AgRISTARS

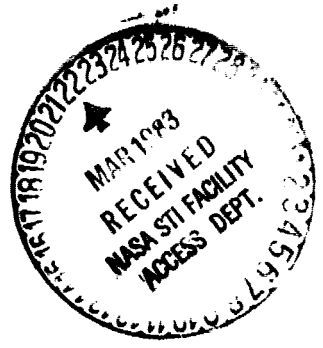
SR-L3-04380
JSC-18891

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

January 1983

ON THE ERROR IN CROP ACREAGE ESTIMATION USING SATELLITE (LANDSAT) DATA



R. Chhikara

Lockheed Engineering and Management Services Company, Inc.

Earth Resources Research Division
Lyndon B. Johnson Space Center
Houston, Texas 77058

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SR-L3-04389, JSC-18591		

4. Title and Subtitle	5. Report Date
On the Error in Crop Acreage Estimation Using Satellite (Landsat) Data	January 1983
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
R. Chhikara Lockheed Engineering and Management Services Company, Inc.	LEMSCO-19021
	10. Work Unit No.

9. Performing Organization Name and Address	11. Contract or Grant No.
Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77258	NAS 9-15800
	13. Type of Report and Period Covered
	Technical Report

12. Sponsoring Agency Name and Address	14. Sponsoring Agency Code
National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Technical Monitor: F. G. Hall	

15. Supplementary Notes

16. Abstract

The problem of crop acreage estimation using satellite data is discussed. Bias and variance of a crop proportion estimate in an area segment obtained from the classification of its multispectral sensor data are derived as functions of the means, variances and covariance of error rates. The linear discriminant analysis and the class proportion estimation for the two-class case have been extended to include a third class of measurement units, where these units are mixed on ground. Special attention is given to the investigation of mislabeling in training samples and its effect on crop proportion estimation. It is shown that the bias and variance of the estimate of a specific crop acreage proportion increase as the disparity in mislabeling rates between two classes increases. Some interaction is shown to take place, causing the bias and the variance to decrease at first and then to increase, as the mixed-unit class varies in size from 0 to 50 percent of the total area segment.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Crop proportion estimate, bias, variance, linear discriminant analysis, error rates, mislabeling, mixed pixels.	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified		

*For sale by the National Technical Information Service, Springfield, Virginia 22161

SR-L3-04389
JSC-18691

ON THE ERROR IN CROP ACREAGE ESTIMATION USING SATELLITE (LANDSAT) DATA

Job Order 71-302

This report describes Error Analysis Research activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

R. S. Chhikara

APPROVED BY

R. K. Lennington, Manager
Supporting Research Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Resources Resarch Division

Space and LIfe Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

January 1983

1. INTRODUCTION

During the last decade, researchers associated with the earth resources program of NASA have been working on the problem of crop acreage and production estimation using LANDSAT data. LANDSAT is a near-earth orbiting satellite equipped with a multispectral scanner (MSS) which measures the reflectance of a target scene in various wavelength bands. The measurement unit is a 1.1 acre square plot of land called a pixel. In estimating acreages devoted to a specific crop of interest in a scene, each pixel is assigned either to the crop or to the class of other ground categories; the pixel classification is based on its spectral response, say a kx1 vector measurement [Heydorn, et al., 1979].

A scene image in the form of a false color composite picture is constructed using its MSS data. Image analysis and pattern recognition techniques are used to correlate the spectral characteristics in the scene to the features on ground. An area segment of several square miles is generally required for an analyst to interpret its scene image and to delineate discernible patterns for identifying possible land-use and land-cover classes. The MSS data for pixels of a class are modeled by a multivariate distribution function. Discriminant analysis techniques are applied to classify the MSS data and to estimate, say, a crop acreage proportion in a segment (Odell, 1976).

As usual, a number of pixels are sampled to estimate the distribution parameters for the distinct classes of pixels and to specify the classification procedure. Sampled pixels are first required to be identified and labeled by their classes on ground. Lack of adequate spectral discrimination between the classes, among others, may cause mislabeling of of some pixels, thus resulting in a biased estimate of the classification

parameters.

Another source of error in estimating a crop acreage proportion is the presence of mixed pixels in a scene. A pixel is defined mixed if it is a boundary unit consisting of areas from more than one category of land use. Otherwise, it is to be called a pure pixel. Often no distinction is made in the handling of mixed and pure pixels in clustering and classification of MSS data for estimating a crop acreage proportion in a segment. Previous empirical studies conducted at the Johnson Space Center have shown that the treatment of mixed pixels as if they are pure causes an additional bias in crop acreage estimation (Carnes and Baird, 1980). A large-scale application of LANDSAT data for wheat estimation in U.S. and U.S.S.R. is described in the Proceedings of Technical Sessions, The LACIE Symposium, NASA (1979).

The problem of estimating the relative acreage of a specific crop in an area segment can be formulated as follows: suppose C_1 and C_0 denote the classes of pixels for the crop of interest and the group of other ground categories, respectively, and C_m denotes the class of mixed pixels in the segment. Considering the segment size to be large, let π_m be the probability of a random pixel to be from C_m and for a spectral measurement $\underline{X}$, let p_i be the conditional probability that it belongs to C_i, i=0, 1, given that the pixel is from either C_1 or C_0 (i.e., it is a pure pixel). Suppose p_{m1} is the proportion of acreages in C_m that are devoted to the crop of interest. Thus if p is the actual acreage proportion for the crop of interest in the segment, then

$$p = \pi_m p_{m1} + (1-\pi_m)\, p_1. \qquad (1.1)$$

Suppose p is estimated by the relative frequency of the segment pixels that are classified into C_1 using a sample-based classification procedure. It is assumed that only pure pixels are sampled for training the classifier and these are subject to mislabeling as C_1 or C_0. Let R_1 and R_0 be the classification regions for C_1 and C_0, respectively. Define the random variable

$$I(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \underset{\sim}{X} \epsilon R_1 \\ 0 & \text{if } \underset{\sim}{X} \epsilon R_0. \end{cases}$$

Then the estimate of p is given by

$$\hat{p} = \frac{1}{N} \sum_{1}^{N} I(\underset{\sim}{X}_i) \tag{1.2}$$

where N is the total number of pixels in the segment (N is assumed to be large). If N_m is the number of mixed pixels, then (1.2) can be written as

$$\hat{p} = \pi_m \hat{p}_{m1} + (1-\pi_m)\hat{p}_1 \tag{1.3}$$

where

$$\pi_m = \frac{N_m}{N}, \qquad \hat{p}_{m1} = \frac{1}{N_m} \sum_{1}^{N_m} I(\underset{\sim}{X}_i)$$

$$\hat{p}_1 = [1/(N-N_m)] \sum_{1}^{N-N_m} I(\underset{\sim}{X}_i). \tag{1.4}$$

In this paper, we investigate $\hat{p}$ for its bias and variance. In section 2, we express these parameters in terms of the expected classification error rates, their variances and covariance, and the first two moments of $\hat{p}_{m1}$. Considering the linear discriminant function for the classification rule, the asymptotic first two moments of error rates and those of $\hat{p}_{m1}$ are obtained

in section 3. Certain numerical results are given in section 4. It is shown that the disparity in mislabeling rates of the two classes C_1 and C_0 has a significant effect on the error rates as well as on the bias and variance of $\hat{p}$. On the other hand, mainly the bias, and not the variance of $\hat{p}$ is affected significantly as the relative size of mixed pixel class, π_m varies.

2. BIAS AND VARIANCE OF $\hat{p}$

Suppose θ_1 is the error of classifying a pixel from C_1 into C_0 and θ_0 is for a pixel from C_0 into C_1. Of course, there is no error committed in the classification of a mixed pixel. Let θ_{m1} be the probability of classifying a mixed pixel into C_1. Suppose the classification rule is determined on the basis of sample means and covariance matrices obtained from a sample of n pure pixels of which n_1 labeled as C_1 and $n_2 = n-n_1$ labeled as C_0. Let $\underset{\sim}{\bar{X}}_1$ and $\underset{\sim}{\bar{X}}_0$ be the sample means, and $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_0$ be the sample covariance matrices for the two groups of labeled samples. Suppose $\hat{\theta}_1$, $\hat{\theta}_0$ and $\hat{\theta}_{m1}$ are the estimates of θ_1, θ_0, and θ_{m1}, respectively, given the sample observations. Then these estimates can be written in terms of the conditional probabilities as follows:

$$\hat{\theta}_1 = P[\underset{\sim}{X}\epsilon R_0 | \underset{\sim}{X}\epsilon C_1 \ , \ \underset{\sim}{\bar{X}}_i, \ \underset{\sim}{S}_i, \ i=0, \ 1]$$

$$\hat{\theta}_0 = P[\underset{\sim}{X}\epsilon R_1 | \underset{\sim}{X}\epsilon C_0 \ , \ \underset{\sim}{\bar{X}}_i, \ \underset{\sim}{S}_i, \ i=0, \ 1]$$

$$\hat{\theta}_{m1} = P[\underset{\sim}{X}\epsilon R_1 | \underset{\sim}{X}\epsilon C_m \ , \ \underset{\sim}{\bar{X}}_i, \ \underset{\sim}{S}_i, \ i=0, \ 1] \qquad (2.1)$$

Bias

For the expected value of $\hat{p}_1$, we have

$$E(\hat{p}_1) = E[E[\hat{p}_1|\underset{\sim}{\bar{X}}_i, \underset{\sim}{S}_i, i=0, 1]]$$

$$= E[P[\underset{\sim}{X} \in R_1|\underset{\sim}{\bar{X}}_i, \underset{\sim}{S}_i, i=0, 1]]$$

Due to (2.1), we can write

$$E(\hat{p}_1) = p_1(1-E(\hat{\theta}_1)) + p_0E(\hat{\theta}_0). \tag{2.2}$$

So the bias of $\hat{p}_1$ given by $E(\hat{p}_1)-p_1$, is

$$B(\hat{p}_1) = -p_1E(\hat{\theta}_1) + p_0E(\hat{\theta}_0). \tag{2.3}$$

As pointed out in the appendix, the expected acreage of C_m devoted to the crop of interest is half of its total size so that $p_{m1}=.5$ and the bias of $\hat{p}_{m1}$,

$$B(\hat{p}_{m1}) = E(\hat{\theta}_{m1}) -.5. \tag{2.4}$$

Accordingly, it follows from (1.3) that the bias of $\hat{p}$,

$$B(\hat{p}) = \pi_m B(\hat{p}_{m1}) + (1-\pi_m)\, B(\hat{p}_1)$$

where $B(\hat{p}_1)$ and $B(\hat{p}_{m1})$ are given by (2.3) and (2.4).

Variance

For the variance of $\hat{p}_1$, we can write

$$\mathrm{Var}(\hat{p}_1) = E[\mathrm{Var}[\hat{p}_1|\underset{\sim}{\bar{X}}_i, \underset{\sim}{S}_i, i=0,1]]$$
$$+ \mathrm{Var}(E[\hat{p}_1|\underset{\sim}{\bar{X}}_i, \underset{\sim}{S}_i, i=0,1])$$

Since the entire segment data are classified to obtain $\hat{p}_1$, the conditional variance of $\hat{p}_1$, given sample data, is zero. Thus, the first term on the right side is zero, and

$$\mathrm{Var}(\hat{p}_1) = \mathrm{Var}(P[\underset{\sim}{X}\epsilon R_1 | \underset{\sim}{\bar{X}}_i, \underset{\sim}{S}_i, i=0,1]).$$

Again, it follows from (2.1) that

$$\mathrm{Var}(\hat{p}_1) = p_1^2 \mathrm{Var}(\hat{\theta}_1) + p_0^2 \mathrm{Var}(\hat{\theta}_0) - 2p_1p_0 \mathrm{Cov}(\hat{\theta}_1,\hat{\theta}_0). \qquad (2.6)$$

Because of (1.3), we have the variance of $\hat{p}$ given by

$$\mathrm{Var}(\hat{p}) = \pi_m^2 \mathrm{Var}(\hat{p}_{m1}) + (1-\pi_m)^2 \mathrm{Var}(\hat{p}_1) + 2\pi_m(1-\pi_m) \mathrm{Cov}(\hat{p}_1\, \hat{p}_{m1}) \qquad (2.7)$$

where $\mathrm{Var}(\hat{p}_1)$ is as given in (2.6), $\mathrm{Var}(\hat{p}_{m1})$ is simply the variance of $\hat{\theta}_{m1}$, and $\mathrm{Cov}(\hat{p}_1, \hat{p}_{m1})$ obtained using the conditional argument, is given by

$$\mathrm{Cov}(\hat{p}_1, \hat{p}_{m1}) = -p_1 \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_{m1}) + p_0 \mathrm{Cov}(\hat{\theta}_0, \hat{\theta}_{m1}) \qquad (2.8)$$

3. LINEAR DISCRIMINANT ANALYSIS

As considered by Heydorn, et al.(1979), we assume that C_0 and C_1 have multivariate normal distributions: $\underset{\sim}{X} \sim N_k(\underset{\sim}{\mu}_i, \underset{\sim}{\Sigma})$ if $\underset{\sim}{X}\epsilon C_i$, $i=0,1$. Without loss of generality, let

$$\underset{\sim}{\mu}_0 = \begin{bmatrix} -\Delta/2 \\ \underset{\sim}{0} \end{bmatrix}, \quad \underset{\sim}{\mu}_1 = \begin{bmatrix} \Delta/2 \\ \underset{\sim}{0} \end{bmatrix}, \quad \underset{\sim}{\Sigma} = \underset{\sim}{I} \qquad (3.1)$$

where

$$\Delta^2 = (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_0)' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_0)$$

Suppose α_i is the probability of mislabeling of a pixel from C_i, i=0,1, for the sample labeling procedure, which is generally manual using visual interpretation of a scene image and knowledge of crop characteristics in the area. The image analyst who makes the labeling decision for sampled pixels uses their spectral information plus his a-priori crop knowledge which is fairly reliable at a somewhat larger level (e.g., crop field) but not at the pixel level. Obviously, labeling of a pixel is partly dependent upon its spectral response and as such it should be taken into account.

In another paper related to this topic, Chhikara and McKeon (1983) have proposed an approach to modeling misallocation, in general, and have discussed analytically the linear discriminant analysis in the presence of misallocation in training samples. In the present context, their model (b) can be considered suitable for the mislabeling of pixels. This is a truncated model for which thresholds are determined for the two classes of pure pixels to assign class labels C_0 and C_1 to sampled pixels. Since the crop information is utilized, these thresholds should be class specific and be functions of mislabeling rates. Given α_0 and α_1, the labeling procedure can be considered as follows:

Suppose X_1 is the component of measurement vector $\underset{\sim}{X}$ in the first dimension along which the class means are aligned. For $\underset{\sim}{X}\varepsilon C_0$, label the pixel as C_0 if $X_1 < -\Delta/2 + Z_{1-2\alpha_0}$ and as C_1 with probability .5, otherwise; and for $\underset{\sim}{X}\varepsilon C_1$, label the pixel as C_1 if $X_1 > \Delta/2 + Z_{2\alpha_1}$ and as C_0 with probability .5, otherwise, where $Z_{1-2\alpha_0}$ and $Z_{2\alpha_1}$ are the $(1-2\alpha_0)$- and $2\alpha_1$- percentage points of the standard normal distribution. Under this rule, a pixel has at most fifty percent chance of misallocation. Of course, one can consider other than half for the

maximum probability of misallocation , say u, where $0<u<1$; but this would require the use of $Z_{1-\alpha_0/u}$ and $Z_{\alpha_1/u}$ for the percentage points so that the mislabeling rates remain as specified.

Chhikara and McKeon (1983) give the mixture distributions of the two classes represented in the labeled training samples and obtain the asymptotic distribution of the sample-based boundary. Their approach is similar to that of Efron (1975) and can be extended to obtain asymptotic first two moments of all three estimators, $\hat{\theta}_1$, $\hat{\theta}_0$ and $\hat{\theta}_{m1}$. As discussed by Efron, the optimum boundary (i.e., the case of known parameters) for the linear discriminant rule is a plane perpendicular to x_1-axis and intersecting it at point τ, whereas when the sample size n is large, the sample-based boundary is a plane intersecting x_1-axis at point $\tau+d\tau$, with normal vector at an angle $d\alpha$ from the x_1-axis, where $d\tau$ and $d\alpha$ represent small deviations. If D_0, D_1 and D_m are the respective distances of the first component means of C_0, C_1 and C_m from the optimum boundary, then their corresponding distances from the sample-based boundary are

$$d_0 = (D_0 + d\tau) \cos d\alpha$$
$$d_1 = (D_1 - d\tau) \cos d\alpha$$
$$d_m = (Dm + d\tau) \cos d\alpha \qquad (3.2)$$

Then, $\hat{\theta}_0 = \Phi(- d_0)$ and $\hat{\theta}_1 = \Phi(-d_1)$. In Appendix we show that when $\Delta < 3.5$, the distribution for C_m can be approximated by normal, with its mean zero and variance in the first dimension, $\sigma_1^2 = 2/3 + \Delta^2/12$. Though it is not discussed, its variance in any other dimension is 1. Thus, the variance along the d_m- direction is $\sigma_1^2 \cos^2 d\alpha + \sin^2 d\alpha$ or $\sigma_1^2 + (1-\sigma_1^2) \sin^2 d\alpha = \sigma_d^2$, say. Accordingly, we have $\hat{\theta}_{m1} = \Phi(-d_m/\sigma_d)$.

By Taylor series expansion, ignoring higher than second order differential terms, it can be shown that

$$\hat{\theta}_0 = \phi(-D_0) - \phi(-D_0)\, d\tau + (1/2)\, D_0\, \phi(-D_0)[(d\tau)^2 + (d\alpha)^2]$$

$$\hat{\theta}_1 = \phi(-D_1) + \phi(-D_0)\, d\tau + (1/2)\, D_1\, \phi(-D_1)[(d\tau)^2 + (d\alpha)^2]$$

$$\hat{\theta}_{m1} = \phi(-D_m/\sigma_1) - \phi(-D_m/\sigma_1)\,(d\tau/\sigma_1) + \tfrac{1}{2}(D_m/\sigma_1)\,\phi(-D_m/\sigma_1)\,[(d\tau/\sigma_1)^2 + (d\alpha/\sigma_1)^2] \qquad (3.3)$$

In (3.3), we have $D_0 = \tau + \Delta/2$, $D_1 = -\tau + \Delta/2$ and $D_m = \tau$ (e.g., refer to the figure given in Efron, 1975). Now by a straight forward extension of the discussion and results of Chhikara and McKeon (1983) the asymptotic first moments of $\hat{\theta}_0$, $\hat{\theta}_1$, and $\hat{\theta}_m$ can be obtained as follows:

$$E(\hat{\theta}_0) = \phi(-D_0) + \tfrac{1}{2n}\, D_0\phi(-D_0)[\sigma_\tau{}^2 + (k-1)\,\sigma_\omega{}^2]$$

$$E(\hat{\theta}_1) = \phi(-D_1) + \tfrac{1}{2n}\, D_1\phi(-D_1)[\sigma_\tau{}^2 + (k-1)\,\sigma_\omega{}^2]$$

$$E(\hat{\theta}_{m1}) = \phi(-D^*) + \tfrac{1}{2n}\, D^*\phi(-D^*)[\sigma_\tau{}^{*2} + (k-1)\,\sigma_\omega{}^{*2}]$$

$$\mathrm{Var}(\hat{\theta}_0) = \tfrac{1}{n}\phi^2(-D_0)[\sigma_\tau{}^2 + \tfrac{1}{2n}D_0^2[\sigma_\tau{}^4 + (k-1)\sigma_\omega{}^4]]$$

$$\mathrm{Var}(\hat{\theta}_1) = \tfrac{1}{n}\phi^2(-D_1)[\sigma_\tau{}^2 + \tfrac{1}{2n}D_1^2[\sigma_\tau{}^4 + (k-1)\sigma_\omega{}^4]]$$

$$\mathrm{Var}(\hat{\theta}_{m1}) = \tfrac{1}{n}\phi^2(-D^*)[\sigma_\tau{}^{*2} + \tfrac{1}{2n}(D^*)^2\,[\sigma_\tau{}^{*4} + (k-1)\sigma_\omega{}^{*4}]]$$

$$\mathrm{Cov}(\hat{\theta}_0, \hat{\theta}_1) = \tfrac{1}{n}\phi(-D_0)\phi(-D_1)[-\sigma_\tau{}^2 + (D_0D_1/2n)\,(\sigma_\tau{}^4 + (k-1)\sigma_\omega{}^4)]$$

$$\mathrm{Cov}(\hat{\theta}_0, \hat{\theta}_{m1}) = \frac{1}{n}\phi(-D^*)\phi(-D_0)\left[(\sigma_\tau^2/\sigma_1) + (D^*D_0/2n\sigma_1^2)\right.$$
$$\left.(\sigma_\tau^4 + (k-1)\sigma_\omega^4)\right]$$

$$\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_{m1}) = \frac{1}{n}\phi(-D^*)\phi(-D_1)\left[-(\sigma_\tau^2/\sigma_1) + (D^*D_1/2n\sigma_1^2)\right.$$
$$\left.(\sigma_\tau^4 + (k-1)\sigma_\omega^4)\right] \qquad (3.4)$$

where

$$D_0 = \tau + \Delta/2 \ , \quad D_1 = -\tau + \Delta/2 \ , \quad D^* = \tau/\sigma_1$$
$$\sigma_\tau^* = \sigma_\tau/\sigma_1 \ , \quad \sigma_\omega^* = \sigma_\omega/\sigma_1 \qquad (3.5)$$

with $\sigma_1^2 = 2/3 + \Delta^2/12$, and σ_τ^2 and σ_ω^2 as the asymptotic variances for the discriminant boundary, which are given in equations (3.10) and (3.11) of Chhikara and McKeon (1983).

4. NUMERICAL RESULTS

In this section we illustrate the bias and variance of $\hat{p}$ numerically by considering k=2, Δ=2, p=.5, and n=100. First, in Table 1, we give values of τ, σ_τ^2 and σ_ω^2 associated with the sample-based discriminant boundary when Δ=2, p_1=.5, .3 and the mislabeling rates, α_1=0 and α_0=0, .1, .2, .3, .4. These values are taken from Table 1 in Chhikara and McKeon (1983) corresponding to their model (b), and are used here to compute asymptotic first two moments of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_{m1}$ as described in (3.4).

It is seen that the discriminant boundary point, τ shifts to the left as α_0 increases. This is expected due to disparity in mislabeling rates for the training samples disfavoring C_0 which is centered to the left on x_1-axis. The variance $\sigma_\tau 2$ increases and $\sigma_\omega 2$ decreases as α_0 increases.

Table 1: Values of τ, σ_τ^2 and σ_ω^2 for the Sample-based Boundary (Δ=2)

(α_0, α_1)	τ		σ_τ^2		σ_ω^2	
	p_1=.5	p_1=.3	p_1=.5	p_1=.3	p_1=.5	p_1=.3
(0,0)	0	.42	1.000	1.360	2.000	2.190
(.1,0)	-.19	.09	1.136	1.308	1.068	.845
(.2,0)	-.40	-.17	1.541	1.717	.747	.488
(.3,0)	-.65	-.44	2.473	2.542	.644	.387
(.4,0)	-1.00	-.82	5.373	5.178	.773	.515

In Table 2, we show the results on mean and variance for different estimators as well as give the bias of $\hat{p}_1$ and that of $\hat{p}_{m1}$ when p_1=.5 and n=100. Based on the values of σ_τ^2 and σ_ω^2 given in Table 1, similar results can be easily computed for the case of p_1=.3. It is seen from these results that $E(\hat{\theta}_0)$ and $Var(\hat{\theta}_0)$ increase, whereas $E(\hat{\theta}_1)$ and $Var(\hat{\theta}_1)$ decrease as α_0 increases, but $E(\hat{\theta}_{m1})$ decreases and $Var(\hat{\theta}_1)$ increases. Again, this can be expected because of a shift to the left in the boundary. $\hat{\theta}_0$ and $\hat{\theta}_1$ are negatively correlated, but $\hat{\theta}_{m1}$ has positive correlation with each of $\hat{\theta}_0$ and $\hat{\theta}_1$. Interestingly, the variance of $\hat{\theta}_{m1}$ is affected only slightly, though it is considerably higher than those of $\hat{\theta}_0$ and $\hat{\theta}_1$ when the disparity between the two mislabeling rates is small. The absolute bias increases for each of $\hat{p}_1$ and $\hat{p}_{m1}$, and so are their variances and covariances.

Next, we combine the two estimators $\hat{p}_1$ and $\hat{p}_{m1}$ by considering the proportion of mixed pixels, π_m = 0, .1, .3, .4, .5, and compute the bias and variance of $\hat{p}$ (Table 3). Both the bias and the variance increase as α_0 increases. But there is an interaction with respect to change in π_m; the absolute bias and variance first decrease then increase as π_m varies from 0 to .5. However, there is only a slight change in the variance due to change in π_m.

Table 2: Means, Variances and Covariances of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_{m1}$, and Biases, Variances and Covariances of $\hat{p}_1$, $\hat{p}_{m1}$.
(Δ=2, p_1=.5, n=100)

Parameter	Mislabeling Rates (α_0,α_1)				
	(0,0)	(.1,0)	(.2,0)	(.3,0)	(.4,0)
$E(\hat{\theta}_0)$	.162	.212	.276	.365	.500
$E(\hat{\theta}_1)$	.162	.119	.084	.052	.026
$E(\hat{\theta}_{m1})$	.500	.422	.346	.256	.151
$Var(\hat{\theta}_0)$	.0006	.0009	.0017	.0035	.0085
$Var(\hat{\theta}_1)$	.0006	.0004	.0004	.0003	.0002
*$Var(\hat{\theta}_{m1})$	.0025	.0017	.0021	.0026	.0032
$Cov(\hat{\theta}_0,\hat{\theta}_1)$	-.0006	-.0006	-.0008	-.0009	-.0012
$Cov(\hat{\theta}_0,\hat{\theta}_{m1})$	.0012	.0013	.0019	.0030	.0052
$Cov(\hat{\theta}_0,\hat{\theta}_{m1})$	.0012	.0009	.0009	.0008	.0007
$B(\hat{p}_1)$	0	.047	.096	.156	.237
$B(\hat{p}_{m1})$	0	-.078	-.154	-.244	-.349
$Var(\hat{p}_1)$	.0006	.0006	.0009	.0014	.0024
$Cov(\hat{p}_1, \hat{p}_{m1})$	0	.0002	.0005	.0011	.0023

*$Var(\hat{p}_{m1}) = Var(\hat{\theta}_{m1})$

Table 3: Bias and Variance of $\hat{p}$ ($p=.5$, $\Delta=2$, $k=2$, $n=100$)

π_m	Mislabeling Rates (α_0, α_1)				
	(0,0)	(.1,0)	(.2,0)	(.3,0)	(.4,0)
		(i) Bias			
0	0	.047	.096	.156	.237
.1	0	.035	.071	.116	.178
.2	0	.022	.046	.076	.120
.3	0	.010	.021	.036	.061
.4	0	-.003	-.004	-.004	.003
.5	0	-.016	-.029	-.044	-.056
		(ii) Variance $x10^{-4}$			
0	6	6	9	14	28
.1	5	5	8	14	27
.2	5	5	8	14	27
.3	5	5	8	14	26
.4	6	5	9	14	26
.5	8	7	10	16	27

5. CONCLUSION

We have investigated theoretically the error in crop acreage estimation using Landsat data and the current methodology of MSS data processing and linear discriminant analysis. Labeling of pixels by an image analyst is modeled and the effect of mislabeling rates on the bias and variance of the crop proportion estimate discussed. In past, investigators have assumed a random model for misallocation in training samples (Lachenbruch, 1966 and McLachlan, 1972), which is not applicable here. Lachenbruch (1974) has discussed two non-random models which are similar to our proposed model. He, however, studied these models only in the context of Fisher linear discriminant function, assuming equal a-priori probabilities and evaluated its performance using a simulation study. Presently no assumption of equal a-priori probabilities is made and the numbers of pixels labeled as C_0 and C_1 are, in fact, treated as random, as one would expect. Only the total sample size n is assumed fixed.

This study extends the usual two-class classification methodology to a third class which presently arises due to mixed pixels in an area segment. Similar situation may also arise in inventorying forest, range, and other land-use and land-cover categories using a fallible measuring device.

Presently we have assumed that the class of mixed pixels is separable from the other two classes, and hence, N_m is known. If N_m is unknown and mixed pixels are delineated using an imperfect boundary detection method, then an estimate of p is obtained by

$$\tilde{p} = \hat{\pi}_m \hat{p}_{m1} + (1 - \hat{\pi}_m) \hat{p}_1$$

where $\hat{\pi}_m$ is an estimate of π_m. Because the number of mixed pixels in a segment is large and because all such pixels are delineated (i.e., no sampling is envolved in the estimation of π_m), the variance of $\hat{\pi}_m$ will be negligible. Thus, we may consider π_m known, assuming that the procedure of delineating mixed pixels is unbiased.

6. ACKNOWLEDGMENT

The research was conducted under NAS/JSC contract, NAS-15800. The author would like to thank Jim McKeon, a graduate student at Old Dominion University for many discussions during the course of this study.

APPENDIX

Distribution of C_m

Let U be the proportion of acreage devoted to the crop of interest in a randomly selected mixed pixel. Then U has the uniform distribution over interval (0, 1). Thus, $p_{m1}=E(U)=.5$.

Dana (1982), and Lambeck and Potter (1979) have shown that the radiance received by Landsat sensor over a target area (resolution element) is almost a linear function of the reflectivity directly transmitted from the target to the sensor. The aerosol optical thickness in atmosphere has a multiplicative effect on the target reflectivity and the convolution of two contrasting surface reflectances for a boundary pixel can well be approximated by their linear combination. So, one may define the spectral measurement of a mixed pixel in a wavelength band by

$$Y = U\,X_0 + (1-U)\,X_1 \qquad (A.1)$$

where X_0 and X_1 are the spectral measurements of two pure pixels representing the two classes, say C_0 and C_1, of the boundary.

In the transformed space, discrimination between C_0 and C_1 is in the first dimension alone. Thus, it is suffice to discuss the distribution of C_m for the univariate case. Suppose X_0 and X_1 are univariate and normally distributed, say $X_0 \sim N(-\theta,\sigma^2)$ and $X_1 \sim N(\theta,\sigma^2)$.

It is easy to see that the conditional distribution of Y given U = u, is normal with mean $(1-2u)\theta$ and variance $[u^2 + (1-u)^2]\sigma^2$. Now writing the joint density, say $f(y,u)$, of Y and U as a product of the conditional density, say $f(y|u)$, and the marginal density of U, and then integrating it with respect to u, the density function of Y can be expressed as follows;

$$f(y) = (1/2\sigma\sqrt{\pi}) \int_{-1}^{1} (1/\sqrt{1+v^2}) \exp\left[-(y-v\theta)^2/2(1+v^2)\sigma^2\right] dv \qquad (A.2)$$

Clearly, the density function $f(y)$ is not of the normal type. To examine its departure from normality, we next obtain its moments and the measures of skewness and kurtosis.

It can be easily verified that

$$\begin{aligned} E(Y) &= 0 \\ E(Y^2) &= \tfrac{2}{3}\sigma^2 + \tfrac{1}{3}\theta^2 \\ E(Y^3) &= 0 \\ E(Y^4) &= \tfrac{1}{5}(7\sigma^4 + 8\sigma^2\theta^2 + \theta^4) \end{aligned} \qquad (A.3)$$

The measure of skewness is zero and the measure of kurtosis is given by

$$\gamma = \frac{9\,(7\sigma^4 + 8\sigma^2\theta^2 + \theta^4)}{5\,(4\sigma^4 + 4\sigma^2\theta^2 + \theta^4)}$$

To evaluate γ, let $\theta = \sigma\Delta/2$, where Δ represents the distance between the distributions of classes being mixed. Then

$$\gamma = \frac{9}{5}\,\frac{(\Delta^4 + 32\Delta^2 + 112)}{(\Delta^4 + 16\Delta^2 + 64)} \qquad (A.5)$$

Thus, for some typical values of Δ, we find γ as shown below.

Δ	0	1	2	3	3.5	4	6	10
γ	3.15	3.22	3.20	3.00	2.87	2.75	2.38	1.87

Since $\gamma = 3$ for a normal distribution, the distribution of the mixed pixel class is not normal. However, its departure from normality is small if $0 \leq \Delta \leq 3.5$. Thus, this distribution can be approximated by normal provided the spectral measurements of classes constituting boundary satisfy the condition of $\Delta \leq 3.5$.

REFERENCES

Carnes, J. G. and Baird, J. E. (1980), "Evaluation of Results of U.S. Corn and Soybeans Exploratory Experiment - Classification Procedures Verification Test" NASA/JSC Technical Report, JSC-16339, Houston, Texas.

Chhikara, Raj S. and McKeon, Jim (1983), "Linear Discriminant Analysis With Misallocation in Training Samples," NASA/JSC Technical Report, JSC-18590, December 1982, Houston, Texas. (submitted for publication)

Dana, Robert W. (1982), "Background Reflectance Effects in Landsat Data," Applied Optics, 21, 4106-4111.

Efron, Bradley (1975), "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis," Journal of the American Statistical Association, 70, 892-898.

Heydorn, R. P., Bizzell, R. M., Quirein, J. A., Abbotteen, K. M. and Sumner, C. A. (1979), "Classification and Mensuration of LACIE Segments," Proceedings of Technical Sessions, The LACIE Symposium, NASA, JSC 16015, Vol. I, 73-86, Houston, Texas.

Lachenbruch, Peter A. (1966), "Discriminant Analysis When the Initial Samples Are Misclassified," Technometrics, 8, 657-662.

_____(1974), "Discriminant Analysis When the Initial Samples are Misclassified II: Non-Random Misclassification Models," Technometrics, 16, 419-424.

Lambeck, P. F. and Potter, J. G. (1979), "Compensation for Atmospheric Effects in Landsat Data," Proceedings of Technical Sessions, The LACIE Symposium, NASA, JSC 16015, Vol. II, 723-738, Houston, Texas.

McLachlan, G. J. (1972), "Asymptotic Results for Discriminant Analysis When the Initial Samples are Misclassified," Technometrics, 14, 415-422.

Odell, Patrick L. (1976) (Ed.) Special Issue on Remote Sensing, Communications in Statistics, A5, 12

The LACIE Symposium, NASA (1979), Proceedings of Technical Sessions, Vol. I & II, JSC 16015, Houston, Texas.