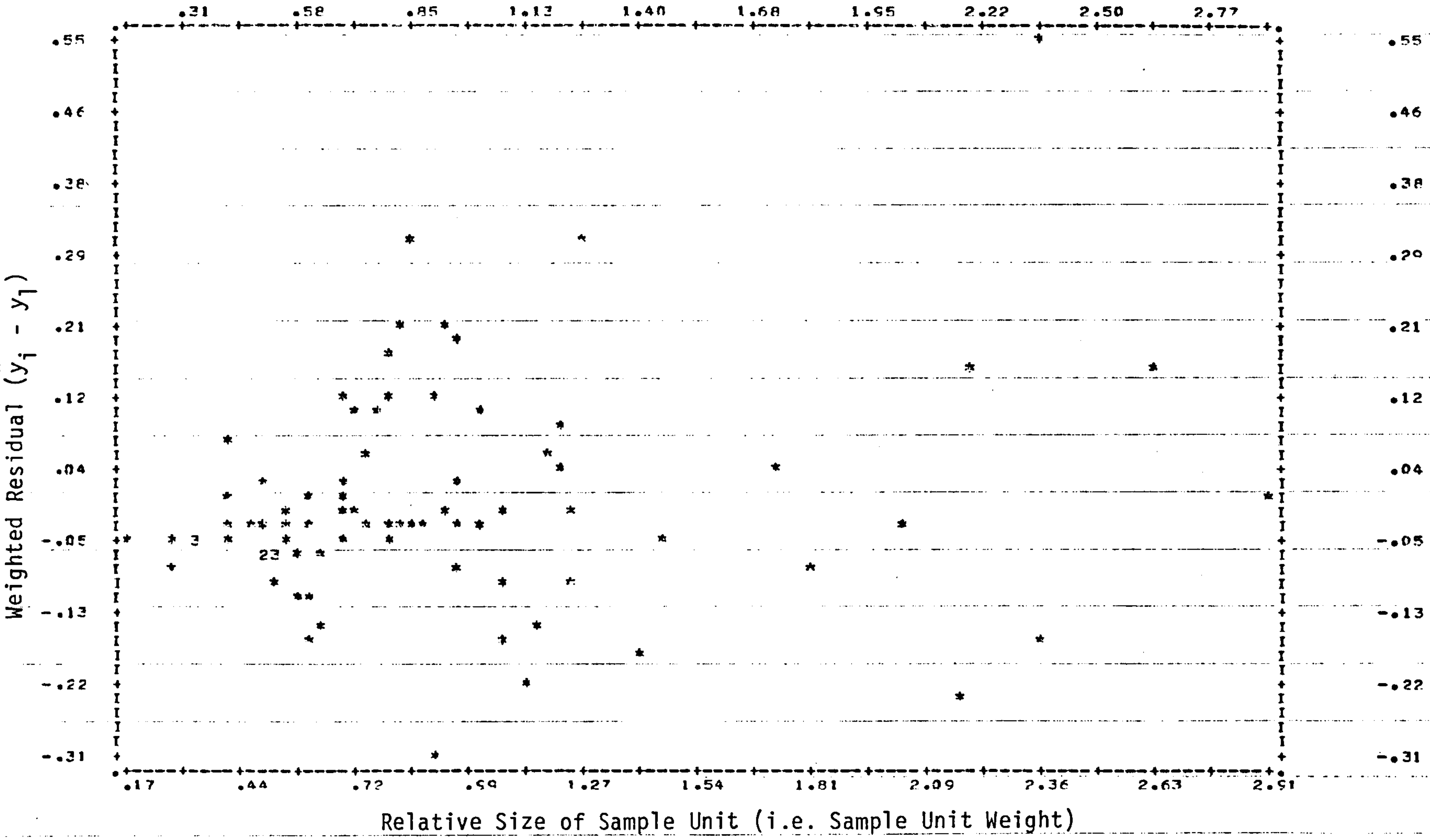
CORRELATION (R)-	.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.12739	INTERCEPT (A) -	-.24155E-08	STD ERROR OF A -	.01900
SIGNIFICANCE A -	.50000	SLOPE (B) -	.14027E-08	STD ERROR OF B -	.03100
SIGNIFICANCE B -	.50000				

PLOTTED VALUES - 81 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-22. Unstratified Residuals Versus Corresponding Relative Size of Sample Unit for the South Coast Basin

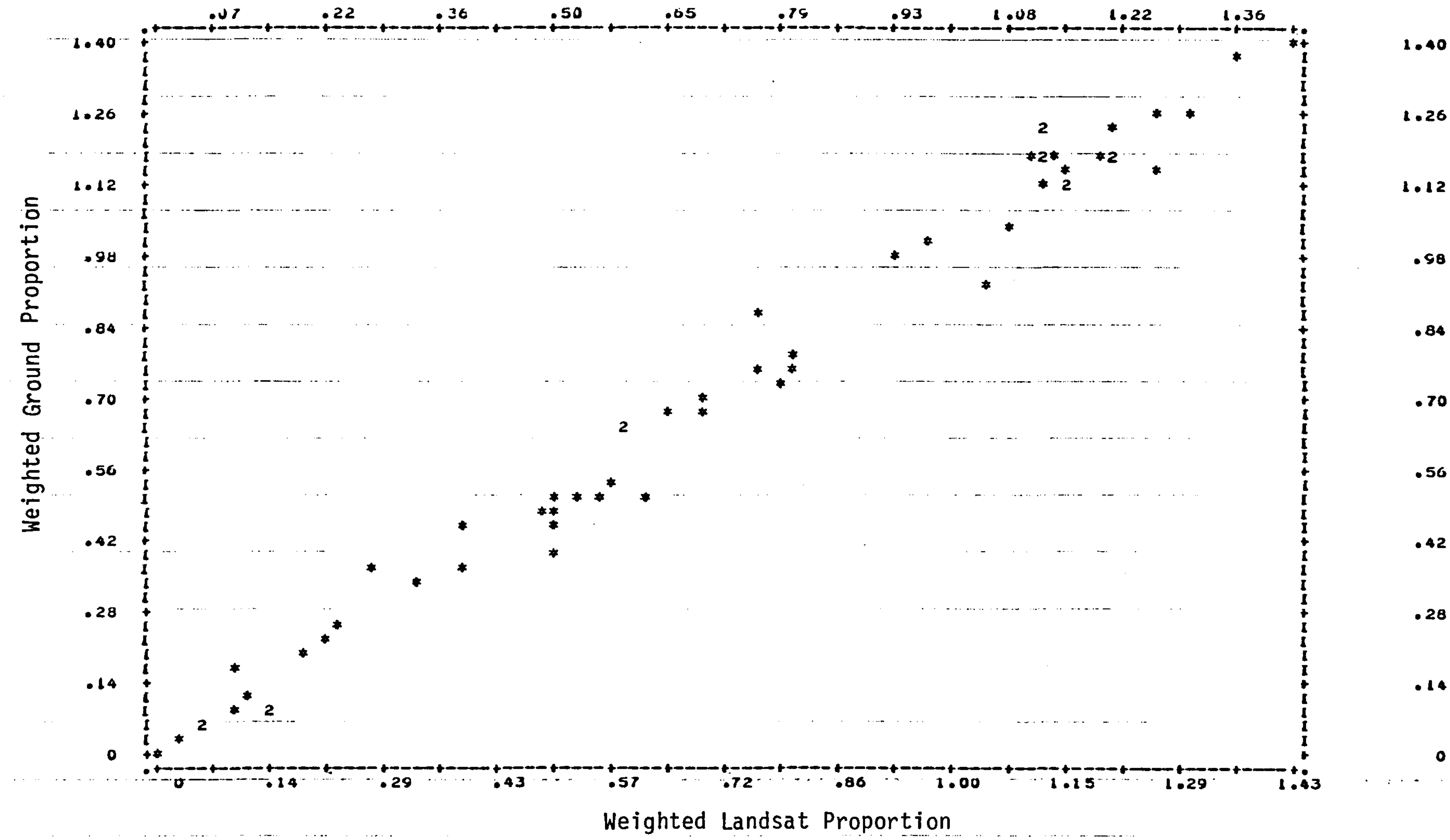
```
FILE    PLOTIT    (CREATION DATE =  01/11/82 )
SUBFILE    SC

SCATTERGRAM OF     (DOWN)  RESIDWT
                   (ACROSS) WT2
```




```
UNSTRATIFIED PLOTS                                        01/11/82    16.45.31.    PAGE   18

STATISTICS..

  CORRELATION (R)-      .18480    R SQUARED      -    .03415    SIGNIFICANCE R -    .04930
  STD ERR OF EST -      .12520    INTERCEPT (A)  -   -.03872    STD ERROR OF A -    .02702
  SIGNIFICANCE A -      .07752    SLOPE (B)      -    .04199    STD ERROR OF B -    .02512
  SIGNIFICANCE B -      .04930

PLOTTED VALUES -   81    EXCLUDED VALUES -    0    MISSING  VALUES -    0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.
```

Figure 3-23. Unstratified Weighted Ground Observations of Irrigated Proportion Versus Matching Weighted Landsat Observations for the Colorado Desert

SCATTERGRAM OF (DOWN) WTG
(ACROSS) WTL



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 38

STATISTICS..

CORRELATION (R)-	.99244	R SQUARED -	.98494	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.05277	INTERCEPT (A) -	.01009	STD ERROR OF A -	.01367
SIGNIFICANCE A -	.23183	SLOPE (B) -	.99902	STD ERROR OF B -	.01651
SIGNIFICANCE B -	.00001				

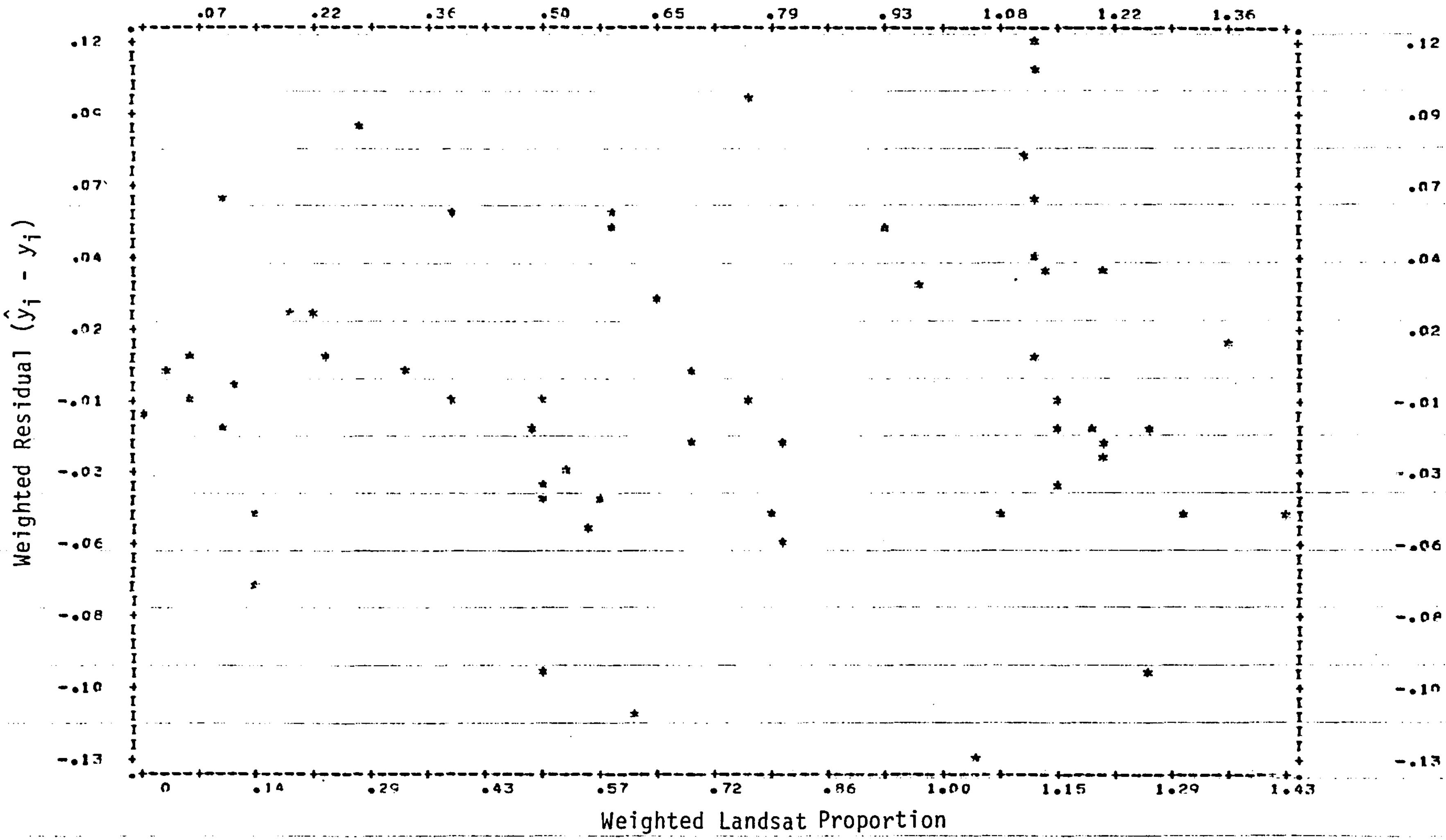
PLOTTED VALUES - 58 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-24. Unstratified Residuals $(\hat{y}_i - y_i)$ Versus the Corresponding Weighted Landsat Observations of Irrigated Proportion for the Colorado Desert Basin

FILE FLOTIT (CREATION DATE = 01/11/82)
SUBFILE CO

SCATTERGRAM OF (DOWN) RESIDWT
(ACROSS) WTL



UNSTRATIFIED PLOTS 01/11/82 16.45.31. PAGE 20

STATISTICS..

CORRELATION (R)-	-.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.05277	INTERCEPT (A) -	.10539E-08	STD ERROR OF A -	.01367
SIGNIFICANCE A -	.50000	SLOPE (B) -	-.53036E-08	STD ERROR OF B -	.01651
SIGNIFICANCE B -	.50000				

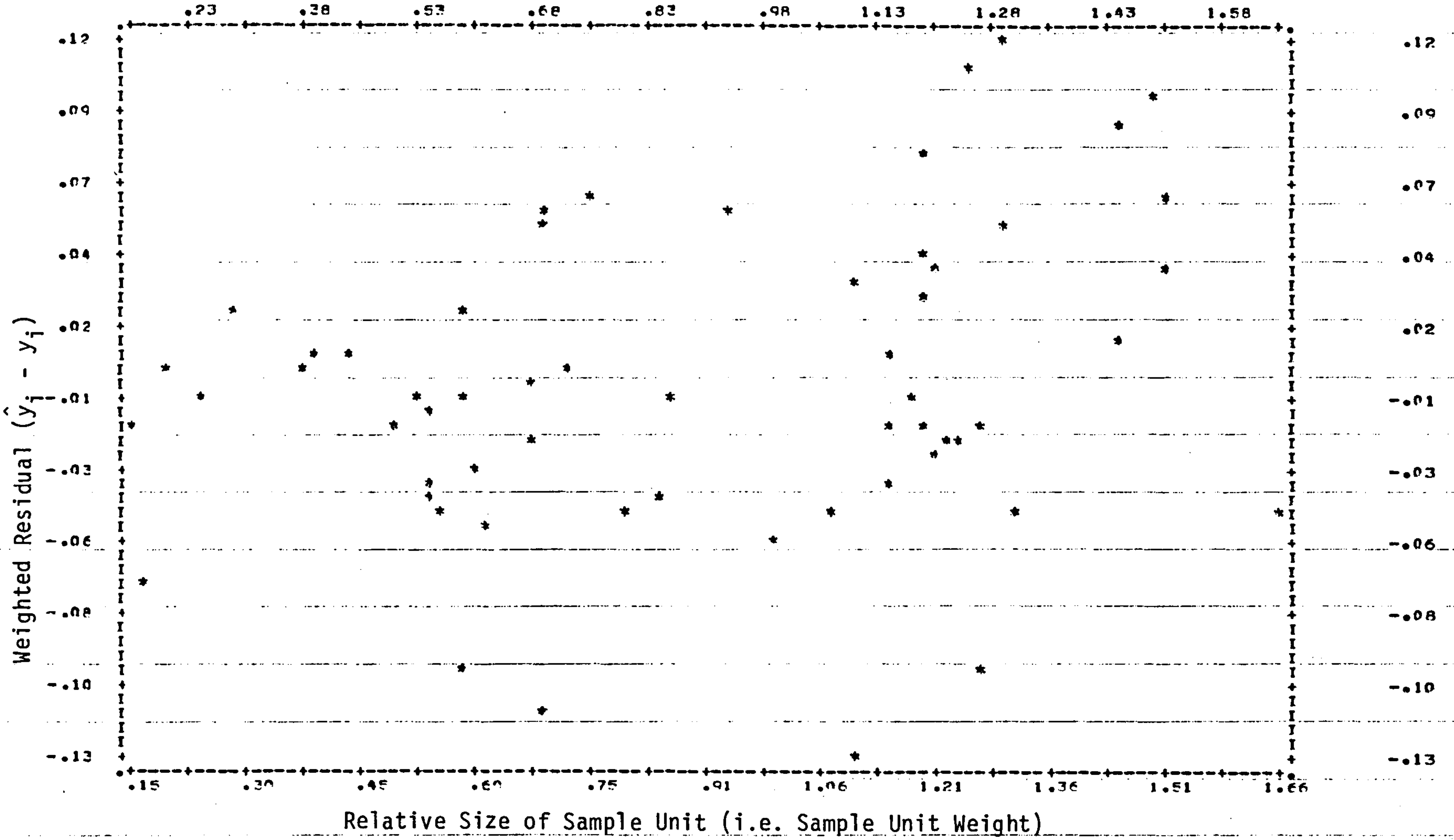
PLOTTED VALUES - 58 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-25. Unstratified Residuals versus Corresponding Relative Size of Sample Unit for the Colorado Desert Basin

FILE PLOTIT (CREATION DATE = 01/11/82)
SUBFILE CD

SCATTERGRAM OF (DOWN) RESIDWT
(ACROSS) WT2



UNSTRATIFIED PLOTS 01/11/82 16.45.31. PAGE 22

STATISTICS..

CORRELATION (R)-	.26218	R SQUARED -	.06874	SIGNIFICANCE R -	.02340
STD ERR OF EST -	.05093	INTERCEPT (A) -	-.03111	STD ERROR OF A -	.01670
SIGNIFICANCE A -	.03365	SLOPE (B) -	.03461	STD ERROR OF B -	.01702
SIGNIFICANCE B -	.02340				

PLOTTED VALUES - 58 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

(2) The regression estimator was superior in basins where the strata slopes differed significantly from one, or where high variability existed between slope values;

(3) Error rankings for the ratio estimators improved somewhat when the variance of Y increased or varied over the range of X - but not enough to give the ratio estimators variance superiority over the regression or difference estimators; and

(4) Stratification tended to reduce estimated error for all estimators when strata comprising a significant proportion of the sample frame had significantly different slopes or intercepts.

Comparison of irrigated proportion estimates by basin showed that in all cases except one, the estimators gave values differing by no more than one percent of the sample frame area. This indicates that relative bias in the estimators is small and that the final choice should be based on flexibility (favoring regression), or sampling efficiency, and on the theoretical justification of estimators of variance.

Comparison of the estimates of irrigated proportion to preliminary DWR figures for the same basins generally agreed with the error pattern ranking obtained from the standard errors and confidence interval half-widths. It was found that regression estimation was an even stronger performer than in the case of estimated error, but the need for verification of DWR figures prevented further analysis in this regard.

Comparison of Unweighted Estimators

Irrigated proportion and associated error estimates were also produced with unweighted sample unit observations. This was done in order to evaluate the relative performance of the four primary estimators (regression, both ratio, and difference) when the effect of sample unit measurement weighting was removed. Particular emphasis was placed on relating the estimated error to the form of the variation in Y over the range of X.

The method used to perform this comparison was similar to that employed in examining the weighted estimators. Table 3-35 presents the standard errors, 95 percent confidence interval half-widths, degrees of freedom, and estimates of proportion irrigated for the four estimators. Ranks for the standard errors within basins are shown in parenthesis. Plots of Y versus X, the residuals $(y_i - \hat{y}_i)$ versus X, and the residuals versus unweighted sample unit size were prepared and examined. Examples are shown in Figures 3-26 to 3-37.

Results of the Unweighted Estimator Comparison

Comparison of the unweighted estimates of error and irrigated proportion showed the following:

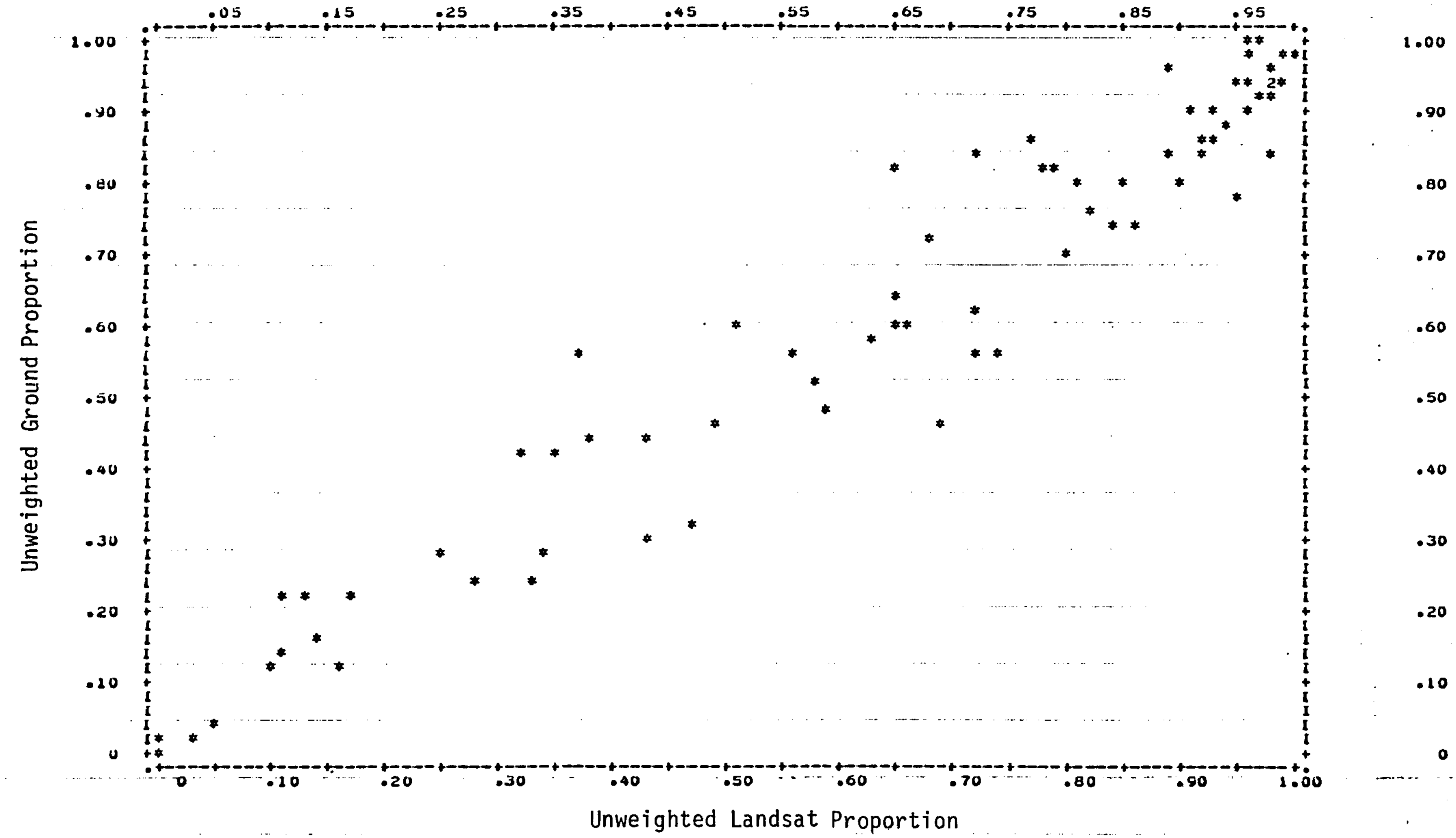
(1) The regression and difference estimators gave the best standard error rankings;

Table 3-35. Results for the various estimators using the unweighted-unstratified observations.

BASIN	REGRESSION (f5)	UNBIASED RATIO	BIASED RATIO	DIFFERENCE
		STANDARD ERRORS		
North Coast	(3) .01686	(4) .01982	(2) .01572	(1) .01560
San Francisco	(2) .01020	(4) .01385	(3) .01052	(1) .00990
Central Coast	(2) .00950	(4) .01225	(3) .01021	(1) .00940
South Coast	(2) .01161	(4) .01358	(3) .01251	(1) .01158
Colorado Desert	(1) .00859	(4) .00929	(3) .00885	(2) .00875
South Lahontan	(3) .01618	(4) .01643	(2) .01523	(1) .01505
North Lahontan	(1) .00931	(4) .00989	(3) .00956	(2) .00941
Sacramento	(1) .00867	(4) .00941	(2) .00889	(3) .00903
San Joaquin	(1) .01276	(4) .01374	(3) .01346	(2) .01333
Tulare	(4) .01191	(2) .01185	(3) .01188	(1) .01182
		95% CONFIDENCE INTERVAL HALF WIDTHS		
North Coast	.00392	.03984	.03160	.03137
San Francisco	.02044	.02776	.02108	.01983
Central Coast	.01891	.02438	.02033	.01871
South Coast	.02310	.02703	.02490	.02305
Colorado Desert	.01720	.01861	.01773	.01754
South Lahontan	.03301	.03346	.03101	.03066
North Lahontan	.01889	.02006	.01940	.01908
Sacramento	.01729	.01876	.01772	.01801
San Joaquin	.02542	.02737	.02681	.02655
Tulare	.02380	.02366	.02373	.02362
		DEGREES OF FREEDOM		
North Coast	47.00	48.00	48.00	48.00
San Francisco	54.00	55.00	55.00	55.00
Central Coast	77.00	78.00	78.00	78.00
South Coast	79.00	80.00	80.00	80.00
Colorado Desert	56.00	57.00	57.00	57.00
South Lahontan	31.00	32.00	32.00	32.00
North Lahontan	35.00	36.00	36.00	36.00
Sacramento	70.00	71.00	71.00	71.00
San Joaquin	74.00	75.00	75.00	75.00
Tulare	63.00	64.00	64.00	64.00
		IRRIGATED PROPORTION		
North Coast	.51175	.54402	.51857	.52115
San Francisco	.21197	.18073	.19938	.20129
Central Coast	.32533	.31519	.32067	.32386
South Coast	.44234	.43908	.44046	.44186
Colorado Desert	.80461	.81078	.80930	.80900
South Lahontan	.25733	.27821	.26546	.27149
North Lahontan	.58080	.57850	.57936	.57956
Sacramento	.66185	.66501	.66392	.66455
San Joaquin	.76043	.76876	.76788	.76747
Tulare	.81463	.81436	.81444	.81454

Figure 3-26. Unstratified Unweighted Ground Observations of Irrigated Proportion Versus Matching Unweighted Landsat Observations for the Sacramento Basin

SCATTERGRAM OF (DOWN) UNG
(ACROSS) UNL



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 60

STATISTICS..

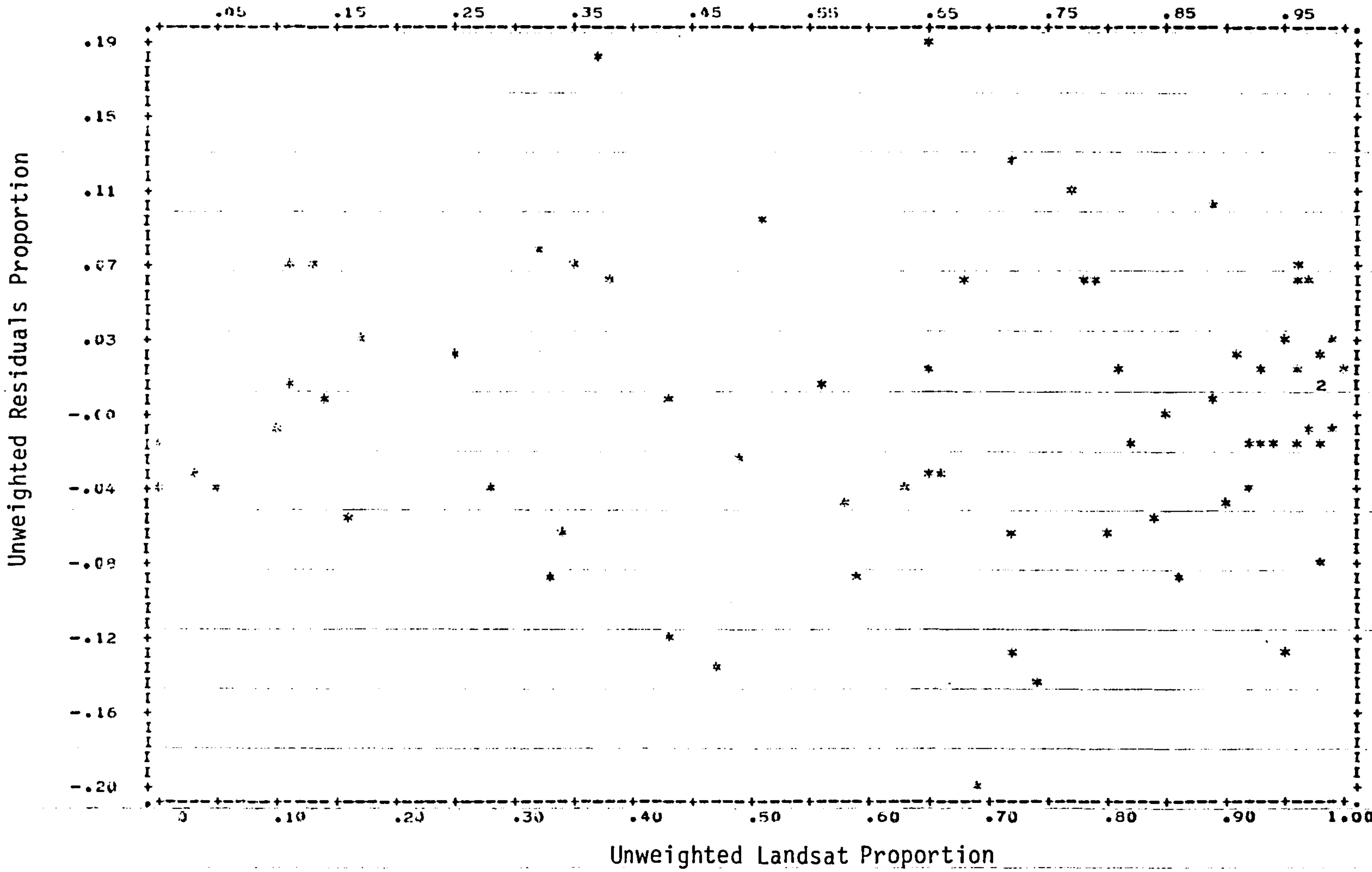
CORRELATION (R)-	.96871	R SQUARED -	.93839	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.07494	INTERCEPT (A) -	.03930	STD ERROR OF A -	.02015
SIGNIFICANCE A -	.02758	SLOPE (B) -	.92026	STD ERROR OF B -	.02818
SIGNIFICANCE B -	.00001				

PLOTTED VALUES - 72 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-27. Unstratified Unweighted Residuals Versus the Corresponding Unweighted Landsat Observations of Irrigated Proportion for the Sacramento Basin

SCATTERGRAM OF (DOWN) RESIDUN
(ACROSS) UNL



UNSTRATIFIED PLOTS 03/24/81 20.51.56. PAGE 8

STATISTICS..

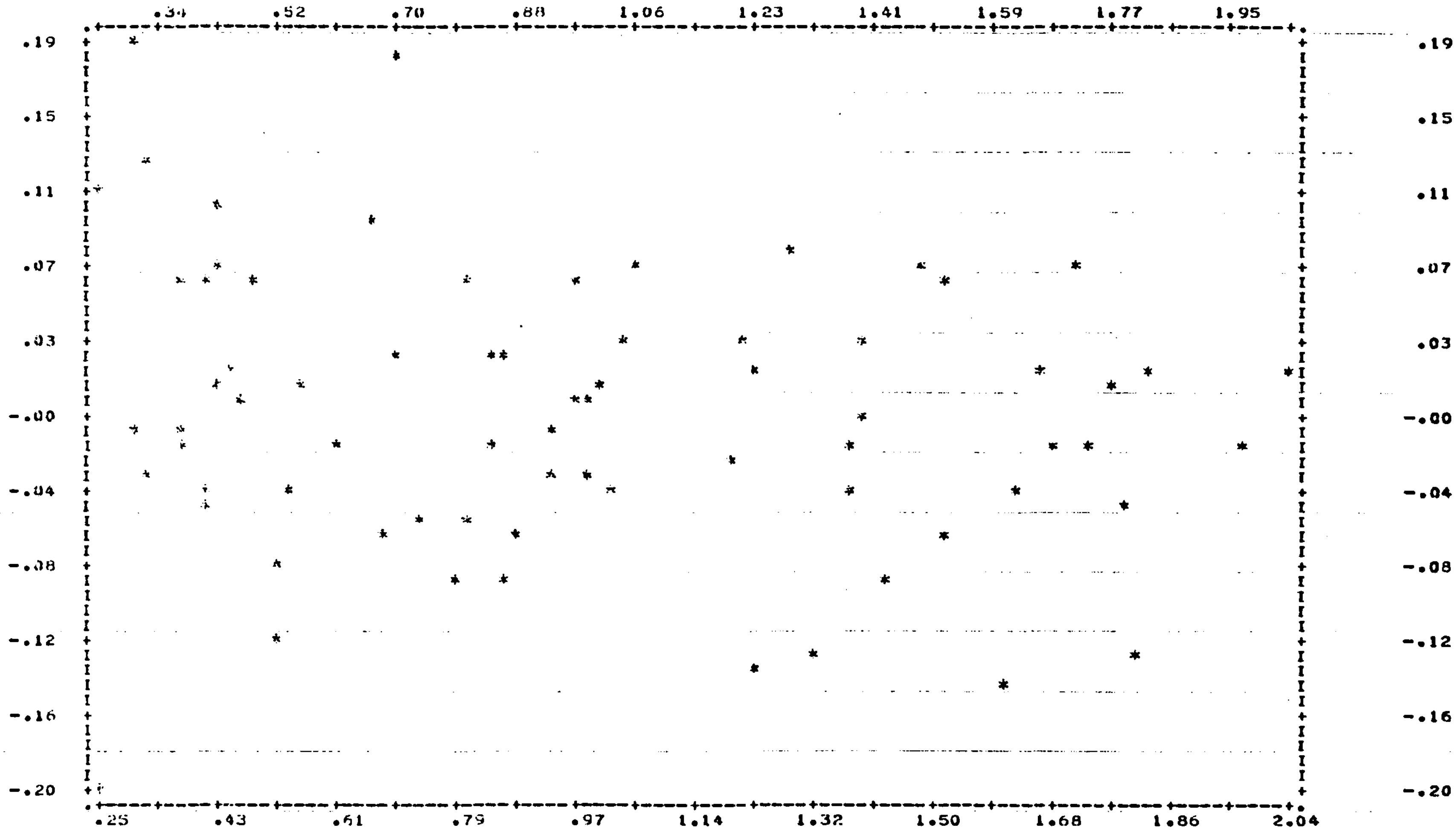
CORRELATION (R)-	-.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.07494	INTERCEPT (A) -	-.64836E-10	STD ERROR OF A -	.02015
SIGNIFICANCE A -	.50000	SLOPE (B) -	-.11309E-08	STD ERROR OF B -	.02818
SIGNIFICANCE B -	.50000				

PLOTTED VALUES - 72 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-28. Unstratified Unweighted Residuals Versus Corresponding Relative Size of Sample Unit for the Sacramento Basin

SCATTERGRAM OF (DOWN) RESIDUN
(ACROSS) WT2



Relative Size of Sample Unit (i.e. Sample Unit Weight)

UNSTRATIFIED PLOTS 03/24/81 20.51.56. PAGE 10

STATISTICS..

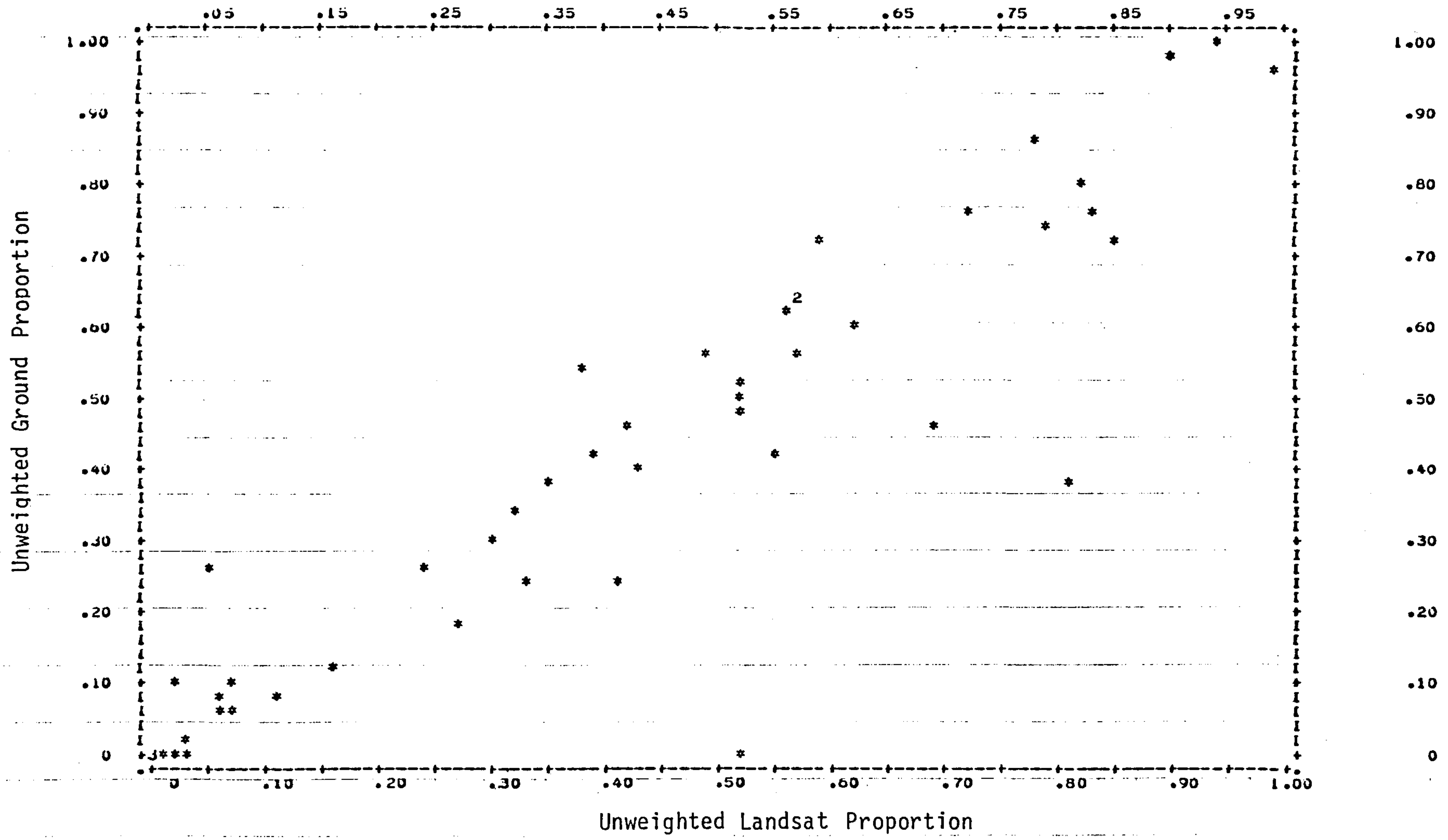
CORRELATION (R)-	-.17402	R SQUARED -	.03028	SIGNIFICANCE R -	.07187
STD ERR OF EST -	.07390	INTERCEPT (A) -	.02501	STD ERROR OF A -	.01902
SIGNIFICANCE A -	.09642	SLOPE (B) -	-.02579	STD ERROR OF B -	.01744
SIGNIFICANCE B -	.07187				

PLOTTED VALUES - 72 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED

Figure 3-29. Unstratified Unweighted Ground Observations of Irrigated Proportion Versus Matching Unweighted Landsat Observations for the North Coast Basin

SCATTERGRAM OF (DOWN) UNG
(ACROSS) UNL



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 4

STATISTICS..

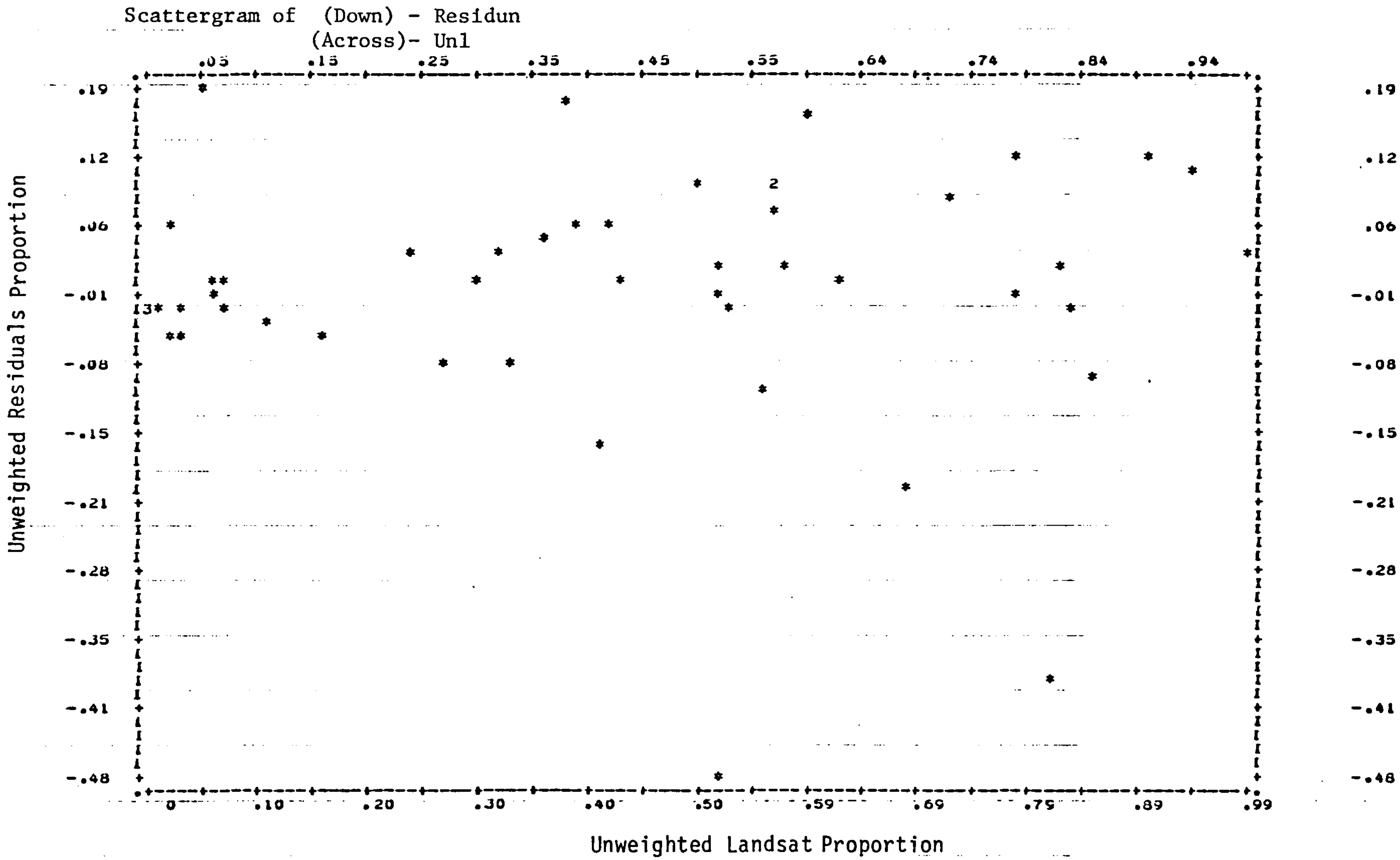
CORRELATION (R)-	.91895	R SQUARED -	.84447	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.12034	INTERCEPT (A) -	.02407	STD ERROR OF A -	.02935
SIGNIFICANCE A -	.20817	SLOPE (B) -	.91977	STD ERROR OF B -	.05758
SIGNIFICANCE B -	.00001				

PLOTTED VALUES - 49 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-30. Unstratified Unweighted Residuals Versus the Corresponding Unweighted Landsat Observations of Irrigated Proportion for the North Coast Basin

1.



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 8

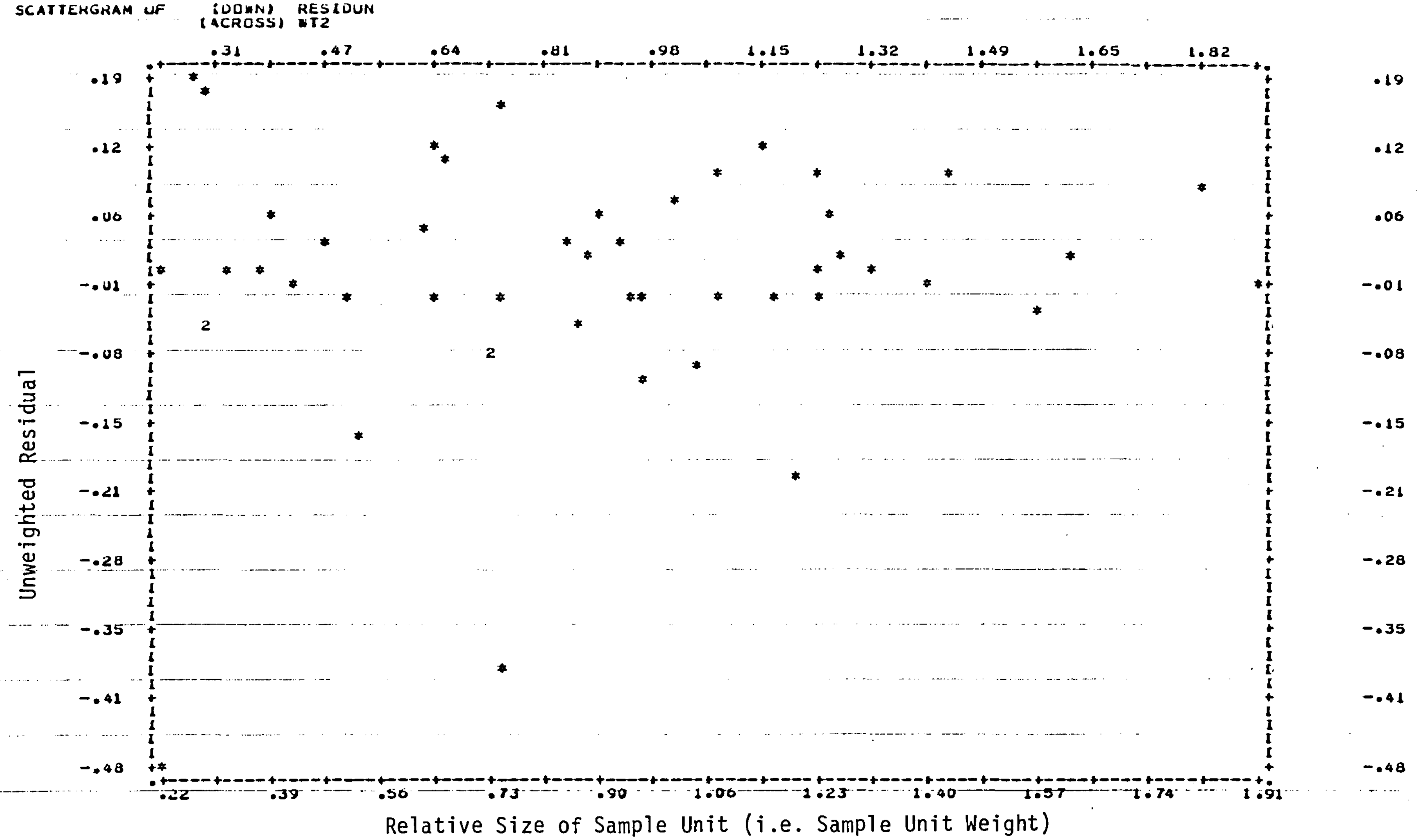
STATISTICS..

CORRELATION (R)-	-.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.12034	INTERCEPT (A) -	.60893E-09	STD ERROR OF A -	.02935
SIGNIFICANCE A -	.50000	SLOPE (B) -	-.34009E-08	STD ERROR OF B -	.05758
SIGNIFICANCE B -	.50000				

PLOTTED VALUES - 49 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-31. Unstratified Unweighted Residuals Versus Corresponding Relative Size of Sample Unit for the North Coast Basin



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 10

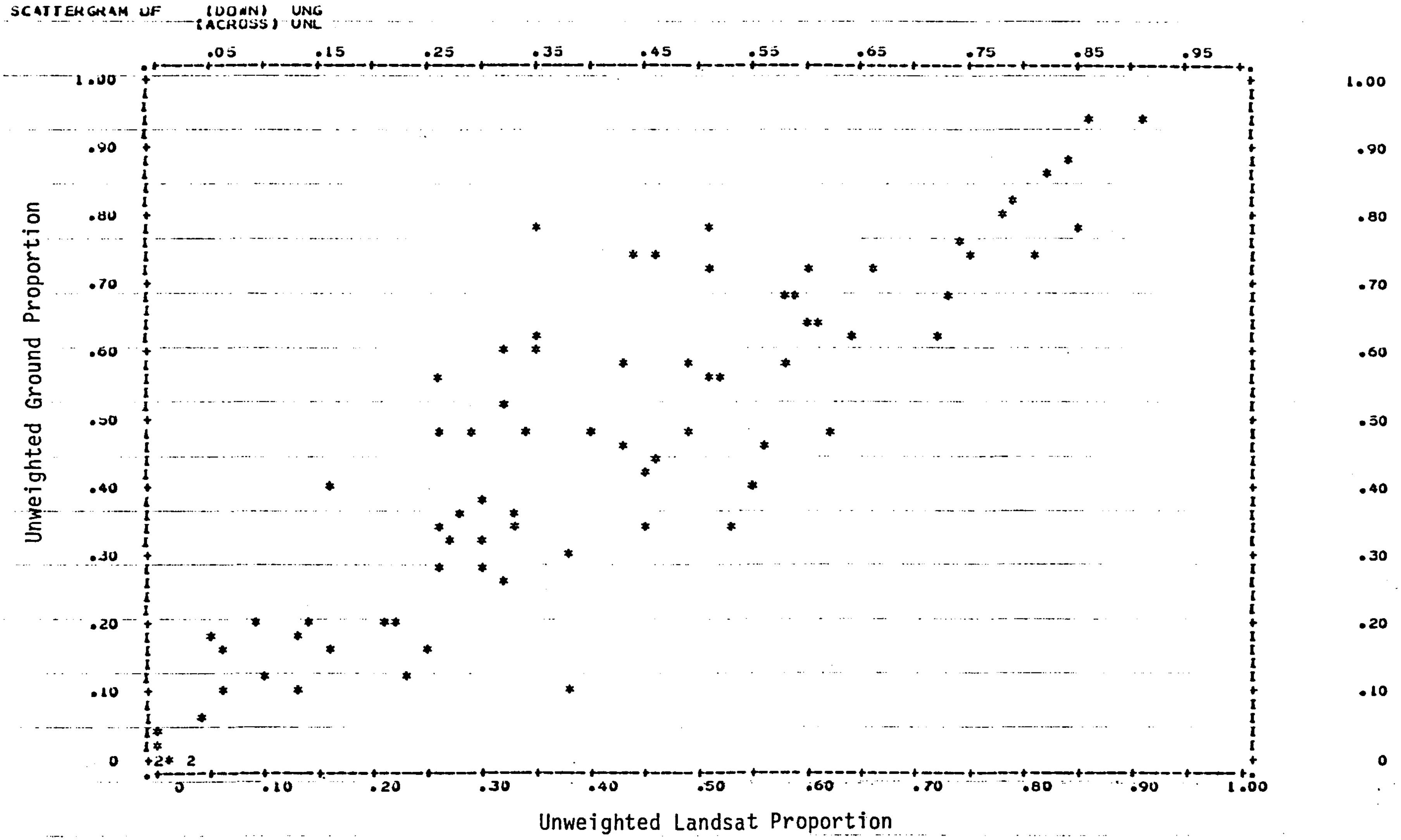
STATISTICS..

CORRELATION (R)-	.13435	R SQUARED -	.01805	SIGNIFICANCE R -	.17869
STD ERR OF EST -	.11925	INTERCEPT (A) -	-.03278	STD ERROR OF A -	.03917
SIGNIFICANCE A -	.20342	SLOPE (B) -	.03722	STD ERROR OF B -	.04004
SIGNIFICANCE B -	.17869				

PLOTTED VALUES - 49 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-32. Unstratified Unweighted Ground Observations of Irrigated Proportion Versus Matching Unweighted Landsat Observations for the South Coast Basin



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 28

STATISTICS..

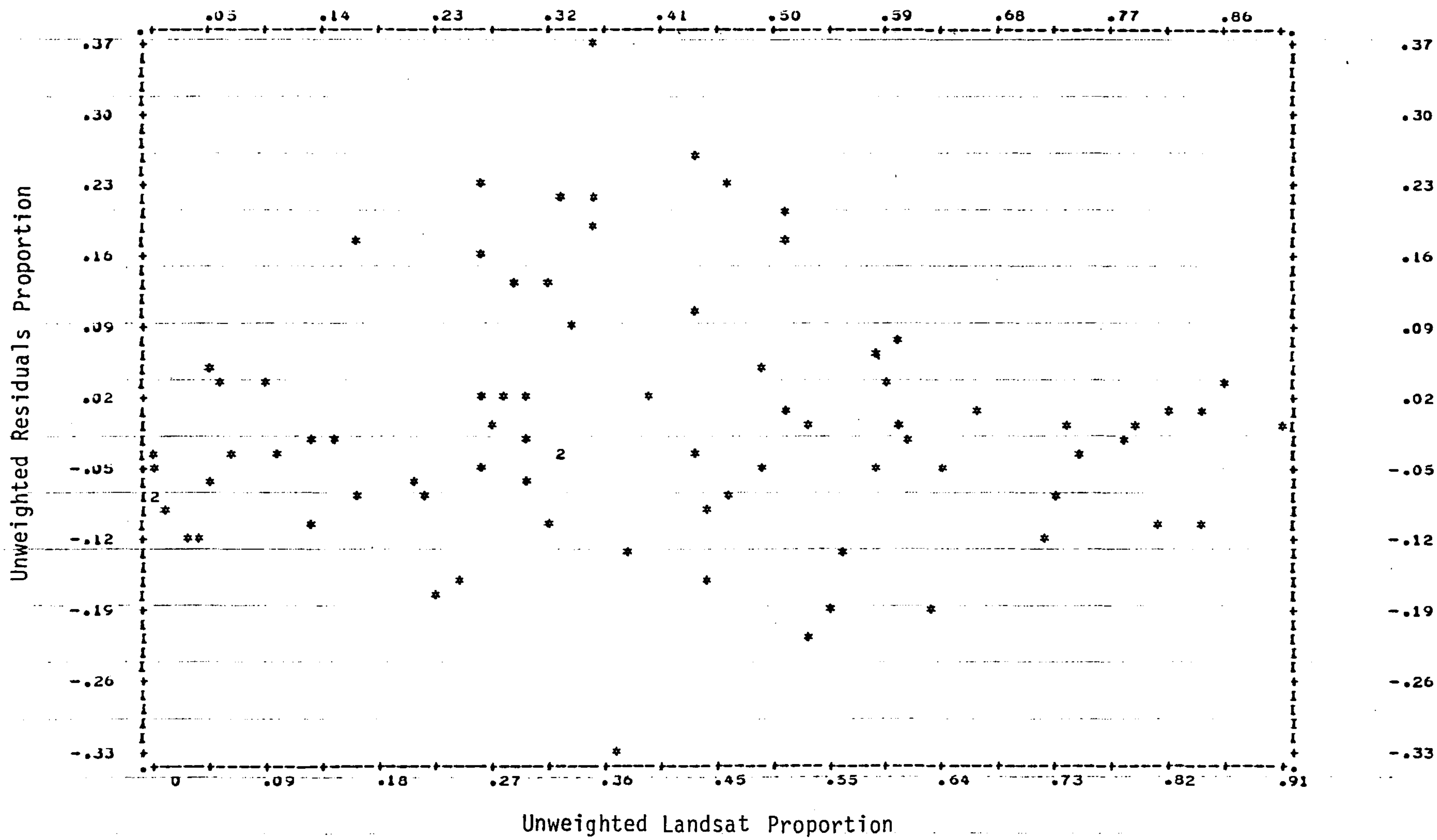
CORRELATION (R)-	.88717	R SQUARED -	.78708	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.12245	INTERCEPT (A) -	.07738	STD ERROR OF A -	.02577
SIGNIFICANCE A -	.00179	SLOPE (B) -	.94938	STD ERROR OF B -	.05556
SIGNIFICANCE B -	.00001				

PLOTTED VALUES - 81 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-33. Unstratified Unweighted Residuals Versus the Corresponding Unweighted Landsat Observations of Irrigated Proportion for the South Coast Basin

SCATTERGRAM OF (DOWN) RESIDUN
(ACROSS) UNL



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 32

STATISTICS..

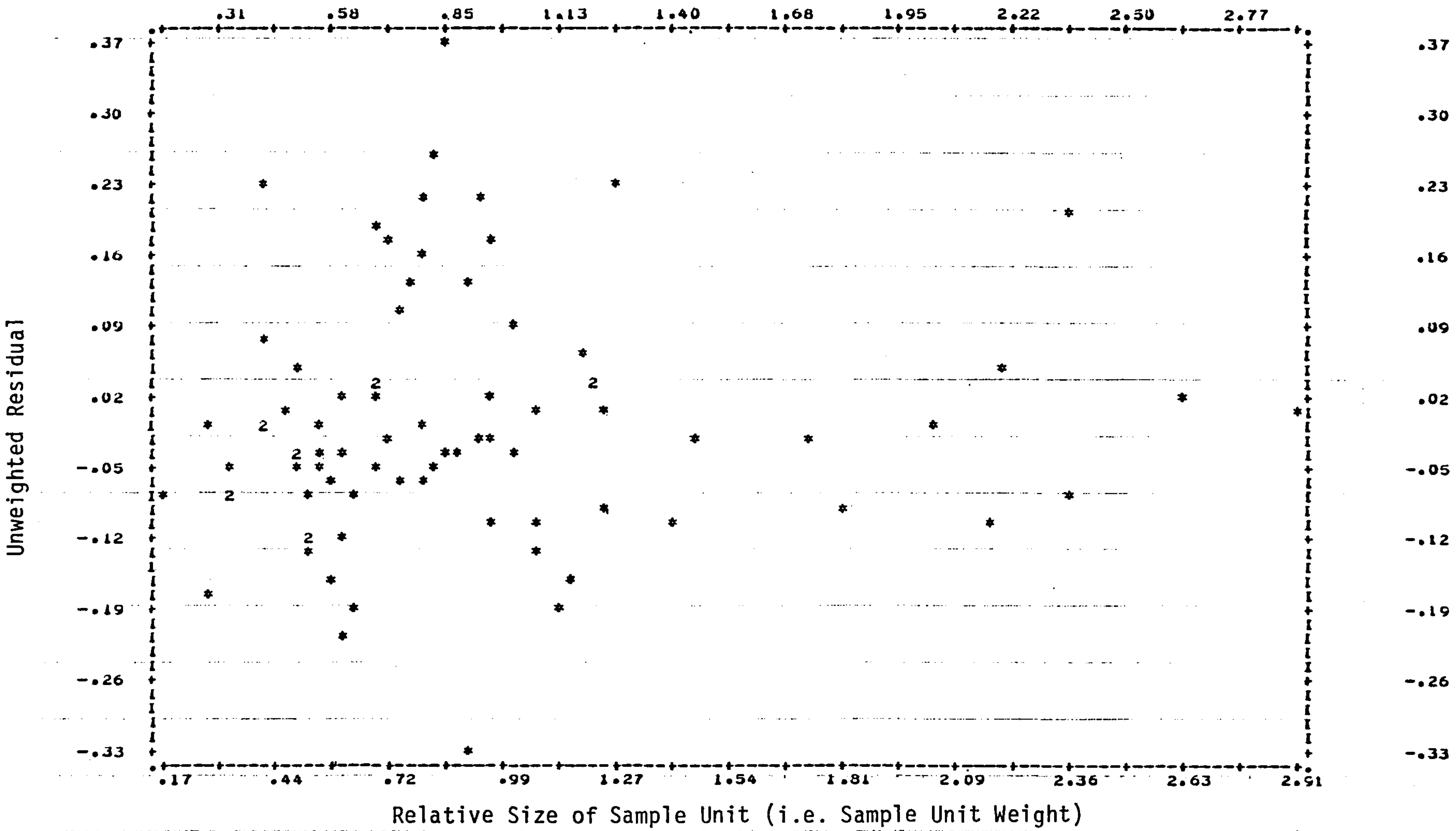
CORRELATION (R)-	.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.12245	INTERCEPT (A) -	-.75803E-09	STD ERROR OF A -	.02577
SIGNIFICANCE A -	.50000	SLOPE (B) -	.55587E-08	STD ERROR OF B -	.05556
SIGNIFICANCE B -	.50000				

PLOTTED VALUES - 81 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-34. Unstratified Unweighted Residuals Versus Corresponding Relative Size of Sample Unit for the South Coast Basin

SCATTERGRAM OF (DOWN) RESIDUN
(ACROSS) WT2



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 34

STATISTICS..

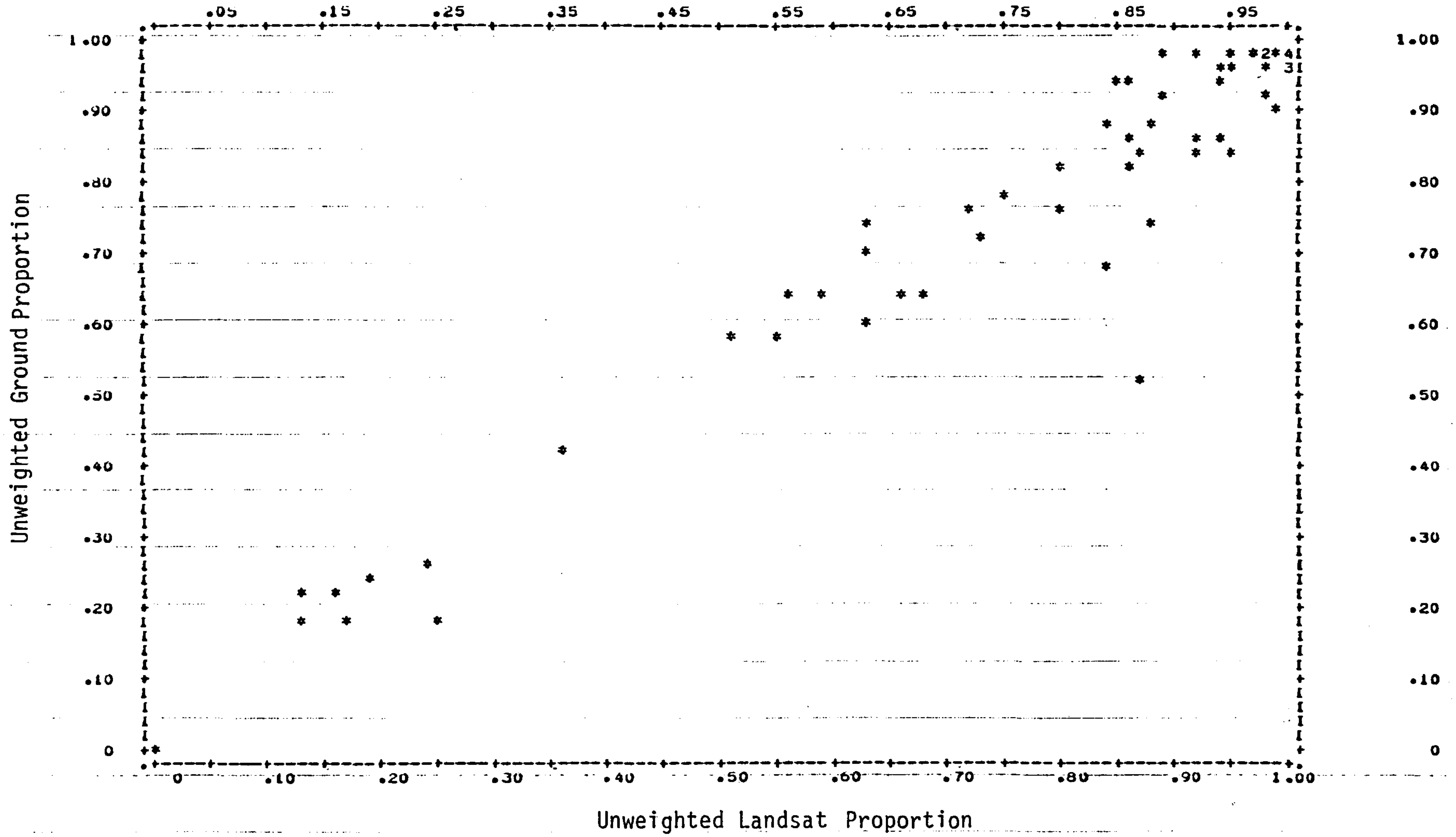
CORRELATION (R)-	.08398	R SQUARED -	.00705	SIGNIFICANCE R -	.22801
STD ERR OF EST -	.12201	INTERCEPT (A) -	-.01691	STD ERROR OF A -	.02633
SIGNIFICANCE A -	.26131	SLOPE (B) -	.01834	STD ERROR OF B -	.02448
SIGNIFICANCE B -	.22801				

PLOTTED VALUES - 81 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-35. Unstratified Unweighted Ground Observations of Irrigated Proportion Versus Matching Unweighted Landsat Observations for the Colorado Desert Basin

SCATTERGRAM OF (DOWN) UNG
(ACROSS) UNL



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 36

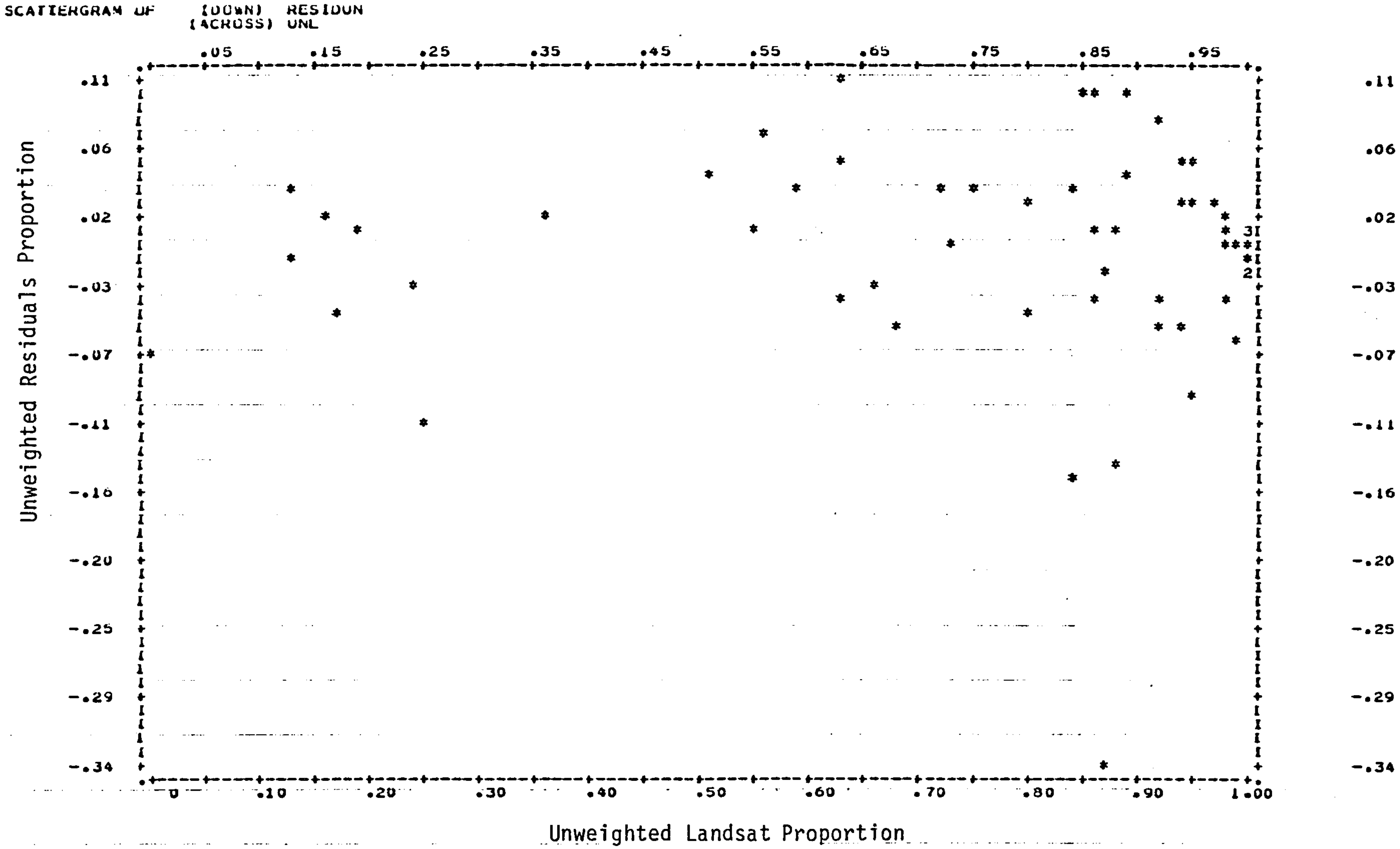
STATISTICS..

CORRELATION (R)-	.96493	R SQUARED -	.93108	SIGNIFICANCE R -	.00001
STD ERR OF EST -	.07082	INTERCEPT (A) -	.06546	STD ERROR OF A -	.02675
SIGNIFICANCE A -	.00877	SLOPE (B) -	.91841	STD ERROR OF B -	.03339
SIGNIFICANCE B -	.00001				

PLOTTED VALUES - 58 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-36. Unstratified Unweighted Residuals Versus the Corresponding Unweighted Landsat Observations of Irrigated Proportion for the Colorado Desert Basin



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 40

STATISTICS..

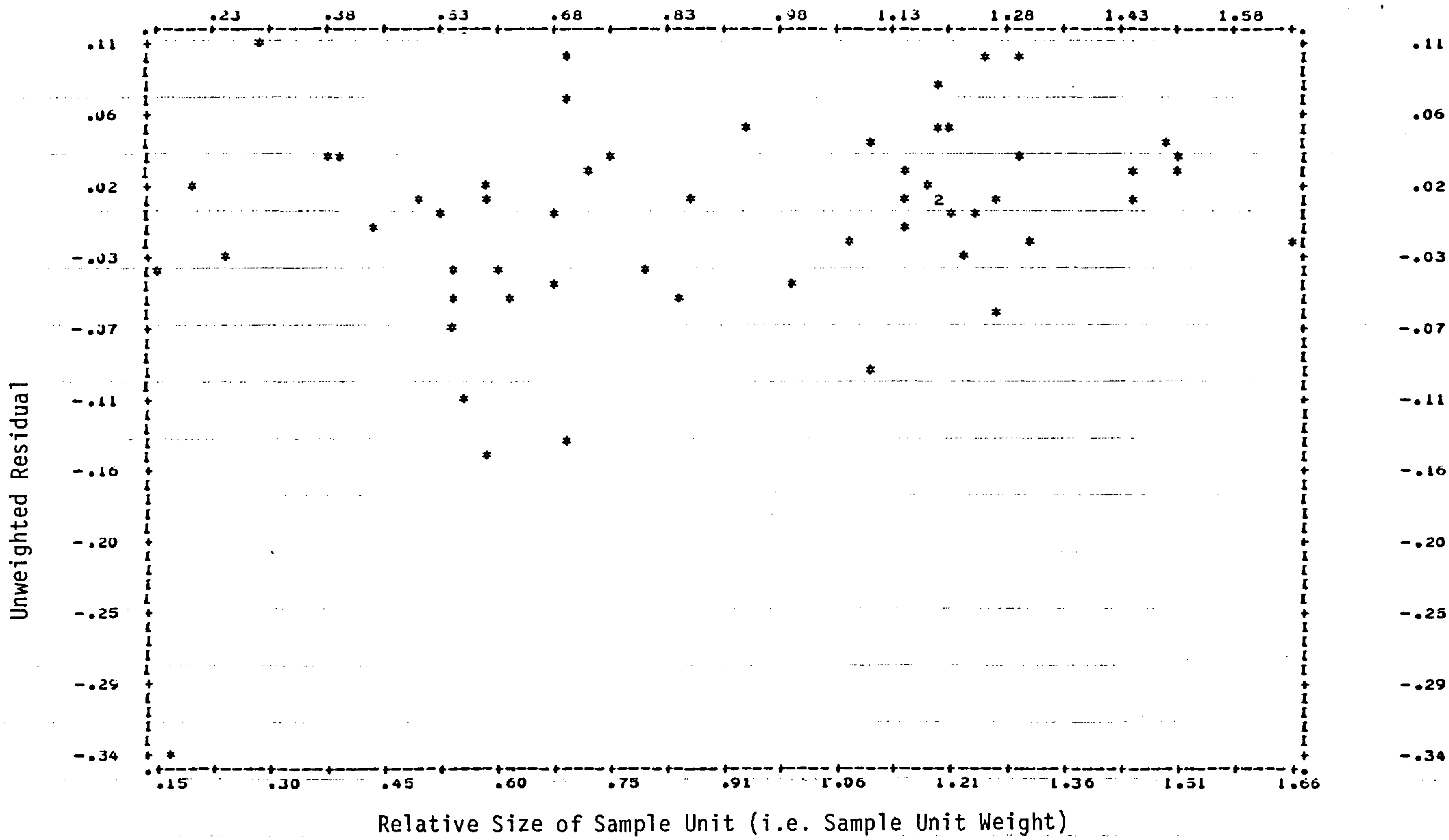
CORRELATION (R)-	-.00000	R SQUARED -	.00000	SIGNIFICANCE R -	.50000
STD ERR OF EST -	.07082	INTERCEPT (A) -	.39178E-08	STD ERROR OF A -	.02675
SIGNIFICANCE A -	.50000	SLOPE (B) -	-.38161E-08	STD ERROR OF B -	.03339
SIGNIFICANCE B -	.50000				

PLOTTED VALUES - 58 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Figure 3-37. Unstratified Unweighted Residuals Versus Corresponding Relative Size of Sample Unit for the Colorado Desert Basin

SCATTERGRAM OF (DOWN) RESIDUN
(ACROSS) WT2



UNSTRATIFIED PLOTS 03/24/81 19.41.33. PAGE 42

STATISTICS..

CORRELATION (R)-	.32090	R SQUARED -	.10298	SIGNIFICANCE R -	.00702
STD ERR OF EST -	.06707	INTERCEPT (A) -	-.05111	STD ERROR OF A -	.02200
SIGNIFICANCE A -	.01191	SLOPE (B) -	.05685	STD ERROR OF B -	.02242
SIGNIFICANCE B -	.00702				

PLOTTED VALUES - 58 EXCLUDED VALUES - 0 MISSING VALUES - 0

******** IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

(2) Ratio estimates (biased and unbiased) of standard error still ranked decidedly below the corresponding estimates produced by the regression and difference estimators. Had assumptions on the form of heteroscedasticity (variation in Y over the range of X) been met for ratio estimation this result might at first appear unexpected; if we recall that a further assumption for (biased) ratio estimation is that the relationship is linear through the origin, the result can be visualized in terms of the relationship shown in Figure 3-38a. However, the observed heteroscedasticity did not take the form of 3-38a but rather that shown in Figures 3-38b, 3-38c, and 3-38d. A well known fact in general linear regression states that if the error model is mis-specified the variance estimators will be inefficient. This is consistent with the observed poor rankings of the biased ratio estimate. The argument also holds for the unbiased ratio estimate.

(3) Differences between the regression estimate of irrigated proportion and the corresponding unbiased ratio estimate varied by approximately three percent of the sample frame in the North Coast and San Francisco hydrologic basins; otherwise, differences between estimators were one percent of the sample frame or less.

Comparison of Weighted Versus Unweighted Estimators

The relative performance of linear estimators using sample unit size-weighted versus unweighted observations was evaluated. This evaluation included a comparison of the relative size of estimated standard errors between corresponding estimators in each basin. In addition, weighted and unweighted plots of Y versus X, and residual versus X were examined to assess (1) the relative variation of Y about the regression line and (2) the pattern of this variation over the range of X. Finally, differences between corresponding weighted and unweighted estimates of irrigated proportion were determined.

Results of the Weighted Versus Unweighted Estimator Comparison

A number of trends were evident:

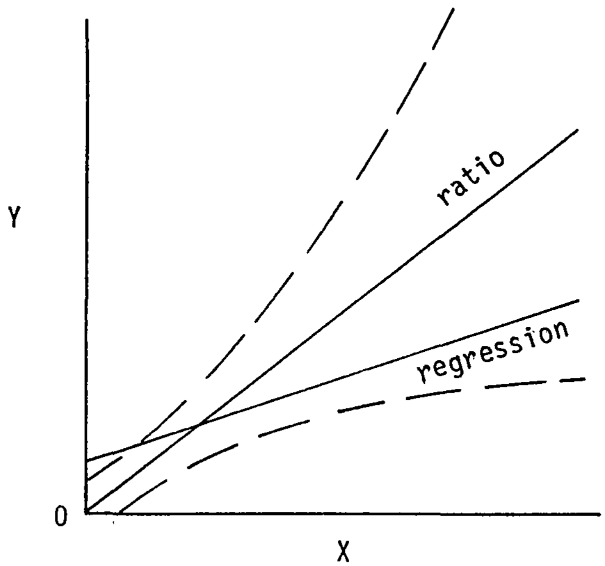
(1) Sample unit weighting showed less variability about the regression line of Y on X; this was reflected visually in the plots and in the higher r^2 (squared correlation between X and Y) given by weighting versus no weighting;

(2) However, larger variance of Y caused by weighting compensated for the increase in r^2 in some cases; thus the net effect was that the smallest standard errors were split between weighted and unweighted approaches on an estimator by estimator basis;

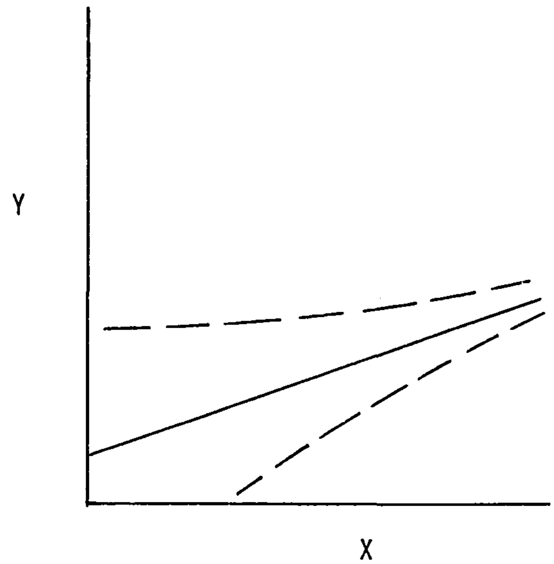
(3) A pattern of larger residuals (unweighted estimator) in small to medium-sized sample units in some basins (e.g. the South Coast and the San Joaquin) tended to support the rationale for a sample unit weighted approach; recall that weighting measurement observations by sample unit size was intended to decrease the relative weight given to smaller sample units - these smaller sample units potentially subject to proportionally greater error due to their small size; and

Figure 3-38. Distributions of Y Versus X Shown Schematically as Dotted Line Error Bounds

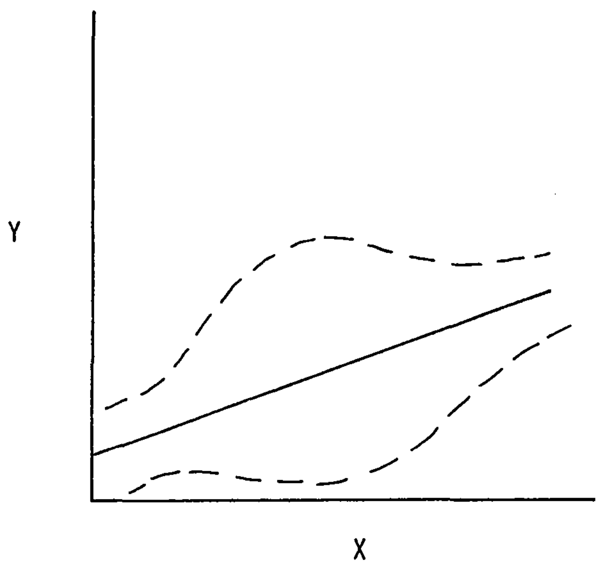
A: Theoretical Distribution Favorable to Ratio Estimation



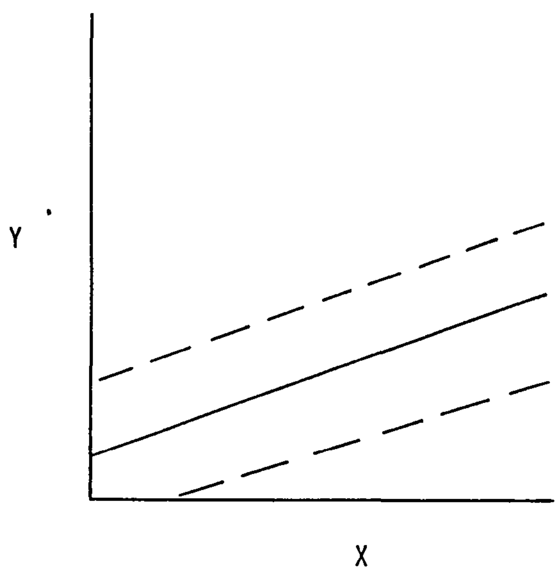
B: Example of a Distribution Type Actually Observed in 1979



C: Example of a Distribution Type Actually Observed in 1979



D: Example of a Distribution Type Actually Observed in 1979



(4) Unweighted estimates of basin irrigated proportion were lower than their weighted counterparts in seven out of ten basins; the three exceptions were the Central Coast, Sacramento, and San Joaquin hydrologic basins; this pattern meant that the weighted estimators would have been closer to preliminary DWR estimates in eight out of ten basins.

Comparison of Estimators: Conclusions

The following conclusions were drawn after reviewing the results presented above and after considering the measurement context in which the 1979 inventory was performed:

(1) The regression and difference estimators produced the smallest estimated errors on the average in all cases evaluated; given the ground sample drawn in the 1979 inventory, these estimators appear to be superior to both the ratio estimators and to the combination regression/ratio estimators;

(2) The difference estimator is simple and often gave somewhat lower estimated standard errors or confidence interval half-widths than regression using the 1979 data; we, however, recommend the regression estimator over the difference estimator because regression should be more robust against inventory-specific changes in intercept or slope caused by (a) the Landsat dates of imagery available, (b) difference image analyst expertise, and (c) changes in ground irrigation practice:

(3) Weighting sample unit observations by size of sample unit reduced estimated standard error in five out of ten basins using the variance formula for regression with factor 5;

(4) Given the present sample size distribution, we would recommend the use of sample size weighting of sample observations to guard against registration, digitizing, or analyst error in smaller sample units; if the nominal size of sample unit were made smaller and the range of size variation was limited to $\pm$ ½ mile^2, then the weighting system employed in the 1979 inventory might not be necessary; and

(5) A better transformation than sample size weighting may be available - e.g. the logit transformation, a weighting scheme appropriate to proportion data regardless of the size of sample unit.

3.7.3 Evaluation of the Impact of Sample Unit Area on Sample Size (n) and Cost

An important factor affecting the ground sample size required to meet given inventory error goals is the size of an individual sample unit. Generally, the larger the size of sample unit, the lower the variation between sample units and therefore the lower the estimated sampling error. Lower variation between sample units with increased size is the result of averaging differences in local irrigated proportion over greater area. This gain in sampling precision with increased size of sample unit is, however, offset by increased measurement cost. Thus one important inventory design goal was to determine the size of sample unit that enables achievement of basin error goals at minimum cost.

Area of Study

Evaluation of the sample unit size problem required an area typical of a large portion of irrigated land in the state. It also required an area for which between-sample unit variance and costs associated with a number of different sample unit sizes could be obtained easily. Such an area was the floor and surrounding terrace lands of the Sacramento Valley. Both University and project-related DWR personnel had significant familiarity and experience with this region. In addition, a Landsat digital irrigation class map was available for a one degree block* covering the north-central portion of the Sacramento basin. This digital map could be easily accessed by the UCB Survey Planning Model to produce estimates of irrigated proportion variance for varying sizes of sample units.

Landsat Data Used

The Landsat irrigated class map was created using the Task II technique (see section 4.0 and Wall et al 1980, page 115) of band 7 to band 5 ratioed data. Eight irrigation classes were recognized according to whether a given pixel was above or below an 'irrigation' 7/5 ratio threshold on a given date. In order to create a digital class map reflecting as closely as possible the manually derived irrigated class map, all digital classes having at least one irrigated date were grouped into a single class. Thus the digital Landsat class map became a map of irrigated versus non-irrigated areas. For purposes of this analysis, a one-to-one relationship was assumed between this digital map and the map produced by the manual Landsat interpretation technique in the 1979 inventory.

Costs Considered

Only ground sample unit costs were considered in this analysis. That is, only ground costs were considered to vary with sample unit size. Costs associated with Landsat sample units, on the other hand, were not assumed to vary for a given measurement procedure, since the entire population of those units was to be measured at each inventory.**

Ground sample unit costs were broken down into personnel and other costs. Personnel costs per sample unit resulted from (1) office preparation time (OP), time mapping the sample unit on the ground (M), time traveling to and between sample units (TR), and time to tabulate the ground data (TAB). Resulting costs were then determined from the equations

---

* 64 seven and one half-minute quadrangles (see Figure 3-39)

** Though "handling" costs associated with Landsat units having matching ground data might vary somewhat with size. Sample frame construction costs were not evaluated here, as these were considered long-term reoccurring costs.

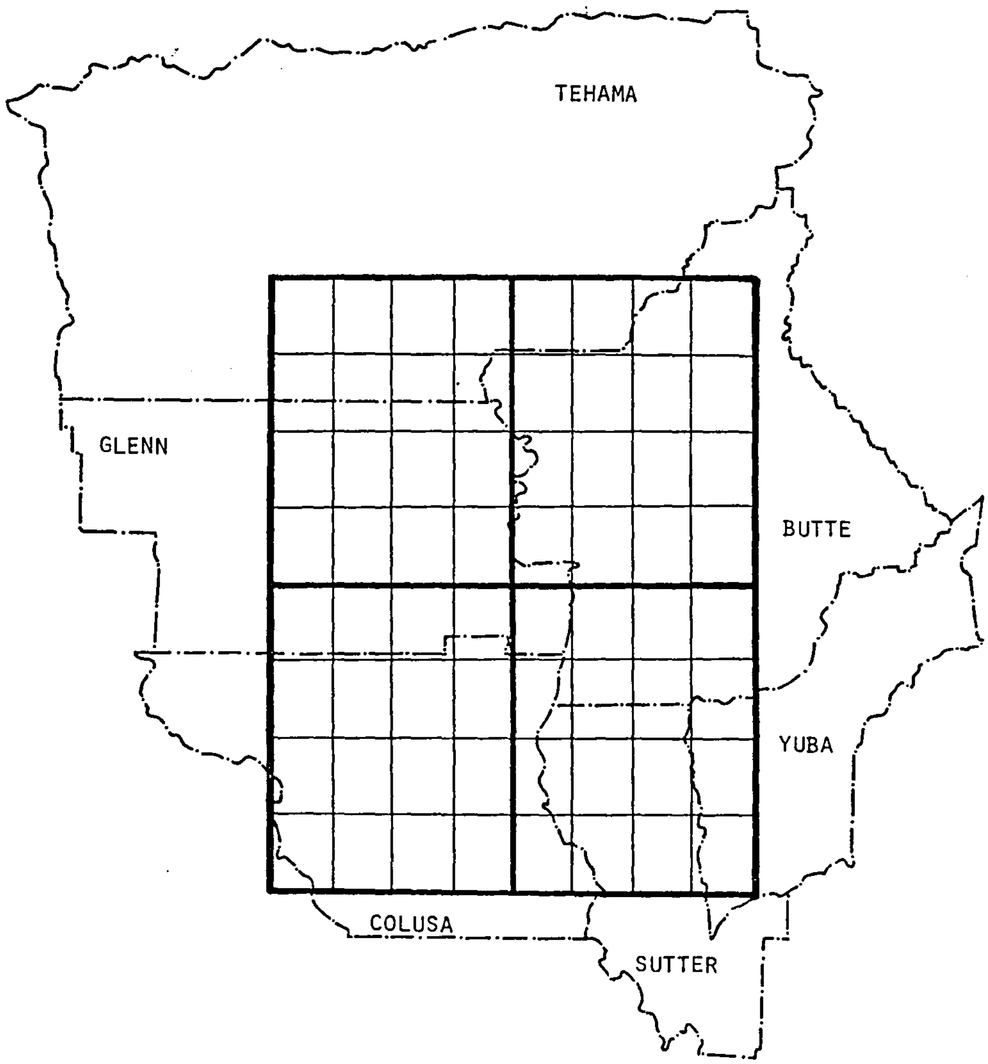


Figure 3-39. Location of the One Degree Block in the Sacramento Valley

$$P_1 = \begin{array}{l}\text{Personnel}\\ \text{Cost for}\\ \text{Measurement}\end{array} = (\text{OP times rate/hr}) + (\text{M times rate/hr}) + (\text{TAB times rate/hr})$$

and

$$P_2 = \begin{array}{l}\text{Personnel}\\ \text{Cost for}\\ \text{Travel}\end{array} = \text{TR times rate/hr.}$$

Average times for each cost component were provided by DWR personnel (Ferchaud 1981). These figures, summarized in Table 3-36, were reconstructed from records kept by individual DWR districts in 1979. For comparison, revised times for the Sacramento Valley and San Joaquin Valley floors were provided by Ferchaud (1981) (Tables 3-37 and 3-38) to show the improvement expected in an operational system. The times reported in Table 3-36 were then multiplied by dollar rates per hour that varied between DWR districts.

Other costs considered included (1) cost of aerial photography for each sample unit (AP), (2) computer tabulation cost per sample unit (CT), (3) per diem per sample unit (PD), and (4) car cost per sample unit (CC). Thus

$$C_1 = \begin{array}{l}\text{Total Ground}\\ \text{Measurement}\\ \text{Cost per SU}\end{array} = P_1 + AP + CT$$

and

$$C_2 = \begin{array}{l}\text{Total Ground}\\ \text{Travel Cost}\\ \text{Per SU}\end{array} = P_2 + PD + CC \quad .$$

The primary time/cost assumptions (set A) used in the sample unit size analysis were defined as follows:

(1) DWR time and cost data reported for the Sacramento basin in 1979 were adjusted to valley floor conditions by reference to corresponding 1979 San Joaquin and Tulare basin data.

(2) A further cost adjustment was made to make cost and time figures specific to the average size of a valley floor unit. The average stratum 4 value of 3.47 mi^2 was taken to represent this size.

(3) The resulting C_1 was based on the per sample unit assumption of 1.25 hours of office preparation time, 1.00 hours field mapping time, and 2.00 hours tabulation time. In addition, a two dollar computer tabulation cost was included as was a per sample unit charge of $22.50 for color 35 mm transparency aerial photography. This aircraft cost was based on the assumption that each frame would be obtained at a scale of 1:62,500 and would cost $2.50 when delivered to DWR.

(4) The resulting C_2 was based on the assumption that travel time per sample unit would be 0.75 hours. A $9.50 per diem and car mileage charge was then added to the resulting personnel travel cost.

Table 3-36. AVERAGE PERSON HOURS PER SAMPLE UNIT FOR DATA COLLECTION AND TABULATION IN THE 1979 STATEWIDE IRRIGATED LANDS INVENTORY

BASIN	NUMBER OF SAMPLE UNITS	AVERAGE TIME PER SAMPLE UNIT FOR DATA COLLECTION (HOURS)			TOTAL AVERAGE COLLECTION TIME (HOURS)	AVERAGE TABULATION TIME PER SU (HOURS)	TOTAL AVERAGE COLLECTION AND TABULATION TIME PER SU (HOURS)
		OFFICE	TRAVEL	FIELD			
NORTH COAST	51	1.57	1.81	2.20	5.57	1.88	7.45
SAN FRANCISCO	56	1.36	1.20	2.95	5.51	1.88	7.39
CENTRAL COAST	79	1.69	1.67	2.88	6.25	1.88	8.13
SOUTH COAST	84	2.60	1.31	2.50	6.41	1.88	8.29
COLORADO DESERT	58	1.74	1.55	2.17	5.46	1.88	7.34
SOUTH LAHONDAN	34	1.38	1.12	.97	3.47	1.88	5.35
NORTH LAHONDAN	37	2.05	2.60	2.33	6.98	1.88	8.86
SACRAMENTO	72	.82	2.01	1.43	4.26	1.88	6.14
SAN JOAQUIN	81	1.29	.84	.77	2.89	1.88	4.77
TULARE	65	1.58	.49	.54	2.61	1.88	4.49
TOTAL SU's	617						
WEIGHTED AVERAGE		1.62	1.41	1.88	4.91	1.88	6.79

Table 3-37. Average Times Per Sample Unit Expected in a Future Operational Inventory Based on a 7 Hour Day

Basin	Number of Sample Units	Average Time Per Sample Unit for Data Collection (Hours)			Total Average Collection Time (Hours)	Average Tabulation Time Per SU (Hours)	Total Average Collection and Tabulation Time (Hours)
		Office	Travel	Field			
Sacramento Valley Floor							
160 Acre SU	59	.32	1.05	.29	1.66	.59	2.25
1 Square Mile SU	59	.48	1.05	.48	2.01	.69	2.70
5 Square Mile SU	59	.61	1.45	1.61	3.67	.97	4.64

Table 3-38. Average Times Per Sample Unit Expected in a Future Operational Inventory Based on a Seven Hour Day

Basin & Subarea	Number of Sample Units	Average Time Per Sample Unit for Data Collection (Hours) Office	Travel	Field	Total Average Collection Time (Hours)	Average Tabulation Time Per SU (Hours)	Total Average Collection and Tabulation Time Per SU (Hours)
Tulare Valley Floor							
160 Acre SU	65	.32	.75	.25	1.32	.59	1.91
1 Square Mile SU	65	.48	.75	1.38	1.61	.65	2.26
5 Square Mile SU	65	.61	.95	1.30	2.86	.88	3.74

For comparison, a second cost set (set B) was defined. This set used the actual times and costs reported for the 72 ground sample units obtained in the Sacramento basin in 1979. No cost for aerial photography was included as previously obtained photography was used where required. Actual costs of sample unit photography in an operational inventory and its amortization remains an issue for further analysis.

In order to obtain a C_1 cost for each size of sample unit, a cost versus size curve was constructed in the following manner. The C_1 cost under cost set A was defined to equal a <u>relative cost</u> of 1.0 at an average sample unit size ($\bar{s}$) of 3.47 square miles. Next, a straight line was drawn between the point just defined and the origin. The relationship between relative cost and sample unit size was assumed, based on a review of factors affecting cost, to lie along this line above an s of 2 mi^2. Below 2 mi^2, the rate of decrease in cost with decrease in sample unit size was assumed to slow to an exponential (curved) form.* This assumption was made to reflect the fact that there would be a fixed cost of measurement even with a very small sample unit and that, between this small sample unit size and 2 mi^2, this fixed cost component would tend to dominate the area-variable component of cost.

Figure 3-40 shows the completed plot of relative measurement cost, C_1, versus sample unit size, s. Two curves are given there. The upper curve, representing cost set A, was constructed according to the method described above. The corresponding curve for cost set B is given for comparison.

Computation of Between Sample Unit Variance for Given Size of Unit

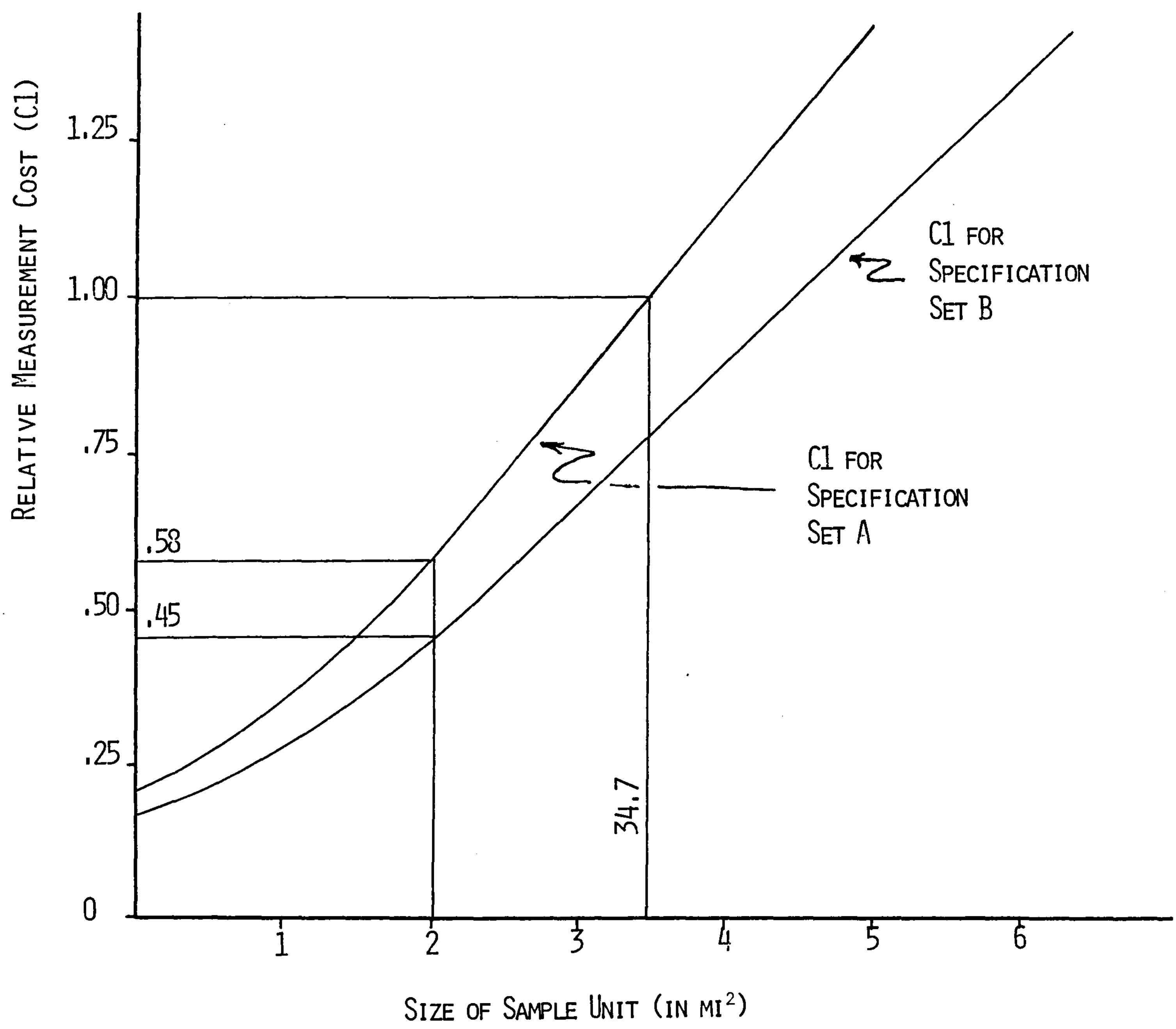
The next piece of information required in the sample unit size analysis was the expected between sample unit variance for given sizes of sample units. This sample variance was obtained through the application of the UCB Survey Planning Model (SPM) to the digital Landsat irrigation class map described earlier. In essence, the SPM was instructed to partition the class map into a grid of sample units having specified size. Each 'active' pixel in the class map was then assigned a 'one' if it belonged to an irrigated class and a 'zero' if it did not.

Active pixels were identified by a digital 'mask' as belonging to agricultural areas not excluded as native vegetation. For each sample unit the 'ones' were summed and divided by the number of 'active' pixels to produce an irrigated proportion value for that unit. The simple variance was then computed among the resulting sample unit proportions to give an estimate of valley floor variance expected for a given size of sample unit.

---

* The decay function chosen, exp (size/2), was a 'best' guess in absence of previous information.

FIGURE 3-40. RELATIVE MEASUREMENT COST VERSUS AREA OF A GIVEN SAMPLE UNIT ESTIMATED FOR THE FLOOR OF THE SACRAMENTO VALLEY IN 1979



Computation of Sample Size (n) to Achieve A Given Error Level for Each Size of Sample Unit

The number of ground sample units required to achieve a valley floor precision of $\pm$ 5 percent 95 times out of 100 was next computed for each size of sample unit. This was done by setting

$$\text{Error} = \text{Student's t times} \left[\hat{V}(\hat{\bar{Y}}_{\text{unstrat,regression}})\right]^{\frac{1}{2}}$$

$$= t_{n-2,.95}\left[\left(\frac{N-n}{Nn}\right)\left(\frac{n-1}{n-3}\right) S_y^2(1 - r^2)\right]^{\frac{1}{2}}$$

equal to .05 (i.e. a 95 percent confidence interval half-width of five percent) and then obtaining the quadratic solution for n. This was also done for an error of 3.5 percent to show the difference in sample size and cost with a higher precision goal. In the formula above, N represented the toal number of sample units used to calculate an estimate of S_y^2 in the SPM. Thus N and $\hat{S}_y^2$ were obtained for each value of sample unit size, with the assumption that the estimate of ground sample unit variance ($\hat{S}_y^2$), obtained by computing the variance among the population of Landsat sample units in the SPM, was in fact unbiased. Figure 3-41 shows the resulting SPM estimate of S_y^2 plotted over the range of 11 acres (a 3 pixel x 3 pixel 'dot') to 16 square miles. This curve was drawn by interpolating between values computed for S_y^2 at 11 acres, 49 acres, 79 acres, quarter mi^2, half mi^2, three quarter mi^2, one mi^2, two mi^2, three mi^2, four mi^2, and 16 mi^2. The curve was discontinuous at a sample unit size of zero mi^2. Note that the variance increased dramatically below two mi^2.

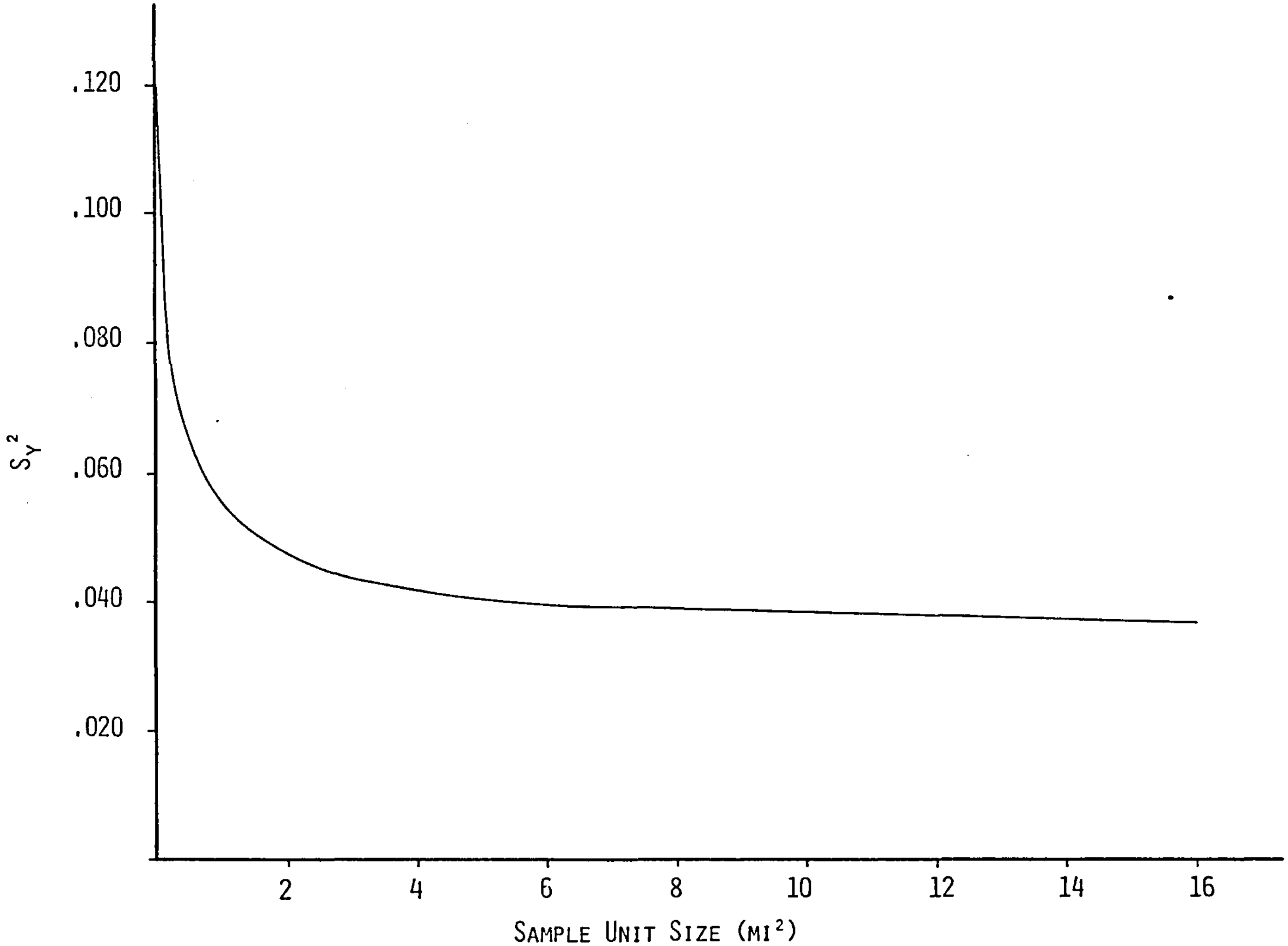
The terms shown to the right of Student's t-statistic in the equation above represent the estimate of sample variance for unstratified regression estimation.* Thus, to determine n, a Landsat-to-ground correlation (r) must be specified. Several values of r were selected in the range of .7 to .995 to show the sensitivity of n to correlation. The resulting sample size required to achieve a sampling precision of $\pm$ 5 percent, 95 times in 100 is plotted against sample unit size in Figures 3-42 and 3-43. The first figure provides results for sample error (absolute) expressed as a percentage of the sample frame and the second for sample error (relative) expressed as a percent of the area-wide estimate of irrigated proportion. Similar solutions for an absolute and relative error of 3.5 percent are provided in Figures 3-44 and 3-45.

These four figures show required sample size increasing significantly below a sample unit size of two mi^2. This reflects the increase in estimated ground sample variance below this threshold. Given an absolute or relative error goal

---

* The stratified version of this variance equation was introduced in Appendix IB as regression variance with factor 3.

FIGURE 3-41. BETWEEN SAMPLE UNIT VARIANCE (S_Y^2) FOR GIVEN AREA OF SAMPLE UNIT (FROM SPM SIMULATION)



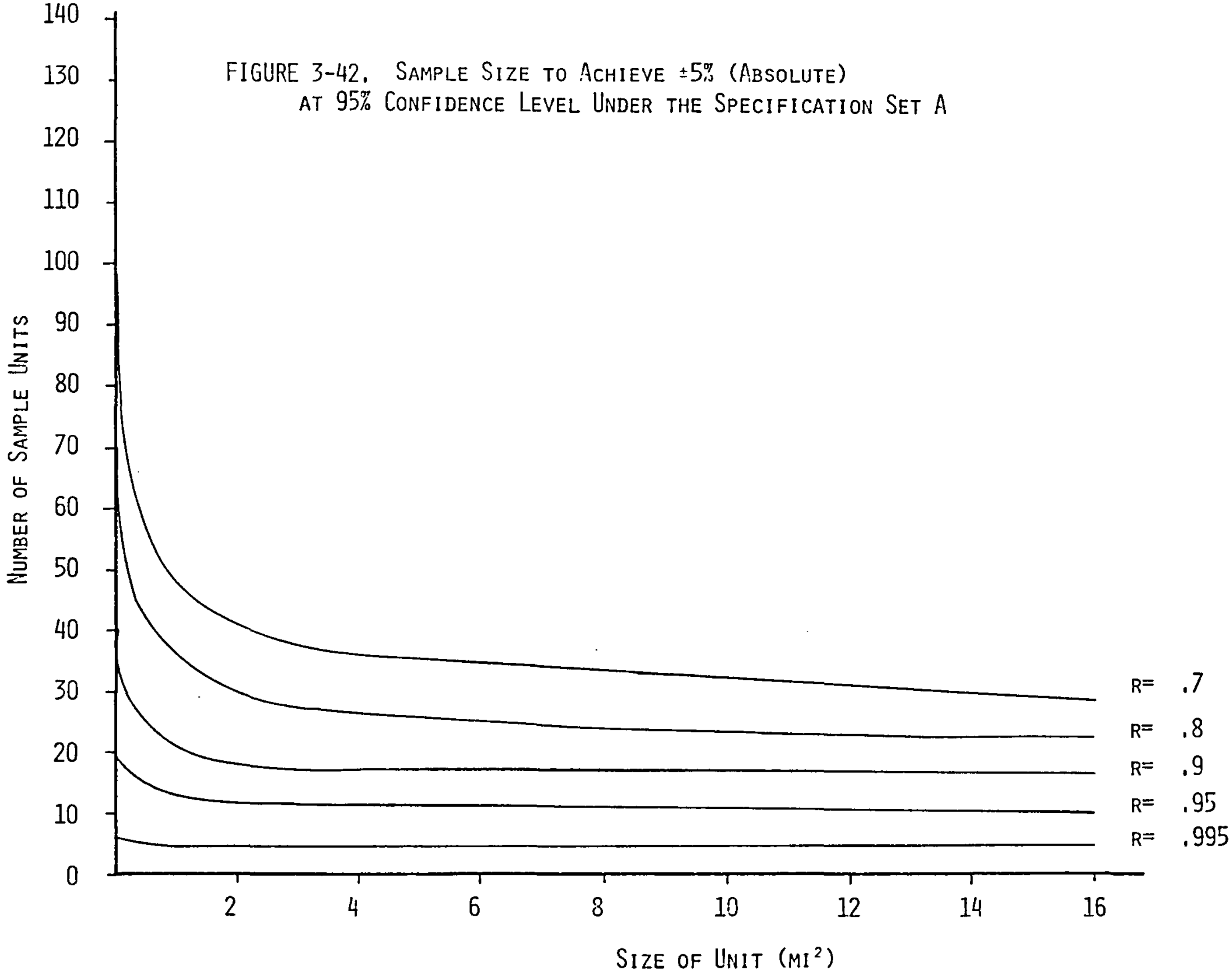
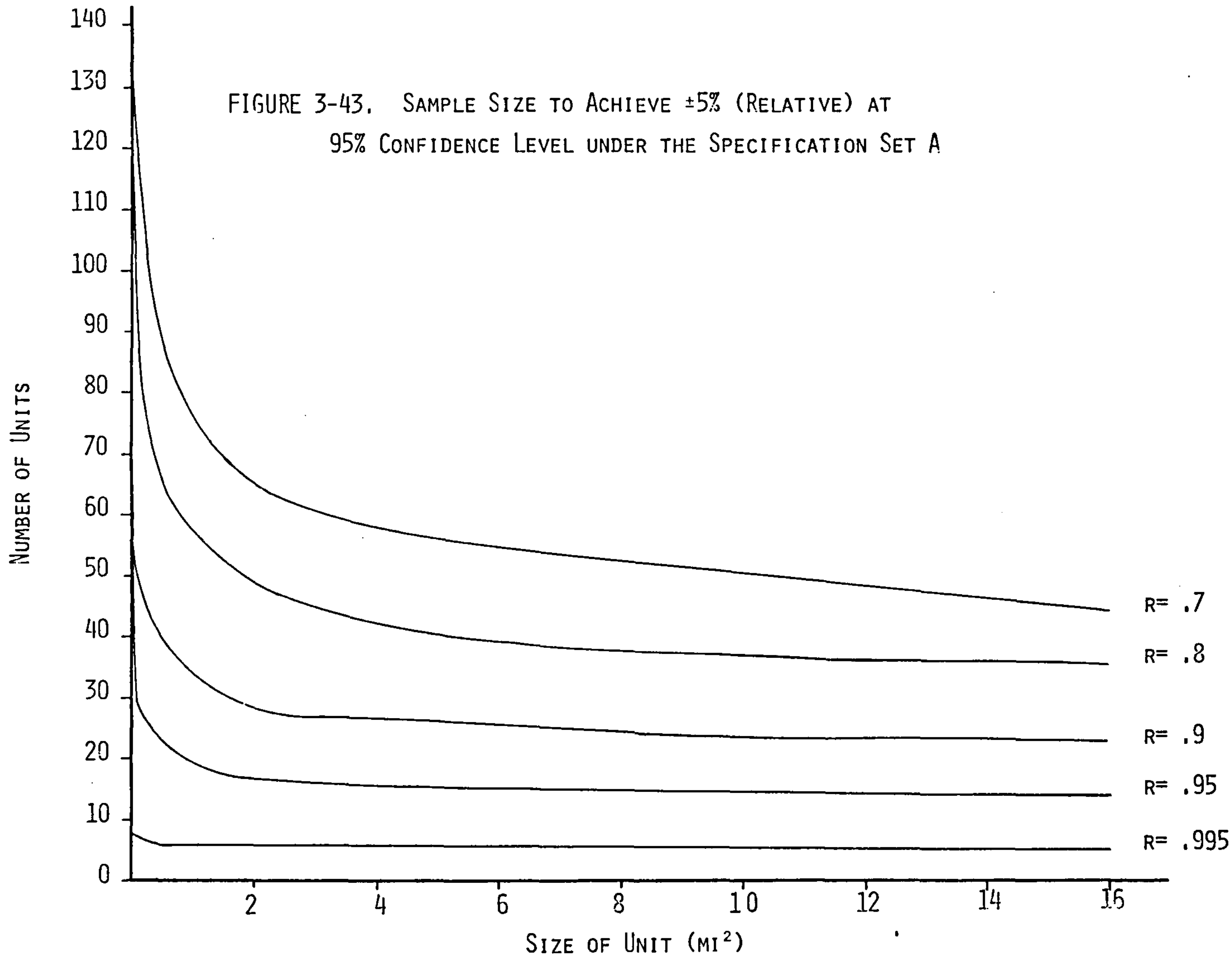


FIGURE 3-42. Sample Size to Achieve ±5% (Absolute) at 95% Confidence Level Under the Specification Set A

FIGURE 3-43. SAMPLE SIZE TO ACHIEVE ±5% (RELATIVE) AT 95% CONFIDENCE LEVEL UNDER THE SPECIFICATION SET A



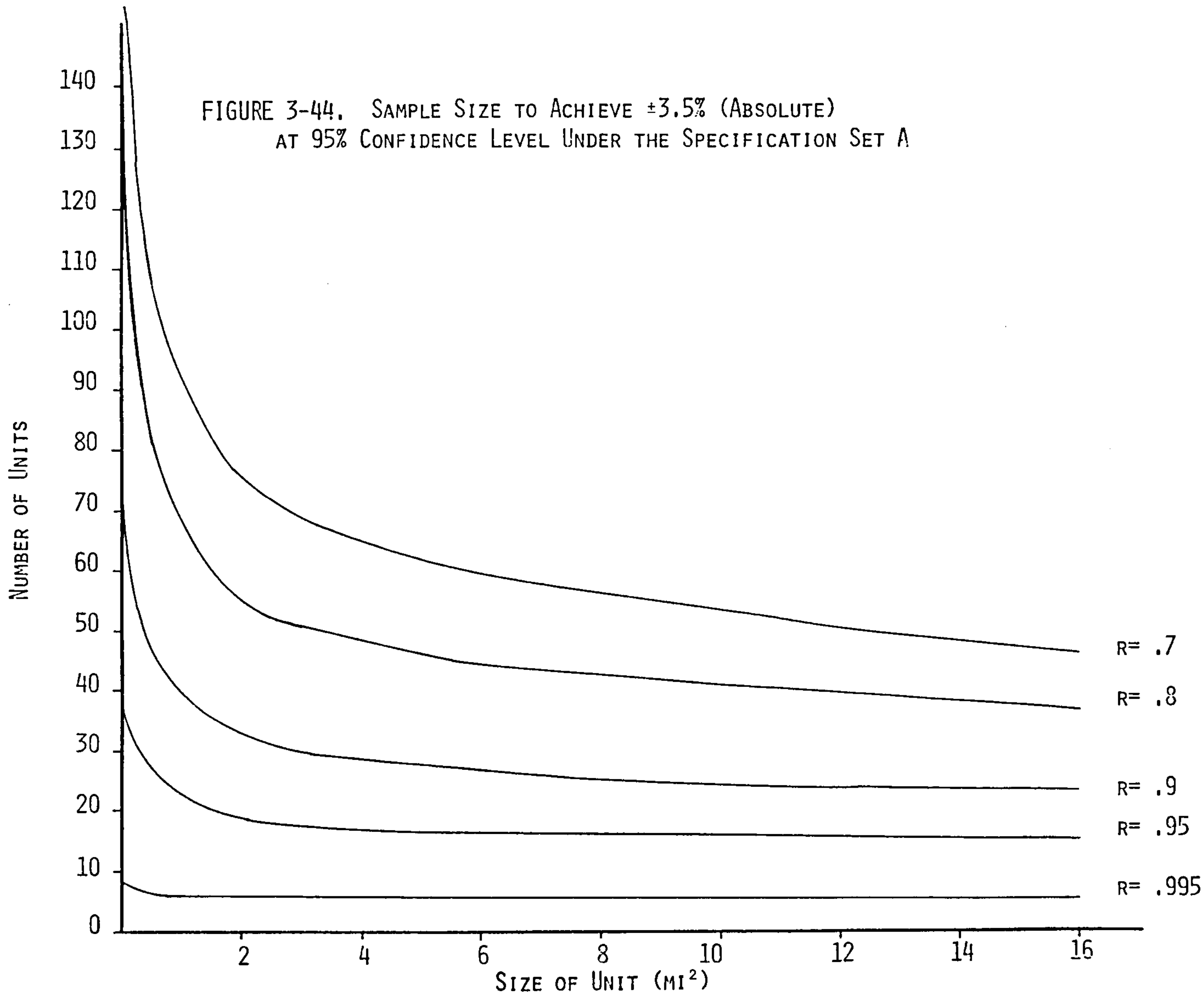
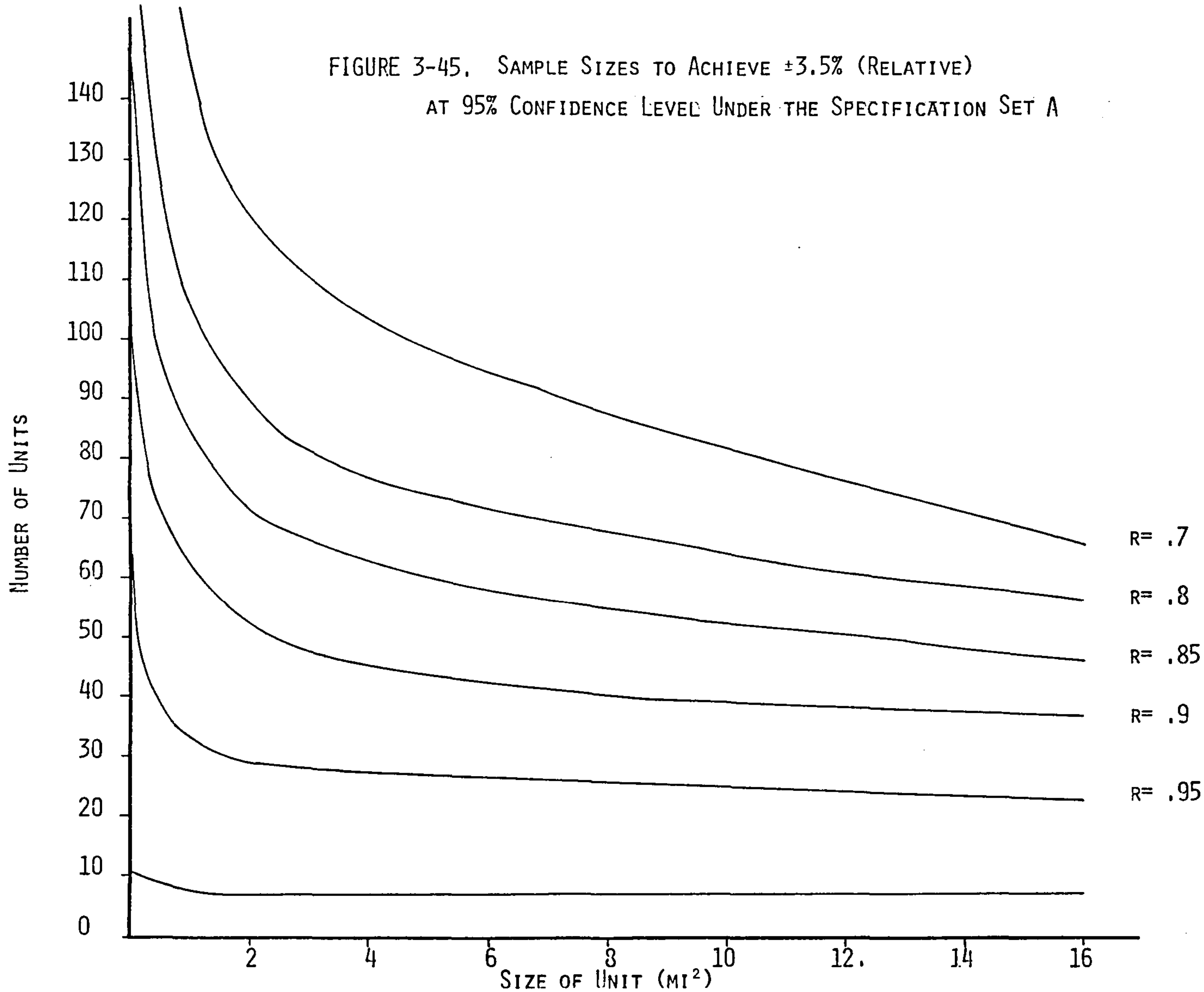


FIGURE 3-44. Sample Size to Achieve ±3.5% (Absolute) at 95% Confidence Level Under the Specification Set A

FIGURE 3-45. SAMPLE SIZES TO ACHIEVE ±3.5% (RELATIVE) AT 95% CONFIDENCE LEVEL UNDER THE SPECIFICATION SET A



of five percent, difference in Landsat-to-ground correlation was found to increase required sample size by a factor of 9 to 10 times when correlation was reduced from .995 to .7 . A corresponding increase of 10 to 15 times was seen to occur when the error goal was 3.5 percent. Below sample unit sizes of two mi^2, differences in sample size between correlations of .7 and .995 were even larger. In most major agricultural areas inland from the coast, an operational system would be expected to achieve Landsat-to-ground correlations in the range of .9 to .995 . The difference in required sample size between these two correlations varies by a factor of approximately three to five above two mi^2.

Computation of the Total Variable Cost to Achieve A Given Error Level

The final step in determining the impact of sample unit size on cost was the computation of total variable cost (TVC) to achieve a given level of sampling precision. This analysis is designed to show which size of sample unit is likely to achieve a desired error goal at minimum cost. A cost measure was defined that would reflect only those costs that varied with the number of sample sample units and with their relative size. This value, termed total variable cost (TVC), was a function of C_1 and C_2 described earlier, viz:

$$TVC = (C_1 n) + (C_2 n)(\sqrt{n_{baseline}} \div \sqrt{n})$$

where n was the sample size calculated to achieve a given error goal for a given size of sample unit and a given Landsat-to-ground correlation. C_1 included all ground sample unit measurement costs exclusive of travel and C_2 included all travel costs associated with ground units. The product of C_2 and n was multiplied by a ratio of square roots to account for the effect of reduced travel times as the number and therefore density, of sample units increased. This meant that C_2 had to be calculated for an initial 'standard' sample size ($n_{baseline}$) before it could be employed in the formula above. A sample size of approximately 50 units was selected to represent this 'initial' density of ground sample units on the floor of the Sacramento Valley. This figure corresponded roughly to that obtained in the actual 1979 irrigated lands inventory. Thus, if $n_{baseline}$ exceeded the n calculated to achieve a given error level, the unit travel cost (C_2) would in effect be increased over what it was with $n_{baseline}$. Intuitively this makes sense, since a lower ground sample size will increase the distance between units to be measured and therefore increase travel cost. If $n_{baseline}$ was less than n, the opposite effect would occur.

The resulting TVC for each size of sample unit (and for given error goal and correlation) was then expressed as a fraction of the TVC required for the average-size valley floor sample unit (3.47 mi^2). These 'relative' TVC's were then plotted versus size of sample unit. Figures 3-46 through 3-49 present the results for error goals (absolute and relative) of 5.0 and 3.5 percent 95 times out of 100 and for correlations ranging from .7 to .995. Examination of these figures showed that the minimum TVC for given error generally occurred between sample unit sizes of .5 to 1.0 mi^2.

Conclusions Relative to the Sample Unit Size Analysis

(1) The lowest calculated total variable cost occurred for sample unit sizes ranging from .5 to 1.0 mi^2 regardless of correlation for both 5.0 and 3.5 error goals.

(2) However, the percent of a sample unit subject to misregistration error increases significantly below a size of one square mile as shown in Figure 3-50.

(3) Thus, based on the results shown above, the University of California team would recommend that the size of the valley floor sample units in the Sacramento basin be set at 1.0 to 1.5 mi^2 in an operational inventory; this size range should allow achievement of overall basin error goals at minimum TVC, subject to specification of optimal sample unit size in land use strata not falling in valley floor areas.

(4) A comparison of land use stratum-specific standard errors presented earlier in Table 3-14, and of plots of Y minus X versus size of sample unit among basins throughout the state of California suggests that:

   (a) In most valley floor areas throughout the state, the 1.0 to 1.5 mi^2 sample unit size may be best from the standpoint of minimizing TVC for a given basin error goal; and

   (b) In other areas of higher standard error (e.g. the dryland stratum) or high Y minus X values at small size, a 2.0 to 2.5 mi^2 sample unit size may be preferable.

(5) Under the assumptions used in this analysis, the Sacramento Valley floor TVC versus sample unit size curves shown in Figures 3-46 to 3-49 suggest a potential ground sample unit cost savings of approximately 30 to 35 percent for the 1.0 to 1.5 mi^2 unit size relative to the corresponding cost of a 3.47 mi^2 unit.

FIGURE 3-46. Estimated Relative Cost to Achieve ±5% Absolute Error at the 95% Level of Confidence With Cost Set A

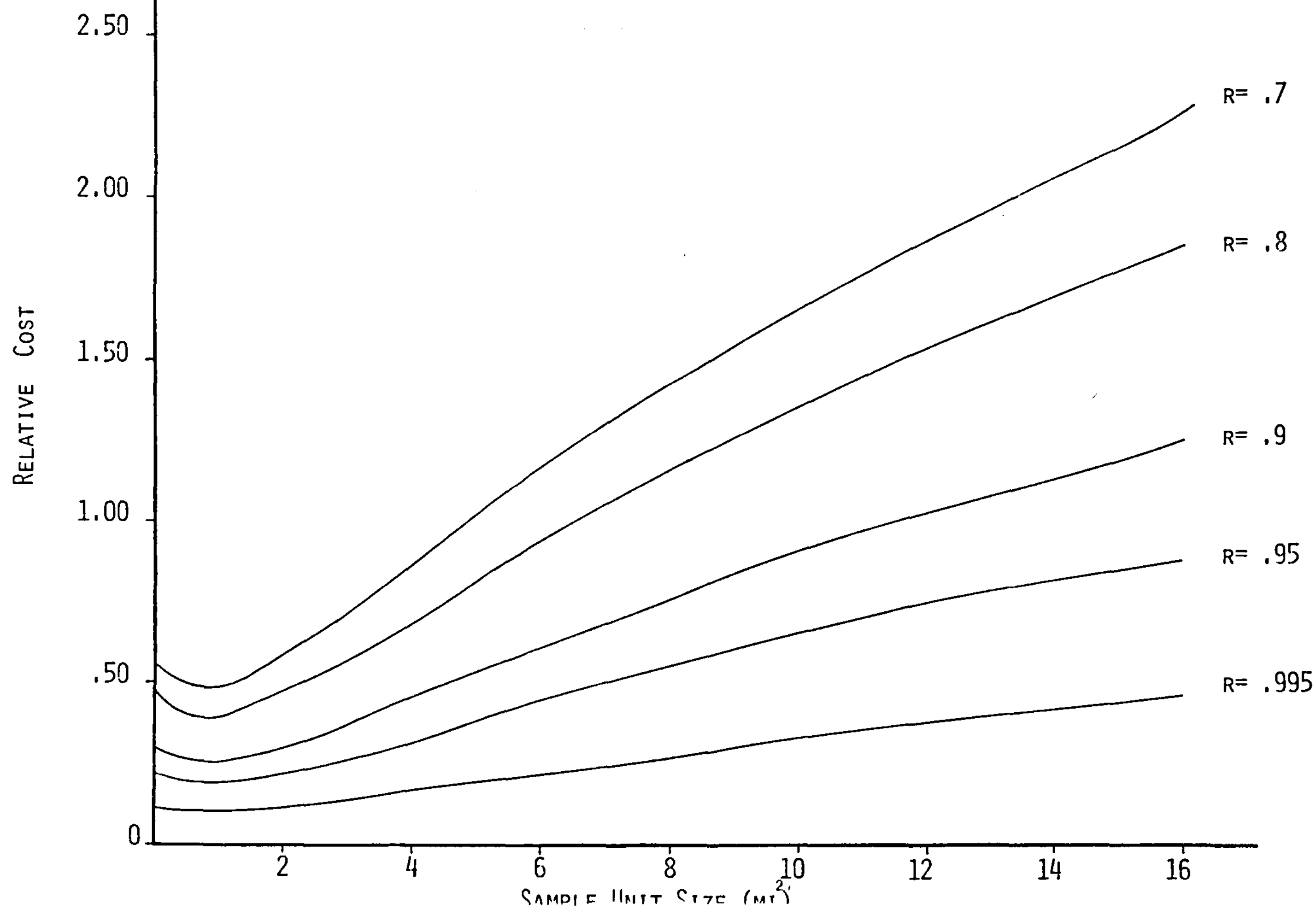


FIGURE 3-47. ESTIMATED RELATIVE COST TO ACHIEVE ±5% RELATIVE ERROR AT THE 95% LEVEL OF CONFIDENCE WITH COST SET A

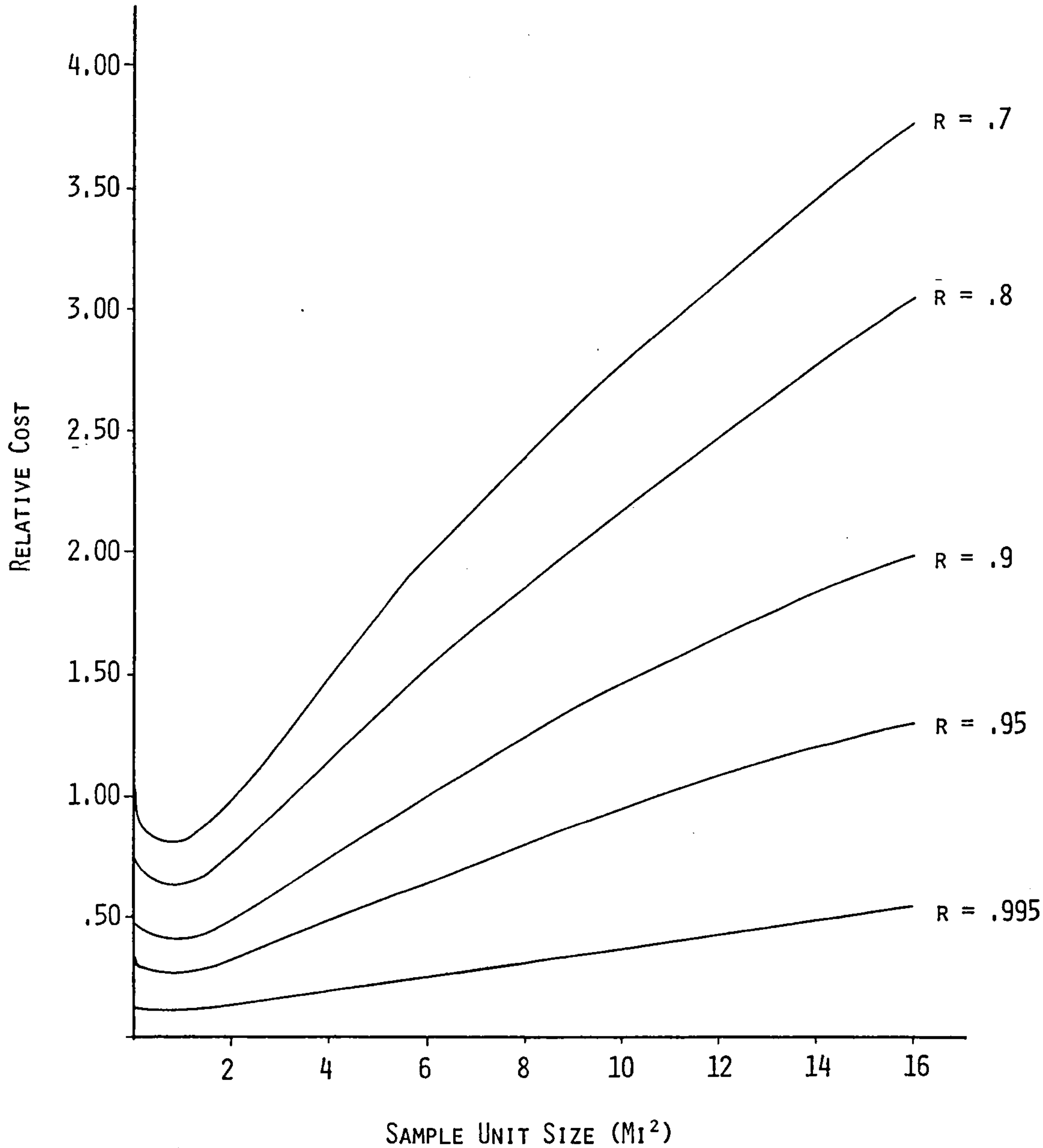
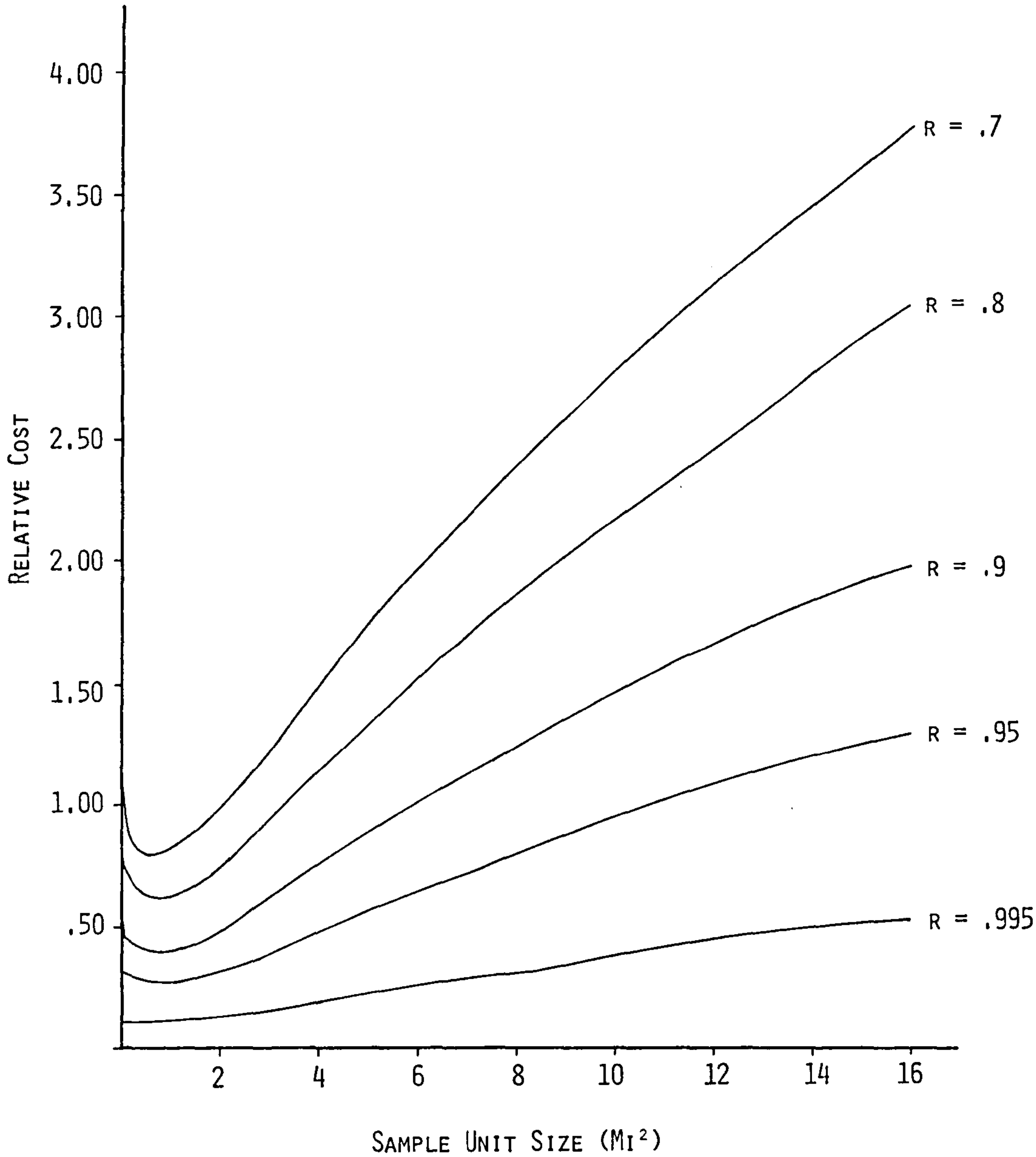


FIGURE 3-48. ESTIMATED RELATIVE COST TO ACHIEVE ±3.5% <u>ABSOLUTE</u> ERROR AT THE 95% LEVEL OF CONFIDENCE WITH COST SET A



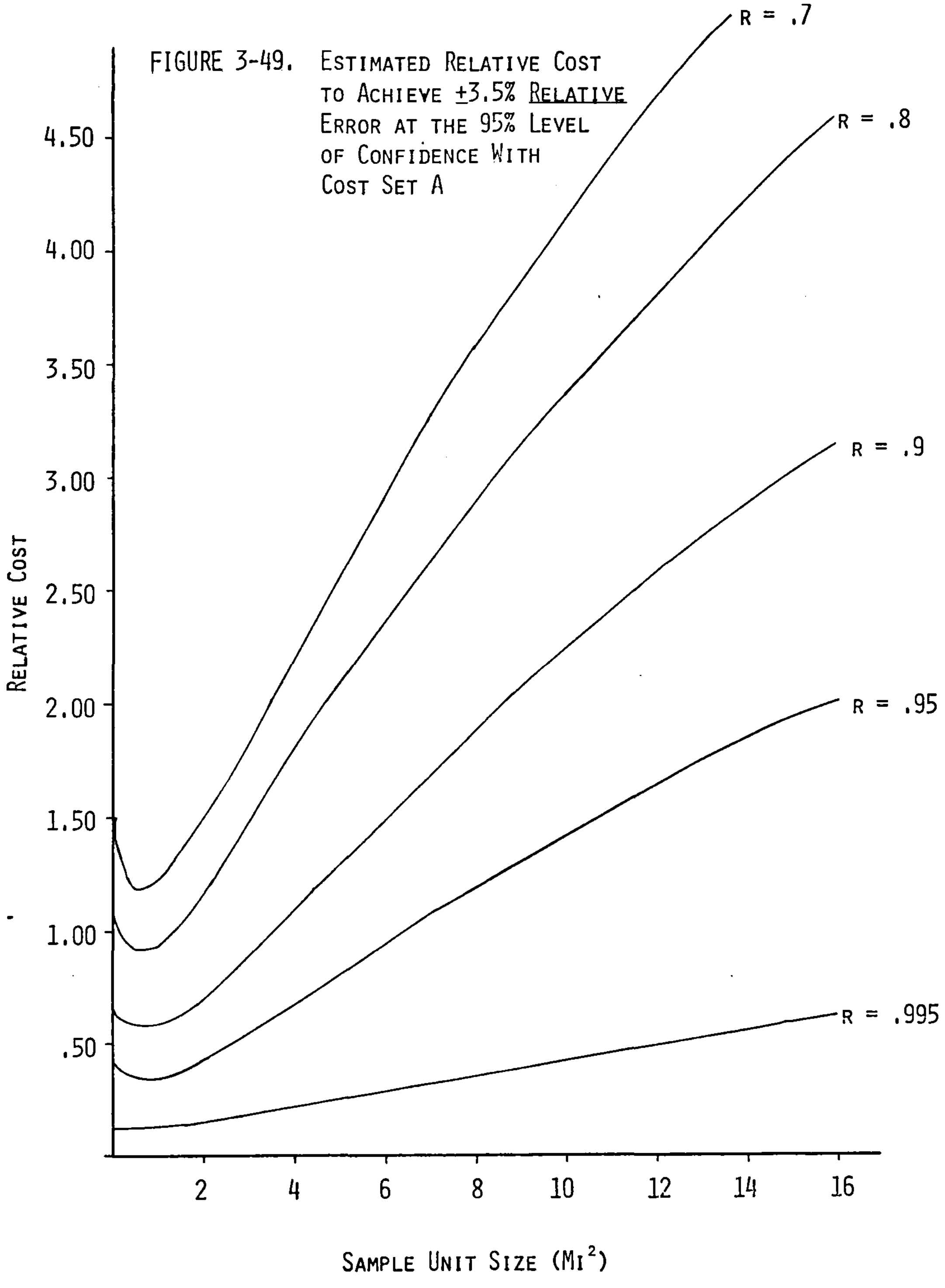
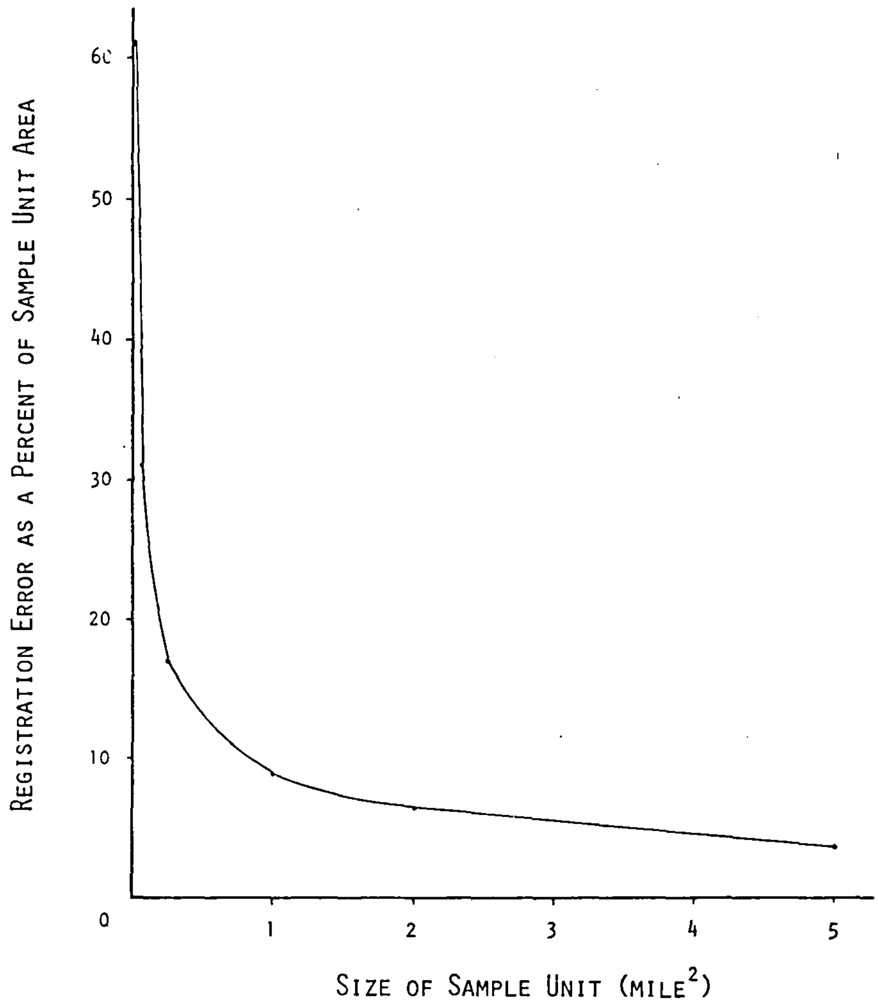


FIGURE 3-49. ESTIMATED RELATIVE COST TO ACHIEVE ±3.5% RELATIVE ERROR AT THE 95% LEVEL OF CONFIDENCE WITH COST SET A
R = .7
R = .8
R = .9
R = .95
R = .995
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
RELATIVE COST
2
4
6
8
10
12
14
16
SAMPLE UNIT SIZE (MI²)

FIGURE 3-50. APPROXIMATE PERCENT OF SAMPLE UNIT AREA MISREGISTERED GIVEN A REGISTRATION ERROR OF ONE CELL (½ HECTARE) IN BOTH X & Y DIRECTIONS



3.7.4 Expected Sample Size and Total Variable Cost Using An Unstratified 1979 Design

Recall that the original computation of required basin sample size was based on "best guesses" of variance, correlation, and cost. No previous design information was available in most basins. Thus an important question regarding the 1979 inventory, and an important consideration in future inventories was: What should the sample size and resulting total variable cost have been in order to meet the $\pm$ 5 percent, 95 times in 100 error goal in each basin in 1979, given the ground sample variances, Landsat-to-ground correlations, and costs actually encountered in 1979? To answer this question, the unstratified design was selected for analysis. Choice of this design was based on (1) simplicity of computations and on (2) its ability to demonstrate the general degree of sample size and cost change expected in operational system. Its use here is not intended as a recommendation for an operational system. The method was as follows:

(1) The regression sample variance expression with factor 3 was used to compute the sample size required to meet a number of allowable error (confidence interval half-width) goals. This variance expression assumed that (a) Landsat measurements having matched ground data were distributed according to a normal distribution and that (b) degrees of freedom were equal to n - 2. The form of this regression variance was described in Appendix II. The required sample size was obtained by quadratic solution for n in the manner presented in the last section. Unstratified basin ground sample variance, Landsat-to-ground correlation, and total sample unit population size actually encountered in 1979 were used in the variance expression. The values for unstratified ground sample unit variance were not adjusted (by Cochran's 1977 formula 5.A.44) for the fact that original sample allocation was based on optimal as opposed to proportional allocation among strata. See Appendix II for a discussion of the effect of the formula 5.A.44.

(2) Next, per sample unit values for C_1 (variable ground unit measurement costs excluding travel) and C_2 (ground travel cost) were computed according to the formulas presented in the last section. Times and costs reported by DWR (Ferchaud 1980) in 1979 were used for each basin.

(3) The total variable cost was computed using the formula

$$TVC = C_1 n + (C_2 n)(\sqrt{n_{baseline}} \div \sqrt{n})$$

for each basin. Note that in this formula, as explained in the last section, C_2 represents the average cost of travel per sample unit at the density of sample units chosen as a baseline - in this case, the density of sample units ($n_{baseline}$) actually used in 1979.

(4) The TVC resulting from each sample size (n) required to meet a given error goal was then expressed as a fraction of the TVC actually incurred in 1979.

(5) The resulting 'relative' TVC's were plotted versus confidence interval half-width (allowable error). Figures 3-51 to 3-60 show the resulting plots for each of the ten basins. Two reference relationships are drawn on each figure. The first represents the absolute allowable error actually achieved in 1979. It can be identified by the horizontal line projecting from the relative TVC value of 1.0 (i.e. the variable cost resulting from the actual sample size (n) used in 1979) to the curve and from this point of intersection by vertical line to the horizontal axis. The resulting value on the horizontal axis represented the allowable error actually achieved in 1979 - for example, .0218 as read off the North Coast figure. The second relationship shown, and the important one in this section, is the TVC expected for an allowable error of .05 . This can be read from the North Coast figure to be .406 TVC, that is approximately 41 percent of the total variable cost actually spent in 1979. If the TVC for a five percent allowable error is less than 1.0, then that fractional TVC value will also represent the lower limit on the potential reduction of sample size (n). This last statement follows from the form of the TVC cost function.

Results and Discussion

(1) The plots of relative TVC versus absolute allowable error in Figures 3-51 through 3-60 suggest that total variable cost for achievement of an irrigated proportion sample error of $\pm$ 5 percent of the sample frame 95 times out of 100 can be reduced significantly. Results in the coastal basins, for example, showed a reduction of roughly 50 to 60 percent might be possible. Similarly, the plotted curves for the Central Valley basins suggest a potential reduction of 65 to 73 percent.

(2) Given the form of the TVC cost function, the curves shown in Figures 3-51 to 3-60 also suggest that ground sample size can be reduced by as much or, perhaps, slightly more.

Whether or not the cost savings or sample size reductions indicated in the figures can be achieved in a future inventory will depend on a number of factors. First, the variance, correlation, and cost estimates obtained in the 1979 inventory may not be the same as those of a future inventory. This could derive from the nature of the sample itself, another sample giving a slightly higher ground variance and lower Landsat-to-ground correlation. Or, alternatively, the method of pooling stratified observations used in this evaluation could have underestimated actual ground variance. Similarly, more experience with cost data might suggest a somewhat different form for the TVC cost function or different values for its coefficients C_1 and C_2.

FIGURE 3-51. ESTIMATED RELATIVE COST VERSUS ABSOLUTE ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE NORTH COAST BASIN

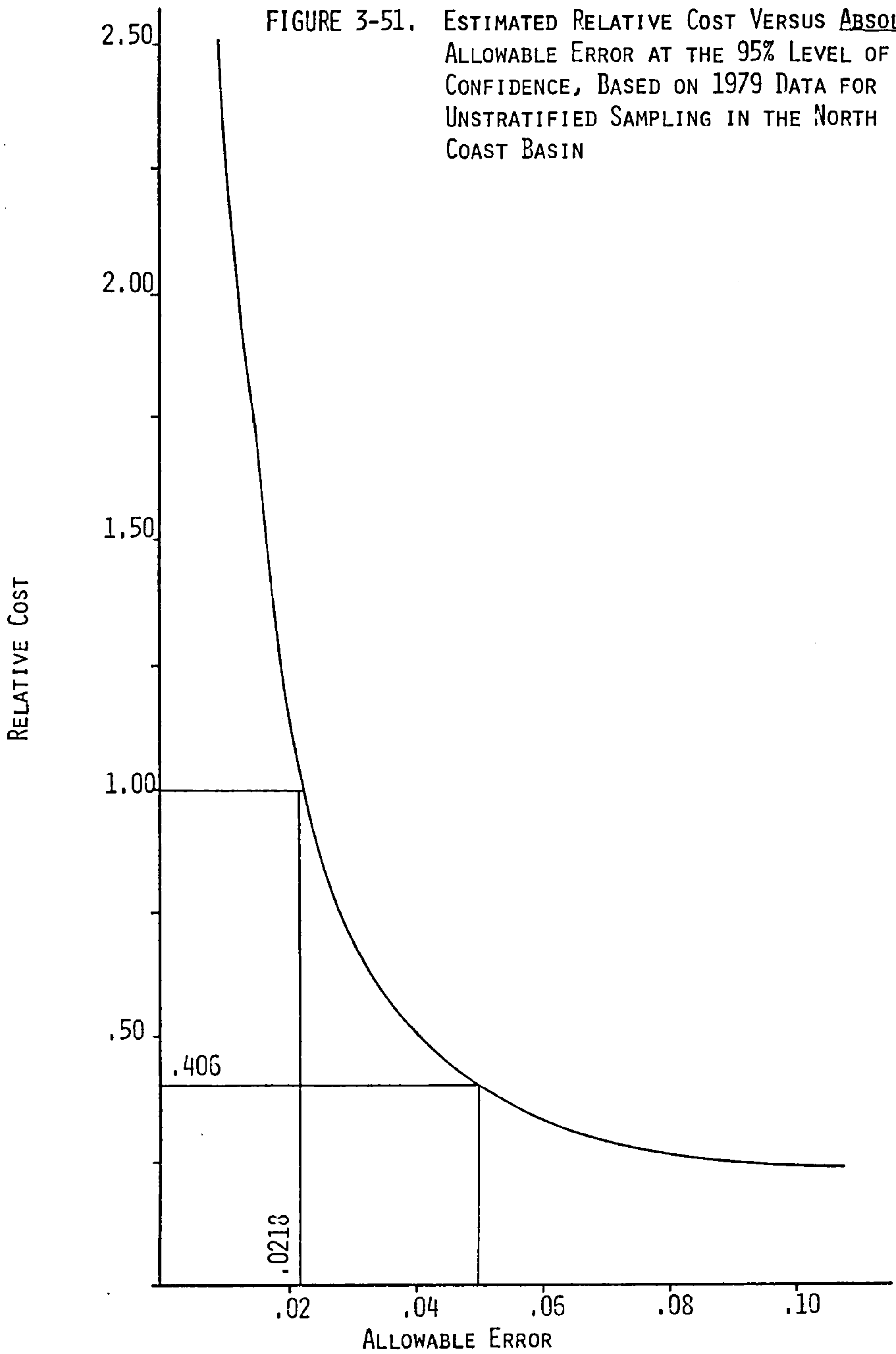


FIGURE 3-52. ESTIMATED RELATIVE COST VERSUS <u>ABSOLUTE</u> ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE SAN FRANCISCO BASIN

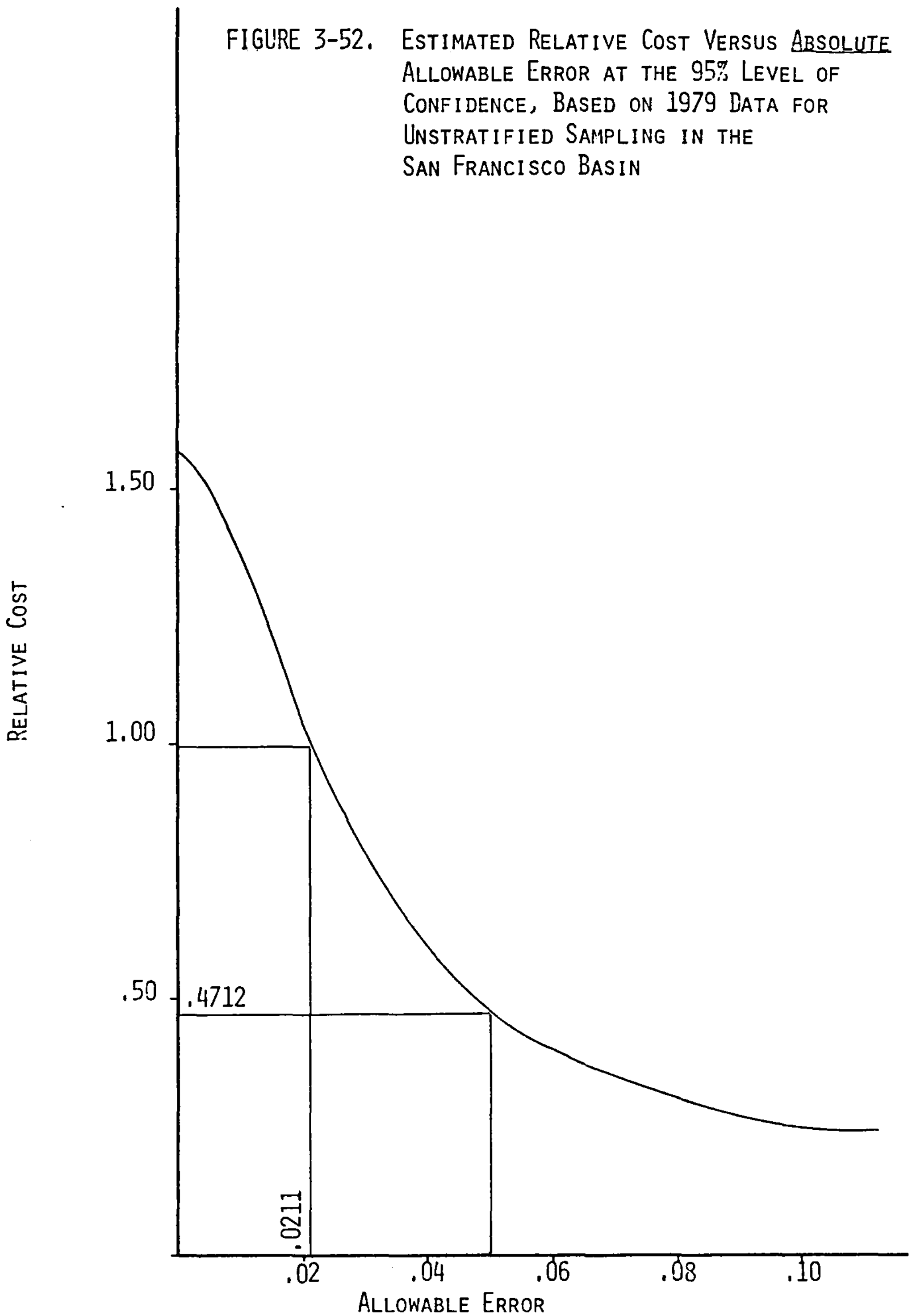
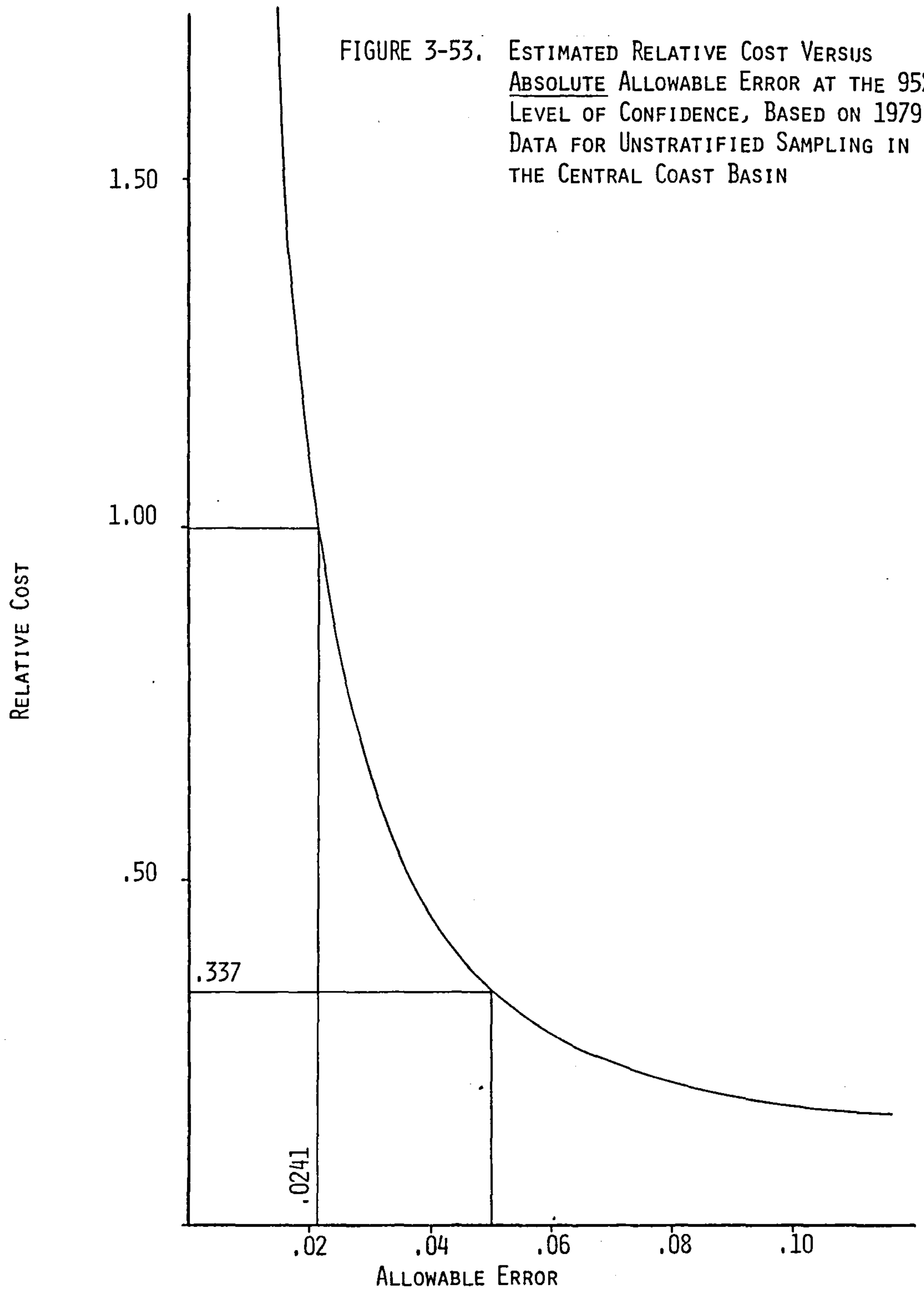


FIGURE 3-53. ESTIMATED RELATIVE COST VERSUS ABSOLUTE ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE CENTRAL COAST BASIN



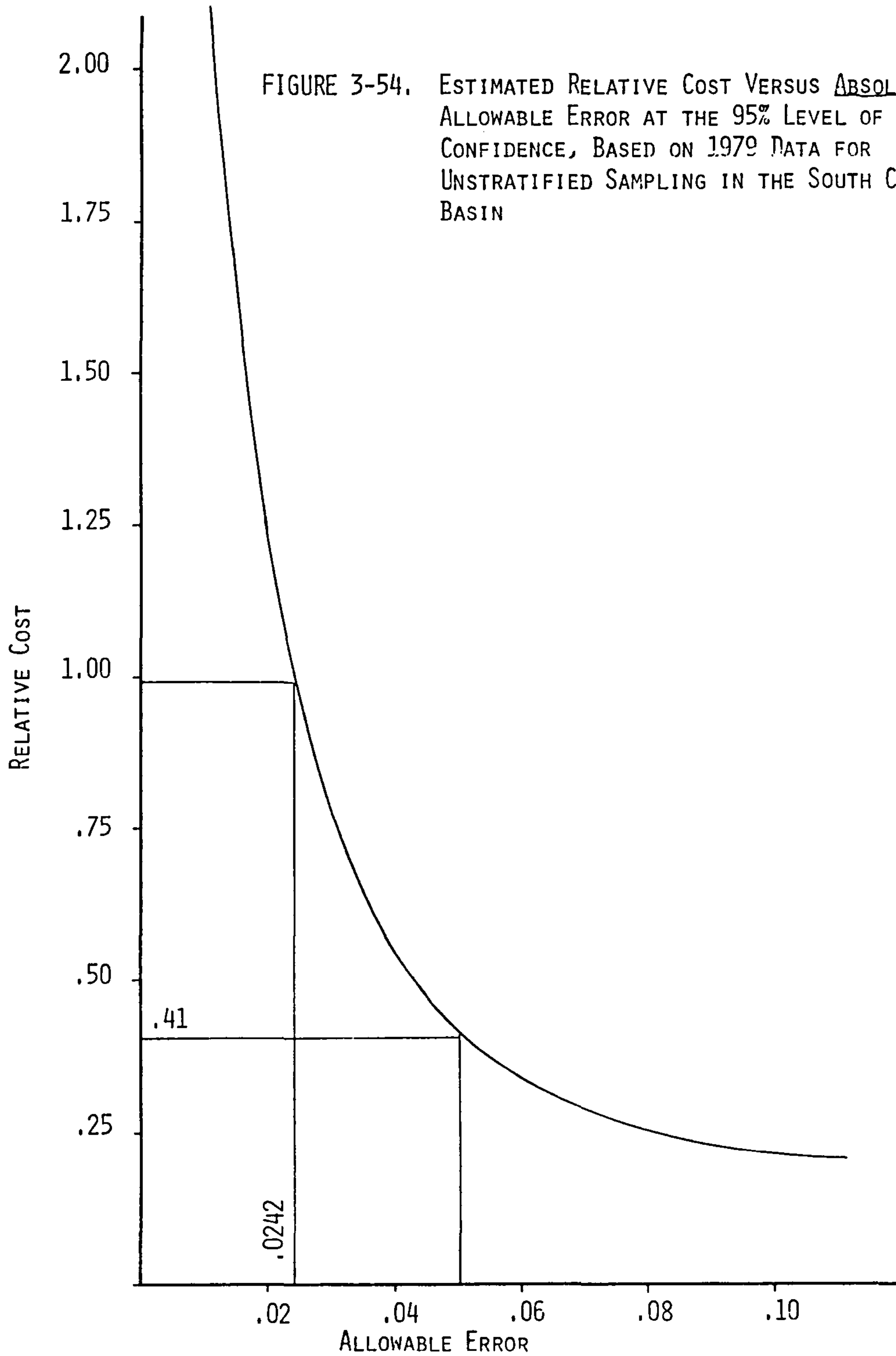


FIGURE 3-54. Estimated Relative Cost Versus Absolute Allowable Error at the 95% Level of Confidence, Based on 1979 Data for Unstratified Sampling in the South Coast Basin

FIGURE 3-55. ESTIMATED RELATIVE COST VERSUS <u>ABSOLUTE</u> ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE COLORADO DESERT BASIN

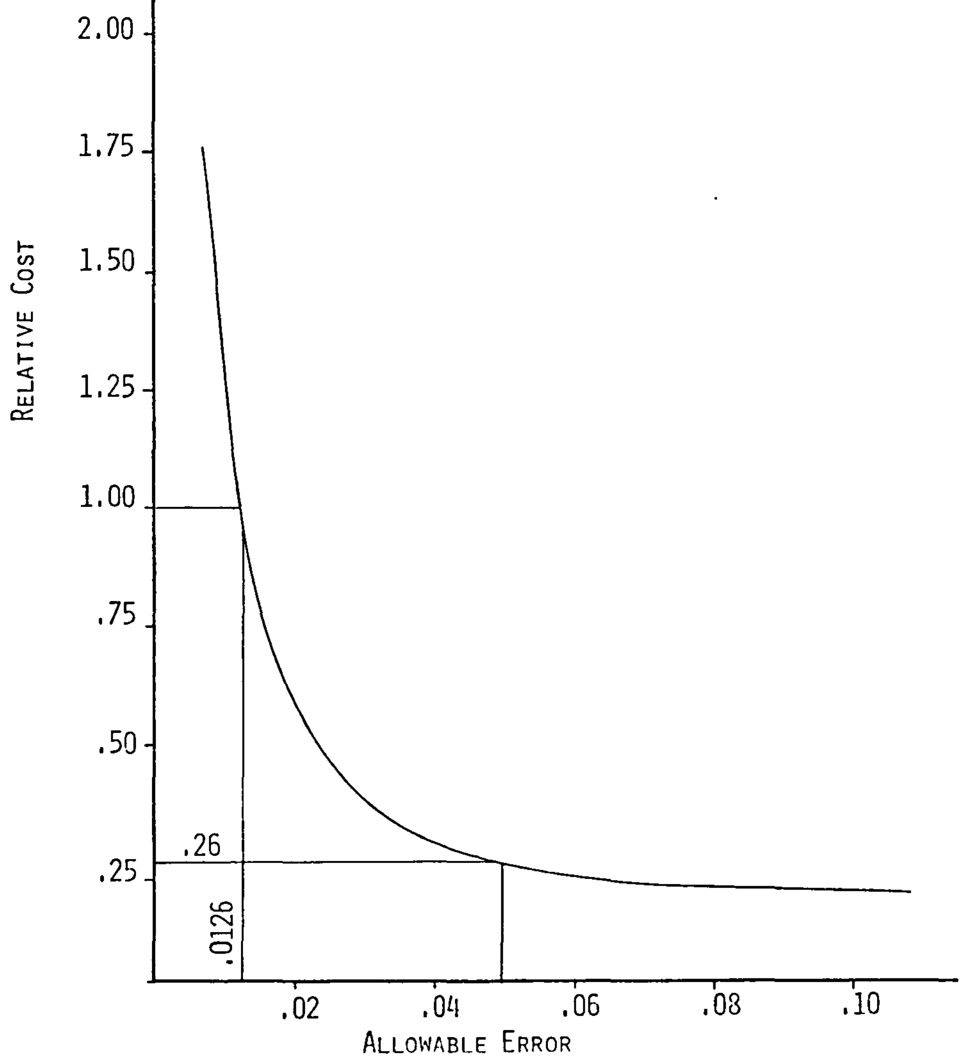


FIGURE 3-56. ESTIMATED RELATIVE COST VERSUS ABSOLUTE ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE SOUTH LAHONTON BASIN

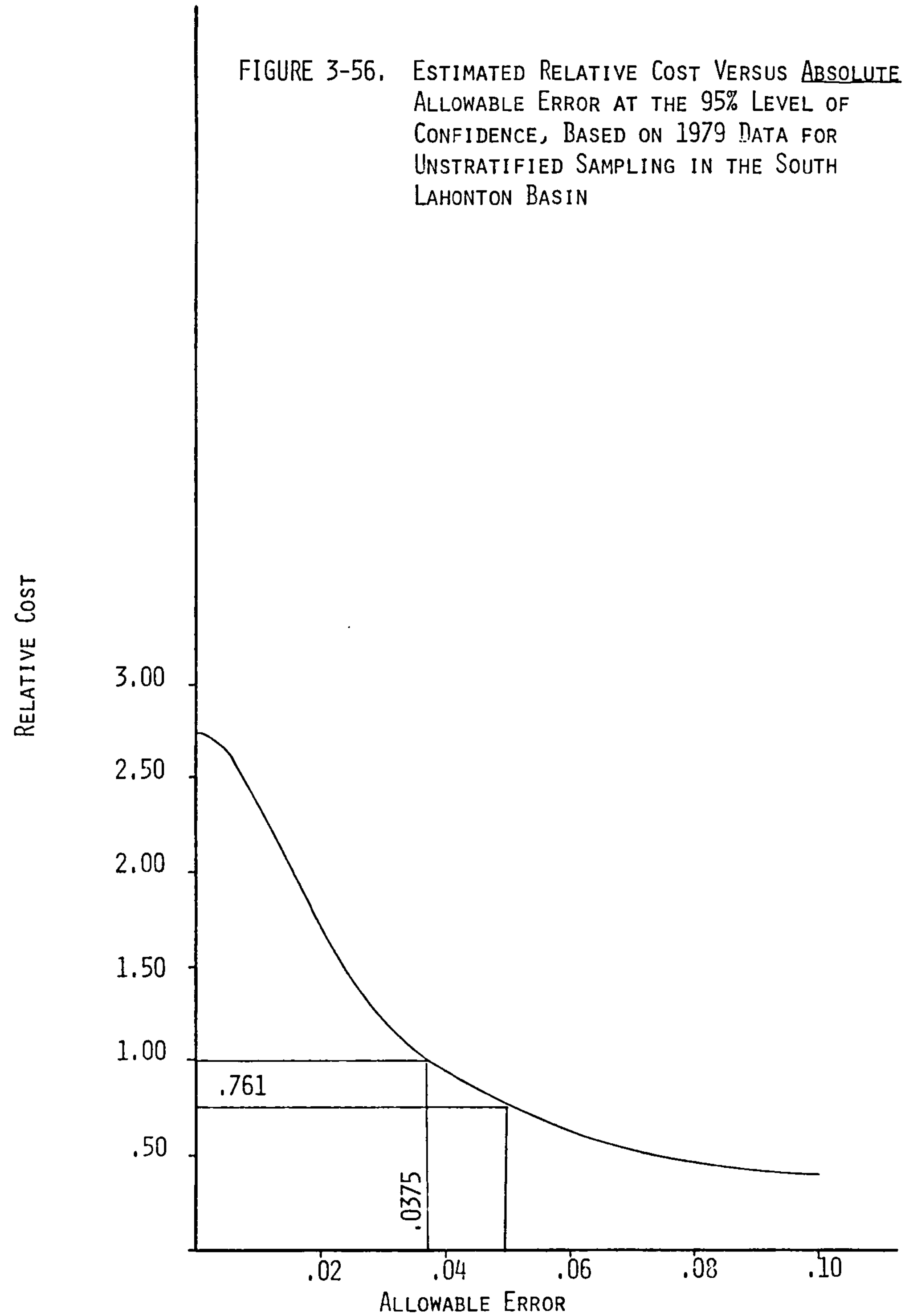


FIGURE 3-57. ESTIMATED RELATIVE COST VERSUS ABSOLUTE ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE NORTH LAHONTON BASIN

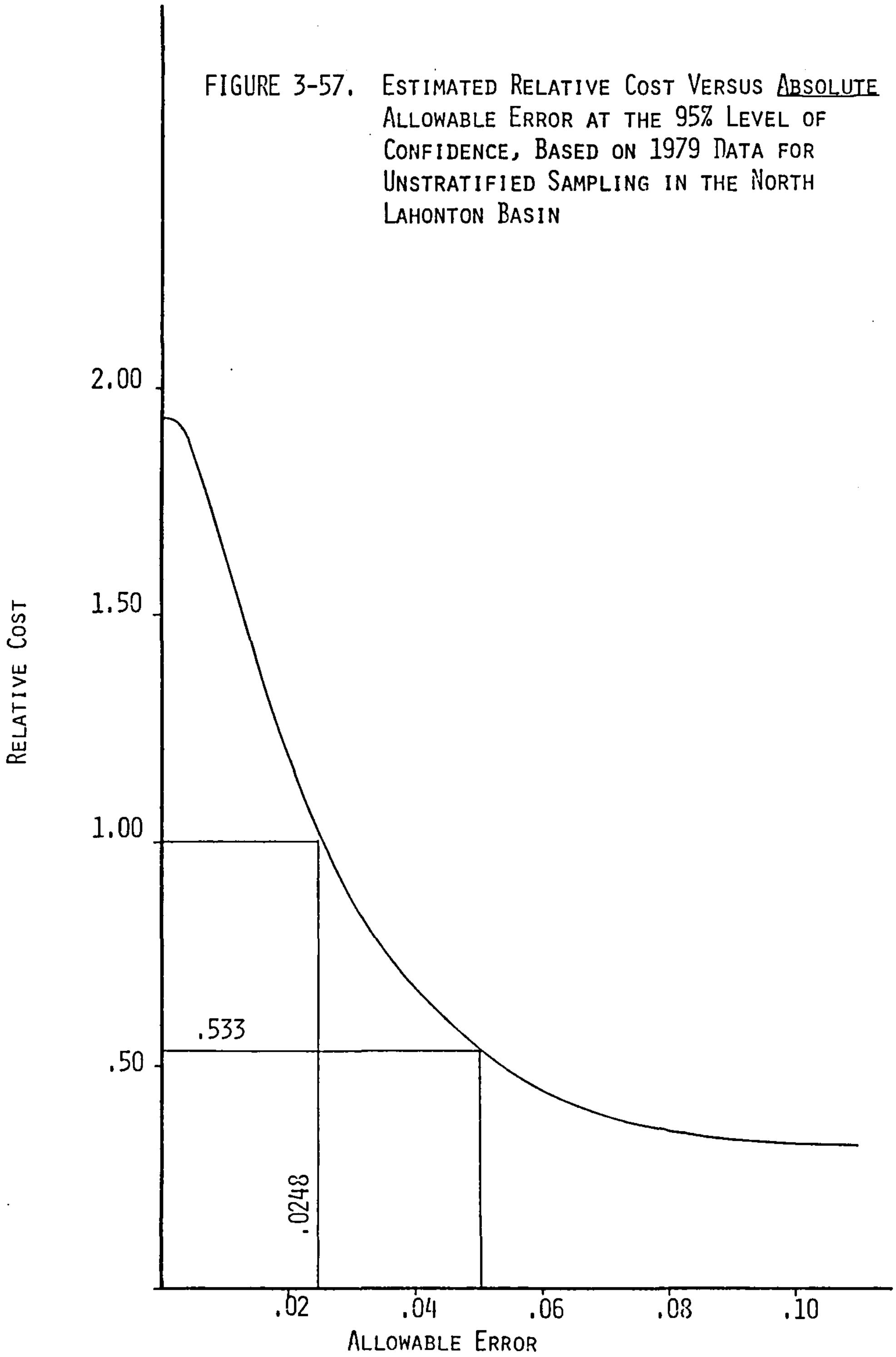
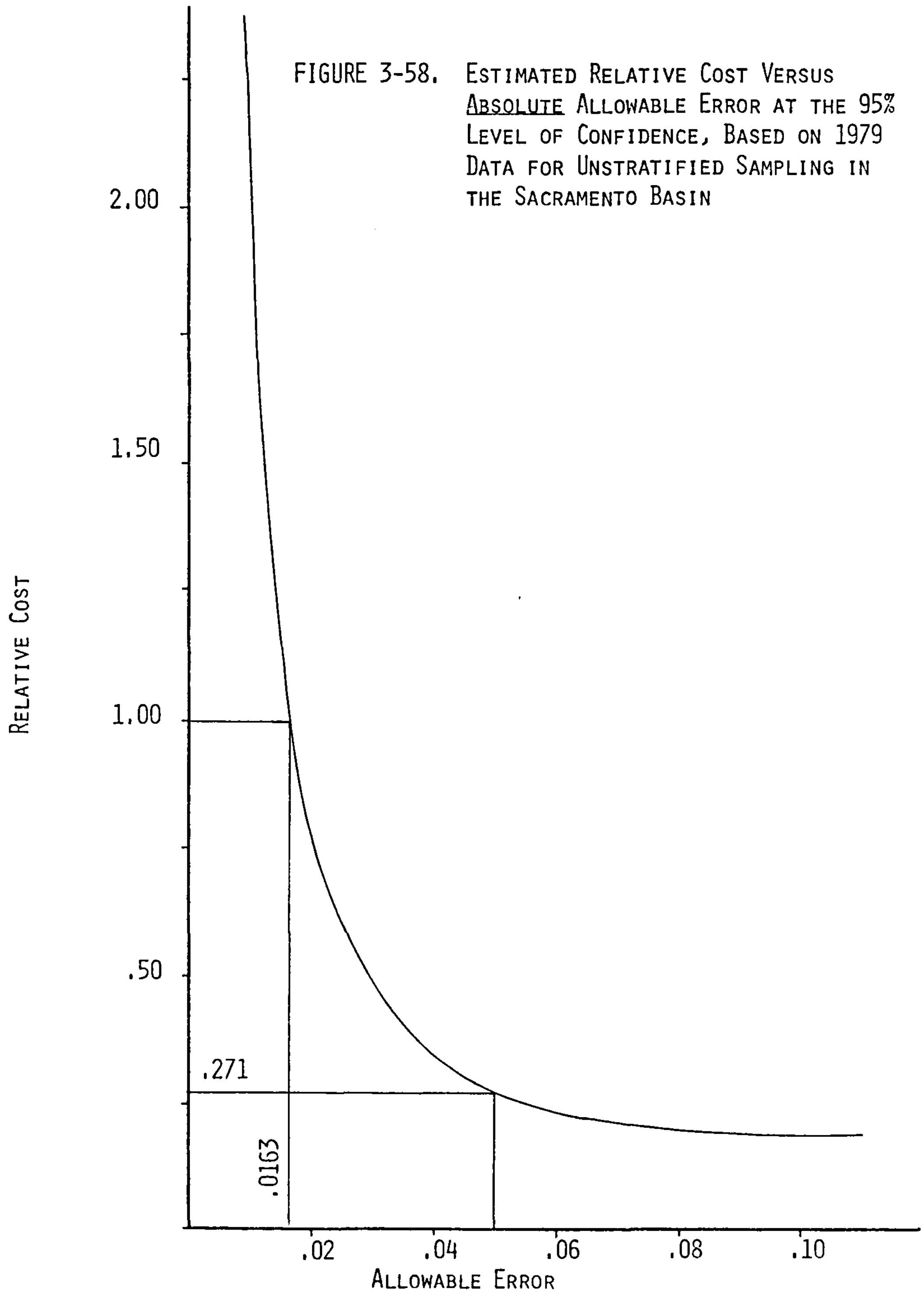


FIGURE 3-58. ESTIMATED RELATIVE COST VERSUS <u>ABSOLUTE</u> ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE SACRAMENTO BASIN



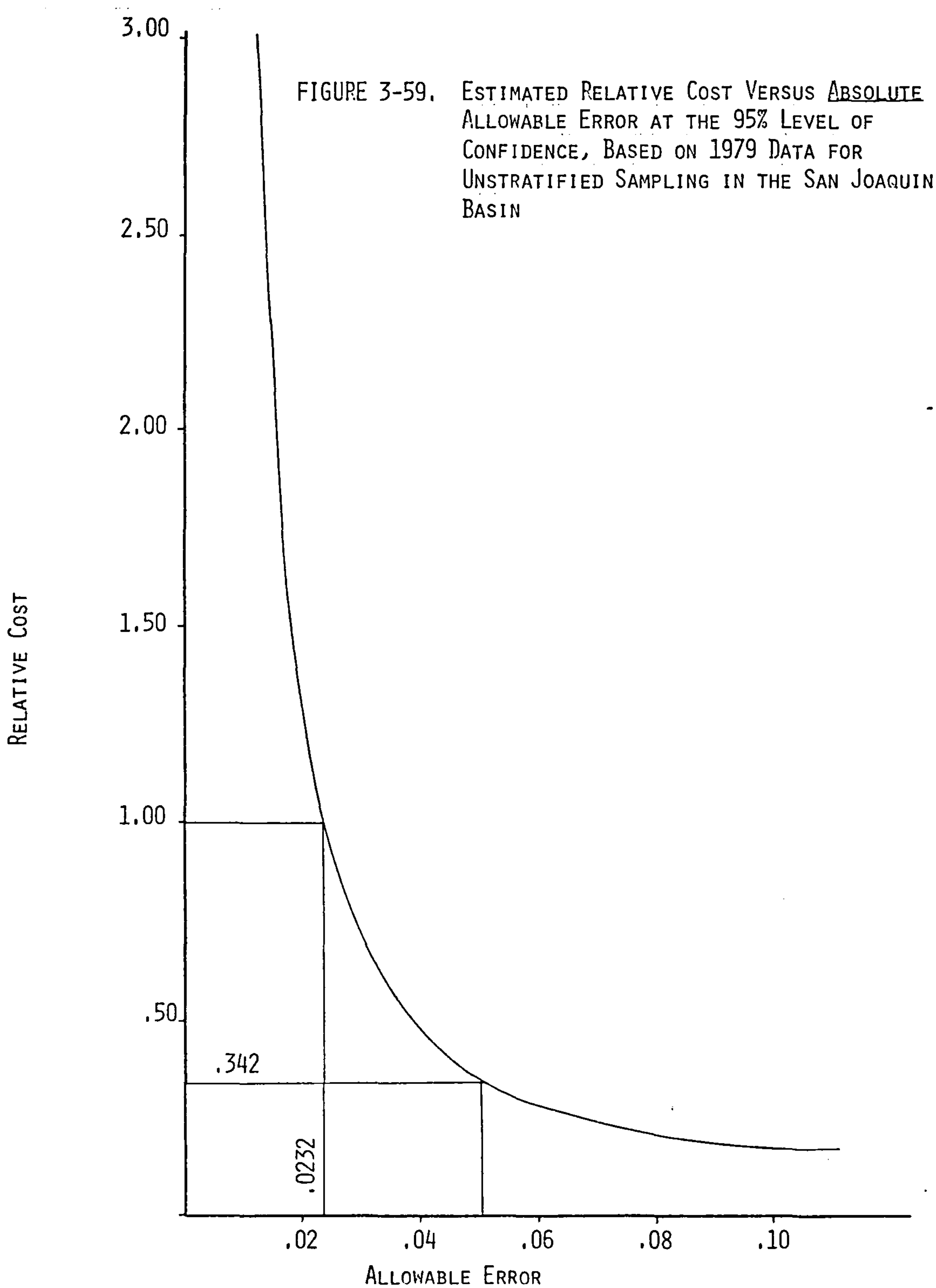
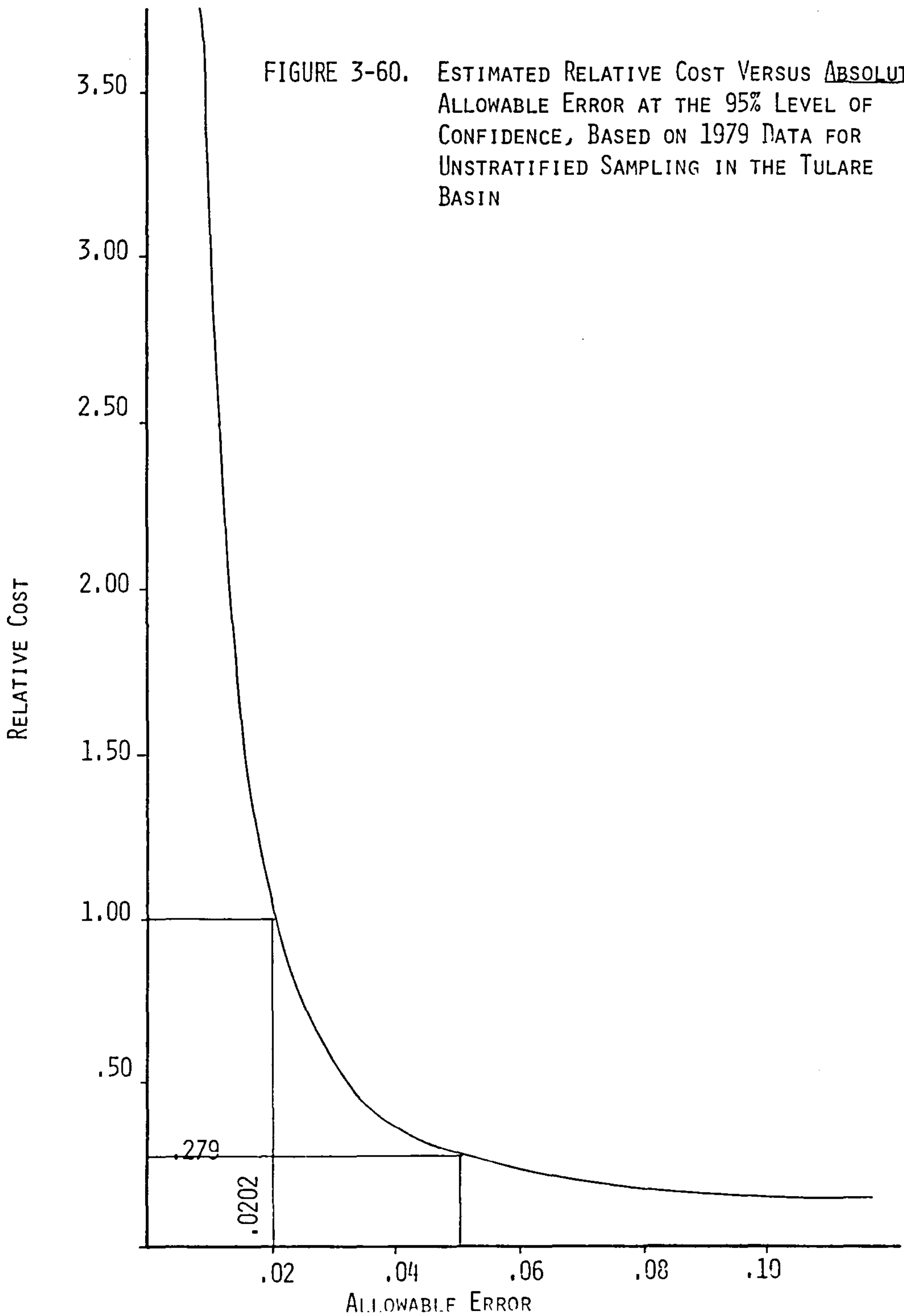


FIGURE 3-59. ESTIMATED RELATIVE COST VERSUS ABSOLUTE ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE SAN JOAQUIN BASIN

FIGURE 3-60. ESTIMATED RELATIVE COST VERSUS ABSOLUTE ALLOWABLE ERROR AT THE 95% LEVEL OF CONFIDENCE, BASED ON 1979 DATA FOR UNSTRATIFIED SAMPLING IN THE TULARE BASIN



A second factor affecting the estimate of potential cost or sample size savings is whether a stratified sample is desired. If it is, minimum strata sample requirments, homogeniety of variance within strata, correlation within strata, and population size within strata will have varying impacts on required sample size and sample distribution, and therefore on cost. In the absence of further information, the savings indicated in the figures must be considered optimistic. A third factor reinforcing this view, would be the use of error goals based on relative error. It was seen, for example, that in some basins the relative error exceeded ± 5 percent 95 times out of 100. In these cases no reduction of sample size would be possible using the design employed in 1979. Only design modifications would allow use of a smaller number of ground sample units in these basins.

For the reasons cited above, actual total variable cost and sample size reductions of 25 percent and perhaps as much as 40 to 50 percent may be possible in several basins. The particular value will depend on the design and allocation strategy employed, and whether absolute or relative error is to be used as the performance measure. It should be noted that if smaller sample unit sizes are used, sample size relative to 1979 may actually increase - yet the TVC savings suggested above may still be achieved through the relationships identified in the previous section.

Conclusions Relative to Projected Sample Sizes and TVC's

(1) Where absolute and relative errors are similar, total variable cost (TVC) can probably be reduced by 25 percent and perhaps as much as 40 to 50 percent in some basins relative to that incurred in 1979. This statement assumes a ± 5 percent, 95 times out of 100 error goal. The actual savings will depend on the sample design and allocation strategy employed.

(2) When relative error is used as the performance measure, potential TVC savings in future inventories will be less. In some basins, an increase in TVC may be necessary to achieve a ± 5 percent 95 times out of 100 error goal.

(3) Percent sample size (n) reductions similar to TVC savings should be possible under the circumstances described above for cost reduction. If sample unit size is reduced below that used in 1979, then sample size (n) may actually increase relative to 1979. However, TVC savings would still be achievable.

(4) Use of a stratified sample may reduce savings suggested above.* The amount of reduction, if any, will depend on the number of strata requiring independent allocation, their relative size, the variability among sample unit measurements of irrigated proportion within strata, and variable cost differences between strata.

---

* Though the effect of stratification was considered when adjusting the projected savings downward.

(5) Conclusions cited above depend upon the variance, correlation, and cost figures obtained in the 1979 inventory. Ground sample variance was not corrected for the effect of pooling observations from independent strata. To account for these effects, the TVC and sample size reductions reported here have been adjusted downward to obtain more conservative figures.

3.7.5 Relative Efficiency Analysis

Relative efficiency is a measure of the gain in ground sampling efficiency of a given design (in this case stratified regression with factor 5) over that of a comparable simple random sampling (SRS) design (in this case stratified SRS). Two alternative interpretations of relative efficiency (RE) are given here.

First, relative efficiency can be defined as the variance "obtained" in simple random sampling relative to that of the alternative design. This is, in fact, a measure of the "relative variances" for a given sample size. Within a stratum,

$$RE_{h_1} = \left(\frac{N_h - n_h}{N_h n_h}\right)(S_y^2)_h \div (SE_{Reg.5})_h^2 \qquad (1)$$

where

RE_{h_1} = relative sampling efficiency within stratum h for method #1,

$(S_y^2)_h$ = simple variance of ground sample unit observations of proportion irrigated within stratum h,

$(SE_{Reg.5})_h^2$ = the variance of the regression estimate for irrigated proportion in stratum h (see Equations 17 and 18 of Appendix 1B, and regression with "factor 5" in Appendix II),

N_h = total number of sample units in stratum h, and

n_h = number of sample units selected for ground measurement within stratum h.

To find the basin-wide, stratified efficiences for the first method, summation over strata is performed. That is,

$$RE_{b_1} = \frac{\sum_h W_h^2\left(\frac{N_h - n_h}{N_h n_h}\right)S_{y_h}^2}{\sum_h W_h^2(SE_{h\cdot Reg.5})^2}, \quad \text{or} \qquad (2)$$

$$RE_{b_1} = \frac{V_{SRS_b}}{V_{Reg.5_b}} .$$

A similar agregation over basins gives

$$RE_{state_1} = \frac{\sum_b W_b^2 (V_{SRS})_b}{\sum_b W_b^2 (V_{Reg.5})_b}$$

$$= \frac{\sum_b W_b^2 \left[\sum_h W_h^2 \left(\frac{N_h - n_h}{n_h N_h} \right) S_{y_h}^2 \right]_b}{\sum_b W_b^2 \left[\sum_h W_h^2 (SE_{h.Reg.5})^2 \right]_b} \qquad (3)$$

for the statewide relative efficiency by the relative variance method.

The second interpretation of relative sampling efficiency ("relative sample sizes") is a measure of the number of samples for simple random sampling to achieve a variance equal to that of the regression sample. It is found by first solving

$$\frac{N_h - n_h'}{N_n n_h'} (S_y)_h^2 = (SE_{Reg.5})_h^2$$

for n_h' (the SRS sample size in stratum h). This gives

$$n_h' = \frac{N_h (S_y^2) N_h}{N_h (SE_{Reg.5})_h^2 + (S_y^2)_h} \quad . \qquad (4)$$

Then relative efficiency on a basin basis can be defined as the sum of SRS ground sample sizes over strata divided by the sum of regression ground sample sizes over strata, both sets of sample sizes giving the same basin-wide sample variance. Thus

$$RE_{b_2} = \frac{\sum_h n_h'}{\sum_h n_h'} \qquad \text{by method \#2.}$$

The statewide relative efficiency by this second method is then given by

$$RE_{state_2} = \frac{\sum_b \sum_h n_{bh}'}{\sum_b \sum_h n_{bh}'} \quad . \qquad (5)$$

Recalling that the purpose in using weighted observations was to stabilize the regressions, there is no justification in weighting the observations in simple random sampling. For this reason, the $S_{y_h}^2$ terms in each of the above equations were evaluated using <u>unweighted</u> values, whereas the $SE_{h \cdot Reg.5}$ terms were evaluated using the weighted values.

RESULTS

The resulting basin and state estimates of efficiency are given in Table 3-39.

Table 3-39: Relative sampling efficiences using methods 1 and 2 for stratified (weighted) regression with factor 5 relative to stratified (un- (unweighted) random sampling.

Basin	RE_1	RE_2
North Coast	13.82	3.84
San Francisco	3.35	1.50
Central Coast	5.78	3.71
South Coast	2.14	1.62
Colorado Desert	8.80	3.55
South Lahonton	2.39	1.70
North Lahonton	5.44	1.76
Sacramento River	6.60	5.04
San Joaquin	2.32	2.18
Tulare	6.79	5.62
State	5.33	3.16

The consistency of RE_2 in Table 3-39 to be less than RE_1, can be explained by the fact that as n' approaches N, a larger proportion of the variance for simple random sampling (within strata) is accounted for by the finite population correction. RE_2, therefore, will be dependent on the number of sample units in the population. If the total number of sample units is increased by decreasing the relative area within each, then RE_2 will tend upward toward the basin RE_1 value. The value of RE_2 is that it gives* the ground sample size necessary to achieve the same variance as the regression design for the given population. It also can be used directly to compute the sample size-dependent cost associated with the stratified simple random sampling design.

Inspection of the values in Table 3-39 shows that RE_1 ranged from a low of 2.14 in the South Coast to a high of 13.82 in the North Coast. In two of the three major agricultural basins, RE_1 exceeded 6.5 (6.60 in Sacramento River and 6.79 in Tulare). In contrast, an RE_1 of only 2.32 was achieved for the San Joaquin basin. This lower RE_1 resulted from a tight clustering of paired ground and Landsat measurements of irrigated proportion within each land use stratum. The computed variance between ground observations ($(S_y)_h$) was therefore low (especially so in

*by multiplying RE_2 times the ground sample size required for regression two strata), giving a stratified simple random sample ground variance only 2.32 times as large as the corresponding regression variance. As noted above, RE_2 tended to be lower than corresponding RE_1 figures, though not substantially in the three prime agricultural basins.

Statewide, RE_1 for the stratified design was computed to be 5.33. That is, the variance reduction was over five times when using regression as opposed to ground sampling only in a stratified design. The value for RE_2 shows that, statewide, more than three times the number of ground units would have to have been measured in a ground-only design to achieve the same sample variance as the Landsat regression design, given the sample frame used and the correlations achieved in 1979.

Appendix III presents RE_1 and RE_2 values for the unstratified case. These tend to be higher than the corresponding stratified values due to the larger variance among ground observations when stratification is not used. The unstratified relative efficiencies should be used for comparison only. Stratified values for RE_1 and RE_2 reported in Table 3-39 are considered to represent the baseline of relative efficiency for the Landsat regression design established in the 1979 irrigated land inventory.

3.8 Task I Recommendations

The results and experience gained in the Irrigated Lands Applications Pilot Test have enabled the formulation of a number of recommendations regarding future operational implementation. These recommendations were based on a review of irrigated area estimates produced by the 1979 inventory, findings from the evaluation, and observations made during the course of planning and conduct of the inventory. These recommendations should not be viewed as final. A more detailed proposal for operational implementation will be forthcoming in the coming year with publication of the Procedural Manual. Recommendations are listed below according to sampling, measurement, and estimation design components.

3.8.1 Sample Frame

(1) Size of Sample Unit:

(a) 1.0 to 1.5 mi^2 in predominantly agricultural areas dominated by field crops, orchards, and vineyards (typically valley floor and terrace lands);

(b) 2.0 to 2.5 mi^2 in other areas.

(2) Shape of Sample Unit:

Rectangular, to include if necessary non-agricultural land.

Rectangular or simple polygon shape designed to minimize sample unit registration, handling, and location problems.

(3) Orientation of Sample Unit:

Aligned with road networks, field boundaries, and other 'natural' boundaries; this should minimize ground access problems, field time, labelling

(4) Location of Sample Unit:

Each sample unit should be constrained to fall on preferably one and not more than two USGS 7½ minute quadrangles.

(5) Number of Land Use Strata:

(a) Retain the small grain and vegetable strata for identifying sample units in which multiple cropping in individual fields is likely to occur during the course of a given calendar year;

(b) Reduce the number of land use strata to which independent allocation of ground samples is required. In some basins (e.g. desert hydrologic units) no stratification may be required. In others, a stratification scheme separating dryland areas, field crop areas, orchard/vineyard areas, agriculture-urban mix areas (formerly 'exclusion' areas), and dispersed agricultural areas may lower estimated basin sampling error. Not all basins will contain all strata. Until further experience is gained, retention of these general strata in most non-desert basins is recommended for maximum design flexibility in future inventories.

(6) Urban Fringe Exclusion Areas:

Include more urban fringe (mixed urban and small-field agriculture) within the sample frame, possibly as a separate stratum. Options include using a simple random ground sample to estimate irrigated acreage in this area; or using the regression technique employed in the 1979 inventory; or using the USDA estimate for this area; or possibly, though this would be the most expensive alternative, performing a complete ground enumeration of land within this present exclusion area.

(7) Agricultural Land Outside the Sample Frame:

Include agricultural land designated as 'outside the sample frame' in 1979 within a future sample frame. Most of this land would probably be assigned to (old) stratum 7, though a smaller portion may be classified to stratum 1 (dryland). Areas assigned to stratum 7 would be subject to the same regression technique for estimation of irrigated land as applied in 1979. Land assigned to the dryland stratum will necessarily be subject to the sampling and estimation technique selected for that stratum. Equal probability selection of ground sample units for use in regression estimation of irrigated area, or variable probability selection of sample units with probability proportional to estimated size (ppes) estimation are the two most likely sampling/estimation candidates for the dryland stratum.

Use of USDA estimates for areas outside the current (1979) irrigated lands sample frame should be evaluated as an alternative solution to this problem. The disadvantage to this option may be a relatively high sampling error for irrigated land in this area.

(8) County and Basin Boundaries

Maintain 1979 system of county and basin boundaries, but recheck county and basin boundaries for possible errors in placement.

3.8.2 Sample Allocation

(1) Early System Use

If it is necessary to use the Landsat-aided APT technique for estimation of irrigated land area before a revised sample frame, stratification, or sample unit list can be constructed, then it is recommended that the sample units selected for ground measurement in 1979 be used again. This should be considered a 'one time only' approach to sample allocation in an operational application. Strategies for selection of new sample units (described below) should be employed when a revised sample frame becomes available.

(2) Proportional Allocation

(a) Once a revised sample frame is available, it is recommended that allocation of sample units to land use strata be proportional to the relative size of each stratum. This procedure represents a simplification of the 1979 system. The relative number of sample units allocated to a given stratum in 1979 was directly proportional to the product of stratum size (actually total number of sample units available) and stratum standard error, and inversely proportional to the square root of measurement cost for individual sample units. This technique, known as optimal allocation, was designed to give the smallest sampling error for number of ground units sampled.

By using proportional allocation instead, the problem of computing sample size for each stratum is simplified considerably (in terms of calculations and, in this case, software required), understandability of the technique is enhanced, and flexibility in using the given sample allocation for estimation of other parameters (e.g. small grain acreage) is increased. The 'cost' of using proportional as opposed to optimal allocation to strata may be an increase in sampling error for some basins. We expect this 'cost' to be more than offset by the simplicity of use during the early application of this system. As experience and cost/variance data are gained over repeated applications of this inventory procedure, use of optimal sample allocation may, in DWR's view, become a more attractive option.

(b) Total basin sample size can be computed by simple formula for the proportional allocation case once the size and sampling variance for each stratum are known. The size of a stratum will be available from digitizing stratum area or, if sample units are of roughly equal size, from a count of sample units within the given stratum of the basin in question. An estimate of regression sample variance can be obtained by classifying old (e.g. 1979) sample units into the new strata and recomputing (1) the variance among sample units falling in each stratum and (2) the Landsat-to-ground correlation for each stratum.

The sample size to be allocated to each stratum can be obtained directly by multiplying the total basin sample size times the proportion of the basin sample frame (alternatively the proportion of sample units) within the given stratum. A minimum limit of ten ground-measured sample units per stratum is recommended to insure the development of reasonably meaningful Landsat-to-ground regression equations.

(c) It is recommended that selection of sample units for ground-measurement in each stratum be random, equal probability, without replacement.

(3) Partial Replacement Sampling

A procedure for replacing only some ground sample units at each new inventory should be considered for implementation in future surveys. This would reduce costs of establishing new ground sample units and would likely reduce overall sampling error. Replacing 25 percent of the sample units used for ground measurement at each inventory might, for example, reduce sampling error 10 to 25 percent depending on year-to-year correlation between sample units, variance, and other assumptions. In addition, error in estimation of difference in irrigated acreage between inventories might be reduced significantly (see, for example, discussion by Cochran 1977).

3.8.3 Landsat Enlargement and Measurement Recommendations

(1) Landsat Enlargement

(a) The base date must be enlarged accurately to the specified scale. Additional dates need not be scale matched as critically.

(b) Need to investigate the use of 1:100,000 and/or 1:250,000 scale enlargements as the base.

- The USGS is now or will be producing maps at both scales. Having Landsat and maps at the same scale would considerably reduce the time spent on merging strata, locating sample units, etc.

- For 1:100,000 scale, this would require a once only redo of boundaries and sample frame

(c) Need to investigate alternative enlargement methods.

(2) Landsat Recording Forms

(a) Preparation of county boundaries by an automatic plotter was done in the 1979 survey. In-house plotting capability at DWR could streamline the process.

(b) Automatic plotting of the 7.5' grid would also greatly reduce preparation time.

(3) Interpretation Procedure

(a) Using analysts familiar with the area being interpreted and the agricultural practices common to the area should reduce analyst time and errors.

(b) Recording the time spent to interpret an area should be done on a sample basis.

(4) Digitization

(a) When locating the sample units on Landsat prior to digitization, the units must be located with reference to the 7.5' ground data maps.

(b) Having regular boundaries on the sample units and strata boundaries would greatly reduce digitizing time.

(c) If more than one digitizer is being used, every attempt should be made to use common measurement systems (i.e. 1000ths of an inch).

(d) Appropriate software should be made available to maximize the efficiency of digitizing.

3.8.4 Estimation Equations for Irrigated Land

(1) Regression Estimator

The stratified regression estimator is recommended for estimation of irrigated land in an operational system. This recommendation is based on the following considerations:

(a) it is fairly simple to understand and to apply;

(b) it was one of the best 'performers' (tended to give smaller confidence interval half-widths) in the evaluation of linear estimators; and

(c) it appears to be reasonably robust against deviations from linear model assumptions experienced in the 1979 inventory.

(2) Sample Unit Weighting

Omit weighting of irrigated proportion observations by sample unit size unless sample units differ significantly in area.

(3) Variance Expression

Use of the expression for regression variance given by Equations 17 and 18 in Appendix IB (denoted as the variance expression with factor 5 in Appendix II) is recommended. This formulation of regression variance makes the least assumptions about the distribution of Landsat measurements of irrigated proportion in the basin population of sample units.

4.0 ESTIMATION AND MAPPING OF IRRIGATED LAND USING DIGITAL ANALYSIS TECHNIQUES (TASK II)

The development of digital analysis techniques for inventorying irrigated land continued during 1980. The information requirements for Task II were similar to those used in the manual analysis task (Table 2-1). Primary emphasis was still on the estimation of irrigated land; however, mapping irrigated land through the classification of digital Landsat data became an important secondary objective. The error goal for the estimation procedure was again set at ± 5% for the 95% level of confidence on a basin basis. Two major sub-tasks were addressed:

(1) the development of a streamlined, precise method of full-frame registration for use with multitemporal Landsat digital data; and

(2) the continued development, testing and evaluation of the MSS band 7-to-MSS band 5 ratio as an accurate discriminant for separating irrigated from non-irrigated land in the major agricultural areas of California. For the second sub-task three test sites were selected for analysis: the Tulare Basin in the southern San Joaquin Valley; Sacramento County in the heart of the Central Valley; and, the Sacramento Basin in the northern Sacramento Valley (Figure 2-1).

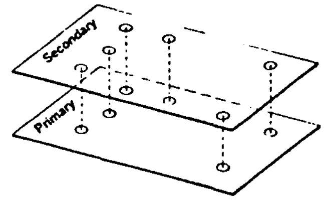
4.1 Registration of Multitemporal Landsat Digital Data - NASA/AMES Research Center

As in the manual analysis procedure, the long growing season in California necessitates the use of multiple dates of Landsat data to insure the detection and identification of irrigated land. Precise registration of Landsat digital data for a number of dates, as well as mosaicking adjoining path and row acquisitions together, is a prerequisite for successful classification and output of results. Both the Tulare and Sacramento Valley Hydrologic Basins occupy portions of two Landsat scenes (Figure 2-1). For both of these basins three dates of Landsat full frame data were registered to each other. Figure 4-1 represents 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 overflight dates are manipulated to overlay those of the primary scene. Any computer compatible tapes may be chosen as the primary scene.

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 64 x 64 pixels in area, and secondary 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

1. ESTABLISH INITIAL CORRESPONDENCE
   - ≈ 10 Corresponding Points: Common points on each image to obtain initial overlay

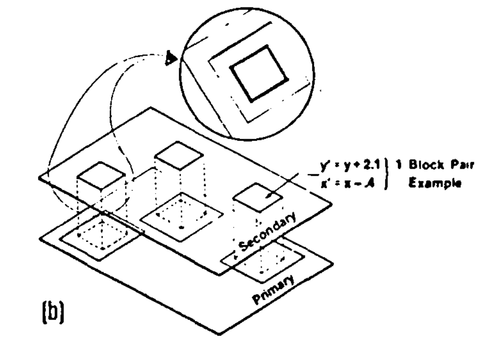
(MANUAL OPERATION)



(a)

2. PERFORM BLOCK CORRELATION
   - 340 Pairs of Corresponding Blocks from primary and secondary scene are evaluated
     - Primary Scene blocks are 64x64 Pixels in area
     - Secondary Scene blocks are 32x32 Pixels in area
   - 1089 Correlations/Block pair: Small window is moved around on large window until best fit is determined
   - LaGrange Interpolation for Exact Peak: Subpixel movements determined and residuals listed for each block pair

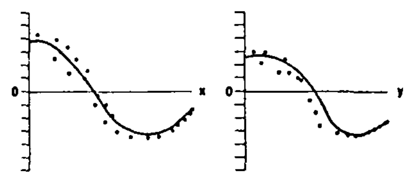
(AUTOMATED)



(b)

3. EDIT BLOCKS
   - Remove Outliers: Because of clouds, etc. some block pairs are removed from analysis
     ≈ 140 Blocks Remain
   - Derive Least Squares 3rd Order Polynomial based on remaining residuals
     Max Error ≈.3 Pixel
     RMS Error ≈.2 Pixel

(ITERACTIVE)

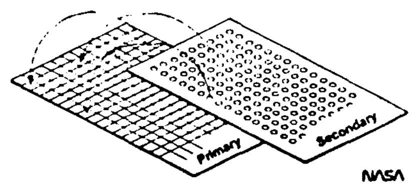


$$y' = A + By + Cx + Dy^2 + Exy + Fx^2 + Gy^3 + Hy^2x + Ix^2y + Jx^3$$
$$x' = K + Ly + Mx + Ny^2 + Oxy + Px^2 + Qy^3 + Ry^2x + Sx^2y + Tx^3$$

(c)

4. "WARP" SECONDARY SCENE TO PRIMARY SCENE
   - Evaluate 3rd Order Polynomial on Each Primary Pixel
   - Map Nearest Neighbor Secondary Pixel: Secondary Scene Pixels are moved to Primary Scene Location
   - New Tape of Secondary Scene created with subpixel accuracy

(AUTOMATED)



(d)

NASA

Figure 4-1. Scene-to-scene registration procedure

LaGrange interpolation at the maxima of this correlation surface to obtain the correspondence for this block pair to sub-pixel 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 attained. (Within 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 calibrated to a map base; however, when the primary scene is calibrated, all the overlaid secondary scenes also become calibrated. Upon completion of the registration, a band of Landsat MSS 7-to-MSS 5 ratioed data in addition to the four Landsat raw data bands was computed.

4.2 Classification of Multitemporal Landsat Digital Data

The second major sub-task of Task II was to continue the development, testing and evaluation of the MSS 7/MSS 5 ratio as a viable discriminant for differentiating irrigated from non-irrigated land in the major agricultural areas of the state. In support of this, each cooperating institution worked on a separate part of the Central Valley: UC Santa Barbara studied the Tulare Basin area in the southern part of the Central Valley; NASA/Ames tested the techniques in Sacramento County - an area located in the center of the Valley and encompassing a portion of the Sacramento-San Joaquin River delta; and, UC Berkeley continued development in the northern (Sacramento Valley) section of the Central Valley.

4.2.1 Tulare Basin Test Site - University of California/ Santa Barbara Campus

Digital analysis of three Landsat dates for the 1979 growing season indicates that comparable results can be obtained with either digital or manual interpretation techniques. Using the estimates of irrigated acreage for a portion of the Tulare Basin (Kings, Tulare, and part of Kern Counties) obtained from Task I as the standard of comparison, a digital classification of the basin was undertaken using a simple band 7/band 5 ratio enhancement with a threshold cutoff.

The purpose of this work was to define a procedure for identifying irrigated acreage using digital techniques that would mimic the procedure developed in Task I. Because the definition of irrigated land is essentially a binary one (red/not red) the digital approach is well suited to this task.

The enhanced scenes were created using three dates of Landsat that had been registered and 7/5-ratioed by NASA-Ames. The procedures used were similar to those developed in earlier work on the 3-quad Kern County study site. The scenes were displayed and a cutoff value was applied. All pixels below the cutoff value were considered to be non-vegetated while all those above it had vegetative cover. Different cutoffs were tried until the analyst felt that the scene on the monitor matched the photograph of the

Landsat image. Generally the critical factor for determining the appropriate cutoff is the identification of that pixel value at which optimal field definition occurs (i.e., the interior of a vegetated field is made up of pixels above the cutoff with few or no pixels below it). Usually some trade-off occurs because as fields "fill in," scattered, isolated pixels begin to show-up as vegetated. The end result is a compromise that results in a product that is visually satisfying to the individual analyst. By displaying Landsat bands 4 and 5 in blue and green, respectively, and the 7/5 ratio band in red, an enhanced color-IR scene was created. The cutoff was then applied to the red channel.

The 256 x 256 image display monitor will hold an area somewhat larger than a 7.5' quad. A number of these subscenes were displayed and a cutoff selected for each of the three counties analyzed - Kern, Kings, and Tulare.

As a first attempt, the boundaries for each of the three counties were digitized, registered, and overlayed onto the Landsat scenes. Also included in the digitized overlay were the major agricultural/non-agricultural boundaries and urban boundaries (Figure 4-2). These were used to mask large areas of non-agricultural activity so that any pixels having values above the selected 7/5 ratio would not be erroneously classified as agriculture.

The acreage for each county was determined using VICAR to assign the appropriate 7/5 ratio cutoff value to all pixels within each of the county boundaries. Tabulation of the results was accomplished by setting all pixels on the overlay mask (Figure 4-3) to zero, except the county in question, in which values were set to 1. The mask and the classified Landsat scene were multiplied together to create a third scene containing only the particular county. Acreage was then computed from the resulting pixel count of all pixels classified as irrigated.

The entire agricultural area of Kings and Tulare Counties was contained in each of the three scenes but the southernmost postion of Kern County was outside of the Landsat frame. Rather than incur the costs of three additional Landsat tapes and the associates registration costs, we elected to use the acreage measurements for this portion of Kern County obtained during Task I. Table 4-1 shows the acreage measurements for Kings, Tulare, and Kern Counties. By way of comparison figures are also included for the measured acreage (bias corrected acreages are shown in parentheses) obtained during Task I, as well as an estimate of irrigated acreage provided by DWR. This latter figure is extrapolated from the most recent DWR land use survey (in some cases over five years old) and current information from the county agricultural commissioner. Historically, this update procedure has resulted in estimates within two percent of actual conditions.

Table 4-1 compares the digital classification results with both the Task I measurement and DWR's estimate. Also shown is a comparison of Task I results with the DWR estimate. In all cases the digital classification resulted in a smaller number of measured irrigated acres indicating that the selection of a lower 7/5 ratio cutoff may be appropriate. As the "correct" threshold value is approached, the analyst increasingly cues on the presence of speckle, resulting in a conservative bias. The implementation of some type of post-

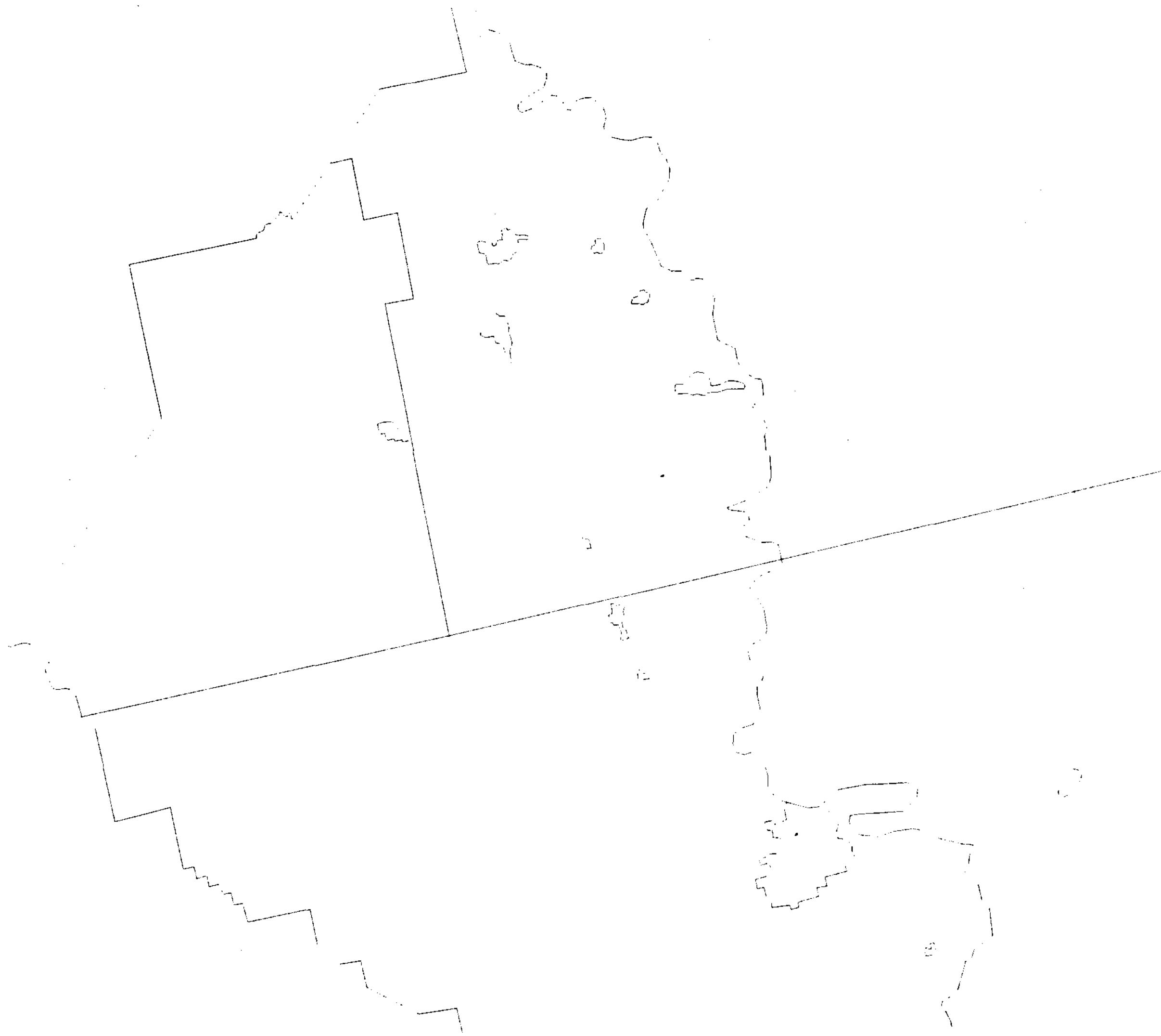


Figure 4-2. County, Urban and Foothill Boundaries Digitized on IBIS

Figure 4-3. Digital Mask Created on VICAR/IBIS

Table 4-1 Comparison of Total County Acreage Estimates Between DWR and Manual and Digital Landsat Measurements

County	DWR Estimate (1000 Acres)	Task I - Manual (1000 Acres)	Task II - Digital (1000 Acres)
KINGS	613.7	568.5	547.7
TULARE	710.9	792.9	759.3
KERN	991.5	1006.2	976.7

classification filtering could be used to remove speckle, allowing the selection of a more liberal cutoff value. During 1981, further work will be done to test various cutoffs and perhaps define stratification criteria that can be used to control error.

Both the manual and digital interpretations had errors ranging from 1.5 to approximately 11 percent. Both procedures did quite well in Kern County (+1.5% and -1.5% for the manual and digital approach, respectively). The error ranged from 6.8% to 11.5% for Kings and Tulare Counties with the magnitude of the error "flip-flopping" between the manual and digital procedures. The nature of the misclassification error will be covered in the Task I evaluation. Further study is needed to see why the magnitude of the error flip-flops between the two procedures - possibly a county boundary separating the two counties is misplaced. We have reviewed the manual interpretation of Kings County and find it difficult to account for up to 7.4% more irrigated land.

During 1981, these questions will be addressed as we attempt to fine-tune the digital classification procedure. We will also register the 1 x 5 mile sample units from Task I allowing for a full comparison with the Task I effort.

4.2.2 Sacramento County Test Site - NASA/Ames Research Center

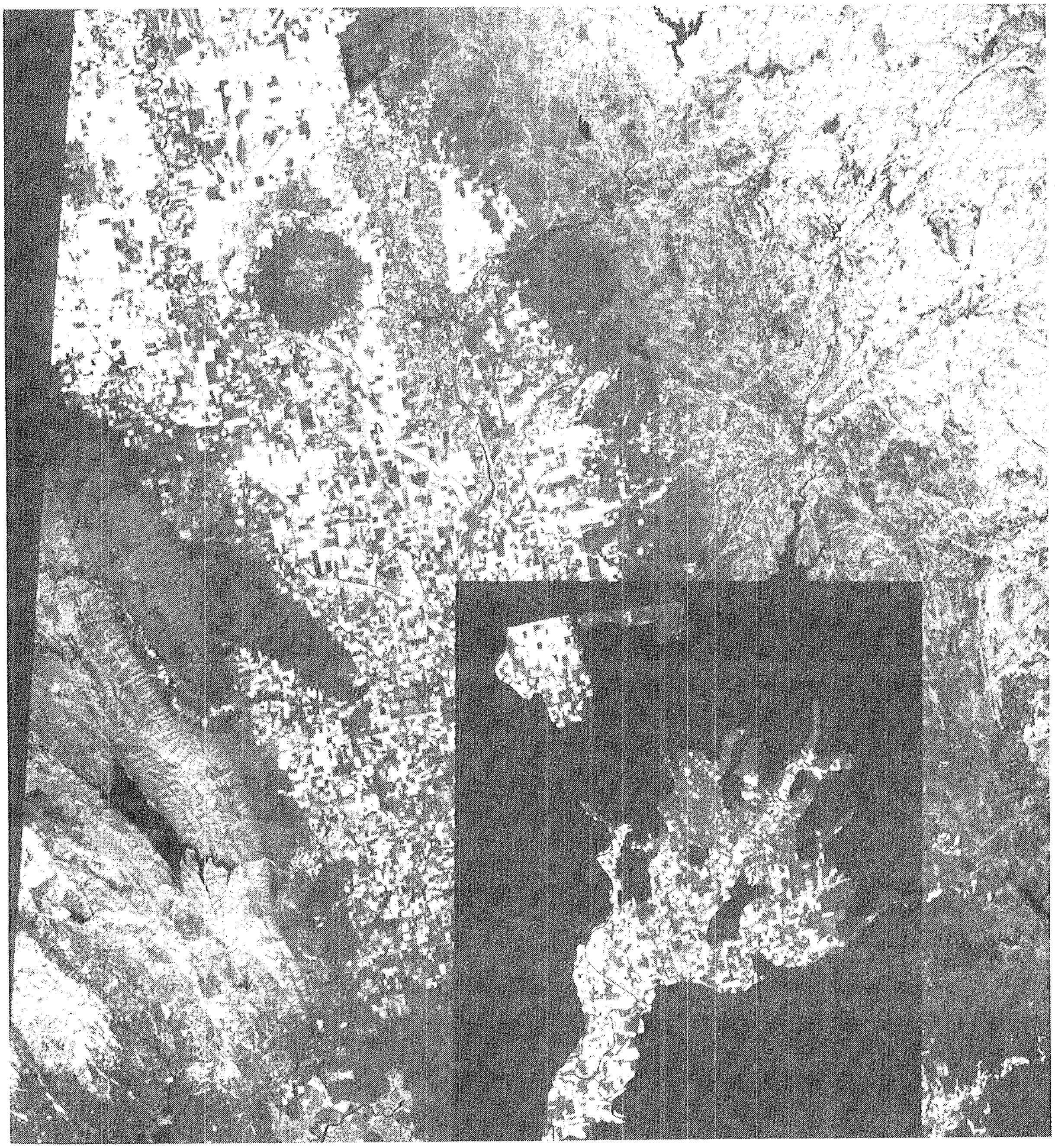
Development and testing of the MSS 7/MSS 5 ratioing technique continued with an experiment conducted in Sacramento County (Figures 2-1 and 4-4). Sacramento County is located between the two test sites being studied by the University of California groups and offered the opportunity to confirm the validity of the ratioing technique in the central part of the valley and in the agriculturally productive Sacramento River delta area. For this test site four dates of 1979 Landsat data were registered (as described in Section 4.1) and 7/5 ratio bands created. The dates selected for study were: June 11, July 8, July 26, and September 18.

As in the work done in the Tulare and Sacramento Valley test sites, the 7/5 ratio images were read into an interactive display and analysis system (IDIMS at Ames) and individually looked at. By referring to a color composite photographic product the analyst scrutinized individual fields on the basis of their red color on the composite and the amount of "shade-in" shown on the television monitor of the IDIMS system. The threshold value of the "shade-in" was adjusted by the analyst and pixels with values occurring below the threshold value were considered non-irrigated. The threshold values arrived at varied from date to date in the following way:

Table 4-2. MSS 7/MSS 5 ratio threshold values used for classifying irrigated land in Sacramento County

	RATIO VALUE
June 11	1.70
July 8	1.70
July 26	2.00
Sept 18	1.40

Figure 4-4. July 8, 1979 full frame 7/5 ratio greyscale image with the Sacramento County agricultural strata overlay. The stratification was the same as used in the Task I manual analysis (Section 3.2).



After the threshold values were determined a pseudo-binary image was made for each date. The mapping scheme used to prepare the binary image allowed the analyst to determine both (1) that any given field had been above the threshold on one of the dates, and (2) on which date or combination of dates that area was above the threshold. Therefore, in the creation of the pseudo-binary images each pixel was assigned the value of 2, 4, 8, or 16 if it was above the threshold value on one or more of the four Landsat overpasses. Given the values assigned in the mapping function there exists the possibility of 16 different values; the summed value allows you to determine the dates when the pixel was above the 7/5 ratio threshold. Table 4-3 shows the sixteen possible values and how they were derived.

Table 4-3. The sixteen possible summed values used to describe the irrigated/non-irrigated sequence in Sacramento County.

June 11	July 8	July 26	Sept 18	SUM
Y(2)	Y(4)	Y(8)	Y(16)	30
Y(2)	Y(4)	Y(8)	N(1)	15
Y(2)	Y(4)	N(1)	N(1)	8
Y(2)	N(1)	N(1)	N(1)	5
N(1)	N(1)	N(1)	N(1)	4
N(1)	Y(4)	N(1)	N(1)	7
N(1)	Y(4)	Y(8)	N(1)	14
N(1)	Y(4)	Y(8)	Y(16)	29
N(1)	N(1)	Y(8)	N(1)	11
N(1)	N(1)	Y(8)	Y(16)	26
N(1)	N(1)	N(1)	Y(16)	19
Y(2)	N(1)	N(1)	Y(16)	20
Y(2)	N(1)	Y(8)	N(1)	12
N(1)	Y(4)	N(1)	Y(16)	22
Y(2)	N(1)	Y(8)	Y(16)	27
Y(2)	Y(4)	N(1)	Y(16)	23

Y=Irrigated, N=Non-irrigated

The number of pixels occurring in each of the sixteen possible classes was then counted. Of the 1,593,102 pixels counted, 487 pixels were classified with values other than the sixteen described above. Upon investigation, these 487 pixels were found to be associated with riparian areas and were subsequently reclassified into the never irrigated class (#4). Figure 4-5 shows the Sacramento County area with all pixels irrigated at least once shown in white and pixels in stratified-out areas shown in black. Table 4-4 shows the resulting pixel counts and subsequent acreage values for the four dates used.

Table 4-4. Pixel counts and acreage values for the Sacramento County test site.

PIXEL VALUE	GREEN DATE	PIXCOUNT	ACREAGE pixels X .81
4		158,184	
5	June 11	5,437	4,403.97
7	July 8	8,045	6,516.45
8	June 11, July 8	6,522	5,282.82
11	July 26	9,064	7,341.84
12	June 11, July 26	744	602.64
14	July 8, July 26	13,876	11,239.56
15	June 11, July 8, July 26	15,033	12,176.73
19	Sept 18	22,924	18,568.44
20	June 11, Sept 18	3,032	2,455.92
22	July 8, Sept 18	14,164	11,472.84
23	June 11, July 8, Sept 18	17,220	13,948.20
26	July 26, Sept 18	11,478	9,297.18
27	June 11, July 26, Sept 18	1,965	1,591.65
29	July 8, July 26, Sept 18	46,829	37,931.49
30	all four dates	51,487	41,704.47

subtotal	184,534.20
acres outside ag strata	1,557.00
number acres on path/row 4734	10,040.76
total irrigated acres	196,131.96

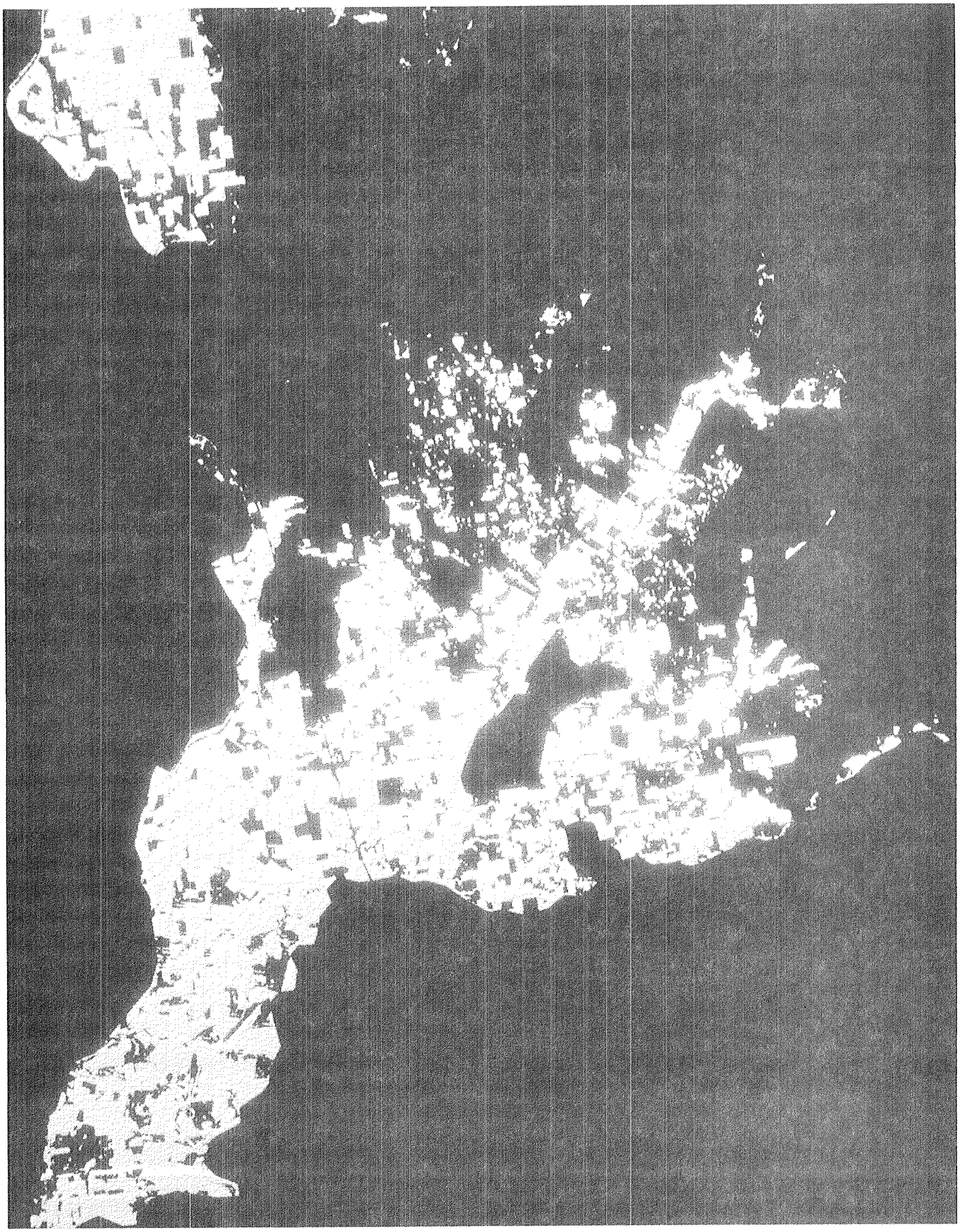


Figure 4-5. Total irrigated pixels in the Sacramento County test site. Black = pixels never classified as irrigated, white = pixels classified as irrigated on at least one of the four dates.

The results of this test are very promising. The Department of Water Resources estimated that there were 196,700 acres of irrigated land in Sacramento County. This estimate differs from that classified as irrigated using the 7/5 ratio technique by approximate 570 acres or .3%.

4.2.3 Sacramento Valley Test Site - University of California/ Berkeley Campus

Development and testing of digital analysis techniques continued in the Sacramento Valley in 1980. The primary task was to perform an inventory of the Sacramento Hydrologic Basin (See Figure 4-6). Digital data from 1979 was used along with the 1979 Task I ground sample units (55 sample units). Significant progress was made on several sub-tasks: (1) determining the optimum sample design minimizing cost for a fixed error; (2) implementing the Survey Planning Model Simulation technique in sample design analysis; and (3) refining the technique for setting the irrigated/non-irrigated discriminant in the greeness indicator band.

Registration of Multi-Temporal Landsat Data

As in Task I, the detection and identification of irrigated land in California necessitates the use of multiple dates of Landsat. Therefore precise registration of multitemporal Landsat digital data is necessary for accurate classification. Since DWR will ultimately require summarization and output in the form of U.S.G.S. 7.5 quadrangles, the registration of Landsat data to that map base is also desirable. The three dates of Landsat used in Task I (June 11, July 8, and Sept. 18) for two scenes (Path 47, Row 33 and Path 48, Row 32) were registered to each other, as described in Section 4.1. This procedure does not include registration to a map base. Upon completion of the registration, NASA/Ames computed a band of Landsat MSS 7 to MSS 5 ratioed data in addition to the four Landsat raw data bands. The registered raw data and 7/5 ratioed bands were then provided to the University for further analysis.

The date-registered data were next registered to the U.S.G.S. 7.5 minute map base. As a first step in this process the date-registered Landsat scenes were displayed on the RSRP interactive image analysis system and were divided into seven blocks for ease of storage, display, and analysis; each block was 30 minutes of longitude by 30 minutes of latitude in size (see Figure 4-6). A multidate, Landsat data file was then created on a computer disk for each of the 30 minute blocks.

Next, a set of control points was selected to initiate registration of the multitemporal data set to the map base. One set of control points was used for each 30' block. These points were distributed as evenly as possible with approximately three points per 7.5 minute quadrangle. Control point coordinates were obtained by displaying the base date for each block, moving the cursor on the TV monitor to the selected point, and recording the x and y coordinates. Control points were selected based on (1) the ease with which they could be located on the Landsat base date and the U.S.G.S. 7.5 minute quadrangle maps, and on (2) the degree to which they contributed to an approximately even distribution of points over the 30 minute block. The x and y coordinates of each

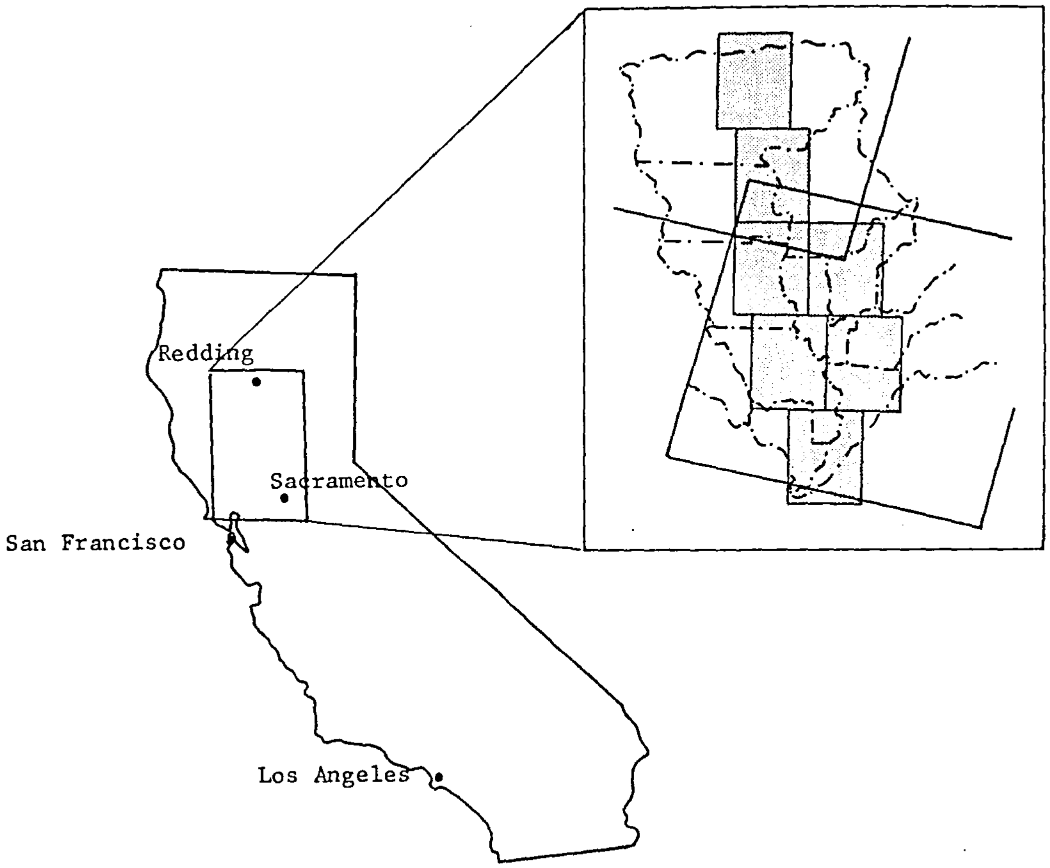


Figure 4-6. Location of 30 minute blocks in the Sacramento Valley

control point were then measured on the 7.5 minute quadrangles in 1/60 inch increments. Measurements were made using the upper left corner of each 7.5 minute quadrangle as the origin.

Dimensions for the 30 minute block computer file were set at 640 points by 800 lines to (1) give a map cell size of approximately 0.5 hectare (1.2 A), and to (2) allow division of each block into various sample unit sizes for design analysis. These dimensions were then used to convert the map coordinates (in inches) to ground computer file coordinates using the following formulas:

$$X_{nf} = \frac{X_m \times 160}{W} + (N \times 160)$$

$$Y_{nf} = \frac{Y_m \times 200}{L} + (M \times 200)$$

where

X_{nf} = X value in new file

X_m = X value on the map

Y_{nf} = Y value in new file

Y_m = Y value on the map

W = 7.5' map width in inches

L = 7.5' map length in inches

N = 0,1,2, or 3 depending on whether the 7.5' map is the first, second, third or fourth from the west side of the 30' block

M = 0,1,2,3 depending on whether the 7.5' map is the first, second, third or fourth map from the north side of the 30' block

The control point coordinates for the Landsat data and the new ground file were run through the regression program DANIEL. This program calculated the equations necessary to transform the Landsat data to the new ground coordinate file. These equations were of the form:

$$X_{Landsat} = b_0 + b_1X_G + b_2X_G^2 + b_3y_G + b_4y_G^2 + b_5X_Gy_G$$

and

$$y_{Landsat} = b_6 + b_7X_G + b_8X_G^2 + b_9y_G + b_{10}y_G^2 + b_{11}X_Gy_G$$

where x_G and y_G are the new ground file coordinates.

The equations from DANIEL were used in the program COTRANS to place the Landsat data into the new ground file. This program resampled the data by using the DANIEL equations and the coordinates for each new file cell to predict the corresponding location in the original Landsat file. The data values for that Landsat pixel were then transferred to the cell in the new file. This was done for the 7/5 ratio bands for each date, the end product being a file 640 points by 800 lines with three bands: June, July, and September. The data had been rotated so that a particular cell represented the same point on the ground for all three dates.

Ancillary Data

To facilitate data summary and calibration it was necessary to prepare a multi-layered data base. The data base contained the Landsat digital data for three dates as well as county boundaries, the hydrologic basin boundary, the land use stratification used in Task I, and the Task I ground sample unit boundaries. (See Figure 4-7), All boundaries were digitized and were transformed to overlay the registered Landsat data.

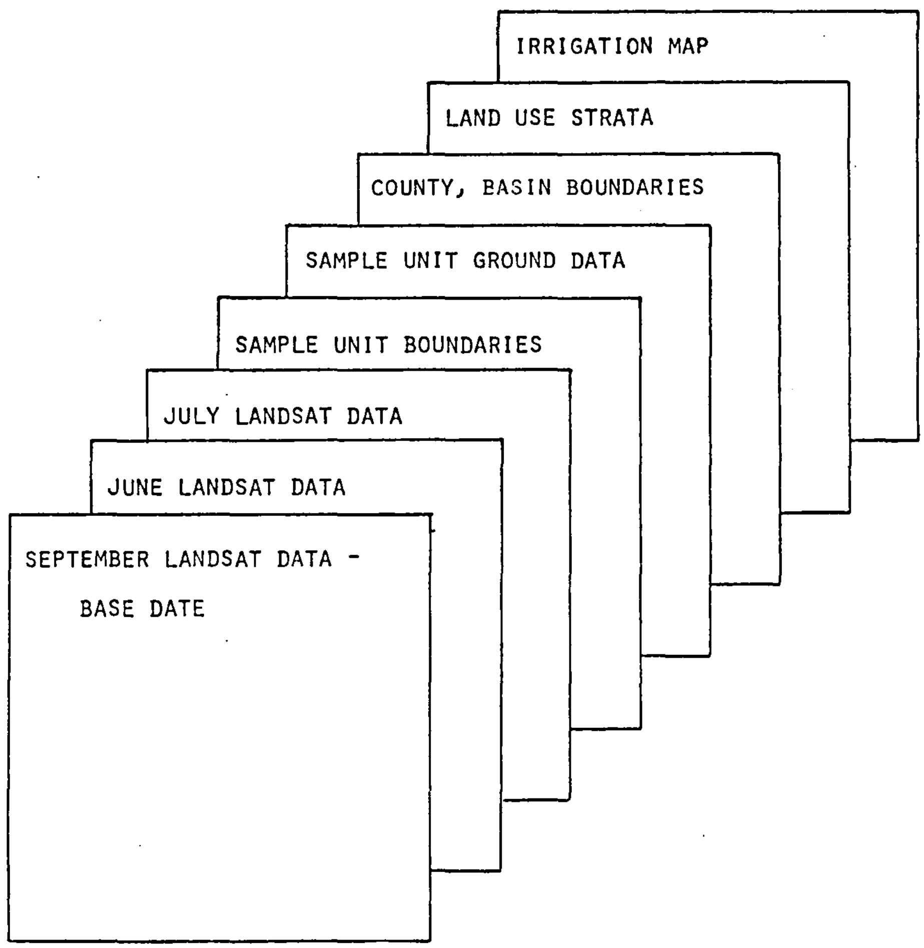


Figure 4-7. Registered Data for Each 30 Minute Block

Classification

To identify irrigated land, a simple vegetative indicator was used. This indicator consisted of the ratio of Landsat MSS band 7 to MSS band 5 (7/5 ratio). Since actively growing vegetation generally has a higher 7/5 ratio than other cover classes, agricultural land, with its healthy vegetation and high percent of canopy cover, should have a higher 7/5 ratio than native vegetation or fallow land. A threshold 7/5 value can be determined to separate irrigated agricultural land from non-irrigated land.

The three dates of Landsat 7/5 ratioed data selected to identify irrigated land corresponded to those used in Task I. The late summer date was selected as the base date since in California, as in most arid or semi-arid areas, only irrigated crops are actively growing during summer. A late Spring date was chosen to monitor small grains and a Fall date was used to detect multiple-cropped fields.

The 7/5 bands for each of the three dates were analyzed by 30 minute block, and a threshold value was selected to separate irrigated from non-irrigated acreage. It was expected that the 7/5 threshold value would vary by date and ground location of the 30 minute block due to: (1) changes in the condition of annual grasslands bordering the area since grasslands are one of the main non-irrigated cover types to be eliminated; (2) changes in type and proportion of crops grown since each crop has unique spectral characteristics; and (3) shifts in crop calendars due to climatic and latitudinal variations since the phenological stage of a crop affects its spectral appearance. Using the RSRP interactive image display system, each 30 minute block was displayed and analyzed separately.

To set the threshold value for a given band (date) the data was displayed on the TV monitor and was compared to the Landsat 1:1,000,000 color composite transparency. Using a real time masking option, file cells with values below a specified 7/5 value were masked out. This 7/5 value was adjusted visually until only what appeared to be actively growing cropland on the transparency was displayed. This value was then used as the threshold value for that date. (See Figure 4-8)

When threshold values had been determined for each date, an irrigation class map was created for each 30 minute block. For a given date, the 7/5 ratio of each cell was compared to the selected threshold value and was labeled as irrigated if its value was greater than the threshold. After each pixel was labeled irrigated or not on all three dates, the bands were combined to create a class map.* The three date pattern of irrigation for each pixel was then labeled as one of eight classes (see Figure 4-9):

---

* This classification technique was developed from an earlier procedure reported by Hay et al (1977, pp 2-8 to 2-33) in which crop group stratification was obtained by using a constant threshold on several dates of Landsat 7/5 ratio data.

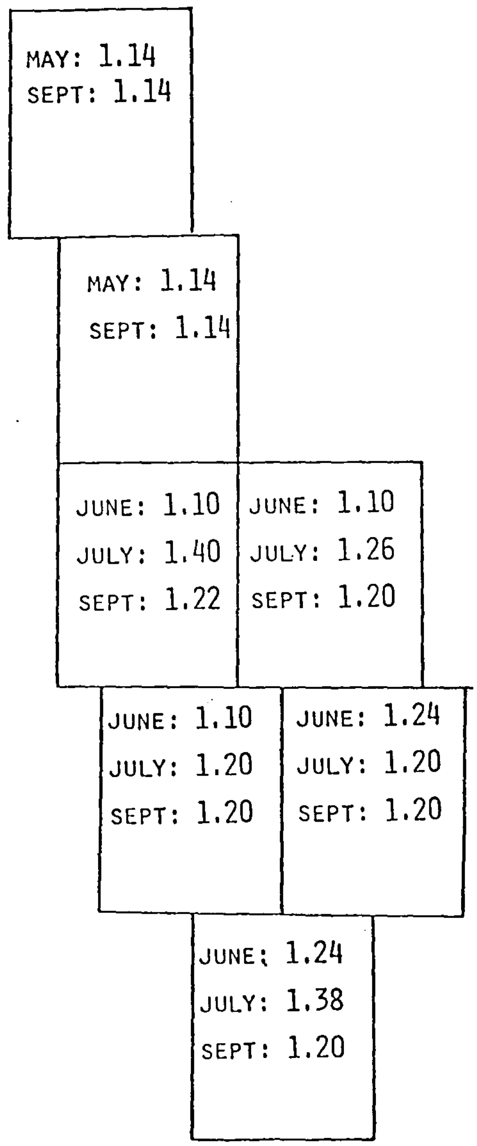


Figure 4-8. 7/5 Threshold Values by 30 Minute Block

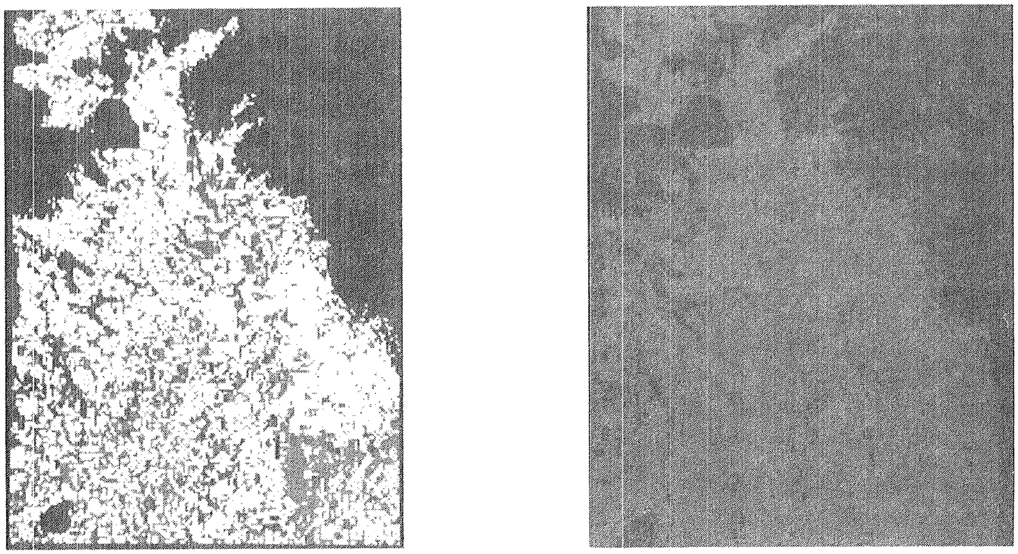


Figure 4-9. Class maps for one 30 minute block in northern Sacramento Valley. Map on left shows 8 classes. Map on right shows 2 classes with red being irrigated, black being not irrigated.

CLASS NUMBER		IRRIGATION PATTERN	COLOR
1	:	not irrigated on any date	black
2	:	irrigated in June only	purple
3	:	irrigated in July only	pink
4	:	irrigated in September only	red
5	:	irrigated in June and July	blue
6	:	irrigated in June and September	tan
7	:	irrigated in July and September	green
8	:	irrigated in June, July and September	white

The classification was then summarized to output a measurement of the proportion irrigated. Within each 30 minute block, irrigated cell counts were summarized for each of the 8 classes. These counts were obtained for each ground sample unit and each land use stratum polygon within the block. Using these counts a proportion labeled irrigated was calculated for each sample unit and for each land use stratum. These proportions were used to assess the accuracy of the 7/5 discriminant and to simulate regression estimates of irrigated proportion.

Estimation

The proportions irrigated from the classification and the ground data collection were input to a computer program, in order to produce an estimate of the proportion of irrigated land. As in Task I, both weighted, shown in IA and unweighted observations were used with the regression estimator equation 5 of Appendix IA to calculate an estimate for each of the seven land use strata described in Table 3-2. Regression coefficients were estimated using matched ground and Landsat class map irrigated proportion data associated with sample unit locations defined in Task I. Ground proportions were obtained from DWR field enumeration for Task I in 1979 and Landsat proportions from ground file cells labeled as irrigated using the 7/5 ratioed digital data. The resulting regression equation for each stratum was then used to compute an estimate of irrigated proportion and standard error for the entire population of sample units within that stratum. The results of the estimation are shown in Tables 4-5 through 4-7.

Summary statistics were computed (see Table 4-5) for the entire area within the Task II sample frame by combining the stratum estimates according to the equations presented in Task I. Thus, for the area within the "pseudo" Sacramento basin sample frame, the regression estimator (using sample unit size-weighted observations) produced an estimate of 72.4 percent irrigated with a relative 95 percent confidence interval half-width of 8.0 percent.

TABLE 4-5. Summary Statistics for the Stratified Weighted and Unweighted Task II Regression Estimates of Irrigated Proportion

Estimator	Proportion	Standard Error	Degrees of Freedom	95% C.I.	Relative Standard Error (%)
Weighted	.72404	.02828	30.14	.05775	7.98
Unweighted	.73821	.02764	21.41	.05749	7.79

TABLE 4-6. Summary statistics for the per stratum regressions of Task II weighted proportions on the weighted ground proportions

Stratum	Proportion	Standard Error	Population Size	Sample Size	Intercept	Slope Coefficient	r^2
1	.35145	.11178	139	7	-.02936	.46108	.47740
2	.59923	.00113	45	4	.04410	.71194	.95692
4	.79651	.03391	587	29	.02597	.85479	.85920
5	.87183	.04997	102	4	-.00443	.94806	.99121
6	.75831	.00590	9	4	.06491	.85282	.99905
7	.24812	.04460	17	4	-.08150	.34469	.69372

TABLE 4-7. Summary statistics for the per stratum regressions of Task II unweighted proportions on the unweighted ground proportions

Stratum	Proportion	Standard Error	Population Size	Sample Size	Intercept	Slope Coefficient	r^2
1	.47373	.13596	139	7	-.14413	.74872	.57592
2	.57848	.02740	45	4	-.53701	1.42514	.95087
4	.80108	.02923	587	29	-.14255	1.04659	.22667
5	.85210	.05379	102	4	.83145	.02236	.00140
6	.65429	.10644	9	4	-1.03065	2.04218	.72929
7	.19938	.04414	17	4	.55913	-.37470	.15176

Map Accuracy Assessment

As expected, best class map accuracies were achieved in areas dominated by agriculture (see Table 4-6). In dryland areas (stratum 1) and in areas where the agriculture is dispersed and mixed with urban and native vegetation areas (stratum 7), the simple 7/5 discriminant is not as effective. These areas may require supplementary spectral information (e.g. a brightness band), or the threshold 7/5 value may need to be set separately by land use stratum. It should be noted, however, that strata one and seven make up a very small proportion of this agricultural area (see Table 4-8), although they could be significant in other areas.

Table 4-8. Strata Weights for Sacramento Hydrologic Basin

STRATUM	SAMPLE UNIT POPULATION	STRATUM WEIGHTS
1	212	.138
2	150	.083
3	---	----
4	951	.664
5	60	.037
6	47	.025
7	73	.053
TOTAL	1493	

The sample units and land use strata were examined for confusion factor and classification problems. Certain factors, such as riparian areas being classified as irrigated, appeared throughout. Others, such as young orchards being classified non-irrigated due to the high proportion of bare soil, were stratum-specific (see Table 4-9).

Table 4-9. Task II Classification

STRATUM 1

- Irrigated versus non-irrigated grain
- Native vegetation
- Riparian areas

STRATUM 2

- Riparian areas
- Small fields surrounded by native vegetation

STRATUM 4

- Riparian areas
- Idle or weedy fields

STRATA 5 AND 6

- Riparian areas
- Young orchards and vineyards

STRATUM 7

- Riparian areas
- Brush and native grasses

The Task II sample unit Landsat measurements of irrigated proportion were compared to the Task I manual Landsat measurements and to the ground data measurements (see Table 4-10). In strata 1 and 7 where the digital analysis was not as effective, quite good results were obtained with the manual analysis. This result can be explained by the fact that the human interpreter was better able to distinguish riparian and native vegetation from irrigated area. Although the non-agricultural areas are vegetated and have high 7/5 values, their texture and appearance (identifiable manually) are less uniform than cropland in agricultural areas.

Continuing Work

Further analysis of the Sacramento Valley digital data set is planned for 1981. This work includes use of UC Berkeley's Survey Planning Model in analysis of the sample design and refinement of the use of the vegetative indicator.

The Survey Planning Model (SPM) (see Section Task 4) is being upgraded to arrow inexpensive simulation of sample frame and irrigated (or crop) proportions by spectral class over very large areas. The SPM will also allow simultaneous summary of irrigated proportion(s) and variance(s) by sampling stratum, measurement error stratum, and reporting unit stratum. Additionally, the SPM will be used to compute sample size and allocation among strata that minimize total variable cost subject to meeting pre-specified sample error requirements for an estimate of irrigated proportion.

Further analysis of the vegetative indicator will be performed. The feasibility of setting the irrigation threshold value by land use stratum, or by some other set of strata such as crop group, will be determined. The effect of masking-out urban areas and/or large areas of native vegetation will be evaluated. Also, time and resources permitting the use of vegetative indicators other than the 7/5 ratio will be examined. Finally, an evaluation of map accuracy on a point-by-point basis by irrigation line threshold, region, and combination of dates will be performed.

TABLE 4-10. Comparison of Digital and Manual Results

STRATUM	SAMPLE UNIT	GROUND	DIGITAL	MANUAL
ONE	CO111	.12	.60	.10
	PL24	.33	.49	.46
	PL32	.45	.56	.37
	YO7	.03	.32	.03
	YO29	.04	.33	.04
TWO	SU123	.87	.95	.77
	TE12	.47	.70	.49
	TE23	.74	.93	.86
	GL7	.65	.84	.65
FOUR	CO170	.86	.98	.91
	CO160	.96	.90	.88
	CO119	.84	.88	.72
	GL139	.95	.92	.97
	GL156	.90	.98	.95
	GL105	.81	.95	.84
	SO60	.46	.69	.68
	SO47	.83	.85	.78
	SU20	.75	.93	.84
	SU59	.92	.79	.91
	SU40	.86	.93	.97
	YO170	1.00	.99	.97
	YO174	.95	.98	.96
	YO83	.77	.80	.82
	YO138	.92	.89	.98
	YO143	.58	.75	.74
	YO89	.71	.71	.80
	YO147	.88	.95	.92
	YO148	.58	.90	.72
	YO195	.95	.92	.95
	YU36	.52	.91	.58
	TE56	.91	.89	.93
	GL149	.96	.84	.98
	GL50	.98	.94	1.00
	GL16	.62	.78	.66
	TE94	.32	.91	.43
	BU22	.83	.99	.65
	GL76	.91	.94	.91
	GL62	.61	.86	.51
FIVE	BU96	.94	.99	.99
	CO172	.84	.68	.72
	YU30	.73	.99	.68
	BU32	.89	.93	.94
SIX	TE90	.86	.94	.88
	BU4	1.00	.96	.96
	SU109	.92	.98	.98
	SU120	.80	.90	.81
SEVEN	PL67	.25	.80	.27
	PL52	.17	.96	.14
	PL60	.29	.96	.24
	PL49	.13	.96	.15

5.0 MANUAL CROP TYPE IDENTIFICATION (TASK III)

The complexity of California's agricultural environment defies any effort at large area crop-type identification using manual interpretation techniques. The sheer volume of the data, the number of crops, and the number of dates required would be too burdensome for the human interpreter. Nevertheless, a number of research questions and limited areas of study present themselves.

During the 1980 project year, three substudies were undertaken in this area by GRSU. The first was the completion of an additional four phenology diagrams. The second involved planning the development of a statewide crop-type data base which would aid in either manual or digital interpretations. The third involved planning for a four county small grains survey during the 1981 season. Each substudy is detailed in the following section.

5.1 Crop Phenology Diagrams

The Geography Remote Sensing Unit completed an additional four crop phenology diagrams during 1981. Phenology diagrams have now been completed for Cotton, Small Grains, Sugar Beets, Melons, Rice, and Alfalfa (see Figures 5-1 to 5-6). Each diagram graphically illustrates the growth cycle, the spectral appearance on color-IR Landsat and an oblique view of the crop in the field during the critical portions of its life cycle.

We have found the diagrams to be an excellent educational tool for illustrating the multitemporal dimension of crop growth and the accompanying variation in spectral signatures. They have additionally served to increase the visibility of the California Irrigated Lands APT project, with copies of the slide sets being sent to individuals in both federal and state agencies, as well as individuals in the private sector and foreign countries.

There are no current plans to make phenology diagrams for any other crops. We have, however, been made aware of an extensive study done by the USDA of cotton in the San Joaquin Valley and are considering the addition of day-degree data to the cotton diagram. Day-degree information is important in phenological studies and its addition should prove valuable in illustrating the linkage between traditional crop ecology techniques and remote sensing technologies.

5.2 California Quad-Based Agricultural Information System

During the 1980 project year, plans were developed for creating a simple statewide crop-type data base - the actual work to be carried out during 1981. The need for such a data base arises from the large area and great crop diversity found in California agriculture. Knowledge of the potential mix of crops in an area allows the interpreter or the computer to quickly focus on the best interpretation strategy by excluding from consideration, or at least,

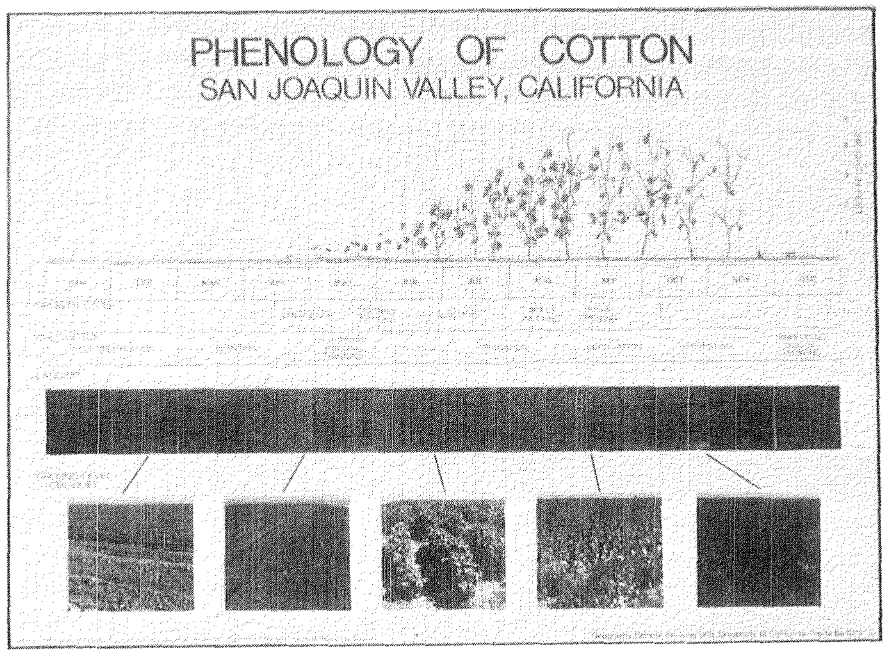


Figure 5-1. Phenology of Cotton in San Joaquin Valley, California

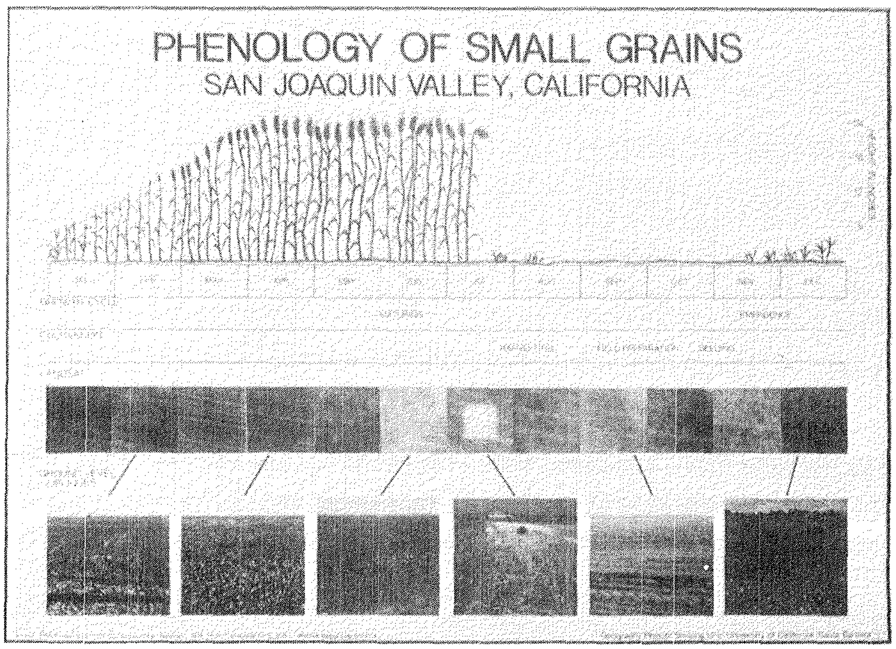


Figure 5-2. Phenology of Small Grains in San Joaquin Valley, California

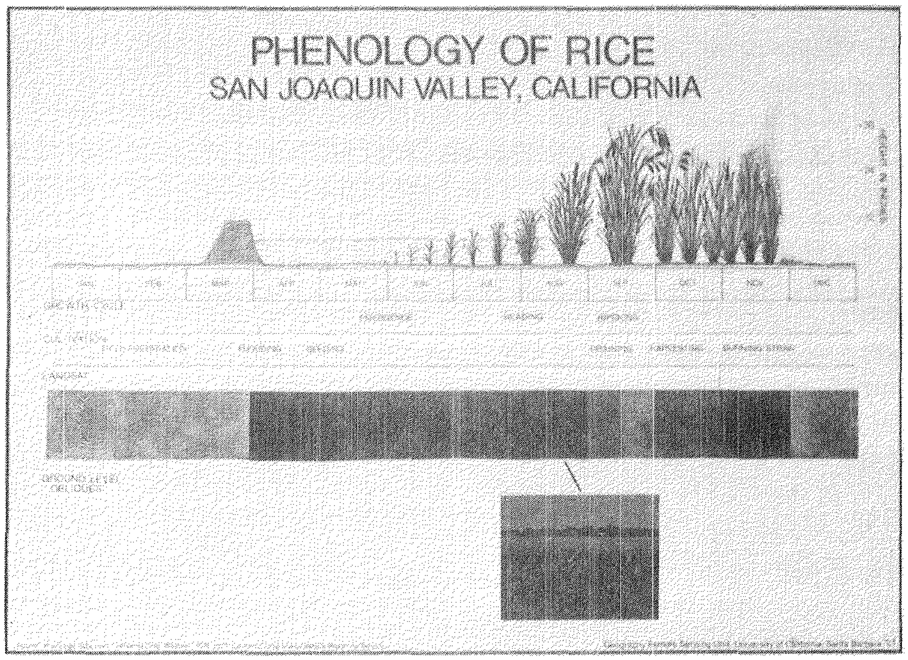


Figure 5-3. Phenology of Rice in San Joaquin Valley, California

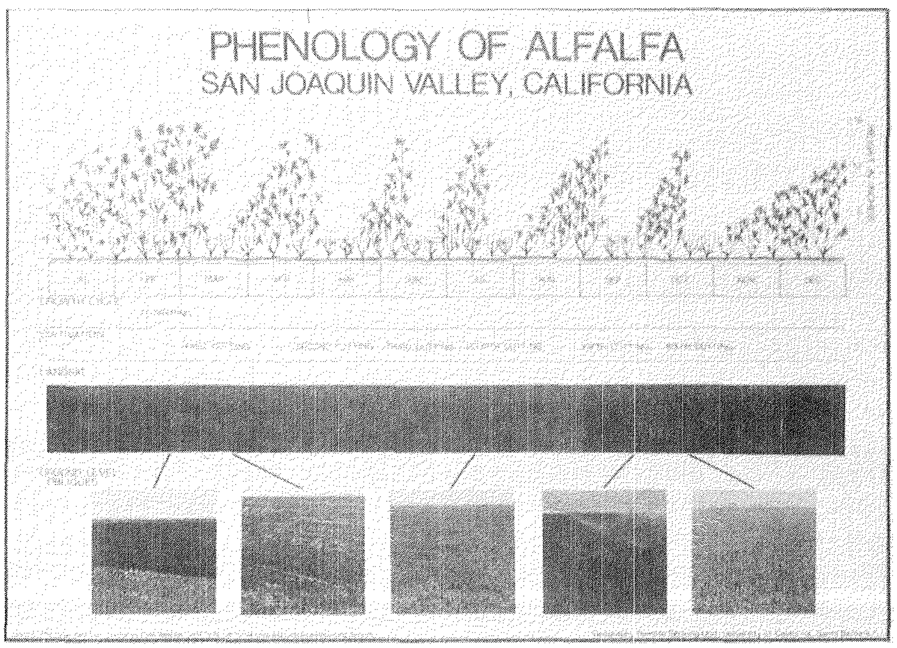


Figure 5-4. Phenology of Alfalfa in San Joaquin Valley, California

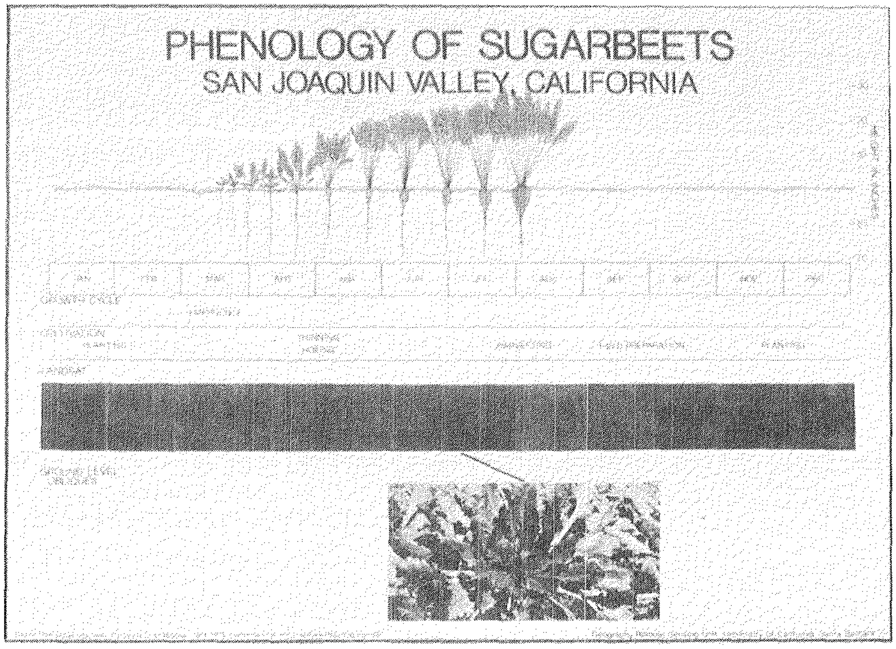


Figure 5-5. Phenology of Sugarbeets in San Joaquin Valley, California

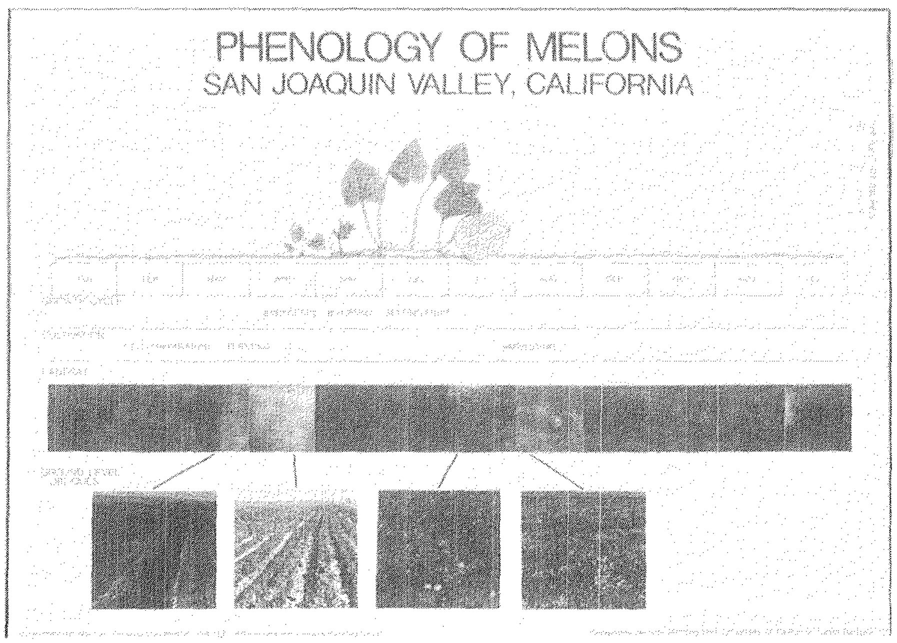


Figure 5-6. Phenology of Melons in San Joaquin Valley, California

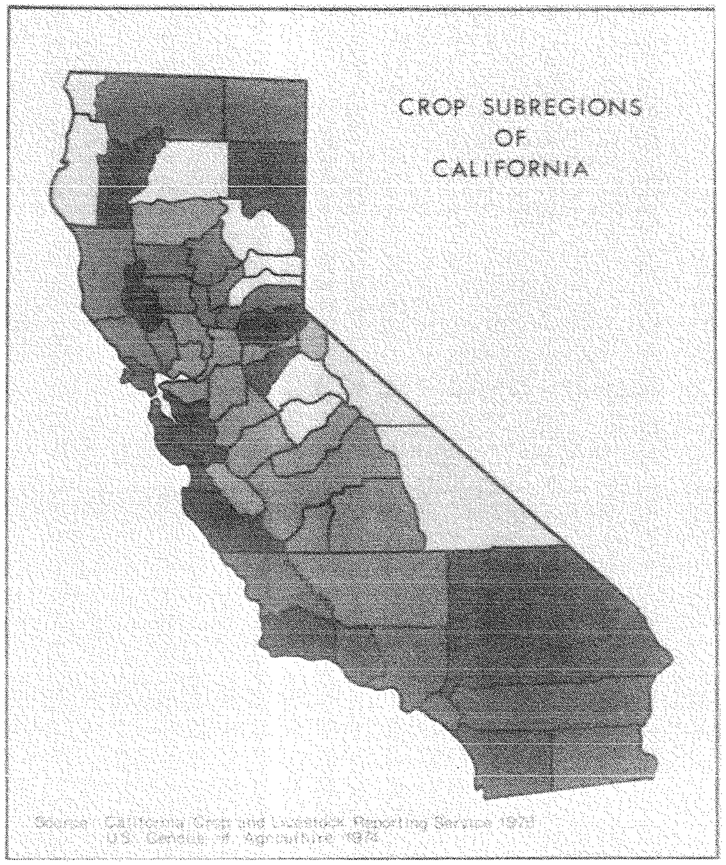


Figure 5-7. Crop Subregions of California

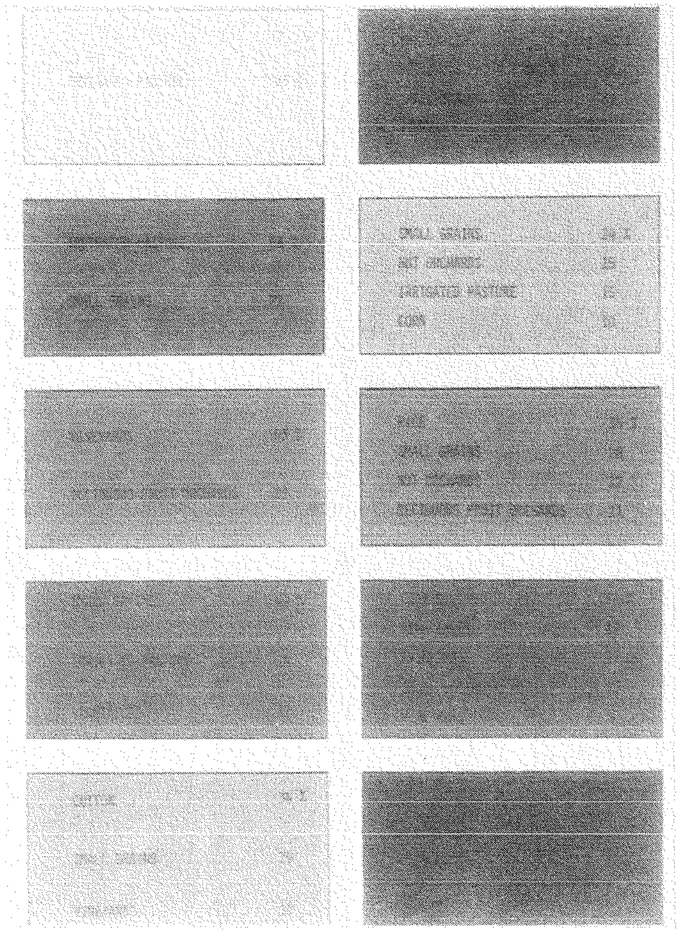


Table 5-1. Crop Subregion Crop Clusters

weighty consideration, those crops that historically have not been grown in the area.

We have examined agricultural statistics gathered by the California Crop and Livestock Reporting Service. As with most such data sets, crop acreages were aggregated at the county level which greatly reduces their value to the user concerned with crop distributions at a finer level. Nevertheless, analyzing such highly aggregated data revealed statewide crop mix patterns which, when displayed in a graphic format, could be used to orient interpreters to the general crop environment (Figure 5-7). Using a clustering algorithm on the Statistical Analysis System (SAS), county crop acreage statistics were analyzed to group those counties together having similar crop composition (both type and proportion) (Table 5-1). A number of groupings of adjacent counties developed, indicating similar crop environments. While acceptable for orientation purposes, it was not adequate for interpretation purposes.

Probably the most potent source of data available is gathered by the California Department of Water Resources. The tabular data, derived from field maps, is available at the 7.5' quad level and covers over 60 crops and field conditions statewide. Its major weakness is that, for a given county, the most recent survey may be five or more years old. In certain areas, where there has been rapid change - urban encroachment, new water supplies, new crop varieties, changing market conditions, etc. - the data's lack of timeliness may reduce its usefulness as a representative picture of current conditions. Nevertheless, most areas probably do not change a great deal on an annual basis. While the proportion of a quad given over to a particular crop may fluctuate considerably from year-to-year, the crop mixture of a particular quad or group of quads probably remains fairly constant. This would be particularly true for crops that require major investments of time and capital such as orchards and vineyards, as well as those characterized by particular soil or climatic requirements. The level of organization required for harvesting and marketing activities will also impact the stability of cropping practices over time. In fact, it would be expected that the expansion of a new crop into an area would be by a type of "osmosis" from surrounding quads and, if the prediction of the probable crop mix in a given quad takes account of the adjacent quads, the change would not be particularly surprising. Major changes in activity are often detectable using the annual reports of various state and local agencies who publish data at the county level. Such ancillary data sources could be used in conjunction with the DWR data to make reasonable predictions about the crop environment that will be encountered during the interpretation phase. In short, we would expect the DWR 7.5' quad data to be a valuable part of the interpretation process and that its added spatial resolution would more than compensate for its poorer temporal resolution.

Incorporating this data into the interpretation procedure - whether manual or digital - allows the analyst to concentrate on the crops most likely present. The reduced dimensionality should actually increase interpretation accuracy. In the digital domain, one can imagine using this data as input to an <u>a priori</u> classifier. Rather than compare each pixel signature with all possible crop signatures, the decision strategy could begin with those crops having the highest likelihood of being in an area as well as those crops not expected but which are known to have a highly similar Landsat signature to one or more of the predominant crops. Combined with a well thought-out procedure for accepting the computer classification at some threshold level of certainty, without actually comparing each pixel with all known crop signatures, the use of DWR and other ancillary data in the interpretation flow may increase accuracy and reduce computing costs.

For the 1981 project year we are proposing to set up a simple statewide crop data base that utilizes DWR data. The tabular data for each county is currently on computer tape at DWR. We will format it in such a way that we can analyze the data using SAS as well as display the data graphically. We already have a program that will generate a grid of all 7.5' quads for California in latitude/longtitude coordinates and plans have been made to obtain the latitude/longtitude coordinates for the county boundaries of California. This will serve as the base for graphic representations of the data.

We feel that this effort will at the very least provide the analyst with a tool that will increase understanding of California's unique crop environment. Graphic representations of this work may also serve the same educational purpose as the crop phenology diagrams. We expect, this effort will allow us to incorporate historical DWR data directly into the interpretation procedure. Conversely, advantage will accrue from setting up a system that allows DWR to logically integrate remote sensing technology into its own data collection procedures.

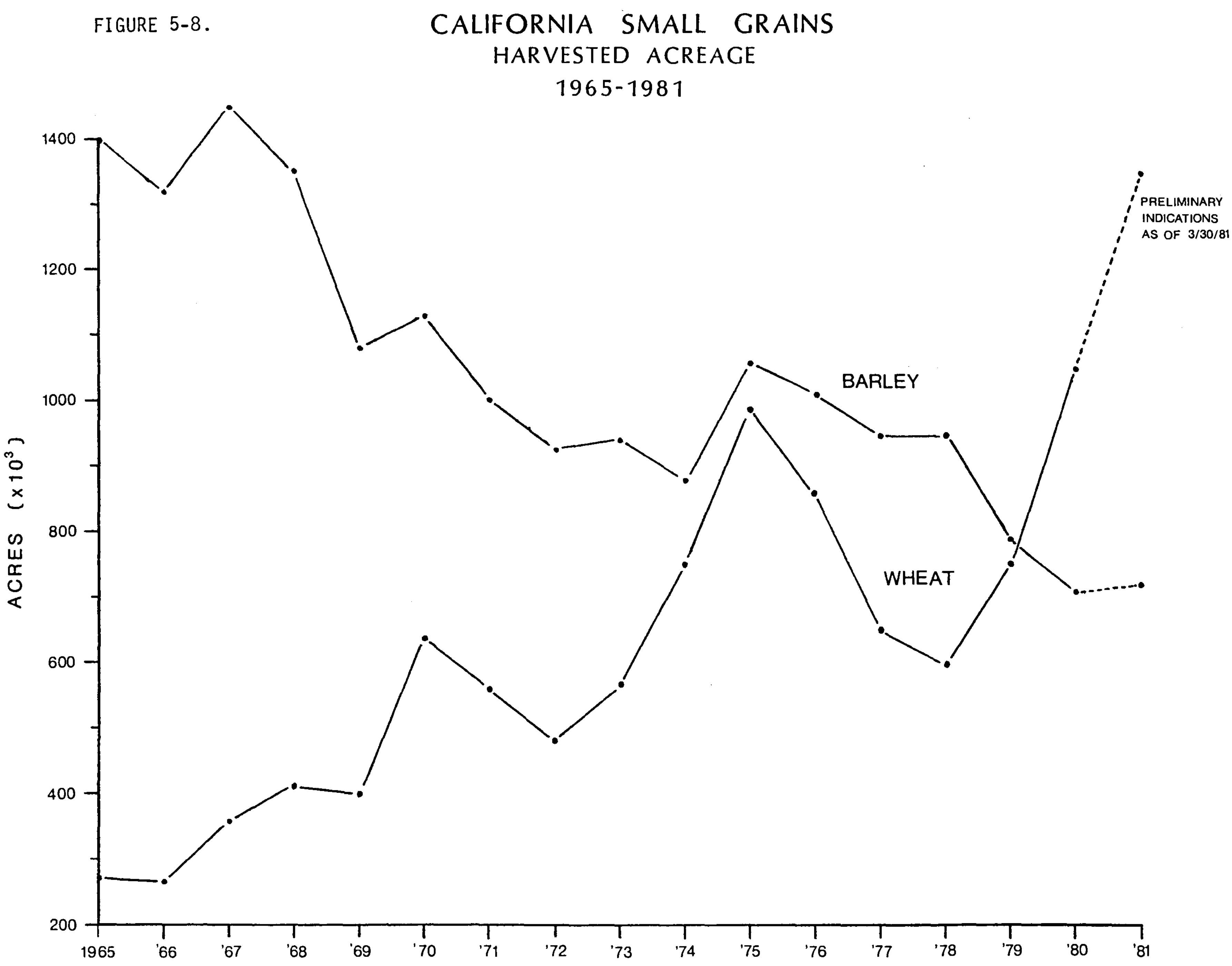
5.3 Estimation of Small Grain Acreage

Small grains occupy more acreage than any other single crop in California. Consisting primarily of wheat and barley, but also including oats, small grains are characterized by variable acreages from year-to-year (Figure 5-8). Over the past 15 years, barley has declined in importance while wheat acreage has been increasing. Superimposed on this changing picture is the fact that small grains are both irrigated and non-irrigated.

The California Department of Water Resources includes small grains in its land use mapping efforts, although wheat, barley or oats are seldom separately annotated. While certain areas have small grains that are definitely dry farmed, non-irrigated grains grown on the valley floor, where extensive irrigation is practiced, are often indistinguishable from those which have been irrigated. In many cases the amount of irrigation water applied is very light. Because the standard DWR survey does not usually get underway until July, and can extend well into September, long after small grains have been harvested, DWR tends to miss some grains. This problem is compounded by the fact that grain's early harvest date allows farmers to use the same fields for second crops, often making it impossible for field crews to detect evidence of a previous grain crop.

During the 1980 project year, the Geography Remote Sensing Unit began preparation for a limited manual survey of small grains to be conducted during the 1981 season. The primary focus to-date has been to improve the manual interpretation techniques developed during Task I and to develop a system which could easily be integrated into DWR's land use survey procedure. Previous

FIGURE 5-8.


CALIFORNIA SMALL GRAINS
HARVESTED ACREAGE
1965-1981
ACRES (x10³)
1400
1200
1000
800
600
400
200
1965
'66
'67
'68
'69
'70
'71
'72
'73
'74
'75
'76
'77
'78
'79
'80
'81
BARLEY
WHEAT
PRELIMINARY
INDICATIONS
AS OF 3/30/81

work conducted by UC Berkeley has already demonstrated Landsat's capability to detect small grains. The unique "spring red/early summer yellow" signature makes grain fairly easy to distinguish from other crops. We will closely follow Berkeley's work, particularly with reference to typical grain signatures, confusion crops, and interpretation strategies.

In order for DWR to operationally utilize Landsat for small grains estimation, a number of research questions will have to be answered - What dates are required to detect grains? What are the potential confusion crops and can wheat, barley, and oats be separately distinguished? Can irrigated and non-irrigated grains be separately distinguished? At what scale should the data be interpreted and measured? What are the areas of potential cost savings that could be realized in an operational setting?

Our work thus far indicates that small grain identification is basically a two-date task. The first date required usually ranges from early May to late June when grains are yellow in color. This phase distinguishes grains from other crops which are usually characterized by a red color on Landsat, or, in some cases, by a bare soil signature when the crop has not adequately emerged. A second date in early spring, usually early March to mid-April, is required to ascertain that a yellow field on the first date had actually been an actively growing field (red signature on spring Landsat) earlier in the year. If not, the yellow field may simply be grain stubble left from the previous season. It is important to realize that native grasslands in California go through a similar and concurrent color change on Landsat - even with optimal data selection, difficulties in interpretation may develop. A third data may possibly be of some value at this stage. An early August date, for example, could be useful for detecting burned stubble, plowed or replanted conditions which may help distinguish questionable grain fields from native vegetation.

While a given grain field can probably be identified with only two dates, not all grain fields can be correctly interpreted with the same two dates. Examination of small grains in the Tulare Basin reveal fields ranging from bright red to golden yellow to bare soil on the same early May date and on the same 7.5' quadrangle! Any "typical grain" signature that is defined will have to be shifted along its temporal axis to be applicable to the variety of conditions that will be encountered. It may prove to be the case that every Landsat date available during the critical yellowing period should be obtained so that all variations can be observed and analyzed.

In addition to selecting Landsat dates that capture the spring red/early summer yellow signature of grains, dates must be selected which enable the interpreter to distinguish grain from other crops and ground covers. In the UC Berkeley study the major confusion crop was safflower, which also turns golden during the summer. Despite its similiarity, safflower could be distinguished by trained interpreters due to slight color differences and the fact the safflower yellowing occurred generally later than grains. Additionally, safflower is declining in importance in California and should be represented by only minimal acreage. Crop confusion is not expected to be a major problem although grasslands may cause some difficulties. The use of DWR land use maps from previous surveys may aid the interpreter to distinguish native vegetation from crop land.

Because of the variety seen in grains as a category it seems unlikely that we could distinguish wheat, barley, and oats; although the variation seen during the preliminary examination of Landsat may be partially attributable to different spectral and temporal responses by grains. These are areas that need further research. A related question is the separability of irrigated and non-irrigated grains. This information is important for DWR's water management responsibilities, particularly during drought years.

The scale of interpretation and measurement, and potential cost savings represent the area given the greatest attention in 1980. In part, this work was carried out to explore means of improving the Task I manual irrigated/non-irrigated procedure. Three scales are being examined - 1/250,000, 1/125,000, and 1/100,000 - for Landsat interpretation.

Interpretation at 1/250,000 would be advantageous because EROS provides Landsat prints at this scale (current cost \$70.00) as a standard product. If usable, this product, although more expensive than 1/1,000,000 transparencies, would greatly reduce the enlargement costs experienced in Task I. An additional argument for this scale is the already existing 1:250,000 USGS map series from which base maps could be made.

EROS also provides, as a standard product, 1:125,000 scale RBV enlargements (current cost \$35.00). The sharper resolution of the RBV makes it an ideal product for verifying boundaries and clarifying the agricultural/urban interface. Although the image is in black and white, an interpreter experienced with panchromatic films may be able to distinguish some crop conditions. RBV coverage is intermittent and not always available and so must be viewed as ancillary information rather than a primary data source. It should be noted that there are relatively few map bases available at this scale which would require a considerable start up effort by DWR. It is our opinion that the availability of RBV as a standard product at 1:125,000 is not in itself an adequate argument for this scale. A better strategy would involve the enlargement of RBV transparencies by DWR to whatever scale is finally deemed appropriate. We have, however, done irrigated cropland mapping for the Kern County Water Agency using a map base (provided by KCWA) at 1:125,000 scale and found it an excellent scale with which to work. Landsat MSS can be enlarged to this scale without extreme fuzziness. An added advantage of this scale is that NASA U-2 coverage, an even higher resolution ancillary data source, usually runs between 1:120,000 and 1:130,000 scale, which is close enough to the 1:125,000 map base for easy verification.

The USGS is presently in the process of completing the 1:100,000 map series. As the planimetric bases are completed data from various series are being added (e.g., Land Use and Land Cover). A number of these bases have been completed and, in some cases, the Soil Conservation Service has used them to create county maps showing important farmlands. The 1:100,000 series may prove to be the best base for statewide and regional planning in the future. In terms of Landsat interpretation our experience indicates that this scale is relatively easy to work with - enlargements maintain sufficient color saturation and edge definition - although with larger counties, the interpretation overlay can become unwieldly because of its larger size. The major arguments for this scale are that its larger format allows more categorical information and smaller fields to be included in the final map product, and that it may become the principal scale of presentation for other resource and land use data in the future. While all areas in California have not been completed, we were fortunate that the four counties (Glenn, Butte, Yolo, and Kings) selected for the 1981 grain survey were available in planimetric form.

The general cartographic rule-of-thumb is that the minimum mapping unit - on the map surface - should be 2 mm on a side. Below we see the acreage represented by a rectangular feature 2 mm on a side (4 mm^2) at a variety of scales.

Scale	Acres
1:1,000,000	984.0
1:250,000	61.5
1:150,000	22.2
1:125,000	15.4
1:100,000	9.9
1:24,000	0.6

Obviously, the larger scales allow the interpreter to include smaller features and a greater number of categories. This will have to be a factor in the final scale selection. While this may be a valid rule for producing maps, our experience indicates that where the purpose is deriving an estimate rather than a publishable map, areas smaller in size can be easily marked (the minimum mapping unit used in Task I resulted in the demarcation of 10 acre fields at 1:150,000 scale - less than one-half the area shown above). Whatever the purpose - mapping or acreage estimation - DWR will have to determine where Landsat fits into their data collection efforts and at what scale it should be analyzed and presented.

In the process of examining potential scales for Landsat interpretation a procedure was developed at the Geography Remote Sensing Unit that we feel will greatly cut the cost of any future manual effort. Costs for Landsat enlargement for the statewide manual irrigated/non-irrigated effort was approximately $13,000. Our experience with Kern County Water Agency is that Cibachrome enlargements for three dates of Landsat to 1:125,000 scale for the Central Valley portion of Kern County costs approximately $450, including materials and the photographer's time. Major problems encountered are consistency of color balance and the precision of enlargement to the desired scale. Generally our experience has been that the enlarged products have always had some deficiency, although the interpreter usually found an adequate means of compensating.

An experiment was undertaken whereby 35 mm color slides were taken of the Landsat 1:1,000,000 transparency and then projected onto a map base for interpretation. The system that was developed is simple to use, allows for archival of the Landsat data by DWR's quad indexing system, requires a minimum amount of the photographer's time, and results in products that are easier to interpret than enlargements and are far less costly.

The viewing area of a 35 mm slide covers an area equivalent to a 15' block of latitude and longtitude (with a one mile wide border on all sides) when the Landsat 1:1,000,000 transparency is photographed at contact scale. Using the USGS 1:1,000,000 scale 7.5' quad index map of California, an overlay of the area of interest is made by tracing the county boundary and the 15' lines of latitude and longtitude onto clear mylar. The county boundary generally has some manifestation on the Landsat scene (e.g., rivers, ridge lines, coast line, roads) so that the overlay can be registered to the Landsat transparency. With one side of the overlay taped to the transparency it is placed on a small light table.

The small light table is placed on a photographic copy stand to which has been secured a vertically aimed automatic 35 mm camera. The camera lens is adjusted for contact scale and precise focusing is obtained by adjusting the height of the camera above the light table surface. The automatic setting is used for timing the exposure with the camera stepped down one full stop to increase color saturation on the slides. Because fluorescent lights found in our light tables is deficient in red light, a magenta filter (81A + 81C + 35M) is used. The film type used to-date is Kodak Ektachrome 64.

Each 15' block is then lined-up within the view finder and photographed. The small light table can easily be repositioned under the stationary camera to change the scene. The camera used by GRSU has a motor drive film advance so that the process at this point consists of merely positioning the Landsat scene and activating the camera with a cable release. Because of its simplicity our photographer only has to set up the camera; the actual photography can be done by anyone (a set of procedural instructions for setting up the camera are currently being drafted so that the photographer's time can be used for more demanding tasks). The time spent photographing the Landsat is minimal. We have, on one occasion, received our Landsat transparencies in the morning mail, photographed the area of interest and began interpretation on the afternoon of the same day.

The slides, when mounted, are archived using the Landsat acquisition date and DWR's quad indexing system which shows the path and row numbers for the four 7.5' quads found in a 15' block. The labelling of the slides is currently the most time consumming part of the process. The use of computer printed labels which could be quickly affixed to the slide mount surface would greatly reduce the time spent.

Once the slides are in hand, an advantage of this approach becomes apparent - the reduced volume of materials that must be physically handled and stored. The enlargements used in the Task I effort frequently presented storage problems, and transportation between UC Berkeley and UC Santa Barbara always required special handling. Using commercially available slide drawers, an array of 38 (there are 76 rows of 7.5' quads in California) by 25 (while there are 83 columns of 7.5' quads in the state, there is no single row with more than 50 columns) such drawers would provide a simple storage system that would allow quick access to Landsat coverage of any 15' block in the state. With each drawer holding approximately 50-100 slides (depending on the type of mount used) multiple years of data could be simply and readily archived.

Figure 5-9 shows a schematic design of a throw-back or back-lit projection system that can be used for image enlargement and interpretation. A simple version of this has been constructed at GRSU to test its usefulness to this project. Using a standard 35 mm slide projector, slides of 15' blocks of Landsat coverage are projected onto the back of a glass surface. The scene is made visible on the mapping material placed on top of the glass. The projected slide is brought to the correct scale by simply changing the distance between the projector and the glass table top. No major distortions are noticeable but an area for futher experimentation would be the use of a flat field projector lens (current cost ≅ $70).

This system requires the use of a base map showing numerous stable ground features which can be seen on Landsat. In the major agricultural areas of the Central Valley, the one square mile sections of the Township and Range Survey are often evident; in more mountainous terrain, deep river valleys and major roads can be seen on Landsat. These features, when placed on the base map, can be used to assure accurate, local registration. The base map represents a long-term capital investment and, once produced, is merely archived until a copy must be made to use in another study. Reproduction can be accurately done using a large format film, such as Cronoflex, which is relatively expensive, or, where some distortion is permissible, low cost Diazzo products. Excellent work copies are the inexpensive blueprint-type paper products which also are good surfaces on which to back-project slides. There are flat plate processes available for Diazzo reproductions which may be usable for relatively inexpensive but planimetrically accurate base map reproductions.

A thorough cost analysis has not been attempted but we would estimate the costs to be less than 20% of those incurred by using enlargements. Further cost savings could be realized by making slides (or enlargements) for only those areas where agriculture can be expected and only briefly examining the mountainous and vast desert regions on the 1:1,000,000 transparencies. While preliminary results indicate that the approach outlined here would fit into the DWR environment, we will continue to test it during the small grains analysis in 1981.

The actual 1981 small grains analysis will take place in Glenn, Butte, Yolo, and Kings County. Ground truth will be taken from a special full county survey conducted by DWR with the aid of NASA and University personnel. We at the Geography Remote Sensing Unit will analyze the imagery as it becomes available and teach DWR personnel the appropriate interpretative techniques. Following the interpretation of Landsat, DWR will provide GRSU with feedback as to any problem areas and suggestions for improvement. A procedural manual for interpreting small grains will follow.

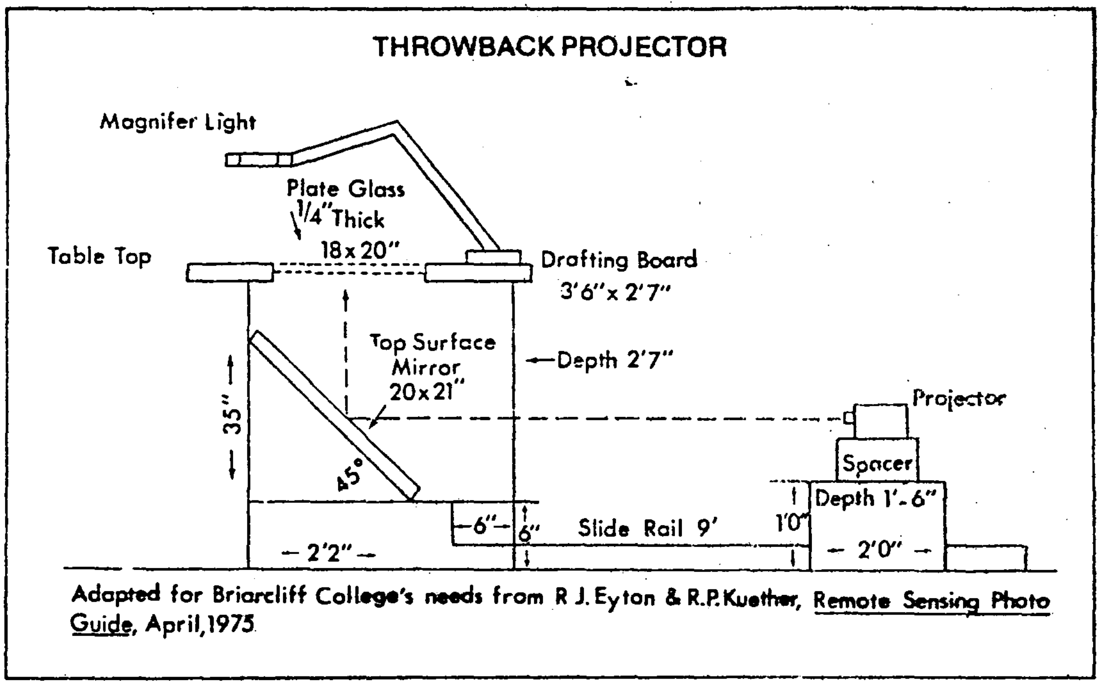


Figure 5-9. Schematic of a Throw-Back or Back-Lit Projection System

6.0 IDENTIFICATION OF CROP TYPE USING DIGITAL ANALYSIS TECHNIQUES (TASK IV)

The development of digital analysis techniques for estimating and mapping specific crop types continued during 1980. The results of previous crop type classification done in Kern County were re-evaluated. Work in the Sacramento Valley included the development of a general approach to crop type classification, development of a baseline classification procedure, and the development of a multicrop sample design.

6.1 Re-evaluation of Kern County Digital Crop Type Classification (U.C. Santa Barbara)

No major digital crop-type effort was undertaken at GRSU during the 1980 project year, but the results of work done in Kern County in 1979 was reassessed.

During 1979, a digital classification of Landsat data for a 3-7.5' quad size study area in the southern San Joaquin Valley was done. Unsupervised clustering of the brightness and greenness channels for four dates (May 1, June 11, August 8, and October 10, 1976) was used to create 100 clusters. The clustered data was then displayed on the video monitor and the analyst, using an interactive masking program and ground truth from one of the 7.5' quads, determined the appropriate label for each cluster. All three quads were then classified and evaluated to determine the extendibility and accuracy of the labels.

Results were based on 200 test pixels per crop or land cover type. The 1979 results indicated an average accuracy of approximately 66 percent. While the accuracy for certain crops and land cover was quite high (cotton 97%, grain 92%, native vegetation 97%) the accuracy for the various vegetable crops was very low. Because there were a number of different kinds of truck crop vegetables and each crop or land cover type was weighted equally in the analysis, the results indicated a poorer performance then had actually occurred.

During 1980, the accuracy assessment was revised using the same 200 test pixels per crop or land cover type. Because the classification approach used had done so poorly on the various vegetables it was decided to group all the vegetables together as a single class. While accuracy for the various vegetables was poor individually, as a group they were correctly identified 57 percent of the time. This increased the average class accuracy to 76.4 percent (Table 6-1). Vegetables, and the closely related melons, represent the major confusion categories.

This was followed by an accuracy assessment that weighted the test pixel results by percentage of the total area of each crop as represented on the ground truth maps (Table 6-2). With the exception of vegetables, any crop or land cover type representing more than 10 percent of the total area had an accuracy of at least 92 percent. Because cotton represented over one-third of the total area and its test pixel accuracy was 97 percent, it greatly influenced the final results. Using a weighted average approach, the accuracy of the cluster labelling would be more correctly stated as 87.9 percent. Table 6-3 shows the percent of the total area in each of the potential categories.

TABLE 6-1

KERN COUNTY 3-QUAD STUDY SITE

LANDSAT MULTICROP CLASSIFICATION - UNWEIGHTED

(BASED ON 200 TEST FIELD PIXELS/CROP)

LANDSAT CLASSIFICATION

GROUND DATA	COT	OR/VN	VEG	GRAN	MEL	POT	SB	ALF	NV/OTH
COTTON	97								3
ORCH/VINE		93	1	6					
VEGETABLE	7	4	57	14	6	3	5		4
GRAIN/SORGHUM			8	92					
MELONS			28		45				27
POTATOES			13			56			31
SUGAR BEETS			10	3			87		
ALFALFA		13						64	23
NV/OTHER		1		2					97

AVERAGE CROP CLASS ACCURACY = 76.4%

TABLE 6-2
KERN COUNTY THREE-QUAD STUDY SITE

LANDSAT MULTICROP CLASSIFICATION - AREA WEIGHTED
(BASED ON 200 TEST FIELD PIXELS/CROP)

	LANDSAT CLASSIFICATION								
GROUND DATA	COT	OR/VN	VEG	GRAN	MEL	POT	SB	ALF	NV/OTH
COTTON	35.5								1.1
ORCH/VINE		16.1	0.2	1.0					
VEGETABLE	0.8	0.5	6.6	1.6	0.7	0.3	0.6		0.5
GRAIN			0.9	9.8					
MELONS			0.7		1.2				0.7
POTATOES			0.3			1.5			0.8
SUGAR BEETS			0.3	0.1			2.2		
ALFALFA		0.2						1.0	0.4
NV/OTHER		0.1		0.3					14.0

AVERAGE CROP CLASS ACCURACY = 87.9

TABLE 6-3
KERN COUNTY THREE-QUAD STUDY SITE
CROP AREA WEIGHTS

CROP	ACRES	PERCENT TOTAL ACRES	TEST FIELD CLASSIFICATION ACCURACY	% TEST AREA CORRECTLY CLASSIFIED
COTTON	46368	36.6	97	35.5
ORCHARD/VINEYARD	21946	17.3	93	16.1
VEGETABLES	14640	11.6	57	6.6
GRAIN/GRAIN SORGHUM	13512	10.7	92	9.8
MELONS	3331	2.6	45	1.2
POTATOES	3308	2.6	56	1.5
SUGAR BEETS	3107	2.5	87	2.2
ALFALFA	2065	1.6	64	1.0
NV/OTHER	18264	14.4	97	14.0
	126541	100.0		87.9

The results of this exercise indicate that an appropriate accuracy assessment procedure will be required by DWR personnel so they can determine the value of Landsat classification over other data collection procedures. Improperly conducted, the accuracy assessment phase can either over-or under-sell Landsat's capability.

Secondly, the level of categorical grouping done during the labelling phase will greatly impact the accuracy results. As greater effort is made in the area of crop-type identification, crop group definition will become necessary - it is highly unlikely that California's 200 different crops could be correctly identified with any precision using current remote sensing data and techniques. This being the case, accuracy results should be carefully analyzed and not taken at face value.

6.2 Sacramento Valley Crop Type Inventory and Classification (U.C. Berkeley)

Crop type classification work in the Sacramento Valley addressed three major subtasks: (1) establishment of an overall framework for multicrop inventory system development responsive to DWR needs; (2) development of a baseline Landsat classification procedure which would work in California's complex agricultural environment; and (3) identification of workable sample designs for multicrop area estimation. These subtasks are discussed in the following sections.

6.2.1 General Approach to the Development of a California DWR Multiple Crop Inventory and Mapping System

The flow charts shown in Figures 6-1a to 6-1g present the approach selected for the development of a multiple crop inventory and mapping system responsive to California DWR's information needs. A first step in this process, in fact a naturally continuing step, is to define specific DWR information requirements. The answer to this question depends not only on the Department's current and projected information needs, but also (as shown in Figure 6-1a) upon the expected cost, accuracy, and timeliness characteristics of feasible inventory systems. In effect, system goals can be seen as dynamic. Information needs and the ability to afford given levels of direct cost, error, and turnaround time change over time. This situation is typical in large organizations and requires the development of inventory and mapping systems flexible enough to allow change and growth to occur in an organized, efficient manner.

Figure 6-1b describes the flow for specifying and developing the data acquisition, registration, preprocessing, stratification, and Landsat classification techniques appropriate to the DWR problem. Here, as elsewhere in this effort, advantage is taken of previous and on-going work in Landsat classification (e.g. as reported in AgRISTARS publications). Figures 6-1c and 6-1d describe two alternative procedures for classification of large, full frame areas. The first, termed multicrop type A classification, emphasizes the integrated use of presently available classification techniques. Reference to Figure 6-1c shows that the first two steps of the type A flow represent the classification procedure already described for irrigated-only mapping. In effect, the irrigated map provides a 'spectral stratification' for efficient multicrop classification within strata. Spectral clusters are defined through currently available techniques (e.g. ISOCLAS) and ground file cell assignment to spectral classes can proceed according to simple distance or more sophisticated maximum likelihood rules.

In contrast, the type B classification flow shown in Figure 6-1d uses more experimental techniques to increase classification accuracy beyond that obtainable in some areas with the type A procedure. The top three 'boxes' in Figure 6-1d refer to a more sophisticated (and more costly) method of defining spectral strata. Separation of crops into spectral strata is expected to be improved and therefore classification confusion between crops reduced with this procedure. This method is based on a technique currently under development in the AgRISTARS Corn/Soybeans project (Cicone _et al_ 1981). From this point, classification would proceed as in type A flow or, alternatively, using an AgRISTARS technique on a sample unit basis if spectral separation of crop types was not sufficient in the simpler procedure. In a development sense, the type B procedure is seen as a 'down stream' capability, not required unless the less expensive and simpler type A methodology is found inadequate for some DWR land use mapping problems. Table 6-4 summarizes the differences between these two approaches to Landsat classification.

Once a class map is available, a procedure for determining the crop/land use composition of each spectral class must be defined. Figure 6-1e shows the proposed approach to this problem. Field data is obtained, digitized, and registered to the north-south ground coordinate file. Landsat class map data, previously registered to this sytem, is then intersected by computer with the ground truth data. Crop/land use proportion data by spectral class results directly. Also as a by-product of this process, an estimate can be obtained of the correlation between Landsat proportion by crop type versus ground truth proportion for the same crop type.

Performance of sample designs for multiple crop estimation are affected by a variety of factors, factors which often interact in complex ways. In order to allow a systematic evaluation of alternative specifications for inventory components, a survey simulation approach has been selected for use in designing a multicrop inventory system for the California DWR. The idea is to simulate a population of sample units and associated crop proportions, and then estimate the cost and error performance of alternative sample designs when applied to that population. When combined with an appropriate experimental design, the cost and/or error attributable to a given design specification can be identified as can the cost/error interactions between design component specifications.

University of California personnel have access to an inventory simulation capability. Known as the Survey Planning Model (SPM), this software package has been developed over the past several years at Berkeley (Titus 1979 and Wensel <u>et al</u> 1979). Basically, the SPM consists of two parts. The first is a module which defines a sample frame and then simulates resource parameter values for each sample unit based on class map data and means/variances for resource parameters associated with each map class. This module enables simulation of a sample unit population according to alternative specifications of (Landsat) classification procedure, sample frame configuration, and stratification strategy. Using the resulting sample unit counts and simulated parameter variances (by stratum and sample stage), the second SPM module estimates the sample size and allocation to strata/stages necessary to simultaneously meet precision goals for each parameter. The expected sample-size-dependent cost for each such estimate is also given.

The plan in this project is to adapt the SPM to the California DWR multicrop estimation problem. Figure 6-1f illustrates the SPM's use in the multicrop sample design process. The resource parameters of interest are represented by the proportion or area for each crop type or group having a given level of water use. As seen in the figure, specifications are made for each sample design component according to a strategy identified by the experimental design. These specifications then direct processing in both SPM modules.

Each of the designs examined in the Survey Planning Model can be ranked according to the total sample size-dependent cost to achieve given error goals. Since a map product is also required, this ranking can be adjusted with reference to the accuracy of the Landsat class map associated with each design.

FIGURE 6-1A. PHILOSOPHICAL APPROACH TO DWR MULTICROP PROBLEM

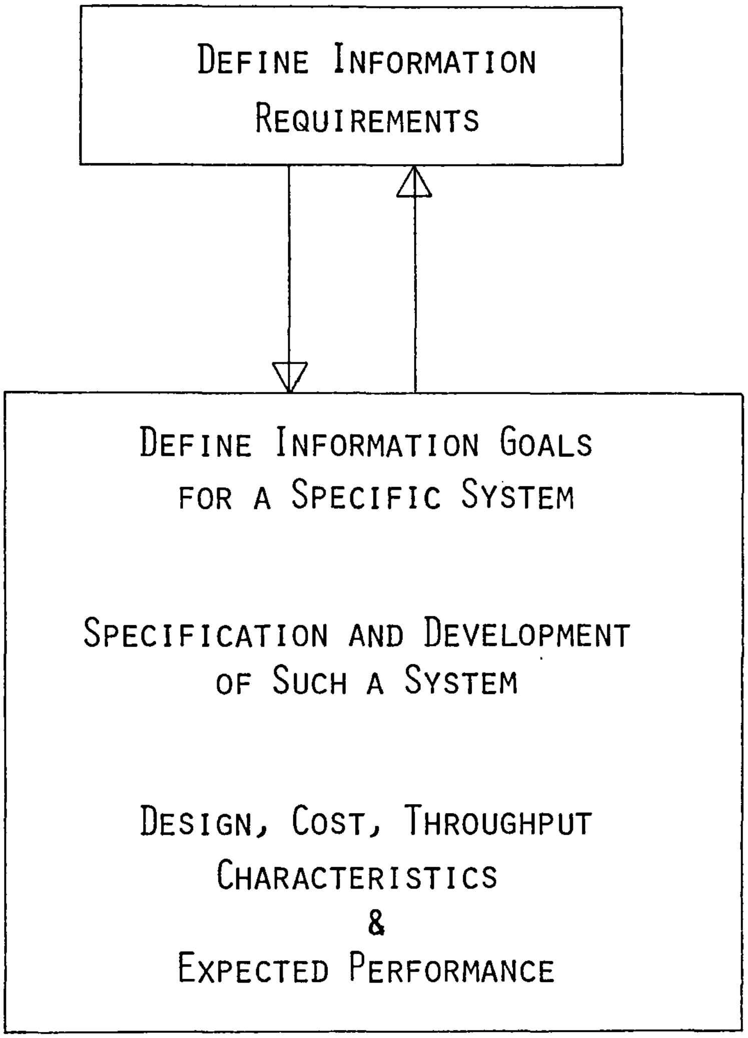


FIGURE 6-1B. APPROACH TO MULTICROP AREA INVENTORY AND MAPPING SYSTEM DEVELOPMENT: LANDSAT/ANCILLARY DATA PREPROCESSING AND CLASSIFICATION

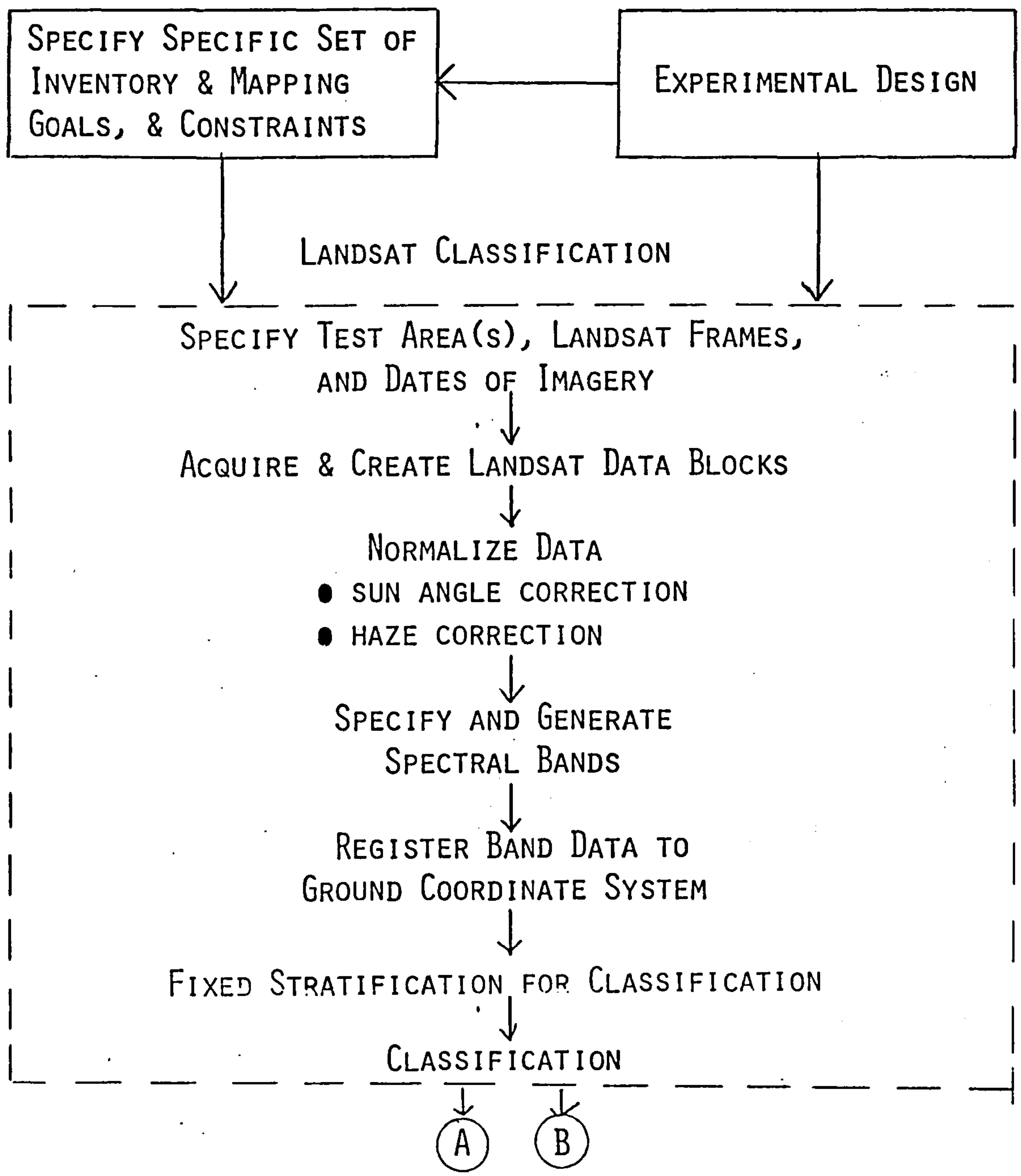


FIGURE 6-1c. MULTICROP TYPE A CLASSIFICATION FLOW

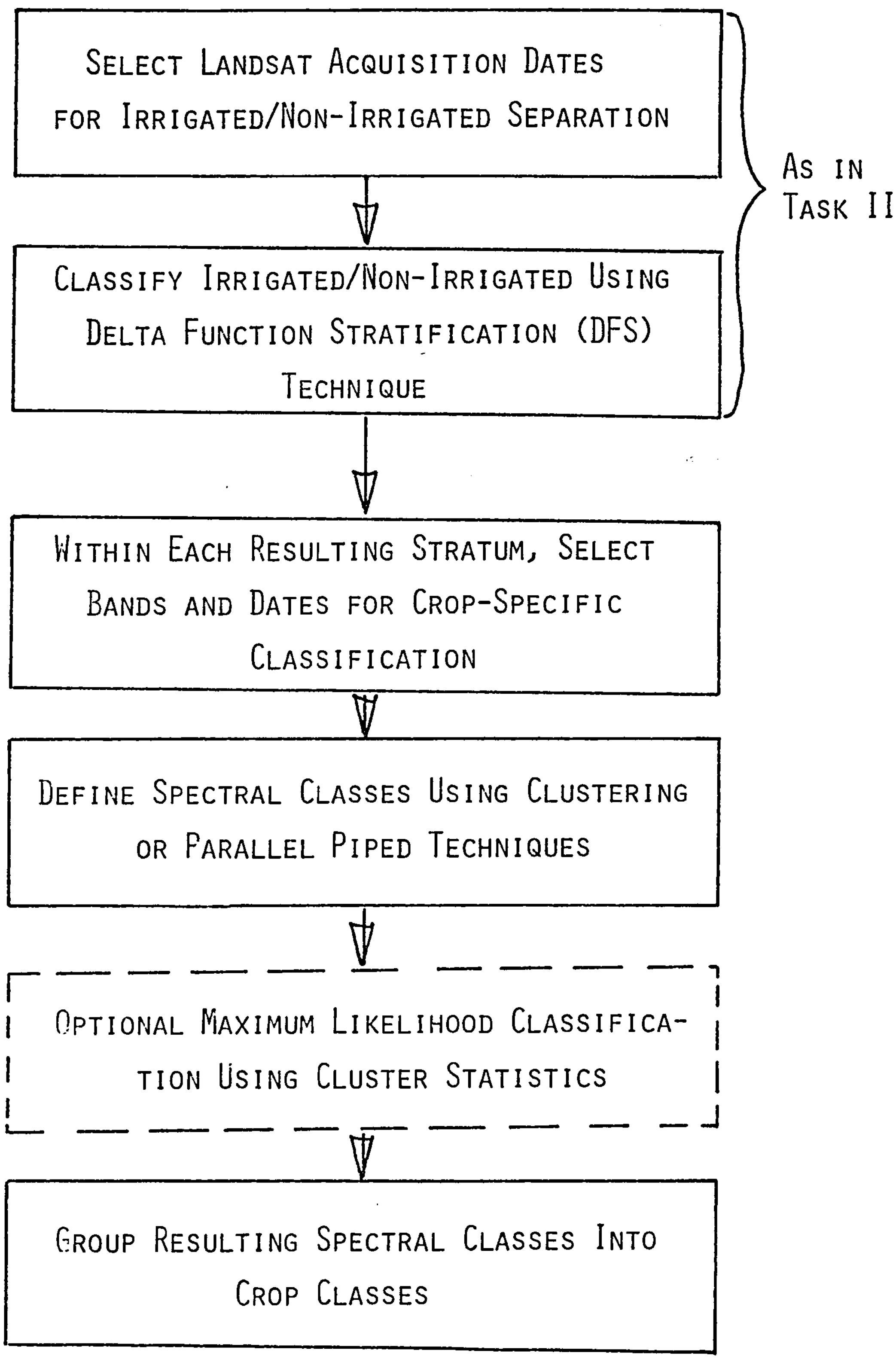


FIGURE 6-1D. EXAMPLE OF A MULTICROP TYPE B CLASSIFICATION FLOW

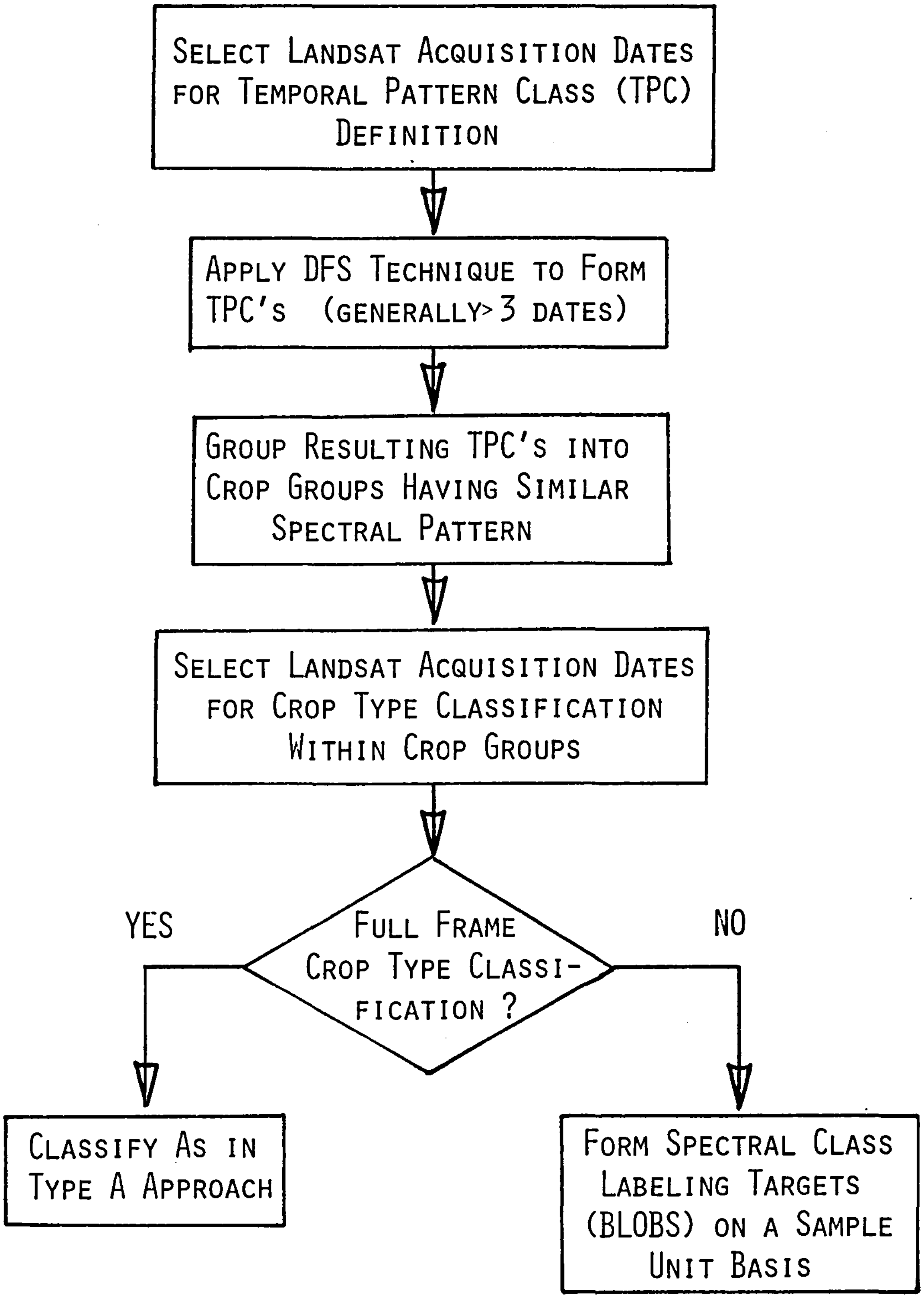


FIGURE 6-1E. LANDSAT CLASS MAP COMPOSITION AND CORRELATION ASSESSMENT

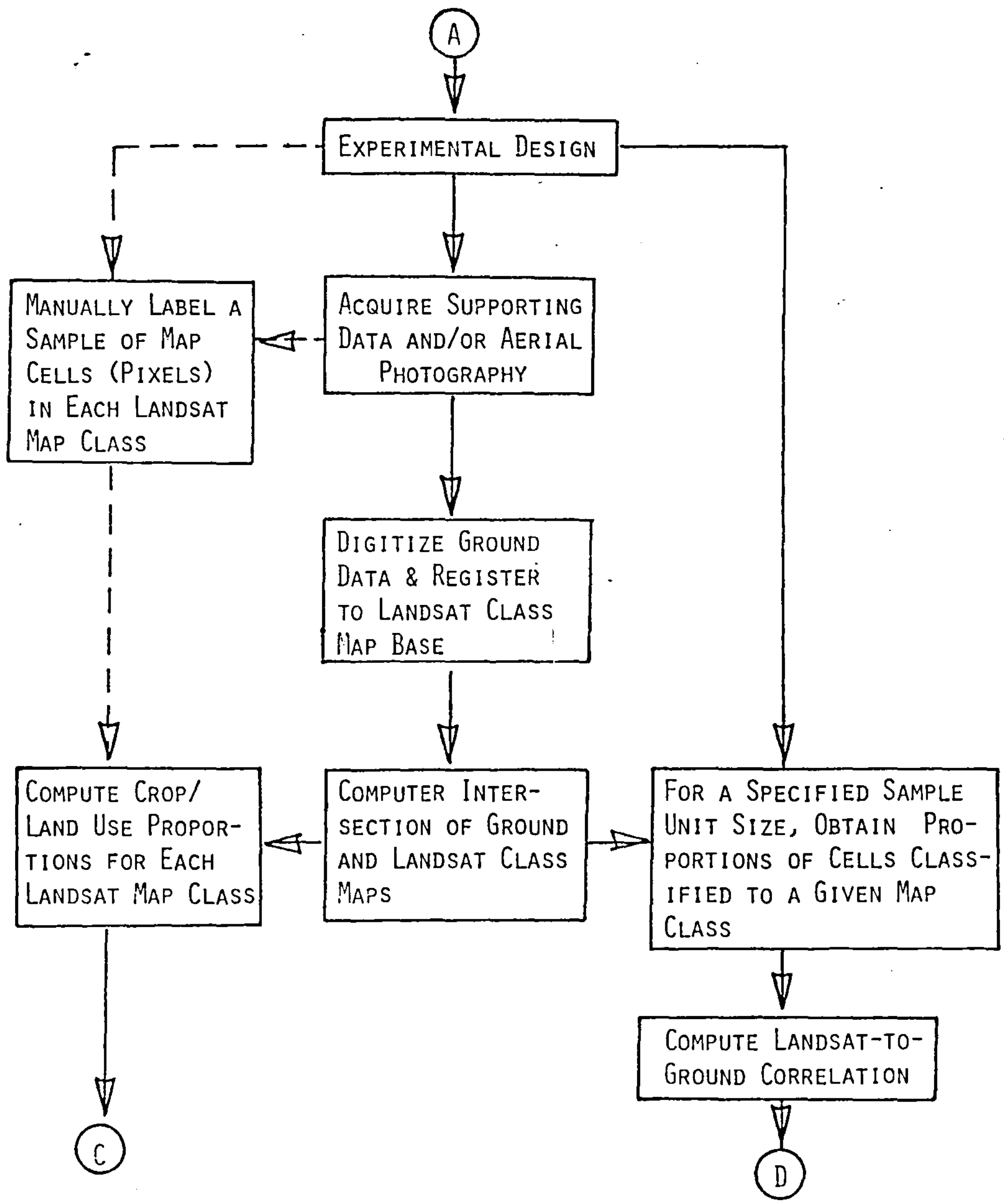


FIGURE 6-1F. SURVEY PLANNING MODEL COMPUTATION OF SAMPLE SIZE AND EXPECTED SAMPLING ERROR

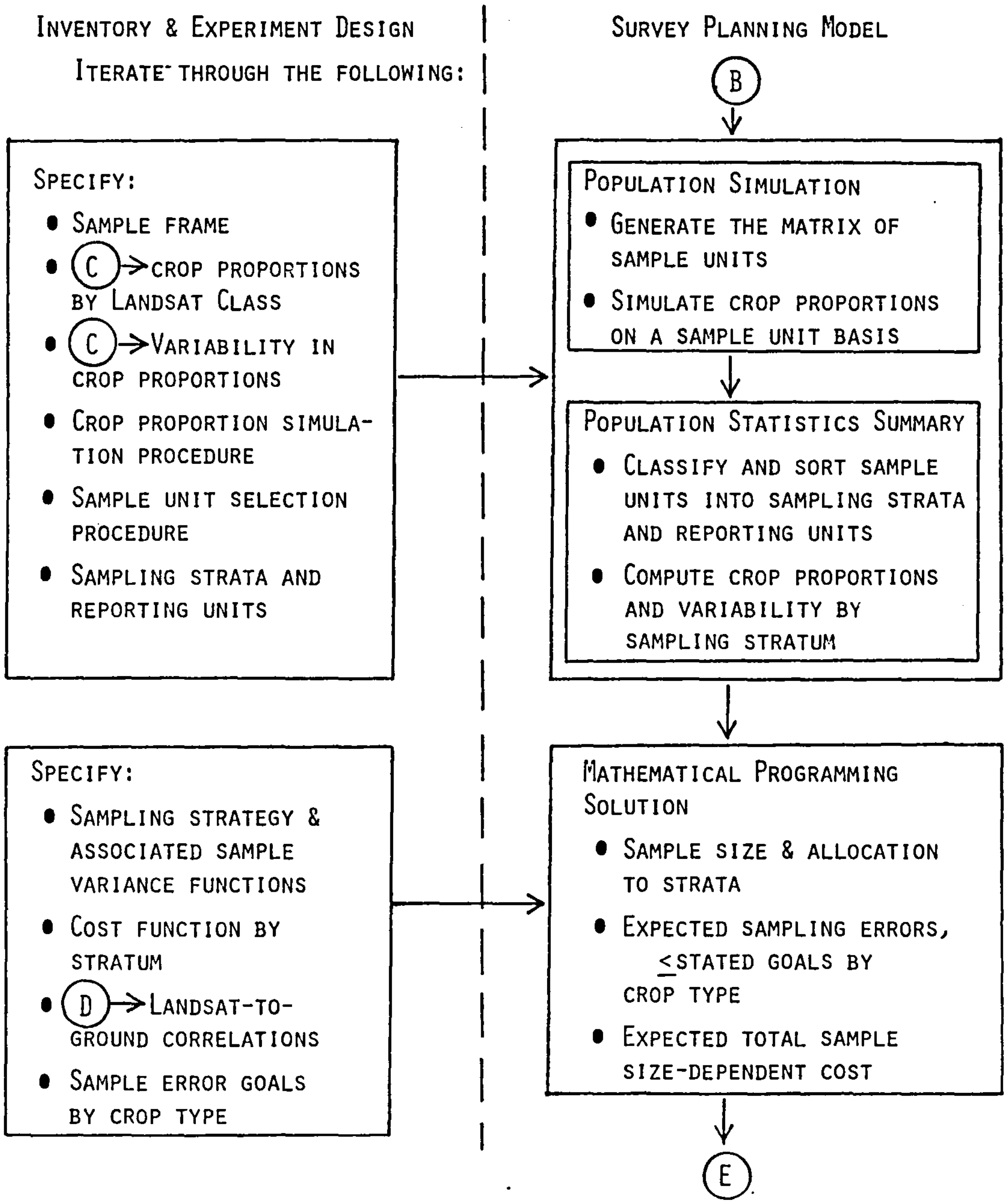


FIGURE 6-1G. RANKING ALTERNATIVE DESIGNS AND SELECTION OF FINAL SYSTEM

(E)

↓

RANK ALTERNATIVE MEASUREMENT/SAMPLING INVENTORY SYSTEMS BY

1. TOTAL VARIABLE COST FOR A RANGE OF ERROR GOALS BY CROP
2. CELL MAP ACCURACY

↓

FINAL RANKING OF DESIGN ALTERNATIVES BY ALSO CONSIDERING:

1. FIXED COST
2. ROBUSTNESS TO MISSING DATA, DATA ERRORS, OR MACHINE FAILURE
3. SPEED IN PRODUCING ESTIMATES
4. FLEXIBILITY FOR CHANGE
5. INSTITUTIONAL CONSTRAINTS

TABLE 6-4. COMPARISON OF THE TWO APPROACHES TO MULTICROP CLASSIFICATION

TYPE A:

- INTEGRATED USE OF PRESENTLY AVAILABLE CLASSIFICATION TECHNIQUES
- CLASSIFICATION BANDS:
  - SIMPLE CONCEPTUALLY
  - LOWEST COST TO GENERATE
- CLASSIFICATION PROCEDURES
  - SIMPLEST TO UNDERSTAND
  - LEAST USER TRAINING REQUIRED
  - LOWEST COST PER UNIT AREA

TYPE B:

- USE OF MORE SOPHISTICATED TECHNIQUES TO INCREASE CLASSIFICATION ACCURACY AND/OR FLEXIBILITY
- CLASSIFICATON BANDS:
  - NORMALIZATION REQUIRED
  - DEFINITION MORE COMPLEX
  - MORE EXPENSIVE TO GENERATE
- CLASSIFICATION PROCEDURES
  - MORE COMPLEX, MORE TRAINING REQUIRED
  - GREATER SAMPLING SOPHISTICATION
  - PROBABLY MORE COSTLY

The final ranking of design alternatives and selection of the design to implement will depend on both objective and subjective criteria. Figure 6-1g lists some of these criteria. To the list of objective criteria must be added an estimate of the fixed cost of implementing the inventory system. Fixed costs of operation should be identified as well. In addition, the turnaround time required to produce crop/land use estimates may be critical, and the robustness of any given inventory system to data problems likely to occur must be weighed in the selection of a system. Subjective considerations include flexibility for short term change and long term system growth to encompass larger information system objectives. The extent to which the procedure can be understood by management and technical personnel is often a key factor in the selection of a design. An associated consideration is the user expertise and hardware/software capability required for successful implementation. Other subjective factors that affect final selection of a system by a user include expected long range benefits accrueing from the planning and management value of information forthcoming from candidate inventory systems.

Obviously, the ranking and selection of inventory system designs based on 'expected performance' is not a simple, straight-forward task. The approach used in this study will be of necessity to develop a large area, crop estimation technology in a stepwise fashion. At each point in this process, the technology can be 'graded' or ranked with respect to the California DWR's evolving information goals and performance criteria. In this way, system development responsive to DWR's short and long term needs can proceed prior to the selection of a specific design for operational use.

6.2.2 Development of An Initial Baseline Multicrop Classification Procedure

Work in the Sacramento Valley addressed two general issues. First, basic spectral/temporal data is needed on the major crops of the area as input for inventory design and classification procedures. Second, DWR ultimately requires output in a map-like form, preferably the 7.5 minute quadrangle, with the capability of recombining the data shown on the map in a number of ways (e.g. by water district, county, etc.).

An area in the Sacramento Valley containing sixty-four U.S.G.S. 7.5 minute quadrangles was selected as the test site. This area was selected for several reasons: (1) the agricultural crop mix is diverse; (2) the mix is representative of much of the agriculture in Northern California; and (3) DWR had collected detailed ground data over the entire site in 1976 (see Figure 6-2).

The first step in pursuing the temporal/spectral pattern of agriculture in this area was to determine the major crops and their spatial distribution. Using County Agricultural Commissioner's reports and DWR's 7.5 minute quadrangle maps and statistical summaries, a detailed analysis of the counties in the test area (Butte, Colusa, Glenn, Sutter, and Tehama) was done. Reported crop acreages were tabulated and a crop was selected for specific study if: (1) its area represented five percent of a single county's total; (2) it occupied five percent of the area of the combined counties in the test area; or (3) was of particular interest to DWR. The crop distribution in the study area is shown in Tables 6-5a and 6-5b. The crops selected for study were rice, small grains, orchard, pasture, sorghum, corn, tomatoes, beans, and sugar beets.

FIGURE 6-2. Location of Task IV Test Site in the Sacramento Valley

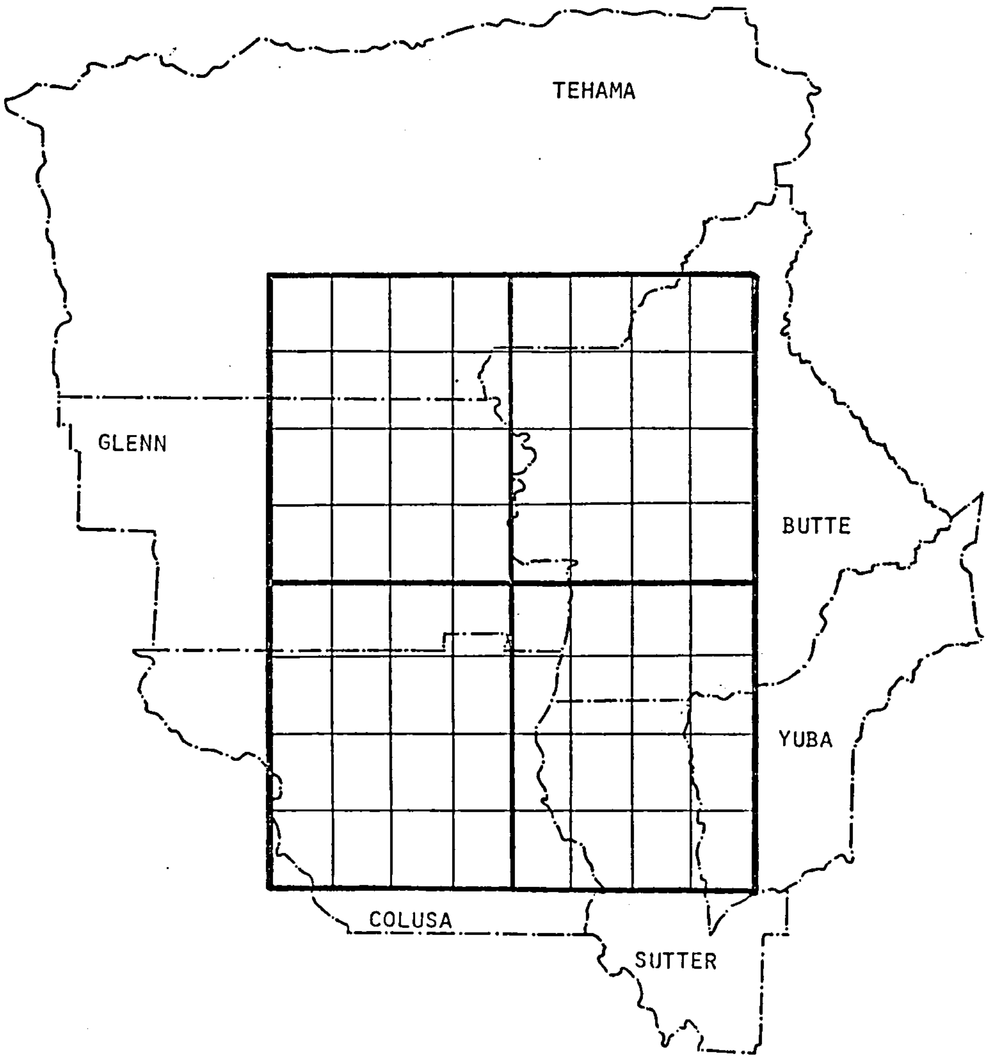


TABLE 6-5A. Crop Distribution in the Study Area

Crops - 1976	Butte	Colusa	Glenn	Sutter	Tehama	Total	% Total	
•Barley	12,000	18,600	3,500	19,000	7,100	60,200	5.1	
Beans	8,100	9,375	2,552	11,576	-	31,603	2.7	
•Corn	8,800	16,000	8,000	15,592	1,220	49,612	4.2	
•Hay, Alfalfa	6,500	3,440	16,000	9,715	4,000	39,655	3.4	
Grain	3,000	1,200	-	7,174	2,900	14,274	1.2	6.8
Other	1,150	-	3,000	20,430	2,000	26,580	2.2	
•Oats	5,000	-	-	1,043	2,500	8,543	.7	
•Pasture, Irr.	19,800	12,000	36,000	24,000	32,200	124,000	10.5	
•Rice	70,000	108,000	53,149	78,964	-	310,000	26.2	
Safflower	-	8,100	888	8,226	-	17,254	1.5	
•Sorghum	11,200	9,150	6,500	23,177	1,850	51,877	4.4	
Sugar Beets	4,040	12,800	7,222	5,648	815	30,525	2.6	
•Wheat	27,600	39,000	22,500	40,000	14,000	143,100	12.1	
•Fruit & Nut Crops	64,976	22,150	21,062	46,288	25,663	180,139	15.2	
Seed Crops	20,108	6,145	5,613	19,301	3,877	55,044	4.7	
Vegetable Crops							3.4	
Melons	-	-	-	1,556	-	1,556		
Pumpkins	-	-	-	1,265	-	1,265		
Squash	-	-	-	322	-	322		
•Tomatoes, Canning	-	8,000	-	24,500	-	32,500		
Fresh	-	-	-	85	-	85		
Corn, Sweet	-	-	-	178	-	178		
Watermelons	-	-	-	215	-	215		
Miscellaneous	1,932	-	1,350	211	95	3,588		

TABLE 6-5B. Crops Representing Approximately Five Percent of the Tabulated Acreage

	Butte	Colusa	Glenn	Sutter	Tehama	All
Barley	-	•	-	•	•	•
Corn		•		-		-
Rice	•	•	•	•		•
Wheat	•	•	•	•	•	•
Hay			•	•	•	•
Pasture	•		•	•		•
Sorghum				•		-
Fruit & Nuts	•	•	•	•	•	•
Tomatoes				•		

• = Crop represented 5% or greater of reported acreage

- = Crop represented between 4 and 5% of reported acreage

General and year-specific crop calendars were generated for each of these crops using the Crop and Livestock Reporting Service Weekly Crop and Weather Reports. These calendars show crop phenology and pertinent cultivation practices throughout the 1975-1976 crop year and were useful in selecting the appropriate Landsat acquisitions.

The dates selected were May 4, May 30, June 26, August 28, and October 3. The dates corresponding to the Task II work were used along with the early May date for small grain identification and the June date for field crop differentiation. The Sutter 30 minute block (see Figure 6-3) was selected for further analysis because of the high proportion of agriculture and the availability of the data.

For purposes of classification we wanted to test the utility of easily and inexpensively computed bands to identify crop type. We also wanted to examine the validity of using a stratification scheme based on the timing of irrigation to reduce potential classification confusion among crops.

Ratioed spectral bands were created for each of the five dates. A ratio band of MSS7 to MSS5 was created to measure the ratio of reflected infrared energy to reflected red energy. Healthy metabolizing vegetation will have a higher 7/5 value than other cover types. A ratio band of MSS5 to MSS4 was created to measure the ratio of reflected red light to reflected green light. This is a measure of vegetation senescence, that is an indication of the end of a plant's growth cycle. A Euclidean albedo band ($EB = ((MSS4)^2 + (MSS5)^2 + (MSS6)^2 + (MSS7)^2)^{1/2}$) was also 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. All fields of a given crop type greater than ten pixels in area (after eliminating border pixels) found in a systematic sample of 7½ minute quadrangles were subjected to this statistics summary. The resulting data were then plotted against time for each crop and were compared for crop separability (Figure 6-4). A crop matrix was prepared showing the best bands and dates for crop differentiation. (Figure 6-5)

The test area was stratified into general crop groupings based on irrigation timing. To identify irrigated land, a simple vegetation indicator, the 7/5 ratio band, was used following the Task II procedure. Since actively growing vegetation generally has a higher 7/5 ratio than other cover classes, a threshold 7/5 value for irrigated land could be determined for each of three dates, May 4, Aug. 28, and Oct. 3. Each point was then labeled irrigated or not on each date and a map was produced showing land irrigated at least once during the year. 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 (see Table 6-6).

FIGURE 6-3. LOCATION OF SUTTER 30' BLOCK IN THE TEST SITE

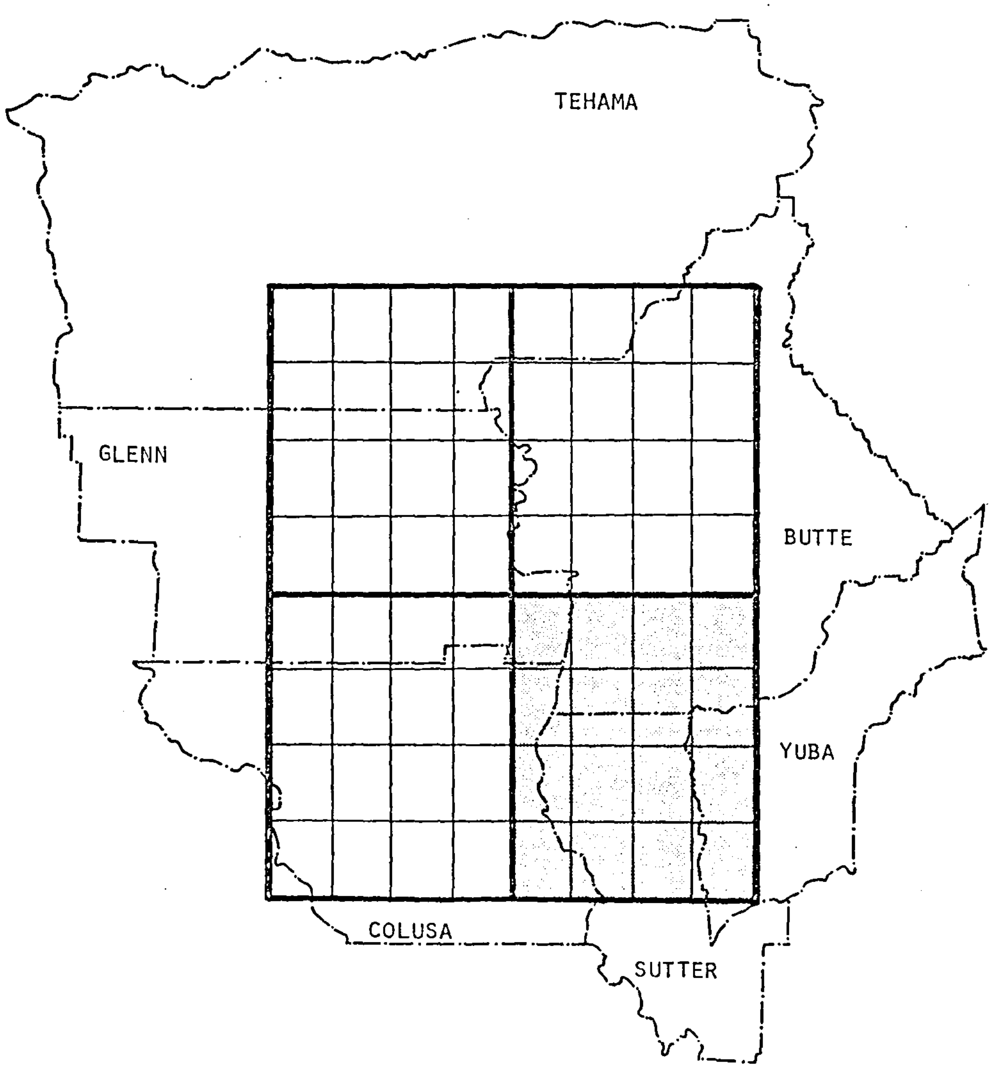


FIGURE 6-4. EXAMPLES OF THE 7/5 AND 5/4 RATIO AND EUCLIDEAN BRIGHTNESS GRAPHED AGAINST THE FIVE DATES STUDIED OVER THE 30′ BLOCK

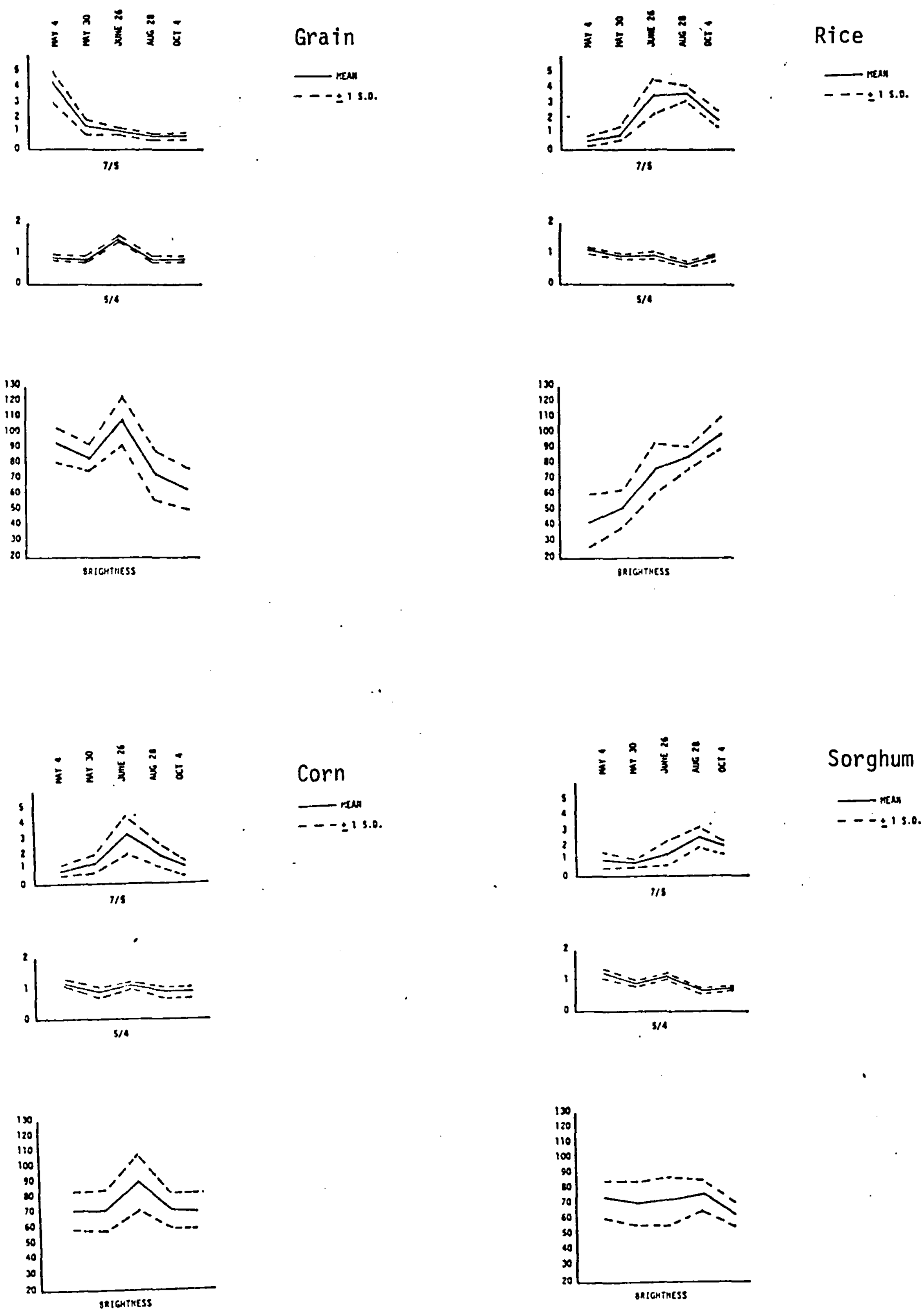


FIGURE 6-5. SPECTRAL/TEMPORAL CROP SEPARATION MATRIX

	GRAIN	RICE	SORGHUM	CORN	PASTURE	ORCHARD	TOMATOES	SUGAR BEETS	BEANS
GRAIN	-	May 4, Aug, Oct & June 7/5. June 5/4. May, June & Oct Brightness	May 4 & Aug 7/5. June, Aug & Oct 5/4. June Bright	May 4 & June 7/5. June 5/4	June, Aug & Oct 7/5. June 5/4	Aug & Oct 7/5. June 5/4. June Brightness	May 4 7/5. May 4 & June 5/4.	May 4, Aug, Oct 7/5	May 4, May 30, Aug 7/5. May 4, May 30, June 5/4. Aug Bright.
RICE	-	-	June 7/5. May 4 & Oct Bright	Aug & Oct 7/5. Oct Brightness	May 4 & May 30 7/5. May 4 & May 30 5/4.	May 4 & Aug 7/5. May 4 & Oct Bright	Aug & Oct 7/5. Oct Brightness	Oct Bright. May 4 7/5.	June & Oct 7/5. May to Aug 5/4. May & Oct Bright
SORGHUM	-	-	-	June & Oct 7/5. June Brightness	May 4 & May 30 7/5. May to June Brightness	May 30 7/5	Oct 5/4. June Bright June & Aug 7/5	June 7/5	Oct 7/5. May 30, June, Oct 5/4. Aug Brightness
CORN	-	-	-	-	May 4 & Oct 7/5. Oct 5/4. May 4 & May 30 Bright	May 4 & Oct 7/5.	June 7/5	Oct 7/5	June 7/5. May to June 5/4. Aug Brightness
PASTURE	-	-	-	-	-	Oct 7/5. May 30, Aug & Oct Bright	May 4, May 30 & Oct 7/5. May 4 Oct 5/4.	May 4 7/5. May 4 &Brightness	May, June & Oct 7/5. May June & Oct Brightness.
ORCHARD	-	-	-	-	-	-	May 4 & Oct 7/5. Aug Brightness	May 4 7/5. Aug & Oct Brightness	May to June 7/5. May 5/4. Aug Bright.
TOMATOES	-	-	-	-	-	-	-	Oct 7/5	Aug & June 7/5
SUGAR BEETS	-	-	-	-	-	-	-	-	May 30 & June 7/5.

Table 6-6. General Crop Groups and Irrigation Patterns.

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

Using the crop separability matrix, bands were chosen for each irrigation stratum to separate the crop types found within that stratum (Figure 6-5). Unsupervised classification was performed within each stratum of the 30 minute block using these input bands. This classification uses all of the digital values from the selected Landsat bands and combines them into spectrally similar groups which can then be labeled as to crop type. The final classification results for each stratum were combined producing a map for the 30 minute block with 139 land cover classes.

Ground data maps for this 30 minute block (sixteen 7.5 minute quadrangles) 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. The detailed ground data classes were then combined into thirty crop groups; for example, irrigated wheat, barley, and oats were called "irrigated small grains" and deciduous fruit varieties were called "orchard". Each spectral class was then given the label of the ground data crop group with the highest proportion within that class. (See Figure 6-6)

Preliminary evaluation indicates that certain crops and crop groups are discernable (see Figure 6-7). Small grains and rice, for instance, both have highly individualized temporal and spectral patterns making them easily identifiable. Orchard as a crop group is quite easily distinguished; however, separating the various fruit and nut varieties is not, at present, feasible with these techniques. Separating individual vegetable crops requires additional research. The crop calendars of the various vegetable crops are quite similar, with planting and harvest times overlapping. Also, many of these crops are quite similar spectrally, causing them to be easily confused.

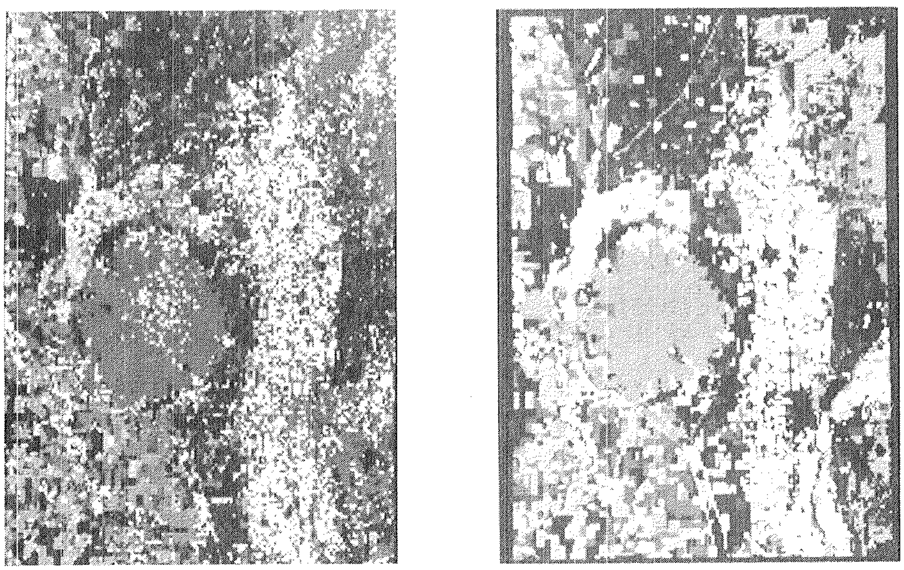


Figure 6-6. Results of Crop Type Classification on the Left. Digitized Ground Data on the Right.

FIGURE 6-7. NUMBER OF PIXELS CLASSIFIED INTO MAJOR CROP TYPES

GROUND

PASTURE	991	1128	2760	0	94	604	4	251	100	6189	10535
ORCHARD	1589	860	2030	0	91	2389	5	1174	747	44985	1054
TOMATO	390	23	971	0	310	562	57	1522	7907	281	11
BEANS	223	197	998	0	73	784	9	5498	1406	601	45
SUG BEETS	675	51	3078	0	472	1042	387	175	748	565	186
SORGHUM	370	1379	1396	0	385	5485	27	474	536	1663	191
CORN	301	273	1252	0	2058	1117	258	531	1461	718	95
SAFFLOWER	396	12	198	0	27	17	3	39	1742	74	0
RICE	939	254	93396	0	847	2653	85	957	1112	2648	337
GRAIN/DBL	496	2118	530	0	20	1273	31	234	135	1448	441
GRAIN	28189	277	453	0	54	203	0	302	391	1398	112
	GRN	GR/DB	RICE	SAFF	CORN	SORG	S BTS	BEANS	TOMA	ORCH	PAST

LANDSAT

6.2.3 Development of a Multicrop Sample Design

A preliminary effort was undertaken during the past year to implement the Survey Planning Model (SPM) portion of the multicrop design effort. The SPM is a software package developed at U.C. Berkeley during the last several years (Wensel et al 1979, 1981). It is designed to assist in sample survey planning through use of population simulation techniques and nonlinear programming procedures for determining sample allocation. When coupled with an appropriate experimental design, use of the SPM enables a systematic evaluation of the cost/error impact of alternative specifications for sample frame, stratification, sample selection, and estimation procedure. In addition, the impact of alternative classification procedures on estimation error can be evaluated at least indirectly with the SPM.

The objectives of this preliminary DWR multicrop effort were to (a) modify the SPM software to allow less expensive and more varied simulation of sample unit population characteristics over large areas, and to (b) apply the SPM to a multiple crop estimation problem on an initial test area. SPM software modifications centered on a new module to create first or second stage rectangular sample units within a digital grid cell data base. This grid cell base would represent the DWR land use data file in an operational system. Once defined, the new SPM module allows rapid simulation of crop proportions for each Landsat spectral class occurring within each sample unit. A vector of crop proportions can then be obtained directly for each sample unit and used to compute within and between sample unit crop covariances. The new SPM module also allows stratification of sample units according to reporting unit and land use strata as well as assignment of sample units to strata on the basis of characteristics (e.g. crop proportions) associated with each sample unit.

In addition to the SPM module described above, an additional sample allocation alternative was implemented. This new module, enabling calculation of sample size for stratified regression estimation, was linked with the nonlinear programming software already included in the SPM. Regression sampling was added as it is presently considered a primary candidate for the DWR multicrop estimation problem. Previously implemented sample allocation alternatives include stratified random sampling, stratified two stage sampling, and stratified two stage, two phase sampling.

Test of Multiple Crop Sample Allocation on the 30' Sutter Block

A test problem was defined for the 30' Sutter block (sixteen 7½' quadrangles covering approximately one half million acres) in the central portion of the Sacramento Valley. The crop mix and Landsat classification procedures for this block were described earlier in Section 6.2.2. The objective of this test was to use the Survey Planning Model to compute the sample allocation required to simultaneously estimate the area of four crop categories, each to within plus or minus 10 percent of the estimate at the 90 percent level of confidence.

To obtain the required sample allocation, the following procedure was employed. First, the 30' Landsat class map was input to the SPM. This map, rectified to a north-south orientation, contained 30 spectral classes defined by grouping the much larger original number of spectral classes according to the dominant ground truth class associated with each. The SPM was then instructed to partition this class map into a contiguous matrix of approximately one square mile (25 cells by 25 cells) sample units.

A simple procedure for simulating crop proportions for each sample unit was then selected. In essence, each group of cells within a given sample unit belonging to a given spectral class was assigned the label of the dominant ground category associated with that Landsat spectral class. Division by the total number of cells in the sample unit gave a proportion value for each ground category. While more elegant simulation methods were available, this one was selected as an inexpensive baseline.

The resulting crop/land use proportions for each sample unit were summed into five crop/land use groups of estimation interest. These were small grains, rice, orchards, combined field and truck crops (corn, grain sorghum, beans, tomatoes, and pasture), and others. Each sample unit was then sorted by the SPM into one of five strata. In this example the five strata were defined in terms of the five crop/land use groups just described. Sample unit assignment to strata was based on the 'plurality rule': a given unit was assigned to a given stratum if the cells belonging to the crop group associated with that stratum out-numbered cells belonging to other crop groups. Table 6-7 shows the resulting crop composition for each of the four agricultural strata as a percent of total stratum area. The diagonal of this table shows that the plurality rule produced strata that were dominated by the crop group of interest.

As a summary function, the SPM population simulation module produced a map (Figure 6-8) and count of sample units falling into each of the five strata. The crop/land use composition of each stratum was also reported (e.g. as in Table 6-7) as was the between sample unit covariance matrix for each of the four crop categories within each stratum.

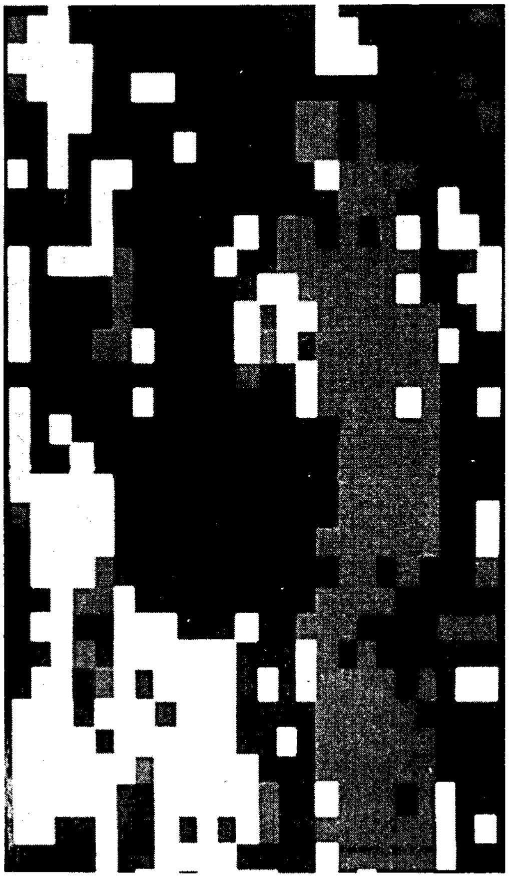
This SPM characterization of the sample unit population was then input into the SPM sample allocation module. The core of this module consists of the Sequential Unconstrained Minimization Technique (SUMT) software for minimization of an objective function subject to some set of constraints. SUMT was developed by Fiaco and McCormick (1968) as a flexible nonlinear programming package applicable to a wide variety of minimization problems. Titus (1977) adapted this package to the sample allocation problem. In effect, SUMT as implemented in the SPM minimizes the cost that varies with the number of sample units taken in the sample, subject to meeting sampling precision goals for each item under consideration.

To use SUMT, the cost function must be specified for the sample design in question. A simple cost function of the form

$$C = c_h\, n_h + c_h'(n_n)^{\frac{1}{2}}$$

was selected to represent the sample size-dependent cost for the stratified regression design. In this equation c_h represented the cost associated with

FIGURE 6-8. SPM CLASSIFICATION OF APPROXIMATELY ONE SQUARE MILE AREAS INTO FOUR SAMPLING STRATA



KEY:

SMALL GRAINS : WHITE
RICE : RED
ORCHARDS : GREEN
FIELD/TRUCK CROPS : BLUE

TABLE 6-7. CROP COMPOSITION OF STRATA AS A PERCENT OF STRATUM AREA

Stratum \ Crop	Small Grains	Rice	Orchards	Field/ Truck
Small Grains	44.7	10.5	9.3	16.5
Rice	7.1	62.6	6.1	11.1
Orchards	6.9	5.4	51.7	13.9
Field/ Truck	11.2	15.3	11.3	46.1

selecting, preparing, mapping, labelling, and digitizing each of the n_h sample units requiring ground measurement in stratum h. c_h' represented the travel cost associated with an initial sample unit. A value of $100 was chosen for c_h based on Task I experience and a best guess adjustment for mapping crop type as opposed to irrigated-only on a sample unit basis. This value was used for all strata pending further data. Similarly, a figure of $21.50 was chosen to represent c_h' in all strata.

Use of SUMT also required specification of the expected Landsat-to-ground correlation for each crop group of interest within each stratum. These correlations comprise part of the formula for regression sample variance - a formula which is in turn used by SUMT as the error constraint function. Time did not permit a direct calculation of these correlations. Thus, for this test example, best guesses of correlations were made based on inspection of preliminary classification accuracy results. The Landsat-to-ground correlations assumed were

.8 for small grains,

.9 for rice,

.85 for orchards, and

.7 for combined field and truck crops.

These correlations were assumed to hold in all strata pending further results.

Given these specifications and several others relating to control of the minimization algorithms, SUMT was requested to minimize the cost function cited above subject to the following constraints. These constraints were that (1) regression variance for each of the four crop groups be held to less than or equal to 10 percent at the 90 percent confidence level, and that (2) sample size within each stratum fall between 4 (in order to give at least one degree of freedom) and the total number of sample units within that stratum.

Sample Allocation Results for the 30' Sutter Block

Table 6-8 presents the SPM estimate of sample allocation required to achieve less than or equal to 10 percent sampling error at the 90 percent level of confidence for each of the four crop categories. The total (population) size for each stratum is listed in the column on the left, while the column on the right gives the required ground sample size.

Of the 545 sample units falling within the four sampling strata of the 30' Sutter block, an estimated 73 required ground measurement in order to achieve the stated error goals. This represents a sampling rate of 13.4 percent. Operationally, this percentage would be lower as the sample would be allocated over a much larger area. It should be noted that the fifth stratum, labelled 'other', was excluded as a sampling stratum in this example. Exclusion resulted from the fact that this stratum was dominated by non-agricultural cover types. In an operational system, the pockets of agricultural area within this stratum would be included within the sampling frame, thereby raising the ground sample size required.

Overall, this test of the SPM indicated that sample design for simultaneous estimation of several crop types and/or groups is feasible. Use of the Landsat class map in developing a meaningful set of multicrop sampling strata appeared

TABLE 6-8. SPM SAMPLE SIZE RESULTS FOR THE FOUR STRATA, FOUR-PARAMETER ESTIMATION PROBLEM IN THE 30' SUTTER BLOCK

A. SAMPLE SIZE

STRATUM	POPULATION SIZE	SAMPLE SIZE TO ACHIEVE ± 10% AT 90% CL FOR EACH CROP GROUP
SMALL GRAINS	57	11
RICE	213	29
ORCHARDS	154	15
FIELD/TRUCK CROP GROUP	121	18
TOTAL	545	73

to be especially helpful. Work during the coming year will seek to expand this effort to a larger area and to examine a number of alternative inventory designs.

Work Planned During the Coming Year

The objective of the Irrigated Lands APT multicrop work during the coming year will be to: develop and demonstrate an initial, end-of-season multicrop estimation and mapping procedure. Estimation and mapping targets will include crop types, or crop groups having significant impact on water use. Sample precision goals will be ± 10 percent at the 90 percent level of confidence on the most significant water use groups. Map product goals include (a) achievement of field labelling accuracies greater than or equal to 70 percent on important target categories, (b) generation of map products in transparency form for easy projection onto USGS 7½' quadrangle sheets, and (c) development of procedures for creating and manipulating digital class maps in an interactive system environment.

The general approach proposed to achieve these goals was outlined in the multicrop flow charts provided earlier. Specific work scheduled for the coming year will focus on a one degree by one degree block covering the heart of the Sacramento Valley. This area has been chosen as the study site due to a fairly complete set of Landsat full frame acquisitions in 1976 and corresponding wall-to-wall California DWR ground data for the same year.

Within the limits of the University's resources for the coming year, emphasis in this effort will be placed on:

(1) Establishment of a data handling pipeline to include Landsat and ancillary data registration, preprocessing, classification, sample system interface, and products generation;

(2) Development of a simple technique for multicrop classification in the Sacramento Valley which will use

    (a) a Landsat greeness indicator to form spectral strata,

    (b) clustering within spectral strata within 30' blocks to define spectral classes, and

    (c) a simple (e.g. Euclidean distance) or more complex (e.g. maximum likelihood) rule to assign ground file cells to spectral classes;

(3) Conduction of a performance assessment which will include

    (a) class map accuracy assessment relative to California DWR ground data, and

    (b) use of the Survey Planning Model to evaluate alternative multicrop inventory designs; this assessment will include

        i) specification of an experimental design for evaluating inventory system components,

ii) completion of SPM modifications to allow evaluation of candidate combinations of sample frame, sample design, classification, and estimation procedure,

iii) SPM analysis of inventory component impact on estimate sampling precision and cost, and

iv) SPM estimation of expected sampling precision (by crop type or group) and total variable cost for specific inventory designs.

LITERATURE CITED

Cicone, R., C. M. Hay, R. Horvath, M. Metzler, O. Mykolenko, J. Odenweller, and D. Rice. 1981. Users Manual for the U.S. Baseline Corn and Soybean Segment Classification Procedure. Environmental Institute of Michigan (R. Horvath, Program Manager) and University of California at Berkeley (R. N. Colwell, Principal Investigator). NASA-Johnson Space Center Publication No. FC-E1-00713. NASA Contracts NAS9-15476 and NAS 9-14565. 267 pp.

Best, M. J. 1972. FCDPAK: A FORTRAN IV Subroutine to Solve Differentiable Mathematical Programmes. Department of Combinaterics and Optimization, University of Waterloo, Waterloo, Canada. 25 pp.

Cochran, W. G. 1977. Sampling Techniques, Third Edition. John Wiley and Sons, Inc., New York. 428 pp.

Ferchaud, C. 1981. Personal Communication.

Fiacco, A.V. and G. P. McCormick. 1968. Non-linear Programming: Sequential Unconstrained Minimization Techniques. Wiley and Sons, New York. 210 pp.

Goodman, L. A. and H. O. Hartley. 1958. The Precision of Unbiased Ratio-Type Estimators. American Statistical Association Journal, June. p. 491-508.

Hay, C. M., R. W. Thomas, and A. S. Benson. 1977. Development of Techniques for Producing Static Strata Maps and Development of Photointerpretive Methods Based on Multitemporal Landsat Data. R. N. Colwell, Principal Investigator. Space Sciences Laboratory Series 18, Issue 1.

Scheffé, H. 1959. The Analysis of Variance. John Wiley and Sons, Inc., New York. 477 pp.

Titus, S. J. 1979. Design of Multivariate Surveys for Sampling Wildland Ecosystems. Ph.D. Thesis, Department of Forestry and Conservation, University of California at Berkeley. 123 pp.

Wall, S.L., S.J. Titus, D. K. Noren, and J. M. Sharp. 1977. An Inventory of Irrigated Lands for Selected Counties within the State of California Based on Landsat and Supporting Aircraft Data. R. N. Colwell, Principal Investigator. Final Report for NASA-Goddard Space Flight Center Contract No. NAS5-20969. University of California at Berkeley, Space Sciences Laboratory Series 18, Issue 50. 52 pp.

Wall, S. L., C. E. Brown, S. J. Daus, and V. L. Vesterby. 1978. Determining the Usefulness of Remote Sensing for Estimating Agricultural Water Demand in California. R. N. Colwell, Principal Investigator. Annual Report for Ames Research Center NASA Grant NSG 2207. University of California at Berkeley, Space Sciences Laboratory.

Wall, S. L., et. al. 1980. Irrigated Lands Assessment for Water Management Applications Pilot Test. R. N. Colwell, Principal Investigator. Final Report for Ames Research Center Grant NSG 2207. University of California, Space Sciences Laboratory Series 21, Issue 5.

Wensel, L. C., S. J. Titus, and R. W. Thomas. 1979. Sample Allocation in Multiresource Inventories. Proceedings, Society of American Foresters Workshop on Forest Resource Inventories. Colorado State University, Ft. Collins.

Wensel, L. C., M. Eriksson, R. W. Thomas, and D. R. Taylor. 1981. Survey Planning Model - User's Manual. Part of final report for USFS, Rocky Mountain Station Cooperative Agreement No. 16-935-CA and NASA Johnson Space Center Contract No. NAS 9-15987. Department of Forestry and Resource Management & Remote Sensing Research Program, University of California, Berkeley.

APPENDIX I

IA - Equations Used for Estimation of Basin and Statewide Irrigated Area

IB - Equations Used for Estimation of Error Associated with Estimates of Irrigated Land

IC - Equations Used for Estimation of County Irrigated Area and Associated Error

APPENDIX IA: Equations Used for Estimation of Basin and Statewide Irrigated Area

Part 1 - Stratified Regression Estimation

Landsat and ground observations (measurements) were expressed in terms of proportion area irrigated, as opposed to area irrigated. This was done to minimize error due to differences in determing total digitized area of matched Landsat and ground sample units. In addition, the measurements of proportion were weighted by the relative size of sample unit with which they were associated. Sample unit weights were expressed as the area in the sample unit relative to the average area in all sample units in the given land use stratum. Thus the sample unit observations could be expressed mathematically as

$$x_{hi} = w_{hi}\, u_{hi} = \frac{A_{hi}}{\bar{A}_h}\, u_{hi} \tag{1}$$

and

$$y_{hi} = w_{hi}\, v_{hi} = \frac{A_{hi}}{\bar{A}_h}\, v_{hi} \tag{2}$$

where

x_{hi} = weighted Landsat-measured proportion irrigated for sample unit i in stratum h,

y_{hi} = weighted ground-measured irrigated proportion for sample unit i in stratum h,

u_{hi} = unweighted Landsat-measured irrigated proportion for sample unit i in stratum h,

v_{hi} = unweighted ground-measured sample unit irrigated proportion in sample unit i in stratum h,

w_{hi} = relative weight for sample unit i in stratum h,

A_{hi} = size, in acres, of sample unit i in stratum h as measured on the Landsat interpretation base, and

$\bar{A}_h$ = average size, in acres, of all sample units in stratum h as measured on the Landsat interpretation base.

Once these values were computed, an estimate of average Landsat and ground irrigated proportion in a given stratum was obtained by taking the simple mean of the matched sample unit observations in that stratum. Hence

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} \qquad (3)$$

and

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \quad , \qquad (4)$$

where

$\bar{x}_h$ = estimate of the average Landsat proportion irrigated in stratum h based on sample units having matched ground data in that stratum,

$\bar{y}_h$ = estimate of the average ground proportion irrigated in stratum h based on ground sample units in that stratum,

n_h = number of spatially-matched Landsat and ground sample units in stratum h; also known as the ground sample size in stratum h,

Σ = a symbol (sigma) that indicates summation, in this case summation of sample unit measurement observations from the first (i=1) to the last ($i=n_h$) in stratum h, and the x_{hi} and y_{hi} are as defined previously.

The stratum-wide regression estimate of ground irrigated proportion was then defined as

$$\hat{\bar{Y}}_h = \bar{y}_h + b_h\,(\bar{X}_h - \bar{x}_h) \quad , \qquad (5)$$

where

$\hat{\bar{Y}}_h$ = estimate of stratum-wide, ground irrigated proportion for stratum h

b_h = estimated slope of the regression line; can be interpreted as the change in y with a unit (i.e. one integer) change in x; mathematically

$$b_h = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{\sum_{i=1}^{n_h} (x_{hi} - x_h)^2} \quad , \qquad (6)$$

$\bar{X}_h$ = average Landsat proportion irrigated based on measurements from all sample units in stratum h, i.e. from the matched sample units plus all the remaining units, and $\bar{x}_h$ and $\bar{y}_h$ are as defined previously.

Equation 5 can be rearranged to read

$$\hat{\bar{Y}}_h = (\bar{y}_h - b_h \bar{x}_h) + b_h \bar{X}_h) \quad . \tag{7}$$

Letting the term in brackets on the right side of Equation 7 equal a_h we obtain

$$\hat{\bar{Y}} = a_h + b_h \bar{X}_h \tag{8}$$

for the estimate of stratum-wide ground irrigated proportion in stratum h. Reference to Figure 1 allows physical interpretation of this equation.

Assume that the vertical axis represents $\hat{\bar{Y}}_h$ and the horizontal axis represents $\bar{X}_h$. Then the constant a_h can be seen to be the value of $\hat{\bar{Y}}_h$ at the point the regression line intercepts the $\hat{\bar{Y}}_h$ axis. That is, if the value of the average, stratum-wide Landsat irrigated proportion ($\bar{X}_h$) was zero, then the stratum-wide ground estimate $\hat{\bar{Y}}_h$ would equal a_h. The slope term, b_h, represents the slope of the regression line as drawn. Thus the regression line is completely determined (and represented) by Equation 8. Once a_h and b_h have been determined in a given basin for a given inventory, the procedure for estimating $\bar{Y}_h$ can be seen to be simply (1) obtaining $\bar{X}_h$ from the digitized Landsat interpretation and (2) substituting it into Equation 8.

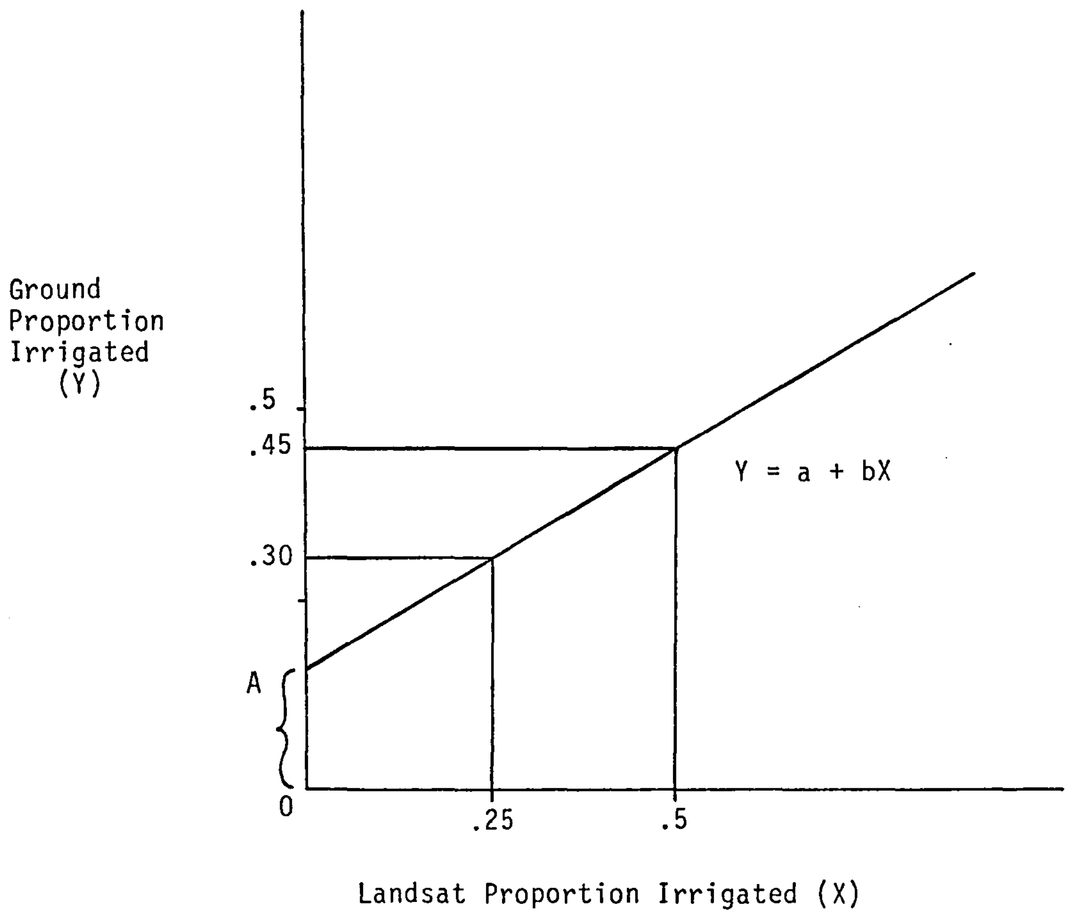
The estimate of acreage irrigated within a stratum was produced by multiplying the resulting $\hat{\bar{Y}}_h$ (a proportion) by the corresponding number of acres in the stratum (A_h). A_h was determined from digitization of stratum boundaries. Finally, the basin-wide, within-sample frame estimate of acreage irrigated at least once during 1979 was obtained by addition of the estimates of irrigated acreage in each stratum. That is

$$\hat{I}_{\substack{\text{within-frame,}\\ \text{stratified}}} = \sum_{h=1}^{L} A_h \hat{\bar{Y}}_h \quad , \tag{9}$$

where

$\hat{I}_{\substack{\text{within frame,}\\ \text{stratified}}}$ = estimate of basin-wide ground irrigated acreage for area within the sample frame,

Figure 1. Schematic for Linear Regression Line



a = regression line intercept

b = regression line slope

L = total number of sample frame strata within the given basin,

A_h = total area, in acres, of stratum h in the given basin as measured on the Landsat interpretation base, and h and $\hat{\bar{Y}}$ are as defined previously.

The within-frame estimate of irrigated acreage can also be computed by forming a weighted sum of stratum irrigated proportion estimates and then multiplying the resulting sum by the total acreage within the frame. Algebraically

$$\hat{I}_{\text{within-frame, stratified}} = A \sum_{h=1}^{L} \left(\frac{A_h}{A}\right) \hat{\bar{Y}}_h \quad , \tag{10}$$

where

A = total area, in acres, within the sample frame in the given basin as measured on the Landsat interpretation base,

i.e. $A = \sum_{h=1}^{L} A_h$,

$\frac{A_h}{A} = W_h$ = a weight for stratum h representing the proportion of the sample frame within the given basin occupied by stratum h, and all other terms are as defined previously.

Removing the A in front of the summation in Equation 10 gives the within-frame estimate of basin-wide irrigated <u>proportion</u> instead of acreage:

$$\hat{\bar{Y}}_{b,\ \text{within-frame, stratified}} = \sum_{h=1}^{L} W_h \hat{\bar{Y}}_h \tag{11}$$

where b is a basin index. Note that Equation 10 can be rearranged to yield Equation 9, viz

$$\hat{I}_{\text{within-frame, stratified}} = A \sum_{h=1}^{L} \left(\frac{A_h}{A}\right) \hat{\bar{Y}}_h$$

$$= \left(\frac{A}{A}\right) \sum_{h=1}^{L} A_h \hat{\bar{Y}}_h$$

$$= \sum_{h=1}^{L} A_h \hat{\bar{Y}}_h \quad .$$

The estimate of total irrigated acreage within a given basin was obtained by adding the within-frame estimate to the area identified by Landsat interpretation as irrigated outside the sample frame. Thus

$$\hat{I}_{\substack{\text{total,}\\ \text{stratified}}} = \hat{I}_{\substack{\text{within-frame,}\\ \text{stratified}}} + I_{\substack{\text{exclusion}\\ \text{area}}} + I_{\substack{\text{outside}\\ \text{frame}}} , \tag{12}$$

where

$\hat{I}_{\text{total, stratified}}$ = estimate of total irrigated area within a given basin,

$\hat{I}_{\text{within-frame, stratified}}$ = estimate of within-frame irrigated area defined in Equation 9,

$I_{\text{exclusion area}}$ = direct measurement of total irrigated area within exclusion areas (areas inside the contiguous boundaries of the sample frame area that have been excluded from the sample frame) based on interpretation and digitization of Landsat imagery; no calibrating ground data available; and

$I_{\text{outside frame}}$ = direct measurement of total irrigated area found in locations outside the contiguous sample frame; based on interpretation and digitization of Landsat imagery; no calibrating ground data available.

A statewide estimate of irrigated acreage was constructed by adding the separate basin estimates together:

$$\hat{I}_{\substack{\text{statewide total,}\\ \text{stratified}}} = \sum_{b=1}^{B} \hat{I}_{\substack{b,\text{total,}\\ \text{stratified}}} \tag{13}$$

where the basin index b has been added to the term on the right obtained in Equation 12, and the summation is taken over the B (= 10) basins. In a similar fashion an estimate of statewide proportion irrigated was produced for only the area within the sample frame. This estimate was obtained by forming a weighted average of the separate basin estimates, where the weight for each basin was proportional to the area within the sample frame in the given basin relative to the total area within the sample frame statewide. Thus

$$\hat{\bar{Y}}_{\substack{\text{statewide,}\\ \text{stratified}}} = \sum_{b=1}^{B} \left(\frac{A_b}{A_s}\right) \hat{\bar{Y}}_{\substack{b,\text{within-frame,}\\ \text{stratified}}} \tag{14}$$

where

$\hat{\bar{Y}}_{\text{statewide, stratified}}$ = estimate of proportion of area within the statewide sample frame irrigated at least once during the calendar year using a stratified sample within basins,

$\hat{\bar{Y}}_{\text{b,within-frame, stratified}}$ = estimate of proportion irrigated within sample frame of basin b from Equation 11,

A_b = area within the sample frame in basin b as measured on the Landsat interpretation base,

A_s = total area within the sample frame statewide obtained by adding the A_b over basins, i.e.

$$A_s = \sum_{b=1}^{B} A_b \text{ , and}$$

$\frac{A_b}{A_s} = W_b$ = weight for a given basin.

Part 2 - Summary for Stratified Regression Estimation

Summarizing for the stratified regression estimation procedure, the following tabled values correspond to the equations just presented:

1) the within-sample frame estimate of basin irrigated proportion shown in Table 1a, column 1 (counting left to right) and column A2 of Table 2a was produced by application of equation 11;

2) the within-sample frame estimate of statewide irrigated proportion shown in the last row of Table 1a, column 1 and in the last row of Table 2a, column A2 was produced by application of Equation 14;

3) the total within-sample frame estimate of basin-wide ground irrigated acreage shown in column A3 of Table 2a was produced by application of Equation 10;

4) the sum of Landsat-measured irrigated acreage in exclusion areas and in areas outside the contiguous sample frame (i.e. $I_{\text{exclusion area}} + I_{\text{outside frame}}$) is reported in Table 2b, column B3;

5) the estimate of total basin-wide acreage irrigated at least once in 1979 as shown in Table 2b, column A3+B3 (rightmost column), was produced by application of equation 12;

6) measurements of number of acres within sample frame, acres in excluded areas, and acres in areas outside the sample frame were obtained by digitization of the Landsat interpretation base and reported in Tables

TABLE 1A. RESULTS OF 1979 STATEWIDE INVENTORY OF IRRIGATED LAND

BASIN	STRATIFIED ESTIMATE (PERCENT OF AREA IRRIGATED IN SAMPLE FRAME)	100 x ABSOLUTE S.E. AT 95% C.L.	RELATIVE S.E. AT 95% C.L.
NORTH COAST	53.52	2.04 *	3.81
SAN FRANCISCO	21.85	1.22 *	5.56
CENTRAL COAST	31.91	1.84 *	5.77
SOUTH COAST	45.79	2.88	6.28
COLORADO DESERT	82.15	1.40 *	1.70
SOUTH LAHONTAN	27.38[A]	3.81[A]	13.91[A]
NORTH LAHONTAN	58.73	2.68	4.56
SACRAMENTO	65.38	1.80 *	2.75
SAN JOAQUIN	74.78	2.55 *	3.41
TULARE	82.04	2.00 *	2.44
STATE	67.09	95: 0.89 99: 1.17	95: 1.32 99: 1.74

TABLE 1B. FOR COMPARISON ONLY - UNSTRATIFIED RESULTS
OF 1978 STATEWIDE INVENTORY OF IRRIGATED LAND

BASIN	UNSTRATIFIED ESTIMATE (PERCENT OF AREA IRRIGATED IN SAMPLE FRAME)	100 x ABSOLUTE S.E. AT 95% C.L.	RELATIVE S.E. AT 95% C.L.
NORTH COAST	53.18	2.39	4.49
SAN FRANCISCO	21.19	2.20	10.37
CENTRAL COAST	32.58	2.15	6.60
SOUTH COAST	45.25	2.40	5.31
COLORADO DESERT	82.25	1.30	1.58
SOUTH LAHONTAN	27.38	3.81	13.91
NORTH LAHONTAN	58.73	2.45	4.17
SACRAMENTO	65.44	1.63	2.49
SAN JOAQUIN	75.16	2.31	3.08
TULARE	81.46	2.09	2.57
STATE	67.04	95: 0.88 99: 1.15	95: 1.31 99: 1.72

Table 2a. Stratified summary statistics for the area within the sample unit frame. Regression with factor 5.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53521	321070	6029	12238
San Francisco	191654	0.21852	41880	1108	2329
Central Coast	1380040	0.31906	440316	12572	25420
South Coast	598866	0.45787	274203	8510	17223
Colorado Desert	818231	0.82147	672152	5670	11447
South Lahontan	235626	0.27383	64522	4402	8977
North Lahontan	175456	0.58726	103038	2297	4695
Sacramento	3388466	0.65381	2215413	30327	60823
San Joaquin	2788914	0.74778	2085494	35419	71145
Tulare	4080305	0.82038	3347400	40721	81769
State	14257457	0.67091	9565489	64477	127020

Table 2b. Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346785
San Francisco	512	2586376	5623	2778543	47503
Central Coast	13475	5789056	24725	7182571	465041
South Coast	62133	6289499	63810	6950499	338013
Colorado Desert	28421	11852213	10830	12698865	682982
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117981
Sacramento	211744	13452904	37823	17053114	2253236
San Joaquin	542467	6704753	51098	10036134	2136592
Tulare	123300	5977461	42352	10181065	3389752
State	1000375	85067824	294255	100325656	9859744

2a and 2b, columns A1, B1, and B2, respectively; total basin acreage was obtained by addition of these three columns and reported in Table 2b, column A1+B1+B2; and

7) the estimate of the total statewide acreage irrigated at least once during 1979 shown in Table 2b, column A3+B3, was produced by application of Equation 13.

Part 3 - Unstratified Regression Estimation

The unstratified regression estimator of basin-wide, within-frame irrigated proportion was

$$\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{unstratified}}} = \bar{y} + b(\bar{X} - \bar{x}) \quad , \tag{15}$$

where $\bar{x}$ and $\bar{y}$ were the average Landsat and ground measured irrigated proportions,* respectively, within sample units in the given basin having spatially-matched Landsat and ground data; b represented the estimated regression line slope based on all spatially-matched sample units within the given basin; and $\bar{X}$ represented the average Landsat-measured irrigated proportion for all sample units within the given basin. Multiplying Equation 15 by the total area (A) within the sample frame within the given basin gave an estimate of $I_{\substack{\text{within-frame,}\\ \text{unstratified}}}$, the irrigated acreage within the sample frame, viz

$$\ddot{I}_{\substack{\text{within-frame,}\\ \text{unstratified}}} = A\,\bar{Y} \quad . \tag{16}$$

Addition of $\hat{I}_{\substack{\text{within-frame,}\\ \text{unstratified}}}$ to $I_{\substack{\text{exclusion}\\ \text{area}}}$ and $I_{\substack{\text{outside}\\ \text{frame}}}$ discussed earlier in the case of stratified sampling gave an estimate of $I_{\substack{\text{total,}\\ \text{unstratified}}}$, the total irrigated acreage in the given basin. Statewide figures were generated in a method analogous to that used in the stratified case.

Estimates of basin irrigated proportion for within-sample frame area produced by application of Equation 15 are shown in column 1 of Table 1b, and in column A2 of Table 3a. Within-frame estimates of irrigated acreage corresponding to Equation 16 are given in column A3 of Table 3a.

---

*where each x_i and y_i was weighted by the ratio of the area of sample unit i to the average area of sample units in the given basin in a manner analogous to Equations 1 and 2 for stratified sampling; see also Appendix II.

Table 3a. Unstratified summary statistics for the area within the sample unit frame. Regression with factor 5.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53182	319037	7121	14320
San Francisco	191654	0.21192	40615	3653	4213
Central Coast	1380040	0.32579	449603	14904	29671
South Coast	598866	0.45251	270993	7228	14385
Colorado Desert	818231	0.82245	672954	5294	10604
South Lahontan	235626	0.27383	64522	4402	8977
North Lahontan	175456	0.58725	103036	2118	4300
Sacramento	3388466	0.65443	2217513	27684	55232
San Joaquin	2788914	0.75164	2096260	32379	64480
Tulare	4080305	0.81458	3323734	42721	85401
State	14257457	0.67040	9558269	63484	125063

Table 3b. Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	344751
San Francisco	512	2586376	5623	2778543	46238
Central Coast	13475	5789056	24725	7182571	474328
South Coast	62133	6289499	63810	6950499	334803
Colorado Desert	28421	11852213	10830	12698865	683784
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117979
Sacramento	211744	13452904	37823	17053114	2255337
San Joaquin	542467	6704753	51098	10036134	2147357
Tulare	123300	5977461	42352	10181065	3366086
State	1000375	85067824	294255	100325656	9852524

APPENDIX IB: Equations Used for Estimation of Error Associated With Estimates of Irrigated Land

Part 1 - Basin-wide Estimates of Error

The following assumptions were made in order to specify the formulas for regression variance:

1) the distribution of X (i.e. the frequency distribution of values for X taken over the whole population of sample units in the given basin) was not constrained to any assumed form,

2) ground sample size was assumed to be greater than or equal to four in any stratum, and

3) the contribution to sample variance due to estimation of the slope of the regression line (stratified or unstratified) was assumed to be conditional on the actual sample chosen.

Given these assumptions, an estimate of regression sample variance was specified using a sample-based formula similar to the expression presented by Cochran (1977: eqn 7.46):

$$var(\hat{\bar{Y}}_h) = s^2_{yh}\ (1 - r^2_h) \left(\frac{N_h - n_h}{N_h n_h}\right) K_h \tag{17}$$

where

$var(\hat{\bar{Y}}_h)$ = estimated variance of the estimate of irrigated proportion in stratum h in the given basin,

$$s^2_{yh} = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$ = estimate of variance among ground sample units in stratum h,

$$r_h = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)}{\left[\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2\right]^{1/2} \left[\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2\right]^{1/2}}$$

= estimated correlation between X and Y,

n_h = number of sample units drawn for measurement of ground proportion irrigated in stratum h,

N_h = total number of sample units in stratum h,

$$K_h = \left[\frac{n_h - 1}{n_h - 2}\right] \left[1 + \left(\frac{N_h n_h}{N_h - n_h}\right) \frac{(\bar{X}_h - \bar{x}_h)^2}{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}\right], \quad (18)$$

K_h = term to (a) convert the degrees from n_h-1 to n_h-2 and to (b) account for the variance introduced by estimation of the regression coefficient,

y_{hi} = measurement of ground irrigated proportion for sample unit i in stratum h from Equation 2,

$\bar{y}_h$ = average ground irrigated proportion for sample units having ground measurements in stratum h from Equation 4,

x_{hi} = Landsat-derived, measurement of irrigated proportion for sample unit i in stratum h from Equation 1,

$\bar{x}_h$ = average Landsat irrigated proportion for sample units having matching ground data in stratum h from Equation 3, and

$\bar{X}_h$ = average Landsat irrigated proportion for all sample units in stratum h.

Equation 17 can be interpreted as follows. The first term on the right side of the equation, s_{yh}^2, represents the simple variance of ground observations (after sample unit size weighting) of irrigated proportion in stratum h. This variance is then reduced by multiplying by $(1 - r_h^2)$, the proportion of variation not accounted for by Landsat data. Note that the square of the correlation, r_h, represents the amount of variation in ground observations accounted for by Landsat data. Thus if the correlation between X and Y was one, all variation in ground observations would be accounted for by Landsat (i.e. $(1 - r_h^2)$ would be zero) and var$(\hat{\bar{Y}}_h)$ would be reduced to zero. This circumstance represents one of the major justifications for the use of Landsat data. If the correlation is reasonably high, and if the per unit area cost of Landsat observations is significantly less than that of corresponding ground observations, then use of Landsat measurement data will reduce var($\hat{\bar{Y}}_h$) below what it would have been had the same amount of money been invested in a ground-only sample.*

The next term in Equation 17 is composed of two parts. One, $(N_h - n_h)/N_h$, is known as a finite population correction. It reduces the expression for var($\hat{\bar{Y}}_h$) by an amount shown in the ratio to correct for the fact that the sample of ground data was drawn without replacement. This reduction from without replacement sampling arose from the probability with which any given unit is included in the sample. It can be explained, intuitively, by the fact that as more units are included in the sample, the information added by drawing another unit diminishes. In the extreme cases, for the first unit drawn, we learn a lot about the population but if $N_n - 1$ units are already in the sample we learn very little more by measuring the N^{th} unit. The second part, $1/n_h$, transforms Equation 17 from an expression for the variance of Y_h (individual observations of irrigated proportion) to an expression for sample variance of $\hat{\bar{Y}}_h$. It is the average stratum proportion that is to be estimated, so the variance to be estimated should be that of $\hat{\bar{Y}}_h$.

K_h is also composed of two terms. The left-most term changes the degrees of freedom (number of statistically independent observations) for computation of within stratum variance from n_h-1 to n_h-2. This modification was suggested by Tikkiwall (1960) to account for the fact that the regression coefficient is estimated from the sample and is not known beforehand. Thus, within a stratum, one degree of freedom is lost to compute the sample variance of Y_h (s^2_{yh}) in Equation 17 and one degree of freedom is lost in computing the sum of squares term ($\sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)^2$) involved in the variance of b_h.

The second term in K_h is composed of three parts. The first, unity (i.e. 1), is simply used to multiply all the preceding terms to the left of K_h in Equation 17. This results in an estimate for the variance of $\hat{\bar{Y}}_h$, less the variance for

---

* Assuming fixed or overhead costs do not differ significantly between Landsat and ground measurements.

the regression coefficient b_h. To the right of the 1 in Equation 18 are two expressions that when multiplied by the terms preceding K_h give the estimate of the variance associated with b_h.

To summarize, the estimated sample variance for $\bar{Y}_h$ shown in Equation 17 was composed of five parts: 1) the variance of the ground observations, 2) a factor reducing the variance of the ground observations to only that proportion not accounted for by Landsat measurements, 3) a finite population correction representing a further reduction in sample variance arising from the sample selection method, 4) the term $1/n_h$ giving variance in terms of the mean and not individual observations, and 5) a term representing the sample variance of b_h.

An estimate of basin-wide sample variance was produced by forming a weighted average of individual stratum sample variances. The usual formula (e.g. Cochran 1977) for estimation of variance from a stratified sample was used. Thus

$$\mathrm{var}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}) = \sum_{h=1}^{L} (W_h)^2 \, \mathrm{var}(\hat{\bar{Y}}_h) \quad , \tag{19}$$

where

$\mathrm{var}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}})$ = estimate of the variance of $\hat{\bar{Y}}$ within the sample frame within the given basin,

$\mathrm{var}(\hat{\bar{Y}}_h)$ = estimate of the variance of $\hat{\bar{Y}}_h$ within stratum h in the given basin from Equation 17,

W_h = area within the sample frame in stratum h relative to the total area in the sample frame in the given basin (this weight was defined in Equation 10), and

L = the total number of strata in the given basin.

Use of Equation 19 assumes that samples were selected independently in each stratum. Justification for squaring the W_h, while developed formally in the literature, can be seen by intuitive argument. Recall from Equation 11 that an estimate of the within-frame $\bar{Y}$ was obtained by forming a linear combination of the $\hat{\bar{Y}}_h$, namely $\hat{\bar{Y}} = \sum_h^L W_h \hat{\bar{Y}}$. Further note that regression sample variance involves a sum of squares in Y, in this case the square of differences between an 'observed' $\bar{Y}$ (after dividing a sample unit observation of irrigated proportion by $1/n_n$) and its predicted value $\hat{\bar{Y}}_h$. Then if basin-wide sample variance is defined in terms of a linear combination in the squares of the $\hat{\bar{Y}}_h$, and if the weights attached to the individual (unsquared) $\hat{\bar{Y}}_h$ are W_h, it follows that the W_h should be squared in the variance formula.

Basin-wide sample variance for acreage irrigated was obtained by multiplying Equation 19 by the square of the total area within the sample frame in the given basin:

$$\text{var}(\hat{I}_{\substack{\text{within-frame,}\\ \text{stratified}}}) = A^2 \text{ var}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}})$$

$$= A^2 \sum_{h=1}^{L} (W_h)^2 \text{ var}(\hat{\bar{Y}}_h) \quad , \tag{20}$$

where A represents the acreage within the sample frame and the other terms are as defined previously. Intuitive justification for this formula follows from the linear combination argument applied to Equation 10.

The variance estimates defined in Equations 19 and 20 are in units of percent squared and acreage squared, respectively. In order to report error in the original units, the square root of the estimated variances must be computed. The resulting value is known as the <u>standard error</u> of the estimate. Thus

$$\text{S.E.}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}) = (\text{var}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}))^{\frac{1}{2}} \times 100 \tag{21}$$

and

$$\text{S.E.}(\hat{I}_{\substack{\text{within-frame,}\\ \text{stratified}}}) = (\text{var}(\hat{I}_{\substack{\text{within-frame,}\\ \text{stratified}}}))^{\frac{1}{2}} \quad , \tag{22}$$

where S.E. stands for standard error and the other terms are as defined previously.

The standard error shown in Equation 21 represents error in the estimate of basin-wide irrigated proportion as a percent of the <u>total area</u> within the sample frame in the given basin. For example, if the estimate of basin irrigated proportion was 60 percent (of the sample frame) and if the standard error computed by Equation 21 was two percent, then the true value of irrigated proportion was expected to fall between 58 and 62 percent of the sample frame area with a certain frequency. Standard error was also computed as a percent of the estimate of irrigated proportion. This second type of percent error, termed relative standard error, was defined by

$$\text{Relative S.E.}(\hat{\hat{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}) = \frac{(\text{var}(\hat{\hat{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}))^{\frac{1}{2}}}{\hat{\hat{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}} \times 100 \quad . \tag{23}$$

Thus a relative standard error of two percent would, given the previous example, mean that the true value of irrigated proportion should fall between 58.8 percent (i.e. 60% minus 2% of 60%) and 61.2 percent with a given frequency. To distinguish the two percent standard errors apart, the statistic calculated in Equation 21 was labeled the absolute standard error. Whether absolute or relative standard error is reported depends upon which is most meaningful to DWR. If error expressed as a percent of the total area in the sample frame is most useful for planning purposes, then the former should be used. On the other hand, if error as a percent of the estimate is desired, then the latter is appropriate. There will be little difference between the two standard errors in basins with high percent irrigated, though the relative error will be somewhat more conservative. However, in basins where the percentage irrigated is low, the reported relative error may be much larger than the reported absolute error expressed in percent.

Another consideration relevant to the discussion of standard error is sample size. Recall that the design goal in the 1979 inventory was to obtain estimates of basin-specific irrigated area that fell within a certain percentage of the true (ground) value with a specified frequency. This percentage range on either side of the estimate was set at five percent. Percent in this case was defined as a function of relative standard error. This was done for two reasons: (1) it was a conventional approach and (2) it was, as has been mentioned, conservative -- a desirable characteristic when results from previous surveys are not available. So, aside from the question of utility of the resulting error statements, it was appropriate to report 1979 results as a function of relative standard error in order to judge the adequacy of the original sample allocation. Sample size calculations in subsequent inventories could, however, be based on measures of either relative or absolute error.

A further point to consider here is the interpretation of Equation 22. As stated, this equation specifies an acreage error on either side of the basin-wide estimate in which the true value should fall with given frequency. Given a within-sample frame basin estimate of 300,000 acres irrigated, for example, a standard error of 5,000 acres would be interpreted to mean that the true value of irrigated acreage within the frame should fall between 295,000 acres and 305,000 acres with a certain frequency. Thus Equation 22 produces a form of absolute error.

When specifying an inventory error goal two parameters are usually stated. The first is a relative or absolute error range, known as the confidence interval, around the estimate of the quantity of interest within which the true value should fall. The second parameter represents the desired frequency (in "repeated sampling") with which the true value will, in fact, fall in that error interval. Furthermore, the two are related. The width of the confidence interval depends on the frequency of inclusion (confidence level) specified - higher frequencies generally requiring wider intervals. Confidence intervals in the 1979 statewide irrigated lands inventory were constructed according to classical (e.g. see Cochran 1977 or Jessen 1978) sample-based procedures. A major assumption used to construct such error intervals is that the population of all possible estimates resulting from all possible samples of given size will follow some known distribution of values, typically the normal distribution. This distribution can be pictured as the form of a bell-shaped histogram (see Figure 2) centered on the true value of irrigated proportion. Estimates based on very large samples should depart from the true value with frequencies specified by the normal distribution. Estimates of irrigated proportion based on smaller samples will tend to follow a 'modified' normal, or Student's t, distribution of error about the true value.

Student's t distribution, an example of which is also shown in Figure 2, is flatter at the center and has heavier tails than the normal distribution. This movement of 'mass' or probability of occurrence to the tails of the distribution reflects the fact that variance estimates based on smaller samples will tend to be more influenced by sample observations far from the mean. There is a different Student's t curve for each sample size (or number of independent observations denoted by degrees of freedom), the shape of the curve approaching the normal distribution as sample size becomes large.

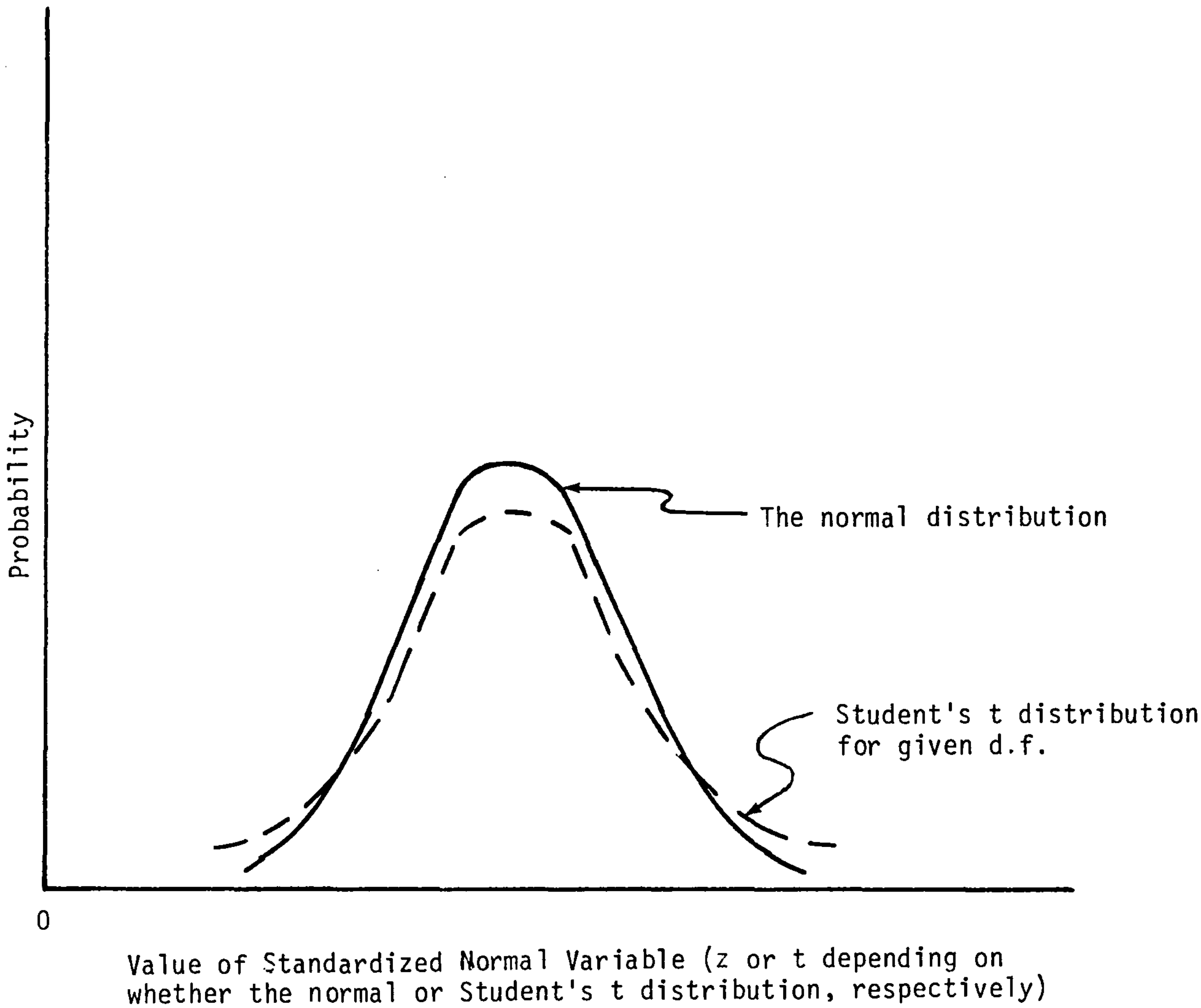
Use of the normal assumption and the related Student's t distribution has given satisfactory results in many inventories. The method used to construct estimate error intervals under this assumption is to multiply an estimate of standard error (i.e. the error due to sampling) by the Student's t value appropriate for the given degrees of freedom and the frequency of inclusion desired. The resulting error value is defined as the confidence interval half-width:

C.I. Half-Width = Standard Error x Student's t (d.f., 1 - alpha) , (24)

where d.f. stands for degrees of freedom and alpha represents the probability that the true value of irrigated proportion does not fall within plus or minus the confidence interval half-width around the estimate. The probability (or frequency) of inclusion is thus 1 - alpha.

In order for this confidence interval to be valid, one further assumption is necessary. This second assumption is that the estimator, in this case the regression estimator, is unbiased. If it is not, then the true value of irrigated proportion or area will not fall within the confidence interval with the stated frequency. The amount of bias would have to be estimated and the confidence interval endpoints adjusted up or down accordingly to obtain an error interval having the desired true value inclusion probability. As was stated earlier, evidence to date indicates that the regression estimator

FIGURE 2. The Normal and Student's t Distributions



should be effectively unbiased for the inventory problem considered here. Therefore no adjustments were made to confidence intervals or estimated values reported for the 1979 inventory.

Three different confidence interval half-widths were calculated and reported for each basin. The first was the absolute confidence interval half-width in percent for a frequency of inclusion of ninety-five percent:

$$\text{C.I.H.W.}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\\text{stratified}}}) = \text{S.E.}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\\text{stratified}}}) \times t_{(\text{d.f.},.95)} \quad . \tag{25}$$

This error interval has an interpretation analagous to that of Equation 21. Recalling the example given with that equation, an absolute standard error of two percent was subtracted and added to an irrigated estimate of 60 percent. This lead to the statement (actually a confidence interval statement) that the true value of irrigated proportion should fall between 58 and 62 percent of the area within the sample frame with a given frequency. If different estimates resulting from repeated sampling followed a normal distribution about the true value, and if the regression estimator was unbiased, then the frequency of inclusion in that example would have been 68 percent (68 times out of 100).* In this case the Student's t value, shown on the right side of Equation 25, would have been equal to one. To compute a confidence interval that should include the true value 95 percent of the time (i.e. with a probability of .95), a Student's t value must be found such that 95 percent of the area under the appropriate Student's t curve would be included between +t and -t. The appropriate Student's t value can be found by simply consulting previously tabulated values of t for the given degrees of freedom and area under the curve. Assuming the degrees of freedom were equal to 50, reference to such a table would show a Student's t value of 2.009 required to produce a 95 percent confidence interval. Thus the confidence interval half-width in units of absolute percent would be 2.009 x 2 percent = 4.018 percent of the area within the sample frame. The confidence interval statement would be that if repeated samples of the same population were taken , the true value of irrigated proportion would be expected to fall between 55.8 percent (60 minus 4.02) and 64.2 percent (60 plus 4.02) of the area within the sample frame 95 times out of 100 under the given assumptions. Alternatively, and more formally, we say that 95 percent of the confidence intervals resulting from repeated sampling would cover the true value of irrigated proportion.

A second confidence interval half-width was computed using the standard error for acreage calculated in Equation 22. In this case

$$\text{C.I.H.W.}(\hat{I}_{\substack{\text{within-frame,}\\\text{stratified}}}) = \text{S.E.}(\hat{I}_{\substack{\text{within-frame,}\\\text{stratified}}}) \times t_{(\text{d.f.},.95)} \quad , \tag{26}$$

where the probability for true value inclusion in the confidence interval has been set at .95. Using the example cited for Equation 22, the standard error of 5,000 acres would be multiplied by a Student's t value of 2.009 (assuming for sake of example that d.f.=50). This would yield a confidence interval

---

* known as the one standard error level of confidence

half-width of 10,045 acres. Thus the resulting confidence interval statement would be that the true value of irrigated area should fall between 289,955 acres (300,000 minus 10,045) and 310,045 acres (300,000 plus 10,045) in the sample frame 95 times out of 100 under the given assumptions.

The third confidence interval half-width reported was based on relative standard error expressed in percent:

$$\text{C.I.H.W.}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}}) = \text{R.S.E.}(\hat{\bar{Y}}_{\substack{\text{within-frame}\\ \text{stratified}}}) \times t_{(d.f.,.95)} \quad , \tag{27}$$

where R.S.E. stands for relative standard error. Returning to the example given with Equation 23, a two percent R.S.E. would translate to a 2.009 x 2 percent - 4.018 percent confidence interval half-width.* This percent is <u>relative</u> to the estimated basin irrigated proportion of 60 percent. That is, the confidence interval half-width is actually 4.02 percent of 60 percent, or 2.41 percent of the sample frame.

While not developed here, unstratified estimators of variance, standard error, and confidence interval half-width were constructed in a manner similar to that used in the stratified case. The unstratified estimators of error differed from their stratified counterparts only in that the stratum subscripts (h) were dropped, weights by stratum were not necessary, and summation over strata was eliminated. Sample unit data was treated as if it had been obtained by random selection from one master stratum covering the entire area within the basin sample frame.

Degrees of freedom used for determining Student's t were calculated as follows. For the unstratified basin estimates of confidence interval half-width, degrees of freedom were set equal to the total ground sample size minus two. The rational for substraction of two was explained in the discussion of stratum-specific variance associated with Equation 17. One degree of freedom was lost in calculating the ground sample variance and one was lost in estimating the variance of the regression coefficient.

The same procedure for specifying degrees of freedom could not be used for the stratified case. This was because the error distribution for $\text{var}(\hat{\bar{Y}}_{\substack{\text{within-frame,}\\ \text{stratified}}})$ is generally too complex to allow use of the sample size minus two rule for determining degrees of freedom associated with Student's t. Instead an approximate method due to Satterthwaite (1946) was used to determine the effective degrees of freedom. The formula employed was

$$m_{\text{basin}} = \frac{\left(\sum_{h=1}^{L} W_h^2 \, \text{var}(\hat{\bar{Y}}_h)\right)^2}{\sum_{h=1}^{L} \left(\frac{W_h^4 (\text{var}(\hat{\bar{Y}}_h))^2}{m_h}\right)} \quad , \tag{28}$$

---

* assuming d.f.=50 for example

where

m_{basin} = number of effective degrees of freedom used to determine Student's t for the stratified estimate of confidence interval half-width in basin b,

m_h = number of degrees of freedom for stratum h in basin b (= n_h - 2 for the regression estimator, and n_h - 1 for other estimators described in the evaluation section),

$var(\hat{\bar{Y}}_h)$ = estimated variance of the estimate of irrigated proportion in stratum h as defined in Equation 17,

W_h = weight for stratum h defined as the area in stratum h relative to the total area in the sample frame in the given basin, and

L = the total number of sample frame strata in the given basin.

The numerator of Equation 28 represents the square of the within-sample frame estimate of basin variance. The denominator consists of the sum of the squared contributions to basin variance from individual strata, each divided by its respective degrees of freedom, m_h. In rough terms this means the denominator represents the sum of the squared contributions to variance by independent sample unit observations over all strata. Dividing the sum of squared variance contributions per independent sample unit into the square of the basin-wide variance (the numerator of Equation 28) then gives the degrees of freedom for the stratified, basin-wide estimate of sample variance.

The value of m_{basin} should always lie between the smallest of the terms (n_h - 2) and their sum. Furthermore, the accuracy of the approximation given by equation 28 depends upon the assumption that sample unit observations are normally distributed within strata. If the observations are distributed with tails that are heavier, and the center more sharply peaked than corresponding features of the normal distribution, then Equation 28 will tend to over estimate the effective degrees of freedom (Cochran, 1977).

Part 2 - Statewide Estimates of Error

Statewide estimates of error were constructed in a manner parallel to that used for stratified basin estimates. The procedure was to treat each basin as an independent sampling stratum and then form a weighted sum of the estimated variances from those 'strata'. Thus the estimate of statewide sample variance for irrigated proportion was expressed as

$$\mathrm{var}(\hat{\bar{Y}}_{\text{statewide}}) = \sum_{b=1}^{B} \left(\frac{A_b}{A_s}\right)^2 \mathrm{var}(\hat{\bar{Y}}_{b,\text{within-frame}}) \quad , \tag{29}$$

where

$\mathrm{var}(\hat{\bar{Y}}_{\text{statewide}})$ = = estimated variance of the statewide estimate of irrigated proportion,

$\mathrm{var}(\hat{\bar{Y}}_{b,\text{within-frame}})$ = estimated variance of the estimate of within-sample frame irrigated proportion for basin b,

$\frac{A_b}{A_s}$ = weight assigned to basin b; defined as the area within the sample frame in basin b (A_b) divided by the total area in the sample frame statewide (A_s), and

B = total number of basins defined for the 1979 statewide inventory (=10).

Equation 29 was used to produce estimates of variance based on either stratified or unstratified estimates of individual basin variance. In the former case, the calculated values for $\mathrm{var}(\hat{\bar{Y}}_{b,\text{within-frame, stratified}})$ were substituted into the right side of Equation 29, and in the latter case values for $\mathrm{var}(\hat{\bar{Y}}_{b,\text{within-frame, unstratified}})$.

An estimate of statewide variance for irrigated acreage was given by

$$\text{var}(\hat{I}_{\text{statewide}}) = A_s^2 \text{ var}(\hat{\bar{Y}}_{\text{statewide}})$$

$$= A_s^2 \sum_{b=1}^{B} \left(\frac{A_b}{A_s}\right)^2 \text{ var}(\hat{\bar{Y}}_{b,\text{within-frame}})$$

$$= \sum_{b=1}^{B} A_b^2 \text{ var}(\hat{\bar{Y}}_{b,\text{within-frame}}) \quad , \qquad (30)$$

where all terms have been defined previously. Equation 30 was used to produce statewide estimates of variance using either stratified or unstratified estimates of basin variance.

Confidence interval half-widths were defined similarly to those described for stratified basin estimates. Three types of statewide standard error were first computed from Equations 29 and 30 using the methodology described in Equations 21, 22, and 23. These represented, respectively, 1) absolute standard error expressed in units of percent of the sample frame, 2) standard error in acres, and 3) relative standard error in units of percent of estimated irrigated proportion. The corresponding confidence interval half-widths were then computed using Equation 24.

When determining Student's t, the confidence probability (1 - alpha) for inclusion of the true value of statewide irrigated area within the confidence interval was set at .99 . That is, it was desired that the true value fall within plus or minus the confidence interval half-width (centered on the estimated value) 99 times out of 100. Since the total statewide ground sample size was very large (607), the degrees of freedom value used to specify the appropriate Student's t distribution was set arbitrarily to 500. This, in effect, made the Student's t distribution nearly identical to the normal distribution - an approximation giving a resulting Student's t value accurate to roughly .01 at the 99 percent level of confidence for degrees of freedom ranging between 400 and infinity.

Part 3 - Summary

To summarize, the values of basin-wide and statewide error reported in Tables 1a, 1b and 2a correspond to the following equations explained above:

1) absolute error, expressed as a percent of the sample frame, is reported in the second numerical column from the left in Table 1a for stratified sampling within basins and in the second column from the left in Table 1b for unstratified estimation within basins; Equation 25 was used to produce the basin figures shown in column two for a true value inclusion probability (level of confidence) of .95; substitution of the statewide variance given in Equation 29 into the confidence interval half-width formula for inclusion probabilities of .95 and .99 gave the figures shown for the state at the

bottom of the second column, Table 1a; a similar procedure was used to give the unstratified basin and statewide results shown in the second column, Table 1b;

2) relative error, expressed as a percent of the estimated value of irrigated proportion, is shown in the third numerical column from the left in Table 1a for stratified sampling within basins and in the third column from the left in Table 1b for unstratified sampling; the procedure was analogous to that for absolute error for both basin and statewide estimates;

3) standard error in acres resulting from stratified regression estimation is reported in column A4 of Table 2a; Equation 22 was used to produce this value for basins and the square root of Equation 30 gave the standard error listed for the state at the bottom of the column;

4) the confidence interval half-width, in acres, is given for an inclusion probability of .95 in column A5 of Table 2a; values for basins were computed according to Equation 26 for basin estimates based on stratified sampling; the same equation was used to calculate the 95 percent confidence interval half-width shown for the state, where the square root of equation 30 was used as the appropriate standard error value.

Note that the errors reported were based on only the area within the sample frame. That is, strictly speaking, the basin and statewide errors listed in the tables should apply to only the estimates of irrigated land within the sample frame. No error statements could be made concerning the roughly three percent of irrigated land in the state that fell in small parcels outside the sample frame.

APPENDIX IC: Equations Used for Estimation of County Irrigated Area and Associated Error

Part 1 - Estimation of Irrigated Area

The equations required to estimate county-specific irrigated land were similar in form to those used in the stratified basin estimation problem. The estimate of irrigated proportion within the county sample frame was given by

$$\hat{\hat{Y}}_{c,\text{within-frame}} = \sum_{j=1}^{B} \left(\frac{A_{cj}}{A_c}\right)(a_j + b_j x_{cj}) \quad , \tag{31}$$

where

$\hat{\hat{Y}}_{c,\text{within-frame}}$ = estimate of the proportion of the sample frame within county c that is irrigated at least once during the calendar year,

j = hydrologic basin index,

B = total number of hydrologic basins defined in the 1979 statewide inventory (= 10),

A_{cj} = total number of acres within the sample frame of county c belonging to basin j,

A_c = $\sum_{j=1}^{B} A_{cj}$ = total number of acres within the sample frame in county c,

a_j = estimated intercept for the unstratified regression equation based on sample unit size-weighted observations in basin j,

b_j = estimated slope for the same regression equation in basin j, and

x_{cj} = proportion irrigated within the sample frame in county c, in basin j determined by digitization of the Landsat interpretation.

Equation 31 can be seen to be a stratified regression estimator, where the strata are now represented by the individual basins covering the county in question. In essence, Equation 31 says that the within-frame county estimate consists of a weighted sum of the individual basin estimates $(a_j + b_j x_{cj})$ of irrigated proportion for that county. The weight for each estimate is proportional to the area within the county inside the sample frame of the given basin. These individual estimates actually represented <u>predictions</u>, since the regression equations used to produce the county estimates were developed over data sets differing (at least in part) from those of the counties to which they were applied.

The estimate of irrigated area within the sample frame of a given county was computed according to

$$\hat{I}_{c,\text{within-frame}} = A_c \hat{\hat{Y}}_{c,\text{within-frame}} \quad , \tag{32}$$

which is simply the product of the irrigated proportion estimate determined in Equation 31 and the total area (A_c) inside the sample frame in the given county. An estimate of the total irrigated acreage in a given county was obtained by adding irrigated area outside the sample frame to $\hat{I}_{c,\text{within-frame}}$. This additional irrigated acreage was located by Landsat interpretation in areas excluded from the interior of the contiguous sample frame as well as in small parcels of agricultural land outside the sample frame.

<u>Part 2</u> - Estimation of Error

Estimated errors for county estimates were reported under the assumption that the estimates of irrigated area could be considered minimally biased.* As a consequence, errors given in Table 4 were defined to represent error

---

* The validity of this assumption will be evaluated by comparison of DWR and regression estimate results during the coming year.

Table 4. County estimates based on the weighted-unstratified model.

County	Inside Acres	Prop Irrig	Acres Irrig	S.E. (acres)	Excl Acres	Outside Acres	Ex & Out Irrig	Total Acres	Total Irrig
Alameda	35808	0.20437	7318	2163	512	482615	5015	518935	12333
Alpine	14071	0.30028	4225	1059	0	451350	688	465421	4913
Amador	10907	0.76512	8345	1055	0	368034	316	378941	8662
Butte	325022	0.71862	233567	22570	14982	720116	2835	1060120	236403
Calaveras	0	0.	0	0	0	655715	1053	655715	1053
Colusa	401541	0.77695	311977	27883	5510	327667	43	734718	312020
Contra Costa	115508	0.52964	61178	9370	0	370500	6034	486008	67212
Del Norte	14919	0.58190	8681	1155	0	627571	0	642490	8681
El Dorado	19616	0.11699	2295	1362	0	1112280	1928	1131896	4223
Fresno	1499786	0.83009	1244957	100606	50435	2057156	7489	3607376	1252446
Glenn	370991	0.67563	250653	25762	10061	461687	860	842740	251512
Humboldt	75657	0.32805	24819	5856	0	2209857	3967	2285514	28786
Imperial	595280	0.85959	511697	22252	9603	2275803	3287	2880686	514984
Inyo	31683	0.34954	11074	3338	0	6439435	2230	6471118	13305
Kern	1208298	0.78631	950097	90248	0	4005342	35980	5213641	986076
Kings	691649	0.80384	555975	53036	35962	157713	486	885324	556461
Lake	44004	0.30721	13518	3056	0	800563	1056	844567	14575
Lassen	118355	0.50783	60104	6080	4314	2878058	11565	3000727	71670
Los Angeles	138798	0.18266	25353	13901	4377	2372842	7565	2516017	32918
Madera	410092	0.65970	270538	39664	38775	913974	8756	1362842	279294
Marin	0	0.	0	0	0	377393	461	377393	461
Mariposa	0	0.	0	0	0	894479	288	894479	288
Mendicino	72478	0.32023	23210	5610	0	2155594	279	2228072	23489
Merced	725187	0.78865	571919	70140	181085	457383	15482	1363655	587401
Modoc	203746	0.74043	150860	8938	6940	2456047	11077	2666732	161937
Mono	60610	0.58197	35273	3816	0	1951702	5544	2012312	40817
Monterey	511204	0.45914	234714	47051	8007	1536931	1125	2056142	235839
Napa	48339	0.31740	15343	3069	0	453412	187	501751	15530
Nevada	30313	0.12329	3737	2105	0	588398	1235	618711	4972
Orange	37611	0.41968	15785	3513	0	470837	2748	508448	18533
Placer	193399	0.33930	65620	13430	406	756248	859	950053	66479
Plumas	62085	0.18388	11416	4311	0	1588495	2333	1650580	13749
Riverside	434233	0.52227	226787	23375	43352	4078911	22781	4556495	249568
Sacramento	356086	0.62629	223013	20144	21945	270650	1557	648681	224570
San Benito	124741	0.37823	47181	9961	0	740074	1428	864815	48609
San Bernadino	95242	0.48689	46372	5343	0	12718372	21759	12813614	68131
San Diego	121421	0.41910	50888	10523	25192	2587831	22068	2734444	72955
San Francisco	0	0.	0	0	0	30443	0	30443	0
San Joaquin	710511	0.75262	534745	68721	121672	72061	5148	904244	539893
San Luis Obispo	534008	0.10701	57144	49150	2135	1578858	9658	2115002	66803
San Mateo	4605	0.48262	2222	306	0	270995	827	275600	3049
Santa Barbara	155176	0.44387	68878	14282	579	1475465	10773	1631220	79651
Santa Clara	55357	0.49159	27213	3377	0	773207	720	828564	27933
Santa Cruz	40236	0.57639	23192	3703	2754	242212	1169	285202	24361
Shasta	112608	0.49302	55518	7819	0	2320549	8426	2433157	63944
Sierra	27840	0.68768	19145	1933	0	581838	634	609678	19780
Siskiyou	299259	0.64113	191864	22884	7006	3902853	16814	4209118	208678
Solano	278318	0.63448	176587	16788	78822	222472	2558	579612	179145
Sonoma	118844	0.21798	25906	7315	0	887805	3642	1006649	29548
Stanislaus	489547	0.82526	404004	47349	191784	277631	5160	958961	409164
Sutter	333264	0.85744	285754	23142	51784	0	0	385048	285754
Tehama	218727	0.49919	109186	15188	8052	1641421	4766	1868200	113952
Trinity	0	0.	0	0	0	2028052	802	2028052	802
Tulare	929336	0.83065	771953	71261	36903	2133106	4163	3099346	776116
Tuolumne	0	0.	0	0	0	1436630	641	1436630	641
Ventura	152542	0.64299	98083	14247	12407	991918	3602	1156868	101685
Yolo	471407	0.71555	337315	32735	19351	155340	146	646099	337462
Yuba	127194	0.71585	91052	8832	5667	273928	2240	406789	93292
State	14257459		9558250		1000375	85067840	294255.	100325656	9852506

primarily due to sampling and were based on the formula for the variance of values predicted by regression. Formally, the estimator of regression variance for county irrigated proportion was given by

$$\mathrm{var}(\hat{\hat{Y}}_{c,\text{within-frame}}) = \sum_{j=1}^{B} \left(\frac{A_{cj}}{A_c}\right)^2 s_{yj}^2 \,(1 - r_j^2) \left(\frac{n_j - 1}{n_j - 2}\right) \cdot \left(1 + \frac{1}{n_j} + \frac{(x_{cj} - \bar{x}_j)^2}{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}\right) , \quad (33)$$

where

$\mathrm{var}(\hat{\hat{Y}}_{c,\text{within-frame}})$ = estimated variance of the estimate of irrigated proportion within the sample frame within county c,

s_{yj}^2 = pooled sample variance of all ground sample units in basin j,

r_j = estimated Landsat-to-ground correlation computed over all ground sample units in basin j,

$(1 - r_j^2)$ = proportion of ground sample unit variance not accounted for by Landsat measurement of proportion irrigated in basin j,

n_j = total ground sample size in basin j,

$\frac{n_j - 1}{n_j - 2}$ = ratio used to give a mean square error about the regression line based on $n_j - 2$ instead of $n_j - 1$ degrees of freedom.

x_{cj} = proportion irrigated within the sample frame in county c, in basin j determined by digitization of the Landsat interpretation,

x_{ij} = Landsat measurement of irrigated proportion for sample unit i in basin j,

$\bar{x}_j$ = average value of Landsat irrigated proportion for sample units having matching ground data in basin j, and

last term in brackets on the right = a multiplier composed of three terms which, when multiplied by the terms to their left in Equation 33, give, respectively, a) the estimated variance for an individual observation about the regression line, b) the estimated variance of the predicted mean (i.e. point on the regression line), and c) the estimated variance of the regression coefficient

Equation 33 differs from the basic form of regression variance (Equation 17), in that an extra term was included in the county equation to account for the variation of individual predicted observations about the regression line. This extra term was necessary since, as has been mentioned, the objective was to 'predict' a county value from regression equations developed, in part, elsewhere. Consequently, county error values tend to be relatively larger than their basin counterparts.

The estimate of county sample variance, expressed in acres, was then obtained directly as

$$\mathrm{var}(\hat{I}_{c,\text{within-frame}}) = A_c^2 \ \mathrm{var}(\hat{\bar{Y}}_{c,\text{within-frame}}) \quad , \tag{34}$$

where terms on the right side were defined previously. Standard error, as reported in Table 4, was calculated by taking the square root of Equation 34. That is,

$$\mathrm{S.E.}(\hat{I}_{c,\text{within-frame}}) = (\ \mathrm{var}(\hat{I}_{c,\text{within-frame}})\ )^{\frac{1}{2}} \quad . \tag{35}$$

County error values were cited at the one standard error level (i.e. Student's t was set equal to one). Thus, in the context of Table 4, the true value of irrigated acreage was expected to fall within plus or minus $\mathrm{S.E.}(\hat{I}_{c,\text{within-frame}})$ acres of the estimated value 68 times out of 100 - assuming a series of estimates themselves would be distributed normally and centered on the true value. As in the case of basin estimates, no error statement was available for irrigated land outside of the sample frame.

Part 3 - Summary

Table 4 summarizes the 1979 estimation and measurement data for each county in the state. This information includes:

1) Figures for total acreage (irrigated and not) obtained by digitizing administrative boundaries on the Landsat interpretation base; these are reported for

   a) the total area within the sample frame in the first numerical column (counting left to right in Table 4),

   b) the total area excluded from the interior of the contiguous sample frame in the fifth column from the left,

   c) the total area outside the contiguous sample frame the sixth column from the left, and

   d) the total area within the county (sum of a, b, and c) in the eighth (second to last) column from the left;

2) an estimate of proportion irrigated within the sample frame calculated according to Equation 31; this value is given in the second column from the left;

3) irrigated acreage figures reported for

   a) the sample frame (as estimated by Equation 32) in the third column from the left,

   b) exclusion areas and/or areas outside the sample frame in the seventh column from the left, and

   c) the entire county (by sum of a and b) in the last column on the right; and

4) an estimate of standard error (by Equation 35) in acres in the fourth column from the left.

LITERATURE CITED

Cochran, W. G. 1977. Sampling Techniques, Third Edition. John Wiley and Sons, Inc., New York. 428 pp.

Jessen. R. J. 1978. Statistical Survey Techniques. John Wiley and Sons, New York. 520 pp.

Satterthwaite, F. E. 1946. An Approximate Distribution of Estimates of Variance Components. Biometrics, (2):110-114.

Tikkiwal, B. D. 1960. On the Theory of Classical Regression and Double Sampling Estimation. Journal of the Royal Statistical Society, Series B, (22):131-138.

APPENDIX II

PRESENTATION OF ESTIMATORS OF IRRIGATED AREA AND TABLES SUMMARIZING RESULTS

APPENDIX II

PRESENTATION OF ESTIMATORS OF IRRIGATED AREA AND TABLES SUMMARIZING RESULTS

The primary irrigated acreage estimators considered were those for the stratified regression design. This is the design for which the samples were allocated and for which the least restrictive assumptions are made on the relationship between the "independent" and the "dependent" variables. However, in order to try to determine if some other estimator might be more efficient for the estimation of irrigated proportions, ratio (biased and unbiased) and difference estimators were also considered. While any statement on the relative efficiencies of these estimators is strictly sample specific, large or consistent differences between estimators should be apparent.

Let u_{hi} be the i-th observation, h-th stratum of the Landsat proportion irrigated and v_{hi} be the i-th observation, h-th stratum of the ground proportion irrigated. Then define the variables of interest, x and y, to be the associated weighted proportions

$$x_{hi} = w_{hi}u_{hi} = \left(\frac{A_{hi}}{\bar{A}_h}\right) u_{hi} = \frac{N_h(A_{hi}u_{hi})}{\sum_{i=1}^{N_h} A_{hi}} \qquad [1]$$

and

$$y_{hi} = w_{hi}v_{hi} = \left(\frac{A_{hi}}{\bar{A}_h}\right) v_{hi} = \frac{N_h(A_{hi}v_{hi})}{\sum_{i=1}^{N_h} A_{hi}}, \qquad [2]$$

where

A_{hi} = acres in the i-th sample unit, h-th stratum as measured on Landsat

$\bar{A}_h$ = population mean sample unit size.

Further, for each stratum h, define

$\bar{x}_h$ = mean over matched units of the weighted Landsat proportion irrigated

$\bar{y}_h$ = mean over matched units of the weighted ground proportion irrigated

$\bar{X}_h$ = mean over all units of the weighted Landsat proportions, and

s^2_{yh} = sample estimate of the variance of the y's.

REGRESSION ESTIMATORS

For the stratified regression estimate of the mean proportion irrigated the classical estimator

$$\hat{\bar{Y}}_h = \bar{y}_h + b_h(\bar{X}_h - \bar{x}_h) \qquad [3]$$

where

$$b_h = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2} \qquad [4]$$

was used. The estimator of the variance of this estimator is, in the literature, not well agreed upon. Various equations have been proposed, however the majority of them are based on rather restrictive assumptions which, typically, are not met in a sampling framework (e.g. infinite populations, large samples and normality). The equations considered were all of the form

$$\mathrm{var}(\hat{\bar{Y}}_h) = s^2_{yh}(1 - r^2_h) * \left(\frac{N_h - n_h}{N_h n_h}\right) * K_h \qquad [5]$$

where five different expressions for the "factor" K_h were evaluated.

Factor 1:

$$K_h = \left(\frac{n_h - 1}{n_h - 2}\right) \qquad [6]$$

With factor 1 the variance equation reduces to that proposed by Cochran (1977). The derivation of this formula assumes infinite populations and large samples. It further assumes that the slope coefficient is a known constant (Sukhatme, 1954). The variance equation is that which would be used if the data were being treated as a general linear model multiplied by the finite population correction. The factor itself is used merely to set the degrees of freedom in the denominator of the error term to n-2 rather than n-1.

Factor 2:

$$K_h = \left(\frac{n_h - 2}{n_h - 3}\right) \qquad [7]$$

Factor 2 was the factor used in the allocation of samples. The resulting variance equation is based on the derivation by Tikkiwal (1960) and assumes infinite-bivariate-normally distributed populations; it is however justified by a super-population argument which says, essentially, that if the population is considered to be a sample from some infinite "super-population" then this is the best estimate. The above expression for the factor results from the reduction of (1 + 1/n-3) where the second part of this expression represents the increase in variance due to estimating the slope coefficient from a sample; the derivation of this term is based on the assumption that there is an infinite-normally distributed population of x's. The degrees of freedom for the error term is assumed to be n-1.

Factor 3:

$$K_h = \left(\frac{n_h - 1}{n_h - 3}\right) \qquad [8]$$

Factor 3 is the same as factor 2 except that the degrees of freedom of the error term is now assumed to be n-2. The resulting variance equation is, in fact, the one given by Tikkiwal (1960) and also by O'Regan and Boyd (1974). O'Regan and Boyd also assume infinite populations, large samples and normality of the x's.

Factor 4:

$$K_h = \left(1 + \frac{n_h N_h}{N_h - n_h} \frac{(\bar{X}_h - \bar{x}_h)^2}{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}\right) \qquad [9]$$

This factor is that derived by Tikkiwal (1960) for a given sample of x's and agrees with the expression given by Cochran (1977, eqn 7.46). It, consequently, does not make assumptions on the distribution of the x's but it is only conditional. The degrees of freedom for error is taken to be n-1.

Factor 5:

$$K_h = \left(\frac{n_h - 1}{n_h - 2}\right) * \left(1 + \frac{n_h N_h}{N_h - n_h} \frac{(\bar{X}_h - \bar{x}_h)^2}{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}\right) \qquad [10]$$

This factor is the same as number four except that the degrees of freedom for the error term is taken to be n-2; it is, actually, the one suggested by Tikkiwal (1960) for the conditional case.

Of the above factors, factor 1 is rather unrealistic and will tend to under estimate the true variance of the estimator (Cochran, 1977). It was never given serious consideration. The third factor was prefered over the second with regard to the assumptions on the degrees of freedom and, similarly, the fifth was prefered over the fourth. Thus, the choice of variance equations reduced to whether it is more realistic to assume normally distributed-infinite populations of the x's or to make the variance conditional on the samples chosen.

The conditional approach (factors 4 and 5) is sample specific and is to be prefered on theoretical grounds 1/. It was therefore chosen as the primary method of analysis. The use of factor 3 should be restricted to sample size computation and as a check on the estimates produced using factor 5.

1/ Visual inspection of the distributions of both x and y indicate that on a per stratum basin the assumption of normality is violated.

The estimators used for the stratified regression estimates were formed via the normal combination of per stratum estimates. Specifically, if, for a given basin, A_h is the total acreage in the h-th stratum and A is the total acreage in the basin

$$\hat{\bar{Y}}_{st} = \sum_{h=1}^{L} \left(\frac{A_h}{A}\right) * \hat{\bar{Y}}_h \qquad [11]$$

where in acres this becomes

$$A\hat{\bar{Y}}_{st} = A \sum_{h=1}^{L} \left(\frac{A_h}{A}\right) \hat{\bar{Y}}_h = \sum_{h=1}^{L} A_h \hat{\bar{Y}}_h$$

and

$$\operatorname{var}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{L} \left(\frac{A_h}{A}\right)^2 * \operatorname{var}(\hat{\bar{Y}}_h) \qquad [12]$$

or on an acreage basis

$$A^2 \operatorname{var}(\hat{\bar{Y}}_{st}) = A^2 \sum_{h=1}^{L} \left(\frac{A_h}{A}\right)^2 \operatorname{var}(\hat{\bar{Y}}_h) = \sum_{h=1}^{L} A_h^2 \operatorname{var}(\hat{\bar{Y}}_h) .$$

This stratified variance will differ from the variance obtained using the data in an unstratified estimation procedure in three ways. First, the degrees of freedom (see below) will be greater for the unstratified model. This will result in a lower Student's t-statistic even for identical variances, and, consequently, a lower relative error. Secondly, as Cochran (1977) points out, when the allocation of sample units to strata is not approximately proportional the estimate

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1} \qquad [13]$$

where n = total sample size

and $\bar{y} = (\Sigma\, y_i) / n$

formed by assuming the sample to have been drawn randomly from the total population may be a poor estimate of the true variance. Consequently, the use of his equation 5.A.44 (due to Rao, 1962) where corresponding sample estimates are used in place of population values should provide a better estimate of the unstratified variance. Specifically, letting

$$\left[s_y^2\right]' = \left[\frac{1}{N}\sum_h^L \frac{N_h}{n_h}\sum_i^{n_h} y_{hi}^2\right] - \hat{\bar{Y}}_{st}^2 + \mathrm{var}(\hat{\bar{Y}}_{st}) \qquad [13']$$

we can estimate the associated regression variance by

$$\mathrm{var}(\bar{Y}) = \left[s^2\right]' (1 - r^2)\left(\frac{N - n}{N\, n}\right) K \qquad [14]$$

where s^2 is from 5.A.44, N is the total population size, and

$$r^2 = \frac{\left[\sum^n (y_i - \bar{y})(x_i - \bar{x})\right]^2}{\Sigma(y_i - \bar{y})^2\,\Sigma(x_i - \bar{x})^2} = \frac{s_{xy}^2}{s_x^2 s_y^2} \qquad [15]$$

represents the square of the correlation between x and y.

It was recently pointed out to us that 5.A.44 can be modified (Sigmund, 1981) to provide an estimate of the unstratified covariance between x and y

$$(s_{xy}^2)' = \left[\frac{1}{N}\sum_h^L \frac{N_h}{n_h}\sum_i^{n_h} x_{hi}y_{hi}\right] - \left[\sum_i^{n_h} w_h\bar{x}_h\right]\left[\sum_i^{n_h} w_h\bar{y}_h\right] + \sum_i^{n_h}\frac{w_h^2 s_{hxy}}{n_h}. \qquad [13'']$$

Using this expression and [13'] (where $(s_x^2)'$ is obtained from an expression in x completely analogous to [13'] in y) we can obtain a revised estimate of r^2

$$(r^2)' = \frac{(s'_{xy})^2}{(s_x^2)'(s_y^2)'} \qquad [15']$$

which can be used in [14] to give

$$(\mathrm{var}(\hat{\bar{Y}}))' = \left(\frac{N-n}{N\ n}\right)(s_y^2)'(1-(r^2)')K \ . \qquad [14']$$

And finally, the unstratified estimates will differ from the stratified estimates in that the "observations" themselves differ between the stratified and the "inherently unstratified" models. This difference is due to the method of weighting. Observations in the unstratified model are weighted according to the size of the sample unit relative to the average size in the whole population, viz

$$x^*_{hi} = w^*_{hi}u_{hi} = \left(\frac{A_{hi}}{\bar{A}}\right) u_{hi} = \frac{N(A_{hi}u_{hi})}{\sum\limits_{h=1}^{L}\sum\limits_{i=1}^{N_h} A_{hi}} \qquad [16.1]$$

and similarly,

$$y^*_{hi} = w^*_{hi}v_{hi} = \left(\frac{A_{hi}}{\bar{A}}\right) v_{hi} = \frac{N(A_{hi}v_{hi})}{\sum\limits_{h=1}^{L}\sum\limits_{i=1}^{N_h} A_{hi}} \qquad [16.2]$$

where $\bar{A}$ is the mean sample unit size (unstratified) over all elements in the population, while observations in the stratified model are weighted relative to $\bar{A}_h$. Thus, being based on a more stable mean, we would expect these unstratified weighted observations to be somewhat less variable.[2/]

RATIO ESTIMATES

For the problem at hand, our model specifies that the relationship between the x's and the y's is linear. Further, because of the nature of the variables being used we know that the regression line may have an intercept near the origin and a slope close to unity. If, rather than using sample estimates of the slope and intercept, we "accept the hypothesis" that the intercept is, in fact, the origin, then we can use a ratio model and gain more degrees of freedom. Two ratio estimators were proposed for use, an unbiased mean of ratios estimator and a biased ratio of means estimator.

BIASED RATIO ESTIMATOR

The biased ratio estimator is

$$\hat{\bar{Y}}_h = \bar{X}_h * r_h = \bar{X}_h * \frac{\bar{y}_h}{\bar{x}_h} \qquad [17]$$

with the population estimate of the variance given approximately by

$$V_h^* = \left(\frac{N_h - n_h}{N_h n_h}\right) (S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{xyh}) \qquad [18]$$

where R_h is the population ratio. This is the commonly recognized expression for the variance; however Cochran (1977) notes that it tends to underestimate the true variance. A modification of this, due to Sukhatme (1954), will produce an expression of order $1/n^2$ (as opposed to $O(1/n)$). Specifically,

2/ Initial tests indicate that this is, indeed, the case. It may be desireable to use this method of weighting for future estimation.

$$V(\hat{\bar{Y}}_h) = V_h^* * \left(1 + \frac{3C_{xxh}}{n_h} + \frac{6C_{xxh}}{n_h} * \frac{\rho_h^2 C_{yyh} + C_{xxh} - 2C_{xyh}}{C_{yyh} + C_{xxh} - 2C_{xyh}}\right) \qquad [19]$$

where $C_{xxh} = S_{xh}^2 \,/\, \bar{X}_h^2$

$C_{yyh} = S_{yh}^2 \,/\, \bar{Y}_h^2$

and $C_{xyh} = S_{xyh} \,/\, \bar{X}_h \bar{Y}_h$

is a better estimator for the population variance. This derivation is, however, based on the assumption of normality and large populations. Substituting the respective sample estimates for each of the terms in this expression gives the estimated variance.

UNBIASED RATIO ESTIMATES

Since the bias of a ratio estimate in a stratified design may be large, a number of unbiased ratio-type estimators have been proposed. The one used here was suggested by Goodman and Hartley(1958) and is given by

$$\hat{\bar{Y}}_h = \bar{r}_h \bar{X}_h + \frac{(N_h - 1)n_h}{(n_h - 1)N_h}\left(\bar{y}_h - \bar{r}_h \bar{x}_h\right) \qquad [20]$$

where $\bar{r}_h$ is the sample mean ratio

$$\bar{r}_h = \frac{1}{n_h} \sum_{h=1}^{n_h} \frac{y_{hi}}{x_{hi}} .$$

The population variance of this estimator is given by

$$\mathrm{Var}(\hat{\bar{Y}}_h) = \left[\frac{N_h - n_h}{N_h n_h}\right] * \left[S_{yh}^2 + \bar{R}_h^2 S_{xh}^2 - 2\bar{R}_h S_{xyh} + \left[\frac{S_{rh}^2 S_{xh}^2 + S_{rxh}^2}{n - 1}\right]\right] \qquad [21]$$

where $\overline{R}_h$ is the population mean ratio, S^2_{rh} is the population variance of the ratios, and so on (Goodman and Hartley, (1958)). While the authors give an expression for the estimated variance there appear to be a number of errors in their component equations. Even after correcting ones that we found, we still obtained negative variance estimates. Lacking the time to rederive the correct formula, we instead substituted the associated sample estimates of the (co)variances and the mean ratio into the population expression to obtain a sample estimate of this variance.

Throughout their development Goodman and Hartley assumed a large or infinite population ($N_h >> n_h$). Clearly, with the population and sample sizes encountered in the current design this assumption does not hold. We multiplied the above variance expression by a finite population correction to obtain an expression that would, to an order of magnitude comparable to the other variance expressions, accurately reflect the underlying population variance.

A second, and more troublesome, problem encountered in the use of this ratio-type estimator is that it is quite possible that a given x_i will be identically zero. When this happens the associated ratio can not be evaluated. This was the case for a number of sample units and can, in general, be expected to occur for the irrigated proportions problem. In an attempt to overcome this problem the following contingency table was suggested:

x_i \ y_i	$=0$	$\neq 0$
$=0$	$r_i = 1$	$r_i = \frac{1}{1 - y_i}$
$\neq 0$	$r_i = 1 - x_i$	$r_i = y_i / x_i$

It will be noted, however, that this algorithm essentially maps the most variable ratios to a value very close to the mean ratio and, consequently, will cause the variance of the ratios and ultimatley the variance of $\overline{Y}$ to be underestimated. The overall effect of this is not yet known.

DIFFERENCE ESTIMATES

If future studies indicate that the slopes within strata and within basins are constant, then the known slopes can be used in a difference model

$$\hat{\overline{Y}}_h = \overline{y}_h + b_{oh}(\overline{X}_h - \overline{x}_h) \qquad [22]$$

where b_{oh} is the assumed slope coefficient. The estimated variance of this estimator is given by Cochran (1977) as

$$\mathrm{var}(\hat{\overline{Y}}_h) = \left(\frac{N_h - n_h}{N_h n_h}\right) * (s^2_{yh} + b^2_{oh} s^2_{xh} - 2b_{oh} s_{xyh}). \qquad [23]$$

Such a model can potentially increase the degrees of freedom and reduce the variance. However, if the assumed slopes are not close to the true ones, serious bias may be introduced into the estimate of $\overline{Y}$. In the current study we had no prior information on the slopes but we expected them to be "close" to unity; consequently, we chose b_{oh} to be 1 for each stratum in each basin.

COMBINED UNBIASED RATIO - REGRESSION ESTIMATORS

Runs of the program MONTE CRISTO with 1976 Sacramento digital Landsat data indicated that for small sample sizes the unbiased ratio estimator is more efficient than the regression estimator. Specifically, Table 1 gives the sample sizes, by stratum, for which each of the estimators would be used. The combined, stratified estimates of irrigated proportion and its variance is then

$$\hat{\bar{Y}}_{st} = \sum_{h=1}^{L} \left(\frac{A_h}{A}\right) \left(i_{h1}\hat{\bar{Y}}_{h.reg} + i_{h2}\hat{\bar{Y}}_{h.rat}\right) \qquad [24]$$

$$\text{var}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{L} \left(\frac{A_h}{A}\right)^2 \left(i_{h1}\text{var}(\hat{\bar{Y}}_{h.reg}) + i_{h2}\text{var}(\hat{\bar{Y}}_{h.rat})\right) \qquad [25]$$

where

$$i_{h1} = \begin{cases} 1 & \text{If sample size for stratum h is large according to Table 1} \\ 0 & \text{Otherwise} \end{cases}$$

and

$$i_{h2} = \begin{cases} 1 & \text{If sample size for stratum h is small according to Table 1} \\ 0 & \text{Othewise} \end{cases}$$

$\hat{\bar{Y}}_{h.reg}$ = regression estimate for the h-th stratum as figured above,

$\text{var}(\hat{\bar{Y}}_{h.reg})$ = estimated regression variance as figured above,

$\hat{\bar{Y}}_{h.rat}$ = unbiased ratio estimate, and

$\text{var}(\hat{\bar{Y}}_{h.rat})$ = estimated unbiased-ratio variance.

Table 1. Sample sizes defining which estimator (ratio or regression) to use in the combined estimator.

Stratum	Sample Size (n_h) Use Ratio	Use Regression
1	$\leqslant$ 3	$\geqslant$ 4
2	$\leqslant$ 10	$\geqslant$ 11
3	$\leqslant$ 5	$\geqslant$ 6
4	$\leqslant$ 5	$\geqslant$ 6
5	$\leqslant$ 6	$\geqslant$ 7
6	$\leqslant$ 6	$\geqslant$ 7
7	$\leqslant$ 5	$\geqslant$ 6

DEGREES OF FREEDOM

For a stratified estimate the approximate degrees of freedom, m_e is given by Wensel (1980) as

$$m_e = \frac{\left(\sum_{h=1}^{L} W_h^2 S_{\bar{Y}h}^2 \right)^2}{\sum_{h=1}^{L} \frac{W_h^4 S_{\bar{Y}h}^4}{m_h}} \qquad [25]$$

where

W_h = weight for the h-th stratum, and

m_h = degrees of freedom for the estimator used in the h-th stratum.

For the regression estimator $m_h = n_h - 2$ and for each of the others $m_h = n_h - 1$. Note that the numerator in the above expression is the square of the stratified variance.

UNSTRATIFIED ESTIMATES

As stated above, the observations for the unstratified models differ from those for the stratified models in the system of weights used. If we define the unstratified sample sizes and population sizes to be the sum over strata of the stratum specific sizes (i.e., $n = \sum_h^L n_h$ and $N = \sum_h^L N_h$) and then take $L = 1$ and $w_h = 1$, the above equations for the stratified system are also applicable to the unstratified system (alternatively, if we drop all the subscripts "h", the resulting unstratified equations are more easily seen). Note that the degrees of freedom, m_e, reduces to m_h, or equivalently to $n - 1$ or $n - 2$, as expected for the unstratified case.

Tables A1 to A7 give the estimates resulting from the stratified models for the regression with factor 3, regression with factor 5, unbiased ratio, biased ratio, difference, and for the combined ratio-regression (with factor 3 and with factor 5) estimators respectively. Tables B1 to B5 give the corresponding unstratified estimates (with the exception of the combined ratio-regression estimates).

Table C1.1 gives the values obtained for $(s_x^2)'$, $(s_y^2)'$ and $(s_{xy})'$ by applying 5.A.44 (i.e., equations [13'] and [13''] to the estimates, and Table C1.2 gives the resulting $(r^2)'$ and standard errors for regression. Sizable increases in standard error relative to corresponding uncorrected values are seen in Table C1.2 for at least five basins. These increases, when translated to confidence interval half-widths, meant that stratified sampling became superior in seven instead of four basins on the basis of estimated error. Tables C2.1 and C2.2 give corresponding "corrected" values for unweighted estimates. Inspection of Table C2.2 shows that increases in standard error, when they occurred, were generally not large in the unweighted case.

COUNTY ESTIMATES

The county estimates were based on the regular prediction equations for the general linear model. Here the basins were, essentially, considered as strata for estimation purposes. Thus, assuming no land use stratification, the prediction equations were

$$\hat{\bar{Y}}_c = \sum_{j=1}^{10} \left(\frac{A_{cj}}{A_c}\right) (a_j + b_j x_{cj}) \qquad [26]$$

and

$$\text{var}(\hat{\bar{Y}}_c) = \sum_{i=1}^{10} \left(\frac{A_{cj}}{A_c}\right)^2 s^2_{yj}\ (1-r^2_j) \left(\frac{n_j-1}{n_j-2}\right) \left[1 + \frac{1}{n_j} + \frac{(x_{cj}-\bar{x}_j)^2}{\sum_{i=1}^{n_j} (x_{ij}-\bar{x}_j)^2}\right] \qquad [27]$$

where the subscripts c and j denote county and basin respectively and where s^2_{yj} (and/or r^2_j and the sums of squares for the x's in the last term) may be improved upon through the use of [13'] and [15']. A similar formulation where, for each county, both land use strata within basins and between basins are considered to be independent strata gives

$$\hat{\bar{Y}}_c = \sum_{h}^{L} \sum_{j=1}^{10} \left(\frac{A_{chj}}{A_c}\right) (a_{hj} + b_{hj} x_{chj})$$

and

$$\text{var}(\hat{\bar{Y}}_c) = \sum_{h}^{L} \sum_{i=1}^{10} \left(\frac{A_{chj}}{A_c}\right)^2 s^2_{yhj}\ (1-r^2_{hj}) \left(\frac{n_{hj}-1}{n_{hj}-2}\right) \left[1 + \frac{1}{n_{hj}} + \frac{(x_{chj}-\bar{x}_{hj})^2}{\sum_{i=1}^{n_{hj}} (x_{hij}-\bar{x}_{hj})^2}\right]$$

Here the county predictor x_{cj} (or x_{chj} in the stratified case) was defined as the total irrigated acreage relative to the total county ("inside") acreage. This definition of the predictor variable lead to some problems. Specifically, for the overall estimate of irrigated proportion (in a given stratum within a given basin) the weighted variable $x_{hi} = w_{hi}u_{hi}$ (see [1]) was used in conjunction with [2] to obtain estimates of the regression coefficients. However in the prediction situation where the corresponding definition of weights is meaningless, the "weighted regressions" can lead to questionable predictions. Therefore, in producing the county estimates, the unweighted regressions (i.e., regressing u_{hi} on v_{hi} rather than x_{hi} on y_{hi} to produce the estimated regression coefficients in [26]) are to be preferred on theoretical grounds.

COUNTY RESULTS

The county estimates based on the weighted regressions are given in Tables G1 (regressions based on the unstratified observations) and G3 (regressions based on the stratified observations). Tables G2 & G4 give the corresponding county results based on the unweighted regressions, and the Tables H1 to H3 show how these estimates (APT) differ from DWR's best (revised to 1979) estimates.

In producing the county estimates, we originally decided that the weighted-unstratified county model would be sufficient for prediction purposes. This decision was based on the following considerations:

(i) The county estimates were predictions based on the basin regressions, and as a result were subject to a potentially large inherent error of prediction. We felt that the bias produced in using the weighted model would be overshadowed by the error of prediction.

(ii) A desire to avoid possible confusion arising from the use of regression equations for the county predictions differing from those used for the basin estimates.

(iii) An error in reasoning (centering around sample sizes) which indicated that the stratified county estimates would be inappropriate.

However, in retrospect, we would recommend the use of unweighted regressions for prediction based on theoretical considerations presented at the end of the previous section. Either the stratified or the unstratified models are nominally appropriate. Since the samples are allocated according to the stratified model however, the regression coefficients for the unstratified model will be biased (this could be "corrected" by following the same sort of reasoning used to produce $(s_y^2)'$ in [13']). Consequently, the unweighted-stratified county prediction model is to be preferred (see results in Table G4).

LITERATURE CITED

Cochran, W. G. (1977),

Sampling Techniques, 3rd Edition. John Wiley and Sons, Inc., New York. 428p.

Goodman, Leo A. and H. O. Hartley (1958),

The precision of unbiased ratio-type estimators. Jour. Amer. Stat. Assoc., (53):491-508.

O'Regan, William G. and Robert W. Boyd (1974),

Regression Sampling: some results for resource managers and researchers. USDA Forest Service, Res. Note PSW-286.

Rao, J.N.K. (1962),

On the estimation of relative efficiency of sampling procedures. Ann. Inst. Math. Stat., (14):143-150.

Sigmund, R. (1981),

Personal communication.

Sukhatme, P.V. (1954),

Sampling Theory of Surveys with Applications. Iowa State College Press; Ames, Iowa.

Tikkiwal, B.D. (1960),

On the theory of classifical regression and double sampling estimation. Jour. Roy. Stat. Soc., Ser. B, (22):131-138.

Wensel, L.C. (1980),

Wildland Resource Sampling. Draft manuscript.

TABLE A1.1 Stratified summary statistics for the area within the sample unit frame. Regression with factor 3.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53521	321070	5891	11974
San Francisco	191654	0.21852	41880	1175	2459
Central Coast	1380040	0.31906	440316	12821	25903
South Coast	598866	0.45787	274203	7618	15289
Colorado Desert	818231	0.82147	672152	5736	11586
South Lahontan	235626	0.27383	64522	4328	8827
North Lahontan	175456	0.58726	103038	2284	4688
Sacramento	3388466	0.65381	2215413	30056	60315
San Joaquin	2788914	0.74778	2085494	37232	75105
Tulare	4080305	0.82038	3347400	40517	81402
State	14257457	0.67091	9565489	65166	128376

TABLE A1.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346785
San Francisco	512	2586376	5623	2778543	47503
Central Coast	13475	5789056	24725	7182571	465041
South Coast	62133	6289499	63810	6950499	338013
Colorado Desert	28421	11852213	10830	12698865	682982
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117981
Sacramento	211744	13452904	37823	17053114	2253236
San Joaquin	542467	6704753	51098	10036134	2136592
Tulare	123300	5977461	42352	10181065	3389752
State	1000375	85067824	294255	100325656	9859744

TABLE A2.1 Stratified summary statistics for the area within the sample unit frame. Regression with factor 5.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53521	321070	6029	12238
San Francisco	191654	0.21852	41880	1108	2329
Central Coast	1380040	0.31906	440316	12572	25420
South Coast	598866	0.45787	274203	8510	17223
Colorado Desert	818231	0.82147	672152	5670	11447
South Lahontan	235626	0.27383	64522	4402	8977
North Lahontan	175456	0.58726	103038	2297	4695
Sacramento	3388466	0.65381	2215413	30327	60823
San Joaquin	2788914	0.74778	2085494	35419	71145
Tulare	4080305	0.82038	3347400	40721	81769
State	14257457	0.67091	9565489	64477	127020

TABLE A2.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346785
San Francisco	512	2586376	5623	2778543	47503
Central Coast	13475	5789056	24725	7182571	465041
South Coast	62133	6289499	63810	6950499	338013
Colorado Desert	28421	11852213	10830	12698865	682982
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117981
Sacramento	211744	13452904	37823	17053114	2253236
San Joaquin	542467	6704753	51098	10036134	2136592
Tulare	123300	5977461	42352	10181065	3389752
State	1000375	85067824	294255	100325656	9859744

TABLE A3.1 Stratified summary statistics for the area within the sample unit frame. Unbiased ratio estimate.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% Width (acres) (A5)
North Coast	599896	0.53451	320651	11848	25124
San Francisco	191654	0.21163	40560	2620	5552
Central Coast	1380040	0.32208	444483	23364	47653
South Coast	598866	0.46696	279647	7707	15415
Colorado Desert	818231	0.82376	674026	5949	11963
South Lahontan	235626	0.27683	65228	4310	8777
North Lahontan	175456	0.58248	102200	2293	4685
Sacramento	3388466	0.65183	2208704	34122	68481
San Joaquin	2788914	0.75390	2102562	42336	84895
Tulare	4080305	0.82246	3355887	39497	79117
State	14257457	0.67291	9593948	72996	143802

TABLE A3.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346365
San Francisco	512	2586376	5623	2778543	46182
Central Coast	13475	5789056	24725	7182571	469209
South Coast	62133	6289499	63810	6950499	343456
Colorado Desert	28421	11852213	10830	12698865	684856
South Lahontan	4377	16668221	17338	16908224	82567
North Lahontan	0	3891697	14942	4067153	117142
Sacramento	211744	13452904	37823	17053114	2246527
San Joaquin	542467	6704753	51098	10036134	2153660
Tulare	123300	5977461	42352	10181065	3398239
State	1000375	85067824	294255	100325656	9888203

TABLE A4.1 Stratified summary statistics for the area within the sample unit frame. Biased ratio estimate.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53600	321544	6305	12826
San Francisco	191654	0.21785	41752	1305	2693
Central Coast	1380040	0.32145	443614	15498	31286
South Coast	598866	0.46408	277922	7558	15127
Colorado Desert	818231	0.82281	673249	5752	11578
South Lahontan	235626	0.27291	64305	4383	8930
North Lahontan	175456	0.58362	102400	2307	4711
Sacramento	3388466	0.65214	2209754	35579	71971
San Joaquin	2788914	0.75258	2098881	38989	78090
Tulare	4080305	0.82197	3353888	38763	77730
State	14257457	0.67244	9587309	68447	134840

TABLE A4.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	347259
San Francisco	512	2586376	5623	2778543	47375
Central Coast	13475	5789056	24725	7182571	468339
South Coast	62133	6289499	63810	6950499	341732
Colorado Desert	28421	11852213	10830	12698865	684078
South Lahontan	4377	16668221	17338	16908224	81643
North Lahontan	0	3891697	14942	4067153	117342
Sacramento	211744	13452904	37823	17053114	2247577
San Joaquin	542467	6704753	51098	10036134	2149979
Tulare	123300	5977461	42352	10181065	3396240
State	1000375	85067824	294255	100325656	9881564

TABLE A5.1 Stratified summary statistics for the area within the sample unit frame. Difference estimate.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53530	321124	6353	12946
San Francisco	191654	0.21807	41794	1152	2396
Central Coast	1380040	0.32074	442634	13028	26317
South Coast	598866	0.45102	270101	7342	14684
Colorado Desert	818231	0.82058	671424	5629	11357
South Lanontan	235626	0.27320	64373	4192	8539
North Lanontan	175456	0.58559	102745	2216	4532
Sacramento	3388466	0.65521	2220157	32800	65770
San Joaquin	2788914	0.75432	2103734	38264	76695
Tulare	4080305	0.82316	3358743	38722	77567
State	14257457	0.67311	9596831	66022	130063

TABLE A5.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346839
San Francisco	512	2586376	5623	2778543	47417
Central Coast	13475	5789056	24725	7182571	467359
South Coast	62133	6289499	63810	6950499	333910
Colorado Desert	28421	11852213	10830	12698865	682254
South Lanontan	4377	16668221	17338	16908224	81711
North Lanontan	0	3891697	14942	4067153	117688
Sacramento	211744	13452904	37823	17053114	2257980
San Joaquin	542467	6704753	51098	10036134	2154832
Tulare	123300	5977461	42352	10181065	3401095
State	1000375	85067824	294255	100325656	9891085

TABLE A6.1 Stratified summary statistics for the area within the sample unit frame. Combined ratio-regression with factor 3.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53472	320777	6215	12568
San Francisco	191654	0.21279	40782	1826	4318
Central Coast	1380040	0.31949	440909	12959	26138
South Coast	598866	0.45787	274203	7618	15289
Colorado Desert	818231	0.82144	672128	5744	11586
South Lahontan	235626	0.27383	64522	4328	8827
North Lahontan	175456	0.58726	103038	2284	4688
Sacramento	3388466	0.65059	2204502	30767	61636
San Joaquin Tulare	2788914	0.74635	2081506	36730	73711
State	14257457	0.67018	9555072	65821	129667

TABLE A6.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346491
San Francisco	512	2586376	5623	2778543	46405
Central Coast	13475	5789056	24725	7182571	465634
South Coast	62133	6289499	63810	6950499	338013
Colorado Desert	28421	11852213	10830	12698865	682957
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117981
Sacramento	211744	13452904	37823	17053114	2242325
San Joaquin	542467	6704753	51098	10036134	2132604
Tulare	123300	5977461	42352	10181065	3395057
State	1000375	85067824	294255	100325656	9849326

TABLE A7.1 Stratified summary statistics for the area within the sample unit frame. Combined ratio-regression with factor 5.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53472	320777	6347	12838
San Francisco	191654	0.21279	40782	1796	4395
Central Coast	1380040	0.31949	440909	12724	25683
South Coast	598866	0.45787	274203	8510	17223
Colorado Desert	818231	0.82144	672128	5679	11463
South Lahontan	235626	0.27383	64522	4402	8977
North Lahontan	175456	0.58726	103038	2297	4695
Sacramento	3388466	0.65059	2204502	30632	47100
San Joaquin	2788914	0.74635	2081506	36116	72456
Tulare	4080305	0.82168	3352705	41578	83401
State	14257457	0.67018	9555072	65621	129274

TABLE A7.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346491
San Francisco	512	2586376	5623	2778543	46405
Central Coast	13475	5789056	24725	7182571	465634
South Coast	62133	6289499	63810	6950499	338013
Colorado Desert	28421	11852213	10830	12698865	682957
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117981
Sacramento	211744	13452904	37823	17053114	2242325
San Joaquin	542467	6704753	51098	10036134	2132604
Tulare	123300	5977461	42352	10181065	3395057
State	1000375	85067824	294255	100325656	9849326

TABLE B1.1 Unstratified summary statistics for the area within the sample unit frame. Regression with factor 3.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53182	319037	6491	13060
San Francisco	191654	0.21192	40615	2014	4036
Central Coast	1380040	0.32579	449603	14794	29464
South Coast	598866	0.45251	270993	7270	14475
Colorado Desert	818231	0.82245	672954	5163	10351
South Lahontan	235626	0.27383	64522	4328	8827
North Lahontan	175456	0.58725	103036	2139	4343
Sacramento	3388466	0.65443	2217513	27684	55198
San Joaquin	2788914	0.75164	2096260	32463	64675
Tulare	4080305	0.81458	3323734	41129	82218
State	14257457	0.67040	9558269	62288	122707

TABLE B1.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	344751
San Francisco	512	2586376	5623	2778543	46238
Central Coast	13475	5789056	24725	7182571	474328
South Coast	62133	6289499	63810	6950499	334803
Colorado Desert	28421	11852213	10830	12698865	683784
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117979
Sacramento	211744	13452904	37823	17053114	2255337
San Joaquin	542467	6704753	51098	10036134	2147357
Tulare	123300	5977461	42352	10181065	3366086
State	1000375	85067824	294255	100325656	9852524

TABLE B2.1 Unstratified summary statistics for the area within the sample unit frame. Regression with factor 5.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53182	319037	7121	14320
San Francisco	191654	0.21192	40615	3653	4213
Central Coast	1380040	0.32579	449603	14904	29671
South Coast	598866	0.45251	270993	7228	14385
Colorado Desert	818231	0.82245	672954	5294	10604
South Lahontan	235626	0.27383	64522	4402	8977
North Lahontan	175456	0.58725	103036	2118	4300
Sacramento	3388466	0.65443	2217513	27684	55232
San Joaquin	2788914	0.75164	2096260	32379	64480
Tulare	4080305	0.81458	3323734	42721	85401
State	14257457	0.67040	9558269	63484	125063

TABLE B2.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	344751
San Francisco	512	2586376	5623	2778543	46238
Central Coast	13475	5789056	24725	7182571	474328
South Coast	62133	6289499	63810	6950499	334803
Colorado Desert	28421	11852213	10830	12698865	683784
South Lahontan	4377	16668221	17338	16908224	81860
North Lahontan	0	3891697	14942	4067153	117979
Sacramento	211744	13452904	37823	17053114	2255337
San Joaquin	542467	6704753	51098	10036134	2147357
Tulare	123300	5977461	42352	10181065	3366086
State	1000375	85067824	294255	100325656	9852524

TABLE B3.1 Unstratified summary statistics for the area within the sample unit frame. Unbiased ratio estimator.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.56517	339043	9814	19737
San Francisco	191654	0.18562	35575	2927	5863
Central Coast	1380040	0.31270	431539	19541	38890
South Coast	598866	0.44940	269131	9312	18535
Colorado Desert	818231	0.82626	676071	5425	10866
South Lahontan	235626	0.27683	65228	4310	8777
North Lahontan	175456	0.58990	103501	2291	4650
Sacramento	3388466	0.66009	2236692	33919	67634
San Joaquin	2788914	0.75685	2110790	38097	75858
Tulare	4080305	0.81593	3329243	40721	81361
State	14257457	0.67311	9596815	69905	137714

TABLE B3.2 Summary if irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	364758
San Francisco	512	2586376	5623	2778543	41198
Central Coast	13475	5789056	24725	7182571	456264
South Coast	62133	6289499	63810	6950499	332940
Colorado Desert	28421	11852213	10830	12698865	686901
South Lahontan	4377	16668221	17338	16908224	82567
North Lahontan	0	3891697	14942	4067153	118444
Sacramento	211744	13452904	37823	17053114	2274515
San Joaquin	542467	6704753	51098	10036134	2161888
Tulare					
State	1000375	85067824	294255	100325656	9891069

TABLE B4.1 Unstratified summary statistics for the area within the sample unit frame. Biased ratio estimator.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53344	320009	6533	13132
San Francisco	191654	0.20481	39253	2162	4331
Central Coast	1380040	0.32014	441806	16064	31976
South Coast	598866	0.45121	270215	7917	15756
Colorado Desert	818231	0.82386	674108	5155	10326
South Lahontan	235626	0.27291	64305	4383	8930
North Lahontan	175456	0.58925	103387	2218	4499
Sacramento	3388466	0.65745	2227747	29378	58553
San Joaquin	2788914	0.75532	2106523	35001	69751
Tulare	4080305	0.81445	3323204	40721	81361
State	14257457	0.67127	9570557	64538	127140

TABLE B4.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	345723
San Francisco	512	2586376	5623	2778543	44875
Central Coast	13475	5789056	24725	7182571	466531
South Coast	62133	6289499	63810	6950499	334024
Colorado Desert	28421	11852213	10830	12698865	684938
South Lahontan	4377	16668221	17338	16908224	81643
North Lahontan	0	3891697	14942	4067153	118330
Sacramento	211744	13452904	37823	17053114	2265570
San Joaquin	542467	6704753	51098	10036134	2157621
Tulare	123300	5977461	42352	10181065	3365556
State	1000375	85067824	294255	100325656	9864811

TABLE B5.1 Unstratified summary statistics for the area within the sample unit frame. Difference estimator.

Basin	Acres Within Frame (A1)	Proportion Irrig (A2)	Acres Irrig (A3)	Standard Error (acres) (A4)	95% C.I. (acres) (A5)
North Coast	599896	0.53461	320711	6377	12820
San Francisco	191654	0.20595	39471	2032	4069
Central Coast	1380040	0.32451	447837	14656	29188
South Coast	598866	0.45289	271221	7228	14379
Colorado Desert	818231	0.82255	673036	5073	10162
South Lahontan	235626	0.27320	64373	4192	8539
North Lahontan	175456	0.58881	103310	2134	4327
Sacramento	3388466	0.65911	2233372	31106	62043
San Joaquin	2788914	0.75565	2107443	35112	69946
Tulare	4080305	0.81677	3332670	40681	81239
State	14257457	0.67287	9593445	64924	127900

TABLE B5.2 Summary of irrigated and total acreages within the sample frame, outside of the sample frame and areas within the frame but not considered (excluded) in the sample design.

Basin	Excluded Acres (B1)	Outside Acres (B2)	Excl & Outside Irrig (B3)	Total Basin Acres (A1+B1+B2)	Total Basin Irrig (A3+B3)
North Coast	13946	11855644	25715	12469486	346425
San Francisco	512	2586376	5623	2778543	45094
Central Coast	13475	5789056	24725	7182571	472562
South Coast	62133	6289499	63810	6950499	335030
Colorado Desert	28421	11852213	10830	12698865	683866
South Lahontan	4377	16668221	17338	16908224	81711
North Lahontan	0	3891697	14942	4067153	118252
Sacramento	211744	13452904	37823	17053114	2271195
San Joaquin	542467	6704753	51098	10036134	2158541
Tulare	123300	5977461	42352	10181065	3375023
State	1000375	85067824	294255	100325656	9887699

Table C1.1 Weighted (co)variances as estimated from the unstratified elements and as "corrected" by equation 5.A.44.

Basin	Uncorrected s_y^2	Corrected s_y^2	Uncorrected s_x^2	Corrected s_x^2	Uncorrected s_{xy}	Corrected s_{xy}
North Coast	.13155	.13742	.12942	.13441	.12706	.13006
San Francisco	.13067	.09358	.13607	.09747	.12534	.08914
Central Coast	.18634	.17373	.18324	.16884	.17963	.16556
South Coast	.24056	.17610	.21108	.14871	.21770	.15202
Colorado Desert	.18165	.10997	.17927	.10469	.17909	.10356
South Lahontan	.06759	.06759	.05163	.05163	.05229	.05229
North Lahontan	.14659	.15147	.15414	.15636	.14548	.14882
Sacramento	.19772	.19869	.22803	.22846	.20969	.20776
San Joaquin	.19329	.17578	.22449	.20835	.20245	.18241
Tulare	.13324	.11862	.13183	.11612	.12917	.11250

Table C1.2. Correlations and standard errors for the regressions with factors 3 and 5 resulting from estimated and "corrected" (co)variances. (Weighted Model).

Basin	Uncor- rected r^2	Cor- rected r^2	Uncor- rected S.E.(3)	Cor- rected S.E.(3)	Uncor- rected S.E.(5)	Cor- rected S.E.(5)
North Coast	.94825	.91578	.01082	.01411	.01187	.01547
San Francisco	.88351	.87116	.01051	.00935	.01096	.00976
Central Coast	.94506	.93450	.01072	.01131	.01080	.01139
South Coast	.93338	.88248	.01214	.01380	.01207	.01372
Colorado Desert	.98494	.93157	.00631	.01047	.00647	.01073
South Lahontan	.78344	.78344	.01837	.01837	.01868	.01868
North Lahontan	.93665	.93516	.01219	.01254	.01207	.01241
Sacramento	.97521	.95094	.00817	.01152	.00817	.01152
San Joaquin	.94461	.90847	.01164	.01427	.01161	.01423
Tulare	.94989	.91881	.01008	.01211	.01047	.01257

%

Table C2.1. Unweighted (co)variances as estimated from the unstratified elements and as "corrected" by equation 5.A.44.

Basin	Uncorrected s_y^2	Corrected s_y^2	Uncorrected s_x^2	Corrected s_x^2	Uncorrected s_{xy}	Corrected s_{xy}
North Coast	.09118	.09072	.09101	.08773	.08371	.08007
San Francisco	.07492	.07086	.08447	.07737	.07269	.06872
Central Coast	.11843	.11681	.11772	.11499	.11404	.11168
South Coast	.06954	.05768	.06072	.04499	.05765	.04345
Colorado Desert	.07150	.01469	.07892	.01448	.07248	.01123
South Lahontan	.03863	.03863	.04193	.04193	.03504	.03504
North Lahontan	.04730	.04978	.04958	.04915	.04552	.04642
Sacramento	.08987	.08705	.09958	.09503	.09164	.08742
San Joaquin	.04888	.03447	.05401	.04152	.04423	.03012
Tulare	.06202	.04479	.05082	.03797	.05168	.03662

%

Table C2.2. Correlations and standard errors for the regressions with factors 3 and 5 resulting from estimated and "corrected" (co)variances. (Unweighted Model).

Basin	Uncorrected r^2	Corrected r^2	Uncorrected S.E.(3)	Corrected S.E.(3)	Uncorrected S.E.(5)	Corrected S.E.(5)
North Coast	.84447	.80550	.01562	.01743	.01686	.01881
San Francisco	.83489	.86150	.00947	.00844	.01020	.00908
Central Coast	.93279	.92862	.00945	.00968	.00950	.00972
South Coast	.78707	.72739	.01167	.01203	.01161	.01196
Colorado Desert	.93108	.59237	.00847	.00934	.00859	.00947
South Lahontan	.75796	.75796	.01468	.01468	.01618	.01618
North Lahontan	.88330	.88071	.00940	.00975	.00931	.00966
Sacramento	.93839	.92386	.00868	.00950	.00867	.00949
San Joaquin	.74103	.63380	.01265	.01264	.01276	.01274
Tulare	.84753	.78851	.01200	.01201	.01191	.01192

%

Table G1. County estimates based on the weighted-unstratified model.

County	Inside Acres	Prop Irrig	Acres Irrig	S.E. (acres)	Excl Acres	Outside Acres	Ex & Out Irrig	Total Acres	Total Irrig
Alameda	35808	0.20437	7318	2163	512	482615	5015	518935	12333
Alpine	14071	0.30028	4225	1059	0	451350	688	465421	4913
Amador	10907	0.76512	8345	1055	0	368034	316	378941	8662
Butte	325022	0.71862	233567	22570	14982	720116	2835	1060120	236403
Calaveras	0	0.	0	0	0	655715	1053	655715	1053
Colusa	401541	0.77695	311977	27883	5510	327667	43	734718	312020
Contra Costa	115508	0.52964	61178	9370	0	370500	6034	486008	67212
Del Norte	14919	0.58190	8681	1155	0	627571	0	642490	8681
El Dorado	19616	0.11699	2295	1352	0	1112280	1928	1131896	4223
Fresno	1499786	0.83009	1244957	100606	50435	2057156	7489	3607376	1252446
Glenn	370991	0.67563	250653	25762	10061	461687	860	842740	251512
Humboldt	75657	0.32805	24819	5856	0	2209857	3967	2285514	28786
Imperial	595280	0.85959	511697	22252	9603	2275803	3287	2880686	514984
Inyo	31683	0.34954	11074	3338	0	6439435	2230	6471118	13305
Kern	1208298	0.78631	950097	90248	0	4005342	35980	5213641	986076
Kings	691649	0.80384	555975	53036	35962	157713	486	885324	556461
Lake	44004	0.30721	13518	3056	0	800563	1056	844567	14575
Lassen	118355	0.50783	60104	6080	4314	2878058	11565	3000727	71670
Los Angeles	138798	0.18266	25353	13901	4377	2372842	7565	2516017	32918
Madera	410092	0.65970	270538	39664	38775	913974	8756	1362842	279294
Marin	0	0.	0	0	0	377393	461	377393	461
Mariposa	0	0.	0	0	0	894479	288	894479	288
Mendicino	72478	0.32023	23210	5610	0	2155594	279	2228072	23489
Merced	725187	0.78865	571919	70140	181085	457383	15482	1363655	587401
Modoc	203746	0.74043	150860	8938	6940	2456047	11077	2666732	161937
Mono	60610	0.58197	35273	3816	0	1951702	5544	2012312	40817
Monterey	511204	0.45914	234714	47051	8007	1536931	1125	2056142	235839
Napa	48339	0.31740	15343	3069	0	453412	187	501751	15530
Nevada	30313	0.12329	3737	2105	0	588398	1235	618711	4972
Orange	37611	0.41968	15785	3513	0	470837	2748	508448	18533
Placer	193399	0.33930	65620	13430	406	756248	859	950053	66479
Plumas	62085	0.18388	11416	4311	0	1588495	2333	1650580	13749
Riverside	434233	0.52227	226787	23375	43352	4078911	22781	4556495	249568
Sacramento	356086	0.62629	223013	20144	21945	270650	1557	648681	224570
San Benito	124741	0.37823	47181	9961	0	740074	1428	864815	48609
San Bernadino	95242	0.48689	46372	5343	0	12718372	21759	12813614	68131
San Diego	121421	0.41910	50888	10523	25192	2587831	22068	2734444	72955
San Francisco	0	0.	0	0	0	30443	0	30443	0
San Joaquin	710511	0.75262	534745	68721	121672	72061	5148	904244	539893
San Luis Obispo	534008	0.10701	57144	49150	2135	1578858	9658	2115002	66803
San Mateo	4605	0.48262	2222	306	0	270995	827	275600	3049
Santa Barbara	155176	0.44387	68878	14282	579	1475465	10773	1631220	79651
Santa Clara	55357	0.49159	27213	3377	0	773207	720	828564	27933
Santa Cruz	40236	0.57639	23192	3703	2754	242212	1169	285202	24361
Shasta	112608	0.49302	55518	7819	0	2320549	8426	2433157	63944
Sierra	27840	0.68768	19145	1933	0	581838	634	609678	19780
Siskiyou	299259	0.64113	191864	22284	7006	3902853	16814	4209118	208678
Solano	278318	0.63448	176587	16788	78822	222472	2558	579612	179145
Sonoma	118844	0.21798	25906	7315	0	887805	3642	1006649	29548
Stanislaus	489547	0.82526	404004	47349	191784	277631	5160	958961	409164
Sutter	333264	0.85744	285754	23142	51784	0	0	385048	285754
Tehama	218727	0.49919	109186	15188	8052	1641421	4766	1868200	113952
Trinity	0	0.	0	0	0	2028052	802	2028052	802
Tulare	929336	0.83065	771953	71261	36903	2133106	4163	3099346	776116
Tuolumne	0	0.	0	0	0	1436630	641	1436630	641
Ventura	152542	0.64299	98083	14247	12407	991918	3602	1156868	101685
Yolo	471407	0.71555	337315	32735	19351	155340	146	646099	337462
Yuba	127194	0.71585	91052	8832	5667	273928	2240	406789	93292
State	14257459		9558250		1000375	85067840	294255	100325656	9852506

Table G2. County estimates based on the unweighted-unstratified model.

County	Inside Acres	Prop Irrig	Acres Irrig	S.E. (acres)	Excl Acres	Outside Acres	Ex & Out Irrig	Total Acres	Total Irrig
Alameda	35808	0.21543	7714	2243	512	482615	5015	518935	12729
Alpine	14071	0.30919	4351	824	0	451350	688	465421	5038
Amador	10907	0.77395	8441	1010	0	368034	316	378941	8758
Butte	325022	0.72631	236067	23554	14982	720116	2835	1060120	238902
Calaveras	0	0.00000	0	0	0	655715	1053	655715	1053
Colusa	401541	0.78468	315081	29100	5510	327667	43	734718	315124
Contra Costa	115508	0.55026	63559	8988	0	370500	6034	486008	69593
Del Norte	14919	0.56539	8435	1627	0	627571	0	642490	8435
El Dorado	19616	0.12423	2437	1422	0	1112280	1928	1131896	4365
Fresno	1499786	0.83178	1247492	107355	50435	2057156	7489	3607376	1254981
Glenn	370991	0.68329	253494	26886	10061	461687	860	842740	254354
Humboldt	75657	0.32757	24783	8249	0	2209857	3967	2285514	28750
Imperial	595280	0.84640	503845	17382	9603	2275803	3287	2880686	507132
Inyo	31683	0.31067	9843	2668	0	6439435	2230	6471118	12073
Kern	1208298	0.78398	947282	96688	0	4005342	35980	5213641	983261
Kings	691649	0.80428	556279	56840	35962	157713	486	885324	556765
Lake	44004	0.31460	13844	3189	0	800563	1056	844567	14900
Lassen	118355	0.51355	60781	5477	4314	2878058	11565	3000727	72347
Los Angeles	138798	0.17495	24283	11119	4377	2372842	7565	2516017	31847
Madera	410092	0.67823	278137	37983	38775	913974	8756	1362842	286893
Marin	0	0.00000	0	0	0	377393	461	377393	461
Mariposa	0	0.00000	0	0	0	894479	288	894479	288
Mendicino	72478	0.32025	23211	7902	0	2155594	279	2228072	23490
Merced	725187	0.79532	576756	67167	181085	457383	15482	1363655	592238
Modoc	203746	0.73451	149653	9619	6940	2456047	11077	2666732	160730
Mono	60610	0.57825	35048	2982	0	1951702	5544	2012312	40591
Monterey	511204	0.45673	233482	42675	8007	1536931	1125	2056142	234607
Napa	48339	0.31635	15292	3188	0	453412	187	501751	15479
Nevada	30313	0.13053	3957	2197	0	588398	1235	618711	5192
Orange	37611	0.42371	15936	3597	0	470837	2748	508448	18684
Placer	193399	0.34671	67053	14016	406	756248	859	950053	67912
Plumas	62085	0.19117	11869	4499	0	1588495	2333	1650580	14202
Riverside	434233	0.52625	228515	23440	43352	4078911	22781	4556495	251296
Sacramento	356086	0.63971	227792	20568	21945	270650	1557	648681	229348
San Benito	124741	0.38641	48201	9050	0	740074	1428	864815	49629
San Bernadino	95242	0.47788	45514	5152	0	12718372	21759	12813614	67273
San Diego	121421	0.42403	51486	11083	25192	2587831	22068	2734444	73554
San Francisco	0	0.00000	0	0	0	30443	0	30443	0
San Joaquin	710511	0.76260	541836	65808	121672	72061	5148	904244	546984
San Luis Obispo	534008	0.10878	58089	44579	2135	1578858	9658	2115002	67748
San Mateo	4605	0.47109	2169	318	0	270995	827	275600	2996
Santa Barbara	155176	0.44164	68532	12954	579	1475465	10773	1631220	79305
Santa Clara	55357	0.48955	27100	3155	0	773207	720	828564	27820
Santa Cruz	40236	0.57260	23039	3359	2754	242212	1169	285202	24208
Shasta	112608	0.50054	56365	8161	0	2320549	8426	2433157	64791
Sierra	27840	0.69535	19359	2018	0	581838	634	609678	19993
Siskiyou	299259	0.62074	185762	32239	7006	3902853	16814	4209118	202576
Solano	278318	0.64149	178538	17517	78822	222472	2558	579612	181096
Sonoma	118844	0.22442	26671	10159	0	887805	3642	1006649	30313
Stanislaus	489547	0.82857	405624	45342	191784	277631	5160	958961	410784
Sutter	333264	0.86524	288353	24152	51784	0	0	385048	288353
Tehama	218727	0.50672	110833	15851	8052	1641421	4766	1868200	115599
Trinity	0	0.00000	0	0	0	2028052	802	2028052	802
Tulare	929336	0.83212	773319	76373	36903	2133106	4163	3099346	777483
Tuolumne	0	0.00000	0	0	0	1436630	641	1436630	641
Ventura	152542	0.62927	95990	14591	12407	991918	3602	1156868	99592
Yolo	471407	0.72325	340945	34163	19351	155340	146	646099	341091
Yuba	127194	0.72354	92030	9218	5667	273928	2240	406789	94270
State	14257459		9594468		1000375	85067840	294255	100325656	9888726

G3. County Estimates Based on the Weighted-Stratified Regression Model.

County	Inside Acres	Prop Irrig	Acres Irrig	S.E. (acres)	Excl Acres	Outside Acres	Ex & Out Irrig	Total Acres	Total Irrig
Alameda	35808	0.21271	7617	1363	512	482615	5015	518935	12632
Alpine	14071	0.28393	3995	1362	0	451350	688	465421	4683
Amador	10907	0.75570	8242	1155	0	368034	316	378941	8559
Butte	325022	0.71515	232439	16898	14982	720116	2835	1060120	235275
Calaveras	0	0.00000	0	0	0	655715	1053	655715	1053
Colusa	401541	0.76777	308291	23611	5510	327667	43	734718	308334
Contra Costa	115508	0.51643	59652	7880	0	370500	6034	486008	65686
Del Norte	14919	0.59679	8904	964	0	627571	0	642490	8904
El Dorado	19616	0.11091	2176	1054	0	1112280	1928	1131896	4104
Fresno	1499786	0.82850	1242573	74359	50435	2057156	7489	3607376	1250062
Glenn	370991	0.67171	249198	19280	10061	461687	860	842740	250058
Humboldt	75657	0.33422	25286	4695	0	2209857	3967	2285514	29253
Imperial	595280	0.85281	507661	32788	9603	2275803	3287	2880686	510948
Inyo	31683	0.34954	11074	3953	0	6439435	2230	6471118	13305
Kern	1208298	0.79780	963980	83445	0	4005342	35980	5213641	999960
Kings	691649	0.81072	560734	51728	35962	157713	486	885324	561219
Lake	44004	0.37471	16489	5185	0	800563	1056	844567	17545
Lassen	118355	0.50972	60328	5769	4314	2878058	11565	3000727	71893
Los Angeles	138798	0.18364	25489	16477	4377	2372842	7565	2516017	33054
Madera	410092	0.65829	269959	31462	38775	913974	8756	1362842	278716
Marin	0	0.00000	0	0	0	377393	461	377393	461
Mariposa	0	0.00000	0	0	0	894479	288	894479	288
Mendicino	72478	0.28751	20838	3930	0	2155594	279	2228072	21117
Merced	725187	0.78281	567684	63106	181085	457383	15482	1363655	583166
Modoc	203746	0.74188	151155	9529	6940	2456047	11077	2666732	162232
Mono	60610	0.57531	34870	4854	0	1951702	5544	2012312	40413
Monterey	511204	0.44126	225574	37998	8007	1536931	1125	2056142	226699
Napa	48339	0.34061	16465	6950	0	453412	187	501751	16652
Nevada	30313	0.11863	3596	1629	0	588398	1235	618711	4831
Orange	37611	0.44287	16657	3430	0	470837	2748	508448	19405
Placer	193399	0.34047	65847	6831	406	756248	859	950053	66705
Plumas	62085	0.17697	10987	4522	0	1588495	2333	1650580	13320
Riverside	434233	0.53803	233630	19032	43352	4078911	22781	4556495	256411
Sacramento	356086	0.62820	223693	16401	21945	270650	1557	648681	225250
San Benito	124741	0.36029	44943	5151	0	740074	1428	864815	46371
San Bernadino	95242	0.51537	49085	6100	0	12718372	21759	12813614	70844
San Diego	121421	0.42802	51971	10701	25192	2587831	22068	2734444	74038
San Francisco	0	0.00000	0	0	0	30443	0	30443	0
San Joaquin	710511	0.75291	534951	49707	121672	72061	5148	904244	540099
San Luis Obispo	534008	0.07288	38919	19806	2135	1578858	9658	2115002	48577
San Mateo	4605	0.48735	2244	445	0	270995	827	275600	3071
Santa Barbara	155176	0.48159	74731	16109	579	1475465	10773	1631220	85504
Santa Clara	55357	0.48496	26846	2878	0	773207	720	828564	27566
Santa Cruz	40236	0.60447	24321	4778	2754	242212	1169	285202	25491
Shasta	112608	0.50415	56771	6325	0	2320549	8426	2433157	65198
Sierra	27840	0.68219	18992	2028	0	581838	634	609678	19627
Siskiyou	299259	0.65820	196972	17776	7006	3902853	16814	4209118	213786
Solano	278318	0.62750	174645	13621	78822	222472	2558	579612	177202
Sonoma	118844	0.19223	22845	9727	0	887805	3642	1006649	26487
Stanislaus	489547	0.81956	401213	37088	191784	277631	5160	958961	406373
Sutter	333264	0.85702	285614	19616	51784	0	0	385048	285614
Tehama	218727	0.50926	111389	11912	8052	1641421	4766	1868200	116155
Trinity	0	0.00000	0	0	0	2028052	802	2028052	802
Tulare	929336	0.83789	778681	54970	36903	2133106	4163	3099346	782845
Tuolumne	0	0.00000	0	0	0	1436630	641	1436630	641
Ventura	152542	0.61130	93249	10428	12407	991918	3602	1156868	96851
Yolo	471407	0.71350	336349	26875	19351	155340	146	646099	336495
Yuba	127194	0.71630	91109	5304	5667	273928	2240	406789	93349
State	14257459		9550923		1000375	85067840	294255	100325656	9845179
%									

G4. County Estimates Based on the Unweighted-Stratified Model.

County	Inside Acres	Prop Irrig	Acres Irrig	S.E. (acres)	Excl Acres	Outside Acres	Ex & Out Irrig	Total Acres	Total Irrig
Alameda	35808	0.24838	8894	2227	512	482615	5015	518935	13909
Alpine	14071	0.28558	4018	1114	0	451350	688	465421	4706
Amador	10907	0.76837	8381	1238	0	368034	316	378941	8697
Butte	325022	0.72317	235046	18689	14982	720116	2835	1060120	237882
Calaveras	0	0.00000	0	0	0	655715	1053	655715	1053
Colusa	401541	0.77818	312471	25815	5510	327667	43	734718	312514
Contra Costa	115508	0.54427	62868	8299	0	370500	6034	486008	68901
Del Norte	14919	0.58558	8736	1829	0	627571	0	642490	8736
El Dorado	19616	0.13373	2623	863	0	1112280	1928	1131896	4552
Fresno	1499786	0.82931	1243787	86343	50435	2057156	7489	3607376	1251277
Glenn	370991	0.67729	251268	21013	10061	461687	860	842740	252128
Humboldt	75657	0.32944	24924	8903	0	2209857	3967	2285514	28892
Imperial	595280	0.84183	501125	45848	9603	2275803	3287	2880686	504411
Inyo	31683	0.31067	9843	3159	0	6439435	2230	6471118	12073
Kern	1208298	0.79582	961588	97570	0	4005342	35980	5213641	997567
Kings	691649	0.81307	562359	60505	35962	157713	486	885324	562845
Lake	44004	0.34806	15316	4840	0	800563	1056	844567	16372
Lassen	118355	0.51382	60813	4658	4314	2878058	11565	3000727	72379
Los Angeles	138798	0.17767	24660	13177	4377	2372842	7565	2516017	32225
Madera	410092	0.67107	275200	31737	38775	913974	8756	1362842	283957
Marin	0	0.00000	0	0	0	377393	461	377393	461
Mariposa	0	0.00000	0	0	0	894479	288	894479	288
Mendicino	72478	0.29115	21102	5048	0	2155594	279	2228072	21381
Merced	725187	0.78895	572136	67479	181085	457383	15482	1363655	587619
Modoc	203746	0.73971	150713	10900	6940	2456047	11077	2666732	161790
Mono	60610	0.57494	34847	3961	0	1951702	5544	2012312	40391
Monterey	511204	0.44558	227782	30780	8007	1536931	1125	2056142	228907
Napa	48339	0.35353	17089	5678	0	453412	187	501751	17277
Nevada	30313	0.13927	4222	1333	0	588398	1235	618711	5457
Orange	37611	0.43231	16260	2831	0	470837	2748	508448	19008
Placer	193399	0.33996	65748	6755	406	756248	859	950053	66607
Plumas	62085	0.17485	10856	4923	0	1588495	2333	1650580	13188
Riverside	434233	0.52858	229527	22315	43352	4078911	22781	4556495	252308
Sacramento	356086	0.64308	228992	17591	21945	270650	1557	648681	230548
San Benito	124741	0.36106	45039	4214	0	740074	1428	864815	46467
San Bernadino	95242	0.49465	47111	5179	0	12718372	21759	12813614	68870.
San Diego	121421	0.42829	52003	8342	25192	2587831	22068	2734444	74071
San Francisco	0	0.00000	0	0	0	30443	0	30443	0.
San Joaquin	710511	0.76406	542873	52805	121672	72061	5148	904244	548021
San Luis Obispo	534008	0.08300	44323	14365	2135	1578858	9658	2115002	53981
San Mateo	4605	0.44931	2069	431	0	270995	827	275600	2896
Santa Barbara	155176	0.49279	76469	13131	579	1475465	10773	1631220	87242
Santa Clara	55357	0.46540	25763	2612	0	773207	720	828564	26483
Santa Cruz	40236	0.60379	24294	3890	2754	242212	1169	285202	25463
Shasta	112608	0.50211	56542	6662	0	2320549	8426	2433157	64963
Sierra	27840	0.68643	19110	2208	0	581838	634	609678	19745
Siskiyou	299259	0.64514	193064	33706	7006	3902853	16814	4209118	209878
Solano	278318	0.63450	176593	14896	78822	222472	2558	579612	179151
Sonoma	118844	0.21134	25116	12605	0	887805	3642	1006649	28758
Stanislaus	489547	0.81826	400577	39340	191784	277631	5160	958961	405737
Sutter	333264	0.86325	287690	21372	51784	0	0	385048	287690.
Tehama	218727	0.50640	110763	11691	8052	1641421	4766	1868200	115529
Trinity	0	0.00000	0	0	0	2028052	802	2028052	802
Tulare	929336	0.83627	777176	63436	36903	2133106	4163	3099346	781339
Tuolumne	0	0.00000	0	0	0	1436630	641	1436630	641
Ventura	152542	0.58209	88793	8027	12407	991918	3602	1156868	92395
Yolo	471407	0.71331	338616	29345	19351	155340	146	646099	338763
Yuba	127134	0.72041	91632	5925	5667	273928	2240	406789	93872
State	14257459		9578813		1000375	85067840	294255	100325656	9873069

%

Table H1: COUNTY ESTIMATES

A Comparison of DWR Estimates with APT Estimates Based on Weighted-Unstratified Values

COUNTY	ESTIMATES BY COUNTY (IN THOUSANDS OF ACRES) DWR	APT	DIFFERENCE (IN THOUSANDS OF ACRES)	DIFFERENCE AS PERCENT OF DWR'S ESTIMATE
Alameda	14.4	12.3	-2.1	-14.6
Alpine	6.3	4.9	-1.4	-22.2
Amador	4.6	8.7	+4.1	+89.1
Butte	252.6	236.4	-16.2	- 6.4
Calaveras	2.7	1.1	- 1.6	-59.3
Colusa	307.5	312.0	+ 4.5	+ 1.5
Contra Costa	58.3	67.2	+ 8.9	+15.3
Del Norte	5.8	8.7	+ 2.9	+50.0
El Dorado	7.1	4.2	- 2.9	-40.8
Fresno	1310.8	1252.4	-58.4	- 4.5
Glenn	240.0	251.5	+11.5	+ 4.8
Humboldt	24.8	28.8	+ 4.0	+16.1
Imperial	527.4	515.0	-12.4	- 2.4
Inyo	16.5	13.3	- 3.2	-19.4
Kern	991.5	986.0	- 5.5	- .6
Kings	613.7	556.5	-57.2	- 9.3
Lake	16.3	14.6	- 1.7	-10.4
Lassen	80.2	71.7	- 8.5	-10.6
Los Angeles	41.2	32.9	- 8.3	-20.1
Madera	353.1	279.3	-73.8	-20.9
Marin	.6	.5	- .1	-16.7
Mariposa	.8	.3	- .5	-62.5
Mendocino	21.7	23.5	+ 1.8	+ 8.3
Merced	492.4	587.4	+95.0	+19.3
Modoc	172.0	161.9	-10.1	- 5.9
Mono	36.8	40.8	+ 4.0	+10.9
Monterey	184.5	235.8	+51.3	+27.8
Napa	18.0	15.5	- 2.5	-13.9
Nevada	11.1	5.0	- 6.1	-55.0
Orange	21.2	18.5	- 2.7	-12.7
Placer	42.3	66.5	+24.2	+57.2
Plumas	37.7	13.7	-24.0	-63.7
Riverside	250.0	249.6	- .4	- .2
Sacramento	196.7	224.6	+27.9	+14.2
San Benito	54.2	48.6	- 5.6	-10.3
San Bernardino	72.1	68.1	- 4.0	- 5.5
San Diego	85.0	73.0	-12.0	-14.1
San Francisco	0.0	0.0	-	-
San Joaquin	573.2	539.9	-33.3	-5.8
San Luis Obispo	58.4	66.8	+ 8.4	+14.4
San Mateo	4.8	3.0	- 1.8	-37.5
Santa Barbara	90.9	79.7	-11.2	-12.3
Santa Clara	42.0	27.9	-14.1	-33.6
Santa Cruz	24.0	24.4	+ .4	+ 1.7
Shasta	53.0	63.9	+10.9	+20.6
Sierra	16.4	19.8	+ 3.4	+20.7
Siskiyou	196.0	208.7	+12.7	+6.5
Solano	179.6	179.1	- .5	- .3
Sonoma	35.0	29.5	- 5.5	-15.7
Stanislaus	402.0	409.2	+ 7.2	+ 1.8
Sutter	298.6	285.8	-12.8	- 4.3
Tehama	97.2	114.0	+16.8	+17.3
Trinity	1.4	.8	- .6	-42.9
Tulare	710.9	776.1	+65.2	+ 9.2
Tuolumne	2.9	.6	- 2.3	-79.3
Ventura	111.9	101.7	-10.2	- 9.1
Yolo	327.0	337.5	+10.5	+ 3.2
Yuba	97.9	93.3	- 4.6	- 4.7
STATE	9894.2	9852.5	-41.7	- 0.4

Table H2. COUNTY ESTIMATES

A COMPARISON OF DWR ESTIMATES WITH APT
ESTIMATES BASED ON UNWEIGHTED-UNSTRATIFIED VALUES

COUNTY	ESTIMATES BY COUNTY (IN THOUSANDS OF ACRES) DWR	APT	DIFFERENCE (IN THOUSANDS OF ACRES)	DIFFERENCE AS PERCENT OF DWR'S ESTIMATE
ALAMEDA	14.4	12.7	- 1.7	-11.8
ALPINE	6.3	5.0	- 1.3	-20.6
AMADOR	4.6	8.8	+ 4.2	+91.3
BUTTE	252.6	238.9	-13.7	- 5.4
CALAVERAS	2.7	1.1	- 1.6	-59.3
COLUSA	307.5	315.1	+ 7.6	+ 2.5
CONTRA COSTA	58.3	69.6	+11.3	+19.4
DEL NORTE	5.8	8.4	+ 2.6	+44.8
EL DORADO	7.1	4.4	- 2.7	-38.0
FRESNO	1310.8	1255.0	-55.8	- 4.3
GLENN	240.0	254.4	+14.4	+ 6.0
HUMBOLDT	24.8	28.8	+ 4.0	+16.1
IMPERIAL	527.4	507.1	-20.3	- 3.8
INYO	16.5	12.1	- 4.4	-26.7
KERN	991.5	983.3	- 8.2	- .8
KINGS	613.7	556.8	-56.9	- 9.3
LAKE	16.3	14.9	- 1.4	- 8.6
LASSEN	80.2	72.3	- 7.9	- 9.9
LOS ANGELES	41.2	31.8	- 9.4	-22.8
MADERA	353.1	286.9	-66.2	-18.7
MARIN	.6	.5	- .1	-16.7
MARIPOSA	.8	.3	- .5	-62.5
MENDOCINO	21.7	23.5	+ 1.8	+ 8.3
MERCED	492.4	592.2	+99.8	+20.3
MODOC	172.0	160.7	-11.3	- 6.6
MONO	36.8	40.6	+ 3.8	+10.3
MONTEREY	184.5	234.6	+50.1	+27.2
NAPA	18.0	15.5	- 2.5	-13.9
NEVADA	11.1	5.2	- 5.9	-53.2
ORANGE	21.2	18.7	- 2.5	-11.8
PLACER	42.3	67.9	+25.6	+60.5
PLUMAS	37.7	14.2	-23.5	-62.3
RIVERSIDE	250.0	251.3	+1.3	+ .5
SACRAMENTO	196.7	229.3	+32.6	+16.6
SAN BENITO	54.2	49.6	- 4.6	- 8.5
SAN BERNARDINO	72.1	67.3	- 4.8	- 6.7
SAN DIEGO	85.0	73.6	-11.4	-15.8
SAN FRANCISCO	0.0	0.0	-	-
SAN JOAQUIN	573.2	547.0	-26.2	- 4.6
SAN LUIS OBISPO	58.4	67.7	+ 9.3	+15.9
SAN MATEO	4.8	3.0	- 1.8	-37.5
SANTA BARBARA	90.9	79.3	-11.6	-12.8
SANTA CLARA	42.0	27.8	-14.2	-33.8
SANTA CRUZ	24.0	24.2	+ .2	+ .8
SHASTA	53.0	64.8	+11.8	+22.3
SIERRA	16.4	20.0	+ 3.6	+22.0
SISKIYOU	196.0	202.6	+ 6.6	+ 3.4
SOLANO	179.6	181.1	+ 1.5	+ .8
SONOMA	35.0	30.3	- 4.7	-13.4
STANISLAUS	402.0	410.8	+ 8.8	+ 2.2
SUTTER	298.6	288.4	-10.2	- 3.4
TEHAMA	97.2	115.6	+18.4	+18.9
TRINITY	1.4	.8	- .6	-42.9
TULARE	710.9	777.5	+66.6	+ 9.4
TUOLUMNE	2.9	.6	- 2.3	-79.3
VENTURA	111.9	99.6	-12.3	-11.0
YOLO	327.0	341.1	+14.1	+ 4.3
YUBA	97.9	94.3	- 3.6	- 3.7
STATE	9894.2	9888.7	-89.0	- .9

Table H3. COUNTY ESTIMATES

A Comparison of DWR Estimates with APT
Estimates Based on Weighted Stratified Values

COUNTY	ESTIMATES BY COUNTY (IN THOUSANDS OF ACRES) DWR	APT	DIFFERENCE (IN THOUSANDS OF ACRES)	DIFFERENCE AS PERCENT OF DWR'S ESTIMATE
Alameda	14.4	12.6	- 1.8	-12.5
Alpine	6.3	4.7	- 1.6	-25.4
Amador	4.6	8.6	+ 4.0	+87.0
Butte	252.6	235.3	-17.3	- 6.8
Calaveras	2.7	1.1	- 1.6	-59.3
Colusa	307.5	308.3	+ 0.8	+ 0.3
Contra Costa	58.3	65.7	+ 7.4	+12.7
Del Norte	5.8	8.9	+ 3.1	+34.8
El Dorado	7.1	4.1	- 3.0	-42.3
Fresno	1310.8	1250.1	-60.7	- 4.6
Glenn	240.0	250.1	+10.1	+ 4.2
Humboldt	24.8	29.3	+ 4.5	+18.1
Imperial	527.4	510.9	-16.5	- 3.1
Inyo	16.5	13.3	- 3.2	-19.4
Kern	991.5	1000.0	+ 8.5	+ 0.9
Kings	613.7	561.2	-52.5	- 8.6
Lake	16.3	17.5	+ 1.2	+ 7.4
Lassen	80.2	71.9	- 8.3	-10.3
Los Angeles	41.2	33.1	- 8.1	-19.7
Madera	353.1	278.7	-74.4	-21.1
Marin	.6	.5	- 0.1	-16.7
Mariposa	.8	.3	- 0.5	-62.5
Mendocino	21.7	21.1	- 0.6	- 2.8
Merced	492.4	583.2	+90.8	+18.4
Modoc	172.0	162.2	- 9.8	- 5.7
Mono	36.8	40.4	+ 3.6	+ 9.8
Monterey	184.5	226.7	+42.2	+22.9
Napa	18.0	16.7	- 1.3	- 7.2
Nevada	11.1	4.8	- 6.3	-56.8
Orange	21.1	19.4	- 1.8	- 8.5
Placer	42.3	66.7	+24.4	+57.7
Plumas	37.7	13.3	-24.4	-64.7
Riverside	250.0	256.4	+ 6.4	+ 2.6
Sacramento	196.7	225.3	+28.6	+14.5
San Benito	54.2	46.4	- 7.8	-14.4
San Bernardino	72.1	70.8	- 1.3	- 1.8
San Diego	85.0	74.0	-11.0	-12.9
San Francisco	0.0	0.0	-	-
San Joaquin	573.2	540.1	-33.1	- 5.8
San Luis Obispo	58.4	48.6	- 9.8	-16.8
San Mateo	4.8	3.1	- 1.7	-35.4
Santa Barbara	90.9	85.5	- 5.4	- 5.9
Santa Clara	42.0	27.6	-14.4	-34.3
Santa Cruz	24.0	25.5	+1.5	+ 6.3
Shasta	53.0	65.2	+12.2	+23.0
Sierra	16.4	19.6	+ 3.2	+19.5
Siskiyou	196.0	213.8	+17.8	+ 9.1
Solano	179.6	177.2	- 2.4	- 1.3
Sonoma	35.0	26.5	- 8.5	-24.3
Stanislaus	402.0	406.4	+ 4.4	+ 1.1
Sutter	298.6	285.6	-13.0	- 4.4
Tehama	97.2	116.2	+19.0	+19.5
Trinity	1.4	.8	- .6	-42.9
Tulare	710.9	782.8	+71.9	+10.1
Tuolumne	2.9	.6	- 2.3	-79.3
Ventura	111.9	96.9	-15.0	-13.4
Yolo	327.0	336.5	+ 9.5	+ 2.9
Yuba	97.9	93.3	- 4.6	- 4.7
STATE	9894.2	9845.2	-49.0	- .5

APPENDIX III

RELATIVE EFFICIENCIES COMPUTED FOR THE UNSTRATIFIED CASE

APPENDIX III: Relative Efficiencies Computed for the Unstratified Case

The values reported in Table 1 below represent measures of relative efficiency for an unstratified, simple random ground sample design versus a corresponding unstratified, Landsat-ground regression design. RE_1 and RE_2 figures were computed according to unstratified versions of formulas presented in Section 3.7.5 .

Two sets of RE_1 and RE_2 are given in the Table. The set on the left give efficiency values when variances were not corrected for the fact that samples were allocated according to optimal as opposed to proportional-to-stratum-size rules. As explained on pages 6-8 of Appendix II, an incorrect estimate of the unstratified variance obtained by combination of strata observations may result when sample units were not originally drawn at random (and therefore with relative stratum sample size approximately proportional to relative stratum area) from the entire population of sample units. To correct for this effect, formulas 13', 13", 14' and 15' of Appendix II were applied to generate "inherently unstratified" estimates of regression and simple random sample ground variance. The resulting RE_1 and RE_2 values based on these "inherently unstratified" variance estimates are given on the right side of Table 1.

The pattern of these results is similar to that seen for the stratified case. Efficiency values tend to be larger in the unstratified case since (1) ground sample-only variance is significantly inflated by the absence of strata and (2) corresponding regression variance inflation is only moderate due to the use of the relationship between Landsat (X) and ground observations (Y). Correction of variances to simulate an "inherently unstratified" original allocation of ground sample units did not significantly affect the size of RE_1 and RE_2 values, with the exception of the Colorado Desert unit.

Table 1: Relative Sampling Efficiency for Unstratified Regression Estimation with Factor 5 Relative to Unstratified, Unweighted Random Sampling.

Basin	Not Corrected for Original Allocation RE_1	Not Corrected for Original Allocation RE_2	Corrected for Original Allocation RE_1	Corrected for Original Allocation RE_2
North Coast	10.67	3.73	10.62	3.73
San Francisco	1.44	1.14	1.36	1.13
Central Coast	11.11	4.68	10.96	4.67
South Coast	4.28	2.26	3.55	2.09
Colorado Desert	24.01	4.57	4.93	2.86
South Lahonton*	2.39	1.70	2.39	1.70
North Lahonton	4.91	1.81	5.17	1.81
Sacramento	17.80	9.83	17.24	9.67
San Joaquin	4.47	3.66	3.15	2.76
Tulare	8.35	6.42	6.03	5.00
Statewide	9.34	4.30	7.68	3.83

* originally an unstratified regression sample

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR 166370		

4. Title and Subtitle

IRRIGATED LANDS ASSESSMENT FOR WATER MANAGEMENT:
TECHNIQUE TEST

5. Report Date
April 1981

6. Performing Organization Code

7. Author(s) Sharon L. Wall,* Catherine E. Brown,* Marian Eriksson,* Carolyn A. Grigg,* Randall W. Thomas,* Robert N. Colwell,* John E. Estes,† Larry R. Tinney,† Jay O. Baggett,† Glenn Sawyer,‡ and Charles Ferchaud‡

8. Performing Organization Report No.

9. Performing Organization Name and Address

*University of California, Berkeley, Calif. 94720
†University of California, Santa Barbara, Calif. 93106
‡California Department of Water Resources, Sacramento, Calif. 95814

10. Work Unit No.
T3813

11. Contract or Grant No.
NCC-2-54

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Washington, D.C. 20546

13. Type of Report and Period Covered
Contractor Report
Annual Report

14. Sponsoring Agency Code
677-60-05

15. Supplementary Notes

Point of Contact: Ethel Bauer
NASA Ames Research Center, M.S. 242-4
Moffett Field, Calif. 94035 (415) 965-5513 or FTS 448-5513

16. Abstract

A procedure for estimation of irrigated land using full frame Landsat imagery was demonstrated. Relatively inexpensive interpretation of multidate Landsat photographic enlargements was used to produce a map of irrigated land in California. Landsat and ground maps were then linked by regression equations to enable precise estimation of irrigated land area by county, basin, and statewide. Land irrigated at least once in California during 1979 was estimated to be 9.86 million acres, with an expected error of less than 1.75 percent at the 99 percent level of confidence. To achieve the same level of error with a ground-only sample would have required 3 to 5 times as many ground sample units statewide. A procedure for relatively inexpensive computer classification of Landsat digital data to irrigated land categories was also developed. The procedure is based upon ratios of Landsat band 7 and 5, and gave good results for several counties in the California Central Valley.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Landsat Agriculture Inventory Irrigated land	Unclassified - Unlimited Subject Category 43

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Classified	Classified	339	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

End of Document