033176

# QUARTERLY REPORT

Reporting Period: January 1, 1976 - March 31, 1976
Contract Number: NAS9-14016
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
Applied Physics Branch
Attn: A. E. Potter, Mail Code TF3
Houston, Texas 77058
Mark for: Contract NAS9-14016

## 2.1 Layered Classifier Adapted to Multitemporal Data Sets

The Multitemporal/Multisegment Problem. We reported earlier that the classifier logic developed under a set of moderately restrictive assumptions about the nature of segment coverage has provided marginally satisfactory results. The data studied in reaching that conclusion was CITARS data from Illinois and Indiana, characterized by very small field sizes. Registered data for the LACIE intensive test sites in Finney, Morton, and Ellis counties are now available. The associated "ground truth" overlays have been used to extract a set of fields to be used for training and test, and analysis of the data is in progress. We are having significant problems, however, because of the low quality of the ground truth data. There is still not a LACIE data set available from NASA of sufficient quality for research purposes. This was highlighted as an important problem at the March, 1976, LACIE project review at JSC.

Layered Classifier Software Improvement. Implementation of a LARSYS-compatible version of the basic layered classifier software was carried out this quarter. A copy of the control card summary is shown in Figure 1. The *XLAYER program will implement a broad class of layered classifier logic. However, no provision is made in this program for deriving such logic. Rather, the program requires as input a description of a "decision tree" which has been developed by an optimal layered classifier design procedure (see Figure 1) or some other method.

At this juncture, a number of important task options remain in the software upgrade effort and we are in the process of prioritizing them. The documentation effort is certainly top priority and is being participated in jointly by data analysis research personnel as well as software development personnel. Other important tasks include (1) developing a results tape format which adequately documents the layered classifier structure used to produce a given result, and (2) improving the software which is used to design the classifier decision logic.

A documentation effort has been initiated and efforts are proceeding to assemble both software-oriented documentation and user-oriented documentation.

Optimal Classifier Design. The goal of this effort is to develop a design procedure for producing layered classifiers which minimize classification cost (computer time) while maximizing classification accuracy.

Previous research has identified four key steps in the design process which we identify as (1) feature selection, (2) cluster analysis, (3) specifying the classifier, and (4) tree search. The first of these steps accomplishes assignment to a given node of the decision tree the features to be used in the class discrimination at that stage of the classification. The second step partitions the set of classes into the class subsets to be discriminated at the node (the first and second steps must be carried out concurrently for optimal design). The third step determines the degree of complexity to be implemented in the decision process at a node. And the fourth step implements the search through all candidate trees for the desired (and, ideally, optimal) tree.

```
REVISED  02/01/76
                            LARSYS CONTROL CARDS

                               LAYERED CLASSIFIER

R
E KEY          CONTROL
O WORD(COL.1)  PARAMETER                     FUNCTION                DEFAULT
- -----------  ---------- --------------------------------- -------------------

+ *XLAYER         (NONE)       SELECT LAYER CLASSIFICATION    (NONE)
                                 FUNCTION.

+ RESULTS                      DESTINATION OF RESULTS         (NONE)
               TAPE(XXX)       PUT ON TAPE XXX,
               FILE(FF)        FILE FF.
               INITIALIZE      INITIALIZE FILE ONE OF A       SEE CONTROL CARD
                                 NEW RESULTS TAPE               DICTIONARY
                                 (REQUIRED WHEN USING A
                                 NEW TAPE).
               DISK            RESULTS WILL BE STORED ON
                                 LARSYS DISK.

  PRINT        STATS           PRINT STATISTICS TO BE USED.   NO STATISTICS
                                                                PRINTED
               MAP             PRINT RESULTS MAP.             NO MAP
                                                                PRINTED
               NOFIELDS        NO TRAINING FIELDS PRINTED.    TRAINING
                                                                FIELDS PRINTED

  CARDS        READSTATS       STATISTICS FILE WILL BE        STATISTICS EXPECTED
                               INPUT ON CARDS.                  FROM DISK

  DATA         -----START OF DATA DECK---------------------------------------
               I                                                              I
               I               PUNCHED STATISTICS FILE FROM STATISTICS        I
               I               FUNCTION IF 'CARDS READSTATS' CONTROL CARD     I
               I               IS INCLUDED.                                   I
               I                                                              I
               ---------------------------------------------------------------

+ DATA         -----START OF DATA DECK---------------------------------
               I
               I DECISION DECK USED FOR THE LAYERED CLASSIFIER
               I
               I CARDS ARE FREE FORMAT ON EACH CARD BUT CARD ORDER
               I MUST BE PRESERVED.  DECK IS ALWAYS REQUIRED.
               I
               I DATA CARD FORMATS----
               I
               I TREE TOP(NODEA,NODEB,NODEC),NODEA(NODE1,NODE3,NODE4)
               I TREE NODEB(NODE2,NODE5), ETC
               I
               I FEATURES TOP(X,Y),NODEA(Y,Z),NODEB(W,X,Y), ETC
               I
               I REPRESENT NODE1(I,J),NODE2(K,L,M),NODE3(N),NODE4(O)
               I REPRESENT NODE5(P), ETC
               I
               I RENAME CLASS1(1/I,J/),CLASS2(2/K/),CLASS3(3/L,M/), ETC
               I
               I THE TREE CARDS DESCRIBE THE NODE SEQUENCE OF THE
               I DECISION.  THE FEATURES CARDS DESCRIBE THE CHANNELS
               I USED AT THE EACH DECISION NODE.  THE REPRESENT
               I CARDS ARE USED TO INDICATE THE REPRESENTATIVE CLASS
               I USED.  THE RENAME CARDS ARE USED THE ASSIGN A NEW NAME
               I TO GROUPS OF INPUT CLASSES.
               I
               I TOP IS A REQUIRED NAME ON THE TREE AND FEATURES
               I CARDS.  THE NODE AND CLASS NAMES MAY BE UP TO EIGHT
               I CHARACTERS.
               I
               ---------------------------------------------------------

+ DATA         -----START OF DATA DECK---------------------------------------
               I                                                              I
               I               FIELD DESCRIPTION CARDS DESCRIBING AREAS TO    I
               I     BE CLASSIFIED (ALWAYS REQUIRED).  EITHER FORM OF THE     I
               I     FIELD DESCRIPTION CARD MAY BE USED.                      I
               I                                                              I
               ---------------------------------------------------------------
```

**Figure 1. Control Card Listing for *XLAYER**

During this quarter, detailed consideration has been given to each of these phases of the decision tree design process and a list developed of promising candidate methods for accomplishing each. (The existing procedure developed by Wu at LARS is only one of many possible and is known to give suboptimal results.) The balance of the contract period will be used to begin to determine which of the candidate methods can be synthesized most effectively into the desired optimal design procedures.

## 2.2 Development of Spectral Strata from Clustering Techniques

Spectral stratification activities during this quarter included work in three areas: (1) partioning and classification of LACIE segments, (2) physical factor analysis, and (3) ancillary data registration. The majority of the effort was devoted to the first of these activities.

Partioning and Classification. The partitioning effort requested by NASA/JSC in November has been completed. The two data sets provided by JSC (LACIE operational and test and evaluation) have been divided by biophase and partitioned within both crop reporting district and larger regions. Partial results have already been sent to UCB and ERIM, and the complete set will be sent to UCB, ERIM, and JSC as soon as it is compiled.

The crucial need in this area at present is an evaluation of the partitions already obtained. Without this, we can not effectively attempt to improve the partitioning algorithm. The effort to evaluate the partitions by comparison of classification performances for local and non-local training has been severely hampered by lack of an adequate data set. Consequently, the results are inconclusive. The present data sets have several problems: too few segments within some strata for an effective evaluation, mislabeled field coordinates, too few data points for training, and data sometimes is not available in the correct form. The acquisition of an adequate data such as previously described by us is strongly recommended.

Physical Factor Analysis. The work on determining the correlation of physical factor (soil, crop maturity, precipitation, yield, and land use) with the spectral strata has been delayed by the partioning and classification effort. As reported last quarter the necessary physical factor data has been gathered. This quarter the design for the regression analysis was completed. A difficulty in implementing the analysis is caused by the fact that factors such as soil association and land use are qualitative, and available on a pixel basis while other factors such as precipitation and yield are quantitative and are observed for discrete geographic points (e.g., weather station and county). During the next quarter the regression analysis will be run on a test basis for one Landsat scene.

Registration of Ancillary Data. Processing required for overlay of physical data onto LANDSAT data for sites in Kansas is proceeding although significant delays are being encountered in the final overlay step. The first physical variable digitized was the generalized soil association map for the state of Kansas. The 1:125,000 scale map was compiled by the Kansas Agricultural Experiment Station and the U.S. Soil Conservation Service in 1973. There are 42 associations in the map compiled into eleven soil groups. The soil boundaries on the map were digitized on a digitizing table, gridded to the LANDSAT pixel size, and the final registration is in progress. The registration is being carried out initially for one frame of LANDSAT data which includes LACIE segments in Central Kansas. This registration task was completed at the end of March. An additional activity was carried out to enable a check of the table digitizer output. A CALCOMP plotter was used to regenerate the soil boundaries from the digital data. A map was drawn by the CALCOMP and this was overlayed on the original map to check the lines.

The second physical variable to be digitized from map data was land use. A Kansas Land Use Patterns map at a scale of 1:1,000,000 generated by the Space Technology Laboratory of the University of Kansas in cooperation with the Planning Division of the Kansas Department of Economic Development was used. This map has twelve classes of land use and was prepared from LANDSAT-1 imagery. All units were digitized for the entire state and the results gridded to the LANDSAT pixel size. A check CALCOMP output was generated and the data verified. This data is currently in the process of being prepared for registration and completion of the overlay is expected by April 30, 1976.

Three other variables are being processed currently. Precipitation, temperature, and yield data are available on a county basis. County boundaries are being digitized for Kansas and the temperature, precipitation, and yield data will be combined with the county data to form three or more channels of data. These overlays will be started when the soil and land use variables are completed. At the present rate of processing the five variables will probably not be completed until the end of May 1976. Overlay of the several other frames originally specified in the plan of work can be done much more rapidly since the digitized physical data will be available for the entire state.

## 2.3 Field Measurements for Remote Sensing of Wheat

Field measurements activities during this quarter have included data processing, data quality evaluation, management of the data library, data analysis, 1976 experiment planning, and technical coordination and leadership of the project. Work in support of the remote sensing experiments for analogous vegetative areas in the U.S. and Societ Union has also been conducted.

Data Processing. During this period, major emphasis was placed on software implementation of solutions to data anomolies, software changes to accommodate modified data processing procedures, and data quality evaluation. Processed data were received from JSC for LANDSAT, MSDS, FSS, and FSAS Systems and from ERL for the Exotech 20D System.

Implementation of an algorithm formulated by Barrett Robinson to correct a sensor offset problem is nearing completion. This correction will effect all Exotech 20C spectroradiometer collected to date for the Field Measurements project. Quality evaluations of FSS data showed a similar offset problem and correction during data library entry is under consideration.

FSS data library entry software has been modified to average the scans collected for a single pass over a field. This requirement developed after preliminary analysis of FSS data. The averaging process is designed to statistically delete atypical scans from the average. This software modification is now being tested.

Exotech 20D spectroradiometer data received from ERL is being trial processed into the field measurements data library for quality evaluation. Associated Exotech 20D ground observation data are being keypunched for library entry. Reflectance calibrated FSAS data collected in Kansas have been received and are currently being processed for entry into the data library.

Data Quality Evaluation. A document describing our recommendations for data quality evaluation and verification has been prepared. We are recommending greater emphasis be placed on this phase of the project during 1976 and believe that if the procedures and techniques recommended are implemented that the quality of data will be substantially increased. In particular, timely data checks will indicate inadequate procedures or poor sensor performance which can be corrected for subsequent missions and the results of the data evaluation can be used by data analysts (users) to determine which data meets their requirements.

We continue to study data from the spectrometer systems with the objective of improving current and future data quality. A procedure to evaluate the FSS data has been developed; preliminary results show a large variation in the calibration data being used.

Data Library. All data acquired during 1974-75 have been received and cataloged. These data are being processed for final entry and archival in the library. Some of the 1975-76 data has also been received. A schematic diagram showing the organization of the data library is shown in Figure 2.

Data Analysis. Analysis of field measurements data is being pursued on three fronts. The first is development of statistical analysis techniques for determination of the effects of agronomic variables such as leaf area index, biomass, and crop maturity on the multispectral reflectance of crops. These techniques and results will be described in the contract's final report.

Secondly, data from an experimental procedure utilizing a laser probe technique have been used to study a wheat canopy. The technique shows considerable promise as a means of characterizing the physical characteristics of crop canopies which when combined with incident radiation measurements will provide valuable information about the interaction of solar energy and crop canopies. In addition, major progress is being made on the analysis of wheat reflectance measurements made as a function of crop maturity (date), time of day (sun angle), azimuth angle, zenity angle, and wavelength.

Thirdly, the analysis of the thermal modeling measurements made last summer at Williston is progressing well. During the past two quarters, the thermal modeling activities have been centered on two areas: the development of simplistic, but useful models to relate canopy temperature profiles to radiometrically or remotely sensed temperatures and the analysis of several thermal experiments, conducted at the Williston test site.

The thrust of the temperature modeling study is to relate canopy geometry parameters with solar absorption and temperature distribution. The radiative transfer equation is being used to describe the phenomena by treating the canopy as non-homogeneous with directional properties. Of particular interest will be the application of these models toward understanding the effects of percentage ground cover and observation direction on the remotely sensed temperature measurements.

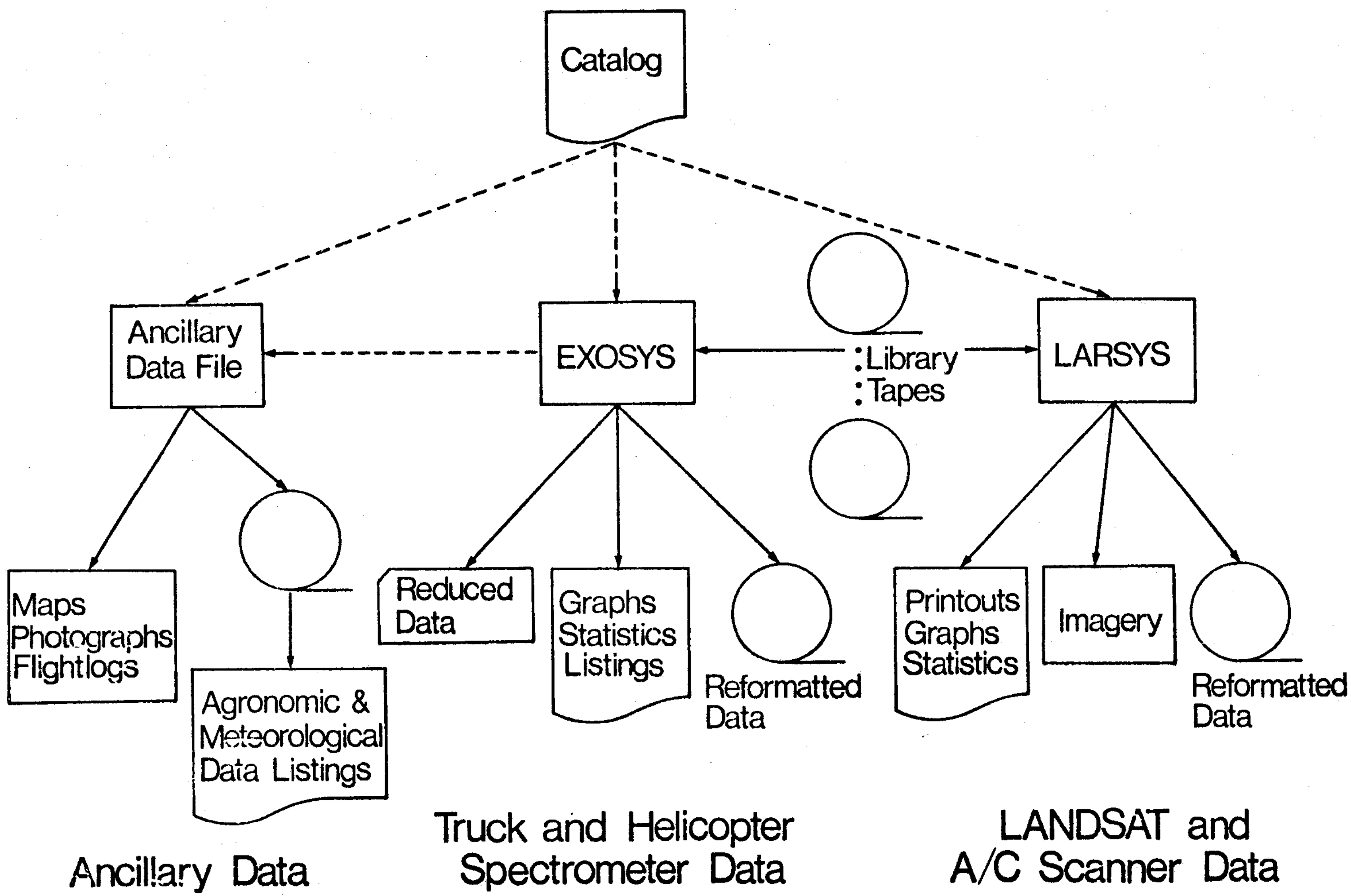
The analysis of the thermal experiment conducted at Williston is nearly 60% completed. The objectives of these experiments were: (1) to measure and identify the spectral characteristics of the radiant of crop canopies from 2.7 to 14μm, that are relevant to remote sensing applications; (2) to measure the daily transient temperature profiles in selected crop canopies and interpret these profiles in terms of agronomic and environmental parameters on a temporal basis; (3) to relate the measured spectral radiance to canopy temperature profiles. It is our intention to briefly summarize the major results of this study within the next few weeks.

Work will continue on these two areas during the next quarter. A major portion of our effort will be devoted to the design of thermal experiments to be conducted during the forthcoming growing season.

Project Leadership and Coordination. A field measurements project status report was prepared and distributed in January. In early March a

Figure 2.

## FIELD MEASUREMENTS DATA LIBRARY



meeting of all project participants was held at Texas A&M University for the purpose of bringing people up-to-date on the status of all phases of the project and to discuss and finalize plans for the project during the coming year. A key issue at the meeting was the discussion of data quality and verification. Following this meeting LARS has prepared a document describing our recommendations for evaluating data quality.

Remote Sensing Experiments for Analogous Vegetative Sites in the United States and Soviet Union. We have continued to support this project during this quarter. A major effect was devoted to preparation of the multispectral scanner data for exchange. The S-191 data and multispectral photography have also been prepared for exchange. In addition, documentation describing the sensor systems, data acquisition procedures, and data processing procedures are being prepared.

## 2.4 Scanner System Parameter Study

Major Activities. During the third quarter activities were concentration principally in the areas of a thematic mapper simulation study, an analytical multispectral scanner system design, and a scanner systems modeling study. Work on the development of an experimental spectral-spatial data array has been temporarily suspended to permit work on the thematic mapper at the request of NASA.

### 1. Thematic Mapper Simulation Study

Preparation of Simulated Data. A total of seven flightlines (see Table 1) of the 24-channel scanner (MSDS) data collected over the Kansas and North Dakota supersites during 1975 were processed to simulate spacecraft data for the purpose of obtaining additional quantitative support for the selection of engineering specifications of the thematic mapper instrument. Each of the seven MSDS flightlines were spatially degraded to simulate four resolutions- 30, 40, 50, and 60 meters. The MSDS channels selected were those that most nearly represent those chosen for the thematic mapper at this time (see Table 2). The most serious discrepancy in aligning channels between those in the MSDS and those in the Thematic Mapper was Thematic Mapper channel 2, .52-.60μm. The .53-.57μm channel in the MSDS was not available, so the .57-.63μm channel had to be substituted as the next nearest channel. The MSDS data was also scaled to simulate the proposed Thematic Mapper dynamic range.

Because of the noise problems in the MSDS data and one of the types of noise present (banding) it was not possible to simulate the proposed Thematic Mapper noise levels exactly. However, the calculated noise present in the simulated data was of the same order as that proposed for the Thematic Mapper. To study the effects of added amounts of noise in the data, seven levels of calibrated noise (.0025, .0050, .0075, .010, .015, .020, .030, NF$\Delta\rho$) were added to the 30 and 40 meter spatially simulated data for two flightlines, flightline #3, 7/6-Finney County and flightline #1, 8/15-Williams County. All of the previously mentioned data sets, a total of 56, were made available to GISS, ERL, and NASA/Goddard who are participating in the analysis of the data with Purdue/LARS. The present processing tasks are concerned with simulating LANDSAT C for four of the MSDS flightlines used before.

Analysis of Simulated Data. Analysis of the simulated data has concentrated on three flightlines (see Table 1), two of which have complete and one which has partially complete first order results. Three parameters have been studied - spatial resolution, noise, and to a limited extent spectral range.

Spatial Resolution Parameter. The analysis technique for the spatial resolution study included selecting training areas to represent the spectral classes from cluster maps which represented one half of the flightline. The multivariate Gaussian statistics were estimated using the previously mentioning training areas. This technique was followed for the analysis of each of the four simulated spatial resolutions. The flightline was then classified using simulated Thematic Mapper channels- 2,3,4,5,6, and 7. The

Table 1

MSDS Data Selected for Simulation

| Site | Date | Flight Line | Comments |
|---|---|---|---|
| * Williams County, N.D. | 6/22/75 | 1 | Poor crop calendar date |
| Williams County, N.D. | 6/22/75 | 3 | Poor crop calendar date |
| Finney County, KS | 7/6/75 | 1 | Serious banding |
| * Finney County, KS | 7/6/75 | 3 | Moderate banding |
| * Williams County, N.D. | 8/15/75 | 1 | Good Set |
| Williams County, N.D. | 8/15/75 | 2 | Good Set |
| Williams County, N.D. | 8/15/75 | 3 | Good Set |

* Those being analyzed presently

Table 2

Correspondence of Thematic Mapper and MSDS Channels

| Channel | Thematic Mapper | MSDS |
|---|---|---|
| 1 | .45 - .52 μm | .46 - .50 μm |
| 2 | .52 - .60 | .57 - .63 |
| 3 | .63 - .69 | .64 - .68 |
| 4 | .74 - .80 | .76 - .80 |
| 5 | .80 - .91 | .82 - .87 |
| 6 | 1.55 - 1.75 | 1.52 - 1.73 |
| 7 | 10.4 - 12.5 | 10.0 - 11.0 + 11.0 - 12.0 |
| 8 | .74 - .91 | .76 - .80 + .82 - .87 |

first order evaluation of the spatial resolution parameter includes four different criterions - training performance, test performance, root mean square (RMS) error of informational class proportion estimates for the flight line and average RMS error of informational class proportion estimate of all sections in the flight line (See Figs. 3, 4, 5, and 6).

The training performance is the overall classification accuracy of the training pixels. The test performance is the overall classification accuracy for field center pixels. The test fields were selected in the original six meter data by choosing the largest rectangular block of pixels that would fit within the agricultural field so that no boundary pixels were included. The test field boundaries were then found in the degraded spatial resolutions such that no "super" pixels containing boundaries were included. Some of the original test fields were discarded in this process because they were too small (ie. there were no pure field center "super" pixels).

The RMS error of informational class proportion estimates for the flightline was found by finding the percent of the flightline classified as a particular class and comparing it with the ASCS ground collected estimate using equation 1.

$$\text{RMS Error} = \sqrt{\frac{\sum_{i=1}^{N} (C_i - C_i')^2}{N}} \qquad (1)$$

where N = number of informational classes
$C_i$ = percent classified as particular informational
$C_i'$ = percent estimated by ASCS ground collected data

The fourth criterion stated was found using equation 1 on a section basis (approximately one square mile) and finding the average RMS error of all the sections in the flightline. Each flightline analyzed included a two by six mile area; so there are twelve sections in each flight The informational classes used for the three flightlines are given in Table 3.

LANDSAT 2 data for two of the flightlines were also analyzed; the results are given in Figures 3 and 4. Test fields have not been selected for the data.

The results indicate that there is little significant change in the training performance across the four resolutions. The test performance tended to go up slightly from 30 to 60 meters for two of the flightlines. This is suspected to be caused by the better signal to noise in the 60 meter data since more original six meter pixels are being combined to produce the 60 meter pixel.

Figure 3.
Spatial Resolution Parameter
Overall Training Performance
channels: 2,3,4,5,6,7
noise: .002-.008 NEΔP, .09 NEΔT for 30m 8/15-1
.004-.023 NEΔP, NEΔT for 30m 7/6-3

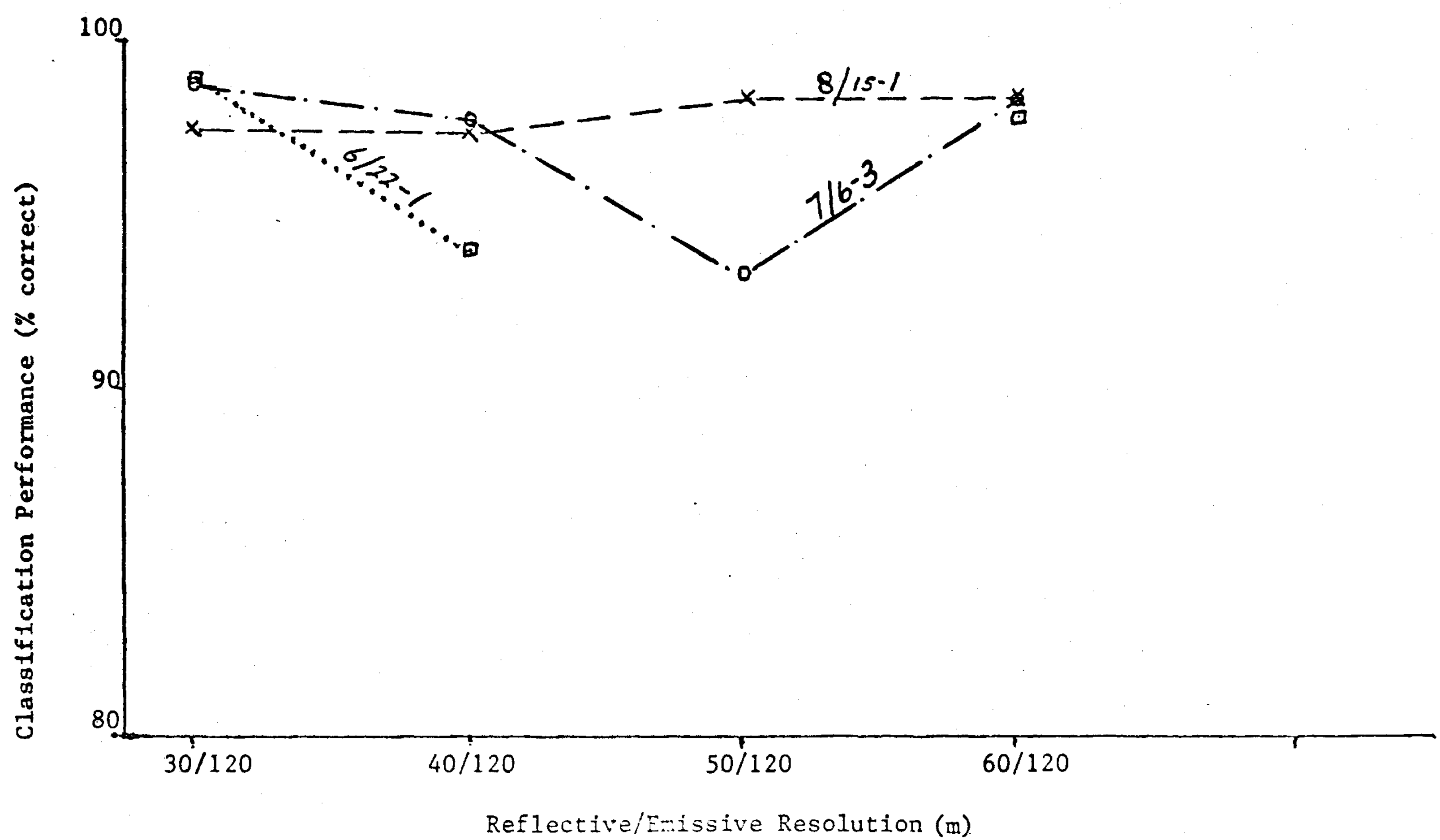


**Figure 4.**
Spatial Resolution Parameter
Overall Test Performance
Channels: 2,3,4,5,6,7.
**noise:** .002-.008 NEΔP, .09 NEΔT for 30m 8/15-1
.004-.023 NEΔP, for 30m 7/6-3

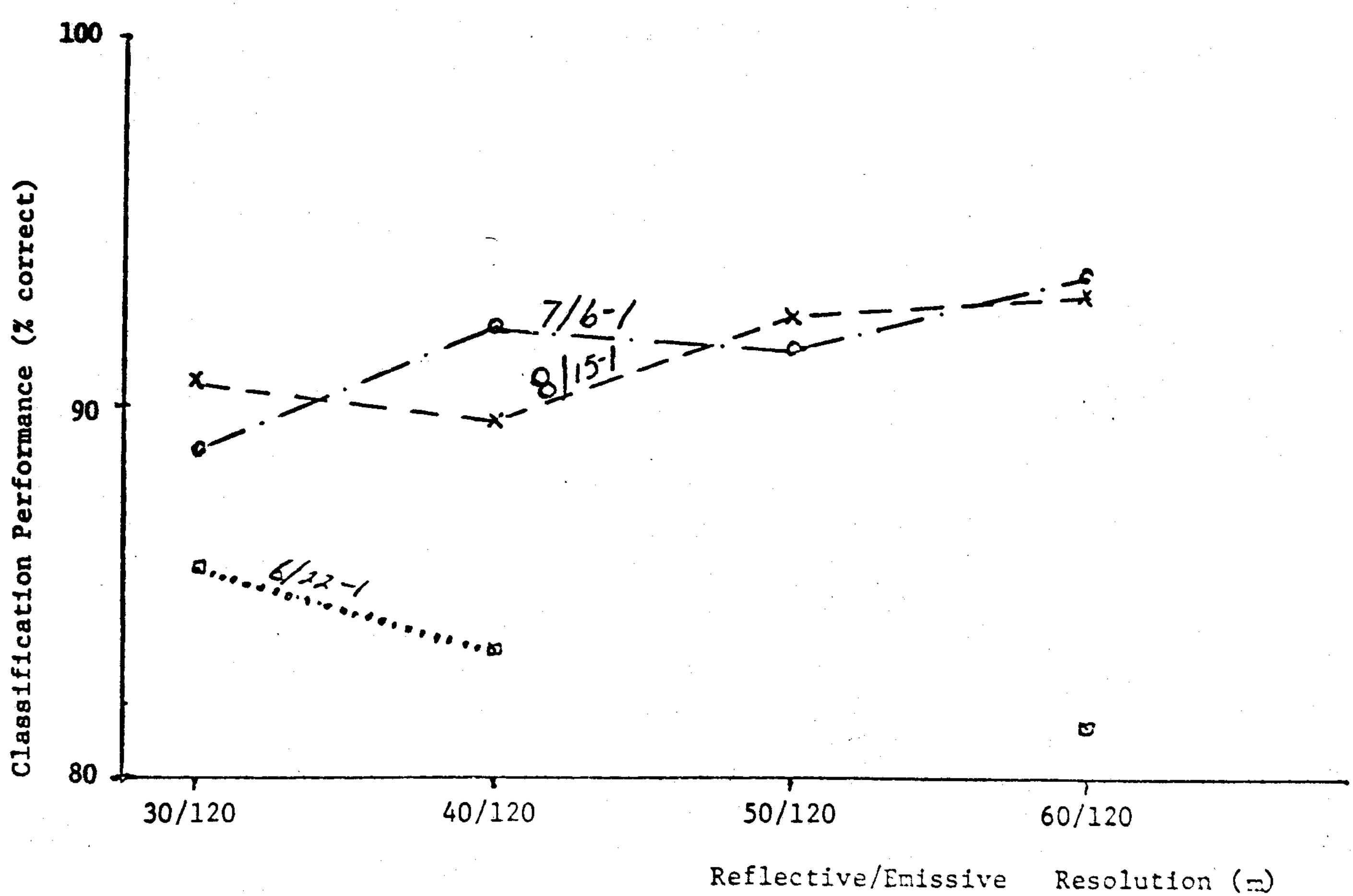


Figure 5.
Spatial Resolution Parameter
RMS Error of Preportion Estimate for Flight Line
channels: 2,3,4,5,6,7
noise: .002-.008 NE$\Delta$P, .09 NE$\Delta$T for 30 m 8/15-1
.004-.023 NE$\Delta$P, for 30 m 7/6-3

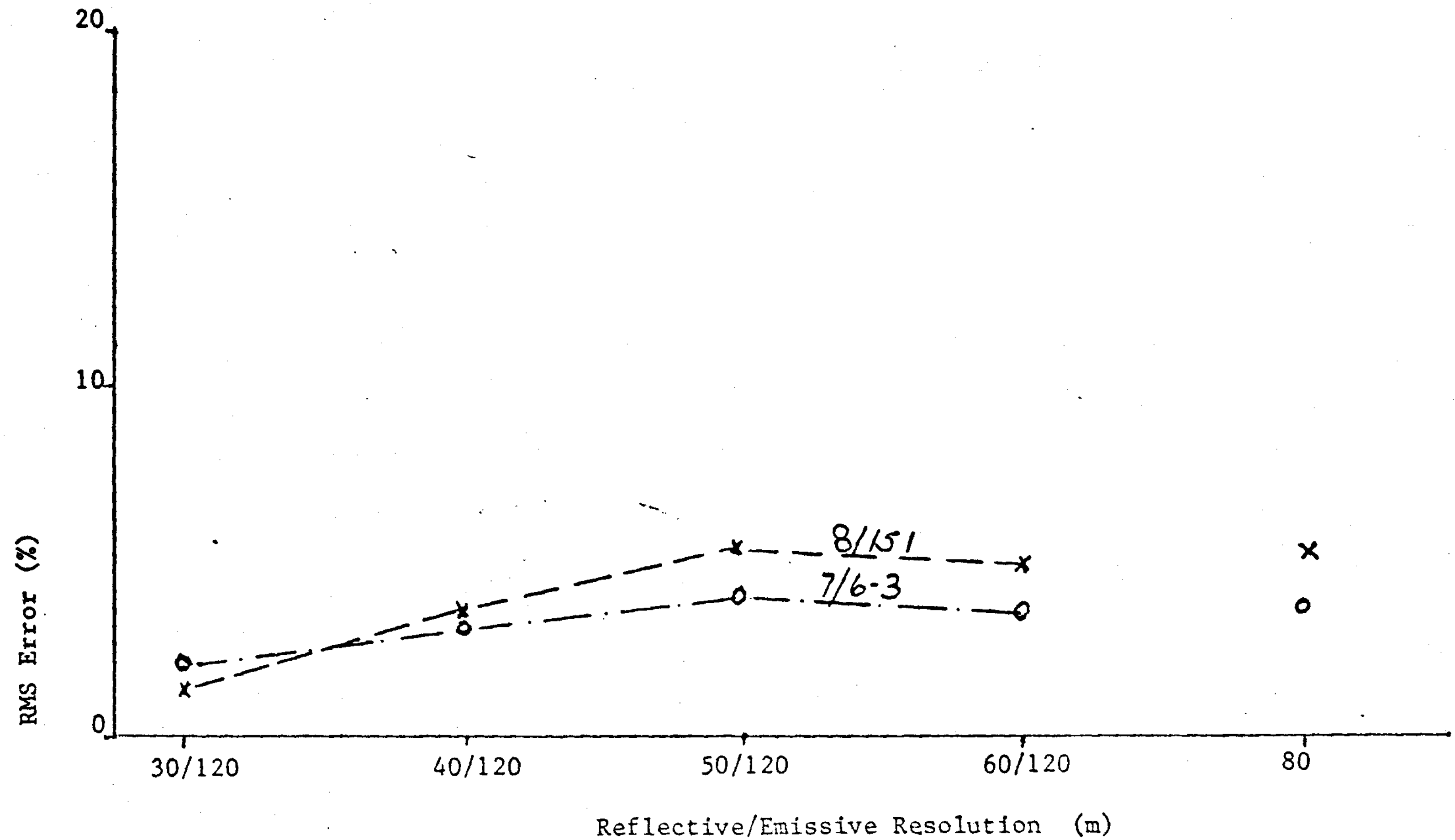


Figure 6.
Spatial Resolution Parameter
Average RMS Error of Proportion Estimates of Sections in Flight Line
channels: 2,3,4,5,6,7
noise: .002 - .008 NEΔP, .09 NEΔP for 30m 8/15-1
.004 - .023 NEΔP, for 30m 7/6-3

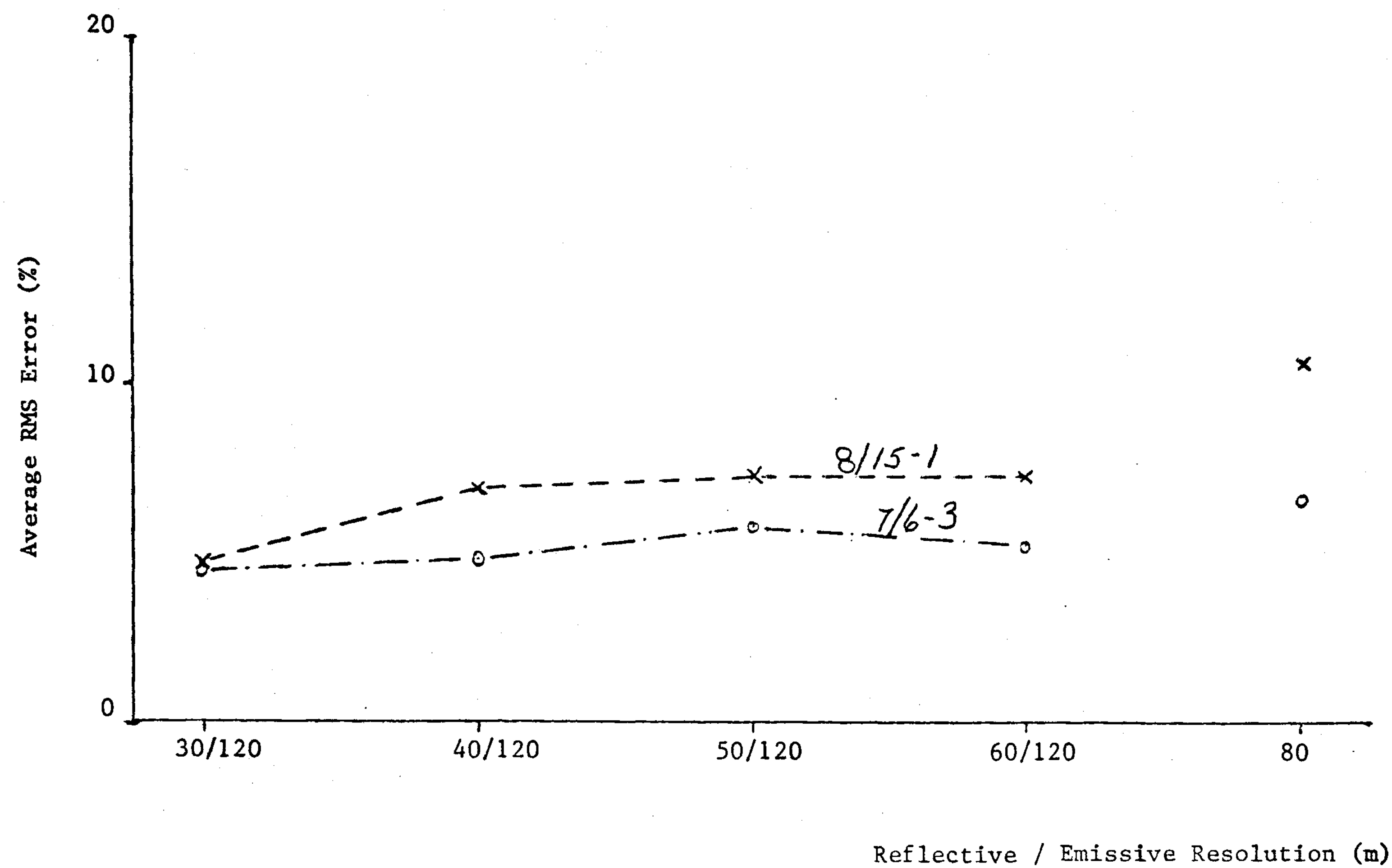


Table 3

Information Classes Used in Analysis

| Site and Date | Informational Classes |
|---|---|
| Williams County, N. Dakota 6/22/75 | Bare Soil, Grasses/Pasture, Small Grain |
| Finney County, Kansas 7/6/75 | Harvested Wheat, Corn, Grain Sorghum, Grass/Pasture, Fallow |
| Williams County, N. Dakota 8/15/75 | Harvested Wheat, Unharvested Wheat, Grass/Pasture, Fallow, Other (corn/oats) |

Care must be taken in evaluating the 6/22 Williams County, North Dakota flightline since the date is a poor one for delineating spring wheat, oats, and barley from fallow ground. Some of the small grain is just emerging from the ground. The bare soil class (see Table 3) included the combination of summer fallow and those small grain fields whose ground cover is principally bare soil. It was difficult to be objective in deciding which small grain fields should be included in the bare soil class and those which should be included in the small grain class. The planting date has been found to be the most consistant criterion. Some of the differences in the 6/22 data may represent the sensitivity in the selection of training fields in the four resolutions.

The RMS error of the proportion estimates for both the flightline and the average of the sections indicate that the least error is obtained with 30 meters. The RMS error increased at 40m and again at 50m and then leveled off. The RMS error for the actual LANDSAT 2 data sets are also shown.

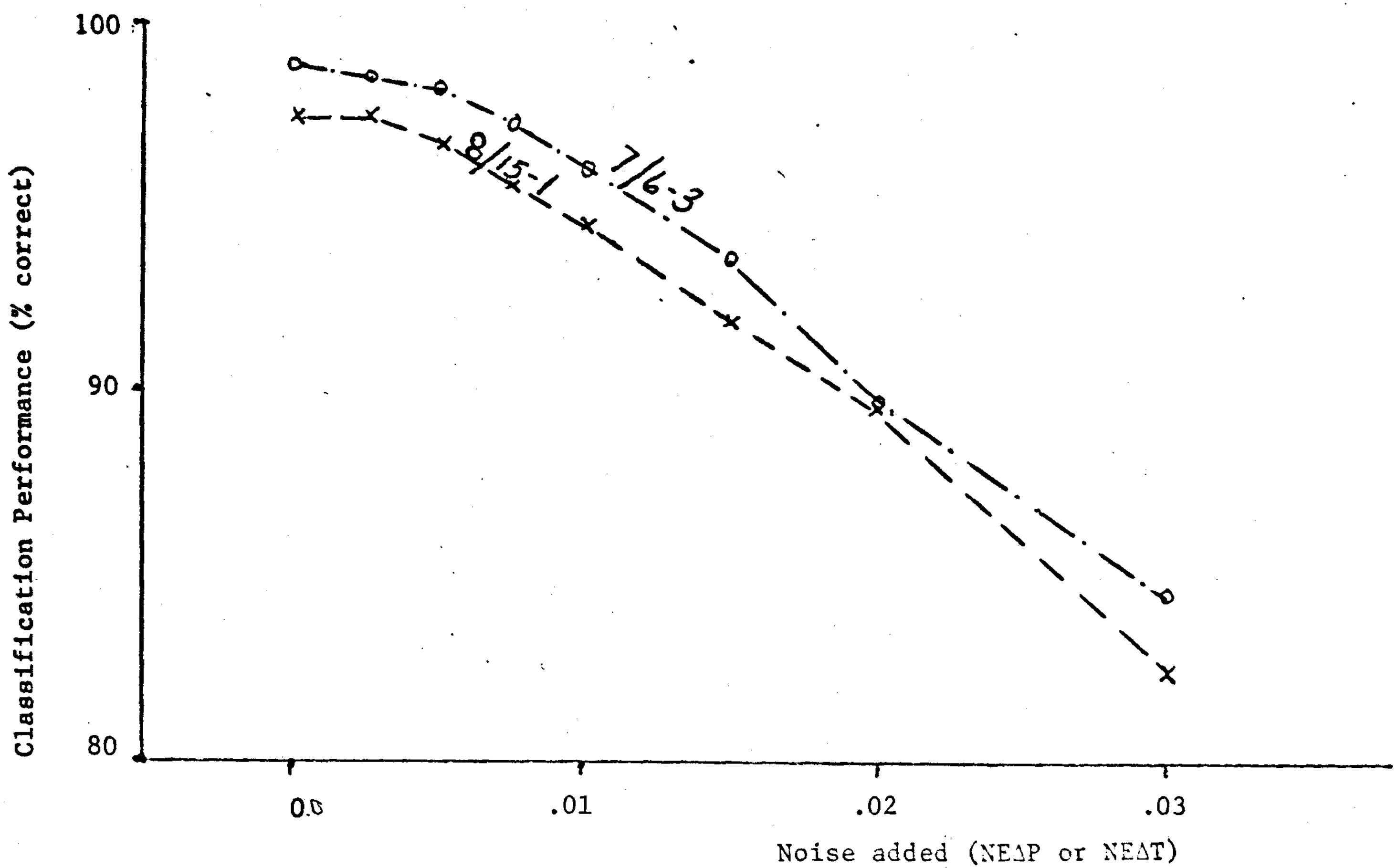
Noise Parameter. The analysis technique of the noise added data included using the training fields selected in the no noise added case and re-estimating the multivariate Gaussion statistics in each of the seven noise added data sets for a particular resolution and flightline. Each of the noise added data sets were then classified using channels 2,3,4,5,6, and 7, and their respective training statistics. To actually simulate different noise levels in data from satellites, the variable of analyst determination of field boundaries, could have been included. This would have necessitated the time consuming routine of the selection of training areas from each noise added data set independently from the other sets. Tests were run which illustrated that the data clustered nearly the same for the noise added levels of .0025 to .015 NE$\Delta$P added. The field boundaries, however, were difficult to distinguish in the cluster maps for the noise levels of .02 and .03 NE$\Delta$P added levels. In light of these results the performances found for the .02 and .03 NE$\Delta$P noise added levels may be optimistic, since the field boundary delineation difficulty is not included.

The evaluation criterions of the noise added levels include the training performance and the test performance, both very similar. The training performance for the 30 meter data of two flightlines is illustrated in Fig. 7. The performance stays relatively constant until the amount of noise added becomes greater than the noise already in the data.

Spectral Range Parameter. Little evaluation of the spectral range parameter has been done and the results are too inconclusive to be presented at this time. Classifications have been done using different sets of channels - 1,2,3,6,7, and 8 (4&5); 1,2,3,4,5,6,7; 2,3,4,5; and 2,3,4,5,6,7. Also the correlation of Thematic Mapper channels 4 and 5 is being studied.

Plans for Next Quarter. Analysis of the data will continue with a more detailed evaluation of the three flightlines analyzed to date. Also the simulated LANDSAT C data will be analyzed and evaluated.

Figure 7.
Noise Parameter
Overall Train Performance
channels: 2,3,4,5,6,7
resolution: 30/120 meters



2. Analytical Multispectral Scanner System Design

During this quarter various aspects of the scanner design problem were considered. A conceptual model of the physical process for which the scanner is to be designed was formulated. From this model a theoretical result was obtained which specifies an optimum system for the problem of classification of the multispectral data into one of several predetermined classes. This optimum system takes into account spectral characteristics of the classes to be identified, spectral bandwidth, location of the spectral bands within the spectrum, and other factors. Thus other multispectral scanner designs can be compared to the optimum scanner to form a basis for scanner system selection.

In the coming quarter this optimum system will be studied further to elaborate on the relation of the parameters of the theoretical optimum system and the parameters of a realizable scanner design.

3. Scanner Systems Modeling

Of the several aspects to the original scanner system study defined in the plan of work only the classification error model is receiving extensive attention. In this activity the state of the art in prediction of classification error was surveyed and an investigation into the M class N dimensional case for an optimum Bayes classifier is in progress.

The evaluation of the performance of various classifiers through calculation of exact probability of misclassification has been under study by many researchers for some time. Although empirical results have been obtained, the precise recognition rate for a Bayes classifier has not been evaluated for more than two normal classes with arbitrary covariance matrices (Anderson[1], Das Gupta[2], Funkunaga[3]). Here, by obtaining the marginal and joint distribution of discriminant functions, an attempt is made to find the exact probability of error.

---

[1] Anderson, T. W., Bahadur, R., "Classification into two Multivariate Normal Distributions with Different Covariance Matrices," Ann. Math. Statistics, Vol. 33, pp. 420-431, 1962.

[2] Das Gupta, S., "Optimum Classification Rules for Classification into two Multivariate Normal Populations," Amm. Math. Statistics, Vol. 36, pp. 1174-1184, 1965.

[3] Fukunaga, K, Krile, T., "Calculation of Bayes Recognition Error for Two Multivariate Gaussian Distributions," IEEE Transactions on Computers, March 1969.

Decision functions are:

$$\ell_i = (\underline{X}-\underline{\Gamma}_i)^T \underline{\Lambda}_i^{-1} (\underline{X}-\underline{\Gamma}_i) - C_i \qquad i = 1 \ldots M$$

choose class $M_i$ if $\ell_i$ is smallest. $\underline{\Gamma}_i$ and $\Lambda_i$: are mean vector and covariance matrix of class i and $C_i$ is a constant depending on a priori class probability.

Then under $M_1$:

$$P(\varepsilon|M_1) = 1-P(c|M_1) = 1-P_r\{\text{all } \ell_i > \ell_1 | \forall i \neq 1\}$$

$$= 1 - \int_0^\infty p(\ell_1) \left[ \int_{\ell_1}^\infty \cdots \int_{\ell_1}^\infty p(\ell_M, \ell_{M-1}, \ldots \ell_2 | \ell_1) \, d\ell_M d\ell_{M-1} \ldots d\ell_2 \right] d\ell_1$$

where: $P(\varepsilon|M_1)$ is the probability of error given the sample comes from $M_1$

$P(c|M_1)$ is the probability that the classification is correct.

To evaluate this expression, one needs the marginal and joint distribution of $\ell_i$'s.

Briefly under $M_1$, $\ell_1$ is distributed as chi-squared witn N degrees of freedom; others don't have well known distributions but their moment generating functions (m.g.f.) can be found using diagonalization of respective quadratic forms. For instance, under $M_1$ m.g.f. of $\ell_2$ has the following functional form

$$M_{\ell_2}(t) = \prod_{i=1}^{N} \left[ \frac{\dfrac{K_i t}{e^{1-2d_i t}}}{1-2d_i t} \right] \qquad |t| \leq \frac{1}{2d_i}$$

where $d_i$ and $k_i$ are known constants. An inverse FFT would supply the density function of $\ell_2$.

Joint distributions are found, also by using m.g.f. This method is currently being tested to see if it reduces to cases for which precise results are known.

## 2.5 Transfer of Computer Image Analysis Techniques (Remote Terminal)

Activities under this section of the work statement include continued support of the JSC remote terminal and the development of technology transfer activities and materials. In accordance with the milestones for this task activities scheduled for this quarter included Materials Development, Interim Training Programs and Remote Terminal Support. Each of these activities are described below.

Materials Development. A simulation exercise dealing with the analysis of LANDSAT data using computer-aided analysis techniques was finished during this quarter. This document, distributed as an interim technical report (LARS Information Note 012376), documents in detail the analysis of remotely sensed data for a forestry applications example. The example uses numerically-oriented pattern recognition techniques and emphasizes the way and manner in which the analyst combines his skills with the computational power of the computer to meet analysis objectives.

Work on a second analysis simulation exercise has begun. This report will describe the procedures used to analyze and map a coastal zone region. Again, the aim of the report will be to document the decisions and trade-offs made by an experienced analyst at each step in the analysis sequence. The report will stress the interaction between man (the analyst) and machine (the computer). The data set used for the simulation exercise has been drawn from among those analyzed in conjunction with the Regional Application Project task of this contract.

Interim Training Programs. No requests for training programs at JSC or LARS were received during this reporting period. However, to better facilitate our ability to respond to such requests, education and training materials which have been developed under NASA and Purdue University support have been cataloged in a consistent format. The catalog presently serves as an internal planning document and provides a mechanism for designing training programs to match the education and experience backgrounds of trainees with the goals of the training program. After suitable internal tryout, the catalog will be supplemented with an explanation of how it fits into the longer technology transfer goal to develop a matrix* of remote sensing educational materials and a description of representative sample training programs. This effort is expected to facilitate implementation of remote sensing training requirements at JSC.

Remote Terminal Support. Support of the JSC remote terminal was continued during the third quarter of the contract period. The operation ran smoothly and no problems arose.

---

* The concept of a matrix of educational materials spanning a variety of technical and application areas and ranging in detail from short briefing aids to detailed case studies was described in the Final Report of NASA Contract NAS9-14016, June 1, 1974 - May 31, 1975.

## 2.6 Research in Remote Sensing Technology

The research under this task has three activities: 1) Development of Ancillary Data Registration Techniques, 2) LANDSAT LACIE Site Imagery Enhancement and 3) Soil Reflectance Property Research. Also, a LACIE temporal registration request was made by JSC.

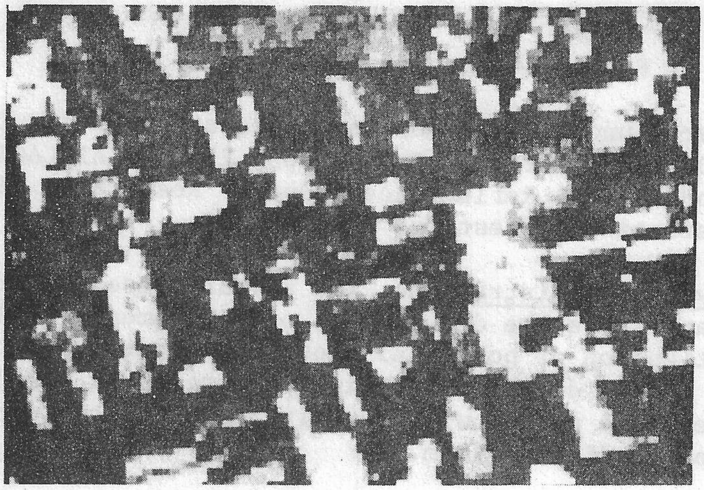
Ancillary Data Registration. The primary research oriented activity during the quarter was the development of refinements in labeling procedures for digitization of polygonal maps. A computer program was implemented to aid in the detailed labeling process for a complex land use map. The experience gained will contribute to the state of the art is converting maps to digital form.

Image Enhancement. The optimum IFOV enhancement filter developed in CY75 and discussed in previous reports was applied to several LACIE segments for further evaluation. In addition a modified program for cubic interpolation was generated and applied to the same data. Also, an existing blocking expansion program was applied to the data sets. Thus a comparison of blocking, cubic interpolation and IFOV filtering and expansion was obtained. The results using the IFOV indicate a significant sharpening while maintaining a "smooth" characteristic to the image as compared to blocking. The cubic causes considerable "blurring" and was considered not very desirable. Figure 8 contains a comparison of a LACIE segment in Barton County, Kansas imaged on June 12, 1974 for blocking, cubic, and IFOV filtering. In each case the expansion factor is 3x for columns and 4x for lines which approximately corrects for the LANDSAT sampling ratio. Six LACIE site enhancements with blocking, cubic, and IFOV for each were transmitted to JSC on March 16, 1976 for reproduction on the production film converter and evaluation by LACIE R&T personnel.

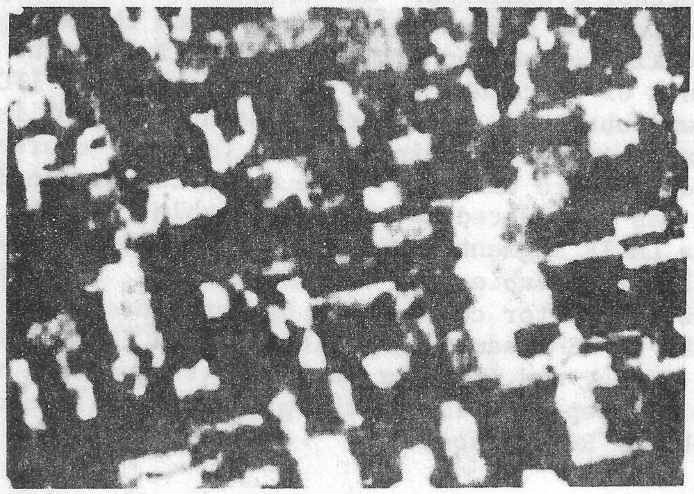
A cost comparison was made while computing the enhancements although the programs are not optimized to the extent they could be. Significant improvements are expected with specialized programming approaches. For a 192x192 point area expanded 3x4 to a 576x768 point area in 4 channels by each of the three algorithms the CPU computing time on an IBM 360 Model 67 computer was 15 sec. for blocking, 338 sec. for cubic and 965 sec. for IFOV filtering. The high cost of IFOV is due to the fact that 81 original points are used to compute each of the new points where with cubic only 16 are used.

Temporal Registration of LACIE Sites. A temporal registration task not in the work statement is progressing and completion of the processing is expected by May 15, 1976. Both nearest neighbor and cubic interpolation will be used in the registrations.

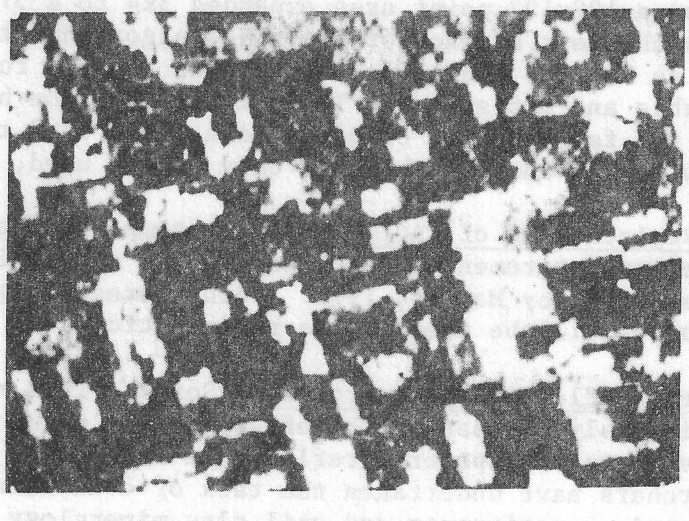
The Spectral Reflectance Properties of Soils. This research endeavor is continuously evaluating the effect of the physico-chemical properties of soil on the spectral reflectance of soils. Along this line the researchers have undertaken the task of evaluating the effect of climate, organic constituents and soil clay mineralogy upon the physical and chemical nature of the measured soil parameters. Initial results show that the significance of the soil parameters vary depending largely on climate, organic constituents and soil clay mineralogy.



Blocked Enlargement



Cubic Interpolation



IFOV Filtering

Figure 8 Comparison of Enhanced LANDSAT Imagery from Band 5 of LACIE Segment in Burton County, Kansas, June 12, 1974.

Recent results indicate that when climatic zones are integrated or combined the significance of the soil parameters decreases significantly as compared to single climatic regions. It is very difficult if not impossible to arrive at a quantitative value for the climatic factor. This is due in part to the independent response of each soil parameter to variation in climate and the interdependence of these parameters. The degree of success in classifying soil may be greater if done within climatic regions.

A comparison was performed between the results obtained in the visible region (0.48-0.71μm) and the near infrared (0.73-2.37μm). The near infrared appears to be the best area for studying soil properties. This may be due to diminishing overtones and molecular vibrations originating in the middle and far infrared. More study is required to confirm this hypothesis.

The results of the organic constituents have just been received. The effect of these parameters on the spectral reflectance of soils are presently being investigated. The results of these investigations will appear in the next report.

## 2.7 Forestry Applications of Computer-Aided Analysis Techniques

Introduction. Work at LARS is supporting the FAP by documenting and refining computer-aided analysis techniques for forestry applications. Under this broad objective, specific subtasks of 1) defining training procedures, 2) statistically evaluating results, 3) selecting optimum wavelength combinations, 4) analyzing multitemporal data, and 5) evaluating change detection methods, have been defined.

Accomplishments to Date. Documentation for the first task - training area selection - has been completed. The publication entitled: "A Forestry Application Simulation of Man-Machine Techniques for Analyzing Remotely Sensed Data" (LARS Information Note 012376) has been submitted for release to the public domain.

Statistical evaluation of classification accuracy activities have been directed toward documentation of procedures and evaluation of results. The foundation for the next stage in the evaluation procedure is being laid. The ultimate goal of any evaluation technique is to achieve a measure of mapping accuracy in which the user can place a high degree of confidence. Such a measure can be attained with the Stage II procedure, at an increase man-power cost. An approach is being sought which can maximize user confidence and minimize analyst time.

The wavelength band analyses completed with SKYLAB data collected over the San Juan National Forest has been augmented with NC-130 aircraft data. Wavelength bands were chosen for analysis that closely compare to the anticipated Thematic Mapper wavelengths. A report outlining the findings of these investigations is being prepared.

The primary effort in the multi-temporal analysis task has been involved with describing procedures to select channel/date combinations. This is a complex area of study requiring much more intensive analysis. A possible recommendation from this task would be to support additional efforts dealing with analyzing multi-temporal data sets. An important question to be asked before additional direction is given this task relates to the amount of improvement that can be expected when a multi-temporal data set is analyzed.

There will be no efforts aimed at attempting a change-detection classification during the remainder of the contract year. However, recommendations regarding procedures will be forthcoming.

Projected Accomplishments. Activities for the remainder of the contract year will be toward writing the final report.

2.8 Regional Applications Project (RAP)

The previous quarterly report discussed the preparation of the four data sets for the Port O'Connor, Pass Cavallo, Port Lavaca and Austwell quadrangles. Minor discrepancies were found and corrected in the four-date overlays for the Austwell and Port O'Connor quadrangles.

A short description of each of the four proposed change detection procedures was included in the previous report. All four procedures, Direct Change Detection, Delta Change Detection, Spectral/Temporal Change Classification, and Layered Spectral-Temporal Change Classification, are being applied to the Texas Coastal Zone data sets. The lack of coincident ground truth still is the major problem involved in locating change and evaluating the performance of the various procedures. This problem will not be overcome until additional data (LANDSAT and coincident ground truth) have been collected and incorporated into the data sets for the four quadrangle areas.

Initially, it had been planned to conduct the preliminary change detection development effort utilizing the November 1972 and December 1973 data sets. Since these two data sets lack only 18 days of being a year apart, the assumption of little or no seasonal changes nor sun angle effects would have been reasonable. This would have greatly simplified the development of procedures to detect permanent change. However, since neither data set had supporting ground truth, this approach was not feasible. Thus, the approach was modified to include the February 1975 data set, which includes supportive ground truth, in place of the December 1973 data. This introduced the problem of working with data sets containing large amounts of seasonal change and sun angle effects in addition to the permanent change of interest.

Because of the availability of a photointerpretive overlay prepared by Lockheed for the Port Lavaca quadrangle, work during this quarter was initially directed toward this quadrangle. The direct change detection procedure did indicate change was occurring within the Port Lavaca quadrangle area. However, the area being categorized as change was due primarily to seasonal changes and tidal variations. Comparison of simulated false color images of this area supported this inference. Since this quadrangle appeared to contain little permanent change, future work was directed toward the Port O'Connor area.

At about the same time, "Spectral Environmental Classification" overlays were obtained from NASA/JSC. These overlays are being used to compare the spectral classes classified by LARS to the photointerpretive classes described by LEC. Frequency distributions of the spectral classes versus photointerpretive classes are being prepared and will be utilized in grouping of the informational classes into the spectral classes. This aggregation will allow for a more manageable number of classes in the initial change detection technique development. If the spectral classes were not grouped, a classification containing 22 spectral classes could yield 462 possible change classes. Grouping to 6-8 classes reduces to 30-56 the number of change classes.

This exercise will also allow the analysts to have a good representation of the February classifications with regard to the relationships between spectral and information classes. However, since there is no concurrent ground truth for the '72 or '73 data, only inferences can be made concerning the combination of these spectral and informational classes.

Also during this quarter, the four quadrangle areas were classified using the "Extraction and Classification of Homogeneous Objects" (ECHO) algorithm in an effort to determine if this classification scheme might be applicable to the change detection. Even though this approach shows promise, study of the classifications led to the decision to restrict initial work to the per point classifier.

During the final quarter, work will be concentrated mainly within the Port O'Connor quadrangle. The four change detection procedures will continue to be evaluated. If time permits, the Austwell quadrangle will be included.

Information Notes Issued
During Period
January-March, 1976

121874 Variance Comparisons for Unbiased Estimators of Probability of Correct Classification by David S. Moore, Stephen J. Whitsitt, and David A. Landgrebe.

Variance relationships among certain count estimators and posterior probability estimators of probability of correct classification are investigated. An estimator using posterior probabilities is presented for use in stratified sampling designs. A test case involving three normal classes is examined.

052175 Experimental Examination of Similarity Measures and Preprocessing Methods used for Image Registration by M. Svedlow, C. D. McGillem and P. E. Anuta.

A variety of image registration algorithms have been developed over the past several years by various authors. This study examines the applicability of several of these to the temporal registration problem of LANDSAT data. The results indicate that most methods work well for highly correlated images but only a few are effective for the low correlation case. The magnitude of the gradient enhancement method coupled with a magnitude of the difference correlator is observed to produce the best results.

072475 Computer-Aided Analysis of LANDSAT-1 MSS Data: A Comparison of Three Approaches, Including a "Modified Clustering" Approach by M. D. Fleming, J. S. Berkebile and R. M. Hoffer.

Three approaches to computer-aided analysis of LANDSAT-1 MSS data were evaluated utilizing data from a test site in rugged, mountainous terrain. The approaches compared included non-supervised (clustering), modified supervised, and modified clustering. The modified clustering approach proved to be the optimal technique because of minimal computer time, highest classification accuracy, and most effective analyst/data interaction. This paper was presented at Purdue University's Symposium on Machine Processing of Remotely Sensed Data, June 1975.

012176 Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS) Project. by R. Bizzell, F. Hall A. Feiveson, M. Bauer, B. Davis, W. Malila and D. Rice.

The CITARS task design and objectives are reviewed and final results are presented, together with conclusions and recommendations. It was found that several factors had a significant effect on crop identification performance: (a) crop maturity and site characteristics, (b) which of several different single-date automatic data processing procedures was used for local recognition, (c) nonlocal recognition, both with and without preprocessing for the extension of recognition signatures, and (d) use of multidate (multitemporal) data. It also was found that classification accuracy for field center pixels was not a reliable indicator of proportion estimation performance for whole areas, that bias was present in proportion estimates, and that training data and procedures strongly influenced crop identification performance.

012376 A Forestry Application Simulation of Man-Machine Techniques for Analyzing Remotely-Sensed Data by J. Berkebile, J. Russell and B. Lube.

The manual is a detailed step-by-step description of an actual analysis of remotely-sensed data performed by a forester for a portion of the Hoosier National Forest. The decsions made during the analysis and their rationale are described. The importance of the man-machine interactions are emphasized. The steps are documented with illustrations and examples. The reader is asked to display a mastery of each step through self-check items.