Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 21 - 23, 1977

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
77CH1218-7 MPRSD

STRATIFIED ACREAGE ESTIMATES IN THE ILLINOIS CROP-ACREAGE EXPERIMENT

RICHARD SIGMAN, CHAPMAN P. GLEASON, GEORGE A. HANUSCHAK, AND ROBERT R. STARBUCK

U. S. Department of Agriculture, Statistical Reporting Service

I. INTRODUCTION

The approach of the Statistical Reporting Service (SRS) for using LANDSAT remote sensor data is to use it as an auxiliary variable with existing operational ground surveys.[1] SRS objectives have been to investigate the use of LANDSAT data to improve crop-acreage estimates at several levels for which acreage statistics are needed; such as counties, groups of counties such as Crop Reporting Districts (CRD's), and entire states.

To determine the feasibility of these objectives, the Illinois crop-acreage experiment was established in 1975.[2] The experiment employs LANDSAT data for the state of Illinois and data from SRS's June Enumerative Survey (JES) for Illinois. The JES data was collected and edited by the Illinois Cooperative Crop Reporting Service. In addition the JES data was supplemented by monthly-updates conducted throughout the growing season and by low-altitude color-infrared photography for 202 of the 300 JES segments in Illinois.

This paper describes:

1. The statistical methodology for the auxiliary use of LANDSAT data to estimate crop acreages,

2. The procedure for designing the pixel classifier which is required by this methodology, and

3. Results obtained by applying this methodology for three LANDSAT frames in western Illinois.

Software systems have been developed jointly by SRS and the Center for Advanced Computation of the University of Illinois which implement the estimation methodology.[3]

The use of LANDSAT data as an auxiliary variable developed from a realization that using LANDSAT data as a survey variable produces biased estimates. The two major types of bias in using LANDSAT data as a survey variable are:

1. Mensuration biases due to the large pixel size of the LANDSAT data (57m x 79m), and

2. Classifier-related procedural biases due to different discrimination functions (linear or quadratic), training sets, prior probabilities, and classification categories used in the design of the classifier.

II. STATISTICAL THEORY AND METHODOLOGY

A. DIRECT EXPANSION ESTIMATION (GROUND DATA ONLY)

Aerial photography obtained from the Agricultural Stabilization and Conservation Service is photo-interpreted using the percent of cultivated land to define broad land-use strata. For example, the stratum definitions for Illinois are given in Table 1.

Within each stratum, the total area is divided into N_h area frame units. This collection of area frame units for all strata is called an area sampling frame. A simple random sample of n_h units is drawn within each stratum. The Statistical Reporting Service then conducts a survey in late May, known as the June Enumerative Survey (JES). In this general purpose survey, acres devoted to each crop or land use are recorded for each field in the sampled area frame units. Intensive training of field statisticians and inteviewers is conducted providing rigid controls to minimize non-sampling errors[4].

The scope of information collected on this survey is much broader than crop acreage alone. Items estimated from this survey include crop acres by intended utilization, grain storage on farms, livestock inventory by various weight categories, and agricultural labor and farm economic data.

Let h = 1, 2,..., L be the L land-use strata. For a specific crop (corn, for example) the estimate of total crop acreage for all purposes and the estimated variance of the total are as follows:

Let Y = Total corn acres for a state (Illinois, for example).

$\hat{Y}$ = Estimated total of corn acres for a state.

y_{hj} = Total corn acres in j^{th} sample unit in the h^{th} stratum.

Then

$$\hat{Y} = \sum_{h=1}^{L} N_h \left(\sum_{j=1}^{n_h} y_{hj} \right) / n_h \qquad (1)$$

The estimated variance of the total is:

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

Note that we have not yet made use of an auxiliary variable such as classified LANDSAT pixels. The estimator in (1) is commonly called a direct expansion estimate, and we will denote this by $\hat{Y}_{DE}$.

As an example, for the state of Illinois in 1975, the direct expansion estimates were:

Corn $\hat{Y}_{DE}$ = 11,408,070 Acres
Relative Sampling Error = 2.4% = $\sqrt{v(\hat{Y})} / \hat{Y}$

Soybeans $\hat{Y}_{DE}$ = 8,569,209
Relative Sampling Error = 2.9% = $\sqrt{v(\hat{Y})} / \hat{Y}$

B. REGRESSION ESTIMATION (GROUND DATA AND CLASSIFIED LANDSAT DATA)

The regression estimator utilizes both ground data and classified LANDSAT pixels. The estimate of the total Y using this estimator 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 corn acres per sample unit from the ground survey for the h^{th} land-use stratum

$$= \sum_{j=1}^{n_h} y_{hj} / n_h$$

$\hat{b}_h$ = the estimated regression coefficient for the h^{th} land-use stratum when regressing ground-reported acres on classified pixels for the n_h sample units.

$$= \frac{\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)(y_{hj} - \bar{y}_h)}{\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)^2}$$

$\bar{X}_h$ = the average number of pixels of corn per frame unit for <u>all</u> frame units in the h^{th} land-use stratum. Thus <u>whole</u> LANDSAT frames must be classified to calculate $\bar{X}_h$. Note that this is the mean for the population and not the sample.

$$= \sum_{i=1}^{N_h} X_{hi} / N_h$$

X_{hi} = number of pixels classified as corn in the i^{th} area frame unit of the h^{th} strata.

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

$$= \sum_{j=1}^{n_h} x_{hj} / n_h \, .$$

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

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

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

where

r_h^2 = sample coefficient of determination between reported corn acres and classified corn pixels in the h^{th} land-use stratum.

$$= \frac{\left[\sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)(x_{hj} - \bar{x}_h) \right]^2}{\left[\sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2 \right] \left[\sum_{j=1}^{n_h} (x_{hj} - \bar{x}_h)^2 \right]}$$

Note that,

$$v(\hat{Y}_R) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h - 2} (1 - r_h^2) \, v(\hat{Y}) \qquad (2)$$

and so $\lim v(Y_R) = 0$ as $r_h^2 \to 1$ for fixed n_h. Thus a gain in lower variance properties is substantial if the coefficient of determination is large for most strata.

The relative efficiency of the regression estimator compared to the direct expansion estimator will be defined as the ratio of the respective variances:

$$R.E. = v(\hat{Y}_{DE}) / v(\hat{Y}_R) \qquad (3)$$

When LANDSAT passes do not cover the entire state on one date, it is necessary to work with analysis districts (domains) which are wholly contained within a LANDSAT scene or pass. In this study the analysis districts were collections of counties wholly contained in a LANDSAT pass. The regression estimate for the i^{th} analysis district is

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

and the entire-state estimate is

$$\hat{Y}_R = \sum_{h=1}^{L_i} N_{hi}\ \bar{y}_{hi(reg)}.$$

When analysis districts are used, degrees of freedom for least squares regression by strata can become small. Under these circumstances it is necessary to pool strata, and the regression estimate for the i^{th} analysis district becomes:

$$\bar{y}^*_{ki(reg)} = \bar{y}^*_{ki} + \hat{b}^*_{ki}\ (\bar{X}^*_{ki} - \bar{x}^*_{ki})$$

for $k = 1, 2, \ldots, L^*_i$, and the entire-state estimate becomes

$$\ddot{Y}_R = \sum_{k=1}^{L^*_i} N^*_{ki}\ \bar{y}^*_{ki(reg)},$$

where L^* = total number of pooled strata for the i^{th} analysis domain and N^*_{ki}, X^*_{ki}, x^*_{ki}, y^*_{ki} are adjusted for varying sizes of the sample units in each stratum. (Thus, h indexes individual stratum; whereas, k indexes pooled stratum. Consequently, the * notation is redundant and will not be used in the next section.)

C. COUNTY ESTIMATES USING A REGRESSION ESTIMATOR

Let $N_{k,c}$ = total number of area frame units in the k^{th} pooled strata for a set of C counties.

$\bar{X}_{k,c}$ = total number of pixels in the set of C counties classified as corn for the k^{th} pooled stratum divided by $N_{k,C}$.

Then an estimate based on the regression estimator of the total corn acreage for the C counties is:

$$\hat{Y}_{REG,c} = \sum_{k=1}^{L} N_{k,c}(\bar{y}_k + \hat{b}_k\ (\bar{X}_{k,c} - \bar{x}_k)) \tag{4}$$

$$v(\hat{Y}_{REG,c}) = \sum_{k=1}^{L} N^2_{k,c}\ \frac{N_k - n_k}{n_k}\ S^2_{k,y}\ \frac{n_k - 1}{n_k - 2}\ .$$

$$(1 - r^2_k)\ (I(C) + \frac{1}{n_k} + \frac{(\bar{X}_{k,c} - \bar{x}_k)^2}{\sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2})$$

where

$I(C)$ = 1 if $O(C)$ < total number of counties wholly contained in the analysis district
= 0 otherwise

$O(C)$ is the cardinality of the set C.

$S^2_{k,y}$ = variance for the corn reported acreage for the k^{th} pooled stratum

$$= \sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2\ /\ (n_k - 1)$$

III. DESIGNING A CLASSIFIER

The pixel classifier is a set of discriminant functions corresponding one-to-one with a set of classification categories. Each discriminant function consists of the category's likelihood probability multiplied by the category's prior probability. If the prior probabilities used are correct for the population of pixels being classified, then the resulting Bayes classifier minimizes the posterior probability of misclassifying a pixel for a 0-1 loss function.[5]

In crop-acreage estimation, however, the objective is to minimize the variance of resulting acreage estimates. Since minimizing the posterior probability of misclassification does not necessarily achieve this objective, optimum acreage estimation may require the use of prior probabilities different than the optimum Bayes set.

For the case of multivariate normal signatures, the category likelihood functions are completely specified by the population means and covariances of the category signatures. Thus, the calculation of category discriminant functions involves the estimation of signature means and covariances and category prior probabilities.

Designing the classifier for this experiment consisted of the following steps:

1. Identification of classification categories.

2. Calculation of signature means and covariances and category prior probabilities from a training set of labeled pixels (called "training the classifier").

3. Measurement of classifier performance on a test set of labeled pixels (called "testing the classifier").

4. Heuristic optimization of the classifier by repeating steps 1 through 3 for different numbers of categories and/or different prior probabilities, and then proceeding to step 5 for the "optimized" classifier.

5. Estimation of classifier performance in classifying the entire pixel population.

Because of the availability of ground data, which supplied the location and cover type of agricultural fields, supervised identification of classification categories was possible. A classification category was created for each cover type in which the number of training pixels exceeded a specified threshold, usually 100 pixels. In addition, a classification category for surface water was created using pixels from rivers, lakes, and ponds.

A classifier was heuristically optimized through a series of classification trials using field-interior pixels to train and all segment-interior pixels to test. The various trials used different combinations of the number of categories and the method of computing prior probabilities. These classification trials, along with additional details on the classifier design procedure, are described in the next section.

IV. ANALYSIS RESULTS FOR WESTERN ILLINOIS

The purpose of the Illinois crop-acreage experiment is to investigate the effectiveness of LANDSAT data to serve as an auxiliary variable for crop acreage estimates. In the analysis of the LANDSAT pass covering western Illinois, referred to simply as the "Western Pass", this investigation had three major objectives. These were:

1. To investigate the influence or lack of influence of various factors, both methodological and geographical, on classifier performance.

2. To compute LANDSAT-based regression estimates for crop acreages in all counties wholly contained in the Western Pass and for the Western Crop Reporting District (CRD) and then compare the precisions of these estimates to JES direct expansion estimates for these areas.

3. To compute crop-acreage regression estimates plus the relative sampling errors of these estimates for the twenty-nine individual counties wholly contained within the Western Pass.

A. CLASSIFIER PERFORMANCE STUDY

The following factors were investigated for their influence or lack of influence on classifier performance:

1. Scene Domain. The northwest Illinois LANDSAT scene, denoted W1 (scene 2194-16035, August 4, 1975), and the west-central scene, denoted W2 (scene 2194-16042, August 4, 1975) were first analyzed separately and then collectively within the Western Pass joined-scene, denoted W123. The southern scene denoted W3 was not analyzed individually since only four segments were on this scene.

2. Number of Classification Categories. This factor investigated the influence of intra-crop clustering to create multiple categories per crop (MCPC) versus straight supervised clustering with a single category per crop (SCPC). The SCPC set of categories consisted of seven categories for W2 and ten categories for W1 and W123. The MCPC set of categories consisted of fifteen categories and was developed by clustering the ten-category SCPC set of covers. This resulted in three categories for alfalfa--cut, uncut, and dried; two categories for hay; and two categories for oat stubble.

3. Prior Probabilities. This factor investigated the effect on classifier performance of using "different prior probabilities" for the classification categories. Strictly speaking, there is only one correct set of prior probabilities for a given geographical region. Using "different prior probabilities" actually means using different weighting factors for the likelihood probabilities in the class discriminant functions. The two sets of prior probabilities which were studied were using priors proportional to expanded reported acres, denoted PER, and using equal priors, denoted EP.

4. Training/test data sets. This factor investigated the data sets on which the classifier was trained and tested. The following methods were employed to allocate the LANDSAT data associated with JES segments between the training and test data sets:

a. Resubstitution, in which all of the segment data, denoted NB for "not background", was used to both train and test the classifier.

b. Sample partition, in which the classifier was trained on a 50% sample of segment fields, denoted FLDS, and then tested on all of the segment data.

c. Jackknifing, denoted JK, in which the training set was 3/4 of the data and the test set was the remaining 1/4. This allocation was repeated four times so that the union of the four test sets was the entire collection of segment data.

The jackknifing technique used was that referred to by Toussaint as the Pi-method.[6] Thus, four separate estimates of classifier performance were obtained and then averaged to yield the jackknife estimate.

There are two reasons why the training/test factor was of interest. The first reason was the desire to minimize the work involved with evaluating a classifier. The resubstitution and sample partition methods are easy to perform but are known to produce biased evaluations of the classifier in small samples. On the other hand, the jackknife is known to give a less biased evaluation but also involves substantially more work to perform. Consequently, if in this investigation the three methods give similar results, then in future experiments of the same size or larger the much-easier-to-apply resustitution and sample partition methods will be compared. If there is no difference between the resubstitution and sample partition methods then these will be used and jackknifing will not be investigated.

The second reason for investigating this factor was to study the sensitivity of the classifier to the selection of the training data. This was the purpose of performing sample partition and then comparing the results with those from the other two methods of classifier evaluation.

5. Strata poolings. Table 2 shows the distribution of JES segments by stratum for W1, W2, and W123. As can be seen, a number of strata have zero or very few segments in them. Thus, it was necessary to pool a number of strata together and then compute $\bar{y}_{h(reg)}$ on the pooled strata. Three different strata poolings were tried and are denoted by the pooled strata given in Table 2.

The purpose of the classifier performance study was to investigate the influence of the above factors on classifier performance. Traditionally, the performance of a classifier has been measured in terms of its confusion matrix of percents correct and commission error rates. However, if a classifier is being used to estimate crop acreages, then it should be evaluated in terms of how well it does exactly that. Thus, the classification objective is to minimize the variance of the resulting regression estimates, and as shown in equation (2) this is accomplished by maximizing the r_h^2's (r-squares). Hence, to compare classifier performance on the same stratum, the respective r-squares were compared. For multi-strata regions, classifier performances were compared in terms of the relative efficiencies (equation (3)) of the resulting estimates. Two types of relative efficiency were calculated. The first type, denoted RE1, was calculated with respect to the direct expansion estimator which uses the same poolings as the regression estimator. RE1 measures the gain in terms of lower variance, of the regression estimate over the pooled JES direct expansion estimate. Of course this doesn't take into account the strata in the direct expansion estimate. However, a second type of relative efficiency, denoted RE2, was calculated with respect to direct expansion over the 11-12-20-30 pooling. Thus RE2 measures the gain, in terms of increased precision, of the regression estimate over the unpooled JES direct expansion estimate.

Counting the different strata poolings as separate trials, thirty-four separate classification trials were performed in the classification performance study. Even this, however, is far short of the number of trials required for a complete factorial analysis. Nevertheless, the influence of each factor on classifier performance can be determined but only on a subset of the levels of other factors. The factor levels for the different trials are summarized in Table 3.

Table 4 compares the r-squares and percents correct for corn in twenty-seven of the classification trials. The MCPC and JK trials are not included in this table. Items of note in this table are:

a. Percents correct are greater for PER priors than for equal priors, but for r-square the opposite is true.

b. Training on a 50% sample of fields yields r-squares very close to those for training on NB.

c. r-square is very small in stratum 20.

d. The r-squares in W1 are generally larger than the corresponding r-squares in W2. W123 is in-between but closer to W2 than W1.

Table 5 presents the relative efficiencies for corn for the same twenty-seven trials. As expected, RE1 and RE2 have the same rankings across factor levels as noted for r-square in Table 4. An interaction between domain location and the optimum strata pooling can be noted. In W1 and W123 the 11-12-20-30 pooling is optimum for RE2, but in W2 the 10-50 pooling is best.

A possible explanation of the effect of domain location on classifier performance is that scenes W1 and W2 are markedly different agriculturally. These differences are exhibited in Table 6 which indicates the amount of land in W1, W2, and W123 devoted to various levels of agricultural activity.

Tables 7 and 8 present results for soybeans for twenty-seven of the classification trials. Unlike corn, the effect of different priors on the classification results for soybeans is very slight, with PER being slightly better than EP. Again, an interaction between location and the optimum strata pooling for RE2 is exhibited, and the nature of this interaction is different from that observed for corn.

Table 9 presents the results of trial JK in which jackknife training and testing is used. Table 10 compares the results of this trial to the corresponding resubstitution trial (Trial W123.2). The jackknife and resubstitution r-square values are quite similar, the major dissimilarities being for those cover types which have large coefficients of variation and small r-squares in Table 9. This suggests that for sufficiently large sample sizes, the resubstitution method will yield r-square values whose biases are acceptably small.

Table 11 compares MCPC versus SCPC. For corn, MCPC is superior; whereas for soybeans an interaction with type of priors can be noted. For the soybeans EP case, SCPC is better. On the other hand, for soybeans PER the MCPC method is superior.

Finally, Table 12 compares classifier performance for all covers and two different priors. Items of note are the low r-square and RE1 values for minor crops and the fact that no single type of prior probability, neither EP nor PER, is optimum for every cover.

B. Large-area Estimates

The relative efficiencies obtained in the classification trials indicated that the auxiliary use of LANDSAT data can reduce the variance of acreage estimates for corn and soybeans. Consequently, the regression estimates for these crops were calculated for the nine-county Western Crop Reporting District (CRD) and for the entire twenty-nine county region contained in the Western Pass. These large-area estimates were then compared to the corresponding direct expansion estimates and to estimates based on the Illinois State Farm Census.

The Western CRD is completely contained in scene W2 and occupies about half of the W2 land area. Regression estimates for the CRD were calculated by first classifying all pixels in W123 with the classifier from classification trial W123.2; i.e., EP, SCPC with ten crops, and training on NB in W1 + W2. The classification results for only those pixels in the Western CRD were then used with a 10-50 strata pooling to compute the $X_{k,c}$ values for equation (4).

Table 13 compares the regression and direct expansion estimates for corn and soybeans in the Western CRD. For each crop the difference between the regression estimate and the direct expansion estimate is less than the standard error of either estimate. For corn the regression estimate C.V. is 54% of the C.V. for direct expansion. For soybeans, however, the regression estimate C.V. is 81% of the direct expansion C.V. Thus, the gain, in terms of lower variance, of the regression estimator over direct expansion is smaller for soybeans than for corn. One reason for this is the fact that an EP classifier was used. The classification trials indicate that EP is optimal for corn but sub-optimal for soybeans.

Table 13 also compares the direct expansion estimates for the Western CRD with acreage estimates based on the Illinois State Farm Census. For each crop the difference between the two estimates exceeds 1.5 times the standard error of the direct expansion estimate. The two estimates, however, measure different quantities--the direct expansion estimate measures standing acres, whereas the State Farm Census measure acres harvested.

Table 14 lists acreage estimates for the entire twenty-nine county region contained in the Western Pass. These estimates were computed using the same classifier as that used for the Western CRD.

C. County Estimates

Regression estimates for corn and soybeans were calculated for the twenty-nine individual counties in joined-scene W123. These are listed in Table 15 and were also computed with the same classifier as that used for the CRD estimates. With two exceptions the C.V.'s for corn ranged between 15 and 20% on a county-by-county basis in northwest Illinois. The exceptions were Jo Davies county (34% C.V.), which is almost entirely stratum 20, and Peoria county (24% C.V.), which is largely urban.

The high C.V.'s in stratum 20 are to be expected due to the very nature of this stratum. Basically, stratum 20 is a "catch-all" stratum in which areas of highly heterogeneous land use are placed.

In west-central Illinois the C.V.'s for corn ranged as high as 33% on a county-by-county basis. Counties with the largest C.V.'s were located on the Mississippi or Illinois rivers.

The C.V.'s for soybeans were considerably larger than those for corn. One reason for this, as was also the case for the CRD estimates, is that the EP classifier is sub-optimal for soybeans.

V. SUMMARY

In order to investigate the effectiveness of LANDSAT data as an auxiliary variable for crop acreage estimates, three LANDSAT frames from an August 4, 1975 satellite pass over western Illinois were analyzed. It was observed that the pixel classifier used in the crop-acreage methodology was influenced by a number of factors, both methodological and geographical.

Large-area corn and soybean acreage estimates were calculated using LANDSAT data as an auxiliary variable for both a twenty-nine county area and a nine-county Crop Reporting District. Significant increases in precision over ground survey estimates were demonstrated.

It was also shown that small-area crop-acreage estimates for individual counties with measurable precision are technically feasible. However, the large coefficients of variation of some of these estimates may make them unsuitable for operational publications.

ACKNOWLEDGEMENTS

The authors wish to acknowledge several individuals for their contribution to this paper and project: Martin Ozga and Walt Donovan under the direction of Robert M. Ray provided heroic programming and systems development in support of our efforts, Richard D. Allen of the Illinois Cooperative Crop Reporting Service for his management of the ground data collection in Illinois, Harold F. Huddleston for his statistical leadership and guidance, and William H. Wigton and fellow members of the New Techniques Section for their efforts and support in this large scale project.

REFERENCES

1. Von Steen, Donald H. and Wigton, William H., "Crop Identification and Acreage Measurement Utilizing LANDSAT Imagery", Statistical Reporting Service, U. S. Department of Agriculture, Washington, D.C., March 1976.

2. Ray, Robert M., III and Huddleston, Harold F., "Illinois Crop-Acreage Estimation Experiment", Proceedings of the 1976 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

3. Ozga, Martin; Donovan, Walter E.; and Gleason, Chapman P., "An Interactive System for Agricultural Acreage Estimates Using LANDSAT Data", Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.

4. 1975 June and December Enumerative Supervising and Editing Manual, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

5. Anderson, T.W., An Introduction to Multivariate Statistical Analysis, New York: Wiley, 1958.

6. Toussaint, Godfried T., "Bibliography on Estimation of Misclassification", IEEE Transactions on Information Theory, Vol. IT-20, No. 4, (July, 1974), pages 472-79.

Table 1. Stratum numbers and definitions

stratum #	stratum description	sub-stratum #	sub-stratum description	
10	intensive agriculture	11	75%+ cultivated	
		12	50% - 75% cultivated	
50	non-intensive agriculture	20	15% - 49% cultivated	
		31	urban	non-agricultural (# 30)
		32	urban	non-agricultural (# 30)
		33	urban	non-agricultural (# 30)
		40	range land	non-agricultural (# 30)
		61	proposed water	non-agricultural (# 30)
		62	water	non-agricultural (# 30)

Table 2. Sample Sizes within Strata and Strata Poolings

original stratum #	# segments* W1	W2	W123	pooled stratum # 0	10-50	11-12-20-30
11	30	16	44	0	10	11
12	6	10	16	0	10	12
20	5	11	17	0	50	20
31	2	1	3	0	50	30
32	1	0	1	0	50	30
33	0	0	0	0	50	30
40	0	1	1	0	50	30
61	0	1	1	0	50	30

*W1 and W2 entries are on an entire scene basis. W123 entries are for the counties wholly contained in W1+W2+W3.

Table 3. Summary of Classifier Performance Study

trial	factor: domain W1	W2	W123	categories SCPC	MCPC	priors EP	PER	train/test NB	FLDS	JK	strata poolings
W1.1	X			X/10			X	X			all 3 poolings
W1.2	X			X/10			X		X		
W1.3	X			X/10		X		X			
W1.4	X			X/10		X			X		
W2.1		X		X/7		X			X		
W2.2		X		X/7			X	X			
W2.3		X		X/7		X		X			
W123.1			X	X/10			X	X			
W123.2			X	X/10		X		X			
W123.3			X		X/15	X			X		all 3 poolings
W123.4			X		X/15	X		X			
JK			X	X/10		X				X	pooling 0

Table 4. Sample coefficients of determination (r-squares) and percents correct for corn in SCPC classifications

analysis district	train/ test	priors	stratum r-square 0	10-50: 10	10-50: 50	11-12-20-30: 11	11-12-20-30: 12	11-12-20-30: 20	11-12-20-30: 30	% correct*
W1	NB	EP	.83	.80	.36	.86	.62	.09	1.00	54
		PER	.64	.56	.50	.65	.60	.06	.95	88
	FLDS	EP	.84	.82	.31	.89	.57	.15	1.00	57
		PER	.70	.62	.51	.72	.56	.07	.97	84
W2	NB	EP	.63	.66	.19	.66	.71	.06	.28	51
		PER	.41	.55	.15	.72	.48	.25	.00	85
	FLDS	EP	.69	.74	.30	.82	.58	.12	.53	54
W123	NB	EP	.70	.72	.21	.78	.54	.00	.58	52
		PER	.52	.56	.18	.67	.57	.00	.20	86

*Based on all segment interior pixels, including field boundaries.

Table 5. Relative efficiencies for corn in SCPC classifications

analysis district	train/ test	priors	RE1 pooling 0	RE1 pooling 10-50	RE2 0	RE2 pooling 10-50	RE2 pooling 11-12-20-30
W1	NB	EP	5.69	3.95	3.03	3.78	4.25
		PER	2.74	2.15	1.46	2.06	2.46
	FLDS	EP	5.97	4.20	3.18	4.02	4.58
		PER	3.26	2.44	1.74	2.33	2.77
W2	NB	EP	2.66	1.68	1.61	1.76	1.27
		PER	1.65	1.47	1.00	1.54	1.15
	FLDS	EP	3.16	2.03	1.91	2.13	1.67
W123	NB	EP	3.34	2.23	1.73	2.00	2.23
		PER	2.08	1.74	1.07	1.56	1.81

Table 6. Distribution of population segments by stratum within analysis districts

stratum	% of population segments in analysis district contained in each stratum: W1	W2	W123
11	53.7	32.5	39.8
12	13.0	16.6	15.7
20	10.9	30.8	23.4
31	11.4	8.6	9.7
32	9.4	5.5	7.2
33	1.0	1.8	1.4
40	.5	3.1	2.0
61	.2	1.1	.8
	100.0	100.0	100.0

Table 7. Sample coefficients of determination (r-squares) and percents correct for soybeans in SCPC classifications

analysis district	train/ test	priors	stratum r-square 0	10-50: 10	10-50: 50	11-12-20-30: 11	11-12-20-30: 12	11-12-20-30: 20	11-12-20-30: 30	% correct*
W1	NB	EP	.81	.82	.83	.82	.70	.98	.98	72
		PER	.82	.83	.83	.83	.72	.98	.98	74
	FLDS	EP	.81	.82	.84	.82	.75	.99	.98	71
		PER	.82	.82	.84	.82	.72	.97	.98	74
W2	NB	EP	.62	.60	.49	.73	.31	.63	.55	65
		PER	.63	.62	.49	.73	.38	.58	.55	63
	FLDS	EP	.63	.61	.51	.73	.34	.63	.02	65
W123	NB	EP	.67	.69	.49	.77	.44	.57	.56	63
		PER	.74	.74	.50	.78	.62	.55	.66	67

*Based on all segment interior pixels, including field boundaries.

Table 8. Relative efficiencies for soybeans in SCPC classifications

analysis district	train/ test	priors	RE1 pooling 0	RE1 pooling 10-50	RE2 pooling 0	RE2 pooling 10-50	RE2 pooling 11-12-20-30
W1	NB	EP	5.25	5.26	4.73	4.81	5.56
		PER	5.42	5.43	4.89	4.97	5.76
	FLDS	EP	5.20	5.25	4.69	4.81	5.62
		PER	5.41	5.42	4.87	4.96	5.74
W2	NB	EP	2.53	2.10	2.26	2.18	1.97
		PER	2.63	2.15	2.34	2.23	1.97
	FLDS	EP	2.60	2.16	1.67	2.13	1.91
W123	NB	EP	2.99	2.56	2.84	2.60	2.52
		PER	3.32	2.78	3.15	2.82	2.91

Table 9. r-squares for jackknifed classification (W123, SCPC, EP, pooling 0)

	pooled-stratum-0 r-square						
	jackknife group						C.V.
cover	1	2	3	4	Ave	S.E.	(%)
Alfalfa	.002	.001	.195	.078	.069	.09	132.7
Corn	.734	.814	.639	.680	.717	.07	10.5
Dense Woods	.097	.003	.030	.213	.086	.09	109.2
Hay	.017	.245	.042	.271	.144	.13	92.2
Oat Stubble	.000	.016	.119	.004	.035	.06	163.9
Oats	.119	.001	.069	.109	.094	.08	87.8
Permanent Pasture	.339	.304	.552	.269	.366	.13	34.8
Soybeans	.578	.745	.843	.520	.671	.15	22.2
Wasteland	.847	.732	.062	.248	.472	.38	79.9

Table 10. Comparison of jackknifed and resubstitution r-squares (W123, SCPC, EP, Pooling 0)

cover	train/test JK	train/test NB
Alfalfa	.069	.09
Corn	.717	.70
Dense Woods	.086	.01
Hay	.144	.25
Oat Stubble	.035	.06
Oats	.094	.15
Permanent Pasture	.366	.36
Soybeans	.671	.67
Wasteland	.472	.81

Table 11. Relative efficiencies for corn and soybeans in W123 classifications

cover	priors	cate- gories	train/ test	RE1 pooling 0	RE1 pooling 10-50	RE2 pooling 0	RE2 pooling 10-50	RE2 pooling 11-12-20-30
Corn	EP	SCPC/10	NB	3.34	2.23	2.00	1.73	2.23
		MCPC/15	FLDS	3.90	2.54	2.02	2.28	2.48
	PER	SCPC/10	NB	2.08	1.74	1.07	1.56	1.81
		MCPC/15	FLDS	2.32	1.86	1.20	1.67	1.91
Soybeans	EP	SCPC/10	NB	2.99	2.56	2.84	2.60	2.52
		MCPC/15	FLDS	2.61	2.29	2.48	2.33	2.31
	PER	SCPC/10	NB	3.32	2.78	3.15	2.82	2.91
		MCPC/15	FLDS	3.39	2.84	3.22	2.89	2.97

Table 12. r-squares and relative efficiencies for all covers (W123, MCPC, FLDS, Pooling 0)

Cover	r-square		RE1	
	EP	PER	EP	PER
Water	.89	.84	8.70	6.23
Waste	.78	.82	4.47	5.45
Soybeans	.62	.71	2.61	3.39
Corn	.75	.57	3.90	2.32
Permanent Pasture	.32	.35	1.44	1.51
Woods	.02	.24	1.01	1.31
Alfalfa	.05	.13	1.04	1.13
Hay	.20	.10	1.24	1.10
Oats	.14	.05	1.15	1.04
Oat Stubble	.01	.03	1.00	1.02

Table 13. Estimated acres of corn and soybeans in the Western CRD

Estimator	Corn		Soybeans	
	Acres	C.V.	Acres	C.V.
Direct Expansion	1,316,000	8.5%	562,000	13.1%
Regression	1,269,000	4.6%	574,100	10.6%
Farm Census	1,121,000		688,700	

Table 14. Estimated acres of corn and soybeans in Western Pass 29-county region

Estimator	Corn		Soybeans	
	Acres	C.V.	Acres	C.V.
Direct Expansion	4,110,150	3.6%	1,539,200	7.7%
Regression	4,125,400	2.5%	1,681,800	5.2%
Farm Census	3,653,800		1,707,400	

Table 15. Regression estimates for corn and soybeans in individual counties in Western Pass

County	Corn		Soybeans	
	Acres	C.V.	Acres	C.V.
Adams	166,600	24.0%	83,600	35.3%
Brown	53,700	33.4	24,300	50.7
Bureau	254,000	18.7	110,600	33.4
Calhoun	56,700	25.1	23,3000	39.9
Carroll	126,500	17.5	57,200	29.6
Cass	91,700	20.3	54,100	25.5
Fulton	172,100	29.0	91,400	37.8
Greene	136,800	19.2	76,000	24.8
Hancock	190,500	19.3	74,800	36.2
Henderson	104,000	17.3	37,100	36.4
Henry	276,800	17.2	79,400	46.6
Jersey	85,700	21.6	48,900	27.0
Jodaviess	108,300	34.1	27,100	94.2
Knox	174,100	19.5	79,600	31.6
Mason	129,100	21.3	76,100	27.9
McDonough	162,500	17.4	82,500	26.3
Mercer	139,800	18.7	43,900	43.4
Morgan	147,200	17.6	93,700	20.9
Ogle	223,000	19.0	51,500	64.2
Peoria	124,000	24.0	65,300	32.6
Pike	160,100	25.7	78,300	37.3
Rock Island	107,000	18.7	27,500	52.7
Schuyler	84,000	29.0	36,650	46.2
Scott	61,100	19.9	31,500	28.6
Stark	92,000	18.2	40,600	32.1
Stephenson	172,100	18.6	30,600	81.8
Warren	161,800	16.5	64,100	32.2
Whiteside	242,800	16.2	62,400	49.0
Winnebago	121,500	21.5	29,600	68.0

RICHARD SIGMAN

B.S., 1971, Mathematics, Colorado State University

M.S., 1976, Mathematics, University of Maryland

Mathematician, 1971-1974, Naval Ship Research and Development Center, Bethesda, Maryland

Mathematical Statistician, 1976-Present, Statistical Reporting Service, U.S.D.A.

GEORGE A. HANUSCHAK

B.A., Mathematics, 1971, Youngstown State University, (Ohio)

M.S., Statistics, 1973, Ohio State University, Columbus, Ohio

Mathematical Statistician with Statistical Reporting Service since April 1, 1973, and has served in the California and Iowa State Statistical Offices prior to assignment in the Research Division.

CHAPMAN P. GLEASON

B.A., 1970, Mathematics, Mansfield State College, Mansfield, Pennsylvania;

M.S., 1972, Mathematics, Ohio State University, Columbus, Ohio;

Mathematical Statistician, 1972-Present, Statistical Reporting Service, USDA, Washington, D.C. 20250

ROBERT R. STARBUCK

B.S., 1971, Mathematics, Miami University, Oxford, Ohio

M.Stat., Statistics, 1972, North Carolina State University, Raleigh, North Carolina

Ph.D., 1975, Statistics, North Carolina State University, Raleigh, North Carolina

Associate Scientist, Northeast Utilities Service Co., Hartford, Connecticut, 1975-1976.

Mathematical Statistician, Statistical Reporting Service, 1976-Present.