N O T I C E

81--10006

AgRISTARS

SR-L0-00465
JSC-16754

NASA CR-160776

AUG 11 1980

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

Supporting Research

July 1980

PROPORTION ESTIMATION USING PRIOR CLUSTER PURITIES

G. R. Terrell

(E81-10006) PROPORTION ESTIMATION USING PRIOR CLUSTER PURITIES (Lockheed Engineering and Management) 17 p HC A02/MF A01 CSCL 05B

N81-12483

Unclas

G3/43 00006

Lockheed Engineering and Management Services Company, Inc.

Houston, Texas 77058

Lyndon B. Johnson Space Center

Houston, Texas 77058

1. Report No. JSC- 16754; SR-LO-00465	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Proportion Estimation Using Prior Cluster Purities		5. Report Date July 1980
		6. Performing Organization Code
7. Author(s) G. R. Terrell, Lockheed Engineering and Management Services Company ,Inc.		8. Performing Organization Report No. LEMSCO-15163
9. Performing Organization Name and Address Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058		10. Work Unit No.
		11. Contract or Grant No NAS 9-15800
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Tech. Monitor: J. D. Erickson SF3		13. Type of Report and Period Covered Technical Report
		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The prior distribution of CLASSY component purities is studied, and this information incorporated into maximum likelihood crop proportion estimators. The method is tested on Transition Year spring small grain segments.

17. Key Words (Suggested by Author(s)) Labeling techniques; Proportion estimation techniques; Bayesian estimation; Beta distribution	18. Distribution Statement

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 16	22. Price*

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

JSC Form 1424 (Rev Nov 75) NASA — JSC

SR-LO-00465
JSC-16754

PROPORTION ESTIMATION USING PRIOR CLUSTER PURITIES

Job Order 73-306

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

PREPARED BY

G. R. Terrell

APPROVED BY

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright, Manager
Development and Evaluation Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division
Space and Life Sciences Directorate
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

July 1980 LEMSCO-15163

CONTENTS

Section	Page
1. INTRODUCTION	1
2. PURITY OF CLASSY COMPONENTS	2
3. ESTIMATION OF POSTERIOR PURITY	5
4. APPLICATION TO SMALL GRAINS ESTIMATION	9
5. CONCLUSIONS	11
6. REFERENCES	12

TABLES

Table | Page

1. ESTIMATES OF SPRING SMALL GRAINS PROPORTION 8

FIGURES

Figure | Page

1. Estimated Component Purities 3

1. INTRODUCTION

The CLASSY algorithm attempts to decompose multivariate Landsat spectral data as a mixture of multivariate normal distributions. It is hoped that this information can be used to increase the accuracy of the estimates of area proportion of certain crops of interest. In order for this to be true, the mixture components of the CLASSY decomposition should each represent a spectral signature overwhelmingly of the crop class of interest. This note will study the purity of CLASSY components, and propose a Bayesian method for using that information to improve maximum likelihood area estimates. The method is then tested on ten LACIE Transition Year segments with the classifier trained by Analyst Interpreter labels.

2. PURITY OF CLASSY COMPONENTS

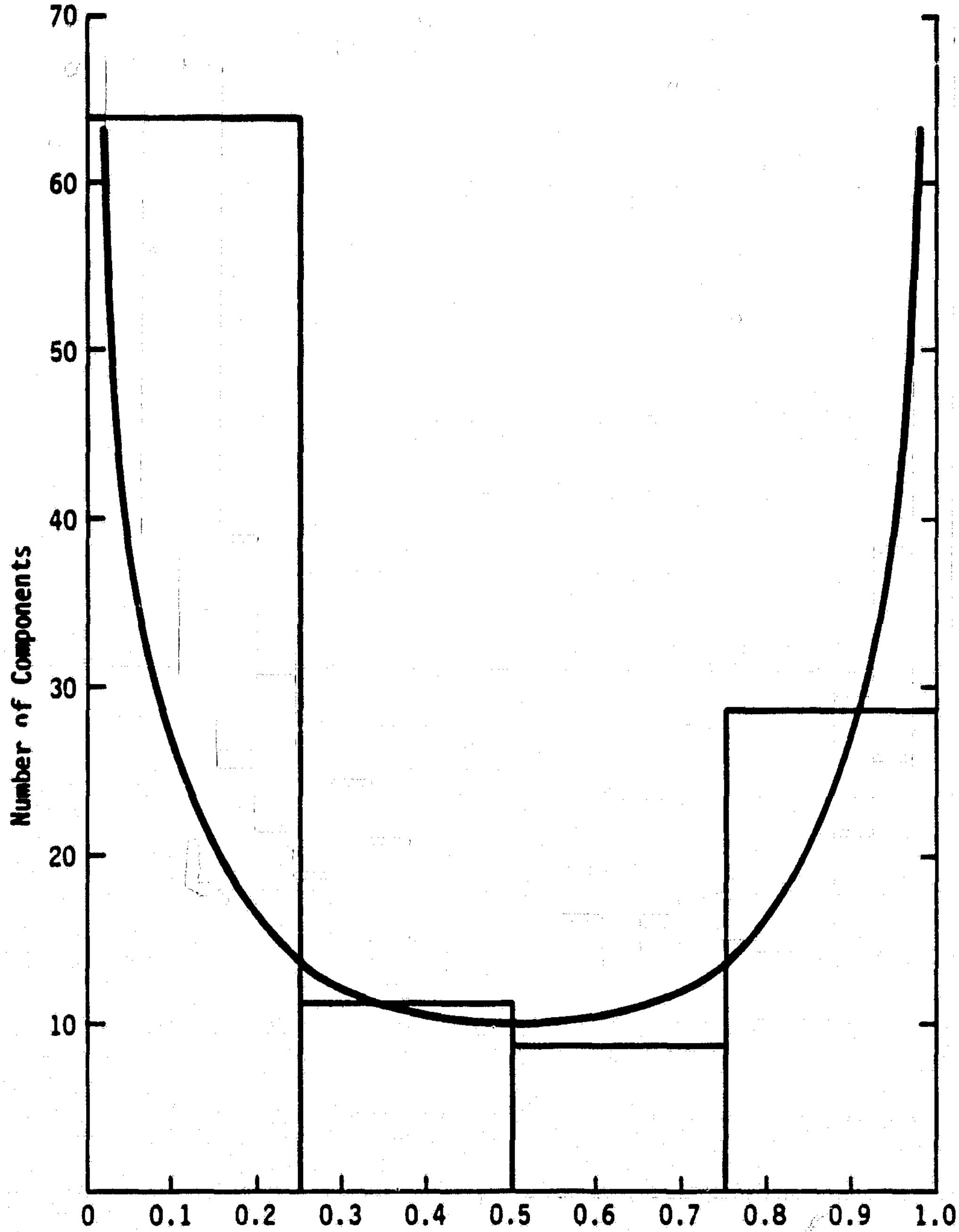
The CLASSY algorithm (Lennington and Rassbach [1979]) estimates a mixture decomposition of continuous multivariate data into multivariate normal components. When the data consists of spectral values from LACIE segments, possibly from several different acquisitions at widely spaced times, it is hoped that CLASSY decomposition may aid in the estimation of proportions of various crops present. If picture elements characteristic of a particular component represent only a random assortment of crop types, clearly nothing has been gained. The hope would be that in many cases a CLASSY component would represent the signature of a particular crop, or a special case of such a signature; therefore, picture elements characteristic of that component would overwhelmingly belong to the crop of interest.

This latter case is that of high component purity. We will measure purity by an index β, $\beta = 0$ means that none of the component is planted in the crop of interest, $\beta = .4$ means that 40% of the component is in the crop, and $\beta = 1.0$ means that all of the component corresponds to the crop. CLASSY is most useful to us when the β's tend to be near zero and one.

The β's may be estimated by the method of maximum likelihood (Lennington and Terrell [1980]). The purity for small grains of 113 CLASSY components of eleven Year Three LACIE segments were estimated using the MAXLABEL program; the estimator was trained using ground truth crop type for approximately 100 pixels per segment. Figure 1 shows a histogram of the β's obtained. It is encouraging that a substantial majority are near zero or one.

In the sequel it will be useful to have a mathematical description of the statistical distribution of the component purities. In the case of two crop types, a Beta distribution with density

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \text{ on } (0,1)$$



n = 113
X = 0.358
S = 0.389

SMALL GRAINS PURITY
FITTED CURVE IS B(0.187, 0.335)

Figure 1.- Estimated component purities.

and zero elsewhere would be convenient. Since the mean $\mu_{a,b} = \frac{a}{a + b}$ and variance $\sigma^2_{a,b} = \frac{a^b}{(a + b)^2(a + b + 1)}$, we may estimate a and b from the sample mean and standard deviation. In our data $\bar{X}_\beta = .358$, $S_\beta = .389$. Now

$$a + b = \frac{\mu(1 - \mu)}{\sigma^2} - 1$$

so that $a = (a + b)\mu$ $b = (a + b)(1 - \mu)$. Substituting population values we got

$$\hat{a} = .187 \ \hat{b} = .335$$

The continuous curve in figure 1 is sufficiently close to the observed distribution to reassure us of the plausibility of the assumption of a Beta distribution (It is simply the density $\beta_{.187,.335}$ rescaled for the histogram).

The fact that $\hat{a}$ is smaller than $\hat{b}$ is related both to the greater purity of non-small-grains components and the fact that small grains elements were in the minority in our sample segments. When a and b were estimated for another population of Transition Year segments, but the estimators were trained with Analyst Interpreter dots, $\hat{a} = .155$, $\hat{b} = .441$ were obtained. These are quite close in practice to the previous results; they seem to reflect in addition the tendency of Analyst Interpreters to make more errors of omission than of commission.

3. ESTIMATION OF POSTERIOR PURITY

We now proceed to utilize prior information to estimate small grains proportion. Two approaches to area estimation have been proposed: If α_j is the prior probability of component j, then a proportion estimate is $\sum_P \alpha_j$ where $j \in P$ if and only if $\beta_j > 1/2$. Alternatively, a proportion estimate is $\Sigma \beta_j \alpha_j$ over all components (Lennington and Terrell (1980)). Notice that the two methods are equivalent if all components are pure.

The generalization of the Beta distribution to a number of simultaneously independently distributed purities is the Dirichlet distribution. In other words, the posterior density of the β_{ij} $i = 1, \cdots, c$, $j = 1, \cdots, d$ where c is the number of crops and d is the number of components is

$$K \Pi \beta_{ij}^{e_i} \text{ with the constraint}$$

that

$$\sum_{i=1}^{c} \beta_{ij} = 1, \; 0 < \beta_{ij} < 1 \text{ and } e_i > -1$$

is a measure of typical prior purity of the i^{th} crop, and K is a constant such that the distribution integrates to one.

Let $\vec{X}_{ik_i}$, where $K_i = 1, \cdots, N_i$ be the spectral vectors of the set of training pixels labeled as crop i. Let f_j be the multinormal density of the j^{th} component estimated by CLASSY. We will assume that the probability that a pixel belongs to a particular crop depends only on its component membership, and not on its spectral location within a component distribution. This simply means that each component is homogeneous; if this were not the case, it would cast doubt on the assumption of multinormality. The likelihood of the observations $\vec{X}_{iK_i}$ is then

$$\prod_{i=1}^{c} \prod_{K_i=1}^{N_i} \sum_{j=1}^{d} \text{Prob } \{\vec{X}_{iK_i}, i, j\}$$

$$= \prod_{i=1}^{c} \prod_{K_i=1}^{N_i} \sum_{j=1}^{d} \text{Prob}\,\{\vec{X}_{iK_i},\, i|j\}\ \text{Prob}\,\{j\}$$

$$= \prod_{i=1}^{c} \prod_{K_i=1}^{N_i} \sum_{j=1}^{d} \text{Prob}\,\{\vec{X}_{iK_i},|i,j\}\ \text{Prob}\,\{i|j\}\ \text{Prob}\,\{j\}$$

$$= \prod_{i=1}^{c} \prod_{K_i=1}^{N_i} \sum_{j=1}^{d} \text{Prob}\,\{\vec{X}_{iK_i}|j\}\ \text{Prob}\,\{i|j\}\ \text{Prob}\,\{j\}$$

$$= \prod_{i=1}^{c} \prod_{K_i=1}^{N_i} \sum_{j=1}^{d} f_j(\vec{X}_{iK_i})\ \beta_{ij}\alpha_j$$

Given the prior information represented by the e_i's, we get a posterior likelihood

$$l = K \prod_{i=1}^{c} \left\{ \left(\prod_{j=1}^{d} \beta_{ij}^{e_i} \right) \left(\prod_{K_i=1}^{N_i} \sum_{j=1}^{d} f_j(\vec{X}_{iK_i})\ \beta_{ij}\ \alpha_j \right) \right\}$$

and taking logarithms

$$L = \log l = \log K + \sum_{i=1}^{c} e_i \sum_{j=1}^{d} \log \beta_{ij}$$

$$+ \sum_{i=1}^{c} \sum_{K_i=1}^{N_i} \log \left[\sum_{j=1}^{d} \alpha_j \beta_{ij} f_j(\vec{X}_{iK_i}) \right]$$

We will estimate cluster purities by taking the maximum posterior likelihood under the constraints $\sum_{i=1}^{c} \beta_{ij} = 1,\ j = 1, \cdots, d$, using the method of Lagrange multipliers. The recursive fixed-point solution was given by

$$\beta_{ij} = \frac{S_{ij}}{\sum_{m=1}^{c} S_{mj}}$$

where

$$S_{ij} = e_i + \sum_{K=1}^{N_i} \frac{\alpha_j \beta_{ij} f_j(\vec{X}_{iK_i})}{\sum_{\ell=1}^{d} \alpha_\ell \beta_{i\ell} f_\ell(\vec{X}_{iK_i})}$$

Notice the parallel between this method and the classical maximum likelihood solution in Lennington and Terrell (1980). This solution becomes the classical solution by setting $e_i = 0$ for each i.

The ease with which the Dirichlet prior enters into this solution may be attributed to the fact that it is conjugate to the multinomial distribution, and the β's have some of the character of multinomial probabilities.

TABLE 1.- ESTIMATES OF SPRING SMALL GRAINS PROPORTION

(1) Segment	(2) Ground Truth %	(3) Maxlabel Stratified %	(4) Maxlabel Cluster %	(5) Empirical Prior Stratified %	(6) Empirical Prior Cluster %
1394	35.45	35.64	27.66	34.80	27.66
1457	47.72	31.20	25.71	30.58	25.71
1518	34.16	26.81	20.67	26.00	20.67
1602	30.42	24.41	21.79	24.18	21.79
1619	47.91	38.32	39.19	38.48	39.19
1668	9.49	7.49	6.34	6.34	6.34
1825	26.69	22.75	19.33	22.63	19.33
1909	22.35	10.18	9.34	9.78	9.34
1918	15.02	18.54	18.16	18.29	18.16
Bias %		-6.01	-9.57	-6.60	-9.57
Mean squared error		.00674	.0134	.00746	.0134

4. APPLICATION TO SMALL GRAINS ESTIMATION

The procedure of the last section was incorporated into a FORTRAN program to run on the LARS system at Purdue University by modifying the MAXLABEL program (Horton and Lennington (1980)). Since the e_i's may be negative, it was necessary to introduce the additional constraint that $0 \le \beta_{ij} \le 1$. This was accomplished by setting an S_{ij} equal to zero whenever it becomes negative during the process of iteration. No effect on the average rate of convergence was noted as a result of this change.

Ten Transition Year LACIE segments in which spring small grains were the major crop of interest were chosen because they had four good quality acquisitions spaced over the growing season, they had been handled by the P1 procedure so that approximately two hundred analyst-labeled pixels were available, and estimates of the true proportion of small grains were available from ground truth surveys. The four spectral values at each acquisition were projected onto the Kauth-Thomas (1976) greenness-brightness plane; thus, there were eight components of the spectral vector for each pixel.

The CLASSY program was run for each of the ten segments listed in column one giving a decomposition into mixtures of eight dimensional multinormal components for each segment. The program described above (called PRELABEL) was then run for the segments with several sets of values of the prior purities. The results are summarized in Table 1. For the third and fourth columns, all e_i's were set to zero, giving the equivalent of a maximum likelihood solution. For columns five and six, $e_1 = \hat{b} - 1 = -.665$ and $e_2 = \hat{a} - 1 = -.813$ were used as these are the empirical values found in section II. Column two is the ground truth proportion of spring small grains for each segment. At the foot of each column are the mean error in percent (called bias) and the mean squared error of each estimate. Columns two and four use the estimator $\sum_{j=1}^{d} \alpha_j \beta_{ij}$ and columns three and five use $\sum_{j \in P} \alpha_j$ where $j \in P_i$, if and only if $\beta_{ij} > 1/2$. It is clear that the introduction of prior purities made no difference to the second estimates and very little difference to the first. A further experiment in which the "diffuse" prior $e_i = -1$ for all i was used made even less difference. The results still reflect the analyst tendency to prefer errors of omission to errors of commission.

5. CONCLUSIONS

A notable result of this study was the success in fitting a Beta distribution to maximum likelihood estimates of component purity. The fact that $\hat{a}$ and $\hat{b}$ are substantially less than one indicates that CLASSY components show more than chance tendency to achieve high degrees of purity in crops of interest. This is the bulk of the evidence in existence that CLASSY actually extracts features of importance from Landsat data.

On the other hand, the methods studied for introducing the prior information into the process of estimation make little difference to the results when used with AI labeled samples, and are not recommended for incorporation into practical estimation procedures.

6. REFERENCES

1. Horton, C.; and Lennington, R. K.: (1980) As-Built Design Specifications for CLASSY, an Adaptive Maximum Likelihood Clustering Method. Lockheed Technical Memorandum 14546.

2. Kauth, R. J.; and Thomas, G. S.: (1976) The Tasselled Cap — A Graphic Description of the Spectral Temporal Development of Agricultural Crops as Seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University.

3. Lennington, R. K., and Rassbach, M. E.: (1979) Mathematical Description and Program Documentation for CLASSY, an Adaptive Maximum Likelihood Clustering Method. LEC Technical Memorandum 12177.

4. Lennington, R. K., and Terrell, G. R.: (1980) Evaluating the Use of Analyst Labels in Maximum Likelihood Cluster Proportion Estimation. AgRISTARS Technical Memorandum JSC-16538.