United States
Department of
Agriculture

Statistical
Reporting
Service

**Statistical
Research
Division**

Washington, D.C.

March 1984

1982 Results for the California Cooperative Remote Sensing Project

Martin L. Holko

1982 Results for the California Cooperative Remote Sensing Project

by

Martin L. Holko
Statistical Reporting Service
U. S. Department of Agriculture
Washington, D.C.

March 1984

1982 Results for the California Cooperative Remote Sensing Project. By Martin L. Holko. Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C. March 1984. SRS Staff Report No. AGES840305.

Abstract

This report presents the results of the California Cooperative Project in Remote Sensing. The report discusses the extension of SRS's Landsat regression estimation procedures (used in mid-western states for estimating the acreages of corn, soybeans, winter wheat and rice) to the multiple crop environment in California. Results of large area estimation, county level estimation, and crop-odds mapping are presented. In addition, results of an independent training procedure, which examines the bias of SRS's estimate of the estimator's precision, are presented. From these results the author recommends the use of SRS's Landsat regression procedures for estimating crops at the state and county level, in California, provided questions on the estimator's precision can be answered.

Acknowledgments

The author would like to thank the following for their strong support:

Richard Sigman, for his supervision and technical support.

The California Crop and Livestock Reporting Service, for their development of the Ground-Data Set.

Jay Baggett, for his devotion in completing the digitization work.

Ed Sheffner, and the staff at NASA Ames Research Center, for their assistance in purchasing and processing the Landsat Data Sets.

Dave Kleweno, for his assistance in developing the Ground-Data Set and his analysis of area AD48/32.

A special thanks goes to Pat Joyce and Patsy Nelson for their assistance in preparing this report.

Table of Contents

List of Tables

List of Figures

Summary

This report presents the results of research conducted by the Statistical Reporting Service for the California Cooperative Remote Sensing Project.

First, an outline of current Landsat analysis procedures used to estimate acreages of winter wheat, corn and soybeans in mid-western states is presented. Results of extending these procedures to obtain land-cover and crop acreage estimates are given showing the potential for increasing the precision of the acreage estimates.

Results of applying the Battese-Fuller Estimator to obtain county crop acreage estimates are discussed. Despite problems with Landsat coverage and sample allocations county estimates for rice, winter wheat and sugarbeets are presented.

An alternative land resource mapping procedure called "crop-odds" mapping is presented. "Crop-odds" mapping uses a calculated posterior probability and a user supplied reliability level to produce land resource maps.

Since little was known about the properties of the Landsat based estimators when used for minor cover types, results of an independent test and training procedure are presented. These results indicated that the estimate of the precision obtained under the current Landsat analysis procedure may be overstated.

Introduction

In March 1982, the Statistical Reporting Service (SRS) signed a Memorandum of Understanding with the following organizations:

- The California Department of Food and Agriculture, California Crop and Livestock Reporting Service (CCLRS)
- The California Department of Water Resources (DWR)
- The National Aeronautics and Space Administration, Ames Research Center
- The University of California-Berkeley, Remote Sensing Research Program (UCB)

The purpose of this agreement was to determine how agricultural remote sensing data can be used in State and Federal programs in California. The first SRS task was to determine if SRS's Landsat procedures for winter wheat, corn and soybeans used in mid-western states could be extended to the multiple crop environment in California. This paper discusses that task.

SRS uses digital data from Landsat to improve crop-area statistics based on ground gathered survey data. This is accomplished by using Landsat digital data as an auxiliary variable in a regression estimator [1]. Recent results from this procedure applied to major crops in the midwest are discussed in papers by Winings, et al. [2] and Mergerson, et al. [3]. Briefly, the SRS Landsat procedure for major crops in mid-western states consists of the following steps:

- ground truth, collected during an operational survey, and corresponding Landsat data are used to develop a discriminant function which in turn is used to classify Landsat pixels as representing a specific ground cover,
- sampled areas are classified and regression relationships developed between classified results and ground truth,
- all of the pixels in the area of interest are classified, and
- crop-area estimates are calculated by applying the regression relationship to the all-pixel classification results.

Through consultation with the cooperators, the author decided to test the ability of the above major crop procedure in meeting the following California information needs:

1) State level area estimates for:

   a. Major land-use types
   b. Areas of individual crops
   c. Irrigated versus non-irrigated crop acreage

2) County level area acreage estimates for individual crops

3) Resource map products to aid in locating user defined land covers (e.g., orchards, vineyards, etc.)

Study Area

Because of the abundance and diversity of agriculture, the 16-county area outlined in Figure 1 was proposed as the primary area of interest. In reviewing the availability of Landsat data, however, it was decided to restrict the study to the areas with sufficient Landsat coverage and minimal cloud cover problems (Figure 2). Analysis district and county level crop acreage estimates were computed for two areas (AD48/32 and AD48/33). (An analysis district is an area of land covered by Landsat imagery from the same overpass date). To reduce the volume of data processed, experimental products such as mapping products, and estimates by land-use and irrigation practice were produced for area AD48/33 only. Area AD48/33 was chosen to coincide with UCB's work.

Since little was known about the properties of the estimates obtained from the major-crop procedure when used for minor cover types, an independent test and training procedure was studied in area AD46. Area AD46 was chosen because of the abundance of ground data in this two-scene analysis district.

Data Requirements

SRS's Landsat regression procedure utilizes the following sources of data:

1) A sample of land areas (Ground Data Set) with:

   a. field boundaries in digital form referenced to a map base, and

   b. field information such as ground cover, field acreage, irrigation practice, etc., plus

2) Landsat digital data also registered to the same map base.

The multi-crop situation in California required that both the ground and Landsat data be available for at least two dates. This is needed in order to make estimates for crops whose growing seasons do not overlap (e.g., winter wheat and corn).

Figure 1

Original 16 County Study Area

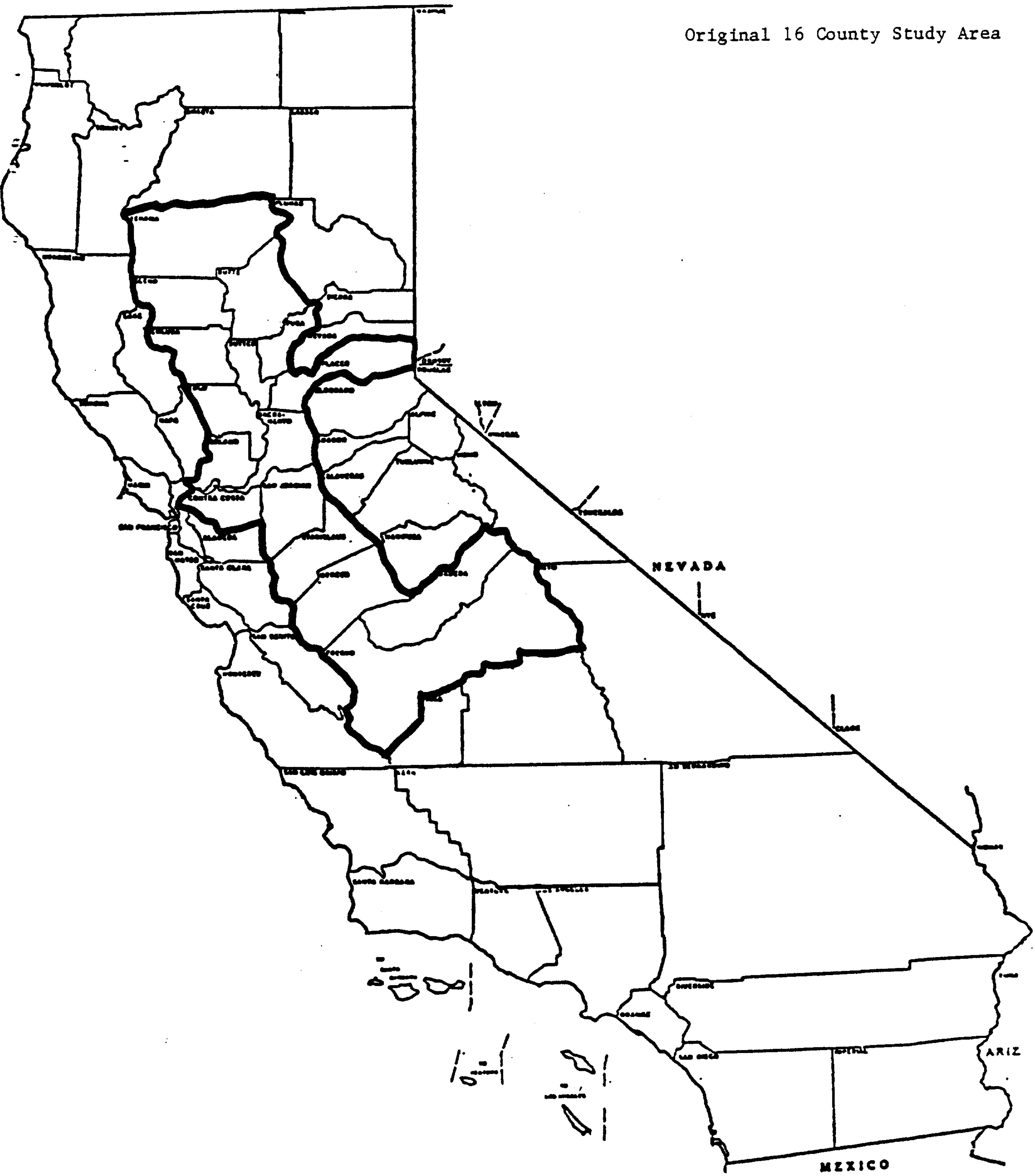
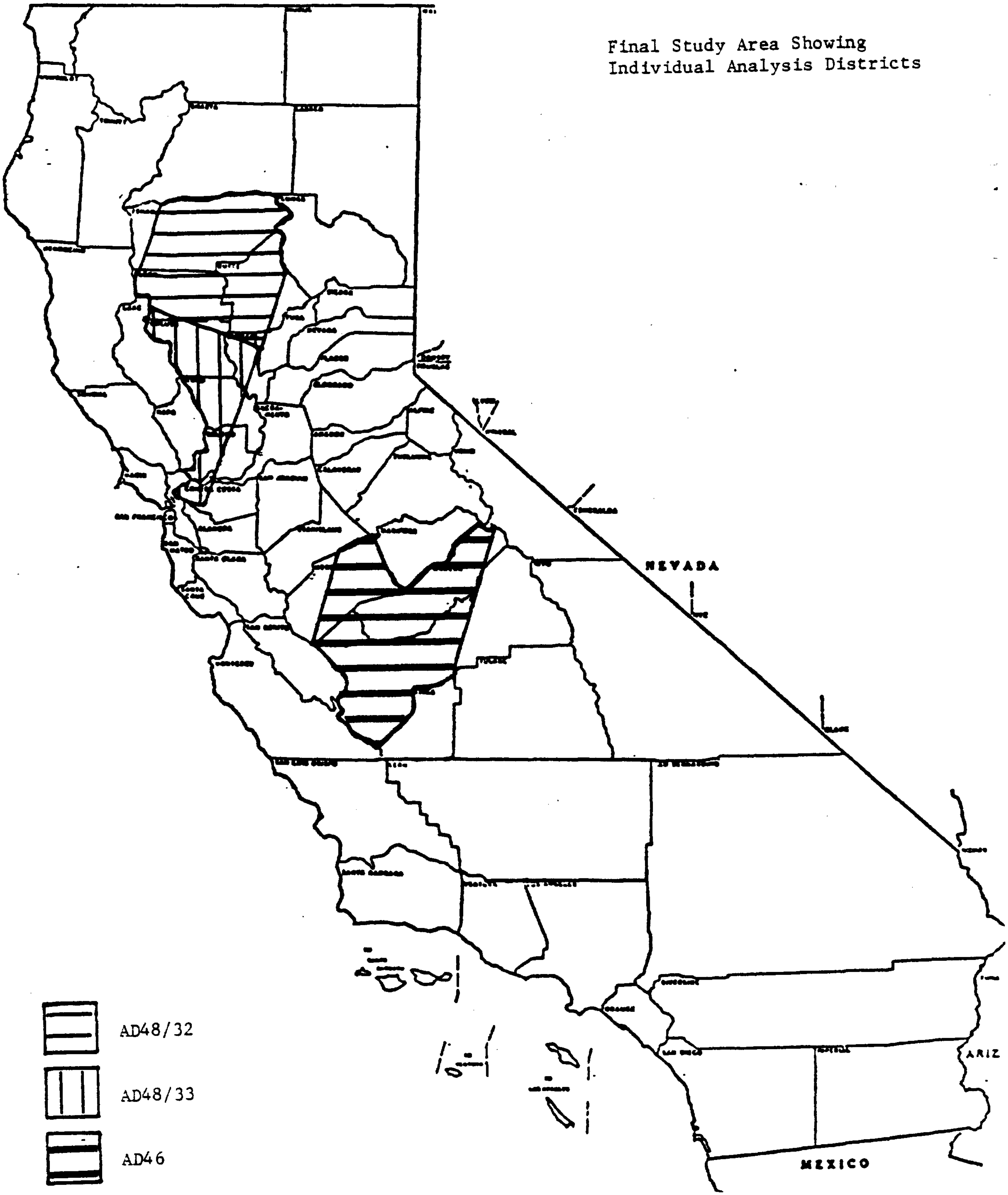


Figure 2



Final Study Area Showing Individual Analysis Districts

Ground Data Set

During late May and early June each year, SRS conducts a nation-wide survey called the June Enumerative Survey (JES). The JES uses an area-frame sampling technique [4], to sample areas of land called segments through stratified sampling. The strata are based on percent cultivation and, in the case of California, predominant crop type [5]. Table 1 lists the stratum definitions for the California area frame. Table 2 shows the number of segments in the population and the sample size by study area. During the JES each segment is visited by an enumerator who records the field boundaries on an aerial photograph. The field acreage, irrigation practice, and cover type are recorded for each field in the segment.

Table 1: California Area Frame Strata Definitions

Stratum 13 — Fifty percent or more cultivated, mostly general crops with less than 10 percent fruit or vegetables.

Stratum 17 — Fifty percent or more cultivated, mostly fruit, tree nuts, or grapes mixed with general crops.

Stratum 19 — Fifty percent or more cultivated, mostly vegetables mixed with general crops.

Stratum 20 — Fifteen to fifty percent cultivated, extensive cropland and hay.

Stratum 31 — Agri-urban, more than 20 dwellings per square mile, residential mixed with agriculture.

Stratum 32 — City, more than 20 dwellings per square mile, heavily residential/commercial, virtually no agriculture.

Stratum 41 — Privately owned range, less than 15 percent cultivated.

Stratum 43 — Desert range, barren areas with less than 15 percent cultivated, virtually no crops or livestock.

Stratum 44 — Public grazing lands, Bureau of Land Management or Forest Service grazing allotments.

Stratum 45 — Public land not in grazing.

Stratum 50 — Nonagricultural, includes State and National Forest, wildlife refuges, military reservations, and similarly designated land.

Stratum 62 — Known water (not sampled), larger than one square mile.

Table 2: California Area Frame

Population (N) and Sample (n) Sizes

Stratum	State N	State n	AD 48/32 N	AD 48/32 n	AD 48/33 N	AD 48/33 n	AD 46 N	AD 46 n
13	6984	240	503	22	975	30	882	30
17	10364	240	781	20	638	17	2350	68
19	3623	100	0	0	308	7	898	23
20	7790	120	377	3	292	0	572	5
31	14779	40	115	0	506	3	531	1
32	23156	10	57	0	648	1	652	1
41	10458	100	508	5	308	1	377	3
43	3994	20	0	0	0	0	0	0
44	1/	25	1/	0	1/	0	1/	0
45	4008	8	17	0	15	0	0	0
50	7090	8	3	0	44	0	77	0
62 2/	1706	0	17	0	113	0	8	0

1/ Sample selection was done by probability proportionate to size.
2/ Units are square miles since no sample was drawn in strata 62.

Data collected from the JES were used to create the ground data set as in the Landsat major-crops procedure. The segment field boundaries were digitized by DWR. Since the JES collects data for all crops for the entire crop year, it was possible to create a ground data set with two observations (visits) for each field within the segments. Visit one corresponded to the ground cover that would appear first during the crop year. Visit two corresponded to the cover that would appear second, if different from visit one. In addition, information on crop acreage and irrigation practice were recorded separately for each visit.

Some fields were recorded as containing crops that the farmer intended to plant when the enumerators visited the JES segments around the first of June. To insure the accuracy of the data, these fields were revisited in September and any discrepancies with the data recorded in June corrected.

Handling the JES data in this manner produced "double-visit" ground data for all crops except vegetables. All vegetables except for tomatoes and sweet corn were called "other vegetables" because it was possible for more than two vegetable crops to be planted in one field in one year.

Landsat Data

Two dates of Landsat data were purchased for each analysis district because of the temporally different phenologies of the crops to be estimated. Table 3 shows the scenes used for each analysis district and the data quality and cloud cover determinations. The earlier date was registered to a map base using a third order polynomial [1]. A multitemporal data set was then created by overlaying the Landsat data from the second date onto the first [6]. The multitemporal data set contained eight channels of Landsat data for each pixel. The first four channels are the reflectance values from the first date, channels five through eight are from the second date.

Table 3:

Landsat Scenes Used for Analysis

Area	Scene ID	Date	Data Quality 1/	Cloud Cover
AD 48/32	831517-18134	4/30/82	8888	10%
	831625-18143	8/16/82	8888	10%
AD 48/33	831535-18142	5/18/82	8888	0%
	831625-18145	8/16/82	8888	10%
AD 46	831515-18030	4/28/82	8888	10%
	831523-18035	8/14/82	8888	10%
	831515-18033	4/28/82	8888	0%
	831523-18041	8/14/82	8888	10%

1/ The data quality number is assigned by the EROS Data Center. Each digit represents the quality of bands 1 through 4 respectively and 8 is the highest quality number assigned.

Data Analysis

After the Landsat data were registered to a map base and the ground data digitized, the segment field boundaries were located in the Landsat digital data. This resulted in a set of pixels labeled by cover type. When a field was doubled cropped (e.g., winter wheat followed by corn) the double cropping was considered as a separate cover type. Also, if a crop appeared in both irrigated and non-irrigated fields, each crop/irrigation practice was considered a separate cover type. This was possible because the SRS processing software permits the selection of fields by combinations of cover type, irrigation practice and date. The pixels for each cover were then clustered using the Classy clustering algorithm [7]. This produced several spectral signatures (categories) for each cover. Each spectral signature consisted of the eight-dimensional mean vector and the covariance matrix of the reflectance values for each category. The statistics for all categories and cover types were then reviewed and combined to form the discriminant functions for a Gaussian maximum likelihood classifier [8]. Table 4 summarizes the training data and number of categories used in each analysis district classifier.

Table 4: Classifier Training Data by Area

AD48/32

Cover Type	Number Pixels	Number Categories
Winter Wheat		
Irrigated	347	2
Non-Irrigated	709	3
and Drybeans	417	5
Rice	2775	9
Cropland Pasture	805	4
Tree Fruits	423	2
Grapes	917	4
Permanent Pasture		
Irrigated	200	2
Non-Irrigated	3506	7
Barley	288	2
Almonds	88	1
Dry Beans	727	2
Sorghum	445	2
Alfalfa	613	1
Other Hay	184	2
Corn	775	2
Sugarbeets	367	4
Sunflowers	104	1
Oats	115	1
Other Crops	624	6
Non-Ag Land	4134	8
TOTAL		70

AD48/33

Cover Type	Number Pixels	Number Categories
Winter Wheat		
Irrigated	891	6
Non-Irrigated	1148	6
and Drybeans	215	1
and Sorghum	69	1
Processing Tomatoes	789	3
Corn	725	3
Rice	2157	7
Sunflowers	122	1
Dry Beans	275	1
Safflower	706	3
Sugarbeets	972	4
Alfalfa	90	3
Other Hay		
Irrigated	158	1
Non-Irrigated	238	2
Tree Fruit	1017	6
Almonds	363	3
Grapes	161	2
Irrigated	98	1
Non-Irrigated	63	1
Permanent Pasture		
Irrigated	203	1
Non-Irrigated	1654	4
Cropland Pasture	80	1
Sorghum	395	2
Sweet Corn	171	1
Other Vegetables	90	1
Other Crops	233	1
Non-Ag Land	2164	9
TOTAL		73

To reduce processing cost, the classification was done in two stages, small-scale and full-frame. In small-scale processing each pixel associated with a segment was classified to a category. The number of pixels classified to each category were summed to segment totals. For each cover or grouping of covers the category totals were summed to segment cover type totals. These cover type totals were used as the independent (auxiliary) variable(s). Correspondingly, the reported acreages were summed to segment totals and used as the dependent variable. The segment totals were used to calculate least squares estimates for the parameters of the selected regression estimator. A separate regression estimator for reported acreage was developed for each cover or cover grouping, for each strata.

In full-frame processing every pixel in the Landsat scene was classified with the classifier selected from small-scale processing. The classified results were then tabulated by category and stratum. For each cover or cover grouping used in small-scale processing, the category totals were summed to stratum cover type totals. From these tabulations the auxiliary variable's stratum population averages per segment were calculated. Using the population averages a regression estimate for the acreage of each cover or cover grouping was made for each stratum that contained at least seven sampled segments. The stratum estimates were then summed to an analysis district estimate.

Crop Acreage Estimation

The single-variable regression estimator was used for estimating individual crop acreages. The theory behind this estimator is discussed by Hanuschak, et al. [1].

Single Variable Regression Estimator

$$\hat{Y}_c = \sum_{s \in S} N_s \bar{Y}_{sc}(\text{reg}) = \text{regression acreage estimate for cover c}$$

where:

$$\bar{Y}_{sc}\,(\text{reg}) = \hat{b}_{osc} + \hat{b}_{1sc}\,(\bar{X}_{sc})$$

$$\bar{X}_{sc} = \frac{X_{sc}}{N_s}$$

S = the set of all strata to be estimated for in this analysis district.

$\hat{b}_{osc}, \hat{b}_{1sc}$ = the least squares estimates of the regression parameters for crop c stratum s when regressing the JES reported acres on classified results.

X_{sc} = the number of pixels classified to any category making up crop c for all the area in stratum s from full-frame.

N_s = the total number of sample units (segments) in stratum s.

Table 5 shows the coefficient of determination (r-squared) and the estimated regression parameters by stratum for the covers in AD48/32 and AD48/33. The wide variation in the r-squared values is explained in part by the differing amounts of classifier training data available (Table 4).

Table 5: Single Variable Regression Parameter Estimates by Analysis District, Strata and Crop

AD48/32	Strata 13			Strata 17		
cover	r^2	b_o	b_1	r^2	b_o	b_1
Winter Wheat (All)	.733	- 8.2	.79	.952	-8.2	.82
Corn	.688	-14.2	1.04		1/	
Drybeans	.845	-.1	.84	.942	-7.1	.79
Rice	.987	-9.5	.82	.891	-2.9	.66
Sorghum	.693	-10.9	.71	.242	-.4	.32
Sugarbeets	.928	-5.1	.83	.989	-1.2	.90
Grapes	.577	-3.1	.42	.851	-7.6	.76
All Hay 2/	.878	-3.5	.97	.777	-4.1	.93
Permanent Pasture	.961	-18.9	1.12	.722	-12.4	.87

AD48/33	Strata 13			Strata 17			Strata 19		
cover	r^2	b_o	b_1	r^2	b_o	b_1	r^2	b_o	b_1
Winter Wheat (IRR)	.948	-10.9	1.54	.037	21.3	1.69	.898	16.5	.77
Winter Wheat (Not IRR)	.389	7.0	3.11	.962	-.8	.91	.018	113.2	.52
Winter Wheat (All)	.917	-15.9	1.24	.964	-12.4	.87	.963	-1.1	.81
Corn	.579	-35.0	1.43	1/			.789	-64.3	2.10
Drybeans	.005	6.8	-.41	.024	1.7	-.33	.789	16.5	.79
Rice	.975	2.9	.84	.823	- 2.8	1.42	.908	-19.7	1.05
Safflower	.107	1.5	.45	1/			.974	-11.0	1.30
Sorghum	.688	-14.2	.96	1/			.928	-3.8	.77
Sugarbeets	.850	-10.0	.84	1/			.644	-29.0	1.20
Tomatoes Processing	.826	-10.4	1.23	1/			.710	-20.7	1.00
Almonds	.822	-3.1	.85	.909	- 5.9	.79	1/		
Grapes	.698	-2.5	.69	.797	- 1.2	1.85	1/		
Other Hay	.915	- 7.3	.78	.389	- 6.8	.86	1/	-.2	.04
Permanent Pasture	.538	- 1.2	.54	.844	-20.7	1.04	1/	-.8	.11

1/ No acreage reported for this crop in the indicated stratum.
2/ Includes Alfalfa and Other Hay.

Tables 6 and 7 show the direct expansion estimates (which use only the JES data) [1] with the regression estimate for areas AD48/32 and AD48/33. A measure of the improved precision of the regression estimate is the ratio of the variance of the direct expansion estimate to the variance of the regression estimate. This statistic is called the relative efficiency (RE). Equivalently, it is the factor by which the sample would have to be increased to produce a direct expansion estimate with the same precision as the regression estimate shown. The tabled relative efficiencies indicate that supplementing the JES with Landsat produces a considerably more precise estimate. Because the JES segments used for estimation were the same ones used to train the classifier, the author feels that the tabled estimates of the precision of the regression estimates are optimistically biased. The independent training and estimating exercise discussed later in this report attempts to examine this bias.

TABLE 6: ACREAGE ESTIMATES BY COVER (AD48/32)1/

COVER	DIRECT EXPANSION ESTIMATE (ACRES)	DIRECT EXPANSION STANDARD DEVIATION (ACRES)	DIRECT EXPANSION CV (%)	REGRESSION ESTIMATE (ACRES)	REGRESSION STANDARD DEVIATION (ACRES)	REGRESSION CV (%)	R.E.
Winter Wheat4/	90,921	52,660	57.9	65,282	6,173	9.5	72.8
Corn 2/	18,119	8,640	47.7	10,160	4,946	48.7	3.1
Drybeans	35,509	12,170	34.3	21,708	3,915	18.0	9.7
Rice	137,616	27,457	20.0	142,890	3,973	2.8	47.8
Sorghum	14,935	5,964	39.9	15,253	3,222	21.1	3.4
Sugarbeets	10,763	5,424	50.4	5,233	1,254	24.0	18.7
Grapes	42,865	15,335	35.8	28,925	6,269	21.7	6.0
All Hay	43,300	13,285	30.7	51,513	5,146	10.0	6.7
Permanent Pasture3/	674,461	273,206	40.5	932,055	207,684	22.3	1.7
Small Grains	50,678	16,838	33.2	39,659	5,920	14.9	8.1
Row Crops	84,169	19,646	23.3	42,867	7,050	16.4	7.6
Permanent Planting	71,900	17,137	23.8	74,356	9,049	12.2	3.6
Pasture	696,554	275,598	39.6	1,026,251	130,889	12.8	4.4

1/ STRATA 13,17.
2/ STRATA 13 ONLY.
3/ STRATA 13,17,20 & 41.
4/ STRATA 13,17, & 20.

TABLE 7: ACREAGE ESTIMATES BY COVER (AD48/33)[1/]

COVER	DIRECT EXPANSION ESTIMATE (ACRES)	DIRECT EXPANSION STANDARD DEVIATION (ACRES)	DIRECT EXPANSION CV (%)	REGRESSION ESTIMATE (ACRES)	REGRESSION STANDARD DEVIATION (ACRES)	REGRESSION CV (%)	R.E.
Winter Wheat	106,041	25,031	23.6	96,909	6,371	6.6	15.4
Corn	62,598	22,642	36.2	22,020	12,326	56.0	3.4
Drybeans	21,939	9,477	43.2	12,280	5,816	47.4	2.7
Rice	274,560	44,590	16.2	281,741	8,224	2.9	29.4
Safflower	38,100	20,703	54.3	26,885	6,600	24.5	9.8
Sorghum	17,106	6,736	39.4	8,121	3,059	37.7	4.8
Sugarbeets	44,603	13,778	30.9	45,315	6,296	13.9	4.8
Tomatoes Processing	50,713	17,753	35.0	42,798	8,292	19.4	4.6
Almonds	24,408	8,376	34.3	26,715	2,963	11.1	8.0
Grapes	18,204	9,188	50.5	13,464	4,336	32.2	4.5
Other Hay	26,217	14,113	53.8	10,406	5,821	55.9	5.9
Permanent Pasture	36,656	14,822	40.4	20,169	7,176	35.6	4.3

1/ Strata 13, 17 & 19.

Major Land-Use Acreage Estimation

Table 8 describes the four studied land-use types and lists a brief description of each.

TABLE 8: LAND-USE TYPES

	Land-Use Type	Description, (Included Covers)
1)	Permanent-Plantings	all orchards and vineyards (includes fruits and nuts)
2)	Pasture	all land used or to be used for grazing (includes cropland and permanent pasture)
3)	Field-Truck	all cropland not included in 1 and 2
4)	Non-Ag	all other land
5)	Total	all land

Several types of regression estimators were evaluated for their ability to estimate the acreages for these major land uses. First, the single variable regression estimator described previously was calculated for each land-use using the following quantities:

$\hat{b}_{osc}$, $\hat{b}_{1sc}$ = the least squares estimates of the regression parameters for land-use c strata s when regressing the JES reported acres on the classified results.

X_{sc} = the number of pixels classified to any category making up land-use c in this analysis district.

The estimates and relative efficiencies for this use of the single variable regression estimator are listed in Table 9.

TABLE 9: ACREAGE ESTIMATES

BY LAND-USE TYPE
(AD48/33*)

		Field-Truck	**Pasture**	**Permanent Planting**	**Non-Ag**	**Total**
	Estimate	678,732	46,643	115,440	135,376	976,192
Direct Expansion	**st.dev.**	25,743	15,842	23,236	18,289	18,219
	cv %	3.8	34.0	20.1	13.5	13.5
Single Variable	**Estimate**	644,579	29,383	97,985	114,335	954,809
Regression	**st. dev.**	13,578	7,949	10,088	14,441	12,386
	cv %	2.1	27.1	10.3	12.6	1.3
	R.E.	3.6	4.0	5.3	1.6	2.2
Multiple Variable	**Estimate**	850,098	33,727	113,209	151,662	1,148,698
Regression	**st. dev.**	12,379	8,831	11,363	16,607	11,997
	cv %	1.5	26.2	10.0	10.9	1.0
	R.E.	4.3	3.2	4.2	1.2	2.3

* Strata 13, 17, & 19

Note in Table 9 that for single variable regression the sum of the estimates for the land types does not equal the estimate for total land. To correct this situation multiple regression estimates of the following form were calculated:

Multiple Regression Estimate

$$\hat{Y}_c = \sum_{s \varepsilon S} N_{sc} \bar{Y}_{sc}(\text{mreg}) = \text{regression acreage estimate for the cover grouping making up land use c.}$$

where:

$$\bar{Y}_{sc}(\text{mreg}) = \hat{b}_{osc} + \sum_{i=1}^{4} \hat{b}_{isc} (\bar{X}_{si})$$

$$\bar{X}_{si} = \frac{X_{si}}{N_s}$$

N_s, S = same as in single variable regression

X_{si} = the number of pixels classified to any category making up land-use i for strata s from full-frame.

$\hat{b}_{osc}, \hat{b}_{isc}$ = the least squares estimates of the multiple regression parameters.

i = 1, 2, 3, 4 (Table 8)

As in single variable regression the segment totals from small-scale processing were used to make least squares estimates for $\hat{b}_{osc}$ and $\hat{b}_{isc}$ for all i. These parameter estimates along with the population totals from full-frame processing were used to estimate the reported acreage per segment by cover. Ultimately the estimates of the segment averages were expanded to strata level totals and then summed over strata. Table 9 also lists the multiple regression estimates. The sum of the different land-use types now agrees with the estimate of total land. The multiple regression estimator can, however, have a significant bias if the X_{si}'s are highly correlated [18]. In this application the X_{si}'s were highly correlated (the correlations ranged from .3 to .7). Because of the possible bias in the tabled multiple regression estimates, the author feels that some other approach may be necessary to make land-use estimates summable to total land. Nevertheless, the relative efficiencies of both the single variable and multiple regression approaches, although possibly optimistic, indicate the potential for improved acreage estimates resulting from supplementing the JES with digital Landsat data.

Estimation By Irrigation Practices

A major goal of the project's cooperators was to estimate crop types by irrigation practice. The primary crop of interest in this regard was winter wheat. To demonstrate this capability the following single and multiple regression estimates were calculated.

Single Variable Regression Estimate

$\hat{Y}_I = \sum_{s \in S} N_s \quad \bar{Y}_{sI}(mreg)$ = regression acreage estimate for winter wheat under irrigation practice I

where:

$$\bar{Y}_{sI}(mreg) = \hat{b}_{osI} + \hat{b}_{1sI}\ (\bar{X}_{sI})$$

$$\bar{X}_{sI} = \frac{X_{sI}}{N_s}$$

S = the set of all strata to be estimated for.

N_s = the total number of sample units in strata s.

X_{sI} = the number of pixels classified to the categories making up winter wheat with irrigation practice I in strata s.

I = irrigation practice (irrigated or not irrigated).

$\hat{b}_{osI}$, $\hat{b}_{1sI}$ = the least squares estimates of the multiple regression parameters.

Multiple Regression Estimate

$$\hat{Y}_I = \sum_{s \epsilon S} N_s \bar{Y}_{sI}(mreg)$$ = regression acreage estimate for winter wheat under irrigation practice I.

where:

$$\bar{Y}_{sI}\ (mreg) = \hat{b}_{osI} + \hat{b}_{1sI}\ (\bar{X}_{s1}) + \hat{b}_{2sI}\ (\bar{X}_{s2})$$

$$\bar{X}_{s1} = \frac{X_{s1}}{N_s}$$

$$\bar{X}_{s2} = \frac{X_{s2}}{N_s}$$

$\hat{b}_{osI}, \hat{b}_{1sI}, \hat{b}_{2sI}$ = the least squares estimates for the regression parameters.

X_{s1} = X_{sI} (from single variable regression) when I is irrigated

X_{s2} = X_{sI} (from single variable regression) when I is not irrigated.

N_s,S = Same as single variable regression.

The estimation procedures followed were identical to those used for land-use estimation except that different dependent and independent variables were used. The direct expansions, single and multiple regression estimates and their associated relative efficiencies are listed in Table 10. As with land use estimates, the parts, using single variable regression estimation do not sum to the estimated total but the multiple regression estimates do. Again, with multiple regression, the correlation between the two independent variables was high (ranging from .2 to .6), indicating a possible bias. The tabled relative efficiencies were again substantial, though possibly overstated.

TABLE 10: ACREAGE ESTIMATES
IRRIGATED VS NON-IRRIGATED WINTER WHEAT

Single Variable	Estimate (acres)	Standard Division (acres)	CV% (%)	R.E.
Irrigated	59,248	8,024	13.5	7.7
Not Irrigated	36,291	5,758	15.9	5.0
Total	96,909	6,371	6.6	15.4
Multiple Variable (2)				
Irrigated	43,965	7,047	16.0	10.0
Not Irrigated	39,990	5,494	13.7	5.5
Total	83,955	5,244	6.3	22.8
Direct Expansion				
Irrigated	74,426	22,253	29.9	---
Not Irrigated	31,614	12,910	40.8	---
Total	106,041	25,031	23.6	---

County Level Estimation

In California a large proportion of the crop statistics requests are for substate areas. Consequently, in this study Landsat-based county estimates based on the Battese-Fuller model were examined for their ability to meet some of these requests. The theory and performance of the Battese-Fuller model is discussed in detail by Walker and Sigman [9]. The form of this model is as follows:

$$Y_{kc} = b_{oc} + b_{1c} \; (X_{kc}) + V_{kc} \; + \; E_{kc}$$

where:

Y_{kc} = acreage of crop c in county k

X_{kc} = number of pixels classified to crop c in county k

V_{kc} = the county effect on the regression for crop c in county k

E_{kc} = random error

b_{0c}, b_{1c} = the analysis district single variable regression parameters

Using the data from small-scale processing, the fitting of constants procedure recommended by Fuller and Battese [10] was used to calculate the best linear unbiased estimates of b_{0c} and b_{1c} for each stratum and crop. County data obtained from full-frame processing were used to estimate the acreage per segment by stratum. These averages were then expanded and summed to the county level in order to estimate total county acreages.

In trying to apply this procedure to the California data set, three types of problems occured:

1) Part of a county was located in AD48/32 and part in AD48/33,

2) Part of a county was in either AD48/32 or AD48/33 but the remaining part was in neither scene, and

3) Some strata in either AD48/32 or AD48/33 had an insufficient sample size to estimate the regression parameters.

To counter these problems any area outside the scenes or any stratum for which there was an insufficient number of segments to calculate a regression estimate was eliminated from the population. This meant that some of the Landsat-based county estimates were incomplete and not comparable to CCLRS non-Landsat estimates. Secondly, if a county was located in two scenes the county was split into two distinct parts and each part was treated as a separate area for estimation. The two sub-county estimates and their associated variances were summed to produce the total county estimate. Table 11 shows the proportions of the agricultural strata contained in the two scenes. For Colusa, Glenn and Tehama counties all the land contained in the major agricultural strata was covered by the estimates. Table 12 gives the county estimates for rice, winter wheat and sugarbeets. These three crops were chosen because they span the scale from the most major crop (rice) to a relatively minor crop (sugarbeets).

TABLE 11: PERCENT COVERAGE

By County

	STRATUM 1/				
County	**13**	**17**	**19**	**20**	**Combined 2/**
Butte	88	78	—	86	85
Colusa	100	100	—	100	100
Glenn	100	100	—	100	100
Solano	42	100	100	59	59
Sutter	44	73	99	100	71
Tehama	100	100	100	100	100
Yolo	58	100	57	100	66

1/ See Table 1 for strata definitions
2/ percent coverage relative to total area in the county

The Battese-Fuller estimator produced estimates with root mean squared errors less than 10% for rice and winter wheat for some of the major counties. As pointed out in the subsequent "Independent Training and Estimation Study," the tabled root mean squared errors may be understated. In addition, minor crops, such as sugarbeets, still present a problem with regard to county estimates. This is primarily due to an insufficient JES sample size resulting in the following:

1) the minor-crop direct expansion estimates are highly variable, and

2) too few pixels are available to develop a classifier capable of distinguishing the minor crops, thus limiting the potential gain from the regression estimator.

The problems found in the county estimates procedure point out the need to develop some technique to handle the following situations:

1) incompleteness due to insufficient Landsat coverage, or

2) insufficient sample sizes resulting in loss of strata and poor performance for minor-crops.

Table 12: County Estimates
Battese-Fuller

County	n	Estimate (Acres)	Root-MSE (Acres)	Relative Root-MSE (%)	CCLR Estimate (Acres)
RICE					
Butte	15	111,556	2,327	2.1	106,000
Colusa	23	180,385	4,495	2.5	120,000
Glenn	19	102,253	2,801	2.7	81,000
Solano	6	787	2,483	315.5	---
Sutter	9	88,888	7,056	7.9	91,000
Tehama	8	2,383	1,449	60.8	2,000
Yolo	13	36,180	3,398	9.4	36,500
WINTER WHEAT					
Butte	15	5,317	2,842	53.5	28,500
Colusa	23	35,426	3,412	9.6	39,000
Glenn	19	31,386	3,794	12.1	41,000
Solano	6	601	3,395	564.9	42,500
Sutter	9	13,079	263	2.0	48,500
Tehama	8	10,851	1,583	14.6	13,000
Yolo	13	49,793	2,403	4.8	87,000
SUGARBEETS					
Butte	15	1,492	735	49.3	2,800
Colusa	23	16,329	3,125	19.1	7,200
Glenn	19	5,409	1,030	19.0	6,900
Solano	6	13,649	1,650	12.1	25,000
Sutter	9	7,101	2,679	37.7	5,700
Tehama	8	1,494	191	12.8	680
Yolo	13	16,950	2,076	12.2	21,100

Resource Map Development

Two cooperators in this project (CCLRS and DWR) expressed interest in the development of resource mapping products -- that is, a printed product which identifies resources of interest such as crop fields, orchards, etc., which could be used in conjunction with existing maps to locate the resources of interest on the earth's surface. The use of Landsat in developing these types of mapping products has been examined by a number of researchers. Most of these studies have examined ways to print Landsat classification in a meaningful manner. For this type of product each pixel in the area of interest is first classified to one of a number of cover types of interest and then the classification results are printed at various scales on different types of media. In contrast, the procedure recommended by Sigman [11], called crop-odds mapping, does not classify each pixel into a specific cover type. Instead for each pixel the procedure calculates the pixel's posterior probability of being any of the possible cover types. The posterior probability [$P(c/X_i)$] is the probability that pixel i is from cover c given that the Landsat reflectance value is the observed vector X_i. When the category signatures are multivariate normal, it follows that:

$$P(c/X_i) = \sum_{k \in L} P(k/X_i)$$

where:

L = the set of all categories making up cover c

$$P(k/X_i) = \frac{P_k \, |\Sigma_k|^{-q} \ \mathrm{EXP}\left[- (X_i - \mu_k)' \, \Sigma_k^{-1} (X_i - \mu_k) \right]}{\sum_{k \in K} P_k \, |\Sigma_k|^{-q} \ \mathrm{EXP}\left[- (X_i - \mu_k)' \, \Sigma_k^{-1} (X_i - \mu_k) \right]}$$

P_k = prior probability of category k
q = the number of channels in the Landsat data
Σ_k = the reflectance covariance matrix for category k
μ_k = the mean reflectance vector of category k
X_i = vector of reflectance values for pixel i
K = the set of all categories possible

The advantage of using the crop-odds procedure for mapping purposes is that the user can specify a reliability level for mapping each cover type. For example if the goal was to identify areas of land that have even a remote chance of being an orchard the user could specify a threshold of 0.1 and create a map that prints all pixels with $P(orchard/X_i)$ greater than 0.1. However if the user's goal was to identify only orchards, and he was not concerned with missing some orchards, he could set the threshold at say 0.9 and print all pixels that have $P(orchard/X_i)$ greater than 0.9. Figure 3 shows a portion of a crop-odds map produced for the Tisdale Weir quadrangle in AD48/33. Figure 4 shows the corresponding portion of the U.S. Geological Survey's 7.5 minute quadrangle map. The cover types chosen were arbitrary and can vary under the procedure (for example, by changing the set L to non-agricultural land, irrigated agricultural land then non-irrigated agricultural land a crop-odds map for irrigated agricultural land could have been produced). This demonstration map points out some interesting results in this area. First, a large proportion of the pixels within this area had a maximum posterior probability greater than .75. Secondly, most of the pixels with a maximum $P(c/X_i)$ less than .75 were located on field boundaries and in some highly confused areas. In addition, it seems that the classifier used was able to identify distinct field patterns over most of the area.

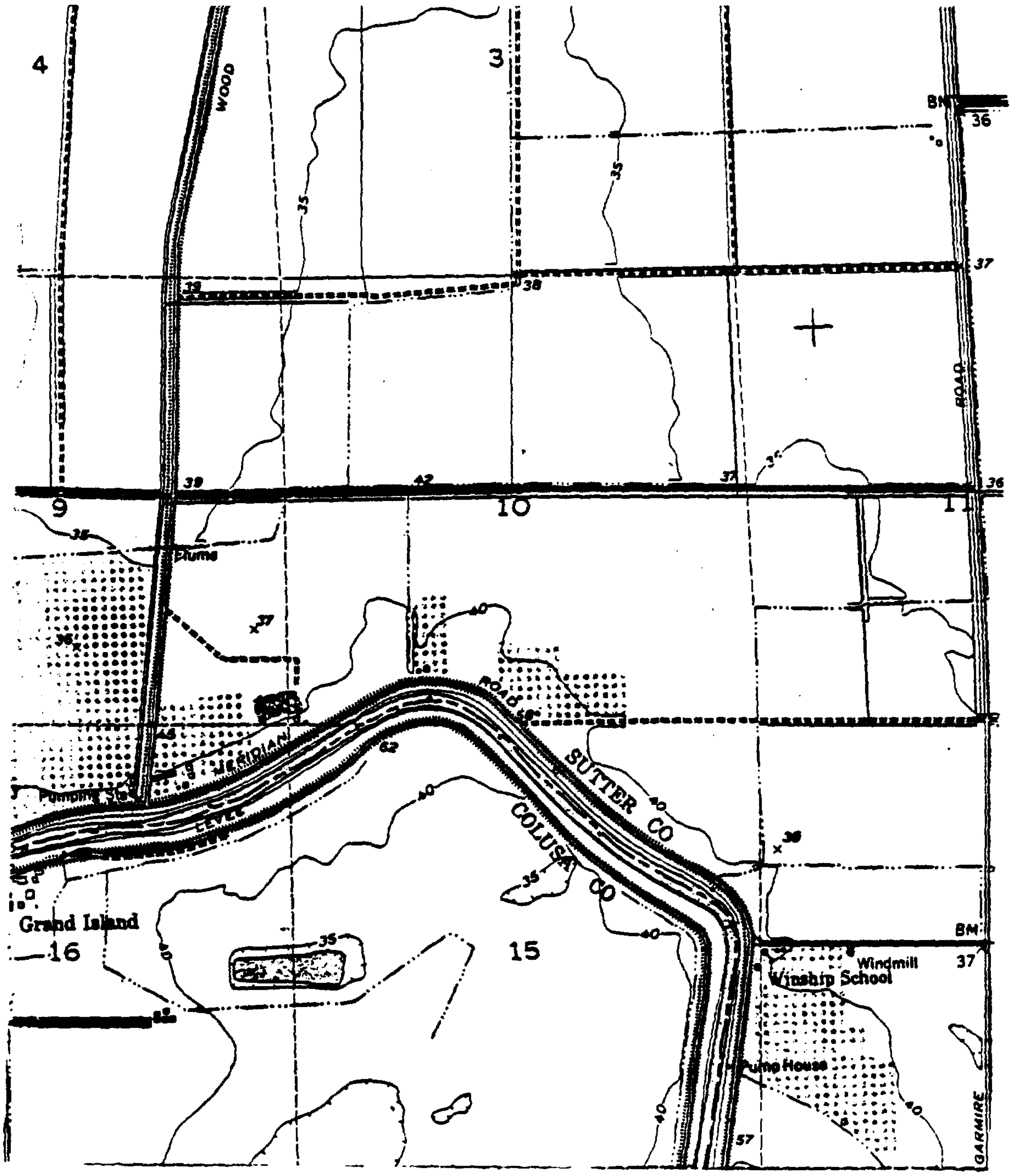
Despite the encouraging results with crop-odds mapping one major disadvantage exists. The present procedure used is prohibitively expensive. However, a new table lookup procedure, equivalent to crop-odds, has been suggested by Sigman. This procedure may make crop-odds mapping reasonable in the future.

Figure 3

Portion of Crop - Odds Map with $P(c/X_i) \geq .75$.

· · · · · · · ·	Small Grains
::::::::::::::::	Other Crops
(symbol)	Rice
(symbol)	Hay
(symbol)	Permanent Plantings
(symbol)	Pasture
(symbol)	Non-Agricultural Land

Figure 4



Portion of U.S.G.S 7.5 Minute Tisdale Weir Quadrangle

Independent Training and Estimation Study

SRS's current small-scale processing of Landsat data for the purpose of developing regression estimates involves two activities. First, the JES segment information is used, through a modified supervised clustering approach, to develop a Gaussian maximum likelihood classifier. Secondly, the same segments are used to calculate sample-based terms of the estimator and to estimate the precision of the resulting estimate. The goal of this entire process is to reduce the variance of crop acreage estimates relative to the original JES direct expansion estimates. The large-sample variance of the single variable regression estimator is given by Cochran [15].

$$V(\bar{Y}(reg)) = \frac{n-1}{n-2} \quad (1-r^2) \quad V(\bar{Y})$$

where

$V(\bar{Y})$ = the variance of the direct expansion estimate
r = the correlation coefficient of the fitted regression line
n = the sample size

From the formula it is evident that the improvement in the precision is dependent on the level of the correlation between the segment reported acreage and the classification results. Under the current approach, however, the calculated r-squared values is one that measures how well the classifier does in classifying the data that was used to create it — that is, training and estimation are dependent. Because of this, one would expect that such a correlation may be optimistically biased if used as an estimate of the population correlation (Sukhatme [16]). Papers by Gleason, et al. [12] and Amis, et al. [13] support this expectation. This part of the study was done to examine in a brief manner if there is a detectable bias of the correlation coefficient relative to the crop mixture found in this area of California.

To get a better understanding of the procedure, the 121 segments in the agricultural strata in area AD46 were divided into two non-overlapping sets. This selection was done in a manner that first equalized the number of segments by strata in each set, as much as possible, and secondly attempted to equalize the crop distribution between the two sets. To meet these objectives the crops found in AD46 were prioritized (Table 13). Then within each stratum the segments containing the first priority crop were listed in descending order of acreage of that crop (i.e., the first segment listed was the segment which had the most winter wheat). Of the remaining segments the ones having the second priority crop (barley) were listed in descending order of that crop's acreage. This procedure was continued until all the segments within each stratum were listed. A one-half systematic sample was drawn from this list separating the segments into two non-overlapping sets (set A and set B).

TABLE 13: Classifier Training Data by Set for AD46

		SET A		SET B	
Priority	Crop	Number Pixels	Number Categories	Number Pixels	Number Categories
1)	Winter Wheat	1897	5	1736	6
2)	Barley	1163	5	1350	6
3)	Alfalfa	1701	5	1671	8
4)	Tomatoes	729	4	803	4
5)	Cotton	7334	10	5220	11
6)	Walnuts	2669	8	3116	6
7)	Grapes	944	4	617	3
8)	Corn	329	2	155	2
9)	Sorghum	377	2	130	2
10)	Sugarbeets	151	1	541	3
11)	Rice	137	1	994	4
12)	Drybeans	81	2	571	4
13)	Pasture	5227	8	1743	7

The current SRS procedure was used to develop a classifier using the segments in set A. The same procedure was used to develop a separate classifier using Set B. Table 13 also shows the number of training pixels and number of categories for each set. To compare the correlation coefficients obtained from a dependent procedure with an independent procedure, classifier A was used to classify the segments in set A and also set B. Classifier B was also used to classify the segments in set A and set B. Classified results for each of the four classifications were used to calculate the sample-based terms of the single variable regression estimator and to estimate the four correlation coefficients. These four sets of results are referred to by the following short hand:

A-setA = the procedure in which the classifier was developed using the segments in set A and the estimation was done using set A (dependent)

A-setB = the procedure in which the classifier was developed using the segments in set A but the estimation was done using set B (independent)

B-setA = the procedure in which the classifier was developed using the segments in set B but the estimation was done using set A (independent)

B-setB = the procedure in which the classifier was developed using the segments in set B and the estimation was done using set B (dependent)

In addition to the independent and dependent sets, the data for the segments used for estimation in A-setB and B-setA were combined to estimate a new regression line called the jackknife estimate [14].

Tables 14, 15 and 16 give the r-squared values for the five regressions by strata. In all cases, except winter wheat in stratum 13 (classifier A) and alfalfa and grapes in stratum 13 (classifier B), the independent training and estimation procedure had a lower correlation coefficient than the dependent procedure. Appendix A reports the results of the Hotelling's T test for the equality of the correlation coefficients. The test suggests that for nearly all cases where the correlation from the independent procedure is less, it is significantly less. However, a review of the tables indicates that the differences in the correlation coefficients are less for major crops such as cotton and walnuts than they are for the minor crops such as sugarbeets. This may indicate that if there is a bias in the estimate of the correlation coefficient it may be influenced by the amount of training data used to develop the classifier. These results agree with those seen by Amis, et al. [13] and Gleason, et al. [12], although the differences shown here are more extreme. The extreme results may be due to using only half of the JES segments in each set. Although there are at least 60 segments in each set, the JES sample allocation of 121 segments is a good indication that the crop area is highly variable and it may be that 60 segments is not enough to develop a classifier in this area.

Table 14: r-Squared Values
Stratum 13

CROP	A-SET A	A-SET B	B-SET B	B-SET A	JACKKNIFE
Winter wheat	.88	.91	.91	.23	.36
Barley	.96	.01	.86	.01	.00
Corn	.00	.01	.69	.03	.01
Drybean	.07	.00	.45	.11	.01
Sugarbeets	.82	.33	.99	.15	.13
Tomatoes	.98	.00	.69	.08	.01
Cotton	.82	.75	.97	.89	.76
Grapes	.82	.20	.29	.84	.42
Alfalfa	.92	.40	.37	.67	.50
Pasture	.68	.54	.92	.59	.35
Small Grains	.96	.50	.94	.91	.65

Table 15: r-Squared Values
Stratum 17

CROP	A-SET A	A-SET B	B-SET B	B-SET A	JACKKNIFE
Winter wheat	.90	.19	.92	.32	.22
Barley	.08	.02	.87	.04	.00
Corn	.16	.01	.88	.01	.01
Drybeans	.56	.01	.87	.01	.01
Cotton	.83	.65	.94	.65	.64
Walnuts	.73	.60	.86	.49	.52
Grapes	.74	.55	.75	.59	.57
Alfalfa	.72	.56	.88	.44	.53
Pasture	.41	.29	.96	.65	.53
Small grains	.29	.11	.82	.16	.12

Table 16: r-Squared Values
Stratum 19

CROP	A-SET A	A-SET B	B-SET B	B-SET A	JACKKNIFE
Winter Wheat	.94	.17	.91	.00	.02
Barley	.83	.02	.65	.00	.00
Sugarbeets	.01	.59	.86	.03	.24
Tomatoes	.99	.95	.98	.10	.47
Alfalfa	.88	.54	.95	.26	.36
Pasture	.01	.01	.01	.01	.01
Small grains	.91	.82	.96	.06	.31

Table 17 shows the effect of the r-squared values on the relative efficiencies of the regression estimators. Even though the bias for the major crops is smaller, the impact on the relative efficiencies is still considerable.

Table 17: Relative Efficiencies by Procedure

CROP	A-SET A	A-SET B	B-SET B	B-SET A	JACKKNIFE
Winter Wheat	3.90	1.67	10.15	1.06	1.16
Barley	5.24	.94	4.15	.93	.96
Tomatoes	271.50	5.42	15.91	1.01	1.63
Cotton	6.83	4.03	18.52	2.84	3.10
Walnuts	3.59	1.55	7.14	1.90	2.05
Grapes	3.71	2.13	3.82	2.35	2.28
Alfalfa	7.71	1.95	4.24	2.17	1.89
Pasture	1.75	1.34	14.21	2.66	1.75

Table 18 compares the estimates obtained from the five regression procedures. The jackknife estimate was obtained by evaluating the jackknife-estimated regression line at the average of the A-setB and B-setA auxiliary-variable population means. Review of these estimates and the associated standard errors indicate that there is little bias in the acreage estimates even though the estimated variances are biased. It must be pointed out, however, that this conclusion may be dependent on the method used to select the two non-overlapping sets.

Since only one half of the JES segments were used, these results may not be directly applicable to other areas. However this study does point out the sensitivity of the correlation coefficients to the training data and how this can result in misleading statements about relative efficiencies. It is apparent that a detailed study is needed to determine if a similar situation exists for areas where the full JES sample is used both for training and estimation.

Conclusions

This study has pointed out some of the weaknessess in the present SRS approach to Landsat analysis when applied to a multiple crop environment. Based on this study the author makes the following conclusions:

1) Given Landsat coverage, scene and ultimately state level estimates can be produced in California for a large number of crop and cover types, but further study is needed to determine the extent of the bias in the estimates of the estimator's precision.

2) The procedure for county level estimates is promising but may be limited by the strata-level sample sizes and the location of the Landsat scenes.

3) The crop-odds mapping procedure is a useful alternative to Landsat classification mapping if the cost can be reduced.

Table 18: Acreage Estimates by Procedure for Individual Crops

CROP	A-SET A Estimate (Acres)	A-SET A Standard Error (Acres)	A-SET B Estimate (Acres)	A-SET B Standard Error (Acres)
Winter Wheat	154,908	19,445	175,036	29,084
Barley	82,112	13,497	85,212	30,419
Tomatoes	38,377	1,611	56,303	12,504
Cotton	556,530	21,638	502,272	36,044
Walnuts	351,418	24,758	428,708	29,715
Grapes	108,742	15,644	116,777	14,931
Alfalfa	170,605	14,478	179,404	25,523
Pasture	176,819	29,813	170,139	35,746

CROP	B-SET B Estimate (Acres)	B-SET B Standard Error (Acres)	B-SET A Estimate (Acres)	B-SET A Standard Error (Acres)
Winter Wheat	182,652	11,796	146,761	37,301
Barley	64,313	14,451	82,900	31,994
Tomatoes	44,503	7,307	56,831	26,412
Cotton	530,583	16,809	510,499	33,549
Walnuts	282,446	17,204	245,649	34,090
Grapes	75,091	11,140	80,792	19,667
Alfalfa	149,125	17,293	135,328	27,305
Pasture	171,355	12,698	139,866	24,175

CROP	JACKKNIFE* Estimate (Acres)	JACKKNIFE* Standard Error (Acres)	Direct Expansion* Estimate (Acres)	Direct Expansion* Standard Error (Acres)
Winter Wheat	148,516	24,406	133,289	26,243
Barley	87,616	21,242	89,017	20,860
Tomatoes	59,192	14,956	47,510	19,112
Cotton	504,984	25,510	509,012	44,945
Walnuts	333,360	22,722	235,175	32,506
Grapes	105,620	12,180	57,153	18,407
Alfafa	160,614	19,391	36,590	26,238
Pasture	155,491	23,314	116,268	30,866

* All 121 Segments.

Recommendations

Based on the knowledge gained from this study the author recommends the following:

1) An experiment be conducted to determine if the bias in the estimate of the estimator's precision is significant when the full JES sample is used for training and estimation.

2) The suggested procedure equivalent to crop-odds mapping be investigated.

3) Given a satisfactory result in (1) above, the procedures outlined in this report should be consolidated into an efficient Landsat processing technique designed to meet a portion of the information needs in California.

4) The processing technique developed in (3) should be used jointly by CCLRS and DWR in a cooperative manner so that both agencies can benefit from a technology that may be cost prohibitive to either separately.

References

[1] Hanuschak, George A., Michael Craig, Martin Ozga, Raymond Luebbe, Paul Cook, David Kleweno, and Charles Miller. "Obtaining Timely Crop Area Estimates Using Ground-Gathered and LANDSAT Data." U.S. Dept. of Agr. ESCS Technical Bulletin No. 1609, August 1979.

[2] Winings, Sherman B., Paul W. Cook, George A. Hanuschak, "AgRISTARS DCLC Applications Project 1982 Corn and Soybeans Area Estimates for Iowa, Illinois." Seventh international Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, May 1983.

[3] Mergerson, J. W., V. B. Johnson, R. A. Kestle. "AgRISTARS DCLC Applications Project: 1982 Winter Wheat Area Estimates For Colorado, Kansas and Oklahoma." Ninth International Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Laboratory for Applications of Remote Sensing, June 1983.

[4] "Scope and Methods of the Statistical Reporting Service." U.S. Dept. of Agriculture, SRS, Miscellaneous Publication Number 1308, September 1983.

[5] Fecso, Ron, Van Johnson. "The New California Area Frame." U.S. Dept. of Agriculture, SRS, SRS-22, September 1981.

[6] Ozga, Martin, Sue R. Faerman, Richard S. Sigman. "Editor Multitemporal System." U.S. Dept. of Agriculture, ESCS, November 1979.

[7] Lennington, R. K., and M. E. Rassback. "Classy - An Adaptive Maximum Likelihood Clustering Algorithm." Proceeding of the LACIE Symposium, NASA Manned Spacecraft Center, Houston, Texas, Jan. 1972.

[8] Swain, Philip H., and Shirley M. Davis. "Remote Sensing The Quantitative Approach." McGraw-Hill, Inc., New York, 1978.

[9] Sigman, Richard, Gail Walker. "The Use of Landsat for County Estimates of Crop Acres: Evaluation of the Huddleston-Ray and Battese-Fuller Estimators." U.S. Dept. of Agr., SRS, September 1982.

[10] Battese, G.E., W.A. Fuller. "Prediction of County Crop Areas Using Survey and Satellite Data." Survey Section Proceedings, 1981 American Statistical Association Annual Meeting, Detroit, Michigan.

[11] Sigman, Richard. Internal Memorandum. "A Proposal for a Probabilistic ESCS Data Product for Use in AgRISTARS Land Use Research." USDA, ESCS, Washington, D.C., 1979.

[12] Gleason, Chapman, Robert R. Starbuck, Richard S. Sigman, George A. Hanuschak, Michael E. Craig, Paul W. Cook, and Richard D. Allen. "The Auxiliary Use of Landsat Data in Estimating Crop Acreages: Results of the 1975 Illinois Crop Acreage Experiment." U. S. Dept. of Agr., SRS, SRS-21, October 1977.

[13] Amis, M.L., R.K. Lennington, M.V. Martin, W. G. McGuire, and S. S. Shen. "Evaluation of Large Area Crop Estimation Techniques Using Landsat and Ground-Derived Data." AgRISTARS Domestic Crop and Land Cover Report, Lyndon B. Johnson Space Center, Houston, Texas, March 1981.

[14] Miller, Rupert G. "The Jackknife - A Review." Biometrika, 61, 1, p.1 (1974).

[15] Cochran, W. G. "Sampling Techniques, Third edition." John Wiley and Sons, Inc., New York, 1977.

[16] Sukhatme, B. V. Discussion Survey Section Proceedings, 1978 American Statistical Annual Association Meeting (pp 175), San Diego, California.

[17] Hotelling, Harold. "The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters." Annals of Math. Stat., Vol. 11 (1940), pp 271-283.

[18] Konijn, H.S. "Statistical Theory of Sample Survey Design and Analysis." American Elsevier Publishing Company, Inc., New York, 1973.

Appendix A: Hotelling's T Test [17]
(conditional on X_{1cs} and X_{2cs})

Y_{cs} = reported acreage for crop c strata s

X_{1cs} = number of pixels classified as crop c using the dependent classifier (A-set A or B-set B) in strata s

X_{2cs} = number of pixels classified as crop c using the independent classifer (B-set A or A-set B) in strata s

r_{ocs} = correlation coefficient of X_{1cs} with X_{2cs}

r_{1cs} = correlation coefficient of Y_{cs} with X_{1cs}

r_{2cs} = correlation coefficient of Y_{cs} with X_{2cs}

n_s = number of segments in strata s for set A or set B.

N_s = n_s = 3

Assumptions:

$$Y_{cs} = B_{00} + B_{01} X_{1cs} + B_{02} X_{2cs} + E_{0cs}$$

$$Y_{cs} = B_{10} + B_{10} X_{1cs} + E_{1cs}$$

$$Y_{cs} = B_{21} + B_{20} X_{2cs} + E_{2cs}$$

where:

E_{0cs}, E_{1cs}, E_{2cs} are distributed $N(0, \sigma_0)$, $N(0, \sigma_1,)$ and $N(0, \sigma_2)$ respectively

Hotellings T test (one sided):

$$H_0: r_{1cs} = r_{2cs} \qquad H_A: r_{1cs} > r_{2cs}$$

$$t = (r_{1cs} - r_{2cs}) \sqrt{\frac{N_s(1+r_{ocs})}{2D}}$$

$$D = \det \begin{vmatrix} 1 & r_{2cs} & r_{0cs} \\ r_{2cs} & 1 & r_{1cs} \\ r_{ocs} & r_{1cs} & 1 \end{vmatrix}$$

The p-value for the one-sided test against the alternative H_A is:

$$P_{cs} = P\ (T_N > t)$$

where:

T_N = T statistic with N degrees of freedom

Table A.1

Values of P_{cs} for the Dependent Independent Study

Stratum (s)

Crop (c)	13		17		19	
	Set A	Set B	Set A	Set B	Set A	Set B
Cotton	.87	.06	.01	.01	—	—
Alfalfa	.00	.51	.00	.00	.00	.00
Walnuts	—	—	.02	.00	—	—
Winter Wheat	.00	.46	.95	.00	.00	.00
Barley	.00	.00	.38	.00	.00	.00

where: Set A is the comparison of A-Set A with B-Set A, and
Set B is the comparison of B-Set B with A-Set B.

Note: Reject H_o in favor of H_A if P_{cs} is less than the desired significance level.