113076

QUARTERLY REPORT

Reporting Period: September 1, 1976-November 30, 1976
Contract Number: NAS9-14970
Title of Investigation:

Research in Remote Sensing of Agriculture, Earth Resources, and Man's Environment

Principal Investigator:

D. A. Landgrebe
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47906

Submitted to:

NASA Lyndon B. Johnson Space Center
Attn: Jon D. Erickson, Mail Code TF3
Houston, Texas 77058
Mark for: Contract NAS9-14970

2.1 Test of Boundary Finding/Per Field Classification (ECHO)

Background. During the previous quarter significant steps toward the projected accomplishments for Task 2.1 were taken. These included:

1. Preparing a detailed implementation plan.
2. Educating the ECHO project staff.
3. Securing the existing research software.
4. Learning how the existing software worked.
5. Selecting the appropriate research programs to refine, alter and further develop to meet task objectives.
6. Putting these selected versions in running order.
7. Determining the appropriate configuration for the supervised processor.
8. Formulating an upgrade plan for the supervised processor.
9. Drafting an experimental design.

A description of these activities and information concerning them is available in the first quarterly report.

Major Activities. This section of the report deals with those sub-tasks which have been completed during the second quarter.

A. User Consultation.

During September a seminar was held at LARS on ECHO. The purpose of the seminar was to collect together as much experience as possible on the use of ECHO and its performance characteristics in as wide a set of circumstances as possible. At that seminar the general concept behind ECHO was presented along with information concerning the Supervised and Non-Supervised implementations which are being developed for Task 2.1. Information on how to get access to the research versions of the Non-Supervised processors was released at the seminar.

In early October a User's Workshop was held involving several analysts at LARS having experience in the use of the research software for the ECHO processors. The goals of this workshop included release of additional information concerning the Non-Supervised processor, examples of the object map recently added as an output and the classification results maps obtained by varying the annexation parameters for the Non-Supervised processor. A general flow-chart of the proposed restructured version was presented and discussed. Certain interesting characteristics of the processors were discussed by these analysts at both the seminar and the Workshop.

One suggestion resulting from these meetings concerned the cell selection parameter for the Non-

Supervised processor. Presently the cell homogeneity is tested by $s_i/\mu_i < T$, where s_i and μ_i are the standard deviation and cell mean in the i^{th} channel, respectively. This use of the coefficient of variation was criticized on the grounds that not all scanners increase in noise as mean spectral response increases and that a zero response (black) level for a scanner is usually represented by a data value greater than zero. In addition, the same threshold was applied to all channels, irrespective of differences in variability. Consequently, the cell selection test for the Non-Supervised processor will be changed to $s_i < T_i$, where T_i can now be specified for each channel.

A second item raised during these meetings concerned the applicability of statistics generated by traditional (per point) training methods for object classifiers is an area where research is needed. Utilizing the object map produced by the Non-Supervised processor may be at least one appropriate way of training object classifiers.

A third point brought to the surface during user consultation was the tendency for ECHO to favor classes with large variances. This occurs when pixels from separate classes are in a single cell which passes the cell homogeneity test. Such a cell will have a high variance, making it look similar to a class with a high variance. Figure 1 illusstrates the situation in one dimension.

Figure 1

B's distribution   A's distribution

D's distribution

C's distribution

```
1  23     564      78
B  CC     AAA      DD
```

Suppose pixels 1-4 fall in a single 2x2 cell. If these pixels were classified on a per point basis, point 1 would be classified as B, points 2 and 3 as C, and point 4 as A. When they are combined, however, they will likely all be classified as C. This occurs because the assumption is implicitly made that the four points are from a single class. The variance of the cell is not much different than the variance expected for class C and, at the mean for the cell, C is the most likely class. Since point 4 is next to points 1, 2 and 3, it is more likely that these points are all of class C. However, consider points 5, 6, 7 and 8. The variance for these points is not much different than the variance for class C and, at the mean of the points, C is the most likely class. We would have preferred points 5 and 6 to be classified as A and points

7 and 8 as D.

Consideration is being given to this problem, though no completely satisfactory solution has yet been found. One method of avoiding the problem is to have classes with nearly equal variances, but such a method is no solution if the real ground cover classes do indeed have differing variances.

B. Supervised Software Restructuring.

Restructuring of the Supervised processor should be completed by mid-December. Processing flow and control cards will be as described in the report for the first quarter. The processor will optionally allow the user to either start or end with an intermediate results tape or start from a multispectral image storage tape and produce a classification directly. The advantage of the intermediate results tape is that it allows multiple classifications to be performed using different annexation parameters without repeating the more time-consuming process of calculating cell covariance matrices.

C. Determine the Non-Supervised Processor Configuration.

A processing flow-chart and a proposed control card listing for the Non-Supervised processor appear in this report. Like the Supervised version, this processor will be converted from a two phase to a single phase processor with the option of producing an intermediate product. The intermediate product will be a tape which contains an object map and is in LARSYS Multispectral Image Storage Tape (MIST) format. This tape will contain the mean value of all the pixels of an object for each channel as well as the covariance matrix for the object. An object map can then be printed by combining the information from the various channels on the object tape. A program converting this LARSYS formatted object map tape to universal format will be written so that these tapes may be transferred to Johnson Spaceflight Center for NASA's analysis of ECHO's utility to a LACIE analyst in training field selection. In addition, an object map generated on the line printer will optionally be produced by the processor.

An alternate version of this processor will be investigated to examine field-to-field annexation (as opposed to the current cell-to-field (CTF) annexation. One advantage of field-to-field (FTF) annexation is that the amount of Type II error (categorizing cells as belonging to separate objects when they really belong to a single object) is reduced. The tradeoff with FTF is:

1) an increase in CPU time required;
2) a probable increase in Type I error (categorizing cells belonging to separate objects as belonging to a single object);
3) annexation parameters will have differing effects in FTF and CTF schemes, so that a separate set of parameter guidelines will be necessary.

An additional advantage of FTF annexation is that disjoint objects which pass the annexation criterion could be assigned like symbols in the object map produced by the processor.

D. <u>Develop Data Normalization Program</u>.

To examine the effects of the large variance class favoritism problem divorced from the confounding effects of the input data failing to meet the underlying classification assumptions of normal data distribution within each class and of the statistics deck adequately describing the data, development of a program which generates a MIST tape from a results tape and a statistics file was completed. Artificial data (satisfying the classifier assumptions) produced by this program will also be useful in the examination of other aspects of the cell selection process.

E. Revise Experimental Design.

The experimental design has been revised to reduce the amount of computer time required to perform the analysis outlined in the first quarterly report. This design has not yet been finalized due to some uncertainty concerning details of the evaluation techniques. The current plan (given in the appendix to this section) is tentative in some aspects, and is presented so that NASA will have an idea of the lines along which we are proceeding.

One reason the design remains soft is the use of field-center-pixel methods as a measure of classification accuracy. Fields must be large to have "pure" center pixels. Consequently, using field-center-pixels as an accuracy measure leaves ECHO's impact on small fields and mixture pixels unmeasured. Using proportion estimates does reduce the problem, but it does not eliminate it. Proportions give no measure of the pixel-by-pixel accuracy of the classifier. It is especially hard to draw strong conclusions concerning an optimal cell selection parameter using these measures, as both reflect the accuracy of the classifier more in homogeneous areas.

A second problem is the measurement of classifier "noise". When asked what advantages they find in using ECHO, analysts mostly uniformaly reply that ECHO reduces classification noise, gets rid of the "salt and pepper effect", and smooths out the image. Analysts seem to call repeated, rapid class changes "classification noise". With the ECHO processors, rapid class changes can take place only if neighboring pixels repeatedly are significantly different from each other; whereas, using pixel-at-a-time classification, this rapid class switching could occur when neighboring pixels differ relatively slightly. By requiring a significant contextual difference between pixels before class changes can occur, ECHO should reduce "classification noise".

Quantitatively defining and measuring "classification noise" in an absolute sense is hard. A comparative measure, however, might be to count the number of class changes in fifty lines taken at random from per point classification results and several ECHO classifications (where cell selection and annexation are varied). The usefulness of this measure is still being questioned, however.

F. Locate and Select Data Sets.

All candidate data sets have been selected. Candidate sets include the 26 runs in the CITARS data base, the 24-channel aircraft and simulated Thematic Mapper data from North Dakota and Kansas, a Corn Blight Watch aircraft data set and a LANDSAT data set over the Platoro Reservoir area of Colorado. The number of these candidate sets which will actually be used in the experiments will depend on the resources which are available. Not all of these sets will be used. Selection of candidate sets to actually be used in the experiments will be made on the basis of:

1) Type of scanner

   A reasonable combination of aircraft, LANDSAT and simulated Thematic Mapper sets should be included.

2) Per Point Classification Accuracy

   Data and classification problems exhibiting a range of classification accuracies should be included.

3) Number of Spectral Classes

   It would be beneficial to observe how ECHO functions under a mix of class inputs.

4) Types of Classes

How ECHO performs on agricultural areas, forested areas, rangelands, etc., should be assessed.

Work in Progress. This section reports on the work which is currently being pursued under Task 2.1.

A. Implement Non-Supervised Restructuring.

Preliminary alterations to the classification algorithm and to the control card reader have been made for the Non-Supervised processor. Restructured array storage problems are being examined prior to the major reprogramming task.

B. Secure Statistics Decks.

Statistics decks for much of the LACIE, CITARS, and Thematic Mapper sites are already available, and their use will save considerable effort and computer time. Statistics decks for the 24-channel data used to produce the simulated Thematic Mapper data were not available but have been generated. Since the LACIE sites are reduced to a two class problem, but ground truth is available in the SRS/LACIE sites for more than two classes, some of the LACIE sites will be expanded to more than a two class analysis. This will require generation of new statistics decks for these areas.

C. <u>Run Normalization Comparisons</u>.

These are experiments run using the simulated data generated by the normalization program which was discussed in the previous section. It was noted in mid-August that what analysts with experience in the use of the Supervised ECHO processor suggested as a reasonable cell selection parameter was much higher than a .999 confidence level for testing the assumption:

H_o: The cell is homogeneous

H_1: The cell is singular

Examination of the distribution of the discriminant values (which theory says should be chi-squared with (cell size x number of features) degrees of freedom) showed that most cells had values falling above the 90% confidence level. Development of this program was begun to determine whether the failure of the test data to meet the classification assumptions (multivariate normal) was contributing to this circumstance. It turned out that most of the cells with high discriminant values contained a pixel which a per point classifier would classify as a different class; however, some of the cells with low discriminant values also contained pixels which would be classified differently by a per point classifier. Since this situation was observed to occur in both synthetic and unaltered data sets, it could not be concluded that the "unusual" values were due to violation of the normality assumption.

D. ECHO Experiments.

Runs of the cell selection, cell size portions of the ECHO experiments have begun. Preliminary to the actual classification of each data set is the selection of test fields. Average field size must be calculated for each data set. This process is accomplished utilizing the test fields and a computer program, which yields the average field size for each class as well as the overall average field size for the specified field description deck. This program is currently being run on some of the SRS/LACIE candidate data sets.

E. User's Guide Preparation.

The User's Guide will follow the general outline of the individual descriptions in the LARSYS User's Manual, Volume II. A general outline for the ECHO User's Guide is:

I. Synopsis of the ECHO Concept

II. Supervised ECHO

A. Purpose

B. Input Description

C. Output Description

D. Classification Algorithm

E. Comments

1. Limitations

2. Parameter Selection Guide

3. Special Uses

III. Non-Supervised ECHO

(same Sub-headings)

Figure 2.

RESTRUCTURED PROCESSING FLOW

Non-Supervised ECHO

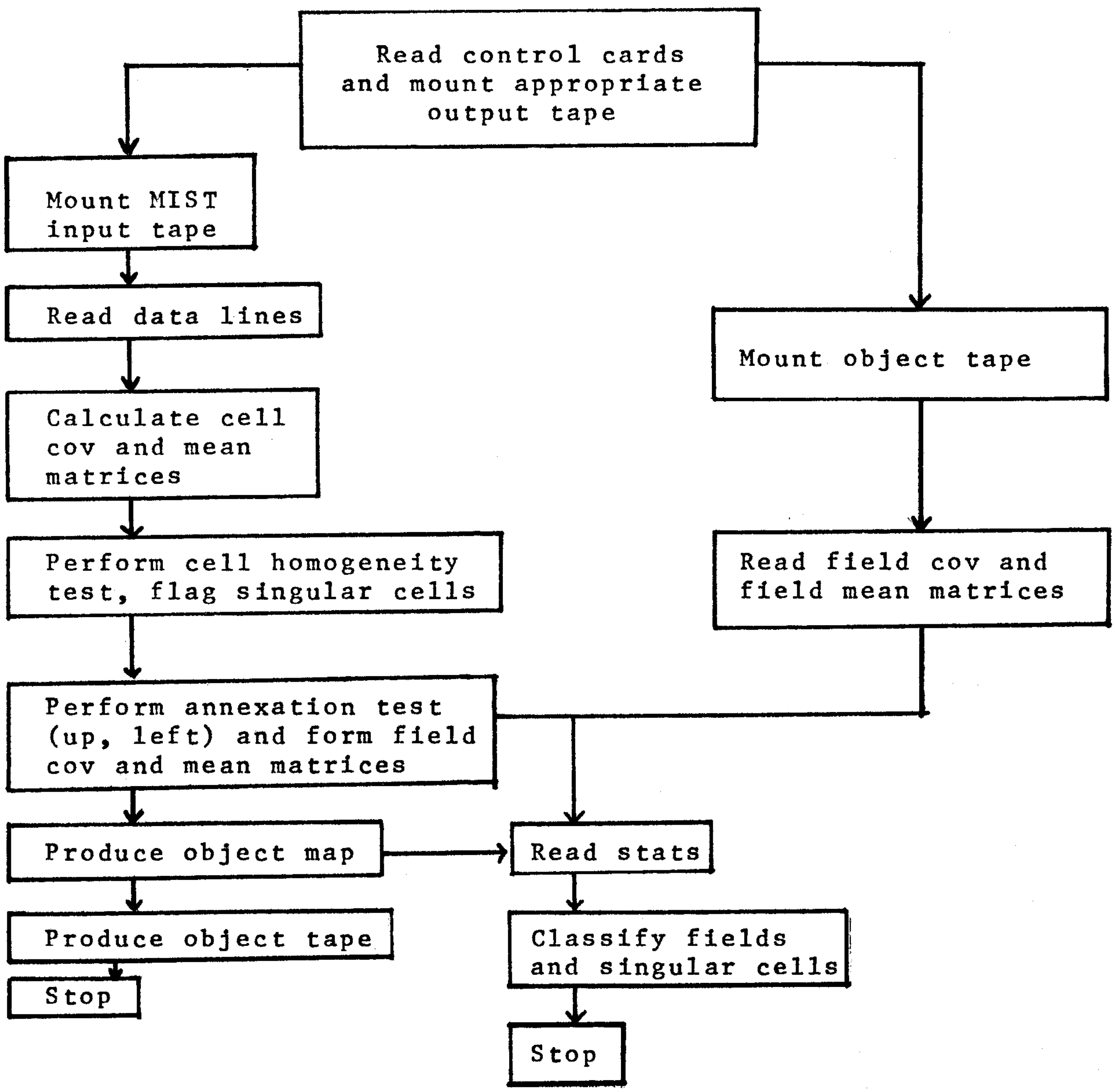


Table 2.1-1

PROPOSED CONTROL CARDS FOR NON-SUPERVISED ECHO

REQ	KEY WORD	CONTROL PARAMETER	FUNCTION	DEFAULT
+	*NSECHO	(NONE)	SELECT NON-SUPERVISED ECHO PROCESSOR	(NONE)
	RESULTS		ESTIMATION OF CLASSIFICATION	CLASSIFICATION RESULTS STORED ON DISK
		TAPE(TTT)	RESULTS ON TAPE(TTT)	
		FILE(FF)	FILE FF.	
		INITIALIZE	INITIALIZE FILE NEW RESULTS TAPE (REQUIRED WHEN USING A NEW TAPE)	
1 3	OBJECT		PRODUCE OR USE A RUN TAPE TTT,	NO RUN TAPE PRODUCED OR USED
		TAPE(TTT)	FILE FF CONTAINING AN OBJECT	
		FILE(FF)	MAP AND CORRELATION MATRICES	
2	CELL	SIZE(N)	PARTITION DATA INTO NXN CELLS AND TEST THESE CELLS FOR	2
		SELECTION($S.S_1,S.S_2,...S_n$)	HOMOGENEITY BY COMPARING WHERE $S.S_J$ IS THE THRESHOLD FOR THE STANDARD DEVIATION FOR CHANNEL J	$S.S_J$-0
2 3	ANNEXATION	MEANS(.MMM)	USE .MMM AS THE SIGNIFICANCE LEVEL TO TEST FOR EQUIVALENT MEANS. (SEE NOTE 1.)	(NONE)
		COVARIANCE(.CCC)	USE .CCC AS THE SIGNIFICANCE LEVEL TO TEST FOR EQUIVALENT COVARIANCE MATRICES. (SEE NOTE 1.)	(NONE)
	CLASSES	NAME (P1/C1,C2/)	POOL STATISTICS FOR TRAINING CLASSES C1,C2,...; ASSIGN NAME AS POOL NAME AND P1 AS THE POOL NUMBER. (SEE NOTE 2.)	INDIVIDUAL CLASSES USED
	OPTIONS	OBJECTAPE	PRODUCE THE CLASSIFICATION FROM THE DATA ON THE OBJECT TAPE	CLASSIFICATION PRODUCED FROM MULTISPECTRAL IMAGE STORAGE TAPE
	PRINT	STATS	PRINTS STATISTICS TO BE USED. (SEE NOTE 3.)	NO STATISTICS PRINTED
		OBJECTS	PRINT AN OBJECT MAP. (SEE NOTES 4 AND 5.)	NO MAP PRINTED
		CLASSIFICATION	PRINT A CLASSIFICATION MAP. (SEE NOTES 3 and 5.)	NO MAP PRINTED

TABLE 2.1-1 CONTINUED

REQ	KEY WORD	CONTROL PARAMETER	FUNCTION	DEFAULT
	SYMBOLS	S1,S2,S3,...	SYMBOLS USED TO PRODUCE CLASSIFICATION MAP	ARBITRARY SYMBOLS ASSIGNED
2 3	CHANNELS	I,J,K,...	CHANNELS I,J,K... ARE SELECTED	(NONE)
	CARDS	READSTATS	STATISTICS FILE WILL BE READ FROM CARDS	STATISTICS FILE EXPECTED FROM DISK
	DATA	START OF DATA DECK		
			PUNCHED STATISTICS FILE FROM STATISTICS FUNCTION IF 'CARDS READSTATS' CONTROL CARD IS INCLUDED.	
2 3	DATA	START OF DATA DECK		
			FIELD DESCRIPTION CARDS DESCRIBING AREAS TO BE CLASSIFIED (ALWAYS REQUIRED). EITHER FORM OF THE FIELD DESCRIPTION CARD MAY BE USED.	
	END			

+ ALWAYS REQUIRED,
1 REQUIRED IF CLASSIFYING FROM OBJECT TAPE,
2 REQUIRED IF CLASSIFYING FROM MULTISPECTRAL IMAGE STORAGE TAPE.
3 REQUIRED TO PRODUCE OBJECT TAPE.

NOTE 1... THE CONFIDENCE LEVELS MUST BE ONE OF .1, .05, .025, .01, .005, or .001. AS THE VALUE BECOMES SMALLER, THE TENDENCY TOWARD ANNEXATION INCREASES.

NOTE 2... THE FORMAT C1, C3,... MAY BE USED FOR THE CLASSES TO INDICATE ONLY A SUBSET OF THE TOTAL CLASSES (C1, C2,...) IN THE STATISTICS FILE SHOULD BE CONSIDERED. IF A CLASSES CARD IS USED, ALL DESIRED CLASSES MUST BE EXPLICITLY SPECIFIED. IF NO CLASSES CARD IS INCLUDED, ALL CLASSES WILL BE USED.

NOTE 3... THESE PRINT OPTIONS ARE AVAILABLE ONLY IF A CLASSIFICATION RESULTS FILE IS BEING CREATED ON TAPE OR ON DISK--THEY ARE NOT AVAILABLE IF ONLY AN OBJECT MAP AND/OR OBJECT TAPE IS BEING PRODUCED.

NOTE 4... THIS OPTION PRINTS THE OBJECTS USING AN ARBITRARY STRING OF SYMBOLS. WHEN THE STRING IS EXHAUSTED, THE PROCESSOR REUSES THE SYMBOLS, ASSIGNING THEM IN ORDER AS NEW OBJECTS ARE ENCOUNTERED.

NOTE 5... ONLY ONE MAP CAN BE PRODUCED FOR A GIVEN RUN.

Appendix to 2.1

Experimental Plan for Task 2.1

Introduction

There are two general types of experiments which are to be performed with the ECHO processors: exploratory experiments and comparative experiments. Exploratory experiments are analyses designed to determine how to most effectively use the ECHO processors. The effect of different choices of parameter values on classification results will be explored to determine an optimum set of parameter values for a data set. The comparative experiments will assess the relative value of the ECHO process as compared to the point classifier. The optimum classifications for several data sets will be compared to the corresponding classifications from conventional analysis procedures using the LARSYS maximum likelihood point classifier.

The value of a classifier can be measured in terms of several criteria. Two important criteria which will be used in both the exploratory and comparative experiments are classification accuracy and the variability present in the classification map. The classification accuracy will be examined in two ways: classification of field center pixels, and proportion estimates. It is not clear that these traditional methods of assessing classification accuracy are appropriate in evaluating a boundary-finding algorithm such as ECHO. These measures are subject to change should a more appropriate measure be discovered.

The variability present in the classification map will be measured by tabulating the number of class changes per line. In addition, the CPU time for each analysis will be recorded, and the classification error for objects of different size will be tabulated. The unsupervised ECHO will also be evaluated on the basis of object map integrity. The object map integrity refers to errors made in finding objects in the scene. The errors consist of either mistakenly combining two distinct objects into one field or not combining sub-fields of a single object.

Exploratory Experiments

The exploratory experiments are meant to assess the effects of parameter variation on the classification results, with the goal of finding a procedure to specify optimum parameter values. One way to do this would be to take a wide range and many values for each parameter and then run analyses using all possible combinations of parameter values. This would give very good information but require large computational resources. Instead, a simpler design will initially be used, and the results from those analyses used to select additional values of the parameters to be used in an analysis. Thus the optimum choices of parameter values for a given data set will be approached sequentially. The initial choices of values for each parameter will be essentially the extremes and midpoint of a reasonable range.

A variety of data sets will be explored, with an emphasis on Kansas LACIE segments and the simulated Thematic Mapper data. Additional LANDSAT data over non-agricultural areas is being considered so that the effects of more complex ground scenes may be examined. The present selection of data sets is: 2 aircraft; 4 LANDSAT, and 8 simulated Thematic Mapper (2 sites and 4 resolutions), for a total of 14 data sets.

The rationale and initial choice of values for each parameter is further explained below.

A. Cell Size (CELW)

It is important that cell size used in the ECHO processor is not larger than the objects of interest in the scene. If the cell were larger, one of two possibilities exist: either a great deal of cell-splitting would take place, with corresponding increases in classification time, or cells which should be split would not be split, placing different objects on the ground in the same cell and leading to misclassifications.

For the data sets which have fields available, the average field size in pixels can be calculated. CELW will range from two to CELMAX, the interger part of $(\text{average field size})^{\frac{1}{2}}$, with the initial values being: 2, the integer part of (CELMAX+3)/2, and CELMAX.

This method of setting cell size assumes that the fields adequately represent the distribution of object sizes. This is in general not true, particularly for analyst-selected fields where the fields would tend to represent larger, more easily seen objects.

Fields are available for the data sets chosen for these experiments; but fields will not always be available to an analyst wishing to use an ECHO processor. Without a set of fields, some measure of the complexity, or size and number of objects, of the scene is needed.

B. Cell Selection (SEL1)

1. Supervised

If all the pixels in the cell are from the same class which has a multivariate normal distribution with mean vector μ and covariance matrix Σ, then the statistic calculated in ECHO to test cell homogeneity is distributed χ^2_{DF} (chi-squared, with DF degrees of freedom), where $DF=(CELW)^2$ x number of channels. For the initial choices of SEL1, χ^2_{DF} values with $\alpha=0.25$, $\alpha=0.001$, and $\alpha=0.0001$ will be used.

2. Non-Supervised (SEL1)

Presently the cell homogeneity is tested by $\frac{s_i}{\mu_i}$

$< T$, where S_i is the cell standard deviation in the i^{th} channel. The use of the coefficient of variation was criticized by users on the grounds that not all scanners have increased noise as the spectral response increases. In addition, the same threshold was used for all four channels, irrespective of the scale of each feature.

For these reasons, the test will be changed to $S_i < T_i$, where T_i can be specified for each channel.

One possible way to compute T_i is to first calculate α_i, the standard deviation in each channel for all the data. If the number of classes desired is n, denote $\alpha_i/\sqrt{n}$ by $\bar{\alpha}_i$. Let T_i equal $C \times \bar{\alpha}_i$ where C is distributed χ^2 with one degree of freedom. Take $C = \chi^2_{1,\leftarrow\alpha=0.5\uparrow}, \chi^2_{1,\leftarrow\alpha=0.1\uparrow}, \chi^2_{1,\leftarrow\alpha=0.3\uparrow}$.

C. Annexation

The optimal cell selection and cell size choices determined in part B will be the basis for the first group of annexation analyses. If the use of annexation does not improve the performance of the ECHO classifier, intermediate results from nonoptimum settings from part B can be chosen and annexation applied to see if annexation can improve nonoptimum choices of cell size and cell selection parameters.

1. Supervised (ANN1)

Initial values chosen for annexation are ANN1= 0, 1, 2. Annexation at levels 1 and 2 will be applied to the intermediate results from the optimum settings of cell size and cell selection. Annexation at level 0 corresponds to the results already obtained for these optimum settings in part B.

2. Non-Supervised (ANN1, ANN2)

Two tests are involved in the annexation of a cell to a field for the unsupervised version of ECHO. One tests the equality of the cell and field means; the other tests the homogeneity of the respective covariance matrices. Both tests involve an F-statistic with degrees of freedom determined by the number of pixels in the cell and in the field. The initial values of ANN1 and ANN2 will be taken from tabulated F distributions with α = 0.1, 0.01, and 0.001.

III. Comparative Experiments

The comparisons among the supervised ECHO classifier, the Non-Supervised ECHO classifier, and the maximum likelihood point classifier will be made using classification results obtained with the optimum parameter values for the ECHO processors determined in the exploratory experiments. The classifiers will be compared over all the data sets chosen for the exploratory experiments.

The classifiers will be compared on the basis of field center pixel accuracy for both training and test fields, average variability per data line, and CPU time divided by area classified. Where available, class proportions of the ECHO results will be compared to proportions obtained from independent sources. The classifiers can be tested for significant differences in these measures by the analysis of variance of a randomized complete block design.

If the training and test fields can be divided into groups of different sizes, the effects of object size and classifier on classification accuracy can be examined. Again, a randomized complete block design would be used with data sets acting as blocks and classifier and object size as factors within blocks.

The eight simulated Thematic Mapper data sets can be analyzed separately to determine the effects of resolution and classifier on classification accuracy, variability in the classification results, and CPU time used. The sites now act as blocks, and the four resolutions and three classifiers are two factors in the analysis of variance.

IV. Additional Tests

If time and resources permit, additional tests may also be run to examine three questions. First, the object map generated by the Non-Supervised ECHO processor may be used to develop new training statistics: the statistics of the objects identified by the Non-Supervised processor may be used for classifier training or the objects identified may be used as training fields. The ECHO classification produced using this new set of statistics will be compared to the ECHO and point classifications produced with the original statistics. These comparisons will indicate the utility of the object map as a source of training data.

A second additional test could be performed with the simulated Thematic Mapper data to examine the effects of increased noise in the MSS data on ECHO classification accuracy and variability in the classification results. Thirdly, the effect on classification accuracy and object map integrity of shifts in the starting line and column should be explored.

2.2 Stratification of Scene Characteristics

2.2a Stratification by Machine Clustering

Effort on this portion of Task 2.2 focused on two activities: consideration of our basic approach to stratification by clustering and correlation-regression analysis of Landsat spectral response and physical factors which may account for differences or changes in spectral response.

Stratification. The first area of effort involves a reconsideration of the approach to partitioning that we are pursuing. As described at the quarterly review at JSC in late November, we are considering some modification of the approach based on results to data. We will provide additional reports during the next quarter as the details of this proposed modification continue to emerge.

The "Blind Site" data set was received in the third and fourth week of November. It is being cataloged and reformatted, and should be available on schedule at the end of February for use in testing stratification procedures as developed.

Physical Factor Analysis. Landsat scene 1689-16382, June 12, 1974, forms the base for one data set to compare spectral response to ancillary variables. The correlations between the Landsat wavelength bands 1 to 4 and the ancillary variables of land use, soil association, yield and two measures of precipitation are shown in Table 1. The highest correlations with spectral response are soil association and precipitation. Later, regression techniques will be used to model the effect of soil and precipitation on spectral response. First, however, some difficulties

Table 1. Correlation of Physical Factors and Spectral Response*

Factor	Band 1	Band 2	Band 3	Band 4
Land Use	-.25	-.37	.15	.27
Soil Assoc.	.56	.53	.52	.43
Precip.-Oct.-June	-.63	.64	.38	-.25
Precip.-March-June	-.57	-.57	-.28	-.16
Yield	-.06	-.08	-.09	-.07

Table 2. Correlation of Physical Factors and Transformed Spectral Response

Factor	Soil Brightness	Green Stuff	Yellow Stuff	None Such
Land Use	-.12	.55	.05	-.35
Soil Assoc.	.59	.02	-.36	-.03
Precip.-Oct.-June	-.58	.25	.38	-.20
Precip.-March-June	-.49	.29	.38	-.27
Yield	-.09	-.01	-.05	.13

* Landsat Scene 1689-16382, June 12, 1974

which arose in the present analysis must be resolved. In analyzing the data, it was discovered that the temperature variables were incorrectly specified. This is being corrected, and temperature will be included in future work.

In addition, the tasselled cap transformation was applied to the spectral response variables so that the relationship of the transformed spectral response and the ancillary variables could be examined. Table 2 shows the correlations between the tasselled cap spectral variables and the ancillary data variables. Again, the correlations are based on 707 segments approximately 3x4 n. miles in size from scene 1689-16382, acquired June 12, 1974. The first tasselled cap feature, soil brightness, is indeed most highly correlated with factors affecting the soils' spectral response: soil association and measures of precipitation. The "green stuff" band is most highly correlated with land use, whose categories relate to the amount and type of vegetation on the ground. As with the previous data set, temperature will be added when it has been correctly coded.

Using a second data set, 98 segment-pass combinations of Landsat data for Kansas LACIE and SRS segments are being analyzed. The same physical factor variables as in the above analysis, plus geographic location (latitude and longitude) and date of acquisition (Julian data and crop maturity stage). In addition, the spectral response of wheat is being added to this regression analysis. This will aid in determining whether the wheat spectral response in a segment is influenced by the same factors as the overall spectral response of the segment.

2.2b Digitization and Registration of Ancillary Data

The ancillary data registration requirement is based on the need to quantitatively relate physical factors to spectral data in the spectral strata research program. During CY76 four key factors were manually digitized and registered with Landsat data for a full frame block in central Kansas. The variables were: soils, land use, temperature and precipitation. Extensive manual interaction was required to complete these operations and the need was observed for automatic methods of conversion of map format data to digital form. Thus a task was defined for CY77 which would explore methods of automatically digitizing certain types of maps containing ancillary data. Three additional tasks were also defined: 1) Digitization of additional variables, 2) Data base coordinate system investigation and 3) Resource evaluation for ancillary data registration.

Five subtasks were to have been active during the quarter as shown in the implementation plan. Task 3 was associated with digitizing additional variables which were to be defined in the first quarter if needed. Tasks 5 and 6 define the advanced map digitization method study. Task 8 is the coordinate system study and Task 9 is the evaluation of resources study. These activities are discussed in the following sections.

Task 3 - Digitization of Selected Data. This task was defined as optional to allow more ancillary variables to be added if needed during this contract year. Completion of details involved in the registration of the four initial variables took place in the first quarter and at that time it was decided not to register any new variables in the second quarter. Topographic data was considered, but due to the low relief in central Kansas it was decided that the cost could not be justified based on the manual methods used last year. Additional versions of county temperature and precipitation variables were added in the first quarter as this was very easy to do once the county boundaries were registered. Further additions to these variables will be made in the third quarter. This task was thus inactive during the quarter and the available resources were directed to the other tasks as these promised greater benefit.

Tasks 5 & 6 - Define and Implement Advanced Map Digitization Methods. The ancillary data digitization methods used in CY76 required extensive and laborious manual use of a table digitizer to convert a map to digital form. Editing and reformatting of the digitized points is an additional task which is tedious and subject to human error. In the case of complicated maps, a lot of detail on the map has to be omitted in order to reduce the burden of editing.

An advanced method is being investigated, in which the table digitizer is replaced by a film scanner. The map must be photo-reduced to a color transparency, which is small enough to be mounted on the drum of the scanner. Digitization is subsequently performed automatically so as to generate red, blue and green digital images of the map; then a classification based on these three color images is used to recover the original color-coded features. This classification is then available for registration with Landsat data. This method is highly automatic and

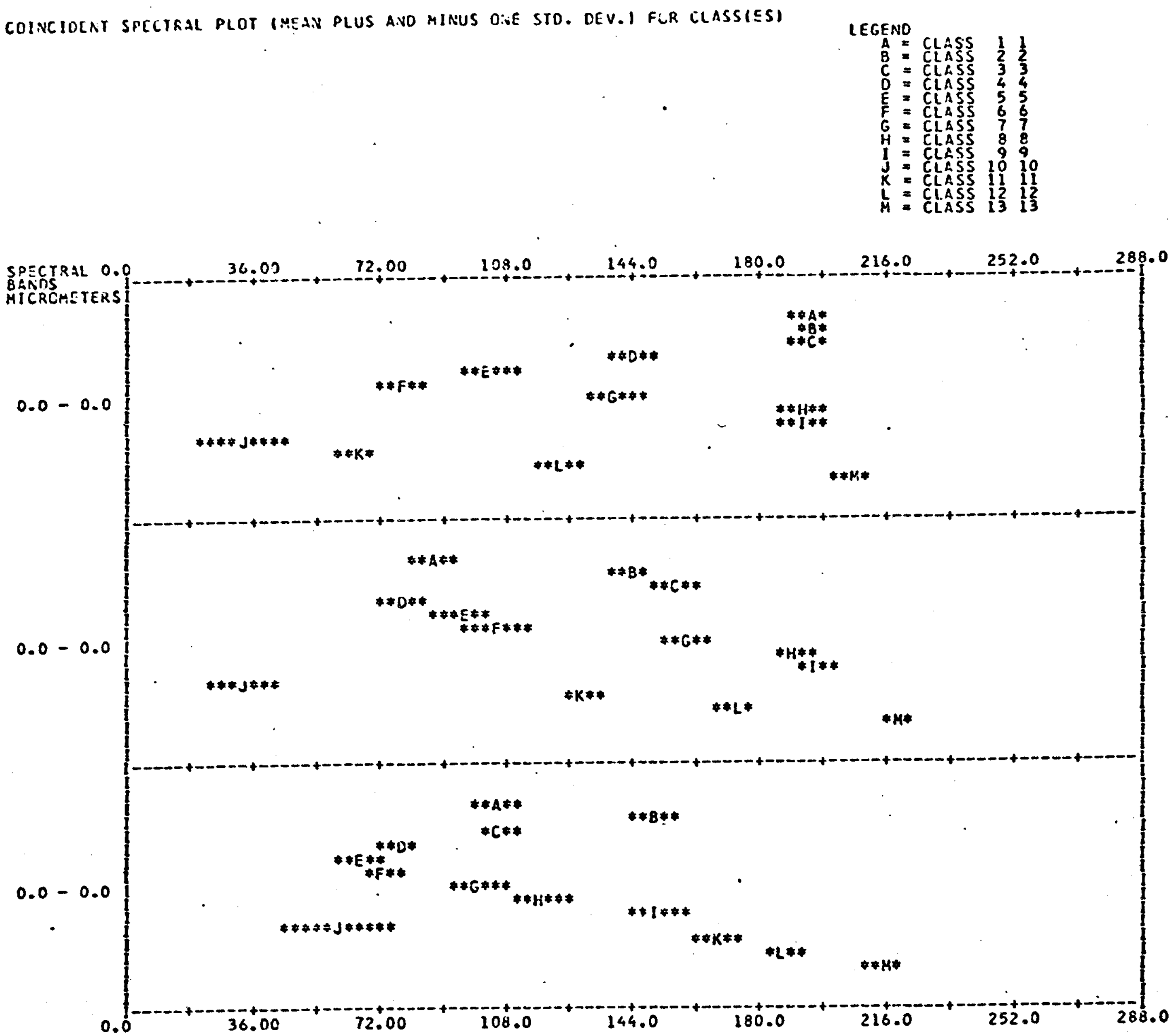


Figure 2: Statistics of Land use map whose map photography involves enlarging a grainy film (Cibrachrome) and photographing under photoflood lamps.

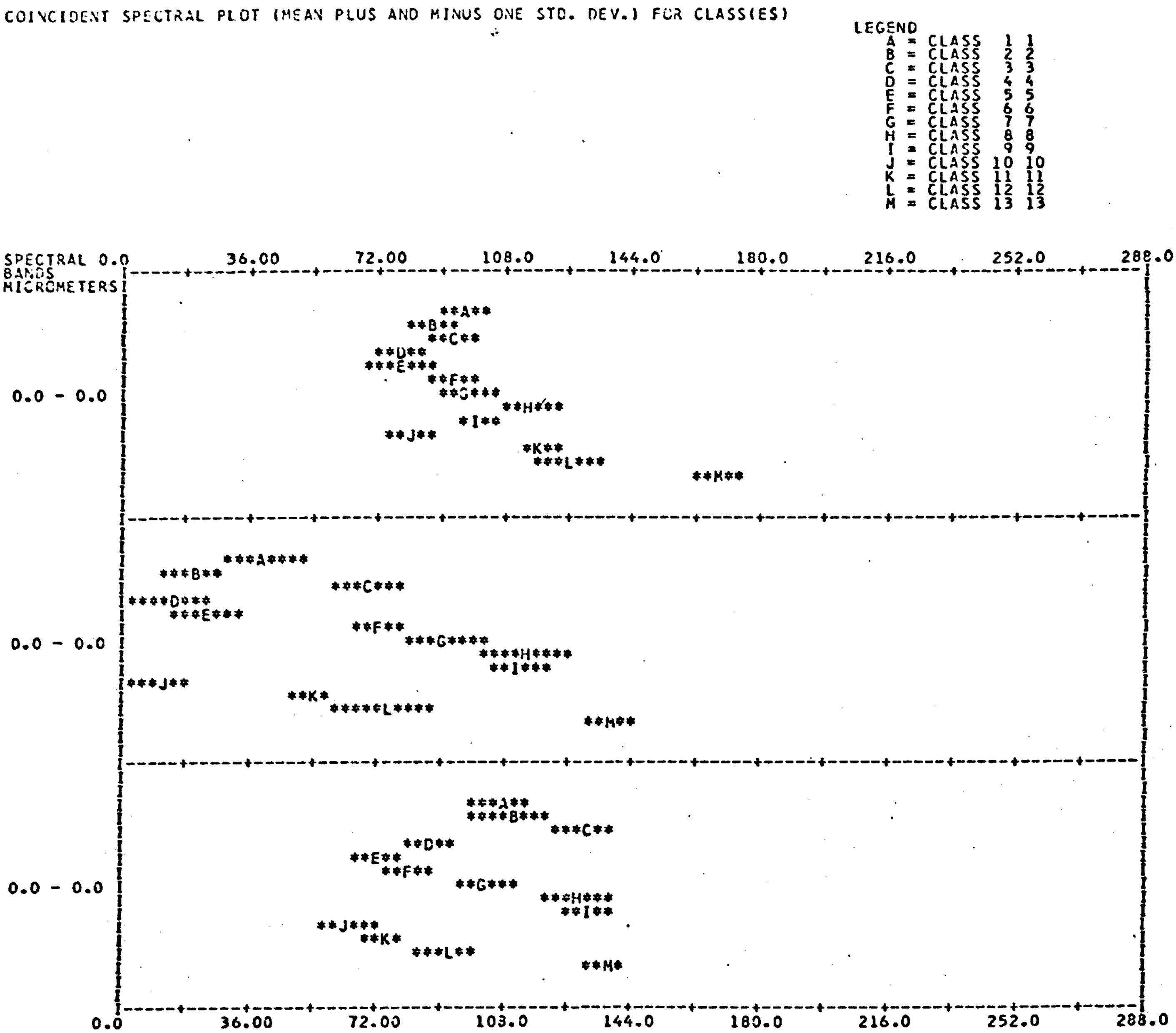


Figure 3: Statistics of Landuse map whose map photography involves using Kodak KM-135-20 film and photographing under bright sunlight.

retains most of the details on the map.

How well the color-coded features on the map can be recovered from its color images depends on how well the classification can be performed. Classification performance is strongly influenced by the quality of the color images, which are ultimately determined by the photo-quality of the transparency. Experimental results show that high photo-quality may be achieved by (1) using high resolution fine grain film such as Kodachrome KM-135-20 and (2) photographing the map under strong uniform sunlight. Figure 2 shows the statistics of the land use map of Kansas which was photographed under photoflood lamps using a relatively grainy film. Figure 2 shows the coincident spectral plot of the same map but photographed with fine grain film under bright sunlight. It is clear that the separation of color types is much better in Figure 3.

Another objective of advanced digitizing method investigation is to retain as much detail of the map as possible. To do this the digitization rate of the scanner has to be increased to bring about high effective spatial resolution. Three factors affecting effective spatial resolution are:

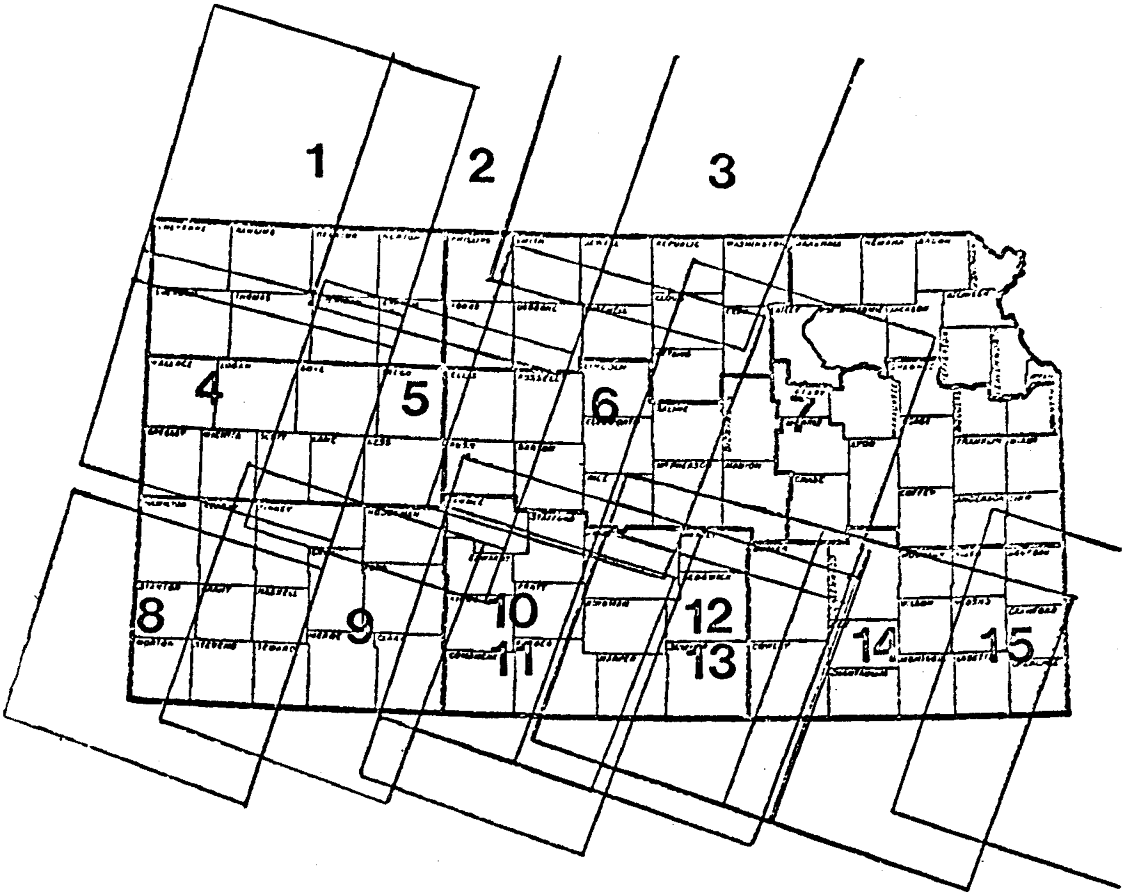
1. aperature size of the scanner and sampling rate
2. ratio of photo-reduction
3. size of the printing dots on the map

Although there is a minimum aperature size (100 μm, 1 μm = 10^{-6} meters) of the scanner used, it is always possible to achieve a desired effective spatial resolution by appropriate change of the photo-reduction ratio. However, spatial resolution cannot be increased without limit beyond the dot density of the map (about 100 dots per inch, with respect to the map). If this happens, the scanner sees one dot or part of a dot at a time and records only the color of the dot, which is unfortunately not the true color representation of the feature on the map. Calculations show that the scanner has to "see" three or more dots in order to record a "reasonably true" color of map features. The variance of the recorded color depends on the number of dots that are averaged.

A reasonable compromise between spatial resolution and variance is achieved as follows: spatial resolution is chosen to be about 30 points per inch with respect to the map, variances within colors to be about 5-10. Since separation between colors is observed to be about 30-50 units, classification accuracy will be expected to be very high. Experimental classification results show visually good recovery of the features on the land use map of Kansas. There is virtually no error in unit centers, but significant misclassifications are seen along field boundaries. This indicates the necessity of post-processing. Post-processing methods will be investigated in the next quarter.

Task 8 - Define Coordinate Systems for Data Base. Work on this task has not progressed as scheduled as clear definition of its intent was not established during the first quarter. Subsequent conversations with JSC during the second quarter disclosed an interest in polygon versus grid cell storage methods and comparison of these methods is planned for the third quarter. Other interest was expressed in terms of geographic versus image system coordinate alternatives and the relationship of these to LACIE segment data files. The most fruitfull comparison appears to be the polygon versus grid cell storage method tradeoff and this will be completed by the end of the contract. The milestone line for this task should be extended to the end of the contract year.

Task 9 - Evaluate Resources for Digitization of Ancillary Data. This task was pursued during the quarter by tabulating manhours, computer costs, digitizer time, and material costs for the manual and automatic methods of map digitization. This process will be continued through the last two quarters and a detailed comparison will be made in the final report at the end of the contract.



Key

	Landsat Scene ID	LARS Run Number	Date
1	2165-16450	75013800	July 6
2	2146-16392	75005800	June 17
3	2163-16334	75006500	July 4
4	2165-16453	75004600	July 6
5	2146-16395	75005900	June 17
6	2163-16340	75006600	July 4
7	2144-16282	75005600	June 15
8	2147-16460	75006200	June 18
9	5032-16310	75007200	May 21
10	2073-16342	75001500	April 15
11	2109-16341	75005000	May 11
12	2072-16284	75000900	April 9
13	2144-16284	75005700	June 15
14	2107-16225	75004900	May 9
15	2142-16171	75005400	June 13

Figure 1: Full-Frame Landsat Classifications of Kansas.

2.2c Crop Inventory Using Full-Frame Classification

Task 1 - Classify Full-Frames. During this quarter, the full-frame classifications of Kansas were completed. The classifications were finished by October 1, as stated in the last quarterly report. These classifications were done on a county basis using every other line and every other column of data. Eighty counties in seven crop reporting districts of Kansas are included. The Landsat scene ID, LARS run number, and date of acquisition for the Landsat frames used in these classifications are given in Figure 1.

Task 2 - Develop Sample Plans. The design of the sampling plans was completed by November 1. The experiment contains two sequential parts.

In part one, the total area covered by the sample is held constant, and the segment size and number varied. Four segment sizes will be used:

Plan	Segment Size	No. of Segments
1	5 x 6 N. miles	84
2	4 x 4	157
3	2 x 2	630
4	pixel	1,926,288

The sampling errors for these plans will be calculated and an analysis of variance will be used to determine if there is a significant difference in sampling error among these plans. The pixel sampling plan is included for comparison.

In the second part, the sampling error will be held constant at the level achieved by using 84 5x6 n. mile segments and the number of segments of each size needed to achieve this error will be calculated. Then samples will be taken to test that the calculated sampling error is correct.

<u>Task 3 - Locate Sample Segments and Compile Classification Results</u>. LACIE procedures for determining the allocation (number) and location (geographic placement) of segments will be followed when possible. However, the method described in the LACIE documentation (LACIE, Volume IV, pp. B-2 to B-3) does not give a consistent system for determining the allocation of sample segments. Since it was not known exactly how LACIE made these decisions, the following method will be used in this project.

The number of segments allocated per county will be used when there is an inconsistency with the CRD number of segments. If this in turn causes the state total to be too high or too low, then the number of segments for all the counties in crop reporting districts which were inconsistent are considered together. One segment from the county which is closest to one of the cut-off points for rounding to the next number of segments will be added or removed. This procedure will be continued until the state total is correct.

At the present time, a program is being written to carry out the sample selection. This will allow a greater number of samples to be taken and give the tests of differences greater validity. It will also allow additional sampling plans to be evaluated (as suggested by JSC/EOD at the SR&T quarterly review in November). Programming should be completed by mid-February. Approximately a month will be required to compile and summarize the classification results and sampling errors, after which the statistical analysis can be performed.

2.3 Field Measurements for Remote Sensing of Wheat

Field measurements activities during this quarter have included: preparation of project plan, data processing and evaluation, management of data library, data analysis, and project leadership and coordination.

Project Leadership and Coordination. The 1976-1977 project plan was prepared and distributed for review to personnel at NASA/JSC, ERIM, and Texas A&M University. A report of field measurements activities and progress was made at the SR&T quarterly review at JSC in late November. During November, two representatives from LARS reviewed the FSS, FSAS, and meteorological data acquisition procedures used by NASA/JSC in Finney County, Kansas. The observations from this review were reported to NASA/JSC. Activities have also included continued communication with personnel at JSC and other centers concerning the project.

Data Processing and Evaluation. The processing and evaluation of the 1974-75 FSS data in the field averaged and individual scan formats were completed. Seven dates of the 1975-76 FSS data were processed. The rest of the ground truth information to be included with the 1975-76 FSS data are being keypunched. The three dates of the FSAS data received from JSC during this quarter are being processed.

Half of the 1976 Exotech 20C data has been digitized. An algorithm to correct the data for the tape recorder WOW present in the data was formulated and implemented. The correction is now being evaluated. Half of the keypunching of the ground truth information for the Exotech 20C data was completed.

All LANDSAT data over the test sites that were received during this quarter have been processed. Registration of the LANDSAT data to form multitemporal data sets over each of the sites is continuing. Nine dates of aircraft multispectral scanner data were converted from universal format to LARSYS format in the continuing effort to condense the large volume of this data and present it in a more usable form to analysts.

Data Library. Computer tapes containing all of the 1974-75 FSS (field averaged format), FSAS, Exotech 20D, and Exotech 20C spectrometer data were distributed to researchers at ERIM and Texas A&M University. All 1975-76 and 1976-77 data received during the quarter have been cataloged.

Suggestions received from the review of Version 2 of the Field Measurements Data Library Catalog were incorporated into the present format of the catalog. The present format consists of separate volumes - one for each crop year of data collection. Volume 1, 1974-75 crop, has been distributed. Preliminary versions of Volume 2 and Volume 3, 1975-76 crop and 1976-77 crop, respectively, have been printed. Procedures for "outside" users to follow to obtain the Field Measurements Data were outlined during communication between JSC and LARS.

Data Analysis. Analysis of the truck spectrometer data acquired during the 1974-75 at the Garden City, Kansas and Williston, North Dakota agriculture experiment station was begun during the quarter. The analysis data on the use of a laser probe to characterize the geometry of crop canopies was completed; a detailed report of this work will be submitted during the next quarter.

Plans for Next Quarter. Processing of the 1975-76 FSS, Exotech 20C and aircraft MSS data will continue. The FSAS data will be processed as it arrives. Volumes 2 and 3 of the Field Measurements Data Library Catalog will be distributed. Corrections or additions to the project plan suggested during the Field Measurements meeting at Houston in December will be incorporated into the plan. During this quarter analysis of the 1974-75 FSS data will begin.

2.4 Scanner System Parameter Study

Background

The scanner parameter study was initiated in CY76 to conduct research on analytical methods of multispectral sensor system design. It was planned that basic information theoretic approaches would be taken to determine optimum performance levels achievable in a given environment against which to test performance of actual or modeled systems. Scanner system modeling techniques were to be developed which would permit explicit evaluation of scanner performance without the use of simulation techniques. The study includes representation of the scene and information extraction process to provide models of the environment in which the scanner would operate. Thus three main activity areas were defined (scene, scanner, and information extraction modeling) and research was begun in CY76.

Shortly into CY76, the direction of the study was changed on request of the sponsor. A conventional simulation study of certain proposed thematic mapper parameters was requested using NASA 24 channel scanner data as input. This activity required almost all the project resources through the end of CY76 and the results were reported in the June 1976 final report and in Information Note 110976. Research on the analytical model program was resumed in CY77. Progress has been made in the three basic areas defined for the study.

With respect to the implementation plan four tasks were active during the quarter. Spectral data base development (Task 2) continued with field spectrometer data being processed, checked and catalogued. The core of the study is Task 4, in which the scene spectra are analyzed with the goal of defining spectral sampling functions to optimize sensor performance. Task 5 addresses the problem of information extraction and its goal is to develop analytical methods of classification error prediction. Task 7 addresses the other parameters to be modeled (e.g., spatial sampling function, signal-to-noise ratio) and is the activity which will integrate the elements of the study into a set of programs and procedures for scanner system analysis. Progress on these tasks is discussed in the following section.

Task 2. Spectral Data Base Development

The activities included in this task relate to the acquisition, reformatting, preprocessing, error checking, cataloging and averaging of field spectrometer data for use in defining the scene model. A large number of spectra are needed for characterizing example classes. Work continued through the quarter to carry out preprocessing of an adequate number of field spectra to support the scanner study. Data collected from Finney County, Kansas; Hand County, South Dakota; and Williams County, North Dakota, is being selected.

A sufficient number of spectra has been processed by the scanner study, and Task 4 has made use of these spectra for defining statistics for example classes for algorithm test purposes. It is expected that allocation of resources to this task will be finished as scheduled on December 31, 1976, and work begun doing statistical analysis of these spectra as scheduled in December of 1976. The major problem foreseen in the spectral data base area is that a limited number of classes is available, i.e., in the wheat crop scene and associated field types. Forest, geological, hydrological, urban scenes are not available. This is considered acceptable for the early phases of the study where basic research is being conducted, but a greater diversity of scenes should be acquired for model testing.

Task 4. Optimum Basis Function Study

Two approaches to the basis function (or spectral band function) specification are being pursued. The first was discussed in the first quarterly report and involves the use of the Karhunen-Loéve (K-L) representation of an ensemble of spectra from a set of classes of interest. The second approach utilizes an information theoretic approach which employs the concept of mutual information to select bands for maximum separation of classes. Work in these two areas is discussed next.

K-L Representation

The theory of K-L expansions is well established and we are applying it here to the case of an ensemble of reflectance/emission spectra from selected classes of areas of the earth's surface. The K-L representation in the case considered here is based on a set of orthogonal basis functions, which are derived from the covariance matrix of the discrete process which is being modeled. Assume the spectral reflectance/emission function $f(\lambda)$ (λ typically in the range of .4 to 14 micrometers) is represented in discrete form as $f(\lambda_i)$ $i=1, N$ and let Λ be the NxN covariance matrix over the N band samples for an ensemble of $f(\lambda_i)$ over the classes of interest.

Then the rows of the matrix T such that $T\Lambda T^T = \Gamma$ are the orthogonal basis functions where Γ is a diagonal matrix of the eigenvalues of Λ. The transformation Y=TX then is used to generate uncorrelated, ordered variance components which reflect the true dimensionality of the original $f(\lambda_i)$ process. The advantage of this representation is that it removes correlation (i.e., redundancy) from an N dimensional vector process and produces a sequence of new vector components which have monotone decreasing variance. The basis functions indicate how the original N variables should be combined to produce the uncorrelated components if indeed there is correlation among the $f(\lambda_i)$.

The work performed during the quarter included programming of algorithms for calculating the K-L representation using field spectrometer data. Dimensionality is a problem here as there are over 1000 samples available from the source instrument over the range of .4 to 14 micrometers. Subsets on the order of 100 samples are being processed in current software. Performance evaluation was pursued through consideration of distance measures such as Bhattacharyya distance and divergence which relate to probability of classification error. Various combinations of the original N samples were selected, numbers of components and different class sets were processed. Some of the specific parameters for the current study are classes: corn, wheat, soybeans in June and July 1974; number of field spectra for each class: 10; input samples: 100; number of principal components being used: 20 to 30. Theoretical consideration of methods of performance evaluation will be made through the next quarter to provide a method for selecting the set of components (basis function).

Information Theoretic Approach

A second analytical technique for spectral band selection for multispectral scanners is being pursued. Given that k classes C_j, $j=1,\ldots,k$ are observed, then M of a possible N spectral bands are to be chosen to provide the maximum amount of information for classification of the k classes.

One criterion for the above problem is the average mutual information between the classes and the chosen spectral bands. This criterion may be expressed in the following form:

$$I(c,\underline{x}) = H(c) - H(c/\underline{x}), \underline{x} = (x_1,\ldots,x_m)$$
$$c = (c_1,\ldots,c_k)$$

where

H(c) = Entropy or a-priori uncertainty of the classes.

or

$$H(c) = \sum_{j=1}^{k} p(c_j) \log p(c_j)$$

where $p(c_j)$ is the a-priori probability of the $j^{\underline{th}}$ class.

$H(c/\underline{x})$ = equivocation or uncertainty about the classes after observing spectral bands $\underline{x}$.

or

$$H(c/\underline{x}) = -\sum_{j=1}^{k} \int_{x_1} \cdots \int_{x_m} p(c_j \underline{x}) \log p(c_j/\underline{x}) \, dx_1 \ldots dx_m$$

Thus the problem is to select m spectral bands to maximize $I(c,\underline{x})$.

It can be shown that

$$o \leq I(c,\underline{x}) \leq H(c)$$

Hence an equivalent problem is the minimization of the equivocation $H(c/\underline{x})$.

The criterion given above can be used to estimate the classification performance in terms of upper and lower bounds on the probability of classification error P_e.

Upper bound: $P_e \leq \frac{1}{2} H(c/\underline{x})$

Lower bound (Fano bound):

$$H(c/\underline{x}) \leq - P_e \log P_e - (1-P_e) \log (1-P_e) + P_e \log (k-1)$$

It has been shown that tighter bounds may be given in terms of "Bayesian distance" $B(c/\underline{x})$ [2].

$$\frac{k-1}{k}\left[1 - \sqrt{\frac{kB(c/x)-1}{k-1}}\right] \leq P_e \leq 1 - B(c/\underline{x})$$

where

$$B(c/\underline{x}) = E_{\underline{x}}\left[\sum_{j=1}^{k} p(c_j/\underline{x})^2\right]$$

$$= \sum_{j=1}^{k} \int_{x_1} \cdots \int_{x_m} p(\underline{x}) \quad p(c_j/\underline{x})^{\,2} \, dx_1 \ldots dx_m$$

The next consideration is actually finding the optimum M out of N spectral bands. The most obvious technique would be an exhaustive search for the M spectral bands that minimize $H(c/\underline{x})$. However, for this technique, the computational load increases very rapidly. For example, finding the best 12 spectral bands out of 24 possible spectral bands gives 2,704,156 combinations to be evaluated.

A technique that gives a reduction in computational load is the branch and bound algorithm [3]. This algorithm is currently under evaluation for the selection of the optimal spectral bands.

Task 6. Explicit Calculation of Classification Error

Computation of the probability of error of any classifier depends on determination of pertinent regions over which the proper density function should be integrated.

Gaussian Bayes classifiers generate quadratic boundaries. Particularly, in N dimensions, the boundaries of interest are the intersection of hyperquadratrics. This is evident from the discriminant functions, i.e.,

$$W_i = (X-M_i)^T \Sigma_i^{-1} (X-M_i) + \ln|\Sigma_i| - 2\ln P(\omega_i) \qquad (1)$$

where M_i and Σ_i, respectively are the mean and the covariance matrix for class i and $P(\omega_i)$ is the a-priori class probability. The decision rule is then choose ω_i, class i if $W_i < W_j$ for all $j \neq i$ $j=1,2,...M$, $j \neq i$, where we have assumed the multivariate normal distribution for each class. Then

$$P(e) = \sum_{i=1}^{M} P(\omega_i)P(e|\omega_i) \tag{2}$$

is the associated probability of error.

Continuous Approach:

The set of inequalities mentioned in the decision-making step defines a hypervolume over which the proper density function must be integrated. So the problem may be viewed as an evaluation of a multiple integral over a given domain. Explicit solutions to multi-dimensional integration are not generally available, and results are known only for fairly simple functions and domains. In our case, formulating the limits of the integral in continuous domain would be quite impractical, if not impossible.

Discrete Approach:

Using discrete random variables, instead of continuous, simplifies the problem, but methods must be developed to account for the error caused by this conversion.

Choice of Discrete Model:

We would like to choose a variate such that, in the limit, it approaches a normal random variable, therefore assuring convergence to the true quantity by decreasing the increments. A normalized binomial random variable approaches a normal random variable as $m \to \infty$, i.e.,

$$Z = \frac{X - mp}{\sqrt{mpq}} \qquad q = (1-p)$$

where X is a binomial random variable with parameters m and p.

The convergence is fastest for p=0.5, thus

$$Z_m = \frac{(X - m/2)\,2}{\sqrt{m}}$$

$$\lim_{m \to \infty} Z_m \sim N(0,1)$$

As m is increased, the distance between two consecutive values of Z_m decreases:

$$\Delta = Z_m\big|_{x=0} - Z_m\big|_{x=1} = \frac{(m/2)^2}{\sqrt{m}} - \frac{(1-m/2)^2}{\sqrt{m}} = \frac{2}{\sqrt{m}} \qquad (3)$$

therefore, as $m \to \infty$ $\Delta \to 0$ and Z approaches a continuous normal random variable.

Application to Calculation of Probability of Error:

As mentioned before, the probability of error is essentially a definite multiple integral over a specified region. When M discriminant functions are available, selection of the $i^{\underline{th}}$ class is based on the decision rule defined under Eq.(1). This inequality defines a region over which the class i density function must be integrated. But now we have a grid of points instead of a continuous variate.

Example:

Assume two classes are characterized by $N(\underline{0},\underline{I})$ and $N(\underline{1},\underline{I})$.

$$W_1 = X_1^2 + X_2^2$$

$$W_2 = (X_1-1)^2 + (X_2-1)^2$$

given class 1, an error is made if $W_1 > W_2$.

The continuous approach requires that precise functional forms of the boundaries be known, which means, usually, solving a set of non-linear equations.

In the discrete model, we test every sample point in the inequality, which rapidly determines whether that point is inside or outside. After all sample points are exhausted, we have divided the grid into two subspaces, inside and outside. With fine quantization we can very closely indicate where the boundary lies, although no explicit mathematical description of it is available. Programming of algorithms to test this approach progressed during the quarter.

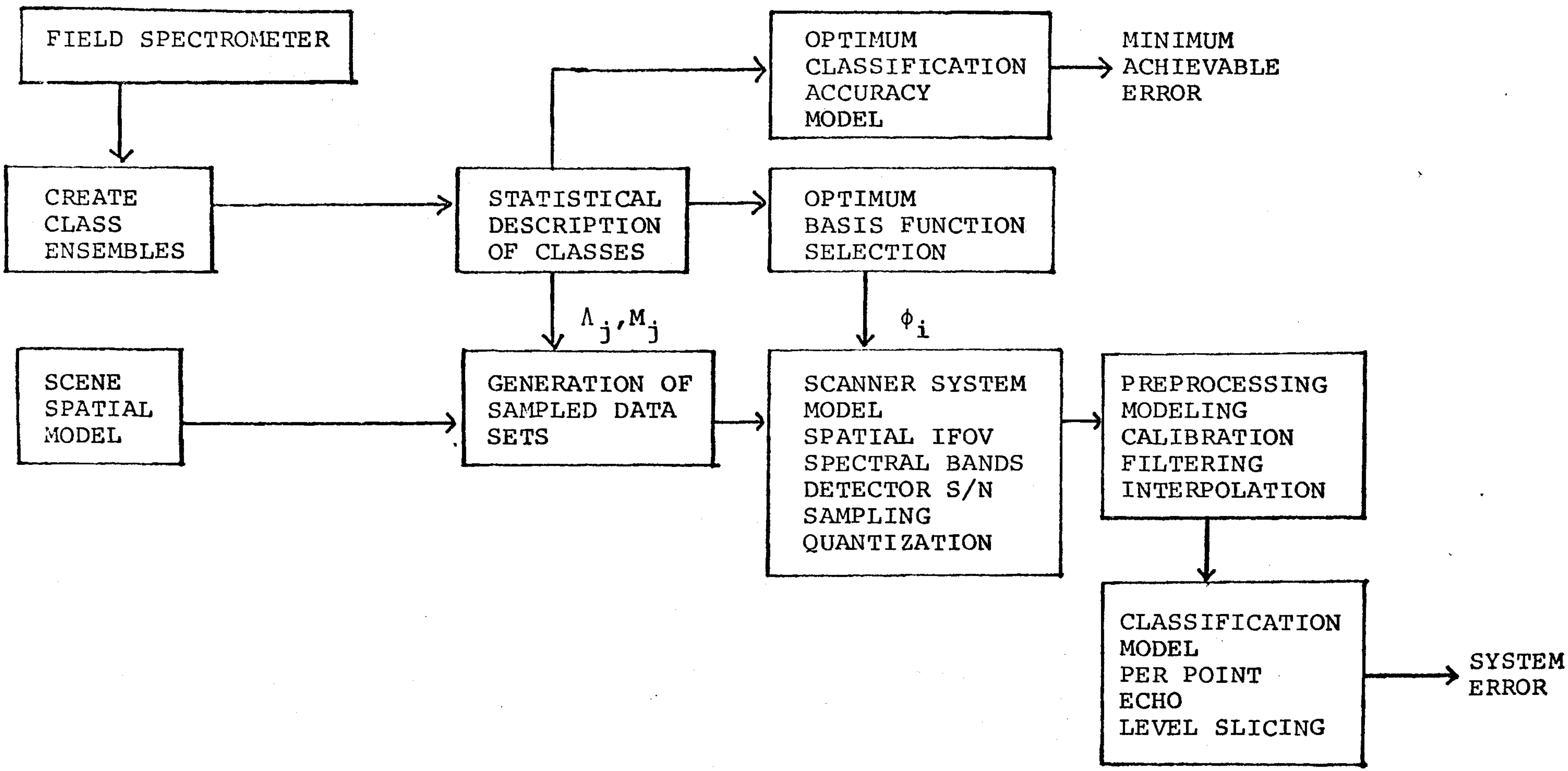
Task 7. Multispectral Scanner Model

The scanner parameter study includes consideration of IFOV (Instantaneous Field of View) effects, signal-to-noise ratio, sampling and quantization effects in addition to the spectral band functions. The scanner model task encompasses the generation of models for these other parameters and the combination of all the component models into a system for predicting scanner performance. The work for the quarter was directed only to the spatial sampling (IFOV) parameter. Methods were considered for modeling the spatial characteristics of a scene and representing the effect of an IFOV function on the spectral data from the scene. The goal here is to generate a statistical representation for various scene types which will enable evaluation of IFOV functions without having to use convolution of real imagery with the candidate functions.

Figure 1 describes the current scanner modeling plan although this is a dynamic structure and is continually modified. Should analytical classification error estimation methods be successful, the classification model block would be replaced by these techniques. This is an ongoing task and is expected to continue throughout the study year.

FIGURE 1.

ANALYTICAL MODEL OF MULTISPECTRAL SCANNER SYSTEM



References

1. C. E. Shannon, The Mathematical Theory of Communication, University of Illinois Press, Urbana, Illinois, 1949.

2. P. A. Devijver, "On a New Class of Bounds on Bayes Risk in Multihypothesis Pattern Recognition", IEEE Trans. Computers, Vol. C-23, No. 1, January 1974, pp. 70-80.

3. P. M. Narendra, "Combinational Problems in Pattern Recognition", Ph.D. Thesis, Purdue University, 1975.

2.5 Transfer of Computer Image Analysis Techniques

Major Activities. This task has been organized into four subtasks as described in the implementation plan. The implementation plan schedule shown on the next page summarizes these subtasks and the major activities under each of these subtasks are described below.

Support of the JSC Remote Terminal. Support of the JSC/LARS remote terminal includes maintaining at the LARS computer facility the necessary hardware and software, providing computer services, providing hardware/software system consultation, and developing concepts for an improved remote terminal system.

Hardware support consists of maintaining communication modems at LARS and a port on the 3705 communication system. Software support includes programming the 3705 to recognize the JSC terminal, CP-67 and LARSYS software and maintaining a virtual machine dedicated to the JSC terminal.

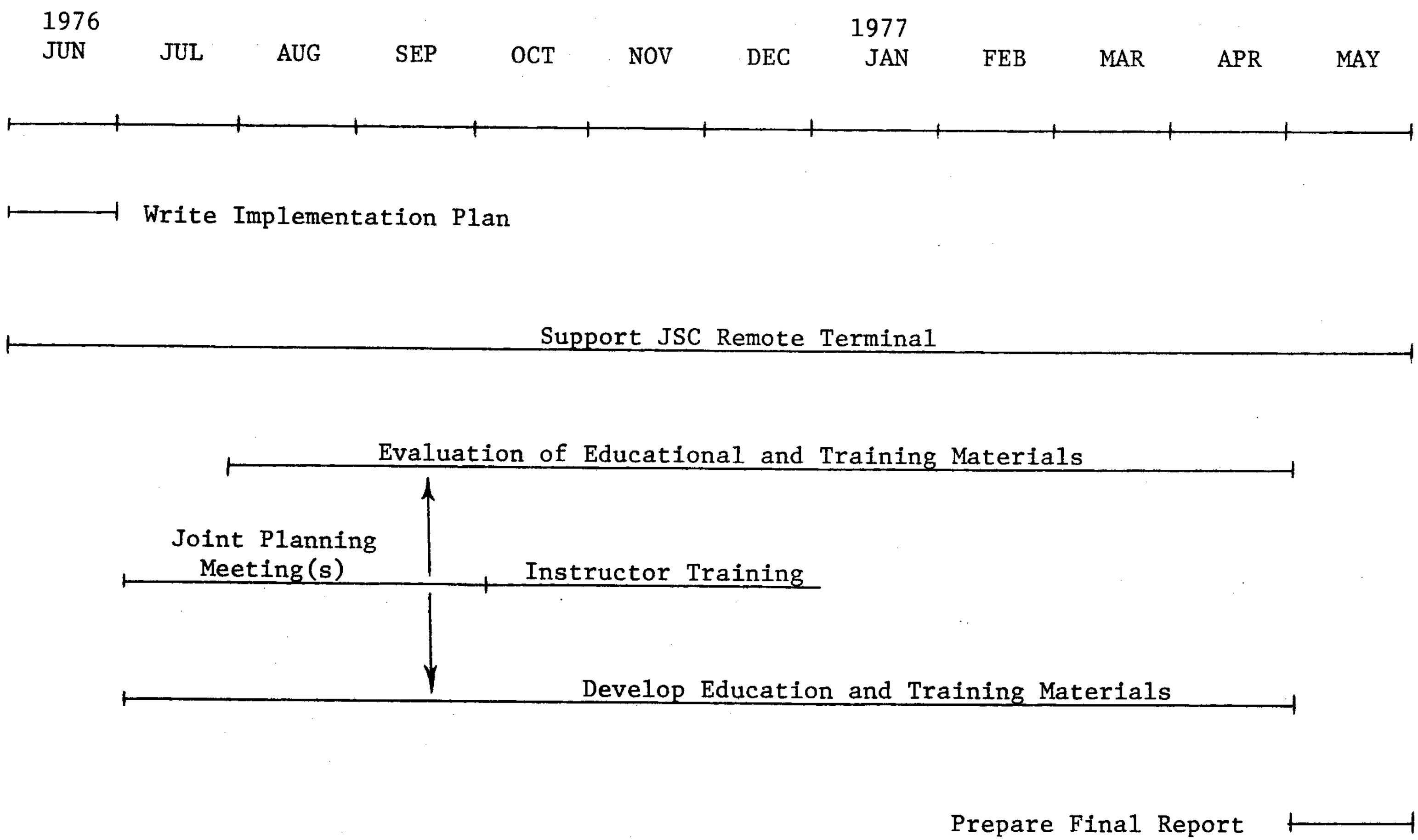
Computer services provided to JSC for September, October, and November are summarized below.

Month	CPU Hours Used	Hours Terminals Attached
September	9.5 hours	374 hours
October	6.8 hours	239 hours
November	6.0 hours	264 hours

An improved remote terminal configuration is being studied. The improved terminal is based around a minicomputer supported by an array of special and general purpose hardware. The concept calls for communication lines connecting the terminal to a high capability

Implementation Schedule

2.5 TRANSFER OF COMPUTER IMAGE ANALYSIS TECHNIQUES



computer which would be logically subordinate to the terminal. During the past quarter the installation of prototype hardware has begun. Testing of the concept is ready to proceed but as a result of the November 22 and 23, 1976 discussions at JSC the priority of testing the concept has been lowered in order to raise the priority of looking into JSC's need to have an improved capability for transmitting data over the remote terminal link.

Evaluation of Education and Training Materials. The revised format of the simulation exercise "Determining Land Use Patterns Through Man-Machine Analysis of LANDSAT Data" was read and critiqued by three individuals representative of the intended student audience and two LARS faculty members who had not worked on the RAP project (Drs. Baumgardner and Hoffer). Their comments were evaluated by the authors, the material was finalized and is presently in press. This document will be issued as LARS Information Note 070676.

Joint Planning Meeting and Instructor Training. A conference telephone call between D. Hay, M. Trichel of NASA/JSC, and J. Lindenlaub of Purdue/LARS was held on October 8, 1976. Instructor training and JSC evaluation of LARS produced education and training materials was discussed briefly. Training of JSC personnel was again discussed at JSC with J. Erickson, M. Trichel and C. Davis on November 22, 1976. At this meeting it was pointed out that there was not a pressing need to familiarize a cadre of JSC people with available education and training materials. As a result of these discussions it was decided not to pursue instructor training activities at this time. Should a need arise, LARS is prepared to carry out such training activities.

Development of Education and Training Materials. a) LARSYS Educational Package. During the past quarter work was begun on modifying Unit III, Demonstration of LARSYS on the 2780 Remote Terminal, and Unit IV, the 2780 Remote Terminal: A "Hands-On Experience" of the LARSYS Educational Package to accommodate a Data 100 Printer-Card Punch-Card Reader Unit. Modification of these materials was motivated by two considerations: 1) LARS has recently installed such terminals and the availability of training materials enhances LARS' ability to efficiently adapt to these hardware changes thus minimizing the temporary productivity loss usually associated with computer system changes--i.e. the productivity rate on other SR&T tasks does not suffer; 2) preparation of these materials is a good preliminary step towards the development of training materials to support the intelligent terminal concept described above.

The modification of these materials is being done in such a way as to accommodate different keyboard terminals as well. Currently, four remote terminal configurations can be accommodated by the LARS computer facility:

1. IBM 2780 with 2741 typewriter terminal
2. IBM 2780 with CRT typewriter terminal
3. DATA 100 with 2741 typewriter terminal
4. DATA 100 with CRT typewriter terminal

The present Educational Package materials were designed for configuration 1. Modifications of Units III and IV have been drafted and tested with several "students" for configuration 3. These will be revised and modified to accommodate configurations 2 and 4 as well.

b) FOCUS Series. Other materials completed during the past quarter include three additions to the FOCUS Series. Each FOCUS is a two-page foldout consisting of a diagram or photograph and an extended caption of three to four hundred words treating a single concept. Their titles and brief descriptions are listed below.

Multispectral-Multitemporal Concept

John C. Lindenlaub

Describes one of the advantages of recording remote sensing data in numerical form as the convenience of merging data from different times to create a multispectral-multitemporal data set.

LARSYS Version 3.1

C. Royal Sand

The LARSYS concept is reviewed and the interrelation of the eighteen processing functions included in LARSYS Version 3.1 are described.

Regional Land Use Inventories

Shirley M. Davis

Procedures used to perform a land use inventory of the United States portion of the Great Lakes Basin derived from LANDSAT data using pattern recognition techniques is described. Examples of tabular and map-like output products are shown.

Three FOCUS titles are in the final review stages: Their titles are Reformatting, Multiband Concept and Snow Cover Mapping. The completion of these additional titles will complete this phase of FOCUS materials development.

c) Advanced Analysis Procedures. As new analysis procedures are developed it is desirable to prepare instructional materials for their use. During the past quarter the initial steps toward the development of case studies utilizing new analysis procedures have been taken. The background literature for the ECHO classifier and the Layered classifier has been reviewed and the format of previous case studies has been examined. Based on these activities a parallel effort will be made during the next quarter toward the development of an ECHO classifier case study. One thrust will be to gain a better understanding of the algorithm, guide lines for analyst selected parameters and analysis situations (average number of picture elements per object, average object size, etc.) in which the classifier works best. This part of the case study development will be coordinated with personnel working on the ECHO classifier task (2.1). The second thrust will be in redesigning the instructional format so that the case study is not dependent upon the use of a 2780-type remote terminal. Essentially a batch processing environment will be assumed. This will give the case study wider utility and suits the current JSC training environment better.

By developing the technical content and instructional strategy in parallel we are able to reduce the time between the introduction of a new analysis procedure and the availability of training materials.

Plans for Next Quarter. Support of the JSC remote terminal will continue as planned. The possibility of up-grading the JSC remote terminal to include a tape drive was discussed at JSC on November 22, 1976 in conjunction with the LARS SR&T review. This possibility will be investigated and a preliminary report listing options will be prepared.

The only evaluation of education and training materials during the next quarter will be that carried out in conjunction with training materials under development.

As explained above, no instructor training activities are planned for the next quarter.

Work on the development of training materials will concentrate on a case study for the ECHO classifier.

2.6 Large Area Crop Inventory Design

Major Activities. Several planning and discussion sessions by the staff were held to identify the 1) approach to be taken in developing plans for multicrop system, 2) elements of the LACIE system which would be applicable, and 3) the "state of the art" in relation to a multicrop inventory system.

A recommendation of a revised implementation plan was received from JSC in October and studied. A revised plan based the initial LARS implementation plan and the JSC recommendations were presented at the SR&T quarterly review in November. We are now pursuing this revised plan.

The major elements of the plan are:

- review LACIE design and results to assess adequacy for multicrop applications,
- develop plan, including recommendations for 1) areas and crops, 2) setting system performance goals, 3) research requirements, and 4) data requirements,
- present strawman plan to NASA and review committee,
- assemble final plan.

Our initial technical recommendations include: 1) inventory of all major crops for important crop production regions (as opposed to surveys of individual crops), 2) criteria for selection of countries and crops should include economic, humanitarian, national value, international value, and probability of success. A key question is the identification of the time frame for phasing in a multicrop system relative to the capabilities of spaceborne sensors.

2.7 Forestry Applications Project

In pursuit of fulfilling the objective to:

Develop inventory methods using remote sensing technology...

our activities during the previous quarter have been focused on reviewing pertainent literature. A considerable part of this effort has been directed toward the 1975 Renewable Resource Assessment prepared by the U.S. Forest Service. Other literature has included: forest mensuration and statistics texts, symposia proceedings, UN-FAO inventory manuals, and material available through the National Technical Information Service. The results of this search, which is expected to be an on-going activity throughout the project, are summarized below.

Inventories are conducted to itemize, catalog, or list assets or stocks. In this situation, assets refer to those attributes of renewable natural resources which can be categorized under physical or socio-economic parameters. Reviewing the requirements of the Resources Planning Act (RPA) of 1974 and the first Assessment in 1975 a list of measures which are amenible to remote sensing can be developed. Table 1 presents a general summary of the 1975 Assessment. The four columns represent: the resource area, a general classification as to units of measure used to assess that resource, the probability of those units being measured by remote sensing techniques and the estimated benefit of using remote sensing to the decision process.

Given, the information from Table 1 we have identified two resource systems where remote sensing inventories are expected to contribute the most. There are, timber and range. An important departure point for

Table 1. Summary Review of Resource Planning Act Assessment (1).

System	Unit of Measure (2)	Measurable by R. S. (3)	Beneficient to Decision (4)
Timber			
Type	Area	+	D
Ownership	Area	0	I
Productivity	Volume	0	I
Range			
Type	Area	+	D
Ownership	Area	0	I
Productivity	Volume	0	I
Water			
Type-Impounded	Area	+	D
Flowing	Area	0	I
Use	Volume	-	I
Fish and Wildlife			
Habitat Type	Area	+	I
Population Estimates	Census	-	I
Recreation and Wilderness			
Participation	Census	-	I
Human and Community Development	Economic	-	N

(1) ref: The Nation's Renewable Resources-An Assessment, 1975, USDA-Forest Service.

(2) General classification of units of measure reported in (1).

(3) Measurable by remote sensing techniques.
Includes all measures possible from satellite and/or aircraft systems.
+ Directly measurable from satellite or aircraft data alone.
0 Measures can be inferred with the support of ancillary type data.
- Measure which is not even indirectly affected by satellite or aircraft inputs.

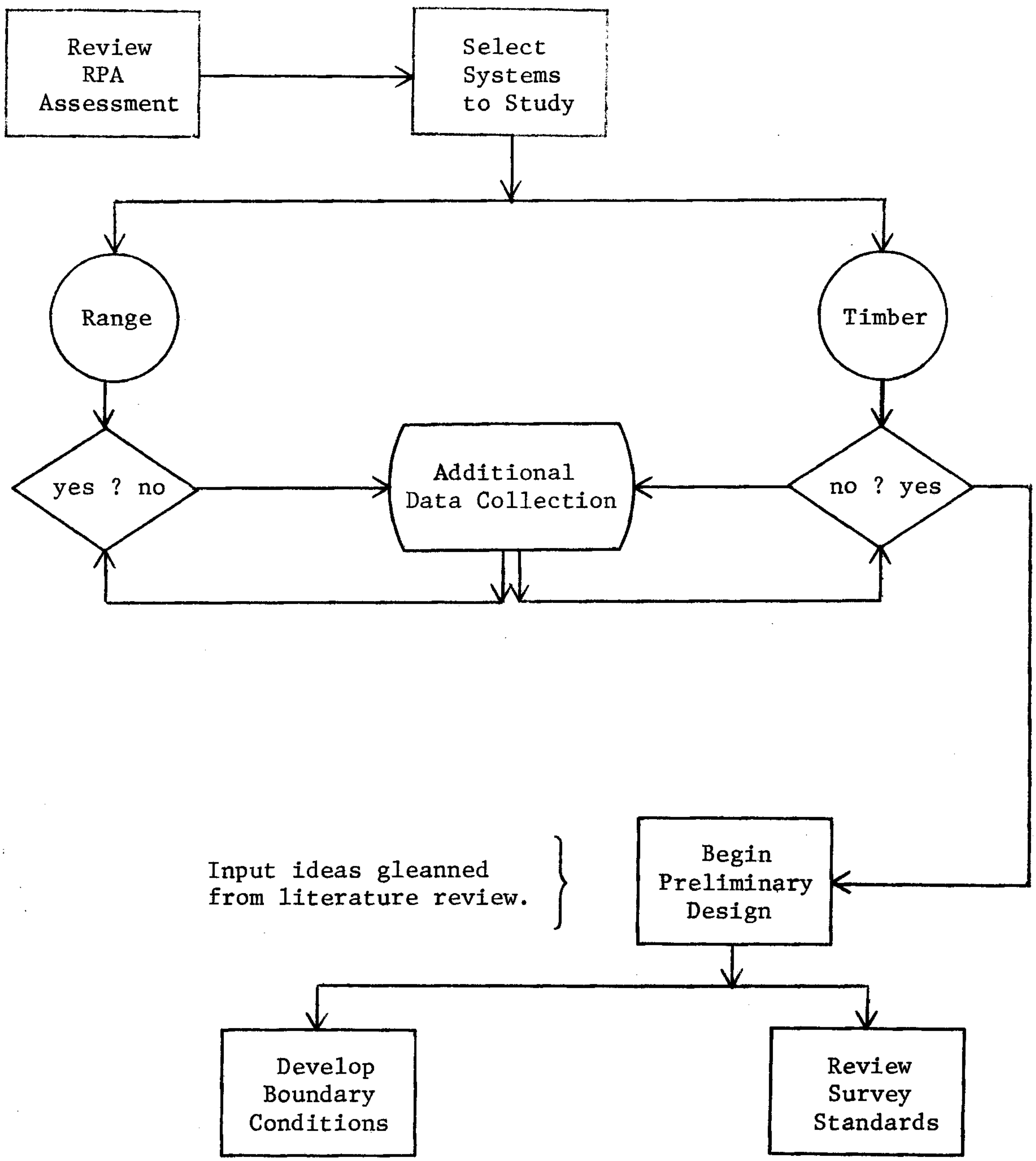
(4) Benefit of Remote Sensing inputs to Decision Process.
Inferred based on inputs.
D - Direct not utilizing other inputs.
I - Indirect utilizing other inputs.
N - Not Obvious.

any timber or rangeland inventory is knowledge of the total areal extent of land base occupied by these systems. If machine-assisted analysis of LANDSAT data will yield reasonably accurate estimates of area, then other inputs (ancillary data and aerial photography) can be used to measure the more specific parameters (species, volume, etc.).

Our next activity will involve reviewing specific survey standards for timber and range inventories. Also we will make some logical assumptions regarding constraints or boundary conditions that a remote sensing inventory design will be expected to meet. A list of conditions that boundaries will be defined for include:

- An assumption regarding overall inventory efficiency.
- A definition of the physical limitations of large area inventory coverage.
- Consideration of how ancillary data inputs can reduce bias associated with the selection of ground sample plots.
- A statement concerning sampling objectivity.
- A definition of reasonable sampling error.
- An approach for minimizing sampling bias.
- An approach to obtaining reliable and repeatable population estimates.

The USDA-FS, Forest Survey Handbook, 1967 (FAP has been requested to provide LARS with a copy of this document) is expected to provide some of the information necessary for timber inventory design. However, we are not aware of a similar publications in the area of range surveys.



Question: Is there sufficient information available concerning specific objectives and precision to begin preliminary design phase?

Figure 1. A schematic of the solution to the inventory design problems.

Therefore, our current plan for the project flow, as shown in Figure 1, is to emphasize timber inventory. This will continue until suitable documentation for the range survey problem is identified.

The approach we intend to pursue will progress from the general to the more specific. Initially our design concern will be focused on partioning the data into homogeneous areas. Succeeding activities would be aimed at further subdividing these areas into meaningful spectral classes. Finally, we will work toward identifying individual ground sample sites within the classes.

Currently, this project is operating in the timeframe detailed in the approved implementation plan. Namely task 1-the literature review-is winding down and task 2-the preliminary design-is commencing. The only problem encountered rests in the lack of information available about range survey procedures. The problem is being addressed and it is not expected to delay the task 2 activities substantially.

2.8 Regional Applications Project

Major Activities

Several tasks were pursued as part of the RAP effort during the second quarter of CY77.

A standardized classification procedure for analyzing the Texas Coastal Zone data was compiled. This procedure was developed in order that all quadrangles in the study area would be classified using similar analysis techniques to allow for meaningful comparisons and evaluations of the classifications.

A second task involved the modification of the spectral hierarchy developed by JSC/LEC and supplied to LARS earlier this year. The original hierarchy was developed using photointerpretative information such as texture, spatial distribution, subjective cultural interpretations and spectral response. Since the Landsat multispectral scanner records one data of a scene, any hierarchy to be used with scanner data must be spectrally derived. Table 1 lists the 35 spectral classes identified in the Austwell Quadrangle area and their corresponding JSC/LEC hierarchical classes. This comparison was accomplished by overlaying the classification with the spectral hierarchical transparency and making a point by point comparison. It is readily apparent that many spectral classes represent more than one hierarchical class. The final version of the modified spectral hierarchy is still to be completed.

The third task accomplished this quarter was the classification of the Port O'Connor and Pass Cavallo Quadrangles using the standardized classification procedures. These classifications will be used as the base data to which classifications developed from the later three dates will be compared in order to analyze the post classification change detection technique and to develop a standardized change detection procedure.

A comparison of various geometric correction functions was also conducted. The mathematical functions tested were the affine, collinearity, and biquadratic geometric correction. The methods were tested by selecting check points in the lower water covered portion of the Port O'Connor Quadrangle where no control points existed. Preliminary results indicate that the affine geometric correction function is the most accurate of the three tested.

The final task accomplished this quarter was the completion of a RAP simulation study. This study consists of a documented procedures package for classifying the Pass Cavallo Quadrangle in the Texas Coastal Zone. This package will be valuable in transferring analysis technique procedures, as applied to coastal zone features, to persons new to data analysis or coastal zone environments.

Problem Areas

Awaiting receipt of Landsat data and supporting reference data for this time t_2. The Post Classification Change Detection procedure will be tested upon receipt of this data.

Spectral Class	Class Definition	Heirarchial Class(es) Represented
22.	sparse vegetation on light colored soil that is wet-few points in the delta	Heg
23, 24	mixed pixel response-bare soil and wet vegetation scattered in the lower delta region and along the coast	---
25.	"edge" water class-occurring along the coast and fringing shallow water areas in the delta	Ws
29, 30	Bare soil tilled and wet	Bst
31.	bright wet surface (concrete) delineating the refinery	MMu
32, 33	very few points picking up the commercial in the city residential class	MMu
34, 35	sea water in the bay area	W
26, 27, 28	shallow water	Ws

SPECTRAL CLASS DEFINITIONS - AUSTWELL QUADRANGLE

Spectral Class	Class Definition	Heirarchial Class(es) Represented
1.	low vegetative cover on light (bright) soil	He, Heg
2.	sparse low vegetation	Het
3.	mixed woody/herbaceous cover-fairly green	Mdg, HGg, some Md.
4.	wet herbaceous-lush cover	He, some Het, Heg.
5.	herbaceous-covered with water	Hei
6.	herbaceous-completely inundated	Hei
7.	mixed-woody and herbaceous-fairly dense	He, Wf
8.	bright bare soil-wet spread all over in the delta	Bst
9.	herbaceous-wet-occurring mainly in the delta	Hei
10.	standing water, dard-low reflectance-some vegetation	some Hei
11.	herbaceous-fairly dense	He
12.	herbaceous emergent wet	Hei, Heg
13.	sparse wet herbaceous-low spectral response	Hei, He
14.	sparse wet herbaceous-low spectral response-few points occurring alongside class 13.	Hei, He
15.	bare soil wet	Bst
16.	water in pond next to the refinery (?)-high reflectance in class 2 indicating high turbidity-or some pollutant-or red suspended sediment	MMw
17, 18 19, 20	Sea water classes	W
21.	bright soil (sandy slightly inundated-fringing around shallow water ponds and in the delta	Bs, Ws

2.9 Interpretation of Thermal Band Data

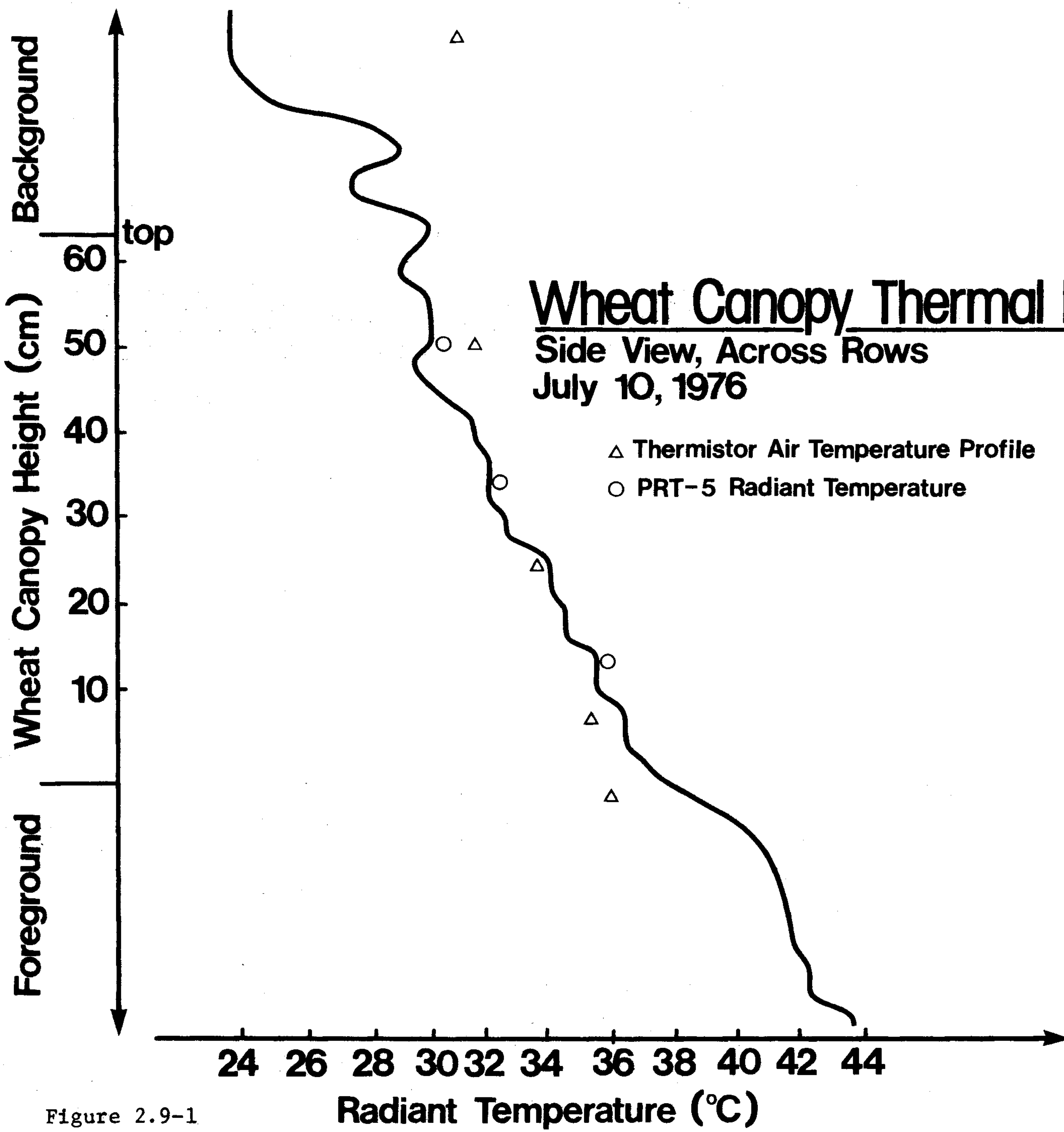
Major Activities. Data acquired during the first quarter of the reporting period were annotated and processed. The data were carefully inventoried and were described in three mission reports that ultimately will become a part of the final report on this project. The thermal scanner that is used as a primary data source is not equipped to produce direct digital output. Therefore, the data were recorded on a video tape recorder in real time, and then subsequently were displayed on the display unit that is part of the thermal scanner system. The thermal displays were then photographed on positive transparency film. The film was placed in a conventional film scanning system and digitized to an 8 bit precision. Naturally, there is a loss of dynamic range in this process but at the present time it is the only process available.

The thermal scanner was operated in one of two basic modes depending upon the particular data being acquired. In one case the thermal image was directly viewed in the normal upright position with the scan lines moving from left to right across the image. Using this approach each scan line can be viewed one at a time in a so called "A-scan" mode so that a radiation temperature profile of any horizontal scene element can be reviewed. In the other mode of operation a mirror-pair was placed in front of the thermal scanner so that a 90-degree rotation of the image occurred. The basic scan line nechanism of the scanner remained unchanged. Therefore, the scan lines appeared to scan the scene vertically from top to bottom. As a result, when the scanner was operated in the "A-scan" mode a vertical radiation temperature profile of any given scene line could be obtained.

The thermal scanner was pointed into the side of the wheat canopy under study. During the thermal scanning process an array of thermistors was used to measure the air temperature as a function of depth into the canopy, the air temperature above the canopy, the temperature at the surface of the soil, and the temperature approximately 1 inch below the soil surface. Several arrays were located at different points in the canopy. The data produced by the thermistors was converted into temperature data and collected with a special multiplexing system that printed the thermistor temperature data onto paper tape. These data were later keypunched and stored on magnetic CCT. After the thermal scanner data were digitized several of the vertical radiation temperature profiles at different portions of the canopy were averaged and correlated with the air temperature profile data mentioned previously. Some of these data are illustrated in Figure 2.9-1.

A special experiment was designed to examine the edge effects in the previously described experiment. A blackbody was positioned near the edge of the canopy and slowly withdrawn into the canopy a row at a time. The images was observed on the display unit on the thermal scanner. The image of the blackbody could be observed from the edge of the canopy when the blackbody was as far as 10 rows deep in the canopy. Therefore, it was concluded that the edge-wise thermal imagery of the canopy essentially represented bulk effects within the main canopy structure.

The thermal scanner was also positioned above the canopy to obtain a vertical view of the detailed structure of the canopy. At a height of approximately 8 meters above the canopy the detailed thermal structure of the soil background and primary canopy elements were observed. Since the



Wheat Canopy Thermal Profile
Side View, Across Rows
July 10, 1976
△ Thermistor Air Temperature Profile
○ PRT-5 Radiant Temperature
Background
top
Foreground
Wheat Canopy Height (cm)
60
50
40
30
20
10
24 26 28 30 32 34 36 38 40 42 44
Radiant Temperature (°C)

Figure 2.9-1

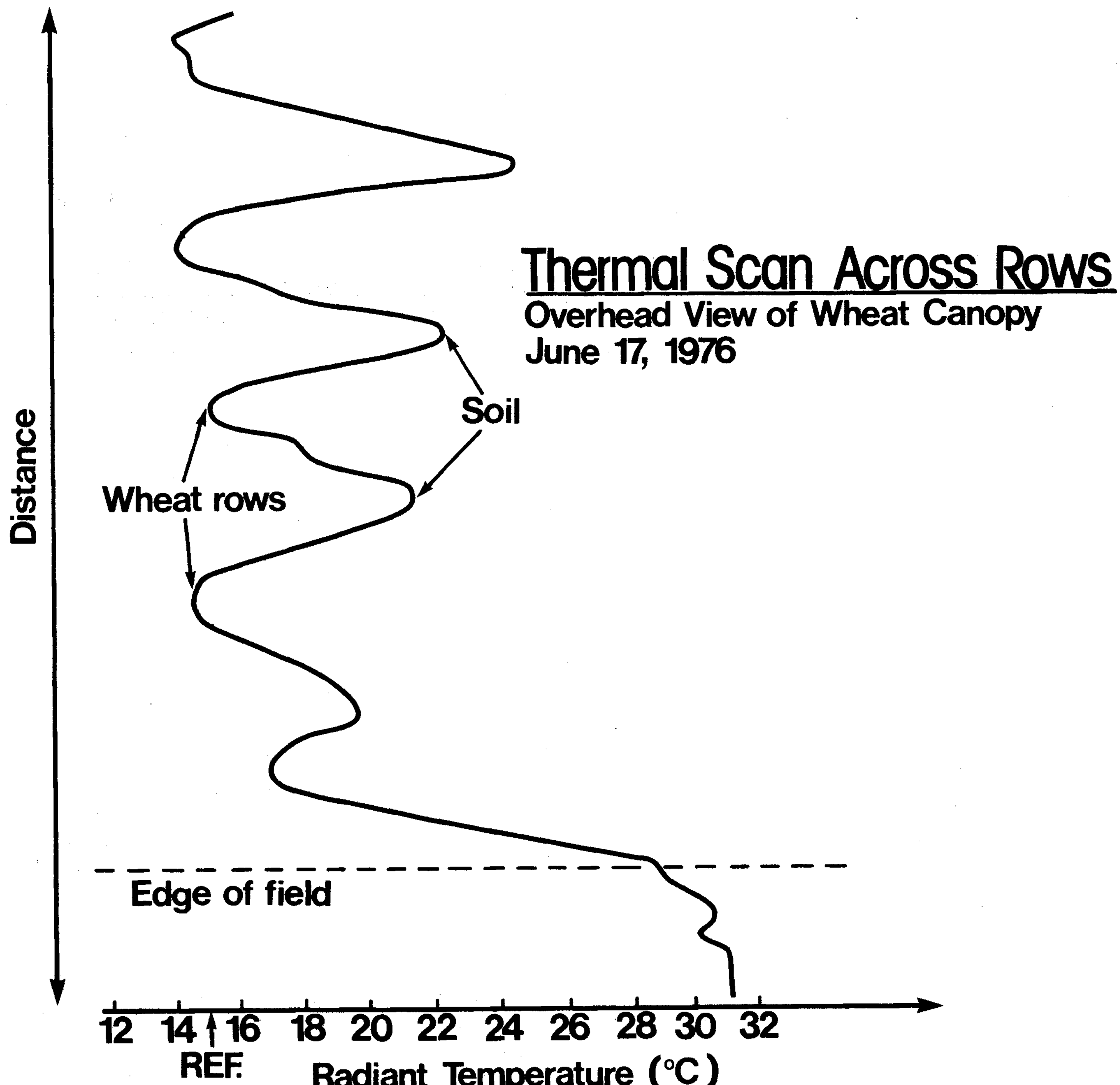

Thermal Scan Across Rows
Overhead View of Wheat Canopy
June 17, 1976
Soil
Wheat rows
Distance
Edge of field
12 14 16 18 20 22 24 26 28 30 32
REF.
Radiant Temperature (°C)

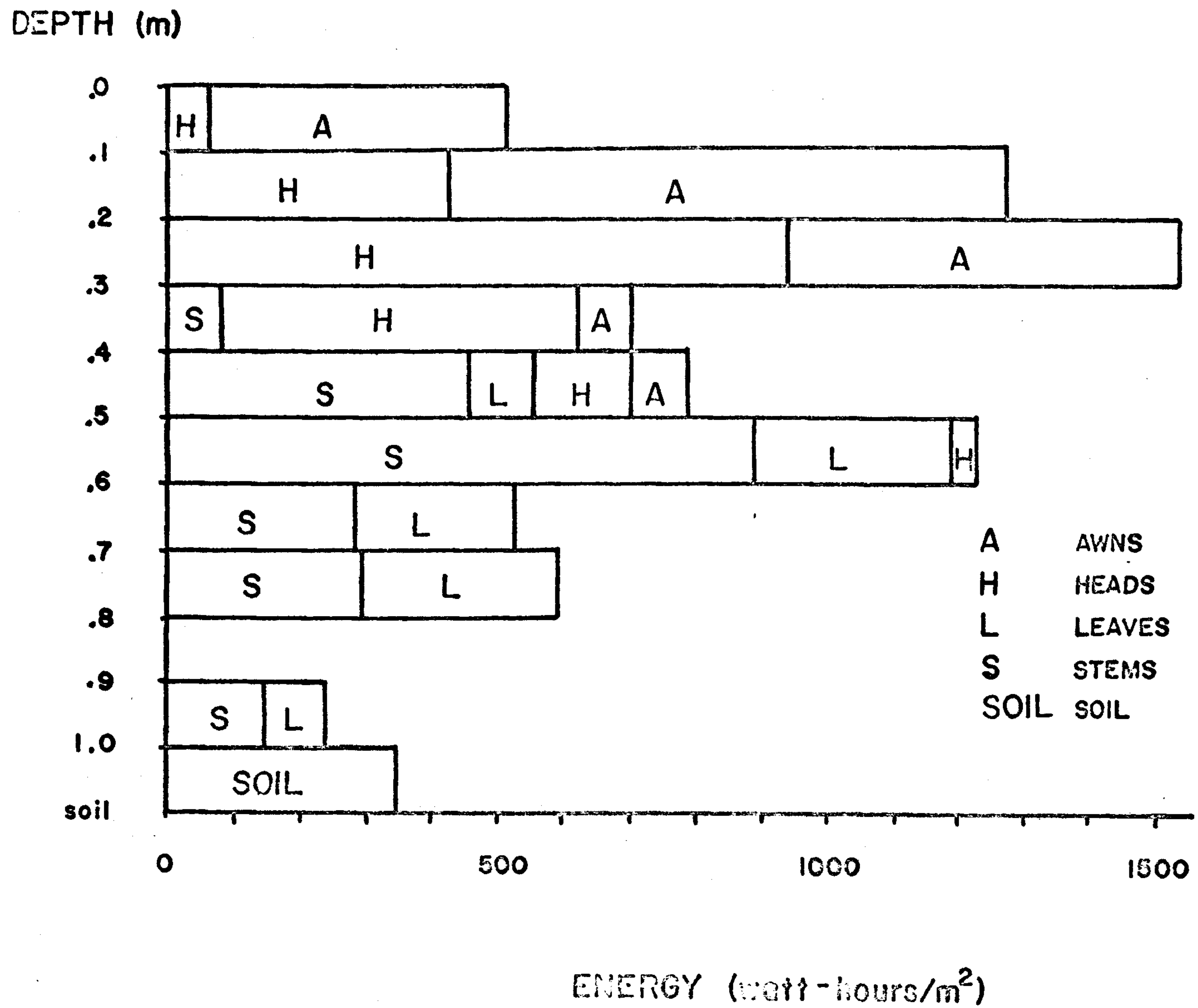
Figure 2.9-2

thermal scanner obtained the thermal imagery in real time the motion of the canopy could be observed in the imagery. During relatively high wind conditions the basic radiation temperature of the canopy elements did not seem to change; however, the canopy would bend over and obscure the soil which, at the time the imagery was taken, was warmer than the canopy elements themselves. A radiation thermometer which was also deployed during all of the thermal measurements would average the temperature measurement shown on the thermal imagery and would indicate that the overall temperature of the scene was changing. However, it is important to realize that the temperature of the canopy elements was relatively stable and that the apparent change in radiation temperature was due principally to geometric changes induced by the ambient wind conditions.

The thermal scanner data form the vertical view was also presented in "A-scan" form. Several of the scan lines were averaged and the results are shown in Figure 2.9-2. Here the variation between the tops of the wheat canopy elements and the soil temperature is clearly delineated. The data were also presented in color-coded image form in which colors of the image correspond to radiant temperature range bands in the imagery. All of the data were calibrated with respect to reference blackbodies located within the thermal imagery. This color-coded calibrated imagery will be presented in the final report.

Development of a simplified radiation model also took place during the quarterly reporting period. In this model the detailed effects of convective and conductive radiation transfer within the canopy are not included in the basic structure of the model. The objective is to produce a model that will predict the radiant temperature of the canopy as viewed from aircraft or satellites in terms of the geometric structure of the canopy. Since radiation temperature is the primary variable of interest, then one can formulate a model in which radiation temperature is the basic experimental variable. The experiments above obtain the radiation temperature profiles which are needed for the model. It turns out that the air temperature profile is closely correlated with the radiation temperature profile obtained from the thermal scanner and in subsequent experiments possibily can be used as the sole data source by experimenters who do not have access to a thermal scanning system. Once the radiation temperature profile is obtained, the model uses these data in combination with geometric data describing the canopy to obtain view factors so that the radiation from each layer in the canopy can be properly combined into an overall radiation exitance from the top of the canopy. The view factors from each layer are obtained from an optical depth measurement that is to be described in LARS Information Note 120776 that will be published in January of 1977. As an example of data that have been produced by this measurement technique the distribution of energy within a wheat canopy is shown in Figure 2.9-3. This figure shows how the energy is distributed vertically through the canopy layer by layer. It also shows how that energy in a particular layer is distributed amongst the wheat plant elements in that layer. This figure shows how the incoming solar flux is distributed and using a simple inversion process this same data can be used to show how the radiation in any given layer appears at the top of the canopy.

Technical Problems and Solutions. No signigicant technical problems have yet been encountered.



2.9-3 Energy Distribution in Bearded Wheat Canopy

Plans for the Next Quarter. The radiation model described in the above report will be programmed and tested during the next quarter. Corrections, if necessary, will be made and much of the data that were acquired during the first quarter should be processed during the coming quarter. A line chart which described the relationship between accomplishments to date and the implementation plan submitted during the first quarter are shown in Figure 2.9-4.

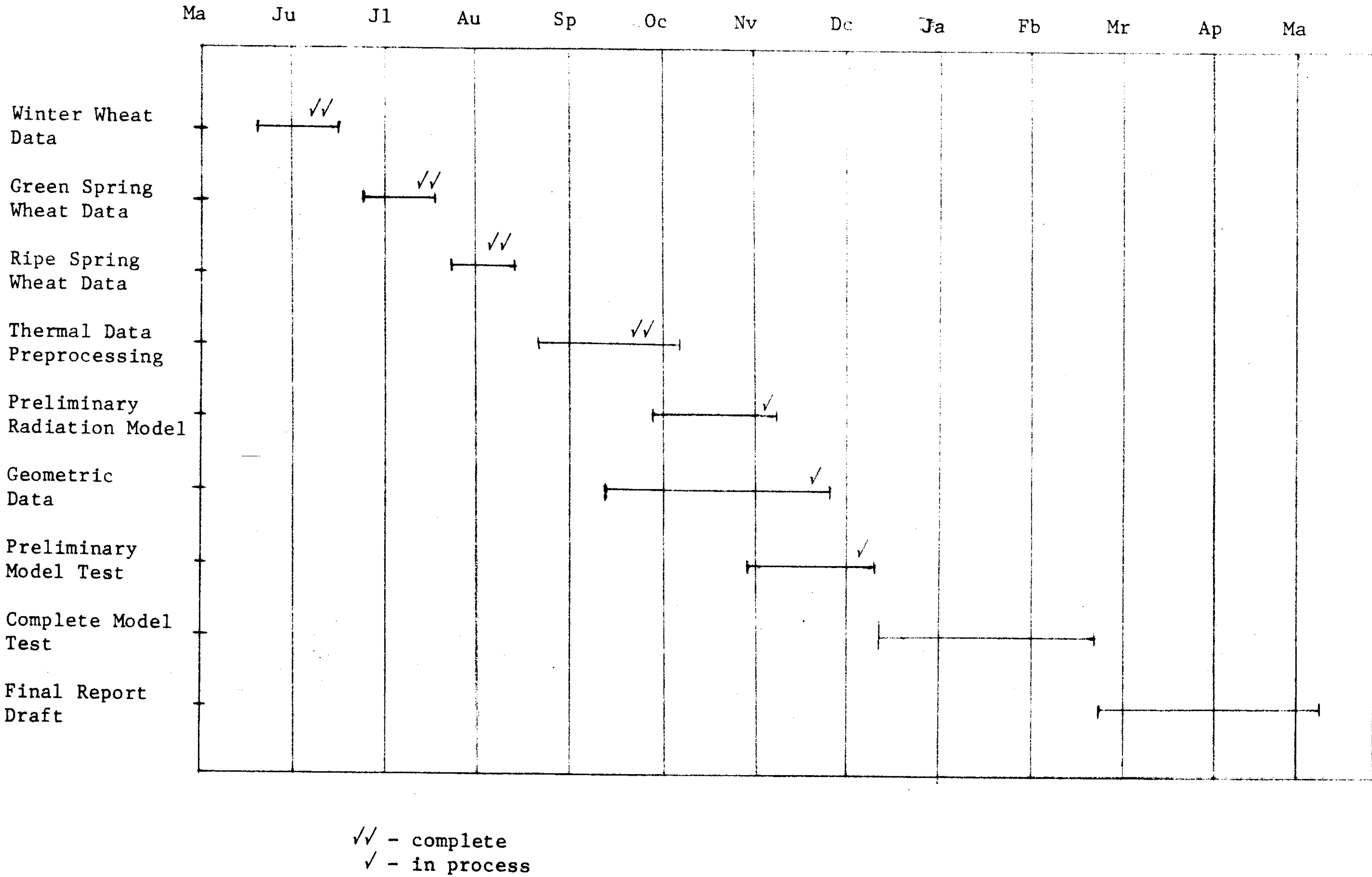


Figure 2.9-4

Implementation Plan Status

2.10 Super Site Data Management

Major Activities. The major activity for the quarter concerned development of the catalog for the LACIE Field Measurements Data. Work dealing with other EOD field data gathering projects was delayed pending a decision as to whether Purdue/LARS should receive and/or manage that data.

The Field Measurements Data Library Catalog was finalized incorporating suggestions obtained from reviews of the first two versions. The Field Measurements Data Library Catalog is divided into separate volumes - one for each crop year during which data is collected. Volume I, for the 1974-1975 crop, has been printed and distributed. Preliminary versions of Volume II, for the 1975-1976 crop, and Volume III, for the 1976-1977 crop, have been printed and will be updated periodically to reflect the current status of data collection and data processing.

Plans for Next Quarter. During the next quarter, work will continue on the management and distribution of LACIE field measurements data. A decision is hoped for during this quarter concerning Purdue/LARS' role with the data from the other EOD field data gathering project.

2.11 Soil Classification and Survey

Preparation of Manuscripts. Preparation was continued on manuscripts for two technical papers entitled "Climatic Effects on the Relationship between Physico-Chemical Properties of Soils and Reflectance" by O. L. Montgomery and M. F. Baumgardner and "An Investigation of the Relationship between Spectral Reflectance and the Chemical, Physical and Genetic Characteristics of Soils" by O. L. Montgomery, R. A. Weismiller, and M. F. Baumgardner. Particular attention was given to a search of the literature and a summary of the effects of climate on soil characteristics, especially on reflectance. Graphs and tables were prepared to illustrate the relationships between climate and soil reflectance.

Indoor Exotech Studies. Planning meetings were held with personnel of the Measurements Research Programs. The Exotech spectroradiometer was tentatively scheduled for indoor acquisition of soils spectra in February and March 1977. Selection of soils to be included in this study is proceeding. It is proposed that soils be selected to provide an adequate statistical sampling for each of several major climatic zones.

Long Range Projection for Soils Research. Considerations of the long range needs for soil research were compiled, taking into account the spectral properties of soils and temporal changes in soil conditions. Soil reflectance is affected by many different parameters, both physical and chemical. Many of the uncertainties inherent in the interpretation of soils reflectance data now will be removed if the relationships between reflectance and physico-chemical properties of soils can be quantitatively defined. In order to obtain these quantitative definitions spectral studies are planned for the following soil parameters and conditions:

A. Soil without Cover

1. Texture
2. Structure
3. Moisture
4. Temperature
5. Chemical Properties
6. Physical Properties
7. Drainage Characteristics
8. Surface Roughness
9. Organic Matter

B. Soil with Cover

1. Organic Residue
   a. Corn Stover
   b. Soybean Residue
   c. Wheat Straw
   d. Decisuous Tree Leaves
   e. Coniferous Tree Residue

2. Living Plants
   a. Species
   b. Stage of Growth
   c. Percent of Ground Cover
   d. Plant Condition

3. Inorganic Cover

Many of the changes in the landscape are of interest to the soil scientist. Present remote sensing technology provides data acquisition and analysis techniques which might be used to identify, map and monitor soil conditions which greatly affect soil productivity and land use. In projecting long range research involving temporal changes in the scene, the following areas have been designated as significant candidates for research:

A. Soil Deterioration

1. Salinization
2. Water-logging
3. Water Erosion
4. Flooding (Inundation)
5. Wind Erosion
6. Desertification
7. Depletion
8. Compaction
9. Subsidence
10. Landslides, Deposition
11. Economic Analysis of Soil Deterioration

B. Watershed Monitoring

1. Surface Drainage Characteristics
2. Soil Loss form Watershed
3. Soil Deposition in Watershed
4. Sedimentation Rate Studies
5. Nutrient Loss from Watershed
6. Changes in Soil Productivity

Basic to all spectral studies of soils is the objective of producing an end product which will classify or map an area of land into classes related to different levels of productivity, different carrying capacities or land use potentials. Both present and future research on the spectral properties of soils and the temporal changes in soil conditions contribute to the application of remote sensing technology in soil classification and survey. Present and long range research must address the following aspects of soil classification and land suitability mapping:

A. Delineation of Soil Boundaries
B. Soil Mapping Unit Homogeneity
C. Soil Characterization
D. Soil Productivity
E. Land Use Capability (Carrying Capacity)
F. Soil Genesis
G. Land Form Characterization
H. Economic Assessment of Digital Analysis of MSS Data for Soil Survey
I. Soil Data Base

These potential areas for soils research are especially dependent upon the development of analytical methods and interpretive skills for effectively using satellite multispectral data.

2.12 Multitemporal Analysis

Introduction. The four main areas that received attention this quarter were data set acquisition, analysis techniques, a preliminary experimental design, and baseline analyses. Four data sets were acquired. Analysis techniques were established for unitemporal analysis and multitemporal analysis. The multitemporal techniques developed so far are variations of ones that analysts have used previously. As the cascade classifier is implemented, techniques will be developed for its use. The preliminary experimental design includes analysis techniques and procedures for classification and comparison of results.

At the conclusion of this quarter, all work on this task was terminated as directed by the sponsor.

Data Acquisition and Preparation. One CITARS data set in Illinois and three LACIE/SRS segments in Kansas have been selected as the first data sets to be analyzed. All contain crops and reflect seasonal changes in one growing season. They were chosen because adequate ground truth is available for them at dates throughout the growing season. A more detailed discussion of the data sets is contained in the preliminary experimental design plan included in this report.

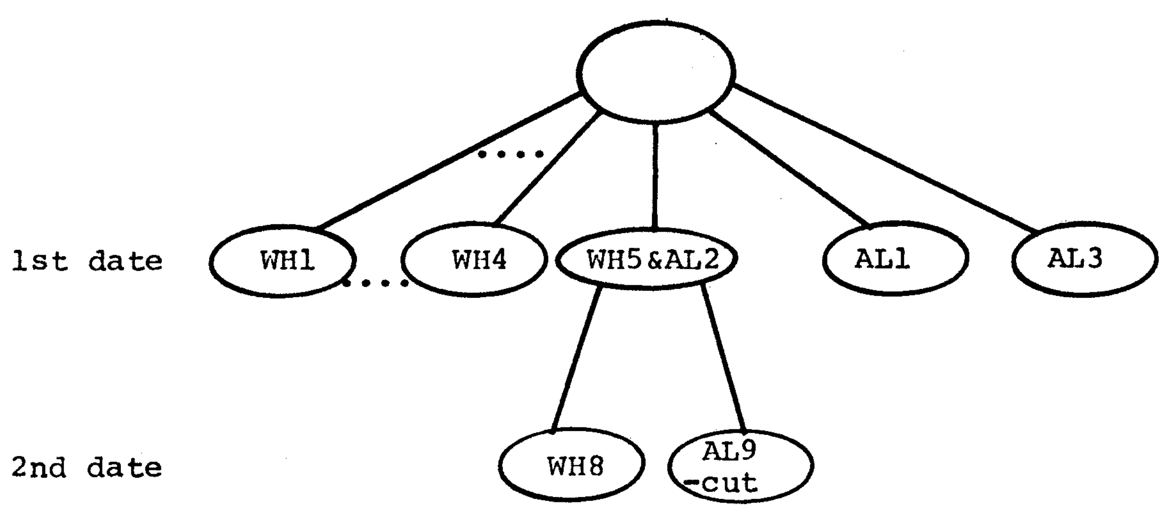
A substantial amount of time was spent searching for data sets and determining which ones could be used for multitemporal analysis. That is, there had to be adequate ground truth available and various cover types that could cause confusion at different times in the year or growing season. Four data sets may be adequate for comparing multitemporal analysis results against unitemporal results, but we are hoping to obtain one more data set for a region in Colorado that contains mountains and forests. This data is registered and has been unitemporally classified by a forestry analyst. He is in the process of obtaining test fields. Since all four of the selected data sets only contain crops, we would like to have at least one more that is non-agricultural. Also, it would certainly be better to have seven or eight data sets instead of four to compare results, so, if time permits, some LACIE intensive test sites in Kansas, or two data sets from the Field Measurements project, could be used. These last two data sets are in the process of being overlayed and geometrically corrected. One is in Finney County, Kansas (4 dates), and the other is in Williams County, North Dakota (2 dates).

Since there may be enough adequate data sets available and others soon to be available, at this time there is no need for this project to preprocess and register any data.

Analysis. Unitemporal baseline analysis has been completed on the four multitemporal data sets. The standard maximum likelihood classifier was used to classify the data. Training and test class performances and proportion estimates from these unitemporal classifications will be used as benchmarks for multitemporal analyses.

A preliminary experimental design plan, described in the next section, was written. It includes analysis techniques to be used as well as the plan for the number and types of classifications to be done. The techniques are not really new; they are variations of fairly standard techniques. As new techniques are developed and the cascade classifier is finished being implemented, the design will change. The main reason for the preliminary experimental design plan at this time was to use it in doing some initial multitemporal analyses. The results of these analyses will show the relative benefits and limitations of fairly "standard" techniques for multitemporal analysis.

An example of one case where multitemporal analysis should prove beneficial is for classifying wheat in Kansas. Wheat can be confused with alfalfa at certain times in the wheat growing season. Alfalfa, though, is cut more than once during its growing season, the first cut being done before any wheat is harvested. In the very early spring, wheat and alfalfa may be distinguishable from everything else on the ground, but some of the wheat may be as green as alfalfa and thus be confused. However, after the alfalfa is cut, wheat can certainly be distinguished from alfalfa; but at this point, pasture and some wheat may be confused and alfalfa and some fallow may be confused. Thus, neither of these two times in the growing season is an optimal time. Now, if the layered classifier or cascade classifier is used, a much better classification should result. For example, a simplified tree design for the layered classifier might be as follows:



The tree does not include pasture, fallow, or any other ground cover, but just shows that for wheat and alfalfa, the wheat subclass that is confused with the alfalfa subclass can be separated on the second date but not on the first. The remaining classes on the second date are not needed, since the first date was good enough for all classes except the ones where wheat and alfalfa were being confused.

There may be potential benefits derived from using three or even four different dates during the growing season. The experimental plan at this time includes only unitemporal and bitemporal classifications. If resources permit, however, 3- and 4-date classifications will be done. All the data sets to date have four different dates available.

A workshop was held at the end of November. Users who have done some multitemporal analysis presented their techniques and discussed the problems they have encountered in working with multitemporal data. These people had previously been contacted concerning data sets and techniques, and the preliminary experimental plan reflects some of their experience.

Experimental Design (Preliminary). The main objectives of this project are to develop effective techniques and algorithms for multitemporal data sets. The data sets available at this time which can be used for multitemporal analysis are from the CITARS project (one data set) and from the LACIE project (three data sets). All of the data sets reflect seasonal changes during one growing season. The analysis of the data can be divided into two parts: obtaining training statistics and classification. For multitemporal analysis, a set of training statistics can be obtained independently for each date. The training subclasses in the set for one date cannot necessarily be paired with subclasses from the next date, but they are based on the same data points. Statistics for a set of classes are also calculated for two dates at once, using all eight channels or a subset of channels from each date. Classification will be done using three types of classifiers: standard maximum likelihood, layered, and cascade. The number and type of classifications in this experimental plan have been kept to the minimum number needed to demonstrate that multitemporal analysis is superior to unitemporal analysis. The emphasis is on showing that a pair of subclasses from different major cover types which are not very separable on one date can be more easily separated if a second date is used. The analyst naturally wants to use the date that will give him the best results, but usually, especially where agricultural crops are concerned, at least one other date will be needed to better differentiate some classes. Results will be evaluated from training and test field performance and proportion estimates, when available.

I. Data Sets

A. The data set being used from CITARS is the one over Fayette County in Illinois. Of the six CITARS data sets, it had the largest number of dates that were nicely spread throughout the growing season. There are six different dates, but two pairs are only one date apart, so there are effectively four periods in the growing season represented. Fields are available, and the Landsat data is geometrically corrected and overlayed. The data, black and white photographs, and acetate overlays with the fields outlined, are all the same scale.

Dates: June 10, June 29, July 17, August 21, 1973

Size: 536 lines, 328 columns

Major Cover Types:

Name	Points
Corn	356
Soybeans	497
Woods	217
Urban	103 (use only for testing as in CITARS)
Other	305

Every other field of each cover type (about half of the points) will be used for training and the remaining for testing.

B. The three LACIE data sets (SRS segments) are in Grant, Stevens, and McPherson counties in Kansas. These counties were chosen from all the SRS segments since they contained an adequate number of points of two or more of the following: wheat, pasture, alfalfa, and "unknown" agriculture. All of these cover types have a good chance of causing confusion with one another during different times in the growing season. Fields are supplied, and the data is overlayed, but not geometrically corrected. Enlarged false-color images of the data and acetate overlays scaled to this imagery are provided. The fields are outlined and identified on the overlays.

Dates:

County	Dates (73-74 wheat growing season)
Grant	Oct. 23, May 9, June 14, July 2
Stevens N	Oct. 23, May 27, June 14, July 20
McPherson	Oct. 20, April 18, May 6, May 24

Size: 200 lines, 200 columns for each

Major Cover Types:

	Points		
Name	Grant	Stevens N	McPherson
wheat	1590	2081	568
fallow	1319	2509	982
corn/ sorghum	874	789	135
pasture, trees	2275	876	936
alfalfa	192	---	217
unknown agriculture	----	468	115

II. Training Set Analysis

In general, the technique will be to use the Cluster processor to obtain subclasses for each of the major cover types and then use the Separability processor to determine which subclasses from different cover types are not very separable and which subclasses from the same cover type should be pooled. Cluster will be run once for each major cover type. As a guideline, use the following procedure to determine how many subclasses to request in Cluster. The procedure was developed for CITARS, but appears here slightly modified for multitemporal data sets: let "x" be number of subclasses to ask for, "s" the estimate of number of distinct subclasses, "t" the number of total points, and "n" the number of channels.

Step 1. Let x=2s if one date; x=2·(s+1) for two dates.

Step 2. If x > (t/10n), set x = (t/10n). This insures that the number of points in each cluster will be approximately 10x number of channels.

Step 3. If x < s, set x = s.

Use these s-values:

Name	s
corn	4
soybeans	3
woods	2
other	6
wheat	3
fallow	3
corn/sorghum	3
pasture, trees	4
alfalfa	3
unknown agriculture	3

From the Separability processor, any class pairs with a transformed divergence less than or equal to 1300 should be labeled as confusion classes if they are from a different major cover type or should be pooled if they are from the same major cover type. Use the Mergestatistics processor to do any pooling and print the bispectral plot of the class means. The plot helps the user become familiar with the data.

A. Unitemporal training sets. Obtain training statistics for each of the four dates in the data set using the above general guidelines.

B. Multitemporal training set - Type I. Use the Separability processor to determine which two channels are the best from each date. That is, use the pair of channels which has the highest minimum divergence for all confusion class pairs. Use this information to merge the statistics for two dates (2 channels for each date). The dates to use are given in Section III.

C. Multitemporal training set - Type II. Use the same two dates and the same two channels from each date as determined by part (B). This training set, though, cannot be obtained from the unitemporal sets. For each pair of dates, use the procedure for Cluster and Separability described in the beginning of this section with x = 2(s+1). Note that the number of classes in this training set will be less than that in (B) since here we have one set of classes for both dates instead of one combined set for each date.

III. Classification

A. Unitemporal. Use the statistics obtained in II-A as input to the standard maximum likelihood classifier. Classify the data set for each date separately.

B. Multitemporal - Type I. The object of any multitemporal classification is to try to improve on a unitemporal classification. Use the Type I multitemporal training statistics. Choose the training statistics from the best date and pair it with each of the other three dates, obtaining three 2-date statistics decks. "Best" is defined as the date with the highest test class performance from the four unitemporal classifications. Classify each of the three pairs of dates twice, once with the layered classifier and once with the cascade classifier.

Now, in order to use the layered classifier, a tree has to be designed. The first level of the tree should contain a node for each training class from the best date. Classes that are not "confusion" classes will be terminal nodes. All confusion classes will branch out into a second tree level. Each of these classes will branch out and contain a terminal node for each subclass of the major cover types being confused. For instance, if an alfalfa subclass on the first date is being confused with wheat on the same date, the alfalfa subclass should have a branch for each of the alfalfa and wheat subclasses from the second date.

Techniques for using the cascade classifier are being developed and will be specified in the next quarterly report, along with some results.

C. Multitemporal - Type II. Use the Type II multitemporal training statistics and, as detailed in III-B, determine which three pairs of dates to use for classification. Classify each of the three pairs of dates three times, once with the standard maximum likelihood classifier, once with the layered, and once with the cascade. Since each class has statistics for channels from both dates, the standard classifier can be used. The tree design for level one is the same as in III-B, but the second level will be different since any class from the first date also has statistics from the second date. So, if there are two subclasses from one date being confused, simply use these same two subclasses with the second date channels as separate nodes in the second level.

Techniques for the cascade classifier are being developed.

IV. Evaluation Criterion

Both training and test field performance will be obtained for each data set. Proportion estimates are available for the CITARS data sets and can be calculated for the LACIE/SRS segments if needed.

Most multitemporal results should be an improvement over the unitemporal results.

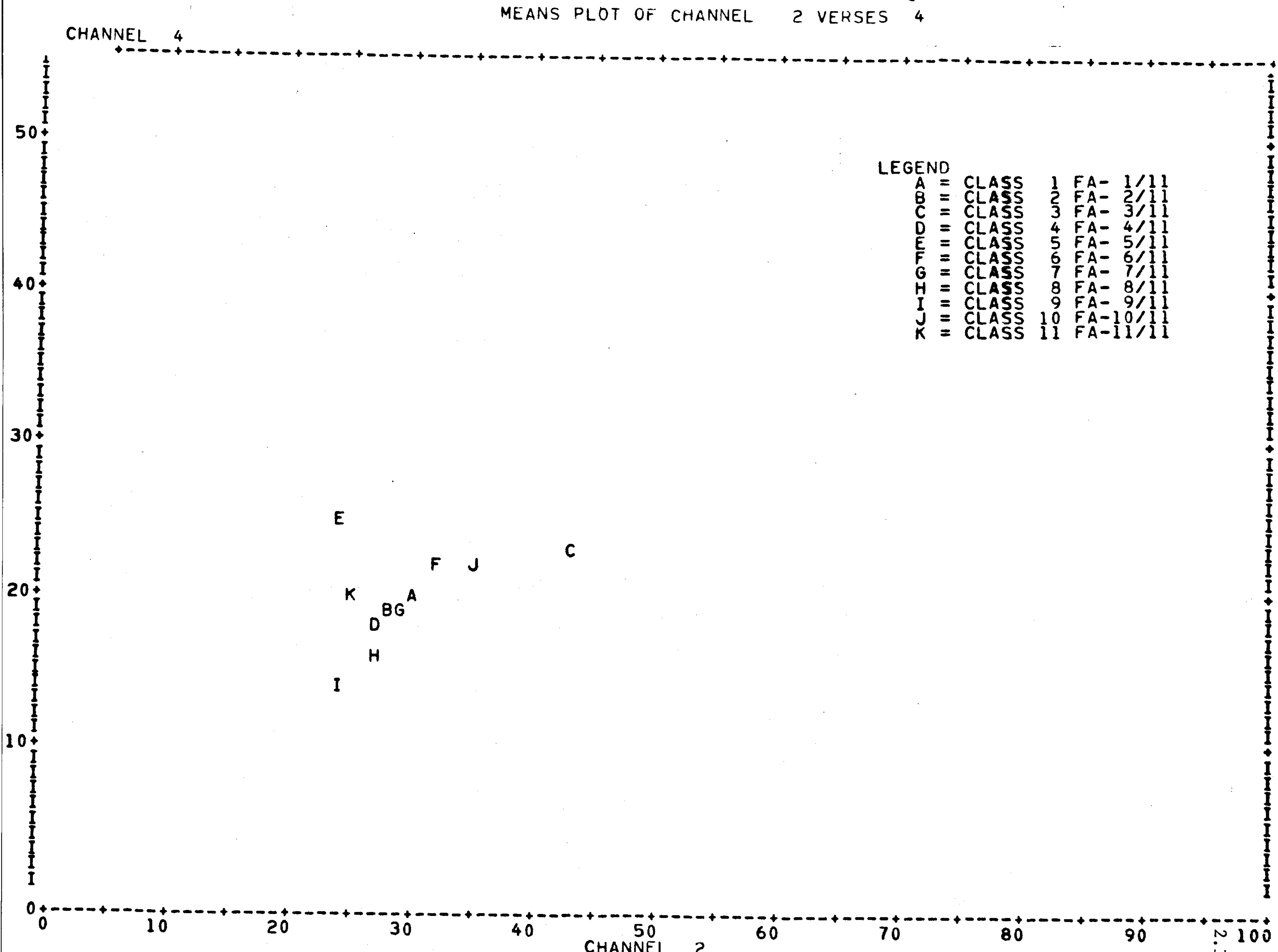


Figure 1. Bispectral Plot

Programming. Implementation of the cascade processor has begun. Details such as control cards are being worked on. The programs for the standard LARSYS maximum likelihood classifier are being used in modified form, since many of the control cards are the same and the algorithm for the cascade classifier can be thought of as an extension and elaboration of the present classifier. Programming for the cascade classifier should be completed and some testing begun during December.

A program already existed for combining statistics decks. This program had to be modified to do some further manipulations on the decks to facilitate the preparation of a statistics deck for input to the layered and cascade classifiers. In addition, a "bispectral plot" of the class means was incorporated into the program. The subroutine needed to print the graph had been previously written by a user from another project and needed only minor modifications to be incorporated into this "merge-statistics" program. The bispectral plot aids the user in becoming familiar with his spectral classes. The plot has one channel as the x-axis and another channel as the y-axis and prints a different character for each class. The character is printed at the point (x,y) where x and y are the mean values for the two channels. An example of a bispectral plot of fallow subclasses is shown in Figure 2. A plot of all subclasses for all cover types could be used to graphically indicate confusion classes.

A couple of rather short programs were also written to aid the user in organizing training and test fields and tabulating points in the different cover types.

LARSYS CONTROL CARDS

MERGESTATISTICS

R E Q	KEY WORD(COL.1)	CONTROL PARAMETER	FUNCTION	DEFAULT
*	*MERGESTATISTICS	(NONE)	SELECT THE MERGESTATISTICS FUNCTION.	(NONE)
	OPTIONS	NOFIELDS	ONE DUMMY FIELD CARD IS PUNCHED FOR EACH CLASS IN THE STATISTICS DECK BEING GENERATED.	FIELD CARDS FROM INPUT DECK ARE PUNCHED.
		COSPEC	PRINT ONE COINCIDENT SPECTRAL PLOT OF ALL POOLS OR OF ALL CLASSES IF POOLS ARE NOT REQUESTED.	NO SPECTRAL PLOTS PRINTED.
		MEANS(C1,C2)	MEANS BI-SPECTRAL PLOT OF CHANNELS C1 VS C2	(NONE)
		MEANS(C1,C2,C3,C4)	MEANS BI-SPECTRAL PLOT OF CHANNELS C1&C2 VS C3&C4	(NONE)
	PUNCH	(NONE)	PUNCH CLASS MEANS AND COVARIANCE MATRICES IN BINARY FORMAT (INTERNAL COMPUTER FORM).	NO PUNCHING
		CHARACTERS	PUNCH CLASS MEANS AND COVARIANCE MATRICES IN CHARACTER FORMAT (EXTERNAL COMPUTER FORM).	BINARY FORMAT
	CHANNELS	I,J,...	CHANNELS I,J,... ARE SELECTED	COMMON CHANNELS TO ALL STAT DECKS
*	CLASSES		INDICATES CLASSES TO BE USED TO GENERATE STATISTICS DECK.	(NONE)
		ENTIRE(D1,D2,...)	ALL CLASSES IN DECKS D1,D2,... ARE INCLUDED IN NEW STATISTICS FILE.	
		DELETE(DN/C1,C2,.../)	ALL CLASSES IN DECK EXCEPT C1,C2,... ARE INCLUDED IN NEW STATISTICS FILE.	
		INCLUDE(DN/C1,C2,.../)	ONLY CLASSES C1,C2,... IN DECK DN ARE INCLUDED IN NEW STATISTICS FILE.	
	POOL	NAME(D1/C1,C2/,D2/C3,C4/,...)	POOL STATISTICS FROM CLASSES C1,C2,... FROM DECK 1 AND CLASSES C3,C4,... FROM DECK 2, ETC. 'NAME' IS THE NAME ASSIGNED TO THE POOL. IF A POOL CARD IS USED, ALL DESIRED POOLS AND CLASSES MUST BE EXPLICITLY REQUESTED.	NO POOLING
	DISK	READSTATS	DECK 1 WILL BE READ FROM DISK.	ALL DECKS READ FROM CARDS.
	SCALE	SPCLOW(D)	SET LOW END OF SPECTRAL PLOT TO D.	D=0
		SPCINT(E)	SET SPECTRAL PLOT INTERVAL TO E.	E=3
		ORIGIN(N,X.XX)	ORIGIN FOR CHANNEL N	AUTO-RANGING

Figure 2. Control Cards

R E Q	KEY WORD(COL.1)	CONTROL PARAMETER	FUNCTION	DEFAULT
		UNIT(N,X.XX)	(MEANS OPTION) INTERVAL FOR CHANNEL N (MEANS OPTION)	AUTO-RANGING
	DATA	-----START OF DATA DECK----- DECK 1 -- LARSYS VERSION 3 STATISTICS DECK.		
	DATA	-----CONTINUE USING DATA CARD IN FRONT OF EACH STATISTICS DECK.-- LARSYS VERSION 3 STATISTICS DECK IN THE ORDER GIVEN BY THE CLASSES CARDS		
*	END	(NONE)	END OF FUNCTION.	(NONE)

NOTE........*MERGESTATISTICS IS AVAILABLE ON LARSLIB 1FF, THE EXPERIMENTAL LARSYS MODULE LIBRARY ESTABLISHED BY THE DATA ANALYSIS RESEARCH GROUP. TO ALLOW USERS EASY ACCESS TO THIS DISK, A MODIFIED EXCOMD EXEC HAS BEEN WRITTEN WHICH AUTOMATICALLY ATTACHES AND LOGS IN THE EXPERIMENTAL PROGRAM DISK. USERS WHO HAVE THIS EXCOMD EXEC ON THEIR P-DISK CAN RUN *MERGESTATISTICS BY THE COMMAND 'RUN MERGE'.

Figure 2. (Continued)

Implementation Plan. An updated copy of the implementation plan is included as Figure 3, showing percentage estimates of subtasks completed.

Some study of temporal characteristics of typical scenes (Task 4.1) was done during selection of data sets. Discussion of some of these characteristics was also discussed during the workshop which was held at the end of November.

Initial multitemporal analysis techniques (Task 4.2) are included in the preliminary experimental plan (Task 5.1).

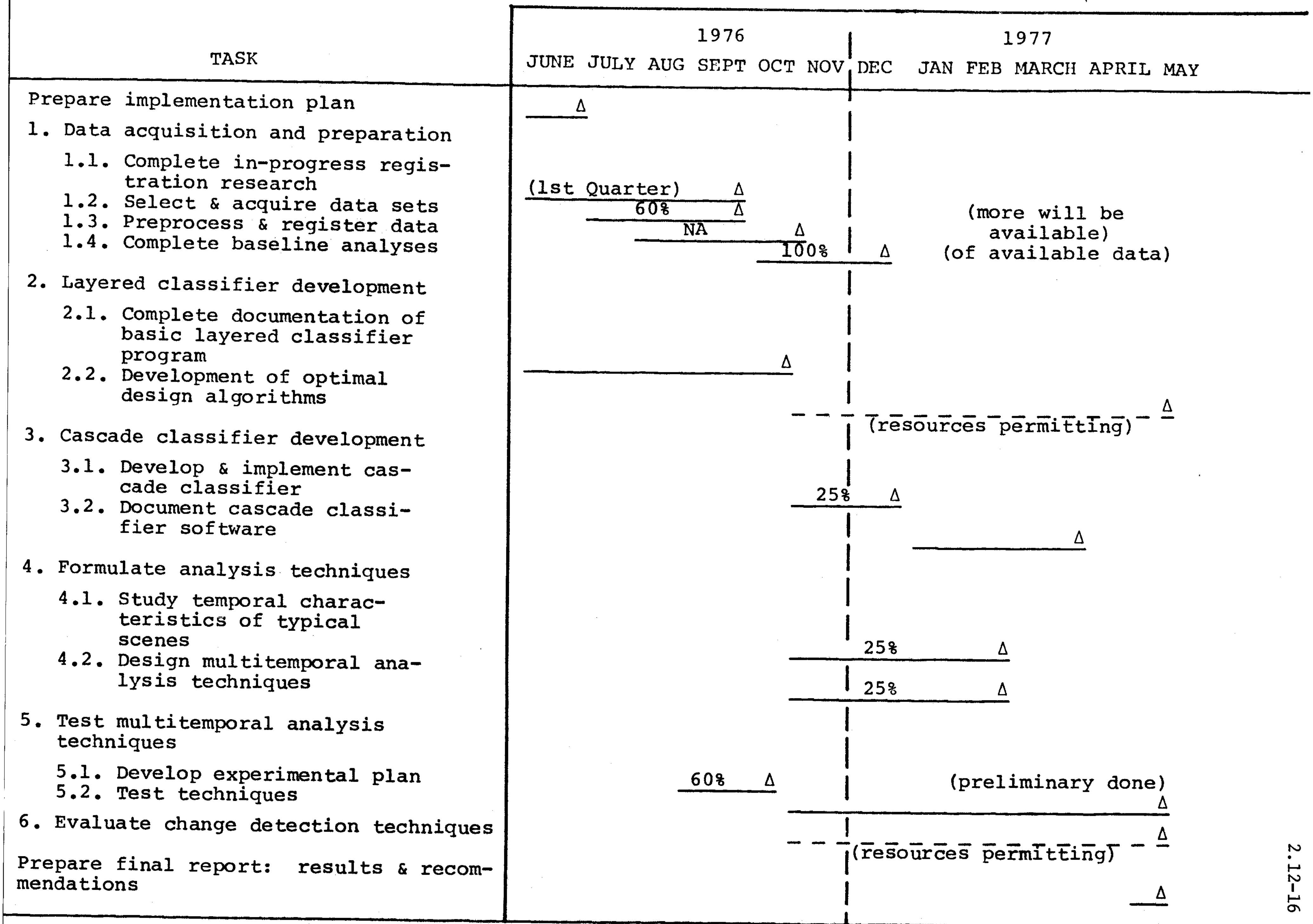


Figure 3. Detailed Implementation Schedule for Task 2.12.

Information Notes Issued
During Period
July-December, 1976

The following Information Notes were published in full or in part under NASA Grant No. NGL 15-005-112.

030576 Evaluation of Surface Water Resources from Machine Processing of ERTS Multispectral Data by P. W. Mausel, W. J. Todd, M. F. Baumgardner, R. A. Mitchell and J. P. Cook.

Water resource data that are useful to environmental scientists and planners frequently are missing, incomplete, or obtained irregularly. A new source of surface hydrological information can be obtained as often as every 18 days in some areas through machine-processing of Earth Resources Technology Satellite (ERTS) multispectral scanner data. This research focused on the surface water resources of a large metropolitan area. Marion County (Indianapolis), Indiana, in order to assess the potential value of ERTS spectral analysis to water resources problems.

062276 Techniques and Applications for Computer-Aided Analysis of Multispectral Scanner Data by R. M. Hoffer.

Several procedures for digitally processing and analyzing data from satellite scanner systems that have been found to be particularly useful are described. The techniques were applied to a mountainous test site of approximately one million hectares in area. In spite of the vegetation and topographic complexity of this test site, coniferous and deciduous forest cover, as well as other major cover types, could be mapped with an accuracy of approximately 85%, using both LANDSAT and SKYLAB data. Individual forest cover types were mapped with approximately 70% accuracy. Accurate acreage estimates of forest cover were obtained through use of these techniques over large geographic areas.

082776 An Investigation of the Relationship Between Spectral Reflectance and the Chemical, Physical, and Genetic Characteristics of Soils by O. L. Montgomery, M. F. Baumgardner, and R. A. Weismiller.

The purpose of this study was to examine the quantitative relationships between some of the most common chemical/physical properties and the reflectance (0.53 - 2.37μm) of 56 soils selected to represent a broad range of parent materials, climate, and drainage characteristics. Step-wise multiple regression analysis revealed that cation exchange capacity and contents of silt, clay, iron oxides and organic response. This study indicated that the middle infrared region of the spectrum is the best region of reflectance for evaluating these relationships.

The following Information Notes were published in full or in part under NASA Contract NAS9-14016.

052075 Description and Operation of a Field Rated ERTS-Band Transmissometer by D. P. DeWitt and B. F. Robinson.

This report describes a field rated instrument for the measurement of normal hemispherical transmittance at four wavelength bands: 0.5-0.6μm, 0.6-0.7μm, 0.7-0.8μm, and 0.8-1.0μm. The instrument consists of a detector system and a transmittance attachment comprised of an integrating sphere with collimator. The principle of measurement permits direct comparison of field results with a laboratory spectrophotometer such as a Bechman DK-2A spectrophotometer with integrating sphere attachment.

031276 An Analysis of Metropolitan Land-Use by Machine Processing of Earth Resources Technology Satellite Data by P. W. Mausel, W. J. Todd and M. F. Baumgardner.

The technology available at Purdue University's Laboratory for Applications of Remote Sensing (LARS) to classify earth surface features from multispectral data is sophisticated. This paper describes the results of a successful application of state-of-the-art remote sensing technology in classifying an urban area into its broad land-use classes. This research proves that numerous urban features are amenable to classification using ERTS multispectral data automatically processed by computer. Furthermore, such automatic data processing (ADP) techniques permit areal analysis on an unprecedented scale with a minimum expenditure of time. Also, classification results obtained using ADP procedures are consistent, comparable, and replicable; hence many spatial analysis problems caused by human errors or decisions are eliminated. The results of classification are compared with the proposed USGS land-use classification system in order to determine the level of classification that is feasible to obtain through ERTS analysis of metropolitan areas. (Anderson, Hardy, and Roach, 1972, 6).

032576 Urban Land Use Monitoring from Computer-Implemented Processing of Airborne Multispectral Data by W. J. Todd, P. W. Mausel, and M. F. Baumgardner.

Machine processing techniques were applied to multispectral data obtained from airborne scanners at an elevation of 600 meters over central Indianapolis in August, 1972. Computer analysis of these spectral data indicate that roads (two types), roof tops (three types), dense grass (two types), sparse grass (two types), trees, bare soil, and water (two types) can be accurately identified. Using computers, it is possible to determine land uses from analysis of type, size, shape, and spatial associations of earth surface images identified from multispectral data. Land use data developed through machine processing techniques can be programmed to monitor land use changes, simulate land use conditions, and provide "impact" statistics that are required to analyze stresses placed on spatial systems.

051576 Stratification of Landsat Data by Clustering by M. E. Bauer and B. J. Davis.

Full realization of the potential advantages of the synoptic coverage provided by Landsat will require the development and use of data analysis techniques which take into account the large variation and diversity of patterns found over many Landsat scenes. Stratification of the scene into units which are internally homogeneous is recommended as a first step in the analysis of data for whole or multiple frames of Landsat data. The use of clustering as an objective and efficient method of dividing scenes into areas which are spectrally similar (strata) is discussed and initial results, including classification performances and comparisons of spectral strata with major physical factors, are presented. Published in the Proceedings of the Third Symposium on Machine Processing of Remotely Sensed Data, June 29-July 1, 1976, Purdue University, W. Lafayette, Indiana.

052576 Matrix of Education and Training Materials in Remote Sensing by J. C. Lindenlaub and B. M. Lube.

Remote sensing education and training materials developed by LARS have been organized in a matrix format. A description of the matrix is followed by three examples designed to illustrate how the matrix can be used to synthesize training programs tailored to meet the needs of individual students. A detailed description of each of the modules in the matrix is contained in a "catalog" section.

062176 The Use of Spatial Characteristics for the Improvement of Multispectral Classification of Remotely Sensed Data by D. J. Wiersma and D. A. Landgrebe.

Two parallel and overlapping approaches to classification of remotely sensed data with the aid of spatial information are underway at the present time. The image processing approach attempts to model after the human visual system, while the second approach is primarily numerical. The technique of texture features[1,2], representing the image processing approach, and the sample classifier ECHO[3,4], representing the numerical approach are compared. The numerical approach is demonstrated to be superior in classification accuracy as well as being more efficient computationally.

090776 Analytical and Experimental Design and Analysis of an Optimal Processor for Image Registration by M. Svedlow, C. D. McGillem, and P. E. Anuta.

The problem of registration of LANDSAT images taken at different times is studied. Several preprocessing methods were evaluated to determine the best method for improving registration accuracy. Alternative correlation methods were also evaluated. Theoretical bounds on registration error were derived and an evaluation of the effects of geometric distortion on registration accuracy are evaluated.

110976 An Empirical Study of Scanner System Parameters by D. A. Landgrebe, W. Simmons, and L. Biehl.

The selection of the correct combination of parametric values (instantaneous field of view, number and location of spectral bands, signal to noise ratio, etc.) of a multispectral scanner is a complex problem due to the strong interrelationship these parameters have with one another. In this paper the results of an empirical study using the proposed Thematic Mapper parameters are presented. The results obtained shows that as the IFOV is decreased, classification accuracy declines slightly but mensuration accuracy improves, among other conclusions.

111076 Systematically Disseminating Technological Information to Potential Users by J. D. Russell.

The article describes the activities of the Technology Transfer program area at LARS. The Matrix of Instructional Materials is presented and each of the types of instructional materials used is described briefly. The instructional development sequence for educational materials is also presented.

The following Information Note was published in part under NASA Contract NAS9-14970.

070676 Determining Land Use Patterns Through Man-Machine Analysis of LANDSAT Data - A Tutorial Simulation by Stevan J. Kristof, James D. Russell, Tina K. Cary, Bruce M. Lube, Richard A. Weismiller.

The tutorial simulation provides a step-by-step description of the typical steps in the analysis of remotely-sensed data for determining land use patterns using numerically-oriented pattern recognition techniques. The material emphasizes the respective roles of the analyst and the computer, describing the decisions made by an analyst and the various computer processing functions. The steps are documented with illustrations and an example analysis from the Texas Coastal Zone.