FRIS

FOREST RESOURCE INFORMATION SYSTEM

NASA ST. REGIS LARS

LARSFRIS USER'S MANUAL

Volume 2

Purdue University
Laboratory for Applications
of Remote Sensing

LARS Contract
Report No. 100280
October 1, 1980

Star Information Form

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle LARSFRIS User's Manual Volume 2		5. Report Date October 1, 1980 6. Performing Organization Code
7. Author(s) LARS Staff, R.P. Mroczynski, editor		8. Performing Organization Report No.
9. Performing Organization Name and Address Laboratory for Applications of Remote Sensing Purdue University West Lafayette, IN 47906		10. Work Unit No. 100280 11. Contract or Grant No. NAS9-15325
12. Sponsoring Agency Name and Address R.E. Joosten/SF5 NASA/Johnson Space Center Houston, TX 77058		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This document contains user instructions for the proper use and application of the Software which comprises the LARSFRIS package. LARSFRIS represents a compilation of software developed over a number of years by the staff at Purdue University's Laboratory for Applications of Remote Sensing. The software packages are designed to help the user analyses digital image data such as that collected by the Landsat Multispectral scanner. This is one of five documents that comprise the LARSFRIS package.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Landsat analysis Digital Image data User Documentation Software packages	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*

*For sale by the National Technical Information Service, Springfield, Virginia 22161

ACKNOWLEDGEMENTS

An undertaking of the magnitude of the LARSFRIS documentation, albeit only a modification to existing materials, depended on the individual dedication of many people. A number of LARS staff contributed to updating LARSYS ver. 3.1 and integrating LARSYSDV (developmental software) into the final LARSFRIS software package.

LARS staff who made significant contributions to either creating new or updating existing program modules included; Sue Schwingendorf, Bill Shelley, Carol Jobusch, Joan Buis, Luis Bartolucci, Louis Lang, and John Cain. Kay Hunt deserves special thanks for coordinating and organizing staff efforts.

Typing of the manuscript for the LARSFRIS documentation was ably handled by; Dee Dee Dexter, Sylvia Johnston, Pam Burroff, and Bonnie Phibbs. Assistance in editorial matters was provided by Doug Morrison and Davida Parks, and Sue Ferringer provided graphic inputs.

Special thanks are also appropriate for members of the FRIS Steering Committee; especially G. R. Barker of the St. Regis Paper Company, and R. E. Joosten of the National Aeronautics and Space Administration, for their patience and sage council during the preparation of these volumes.

Preparation of this documentation was supported by NASA Contract NAS 9-15325.

PREFACE

The documentation of the LARSFRIS system closely parallels existing LARSYS Version 3.1 documentation. The major differences are in the addition of certain program modules which provide the user greater flexibility in the analysis of multispectral data. The LARSFRIS documentation exists in three parts: LARSFRIS Program Abstracts, LARSFRIS System Manual, and LARSFRIS User's Manual.

The first of these contains the documentation of each Fortran and Assembler routine and each CMS Executive routine in LARSFRIS. These program abstracts are provided for programmers who are required to revise and/or maintain these routines.

The second manual, LARSFRIS System Manual, is directed primarily to programmers and analysts who maintain or revise the system or write new functions that must be interfaced with LARSYS. It contains detailed information of (and references to) the hardware and software framework upon which the system was built, the internal organization of the software, the organization of the data fields, and a discussion of special techniques that were used in the implementation of LARSFRIS.

This manual, LARSFRIS User's Manual, contains a comprehensive description of the functional organization of the system, the processing functions provided, and the manner in which the functions are invoked and controlled. While it is written primarily for the system's user, a good knowledge of its contents is essential for any individual who intends to work with the system -- be he a user, an analyst, or a programmer.

Table of Contents

SECTION 6. LARSFRIS PROCESSING FUNCTIONS (cont't)

SECTION 6. LARSFRIS PROCESSING FUNCTIONS (con't)

SECTION 6

LARSFRIS PROCESSING FUNCTIONS

SECTION 6

LARSFRIS PROCESSING FUNCTIONS

This section of the User's Manual contains detailed descriptions of the LARSFRIS Processing Functions. The functions are ordered in alphabetical order to aid the reader in locating the description of a particular function. The page numbers for the chapter always begin with the first three letters of the function (which are unique) followed by the page number within the particular function description.

Each function description begins with introductory material describing the major aspects of the function. This is usually followed by a description of the inputs and outputs of the function using an example input deck and figures showing the outputs that are produced by the deck. In some functions a description of the algorithm is also included.

Many of the examples used for the twenty-three Processing Functions together form a single simulated data analysis task which uses data from the same area. The area chosen for this simulated analysis is a portion of segment 210 from the 1971 Corn Blight Experiment, shown in Figure 1. The upper and lower limits of the area is marked on the figure by arrows. The area is bounded on its four corners by field C6 in the upper left hand corner, field el in the upper right hand corner, field L3 in the lower left hand corner, and field h4 in the lower right hand corner.

BIPLOT

The BIPLOT function is a capability that allows plotting in a two-dimensional feature space, 1.) the means, 2.) the ellipses of concentration, and 3.) the classifications of feature space for each class in a LARSFRIS statistics file. It also provides additional information about the statistical distribution of each class, such as its generalized variance, ellipsoid volume, shape factor, and cut-off probability.

Input to the BIPLOT function consists of:

1. Control cards to select the processing and output options. (See control card listing).
2. A LARSFRIS statistics file, which can be expected from either cards if the CARDS READSTATS option is specified, or from the user's D-disk if the default option is selected.

The BIPLOT function allows the user to specify the plotting scales for both the abscissa and ordinate axes by specifying the desired origins and interval sizes. The abscissa and ordinate axes are subdivided into 100 equal intervals or units. If the scale is not specified, the function will automatically set the origin to 0.00 (zero) and the interval to 1.00 for both axes. One also can specify a desired origin (X.XX) and interval size (X.XX) for all the requested plots (all possible combinations of channels) by setting the channel number N of the control parameters ORIG (N,X.XX) and UNIT (N, X.XX) equal to zero. It should be remembered that the aspect ratio of the characters in a standard line printer is 10 characters per inch in the horizontal direction (abscissa) and 8 characters per inch in the vertical direction (ordinate), therefore it is recommended that the abscissa scale unit be set eight tenths (0.8) of the ordinate scale unit. The lower number channel will always be plotted as the abscissa and the larger number channel will always be plotted as the ordinate, regardless of the sequence in which they are specified in the PLOT control card. For example, PLOT MEANS (3,2) is equivalent to PLOT MEANS (2,3) in that in both instances channel 2 will be plotted along the abscissa and channel 3 along the ordinate.

Impossible channel combinations, such as requesting a plot of the same channel in both coordinate axes, will be ignored. Also, if one requests a plot of channels that are not contained in the input statistics file, the request will be ignored and an error message will be printed both at the terminal and the printer output: (E0000 PLOT OF CLAS (X,Y) SILLY -- IGNORED).

The BIPLOT function allows the user to request up to 30 different plots.

Class Mean Plots. These two-dimensional feature space plots show the positional relationships among class means for any combination of two channels. It also shows the statistical separability between all pairs of classes through the calculation of the pairwise transformed divergence and plotting a connecting line between the class pairs. Although the class means are plotted in a two-dimensional feature space, the transformed divergence is calculated on the basis of all the channels present in the input statistics file. These connecting lines and/or their absence provide information on five different ranges of transformed divergence values, i.e., if this line is constructed by plotting the symbol (=) between two classes, it indicates that the transformed divergence value between these two classes is somewhere between zero and 1000. Other symbols and corresponding ranges of transformed divergence values are: (+) for the range between 1001 and 1250, (-) for the range between 1251 and 1500, and (:) for the range from 1501 to 1750. The absence of the connecting line between two classes indicates that the transformed divergence value between those two classes is greater than 1750. Figure 1 shows an example of this type of two-dimensional plot.

Ellipse of Concentration Plots. The mean plots as described in the preceding section provide positional information among classes in a two-dimensional feature space, however they do not provide any class-variance information.

In order to visualize the magnitude and direction of the class variances for any two-channel combination, one may plot the ellipse of concentration for each class. The ellipses of concentration plotted in the two-dimensional feature space are in fact the projections of the n-dimensional ellipsoids of concentration onto the requested two-dimensional feature space. The "n" refers to the total number of channels present in the input statistics file.

Two-dimensional plot

```
EXECUTION BEGINS...

     ERIS                     LABORATORY FOR APPLICATIONS OF REMOTE SENSING          SEPT 22,1980
     CAROL                                 PURDUE UNIVERSITY                         01:03:44 PM
                                                                                     LARSYS VERSION 3

   *BIPLOT
   CARDS READSTATS
   PLOT MEANS(1,3),ELLIPSE(1,3),CLASS(1,3)
   DATA
I0000 READ STATISTICS COMPLETED                  (BIPRDR)

     ERIS                     LABORATORY FOR APPLICATIONS OF REMOTE SENSING          SEPT 22,1980
     CAROL                                 PURDUE UNIVERSITY                         01:03:44 PM
                                                                                     LARSYS VERSION 3
```

FEATURE INFORMATION

CHANNEL 1,	FROM 0.500	TO 0.600	MICRONS,	CODE 1	ORIG = 0.0	UNIT = 1.00
CHANNEL 2,	FROM 0.600	TO 0.700	MICRONS,	CODE 1	ORIG = 0.0	UNIT = 1.00
CHANNEL 3,	FROM 0.700	TO 0.800	MICRONS,	CODE 1	ORIG = 0.0	UNIT = 1.00
CHANNEL 4,	FROM 0.800	TO 1.100	MICRONS,	CODE 1	ORIG = 0.0	UNIT = 1.00

```
     ERIS                     LABORATORY FOR APPLICATIONS OF REMOTE SENSING          SEPT 22,1980
     CAROL                                 PURDUE UNIVERSITY                         01:03:44 PM
                                                                                     LARSYS VERSION 3
```

CLASS INFORMATION

SYMBOL	CLASS NAME	GENERALIZED VARIANCE	ELLIPSOID VOLUME	SHAPE FACTOR	CUT-OFF PROBABILITY
A	BOSQCOCO	0.10593E 02	32.122	0.6356	0.997
B	AGUA	0.77769E-01	2.752	0.7101	0.999
C	ALGODON	0.16322E 01	12.609	0.7109	0.998
D	NUBDENSA	0.17553E 04	413.502	0.0101	0.990
E	ARENA	0.21606E 04	458.760	0.1454	0.990
F	LODOAGUA	0.15255E 03	121.901	0.1652	0.994
G	2	0.94647E 03	303.635	0.1928	0.991
H	11	0.58958E 03	239.646	0.1100	0.992
I	22	0.20383E 03	140.906	0.2085	0.994

Figure 1

CAROL

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

SEPT 22,1980
LARSYS VERSION 3

MEANS PLOT OF CHANNEL 1 VERSUS 3

TRANSFORMED DIVERGENCE CODING
= 0000 TO 1000 - 1201 TO 1500
+ 1001 TO 1250 . 1501 TO 1750

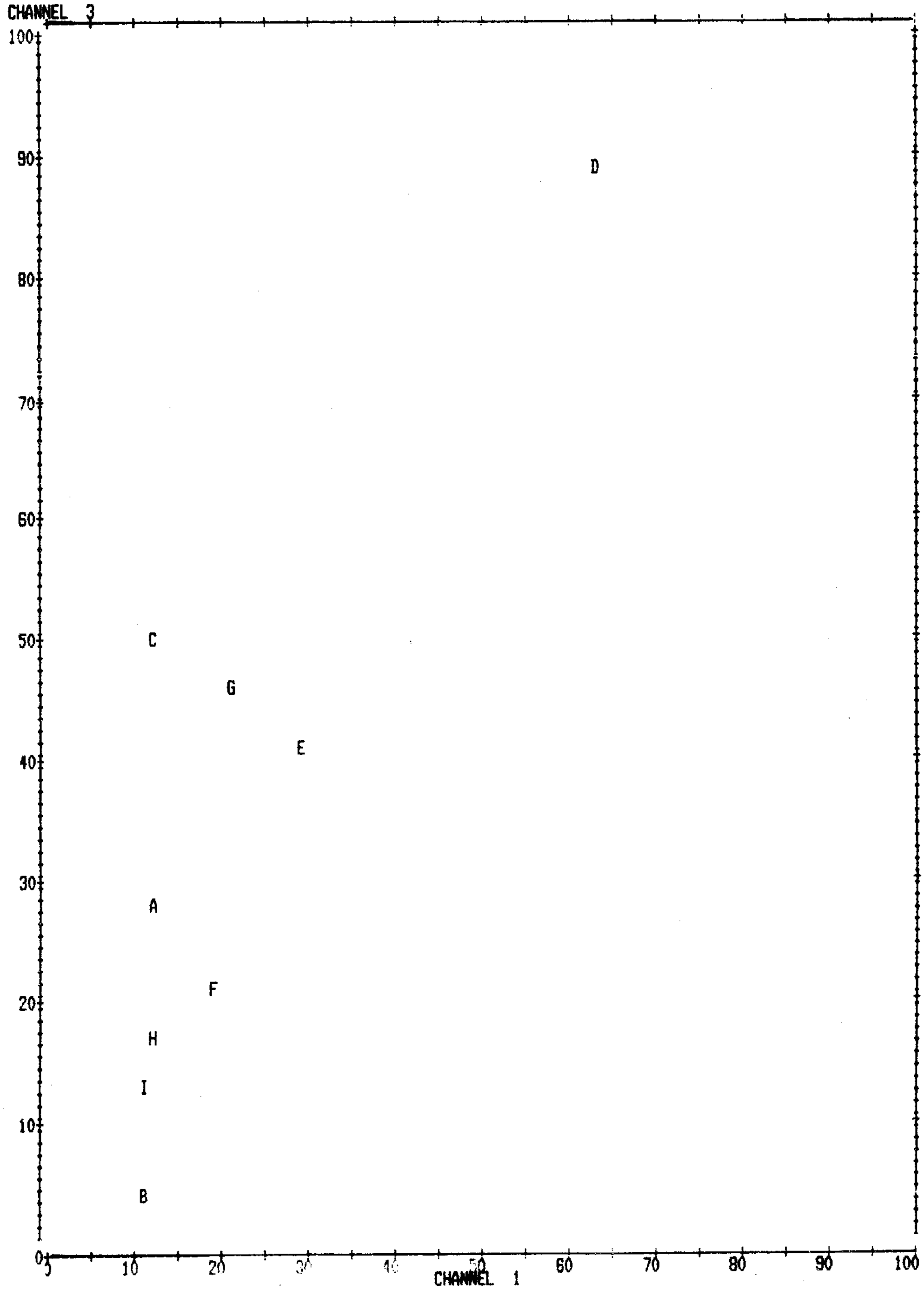


Figure 1 (continued)

The ellipses obtained from the BIPLOT function are the graphic representation of the class distributions at one standard deviation. Figure 2 shows an example of the ellipses of concentration for the same classes shown in Figure 1.

Classification of Feature Space Plots. These plots provide a visual representation of the two-dimensional or bivariate normal density function for each class with three standard deviations. In other words, each point plotted in the feature space represents a "maximum likelihood" classification of the two dimensional feature space, and it provides in a visual manner, valuable information on the shape of the boundaries among classes. Figure 3 shows an example of the classification of feature space for the same classes shown in Figures 1 and 2.

The BIPLOT function also provides additional information on the statistical distribution of each class as illustrated in Figure 4. The definitions regarding this additional information are:

- Generalized Variance which is the value of the determinant of the covariance matrix $|\Sigma|$.
- Ellipsoid Volume which is defined by,

  $$E.V. \triangleq \pi^{\left(\frac{\#\ channels}{2}\right)} \sqrt{\text{Generalized Variance}}$$
- Shape Factor which is defined by,

  $$S.F. \triangleq \left[\frac{\text{Generalized Variance}}{S_1, S_2, S_3 \ldots\ldots\ldots S_\eta}\right]^{\frac{1}{2}}$$

FRIS LABORATORY FOR APPLICATIONS OF REMOTE SENSING PURDUE UNIVERSITY SEPT 22,1980 LARSYS VERSION 3

ELLIPSE PLOT OF CHANNEL 1 VERSUS 3

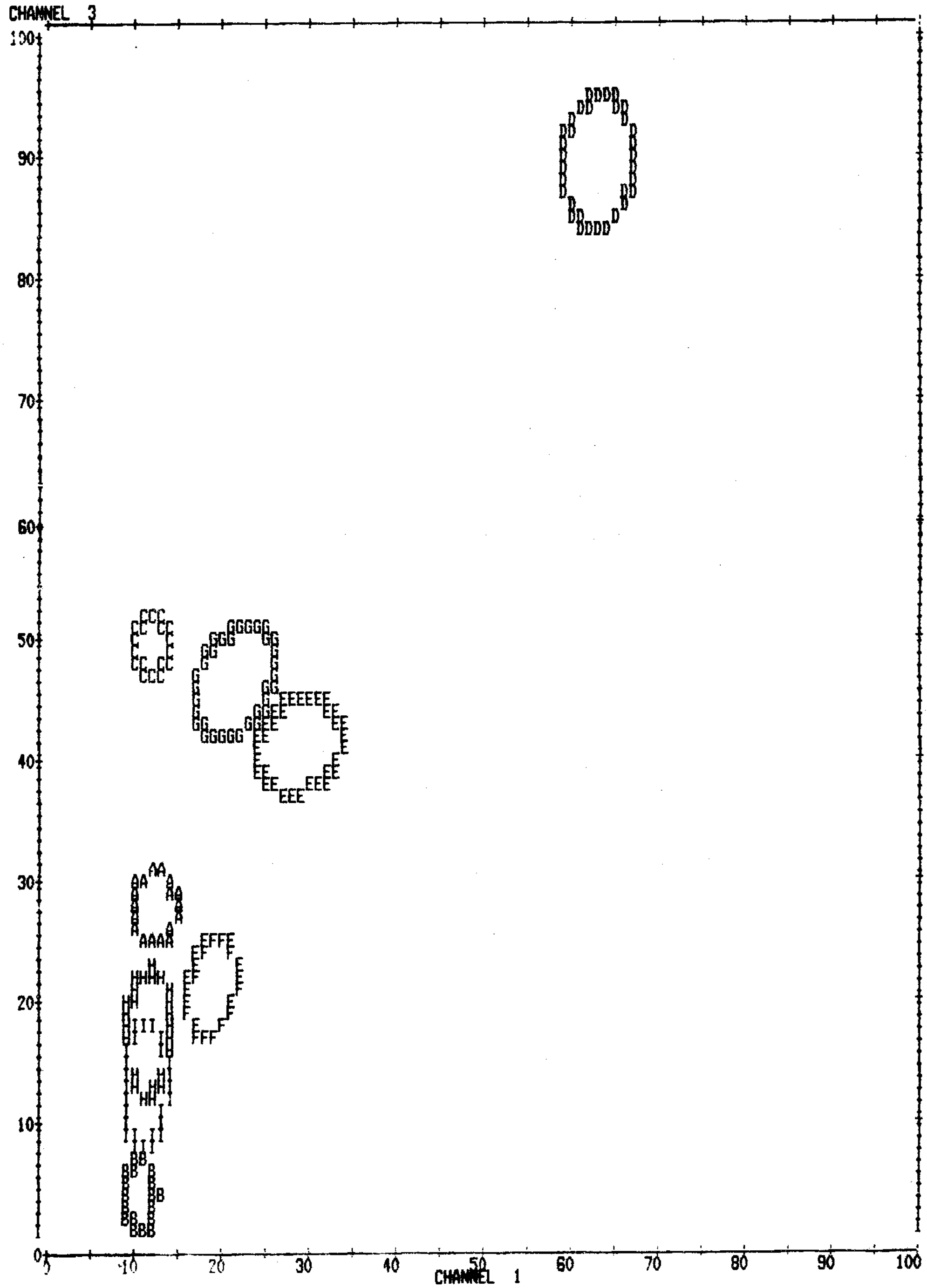
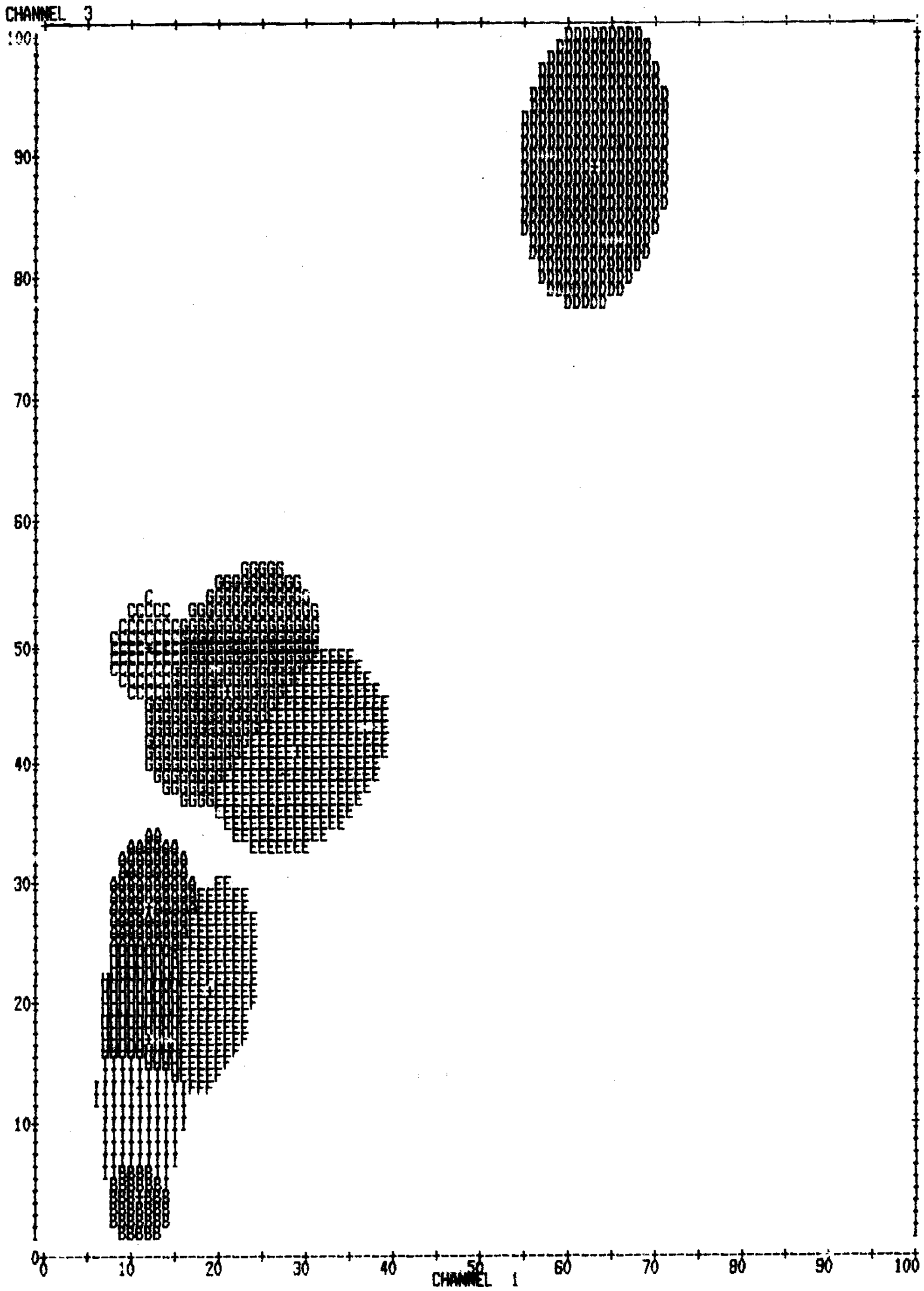


Figure 2

ERIS LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY
SEPT 22,1980
01:03:44 PM
LARSYS VERSION 3

BIP-8

CLASSIFY PLOT OF CHANNEL 1 VERSUS 3



I0103 CPU TIME USED WAS 3.831 SECONDS. (LARSMN)

Figure 3

where:

S_i = variance of a class in channel i

η = number of channels.

- Cut-off Probability

Principal Inputs

The required input to the BIPLOT function consits of a LARSFRIS statistics file and the control cards that select the function and specify the output options. Figure 4 shows a sample BIPLOT input deck.

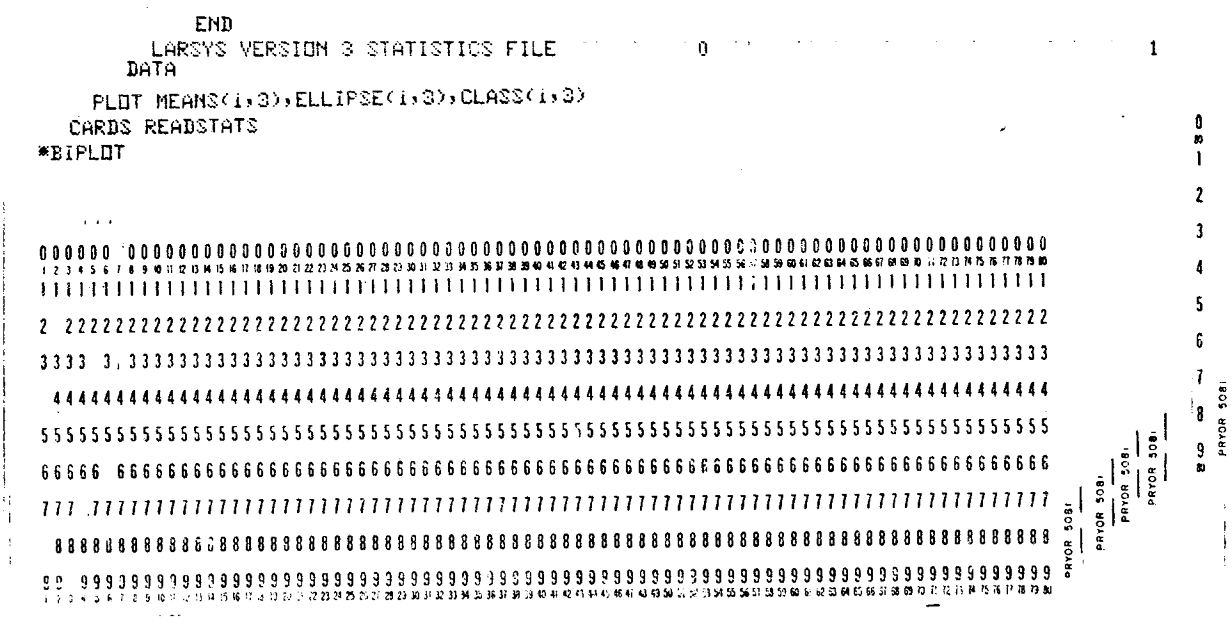


Figure 4

CHANNELTRANSFORM

The Channeltransform function copies a run from a Multispectral Image Storage Tape to another tape. An exact copy can be made or the user can specify the channels to be included in the new copy. Deletion of unneeded channels can save space and time during future processing of the data. Also, new channels can be created as algebraic combinations of existing channels.

For example, analysis of agricultural data offers uses "Brightness" and "Green," which are linear combinations of the four Landsat bands derived using modified principal components techniques. The equations for Landsat 2 data are

$$\begin{bmatrix} \text{BRIGHTNESS} \\ \text{GREEN} \end{bmatrix} = \begin{bmatrix} .332 & .603 & .676 & .263 \\ -.283 & -.660 & .577 & .388 \end{bmatrix} \begin{bmatrix} \text{Band 4} \\ \text{Band 5} \\ \text{Band 6} \\ \text{Band 7} \end{bmatrix} + \begin{bmatrix} 32 \\ 32 \\ 32 \\ 32 \end{bmatrix}$$

Starting from a standard four channel Landsat run, a copy can be created with two new channels for BRIGHTNESS and GREEN by using the following control cards:

```
*CHANNELTRANSFORM
FROM RUN(XXXXXXXX)
TO TAPE(XXX),FILE(XXX)
CHANNELS 1,2,3,4
NEWCHANNELS 6
DATA
C1=C1
C2=C2
C3=C3
C4=C4
C5=.332*C1+.603*C2+.676*C3+.263*C4+32
C6=-.283*C1-.660*C2+.577*C3+.388*C4+32
END
```

Note that if the user had wanted a new copy with only the values for Brightness and Green, there would be a problem since the equation

```
C1=.332*C1+.603*C2+...
```

would redefine the value of C1 immediately, and the new value would then be used in the equation

```
C2=-.283*C1-.660*C2+...
```

which is not what was wanted.

However, the following control cards will produce the desired data:

```
*CHANNELTRANSFORM
FROM RUN(XXXXXXXX)
TO TAPE(XXX),FILE(XXX)
CHANNELS 1,2,3,4
NEWCHANNELS 2
DATA
C5=C1
C1=.332*C1+.603*C2+.676*C3+.263*C4+32
C2=-.283*C5-.660*C2+.577*C3+.388*C4+32
END
```

Since the user specified NEWCHANNELS 2, only the values for C1 and C2 will be on the new run. C5 is used as a "dummy" channel to preserve the original value of C1 for use during the calculation of "Green" -- the new C2.

The expression to the right of the equals sign may include any or all of the FORTRAN operators

+	addition
-	subtraction
*	multiplication
/	division
**	exponentiation

Parentheses may also be used to specify the order in which operations are to be performed.

The Channeltransform function can also be used to copy a run to another tape without changes. It can be used this way to create a private copy of a run from the system copy. In all cases where a user expects to use the data from a run repeatedly, he should use this function to create his own copy of the run. There are a number of advantages to doing so.

- Easy access to the data. If he does not have his own copy of the run, he may unexpectedly have to wait for access to the system copy while another user completes his use of the run (or of another run that happens to reside on the same tape reel).

- Less impact on system resources. If a user must wait for a particular tape while in the middle of a run, he will usually tie up one or more tape drives and prevent other users from having access to the system.

- Protection from data loss. When a single run is used repeatedly, the recording surface of the tape may become worn. This can cause difficulty in reading the data and, in some cases, permanent loss of the data.

- Faster access to data. If the run is one of the later runs on a multi-run tape, the user will get much faster access to the data by placing it in the first file of his own tape. This eliminates the need to read down the tape to find the run.

Input

CHANNELTRANSFORM will copy just one run in each execution of the function. Input to the function is a Multispectral Image Storage Tape and a deck of control cards. Besides the Function Selector Card and the 'END' card, there is a 'FROM' card that specifies the run number of the run to be copied, a 'TO' card that identifies the tape and file numbers where the run is to be copied to. If channel selection or transformation is to be performed, there should be a 'CHANNELS' card to specify channels from which data is to be used; a 'NEWCHANNELS' card to specify the number of channels to be generated, and a data deck with algebraic equations defining the new channels. Note that only a run number is supplied on the 'FROM' card. The function uses either the System Runtable or a user Runtable to determine the tape and file number of the run to be copied. Tapes that are not cataloged in the system catalog may be copied by first using IDPRINT to determine the pertinent tape and file information and then including this information in user Runtable cards preceding the function.

Figure 1 shows a sample input deck that will copy three runs to the user's tape from the system copy of the Multispectral Image Storage Tape(s). Runs 71053900, 71053001, and 66000600 are to be copied. Since no user runtable is provided, LARSFRIS will search the System Runtable to find the tape and file numbers. These three runs will be copied to a single tape (tape number 0068) in three successive files, 1 through 3.

Output

Output from the Channeltransform function is a listing of the input deck and a duplicated data run. The copied run has the same run number as the run that was on the input tape but the tape and perhaps the file number will be different on the new copy from that of the system copy. As a result, the run must be defined in a user Runtable in order to be used as input to the LARSYS functions.

Figure 2 shows the input deck listing for the first of the three sample CHANNELTRANSFORM executions. Note that the function interprets the 'FROM' and 'TO' cards and prints an "end of copy" message as well as the listing of the input cards. A similar listing will be produced for each of the other two executions in the sample deck.

Figure 3 shows the input deck listing for a copy that creates new channels for Brightness and Green.

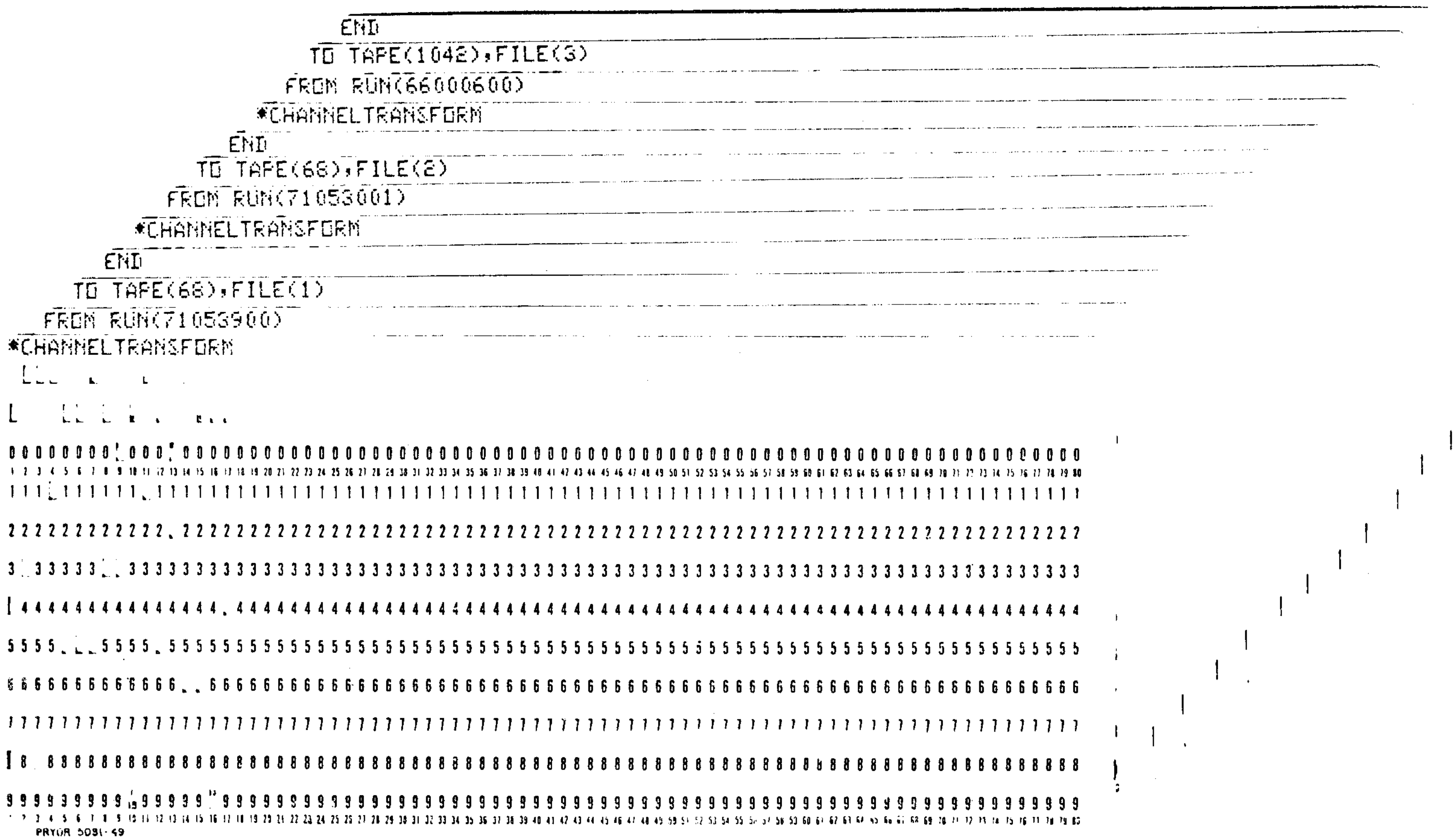


Figure 1 - Sample Input Deck

```
     FRIS                   LABORATORY FOR APPLICATIONS OF REMOTE SENSING          SEPT 22,1980
     CAROL                               PURDUE UNIVERSITY                         01 52 03 PM
                                                                                   LARSYS VERSION 3

   *CHANNELTRANSFORM
   FROM RUN(71053900)
   TC TAPE(68),FILE(1)
   END
INNNN  DATA IS IN LARSYS FORMAT   (GADRUN)
INNNN  DATA IS IN LARSYS FORMAT   (GADRUN)
I0068  RUN 71053900 WILL BE COPIED ONTO TAPE   68, FILE   1.  (CHASUP)
I0119  FILE COPIED AND ENDED PROPERLY.  (CHASUP)

I0103  CPU TIME USED WAS    8.035 SECONDS.     (LARSMN)
```

Figure 2 - Input Deck Listing

```
       FRIS                   LABORATORY FOR APPLICATIONS OF REMOTE SENSING            SEPT 22,1980
       CAROL                               PURDUE UNIVERSITY                           01 39 40 PM
                                                                                        LARSYS VERSION 3

      *CHANNELTRANSFORM
      FROM RUN(78232000)
      TO TAPE(69),FILE(1)
      CHANNELS 1,2,3,4
      NEWCHANNELS 6
      DATA
        CHANNEL CONVERSION CARDS

       C1=C1
       C2=C2
       C3=C3
       C4=C4
       C5=.332*C1 + .603*C2 + .676*C3 + .263*C4 + 32
       C6=-.283*C1 - .660*C2 + .577*C3 + .388*C4 + 32
       END
 INNNN  DATA IS IN UNIVERSAL FORMAT    (GADRUN)
 I0068  RUN 78232000 WILL BE COPIED ONTO TAPE   69, FILE   1.  (CHASUP)
 I0119  FILE COPIED AND ENDED PROPERLY.  (CHASUP)

 I0103  CPU TIME USED WAS   11.864 SECONDS.     (LARSMN)
```

Figure 3 - Input Deck Listing

CLASSIFYPOINTS

The Classifypoints function is an implementation of classification rules which classify multispectral data on a point-by-point basis. The primary (default) rule implemented is a maximum likelihood classification rule which assumes that each training class may be characterized by a multivariate Gaussian (or multivariate normal) distribution, or, equivalently, by the class mean vector and covariance matrix. The program used the class means and covariance matrices (computed by the Statistics or Cluster function) and the data from each point to be classified to calculate the probability that the point belongs to each of the training classes. It then assigns the point to the most probable class and writes the classification (together with a value indicating the associated probability) on an output file for later use. Other classification rules implemented use minimum L1 or L2 (Euclidean) distances as a basis for assigning points to classes.

For L1 distance, the program uses the class means and the data from each point to be classified to compute the distance of the point to each class mean along the channel axes (i.e. the program sums the distance from the point to the class mean in each channel). It then assigns the point to the closest class. The L2 distance also uses the class means and the data points, but computes the Euclidean distance from the point to each class mean and again assigns the point to the closest class (i.e. with the minimum distance to the class mean).

Input to the function is:

- Data from the Multispectral Image Storage Tape(s)
- Control Cards to select the 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 data to be processed and the disposition of results. The principal output is the Classification Results File, which may be placed on tape or disk.

The results file is normally used as input to the Printresults function to produce a variety of printed output for evaluation of the classification. It is also the primary input to the Copyresults, Listresults, and Punchstatistics functions, but it must reside on tape to be used by the latter two of these functions.

CLASSIFYPOINTS produces three standard and two optional printed outputs. The standard outputs are a list of the classes and channels considered, a table of the training fields that were used, and an identification table giving the main characteristics of the classification run. The user may optionally request a statistical summary for the classes considered and a printed map of the classified area. More detailed descriptions of all of these outputs are given later in this subsection.

Inputs

The main inputs to the Classifypoints function are the Multispectral Image Storage Tape and the Statistics File. The function will use the identification information on the Field Description Cards, along with the system (or a user) Runtable, to identify the appropriate input tape and have it mounted. The content and form of this primary LARSFRIS input file is described in Appendix IV.

The Classifypoints function requires that the input statistics be placed in the Statistics File before it begins execution. They can be placed there in one of several ways; by executing a function (STATISTICS, SEPARABILITY, SAMPLECLASSIFY) or by using 'STATDECK USE' command before the use of CLASSIFYPOINTS. Another choice is by including them in his input deck. It should be noted that storage of the Statistics File by any of these methods is temporary. If the user re-initiates LARSFRIS (issues the 'i larsys' command) or logs off the terminal (issues the 'quit' command), the statistics will be erased. To save them, the user must either move them to his permanent disk via the 'STATDECK SAVE' command or have them punched in card form by including the appropriate control card in the Statistics function input deck.

If the user choses to include the statistics in his CLASSIFYPOINTS input deck, he must also include a 'CARDS READSTATS' control card in the deck. He then inserts the statistics card deck in the input deck as the first group of data cards in the deck; i.e., it must come after the Classifypoints control cards and before the Field Description Card group of data cards. Note that the example in Section 3 used this method of storing the Statistics File.

In addition to the principal inputs, the user is expected to provide an input deck which further defines the data to be used and the input/output options. More specifically, he employs control cards to designate the channels to be used in the classification, to specify the desired classification rule, to select training classes to be used and to order them into groups is he wishes (usually based on a priori probabilities if they are available), to select tape or disk for the output Classification Results File and to request one or both of the two output printer options. He may use data cards in the input deck to provide the required statistical input (the Statistics Deck) and must use data cards to define the area or areas to be classified (Field Description Cards). The order of these data decks when both are used must be as indicated (statistics - classification areas).

A sample input deck that illustrates a number of these options is shown in Figure 1. The discussion that follows will give more details about the Classifypoints run that this particular deck defines.

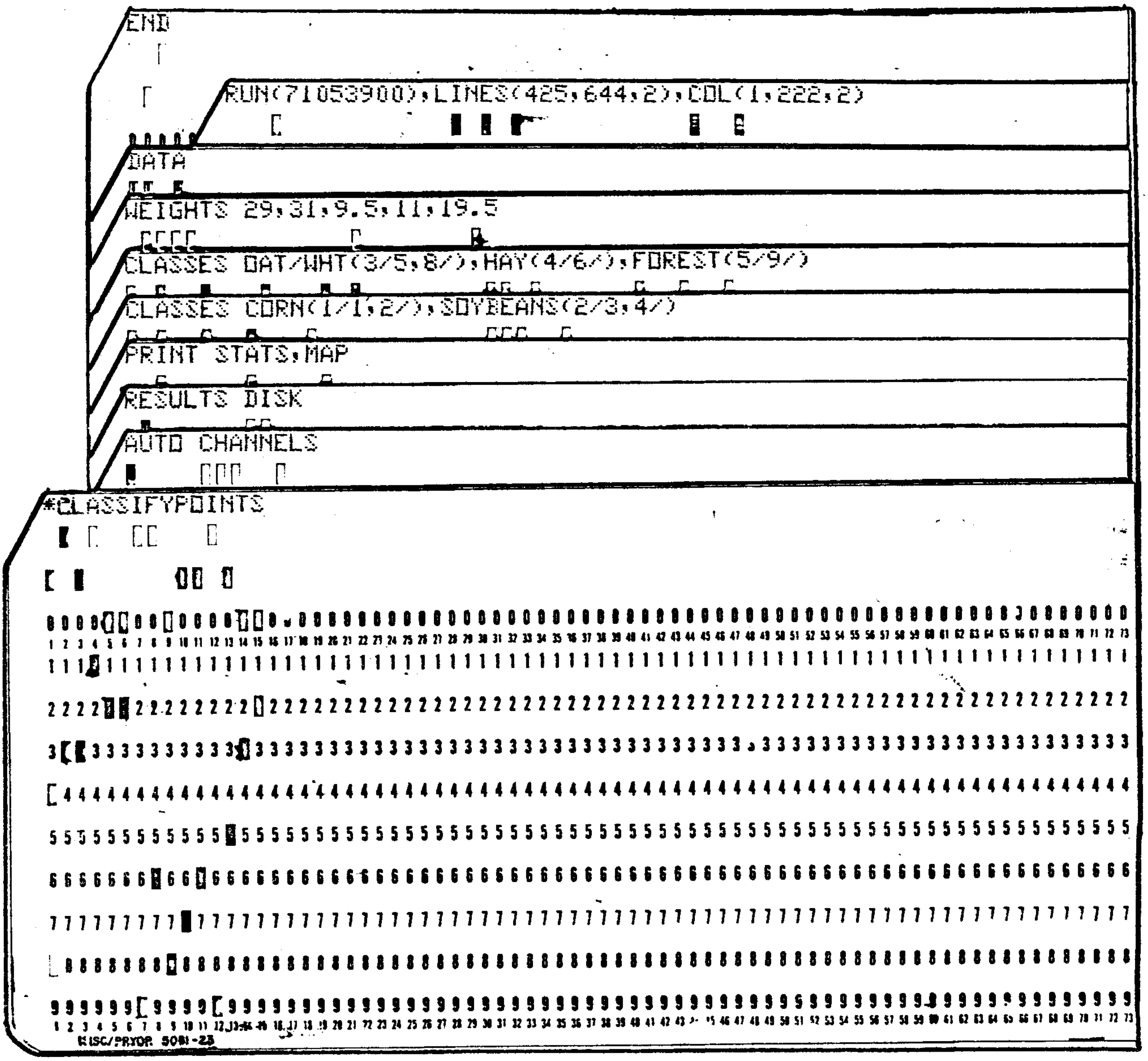


Figure 1. Input Deck for CLASSIFYPOINTS

<u>Specification of Channels:</u> The channels to be used by the Classifypoints function must be specified by including one of two control cards.

- The user may specify channels with the CHANNELS control card in the form:

  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. Appendix IV contains information on how this card may also be used to calibrate data from the Multispectral Image Storage Tape.
- If the Separability function was run prior to CLASSIY-POINTS (in the same terminal session), the channels specification is available from SEPARABILITY. In this case, the AUTO CHANNELS control card is used to indicate that the channels information is specified internally. Figure 1 shows the use of an AUTO CHANNELS card. See the Separability function description for information as to which channels are specified.

<u>Optional Selection of Training Classes:</u> The user may select the training classes from the Statistics File that are to be used by CLASSIFYPOINTS, and he may combine training classes into groups. 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

In this case, the class name assigned by the Statistics function to classes 1, 3, and 5 will be retained by CLASSIFY-POINTS.

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

- GRPA (Group A) will be classes 1 and 2.
- GRPB will be classes 4, 6, and 7.
- GRPC will be class 5 only.
- Classes 3 and 8 will be ignored.

The control card format to specify this option will be:

CLASSES GRPA(1/1,2/),GRPB(2/4,6,7/),GRPC(3/5/)

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

The example in Figure 1 shows the use of the CLASSES control card to group classes. The nine classes identified in the input Statistics File (generated by the second example in the Statistics function description) are grouped here into five classes - CORN, SOYBEANS, OAT/WHT, HAY and FOREST. Note that one of the Statistics File classes (Class 7: "SA") has been dropped from the classification simply by not including it on a CLASSES card.

Optional Specification of Class Weights:
The user may assign a weight to each class or class group, if training classes were grouped, to be used by CLASSIFYPOINTS. The weights are specified by use of the WEIGHTS card. The control card format will be:

WEIGHTS 0.29, 0.31, 0.095, 0.11, 0.195

The sequence of weights on the card corresponds to the group sequence. That is, the first number is the weight for group 1, the second for group 2, and so on. The weights may be either integer or real.

Specification of Areas to be Classified: The user must provide 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.

These Field Description Cards identify the specific portion of data from the Multispectral Image Storage Tape(s) that is to be classified. The information is used by the processor to request the appropriate tapes and access the specified runs.

Outputs

Classification Results File: The principal output of the Classifypoints function is the Classification Results File, which is, in turn, the primary input to four other LARSFRIS functions: PRINTRESULTS, COPYRESULTS, LISTRESULTS, and PUNCHSTATISTICS. The file may be output 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 LISTRESULTS or PUNCHSTATISTICS, he must place it on tape or have it copied to tape by the Copyresults function.

The user must specify where this file will be stored by using a RESULTS control card in one of three forms:

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

The first control card is used to put the file on a tape already containing classification result 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 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 before any files can be written. A new tape must always be initialized before it can be used to store classification results. The third control card specifies that the classification results are to be written on the disk.

When the Classification Results File is placed on disk, it is only stored there temporarily. Any of the actions listed below will cause it to be erased from the disk by the system. If the user wishes to save the file he must copy it to tape with the Copyresults function. The actions which cause the file to be erased are:

- Another execution of the Classifypoints function.
- Re-initiating LARSFRIS, 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 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 is the point-by-point results of 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 and a likelihood code for the point. The likelihood code indicates the probability that the point is correctly classified. These point-by-point 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 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 Punchstatistics function.
- Summary information about the classification, the channels and classes that were used, and the class weights if any were assigned. 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 statistics (mean vectors and covariance matrices) for the classes and channels used in the classification.

<u>Standard Printer Output</u>: The Classifypoints function 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 classes and channels used, the channel numbers used, etc. An example of this output is shown in Figure 2.

Several items listed under "CLASSIFYPOINTS SUPERVISOR INFORMATION...." in Figure 2 are of particular interest. The list is always headed by the Classification Study number, the unique identification number for the particular classification. The next two lines indicate the number of classes defined and the number of channels used in the classification. The number of fields used to generate the statistics for the classifier are given next. Note in this case 16 fields were used to generate the input Statistics File but only 15 fields are used to derive the statistics for this classification due to the deletion of one class.

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. In the example, however, the user has chosen to use the

```
DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING         AUG  30,1974
PAUL SPENCER                        PURDUE UNIVERSITY                           2 52 00 PM
                                                                              LARSYS VERSION 3

    *CLASSIFYPOINTS
    AUTO CHANNELS
    RESULTS DISK
    PRINT STATS,MAP
    CLASSES CORN(1/1,2/),SOYBEANS(2/3,4/)
    CLASSES OAT/WHT(3/5,8/),HAY(4/6/),FOREST(5/9/)
    WEIGHTS 29,31,9.5,11,19.5
    DATA

YOU HAVE SELECTED THE FOLLOWING CLASSIFY OPTIONS

    PRINT MULTISPECTRAL STATISTICS
    PRINT A CLASSIFICATION MAP
    USE DISK FOR CLASSIFICATION RESULTS

CLASSIFYPOINTS SUPERVISOR INFORMATION....

    CLASSIFICATION STUDY.... 424253584
    NO. OF CLASSES...............    5
    NO. OF CHANNELS..............    3
    NO. OF FIELDS................   15
    CHANNELS SELECTED ARE........    2,   9, 12,
```

Figure 2: Control Card Summary for CLASSIFYPOINTS.

'AUTO CHANNELS' card. This will cause CLASSIFYPOINTS to automatically use best channel combination that was defined by the Separability function. In the example this was channels 2, 8 and 12.

There are three other standard printed outputs. They are:

1. A Classes and Channels Table. This shows the symbol that will be used on the map, the class name for each of the training classes (as defined in the Classify-points input deck), the weight assigned to each class if the user specified weights, and the channel number, spectral band, and calibration code for each channel (taken from the Statistics File). A sample is shown in Figure 3.
2. A Listing of Saved Training Fields. This shows, for each training field, the run number, field designation, line and column coordinates, field type, other information (supplied on the Field Description Card), class name assigned in the STATISTICS run, and class name assigned in this CLASSIFYPOINTS run. A sample of this table is shown in Figure 4.
3. A Classification Run Identification Table. This table shows the run information obtained from the input tape ID record, the spectral band and calibra-tion code for each channel, and the coordinates for

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
2 54 55 PM
LARSYS VERSION 3

CLASSIFICATION STUDY 424253584 ON DISK

TRAINING CLASSES

SYMBOL	CLASS	WEIGHT
1	CORN	0.290
2	SOYBEANS	0.310
3	OAT/WHT	0.095
4	HAY	0.110
5	FOREST	0.195

CHANNELS FROM STATISTICS

CHANNEL NO.	SPECTRAL BAND		CAL. CODE
2	0.48	0.51	1
9	1.00	1.40	1
12	9.30	11.70	1

Figure 3: Classes and Channels Table

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 28,1974
8 19 12 AM
LARSYS VERSION 3

CLASSIFICATION STUDY 417929788 ON DISK

SAVED TRAINING FIELDS

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION	STAT. CLASS	POOLED CLASS
1	71053900	L-6	610	628	1	35	52	1	CORN		CORN1	CORN
2	71053900	L-7	619	628	1	58	94	1	CORN		CORN1	CORN
3	71053900	D*5	618	641	1	100	132	1	CORN		CORN2	CORN
4	71053900	H*1	598	612	1	138	169	1	CORN		CORN2	CORN
5	71053900	L-5	619	628	1	23	29	1	SOYBEANS		SOYBEAN1	SOYBEANS
6	71053900	L-8	606	614	1	58	94	1	SOYBEANS		SOYBEAN1	SOYBEANS
7	71053900	D*6	592	612	1	101	131	1	SOYBEANS		SOYBEAN2	SOYBEANS
8	71053900	H*2	591	593	1	137	170	1	SOYBEANS		SOYBEAN2	SOYBEANS
9	71053900	F-9	461	473	1	64	98	1	OATS		OATSCUT	OAT/WHT
10	71053900	K-4	541	559	1	35	38	1	OATS		OATSCUT	OAT/WHT
11	71053900	G*2	565	582	1	144	149	1	OATS		OATSCUT	CAT/WHT
12	71053900	R-1	482	492	1	40	56	1	HAY		HAY	HAY
13	71053900	G*3	562	582	1	156	164	1	WHEAT		WHEAT	OAT/WHT
14	71053900	H-6	543	562	1	63	72	1	WOODS		FOREST	FOREST
15	71053900	E*2	523	528	1	147	197	1	WOODS		FOREST	FOREST

Figure 4. Saved Training Fields

the area to be classified. If a map is requested, this table will be printed as a header for the map.

Optional Printer Output: Two 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 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, and a correlation matrix of the channels. Figure 5 shows the sample output for two of the classes (CORN and SOYBEANS) from the example input deck in Figure 1.

2. A Pictorial Map. The map obtained from CLASSIFYPOINTS is an image of the entire classified area, with each point represented by an alphanumeric symbol (a number, character or special symbol). Figure 6 shows the map that was produced by the input deck in Figure 1.

   The symbol that is used to represent each class on the map is recorded on the Classes and Channels

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 28,1974
8 19 14 AM
LARSYS VERSION 3

CLASSIFICATION STUDY 417929788 ON DISK

CLASS....CORN

CHANNEL	2	8	12
SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
MEAN	96.92	107.01	96.25
STD. DEV.	6.09	5.01	7.98

CORRELATION MATRIX

SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
0.48-0.51	1.00		
0.72-0.92	0.42	1.00	
9.30-11.70	0.71	0.23	1.00

CLASS....SOYBEANS

CHANNEL	2	8	12
SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
MEAN	86.95	140.32	86.01
STD. DEV.	4.76	6.13	5.94

CORRELATION MATRIX

SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
0.48-0.51	1.00		
0.72-0.92	-0.18	1.00	
9.30-11.70	-0.21	0.37	1.00

Figure 5. Statistical Summary For Classes Corn and Soybeans

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 28,1974
8 21 00 AM
LARSYS VERSION 3

CLASSIFICATION STUDY 417929788 ON DISK

RUN NUMBER............ 71053900	DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CRN BLT LO FL21C	TIME DATA TAKEN..... 1202 HOURS
DATA TAPE/FILE NUMBER.. 1305/ 2	PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971	GROUND HEADING..... 180 DEGREES

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1	CO = 23.10
CHANNEL 8	SPECTRAL BAND 0.72 TO 0.92 MICROMETERS	CALIBRATION CODE = 1	CO = 24.85
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO = 20.60

AREA CLASSIFIED.....LINES 425- 644 (BY 2)
COLUMNS 1- 222 (BY 2)

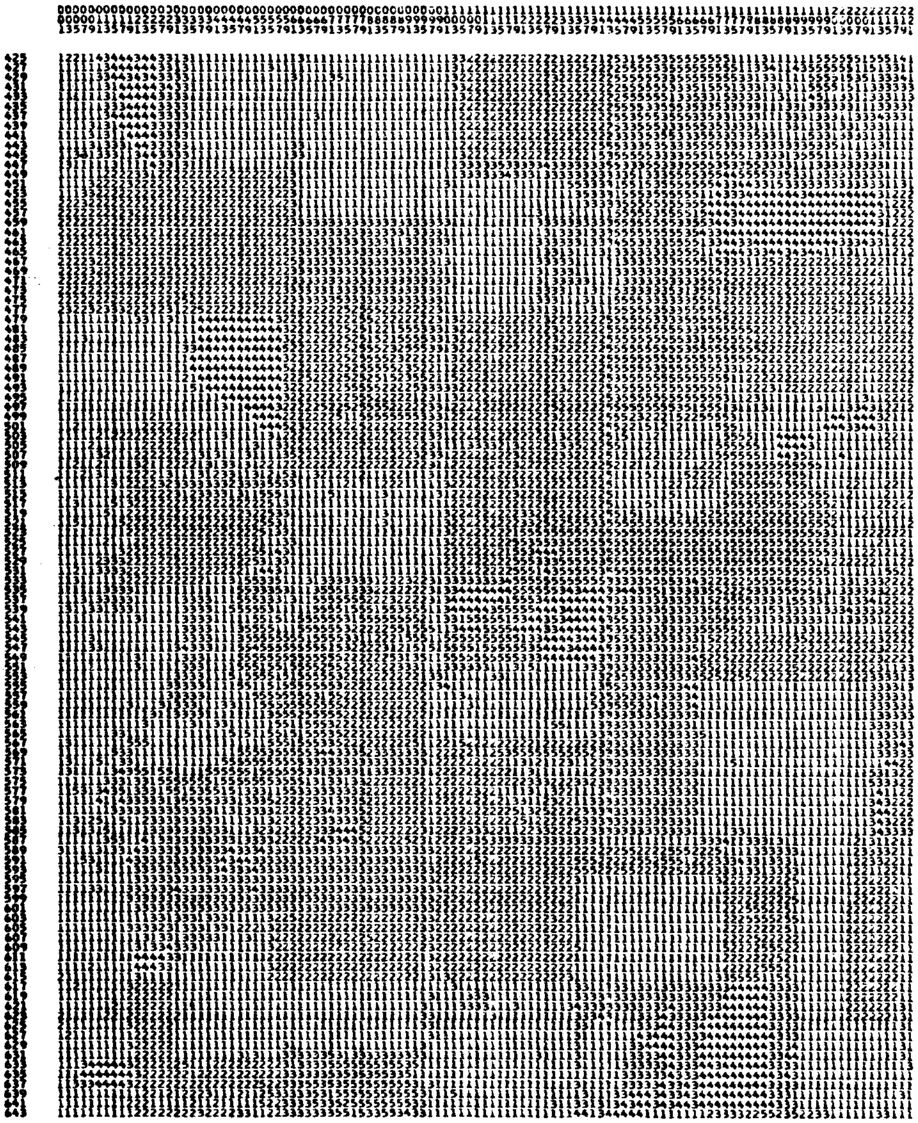


Figure 6. Pictorial Classification Map

Listing (Figure 3). These symbols are assigned by the function to each class (or pooled class) based solely on the class number. They are assigned 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

More comprehensive and flexible mapping capilities are available through the Printresults function. The reader should refer to the description of that function for an example of the map that was produced by PRINTRESULTS from this same classification.

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

Use of the SUSPEND Command

The SUSPEND command has been implemented in CLASSIFYPOINTS to enable the user to interrupt the classification in such a way that he may later restart it at the point at which it was interrupted. For more details on how this command is issued, see the description in Section 4.

The SUSPEND command may be issued in CLASSIFYPOINTS any time after the information message (I0036) appears. This message states that the run being classified has been located. CLASSIFY-POINTS will immediately type a message identifying the last line that was classified before the command was accepted. It will then suspend the Classification function and will save special information that is required for restarting. The manner in which this is done differs depending upon whether the results were being written on tape or on disk.

If the results were being written on tape, the procedure is simple. The function first "closes" the Classification Results File, thereby creating an abbreviated version of the file. It then writes the additional "restart information" as a separate file behind the suspended results file. Finally, it terminates execution in the same way as if the function had normally finished the entire classification.

If the results were being written on disk at the time of suspen-sion, they must be saved on a tape. CLASSIFYPOINTS first types a message asking for a typed RESULTS card to identify the tape to which the file is to be copied. When the user has entered a RESULTS card, the function will attach a tape drive and mount the specified tape reel. The Classification Results File disk is then copied to the tape as a single file, the restart information is written after it as a separate file and the function is terminated as above. In the process of doing this, the results file on the disk is partially destroyed and is no longer usable. The suspended Classification Results File on the tape may, however, be used wherever a normal results file could be used - i.e., as input to the Printresults, Copyresults, Listresults or Punchstatistics functions. This is true regardless of whether the original file was on tape or on disk.

It is important to remember that a suspended Classification Results File actually consists of two files, with the normal results data in the first file and the restart data in the second file. If the user destroys the restart data (for example, by writing another Classification Results File behind the suspended one), the suspended classification is no longer restartable. By the same token, when a suspended classification is restarted, the new classification results will be written on the tape immediately behind the results from the earlier part of the classification. This will destroy anything that has subsequently been written on the tape behind the suspended results file.

To restart a suspended classification, the user submits an input deck consisting of just four control cards; the standard Function Selector Card and END cards, a special RESTART control card and a RESULTS control card. The RESTART card simply identifies the execution of the function as a restart, and the RESULTS card identifies the tape and file numbers of the tape containing the suspended results. The new results must continue to be written on this tape.

CLASSIFYPOINTS automatically mounts the specified tape, restores the required restart information, and resumes the classification at the point of suspension. At the end of the restarted classification the results file will be identical to one created by a normal execution of the function.

A restarted CLASSIFYPOINTS classification has certain unusual characteristics that should be noted.

- Any user Runtable that was in use at the time of suspension will be restored when the classification is restarted. It will replace any Runtable that the user might have specified via RUNTABLE System Initialization cards in the input deck.
- Any special headings that were in effect as a result of the HD1, HD2 or the COMMENT System Initialization Cards in the original input deck will be restored for the restarted classification.

- The date that is placed on the printed output from the restarted classification will be the same as the date on the original output.
- A restarted classification will not leave a Statistics File on the disk, as would a normal execution of CLASSIFYPOINTS. If a Statistics File existed on the disk before the classification was restarted, it will not be used and will remain unchanged.

The Classification Algorithm

The specific implementation of the maximum likelihood classification rule in LARSFRIS is presented in this subsection. One set of discriminant functions g_i (X) (where X is an unknown data vector) given in the references is:

$$g_i\ (X)=\log\ p(\ _i)\frac{n}{2}\ \log\ (2\pi)-\frac{1}{2}\ \log\ |K_i|-\frac{1}{2}\ (X-M_i)^T K_i^{-1}(X-M)_i$$

where M_i and K_i are, respectively, the mean vector and covariance matrix for class ω_i, $p(\omega_i)$ is the a priori probability of class ω_i, and n is the number of features (channels) to be used in the classification.

The decision rule for classification is given as:

Classify X as belonging to class ω_i if

$$q_i\ (X) \geq g_j\ (X) \text{ for all } j \neq i \ *$$

The above decision rule is implemented as follows: If a priori probabilities are assigned by the user, the above discriminant functions, Eq. (1), are used in the decision rule, Eq. (2). If the user does not assign a priori probabilities, the log $p(\omega_i)$ terms are set equal to zero before calculating the discriminant functions. This is equivalent to assigning equal probabilities to all classes used.

As implemented, the a priori probabilities are specified by the user as "weights" (on the WEIGHTS card for CLASSIFYPOINTS). In fact, the class weights card may be determined by the user based on factors other than or in addition to a priori class probabilities. For instance, the relative "costs" of various classification errors may be taken into account. To use this facility to maximum advantage, the user should consult a text on statistical decision theory or statistical pattern recognition (or see references [2] and [3] cited in this section).

---

* In the case of equality in (2), ties may be arbitrarily decided by, say, always deciding $X\varepsilon\omega_i$ if $g_i(X) = g_j(X)$ an and i > j.

At classification time the classes have been defined and the statistics for each class have been calculated.

Therefore the inputs to the classification algorithm are:

- The number of features (channels) used in the classification process (specified by the user in the control cards).
- The number of classes (specified by the Statistics File and the control cards).
- A set of constants, $\log p(\omega_i) - \frac{n}{2} \log (2\pi) - \frac{1}{2} \log K_i$, one for each class i (computed in the initial part of CLASSIFYPOINTS).
- The elements of the inverse covariance matrix (K_i^{-1}) for each class i (computed in the initial part of CLASSIFY-POINTS). Since this matrix is symmetric, only the lower triangular portion is stored and used.
- The statistical mean vectors (M_i), one component for each feature, for each class i (contained in the final class statistics).
- A data vector (X) with one component for each feature taken from a Multispectral Image Storage Tape.

The classification algorithm calculates the quantity:

$$\log p(\omega_i)-\frac{n}{2}\log 2\pi -\frac{1}{2}\log |K_i|-\frac{1}{2}(X-M_i)^T K_i^{-1}(X-M_i) \qquad (3)$$

for each class i in the following manner. Let $Y = (X - M_i)$ and K_{jk} equal a component of the matrix K_i^{-1}. Then expression (3) becomes:

$$\log p(\omega_i)-\frac{n}{2}\log 2\pi -\frac{1}{2}\log |K_i|-\frac{1}{2}\, [Y_1K_{11}Y_1+Y_1K_{12}Y_2 + \ldots$$
$$+Y_1K_{13}Y_3 + Y_2K_{22}Y_2 + Y_2K_{23}Y_3 + \ldots \qquad (4)$$
$$+Y_3K_{31}Y_1 + Y_3K_{32}Y_2 + Y_3K_{33}Y_3 + \ldots]$$

And since $K_{21} = K_{12}$, expression (4) becomes:

$$\log p(\omega_i)-\frac{n}{2}\log 2\pi -\frac{1}{2}\log |K_i|+ [-\frac{1}{2}Y_1K_{11}Y_1$$
$$- Y_1K_{21}Y_1 \quad -\frac{1}{2}Y_2K_{22}Y_2 \qquad (5)$$
$$- Y_3K_{31}Y_1 \quad - Y_3K_{32}Y_2 - \frac{1}{2}Y_3K_{33}Y_3\ldots]$$
$$\vdots$$

This discriminant is calculated for each class i and compared to other discriminant calculations. The largest discriminant value indicates the classified class for the input data vector.

Often there are points in the area classified which in reality do not belong to any of the training classes. The classification rule necessarily assigns these points to one of the training classes, but typically these points yield very small discriminant values. The latter fact can be used to detect them, and the Printresults processor can be instructed to "reject" them. This is the principal reason for saving the discriminant value.

However, experience has shown that the discriminant value can be quantized and coded in such a way as to save space in the Classification Results File. This is accomplished by first converting the discriminant value into the rate of rejection which would cause the Printresults processor to reject or "threshold" the point. This involves the generation of a table of Chi-square values because, under the assumption that the data values are Gaussianly distributed, the discriminant values have a distribution related to the Chi-square distribution.

Specifically, the following table is first constructed:

	Rejection Rate	Chi-square Value
interval = 0.1%	0.1%	C_1
	0.2%	C_2
	.	.
	.	.
	.	.
	.	.
	14.9%	C_{149}
interval = 1.0%	15.0%	C_{150}
	16.0%	C_{150}
	.	.
	.	.
	.	.
	.	.
	98.0%	C_{233}
	99.0%	C_{234}

Then, rather than writing the discriminant value on the results file, the number j is written, where j satisfies:

$$C_{j-1} > 2(\log p(\omega_i) - \frac{n}{2} \log 2\pi - \frac{1}{2}\log|K_i| - d_i(X)) \geq C_j$$

where, $d_i(X)$ is the discriminant value and K_i is the covariance matrix associated with the class into which the point was classified.

When at a later time, it is desired to determine whether the classification should have been "rejected" or "thresholded" at a specified percentage level, this can be done simply by converting the number j back into the associated rejection rate. For example, if the user specifies a threshold of 0.2% (i.e., 0.2 on the THRESHOLD card), then any point for which j is 2 or smaller will be rejected. Note that the smallest threshold which can be meaningfully specified on the THRESHOLD card is 0.1.

L1 Distance Calculation

The L1 distance of a point to a class is computed by summing, for all channels, the distance from the point to the class mean in each channel. In other words, the distance of point X to class j is:

$$\sum_{i=1}^{NC} \left| X_i - M_{ij} \right|$$

Where: NC is the number of channels (features) in the feature space

X_i is the data value of point X in channel i

M_{ij} is the value of the mean for class j in channel i.

The class for which this distance is a minimum is selected as the classified class for data vector X.

L2 Distance Calculation

The L2 distance of a point X to class j is determined by claculating the Euclidean or "straight-line" distance from the point to the mean of the class. The actual quantity computed is:

$$\sum_{i=1}^{NC} \left(X_i - M_{ij} \right)^2$$

Where: NC is the number of channels

X_i is the data value of point X in channel i

M_{ij} is the value of the mean for class j in channel i.

Again, the class for which this distance is a minimum is designated as the output class for data vector X.

References

1. K. S. Fu, D. A. Landgrebe, T. L. Phillips, "Information Processing of Remotely Sensed Agricultural Data," Proceedings of the IEEE, Vol. 57, No. 4, April 1969, pp. 639-653.

2. N. J. Nilsson, Learning Machines, New York: McGraw-Hill, 1965.

3. P. H. Swain, "Pattern Reocgnition: A Basic for Remote Sensing Data Analysis," Information Note 111572. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, January 1973.

CLUSTER

The Cluster function implements an unsupervised classification (clustering) algorithm which classifies individual data points into a predefined number of clusters. The algorithm is based upon the distance relationships between each point and the centers of groups of points (clusters). It requires that the user estimate the number of clusters (or classes) that are to be produced. After arbitrarily assigning a location in the feature space as the initial center of each cluster, the function begins the following two-step iteration:

- It calculates the euclidean distance between each sample and each cluster center and assigns the sample to the cluster with the minimum distance. This causes the samples to be partitioned into groups of samples around each cluster center.

- It determines new cluster centers by calculating the mean of the samples assigned to each cluster. It then proceeds back to the first step to reassign each sample to the closest newly defined cluster center. It continues this operation until the clusters do not change from one iteration to the next, or upon reading a user-defined stopping point.

After completing the above process CLUSTER calculates the separability of each pair of the resulting classes.

CLUSTER permits the user to control this classification scheme in several ways. He may specify the initial number of classes that are to be clustered. He may limit the processing to include only a subset of the samples in the area to be clustered. He may stop the recalculation of cluster centers when the percentage of samples that are not assigned to a new cluster center equals the value he has specified. These controls and their defaults will be described further in the discussion of the sample input deck.

The required input to the function is:

- Data from the Multispectral Image Storage Tape(s).
- An input deck which includes control cards to select the processing and output options and Field Description Cards to identify the area or areas to be clustered.

The content and format of the Multispectral Image Storage Tape are described in Appendix IV with a discussion of how the data on the tape may be calibrated for use in the LARSFRIS functions. The Field Description Cards (see Appendix I) in the input deck may be either of two forms and may define as many areas as desired. The data from all of these areas is consolidated together for the cluster analysis.

The outputs of the Cluster function include six standard printed outputs, one optional printed output, and two optional punched outputs. The printed outputs are:

- A Listing of the Control Cards that were supplied by the user.
- An Input Fields Listing, which lists the run number, and line and column coordinates of the fields to be clustered, as well as the run identification information from the Multispectral Image Storage Tape ID record.
- A Clustering Statistics Listing, which contains processing specifications, spectral bandwidth and calibration information, the number of clusters, and points in each cluster and the cluster means and variances.
- A Cluster Map, which pictorially represents each area that was clustered. Included on the map is field identification information and a summary of the number of points assigned to each cluster.
- Separability Information Listing, which shows the results of the separability calculations for each pair of classes (clusters).
- Results of Cluster Grouping, which show the recommended grouping of classes based on separability quotient values.
- Histograms of data values for each class for all requested channels (optional).

The optional punched outputs are a deck of Field Description Cards, which identify each of the clustered areas, and a statistics deck which may be used in Classifypoints. Only one of these punched outputs may be requested.

Control Card Input

The user must supply an input deck which includes at least the Function Selector Card and a CHANNELS control card followed by a DATA card, Field Description Cards, and an END card. He may also use OPTIONS, PUNCH, PRINT, IDNAME, and SYMBOLS cards to control the processing and the printing of output. The sample input deck that appears in Figure 1 includes all of the control cards that are used by this function except PUNCH STAT or PUNCH STAT, CHAR which may not be requested when PUNCH FIELD is specified.

Data to be Clustered: Two types of cards in the input deck, Field Description Cards and the CHANNELS control card, define the data to be clustered and control the number of samples that can be clustered. The Description Cards (in the data deck) specifies the number of vectors requested to be clustered. The parameters on these cards identify the Multispectral Image Storage Tapes that are to be used and the specific lines and columns (and line and column intervals) from which the data is to be selected.

The CHANNELS control card identifies the channels that are to be used for the clustering. The user should determine beforehand that the channels he specifies exist for the run or runs

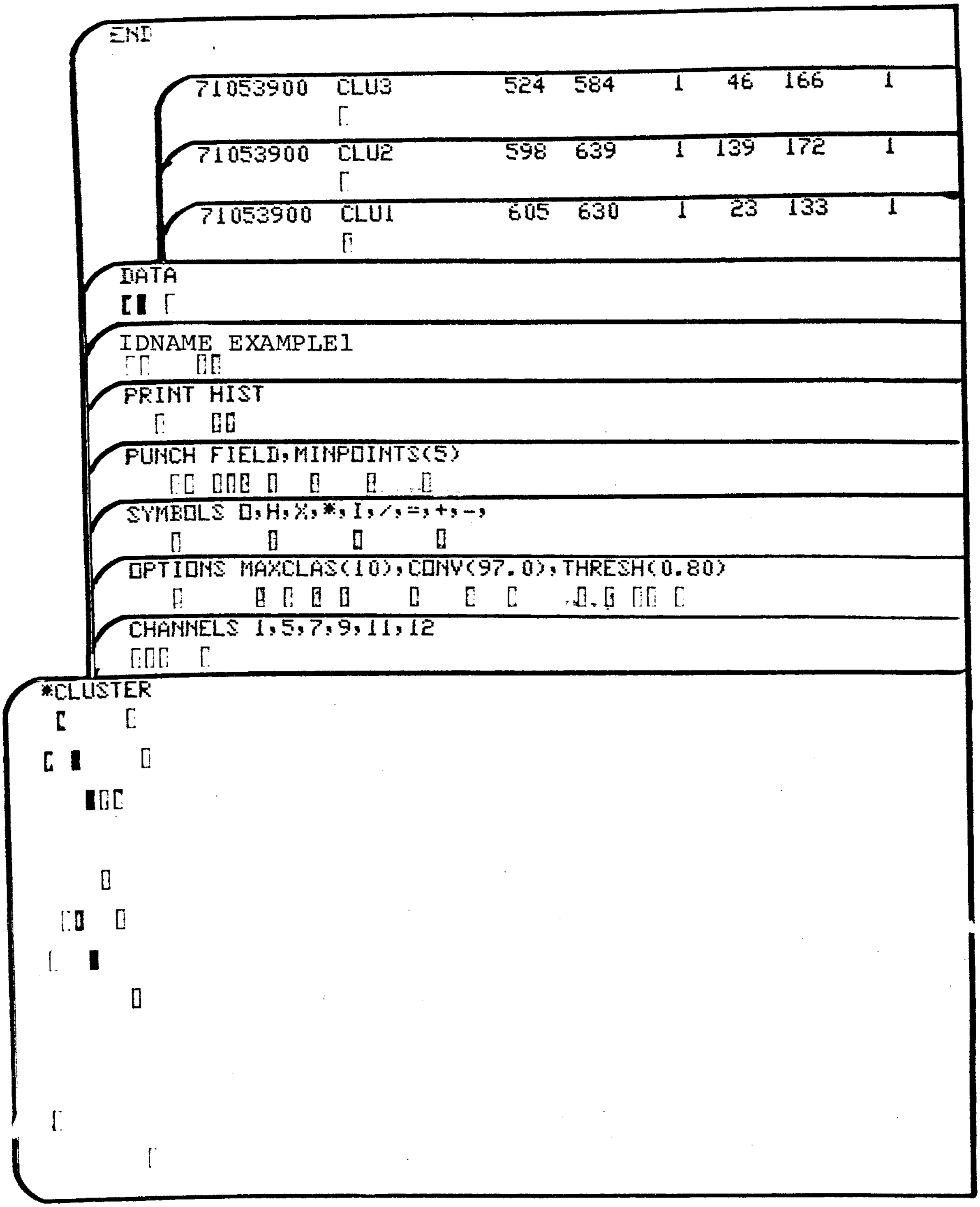


Figure 1. Input Deck for CLUSTER

indicated in his Field Description Cards. The function checks for this, and appropriate messages are printed, but in some cases the function will be terminated if a channel does not exist for one of the requested runs.

The number of vectors that can be clustered is, slightly less than 40,000/n with a maximum of 25,000, where n is the number of channels used. If the maximum storage size is exceeded, a message is printed out and the function reduces the number of vectors to be clustered until a fit is achieved. This is done by increasing the line interval and then the column interval alternately until the requested samples can be fitted into available storage. The sample input deck in Figure 1 includes a CHANNELS control card which specifies that channels 1,5,7,9,11, and 12 are to be used for the clustering calculations.

Control of Processing: The OPTIONS control card is used to control the clustering algorithm. Two of the five parameters in this card function together to control the number of major iterations that are performed. The 'MAXCLAS (n)' parameter establishes the beginning (and maximum) number of clusters that are to be produced by the algorithm, the 'MINCLAS (n)' parameter establishes the minimum number of clusters to be produced. The INTV(n) and CONV(xx.x) parameters can be used to reduce the amount of time required to complete the clustering process. The THRESH(x.xx) parameter can be used to set the threshold value for separability quotients in computing recommended Cluster grouping.

The MAXCLAS(n) parameter specifies the maximum number of classes to be clustered. No more than 40 classes can be requested. The MINCLAS(n) parameter can be used to request clustering iterations for more than 1 specification for a number of classes. For example, a user might want to cluster his data points into 10, 9, 8, 7, 6, and 5 classes. This can be done by setting MAXCLAS equal to 10, and MINCLAS equal to 5. Unless otherwise specified MINCLAS will be set equal to MAXCLAS.

The user may employ the INTV and CONV parameters on the OPTIONS control card to reduce the amount of time required to complete the clustering process. The 'INTV(n)' parameter specifies that every "nth" sample, rather than every sample, be included in determining the clustering centers, however, each and every sample is classified into one of the clusters after the cluster centers are determined. The 'CONV(xx.x)' specifies a relaxed conversion parameter. This parameter is defined in terms of the percentage of samples whose classification is unchanged from the last iteration. The default value used in CLUSTER is 100%; therefore, clustering will proceed until all samples remain unchanged. In the sample input deck, a request was made to cluster until 97% of the samples remained unchanged.

The 'THRESH(x.xx)' parameter specifies the separability quotient value below which Clusters may be grouped together in the Cluster Grouping Table. This value defaults to 0.75. In the sample input deck this value is changed to 0.80.

Control of Output: The user may employ the IDNAME control card to specify an identification number to be placed on his output, a SYMBOLS control card to supply his own set of symbols to be used on the cluster maps, and the PUNCH control card to produce the optional punched output.

The sample input deck uses these control cards to assign an identification name of EXAMPLE1 to all output and to assign ten specific symbols to be used for each of the ten clusters that are to be produced.

The PUNCH control card has four parameters. The 'FIELD' parameter cuases Field Description Cards to be punched that describe the clusters, and the 'MINPOINTS(n)' parameter is used to limit the number of these cards that are produced. Cards are punched only for fields that contain the 'MINPOINTS', the number of consecutive points or more. The term field in this case means a "mini-field" consisting of that portion of a single line assigned to the same class on a CLUSTER map. The system default value for 'MINPOINTS' is four consecutive points. The sample input deck requests that Field Description Cards be punched and specifies the minimum number of points per field at 5.

The 'STAT' parameter on the PUNCH control card causes a hexadecimal (machine-coded) statistics deck to be punched. The 'CHAR' parameter together with the 'STAT' parameter specifies that the deck be punched in character (EBCDIC) representation. This deck is similar in format to the deck produced by the

STATISTICS function except that the training field description cards use dummy fields. This deck may be used directly in the CLASSIFYPOINTS, SEPARABILITY, and SAMPLECLASSIFY functions.

This Statistics File is always calculated and written onto the user's temporary disk. It may be used directly from the disk during the same terminal session. The file may be saved on the user's permanent disk by using the 'statdeck save' command. The file is erased from the temporary disk when the user either logs off the system or re-initiates LARSFRIS.

Since training field cards are not necessary when statistics are obtained directly from CLUSTER, including both PUNCH FIELDS and PUNCH STAT is an invalid request. If both are requested, only statistics will be punched.

The PRINT control card may be used with the 'HIST' parameter to request histograms to be printed for each cluster and every channel. Caution should be used in requesting this option since for a large number of clusters and many channels, this results in a considerable amount of printed output.

Description of Output

The Cluster function always prints a listing showing the control cards and a summary of how it has interpreted the user's input deck. A sample produced by the input deck in Figure 1 is shown in Figure 2.

Besides this preliminary information, there are five additional printed outputs. They are:

Input Fields Listing

This is a listing of the fields in the order supplied by the user, identification information for each different run number, and a listing of the fields again, this time in order processed. A sample is shown in Figure 3.

Clustering Statistics Listing

The number of classes, total number of vectors, clustering interval, and channel information are listed. Then the number of points per cluster and the means and variances of each cluster for each channel are printed out. Figure 4 shows the listing that was produced by the clustering for the sample input deck.

```
   BUIS                         LABORATORY FOR APPLICATIONS OF REMOTE SENSING              OCT  14,1980
   J  BUIS                                  PURDUE UNIVERSITY                              11 33 50 AM
                                                                                           LARSYS VERSION 3


  *CLUSTER
  CHANNELS 1,5,7,9,11,12
  OPTIONS MAXCLAS(10),CONV(97.0),THRESH(0.80)
  SYMBOLS O,H,X,*,I,/,=,+,-,
  PUNCH FIELD,MINPOINTS(5)
  PRINT HIST
  IDNAME EXAMPLE1
  DATA
YOU HAVE SELECTED THE FOLLOWING CLUSTER OPTIONS

  PRINT HISTOGRAMS OF CLUSTER CLASSES
  PUNCH FIELD DESCRIPTION CARD DECK


CLUSTER PROCESSOR INFORMATION....

  IDNAME........EXAMPLE1
  MAXIMUM CLASSES..... 10
  CONVERGENCE......  97.0
  MINIMUM FIELD SIZE..  5
  INTERVAL............  1
```

Figure 2: Control Card Summary for CLUSTER

FIELDS TO BE CLUSTERED
LISTED IN ORDER SUPPLIED

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION
1	71053900		605	630	1	23	133	1		
2	71053900		598	639	1	139	172	1		
3	71053900		524	584	1	46	166	1		

RUNS PROCESSED

INNNN DATA IS IN LARSYS FORMAT (GADRUN)
INNNN DATA IS IN LARSYS FORMAT (GADRUN)

IDNAME.......... EXAMPLE1

RUN NUMBER............ 71053900

FLIGHT LINE... CRN BLT LO FL210

DATA TAPE/FILE NUMBER.. 1005/ 2

REFORMATTING DATE. AUG 16,1971

DATE DATA TAKEN... AUG 13,1971

TIME DATA TAKEN..... 1202 HOURS

PLATFORM ALTITUDE.. 5000 FEET

GROUND HEADING..... 180 DEGREES

FIELDS TO BE CLUSTERED
LISTED IN ORDER PROCESSED

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION
1	71053900		524	584	2	46	166	1		
2	71053900		598	639	2	139	172	1		
3	71053900		605	630	2	23	133	1		

Figure 3. Input Field Listing

FRIS
CAROL

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

SEPT 22,1980
10 06 57 AM
LARSYS VERSION 3

CLUSTERING INFORMATION

NUMBER OF CLUSTERS = 10   CLUSTERING UNIT SIZE = 5908   CLUSTERING INTERVAL = 1

CHANNEL NUMBER 1 SPECTRAL RANGE 0.46 TO 0.49 MICROMETERS CALIBRATION CODE 1

CHANNEL NUMBER 5 SPECTRAL RANGE 0.54 TO 0.60 MICROMETERS CALIBRATION CODE 1

CHANNEL NUMBER 7 SPECTRAL RANGE 0.61 TO 0.70 MICROMETERS CALIBRATION CODE 1

CHANNEL NUMBER 9 SPECTRAL RANGE 1.00 TO 1.40 MICROMETERS CALIBRATION CODE 1

CHANNEL NUMBER 11 SPECTRAL RANGE 2.00 TO 2.60 MICROMETERS CALIBRATION CODE 1

CHANNEL NUMBER 12 SPECTRAL RANGE 9.30 TO 11.70 MICROMETERS CALIBRATION CODE 1

CLUSTER	POINTS	MEANS CH(1)	CH(5)	CH(7)	CH(9)	CH(11)	CH(12)
1	147	211.27	215.05	206.24	108.55	187.95	154.32
2	311	156.73	155.62	140.23	99.52	159.86	166.14
3	507	136.38	138.95	115.60	96.87	133.72	147.86
4	324	144.26	146.30	119.59	107.97	131.93	110.81
5	265	121.32	114.39	93.46	87.54	112.32	144.89
6	939	127.77	131.54	98.90	105.00	117.31	106.89
7	1189	121.36	117.20	83.05	129.10	125.93	85.49
8	1310	126.69	122.79	94.27	99.05	101.93	93.21
9	501	113.53	110.55	80.33	105.41	105.50	92.49
10	415	107.48	98.21	72.63	93.43	90.34	81.15

CLUSTER VARIANCES

	CH(1)	CH(5)	CH(7)	CH(9)	CH(11)	CH(12)
1	600.95	486.27	598.92	64.17	1038.52	393.69
2	215.68	201.37	170.80	62.54	270.40	426.65
3	56.89	63.19	79.92	58.33	69.43	207.00
4	116.80	68.11	105.15	59.29	143.69	105.12
5	30.31	45.61	32.96	90.05	35.38	158.32
6	47.17	42.50	44.64	47.81	63.80	104.08
7	37.49	54.34	26.97	51.87	63.91	35.62
8	31.69	25.13	16.90	23.00	25.13	20.72
9	23.63	43.41	18.66	47.77	51.00	69.94
10	27.47	45.01	23.52	79.87	35.96	43.21

Figure 4. Clustering Statistics Summary

Cluster Map

This classification map is printed for each field that the user identified on Field Description Cards. At the top of the map, field information (run number, number of samples, and line and column coordinates) is printed. This is followed by the pictorial map which uses alphanumeric symbols to represent classes. At the bottom is a table listing the symbols and the number of vectors for each class for the field. Figure 5 shows one of the maps produced by the input deck for the ten-class clustering requested in the sample input deck.

Separability Information Listing

This listing summarizes the class pair separability calculations that were made. Refer to the sample in Figure 6, which shows the information produced by the sample input deck for clustering ten classes. The columns entitled I and J identify the classes involved. D(I,J) is the euclidean distance between the means of classes I and J. D(I) is a measure of the spread of class I in the direction of the mean of class J.

D(J) is the corresponding quantity for class J. D(I) + D(J) is the sum of these two quantities, and QUOT is calculated by the expression: D(I,J)/[D(I) + D(J)]. It is the measure of separability between classes I and J.

Just below the separability listing is the average quotient value which is the average of all the separability quotients.

Results of Cluster Grouping Table

This table gives a suggested grouping of clusters which will minimize the total number of subclasses while ensuring that multimodal class distributions are avoided. Refer to Figure 7 showing the grouping produced by the sample input deck. The threshold value is the threshold specified on the OPTIONS control card. The column entitled GROUP gives the suggested new group numbers. The CLUSTER column gives the original cluster numbers in each new group. The last column gives the number of points in each of the original clusters. These groups of clusters may be used as the classes for classification purposes. (Note that the Statistics File and punched statistics deck generated by CLUSTER contain statistics for the original clusters. No account is taken here for the suggested grouping.)

FRIS
CAROL

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

SEPT 22,1980
10 08 55 AM
LARSYS VERSION 3

FIELD INFORMATION

FIELD
RUN NO. 71053900
OTHER INFORMATION

TYPE
NO. OF SAMPLES 3751

LINES 524- 584 (BY 2)
COLUMNS 46- 156 (BY 1)

NUMBER OF POINTS PER CLUSTER

CLUSTER	1	2	3	4	5	6	7	8	9	10
SYMBOL	O	H	X	*	I	/	=	+	-	
POINTS	51	101	327	160	140	604	702	508	472	376

Figure 5. Cluster Map

FRIS
CAROL

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

SEPT 22,1980
10 09 39 AM
LARSYS VERSION 3

SEPARABILITY INFORMATION

I	J	D(I,J)	D(I)	D(J)	D(I)+D(J)	QUOT
1	2	108.975	92.449	59.219	151.668	0.719
1	3	150.777	93.993	34.511	128.505	1.173
1	4	147.505	89.686	41.875	131.561	1.121
1	5	192.855	91.443	22.959	114.402	1.686
1	6	180.887	90.512	23.122	113.634	1.592
1	7	204.546	78.400	19.215	97.615	2.095
1	8	198.560	88.423	17.996	106.419	1.866
1	9	216.696	89.974	18.760	108.734	1.993
1	10	239.553	86.471	19.932	106.404	2.251
2	3	48.193	56.408	32.678	89.086	0.541
2	4	67.685	50.460	31.602	82.062	0.825
2	5	89.416	58.539	23.360	81.900	1.092
2	6	92.070	51.417	23.096	74.513	1.236
2	7	120.522	37.874	17.290	55.164	2.185
2	8	112.991	53.271	16.228	69.499	1.626
2	9	126.090	49.559	21.739	71.298	1.768
2	10	149.609	51.759	20.010	71.768	2.085
3	4	40.383	40.279	26.693	66.972	0.603
3	5	43.293	29.637	23.969	53.605	0.808
3	6	49.211	34.874	24.299	59.173	0.832
3	7	82.139	29.931	16.892	46.823	1.754
3	8	69.370	30.981	15.155	46.136	1.504
3	9	80.671	30.845	22.945	53.790	1.500
3	10	103.368	28.942	20.870	49.811	2.075
4	5	64.733	27.114	24.394	51.508	1.257
4	6	33.992	41.175	23.382	64.557	0.527
4	7	61.882	37.458	19.023	56.480	1.096
4	8	52.833	38.017	18.093	56.111	0.942
4	9	69.310	40.244	19.415	59.658	1.162
4	10	93.224	37.441	21.901	59.341	1.571
5	6	46.253	31.794	25.147	56.941	0.812
5	7	74.544	28.143	17.029	45.172	1.650
5	8	54.882	32.411	13.245	45.656	1.202
5	9	57.963	35.276	22.605	57.881	1.001
5	10	73.944	33.640	18.146	51.786	1.428
6	7	40.132	25.631	18.990	44.621	0.899
6	8	23.632	23.434	16.076	39.510	0.598
6	9	36.540	22.786	22.048	44.834	0.815
6	10	61.127	22.227	24.288	46.515	1.314
7	8	41.524	24.345	15.385	39.730	1.045
7	9	33.771	23.366	22.905	46.271	0.730
7	10	56.743	25.932	21.455	47.387	1.197
8	9	23.900	16.900	15.892	32.792	0.729
8	10	41.871	19.303	22.661	41.964	0.998
9	10	27.387	26.289	24.775	51.063	0.536

AVERAGE QUOTIENT 1.254

Figure 6. Separability Information Listing

Punched Output: The Cluster processor has two optional punched outputs - field description cards and a statistics deck. Only one of these may be requested in a single run.

Field Description Cards: This output consists of a punched deck of Field Description Cards in the format required for input to the Statistics function. That is, fields (in this case "mini-fields") are segregated according to classes with a CLASS card designating the start of each class. The value of the 'MINIPOINTS' parameter determines the minimum size (number of columns) of the mini-fields, and therefore, how many cards are punched. For example, if the value is 5 (as in the sample deck), the program searches each line of the stored cluster maps for an occurrence of 5 or more consecutive columns within a line (a mini-field) belonging to the same class. A field description card which specifies the line and the columns of the mini-field is punched for each such occurrence.

For example, referring to the Cluster Map in Figure 5, there will be a card punched for class 7 for line 524, columns 46 through 52. A total of 7 cards will be punched for line 524, and 270 cards will be punched for the entire example input deck. In more typical clustering requests, this deck can become quite large (several boxes of cards), and the user should be cautious about requesting it. When the MINCLAS and MAXCLAS options are in effect, a field deck is punched only for the number of classes in the MINCLAS parameter.

Statistics File: The statistics file produced by CLUSTER is identical in format to the file produced by the STATISTICS processor. The training field description cards in the statistics file produced by CLUSTER contain artificial line and column numbers. A detailed description of the statistics file is given in the description of the Statistics function.

The Clustering Algorithm

The manner in which a set of vectors is clustered is outlined in detail below. The procedure is essentially a variation of the well-known ISODATA method of Ball and Hall.[1] The procedure is described below in four steps.

Step 1 - Initialization

Let $x_1, x_2, \ldots, x_N$ be N vectors which are L-dimensional, and assume they are to be grouped into M clusters. Let C_j, μ_j, and σ_j^2 (all L-dimensional vectors) be the "center", mean, and variance, respectively, for the j^{th} cluster. Thus we have:

$$x_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Lj} \end{bmatrix} \quad j = 1, 2, \ldots, M$$

and;

$$C_j = \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{Lj} \end{bmatrix} \quad \mu_j = \begin{bmatrix} \mu_{1j} \\ \mu_{2j} \\ \vdots \\ \mu_{Lj} \end{bmatrix}, \quad \sigma_j^2 = \begin{bmatrix} \sigma_{1j}^2 \\ \sigma_{2j}^2 \\ \vdots \\ \sigma_{Lj}^2 \end{bmatrix}, \quad j = 1, 2, \ldots M$$

To start the algorithm, an initial set of cluster centers must be established. How this is done has arbitrarily been specified as follows:

Compute the mean vector m and variance vector s^2 for the <u>entire</u> <u>set</u> of N measurement vectors. A rectangular parallelepiped, which usually will contain a large percentage of the measurement vectors, has edges oriented parallel to the coordinate axes and given by $m_i \pm s_i$ (s_i is the square root of the variance component s_i^2). The initial cluster centers are chosen to be uniformly spaced along a principal diagonal of this parallelepiped. Accordingly, the j^{th} initial cluster center is given by

$$c_{ij} = m_i + s_i \left[\frac{2(j-1)}{M-1} - 1 \right] \qquad i = 1,2, \ldots, L$$

for j = 1, 2, ..., M.

<u>Step 2 - Mode Assignment</u>

Determine the euclidean distance from each vector to each of the cluster centers. Assign each vector to the cluster with the nearest center.

<u>Step 3 - Mode Migration</u>

If step 2 did not change the cluster assignment of a definite percentage N vectors, go to step 4; otherwise, replace the old cluster centers by the means of the new clusters resulting from step 2, and then go to step 2.

Step 4 - Compute Cluster Separability

The clustering algorithm has terminated and there remains the task of determining the appropriateness of the results produced. As an aid, the program computes and prints

(a) the cluster means (each spectral channel)
(b) the cluster variances (each spectral channel)
(c) a measure of cluster separability based on both means and covariances. Roughly, this measure is the distance between pairs of cluster means, normalized by the "sizes" of the clusters (for details, see the reference by Swain). For the r^{th} and s^{th} clusters it is given by

$$\gamma_{rs} = \frac{D_{rs}}{D_r + D_s}$$

where

$$D_{rs} = \left| \mu_r - \mu_s \right| = \left[\sum_{i=1}^{L} (\mu_{ir} - \mu_{is})^2 \right]^{\frac{1}{2}}$$

and

$$D_\kappa = \left[\frac{(L + 2) \left| \mu_r - \mu_s \right|^2}{tr\{\Sigma_\kappa^{-1} (\mu_r - \mu_s) (\mu_r - \mu_s)^t\}} \right]^{\frac{1}{2}}$$

$$= \frac{(L + 2)^{\frac{1}{2}} \left[\sum_{i=1}^{L} (\mu_{ir} - \mu_{ir} - \mu_{is})^2 \right]^{\frac{1}{2}}}{\left[\sum_{i=1}^{L} \sum_{j-1}^{L} (\sigma_\kappa^{-1})_{ij} (\mu_{ir} - \mu_{is}) (\mu_{jr} - \mu_{js}) \right]^{\frac{1}{2}}}$$

where $tr\{A\}$ = trace of matrix A

Σ_κ^{-1} = inverse of the covariance matrix for cluster κ

$(\sigma_\kappa^{-1})_{ij}$ = i,j element of Σ_κ^{-1}

μ_κ = mean vector for cluster κ

for $\kappa = r,s$. If γ_{rs} is greater than 1 the clusters are "well separated". For γ less than 0.75, the clusters can probably be merged without resulting in a multimodal class density. The situation is unclear for γ between 0.75 and 1.0.

Step 5 - Compute Suggested Cluster Grouping

The final calculation made by the CLUSTER function is the computation of optimal grouping of cluster classes to minimize the total number of classes while ensuring that multimodal class distributions are avoided. This grouping of clusters is printed only as an aid to the user. The algorithm uses the quotient values from the Separability Table to compute the optimal grouping. The quotient values are denoted by D_{ij}, i, j=1, ..., ND. The original cluster groups are denoted by C_i, i=1, ..., NG, $NG \leq ND$. The algorithm is as follows:

(1) Assign each cluster to its own group, $C_1, \ldots, C_n$.
(2) Order the D_{ij}'s from smallest to largest using the ORDR array and work through the list beginning with the smallest as follows.
(3) If D_{xy} > THRESH, stop. (Merging is complete.)
(4) If $C_x = C_y$ (i.e., belong to the same group) get next value of D_{ij} and go to step 3.
(5) Compute the average distance between C_x and each other group C_u for which $D_{ab} \leq$ THRESH for all a in C_x and all b in C_u. Similarly, compute the average distance between C_y and each other group C_u. If the distance between C_x and C_y is less than all of the other intergroup distances, assign C_x and C_y to the same group.
(6) Select the next D_{ij} and return to step 3.

References

1. Ball, G.H. and D.J. Hall, "ISODATA, A Novel Method of Data Analysis and Pattern Classification," Stanford Research Institute, Melno Park, California, 1965, pp. 1-16

2. Swain, P.H. "Pattern Recognition: A Basis for Remote Sensing Analysis", LARS Information Note 111572, Purdue University, 1972.

COLUMNGRAPH

The Columngraph function graphs data values for requested columns from the Multispectral Image Storage Tape. COLUMNGRAPH and LINEGRAPH (which graphs lines) both graph the data values as a function of spatial position. The graphs also contain standard header and indentification information. The use of these two functions will aid in data quality analysis studies.

Input

The required inputs are the Multispectral Image Storage Tape and control cards. In addition to the standard Function Selector Card and END control card, both a PRINT and a CHANNELS card is required. A third control card, the SCALE card is optional.

The PRINT control card specifies the Multispectral Image Storage Tape run, and the specific subset of data, to be used. In addition, printing of the roll parameter for each line may be requested. A description of the data values, calibration of data, and the roll parameter is included in Appendix IV.

Two mutually exclusive subsets of data may be defined on the 'PRINT' card for plotting. Either the columns of data values or the calibration values for the lines may be requested. Graphs of any or all three of the calibration values may be requested, and one graph will be provided for each. If both columns and calibration values are requested, the function will ignore the calibration request.

The CHANNELS control card specifies the data channels that are to be displayed on the graph. Each channel is represented on the graph by an alphanumeric symbol assigned by the function. The symbol that is assigned to each channel is shown in the header for the graph.

The optional SCALE card controls the "low value" (or beginning point), as well as the increment, for the magnitude axis. The axis normally has a low value of 32.0 and each succeeding point on the axis represents an increment of 2.0. The user may specify other values for these by using the 'XLOW' and 'BINSIZ' parameters on this card.

The sample input deck in Figure 1 requests that column 115 of run 71053900 be graphed for lines 532 through 644 with a line interval of 1. Additional columns could have been requested by including a starting and stopping column and a column interval on the PRINT card. The CHANNELS card in the example requests that channels 7 and 9 be graphed. The omission of a SCALE card will cause the low value and the interval of the magnitude axis to assume the default values of 32.0 and 2.0, respectively.

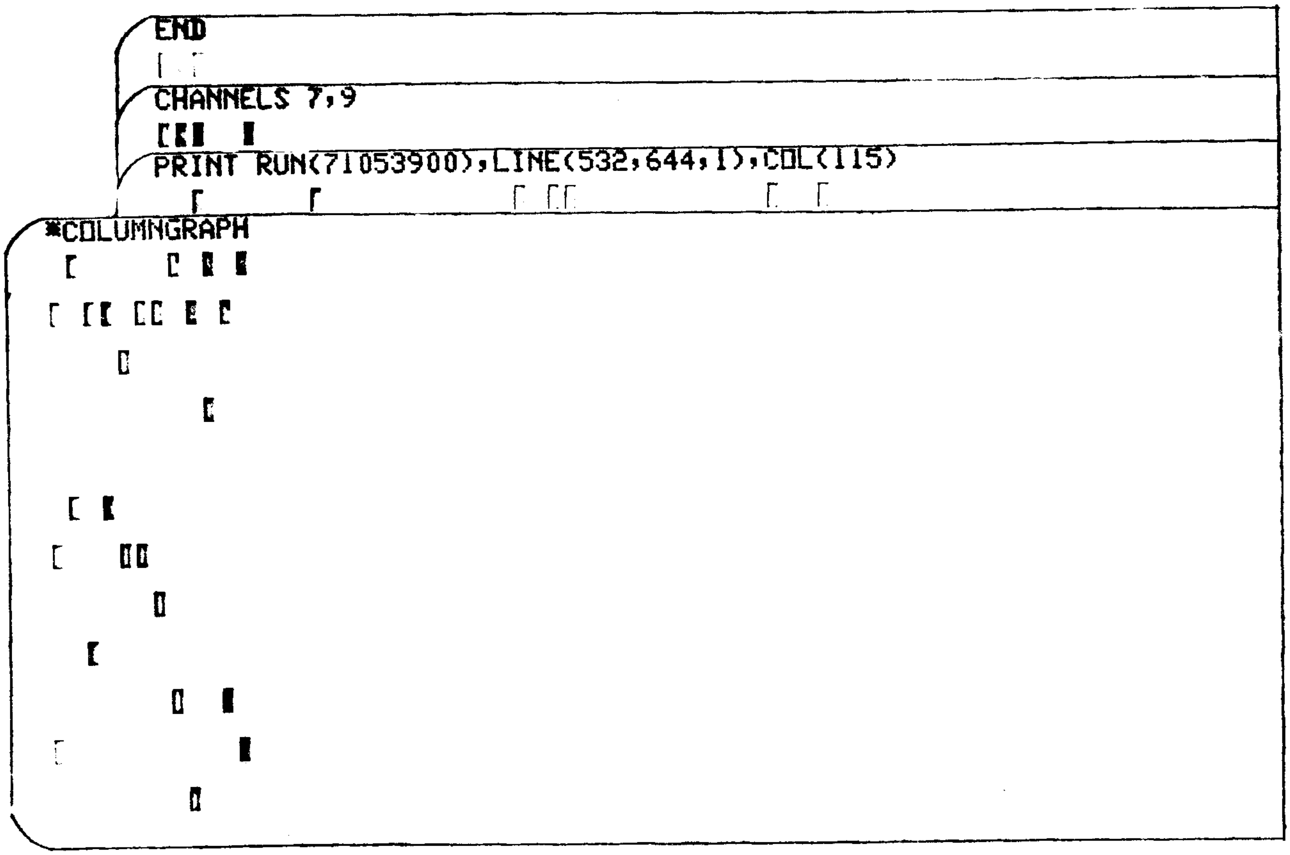


Figure 1. Input Deck for COLUMNGRAPH.

Output

The output of the Columngraph function will be a control card summary and a graph for each column or each calibration value specified. The outputs from the example input deck in Figure 1 are shown in Figures 2 and 3.

The magnitude of each data value for each channel will normally be shown on the graph by its assigned alphanumeric character. However, if two or more channels have the same value, only the highest numbered channel will be printed. This is illustrated in Figure 3 by the plot for line 620 which shows only the symbol for channel 9.

```
DEMO                LABORATORY FOR APPLICATIONS OF REMOTE SENSING        MAR  7,1973
PHILLIPS                         PURDUE UNIVERSITY                       11 14 08 AM
                                                                         LARSYS VERSION 3

*COLUMNGRAPH
PRINT RUN(71053900),LINE(532,644,1),COL(115)
CHANNELS 7,9
END

                                                                               COL-2
```

Figure 2. Control Card Summary for COLUMNGRAPH

DEMO
PHIL

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 27,1974
3 25 04 PM
LARSYS VERSION 3

****** GRAPH OF COLUMN 115 ******

RUN NUMBER............. 71053900 | DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CKN BLT LO FL21C | TIME DATA TAKEN..... 1202 HOURS
DATA TAPE/FILE NUMBER.. 1342/ 1 | PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971 | GROUND HEADING..... 160 DEGREES

CHANNEL 7 SPECTRAL BAND 0.61 TO 0.70 MICROMETERS DISPLAYED AS.. 7 CALCODE = 1 CO = 28.00
CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS DISPLAYED AS.. 9 CALCODE = 1 CO = 22.90

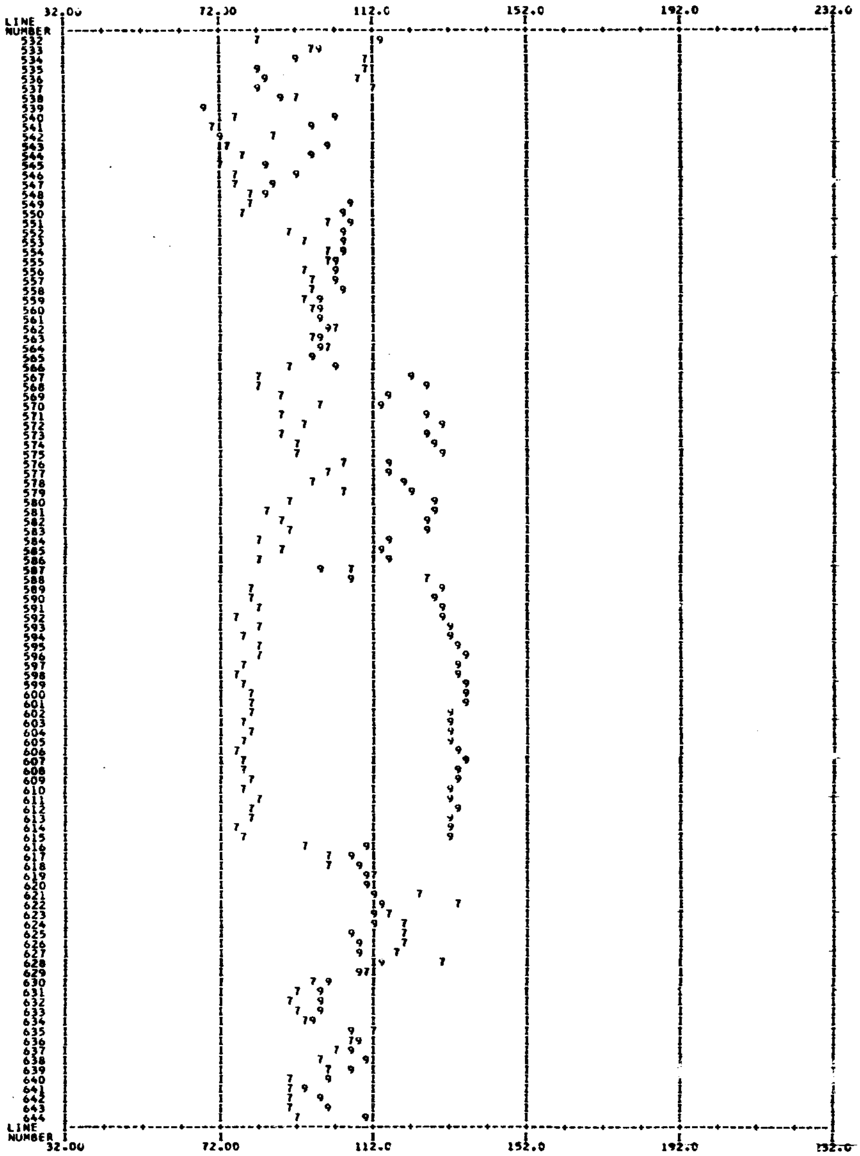


Figure 3. Graph of Column 115

for Run 71053900 Lines 532 through 644

COMPARERESULTS

The COMPARERESULTS function compares two separate LARSFRIS Classification Results Files, pixel for pixel, and produces an output to disk or tape in LARSFRIS Classification Results file format. It can be used to detect ground cover change from temporally overlayed data, or to compare classification results obtained via different channels of data, different training methods, or using different classifiers.

Input to the function consists of two Classification Results Files and in input deck of control cards. The main output is another Classification Results File consisting of user-defined "change" classes. To analyze the results of the COMPARERESULTS job the user will probably want to run the PRINTRESULTS processor. The user should note that the COMPARERESULTS function could require the use of three tape drives (2 for input and 1 for output). If this is the case, the user should contact Purdue/LARS computer operator before running the COMPARERESULTS job to be sure that the 3 required tape drives are available.

Figure 1 pictures a typical input deck for the COMPARERESULTS function. In general, the input deck will cause the classification results files designated on the FIRSTRESULTS and SECONDRESULTS card to be compared, point by point, and new class numbers output to the tape (or disk) designated by the NEWRESULTS card. The cards following the DATA card determine these new class numbers. The

```
*COMPARE
FIRSTRESULTS TAPE (195), FILE(3)
SECONDRESULTS DISK
NEWRESULTS TAPE (9999), FILE(2)
BLOCK RUN (77010200), LINE (1710,1885,1), COL(2900,3040,1)
DATA
CLASS PINE
FIRST 1,2,6,9
SECOND 1
CLASS MIX P/H
FIRST 3,4
SECOND 2
CLASS SLH/CYP
FIRST 7,8,11,12
SECOND 3
CLASS LOWSTOCK
FIRST 5,10
SECOND 4
CLASS NONSTOCK
FIRST 13,14,15
SECOND 5
CLASS WETLAND
FIRST 16,17,18
SECOND 6
CLASS PINE-OTH
FIRST 1,2,6,9
SECOND 2,3,4,5,6
END
```

*Note: To duplicate this example, run *CLASSIFY with the PROCEDURE L2 option on run 77010200, lines 1710 to 1885, and columns 2900 to 3040. Use the statistics file FLEMING STATDECK and store the results on the FIRSTRESULTS tape. Then run the *SMOOTH job illustrated in the SMOOTHRESULTS section, storing that output file on disk. Be sure to substitute your own tape and file numbers for the ones in this example.*

Figure 1. Typical input deck for the COMPARERESULTS function

user may assign a class name to any class combinations from the input tapes and those combinations not specifically mentioned will be placed in a class named *CHANGE* by the program.

Required Control Cards

All of the control cards for COMPARERESULTS are required. The FIRSTRESULTS and SECONDRESULTS cards specify the locations of the two Classification Results Files that are used for input, and the NEWRESULTS card indicates the location for the output Results File. The BLOCK card displays which part of the area is to be compared, and the user (analyst) can define which class combinations are of interest to him with the CLASS, FIRST and SECOND cards.

FIRSTRESULTS Card

The first LARSFRIS Classification Results File to be input may reside on tape or disk. If it is on tape, the FIRSTRESULTS card will specify the tape and file numbers. In this example, the results file is on tape 195, file 3. The DISK parameter should be used if the results file is to be taken from disk. The results file from the last run of CLASSIFYPOINTS (subsequent to the 'i larsys' command) will be available on disk.

SECONDRESULTS Card

The second Classification Results File to be input may reside on tape, or on disk if the first results file is not on disk. If it is on tape, the SECONDRESULTS card will give the tape and file numbers. The DISK parameter specifies that the results file is to be taken from disk as in Figure 1. Remember that only one of the results files may use the DISK option. An error message will be printed if the DISK option is specified for more than one file.

NEWRESULTS Card

The NEWRESULTS card specifies the location for the output LARSFRIS Classification Results File which, like the two input files, may be placed on tape or disk. If it is to go on tape, the NEWRESULTS card will specify the tape number and either the file number or INIT. The INIT option is used to initialize a tape for LARSFRIS Results Files and write the Classification Results File in file 1 of the tape.

If PRINTRESULTS immediately follows a COMPARERESULTS run, the output classification Results File tape which was used will remain mounted, but the tape and file numbers still must be specified on the RESULTS card for PRINTRESULTS. If the file is to go on a scratch tape, a tape number of '000' may be used. If scratch tapes are used when COMPARERESULTS and PRINTRESULTS are run successively, the tape remains mounted for PRINTRESULTS and no further identification of the tape number other than '000' is needed.

The DISK parameter may be used to specify that the results file is to be written on disk, provided that neither of the two input Results Files are on disk. Remember that, when the Classification Results File is placed on disk, it is only stored there temporarily. Any of the following actions will cause it to be erased from the system.

* Executing another function which creates a Classification Results File (e.g. CLASSIFYPOINTS, COMPARERESULTS,...)
* Re-initiating LARSFRIS (i.e., typing the 'i larsys' or 'i larsysdv' control command)
* Logging off the system (i.e., typing the 'quit' control command)

If the user wishes to save the file, he may copy it to tape with the COPYRESULTS function.

BLOCK Card

The user defines the area to be compared by using a BLOCK card. This card specifies the run number and the line and column coordinates for the area. It is assumed that the two Results Files to be compared will have come from the same data run because the data used for the two input Classification Results Files should be registered. If the program finds different run numbers in the results files, a warning message is printed but execution continues.

Defining Classes for COMPARERESULTS

The user must supply a data deck defining the class changes (or non-changes) which he is interested in identifying. Grouping of classes is desirable in this implementation of results comparison because of the prohibitive number of possible class combinations generated from comparing the two input classifications. With 'M' spectral classes from the first classification and 'N' spectral classes in the second classification, it is possible to have 'M*N' classes in the output. It is quite probable that a user will be more interested in a few particular types of changes than in every combination. For this reason, the user must define each new class. Any combination that is not defined is thrown into a general class called *CHANGE*. Therefore, the total number of classes on the output results file is C + 1, where C is the number of classes the user defined.

To define the classes, a DATA card is first needed to designate the start of the data deck. Then there will be a set of three cards (CLASS, FIRST and SECOND) for each class the user is defining. For the example in Figure 1, the FIRSTRESULTS file on tape 195, file 3 consisted of 18 classes and the SECONDRESULTS file on disk had 6 classes, as defined below:

	Class numbers	
Type	FIRSTRESULTS File	SECONDRESULTS File
Pine	1,2,6,9	1
Mixed Pine/Hardwood	3,4	2
Slash Pine & Cypress	7,8,11,12	3
Low-stocked	5,10	4
Non-stocked	13,14,15	5
Wetland	16,17,18	6

In the data deck, the CLASS card contains the name of the class; in the example in Figure 1, the first class is being given the name PINE. The CLASS card is followed by a FIRST card, which contains the class numbers (separated by commas) from the first classification Results File (specified on the FIRSTRESULTS card) for the class being defined. The SECOND card contains class numbers from the second Classification Results File (specified on the SECONDRESULTS card) for the class being defined. Hence, if the computer is looking at a pixel on the first Classification Results File which belongs to one of the classes listed on the FIRST card for user-defined class one, and on the second Classification Results File the same pixel belongs to one of the classes listed on the SECOND card for user-defined class one, then the computer will output class number one for that pixel. In the example from Figure 1, the user's first class is named PINE. A pixel will be assigned to this class if it is in one of classes 1,2,6, or 9 on the first Classification Results File, and in class 1 on the second Classification Results File. Presumably this will identify those points which were pine in both classifications. The last class, PINE-OTH, will identify those classes that were pine on the base classification (classes 1,2,6,9) and changed to something else on the second classification (classes 2,3,4,5,6).

END Card

As in all LARSFRIS functions, the END card is required to indicate the end of the data and control cards for the COMPARERESULTS function.

Standard Output

Three standard outputs are produced by COMPARERESULTS. They are:

*an input deck summary

*a table of output class numbers and names

*a LARSFRIS Classification Results File

A sample of the input deck summary is shown in Figure 2. The class table is illustrated in Figure 3. In order to see the results of running the COMPARERESULTS job, the user should execute the PRINTRESULTS function to get a map and/or table(s) of the COMPARERESULTS Classification Results File which was created. The format of the serial number of a Results file produced by COMPARERESULTS has been altered to TTTTFttttf where 'TTTT' is the tape number of the first results file, 'F' is the file number, 'tttt' is the tape number of the second results file and 'f' is the corresponding file number. Should either file number be greater than ten, only the right-most digit will be recorded. Also, if either file was located on disk, the tape and file numbers would be zero.

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

```
*COMPARERESULTS
FIRSTRESULTS TAPE(195),FILE(3)
SECONDRESULTS DISK
NEWRESULTS TAPE(9999), FILE(2)
BLOCK RUN(77010200),LINE(1710,1885,1), COL(2900,3040,1)
DATA
CLASS PINE
FIRST 1,2,6,9
SECOND 1
CLASS MIX P/H
FIRST 3,4
SECOND 2
CLASS SLH/CYP
FIRST 7,8,11,12
SECOND 3
CLASS LOWSTOCK
FIRST 5,10
SECOND 4
CLASS NONSTOCK
FIRST 13,14,15
SECOND 5
CLASS WETLAND
FIRST 16,17,18
SECOND 6
CLASS PINE-OTH
FIRST 1,2,6,9
SECOND 2,3,4,5,6
END
```

Figure 2. Sample Input Deck Summary

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

	CLASS		CLASS
1	PINE	5	NONSTOCK
2	MIX P/H	6	WETLAND
3	SLH/CYP	7	PINE-OTH
4	LOWSTOCK	8	*CHANGE*

Figure 3. Class table

COPYRESULTS

The Copyresults function copies selected Classification Results Files to tape for selectively storing results. A results file may be copied on tape from disk or another tape.

Input

The input to COPYRESULTS is the Classification Results File on tape or disk and the function input deck. The input deck consists of required FROM and TO cards and an optional PRINT card. Figure 1 is a sample input deck for this function.

The FROM card specifies the particular file to be copied and may be in one of the following forms:

FROM TAPE(xxx),FILE(xx) - tape and file number to be copied

FROM TAPE(xxx),ALL - all files on the tape are to be copied

FROM FILE(xx) - file to be copied from a scratch tape

FROM ALL - all files on a scratch tape are to be copied

FROM DISK - file currently on disk to be copied

In the figure, the FROM card requests a file be copied from disk.

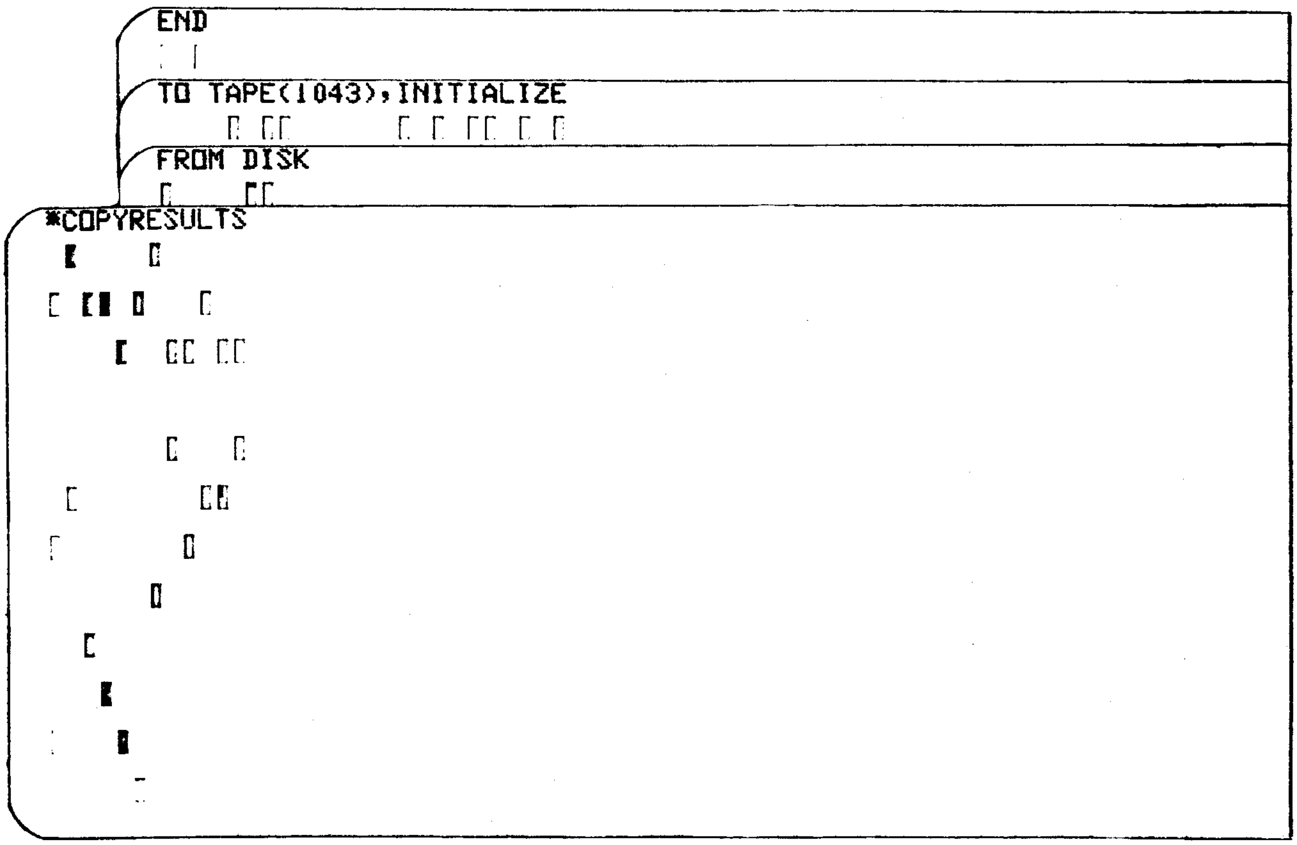


Figure 1. Input Deck for COPYRESULTS

The destination of the copied output is specified on the TO card. The tape specified may not be a scratch tape (TAPE(0)). The card may be in one of the following forms:

TO TAPE(xxx),FILE(xx)

TO TAPE(xxx),INITIALIZE

The first form of the card specifies the tape and file to which the results file is to be copied. The second indicates a new tape is to be used. The 'INITIALIZE' parameter requests that the proper header information be placed at the beginning of the new tape, and the results file be automatically written into file 1. The sample card in the figure indicates the results file is to be copied onto tape 1043 in file 1.

The user may elect to suppress the standard listing provided by this function, by including the 'PRINT NOLIST' card in the input deck. If this optional card is omitted, the complete listing will be produced.

Output

The principal output of this function is the Classification Results File copied onto tape. The new tape will have a tape and file number as requested by the 'TO' control card.

COPYRESULTS prints a list of the control cards entered and a table of how they have been interpreted. An example of this output is shown in Figure 2.

The other printed output is the Classification Results File ID listing, which is the same as that produced by the Listresults function. An example of this listing is shown in Figure 3. If the 'PRINT NOLIST' card is used, the listing will not be printed.

```
     DEMO                  LABORATORY FOR APPLICATIONS OF REMOTE SENSING       MAR  7,1973
     PHILLIPS                        PURDUE UNIVERSITY                         12 19 57 PM
                                                                               LARSYS VERSION 3

    *COPYRESULTS
    FROM DISK
    TO TAPE(1043),INITIALIZE
    END

 YOU HAVE SELECTED THE FOLLOWING OPTIONS

    INITIALIZE OUTPUT RESULTS TAPE
    RESULTS WILL BE COPIED FROM DISK TO TAPE 1043 FILE  1

                                                                                      COP-2
```

Figure 2. Control Card Summary for COPYRESULTS

```
DEMO                           LABORATORY FOR APPLICATIONS OF REMOTE SENSING                    AUG  30,1974
PAUL SPENCER                                 PURDUE UNIVERSITY                                   2 56 29 PM
                                                                                                LARSYS VERSION 3


                TAPE-- 1043            FILE--  1           CLASSIFIED- AUG  30,1974
                RESULTS FROM DISK


                                               CHANNELS USED
                                            -----------------
            CHANNEL  2    SPECTRAL BAND  0.48 TO 0.51 MICROMETERS    CALIBRATION CODE =  1
            CHANNEL  9    SPECTRAL BAND  1.00 TO 1.40 MICROMETERS    CALIBRATION CODE =  1
            CHANNEL 12    SPECTRAL BAND  9.30 TO11.70 MICROMETERS    CALIBRATION CODE =  1


                                            CLASSES AND WEIGHTS
                                         ----------------------
         CLASS          CLASS          CLASS          CLASS          CLASS          CLASS          CLASS
         WEIGHT         WEIGHT         WEIGHT         WEIGHT         WEIGHT         WEIGHT         WEIGHT

         CORN           SOYBEANS       OAT/WHT        HAY            FOREST
         0.290          0.310          0.095          0.110          0.195


      RUN                                        AREA CLASSIFIED                              FLIGHT LINE
    --------                                  -------------------                          ---------------
    71053900         LINES----    425 TO    644 BY  2     COLUMNS--    1 TO  222 BY  2     CRN BLT LO FL210

                                               CALIBRATION VALUES
                                            ---------------------
                                           CHANNEL  2 CO =  23.10
                                           CHANNEL  9 CO =  22.90
                                           CHANNEL 12 CO =  20.60
                                    NUMBER OF LINES CLASSIFIED IS    110
```

Figure 3: Listing of Results File from COPYRESULTS

GRAPHHISTOGRAM

The Graphhistogram function generates graphs from a Histogram file computed by the Histogram or Pictureprint function. It is normally used only in conjunction with the Histogram function, however, since the other function has it's own graphing capability. The Histogram File may be computed immediately before the graphs are printed by GRAPHHISTOGRAM (i.e., since the last 'i larsys' command), or it may have been computed earlier and stored on the user's permanent disk via the 'HISTDECK SAVE' command. In this case, it is moved back to the temporary disk via the 'HISTDECK USE' command, before executing GRAPHHISTOGRAM.

Input to the function is the Histogram file and control cards. There are three control cards, all of which are required. These are a function Selector Card, a CHANNELS card, and an END card. The CHANNELS card specifies the channels to be graphed. If a channel is specified for which no histogram data was computed, a message will be printed to that effect in lieu of the graph for the channel.

The outputs from the function are an input deck summary and the printed graphs of histograms for the selected channels. An example input deck that requests graphs for channels 7 and 9 is shown in Figure 1. The summary listing of the deck is shown in Figure 2, and the two graphs that are produced are shown in Figure 3.

Each of the two graphs has a header identifying the channel number; spectral band definition; run number; calibration code; the lines, columns and number of samples represented; the total number of samples represented by each asterisk (*) on the graph; and the total number of samples represented on the graph. Note that special symbols are used instead of asterisks along the edges of the graphs to represent less than a "full sample count". If the count of samples is 1 to 9, those numbers are used on the graph. If the count is 10 to 35, the letters A to Z are used, and if the count exceeds 35 (i.e., 36 to whatever the full count for an asterisk is), then the dollar symbol is used.

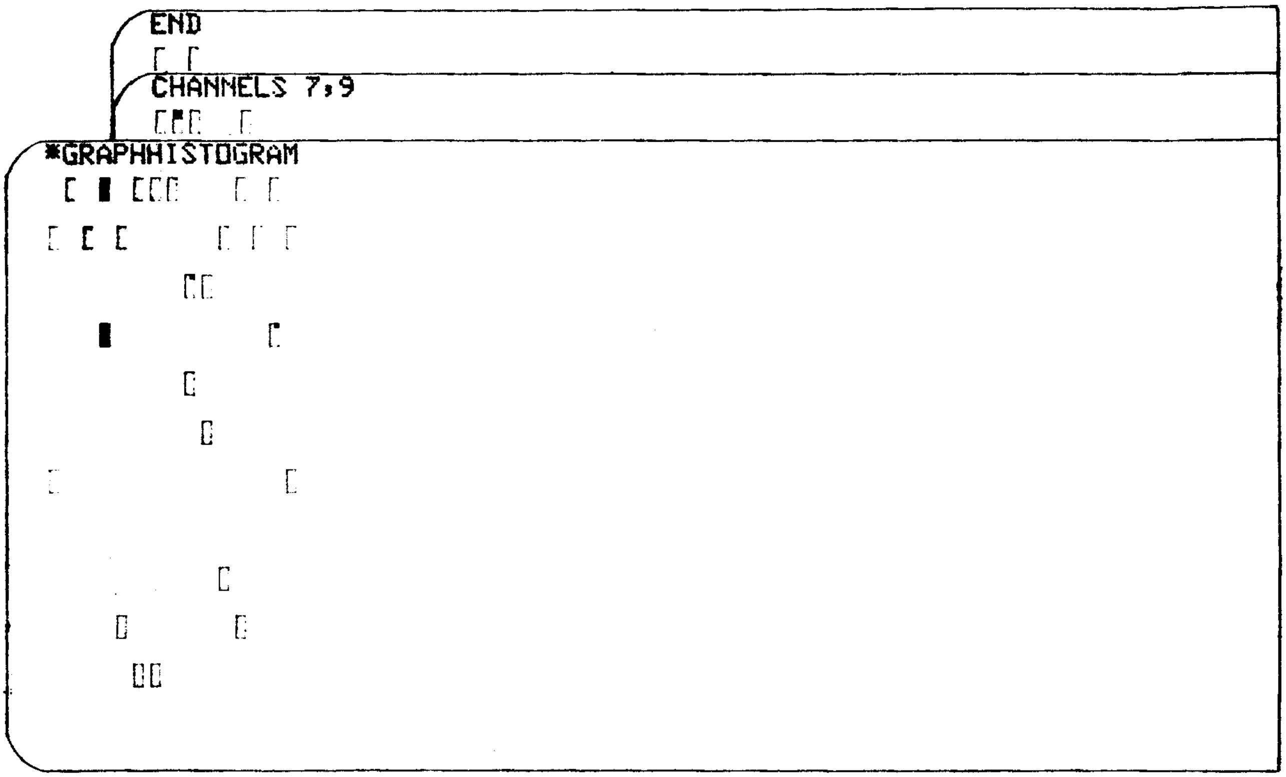


Figure 1. Input Deck for GRAPHHISTOGRAM

```
     DEMO                         LABORATORY FOR APPLICATIONS OF REMOTE SENSING             MAR   7,1973
     PHILLIPS                                   PURDUE UNIVERSITY                           11 05 50 AM
                                                                                            LARSYS VERSION 3

   *GRAPHHISTOGRAM
   CHANNELS 7,9
   END

                                                                                         GRA-2
```

Figure 2. Control Card Summary for GRAPHHISTOGRAM

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

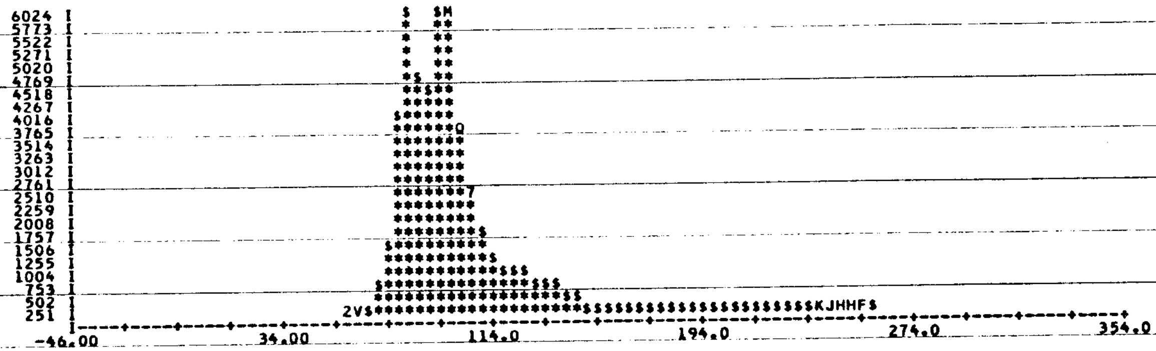
MAR 7,1973
11 11 43 AM
LARSYS VERSION 3

****** HISTOGRAM GRAPH FOR CHANNEL 7 ******

SPECTRAL BAND 0.61 TO 0.70 MICROMETERS

RUN... 71053900 CALIBRATION CODE... 1 LINES... (420, 644, 1) COLUMNS... (1, 221, 1)

EACH * REPRESENTS... 251 SAMPLES. TOTAL NUMBER OF SAMPLES... 49725

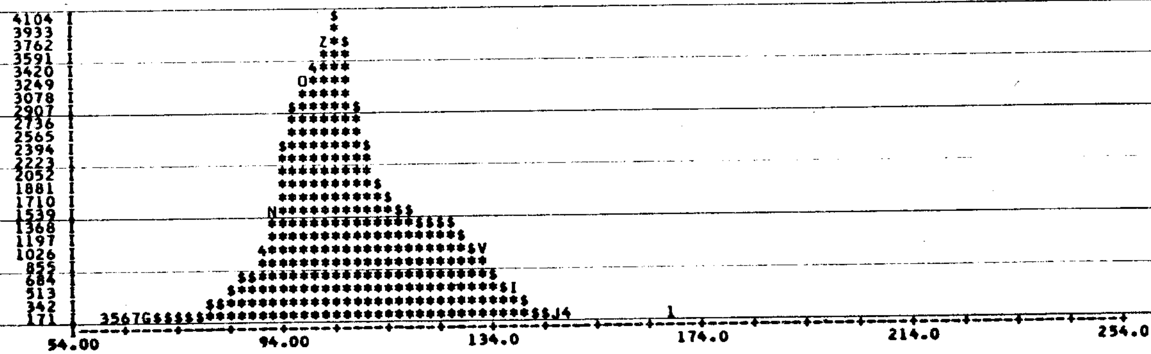


****** HISTOGRAM GRAPH FOR CHANNEL 9 ******

SPECTRAL BAND 1.00 TO 1.40 MICROMETERS

RUN... 71053900 CALIBRATION CODE... 1 LINES... (420, 644, 1) COLUMNS... (1, 221, 1)

EACH * REPRESENTS... 171 SAMPLES. TOTAL NUMBER OF SAMPLES... 49725



GRA-3

Figure 3. Histogram Graph for Channels 7 and 9

HISTOGRAM

The Histogram function produces a histogram of the spectral values for selected channels from the Multispectral Image Storage Tape. The data from this input tape is reduced to the summation of the total number of data values which fall within the range of any of 100 equal-size bins. The range of the bins and the starting (lowest) data value for the histogram is adjusted depending upon the actual distribution of the data, and the histogram is centered in the 100 bins to provide maximum resolution.

The user may specify the Multispectral Image Storage Tape run number and the lines and columns that are to be histogrammed, or he may use the system's default values. The line and column defaults may always be used, but the run number default (which is the last run that was used) is only applicable if another function that used a Multispectral Image Storage Tape was run since the last 'i larsys' command. The user must always specify the channels that are to be histogrammed, and he may use an alternate form of the CHANNELS control card to also specify the calibration of the data from the channels.

The file of histograms that is produced by this function is always stored on the temporary disk and may also be punched on cards. The user may employ the 'HISTOGRAM SAVE' command after running the Histogram function to save this file on his private A-disk. The standard printer output reproduces the user's input deck and summarizes the data that was histogrammed and the statistics for the histograms.

The histograms that are stored on disk may be used as input to GRAPHHISTOGRAM to produce graphs of the histograms. The histograms (whether on disk or cards) may also serve as input to the Pictureprint function. This function uses it to allocate symbols or gray levels to data ranges. (Pictureprint can also compute the histograms using the same procedures as the Histogram function.)

An optional feature of the Histogram function allows the user to include the results of previous histogramming in the new histogram. The function uses the file of histograms stored on the temporary disk, updates them with the new calculations, and places the combined results back on the disk.

Specification of Area to be Histogrammed

The input data file is a Multispectral Image Storage Tape. The function normally uses the information on the BLOCK card in conjunction with the system or a user Runtable to identify the appropriate tape and have it mounted. The content and form of these tapes are described in Appendix IV.

If the BLOCK card is omitted or the RUN number not specified on the BLOCK card. the function will histogram data from the last Multispectral Image Storage Tape run that was used since the last 'i larsys' command. If the line and column numbers and the intervals are not specified the system will use every tenth column from every tenth line on the run. Figure 1 shows a sample input deck that includes a typical BLOCK control card. It requests histograms based on lines 420 through 644 and, columns 1 through 221, of run 71053900.

If the user wants to include previously histogrammed data in the new calculations, he includes an 'OPTIONS ACCUM' control card in his deck. This process is called "accumulation". HISTOGRAM will then use as input whatever histograms are present on the temporary disk. This will consist of the results of the last histogramming performed since the 'i larsys' command; histograms placed there as a result of a 'HISTOGRAM USE' command.

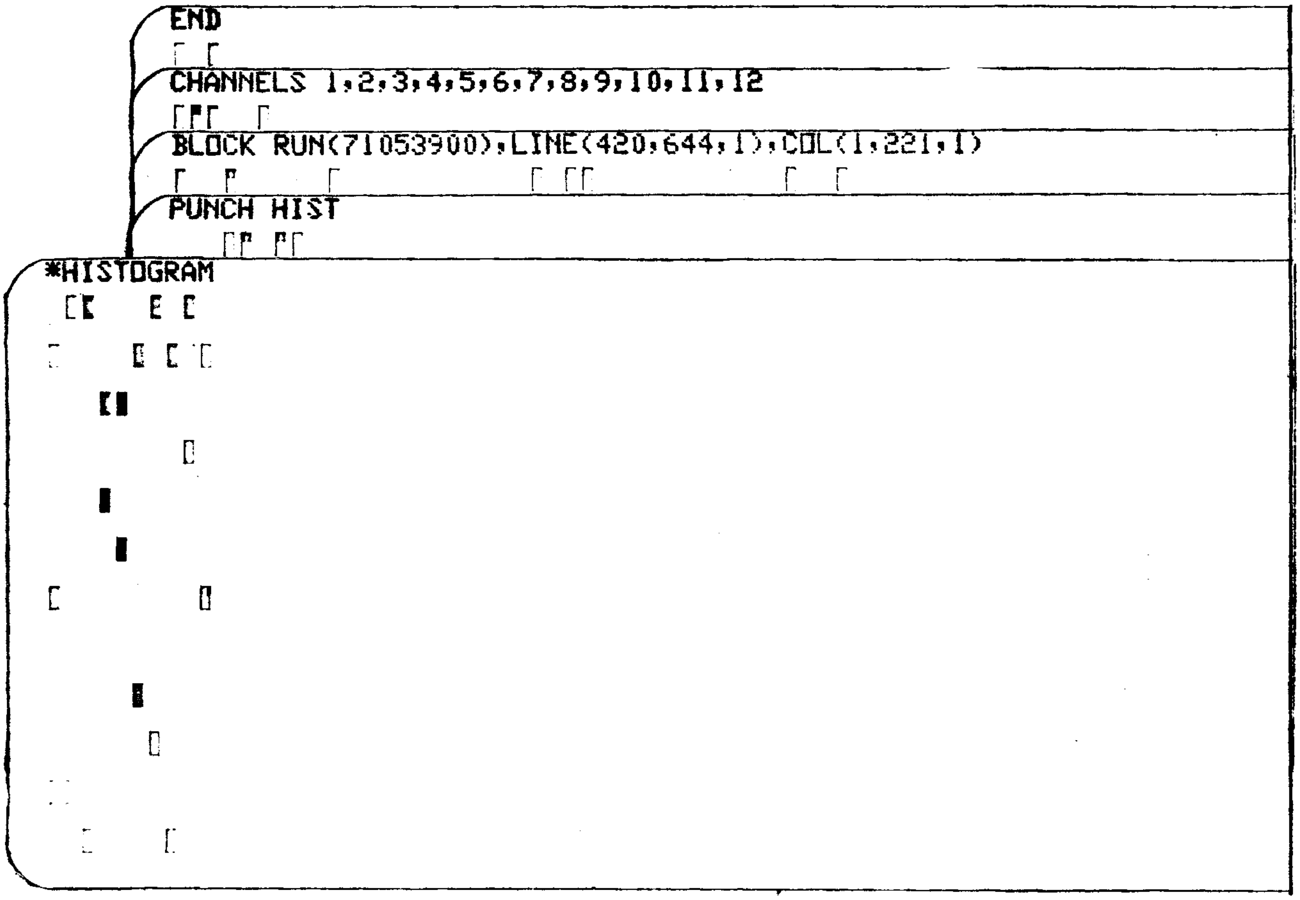


Figure 1. Input Deck for HISTOGRAM

Specification of Channels

The channels that are to be histogrammed must be specified on a 'CHANNELS' control card. A separate histogram is produced for the data from each requested channel. The 'CHANNELS' card in Figure 1 requests channels 1 through 12. For more information on the CHANNELS card and calibration of the data values refer to Appendix I and Appendix IV.

Output Files

The Histogram function always writes its histograms on the temporary disk. The user may also get the histograms punched on cards by using the PUNCH HIST control card. The card deck is useful for saving histograms for later use (a single Histogram File may be saved on the user's permanent disk and used in later commands). The user may want to check to see if he has a complete histogram deck. The first card should have the words LARS HISTO beginning in column 1. The last card should have

    EOH ***** LAST DATA CARD OF HISTOGRAM DECK *****

in the first 48 columns and a sequence number in columns 73 - 80. All of the preceeding cards in the deck except the first card, also have this sequence number.

Standard Printer Output

The Histogram function always prints a list of the control cards, the options selected, and particular characteristics about the run such as the number of channels requested. An example of this output is shown in Figure 2.

In addition, HISTOGRAM prints histogram summary information and statistics. This listing identifies the histogrammed data and lists the data range, mean, standard deviation and the normalized range (mean plus and minus the standard deviation) for each histogrammed channel. A sample of this output is shown in Figure 3.

```
DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING           MAR  7,1973
PHILLIPS                         PURDUE UNIVERSITY                              11 04 48 AM
                                                                                LARSYS VERSION 3

   *HISTOGRAM
   PUNCH HIST
   BLOCK RUN(71053900),LINE(420,644,1),COL(1,221,1)
   CHANNELS 1,2,3,4,5,6,7,8,9,10,11,12
   END

YOU HAVE SELECTED THE FOLLOWING HISTOGRAM OPTIONS
   PUNCH HISTOGRAMS

HISTOGRAM INFORMATION....
   NUMBER OF CHANNELS... 12

                                                                                HIS-2
```

Figure 2. Control Card Summary for HISTOGRAM

DEMO
PHIL

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 27,1974
2 35 07 PM
LARSYS VERSION 3

RUN NUMBER............ 71053900
FLIGHT LINE... CRN BLT LO FL210
DATA TAPE/FILE NUMBER.. 1042/ 1
REFORMATTING DATE. AUG 16,1971

DATE DATA TAKEN... AUG 13,1971
TIME DATA TAKEN..... 1202 HOURS
PLATFORM ALTITUDE.. 5000 FEET
GROUND HEADING..... 180 DEGREES

CHANNEL 1 SPECTRAL BAND 0.46 TO 0.49 MICROMETERS CALIBRATION CODE = 1 CO = 26.25
CHANNEL 2 SPECTRAL BAND 0.48 TO 0.51 MICROMETERS CALIBRATION CODE = 1 CO = 23.10
CHANNEL 3 SPECTRAL BAND 0.50 TO 0.54 MICROMETERS CALIBRATION CODE = 1 CO = 26.80
CHANNEL 4 SPECTRAL BAND 0.52 TO 0.57 MICROMETERS CALIBRATION CODE = 1 CO = 22.95
CHANNEL 5 SPECTRAL BAND 0.54 TO 0.60 MICROMETERS CALIBRATION CODE = 1 CO = 26.25
CHANNEL 6 SPECTRAL BAND 0.56 TO 0.65 MICROMETERS CALIBRATION CODE = 1 CO = 24.00
CHANNEL 7 SPECTRAL BAND 0.61 TO 0.70 MICROMETERS CALIBRATION CODE = 1 CO = 28.00
CHANNEL 8 SPECTRAL BAND 0.72 TO 0.92 MICROMETERS CALIBRATION CODE = 1 CO = 24.85
CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS CALIBRATION CODE = 1 CO = 22.90
CHANNEL 10 SPECTRAL BAND 1.50 TO 1.80 MICROMETERS CALIBRATION CODE = 1 CO = 24.90
CHANNEL 11 SPECTRAL BAND 2.00 TO 2.60 MICROMETERS CALIBRATION CODE = 1 CO = 25.95
CHANNEL 12 SPECTRAL BAND 9.30 TO 11.70 MICROMETERS CALIBRATION CODE = 1 CO = 20.60

DATA BLOCK(S) HISTOGRAMMED

CHANNEL	RUN NUMBER	CALIBRATION CODE	FIRST LINE	LAST LINE	LINE INTERVAL	FIRST SAMPLE	LAST SAMPLE	SAMPLE INTERVAL
1	71053900	1	420	644	1	1	221	1
2	71053900	1	420	644	1	1	221	1
3	71053900	1	420	644	1	1	221	1
4	71053900	1	420	644	1	1	221	1
5	71053900	1	420	644	1	1	221	1
6	71053900	1	420	644	1	1	221	1
7	71053900	1	420	644	1	1	221	1
8	71053900	1	420	644	1	1	221	1
9	71053900	1	420	644	1	1	221	1
10	71053900	1	420	644	1	1	221	1
11	71053900	1	420	644	1	1	221	1
12	71053900	1	420	644	1	1	221	1

HISTOGRAM STATISTICS

CHANNEL	DATA RANGE	MEAN	STANDARD DEVIATION	NORMALIZED RANGE (MEAN + AND - 3 STD DEV)
1	90.8 TO 257.	130.	19.0	73.3 TO 187.
2	65.6 TO 256.	96.5	15.2	50.9 TO 142.
3	65.3 TO 257.	99.6	14.3	56.6 TO 143.
4	81.4 TO 257.	137.	20.3	76.6 TO 198.
5	74.8 TO 257.	126.	22.0	60.0 TO 192.
6	60.5 TO 257.	107.	24.5	33.2 TO 180.
7	53.5 TO 258.	97.9	26.6	18.2 TO 178.
8	58.3 TO 186.	115.	18.0	60.6 TO 169.
9	58.4 TO 168.	107.	13.2	67.4 TO 147.
10	60.4 TO 258.	136.	23.3	66.1 TO 206.
11	52.4 TO 260.	117.	22.8	48.7 TO 185.
12	-1.90 TO 262.	103.	27.8	19.4 TO 186.

Figure 3. Histogram summary from HISTOGRAM.

The Histogram Algorithm

The histogram algorithm creates a 100-bin histogram for the data from each specified channel on the Multispectral Image Storage Tape. One histogram is produced for each channel. The location of the bin edges and the data values represented by each bin (all bins have equal width) are determined dynamically based on the actual range and distribution of the calibrated data.

The recorded values on the tape are integers that may vary between 0 and 255. However, when the data is calibrated, the values are converted to floating point numbers and can have any value. The basic equation for calibration is:

$$D_C = A(D_T) + B$$

where

D_C = the calibrated data value for a sample.

D_T = the digitized data value stored on the tape.

A and B = calibrating constants that are calculated for each line.

A and B may change slightly from line to line but are constant within a line. See Appendix IV for more detailed information on calibration.

The histogram algorithm usually assumes an initial bin size of A and an initial lower limit for the histogram of A(14.5) + B. The only exception to this is when the histograms are to be accumulated with previous histograms. In this latter case, the bin size and lower limit are set to those of the previous (stored) histogram. After setting these values, the Histogram function follows this procedure:

1. The bin number for the data value is calculated as:

$$\text{Bin Number} = \frac{\text{(Data Value)} - \text{(Histogram Lower Limit)}}{\text{(Bin Size)}}$$

2. If the calculated bin number is 1 through 100, the count of samples occurring in the data range represented by that bin is incremented by one, and the algorithm proceeds to the next sample.

3. If the calculated bin number is less than 1 or greater than 100, the histogram must be adjusted to accommodate the new data range that the sample requires. The algorithm first attempts to accomplish this by shifting the upper and lower limits of the histogram while keeping the individual bin size constant. It can do this if there are enough bins at the opposite end of the histogram that have zero counts.

   For example, if the calculated bin number for a new sample is 102, the function checks to determine if bins 1 and 2 of the histogram have zero counts, i.e., no samples have occurred in the range that they represent.

If this is true, the range of the histogram is shifted higher. Bin 3 will become the new bin 1, bin 4 the new bin 2, etc., until the data values required for the calculated bin 102 are represented by the new bin 100. Its count of sample occurrences is incremented by one, and processing passes to the next sample. In this way, two new bins are added, and the range increased at the upper end of the histogram. Two unused bins are dropped, and the range is decreased at the lower end.

4. If it is not possible to accommodate the new data value by shifting the histogram, then it must be accomplished by increasing the overall range of data values represented by the histogram. This is done by doubling the size of each bin, or in practice by consolidating the data range represented by the 100 bins into 50 larger bins, and adding 50 new bins to increase the data range.

   The algorithm first calculates the actual data range (difference between the high and low values in the histogram) that is required to accommodate the new sample and the number of times that the old range must be doubled in order to achieve it. The range is then doubled that number of times, the counts of sample occurrences are added into the appropriate new bin, and the procedure resumes at step 1 above with the calculation of a new bin number for the sample. In some cases the new data value may still lie outside the high and low values represented by the histogram, and the procedure must go through the shifting of the range in step 3 in order to accommodate it.

IDPRINT

The Idprint function prints a listing of the identification record of Multispectral Image Storage Tapes runs. Requests may be made for the ID record for a specific run, for all runs on one tape, or for all runs maintained by LARSFRIS. The ID records for runs that are maintained by LARSFRIS are stored on the System Runtable, and there is no need to actually locate and mount these tapes. To print the ID record of runs that are not cataloged in the System Runtable, however, the actual tapes must be mounted and the ID record read from them. The reader should refer to Appendix IV for more information on the content of the ID record and on the use of the System Runtable.

Input

Input to IDPRINT consists of the function input deck and one or more Multispectral Image Storage Tape ID records. The same three control cards, and no others, are always required for each execution of the Idprint function. These are a Function Selector Card, a 'PRINT' card and an 'END' card. There are three possible parameters on the 'PRINT' control card, only one of which may be used for each execution of the function:

- RUN(xxxxxxxx) - requests the ID record for a specific run from the System Runtable.

- ALL - requests ID records for all runs cataloged in the System Runtable
- TAPE(xxx) - requests ID records for all runs on the specified tape

A tape will have to be mounted only for requesting ID records using the 'TAPE(xxx)' parameter. This is done to provide the same run sequence as that of the tape, or to print ID records from uncataloged tapes.

An example input deck that requests two executions of the Idprint function is shown in Figure 1. The first of these execution will print the ID record for run 71053900, a run that is cataloged in the System Runtable. Following this, the second execution of the function will print the ID records for all of the runs that are recorded on tape number 1042. The runs on this tape may or may not be cataloged in the System Runtable. In either case, the tape will be mounted and the ID records printed from it in the sequence that they appear there.

The reader should note that the only way to get a completely accurate listing of the ID records of <u>private copies</u> of runs that are also cataloged in the System Runtable is to use the 'TAPE' parameter and have the tape actually mounted. If the user were to use the 'RUN' parameter option, he would get a

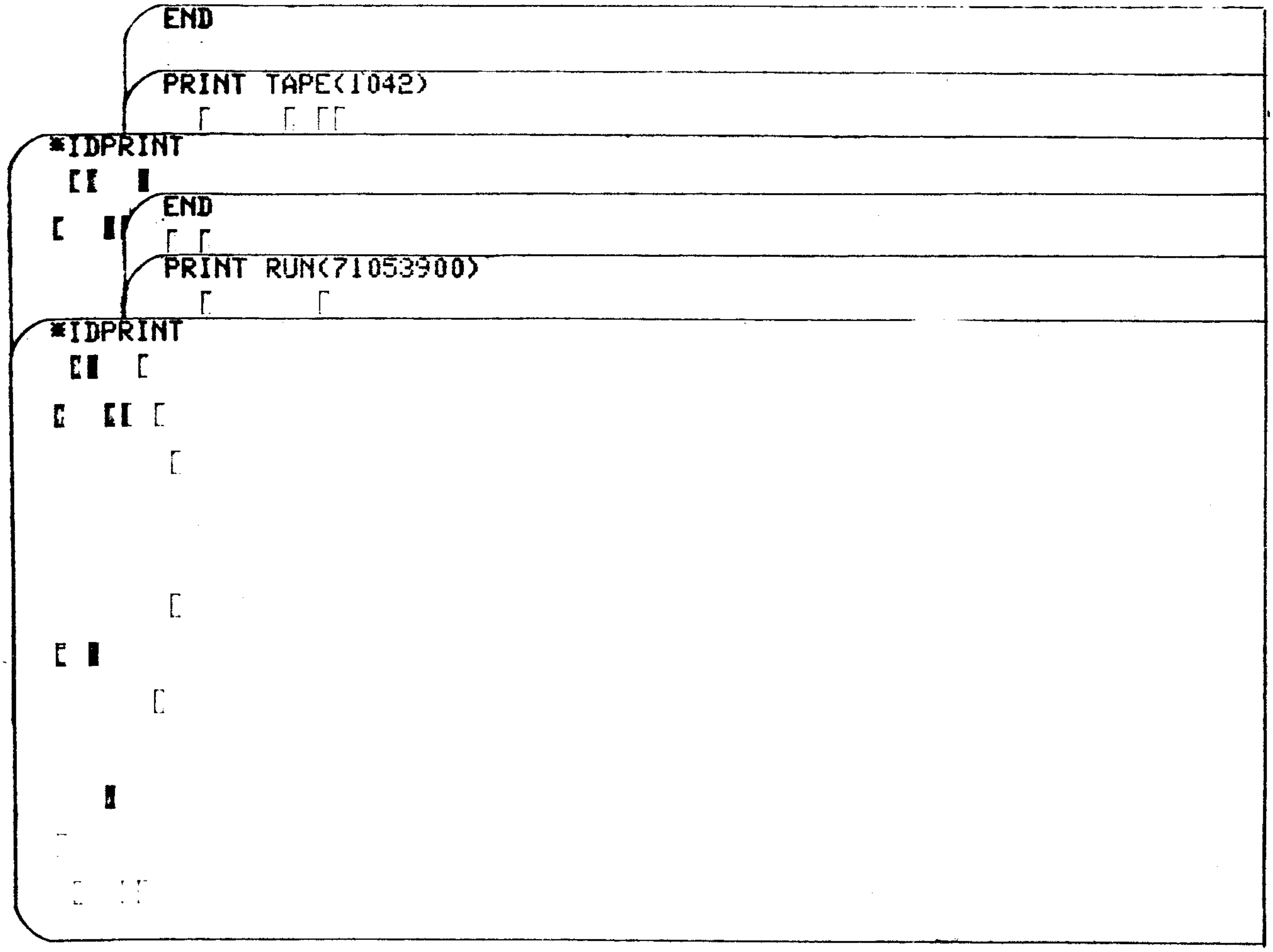


Figure 1. Input Deck for IDPRINT

listing of the ID record of the system copy of the run. This would show a different tape and file number than his private copy.

Output

The only outputs from this function are the listings of the function input cards, and of the ID records themselves. Figure 2 shows the input card listing from the first execution of the function from the example input deck in Figure 1. It consists simply of a listing of the cards that were input to the function. Figure 3 shows the ID record listing that was printed for this same execution, i.e., for run 71053900. Figure 4 shows the first ID record listing that was printed for the second execution, which requested all of the ID records from tape 1042. Note that the first run on this tape was also run 71053900. These two listings are identical except for the different tape and file numbers.

```
    DEMO                          LABORATORY FOR APPLICATIONS OF REMOTE SENSING              MAR  7,1973
    PHILLIPS                                   PURDUE UNIVERSITY                             10 24 15 AM
                                                                                             LARSYS VERSION 3

   *IDPRINT
   PRINT RUN(71053900)
   END

                                                                                                  ID-2
```

Figure 2. Control Card Summary for IDPRINT

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 02 05 AM
LARSYS VERSION 3

TAPE NUMBER.............. 1005	FILE NUMBER............... 2	RUN NUMBER........... 71053900
CONTINUATION CODE.......... 0	NUMBER OF DATA CHANNELS.... 12	NUMBER OF DATA SAMPLES... 228
FLIGHT LINE.. CRN BLT LO FL210	DATE DATA TAKEN...... 8/13/71	TIME DATA TAKEN.... 1202 HOURS
PLATFORM ALTITUDE. 5000 FEET	GROUND HEADING.... 180 DEGREES	REFORMATTING DATE.AUG 16,1971
	NUMBER OF LINES......... 1161	

	SPECTRAL BAND		CALIBRATION PULSE VALUES		
CHANNEL	LOWER	UPPER	CO	C1	C2
1	0.46	0.49	26.25	178.25	207.55
2	0.48	0.51	23.10	144.20	163.30
3	0.50	0.54	26.80	132.20	159.00
4	0.52	0.57	22.95	166.10	209.10
5	0.54	0.60	26.25	179.70	209.00
6	0.58	0.65	24.00	186.55	203.80
7	0.61	0.70	28.00	168.15	205.50
8	0.72	0.92	24.85	118.55	69.60
9	1.00	1.40	22.90	31.30	56.70
10	1.50	1.80	24.90	74.85	135.25
11	2.00	2.60	25.95	94.05	140.60
12	9.30	11.70	20.60	137.00	0.0

ID-3

Figure 3. ID Record Listing from System Runtable

DEMO PHILLIPS	LABORATORY FOR APPLICATIONS OF REMOTE SENSING PURDUE UNIVERSITY	MAR 7,1973 11 03 11 AM LARSYS VERSION 3

TAPE NUMBER............... 1042	FILE NUMBER............... 1	RUN NUMBER........... 71053900
CONTINUATION CODE........... 0	NUMBER OF DATA CHANNELS.... 12	NUMBER OF DATA SAMPLES... 228
FLIGHT LINE.. CRN BLT LO FL210	DATE DATA TAKEN...... 8/13/71	TIME DATA TAKEN.... 1202 HOURS
PLATFORM ALTITUDE. 5000 FEET	GROUND HEADING.... 180 DEGREES	REFORMATTING DATE.AUG 16,1971
	NUMBER OF LINES......... 1161	

	SPECTRAL BAND		CALIBRATION PULSE VALUES		
CHANNEL	LOWER	UPPER	CO	C1	C2
1	0.46	0.49	26.25	178.25	207.55
2	0.48	0.51	23.10	144.20	163.30
3	0.50	0.54	26.80	132.20	159.00
4	0.52	0.57	22.95	166.10	209.10
5	0.54	0.60	26.25	179.70	209.00
6	0.58	0.65	24.00	186.55	203.80
7	0.61	0.70	28.00	168.15	205.50
8	0.72	0.92	24.85	118.55	69.60
9	1.00	1.40	22.90	31.30	56.70
10	1.50	1.80	24.90	74.85	135.25
11	2.00	2.60	25.95	94.05	140.60
12	9.30	11.70	20.60	137.00	0.0

ID-4

Figure 4. ID Record Listing from a Tape

LINEGRAPH

The Linegraph function graphs data values for requested lines from the Multispectral Image Storage Tape. LINEGRAPH and COLUMNGRAPH (which graphs columns) both graph the data values as a function of spatial position. The graphs also contain standard header and identification information. The use of these two functions will aid in data quality analysis studies.

Input

The required inputs are a Multispectral Image Storage Tape and control cards. The PRINT control card identifies the specific subset of data to be used from the tape, and the CHANNELS card specifies the channels to be graphed. A sample input deck is illustrated in Figure 1.

Note that in this case the PRINT card in the figure includes a run number. If the 'RUN' parameter is omitted, the Linegraph function uses the last run specified in a previous function.

The PRINT card illustrated in Figure 1 specifies that line 600 of run 71053900 be plotted from columns 50 through 160 and that each column in this interval be plotted. The CHANNELS card in the figure designates channels 7 and 9.

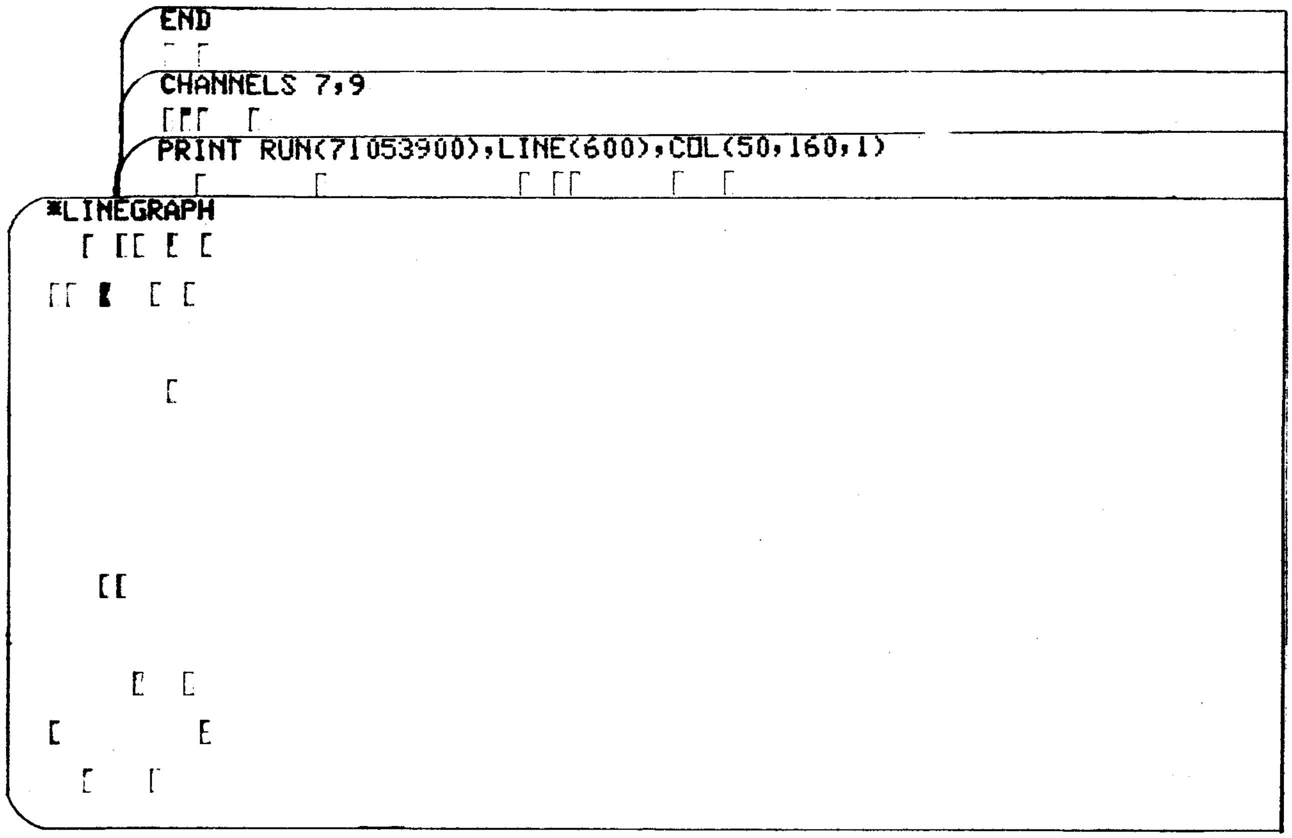


Figure 1. Input Deck for LINEGRAPH

It is possible to request graphs of more than one line but with the same columns and channels by indicating the appropriate starting and stopping line and interval (the interval default is 10) on the PRINT control card.

The SCALE card is an optional control card that controls the "low value" (or beginning point), as well as the increment, for the magnitude axis. The axis normally has a low value of 32.0 and each succeeding point on the axis represents an increment of 2.0. The user may specify other values for these by using the 'XLOW' and 'BINSIZ' parameters on this card.

Output

The output of the Linegraph function is a control card summary and a graph of each line specified. Figures 2 and 3 show the outputs that were produced by the example input deck in Figure 1. Note that since the SCALE card was omitted from the input deck, the graph in Figure 3 uses the default values of 32.0 and 2.0 for the low value and increment of the magnitude axis.

The magnitude of each data value for each channel will normally be shown on the graph by its assigned alphanumeric character. However, if two or more channels have the same value for a particular column, only the highest numbered channel will be shown. This is illustrated in Figure 3 by the plot for column 151, which shows only the symbol for channel 9.

A description of data on the Multispectral Image Storage Tape is contained in Appendix IV.

```
DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING          MAR  7,1973
PHILLIPS                            PURDUE UNIVERSITY                          11 11 43 AM
                                                                               LARSYS VERSION 3

*LINEGRAPH
PRINT RUN(71053900),LINE(600),COL(50,160,1)
CHANNELS 7,9
END

                                                                                  LIN-2
```

Figure 2. Control Card Summary for LINEGRAPH

DEMO
PHIL

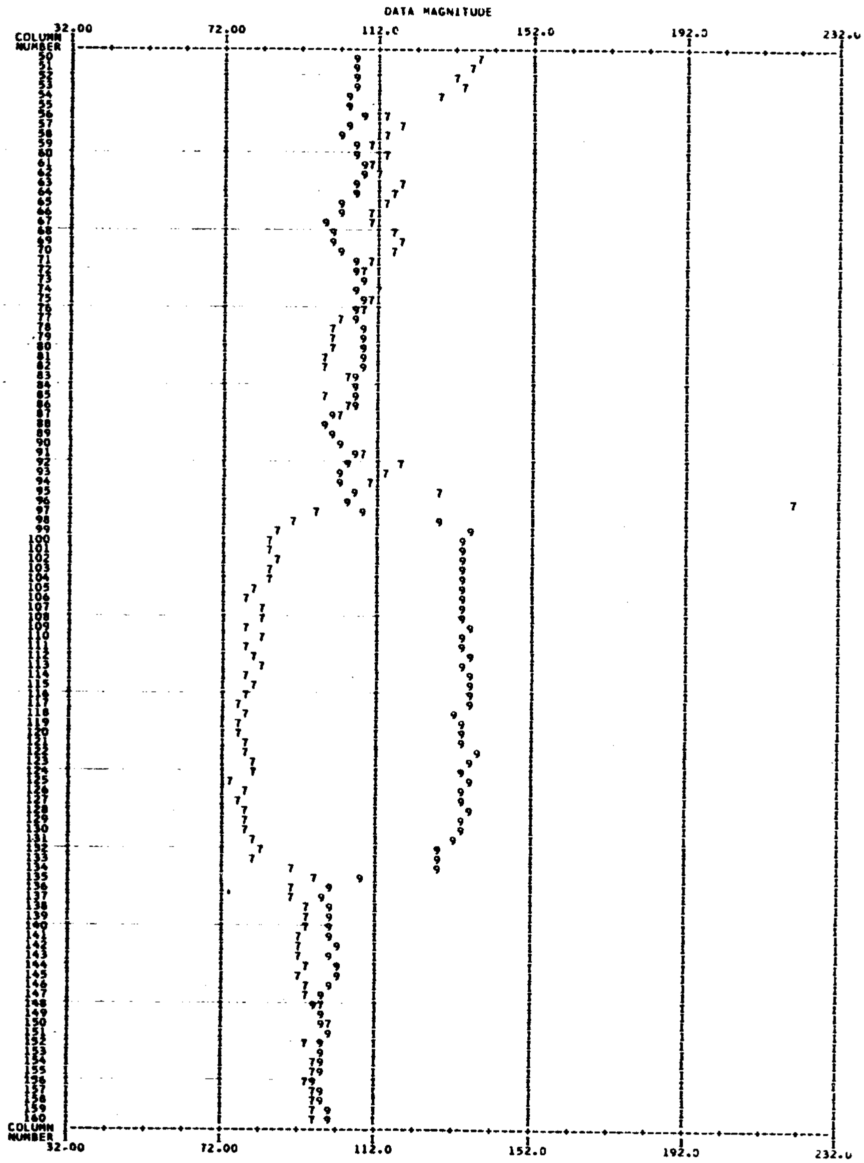
LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 27,1974
3 23 20 PM
LARSYS VERSION 3

****** GRAPH OF LINE 600 ******

RUN NUMBER............. 71053900	DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CRN BLT LO FL210	TIME DATA TAKEN..... 1202 HOURS
DATA TAPE/FILE NUMBER.. 1042/ 1	PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971	GROUND HEADING..... 180 DEGREES

CHANNEL 7 SPECTRAL BAND 0.61 TO 0.70 MICROMETERS DISPLAYED AS..7 CALCODE = 1 CO = 28.00
CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS DISPLAYED AS..9 CALCODE = 1 CO = 22.90



CALIBRATION FOR LINE 600

CHANNEL	C0 MEAN	C0 VARIANCE	C1 MEAN	C1 VARIANCE	C2 MEAN	C2 VARIANCE
7	28.0	1.0	167.0	250.0	210.0	101.0
9	22.9	1.0	30.9	80.0	56.9	118.0

ROLL WAS NOT CALCULATED FOR THIS DATA.

Figure 3.

LISTRESULTS

The Listresults function produces a listing that summarizes the contents of a tape containing Classification Results Files. The information may be extracted from any results file that resides on tape.

Input

Input to the function consists of a tape containing Classification Results Files and the function input deck. The results file must be on tape, not on disk. Also, a scratch tape (TAPE=0) is invalid as input. The 'FROM' card is the required function control card that identifies the tape and file numbers of the results file being specified. If the file number is not included on the 'FROM' card, all files on the tape will be listed.

A sample input deck is shown in Figure 1. The FROM card requests tape 102 and file 1 for the listing.

Output

The Listresults function prints a list of the control cards entered and a table of how the cards are interpreted. An example of this output is shown in Figure 2.

The output is the Classification Results File ID listing. The listing consists of the tape and file numbers, date of the classification, channels used in the classification, the class names, the class weights if they were assigned, the area classified, calibration values, and the number of lines classified. An example of this output is shown in Figure 3.

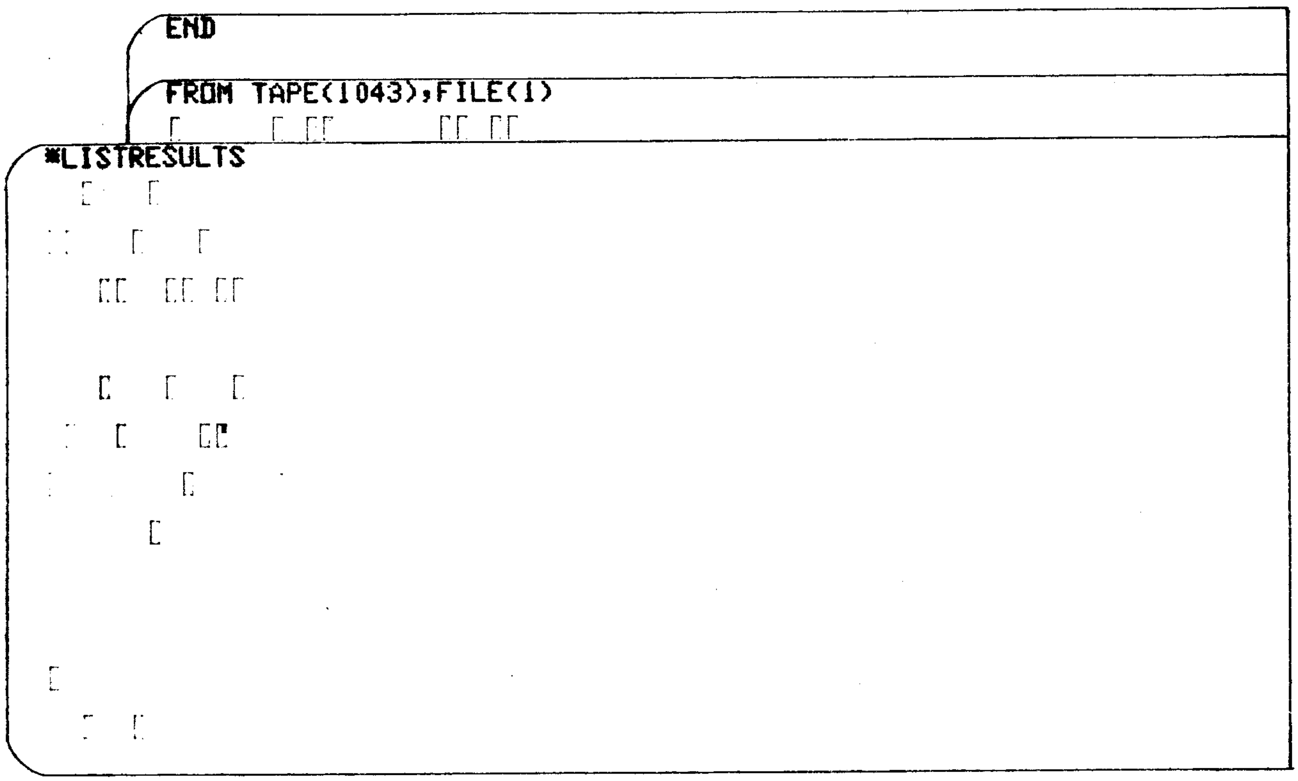


Figure 1. Input Deck for LISTRESULTS

```
     DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING            MAR   7,1973
     PHILLIPS                          PURDUE UNIVERSITY                              12 19 57 PM
                                                                                      LARSYS VERSION 3

    *LISTRESULTS
    FROM TAPE(1043),FILE(1)
    END

  YOU HAVE SELECTED THE FOLLOWING OPTIONS
    TAPE 1043 FILE  1 WILL BE USED

                                                                                            LIS-2
```

Figure 2. Control Card Summary for LISTRESULTS

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
2 56 29 PM
LARSYS VERSION 3

TAPE-- 1043 FILE-- 1 CLASSIFIED- AUG 30,1974

CHANNELS USED

CHANNEL 2	SPECTRAL BAND	0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1
CHANNEL 9	SPECTRAL BAND	1.00 TO 1.40 MICROMETERS	CALIBRATION CODE = 1
CHANNEL 12	SPECTRAL BAND	9.30 TO11.70 MICROMETERS	CALIBRATION CODE = 1

CLASSES AND WEIGHTS

CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT
CORN 0.290	SOYBEANS 0.310	OAT/WHT 0.095	HAY 0.110	FOREST 0.195		

RUN	AREA CLASSIFIED	FLIGHT LINE
71053900	LINES---- 425 TO 644 BY 2 COLUMNS-- 1 TO 222 BY 2	CRN BLT LO FL210

CALIBRATION VALUES

CHANNEL 2 CO = 23.10
CHANNEL 9 CO = 22.90
CHANNEL 12 CO = 20.60
NUMBER OF LINES CLASSIFIED IS 110

Figure 3: Sample File Listing from LISTRESULTS

MERGESTATISTICS

The MERGESTATISTICS function mathematically combines two or more LARSFRIS Statistics Files into a single Statistics File. It can also be used to modify (delete or pool) individual classes within a file. The newly created Statistics File is stored on the user's temporary disk. It contains statistical descriptions of the modified set of training classes, specifically, the mean value of each class in each channel, the covariance matrix, and the number of data points in each of the classes.

This Statistics File may be used in LARSFRIS in the same way as the Statistics Files produced by the *STATISTICS function. It may be used as input to such LARSFRIS functions as SEPARABILITY, CLASSIFYPOINTS, and SAMPLECLASSIFY. It may be used by them directly from the temporary disk. The file is erased, however, if the user either logs off the system (issues the 'QUIT' command) or reinitiates LARSFRIS (issues the 'I LARSYS' command). If the user wishes to save the file longer than this, the 'STATDECK SAVE' command will transfer it to the user's permanent disk. Alternatively, the user may include a control card that directs the function to punch the file in binary or character format.

Additional types of output the user can request are a printed listing of the training fields, the means, standard deviations, correlation matrices, and histograms for every class in the new statistics file, coincident spectral plots for all classes in the file as well as a bi-spectral plot of all classes for any two or four channels.

Input to the MERGESTATISTICS function consists of:

*Control cards to select the processing and output options.

*One or more (up to 30) LARSFRIS statistics files.

The MERGESTATISTICS function always produces 4 standard printed outputs:

*A listing of the control cards that were supplied by the user.

*A list of options selected by the user.

*A list of all classes and pools selected by the user.

*Channel information for the channels included in the modified statistics file.

In addition to the standard printed outputs, the user may specify the following optional printed outputs:

*A list of training fields.

*A summary of the modified statistics deck.

*One coincident spectral plot of all classes.

*Bi-spectral plots. Channel I vs. Channel J or the average of Channels I and J vs. the average of Channels K and L. After each bi-spectral plot there will be printed a table with the following information for each class: class number, symbol, mean values, class name and number of points in each class.

Principal Inputs

The required input to the MERGESTATISTICS function is one or more LARSFRIS statistics file(s) and the control cards that select the function and specify the output options. Figure 1 shows a sample MERGESTATISTICS input deck.

```
*MERGESTATISTICS
 PRINT STATS, MEANS(1,2,7,8)
 PUNCH CHARACTERS, ONEFIELD
 CLASSES ENTIRE(1,2,4), DELETE(3/1,4-9,12/,5/1-5/)
 POOL NEW1(1/1,3-5/,2/1,4/), NEW2(2/2/,3/4-9/)
 SCALE ORIGIN(1,10.0), ORIGIN(7,25.0), UNIT(7,0.5)
 DATA
    STATDECK
 DATA
    STATDECK
 DATA
    STATDECK
 DATA
    STATDECK
 DATA
    STATDECK
 END
```

Figure 1. Sample Input Deck for MERGESTATISTICS

Pooling Algorithm

When two classes are to be pooled, the following operations are performed on their means and covariances.

Let x be a class of n elements with mean M_i^x in channel i and covariance $C_i{}^x{}_j$ between channels i and j. Similarly, let y be another class of m elements with means M_i^y in channel i and covariance $C_i{}^y{}_j$ between channels i and j.

If classes x and y are pooled, the mean of the resulting pooled class is:

$$M_i^{x+y} = \frac{nM_i^x + mM_i^y}{m + n}$$

and the covariance of the resulting pooled class is:

$$C_{ij}^{x+y} = \frac{\left[\sum_{k=1}^{n} x_{ik}x_{jk} + \sum_{k=1}^{m} y_{ik}y_{jk}\right] - \frac{\left(\sum_{k=1}^{n} x_{ik} + \sum_{k=1}^{m} y_{ik}\right)\left(\sum_{k=1}^{n} x_{jk} + \sum_{k=1}^{m} y_{jk}\right)}{m + n}}{m + n - 1}$$

where the covariances of classes x and y are respectively:

$$C_{ij}^{x} = \left[\sum_{k=1}^{n} x_{ik}x_{jk} - \frac{\sum_{k=1}^{n} x_{ik}\sum_{k=1}^{n} x_{jk}}{n}\right] \Big/ n-1$$

$$C_{ij}^{y} = \left[\sum_{k=1}^{m} y_{ik}y_{jk} - \frac{\sum_{k=1}^{m} y_{ik}\sum_{k=1}^{m} y_{jk}}{m}\right] \Big/ m-1$$

In the subroutine REDSAV the calculations are as follows:

$$\begin{aligned} C_{ij}^{(x+y)} &= \left[C_{ij}^{x}(n-1) + nM_i^xM_j^x\right] + \left[C_{ij}^{y}(m-1) + mM_i^yM_j^y\right] \\ &= \left[C_{ij}^{x}(n-1) + n\left(\frac{\sum_{k=1}^{n} x_{ik}}{n}\right)\left(\frac{\sum_{k=1}^{n} x_{jk}}{n}\right)\right] + \left[C_{ij}^{y}(m-1) + m\left(\frac{\sum_{k=1}^{m} y_{ik}}{m}\right)\left(\frac{\sum_{k=1}^{m} y_{jk}}{m}\right)\right] \\ &= \sum_{k=1}^{n} x_{ik}x_{jk} + \sum_{k=1}^{m} y_{ik}y_{jk} \end{aligned}$$

where $M_i^x = \frac{\sum_{k=1}^{n} x_{ik}}{n}$ and similarly for M_j^x, M_i^y, and M_j^y.

and then

$$C_{ij}^{x+y} = C_{ij}^{(x+y)} - (m+n)\left(\frac{\sum_{k=1}^{n} x_{ik} + \sum_{k=1}^{m} y_{ik}}{m + n}\right)\left(\frac{\sum_{k=1}^{n} x_{jk} + \sum_{k=1}^{m} y_{jk}}{m + n}\right) \Big/ m+n-1$$

Statistics File

The Statistics File is always written onto the user's temporary disk and may be punched on cards or written on a permanent disk. The file may be input in either form to LARSFRIS functions such as the SEPARABILITY, CLASSIFYPOINTS and SAMPLECLASSIFY. The file contains the following information:

*Training field descriptions supplied by input files.

*The number of classes, number of fields and number of channels.

*The channel number, wavelength of the spectral band and calibration code for each channel used in the input files.

*The number of points in each class in the modified file.

*The mean for each channel for each class in the modified file.

*The covariance matrix of all specified channels for each class in the modified file.

As in the Statistics File produced by the STATISTICS processing function, the first and last cards are easily identified. The first card is:

LARSYS VERSION3.1 STATISTICS DECK

The last card is:

EOS ***** LAST CARD OF STATISTICS DECK *** nnnnnnnn

The n's are a sequence number in columns 73-80.

The Statistics File can be punched on cards by including a PUNCH control card in the input deck, as shown in Figure 1. Note that in the example, the CHARACTER control parameter requests that the deck be punched in character form instead of binary form. The character form deck requires a significantly greater number of cards than the binary (hexadecimal) form to contain the same information; however, the binary numbers are not as easily understood by the user.

The option to punch a single dummy Field Description Card (instead of all field description cards contained in input Statistics Files) is activated by the control parameter 'ONEFIELD'.

PICTUREPRINT

The Pictureprint function reads data from the Multispectral Image Storage Tape and produces alphanumeric pictorial printouts of the data for each channel that is specified. The pictorial printout is normally used to check data quality and to select subsets (rectangular fields) of the data that will be used as Test and Training fields for other functions, usually STATISTICS, SAMPLE-CLASSIFY, or PRINTRESULTS. The fields are identified by comparing a photograph to the Pictureprint pictorial printout and selecting the line and column (sample) numbers that define the field corners. These line and column numbers are then keypunched onto Field Description Cards for use as input to the other functions. PICTUREPRINT also has the capability of reading Field Description Cards and printing the field boundaries on the pictorial printout so the user can check his field selections.

The pictorial map uses alphanumeric symbols to simulate gray scale tones. Experience has shown that 10 symbols, each of which has an equal probability of appearing, best accomplishes this purpose. PICTUREPRINT allows for the use of from 2 to 16 symbols. The symbols may be taken from a pre-defined set or may be specified by the user. The pre-defined symbols programmed into PICTUREPRINT are included in the description of the SYMBOLS control card in the Control Card Dictionary (Appendix I). The symbols may be automatically assigned to a range of data values based on "equal likelihood" of appearance, or the ranges of data values may be specified by the user.

Unless instructed otherwise, the function will automatically use a subset of 10 pre-defined symbols and will assign these ten symbols to "equally active" ranges of data. It accomplishes this as follows:

- It histograms the data for each channel that is to be mapped, or uses a histogram that has been previously stored on disk or read in on cards.
- It integrates the histogram to find equal distribution cut-offs for the ten symbols that are to appear on the map.
- It then assigns each of the ten ranges to one of the ten symbols in order of increasing magnitude.

If the user wishes to vary this procedure, he has three options at his disposal. He may specify alternate symbols to be used instead of the pre-defined set; he may specify a different number of symbols to be used; and/or he may assign the actual data ranges to be used for each symbol. If he specifies the symbols to be used (on a SYMBOLS control card), they replace the function's pre-defined symbols, and the number that is used will be the number that he supplies. The data for each channel will still be histogrammed, and the histograms will be partitioned so that each of these symbols has an equal probability of occurring.

The user may specify that a specific number of the pre-defined symbols be used on the map. He accomplishes this through the 'NLEV' parameter on a SYMBOLS card. He may specify from 2 to 16 symbols. The data will again be histogrammed and the symbols assigned as above.

If the user wishes to deviate from the "equal likelihood" assignment of symbols, he may do so by including a HISTOGRAM LEVELSCARDS control card and levels data cards in the input deck. The levels cards specify, for each channel, the data ranges that are to be used in assigning the symbols. The symbols that are assigned will be either the pre-defined symbols that PICTUREPRINT normally uses or symbols that the user has specified on a SYMBOLS control card.

Figure 1 shows a sample input deck for the Pictureprint function. Included in the deck is a SYMBOLS control card to specify a particular set of symbols to appear on the pictorial map. Since user-defined levels (HISTOGRAM LEVELSCARDS) are not specified, the 11 symbols defined on the SYMBOLS card (the 10th and 11th are blanks) will be assigned to 11 equally active ranges of data values. The pictorial map in Figure 2 illustrates how this information is documented in the map header.

```
END
71053900 SEG210     425 642  1  25 172  1SEG210  TOTAL TEST
71053900 H*3/4      618 640  1 139 169  1OTHER   WHT/HAY
71053900 I-K-123    573 585  1  25  53  1OTHER   SEV FLDS
71053900 G*1        537 556  1 145 165  1OTHER   HAY
71053900 F*1/2/6    537 547  1 104 139  1OTHER   SA/NF
71053900 D*4        450 475  1 128 142  1OTHER   SET ASIDE/NON
71053900 E*2        427 492  1 149 166  1OTHER   WDS/PAS
71053900 L-2        592 600  1  25  94  1PASTURE
71053900 L-4        634 641  1  68  94  1WOODS
71053900 D-3        525 529  1 122 140  1WOODS
71053900 C-6        428 447  1  30  33  1OATS
71053900 L-3        634 642  1  23  60  1SOYBEANS
71053900 H-5        580 585  1  59  65  1SOYBEANS
71053900 H-4        577 584  1  83  93  1SOYBEANS
71053900 F*5        569 584  1 102 137  1SOYBEANS
71053900 H-3        555 564  1  78  94  1SOYBEANS
71053900 D-2        523 530  1 104 116  1SOYBEANS
71053900 R-3        520 529  1  23  49  1SOYBEANS
71053900 E*3        505 515  1 147 169  1SOYBEANS
71053900 D-2        481 518  1 106 140  1SOYBEANS
71053900 F-10       479 507  1  63  98  1SOYBEANS
71053900 C-7        451 475  1  38  58  1SOYBEANS
71053900 D*1        426 443  1 109 139  1SOYBEANS
71053900 F*3        554 564  1 102 139  1CORN
71053900 K-5        540 564  1  26  29  1CORN
71053900 R-11       517 529  1  62  98  1CORN
71053900 R-2        502 514  1  31  48  1CORN
71053900 D*3        454 475  1 106 122  1CORN
71053900 F-8        425 456  1  65  99  1CORN
71053900 C-5        428 445  1  38  59  1CORN
TEST
71053900 E*2        523 528  1 147 197  1WOODS
71053900 H-6        543 562  1  63  72  1WOODS
71053900 E*1        426 441  1 179 193  1OTHER   SET ASIDE
71053900 G*3        562 582  1 156 164  1WHEAT
71053900 R-1        482 492  1  40  56  1HAY
71053900 G*2        565 582  1 144 149  1OATS
71053900 K-4        541 559  1  35  38  1OATS
71053900 F-9        461 473  1  64  98  1OATS
71053900 H*2        591 593  1 137 170  1SOYBEANS
71053900 D*6        592 612  1 101 131  1SOYBEANS
71053900 L-8        606 614  1  58  94  1SOYBEANS
71053900 L-5        619 628  1  23  29  1SOYBEANS
71053900 H*1        598 612  1 138 169  1CORN
71053900 D*5        618 641  1 100 132  1CORN
71053900 L-7        619 628  1  58  94  1CORN
71053900 L-6        610 628  1  35  52  1CORN
CLASS
DATA
SYMBOLS M,$,X,Z,8,I,/,1,-, ,
PUNCH HIST
PRINT HIST
CHANNELS 9
BLOCK LINE(420,644,1),COL(1,221,1)
DISPLAY RUN(71053900),LINE(420,644,2),COL(1,221,2)
BOUNDARY DELETE,STORE,OUTLINE
*PICTUREPRINT
```

Figure 1. Input Deck for PICTUREPRINT

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 7,1974
2 29 15 PM
LARSYS VERSION 3

RUN NUMBER............ 71053900	DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CRN BLT LD FL210	TIME DATA TAKEN..... 1202 HOURS
DATA TAPE/FILE NUMBER.. 1005/ 2	PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971	GROUND HEADING..... 180 DEGREES

CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS CALIBRATION CODE= 1 C0 = 22.90

THE CHARACTER SET USED FOR DISPLAY IS

```
BELOW  58.4              DISPLAYED AS M
FROM   58.4  TO   90.4   DISPLAYED AS M
FROM   90.4  TO   94.4   DISPLAYED AS $
FROM   94.4  TO   98.4   DISPLAYED AS X
FROM   98.4  TO  100.4   DISPLAYED AS Z
FROM  100.4  TO  102.4   DISPLAYED AS 8
FROM  102.4  TO  104.4   DISPLAYED AS I
FROM  104.4  TO  108.4   DISPLAYED AS /
FROM  108.4  TO  112.4   DISPLAYED AS 1
FROM  112.4  TO  120.4   DISPLAYED AS -
FROM  120.4  TO  126.4   DISPLAYED AS
FROM  126.4  TO  168.4   DISPLAYED AS
ABOVE            168.4   DISPLAYED AS
TRAIN FIELD BOUNDARIES MARKED WITH .
TEST  FIELD BOUNDARIES MARKED WITH ;
SHARED  FLD BOUNDARIES MARKED WITH '
```

HISTOGRAM BLOCK(S)

```
RUN NUMBER........... 71053900
LINES....... (  420,  644,  1)
COLUMNS..... (    1,  221,  1)
CALIBRATION CODE........... 1
```

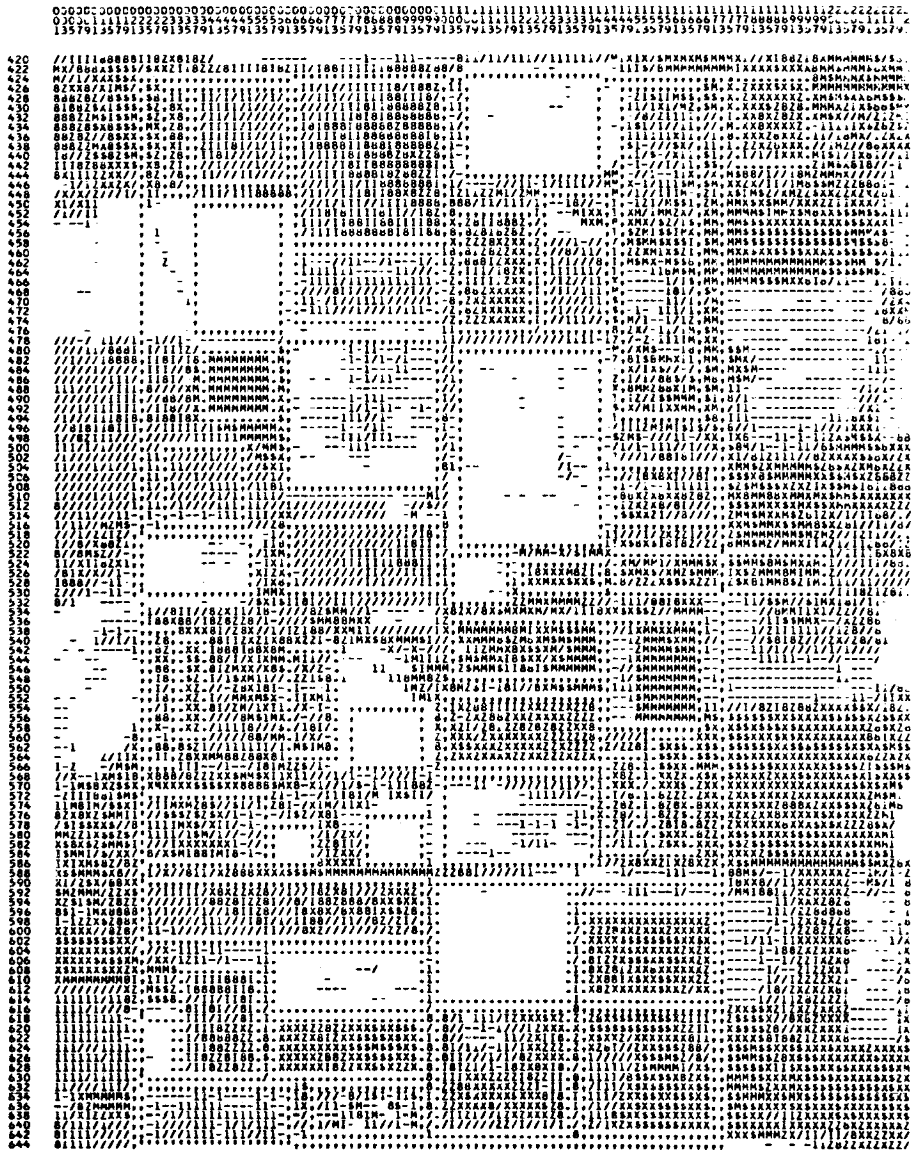


Figure 2. Pictorial Map

PRINT is the primary data quality and image editing function for the user. Once the data quality is approved and Field Description Cards are produced, the user can proceed to the pattern recognition functions of STATISTICS, SEPARABILITY, CLASSIFYPOINTS, SAMPLECLASSIFY and CLUSTER.

Input Files

The primary input data files are the Multispectral Image Storage Tapes. Pictureprint uses data from these tapes for two purposes: to compute histograms that will be used to assign "equally active" data ranges to symbols, and to display a pictorial map of the data based on some symbol assignment. The data that is used for histogramming need not be from the same run that is being mapped, and if it is from the same run, it need not be the same data samples.

Two control cards specify the data that is to be used for these two purposes. The DISPLAY card defines the area that is to be mapped, and the BLOCK card the data that is to be histogrammed.

A DISPLAY card containing a valid run number is always required. In addition, the user must specify the line and column numbers, and their respective intervals, or else allow the default values for these parameters to be used. The standard ways of specifying line and column information are to specify on the DISPLAY card the LINE(x,y,z),COL(x,y,z) information (where x equals the first line or column, y the last, and z the interval), or to specify a particular width with the 'WIDTH(x)' parameter. The 'WIDTH' parameter will cause the column interval to be adjusted to provide a pictorial map as wide in pages as the number specified. There are a number of combinations of default values for line and column information, and the user should refer to the description of the DISPLAY card in the Control Card Dictionary in Appendix I for complete information.

A BLOCK control card is used only if the user wishes to compute histograms but does not wish to follow the standard procedure for selecting the data to be histogrammed. The standard procedure (if the BLOCK card is omitted), is to select the data to be histogrammed from the same data run that is specified on the DISPLAY card, and to take the samples from every tenth column of every tenth line of the area specified on that card.

The sample input deck in Figure 1 illustrates the use of these two control cards. The BLOCK card was included in this case in order to histogram samples from every line and column rather than every tenth line and column.

Specification of Histogramming

The user controls the selection of histograms that will be used to assign symbols by using the HISTOGRAM control card. There are four possible choices in this area:

- Histograms may be calculated from the data area that is specified on the DISPLAY (or BLOCK) card
- Histograms that were previously stored on the user's A-disk may be used
- Histograms may be read in card form
- The user may assign his own data ranges for each symbol.

These alternatives are implemented, respectively, through the 'COMPUTE', 'DISK', 'HISTOCARDS', and 'LEVELSCARDS' parameters on the HISTOGRAM card. If this control card is omitted, PICTUREPRINT assumes the 'COMPUTE' parameter. If the user specifies 'DISK', he may use either a set of histograms that were calculated in his current terminal session, or he may (if he has issued the 'HISTOGRAM USE' command prior to activating the Pictureprint function) use a set of histograms that were stored on his A-disk. The sample input deck in Figure 1 does not include a HISTOGRAM card, so the histograms will be calculated.

Specification of Channels

The channels to be printed out by this function are usually specified by using a CHANNELS card. The CHANNELS card is required unless 'HISTOGRAM LEVELSCARDS' is specified, which requires the channels be identified on the Levels data cards. A separate pictorial printout and histogram (optional) are generated for each channel. The CHANNELS card in Figure 1 specifies that channel 9 should be used. For more complete information on the CHANNELS card refer to Appendix I.

Specification of Boundaries

The user may enter Field Description Cards to define boundaries that are to appear on the map. LARSFRIS maintains a disk file (FIELD BNDRIES) which stores these boundaries for the duration of a terminal session. The file is empty at Login time, or after issuing an 'i larsys' command.

The user may control the use of these field boundaries with the BOUNDARY control card. It has three parameters:

- 'DELETE' to delete all stored boundary information
- 'STORE' to read Field Description Cards and store the boundary information on the disk
- 'OUTLINE' to outline stored boundaries on the pictorial printout.

The sample BOUNDARY card in Figure 1 requests 'DELETE', 'STORE' and 'OUTLINE'. This will cause PICTUREPRINT to read the Field Description Cards that are included in the deck behind the DATA card, to store the boundary information on the disk, and to outline the fields on the pictorial printout. The printout in Figure 2 shows how these fields are outlines.

Output Files

When PICTUREPRINT calculates a histogram, it writes the results on the Histogram File on the temporary disk. This file can be used as input by HISTOGRAM or GRAPHHISTOGRAM, and may be saved for future use with the 'HISTOGRAM SAVE' command. The user may also get the results punched into card form by using a 'PUNCH HIST' card as shown in Figure 1. This deck is useful for histograms that should be saved beyond the current terminal session. The cards may then be used later for input to PICTUREPRINT.

The user should check the punched deck to ensure that he has a complete Histogram Deck. The first card should be "LARS HISTO", and the last card should be

"EOH ***** LAST DATA CARD OF HISTOGRAM DECK *****"

The text on each card begins in column 1.

Standard Printer Output

The Pictureprint function prints a summary of the user's input deck. This includes a list of the input control cards, a list of the options selected, and other particular characteristics about the run, such as the number of gray levels (symbols) selected. An example of this output is shown in Figure 3.

In addition, the function will print one pictorial printout for each channel that was specified. A pictorial map that was created by the sample deck for channel 9 is shown in Figure 2. The printout includes information about the data run and the channel, the character set (symbols) used for display, the data that was histogrammed, and then the map itself.

Optional Printer Output

There are three optional printer outputs that are produced based on the use of the HISTOGRAM, BOUNDARY, and PRINT cards.

Histogram Summary

This summary shows the data that was histogrammed and the following statistics for each channel histogrammed: data range, mean, standard deviation and the normalized range (the mean plus and minus three standard deviations). The output is produced only if the HISTOGRAM COMPUTE card is supplied. Figure 4 shows a sample of this output. This output can also be produced by HISTOGRAM.

```
    DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING           MAR  14,1973
    PHILLIPS                             PURDUE UNIVERSITY                            8 50 56 PM
                                                                                  LARSYS VERSION 3

   *PICTUREPRINT
   BOUNDARY DELETE,STORE,OUTLINE
   DISPLAY RUN(71053900),LINE(420,644,2),COL(1,221,2)
   BLOCK LINE(420,644,1),COL(1,221,1)
   CHANNELS 9
   PRINT HIST
   PUNCH HIST
   SYMBOLS M,$,X,Z,8,I,/,1,-, ,
   DATA

YOU HAVE SELECTED THE FOLLOWING PICTORIAL PRINT OPTIONS
   CALCULATE A NEW HISTOGRAM FOR SETTING GRAY LEVELS
   PUNCH HISTOGRAMS
   DELETE ALL PREVIOUSLY STORED FIELD BOUNDARIES
   READ AND STORE MORE FIELD BOUNDARIES
   OUTLINE ALL FIELDS FOR WHICH BOUNDARIES HAVE BEEN STORED

PICTUREPRINT INFORMATION
   NUMBER OF CHANNELS...  1
   THE NUMBER OF GRAY LEVELS SELECTED... 11

                                                                                   PIC-3
```

Figure 3. Control Card Summary for PICTUREPRINT

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
2 50 44 PM
LARSYS VERSION 3

RUN NUMBER............ 71053900	DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CRN BLT LO FL210	TIME DATA TAKEN..... 1202 HOURS
DATA TAPE/FILE NUMBER.. 1005/ 2	PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971	GROUND HEADING..... 180 DEGREES

CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS CALIBRATION CODE = 1 CO = 22.90

DATA BLOCK(S) HISTOGRAMMED

CHANNEL	RUN NUMBER	CALIBRATION CODE	FIRST LINE	LAST LINE	LINE INTERVAL	FIRST SAMPLE	LAST SAMPLE	SAMPLE INTERVAL
9	71053900	1	420	644	1	1	221	1

HISTOGRAM STATISTICS

CHANNEL	DATA RANGE	MEAN	STANDARD DEVIATION	NORMALIZED RANGE (MEAN + AND - 3 STD DEV)
9	58.4 TO 168.	107.	13.2	67.4 TO 147.

Figure 4: Histogram Summary Listing

Field Cards Stored

This output is produced when the 'BOUNDARY STORE' card is used. It provides a list of the fields added to the disk file during the current function. An example is shown in Figure 5.

Histogram Graphs

These graphs are produced when the 'PRINT HIST' card is used. One graph per channel histogrammed is printed. The graphs are identical to those produced by GRAPHHISTOGRAM. Figure 6 shows an example of the histogram for channel 9.

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 14,1973
8 53 48 PM
LARSYS VERSION 3

FIELD CARDS ADDED TO FIELD BOUNDARY STORAGE

FIELD NUMBER	RUN NUMBER	FIELD DESIGNATION	FIRST LINE	LAST LINE	LINE INTERVAL	FIRST SAMPLE	LAST SAMPLE	SAMPLE INTERVAL	FIELD TYPE	OTHER INFORMATION
1 *	71053900	L-6	610	628	1	35	52	1	CORN	
2 *	71053900	L-7	619	628	1	58	94	1	CORN	
3 *	71053900	D*5	618	641	1	100	132	1	CORN	
4 *	71053900	H*1	598	612	1	136	169	1	CORN	
5 *	71053900	L-5	619	628	1	23	29	1	SOYBEANS	
6 *	71053900	L-8	606	614	1	58	94	1	SOYBEANS	
7 *	71053900	D*6	592	612	1	101	131	1	SOYBEANS	
8 *	71053900	H*2	591	593	1	137	170	1	SOYBEANS	
9 *	71053900	F-9	461	473	1	64	98	1	OATS	
10 *	71053900	K-4	541	559	1	35	38	1	OATS	
11 *	71053900	G*2	565	582	1	144	149	1	OATS	
12 *	71053900	R-1	482	492	1	40	56	1	HAY	
13 *	71053900	G*3	562	582	1	156	164	1	WHEAT	
14 *	71053900	E*1	426	441	1	179	193	1	OTHER	SET ASIDE
15 *	71053900	H-6	543	562	1	63	72	1	WOODS	
16 *	71053900	E*2	523	528	1	147	197	1	WOODS	
17	71053900	C-5	428	445	1	38	59	1	CORN	
18	71053900	F-8	425	456	1	65	99	1	CORN	
19	71053900	D*3	454	475	1	106	122	1	CORN	
20	71053900	R-2	502	514	1	31	48	1	CORN	
21	71053900	R-11	517	529	1	62	98	1	CORN	
22	71053900	K-5	540	564	1	26	29	1	CORN	
23	71053900	F*3	554	564	1	102	139	1	CORN	
24	71053900	D*1	426	443	1	109	139	1	SOYBEANS	
25	71053900	C-7	451	475	1	38	58	1	SOYBEANS	
26	71053900	F-10	479	507	1	63	98	1	SOYBEANS	
27	71053900	O-2	481	518	1	106	140	1	SOYBEANS	
28	71053900	E*3	505	515	1	147	169	1	SOYBEANS	
29	71053900	R-3	520	529	1	23	49	1	SOYBEANS	
30	71053900	O-2	523	530	1	104	116	1	SOYBEANS	
31	71053900	H-3	555	564	1	78	94	1	SOYBEANS	
32	71053900	F*5	569	584	1	102	137	1	SOYBEANS	
33	71053900	H-4	577	584	1	83	93	1	SOYBEANS	
34	71053900	H-5	580	585	1	59	65	1	SOYBEANS	
35	71053900	L-3	634	642	1	23	60	1	SOYBEANS	
36	71053900	C-6	428	447	1	30	33	1	OATS	
37	71053900	O-3	525	529	1	122	140	1	WOODS	
38	71053900	L-4	634	641	1	68	94	1	WOODS	
39	71053900	L-2	592	600	1	25	94	1	PASTURE	
40	71053900	E*2	427	492	1	149	166	1	OTHER	WDS/PAS
41	71053900	D*4	450	475	1	128	142	1	OTHER	SET ASIDE/NUN
42	71053900	F*1/2/6	537	547	1	104	139	1	OTHER	SA/NF
43	71053900	G*1	537	556	1	145	165	1	OTHER	HAY
44	71053900	J-K-123	573	585	1	25	53	1	OTHER	SEV FLDS
45	71053900	H*3/4	618	640	1	139	169	1	OTHER	WHT/HAY
46	71053900	SEG210	425	642	1	25	172	1	SEG210	TOTAL TEST

* - INDICATES TRAINING FIELDS

Figure 5. Test and Training Fields Listing

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

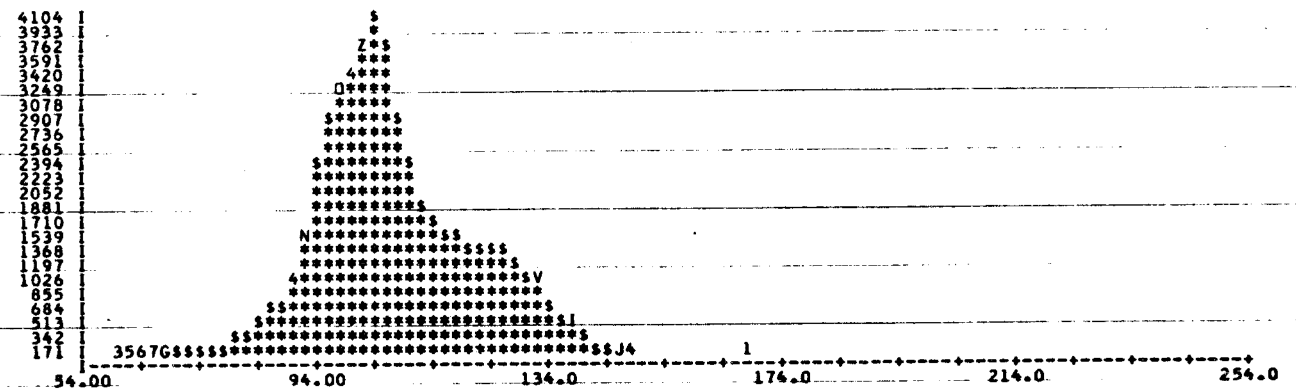
MAR 14,1973
8 53 50 PM
LARSYS VERSION 3

****** HISTOGRAM GRAPH FOR CHANNEL 9 ******

SPECTRAL BAND 1.00 TO 1.40 MICROMETERS

RUN... 71053900 CALIBRATION CODE... 1 LINES... (420, 644, 1) COLUMNS... (1, 221, 1)

EACH * REPRESENTS... 171 SAMPLES. TOTAL NUMBER OF SAMPLES... 49725



PIC-6

Figure 6. Histogram Graph for Channel 9

PRINTRESULTS

The PRINTRESULTS function provides a variety of printed outputs describing the classification results produced by the CLASSIFY-POINTS function. The function provides the researcher with a flexible capability to display the results of a classification in the form of a map image and/or tabular outputs.

Input to the function consists of the Classification Results File and an input deck of control cards and (in some cases) data cards. Four types of output are produced:

- A map image of the classified area composed of printed symbols representing the selected classes, or ranges of the coded discriminant values.
- Test and training fields listings to describe these fields.
- Performance tables for the test and training fields and test and training classes to show how points belonging to these fields and classes were actually classified.
- A summary listing of the statistical data used in the classification.

All performance tables and the statistics summary are optional and must be requested by the user. In addition, certain other outputs may be suppressed, or additional copies of all of the above may be requested.

Options are available to control the form and content of the output:

- Different classes may be grouped into a common class for performance calculations.
- A subset of the classified area may be designated for processing.
- The user may assign any symbols available to represent the classes and groups that are shown on the map.
- The user may apply threshold values to points on the map to eliminate those points with a low probability associated with the class to which they were classified. Points which are "thresholded" are not printed on the map. Any or all classes may have thresholding applied.
- The map can be created based on the coded discriminant values recorded on the classification results file.

Figure 1 shows a typical Printresults input deck. This deck will display results of the Classification Results File created by the example classification used in the description of the Classify-points function.

Required Control Cards

The RESULTS card specifies the location of the Classification Results File that is to be used for input, and the SYMBOLS card specifies the symbols (alphabetic, numeric, or punctuation) that are to be used on the map to represent the classes.

```
END
71053900  H*3/4      618  640   1  139  169   1OTHER   WHT/HAY
71053900  G*1        537  556   1  145  165   1OTHER   HAY
71053900  F*1/2/6    537  547   1  104  139   1OTHER   SA/NF
71053900  I-K-123    573  585   1   25   53   1OTHER   SEV FLDS
71053900  E*2        427  492   1  149  166   1OTHER   WDS/PAS
71053900  D*4        450  475   1  128  142   1OTHER   SET ASIDE/NON
71053900  E*1        426  441   1  179  193   1OTHER   SET ASIDE
71053900  L-4        634  641   1   68   94   1WOODS
71053900  O-3        525  529   1  122  140   1WOODS
71053900  C-6        428  447   1   30   33   1OATS
71053900  L-2        592  600   1   25   94   1PASTURE
TEST 3
71053900  R-3        520  529   1   23   49   1SOYBEANS
71053900  E*3        505  515   1  147  169   1SOYBEANS
71053900  O-2        523  530   1  104  116   1SOYBEANS
71053900  O-2        481  518   1  106  140   1SOYBEANS
71053900  F-10       479  507   1   63   98   1SOYBEANS
71053900  C-7        451  475   1   38   58   1SOYBEANS
71053900  D*1        426  443   1  109  139   1SOYBEANS
71053900  H-3        555  564   1   78   94   1SOYBEANS
71053900  F*5        569  584   1  102  137   1SOYBEANS
71053900  H-4        577  584   1   83   93   1SOYBEANS
71053900  H-5        580  585   1   59   65   1SOYBEANS
71053900  L-3        634  642   1   23   60   1SOYBEANS
TEST 2
71053900  F*3        554  564   1  102  139   1CORN
71053900  K-5        540  564   1   26   29   1CORN
71053900  R-11       517  529   1   62   98   1CORN
71053900  R-2        502  514   1   31   48   1CORN
71053900  D*3        454  475   1  106  122   1CORN
71053900  F-8        425  456   1   65   99   1CORN
71053900  C-5        428  445   1   38   59   1CORN
TEST 1
DATA
BLOCK RUN(71053900),CALC
GROUP CORN(1/1/),SOYBEANS(2/2/),OTHER(3/3,4,5/)
THRESHOLD 2*0.1,3*0.0
SYMBOLS C,S,O,H,F
PRINT TRAIN(F,C),TEST(F,C,P)
PRINT STATS,MAPS(2),OUTLINE(TRAIN,TEST)
RESULTS DISK
*PRINTRESULTS
```

Figure 1. Input Deck for PRINTRESULTS

RESULTS Card

The Classification Results File may reside on tape or disk. If it is on tape, the RESULTS card will specify the tape and file numbers. If PRINTRESULTS immediately follows a Classifypoints run, the results file tape used will still be mounted, but the tape still must be specified on the RESULTS card.

If the file is on a scratch tape, a tape number of '000' is used. Scratch tapes may be used when CLASSIFYPOINTS and PRINTRESULTS are run successively. In this case, the tape remains mounted for the Printresults run, and no further identification of the tape number is needed.

The 'RESULTS DISK' form of the card (as in the example in Figure 1) specifies that the results file is to be taken from disk. The results file from the last run of CLASSIFYPOINTS (subsequent to the 'i larsys' command) will be available on disk.

SYMBOLS Card

The user must designate on the SYMBOLS card the symbols to be used to represent each class on the PRINTRESULTS classification map. These symbols or characters are assigned to every class defined to the Classifypoints function and stored in the results file. The same symbol may be assigned to two more classes, thus grouping them for map-image purposes. The sample deck in the figure specifies the symbols for classes 1 to 5, respectively, to be 'C,S,O,H,F'. The class numbers are those assigned in CLASSIFYPOINTS.

Optional Input

Four optional control cards may be used to control PRINTRESULTS processing. The cards and their purpose are as follows:

- BLOCK -defines a subset of the classified area for mapping and calculations.
- GROUP -groups two or more classes into a single class for performance calculations.
- THRESHOLD -applies thresholding to the printing of points on the classification map.
- PROBABILITY -defines ranges of values related to the probability of correct classification and requests maps and/or tables based on these ranges.
- PSYMBOLS -assigns symbols to the ranges defined above for printing of maps.
- Field Description Cards -specifies test fields.

The use of these cards is described below.

Defining Subsets to be Mapped and Calculated

The user may define a subset of the classified area by using the BLOCK card. This card specifies the run number and the lines and column coordinates for the area. The BLOCK card in Figure 1 specifies that all data points classified from run 71053001 be used for the map and for calculations.

Normally, performance calculations are based on the entire results file. However, by including the 'CALC' parameter on the BLOCK card (as shown in the figure), they will be based on just the area specified on the BLOCK card.

Grouping Classes

A GROUP card is used to group classes for performance calculations. For example, when computing the number of samples classified into each class, the user may wish to consider OAT/WHT, HAY, and FOREST as a single class of OTHER. Grouping done with this card will have no effect on the map output, so all classes still must be assigned symbols.

The sample GROUP card in Figure 1 places the five classes on the results file into three classes for performance calculations. Note that the format is the same as for the CLASSES card in the Classifypoint function (except that the keyword is GROUP instead of CLASSES). However, single classes need not be included on the GROUP card. For instance, 'SOYBEANS(2/2/)' could be omitted, and OTHER designated as '(2/3,4,5/)'. The Printresults function will then automatically assign the remaining class (SOYBEANS) to class 3 and assign the same class name that is provided in the results file.

Applying Threshold Values

Thresholding is the assignment of points to a "null" class by comparing each point's probability of belonging to its assigned class to a "threshold" probability for that class. If its probability of belonging to the assigned class is smaller than the threshold probability, the point is considered to be a member of the null class, i.e., "is thresholded". All thresholded points are left blank on the map-image of the classified area.

Thresholding is requested by a THRESHOLD control card in the input deck. The user must specify on the card a threshold percentage for each class that was defined in the classification. If he does not wish a particular class to have thresholding applied, he must specify a zero threshold value. The threshold percentage is entered on the card as a decimal number that represents, roughly, the percentage of points the user expects to be thresholded in the training fields for that class.

The number of points thresholded is listed in the performance tables, and the thresholded percentages that were used are printed along with the list of channels used and classes considered. It is important to note that points thresholded within training and test fields are counted as classification errors in all performance tables.

The range of threshold values, from 0.0 to 14.9 and from 15 to 99, provides accuracies in the tenths of percentages below 15 and whole percentages above 15. Each value is entered on the card in the same order as the class numbers of the classes to which they apply, i.e., the first value applies to Class 1, the second to Class 2, etc.

As an example, assume that a user wants to specify the following threshold percentages to four classes in the Classification Results File:

Class 1 - 1.5%

Class 2 - 20.0%

Class 3 - 0.1%

Class 4 - no thresholding applied.

The control card will be:

```
THRESHOLD 1.5,20.0,0.1,0.0
```

An optional form may be used to advantage to denote the same threshold percentage for several classes that are adjacent in the positional sequence. For instance, four classes above could all be specified at 5.0% by using:

```
THRESHOLD 4*5.0
```

Class sequences can be grouped with this form also:

```
THRESHOLD 2*33,2*67.0
```

This statement will threshold classes 1 and 2 at 33.0% and classes 3 and 4 at 67.0%.

The THRESHOLD card in Figure 1 specifies that the first two classes are to be thresholded at 0.1% and no thresholding is to be applied to the last three classes.

A more complete discussion of the thresholding algorithm is included in the description fo CLASSIFYPOINTS.

Requesting Maps and/or Tables Based on the Coded Discriminant Values

As discussed in the description of the Classifypoints algorithm, the discriminant values which are calculated to assign a pixel to the most likely class are coded and stored on the Classification Results File. If a PROBABILITY card is included in the input deck to the Printresults processor, then these coded numbers are converted back to the associated percent rejection rate for a Chi-Square distribution. These percentages represent, roughly, the percentage of points which are less likely to belong to that class (in other words, the percentage of points that would be thresholded) in the training fields for that class. This percent is often related to the point's probability of belonging to its assigned class. On the PROBABILITY card, the user may also specify the intervals of these percentages into which the points are to be grouped. One such card might appear as follows:

PROBABILITY 90., 75., 50., 25., 3., 0

This would define ranges of:

Range 1: 90% through 100%

Range 2: 75% through 89%

Range 3: 50% through 74%

Range 4: 25% through 49%

Range 5: 3% through 24%

Range 6: 0% through 2%

A related card, PSYMBOLS, could be supplied to assign symbols to the above percent ranges for printing a map. If the card PROBABILITY is specified with no ranges, 8 default ranges are provided. In addition, 8 symbols are pre-defined to go along with these ranges. These defaults correspond to the following input cards:

```
PROBABILITY  80,60,45,30,20,10,3,0
PSYMBOLS     M, X, O, I, /, -, .
```

Specifying Test Fields

A test field is a subset of the classified area that is used to test the accuracy of the classification. Each test field is identified as belonging to one of the groups defined on the GROUP card, and is assigned a "group number" on this basis. If grouping is not used (no GROUP card in the input deck) these numbers will correspond to the class numbers assigned in the classification function for the classification run.

Test fields are defined, and are assigned group numbers, by a series of data cards in the following sequence:

```
DATA
  TEST n
    Field Description Cards
  TEST m
    Field Description Cards
```

The class of each test field is specified by the 'TEST n' and 'TEST m' card, where n and m are the numbers assigned to the groups or classes. Field Description cards supply the coordinates (run, line and column numbers) for each field that is included in the group. Any number of fields may be included in each group. Performance tables for these test fields and classes are produced as a result of PRINT card options, which are explained below.

The example deck in Figure 1 shows the definition of several test fields, which are identified as belonging to three groups.

Data cards which define the test fields, and assign group numbers to them, must be included in the input deck if any of the following options are exercised:

- Outlines of test fields
- Performance tables for test fields and/or test classes
- Tables of test field performance percentages or acreages

If none of these options are requested, this deck must not be included.

Standard Output

Three standard outputs are provided by PRINTRESULTS:

- An Input Deck Summary
- A Map of the Classified Area
- A List of all Training Fields and Test Fields

Figure 2 is the Input Deck Listing produced by the sample input deck in Figure 1. Two of the items listed under the heading "PRINTRESULTS SUPERVISOR INFORMATION...." should be noted. The first is the "Classification Study" number that appears on the first line. This is the identification number that is assigned to each classification by the Classifypoints function. Since the Classification Results File that is used in this example was produced by the example input deck for the Classifypoints function description, the identification number is the same as the one shown there. The "NUMBER OF TRAINING FIELDS SAVED..." line shows the number of training fields from the Statistics Deck that were used in the classification. Since CLASSIFYPOINTS need not use all of the training fields defined to the Statistics function, this number need not represent the total number of training fields in the Statistics Deck.

Figure 3 is the map-image produced by running the example deck. Note that most of the information that is needed for interpreting the map is included in a table under the heading of "CLASSES". This map should be compared to the much simpler map of the same area that was produced by the Classifypoints function.

Figure 4 is the Training Fields Listing and Figure 5 is the Test Fields Listing. The former lists the training fields actually used in the classification and the latter the test fields that were defined to PRINTRESULTS in its input deck.

```
DEMO                  LABORATORY FOR APPLICATIONS OF REMOTE SENSING          MAR  7,1973
PHILLIPS                           PURDUE UNIVERSITY                         12 18 49 PM
                                                                             LARSYS VERSION 3

    *PRINTRESULTS
    RESULTS DISK
    PRINT STATS,MAPS(2),OUTLINE(TRAIN,TEST)
    PRINT TRAIN(F,C),TEST(F,C,P)
    SYMBOLS C,S,O,H,F
    GROUP CORN(1/1/),SOYBEANS(2/2/),OTHER(3/3,4,5/)
    BLOCK RUN(71053900),CALC
    THRESHOLD 2*0.1,3*0.0
    DATA

YOU HAVE SELECTED THE FOLLOWING DISPLAY OPTIONS

    CONSIDER A SUBSET OF CLASSIFIED AREA FOR MAP
    CONSIDER A SUBSET OF CLASSIFIED AREA FOR CALCULATIONS
    PRINT REDUCED STATISTICS
    OUTLINE TRAINING FIELDS
    OUTLINE TEST FIELDS
    PRINT TRAINING CLASS PERFORMANCE
    PRINT TRAINING FIELD PERFORMANCE
    PRINT TEST FIELD PERFORMANCE
    PRINT TEST CLASS PERFORMANCE
    PRINT TEST FIELD PERCENTAGES
    APPLY THRESHOLDS
    USE CLASSIFICATION RESULTS FROM DISK

PRINTRESULTS SUPERVISOR INFORMATION....

    CLASSIFICATION STUDY.........  306644242
    NUMBER OF MAPS.......................  2
    NUMBER OF TRAINING FIELDS SAVED....   15
    NUMBER OF TEST FIELDS SAVED........   30
    NUMBER OF COPIES OF TABLES..........   1

                                                                  PRI-2
```

Figure 2. Control Card Summary for PRINTRESULTS

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
3 00 51 PM
LARSYS VERSION 3

CLASSIFICATION STUDY 424253584	CLASSIFIED AUG 30,1974
RUN NUMBER............ 71053900	DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CRN BLT LO FL210	TIME DATA TAKEN..... 1202 HOURS
DATA TAPE/FILE NUMBER.. 1005/ 2	PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971	GROUND HEADING..... 180 DEGREES

CLASSIFICATION WRITTEN ON DISK

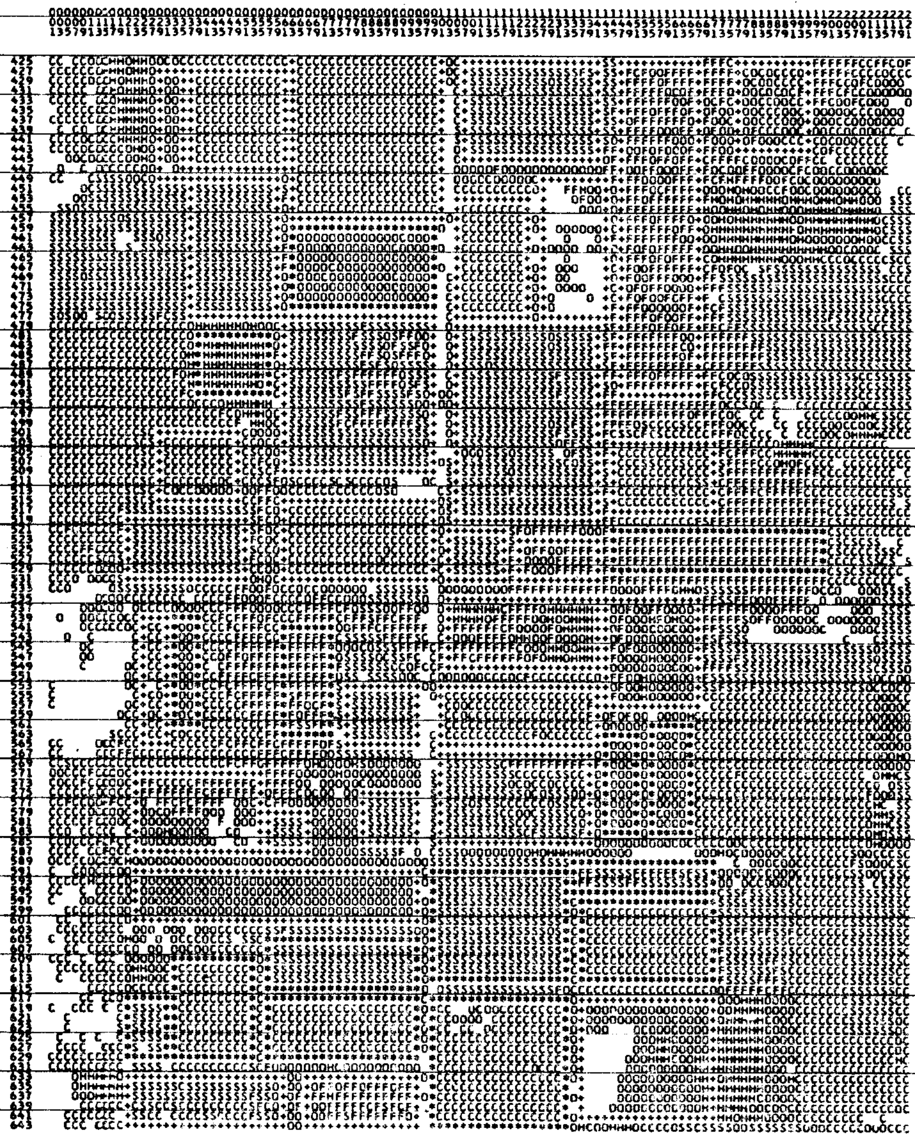
CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1	CO = 23.10
CHANNEL 9	SPECTRAL BAND 1.00 TO 1.40 MICROMETERS	CALIBRATION CODE = 1	CO = 22.90
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO = 20.60

CLASSES

SYMBOL	CLASS	WEIGHT	GROUP	THRES PCT	SYMBOL	CLASS	WEIGHT	GROUP	THRES PCT
C	CORN	0.290	CORN	0.10	H	HAY	0.110	OTHER	0.0
S	SOYBEANS	0.310	SOYBEANS	0.10	F	FOREST	0.195	OTHER	0.0
O	OAT/WHT	0.095	OTHER	0.0					

TRAINING FIELDS OUTLINED WITH A *
TEST FIELDS OUTLINED WITH A +
SHARED BOUNDARIES OUTLINED WITH A $



NUMBER OF POINTS DISPLAYED IS 12210

Figure 3: Classification Map

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 28,1974
8 23 31 AM
LARSYS VERSION 3

CLASSIFICATION STUDY 417929788 CLASSIFIED JUNE 28,1974

CLASSIFICATION WRITTEN ON DISK

SAVED TRAINING FIELDS

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION	CLASSIFY CLASS	DISPLAY CLASS
1	71053900	L-6	610	628	1	35	52	1	CORN		CORN	CORN
2	71053900	L-7	619	628	1	58	94	1	CORN		CORN	CORN
3	71053900	D*5	618	641	1	100	132	1	CORN		CORN	CORN
4	71053900	H*1	598	612	1	138	169	1	CORN		CORN	CORN
5	71053900	L-5	619	628	1	23	29	1	SOYBEANS		SOYBEANS	SOYBEANS
6	71053900	L-8	606	614	1	58	94	1	SOYBEANS		SOYBEANS	SOYBEANS
7	71053900	D*6	592	612	1	101	131	1	SOYBEANS		SOYBEANS	SOYBEANS
8	71053900	H*2	591	593	1	137	170	1	SOYBEANS		SOYBEANS	SOYBEANS
9	71053900	F-9	461	473	1	64	98	1	OATS		OAT/WHT	OTHER
10	71053900	K-4	541	559	1	35	38	1	OATS		OAT/WHT	OTHER
11	71053900	G*2	565	582	1	144	149	1	OATS		OAT/WHT	OTHER
12	71053900	R-1	482	492	1	40	56	1	HAY		HAY	OTHER
13	71053900	G*3	562	582	1	156	164	1	WHEAT		OAT/WHT	OTHER
14	71053900	H-6	543	562	1	63	72	1	WOODS		FOREST	OTHER
15	71053900	E*2	523	528	1	147	197	1	WOODS		FOREST	OTHER

Figure 4. Training Fields Listing

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 28,1974
8 23 31 AM
LARSYS VERSION 3

CLASSIFICATION STUDY 417929788 CLASSIFIED JUNE 28,1974

CLASSIFICATION WRITTEN ON DISK

SAVED TEST FIELDS

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION	DISPLAY CLASS
1	71053900	C-5	428	445	1	38	59	1	CORN		CORN
2	71053900	F-8	425	456	1	65	99	1	CORN		CORN
3	71053900	D*3	454	475	1	106	122	1	CORN		CORN
4	71053900	R-2	502	514	1	31	48	1	CORN		CORN
5	71053900	R-11	517	529	1	62	98	1	CORN		CORN
6	71053900	K-5	540	564	1	26	29	1	CORN		CORN
7	71053900	F*3	554	564	1	102	139	1	CORN		CORN
8	71053900	L-3	634	642	1	23	60	1	SOYBEANS		SOYBEANS
9	71053900	H-5	580	585	1	59	65	1	SOYBEANS		SOYBEANS
10	71053900	H-4	577	584	1	83	93	1	SOYBEANS		SOYBEANS
11	71053900	F*5	569	584	1	102	137	1	SOYBEANS		SOYBEANS
12	71053900	H-3	555	564	1	78	94	1	SOYBEANS		SOYBEANS
13	71053900	D*1	426	443	1	109	139	1	SOYBEANS		SOYBEANS
14	71053900	C-7	451	475	1	38	58	1	SOYBEANS		SOYBEANS
15	71053900	F-10	479	507	1	63	98	1	SOYBEANS		SOYBEANS
16	71053900	O-2	481	518	1	106	140	1	SOYBEANS		SOYBEANS
17	71053900	O-2	523	530	1	104	116	1	SOYBEANS		SOYBEANS
18	71053900	E*3	505	515	1	147	169	1	SOYBEANS		SOYBEANS
19	71053900	R-3	520	529	1	23	49	1	SOYBEANS		SOYBEANS
20	71053900	L-2	592	600	1	25	94	1	PASTURE		OTHER
21	71053900	C-6	428	447	1	30	33	1	OATS		OTHER
22	71053900	O-3	525	529	1	122	140	1	WOODS		OTHER
23	71053900	L-4	634	641	1	68	94	1	WOODS		OTHER
24	71053900	E*1	426	441	1	179	193	1	OTHER	SET ASIDE	OTHER
25	71053900	D*4	450	475	1	128	142	1	OTHER	SET ASIDE/NON	OTHER
26	71053900	E*2	427	492	1	149	166	1	OTHER	WDS/PAS	OTHER
27	71053900	I-K-123	573	585	1	25	53	1	OTHER	SEV FLDS	OTHER
28	71053900	F*1/2/6	537	547	1	104	139	1	OTHER	SA/NF	OTHER
29	71053900	G*1	537	556	1	145	165	1	OTHER	HAY	OTHER
30	71053900	H*3/4	618	640	1	139	169	1	OTHER	WHT/HAY	OTHER

Figure 5. Test Fields Listing

Optional controls may be imposed on the last three of these outputs with the PRINT control card. One map is always produced unless the 'PRINT MAPS(n)' specifies either no map (n=0) be printed or additional copies (n greater than 1) be printed. A 'PRINT NOLIST' card deletes the training and test fields list.

Optional Output

The PRINT card provides control of optional printed output. Besides affecting the standard output, as explained above, this control card enables the user to:

- Request a listing of the Statistics used in the classification.
- Request performance tables for training fields and classes and test fields and classes. These tables also include the standard channels and classes header information that appears on the map.
- Designate test and training fields to be outlined on the map.

The example input deck (Figure 1) produced the outputs shown in Figures 6 through 11 on the following pages.

Requesting the Statistics Listing

A listing of statistics used in the classification may be requested by the 'STATS' parameter on the PRINT control card. This listing contains the means and standard deviations of each channel and the correlation matrix of the channels for each class. The sample output is shown in Figure 6.

Requesting Performance Tables

Performance tables can be requested for training and/or test fields and/or classes. These tables list the number of samples from each field or class that was classified correctly and the number of samples that were classified into each class. For test fields, a table of percentages also may be requested.

A combination of parameters enables the user to select the performance tables to be printed. The parameters specify either test or training samples (TEST or TRAIN) as well as whether performance tables for fields (F), classes (C), or both (F,C) are to be printed. In addition, a percentage table for test field performance may be specified by a 'P', and a table of acres and hectares for test fields may be specified by an 'A'.

DEMO
PHIL ALENDUFF

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 28,1974
8 19 14 AM
LARSYS VERSION 3

CLASSIFICATION STUDY 417929788 ON DISK

CLASS....OAT/WHT

CHANNEL	2	8	12
SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
MEAN	97.02	111.95	117.03
STD. DEV.	4.50	8.59	10.06

CORRELATION MATRIX

SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
0.48-0.51	1.00		
0.72-0.92	-0.30	1.00	
9.30-11.70	0.58	-0.74	1.00

CLASS....HAY

CHANNEL	2	8	12
SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
MEAN	106.72	80.54	163.11
STD. DEV.	4.19	4.36	13.71

CORRELATION MATRIX

SPECTRAL BAND	0.48 - 0.51	0.72 - 0.92	9.30 - 11.70
0.48-0.51	1.00		
0.72-0.92	-0.67	1.00	
9.30-11.70	0.80	-0.71	1.00

Figure 6. Statistics Summary for Classes OAT/WHT and HAY

The user may request more than one copy of these tables by using the 'TABLES(n)' parameter on the PRINT card. For the sample run shown in Figure 1, a single copy of all tables was requested by default. These performance table listings are shown in Figures 7 to 11.

<u>Designating Outlines for the Map</u>

Training and test fields can be outlined on the map by using another option on the PRINT card. Training fields are outlined with asterisk symbols (*), test fields with plus signs (+), and boundaries that are common to both test and training fields with dollar signs ($). Since the fields are defined either in the results file or in the Test Field Data Deck, it is only necessary to specify that outlining be done. Three options are offered:

PRINT OUTLINE(TRAIN) - outline training fields

PRINT OUTLINE(TEST) - outline test fields

PRINT OUTLINE(TRAIN,TEST) - outline both

Since the outlines are printed outside the field boundary, if the field and the map share boundaries or if the field is not completely shown by the area displayed, the outline will not be closed.

DEMO PAUL SPENCER	LABORATORY FOR APPLICATIONS OF REMOTE SENSING PURDUE UNIVERSITY	AUG 30,1974 3 01 14 PM LARSYS VERSION 3

CLASSIFICATION STUDY 424253584 CLASSIFIED AUG 30,1974

CLASSIFICATION WRITTEN ON DISK

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1	CO = 23.10
CHANNEL 9	SPECTRAL BAND 1.00 TO 1.40 MICROMETERS	CALIBRATION CODE = 1	CO = 22.90
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO = 20.60

CLASSES

	CLASS	WEIGHT	GROUP	THRES PCT		CLASS	WEIGHT	GROUP	THRES PCT
1	CORN	0.290	CORN	0.10	4	HAY	0.110	OTHER	0.0
2	SOYBEANS	0.310	SOYBEANS	0.10	5	FOREST	0.195	OTHER	0.0
3	OAT/WHT	0.095	OTHER	0.0					

TRAINING FIELD PERFORMANCE

FIELD DESIG.	GROUP	NO OF SAMPS	PCT. CORCT	NUMBER OF SAMPLES CLASSIFIED INTO CORN	SOYBEANS	OTHER	THRSHOLD
L-6	CORN	81	100.0	81	0	0	0
L-7	CORN	90	100.0	90	0	0	0
D*5	CORN	192	92.7	178	0	9	5
H*1	CORN	112	100.0	112	0	0	0
L-5	SOYBEANS	20	95.0	0	19	0	1
L-8	SOYBEANS	72	100.0	0	72	0	0
D*6	SOYBEANS	160	99.4	0	159	1	0
H*2	SOYBEANS	34	61.8	0	21	13	0
F-9	OTHER	119	96.6	4	0	115	0
K-4	OTHER	20	95.0	1	0	19	0
G*2	OTHER	27	100.0	0	0	27	0
B-1	OTHER	40	100.0	0	0	40	0
G*3	OTHER	40	97.5	1	0	39	0
H-6	OTHER	50	92.0	1	3	46	0
E*2	OTHER	78	100.0	0	0	78	0
	TOTAL	1135		468	274	387	6

OVERALL PERFORMANCE(1096/ 1135) = 96.6

Figure 7: Training Field Performance Table

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
3 01 16 PM
LARSYS VERSION 3

CLASSIFICATION STUDY 424253584 CLASSIFIED AUG 30,1974
CLASSIFICATION WRITTEN ON DISK

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1	CO = 23.10
CHANNEL 9	SPECTRAL BAND 1.00 TO 1.40 MICROMETERS	CALIBRATION CODE = 1	CO = 22.90
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO = 20.60

CLASSES

	CLASS	WEIGHT	GROUP	THRES PCT		CLASS	WEIGHT	GROUP	THRES PCT
1	CORN	0.290	CORN	C.10	4	HAY	0.110	OTHER	0.0
2	SOYBEANS	0.310	SOYBEANS	C.10	5	FOREST	0.195	OTHER	0.0
3	OAT/WHT	0.095	OTHER	C.0					

TEST FIELD PERFORMANCE

FIELD DESIG.	GROUP	NO OF SAMPS	PCT. CORCT	NUMBER OF SAMPLES CLASSIFIED INTO: CORN	SOYBEANS	OTHER	THRSHOLD
C-5	CORN	99	99.0	98	1	0	0
F-8	CORN	288	100.0	288	0	0	0
D*3	CORN	88	100.0	88	0	0	0
R-2	CORN	54	85.2	46	1	7	0
R-11	CORN	126	98.4	124	0	2	0
K-5	CORN	24	87.5	21	0	2	1
F*3	CORN	95	92.6	88	C	7	0
L-3	SOYBEANS	76	57.9	24	44	7	1
H-5	SOYBEANS	12	100.0	0	12	0	0
H-4	SOYBEANS	24	100.0	0	24	0	C
F*5	SOYBEANS	144	61.1	31	88	25	0
H-3	SOYBEANS	40	100.0	0	40	0	0
D*1	SOYBEANS	144	97.9	0	141	3	0
C-7	SOYBEANS	130	100.0	0	130	0	0
F-10	SOYBEANS	270	78.1	1	211	58	0
D-2	SOYBEANS	323	93.2	1	301	20	1
O-2	SOYBEANS	24	100.0	0	24	0	0
E*3	SOYBEANS	72	8.3	66	6	0	0
R-3	SOYBEANS	70	100.0	0	70	0	0
L-2	OTHER	140	100.0	0	C	140	0
C-6	OTHER	20	95.0	0	0	19	1
D-3	OTHER	27	100.0	0	0	27	0
L-4	OTHER	52	92.3	2	2	48	0
E*1	OTHER	64	51.6	31	0	33	0
D*4	OTHER	91	42.9	0	0	39	52
E*2	OTHER	297	99.C	3	0	294	0
I-K-123	OTHER	105	76.2	13	0	80	12
F*1/2/6	OTHER	108	96.3	4	0	104	0
G*1	OTHER	110	100.0	0	0	110	0
H*3/4	OTHER	176	77.3	0	0	136	40
	TOTAL	3293		929	1095	1161	108

OVERALL PERFORMANCE(2874/ 3293) = 87.3

Figure 8: Test Field Performance Table

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
3 01 17 PM
LARSYS VERSION 3

CLASSIFICATION STUDY 424253584 CLASSIFIED AUG 30,1974

CLASSIFICATION WRITTEN ON DISK

CHANNELS USED

CHANNEL 2 SPECTRAL BAND 0.48 TO 0.51 MICROMETERS CALIBRATION CODE = 1 CO = 23.10
CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS CALIBRATION CODE = 1 CO = 22.90
CHANNEL 12 SPECTRAL BAND 9.30 TO 11.70 MICROMETERS CALIBRATION CODE = 1 CO = 20.60

CLASSES

	CLASS	WEIGHT	GROUP	THRES PCT		CLASS	WEIGHT	GROUP	THRES PCT
1	CORN	0.290	CORN	0.10	4	HAY	0.110	OTHER	0.0
2	SOYBEANS	0.310	SOYBEANS	0.10	5	FOREST	0.195	OTHER	0.0
3	OAT/WHT	0.095	OTHER	C.0					

TEST FIELD PERCENTAGE (PCT.)

FIELD DESIG.	GROUP	NO OF SAMPS	PERCENTAGE OF SAMPLES CLASSIFIED INTO CORN	SOYBEANS	OTHER	THRSHOLD
C-5	CORN	99	99.0	1.0	0.0	0.0
F-8	CORN	288	100.0	0.0	0.0	0.0
D*3	CORN	88	100.0	0.0	0.0	0.0
R-2	CORN	54	85.2	1.9	13.0	0.0
R-11	CORN	126	98.4	0.0	1.6	0.0
K-5	CORN	24	87.5	0.0	8.3	4.2
F*3	CORN	95	92.6	0.0	7.4	0.0
L-3	SOYBEANS	76	31.6	57.9	9.2	1.3
H-5	SOYBEANS	12	0.0	100.0	0.0	0.0
H-4	SOYBEANS	24	0.0	100.0	0.0	0.0
F*5	SOYBEANS	144	21.5	61.1	17.4	0.0
H-3	SOYBEANS	40	0.0	100.0	0.0	0.0
D*1	SOYBEANS	144	0.0	97.9	2.1	0.0
C-7	SOYBEANS	130	0.0	100.0	0.0	0.0
F-10	SOYBEANS	270	0.4	78.1	21.5	0.0
O-2	SOYBEANS	323	0.3	93.2	6.2	0.3
D-2	SOYBEANS	24	0.0	100.0	0.0	0.0
E*3	SOYBEANS	72	91.7	8.3	0.0	0.0
R-3	SOYBEANS	70	0.0	100.0	0.0	0.0
L-2	OTHER	140	0.0	0.0	100.0	0.0
C-6	OTHER	20	0.0	0.0	95.0	5.0
O-3	OTHER	27	0.0	0.0	100.0	0.0
L-4	OTHER	52	3.8	3.8	92.3	0.0
E*1	OTHER	64	48.4	0.0	51.6	0.0
D*4	OTHER	91	0.0	0.0	42.9	57.1
E*2	OTHER	297	1.0	0.0	99.0	0.0
I-K-123	OTHER	105	12.4	0.0	76.2	11.4
F*1/2/6	OTHER	108	3.7	0.0	96.3	0.0
G*1	OTHER	110	0.0	0.0	100.0	0.0
H*3/4	OTHER	176	0.0	0.0	77.3	22.7
	TOTAL	3293	28.2	33.3	35.3	3.3

OVERALL PERFORMANCE(2874/ 3293) = 87.3

Figure 9: Test Field Percentage Table

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
3 01 19 PM
LARSYS VERSION 3

CLASSIFICATION STUDY 424253584 CLASSIFIED AUG 30,1974

CLASSIFICATION WRITTEN ON DISK

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1	CO =	23.10
CHANNEL 9	SPECTRAL BAND 1.00 TO 1.40 MICROMETERS	CALIBRATION CODE = 1	CO =	22.90
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO =	20.60

CLASSES

	CLASS	WEIGHT	GROUP	THRES PCT		CLASS	WEIGHT	GROUP	THRES PCT
1	CORN	0.290	CORN	0.10	4	HAY	0.110	OTHER	0.0
2	SOYBEANS	0.310	SOYBEANS	0.10	5	FOREST	0.195	OTHER	0.0
3	OAT/WHT	0.095	OTHER	0.0					

TRAINING CLASS PERFORMANCE

GROUP	NO OF SAMPS	PCT. CORCT	NUMBER OF SAMPLES CLASSIFIED INTO CORN	SOYBEANS	OTHER	THRSHOLD
1 CORN	475	97.1	461	0	9	5
2 SOYBEANS	286	94.8	0	271	14	1
3 OTHER	374	97.3	7	3	364	0
TOTAL	1135		468	274	387	6

OVERALL PERFORMANCE(1096/ 1135) = 96.6
AVERAGE PERFORMANCE BY CLASS(289.1/ 3) = 96.4

Figure 10: Training Class Performance Table

DEMO
PAUL SPENCER

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

AUG 30,1974
3 01 21 PM
LARSYS VERSION 3

CLASSIFICATION STUDY 424253584 CLASSIFIED AUG 30,1974

CLASSIFICATION WRITTEN ON DISK

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.48 TO 0.51 MICROMETERS	CALIBRATION CODE = 1	CO =	23.10
CHANNEL 9	SPECTRAL BAND 1.00 TO 1.40 MICROMETERS	CALIBRATION CODE = 1	CO =	22.90
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO =	20.60

CLASSES

	CLASS	WEIGHT	GROUP	THRES PCT		CLASS	WEIGHT	GROUP	THRES PCT
1	CORN	0.290	CORN	0.10	4	HAY	0.110	OTHER	0.0
2	SOYBEANS	0.310	SOYBEANS	0.10	5	FOREST	0.195	OTHER	0.0
3	OAT/WHT	0.095	OTHER	0.0					

TEST CLASS PERFORMANCE

GROUP	NO OF SAMPS	PCT. CORCT	CORN	SOYBEANS	OTHER	THRSHOLD
			NUMBER OF SAMPLES CLASSIFIED INTO			
1 CORN	774	97.3	753	2	18	1
2 SOYBEANS	1329	82.1	123	1091	113	2
3 OTHER	1190	86.6	53	2	1030	105
TOTAL	3293		929	1095	1161	108

OVERALL PERFORMANCE(2874/ 3293) = 87.3

AVERAGE PERFORMANCE BY CLASS(265.9/ 3) = 88.6

Figure 11: Test Class Performance Table

PUNCHSTATISTICS

The Punchstatistics function produces a punched deck of the Statistics File that is contained in a Classification Results File. The Statistics File on the results tape is a duplicate of that which was used as input to the Classifypoints function.

The only input data to PUNCHSTATISTICS is a Classification Results File specified by the user. The file must reside on tape to be accepted by the function. The outputs are the punched Statistics File and a listing identifying the results file that was used. The latter output is optional.

Control Cards

Other than the Function Selector Card and END card, which are required for all functions, there are only two Punchstatistics control cards - the FROM card and the PRINT card. The FROM Control Card is always required. It specifies the tape and file numbers of the file to be used. Figure 1 illustrates a sample input deck in which the FROM card specifies Tape 1043 and File 1. Scratch tapes ('TAPE(0)') are not usable for this function.

The user may suppress most of the information from the standard listing by using the optional 'PRINT NOLIST' card. The result of this option is explained below in the output description.

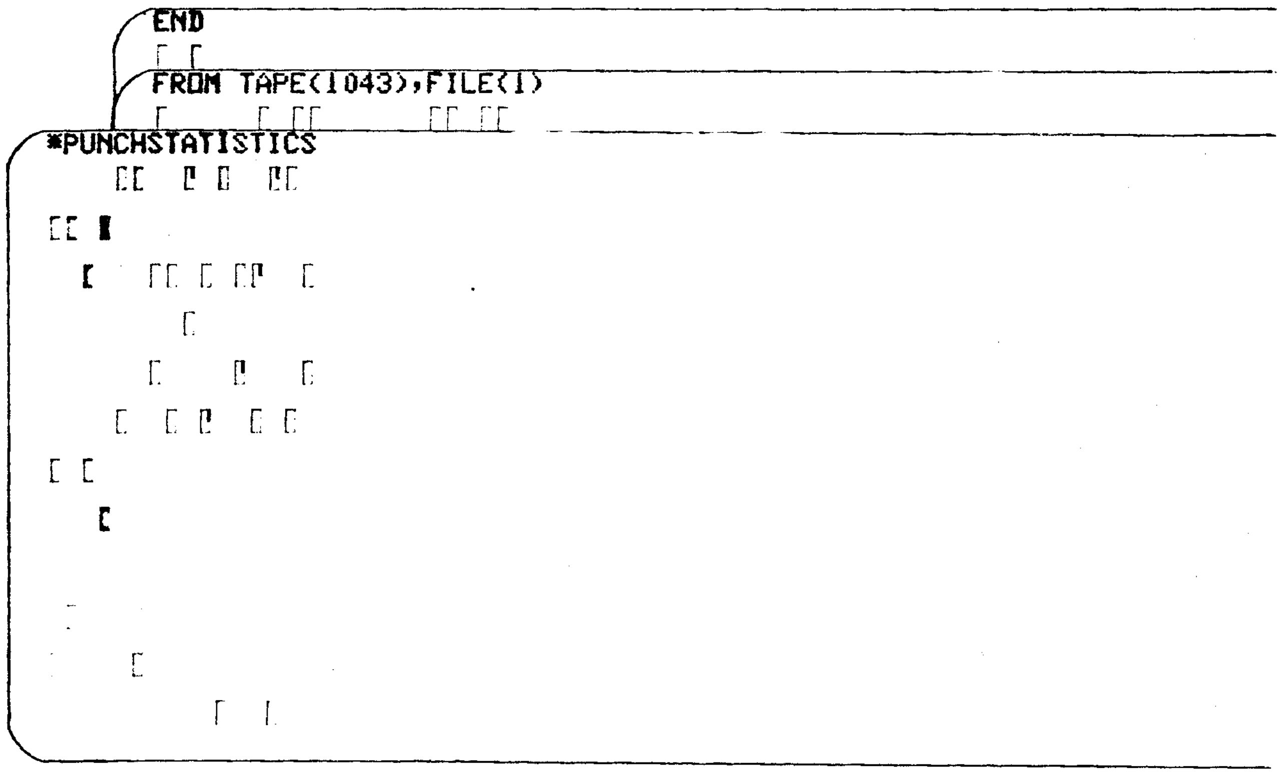


Figure 1. Input Deck for PUNCHSTATISTICS

Punched Card Output

The principal output of PUNCHSTATISTICS is a Statistics Deck which is identical to that produced by the Statistics function. The deck contains the Field Description Cards, number of points in each class, number of classes and fields, channels information, mean vectors, and covariance matrices.

Printed Output

In addition to the punched output, the function produces a Control Card summary listing and a Classification Results File ID Listing. An example of the Control Card summary is shown in Figure 2.

The second standard printed output, the Classification Results File ID Listing, contains the tape and file numbers, channels used in the classification, the class names, the class weights if any were assigned, and the calibration and channels information for each area classified. Figure 3 is an example of this output. The user may suppress this listing by including the 'PRINT NOLIST' control card in his input deck.

```
DEMO                  LABORATORY FOR APPLICATIONS OF REMOTE SENSING          MAR  7,1973
PHILLIPS                          PURDUE UNIVERSITY                          12 19 57 PM
                                                                             LARSYS VERSION 3

    *PUNCHSTATISTICS
    FROM TAPE(1043),FILE(1)
    END

 YOU HAVE SELECTED THE FOLLOWING OPTIONS
    TAPE 1043 FILE  1 WILL BE USED

                                                                             PUN-2
```

Figure 2. Control Card Summary for PUNCHSTATISTICS

DEMO PAUL SPENCER	LABORATORY FOR APPLICATIONS OF REMOTE SENSING PURDUE UNIVERSITY	AUG 30,1974 2 56 20 PM LARSYS VERSION 3

TAPE-- 1043 FILE-- 1 CLASSIFIED- AUG 30,1974

CHANNELS USED

CHANNEL 2 SPECTRAL BAND 0.48 TO 0.51 MICROMETERS CALIBRATION CODE = 1
CHANNEL 9 SPECTRAL BAND 1.00 TO 1.40 MICROMETERS CALIBRATION CODE = 1
CHANNEL 12 SPECTRAL BAND 9.30 TO11.70 MICROMETERS CALIBRATION CODE = 1

CLASSES AND WEIGHTS

CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT	CLASS WEIGHT
CORN 0.290	SOYBEANS 0.310	OAT/WHT 0.095	HAY 0.110	FOREST 0.195		

RUN	AREA CLASSIFIED	FLIGHT LINE
71053900	LINES---- 425 TO 644 BY 2 COLUMNS-- 1 TO 222 BY 2	CRN BLT LO FL210

CALIBRATION VALUES

CHANNEL 2 CO = 23.10
CHANNEL 9 CO = 22.90
CHANNEL 12 CO = 20.60
NUMBER OF LINES CLASSIFIED IS 110

Figure 3: Classification Results File ID Listing

RATIOMEANS

The RATIOMEANS function calculates the ratio of class means between two sets of channels (visible/infrared) . This function can be used also to transfer a LARSFRIS Statistics file from cards to the user's T disk.

Input to the RATIOMEANS function consists of:

- Control cards to select the processing function and output options.
- A LARSFRIS statistics file, which may be obtained from either cards, disk, or a classification results tape.

Figure 1 shows a sample RATIOMEANS input deck.

The RATIOMEANS function always produces the following standard printed output.

- A listing of the control cards that were supplied by the user.
- A list of the options selected by the user.
- Tables containing the identification of each class (class number, class symbol, and class name), the class mean values in every channel, the sum of the means in all channels (Magnitude), the ratio (visible/infrared) of every class, the percent response corresponding to the visible and infrared portions of the spectrum, the class variance, and the number of points in each class. These tables are ordered by either class number, by magnitude, or by ratio values. (See Figure 2.)

The user may specify any number of channels (up to 60 channels) to be treated as visible or infrared. If an OPTIONS control card is used, both the visible and the infrared options have to be specified. It is possible to have several OPTIONS cards for each option, or one OPTIONS card for both the visible and infrared options. In order to allow for differential weighting of selected channels, the user may repeat the selected channels any number of times. The inclusion of an OPTIONS card suppresses the printing of the variances, number of points in each class, and the tables ordered by magnitude and ratio values. If an OPTIONS control card is not used, the RATIOMEANS function will automatically select the first two channels in the statistics file to be treated as visible channels and the second two channels as infrared channels. This default is only valid for Landsat MSS data.

The NORATIOS control card allows the user to transfer the LARSFRIS statistics file from either cards or the classification results tape to the user's T disk.

The PRINT control card enables the user to obtain a listing of the classification results file information when the LARSFRIS statistics file is requested from a classification results tape [FROM TAPE (TTT), FILE (FF).] The classification results information is identical to that obtained from the LARSFRIS *LISTRESULTS function. See Figure 3.

The spectral band ratios have been utilized by a number of remote sensing data analysts as an aid to identify ground cover types.

For further discussion on the applications of spectral band ratios refer to: Kristof and Zachary (1971), Gorsuch (1974), Kristof and Baumgardner (1975), Thie et al. (1975), Tarnocai and Krsitof (1976), Weismiller et al. (1976) and Singh et al. (1979).

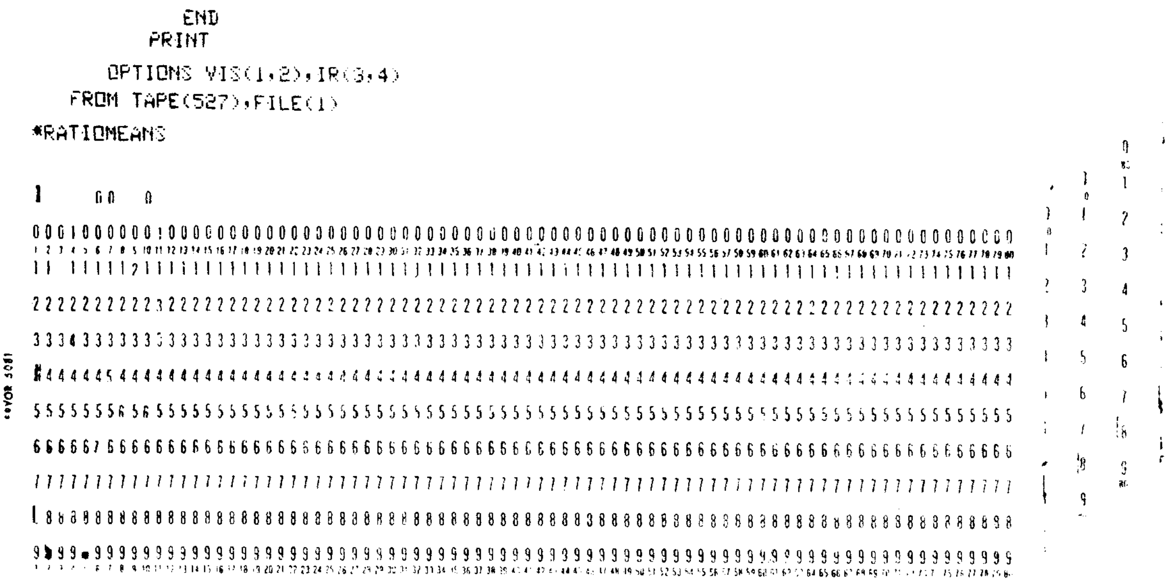


Figure 1. Sample RATIOMEANS input deck.

PRIS LABORATORY FOR APPLICATIONS OF REMOTE SENSING [illegible]
PURDUE UNIVERSITY
09 34 08 AM
LARSYS VERSION 3

****************************** ORDERED BY CLASS ******************************

NO.	SYMBOLS	CLASS	MAGNITUDE	VIS/IR	PCT.VIS	PCT.IR
1	A	01	29.93	6.0438	85.80	14.20
2	B	02	35.18	4.8094	82.79	17.21
3	C	03	50.62	1.4234	58.74	41.26
4	D	04	69.89	1.9295	65.86	34.14
5	E	05	146.55	0.9120	47.70	52.30
6	F	06	126.87	0.7139	41.65	58.35
7	G	07	127.12	1.0125	50.31	49.69
8	H	08	110.63	0.9353	48.33	51.67
9	I	09	94.47	1.0613	51.49	48.51
10	J	10	93.21	0.7766	43.71	56.29
11	K	11	105.15	0.6084	37.83	62.17
12	L	12	111.86	0.4952	33.12	66.88
13	M	13	66.28	0.8486	45.90	54.10
14	N	14	82.79	0.5354	34.87	65.13
15	O	15	98.19	0.4669	31.83	68.17
16	P	16	128.06	0.3146	23.93	76.07

I0103 CPU TIME USED WAS 38.227 SECONDS. (LARSMN)

Figure 2. Sample RATIOMEANS standard output

FRIS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

OCT 10,1980
09 34 08 AM
LARSYS VERSION 3

TAPE-- 527   FILE-- 1   CLASSIFIED- JULY 19,1978

CHANNELS USED

CHANNEL 1	SPECTRAL BAND	0.50 TO 0.60 MICROMETERS	CALIBRATION CODE = 1
CHANNEL 2	SPECTRAL BAND	0.60 TO 0.70 MICROMETERS	CALIBRATION CODE = 1
CHANNEL 3	SPECTRAL BAND	0.70 TO 0.80 MICROMETERS	CALIBRATION CODE = 1
CHANNEL 4	SPECTRAL BAND	0.80 TO 1.10 MICROMETERS	CALIBRATION CODE = 1

CLASSES

CLASS	CLASS	CLASS	CLASS	CLASS	CLASS	CLASS
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16					

RUN	AREA CLASSIFIED		FLIGHT LINE
76020401	LINES---- 877 TO 1661 BY 1	COLUMNS-- 1 TO 1000 BY 1	240515402 GUAT

CALIBRATION VALUES

CHANNEL 1 CO = 0.0

CHANNEL 2 CO = 0.0

CHANNEL 3 CO = 0.0

CHANNEL 4 CO = 0.0

NUMBER OF LINES CLASSIFIED IS 785

Figure 3. Sample RATIOMEANS output when the LARSYS statistics file is requested from a classification results tape and the PRINT control card is included.

REFERENCES

Kristof, S.J. and A.L. Zachary (1971), "Mapping Soil Features from Multispectral Scanner Data", LARS Information Note 061471, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

Gorsuch, Joseph (1974), "A Suggested Range of Ratios of Covertypes for Great Lakes Analysis", LARS Internal Report.

Kristof, S.J. and M.F. Baumgardner (1975), "Changes of Multispectral Soils Patterns with Increasing Crop Canopy", Agronomy Journal Vol. 67, pp: 317-321.

Thie, J. C. Tarnocai, G.E. Mills, and S.F. Kristof (1975), "A Rapid Resource Inventory for Canada's North by Means of Satellite and Airborne Remote Sensing", LARS Information Note 031775, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

Tarnocai, C. and S.J. Kristof (1976), "Computer-Aided Classification of Land and Water Bodies Using Landsat Data, Mackenzie Delta Area, N.W.T., Canada", Arctic and Alpine Research Journal, Vol. 8, No. 2, pp: 151-159.

Weismiller, R.A., S.J. Kristof, D.K. Scholz, P.E. Anuta and S.M. Momin (1977), "Evaluation of Change Detection Techniques for Monitoring Coastal Zones Environments", LARS Information Note 062277, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

Singh, A.N., S.J. Kristof, and M.F. Baumgardner (1979), "Delineating Salt-Affected Soils in the Ganges Plain by Digital Analysis of Landsat Data", Journal of the Indian Society of Photo-interpretation and Remote Sensing, Vol. VII, No. 1.

SAMPLECLASSIFY

The Sampleclassify function implements a statistical distance measure for classifying "samples" (test fields) from the Multispectral Image Storage Tape. It provides an experimental system for testing a classification scheme which uses spatial information (user input field boundaries). It employs mean vectors and covariance matrices for the training classes (computed by the Statistics function) and for the test fields (computed by the Sampleclassify function) to calculate the distance between the probability density function of each test field being classified and the probability density function of each of the training classes. It then assigns the field to the closest training class.

Input to the function is:

- Data from the Multispectral Image Storage Tape(s).
- Control Cards to select the processing and output options.
- A Statistics File from the Statistics function to describe the training classes.
- Field Description Cards to identify the area or areas to be classified (test fields).

SAMPLECLASSIFY produces six standard and one optional printed outputs. The standard outputs are:

- A control card summary
- A list of channels and training classes considered.
- A list of the training fields that were used.
- A list showing identification information from the Multispectral Image Storage Tape with a list of the fields classified from the tape.
- A list of the classified fields in the order supplied in the input deck.
- A performance summary table.

The optional output is a listing of the statistics for each of the training classes. More detailed information on these listings, and an example of each of them, is given later in this description.

Inputs

The major inputs to the Sampleclassify function are the statistical description of the training classes, a description of the areas to be classified, data from the Multispectral Image Storage Tape, and an evaluation procedure. Data for the statistical description of the training classes originated in a Statistics File generated by the Statistics function. By specifying a subset of the channels with a CHANNELS card or by using an optional classes card, the user can specify that only a subset of the data on the Statistics File will be used. The CLASSES card is used to pool the statistics of two or more classes in the file into a single "pooled class", and the CHANNELS card identifies the specific wavelength bands to be used in the classification.

The Sampleclassify function uses these inputs to compute the statistical description of each of the test fields. These test fields are defined by Field Description Cards that specify a data subset from a Multispectral Image Storage Tape to be used in the calculation. The evaluation procedure is defined by a GROUP card, or default options, which identify the evaluation classes and by the location of the Field Description Cards (test fields) in the input deck.

Statistics File Input: A Statistics File, generated by the LARSFRIS Statistics function, must be available to the Sampleclassify function and may be input as a disk file or as punched cards. If the file that is to be used is on disk, there is no need to specify such. SAMPLECLASSIFY assumes that it is going to be input from disk and will automatically look for it there. If, however, the disk file that is to be used was generated in a previous terminal session and saved on the user's A-disk via the STATDECK SAVE command, then the user must notify LARSYS that this file will be used by issuing the STATDECK USE command before starting the Sampleclassify run.

If the file is to be input from punched cards, the CARDS READSTATS control card must be included in the input deck, and the Statistics File on punched cards must be the first group of data cards in the deck, i.e., they must come after all control cards and before the Field Description Cards.

The class statistics that are contained in the Statistics File may be used to classify fields (test fields); however the user also has the option of further defining the training classes with a CLASSES card. This card is used to pool the statistics of different classes from the Statistics File to form new classes ("pooled classes"). Any number of the original classes in the Statistics File may be pooled in this manner. If the user employs this option he must specifically include on a CLASSES card all of the classes he wishes to use. Any that are not included will be dropped from consideration in the classification.

Use of CLASSES Card: The CLASSES control card is used either to select a subset of the classes present in the Statistics File, or to combine classes into pooled classes. In the first case, the user need only specify the class number of the classes he wishes to employ. For example, if there were 10 classes available from the Statistics File, the card;

```
CLASSES 1,3,5,6
```

would indicate that only classes 1,3,5, and 6 are to be employed for classification; classes 2,4,7,8,9, and 10 are ignored. For evaluation purposes, discussed later, classes 1,3,5 and 6 will also be reassigned to training class numbers 1,2,3 and 4.

To combine two or more classes into one pooled class, the user assigns a name and class number to the pool and specifies the classes to be included. For example, assume there are eleven classes defined in the training class statistics, and the user wishes to process the following combinations:

- A pooled class called CORN created from classes 1 and 2 in the Statistics File.
- A pooled class called SOYBEANS created from classes 3 and 4 in the Statistics File.

The control card format to specify this option is:

```
CLASSES CORN(1/1,2/),SOYBEANS(2/3,4/)
```

Note that the number immediately following a left parenthesis specifies the pooled class number. These must be in ascending order and are given the name to the left of the parenthesis. Note also that the class numbers, defined in the Statistics File, to be pooled are enclosed by slashes(/).

The example that is shown above is the same as the first CLASSES card in the sample input deck in Figure 1. This input deck assumes that a Statistics File that contains nine training classes exists on disk. The nine classes have here been re-defined with CLASSES cards into five "pooled classes". For instance, the first CLASSES card defines the pooled class CORN and specifies that the statistics from the first two classes in the Statistics File are to be pooled together to create the new training class. The same card also defines the new training class SOYBEANS (equating it to Statistics File class numbers 3 and 4). Note that class 7 from the Statistics File is not used in this classification.

The training classes defined for the example classification will be CORN, SOYBEANS, OW(Oats and Wheat), HAY and FOREST. These classes are assigned class numbers 1 through 5 respectively. Since three channels (7,8 and 12) were specified on the CHANNELS card, the statistical descriptions of these training classes will contain statistics for these three wavelength bands. A PRINT STATS card is included in the input deck to print the resulting statistics.

<u>Specification of Channels:</u> The channels to be used by SAMPLE-CLASSIFY for both training and test statistical descriptions must be specified on a CHANNELS card. The Control Card

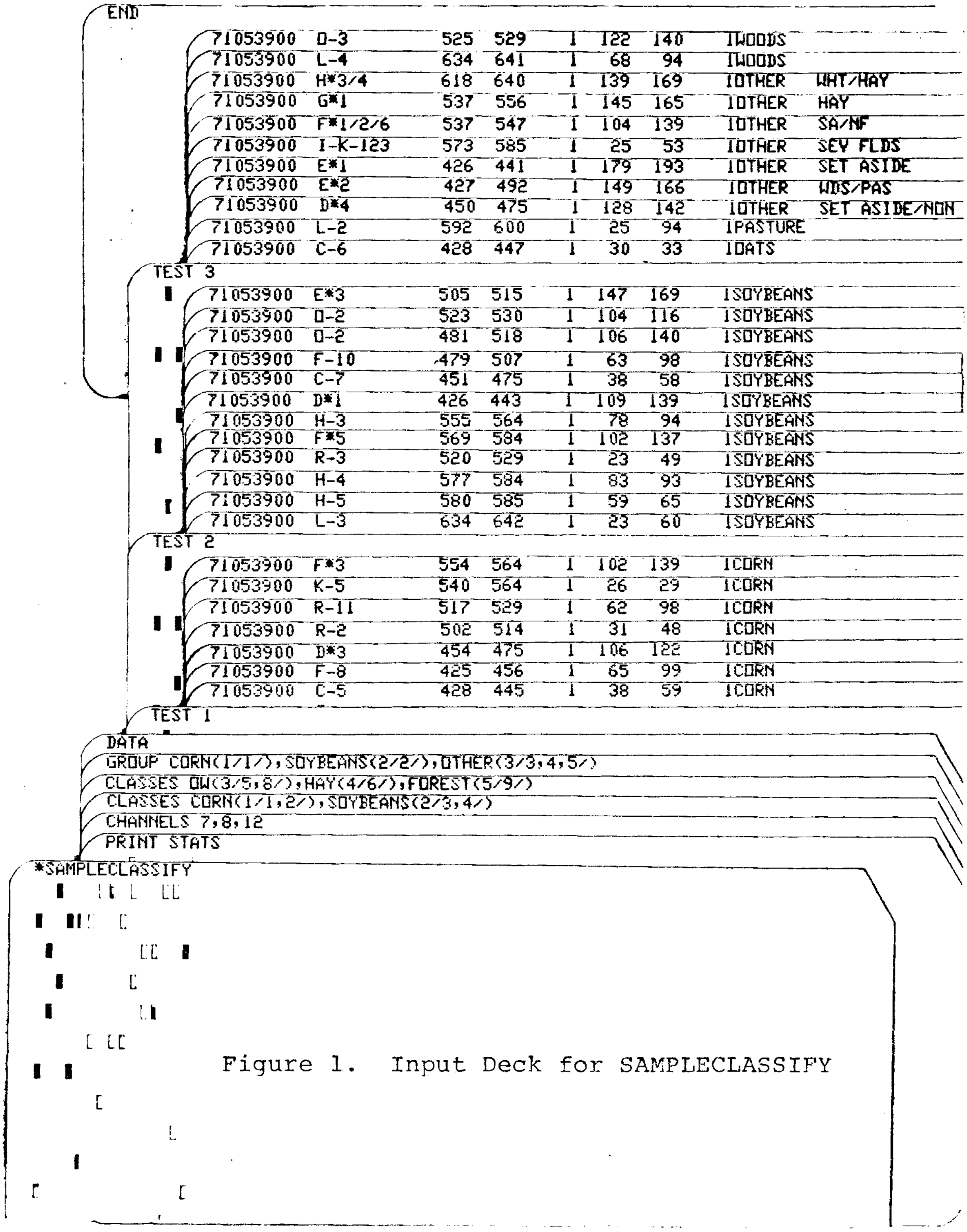


Figure 1. Input Deck for SAMPLECLASSIFY

Dictionary (Appendix I) describes in detail two forms of this card and how to use each of them. The CHANNELS card shown in Figure 1 indicates that channels 7,8, and 12 are to be used.

Evaluation of Classification: The Sampleclassify function will classify each test field by calculating the "distance" between the probability density function of the "field" and the probability density function of each training class and will assign the field to the closest training class. Training classes are selected because of their ability to classify fields and not necessarily because of their informational value. Two or more subclasses might, therefore, be defined for a class of interest; and an evaluation of the classifier requires this subclass information.

In Figure 1, for example, the classifier was designed to classify CORN, SOYBEANS and OTHER. In the final training class definition three subclasses of OTHER are defined. The evaluation should consider only the three principal classes, however, and the GROUP card is provided for this purpose.

The GROUP control card allows the user to combine training classes into performance groups. This grouping does not affect the training class statistics in any way. The classes that are identified on the GROUP card are merely gathered together into a common group for the purpose of calculating classification performance. If the training classes are the same as the performance classes, a GROUP card is not required and the performance class names and numbers are set equal to the training class names and numbers.

The GROUP card in Figure 1 creates three performance classes, CORN, SOYBEANS and OTHER; and assigns a name and a number to each of these. The performance class number is the number to the right of the left parenthesis following the performance class name. The numbers enclosed in slashes following performance class number are the training class numbers of the training classes to be grouped into the performance class. The CORN and SOYBEANS performance classes are to be formed from only a single training class each - classes 1 and 2 respectively. However, all of the fields classified as training class 3, 4 or 5 (OW HAT or FOREST) are to be considered in the OTHER performance class - performance class number 3.

Specification of Test Fields: A deck of Field Description cards is required in the input deck. These cards identify the location of the data for each field to be classified - each "test field". Either of two forms of the Field Description Card may be used. They are both described in the Control Card Dictionary in Appendix I.

The function will use the run number specified on the Field Description Card to search the User or System Runtable to identify the appropriate Multispectral Image Storage Tape and have it mounted. (The form and content of this data tape is described in Appendix IV.) Data from the lines and columns specified on the Field Description Card will then be used to calculate the probability density function for the field.

The user must also specify the performance class, number to which the field is assigned. This information is necessary in order to calculate classification performance tables. It is supplied on TEST cards, which contain the keyword TEST in the first four columns, followed by a blank and the performance class number. The TEST card for each field (or group of fields) must immediately precede the Field Description Cards(s) for the field(s).

Figure 1 illustrates the use of TEST cards. The Field Description Cards that define the test fields for each of the three performance classes are grouped behind their appropriate TEST card i.e., the TEST card bearing the appropriate performance class number. The test fields known from ground observations to be corn are grouped behind the 'TEST 1' card, those known to be soybeans behind the 'TEST 2' card and those known to be anything other than corn or soybeans behind the 'TEST 3' card.

If the fields have not been observed on the ground and/or only classification, not performance, is desired all fields can be placed after a single 'TEST 1' card.

The user is cautioned that there are three levels of class names and numbers used in the Sampleclassify function. They are the class names and numbers assigned in the input Statistics File, those specified on CLASSES cards, and those specified on GROUP cards. The following points summarize relationships between these three:

- Training class names and numbers are automatically defined by the Statistics File unless CLASSES cards are used; in which case they are defined by both the Statistics File <u>and</u> the CLASSES cards. If CLASSES cards are used, all training classes to be considered in the classification must be specified on one of these cards; whether or not they are being pooled with other classes.
- Performance class names and numbers are equal to training class names and numbers unless defined differently on a GROUP card.
- Test fields are arranged in the input deck in groups corresponding to the performance class numbers they are assigned to.

- Final class numbers (of any of the three types) are always consecutive integers starting with 1.

Outputs

Standard Printer Output: There are six standard output listings generated by SAMPLECLASSIFY:

1. An Input Deck. The Sampleclassify function first prints a table of information showing how it has interpreted the user's input deck control cards, a list of options the user has selected, and particular characteristics about the run; such as the number of training classes and channels used, the channel numbers used, etc. An example of this output is shown in Figure 2.

2. A classes and channels table. This shows the class name for each training class defined; and the channel number, spectral band, and calibration code for each channel used from the Statistics File. A sample is shown in Figure 3.

3. A Table of Saved Training Fields. This shows training field information for each defined training class. The information includes the run number, field designation, line and column coordinates, field type,

```
DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING          MAR  7,1973
PHILLIPS                             PURDUE UNIVERSITY                          12 19 57 PM
                                                                                LARSYS VERSION 3

      *SAMPLECLASSIFY
      PRINT STATS
      CHANNELS 7,8,12
      CLASSES CORN(1/1,2/),SOYBEANS(2/3,4/)
      CLASSES OW(3/5,8/),HAY(4/6/),FOREST(5/9/)
      GROUP CORN(1/1/),SOYBEANS(2/2/),OTHER(3/3,4,5/)
      DATA

  YOU HAVE SELECTED THE FOLLOWING SAMPLECLASSIFY OPTIONS

      PRINT MULTISPECTRAL STATISTICS

  SAMPLECLASSIFY SUPERVISOR INFORMATION....

      NO. OF POOLS....................    5
      NO. OF GROUPS...................    3
      NO. OF CHANNELS.................    3
      NO. OF TRAINING FIELDS..........   15
      NO. OF FIELDS TO BE CLASSIFIED..   30
      CHANNELS SELECTED ARE........    7,  8, 12,

                                                                                SAM-2
```

Figure 2. Control Card Summary for SAMPLECLASSIFYPOINTS

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
12 27 14 PM
LARSYS VERSION 3

PERFIELD CLASSIFICATION STUDY ..

TRAINING CLASSES

CHANNELS FROM STATISTICS

CLASS	CHANNEL NO.	SPECTRAL BAND		CAL. CODE
CORN	7	0.61	0.70	1
SOYBEANS	8	0.72	0.92	1
OW	12	9.30	11.70	1
HAY				
FOREST				

SAM-3

Figure 3. Classes and Channels Table

and "other information" (supplied on the Field Description Card in the STATISTICS run). The class name assigned in the STATISTICS run, and class name assigned in this SAMPLECLASSIFY run is also shown for each field. A sample of this table is shown in Figure 4.

4. <u>Classification Results in Data Tape Order</u>. This is the first of two listings that give the classification results. The table lists, for each field, the input group name (performance class name), the group it was classified into (training class name), run number, line and column boundaries and the number of points. This information is ordered by increasing line number on the data tape. Also printed as a heading on this output is a Run Identification Table which shows the run information obtained from the ID record on the data tape, and the spectral bands and calibration codes used in the classification. Figure 5 shows a sample of this output.

5. <u>Classification Results in User Input Order</u>. This is the second listing that gives classification results. This one is ordered by field in the same sequence they

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
12 27 15 PM
LARSYS VERSION 3

PERFIELD CLASSIFICATION STUDY ..

SAVED TRAINING FIELDS

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION	STAT. CLASS	POOLED CLASS
1	71053900	L-6	610	628	1	35	52	1	CORN		CORN1	CORN
2	71053900	L-7	619	628	1	58	94	1	CORN		CORN1	CORN
3	71053900	D*5	618	641	1	100	132	1	CORN		CORN2	CORN
4	71053900	H*1	598	612	1	138	169	1	CORN		CORN2	CORN
5	71053900	L-5	619	628	1	23	29	1	SOYBEANS		SOYBEAN1	SOYBEANS
6	71053900	L-8	606	614	1	58	94	1	SOYBEANS		SOYBEAN1	SOYBEANS
7	71053900	D*6	592	612	1	101	131	1	SOYBEANS		SOYBEAN2	SOYBEANS
8	71053900	H*2	591	593	1	137	170	1	SOYBEANS		SOYBEAN2	SOYBEANS
9	71053900	F-9	461	473	1	64	98	1	OATS		OATSCUT	OW
10	71053900	K-4	541	559	1	35	38	1	OATS		OATSCUT	OW
11	71053900	G*2	565	582	1	144	149	1	OATS		OATSCUT	OW
12	71053900	R-1	482	492	1	40	56	1	HAY		HAY	HAY
13	71053900	G*3	562	582	1	156	164	1	WHEAT		WHEAT	OW
14	71053900	H-6	543	562	1	63	72	1	WOODS		FOREST	FOREST
15	71053900	E*2	523	528	1	147	197	1	WOODS		FOREST	FOREST

SAM-4

Figure 4. Saved Training Fields

PERFIELD CLASSIFICATION STUDY ..

RUN NUMBER............ 71053900 DATE DATA TAKEN... AUG 13,1971
FLIGHT LINE... CRN BLT LO FL210 TIME DATA TAKEN..... 1202 HOURS
TAPE/FILE NUMBER....... 1042/ 1 PLATFORM ALTITUDE.. 5000 FEET
REFORMATTING DATE. AUG 16,1971 GROUND HEADING..... 180 DEGREES

CHANNELS USED

CHANNEL 7	SPECTRAL BAND 0.61 TO 0.70 MICROMETERS	CALIBRATION CODE = 1	CO = 28.00
CHANNEL 8	SPECTRAL BAND 0.72 TO 0.92 MICROMETERS	CALIBRATION CODE = 1	CO = 24.85
CHANNEL 12	SPECTRAL BAND 9.30 TO 11.70 MICROMETERS	CALIBRATION CODE = 1	CO = 20.60

	INPUT GROUP	CLASS'D GROUP	RUN NUMBER	FIRST LINE	LAST LINE	LINE INT.	FIRST COL	LAST COL	COL INT.	NUMBER POINTS
1	CORN	CORN	71053900	425	456	1	65	99	1	1120
2	OTHER	CORN	71053900	426	441	1	179	193	1	240
3	SOYBEANS	SOYBEANS	71053900	426	443	1	109	139	1	558
4	OTHER	OTHER	71053900	427	492	1	149	166	1	1188
5	OTHER	OTHER	71053900	428	447	1	30	33	1	80
6	CORN	CORN	71053900	428	445	1	38	59	1	396
7	OTHER	OTHER	71053900	450	475	1	128	142	1	390
8	SOYBEANS	SOYBEANS	71053900	451	475	1	38	58	1	525
9	CORN	CORN	71053900	454	475	1	106	122	1	374
10	SOYBEANS	SOYBEANS	71053900	479	507	1	63	98	1	1044
11	SOYBEANS	SOYBEANS	71053900	481	518	1	106	140	1	1330
12	CORN	CORN	71053900	502	514	1	31	48	1	234
13	SOYBEANS	CORN	71053900	505	515	1	147	169	1	253
14	CORN	CORN	71053900	517	529	1	62	98	1	481
15	SOYBEANS	SOYBEANS	71053900	520	529	1	23	49	1	270
16	SOYBEANS	SOYBEANS	71053900	523	530	1	104	116	1	104
17	OTHER	OTHER	71053900	525	529	1	122	140	1	95
18	OTHER	OTHER	71053900	537	556	1	145	165	1	420
19	OTHER	OTHER	71053900	537	547	1	104	139	1	396
20	CORN	CORN	71053900	540	564	1	26	29	1	100
21	CORN	CORN	71053900	554	564	1	102	139	1	418
22	SOYBEANS	SOYBEANS	71053900	555	564	1	78	94	1	170
23	SOYBEANS	SOYBEANS	71053900	569	584	1	102	137	1	576
24	OTHER	OTHER	71053900	573	585	1	25	53	1	377
25	SOYBEANS	SOYBEANS	71053900	577	584	1	83	93	1	88
26	SOYBEANS	SOYBEANS	71053900	580	585	1	59	65	1	42
27	OTHER	OTHER	71053900	592	600	1	25	94	1	630
28	OTHER	OTHER	71053900	618	640	1	139	169	1	713
29	OTHER	OTHER	71053900	634	641	1	68	94	1	216
30	SOYBEANS	SOYBEANS	71053900	634	642	1	23	60	1	342

SAM-5

Figure 5. Classification Results in Data Tape Order

were ordered in the input deck. It lists the field identification, the name of the performance class to which it was assigned in the input deck, the name of the performance class into which it was classified, the name of the training class (or pooled class) into which it was classified, run number, and line and column information of the field. See Figure 6 for a sample of this output.

6. Classification Performance Summary. Summary information is printed for each performance class showing the number of fields that were input for the class, the percentage of those fields classified correctly, the number of data points in the class and the percentage of data points classified correctly. A matrix is also printed to show the number of fields that were classified into each performance class. For example, the listing in Figure 7 shows that 11 of 12 test fields that were designated as soybeans were classified as soybeans and 1 was classified as corn.

PERFIELD CLASSIFICATION STUDY ..

SUMMARY OF RESULTS. FIELDS IN ORDER SUPPLIED.

FIELD	INPUT GROUP	CLASS'D GROUP	CLASS'D POOL	RUN NUMBER	FIRST LINE	LAST LINE	LINE INT.	FIRST COL	LAST COL	COL INT.
C-5	CORN	CORN	CORN	71053900	428	445	1	38	59	1
F-8	CORN	CORN	CORN	71053900	425	456	1	65	99	1
D*3	CORN	CORN	CORN	71053900	454	475	1	106	122	1
R-2	CORN	CORN	CORN	71053900	502	514	1	31	48	1
R-11	CORN	CORN	CORN	71053900	517	529	1	62	98	1
K-5	CORN	CORN	CORN	71053900	540	564	1	26	29	1
F*3	CORN	CORN	CORN	71053900	554	564	1	102	139	1
L-3	SOYBEANS	SOYBEANS	SOYBEANS	71053900	634	642	1	23	60	1
H-5	SOYBEANS	SOYBEANS	SOYBEANS	71053900	580	585	1	59	65	1
H-4	SOYBEANS	SOYBEANS	SOYBEANS	71053900	577	584	1	83	93	1
R-3	SOYBEANS	SOYBEANS	SOYBEANS	71053900	520	529	1	23	49	1
F*5	SOYBEANS	SOYBEANS	SOYBEANS	71053900	569	584	1	102	137	1
H-3	SOYBEANS	SOYBEANS	SOYBEANS	71053900	555	564	1	78	94	1
D*1	SOYBEANS	SOYBEANS	SOYBEANS	71053900	426	443	1	109	139	1
C-7	SOYBEANS	SOYBEANS	SOYBEANS	71053900	451	475	1	38	58	1
F-10	SOYBEANS	SOYBEANS	SOYBEANS	71053900	479	507	1	63	98	1
O-2	SOYBEANS	SOYBEANS	SOYBEANS	71053900	481	518	1	106	140	1
O-2	SOYBEANS	SOYBEANS	SOYBEANS	71053900	523	530	1	104	116	1
E*3	SOYBEANS	CORN	CORN	71053900	505	515	1	147	169	1
C-6	OTHER	OTHER	OW	71053900	428	447	1	30	33	1
L-2	OTHER	OTHER	OW	71053900	592	600	1	25	94	1
D*4	OTHER	OTHER	OW	71053900	450	475	1	128	142	1
E*2	OTHER	OTHER	FOREST	71053900	427	492	1	149	166	1
E*1	OTHER	CORN	CORN	71053900	426	441	1	179	193	1
I-K-123	OTHER	OTHER	OW	71053900	573	585	1	25	53	1
F*1/2/6	OTHER	OTHER	FOREST	71053900	537	547	1	104	139	1
G*1	OTHER	OTHER	OW	71053900	537	556	1	145	165	1
H*3/4	OTHER	OTHER	HAY	71053900	618	640	1	139	169	1
L-4	OTHER	OTHER	FOREST	71053900	634	641	1	68	94	1
O-3	OTHER	OTHER	FOREST	71053900	525	529	1	122	140	1

SAM-6

Figure 6. Classification Results in User's Input Order

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
12 30 05 PM
LARSYS VERSION 3

PERFIELD CLASSIFICATION STUDY ..

PERFORMANCE SUMMARY

	CLASS	NO OF FLDS	PCT FLD CORRECT	NO OF SAMS	PCT SAM CORRECT	NUMBER OF FIELDS CLASSIFIED AS CORN	SOYBEANS	OTHER
1	CORN	7	100.0	3123	100.0	7	0	0
2	SOYBEANS	12	91.7	5302	95.2	1	11	0
3	OTHER	11	90.9	4745	94.9	1	0	10
	TOTALS	30	93.3	13170	96.3	9	11	10

SAM-7

Figure 7. Classification Performance Summary

Optional Printer Output: The only optional printer output in the Sampleclassify function is the Statistics Summary for the training classes. It is generated by the use of a PRINT STATS control card in the input deck. This summary shows the mean and the standard deviation of the spectral value for each channel, and a correlation matrix of the channels for each class used in the SAMPLECLASSIFY run. Figure 8 shows a sample of this output for the classes CORN and SOYBEANS.

PERFIELD CLASSIFICATION STUDY ..

CLASS....CORN

CHANNEL	7	8	12
SPECTRAL BAND	0.61 - 0.70	0.72 - 0.92	9.30 - 11.70
MEAN	99.57	107.01	96.25
STD. DEV.	11.20	5.01	7.98

CORRELATION MATRIX

SPECTRAL BAND	0.61 - 0.70	0.72 - 0.92	9.30 - 11.70
0.61-0.70	1.00		
0.72-0.92	0.35	1.00	
9.30-11.70	0.80	0.23	1.00

CLASS....SOYBEANS

CHANNEL	7	8	12
SPECTRAL BAND	0.61 - 0.70	0.72 - 0.92	9.30 - 11.70
MEAN	79.47	140.32	86.01
STD. DEV.	5.53	6.13	5.94

CORRELATION MATRIX

SPECTRAL BAND	0.61 - 0.70	0.72 - 0.92	9.30 - 11.70
0.61-0.70	1.00		
0.72-0.92	-0.21	1.00	
9.30-11.70	-0.15	0.37	1.00

Figure 8. Statistics Summary

The Classification Algorithm

In SAMPLECLASSIFY or "per-field classification", a statistical characterization of the data points in a field (any rectangular area on the ground) is calculated and compared against the statistical characterizations of the training classes. Then the field (i.e., the aggregate of points in the field) is classified as a single unit. This is in contrast to the point-by-point classification method in which each observation is given a classification which is assigned independently of all other observations.

Formally, the sample classification procedure may be defined as follows:

> Let $d(\bullet,\bullet)$ be a measure defining the distance between two probability density functions and let $\{p(X|\omega_i),\ i = 1, 2, \ldots, m\}$ be a set of probability density functions corresponding to the classes $\omega_1, \omega_2, \ldots, \omega_m$. If X is a sample (a set of observations) with estimated probability density $p(X|\omega_x)$ then:
>
> Decide $\{X\} \varepsilon\ \omega_i$ if and only if
>
> $d[p(X|\omega_x),\ p(X|\omega_i)] \leq d[p(X|\omega_x),\ p(X|\omega_j)]$
>
> for all $i, j, = 1, 2, \ldots, m$.

For writing the definition of the distance measure used (Jeffries-Matusita distance [JM distance], see LARS INFO NOTE 100771), it is convenient to use an abbreviated notation for the density functions. Let

$$p_i(X) = p\ (X|\omega_i).$$

Then the JM distance between density functions $p_1(X)$ and $p_2(X)$ is given by

$$d[p_1(X),p_2(X)] \triangleq \left[\int_X \left(\sqrt{p_1(X)} - \sqrt{p_2(X)}\right)^2 dX\right]^{1/2} \tag{1}$$

where the integral is over the entire multi-dimensional space of X. By defining

$$\rho(p_1,p_2) = \int_X \sqrt{p_1(X)}\ .\ \sqrt{p_2(X)}\quad dX \tag{2}$$

the JM distance can be expressed as

$$d[p_1(X),p_2(X)] = [2(1-\rho(p_1,p_2))]^{1/2} \tag{3}$$

In the case of gaussian distributions with class mean vectors M_i, covariance matrices K_i, and a sample with mean M_x and covariance matrix K_x, Eq. (2) can be written in the form

$$\rho(p_x,p_i) = \frac{|K_x^{-1}K_i^{-1}|^{1/4}}{|\frac{1}{2}(K_x^{-1}+K_i^{-1})|^{1/2}} \cdot$$

$$\exp\left[-\frac{1}{4}\ \{-[(K_x^{-1} + K_i^{-1})^{-1}(K_x^{-1}M_x + K_i^{-1}M_i)]^T[K_x^{-1}M_x + K_i^{-1}M_i] + M_x^T\ K_x^{-1}M_x + M_i^TK_i^{-1}\ M_i\}\right] \tag{4}$$

It is significant that this expression can be evaluated without performing explicit integration.

The implementation of the algorithm in SAMPLECLASSIFY is such that the JM distance $-d[p_1(X),p_2(X)]$ - is not actually calculated. Instead the value of $\ell n(\rho(p_x,p_i))$ is computed and used. This is possible since the only use of the distance function is in comparisons, and the larger the value of $\ell n(\rho(p_x,p_i))$ the smaller the value of the JM distance. The computation takes the following form in the program, with n representing the number of degrees of freedom (dimension of K).

$$\ell n(\rho(p_x,p_i)) = \ell n\left(\frac{2^{\frac{n}{2}}}{(|K_x K_i|)^{\frac{1}{4}}}\right) - \ell n\left(\left|K_x^{-1}+K_i^{-1}\right|^{\frac{1}{2}}\right) + \left[-\frac{1}{4}\{-[(K_x^{-1}+K_i^{-1})^{-1}(K_x^{-1}M_x+K_i^{-1}M_i)]^T[K_x^{-1}M_x+K_x^{-1}M_i] + M_x^T K_x^{-1} M_x + M_i^T K_i^{-1} M_i\}\right] \quad (5)$$

In practice, the M's and K's are usually not known, and estimates are used which are obtained from training patterns and from the sample to be classified.

Reference:

Wacker A.G., Landgrebe D.A., "The Minimum Distance Approach to Classification", LARS Information Note 100771

SECHO (Supervised Echo Classifier)

The SECHO 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 classification algorithm and thereby increase the classification accuracy.

The SECHO processor benefits 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 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. SECHO makes use of pre-specified multivariate normal class distributions to identify homogeneous objects.

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 step approach a Classification Results File, which may be placed either on tape or on disk. When the two step 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 step approach utilizes the Intermediate Results Tape and the user-specified annexation criteria* to produce the Classification Results File. The advantage of the two step 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 step approach, all processing listed above is accomplished without the need of an Intermediate Results Tape.

* Analyst experience is required when specifying the homogeneity and annexation parameters. These two parameters greatly affect the results.

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 sorted on tape for use by the latter two LARSFRIS 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 non-homogeneous (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 LARSFRIS 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 LARSFRIS Runtable File can be found in the LARSFRIS 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 LARSFRIS functions that uses the statistics information or by including the statistics information in the control card file. Any of LARSFRIS 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 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 1.

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 1 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 2 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.

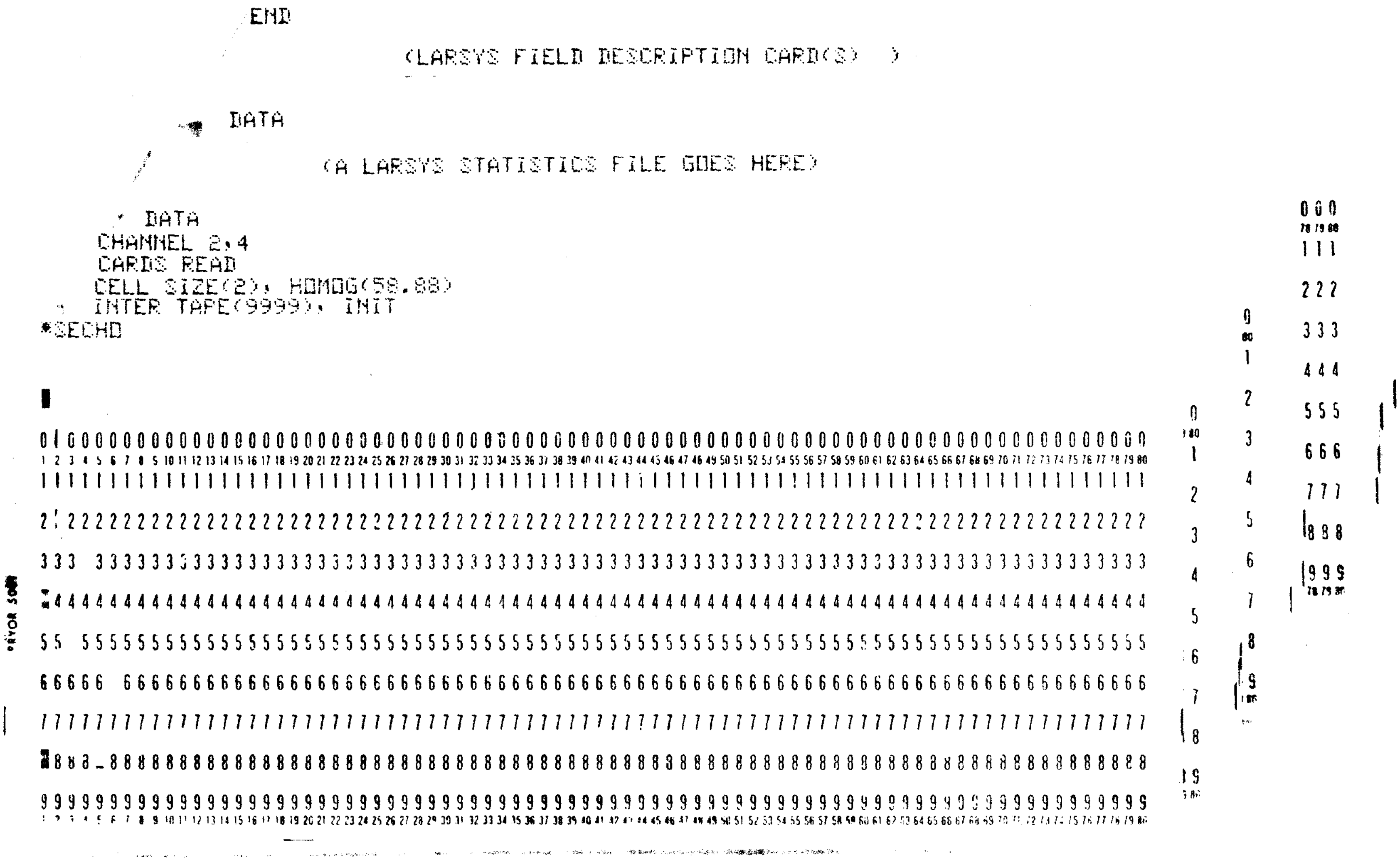


Figure 1. Example Control Cards for executing the SECHO Cell Processing Phase (Phase 1).

Figure 2. Example Control Cards for executing the single step approach of SECHO (Phase 1 and Phase 2).

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 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 3 is an example of the control cards necessary to run phase 2 ECHO classification. Execution of the control cards in Figure 1 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 chase 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 LARSFRIS User's Manual [8] contains information on how this card may also be used to calibrate data from the Multispectral

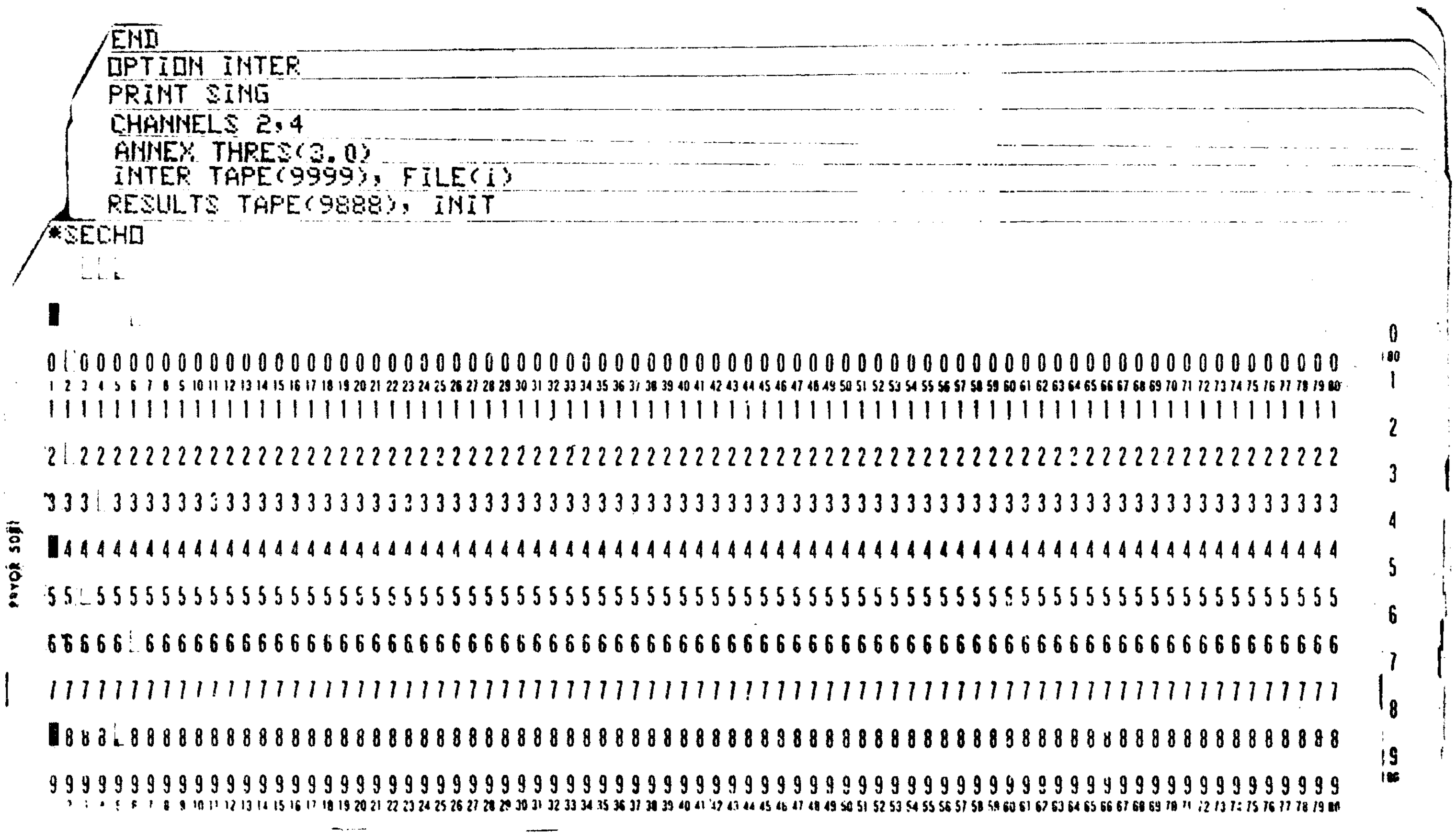


Figure 3. Example Control Cards for executing the SECHO Annexation Phase (Phase 2.)

Image Storage Tape.

Optional Selection of Training Classes

The user may select the training classes from the Statistics File that 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

In this case, the class name assigned by the statistics function as 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 for both object identification and for classification.

Specification of Annexation Parameter

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)

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 of 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 LARSFRIS 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 LARSFRIS 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 is 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 LARSFRIS LISTRESULTS or PUNCHSTATISTICS functions, he must place it on tape or have it copied to tape by the COPY-RESULTS 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

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 year001-365), and sssss is the total number of seconds since the previous mid-night.

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 LARSFRIS 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 LARSFRIS 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 statistics (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 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 LARSFRIS 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 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.

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 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 INFORMATION" in Figure 4 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 also given. Note that in this case 27 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:

Figure 4. SECHO Summary Information.

```
     BUIS                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING            SEPT  9,1980
     JB                                  PURDUE UNIVERSITY                            10 05 24 AM
                                                                                      LARSYS VERSION 3

   *SECHO
   ANNEXATION   THRES(3.0)
   RESULTS TAPE(9999), INIT
   CELL SIZE(2), HOMOG(58.88)
   CARDS READ
   PRINT SING
   SYMBOLS  ,S, ,D, ,W
   CHANNELS 2,4
   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........... 25336403
   NO.OF POOLS....................    6
   NO.OF CHANNELS.................    2
   NO. OF TRAINING FIELDS........  27

-  -  -  -  -
```

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 Figure 5.

2. A Processing Parameters List. Figure 5 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. 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.

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

SEPT 9,1980
10 06 46 AM
LARSYS VERSION 3

TRAINING CLASSES

CHANNELS FROM STATISTICS

CLASS	CHANNEL NO.	SPECTRAL BAND		CAL. CODE
PASMEDAS	2	0.0	0.0	1
SMARSH	4	50.00	0.60	1
MUDSHOR				
DMARSH				
(				

INNNN DATAWIS IN LARSYS FORMAT (GADRUN)

PROCESSING PARAMETERS

CELL WIDTH = 2

NUMBER OF CHANNELS = 2

NUMBER OF POOLED CLASSES = 6

ANNEXATION THRESHOLD = 3.0000E 00

CELL HOMOGENEITY THRESHOLD = 5.8880E 01

NUMBER OF CELL LINES IN BUFFER = 130

\- - - - -

Figure 5. Example SECHO Classes and Channels Table and Processing Parameter List.

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 LARSFRIS 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 represented by an alphanumeric symbol (a number, character or special symbol). Figure 6 presents the classification map which would result from Figure 4 if "class" was specified instead of "sing". 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 LARSFRIS PRINTRESULTS function. The reader should refer to the description of that function in the LARSFRIS User's Manual 8 for an example PRINTRESULTS output.

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 7 is a Singular Cell Map of the same

RUN NUMBER............ 79005600
FLIGHT LINE... 17092 SASK. CAN.
DATA TAPE/FILE NUMBER.. 3780/ 1
REFORMATTING DATE. JULY 7,1980

DATE DATA TAKEN... JULY 9,1979
TIME DATA TAKEN..... 1709 HOURS
PLATFORM ALTITUDE..3062000 FEET
GROUND HEADING..... 0 DEGREES

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.0 TO 0.0 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 4	SPECTRAL BAND 50.00 TO 0.60 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0

AREA PROCESSED.....LINES 575- 624 (BY 1)
COLUMNS 2475- 2570 (BY 1)

EOF RECORD WRITTEN.

I0103 CPU TIME USED WAS 7.073 SECONDS. (LARSMN)

Figure 6. Pictorial Classification Map.

area as that represented on the Pictorial Classification Map in Figure 6. 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 7, 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.

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.

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

SEPT 9,1980
10 07 30 AM
LARSYS VERSION 3
SEC-25

RUN NUMBER........... 79005600	DATE DATA TAKEN... JULY 9,1979
FLIGHT LINE... 17092 SASK. CAN.	TIME DATA TAKEN..... 1709 HOURS
DATA TAPE/FILE NUMBER.. 3780/ 1	PLATFORM ALTITUDE..3062000 FEET
REFORMATTING DATE. JULY 7,1980	GROUND HEADING..... 0 DEGREES

CHANNELS USED

CHANNEL 2	SPECTRAL BAND 0.0 TO 0.0 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 4	SPECTRAL BAND 50.00 TO 0.60 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0

```
AREA PROCESSED.....LINES    575-  624  (BY  1)
                   COLUMNS 2475- 2570  (BY  1)

    222222222222222222222222222222222222222222222222
    444444444444455555555555555555555555555555555555
    777888889999900000111112222233333444445555566666
    579135791357913579135791357913579135791357913579

575                                 0     0
577                                 0           00
579                                00       000000
581                               00
583                                    0
585                                           0
587                                  00
589
591                        0
593
595         00        0
597           00
599        0 0
601                                      0
603     0    0                            0
605                                    0  0
607   0                0                0
609  0                             0
611  00                   0
613  00                                       0
615   0                   0
617                        0               0
619         00
621                                             0
623        0                    0  0
```

EOF RECORD WRITTEN.

I0103 CPU TIME USED WAS 6.833 SECONDS. (LARSMN)

Figure 7. Example SECHO Singular Cell Map.

SEPARABILITY

The Separability function helps the user select the set of channels that will produce the most accurate classification by the Classifypoints function. It uses class statistics to calculate measurements of how well the individual classes may be distinguished from one another, or the degree of "separability" between the classes.

This function is implemented by using the mean vectors and covariance matrices for classes to calculate the divergence (a measure of distance between class densities) of all class pairs for each set of channels. The transformed divergence, whose saturation effect reduces the influence of very widely separated class pairs, is computed. The function then ranks each set of channels by either average or minimum transformed divergence of the class pairs, depending on the choice of the user. In either case, the ranking will be by greatest transformed divergence value, i.e., the class pair having the largest average transformed divergence or the pair having the largest minimum transformed divergence will be ranked first.

In the pages that follow, "divergence" is used to refer to either the divergence or to the transformed divergence. Distinction between the two is made only when it is necessary to avoid confusion.

Inputs to the function are:

- Control cards to select the processing and output options.
- A Statistics File from the Statistics function to provide class means and covariance matrices.

The initial divergence calculations are written on a scratch (or work) file. This is normally a disk file, but the function estimates the space requirements before starting processing; and if the estimated requirement is greater than the available disk storage, a tape is used instead. A number of options are available to control the type and amount of printed output and the classes and channels that are used to calculate divergences.

The user may also use the interactive mode of this function to rerank the divergences after varying one or more of the optional parameters. This feature is implemented by the 'OPTIONS TYPE' control card which causes a programmed pause at the end of the processing for a set of channels. The keyboard is unlocked, and the user may then type in any of a specific set of "control cards" which modify the processing options. This feature will be discussed in greater detail later in this functional description.

The principal output of SEPARABILITY is the Separability Results Listing. An example of part of this listing is shown in Figure 1. The listing resulted from a user's request for the best thirty combinations of channels, taken three at a time, for nine classes. For each channel combination it shows the minimum weighted divergence (DIJ(MIN)), the average of the weighted divergences (D(AVE)), and the weighted interclass divergence for each class pair (DIJ). In this particular listing the channel combinations were ranked in order of decreasing average interclass divergence (D(AVE)). An option of ranking them by decreasing minimum divergence (DIJ(MIN)) is also available.

Since there were nine classes considered (represented by the symbols A,B,1,2,O,H,S,W, and F) in this example, there were thirty-six class pairs. This required two more pages of output to list all of the interclass divergence values. The last two pages are not shown.

A record of the highest ranking combination of channels (ranked number 1) is saved at the end of the function and may be automatically used in place of the "CHANNELS" specification in the Classifypoints function. This is accomplished by including an AUTO CHANNELS card in the input deck for the Classifypoints function. This feature is usable only when the execution of CLASSIFYPOINTS immediately follows that of SEPARABILITY.

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 24 09 AM
LARSYS VERSION 3

RETENTION LEVEL .. 220 MAXIMUM 30000
MINIMUM 0

DIVERGENCE **WITH** SATURATING TRANSFORM

	CHANNELS	DIJ(MIN)	D(AVE)	AB (10)	A1 (10)	A2 (10)	AO (10)	AH (10)	AS (10)	AW (10)	AF (10)	B1 (10)	B2 (10)
				WEIGHTED INTERCLASS DIVERGENCE (DIJ)									
1.	2 9 12	679.	1840.	979	2000	2000	1984	2000	1541	1928	1960	2000	2000
2.	7 9 12	734.	1840.	907	1998	2000	1983	2000	1396	1933	1957	2000	2000
3.	7 8 12	738.	1837.	738	1994	2000	1976	2000	1289	1934	1991	2000	2000
4.	2 8 12	763.	1832.	763	1997	2000	1971	2000	1361	1923	1976	2000	2000
5.	6 9 12	663.	1825.	913	1999	2000	1976	2000	1187	1898	1926	2000	2000
6.	4 9 12	720.	1818.	898	1998	2000	1946	2000	1005	1878	1836	2000	2000
7.	1 9 12	620.	1815.	925	1999	2000	1965	2000	1276	1864	1909	2000	2000
8.	6 8 12	732.	1813.	732	1994	2000	1965	2000	967	1890	1943	2000	2000
9.	5 9 12	586.	1812.	888	1998	2000	1947	2000	1052	1843	1823	2000	2000
10.	1 8 12	736.	1812.	736	1995	2000	1955	2000	1161	1881	1945	2000	2000
11.	4 8 12	726.	1812.	726	1994	2000	1926	2000	908	1897	1929	2000	2000
12.	5 8 12	714.	1809.	714	1993	2000	1941	2000	903	1870	1948	2000	2000
13.	3 9 12	572.	1809.	911	1999	2000	1942	2000	1014	1802	1903	2000	2000
14.	8 9 12	824.	1808.	926	1997	1999	1943	2000	824	1739	1874	2000	2000
15.	7 10 12	544.	1805.	544	1995	2000	1977	2000	1281	1939	1958	1997	1999
16.	3 8 12	720.	1805.	720	1995	2000	1946	2000	872	1845	1971	2000	2000
17.	2 10 12	454.	1791.	454	1995	1999	1914	2000	1340	1892	1941	1995	1990
18.	9 10 12	795.	1791.	860	1996	2000	1916	2000	1152	1772	1645	2000	2000
19.	9 11 12	647.	1784.	834	1998	2000	1945	2000	1185	1861	1636	2000	2000
20.	8 10 12	740.	1783.	740	1993	1999	1933	2000	883	1772	1716	2000	2000
21.	8 11 12	732.	1781.	732	1996	1999	1931	2000	931	1827	1737	2000	2000
22.	6 10 12	540.	1777.	540	1993	2000	1917	2000	1010	1829	1866	1996	1998
23.	4 10 12	459.	1773.	459	1978	1995	1800	2000	1027	1802	1749	1995	1987
24.	5 10 12	482.	1765.	482	1985	1999	1807	2000	999	1744	1751	1995	1994
25.	7 11 12	402.	1761.	402	1994	1999	1860	2000	1066	1849	1924	1991	1995
26.	1 10 12	407.	1759.	407	1985	1992	1852	2000	1091	1798	1858	1995	1980
27.	4 8 10	453.	1757.	453	1994	2000	1754	1993	694	1238	1880	2000	2000
28.	3 10 12	452.	1753.	452	1987	1998	1760	2000	889	1671	1874	1995	1988
29.	4 9 10	524.	1751.	524	1996	2000	1753	1999	561	1270	1795	1999	2000
30.	4 11 12	417.	1745.	417	1981	1987	1615	2000	1033	1825	1711	1986	1947

SEP-1

Figure 1. Sample Output of the SEPARABILITY Results Listing

There are three standard printed outputs in addition to the Separability Results Listing. They are: a listing of the control cards and options selected, a Classes and Channels Table, and a Saved Training Fields Listing. The user may also request a Statistics Summary, which lists the statistics for the classes considered, and a special listing of all class pairs whose maximum divergence for every channel combination printed on the Separability Results Listing is less than a specified value. All of these outputs are described in greater detail later in this function description.

Inputs

Principal Inputs: The Statistics File, which contains a statistical description of classes, is the primary input to the Separability function. It may be input as a card deck or as a disk file. SEPARABILITY assumes the statistics are on disk and will automatically look there, unless the 'CARD READSTATS' control card is present in the input deck. If the Statistics File is to be input on cards, it must be preceded by a 'DATA' card and must come after all control cards and before the 'END' card.

The user must provide an input deck which selects the Separability function and specifies where the statistics data may be found, how the processing is to be performed, and what outputs are to be generated. It is always terminated by an END card. He may use control cards to select the channels to be considered, select the training classes to be used and/or group the classes into pooled classes, indicate the Statistics File which will be entered on punched cards, request a specific channel combination be printed regardless of its rank, provide weights for class pairs, and request various types and amounts of printed output. Data cards may be used in the input deck to provide the required Statistics File.

A sample input deck, which uses two SEPARABILITY runs to illustrate several of these options, is shown in Figure 2. It will be described further in the discussion which follows.

Selection of Classes: By using the CLASSES control card, the user may select the specific classes or pool classes from the Statistics File, and have them be considered by the function. If the CLASSES card is omitted, all training classes in the Statistics File are used. The statistics for pooled classes are combined to form each new "class". For instance, if the user wishes only to include classes 2, 3 and 5 in a run, he would use the control card:

CLASSES 2, 3, 5

A different format of the CLASSES card is used, however, to group statistics classes into pooled classes. The second example in Figure 2 shows two CLASSES cards that create five pooled classes

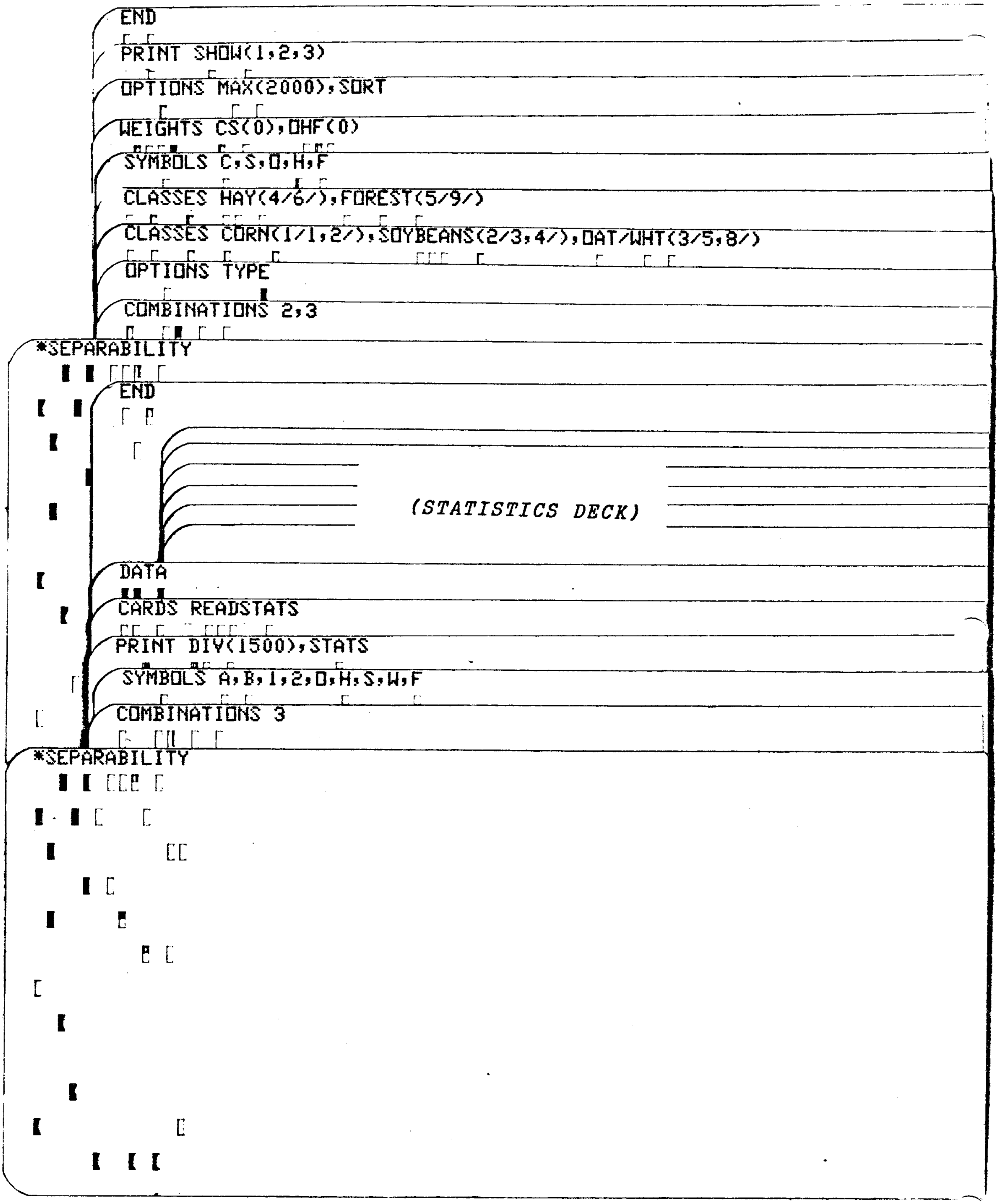


Figure 2. Input Deck for SEPARABILITY

from nine classes that are present in the Statistics File. When the cards are processed, the following classes will result:

- Class 1, named CORN, will be defined by pooled statistics for Classes 1 and 2 from the Statistics File.
- Class 2, SOYBEANS, will be defined from original classes 3 and 4.
- Class 3, OAT/WHT, will be defined from classes 5 and 8.
- Class 4, HAY, will be defined from class 6.
- Class 5, FOREST, will be defined from class 9.

Note that in the CLASSES card in Figure 2, the numbers immediately following a left parenthesis specify pooled class numbers and must be in ascending order. The statistics classes to be pooled (and named) are enclosed by slashes(/). Class 7 in the Statistics File was not used.

Specification of Symbols: The user May supply a symbol to be used on the printed output to represent each defined class. The format for the SYMBOLS card is:

SYMBOLS $S_1,S_2,S_3,...$

where S_1 corresponds to the symbol to be used for the first class (or pooled class), S_2 the symbol for the second, etc. Any symbol may be used. The SYMBOLS cards in Figure 2 use nine symbols to represent nine statistics classes for the first run, and five symbols to represent five pooled classes in the second run. These symbols appear on the Separability Results Listings to identify the class pairs. Figure 1 shows this usage for some of the class pairs in the first run in Figure 2. If symbols are not specified, they will default to A,B,C,..... .

Optional Specification of Weights: Each class pair has a weight assigned to it that is used to calculate the weighted divergence values for the Separability Results Listing. All class pairs are assigned a default weight of 10. If the user would like certain class pairs to be weighted more or less, he may assign them the weight of his choice. The 'WEIGHTS' card is used for this. The format is:

WEIGHTS $S_iS_jS_k$,...(weight 1),S_lS_m...(weight 2),...

The weight specified as 'weight 1' is assigned to all class pairs formed from the classes represented by symbols $S_iS_jS_k$,...; the weight specified as 'weight 2' is assigned to all class pairs formed from the classes represented by symbols S_l,S_m,...; etc.

The WEIGHTS card in Figure 2 will cause the class pairs represented by symbol pairs CS,OH,OF and HF to have a weight of zero. The zero means that their divergences will not be considered in any succeeding calculations such as the average weighted interclass divergence (D(AVE)), and the individual interclass divergences will not be printed. All other class pairs will have the default weight of ten.

<u>Specification of Channels</u>: The channels to be used by the function may be specified with the CHANNELS control card:

CHANNELS K,J,...

where I,J,... are the channel numbers of the channels to be used.

If no CHANNELS card is included, all of the channels in the Statistics File will be used and only the channels that are represented in the Statistics File can be selected. A CHANNELS card was not included in the example input deck in Figure 2. As a result, all twelve channels in the input Statistics File will be used in the calculations. Note that since calibration of data cannot be defined for this function, only the short form of the CHANNELS card can be used. If the user wishes to experiment with calibration, he must do so in the execution of the Statistics function that creates the Statistics File he uses in SEPARABILITY.

<u>Specification of Channel Combinations</u>: The COMBINATIONS control card, which must be provided, is used to specify the number of channels to be considered (out of the total number specified on the CHANNELS card) for each calculation.

For instance, in the input deck in Figure 2, there are twelve channels specified on the CHANNELS card. The COMBINATIONS card in the second run contains a 2 and a 3. This specifies that all combinations of 2 channels out of 12 are to be evaluated and ranked and the listing produced. The process is then to be repeated for all combinations of 3 out of 12 channels.

Other Processing and Printing Options: The OPTIONS control card provides further control of the Separability Results Listing and the processing that produces it. Five of the six possible control parameters on the card are for this purpose. The sixth parameter ('OPTIONS TYPE') activates the interactive mode of processing, which will be described later under "Interactive Control".

SEPARABILITY uses specified "maximum" and "minimum" divergence values to control the contributions of weighted interclass divergences on the ranking of channel combinations. If, for any channel combination, the divergence for any class pair is equal to or greater than the specified "maximum" value, then the "maximum" value is substituted for the actual divergence value in the ranking algorithm, the divergence value for that pair is not printed, and a series of three periods (...) is printed instead. If a particular channel combination produces a divergence for one or more of the class pairs that is less than the "minimum" value, the entire line of divergence values for that channel combination is blanked out on the listing. The identification of the channel combination is printed, however, in its normal position in the list.

The function uses default values of 30,000 for the maximum divergence and zero for the minimum divergence - effectively printing all values, as shown in Figure 1. The user may vary these values through two parameters on the 'OPTIONS' control card. The 'MAX(value)' and the 'MIN(value)' parameters set the maximum and minimum to the specified values. For instance, if the user were to include a control card:

```
OPTIONS MAX(2000),MIN(1800)
```

the divergence values equal to or greater than 2,000 would be replaced with the three periods on the listing, and any channel combination having at least one class pair divergence that was less than 1,800 would be blanked out.

The 'EXCLUDE' parameter on the OPTIONS card excludes from consideration any combinations of channels that include the channels that are specified. For example, if the user had a control card in his input deck that specified:

```
OPTIONS EXCLUDE(7,8)
```

then any set of channels that contain both channel 7 and channel 8 would be excluded from consideration for each calculation of channel combinations.

The 'OPTIONS SORT' control card changes the basis for ranking the channel combinations and printing the results on the Separability Results Listing. The channel combinations are normally ranked in the order of decreasing average weighted interclass divergence, D(AVE) on the listing. The 'OPTIONS SORT' control card causes the channel combinations to be ranked instead by decreasing minimum divergence, "DIJ(MIN)" on the listing.

SEPARABILITY normally uses the transformed divergence results for all calculations and printing. If the user includes the 'OPTIONS UNTRANS' control card, the untransformed divergence will be used instead.

The PRINT Control Card: The 'BEST' and 'SHOW' parameters on the PRINT control card enable the user to control channel combinations that are printed on the Separability Results Listing. A third parameter, the 'DIV(value)' parameter, prints as an optional extension of the results listing, a list of class pairs with low separability. The last parameter, 'STATS', produces an optional Statistics Summary which will be described later.

The Separability Results Listing normally contains the best 30 channel combinations, or all of them if the total number of combinations is less than 30. If the user wishes to vary this default value, he can use the 'PRINT BEST(n)' control card, where "n" is the number of combinations he wishes printed. Since there is no such control card in the sample input deck in Figure 2, the outputs will show the best thirty combinations.

The user also may cause one or more specific channel combinations to be printed, with its ranking, by including the combination on a 'PRINT SHOW(f1,f2,.....)' control card, where "f1" is the number of the first channel, "f2" the number of the second, etc., until all channels in the combination are specified. The sample input deck in Figure 2 specifies the printing of the combination of channels 1, 2 and 3 regardless of its rank.

Outputs:

Divergence Scratch File: The initial divergence calculations are output on disk if there is enough space and on tape if there is not. The calculations written are the divergences for each pair of classes for each channel combination. A separate record is written for each channel combination and contains the set of channels, the sum of the divergences for that combination, and the individual divergences for each class pair. These results are later read and processed by the function for each set of options specified by the user. For example, the pairwise divergences will be multipled by the designated weights and these results will be ranked for printing.

Standard Printer Output: The Separability function always prints a list showing the input control cards, the options selected by the user, and a summary of the function's interpretation of the input deck. An example of this output is illustrated in Figure 3. This is always followed by two other standard printed outputs:

A Classes and Channels Table

This output gives the symbol and class name for each training class; and the channel number, spectral band, and calibration code from the Statistics File for each channel used. Figure 4 illustrates the table that was produced by the sample input deck.

Saved Training Fields Listing

This listing shows the training fields from the Statistics File and their run number, field designation, line and column coordinates, field type, the class name assigned in the Statistics run, and the pooled class name assigned if a CLASSES card is included in the input deck. A sample of the listing that was produced by the first run in the sample input deck is shown in Figure 5. The second run produced a similar listing except that field E*1 (class 7) was omitted.

```
     DEMO                     LABORATORY FOR APPLICATIONS OF REMOTE SENSING          MAR  7,1973
     PHILLIPS                               PURDUE UNIVERSITY                        11 23 42 AM
                                                                                     LARSYS VERSION 3

    *SEPARABILITY
    COMBINATIONS 3
    SYMBOLS A,B,1,2,O,H,S,W,F
    PRINT DIV(1500),STATS
    CARDS READSTATS
    DATA

 YOU HAVE SELECTED THE FOLLOWING OPTIONS

    PRINT MULTISPECTRAL STATISTICS
    PRINT CLASS PAIRS WITH DIVERGENCE LESS THAN 1500

  SEPARABILITY SUPERVISOR INFORMATION....

    NO. OF CLASSES.................   9
    NO. OF CHANNELS................  12
    NO. OF COMBINATIONS REQUESTS...   1
    NO. OF SHOW REQUESTS...........   0
    NO. OF WEIGHT REQUESTS.........   0
    CHANNELS SELECTED ARE..........   1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12,
```

SEP-3

Figure 3. Control Card Summary for SEPARABILITY

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 24 04 AM
LARSYS VERSION 3

SEPARABILITY STUDY

CLASSES CONSIDERED		CHANNELS CONSIDERED			
SYMBOL	CLASS	CHANNEL NO.	SPECTRAL BAND		CAL. CODE
A	CORN1	1	0.46	0.49	1
B	CORN2	2	0.48	0.51	1
1	SOYBEAN1	3	0.50	0.54	1
2	SOYBEAN2	4	0.52	0.57	1
O	OATSCUT	5	0.54	0.60	1
H	HAY	6	0.58	0.65	1
S	SA	7	0.61	0.70	1
W	WHEAT	8	0.72	0.92	1
F	FOREST	9	1.00	1.40	1
		10	1.50	1.80	1
		11	2.00	2.60	1
		12	9.30	11.70	1

SEP-4

Figure 4. Classes and Channels Table

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 24 04 AM
LARSYS VERSION 3

SEPARABILITY STUDY

SAVED TRAINING FIELDS

	RUN NUMBER	FIELD DESIG.	FIRST LINE	LAST LINE	LINE INT.	FIRST COLUMN	LAST COLUMN	COLUMN INT.	FIELD TYPE	OTHER INFORMATION	STAT. CLASS	POOLED CLASS
1	71053900	L-6	610	628	1	35	52	1	CORN		CORN1	CORN1
2	71053900	L-7	619	628	1	58	94	1	CORN		CORN1	CORN1
3	71053900	D*5	618	641	1	100	132	1	CORN		CORN2	CORN2
4	71053900	H*1	598	612	1	138	169	1	CORN		CORN2	CORN2
5	71053900	L-5	619	628	1	23	29	1	SOYBEANS		SOYBEAN1	SOYBEAN1
6	71053900	L-8	606	614	1	58	94	1	SOYBEANS		SOYBEAN1	SOYBEAN1
7	71053900	D*6	592	612	1	101	131	1	SOYBEANS		SOYBEAN2	SOYBEAN2
8	71053900	H*2	591	593	1	137	170	1	SOYBEANS		SOYBEAN2	SOYBEAN2
9	71053900	F-9	461	473	1	64	98	1	OATS		OATSCUT	OATSCUT
10	71053900	K-4	541	559	1	35	38	1	OATS		OATSCUT	OATSCUT
11	71053900	G*2	565	582	1	144	149	1	OATS		OATSCUT	OATSCUT
12	71053900	R-1	482	492	1	40	56	1	HAY		HAY	HAY
13	71053900	E*1	426	441	1	179	193	1	OTHER	SET ASIDE	SA	SA
14	71053900	G*3	562	582	1	156	164	1	WHEAT		WHEAT	WHEAT
15	71053900	H-6	543	562	1	63	72	1	WOODS		FOREST	FOREST
16	71053900	E*2	523	528	1	147	197	1	WOODS		FOREST	FOREST

Figure 5. Saved Training Fields Listing

The principal output from the Separability function is the Separability Results Listing. It shows the channel combinations ordered according to degree of separability (most separable listed first), with the associated divergences for each class pair for each channel combination. In a typical run the listing shows the best 30 channel combinations and class divergences listed by channel combinations ranked according to average pairwise divergence between classes. A separate listing will be produced for each set of channel combinations that was specified.

Four control cards can be used to vary the processing that is performed and the form and content of this listing. These are the SYMBOLS, WEIGHTS, PRINT and OPTIONS cards. The use of these control cards have been explained below in the "Input" subsection. Examples of Separability Results Listings that were produced by the second run in the sample input deck are shown in Figures 6 and 7. Figure 6 is the listing of the best two channels out of twelve, and Figure 7 is the listing of the best three channels out of twelve.

Listing Classes with Low Separability: The user may request an extension to the Separability Results Listing that identifies all class pairs with a low (as defined by the user) degree of separability. The 'PRINT DIV(value)' control card causes a special listing to be produced that shows the class pair identification and the average pairwise divergence for all class pairs whose greatest divergence for all channel combinations printed was less than or equal to the 'value' specified by the user. This listing appears directly after the Separability Results Listing and is identified by a header line that reads:

CLASSES THAT MAY BE COMBINED-MAX DIV. = value

Figure 8 shows the listing that was produced by the 'DIV (1500)' parameter on the PRINT control card in the first example in the sample input deck in Figure 2. It identifies four class pairs, AB, 12, OW and SW, whose maximum divergence for all channel combinations printed was less than or equal to 1,500. The user may wish to verify these results for pair AB by referring to Figure 1. The number printed after the class pair symbols is the average of the weighted divergence values for all printed channel combinations for that class pair.

Optional Printer Output: The user may have a Statistics Summary printed by including a 'PRINT STATS' control card in his input deck. This listing includes, for each class or pooled class that is being considered, the mean and standard deviation of the spectral response for each channel and each class, and a channels correlation matrix for each class. Figure 9 shows one of the nine Statistics Summaries (one for each class) that was produced by the first run in the sample input deck in Figure 2.

DEMO PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 25 12 AM
LARSYS VERSION 3

RETENTION LEVEL .. 66 MAXIMUM 2000
MINIMUM 0

DIVERGENCE **WITH** SATURATING TRANSFORM

*** RESULTS ORDERED ACCORDING TO DIJ(MIN) ***

	CHANNELS	DIJ(MIN)	D(AVE)	WEIGHTED INTERCLASS DIVERGENCE (DIJ)									
				CS (0)	CO (10)	CH (10)	CF (10)	SO (10)	SH (10)	SF (10)	OH (0)	OF (0)	HF (0)
1.	4 10	1415.	1597.		1538	1609	1444	1851	1415	1722			
2.	7 10	1377.	1656.		1399	1377	1922	1518	...	1721			
3.	5 10	1355.	1636.		1355	1478	1531	1777	1975	1699			
4.	4 11	1328.	1555.		1328	1605	1390	1851	1412	1746			
5.	6 10	1275.	1627.		1275	1326	1761	1708	...	1694			
6.	4 9	1238.	1763.		1238	1995	1520	1955	...	1870			
7.	4 8	1184.	1798.		1184	1986	1695	1981	...	1944			
8.	3 10	1141.	1558.		1141	1315	1712	1561	1887	1734			
9.	2 4	1063.	1561.		1320	1468	1731	1861	1920	1063			
10.	4 12	1061.	1632.		1405	...	1342	1985	...	1061			
11.	2 10	1052.	1515.		1317	1136	1841	1052	1977	1764			
12.	5 9	1043.	1726.		1043	1997	1547	1910	...	1858			
13.	1 4	1032.	1488.		1276	1305	1633	1867	1812	1032			
14.	4 6	1029.	1584.		1216	1708	1662	1887	1999	1029			
15.	2 12	1004.	1688.		1458	...	1750	1916	...	1004			
16.	5 11	988.	1557.		988	1445	1445	1780	1953	1729			
17.	4 7	982.	1671.		1394	1878	1886	1883	...	982			
18.	3 4	971.	1400.		1146	1028	1626	1847	1781	971			
19.	5 8	963.	1775.		963	1990	1796	1951	...	1951			
20.	4 5	939.	1399.		1086	1178	1409	1863	1919	939			
21.	8 12	926.	1741.		1579	...	926	1995	...	1946			
22.	3 12	924.	1618.		1228	...	1595	1959	...	924			
23.	7 11	920.	1578.		920	1392	1886	1527	...	1740			
24.	1 12	868.	1616.		1340	1999	1588	1899	...	868			
25.	2 11	867.	1438.		867	1151	1810	1065	1958	1776			
26.	2 5	840.	1427.		955	1301	1712	1820	1935	840			
27.	9 10	837.	1644.		1145	1999	837	1985	...	1896			
28.	6 9	836.	1730.		836	1998	1807	1870	...	1868			
29.	8 11	833.	1590.		833	1993	860	1916	...	1935			
30.	9 12	816.	1710.		1562	...	816	1998	...	1881			

Figure 6. SEPARABILITY Results Listing for the Best Two Channels out of Twelve

DEMO PHILLIPS | LABORATORY FOR APPLICATIONS OF REMOTE SENSING PURDUE UNIVERSITY | MAR 7,1973 11 25 12 AM LARSYS VERSION 3

RETENTION LEVEL .. 220 MAXIMUM .... 2000 MINIMUM .... 0 DIVERGENCE **WITH** SATURATING TRANSFORM

*** RESULTS ORDERED ACCORDING TO DIJ(MIN) ***

	CHANNELS	DIJ(MIN)	D(AVE)	CS (0)	CO (10)	CH (10)	CF (10)	SO (10)	SH (10)	SF (10)	OH (0)	OF (0)	HF (0)
				WEIGHTED INTERCLASS DIVERGENCE (DIJ)									
1.	7 8 12	1872.	1969.		1872	...	1971	1998	...	1972			
2.	7 9 12	1866.	1951.		1866	...	1904	1999	...	1938			
3.	2 10 12	1854.	1927.		1857	...	1854	1979	...	1874			
4.	7 10 12	1847.	1945.		1905	...	1929	1990	...	1847			
5.	7 11 12	1823.	1931.		1823	...	1904	1993	...	1868			
6.	2 8 12	1822.	1948.		1822	...	1898	1997	...	1971			
7.	2 9 12	1814.	1935.		1814	...	1849	1999	...	1947			
8.	6 8 12	1793.	1929.		1793	...	1821	1999	...	1960			
9.	6 10 12	1772.	1896.		1772	...	1775	1994	...	1834			
10.	6 9 12	1765.	1920.		1765	...	1829	1999	...	1927			
11.	1 8 12	1758.	1915.		1758	...	1769	1997	...	1967			
12.	4 7 10	1751.	1891.		1861	1887	1953	1892	...	1751			
13.	4 7 11	1739.	1874.		1739	1899	1935	1891	...	1777			
14.	2 11 12	1733.	1905.		1733	...	1829	1982	...	1888			
15.	5 7 10	1729.	1887.		1855	1913	1964	1862	...	1729			
16.	4 6 10	1717.	1822.		1717	1762	1812	1891	...	1750			
17.	1 9 12	1710.	1901.		1755	...	1710	1999	...	1939			
18.	1 10 12	1701.	1889.		1795	...	1701	1974	...	1862			
19.	5 7 11	1690.	1862.		1690	1921	1954	1855	...	1753			
20.	4 8 12	1688.	1895.		1688	...	1724	...	...	1960			
21.	2 4 11	1684.	1815.		1684	1694	1837	1888	1993	1793			
22.	5 8 12	1682.	1910.		1682	...	1812	...	...	1964			
23.	6 11 12	1680.	1885.		1680	...	1771	1997	...	1862			
24.	3 8 12	1678.	1921.		1678	...	1884	1999	...	1964			
25.	2 4 10	1672.	1840.		1839	1672	1866	1888	1995	1779			
26.	1 4 10	1659.	1807.		1803	1659	1763	1894	1952	1772			
27.	3 4 10	1652.	1770.		1665	1652	1748	1861	1948	1748			
28.	1 11 12	1649.	1862.		1662	...	1649	1977	...	1881			
29.	3 10 12	1648.	1870.		1648	...	1730	1991	...	1849			
30.	3 9 12	1643.	1883.		1643	...	1717	1999	...	1938			
*** 209.	1 2 3	646.	1309.		646	1125	1765	1579	1876	862			

SEP-7

Figure 7. SEPARABILITY Results Listing for the Best Three Channels out of Twelve

```
CLASSES THAT MAY BE COMBINED-MAX DIV. = 1500
AB        693.
12       1268.
OW       1012.
SW        791.
END OF LIST
```


Figure 8. Listing of Classes That May Be Combined

```
LABORATORY FOR APPLICATIONS OF REMOTE SENSING       SEPT 22,1980
             PURDUE UNIVERSITY                        11 24 19 AM
                                                     LARSYS VERSION 3

RESULTS OF SEPARABILITY GROUPING

THRESHOLD =  1500.
```

GROUP	CLASSES	SYMBOL	NO. PTS.
1			
	1	A	405
	2	B	1253
2			
	3	C	2106
3			
	4	D	20
4			
	5	E	80
5			
	6	F	87
6			
	7	G	99

```
I0103  CPU TIME USED WAS   1.255 SECONDS.   (LARSMN)
```

Figure 8A. Example Separability Grouping Table
From a Different Separability Run

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 24 09 AM
LARSYS VERSION 3

SEPARABILITY STUDY

CLASS....FOREST

CHANNEL	1	2	3	4	5	6	7	8	9	10	11	12
SPECTRAL BAND	0.46 - 0.49	0.48 - 0.51	0.50 - 0.54	0.52 - 0.57	0.54 - 0.60	0.58 - 0.65	0.61 - 0.70	0.72 - 0.92	1.00 - 1.40	1.50 - 1.80	2.00 - 2.60	9.30 - 11.70
MEAN	108.01	79.03	81.41	108.37	98.41	79.72	71.75	102.11	97.08	108.42	92.39	83.11
STD. DEV.	5.54	3.89	4.76	9.65	8.67	6.67	5.42	12.31	9.96	16.34	10.84	8.61

CORRELATION MATRIX

SPECTRAL BAND	0.46 - 0.49	0.48 - 0.51	0.50 - 0.54	0.52 - 0.57	0.54 - 0.60	0.58 - 0.65	0.61 - 0.70	0.72 - 0.92	1.00 - 1.40	1.50 - 1.80	2.00 - 2.60	9.30 - 11.70
0.46-0.49	1.00											
0.48-0.51	0.64	1.00										
0.50-0.54	0.68	0.71	1.00									
0.52-0.57	0.63	0.69	0.89	1.00								
0.54-0.60	0.61	0.75	0.88	0.90	1.00							
0.58-0.65	0.46	0.67	0.60	0.64	0.71	1.00						
0.61-0.70	0.63	0.78	0.82	0.82	0.90	0.75	1.00					
0.72-0.92	0.51	0.62	0.74	0.77	0.79	0.55	0.72	1.00				
1.00-1.40	0.32	0.42	0.34	0.40	0.37	0.53	0.42	0.41	1.00			
1.50-1.80	0.54	0.61	0.62	0.63	0.65	0.57	0.67	0.64	0.53	1.00		
2.00-2.60	0.52	0.64	0.59	0.60	0.61	0.67	0.67	0.55	0.70	0.70	1.00	
9.30-11.70	0.46	0.58	0.52	0.55	0.59	0.61	0.62	0.56	0.45	0.65	0.61	1.00

SEP-9

Figure 9. Statistics Summary for One Class

Interactive Control:

SEPARABILITY gives the user the capability to create additional variations of the initial ranking of channel combinations and the Separability Results Listing output produced by the divergence calculations for each "combination" of channels. If an 'OPTIONS TYPE' control card is included in the input deck, the function will pause after completing the processing for each "combination" of channels and will type the message:

I0024 ENTER OPTIONS (USER)

It will then accept inputs from the typewriter that vary the calculations for ranking and will print a new version of the Separability Results Listing. The inputs, which correspond in effect to control cards, are:

- WEIGHTS
- PRINT, but only those parameters that affect the Separability Results Listing are operative. These are the 'BEST', 'SHOW' and 'DIV' parameters.
- OPTIONS, all of the parameters that may be used in the input deck are valid (except 'TYPE' of course), and five new ones are available. These will be described below.
- END.
- STOP, also described below.

All of the control keywords and parameters that may be used in the input deck have the same effect when entered from the typewriter.

The new parameters that may be used with the OPTIONS keyword are:

- 'TRANS' - Causes the function to use transformed results, which is the normal system default. This parameter is used to restore operation to the default value after the 'UNTRANS' parameter has been in effect.
- 'NOSORT' - Causes the function to order the results by D(AVE) rather than DIJ(MIN). Ordering by D(AVE) is the system default, and this card is used to restore the parameter after a 'SORT' parameter had caused results to be ordered by DIJ(MIN).
- 'RESET' - Causes all parameters to be restored to the input deck values or to their default values if they were not specified on a card.
- 'HELP' - Prints a listing of examples for SEPARABILITY control keywords and parameters that may be input from the typewriter.
- 'TABLE' - Prints a table of all class names and their assigned symbols.

The END keyword entered from the typewriter causes the function to rerank the channel combinations according to the input parameters and produce a new results output listing. The STOP keyword is entered in response to the "ENTER OPTIONS" message to indicate that the function should begin calculations for the next specification on the COMBINATIONS control card; or, if there are no more specifications, terminate execution of the Separability function. The new calculations and ranking will be according to the options specified in the input deck.

Figure 10 shows the typewriter terminal sheet from an interactive session that was originated by the sample input deck in Figure 2. Figure 10 shows all communications beginning at the point when the user initiated LARSFRIS and continuing to the end of both executions of SEPARABILITY. The first messages, up to the first I0024 message, document normal non-interactive processing. The I0024 message requests typewriter input of additional options. The user did not want additional calculations of the 2- channel combinations, and he entered the 'stop' keyword. The function then proceeded to calculate the 3- channel combinations according to the options specified in the input deck. The Separability Results Listing shown in Figure 7 was produced from these calculations.

```
run larsys
EXECUTION BEGINS...

 10111  SEPARABILITY FUNCTION REQUESTED     (SEPSUP)
 10032  REDUCED STATISTICS COMPUTED.     (REDSAV)
 10034  ALL CONTROL AND DATA CARDS HAVE BEEN READ     (SEPINT)
 10021  100 OF 220 CALCULATIONS COMPLETED.     (DIVRG2)
 10021  200 OF 220 CALCULATIONS COMPLETED.     (DIVRG2)
 10022  DIVERGENCE CALCULATIONS COMPLETED--READY TO ORDER AND PRINT RESULTS
 10011  SEPARABILITY FUNCTION COMPLETED     (SEPSUP)

 10111  SEPARABILITY FUNCTION REQUESTED     (SEPSUP)
 10032  REDUCED STATISTICS COMPUTED.     (REDSAV)
 10034  ALL CONTROL AND DATA CARDS HAVE BEEN READ     (SEPINT)
 10022  DIVERGENCE CALCULATIONS COMPLETED--READY TO ORDER AND PRINT RESULTS
 10024  ENTER OPTIONS     (USER)
stop
 10021  100 OF 220 CALCULATIONS COMPLETED.     (DIVRG2)
 10021  200 OF 220 CALCULATIONS COMPLETED.     (DIVRG2)
 10022  DIVERGENCE CALCULATIONS COMPLETED--READY TO ORDER AND PRINT RESULTS
 10024  ENTER OPTIONS     (USER)
options table,help
SYMBOLS  CLASS

  C      CORN
  S      SOYBEANS
  O      OAT7WHT
  H      HAY
  F      FOREST

         EXAMPLES OF INPUT

WEIGHT ABCD(100),CF(0)
PRINT SHOW(F1,F2,F3)
OPTIONS HELP
OPTIONS MAX(300),MIN(10)
OPTIONS SORT
OPTIONS EXCLUDE(2,9),UNTRANS,TABLE
OPTIONS RESET,NOSORT,TRANS
PRINT BEST(20),DIV(20)
END
STOP
options max(2500),min(1800),exclude(7,8),sort
 10027  SORT ON DIJ(MIN) ALREADY IN EFFECT     (USER)
end
 10024  ENTER OPTIONS     (USER)
weight co(11),cf(11)
end
 10024  ENTER OPTIONS     (USER)
stop
 10011  SEPARABILITY FUNCTION COMPLETED     (SEPSUP)
 10004  END OF INPUT DECK - RUN COMPLETED     (LARSMN)
T=68.72/77.51 10.53.06
```

Figure 10. Typewriter Terminal Sheet from an Interactive Session

At the end of processing for the 3- channel combinations, the I0024 message was printed again. An 'options' keyword with the parameters 'table' and 'help' was then entered. These cause the table of class symbols and the list of possible interactive input keywords to be printed on the typewriter.

Following this output, the specification 'options max(2500), min(1800),exclude(7,8),sort' was entered. The system responded with an information message (I0027) that the sort parameter was already in effect. The user then entered the 'end' keyword to terminate input and initiate processing with the new specification. The Separability Results Listing that resulted is shown in Figure 11.

When the above processing was complete, the function again typed the I0024 message requesting additional options. This time the user chose to vary the weights for two class pairs, by entering 'weight co(11),cf(11)'. This set the weights for class pairs co and cf at 11 rather than the previous (default) value of 10. Next an 'end' keyword was entered and processing was initiated again. The Separability Results Listing produced by this processing is shown in Figure 12. Note that with the change in weights the channel combination 2-8-12 moved from fifth best (in Figure 11) to first (in Figure 12).

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 25 12 AM
LARSYS VERSION 3

RETENTION LEVEL .. 210 MAXIMUM 2500
MINIMUM 1800
EXCLUDED CHANNEL COMBINATION 7 8

DIVERGENCE **WITH** SATURATING TRANSFORM

*** RESULTS ORDERED ACCORDING TO DIJ(MIN) ***

	CHANNELS	DIJ(MIN)	D(AVE)	CS (0)	CO (10)	CH (10)	CF (10)	SO (10)	SH (10)	SF (10)	OH (0)	OF (0)	HF (0)
				WEIGHTED INTERCLASS DIVERGENCE (DIJ)									
1.	7 9 12	1866.	1951.		1866	2000	1904	1999	2000	1938			
2.	2 10 12	1854.	1927.		1857	2000	1854	1979	2000	1874			
3.	7 10 12	1847.	1945.		1905	2000	1929	1990	2000	1847			
4.	7 11 12	1823.	1931.		1823	2000	1904	1993	2000	1868			
5.	2 8 12	1822.	1948.		1822	2000	1898	1997	2000	1971			
6.	2 9 12	1814.	1935.		1814	2000	1849	1999	2000	1947			
7.	6 8 12												
8.	6 10 12												
9.	6 9 12												
10.	1 8 12												
11.	4 7 10												
12.	4 7 11												
13.	2 11 12												
14.	5 7 10												
15.	4 6 10												
16.	1 9 12												
17.	1 10 12												
18.	5 7 11												
19.	4 8 12												
20.	2 4 11												
21.	5 8 12												
22.	6 11 12												
23.	3 8 12												
24.	2 4 10												
25.	1 4 10												
26.	3 4 10												
27.	1 11 12												
28.	3 10 12												
29.	3 9 12												
30.	1 4 11												
*** 199.	1 2 3												

SEP-11

Figure 11. SEPARABILITY Results Listing After Processing the New 'OPTIONS' Spedification

RETENTION LEVEL .. 210 MAXIMUM 2500
MINIMUM .... 1800 DIVERGENCE **WITH** SATURATING TRANSFORM
EXCLUDED CHANNEL COMBINATION 7 8
*** RESULTS ORDERED ACCORDING TO DIJ(MIN) ***

	CHANNELS	DIJ(MIN)	D(AVE)	WEIGHTED INTERCLASS DIVERGENCE (DIJ)									
				CS (0)	CO (11)	CH (10)	CF (11)	SO (10)	SH (10)	SF (10)	OH (0)	OF (0)	HF (0)
1.	2 8 12	1971.	2010.		2004	2000	2088	1997	2000	1971			
2.	6 8 12	1960.	1989.		1973	2000	2003	1999	2000	1960			
3.	2 9 12	1947.	1996.		1996	2000	2034	1999	2000	1947			
4.	7 9 12	1938.	2014.		2052	2000	2094	1999	2000	1938			
5.	1 8 12	1934.	1974.		1934	2000	1946	1997	2000	1967			
6.	6 9 12	1927.	1980.		1941	2000	2012	1999	2000	1927			
7.	2 11 12	1888.	1965.		1906	2000	2012	1982	2000	1888			
8.	1 9 12	1881.	1958.		1931	2000	1881	1999	2000	1939			
9.	2 10 12	1874.	1989.		2042	2000	2039	1979	2000	1874			
10.	7 11 12	1868.	1994.		2006	2000	2094	1993	2000	1868			
11.	1 10 12	1862.	1947.		1974	2000	1871	1974	2000	1862			
12.	4 8 12	1856.	1952.		1856	2000	1897	2000	2000	1960			
13.	5 8 12	1851.	1968.		1851	2000	1993	2000	2000	1964			
14.	6 11 12	1848.	1943.		1848	2000	1948	1997	2000	1862			
15.	7 10 12	1847.	2009.		2096	2000	2122	1990	2000	1847			
16.	3 8 12	1846.	1980.		1846	2000	2073	1999	2000	1964			
17.	6 10 12	1834.	1955.		1950	2000	1953	1994	2000	1834			
18.	1 11 12	1814.	1917.		1828	2000	1814	1977	2000	1881			
19.	3 10 12	1812.	1926.		1812	2000	1903	1991	2000	1849			
20.	3 9 12	1807.	1939.		1807	2000	1889	1999	2000	1938			
21.	4 7 11												
22.	4 8 10												
23.	5 7 11												
24.	4 7 10												
25.	4 6 10												
26.	4 9 10												
27.	5 7 10												
28.	5 9 12												
29.	3 11 12												
30.	4 6 11												
*** 198.	1 2 3												

SEP-12

Figure 12. SEPARABILITY Results Listing After Processing New 'WEIGHT Control Card

When the function finished this processing, and again typed the I0024 message, the user responded with the 'stop' keyword. This terminated processing for the 3-channel combinations and, since 3 was the last specification on the COMBINATIONS control card, it also terminated execution of the Separability function.

Separability Algorithm

The Separability function uses divergence as a measure of class separability. Divergence is defined for two density functions. In the case of normal (Gaussianly distributed) variables, the divergence is given by

$$D(i,j) = \frac{1}{2}tr[(K_i - K_j)(K_j^{-1} - K_i^{-1})]$$
$$+\frac{1}{2}tr[(K_i^{-1} + K_j^{-1})(M_i - M_j)(M_i - M_j)^T]$$

for classes i and j, where

K_i and K_j are the class covariance matrices

M_i and M_j are the class mean vectors

tr[ ] denotes the trace of a matrix (the sum of the elements on the main diagonal)

$(\ )^T$ denotes the transpose of a vector.

Since divergence is defined only for two classes, it does not in itself provide a measure of separability for $n > 2$ classes. However, it has been demonstrated experimentally that two useful strategies employing divergence for maximizing overall separability (and hence maximizing classification accuracy) are:

1. Maximize the average pairwise divergence (D(AVE) in the Separability function); or
2. Maximize the minimum pairwise divergence (DIJ(MIN) in the Separability function).

The Separability function uses the first of these strategies to rank feature combinations unless the SORT option is in effect; in the latter case, the second strategy is used.

However, because of an effect studied by Swain, Robertson, and Wacker[1], the Separability function normally computes a saturating transformation of divergence before applying either of the strategies noted above. The transformation is defined by

$$D_T(i,j) = 2[1 - \exp(-D(i,j)/8)]$$

where D(i,j) is the divergence as defined previously. For convenience in printing the results of the calculations, the transformed divergence has been implemented in the form

$$D_T(i,j) = 2000.0[1.0 - \exp(-D(i,j)/8.0)]$$

The effect of the transformation is to reduce the influence of very widely separated classes when computing the average pairwise divergence (D(AVE)), and has been found a much more reliable criterion than the "ordinary" divergence for selecting optional features for classification.

The inputs required for this separability algorithm are:

- The number of channels to be used.
- The number of classes.
- The covariance matrix K_i for each class i.
- The elements of the inverse covariance matrix K_i^{-1} for each class i computed in the initial part of SEPARABILITY. Since this matrix is symmetric, only the lower traingular portion is stored and used in the calculations.
- The statistical mean vectors (M_i) (one component for each channel) for each class i.

The outputs of the separability algorithm are:

- The untransformed divergences for all class pairs for each channel combination.
- The sum of the pairwise divergences for each channel combination.

---

[1] Swain P.H., Robertson T.V., Wacker A.G., "Comparison of the Divergence and B-Distance in Feature Selection", LARS Information Note 020871

SMOOTHRESULTS

The SMOOTHRESULTS function processes blocks (or cells) of classified data from a LARSFRIS Classification Results File and modifies each cell according to user-supplied parameters to produce as output a "smoothed" Classification Results File. In general, SMOOTHRESULTS will compute the percentage of each class present within a cell and assign the dominant class to every point of the cell. This processor was originally developed for the purpose of reassigning issolated classified points to match those around it, so that maps of the classified results would more nearly match the labeling conventions of the end user.

Input to the function consists of the Classification Results File and an input deck of control cards. The principal output is another Classification Results File, which may be placed on tape or disk. This results file is normally used as input to the PRINTRESULTS function to produce a variety of printed output for evaluation of the classification. It may also be used as input to the COMPARE-RESULTS, COPYRESULTS, LISTRESULTS, and PUNCHSTATISTICS function, but it must be located on tape to be used by the latter two functions.

The standard printed output includes a listing of the input control cards, a list of options selected by the user and a table of output class numbers, names and weights (if WEIGHTS were specified).

Inputs

The principal input to the SMOOTHRESULTS function is a LARSFRIS Classification Results File, which may reside on either tape or disk. The results file is obtained by previously executing one of the LARSFRIS processors which produces a Classification Results File as output (e.g. CLASSIFYPOINTS). The user must specify the location of this file on an INRESULTS card. Remember that this input Classification Results File will reside on disk only if a previously executed LARSFRIS function wrote the file to disk during this terminal session.

In addition to the LARSFRIS Classification Results File, the user must also provide an input deck which further defines the output classes and classified data to be used. In particular, there are control options to specify the locations of the input and output Classification Results Files, to define the dimensions of the cell to be examined, to group classification pools, to specify priority groups, to assign weights to groups, to specify the area to be processed and to define new mixture classes.

A sample input deck is shown in Figure 1. Let's examine the details of this particular job.

Location of input Classification Results File

The INRESULTS card is required to specify the location of the input Classification Results File, which may be on disk or tape. If it is on tape, then the FILE parameter must also be specified. In

```
*SMOOTHRESULTS
INRE TAPE(195),FILE(3)
CELL 2,2
PRIOR(6)
GROUP PINE(1/1,2,6,9/),MIX P/H(2/3,4/),SLH/CYP(3/7,8,11,12/)
GROUP LOWSTOCK(4/5,10/),NONSTOCK(5/13,14,15/)
GROUP WETLAND(6/16,17,18/)
WEIG 7,7,7,7,8,1
OUTR DISK
BLOCK  RUN(77010200),LINE(1710,1885,1),COL(2900,3040,1)
END
```

Figure 1: Control Card Deck

this example, the input file is on file 3 of tape 195.

Optional definition of cell dimensions

The CELLSIZE card may be included to define the dimensions of the block of points (cell) which is processed at one time by the SMOOTHRESULTS processor. The user must specify both the number of lines and the number of columns, separated by a comma, to be used as a cell. If this card is not included, a 2 line by 2 column cell (block of 4 points) will be assumed. In Figure 1, the cellsize of 2 lines by 2 columns was requested.

Optional grouping of classification pools

The user may wish to redefine the classes for the "smoothed" Classification Results File which is to be created, by grouping together similar classes from the input Classification Results File. This may be accomplished through the use of a GROUP card of the format:

GROUP   name1(1/p1,p2,.../), name2(2/p3,p4,.../)

where "name1" is the name of the first group and consists of points belonging to classification pool numbers (or classes) p1, p2 or ... from the input Classification Results File. The second group, called "name2" will consist of points assigned classification pool numbers p3 or p4 or ... on the input Classification Results File. Any classes from the input Classification Results File not listed on the GROUP card will each be assigned to separate groups, with the same name as the pool or class name. For this job, there will be six output classes, with the input classes grouped as follows:

Output Class Name	Output Class Number	Input Pool or Class numbers
PINE	1	1,2,6,9
MIX P/H	2	3,4
SLH/CYP	3	7,8,11,12
LOWSTOCK	4	5,10
NONSTOCK	5	13,14,15
WETLAND	6	16,17,18

Note that the group names may be up to eight characters long, including blanks within the name.

Optional designation of priority classes

Once the user has defined what classes will exist on the output classification (by deciding whether or not to group input classes) he may choose to designate some of the output classes (groups) as priority classes using the PRIORITY card. This would mean that as a cell is processed, any priority classes will not be reassigned. The user might want to use the priority class designation if certain cover types in the area of study are known to cover only small areas and their location is of importance, such as in pinpointing the location of small lakes. It could also be used to designate the null class which may be surrounding an irregularly shaped area, in order to preserve the boundaries. In Figure 1, group 6 is designated as a priority class because it will mark the location of some small lakes.

Optional assignment of weights

Using the WEIGHT card, the user can assign a relative weight to each output class, which is later applied to the cell tabulations of the number of points in each class. This card might be used to reflect the relative importance of certain classes to the end user in assigning one class name to a mixture sell. Again, the user must decide how many output classes there will be, by deciding whether to request any grouping, before he can assign weights.

The sequence of weights on the card corresponds to the group sequence, and the numbers may be either integer or real. In this example, the non-stocked class is given a slightly increased weight, possibly because an area which is non-stocked may very easily have scattered trees (i.e. points belonging to other classes) in it.

Optional Specifications of Area to be Processed

The user may provide a BLOCK card to define a subset of the classified area to be considered by the SMOOTHRESULTS processor. This card specifies a run number and the line and column co-ordinates for the area. The BLOCK card in Figure 1 specifies that all points from lines 1728 through 1881 and columns 2908 through 3027 of the classification on run 77010-200 be processed. If this card is not included, the entire Classification Results File requested on the INRESULTS card will be used.

Optional Definition of New Classes

If the user would like to assign a different class name to cells containing certain mixtures of the defined classification groups, the MIXCLASS card should be used. The format of this card is:

MIXCLASS name1(p1,p2-p3,...), name2(p4-p5,...),...

where "name1" is the name assigned to the first new mixture class. The points of a cell will be assigned to this class if p1% of the cell's points are in output class or group 1 and between p2% and p3% of the cell's points are in group 2, etc. After each new class name, there must be a percent or range of percents for every output class or group. Up to ten new mixture classes could be defined.

Outputs

The principal output of the SMOOTHRESULTS function is the Classification Results File, for which the OUTRESULTS card must be used to specify where this file will be stored. The three forms of this control card are:

OUTRESULTS TAPE(XXX), INITIALIZE

OUTRESULTS TAPE(XXX), FILE(nn)

OUTRESULTS DISK

The first two forms of the control card are used to write the file onto a tape. If a new results tape is being used, the INITIALIZE parameter must be specified so that proper header information is placed at the beginning of the new tape before any files are written. The current Classification Results File will then be written onto the first file of the tape. If the tape already contains Classification Results Files, then the FILE parameter should be used. If a results file already exists on the specified file of the tape, the user will be notified by a message. He then has the option of writing over the old file, specifying new tape and file numbers, or stopping execution. The DISK parameter, specified in Figure 1, indicates that the results are to be written on the disk. However, the disk is only a temporary storage area. Any of the following actions will cause it to be erased by the system:

* Execution of another classification function such as CLASSIFYPOINTS or SMOOTHRESULTS.
* Re-initiating LARSFRIS, i.e., issuing the 'i larsys' command.
* Logging off the system, i.e., issuing the 'quit' command.

If the user wishes to save the file, he may copy it to tape with the COPYRESULTS function.

The principal data on the results file is the point-by-point classification results of the SMOOTHRESULTS function, obtained from processing the input results file a cell at a time. A separate record is written for each line of the classification and contains the output class number to which each point was assigned. A likelihood code which is generally associated with each point in a Classification Results File is set to zero. These results may be used by the PRINTRESULTS function to produce maps of the classified area and/or tables of the class assignments.

For more information, refer to the description of the PRINTRESULTS function.

In addition to the classification results for each point, the file contains other data related to the classification run, in the format required by LARSFRIS. However, because grouping may have occurred, or points may have been reclassified based on how surrounding points were classified, the Statistics File will not reflect the statistical characteristics of this output Classification Results File. Similarly, other related information may also be invalid.

Standard Printer Output

The SMOOTHRESULTS function always prints a summary of the user's input deck. This summary includes a listing of the input control cards, and a reminder of the options selected by the user. An example of this output is shown in Figure 2. The other standard printed output is a table of class names and weights, as shown in Figure 3.

Algorithm to Determine Cell Classification

Within a cell, the number of points in each group (output class) is tallied and multiplied by the normalized weight for that output class. These numbers are then divided by the sum of all the numbers (times 100) to get the weighted percent each output class is of the total points in the cell. If these percents fall within the ranges for any new classes defined on a MIXCLASS card, then that new class number is assigned to all points in the cell to be output.

Otherwise, the group number of the group with the greatest weighted percent, that is not a priority group, is assigned to all points of the output cell. Finally, any points that were designated priority are overlayed onto the output cell.

```
                         LABORATORY FOR APPLICATIONS OF REMOTE SENSING          OCT  10,1980
                                      PURDUE UNIVERSITY                         10 23 39 AM
                                                                                LARSYS VERSION 3

    *SMOOTHRESULTS
    INRE TAPE(195),FILE(3)
    CELL 2,2
    PRIOR(6)
    GROUP PINE(1/1,2,6,9/),MIX P/H(2/3,4/),SLH/CYP(3/7,8,11,12/)
    GROUP LOWSTOCK(4/5,10/),NONSTOCK(5/13,14,15/)
    GROUP WETLAND(6/16,17,18/)
    WEIG  7,7,7,7,8,1
    OUTR DISK
    BLOCK  RUN(77010200),LINE(1710,1885,1),COL(2900,3040,1)
    END

YOU HAVE SELECTED THE FOLLOWING OPTIONS

    CONSIDER A SUBSET OF CLASSIFIED AREA FOR POLLING
    APPLY GROUPING
    WEIGHT THE CLASSES
    OUTPUT POLLED CLASSIFICATION RESULTS TO DISK
    USE A CELL SIZE OF  2 LINES BY  2 COLUMNS
    CONSIDER THE FOLLOWING AS PRIORITY CLASSES:
      6
```

Figure 2: Control Card Summary for SMOOTHRESULTS

```
                         LABORATORY FOR APPLICATIONS OF REMOTE SENSING          OCT  10,1980
                                      PURDUE UNIVERSITY                         10 23 39 AM
                                                                                LARSYS VERSION 3
```

	OUTPUT CLASS	WEIGHT		OUTPUT CLASS	WEIGHT
1	PINE	0.189	4	LOWSTOCK	0.189
2	MIX P/H	0.189	5	NONSTOCK	0.216
3	SLH/CYP	0.189	6	WETLAND	0.027

```
I0103  CPU TIME USED WAS    7.664 SECONDS.     (LARSMN)
```

Figure 3: Table of Class Names and Weights

STATISTICS

The Statistics function calculates statistics of subsets of data values from the Multispectral Image Storage Tape. The subsets are in the form of training fields; which, in general, are grouped to define the training classes. The training classes are specified by the user on Field Description Cards.

The function calculates the mean and standard deviation, a covariance matrix, and a correlation matrix of the data values for the channels specified by the user. These statistics are calculated for each of the training fields and training classes the user defines. A Statistics File is then generated and written on the user's temporary disk. It contains statistical descriptions of the specified training classes; specifically, the mean vectors, covariance matrix, and the number of data vectors for each of these classes.

This Statistics File is an important input to the Separability, Classifypoints, and Sampleclassify functions. It may be used by them directly from the temporary disk. The file is erased, however, if the user either logs off the system (issues the 'quit' command), or re-initiates LARSFRIS (issues the 'i larsys' command). If the user wishes to save the file longer than this, he must either use the 'statdeck save' command to have it saved on his permanent disk, or must include a control card in the input deck to direct the function to punch the file in cards.

The printer output options of STATISTICS generate a large number of numerical and graphical outputs for each training field and class. It is, in fact, easy to produce literally hundreds of sheets of printer output from a single execution of the function. The user should, therefore, carefully evaluate his actual analysis requirements before specifying optional outputs.

Input to the function consists of:

- Data from the Multispectral Image Storage Tape(s)
- Control cards to select the output options,
- Field Description Cards to define the training fields.

The Statistics function always produces two standard printed outputs:

- A listing of the control cards that were supplied by the user.
- A coincident spectral plot, which shows the normalized range of values (the mean plus-and-minus one standard deviation) of all channels for all classes, or a subset of these.

In addition to these listings, the user may specify a number of optional printed outputs:

- A Statistics Summary containing the mean and standard deviation vectors and correlation matrices of the channels for each field and/or class.
- Histograms of the data values for requested channels for each field and/or class.
- Spectral plots which show the normalized range of values for all channels for each field and/or class.
- Coincident spectral plots for particular groups of classes.

Principal Inputs

The required input to the Statistics function is data from the Multispectral Image Storage Tape and an input deck containing a CHANNELS card to identify specific data to be used to produce the statistics for each class. (It must also, of course, contain the Function Selector Card and END card that all input decks require.) The function uses the Field Description Cards to identify the appropriate input tape. The content and form of the input tapes are described in Appendix IV.

The Field Description Cards also serve to specify the training fields and training classes. Each card specifies a single training field by indicating the run number on the Multispectral Image Storage Tape and the line and column number coordinates of the data to be used. Each card may define a single training class, although several of them are usually grouped together to define the class. In either case, the card(s) for each class are preceded by a particular type of data card that contains the word "CLASS", followed by a blank, in the first six columns and the name of the class in the succeeding columns.

Figure 1 shows a typical input deck that requests two successive executions of the Statistics function. The first of these identifies a single training class named 'HAY' which is defined by three Field Description Cards, each of which represents a different training field. The second execution is a little more complicated. It identifies nine training classes, which are defined by a total of sixteen Field Description Cards (sixteen training fields). Note that all of the Field Description Cards in this sample input deck are in the 'fixed form'. Either this or the 'free form' of the card could have been used. Refer to the Control Card Dictionary in Appendix I for more information of the two forms of this card.

The training fields may request data from the same data runs (as in the example) or from different runs. The fields are not sorted by the program, so it is recommended that those for a given class that use the same data run be adjacent within the cards for the class. The maximum number of training classes

```
END
        71053900  E*2        523  528   1  147  197   1WOODS
        71053900  H-6        543  562   1   63   72   1WOODS
CLASS FOREST
        71053900  G*3        562  582   1  156  164   1WHEAT
CLASS WHEAT
        71053900  E*1        426  441   1  179  193   1OTHER   SET ASIDE
CLASS SA
        71053900  R-1        482  492   1   40   56   1HAY
CLASS HAY
        71053900  G*2        565  582   1  144  149   1OATS
        71053900  K-4        541  559   1   35   38   1OATS
        71053900  F-9        461  473   1   64   98   1OATS
CLASS OATSCUT
        71053900  H*2        591  593   1  137  170   1SOYBEANS
        71053900  D*6        592  612   1  101  131   1SOYBEANS
CLASS SOYBEAN2
        71053900  L-8        606  614   1   58   94   1SOYBEANS
        71053900  L-5        619  628   1   23   29   1SOYBEANS
CLASS SOYBEAN1
        71053900  H*1        598  612   1  138  169   1CORN
        71053900  D*5        618  641   1  100  132   1CORN
CLASS CORN2
        71053900  L-7        619  628   1   58   94   1CORN
        71053900  L-6        610  628   1   35   52   1CORN
CLASS CORN1
DATA
PUNCH
SCALE SPCLOW(20)
SCALE SPCINT(2)
PRINT HIST(C)
OPTIONS HIST(5,7,9)
CHANNELS 1,2,3,4,5,6,7,8,9,10,11,12
*STATISTICS
END
        71053900  R-1        482  492   1   40   56   1HAY
        71053900  E*1        426  441   1  179  193   1OTHER   SET ASIDE
        71053900  G*3        562  582   1  156  164   1WHEAT
CLASS HAY
DATA
SCALE SPCLOW(20)
SCALE SPCINT(2)
PRINT HIST(F,C),CORRE(F,C),SPECTRL(F,C)
OPTIONS HIST(5,7,9)
CHANNELS 1,2,3,4,5,6,7,8,9,10,11,12
*STATISTICS
```

Figure 1. Input Deck for STATISTICS

allowed is 60. If more classes are supplied, the first 60 will be used, and the remainder will be ignored. There is no explicit limit on the number of training fields in a training class.

Channels Specification

The channels for which statistics are to be computed must always be specified. This is accomplished through the CHANNELS control card in the input deck. The channels that are specified on the CHANNELS card will be used for all training classes. If the data run for a training field does not contain a requested channel, then statistics for that channel will not be computed, and the user will be informed via an information message. Both examples in Figure 1 show a CHANNELS card which specifies that channels 1 through 12 be used. A more complex form of the CHANNELS card that provides calibration experimentation is described in the Control Card Dictionary in Appendix I.

The Statistics function uses four optional control cards, each of which controls the output in different ways:

- The OPTIONS card allows the user to histogram only a subset of the channels specified on the CHANNELS card, rather than all of them.

- The PRINT card provides a variety of control over the printed outputs. These will be described in detail below.
- The PUNCH control card gives the user the ability to punch the Statistics File on cards and to control the format in which it is punched.
- The SCALE card enables the user to control the beginning point and the data intervals for the spectral plots.

Statistics File

The Statistics File is always written onto disk and may be punched on cards. It can be input in either form to the Separability, Classifypoints and Sampleclassify functions. The file contains the following:

- Training field information supplied by the user.
- A record containing the number of classes, number of fields and number of channels.
- Records containing the channel number, wavelength of the spectral band and calibration code for each channel used in this statistics calculation.
- A record containing the number of points in each class.
- Records containing the mean for each channel for each class.
- Records containing the covariance matrix of all specified channels for each class.

The Statistics File may be punched into cards by including a PUNCH control card in the input deck. Note that the control cards for the first example of STATISTICS in the sample deck (Figure 1) do not include this card and those for the second example do. The file that is punched in cards contains the same information as the disk file; however, the means and covariance matrices in the deck can be punched in either of two forms: hexadecimal (binary) representation or character representation. The default is the hexadecimal form, since this packs more information on each card and provides a smaller deck. The user gets this form if no parameter is placed on the PUNCH card.

The hexadecimal form requires four card columns per number (e.g., one mean or one value of the covariance matrix). The character form requires 14 card columns per number. The number in character format is of the form n.nnnnnnnnE+mm, which is the number n.nnnnnnnn times ten raised to the mm power. The character representation is selected by using the 'CHARACTEPS' control parameter on the PUNCH card.

The first and last cards of the Statistics File are easily identified.
The first card is:

LARSYS VERSION 3.1 STATISTICS DECK

It contains a 1 or a 0 (which indicates hexadecimal or character format) in column 40, and a 1 in column 80. These numbers are essential to the functions that read the Statistics File.

The last card is:

EOS ***** LAST CARD OF STATISTICS DECK ***** nnnnnnnn

The n's are a sequence number in columns 73-80. (The entire deck includes sequence numbers in these columns.)

Standard Printed Output

A Control Card Summary listing is the first printed output from the Statistics function. The listing that was produced by the first example in the input deck (Figure 1) is shown in Figure 2.

In addition, the Statistics function always produces a Coincident Spectral Plot for all, or a subset, of the classes for which statistics were calculated. This shows, for each channel, the "range" of data values for each class, with the classes plotted on the same listing. The "range" of values that is plotted represents the mean spectral value of the channel (for that class) plus and minus one standard deviation.

Figure 3 shows part of the plot that was produced by the second example in the input deck. The total Spectral Plot shows the range of each of the nine classes in each of twelve channels. For purposes of illustration Figure 3 has been truncated and shows only the first nine channels.

```
DEMO                    LABORATORY FOR APPLICATIONS OF REMOTE SENSING            MAR  7,1973
PHILLIPS                           PURDUE UNIVERSITY                             11 15 53 AM
                                                                                 LARSYS VERSION 3

     *STATISTICS
     CHANNELS 1,2,3,4,5,6,7,8,9,10,11,12
     OPTIONS HIST(5,7,9)
     PRINT HIST(F,C),CORRE(F,C),SPECTRL(F,C)
     SCALE SPCINT(2)
     SCALE SPCLOW(20)
     DATA

 YOU HAVE SELECTED THE FOLLOWING STATISTICS OPTIONS

     PRINT MEAN VECTOR AND CORRELATION MATRIX FOR EACH FIELD
     PRINT MEAN VECTOR AND CORRELATION MATRIX FOR EACH CLASS
     PRINT SPECTRAL PLOT FOR EACH FIELD
     PRINT SPECTRAL PLOT FOR EACH CLASS
     PRINT A HISTOGRAM FOR EACH FIELD
     PRINT A HISTOGRAM FOR EACH CLASS

 STATISTICS CONTROL INFORMATION....

     NUMBER OF CLASSES IS...  1
     CHANNELS SELECTED ARE.  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12,
     CALIBRATION CODES ARE.  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
     HISTOGRAM CHANNELS      5,  7,  9,

                                                                          STA-2
```

Figure 2. Control Card Summary for STATISTICS

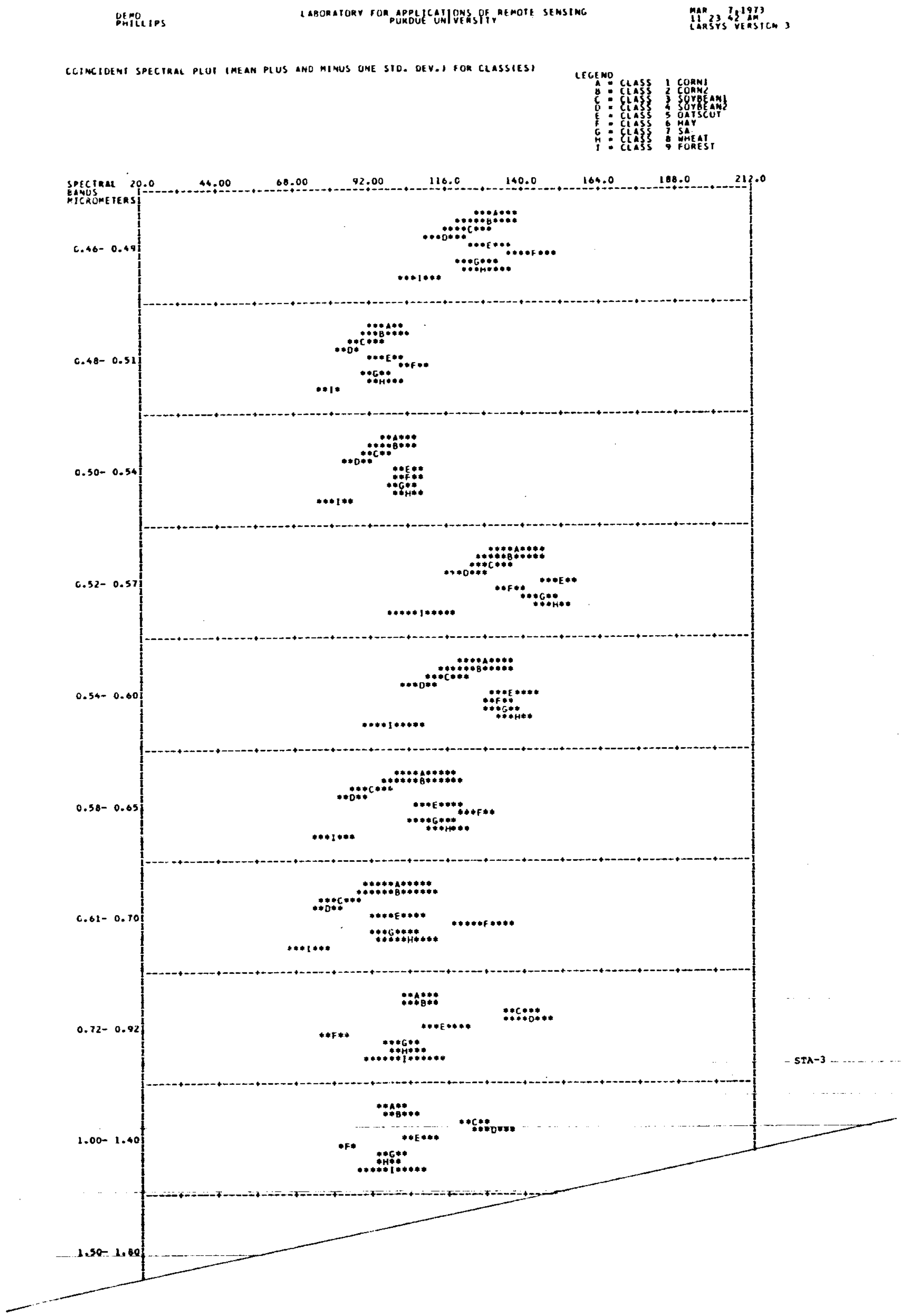


Figure 3. Coincident Spectral Plot for Twelve Channels and Nine Classes

In addition to these standard outputs, there is a Field Information Listing that is produced whenever the user does not select any of the optional printed outputs for training fields. (These optional outputs are discussed later in the narrative). A sample of a listing that was produced by the second example in the sample input deck is shown in Figure 4. The listing shows the identifying information and the number of data samples for each of the training fields for the 'OATSCUT' class. A page similar to this was printed for each of the other eight training classes in the sample deck.

Optional Printed Output

Three optional outputs may be specified with the PRINT control card.

Histograms

Histograms may be requested for selected channels for all fields and/or classes. They are requested by the 'HIST' control parameter on the PRINT card. 'HIST(F)' requests field histograms, 'HIST(C)' requests class histograms, and 'HIST(F,C)' requests both. Figures 5 and 6 are sample outputs produced by the 'HIST(F,C)' parameter on the PRINT card in the first example in the sample input deck. Figure 5 is the histogram for field 'R-1'. Histograms for both of the other two fields defined in that example were

DEMO PHILLIPS	LABORATORY FOR APPLICATIONS OF REMOTE SENSING PURDUE UNIVERSITY	MAR 7,1973 11 22 15 AM LARSYS VERSION 3

CLASS....OATSCUT

FIELD F-9 RUN NO. 71053900 OTHER INFORMATION	TYPE OATS NO. OF SAMPLES	 455	LINES COLUMNS	461- 64-	473 98	(BY (BY	1) 1)
FIELD K-4 RUN NO. 71053900 OTHER INFORMATION	TYPE OATS NO. OF SAMPLES	 76	LINES COLUMNS	541- 35-	559 38	(BY (BY	1) 1)
FIELD G*2 RUN NO. 71053900 OTHER INFORMATION	TYPE OATS NO. OF SAMPLES	 108	LINES COLUMNS	565- 144-	582 149	(BY (BY	1) 1)
	TOTAL	639					

STA-4

Figure 4. Field Information Listing

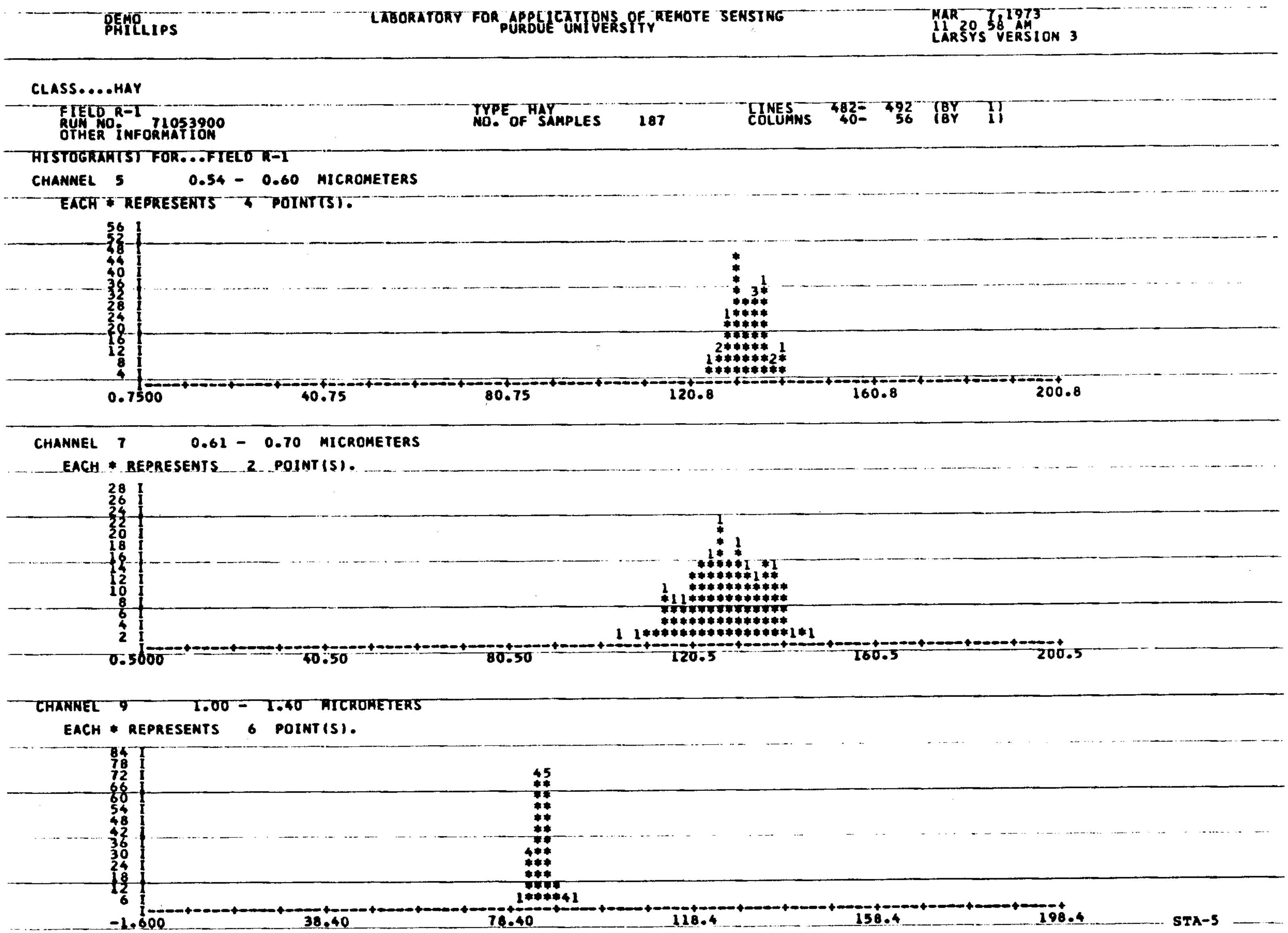


Figure 5. Sample Output of Histogram for a Training Field

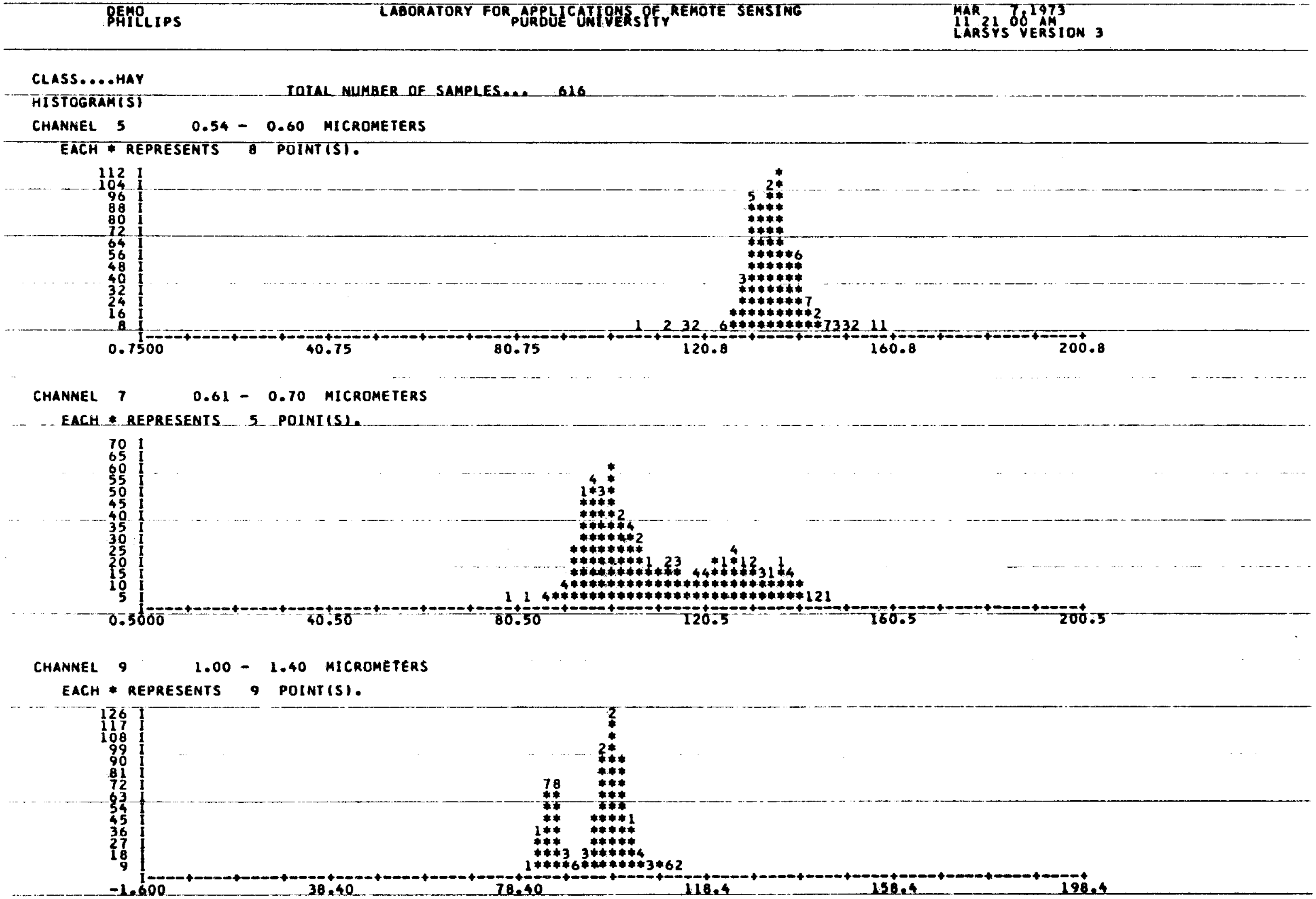


Figure 6. Sample Output of a Histogram for a Class

also produced in this run, but are not shown here. Figure 6 is the histogram for class 'HAY', the only class defined in the example.

Normally a histogram is produced for each channel that was specified on the CHANNELS card. However, it is possible to request only a subset of the channels be histogrammed. This is accomplished by including the OPTIONS control card with its 'HIST' parameter, which identifies the channels to be histogrammed. The channels selected in this way must always be a subset of those used for statistics calculation. In both of the examples shown, an 'OPTIONS HIST(5,7,9)' control card was included and histograms were calculated only for channels 5, 7 and 9.

Statistics Summary

Printing of the mean and standard deviation of each channel, and the correlation matrices for these channels, may be requested for fields and/or classes by the 'CORRE' parameter on the PRINT control card. 'CORRE(F)' requests statistics be printed for each field, 'CORRE(C)' requests classes, and 'CORRE(F,C)' requests both. Figures 7 and 8 show part of the statistics output from the 'CORRE(F,C)' parameter in the first example in the input deck. Only two of the outputs (field R-1 and class HAY that were produced are shown here.

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 20 50 AM
LARSYS VERSION 3

CLASS....HAY

FIELD R-1
RUN NO. 71053900
OTHER INFORMATION

TYPE HAY
NO. OF SAMPLES 187

LINES 482- 492 (BY 1)
COLUMNS 40- 56 (BY 1)

THE MEAN AND STANDARD DEVIATION VECTORS FOR FIELD R-1

CHANNEL	1	2	3	4	5	6	7	8	9	10	11	12
SPECTRAL BAND	0.46 - 0.49	0.48 - 0.51	0.50 - 0.54	0.52 - 0.57	0.54 - 0.60	0.58 - 0.65	0.61 - 0.70	0.72 - 0.92	1.00 - 1.40	1.50 - 1.80	2.00 - 2.60	9.30 - 11.70
MEAN	143.04	106.72	103.96	135.82	131.97	125.28	127.03	80.54	84.28	149.21	131.17	163.11
STD. DEV.	6.26	4.19	3.21	3.21	3.75	5.29	8.18	4.36	1.83	8.12	7.83	13.71

CORRELATION MATRIX

SPECTRAL BAND	0.46 - 0.49	0.48 - 0.51	0.50 - 0.54	0.52 - 0.57	0.54 - 0.60	0.58 - 0.65	0.61 - 0.70	0.72 - 0.92	1.00 - 1.40	1.50 - 1.80	2.00 - 2.60	9.30 - 11.70
0.46-0.49	1.00											
0.48-0.51	0.65	1.00										
0.50-0.54	0.57	0.65	1.00									
0.52-0.57	0.30	0.31	0.39	1.00								
0.54-0.60	0.62	0.59	0.65	0.51	1.00							
0.58-0.65	0.64	0.75	0.68	0.41	0.67	1.00						
0.61-0.70	0.72	0.81	0.68	0.35	0.72	0.80	1.00					
0.72-0.92	-0.57	-0.67	-0.49	-0.11	-0.44	-0.62	-0.79	1.00				
1.00-1.40	-0.30	-0.33	-0.22	0.07	-0.18	-0.39	-0.40	0.46	1.00			
1.50-1.80	0.67	0.76	0.69	0.40	0.65	0.77	0.83	-0.65	-0.32	1.00		
2.00-2.60	0.69	0.77	0.68	0.46	0.64	0.81	0.80	-0.64	-0.37	0.86	1.00	
9.30-11.70	0.67	0.80	0.65	0.34	0.60	0.83	0.86	-0.71	-0.51	0.88	0.89	1.00

STA-7

Figure 7. Sample Output of the Statistical Summary for a Training Field

DEMO
PHILLIPS

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1973
11 20 59 AM
LARSYS VERSION 3

CLASS....HAY   TOTAL NUMBER OF SAMPLES... 616

THE MEAN AND STANDARD DEVIATION VECTORS

CHANNEL	1	2	3	4	5	6	7	8	9	10	11	12
SPECTRAL BAND	0.46 - 0.49	0.48 - 0.51	0.50 - 0.54	0.52 - 0.57	0.54 - 0.60	0.58 - 0.65	0.61 - 0.70	0.72 - 0.92	1.00 - 1.40	1.50 - 1.80	2.00 - 2.60	9.30 - 11.70
MEAN	132.07	98.65	103.12	143.69	134.09	116.86	108.68	95.44	93.69	137.84	118.59	129.45
STD. DEV.	9.63	6.92	3.55	7.24	5.31	8.41	14.67	10.96	6.94	11.81	12.01	25.44

CORRELATION MATRIX

SPECTRAL BAND	0.46 - 0.49	0.48 - 0.51	0.50 - 0.54	0.52 - 0.57	0.54 - 0.60	0.58 - 0.65	0.61 - 0.70	0.72 - 0.92	1.00 - 1.40	1.50 - 1.80	2.00 - 2.60	9.30 - 11.70
0.46-0.49	1.00											
0.48-0.51	0.85	1.00										
0.50-0.54	0.49	0.52	1.00									
0.52-0.57	-0.41	-0.40	0.26	1.00								
0.54-0.60	0.14	0.19	0.65	0.65	1.00							
0.58-0.65	0.77	0.86	0.56	-0.21	0.34	1.00						
0.61-0.70	0.88	0.94	0.51	-0.43	0.18	0.88	1.00					
0.72-0.92	-0.80	-0.82	-0.26	0.65	0.13	-0.73	-0.89	1.00				
1.00-1.40	-0.72	-0.76	-0.23	0.58	0.11	-0.72	-0.83	0.90	1.00			
1.50-1.80	0.74	0.80	0.50	-0.28	0.19	0.76	0.81	-0.64	-0.60	1.00		
2.00-2.60	0.79	0.86	0.47	-0.34	0.17	0.83	0.86	-0.71	-0.66	0.86	1.00	
9.30-11.70	0.83	0.89	0.38	-0.54	0.01	0.83	0.94	-0.88	-0.86	0.84	0.86	1.00

STA-8

Figure 8. Sample Output of the Statistical Summary for a Class

Spectral Plots

Spectral Plots for each field, class or both may be requested with the 'SPECTRL' parameter on the PRINT card. 'SPECTRL(F)' requests plots for fields, 'SPECTRL(C)' for classes, and 'SPECTRL(F,C)' for both. Figures 9 and 10 illustrate some of the results of the 'SPECTRL(F,C)' parameter that appears in the first example in the sample input deck. The outputs are included only for field 'R-1' and class HAY.

In addition, Coincident Spectral Plots for selected classes may be printed on the same plot. The standard output, which was described earlier, produces a single plot which includes all classes for which statistics were calculated. A PRINT control card parameter enables the user to request one or more Coincident Spectral Plots, each of which includes a specified subset of classes. These are requested with the 'COSPEC' parameter on the PRINT card. For example, 'PRINT COSPEC(2,4,6),COSPEC(3,5,6)' requests two plots - one plotting classes 2,4,and 6 together and the other plotting classes 3,5 and 6 together. The resulting Spectral Plots are similar to the Coincident Spectral Plot for all classes that is shown in Figure 3. Use of the 'COSPEC' parameter will cause the plot for all classes to be suppressed unless specifically requested by the use of another 'COSPEC' parameter.

```
     DEMO                                  LABORATORY FOR APPLICATIONS OF REMOTE SENSING                    MAR  7,1973
     PHILLIPS                                            PURDUE UNIVERSITY                                  11 20 57 AM
                                                                                                            LARSYS VERSION 3

CLASS....HAY
   FIELD R-1                                  TYPE  HAY                           LINES    482-  492  (BY   1)
   RUN NO.   71053900                         NO. OF SAMPLES     187              COLUMNS   40-   56  (BY   1)
   OTHER INFORMATION

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.) FOR FIELD R-1

SPECTRAL   20.0       44.00       68.00       92.00      116.0       140.0      164.0       188.0       212.0
BANDS      I-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----I
MICROMETERSI                                                                                                I
0.46- 0.49I                                                            ++++++++                            I
          I                                                                                                I
0.48- 0.51I                                         +++++                                                  I
          I                                                                                                I
0.50- 0.54I                                        +++++                                                   I
          I                                                                                                I
0.52- 0.57I                                                          +++++                                 I
          I                                                                                                I
0.54- 0.60I                                                        +++++                                   I
          I                                                                                                I
0.58- 0.65I                                                    ++++++                                      I
          I                                                                                                I
0.61- 0.70I                                                   ++++++++++                                   I
          I                                                                                                I
0.72- 0.92I                              +++++                                                             I
          I                                                                                                I
1.00- 1.40I                                 +++                                                            I
          I                                                                                                I
1.50- 1.80I                                                              +++++++++                         I
          I                                                                                                I
2.00- 2.60I                                                      +++++++++                                 I
          I                                                                                                I
9.30-11.70I                                                                  ++++++++++++++                I
          I                                                                                                I
          I-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----I
          20.0       44.00       68.00       92.00      116.0       140.0      164.0       188.0       212.0
```

STA-9

Figure 9. Sample Output of a Spectral Plot of a Training Field

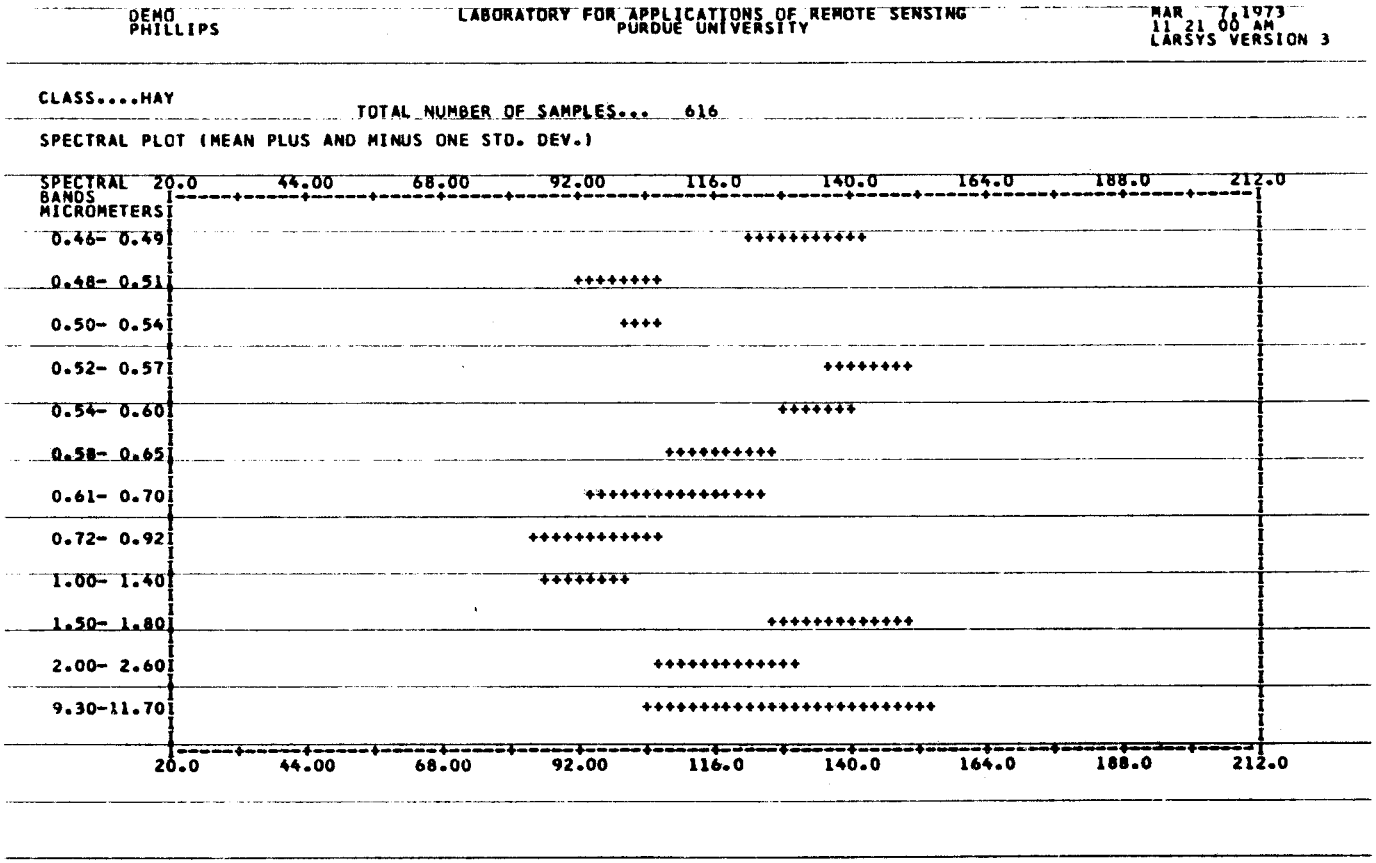


Figure 10. Sample Output of a Spectral Plot of a Class

The user has additional control over the printing of spectral plots. He may use the SCALE control card to specify both the low end of the plot (origin of the magnitude axis), and the interval of data values that will be used in the plot (the interval on the magnitude axis). The values that are specified on this control card apply both to the individual field and class plots that are requested with the 'SPECTRL' parameter on the PRINT card, and to the coincident plots of classes that are produced as standard output and by the 'COSPEC' parameter of the PRINT card.

The values for these constants that are used if the SCALE card is not included are zero for the low end of the plot and three for the interval. In both examples in the input deck, two SCALE control cards were used to specify a beginning value of 20 and an interval of 2. The plots shown in the previous outputs (Figures 3, 9 and 10) all reflect this choice.

TRANSFERDATA

The Transferdata function prints, punches, and/or records on tape data values from the Multispectral Image Storage Tapes. The function is specifically designed to copy data values from these tapes for use in programs other than those in the LARSFRIS system. It converts them to an easily read Fortran format in order to simplify their input to these programs. The data is not acceptable in this new format as input to any of the present LARSFRIS functions.

The data values are packed quite densely on the Multispectral Image Storage Tapes. When they are converted to the Fortran format, however, they are unpacked and their storage requirements increase considerably. As a result, the user should carefully analyze his actual data volume requirements, and should use this function only for very small samples of data.

Since all output from the function is a copy of data stored on a Multispectral Image Storage Tape, the user will need to understand the nature of this data and the manner in which it is stored on the tape. This information, as well as information on how the data is calibrated for use in LARSFRIS, is described in Appendix IV.

Input

The inputs to TRANSFERDATA are a Multispectral Image Storage Tape and an input deck. The input deck must contain at least:

- A Function Selector card to select the Transferdata function.
- A CHANNELS card to specify the channels to be copied.
- A DATA card, followed by one or more Field Description Cards to specify the data by giving the run number and line and column coordinates. Each Field Description Card identifies a single "module" of data which will be assigned an identification number called a "module number".
- An END card.

In addition, the deck can include four optional control cards - the TAPE, PUNCH and PRINT cards, which specify the type of output, and the OPTIONS card, which permits the user to assign his own module numbers to the modules. Figure 1 illustrates an example input deck for the Transferdata function. The purpose of each of these cards is described below.

The CHANNELS card in the input deck specifies that data from channels 2,5,7,9 and 11 on the input tape be copied. These channels will be copied for all fields that are specified in

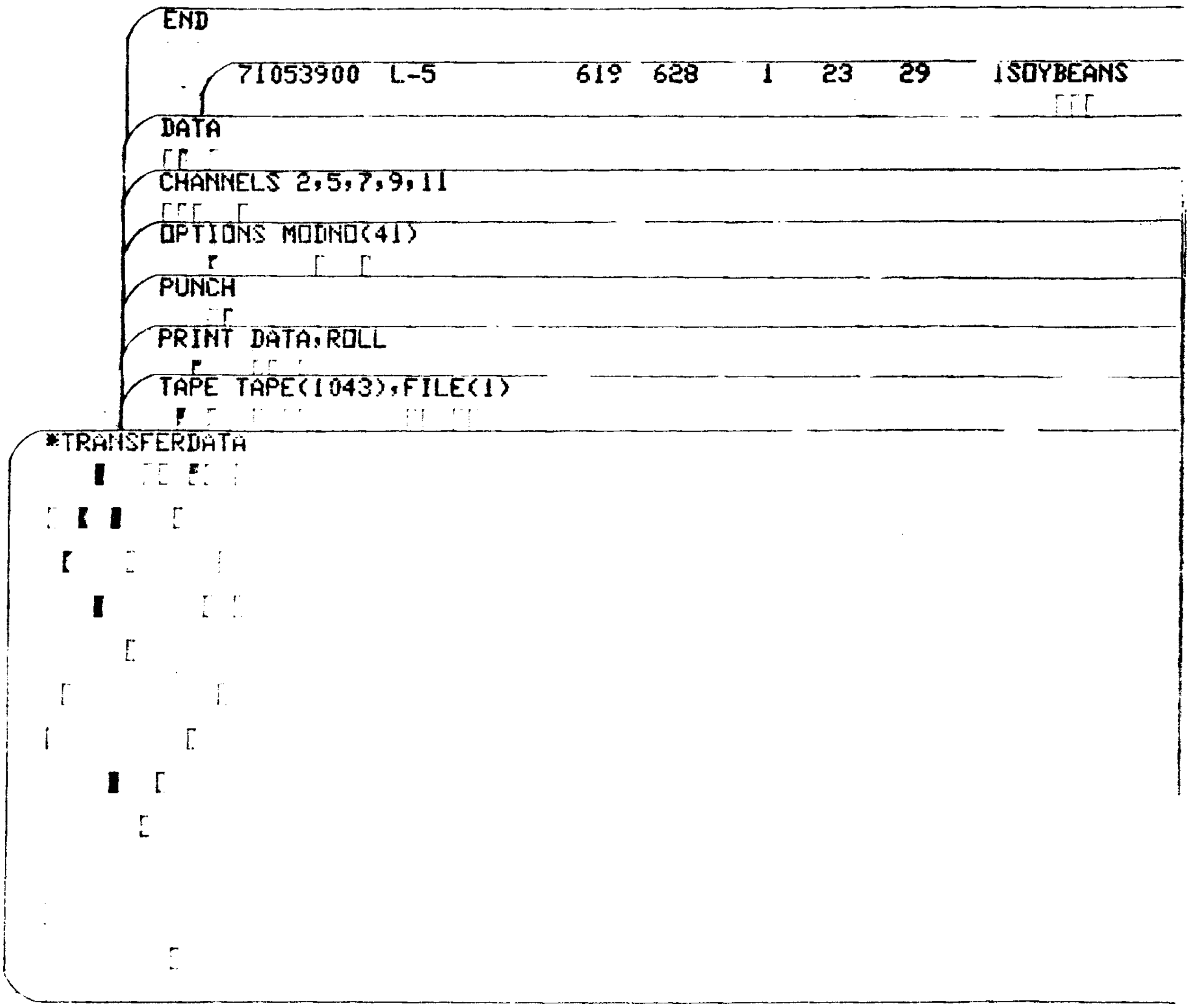


Figure 1. Input Deck for TRANSFERDATA

this execution of the function. The user could have also used the alternate form of the CHANNELS card which allows him to control the calibration of the data. Consult the Control Card Dictionary in Appendix II for a description of how this card is used.

The PRINT card in Figure 1 illustrates the use of both of the two options that it provides. The forms of this card are:

PRINT DATA - Data values are printed.

PRINT ROLL - Only the roll parameters for each line of data are printed.

PRINT DATA, ROLL - Both data values and the roll parameters are printed.

The PUNCH control card will cause the punched card version of the data to be produced. The exact format of these records is described later in this subsection.

The TAPE control card requests that the data be copied to tape. The same Fortran format that is used for punching cards is also used for recording the data values on tape. A tape number must always be specified and the file number must be specified unless file 1 is desired. File 1 is assumed if the File parameter is not used.

The OPTIONS control card in Figure 1 specifies that the output module is to be assigned a module number of 41. If more than one module had been requested, the subsequent modules would be assigned numbers in ascending sequence (42,43,etc.,) until all requested modules were assigned a number. If the OPTIONS card had been omitted, the module numbers would start with 1.

The DATA card identifies the beginning of the Field Description Cards which specify the fields of data values to be copied. The END card must be the last card in the deck.

Outputs

Three types of data output, cards, tape or printer, can be generated by the Transferdata function. The card deck contains the data values and certain ancillary information. One module is punched for each Field Description Card in the input deck. If a tape is generated, it will have the same format as the card deck. These formats are described at the end of this subsection.

Two types of printer output, a control card summary and a data listing, may be generated. The data listing is a line printer listing of the information punched on cards or written on tape. It is intended to provide the user with a record of the information he obtained on tape or cards. Figures 2 and 3 show the line printer output from the example input deck in Figure 1.

```
DEMO              LABORATORY FOR APPLICATIONS OF REMOTE SENSING       MAR  7,1973
PHILLIPS                     PURDUE UNIVERSITY                        11 15 26 AM
                                                                      LARSYS VERSION 3

*TRANSFERDATA
TAPE TAPE(1043),FILE(1)
PRINT DATA,ROLL
PUNCH
OPTIONS MODNO(41)
CHANNELS 2,5,7,9,11
DATA

                                                                      TRA-2
```

Figure 2. Control Card Summary for TRANSFERDATA

```
 DEMO                       LABORATORY FOR APPLICATIONS OF REMOTE SENSING                        MAR  7,1973
 PHILLIPS                                  PURDUE UNIVERSITY                                      11 15 53 AM
                                                                                                  LARSYS VERSION 3

RUN NUMBER... 71053900    LINES... (  619,  628,    1)    COLUMNS... (   23,   29,     1)
FIELD........ L-5         FIELD TYPE.......... SOYBEANS    OTHER INFORMATION
MODULE NUMBER.....  41    NUMBER OF CHANNELS.......   5    SAMPLE VECTORS IN MODULE        70

CHANNEL  2     SPECTRAL BAND  0.48 TO  0.51 MICROMETERS    CALIBRATION CODE = 1    CO =  23.10
CHANNEL  5     SPECTRAL BAND  0.54 TO  0.60 MICROMETERS    CALIBRATION CODE = 1    CO =  26.25
CHANNEL  7     SPECTRAL BAND  0.61 TO  0.70 MICROMETERS    CALIBRATION CODE = 1    CO =  28.00
CHANNEL  9     SPECTRAL BAND  1.00 TO  1.40 MICROMETERS    CALIBRATION CODE = 1    CO =  22.90
CHANNEL 11     SPECTRAL BAND  2.00 TO  2.60 MICROMETERS    CALIBRATION CODE = 1    CO =  25.95
                                          CHANNELS

                 2       5       7       9      11.
LINE  COL                                     DATA

 619        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 619   23    95.1   118.3    85.0   128.9   118.9
 619   24    99.1   116.3    88.0   127.9   118.9
 619   25    94.1   119.3    86.0   122.9   119.9
 619   26    96.1   118.3    87.0   124.9   117.9
 619   27    97.1   120.3    85.0   123.9   119.9
 619   28    98.1   129.3    95.0   124.9   114.9
 619   29    95.1   122.3    89.0   123.9   121.9
 620        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 620   23    96.1   114.3    86.0   127.9   117.9
 620   24    99.1   121.3    89.0   123.9   116.9
 620   25    97.1   119.3    84.0   121.9   119.9
 620   26    91.1   115.3    87.0   123.9   118.9
 620   27    94.1   117.3    87.0   122.9   115.9
 620   28    93.1   115.3    86.0   122.9   118.9
 620   29    96.1   123.3    87.0   122.9   114.9
 621        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 621   23    96.1   119.3    84.0   126.9   125.9
 621   24    95.1   122.3    89.0   126.9   123.9
 621   25    94.1   114.3    84.0   122.9   123.9
 621   26    95.1   117.3    86.0   120.9   117.9
 621   27    94.1   116.3    85.0   121.9   117.9
 621   28    96.1   112.3    83.0   122.9   118.9
 621   29    93.1   119.3    85.0   120.9   117.9
 622        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 622   23    94.1   121.3    86.0   127.9   119.9
 622   24    90.1   118.3    85.0   129.9   117.9
 622   25    91.1   117.3    84.0   126.9   117.9
 622   26    94.1   116.3    84.0   125.9   117.9
 622   27    96.1   119.3    88.0   124.9   114.9
 622   28    93.1   118.3    84.0   125.9   117.9
 622   29    93.1   122.3    83.0   127.9   118.9
 623        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 623   23    93.1   122.3    83.0   125.9   111.9
 623   24    98.1   121.3    87.0   126.9   112.9
 623   25    95.1   122.3    88.0   126.9   115.9
 623   26    97.1   121.3    86.0   127.9   114.9
 623   27    96.1   127.3    88.0   127.9   116.9
 623   28    97.1   122.3    85.0   127.9   114.9
 623   29    93.1   121.3    87.0   126.9   114.9
 624        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 624   23    93.1   116.3    83.0   122.9   114.9
 624   24    96.1   123.3    90.0   121.9   116.9
 624   25    96.1   120.3    88.0   122.9   119.9
 624   26    98.1   121.3    83.0   123.9   119.9
 624   27    96.1   121.3    87.0   127.9   119.9
 624   28    97.1   119.3    83.0   126.9   121.9
 624   29    90.1   121.3    86.0   126.9   119.9
 625        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 625   23    92.1   115.3    89.0   122.9   119.9
 625   24   102.1   128.3    96.0   120.9   118.9
 625   25    99.1   117.3    85.0   119.9   118.9
 625   26    94.1   120.3    85.0   120.9   121.9
 625   27    97.1   119.3    84.0   123.9   118.9
 625   28    96.1   119.3    86.0   127.9   116.9
 625   29   100.1   122.3    87.0   125.9   117.9
 626        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 626   23   100.1   119.3    93.0   121.9   119.9
 626   24   103.1   133.3   106.0   120.9   125.9
 626   25    93.1   119.3    89.0   118.9   128.9
 626   26    94.1   115.3    83.0   119.9   121.9
 626   27    94.1   112.3    83.0   121.9   118.9
 626   28    96.1   117.3    82.0   124.9   117.9
 626   29    95.1   118.3    84.0   126.9   121.9
 627        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 627   23   103.1   135.3   104.0   118.9   129.9
 627   24   107.1   138.3   111.0   118.9   134.9
 627   25   107.1   129.3   108.0   118.9   136.9
 627   26    98.1   116.3    92.0   120.9   126.9
 627   27    94.1   115.3    84.0   120.9   118.9
 627   28    93.1   119.3    86.0   123.9   118.9
 627   29    95.1   116.3    86.0   125.9   119.9
 628        ROLL PARAMETER.... 32767 MEANS THE ROLL WAS NOT CALCULATED.
 628   23   102.1   134.3   101.0   119.9   124.9
 628   24   126.1   156.3   136.0   120.9   139.9
 628   25   109.1   137.3   110.0   120.9   135.9
 628   26   106.1   130.3   105.0   120.9   126.9
 628   27    94.1   113.3    81.0   120.9   122.9
 628   28    97.1   120.3    85.0   125.9   121.9
 628   29    97.1   120.3    87.0   126.9   122.9
```

TRA-3

Figure 3. TRANSFERDATA Printed Output

Figure 2 is a listing of the control card input and Figure 3 is the data that was requested. Note that the roll parameter was requested on the PRINT control card, but was not calculated for data in run 71053900.

Tape and Card Record Formats

The record format for tape and card output is identical. The records are formatted in "card image", i.e., in 80 byte records. The three types of records for each module of data are described below. The FORTRAN format used by the Transferdata function is given under the heading 'Format'. The user may wish to refer to the listing in Figure 4 of the output deck that was generated by the example input deck.

Field Identification Record

One of these records is produced as the first record for each module of data that is copied. The record identifies and describes the field.

Card Columns	Format	Description
1-8	I8	Run number from the Multispectral Image Storage Tape
11-18	Alphabetic	Name of the data module.
21-25	I5	Line number for the first line of of the module.

```
71053900  L-5          619  628    1   23    29    1SOYBEANS    5       70       41   20
  2  1  0.48   0.51     23.10E 00    14.42E 01    16.33E 01                      41    1
  5  1  0.54   0.60     26.25E 00    17.97E 01    20.90E 01                      41    2
  7  1  0.61   0.70     26.00E 00    16.81E 01    20.55E 01                      41    3
  9  1  1.00   1.40     22.90E 00    31.30E 00    56.70E 00                      41    4
 11  1  2.00   2.60     25.95E 00    94.05E 00    14.06E 01                      41    5
 95118 85129119 99116 88128119 94119 86123120 96118 87125118 97120 85124     41    6
120 98129 95125115 95122 89124122 96114 86128118 97121 89124117 97119 84     41    7
122120 91115 87124119 94117 87123116 93115 86123119 96123 87123115 96119     41    8
 84127126 95122 89127124 94114 84123124 95117 86121118 94116 85122118 96     41    9
112 83123119 93119 85121118 94121 86128120 90118 85130118 91117 84127118     41   10
 94116 84126118 96119 84125115 93118 84126118 93122 85128119 93122 83126     41   11
112 96121 87127113 95122 88127116 97121 86126115 96127 88128117 97122 85     41   12
128115 93121 87127115 93116 83123115 96123 90122117 96120 86123120 98121     41   13
 83124120 96121 87128120 97119 83127122 90121 86127122 92115 89123120102     41   14
128 96121119 99117 85120119 94120 85121122 97119 84124119 96119 86128117     41   15
100122 87126118100119 93122120103133106121126 93119 83119129 94115 83120     41   16
122 94112 83122119 96117 82125118 95118 84127122103135104119130107136111     41   17
119135107129108119137 96116 92121127 94115 84121119 93119 86124119 95116     41   18
 86126120102134101120125126156136121140109137110121130106130105121127 94     41   19
113 81121123 97120 85126122 97120 87127123                                   41   20
```

Figure 4. TRANSFERDATA Punched Output

Card Columns	Format	Description
26-30	I5	Line number for the last line of the module.
31-35	I5	Line interval for the module.
36-40	I5	Sample number for the first sample on each line.
41-45	I5	Sample number for the last sample on each line.
46-50	I5	Sample interval for the module.
51-58	Alphabetic	Class name of the class to which the module has been assigned.
59-60	--	Blank

Calibration Records

There is one calibration record for each channel that was requested. They immediately follow the Field Identification Record. The information in columns 7 through 54 is taken from the ID record on the Multispectral Image Storage Tape.

Card Columns	Format	Description
1-3	I3	Channel number
4-6	I3	Calibration code used to calibrate data
7-12	F6.2	Lower spectral band limit for this channel

Card Columns	Format	Description
13-18	F6.2	Upper spectral band limit for this channel
19-30	2PE12.2	Calibration value for C0
31-42	2PE12.2	Calibration value for C1
43-54	2PE12.2	Calibration value for C2
55-72	———	Blank
73-76	I4	Module number
77-80	I4	Card sequence number

Data Records

These records contain the calibrated data values in integer form. They follow the channel records. All the values requested for a field are punched consecutively, 24 to a record, in order by line, column, and channel in the following manner:

$$L_1C_1K_1, L_1C_1K_2, \ldots, L_1C_1K_n, L_1C_2K_1,$$

$$\ldots, L_{.}C_mK_n, L_2C_1K_1, 111, L_kC_mK_n$$

where L = line, C = column, K = channel, n = number of channels requested, m = samples per line and k = number of requested lines. Note that k·m = samples per channel in the module (columns 64-69 on first card).

Card Columns	Format	Description
1-72	I3	24 calibrated data values.
73-76	I4	Module number
77-80	I4	Card sequence number in module after Field Identification Record.