United States
Department of
Agriculture

Economics and
Statistics
Service

**Statistical
Research
Division**

**ESS
Staff Report
No. AGESS 810519**

Washington, D.C. 20250

May 1981

Estimating Potatoes and Other Crops

In the Red River Valley of North Dakota and Minnesota Using 1980 LANDSAT Imagery

Sherman B. Winings

ESTIMATING POTATOES AND OTHER CROPS IN THE RED RIVER VALLEY OF NORTH DAKOTA AND MINNESOTA USING 1980 LANDSAT IMAGERY. By SHERMAN B. WININGS, Statistical Research Division, Economics and Statistics Service, U. S. Department of Agriculture, Washington, D.C. 20250 ESS Staff Report No. AGESS810519, May 1981.

ABSTRACT

This report describes how data from LANDSAT was used in conjunction with ground data to estimate selected crops in the Red River Valley. An extra ground sample was used to increase training data and precision of the ground data estimates for minor crops. Estimates using LANDSAT and the combined ground data samples were generally more precise than the usual ground data alone. The timing of the available LANDSAT scene and inadequate auxiliary information for stratification hampered the complete success of the project.

Key Words: LANDSAT, Red River Valley, potatoes, sugarbeets, EDITOR, Regression , Combined estimates.

ACKNOWLEDGMENTS

The author would like to give special thanks to Marty Holko and Richard Sigman for their technical assistance; Sandy Stutson, Eric Hendry, George Harrell, Pearl Jackson, Lillian Schwartz, and Tjuana Fisher for for their help in digitizing; George Roney and Ed Camera for preparing the maps and cover; and Paul Hopkins and Bessie Johnson for their aid in word processing.

TABLE OF CONTENTS

Red River Valley

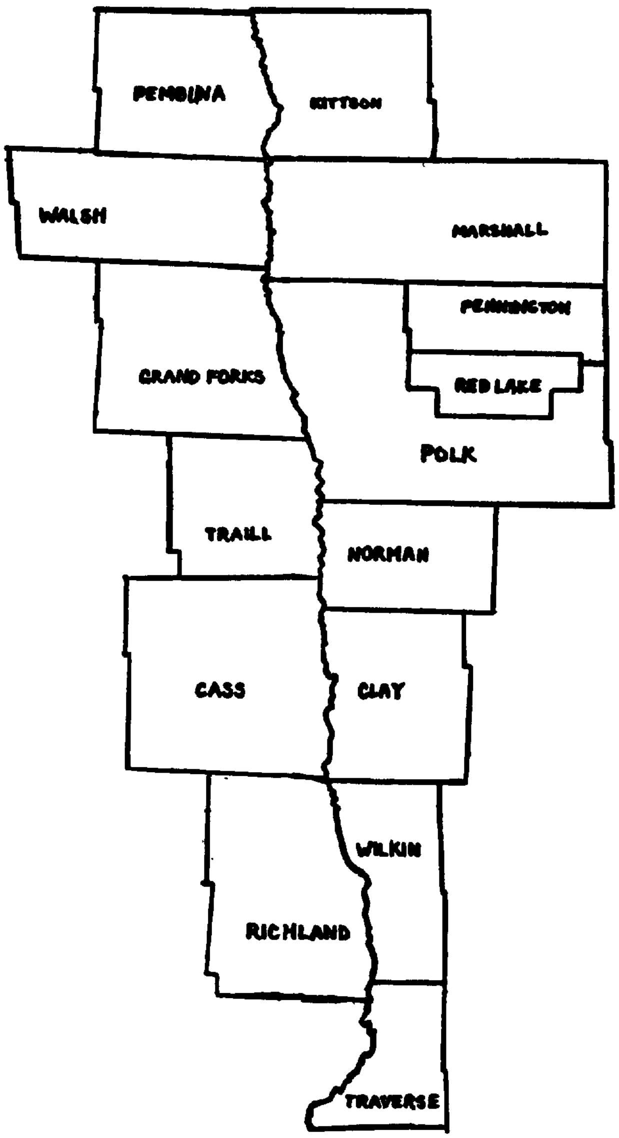


Figure 1. Outline Map Showing Study Area Counties.

General

The purpose of this study was to apply existing ESS procedures for using LANDSAT data in estimating crop areas to that of a specialty crop (potatoes), plus extend these procedures to a broader range of spring planted crops than those experienced in previous ESS LANDSAT projects. Existing ESS LANDSAT techniques consist of supervised training of a maximum-likelihood pixel classifier using as ground truth ESS's June Enumerative Survey (JES) data and then calculating crop-area estimates by applying regression estimation to full scene classification results. This study differed from previous ESS LANDSAT studies in that (1) in order to get satisfactory sampling errors specialty crops require more ground data than that available from the JES; and (2) the study site was located further north than previous studies which indicates temporal shifts in desirable LANDSAT windows.

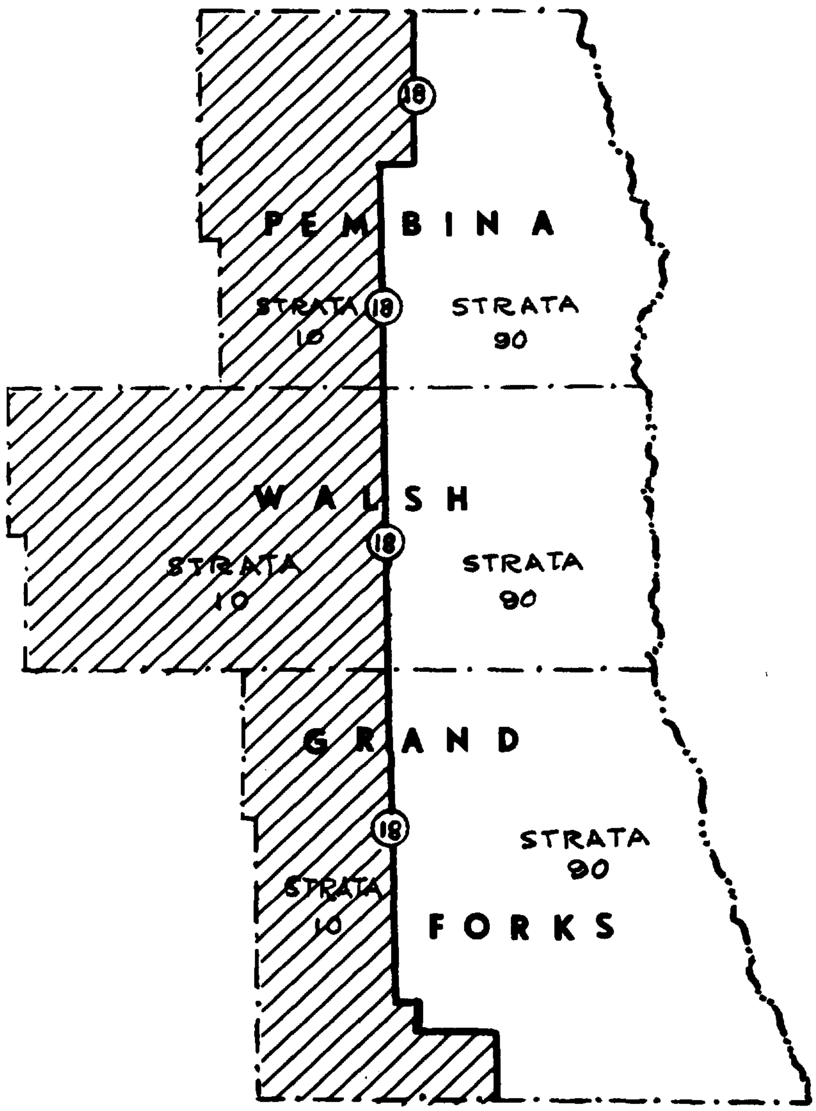
The Red River Valley consists of eight counties along the North Dakota border in Minnesota and the six easternmost counties of North Dakota, see Figure 1. To the east of the valley is the "pot hole and kettle" terrain of the great north woods of Minnesota which is considerably different from the intensive agriculture of the valley. To the west the valley gives way to glacial-till prairie which, although the land is not as productive as the valley, is still intensively cultivated.

Ground Truth Data Sample

An annual ESS survey of a sample of land areas is conducted each year about June first. This survey is called the June Enumerative Survey and will be referred to as JES. The entire area of each state is stratified according to land use and cultivation intensity. The strata then are subdivided into count units or primary sampling units containing 7 to 11 square miles in the most intensively cultivated stratum and various other sizes in other strata. The PSU's are selected randomly for the JES sample. Since a PSU is cumbersome to enumerate, a section sized segment is selected randomly within the sample PSU.

In the North Dakota counties of Pembina, Walsh and Grand Forks, count units (PSU's) of the most intensive cultivated stratum was placed in sub-strata depending on whether or not they were mostly on the valley floor. The sub-stratum more likely to contain potatoes was bounded on the west by North Dakota state highway 18, see Figure 2. The other bounds were the original stratum bounds. Conversely, the other sub-stratum was bounded on the east by this highway and its other bounds were also the original stratum bounds. Although the highway was the general bound the exact bounds followed count unit bounds such that each count unit was contained wholly within a single sub-stratum.

Grand Forks, Pembina and Walsh Counties
of North Dakota



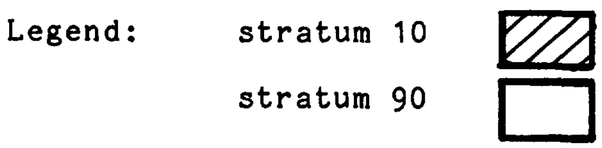


Figure 2 Sub-stratum Boundaries in Grand Forks,
Pembina and Walsh Counties

In the 14 counties 139 JES segments were enumerated in the regular ESS program. This JES data was edited and entered into the data processing stream by field. In addition 40 additional segments one-fourth the size of the JES segments were selected in the most intensively cultivated strata in Minnesota and on the valley floor in North Dakota. In Pembina, Walsh and Grand Forks counties these "mini" segments were placed in the easternmost sub-stratum. These additional segments were selected with probability proportional to the historical county potato acreage. These mini segments were enumerated by observation.

It was hoped that current year ASCS photography would be available for the observations, which were planned to take place the week of July 28 to August 1. The observations were delayed one week (August 4-8), to wait for the current photography which then became available on only a limited basis. Only one county had current photography for the observation work. All other segments were recorded by sketching the segment on paper and later transferring the observations to the current photography.

Current photography became available for all counties about September 1. For two segments current photography was never received and scale sketches of these two segments were made on 7-1/2 minute quadrangle maps. Current photography was also used to edit the JES segments. Photocopies of JES photography and current photography were used to digitize the segments.

In future studies of this type it is recommended that photography one year old be obtained from ASCS for the mini segments and that observations be made earlier in July. The observed field boundaries then would be transferred to current year photography and be used to digitize the segments.

The ASCS photography was obtained in two formats. The first was one section per 35mm frame and the other was two sections per 35mm frame. The two section format is more desirable since it was easier to locate calibration points on it. In addition, the "fish eye" effect was much more evident in the one section(1 square mile) format. "Fish eye" effect is distortion of the photographic scale such that the image looks as if it was stretched over a sphere.

Enlargements were obtained at 5X7 and 8x10 inch sizes. The 8x10's were approximately twice as expensive as the 5X7's. The 8X10's can be digitized more accurately than the 5X7's. In the future it is recommended that 8x10's be used for the mini segments that are digitized from the enlargements and 5X7's be used for the JES segments. Current year photography is used to edit rather than digitize the JES segments. In using these enlargements many fields can be used as candidate training data in cases where the enumerator has erred in sketching the fields.

LANDSAT Imagery

Study of Burkhead, et. al. suggests a LANDSAT data window from July 14 to September 17 with August 10 being optimum. Six LANDSAT scenes are required to cover the entire area. Considering both LANDSAT II and III there were eight passes per

Table 1. LANDSAT Coverage of Study Area

			EROS Data Base			Percent Clouds		
Date	LANDSAT	Total	Missing	Late	On	0-15%	16-50%	51-100%
7/14-16	II	6	3	0	3	0	0	3
7/23-25	III	6	0	0	6	3	1	2
8/1-3	II	6	3	0	3	1	1	1
8/10-12	III	6	1	0	5	0	1	4
8/19-21	II	6	3	1	2	0	0	2
8/28-30	III	6	3	0	3	1	0	2
9/6-8	II	6	0	3	3	2	1	0
9/15-17	III	6	0	6	0	0	0	0
Totals		48	13	10	25	7	4	14
% of Total		100%	27%	20%	53%			
% of EROS Database						28%	16%	56%
Add Lates					10	1	3	6
Total Including Lates					35	8	7	20
% of Total					73%			
% of EROS Data Base						23%	20%	57%

scene during the data window. Thus, there was a potential total of 48 scenes within the LANDSAT window. As of October 15, 23 scenes were not in the data base: 4 of these were clear, 7 were cloudy and 12 could not be determined from NOAA satellite imagery.

Twenty five of these scenes were in the EROS data base, see Table 1. Of these 25, seventeen (of the 4 scenes with 16 to 50 percent cloud cover scene only one was useable) were too cloudy to use, 3 LANDSAT III had "sawtooth", 1 LANDSAT II scene had the "salt and pepper" effect, and 4 scenes were usable. "Sawtooth" is a data defect causing improper alignment of adjacent rows of pixels in the scene. "Salt and pepper" is small white and black dots appearing on the image. The 3 useable LANDSAT II scenes had some striping. The single LANDSAT III scene is missing the western 1/3 of the scene; however, this is out of the study area. This scene also is somewhat cloudy in the Pembina county portion. On October 27 the CCT and 1:1,000,000 film negatives of the LANDSAT III scene were received. Upon examination of the negatives, striping was observed in this scene also. Greyscales were examined to determine if the striping is in the digital tape as well as the film negative. The greyscales did not show any appreciable striping.

Analysis Districts

LANDSAT coverage of the counties of Richland, Cass, and Traill in North Dakota and all Minnesota counties (Clay, Kittson, Marshall, Norman, Pennington, Polk, Red Lake, Traverse, and Wilkin) were defined to be an analysis district, which is to be analyzed using one set of LANDSAT training data. LANDSAT coverage was available for portions of Pembina, Walsh and Grand Forks counties. There were few segments in the unimaged portions of these three counties. Consequently, it was doubtful that much gain would result from LANDSAT analysis, and thus direct expansions were used for these counties. It was unfortunate that these 3 counties were missed because most of the potatoes in North Dakota are grown in these three counties.

Timing of LANDSAT Coverage

The three LANDSAT II scenes were all dated September 6, 1980. This is well beyond the crop-growth window for summer harvested crops (wheat, oats, barley, and rye) and thus

```
Planting
///////XXXX\\\\
    Emergence
    ///////XXXXX\\\\
        Partial Crop Canopy
        ///////////X\\\\\\\\\\\
               Full Crop Canopy
               ///////////XXXXXXXX\\\\\\\\\\\\\\\\\\
                             Crop Turning Color
                             /////////          \\\\\\\\
                                   Harvest
                                   //////                \\\\\\\\
----¦----------¦---------¦----------¦---------¦----------¦--------
20     10  20     10  20    10  20     10  20    10  20    10  20
April   May        June      July      August  September  October

Legend
   Summer Harvested Crops (wheat, barley,          //////////
   oats and rye)
   Row Crops (corn, soybeans, sunflower,           \\\\\\\\\\
   sugarbeets and potatoes)
```

Figure 3. Crop Growth Calendar
Red River Valley of North Dakota and Minnesota

summer-harvested crops were not analyzed, see Figure 3. Of the five row-crops, two (corn and sugarbeets) had good reflectance, one (soybeans) had fair reflectance and two (potatoes and sunflower) had less than optimum reflectance for analysis.

EDITOR Analysis(1)

At the time of this study, the EDITOR software system which ESS uses for analysis had not implemented multi-state capabilities. EDITOR was tricked such that it processed the two states as if it were one state. This was done by renaming all files containing Minnesota data from extension "MN80" to "ND80", adding 500 to all segment numbers in Minnesota, and renumbering all Minnesota strata.

Upon examination of the ground truth files, it was found that some covers were not in standard nomenclature. For example dry beans were listed as beans, dry edible. Covers were edited to standard names. Fields greater than 10 acres with digitized acres differing from reported acres by more than 10 percent were noted as "bad-field" (not to be used for training). Fields greater than 10 acres containing more than 5 percent waste or second crop were also noted "bad-field".

Each of the six row-crops were packed separately (Dry beans, also a row crop was packed but not analyzed). Wheat, oats, barley and rye were packed into a single cover. Alfalfa, other hay, flax, other crops, pasture, crop land pasture, wasteland, and dense woodland were also packed as a single cover. Eight statistics files were obtained from the 8 packed files. A single category per crop file was formed by adding water statistics to the statistics (means and variances) of each of the packed files (9gps).

Each of the 8 packed files was clustered (ordinary cluster) and the resulting clusters were added to make a statistics file of 21 clusters. The statistics of the clusters and water statistics were combined to give the second file (22gps). Upon examination of the 22 group file 6 clusters of various covers were not separable. By removing the 6 clusters a 16 group file was created (16gps). Probabilities obtained in packing were added to each of these files giving three more classifications using prior probabilities.

Small scale regression estimates were run for the five crops using all six statistics files. The statistics file giving the "best" R-square for each crop was used in creating the estimator parameter file that was used in the large scale (full scene) regression estimation. Strata having more than one type of segment (JES or "mini") had their JES and "mini"-segment estimates combined using the reciprocals of the estimated

(1) This section is intended primarily for users of the EDITOR software system. Non-EDITOR users can skip to the next section.

variance as weighting factors (see appendix). Relative efficiency was calculated by dividing the variance of the (combined per appendix) direct expansion estimator by the corresponding variance of the regression estimator.

Table 2. Evaluation of the Effect of Additional Segments For Direct Expansion Estimates

Cover	Sample	Standard Deviations	Relative Efficiency
Corn	JES	4809.44	
	Mini	47277.46	
	Comb	4784.74	1.01
Sunflower	JES	33128.08	
	Mini	93892.49	
	Comb	31240.50	1.12
Sugarbeets	JES	61739.50	
	Mini	134537.19	
	Comb	56018.09	1.21
Potatoes	JES	41488.82	
	Mini	85996.88	
	Comb	32155.66	1.66

Results

Five crops were analyzed (corn, sunflower, sugarbeets, soybeans, and potatoes). The additional sampling reduced the direct expansion variances of the strata involved by the expected amount, see Table 2. That is, to reduce the variance by a factor of 2 increase the sample size by 4. The additional sampling did not give the desired increase in precision for potatoes.

Classification results were somewhat inconsistent. One might expect that the classifier having the best percent correct would have the best R-square and the best relative efficiency. However, this was not the case for most of the crops, see Table 3. For example, the classifier for potatoes having the poorest percent correct had the best R-square. The 16 group classifier results were very near the 22 group classifier results and are not shown in the table.

Improvement from using LANDSAT data with ground data over direct expansion estimates for the entire Red River Valley was relatively low. Relative efficiency's ranged from 1.46 for potatoes to 3.82 for corn, see Table 4. Overall relative efficiency's were higher than the classifier efficiency's because classifiers were selected by stratum for each crop. In other studies relative efficiency's have usually ranged from 2 to 6.

Table 3. Classifier Evaluation

Cover	Classifier	Percent Correct	R Square	Relative Efficiency
Corn	9gps/ep	67.75	.455061	2.98
	9gps/pur	12.85	.607506	2.04
	22gps/ep	20.58	.509800	2.67
	22gps/pur	4.35	.348471	1.20
Soybeans	9gps/ep	20.59	.119982	1.20
	9gps/pur	2.98	.144983	1.19
	22gps/ep	17.30	.069802	1.32
	22gps/pur	6.36	.175098	2.01
Sunflower	9gps/ep	24.10	.386312	1.49
	9gps/pur	19.02	.278108	1.34
	22gps/ep	17.29	.459357	1.74
	22gps/pur	38.26	.249006	1.36
Sugarbeets	9gps/ep	72.32	.631736	2.79
	9gps/pur	66.86	.728740	3.21
	22gps/ep	41.97	.721572	2.92
	22gps/pur	59.13	.723540	2.93
Potatoes	9gps/ep	50.86	.105961	1.46
	9gps/pur	.43	.382046	.98
	22gps/ep	34.14	.105525	1.29
	22gps/pur	8.19	.380664	1.48

ep--classifier using statistics files with equal probabilities
pur--classifier using statistics files with prior probabilities

Conclusions

The stratification for potatoes did not give the desired increase in precision. Only 2 of the 40 segments in the extra sample contained potatoes. If the potato stratification is to be successful; about one-half or more of the extra segments should contain potatoes. Finer stratification using potato grower headquarter information will be attempted in 1981. LANDSAT analysis for minor, fall-harvested crops was inconclusive. In 1981, LANDSAT data earlier in the crop growth window should be obtained if possible.

Table 4. Relative Efficiency
Total Study Area

Cover	Estimate	Standard Deviation	Relative Efficiency
Corn	LANDSAT Combined	55336.3	3.82
	Direct Expansion Combined	108107.0	
Sugarbeets	LANDSAT Combined	33892.4	3.58
	Direct Expansion Combined	64118.6	
Soybeans	LANDSAT Combined	63542.5	2.04
	Direct Expansion Combined	90851.3	
Sunflower	LANDSAT Combined	102304.0	1.77
	Direct Expansion Combined	135993.0	
Potatoes	LANDSAT Combined	26564.1	1.47
	Direct Expansion Combined	32155.6	

References

Hanuschak G., R. Sigman, M. Craig, M. Ozga, R. Lubbe, P. Cook, D. Kleweno, C. Miller; Obtaining Timely Crop Area Estimates Using Ground Gathered and LANDSAT Data. Economics, Statistics and Cooperatives Service, U. S. Department of Agriculture, August 1979.

Craig, M. E. and C. E. Miller; Area Estimates by LANDSAT--Arizona 1979. Economics, Statistics and Cooperatives Service, U. S. Department of Agriculture, January 1980.

Kleweno, D. D. and C. E. Miller; 1980 AgRISTARS DC/LC Project Summary--Crop Area Estimates for Kansas and Iowa. Economics and Statistics Service, U. S. Department of Agriculture, March 1981.

Burkhead, C. E., R. C. Max, R. B. Karnes, E. Reid; Usual Planting and Harvesting Dates, Agriculture Handbook no. 283. Statistical Reporting Service, U. S. Department of Agriculture, March 1972.

Rock C., et. al.; Minnesota Agricultural Statistics 1979. Minnesota Crop and Livestock Reporting Service of Economics, Statistics and Cooperatives Service, U. S. Department of Agriculture, June 1979.

Carver R., et. al.; North Dakota Agricultural Statistics 1979. North Dakota Crop and Livestock Reporting Service of Economics, Statistics and Cooperatives Service, U. S. Department of Agriculture, June 1979.

A P P E N D I X

APPENDIX

Combining estimates of "mini" segments with JES segments

Let

Independent estimators $\hat{Y}_1$ and $\hat{Y}_2$ be the estimators derived from the "mini" and JES segments respectively, with variances V_1 and V_2.(2)

$\hat{Y}.$ be the combined estimator for a single stratum.

Then

$$\hat{Y}. = p\hat{Y}_1 + (1-p)\hat{Y}_2 \quad 0<p<1$$

$$V. = p^2V_1 + (1-p)^2V_2$$

By

setting the partial differential of V. with respect to p equal to zero.

Then

$$p(opt) = \frac{V_2}{V_1 + V_2} \quad \text{and} \quad (1-p(opt)) = \frac{V_1}{V_1 + V_2}$$

$$\hat{Y}.(opt) = \frac{V_2\hat{Y}_1}{V_1 + V_2} + \frac{V_1\hat{Y}_2}{V_1 + V_2}$$

$$V.(opt) = \frac{V_1V_2}{V_1 + V_2}$$

---

(2)
For further explanation of combining estimators which may be dependent see "Multiple Frame Estimation With Stratified Overlap Domain"; Bosecker R. R., and B. L. Ford, Statistical Reporting Service, U. S. Department of Agriculture, July 1976.