1979

79-06

USE OF LANDSAT CLASSIFIED PIXELS FOR ESTIMATING ANNUAL LIVESTOCK AND CROP INVENTORIES

Harold F. Huddleston and Ronald Steele

Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture
Washington, D.C. 20250

1979

ABSTRACT

The U.S. Department of Agriculture (USDA) has been exploring methods of using remote sensing as a basis for improving survey methodology. This paper discusses digital techniques employing computer classification of pixels with ground enumerated livestock inventories for the State of Iowa during 1978. The methods of analyses include discriminant functions for classification of LANDSAT tapes and regression methods for making estimates of livestock numbers based on double sampling employing an area sampling frame. The results were much less promising than for acreage estimates, but similar to results for crop yield forecasts. These small gains in estimating efficiencies may have, in part, been due to the time interval between the dependent variable (livestock numbers) and independent variables (classified pixels). However, the combined gains in crop acreages, yields, and production when added to modest gains expected for livestock indicate that the combined economic benefits for agriculture are important.

1. INTRODUCTION

This paper reports on a research activity in the Economics, Statistics, and Cooperatives Service (ESCS), USDA directed at utilizing remote sensing information. These efforts are directed at increasing the efficiency of collecting current agricultural statistics and providing greater geographic detail for small area estimation.[1]

The first major effort by USDA involving remotely sensed information counts for livestock was in California in 1967.[2] The method employed high resolution aerial photographs to obtain counts of animals by species for randomly selected area segments. Image counts were compared to ground enumerated inventories of livestock on the day of the flight. A major problem existed in detecting all the livestock because some animals were hidden from the camera by trees and buildings. The techniques currently being developed do not depend on high resolution remotely sensed information in order that the animals may be discriminated from their background or that the imagery be for the same day. Instead, the methods employed depend on the number of animals being correlated with the acreage inventories for crops and agricultural cover types (i.e., land use).

2. CROP ACREAGE CLASSIFICATION

The classification and acreage estimation methods have been outlined in a series of papers,[3,4,5] by staff members of ESCS in recent years. The LANDSAT tapes for the individual scenes are obtained. The four tapes are reformatted and the spectral values are interfolded for each pixel. Classification of land is done in the computer by use of discriminant functions. The procedure must differentiate between crops and cover types on the basis of reflected energy. Before starting, a sample of data from each of the crops or cover types of interest must be available that represents how they reflect energy. The problem is to set up rules, using samples of pixels for each crop or cover type. The procedure enables unknown land pixels to be allotted to a crop or land cover type given only the amount of reflected energy of that pixel. A maximum likelihood quadratic classifier using the prior probabilities for a category is used for classifying each pixel into one of a series of land uses important for livestock for the entire analysis district contained in the LANDSAT scene. Each LANDSAT pixel within the boundaries of randomly selected area segments is classified based on sample fields selected from known cover types obtained from ground enumeration information. The livestock numbers are also collected for the same segments as the crop acreage data. Based on the classified pixels for the segments and the ground enumerated livestock numbers, a regression estimator is obtained as a basis for making inferences to the total population of classified pixels for the entire area.

3. REGRESSION ESTIMATION OF LIVESTOCK INVENTORIES

The regression estimator utilizes both livestock numbers from an enumerated random sample and classified LANDSAT pixels for the same area sampling units. The estimate of the total Y (livestock inventory by specie) using this estimator[6] is:

$$\hat{Y}_R = \sum_{h=1}^{L} N_h \cdot \bar{y}_{h(reg)}$$

where $$\bar{y}_{h(reg)} = \bar{y}_h + \hat{b}_h(\bar{X}_h - \bar{x}_h)$$

and $\bar{y}_h$ = the average number of livestock by specie per sample unit from the ground survey for the h^{th} land use stratum.

$\hat{b}_h$ = the estimated regression coefficient for the h^{th} land use stratum when regressing ground reported livestock inventories (say, cattle) on classified pixels of a crop type for the n_h sample units.

$\bar{X}_h$ = the average number of pixels of crop acres or land use acres (corn acres) per frame unit in the h^{th} land use stratum. Thus the entire LANDSAT scene must be classified to calculate $\bar{X}_h$.

X_{hi} = number of pixels classified into the crop or land use in the i^{th} area frame unit of the h^{th} stratum.

$\bar{x}_h$ = the average number of classified pixels of crop per sample unit in the h^{th} land use stratum.

x_{hj} = number of pixels classified as the crop in the j^{th} sample unit in the h^{th} stratum.

Generally, livestock production depends on both feed grains and forage crops; consequently, it is reasonable to use a multiple regression estimator employing several crops and cover types as "independent" variables which may be related to livestock production. Hence, $\bar{y}_{h(reg)}$ is rewritten to reflect the use of several independent variables, or

$$\bar{y}_{h(reg)} = \bar{y}_h + {}_1\hat{b}_h({}_1\bar{X}_h - {}_1\bar{x}_h) + {}_2\hat{b}_h({}_2\bar{X}_h - {}_2\bar{x}_h) + \dots + {}_k\hat{b}_h({}_k\bar{X}_h - {}_k\bar{x}_h)$$

where k independent variables are employed.

The estimated (large sample) variance for the regression estimator is:

$$v(\hat{Y}_R) = \sum_{h=1}^{L} \frac{N_h^2}{n_h} \frac{N_h - n_h}{N_h} \frac{1 - R_h^2}{n_h - k - 1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2 \quad .$$

Thus the reduction in variance is substantial if the coefficient of correlation squared is large for most strata.

4. 1978 EXPERIMENTAL RESULTS FOR IOWA

The entire State of Iowa was included in the acreage study, but only results from three of these LANDSAT frames were used for the livestock study. The best available summer images were selected for the study (i.e., August and September 1978). The livestock inventory data were collected in early June from 51, 55, and 21 segments in the three scenes, respectively. The following crop and cover types were classified on an individual pixel basis and used to derive the livestock results: (1) corn, (2) oats, (3) pasture, (4) woods, and (5) all other land. The regression results for hogs and cattle are summarized in tables 1 and 2.

An inspection of the R^2 values suggests the potential for improving the efficiency of ground enumerated inventory numbers may be small based on these sample data. Also, the best variable(s) is likely to depend on the specie being estimated. Possibly, higher correlations would be expected if the time interval between the livestock inventory data and the LANDSAT images were two weeks rather than the two months. A shorter time interval should be an important objective in future work.

The best relationships which might be expected, if there were no difference in time between the livestock inventories and crop acres, are shown in tables 3 and 4. These relationships would be equivalent to those found if there were perfect classification of the pixels. In general, the correlation coefficients squared are not too different from those in tables 1 and 2 if soybeans is excluded from the cattle results. This suggests that the difference in time was not an important factor for these data. The potential gains appear to be greater for cattle than hogs based on the multiple linear model in tables 3 and 4. The inclusion of soybeans as a variable improves the correlation for cattle by indicating an absence of cattle.

Because only a limited number of variables were examined, further study of the selection of variables (or classification of pixels by crop or cover types) should be investigated in other States. The following are types of variables which would be derived from classified pixels that merit study.

For cattle:

(1) Pasture (or grazing cover) pixels

(2) Oat pixels, or the combining of classified pixels for several feed grains

(3) Hay or alfalfa (or combined) pixels

(4) Pixels for a specialty or cash crop, such as soybeans or cotton, which might indicate the absence of cattle

(5) Possibly pixels for the total area of the unit if not fixed by sample design

For hogs:

(1) Corn (or primary feed grain) pixels

(2) Small grain pixels, or secondary feed grain, such as oats, or the pixels of several grains combined

(3) Pixels for a specialty or cash crop, such as soybeans or cotton, which might indicate the absence of hogs

(4) Possibly pixels for the total area of the unit if not fixed by sample design

The choice of actual variables (or specific crops) would depend on the farm or ranch practices in the State or portion of the State being estimated for using the LANDSAT frame(s).

5. SOME BENEFITS TO AGRICULTURAL STATISTICS USING LANDSAT

The gains of information in estimating acreages have varied from as high as 7-10 for major crops by individual LANDSAT frames which usually translate into information gain factors of about 2-3 at a State level.[1,3,4] The use of spectral data with ground data for yield estimation suggests gains of the order of 1.3 to 1.5. These gains when coupled with the possible gains for livestock of 1.2-1.4 indicate cumulative gains at a State level perhaps as high as 5.0. Based on costs derived earlier for Illinois,[3] as well as Iowa, additional resources of approximately $75,000 would be needed using the current area sampling system, to reduce sampling errors for the production of several major crops by one-half and livestock inventories by one-third. The additional costs of employing the remote sensing in conjunction with current surveys for a State are presently estimated at approximately $225,000 (about 50 percent for data processing and 50 percent for personnel) for annual statistics. That is, the improved precision at the State level could probably be obtained by increasing the area sample for about one-third the cost of using remote sensing. However, the benefits using remote sensing are largely in terms of greater precision for a relatively few crops and livestock species and do not necessarily translate into lower total costs for the area sample because many crop, livestock, and farm statistics cannot be obtained by remote sensing. Consequently, nearly the same ground collection resources will be needed to collect agricultural statistics for a multiplicity of items. The principal benefits from remote sensing for current agricultural statistics are smaller sampling errors obtainable as an alternative to increasing the ground enumeration of sample units (i.e., fewer farmers need to report) and more geographic detail for smaller political or economic units obtainable from the classified pixels. While the benefits from smaller sampling errors can be obtained more economically from ground surveys, the economic value of greater geographic detail is difficult to assess because such information is generally not being provided on an annual basis. However, the value of the greater geographic detail is probably potentially much greater than the benefits due to reduced sampling errors. Consequently, the increased costs ($225,000 - $75,000) using remote sensing to obtain the geographic detail for smaller political or economic units are quite reasonable if the information is needed. In the future, the availability of registered LANDSAT tapes may be expected to reduce these costs.

REFERENCES

1. Wigton, W. H. and H. F. Huddleston, "A Land Use Information System Based on Statistical Inference," Twelfth International Symposium on Remote Sensing of Environment, 1978.

2. Huddleston, H. F. and E. H. Roberts, "Use of Remote Sensing for Livestock Inventories," Fifth International Symposium on Remote Sensing of Environment, 1968.

3. Ray, R. M. III and H. F. Huddleston, "Illinois Crop-Acreage Estimation Experiment," Third Symposium on Machine Processing of Remotely Sensed Data, 1976.

4. Hanuschak, George, et al, "Crop-Area Estimates from Landsat: Transition from Research and Development to Timely Results," 1979 Machine Processing of Remotely Sensed Data Symposium.

5. Gleason, C., et al, "The Auxiliary Use of Landsat Data in Estimating Crop Acreages," Statistical Reporting Service, USDA, 1977.

6. Cochran, William G., "Sampling Techniques," (3rd Ed.) John Wiley and Sons, 1977.

Table 1. Regression Summary for Hogs Based on June Inventories and August LANDSAT Classified Pixels

Model	Scene (Date)		
	1 (8/6)	2 (8/9)	3 (8/9
Best 1 Variable Model	corn pixels	corn pixels	corn pixels
(1) R^2	.161	.045	.046
(2) Intercept (animals)	-105.2	70.85	37.98
(3) Regression coefficient (animals/pixel)	2.446	.900	0.618
(4) Degrees of freedom	49	53	19
(5) F-statistic	9.40	2.49	0.93
All Variables in Model			
(1) R^2	.280	.111	.126
(2) Intercept (animals)	5.475	-10.41	-1.972
(3) Regression coefficients:			
Corn (animals/pixel)	2.601	1.071	0.770
Oats (animals/pixel)	13.899	-2.466	2.422
Pasture (animals/pixel)	-4.137	7.683	-0.369
Woods (animals/pixel)	-2.175	-5.142	-5.031
All other land (animals/pixel)	-2.179	-0.574	-0.083
(4) Degrees of freedom	45	49	15
(5) F-statistic	3.50	1.22	0.43

Table 2. Regression Summary for Cattle Based on June Inventories and August LANDSAT Classified Pixels

Model	Scene (Date)		
	1 (8/6)	2 (8/9)	3 (8/9)
Best 1 Variable Model	pasture pixels	woods pixels	pasture pixels
(1) R^2	.205	.134	.545
(2) Intercept (animals)	44.99	51.523	34.71
(3) Regression coefficient (animals/pixel)	1.673	7.859	1.052
(4) Degrees of freedom	50	53	19
(5) F-statistic	12.64	8.23	22.78
All Variables in Model			
(1) R^2	.277	.210	.662
(2) Intercept (animals)	-74.26	-44.96	34.93
(3) Regression coefficients:			
Corn (animals/pixel)	.283	.358	-0.274
Oats (animals/pixel	1.205	-0.882	-1.621
Pasture (animals/pixel)	1.261	2.904	1.645
Woods (animals/pixel)	.175	5.050	1.979
All other land (animals/pixel)	.929	-0.174	1.303
(4) Degrees of freedom	45	49	15
(5) F-statistics	3.46	2.61	5.88

Table 3. Regression Summary for Hogs Based on June Inventories and June Crop Acres (State of Iowa)

Model	Scene (Date)		
	1 (8/6)	2 (8/9)	3 (8/9)
Best 1 Variable Model	corn acres	total acres	oat acres
(1) R^2	.111	.071	.037
(2) Intercept (animals)	158.549	-212.809	106.251
(3) Regression coefficient (animals/acre)	1.306	0.798	1.393
(4) Degrees of freedom	49	53	19
(5) F-statistic	6.13	4.05	0.74
All Variables in Model			
(1) R^2	0.260	0.109	.250
(2) Intercept (animals)	-296.103	-205.390	18.112
(3) Regression coefficients:			
Corn (animals/acre)	0.375	0.556	2.032
Oats (animals/acre)	2.681	0.763	5.741
Pasture (animals/acre)	-28.441	10.851	17.233
Soybeans (animals/acre)	-1.449	0.012	1.483
Wheat (animals/acre)	-13.139	-10.614	(none)
Total area in segment (animals/acre)	1.570	0.466	-1.472
(4) Degrees of freedom	44	48	15
(5) F-statistic	2.57	0.98	1.00

Table 4. Regression Summary for Cattle Based on June Inventories and June Crop Acres (State of Iowa)

Model	Scene (Date)		
	1 (8/6)	2 (8/9)	3 (8/9)
Best 1 Variable	soybean pixels	pasture pixels	pasture pixels
(1) R^2	.199	.250	0.498
(2) Intercept (animals)	170.317	31.316	65.079
(3) Regression coefficient (animals/acre)	-0.491	18.591	4.737
(4) Degrees of freedom	49	53	19
(5) F-statistic	12.18	17.70	18.87
All Variables in Model			
(1) R^2	.516	.451	.620
(2) Intercept (animals)	-223.635	-86.283	0.244
(3) Regression coefficients:			
Corn (animals/acre)	-0.337	-1.084	-0.627
Oats (animals/acre)	0.512	-1.820	-0.739
Pasture (animals/acre)	-2.273	2.561	-0.071
Soybeans (animals/acre)	-0.651	-1.314	-0.324
Wheat (animals/acre)	-3.968	-0.255	(None)
Total area in segment (animals/acre)	0.813	1.309	0.535
(4) Degrees of freedom	44	48	15
(5) F-statistic	7.80	6.57	4.89