General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

LARS Publication 083077

7.8-10172
NASA CR-157289
C

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

ECHO User's Guide

by J. L. Kast
P. H. Swain
B. J. Davis
P. W. Spencer

(E78-10172) ECHO USER'S GUIDE (Purdue Univ.) 79 p HC A05/MF A01 CSCL 05B
N78-28577
Unclas
G3/43 00172

AUG 1978
RECEIVED
NASA STI FACILITY

Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
1977

LARS Publication 083077

ECHO USER'S GUIDE

BY:

J. L. KAST

P. H. SWAIN

B. J. DAVIS

P. W. SPENCER

TABLE OF CONTENTS

Section	Page
List of Table	ii
Introduction	1
Supervised ECHO Classifier (SECHO)	5
Nonsupervised ECHO: Field Extraction Phase (NS1ECHO)	29
Nonsupervised ECHO: Classification Phase (NS2ECHO)	39
A Discussion of the ECHO Algorithms	47
ECHO Parameter Selection Guidelines	57
References	72

LIST OF FIGURES

		Page
Figure 1	Flow Diagram for the ECHO Classifier	3
Figure 2	Example Control Cards for Executing the SECHO Cell Processing Phase	10
Figure 3	Example Control Cards for Joint Execution of Both Phases of SECHO	11
Figure 4	Example Control Cards for Execution of the Annexation Phase of the SECHO Processor	13
Figure 5	SECHO Summary Information	20
Figure 6	Example SECHO Classes and Channels Table and Processing Parameter List	22
Figure 7	Pictorial Classification Map	25
Figure 8	Example SECHO Singular Cell Map	27
Figure 9	Example Control Cards for NS1ECHO Processor	31
Figure 10	Standard Printer Output for NS1ECHO	36
Figure 11	Nonsupervised ECHO Field Map	38
Figure 12	Example Control Cards for the Nonsupervised ECHO Classification Phase	41
Figure 13	Standard Printer Output for the Classification Phase of Nonsupervised ECHO (NS2ECHO)	46
Figure 14	Supervised ECHO Landsat Parameter Settings to Optimize Six Variables	63
Figure 15	Supervised ECHO Simulated Thematic Mapper Parameter Settings to Optimize Six Variables	66
Figure 16	Nonsupervised ECHO Landsat Parameter Settings to Optimize Six Variables	68
Figure 17	Nonsupervised ECHO Simulated Thematic Mapper Parameter Settings to Optimize Six Variables	71

ECHO USER'S GUIDE

INTRODUCTION

The ECHO* classification functions are designed to identify objects in multispectral data, gather the statistics of the identified objects, and where possible, to classify the data on an object-by-object basis.

The motivation for this approach to classification is to include spatial as well as spectral information in the classfication algorithm and thereby increase the classification accuracy. One by-product of one ECHO implementation is that ECHO classifications require less CPU time than the standard point-by-point classifier.

Point-by-point classifiers, such as the LARSYS CLASSIFY-POINTS function, compare spectral measurements from each feature of each point to class statistics, computing a likelihood or discriminant function associated with each class, and categorizing the point as belonging to the class with the largest discriminant function value. Each point is classified individually, on the basis of spectral measurement alone. One premise of this technique is that the objects of interest are large in comparison to the size of the point. If this were not so, a large portion of points would be composites of several classes, making statistical pattern classification unreliable since pre-specified catagories would be inadequate to describe

---

ECHO stands for Extraction and Classification of Homogeneous Objects

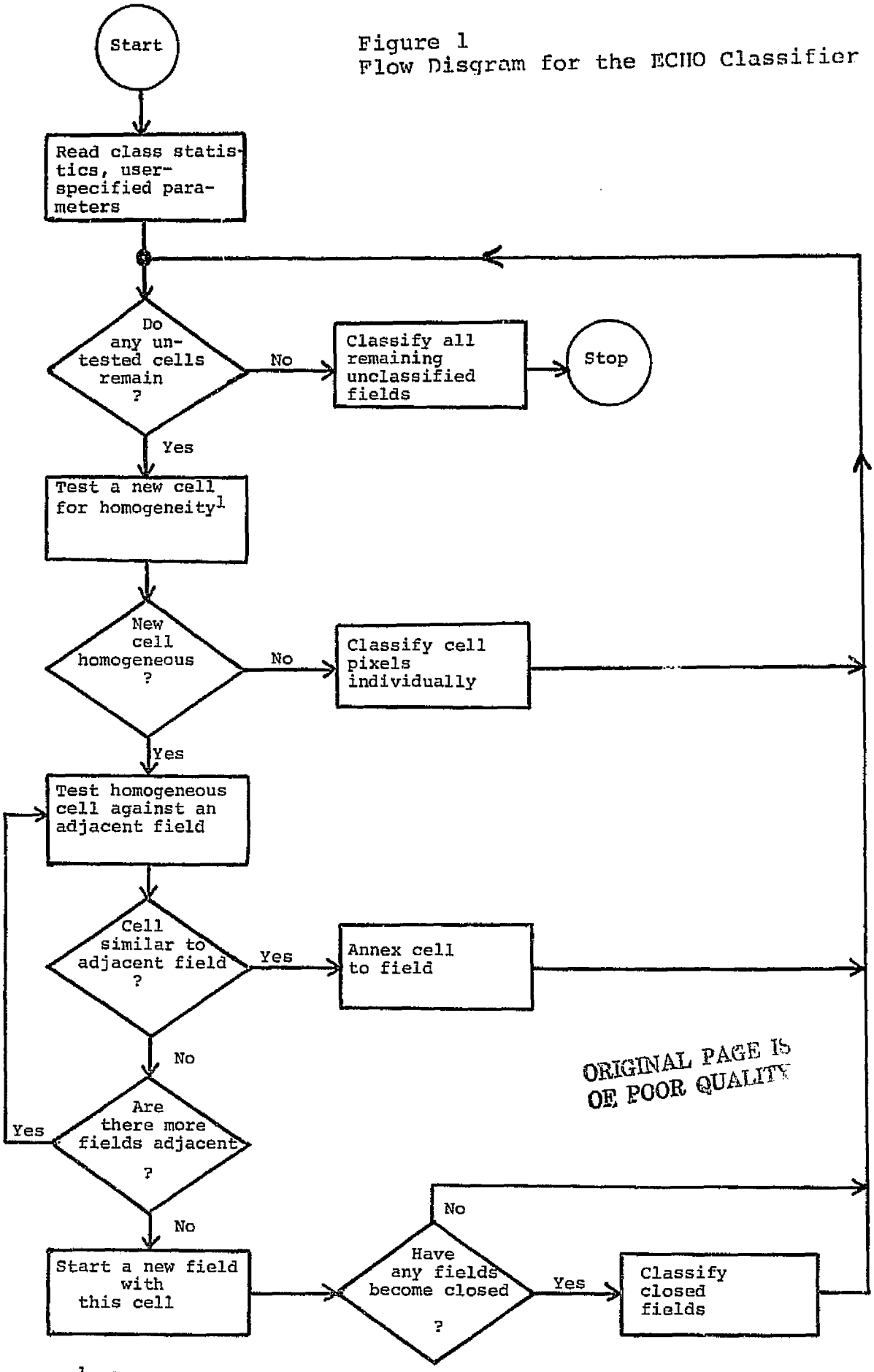
actual states of nature. From this premise it follows that objects are represented by arrays of point, and that a statistical dependence exists between consecutive points. Point-by-point classifiers fail to exploit the statistical dependence between adjacent points when assigning classes.

The ECHO processors benefit from spatial information by aggregating into groups points whose spectral responses are not significantly different in a statistical sense, and then applying a maximum likelihood classification rule to these homogeneous groups. Homogeneous objects are identified in a three step process. First, cells are formed by systematically partitioning the data into N by N sized blocks of pixels. The statistics of each cell are then compared to a homogeneity threshold. Points which do not comprise homogeneous cells (that is, constituent points of cells not meeting the homogeneity criterion) are classified on a point-by-point basis, just as contemporary classifiers catagorize all points. Statistics of adjoining homogeneous cells are then compared. Adjoining cells which appear to belong to the same statistical population on the basis of user-supplied annexation thresholds are combined into a single object and sample classified. To perform both the sample and the point-by-point classifications, Gaussian (or multivariate normal) class distributions (class mean and covariance matrices) are required. A flow diagram of this process is presented in Figure 1.

Two separate ECHO approaches have been developed. The first, supervised ECHO, makes use of pre-specified multivariate normal class distributions to identify homogeneous objects.

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 1
Flow Disgram for the ECHO Classifier



ORIGINAL PAGE IS
OF POOR QUALITY

[1]Cells are tested sequentially left-to-right and top-to-bottom in the image data.

The other, nonsupervised ECHO, identifies objects without the use of class statistics. Both processors require pre-specified class statistics (class mean and covariance matrices) to classify those objects identified. Objects identified by the nonsupervised field extraction algorithm (without the benefit of class statistics) may be used as an aid in obtaining the class statistics needed for the classification phase of the algorithm.

The two succeeding sections of this User's Guide deal with the supervised and nonsupervised algorithms, respectively. The supervised processor tends to be somewhat more accurate than the nonsupervised processor due to the use of the class statistics in the identification of homogeneous objects. On the other hand, since the nonsupervised processor does not require class statistics for object identification, the object map which it produces can be used to aid in developing the class training statistics.

Additional background information on ECHO may be found in the LARS Final Report to JSC in May 1975[1], R. L. Kettig's doctoral thesis[2], a LARS Information Note[3], symposium proceedings[4], the LARS Final Report to JSC in May 1977[5], and the LARS Final Report to JSC in November 1977[6].

SUPERVISED ECHO CLASSIFIER (SECHO)

Input to the function is:

- Data from a Multispectral Image Storage Tape,
- Control cards to select processing and output options,
- A statistics file containing the statistical description of the training classes,
- A data deck containing Field Description Cards to identify the area or areas to be classified.

The user has a wide range of control over the actual parameters used when processing data. He may elect to produce in either a one or two phase approach a Classification Results File, which may be placed either on tape or on disk. When the two phase approach is selected, the data is partitioned into N by N cells of user specified size, statistics are gathered for the cells, and those cells whose statistics do not pass the user-specified homogeneity criterion are identified. This cell processing information is then written to an Intermediate Results Tape. The second phase of the two phase approach utilizes the Intermediate Results Tape and the user-specified annexation criteria to produce the Classification Results File. The advantage of the two phase approach is that it allows the user to produce results utilizing different cell-to-cell annexation parameters without needing to repeat the expensive process of gathering cell statistics each time. When running the supervised ECHO processor in a single phase approach, all processing listed above is accomplished without the need of an Intermediate Results Tape.

ORIGINAL PAGE IS
OF POOR QUALITY

Although the Intermediate Results File has the same basic format as the Classification Results File, it is used only for storing information produced from the cell processing phase (where cell refers to a N by N sized block of data points). This file is used as input to the cell annexation phase which joins cells with similar characteristics and produces classification results.

Note: The Intermediate Results File produced by supervised ECHO processor is not compatible with the Intermediate Results File produced by the nonsupervised ECHO processor. The Intermediate Results Files generated by the two ECHO implementations should not be stored on the same tape.

The Classification Results File is normally used as input to the PRINTRESULTS function to produce a variety of printed output for the evaluation of the classification. It is also the primary input to the COPYRESULTS, LISTRESULTS, and PUNCHSTATISTICS functions. The file must be stored on tape for use by the latter two LARSYS functions.

SECHO produces four standard and three optional printer output products. Standard printer outputs include a control card listing, a list of the channels considered, a list of classes to be used, and an identification header listing characteristics of the run. The optional printer outputs are statistical summaries for the classes considered, a singular cell map, and a classification summary map. Only one of the latter two map outputs may be requested for a single execution of the processor. More detailed descriptions of these outputs appear later.

Inputs

The supervised ECHO classifier, as mentioned above, consists of two main parts: (1) the cell processing phase, carried out first, in which cell statistics are gathered and the screening of nonhomogeneous (singular) cells is performed, and (2) the cell annexation phase, where the cell information is used to join or annex neighboring cells with sufficiently similar spectral characteristics into fields (or groups of cells) and classify each entire field. These processing steps can be conducted either sequentially in a single execution of the processor or independently in two separate SECHO executions. Consequently, the input data required for each step of processing will be discussed separately.

Cell Processing Phase

The initial cell processing phase requires input of control cards, Field Description Cards for the areas to be classified, a Statistics Deck for training the classifier and for object identification, and the Multispectral Image Storage Tape. The supervised ECHO processor uses the identification information on the LARSYS Field Description Cards, along with the System or User Runtable File to identify and request the appropriate Multispectral Image Storage Tape. The format of the Multispectral Image Storage Data File and the LARSYS Runtable File can be found in the LARSYS System Manual [7].

Input statistics must be placed in the Statistics File before being used by the supervised ECHO classifier. A Statistics File is made available to the ECHO classifier either by executing one of the LARSYS functions that uses the statistics information or by including the statistics information in the

control card file. Any of LARSYS functions CLASSIFYPOINTS, STATISTICS, SEPARABILITY, CLUSTER, or SAMPLECLASSIFY may be used to transfer the statistics into the Statistics File.

The 'STATDECK USE' command may also be issued to transfer to the supervised ECHO processor a previously saved Statistics File.

If the user chooses to include the statistics in his supervised ECHO input deck, he must also include a 'CARDS READSTATS' control card in the deck. The statistics card deck is inserted into the input deck as the first group of data cards, preceding the Field Description Cards which describe the areas to be classified. Otherwise, the Statistics File is assumed to reside on the user's Temporary Disk.

Several control card parameters are required by the cell processing phase. The channel numbers of the data to be processed are required; the cell width (number of data points on each side of a square cell) must be declared; the cell homogeneity threshold (for differentiating homogeneous cells from singular cells) must be specified; optional selection of a subset of the training classes represented in the Statistics File may be specified; and declaration of the areas to be classified must be made.

Another required input is the destination of the results. As has been pointed out, the cell processing phase and the cell annexation phase may be carried out either jointly, in a single execution of SECHO, or independently, in two separate executions of SECHO. When the two phases are to be executed independently, an Intermediate Results File must be specified as the destination

ORIGINAL PAGE IS
OF POOR QUALITY

of the cell processing output. When the cell processing phase and the cell annexation phase are to be run jointly in a single execution of the processor, a destination for the final results must be included. The Classification Results File may either be placed on disk or on a Results Tape.

An example control card deck for executing the cell processing phase (phase 1) of the supervised ECHO processor is presented in Figure 2.

Cell Processing and Annexation

When all processing is to be accomplished in one step, (both phases run in a single execution) only the annexation threshold and final results location need to be added to the information required by the cell processing phase. When the 'INTERMEDIATE TAPE' control card in Figure 2 is replaced by a 'RESULTS' control card and an 'ANNEXATION' control card is added, cell processing and annexation occur in one step and a Classification Results File is produced. Figure 3 is an example of the control cards necessary for the execution of both the cell processing (phase 1) and the cell annexation (phase 2) algorithms in a single step. Note: No 'INTERMEDIATE' control card may be used when single step processing is desired.

Cell Annexation Phase

When independent execution of the cell annexation phase (phase 2) is desired, the 'INTERMEDIATE' control card is required to specify input from the Intermediate Tape File, produced by the previously executed cell processing phase (phase 1). An 'OPTIONS INTERMEDIATE' control card must appear in the card deck to indicate that only the cell annexation algorithm is desired. In addition, a Classification Results File destination

Figure 2

Example Control Cards for Executing the SECHO Cell Processing Phase

```
END
        (LARSYS FIELD DESCRIPTION CARD(S) )
/DATA
        (A LARSYS STATISTICS FILE GOES HERE)
  DATA
  CHANNELS 5,6,7,8
  CARDS READSTATS
  INTERMEDIATE TAPE(990), FILE(3)
  CELL SIZE(2), HOMOGENEITY(273)
*SECHO
```

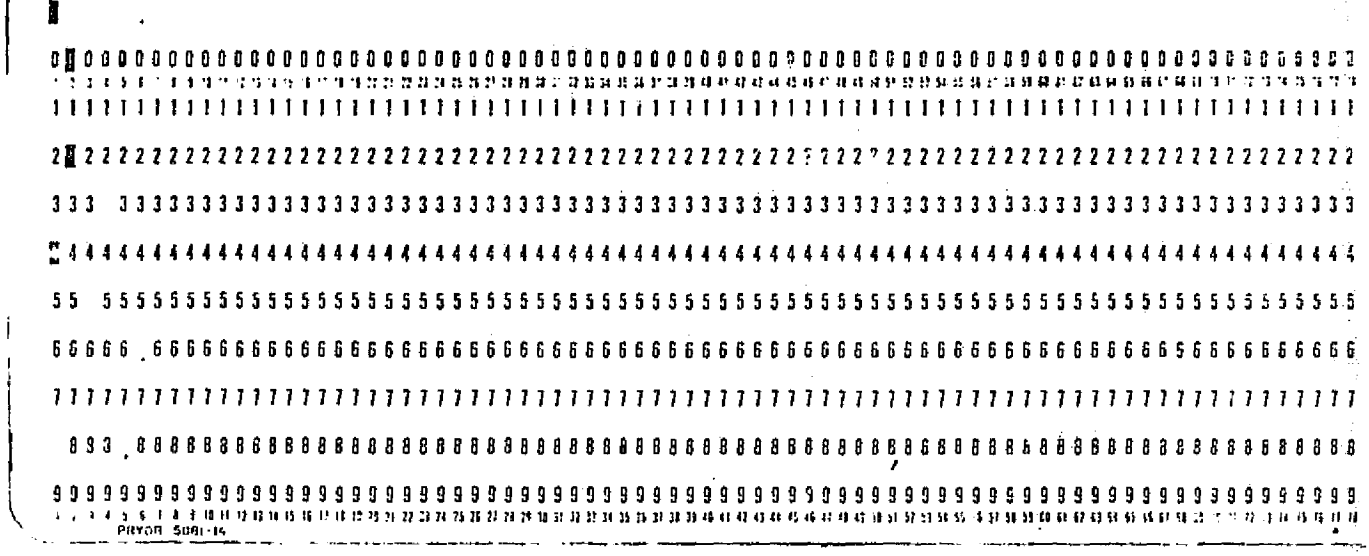


Figure 3

Example Control Cards for Joint Execution of Both Phases of SECHO

```
END
        (LARSYS FIELD DESCRIPTION CARD(S) )
/DATA
 II
        (A LARSYS STATISTICS FILE GOES HERE)
/DATA
/CHANNELS 5,6,7,8
/CARDS READSTATS
/RESULTS TAPE(991), INITIALIZE
/CELL SIZE(2), HOMOGENEITY(273)
/ANNEXATION THRESHOLD(1.0)
*SECHO
```

PAYOR 5081-14

ORIGINAL PAGE IS
OF POOR QUALITY

must be specified. All cell width, channel calibration, and optional selection of training classes information is extracted from the Intermediate Tape and need not be respecified. Figure 4 is an example of the control cards necessary to complete an ECHO classification. Execution of the control cards in Figure 2 would have supplied the Intermediate Results Tape which contains the cell processing input for the annexation phase.

Specification of Channels: The multispectral data channels to be used by the supervised ECHO classifier must be specified by including the CHANNELS control card. This control card must appear whenever the cell processing is to be performed (either for execution of the cell processing phase or for joint execution of both ECHO phases). The user specifies channels in this manner:

CHANNELS I, J. . .

where I, J, . . . are the channel numbers to be used. Appendix IV of the LARSYS User's Manual[8] contains information on how this card may also be used to calibrate data from the Multispectral Image Storage Tape.

Optional Selection of Training Classes: The user may select the training classes from the Statistics File that are are to be used by supervised ECHO's cell processing phase (phase 1), and he may combine training classes into pools. These options are exercised by using the 'CLASSES' control card. For example, if the user wishes to use only classes 1, 3, and 5 of seven training classes previously defined by the Statistics function the control card entry would be:

CLASSES 1, 3, 5

Figure 4

Example Control Cards for Execution of the Annexation Phase of the SECHO Processor

```
END
ANNEXATION THRESHOLD(1.0)
RESULTS TAPE(991), FILE(2)
INTERMEDIATE TAPE(990), FILE(3)
OPTIONS INTERMEDIATE
*SECHO
```

PRYOR 5081-14

In this case, the class name assigned by the statistics function at classes 1, 3, and 5 will be retained by SECHO and the other classes will be totally ignored.

To combine two or more classes into one class, the user assigns a name (up to eight characters) to the pooled class to be created and specifies the classes to be included in the pooled class. For example, assume there are eight classes available in the training statistics, and the user wishes to process the following combinations:

- POOLA (Pool A) will be classes 1 and 2.
- POOLB will be classes 4, 6, and 7.
- POOLC will be class 5 only.
- Classes 3 and 8 will be ignored.

The control card format to specify this option will be:

CLASSES POOLA(1/1,2/), POOLB(2/4,6,7/), POOLC(3/5/)

Note that the number immediately following a left parenthesis specifies the pool sequence. Pool sequence numbers must be in ascending order. Note also that the classes to be pooled (and named) are enclosed by slashes (/).

When no 'CLASSES' card is specified, all the classes in the statistics deck will be considered by the supervised ECHO processor both object identification and for classification.

<u>Specification of Annexation Parameter</u>: The annexation parameter is required for execution of the SECHO processor when the two ECHO phases are to be run jointly or when the cell annexation phase is to be run. The form of this card is:

ANNEXATION THRESHOLD (X.X)

ORIGINAL PAGE IS
OF POOR QUALITY

where X.X is a floating point threshold for the generalized likelihood ratio criterion for annexing to fields adjoining homogeneous cells. The higher the annexation threshold, the more likely it is that annexation will occur.

Specification of Cell Parameters: The cell width and homogeneity parameters are required by the SECHO processor for execution of the cell processing phase or joint execution of both SECHO phases. These parameters are specified with a control card of the form:

CELL WIDTH(N), HOMOGENEITY(XX.X)

The width parameter represents the "width" of the cell in pixels. Each cell is made up on N^2 pixels of N columns and N lines. The homogeneity parameter is used as a threshold for differentiating homogeneous cells from singular (non-homogeneous) cells. As the homogeneity parameter increases, the likelihood that a cell will be identified as homogeneous increases.

Specification of Areas to be Classified: The user must provide the cell processing phase (phase 1) of the supervised ECHO processor with Field Description Cards to define the area or areas to be classified. These are included in the input deck following a DATA CARD. Either of two forms of this card may be used. The formats are described in the Control Card Dictionary for CLASSIFYPOINTS in appendix I of LARSYS User's Manual [8]. These Field Description Cards identify the specific portion of data from the Multispectral Image Storage Tape that is to be classified. The information is used by the processor to request the appropriate tapes and access the desired segment(s) of the specified data run.

Outputs

Classification Results File: The principle output of the Supervised ECHO Classifier is the Classification Results File, which, in turn, is the primary input to other LARSYS functions: PRINTRESULTS, COPYRESULTS, LISTRESULTS, and PUNCHSTATISTICS. The location of this file must be specified when either the single-step (phase 1 and 2 executed jointly) or the cell-to-cell annexation phase are to be executed. The location of this file is not specified when only the cell processing phase is to be executed. The file may reside on either tape or disk, and the user must specify one or the other on a RESULTS control card. However, if the user wishes to save the results file, or if he wishes to use it as input to the LARSYS LIST-RESULTS or PUNCHSTATISTICS functions, he must place it on tape or have it copied to tape by the COPYRESULTS function.

The user specifies where the Results File will reside by using a RESULTS control card in one of three forms:

RESULTS TAPE (xxx), FILE(nn)

RESULTS INITIALIZE, TAPE(xxx)

RESULTS DISK

ORIGINAL PAGE IS OF POOR QUALITY

The first control card is used to add the file to a tape already containing classification Results Files. If a file of the specified number already exists on the tape, the user will be notified by a message. He then has the option of writing over the old file, specifying a new tape and file number, or stopping execution.

The second 'RESULTS' control card example specifies that a new results tape is to be used, and the 'INITIALIZE' parameter

requests that the proper header information be placed at the beginning of the new tape. A tape must always be initialized before it can be used to store classification results.

Execution of the third 'RESULTS' control card would cause the Classification Results File to be written on the disk. When the Classification Results File is placed on disk, it is only stored there temporarily. If the user wishes to save the file, he must copy it to tape with the COPYRESULTS function. Any of the actions listed below will cause the Classification Results File to be erased from the disk by the system:

- Another execution of a classification function.
- Re-initiation LARSYS, i.e., issuing the 'I LARSYS' control command.
- Logging off the system, i.e., issuing the 'QUIT' control command.

A unique "classification study number", based on the date and time of the run, is part of each Results File. The number, identified as "classification study", is included on any outputs that are subsequently derived from the results file. The form of the identification number is "ydddsssss"; where y is the last digit of year, ddd is the Julian date (day of the year 001-365), and sssss is the total number of seconds since the previous midnight.

The principal data on the Classification File is the categorization of each input point made during the classification run. A separate record is written for each line of the

classification. This record contains, for each point in the line, the class number associated with the class to which the point was assigned. The likelihood code, which is set by the LARSYS per point classifier, is not assigned a value by the SECHO processor. The classification results are used by the PRINTRESULTS function to produce detailed maps of the classified area as well as tables of the test fields, training fields, and class performance. For more information on these products, refer to the description of the PRINTRESULTS function in the LARSYS User's Manual[8].

In addition to the classification results, the file contains other data related to the Classification run:

- A complete copy of the Statistics File that was used as input to the run. This file may be punched on cards by using the PUNCHSTATISTICS function.
- Summary information about the classification and the channels and classes that were used. A formatted listing of this information may be produced by using the LISTRESULTS function. This listing is also a secondary product of both the PUNCHSTATISTICS and the COPYRESULTS function.
- Reduced satellite (mean vectors and covariance matrices) for the classes and channels used in the classification.

Intermediate Results File: A secondary output is the Intermediate Results File, used only when cell processing and cell annexation are to be performed independently by two separate executions of the SECHO processor. The same control cards

ORIGINAL PAGE IS
OF POOR QUALITY

are used for specifying the Intermediate Results location as for specifying the classification results location except the card is labeled 'INTERMEDIATE' rather than 'RESULTS' and the 'DISK' is not a valid location. A tape file must be used for Intermediate Results storage. The format of the Intermediate Results File is similar to that of the Classification Results File. The class catagorizations and associated probabilities which appear for each line of input data in the Classification Results File (see LARSYS SYSTEMS MANUAL [7]) are replaced, in the supervised ECHO Intermediate Results File, by the class numbers and cell likelihood values for each row of N by N point cells. When processing is to be carried out in two independent phases, the 'INTERMEDIATE' card must appear in the control card decks of both the cell processing and the cell annexation phase. The 'INTERMEDIATE' card identifies the destination of the principal results of the cell processing phase when that phase is executed independently. It identifies the location of the principal input when the cell annexation phase is executed.

Standard Printer Output: The supervised ECHO classifier always prints a summary of the user's input deck. The summary includes a reproduction of the input deck control cards, a list of options the user has selected, and particular characteristics about the run, such as the number of class and channels used, the channel numbers, etc. An example of this output is shown in Figure 5.

In this case, the 'CARDS READSTATS' option indicates that the Statistics Deck specifying the mean and covariance matrices

Figure 5. SECHO Summary Information

```
BATMED                          LABORATORY FOR APPLICATIONS OF REMOTE SENSING          AUG  24,1977
PAUL SPENCER                                  PURDUE UNIVERSITY                        10 51 21 AM
                                                                                       LARSYS VERSION 3

*SECHO
ANNEXATION THRESHOLD(1.0)
SYMBOLS ,/,I,I,-,h,-,G,-,G,=
CELL SIZE(2),HOMOGENEITY(27.3)
PRINT SINGULAR
RESULTS TAPE(990),FILE(4)
CARDS READSTATS
CHANNELS 5,6,7,8
DATA

YOU HAVE SELECTED THE FOLLOWING SUPERVISED ECHO OPTIONS

  USE STATISTICS FROM CONTROL CARD DECK
  PRINT SINGULAR CELL MAP
  PERFORM INITIAL CELL PROCESSING OF AREA
  PERFORM CELL ANNEXATION AND PRODUCE CLASSIFICATION RESULTS

SUPERVISED ECHO INFORMATION....

  CLASSIFICATION STUDY...........723639386
  NO.OF POOLS.....................  11
  NO.OF CHANNELS..................   4
  NO. OF TRAINING FIELDS........  30
```

ORIGINAL PAGE IS
OF POOR QUALITY

of the training classes appears as part of the control card deck. The 'PRINT SINGULAR' causes a Singular Cell Map to be generated. The absence of an 'INTERMEDIATE' card indicates that both the cell processing and the cell-to-cell annexation phases are to be executed.

Several items listed under "SUPERVISED ECHO CLASSIFIER INFORMATION" in Figure 5 are of particular interest. The list is always headed by the Classification Study Number (the unique identification number for the particular classification). The number of fields used to generate the statistics for the classifier are given next. Note that in this case 30 fields were used to generate the input Statistics File.

The last item in the list, ("CHANNELS, SELECTED ARE...") identifies the channels that will be used in the classification. If the user had included a CHANNELS control card in his input deck, the channels that were specified there would be listed.

There are three other standard printer outputs. They are:

1. A Classes and Channels Table. This shows the class name for each of the training classes (as defined in the Supervised ECHO Classifier input deck) and the channel number, spectral band, and calibration code for each channel (taken from the Statistics File). A sample is shown in the attached Figure 6.
2. A Processing Parameters List. Figure 6 also contains a list of the processing parameters. The cell width, the annexation threshold, and the cell homogeneity threshold are a recapitulation of control card inputs, the number of channels and the number of pools results from the information contained in the Statistics Deck.

Figure 6. Example SECHO Classes and Channels Table and Processing Parameter List.

BATHED
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 24,1977
10 56 29 AM
LARSYS VERSION 3

TRAINING CLASSES

CLASS
HWYPKG
COMI
IND1
IND2
RES1
RIVER
RES2
GRAS1
RES3
GRAS2
ARPT

CHANNELS FROM STATISTICS

CHANNEL NO.	SPECTRAL BAND		CAL. CODE
5	0.50	0.60	1
6	0.60	0.70	1
7	0.70	0.80	1
8	0.80	1.10	1

PROCESSING PARAMETERS

CELL WIDTH = 2

NUMBER OF CHANNELS = 4

NUMBER OF POOLED CLASSES = 11

ANNEXATION THRESHOLD = 1.0000E 00

CELL HOMOGENEITY THRESHOLD = 2.7300E 01

NUMBER OF CELL LINES IN BUFFER = 113

These two parameters may be modified by the 'CLASSES' and 'CHANNELS' control cards. The number of cell lines which the program will hold based on the other input requirements (classes, length of line, size of cell) is specified. This value must be at least 2.

3. A Classification Run Identification Table. This table shows the run information obtained from the input tape ID record, the spectral band and calibration code for each channel, and the coordinates for the area to be classified. If a map is requested, this table will be printed as a header for the map. An example of this table is above the example printer map (Figure 7) which appears in the description of optional printer output.

Optional Printer Output: Three optional printer outputs may be selected with the PRINT control card:

1. Statistics Summary. This output is produced for each of the classes (or pooled classes) used in the classification. Its form and content is the same as that produced in the LARSYS STATISTICS function, except that it covers only the actual channels that are to be used in the classification. It shows, for each of the classes, the mean and the standard deviation of the response for each channel of data, and a correlation matrix of channels.

2. A Pictorial Classification Map. This map, generated during the cell annexation phase of SECHO, is an image of the entire classified area, with each point

ORIGINAL PAGE IS
OF POOR QUALITY

represented by an alphanumeric symbol (a number, character or special symbol). Figure 7 presents the classification map which results from the control card input presented in Figure 5. Note the standard Run Identification Output appears as a header to the Classification Map. The symbol that is used to represent each class on the map is recorded on the Classes and Channels Listing. These symbols default to assignment to each class (or pooled class) based solely on the class number. Default assignments are as follows:

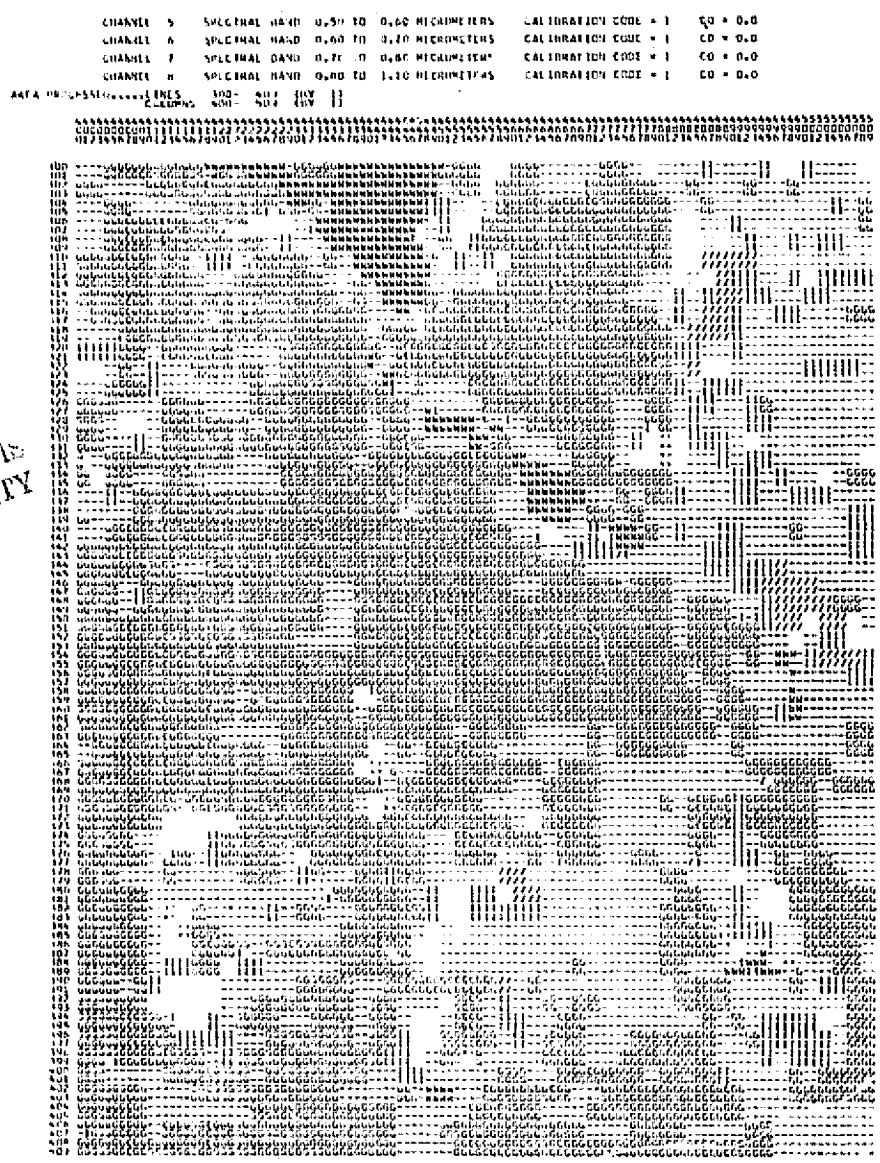
Class Number	Symbol
1 through 9	numbers 1 through 9
10 through 35	characters A through Z
36	number 0
37 through 44	symbols +,=,*,$,/,&,(,and)
45 through 53	numbers 1 through 9
54 through 60	characters A through G

Alternatively, the user may specify symbols assignments by use of a 'SYMBOLS' control card. For example:

SYMBOLS A,A,A,B,W,A,

would cause the first, second, third, and sixth classes to be represented by an A on the classification map, the fourth class by a B and the fifth class by a W. More comprehensive and flexible mapping capabilities are available through the LARSYS PRINTRESULTS function. The reader should refer to the description of that function in the LARSYS User's Manual [8] for an example PRINTRESULTS output.

Figure 7. Pictorial Classification Map.



ORIGINAL PAGE IS OF POOR QUALITY

The user may use the PRINT control card to request either or both outputs discussed to this point. A 'PRINT STATS' card will print only the statistics summary, a 'PRINT CLASSIFICATION' card will print only a map, and a 'PRINT STATS, CLASSIFICATION' card will print both of them.

3. Singular Cell Map. This map is obtained from the cell processing phase of the Supervised ECHO Classifier function. Figure 8 is a Singular Cell Map of the same area as that represented on the Pictorial Classification Map in Figure 7. By applying the cell selection threshold supplied in the input control cards, non-homogeneous cells are detected and screened out. The singular cell map places a symbol ('0') at the coordinator of each singular cell. Note that a character on this map represents a cell of data, not a single point. Hence, in Figure 8, since the cells are two by two sized blocks of pixels, line and column headers are incremented by two. This map is useful in detecting a very non-homogeneous area, too high a value for the cell selection parameter, or classes missing in the statistics information.

   Large groups of contiguous singular cells will occur when one or more spectral classes have been omitted. For example, in Figure 8 there is a large group of singular cells between lines 300 to 322 and columns 424-448. Part of a reservoir is contained in this area. Though water is a class contained in

Figure 8. Example SECHO Singular Cell Map.

```
AREA PROCESSED.....LINES     300-  409  (BY  1)
                   COLUMNS   400-  509  (BY  1)

        444444444444444444444444444444444444444444444444444455555
        000011111222223333344444555556666677777888885999900000
        024680246802468024680246802468024680246802468024680246802468

    300 000 0     00000000000000000 00000000000000000 0 0         0
    302         000 00000000000000 0 00000 0 0 00 000 0 0
    304       0 0000 00000000000000   00000     0      0     0  0
    306             00     00000000000  00000 00       000          0
    308          0     000000000000    0      0     00 00 0         0
    310 0  0  00000 000 00000000000000 00000000   000   0 0
    312   0000 0 0    0 000 000000   00  00000    0  0              0
    314   000 000      00000  000      0 0000    0   00  0        00
    316 0                 00 0000      00  0 00        0          0
    318 000 0 000        00000000        0  0000        0          00
    320         0 0        00000000 0      00     00  0
    322   0 0   00000       00000000  000 0 0 0       0  0
    324 0          000  0  00000    00      0  000    0 0     0
    326     0  000      00 000     0        0      0   0       0
    328   0 00 00           00     000000000 000     0                0
    330   000                    00 000 00 0  0
    332 0                  0          00 00000  0000      0   00  0
    334   0 0            00                 000000000     0      0  0
    336 0 0        0  0000     00       0  000000 00    0 0       0
    338 000   00     000           00    000 0000 0            0
    340  0     0  00000 0      00      00000 0 00000    0         0
    342 0 0                0 00 0000    0000     000 0 00     0000
    344     00        00        0 0  00      0  00000
    346     0         0       00 0            000  00000
    348 0  00                          000     00000     00      0
    350          0       00  0 0                0 000 00000000       00
    352            0    00 0      0  000  0     00000 0 00  0 00 0
    354                          00  0000       0000   0  00000 0  0
    356   0     0      0       0      00000000 0  0  00     00  0  0 0
    358                0       0      0       0     0 000  0     0
    360   0        00      0 0    00       0        00     00  00
    362  00000 0     0    00            0  0        0   0  0  0      0
    364 0 00  0    000        00         0           0000          0
    366   00     0 00000000    00  0        0                      0
    368 000      00000  00000  00   0     0   0                0
    370 000    000000 0  00000 0      0 00                        00
    372 0          0  00000  000 00     0 0              00      0
    374  00 0      0     000    0 0       0  0           0      0
    376      0000  0    000 0          0 0          00    0 0 00
    378      0 0000  00 00 0         00000   0        0      00 00 00
    380 00    0 00        00  00 0  0    000        000  0 00
    382 00    0 0         000 00 0    0 0 0           0 00000    000
    384 00        0  00    0 000 00    0              00     0  0
    386 00     00    0       00   000  0      0         0 00 00
    388 00       00000 00        0          0    0 0  00000000
    390 0000     00000 0       00      00 0                00 0
    392 0000        0        00  00    0  0                  0  0
    394 00000000  0 0         0000 0  0    00     0    0        0 0
    396 00000  00  0     0  0  0      0  0       0  0       0  0
    398 00000   00          0  00    000000      0   0000 0     0
    400   0     0              0 0 0    0                0    000
    402  0                 00     00  0  0                     0
    404 00 00     0 0      0     0                          00000
    406 000 0 000000       0      00       0
    408  00000      00    000    0000       0000000    0

EOF RECORD WRITTEN.

I0103  CPU TIME USED WAS    79.892 SECONDS.      (LARSMN)
```

ORIGINAL PAGE IS
OF POOR QUALITY

the statistics deck for this run, the statistics for the water class were gathered over a river rather than the reservoir. Statistics of the water in the reservoir are different enough from the statistics of the water in the river for these cells to appear unrecognizable, and hence, to be identified as singular. When the cell homogeneity parameter is very high, no cells will be identified as singular. Unless the analyst desires all cells to be classified as small samples, a cell map with few symbols indicates the homogeneity parameter is too high.

Only one map can be produced by a single execution of supervised ECHO classifier. Either a Classification Map or the Singular Cell Map may optionally be produced, but not both.

NONSUPERVISED ECHO: FIELD EXTRACTION - PHASE 1 (NS1ECHO)

The NS1ECHO function is an implementation of the field extraction phase of the nonsupervised Extraction and Classification of Homogeneous Objects (ECHO) algorithm. It partitions the data into N by N sized cells of pixels, performs cell-to-cell annexation to form fields, computes statistics of these fields, and saves the results on an Intermediate Tape. In addition, this function creates an object map by replacing the data vectors of those pixels identified as falling within a field with a data vector of the channel-means of the field. The program flags those cells which it identifies as "singular" (containing pixels from more than one class). Information is stored on the Intermediate Results Tape to be later used as input to the nonsupervised ECHO Classification Phase (NS2ECHO function).

Input to the function is:

- Data from the Multispectral Image Storage Tape
- Control cards to select the processing and output options.
- A data deck containing a Field Description card to identify the area to be processed.

The user has a great deal of control over the data to be processed by means of the control cards. The results are placed on an Intermediate Results Tape for later processing by the

nonsupervised ECHO Classificatin Phase. Note: The format of nonsupervised Intermediate Results File is not compatible with the supervised ECHO Intermediate File format. Intermediate results generated by the nonsupervised ECHO processor should be kept on a separate tape from intermediate results produced by the supervised ECHO processor. NS1ECHO produces, besides general information about the ECHO run, an optional field map. A detailed description of how this map is requested appears below.

Inputs

The main input to the NS1ECHO function is the Multispectral Image Storage Tape. The function will use the identification information on the Field Description Cards, along with the system (or user) Runtable, to identify the appropriate input tape and have it mounted. The content and form of this primary LARSYS input file is described in Appendix IV of the LARSYS System Manual [7].

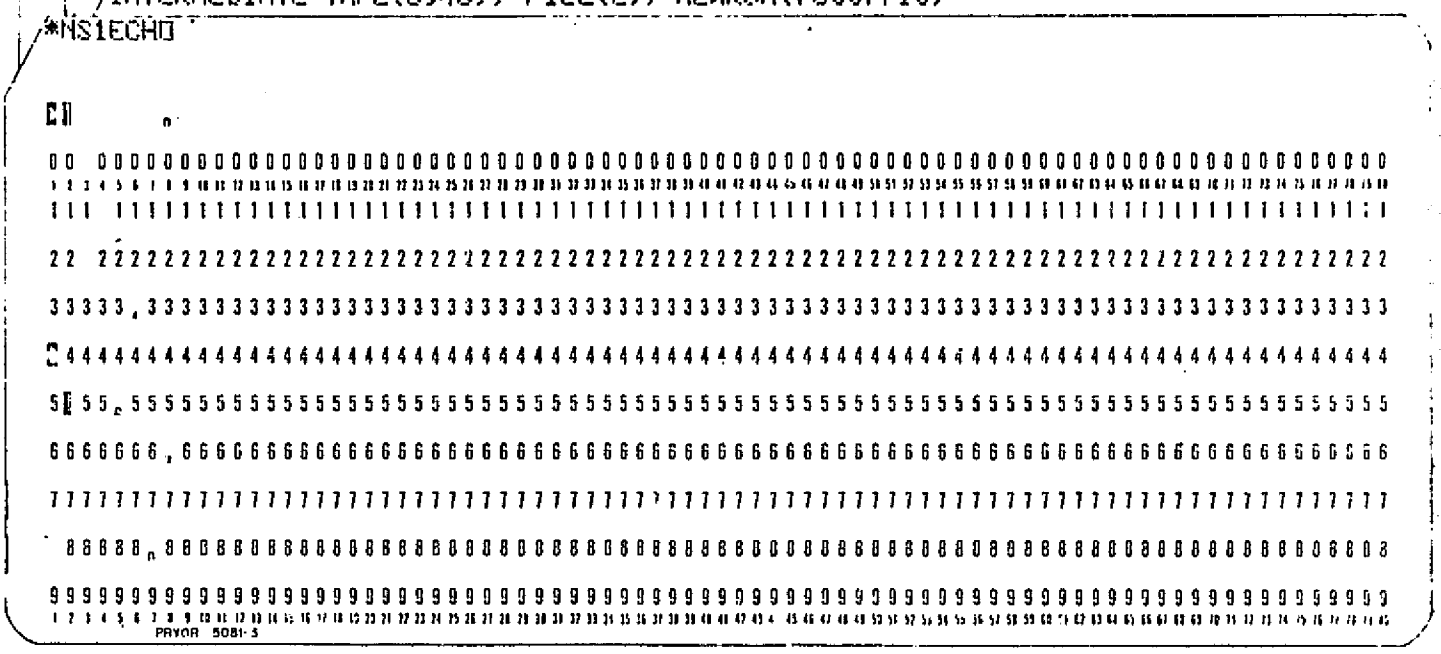
In addition to the principal input, the user is expected to provide an input deck which further defines the data to be used, the processing parameters, and the input/output options. More specifically, he employs control cards to designate the channels to be used, the annexation, cell selection and cell width parameters, and the intermediate tape, file, and run number. He also must provide a data card (a LARSYS Field Description Card) which specifies the area to be processed.

The sample input deck shown in Figure 9 illustrates the use of these inputs. The discussion that follows provides details about the specifications of these inputs.

Figure 9

Example Control Cards for NS1ECHO Processor

```
/END
//
//        (LARSYS FIELD DESCRIPTION CARD(S) )
//
/DATA
/CELL WIDTH(2), HOMOGENEITY(0.25)
/ANNEXATION MEAN(0.01), VARIANCE(0.01)
/INTERMEDIATE TAPE(3943), FILE(2), NEWRUN(75007710)
*NS1ECHO
```



ORIGINAL PAGE IS
OF POOR QUALITY

Specification of Channels: The channels to be used by the NS1ECHO function must be specified in a CHANNELS card. The form is:

CHANNELS I, J, . . .

where I, J, . . . are the channel numbers of the channels to be used. An example of the use of this card is shown in Section 3.1 of the LARSYS User's Manual[8]. Appendix IV of the User's Manual contains information on how this card may also be used to calibrate data from the Multispectral Image Storage Tape.

Specifications of Annexation Parameters: The annexation parameters are required and must be specified on an ANNEXATION card. The form of this card is:

ANNEXATION MEAN (X.XX), VARIANCE (Y.YY)

where X.XX and Y.YY are floating point numbers and represent annexation thresholds for the mean and for the covariance matrix respectively. They must be one of the following values: .1, .05, .025, .01, .005, .001. These parameters are used as thresholds in comparisons between adjacent homogeneous cells. A cell is annexed to a field if it pass both the mean threshold test and the covariance threshold test. As the annexation thresholds become smaller, the likelihood of annexation increases.

Specification of Cell Parameters: The cell width and homogeneity parameters are required and are supplied by means of a CELL card. The form of this card is:

CELL WIDTH(N), HOMOGENEITY(Y.YY, Z.ZZ, . . .).

The width parameter represents the "width" of a cell in pixels. Each cell is made up on N^2 pixels of N columns and N lines. If not specified, the cell width defaults to 2. The cell homogeneity parameter is a threshold for the cell mean and variance. If the variance divided by the mean of the cell is greater than the homogeneity threshold for any selected channel, the cell is split and each constituent pixel classified separately. The cell homogeneity threshold can be any value. As the homogeneity parameter increases, the likelihood that a cell will be identified a 'singular' and its pixels classified individually decreases. If only one homogeneity parameter is specified, it will be applied to the ratio of cell variance to cell mean for each requested channel. When two or more homogeneity parameters are specified, the first threshold will correspond to the first channel selected, the second threshold to the second selected channel, and so on. When more thresholds than channels are specified, the trailing thresholds are ignored; when more channels than thresholds are requested, the last specified threshold will be used for the trailing channels.

Specification of Areas to be Classified: The user must provide a Field Description Card to define the portion of the selected LARSYS run that the field extraction phase of nonsupervised ECHO is to process. This card follows the 'DATA' card in the input card deck. Either of two forms of this card may be used. The formats are described in the control card dictionary for CLASSIFYPOINTS in Appendix I of the LARSYS User's Manual[8].

The Field Description Card identifies the specific portion

of data from the Multispectral Image Storage Tape that is to be used. The information is used by the processor to request the appropriate tapes and position those tapes so as to access the requested lines and columns of the specified runs.

Optional Specification of Field Map: The user may request to have a map printed showing the annexation of cells into fields as well as singular cells. It is specified by the following card:

PRINT MAP

In addition, this option sets up the intermediate tape for conversion to a map tape output by replacing individual pixel value by the mean value of the field which the pixel is associated with. This option will cause computer time to increase, so it should be used only when an object (field) map is desired.

Outputs

Intermediate Results File: The principal output of the NS1ECHO function is the Intermediate Results File, which is, in turn the primary input to the NS2ECHO function. The file must reside on tape which is specified by the user on the INTERMEDIATE card. The nonsupervised Intermediate Results File is not compatible with and may not reside on a LARSYS Classification Results Tape.

The user must specify where the file is to be stored by using an 'INTERMEDIATE' control card in one of two forms:

INTERMEDIATE NEWRUN(XXXXXXXX), TAPE(YYYY), FILE(ZZ)

INTERMEDIATE NEWRUN(XXXXXXXX), TAPE(YYYY), INITIALIZE

The first control card is used to place the file on a tape already containing Intermediate Files. If a file of the

specified number already exists on the tape, the user will be notified by a message. He then has the option of writing over the old file, specifying a new tape and file, or stopping execution. The second control card specifies that a new tape be used, and the 'INITIALIZE' parameter requests that the proper header information be placed at the beginning of the new tape before a new file is written. A new tape must always be initialized before it can be used to store intermediate results. The NEWRUN parameter specifies a unique eight digit number to be placed in the run slot on the file ID record. In addition, point-by-point means of annexed fields (or original data values if the cell was singular) and an array which gives a field number for each cell are contained on the Intermediate Results File. The nonsupervised Intermediate Results File contains statistics for each of the homogeneous fields identified. These statistics are used in NS2ECHO to sample classify the fields.

Standard Printer Output: The NS1ECHO function always prints a summary of the user's input deck. The summary includes a reproduction of the input deck, and a set of parameters selected. This set of information includes the cell width, the number of channels, and the annexation and homogeneity parameters. Figure 10 shows an example of this output for the control cards appearing in Figure 9.

Optional Printer Output: The Field Map is an optional printer output which may be selected by means of the 'PRINT' card.

Figure 10

Standard Printer Output for NS1ECHO

```
ECHOP6                          LABORATORY FOR APPLICATIONS OF REMOTE SENSING          NOV  10,1977
CARLOS ARTURO                              PURDUE UNIVERSITY                           02 15 20 PM
                                                                                       LARSYS VERSION 3

  -COMM AIRCRAFT
  *NS1SUP
  CHANNELS 1,2,3,5,6,7
  INTERMEDIATE TAPE(3943),FILE(3),NEWRUN(75003710)
  ANNEXATION MEAN(.01),VARIANCE(.01)
  CELL WIDTH(2),HOMOGENEITY(.25)
  DATA

READER INFORMATION 0
     NUMBER OF CHANNELS =  6
     CELL WIDTH = 2
     NEW RUN NUMBER = 75003710  TAPE = 3943  FILE =  3
     MEAN AND VARIANCE THRESHOLDS ARE    0.01000  AND   0.01000
     CELL HOMOGENEITY THRESHOLDS ARE     0.25000   0.25000   0.25000   0.25000   0.25000
                                         0.25000

PROCESSING PARAMETERS
  NUMBER OF ROWS OF CELLS IN BUFFER =    194
  MAXIMUM FIELD SIZE (NO. OF POINTS) =   90000
```

ORIGINAL PAGE IS
OF POOR QUALITY

This output is a map showing how cells were annexed into individual fields. Each field which the nonsupervised ECHO processor identifies is arbitrarily assigned a symbol. Singular (nonhomogeneous) cells are assigned blanks. Figure 11 shows an example Field Map which was generated by adding a 'PRINT MAP' control card to the control card deck listed in Figure 9.

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 11

Nonsupervised ECHO Field Map

ECHOP6
CARLOS ARTURO
AIRCRAFT

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

NOV 10,1977
02 18 38 PM
LARSYS VERSION 3

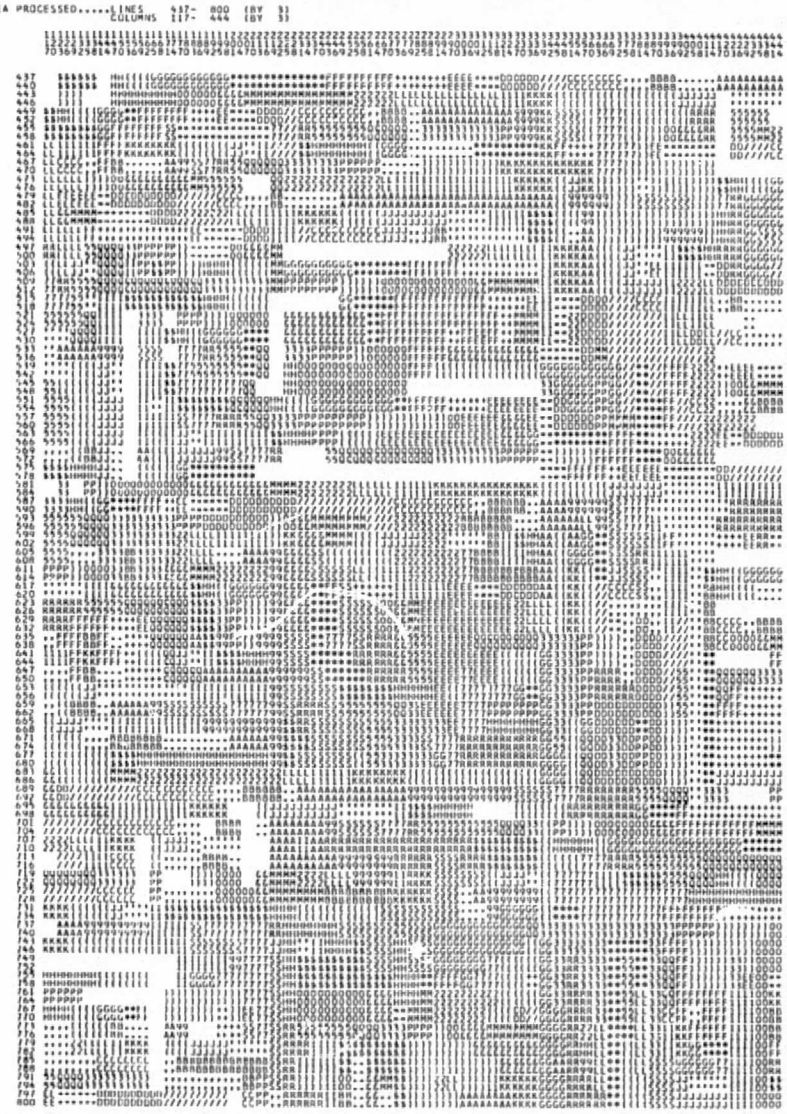
SAMPLE CLASSIFICATION STUDY ..

RUN NUMBER..... 75003710	DATE DATA TAKEN... JULY 6,1975
FLIGHT LINE... LINE 3 R1 KANSAS	TIME DATA TAKEN..... 1553 HOURS
DATA TAPE/FILE NUMBER.. 3230/ 1	PLATFORM ALTITUDE.. 10000 FEET
REFORMATTING DATE. APR 5,1976	GROUND HEADING..... 0 DEGREES

CHANNELS USED

CHANNEL 1	SPECTRAL BAND 0.59 TO 0.64 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 2	SPECTRAL BAND 0.65 TO 0.69 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 3	SPECTRAL BAND 0.72 TO 0.76 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 5	SPECTRAL BAND 0.82 TO 0.88 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 6	SPECTRAL BAND 0.98 TO 1.04 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 7	SPECTRAL BAND 10.10 TO 11.00 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0

AREA PROCESSED.....LINES 437- 800 (BY 3)
COLUMNS 117- 444 (BY 3)



10103 CPU TIME USED WAS 189.347 SECONDS. (LARSMN)

ORIGINAL PAGE IS
OF POOR QUALITY

NONSUPERVISED ECHO: CLASSIFICATION - PHASE 2

(NS2ECHO)

The NS2ECHO function is an implementation of the classification phase of the nonsupervised Extraction and Classification of Homogeneous Objects (ECHO) algorithm. It performs maximum likelihood sample classification of objects that were identified during the nonsupervised field extraction phase (NS1ECHO) and a point-by-point maximum likelihood classification of the constituent points of cells which were labeled singular by the NS1ECHO function. After performing the classification, it writes the results on a Classification Results File to be printed later.

Input to the function:

- An Intermediate Results Tape containing statistics and portions of the fields identified by NS1ECHO and the data vectors of pixels from singular (non-homogeneous) cells.
- Control cards to select the processing options.
- A Statistics File containing the statistical description of the training classes.

The principal output is a LARSYS Classification Results File, which is placed on tape. This file is normally used as input to the LARSYS PRINTRESULTS function for production of a variety of printed map and tabular outputs for display of results and evaluation of the classification. The Classification Results File is also the primary input to the LARSYS COPYRESULTS, LISTRESULTS, and PUNCHSTATISTICS functions.

ORIGINAL PAGE IS
OF POOR QUALITY

Inputs

The principal inputs to the nonsupervised ECHO classification function (NS2ECHO) are the nonsupervised Intermediate Results File which has been produced by the nonsupervised ECHO field extraction algorithm and the LARSYS Statistics File. The Statistics File must be included as card deck input to this function. In addition to the Statistics File, the user must provide an input deck designating the location of the Intermediate Results File and the desired destination of the Classification Results. An example of the use of the control cards and the correct location for the LARSYS Statistics File is shown in Figure 12.

Specification of Intermediate Results Location: The user must specify the tape and file containing the Intermediate Results. This is done by means of an INTERMEDIATE card. The form of this card is:

INTERMEDIATE TAPE(XXX), FILE(YY)

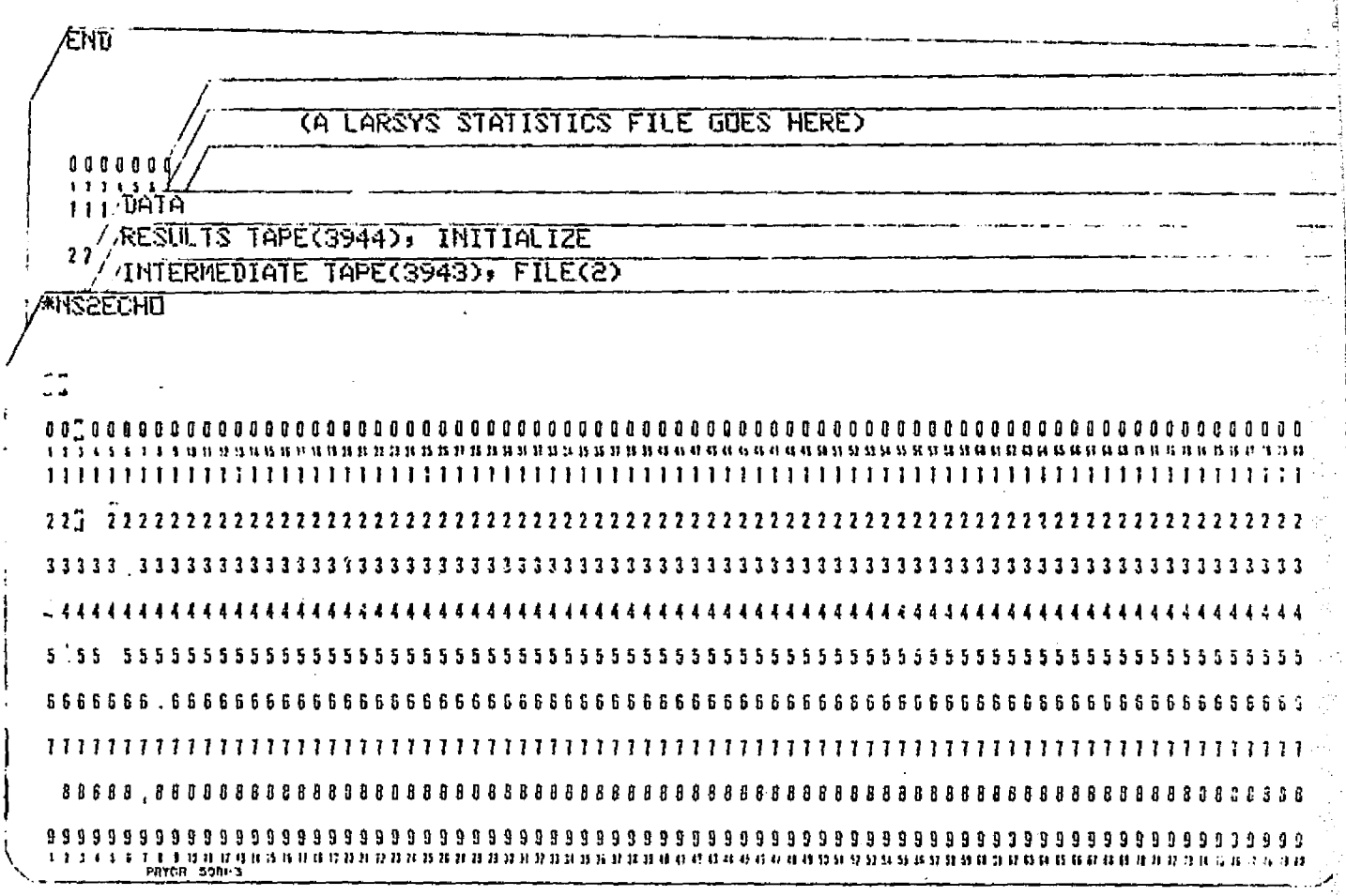
where XXX is the number of an Intermediate Results Tape and YY is the file containing the desired results. Note: Only Intermediate Results File produced by the nonsupervised ECHO field extraction algorithm (NS1ECHO) may be used by NS2ECHO.

Optional Selection of Training Classes: The user may select the training classes from the Statistics File that are to be used by nonsupervised ECHO's classification phase (phase 2), and he may combine training classes into pools. These options are exercised by using the 'CLASSES' control card. For example, if the user wished to use only classes 1, 3, and 5 of seven

Figure 12

Example Control Cards for the Nonsupervised ECHO

Classification Phase



ORIGINAL PAGE IS
OF POOR QUALITY

training classes previously defined, the control card entry would be:

CLASSES 1, 3, 5

In this case, the class name assigned by the statistics function at classes 1, 3, and 5 will be retained by SECHO and the other classes will be totally ignored.

To combine two or more classes into one class, the user assigns a name (up to eight characters) to the pooled class to be created and specifies the classes to be included in the pooled class. For example, assume there are eight classes available in the training statistics, and the user wishes to process the following combinations:

- POOLA (Pool A) will be classes 1 and 2.
- POOLB will be classes 4, 6, and 7.
- POOLC will be class 5 only.
- Classes 3 and 8 will be ignored.

The control card format to specify this option will be:

CLASSES POOLA(1/1,2/),POOLB(2/4,6,7/),POOLC(3/5/)

Note that the number immediately following a left parenthesis specifies the pool sequence. Pool sequence numbers must be in ascending order. Note also that the classes to be pooled (and named) are enclosed by slashes (/).

When no 'CLASSES' card is specified, all the classes in the statistics deck will be considered by the nonsupervised ECHO processor.

<u>Specification of Class Statistics:</u> Class statistics must be supplied to the nonsupervised ECHO classification phase before

classification may proceed. Unlike the supervised ECHO classifier and the LARSYS CLASSIFYPOINTS algorithm, the Statistics File must be provided to the nonsupervised ECHO processor in the control card file. The LARSYS Statistics File is inserted into the input deck immediately before the 'END' card. The Statistics file must be preceeded by a 'DATA' card (see Figure 12).

Outputs

Classification Results File: The principal output of the NS2ECHO function is the Classification Results File, which is, in turn, the primary input to four other LARSYS functions: PRINTRESULTS, COPYRESULTS, LISTRESULTS, and PUNCHSTATISTICS. The user must specify where this file will be stored by using a 'RESULTS' control card in one of two forms:

```
RESULTS TAPE (xxx), FILE(nn)
RESULTS INITIALIZE, TAPE(xxx)
```

The first control card is used to add the file to a tape already containing Classification Results Files. If a file in the specified destination already exists on the tape, the user will be notified by a message. He then has the option of writing over the old file, specifying a new tape and file number, or stopping execution. The second control card specifies that a new results tape be used, and the 'INITIALIZE' parameter requests that the proper header information be placed at the beginning of the new tape so a file may be written. A new tape must always be initialized before it can be used to store classification results.

A unique "Classification Study Number", based on the date and time of the run, is part of each Classification Results File. The number, identified as "Classification Study", is included on any outputs that are subsequently derived from the results file. The form of the identification number is "ydddsssss"; where y is the last digit of the year, ddd is the Julian date (day of the year, 001-365), and sssss is the total number of seconds since the previous midnight.

The principal data on the file are the class assignments for each point of the classification run. A separate record is written for each data line classified. This record contains, for each point in the line, the class number associated with the class to which the point was assigned. These classification results are used by the PRINTRESULTS function to produce detailed maps of the classified area as well as tables of the test fields, training fields, and class performance. For more information on these products, refer to the description of the Printresults function in the LARSYS User's Manual[8].

In addition to the point-by-point classification results, the file contains other data related to the classification run:

- A complete copy of the Statistics File that was used as input to the run. This file may be punched on cards by using the LARSYS Punchstatistics function.
- Summary information about the classification, the channels and classes which were used. A formatted listing of this information may be produced by using the Listresults function. This listing is also a secondary product of both the Punchstatistics and the Copyresults function.

ORIGINAL PAGE IS
OF POOR QUALITY

- Results statistics (mean vectors and covariance matrices) for the classes and channels used in the classification.

Standard Printer Output: Figure 13 presents the standard printer output produced by the classification phase of the nonsupervised ECHO processor. The nonsupervised ECHO classification phase (phase 2) has only two printer outputs, a reproduction of the user's control card deck and a summary of the particular characteristics of the classification, the Classification Study Number, the number of pooled classes, the number of channels, the number of fields and the channels selected.

Figure 13

Standard Printer Output for the Classification Phase of Nonsupervised ECHO (NS2ECHO)

```
ECHOP6                          LABORATORY FOR APPLICATIONS OF REMOTE SENSING          NOV  10,1977
CARLOS ARTURO                                PURDUE UNIVERSITY                         03 27 25 PM
                                                                                       LARSYS VERSION 3

    -COMM AIRCRAFT
    *NS2
    INTERMEDIATE TAPE(3943),FILE(3)
    RESUL TAPE(0),INITIA
    DATA

NONSUPERVISED ECHO (PHASE2) INFORMATION....

    CLASSIFICATION STUDY.... 731455868
    NO. OF CLASSES................   11
    NO. OF CHANNELS...............    6
    NO. OF FIELDS.................    0
    CHANNELS SELECTED ARE........    1,  2,  3,  5,  6,  7,

10103  CPU TIME USED WAS   61.554 SECONDS.     (LARSHN)
```

ORIGINAL PAGE IS
OF POOR QUALITY

A DISCUSSION OF THE ECHO ALGORITHMS

The following material assumes that the reader is already aware of the general nature of the ECHO process, including data and parameter inputs required and the outputs produced by the programs which are discussed.

Background

As we have noted, the ECHO process consists of two phases: object finding and sample classification. Furthermore, there are both "supervised" and "nonsupervised" versions of the process, the principal difference in the two versions being determined by whether or not a set of precalculated class statistics is used in the object-finding phase. The purpose of this section is to outline the mathematical basis for the supervised ECHO process and to describe its implementation in the form of an algorithm compatible with LARSYS-like data analysis. In a later section we shall do the same for the unsupervised ECHO process.

In all that follows, it is implicitly assumed that the class-conditional density functions are multivariate normal; i.e., for the ith class and for pixel vector X, the n-variate probability density function can be written as:

$$p(X|\omega_i) = \frac{1}{(2\pi)^{n/2}|K_i|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(X - M_i)^t K_i^{-1}(X - M_i)\right]$$

where

K_i is the covariance matrix for class ω_i

M_i is the mean vector for class ω_i

n is the dimensionality of the data (pixel vector X).

In general, the covariance matrices and mean vectors will be estimates from collections of pixels assumed to belong to a given class.

It will also be assumed that the data from adjacent or nearby pixels are class-conditionally independent. This will allow the joint probability density function for a collection of such pixels, all assumed to belong to the same class, to be written in product form:

$$p(\underline{X}|\underline{\omega}) = p(X_1|\omega_{i_1})p(X_2|\omega_{i_2}) \cdots p(X_s|\omega_{i_s})$$

where $\underline{X} = \{X_1, X_2, \ldots, X_s\}$ is such a collection (sample) consisting of s pixels belonging respectively to classes $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_s}$.

<u>Supervised Object Finding (SECHO)</u>: The object-finding process is in itself a two-phase process. In the first phase, referred to as "cell selection", the scene is partitioned into a rectangular grid of small groups of pixels, called "cells". As implemented in ECHO, each group or cell is a sqaure with N pixels on a side (N is an input to the program). To remain a cell, the group must satisfy a statistical homogeneity criterion, described in more detail below. A cell failing to satisfy this criterion is called "singular", and its pixels will be classified individually.

The supervised cell selection homogeneity test used in ECHO is performed as follows. Define the quantity

$$Q_j(\underline{Y}) = \sum_{i=1}^{s} (Y_i - M_j)^t K_j^{-1} (Y_i - M_j)$$

$$= tr(K_j^{-1} \sum_{i=1}^{s} Y_i Y_i^t) - 2M_j^t K_j^{-1} \sum_{i=1}^{s} Y_i + sM_j^t K_j^{-1} M_j$$

where

Y_i is the i<u>th</u> pixel vector in the cell being tested

s is the number of pixels in the cell ($s = N^2$)

K_j is the sample covariance matrix for the j<u>th</u> training class

M_j is the sample mean vector for the j<u>th</u> training class.

This quadratic form is a measure of the statistical distance of the collection of data contained in the cell from the distribution of the training data for the j<u>th</u> class. Now let ω^* be the class for which the "log-likelihood" of the cell is maximum; i.e., $\ln p(\underline{Y}|\omega^*) = \max \ln p(\underline{Y}|\omega_j) = \max_j\left[-\frac{s}{2} \ln|2\pi K_j| - \frac{1}{2}Q_j(\underline{Y})\right]$ and let $Q^*(\underline{Y})$ be the value of the corresponding quadratic form. A cell is defined to be singular (and its pixels will be classified individually) if $Q^*(\underline{Y}) > c$, where c is a user-specified threshold value. Otherwise, we accept the hypothesis that the cell $\underline{Y}$ is homogeneous and treat it as a unit.

This criterion has the particular advantage that it tends to "reject" not only inhomogeneous cells, but "unrecognizable" cells as well (cells very unlikely to belong to any of the training classes). Another advantage is that the computations involved are particularly compatible with the supervised annexation criterion and the maximum likelihood sample classifier.

Also of importance, the distribution of the Q_j values can be shown to be chi-squared with $s \cdot n$ degrees of freedom. This fact is used in determining appropriate values of the threshold parameter c.

In the second object-finding phase, called "annexation", a cell is compared to an adjacent "field", which is simply a group

of one or more spatially connected cells which have already been merged. If the two samples are statistically similar, according to a test we shall detail below, then the cell is merged or "annexed" into the field. Otherwise the cell is compared to another adjacent field, if one exists, or it becomes a new field by itself.

In ECHO, the supervised annexation similarity test is based on the statistic

ORIGINAL PAGE IS
OF POOR QUALITY

$$\Lambda = \frac{\max_i p(\underline{X}|\omega_i)\ p(\underline{Y}|\omega_i)}{\max_i p(\underline{X}|\omega_i)\ \max_j p(\underline{Y}|\omega_j)}$$

where $\underline{X}$ is the collection of pixels forming the field and $\underline{Y}$ is the collection of pixels in the cell. Notice that Λ has a value between 0 and 1. It is closest to 1 when both $p(\underline{X}|\omega_i)$ and $p(\underline{X}|\omega_j)$ have their maximum value for the same class.

Thus the annexation criterion may be stated as follows: The cell is assumed to belong to the same class as the field and is annexed to the field if $\Lambda \geq T$, where T is a threshold value ($0 \leq T \leq 1$). Otherwise the cell is considered significantly different from the field and no annexation takes place.

For purposes of computational efficiency it is preferable to work with the logarithm of Λ. This not only converts the statistic into a difference of sums (rather than a quotient of products) but also simplifies computation of the $p(\underline{X}|\omega_i)$, etc., under the multivariate normal assumption noted earlier. We restate the annexation criterion as: Assume the cell belongs to the same class as the field and annex the two provided $-\log \Lambda \leq t$, where t is a user-specified threshold value ($t \geq 0$). Note that t can be related to the parameter T by the expression $T = 10^{-t}$.

Maximum Likelihood Sample Classification (SECHO and NS2ECHO): Regardless of whether the object-finding method used was supervised or nonsupervised, the resulting objects are always classified by a supervised classification rule. In other words, training class statistics must be provided for use by the classification rule.

Therefore, let K_j and M_j be the covariance matrix and mean vector, respectively, for the jth training class (estimated from training data), and let s be the number of pixels in an object to be classified. The maximum likelihood sample classification rule is:

Decide $\underline{X} = \{X_1, X_2, \ldots, X_s\}$ belongs to class ω^* if and only if $p(\underline{X}|\omega^*) = \max_j p(\underline{X}|\omega_j)$

or equivalently

$$\ln p(\underline{X}|\omega^*) = \max_j \ln p(\underline{X}|\omega_j)$$

Under the assumption noted earlier of class-conditional independence of pixels within an object, we have

$$p(\underline{X}|\omega_j) = p(X_1|\omega_j)\ p(X_2|\omega_j) \cdots p(X_s|\omega_j)$$

or

$$\ln p(\underline{X}|\omega_j) = \sum_{k=1}^{s} \ln p(X_k|\omega_j)$$

Taking into account the multivariate normal assumption, this becomes, after some manipulation:

$$\ln p(\underline{X}|\omega_j) = -\frac{s}{2} \ln|2\pi K_j| - \frac{1}{2}Q_j(\underline{X})$$

$$= -\frac{1}{2} \operatorname{tr}(K_j^{-1}S_2) + M_j^t K_j^{-1} S_1 - \frac{s}{2} M_j^t K_j^{-1} M_j - \frac{s}{2} \ln|2\pi K_j|$$

ORIGINAL PAGE IS
OF POOR QUALITY

where

$$S_1 = \sum_{i=1}^{s} X_i \text{ and } S_2 = \sum_{i=1}^{s} X_i X_i^t,$$

the sums taken over all pixels in the object to be classified. Notice that S_1 is a vector and S_2 is a matrix.

Expressed in this way, two terms in the "log-likelihood" depend on the data to be classified and the training statistics, whereas the third and fourth terms depend only on the training statistics. Thus the latter two terms need to be evaluated once, whereas the first two terms need to be re-evaluated for each data point to be classified.

The expression above for the log-likelihood is perfectly valid for the case s = 1. It provides the computation necessary for classifying the individual pixels resulting from cells which fail to pass the cell selection homogeneity test.

Nonsupervised Object Finding (NS1ECHO): It was noted in the discussion of the supervised ECHO algorithms that the only difference between the supervised and nonsupervised ECHO processes is in the approach used for object finding. The supervised ECHO process utilizes predetermined class statistics in partitioning the image data into objects. The nonsupervised process must accomplish the partitioning without benefit of predetermined class statistics.* Both processes utilize the same maximum likelihood sample classification algorithm.

---

*Since the supervised object-finding process uses more *a priori* knowledge about the data, it might be expected that it would perform somewhat more reliably than the nonsupervised version. In fact, this has been demonstrated experimentally [Kettig, R. L. and D. A. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects," IEEE Trans. Geoscience Electronics, vol. GE-14 no. 1, January 1976].

The nonsupervised object-finding process, like the supervised version, is a two-phase process involving "cell selection" and "annexation". In the cell selection phase, the scene is partitioned into a rectangular grid of small groups of pixels, called "cells". Each cell is a square with N pixels on a side (the cell width, N, is an input to the program). To remain a cell, the group must satisfy a statistical homogeneity criterion (described below). A cell failing to satisfy this criterion is called "singular", and its pixels will be classified individually.

The nonsupervised cell selection homogeneity test used in ECHO is quite simple. The sample variance of the data in each channel divided by the corresponding channel mean is compared to a user-specified threshold which is an input to the program. If the threshold is exceeded in <u>any</u> channel, the cell is considered singular and its pixels dealt with accordingly, i.e., classified individually. Although more powerful statistical tests have been investigated for cell selection purposes, none have been found more effective than the one described here. Furthermore, the more powerful tests often impose undesirable requirements on the minimum usable cell size.

In the annexation phase of the nonsupervised object-finding process, a cell is compared to an adjacent "field", which is simply a group of one or more spatially connected cells which have already been merged. If the two samples are statistically similar, according to a test described below, then the cell is merged or annexed into the field. Otherwise, the cell is

compared to another adjacent field, if one exists, or it becomes a new field by itself.

The test implemented for annexation in ECHO is a "multi-univariate" test rather than a truly multivariate test. That is, the test is based on examining sequentially the statistics associated with each data channel rather than examining the multivariate statistics for all channels combined. Extensive testing has shown that this approach is best when the cell size is small, because the number of pixels in the cell may not be sufficient to provide a good estimate of the multivariate statistics (particularly the cell covariance matrix).

In this case, the means and the variances are tested independently. First the cell and field means are tested for similarity based on the statistic

$$\Lambda_{1i} = \frac{(T-2)rs}{T} \frac{(\bar{x}_i - \bar{y}_i)^2}{a_i}, \quad i = 1, 2, \ldots, n$$

where

$\bar{x}_i$ is the field mean in channel i

$\bar{y}_i$ is the cell mean in channel i

r is the number of pixels in the field

s is the number of pixels in the cell

$T = r + s$

$$a_i = \sum_{j=1}^{r} (x_{ij} - \bar{x}_i)^2 + \sum_{j=1}^{s} (y_{ij} - \bar{y}_i)^2.$$

Under the hypothesis that field and cell have the same distribution, this statistic has an F distribution with 1 and (T-2) degrees of freedom. Large values of Λ_{1i} indicate that the hypothesis is not true. The field and cell will not be merged

if any component of the means fails to pass this test at a level of significance defined by a user-supplied threshold constant.

If the means pass the similarity test, then the channel variances are tested. The cell and field variances are tested for similarity based on the statistic

$$\Lambda_{2i} = \frac{(1 - g + \frac{2}{3} g^2) G_i}{1 - (1 - g + \frac{2}{3} g^2) \frac{g^2}{3} G_i}, \qquad i = 1, 2, \ldots, n$$

where

$$g = \frac{1}{3}\left(\frac{1}{r-1} + \frac{1}{s-1} - \frac{1}{T-2}\right)$$

r, s, T are as defined above

and

$$G_i = \log_e \left\{ \left|\frac{a_i}{T-2}\right|^{(T-2)} \left[\left|\frac{a_{xi}}{r-1}\right|^{r-1} \left|\frac{a_{yi}}{s-1}\right|^{(s-1)}\right]^{-1} \right\}$$

where

$$a_{xi} = \sum_{j=1}^{r} (x_{ij} - \bar{x}_i)^2$$

$$a_{yi} = \sum_{j=1}^{s} (y_{ij} - \bar{y}_i)^2$$

$$a_i = a_{xi} + a_{yi}$$

Under the hypothesis that the field and cell have the same distribution, Λ_{2i} has an F distribution with 1 and (3, degrees of freedom. The field and cell will not be merged if the data in any of the channels fails to pass this test at a level of significance defined by a user-supplied threshold constant.

Maximum Likelihood Sample Classification (NS2ECHO): The objects defined by the nonsupervised objected-finding process may be subsequently classified by a sample classification rule. This is a logical step to perform only if it is done by a supervised sample classifier, however, and we have already noted that the supervised classifier used is the same as that used following supervised object finding.

ECHO PARAMETER SELECTION GUIDELINES

This section discusses settings of the object extraction parameters required by the supervised ECHO (SECHO) and non-supervised ECHO (NS2ECHO) processors. These comments have their origin in the test and evaluation of the ECHO processors performed between June 1976 and August 1977 and reported in the LARS Final Technical Report to JSC in May 1977[5] and the LARS Final Technical Report to JSC in November 1977[6].

The results of Landsat and simulated Thematic Mapper data are discussed. The Landsat data were drawn from two sources, LACIE/SRS data sets collected over Kansas where the principal information classes (wheat and other) are in relatively large fields and CITARS data sets collected over Indiana and Illinois where the principal information classes (corn, soybeans, and other) occur in relatively small fields.

The simulated Thematic Mapper data collected over Kansas and North Dakota has relatively large fields and is simulated at 30, 40, 50 and 60 meter resolutions.

Six variables were monitored to evaluate the ECHO algorithms:

- CPU time,
- Field center pixel classification performance,
- Training field classification performance,
- Full field classification performance
- RMS proportion estimate error, and
- Classification variability.

These variables are related to reasons for adopting a new classification technique: cost, accuracy, and usability of results. The CPU time required to perform a classification is one way to measure the cost of classification. Field center pixel, full field, and training field performances and RMS proportion estimate error are all ways to evaluate the accuracy of the classifier. Classification variability is a measure of "salt and pepper effect" in classification results.

The CPU time required to execute each of the ECHO classifications has bee recorded so that the effects of varying the cell homogeneity and annexation thresholds may be monitored. The CPU time required to perform the perpoint classifications have been adjusted to reflect the increased efficiency of the LARSYS perpoint classifier which is coded in assembly language. Thus, the CPU time recorded for a perpoint classification is what a FORTRAN classifier would have required to perform the classification.

The indices of classification performance were applied in several ways. Classification accuracy (identification) was evaluated utilizing field center pixel, "full field" and test field sample performances for all data sets. Proportion estimation was carried out for the Landsat and Simulated Thematic Mapper data sets.

The training performance is the overall classification accuracy (number of training pixels correctly classified divided by the total number of training pixels) of the pixels used to calculate the class statistics. Field center pixel performance is the overall classification accuracy of pixels inset at least one pixel from the field boundary. For the registered

LACIE/SRS data the field center pixels are inset at least two pixels from the field boundary. Although this procedure insures that the pixels examined are not mixture pixels, it has the unfortunate effect of eliminating smaller fields from consideration. The third measure of classification accuracy, "full field" performance, includes those pixels on the boundaries of the fields in the classification performance. The "full field" pixels were generated by expanding the field center pixel boundaries one pixel in all directions.

The RMS error of informational class proportion estimates for each flightline was found by calculating the percent of the flightline classified as a particular class and comparing it with the ground-collected estimage 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 informational class i, and

C_i = percent of class i estimated from ground-collected data.

RMS error is calculated for the Landsat and Thematic Mapper data runs. The Agricultural Stabilization and Conservation Service (ASCS) provided the ground truth proportion estimates for the simulated Thematic Mapper data set. Proportion estimates for the 1974 LACIE/SRS segments were provided in ground truth packets received from JSC. The SRS county proportion estimates were used to calculate RMS proportion error for the CITARS data set.

Average variability is a measure of the rate of change from one information class to another. It should reflect the degree to which ECHO reduces the "salt and pepper effect" which is sometimes present in perpoint classifications. Variability is calculated by systematically selecting 50 lines of the classified area, counting the number of information class changes, and dividing by the number of opportunities for class changes.

$$\text{Variability} - NCC/(50*(NS-1)) \qquad (2)$$

Where:

NCC = the number of class changes over the 50 selected lines, and

NS = the number of classified pixels/lines.

Supervised ECHO Parameters

Landsat Parameter Selection for the Supervised ECHO Processor: The LACIE and the CITARS data sets appear to have different requirements in parameter settings. This is very reasonable as the two data sets have a very different range of average field sizes and different ground cover types. The average field size in the CITARS data sets range from 17 in Shelby to 23 in Livingston; the average field size in the LACIE data sets range from 78 in Haskell to 91 in Graham. The LACIE data sets are composed of classes of wheat and other while CITARS data sets are corn, soybeans, and other.

The cell width setting which optimizes the field center pixel and full field performances varies over the data sets with cell width 2 most frequently providing the optimal results. There appears to be a slight tendency toward larger values of cell width showing superior performance at smaller values of

average field size which is not consistent with our expectations and difficult to justify theorically. The training performance, however, is consistently optimum at a cell width of 2. The proportion estimate error follows a different pattern for the CITARS data sets than for the LACIE data sets. For the CITARS data sets a cell width of 4 is best when the number of spectral classes is less than 10; when the number of spectral classes is greater than or equal to 10, a cell width of 2 is better. The opposite pattern holds for the LACIE data sets. A cell width of 2 is best when the number of spectral classes is less than 10; and a value of 4 or 5 is better when the number of spectral classes is greater than or equal to 10. For both CPU time and classification variability, cell width settings of 4 for the CITARS data sets and from 2 to 4 for the LACIE data sets will give optimal results.

ORIGINAL PAGE IS OF POOR QUALITY

The optimal cell homogeneity settings are rather scattered and inconsistent for field center pixel, full field, and training performances as well as proportion estimate error. There appears to be a slight tendency toward larger values of the cell homogeneity parameter optimizing field center pixel, full field and training field performances as the average field size increases. For field center pixel performance, no one value consistently yields superior results for the CITARS; however, a homogeneity setting 79 is most often optimum for the LACIE data sets. For full field performance values around 15 and around 118 appear often as the optimal cell homogeneity setting for the CITARS data sets; for the LACIE data sets, homogeneity settings around 40 and 80 often give optimal values. For training performance, homogeneity settings between 60 and 120 appear equally often as

the optimal performance settings; a more narrow recommendation is difficult to make. A recommendation of cell homogeneity setting is less difficult to make to optimize CPU time required or classification variability produced, as a setting of 120 or more always minimized both.

The optimal cell annexation parameter settings are somewhat inconsistent for field center pixel performance. There is a slight tendency for larger annexation values (2-4) to yield improved field center pixel performances for runs having large average field sizes (above 60 pixels). The CITARS data sets have optimal field center pixel performances with settings of 0 or 1 while the LACIE data sets, with large average field sizes, have optimal performances for annexation of 2 or 4. Similarly, for full field performance, the CITARS data sets generally perform best with an annexation setting of 1 and LACIE data sets perform best with an annexation setting of 2. For training performance, a setting of 2 gives the optimum for most Landsat data sets. Both proportion estimate error and classification variability are minimized with an annexation setting of 4; while CPU time is lowest with annexation settings of 1 or 4.

Figure 14

Supervised ECHO Landsat Parameter Settings
to Optimize Six Variables

Dependent Variables	Cell Width Setting	Homogeneity Parameter Setting	Annexation Threshold Setting
Field Center Pixel Performance	2	20-80 (higher as AFS increases)	AFS<40, 0 to 2 AFS>40, 2 to 4
Full Field Performance	AFS<40, 3 AFS>40, 2	AFS<40, 15-30 or 100-130 AFS>40, 35-130	AFS<40, 1 AFS>40, 2
Training Field Performance	2	80-120	2
Proportion Estimate Error	SP<10 SP>10 AFS<40 4 2 AFS>40 2 4-5	20-80 (higher as AFS increases)	4
Classification Variability	Largest Possible	∞	∞
CPU Time	2-4 (higher as AFS increases)	∞	∞

AFS = Average Field Size (AFS<40, CITARS data;
AFS>40, LACIE data)

SP = Spectral Classes

Thematic Mapper Parameter Selection for the Supervised ECHO Processor: The results are fairly consistent except at the 50 meter resolution. For the supervised ECHO processor, a cell width of 2 is best for field center pixel performance and training field performance. This choice is also best for full field performance, CPU time, and variability except at the 50 meter resolution. The root mean square error is minimized when the cell width is the integer part of the square root of the average field size.

The six settings of cell homogeneity which have been tested for the Supervised ECHO on simulated Thematic Mapper data are to 19, 32, 45, 68, 91 and 136. Field center pixel performance is highest at settings between 68 and 91, except at the 50 meter resolution where a setting between 19 and 32 does better. Similarly, for full field performance, the best cell homogeneity setting is around 68, except at the 50 meter resolution where values between 19 and 32 are better. Training field performance is best when the homogeneity parameter is set around 19 when the average field size is less than 75 pixels; otherwise homogeneity values between 32 and 45 yield higher training field performances. The cell homogeneity setting is less important in optimizing proportion estimates; when the cell width is the integer part of the square root of the average field size. All cell homogeneity settings between 32 and 91 produced very similar results. For both variability and CPU time, the highest homogeneity setting tested (136) yielded the optimal results.

For all the measures except proportion estimate error, a cell annexation setting of 4 yielded superior results. With

ORIGINAL PAGE IS
OF POOR QUALITY

respect to proportion estimation error, all settings produced similar results.

Figure 15

Supervised ECHO Simulated Thematic Mapper Parameter Settings to Optimize Six Variables

Dependent Variable	Cell Width Setting	Homogeneity Parameter Setting	Annexation Threshold
Field Center Pixel Performance	2	60-95**	4
Full Field Performance	2*	60-70**	4
Training Field Performance	2	AFS<75 15-25 AFS>75 30-50	4
Proportion Estimate Error	4-6 (AFS	>30	any
Classification Variability	2-4 (Larger as AFS increases)	∞	∞
CPU Time	2-4 (Larger as AFS increases)	∞	∞

* Except at resolution 50
**Except at resolution 50 where 20-40 is optimal

AFS = Average Field Size

Nonsupervised ECHO Parameters

Nonsupervised ECHO Landsat Parameter Selection: Only a cell width of 2 was used on the nonsupervised ECHO data sets. The recommendations are thus made only on cell homogeneity and cell annexation parameter settings.

The optimal cell homogeneity settings are not very consistent for field center pixel, full field and training performances, where the optimum tends to alternate between 0.05 and 0.25. For proportion estimate error, a setting of 0.05 is best for CITARS data sets while a setting of 0.10 is best for the LACIE data sets. For variability and CPU time, a setting of 0.25 is the optimum for almost all data sets. Cell annexation settings of 0.010 give optimal results for field center pixel, full field, and training performances and for proportion estimate error. A cell annexation setting of 0.001 yield classification results with the lowest classification variability ("salt and pepper" effect) and requires the least CPU time to execute for a given area.

Figure 16

Nonsupervised ECHO Landsat Parameter Settings
to Optimize Six Variables

Dependent Variable	Cell Width Setting	Homogeneity Parameter Setting	Annexation Parameter Setting
Field Center Pixel Performance	Only a cell width of 2 was tested	.05 - .25	.005, .01, .025
Full Field Performance		.05 - .25	.005, .01, .025
Training Field Performance		.05 - .10	.005, .01, .025
Proportion Estimate Error		.05 - .10 (larger as AFS increases)	.005, .01, .025
Classification Variability		∞	.001
CPU Time		∞	.001

AFS = Average Field Size

ORIGINAL PAGE IS
OF POOR QUALITY

Nonsupervised ECHO Thematic Mapper Parameter Selection: Many of the parameter settings for the nonsupervised ECHO algorithm appear to be related to the number of spectral classes. For both field center pixel and full field performance, a cell width of 2 is better where the number of spectral classes is less than 30, a cell width of 3 is better when the number of spectral classes is greater than 30. For training performance, the same pattern holds except that the dividing value is 20 spectral classes. The reverse pattern appears for variability, with the cell width of 2 minimizing variability in the classification results if the number of spectral classes is greater than 30, and cell width 3 minimizing classification variability when the number of spectral classes is less than 30. A cell width parameter setting of 3 minimizes the proportion estimate error or the CPU time required.

The optimal cell homogeneity parameter settings also appear to be related to the number of spectral classes. For both field center pixel and full field performance, a cell homogeneity setting of 0.05 is best when the number of spectral classes is greater than 30 while a value of 0.10 is better when the number of spectral classes is less than 30. For both training performance and root mean square error, a cell homogeneity setting of 0.05 gives optimal results while both CPU time and the variability of the classification are minimized with a cell homogeneity parameter of 0.25.

A cell annexation parameter setting around 0.10 yields the optimal field center pixel performance, while values of between

0.010 and 0.100 optimize full field performance and minimize proportion estimation error. Training performance, CPU time required, and variability present are all optimized by a cell annexation setting of 0.001.

Figure 17

Nonsupervised ECHO Simulated Thematic Mapper
Parameter Setting to Optimize Six Variables

Dependent Variable	Cell Width Setting	Homogeneity Parameter Setting	Annexation Threshold
Field Center Pixel Performance	SP<30, 2 SP>30, 3	SP<30, .08-.12 SP>30, .03-.07	.005, .01, .025
Full Field Performance	SP<30, 2 SP>30, 3	SP<30, .08-.12 SP>30, .03-.07	.01, .025, .05, .1
Training Field Performance	SP<20, 2 SP>20, 3	.03-.07	.001, .025
Proportion Estimate Error	3	.03-.07	.01, .025, .05, .10
Classification Variability	SP<30, 3 SP>30, 2	∞	.001
CPU Time	As large as possible	∞	.001

SP = Spectral Classes

ORIGINAL PAGE IS
OF POOR QUALITY

REFERENCES

1. Landgrebe, D. A. 1975.
   Final Report, NASA Contract NAS9-14016, June 1, 1974-May 31, 1975. pp. x-1 - x-11.

2. Kettig, R. L. 1975.
   Computer Classification of Remotely Sensed Multispectral Image Data by Extraction and Classification of Homogeneous Objects. Ph.D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, Indiana.

3. Kettig, R. L. and Landgrebe, D. A. 1973.
   Automatic Boundary Finding and Sample Classification of Remotely Sensed Multispectral Data.
   LARS Information Note 041773, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

4. Kettig, R. L. and Landgrebe, D. A. 1975.
   Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects.
   Proceedings Symposium on Machine Processing of Remotely Sensed Data, Purdue University, IEEE Catalog No. 75CH1009-0-C.

5. Landgrebe, D. A. 1977.
   Final Technical Report, NASA Contract NAS9-14970, June 1, 1976-May 31, 1977, vol.1.

6. Landgrebe, D. A. , Principal Investigator, 1977.
   Final Report, NASA Contract NAS9-14970, November 28, 1977, Volume III, Test of Spectral/Spatial Classifier, J. L. Kast and B. J. Davis.

7. S. K. Hunt, Editor, 1973.
   LARSYS System Manual, Purdue University, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana. Copyright © 1973.

8. T. L. Phillips, Editor, 1973.
   LARSYS Version 3 User's Manual, Purdue University, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana. Copyright © 1973.