E85-10054

NASA CONTRACTOR REPORT 177327

(E85-10054 NASA-CR-177327) MULTI CROP AREA ESTIMATION IN IDAHO USING EDITOR (Technicolor Government Services, Inc.) 37 p HC A03/MF A01 CSCL 06C

N85-15251

Unclas G3/43 00054

Crop Area Estimation in Idaho Using EDITOR

E.J. Sheffner

CONTRACT NAS2-11101
December 1984

NASA

NASA CONTRACTOR REPORT 177327

Crop Area Estimation in Idaho Using EDITOR

E.J. Sheffner
Technicolor Government Services
Moffett Field, California

Prepared for
Ames Research Center
under Contract NAS2-11101

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

Table of Contents

List of Tables

List of Figures

Introduction

This report desribes work performed at NASA Ames Research Center on a multi- crop area estimation project in the central Snake River Plain of Idaho in partial fulfillment of contract NAS2-11101.

In 1982, H.A. Anderson, Director of the Idaho Image Analysis Facility (IIAF) of the Idaho Department of Water Resources (IDWR), approached the National Aeronautics and Space Administration (NASA) with a request for assistance. Two of IDWR's responsibilities are to monitor the extent of irrigated land and to estimate, on an annual basis, the demand for water for irrigation from ground and surace water sources. In order to meet the Department's information needs, IIAF has been using remotely sensed data, especially Landsat digital data and image products. The first large scale project incorporating Landsat multispectral scanner (MSS) data undertaken by IIAF was an irrigated lands survey of the Snake River Plain - the premier agricultural region in Idaho and one of the major areas of irrigated agriculture in the United States. The irrigated lands survey was completed in 1982. By the end of that year, IIAF was seeking ways to increase the amount of information it could obtain through digital data processing and requested that NASA assist it in the research effort required.

Anderson approached NASA with a research plan consisting of five tasks including:

- develop a procedure to detect change in irrigated lands, i.e. land going into and out of irrigation,
- develop a procedure for multi-crop labelling and area estimation in southern Idaho,
- determine the separability in spectral response of irrigated and non-irrigated grain and alfalfa,
- determine the separability in spectral response of irrigated land and natural wetland, and
- determine the relationships among biomass, spectral response and water-consumptive use and the extent to which those relationships can be used to improve the accuracy of water-consumptive use models.

The first task addresses the need for IDWR to locate sources of water application. The latter four tasks are directed at improving the accuracy of the input data in the water-consumptive use model utilized by IIAF to estimate demand for irrigation water.

NASA responded to IIAF's request for assistance by providing research support in FY-84 on the change detection and multi-crop area estimation tasks. The research was conducted at the Ames Research Center (ARC) by the Technology Application Branch with funding from the Western Regional Test and Evaluation (WRETE) program. The following report describes the work undertaken at

ARC between January 1983 and June 1984 on the multi-crop area estimation task.

Although the area estimation task was initiated by IIAF, the character of the work was influenced by the US Department of Agriculture (USDA). USDA influence came from two sources. Concurrent with the Idaho WRETE project, the Remote Sensing Branch (RSB)-Statistical Research Division, Statistical Reporting Service of the USDA was working with ARC, the California Department of Water Resources and the Remote Sensing Research Program (RSRP) of the University of California on the California Cooperative Remote Sensing Project - a program to improve multi-crop area estimation and mapping in California. Because of the existing cooperation between ARC and RSB, the Idaho-WRETE project had access to the USDA sampling frame for Idaho and assistance with the operation of the EDITOR software system. The second source of USDA influence came from the Boise, Idaho office of the Statistical Reporting Service (SRS). SRS supplied the ground data for the multi-crop task as part of a working agreement between SRS and IDWR to determine if the same multi-crop area estimation procedure could meet the needs of both agencies and, therefore, save each agency the expense of generating the data independently.

Approach

The approach to the multi-crop area estimation task evolved as decisions were made on the nature of key elements in the data analysis procedure. These elements included stratification, study area, cover types, data processing system and area estimation procedure.

Stratification

The IIAF plan recommended evaluation of two candidate stratification and sampling schemes as the first phase in the development of a research approach to the multi-crop task. The recommendation was made because stratification prior to sampling can improve the efficiency of a crop area estimation procedure by reducing the number of samples required and the variance of the estimates. The two stratifications evaluated were the IIAF stratification and the USDA sampling frame for Idaho.

The IIAF stratification is based on land use only or location and field size. It was used in the 1980 irrigated lands survey and IIAF was hopeful that the it would be suitable for multi-crop area estimation. Twenty-four strata were defined in the Snake River Plain. Strata definitions included large fields - lower Snake River Plain, small fields- Boise region, rangeland, wetlands, etc. Strata boundaries were drawn by inspection on 1:250000 scale Landsat color composite images. Sample units were 6.25 sq.km. (2.5 sq.mi.).

The USDA sample frame is a stratification based on agricultural land use. Sample units, called segments, are generally one square mile in size, but the sample unit size may vary by stratum. Strata boundaries extend beyond the area of interest for estimation, and strata definitions are not state-specific.

Evaluation of the USDA and IIAF stratifications was performed under the direction of Dr. R.W. Thomas (RSRP) using a tool for experimental design developed at the University of California, Berkeley, called the survey planning model (SPM) [11,12]. SPM determines the number of sample units required and the cost to generate an area estimate given the requirements for estimate accuracy, Landsat-to-ground correlations for the crop types of interest, sample unit size, a digital map of the strata boundaries, a digital map of the class boundaries, and estimated cost per sample unit to collect ground data.

A complete data set for analysis with SPM was available for the IIAF stratification only, because the USDA sampling frame for Idaho was revised in 1982 and strata boundaries for the revised frame were not available at the time of the SPM evaluation. Consequently, sample size and number of sample units were evaluated using the IIAF stratification and class map. The output from SPM indicated that estimate costs would not vary significantly if either the USDA or IIAF sample unit size were used [10]. Nevertheless, the number of sample units required to generate crop acreage estimates within acceptable limits of error was too large to be economical using the IIAF stratification. Although the USDA stratification was not fully evaluated with SPM, the stratification and sample size were selected for the multi-crop task.

Study Area

The Snake River Plain is an elongated U-shaped region in southern Idaho (Figure 1). The region increases in elevation from the west, about 2000' at the Oregon border, to the east, approximately 5200' at Ashton. The plain is confined within high mountains to the north and south and varies in width between four and sixty miles. Sagebrush and grass are the predominant natural vegetation. The climate is arid to semi-arid, and the soils have developed on highly permeable extrusive igneous rock. Irrigation, from ground water and surface sources, is extensive.

Agriculture in the plain varies by field size and crop type. IIAF selected three study sites that captured the variation in the region (Figure 1). In the Boise area, site I, field size is small, cover types are numerous and "ranchettes," houses on 5 to 40 acres of idle land or land used for pasture or crops, are common. Site II, near Twin Falls, is a mixture of large and small fields. It has less variation in cover types than site I and more field crops such as sugarbeets and potatoes. Field size is largest and variation in cover types least in site III, centered on Blackfoot.

The research plan from IIAF suggested developing and testing a multi-crop area estimation procedure using data from the three study sites. For the research performed in 1983, a four county study site was selected that included most of site II. The four county site was selected for the following reasons:

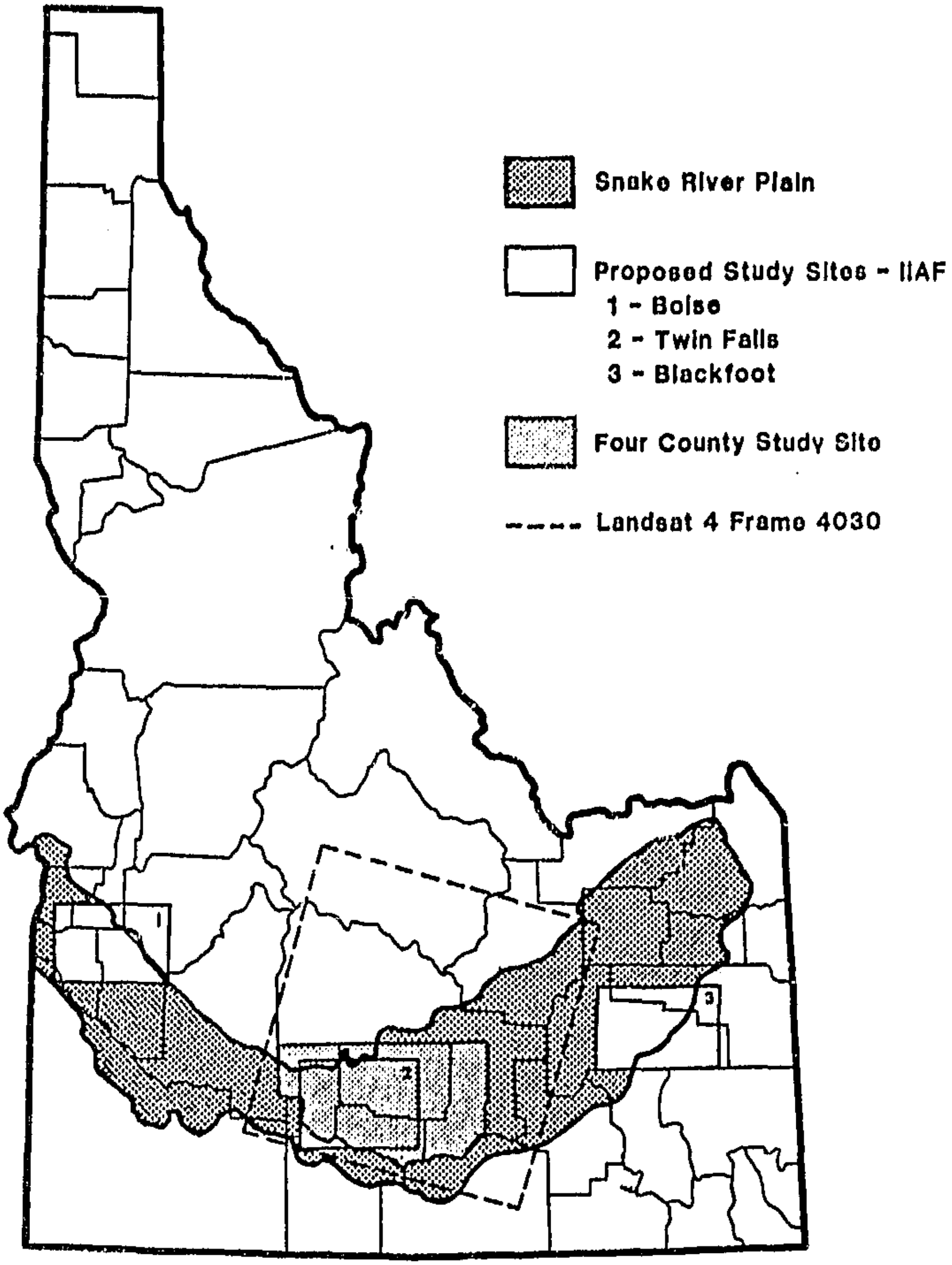
- in the midst of the research year, the Technology Applications Branch at Ames retired its two primary interactive computing systems and brought new systems on-line. Scheduling delays caused by the systems switch made it doubtful that work on three study sites could be completed on time.
- site II contained a good mix of field sizes and a good representation of the crop types of interest to IDWR and USDA.
- SRS collected ground sample data in 1983 from the three sites as part of its annual June Enumerative Survey (JES). SRS was able to collect additional data in site II so that there was sufficient ground data from that site to develop independent training and test sets for digital data processing.

The boundary of the study site was determined by the estimation unit of interest to SRS and the location of the Landsat frame covering the site. Because SRS was interested primarily in county level acreage estimates, the study site was defined as Gooding, Jerome, Lincoln and Minidoka Counties (Figure 1). The four counties fall entirely within Landsat 4 path/row 40/30.

Cover Types

The IIAF research proposal specified five cover types of interest: potatoes, small grains (wheat, barley and oats), alfalfa, row crops (corn, sugarbeets, and beans), and other irrigated crops. Because SRS was also interested in potatoes and sugarbeets, those two crops were designated the crops of primary interest. The multi-crop task was able to generate estimates for potatoes, sugarbeets, alfalfa, small grains, corn and beans.

Figure 1.
Snake River Plain and the Four County Study Site



Data Processing System

The IIAF plan did not specify a software system to use in the area estimation procedure but required the software to be compatible with the IIAF data processing system. EDITOR was the software system chosen for the task.

EDITOR was developed by USDA in the mid-1970's to improve the accuracy of its large scale acreage estimates of major crops in the Midwest. The system generates a regression estimate using ground sample units (segments), and corresponding Landsat MSS data from a maximum of two acquisitions. Segments are selected through a stratified random sampling procedure, and a Landsat pixel classifier is compiled and tested using data from all the segments. Descriptions of the development and operation of EDITOR are given in Ozga, et. al.[8] and Hanuschak, et. al. [3,4]. During the time that work on the multi-crop task was performed, EDITOR resided on a PDP-10 at Bolt, Berenek, and Newman (BBN) in Cambridge, Mass. and was accessible to ARC through the ARPANET.

EDITOR was selected for data processing and area estimation in the multi-crop task because it met the project specifications and was the best system available for the task at ARC in FY-84. EDITOR software is currently being re-written in a "portable" format, and it is intended that most, if not all, EDITOR functions will be transferrable to the IIAF system. Due to the system changes at ARC described above, EDITOR was the only reliable software system available that could generate the estimates for the four county study site and meet the schedule given in the research plan.

Use of EDITOR in Idaho was encouraged by RSB and SRS in order to evaluate the performance of standard EDITOR processing and the use of EDITOR for acreage estimation at the county level. The standard EDITOR procedure uses data from all sampled segments to build the classifier and test the results. Recent work on the California Cooperative Remote Sensing Project suggested that a significant bias is introduced by training and testing with the same data [5]. The experimental design for the multi-crop task in Idaho was modified to include an EDITOR-based procedure that used independent training and testing.

Approach Summary

Through the efforts of personnel from IIAF, USDA, ARC and RSRP, a plan was developed for multi-crop area estimation that contained the following elements:

- Stratification and sampling were accomplished using the USDA sampling frame and sample segments.
- The study site was in the central Snake River Plain - Gooding, Jerome, Lincoln and Minidoka Counties.

• Two dates of landsat MSS data from the 1983 growing season were required.

• Two procedres for data processing would be tested. One procedure would follow standard EDITOR processing for a multi-date data set. The other procedure would use EDITOR software but would employ independent training and test data sets.

• County level and study site acreage estimates would be generated for potatoes, sugarbeets and other crops of interest with sufficient representation to make estimation feasible.

Data Processing

Ground Data Preparation

The EDITOR system requires both ground data and Landsat MSS data. The ground data consisted of digital maps of the segments and stratification, and digital data files containing descriptions of the contents of each field within each each segment, a segment catalog file identifying the stratum from which each segment was drawn and a frame unit file listing the number of sample segments in each stratum by county.

Stratification

Six strata were defined by USDA in the area frame revised in 1982. Strata definitions are given in Table 1. Virtually all agriculture in the four county test site was located in strata 15 and 25. Table 2. lists the percent of land in each county occupied by the each stratum. Strata boundary information was recieved at ARC on maps, scale 1 inchm = 2 miles. The boundaries were digitized using EDITOR software and a standard USDA digitizing tablet linked to BBN via the ARPANET.

Table 1.
USDA Strata Definitions for Idaho Area Frame

Stratum 15 - General crops; 50% or more cultivated along the Snake River. All irrigated, intensively cultivated land in Canyon, Ada, Owyhee, Elmore, Gooding, Twin Falls, Lincoln, Jerome, Minidoka, Cassia, Power, Bannock, Caribou, Bingham, Bonneville, Teton, Madison, Jefferson, Fremont, Clark, and Butte Counties. This stratum should contain virtually all potatoes and sugarbeets.

Stratum 25 - 15-49% cultivated used in conjunction with stratum 15.

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

Stratum 32 - Residential/Commercial; more than 20 dwellings per square mile. No agriculture present.

Stratum 42 - Rangeland; less than 15% cultivated.

Stratum 62 - Water bodies larger than one square mile.

Table 2.
Number of Pixels and
Percent Cover by Stratum and County

County	Stratum	Pixels	% Cover
Gooding	15	171924	29.2
	25	69199	11.8
	31/32	2160	.4
	42	345192	58.6
Jerome	15	212519	44.2
	25	127406	26.5
	31/32	4998	1.0
	42	135221	28.1
	62	601	.1
Lincoln	15	106224	13.8
	25	78762	10.3
	31/32	1945	.2
	42	580289	75.6
Minidoka	15	257674	62.4
	25	52428	12.7
	31/32	9145	2.2
	42	93765	22.7

Segment Data

The ground data compiled for the multi-crop task differed from a standard EDITOR data set in two ways that affected the results of the task - the sampling rate at the county level was higher, and the type of information collected from each segment varied.

All sample unit (segment) data for the task was collected by, and supplied to ARC through, the SRS office in Boise. SRS collects segment data every June as part of the June Enumerative Survey (JES). The JES was not designed for county level estimation, and the number of segments sampled in the study site in 1983 was only twenty. In order to improve the chances of generating reasonable regression estimates at the county level, SRS agreed to collect data from twelve additional segments in the study site. The additional segments were JES segments from previous surveys; however, the amount of information collected on the additional segments was not as extensive as that collected on the JES segments from 1983 and included field boundaries and crop labels only.

Table 3. lists the 32 segments from which data was collected for the multi-crop task with their stratum and county location. The segments numbered 2000 and 3000 were JES segments in 1983. The segments numbered 6000 and 9000 were JES segments in previous years, rotated out in 1983, and surveyed for use in the multi-crop task only.

Table 3.
Segments in the Four County Study Site

Segment number	County	Stratum
2078	Gooding	15
2081	Minidoka	15
2082	Jerome	15
2090	Gooding	15
2093	Minidoka	15
2094	Lincoln	15
2105	Jerome	15
2106	Jerome	15
2141	Minidoka	15
2142	Lincoln	15
2153	Jerome	15
2165	Minidoka	15
2174	Gooding	15
2177	Minidoka	15
2253	Lincoln	25
2254	Lincoln	25
2263	Minidoka	25
3129	Minidoka	15
3130	Jerome	15
3189	Jerome	15
6102	Minidoka	15
6213	Lincoln	15
6216	Gooding	15
6217	Gooding	15
6218	Gooding	15
6245	Lincoln	25
9088	Jerome	25
9091	Jerome	15
9093	Jerome	15
9098	Minidoka	15
9101	Minidoka	15
9102	Minidoka	15

Total by Stratum:
Stratum 15 = 27
Stratum 25 = 5

Total by County:
Gooding = 6
Jerome = 9
Lincoln = 6
Minidoka = 11

Field boundary data was inscribed on large scale aerial photography and the boundaries transferred by IIAF to 1:24000 scale maps. Copies of the maps were sent to ARC for digitizing. Segment data was digitized directly from the maps using EDITOR software to create segment network files.

Cover type information and reported field size were sent to ARC separately. Ground truth files were created manually using the ground truth file creating and editing routine in EDITOR. Reported field size was included in the file when available, i.e. only for the 1983 JES segments. Digitized acreage was substituted for reported acreage for the previous years segments.

A segment catalog file and a frame unit file were compiled manually using EDITOR routines. The frame unit counts were supplied by SRS-RSB. Table 4. lists the frame unit information.

Table 4.
Population (N) and Sample Size (n) - Strata 15 and 25

Stratum	County								Total	
	Gooding		Jerome		Lincoln		Minidoka			
	N	n	N	n	N	n	N	n	N	n
15	218	6	272	8	140	3	331	10	961	27
25	42	0	81	1	54	3	32	1	209	5
Total	260	6	353	9	194	6	363	11	1170	32

Landsat Data Preparation

Date Selection

Landsat MSS data have been shown to be most effective for discrimination of crop types when the data are obtained at or near the period in the crop phenology when infrared response is at its peak [2]. The crops of interest in the multi-crop task reach peak infrared response at two different times in the growing season. Small grains (wheat, barley, and oats) are planted in early spring, reach maximum infrared response in early summer and are harvested by the end of July. Other field crops, such as potatoes and sugarbeets, are planted in early summer and reach maximum infrared response in mid-summer. Consequently, it was desirable to obtain at least two acquisitions of MSS data to perform the multi- crop task. Since EDITOR software can process up to eight channels of data, a maximum of two Landsat acquisitions was sought. The preferred dates were late May and mid-August.

A satisfactory May acquisition for frame 4030 was obtained, but the earliest mid-summer acquisition with a low percentage of cloud cover was September 1. The September date was used although it was considered a marginal date and, subsequently, proved to be too late in the season for good summer crop separability. The Landsat acquisitions used for the multi-crop task were:

Frame	Scene ID	Date
4030	4031617491	28 May 83
	4041217491	1 Sept 83

Registration

Before the Landsat MSS data were extracted for classifier training, the two Landsat acquisitions were registered to a common point-line coordinate system. The scene-to-scene registration was accomplished using a block correlation technique in EDITOR. The May acquisition was selected as the base date. Output from the registration procedure was two eight channel tapes containing the eight registered bands. The data set was split at column 1800 so that the record would not exceed BBN system limitations, and the first 1599 lines removed because they were north of the study site. Registration was confirmed by visual inspection of gray scale maps of blocks of data from bands on both dates. Field patterns were observed to ensure that a proper overlay had been achieved.

Landsat to Ground Calibration

The final step in Landsat data preparation was to generate a Landsat-to- ground calibration file. The file is used to locate the digitized ground segment and stratum data in the Landsat scene. MSS data from the 28 May acquisition was displayed on a CRT using a HP3000 system and IDIMS software. At least one point was located on each 7.5' quad in the four county area and the point/column coordinates recorded. The points were digitized and the corresponding Landsat coordinates entered. EDITOR software was used to generate a second order polynomial relating the ground data to Landsat coordinates with a mean square error of less than .5 pixels.

Selection of Test and Training Sets

The approach to the multi-crop task specified using all segments to train the classifier and test the results and using independent training and test sets of segments to accomplish the same. Because there were only 32 segments in the sample, dividing the sample in half, and using half for training and half for testing, was not considered feasible, and a "modified jackknife" approach was adopted to define the independent training and test set.

The available segments were divided into three groups, and three training and three test sets were created. Each training set contained two thirds of the segments, while the remaining third were reserved for testing. Each test set contained a unique list of segments, but half the segments in each training set were also used in one other training set. Following classification, the segment results from the three test sets were combined to form the independent test set used in the regression estimates.

Sugarbeets and potatoes were the crops of primary interest and served as the starting point for development of the training sets. Three segments contained at least one potato field larger than 100 acres and five segments contained at least one sugarbeet field of similar size. The three segments with the large potato fields were placed one in each training set using a random number table to select the segments. The five segments with large sugarbeet fields (there was no duplication with the first three segments) were placed similarly in the three training sets, but the segment with the largest field of sugarbeets was placed on its own. Nine segments contained at least one sugarbeet field with no field larger than 100 acres, and seven segments contained potato fields with the same size limit (three segments were on both lists). The segments with the small sugarbeet fields were randomly selected and placed into the three sets, three to a set. The segments containing potatoes which had not already been assigned by virtue of their sugarbeet fields were randomly selected and placed in the three sets. The remaining segments, none of which contained either sugarbeets or potatoes, were randomly selected and assigned, in turn, to one of the sets. Segment assignments for the training and test sets are shown in Table 5.

Table 5.
Segment Assignments - Independent Training and Test Sets

Training Set A		Test Set A	Training Set B		Test Set B
2078	2093	6213	2105	6213	2078
2105	2165	2253	2093	2253	2165
2177	2142	2081	2142	2081	2177
9102	9091	2141	9091	2141	9102
9088	3129	9101	3129	9101	9088
2106	6102	2153	6102	2153	2106
2090	6217	2254	6217	2254	2090
9093	2082	3130	2082	3130	9093
2263	3189	6218	3189	6218	2263
2094	6245		6245	2174	2094
6216	2174		9098		6216
9098					

Training Set C		Test Set C	
2078	6213	2105	2174
2165	2253	2093	9098
2177	2081	2142	
9102	2141	9091	
9088	9101	3129	
2106	2153	6102	
2090	2254	6217	
9093	3130	2082	
2263	6218	3189	
2094	6216	6245	

Digital Data Processing

Clustering

Landsat data from all cover types was packed using EDITOR software. In the packing process, the data for a given cover type is gathered from all segments and placed in a disk file. The geometric relationship among the pixels is lost. Packed files were created for each cover type by training set. Two packed files were created for each test set; one file contained the data from all cover types within the segments excluding background pixels (the NB file), and the second file contained the data from all cover types within the segments excluding the background pixels, a one pixel border on the edge of each field, and fields identified as bad fields in the ground data files (the -NB file). Table 6. lists the cover types found in the training sets and the number of pixels packed by cover type and training group.

Table 6.
Number of Pixels Packed for
Training by Cover Type and Training Set*

Cover	Standard Procedure	Training Set A	Training Set B	Training Set C
Alfalfa	1232	694	874	896
Corn	680	503	378	479
Farmstead	19	19	16	3
Sugarbeets	974	695	656	597
Wasteland	140	101	94	85
Wheat**	3437	2176	2505	2035
Pasture	989	831	845	302
Idle	1337	991	613	742
Potatoes	721	462	517	432
Wild hay	15	15	15	5
Other crops	5	0	5	7
Other hay	7	1	6	0
Range	1990	1454	563	1963
Beans	328	244	271	141
Onions	63	63	63	0
Unknown	155	134	63	113
Total	12092	8383	7484	7800

*background and field border pixels not packed
**small grains

Standard EDITOR processing uses the algorithm CLASSY to cluster data, i.e. to group pixels with similar spectral characteristics [7]. USDA experience with CLASSY has shown that the algorithm does not perform well with fewer than 100 pixels; consequently, 100 pixels was used as the minimum number for CLASSY clustering. CLASSY does not require the operator to specify any parameter settings other than the maximum number of iterations to perform on the data. The program was run on the Cray 1S at ARC.

Wasteland, farmstead and onions fell below the minimum pixel count for CLASSY clustering but were felt to be represented sufficiently in the data set to warrant inclusion in the classifier. Clustering was performed on those three cover types at BBN using the EDITOR version of ISOCLAS. Unlike CLASSY, ISOCLAS requires the operator to set three parameters - the number of clusters to create in the data set, the number of iterations to perform, and the minimum cluster distance acceptable before clusters will be combined. The cover types wild hay, other hay and other crops were not clustered because too few pixels remained in the training set after field border pixels were removed. All clustering was performed by cover type.

Statistics from the CLASSY clustering -- means, variances and covariances -- were file transferred to BBN and reformatted into EDITOR format statistics files, one statistics file for each cover type in each training group. The crop specific statistics files from CLASSY and ISOCLAS were combined by training group using the statistics file editing program in EDITOR. Clusters with fewer than 10 pixels were deleted from the file. Four classifiers were compiled in the above manner - one for the standard procedure and one for each of the three training groups in the modified jackknife procedure. The cover types included in the classifiers and the number of clusters in each category are listed in Table 7.

Table 7.
Number of Training Categories
by Cover Type and Training Set

Cover	Standard Procedure	Training Set A	Training Set B	Training Set C
Alfalfa	7	2	3	6
Corn	6	5	4	6
Sugarbeets	5	6	8	6
Wasteland	2	2	2	2
Wheat*	13	11	10	10
Pasture	7	2	9	2
Idle	6	2	6	5
Potatoes	6	6	2	2
Range	8	8	2	9
Beans	6	4	5	3
Onions	3	3	3	3
Farmstead	1	1	1	1
Total	70	53	55	55

*small grains

Classification and Aggregation

Maximum likelihood classifications were performed on the NB and -NB files for all test groups using the appropriate training statistics. In addition, each of the three sets of training statistics was used to generate a maximum likelihood classification of the entire scene.

Percent correct classified was calculated from the classifications of the -NB files. The results from the standard procedure and the independent training sets are given in Table 8. The test with the standard procedure performed better than the modified jackknife (as expected) because, in the standard set, all segments were used for training and testing. The percent correct and commission error give an impression of the amount of confusion among cover types. The variation in percent correct among the three independent test groups may be indicative of the limited number of pixels available for training.

Table 8.
Percent Correct/Percent Commission Error

Cover	Standard Procedure	Independent Training and Testing Group A	Group B	Group C
Alfalfa	87.2/8.5	74.4/6.6	91.3/36.7	83.0/16.2
Corn	80.4/28.4	61.0/63.1	38.0/36.4	20.0/79.6
Farmstead	73.7/72.0	----	66.7/89.5	87.5/64.1
Sugarbeets	84.1/14.5	78.9/12.4	80.0/34.5	76.3/38.0
Wasteland	31.4/75.4	0.0/100.0	27.3/97.2	52.7/83.9
Small grain	89.2/5.8	90.4/11.5	62.5/22.3	77.1/5.8
Pasture	76.7/26.5	55.1/70.0	44.4/93.7	3.1/78.5
Idle	76.9/11.7	27.5/25.8	67.7/27.4	65.5/38.9
Potatoes	77.9/38.6	60.1/53.6	22.4/74.0	39.5/51.1
Range	82.9/16.1	65.9/29.6	9.0/25.0	37.0/98.6
Beans	66.5/41.1	51.2/29.6	50.0/88.0	20.3/73.4
Onions	85.7/51.4	----	----	92.1/46.3
Overall % Correct	81.4	70.6	42.0	55.2

The total number of pixels classified in each cover type by county and stratum was required prior to estimating acreages. The EDITOR aggregation program was employed to intersect the full scene classified images with the mask files containing county and stratum boundaries. The output from the aggregation program was a table listing the number of pixels from each spectral class in each stratum within each county. The class totals were summed by cover type to generate the required numbers.

The total acreage numbers for the modified jackknife procedure required an additional step because there were three separate classifications of the four county area, and, therefore, three "independent" totals for each cover type. The totals were averaged to obtain the number used as input in the regression estimator.

The results of the aggregation for the independent and dependent data sets are listed in Tables 9 to 12.

Table 9.
Aggregation Totals, in Acres,
for Strata 15 and 25 - Gooding County

Cover	Standard Procedure		Independent Training and Testing Group A		Group B		Group C		Jackknife (Mean A+B+C)	
	15	25	15	25	15	25	15	25	15	25
Alf.	27365	4896	25035	4312	29401	5246	31766	5657	28734	5071
Corn	9742	2347	11868	3398	9172	1746	10696	2175	10578	2439
Sgb.	3698	410	4722	564	8169	1134	5512	741	6134	813
Bns.	5357	766	3762	474	5888	1271	2426	387	4025	2052
Smgr.	22881	3768	24905	4007	21616	3856	25642	4374	24057	4079
Pot.	11796	2579	14816	3216	6181	1152	2967	523	7988	1630

Table 10.
Aggregation Totals, in Acres,
for Strata 15 and 25 - Jerome County

Cover	Standard Procedure		Independent Training and Testing Group A		Group B		Group C		Jackknife (Mean A+B+C)	
	15	25	15	25	15	25	15	25	15	25
Alf.	33497	5936	31751	5561	33930	6210	36421	6520	34034	6097
Corn	9445	1857	12105	2897	9393	1620	10382	2132	10627	2219
Sgb.	8352	1934	9431	1991	13758	2895	10198	2378	10303	1601
Bns.	9681	1503	7114	1057	10349	2223	7175	787	8212	1355
Smgr.	42203	3695	44271	10851	41034	10833	44391	10602	43232	10750
Pot.	13179	3131	15569	2771	7739	1366	5590	1015	9633	1717

Table 11.
Aggregation Totals, in Acres,
for Strata 15 and 25 - Lincoln County

Cover	Standard Procedure		Independent Training and Testing Group A		Group B		Group C		Jackknife (Mean A+B+C)	
	15	25	15	25	15	25	15	25	15	25
Alf.	15484	3830	11841	2870	17633	4318	18194	4378	15899	3865
Corn	3574	1129	5323	2186	3544	810	3591	1202	4152	1399
Sgb.	3228	519	3350	517	4687	893	3746	692	3927	2104
Bns.	1192	510	1394	350	2422	809	1330	340	1715	500
Smgr.	14194	6109	16091	5181	14810	4928	15870	5154	15590	5088
Pot.	5009	1526	6467	525	2765	700	1296	368	2509	531

Table 12.
Aggregation Totals, in Acres,
for Strata 15 and 25 - Minidoka County

Cover	Standard Procedure		Independent Training and Testing Group A		Group B		Group C		Jackknife (Mean A+B+C)	
	15	25	15	25	15	25	15	25	15	25
Alf.	22303	982	20459	897	21919	1130	23872	1061	22083	1020
Corn	8187	764	11128	1570	7249	520	9327	952	9234	1014
Sgb.	27772	2048	28495	1963	33967	2509	31172	2173	31210	1562
Bns.	10751	560	8141	453	11657	665	6696	382	8831	385
Smgr.	69994	5715	75013	6233	69988	6317	72884	6047	72629	6199
Pot.	14851	1017	16127	1189	9707	728	7277	583	11037	833

Estimation

Two types of estimates were generated for each cover type -- direct expansion of the ground data and regression. The direct expansion estimates were generated using EDITOR software and the ground truth files. Since digitized acreage was substituted for crop acreage in the twelve segments not from the JES in 1983, the direct expansion estimate, given in Table 13. may be slightly inflated.

Table 13.
Acreage Estimates - Direct Expansion of Ground Data

County	Cover	Stratum 15 Estimate	Stratum 15 CV(%)	Stratum 25* Estimate	Stratum 25* CV(%)	Strata 15+25* Estimate	Strata 15+25* CV(%)
Gooding	Alfalfa	18075	28.23	--	--	21558	35.83
	Corn	20695	19.54	--	--	24693	29.45
	Sugarbeets	0	--	--	--	0	--
	Beans	1231	62.67	--	--	1469	66.55
	Small grain	28412	37.36	--	--	33886	45.42
	Potatoes	11314	50.57	--	--	13494	55.25
Jerome	Alfalfa	32755	11.50	--	--	39284	22.46
	Corn	8381	53.01	--	--	9668	61.09
	Sugarbeets	3672	69.93	--	--	6577	32.37
	Beans	13319	47.78	--	--	21132	56.00
	Small gr.	39276	25.30	--	--	40561	20.10
	Potatoes	3889	91.08	--	--	4487	99.64
Lincoln	Alfalfa	13351	49.97	3925	63.64	17277	41.23
	Corn	2613	98.92	86	97.18	2699	95.81
	Sugarbeets	7000	98.92	4655	97.18	11755	70.82
	Beans	5292	98.92	0	--	5292	98.92
	Small gr.	25223	30.71	11509	79.21	36732	32.57
	Potatoes	5338	98.92	0	--	5338	98.92
Minidoka	Alfalfa	28906	29.08	--	--	28818	33.31
	Corn	8162	42.24	--	--	8137	46.28
	Sugarbeets	30150	31.84	--	--	30059	35.98
	Beans	6193	55.67	--	--	6174	59.90
	Small gr.	62012	19.24	--	--	67306	19.01
	Potatoes	12952	30.11	--	--	12912	34.30
Above Four Combined	Alfalfa	93263	12.61	10713	57.46	103976	12.77
	Corn	39781	22.10	200	98.80	39982	22.00
	Sugarbeets	41604	32.11	13539	78.05	55143	30.89
	Beans	31079	35.42	0	--	31079	35.42
	Small gr.	154870	12.76	41377	47.73	196248	14.24
	Potatoes	33154	25.97	0	--	33154	29.97

*If no estimate is listed for stratum 25 (--), less than 2 segments were present from that stratum in that county, and the combined estimate was generated by pooling strata 15 and 25.

The Landsat to ground correlations for the regression estimates came from the classifications of the NB files. Following classification, segment total files were created for each of the four classifications. The segment total files were crop specific. Each file contained a list of segments, the stratum assignment for each segment, and the number of pixels of the given cover type in the segment from the ground data and the Landsat classification. The segment total files from the modified jackknife procedure contained only about 1/3 of the segments and had to be combined using a text editor to put the data in a format suitable for estimation.

The regression estimates were generated using a procedure under study by USDA. The procedure employs the Battesse/Fuller estimator to calculate small area (county) estimates from data covering a larger region [1]. The Battesse/Fuller model is:

$$Y[kc] = b[0c] + b[1c] (X[kc]) + V[kc] + E[kc]$$

where:

Y[kc]	=	acreage of crop c in county k
b[0c],b[1c]	=	analysis district single variable regression parameters
X[kc]	=	number of pixels classified to crop c in county k
V[kc]	=	county effect on the regression for crop c in county k
E[kc]	=	random error

The characteristics of the model are described in Sigman and Walker [9], and its use in an agricultural inventory similar to the one described in this paper is given in Holko [4]. The Battese/Fuller program was applied to the four county area because county estimates were desired. The program works on one cover type at a time. It requires as input an EDITOR segment total file, the segment catalog and frame unit files, the number of frame units (sample units) by county and stratum, and the aggregated total of pixels (or acres) of the cover type of interest by county and stratum. Output from the program includes the following options: estimate of the mean, estimate of the total, b0, b1 and r-square by county and combined by stratum. The Battesse/Fuller estimates were generated for the standard and modified jackknife procedures. Results are given in tabels 14 and 15.

Table 14.
Acreage Estimates - Standard Procedure*

		Stratum 15		Stratum 25		Strata 15+25	
County	Cove	Estimate	CV(%)	Estimate	CV(%)	Estimate	CV(%)
Gooding	Alfalfa	29376	7.05	6066	6.69	35443	5.95
	Corn	11654	14.18	672	84.67	12372	14.17
	Sugarbeets	1140	237.72	586	40.96	1727	157.50
	Beans	3755	62.23	--	--	3755	--
	Small gr.	24912	11.88	3446	19.41	28378	10.69
	Potatoes	9281	28.20	--	--	9281	--
Jerome	Alfalfa	34091	6.47	6415	.85	40507	5.45
	Corn	2919	59.68	477	90.57	3396	52.86
	Sugarbeets	8109	35.65	2930	1.96	11040	26.20
	Beans	16423	15.41	--	--	16423	--
	Small gr.	41443	7.60	3642	72.46	45086	9.11
	Potatoes	5724	48.57	--	--	5724	--
Lincoln	Alfalfa	17825	10.35	4858	.42	22684	8.13
	Corn	2008	70.42	287	81.18	2296	62.41
	Sugarbeets	3915	63.65	527	4.65	4488	55.53
	Beans	--	--	--	--	--	--
	Small gr.	18043	14.91	5419	18.62	23463	12.25
	Potatoes	819	339.44	--	--	819	--
Minidoka	Alfalfa	24672	9.80	1279	1.64	25952	9.32
	Corn	4065	46.27	229	75.11	4295	43.98
	Sugarbeets	35009	9.69	2360	.96	37369	9.08
	Beans	11494	24.25	--	--	11494	--
	Small gr.	69911	5.08	6032	16.10	75943	4.85
	Potatoes	6495	48.25	--	--	6495	--
Above Four Combined	Alfalfa	105669	4.07	18619	2.23	124586	3.48
	Corn	20684	17.85	1667	80.98	22316	17.58
	Sugarbeets	48174	11.93	6450	3.92	54625	10.44
	Beans	29561	6.53	0	--	29561	--
	Small gr.	154311	3.95	18561	19.95	172872	4.13
	Potatoes	22320	25.63	0	--	22320	--

*Estimates and errors were calculated using the Battese/Fuller estimator in the USDA-SRS program DFB; delta=1.

Table 15.
Acreage Estimates - Jackknife Procedure*

County	Cover	Stratum 15 Estimate	Stratum 15 CV(%)	Stratum 25 Estimate	Stratum 25 CV(%)	Strata 15+25 Estimate	Strata 15+25 CV(%)
Gooding	Alfalfa	25605	12.77	7049	12.26	32655	10.36
	Corn	19991	17.45	244	.81	20236	17.24
	Sugarbeets	294	1274.83	875	29.49	1169	321.39
	Beans	--	--	--	--	--	--
	Small gr.	29434	15.29	3994	8.89	33428	13.50
	Potatoes	9769	4.25	--	--	9769	--
Jerome	Alfalfa	32987	10.43	6675	17.25	39662	9.15
	Corn	7463	50.90	216	1.90	7679	53.21
	Sugarbeets	5892	67.00	1159	1.55	7142	--
	Beans	17505	30.54	--	--	17505	--
	Small gr.	42519	11.49	11210	3.61	53729	9.12
	Potatoes	3616	123.64	--	--	3616	--
Lincoln	Alfalfa	16120	17.97	5824	12.29	21945	13.37
	Corn	2315	137.54	112	.14	2427	131.19
	Sugarbeets	3666	93.94	2633	.27	6299	54.69
	Beans	1651	285.71	--	--	1651	--
	Small gr.	20055	20.67	5201	3.27	25257	16.43
	Potatoes	2229	169.54	--	--	2229	--
Minidoka	Alfalfa	25893	14.72	1455	28.59	27348	14.02
	Corn	5375	81.30	103	.15	5479	79.76
	Sugarbeets	35245	13.34	2015	.35	37261	12.62
	Beans	7042	82.28	--	--	7042	--
	Small gr.	71480	7.85	6072	2.64	77508	7.24
	Potatoes	10694	48.80	--	--	10694	--
Above Four Combined	Alfalfa	100605	8.23	21003	11.49	121611	5.79
	Corn	35164	21.89	677	1.11	35823	21.48
	Sugarbeets	45188	17.43	6683	5.49	51872	15.20
	Beans	24309	-4.87	--	--	24309	--
	Small gr.	163489	5.81	26434	2.58	189923	5.01
	Potatoes	26304	35.57	--	--	26304	--

*Estimates and errors were calculated using the Battese/Fuller estimator in the USDA-SRS program DFB; delta=1.

Results

Discussion

Acreage estimates at the county level and for the four counties combined were generated using three methods -- direct expansion of the ground data, the standard procedure and the modified jackknife procedure. The standard procedure performed the best for the four county combined estimates i.e. errors for all crops were lowest and Landsat-to-ground correlations (R-squares) were best (tables 13-16). Acreage estimates, errors and R-squares from the modified jackknife procedure were similar to those from the standard procedure for three cover types -- sugarbeets, alfalfa and small grain.

Table 16.
Regression Estimate Parameters and Relative Efficiency (RE)*
- Gooding, Jerome, Lincoln, Minidoka Counties Combined

Standard Procedure

	Stratum 15				Stratum 25				Strata 15+25		
Cover	b0	b1	r2	RE	b0	b1	r2	RE	b0	b1	r2
Alf.	12.1	.94	.89	8.41	4.3	1.29	.99	29.48	9.2	.98	.90
Corn	16.4	1.24	.84	5.53	-.8	.30	.51	1.02	-22.1	1.31	.84
Sgb.	-1.6	1.23	.88	8.35	1.6	1.15	.99	827.91	-.8	1.20	.91
Bns.	-29.4	2.10	.81	4.77	--	--	--	--	-29.4	2.10	.81
Smgr.	6.3	.99	.91	25.29	-12.3	1.06	.99	73.50	1.6	1.01	.94
Pot.	-24.2	1.02	.65	2.18	--	--	--	--	-24.2	1.02	.65

Jackknife Procedure

	Stratum 15				Stratum 25				Strata 15+25		
Cover	b0	b1	r2	RE	b0	b1	r2	RE	b0	b1	r2
Alf.	23.5	.77	.74	3.48	-1.7	1.44	.95	11.12	20.3	.81	.76
Corn	6.6	.79	.16	1.09	-.6	.11	.97	15.88	1.8	.84	.22
Sgb.	-12.7	1.11	.68	2.84	-3.0	1.29	.99	352.30	-12.4	1.17	.76
Bns.	-1.6	1.14	.14	1.06	--	--	--	--	-2.9	1.14	.17
Smgr	16.9	.94	.80	5.64	-8.13	1.08	.99	113.15	9.2	.99	.87
Pot.	6.2	.63	.12	1.05	--	--	--	--	.8	.69	.17

*RE = ground variance/regression variance
-- = could not be calculated due to zero value in one term

Although the errors and R-squares from the modified jackknife procedure for corn, beans and potatoes were less satisfactory than those from the standard procedure, they may be more representative of the informational content of Landsat data, because the same data was not used for training and testing the classifier. In that regard, the relative efficiencies for the jackknife procedure given in Table 16. are more conservative indicators of what can be expected when Landsat data is used in this type of an inventory than the relative efficiencies shown for the standard procedure.

Both regression procedures improved upon the precision of direct expansion of the ground data for the large acreage crops in the four county combined estimates. When discussing the results from a similar jackknife procedure performed with Landsat data from California, Holko [5] noted that the best results were obtained when independent training was done on cover types with large acreage. That result was duplicated in this experiment.

The poor R-squares and large error terms for corn, beans and potatoes in the modified jackknife procedure may be attributable to inadequate training or the date of the second Landsat acquisition. Display of segment data from the September 1 acquisition revealed that most of the potato fields were either turned (vine death), partially turned, or harvested - well beyond the biostage desirable for acreage estimation with Landsat. An earlier acquisition might improve the performance of the classifier, but a definitive statement to that effect cannot be made based on the results from this study.

Regression parameters and relative efficiency for county level estimates, strata 15, are shown in Table 17. The results, at the county level, parallel those for the four counties combined, i.e. both regression procedures improved upon direct expansion for the large acreage crops. However, even in stratum 15, only Minidoka and Jerome Counties had enough segments, 10 and 8 respectively, to attempt comparisons between the standard and modified jackknife procedures.

Regression parameters at the county level could not be generated for stratum 25 because of the limited sample size.

Table 17.
Regression Estimate Parameters and Relative Efficiency (RE)*
by County - Strata 15

County	Cover	Standard Procedure b0	b1	r2	RE	Jackknife Procedure b0	b1	r2	RE
Gooding	Alfalfa	27.7	.78	.74	2.28	50.4	.37	.38	.97
	Corn	-3.5	1.27	.73	2.24	150.0	-1.04	.43	1.06
	Sugarbt	--	--	--	--	--	--	--	--
	Beans	-1.5	.38	.58	1.43	5.4	.01	.00	.60
	Smallgrn	-18.9	1.23	.96	14.54	3.1	1.21	.78	2.74
	Potato	-36.1	1.39	.78	2.73	34.9	.36	.03	.62
Jerome	Alfalfa	17.0	.88	.89	6.69	5.0	.93	.92	8.64
	Corn	-38.6	1.37	.84	4.46	-55.8	1.95	.76	3.01
	Sugarbt	7.7	.34	.06	.76	9.4	.14	.01	.72
	Beans	-31.6	2.54	.98	30.76	-12.8	2.44	.32	1.05
	Smallgrn	-44.7	1.29	.98	55.13	-78.7	1.53	.80	3.68
	Potato	-32.8	1.14	.91	8.02	-16.5	.83	.22	.91
Lincoln	Alfalfa	13.0	1.05	.87	--	-4.3	1.13	.98	--
	Corn	-21.6	1.39	.94	--	6.8	.36	.01	--
	Sugarbt	-3.1	1.29	.99	--	-9.9	1.21	1.00	--
	Beans	-18.6	1.42	.98	--	-44.0	2.28	.90	--
	Smallgrn	72.1	.70	.78	--	55.9	.83	.72	--
	Potato	-23.9	.92	.32	--	7.2	.51	.32	--
Mini-doka	Alfalfa	3.2	1.04	.96	17.22	7.5	1.02	.93	12.12
	Corn	-11.7	1.05	.93	10.39	-13.7	.93	.57	1.79
	Sugarbt	10.7	1.12	.83	4.38	5.1	1.07	.65	2.20
	Beans	-53.1	2.91	.49	1.53	30.5	-.48	.05	.82
	Smallgrn	21.3	.89	.96	21.03	29.5	.84	.95	15.14
	Potato	-7.5	.73	.60	1.96	10.7	.64	.19	.96

*RE = ground sample variance/regression variance
-- = could not be calculated due to zero value in one term

County level estimates would probably improve if additional training were made available either by increasing the JES sampling rate further or by devising an alternative procedure for collecting training and test data. The California Cooperative Remote Sensing Project is preparing an experiment in which JES segments will be used for testing the classifier, but the training data will be collected along transects. The transect data will provide more training at a lower cost than JES data and will allow all the JES data to be reserved for testing. A similar methodology could be applied in Idaho where a shorter growing season and less phenological variability in a given crop at a given time should make collecting an adequate training set a manageable problem.

Summary

1. Two methods for generating regression estimates using Landsat data from four counties in the central Snake River Plain produced significantly better estimates for large acreage crops than direct expansion of the ground data. The improvement was noted in the estimates for the counties combined and for county level estimates, although the county level estimates are suspect because of the small number of samples.

2. The results of this test support the contention that training and testing on the same data set introduces a bias in the results.

3. A 50% increase in the JES sampling rate was required to obtain independent training and testing data sets that were, at best, marginally sufficient in size. If additional work of this type is to be performed in Idaho, the acquisition of adequate training and test data sets will have to be addressed.

4. A late August, early September Landsat acquisition appears to be too late in the growing season for differentiation of summer crops in the central Snake River Plain.

References

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

[2] Colwell, R.N. et. al. (editor). Manual of Remote Sensing (2nd edition). American Society of Photogrammetry, Falls Church Virginia. 1983

[3] Hanuschak, G.A., R.D. Allen and W.H. Wigton. "Integration of Landsat Data into the Crop Estimation Program of USDA's Statistical Reporting Service." Machine Processing of Remotely Sensed Data Symposium, Purdue University, West Lafayette, Indiana. 1982

[4] Hanuschak, G.A., M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno and C. Miller. "Obtaining Timely Crop Area Estimates Using Ground-Gathered and LANDSAT Data." USDA ESCS Technical Bulletin No. 1609. 1979

[5] Holko, M.L. "1982 Results for the California Cooperative Remote Sensing Project." Statistical Reporting Service Staff Report No. AGES840305, Washington, D.C. 1984

[6] Idaho Image Analysis Facility - Project Plan for FY83

[7] Lennington,R.K, and M.E. Rassback. "CLASSY - An Adaptive Maximum Likelihood Clustering Algorithm." Proceedings of the LACIE Symposium, Houston, Texas. 1972

[8] Ozga, M., S.R. Faerman and R.S. Sigman. "Editor Multitemporal System." USDA, ESCS. 1979

[9] Sigman, R. and G. Walker. "The Use of Landsat for County Estimates of Crop Acres: Evaluation of the Huddleston-Ray and Battese-Fuller Estimators." USDA, SRS, Washington, D.C. 1982

[10] Thomas, R.W. personal communication

[11] Wensel, L.C., S.J. Titus and R.W. Thomas. "Sample Allocation in Multiresource Inventories." Proceeding, Society of American Foresters Workshop on Forest Resource Inventories. Colorado State University, Ft. Collins. 1979

[12] Wensel, L.C., M. Eriksson, R.W. Thomas and D.R. Taylor. "Survey Planning Model - User's manual. Part of the final report for USFS, Rocky Mountain Station Cooperative Agreement No.16-935-Ca and NASA Johnson Space Center Contract No. NAS9-15987. Department of Forestry and Resource Management and Remote Sensing Research Program, University of California, Berkeley. 1981

Acknowledgments

The author wishes to express his gratitude to:

The National Aeronautics and Space Administration, under whose auspices this research was done.

Hal N. Anderson, Director of the Idaho Image Analysis Facility, for initiating this project and providing the inspiration to bring it to a successful conclusion.

Ethel Bauer, NASA-Ames Research Center, for making this project happen and giving the author the opportunity to work on it.

Richard Sigman, Marty Holko, Paul Zuttermeister, and Martin Ozga, USDA, Statistical Reporting Service, Remote Sensing Branch, Washington, for their support, for supplying the USDA sample frame data, and for their willingness to answer innumerable questions about EDITOR.

Dick Max and Doug Hensley, Statistical Reporting Service, Boise, for providing the ground data and reviewing the work while in progress.

Tony Morse, Idaho Image Analysis Facility, for reviewing the final draft.

Dr. Randy Thomas, Remote Sensing Research Program, University of California, Berkeley, for sharing his expertise in the design of agricultural inventories while supervising the design of this project, and for always being there when called upon.