EVALUATION OF THE USDA LARGE AREA CROP ESTIMATION PROCEDURE USING ALTERNATIVE CLUSTERING AND CLASSIFICATION TECHNIQUES

M. L. Amis, R. K. Lennington, M. V. Martin,
W. G. McGuire, S. S. Shen
Lockheed Engineering and Management Services Company, Inc.

1. OBJECTIVES AND EXPERIMENTAL DESIGN

This paper describes the results of the National Aeronautics and Space Administration (NASA) Domestic Crops and Land Cover Classification and Clustering Study on crop area estimation.[1,9] The objective was to evaluate the current crop area estimation approach of the Economics and Statistics Service (ESS) of the United States Department of Agriculture (USDA) in terms of the factors that are likely to influence the bias and variance of the estimator, and to recommend improvements.

The system used by ESS in their crop area estimation is called EDITOR. The EDITOR crop area estimator is a regression estimator. Individual small geographic areas called segments, for which the ground truth crop proportions are known, are clustered and then classified using a Gaussian maximum likelihood classifier. The ground truth hectarages are then regressed onto the number of pixels classified per segment, and regression lines are obtained for each crop. The r^2 of the regression provides a measure of confidence in the estimator obtained. For large area estimation the whole area is classified, and the average number of pixels classified per segment-sized area for each crop is calculated. The corresponding ground truth hectarage from the regression line provides the estimate of the ground truth mean.

In the current ESS estimation procedure, all segments for which ground truth is available are used to train the classifier. Those same segments are then classified and used to obtain the regression estimator. Ideally, independent data sets should be used for developing the regression equation and evaluating area estimates. One way to accomplish this is to divide the available data into training and test portions. Alternatively, quasi-independent segments for regression could be generated using a jackknifing technique. Both methods were employed in this study to better evaluate the performance of the estimator.

The evaluation portion of the study was divided into three levels. The first level consisted of training and testing the regression estimator on all thirty-three segments. This corresponds to the current ESS estimation procedure. The entire estimation process was carried out for both unitemporal and multitemporal data. Summary statistics were collected, and the Hotelling's T^2 test was used to determine if multitemporal data produced significantly better estimates than unitemporal.

In the second level the thirty-three segments were partitioned into a training set and a test set. The classifier was trained on the training set, and then both sets were classified. Summary statistics were collected, and regression equations developed on both sets were compared for homogeneity of variances and equality. These tests indicate whether the regression lines developed on the training set are extendible to the test set.

Jackknifing techniques were used in the third level as a means of obtaining a quasi-independent test set which is larger than that obtainable by using a single partitioning of the data. Summary statistics were collected, and regression estimators obtained from the jackknifed test set were compared to those obtained in the first level of this study.

A second objective was to investigate procedures that would provide improved crop area estimation. The emphasis was on alternative clustering and classification algorithms which would reduce the variance of the resulting regression estimates. Consideration of these alternative methods was principally motivated by two factors. First, it was believed that an improved clustering algorithm for EDITOR would be appropriate. The current EDITOR clustering algorithm is a modified K-means method using Swain-Fu distance as a cluster merge criterion.[2] The CLASSY algorithm had performed well in tests at the Lyndon B. Johnson Space Center (JSC) and was the candidate clustering algorithm replacement. CLASSY is an adaptive maximum likelihood clustering algorithm which models the overall data distribution as a mixture of multivariate normals.[3,4,5,6] A second factor was the belief that the EDITOR procedure should ideally use independent data sets for developing the regression equations and evaluating area estimates. The Mean Square Error (MSE) classifier lends itself well to this use, as it makes no parametric assumptions; thus, there are fewer parameters estimated with the algorithm, implying more stable parameter estimates.[7,8] It was felt that due to this robust nature, the MSE classifier would be more extendible to an independent test set than the Gaussian maximum likelihood classifier currently used. The Mean Square Error Classifier is a nonparametric, least squares classifier that can be weighted through the input of a loss matrix. A quadratic discriminant function was used.

The Landsat data used in this study included 33 segments in northwest Missouri, each having an area of approximately 1 square mile (259 hectares). Data were available for two dates: May 14 and August 3, 1979. Unless explicitly stated, all analysis is with multitemporal data.

The major crops in this study were corn, soybeans, and pasture, which represented about 12, 25, and 30 percent of the crops present in each segment, respectively. Three additional crops were also studied: winter wheat (3 percent), dense woodland (8 percent), and other hay (7 percent). About 15 percent of the segments consisted of other crops, mainly wasteland.

2. TRAINING AND TESTING ON ALL 33 SEGMENTS

The current ESS method of training on a sample and developing the regressions on the training set was performed using all 33 segments. The following comparisons were made:

a. Comparison of unitemporal versus multitemporal -- The entire estimation process was carried out for unitemporal data and for multitemporal data within the current EDITOR system. Summary statistics were collected. The Hotelling's T^2 test was used to determine if multitemporal data produced significantly better estimates than unitemporal.

The computed Hotelling's T^2 was 44.8324 when comparing multitemporal estimates with August estimates (the better of the unitemporal dates). The T^2 statistic at 0.05 level of significance was 17.4. The computed T^2 was greater than $T^2_{0.05}$ (6, 32), and the mean vector of absolute differences between ground truth and estimated hectarage for all six crops was uniformly larger for unitemporal than for multitemporal. Thus it was concluded that the use of multitemporal data over unitemporal significantly improved crop hectarage estimates.

b. Comparison of the current ESS procedure versus the CLASSY clustering algorithm -- The entire estimation process was performed using the current ESS procedure with the CLASSY clustering algorithm substituted into the EDITOR system. Summary statistics were collected. The Hotelling's T^2 test was used to determine if the use of CLASSY produced significantly better estimates on the training set than the current USDA procedure. The results are presented in table 1.

The T^2 calculated on the mean vectors was 44.1959 and the $T^2_{0.05}$ (6, 32) was 17.4. Since the T^2 was greater than the $T^2_{0.05}$ the regression estimates obtained by using CLASSY were closer to the ground truth for all crops than those obtained by using the current procedure, it was concluded that the use of CLASSY did indeed provide improved crop estimates.

c. Comparison of the current ESS procedure versus the MSE classifier -- The entire estimation process was performed using the MSE classifier software. Summary statistics were collected. The Hotelling's T^2 test was used to determine if the use of the MSE classifier produced significantly better estimates on the training set than the current ESS procedure. The results are shown in table 2.

The computed T^2 was 21.777 and the $T^2_{0.05}$ (6, 32) was 17.4 and we concluded that the two procedures do not perform equally. However, since the results were inconsistent across crop types (the MSE classifier provided better results for some crops and worse for others) it cannot be concluded that one classifier performed better than the other.

3. TRAINING ON 25 SEGMENTS AND TESTING ON 8 SEGMENTS

The data were divided into two sets: a training set of 25 and a test set of 8 segments. The classifier developed on the training set was used to classify both the training and test sets. Regressions for the six crops of interest were developed on the training set and also on the test set. This was carried out with the standardized ESS procedure and again with CLASSY as a component of the EDITOR system, and finally with the MSE classifier software. The following tests and comparisons were made:

a. For each of three classification choices, an F-test was performed to determine if the regression line developed on the training set for a given crop was equal to the regression line developed on the test set. (A preliminary test for homogeneity of variance must be carried out first.) This test indicates if the regression developed on the training set is extendible to the test set.

b. The Hotelling's T^2 test was performed to determine if CLASSY produced significantly better estimates on an independent set than the ESS procedure.

c. The Hotelling's T^2 test was performed to determine if the use of the MSE classifier produced significantly better estimates on an independent set than the ESS procedure.

Performance measures for the current procedure are given in table 3.

This table shows that the training set had lower omission and commission errors for each crop than did the test set in general, and the training set yielded higher r^2's. Finally, the overall percent correct was much higher for the training set.

To determine whether the regression lines fitted to the 25 segments in the training set were appropriate for predicting ground truth in the 8 independent test segments, a two-stage F-test was performed for each crop. The structure of this test requires that the residual sum of squares for each regression line be pooled to form a common variance estimate. Thus, homogeneity tests for the error variances of the training and test sets must first be performed.

Results from the F-tests for homogeneity of variances and for equality of regression lines in the current procedure showed that homogeneity of variances was rejected for the three major crops (corn, pasture, soybeans). Of the three remaining crops, the equality of the training and test set regression lines was rejected for the crop other hay. Thus, since regression lines developed on the training set were not appropriate for the test set, the lower r^2's from the test set are a better indication of the reliability of the estimator than r^2's from the training set.

See table 4 for performance measures when using CLASSY in the current procedure.

A Hotelling's T^2 was calculated comparing the performance of CLASSY versus the current procedure on an independent test set. The computed T^2 was 11.035 and the $T^2_{0.05}$ (6, 7) was 405.92.

Table 1.- CLASSY procedure Train and Test

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.9308	72.31	27.69	29.47	23.33
Winter wheat	.4427	38.05	61.95	58.35	22.07
Permanent pasture	.8435	75.45	24.55	45.50	239.79
Soybeans	.8877	81.57	18.43	34.01	85.95
Dense woodland	.7195	49.74	50.26	51.50	62.53
Other hay	.4845	26.14	73.86	63.05	59.45
Overall percent correct = 58.10.					

Table 2.- MSE Classifier Procedure Train and Test on 33 Segments

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.8460	65.61	34.39	24.12	51.90
Winter wheat	.3781	20.13	79.87	40.13	24.63
Permanent pasture	.7643	85.34	14.66	50.01	361.16
Soybeans	.8478	83.48	16.52	35.73	128.02
Dense woodland	.5733	33.98	66.02	47.65	95.15
Other hay	.0005	1.87	98.13	47.06	115.38
Overall percent correct = 57.04					

Table 3.- Current Procedure

(a) Train on 25 Segments					
Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.91	74.24	25.76	31.13	28.805
Winter wheat	.50	35.46	64.54	66.05	21.628
Permanent pasture	.88	66.56	33.44	44.56	176.736
Soybeans	.86	83.69	16.31	31.09	119.426
Dense woodland	.66	54.55	45.45	50.12	72.595
Other hay	.37	32.52	67.48	71.89	79.541
Overall percent correct = 57.70					

(b) Test on an independent set (8 segments)

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.61	54.98	45.02	42.89	202.865
Winter wheat	.00	32.97	67.03	71.15	37.275
Permanent pasture	.39	51.76	48.24	47.87	1268.635
Soybeans	.40	71.74	28.26	63.17	395.029
Dense woodland	.88	27.04	72.96	55.80	36.662
Other hay	.24	39.81	60.19	88.64	52.365
Overall percent correct = 42.00					

Table 4.- CLASSY Procedure

(a) Train on 25 Segments

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.9245	77.12	22.88	30.48	24.06
Winter wheat	.5808	35.73	64.27	67.34	18.22
Permanent pasture	.8437	73.12	26.88	44.66	230.88
Soybeans	.8721	84.15	15.85	31.43	112.10
Dense woodland	.7681	50.19	49.81	42.44	49.45
Other hay	.5760	32.29	67.71	62.13	53.90
Overall percent correct = 59.62					

(b) Test on an indepentdent set (8 segments)

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.3966	55.97	44.03	48.04	313.48
Winter wheat	.3432	41.76	58.24	53.09	24.48
Permanent pasture	.4449	64.20	35.80	48.84	1162.13
Soybeans	.7148	70.43	29.57	59.70	186.40
Dense woodland	.8270	19.15	80.85	55.56	52.15
Other hay	.2081	26.21	73.79	87.32	54.47
Overall percent correct - 45.38					

With the test sample of 8 segments there was no statistical evidence to show any difference between procedures of an independent test set. A larger independent test set would be more appropriate because the critical value T^2 (p,N-1) decreases rapidly as the sample size N increases.

Results from the F-tests for homogeneity of variances and for equality of regression lines using CLASSY showed that homogeneity of variances was rejected for two of the major crops (corn and pasture). Of the 4 remaining crops, the equality of the training and test set regression lines was rejected for dense woodland and other hay. Again, regression lines developed on the training set are not appropriate for the test set, and r^2's from the test set provide more reliable estimator performance measures than those from the training set.

Performance measures when using a simpler classifier in the current procedure are shown in table 5.

When comparing the MSE classifier with the current procedure on an independent set, the T^2 was 25.1924 and the $T^2_{0.05}$ (6, 7) was 405.92. Again, there was not enough evidence to show any improvement.

Corn and pasture failed the homogeneity of variances test. Equality of the regression lines was not rejected for any crop which passed the homogeneity of variances test. In general, the regression lines fitted to the training set were appropriate for the test set.

4. JACKKNIFING WITH THE EDITOR SYSTEM

When it is impossible to have a large training sample as well as a large sample with which to develop the regression lines, a jackknifing procedure can be employed. The jackknifing, which is now described, simulates the method of training a classifier on a sample and then developing a regression on an independent sample.

The 33 segments were grouped into 11 sets containing 3 segments each. One set of 3 segments became the test set, while the remaining 10 sets were pooled and used to train a classifier. The test set containing three segments was then classified. This procedure was repeated 10 more times, with each set of 3 segments being the test set exactly once, and the remaining 30 segments being used to train a classifier. The 11 test sets were then combined, resulting in a sample of 33 segments, each having ground truth (Y) and a classification variable (X).

Regression equations for the six crops of interest were developed on this combined set of 33 segments. The regression MSE'S, r^2's, and classification performance measures are given in table 6 for this combined set. With only one exception, the omission and commission error rates are higher in the jackknifed set than in the set where all 33 segments were used in the training. Also, the r^2's are lower in the jackknifed set. For the major crops of corn, permanent pasture, and soybeans, the decrease in r^2 is 0.15, 0.23, and 0.14, respectively. The

TABLE 5.- MSE Classifier Procedure

(a) Train on 25 segments

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.92	71.36	28.64	24.22	26.37
Winter wheat	.49	16.90	83.10	44.74	22.09
Permanent pasture	.79	87.13	12.87	46.13	316.55
Soybeans	.87	86.02	13.98	32.43	109.97
Dense woodland	.56	32.91	67.09	49.55	93.93
Other hay	.07	2.79	97.21	59.32	111.99
Overall percent correct = 44.69					

(b) Test on an independent set (8 segments)

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.40	54.73	45.27	45.85	313.74
Winter wheat	.05	32.79	67.03	36.17	35.38
Permanent pasture	.39	76.46	23.53	47.64	1269.16
Soybeans	.67	74.78	25.22	59.10	213.21
Dense woodland	.88	18.59	81.41	53.85	37.38
Other hay	.01	2.91	97.09	70.00	68.24
Overall percent correct = 49.24					

Table 6.- Current Procedure - Results for a Jackknifed Test Set of 33 Segments

Crop	r^2	Percent correct	Omission error	Commission error	MSE
Corn	0.75	67.50	32.50	37.51	83.106
Winter wheat	.13	23.19	76.81	74.76	34.538
Permanent pasture	.56	62.75	37.25	51.20	680.577
Soybeans	.71	78.45	21.55	37.26	243.650
Dense woodland	.59	48.24	51.76	59.62	92.173
Other hay	.02	15.48	84.52	80.74	113.273
Overall percent correct = 51.62					

results of this jackknifing study indicate that performance measures for the current procedure are overly optimistic, and that more realistic performance measures are obtained from a separate test set.

Due to the overlap of the training sets, no statistical tests were performed.

5. CONCLUSIONS

Results from the Hotelling's T^2 test showed that the use of multitemporal data over unitemporal significantly improved crop acreage estimates. Performance measures on an independent test set and a jackknifed test set were poorer than those obtained using the current procedure. Performance measures decreased by an average 15% indicating that the current ESS procedure of using a single data set for developing the classifier, performing the regression and evaluating the results leads to overoptimistic performance measures. The CLASSY clustering algorithm, when substituted for the current ESS clustering method, produced significantly improved hectarage estimates when testing and training were done on all 33 segments. The independent test set of eight segments was not large enough to allow the detection of any significant difference between CLASSY and the current ESS procedure; however, the performance measures indicate an improvement when using CLASSY clustering.

It is worthwhile to note that CLASSY requires no decisions from an analyst concerning the number of clusters, separability thresholds, or other arbitrary parameters as does the current clustering method.

The MSE classifier did not produce significantly better hectarage estimates than the standardized USDA procedure when evaluated on either the training set or the independent test set.

However, this classifier showed less sensitivity to the training/test degradation discussed earlier. Also the overall percent correct on the independent test set decreased least when using the MSE classifier. This greater extendibility might be expected due to the fewer parameters required to be estimated in using this classifier. In addition the classifier requires no analyst interaction, and is efficient with respect to CPU usage.

7. RECOMMENDATIONS

Several recommendations seemed appropriate at the conclusion of this study. First, the use of CLASSY clustering in place of the current EDITOR clustering algorithm was recommended. CLASSY seems to offer a tangible improvement to the current EDITOR system in terms of increased performance and decreased analyst interaction. It was recommended that some form of jackknifing also be implemented to obtain more reliable performance measures.

8. REFERENCES

1. Amis, M. L.; Lennington, R. K.; Martin, M. V.; McGuire, W. G.; and Shen, S. S.: Evaluation of Large Area Crop Estimation Techniques Using Landsat and Ground Derived Data, Lockheed/EMSCO Domestic Crops/Land Cover Publication. LEMSCO-15763, February 1981.
2. Swain, P.: Pattern Recognition: A Basis for Remote Sensing Data Analysis; PURDUE University LARS Information note 111572.
3. Lennington, R. K.; and Malek, H.: The CLASSY Clustering Algorithm - Description, Evaluation, and Comparison with the Iterative Self-Organizing Clustering System (ISOCLS). LEC-11289, March 1978.
4. Lennington, R. K.; and Rassbach, M. E.: CLASSY - An Adaptive Maximum Likelihood Clustering Algorithm. LEC-12145, May 1978, [presented at the Ninth Annual Meeting of the Classification Society (North American Branch), Clemson University (Clemson, South Carolina), May 21-23, 1978].
5. Lennington, R. K.; and Rassbach, M. E.: Mathematical Description and Program Documentation for CLASSY, An Adaptive Maximum Likelihood Clustering Method. LEC-12177 (JSC-14621), April 1979.
6. Lennington, R. K.; and Rassbach, M. E.: CLASSY - An Adaptive Maximum Likelihood Clustering Algorithm. Proceedings of Technical Sessions, Volume II, LACIE Symposium, October 1978, JSC-16015 (Houston, Texas), July 1979.
7. Duda, R. O.; and Hart, P. E.: Pattern Classification and Scene Analysis, John Wiley and Sons (New York), 1973.
8. Thadani, S. G.: A Nonparametric Loss-Optimal Pattern Classification System. LEC-11451 (JSC-13901), April 13, 1978.
9. Poster presentation of results made at Fifteenth International Symposium on Remote Sensing of Environment, Ann Arbor, Mich., May 1981.