LOCKHEED ELECTRONICS COMPANY, INC.
AEROSPACE SYSTEMS DIVISION
16811 EL CAMINO REAL • HOUSTON, TEXAS 77058 • TELEPHONE (AREA CODE 713) 488-0080

NASA CR-
160344

TECHNICAL MEMORANDUM

THE CLASSY CLUSTERING ALGORITHM – DESCRIPTION, EVALUATION, AND COMPARISON WITH THE ITERATIVE SELF-ORGANIZING CLUSTERING SYSTEM (ISOCLS)

By

R. K. Lennington
and H. Malek

Approved By: J. C. Minter

T. C. Minter, Jr., Supervisor
Techniques Development Section

Distribution:

JSC/L. F. Childs
J. D. Erickson (4)
G. E. Graybeal

LEC/M. L. Bertrand
B. L. Carroll
J. E. Davis
D. G. Saile
P. C. Swanzy
R. E. Tokerud
Data Research and Control (3)
Technical Library (2)
Job Order File

March 1978 LEC-11289

ABSTRACT

A new clustering method called CLASSY has been developed, which alternates maximum likelihood iteration with a procedure for splitting, combining, and eliminating the resulting statistics. The objectives are to maximize the fit of a mixture of normal distributions to the observed first through fourth central moments of the data and to produce an estimate of the proportions, means, and covariances in this mixture. This document describes the mathematical model which is the basis for CLASSY and the actual operation of the algorithm and compares the results of CLASSY with those produced by ISOCLS, which currently performs these functions. Simulated and actual LACIE data are used in the comparisons.

CONTENTS

Section		Page
1.	INTRODUCTION	1-1
2.	MATHEMATICAL DESCRIPTION	2-1
	2.1 ASSUMPTIONS AND PROBLEM DEFINITION	2-1
	2.2 SOLUTION PROCEDURE	2-2
	2.3 FLOW DIAGRAM	2-7
3.	DATA, PROCEDURES, AND RESULTS	3-1
	3.1 DATA SETS	3-1
	3.2 EVALUATION METHOD AND PROCEDURES	3-1
	3.3 RESULTS	3-3
4.	CONCLUSIONS AND RECOMMENDATIONS	4-1
	4.1 CONCLUSIONS	4-1
	4.2 RECOMMENDATIONS	4-2
5.	REFERENCES	5-1

TABLES

Section		Page
1	DESCRIPTION OF LACIE SAMPLE SEGMENTS.	3-7
2	DISTRIBUTION OF CLASSES IN SIMULATED SEGMENT.	3-7
3	COMPARISON OF THE NUMBER OF CLUSTERS AND THE ESTIMATED PROBABILITY OF CORRECT CLASSIFICATION USING SINGLE-PASS SEGMENT DATA.	3-8
4	COMPARISON OF WHEAT PROPORTION ESTIMATES FOR LABELED CLUSTERS USING SINGLE-PASS SEGMENT DATA	3-9
5	ACQUISITIONS USED IN CREATING FOUR-CHANNEL GREEN IMAGES . . .	3-10
6	COMPARISON OF THE NUMBER OF CLUSTERS AND THE ESTIMATED PROBABILITY OF CORRECT CLASSIFICATION USING THE FOUR-CHANNEL GREEN IMAGE DATA.	3-11
7	COMPARISON OF WHEAT PROPORTION ESTIMATES FOR LABELED CLUSTERS USING FOUR-CHANNEL GREEN IMAGE DATA.	3-12
8	COMPARISON OF THE NUMBER OF CLUSTERS AND THE ESTIMATED PROBABILITY OF CORRECT CLASSIFICATION USING SINGLE-PASS SIMULATED DATA.	3-13
9	COMPARISON OF THE WHEAT PROPORTION ESTIMATES FOR LABELED CLUSTERS USING SINGLE-PASS SIMULATED DATA	3-14
10	PROBABILITY OF MISCLASSIFICATION USING MULTIPASS SIMULATED DATA.	3-14
11	COMPARISON OF CLUSTER STATISTICS FOR PASS 2 SIMULATED DATA. .	3-15
12	COMPARISON OF CLUSTER STATISTICS FOR BAND 1 FOR EACH OF FOUR PASSES OF THE SIMULATED DATA	3-16

FIGURES

Figure		Page
1	Flow diagram for CLASSY algorithm.	2-8
2	Example of the ISOCLS cluster map – segment 1181	3-17
3	Example of the CLASSY cluster map – segment 1181	3-19

1. INTRODUCTION

The Large Area Crop Inventory Experiment (LACIE) is dependent upon clustering for the determination of spectral classes within a scene. Currently, the Iterative Self-Organizing Clustering System (ISOCLS) is used for this purpose (ref. 1). ISOCLS is basically a variation of the k-means or ISODATA algorithm of Ball and Hall (ref. 2). Although this algorithm may be interpreted as a simplified maximum likelihood procedure, it is fundamentally a heuristic algorithm for breaking a data set into fairly homogeneous compact clusters.

The purpose of this study was to compare ISOCLS as a clustering method with a new clustering method called CLASSY.[1] CLASSY operates by alternating maximum likelihood iteration with a procedure for splitting, combining, and eliminating the resultant statistics in order to maximize the fit of a mixture of normal distributions to the observed first through fourth central moments of the data. It is based on a formal mathematical model of the data as a mixture of multivariate normal distributions. CLASSY produces an estimate of the proportions, means, and covariances in this mixture. It differs from standard maximum likelihood procedures in that it also generates an estimate of the number of components of the mixture via the split, combine, and eliminate operations.

Section 2 of this report describes the mathematical model which is the basis for CLASSY and provides a brief description of the actual operation of the algorithm. The results section (3.3) presents data comparing the performances of CLASSY and ISOCLS on simulated data and on actual LACIE data. Finally, these results are evaluated, and conclusions and recommendations are developed (section 4).

---

[1] CLASSY was developed by Dr. M. E. Rassbach while he was a National Research Council postdoctoral fellow working at the Lyndon B. Johnson Space Center.

2. MATHEMATICAL DESCRIPTION

2.1 ASSUMPTIONS AND PROBLEM DEFINITION

The fundamental mathematical assumption underlying CLASSY is that the data may be represented by a mixture of multivariate normal densities. That is, if p denotes probability and $\underline{x}$ is an observation vector,

$$p(\underline{x}|m,\underline{\pi}) = \sum_{i=1}^{m} a_i p_i(\underline{x}|\underline{\mu}_i,\Sigma_i) \qquad (1)$$

where

a_i = the *a priori* probability of occurrence of class i

$p_i(\underline{x}|\underline{\mu}_i,\Sigma_i)$ = the multivariate normal probability density function for class i

m = the total number of classes

$\underline{\pi}$ = the vector of parameters
= $\{a_1,\cdots,a_m, \underline{\mu}_1,\cdots,\underline{\mu}_m, \Sigma_1,\cdots,\Sigma_m\}$

Given a set of unlabeled sample vectors $\{\underline{x}_j\}$, we may form the likelihood function in the following manner.

$$L(\{\underline{x}_j\}|m,\underline{\pi}) = \prod_{j=1}^{N} \left[\sum_{i=1}^{m} a_i p_i(\underline{x}_j|\underline{\mu}_i,\Sigma_i)\right] \qquad (2)$$

where N = the total number of samples.

So far, the assumptions and equations parallel the usual maximum likelihood development. CLASSY makes the additional assumption that each value of the parameters m and $\underline{\pi}$ occurs with an *a priori* probability $A(m,\underline{\pi})$. The objective of CLASSY, then, is to determine the discrete parameter m and the continuous parameter vector $\underline{\pi}$ so as to maximize the following function.

$$L(\{\underline{x}_j\},m,\underline{\pi}) = A(m,\underline{\pi}) \prod_{j=1}^{N} \left[\sum_{i=1}^{m} a_i p_i(\underline{x}_j|\underline{\mu}_i,\Sigma_i) \right] \tag{3}$$

Of course, $A(m,\underline{\pi}_m)$ must be chosen so that it satisfies the normalization constraint

$$\sum_{m=1}^{\infty} \int A(m,\underline{\pi}) d\underline{\pi} = 1 \tag{4}$$

Typically, in the absence of other information, the *a priori* probabilities may be chosen as

$$A(m,\underline{\pi}) = \prod_{i=1}^{m} C_i \tag{5}$$

where C_i = a constant. With this choice for $A(m,\underline{\pi})$, the function to be maximized becomes

$$L(\{\underline{x}_j\},m,\underline{\pi}) = \left(\prod_{i=1}^{m} C_i \right) \prod_{j=1}^{N} \left\{ \sum_{i=1}^{m} \frac{a_i}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{(\underline{x}_j - \underline{\mu}_i)\Sigma_i^{-1}(\underline{x}_j - \underline{\mu}_i)}{2} \right] \right\} \tag{6}$$

where d = dimensionality of the samples.

2.2 SOLUTION PROCEDURE

Many approaches may be taken in maximizing equation (6). The approach chosen in CLASSY is to interleave the standard maximum likelihood iteration [designed to maximize $L(\{\underline{x}_j\},m,\underline{\pi})$ with respect to the continuous parameter vector $\underline{\pi}$] with a discrete split, join, and combine process [designed to maximize $L(\{\underline{x}_j\},m,\underline{\pi})$ with respect to the discrete parameter m]. It is expected that, by alternating these two techniques, values of m and $\underline{\pi}$ corresponding to at least a local maxima of $L(\{\underline{x}_j\},m,\underline{\pi})$ will be determined.

The maximum likelihood iteration is carried out in the standard manner. The data are first scrambled to ensure that a true random sample is obtained. This is especially important in the CLASSY algorithm since any correlation in the data may cause the maximum likelihood procedure to converge to a very poor local minimum or perhaps to fail to converge at all. The initial values assumed are

$$\left.\begin{aligned} m &= 1 \\ a_1 &= 1 \\ \mu_1 &= \begin{bmatrix} 0.04 \\ \vdots \\ 0.04 \end{bmatrix} \\ \Sigma_1 &= \begin{bmatrix} 10 & & 0 \\ & \ddots & \\ 0 & & 10 \end{bmatrix} \end{aligned}\right\} \tag{7}$$

The data are then examined point by point, and the parameter vector $\underline{\pi}$ is iteratively adjusted according to the iterative maximum likelihood equations which may be expressed as follows.

$$p_{(j+1)}[i|\underline{x}_k, \underline{\pi}(j)] = \frac{a_i(j)p_i[\underline{x}_k|\underline{\mu}_i(j), \Sigma_i(j)]}{\sum_{i=1}^{m} a_i(j)p_i[\underline{x}_k|\underline{\mu}_i(j), \Sigma_i(j)]} \tag{8}$$

$$a_i(j+1) = \frac{1}{N}\sum_{k=1}^{N} p_{(j)}(i|\underline{x}_k, \underline{\pi}) \tag{9}$$

$$\underline{\mu}_i(j+1) = \frac{\sum_{k=1}^{N} p_{(j)}[i|\underline{x}_k, \underline{\pi}(j)]\underline{x}_k}{\sum_{k=1}^{N} p_{(j)}[i|\underline{x}_k, \underline{\pi}(j)]} \tag{10}$$

$$\Sigma_i(j+1) = \frac{\sum_{k=1}^{N} P_{(j)}[i|\underline{x}_k, \underline{\pi}(j)][\underline{x}_k - \underline{\mu}(j)][\underline{x}_k - \underline{\mu}(j)]^T}{\sum_{k=1}^{N} P_{(j)}[i|\underline{x}_k, \underline{\pi}(j)]} \qquad (11)$$

where

$P_{(j+1)}[i|\underline{x}_k, \underline{\pi}(j)]$ = the posterior probability of class i on iteration j + 1, given the k*th* sample vector and value of the parameters on the j*th* iteration

$a_i(j)$, $\underline{\mu}_i(j)$, and $\Sigma_i(j)$ = the values of the parameters on the j*th* iteration

In addition to iterating on these parameters, the program also accumulates the third- and fourth-order moments and the logarithm likelihood for each cluster. These statistics are computed on a point-by-point basis simultaneously with the parameter iteration. This means that the parameters are evolving as the moments and the logarithm likelihood are accumulated; and thus, only approximate values are generated.

As each point is considered, the probability that it belongs to each class is computed. These probabilities may be thought of as the fractional part of each data point which is assigned to each cluster. These probabilities are accumulated as the "weights" for each cluster. When the weight for a given cluster exceeds a threshold value, which increases each time it is exceeded, the maximum likelihood iteration is stopped; and the program then checks the fit of the normal distribution to the data for that cluster.

The fit of the hypothesized normal distribution to the data for a cluster is evaluated by examining the third- and fourth-order moments, which represent measures of skewness and kurtosis. The statistics which are generated are given by

$$S_1 = (S\Sigma^{-1}S^T) \qquad (12)$$

where

S = the skewness vector

S_1 = a scalar measure of skewness

S^T = transpose of S

Σ^{-1} = the inverse covariance matrix

$$K_1 = Tr(K\Sigma^{-1}) \tag{13}$$

$$K_2 = Tr(K\Sigma^{-1}K\Sigma^{-1}) - \frac{1}{d}[Tr(\Sigma^{-1}K)]^2 \tag{14}$$

where

K = matrix of kurtosis values

$K_1, K_2, \cdots$ = scalar measures of kurtosis

In CLASSY, these three statistics are tested against their approximate sampling distributions computed under the hypothesis that the samples were drawn from the normal distribution specified by the current values of the parameters. If any one of these three statistics exceeds the threshold value, the cluster is split into two parts. The parameters for each of the two new clusters are determined in order to minimize the difference between the observed covariance matrix, the skewness vector, and the kurtosis matrix and the corresponding quantities for the mixture distribution composed of the two new normal distributions.

Following a split, the parent cluster is not discarded immediately. When the maximum likelihood iteration cycle is begun again, it is carried out for the previously existing clusters, including the parent cluster and the new subclusters (with the new parameters and a weight of 40 points each). Thus, a hierarchical structure or cluster tree evolves as this process is repeated.

At the same time in the processing that a cluster is checked to see if it needs to be split, certain other tests are performed. If a cluster has subclusters (i.e., has been previously split), it is not split again; but the

likelihood ratio of the daughter clusters to the parent cluster is examined. If this ratio is larger than a given threshold, then the parent cluster is eliminated and the daughter clusters take its place. On the other hand, if this ratio is too small, the daughter clusters are eliminated in favor of the parent. In addition, a cluster may be eliminated if its prior probability becomes too small. The program also checks the degree of overlap between clusters at the same level in the cluster tree. If the degree of overlap is too great and the two clusters are not the only subclusters of a given parent cluster, the parameters and other statistics for the two clusters are joined. All of these tests allow for periodic restructuring of the cluster tree at certain intervals; namely, when the weight (or number of points assigned to a given cluster on a fractional probabilistic basis) has accumulated to a certain point in the maximum likelihood iteration portion of the program.

After tests have been made to determine if a cluster needs to be split or if the cluster tree needs to be restructured, the skewness vector and the kurtosis matrix for that cluster are reset to zero. The program then continues the process of maximum likelihood iteration. If a complete pass through the data set is made before a cluster is tested for possible adjustment, then the values of the means at that time are used in equation (11) until another pass through the data set has been completed.

The program recycles through the data a fixed number of times. The number of passes through the data is controlled by an external parameter. When the desired number of passes is complete, the program goes through the data point by point and assigns each data point to the cluster in the cluster tree for which the probability of occurrence of this data point is the greatest. This is the only time in the program that points are assigned to clusters. When all of the points have been assigned, a cluster map showing the cluster symbol for each point is printed out. The program also prints out the final values for the parameters for each cluster in the cluster tree.

2.3 FLOW DIAGRAM

This section gives a general flow diagram for the CLASSY program (fig. 1). This is not a detailed flow diagram for the program but merely serves to summarize the information given in section 2.2 in a convenient manner.

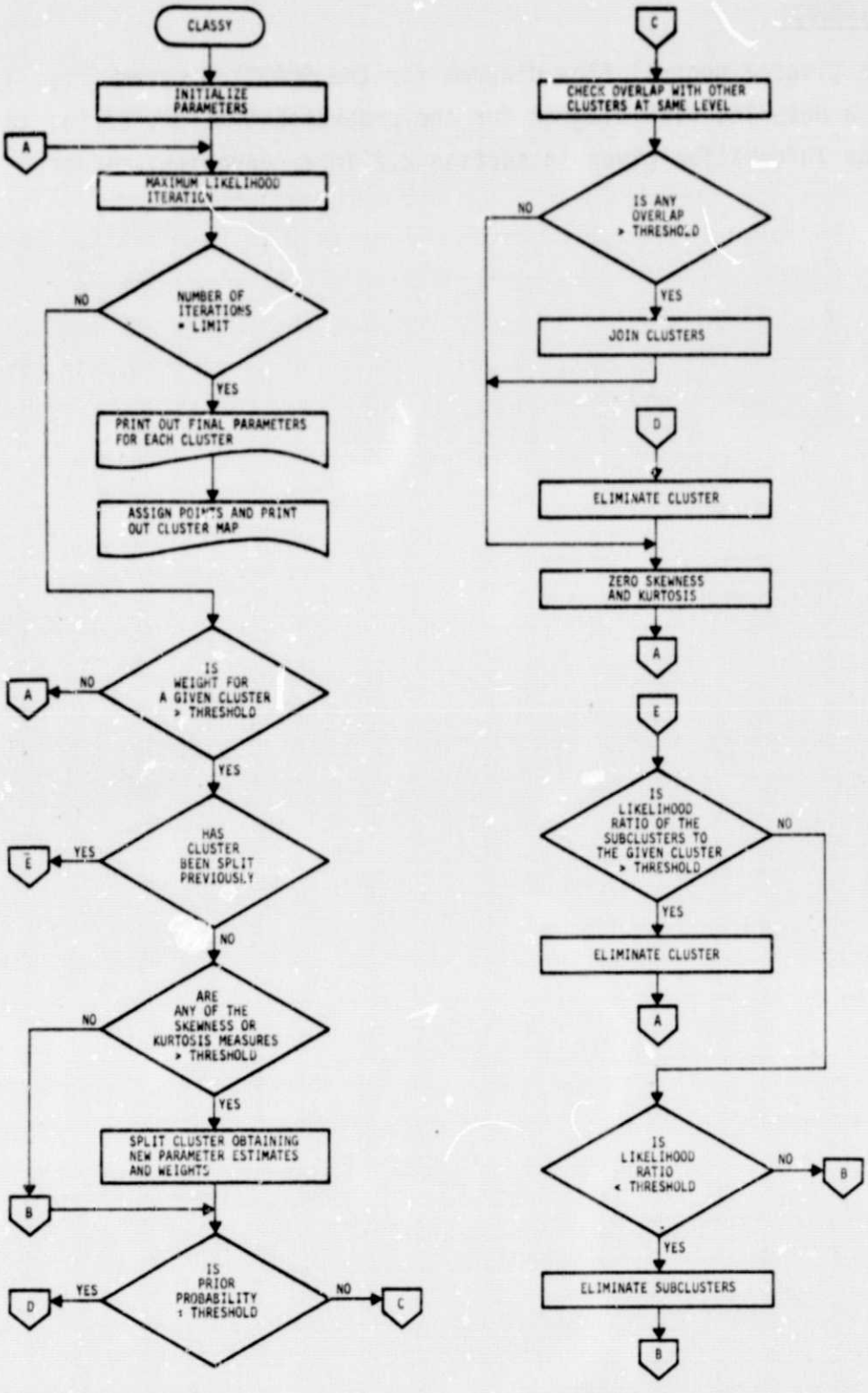

Figure 1.— Flow diagram for CLASSY algorithm.

3. DATA, PROCEDURES, AND RESULTS

3.1 DATA SETS

Two different data sets were used in this study. The first was a set of acquisitions of four different LACIE segments. The second was a set of four different simulated acquisitions of a simulated LACIE segment. Each of these data sets is described separately in the following paragraphs.

The four LACIE segments were selected on the basis of the availability of ground-truth grid-intersection dots and to provide a representative sampling of LACIE segments in terms of field structure and the proportion of wheat present. Once the segments had been chosen, the acquisition which had the largest Bhattacharyya distance of any of the available acquisitions was selected. The segment number, location, acquisition used, and the ground-truth percentages of wheat and small grains for each segment are given in table 1.

The simulated data set consisted of four simulated acquisitions. Each acquisition was derived first by specifying the mean vector and covariance matrix for each of 10 different classes. The class statistics for each class were specified so as to simulate the LACIE data for two wheat classes (W_1 and W_2), two barley classes (B_1 and B_2), two classes of grass (G_1 and G_2), two stubble classes (S_1 and S_2), and two classes of fallow (F_1 and F_2). Once the statistics were specified, samples were generated from a normal distribution having the statistics of a given class. These samples were then placed in rectangular fields arranged over the simulated segment. This process was repeated for each class and for each of the four acquisitions. The arrangement of the simulated fields over the segment was the same for each acquisition. The pattern of the simulated fields is given in table 2.

3.2 EVALUATION METHOD AND PROCEDURES

CLASSY was evaluated using a comparative analysis method in which the clustering results of CLASSY were compared with those of ISOCLS using the ground truth as a reference. The evaluation procedure followed in three steps.

a. The CLASSY and ISOCLS algorithms were applied to each segment in each data set. The clustering results were then obtained in line-printer cluster-map form.

b. The clusters in each map were labeled first by tabulating the cluster symbol and the corresponding ground-truth label (as either wheat or nonwheat) for each grid intersection where ground truth was available. These results were tabulated, and the number of ground-truth wheat pixels and ground-truth nonwheat pixels falling in each cluster was computed.

c. The clusters were then labeled wheat or nonwheat by majority rule.

A measure of the accuracy of each clustering algorithm in separating wheat from nonwheat (or a measure of the overall purity of the wheat and nonwheat clusters) was computed by estimating the probability of correct classification (PCC) for the labeled clusters. This estimate was computed in the following manner.

$$PCC = \sum_{i=1}^{m_1} P(O_i|O)P(O) + \sum_{i=1}^{m_2} P_i(W_i|W)P(W) \qquad (15)$$

where

m_1 = number of clusters labeled "other"

m_2 = number of clusters labeled wheat

$P(O_i|O)$ = probability that a pixel falls in the i*th* other cluster, given that it is other than wheat

$P_i(W_i|W)$ = probability that a pixel falls in the i*th* wheat cluster, given that it is wheat

$P(W)$ = the *a priori* probability that a pixel is wheat

$P(O)$ = the *a priori* probability that a pixel is other than wheat

If empirical proportions are used to estimate these probabilities and *a priorıs*, the resulting expression is as follows.

$$\hat{PCC} = \frac{1}{N_T}\left(\sum_{i=1}^{m_1} N_{O_i|O} + \sum_{i=1}^{m_2} N_{W_i|W}\right) \quad (16)$$

where

N_T = total number of ground-truth pixels

$N_{O_i|O}$ = number of ground-truth other pixels falling in the i*th* other cluster

$N_{W_i|W}$ = number of ground-truth wheat pixels falling in the i*th* wheat cluster

It is noteworthy that, to obtain an accurate estimate of PCC using equation (16), it is necessary that several ground-truth pixels fall in each cluster. Specifically, if there are clusters which have only one or two ground-truth grid pixels, the estimate of PCC will be biased on the high side.

As a part of the analysis, the proportion of wheat also was estimated for the labeled clusters and compared to the ground-truth value. The equation used for this estimate is

$$\hat{P}(W) = \frac{1}{N_T}\sum_{i=1}^{m_2} N_{W_i} \quad (17)$$

where N_{W_i} = the total number of ground-truth pixels (wheat and other) falling in the i*th* wheat cluster.

3.3 RESULTS

The results of these computations and the acquisitions used are given in tables 3 through 12. Tables 3, 4, 6, and 7 compare CLASSY and ISOCLS results for the LACIE segments examined; the corresponding results for simulated segment data are given in tables 8 through 12.

Table 3 compares the number of clusters and the PCC estimates for ISOCLS ($\hat{PCC}_I$) and for CLASSY ($\hat{PCC}_C$) as a result of clustering each of the four LACIE segments examined using both methods. The PCC estimates for CLASSY are, on

the average, about 4 percentage points lower than those for ISOCLS. However, since ISOCLS generates a factor of 4 to 6 more clusters than CLASSY, many of the ISOCLS clusters contain only one or two ground-truth grid-intersection points. As discussed in section 3.2, this means that the PCC estimates for ISOCLS will be biased high relative to CLASSY. In the light of this built-in bias, CLASSY compares very favorably to ISOCLS.

It should be noted that the reduced number of clusters generated by CLASSY results in a dramatic increase in the ease with which the cluster maps may be interpreted visually. Examples of a portion of the cluster map generated by each algorithm are given in figures 2 and 3.

The LACIE segments used in this study contained varying amounts of wheat. The ground-truth percentages of wheat [P(W)] and small grains [P(SG)] are given in table 4. The estimate of the proportion of wheat computed using the ground-truth grid-intersection dots [$\hat{P}_D(W)$] is also included. An estimate of the proportion of wheat from the ground-truth labeled clusters can be obtained using equation (17). The wheat proportion estimates resulting from applying this equation to the CLASSY results (D_C) and ISOCLS results (D_I) are also given in table 4. Comparing these percentages to the ground-truth wheat proportions shows that with the exception of segment 1965 the wheat proportion estimates are about 4 to 6 percent higher than the ground-truth wheat proportion values. These slightly high estimates may be due to the fact that, even though only wheat ground-truth dots were used to label clusters, labeled wheat clusters may reasonably be assumed to include some small grains. The last column in table 4 shows that the ISOCLS estimate was closer to the ground-truth wheat proportion for two segments and the CLASSY estimate was closer for the other two segments.

The imagery for segment 1965 was examined in detail because the wheat proportion estimates for both CLASSY and ISOCLS deviated considerably from the ground truth and the PCC estimates for both algorithms were correspondingly low for this segment. This segment contained numerous small strip fields. Typically, small-fields regions accentuate misregistration problems, which

appear to be the case for this segment. The misregistration of the ground-truth reference acquisition relative to the acquisition clustered reduced PCC values and distorted the proportion of wheat estimates for both algorithms.

In order to obtain an idea about the relative performance of CLASSY and ISOCLS when applied to multitemporal data, four-channel green images were formed for each segment by applying the Kauth transformation to each of four acquisitions for a given segment and then selecting the green channel from each acquisition. It was necessary to reduce the 16-dimensional data to 4 dimensions since CLASSY is limited to 4 dimensions at the present time. Table 5 lists the four acquisitions used for each segment. The results of comparing the PCC values and the wheat proportion estimates for the two algorithms are given in tables 6 and 7, respectively. Comparing table 6 and table 3 shows that the PCC values for both algorithms remained about the same for segments 1181 and 1961 and that they increased significantly for segments 1958 and 1965. The average difference between the CLASSY and ISOCLS PCC values remained about 4 percent. However, the CLASSY PCC equaled the ISOCLS PCC for segment 1988, and the difference was very small for segment 1961. The last column of table 7 shows that, when the four-channel green images were used, the wheat proportion estimates from the CLASSY clusters were closer to the ground-truth values than were the ISOCLS estimates in every case.

Tables 8 and 9 are analogous to tables 3 and 4, except that they give the results for the single-pass simulated data. The column labeled maximum likelihood PCC (PCC_M) gives the overall PCC when using standard maximum likelihood parameter estimates and classification with the number of classes known. Note that the PCC estimates for CLASSY were higher than those for ISOCLS in two of the four passes. In fact, on pass 2, where the separability was greatest, the PCC for CLASSY equaled the maximum likelihood PCC. On the average, the PCC for CLASSY was 1.4 percent higher than that for ISOCLS.

The proportion estimate computed from the labeled clusters is given in table 9. Again, the estimate from CLASSY was closer to the true value in two of the four passes. However, the average individual ISOCLS estimate was about 2 percent closer to the true value.

The results for the simulated data using band 1 from each of the four passes are given in table 10. Band 1 was selected arbitrarily to assess the use of multitemporal data. Note that the PCC estimate for CLASSY was 1.0, meaning that none of the CLASSY clusters contained a mixture of wheat and nonwheat grid intersection dots.

Using the simulated data makes it possible to identify a cluster with a certain class in the data by determining which class contributes the majority of pixels to the cluster. After such an identification, the generating statistics for the subclass may be compared with the cluster statistics produced by CLASSY. Table 11 presents the results of such a comparison for the pass 2 simulated data, whereas table 12 gives similar results for the clustering using band 1 from each of the four passes.

In the pass 2 CLASSY results, four of the five clusters could be clearly identified with one of the generating classes or distributions. A comparison of the mean vector and covariance matrices shows a remarkable correspondence between the CLASSY statistics and the generating statistics. Cluster 3 was about equally divided between grass 1 and grass 2. The statistics for grass 1 are given. Similarly, cluster 2 is a mixture of stubble, fallow, and barley 2. The statistics for each of these classes are very similar for this pass. The statistics for stubble 1 are given as a representative example.

The data from band 1 of each of the four simulated passes had more separability; thus, CLASSY was able to distinguish more classes. The comparison of the generating statistics and the CLASSY statistics is presented in table 12.

Only the variance terms from the multipass covariance matrix were available. Again there is remarkable correspondence between the CLASSY statistics and the generating statistics.

TABLE 1.– DESCRIPTION OF LACIE SAMPLE SEGMENTS

Segment	Location	Acquisition	Ground truth, % wheat	Ground truth, % small grains
1181	Kans.	76070	23.4	29.0
1988	Kans.	75312	33.0	33.0
1961	Kans.	76200	8.2	8.2
1965	N. Dak.	76221	41.6	47.0

TABLE 2.– DISTRIBUTION OF CLASSES IN SIMULATED SEGMENT

W_1	G_2	B_1	S_1	W_2	S_2	W_2	W_1	G_1	B_2
F_2	W_2	G_1	W_1	S_1	S_2	G_2	B_2	W_1	B_1
W_1	G_1	S_2	G_2	S_1	W_2	B_2	W_2	B_1	F_1
G_2	S_1	W_2	B_1	S_2	W_1	W_2	G_1	F_1	B_2
W_2	W_1	G_1	B_1	W_1	S_1	G_2	S_2	B_2	W_2

TABLE 3.— COMPARISON OF THE NUMBER OF CLUSTERS AND THE ESTIMATED PROBABILITY OF CORRECT CLASSIFICATION USING SINGLE-PASS SEGMENT DATA

Segment	ISOCLS		CLASSY		$\hat{PCC}_c - \hat{PCC}_I$
	Number of clusters	$\hat{PCC}_I$	Number of clusters	$\hat{PCC}_c$	
1181	40	0.8410	7	0.8052	-0.0358
1988	40	.8070	8	.7661	-.0409
1961	40	.9236	11	.9028	-.0208
1965	40	.7419	9	.6774	-.0645
Average	40	.8284	8.75	.7875	-.0405

TABLE 4.— COMPARISON OF WHEAT PROPORTION ESTIMATES FOR LABELED CLUSTERS USING SINGLE-PASS SEGMENT DATA

Segment	Ground truth		Ground-truth dots $\hat{P}_D(W)$	ISOCLS $\hat{P}_I(W)$	CLASSY $\hat{P}_C(W)$	$D_I = \hat{P}_I(W) - \hat{P}(W)$	$D_C = \hat{P}_C(W) - \hat{P}(W)$	$\lvert D_I \rvert - \lvert D_C \rvert$
	P(W)	P(SG)						
1181	0.234	0.290	0.333	0.287	0.303	0.053	0.069	-0.016
1988	.330	.330	.322	.397	.287	.067	-.043	.024
1961	.082	.082	.097	.042	.069	-.040	-.013	.027
1965	.416	.470	.516	.526	.645	.110	.229	-.119
Average	.266	.293	.317	.313	.326	.047	.061	-.021

TABLE 5.— ACQUISITIONS USED IN CREATING FOUR-CHANNEL GREEN IMAGES

Segment	Acquisitions
1181	76070
	76107
	76124
	76196
1988	75293
	76127
	76164
	76272
1961	75227
	76164
	76236
	76254
1965	76132
	76203
	76221
	76258

TABLE 6.– COMPARISON OF THE NUMBER OF CLUSTERS AND THE ESTIMATED PROBABILITY OF CORRECT CLASSIFICATION USING THE FOUR-CHANNEL GREEN IMAGE DATA

Segment	ISOCLS		CLASSY		$\hat{PCC}_c - \hat{PCC}_I$
	Number of clusters	$\hat{PCC}_I$	Number of clusters	$\hat{PCC}_c$	
1181	40	0.8667	4	0.8000	-0.0667
1988	40	.9357	16	.9357	0
1961	40	.9167	23	.9097	-.0070
1965	40	.8065	13	.7290	-.0775
Average	40	.8814	14	.8436	-.0378

TABLE 7.– COMPARISON OF WHEAT PROPORTION ESTIMATES FOR LABELED CLUSTERS USING FOUR-CHANNEL GREEN IMAGE DATA

Segment	Ground truth		ISOCLS $\hat{P}_I(W)$	CLASSY $\hat{P}_C(W)$	$D_I = \hat{P}_I(W) - \hat{P}(W)$	$D_C = \hat{P}_C(W) - \hat{P}(W)$	$\lvert D_I \rvert - \lvert D_C \rvert$
	P(W)	P(SG)					
1181	23.4	23.0	29.2	24.1	5.8	0.7	5.1
1988	33.0	33.0	31.6	34.2	-1.4	1.2	0.2
1961	8.2	8.2	6.6	6.9	-1.6	-1.3	0.3
1965	41.6	47.0	62.5	56.5	20.9	14.9	6.0
Average	26.6	29.3	32.5	30.4	5.9	3.9	2.9

TABLE 8.— COMPARISON OF THE NUMBER OF CLUSTERS AND THE ESTIMATED PROBABILITY OF CORRECT CLASSIFICATION USING SINGLE-PASS SIMULATED DATA

Pass	PCC_M	ISOCLS		CLASSY		$PCC_M - \hat{PCC}_I$	$PCC_M - \hat{PCC}_C$	$\hat{PCC}_C - \hat{PCC}_I$
		Number of clusters	$\hat{PCC}_I$	Number of clusters	$\hat{PCC}_C$			
1	0.935	40	0.9139	5	0.9043	0.021	0.030	-0.0096
2	.986	40	.9713	5	.9857	.015	.000	.0144
3	.970	40	.9761	8	.9522	-.006	.018	-.0239
4	.928	40	.8852	7	.9187	.043	.009	.0335
Average	.955	40	.9366	6.25	.9402	.018	.014	.0144

TABLE 9.— COMPARISON OF THE WHEAT PROPORTION ESTIMATES FOR LABELED CLUSTERS USING SINGLE-PASS SIMULATED DATA

Pass	P(W)	$\hat{P}_I(W)$	$\hat{P}_C(W)$	$D_I = \hat{P}_I(W) - \hat{P}(W)$	$D_C = \hat{P}_C(W) - \hat{P}(W)$	$\lvert D_I \rvert - \lvert D_C \rvert$
1	0.3398	0.3301	0.2536	-0.0097	-0.0862	-0.0765
2	.3398	.3254	.3541	-.0144	.0143	.0001
3	.3398	.3636	.2917	.0238	-.0481	-.0243
4	.3398	.3254	.3349	-.0144	-.0049	.0095
Average	.3398	.3361	.3086	-.0147	-.0312	-.0228

TABLE 10.— PROBABILITY OF MISCLASSIFICATION USING MULTIPASS SIMULATED DATA

Data	ISOCLS		CLASSY		$\hat{PCC}_C - \hat{PCC}_I$
	Number of clusters	$\hat{PCC}_I$	Number of clusters	$\hat{PCC}_C$	
Band 1 from each of 4 passes	40	0.9809	7	1.0000	0.0191

TABLE 11.- COMPARISON OF CLUSTER STATISTICS FOR PASS 2 SIMULATED DATA

Cluster number	Identification	Generating statistics		CLASSY statistics	
		Mean vector	Covariance matrix	Mean vector	Covariance matrix
4	Wheat 1	[20.36 20.19 27.29 28.14]	[0.91 1.21 0.34 -0.01 1.21 3.24 .24 -.65 .34 .24 1.77 1.75 -.01 -.65 1.75 3.15]	[20.56 20.67 27.45 28.26]	[1.21 1.04 0.13 -0.19 1.04 2.87 -.10 -.95 .13 -.10 1.34 1.76 -.19 -.95 1.76 3.50]
5	Wheat 2	[18.55 17.02 26.35 28.00]	[0.82 0.69 -0.01 -0.47 .69 1.11 -.48 -1.19 -.01 -.48 1.23 1.41 -.47 -1.19 1.41 3.25]	[18.76 17.13 26.36 27.97]	[1.08 0.80 -0.03 -0.50 .80 1.54 -.47 -1.20 -.03 -.47 1.46 1.50 -.50 -1.20 1.50 3.51]
1	Barley 1	[23.30 25.80 25.98 24.19]	[1.55 1.74 1.22 0.96 1.74 3.16 1.52 1.12 1.22 1.52 1.65 .91 .96 1.12 .91 1.19]	[22.97 25.45 25.27 23.50]	[2.00 1.97 1.83 1.41 1.97 3.59 2.36 1.77 1.83 2.36 2.92 1.84 1.41 1.77 1.84 2.22]
3	Grass 1 (grass 2)	[20.83 20.86 23.37 22.50]	[1.31 2.07 0.54 0.11 2.07 4.70 .91 -.29 .54 .91 1.10 .70 .11 -.29 .70 1.23]	[20.71 20.54 23.18 22.52]	[1.15 1.48 0.55 0.22 1.48 4.10 1.01 .28 .55 1.01 1.40 .64 .22 .28 .64 1.24]
2	Stubble 1 (stubble 2, fallow, barley 2)	[21.90 23.64 24.22 23.12]	[0.97 0.62 0.77 0.69 .64 1.12 .70 .66 .77 .70 1.51 1.40 .69 .66 1.40 2.31]	[22.40 24.43 24.18 22.77]	[0.96 0.44 0.31 0.22 .44 1.17 .38 .29 .31 .38 1.41 .92 .22 .29 .92 1.68]

TABLE 12.– COMPARISON OF CLUSTER STATISTICS FOR BAND 1 FOR EACH OF FOUR PASSES OF THE SIMULATED DATA

Cluster number	Identification	Generating statistics		CLASSY statistics	
		Mean vector	Covariance matrix	Mean vector	Covariance matrix
5	Wheat 1	26.93 20.36 17.39 17.27	1.06 0.91 2.15 3.30	26.84 20.27 17.22 17.02	1.27 0.69 1.42 1.61 .69 1.21 1.25 1.62 1.42 1.25 2.32 2.65 1.61 1.62 2.65 3.49
2	Wheat 2	25.79 18.55 16.85 18.12	1.03 0.82 0.47 1.76	25.90 18.76 16.88 17.97	1.22 0.94 0.78 0.98 .94 1.23 .78 .87 .78 .78 .85 .67 .98 .87 .67 1.80
4	Barley 1	28.41 23.30 22.01 17.01	2.16 4.86 4.15 4.47	28.40 22.71 22.56 17.44	2.30 1.56 3.03 2.18 1.56 1.81 2.69 2.17 3.03 2.69 5.33 3.80 2.18 2.17 3.86 3.58
3	Barley 2	28.23 22.78 22.37 17.34	1.33 0.77 1.88 1.61	28.40 22.71 22.56 17.44	1.63 -0.08 1.79 1.05 -.08 .79 -.40 -.09 1.79 -.40 2.54 1.23 1.05 -.09 1.23 1.86
1	Grass 1 (grass 2, stubble 1)	25.67 20.83 20.10 20.60	1.81 1.31 1.80 1.62	25.82 21.20 20.35 20.72	2.69 0.87 1.76 2.17 .87 1.39 .74 .98 1.76 .74 1.71 1.65 2.17 .98 1.65 2.43
6	Fallow 1	24.59 22.48 23.22 21.56	0.67 0.52 0.90 .66	24.68 22.45 23.21 21.67	0.75 0.38 0.42 0.48 .38 .72 .68 .09 .42 .68 1.06 .04 .48 .09 .04 .75
7	Stubble 2 (fallow 2)	24.33 22.21 22.69 28.63	1.17 0.67 0.74 1.04	24.34 22.25 22.70 28.63	1.31 0.38 -0.01 -0.14 .38 .86 .09 -.15 -.01 .09 1.01 .84 -.14 -.15 .84 1.35

Figure 2.— Example of the ISOCLS cluster map — segment 1181.

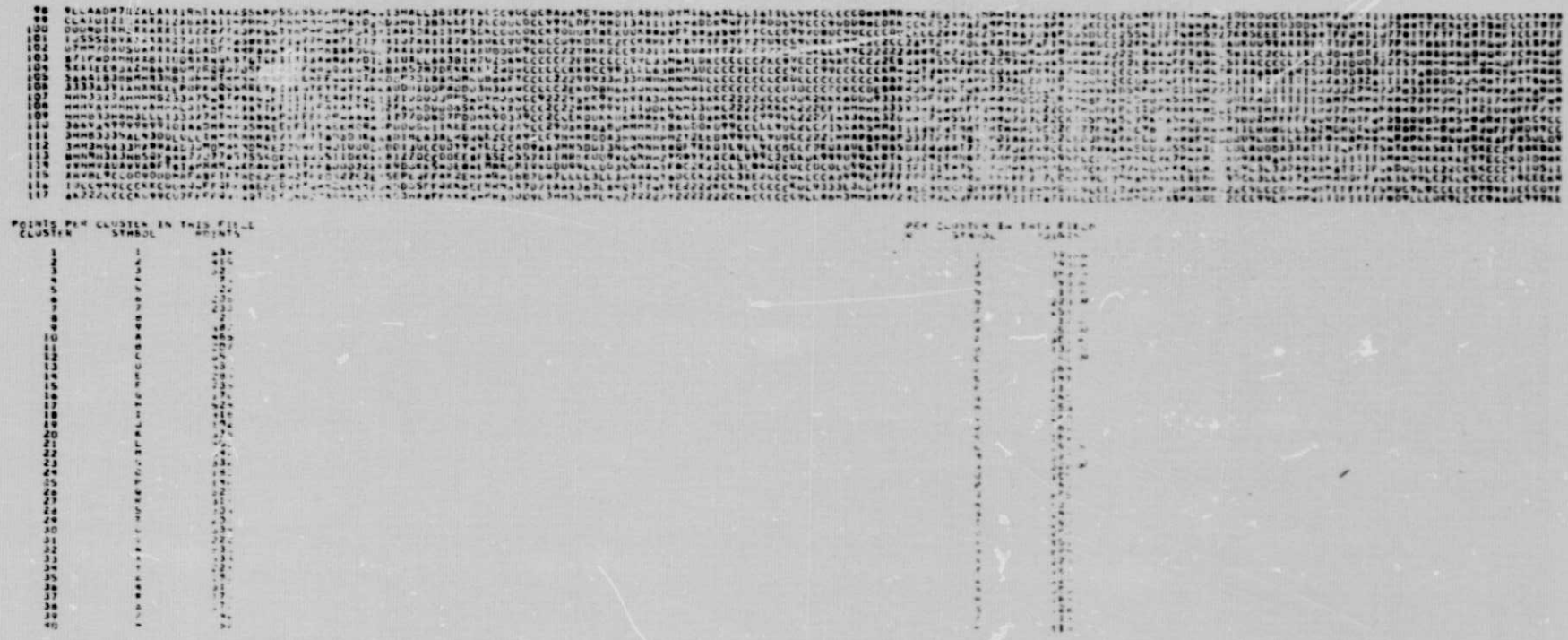

Figure 2.— Concluded.

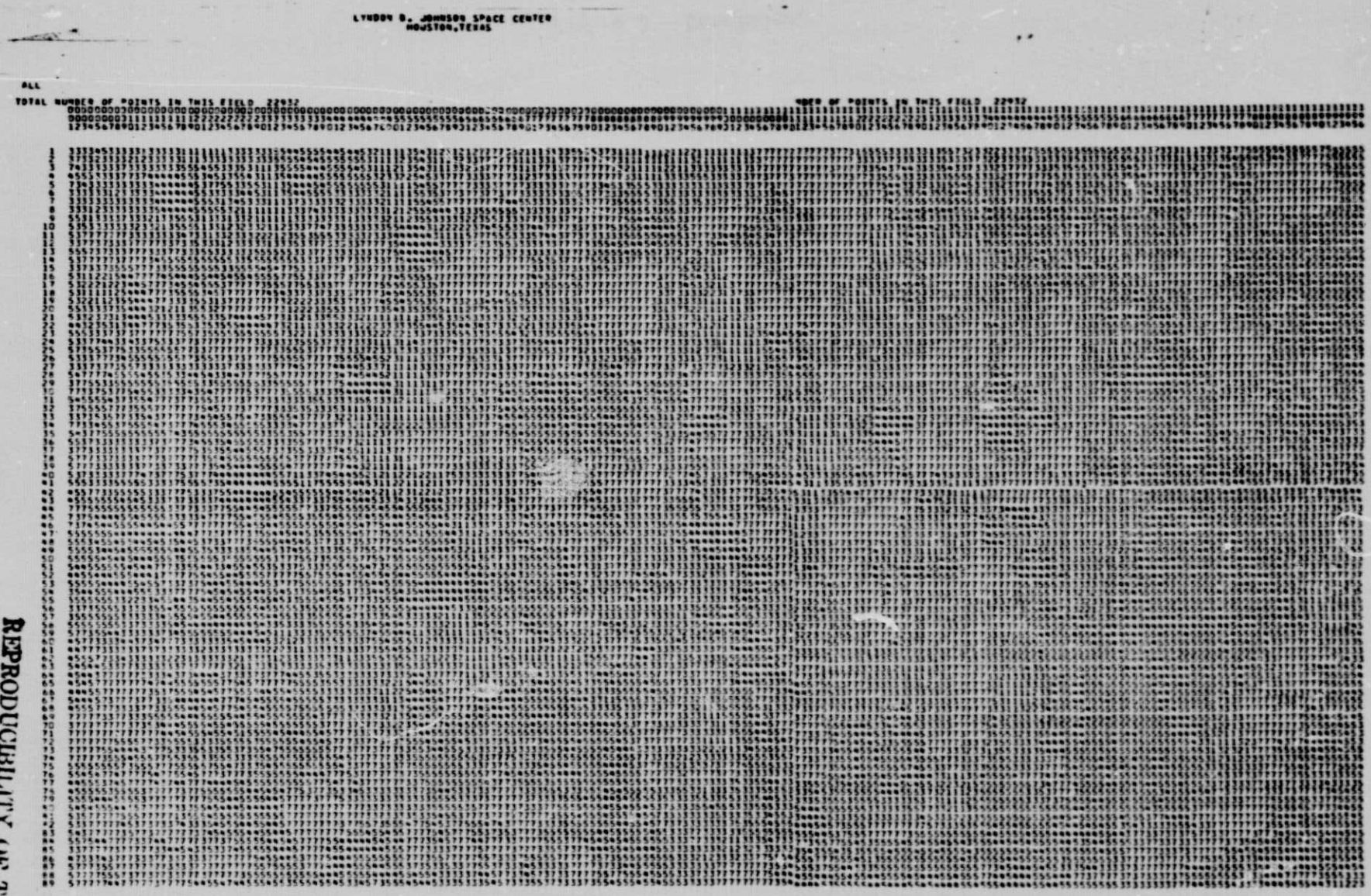

Figure 3.– Example of the CLASSY cluster map – segment 1181.

Figure 3.— Concluded.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 CONCLUSIONS

The main conclusion of this study is that the performance of the CLASSY clustering algorithm compares favorably with ISOCLS on both the real and simulated LACIE segment data. In terms of performance, these results were obtained despite the fact that CLASSY reduces the number of clusters by a factor of 4 to 6 as compared to ISOCLS. This would indicate that CLASSY is indeed approximating the empirical mixture density rather than just breaking up the data space into small homogeneous areas as does ISOCLS. This conclusion is further substantiated by noting the high degree of correspondence between the CLASSY cluster statistics and the generating statistics of classes in the simulated data. It appears that the CLASSY algorithm may well provide a solution to the fundamental problem of maximum likelihood clustering — the determination of the inherent number of classes in the data.

A detailed examination of the results indicates that, in general, the PCC estimates for ISOCLS were slightly higher than those for CLASSY. (However, CLASSY did actually have higher PCC estimates on two of the simulated data passes.) It should be remembered in viewing these results that, because ISOCLS had many more clusters than CLASSY, there were always ISOCLS clusters which contained only one or two ground-truth dots. As discussed in section 3.2, this tends to bias the PCC estimate for ISOCLS on the high side.

The wheat proportion estimates for both CLASSY and ISOCLS were comparable. Again, ISOCLS is usually a little closer to the ground-truth value. However, the proportion estimates are also biased when the clusters are mixed. So, again, it is to be expected that ISOCLS, with its larger number of clusters, would generate better estimates. The fact that the estimates are only slightly better and sometimes worse indicates again that CLASSY is determining the distributional structure of the data.

Finally, it should be noted that ISOCLS typically requires 3 to 5 minutes to process a real LACIE segment; whereas CLASSY, iterating through the data three times, typically requires 9 to 16 minutes of central processing unit time.

4.2 RECOMMENDATIONS

On the basis of these tests, it is recommended

a. That further tests be conducted using CLASSY, particularly on multiple-pass LACIE data

b. That the CLASSY program be completely documented, including the revision of certain parts of the program to improve the performance or speed of the algorithm

c. That methods for incorporating the CLASSY algorithm into LACIE Procedure 1 be developed and tested

5. REFERENCES

1. Kan, E. P.: The JSC Clustering Program ISOCLS and Its Applications. LEC-0483, NASA/JSC (Houston), July 1973.

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