80-10058
NASA CR-
160436
JSC-16012

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

Ref: 642-7677
Contract NAS 9-15800
Job Order 73-705-18

TECHNICAL MEMORANDUM

EVALUATION OF THREE-CATEGORY CLASSIFICATION

By

K. A. Havens and K. M. Abotteen

(E80-10058) LARGE AREA CROP INVENTORY EXPERIMENT (LACIE). EVALUATION OF THREE-CATEGORY CLASSIFICATION (Lockheed Electronics Co.) 45 p HC A03/MF A01 CSCL 02C G3/43
N80-18509
Unclas 00058

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

August 1979

LEC-13498

PREFACE

Documented in this report are the results of a study to evaluate the Large Area Crop Inventory Experiment clustering and classification procedures in terms of variance of the proportion estimates and the probabilities of correct classification for three categories. The categories of interest were corn, soybeans, and other.

Timely preparation of the data and experiment design for this study would not have been possible without the aid of several coworkers. K. Lennington and D. Register wrote the initial experiment design. R. Abotteen and J. Johnson helped to verify the ground-truth labels and to prepare the initial machine processing runs. Their assistance with this study was greatly appreciated.

CONTENTS

Section		Page
1.	INTRODUCTION	1-1
2.	EXPERIMENT OBJECTIVES AND DESCRIPTION	2-1
	2.1 OBJECTIVES	2-1
	2.2 GENERAL DESCRIPTION	2-1
3.	PROCEDURAL DESCRIPTION	3-1
4.	RESULTS	4-1
5.	CONCLUSIONS AND RECOMMENDATIONS	5-1
6.	REFERENCES	6-1
Appendix		A-1

TABLES

Table		Page
2-1	DATA SET	2-2
4-1	RAW DATA FOR 60, 45, AND 30 TYPE 1 DOTS	4-2
	(a) 60 type 1 dots	4-2
	(b) 45 type 1 dots	4-3
	(c) 30 type 1 dots	4-4
4-2	PROPORTION ESTIMATES ANOVA USING 60, 45, AND 30 STARTING DOTS	4-5
	(a) Using 60 starting dots	4-5
	(b) Using 45 starting dots	4-5
	(c) Using 30 starting dots	4-5
4-3	STARTING DOTS ANOVA USING MACHINE CLUSTERING PROPORTIONS, SAE CLUSTERING PROPORTIONS, MACHINE CLASSIFICATION PROPORTIONS AND SAE CLASSIFICATION PROPORTIONS	4-6
	(a) Machine clustering proportions	4-6
	(b) SAE clustering proportions	4-6
	(c) Machine classification proportions	4-6
	(d) SAE classification proportions	4-6
4-4	VARIANCES OF THE PROPORTION ESTIMATES	4-8
4-5	VARIANCES OF THE PCC	4-8
4-6	THE R-VALUES AVERAGED OVER 33 THREE-CATEGORY RUNS	4-10
A-1	DIFFERENCES BETWEEN PROPORTIONS AND GROUND TRUTH USING 60, 45, AND 30 STARTING DOTS	A-2
	(a) Using 60 starting dots	A-2
	(b) Using 45 starting dots	A-3
	(c) Using 30 starting dots	A-4
A-2	VARIANCES OF THE PROPORTION ESTIMATES USING 60, 45, AND 30 STARTING DOTS	A-5
	(a) Using 60 starting dots	A-5

Table		Page
	(b) Using 45 starting dots	A-6
	(c) Using 30 starting dots	A-7
A-3	VARIANCE OF THE PCC USING 60, 45, AND 30 STARTING DOTS	A-8
	(a) Using 60 starting dots	A-8
	(b) Using 45 starting dots	A-9
	(c) Using 30 starting dots	A-10
A-4	THE R-VALUES FOR 60, 45, AND 30 STARTING DOTS USING LABELING PROPORTIONS	A-11
	(a) R-values for 60 dots	A-11
	(b) R-values for 45 dots	A-12
	(c) R-values for 30 dots	A-13
A-5	THE R-VALUES FOR 60, 45, AND 30 STARTING DOTS USING GROUND-TRUTH PROPORTIONS	A-14
	(a) R-values for 60 dots	A-14
	(b) R-values for 45 dots	A-15
	(c) R-values for 30 dots	A-16

ACRONYMS

ANOVA	analysis of variance
LACIE	Large Area Crop Inventory Experiment
Landsat	land satellite
ME	machine estimate
P1	Procedure 1
PCC	probability of correct classification
PCC1	probability of correct classification for type 1 dots
PCC2	probability of correct classification for type 2 dots
pixels	picture elements
R	reduction coefficient
SAE	stratified areal estimate
TY	Transition year

1. INTRODUCTION

In transition from the Large Area Crop Inventory Experiment (LACIE) to the LACIE Transition Project (FY79), the basic components of Procedure 1 (P1) required investigation. P1, as used in LACIE, was a two-category procedure estimating wheat and nonwheat. In the mixed wheat states, a three-category classification was used to estimate winter wheat, spring wheat, and other in LACIE Phase III, but no investigation of the appropriate number of type 1 dots was made. This study was initiated to test a three-category classifier using corn, soybeans, and other as categories to examine the appropriate number of type 1 dots. Since a machine estimate (ME) and a stratified areal estimate (SAE) were produced by both nearest-neighbor clustering and maximum-likelihood classification in a standard P1 run, all four estimates were compared to ground-truth proportions. Each of these four proportion estimates were also analyzed in terms of the variance of the estimates and the variance of the corresponding probabilities of correct classification (PCC). The reduction coefficients, R-values, were calculated for all processing runs and were compared to previous two-category calculations.

2. EXPERIMENT OBJECTIVES AND DESCRIPTION

2.1 OBJECTIVES

The experiment was designed to test the three-category classifier using corn, soybeans, and other as the categories. The objectives of the experiment were:

a. To examine a three-category classifier proportion estimate in terms of the number of type 1 dots used

b. To examine a three-category classifier in terms of the variance of the estimate

c. To examine the evaluation criterion (the PCC) in terms of its variance

2.2 GENERAL DESCRIPTION

The experiment was planned to include the processing of 12 test segments using varying numbers of starting dots (type 1) and 105 bias correction dots (type 2). Of these 12 segments, 6 were obtained from the corn and soybean allocation and 6 from the LACIE Transition Year (TY) allocation. Detailed information about these data is described in table 2-1.

The crops chosen for both the six segments obtained from the corn and soybean allocation and the six segments obtained from the LACIE TY allocation were the major crops in the segment. The primary purpose of the test was to process corn and soybeans; however, if one of these crops were not adequately represented in a segment, another major crop was chosen to replace it as a crop of interest.

The number of type 1 dots (sets of 30, 45, and 60 dots) was varied in order to examine the effect of the number of dots used in a three-category classifier. To estimate the variance of the proportion estimates and the PCC, three independent sets of dots were selected from the 209 grid intersections for a fixed number of type 1 dots. Thus, a total of 108 processing runs was possible. To make the initial type 1 dot selections for a segment, three independent sets of 60 dots were randomly selected in the usual manner of skipping all border

TABLE 2-1.— DATA SET

Segment		Acquisition used	Major crops
Number	Location		
146	Kentucky	8180, 8198, 8234, 8270	Corn (C), Soybeans (X)
185	Minnesota	8169, 8197, 8205, 8224	Corn (C), Soybeans (X), Spring wheat, Sunflowers
804	Iowa	8229, 8247, 8274, 8292	Corn (C), Soybeans (X) Oats
812	Mississippi	8199, 8235, 8280, 8289	Soybeans (X), Cotton (K), Rice
824	Illinois	8163, 8235, 8271, 8307	Corn (C), Soybeans (X)
883	Iowa	8186, 8213, 8222, 8293	Corn (C), Soybeans (X)
1075	Nebraska	8133, 8206, 8259, 8296	Corn (C), Alfalfa (A)
1253	Oklahoma	8165, 8184, 8274, 8291	Soybeans (X), Alfalfa (A)
1341	Kansas	8113, 8167, 8186, 8293	Corn (C), Soybeans (X) Sorghum
1502	Colorado	8138, 8246, 8282, 8300	Corn (C), Sugar beets (Y) Winter wheat, Alfalfa
1572	Nebraska	8153, 8206, 8279, 8296	Corn (C), Pasture (P)
1591	Nebraska	8134, 8241, 8259, 8278	Corn (C), Sorghum (E)

or edge picture elements (pixels) described in reference 1. Next, 15 dots were randomly deleted to produce three sets of 45 type 1 dots. And again, 15 dots were randomly deleted for three sets of 30 type 1 dots. For both of these random deletions, each category was guaranteed to have at least one type 1 dot, thus restricting the deletion process.

Each processing consisted of a three-category version of the standard P1 clustering and classification. Proportion and PCC estimates were obtained from the automatically labeled clusters and from the maximum-likelihood classifier output. The SAE were also calculated for each ME using a set of 105 type 2 dots. For the three replications, using a fixed number of starting dots (30, 45, or 60), the type 2 dots were selected independently where the overlap (between sets of type 2 dots only) occurred from necessity. The abundance of border and edge pixels in the type 1 selections prohibited a third set of dots for three of the twelve test segments. This caused the total number of processing runs to be decreased to 99.

2.3 EXPERIMENT DESIGN

An analysis of variance (ANOVA) was planned for each set of starting dots (set of 60, set of 45, and set of 30) to determine differences between the proportion estimation procedures. The signed difference between each proportion estimate and the ground-truth estimate was used for the response variable. The linear model for the three analyses was as follows:

$$Y_{ijkm} = \mu + \sigma_i + \tau_j + C_k + \varepsilon_{ijkm}$$

where

μ = the overall mean of the observations

σ_i = the segment effect ($i = 1, 2, \cdots, 12$)

τ_j = the treatment or procedure effect ($j = 1, 2, 3, 4$)

C_k = the crop effect ($k = 1, 2$)

ε_{ijkm} = the random error for each observation (m represents the repetitions performed for each observation and is a function of i, j, and k.)

Y_{ijkm} = the response variable

An ANOVA was also planned for each proportion estimation procedure to determine differences between the number of starting dots used (60, 45, and 30). The response variable was again the signed difference between the proportion estimate and ground truth. The linear model for these four ANOVAs was as follows:

$$Y_{ijkm} = \mu + \sigma_i + \tau_j + C_k + \varepsilon_{ijkm}$$

where

τ_j = the treatment effect representing the number of starting dots used (j = 1, 2, 3)

μ = the overall mean of the observations

σ_i = the segment effect (i = 1, 2, ···, 12)

C_k = the crop effect (k = 1, 2)

ε_{ijkm} = the random error for each observation (m represents the repetitions performed for each observation and as a function of i, j, or k.)

A general linear model ANOVA program was used to generate the ANOVA tables (ref. 2).

To examine the variability in the performance of P1, estimates of the variance of the proportion estimates and the variance of the PCC estimates were to be computed. The variances were estimated by pooling the within-segment variances over each segment for each case of 30, 45, and 60 type 1 dots. These variances were then pooled over all segments for each case of 30, 45, and 60 type 1 dots. The equations for computing these variance estimates are as follows.

$$\mathrm{Var}\,[X] = \frac{\sum_{j=1}^{N}\sum_{i=1}^{M_j} (X_{ij} - \overline{X}_j)^2}{2N}$$

where

X_{ij} = the variable, proportion estimate, or PCC as measured for the *ith* sampling and *jth* segment

$\overline{X}_j$ = the average value of X_{ij} for the *jth* segment

N = total number of segments and range of j

M_j = total number of samplings which are dependent upon j and are in the range of i

Separate comparisons were planned for the variance of the proportion estimates and the variance of the PCC estimates. In each case, ratios between the variances for the estimates of the set of 45 type 1 dots and the set of 60 type 1 dots and between the variances for the estimates for the set of 30 and the set of 60 type 1 dots were to be calculated. These ratios were approximately distributed as F-statistics and, therefore, may be tested for statistically significant departures from unity. Statistical tables indicated that ratios with a value of approximately two or larger were significant at the 5-percent level if a total of 12 segments was used.

3. PROCEDURAL DESCRIPTION

This study was performed using ground-truth labels that were manually verified with an annotated aerial photograph and registered grid overlay. The grid overlay corresponds to the grid intersections on the land satellite (Landsat) film products. Border (spectrally mixed pixels) and edge (spatially misregistered pixels from acquisition to acquisition) were also identified and documented at this time since these types of pixels are not used as type 1 dots.

Standard P1 processing was performed. The type 1 dots started the nearest-neighbor clustering algorithm (ref. 3) with the following parameters.

a. CO = 60

b. Percent = 0

c. SEP = 1

d. STDMAX = 20

e. DLMIN = 0

f. R2 = 8191

g. NMIN2 = 18

h. ITMAX = 0

i P of N = 1

j. SC Seq. = S

k. Distance measure = L2 (Euclidean)

The NMIN2 parameter was changed from the standard value of 100 to 18 in order to prevent the deletion of small clusters.

The clusters were automatically labeled by the closest type 1 dot using an L2 distance criterion. The cluster statistics were then used in a maximum-likelihood classifier to classify the segment. Output reports included cluster proportion estimates, classification proportion estimates, their corresponding SAEs, type 1 PCCs, and type 2 PCCs.

Initially, the three-category version of P1 was run using each set of 60 type 1 dots for each segment. Following the completion of these runs, 15 type 1 dots were deleted at random from each set of type 1 dots and the processing was repeated using the 45 remaining dots. Finally, 15 more type 1 dots were deleted at random, and the 30 remaining dots were used to make the final runs.

4. RESULTS

The estimates obtained from the study are shown in tables 4-1a, 4-1b, and 4-1c for the sets of 60, 45, and 30 type 1 dots, respectively. The C1 and C2 are the two categories of interest that were processed with other (N). The ground-truth estimates of these categories are 400-random-dot counts, taken from annotated aircraft photography because digitized ground-truth maps were not available. The ME, SAE, type 1 dot PCC (PCC1), and type 2 dot PCC (PCC2) are shown for both cluster and classification results. The SAEs were computed on a category level for both MEs. Note that the PCC values are computed for the ME only.

The raw proportion estimates were differenced with the ground truth before analysis, and these signed differences appear in the appendix, table A-1a, A-1b, and A-1c for 60, 45, and 30 type 1 dots, respectively.

The first set of ANOVA tests was performed on the signed differences between proportions and ground truth for each different number of type 1 dots: 60, 45, and 30. This was to determine if any significant differences existed between the four methods of achieving a proportion estimate. These ANOVA tests appear in table 4-2a, 4-2b, and 4-2c. For each separate set of starting dots, no significant differences were found between the proportion estimates.

The ANOVA tests were also performed to detect differences between 30, 45, and 60 starting dots, based on the signed differences between the proportions and ground truth. These ANOVA tests appear in tables 4-3a, 4-3b, 4-3c, and 4-3d. No significant differences were found between numbers of starting dots (60, 45, and 30) for each proportion estimation technique: machine clustering, SAE clustering, machine classification, and SAE classification.

The variance of the proportion estimates on a per segment basis appear in the appendix, tables A-2a, A-2b, and A-2c for 60, 45, and 30 starting dots, respectively. Comparisons were more readily made when these variances were pooled over all the test segments for each number of starting dots, as in

TABLE 4-1.– RAW DATA ESTIMATES FOR 60, 45, AND 30 TYPE 1 DOTS

(a) 60 type 1 dots

Segment		Ground-truth data				Cluster						Classify					
						Machine estimate		Stratified areal estimate		PCC1	PCC2	Machine estimate		Stratified areal estimate		PCC1	PCC2
Number	Location	Crop 1	Percent	Crop 2	Percent	Crop 1	Crop 2	Crop 1	Crop 2			Crop 1	Crop 2	Crop 1	Crop 2		
1075	Nebraska	C	29	A	0	16	4	23	14	100.0	72.4	16	3	25	14	100.0	69.5
						25	8	36	11	100	75.2	25	2	36	11	100	72.1
						25	6	34	12	100	77.1	25	6	33	12	100	79.8
1341	Kansas	C	40	X	14	46	15	35	17	98.3	55.3	47	15	35	18	98.3	59.2
						41	13	39	16	98.3	56.2	41	17	37	16	98.3	57.1
						42	20	36	17	100	55.2	42	20	37	17	100	55.9
1591	Nebraska	C	15	E	7	15	5	19	4	98.3	83.3	16	4	16	3	98.3	83.2
						10	3	12	5	100	79.6	10	3	17	5	100	79.6
						14	7	15	6	98.3	79.9	14	5	16	7	100	81.6
146	Kentucky	C	17	X	46	15	44	20	46	98.3	73.8	15	44	20	46	98.3	79.8
						17	45	16	44	100	73.7	16	46	17	45	98.3	78.7
185	Minnesota	C	8	X	8	4	6	7	6	100	87.6	4	6	7	6	100	87.3
						3	2	6	10	100	85.7	3	2	5	10	100	85.6
						1	4	8	6	100	82.9	1	4	8	7	100	83.7
804	Iowa	C	45	X	29	47	26	40	34	100	80.0	47	26	41	33	100	79.0
						45	27	44	29	100	79.5	45	27	42	29	100	78.8
						43	27	41	27	100	82.9	43	27	41	27	100	82.7
212	Mississippi	X	48	K	7	49	5	50	5	100	80.0	48	5	50	6	100	83.8
						50	3	52	6	98.3	81.0	49	3	53	6	98.3	84.6
						51	5	42	7	100	78.1	50	5	42	7	100	78.1
824	Illinois	C	52	X	41	55	41	52	37	100	75.0	55	41	52	36	100	77.5
						52	43	46	44	96.7	81.6	52	43	48	42	96.6	82.5
883	Iowa	C	37	X	33	38	30	40	34	98.3	63.5	38	31	41	34	96.6	65.6
						40	31	36	33	98.3	69.8	39	31	38	33	96.7	71.8
1253	Oklahoma	X	33	A	3	32	1	29	2	96.7	87.7	32	1	31	1	96.6	92.3
						33	1	33	2	98.3	83.8	33	1	33	3	100	85.6
						32	1	26	2	100	81.0	32	1	27	2	100	85.1
1502	Colorado	C	11	Y	4	9	1	6	8	100	83.8	9	1	6	6	98.3	86.4
						3	1	12	4	98.3	86.7	3	1	12	5	100	86.3
						10	5	8	7	100	86.7	10	5	9	7	100	86.5
1572	Nebraska	C	14	P	64	6	76	14	59	98.3	72.4	5	76	14	59	96.7	73.3
						14	75	14	54	95.0	70.5	14	75	14	54	96.7	71.4
						17	69	15	63	98.3	74.3	6	70	14	65	98.3	73.3

Symbol definitions:

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE 4-1.— Continued.

(b) 45 type 1 dots

Segment		Ground-truth data				Cluster						Classify					
						Machine estimate		Stratified areal estimate		PCC1	PCC2	Machine estimate		Stratified areal estimate		PCC1	PCC 2
Number	Location	Crop 1	Percent	Crop 2	Percent	Crop 1	Crop 2	Crop 1	Crop 2			Crop 1	Crop 2	Crop 1	Crop 2		
1075	Nebraska	C	29	A	8	17	4	28	14	100.0	72.4	17	4	27	14	100.0	70.9
						25	7	33	12	100	73.1	24	7	34	12	100	76.7
						26	6	33	10	100	78.1	26	6	33	10	100	79.8
1341	Kansas	C	40	X	14	47	17	35	17	97.8	53.3	47	17	36	17	97.8	55.9
						37	21	39	15	100	53.3	37	20	39	16	100	58.1
						44	23	38	15	100	50.5	44	22	39	16	100	59.2
1591	Nebraska	C	15	E	7	17	4	17	3	93.3	79.4	17	4	17	3	88.9	80.4
						10	1	17	5	100	79.0	9	1	18	5	100	81.0
						14	8	16	7	97.7	78.8	14	7	17	6	100	80.4
146	Kentucky	C	17	X	46	12	50	18	47	97.8	81.0	12	50	19	45	97.8	81.9
						17	45	18	44	100	82.0	16	45	17	43	100	83.1
185	Minnesota	C	8	X	8	4	7	5	6	95.5	84.8	4	7	5	6	97.7	84.2
						3	2	5	11	100	85.7	3	2	5	11	100	85.4
						1	8	8	7	100	80.0	1	8	8	7	100	81.6
804	Iowa	C	46	X	29	44	26	40	34	100	79.0	44	26	40	34	100	78.2
						45	27	42	30	100	81.0	45	27	42	29	100	80.8
						45	28	41	28	100	81.9	44	28	41	28	97.8	81.7
812	Mississippi	X	48	K	7	48	7	50	5	95.6	80.0	48	8	49	6	95.5	81.7
						48	3	49	4	100	83.8	48	3	53	5	100	80.8
						50	6	43	7	100	78.1	49	6	43	7	97.7	78.8
824	Illinois	C	52	X	41	54	42	54	34	97.8	77.5	54	42	53	35	97.8	78.8
						51	43	47	44	100	81.6	51	43	49	43	100	82.3
883	Iowa	C	37	X	33	35	31	41	34	97.8	66.7	35	32	42	32	95.6	66.3
						38	30	38	33	100	70.9	38	30	38	33	97.8	72.9
1253	Oklahoma	X	33	A	3	35	1	29	2	97.8	86.8	35	1	32	3	100	88.5
						27	1	33	3	95.6	81.9	27	1	33	3	95.6	82.5
						33	1	28	2	100	84.8	33	1	28	1	100	86.1
1502	Colorado	C	11	Y	4	7	-	6	7	100	82.9	7	2	6	6	100	82.4
						3	1	12	5	100	87.6	3	1	12	5	100	87.4
						11	7	8	8	100	85.7	10	7	9	7	100	85.6
1572	Nebraska	C	14	P	64	6	81	14	58	95.6	67.6	6	81	14	59	95.6	68.6
						14	75	14	54	95.6	70.5	14	75	12	56	93.3	72.8
						5	72	14	65	93.3	69.5	6	72	14	64	93.2	70.5

Symbol definitions:

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE 4-1.— Concluded.

(c) 30 type 1 dots

Segment		Ground-truth data				Cluster						Classify					
						Machine estimate		Stratified areal estimate		PCC1	PCC2	Machine estimate		Stratified areal estimate		PCC1	PCC2
Number	Location	Crop 1	Percent	Crop 2	Percent	Crop 1	Crop 2	Crop 1	Crop 2			Crop 1	Crop 2	Crop 1	Crop 2		
1075	Nebraska	C	29	A	8	17	2	29	13	100.0	67.6	17	1	29	13	100.0	62.3
						26	10	33	10	100	74.0	26	10	33	12	100	76.7
						23	6	32	8	100	74.3	22	6	31	8	100	79.6
1341	Kansas	C	40	X	14	53	18	35	18	90.0	56.2	52	18	37	17	90.0	56.3
						31	29	43	14	96.7	52.4	31	22	44	14	100	56.2
						46	31	38	15	100	46.7	46	31	37	16	96.6	47.6
1591	Nebraska	C	15	E	7	15	3	15	3	96.7	76.5	15	2	17	3	96.7	80.0
						10	1	17	5	100	78.1	10	1	16	5	100	79.6
						10	5	16	7	100	75.0	11	4	16	7	100	75.7
146	Kentucky	C	17	X	46	14	53	20	47	80.0	71.4	14	54	21	45	86.7	69.9
						16	45	17	42	100	79.8	16	45	18	43	100	83.1
185	Minnesota	C	8	X	8	0	5	5	6	96.6	83.3	0	5	5	6	96.6	85.1
						2	3	3	9	100	83.3	2	3	3	10	100	84.0
						5	3	8	6	96.7	76.2	5	3	8	6	89.7	76.2
804	Iowa	C	46	X	29	44	26	41	35	100	76.2	45	27	40	34	100	77.7
						49	27	42	29	100	79.0	48	27	43	29	100	79.6
						48	26	42	22	100	81.0	47	27	41	22	100	81.6
812	Mississippi	X	48	K	7	43	5	49	6	100	78.1	42	5	49	6	100	81.7
						50	3	47	6	100	82.9	50	4	52	7	100	80.4
						55	7	43	8	93.3	78.1	54	7	42	6	96.6	78.8
824	Illinois	C	52	X	41	57	38	50	37	100	77.5	57	39	49	38	100	75.0
						51	41	47	42	100	75.5	50	42	49	41	100	78.9
883	Iowa	C	37	X	33	42	32	39	34	100	64.6	41	32	40	35	96.7	66.0
						44	32	34	37	100	72.7	43	32	36	34	100	72.1
1253	Oklahoma	X	33	A	3	32	1	29	2	96.7	86.8	33	1	31	3	100	86.7
						24	3	33	3	100	80.0	24	3	33	4	100	81.7
						33	1	22	2	100	84.8	33	1	22	1	100	86.1
1502	Colorado	C	11	Y	4	4	2	6	7	100	82.9	4	2	7	6	100	82.8
						3	1	12	4	100	87.5	4	1	12	5	100	87.5
						12	3	9	7	100	89.5	12	3	9	7	100	89.9
1572	Nebraska	C	14	P	64	7	88	15	58	83.3	70.5	7	87	15	59	83.3	72.4
						9	81	12	57	96.8	68.6	9	81	11	58	96.6	69.9
						6	71	14	66	93.3	68.6	6	72	13	66	90.0	70.5

Symbol definitions:

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE 4-2.– PROPORTION ESTIMATES ANOVA USING 60, 45, AND 30 STARTING DOTS

Source of variation	Degrees of freedom	Sum of squares	Mean square	F-value
(a) Using 60 starting dots				
Mean, segment, and crop effect	13	505.31		
Estimation procedure effect	3	32.59	10.86	0.85[a]
Error	248	3149.10	12.70	
Total	264	3687.00		
(b) Using 45 starting dots				
Mean, segment, and crop effect	13	697.02		
Estimation procedure effect	3	10.80	3.60	0.25[a]
Error	248	3558.17	14.34	
Total	264	4266.00		
(c) Using 30 starting dots				
Mean, segment, and crop effect	13	1223.56		
Estimation procedure effect	3	1.83	0.61	0.03[a]
Error	248	5564.61	22.44	
Total	264	6790.00		

[a]Indicates nonsignificance at the α = 5-percent level.

TABLE 4-3.— STARTING DOTS ANOVA USING MACHINE CLUSTERING PROPORTIONS, SAE CLUSTERING PROPORTIONS, MACHINE CLASSIFICATION PROPORTIONS, AND SAE CLASSIFICATION PROPORTIONS

Source of variation	Degrees of freedom	Sum of squares	Mean square	F-value
(a) Machine clustering proportions				
Mean, segment, and crop effect	13	1918.96		
Starting dot effect	2	14.92	7.46	0.38[a]
Error	183	3569.12	19.50	
Total	198	5503.00		
(b) SAE clustering proportions				
Mean, segment, and crop effect	13	461.60		
Starting dot effect	2	3.67	1.83	0.22[a]
Error	183	1550.73	8.47	
Total	198	2016.00		
(c) Machine classification proportions				
Mean, segment, and crop effect	13	1968.49		
Starting dot effect	2	17.04	8.52	0.46[a]
Error	183	3394.47	18.55	
Total	198	5380.00		
(d) SAE classification proportions				
Mean, segment, and crop effect	13	487.19		
Starting dot effect	2	0.39	0.20	0.03[a]
Error	183	1356.42	7.41	
Total	198	1844.00		

[a]No significant difference was found at the α = 5-percent level.

table 4-4. The variances of the MEs, both clustering and classification, were significantly decreased by increasing the number of starting dots from 30 to 45 for both categories of interest. These variances were again decreased, but not significantly, by increasing the numbers of starting dots from 45 to 60. The variances of the SAE estimates, both clustering and classification, did not significantly differ for any change in the number of starting dots. For 60 starting dots, no significant differences were found between any of the proportion estimation techniques for either crop.

The variances of the PCCs on a per segment basis appear in the appendix, tables A-3a, A-3b, and A-3c for 60, 45, and 30 starting dots, respectively. These variances are pooled over all test segments for each number of starting dots and appear in table 4-5. The variance of the PCC1 for clustering and classification was significantly decreased when the number of starting dots changed from 30 to 45 and from 45 to 60. The variance of the PCC2 for clustering and classification was significantly decreased when the number of starting dots changed from 30 to 45. When the number of starting dots changed from 45 to 60, the variance of PCC2 increased for both clustering and classification, but the increase was not significant. For 60 starting dots, the variance of PCC1 was significantly different from the variance of PCC2 for both clustering and classification. This significance can be attributed to the difference between training and test data.

The reduction coefficient, R, has been presented as a method of observing how much the machine classification reduces the variance of the SAE proportion estimation (ref. 3) in comparison with the variance of a simple random sample estimate. In the computation of the R-values, the omission and commission rates are computed by comparing the machine labels of the type 2 dots to the ground truth. The sampling error can be computed using the ground truth proportions or the labeling proportion from the type 2 dots denoted herein as π. The R-values were computed for both clustering and classification for all of the three-category runs. The omission and commission rates and the R-values are presented, for both ground truth and labeling proportions, in the appendix, tables A-4a, A-4b, and A-4c. The R-values were then averaged over the 33 runs for a particular number of starting dots (60, 45, or 30) and using

TABLE 4-4.— VARIANCES[a] OF THE PROPORTION ESTIMATES

Number of starting dots	Crop	Cluster		Classify	
		Machine	Bias	Machine	Bias
30	1	23.389	5.562	22.437	6.549
	2	14.021	6.312	12.875	4.597
45	1	11.285	5.181	10.625	6.111
	2	6.479	7.431	6.181	5.076
60	1	7.715	8.194	8.007	6.542
	2	3.493	5.125	3.083	5.264

[a]Statistical significance at the 5-percent level is found whenever the ratio of variances is at least two.

TABLE 4-5.— VARIANCES OF THE PCC

Number of starting dots	Cluster		Classify	
	PCC1	PCC2	PCC1	PCC2
30	16.751	10.992	12.782	14.446
45	3.053	4.590	5.254	5.484
60	1.138	5.725	0.954	7.408

both the ground-truth and the labeling proportion. The average R-values and their corresponding variances are presented in table 4-6. The averaged R-values were only slightly higher than those for the two-category P1 runs (0.718 and 0.714) as documented in reference 3.

TABLE 4-6.— THE R-VALUES AVERAGED OVER 33 THREE-CATEGORY RUNS

Number of starting dots	π	Cluster		Classify	
		R-mean	R-variance	R-mean	R-variance
60	Ground truth	0.738	0.024	0.731	0.034
60	Labeling proportion	.739	.025	.729	.032
45	Ground truth	.736	.032	.730	.029
45	Labeling proportion	.735	.032	.729	.030
30	Ground truth	.775	.029	.746	.034
30	Labeling proportion	.774	.029	.746	.034

5. CONCLUSIONS AND RECOMMENDATIONS

Upon examination of both ME and SAE proportion estimates produced by clustering and classification, no significant differences were found between the proportion estimates and ground-truth estimates. Since this was the case in previous two-category studies (ref. 4), it is not considered unusual in the three-category case, but instead, indicates that conclusions should be made on the basis of the consistency or variance of the estimates as well as the accuracy.

When testing the variances of the ME proportion estimates, a significant reduction in the variances was found when the number of starting dots was increased from 30 to 45. The variances were again reduced, although not significantly, when the number of starting dots was increased from 45 to 60. From these results, 60 starting dots are recommended for a three-category classifier.

When examining the variances of the estimates for the four estimation procedures (using 60 starting dots), no significant differences were found between procedures. Thus, only the machine clustering may be used to produce an estimate and the SAE computations and maximum-likelihood classification can be deleted. This will allow two advantages over P1: (1) using only clustering will eliminate the additional machine time required by classification, and (2) deleting the SAE will minimize the analyst-labeling-time required because only type 1 dot labeling will be necessary.

The variance of the PCC1 was significantly lower for 60 starting dots than for either the set of 45 or 30 dots. Since the type 1 dots are the training data, an increase in the training sample size is expected to produce significant decreases in the variance of the PCC1. For the PCC2, a significant reduction in the variance was observed when the number of starting dots was increased from 30 to 45. No significant differences were observed when the number of starting dots was increased from 45 to 60. Thus, the variance of

PCC1 decreased when the number of starting dots increased up to 60, and the variance of PCC2 decreased when the number of starting dots increased to 45 and then statistically stabilized. This further reinforces the choice of 60 starting dots.

The efficiency of P1 in reducing the variance of a proportion estimate obtained from SAE has been presented in reference 3. In this experiment, virtually no difference existed between the R-values, regardless of the number of starting dots used or the proportion estimation procedure. There were no cases where the R-value was lower for clustering than for classification. This would indicate that classification was better than clustering, but the differences between the R-values were consistently very small. As in the two-category case, these R-values indicate that not much is being gained by classification or clustering over a simple random sample. Since these R-values are the product of the best possible labeling of dots (ground truth), an improved procedure to P1 seems desirable to improve the cost-effectiveness of this machine processing.

To summarize, the recommendations resulting from this study are as follows:

a. A set of 60 starting dots should be used in a three-category classifier.

b. The ME produced by nearest-neighbor clustering is an adequate estimator.

c. More study is needed in the area of an alternative for P1.

6. REFERENCES

1. LACIE Transition Project (FY79) Detailed Analysis Procedures. LACIE-LACIE-00724, JSC-13756, March 1979.

2. Graybill, Franklin A.: Theory and Application of the Linear Model. Duxbury Press, North Scituate, Mass., 1976, pp. 247-252.

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

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

APPENDIX

APPENDIX

Tables A-1 through A-5, included in the appendix, are supplemental material referred to in section 4 of this document.

TABLE A-1.— DIFFERENCES BETWEEN PROPORTIONS AND GROUND TRUTH USING 60, 45, AND 30 STARTING DOTS

(a) Using 60 starting dots

Segment		Crop 1	Crop 2	Cluster				Classify			
				Machine estimate		Stratified areal estimate		Machine estimate		Stratified areal estimate	
Number	Location			Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2
1075	Nebraska	C	A	-13	-4	-1	6	-13	-5	-1	6
				-3	0	7	3	-4	0	7	3
				-3	-2	5	4	-3	-2	4	4
1341	Kansas	C	X	6	1	-5	3	7	1	-5	4
				1	4	-1	2	1	3	-3	2
				2	6	-4	3	2	6	-3	3
1591	Nebraska	C	E	0	-2	4	-3	1	-3	1	-4
				-5	-4	3	-2	-5	-4	2	-2
				-1	0	0	-1	-1	-2	1	0
146	Kentucky	C	X	-2	-2	3	0	-2	-2	3	0
				0	0	-1	-2	-1	0	0	-1
185	Minnesota	C	X	-4	-2	-1	-2	-4	-2	-1	-2
				-5	-6	-2	2	-5	-6	-2	2
				-7	-4	0	-2	-7	-4	0	-1
804	Iowa	C	X	1	-3	-6	5	2	-3	-5	4
				-1	-2	-2	0	-1	-2	-4	0
				-3	-2	-5	-2	-3	-2	-5	-2
812	Mississippi	X	K	1	-2	2	-2	0	-2	2	-1
				2	-4	4	-1	1	-4	5	-1
				3	-2	-6	0	2	-2	-6	0
824	Illinois	C	X	3	0	0	-4	3	0	0	-5
				0	2	-6	3	0	2	-4	1
883	Iowa	C	X	1	-3	3	1	1	-2	4	1
				3	-2	-1	0	2	-2	1	0
1253	Oklahoma	X	A	-1	-2	-4	-1	-1	-2	-2	-2
				0	-2	0	0	0	-2	0	0
				-1	-2	-7	-1	-1	-2	-6	-1
1502	Colorado	C	Y	-2	-3	-5	4	-2	-3	-5	2
				-8	-3	1	0	-8	-3	1	1
				-1	1	-3	3	-1	1	-2	3
1572	Nebraska	C	P	-8	12	0	-5	-8	12	0	-5
				0	11	0	-10	0	11	0	-10
				-7	5	1	-1	-8	6	0	1

Symbol definitions:

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-1.— Continued.

(b) Using 45 starting dots

Segment		Crop 1	Crop 2	Cluster				Classify			
				Machine estimate		Stratified areal estimate		Machine estimate		Stratified areal estimate	
Number	Location			Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2
1075	Nebraska	C	A	-12 -4 -3	-4 -1 -2	-1 -4 -4	6 4 2	-12 -5 -3	-4 -1 -2	-2 5 4	6 4 2
1341	Kansas	C	X	7 -3 4	3 7 9	-5 -1 -2	3 1 1	7 -3 4	3 6 8	-4 -1 -1	3 2 2
1591	Nebraska	C	E	2 -5 -1	-3 -6 1	2 2 1	-4 -2 0	2 -6 -1	-3 -6 0	2 3 2	-4 -2 -1
146	Kentucky	C	X	-5 0	4 -1	1 1	1 -2	-5 -1	4 -1	2 0	-1 -3
185	Minnesota	C	X	-4 -5 -7	-1 -6 0	-3 -3 0	-2 3 -1	-4 -5 -7	-1 -6 0	-3 -3 0	-2 3 -1
804	Iowa	C	X	-2 -1 -1	-3 -2 -1	-6 -4 -5	5 1 -1	-2 -1 -2	-3 -2 -1	-6 -4 -5	5 0 -1
812	Mississippi	X	K	0 0 2	0 -4 -1	2 1 -5	-2 -3 0	0 0 1	1 -4 -1	1 5 -5	-1 -2 0
824	Illinois	C	X	2 -1	1 2	2 -5	-7 3	2 -1	1 2	1 -3	-6 2
883	Iowa	C	X	-2 1	-2 -3	1 1	1 0	-2 1	-1 -3	5 1	-1 0
1253	Oklahoma	X	A	2 -6 0	-2 -2 -2	-4 0 -5	-1 0 -1	2 -6 0	-2 -2 -2	-1 0 -5	0 0 -2
1502	Colorado	C	Y	-4 -8 0	-2 -3 3	-5 1 -3	3 1 4	-4 -8 -1	-2 -3 3	-5 1 -2	2 1 3
1572	Nebraska	C	P	-8 0 -9	17 11 8	0 0 0	-6 -10 1	-8 0 -8	17 11 8	0 -2 0	-5 -8 0

Symbol definitions:

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-1.– Concluded.

(c) Using 30 starting dots

Segment		Crop 1	Crop 2	Cluster				Classify			
				Machine estimate		Stratified areal estimate		Machine estimate		Stratified areal estimate	
Number	Location			Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2
1075	Nebraska	C	A	-12 -13 -1	-6 2 -2	0 4 3	5 2 0	-12 -3 -1	-7 2 -2	0 4 2	5 4 0
1341	Kansas	C	X	13 -9 6	4 16 17	-5 3 -2	4 0 1	12 -9 6	4 14 17	-3 4 -3	3 0 2
1591	Nebraska	C	E	0 -5 -5	-4 -6 -2	0 2 1	-4 -2 0	0 -5 -4	-5 -6 -3	2 1 1	-4 -2 0
146	Kentucky	C	X	-3 1	7 -2	3 0	1 -4	-3 0	8 -2	4 2	-1 -4
186	Minnesota	C	X	-8 -6 -3	-3 -5 -5	-3 -5 0	-2 1 -2	-8 -6 -3	-3 -5 -5	-3 -5 0	-2 2 -2
804	Iowa	C	X	-2 3 2	-3 -2 -3	-5 -4 -4	6 0 -1	-1 2 1	-2 -2 -2	-6 -3 -5	5 0 -1
812	Mississippi	X	K	-5 -1 7	-2 -4 0	1 0 -5	-1 -2 1	-6 -1 6	-2 -3 0	1 3 -6	-1 0 -1
824	Illinois	C	X	5 6	-3 -2	-2 -3	-4 1	5 5	-2 -2	-3 -2	-3 -1
883	Iowa	C	X	5 7	-1 -1	2 -3	1 4	4 6	-1 -1	3 -1	2 1
1253	Oklahoma	X	A	-1 -9 0	-2 0 -2	-4 0 -5	-1 0 -1	0 -9 0	-2 0 -2	-2 0 -5	0 1 -2
1502	Colorado	C	Y	-7 -8 1	-2 -3 -1	-5 1 -2	3 0 3	-7 -7 1	-2 -3 -1	-4 1 -2	2 1 3
1572	Nebraska	C	P	-7 -5 -8	24 17 7	1 -2 0	-6 -7 2	-7 -5 -8	23 17 8	1 -3 -1	-5 -6 2

Symbol definitions:

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-2.— VARIANCES OF THE PROPORTION ESTIMATES USING 60, 45, AND 30 STARTING DOTS

(a) Using 60 starting dots

Segment number	Cluster				Classify			
	Machine estimate		Stratified areal estimate		Machine estimate		Stratified areal estimate	
	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2
1075	33.333	4.000	17.333	2.333	30.333	6.333	16.333	2.333
1341	7.000	6.333	4.333	0.333	10.333	6.333	1.333	1.000
1591	7.000	4.000	4.333	1.000	9.333	1.000	0.333	4.000
185	2.333	4.000	1.000	5.333	2.333	4.000	1.000	4.333
804	4.000	0.333	4.333	13.000	4.000	0.333	0.333	9.333
812	1.000	1.333	28.000	1.000	1.000	1.333	32.333	0.333
1253	0.333	0.0	12.333	0.333	0.333	0.0	9.333	1.000
1502	14.333	5.333	9.333	4.333	14.333	5.333	9.000	1.000
1572	19.000	14.333	0.333	20.333	21.333	10.333	0.0	30.333
146	1.000	1.000	4.000	1.000	0.250	1.000	2.250	0.250
824	2.250	1.000	9.000	12.250	2.250	1.000	4.000	9.000
883	1.000	0.250	4.000	0.250	0.250	0.0	2.250	0.250

TABLE A-2.— Continued.

(b) Using 45 starting dots

Segment number	Cluster				Classify			
	Machine estimate		Stratified areal estimate		Machine estimate		Stratified areal estimate	
	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2
1075	24.333	2.333	8.333	4.000	23.333	2.333	14.333	4.000
1341	26.333	9.333	4.333	1.333	26.333	6.333	3.000	0.333
1591	12.333	12.333	0.333	4.000	16.333	9.000	0.333	2.333
185	2.333	10.333	3.000	7.000	2.333	10.333	3.000	7.000
804	0.333	1.000	1.000	9.333	0.333	1.000	1.000	10.333
812	1.333	4.333	14.333	2.333	0.333	6.333	25.333	1.000
1253	17.333	0.0	7.000	0.333	17.333	0.0	7.000	1.333
1502	16.000	10.333	9.333	2.333	12.333	10.333	9.000	1.000
1572	24.333	21.000	0.0	31.000	21.333	21.000	1.333	16.333
146	6.250	6.250	0.0	2.250	4.000	6.250	1.000	1.000
824	2.250	0.250	12.250	25.000	2.250	0.250	4.000	16.000
883	2.250	0.250	2.250	0.250	2.250	1.000	4.000	0.250

TABLE A-2.— Concluded.

(c) Using 30 starting dots

Segment number	Cluster				Classify			
	Machine estimate		Stratified areal estimate		Machine estimate		Stratified areal estimate	
	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2	Crop 1	Crop 2
1075	34.333	16.000	4.333	6.333	34.333	20.333	4.000	7.000
1341	126.333	49.000	16.333	4.333	117.000	46.333	16.333	2.333
1591	8.333	4.000	1.000	4.000	7.000	2.333	0.333	4.000
185	6.333	1.333	6.333	3.000	6.333	1.333	6.333	5.333
804	7.000	0.333	0.333	14.333	2.333	0.0	2.333	10.333
812	36.333	4.000	9.333	1.333	37.333	2.333	26.333	0.333
1253	24.333	1.333	7.000	0.333	27.000	1.333	6.333	2.333
1502	24.333	1.000	9.000	3.000	21.333	1.000	6.333	1.000
1572	2.333	73.000	2.333	24.333	2.333	57.000	4.000	19.000
146	1.000	16.000	2.250	6.250	1.000	20.250	2.250	1.000
824	9.000	2.250	2.250	6.250	12.250	2.250	0.0	2.250
883	1.000	0.0	6.250	2.250	1.000	0.0	4.000	0.250

TABLE A-3.– VARIANCE OF THE PCC USING 60, 45, AND 30 STARTING DOTS

(a) Using 60 starting dots

Segment number	Cluster		Classify	
	PCC1	PCC2	PCC1	PCC2
1075	0.0	12.423	0.0	28.690
1341	0.963	0.333	0.963	4.680
1591	0.963	5.230	0.963	3.253
185	0.0	5.590	0.0	3.243
804	0.0	4.103	0.0	4.823
812	0.963	2.170	0.963	10.663
1253	2.723	11.323	3.853	16.163
1502	0.963	2.803	0.963	0.010
1572	3.630	3.610	0.853	1.203
146	0.722	0.303	0.0	0.303
824	2.723	10.890	2.890	6.250
883	0.0	9.923	0.002	9.610

TABLE A-3.— Continued.

(b) Using 45 starting dots

Segment number	Cluster		Classify	
	PCC1	PCC2	PCC1	PCC2
1075	0.0	9.663	0.0	20.410
1341	1.613	2.613	1.613	2.823
1591	11.590	0.093	41.070	0.120
185	6.750	9.390	1.763	3.773
804	0.0	2.203	1.613	2.203
812	6.453	8.423	5.063	2.203
1253	4.840	6.070	6.453	9.120
1502	0.0	5.590	0.0	6.413
1572	1.763	2.170	1.843	4.423
146	1.210	0.250	1.210	0.360
824	1.210	4.203	1.210	3.063
883	1.210	4.410	1.210	10.890

TABLE A-3.– Concluded.

(c) Using 30 starting dots

Segment number	Cluster		Classify	
	PCC1	PCC2	PCC1	PCC2
1075	0.0	13.890	0.0	34.443
1341	25.963	22.863	25.853	24.943
1591	3.630	2.403	3.630	5.643
185	3.743	19.253	27.543	23.543
804	0.0	5.813	0.0	3.803
812	14.963	7.680	3.853	2.110
1253	3.630	12.213	0.0	7.453
1502	0.0	11.543	0.0	13.043
1572	49.083	1.203	45.563	1.703
146	100.000	17.640	44.223	43.560
824	0.0	1.000	0.0	3.802
883	0.0	16.402	2.723	9.303

TABLE A-4.— THE R-VALUES FOR 60, 45, AND 30 STARTING DOTS USING LABELING PROPORTIONS

(a) R-values for 60 starting dots

Segment		Crop 1	Crop 2	Cluster [Mean (μ) = 0.73852; variance (σ^2) = 0.025]				Classify [Mean (μ) = 0.72888; variance (σ^2) = 0.032]			
Number	Location			π_{10}	π_{01}	π	R	π_{10}	π_{01}	π	R
1075	Nebraska	C	A	0.033 .125 .067	0.600 .479 .444	0.429 .462 .429	0.786 .817 .707	0.067 .107 .050	0.622 .438 .409	0.429 .462 .423	0.852 .762 .646
1341	Kansas	C	X	.469 .408 .460	.023 .321 .345	.533 .533 .524	.666 .926 .961	.458 .429 .447	.250 .286 .345	.538 .533 .539	.910 .917 .966
1691	Nebraska	C	E	.059 .076 .111	.588 .577 .436	.167 .248 .221	.833 .835 .787	.071 .077 .062	.588 .560 .500	.168 .243 .214	.855 .824 .756
146	Kentucky	C	X	.164 .344	.155 .123	.690 .640	.561 .697	.154 .313	.172 .140	.690 .640	.588 .591
186	Minnesota	C	X	.043 .011 .034	.667 .867 .938	.114 .143 .162	.879 .934 .997	.044 .011 .023	.667 .867 .934	.118 .144 .164	.879 .934 .992
804	Iowa	C	X	.269 .233 .194	.089 .107 .054	.752 .714 .706	.608 .573 .411	.269 .267 .194	.114 .096 .066	.752 .712 .072	.637 .586 .413
812	Mississippi	X	K	.191 .136 .250	.190 .197 .113	.662 .581 .505	.619 .566 .586	.174 .093 .250	.190 .180 .113	.556 .587 .505	.599 .487 .586
824	Illinois	C	X	.778 .500	.014 .011	.888 .898	.880 .617	.778 .500	.014 .011	.888 .897	.880 .616
883	Iowa	C	X	.308 .300	.243 .196	.729 .651	.828 .753	.269 .241	.267 .196	.729 .659	.813 .700
1253	Oklahoma	X	A	.016 .117 .183	.126 .178 .176	.377 .429 .324	.419 .500 .622	.046 .102 .104	.103 .166 .147	.376 .433 .337	.268 .446 .454
1502	Colorado	C	Y	.055 .000 .068	.714 .737 .412	.133 .181 .162	.921 .778 .716	.044 .000 .069	.667 .737 .412	.117 .186 .163	.878 .775 .717
1572	Nebraska	C	P	.636 .867 .455	.091 .013 .146	.733 .714 .790	.823 .935 .850	.571 .867 .455	.078 .013 .169	.733 .714 .790	.830 .935 .874

Symbol definitions:

R = reduction coefficient

π = the probability a pixel is labeled wheat

π_{10} = the probability a pixel is classified wheat and labeled nonwheat

π_{01} = the probability a pixel is classified nonwheat and labeled wheat

<u>Crop codes</u>

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-4.– Continued.

(b) R-values for 45 starting dots

Segment		Crop 1	Crop 2	Cluster [Mean (μ) = 0.73476; variance (σ^2) = 0.032]				Classify [Mean (μ) = 0.72915; variance (σ^2) = 0.030]			
Number	Location			π_{10}	π_{01}	π	R	π_{10}	π_{01}	π	R
1075	Nebraska	C	A	0.060	0.578	0.429	0.796	0.068	0.591	0.427	0.831
				.125	.438	.462	.784	.071	.426	.456	.701
				.100	.378	.429	.695	.067	.386	.423	.651
1341	Kansas	C	X	.551	.232	.533	.947	.489	.218	.539	.907
				.367	.339	.533	.914	.347	.339	.533	.902
				.560	.345	.524	.991	.417	.309	.534	.924
1591	Nebraska	C	E	.106	.588	.167	.902	.094	.647	.167	.921
				.063	.654	.248	.871	.038	.654	.248	.825
				.099	.522	.221	.834	.075	.500	.216	.780
146	Kentucky	C	X	.192	.121	.690	.546	.231	.105	.687	.559
				.219	.140	.640	.594	.250	.105	.640	.571
185	Minnesota	C	X	.065	.583	.114	.867	.067	.583	.119	.867
				.011	.800	.143	.880	.011	.800	.146	.880
				.079	.875	.152	.997	.057	.875	.156	.990
804	Iowa	C	X	.231	.127	.752	.624	.269	.128	.750	.661
				.200	.093	.714	.509	.233	.095	.712	.549
				.229	.054	.706	.452	.226	.055	.702	.450
812	Mississippi	X	K	.213	.172	.552	.622	.196	.138	.558	.554
				.182	.148	.581	.554	.140	.230	.587	.615
				.250	.132	.505	.612	.235	.132	.510	.594
824	Illinois	C	X	.770	.014	.888	.880	.778	.014	.888	.880
				.400	.011	.898	.519	.300	.023	.896	.496
883	Iowa	C	X	.192	.283	.625	.741	.269	.232	.726	.787
				.020	.214	.651	.465	.172	.196	.659	.629
1253	Oklahoma	X	A	.136	.100	.377	.439	.109	.100	.385	.389
				.067	.218	524	.483	.069	.244	.437	.501
				.127	.176	.324	.534	.132	.121	.327	.477
1502	Colorado	C	Y	.077	.786	.133	.975	.079	.846	.127	.992
				.000	.684	.181	.725	.000	.684	.184	.726
				.080	.412	.162	.742	.080	.412	.163	.742
1572	Nebraska	C	P	.714	.065	.733	.912	.679	.065	.733	.889
				.833	.013	.714	.910	.828	.014	.718	.908
				.545	.181	.790	.931	.545	.157	.790	.914

Symbol definitions:

R = reduction coefficient

π = the probability a pixel is labeled wheat

π_{10} = the probability a pixel is classified wheat and labeled nonwheat

π_{01} = the probability a pixel is classified nonwheat and labeled wheat

<u>Crop codes</u>

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-4.— Concluded.

(c) R-values for 30 starting dots

Segment		Crop 1	Crop 2	Cluster [Mean (μ) = 0.77448; variance (σ^2) = 0.029]				Classify [Mean (μ) = 0.74633; variance (σ^2) = 0.034]			
Number	Location			π_{10}	π_{01}	π	R	π_{10}	π_{01}	π	R
1075	Nebraska	C	A	0.050	0.689	0.429	0.877	0.050	0.682	0.423	0.872
				.196	.333	.462	.773	.125	.340	.456	.695
				.163	.356	.429	.779	.133	.302	.417	.667
1341	Kansas	C	X	.510	.161	.467	.879	.511	.143	.544	.859
				.306	.339	.533	.875	.246	.321	.533	.812
				.680	.273	.524	.997	.688	.236	.534	.993
1691	Nebraska	C	E	.118	.706	.167	.966	.096	.647	.170	.922
				.076	.654	.248	.889	.064	.640	.243	.863
				.086	.739	.221	.952	.088	.696	.223	.931
146	Kentucky	C	X	.385	.155	.690	.783	.462	.158	.687	.845
				.313	.105	.640	.638	.219	.105	.640	.536
185	Minnesota	C	X	.075	.833	.114	.989	.056	.833	.119	.900
				.044	.800	.143	.952	.035	.800	.150	.938
				.101	1.000	.152	.983	.112	.938	.162	.997
804	Iowa	C	X	.269	.177	.752	.734	.308	.143	.748	.719
				.300	.067	.714	.561	.276	.068	.718	.539
				.258	.041	.705	.452	.233	.027	.709	.385
812	Mississippi	X	K	.170	.259	.552	.677	.109	.241	.558	.582
				.205	.115	.581	.531	.195	.197	.598	.640
				.308	.075	.505	.596	.294	.038	.510	.519
824	Illinois	C	X	.889	.014	.888	.961	.889	.014	.888	.961
				.500	.045	.898	.772	.400	.047	.895	.694
883	Iowa	C	X	.346	.157	.729	.764	.280	.174	.734	.732
				.267	.086	.659	.558	.241	.070	.663	.497
1263	Oklahoma	X	A	.121	.125	.377	.446	.108	.150	.381	.455
				.100	.289	.429	.604	.085	.289	.433	.580
				.127	.176	.324	.534	.132	.121	.327	.477
1502	Colorado	C	Y	.066	.857	.133	.990	.069	.917	.121	.9996
				.000	.684	.181	.725	.000	.684	.183	.726
				.045	.353	.162	.597	.048	.313	.162	.569
1572	Nebraska	C	P	.821	.000	.733	.862	.760	.000	.733	.804
				.867	.027	.714	.950	.862	.027	.718	.955
				.455	.205	.790	.904	.455	.181	.790	.885

Symbol definitions:

R = reduction coefficient

π = the probability a pixel is labeled wheat

π_{10} = the probability a pixel is classified wheat and labeled nonwheat

π_{01} = the probability a pixel is classified nonwheat and labeled wheat

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-5.— THE R-VALUES FOR 60, 45, AND 30 STARTING DOTS USING GROUND-TRUTH PROPORTIONS

(a) R-values for 60 starting dots

Segment		Crop 1	Crop 2	Cluster [Mean (μ) = 0.73779; variance (σ^2) = 0.024]				Classify [Mean (μ) = 0.73136; variance (σ^2) = 0.034]			
Number	Location			π_{10}	π_{01}	π	R	π_{10}	π_{01}	π	R
1075	Nebraska	C	A	0.033 .125 .067	0.600 .479 .444	0.37	0.776 .815 .701	0.067 .107 .050	0.622 .438 .409	0.37	0.849 .758 .636
1341	Kansas	C	X	.469 .408 .460	.023 .321 .346	.54	.664 .926 .962	.468 .429 .447	.250 .286 .345	.54	.910 .917 .956
15[illegible]1	Nebraska	C	E	.059 .076 .111	.588 .577 .436	.22	.819 .840 .787	.071 .077 .062	.588 .560 .500	.22	.840 .829 .753
146	Kentucky	C	X	.164 .344	.165 .123	.63	.540 .696	.154 .313	.172 .140	.63	.566 .690
[illegible]35	Minnesota	C	X	.043 .011 .034	.667 .867 .938	.16	.861 .932 .997	.044 .011 .023	.667 .867 .934	.16	.863 .932 .991
804	Iowa	C	X	.269 .233 .194	.089 .107 .054	.75	.587 .588 .422	.269 .267 .194	.114 .095 .055	.75	.636 .598 .425
812	Mississippi	X	K	.191 .136 .250	.190 .197 .113	.55	.619 .560 .581	.174 .093 .250	.190 .180 .113	.55	.599 .477 .581
824	Illinois	C	X	.778 .500	.014 .011	.93	.898 .640	.778 .500	.014 .011	.93	.997 .640
883	Iowa	C	X	.308 .300	.243 .196	.70	.820 .765	.269 .241	.257 .196	.70	.803 .713
1253	Oklahoma	X	A	.106 .117 .183	.125 .178 .176	.36	.423 .509 .610	.046 .102 .104	.103 .156 .147	.36	.269 .455 .448
1502	Colorado	C	Y	.055 .030 .068	.714 .737 .412	.15	.917 .767 .723	.044 .000 .069	.667 .737 .412	.15	.866 .767 .726
1572	Nebraska	C	P	.536 .867 .465	.091 .013 .145	.78	.828 .934 .840	.571 .867 .455	.078 .013 .169	.78	.839 .935 .871

Symbol definitions:

R = reduction coefficient

π = the probability a pixel is labeled wheat

π_{01} = the probability a pixel is classified wheat and labeled nonwheat

π_{10} = the probability a pixel is classified nonwheat and labeled wheat

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-5.— Continued.

(b) R-values for 45 starting dots

Segment		Crop 1	Crop 2	Cluster [Mean (μ) = 0.73579; variance (σ^2) = 0.032]				Classify [Mean (μ) = 0.72982; variance (σ^2) = 0.029]			
Number	Location			π_{10}	π_{01}	π	R	π_{10}	π_{01}	π	R
1075	Nebraska	C	A	0.050 .125 .100	0.578 .438 .378	0.37	0.788 .782 .693	0.068 .071 .067	0.691 .426 .386	0.37	0.827 .691 .646
1341	Kansas	C	X	.551 .367 .560	.232 .339 .345	.54	.947 .914 .991	.489 .347 .417	.218 .339 .309	.54	.907 .902 .924
1591	Nebraska	C	E	.106 .063 .099	.588 .654 .522	.22	.888 .875 .835	.094 .038 .075	.647 .654 .600	.22	.910 .828 .779
146	Kentucky	C	X	.192 .219	.121 .140	.63	.531 .592	.231 .250	.105 .105	.63	.549 .570
185	Minnesota	C	X	.065 .011 .079	.583 .800 .875	.16	.844 .879 .996	.067 .011 .057	.583 .800 .875	.16	.847 .879 .990
804	Iowa	C	X	.231 .200 .229	.127 .093 .054	.75	.623 .524 .461	.269 .233 .226	.128 .095 .055	.75	.661 .563 .460
812	Mississippi	X	K	.213 .182 .250	.172 .148 .132	.55	.622 .551 .609	.196 .140 .235	.138 .230 .132	.55	.554 .607 .592
824	Illinois	C	X	.778 .400	.014 .011	.93	.898 .544	.778 .300	.014 .023	.93	.898 .544
883	Iowa	C	X	.192 .020	.283 .214	.70	.765 .501	.269 .172	.232 .196	.70	.778 .646
1253	Oklahoma	X	A	.136 .067 .127	.100 .218 .176	.36	.444 .463 .524	.109 .069 .132	.100 .244 .121	.36	.396 .497 .465
1502	Colorado	C	Y	.077 .000 .080	.786 .684 .412	.15	.973 .718 .750	.079 .000 .080	.846 .684 .412	.15	.991 .718 .750
1572	Nebraska	C	P	.714 .833 .545	.065 .013 .181	.78	.917 .909 .930	.679 .828 .545	.065 .014 .157	.78	.895 .908 .912

Symbol definitions:

R = reduction coefficient

π = the probability a pixel is labeled wheat

π_{10} = the probability a pixel is classified wheat and labeled nonwheat

π_{01} = the probability a pixel is classified nonwheat and labeled wheat

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets

TABLE A-5.— Concluded.

(c) R-values for 30 starting dots

Segment		Crop 1	Crop 2	Cluster [Mean (μ) = 0.77509; variance (σ^2) = 0.029]				Classify [Mean (μ) = 0.74636; variance (σ^2) = 0.034]			
Number	Location			π_{10}	π_{01}	π	R	π_{10}	π_{01}	π	R
1075	Nebraska	C	A	0.050 .196 .183	0.609 .333 .356	0.37	0.873 .778 .783	0.050 .125 .133	0.682 .340 .302	0.37	0.868 .695 .669
1341	Kansas	C	X	.510 .306 .680	.161 .339 .273	.54	.875 .876 .997	.511 .245 .588	.143 .321 .236	.54	.859 .813 .993
1591	Nebraska	C	E	.118 .076 .086	.705 .654 .739	.22	.960 .893 .952	.096 .064 .088	.647 .640 .696	.22	.912 .866 .932
146	Kentucky	C	X	.385 .313	.155 .105	.63	.775 .637	.462 .219	.158 .105	.63	.839 .535
185	Minnesota	C	X	.075 .044 .101	.833 .800 1.000	.16	.986 .949 .982	.056 .035 .112	.833 .800 .938	.16	.976 .937 .996
804	Iowa	C	X	.269 .300 .258	.177 .067 .041	.75	.734 .569 .456	.308 .276 .233	.143 .068 .027	.75	.720 .547 .385
812	Mississippi	X	K	.170 .205 .308	.259 .115 .075	.55	.677 .531 .587	.109 .195 .294	.241 .197 .038	.55	.580 .633 .507
824	Illinois	C	X	.889 .500	.014 .045	.93	.970 .810	.889 .400	.014 .047	.93	.970 .746
883	Iowa	C	X	.346 .267	.157 .086	.70	.756 .564	.280 .241	.174 .070	.70	.720 .502
1253	Oklahoma	X	A	.121 .100 .127	.125 .289 .176	.36	.451 .605 .524	.108 .085 .132	.150 .289 .121	.36	.459 .578 .465
1502	Colorado	C	Y	.066 .000 .045	.857 .604 .353	.15	.989 .718 .605	.069 .000 .048	.917 .864 .313	.15	.9996 .718 .577
1572	Nebraska	C	P	.821 .867 .455	.000 .027 .205	.78	.855 .960 .902	.750 .862 .455	.000 .027 .181	.78	.794 .957 .882

Symbol definitions:

R = reduction coefficient

π = the probability a pixel is labeled wheat

π_{10} = the probability a pixel is classified wheat and labeled nonwheat

π_{01} = the probability a pixel is classified nonwheat and labeled wheat

Crop codes

A = alfalfa

C = corn

E = sorghum

K = cotton

P = pasture

X = soybeans

Y = sugar beets