# AgRISTARS

SR-LO-00442
JSC-16358 80-10295

NASA CR-
160729

JUN 26 1980



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

## Supporting Research

May 1980

# AN EXPLORATORY STUDY TO DEVELOP A CLUSTER-BASED AREA ESTIMATION PROCEDURE

R. K. Lennington

(E80-10295) AN EXPLORATORY STUDY TO DEVELOP A CLUSTER-BASED AREA ESTIMATION PROCEDURE (Lockheed Engineering and Management) 102 p HC A06/MF A01 CSCL 02C

N80-30863

Unclas
G3/43 00295

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
1830 NASA Road 1, Houston, Texas 77058

SR-LO-00442
JSC-16358

# AN EXPLORATORY STUDY TO DEVELOP A CLUSTER-BASED AREA ESTIMATION PROCEDURE

Job Order 73-302

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

PREPARED BY

R. K. Lennington

APPROVED BY

J. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright

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

May 1980

LEMSCO-14670

CONTENTS

# 1. INTRODUCTION

Significant research effort has been devoted to the development of an improved crop area estimation procedure. This procedure would be a replacement for Procedure 1, which was used extensively for crop area estimation in the Large Area Crop Inventory Experiment (LACIE) at the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center (NASA/JSC).

In view of the deficiencies of Procedure 1 (ref.), the goal of this research has been to develop a procedure which is efficient in the sense of having a small mean squared error relative to simple random sampling and which, at the same time, uses a minimum number of labeled pixels. These two goals are in a sense complementary. An efficient procedure is one which obtains a specified acceptable variance with a minimum number of labeled or training pixels.

# 2. CLUSTER-BASED PROPORTION ESTIMATION

As a result of evaluations of Procedure 1 by Jess Carnes (ref.), it became clear at the beginning of the development effort that the classification which followed the clustering in Procedure 1 did not significantly improve the stratification of the scene. Thus, from the outset, cluster-based procedures were developed. That is, the candidate procedures were of the stratified sampling variety, where the strata would be obtained by using an unsupervised clustering procedure. This approach had the advantage of eliminating the type 1 dots used for initiating and labeling the clusters in Procedure 1. In addition, stratifying with clusters was expected, on theoretical grounds, to be more efficient than stratifying with the two strata produced by the classifier.

In order to begin development of the procedure, it was necessary to choose an unsupervised clustering algorithm. Three algorithms, the Iterative Self-Organizing Clustering System (ISOCLS), the Texas A&M University-developed program (AMOEBA), and the CLASSY program, were tested by applying them to 21 LACIE Phase III blind sites and evaluating the average purity of the resultant clusters and the theoretical reduction in variance for the stratification. These evaluations were made using the ground-truth label for every pixel in the image. A complete statement of the results of this is found in appendix A. The basic finding was that the average performance for the three algorithms tested was remarkably similar. The only significant difference was in the number of clusters generated. CLASSY generated an average of about 9, AMOEBA had an average of about 17, and ISOCLS had about 37 clusters. It was concluded that the similarity of performance probably indicated that a limit had been reached in the separability of the data. The fact that this parallel performance was obtained with very few clusters was seen as an advantage for the CLASSY and AMOEBA algorithms.

The next stage in the development was to test each of the candidate clustering algorithms in combination with various schemes for forming proportion estimates. Six different proportion estimation techniques were chosen for testing. Three of these were techniques which resulted in the labeling of entire clusters. They may be described as (1) proportional allocation followed by majority-rule labeling, (2) a sequential allocation technique for labeling with a fixed degree of confidence, and (3) a Bayesian sequential technique for labeling with a fixed degree of confidence. Three techniques for stratified proportion estimation using clusters as the strata were also tested. They may be described as (4) proportional allocation followed by stratified proportion estimation, (5) a sequential allocation technique for minimizing the estimated mean squared error of the proportion estimate at each step, and (6) a Bayesian sequential allocation technique for minimizing the estimated mean squared error of the proportion estimate of each step. Each of these techniques is described in detail in appendix A.

The evaluation involved testing each of these six techniques in combination with each of the three clustering algorithms. Each combination of clustering algorithm and estimation technique was used with 100 different psuedorandom allocations of ground-truth-labeled pixels for each segment. Initially, each technique was evaluated using five segments. Promising techniques were subsequently evaluated using all of the 21 Phase III blind sites used in evaluating the clustering algorithms. The results of this study, as presented in appendix A, were that only two of the techniques appeared to perform consistently better than simple random sampling. These were proportional allocation followed by stratified proportion estimation and sequential Bayesian allocation for minimizing the mean squared error of the stratified proportion estimate at each step. The proportional allocation technique had a reduction in mean squared error over simple random sampling of about 0.65 for each of the three clustering algorithms. The Bayesian sequential allocation technique had a reduction in mean squared error of about 0.51 for CLASSY and ISOCLS and about 0.73 for AMOEBA. Because the CLASSY program generated many fewer clusters than ISOCLS, it was possible to estimate the purity of each cluster using a much smaller number of total labeled pixels. Hence, the Bayesian sequential stratified proportion estimate, using CLASSY clusters as the strata, emerged as the best technique of those tested with respect to the goals of this study.

## 3. ANALYST LABELING WITH BAYESIAN SEQUENTIAL TECHNIQUE

Because all of the preliminary testing had been done with ground-truth labeled pixels, it was desirable to test this new Bayesian sequential technique with CLASSY clusters as the strata using analyst-interpreter (AI) labels. This was the focus of the second study, which is reported in appendix B. In this test, each of 10 LACIE Phase III blind sites was evaluated using the Bayesian sequential procedure. A total of 45 AI-labeled dots were allocated to each segment. The result was that the Bayesian sequential procedure performed significantly better than either Procedure 1 or simple random sampling. The reduction in mean squared error adjusted for sample size was approximately 0.51 for the Bayesian sequential technique compared to simple random sampling and approximately 0.29 for the Bayesian sequential technique compared to Procedure 1. In addition, the Bayesian sequential procedure obtained a lower average bias than either simple random sampling or Procedure 1. This led to the investigation of the AI error rate on the sequentially labeled pixels versus the pixels allocated as random samples. In each of the segments tested, the AI error rate for small-grain pixels was lower for the dots allocated using the Bayesian sequential technique. This phenomenon appears to be due to the influence of the prior distribution on cluster purities used in the Bayesian scheme. In effect, the prior distribution considers pure small-grain clusters to be fairly rare. Hence, if they occur, they are sampled more heavily to verify their reliability. Since pure small-grain clusters are more accurately labeled, this reduces the overall AI error rate.

## 4. CONCLUSION

Based on the results of tests using both ground-truth and AI labels, it is the conclusion of these studies that stratified proportion estimation using CLASSY clusters as the strata and Bayesian sequential allocation as the allocation and estimation technique for minimizing the mean squared error of the proportion estimate offers significant advantages over Procedure 1. It is the recommendation of these studies that this new technique be considered as a replacement for Procedure 1 and further tested in a semioperational environment.

## 5. REFERENCE

Carnes, J. G.: Detailed Analysis of CAMS Procedures for Phase III Using Ground Truth Inventories. JSC-14845, LEC-13343, NASA/JSC (Houston), April 1979.

APPENDIX A

CLUSTERING ALGORITHM EVALUATION AND THE
DEVELOPMENT OF A REPLACEMENT FOR
PROCEDURE 1

APPENDIX A

Lockheed Electronics Company, Inc.
A SUBSIDIARY OF LOCKHEED CORPORATION
1830 NASA Road 1, Houston, Texas 77058
Tel. 713-333-5411

JSC-16232

Ref: 643-7863
Job Order 73-302
Contract NAS 9-15800

TECHNICAL MEMORANDOM

# CLUSTERING ALGORITHM EVALUATION AND THE DEVELOPMENT OF A REPLACEMENT FOR PROCEDURE 1

By

R. K. Lennington and J. K. Johnson

Approved By: T. C. Minter
T. C. Minter, Supervisor
Techniques Development Section

November 1979

LEC-13945

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| JSC-16232 | | |

**4. Title and Subtitle**
CLUSTERING ALGORITHM EVALUATION AND THE DEVELOPMENT OF A REPLACEMENT FOR PROCEDURE 1

**5. Report Date**
November 1979

**6. Performing Organization Code**


**7. Author(s)**
R. K. Lennington and J. K. Johnson
Lockheed Electronics Company, Inc.


**8. Performing Organization Report No.**
LEC-13945

**9. Performing Organization Name and Address**
Lockheed Electronics Company, Inc.
Systems and Services Division
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
Technical Monitor: J. D. Erickson/SF3

**13. Type of Report and Period Covered**
Technical Memorandum

**14. Sponsoring Agency Code**

**15. Supplementary Notes**


**16. Abstract**

This study was designed as a response to observed deficiencies in Procedure 1. A more efficient procedure would be to simply cluster the data using a completely unsupervised clustering algorithm and then use labeled pixels to either label the resulting clusters directly or to perform a stratified estimate using the clusters as the strata.

In the new procedure, clustering is the primary machine processing step, and the most efficient clustering algorithm available was needed. Three algorithms, CLASSY, AMOEBA, and Iterative Self-Organizing Clustering System (ISOCLS), were chosen for testing.

An equally important part of defining a new proportion estimation procedure was the selection of a scheme for obtaining a stratified estimate or a method of labeling each cluster. Three stratified estimation schemes and three labeling schemes were considered.

The evaluation and comparison of the algorithms and the six techniques for proportion estimation are documented in this report with recommendations.



**17. Key Words (Suggested by Author(s))**
clustering algorithms
stratified proportion estimation
cluster labeling
CLASSY, AMOEBA, and ISOCLS


**18. Distribution Statement**

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price* |
|---|---|---|---|
| Unclassified | Unclassified | 79 | |

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

JSC Form 1424 (Rev Nov 75) NASA — JSC



A-3

7

CONTENTS

## TABLES



A-6



9

## Figures





10

## ACRONYMS

| | |
|---|---|
| AA | Accuracy Assessment |
| JSC | Lyndon B. Johnson Space Center |
| ISOCLS | Iterative Self-Organizing Clustering System |
| LACIE | Large Area Crop Inventory Experiment |
| LSD | least significant difference |
| NASA | National Aeronautics and Space Administration |
| Pixel | picture element |
| PCC | percent of correct classification |
| R | the variance reduction criterion |

## 1. BACKGROUND AND INTRODUCTION

In performing machine classification of remotely sensed data, clustering has typically been used to analyze and determine the inherent data signatures. In the proportion estimation system developed during the Large Area Crop Inventory Experiment (LACIE) and called Procedure 1, the multispectral land satellite (Landsat) data was first clustered to obtain the spectral signatures. These signatures were then labeled and used to train a maximum likelihood classifier which classified each picture element (pixel) in the image into one of the labeled classes. The final step was to evaluate the performance of this classifier on an independent labeled data set and to use the estimates of the omission and commission errors resulting from this evaluation to correct the bias in the classified data. Procedure 1, thus, required two sets of labeled data. A set of approximately 40 labeled pixels, called type 1 dots, was used to initiate the clustering and to label the resulting clusters. Another set of approximately 60 labeled pixels, called type 2 dots, was used to evaluate the classifier and correct any bias in the overall proportion estimates for the labeled classes.

Within the past year, different investigations have resulted in several important conclusions regarding the Procedure 1 system. One study (ref. 1) concluded that the labeled clusters agreed very closely with corresponding classifier results. This seems to imply that the classification is unnecessary. In a second series of studies (refs. 2 and 3), it was found that the overall variance of the proportion estimates, resulting from Procedure 1, were only smaller by a factor of about 0.7 (on the average) than the proportion estimates resulting from a simple random sample of 60 labeled pixels. The conclusion was that the machine processing, which comprised Procedure 1, was relatively inefficient.

The current study was designed as a response to the observed deficiencies in Procedure 1. It appeared that the classification step was unnecessary and that a more efficient procedure would be to simply cluster the data using a completely unsupervised clustering algorithm and then use any labeled pixels to either label the resulting clusters directly or to perform a stratified estimate using the clusters as the strata. Such an approach would have the advantage of eliminating the need for the type 1 dots as well as the machine classification step.

Since clustering was to be the primary machine processing step in the new procedure, it was important to choose the most efficient clustering algorithm available. Three algorithms were ultimately chosen for testing. These algorithms were:

a. CLASSY (refs. 4, 5, and 6) — an adaptive maximum likelihood algorithm developed at the National Aeronautics and Space Administration (NASA), Lyndon B. Johnson Space Center (JSC)

b. AMOEBA (ref. 7) — an algorithm developed at Texas A&M University, employing both spectral and spatial information

c. The Iterative Self-Organizing Clustering System (ISOCLS), (ref. 8) — a variant of the ISODATA algorithm of Ball and Hall (ref. 9), and the algorithm used in Procedure 1

These algorithms were applied to each of 25 LACIE segments collected during the 1976-77 crop year. The details of the clustering algorithms and the measures used in evaluating the clustering results are discussed in section 2 of this report.

An equally important part of defining a new proportion estimation procedure was the selection of a scheme for obtaining a stratified estimate or a method of labeling each cluster. In this regard, three stratified estimation schemes and three labeling schemes were considered. The details of these schemes are described in section 3. A description of the data set and the experimental design is included in section 4. In section 5 is a summary of the primary results, and section 6 consists of the conclusions drawn from the observed results with appropriate recommendations.



\13

## 2. CLUSTERING ALGORITHMS AND EVALUATION CRITERIA

The clustering evaluation portion of the study consisted of running each of three different clustering algorithms on each of the 25 LACIE segments selected. The clustering algorithms tested were CLASSY, AMOEBA, and ISOCLS.

CLASSY was run using three complete passes through the data where the data set consisted of every other pixel in the image. Clusters smaller than 2 percent of the scene were eliminated.

ISOCLS was run with the standard iterative parameter set recommended by Wylie and Bean (ref. 10) and known as the MPAD cluster parameter set. The values of these parameters are given in table 2-1. The algorithm was started with 40 randomly selected and unlabeled pixels from each image.

AMOEBA was run with parameters specified by its developers at Texas A&M University. The minimum number of clusters was set at five.

Both CLASSY and AMOEBA were run on data which had been transformed to Kauth brightness and greenness coordinates on each pass (ref. 11). This reduced the dimensionality of the data by a factor of 2. ISOCLS was run on the full dimensional data in accordance with the standard practice during LACIE Phase III.

Each of the algorithms tested produced cluster maps which were subsequently compared with digitized ground-truth maps. The ground-truth maps were prepared from ground-truth images having a resolution six times that of Landsat imagery. The higher resolution ground truth was converted to Landsat resolution by applying majority rule to each six-subpixel area corresponding to one Landsat pixel. In the event of ties, the first label to receive the tying number of subpixels was chosen as the Landsat pixel label.

By comparing the digitized ground truth with a cluster image, the proportion of each ground-truth class, making up each cluster, was determined. The proportions for the small-grains classes were then combined to give the proportion

TABLE 2-1.— MPAD CLUSTER PARAMETER SET

| Parameter | Number of channels | | |
|---|---|---|---|
| | 8 | 12 | 16 |
| CLUSTERS | 60.0 | 60.0 | 60.0 |
| THRESHOLD | 8191 | 8191 | 8191 |
| SEP | 1 | 1 | 1 |
| PERCENT | 100 | 90 | 90 |
| STDMAX | 3.6 | 3.6 | 3.6 |
| DLMIN | 3.9 | 4.1 | 4.5 |
| NMIN | 50 | 50 | 50 |
| ISTOP | 8 | 8 | 8 |
| SEQUEN | Split-combine | Split-combine | Split-combine |
| DOTFIL | (a) | (a) | (a) |

[a]Randomly selected starting dots.



15

of small grains ($P_i$) in each cluster. These data were used to calculate two different evaluation criteria for each clustered image. These criteria are called the variance reduction criterion (R) and the percent of correct classification (PCC), using majority rule labeling.

The R criterion represents the ratio of the variance of a proportion estimate based on a stratified random sample allocation (in which strata are the clusters) to the variance of a simple random sample proportion estimate. The equation for this ratio (when samples that are allocated to clusters are proportional to the size of the cluster) follows:

$$R = \frac{\sum_{i=1}^{c} \frac{N_i}{N_T} P_i(1 - P_i)}{P(1 - P)} \qquad (1)$$

where

$c$ = total number of clusters

$N_i$ = total number of pixels in cluster i

$N_T$ = total number of pixels in the segment

$P_i$ = the proportion of small grains in cluster i

$P$ = the overall proportion of small grains in the segment.

The parameters $P_i$ and P were evaluated using the Accuracy Assessment (AA) digitized ground-truth data for each segment.

The PCC criterion measures the proportion of pixels that would be correctly labeled or classified if each cluster were labeled by majority rule. The equation for computing the PCC criterion may be written as follows:

$$PCC = \sum_{P_i \geq 0.5} P_i\left(\frac{N_i}{N_T}\right) + \sum_{P_i \leq 0.5} (1 - P_i)\left(\frac{N_i}{N_T}\right) \qquad (2)$$

where $P_i$, $N_i$, and $N_T$ are defined above. The first term represents the summation over all clusters having $P_i \geq 0.5$. These clusters would be labeled "small



16

grains" by majority rule. The second term represents the summation over all clusters having $P_i \leq 0.5$. These clusters would be labeled "other" by majority rule.

The R criterion serves as a measure of the efficiency of a clustering algorithm as used in a stratified sampling proportion estimation scheme. The PCC criterion, on the other hand, serves as an overall indicator of cluster purity and of the quality of a proportion estimate obtained by labeling clusters.

The results of evaluating these criteria for each of the three clustering algorithms as applied to the 25 LACIE segments are given in section 5.

## 3. TECHNIQUES FOR CLUSTER-BASED PROPORTION ESTIMATION

The objective of performing clustering in the context of Procedure 1 replacement is to use the results of the clustering as a basis for obtaining a proportion estimate for a crop of interest. In this study, six different techniques for obtaining proportion estimates by labeling a subset of pixels from the image were explored. Three of these techniques result in a labeling of each cluster, whereas the other three produce estimates of the proportion of the crop of interest in each cluster. We will refer to the first three techniques as cluster-labeling techniques and the last three as stratified proportion estimation techniques.

The various cluster-labeling techniques differ from one another in the manner in which the subset of pixels to be labeled is selected. In one technique, pixels are allocated to each cluster, proportionally to the size of that cluster; that is, if $n_T$ total pixels are to be labeled, then

$$n_i = \frac{N_i}{N_T} n_T \tag{3}$$

is the number of pixels to be labeled from each cluster. It should be noted that if $n_i$ is not an integer, it is rounded up or down. If this produces a total number of pixels less than n, the remaining pixels are selected first from the largest cluster, then the next largest, continuing in this manner. Clusters too small to receive a single pixel are lumped together, and an allocation is made to that lumped group. Following the pixel allocation, majority rule may be applied to label the cluster; that is, if

$$\hat{P}_i = \frac{x_i}{n_i} \tag{4}$$

where $x_i$ = the number of pixels out of the $n_i$ pixels labeled in cluster i that are the crop of interest.

Then the labeling rule is as follows:

a. Label cluster i as the crop of interest if

$$P_i \geq \frac{1}{2}$$

b. Otherwise, label cluster i as being other than the crop of interest.

The proportion estimate is obtained as

$$\hat{p} = \sum_{P_i \geq \frac{1}{2}} \frac{N_i}{N_T} \tag{5}$$

The procedure just described will be called cluster labeling by proportional allocation.

The other two cluster-labeling procedures tested were developed by M. D. Pore of Lockheed Electronics Company, Inc. (ref. 12). One approach, called cluster labeling by sequential allocation, labels pixels, selected at random, from a given cluster until a confidence interval for the estimated proportion of the crop of interest no longer contains one-half.

The final cluster-labeling approach tested is called cluster labeling by sequential Bayesian allocation. In this approach a Bayesian estimate for $P_i'$, the probability that the true proportion of the crop of interest is less than or equal to one-half is developed. The formal equation is

$$P_i' = \text{Prob}\left[0 \leq \theta_i \leq \frac{1}{2}\right] = \int_0^{1/2} f(\theta_i | x_i) d\theta$$

$$= \frac{1}{f(x_i)} \int_0^{1/2} f(x_i | \theta_i) g(\theta_i) d\theta_i \tag{6}$$

where $\theta_i$ = the true proportion of the crop of interest in cluster i, $g(\theta_i)$ = the unknown prior distribution for the $\theta_i$'s and as before $x_i$ = the number of pixels out of the $n_i$ pixels labelled in cluster i that are the crop of interest.



19

The strategy is to select a form for $g(\theta_i)$ and calculate the form of $P_i'$. Then one may continue sampling at random and labeling the samples selected until $P_i'$ is smaller or larger than a fixed threshold. If $P_i'$ is smaller than $\alpha$, then label cluster i as other than the crop of interest. If $P_i'$ is greater than $1 - \alpha$, then label the cluster as the crop of interest. Thus, in both cluster labeling by sequential allocation and cluster labeling by Bayesian sequential allocation, labeling from a given cluster continues until a specified confidence on the label of that cluster is obtained. The Bayesian scheme uses the additional information of an estimated prior distribution on the true cluster purities produced by a given algorithm. The necessary labeling rules and equations for these two techniques are developed in (ref. 12) and repeated here.

For cluster labeling by sequential allocation, the labeling rule is as follows:

a. Continue labeling if

$$x_i = \left(\frac{x_i}{n} - 1.534\hat{\sigma}_i, \frac{x_i}{n_i} + 1.534\hat{\sigma}_i\right)$$

where

$$\sigma_i = \sqrt{\frac{x_i(n_i - x_i)}{n_i^2(n_i - 1)}}$$

or until 35 samples have been allocated.

b. Otherwise, label by majority rule

This interval provides an approximate confidence of 1 - 1/8 = 0.875 in the label for each cluster.

For cluster labeling by sequential Bayesian allocation, the labeling rule is as follows:

a. Label two pixels from a given cluster. If $x_i = 0$ or 2, stop and label by majority rule. Otherwise, go to step b.

b. Label three more pixels. If $x_i = 1$ or 4, stop and label by majority rule. Otherwise, go to step c.

c. Label two more pixels. If $x_i = 2$ or 5, stop and label by majority rule. Otherwise, go to step d.

d. Label three more pixels. If $x_i = 3$ or 7, stop and label by majority rule. Otherwise, go to step e.

e. Label three more pixels and label the cluster by majority rule.

This labeling rule is derived using a uniform prior for $g(\theta)$ and also provides an approximate probability of correct labeling of $1 - 1/8 = 0.875$.

The three techniques for stratified proportion estimation parallel the three cluster-labeling techniques just discussed. One possibility is to allocate a total of $n_T$ pixels such that each cluster receives an allocation proportional to its size. This proportional allocation is accomplished as described earlier in this section. The proportion estimate is then computed as

$$\hat{p} = \sum_i \left(\frac{N_i}{N_T}\right)\left(\frac{x_i}{n_i}\right) \tag{7}$$

The term $\frac{x_i}{n_i}$ represents an estimate of the proportion of cluster i which is the crop of interest. The remaining two techniques for stratified proportion estimation differ in the rules used for allocating pixels to cluster and in the equation used for obtaining the final estimate. As was the case for cluster labeling, both techniques are sequential in nature with one employing a Bayesian prior distribution. Both techniques were developed by M. D. Pore (ref. 13).

The concept of sequential sampling as it is used in these two techniques is to apply information obtained from previously allocated samples in determining which cluster should receive the new sample. Suppose $n_i$ pixels have been allocated to cluster i, and $x_i$ of these pixels are of the crop of interest. Then

$$\hat{\sigma}_n^2 = \sum_i \left(\frac{N_i}{N_T}\right)^2 \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i - 1} \tag{8}$$

where

$$\hat{p}_i = \frac{x_i}{n_i}$$

is an estimate of the variance of the usual stratified proportion estimator as given in equation (7). Now the estimated expected value of $\hat{\sigma}_n$ is (if one more sample from the $ith$ cluster is taken)

$$\hat{E}\left[\hat{\sigma}^2_{n+1}\right] = \hat{p}_i\sigma^2_{n+1}(x_i + 1) + (1 - \hat{p}_i)\sigma^2_{n+1}(x_i) \tag{9}$$

where $\sigma^2_{n+1}(x_i + 1)$ is the variance based on n + 1 total samples if the last sample selected is from cluster i and is also the crop of interest, and $\sigma^2_{n+1}(x_i)$ is the variance if the last sample selected is from cluster i and is other than the crop of interest.

The expected change in the estimated segment proportion variance due to an additional labeled sample from cluster i is then

$$\Delta\sigma^2_i = \hat{\sigma}^2_n - \hat{E}\left[\hat{\sigma}^2_{n+1}\right] \tag{10}$$

Written in terms of the basic variables this equation becomes

$$\Delta\sigma^2_i = \left(\frac{N_i}{N_T}\right)^2 \frac{n_i + 3}{(n_i - 1)n^2(n_i + 1)^2} x_i(n_i - x_i) \tag{11}$$

The strategy for the first technique, which we shall call stratified proportion estimation using sequential allocation, is to first allocate at random a fixed number of pixels to each cluster for the purpose of obtaining an initial estimate of the proportion of each cluster which is the crop of interest. Then $\Delta\sigma^2_i$ is computed for each cluster, and the next sample to be labeled is allocated to the cluster with the largest value of $\Delta\sigma^2_i$. This process continues until a fixed number of pixels have been labeled. The proportion estimate is then

$$\hat{p} = \sum_i \left(\frac{N_i}{N_T}\right)\left(\frac{x_i}{n_i}\right) \tag{12}$$

The last technique, which is called stratified proportion estimation using Bayesian sequential allocation, is similar to the technique just described except that the additional information of a prior distribution on cluster purities is used. In this case we use the posterior Bayes estimate

$$\hat{\theta}_i = E(\theta_i|x_i) = \frac{1}{f(x_i)}\int_0^1 \theta f(x_i|\theta_i)g(\theta_i)d\theta_i \tag{13}$$

in place of the minimum variance unbiased estimator

$$\hat{P}_i = \frac{x_i}{n_i}$$

Although $\hat{\theta}_i$ is not unbiased, it is the minimum mean-square-error estimator. Following an initial fixed allocation to each cluster, one may then use $\hat{\theta}_i$ in place of $\hat{P}_i$ in equations (8) and (9) to calculate $\Delta\sigma_i^2$ for each cluster and proceed to allocate sequentially as before. The only difficulty is in the selection of a prior distribution on cluster purities.

The prior distribution on cluster purities was chosen following an examination of the empirical distribution for each of the three clustering algorithms on a subset of 10 segments. These histograms representing percentage of clusters versus ground-truth percentage of small grains are given in figures 3-1, 3-2, and 3-3. The similarity of these histograms and their general shape led to the belief that at least for segments having a moderate to large amount of small grains, a prior distribution which was quadratic in form would be appropriate.

It seemed reasonable that the prior distribution, $g(\theta)$, satisfy the following criteria.

$$g(\theta) \geq 0 \text{ for all } 0 \leq \theta \leq 1$$

$$\int_0^1 g(\theta)d\theta = 1 \tag{14}$$

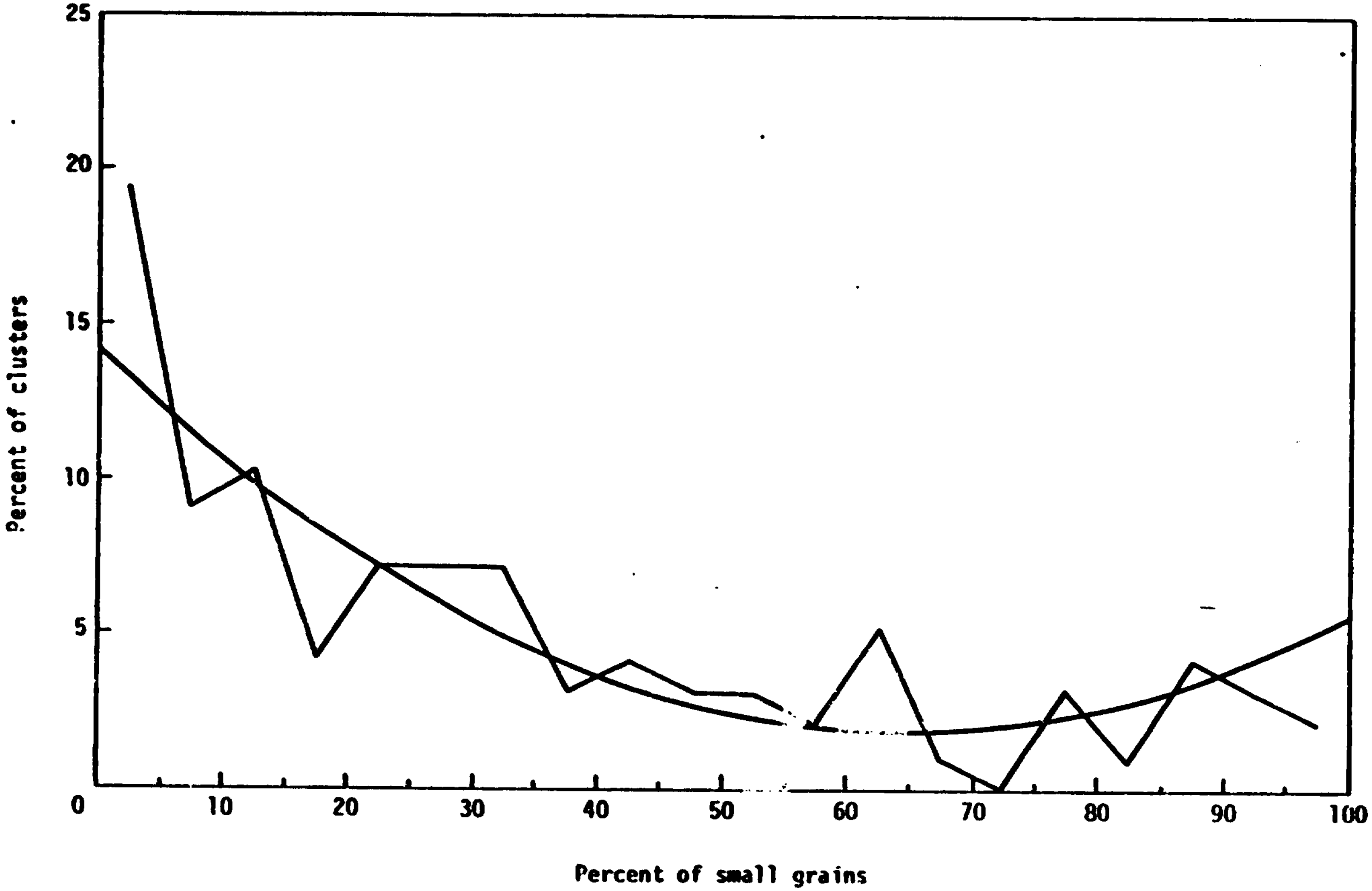


Figure 3-1.– Empirical purity distribution for CLASSY clusters over 10 segments compared with quadratic prior.

25

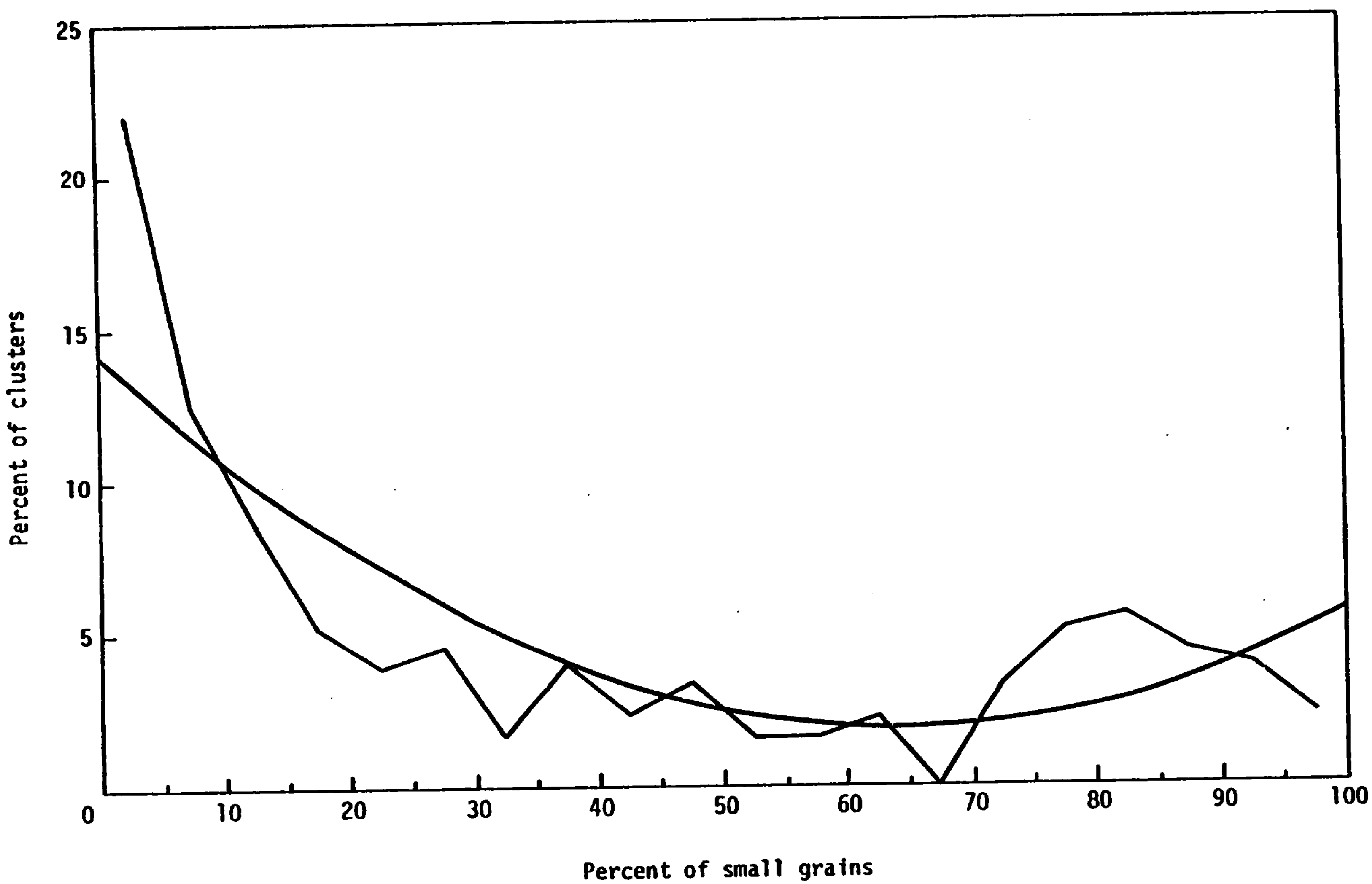


Figure 3-2.– Empirical purity distribution for AMOEBA clusters over 10 segments compared with quadratic prior.

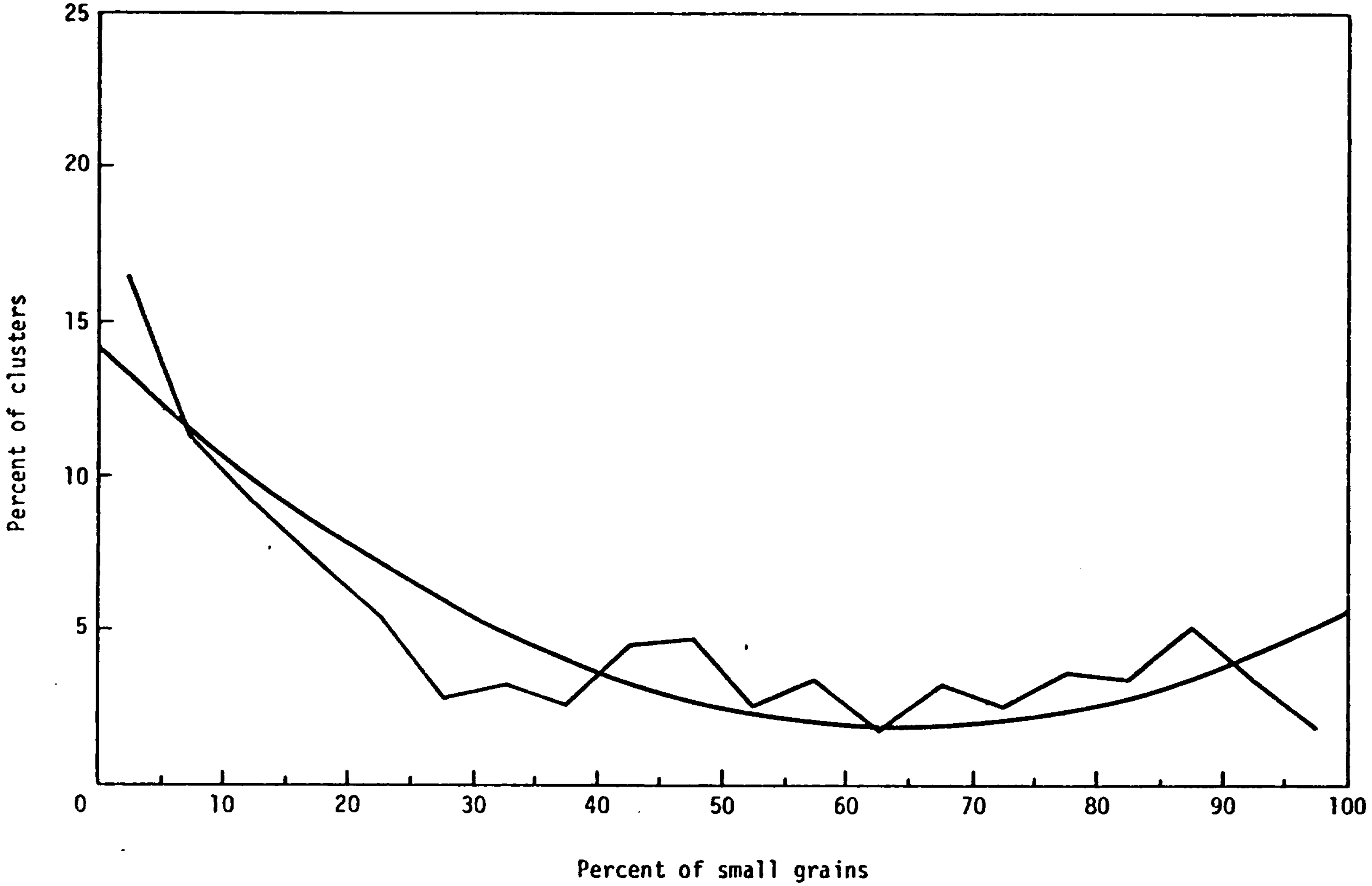


Figure 3-3.– Empirical purity distribution for ISOCLS clusters over 10 segments compared with quadratic prior.

and

$$\int_0^1 \theta g(\theta) d\theta = \hat{P}$$

where

$$\hat{P} = \sum_i \left(\frac{N_i}{N_T}\right) \frac{x_i}{n_i}$$

and is computed following the fixed allocation of pixels to clusters.

These three conditions allow the specification of the three coefficients in the equation

$$g(\theta) = a\theta^2 + b\theta + c$$

These coefficients are

$$\left.\begin{array}{l} a = 6 \\ b = 12(\hat{P} - 1) \\ c = 5 - 6\hat{P} \end{array}\right\} \text{ for } 0.211 \le \hat{P} \le 0.789 \qquad (15)$$

It should be noted that the b and c coefficients are only appropriate for a specified range of $\hat{P}$ values. If $\hat{P}$ is not in this range, then $g(\theta)$ will be negative at some point.

The fact that a quadratic prior is only appropriate over a limited range of P values also seemed to be validated by empirical evidence. Figures 3-4 and 3-5 show histograms of cluster purity for eight segments which had low ground-truth proportions of small grains. Clearly a quadratic prior is not appropriate. On this basis, it was decided to select an alternate prior for segments which had a small portion of the crop of interest. The prior for segments with a very large proportion of the crop of interest might reasonably be thought to be like a "flipped" version of the prior for small proportion segments.

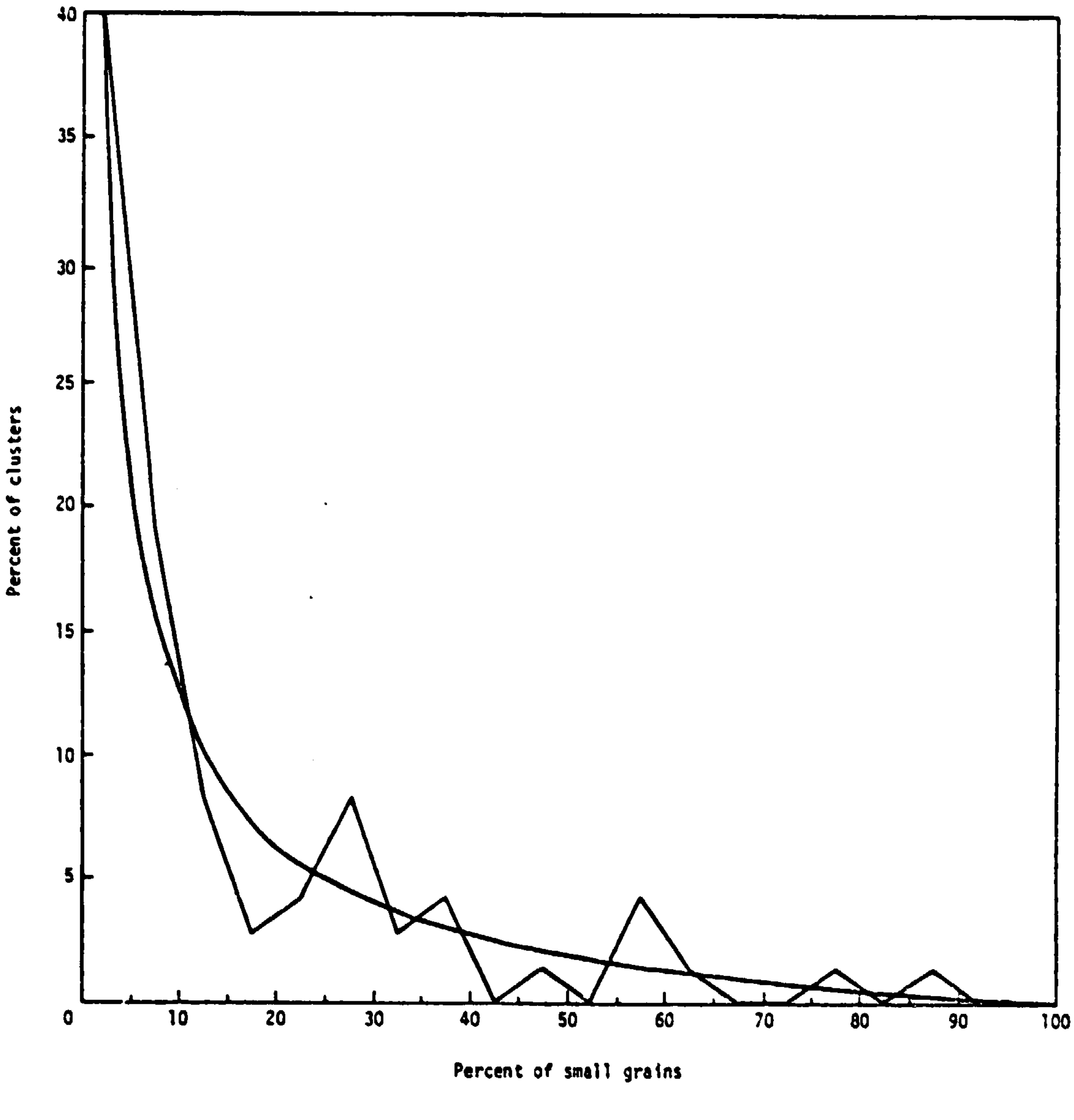


Figure 3-4.— Empirical purity distribution for CLASSY clusters over eight small proportion segments compared with exponential prior.

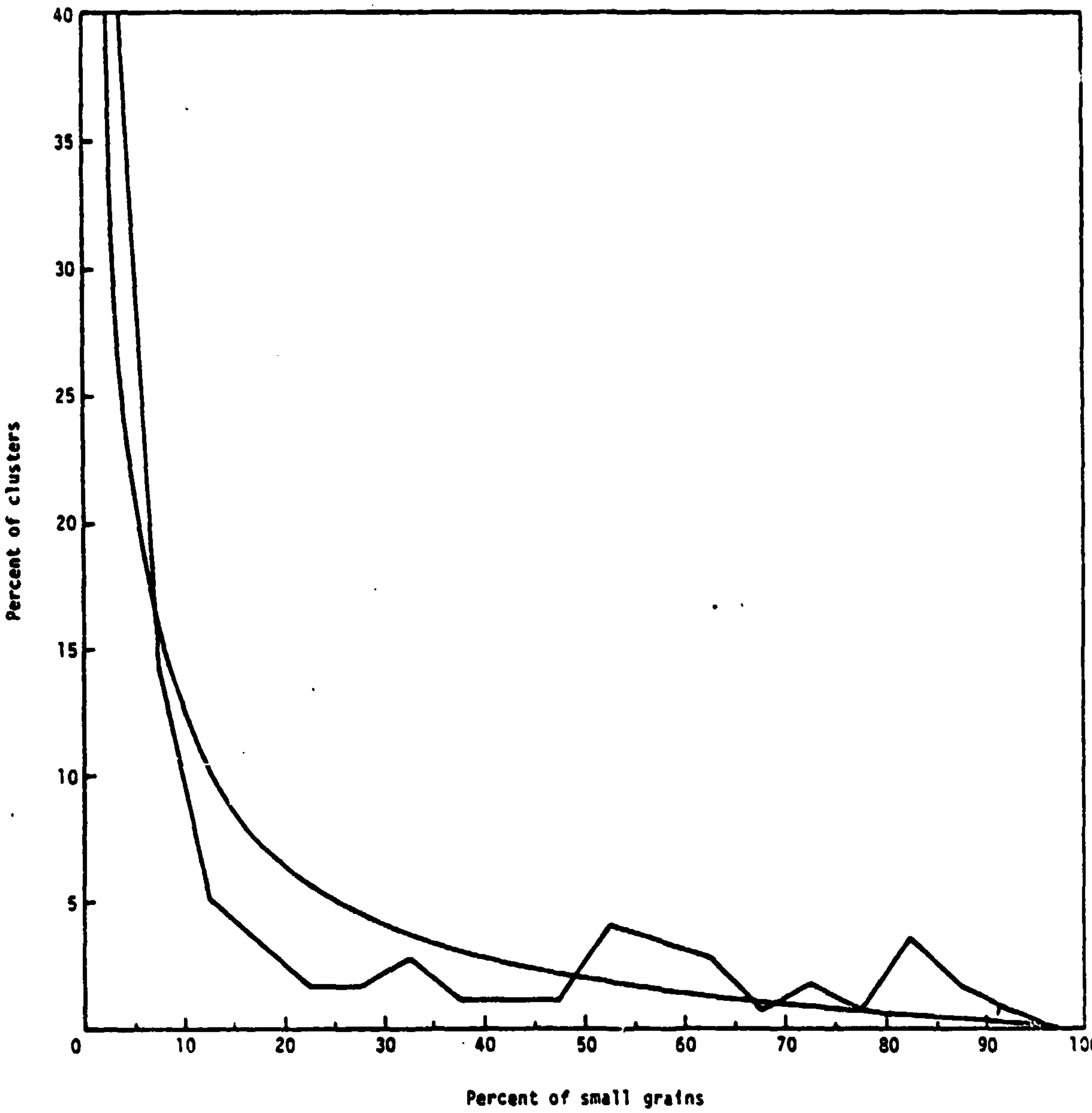


Figure 3-5.— Empirical purity distribution for AMOEBA clusters over eight small proportion segments compared with exponential prior.

It was decided that the form of the prior for small proportion segments would be

$$g(\theta) = \beta\theta^{-\alpha} - \beta = \beta(\theta^{-\alpha} - 1) \tag{16}$$

and that this distribution should satisfy the following constraints

$$g(\theta) \geq = 0 \text{ for all } 0 \leq \theta \leq 1$$

$$\int_0^1 g(\theta)d\theta = 1$$

$$g(1) = 0$$

$$\int_0^1 \theta g(\theta)d\theta = \hat{P} \tag{17}$$

These constraints may be used to determine the parameters $\alpha$ and $\beta$ which are

$$\alpha = \frac{1 - 4\hat{P}}{1 - 2\hat{P}} \Bigg\} \text{ for } 0 < \hat{P} \leq 0.25$$

$$\beta = \frac{1 - \alpha}{\alpha} \tag{18}$$

This prior will be called the exponential prior. In order to see how well the quadratic and exponential priors fit the empirical cluster purity histograms, the following calculations were made:

a. The average ground-truth proportion of small grains in the 10 segments used to obtain the data reflected in figures 3-1, 3-2, and 3-3 was computed.

b. The average ground-truth proportion of small grains in the eight segments used to obtain the data reflected in figures 3-4 and 3-5 was computed.

The first proportion, call it $P_1$, was then used to calculate the coefficients a, b, and c [equation (15)] specifying a quadratic prior. This prior is plotted in figures 3-1, 3-2, and 3-3 as a smooth curve for comparison with the empirical histograms. Similarly, the average ground-truth proportion for the eight small proportion segments, call it $P_2$, was used to calculate the coefficients $\alpha$ and $\beta$ for an exponential prior. This prior is plotted as a smooth curve on figures 3-4 and 3-5. It is evident from examining figures 3-1 through 3-5 that both prior distibutions seem to fit the empirical cluster purity distributions well.

In actual practice, both the sequential and the Bayesian sequential procedure were initiated with random allocation of two pixels per cluster. Following this allocation, the Bayesian sequential procedure computes two different estimates of the segment proportion. One is given by

$$\hat{p} = \sum_i \left(\frac{N_i}{N_T}\right) \frac{x_i}{n_i} \tag{19}$$

whereas the other is the Bayes posterior estimate based on a quadratic prior and an average proportion estimate of P = 0.34. The equation for this estimate is

$$\hat{\theta} = \sum_i \left(\frac{N_i}{N_T}\right) \hat{\theta}(n_i, x_i) \tag{20}$$

where

$$\hat{\theta}(n_i, x_i) = \frac{a[(x_i + 1)(x_i + 2)(x_i + 3)] + b[(x_i + 1)(x_i + 2)(n_i + 4)] + c[(x_i + 1)(n_i + 3)(n_i + 4)]}{a[(x_i + 1)(x_i + 2)(n_i + 4)] + b[(x_i + 1)(n_i + 3)(n_i + 4)] + c[(n_i + 2)(n_i + 3)(n_i + 4)]} \tag{21}$$

If $0.211 \leq \hat{P}$, then the quadratic prior is selected and $\hat{\theta}$ is used to reset the parameters a, b, and c. Sequential selection then proceeds with

$$\Delta\sigma_i^2 = \left(\frac{N_i}{N_T}\right)^2 \left[\frac{\hat{\theta}(n_i, x_i)[1 - \hat{\theta}(n_i, x_i)]}{n_i - 1} - \frac{\hat{\theta}(n_i, x_i)\hat{\theta}(n_i + 1, x_i + 1)[1 - \hat{\theta}(n_i + 1, x_i + 1)]}{n_i} - \frac{[1 - \hat{\theta}(n_i x_i)]\hat{\theta}(n_i + 1, x_i)[1 - \hat{\theta}(n_i + 1, x_i)]}{n_i}\right] \tag{22}$$

After a number of dots have been allocated, an overall proportion estimate is obtained via equation (20), using the current values of the $\hat{\theta}(n_i, x_i)$ estimates. If $0.211 > \hat{P}$, then the exponential prior is used to calculate the parameters $\alpha$

and $\beta$. Sequential selection then proceeds with $\Delta\sigma_i^2$ given by equation (22), using

$$\hat{\theta}(n_i, x_i) = \frac{\left(\frac{x_i + 1 - \alpha}{n_i + 2 - \alpha}\right) - \left(\frac{x_i + 1}{n_i + 2}\right)\gamma_2}{\gamma_1 - \gamma_2} \tag{23}$$

where

$$\gamma_1 = (n_i + 1)(n_i)(n_i - 1) \cdots (x_i + 1)$$

$$\gamma_2 = (n_i + 1 - \alpha)(n_i - \alpha) \cdots (x_i + 1 - \alpha)$$

After a number of dots have been allocated, an overall proportion estimate is obtained as before using equation (20).

Figure 3-6 shows a comparison of the quadratic and exponential priors at the value $\hat{P} = 0.211$, where the switch occurs from one to the other. The curves are close enough for this value of $\hat{P}$ that the decision as to which one to use is not critical.

Outlined in this section are six different techniques for cluster based proportion estimation. As a way of summarizing these developments, a brief discussion on some of the expected characteristics of these techniques follows.

Three cluster-labeling and three stratified proportion-estimation schemes have been considered. If the clusters are very pure, then cluster labeling should produce proportion estimates with small bias and very small variance. In addition, relatively few labeled pixels should be required to obtain these estimates, and the estimates themselves should not be very sensitive to occasional labeling errors. Cluster labeling using sequential allocation or Bayesian sequential allocation provides a specified confidence in the labels of clusters. These techniques should require fewer dots to be labeled on the average than does cluster labeling using proportional allocation.

33

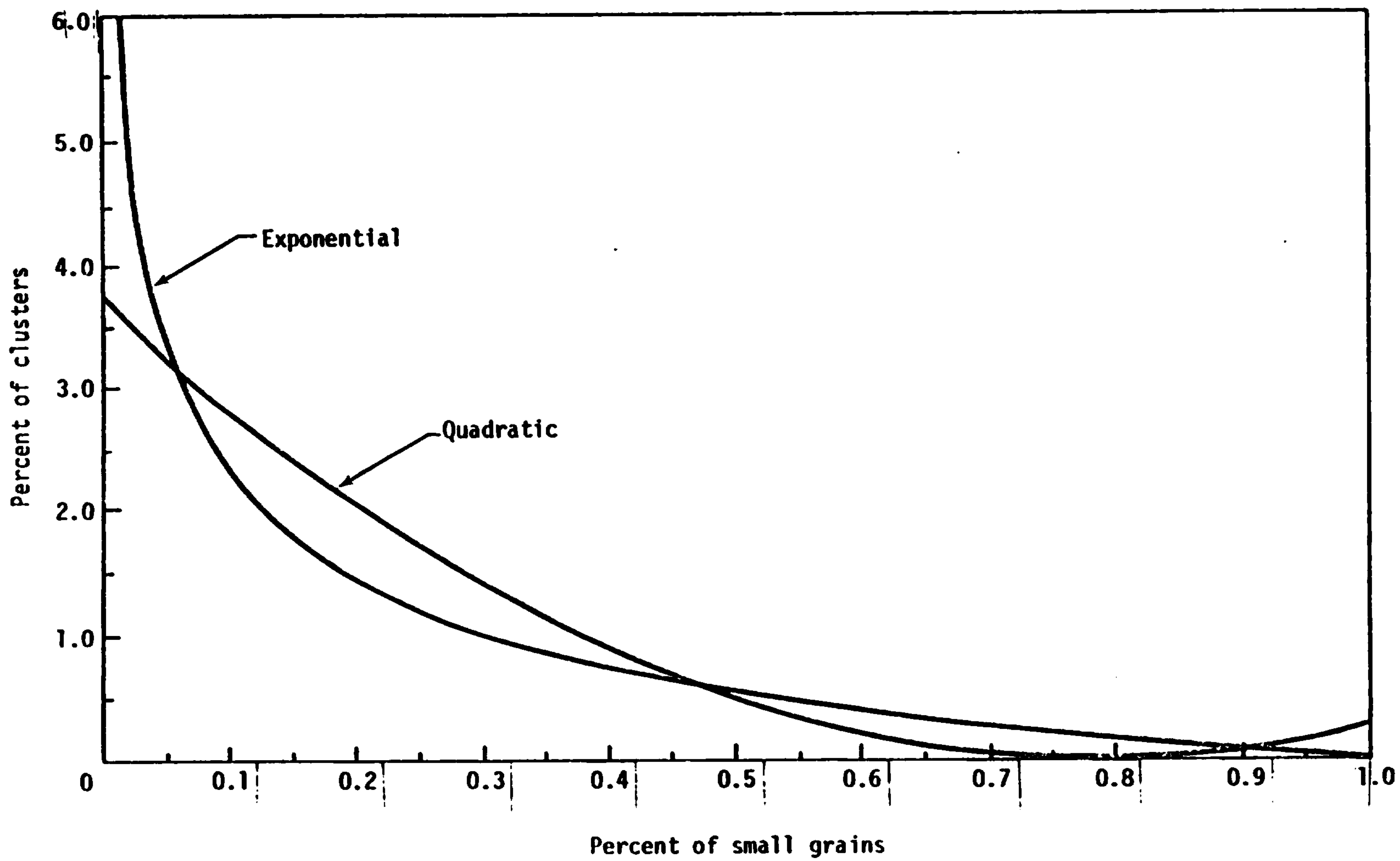


Figure 3-6.— Comparison of quadratic and exponential oriors at the value of P = 0.211.

If the clusters are significantly mixed, all of the cluster-labeling schemes will suffer. In this case, a more appropriate technique is provided by stratified proportion estimation. Stratified proportion estimation, using proportional allocation, provides theoretically unbiased estimates. The stratified proportion estimation, using sequential and Bayesian sequential allocation, are not theoretically unbiased but should produce estimates with a lower mean-square error for a given number of dots allocated than the proportional allocation approach. Both of the sequential techniques incorporate information about both the size and the estimated purity of clusters in performing the dot allocation.

## 4. DATA SET AND EXPERIMENTAL DESIGN

The data set for this study consisted of 25 LACIE segments selected at random from the Phase III (1976-1977) blind site data base. Eighteen of the segments are the same as those used in the secondary error analysis study (refs. 2 and 3). Seven substitutions in the secondary error analysis data set were necessary because the original segments were not well registered to the digitized ground truth. The segments selected represent a cross section of the U.S. Great Plains. Both winter- and spring-wheat segments were included.

Three segments in the data set were discovered to have significant amounts of strip fallow small grains where the strips were not resolved in the ground truth. These segments, 1648, 1739, and 1544, were clustered but were not evaluated using the proportion-estimation schemes because reliable labels were not available for the strip fallow area. One other segment, 1079, was not evaluated using the proportion-estimation schemes because it was found to contain 27 percent abandoned winter wheat and was, thus, a very atypical segment. In table 4-1 is a listing of the 21 segments actually used in the testing, their location, the acquisitions used, and the proportion of small grains from the digitized ground truth.

The experimental design for the evaluation of the six proportion-estimation techniques was that each of them were evaluated on a subset of five segments selected from the set of 21 acceptable segments. The subset that was selected consisted of segments 1005, 1853, 1520, 1231, and 1060. After evaluating these preliminary results, the most promising techniques were selected and run on the remainder of the 21 segments.

Each proportion-estimation technique — clustering algorithm combination — was repeated 100 times for each segment. Each repetition used a different pseudo random sequence in selecting pixels. Thus, it was possible to calculate the average bias in the proportion estimate, the mean-square error of the estimate, and the R factor as compared to simple random sampling. These results are reported in the appendix. Averages and variances of these results over segments were also calculated. These results appear in section 5.





36

TABLE 4-1.— DESCRIPTION OF THE TWENTY-ONE SEGMENTS USED IN THE STUDY

| Segment | Location | Acquisitions used | Ground-truth proportion of small grains |
|---|---|---|---|
| 1005 (W) | Cheyenne, Colorado | 7177, 7159, 6326, 6254 | 0.348 |
| 1032 (W) | Wichita, Kansas | 7194, 7086, 6326, 6254 | .371 |
| 1033 (W) | Clark, Kansas | 7156, 6288 | .095 |
| 1853 (W) | Ness, Kansas | 7193, 7067, 6253 | .306 |
| 1166 (W) | Lyon, Kansas | 7190, 7154, 7082, 6286 | .066 |
| 1512 (S) | Clay, Minnesota | 7193, 7156 | .340 |
| 1520 (S) | Big Stone, Minnesota | 7174, 7156, 7120 | .301 |
| 1577 (W) | Platte, Nebraska | 7120, 6306 | .029 |
| 1604 (S) | Renville, North Dakota | 7143, 7125 | .524 |
| 1606 (S) | Ward, North Dakota | 7197, 7125 | .330 |
| 1661 (S) | McIntosh, North Dakota | 7159, 7123 | .414 |
| 1899 (S) | Walsh, North Dakota | 7193, 7175, 7157, 7122 | .596 |
| 1231 (W) | Jackson, Oklahoma | 7156, 7066, 6288 | .744 |
| 1239 (W) | Noble, Oklahoma | 7155, 7082, 6268 | .167 |
| 1367 (W) | Major, Oklahoma | 7155, 7101, 6287 | .606 |
| 1675 (S) | McPherson, South Dakota | 7230, 7176, 7123, 6254 | .291 |
| 1686 (S) | Beadle, South Dakota | 7194, 7140, 6307, 6254 | .194 |
| 1803 (W) | Shannon, South Dakota | 7178, 7159, 7123, 6255 | .032 |
| 1805 (M) | Gregory, South Dakota | 7211, 7158, 6307, 6290 | .164 |
| 1059 (W) | Ochiltree, Texas | 7157, 7121, 6325, 6307 | .437 |
| 1060 (W) | Sherman, Texas | 7158, 7068 | .231 |

Symbol definition:

M = Mixed

S = Spring wheat

W = Winter wheat

## 5. RESULTS

The results of the study are summarized in two parts. The first part pertains to the evaluation of the clustering algorithms, and the second part is an evaluation and comparison of the six techniques for proportion estimation.

The R, as compared to simple random sampling, and the PCC, using majority rule labeling, are given in table 5-1 for each of the three algorithms tested as applied to each of the 21 segments. Averages for each measure over segments are given at the bottom of the table along with an estimate of the standard deviation over segments. None of the averages are significantly different. In fact, it is striking how similar the average results are in view of the differences in the algorithms. This similarity will be further discussed in section 6.

One significant difference is in the number of clusters produced by each algorithm. At the bottom of table 5-1, the average number of clusters and the standard deviation in the number of clusters are indicated. The average number of clusters nearly doubles when going from CLASSY to AMOEBA and doubles again in going from AMOEBA to ISOCLS. Economy in the number of clusters produced is generally considered a distinct advantage for a clustering algorithm. It is clearly an advantage in the stratified proportion-estimation techniques. Indeed the sequential stratified techniques require that a fixed number of pixels (usually 2) be allocated to each cluster initially. Thus, a large number of clusters means that a large number of pixels must be allocated before sequential allocation even begins.

Presented in tables 5-2, 5-3, and 5-4 are the results for the three cluster-labeling schemes; and in tables 5-5, 5-6, and 5-7 are the results for the three stratified proportion-estimation schemes. The results presented in each table are averages and variances over the segments processed for each of the measures recorded, using a given scheme. For each scheme, with the exception of stratified proportion estimation using proportional allocation, the measures recorded were the average bias, the mean-square error, and the reduction

TABLE 5-1.— PCC VALUES USING MAJORITY RULE LABELING AND R VALUES FOR CLASSY, AMOEBA, AND ISOCLS

| Segment | CLASSY | | AMOEBA | | ISOCLS | |
|---|---|---|---|---|---|---|
| | PCC | R | PCC | R | PCC | R |
| 1005 (W) | 0.8398 | 0.5[illegible]71 | 0.9132 | 0.6372 | 0.8659 | 0.6571 |
| 1032 (W) | .8975 | .3450 | .8541 | .4585 | .8367 | .4978 |
| 1033 (W) | .9050 | .8208 | .9151 | .7363 | .9247 | .6247 |
| 1853 (W) | .8948 | .4073 | .7926 | .6966 | .8859 | .4655 |
| 1166 (W) | .9333 | .8287 | .9388 | .7857 | .9386 | .6994 |
| 1512 (S) | .7110 | .8269 | .7621 | .7481 | .7576 | .7767 |
| 1520 (S) | .8361 | .5758 | .8522 | .5213 | .8546 | .5735 |
| 1577 (W) | .9678 | .9055 | .9678 | .9076 | .9684 | .8814 |
| 1604 (S) | .6877 | .8419 | .7318 | .7538 | .6749 | .7893 |
| 1606 (S) | .8229 | .6071 | .8002 | .6511 | .7958 | .7201 |
| 1661 (S) | .7260 | .7395 | .7523 | .6745 | .7184 | .7767 |
| 1899 (S) | .8427 | .4852 | .8555 | .4684 | .8426 | .5196 |
| 1231 (W) | .8773 | .4849 | .8926 | .4450 | .8788 | .4941 |
| 1239 (W) | .8508 | .7175 | .8702 | .6586 | .8601 | .7322 |
| 1367 (W) | .8023 | .5654 | .8198 | .5644 | .8051 | .6238 |
| 1675 (S) | .7929 | .7056 | .8060 | .6243 | .7890 | .7282 |
| 1686 (S) | .8352 | .7847 | .8485 | .6933 | .8400 | .8128 |
| 1803 (W) | .9681 | .8313 | .9701 | .7339 | .9733 | .6502 |
| 1805 (M) | .9052 | .5007 | .9199 | .4680 | .9219 | .4839 |
| 1059 (W) | .8448 | .4515 | .8667 | .4126 | .8758 | .4062 |
| 1060 (W) | .8583 | .5984 | .8824 | .5227 | .8757 | 6002 |
| Average | .8476 | .6472 | .8521 | .6268 | .8488 | .6435 |
| Standard deviation | .0754 | .1663 | .0688 | .1333 | .0771 | .1316 |
| Average number of clusters, ± 1 standard deviation | 9.32 ± 2.15 | | 17.46 ± 10.15 | | 36.84 ± 2.32 | |

A-36



34

TABLE 5-2.— MAJORITY RULE LABELING USING PROPORTIONAL ALLOCATION RESULTS FOR FIVE SEGMENTS

| Number of pixels allocated | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|---|---|---|
| | Average bias | | | Variance of bias | | |
| 30 | -0.009508 | -0.015600 | 0.013634 | 0.000839 | 0.001[illegible] | 0.000202 |
| 60 | .001838 | -.026056 | -.024830 | .002620 | .000596 | .000195 |
| 90 | -.071312 | -.034964 | -.026952 | .022647 | .000651 | .000371 |
| 120 | -.016828 | -.033568 | -.016600 | .001955 | .000800 | .001039 |
| | Average mean-square error | | | Variance of mean-square error | | |
| 30 | 0.024594 | 0.057056 | 0.011561 | 0.000188 | 0.002791 | 0.000050 |
| 60 | .054702 | .038171 | .029260 | .002262 | .000619 | .000205 |
| 90 | .062212 | .050078 | .029656 | .005637 | .002679 | .000463 |
| 120 | .047929 | .049945 | .033015 | .003409 | .002345 | .001398 |
| | Average reduction in mean-square error | | | Variance of reduction in mean-square error | | |
| 30 | 3.585012 | 8.984081 | 1.747804 | 3.608364 | 83.195801 | 1.329618 |
| 60 | 16.227676 | 11.904598 | 8.945074 | 207.806641 | 71.670441 | 25.393509 |
| 90 | 27.270935 | 24.033615 | 13.662670 | 1017.292236 | 719.822998 | 115.909088 |
| 120 | 27.489548 | 32.010651 | 20.962250 | 1101.703857 | 1113.502686 | 631.753662 |

TABLE 5-3.— MAJORITY RULE LABELING USING SEQUENTIAL ALLOCATION RESULTS FOR FIVE SEGMENTS, THREE-PIXEL PER CLUSTER INITIAL ALLOCATION

| CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|---|---|
| Average bias | | | Variance of bias | | |
| -0.04449496 | -0.03424257 | -0.03201438 | 0.00107109 | 0.00053136 | 0.00094198 |
| Average mean-square error | | | Variance of mean square-error | | |
| 0.00574680 | 0.00254860 | 0.00266640 | 0.00000913 | 0.00000660 | 0.00000073 |
| Average reduction in mean-square error | | | Variance of reduction in mean-square error | | |
| 1.67606068 | 1.24144173 | 3.41460514 | 0.90543842 | 1.75853252 | 1.39696312 |
| Average number of pixels allocated | | | Variance of number of pixels allocated | | |
| 57.648 | 75.286 | 257.475 | 68.674 | 2042.372 | 308.177 |



41

TABLE 5-4.— MAJORITY RULE LABELING USING BAYESIAN SEQUENTIAL ALLOCATION RESULTS FOR FIVE SEGMENTS, TWO-PIXEL PER CLUSTER INITIAL ALLOCATION

| CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|---|---|
| Average bias | | | Variance of bias | | |
| -0.03277557 | -0.02864778 | -0.02584878 | 0.00060669 | 0.00038843 | 0.00079368 |
| Average mean-square error | | | Variance of mean-square error | | |
| 0.00604460 | 0.00682659 | 0.00267940 | 0.00000393 | 0.00000916 | 0.00000062 |
| Average reduction in mean-square error | | | Variance of reduction in mean-square error | | |
| 0.91108280 | 1.38561249 | 1.65233707 | 0.13923180 | 0.85401917 | 0.18573952 |
| Average number of pixels allocated | | | Variance of number of pixels allocated | | |
| 29.930 | 43.074 | 125.996 | 23.486 | 566.810 | 47.896 |

TABLE 5-5.– STRATIFIED PROPORTION ESTIMATION USING PROPORTIONAL ALLOCATION

RESULTS FOR TWENTY-ONE SEGMENTS

| Number of pixels allocated | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|---|---|---|
| | Average variance | | | Variance of variance | | |
| 30 | 0.003852895 | 0.003591756 | 0.003565516 | 0.000004197 | 0.000002433 | 0.000002063 |
| 60 | .001815951 | .001814903 | .001715998 | .000000648 | .000000738 | .000000464 |
| 90 | .001301855 | .001269474 | .001444855 | .000000391 | .000000339 | .000000871 |
| 120 | .000884570 | .000945522 | .000986570 | .000000143 | .000000164 | .000000350 |
| | Average reduction in variance | | | Variance of reduction in variance | | |
| 30 | 0.687449038 | 0.627526164 | 0.636414111 | 0.053946018 | 0.019914806 | 0.025356948 |
| 60 | .636317074 | .626016080 | .629446924 | .023804247 | .031225204 | .042545319 |
| 90 | .688710690 | .656349719 | .694832742 | .041802645 | .024449527 | .042262435 |
| 120 | .636751771 | .662965417 | .624346912 | .028034508 | .024315834 | .023863912 |

TABLE 5-6.— STRATIFIED PROPORTION ESTIMATION USING SEQUENTIAL ALLOCATION RESULTS FOR FIVE SEGMENTS, THREE-PIXEL PER CLUSTER INITIAL ALLOCATION

| Number of pixels allocated | CLASSY | AMOEBA | ISCCLS | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|---|---|---|
| | Average bias | | | Variance of bias | | |
| 30 | -0.00088333 | -0.00585000 | 0.0 | 0.00015393 | 0.00003784 | 0.0 |
| 60 | -.01415999 | -.02248665 | .0 | .00036671 | .00009266 | .0 |
| 90 | -.01781999 | -.02010199 | .0 | .00045373 | .00013612 | .0 |
| 120 | -.01948998 | -.02173998 | -.00385000 | .00046703 | .00017864 | .00007823 |
| | Average mean-square error | | | Variance of mean-square error | | |
| 30 | 0.00345100 | 0.00513500 | 0.0 | 0.00000020 | 0.00000001 | 0.0 |
| 60 | .00296520 | .00325900 | .0 | .00000024 | .00000002 | .0 |
| 90 | .00277940 | .00298240 | .0 | .00000030 | .00000090 | .0 |
| 120 | .00274540 | .00276980 | .00124575 | .00000035 | .00000087 | .00000015 |
| | Average reduction in mean-square error | | | Variance of average reduction in mean-square error | | |
| 30 | 0.54175025 | 0.72903204 | 0.0 | 0.01088542 | 0.00039721 | 0.0 |
| 60 | .87602842 | .98629665 | .0 | .01731825 | .01368725 | .0 |
| 90 | 1.23414421 | 1.30850601 | .0 | .05600834 | .13552380 | .0 |
| 120 | 1.62500954 | 1.61916065 | .70379806 | .11822701 | .24639034 | .038685[illegible]4 |

TABLE 5-7.— STRATIFIED PROPORTION ESTIMATION USING BAYESIAN SEQUENTIAL ALLOCATION RESULTS FOR TWENTY-ONE SEGMENTS, TWO-PIXEL PER CLUSTER INITIAL ALLOCATION

| Number of pixels allocated | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|---|---|---|
| | Average bias | | | Variance of bias | | |
| 30 | 0.00036809 | -0.00841666 | 0.0 | 0.00010890 | 0.00051509 | 0.0 |
| 60 | .00006095 | -.00430625 | .0 | .00012138 | .00013838 | .0 |
| 90 | -.00037000 | -.00495141 | -.00323619 | .00008227 | .00020197 | .00007368 |
| 120 | -.00040190 | -.00451095 | -.00324428 | .00006833 | .00017815 | .00007746 |
| | Average mean-square error | | | Variance of mean-square error | | |
| 30 | 0.00285286 | 0.00522211 | 0.0 | 0.00000367 | 0.00000503 | 0.0 |
| 60 | .00148009 | .00212906 | .0 | .00000065 | .00000119 | .0 |
| 90 | .00099690 | .00140800 | .00099719 | .00000030 | .00000059 | .00000021 |
| 120 | .00073538 | .00106862 | .00075933 | .00000015 | .00000035 | .00000012 |
| | Average reduction in mean-square error | | | Variance of reduction in mean-square error | | |
| 30 | 0.48676664 | 0.76839358 | 0.0 | 0.04504710 | 0.10229522 | 0.0 |
| 60 | .51693314 | .72288340 | .0 | .03661084 | .06289172 | .0 |
| 90 | .52017057 | .72251660 | .51264614 | .03732508 | .07170510 | .01777804 |
| 120 | .51932829 | .73885107 | .52794492 | .03581393 | .08057529 | .02143240 |

in mean-square error as compared to simple random sampling. Because stratified proportion estimation (using proportional allocation) is theoretically unbiased, the bias was not recorded; the variance and the R, rather than the mean-square error and reduction in mean-square error, were recorded. The techniques using sequential allocation for majority-rule labeling did not allocate a fixed number of pixels, and hence, only the average number of pixels allocated is reported. The sequential Bayesian technique used an initial allocation of two pixels per cluster, whereas the sequential technique without prior used a three-pixel cluster initial allocation. The same initial allocation was used for the Bayesian and "no prior" sequential techniques that were used in stratified proportion-estimation. The missing values in tables 5-6 and 5-7 indicate that in some cases sequential allocation could not begin until a larger number of dots had been allocated.

After examining the results for the subset of five segments, it was clear that all of the cluster-labeling schemes as well as the stratified proportion estimation using sequential allocation were not competitive with stratified proportion estimation using either proportional allocation or Bayesian sequential allocation. This is most readily apparent in a comparison of the reduction in mean-square error or R results.

The technique using sequential allocation in obtaining stratified proportion estimates does look competitive at an allocation of 30 pixels. Because it was not significantly better than stratified proportion estimation using Bayesian sequential allocation, it was decided to place the most emphasis on a comparison of the Bayesian sequential and the proportional allocation techniques as used in obtaining stratified proportion estimates. Consequently, tables 5-5 and 5-7 represent results for the full 21 segments, whereas 5-2, 5-3, 5-4, and 5-6 represent the results for five segments.

Figures 5-1 and 5-2 are a presentation in histogram form of the same data which are summarized in tables 5-5 and 5-7. Figure 5-3 is a comparative histogram plot of R values for Procedure 1, which are reported in reference 3. In this plot, it is assumed that there is an allocation of pixels equal to the

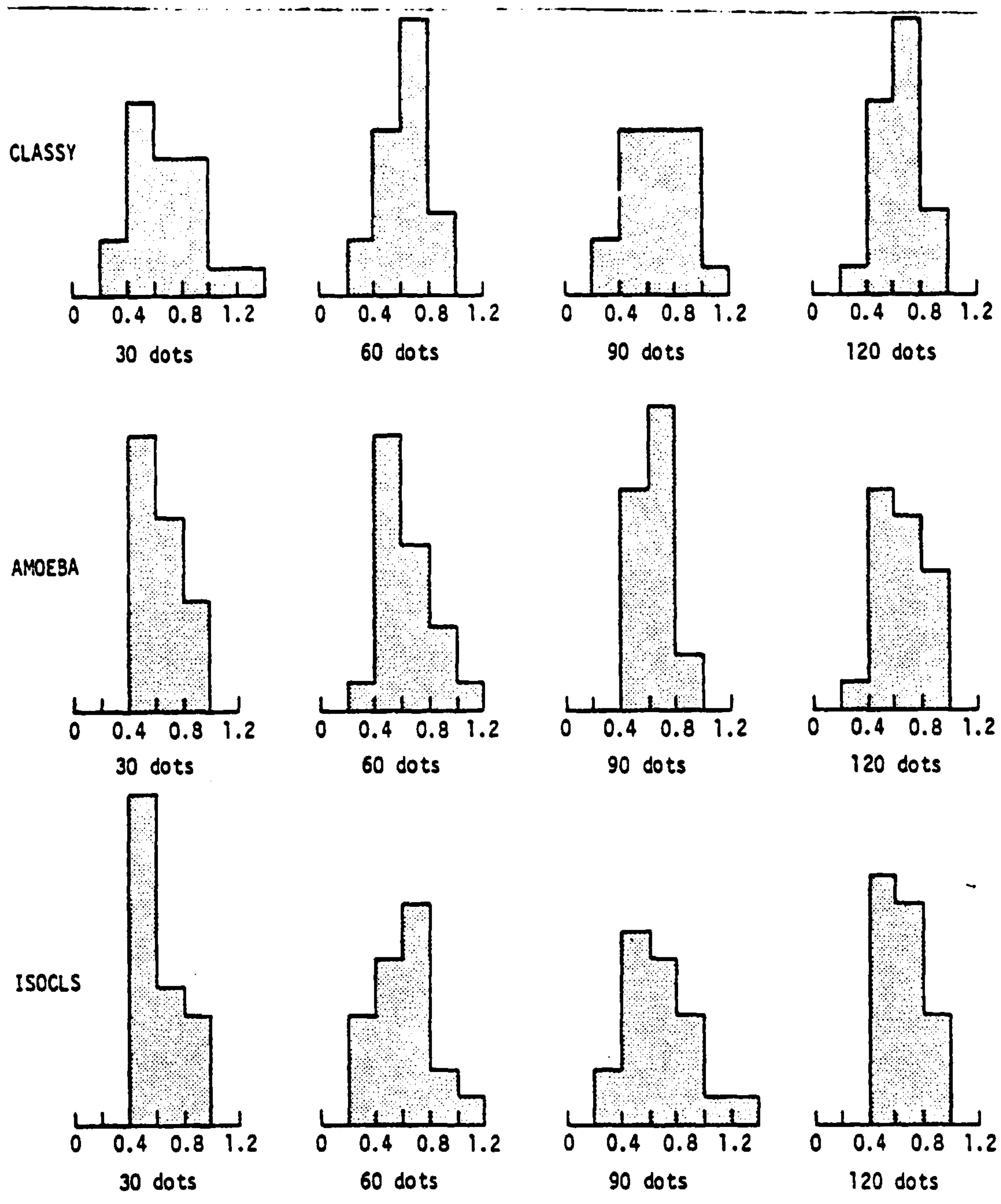


Figure 5-1.— Histogram plots of the R for stratified proportion estimation using proportional allocation.

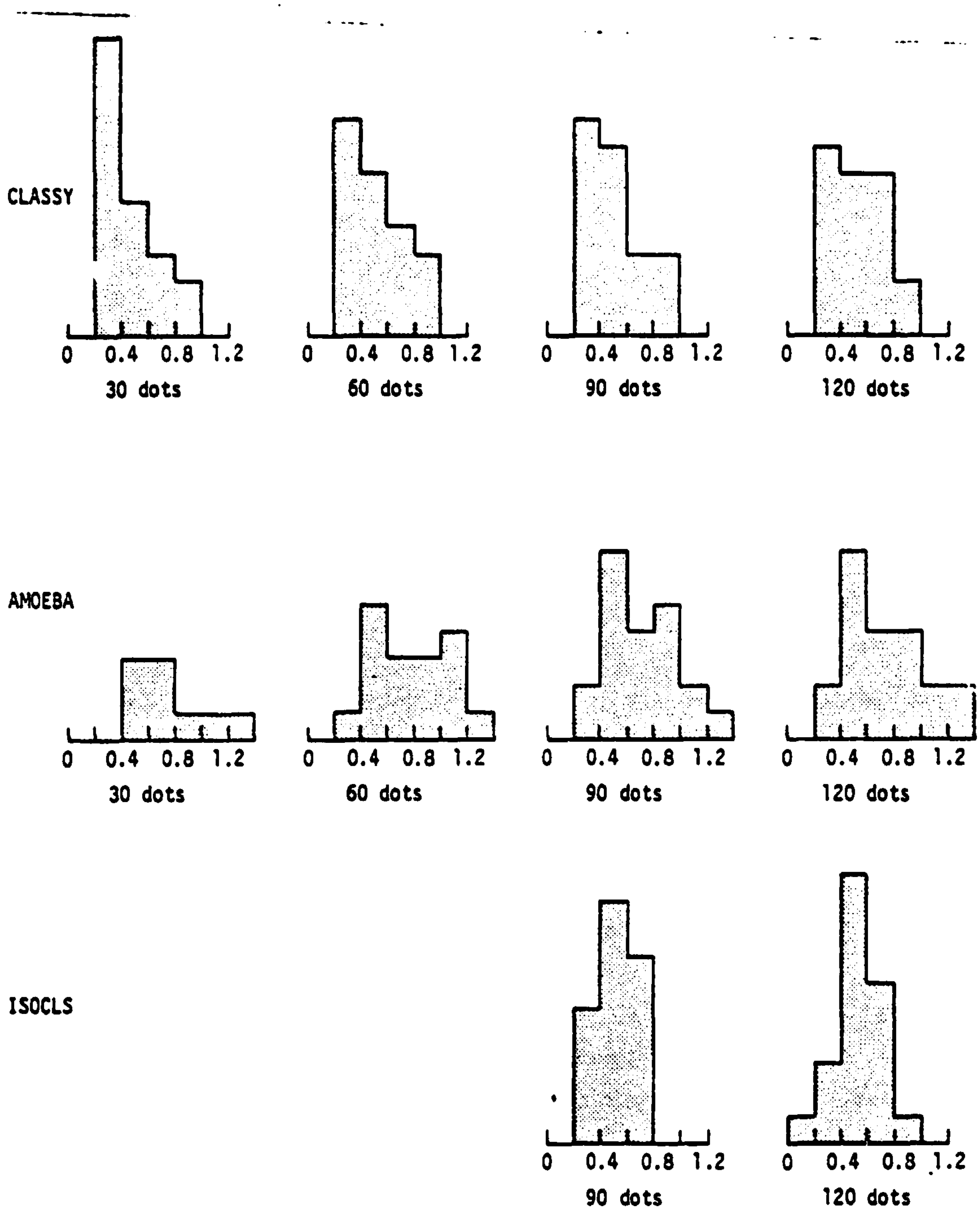


Figure 5-2.— Histogram plots of the reduction in mean-square error for stratified proportion estimation using Bayesian sequential allocation.



48

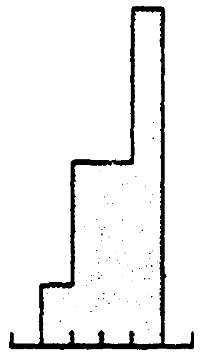


Figure 5-3.— Histogram plot of the R for Procedure 1 based on approximately 60 pixels (type 2) per estimate.

number of type 2 dots used in each estimate (approximately 60 pixels). The complete data for each of the six proportion-estimation techniques studied are in the appendix of this report.

The results in table 5-5 are essentially an empirical verification of the results in table 5-1. In particular, the R averages may be compared. In theory, the R (using this technique) should be independent of the number of dots allocated. Indeed, there are no significant differences among the values of average R calculated for 30, 60, 90, or 120 dots. In addition, the averages for each algorithm tend to agree well with the theoretical average R values appearing in table 5-1.

In examining table 5-7, it is clear that the Bayesian sequential allocation technique, as used in obtaining stratified proportion estimates, has an extremely low bias for all three algorithms even though the procedure itself is not theoretically unbiased. None of the average bias results in this table for any of the algorithms are significantly different from zero.

A comparison of the average reduction in mean-square error for the Bayesian sequential allocation technique (table 5-7) with the average R for the proportional allocation technique (table 5-5) shows that using the Bayesian sequential approach with the CLASSY algorithm gives results which are consistently lower than proportional allocation for all numbers of pixels allocated. If the variances for each technique-algorithm combination are pooled over the various numbers of pixels allocated, the results are given in table 5-8.

TABLE 5-8.— POOLED VARIANCES FOR SEQUENTIAL ALLOCATION TECHNIQUES

| Pool Variances | Bayesian sequential allocation | | | Proportional allocation | | |
|---|---|---|---|---|---|---|
| | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
| | 0.038699 | 0.079350 | 0.019605 | 0.036897 | 0.024976 | 0.033507 |

In table 5-9 are the least significant differences (LSD) for comparisons between the two sequential techniques within the results for a given family. The LSD is computed as

$$\mathrm{LSD} = t\left(\frac{\hat{S}_1^2 + \hat{S}_2^2}{21}\right)^{1/2} \tag{24}$$

where $\hat{S}_1$ and $\hat{S}_2$ are the pooled variance estimates of the groups to be compared and t is the 0.975 percentage point of the Student's-t distribution with 80 degrees of freedom ≈ 1.99.

TABLE 5-9.— LEAST SIGNIFICANT DIFFERENCES FOR COMPARISONS BETWEEN BAYESIAN SEQUENTIAL AND PROPORTIONAL ALLOCATION TECHNIQUES FOR STRATIFIED PROPORTION ESTIMATION

| LSD in R values | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | 0.119397 | 0.140262 | 0.100078 |

The differences between the corresponding R values for tables 5-5 and 5-7 are given in table 5-10.

TABLE 5-10.— VALUES FOR $\overline{R}_{\text{Proportional}} - R_{\text{Bayes sequential}}$

| Pixels | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| 30 | [a]0.200682 | [b]-0.140867 | |
| 60 | [b]0.119384 | -.086566 | |
| 90 | [a]0.168540 | -.066167 | [a]0.182187 |
| 120 | [b]0.116789 | -.075886 | [b]0.096402 |

[a]Significant at the 0.05-percent level.

[b]Marginally significant at the 0.05-percent level.

An examination of table 5-9 shows that the CLASSY results for each number of pixels and the ISOCLS results for 90 and 120 pixels are either significant or very nearly significant at the 0.05-percent level. ISOCLS results are not available for 30 and 60 pixels as there were more pixels than 60 allocated following the two-pixel per cluster allocation in the Bayesian sequential procedure. The AMOEBA results for the Bayesian procedure are consistently higher than for the proportional allocation procedure, and in the case of 30 pixels allocated, the reduction in mean-square-error value was significantly higher.

## 6. CONCLUSIONS AND RECOMMENDATIONS

The clustering algorithms CLASSY, AMOEBA, and ISOCLS performed comparably with respect to the PCC using majority-rule labeling and the R measures. The fact that the average results for all three algorithms were so similar and that the average R value for Procedure 1 has been reported in several independent studies to be about this same value (0.65 - 0.70) suggests there is a fundamental limitation in the separability of the data which precludes better performance. This idea should be tested further in later studies. The fact that CLASSY had, on the average, only about 9 clusters, whereas AMOEBA had about 17, and ISOCLS had almost 37 is seen as important. Given the same overall level of performance, an economy in the number of clusters produced is to be preferred.

The cluster-labeling techniques appear to suffer from the same fate. The proportion estimates obtained using these techniques were generally biased; the R-values were always greater than 0.9 and typically they were greater than 1. This poor performance for all of the clustering algorithms indicates that clusters were simply not pure enough for cluster labeling to function efficiently as a proportion-estimation technique. For all three clustering algorithms, the average PCC value, which may be thought of as a measure of cluster purity, was about 0.85. Apparently, much greater cluster purity is needed for cluster labeling to be a viable approach.

The stratified proportion-estimation techniques generally worked well. The sequential allocation approach with no prior distribution on cluster purities produced good results for an allocation of 30 pixels; however, the results for allocations of 60, 90, and 120 pixels were biased and had much larger reduction in mean-square error values for all of the clustering algorithms. In addition, these results were obtained with an initial allocation of three pixels per cluster, which means that in many cases, sequential allocation did not begin until more than 30 pixels had been allocated.

The study eventually focused on a comparison of the Bayesian sequential allocation technique and the proportional allocation technique for stratified proportion estimation. Both of these techniques are unbiased. The proportional allocation technique has an R value of about 0.67 which does not differ significantly from algorithm to algorithm or for different numbers of pixels allocated. This result is also not much different from the Procedure 1 value. However, the Bayesian sequential allocation technique, when used with the CLASSY or ISOCLS clustering algorithm, has significantly lower reduction in mean-square-error values than does proportional allocation. The fact that CLASSY has many fewer clusters than ISOCLS and, thus, is able to begin allocating sequentially at a much lower number of dots makes it the preferred algorithm.

The recommendation of this report is that studies be undertaken to determine how best to implement stratified proportion estimation using CLASSY clusters as the strata and the Bayesian sequential technique for pixel allocation. It appears that a total allocation of 30 pixels would achieve the minimum R. The average mean-square error for this number of pixels is 0.002853, which compares very favorably with the average variance of 0.002515 calculated from the results of the Procedure 1 secondary error analysis study (ref. 3). This variance for Procedure 1 was obtained with about 100 labeled pixels for each estimate ($\approx$ 40 type 1 pixels plus $\approx$ 60 type 2 pixels). Thus, an allocation of only 30 total dots represents a very clear advantage for the proposed replacement procedure for Procedure 1.

## 7. REFERENCES


1. Carnes, J. G.: Detailed Analysis of CAMS Procedures for Phase III Using Ground-truth Inventories. LEC-13343, April 1979, pp. 18.

2. Havens, K. A.: Secondary Error Analysis: The Evaluation of Analyst Dot Labeling. LEC-12380, September 1978, pp. 17.

3. Havens, K. A.: Further Evaluation of Procedure 1 Secondary Error Analysis. LEC-13180, May 1979, pp. 39.

4. 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, pp. 37.

5. Lennington, R. K. and Rassbach, M. E.: CLASSY — An Adaptive Maximum Likelihood Clustering Algorithm. LEC-12145, May 1978, pp. 37 [Presented at the Ninth Annual Meeting of the Classification Society (North American Branch), Clemson University (Clemson, South Carolina), May 21-23, 1978].

6. Lennington, R. K. and Rassbach, M. E.: Mathematical Description and Program Documentation for CLASSY, An Adaptive Maximum Likelihood Clustering Method. LEC-12177, April 1979, pp. 63.

7. Bryant, J.: On the Clustering of Multidimensional Pictoral Data. Pattern Recognition, vol. 11, 1979, pp. 115-125.

8. Kan, E. P.: The JSC Clustering Program ISOCLS and ITS Applications. LEC-0483, July 1973, pp. 57.

9. Ball, G. H. and Hall, D. J.: A Clustering Technique for Summarizing Multivariate Data. Behavioral Science, vol. 12, March 1967, pp. 153-155.

10. Wiley, A. D. and Bean, W. C.: MPAD LACIE Clustering Parameter Study, JSC Internal Note 76-FM-116.

11. Kauth, R. J. and Thomas, G. S.: The Tasselled Cap — A Graphic Description of the Spectral Temporal Development of Agricultural Crops as Seen by Landsat. Proceedings of the Symposium on the Machine Processing of Remotely Sensed Data, Purdue University, June 29 - July 1, 1976.

12. Pore, M. D.: On Evaluating Clustering Procedures for Use in Classification. 1978 Annual Meeting of the American Statistical Association, LEC-12171, August 1978, pp. 23.

13. Pore, M. D.: Bayesian Techniques in Stratified Proportion Estimation. 1979 Annual Meeting of the American Statistical Association, LEC-13490, August 1979, pp. 22.

# APPENDIX

## CALCULATION RESULTS OF THE AVERAGE BIAS IN THE PROPORTION ESTIMATE, THE MEAN-SQUARE ERROR OF THE ESTIMATE, AND THE VARIANCE REDUCTION FACTOR AS COMPARED TO SIMPLE RANDOM SAMPLING

MAJORITY RULE LABELING USING PROPORTIONAL ALLOCATION

| 1005 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.022320 | 0.025340 | 0.007720 |
| 60 | 0.030240 | 0.015500 | -0.025220 |
| 90 | 0.014950 | -0.014170 | -0.023670 |
| 120 | 0.011620 | 0.002390 | 0.014130 |
| | | MSE | |
| 30 | 0.025267 | 0.030218 | 0.008555 |
| 60 | 0.014734 | 0.019830 | 0.024368 |
| 90 | 0.009784 | 0.024408 | 0.020461 |
| 120 | 0.010105 | 0.012791 | 0.004421 |
| | | REL.MSE | |
| 30 | 3.339663 | 3.944055 | 1.130924 |
| 60 | 3.894919 | 5.242114 | 6.441793 |
| 90 | 3.879445 | 9.678555 | 8.113338 |
| 120 | 5.342396 | 6.762604 | 2.337595 |

| 1853 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -0.026150 | -0.015330 | -0.003830 |
| 60 | -0.035980 | -0.015740 | -0.003150 |
| 90 | -0.353400 | -0.015230 | -0.002120 |
| 120 | -0.047820 | -0.016170 | -0.004230 |
| | | MSE | |
| 30 | 0.016611 | 0.010254 | 0.010463 |
| 60 | 0.031794 | 0.009045 | 0.009296 |
| 90 | 0.029800 | 0.008451 | 0.005124 |
| 120 | 0.052291 | 0.009588 | 0.006041 |
| | | REL.MSE | |
| 30 | 2.346127 | 1.448185 | 1.477779 |
| 60 | 8.990927 | 2.569089 | 2.625855 |
| 90 | 12.626630 | 3.580799 | 2.171080 |
| 120 | 29.541494 | 5.416903 | 3.412844 |

| 1520 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -0.047980 | -0.045070 | 0.007990 |
| 60 | -0.068670 | -0.031460 | -0.035680 |
| 90 | -0.100460 | -0.029430 | -0.042900 |
| 120 | -0.088860 | -0.033320 | -0.041500 |
| | | MSE | |
| 30 | 0.050586 | 0.059013 | 0.005191 |
| 60 | 0.113270 | 0.026398 | 0.034730 |
| 90 | 0.211542 | 0.023371 | 0.046732 |
| 120 | 0.159811 | 0.030093 | 0.048556 |
| | | RED.MSE | |
| 30 | 7.218562 | 8.421125 | 0.740695 |
| 60 | 32.326459 | 7.533875 | 9.911716 |
| 90 | 90.560196 | 10.005219 | 20.005707 |
| 120 | 91.219086 | 17.177032 | 27.715195 |

| 1231 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.026170 | 0.039290 | 0.038580 |
| 60 | 0.078530 | -0.044910 | -0.017190 |
| 90 | 0.046750 | -0.032100 | -0.011820 |
| 120 | 0.024800 | -0.038390 | 0.014950 |
| | | MSE | |
| 30 | 0.011585 | 0.027813 | 0.025340 |
| 60 | 0.110459 | 0.071193 | 0.025126 |
| 90 | 0.031956 | 0.042902 | 0.013481 |
| 120 | 0.009694 | 0.056304 | 0.005102 |
| | | RED.MSE | |
| 30 | 1.826858 | 4.365946 | 3.996099 |
| 60 | 34.837662 | 22.453934 | 7.924449 |
| 90 | 15.117999 | 20.296449 | 6.377600 |
| 120 | 6.115053 | 35.515686 | 3.849034 |

| 1060 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -0.021900 | -0.082230 | 0.017710 |
| 60 | 0.005070 | -0.053670 | -0.042910 |
| 90 | 0.035600 | -0.083890 | -0.054250 |
| 120 | 0.016120 | -0.042350 | -0.066400 |
| | | MSE | |
| 30 | 0.013919 | 0.157984 | 0.008255 |
| 60 | 0.003252 | 0.064340 | 0.052782 |
| 90 | 0.027979 | 0.151259 | 0.062433 |
| 120 | 0.007745 | 0.140949 | 0.049953 |
| | | RED.MSE | |
| 30 | 3.193859 | 26.671066 | 1.343622 |
| 60 | 1.098066 | 21.723984 | 17.821579 |
| 90 | 14.170469 | 76.607086 | 31.645645 |
| 120 | 5.229753 | 95.181076 | 67.496613 |

| AVERAGES | | | VARIANCES | | |
|---|---|---|---|---|---|
| CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
| BIAS | | | | | |
| -0.009508 | -0.015600 | 0.013634 | 0.000839 | 0.001999 | 0.000202 |
| 0.001838 | -0.026056 | -0.024830 | 0.002620 | 0.000546 | 0.000195 |
| -0.071312 | -0.034964 | -0.026952 | 0.022647 | 0.000651 | 0.000371 |
| -0.016828 | -0.033568 | -0.016600 | 0.001955 | 0.000800 | 0.001039 |
| MSE | | | | | |
| 0.024594 | 0.057056 | 0.011561 | 0.000188 | 0.002791 | 0.000050 |
| 0.054702 | 0.038171 | 0.029260 | 0.002262 | 0.000619 | 0.000205 |
| 0.062212 | 0.050078 | 0.029656 | 0.005637 | 0.002679 | 0.000463 |
| 0.047929 | 0.049945 | 0.033015 | 0.003409 | 0.002345 | 0.001398 |
| RED.MSE | | | | | |
| 3.585012 | 8.984081 | 1.747804 | 3.608364 | 83.195801 | 1.329618 |
| 16.227676 | 11.904598 | 8.945074 | 207.806641 | 71.670441 | 25.393509 |
| 27.270935 | 24.033615 | 13.662670 | 1017.242236 | 719.822998 | 115.909088 |
| 27.489548 | 32.010651 | 20.962250 | 1101.703857 | 1113.502686 | 631.753662 |

MAJORITY RULE LABELING USING SEQUENTIAL ALLOCATION

| 1005 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.091846 | -0.062759 | -0.053459 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.004277 | 0.006663 | 0.003432 |
| | | RED.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 2.740766 | 3.629826 | 3.652670 |

| 1853 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.043749 | -0.007239 | -0.036802 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.002467 | 0.000052 | 0.003061 |
| | | RED.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.623225 | 0.007596 | 3.970898 |

| 1520 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.062555 | -0.013257 | -0.044603 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.009528 | 0.000176 | 0.002624 |
| | | REL.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 2.874748 | 0.023190 | 3.340642 |

| 1231 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.006490 | -0.028416 | 0.028109 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.003620 | 0.001527 | 0.001045 |
| | | REL.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.830390 | 1.073611 | 1.272033 |

| 1060 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.030815 | -0.059542 | -0.053317 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.003942 | 0.004325 | 0.003170 |
| | | REL.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 1.311178 | 1.472968 | 4.836786 |

| | AVERAGES | | | VARIANCES | | |
|---|---|---|---|---|---|---|
| | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
| | | | BIAS | | | |
| 30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 120 | -.04449496 | -.03424257 | -.03201438 | 0.00107109 | 0.00053136 | 0.00094198 |
| | | | MSE | | | |
| 30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 120 | 0.00574680 | 0.00254860 | 0.00266640 | 0.00000913 | 0.00000660 | 0.00000073 |
| | | | RED.MSE | | | |
| 30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 120 | 1.67406068 | 1.24144173 | 3.41440514 | 0.90543842 | 1.75853252 | 1.39696312 |

## MAJORITY RULE LABELING USING BAYESIAN SEQUENTIAL ALLOCATION

| 1005 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.067757 | -0.060338 | -0.049491 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.007431 | 0.009865 | 0.003938 |
| | | REL.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 1.288348 | 2.939078 | 2.139337 |

| 1853 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.025905 | -0.009112 | -0.024826 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.003156 | 0.003551 | 0.002482 |
| | | REL.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.420471 | 0.333351 | 1.481780 |

| 1520 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.043679 | -0.023170 | -0.039418 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.008590 | 0.008494 | 0.002910 |
| | | RED.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 1.384761 | 0.725148 | 1.706895 |

| 1231 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.007739 | -0.009436 | 0.028156 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.005074 | 0.002757 | 0.001490 |
| | | RED.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.621741 | 1.077502 | 0.923436 |

| 1060 | CLASSY | AMOEBA | ISOCLS |
| --- | --- | --- | --- |
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | -0.034276 | -0.041143 | -0.044665 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.005470 | 0.004066 | 0.012577 |
| | | REL.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 |
| 120 | 0.840094 | 1.852485 | 2.010332 |

| | AVERAGES | | | VARIANCES | | |
|---|---|---|---|---|---|---|
| | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
| | | | BIAS | | | |
| 30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 120 | -.03277557 | -.02864778 | -.02584878 | 0.00060669 | 0.00038843 | 0.00079368 |
| | | | MSE | | | |
| 30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 120 | 0.00604460 | 0.00682659 | 0.00267940 | 0.00000393 | 0.00000916 | 0.00000062 |
| | | | RED.MSE | | | |
| 30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 60 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 90 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 120 | 0.91108280 | 1.38561249 | 1.65233707 | 0.13923180 | 0.85401917 | 0.18573952 |



ll

# STRATIFIED PROPORTION ESTIMATION USING PROPORTIONAL ALLOCATION

| SEG | DOTS | CLASSY VAR | AMOEBA VAR | ISOCLS VAR | CLASSY VAR RD | AMOEBA VAR RD | ISOCLS VAR RD |
|---|---|---|---|---|---|---|---|
| 1005 | 30 | 0.003293 | 0.004616 | 0.006920 | 0.435271 | 0.610079 | 0.914715 |
| 1005 | 60 | 0.002272 | 0.002301 | 0.002033 | 0.600677 | 0.608259 | 0.537301 |
| 1005 | 90 | 0.001636 | 0.001997 | 0.001944 | 0.648707 | 0.791924 | 0.770715 |
| 1005 | 120 | 0.001167 | 0.001438 | 0.001147 | 0.616830 | 0.760133 | 0.606595 |
| 1853 | 30 | 0.002134 | 0.003614 | 0.004163 | 0.301395 | 0.510395 | 0.588025 |
| 1853 | 60 | 0.001522 | 0.001484 | 0.002189 | 0.429885 | 0.419259 | 0.618401 |
| 1853 | 90 | 0.000870 | 0.001064 | 0.001479 | 0.368437 | 0.450880 | 0.626344 |
| 1853 | 120 | 0.000745 | 0.000990 | 0.001147 | 0.420801 | 0.559047 | 0.647838 |
| 1231 | 30 | 0.003331 | 0.003090 | 0.003341 | 0.525366 | 0.487349 | 0.526881 |
| 1231 | 60 | 0.001075 | 0.001521 | 0.001073 | 0.338912 | 0.479664 | 0.338534 |
| 1231 | 90 | 0.000945 | 0.001015 | 0.001238 | 0.446852 | 0.480233 | 0.585654 |
| 1231 | 120 | 0.000870 | 0.000586 | 0.000711 | 0.548598 | 0.369457 | 0.448341 |
| 1060 | 30 | 0.004071 | 0.003369 | 0.003648 | 0.687223 | 0.568801 | 0.615791 |
| 1060 | 60 | 0.001718 | 0.002044 | 0.001465 | 0.579929 | 0.690218 | 0.494569 |
| 1060 | 90 | 0.001079 | 0.001329 | 0.001079 | 0.546337 | 0.673256 | 0.546269 |
| 1060 | 120 | 0.000919 | 0.000844 | 0.000810 | 0.620411 | 0.569787 | 0.546967 |
| 1520 | 30 | 0.003945 | 0.003094 | 0.004698 | 0.562938 | 0.441525 | 0.670451 |
| 1520 | 60 | 0.002034 | 0.001590 | 0.001798 | 0.580486 | 0.453894 | 0.513131 |
| 1520 | 90 | 0.001254 | 0.001031 | 0.001375 | 0.536647 | 0.441337 | 0.588680 |
| 1520 | 120 | 0.000952 | 0.000818 | 0.000835 | 0.543179 | 0.466786 | 0.476820 |
| 1604 | 30 | 0.010260 | 0.005781 | 0.005145 | 1.234115 | 0.695336 | 0.618851 |
| 1604 | 60 | 0.002811 | 0.002857 | 0.002485 | 0.676191 | 0.687406 | 0.597702 |
| 1604 | 90 | 0.002449 | 0.002119 | 0.002226 | 0.883549 | 0.764716 | 0.803099 |
| 1604 | 120 | 0.001959 | 0.001585 | 0.001566 | 0.894510 | 0.762431 | 0.801775 |
| 1675 | 30 | 0.004483 | 0.005450 | 0.004363 | 0.652401 | 0.793106 | 0.635603 |
| 1675 | 60 | 0.002319 | 0.002886 | 0.002347 | 0.674841 | 0.839983 | 0.682925 |
| 1675 | 90 | 0.001687 | 0.002028 | 0.001889 | 0.736526 | 0.885288 | 0.824630 |
| 1675 | 120 | 0.001077 | 0.001516 | 0.001066 | 0.626592 | 0.882610 | 0.620400 |
| 1805 | 30 | 0.002646 | 0.001876 | 0.001937 | 0.578474 | 0.410208 | 0.423861 |
| 1805 | 60 | 0.001222 | 0.001307 | 0.000868 | 0.534477 | 0.571459 | 0.379362 |
| 1805 | 90 | 0.000764 | 0.000646 | 0.000782 | 0.501006 | 0.423434 | 0.512904 |
| 1805 | 120 | 0.000533 | 0.000662 | 0.000494 | 0.466463 | 0.578475 | 0.431621 |
| 1577 | 30 | 0.000951 | 0.000827 | 0.000879 | 1.005363 | 0.875085 | 0.929719 |
| 1577 | 60 | 0.000383 | 0.000434 | 0.000550 | 0.810335 | 0.918121 | 1.164354 |
| 1577 | 90 | 0.000349 | 0.000313 | 0.000391 | 1.107220 | 0.994042 | 1.239814 |
| 1577 | 120 | 0.000229 | 0.000195 | 0.000206 | 0.969019 | 0.828247 | 0.870670 |
| 1606 | 30 | 0.004854 | 0.004377 | 0.004220 | 0.658850 | 0.594103 | 0.572803 |
| 1606 | 60 | 0.002851 | 0.002297 | 0.002407 | 0.773946 | 0.623534 | 0.653471 |
| 1606 | 90 | 0.001752 | 0.001644 | 0.001969 | 0.713459 | 0.669295 | 0.801968 |
| 1606 | 120 | 0.001236 | 0.001306 | 0.001756 | 0.671128 | 0.709219 | 0.953581 |
| 1661 | 30 | 0.007599 | 0.006341 | 0.004819 | 0.939657 | 0.784108 | 0.595749 |
| 1661 | 60 | 0.003084 | 0.002262 | 0.002711 | 0.762752 | 0.559506 | 0.670387 |
| 1661 | 90 | 0.002590 | 0.002064 | 0.004818 | 0.960878 | 0.765827 | 0.595749 |
| 1661 | 120 | 0.001304 | 0.001549 | 0.002711 | 0.645045 | 0.766334 | 0.670387 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1686 | 30 | 0.004382 | 0.003372 | 0.002527 | 0.842327 | 0.648202 | 0.465787 |
| 1686 | 60 | 0.001870 | 0.002357 | 0.001870 | 0.718877 | 0.905936 | 0.718908 |
| 1686 | 90 | 0.001239 | 0.001387 | 0.001275 | 0.714290 | 0.799922 | 0.735030 |
| 1686 | 120 | 0.000896 | 0.001136 | 0.000754 | 0.688625 | 0.873790 | 0.579641 |
| 1803 | 30 | 0.000999 | 0.000831 | 0.000936 | 0.966918 | 0.805070 | 0.906503 |
| 1803 | 60 | 0.000425 | 0.000308 | 0.000383 | 0.922259 | 0.595521 | 0.741064 |
| 1803 | 90 | 0.000311 | 0.000261 | 0.000248 | 0.903196 | 0.759477 | 0.719147 |
| 1803 | 120 | 0.000179 | 0.000208 | 0.000161 | 0.691385 | 0.806631 | 0.622639 |
| 1899 | 30 | 0.003756 | 0.004340 | 0.004334 | 0.468194 | 0.540992 | 0.540168 |
| 1899 | 60 | 0.001873 | 0.001662 | 0.001547 | 0.466976 | 0.414184 | 0.398066 |
| 1899 | 90 | 0.001350 | 0.001368 | 0.001194 | 0.504684 | 0.511425 | 0.446331 |
| 1899 | 120 | 0.000882 | 0.001044 | 0.000926 | 0.439648 | 0.520342 | 0.461794 |
| 1032 | 30 | 0.002914 | 0.003676 | 0.004294 | 0.374400 | 0.472428 | 0.551776 |
| 1032 | 60 | 0.001353 | 0.002096 | 0.002175 | 0.347800 | 0.538653 | 0.559013 |
| 1032 | 90 | 0.000906 | 0.001264 | 0.001003 | 0.349293 | 0.487394 | 0.386791 |
| 1032 | 120 | 0.000749 | 0.000969 | 0.000648 | 0.384830 | 0.497900 | 0.435841 |
| 1033 | 30 | 0.002615 | 0.001851 | 0.001998 | 0.911555 | 0.645327 | 0.696547 |
| 1033 | 60 | 0.001079 | 0.000780 | 0.001343 | 0.752289 | 0.543604 | 0.971323 |
| 1033 | 90 | 0.000494 | 0.000626 | 0.000764 | 0.935079 | 0.654877 | 0.799077 |
| 1033 | 120 | 0.000522 | 0.000531 | 0.000476 | 0.727264 | 0.740112 | 0.663843 |
| 1059 | 30 | 0.003643 | 0.003368 | 0.003593 | 0.444114 | 0.410574 | 0.438039 |
| 1059 | 60 | 0.001900 | 0.001553 | 0.001415 | 0.463259 | 0.378579 | 0.345070 |
| 1059 | 90 | 0.001480 | 0.001390 | 0.001139 | 0.541455 | 0.508307 | 0.416753 |
| 1059 | 120 | 0.000856 | 0.000875 | 0.000969 | 0.417471 | 0.426932 | 0.472455 |
| 1166 | 30 | 0.001590 | 0.001656 | 0.002025 | 0.772634 | 0.804740 | 0.984205 |
| 1166 | 60 | 0.000776 | 0.000632 | 0.000906 | 0.753901 | 0.614148 | 0.980271 |
| 1166 | 90 | 0.000444 | 0.000479 | 0.000688 | 0.647189 | 0.698611 | 1.002699 |
| 1166 | 120 | 0.000402 | 0.000462 | 0.000342 | 0.780424 | 0.898565 | 0.665585 |
| 1239 | 30 | 0.003817 | 0.003167 | 0.002754 | 0.823615 | 0.683444 | 0.594169 |
| 1239 | 60 | 0.001780 | 0.001631 | 0.001696 | 0.767967 | 0.703992 | 0.732087 |
| 1239 | 90 | 0.001300 | 0.000850 | 0.001134 | 0.841456 | 0.550304 | 0.734311 |
| 1239 | 120 | 0.000926 | 0.000789 | 0.000614 | 0.799102 | 0.681252 | 0.530059 |
| 1367 | 30 | 0.004419 | 0.004675 | 0.003899 | 0.555033 | 0.587196 | 0.489705 |
| 1367 | 60 | 0.002534 | 0.002101 | 0.001727 | 0.636640 | 0.527727 | 0.433927 |
| 1367 | 90 | 0.001805 | 0.001867 | 0.001320 | 0.680055 | 0.703660 | 0.497247 |
| 1367 | 120 | 0.000978 | 0.001039 | 0.001318 | 0.491563 | 0.521841 | 0.662124 |
| 1512 | 30 | 0.005209 | 0.006056 | 0.004377 | 0.696581 | 0.809989 | 0.585357 |
| 1512 | 60 | 0.003254 | 0.004010 | 0.002948 | 0.870270 | 1.072700 | 0.788530 |
| 1512 | 90 | 0.002235 | 0.001917 | 0.002388 | 0.896620 | 0.769146 | 0.958286 |
| 1512 | 120 | 0.001295 | 0.001313 | 0.001761 | 0.928910 | 0.702395 | 0.942318 |

AVERAGES

| SEG DOTS | CLASSY VAR | AMOEBA VAR | ISOCLS VAR | CLASSY VAR RO | AMOEBA VAR RO | ISOCLS VAR RO |
|---|---|---|---|---|---|---|
| 30 | .003852895 | .003591756 | .003565516 | .687449038 | .627526164 | .636414111 |
| 60 | .001815951 | .001814903 | .001715998 | .636317074 | .626016080 | .629446924 |
| 90 | .001301855 | .001269474 | .001444855 | .684710690 | .656349719 | .694832742 |
| 120 | .000884570 | .000945522 | .000986570 | .636751771 | .662965417 | .624346912 |

VARIANCES

| SEG DOTS | CLASSY VAR | AMOEBA VAR | ISOCLS VAR | CLASSY VAR RO | AMOEBA VAR RO | ISOCLS VAR RO |
|---|---|---|---|---|---|---|
| 30 | .000004197 | .000002433 | .000002063 | .053946018 | .019914806 | .025356948 |
| 60 | .000000648 | .000000738 | .000000464 | .023404247 | .031225204 | .042545319 |
| 90 | .000000391 | .000000339 | .000000871 | .041802545 | .024449527 | .042262435 |
| 120 | .000000143 | .000000164 | .000000350 | .028034508 | .024315834 | .023863912 |

## STRATIFIED PROPORTION ESTIMATION USING SEQUENTIAL ALLOCATION

| 1005 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | -.027710 | 0.0 | 0.0 |
| 90 | -.034370 | -.021960 | 0.0 |
| 120 | -.036320 | -.027500 | -.012560 |
| | | MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.003363 | 0.0 | 0.0 |
| 90 | 0.002956 | 0.004429 | 0.0 |
| 120 | 0.003031 | 0.004066 | 0.001418 |
| | | RED.MSE | |
| 30 | 0.0 | 0.0 | 0.0 |
| 60 | 0.888919 | 0.0 | 0.0 |
| 90 | 1.172315 | 1.756413 | 0.0 |
| 120 | 1.602329 | 2.149695 | 0.749571 |

| 1853 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.006800 | -.001500 | 0.0 |
| 60 | -.021500 | -.012080 | 0.0 |
| 90 | -.024060 | -.012430 | 0.0 |
| 120 | -.024110 | -.012680 | -.012360 |
| | | MSE | |
| 30 | 0.003204 | 0.005062 | 0.0 |
| 60 | 0.002274 | 0.003084 | 0.0 |
| 90 | 0.002045 | 0.003027 | 0.0 |
| 120 | 0.001935 | 0.002862 | 0.001736 |
| | | RED.MSE | |
| 30 | 0.452484 | 0.714945 | 0.0 |
| 60 | 0.642237 | 0.871107 | 0.0 |
| 90 | 0.866479 | 1.282661 | 0.0 |
| 120 | 1.092956 | 1.617064 | 0.980708 |

| 1520 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.0 | -.010200 | 0.0 |
| 60 | -.021030 | -.020090 | 0.0 |
| 90 | -.025329 | -.021040 | 0.0 |
| 120 | -.028440 | -.023090 | 0.001900 |
| | | MSE | |
| 30 | 0.0 | 0.005203 | 0.0 |
| 60 | 0.003670 | 0.003297 | 0.0 |
| 90 | 0.003699 | 0.002866 | 0.0 |
| 120 | 0.003668 | 0.002810 | 0.001142 |
| | | RED.MSE | |
| 30 | 0.0 | 0.743120 | 0.0 |
| 60 | 1.047401 | 0.941039 | 0.0 |
| 90 | 1.583390 | 1.227132 | 0.0 |
| 120 | 2.093648 | 1.603925 | 0.651783 |

| 1231 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.016340 | 0.0 | 0.0 |
| 60 | 0.023840 | 0.0 | 0.0 |
| 90 | 0.024200 | -.005140 | 0.0 |
| 120 | 0.022940 | -.003040 | 0.007620 |
| | | MSE | |
| 30 | 0.003073 | 0.0 | 0.0 |
| 60 | 0.002779 | 0.0 | 0.0 |
| 90 | 0.002505 | 0.001447 | 0.0 |
| 120 | 0.002319 | 0.001146 | 0.000687 |
| | | RED.MSE | |
| 30 | 0.484640 | 0.0 | 0.0 |
| 60 | 0.876588 | 0.0 | 0.0 |
| 90 | 1.185272 | 0.684467 | 0.0 |
| 120 | 1.462873 | 0.722813 | 0.433132 |

| 1060 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.012230 | 0.0 | 0.0 |
| 60 | -.024400 | -.035290 | 0.0 |
| 90 | -.029050 | -.039940 | 0.0 |
| 120 | -.031570 | -.042390 | 0.0 |
| | | MSE | |
| 30 | 0.004076 | 0.0 | 0.0 |
| 60 | 0.002740 | 0.003396 | 0.0 |
| 90 | 0.002692 | 0.003143 | 0.0 |
| 120 | 0.002774 | 0.002965 | 0.0 |
| | | RED.MSE | |
| 30 | 0.689127 | 0.0 | 0.0 |
| 60 | 0.924998 | 1.146745 | 0.0 |
| 90 | 1.363269 | 1.591858 | 0.0 |
| 120 | 1.873245 | 2.002310 | 0.0 |

| | AVERAGES | | | VARIANCES | | |
|---|---|---|---|---|---|---|
| | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
| | | | BIAS | | | |
| 30 | -.00089333 | -.00585000 | 0.0 | 0.00015393 | 0.00003784 | 0.0 |
| 60 | -.01415999 | -.02249665 | 0.0 | 0.00036671 | 0.00009266 | 0.0 |
| 90 | -.01781999 | -.02010199 | 0.0 | 0.00045373 | 0.00013612 | 0.0 |
| 120 | -.01948998 | -.02173998 | -.00385000 | 0.00046703 | 0.00017864 | 0.00007823 |
| | | | MSE | | | |
| 30 | 0.00345100 | 0.00513500 | 0.0 | 0.00000020 | 0.00000001 | 0.0 |
| 60 | 0.00296520 | 0.00325900 | 0.0 | 0.00000024 | 0.00000002 | 0.0 |
| 90 | 0.00277940 | 0.00298240 | 0.0 | 0.00000030 | 0.00000090 | 0.0 |
| 120 | 0.00274540 | 0.00276980 | 0.00124575 | 0.00000035 | 0.00000087 | 0.00000015 |
| | | | RED.MSE | | | |
| 30 | 0.54175025 | 0.72903204 | 0.0 | 0.01088542 | 0.00034721 | 0.0 |
| 60 | 0.87602842 | 0.98629665 | 0.0 | 0.01731825 | 0.01368725 | 0.0 |
| 90 | 1.23414421 | 1.30850601 | 0.0 | 0.05600834 | 0.13552380 | 0.0 |
| 120 | 1.62500954 | 1.61916065 | 0.70379806 | 0.11672701 | 0.24639034 | 0.03868544 |

## STRATIFIED PROPORTION ESTIMATION USING BAYESIAN SEQUENTIAL ALLOCATION

| 1005 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.009360 | 0.0 | 0.0 |
| 60 | 0.006340 | -.012140 | 0.0 |
| 90 | 0.002710 | -.005480 | -.008870 |
| 120 | 0.003610 | 0.000030 | -.005340 |
| | | MSE | |
| 30 | 0.002561 | 0.0 | 0.0 |
| 60 | 0.001308 | 0.003162 | 0.0 |
| 90 | 0.000945 | 0.001211 | 0.001180 |
| 120 | 0.000683 | 0.000852 | 0.000882 |
| | | REL. MSE | |
| 30 | 0.338534 | 0.0 | 0.0 |
| 60 | 0.345714 | 0.835860 | 0.0 |
| 90 | 0.335097 | 0.430354 | 0.467731 |
| 120 | 0.361120 | 0.450488 | 0.466577 |

| 1853 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.021540 | 0.017160 | 0.0 |
| 60 | 0.020110 | 0.017620 | 0.0 |
| 90 | 0.016380 | 0.012260 | 0.004750 |
| 120 | 0.012170 | 0.012360 | 0.005440 |
| | | MSE | |
| 30 | 0.002295 | 0.003676 | 0.0 |
| 60 | 0.001037 | 0.001712 | 0.0 |
| 90 | 0.000756 | 0.001091 | 0.001366 |
| 120 | 0.000520 | 0.000896 | 0.001021 |
| | | REL. MSE | |
| 30 | 0.324193 | 0.519185 | 0.0 |
| 60 | 0.292972 | 0.486245 | 0.0 |
| 90 | 0.320502 | 0.462054 | 0.574828 |
| 120 | 0.293555 | 0.506083 | 0.576932 |

| 1231 | CLASSY | AMOEBA | ISOCLS |
| --- | --- | --- | --- |
| | | BIAS | |
| 30 | 0.003380 | 0.0 | 0.0 |
| 60 | -.001640 | -.007540 | 0.0 |
| 90 | -.001800 | -.005470 | -.024740 |
| 120 | -.003350 | -.001700 | -.024110 |
| | | MSE | |
| 30 | 0.002027 | 0.0 | 0.0 |
| 60 | 0.001101 | 0.001837 | 0.0 |
| 90 | 0.000669 | 0.001120 | 0.001295 |
| 120 | 0.000464 | 0.000753 | 0.001049 |
| | | RED.MSE | |
| 30 | 0.319673 | 0.0 | 0.0 |
| 60 | 0.347333 | 0.579233 | 0.0 |
| 90 | 0.316647 | 0.529677 | 0.612437 |
| 120 | 0.292488 | 0.475198 | 0.661553 |

| 1060 | CLASSY | AMOEBA | ISOCLS |
| --- | --- | --- | --- |
| | | BIAS | |
| 30 | 0.010300 | -.025250 | 0.0 |
| 60 | 0.005400 | -.015290 | 0.0 |
| 90 | 0.002320 | -.010150 | -.009080 |
| 120 | 0.001130 | -.007250 | -.014980 |
| | | MSE | |
| 30 | 0.001862 | 0.003832 | 0.0 |
| 60 | 0.001002 | 0.001539 | 0.0 |
| 90 | 0.000670 | 0.001135 | 0.001010 |
| 120 | 0.000640 | 0.000955 | 0.000908 |
| | | RED.MSE | |
| 30 | 0.314344 | 0.646961 | 0.0 |
| 60 | 0.338304 | 0.519750 | 0.0 |
| 90 | 0.330151 | 0.574634 | 0.511478 |
| 120 | 0.432498 | 0.645195 | 0.613482 |

| 1520 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.004220 | -.014120 | 0.0 |
| 60 | 0.007260 | -.004610 | 0.0 |
| 90 | 0.007530 | -.003650 | 0.009590 |
| 120 | 0.009360 | -.002580 | 0.011520 |
| | | MSE | |
| 30 | 0.002769 | 0.003039 | 0.0 |
| 60 | 0.001418 | 0.001665 | 0.0 |
| 90 | 0.000844 | 0.001287 | 0.000739 |
| 120 | 0.000642 | 0.000921 | 0.000646 |
| | | REL.MSE | |
| 30 | 0.395182 | 0.433603 | 0.0 |
| 60 | 0.404696 | 0.475062 | 0.0 |
| 90 | 0.361107 | 0.550816 | 0.316236 |
| 120 | 0.366615 | 0.525868 | 0.368449 |

| 1604 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.010960 | 0.023940 | 0.0 |
| 60 | 0.011720 | 0.021550 | 0.0 |
| 90 | 0.002640 | 0.013670 | -.004570 |
| 120 | 0.000940 | 0.008080 | -.000880 |
| | | MSE | |
| 30 | 0.007777 | 0.007740 | 0.0 |
| 60 | 0.003058 | 0.003494 | 0.0 |
| 90 | 0.002223 | 0.002467 | 0.001906 |
| 120 | 0.001725 | 0.001764 | 0.001298 |
| | | REL.MSE | |
| 30 | 0.935491 | 0.930942 | 0.0 |
| 60 | 0.735613 | 0.840631 | 0.0 |
| 90 | 0.802331 | 0.890052 | 0.720130 |
| 120 | 0.829968 | 0.948858 | 0.624377 |

| 1675 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.000450 | 0.0 | 0.0 |
| 60 | -.003150 | 0.0 | 0.0 |
| 90 | -.004060 | -.021930 | -.001770 |
| 120 | -.002740 | -.022030 | -.005100 |
| | | MSE | |
| 30 | 0.003561 | 0.0 | 0.0 |
| 60 | 0.001845 | 0.0 | 0.0 |
| 90 | 0.001214 | 0.002047 | 0.001303 |
| 120 | 0.000588 | 0.001702 | 0.000923 |
| | | REL.MSE | |
| 30 | 0.518149 | 0.0 | 0.0 |
| 60 | 0.536885 | 0.0 | 0.0 |
| 90 | 0.530038 | 0.893594 | 0.569016 |
| 120 | 0.400639 | 0.990663 | 0.537458 |

| 1805 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.000230 | 0.0 | 0.0 |
| 60 | -.001510 | 0.0 | 0.0 |
| 90 | -.000920 | -.025350 | -.004170 |
| 120 | -.000590 | -.026230 | -.005540 |
| | | MSE | |
| 30 | 0.001757 | 0.0 | 0.0 |
| 60 | 0.001027 | 0.0 | 0.0 |
| 90 | 0.000571 | 0.001664 | 0.000641 |
| 120 | 0.000560 | 0.001310 | 0.000488 |
| | | REL.MSE | |
| 30 | 0.384140 | 0.0 | 0.0 |
| 60 | 0.448429 | 0.0 | 0.0 |
| 90 | 0.440080 | 1.091645 | 0.420153 |
| 120 | 0.490000 | 1.145226 | 0.427027 |

| 1577 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.002710 | 0.0 | 0.0 |
| 60 | -.007480 | -.005540 | 0.0 |
| 90 | -.008090 | -.007680 | -.001900 |
| 120 | -.009020 | -.009840 | -.004470 |
| | | MSE | |
| 30 | 0.000560 | 0.0 | 0.0 |
| 60 | 0.000300 | 0.000474 | 0.0 |
| 90 | 0.000190 | 0.000296 | 0.000163 |
| 120 | 0.000176 | 0.000239 | 0.000133 |
| | | RED.MSE | |
| 30 | 0.592409 | 0.0 | 0.0 |
| 60 | 0.634560 | 1.003359 | 0.0 |
| 90 | 0.603320 | 0.439086 | 0.516164 |
| 120 | 0.743240 | 1.011619 | 0.563520 |

| 1606 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.002880 | -.004400 | 0.0 |
| 60 | 0.012060 | -.000200 | 0.0 |
| 90 | 0.010620 | -.001450 | -.002900 |
| 120 | 0.012200 | 0.000470 | 0.001060 |
| | | MSE | |
| 30 | 0.003007 | 0.003849 | 0.0 |
| 60 | 0.002066 | 0.002360 | 0.0 |
| 90 | 0.001471 | 0.001416 | 0.001144 |
| 120 | 0.001202 | 0.001094 | 0.000979 |
| | | RED.MSE | |
| 30 | 0.408142 | 0.522492 | 0.0 |
| 60 | 0.560762 | 0.540465 | 0.0 |
| 90 | 0.598887 | 0.576668 | 0.484150 |
| 120 | 0.652907 | 0.596542 | 0.531742 |

| 1661 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.014400 | -.014650 | 0.0 |
| 60 | -.016040 | -.016760 | 0.0 |
| 90 | -.007480 | -.014470 | -.004010 |
| 120 | -.007190 | -.012930 | -.007140 |
| | | MSE | |
| 30 | 0.006277 | 0.009354 | 0.0 |
| 60 | 0.003242 | 0.034513 | 0.0 |
| 90 | 0.002048 | 0.002651 | 0.001649 |
| 120 | 0.001272 | 0.001944 | 0.001074 |
| | | REL.MSE | |
| 30 | 0.776205 | 1.157222 | 0.0 |
| 60 | 0.814183 | 1.116019 | 0.0 |
| 90 | 0.759891 | 0.983292 | 0.611720 |
| 120 | 0.629021 | 0.961625 | 0.531200 |

| 1686 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.021870 | -.046420 | 0.0 |
| 60 | -.020750 | 0.0 | 0.0 |
| 90 | -.019020 | -.027160 | -.011150 |
| 120 | -.016930 | -.023240 | -.003820 |
| | | MSE | |
| 30 | 0.003743 | 0.007504 | 0.0 |
| 60 | 0.002246 | 0.0 | 0.0 |
| 90 | 0.001545 | 0.002367 | 0.000471 |
| 120 | 0.001144 | 0.001727 | 0.000858 |
| | | REL.MSE | |
| 30 | 0.729088 | 1.443249 | 0.0 |
| 60 | 0.878717 | 0.0 | 0.0 |
| 90 | 0.920284 | 1.364932 | 0.560204 |
| 120 | 0.876505 | 1.327730 | 0.659524 |

| 1803 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.000910 | 0.0 | 0.0 |
| 60 | -.002770 | 0.0 | 0.0 |
| 90 | -.002630 | -.005140 | -.002720 |
| 120 | -.002070 | -.006270 | -.004740 |
| | | MSE | |
| 30 | 0.000451 | 0.0 | 0.0 |
| 60 | 0.000244 | 0.0 | 0.0 |
| 90 | 0.000180 | 0.000204 | 0.000142 |
| 120 | 0.000118 | 0.000149 | 0.000124 |
| | | REL.MSE | |
| 30 | 0.436466 | 0.0 | 0.0 |
| 60 | 0.477273 | 0.0 | 0.0 |
| 90 | 0.523047 | 0.533267 | 0.412729 |
| 120 | 0.457317 | 0.575849 | 0.461914 |

| 1899 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.000610 | 0.0 | 0.0 |
| 60 | -.003120 | 0.0 | 0.0 |
| 90 | -.000750 | 0.035240 | 0.012430 |
| 120 | -.002900 | 0.034410 | 0.013240 |
| | | MSE | |
| 30 | 0.002502 | 0.0 | 0.0 |
| 60 | 0.001066 | 0.0 | 0.0 |
| 90 | 0.000415 | 0.002785 | 0.000498 |
| 120 | 0.000491 | 0.002426 | 0.000431 |
| | | REL.MSE | |
| 30 | 0.311432 | 0.0 | 0.0 |
| 60 | 0.270761 | 0.0 | 0.0 |
| 90 | 0.304781 | 1.041327 | 0.373180 |
| 120 | 0.344645 | 1.209433 | 0.414249 |

| 1032 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.016010 | 0.0 | 0.0 |
| 60 | 0.025430 | 0.001150 | 0.0 |
| 90 | 0.022910 | 0.005330 | 0.000930 |
| 120 | 0.019940 | 0.002540 | 0.004760 |
| | | MSE | |
| 30 | 0.001661 | 0.0 | 0.0 |
| 60 | 0.001399 | 0.001630 | 0.0 |
| 90 | 0.000957 | 0.000907 | 0.000862 |
| 120 | 0.000760 | 0.000633 | 0.000668 |
| | | RED.MSE | |
| 30 | 0.216039 | 0.0 | 0.0 |
| 60 | 0.359566 | 0.419009 | 0.0 |
| 90 | 0.384435 | 0.349774 | 0.332384 |
| 120 | 0.390664 | 0.325306 | 0.343243 |

| 1033 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.010440 | 0.0 | 0.0 |
| 60 | -.008620 | -.007710 | 0.0 |
| 90 | -.007740 | -.012970 | -.002810 |
| 120 | -.007590 | -.013260 | -.005710 |
| | | MSE | |
| 30 | 0.001661 | 0.0 | 0.0 |
| 60 | 0.000994 | 0.001363 | 0.0 |
| 90 | 0.000562 | 0.000600 | 0.000587 |
| 120 | 0.000526 | 0.000554 | 0.000460 |
| | | RED.MSE | |
| 30 | 0.578835 | 0.0 | 0.0 |
| 60 | 0.692973 | 0.950065 | 0.0 |
| 90 | 0.587514 | 0.627757 | 0.613620 |
| 120 | 0.733801 | 0.771793 | 0.641213 |

| 1059 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.007100 | 0.0 | 0.0 |
| 60 | 0.000080 | -.000020 | 0.0 |
| 90 | 0.000340 | 0.005500 | 0.008470 |
| 120 | 0.001970 | 0.005660 | 0.007440 |
| | | MSE | |
| 30 | 0.002153 | 0.0 | 0.0 |
| 60 | 0.001375 | 0.001158 | 0.0 |
| 90 | 0.000824 | 0.000736 | 0.000663 |
| 120 | 0.000618 | 0.000529 | 0.000347 |
| | | RED.MSE | |
| 30 | 0.262437 | 0.0 | 0.0 |
| 60 | 0.335275 | 0.282373 | 0.0 |
| 90 | 0.301263 | 0.269240 | 0.244434 |
| 120 | 0.301517 | 0.258177 | 0.193431 |

| 1166 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.008560 | 0.0 | 0.0 |
| 60 | -.010440 | -.007920 | 0.0 |
| 90 | -.009740 | -.012650 | -.006310 |
| 120 | -.008790 | -.012280 | -.006810 |
| | | MSE | |
| 30 | 0.001435 | 0.0 | 0.0 |
| 60 | 0.000856 | 0.001070 | 0.0 |
| 90 | 0.000591 | 0.000468 | 0.000507 |
| 120 | 0.000407 | 0.000368 | 0.000390 |
| | | RED.MSE | |
| 30 | 0.697536 | 0.0 | 0.0 |
| 60 | 0.831660 | 1.039394 | 0.0 |
| 90 | 0.860995 | 0.681686 | 0.739375 |
| 120 | 0.791788 | 0.714966 | 0.757564 |

| 1239 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | 0.009660 | -.026530 | 0.0 |
| 60 | 0.002180 | -.023130 | 0.0 |
| 90 | 0.001590 | -.016880 | -.001730 |
| 120 | 0.001130 | -.015140 | -.002680 |
| | | MSE | |
| 30 | 0.001614 | 0.002858 | 0.0 |
| 60 | 0.000858 | 0.001564 | 0.0 |
| 90 | 0.000626 | 0.001057 | 0.000942 |
| 120 | 0.000468 | 0.000839 | 0.000533 |
| | | RED.MSE | |
| 30 | 0.348243 | 0.616581 | 0.0 |
| 60 | 0.370272 | 0.675125 | 0.0 |
| 90 | 0.404907 | 0.694474 | 0.609777 |
| 120 | 0.403667 | 0.724419 | 0.545926 |

| 1367 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.006120 | 0.018520 | 0.0 |
| 60 | -.009970 | 0.007600 | 0.0 |
| 90 | -.006720 | 0.005100 | -.001660 |
| 120 | -.003720 | 0.003360 | -.001430 |
| | | MSE | |
| 30 | 0.003030 | 0.005137 | 0.0 |
| 60 | 0.001756 | 0.002561 | 0.0 |
| 90 | 0.001312 | 0.001636 | 0.001018 |
| 120 | 0.000839 | 0.001138 | 0.000646 |
| | | RED.MSE | |
| 30 | 0.380562 | 0.645300 | 0.0 |
| 60 | 0.441206 | 0.643461 | 0.0 |
| 90 | 0.494512 | 0.616503 | 0.383593 |
| 120 | 0.421632 | 0.571995 | 0.324716 |

| 1512 | CLASSY | AMOEBA | ISOCLS |
|---|---|---|---|
| | | BIAS | |
| 30 | -.002900 | 0.0 | 0.0 |
| 60 | -.005110 | -.015640 | 0.0 |
| 90 | -.005910 | -.011450 | -.021740 |
| 120 | -.005950 | -.009990 | -.018350 |
| | | MSE | |
| 30 | 0.007147 | 0.0 | 0.0 |
| 60 | 0.002774 | 0.003963 | 0.0 |
| 90 | 0.001831 | 0.002423 | 0.001705 |
| 120 | 0.001300 | 0.001643 | 0.001538 |
| | | RED.MSE | |
| 30 | 0.954468 | 0.0 | 0.0 |
| 60 | 0.742057 | 1.059890 | 0.0 |
| 90 | 0.734764 | 0.972015 | 0.694028 |
| 120 | 0.695696 | 0.878448 | 0.822710 |



84

| | AVERAGES | | | VARIANCES | | |
|---|---|---|---|---|---|---|
| | CLASSY | AMOEBA | ISOCLS | CLASSY | AMOEBA | ISOCLS |
| | | | BIAS | | | |
| 30 | 0.00036834 | -.00941666 | 0.0 | 0.00010890 | 0.00051509 | 0.0 |
| 60 | 0.00006095 | -.00430625 | 0.0 | 0.00012138 | 0.00013838 | 0.0 |
| 90 | -.00037000 | -.00495141 | -.00323619 | 0.00008227 | 0.00020197 | 0.00007368 |
| 120 | -.00040130 | -.00451095 | -.00324428 | 0.00006833 | 0.00017815 | 0.00007746 |
| | | | MSE | | | |
| 30 | 0.00285246 | 0.00522211 | 0.0 | 0.00000367 | 0.00000503 | 0.0 |
| 60 | 0.00148009 | 0.00212906 | 0.0 | 0.00000065 | 0.00000119 | 0.0 |
| 90 | 0.00099690 | 0.00140800 | 0.00099719 | 0.00000030 | 0.00000059 | 0.00000021 |
| 120 | 0.00073538 | 0.00106862 | 0.00075933 | 0.00000015 | 0.00000035 | 0.00000012 |
| | | | RED.MSE | | | |
| 30 | 0.49676664 | 0.76839358 | 0.0 | 0.04504710 | 0.10229522 | 0.0 |
| 60 | 0.51693344 | 0.72288340 | 0.0 | 0.03661084 | 0.06289172 | 0.0 |
| 90 | 0.52017057 | 0.72251660 | 0.51264614 | 0.03732508 | 0.07170510 | 0.01777804 |
| 120 | 0.51962829 | 0.73885107 | 0.52794492 | 0.03581393 | 0.08057529 | 0.02143240 |

APPENDIX B

EVALUATION OF BAYESIAN SEQUENTIAL PROPORTION ESTIMATION
USING ANALYST LABELS

## APPENDIX B

# EVALUATION OF BAYESIAN SEQUENTIAL PROPORTION ESTIMATION USING ANALYST LABELS*

By R. K. Lennington and K. M. Abotteen

## 1. INTRODUCTION

A previous study by R. K. Lennington and J. K. Johnson (ref. 1) concluded by recommending a new procedure for crop proportion estimation. The procedure consisted of two steps. First, the Landsat data were to be clustered using the CLASSY clustering algorithm. Then, picture elements (pixels) were to be allocated to each cluster strata and labeled using a sequential Bayesian allocation scheme developed by M. D. Pore (ref. 2). The labeled pixels were used to form a posterior distribution Bayes estimate of the proportion of the class of interest. In tests involving ground-truth data from 21 blind sites used in Phase III of the Large Area Crop Inventory Experiment (LACIE), this procedure was unbiased and had an estimated mean squared error (MSE) approximately equal to that of a procedure called Procedure 1 (which is based on the sampling of individual pixels) and uses only one-third of the total number of labeled pixels (ref. 1).

In order to explore the feasibility of the new procedure in an actual labeling situation and to perform a preliminary evaluation of its characteristics using analyst labels, a test involving 10 Phase III segments was undertaken. Section 2 describes the procedure used for selecting pixels to be labeled and the method for obtaining proportion estimates. The data set used in the experiment is described in section 3, while the results pertaining to the accuracy of the analyst labels and the bias and MSE of the proportion estimates obtained using these labels are described in section 4. Section 4 also presents the conclusion and recommendations.

---

*Published by Lockheed Engineering and Management Services Company, Inc., LEMSCO-14355, NASA/JSC (Houston), April 1980.

## 2. LABELING PROCEDURE

For the purposes of this test, the Bayesian sequential allocation procedure was implemented on a Texas Instruments TI159 programmable calculator. The version of the allocation procedure implemented was slightly different from the procedure used in the previous study (ref. 1) in that a beta distribution was used for the prior distribution of cluster purities rather than a quadratic or exponential distribution. The form of the distribution used was as follows.

$$g(\theta_i) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(\theta_i)^{a-1} (1-\theta_i)^{b-1} \tag{1}$$

where

$b = 1$

$a = \frac{\hat{p}}{1-\hat{p}}$

$\hat{p}$ = the estimated proportion of the class of interest in the whole segment

$\theta_i$ = the proportion of the class of interest in cluster i

g = the prior distribution of cluster purities

The choice of the parameters a and b ensures that the mean of the distribution will be $\hat{p}$. The parameter b was chosen to be fixed at a value of 1 because that value seemed to give the best fit to the previously obtained empirical prior distributions (ref. 1). Initially, the parameter a was chosen to be 0.515, corresponding to a $\hat{p}$ of 0.34.

The beta prior distribution, although not identical to the prior distributions used in the previous study, is not greatly different and does offer some advantages. It may be used over the entire range of segment proportions; hence, the use of a prior distribution for large proportion segments and another for small proportion segments is unnecessary. Also, the similarity of



88

the beta distribution to the binomial distribution allows the calculation of the Bayes posterior distribution estimator for $\theta_i$ and the expressions for the bias and variance of this estimator with comparative ease. In fact, the beta distribution is called a "natural conjugate prior distribution" to the binomial distribution for this reason. In addition, tests performed subsequent to the work reported in reference 1 showed that use of the beta prior distribution with ground-truth labels produced results which were at least as good as those produced using the combination of a quadratic and exponential prior distribution.

Using the beta prior distribution for $\theta_i$, the Bayes posterior distribution estimator for $\theta_i$ becomes

$$\hat{\theta}_i = \frac{X_i + a}{n_i + a + b} \tag{2}$$

where

$n_i$ = the total number of pixels sampled from cluster i

$X_i$ = the number of sampled pixels which belong to the class of interest

The bias and MSE of this estimator are

$$\text{Bias}_i = E(\hat{\theta}_i - \theta_i) = \frac{a(1-\theta_i) + b\theta_i}{n_i + a + b} \tag{3}$$

$$\text{MSE}_i = \frac{n_i\theta_i(1 - \theta_i) + [a(1 - \theta_i) - b\theta_i]^2}{(n_i + a + b)^2} \tag{4}$$

where E = the expected value operator.

The allocation procedure begins with the allocation of two random pixels to each cluster. At this point, $\hat{p}$ is calculated as

$$\hat{p} = \sum_{i=1}^{c} \left(\frac{N_i}{N_t}\right) \hat{\theta}_i \tag{5}$$

where

$N_i$ = the number of pixels in cluster i

$N_t$ = the total number of pixels in the segment

c = the number of clusters

The parameter a is then reset using the equation

$$a = \frac{\hat{p}}{1 - \hat{p}}$$

At this point, the sequential allocation of pixels begins. Succeeding pixels are allocated to clusters which will minimize the expected value of an estimator of the overall MSE for the segment proportion estimate $\hat{p}$.

The MSE for $\hat{p}$ may be written as

$$MSE_{\hat{p}} = \sum_{i=1}^{c} \left(\frac{N_i}{N_t}\right)^2 MSI_i \tag{6}$$

By using $\hat{\theta}_i$ in place of $\theta_i$ in equation (4), $MSE_i$ may be estimated. We will denote this estimator as $\hat{MSE}_i(x_i, n_i)$.

The expected reduction in the estimated MSE by labeling another pixel from cluster i becomes

$$\Delta\hat{MSE}_i = \left(\frac{N_i 2}{N_t}\right)\left\{\hat{MSE}_i(x_i, n_i) - \left[\theta_i \hat{MSE}_i(x_i + 1, n_i + 1) + (1 - \hat{\theta}_i) MSE_i(x_i, n_i + 1)\right]\right\} \tag{7}$$

Thus, each successive pixel is chosen at random from the cluster having the largest value of $\Delta\hat{MSE}_i$.

In practice, the CLASSY clustering algorithm was first run on a given segment. Then each of the 209 grid intersection pixels was associated with the cluster in which it was placed, and the grid intersection pixels falling in each cluster were listed in a randomized order. The randomized list also



90

contained the label of each pixel that had been previously labeled by an analyst and indicated whether the labeled pixel was a type I or type II dot.

In selecting pixels from clusters, the first to be selected from the randomized list were the type II dots for which analyst labels were available. When these pixels were exhausted, others were chosen according to the randomized order within clusters. If a type I dot fell in this sequence, its label was used. Dots other than type I were labeled by one of the authors (K. Abotteen) using standard analyst procedures. A total of 45 pixels were allocated and labeled for each segment.

## 3. DATA SET AND EXPERIMENTAL DESIGN

The data set for this experiment consisted of 10 phase III blind sites chosen as a subset of the 21 segments used in the previous study (ref. 1). These segments were chosen to be representative of the previously used, larger data set with regard to geographical location and range of segment proportions of small grains. These segments and acquisitions along with their location and the ground-truth proportion of small grains in each segment are given in table 1.

The experimental design consisted of selecting and labeling 45 grid intersection dots from each segment. Repeated processings were not attempted due to the limited number of analyst labels available.

## 4. RESULTS

This study provides the data for answering two important questions relative to the use of analyst labels with the Bayesian sequential allocation procedure. The first question concerns analyst accuracy in labeling pixels. Since in the Bayesian sequential procedure more pixels are allocated to mixed clusters, it was thought that the analyst labeling accuracy might decrease. The second question concerns the bias and MSE of the proportion estimate resulting from the procedure as compared to the bias and MSE of a simple random sample of the

TABLE 1.- DESCRIPTION OF THE DATA SET

| Segment | Location | Acquisitions used | Ground-truth proportion of small grains |
| --- | --- | --- | --- |
| 1005(w) | Cheyenne, Colorado | 7177, 7159, 6326, 6254 | 0.348 |
| 1033(w) | Clark, Kansas | 7156, 6288 | .095 |
| 1060(w) | Sherman, Texas | 7158, 7068 | .231 |
| 1231(w) | Jackson, Oklahoma | 7156, 7066, 6288 | .744 |
| 1520(w) | Big Stone, Minnesota | 7174, 7156, 7120 | .301 |
| 1604(s) | Renville, North Dakota | 7143, 7125 | .524 |
| 1675(s) | McPherson, South Dakota | 7230, 7176, 7123, 6254 | .291 |
| 1803(w) | Shannon, South Dakota | 7178, 7159, 7123, 6255 | .032 |
| 1805(m) | Gregory, South Dakota | 7211, 7158, 6307, 6290 | .164 |
| 1853(w) | Ness, Kansas | 7193, 7067, 6253 | .306 |

Symbol definition:

w = winter wheat
s = spring wheat
m = mixed wheat



92

same size. Analyst accuracy will be examined first, followed by results concerning the proportion estimate itself.

Table 2 shows the error rate in labeling small grains (percentage of ground-truth small grain pixels labeled "other") and the error rate in labeling "other" (percentage of ground-truth "other" pixels labeled small grains) for the 45 pixels that were sequentially allocated to each segment. The corresponding error rates for the type II dots that are selected as a simple random sample are also given. It should be noted that in every case the error rate in labeling small grain pixels was lower for the sequentially allocated pixels than for the type II dots. The error rate in labeling "other" pixels was lower in two cases for the sequentially allocated pixels; however, the error rate in labeling "other" pixels was generally fairly low for both types of allocations.

As another test, one may examine the total number of labeling errors using a sequential Bayesian allocation and compare this to the expected total number of errors based on the error rate for the type II dots. The expected number of errors was calculated by multiplying the total error rate calculated from the type II dots by 45. These data are given in table 3. A chi-square test of these observed and expected number of errors yields a value of

$$X^2 = 14.811$$

With 9 degrees of freedom, the 5 percent significance level of the $X^2$ random variable is 16.9. Hence, at this level of significance, we fail to reject the hypothesis that the observed number of errors are not different than the expected number of errors based on the simple random sample of type II dots. It should be noted that the chi-square test may fail to hold since three of the segments have an expected number of errors less than five. However, the test may be taken as an indication of very little difference in the error rates for the two labeling procedures.

TABLE 2.- ANALYST ERROR RATES FOR SEQUENTIALLY ALLOCATED DOTS VERSUS THE TYPE II DOTS

| Segment | Sequentially allocated dots | | Type II dots | |
|---|---|---|---|---|
| | Error rate for spring grains | Error rate for "other" | Error rate for spring grains | Error rate for "other" |
| 1005 | 0.4286 | 0.0417 | 0.5000 | 0.0270 |
| 1033 | .7000 | .0286 | .8571 | .0189 |
| 1060 | .2778 | .0370 | .2857 | .0000 |
| 1231 | .0294 | .0909 | .0851 | .1818 |
| 1520 | .2353 | .1429 | .2500 | .0909 |
| 1604 | .4800 | .2000 | .4839 | .3158 |
| 1675 | .3571 | .0323 | .8333 | .0208 |
| 1803 | .2500 | .0244 | .5000 | .0000 |
| 1805 | .2000 | .0857 | .3636 | .0460 |
| 1853 | .1429 | .1613 | .2000 | .0889 |
| Averages | 0.3101 | 0.0845 | 0.4359 | 0.0790 |



94

TABLE 3.- OBSERVED AND EXPECTED TOTAL NUMBER OF ANALYST LABELING ERRORS

| Segment | Total number of errors | |
|---|---|---|
| | Observed[a] | Expected[b] |
| 1005 | 10 | 9.135 |
| 1033 | 8 | 5.265 |
| 1060 | 6 | 3.015 |
| 1231 | 2 | 4.635 |
| 1520 | 8 | 5.985 |
| 1604 | 16 | 15.750 |
| 1675 | 6 | 8.235 |
| 1803 | 3 | 0.765 |
| 1805 | 5 | 3.690 |
| 1853 | 7 | 5.265 |

[a]Number of errors observed out of 45 sequentially allocated pixels.

[b]Number of errors expected based on the error rate on the type II dots.

Regarding the actual proportion estimates, table 4 shows the posterior distribution Bayes proportion estimates produced following the sequential allocation of 45 pixels, the proportion estimates based on the type II dots used as a simple random sample, and the Phase III Procedure I estimates. The deviation of each of these estimates from the ground-truth proportion of small grains for each segment also appears in this table.

Several observations may be made from table 4. First, the average bias computed over segments is smaller for the Bayesian sequential estimates than for the simple random sample estimates or the Procedure I estimates. Thus, the Bayesian sequential estimates appear to be somewhat less sensitive to the effects of analyst bias. Also, the MSE computed over segments is smaller for the Bayesian sequential procedure than for the other two procedures. In fact, if we correct the MSE for the type II dot estimates and the Procedure I estimates to reflect an average sample size of 45 pixels rather than the average sample size of 63.5 or 105.5 pixels as given in table 4, we obtain

$$MSE_{\text{Type II adjusted}} = \frac{63.5}{45}(.0118325) = 0.0166970$$

$$MSE_{\text{PI adjusted}} = \frac{105.5}{45}(.0126021) = 0.0295449$$

These values, when compared to the MSE for the Bayesian sequential procedure, yield the following reduction in MSE values.

$$\frac{MSE_{\text{Bayes Seq}}}{MSE_{\text{Type II adjusted}}} = 0.5137 = R_1$$

$$\frac{MSE_{\text{Bayes Seq}}}{MSE_{\text{PI adjusted}}} = 0.2903 = R_2$$

The reduction in the MSE for the type II dots, $R_1$, is very close to the value reported in reference 1 for the reduction in the MSE of the Bayesian sequential procedure as compared to a simple random sample of the same size using ground-truth labels. Both $R_1$ and $R_2$ represent very favorable reductions in MSE values and tend to validate the results of the previous study obtained using the ground truth.



ហu

TABLE 4.- SMALL GRAIN PROPORTION ESTIMATES USING THREE DIFFERENT PROCEDURES

| Segment | $P_{G.T.}$ | Bayesian sequential allocation | | Simple random sample of type II dots | | | Procedure I | | |
|---|---|---|---|---|---|---|---|---|---|
| | | p | $p - P_{G.T.}$ | p | $p - P_{G.T.}$ | Number of Type II dots | p | $p - P_{G.T.}$ | Number of type I and type II dots |
| 1005 | 0.348 | 0.221 | -0.127 | 0.203 | -0.145 | 59.0 | 0.199 | -0.149 | 96.0 |
| 1033 | .095 | .061 | -.034 | .033 | -.062 | 60 | .020 | -.075 | 110 |
| 1060 | .231 | .196 | -.035 | .167 | -.064 | 60 | .170 | -.061 | 106 |
| 1231 | .744 | .755 | +.011 | .776 | -.032 | 58 | .720 | -.024 | 96 |
| 1520 | .301 | .309 | +.008 | .267 | -.034 | 60 | .260 | -.041 | 91 |
| 1604 | .524 | .326 | -.198 | .367 | -.157 | 60 | .350 | -.174 | 101 |
| 1675 | .291 | .128 | -.163 | .050 | -.241 | 60 | .050 | -.241 | 106 |
| 1803 | .032 | .056 | -.024 | .017 | -.015 | 60 | .020 | -.012 | 109 |
| 1805 | .164 | .150 | -.014 | .112 | -.052 | 98 | .124 | -.040 | 149 |
| 1853 | .306 | .329 | +.023 | .267 | -.039 | 60 | .260 | -.046 | 91 |
| Averages | | | -0.051 | | -0.078 | 63.5 | -0.086 | 105.5 | |

$MSE_{Bayes\ Seq.} = .008577$ $MSE_{Type\ II} = .0118325$ $MSE_{Proc\ I} = .0126021$



97

## 5. CONCLUSIONS AND RECOMMENDATIONS

This study indicates that the Bayesian sequential dot allocation and proportion estimation procedure does not significantly increase the analyst labeling error rate. In addition, as compared to a simple random sample, the procedure reduces the MSE by a factor of two. When compared to Procedure I, it reduces the MSE by a factor of approximately three. These results validate the advantages to be obtained in using this procedure with analyst labels.

The fact that the procedure was implemented on a small programmable calculator indicates that it is operationally feasible. However, it should be mentioned that the dot selection part of the program was slower than the normal analyst dot-labeling rate. Another yet-to-be-resolved issue is the development of a technique for selecting pixels from clusters without revealing to the analyst the identity of the cluster in which the pixels fall. It is felt that the knowledge that pixels fall in the same or different clusters may bias the analyst decision. One obvious solution to the computer-time problem and the cluster identity problem would be to implement the procedure on a main-frame computer with interactive analyst access via a terminal. Using this approach, the cluster identities of all the grid intersection pixels could be retained in the computer and therefore would not have to be revealed to the analyst. A larger computer should also be able to select pixels faster than an analyst can label them.

In conclusion, it is recommended that steps be initiated for incorporating this procedure in a large-scale test using fully developed analyst procedures.

## 6. REFERENCES

1. Lennington, R. K.; and Johnson, J. K.: Clustering Algorithm Evaluation and the Development of a Replacement for Procedure 1. Lockheed Electronics Company, Inc. Tech. Memo LEC-13945, November 1979.

2. Pore, M. D.: Bayesian Techniques in Stratified Preparation Estimation. Lockheed Electronics Company, Inc. Tech. Report LEC-13490, August 1979, p. 22.