LARS Information Note 072175

NASA CR-147389

(NASA-CR-147389) CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS) (Purdue Univ.) 59 p HC $4.50 CSCL 02C

N76-15528

Unclas
G3/43 07458

Crop Identification Technology Assessment for Remote Sensing (CITARS)

Results of CITARS Experiments Performed by LARS

Marvin E. Bauer
Tina K. Cary
Barbara J. Davis
Philip H. Swain

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1975

LARS Information Note 072175
T-1039/4

CROP IDENTIFICATION TECHNOLOGY ASSESSMENT

FOR REMOTE SENSING (CITARS)

RESULTS OF CITARS EXPERIMENTS PERFORMED BY LARS

Marvin E. Bauer, Tina K. Cary, Barbara J. Davis, and Philip H. Swain

Laboratory for Applications of Remote Sensing

Purdue University

West Lafayette, Indiana

This work was sponsored by the National Aeronautics and Space Administration under Contract No. NAS9-14016.

ABSTRACT

This report summarizes the results of classifications and experiments performed by LARS/Purdue University for the Crop Identification Technology Assessment for Remote Sensing (CITARS) project. Background information describing the experimental design and procedures may be found in reference 4 or 11.

Fifteen data sets were classified using two analysis procedures. One procedure used class weights while the other assumed equal probabilities of occurrence for all classes. In addition, 20 data sets were classified using training statistics from another segment or date. The results of both the local and non-local classifications in terms of classification and proportion estimation are presented in Part 1.

Part 2 of the report describes several additional experiments performed to provide additional understanding of the CITARS results. These experiments investigated alternative analysis procedures, training set selection and size, effects of multitemporal registration, the spectral discriminability of corn, soybeans, and "other," and analysis of aircraft multispectral data.

Part 3 of the report summarizes the results and presents our overall conclusions.

RESULTS OF CITARS EXPERIMENTS PERFORMED BY LARS

Table of Contents

Abstract i

Table of Contents ii

Part 1. CITARS Analyses 1

I. Introduction 1

II. Data Analysis Procedures 1

III. Classification Results 4

Part 2. Additional Investigations 15

I. Introduction 15

II. Factors Affecting Classification Performance . . 15

III. Statistical Analysis of Results 20

IV. Investigation of Alternative Analysis Procedures 21

V. Comparison of Training Sets 30

VI. Effect of Multitemporal Registration on Classification Performance 37

VII. Spectral Discriminability of Corn, Soybeans, and "Other" 40

VIII. Analysis of Aircraft Multispectral Scanner Data . 43

Part 3. Summary and Conclusions 50

References 54

Part 1. CITARS Analyses

I. Introduction

This section briefly describes the two analysis procedures followed by LARS in classifying the ERTS data for CITARS and presents the results of the classifications as measured by classification accuracy and proportion estimation.

II. Data Analysis Procedures

The CITARS data analysis procedures used by LARS were designed to be automated (capable of being programmed) and repeatable with the intent of minimizing the amount of subjective decision making on the part of the analysts. Subsequent tests have shown that different analysts following the procedures obtained identical results. This has the advantage of allowing comparison of results obtained by different analysts which is an important consideration in evaluating different data collection or data processing technologies as in CITARS. It also has the potential for increasing the speed and volume of data analysis relative to procedures involving the analyst to a greater degree. On the other hand, some performance may be sacrificed when the analyst is not permitted to tailor the analysis procedure to the particular problem and data set.

The analysis techniques used by LARS utilized the LARSYS Version 3 multispectral data analysis system. Its theoretical basis and details of the algorithm implementation are described by Swain [1] and Phillips [2]. The analysis procedure was described in detail by Davis and Swain [3] and in Volume I of the CITARS final report [4]. The procedures are designed to provide repeatable results, i.e., variation due to analysts is minimized. Briefly, the analysis procedures consist of:

A. Class Definition and Refinement

Four major classes, corn, soybeans, wheat (for selected missions) and all "other" ground covers were defined. These major classes were divided into subclasses where spectral variability within a class was so great as to result in multimodal probability distributions for that class. Clustering quarter-section field centers was used to isolate the subclasses. For clustering all four ERTS bands are used. A systematic method which minimized the total number of subclasses while avoiding multimodal subclass distributions was used for interpreting information on the separability of subclasses [Davis and Swain (3)].

B. Classification

Each data set was analyzed using two versions of the maximum likelihood classification algorithm. Gaussian probability density functions were assumed for both procedures. The first classification method, LARS/SP1, was the maximum likelihood classification rule assuming equal prior probabilities for all classes and subclasses. This is the rule which has been in common usage for remote sensing data analysis for some time.

The second method, LARS/SP2, used "class weights" proportional to the class prior probabilities. This approach is more nearly optimal given that the Bayesian error criterion (minimum expected error) is preferred. Class weights may be based on any reasonably reliable source of information. In CITARS the class weights were computed from county acreage estimates made by the USDA the previous year. Class weights were divided among the subclasses in proportion to the number of points in each subclass as determined by the clustering procedure.

C. Results Display and Tabulation

The results of the classification were displayed using a discriminant threshold of 0.1%. This low threshold eliminated only those data points very much different from the major class characterizations. Thresholded points were counted in the "other" category. A computer program was used to generate results tabulations, in both printed and punchcard form, for training fields, test fields, and test sections.

III. Classification Results

The classification results obtained by LARS are summarized in Tables 1-8. Classification accuracy (average and overall) and class bias and root mean square errors of proportion estimates are presented. Tables 1-4 present the results of the local recognition and Tables 5-8 show the non-local classification results. The statistical analyses of the classification results, along with those of EOD and ERIM, are presented and discussed in Volume IX and X of the CITARS final report and will not be repeated here, except for the comparison of the two analysis procedures used by LARS.

The LARS/SP1 procedure used a maximum likelihood Gaussian classifier which assumed that the frequency of occurrence of each class was the same for all classes. The LARS/SP2 procedure was identical to the SP1 procedure except unequal class weights (i.e., prior probability information) was used. The use of the "correct" values for the frequency of occurrence of each class will theoretically maximize the overall performance; that is, the proportion of the test pixels which are correctly classified. LARS/SP2 was designed to attempt to maximize overall performance.

Statistical comparison of the overall results of the equal (SP1) and unequal (SP2) prior probability procedures indicated that the use of historical data as a basis for prior probabilities did not affect proportion estimation or classification accuracy significantly for either local or non-local recognition on the basis of average performance. However, in interpreting this result it must be remembered that LARS/SP2 was an attempt to maximize overall performance rather than average performance. However, in the case of CITARS the two procedures were not significantly different as measured by either overall or average classification accuracy. Therefore, the quality of the prior probabilities used should be examined.

The unequal prior probabilities were based on the 1972 crop acreage estimates made by the USDA, Statistical Reporting Service for each county. While it was expected that the probabilities derived from these figures would not be the true probabilities for 1973, it was expected that there would no be major change.

The USDA figures were available only on a county basis, while CITARS examined only a 5 x 20 mile segment of each county. Furthermore, performance was examined on only 20 of the 100 sections in the segment. Since the crop proportions varied significantly from section to section, the crop proportions based on county estimates may not apply. Table 9 presents the actual proportions in the 20 sections of each segment and the class weights used in LARS/SP2. Examination of the data in Table 9 shows that there was considerable difference between the two. A final observation is that the classifier may not be very sensitive to the differences between equal and non-equal weights which were actually present in the CITARS data.

Our conclusion is that while prior probability information in the form of class weights should be used when available (as such usage has a sound theoretical basis), it may not in practice give much, if any, improvement in performance. Further tests to determine the sensitivity of the classifier to class weights are recommended.

TABLE 1. BIAS AND ROOT MEAN SQUARE ERROR OF PROPORTION ESTIMATES USING LARS/SP1 FOR LOCAL RECOGNITION.

SEGMENT (PASS)	CLASS BIAS: CORN	CLASS BIAS: SOYBEAN	CLASS BIAS: 'OTHER'	ROOT MEAN SQUARE ERROR: OVERALL SEGMENT ESTIMATES	ROOT MEAN SQUARE ERROR: AVERAGE OVER SECTIONS
HU(6)	0.157	0.302	-.459	0.330	0.292
HU(13)	0.061	0.121	-.182	0.131	0.157
SH(12)	0.014	-.038	0.024	0.027	0.129
SH(13)	0.206	-.057	-.149	0.151	0.207
WH(10)	-.058	0.091	-.033	0.065	0.109
WH(11)	-.046	0.080	-.034	0.057	0.150
LI(5)	0.004	-.005	0.001	0.004	0.112
LI(7)	-.013	0.017	-.004	0.013	0.097
FA(4)	0.127	-.152	0.025	0.115	0.180
FA(5)	0.185	-.020	-.165	0.144	0.192
FA(6)	0.179	0.017	-.196	0.154	0.178
FA(9)	0.076	0.145	-.220	0.158	0.136
LE(5)	0.014	0.015	-.029	0.020	0.111
LE(6)	0.011	-.034	0.023	0.025	0.110
LE(8)	0.029	0.018	-.047	0.034	0.118
MEANS OVER SEGMENTS	0.063	0.033	-.096	0.095	0.152

BIAS = ESTIMATED - PHOTOINTERPRETED PROPORTION

TABLE 2. BIAS AND ROOT MEAN SQUARE ERROR OF PROPORTION ESTIMATES USING LARS/SP2 FOR LOCAL RECOGNITION.

	CLASS BIAS			ROOT MEAN SQUARE ERROR	
SEGMENT (PASS)	CORN	SOYBEAN	'OTHER'	OVERALL SEGMENT ESTIMATES	AVERAGE OVER SECTIONS
HU(6)	0.227	0.229	-.456	0.322	0.281
HU(13)	0.177	0.006	-.183	0.147	0.182
SH(12)	0.125	-.069	-.056	0.089	0.163
SH(13)	0.044	0.051	-.095	0.067	0.148
WH(10)	-.041	-.002	0.042	0.034	0.094
WH(11)	-.062	-.072	0.134	0.095	0.146
LI(5)	0.014	0.016	-.031	0.022	0.131
LI(7)	0.097	-.098	0.001	0.079	0.150
FA(4)	0.078	0.014	-.091	0.070	0.139
FA(5)	0.086	0.140	-.226	0.162	0.175
FA(6)	0.180	-.007	-.173	0.144	0.172
FA(9)	0.092	0.140	-.232	0.165	0.141
LE(5)	0.075	0.219	-.294	0.216	0.203
LE(6)	0.069	0.117	-.187	0.133	0.142
LE(8)	0.007	0.125	-.132	0.105	0.147
MEANS OVER SEGMENTS	0.078	0.054	-.132	0.123	0.161

BIAS = ESTIMATED - PHOTOINTERPRETED PROPORTION

TABLE 3. CLASSIFICATION ACCURACY USING LARS/SP1 FOR LOCAL RECOGNITION.

SEGMENT (PASS)	CLASSIFICATION ACCURACY				
	CORN	SOYBEAN	'OTHER'	AVERAGE	OVERALL
HU(6)	0.599	0.910	0.313	0.607	0.448
HU(13)	0.478	0.471	0.505	0.484	0.496
SH(12)	0.498	0.482	0.527	0.502	0.498
SH(13)	0.640	0.266	0.245	0.384	0.485
WH(10)	0.748	0.841	0.639	0.742	0.751
WH(11)	0.545	0.810	0.471	0.609	0.612
LI(5)	0.618	0.632	0.512	0.588	0.599
LI(7)	0.691	0.633	0.777	0.700	0.673
FA(4)	0.745	0.235	0.651	0.544	0.531
FA(5)	0.864	0.425	0.325	0.538	0.511
FA(6)	0.968	0.458	0.433	0.620	0.592
FA(9)	0.790	0.950	0.652	0.797	0.796
LE(5)	0.570	0.634	0.413	0.539	0.576
LE(6)	0.641	0.573	0.462	0.559	0.583
LE(8)	0.568	0.536	0.549	0.551	0.550
MEANS OVER SEGMENTS	0.664	0.590	0.498	0.584	0.580

ACCURACY = PROPORTION OF CORRECTLY CLASSIFIED PIXELS IN A CLASS

AVERAGE = AVERAGE CLASS ACCURACY

OVERALL = PROPORTION OF CORRECTLY CLASSIFIED PIXELS OF ALL PIXELS CLASSIFIED

TABLE 4. CLASSIFICATION ACCURACY USING LARS/SP2 FOR LOCAL RECOGNITION.

SEGMENT (PASS)	CLASSIFICATION ACCURACY				
	CORN	SOYBEAN	'OTHER'	AVERAGE	OVERALL
HU(6)	0.681	0.889	0.317	0.629	0.458
HU(13)	0.669	0.249	0.513	0.477	0.491
SH(12)	0.623	0.441	0.463	0.509	0.551
SH(13)	0.528	0.367	0.340	0.412	0.459
WH(10)	0.721	0.808	0.773	0.767	0.764
WH(11)	0.489	0.659	0.618	0.589	0.579
LI(5)	0.582	0.674	0.510	0.589	0.607
LI(7)	0.803	0.552	0.763	0.706	0.663
FA(4)	0.513	0.444	0.549	0.502	0.502
FA(5)	0.850	0.567	0.292	0.570	0.546
FA(6)	0.958	0.489	0.535	0.660	0.638
FA(9)	0.762	0.944	0.615	0.774	0.772
LE(5)	0.686	0.825	0.141	0.551	0.669
LE(6)	0.633	0.716	0.255	0.535	0.615
LE(8)	0.555	0.641	0.435	0.543	0.579
MEANS OVER SEGMENTS	0.670	0.618	0.475	0.588	0.593

ACCURACY = PROPORTION OF CORRECTLY CLASSIFIED PIXELS IN A CLASS

AVERAGE = AVERAGE CLASS ACCURACY

OVERALL = PROPORTION OF CORRECTLY CLASSIFIED PIXELS OF ALL PIXELS CLASSIFIED

TABLE 5. BIAS AND ROOT MEAN SQUARE ERROR OF PROPORTION ESTIMATES USING LARS/SP1 FOR NONLOCAL RECOGNITION.

	CLASS BIAS			ROOT MEAN SQUARE ERROR	
TRAINING--CLASSIFIED	CORN	SOYBEAN	'OTHER'	OVERALL SEGMENT ESTIMATES	AVERAGE OVER SECTIONS
FA(5)--FA(6)	0.129	-.031	-.098	0.095	0.159
FA(6)--FA(5)	0.189	0.051	-.240	0.179	0.186
LE(5)--LE(6)	-.007	0.094	-.087	0.074	0.128
LE(6)--LE(5)	-.113	0.002	0.111	0.092	0.149
HU(6)--LI(5)	0.185	0.030	-.215	0.164	0.268
HU(6)--LE(6)	-.117	0.298	-.182	0.213	0.260
LE(6)--LI(5)	-.267	-.070	0.337	0.252	0.268
LE(6)--HU(6)	-.126	0.108	0.018	0.097	0.204
LI(7)--LE(8)	0.093	0.167	-.259	0.186	0.181
LE(8)--LI(7)	-.037	0.005	0.032	0.029	0.151
LI(5)--FA(5)	-.075	-.240	0.315	0.233	0.273
FA(5)--LI(5)	-.225	0.053	0.173	0.167	0.257
WH(11)--SH(12)	0.017	-.105	0.088	0.080	0.143
SH(12)--WH(11)	-.036	-.035	0.071	0.050	0.122
SH(13)--HU(13)	0.306	-.038	-.269	0.236	0.264
HU(13)--SH(13)	0.068	0.103	-.171	0.121	0.146
FA(6)--HU(6)	0.119	0.140	-.259	0.183	0.254
HU(6)--FA(6)	0.174	0.241	-.415	0.294	0.261
WH(10)--FA(9)	-.142	-.116	0.257	0.182	0.236
FA(9)--WH(10)	-.221	-.073	0.294	0.216	0.195
MEANS OVER RECOGNITIONS	-.004	0.029	-.025	0.157	0.205

BIAS = ESTIMATED - PHOTOINTERPRETED PROPORTION

TABLE 6. BIAS AND ROOT MEAN SQUARE ERROR OF PROPORTION ESTIMATES USING LARS/SP2 FOR NONLOCAL RECOGNITION.

	CLASS BIAS			ROOT MEAN SQUARE ERROR	
TRAINING--CLASSIFIED	CORN	SOYBEAN	'OTHER'	OVERALL SEGMENT ESTIMATES	AVERAGE OVER SECTIONS
FA(5)--FA(6)	0.066	0.084	-.149	0.106	0.136
FA(6)--FA(5)	0.177	0.055	-.233	0.172	0.177
LE(5)--LE(6)	-.043	0.318	-.275	0.244	0.254
LE(6)--LE(5)	-.092	0.114	-.021	0.086	0.168
HU(6)--LI(5)	0.288	-.074	-.213	0.211	0.309
HU(6)--LE(6)	0.037	0.129	-.166	0.123	0.155
LE(6)--LI(5)	-.277	0.032	0.245	0.214	0.292
LE(6)--HU(6)	-.141	0.161	-.020	0.124	0.228
LI(7)--LE(8)	0.295	-.091	-.205	0.214	0.243
LE(8)--LI(7)	-.159	0.232	-.073	0.168	0.239
LI(5)--FA(5)	-.112	-.265	0.377	0.274	0.282
FA(5)--LI(5)	-.135	0.141	-.006	0.113	0.245
WH(11)--SH(12)	-.025	-.200	0.224	0.174	0.189
SH(12)--WH(11)	0.014	-.042	0.028	0.031	0.117
SH(13)--HU(13)	0.071	0.122	-.193	0.138	0.185
HU(13)--SH(13)	0.278	-.095	-.183	0.200	0.234
FA(6)--HU(6)	0.217	0.076	-.293	0.215	0.267
HU(6)--FA(6)	0.197	0.209	-.405	0.287	0.253
WH(10)--FA(9)	-.141	-.205	0.346	0.246	0.256
FA(9)--WH(10)	-.190	-.097	0.287	0.207	0.188
MEANS OVER RECOGNITIONS	0.016	0.030	-.046	0.177	0.221

BIAS = ESTIMATED - PHOTOINTERPRETED PROPORTION

TABLE 7. CLASSIFICATION ACCURACY USING LARS/SP1 FOR NONLOCAL RECOGNITION.

TRAINING--CLASSIFIED	CLASSIFICATION ACCURACY CORN	SOYBEAN	'OTHER'	AVERAGE	OVERALL
FA(5)--FA(6)	0.885	0.430	0.487	0.600	0.579
FA(6)--FA(5)	0.934	0.545	0.418	0.632	0.609
LE(5)--LE(6)	0.634	0.664	0.212	0.503	0.584
LE(6)--LE(5)	0.166	0.620	0.456	0.414	0.421
HU(6)--LI(5)	0.777	0.413	0.082	0.424	0.433
HU(6)--LE(6)	0.513	0.774	0.103	0.463	0.573
LE(6)--LI(5)	0.020	0.389	0.583	0.331	0.333
LE(6)--HU(6)	0.172	0.302	0.576	0.350	0.478
LI(7)--LE(8)	0.687	0.643	0.168	0.499	0.589
LE(8)--LI(7)	0.644	0.509	0.856	0.670	0.604
LI(5)--FA(5)	0.024	0.031	0.639	0.231	0.248
FA(5)--LI(5)	0.147	0.429	0.244	0.273	0.302
WH(11)--SH(12)	0.594	0.377	0.635	0.535	0.557
SH(12)--WH(11)	0.329	0.663	0.482	0.491	0.478
SH(13)--HU(13)	0.541	0.349	0.428	0.440	0.431
HU(13)--SH(13)	0.635	0.359	0.365	0.453	0.526
FA(6)--HU(6)	0.771	0.275	0.349	0.465	0.394
HU(6)--FA(6)	0.874	0.737	0.192	0.601	0.576
WH(10)--FA(9)	0.024	0.134	0.687	0.282	0.306
FA(9)--WH(10)	0.089	0.608	0.529	0.409	0.377
MEANS OVER RECOGNITIONS	0.473	0.463	0.425	0.453	0.470

ACCURACY = PROPORTION OF CORRECTLY CLASSIFIED PIXELS IN A CLASS

AVERAGE = AVERAGE CLASS ACCURACY

OVERALL = PROPORTION OF CORRECTLY CLASSIFIED PIXELS OF ALL PIXELS CLASSIFIED

TABLE 8. CLASSIFICATION ACCURACY USING LARS/SP2 FOR NONLOCAL RECOGNITION.

TRAINING-- CLASSIFIED	CLASSIFICATION ACCURACY CORN	SOYBEAN	'OTHER'	AVERAGE	OVERALL
FA(5)--FA(6)	0.892	0.626	0.452	0.656	0.637
FA(6)--FA(5)	0.920	0.603	0.494	0.672	0.653
LE(5)--LE(6)	0.657	0.855	0.065	0.526	0.660
LE(6)--LE(5)	0.181	0.751	0.293	0.408	0.464
HU(6)--LI(5)	0.835	0.303	0.082	0.407	0.399
HU(6)--LE(6)	0.598	0.651	0.109	0.453	0.549
LE(6)--LI(5)	0.018	0.449	0.305	0.257	0.291
LE(6)--HU(6)	0.166	0.376	0.533	0.358	0.458
LI(7)--LE(8)	0.870	0.419	0.304	0.531	0.575
LE(8)--LI(7)	0.440	0.745	0.823	0.669	0.659
LI(5)--FA(5)	0.014	0.014	0.803	0.277	0.300
FA(5)--LI(5)	0.311	0.536	0.128	0.325	0.370
WH(11)--SH(12)	0.525	0.154	0.719	0.466	0.483
SH(12)--WH(11)	0.391	0.687	0.417	0.498	0.494
SH(13)--HU(13)	0.280	0.630	0.545	0.485	0.523
HU(13)--SH(13)	0.824	0.114	0.335	0.424	0.580
FA(6)--HU(6)	0.802	0.386	0.369	0.519	0.430
HU(6)--FA(6)	0.888	0.732	0.233	0.617	0.592
WH(10)--FA(9)	0.031	0.081	0.799	0.304	0.331
FA(9)--WH(10)	0.105	0.585	0.514	0.401	0.372
MEANS OVER RECOGNITIONS	0.487	0.485	0.416	0.463	0.491

ACCURACY = PROPORTION OF CORRECTLY CLASSIFIED PIXELS IN A CLASS

AVERAGE = AVERAGE CLASS ACCURACY

OVERALL = PROPORTION OF CORRECTLY CLASSIFIED PIXELS OF ALL PIXELS CLASSIFIED

TABLE 9. WEIGHTS USED IN LARS/SP2 AND PHOTOINTERPRETED PROPORTIONS

	WEIGHTS USED IN LARS/SP2		
SEGMENT	CORN	SOYBEAN	'OTHER'
HUNTINGTON	23.72	23.92	52.36
SHELBY	34.69	22.16	43.15
WHITE	31.45	26.70	41.85
LIVINGSTON	38.59	37.75	23.66
FAYETTE	14.15	23.76	62.09
LEE	37.91	21.92	40.17

	PHOTOINTERPRETED PROPORTIONS		
SEGMENT	CORN	SOYBEAN	'OTHER'
HUNTINGTON	18.59	22.07	59.34
SHELBY	38.29	24.30	37.41
WHITE	36.28	31.08	32.64
LIVINGSTON	32.46	37.75	29.79
FAYETTE	19.43	29.34	51.22
LEE	33.22	28.70	38.07

Part 2. Additional Investigations

I. Introduction

Classification performances of 55 to 75 percent for test fields were obtained for CITARS; whereas, in previous ERTS investigations 75 to 95 percent correct crop identifications were reported [5,6,7,8]. Several additional special experiments were performed by LARS to determine the cause of unexpectedly low classification performance and to determine possible methods for improving the performance. Those experiments and results are discussed in this section.

II. Factors Affecting Classification Performance

Before describing the various experiments that were conducted, it may be useful to summarize possible factors affecting classification performances. They include: (1) the method of evaluation used, (2) the data analysis and classification procedures used, (3) availability of training data, (4) registration accuracy, (5) spectral characteristics of the scene, and (6) characteristics of the ERTS data.

A. Evaluation Method

While actual ground observations of crop identification were available for the fields used for training the classifiers, crop identifications for the test fields used to evaluate the classifications were determined by photointerpretation. Accurate identifications are, of course, required if a reliable measure of classification performance is to be obtained. Tests of the photointerpretation accuracy were conducted and results indicated that the crops in 95-98 percent of the fields were correctly identified [4]. Even this small percentage of errors, however, likely led to some reduction in the estimate of classification performance, perhaps on the order of two to three percent. However, no further work has been done by LARS to determine either the magnitude of photointerpretation errors or their effect on classification performance.

B. Data Analysis and Classification Procedures

A second factor which may have influenced classification performance was the data analysis procedures used to develop training statistics. While CITARS was intended to evaluate the adequacy of currently available technology; in fact, in response to the requirement for using repeatable procedures capable of being programmed, it resulted in the use of new and unproven analysis techniques [3]. Although these procedures were well-thought out and based on several years' experience in analyzing multispectral scanner, they were first used on the CITARS data. The primary question concerning the procedures used by LARS was whether using automatic and repeatable procedures which reduced the number of decisions made by the analyst may have adversely affected classification performance. To answer this question several alternative analysis procedures were evaluated with the CITARS data.

C. Availability of Training Data

The supervised classification methods used for CITARS require that fields with known crop identities be available for training. In the case of CITARS, fields from 20-quarter sections were potentially available for training purposes. This represented 20 percent of the total area for which the ground cover type was identified, but the amount of training data available is generally more critical than the percentage since a minimum number of points is required to adequately represent a class. As a rule of thumb the minimum is 10 times the number of features (channels) to be used in the classification or 40 for the CITARS data. While the original calculations of the number of points that would be available for training indicated that there would be adequate numbers of points, the number actually available was considerably smaller than anticipated.

The acres, number of fields, and average field size for the 20-quarter sections are shown in Table 10. It can be seen that with average field sizes of only 15 to 35 acres that the maximum number of pure pixels from an individual field will generally be small. This problem was compounded by: (1) the criteria for sampling pixels from field centers (at least one whole pixel between the field boundary and any sampled pixel), (2) clouds and cloud shadows, (3) bad data lines, and (4) segments only partially in the ERTS data. As a result of these conditions many training sets contained fewer data points than would have been desirable. And, in some instances classes had to be deleted because too few points were available to represent them. Therefore, an experiment to determine the effects of training set size and variability was performed.

Table 10. Summary of acres, hectares, number of fields, and average field size in the quarter-sections.

Segment		Corn	Soybeans	Wheat	Other
Huntington	Acres	831	618	63	986
	Hectares	336	250	25	399
	No. Fields	39	25	6	54
	Avg. Size				
	(Acres)	21.2	24.7	10.4	18.3
	(Hectares)	8.6	10.0	4.2	7.4
Shelby	Acres	1888	540	323	753
	Hectares	764	218	131	305
	No. Fields	71	24	15	61
	Avg. Size				
	(Acres)	26.5	22.5	21.5	12.3
	(Hectares)	10.8	9.1	8.7	5.0
White	Acres	1836	510	38	954
	Hectares	743	206	15	386
	No. Fields	42	13	2	41
	Avg. Size				
	(Acres)	43.7	39.2	19.0	23.3
	(Hectares)	17.7	15.9	7.6	9.4
Livingston	Acres	1239	1073	39	569
	Hectares	501	434	16	230
	No. Fields	33	27	2	33
	Avg. Size				
	(Acres)	37.5	39.7	19.5	17.2
	(Hectares)	15.2	16.1	7.9	7.0
Fayette	Acres	733	287	416	1358
	Hectares	297	116	168	550
	No. Fields	37	11	26	92
	Avg. Size				
	(Acres)	19.8	26.0	16.0	14.7
	(Hectares)	8.0	10.6	6.5	6.0
Lee	Acres	1498	813	36	620
	Hectares	606	329	15	251
	No. Fields	42	31	2	34
	Avg. Size				
	(Acres)	35.6	26.2	18.0	18.2
	(Hectares)	14.4	10.6	7.4	7.4

D. Registration Accuracy

To alleviate locating field and section coordinates in all data sets and to permit multitemporal data analysis, ERTS data from all available passes over each segment were spatially registered. For CITARS, the maximum allowable error in registration was 0.5 pixels as measured by the root mean squares of checkpoint residuals. With the guard row and column pixels of one whole pixel between actual field boundaries and selected sample pixels any error in spatial registration should not affect classification performance of field center pixels. Any registration error, however, could affect the proportion estimates obtained from classifications of entire sections. To determine if there was any significant effect of registration on classification performance, comparisons were made between registered and non-registered data for five segment-date combinations.

E. Spectral Characteristics of Crops

Accurate identification of crops by the methods used for CITARS requires that the crops and other cover types are separable based on their spectral characteristics. Classification performance, then, depends on the spectral separability of the cover types. An experiment was performed to evaluate the spectral discriminability of the cover types involved.

F. Characteristics of ERTS Data

Since accurate identification of crops by the methods used for CITARS requires that the cover types are separable based on their spectral characteristics, classification performance depends not only on the spectral separability of the cover types but also on the ability of the scanner to measure spectral differences. An experiment was performed with aircraft scanner data having greater number, width, and dynamic range of spectral bands than the ERTS bands were to determine whether classification performance would be increased.

III. Statistical Analysis of Results

The statistical analyses used for the principal CITARS results were applied to the results of the additional investigations. Briefly, analysis of variance was used to determine if any differences in results were statistically significant and the Newman-Keuls Multiple Range Test was applied to determine which treatments were different.

For the analysis of test field classification performance results, the non-diagonal elements of the classification performance matrix were used. Since the elements of the estimated performance matrix are distributed binomially, the variance of the sum of the non-diagonal elements will be less dependent on the mean if the individual elements of the performance matrix are transformed [9]. A summation of transformed values was used as the variable for analysis of variance. The value of the variable was found by:

$$\sum_{\substack{i,j=1 \\ i \neq j}}^{3} \frac{2}{\pi} \arcsin \left(e_{ij}^{1/2}\right)$$

where e_{ij} is an element of the classification performance matrix. (Summation is from 1 to 3 for the three cover types.)

To evaluate the proportion estimates for the sections the classification results proportions were compared to the proportions as determined by photointerpretation. The accuracy of the proportion estimation is measured by

$$\sum_{i=1}^{k} (\hat{P}_i - P_i)^2$$

where k is the number of classes, $\hat{P}_i$ is the computer-estimated proportion of class i, and P_i is the proportion of class i as determined by photointerpretation. In order to obtain more homogeneous variances, the variable was transformed [9]. The variable used for the analysis of variance was

$$\ln[100 \sum_{i=1}^{k} (\hat{P}_i - P_i)^2 + .02]$$

A detailed discussion of the statistical analysis of results can be found in Volume IX of this report [4].

IV. Investigation of Alternative Analysis Procedures

A. Introduction

To accomplish the objectives of the CITARS experiment, the ADP procedures used to obtain classification results had to be well-defined (capable of being automated) and repeatable. Procedures meeting these criteria would not be biased by analyst subjectivity. While this approach has certain advantages, it has the disadvantage that the analyst(s) could not tailor the procedure to the particular problem and data set. The objective of this study was to determine if classification performance was adversely affected by the automated and repeatable data analysis procedure used for CITARS.

To answer this question, several variations in the procedure were applied to the same data set. Data for Lee County, Illinois collected August 5, 1973 (run 73120202) were used. This particular data set was chosen because the original classification accuracy (60 percent) indicated that there was potential for improvement.

B. Description of Analysis Procedures

Seven variations of the analysis procedure were applied. They are described in the following paragraphs and are summarized in Table 11.

Procedure 1. The initial procedure is the one which was utilized for CITARS and consists of the following steps: Three cover type classes were defined: corn, soybeans, and all "other" ground covers. When the major cover type classes were multimodal, clustering was used to divide the classes into subclasses. The clustering algorithm used requires that the analyst specify the number of clusters to be found. The following rules were used to determine the number of clusters to request: for corn, request 5, for soybeans 5, agricultural "other" 10, and nonagriculturla "other" 3 for each identifiable subclass. There are two exceptions: determine the maximum number of clusters to request for each major class by dividing the number of data points available for clustering by 40; for the agricultural "other" or the non-agricultural "other," the minimum number of clusters is the number of identifiable subclasses, even if this minimum is greater than the maximum found in the previous exception.

All four channels were used for clustering, and a statistics deck was punched from each cluster analysis, to be merged later. Any cluster group having fewer than 25 points total was deleted from further consideration. After the classes were refined and the statistics decks merged into one, the data was classified using a Gaussian maximum-likelihood classification rule. Equal prior probabilities for all subclasses were assumed.

The classification results were displayed in the form of maps and tables. Performances were tabulated for training fields, test fields, and test sections. Pilot and test fields were combined for this investigation.

In the remainder of this investigation, the procedures for class definition and refinement were varied. The same classification algorithm was used throughout and results were always tabulated for the same fields and sections.

Procedure 2. The second test was verification of the repeatability of the analysis. Given the original training fields and the number of clusters to request, the analyst carried out the specified procedure. The results, as expected, did duplicate the results obtained the first time. The overall classification performance for test fields was 55.2 percent.

Procedure 3. For the next procedure, the only variation from the defined procedure was in the number of clusters requested. The guideline for the maximum number of clusters to request is to divide the number of data points for the class by 40. The quotients were 3.3 for corn, 2.75 for soybeans, and 9.9 for "other." Originally, three corn, two soybean, and nine "other" clusters were requested. The same quotients could have been interpreted to request three corn, three soybean, and 10 "other" clusters.

When these clusters were requested and the defined procedure was followed, overall performance was 55.3 percent.

Procedure 4. The next factor investigated was number of training points. The number of training points originally provided was 131 corn, 110 soybeans, and 396 "other." The analyst went back to an aerial photograph, an overlay defining fields, and field identification information to select more training points. The original criteria of using only points inside a buffer zone of one line or column was relaxed. The total number of training points used was 416 corn, 350 soybean, and 788 "other." The defined procedure was followed for the classification using these points for training. Overall performance was 56.4 percent.

Procedure 5. The next procedure varied from the defined procedure in several ways. One half of the original corn training fields, one half of the original corn pilot fields, one half of the original corn test fields were randomly selected for training; also, one half of the original soybean training, test, and pilot fields were similarly selected. All of the additional training points selected in the previous procedure were also included. For clustering, five corn clusters and five soybean clusters were requested as before, but the "other" was handled differently.

For clustering the class of "other," the analyst first divided the training points into the following categories: woods; urban, freeway, and other bare; pasture, small grain, and woods-pasture; and water. Each of these subclasses of "other" was clustered separately. The number of clusters to request was determined by dividing the number of data points by 40 (and rounding). Then the statistics from these six clustering jobs were merged into a single statistics deck.

The analyst next ran the SEPARABILITY processor which calculates the statistical distance known as transformed divergence for all pairs of classes. The analyst then looked for class pairs having a transformed divergence less than 1000 (the maximum possible value is 2000). There were three such class pairs. The class pairs were (1) corn-2/woods-1, (2) corn-5/woods-2, and (3) soybean-4/small grain-2, where corn-2/woods-1 means subclass 2 of class corn and subclass 1 of class woods. Since in each case the classes were from two different cover types, one of the classes was deleted from each pair. The criterion for deletion of subclasses was: delete the subclass of the cover type having more subclasses. That is, corn had five subclasses, and woods two, so for both corn-woods class pairs, the corn class was deleted. Soybeans had five subclasses and small grain two, so for that class pair soybeans was deleted. This left three subclasses of corn, four soybean, two small grain, three woods, three urban and bare, and one water class, and none of these class pairs had a transformed divergence less than 1000. The area was then classified following the original CITARS procedures. Overall test field performance was 57.1 percent.

Procedure 6. The next procedure differed rather drastically from the standard CITARS procedure. The quarter sections were used as the basis for training. Due to computer core size limitations, not all quarter sections could be clustered at once, so the quarter sections were arbitrarily divided into three groups.

Again the problem of number of clusters to request had to be solved. The problem was approached in the following way: for each group of quarter sections, clustering was run several times with various numbers of clusters requested, SEPARABILITY was run on the statistics of those clusters, and the set of clusters having the greatest pairwise minimum distance was chosen.

For the first group of quarter sections, 16 clusters were requested; for the second, 12 clusters; and for the third, 16 clusters. Statistics were calculated for each cluster and punched on cards for further use.

The map output from CLUSTER was used in conjunction with aerial photography, an overlay of field boundaries, and field identification information to identify the cover type associated with each cluster. The statistics from all the clusters were put into the SEPARABILITY processor, and again the transformed divergence measure was used as the criterion for pooling and deleting subclasses. The data was then classified in the normal way. Overall performance was 61.4 percent.

Procedure 7. Procedure 6 had achieved the best overall performance, and the best performance for the class corn, but procedure 5 had the best performance for soybeans, and the best training field performance for "other." For procedure 7 training classes from the procedure in which they had performed best were combined into a new training statistics deck. Again SEPARABILITY was run and transformed divergence used as a basis for pooling or deleting subclasses. Overall classification performance for this procedure was 47.4 percent.

C. Results and Discussion

The classification results are summarized in Tables 12 and 13. None of the five alternative analysis procedures resulted in any significant improvement in classification performance as measured by proportion estimates for sections. The sixth procedure which involved clustering the quarter-sections gave improved performance for corn and "other" test fields, but at the expense of soybean performance. Further investigation of that result, however, shows that too many pixels in the sections were classified as corn, too few as soybeans, and too few as "other." The seventh procedure gave improved performance for "other" but low performances for both corn and soybeans.

The conclusions drawn from these results are that (1) the CITARS procedures used by LARS produce repeatable results and (2) none of the alternative procedures tried resulted in any improvement in classification performance. While these results and conclusions are based on a relatively limited sample, it is probably safe to conclude that little if any of the generally low classification performances obtained in CITARS can be attributed to the data analysis procedures used. In the context of LACIE which will involve many analysts these results indicate that it is possible to use repeatable and relatively automatic analysis procedures without sacrificing classification performance.

Table 11. Summary descriptions of analysis procedures.

Procedure	Description
1.	Original analysis following defined procedure.
2.	Verification of repeatability.
3.	Defined procedure, requesting different number of clusters for soybeans and other.
4.	Additional training points selected, then defined procedure followed.
5.	Extended set of training points, classes of other separated before clustering, transformed divergence calculated for class pairs, one class of pair deleted for distances below threshold (1000).
6.	Quarter sections clustered, cluster maps used to identify clusters; transformed divergence used as criterion for pooling or deleting subclasses.
7.	Corn training from procedure 6 and soybeans and other training from procedure 5 used for training, transformed divergence criterion used for pooling or deleting subclasses.

Table 12. Summary of classification performances (% correct classification of test fields) for seven analysis procedures.

Procedure	Corn	Soybeans	Other	Overall
1	57.1	53.6	55.4	55.2
2	57.1	53.6	55.4	55.2
3	55.8	53.1	60.8	55.3
4	68.8	50.8	42.5	56.4
5	47.9	63.6	60.2	57.1
6	87.6	37.1	69.9	61.4
7	37.2	42.6	88.2	47.4

Table 13. Average proportions of corn, soybeans, and "other" present in 20 test sections as determined from seven analyses.

Procedure	Corn	Soybeans	Other
1	36.1	30.5	33.4
2	36.1	30.5	33.4
3	36.0	28.5	35.5
4	46.6	24.2	29.2
5	25.2	31.2	43.6
6	48.0	12.4	39.7
7	21.8	15.7	62.5
Photointerpreted Proportion	31.3	21.8	46.9

V. Comparison of Training Sets

A. Introduction

One of the objectives of CITARS was an examination of the effect of varying the training set selection on classification performance. To meet this objective, two training sets, each containing 10 quarter-sections, were to have been available for comparison. However, as training fields were selected, it became obvious that 10 quarter-sections would not provide an adequate training sample, and the two sets were combined to provide the 20 quarter-section training set.

In this experiment, two training sets were used to train the classifier - the ten "pilot" sections the the ten "test" sections. The classification performance for each of these training sets was compared to the classification performance of the 20 quarter-section training set.

B. Procedures

The ten data sets described in Table 14 were selected for this experiment. They were first classified using the 10 "pilot" sections as the basis for training the classifier, and then classified again using the 10 "test" sections as the basis for training. The analysis procedures were the same as for other classifications of ERTS data performed by LARS (i.e. LARS/SP1 and LARS/SP2). The classifications based on "pilot" sections were compared to the regular CITARS classifications (based on "training" quarter-sections) by examining the overall classification performance of field center pixels from the 10 "test" sections. Similarly, the classifications based on the "test" sections were

Table 14. Summary of data analyzed to determine effect of varying training set on classification performance

Segment-Period-Pass	Date	ERTS Scene ID
Huntington-III	July 15	1357-15590
Livingston-III	July 16	1358-16045
Fayette-III-2	July 17	1359-16105
Lee-III-2	July 18	1360-16155
Lee-IV	August 5	1378-16153
White-V	August 21	1394-16042
Fayette-V	August 21	1394-16044
Shelby-VI	September 7	1411-15581
Huntington-VII	September 24	1428-15520
Shelby-VII	September 24	1428-15523

compared to the regular CITARS classifications by examining the overall classification performance of field center pixels from the 10 "pilot" sections. The comparisons were made in this way to avoid biasing classification performance by testing on samples which were used in training the classifier. The variability of proportion estimation accuracy was evaluated using analysis of variance.

C. Results and Discussion

Overall performances obtained from the CITARS classifications based on the "training" quarter sections and overall performances obtained from the classifications based on the ten "pilot" sections are shown in Table 15. For seven of ten cases, the "pilot" classifications had higher overall test performance (column 5) than the CITARS classifications (column 3). In only four instances (i.e. HU-III, LI-III, WH-V, and FA-V) could "pilot" overall test performance (column 5) be considered reasonably high (greater than 75%). Two of these instances (HU-III and FA-V) had reasonably high CITARS overall test performance (column 3).

Table 15 also shows the overall performances obtained from the classifications based on the ten "test" sections. The "test" overall test performance (column 7) was less than the CITARS overall test performance (column 2) were above 75%.

The same random sampling scheme was used to choose the "pilot" and the "test" sections. Thus both sets of sections should represent the same population. However, comparisons between the second and third columns of Table 15 suggest that

Table 15. Comparison of different samples of training and test fields.

Segment-Period	Source of Training Data						
	Training Fields			Pilot Fields		Test Fields	
	Classification Performance (%)			Classification Performance (%)		Classification Performance (%)	
	Training Fields	Pilot Fields	Test Fields	Pilot Fields	Test Fields	Test Fields	Pilot Fields
HU-III	92.3	28.4	80.1	89.7	78.7	87.1	72.7
LI-III	78.1	58.8	60.6	81.4	76.1	75.2	71.7
FA-III-2	77.8	52.9	63.7	86.8	69.7	89.8	73.7
LE-III-2	80.2	53.2	61.7	58.8	64.3	79.7	54.8
LE-IV	75.5	62.4	49.9	71.0	57.0	75.9	59.2
WH-V	87.9	75.8	74.3	88.3	80.7	84.1	67.0
FA-V	90.5	79.7	79.5	84.4	86.3	90.5	85.2
SH-VI	77.1	48.0	51.8	76.4	49.2	76.9	58.0
HU-VII	81.2	40.9	68.2	87.1	66.8	78.2	60.5
SH-VII	73.5	52.9	43.8	64.7	51.6	71.6	61.3
Mean	81.4	55.3	63.4	78.9	68.0	80.9	66.4

this conclusion is not always true. In 4 cases (HU-III, FA-III-2, LE-IV, HU-VII), the entries in column 2 and column 3 of Table 15 show differences in performance greater than 10%. In two additional cases (LE-III-2 and SH-VII), the differences are greater than 8%. These differences suggest that the "pilot" fields and the "test" fields were not always representative samples of the same population.

The "pilot" fields, and also the "test" fields, were obtained from ten sections. Since ten sections have twice the area of twenty quarter-sections, one could expect the "pilot" fields (or the "test" fields) to contain twice as many pixels as the "train" fields. However, this was not the case.

Table 16 gives the number of data points in each training set of the ten data sets used in this investigation. In only four cases, HU-III "pilot", LI-III "test", SH-VI "pilot", and HU-VII "pilot" were the number of points more than twice the number of points in the regular CITARS training set. Thus, the effect of training set size can not be fully evaluated.

It is interesting to examine these four cases (HU-III, LI-III, SH-VI, and HU-VII) Table 15 in light of the number of points in each training set. For example, though the "pilot" training set of HU-III was more than twice the size of the "train" training set, the "pilot" overall test performance was 78.7%, 1.4% less than the CITARS overall test performance of 80.1% (column 3). The "test" training set of HU-III was less than 50 points bigger than the "train" training set, but the

Table 16. Number of Points in Each Training Set.

Segment-Period-Pass	Source of Training Data		
	Training	Pilot	Test
HU-III	325	799	371
LI-III	544	738	1018
FA-III-2	460	418	600
LE-III-2	637	500	729
LE-IV	637	500	725
WH-V	812	871	673
FA-V	454	418	600
SH-VI	271	550	490
HU-VII	325	799	371
SH-VII	291	569	525

"test" overall test performance was 72.7%, a gain of 44.3% over the CITARS overall test performance of 28.4% (column 2). These results suggest that the representativeness and adequacy of the training set is not a function of the training set size along.

The proportion estimation accuracy was examined through analyses of variance. The "pilot" and the "train" training sets were not significantly different; however, the "test" and the "train" training sets were significantly different. Since both the "test" and the "pilot" training sets were chosen in the same way, the results of the analyses of variance suggest that the choice of training set can significantly affect proportion estimation accuracy.

VI. Effect of Multitemporal Registration on Classification Performance

A. Introduction

To enable classifications of multitemporal ERTS data and to alleviate having to locate section and field coordinates in each segment-date combination of data, the satellite passes over each segment were registered as part of the data preparation phase [4, Volume 5, "ERTS-1 Data Preparation."] This experiment was performed to determine if registration had any effect on classification performance and if so, the magnitude of the effect.

B. Procedures

The experiment consisted of a comparison of crop classification performances obtained with registered and non-registered forms of ERTS data. Both forms of the data were geometrically corrected. Five segment-date combinations of data were selected for analysis. The coordinates of sections and fields used for the registered data were the same as used in the regular CITARS data classifications. The coordinates from approximately the same fields were located in the non-registered data by manually overlaying the photo overlays onto the ERTS imagery. A one-to-one correspondence of fields in both data sets was not used because to do so would have eliminated several fields which were needed for training. However, about 80 percent of the fields were common to both data sets. The same procedure for selecting pixels from fields, i.e. one "guard" pixel between field boundary and any selected pixel, was followed in both cases.

The same classification procedures, i.e., LARS/SP1 and SP2, were applied to both the registered and non-registered data sets for all five segment-date combinations. Also, the non-registered data was classified with statistics from the registered data, and the registered data was classified with statistics from the non-registered data. Test and pilot fields were combined into a single test set, and test and pilot sections were combined. Recognition performances for fields and proportion estimates for sections were tabulated, and an analysis of variance was performed to determine if any significant difference existed between the registered and non-registered data.

C. Results

Overall classification performances for test and pilot fields combined are shown in Table 17 for the five segment-date combinations. The results of the analysis of variance (a conservative test) indicated that there was no significant difference between the performance of registered and non-registered data. However, inspection of overall classification performances for test and pilot fields combined, summarized in Table 17, shows that Fayette-III-1 and Huntington-III had differences in performance of approximately 20% between registered and non-registered results. Huntington and Fayette had the smallest average field sizes, and it would be expected that the effect of any registration errors would be magnified for small fields. From this, it appears that average field size may be one factor affecting classification performance in registered data sets.

Table 17. Overall classification performance of registered and non-registered forms of ERTS data.

segment-date	average field size (acres)	without weights				with weights	
		Non-Reg.	Reg.	Non-Reg. w/Reg. Stats	Reg. w/Non-Reg. Stats	Non-Reg.	Reg.
Fayette-II	16.8	42.4	53.1	49.5	48.1	41.8	50.2
Fayette-III-1	16.8	71.0	51.1	51.2	69.7	74.1	54.6
Livingston-IV	30.7	70.1	67.3	68.2	68.4	73.3	66.3
White-V	34.1	76.2	75.1	76.7	74.5	78.3	76.4
Huntington-III	20.1	66.1	44.8	48.0	65.7	63.3	45.8

VII. Spectral Discriminability of Corn, Soybeans, and "Other"

A. Introduction

In Section V the effects on classification performance of training set variation were discussed. In this section the potential spectral discriminability of corn, soybeans, and "other" will be examined in the context of the level of classification performance which would be possible if the number of training points were not limited (i.e. if all fields were used for training the classifier). Using all fields for training the classifier should provide an optimistic upper limit on classification performance and an indication of the true spectral discriminability of the cover types of interest under the CITARS conditions (i.e. ERTS data for selected locations and times). By comparing these results to the original classifications it should also be possible to determine if classification accuracy was severely affected by the limitation of available training data.

B. Procedures

Ten data sets, described in Table 14 were selected for classification using all training, test, and pilot fields for training. The analysis procedure was the basic procedure used by LARS for CITARS classifications of ERTS data (i.e. LARS/SP1). Overall correct classification of field center pixels was used as the measure of classification performance.

C. Results and Discussion

Classification results obtained with the original training sets (fields from 20 quarter-sections) are compared in Table 18

Table 18. Comparison of overall classification performance for classifications based on training statistics from training fields versus all fields classified.

Segment-Period-Pass	"Source of Training Data"				
	Training Fields		All Fields		
	Classification Results		Classification Results		
	Training	Test*	Training	Test*	All Fields
HU-III	92.3	44.8	83.1	82.9	82.9
LI-III	78.1	59.9	66.9	70.8	69.9
FA-III-2	77.8	59.3	72.9	74.0	73.6
LE-III-2	80.2	58.3	72.4	44.3	53.9
LE-IV	75.5	55.0	68.3	65.2	66.3
WH-V	87.9	75.1	78.9	77.1	77.7
FA-V	90.5	79.6	83.5	84.3	84.0
SH-VI	77.1	49.8	71.5	65.9	67.1
HU-VII	81.2	49.6	72.6	78.6	77.3
SH-VII	73.5	48.5	48.5	48.4	48.4

*Test = test + pilot fields as defined for CITARS with results obtained using all fields for training. The classification results for all fields show that in some instances (i.e. HU-III, FA-V, WH-V, and HU-VII) reasonably high classification performance (greater than 75%) would be possible if adequate training data were available. In the remainder of data sets classified the low performances indicate that the cover types of interest are not spectrally separable in the ERTS bands.

Comparison of the results for the four best classifications to the results of the original classifications of test + pilot fields shows that WH-V and FA-V (75.1 and 79.6, respectively) were classified reasonably well with the original training fields, but HU-III (44.8) and HU-VII (49.6) were not. This means that in at least two cases the original training fields were not representative of all fields in the segment and that performance was adversely affected by inadequate or non-representative training sets.

The results indicate that there were two different situations present: (1) For the available spectral bands, the spectral characteristics of the cover types of interest were potentially different enough to enable "good" classifications to be made; and (2) the cover types were sufficiently similar that accurate classifications could not be obtained by methods currently available which rely only on the spectral information content of ERTS multispectral scanner data. In the former case the level of classification accuracy actually achieved depends on the quantity and quality of training data; whereas, in the latter case performance is low (< 75 percent overall correct classification of test pixels) regardless of the amount and kind of training data available.

Of course, recognition might be improved in both cases by the use of temporal and/or spatial information.

These conclusions are necessarily limited to the ERTS data, cover types, locations, and times considered in the CITARS experiment. In particular, it should be noted that the conclusions about the spectral separability of the cover types are based on the measurements made by the ERTS multispectral scanner. Evidence exists indicating that if the ERTS data had more spectral bands and/or greater dynamic range the separability of the cover types would be increased [10]. This question was investigated by analyzing aircraft multispectral scanner data having more spectral bands and greater dynamic range for one of the CITARS segments. Results of that investigation are presented in the following section of this report.

VIII. Analysis of Aircraft Multispectral Scanner Data

A. Introduction

One of the original objectives of CITARS was to compare classification performances of ERTS-1 MSS data to aircraft-acquired MSS data. Aircraft scanner data was acquired by the Bendix M^2S system for six missions and by the ERIM M-7 system for two missions. Subsequent resource and time constraints limited the analysis primarily to the ERTS data. The comparison, however, is still an important one to be made, particularly in light of the unexpected low performances obtained for the ERTS data classifications. With this background, one of the flightlines of M-7 scanner data over the Fayette Co., Illinois segment was analyzed by LARS.

B. Procedures

Both the ERTS and aircraft scanner data were collected over the Fayette Co. segment on August 21, 1973. The Fayette data was selected primarily because of its availability for analysis (no Bendix M^2S data was available to LARS and only the data for the ERIM M-7 mission over Fayette Co. on August 21 had been digitized at the time of this analysis). The M-7 scanner data analyzed was collected over the western two-thirds of the segment (two passes were required to cover the entire segment) from an altitude of approximately 4,650 meters at 8:30a.m. local time. The low solar elevation at the time of data collection caused severe sun angle effects readily apparent in the data. Therefore, a preprocessing algorithm for mean angle response correction was applied to the data before analysis. Also, because the flight was flown so early in the morning the utility of the thermal channel for providing crop discriminability information was probably limited. The aircraft scanner data had 12 wavelength bands and an instantaneous field of view of approximately 12 meters compared to 80 meters for ERTS data. The 12 wavelength bands are shown in Table 19.

Sixteen of the 20 quarter-sections and 19 of the 20 sections in the segment were contained in the aircraft data. Coordinates were obtained for a majority of fields present in the quarter-sections and sections taking care to insure that only "pure" field center pixels were sample. Training statistics were developed in the same manner as for the ERTS data analyses (i.e. LARS/SP1 and LARS/SP2 were used). The only exception was that four of the 12 available channels for classification were chosen based on the

Table 19. Wavelength bands of the M-7 scanner.

Channel	Wavelength Band (micrometers)	Spectral Region
1	.41-.48	visible
2	.48-.52	visible
3	.50-.54	visible
4	.52-.57	visible
5	.55-.60	visible
6	.58-.64	visible
7	.62-.70	visible
8	.67-.94	near infrared
9	.71-.73	near infrared
10	1.00-1.40	near infrared
11	2.00-2.60	middle infrared
12	9.30-11.70	thermal infrared

maximum average pairwise transformed divergence of the classes. The four channels with the greatest average pairwise divergence were .58-.64, .71-.73, 1.00-1.40, and 2.00-2.60 μm. The number of subclasses of corn, soybeans, ag "other" and non-ag "other" was two, two, five, and four, respectively, for the aircraft data. The number of subclasses of corn, soybeans, and "other" was two, four, and four, respectively, in the ERTS data. The classifications were performed with and without class weights and classification performance tabulated for training, test, and pilot fields.

C. Results and Discussion

Classification performance for field center pixels (test fields) for the ERTS and aircraft data are shown in Table 20. Although there were substantial differences for individual classes between the ERTS and aircraft data classifications, overall performance for the two data sets was nearly identical; performance for with weights and without weights classifications averaged 78 percent for ERTS vs. 77 percent for aircraft. Use of class weights did not significantly affect performance for either the ERTS or aircraft data classifications.

Another topic of interest is the wavelength bands indicated by the feature selection algorithm as best for discriminating among the training classes for the aircraft data. Table 21 shows the best five combinations of four, five and six channels. Every channel combination in the table includes at least one visible and two near infrared bands. In the combination of four channels, the remaining band was middle infrared, four out of five times. For

Table 20. Classification performance (percent correct) for field center pixels of ERTS-1 MSS data and aircraft MSS data, Fayette Co., Illinois, August 21, 1973.

	Training Fields		Test Fields*	
Class	W/ Wts.	W/O Wts.	W/ Wts.	W/O Wts.
		ERTS-1 MSS data		
Corn	77.1	80.0	79.0	76.2
Soybeans	89.6	89.1	95.0	94.4
"Other				
"Other"	96.4	96.9	65.2	61.5
Overall	90.5	91.0	79.6	77.2
		Aircraft MSS data		
Corn	83.7	86.6	69.1	71.3
Soybeans	84.9	85.9	76.0	76.0
"Other"	91.6	91.3	83.4	83.3
Overall	86.7	87.7	76.9	77.4

*Test = test + pilot fields

Table 21 Rank of channel combinations on basis of average divergence.

Rank	Channels	Minimum Divergence	Average Divergence	Spectral Regions
a. Best five combinations of four channels.				
1	2,9,10,11	1390	1939	V,NIR,NIR,MIR
2	7,9,10,11	1363	1932	V,NIR,NIR,MIR
3	5,9,10,11	1345	1931	V,NIR,NIR,MIR
4	6,8,9,10	1132	1930	V,NIR,NIR,NIR
5	2,9,10,11	1278	1925	V,NIR,NIR,MIR
b. Best five combinations of five channels.				
1	6,8,9,10,11	1457	1963	V,NIR,NIR,NIR,NIR
2	7,8,9,10,11	1456	1960	V,NIR,NIR,NIR,MIR
3	5,8,9,10,11	1450	1958	V,NIR,NIR,NIR,MIR
4	2,8,9,10,11	1468	1956	V,NIR,NIR,NIR,MIR
5	3,8,9,10,11	1417	1954	V,NIR,NIR,NIR,MIR
c. Best five combinations of six channels.				
1	6,8,9,10,11,12	1499	1969	V,NIR,NIR,NIR,MIR,FIR
2	2,6,8,9,10,11	1493	1968	V,V,NIR,NIR,NIR,MIR
3	4,6,8,9,10,11	1498	1968	V,V,NIR,NIR,NIR,MIR
4	1,6,8,9,10,11	1508	1968	V,V,NIR,NIR,NIR,MIR
5	4,7,8,9,10,11	1491	1967	V,V,NIR,NIR,NIR,MIR

the combinations of five channels, the five best combinations all included the available reflective infrared (three near and one middle), and the fifth channel was a visible band. The best five combinations of six channels also included the four reflective infrared bands and a visible band. The remaining band was another visible four out of five times. Caution should be exercised in making any conclusions about the utility of the far infrared (emissive infrared, or thermal) due to the fact that the data was collected at 8:30a.m.

This comparison for one segment and time of ERTS and aircraft data classification performance indicates that there was little if any difference between the two. However, this conclusion was based on analysis of only one segment and time. Further, the ERTS data classification had the highest classification accuracy of all the CITARS classifications and the aircraft scanner data was collected under suboptimal conditions with very low sun angle. In spite of attempts to "correct" or compensate for the sun angle problem, this is likely (because of its severity) to have had an adverse effect on classification performance. The combination of these two effects may have brought the ERTS and aircraft data classifications closer together than they might be under other conditions. The classification performances obtained in this experiment with aircraft data do not approach those obtained in previous classifications of aircraft data (i.e., 1971 CBWE). To better determine the level of classification accuracy which could be anticipated from aircraft data in the CITARS context, performance of additional analyses is recommended.

Part 3. Summary and Conclusions

The classification results obtained by LARS were presented in Parts 1 and 2 of this report. Part 1 contains the "regular" CITARS classification results and Part 2 describes the results of several additional investigations. Since the results of the statistical analyses are presented in Volume IX and discussed in Volume X of the final report along with results from EOD and ERIM, only the results specific to LARS have been discussed in this report.

One of the important results of CITARS at LARS has been the definition, implementation, and evaluation of an automatable and repeatable data analysis procedure. The newly defined procedure was first used for CITARS, but it performed very well relative to other procedures both in terms of data analysis efficiency and classification performance. The efficiency of the procedure is indicated by the fact that the 15 local and 20 non-local classifications using both the SP1 and SP2 procedures were all completed by two part-time analysts in three months. The procedure was also shown to yield nearly identical results when used by several analysts on the same data sets. Subsequent tests showed that the performances obtained using the procedure were similar to those obtained using analyst dependent procedures.

Statistical comparisons of the two LARS procedures, SP1 and SP2, showed no significant difference between them as measured by either classification accuracy or proportion estimation. The procedure identified as SP1 used equal prior probabilities, while SP2 used unequal prior probabilities based on 1972 county acreage estimates by the Statistical Reporting Service of the U.S. Department of Agriculture.

There are three possible reasons why unequal prior probabilities did not produce significantly better results than equal prior probabilities: (1) the weights came from 1972, while data was from 1973, and the true proportions could have changed from one year to the next; (2) the weights pertain to counties but were applied to segments, which are fractions of counties and might therefore have different true proportions; (3) the analysis of variance was performed on results for sections, and sections vary within segments.

Classification performances for CITARS were generally lower than originally anticipated. For this reason, several experiments were performed to investigate the effect of various factors, and the results were presented in Part 2 of this report. Six factors which may have affected the performance were identified and investigated: (1) method of evaluation used, (2) data analysis and classification procedures used, (3) availability of training data, (4) registration accuracy, (5) spectral characteristics of the scene, and (6) characteristics of the ERTS data.

Evaluation of the classifications was based on crop identifications determined by photointerpretation. These identifications must be accurate if performance evaluation are to be reliable. Tests of photointerpretation accuracy indicated that the crops in 95-98 percent of the fields were correctly identified (5). It was therefore concluded that photointerpretation errors did not substantially influence classification performance.

To investigate the effects of the data analysis procedures used, an experiment was conducted using several alternative procedures. The alternative procedures did not result in improved classification performances, indicating that the generally low classification performances obtained in CITARS cannot be attributed to the data analysis procedures used.

Another experiment was conducted to determine the effects of training set size and selection. Results showed that significant differences in classification performance can be obtained with different training sets, and that training set size alone does not determine the representativeness of a training set.

Comparisons of classification performance for registered and non-registered data showed that there was no significant difference between the two forms of ERTS data.

Classification performance depends largely on the degree of spectral separability of the cover types of interest. An investigation of the data characteristics showed that there were some cases in which the cover types of interest were spectrally different enough to enable discrimination among them (provided adequate training data was available). However, in other instances the cover types were so spectrally similar (as measured by the ERTS system) that they could not be discriminated regardless of the amount of training data used.

Since accurate identification of crops requires spectral separability, classification performance depends not only on the spectral characteristics of the cover types but also on the ability of the scanner to detect and measure spectral differences. To study the effect of the ERTS scanner on classification performance, a data set collected by an airborne multispectral scanner system having more wavelength bands over a wider region of the spectrum and greater sensitivity, and dynamic range was analyzed for comparison. Although there were substantial differences in performance for individual classes between the ERTS and aircraft data analyses, overall performance for the two data sets was nearly identical.

REFERENCES

1. Swain, P.H. 1972. "Pattern Recognition. A basis for Remote Sensing Data Analysis", LARS Information Note 111572. Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana.

2. Phillips, T.L. (ed.). 1973. "LARSYS Version 3 User's Manual." Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana.

3. Davis, B.J. and Swain, P.H. 197-. "An Automated and Repeatable Data Analysis Procedure for Remote Sensing Applications". Proc. Ninth Int'l. Symposium on Remote Sensing of Environment. pp. 771-774.

4. National Aeronautics and Space Administration, Johnson Sqace Center. 1975. "Crop Identification Technology Assessment for Remote Sensing (CITARS), Final Report."

Volume 1. Task Design Plan.
2. Ground Data Acquisition
3. ERTS-1 Multispectral Scanner Data Acquisition.
4. Photointerpretation.
5. ERTS-1 Data Preparation.
6. Classification Results, Laboratory for Applications of Remote Sensing, Purdue University.
7. Classification Results, Environmental Research Institute of Michigan.
8. Classification Results, Earth Observations Division, Johnson Space Center, National Aeronautics and Space Administration.
9. Statistical Analyses of Results.
10. Interpretation of Results.

5. Williams, D.L., Morain, S.A., Barker, B., and Coiner, J.C. 1973. "Identification of Winter Wheat from ERTS-1 Imagery". NASA SP-327. Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1. Volume I: pp. 11-18.

6. Thomson, F.J. 1973. "Crop Species Recognition and Mensuration in the Sacramento Valley". NASA SP-327. Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1. Volume I: pp. 181-188.

7. Draeger, W.C. 1973. "Agricultural Applications of ERTS-1 Data". NASA SP-327. Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1. Volume I: pp. 197-204

8. Bauer, M.E. and Cipra, J.E. 1973. "Identification of Agricultural Crops by Computer Processing of ERTS MSS Data". NASA SP-327. Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1. Volume I: pp. 205-212.

9. Snedecor, B.W. and Cochran, W.G. 1967. Statistical Methods. The Iowa State University Press, Ames, Iowa. Sixth ed.

10. National Aeronautics and Space Administration, Johnson Space Center. 1974. "Corn Blight Watch Experiment Final Report. Volume III: Experimental Results".

11. Hall, F.G., Bauer, M.E., and Malila, W.A. 1974. "First Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS)". Proc. Ninth Int'l. Symp. on Remote Sensing of Environment. pp. 1171-1191.