1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-13821		

4. Title and Subtitle	5. Report Date
Earth Observations Division Version of the Laboratory for Applications of Remote Sensing System (EOD-LARSYS) User Guide for the IBM 370/148 - Volume II, Detailed User Documentation	December 1978
	6. Performing Organization Code: SF3
7. Author(s): J. Stewart and P. J. Aucoin, Lockheed Electronics Comapny, Inc.	8. Performing Organization Report No.: LEC-12564
	10. Work Unit No.
9. Performing Organization Name and Address: Lockheed Electronics Company, Inc., Systems, and Services Division, Houston, Texas 77058	11. Contract or Grant No.: NASA 9-15200
	13. Type of Report and Period Covered: User Guide
12. Sponsoring Agency Name and Address: National Aeronautics and Sapce Administration, Lyndon B. Johnson Space Center, Houston, Texas 77058 (J. M. Sulester, Tech. Monitor)	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This document presents instructions for analysts who use EOD-LARSYS as programmed on the Purdue University IBM-370/148 computer. It presents sample applications, control cards, and error messages for all processors in the system and gives detailed descriptions of the mathematical procedures and the information needed to execute the system and obtain the desired output. EOD-LARSYS is the JSC version of an integrated batch system for analysis of multispectral scanner imagery data. This volume is designed for use with the As-Built Documentation (Volume III) and the EOD-LARSYS program listings (Volume IV). The system is operational from remote terminals at Johnson Space Center, under the Virtual Machine/Conversational Monitor System environment.

17. Key Words (Suggested by Author(s))		18. Distribution Statement
Classification	Mapping	
Clustering	Scatter plots	
Display	Transformations	
Feature selection		
Histograms		
Landsat imagery		

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

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

EARTH OBSERVATIONS DIVISION VERSION OF THE LABORATORY FOR APPLICATIONS OF REMOTE SENSING SYSTEM (EOD-LARSYS) USER GUIDE FOR THE IBM 370/148

VOLUME II - USER'S REFERENCE MANUAL

Job Order 71-475

PREPARED BY

P. J. Aucoin and J. Stewart
Earth Observations Data Products Department

APPROVED BY

for James A Wilkinson
P. L. Krumm, Acting Supervisor
Scientific Applications Section

Prepared By

Lockheed Electronics Company, Inc.

For

Earth Observations Division

Space and Life Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

December 1978 LEC-12564

CONTENTS

Section | Page

Section | Page

Section | Page

Section Page

Section | Page

Section | Page

TABLES

FIGURES

1. INTRODUCTION

For several years, the Earth Observations Division (EOD) of the National Aeronautics and Space Administration (NASA), Lyndon B. Johnson Space Center (JSC), has supported research for the development of techniques to be used in processing remotely sensed imagery data obtained from the multispectral scanner (MSS) placed aboard various aircraft and satellites. One of the earliest operational computer systems to use pattern recognition techniques in the analysis of these data was developed at Purdue University's Laboratory for Applications of Remote Sensing (LARS). The earliest version of the LARS system (LARSYS) was converted in 1970 to a batch program for execution on the Univac 1108 EXEC 2 system at NASA/JSC (ref. 1).

The computer system described in this document originated from this early version of the LARSYS. However, since 1970, personnel of the EOD, Lockheed Electronics Company, Inc. (LEC), and other EOD support contractors have made many modifications and improvements to the Univac 1108 version of the LARSYS; thus, new techniques have been developed and programmed to perform additional functions in the evaluation of the data.

Although the basic structure of the system remains the same, a large portion of it has been reprogrammed. Modifications to existing techniques and the addition of new techniques have expanded the capabilities of the system. The current version is called the EOD-LARSYS. Recently, it has been converted from Univac 1108 Fortran to IBM 370/148 Fortran. The EOD LARSYS system is now operational on the IBM 370/148 computer located at LARS Purdue University, West Lafayette, Indiana.
It may be accessed both on and off the JSC site by remote terminal.

The purposes of this document are to define the capabilities and limitations of the system and to provide the user with the information needed to execute the program and obtain the output desired. It is assumed throughout the document that the user is familiar with the terminology and the pattern recognition techniques involved. No attempt is made to assist the user in the analysis of data output by the system.

This EOD-LARSYS User Guide for the IBM 370/148 is being issued in four volumes:

Volume I	--System Overview, August 1978 (JSC 13821; LEC 12563)
Volume II	--User's Reference Manual, December 1978 (JSC 13821; LEC 12564) (this document)
Volume III	--"As-Built" Design Specification (to be issued)
Volume IV	--Program Listings (to be issued)

2. GENERAL SYSTEM DESCRIPTION

The EOD-LARSYS is a batch processing program operational on the IBM 370/148 system at Purdue University. The system is composed of a system monitor and a set of processors, each of which performs a specific function in the analysis of MSS imagery data. Linkage between processors is accomplished by the use of files in the computer; by files on disk or tape; or, less commonly, by card decks. The execution of a particular batch job may begin or end with any processor, provided the appropriate files are furnished.

The data are preprocessed by the Earth Resources Interactive Processing System (ERIPS). (ERIPS merges tapes received from Goddard Space Flight Center and creates a multi-temporal/multi-pass tape.)

Two pattern recognition classification schemes are provided by the system. One, the supervised classification algorithm known as the maximum likelihood classifier, is embodied in the CLASSIFY processor (ref. 2). The other, an unsupervised classification or clustering algorithm, is embodied in the Iterative Self-Organizing Clustering System (ISOCLS) processor (ref. 3). ISOCLS, along with other processors, may be used to "train" the maximum likelihood classifier or to display the results of classification.

Having obtained an MSS image data tape (DATAPE) in one of the allowable formats (see section 3.2), the data analyst must train the classifier. The maximum likelihood classification algorithm is based on the assumption that the samples within a given class are distributed according to a multivariate normal probability density function. Such distribution is specified completely in terms of a mean vector and a covariance matrix, which must be computed from known samples of the class being represented. This implies that the data analyst must have some prior knowledge (i.e., ground-truth information) of specific areas within the MSS image. Using this ground truth, the analyst must identify training samples for computation of statistics. The histogram (HIST) and gray map (GRAYMAP) processors may be used to aid the analyst.

The HIST processor provides a histogram of data values from the MSS image for use by the GRAYMAP processor. HIST may also be used independently to provide the analyst with information on the distribution of data values within specific user-defined blocks (or fields) of the image. The mean, standard deviation, and range of data within each user-defined field are standard outputs from the HIST processor for each requested channel. Histogram plots may be obtained optionally. With the histogram information, a file (HISFIL) is written automatically for the GRAYMAP processor.

The GRAYMAP processor provides the analyst with a pictorial gray-scale map of any channel of the MSS image for use in obtaining training field coordinates. The map is labeled by sample and scan-line numbers. From this map, the analyst may locate the fields within the image for which he has ground-truth data. Having identified the fields, coordinates (sample and scan-line numbers at each vertex) must be noted for defining the fields to the statistics (STAT) or ISOCLS processor.

Alternatively, the analyst may proceed using Procedure 1 methodology. Using an ERIPS tape as input to the DOTDATA processor, the analyst writes a dot data file. A dot data file contains both type 1 and type 2 dots. Type 1 dots are used both as starting vectors for the clustering processor (ISOCLS) and as labeling vectors for the labeling processor (LABEL).

Type 2 dots are used as a bias correction factor in computing the classification results output by the DISPLAY processor.

The coordinates for training fields may be input to either the STAT or ISOCLS processor for computation of statistics for the classifier. The dot data file is input to ISOCLS to compute statistics. Both processors save the statistics and training field or dot information on a file (SAVTAP, section 4.1) for use in other processors.

In using the STAT processor, the user must group training fields into statistically similar subclasses. Subclasses may be grouped further into classes. For example, three statistically similar subclasses of spring wheat, winter wheat, and harvested wheat may be grouped into one wheat class. Statistics for each subclass are maintained on the SAVTAP file, along with the class grouping. Class groupings are maintained simply for convenience in defining categories in the CLASSIFY processor and for performance reports by the DISPLAY processor. The analyst may obtain the following output for each training field and/or subclass.

- Mean vector
- Covariance matrix
- Correlation coefficient
- Histogram plots
- Spectral plots

In using the ISOCLS processor with training classes, the user must group the training fields into classes. The clustering process breaks the class data into statistically similar subclasses (clusters). Subclasses are given names by taking the first two characters of the class name and two digits indicating the number of the subclass within the class. Again, the statistics for each subclass are saved on the SAVTAP file for use in other processors. ISOCLS is an iterative self-organizing clustering procedure which uses the measure of absolute (L1) distance from a picture element (pixel) to the cluster center to determine the similarity of pixels. At each iteration the user may obtain a cluster summary and map. Optionally, a cluster image data tape (MAPUNT) may be output in either LARSYS III or Universal format (appendixes B and C, respectively).

Under Procedure 1, designated other/designated unidentifiable (DO/DU) fields are delineated by card input. ISOCLS clusters the LACIE segment using the starting vectors from the dot data file to initialize the clustering process. Sun angle correction is provided. An unconditional cluster map and a set of "unlabeled" statistics are output.

The "unlabeled" statistics, cluster map, and dot data file are input to the LABEL processor. Using one of two procedures, k-nearest-neighbor or all-of-a-kind, the statistics are "labeled." A conditional or mixed cluster map may be output to be displayed later on the Passive Microwave Imaging System/Data Analysis Station (PMIS/DAS) or the Interactive Multispectral Image Analysis System, model 100 (Image 100).

By the use of the transform statistics (TRSTAT) processor, the statistics on the SAVTAP file output by the processor STAT or ISOCLS may be transformed according to

$$\left.\begin{aligned} \mu' &= A\mu + b \\ K' &= AKA^T \end{aligned}\right\} \qquad (2\text{-}1)$$

where

μ' = transformed means

A = a matrix

μ = means from SAVTAP file

b = a vector

K' = transformed covariance matrix

K = covariance matrix from SAVTAP file

A^T = transpose of matrix A

The transformed statistics are output as a new file on SAVTAP.

Before proceeding to classification of the MSS image, it may be desirable to reduce the dimensionality of the data vectors by selecting a smaller set of channels or a linear combination of the channels which maximizes some class separability measure.

The SELECT processor (ref. 4) provides this capability. In order to compute the value of the separability measures, the statistics calculated using the STAT or ISOCLS processor must be made available to the SELECT processor by card deck, tape, or disk file, usually the last. The SELECT processor allows the analyst to work with subsets of the statistics on the file, if he desires. Subsets of the statistics are indicated by the CHANNELS and SUBCLASS control cards, which are defined further in section 10.4.3.

In addition, the statistics for two or more subclasses may be grouped together and considered as one subclass. Grouping the statistics for two or more subclasses is equivalent to going back through the STAT processor and combining all training fields for those subclasses being grouped into one subclass. The grouping option is exercised via the GROUP control card defined in table 10.1. The subsets and groupings of the statistics provided to the SELECT processor for computation are used only in SELECT and are not passed on to other processors.

The SELECT processor also allows the analyst to evaluate a given set of channels using one of three different separability measures or to select the best set of channels (k) out of the total channels (n) based on one of these separability measures. The three separability measures provided are:

a. Weighted average interclass divergence

b. Weighted average transformed divergence

c. Weighted average Bhattacharyya distance

To select the best set of k out of n channels, the analyst may use either the Without Replacement Procedure or the Exhaustive Search Procedure. A third procedure will find k linear combinations of n measurements which extremize a given separability measure. This procedure, known as the Davidon-Fletcher-Powell Procedure, outputs the linear combinations in matrix form. All the procedures and equations for separability measures referred to above are discussed in detail in reference 5.

After the SELECT processor has determined the subset or the linear combination of channels which maximizes subclass separability, the supervised classification of the imagery data is performed by the CLASSIFY processor. The options available in SELECT for grouping and selecting subsets of the SAVTAP file are also available in CLASSIFY.

However, once the statistics for classification have been specified, the classes and subclasses are renumbered and referred to in the DISPLAY processor by the new numbers.

The CLASSIFY processor allows the user to group classes previously defined by the STAT or ISOCLS processor into categories for the sum-of-densities classification. When a category is defined (by class names), all subclasses in each class are assigned to the category. The density function for category i, $P_i(x)$, is the sum of densities for all subclasses in the category; that is,

$$P_i(x) = \sum_{j=1}^{k_i} p_j(x) \qquad (2-2)$$

where

p_j = the probability density function for subclass j

k_i = the number of subclasses in category i

(Note: More detailed equations are given in section 11.)

Pixel x is assigned to category i if $P_i(x) > P_k(x)$ for all categories $k \neq i$. Pixel x is further assigned to subclass j if (1) j belongs to category i and (2) $p_j(x) > p_k(x)$ for all subclasses k, $k \neq j$, and k belongs to category i.

Obviously a one-to-one correspondence between categories and subclasses reduces the above equation to:

$$P_i(x) = p_j(x) \qquad (2-3)$$

When this is the case, the amount of computation required for classification can be greatly reduced by the use of thresholds. CLASSIFY then has two procedures for classification which use this computational reduction to advantage. The sum-of-densities rule is used only when categories are defined by the user. Otherwise, the classification-by-thresholding procedure detailed in reference 5 is used.

The CLASSIFY processor writes a file (MAPTAP, section 4.4 and appendix C) containing the subclass number and confidence level for each pixel classified; the training fields and statistics for the classes and subclasses actually used in classification; and the correspondence between categories, classes, and subclasses.

The DISPLAY processor accepts the file output by CLASSIFY and generates a line-printer map of the classified data, along with several performance tables. In the map, each subclass has a symbol associated with it. A threshold option is provided for the analyst to print no symbol (blank) for samples classified with a confidence level less than some specified threshold value.

Performance summaries are provided on subclass, class, and category levels for pixels within each classified field, training field, and test field which are input to the DISPLAY processor. The training or test field performance summaries may be obtained by fields and/or classes. The DISPLAY processor also provides optional output of a classification map (MAPUNT) on tape in either Universal or LARSYS III format.

The data-transformation (DATA-TR) processor allows the analyst to use the linear transformation matrix computed by SELECT to create a new image data tape (TRFORM). Since the matrix is computed to extremize subclass separability, the k linear combinations out of n channels represented by the matrix produce better class contrast when the image is displayed; that is, on the DAS. In addition, the best linear combination of the data can be used to enhance the image.

The TRFORM tape may be output in either the LARSYS III or Universal format.

The NDHIST processor performs an n-dimensional histogram of areas on the MSS data tape (DATAPE), for which the user wishes to create scatter plots. The fields may be histogrammed on a class, subclass, or per-field basis. A line-printer summary of the fields, the number of data vectors in each field, and the number of unique data vectors histogrammed is given.

Optionally, if a scatter plot of a classified or clustered area is requested, a classification or cluster image data tape (MAPUNT) from the DISPLAY or ISOCLS processor must be input to NDHIST. If this option is exercised, the field or fields input to this processor and their order of input must be the same as those input to CLASSIFY or ISOCLS.

Information such as the field, cluster or subclass number, the frequency of occurrence, and color code for each histogrammed radiance vector is written on the n-dimensional histogram (NHSTUN) file.

The SCTRPL processor reads the NHSTUN file, and a two-axis color-coded spectral plot (SCTRUN) is output in the Universal format. The background for the plot may be black or white.

If more than two channels were histogrammed by the NDHIST processor, the data vector is reduced to two components by

$$y' = Ax + b \tag{2-4}$$

where

y' = transformed image

A = matrix

x = data vector

b = bias vector

Columns	Type/format	Definition
11-15	Integer/I5	Sample number of first vertex
16-20	Integer/I5	Line number of first vertex
21-25	Integer/I5	Sample number of second vertex
26-30	Integer/I5	Line number of second vertex
⋮	⋮	⋮
76-80	Integer/I5	Line number of the seventh vertex

- Card type 6 – Class names, nine names per card, left justified in field: The number of cards is determined by the number of classes.

Columns	Type/format	Definition
11-16	Alphanumeric/ A4	Four-character class name for first class
19-24	Alphanumeric/ A4	Four-character class name for second class
27-32	Alphanumeric/ A4	Four-character class name for third class
⋮	⋮	⋮
75-80	Alphanumeric/ A4	Four-character class name for ninth class

- Card type 7 – Number of subclasses in each class, 24 per card: The number of cards is determined by the number of subclasses.

Columns	Type/format	Definition
9-10	Integer/I2	Number of subclasses in first class
12-13	Integer/I2	Number of subclasses in second class
15-16	Integer/I2	Number of subclasses in third class
⋮	⋮	⋮
78-79	Integer/I2	Number of subclasses in 24th class

NASA
National Aeronautics and
Space Administration

80-10138

NASA CR-
160594

Lyndon B. Johnson Space Center
Houston, Texas 77058

JSC-13821

December 1978

EARTH OBSERVATIONS DIVISION VERSION OF THE LABORATORY FOR APPLICATIONS OF REMOTE SENSING SYSTEM (EOD-LARSYS) USER GUIDE FOR THE IBM 370/148

VOLUME II - USER'S REFERENCE MANUAL

Job Order 71-475

(E80-10138) EARTH OBSERVATIONS DIVISION VERSION OF THE LABORATORY FOR APPLICATIONS OF REMOTE SENSING SYSTEM (EOD-LARSYS) USER GUIDE FOR THE IBM 370/148. VOLUME 2: USER'S REFERENCE (Lockheed Electronics Co.) HC A14/MF A01 N80-20725 Unclas G3/43 00138

Prepared By

Lockheed Electronics Company, Inc.

Systems and Services Division

Houston, Texas

Contract NAS 9-15200

For

EARTH OBSERVATIONS DIVISION

SPACE AND LIFE SCIENCES DIRECTORATE

LEC-12564

- Card type 8 — Subclass names, 10 per card: The number of cards is determined by the number of subclasses.

Columns	Type/format	Definition
9-14	Alphanumeric/ A4	Four-character subclass name for first subclass
16-21	Alphanumeric/ A4	Four-character subclass name for second subclass
23-28	Alphanumeric/ A4	Four-character subclass name for third subclass
⋮	⋮	⋮
72-77	Alphanumeric/ A4	Four-character subclass name for 10th subclass

To complete the set of statistics for one subclass, the following three types of cards are grouped together. The number of sets of cards is determined by the number of subclasses.

- Card type 9 — Number of points in this subclass.

Columns	Type/format	Definition
13-20	Integer/I8	Number of points in this subclass

- Card type 10 — Mean vector for this subclass, five values per card: The number of cards is determined by the number of channels.

Columns	Type/format	Definition
6-20	Real/E15.8	Mean for first channel for this subclass
21-35	Real/E15.8	Mean for second channel for this subclass
36-50	Real/E15.8	Mean for third channel for this subclass

Columns	Type/format	Definition
51-65	Real/E15.8	Mean for fourth channel for this subclass
66-80	Real/E15.8	Mean for fifth channel for this subclass

- Card type 11 – Covariance matrix for this subclass: Only the lower triangular portion of the matrix is output; the number of values input for this matrix is equal to (number of channels) × (number of channels + 1)/2. Five values are written on each card image in the order indicated.

$$\begin{bmatrix} 1 & & & & & & \\ 2 & 3 & & & & & \\ 4 & 5 & 6 & & & & \\ 7 & 8 & 9 & 10 & & & \\ - & - & - & - & - & & \\ - & - & - & - & - & - & \\ - & - & - & - & - & - & - \end{bmatrix}$$

Columns	Type/format	Definition
6-20	Real/E15.8	Element 1 of matrix
21-35	Real/E15.8	Element 2 of matrix
36-50	Real/E15.8	Element 3 of matrix
51-65	Real/E15.8	Element 4 of matrix
66-80	Real/E15.8	Element 5 of matrix

3.1.4.2 B-Matrix File

This file is an optional output of the SELECT processor when the Davidon-Fletcher-Powell Procedure is used. The file contains a transformation matrix which extremizes a given separability measure for the subclasses being used. The matrix is optimized using the Davidon-Fletcher-Powell Procedure. The linear transformation of the original measurements can be used in the CLASSIFY, SCTRPL, TRSTAT, or DATA-TR processor to reduce the dimensionality of the data and/or statistics.

The B-matrix deck, or corresponding file, is an optional input to SELECT, SCTRPL, and CLASSIFY and a required input to DATA-TR and TRSTAT (see A-MATRIX control card in table 14-1). When input to SELECT, the matrix is used to evaluate a specific separability measure or it is used as a first guess for the Davidon-Fletcher-Powell Procedure, depending on the user's request. When input to CLASSIFY, classification is performed using the linear transformation. When input to SCTRPL, the dimension of the data from the MSS data tape (DATAPE) is reduced to two linear combinations. When input to TRSTAT, a new file containing the transformed statistics is created on SAVTAP. The DATA-TR processor uses the matrix to create a new image file with the reduced dimensionality.

The keyword B-MATRIX on a control card indicates that the B-matrix is being input. Since the matrix may be on cards or be a disk file, the parameter CARDS or FILE must be specified on the same card in columns 11 through 72. The entire file is defined below by card types.

- Card type 1 — The keyword B-MATRIX in columns 1 through 10 and CARDS or FILE in columns 11 through 72 initialize input of the file.
- Card type 2 — One card of this type.

Columns	Type/format	Definition
6-7	Integer/I2	Number of linear combinations.
13-14	Integer/I2	Number of channels.

Columns	Type/format	Definition
18-80	Integer/30I2	The remainder of this card lists by number the channels for which the matrix was computed; e.g., columns 18 through 19, first channel, etc., for a maximum of 30 channels right justified in the field.

- Card type 3 — The number of these cards is determined by the size of the matrix. The values are input by column as indicated below, five values per card.

B(k,n), k = linear combinations; n = channels

$$\begin{bmatrix} 1 & (k+1) & \cdots & [nk-(k-1)] \\ 2 & (k+2) & \cdots & [nk-(k-2)] \\ 3 & (k+3) & \cdots & [nk-(k-3)] \\ \vdots & \vdots & & \vdots \\ k & 2k & & nk \end{bmatrix}$$

Columns	Type/format	Definition
6-20	Real/E15.8	Element 1 of matrix
21-35	Real/E15.8	Element 2 of matrix
36-50	Real/E15.8	Element 3 of matrix
51-65	Real/E15.8	Element 4 of matrix
66-80	Real/E15.8	Element 5 of matrix (Continued on next card)

3.1.4.3 Cluster Means File

This file is an optional input to the clustering processor ISOCLS. It may be used to initialize the clustering process by estimating cluster centers (means). The means can be taken from the module STAT file (see section 3.1.4.1) created by either TRSTAT, STAT, ISOCLS, or the user. Means may be input for up to 30 channels for each cluster center, and a subset of the channels to be used may be indicated on the CHANNELS control card.

The keyword MEANS in the control cards for ISOCLS indicates initial cluster means are being input. Since the means may be input on cards or from a disk file, the keyword CARDS or FILE must be punched on the same card in columns 11 through 72. If on cards, CARDS initializes input of the cluster means deck which must immediately follow. The format for the entire file is indicated below.

- Card type 1 — Control card keyword MEANS is left justified in columns 1 through 5. The keyword CARDS in columns 11 through 72 initializes input of the card deck.
- Card type 2 — Number of clusters and channels.

Columns	Type/format	Definition
6-10	Integer/I5	Number of initial clusters for which means are provided
25-30	Integer/I5	Number of channels for which means are provided

- Card type 3 — Actual channels used in computation of means.

Columns	Type/format	Definition
6-7	Integer/I2	Channel 1
8-9	Integer/I2	Channel 2
10-11	Integer/I2	Channel 3
⋮	⋮	⋮
64-65	Integer/I2	Channel 30

- Card type 4 — Mean vectors for the initial clusters: These cards are in the same format as the means cards (card type 10) in the module STAT file. The first mean for each cluster always begins on a new card. The number of cards depends on the number of channels and the number of clusters. Five values are placed on each card.

Columns	Type/format	Definition
6-20	Real/E15.8	Mean for channel 1
21-35	Real/E15.8	Mean for channel 2
36-50	Real/E15.8	Mean for channel 3
51-65	Real/E15.8	Mean for channel 4
66-80	Real/E15.8	Mean for channel 5 (Continued on consecutive cards of the same format)

3.2 MSS IMAGE DATA TAPES

Every processor except SELECT, LABEL, TRSTAT, and SCTRPL uses an MSS data tape (DATAPE). The tape assignment defaults to logical unit 11, but the user may assign any unit available by input of the DATAFILE control card. For details, see the file assignment chart in section 5 and the control card section for each processor. The tape may be in either the LARSYS III format or the Universal format. These formats are defined in appendixes B and C, respectively.

The control card DATAFILE allows the user to communicate the file number of the MSS data tape (DATAPE) to be processed and the logical unit assignment. This is optional input to every processor that requires the MSS data tape. The first file of the tape will be processed unless otherwise specified by the DATAFILE control card. In executing the same and/or different processors back to back, the DATAFILE control card may be input only to the first processor executed if the same file and logical unit are to be used throughout the execution. For example,

```
$HIST                            [File 2 of the MSS data tape
DATAFILE  UNIT=11,FILE=2         (DATAPE) assigned to unit 11
:  (Other control cards)         is processed by GRAYMAP as
.                                well as HIST.]
*END
(Field definition)
$END
$GRAYMAP
CHANNELS   5,6
*END
(Field definitions)
$END
```

4. SYSTEM INTERNAL FILES

The files described in this section are used internally by the system to pass information between processors. It is the user's responsibility to assign the necessary files for his particular job. In the discussion that follows, the names of the units on which the files are written are also used to identify the files.

4.1 STATISTICS FILE (SAVTAP, UNIT 20)

This file must be assigned either to disk or to tape, normally the former, whenever one or more of the processors STAT, SELECT, CLASSIFY, ISOCLS, LABEL, NDHIST, SCTRPL, TRSTAT, or DATA-TR is executed. One file is written on this unit for each execution of STAT, TRSTAT, or ISOCLS or for input of a module STAT file to some other processor. The file contains the same information as itemized in section 3.1.4.1 for the module STAT file.

Multiple files may be written on a single unit, usually disk or tape, and may be accessed by using the STATFILE control card. This control card communicates the file number for positioning the unit and the logical unit assignment. The first file is always assumed unless otherwise specified by the user, and the unit assignment assumes logical unit 20 unless otherwise specified by the STATFILE control card. In executing several processors back to back and in referencing the same file, only one STATFILE control card need be submitted. If different file numbers are to be referenced during one execution, then the file number may be changed from one processor to the next by input of the STATFILE control card to each processor.
For example,

```
$STAT
STATFILE  UNIT=20,FILE=2
(Other control cards)
*END
(Class, subclass, and field definitions)
```

```
$END
$CLASSIFY
*END
(Fields to be classified)
$END
```

The STAT processor will write the training statistics for this run on file 2 of the SAVTAP file (unit 20). (The system files, their logical units, and assignments are set out in table 4-1.) CLASSIFY will use all of the statistics on file 2 of the tape for classification.

The following example shows assignments for back-to-back executions of STAT, ISOCLS, and SELECT.

```
$STAT
STATFILE  UNIT=20,FILE=2
(Other control cards)
*END
(Class, subclass, and field definitions)
$END
$ISOCLS
STATFILE  UNIT=20,FILE=3
(Other control cards)
*END
(Class and field definitions)
$END
$SELECT
STATFILE  UNIT=20,FILE=2
BEST      4
*END
$END
```

STAT will write on file 2 of unit 20, ISOCLS will write on file 3 of the same unit, and SELECT will go back to file 2 of unit 20 for the statistics computed by STAT.

4.2 B-MATRIX FILE (BMFILE, UNIT 10)

The file written to the BMFILE unit contains the transformation matrix which corresponds to the B-matrix file (section 3.1.4.2). The file is an optional input to SELECT, SCTRPL, and CLASSIFY and a required input to DATA-TR and TRSTAT. When the card deck is input to any of these processors, this file is automatically written. The B-matrix is computed by the SELECT processor and is automatically output to the file when the Davidon-Fletcher-Powell Procedure is executed.

4.3 ONE-DIMENSIONAL HISTOGRAM FILE (HISFIL, UNIT 13)

On logical unit 13, the HIST processor creates the HISFIL file, which is used by the GRAYMAP processor.

4.4 CLASSIFICATION MAP FILE (MAPTAP, UNIT 2)

The MAPTAP file (appendix D), which is output by CLASSIFY, contains the statistics actually used in the classification, the training field information, and all of the classification results.

4.5 N-DIMENSIONAL HISTOGRAM FILE (NHSTUN, UNIT 4)

The NDHIST processor writes a file to the NHSTUN unit to be used as an interface to the SCTRPL processor. The device default assignment is unit 4, but the user may assign any available unit. The NHSTUN file format is defined in appendix E. In earlier documentation of EOD-LARSYS, the file written on NHSTUN, unit 4, is referred to as the NDIM file.

4.6 TRANSFORMED STATISTICS FILE (SAVTAP, UNIT 20)

The TRSTAT processor writes the transformed statistics on the SAVTAP file. (See section 4.1 for further information.)

4.7 DOT DATA FILE (DOTUNT, UNIT 19)

The DOTDATA processor writes a multifile unformatted file on the DOTUNT unit. The files contain information extracted from the MSS data tape, using all or a subset of 209 possible grid points (dots). The file created on the DOTUNT unit is an interface for the processors ISOCLS, LABEL, and DISPLAY. The format of the dot data file is defined in appendix H.

5. SYSTEM OUTPUT FILES

5.1 CLUSTER MAP FILE (MAPUNT, UNIT 16)

On logical unit 16 the DISPLAY processor optionally outputs a multifile data tape (MAPUNT) containing the subclass number to which each corresponding pixel was assigned during classification by CLASSIFY. Also, on logical unit 16, the ISOCLS processor outputs a file containing either the cluster number (OPTION CLUSTER control card) or the mean vector to which each corresponding pixel was assigned during clustering. A color key containing the color code for each cluster is given for the mean vectors. The color codes optionally may be ordered according to the cluster number or to greenness (OPTION ORDER control card). (See section 9.5.3, table 9-1, for ISOCLS control cards.)

The results of the classification/clustering may be displayed on the PMIS or the Bendix 100 DAS. The tape necessary must be mounted on a nine-track tape drive compatible with the DAS and may be output in either the LARSYS III or Universal format. The display may be made without the color keys (appendices B and C) or with color keys (see appendix F for tape format). To exercise this option, see FORMAT control card (table 9-1) for the ISOCLS and section 12 (table 12-1) for the DISPLAY processor.[1]

One file is written on the output tape for each field classified or clustered. In earlier documentation of EOD-LARSYS, the file written to the MAPUNT unit is referred to as MAPFIL.

---

[1]These data are available as input to NDHIST via seven- or nine-track tape or disk.

5.2 SCATTER PLOT DATA FILE (SCTRUN, UNIT 12)

The SCTRPL processor outputs two-axis color-coded spectral plots on a multifile Universal-formatted tape. The file default assignment is unit 12, but the user may assign any available unit. (See file assignment chart, table 5-1.) The relevant tape format is defined in appendix G. In earlier documentation of EOD-LARSYS, the file written to the SCTRUN unit is referred to as PLOTAP.

5.3 TRANSFORMED DATA FILE (TRFORM, UNIT 14)

The DATA-TR processor outputs a multifile image file of transformed data. The image file may be produced in either the LARSYS III or Universal format defined in appendixes B and C, respectively. The tape must be assigned to logical unit 14.

TABLE 5-1.— OVERVIEW OF EOD-LARSYS FILES

Processor	Input			Output		
	Unit (a)		File	Unit		Comments
	No.	Name		No.	Name	
DOTDATA	11	DATAPE	Field cards (dots)	19	DOTUNT	
ISOCLS	11 10 20 19	DATAPE BMFILE SAVTAP DOTUNT	Field cards	20 16	SAVTAP MAPUNT	
LABEL	20 19 16	SAVTAP DOTUNT MAPUNT	Field cards from ISOCLS	16 2 20 19	MAPUNT MAPTAP SAVTAP DOTUNT	Conditional or mixed cluster map for DISPLAY Relabeled Relabeled
CLASSIFY	11 10 20	DATAPE BMFILE SAVTAP	Field cards	2	MAPTAP	
DISPLAY	11 19 2	DATAPE DOTUNT MAPTAP	Field cards — DO/DU or test fields	16	MAPUNT	
SELECT	20 10	SAVTAP BMFILE		10	BMFILE	
STAT	11	DATAPE	Field cards	20	SAVTAP	
DATATR	11 10	DATAPE BMFILE	Field cards	14	TRFORM	
TRSTAT	20 21	SAVTAP CRDUNT	Affine transformation	20	SAVTAP	
NDHIST	11 16	DATAPE MAPUNT	Field cards	4	NHSTUN	Histogrammed by class, subclass, or field
SCTRPL	20 4 10	SAVTAP NHSTUN BMFILE		12	SCTRUN	
HIST	11	DATAPE	Field cards	13	HISFIL	
GRAYMAP	11 13	DATAPE HISFIL	Field cards			

[a]Other logical units are PCHUNT (punch unit, 7), PRTUNT (printer, 6), and RANDIO (direct access temporary file, 22). Hand-coded units are: RANDIO (22), CRDUNT (21), and PRTUNT (6).

The location on the scatter plot for each vector in the NHSTUN file is determined by its radiance values (if only two channels were histogrammed) or by two linear combinations of radiance values (if more than two channels were histogrammed).

The color for the pixel is assigned by

- Original radiance values
- Mean value of the subclass or cluster to which the pixel was assigned during classification or clustering
- Mean value of the test or training field from which the pixel was extracted
- User-defined colors
- Color extraction from a different pass when using multiregistered Landsat data.

Optionally, for pixel color assignment, the SAVTAP file created by STAT or ISOCLS may be input.

Optionally, a line-printer pixel frequency or log of pixel frequency (base 2) plot is given. The plot is printed with up to 16 different symbols.

Figure 2-1 is a diagram showing the principal processing options and paths in EOD-LARSYS.

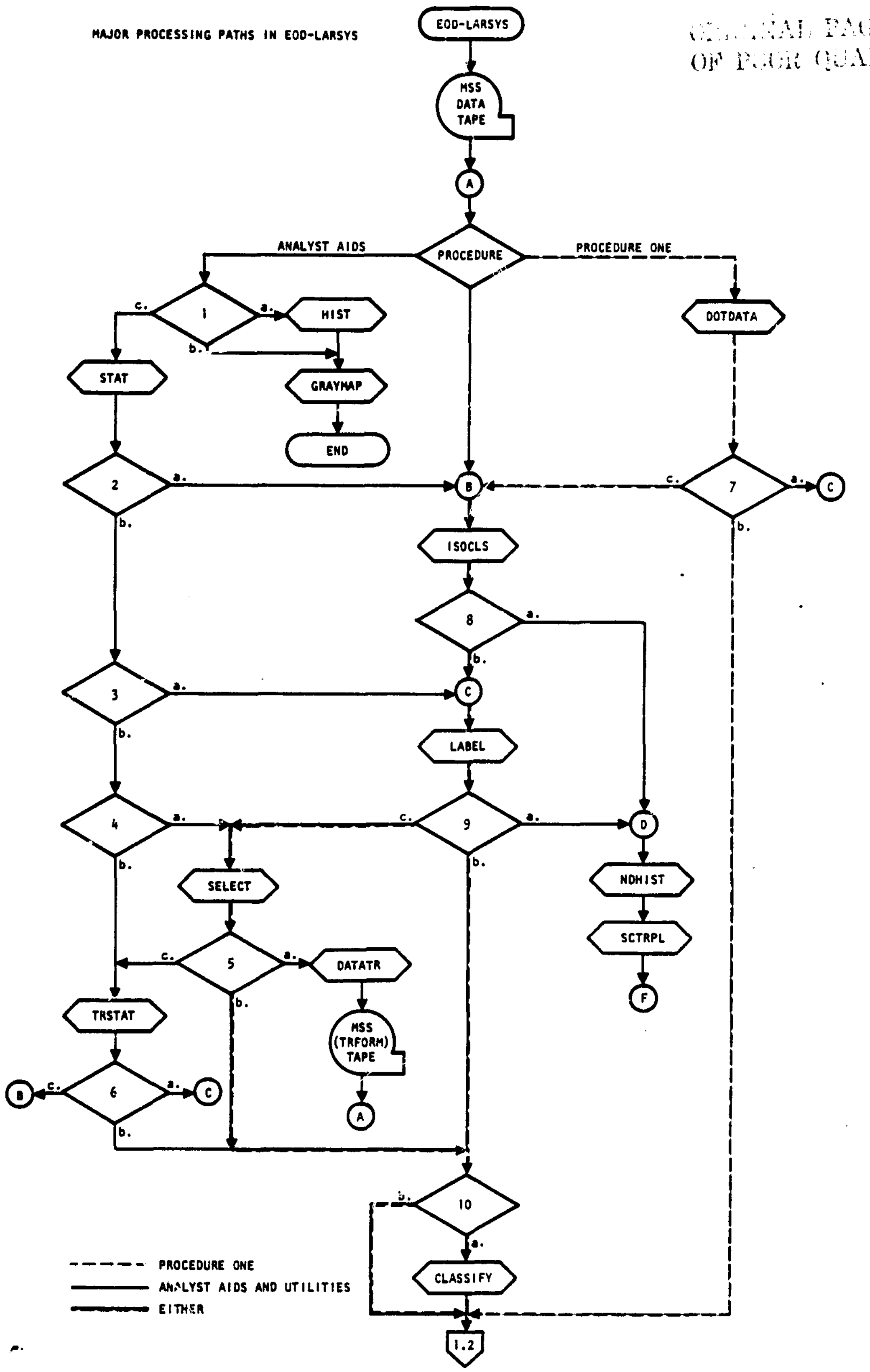
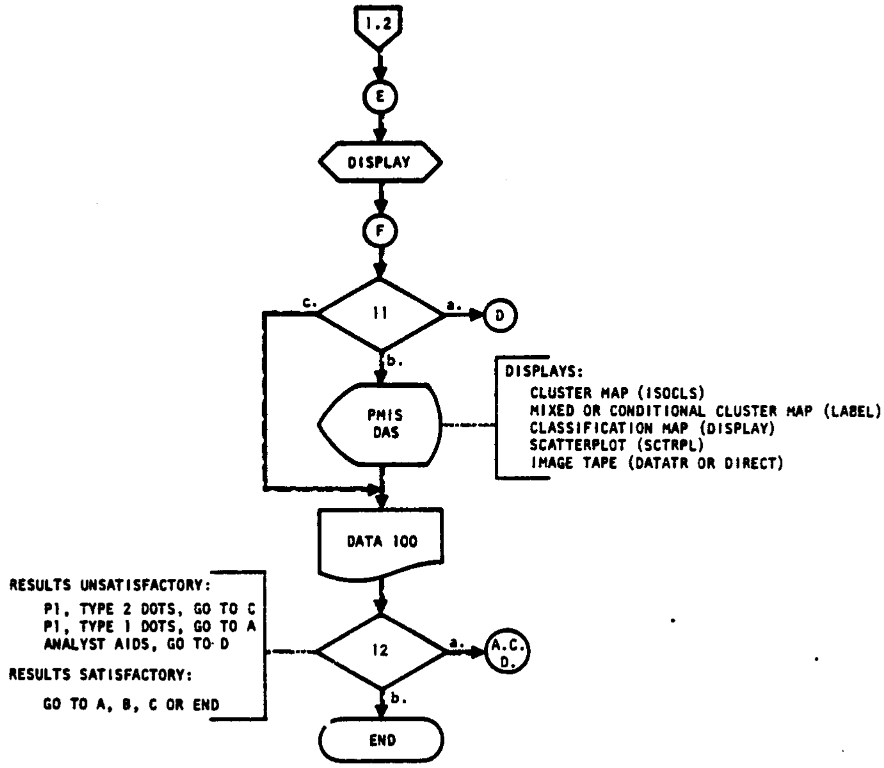


Figure 2-1. - Major processing paths in EOD-LARSYS.



Key to decision points 1 through 12.

1 a. Compute histograms and print pictorial gray scale map of data from any channel, using HIST and GRAYMAP.
b. Print pictorial gray scale map only, using GRAYMAP.
c. Compute training field statistics, write SAVTAP file, using STAT.

2 a. Group training fields into classes, using ISOCLS.
b. Omit clustering.

3 a. (Re)label training field statistics.
b. Omit labeling.

4 a. Determine subset or linear combination of channels that maximizes subclass separability, using SELECT.
b. Transform training field statistics using TRSTAT.

5 a. Create new image data tape, applying linear transformation matrix computed by SELECT, using DATATR.
b. Perform supervised classification of image using CLASSIFY.

6 a. Relabel previously labelled statistics file SAVTAP, using LABEL.
b. Proceed to classification, using CLASSIFY.
c. Use statistics file SAVTAP to provide starting cluster mean vectors for ISOCLS.

7 a. Relabel dots in dot data file using LABEL.
b. Display dots using DISPLAY.
c. Cluster image using ISOCLS.

8 a. Compute n-dimensional histogram of selected data areas, using NDHIST, and scatter plot, using SCTRPL.
b. Proceed to labeling, using LABEL.

9 a. Compute n-dimensional histogram of selected data areas, using NDHIST, and scatterplot, using SCTRPL.
b. Proceed to classification, using CLASSIFY.
c. Evaluate discriminatory capability of channels, using SELECT.

10 a. Proceed to classification, using CLASSIFY.
b. Proceed to classification summary, using DISPLAY.

11 a. Compute n-dimensional histogram of selected data areas, using NDHIST, and scatterplot, using SCTRPL.
b. Display image on display station and print results.

12 a. Results unsatisfactory (see annotation on flowchart).
b. Results satisfactory (see annotation on flowchart).

Figure 2-1. - Concluded.

3. SYSTEM INPUT/OUTPUT FORMATS

3.1 CARD INPUT/OUTPUT

Card input to and output from the system must be one of the types listed below. It should be noted that card image files are normally used in remote processing applications. In the discussion following, "card image" should be understood for "card."

3.1.1 PROCESSOR CARDS

Processor cards identify the processor that is to be executed. The system monitor routine calls the appropriate processor, which initiates the loading of all routines used by the processor. The processor card is always a $ symbol followed by the processor name and must always be punched left justified beginning in column 1. No blanks are allowed. The $ symbol and the first five characters are the unique processor identification used by the system monitor routine.

Below is a list of all processor cards recognized by the system, along with the section in which each processor is described.

$HIST	Section 6
$GRAYMAP	Section 7
$STAT	Section 8
$ISOCLS	Section 9
$SELECT	Section 10
$CLASSIFY	Section 11
$DISPLAY	Section 12
$DATA-TR	Section 13
$TRSTAT	Section 14
$NDHIST	Section 15
$SCTRPL	Section 16

$DOTDATA	Section 17
$LABEL	Section 18
$EXIT	Execution terminates when this card is encountered.

3.1.2 CONTROL CARDS

Each processor has its own set of control cards which allow the user to exercise various options in the particular processor or to change the default value assigned to certain parameters in the system. These cards must immediately follow a processor card. The control cards are identified by a keyword in columns 1 through 10 of the card. Only the first six characters are used for testing. In columns 11 through 72, the parameter values or options are indicated. These columns are free form, blanks are ignored (unless of legitimate parameter value), and multiparameter values or options are separated by commas. Columns 73 through 80 of the card are not used. With the exceptions of the FORMAT, *END, $END, and in some cases the STATFILE cards, control cards may occur in any order. (The STATFILE control card exception is noted in the section for the appropriate processor.) If the list of parameter values for a given keyword is too long for one card, the remaining values can be input on another card with the same keyword. (The continuation of a CATEGORY control card is slightly different; see section 11, table 11-1.) In every processor, the *END control card indicates the end of a set of control cards, and the $END card indicates the end of field definition card input. The FORMAT card defines the format of the MSS data tape by a 1 (Universal) or by a 2 (LARSYS III) in column 11 (or subsequent columns). This card precedes all others in a job setup. The user should ensure that all files written in the run are consistent in format.

3.1.3 CLASS, SUBCLASS, AND FIELD DEFINITIONS

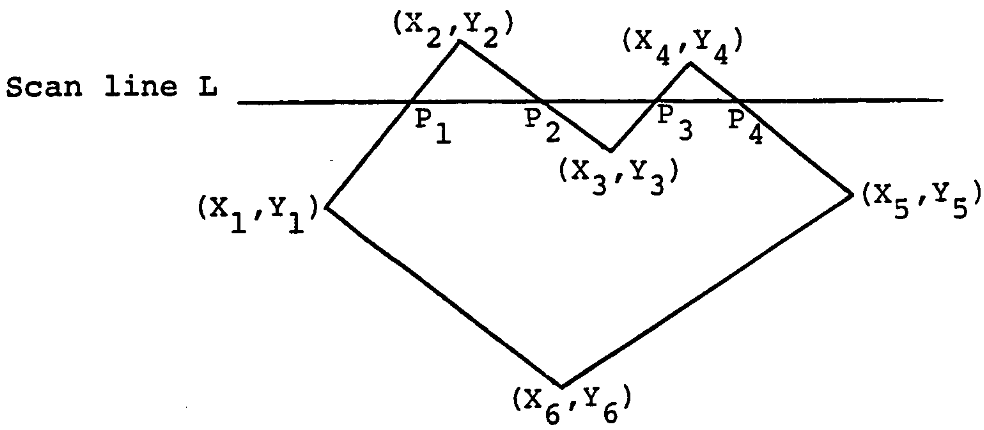
A field is a specific block of data to be extracted from the input MSS data tape (DATAPE) and processed. It is defined by a sample increment, a line increment, and from 1 to 10 vertices. Optionally, the user may associate a name with each field. The alphanumeric field description is located in columns 1 through 6. In columns 11 through 72, sample and line increments are separated by a comma and enclosed in parentheses. A comma separates the increments and each of the following vertices. The vertices must be arranged in clockwise order. Sample and line numbers which describe a vertex are separated by a comma and enclosed in parentheses. The sample number must be given first for each vertex. More than one card may be used to describe a field. An asterisk occurring after a vertex indicates a continuation card is to be read beginning in column 11. A vertex must be completed on a card and cannot be split between two cards. The numbers which describe the increments and vertices must be integers.

It is the user's responsibility to ascertain that all defined fields are within the bounds of the MSS image. In determining which pixels belong in a particular field, the EOD-LARSYS examines the pixel intercepts of each scan line with each side of the field. The pixel intercept X, with the scan line L and the side defined by vertices (X_1,Y_1) and (X_2,Y_2), is calculated by the equation:

$$X = \frac{(L - Y_1)(X_2 - X_1)}{(Y_2 - Y_1)} + X_1 \qquad (3\text{-}1)$$

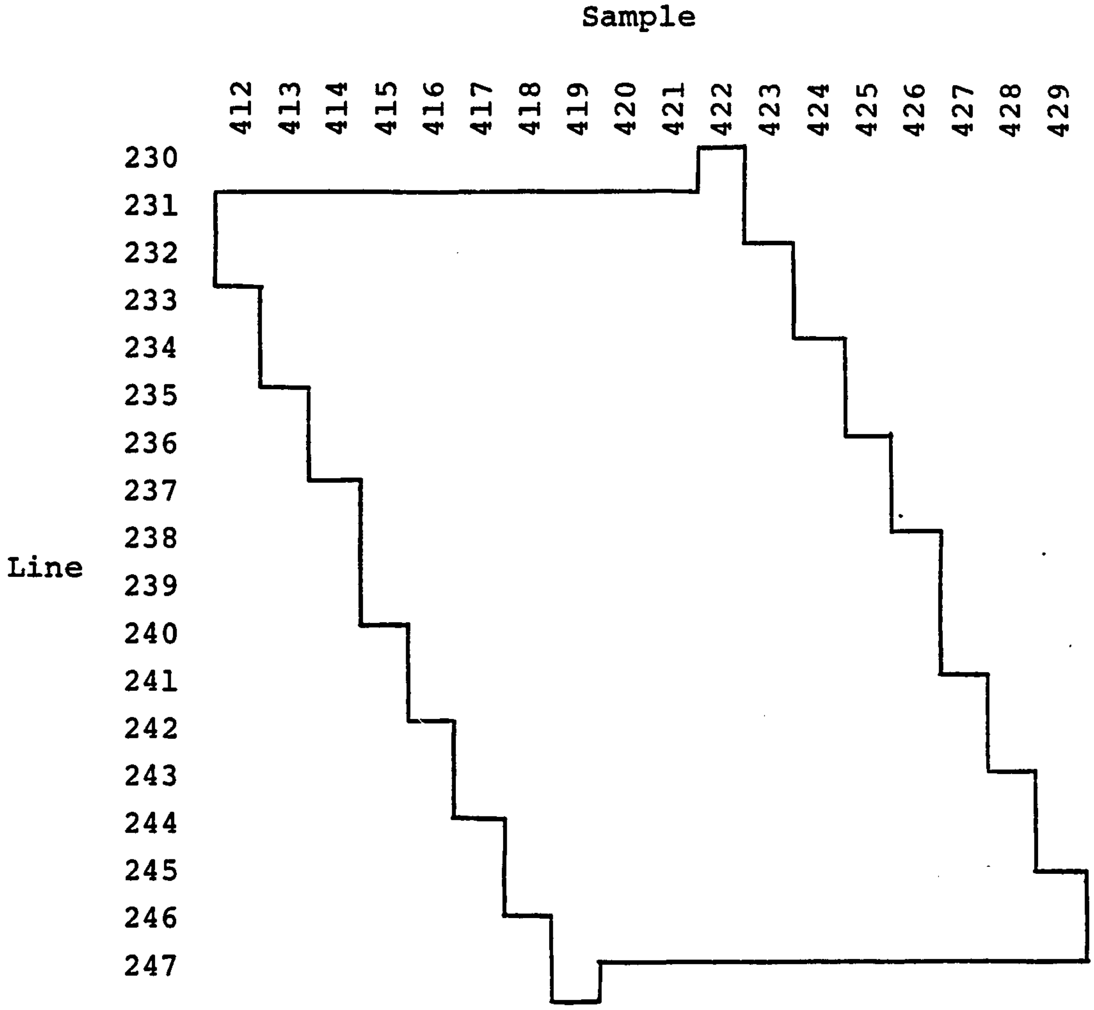
The value of X is computed as a floating-point number; however, the actual pixel intercept must be an integer number. Therefore, if the fractional part of X is greater than one-half, the pixel intercept is the next higher integer number. If the fractional part of X is less than one-half, the pixel intercept will be the next lower integer number. When the fractional part of X is exactly one-half, the integer pixel intercept depends on the direction of movement from the point (X_1,Y_1) to (X_2,Y_2). If Y_1 is less than Y_2, the pixel intercept is the next higher integer. If Y_1 is greater than Y_2, the pixel intercept is the next lower integer number.

After all pixel intercepts for a given scan line have been determined, the intercepts are taken in pairs and all pixels between and including the pair of intercepts are included in the field. In the following example for scan line L, all pixels between and including P_1 and P_4 are included, and all pixels between and including P_3 and P_4 are included.



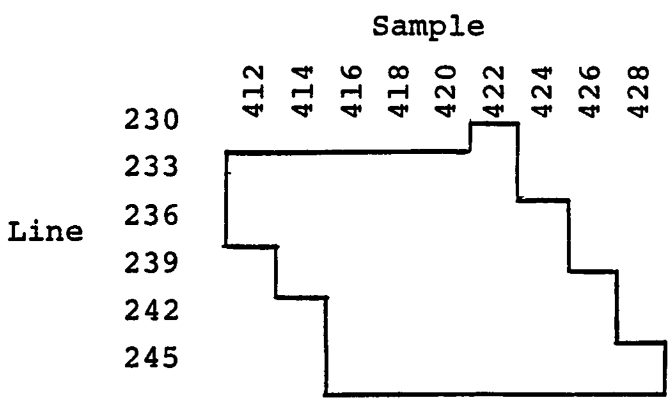
The following three examples describe field definition cards and the fields they describe. In example 1, the sample and line increments are equal to 1 for field F1, and there are four vertices.

F1 (1,1),(412,231),(422,230),(429,246),(419,247)



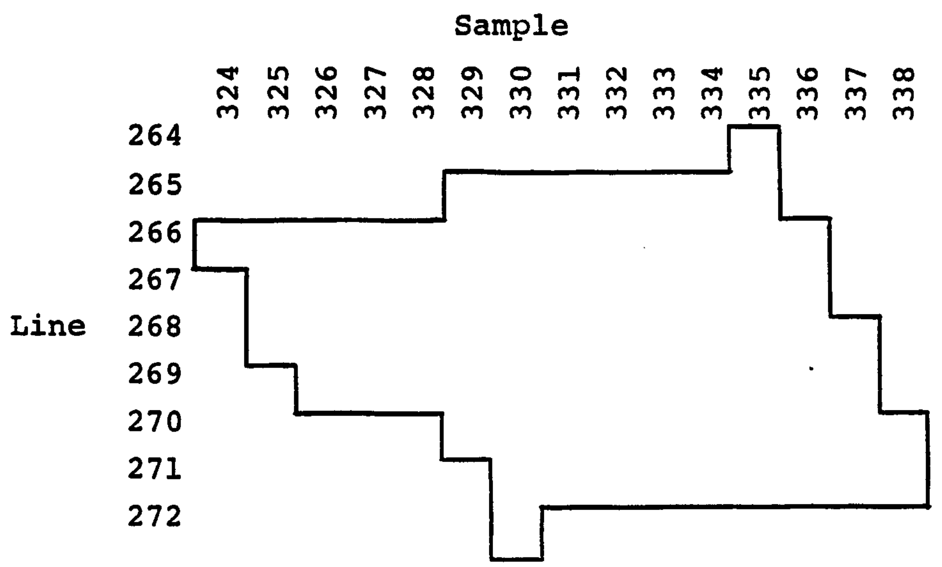
In example 2, the field F2 has the same vertices as F1; however, the sample increment is 2 and the line increment is 3.

F2 (2,3),(412,231),(422,230),(429,246),(419,247)



In example 3, the sample and line increments for field W187 are equal to 1, and there are six vertices.

W187 (1,1),(324,266),(335,264),(338,271),
(330,272),(329,269),(326,269)



Except for SELECT, TRSTAT, and Scatter Plot (SCTRPL), every processor accepts the input of field definition cards. Field definitions are always input between the *END and the $END control cards. In the STAT and ISOCLS processors, fields must be associated with a class or subclass name. In the DISPLAY processor, fields may be test fields or designated fields. In the NDHIST processor, fields are associated with class, subclass, test or training, or any user-defined field.

The fields defined in STAT and ISOCLS are called training fields, and the data within these fields are used for computing statistics. Training fields are grouped into subclasses, and subclasses are further grouped into classes, using the STAT processor. In ISOCLS, training fields are grouped into classes, and the clustering procedure breaks the class data into subclasses (clusters). To allow for these groupings, cards bearing a class name and a subclass name are necessary.

A class name card has the keyword CLASSNAME beginning in column 1 and the alphanumeric name of the class left justified in columns 11 through 14 of the card. Blanks should not be embedded in the class or subclass names.

A subclass name card has the keyword SUBCLASS beginning in column 1 and the alphanumeric name of the class left justified in columns 11 through 14 of the card.

In STAT, a CLASSNAME card must immediately follow the *END control card. The CLASSNAME card is followed by one or more SUBCLASS cards, each of which must be followed by one or more field definition cards. See the example for STAT (section 8.4.4).

In ISOCLS, a CLASSNAME card must immediately follow the *END control card. The CLASSNAME card is immediately followed by one or more field definition cards. The data from the fields associated with a given class name are clustered as one data set. The class is broken into subclasses (clusters) which do not have field boundaries. So, even though statistics are computed on a subclass level, training fields cannot be associated with subclasses in ISOCLS. See the example for ISOCLS (section 9.5.4).

In DISPLAY, test fields (if input) must be identified by a previously defined class or subclass name. When associated with classes, a CLASSNAME card should immediately follow the *END control card. Test fields for that class should follow immediately. When associated with subclasses, a SUBCLASS card should immediately follow the *END card, followed by the test fields for that subclass.

Designated fields are the other type of field input to DISPLAY. Fields may be DU or DO. For input of designated fields, a card with the keyword DESIGNATE beginning in column 1 and the keyword OTHER or UNIDENTIFIABLE beginning in column 11 must precede the field definition cards. See section 12.4.4 for sample input of test and designated fields.

3.1.4 SPECIAL SYSTEM FILES

The card image files described in this section are special files normally output from one processor to be used at some future time for input to another processor. However, if the user can obtain all of the information needed for any of the special card image files from some other source, such information may be input directly to the processor if the formats described in this section are followed.

These files are always referred to in the job setup with the control cards for the particular processor. The first card image of each file acts as a keyword which initiates the input of the file. It is not necessary to input the same file to more than one processor in the same run.

3.1.4.1 Module STAT File

The module STAT file is optional output from the STAT, ISOCLS, and TRSTAT processors. It contains either the statistics (mean vectors and covariance matrices) for all the subclasses input to STAT or for clusters computed by ISOCLS or the transformed statistics for all subclasses or clusters input to TRSTAT. These statistics are needed in the computation of the probability density function in CLASSIFY and the computation of separability measures in SELECT.

This file also contains all the training field boundaries, the class and subclass numbers to which the training fields belong, the class and subclass names, the number of subclasses in each class, and the number of points in each subclass or cluster. By defining the required training fields in STAT, the user has absolute control over the data samples which will define a subclass from the MSS data tape (DATAPE). Every data sample occurring in any one of the training fields defined by a particular subclass is used in computing the mean vector and covariance matrix for that subclass.

In the clustering processor ISOCLS, the user has no control over the specific samples which comprise a cluster. The processor determines which data samples are used in computing the mean vector and covariance matrix for each cluster. Because of the desirability of using these cluster statistics in other processors, the ISOCLS processor creates a file in the same format as the STAT processor. The file may be punched if desired. Training fields are associated with classes rather than subclasses. Clusters are given a six-character name. The first three characters are the first three characters of the class name associated with the cluster, and the last three characters are digits. The digits for the subclasses are in sequential order.

When the module STAT file is input to the CLASSIFY or SELECT processor, the user may request subsets of the statistics to be used for classification or channel selection via the CHANNELS and SUBCLASS control cards in both processors. Subclasses are numbered as they were input to STAT, and clusters are numbered as they were created in ISOCLS. The channels are numbered as they occur on the MSS data tape (DATAPE). To select a subset of the statistics in the module STAT file, the user should indicate by number the subclasses and/or channels he wishes to use. (Unless the user has previous knowledge of the number of clusters in the module STAT file, he or she cannot accurately select a subset of the clusters when executing ISOCLS back to back with another processor.)

The first card in the module STAT file acts as a control card, with the keyword MODULE initializing the input of the remainder of the file. The entire file is composed of the card image types listed below. All integers should be right justified in the specified field, and alphanumeric characters should be left justified in the specified field.

- Card type 1 — Keyword MODULE in columns 1 through 6.
- Card type 2 — Number of classes, subclasses, channels, fields, and vertices for training fields.

Columns	Type/format	Definition
7-10	Integer/I4	Number of training classes from STAT or ISOCLS
19-20	Integer/I2	Number of training subclasses from STAT (clusters from ISOCLS)
29-30	Integer/I2	Number of channels used in computation of statistics
38-40	Integer/I3	Number of training fields input to STAT or ISOCLS
49-52	Integer/I4	Number of vertices in all the training fields

- Card type 3 — Actual channels used in computation of statistics.

Columns	Type/format	Definition
11-12	Integer/I2	Channel 1
13-14	Integer/I2	Channel 2

Columns	Type/format	Definition
15-16	Integer/I2	Channel 3
⋮	⋮	⋮
69-70	Integer/I2	Channel 30

- Card type 4 — Training field information: The first card of the set.[1] Names should not exceed four characters.

Columns	Type/format	Definition
1-6	Alphanumeric/ A4	Field name
11-12	Integer/I2	Number of the class associated with this field
21-22	Integer/I2	Number of the subclass associated with this field input to STAT (Since ISOCLS associates fields with classes, ISOCLS dummies this information by setting it equal to zero.)
31-32	Integer/I2	Number of vertices for this field, including closure point

- Card type 5 — Vertices for the training field: Up to 10 vertices plus the closure point are allowable for each training field, 7 vertices per card with coordinates ordered (sample, line). The coordinates are listed in a clockwise manner, with the coordinate having the smallest sample number listed first.[1]

---

[1]Card types 4 and 5 define a training field. To complete the set of information for one training field, one card of type 4 and one or two cards of type 5 are required. The number of card sets is determined by the number of training fields.

6. ONE-DIMENSIONAL HISTOGRAM PROCESSOR — HIST

The processor HIST computes individual field histograms and a total histogram for all the fields and channels defined by the user. An individual statistics report is printed for every field histogrammed. The report contains field descriptions, data ranges, means, standard deviations, and normalized ranges (mean ±3 standard deviations).

A cumulative histogram of all the fields is calculated and written on an internal file to be read later by the GRAYMAP processor. Like the field histograms, a statistics report is printed for the combined fields.

The input DISPLAY control card allows the user to obtain a line-printer plot of the histograms. A histogram for each channel on the DISPLAY card (described in table 6-1) is displayed for each field, along with a cumulative histogram for all the fields.

6.1 INPUT FILES

An MSS data tape (DATAPE) must be input. The tape assignment defaults to logical unit 11; but, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information on format.)

6.2 OUTPUT FILES

The HIST processor writes a file for the GRAYMAP processor on logical unit 13. This file (HISFIL) contains the histogram data for each channel requested.

6.3 SCRATCH FILES

The HIST processor does not require an additional scratch file.

6.4 CARD INPUT

The formats for all system card input are defined in section 3.1.

6.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$HIST

This card directs the system monitor routine to select the HIST processor and causes all the routines used by the HIST processor to be loaded into the system.

6.4.2 SPECIAL SYSTEM FILES

The HIST processor does not use any special input files.

6.4.3 CONTROL CARDS

Control cards allow the user to input various options. These cards are identified by a keyword left justified in columns 1 through 10 of the card, with the parameter values or additional keywords in columns 11 through 72 (beginning in any column after column 10). These control cards may be in any order, but they must be the first cards after the processor card $HIST. Table 6-1 lists all available options, along with their default values.

6.4.4 FIELD DEFINITIONS

Fields to be histogrammed are input immediately following the *END control card. The card column format for field definitions is defined in section 3.1.2. Input of field definition cards is terminated by the $END control card.

6.5 CARD OUTPUT

This processor does not output any card decks.

6.6 RESTRICTIONS

a. The maximum number of channels is 30.

b. The number of histograms requested to be plotted may be limited if internal dimensions are too small for all user requests. (For example, if the user requests 30 channels to be histogrammed, only 14 of those histograms may be plotted; however, all 30 will be histogrammed.)

   This limitation is a function of the number of channels requested on the CHANNELS control card. If too many channels are indicated on the DISPLAY control card, a diagnostic is printed but execution continues.

c. The DISPLAY card must be a subset of the CHANNELS card.

d. The data for all channels for one scan line are unpacked into an array dimensioned 12 000. If the number of channels times [(sample end - sample begin)/sample increment] exceeds 12 000, a diagnostic message is printed. Sample end is reset to fit the dimensions and execution continues.

6.7 DIAGNOSTIC MESSAGES

The diagnostic messages and the subroutines in which they appear are as follows.

6.7.1 SUBROUTINE HISTGM

Message	Explanation
TOO MUCH DATA REQUESTED -- SAMPLE END WAS RESET TO ________.	The data for all channels for one scan line are unpacked into an array dimensioned 12 000. If the number of channels times [(sample end - sample begin)/sample increment] exceeds 12 000, this diagnostic is printed. Sample end is reset to fit the dimensions and execution continues.

6.7.2 SUBROUTINE HISTIC

Message	Explanation
ONLY THE FIRST 50 FIELD DESCRIPTIONS WERE PRINTED, BUT ALL THE FIELDS WERE INCLUDED IN THE TOTAL HISTOGRAMMED STATS.	The user has input more than 50 fields, and only the first 50 field descriptions will be printed in the "Data Blocks Histogrammed" portion of the total report; however, all the input fields were included in the calculations of the "Total Histogrammed Statistics."

6.7.3 SUBROUTINE SETUP5

Message	Explanation
a. ERROR ON DATAFILE CARD.	Check unit assignment and file number.

Message	Explanation
b. TOO MANY CHANNELS ARE BEING HISTOGRAMMED AND PLOTTED, NO. OF CHANNELS WAS RESET TO __________.	User requested too many histograms to be plotted. The number of histograms plotted varies according to the number of channels histogrammed.
c. INVALID CARD -- IGNORED.	Inappropriate or defective card read. Make sure cards are punched correctly.
d. XHIGH - XLOW WAS LESS THAN 100. XHIGH WAS RESET TO XXX, OR XLOW WAS RESET TO XXX.	Range of pixel radiances required to be $\geq$ 100.
e. CHANNEL, I2, IS NOT A SUBSET OF THE CHANNELS GIVEN ON CHANNELS CARD.	A channel on the DISPLAY card is not a member of the set of channels on the CHANNELS card.
f. BAD SUPERVISOR CONTROL CARD.	Check spelling of keywords.

TABLE 6-1.– HIST PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNELS	$C_1,C_2,C_3,\cdots,C_k$ $k\leq 30$ Default: None	Channels to be histogrammed, $C_1,C_2,C_3,\cdots,C_k$, should be integer numbers separated by commas.
SIZE	XHIGH=K 0<K<255 Default: XHIGH=255	K is an integer which sets the maximum radiance value which will be histogrammed. XHIGH becomes X_{max} on the X-axis of the histogram plot.[c]
SIZE	XLOW=J 0<J<XHIGH Default: XLOW=0	J is an integer which sets the minimum radiance value which will be histogrammed. XLOW becomes X_{min} on the X-axis of the histogram plot.[c]
SIZE	YSIZ=L $0<L<f(x)_{max}$ Default: YSIZ=15	L is an integer which sets the height of the Y-axis (number of print lines). Using the input YSIZ, the Y-axis scale for the histogram plot will be determined by the processor to be: $f(x)_{max}$+(YSIZ-1)/YSIZ.

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

[c]The difference between XHIGH and XLOW must be at least 100.

TABLE 6-1.– Continued.

Keyword	Parameter and default values	Function
DISPLAY	$C_1, C_2, C_3, \cdots, C_k$ $k \leq 30$ Default: No plots	Channels for which histograms will be plotted. $C_1, C_2, C_3, \cdots, C_k$ must be a subset of the CHANNELS card.
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	N is the Fortran logical unit number to which the MSS data tape (DATAPE) has been assigned; M is the file number for the tape to be processed. For back-to-back executions of more than one processor, if using the same file number, only one DATAFILE control card need be submitted.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.

TABLE 6-1.– Concluded.

Keyword	Parameter and default values	Function
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of all card input for the processing function.

7. GRAYMAP PROCESSOR

The chief purpose of GRAYMAP is to produce alphanumeric pictorial printouts of digitized MSS data. Each data sample is eight bits, providing 256 possible gray levels for each MSS data channel. To allow a meaningful distinction in gray-scale tones, GRAYMAP assigns each of the 256 levels to 1 of as many as 16 possible symbols. These symbols may be preassigned or arbitrarily assigned for each run. The specifications for the bin edges for each symbol may be assigned arbitrarily by the user for each run or computed from the histogram data in order to result in equal activity for each of the symbols. In any case, the data are subsequently output in terms of symbols, and each symbol represents a range of data values in which the corresponding data points fall.

7.1 INPUT FILES

An MSS data tape (DATAPE) must be input to the GRAYMAP processor. The tape assignment defaults to logical unit 11; however, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information on format.)

The GRAYMAP processor requires the bin levels to be input on a control card or computed from the histograms output by the HIST processor on the HISFIL file. When the bin levels are to be computed, logical unit 13 may be assigned to either disk or tape or allowed to default to disk (no assignment). If the HIST processor has not been executed prior to running GRAYMAP and bin levels have not been input, a default histogram of every 10th line for 500 lines and every 10th sample for 200 samples is computed, and HISFIL is created on logical unit 13.

Figure 7-1 shows the interaction of the HIST and GRAYMAP processors.

7.2 OUTPUT FILES

No files are output by the GRAYMAP processor.

7.3 SCRATCH FILES

The GRAYMAP processor does not require additional scratch files.

7.4 CARD INPUT

All system card input formats referred to in this section are defined in section 3.1.

7.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$GRAYMAP

This card directs the system monitor routine to select the GRAYMAP processor and initiates loading of routines used by GRAYMAP.

7.4.2 SPECIAL SYSTEM FILES

None of the special system files is required for this processor.

7.4.3 CONTROL CARDS

Table 7-1 lists all available options and control cards recognized by GRAYMAP, along with their default values.

7.4.4 FIELD DEFINITIONS

Fields for which gray-scale maps are desired must follow the *END control card. See section 3.1.3 for the format of field definition cards. Field definition input is terminated by the $END control card.

7.5 CARD OUTPUT

The GRAYMAP processor produces no card output.

7.6 RESTRICTIONS

The system-related restrictions in section 19 apply to this processor.

The maximum number of channels allowed is 30, and the maximum number of bin levels is 16.

7.7 DIAGNOSTIC MESSAGES

7.7.1 SUBROUTINE PICT

	Message	Explanation
a.	THE NO. OF CHANNELS FOR THIS FIELD HAS BEEN REDUCED TO XXX SO ALL THE INFORMATION WILL FIT ON DISK. MAKE ANOTHER RUN TO GRAYMAP OTHER CHANNELS.	Self explanatory.
b.	FIELD TOO LARGE, TERMINATING.	Data exceed allocated storage.
c.	YOU HAVE ASKED FOR TOO MANY SAMPLES. THE LAST SAMPLE IS ____________.	The last sample is reset to the last sample on the data tape.

7.7.2 SUBROUTINE SETUP6

	Message	Explanation
a.	BAD SUPERVISOR CONTROL CARD.	Check spelling of keyword.
b.	THIS CHANNEL IS OUT OF NUMERICAL RANGE AND WAS IGNORED.	All channels requested must be in the range 1 to 30.
c.	ONLY 16 BINLEVELS PERMITTED.	Reduce the number of bin levels to 16.
d.	ERROR ON DATAFILE CARD.	Check for format error and unit assignment.
e.	THIS CHANNEL IS NOT HISTOGRAMMED.	Check CHANNELS control card and make sure all channels requested have been histogrammed.

TABLE 7-1.— GRAYMAP PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNELS	$C_1, C_2, C_3, \cdots, C_k$ $k \leq 30$ Default: Gray map for all channels on HISFIL (created by a previous execution of HIST)	Provides pictorial printout for requested channels.
BINLEVEL	$N_1, N_2, N_3, \cdots, N_k$ $k \leq 16$ Default: Histograms used to set bin levels	Upper bin edges for gray-scale levels with a range of 0 to 255 and a maximum of 16 levels; the last bin level should always be 255.
SYMBOLS	$S_1, S_2, S_3, \cdots, S_k$ $k \leq 16$ Default: Two sets of 10 symbols overprinted, resulting in one of $,X,θ,0,*,=,•,-,/,b	Character set separated by commas, with a maximum of 16 symbols per SYMBOL card. If 2 sets are input, the second overprints the first. The number of symbols input on one card determines the number of bin levels when using the histograms to set the levels. Blank is a legitimate character.

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 7-1.– Continued.

Keyword	Parameter and default values	Function
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	N is the Fortran logical unit number to which the image data tape has been assigned; M is the file number on the tape to be processed. For back-to-back executions of more than one processor, if using the same file number, only one DATAFILE control card need be submitted.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Signals the end of the control cards.

TABLE 7-1.– Concluded.

Keyword	Parameter and default values	Function
$END	Blank	Signals the end of all card input for the processing function.

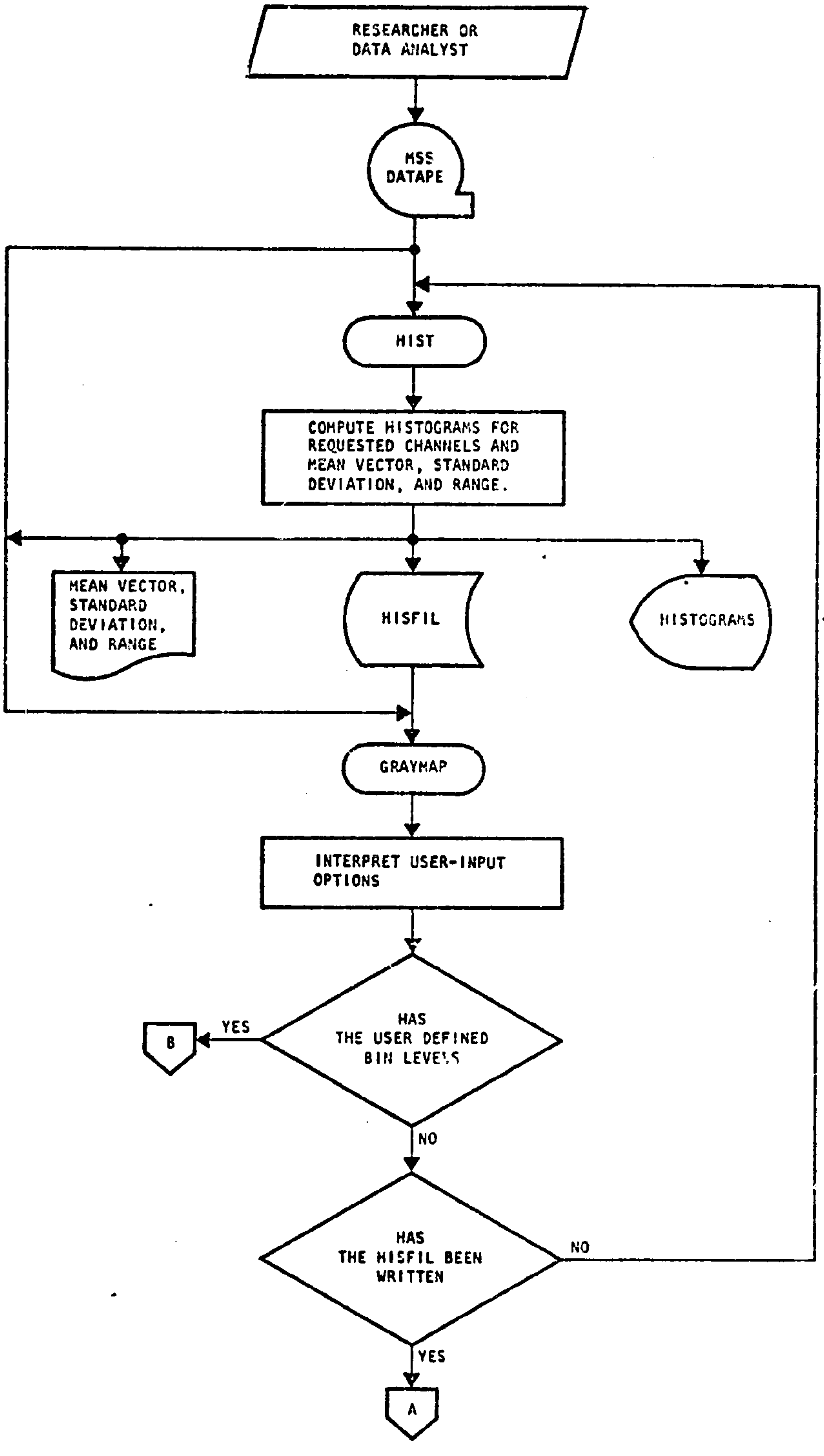


Figure 7-1.– Functional diagram showing interaction of the HIST and GRAYMAP processors.

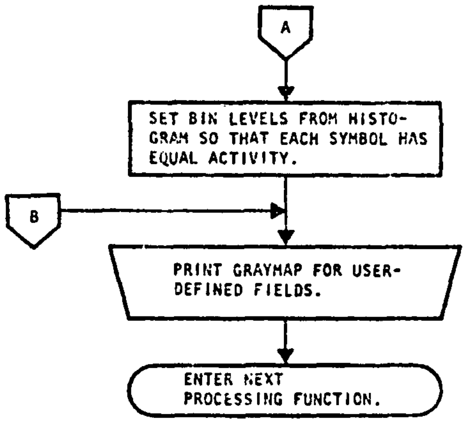


Figure 7-1.— Concluded.

8. STATISTICS PROCESSOR — STAT

The statistics processor STAT computes the multichannel means, standard deviations, covariance matrix, and correlation coefficient for each training field and all training subclasses which are defined through user input to the processor. In addition, at the user's option, histograms and spectral plots may be computed for each field and/or subclass.

The STAT processor requires user input of both card images and an MSS data tape (DATAPE). Card image input consists of an optional number of cards from the set of control cards defined in table 8-1 and the training field definitions described in section 3.1.3. The required input MSS data must encompass the area of interest specified in the training field definitions. The processor is activated and initialized by a specific processor card defined in section 8.4.1. All processor functions which are available as options to the user are directed by means of the input control cards or the built-in defaults for any control card which is not input.

In addition to the optional printouts under the direction of the control cards, the STAT processor creates the output file SAVTAP, which contains the computed statistics (mean vector and covariance matrix) for each training subclass. The training subclass statistics optionally may be output on punched cards (the module STAT file). Both the output statistics file SAVTAP and the output module STAT file are in a format acceptable to the statistics input requirements of other processors in the EOD-LARSYS. Figure 8-1 gives the functional flow of the STAT processor.

The mean vector for the *ith* subclass is computed as follows:

$$\mu_i = \overline{X}_{1i}, \overline{X}_{2i}, \cdots, \overline{X}_{pi}, \cdots, \overline{X}_{Pi} \qquad (8\text{-}1)$$

where

$\overline{X}_{pi} = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{pj}$ = average value in channel p for subclass i

p = channel number.

P = largest channel number.

N_i = number of samples in all training fields for subclass i.

X_{pj} = the *jth* sample of the MSS data for channel p (a value between 0 and 255).

μ_i = mean vector for the *ith* subclass.

The covariance matrix for the *ith* subclass is computed as follows:

$$K_1 = \begin{bmatrix} k_{11i} & k_{12i} & \cdots & k_{1Pi} \\ k_{21i} & k_{22i} & \cdots & \\ \vdots & & k_{pqi} & \vdots \\ & & \vdots & \\ k_{P1i} & & \cdots & k_{PPi} \end{bmatrix} \qquad (8\text{-}2)$$

where

$$k_{pqi} = \frac{1}{N_i - 1}\left[\sum_{1}^{N_i} x_p x_q - \frac{1}{N_i} \sum_{1}^{N_i} x_p \sum_{1}^{N_i} x_q\right]$$

q = channel number.

Closely related statistics are the standard deviation and correlation coefficient for the *ith* subclass, which are computed as follows:

$$\left.\begin{aligned} \sigma_{pi} &= (k_{ppi})^{1/2} \\ \rho_{pqi} &= \frac{k_{pqi}}{(k_{ppi}k_{qqi})^{1/2}} \end{aligned}\right\} \qquad (8\text{-}3)$$

where

k_{pqi} = element of the covariance matrix for subclass i; the variance between channels p and q.

σ_{pi} = standard deviation in channel p for subclass i; p = q.

ρ_{pqi} = correlation coefficient between channels p and q for subclass i.

8.1 INPUT FILES

An MSS data tape (DATAPE) must be input to the STAT processor. The tape assignment defaults to logical unit 11; however, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information on format.)

8.2 OUTPUT FILES

The STAT processor always outputs the statistics on the SAVTAP file and, optionally (by means of the OPTION PUNCH control card), provides the module STAT file on cards. (See section 3.1.4.1 for format of card file.)

The required output file SAVTAP will contain the class names, subclass names, and the subclass statistics (mean vector and covariance matrix) computed by the STAT processor for every subclass defined. The output statistics file must be assigned to either disk or tape. The tape assignment defaults to logical unit 20; however, by input of the STATFILE control card, the user may assign any available unit. (See table 3-1 for file assignments and table 8-1 for control card description.)

If the STATFILE control card is used, the statistics from more than one execution of STAT may be saved on the same tape.

8.3 SCRATCH FILES

The STAT processor does not require the use of a separate scratch file.

8.4 CARD INPUT

The specific card column formats for the information to be input on the processor and control cards are given in sections 3.1.1 and 3.1.2. Table 8-1 describes the complete set of keywords and option parameters recognized and acted upon by the STAT processor.

If possible, each keyword and its option parameters are to be completely contained on one control card. However, if more parameters are required than can be contained on one card, the control card may be repeated and the parameters continued on the next control card. The parameters for a control card of a given type will be cumulative over all cards of that type.

The control cards follow the $STAT processor card. All options available on the STAT processor have a default setting which is used by the processor for those option parameters not input via control card. The control card *END must be input to signify the end of the set of control cards. Immediately following the *END card, a set or sets of CLASSNAME, SUBCLASS, and training field definition cards must be input. See section 8.4.4 for further details on training field definitions.

8.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$STAT

This card directs the system monitor routine to select the STAT processor and initiates the input of STAT processor control card image file.

8.4.2 SPECIAL SYSTEM FILES

The processor does not expect input of any of the system-generated files described in section 3.1.4.

8.4.3 CONTROL CARDS

Table 8-1 gives the complete set of control cards which the user may input to direct the STAT processor functions and the default functions performed by the processor. With the exception of the *END and $END control cards, the sequence of the control cards is optional. The *END card must immediately follow the last control card, if any; the CLASSNAME, SUBCLASS, and training field definition cards must immediately follow the *END card; and the $END card must immediately follow the last card of the input training field definitions.

8.4.4 CLASS, SUBCLASS, AND FIELD DEFINITIONS

All CLASS, SUBCLASS, and field definition cards occur between the *END and $END control cards. The formats for these cards are given in section 3.1.3. Training fields are grouped into statistically similar subclasses, and subclasses are grouped further into classes.

A training class is defined to the processor by one card containing the keyword CLASSNAME in columns 1 through 9. The user-determined alphanumeric name to be assigned to the class begins in column 11 and may contain a maximum of four characters (through column 14). At least one CLASSNAME card must be input.

A CLASSNAME card must be followed by at least one subclass grouping. A subclass grouping is on a SUBCLASS card followed immediately by one or more field definition cards. All fields defined by field definition cards following the SUBCLASS card will contribute a cumulative sample set from which the training subclass statistics will be computed for the named subclass. The set of cards – one SUBCLASS card followed by one or more field definition cards – generates the statistics for one training subclass. The number of sets of SUBCLASS and field definition cards is determined by the number of sets of training subclass statistics. The number of training fields to be defined for one given subclass is not restricted. The following example shows the grouping of subclasses into classes.

$STAT

⋮ (Control cards)

*END

Classes	Subclasses		
		CLASSNAME	CLASS1
		SUBCLASS	SUB 11
	1	⋮ (Fields)	SUB 11 (of CLASS1)
1			
		SUBCLASS	SUB 12
	2	⋮ (Fields)	SUB 12 (of CLASS1)
		CLASSNAME	CLASS2
		SUBCLASS	SUB 21
2	3	⋮ (Fields)	SUB 21 (of CLASS2)
		CLASSNAME	CLASS3
		SUBCLASS	SUB 31
	4	⋮ (Fields)	SUB 13 (of CLASS3)
		SUBCLASS	SUB 32
3	5	⋮ (Fields)	SUB 32 (of CLASS3)
		SUBCLASS	SUB 33
	6	⋮ (Fields)	SUB 33 (of CLASS3)

$END

8.5 CARD OUTPUT

The STAT processor will optionally output on punched cards the module STAT file (section 3.1.4.1). The module STAT file contains the training field vertices, the subclass names for each training subclass, the subclass numbers assigned to each training subclass, the association of training fields to each subclass and class, and the computed statistics for each training subclass.

The module STAT file is output by the processor on cards only when specified by user input of the OPTION PUNCH control card.

8.6 RESTRICTIONS

The system-related restrictions in section 19 apply to the STAT processor.

In addition, a core storage limitation is associated with the total storage required by the training subclass statistics and the various options (i.e., producing histograms and spectral plots). The upper limit on core storage available for all requirements generated by the input to the STAT processor is 10 600 locations. Each subclass covariance matrix requires approximately 1/2(number of channels)^2 locations; each subclass mean vector requires locations equal to the number of channels; and each training field requires seven locations. If a large number of subclasses, channels, and training fields in combination with one or more of the options available by means of the OPTION control cards causes the core storage limits to be exceeded, the STAT processor prints a diagnostic message requesting the user to decrease options, after which it terminates execution (see section 8.7).

The following formula determines the maximum number of fields that can be input for a case without any histograms (eq. 8-4) and another case with subclass histograms (eq. 8-5).

$$\text{NOFLD} = \frac{10\ 600 - 5\text{NOSPEC} + 7\text{MAXSUB} + \frac{4 + 2\text{MAXSUB} + 5}{2}\text{NOFEAT} + 1\ \text{NOFEAT} + 40}{32} \tag{8-4}$$

where

NOFLD = number of fields

NOSPEC = number of subclasses grouped together for spectrogram (maximum of 20)

MAXSUB = maximum number of subclasses

NOFEAT = number of channels

$$\text{NOFLD} = \frac{10\ 600 - (5\text{NOSPEC} + 7\text{MAXSUB} + \left(\frac{4 + 2\text{MAXSUB} + 5}{2}\text{NOFEAT} + 1\right)\text{NOFEAT} + 40 + \text{XSIZ}}{32} \tag{8-5}$$

If fields and subclasses need to be histogrammed, a value for XSIZ(NOHIST) + 1 should be added to the numerator of equation (8-5), where

XSIZ = range of histogram (maximum of 101)

NOHIST = number of channels histogrammed

8.7 DIAGNOSTIC MESSAGES

The diagnostic messages provided by the STAT processor are listed, along with probable cause and remedy of the condition which prompted the message. During statistical computations, other messages also may be output by utility routines common to both STAT and other processors in the EOD-LARSYS. See the system-related messages in section 6 for additional messages obtained from a STAT execution.

8.7.1 SUBROUTINE SETUP1

	Message	Explanation
a.	///// FROM SUBR. SETUP1 --- BAD CONTROL CARD ENCOUNTERED --- INPUT CARD IS ________, 'CCCC ··· CCC'	The input card which was read has none of the legitimate keywords to identify it as a recognizable control card. The card which caused the message is printed out as part of the message. Although the processor will continue to read more control cards, this is an indication of an error in the deck setup. The deck should be checked for proper control cards and proper sequence of cards.
b.	*** STAT/SETUP1 -- ERROR IN OPTION(S) REQUESTED - SCAN OF OPTION(S) DISCONTINUED AT CARD COLUMN XX ***	An OPTION control card is not acceptable to the processor. The scan of the options will be discontinued by the processor, and any options specified beyond the erroneous one will not be activated for the run. The processor continues with reading of the next control card. (See section 3.1.2 and table 8-1 for correct OPTION control card usage.)
c.	ERROR ON DATAFILE CARD	Check format and unit assignment.
d.	ERROR ON STATFILE CARD	Check format and unit assignment.

	Message	Explanation
e.	***MAXSUB=XX --- MAX NO. OF SUBCLASSES CANNOT BE GREATER THAN YY MAXSUB SET=YY PROCEEDING TO NEXT OPTIONS(S)***	The maximum subclass number input on the OPTION MAXSUB control card exceeds the maximum number of subclasses that can be handled by the EOD-LARSYS. The processor will set the maximum number of subclasses, which will apply to subclasses read in from the input subclass/field definition deck.
f.	/////FROM SUB. SETUP1 --- DECREASE OPTIONS ***** TERMINATING PROGRAM EXECUTION FROM SUBR. SETUP1*****	The STAT processor has run out of internal storage to handle the combination of the quantities of input training fields, subclasses, and channels. Internal storage is fixed at 10 600 locations. Each subclass required roughly $1/2(\text{number of channels})^2$ locations for the subclass statistics. If histograms or spectral plots of subclasses and/or fields are requested, additional internal storage is required. The options specified in the run deck (i.e., histograms and spectral plots) and possibly the quantities of subclasses, channels, and training fields must be decreased or eliminated in

	Message	Explanation
		order to get a successful run within the core storage limitation.
g.	CHECK CHANNELS OR CLASS NOS. REQUESTED -- CANNOT BE LESS THAN OR EQUAL ZERO, OR GREATER THAN 30 ***** TERMINATING PROGRAM EXECUTION FROM SUBR. SETUP1 *****	If the channel numbers specified on a HISTO or CHANNELS control card are not integers within the range 1 through 30, this message results. The processor terminates execution after printing this message. Check the format of the applicable processor control cards (see section 3.1.2 and table 8-1).
h.	***** STAT/LEARN--MAX. OF XX SUBCLASSES EXCEEDED -- FIRST XX SUBCLASSES USED -- REMAINDER IGNORED	The processor has read the maximum allowable number of subclass names and training fields to be associated with each subclass, and the next subclass name encountered in the training field/subclass definition deck caused this diagnostic message. The first MAXSUB subclasses and associated training fields input are computed and the remainder are ignored by the processor.

Message	Explanation
i. ***** STAT/LEARN -- MAX. OF XX SUBCLASSES EXCEEDED -- FIRST XX SUBCLASSES USED -- REMAINDER IGNORED	The STAT processor has read the maximum number of subclass names and associated training fields from the input training field/subclass definition deck. The available internal storage has been filled, and no further training fields can be accepted. Training statistics will be computed for the subclasses and fields which have been read to this point, and the remainder are ignored by the processor.

TABLE 8-1.- STAT PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNELS	$N_1,N_2,N_3,\cdots,N_k$ $1<k\leq 30$ Default: k=30; unless the MSS data tape (DATAPE) has exactly 30 channels, the default should not be taken.	N's are the integer channel numbers used by the processor in computing training subclass and training field statistics; must be from the set of channels available on the MSS DATAPE file.
OPTION	PUNCH Default: The module STAT card file is not punched, in which case statistics are output on the SAVTAP file only.	The subclass mean vector and covariance matrix for every subclass defined by user input will be punched on cards in a format acceptable as input to other processors in the system. This punched card deck is the module STAT file defined in section 3.1.4.1.
OPTION	MAXSUB=N Default: MAXSUB=15	Informs the processor as to the maximum number of subclasses which will be input. The parameter value is used for dimensioning purposes and reflects the maximum number of available computer storage locations being utilized for other options allowed by the STAT processor. This parameter must be set by

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 8-1.- Continued.

Keyword	Parameter and default values	Function
		the user if the number of subclasses he is about to define will exceed the default. It is advisable to use this option when a large number of training fields are to be processed or when histograms have been requested.
OPTION	COVAR Default: Statistics are not printed.	The multichannel means, standard deviations, and covariance matrix (lower triangular portion) are printed out for each training subclass and training field defined in the input training field definition deck.
OPTION	COVAR=C Default: Statistics are not printed.	The multichannel means, standard deviations, and covariance matrix (lower triangular portion) are printed out for each training subclass defined in the input training field definition deck.
OPTION	COVAR=F Default: Statistics are not printed.	The multichannel means, standard deviations, and covariance matrix (lower triangular portion) are printed out for each training field defined in the input training field definition deck.

TABLE 8-1.— Continued.

Keyword	Parameter and default values	Function
OPTION	NOCOVAR	No training subclass or training field statistics are printed out.
HISTO	$N_1, N_2, N_3, \cdots, N_k$ $1<k\leq 30$ Default: k=30	N's are integers which provide a list of channel numbers for use in the histogram options. The channel numbers must be from the set designated on the CHANNELS control card. Note: This control card does not initiate the histogram option.
OPTION	HIST Default: No histograms	A histogram showing frequency distribution of pixels (resolution elements or radiance values) is printed out for every training field and every training subclass defined in the input training field definition deck. For each subclass (or field), a histogram is provided for every channel designated on the HISTO control card.
OPTION	HIST=C Default: No histograms	A histogram printout is provided for every training subclass defined in the input training field definition card. For each subclass, a histogram is provided for every channel

TABLE 8-1.— Continued.

Keyword	Parameter and default values	Function
		designated on the HISTO control card.
OPTION	HIST=F Default: No histograms	A histogram printout is provided for every training field defined in the input training field definition deck.
SIZE	XHIGH=K $0<K\leq 255$ Default: XHIGH=220	K is an integer which sets the maximum radiance value which will be histogrammed. XHIGH becomes X_{max} of the X-axis of the histogram plot.
SIZE	XLOW=L $0\leq L<XHIGH$ Default: XLOW=120	L is an integer which sets the minimum radiance value which will be histogrammed. XLOW becomes X_{min} of the X-axis of the histogram plot.
SIZE	YSIZ=J $0<J\leq f(x)_{max}$ Default: YSIZ=14	J is an integer which sets the number of increments on the Y-axis of the histogram plot; therefore, it is the height (number of print lines) of the Y-axis. Using the input YSIZ, the processor will determine the Y-axis scale for the histogram plot to be $f(x)_{max}+(YSIZ-1)/YSIZ$.
SIZE	XSIZ=K Default: XHIGH-XLOW	Sets the range which will be histogrammed; maximum range is 101.

TABLE 8-1.- Continued.

Keyword	Parameter and default values	Function
SPECTRAL	M_1, M_2, M_3, M_4 $1 \leq M_i \leq 30$ Default: 4 subclasses per spectral plot; subclasses 1, 2, 3, and 4 on the first plot; 5, 6, 7, and 8 on the second plot; etc.	M's are integers which provide a list of from one to four subclass numbers for the subclasses which are to be plotted on one single composite spectral plot. The subclass numbers must be obtained from the set of subclasses defined in the input training field definition deck. Subclass 1 is the first subclass defined in the deck, and subsequent subclass numbers are obtained by sequentially numbering the subclasses as they occur in the training field definition deck.
OPTION	SPECTRAL Default: Spectral plots for subclasses	A spectral plot is printed out for every training subclass and training field defined in the input training field definition deck. The plot consists of the subclass (or field) mean radiance value, mean standard deviation (σ), and mean $-\sigma$ plotted versus the channel (spectral band) for every channel designated on the CHANNELS control card.

TABLE 8-1.– Continued.

Keyword	Parameter and default values	Function
OPTION	SPECTRAL=C Default: Spectral plots for subclasses	A spectral plot will be printed out for every subclass defined in the input training field definition deck.
OPTION	SPECTRAL=F Default: Spectral plots for subclasses	A spectral plot will be printed out for every field defined in the input training field definition deck.
SIZE	SPECBAS=I $0 \leq I \leq 105$ Default: SPECBAS=75	I is an integer which sets the minimum radiance value on the Y-axis of the spectral plot (i.e., Y_{min}). The processor has a fixed Y-axis increment (3) and a fixed number of Y-axis values (50). Using SPECBAS, the processor determines the Y-axis range to be: Y_{min}=SPECBAS, Y_{max}=SPECBAS+150.
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	N is the Fortran logical unit number to which the MSS data tape (DATAPE) has been assigned; M is the file number on the tape to be processed. For back-to-back executions of several processors if the same file number is used, only one DATAFILE control card need be input.

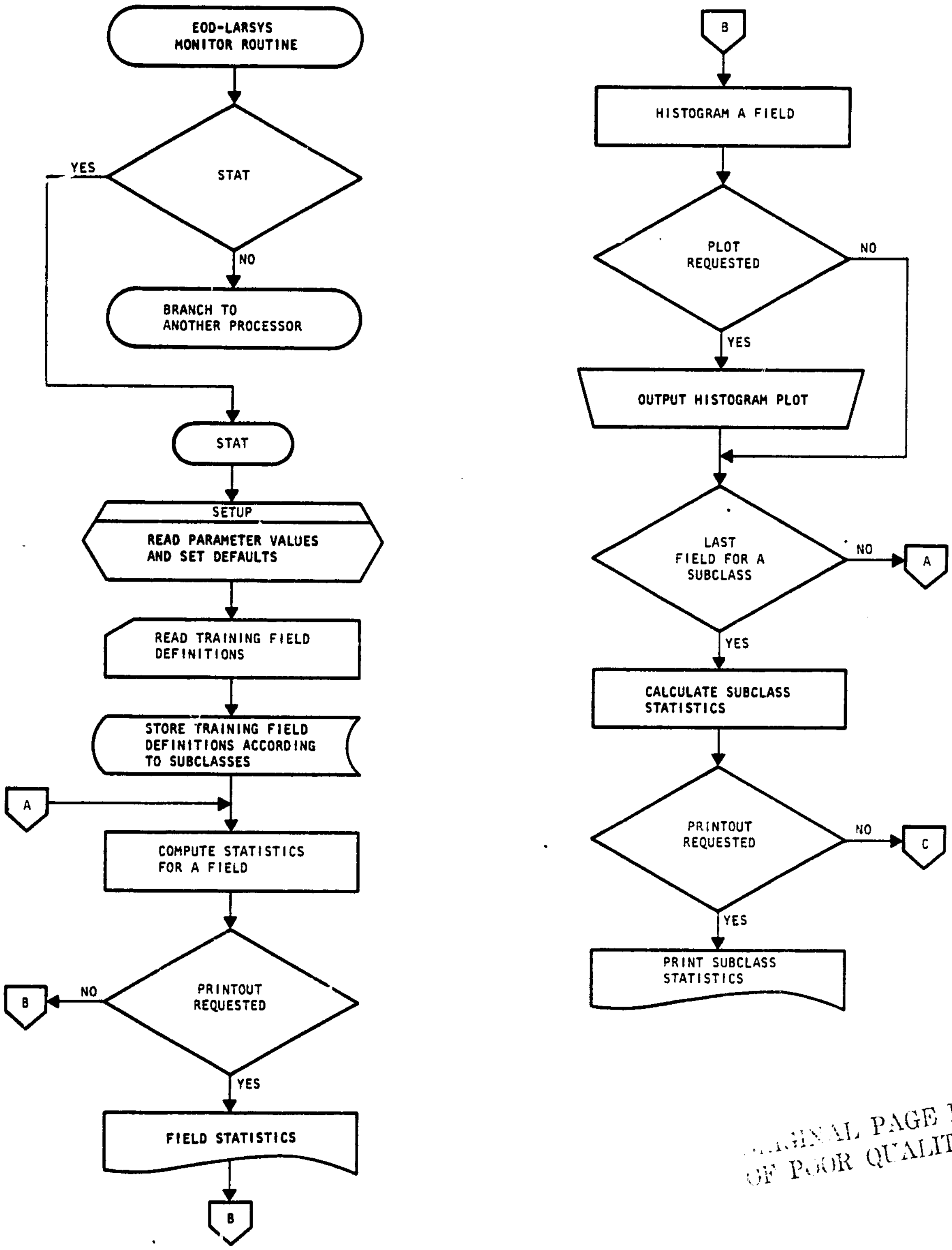


Figure 8-1.— Functional flow chart for the STAT processor.

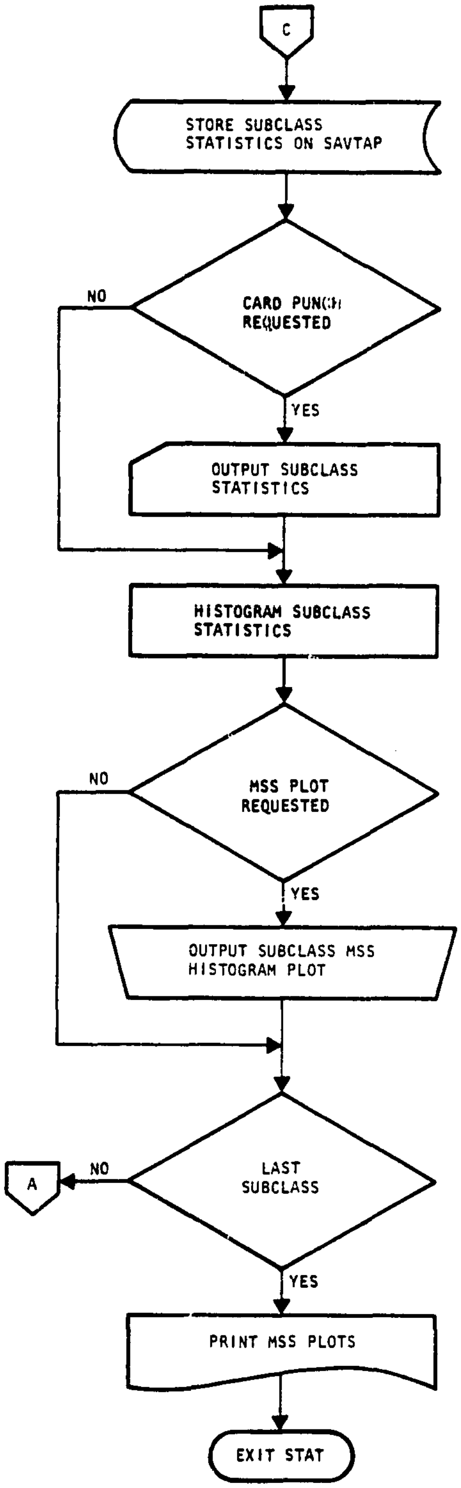


Figure 8-1.— Concluded.

9. ITERATIVE SELF-ORGANIZING CLUSTERING SYSTEM PROCESSOR — ISOCLS

A data set to be clustered by ISOCLS is defined by a class consisting of one or more fields from the MSS data tape (DATAPE). Any number of classes may be defined and clustered as individual data sets with one entry into the ISOCLS processor. The user has control over the maximum number of clusters allowed per class via the CLUSTERS control card. However, the procedure may find fewer clusters than the maximum allowed. If the user plans to use the statistics generated from the clusters in later CLASSIFY or SELECT runs, he or she must exercise control over the maximum number of clusters. The SAVTAP file may contain statistics for up to 75 clusters (or subclasses), but only 60 can be used for processing at any one time in CLASSIFY or SELECT. (Control cards are set out in table 9-1.)

The clustering procedure used in ISOCLS (ref. 3) is an iterative procedure which assigns each MSS data sample to a specific cluster by determining the nearest (in absolute distance) cluster center and assigning the sample to it. At the end of each iteration (i.e., when all samples have been assigned to a cluster), new cluster centers are defined by computing the mean vector for the data samples actually assigned to the cluster. After the initial split sequence, the iterative procedure terminates when the user-specified sequence of splits and combinations is exhausted. See the SEQUEN control card. The criteria for splitting or combining a cluster are user-specified by the STDMAX and DLMIN control cards.

After the final iteration, the covariance matrix for each cluster is computed and, at the user's option, is printed. All cluster statistics for the class are saved on a scratch file until all classes have been clustered, at which time the SAVTAP file is written. The chaining of clusters for the final map printout is performed, if the user has requested the option (see CHAIN control card, table 9-1). Statistics for the chained clusters are not computed.

The processor allows the user to control the amount of line-printer output he receives via the KRN and MAP control cards. A final map of the clustered data is always output along with a statistical summary of the clusters, which includes mean and standard deviation vectors, total points assigned to each cluster, and intercluster distances.

Optionally the user may (1) input initial cluster centers to hasten the clustering process or (2) allow the program to initialize the process by assigning all the data to one cluster, obtaining the mean and standard deviation, and then splitting. Initial means may be input (1) by cards (see control card MEANS and Cluster Means Deck, section 3.1.4.3) or (2) by the SAVTAP file (see control card STATFILE). Input of the initial means causes a scratch file to be written so that the means can be used repeatedly. Successive classes may or may not use the same means to initialize cluster centers for a new class. The control card MEANS allows the user to request cluster centers from the last class to be read from the scratch file and used as initial centers for a new class. Input of a new set of initial means will cause the scratch file to be overwritten with new cluster centers.

Several additions have been made in support of Procedure 1 requirements. These additions are described in a general sense as follows.

a. Optionally, starting dots (pixels) from the dot data file on the DOTUNT unit may be used to begin clustering.

b. The analyst may identify DO and DU pixel sets (fields) by field card input. The pixels in these fields are not included as inputs to the clustering algorithm. They are assigned special cluster numbers and mean vectors for display purposes.

c. Using a Sun angle correction table, the pixel radiance values may be modified. (The correction table is built in.) The radiance value correction applies only for clustering purposes.

The user may input the Sun angles by cards or request that these angles be extracted from the header record of a Universal-formatted MSS data tape (ERIPS unload tape).

The clustering algorithm embodied in ISOCLS is detailed step by step in the following subsections. This entire procedure is repeated for each class (or data set).

See the functional flow chart for ISOCLS (fig. 9-1).

9.1 PROCEDURES

9.1.1 NOTATION DEFINITIONS

Symbol	Fortran name	Definition
CLD_{ij}	CLD(I,J)	Intercluster distance between clusters i and j.
$d[X_k,\mu^{(i)}]$	DIST	Distance from the data point k to the center of cluster i.
DLMIN	DLMIN	Threshold value for combining clusters.

TABLE 8-1.– Concluded.

Keyword	Parameter and default values	Function
STATFILE	UNIT=N,FILE=M Default: N=20,M=1	N is the logical Fortran unit number to which the SAVTAP file has been assigned; M is the file number on the tape to be processed.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of all card input for the processing function.

Symbol	Fortran name	Definition
ISTOP	ISTOP	Maximum number of initial split iterations.
CHNTHS	CHNTHS	Chaining threshold value.
LNCAT	LNCAT INCAT	Number of existing clusters at a given time.
N(i)	N(I) DN(I)	Total number of data points assigned to cluster i.
SEQUEN	SEQUEN	User-specified sequence of split and combine iterations.
NMIN	NMIN	Minimum number of data points allowed per cluster for both the initial split iterations and for one through (NOSEQ-1) SEQUEN iterations.
PMIN	PMIN	Minimum of (PMIN+NOFEAT) number of data points allowed per cluster for the NOSEQ*th* SEQUEN iteration.
NOSEQ	NOSEQ	Maximum number of SEQUEN iterations.
ISEQ	ISEQ	Number of SEQUEN iterations at a given time.
NOFEAT	NOFEAT	Number of coordinates (channels) in a data vector.
STDMAX	STDMAX	Threshold for splitting clusters.
X_k	C(I,K)	Data vector k, $C(I,K)=(X_1,X_2,\cdots,X_{NOFEAT})_k$
$\mu_j^{(i)}$	MEANS(J,I)	Mean of the j*th* coordinate of the i*th* cluster.

Symbol	Fortran name	Definition
$\gamma_j^{(i)}$	AVP(J,I)	Temporary summing variable for the calculation of the standard deviation of the *jth* coordinate of the *ith* cluster.
$M_j^{(i)}$	AMN(J,I)	Summing variable for computation of new means. After all data have been assigned to clusters on each iteration, AMN(J,I) is the new mean of the *jth* coordinate of the *ith* cluster.
$\sigma_j^{(i)}$	STDEV(J,I)	Standard deviation of the *jth* coordinate of the *ith* cluster.

9.1.2 INITIALIZING THRESHOLD VALUES

Initialize threshold values for splitting clusters (STDMAX), combining clusters (DLMIN), and deleting clusters (NMIN and PMIN). Then begin the following iterative procedure.

9.1.3 ITERATIVE PROCEDURE

9.1.3.1 Classify and Calculate New Statistics

Assign each data point to a cluster and at the same time collect the means, standard deviations, and point counts of the newly developing clusters. If zero clusters, set i = 1 and go to iteration b. If more than zero clusters, go to iteration a.

a. Assign the data point $X_k = (X_1, X_2, \cdots, X_{NOFEAT})_k$ to the *ith* cluster if $d\left[X_k, \mu^{(i)}\right] \leq d\left[X_k, \mu^{(j)}\right]$ for all $j \neq i$, where $d\left[X_k, \mu^{(i)}\right]$ is defined as

$$d[X_k, \mu^{(i)}] = \sum_{j=1}^{NOFEAT} \left| X_{jk} - \mu_j^{(i)} \right| \qquad (9\text{-}1)$$

b. $$N(i) = N(i) + 1 \quad (9\text{-}2)$$

c. $$M_j^{(i)} = \frac{N(i) - 1}{N(i)} M_j^{(i)} + \frac{1}{N(i)} X_{jk} \quad (9\text{-}3)$$

d. $$\gamma_j^{(i)} = \frac{N(i) - 1}{N(i)} \gamma_j^{(i)} + \frac{1}{N(i)} X_{jk}^2 \quad (9\text{-}4)$$

e. $$\sigma_j^{(i)} = \left\{ \gamma_j^{(i)} - \left[\mu_j^{(i)} \right]^2 \right\}^{1/2} \quad ; \quad j = 1, \text{NOFEAT} \quad (9\text{-}5)$$

Return to step a and repeat iterations a through e until all data points have been classified.

9.1.3.2 Delete Clusters

For the initial split iterations and one through (NOSEQ-1) user-specified SEQUEN iterations, delete all clusters which have fewer than NMIN members. For the NOSEQ*th* user-specified iteration (last user-input sequence), delete all clusters which have fewer than PMIN members. A cluster is deleted simply by removing the statistics for that cluster and reducing the number of clusters (specified by LNCAT) accordingly.

9.1.3.3 Test for Completion

If this is not the last iteration, proceed to 9.1.3.4. If this is the last iteration and no clusters were deleted, the procedure is finished. If one or more clusters were deleted for having less than PMIN members, go back to 9.1.3.1 and reassign the data to the clusters obtained from iteration (NOSEQ-1).

9.1.3.4 Determine Type of Iteration

Determine whether this is to be a split iteration or a combine iteration and proceed to the appropriate step.

The sequence of iterations will be as follows:

$$\underbrace{\text{SSSS}}_{\substack{\text{ISTOP}\\ \text{and/or}\\ \text{PERCENT}}} \quad \underbrace{\text{CCSCSC}}_{\text{SEQUEN}}$$

where

S = split iteration

C = combine iteration

The beginning sequence of split iterations is terminated either (1) when the standard deviations for the user-input percentage of clusters (see PERCENT control card, table 9-1) are less than the STDMAX threshold parameter or (2) when ISTOP iterations have been reached. At that point, the type of iteration (split or combine) and number of iterations (NOSEQ) are determined by the SEQUEN parameter.

The initial split iterations are for the automatic initialization of cluster centers in the event they are not input. The sequence is shortened considerably if initial cluster centers are input.

9.1.3.5 Split Clusters

A cluster is split along the *jth* coordinate (1) if the *jth* coordinate has the maximum standard deviation for the cluster, (2) if the standard deviation along the *jth* coordinate is greater than the STDMAX threshold parameter, and (3) if the cluster has more than 2(NMIN+1) data points.

If conditions (1) through (3) are met, two new clusters are created and the parent cluster is deleted. A cluster is created merely by defining its centers (means) for each coordinate. If the *ith* cluster is split in the *jth* coordinate, the two new clusters will have centers at $\left[\mu_1^{(i)}, \mu_2^{(i)}, \cdots, \mu_j^{(i)} \pm \alpha, \cdots, \mu_{NOFEAT}^{(i)}\right]$, where α will normally be $\sigma_j^{(i)}$ but can be a user-input constant (see SEP control card). On a given split iteration, if the maximum number of clusters (CLUSTER) has not been reached, all clusters having a standard deviation greater than the STDMAX parameter will be split. To ensure that the clusters with the largest standard deviations receive the highest priority for splitting, when 2 × LNCAT > CLUSTER, the standard deviations are ordered along the *jth* coordinate in descending order. Return to 9.1.3.1 after splitting clusters.

9.1.3.6 <u>Combine Clusters</u>

Two clusters are combined if the distance between them is less than the DLMIN threshold parameter. The distance between clusters i and j is calculated as

$$CLD_{ij} = \left(\sum_{k=1}^{NOFEAT} \frac{\mu_k^{(i)} - \mu_k^{(j)}}{\alpha_k^{(i)} \alpha_k^{(j)}} \right)^{1/2} \qquad (9\text{-}6)$$

If CLD_{ij} < DLMIN and CLD_{ij} = MIN(CLD_{ij}) for all i = 1,LNCAT and j = 1,LNCAT for all i ≠ j, clusters i and j will be merged to form a new cluster L with means

$$\mu_k^{(L)} = \frac{N(i)\mu_k^{(i)} + N(j)\mu_k^{(j)}}{N(i) + N(j)} \quad ; \quad k = 1,NOFEAT \qquad (9\text{-}7)$$

The clusters i and j are deleted. The new cluster L is not considered as a candidate for merging with any other cluster on the iteration in which it was formed. Return to 9.1.3.1 after combining clusters.

9.1.4 CHAINING

A final optional step in the clustering procedure groups all clusters which have intercluster distances less than the chaining threshold (CHNTHS) to form one cluster. The chaining procedure was adopted because the minimum variance criterion used in the iterative procedure above tends to group the data into spherical (or ellipsoidal) groupings with Gaussian distributions. This type of grouping is certainly a natural grouping and would quite often be completely satisfactory.

Some natural groupings of the data are odd shaped and cannot be approximated by Gaussian distributions. Two examples are given in figure 9-2. At the end of the sequence of split and combine iterations, groupings of the type in figure 9-2 are likely to be separated into subclusters as illustrated in figure 9-3. The chaining algorithm will indicate that the subclusters 1, 2, and 3 (fig. 9-3) could be grouped into one composite cluster; likewise, subclusters 4, 5, 6, and 7 could be grouped together to form one cluster.

The algorithm scans the intercluster distance table (CLD) and begins a chain with the first appearance of two clusters within a distance of CHNTHS units. Once a subcluster is in the chain, all clusters which are within CHNTHS units of the subcluster are added to the chain. See figure 9-4.

The statistics (means, standard deviations, and covariance matrices) of the clusters resulting from chaining are not calculated by the program because, in many cases, the chained cluster cannot be represented by a Gaussian distribution.

There are, of course, instances where one can safely combine those subclusters that are chained by the program into one composite (Gaussian) cluster. For example, subclusters 1, 2, and 3 in figure 9-5 can safely be combined into one final cluster. This is indicated by the fact that, pairwise, these three subclusters are all close to one another. In this case, the following formulas (ref. 3) can be used iteratively to compute the composite statistics.

Assuming that two clusters (n_1, m_1, C_1) and (n_2, m_2, C_2) are to be considered as one cluster (n, m, C), where all n, m, and C are the number of points, the mean vectors, and the covariance matrices, respectively, and m^T is the transpose of m then

$$\left.\begin{aligned} n &= n_1 + n_2 \\ m &= \left(\frac{n_1}{n_1 + n_2}\right)m_1 + \left(\frac{n_2}{n_1 + n_2}\right)m_2 \\ C &= \left(\frac{n_1}{n_1 + n_2}\right)C_1 + \left(\frac{n_2}{n_1 + n_2}\right)C_2 + \left(\frac{n_1}{n_1 + n_2}\right)m_1 m_1^T \\ &\quad + \left(\frac{n_2}{n_1 + n_2}\right)m_2 m_2^T - mm^T \end{aligned}\right\} \qquad (9\text{-}8)$$

9.2 INPUT FILES

An MSS data tape (DATAPE) must be input to the ISOCLS processor. The tape assignment defaults to Fortran unit 11; however, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information on format.)

In support of Procedure 1, the following inputs are also required: a statistics file (SAVTAP) to provide starting cluster mean vectors; initial cluster "centers" on cards; and the starting dots from the DOTUNT file to initialize the cluster processing.

Format descriptions of these files are included in sections 4.1 and 3.1.4.3.

9.3 OUTPUT FILES

Statistics are output by ISOCLS to the SAVTAP file (section 4.1). The file assignment defaults to logical unit 20; but, by input of the STATFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and STATFILE control card, table 9-1, for further information.)

A cluster map tape (MAPUNT) may be generated optionally for displaying the results of the clustering on the Bendix 100 or the PMIS DAS. The FORMAT control card initiates the option and names the desired format of the tape. Logical unit 16 should be assigned to a nine-track tape drive when this option is exercised (see section 5.1).

A printout of the cluster results consists of the following data items by class: cluster numbers and symbols; cluster mean vectors (by channel); cluster standard deviations by channel; intercluster distances; number of pixels per cluster; number of clusters; and cluster map by field for each class.

9.4 SCRATCH FILES

The program dynamically assigns random access disk storage for scratch files. ISOCLS uses the disk storage for temporary storage of cluster statistics, the data to be clustered, and the classification of each pixel.

9.5 CARD INPUT

9.5.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$ISOCLS

This card directs the monitor routine to call the ISOCLS processor and causes all routines used by the ISOCLS processor to be loaded into the system.

9.5.2 SYSTEM CARD DECKS

The processor will read a cluster MEANS deck in the format defined in section 3.1.4.3. The deck may be used to initialize cluster centers for the clustering procedure.

9.5.3 CONTROL CARDS

Control cards allow the user to input various options. They are identified by a keyword that is left justified in columns 1 through 10 of the card, with parameter values or additional keywords in columns 11 through 72. These control cards may be in any order, but they must be the first cards after the processor card $ISOCLS. Table 9-1 lists all available options, along with their parameter values.

9.5.4 CLASS AND FIELD DEFINITIONS

A CLASSNAME card, followed by at least one field definition card, must immediately follow the *END control card. The formats for these cards are defined in section 3.1.3.

The pixels from all fields for one class are extracted from the MSS data tape (DATAPE) and stored on a disk file. The data from all fields for one class are clustered as one data set. The statistics for all clusters in that class are saved on a scratch file, and the next class is clustered. When all classes have been clustered, the statistics are written on the SAVTAP file. The SETUP routine may be entered after each class to change parameter values. The input for definition of classes and fields is explained as follows.

```
Write the SAVTAP file      }      $ISOCLS
after five classes have    } ---> CLASSES    5
been clustered.            }
                                 {  :  (Other control cards)
                                 {
                                  *END
WHEAT 1 clustered as             { CLASSNAME            WHT1
one data set.                    { FIELD1                 |
                              1  { FIELD2                 |
                                 { FIELD3                 v
                                 { CLASSNAME            WHT2
                              2  { FIELD4                 |
Enter SETUP routine        }     { FIELD5                 v
again for new parameter    } ---> $END
values.                    }
                                 {  :  (New parameter values)
                                 {
                                  *END
                                 { CLASSNAME            NWH1
                              3  { FIELD6                 |
                                 { FIELD7                 |
                                 { FIELD8                 v
                              4  { CLASSNAME            NWH2
                                 { FIELD9                 v
                              5  { CLASSNAME            NWH3
Now write statistics       }     { FIELD10                v
file.                      } ---> $END
                                  $(Next processor)
```

Note that actual names may not exceed four characters. ISOCLS recognizes DO/DU fields. All the DO/DU field cards (for all classes) must be input before the fields to be clustered. These fields must immediately follow the *END card. The CLASSNAME card follows the last DO/DU field card.

Example:

- If DO/DU fields are being defined:

```
*END
DESIGNATED    OTHER
OTHER         (1,1),(1,1),(40,1),(40,20),(1,20)
DESIGNATED    UNIDENTIFIABLE
UNIDEN        (1,1),(5,7),(8,7),(8,10),(5,10)
CLASSNAME     WHT
```

```
WHT1            (1,1),(1,1),(196,1),(196,117),(1,117)
$END
```

- If no DO/DU fields are being defined:

```
*END
CLASSNAME       WHT
WH1             (1,1),(1,1),(196,1),(196,117),(1,117)
$END
```

9.6 CARD OUTPUT

A module STAT file (see section 3.1.4.1) may be punched and used as an interface between ISOCLS and SELECT or CLASSIFY. This option is exercised via the OPTION PUNCH control card.

9.7 RESTRICTIONS

The ISOCLS processor uses disk for a temporary scratch file. There are approximately 750 000 words of storage available on disk. The data to be clustered for one class are stored on this file, along with other information. To compute the maximum number of pixels per class, use the following formula.

$$\text{Maximum pixels} = \frac{750\ 000 - 30\left\{\begin{matrix}\text{number of}\\ \text{classes}\end{matrix}\left[\left(\begin{matrix}\text{number of}\\ \text{channels}\end{matrix}\right)^2 + 3\left(\begin{matrix}\text{number of}\\ \text{channels}\end{matrix} + 2\right)\right] - 1800\right\}}{\text{number of channels} + 1} \tag{9-9}$$

The maximum number of clusters per class is 60, and the maximum number of channels is 30. The covariance matrices for all clusters in one class must be stored in core at one time. They are stored in an array dimensioned 11 500. The following formula may be used to see if enough storage is available for the covariances.

$$11\ 500 \geq \begin{matrix}\text{number of}\\ \text{clusters}\end{matrix}\left[\begin{matrix}\text{number of}\\ \text{channels}\end{matrix}\left(\frac{\begin{matrix}\text{number of}\\ \text{channels}\end{matrix} + 1}{2}\right)\right] \tag{9-10}$$

9.8 DIAGNOSTIC MESSAGES

9.8.1 SUBROUTINE DSTAPE

Message	Explanation
THE NUMBER OF CHANNELS TIMES THE NUMBER OF SAMPLES HAS EXCEEDED 11500. DECREASE THE NUMBER OF CHANNELS OR THE NUMBER OF SAMPLES. TERMINATING RUN FROM DSTAPE.	Storage available has been exceeded.

9.8.2 SUBROUTINE ISOCLS

Message	Explanation
a. NO. CHANNELS FOR STARTING NOT EQUAL THAT FOR CLUSTER	The number of channels of starting vectors from the STAT file must equal the number of requested data channels.
b. DIMENSION LIMITS EXCEEDED IN ISOCLS BY ____. REDUCE CHANNELS OR MAX. CLUSTERS.	The user has exceeded storage. The number of channels or maximum clusters per class should be reduced.
c. DIMENSION LIMIT OF ____ FOR COVARIANCES EXCEEDED.	

9.8.3 SUBROUTINE PSPLIT

Message	Explanation
ERROR READING DISK-ISTAT=XXXX	

9.8.4 SUBROUTINE RANK

Message	Explanation
THE NUMBER OF CHANNELS ARE NOT A MULTIPLE OF 4. THE COLOR KEYS WILL BE ORDERED BY CLUSTER NUMBER.	Currently used greenness/ brightness transformations require four channels per pass.

9.8.5 SUBROUTINE RDDATA

Message	Explanation
a. TOO MANY DO OR DU FIELDS. THESE IGNORED.	There can be up to 10 DO fields and 10 DU fields.
b. TOO MUCH DATA REQUESTED -- PIXELS*(CHANNELS +1) CANNOT EXCEED XXXXXXXXXX.	Self-explanatory.
c. STORAGE REQUIRED FOR FIELD DEFINITION INFORMATION EXCEEDS THE DIMENSION LIMIT OF ______.	Reduce the number of fields. All vertices, names, and rectangular coordinates are saved for each field. The user has exceeded storage.
d. END-OF-TAPE REACHED BEFORE END-OF-FIELD.	A field has been defined beyond the limits of the MSS DATAPE.
e. INPUT ERROR — A CLASSNAME CARD MUST BE INPUT BEFORE A GROUP OF FIELDS.	See section 9.5.4 on defining classes and fields.
f. NO. OF PIXELS TO BE UNPACKED PER SCAN EXCEEDS THE DIMENSION LIMIT OF ______.	Decrease the number of channels or pixels per scan in the field.
g. TOO MUCH DATA REQUESTED -- PIXELS *(CHANNELS + 1) CANNOT EXCEED ______.	Disk file will not hold all of the data for one class. Reduce channels or size of fields.

9.8.6 SUBROUTINE RDMEAN

Message	Explanation
a. MEANS FOR CHANNEL XXXX ARE NOT ON FILE--DUMMY VALUES WILL BE USED.	Self-explanatory.

9.8.7 SUBROUTINE SETUP7:

Message	Explanation
a. ERROR ON CHANNEL CARD.	Check format.
b. ERROR ON DATAFILE CARD.	Check format.
c. ERROR ON STATFILE CARD.	Check format.
d. ERROR ON DOTFIL CARD.	Check format.
e. INVALID INPUT CARD _____ IGNORED.	Check table 9-1 for correct spelling of keywords for card input and make sure the keyword is left justified in the field.
f. CHANNELS CANNOT BE CHANGED UNTIL THIS EXECUTION OF ISOCLS IS COMPLETED.	The channels to be used should be set in the first set of control cards input after the ISOCLS card. That set of channels will be used for all classes. If the user attempts to input a CHANNELS card into the SETUP routine on a later entry, the card will be ignored.
g. NO. OF CLASSES CANNOT BE CHANGED UNTIL THIS EXECUTION OF ISOCLS IS COMPLETED.	The number of classes to be clustered must be input only in the first set of control cards input after the ISOCLS

	card. If the user attempts to change this parameter, the input will be ignored.
h. **WARNING** NMIN IS LESS THAN NO. OF CHANNELS, COVARIANCES? WILL NOT BE INVERTIBLE.	NMIN should be increased to greater than total number of channels.

TABLE 9-1.— ISOCLS PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNELS	DATA=$C_1, C_2, C_3, \cdots, C_k$, STAT=$A_1, A_2, A_3, \cdots, A_k$ $k \leq$ number of channels on SAVTAP≤ 30 Default: None	C's are integer channel numbers that (1) will be used in clustering and (2) refer to the MSS data tape (DATAPE). A's are integer channel numbers that (1) will be the starting vectors (initial means), (2) refer to the SAVTAP file, and (3) must be a subset of the channels on the SAVTAP file. The same channels must be used throughout one execution of ISOCLS. If a cluster MEANS card file is input, the channels on this card must be a subset of the channels in the MEANS card file.
OPTION	ORDER Default: The color keys will be ordered according to cluster numbers.	The color keys on the MAPUNT tape will be ordered according to greenness. See section 5.1 for further details of color keys.
OPTION	PUNCH=N Default: If PUNCH is omitted, no cards are punched; if N is omitted, it defaults to 1.	Punches the means and covariance matrix for each cluster in the module STAT card file format defined in section 3.1.4.1. N=1 punches module STAT card file; N=2 punches

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 9-1.– Continued.

Keyword	Parameter and default values	Function
		ERIPS interface card file; and N=3 punches both card files.
OPTION	STATS	Prints the covariance matrix for each cluster.
SEQUEN	AA•••A Default: SC	A represents the sequence of S and C characters used for iteration control after the initial split sequence. A maximum of 19 characters may be input.
OPTION	ERCOMP	Prints an error criterion for each iteration.
SYMBOLS	$S_1, S_2, S_3, \cdots$ Default: 1,2,•••9, A,B,•••Z,%,#,Δ/,-,*,+, $,@,=,0,?, , ,),(,:,!, ,;, ,',comma,period, blank,	Symbols used to identify clusters in the printout.
FORMAT	UNIVERSAL Default: Output MAPUNT tape is not generated.	Generates the output cluster MAPUNT tape in Universal format (see section 5.1 for further information).
FORMAT	LARSYS Default: Output MAPUNT tape is not generated.	Generates the output cluster MAPUNT tape in LARSYS format.
OPTION	CLUSTER Default: If the FORMAT control card	The output cluster MAPUNT tape will contain the cluster number to which the corresponding

TABLE 9-1.- Continued.

Keyword	Parameter and default values	Function
	is input, the output cluster MAPUNT tape will contain the mean vector of the cluster to which the corresponding pixel was assigned.	pixel was assigned. When selecting this option, the FORMAT control card must be input also.
NMIN	N	Deletes any cluster with fewer than N members on the first through next-to-last iteration (see section 9.1.1).
DLMIN	X Default: 3.2	On a combine iteration, combines any two clusters whose means are closer than X units.
PMIN	N	Deletes any cluster with fewer than N members on last iteration (see section 9.1.1).
SEP	X Default: Maximum of the channel standard deviations in the cluster	When splitting a cluster, separates the new clusters by a distance of X units.
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	N is the Fortran logical unit number to which the MSS data tape (DATAPE) has been assigned; M is the file number on the tape to be processed. For back-to-back executions, if

TABLE 9-1.— Continued.

Keyword	Parameter and default values	Function
		the same data file is to be processed throughout the execution, only one DATAFILE card need be submitted.
STATFILE	INPUT/UNIT=N,FILE=M, OUTPUT/UNIT=L,FILE=S Default: No defaults for INPUT; L=20,S=1 for OUTPUT.	N is the Fortran logical unit number to which the SAVTAP file containing the initial means has been assigned; M is the file number of the tape to be processed; L is the Fortran logical unit number to which the SAVTAP file containing the generated statistics will be output; S is the file number on the tape for saving the clustered statistics. (Input of initial means from SAVTAP file is illustrated in section 16, figure 16-4.)
ISTOP	N Default: 10	A maximum of N iterations is performed in the initial split sequence.
PERCENT	N Default: 80	N, an integer number, is the test variable for the percentage of stabilized clusters with standard deviations less than the threshold parameter STDMAX in the initial split iteration sequence.

TABLE 9-1.— Continued.

Keyword	Parameter and default values	Function
STDMAX	X Default: 4.5	On a split iteration, splits any cluster whose maximum standard deviation is greater than X units.
CLASSES	N Default: 1	Number of classes to be clustered (see section 9.5.4 for defining classes).
CLUSTERS	N Default: 60	Maximum number of clusters per class; N must be ≤60.
KRN	N Default: 20	Prints out a summary of the clusters at every N*th* iteration.
MAP	N Default: 20	Prints out a map of the clustered data along with the summary for every N*th* iteration. A final cluster map is printed regardless of this parameter.
CHAIN	X Default: Chaining not performed	Chains all clusters within X units of each other to form one cluster. Chaining of clusters affects only the final map printout and MAPUNT tape.
SUBCLASS	$C_1, C_2, C_3, \cdots, C_k$ $k \leq 60$ Default: All subclasses/clusters on SAVTAP file will be used in initializing the clustering.	C's are integer subclass or cluster numbers that (1) will be used in the initial means, (2) refer to the SAVTAP file, and (3) must be a subset of the subclasses or clusters on the SAVTAP file.

TABLE 9-1.— Continued.

Keyword	Parameter and default values	Function
MODULE	Blanks	Initializes the reading of the module STAT card file that immediately follows this card.
MEANS	CARDS Default: Clustering procedure is automatically initialized if this deck or MEANS file is not input.	Initializes input of the cluster MEANS deck defined in section 3.1.4.3. This deck is used to initialize cluster centers for the clustering procedure.
MEANS	FILE Default: Cluster centers are automatically initialized if this card or the MEAN card deck is not input.	Indicates means for initial clusters have been input previously from cards and stored on file. The same initial means are to be used again for initializing the process for a new data set.
DOTFILE	INPUT/UNIT=n,FILE=m Default: Self-initializing starting.	Defines the Fortran unit number n and file number m of the dot data file (DOTUNT unit) — containing the starting vectors.
DOTS	$n_1,n_2,\cdots,n_{60}$ Default: Dots will not be used for starting vectors.	n_i are integer numbers separated by a comma specifying the dots to be used as starting vectors.
SUNANG	TAPE Default: No sun angle correction applied.	Sun angles are extracted from the ERIPS unload MSS data tape.

TABLE 9-1.— Continued.

Keyword	Parameter and default values	Function
SUNANG	$n_1, n_2, \cdots, n_j$ n_j are integer numbers, $j \leq 8$. Default: No sun angle correction applied.	n_j are the sun angles to be used in computing the sun angle corrections for use in the clustering algorithm. A sun angle must be input for each set of 4 channels input on the CHANNEL control card.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11. Default: No comments printed.	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Indicates the end of control cards.

TABLE 9-1.— Concluded.

Keyword	Parameter and default values	Function
$END	Blank	Indicates the end of all classes to be clustered for this set of control cards. The SETUP routine will be reentered to read new control cards for the next class until all classes have been clustered.

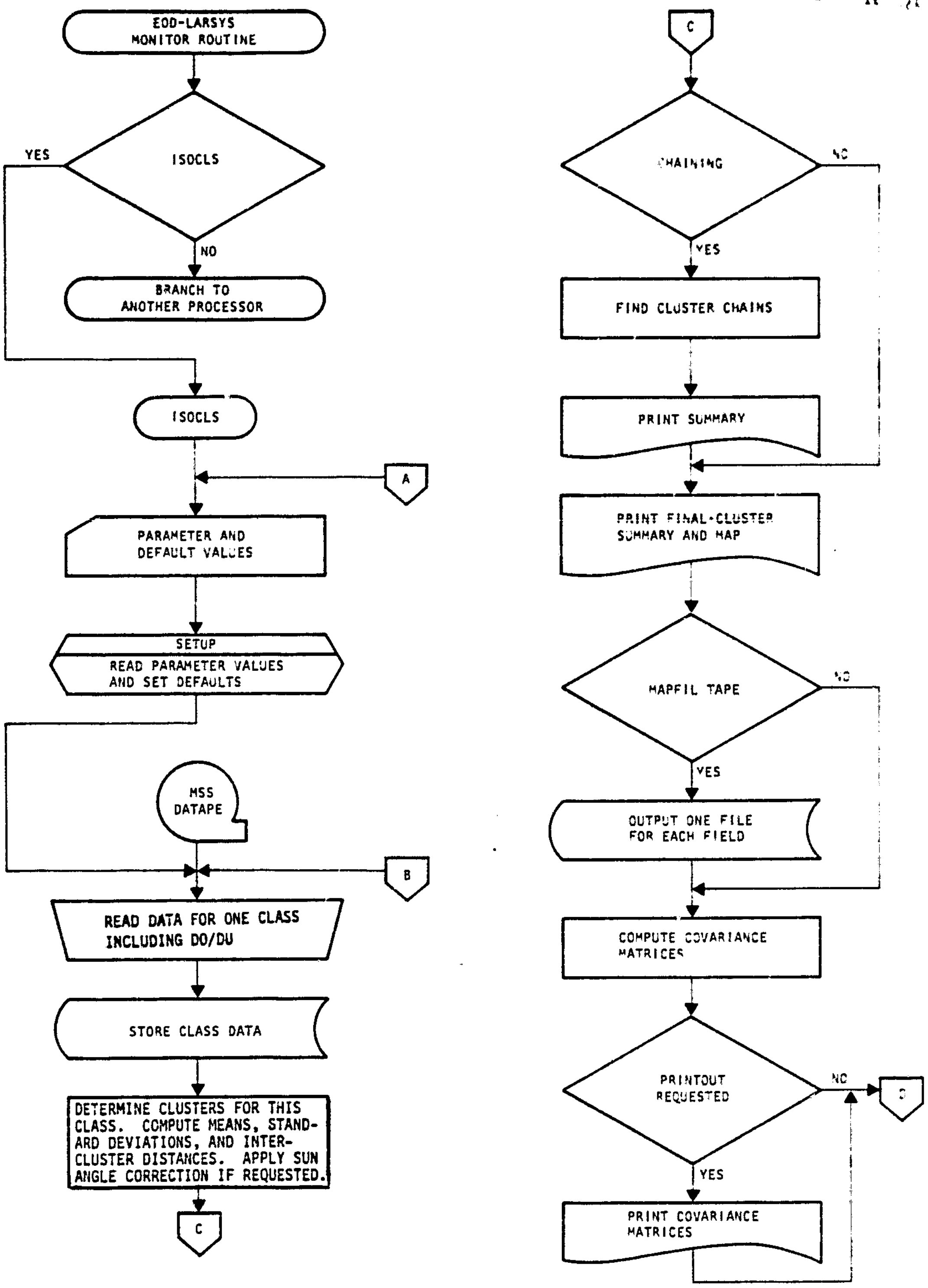


Figure 9-1.— Functional flow chart for the ISOCLS processor.

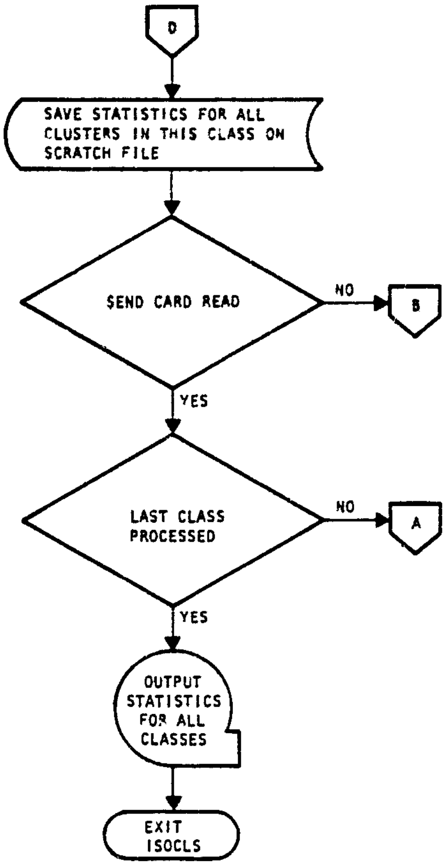
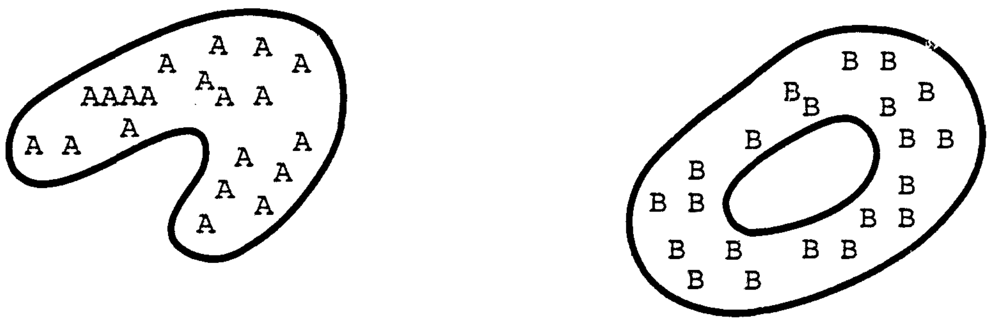


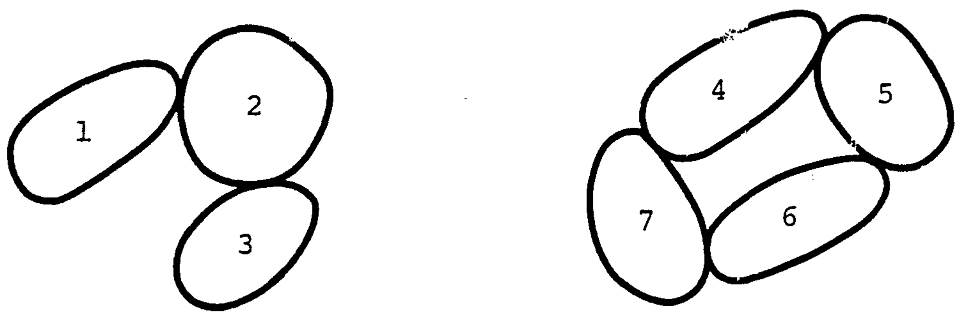
Figure 9-1.– Continued.



(a) The boomerang-shaped cluster.

(b) The donut-shaped cluster.

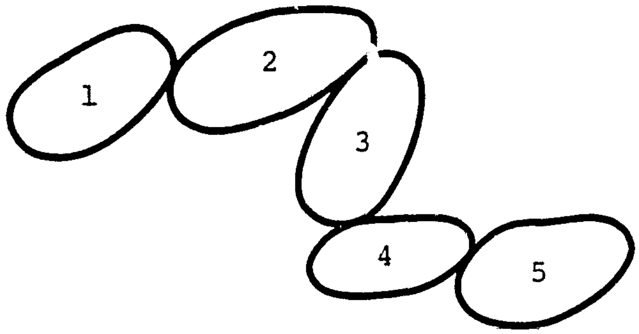
Figure 9-2.— Odd-shaped clusters.



(a) Subclustering of the boomerang-shaped cluster.

(b) Subclustering of the donut-shaped cluster.

Figure 9-3.— Breaking up of the clusters into subclusters.



(a) Cluster structure.

j \ i	1	2	3	4	5
1	0.0	7.5	6.2	3.2	11.8
2	7.5	0.0	3.1	5.6	3.0
3	6.2	3.1	0.0	3.1	6.3
4	3.2	5.6	3.1	0.0	9.7
5	11.8	3.0	6.3	9.7	0.0

CHNTHS = 3.2

(b) Intercluster distance table.

Figure 9-4.— Example of chaining.

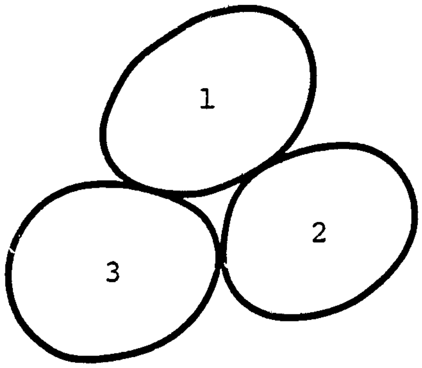


Figure 9-5.– Example in which chained subclusters can be combined safely into one composite cluster.

10. FEATURE SELECTION PROCESSOR — SELECT

The feature selection processor SELECT provides a means of measuring the relative importance of the individual channels and obtaining the set of channels which provides the best discrimination between subclasses. The processor allows the user to choose one of the following three criteria for measuring the separability of the subclasses for a set of channels or for linear combinations of the channels.

- Weighted average interclass divergence
- Weighted average transformed divergence
- Weighted average Bhattacharyya distance

Either the Exhaustive Search or the Without Replacement Procedure can be used with one of the above criteria to select a "best" set of channels. The Exhaustive Search Procedure determines the best set of k out of n channels by computing the separability measure for every possible combination of k channels. This results in $n!/k!(n - k)!$ computations of the separability measure. The computer time required for this procedure is prohibitive for a large n. In such cases, the Without Replacement Procedure could be used.

The Without Replacement Procedure determines the best k out of n channels in the following manner. First, the single channel which extremizes the separability measure is selected. Each of the remaining (n - 1) channels are paired with the best single channel in selecting the best pair of channels. The best triplet is determined by combining the remaining (n - 2) channels with the best pair. The process continues until the best set of k channels has been selected. The number of times the separability measure must be computed is $n + (n - 1) + (n + 2) + \cdots (n - k + 1)$.

A third procedure, the Davidon-Fletcher-Powell Procedure, is a powerful iterative descent method for finding a local minimum of a function of several variables. The procedure is discussed in reference 6. How the procedure applies to the problem of channel selection or dimensionality reduction is discussed in reference 7. In SELECT, the Davidon-Fletcher-Powell Procedure computes a k-by-n linear transformation matrix which extremizes a given separability measure. This matrix, referred to as the B-matrix, is saved on the BMFILE (section 4.2) and optionally is punched on cards (B-Matrix file, section 3.1.4.2) for later input to the CLASSIFY, SELECT, TRSTAT, SCTRPL, or DATA-TR processors.

An initial guess for the B-matrix must be provided for the Davidon-Fletcher-Powell routines and may be input via the B-matrix file on cards or BMFILE. If the initial guess is not provided by the user, SELECT will execute the Without Replacement Procedure first to obtain a best set of channels, which it will use to initialize a first-guess B-matrix for the Davidon-Fletcher-Powell Procedure.

In addition to selecting a best set of channels and/or linear combinations, the processor will evaluate any one of the three separability measures for a specified linear combination of the channels. The linear combination must be input via the B-matrix deck or the BMFILE if SELECT has been executed previously in the same run. This option is the fourth procedure defined under the PROCEDURE control card.

The processor will also evaluate any of the separability measures for specified sets of channels. This request is made using the EVALUATE and PROCEDURE control cards. This is the fifth option defined under the PROCEDURE control card.

For Procedure 1 applications, the SELECT processor optionally provides an option of an automatic assignment of interclass weights. The weights for $class_i$, $class_j$ = 1.0 for $i \neq j$ and the weights for $class_i$, $class_j$ = 0.0 for $i = j$. The breakdown of $class_i$, $class_j$ pairs into the correct set of intersubclass pairs is provided by the processor. Information concerning the class-subclass association is extracted from the input statistics file SAVTAP.

See the functional flow chart for the SELECT processor (fig. 10-1).

10.1 INPUT FILES

The SELECT processor requires the statistics output from either STAT or ISOCLS. Both STAT and ISOCLS write the SAVTAP file and optionally punch the module STAT file on cards (see section 3.1.4.1 for format) which may be used as input to SELECT.

10.2 OUTPUT FILES

The BMFILE is output by SELECT when the Davidon-Fletcher-Powell Procedure is used (see appendix I for sample execution). The file is written on logical unit 10.

The corresponding B-matrix card file is punched if the OPTION PUNCH control card is included in the deck setup.

10.3 SCRATCH FILES

Disk files are used as scratch files in SELECT. No assignment is necessary.

10.4 CARD INPUT

All system card input formats referred to in this section are defined in section 3.

10.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$SELECT

This card directs the monitor routine to execute SELECT and initiates loading of routines used by SELECT.

10.4.2 SYSTEM CARD FILES

The processor will read and process the module STAT file and the B-matrix file.

10.4.3 CONTROL CARDS

Table 10-1 lists the control cards which are recognized by SELECT.

10.4.4 FIELD DEFINITIONS

Field definitions do not apply to the SELECT processor.

10.5 CARD OUTPUT

SELECT outputs the B-matrix file on cards as an option (see control card OPTION PUNCH). This is optional output only when the Davidon-Fletcher-Powell Procedure is executed.

10.6 RESTRICTIONS

The system-related restrictions in section 19 apply to the SELECT processor.

Two large arrays are dimensioned in SELECT and used for the variable dimensioning of several smaller arrays. Storage in one array is a function of the number of subclasses and channels requested. That is,

$$\left[\begin{matrix}\text{Number of}\\ \text{channels}\end{matrix}\left(\frac{\begin{matrix}\text{Number of}\\ \text{channels}\end{matrix}}{2} + 3\right)\right]\begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix} + \left[\begin{matrix}\text{Number}\\ \text{of best}\end{matrix}\left(\begin{matrix}\text{Number}\\ \text{of best}\end{matrix} + 3\right)\right]\begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix}$$

$$+ \left(\begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix}\right)\left(\begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix} + 3\right) \leq 10\ 600 \qquad (10\text{-}1)$$

Storage requirements in the other array are dependent on the procedure and criterion being used. The Davidon-Fletcher-Powell Procedure requires much more storage than the other procedures, and the weighted average interclass divergence requires more storage than the other criteria. Requirements for each are as follows.

10.7 DIAGNOSTIC MESSAGES

10.7.1 SUBROUTINE AVEDIV

	Message	Explanation
a.	REDUCED COVARIANCE MATRIX FOR CLASS XXX IS NOT POSITIVE DEFINITE.	Check subclass/cluster statistics for singularity.
b.	MORE STORAGE NEEDED IN SUB. AVEDIV FOR WORK ARRAY-WORK SIZE=XXXXXXX.	Storage inadequate; adjust parameters.

10.7.2 SUBROUTINE BHTCHR

	Message	Explanation
a.	COVAR FOR CLASS XXX IS NOT POSITIVE DEFINITE.	Check subclass/cluster statistics for singularity.

Message	Explanation
b. COVAR FOR SUM OF CLASSES XXXX,XXXX IS NOT POSITIVE DEF.	Same as above.
c. NOT ENOUGH WORK AREA AVAILABLE IN BHTCHR -- IWRKSZ= XXXXX.	Storage inadequate; adjust parameters.

10.7.3 SUBROUTINE BSTCHK

Message	Explanation
a. AA0760 "BEST" XXX IS GREATER THAN OR EQUAL TO NO. OF FEATURES IN GIVEN DATA...IGNORED.	The channels included in "Best" must be a subset of the input channels.
b. INVALID EVALUATE REQUEST...	The channels whose separabilities are to be evaluated must be a subset of total input channels.

10.7.4 SUBROUTINE DAVDN1

Message	Explanation
ERROR ON DISK FILE - SUBROUTINE DAVDN1---ISTAT = XXX.	Tape hardware read error.

10.7.5 SUBROUTINE DAVDN2

Message	Explanation
MINIMUM IS AT ORIGIN - PROGRAM CANNOT CONTINUE.	Davidon-Fletcher-Powell iteration has reached an unusable minimum.

10.7.6 SUBROUTINE DAVDN3

Message	Explanation
DAVDN3--EITHER SIGYI OR YHY TOO CLOSE TO ZERO TO UPDATE H MATRIX--SIGYL=XXXXXXX. XXXXXXX. YHY=XXXXXXX.XXXXXXX.	Davidon-Fletcher-Powell iteration cannot continue.

10.7.7 SUBROUTINE DAVIDN

Message	Explanation
a. NOT ENOUTH WORK AREA AVAILABLE IN DAVIDN--IWRKSZ= XXXXXX.	Storage inadequate; adjust parameters.
b. ERROR ON DISK FILE - SUBROUTINE DAVIDN--- LSTAT=XXX.	Read error on disk file.

10.7.8 SUBROUTINE DIVERG

Message	Explanation
a. COVAR FOR CLASS XXXX IS NOT POSITIVE DEFINITE.	Check subclass/cluster for singularity.
b. NOT ENOUGH WORK AREA AVAILABLE IN DIVERG -- IWRKSZ=XXXXX.	Storage inadequate; adjust parameters.

10.7.9 SUBROUTINE GTSTAT

Message	Explanation
NOT ENOUGH WORK AREA IN GSTAT -- IWRKSZ=XXXXX.	Storage inadequate; adjust parameters.

10.7.10 SUBROUTINE SELECT

Message	Explanation
a. ERROR IN INPUT CHANNELS.	
b. CORE OVERFLOW IN SUBRAY - NN - STORAGE LOCATIONS NEEDED FOR THIS PROBLEM.	User might reduce the number of subclasses or channels or try another procedure. The SUBRAY array is used for temporary storage in SELECT only. (See restrictions, section 10.6.)
c. CORE OVERFLOW IN ARRAY - NN*2 - STORAGE LOCATIONS NEEDED FOR THIS PROBLEM.	See suggestions for first diagnostic message. The ARRAY array is used throughout the system for variably dimensioned storage.

10.7.11 SUBROUTINE SETUP4

Message	Explanation
a. ERROR ON STATFILE CARD.	Check format.
b. TOO MANY EVALUATE REQUESTS -- REMAINDER IGNORED.	The buffer to hold EVALUATE requests is dimensioned 100. The number of channels and channels to be evaluated for each EVALUATE request are stored in this array.
c. GROUP CARD IN ERROR - IGNORED.	Check format of GROUP option.
d. PROGRAM CANNOT PROCESS LESS THAN 2 CLASSES.	At least two channels must be input.
e. PROGRAM CANNOT PROCESS LESS THAN 2 CLASSES.	At least two classes must be input.
f. INVALID CONTROL CARD - IGNORED.	Check spelling of keyword.

Message	Explanation
g. CORE NEEDED IN ARRAY FOR THIS PROBLEM IS XXXXXX WORDS.	Storage is inadequate; adjust parameters.

10.7.12 SUBROUTINE TRNDIV

Message	Explanation
a. REDUCED COVARIANCE MATRIX FOR CLASS N IS NOT POSITIVE DEFINITE.	The indicated covariance matrix cannot be inverted.
b. NOT ENOUGH WORK AREA IN TRNDIV -- IWRKSZ=XXXXX.	Storage is inadequate; adjust parameters.

10.7.13 SUBROUTINE WGTCHK

Message	Explanation
SUBCLASS IS NOT AMONG INPUT SUBCLASSES - WEIGHT INPUT IGNORED.	Self-explanatory.

10.7.14 SUBROUTINE WGTSCN

Message	Explanation
a. SYNTAX ERROR ON WEIGHT CARD - REMAINDER OF CARD IGNORED.	Self-explanatory.
b. WEIGHT BUFFER IS FILLED - ONLY XXXXX CLASS NAME PAIRS ALLOWED.	Buffer storage is inadequate for all class name pairs.

10.7.15 SUBROUTINE WHRPLC

Message	Explanation
THE INCLUDE REQUEST FOR CHANNEL N IS NOT A LEGITIMATE REQUEST - IGNORED.	The indicated channel to be included is not among the input channels.

TABLE 10-1.— SELECT PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
BEST	$N_1, N_2, \cdots$ Default: None	Finds the best set of N_1, $N_2, \cdots$ channels if procedure 1 or 2 is indicated. If procedure 3 is indicated, the best N_1, N_2 linear combinations of the channels are found. $N_1, N_2, \cdots$ are integers separated by commas. A request can be made for a maximum of 10 best in one call to SELECT.
SUBCLASSES	$C_1, C_2, \cdots, C_k$ $k \leq$ number of subclasses on SAVTAP≤ 60 Default: All subclasses on the SAVTAP file	Provides for use of only subclasses $C_1, C_2, \cdots$ statistics for computation of separability measure; allows the user to select a subset of the statistics on the SAVTAP file for use in computing the subclass separabilities. $C_1, C_2, \cdots$ are integers representing the subclass number as it occurs in the SAVTAP file.
CHANNELS	$C_1, C_2, \cdots, C_k$ $k \leq$ number of channels on SAVTAP≤ 30	Selects the best set of channels from those indicated on this card. Must be a subset

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 10-1.– Continued.

Keyword	Parameter and default values	Function
	Default: All channels on the SAVTAP file	of the channels for which statistics are input via the SAVTAP file or module STAT card file. $C_1, C_2, \cdots$ are integers separated by commas.
OPTION	STATS Default: No statistics printed	Prints a summary of the statistics for the subclasses and channels actually used in SELECT.
WEIGHTS	C1=XX, (C1,C2)=YY, OTHERS=ZZ Default: All weights set to 1.0 for criteria 2 and 3. For criterion 1, weights for subclass pair (i,j) are $W_{ij}=e^{-D_{ij}/16}$, where D_{ij} is the divergence for subclass pair (i,j).	Sets weights for all subclass pairs of subclasses C1 to XX, then sets subclass pair (C1, C2) to YY; sets all other subclass pairs to ZZ. Subclass names C1, C2, etc., must match a subclass name from the module STAT card file, the SAVTAP file, or a GROUP name.[c] See OPTION-CLSWT for constraint.

[c]Consider the problem of selecting channels which best separate wheat from nonwheat classes, where wheat is divided into subclasses W1, W2, and W3, and nonwheat is divided into the subclasses NW1, NW2, NW3, and NW4. It is desirable to set all weights between subclasses in each class to zero, whereas wheat/nonwheat class pair weights are set to 1. This can be accomplished by the following WEIGHTS control cards: W1=1.,W2=1.,W3=1., (W1,W2)=0.; and (W1,W3)=0.,(W2,W3)=0.,OTHERS=0. W1=1. will set weights for the following subclass pairs equal to 1: (W1,NW1),(W1,NW2),(W1,NW3),(W1,NW4),(W1,W2),(W1,W3). W2=1. will set the weights for the following subclass pairs equal to 1: (W2,NW1),(W2,NW2),(W2,NW3),(W2,NW4),(W2,W3),(W2,W1). W3=1. will set weights for the following subclass pairs equal to 1: (W3,NW1),(W3,NW2),(W3,NW3),(W3,NW4),(W3,W1),(W3,W2). (W1,W2)=0. resets this subclass pair weight to 0. OTHERS=0. sets all other subclass pair weights to 0.

TABLE 10-1.— Continued.

Keyword	Parameter and default values	Function
B-MATRIX	CARDS Default: None	Indicates that the B-matrix card file immediately follows; results in the evaluation of the separability measure using the linear combinations defined by the B-matrix if procedure 4 is indicated. If procedure 3 is indicated, the B-matrix will be used as a first guess for the Davidon-Fletcher-Powell Procedure.
B-MATRIX	FILE Default: None	Indicates that a previous execution of SELECT has written the BMFILE. Depending on the PROCEDURE card, the B-matrix on file will be used as an initial guess for the Davidon-Fletcher-Powell Procedure or in evaluating the separability measure.
EVALUATE	$C_1, C_2, \cdots$ Default: None	Evaluates the separability measure indicated on the CRITERION card for channels $C_1, C_2, \cdots$. The set of channels to be evaluated must be (1) a subset of the channels on CHANNEL card and (2) must be on one card. Several sets of channels may

TABLE 10-1.– Continued.

Keyword	Parameter and default values	Function
		be input by using more than one EVALUATE card.
MODULE	Blank	Indicates that the module STAT card file immediately follows. The SAVTAP file will be written as this card file is read.
GROUP	NAME,I,J,··· Default: No grouping; individual subclasses are used.	Groups the training subclasses I,J,···, pools their statistics, and assigns NAME as the group name. NAME may be any six characters. Integers I,J,··· must correspond to the subclasses as they occur in the module STAT card file or the SAVTAP file.
PROCEDURE	N	N=1: The Exhaustive Search Procedure is used; N=2: The Without Replacement Procedure is used; N=3: The Davidon-Fletcher-Powell Procedure is used; N=4: The user-input B-matrix is used to evaluate the separability measure; N=5: The Evaluate Channels Procedure is used; and N=6: Invokes the best k of N options.

TABLE 10-1.– Continued.

Keyword	Parameter and default values	Function
OPTION	CLSWT Default: The weights are assigned to intersubclass pairs.	The processor determines the class/subclass correspondence (after any grouping of subclasses to form new subclasses if the GROUP control card is used) and assigns a weight of 1.0 to the subclass pairs associated with all interclass pairs. Intraclass subclass pairs are given a weight of 0.0. (NOTE: The WEIGHTS control card remains available to allow the user to set weights for specific subclass pairs. If used, the input subclass pair weights override the processor set subclass pair weights. The WEIGHTS OTHERS capability is not available when this option is exercised. If input, it is ignored by the processor.)
BSPASS	N Default: None	N is the number of passes to be included in the best set.
NCPASS	N Default: 4	N = number of channels per pass (acquisition).
APRIOR	Default: Omit card	This card sets the switch to modify intersubclass weights.

TABLE 10-1.– Continued.

Keyword	Parameter and default values	Function
CRITERION	N Default: N=1	The indicated criterion is used to measure the separability between subclasses. N=1 for weighted average divergence; N=2 for weighted transformed divergence; and N=3 for weighted average Bhattacharyya distance.
INCLUDE	$C_1, C_2, \cdots$ Default: None	Includes channels $C_1, C_2, \cdots$ in the best set; meaningful only for the Without Replacement Procedure. $C_1, C_2, \cdots$ must be a subset of channels on CHANNELS card.
STATFILE	UNIT=N,FILE=M	N is the Fortran logical unit number to which the SAVTAP file has been assigned; M is (1) the file number from which the training statistics are to be retrieved and (2), if the module STAT deck is input, the file number on which the statistics are to be stored. If M≠1, this control card must precede the module STAT card file in the control card deck setup.

TABLE 10-1.– Concluded.

Keyword	Parameter and default values	Function
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
ICOUNT	N Default: N=300	Number of iterations for the Davidon-Fletcher-Powell Procedure.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comment printed	Prints a comment line using the 60 characters found in the parameter field.
END	Blank	Signals the end of the control cards.
$END*	Blank	Signals the end of all card input for the processing function.

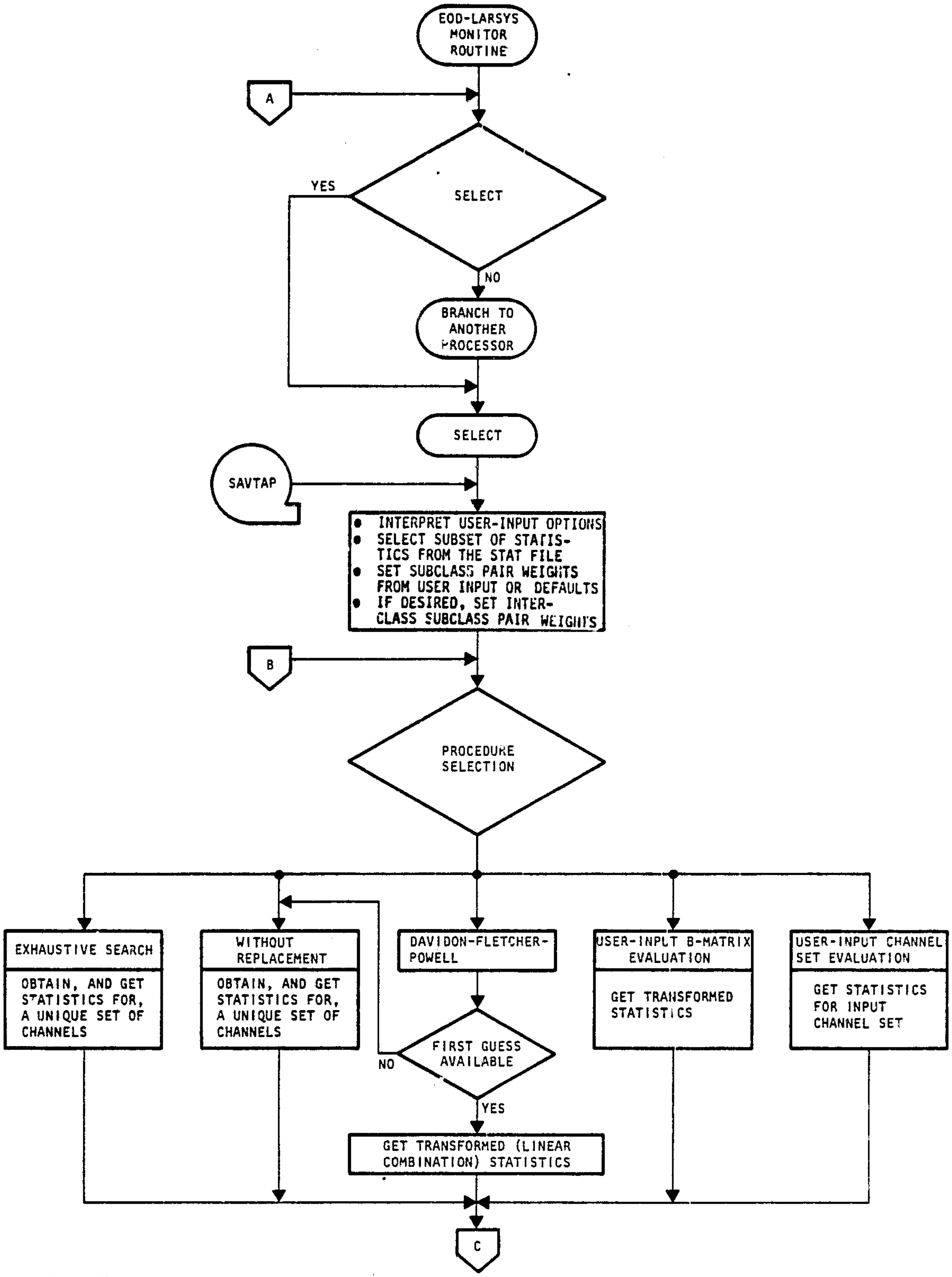


Figure 10-1.— Functional flow chart for the SELECT processor.

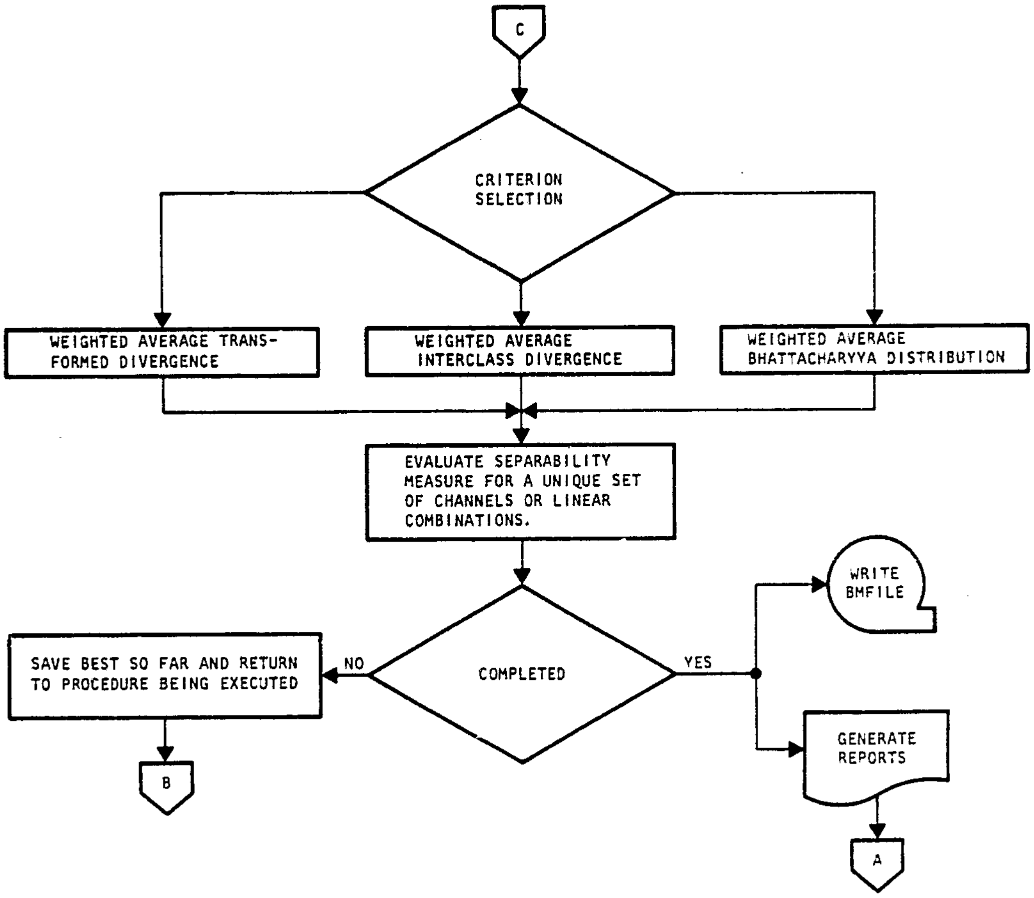


Figure 10-1.— Concluded.

11. CLASSIFICATION PROCESSOR - CLASSIFY

The classification processor CLASSIFY classifies the MSS image data based on statistics (mean vectors and covariance matrices) computed from the training fields.

11.1 PROCEDURES

Given the statistics for each subclass of interest, each data point within the defined classification field from the MSS data tape (DATAPE) is assigned to a subclass by one of two procedures.

In the first procedure, the user does not define categories in his input, and the standard m-class maximum likelihood classification rule is followed. However, to decrease the number of times the density function must be computed, the classification-by-thresholding procedure proposed by Hallum and Minter (ref. 8) and improved upon and implemented by Feiveson (ref. 9) is used. The standard classification rule (i.e., when no categories are defined by the user) is outlined in section 11.1.1.

In the second procedure, the user defines categories in his input, and the sum-of-normal-densities classification rule is followed, as set out in section 11.1.2.

11.1.1 STANDARD M-CLASS CLASSIFICATION

Assuming multivariate normal probability density functions and using the maximum likelihood classification rule, the data vector $X^T = (X_1, X_2, X_3, \cdots, X_N)$ is assigned to subclass i in the following manner.

The assumed joint probability density function of X, when material of type i fills the MSS field of view, is given in the following equation.

$$p_i(X) = \frac{P_i}{(2\pi)^{N/2}|K_i|^{1/2}} e^{-0.5Q_i} \qquad (11\text{-}1)$$

where:

P_i = *a priori* probability for subclass i

N = number of channels used for classification

K_i = covariance matrix for subclass i

X = data vector $(X_1, X_2, X_3, \cdots, X_N)$

$Q_i = (X - \mu_i)^T K_i^{-1} (X - \mu_i)$

μ_i = mean vector for subclass i

Because of the exponential form of p_i and because $\ell n(p_i)$ is a monotonically increasing function of p_i, for computational purposes it is convenient to define a new function V_i by

$$V_i = \ell n(p_i) = \ell n(P_i) - \frac{N}{2}\ell n(2\pi) - \frac{1}{2}\ell n|K_i| - \frac{1}{2}(X - \mu_i)^T K_i^{-1}(X - \mu_i) \qquad (11\text{-}2)$$

The data vector X is classified as belonging to subclass i if $V_i > V_j$ for all $i \neq j$, where $j = 1,2,3,\cdots,n$ and n = number of subclasses.

The number of times the function V_i must be computed may be reduced by the use of thresholds; i.e., real numbers γ_{ij} (independent of X) such that

$$\left.\begin{array}{ll} & V_i(X) > \gamma_{ij} \text{ implies } V_i(X) > V_j(X) \\ \text{and} & V_j(X) > \gamma_{ij} \text{ implies } V_j(X) > V_i(X) \end{array}\right\} \qquad (11\text{-}3)$$

where $i,j = 1,2,3,\cdots,n$ and $i \neq j$.

The utility of these thresholds is that, if $V_i(X) > \gamma_{ij}$, $V_j(X)$ need not be computed. Once the values for γ_{ij} have been determined, they may be used for each observation vector X.

11.1.2 SUM-OF-NORMAL-DENSITIES CLASSIFICATION

Also, assuming multivariate normal probability density functions, the category classifier classifies the data vector $X^T = (X_1, X_2, X_3, \cdots, X_N)$ to category j and subclass i in the following manner.

The probability density function for each category j is computed by the following equation.

$$p_j(X) = \sum_{i=1}^{k_j} \frac{P_i}{(2\pi)^{N/2}|K_i|^{1/2}} e^{-0.5Q_i} \qquad (11\text{-}4)$$

where

i = subclass number

j = category number

k_j = number of subclasses in category j

P_i = *a priori* probability for subclass i in category j

N = number of channels used for classification

Having computed the probability density function for all categories, the data vector X is classified as belonging to category j if $p_j > p_\ell$, where $\ell = 1,2,3,\cdots,q$ for all $j \neq \ell$ and q = number of categories.

The data vector is classified as belonging to subclass i if the probability density function for subclass i in category j is such that $p_{ji} > p_{jm}$ for all $i \neq m$, where $m = 1,2,3,\cdots,k_j$. In the computation of p_j, if the value of the quadratic form Q_i is smaller than -88, the computer cannot store the computed value of e^{Q_i}. Thus, $e^{Q_i} = 0$ for all $Q_i \leq -88$. In the case of all $P_j = 0$ for $j = 1,2,3,\cdots,q$, the data point will not be classified; it will be assigned to a null subclass.

When the line printer map of the classified data is displayed, each data point is printed with the symbol representing the legitimate subclass to which the data point belongs, and the null subclass is printed with the blank symbol. Figure 11-1 gives the functional flow of the CLASSIFY processor.

11.1.3 PROCEDURE 1

For Procedure 1 applications, the CLASSIFY processor allows an option for obtaining subclass *a priori* values using subclass population data from the input file, SAVTAP, and also allows the system to assign the category names using the class names from the input statistics file, SAVTAP, as the assigned category names.

Both options are an addition to the usual capability of analyst input of *a priori* probability values at the subclass, class, or category level via the APRIORI control card and of a category name input via the CATEGORY control card.

11.2 INPUT FILES

An MSS data tape (DATAPE) must be input to the CLASSIFY processor. The tape assignment defaults to logical unit 11; however, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information on format.)

11.3 OUTPUT FILES

The classification results are output on the MAPTAP file (see appendix D) which is assigned to logical unit 2. In the event of card input of the module STAT file, the statistics will be output on the SAVTAP file (see section 4.1).

11.4 SCRATCH FILES

The processor requires no scratch file.

11.5 CARD INPUT

All required input card types are described below.

11.5.1 PROCESSOR CARD

The processor keyword is left justified beginning in column 1; the parameter FILE is punched starting in column 11. For example,

$CLASSIFY FILE=N

This card directs the system monitor routine to load all routines used by the CLASSIFY processor. The parameter FILE=N informs the processor to output the current classification results on file N of MAPTAP. If no integer file number is specified, the processor defaults to file 1 of MAPTAP.

11.5.2 SPECIAL SYSTEM FILES

The training statistics may be input by means of the module STAT file. The B-transformation matrix may be input by means of the B-matrix file. The EOD-LARSYS deck formats are described in section 3.1. The use of card input is an option; normally, card image files are used.

11.5.3 CONTROL CARDS

The cards described in table 11-1 are the complete set of control cards recognized by the CLASSIFY processor. All options available to the user are exercised by use of the appropriate processor control card. If a default condition is specified, the control card is optional. If no default condition is specified, the control card is mandatory input. Ordering of the sequence of processor control cards is unnecessary, with the exceptions that (1) the *END card must follow the last processor control card, (2) the $END card must follow the last field definition card for an area to be classified, and (3) the STATFILE control card must precede the input of the module STAT file in some cases. See table 11-1.

11.5.4 FIELD DEFINITIONS

Areas to be classified are communicated to the classification processor by using the field definition data card described in section 3.1.3, which contains the scan line and sample coordinates for the area over which classification is to be performed. At least one field definition card must be in the run deck immediately following the *END control card. As many field definition cards as there are areas to be classified may be input. The processor will classify each field in the order it is identified, will print on the line printer the first 110 samples of the classification map, and will print any optional output prescribed by the control cards for each field classified. The scan line and sample coordinates specified on the field definition card must be available on the input MSS data tape (DATAPE).

11.6 CARD OUTPUT

The classification processor outputs no punched cards.

11.7 RESTRICTIONS

The system-related restrictions described in section 19, along with the following, apply to the CLASSIFY processor.

The category, class, and subclass relationship is as follows:

$$\text{Number of categories} \leq \text{Number of classes} \leq \text{Number of subclasses} \leq 60 \qquad (11\text{-}5)$$

The size of the B-matrix cannot exceed 450 locations:

$$\begin{pmatrix}\text{Number of linear}\\ \text{combinations}\end{pmatrix}\begin{pmatrix}\text{Number of chan-}\\ \text{nels in B-matrix}\end{pmatrix} \le 450 \qquad (11\text{-}6)$$

The channels used in computing the B-matrix automatically replace the channels, if any, on the CHANNELS control card.

The largest sample number of the classification field minus the smallest sample number of the classification field cannot exceed 1000.

Beginning with the smallest sample number of the classification field, only the next 110 samples are displayed on the line-printer map output by CLASSIFY, but the entire classified field is displayed on the line-printer map output by DISPLAY.

When applying the category classifier option, 12 500 storage locations are reserved for the data such that

$$\left(\frac{\text{Sample end - sample start}}{2 + 1}\right)\begin{pmatrix}\text{Number of}\\ \text{channels}\end{pmatrix} \le 12\ 500 \qquad (11\text{-}7)$$

When applying the standard classifier option, the table computed for the class-pair thresholding procedure shares this storage of 12 500 locations reserved for the data such that

$$\left[\begin{pmatrix}\text{Number of}\\ \text{subclasses}\end{pmatrix} - 1\right) \frac{\begin{pmatrix}\text{Number of}\\ \text{subclasses}\end{pmatrix} - 2)}{2} + \begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix}\right]$$

$$+ \left(\frac{\text{Sample end - sample start}}{2 + 1}\right)\begin{matrix}\text{Number of}\\ \text{channels}\end{matrix} \le 12\ 500 \qquad (11\text{-}8)$$

11.8 DIAGNOSTIC MESSAGES

The diagnostic messages and the routines in which they appear are as follows.

11.8.1 SUBROUTINE CLSFY1

	Message	Explanation
a.	***** CLSFY/CLSFY1 --- THE COVARIANCE MATRIX FOR SUBCLASS NO. XX IS EITHER SINGULAR OR NOT POSITIVE DEFINITE - THE DETERMINANT = XXXX.XXXX ***** TERMINATING PROGRAM EXECUTION *****	The determinant of each subclass covariance matrix is checked by CLASSIFY to see that it is a positive nonzero value. A zero value indicates a singular matrix, and a negative value i indicates a nonpositive definite matrix. If either condition occurs for any subclass covariance matrix to be used in classification, the processor will stop.[1]

11.8.2 SUBROUTINE CLSFY2

	Message	Explanation
a.	WIDTH OF RECTANGULAR FIELD SURROUNDING CLASSIFICATION FIELD CANNOT EXCEED 1000 POINTS.	The largest sample of the classification field minus the smallest sample of the classification field cannot exceed 1000 samples. Reduce amount of samples per scan line.

---

[1]A probable source of an invalid covariance matrix is a module STAT file which has been incorrectly formatted and thus is not producing good training class statistics. Another possible source is that the SAVTAP file does not contain valid statistical data.

Message	Explanation
b. AS THE COMPUTER CANNOT EXPONENTIATE A NUMBER SMALLER THAN EXP(-88), XXXXXX PTS WERE NOT CLASSIFIED IN THIS FIELD.	Self-explanatory.
c. TOO MUCH DATA REQUESTED.	When too much data has been requested, (1) for the standard classifier, reduce parameters so that $\left(\begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix} - 1\right)\left(\begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix} - 2\right) + \begin{matrix}\text{Number of}\\ \text{subclasses}\end{matrix} + \left(\begin{matrix}\text{Points per}\\ \text{scan line}\end{matrix}\right) \times \left(\begin{matrix}\text{Number of}\\ \text{channels}\end{matrix}\right) \le 12\ 500$; or (2) for category classifier, reduce data so that the number of points per scan line × number of channels $\le$ 12 500.

11.8.3 SUBROUTINE REDIF2

Message	Explanation
a. ERROR ON CHANNELS CARD.	Check job setup and unit assignments.
b. **** CLSFY/REDIF2 --- B-MATRIX INPUT STIPULATED BY CONTROL CARD... ****UNABLE TO INTERPRET TYPE OF B-MATRIX INPUT --- PROGRAM EXECUTION TERMINATED FROM REDIF2****	

	Message	Explanation
c.	** CLSFY/REDIF2 -- B-MATRIX INPUT FROM BMFILE - BAD INPUT VALUES DETECTED: NO. COMBINATIONS (BMCOMB) = ________, NO. CHANNELS (BMFEAT) = ______, CHANNEL VECTOR (BMVEC) = __. **** TERMINATING PROGRAM EXECUTION FROM REDIF2 *****	The input B-MATRIX control card is printed out as part of the error message. One of the data cards following it is incorrectly formatted. Check deck setup and B-matrix card file. Invalid data from the BMFIL has been deleted.
d.	*** CLSFY/REDIF2 - BAD CARD INPUT ON APRIORI CARD - DEFAULT APRIORI PROBABILITY VALUES WILL BE USED.	Check format of APRIORI card.
e.	AT LEAST TWO (2) CATEGORIES MUST BE ASSIGNED. EXITING FROM REDIF2.	In exercising the category option, two or more categories must be used.
f.	**** CLSFY/REDIF2 --- BAD PROCESSOR CONTROL CARD **** TERMINATING PROGRAM EXECUTION FROM REDIF2 ****	Check spelling of keyword.

11.8.4 SUBROUTINE SETUP2

	Message	Explanation
a.	***** CLSFY/SETUP2 ··· ERROR CONDITION ON ATTEMPT TO POSITION MAPTAP TO FILE NO. XX *****ERROR STATUS CODE=YY --- ABORTING THE RUN*****	The CLASSIFY processor attempted to position the output classification results file (MAPTAP) to the file number specified on the $CLASSIFY processor card. Possibly (1) more files were indicated than currently

	Message	Explanation
		existed on the MAPTAP file, (2) bad tape if the file is assigned to tape, or (3) the format of the $CLASSIFY processor card is incorrect.
b.	AN ERROR HAS OCCURRED IN GROUPING CLASSES INTO CATEGORIES. CHECK THE FOLLOWING: NOT ALL OF THE CLASSES HAVE BEEN ASSIGNED TO A CATEGORY. A CLASS NAME ON THE CATEGORY CARD HAS BEEN MISSPELLED. CLASS NAMES FROM SAVTAP FILE ARE: ________. CLASS NAMES FROM CATEGORY CARDS ARE: ______________.	When an error occurs in grouping classes into categories, either one or more class names (1) have not been assigned or (2) have been misspelled. The program lists the class names as submitted from the SAVTAP file or cards. Check these for errors. If neither (1) nor (2) is applicable, check the module STAT file to assure that class names are left justified in the field.
c.	USER INPUT A PRIORI VALUES DO NOT SUM TO 1.0. INPUT VALUES WERE NORMALIZED.**	Self-explanatory.
d.	** ERROR IN A PRIORI CONTROL CARD. USER INPUT VALUES IGNORED.**	Check format of APRIORI card.
e.	NO. OF CHANNELS REQUESTED FOR DATA TAPE AND NO. OF CHANNELS ON STAT FILE MUST BE EQUAL.	Self-explanatory.

TABLE 11-1.— CLASSIFY PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
SUBCLASS	$K_1, K_2, \cdots, K_i$ $1 \leq K_i \leq 60$; i = number of subclasses in training statistics Default: All the training subclasses are used.	K's are integers comprising the set of subclass numbers used by the processor to classify the unknown data points; must be a subset of training subclasses designated as they occur on the SAVTAP file.
CHANNELS	STAT=$N_1, N_2, \cdots, N_k$, DATA=$M_1, M_2, \cdots, M_k$ $k \leq 30$ Default: (1) If executed back to back with SELECT, the channels selected by the SELECT processor are used; (2) If a B-matrix is input, the channels used in computing the matrix are used; (3) Otherwise, all channels in the training statistics are used.	$N_1, N_2, \cdots, N_K$ are the channel numbers (integers) from the SAVTAP file to be used in classification; $M_1, M_2, \cdots, M_K$ are the channel numbers (integers) from the MSS data tape (DATAPE). The number of channels selected from SAVTAP and DATAPE must be equal.
CATEGORY	CATNAM/$NAME_1$,$NAME_2$,$\cdots$ Default: If no	Informs the processor that the category classifier option

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 11-1.— Continued.

Keyword	Parameter and default values	Function
	categories are defined, the M-class standard classifier is applied.	will be applied and defines one category name (CATNAM) and the class names ($NAME_i$'s) for this category. All subclasses for a class are assigned to this category. CATNAM and $NAME_i$ may be up to four characters, and $NAME_i$ must match a class name on the SAVTAP file. A slash (/) separates the category name from the class name. Note: (1) Every class must be assigned to a category unless the class was eliminated by omitting all of its subclasses on the SUBCLASS control card; (2) At least two categories must be defined; (3) Continuation of the list of class names in the category on another card is indicated by an asterisk after the last class name of that card. The next card should continue the list of class names in columns 11-72. (See test example in section 11.6.)
CATEGORY	FILE Default: No categories are defined	Initiates the assigning of the category names using the class names from the input

TABLE 11-1.– Continued.

Keyword	Parameter and default values	Function
	and the standard classifier will be applied.	statistics file, SAVTAP, and invokes the category classifier.
GROUP	SUBNAM,K_1,K_2,···,K_i $1 \leq K_i \leq 60$; i=number of training subclasses. Default: Subject to the SUBCLASS control card, each individual training subclass is used as a possibility for unknown data sample classification.	K_i's are integer subclass numbers taken from the set of available training subclasses. The processor creates a new training subclass by combining the statistics of the training subclasses listed. The training subclasses used are not thereafter available as individual subclass possibilities for an unknown data sample. The set of training subclasses to be used is renumbered by the processor to account for the new grouped subclass and the training subclasses deleted by grouping. The revised set of training subclasses is used for all processor output. SUBNAM may be from one to four characters and will become the name for a new training subclass.
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	N is the Fortran logical unit number to which the MSS data tape (DATAPE) has been assigned; M is the file number of the tape to be processed. For back-to-

TABLE 11-1.— Continued.

Keyword	Parameter and default values	Function
		back executions of several processors, if using the same file number, only one DATAFILE control card need be input.
STATFILE	UNIT=N,FILE=M Default: N=20,M=1	N is the Fortran logical unit number to which the SAVTAP file has been assigned; M is: (1) the number of the file to be processed or (2), if the module STAT file is input, the number of the file for storing the statistics. If M≠1, this control card must precede the module STAT file in the control card deck or card image file setup.
MODULE	Blank Default: Training subclass statistics are read from the input file SAVTAP.	Indicates to the processor that the training subclass statistics will be input on cards. The module STAT file must immediately follow this control card. See section 3.1.4.1 for further details.
B-MATRIX	CARDS or FILE Default: No transformation of training subclass covariance matrices	Informs the processor that the B-transformation matrix is to be input and applied to the training subclass statistics prior to classification. If FILE is placed in the parameter field, the mode of B-matrix input will be from BMFILE; if

TABLE 11-1.— Continued.

Keyword	Parameter and default values	Function
		CARDS is specified, the B-matrix card deck must immediately follow this control card. The channels which were used to derive the B-transformation matrix will be the channels used by the processor in classification. (See section 3.1.4.2 for further details.)
APRIORI	$A_1, A_2, \cdots, A_M$ or $N{*}A_i, K{*}A_{N+1}, \cdots, A_M$ $M \leq 60$ Default: If executing the standard classifier, each subclass is given an equal *a priori* value. If executing the category classifier, each category is given an equal *a priori* value which is divided equally among the subclasses in that category.	*A priori* values may be input by subclass, class, or category. N and K are arbitrary repetition factors, and A_i's are decimal numbers such that $\sum_{i=1}^{M} A_i = 1.0$ M = number of training subclasses, training classes, or categories. If input by class or category, the setup routine will distribute the *a priori* values among the subclasses in the following manner: By class = $\frac{\text{Class } a\ priori \text{ values}}{\text{Number of subclasses in that class}}$ By category = $\frac{\text{Category } a\ priori \text{ values}}{\text{Number of subclasses in that category}}$

TABLE 11-1.– Continued.

Keyword	Parameter and default values	Function
		The order in which the A_i's are input must be the order in which the category, class, or subclass was defined.
APRIORI	FILE Default: Subclass *a priori* is not to be computed from the statistics file, SAVTAP.	The subclass *a priori* probability values are computed using subclass or cluster point populations from the statistics file, SAVTAP.
OPTION	STATS Default: No training subclass statistics printout	Training statistics will be printed out for each subclass, reflecting the B-transformation, if any, and the Cholesky factorization of the covariance matrices.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.

TABLE 11-1.– Concluded.

Keyword	Parameter and default values	Function
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of card input for the processing function.

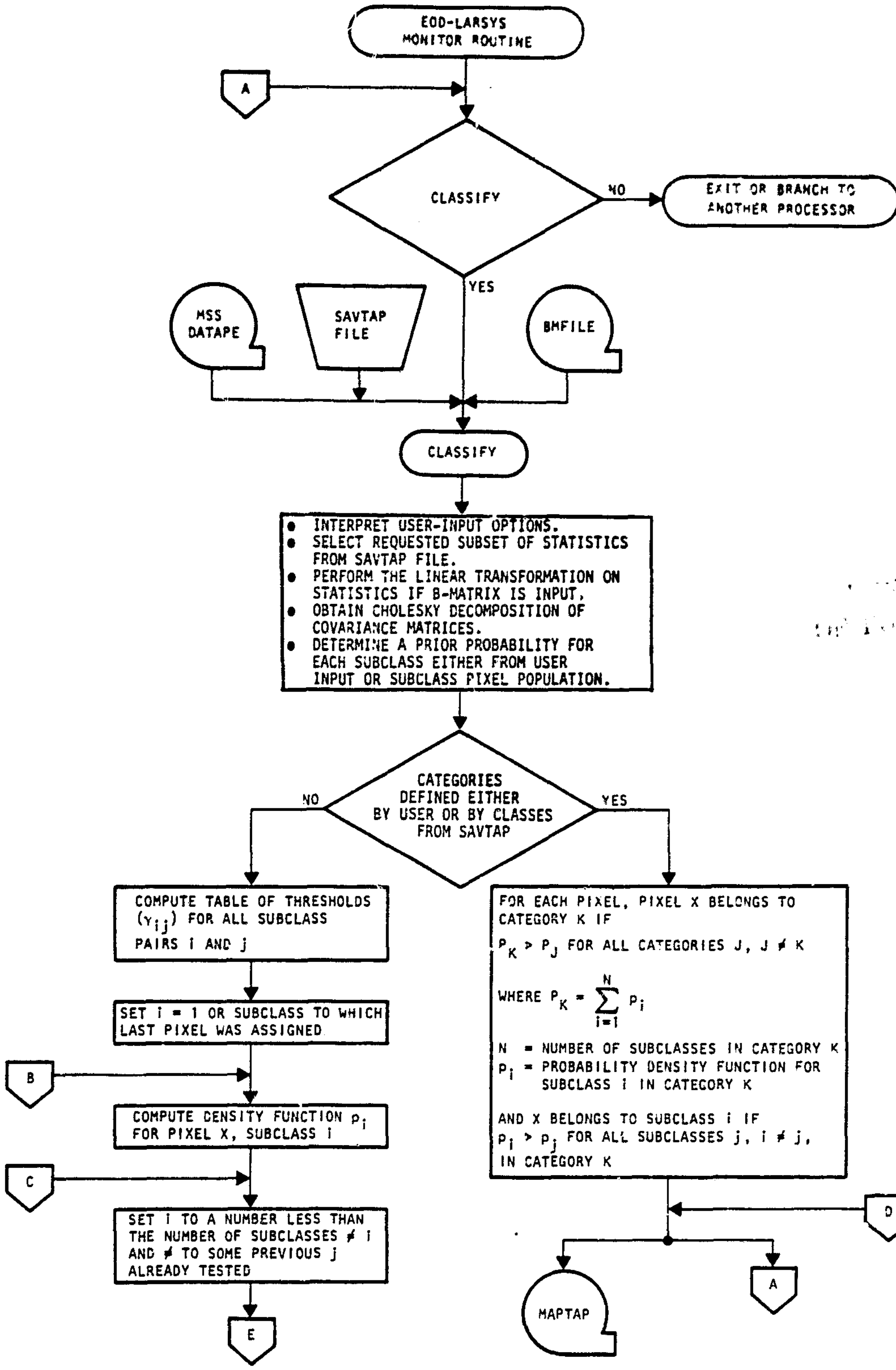


Figure 11-1.— Functional flow chart for the CLASSIFY processor.

ORIGINAL PAGE IS OF POOR QUALITY

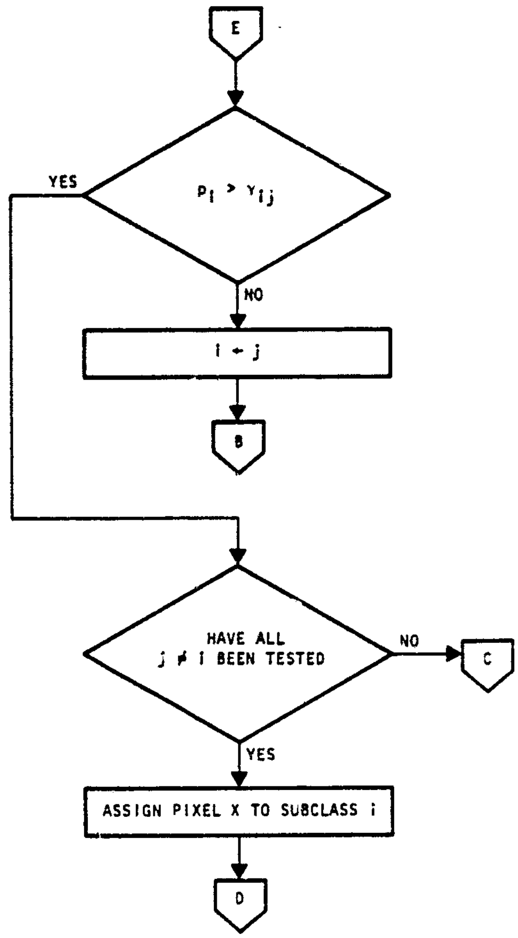


Figure 11-1.— Concluded.

12. PERFORMANCE DISPLAY PROCESSOR - DISPLAY

The DISPLAY processor reads the MAPTAP tape output by CLASSIFY and performs the following functions.

a. Provides a line printer map of each classified field on MAPTAP. The training and test fields within the classified image are outlined.

b. Produces classification summaries for each classified field, which gives a breakdown on the number of pixels classified into, and the number of pixels thresholded from, each subclass, class, and category.

c. Produces (optionally) an intensive test site (ITS) classification summary for one crop type versus all other crop types; the user-specified crop may be a category, class, or subclass.

d. Allows the user to designate fields to be excluded from the classification summaries. Fields may be designated "unidentifiable" or "other." Pixels within the unidentifiable fields are counted and are not considered in the classification summaries. Pixels within the designated "other" fields are counted as a separate crop type regardless of how they were classified. These pixels are included in category "other" in the ITS report. (See section 12.4.4 for sample input of designated fields.)

   All pixels within the designated areas are printed with the pound (#) symbol.

e. Assigns a pixel to the threshold class if thresholding is requested and if $Q_i > t_i$, where

   Q_i = the value of the quadratic form $(X - \mu_i)^T K_i^{-1} (X - \mu_i)$ as computed by CLASSIFY (section 11.1.1) for subclass i

   μ_i = mean vector for subclass i

K_i = covariance matrix for subclass i

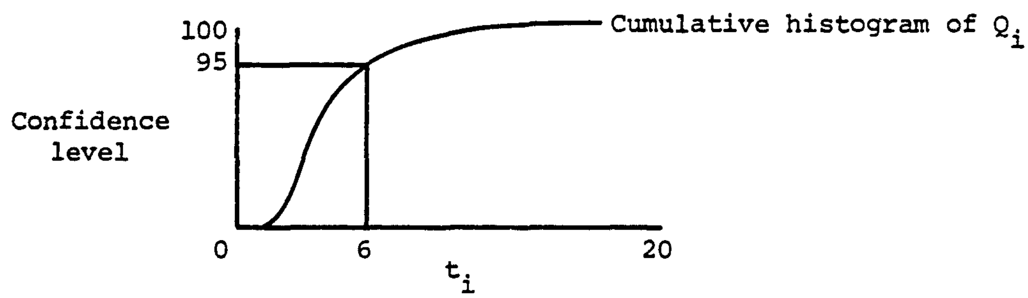
t_i = threshold value for subclass i

f. Allows t_i to be determined in one of four ways:

- User input — The user inputs the exact threshold value. See control cards THRESHOLD and OPTION THRESHOLD VALUE.

- Chi-square option — The user inputs confidence levels for each subclass on the THRESHOLD card and includes the OPTION CHI SQUARE. The program obtains the chi-square threshold value from an internal chi-square functional routine.

- Empirical option — The user inputs confidence levels for each subclass on the THRESHOLD card and includes the OPTION EMPIRICAL card. The program determines the empirical distribution function for each subclass from the cumulative histogram of Q_i for correctly classified pixels in the ground truth areas (i.e., training or test fields), as shown in the following example.



From this curve, the user input of a 95-percent confidence level for subclass i would result in a threshold value of 6.0.

See reference 10 for more information on the use of empirically computed thresholds.

- Fisher F-distribution option — The user inputs confidence levels for each subclass on the THRESHOLD card and includes the OPTION FISHER card. The program obtains the F-distribution threshold values from an internal routine. If a computational overflow occurs in the routine, the threshold value for that subclass is set equal to 999.999.

g. Produces plots of the empirical distribution function when OPTION PLOT is exercised.

h. Performs (optionally) a four-nearest-neighbors spatial filtering on the classified image. This algorithm takes into consideration that, in many instances, a pixel is most likely to be like its nearest neighbors. When the option is exercised via the OPTION FILTER control card, the four nearest neighbors of each pixel are examined. If all the neighbors are classified the same and the pixel in question is classified differently, then it is assumed that the pixel was classified incorrectly and its classification is changed. In the following example, the pixel classified as X will be changed to C. (See reference 11 for more information on this algorithm.)

Line 1		C	
2	C	X	C
3		C	

i. Outputs (optionally) the classified image onto tape (MAPUNT) in either LARSYS III or Universal format via the FORMAT NAME control card.

j. Provides classification performance summaries for ground truth areas within the classified image. The following six performance summaries are available to the user. The fields in these reports are either training fields used in the STAT or ISOCLS processor and transmitted to DISPLAY via the MAPTAP file or test fields input directly to DISPLAY (see section 12.4.4).

- Field by subclass

- Field by class
- Field by category
- Class by subclass
- Class by class
- Class by category

k. In Procedure 1 applications, DISPLAY is able:

1. To accept the dot data file, DOTUNT.
2. To provide a dot classification performance summary by dot categories which also includes:
   - A tabulation of both the uncorrected proportion and the bias corrected proportion of each dot category in the total area classified.
   - An "alpha" table which tabulates proportions for each labeled category of bias correction dots (type 2 dots), which, for each category, are the ratios of the number of dots with the given analyst-labeled category that were classified into each possible category, to the total number of dots classified into each category.
3. To provide a dot classification performance summary for each dot on the analyst's specified file (DOTUNT).

The functions of the DISPLAY processor are such that the analyst may either exercise the initial processor capabilities (a) through (j) or the LACIE Procedure 1 capabilities (k). The difference between the two capabilities is in the type and format of classification performance tables output.

Figure 12-1 shows a function flow chart of the DISPLAY processor.

12.1 INPUT FILES

The only input file required for DISPLAY is the MAPTAP (section 4.4) file output by CLASSIFY. This file must be assigned to logical unit 2.

For Procedure 1, the DISPLAY processor optionally accepts as input the dot data file, DOTUNT, created by the DOTDATA processor. The file is assigned either to logical unit 19 or to a user-specified unit.

12.2 OUTPUT FILES

The DISPLAY processor will optionally generate a tape of the classified image on the MAPUNT unit for display on the DAS. The control card FORMAT allows the user to exercise this option. When requested, the tape should be assigned to a nine-track tape drive for compatibility with the DAS tape drives. The tape assignment must be made to logical unit 16.

For Procedure 1, the DISPLAY processor optionally provides dot data classification performance summaries (instead of the normal output classification summaries described above) if the DOTUNT control card is input to the DISPLAY processor.

12.3 SCRATCH FILES

The disk provides random access storage for a scratch file in DISPLAY. No assignment is necessary.

12.4 CARD INPUT

All system formats referred to in this section are defined in section 3.

12.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1; the parameter FILE is punched starting in column 11. For example,

$DISPLAY FILE=N

This card directs the monitor routine to select the DISPLAY processor and initiates the loading of routines used by DISPLAY. Parameter value N is the file number on the MAPTAP file to be processed; if not input, default is to file 1 of MAPTAP.

12.4.2 SPECIAL SYSTEM FILES

No special files are required for the DISPLAY processor.

12.4.3 CONTROL CARDS

Table 12-1 lists the control cards and available options for the DISPLAY processor.

12.4.4 CLASS, SUBCLASS, AND FIELD DEFINITIONS

Both test and designated fields are optional input to DISPLAY. However, both types of fields cannot be input in the same execution of DISPLAY. If no test fields are input, the ground-truth summaries will be for training fields. When input, test fields must be identified with a previously defined class or subclass. All test class, subclass, and field definitions begin immediately following the *END control card and are terminated by the $END control card. Formats for the CLASSNAME, SUBCLASS, and field definition cards are defined in section 3.1.3. The following example shows test field input to DISPLAY. Note that test fields are identified with classes; that is, each NAME1, NAME2, NAME3, etc., must match the name of a class defined in either STAT or ISOCLS and used in CLASSIFY.

```
$DISPLAY
(Control cards)
*END
CLASSNAME    NAME1
FIELD1        ---
FIELD2        ---
CLASSNAME    NAME2
FIELD3        ---
FIELD4        ---
FIELD5        ---
CLASSNAME    NAME3
FIELD6        ---
$END          ---
```

In the following example, test fields are identified with subclasses, in which case each NAME1, NAME2, NAME3, etc., must match the name of a subclass used in CLASSIFY.

```
$DISPLAY
(Control cards)
*END
SUBCLASS     NAME1
TEST FIELDS  FOR SUBCLASS NAME1
SUBCLASS     NAME2
TEST FIELDS  FOR SUBCLASS NAME2
$END
```

Note that actual name must not exceed four characters.

Designated fields are large areas within the classified area which are either unidentifiable or can be specifically identified as being other than the crop type of interest. This type of field input is meaningful only when the ITS summary report is being generated for one specific crop type. Pixels within unidentifiable areas are removed from the summaries altogether. Pixels within the designated "other" areas are counted as other regardless of how they were classified.

An example of input designated fields follows.

```
$DISPLAY
(Control cards)
*END
DESIGNATE UNIDENTIFIABLE
(Field definitions)
DESIGNATE OTHER
(Field definitions)
$END
```

Either one, both, or neither of the two types of designated fields may be input.

If the Procedure 1 option is to be exercised, the only kinds of fields that may be input are DO/DU fields. No test fields may be input; the training fields are the bias correction dots on the dot data file, DOTUNT. The format of the DO/DU field cards and the method of input are given in section 3.1.3.

12.5 CARD OUTPUT

No cards are output by the DISPLAY processor.

12.6 RESTRICTIONS

The system-related restrictions given in section 5 apply to this processor.

12.7 DIAGNOSTIC MESSAGES

See section 6 for further diagnostic messages.

12.7.1 SUBROUTINE DISTCV

Message	Explanation
OVERFLOW.	Error flag set by CHIN indicates overflow condition.

12.7.2 SUBROUTINE DSPLAY

Message	Explanation
DISPLAY - FIELDS MUST BE DEFINED FOR SUBCLASSES FOR EMPIRICAL THRESHOLDS.	Self-explanatory.

12.7.3 SUBROUTINE DSPLY1

Message	Explanation
NOT ENOUGH STORAGE FOR COVARIANCE MATRICES - DSPLY1.	Adjust parameters.

12.7.4 SUBROUTINE DSPLY2

Message	Explanation
a. DISPLAY WILL ACCEPT ONLY 1000 PTS/SCAN LINE.	Adjust parameters.

Message	Explanation
b. **** DSPLY2/DOTSUM - - DISCREPANCY IN DOTFILE INFORMATION **** NO. OF DOT CATEGORY LABELS MATCHING MAPTAP CATEGORY NAMES = XXXXXX. NO. OF DOT CATEGORIES IS GIVEN AS = XXXXXX.	Check if correct DOTFILE is used.
c. **** DSPLAY/DSPLY2 **** NO. OF DOTS = XXXXXX --- EXCEEDS THE MAX. ALLOWABLE (250). **** DOT PERFORMANCE SUMMARIES WILL NOT BE PRODUCED.	Maximum number of dots exceeded.

12.7.5 SUBROUTINE EMTHRS

Message	Explanation
ERROR BACK SPACING MAPTAP. ISTAT = XXXXX.	Hardware tape-read error.

12.7.6 SUBROUTINE FDIST

Message	Explanation
FDIST-OVERFLOW CONDITION IN FISHIN ROUTINE FOR SUBCLASS = XXXX. THRESHOLD SET TO 999.999.	The FISHIN system subroutine has returned an overflow condition. The threshold value is set to 999.999 by the program.

12.7.7 SUBROUTINE PRTSUM

Message	Explanation
THE CROP NAME XXXX DOES NOT MATCH A CATEGORY, CLASS OR SUBCLASS NAME.	Check spelling on CROP cards.
THE INTENSIVE TEST SITE SUMMARY REPORT CANNOT BE PRINTED.	

12.7.8 SUBROUTINE REDIF3

	Message	Explanation
a.	**** DSPLAY/REDIF3 --- ERROR IN 'OPTION' CARD ••• **** SCAN OF THIS CARD DISCONTINUED --- PROCEEDING TO NEXT CARD ****	Check format and spelling of parameter.
b.	*****FISHER THRESHOLD REQUESTED - NOT PERFORMED •••NO. SAMPLES FOR SUB-CLASS NAME (=N) IS LESS THAN OR EQUAL TO NUMBER OF CHANNELS (=M).	The program compares the number of samples to the number of channels. If the number of samples $\leq$ number of channels, the threshold request is bypassed.
c.	ERROR IN ACREAGE CARD — CARD IGNORED.	Check format.
d.	ERROR ON DOTFILE CARD.	Check format and parameters.
e.	*** A THRESHOLD VALUE IS OUTSIDE THE ALLOWABLE RANGE 0 - 1, THEREFORE NO THRESHOLDING HAS BEEN DONE IN THIS RUN *** XXXXX, XXXXXXXXX.XXXXX	Check format of THRESHOLD card.

	Message	Explanation
f.	* ERROR ON SUBCLASS NAME CARD – XXXX DOES NOT MATCH A SUBCLASS FROM THE MAPTAP FILE.	Self-explanatory.
g.	* ERROR ON CLASSNAME CARD – XXXX DOES NOT MATCH A CLASS NAME FROM THE MAPTAP. FILE.*	Self-explanatory.
h.	TEST FIELDS AND DESIGNATED FIELDS CANNOT BE INPUT TOGETHER.	Self-explanatory.
i.	INVALID CONTROL CARD- CHECK SPELLING OF KEY WORD.	Self-explanatory.

12.7.9 SUBROUTINE SETUP3

	Message	Explanation
a.	*****DSPLAY/SETUP3 ••• ERROR CONDITION ON ATTEMPT TO POSITION MAPTAP OVER _____ FILES. *****FSBSFL STATUS CODE = _______ --- ABORTING RUN ***	The system routine for positioning files (FSBSFL) has encountered difficulties in positioning MAPTAP to the correct file. Error occurred in SETUP3 routine for DISPLAY. User should make sure that the correct file number for the MAPTAP has been indicated and that MAPTAP does in fact have the correct number of files.

	Message	Explanation
b.	****** DISPLAY/SETUP3 — CORE OVERFLOW (TOP-TOP2) BY XXXXXX -- EXECUTION TERMINATED ******	Subroutine SETUP3 has computed the storage needed for the specific problem; if more is needed than is available, this diagnostic is printed.
c.	CLASSIFICATION BY CATEGORY (ON MAPTAP) IS REQUIRED IN ORDER TO PROCESS THE DOT DATA *** *** DOT PERFORMANCE SUMMARIES WILL NOT BE OUTPUT ****	Procedure 1 uses category classifier only.
d.	*** CLASSIFICATION BY CATEGORY (ON MAPTAP) IS REQUIRED IN ORDER TO PROCESS THE DOT DATA *** *** DOT PERFORMANCE SUMMARIES WILL NOT BE OUTPUT ***	Self-explanatory.
e.	**** SETUP3** FROM DOTFIL, THE NO. OF DOT CATEGORIES = XXXXXX. **** DOT PERFORMANCE SUMMARIES WILL NOT BE PROVIDED ****	Error in DOTFIL - reset DOTKEY and TRNKEY - turn on DOTERR.
f.	*** DOT DATA PERFORMANCE SUMMARIES WILL NOT BE PRODUCED — THE CATEGORY NAMES FROM MAPTAP AND DOTFILE DO NOT MATCH.	Change appropriate category names. Correct as required.

Message	Explanation
CATEGORY NAMES FROM DOTFILE ARE –	
**** SETUP3/DTCHK --- FROM DOTFILE, THE MIN. LINE NO. = XXXXXX. MAX. LINE NO. = XXXXXX.	
***** DOT PERFORMANCE SUMMARIES WILL NOT BE PRODUCED *****	

TABLE 12-1.– DISPLAY PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
OPTION	PLOT	Plots the empirical distribution functions obtained from the cumulative histograms of Q_i for each subclass.
OPTION	CHI SQUARE Default:[c]	Computes thresholds from the chi-square distribution using the confidence levels input on the THRESHOLD control card.
OPTION	FISHER Default:[c]	Computes thresholds from the Fisher F-distribution using the confidence levels input on the THRESHOLD control card.
OPTION	EMPIRICAL Default:[c]	Computes the empirical threshold values using the percentages input on the THRESHOLD control card.
		Uses the numbers input on the THRESHOLD control card for the actual threshold value.

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

[c]If the THRESHOLD control card is input, one of the four options (CHI SQUARE, FISHER, EMPIRICAL, or THRESHOLD VALUE) should be input also. If the OPTION card is omitted and the THRESHOLD card is input, chi square is assumed. If more than one THRESHOLD option is input, only the last one read will be performed.

TABLE 12-1.— Continued.

Keyword	Parameter and default values	Function
SYMBOLS	$S_1, S_2, \cdots S_k$ k=number of subclasses on MAPTAP Default: 1,2,•••9 A,B,C,D,•••,Z, 1,2,3,4	Assigns symbols $S_1, S_2, \cdots$ to subclasses 1,2,•••, respectively.
OPTION	STAT Default: No statistics printed	Prints statistics for subclasses used in the previous CLASSIFY run. These statistics are saved on the MAPTAP.
OPTION	PCT Default: Performance summary printed for classes only	Prints a performance summary on a per-field as well as a per-class basis for ground-truth fields (i.e., training or test fields within the classified image).
OPTION	OUTLINE Default: Training fields are not outlined.	Outlines training fields with asterisks; has no effect on test fields. (Test fields are always outlined with "+" symbol.)
OPTION	NOMAP Default: Map printed	Instructs the processor not to print a map of the data; only a performance summary is printed.
OPTION	FILTER Default: Spatial filtering is not performed.	Performs four-nearest-neighbors spatial filtering on the classified image.

TABLE 12-1.— Continued.

Keyword	Parameter and default values	Function
THRESHOLD	$T_1, T_2, \cdots T_k$ k=number of subclasses on MAPTAP Default: No thresholding	Uses the threshold values $t_1, t_2, \cdots$ for subclasses 1, 2, ···, respectively; thresholds must be positive floating-point numbers. One value must be specified for each subclass on the MAPTAP file. Thresholds may be specified also in the following format: $N_1{*}t_1, N_2{*}t_2, \cdots$ where N_1 and N_2 are integers which specify how many consecutive times the corresponding thresholds should be used. For the CHI SQUARE and the EMPIRICAL options, the numbers input on these cards are the confidence levels (i.e., t_1=0.99 means that the user wants to maintain 99% or reject 1%). The numbers input on the OPTION THRESHOLD VALUE card are the actual values to be used for thresholding (i.e., t_1=10.02 means that the threshold value for subclass 1 is 10.02).

TABLE 12-1.— Continued.

Keyword	Parameter and default values	Function
CROP	NAME Default: No ITS report	Initiates the option for printing the ITS summary report for the crop indicated. NAME must match a category, class, or subclass name used in CLASSIFY.
ACREAGE	TOTAL=X, CROP=Y, OTHER=Z	The total acreage in the ITS is X; acreage of the crop named on the CROP control card is Y; and the acreage of all other crop types in the ITS is Z. X, Y, and Z are floating-point numbers. This input is meaningful only if the CROP control card is input.
SITE	Any 24 characters Default: Blanks	Name of the ITS; used in printing the heading for the ITS summary report.
ANALYST	Any 18 characters Default: Blanks	Name of the data analyst printed in the heading for the ITS summary report.
PROCEDURE	Any 60 characters Default: Blanks	Procedure used in classification of the ITS; printed in the heading for the ITS summary report.

TABLE 12-1.– Concluded.

Keyword	Parameter and default values	Function
FORMAT	NAME Default: No output classification map tape is generated by DISPLAY.	If NAME=UNIVERSAL, the output classification tape (MAPUNT) will be generated in the Universal format. If NAME=LARSYS, the MAPUNT tape will be generated in the LARSYS III format.
DOTFILE	INPUT/UNIT=n,FILE=m Default: UNIT= 19, FILE=1. No dot data classification performance summaries.	Initiates the input of the dot data file, DOTUNT, from the designated (or default, if not designated) unit and/or file, and initiates the output of the dot data classification performance summaries. The parameter, m, designates the file to be processed by DISPLAY, and n designates the Fortran unit number assigned to the input file, DOTUNT.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.

TABLE 12-1.— Concluded.

Keyword	Parameter and default values	Function
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of all card input for the processing function.

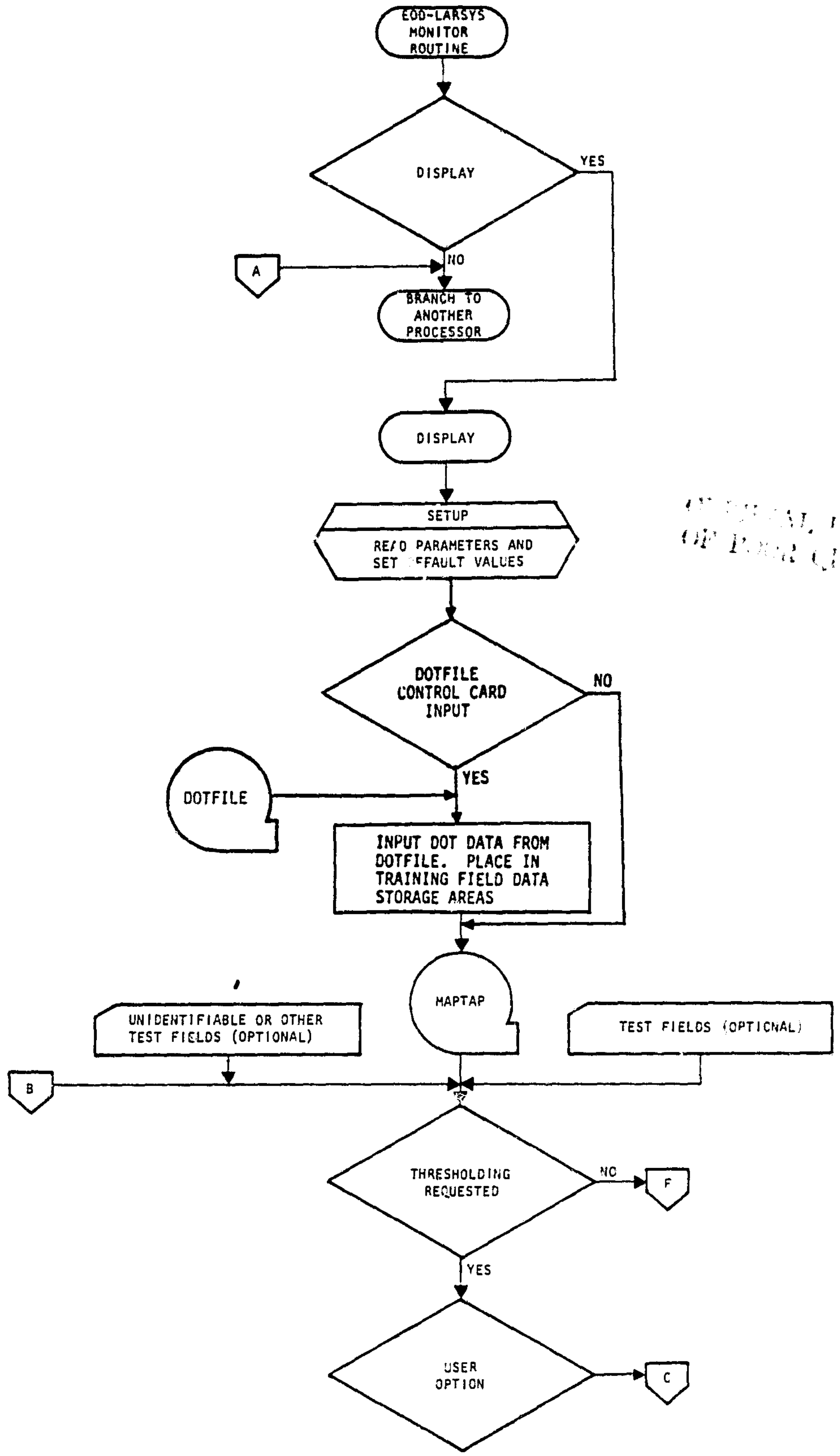


Figure 12-1.— Functional flow chart for the DISPLAY processor.

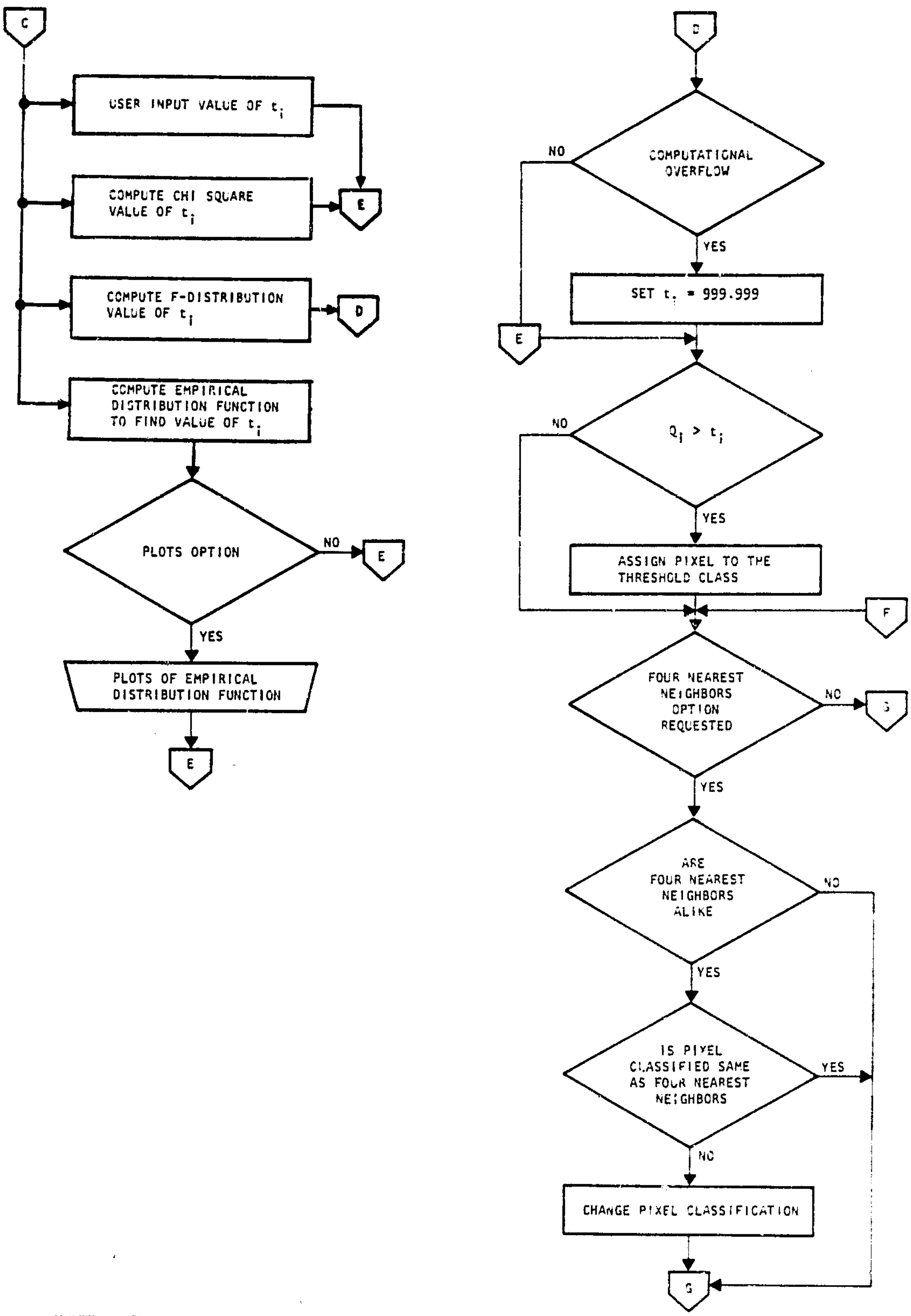


Figure 12-1.— Continued.

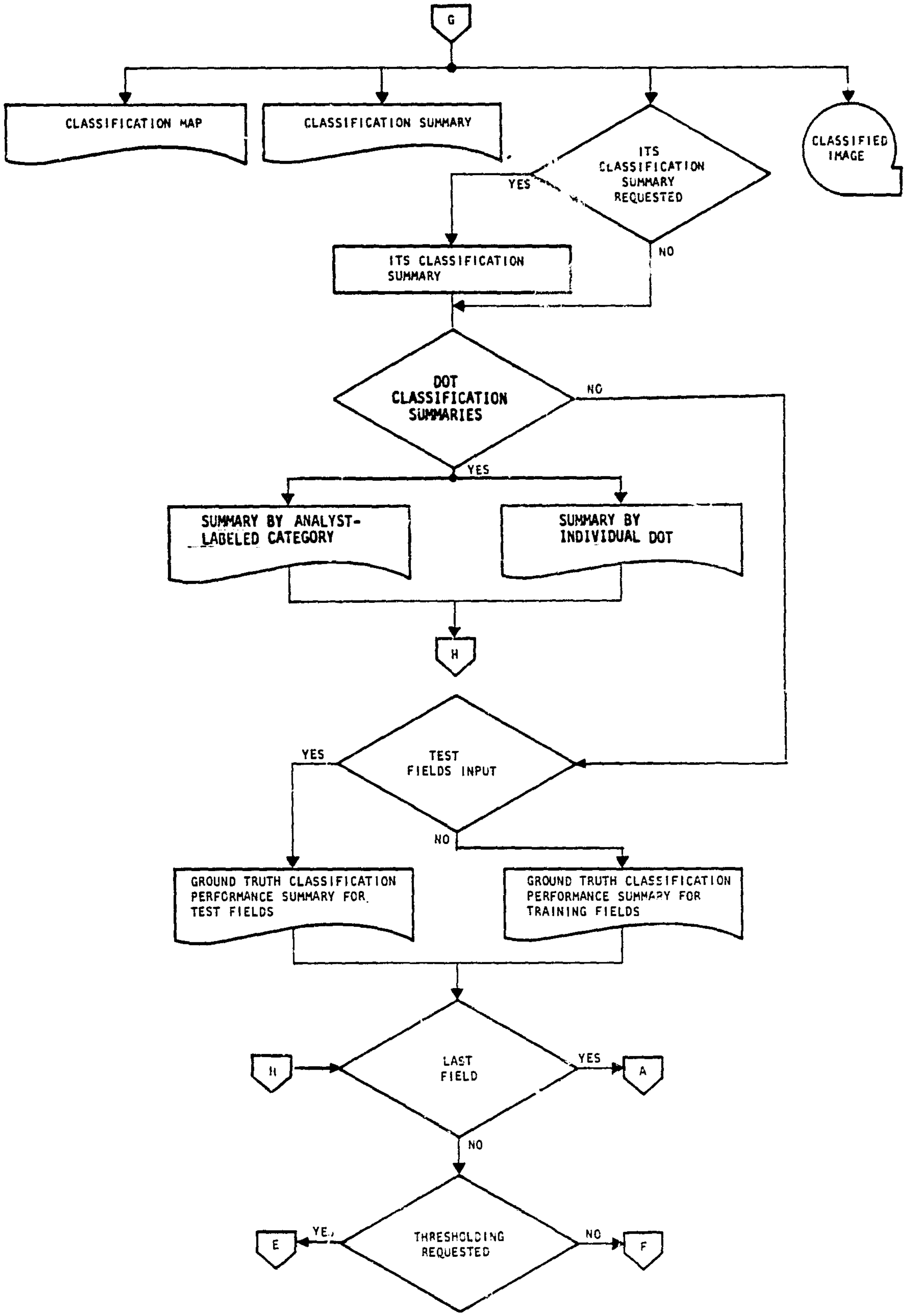


Figure 12-1.– Concluded.

13. DATA-TRANSFORMATION PROCESSOR – DATA-TR

13.1 PROCEDURES

The DATA-TR processor transforms images from the MSS data tape (DATAPE). The linear transformation is performed on user-defined fields according to the following formula:

$$z = B\overline{x} + b \qquad (13\text{-}1)$$

where

$k \leq 16$

$n \leq 30$

B = a k-by-n input transformation matrix (see section 3.1.4.2)

$\overline{x}$ = an n-by-1 data vector

b = a k-by-1 bias vector (see BIAS control card, table 13-1)

z = a k-by-1 transformed data vector

The user has the option of rescaling the transformed data via the RESCALE control card. (Control cards are listed in table 13-1.) If no rescaling is performed, it is assumed that the data can be represented in eight bits ($0 \leq z \leq 255$). For rescaling the data to be within the range of 0 to 255, the following equation is computed.

$$Y_i = \frac{255}{R_i} \times |MIN_i - Z_i| \qquad (13\text{-}2)$$

where

MIN_i = minimum value for component i

Z_i = transformed data point for channel i

R_i = range of component i $[MAX_i - MIN_i]$

MAX_i = maximum value for component i

Y_i = rescaled transformed data point for channel i

The user may obtain the parameters R_i and MIN_i in one of three ways: the histogram method, the statistical method, or user input. The method and control cards associated with each method are defined below.

13.1.1 HISTOGRAM (DEFAULT) METHOD

A histogram of a segment of the transformed image is performed to find the R_i and MIN_i for each component of the transformed data. If the user-defined field is smaller than 2000 pixels, all pixels are used in the histogram; otherwise, the following formula is used to determine the line and sample increments needed to obtain 2000 points for the histogram.

$$\alpha = \left(\frac{MN}{2000}\right)^{1/2} \qquad (13\text{-}3)$$

where

M = number of samples

N = number of lines

α = increment (integer)

In deriving an approximate range for the transformed data, the user may specify a percentage of points to be excluded from the upper and lower tails of the histogram by using the PEROUT control card. If not so specified, 2.5 percent of the points on the tails are excluded when determining the MAX_i and MIN_i values of the central 95 percent of the transformed data distribution.

Optionally, the user may specify the maximum expected data value for each channel n of the input data vector $\bar{x}$. Otherwise, the maximum data value for each channel is set equal to 255.

13.1.2 STATISTICAL METHOD

Activated by the RESCALE and MODULE or STATFILE control card, the statistical method is applied for deriving an approximate MAX_i and MIN_i value for each component i. Using the subclass statistics, an approximate R_i is computed using equations (13-4) and (13-5).

Let
$$\alpha_i = MAX_j\left(\hat{\beta}_i^j + k\hat{\sigma}_i^j\right) \tag{13-4}$$

and
$$\delta_i = MIN_j\left(\hat{\beta}_i^j - k\hat{\sigma}_i^j\right) \tag{13-5}$$

where

i = 1,···,m components of Z

j = 1,···,w subclasses

k = an integer specified by the user (see LAM control card)

$\hat{\beta}_i^j$ = transformed mean of the *ith* component of subclass j

$\hat{\sigma}_i^j$ = standard deviation of the *ith* component of subclass j computed from the transformed covariance matrix for subclass j.

The approximate range of each component will be

$$R_i = \alpha_i - \delta_i \quad ; \quad i = 1,\cdots,m \tag{13-6}$$

and
$$MIN_i = \delta_i \tag{13-7}$$

Letting the scale factor $\varepsilon_i = \frac{255}{R_i}$, i = 1,···,m, the complete transformation, including rescaling, to be performed on the original image (each pixel) is:

$$Y_i = \varepsilon_i \times \left|\delta_i - (BX + b_i)\right| \tag{13-8}$$

where

b_i = *ith* element of the bias vector b .

Optional control cards that may be used in conjunction with the statistical method are SUBCLA, LAM, PEROUT, OPTION ORIG, and OPTION TRANSF. Their functions, as well as those of other control cards, are described in table 13-1.

13.1.3 USER INPUT METHOD

The user may input his own scaling parameters via the OPTION SCAFAC control card or use input from a previous execution of DATA-TR in which the computed scaling parameters ($255/R_i$, MIN_i) were punched on cards via the OPTION PUNCH control card. When inputting the previously executed deck, the user should note that the values punched were computed using the MIN_i and MAX_i determined after a certain percentage of the tails of the histograms (see PEROUT control card) were discarded.

The transformed/rescaled data are output in either the Universal or LARSYS III format. The otpion is controlled by the FORMAT control card.

A line printer plot of the histogram (frequency distribution) of the transformed rescaled data is printed. If applicable, the MAX_i and MIN_i are printed.

13.2 INPUT FILES

An MSS data tape (DATAPE) must be input to the DATA-TR processor. The tape assignment defaults to logical unit 11; however, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information on format.)

13.3 OUTPUT FILES

The transformed/rescaled data are output on the TRFORM file, logical unit 14, in either the Universal or LARSYS III format.

13.4 SCRATCH FILES

The DATA-TR processor does not use scratch files.

13.5 CARD INPUT

All system card file input formats referred to in this section are defined in section 3.

13.5.1 PROCESSOR CARD

The keyword for the processor card is left justified beginning in column 1. For example,

$DATA-TR

This keyword directs the system monitor routine to select the DATA-TR processor and initiates loading of routines used by DATA-TR.

13.5.2 SPECIAL SYSTEM FILES

The B-matrix file discussed in section 3.1.4.2 must be input to this processor. The deck may be obtained from a previous execution of SELECT. The module STAT file (section 3.1.4.1) is optional input. If input, the second method for rescaling the input data (given in section 13.1.2) will be used.

13.5.3 CONTROL CARDS

Table 13-1 lists the control cards and available options for the DATA-TR processor.

13.5.4 FIELD DEFINITIONS

See section 3.1.3 for the format of field definition cards. At least one field definition card must immediately follow the *END* control card. An output file is created for each field definition input and is written on unit 14. Each of these fields consists of a rectangular field which surrounds the vertices of the input field. All pixels outside the input field and within the rectangular output field are set equal to zero. The line and sample number will be numbered sequentially from 1.

13.6 CARD OUTPUT

The DATA-TR processor, via the OPTION PUNCH control card, outputs the computed scaling parameters on cards. Two pairs of scaling parameters are punched on each card; i.e., each punched card contains the scaling parameters for two components of the transformed data. The cards must be used in the same order as punched. Their formats and definitions are as follows The number of cards is determined by the number of components.

Columns	Type/format	Definition
1-6	A6	OPTION
11-17	A7	SCAFAC=
18-27	A1,F9.3,F9.3,A1	(CON_1,MIN_1) where CON_1=255/R_1, R_1 is the range of component 1, and MIN_1=minimum value for component 1. Parentheses must be input.
28-37	A1,F9.3,F9.3,A1	(CON_2,MIN_2) where CON_2=255/R_2, R_2 is the range of component 2 and MIN_2=minimum value for component 2. Additional pairs are continued on succeeding cards.

13.7 RESTRICTIONS

The system-related restrictions in section 19 apply to this processor.

The maximum number of channels allowed is 30, and the maximum number of components in the transformed vector is 16.

13.8 DIAGNOSTIC MESSAGES

13.8.1 SUBROUTINE LNTRAN

	Message	Explanation
a.	*** THE NUMBER OF COMPONENTS IN Y-VECTOR TIMES THE NUMBER OF SAMPLES EXCEEDS THE SIZE OF STORAGE AREA — TERMINATE ***.	Self-explanatory.
b.	*** NUMBER OF CHANNELS TIMES NUMBER OF SAMPLES EXCEEDS 10 600.***	Storage exceeded. Adjust parameters.
c.	***** DATATR/LNTRAN ***** ERROR ON INPUT FIELD DEFINITION CARD, FOR FIELD NAME XXXX. ***** CONTINUING TO NEXT FIELD DEFINITION CARD(S).	Check format and parameters.

13.8.2 SUBROUTINE SETREM

Message	Explanation
SETREM ERROR - THERE WERE XX SCALE FACTORS AND MINIMUM VALUES INPUT THROUGH THE SCAFAC OPTION. YY LINEAR COMBINATIONS WERE REQUESTED.	This message indicates that the input scaling parameter pairs are not in one-to-one correspondence with the number of components of the transformed

Message	Explanation
THERE MUST BE A SCALE FACTOR AND A MINIMUM VALUE FOR EACH LINEAR COMBINATION. THE PROGRAM WILL TERMINATE THROUGH CMERR.	data. Too many or too few pairs were input.

13.8.3 SUBROUTINE SETUP8

	Message	Explanation
a.	*** BAD CONTROL CARD - DATATR/SETUP8 ***	Check spelling of keyword.
b.	***** DATATR/SETUP8 ***** ERROR ON INPUT DATAFILE CARD --- CONTINUING TO PROCESS INPUT *****	Check control card, correct, and resubmit.
c.	***** DATATR/SETUP8 ***** ERROR ON INPUT STATFILE CARD --- CONTINUING TO PROCESS INPUT *****	Check control card, correct, and resubmit.
d.	*** INVALID CONTROL CARD REJECTED BY DATATR/ SETUP8 ***	Check spelling of parameter.

TABLE 13-1.– DATA-TR PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
B-MATRIX	CARDS or FILE Default: None	CARDS indicates that the B-matrix is on cards immediately following. FILE indicates that the B-matrix is on file and initiates input of the BMFILE.
FORMAT	OUTPUT=UNIVERSAL Default: LARSYS III	The transformed data will be output in Universal format.
FORMAT	OUTPUT=LARSYS Default: LARSYS III	The transformed data will be output in LARSYS III format.
RESCALE	Blanks Default: No rescaling	Initiates rescaling of the transformed data to the range of 0 to 255.
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	11 is the Fortran default logical unit number to which the MSS data tape (DATAPE) has been assigned; M is the file number on the tape to be written.
STATFILE	UNIT=N,FILE=M Default: N=20,M=1	20 is the default Fortran logical unit number to which the SAVTAP file has been assigned; M is (1) the file number on the tape to be processed or (2), if a module STAT file is input, the number of the file

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 13-1.– Continued.

Keyword	Parameter and default values	Function
		on which to store the training statistics. If $M \neq 1$, this control card must precede the module STAT deck.
OPTION	PUNCH Default: No cards punched	Directs the program to punch the scaling parameters (CON_i, MIN_i) on cards.
OPTION	$SCAFAC=(CON_1,MIN_1),(CON_2,MIN_2),\cdots,(CON_i,MIN_i)$ Default: Histogram method of rescaling	CON and MIN are floating point values separated by a comma. Blanks between the two values are ignored. The scaling parameters should be ordered according to the transformed data vector components.
MODULE	FILE Default: If RESCALE is input, the histogram method is assumed.	Initiates reading of the SAVTAP file; if rescaling is performed, it initiates the statistical method.
MODULE	CARDS Default: If RESCALE is input, the histogram method is assumed.	Initiates reading of the module STAT file that must immediately follow this card; if rescaling is performed, it initiates the statistical method.
SUBCLASS	$S_1,S_2,S_3,\cdots,S_k$ $k \leq$ number of subclasses on SAVTAP≤ 60	Integers which define a subset of subclasses $S_1,S_2,S_3,\cdots$ from the input statistics

TABLE 13-1.– Continued.

Keyword	Parameter and default values	Function
	Default: Statistics for all subclasses defined are used in calculating the scaling factors.	(SAVTAP) to be used in calculating the scaling factors and approximating R_i.
LAM	N Default: N=2	An integer multiplied by the standard deviations of the input subclass statistics to derive an approximate range for rescaling the transformed data.
PEROUT	N Default: N=5, in which case 5% of the total distribution will be deleted from both the upper and lower tails of the transformed data set.	An integer which specifies the percentage of points to be deleted from the upper and lower tails of the transformed data distribution in computing an approximate range for rescaling. For the histogram method of rescaling, N/2% is deleted from both the upper and lower tails of the histogram. For the statistical method of rescaling, N% is deleted from both the upper and lower tails of the histogram.
MAXPT	$M_1,M_2,M_3,\cdots,M_k$ $k\leq 30$ Default: 255,255,···	Maximum expected value of MSS data tape (DATAPE) input for each channel. M's are integers

TABLE 13-1.– Continued.

Keyword	Parameter and default values	Function
		used in deriving an approximate range (MIN_i,MAX_i) of the transformed data set for the histogram method of rescaling.
OPTION	ORIG Default: No statistics printout	Initiates the printout of the original (untransformed) statistics for the subclasses input for the statistical rescaling method.
OPTION	TRANSF Default: No statistics printout	Initiates the printout of the transformed statistics.
BIAS	$b_1,b_2,b_3,\cdots,b_k$ or $N*b_1,b_{N+1},\cdots,b_k$ k=number of components in the transformed data set and N=an integer repetition factor for b_i $i \leq 16$ Default: b_i=0.0	All b's are decimal (floating point) numbers, separated by commas; they comprise the bias vector to be applied in the transformation of the input data set: $\vec{Z}=B(\vec{x}+b)$
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.

TABLE 13-1.– Concluded.

Keyword	Parameter and default values	Function
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of all card input for the processing function.

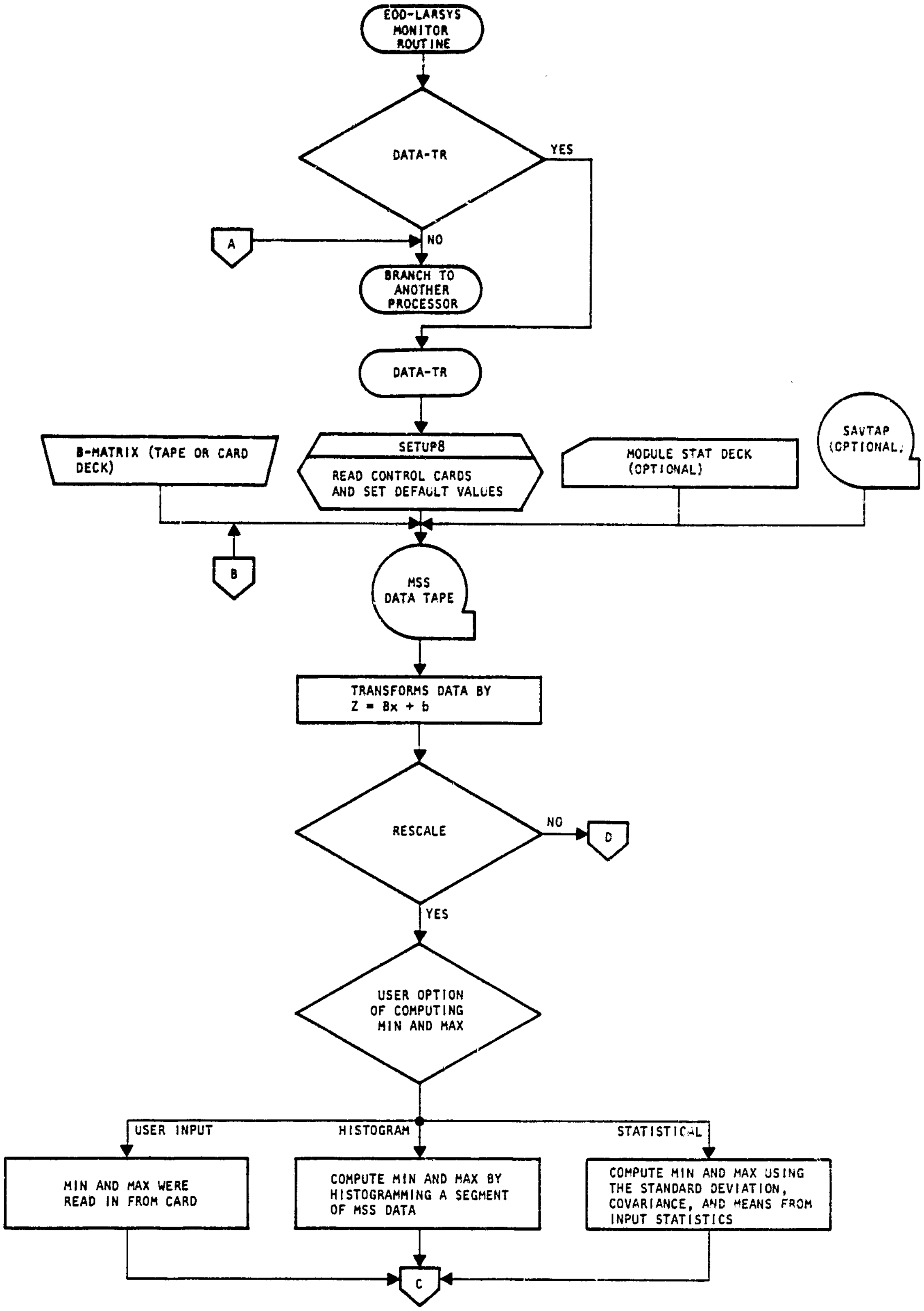


Figure 13-1.– Functional flow chart for the DATA-TR processor.

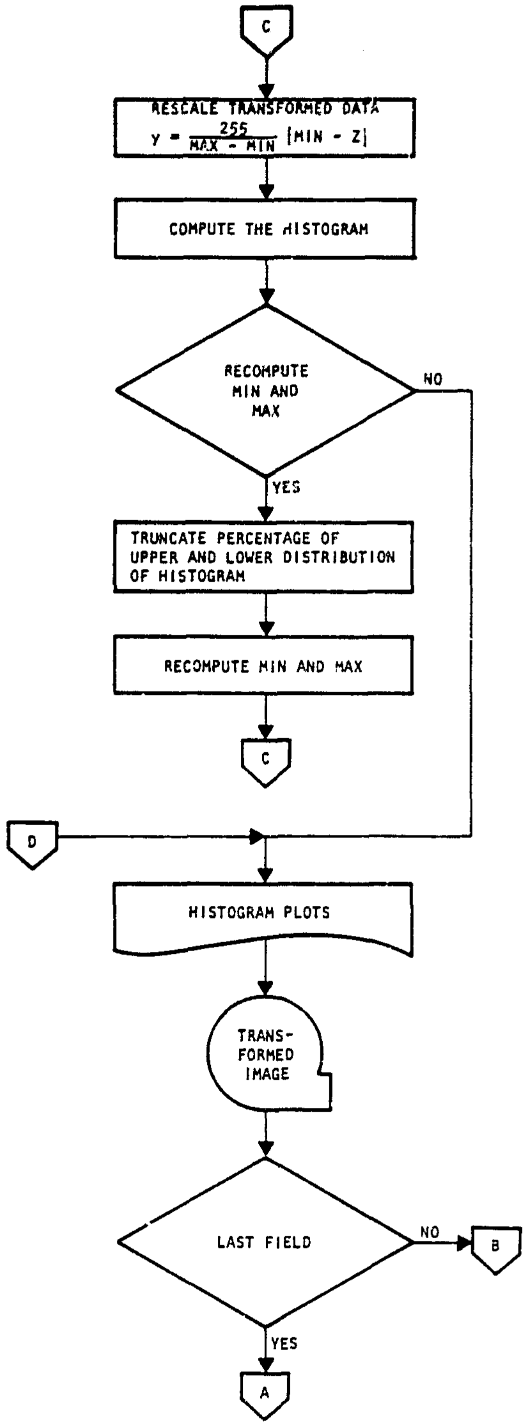


Figure 13-1.– Concluded.

14. STATISTICS TRANSFORMATION PROCESSOR – TRSTAT

The TRSTAT processor will read a SAVTAP file or card deck generated by STAT or ISOCLS, perform a linear transformation on the means and covariances, and output the transformed statistics on a new file. The equation for the linear transformation of the means is as follows:

$$\mu' = A_{\mu} + b \qquad (14\text{-}1)$$

where

A = a k-by-n matrix (see section 14.4.2); $k \leq 15$ and $n \leq 30$

μ = an n-by-1 mean vector

b = a k-by-1 bias vector (see card type 4, section 14.4.2)

μ' = a k-by-1 transformed mean vector

The equation for the linear transformation of the covariances is as follows:

$$K' = AKA^{T} \qquad (14\text{-}2)$$

where

K = an n-by-n covariance vector

A^{T} = an n-by-k transpose of A

K' = a k-by-k transformed covariance matrix

14.1 INPUT FILES

A set of statistics must be input either from the SAVTAP file or by cards. (See STATFILE or MODULE control card, table 14-1.)

14.2 OUTPUT FILES

The transformed statistics are output on a file in the SAVTAP format. (See section 4.1 and STATFILE control card.)

14.3 SCRATCH FILES

The TRSTAT processor uses no scratch files.

14.4 CARD INPUT

All system formats referred to in this section are defined in sections 3 and 14.4.2.

14.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$TRSTAT

This card directs the monitor routine to execute the TRSTAT processor and initiates loading of routines used by TRSTAT.

14.4.2 A-MATRIX FILE

The A-matrix file is composed of a transformation matrix and an additive bias vector. Its format is shown below. For additional information on the transformation matrix, see section 3.1.4.2.

Card type	Columns	Type/format	Definition
1	1–8	Alphanumeric	Keyword A-MATRIX.
2	6–7	Integer/I2	Number of linear combinations.
	13–14	Integer/I2	Number of channels.
	17–80	Integer/I2	Actual channels used in computation.[a]
3	6–20	Real/E15.8	Element 1
	21–35	Real/E15.8	Element 2
	⋮		⋮
	66–80	Real/E15.8	Element 5 of A-matrix — the full matrix is punched,

[a]The channels in the B-matrix described in section 3.1.4.2 begin in column 18.

Card type	Columns	Type/format	Definition
			5 values per card. Additional elements are continued on succeeding cards.
4	6–20	Real/E15.8	Element 1[a]
	21–35	Real/E15.8	Element 2
	⋮		⋮
	66–80	Real/E15.8	Element N of b-vector, 5 values per card, with additional values continued on succeeding cards; N=number of linear combinations.

14.4.3 CONTROL CARDS

Table 14-1 lists the control cards and available options for the TRSTAT processor.

14.4.4 FIELD DEFINITIONS

No field definition cards are input to TRSTAT.

14.5 CARD OUTPUT

The transformed statistics deck will be output in the same format as the module STAT card file.

14.6 RESTRICTIONS

The system-related restrictions in section 19 apply to this processor.

---

[a]Unlike the B-matrix described in section 3.1.4.2, the A-matrix file contains the additive vector.

The maximum dimension of the A-MATRIX is 15 by 30, and the maximum number of elements in the additive B-vector is 30.

14.7 DIAGNOSTIC MESSAGES

14.7.1 SUBROUTINE SETUP9

Message	Explanation
a. *** BAD SUPERVISOR CONTROL CARD SETUP9***	Invalid control card. Check spelling of keyword.
b. ERROR ON STATFILE CARD.	Check spelling of keyword and parameters.
c. NUMBER OF CHANNELS FROM STAT FILE DOES NOT EQUAL THE NUMBER OF CHANNELS ON A-MATRIX FILE. CHANNELS ON STAT FILE = ________. CHANNELS ON A-MATRIX = ____________.	Self-explanatory.
d. INVALID CONTROL CARD REJECTED *** SETUP9 ***	The parameter field of the control card is in error.

14.7.2 SUBROUTINE TRAMTX

Message	Explanation
ERROR IN TRYING TO POSITION TRANSFORMED STAT FILE TO BEGINNING OF FILE XXX.	Check file assignment.

TABLE 14-1.– TRSTAT PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
A-MATRIX	Blank Default: File must be input	Initiates input of the A-matrix and b-vector. The A-matrix card file (see section 14.4.2 for format) immediately follows this card.
OPTION	P,O,T	P punches the transformed statistics; O prints the original statistics; and T prints the transformed statistics.
MODULE	Blank	Initiates input of the module STAT card file, which immediately follows this card. (See section 3.1.4.1 for module STAT file format.)
STATFILE	INPUT/UNIT=N,FILE=M, OUTPUT/UNIT=L,FILE=S Default: N=1; M=1; L=14; S=1	N is the Fortran logical unit number to which the tape containing the statistics to be transformed has been assigned; M is the number of the file to be processed; L is the Fortran logical unit number to which the transformed statistics are to be output; and S is the number of the next file to be created on output SAVTAP file.

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 14-1.– Continued.

Keyword	Parameter and default values	Function
CHANNELS	$N_1, N_2, N_3, \cdots, N_k$ k=number of matrix channels≤30 Default: None	N's are integer channel numbers referring to the SAVTAP file. The number of channels requested from SAVTAP must be equal to the number of channels on the A-matrix file.
SUBCLASS	$S_1, S_2, S_3, \cdots, S_k$ k≤number of subclasses on SAVTAP≤60 Default: Statistics for all subclasses defined	Transforms statistics for only subclasses $S_1, S_2, S_3, \cdots$.
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.

TABLE 14-1.– Concluded.

Keyword	Parameter and default values	Function
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of all card input for the processing function.

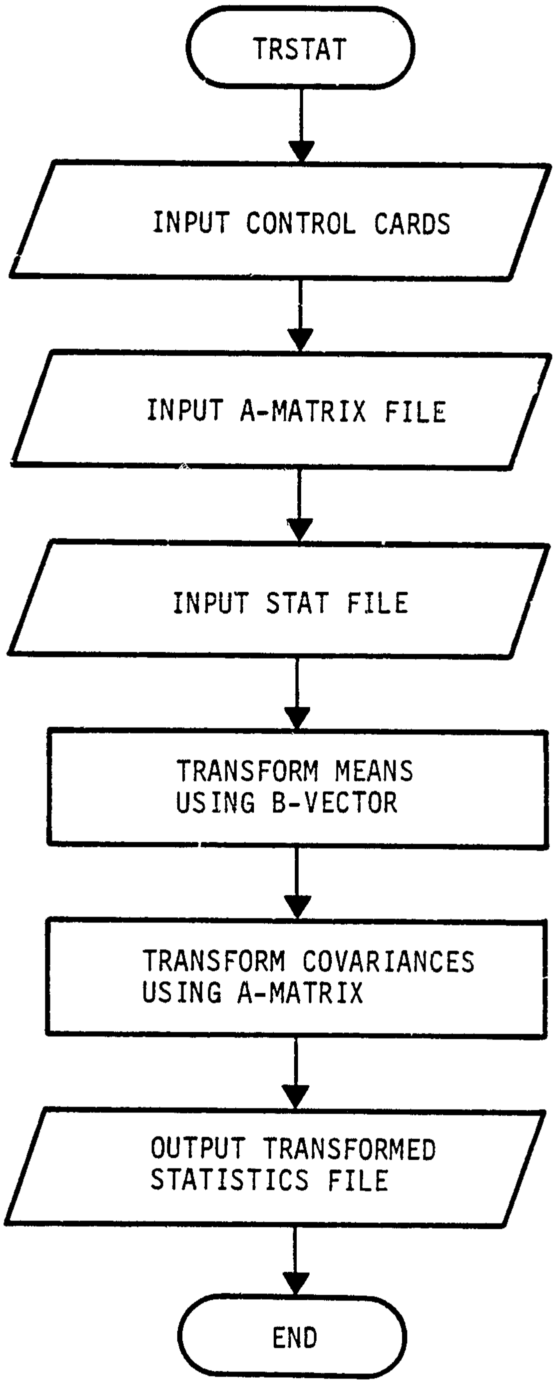


Figure 14-1.— Functional flow chart for the TRSTAT processor.

15. N-DIMENSIONAL HISTOGRAM PROCESSOR — NDHIST

The NDHIST processor computes an n-dimensional histogram of areas of the MSS data tape (DATAPE) of which the user has requested scatter plots. The pixel dimensions are user specified by the plotting channels, and the histogrammed pixels are output on the NHSTUN file. The file is written as an interface to the SCTPRL processor.

15.1 PROCEDURES

The number of channels (dimensions) used in histogramming is specified by means of the CHANNELS control card defined in table 15-1. The plotting channels are the primary input channels. The color channels are for further delineation of the frequency determination.

In the case of $n > 2$, the dimensionality is reduced to 2 in the SCTRPL processor by means of a linear transformation.

Composed of the plotting channels, each unique (positional) pixel within the field designates the position of a pixel on the scatter plot tape, SCTRUN. The frequency of each positional pixel is determined as a function of the color channels (if input) and the plotting channels.

If only plotting channels are input, the positional pixel and its frequency are calculated using plotting channel data. If both plotting and color channels are input, the frequency is a function of both sets of channels.

The color assignment for each plot pixel may be optionally set by the NDHIST or the SCTRPL processor. If applicable, the color codes are output on the NHSTUN file. The color codes may be set using the following information.

a. The original radiance value of the pixel (see CHANNELS control card, table 15-1).

b. The mean value of the cluster or subclass to which the pixel was assigned during clustering or classification. In exercising this option, the user must input a classification or cluster map (see MAPUNT control card, table 15-1) to this processor. To execute the SCTRPL processor, a SAVTAP file related to the MAPUNT must be input (see CHANNELS and STATFILE control cards, section 16, table 16-1). The subclass or cluster numbers assigned to the pixel during classification or clustering are stored on the NHSTUN file, passed to the SCTRPL processor, and used for retrieving the means from the SAVTAP file.

c. The mean value of the test or training field from which the pixel was extracted.

d. User-defined colors (see COLOR control card, table 16-1).

e. From any pass on the MSS data tape when using multiregistered Landsat data (see CHANNELS control card, table 15-1).

The areas selected for histogramming are defined by test and/or training fields. The manner in which the fields are collected or grouped for histogramming is user controlled by input parameters. The data vectors may be histogrammed collectively on a class, subclass, or per-field level. The maximum number of fields input on any level is 200, and the maximum number of unique data vectors accumulated on any level is 12 000 divided by one-fourth the number of plotting channels.

A functional flow diagram of the NDHIST processor is given in figure 15-1.

15.2 INPUT FILES

An MSS data tape (DATAPE) must be input to the NDHIST processor. The tape assignment defaults to logical unit 11; however, by input of the DATAFILE control card, the user may assign any available logical unit. (See table 5-1 for file assignments and section 3.2, MSS Image Data Tapes, for further information.) Optionally, a classification or cluster MAPUNT tape may be input (see MAPUNT control card).

15.3 OUTPUT FILES

A multifile tape is always output: NHSTUN file. It is an interface to the SCTRPL processor and must be assigned to tape or Fastrand. No file-skipping capability is available; the first file created is always file 1. (See HISFIL control card, table 15-1, and appendix I.)

15.4 SCRATCH FILES

The NDHIST processor dynamically assigns random access drum storage for the histogram counters, color codes, identification information, and (optionally) the pixel assignment from the classified or clustered image tape (MAPUNT).

15.5 CARD INPUT

15.5.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$NDHIST

This card directs the system monitor routine to select the NDHIST processor and initiates loading of all the NDHIST routines into the system.

15.5.2 SYSTEM CARD FILES

No special system card decks are required for the NDHIST processor.

15.5.3 CONTROL CARDS

Table 15-1 lists the control cards and available options for the NDHIST processor.

15.5.4 FIELD DEFINITIONS

The field cards, which immediately follow the *END control card, define the areas to be histogrammed, and the OPTION control card determines the level of histogramming. The fields may be ordered in one of four ways:

a. As input to STAT (section 8.4.4)

b. As input to ISOCLS (section 9.5.4)

c. As input to CLASSIFY (section 11.5.4)

d. As a user-defined field (section 3.1.3)

For example:

```
*END
CLASSNAME    WHT
SUBCLASS     WHT1
(Field card 1)
(Field card 2)
SUBCLASS     WHT2
(Field card 3)
CLASSNAME    NWHT
SUBCLASS     NWH1
(Field card 4)
SUBCLASS     NWH2
(Field card 5)
SUBCLASS     NWH3
(Field card 6)
(Field card 7)
$END
```

If the histogram is accumulated on class bases, fields 1, 2, and 3 are histogrammed collectively and output as data file 1; and fields 4, 5, 6, and 7 are histogrammed collectively and output as data file 2.

If the histogram is accumulated on subclass bases, fields 1 and 2 are histogrammed collectively and output as data file 1; field 3 is histogrammed and output as data file 2; field 4 is histogrammed and output as data file 3; field 5 is histogrammed and output as data file 4; and fields 6 and 7 are histogrammed collectively and output as data file 5.

If the histogram is performed on per-field bases, each field is histogrammed separately and output to a file, making a total of seven data files created.

On a cumulative histogram, a maximum of 200 fields may be input.

See section 3.1.3 for format of the field definition card.

15.6 CARD OUTPUT

The NDHIST processor does not provide punched card output.

15.7 RESTRICTIONS

The system-related restrictions in section 19 apply to this processor. Other restrictions are as follows.

- A maximum of 16 channels may be histogrammed.
- A maximum of 4 channels may be used for color codes.
- The maximum of unique vectors to be histogrammed is

$$n \leq \frac{12\ 000}{1/4(\text{number of channels})} \tag{15-1}$$

- A maximum of 4000 words of storage is allowed for storing the MSS data. The equation for computing maximum number of pixels is

$$n \leq \frac{4000}{[\text{number of channels}(\text{number of samples per scan line})]} \qquad (15\text{-}2)$$

15.8 DIAGNOSTIC MESSAGES

15.8.1 SUBROUTINE ADDRES

Message	Explanation
TOO MUCH DATA REQUESTED. REDUCE NO. OF SAMPLES PER SCAN LINE AND/OR NO. OF CHANNELS.	Self-explanatory.

15.8.2 SUBROUTINE NDHST1

Message	Explanation
a. -- VECTORS WERE NOT HISTOGRAMMED, BUT USED IN COMPUTING FIELD MEANS, IF APPLICABLE.	The histogrammed vector table is full. N number of unique vectors were not histogrammed.
b. ERROR IN FIELD CARD. ABORTING.	Check format and parameters.

15.8.3 SUBROUTINE NDHST2

Message	Explanation
CORE LIMITS EXCEEDED. MAXIMUM NO. OF VECTORS ACCEPTED IS ___.	Self-explanatory.

15.8.4 SUBROUTINE RESTO

Message	Explanation
ERROR READING DISK.	Self-explanatory.

15.8.5 SUBROUTINE SET10

Message	Explanation
a. INVALID CONTROL CARD - IGNORED.	Check spelling of keyword.
b. ERROR ON CHANNELS CARD.	Check parameter field of CHANNELS control card.
c. ERROR ON DATA FILE CARD.	Check parameter field of DATAFILE control card.
d. ERROR ON DAS FILE CARD.	Check parameter field of DAS control card.
e. ERROR ON N-DIM HISTOGRAM FILE CARD.	Check parameter field of HISFIL control card.
f. ERROR ON OPTION CARD.	Check parameter field of OPTION control card.

15.8.6 SUBROUTINE STODAT

Message	Explanation
NOT ENOUGH DISK SPACE TO STORE DAS TAPE DATA.	Self-explanatory.

TABLE 15-1.– NDHIST PROCESSOR OPTIONS AND CONTROL CARDS

[All m and n are integers.]

Keyword (a)	Parameter and default values (b)	Function
CHANNELS	PLOT=$n_1, n_2, n_3, \cdots, n_i$, COLOR=$m_1, m_2, m_3, \cdots, m_j$ Default: None	The n's are the channels for determining the position (PLOT) of the pixels to be output on the SCTRUN tape ($i \leq 16$). If $i=2$, n_1 defines the sample location and n_2 the line location on the scatter plot tape (SCTRUN). If $i>2$, the pixels will be transformed to two components in the SCTRPL processor; component 1 will define the sample location and component 2 the line location on the scatter plot tape (SCTRUN). The m's represent the channels for the color codes ($j \leq 4$). If the COLOR channels are input, the histogram is a function of both the PLOT and COLOR channels; if the COLOR channels are omitted the histogram is a function of only the PLOT channels. (See section 16 for further information.)

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 15-1.– Continued.

Keyword	Parameter and default values	Function
DATAFILE	UNIT=N,FILE=M Default: N=11,M=1	N is the logical unit number assigned to the MSS data tape (DATAPE); M is the file number of the data to be processed.
MAPUNT	UNIT=N,FILE=M Default: None	N is the logical unit number assigned to the MAPUNT tape; M is the file number of the data to be processed. (The order of the fields to be histogrammed must correspond to the order of the clustered or classified fields on the input MAPUNT tape.)
HISFIL	UNIT=N Default: N=4	The logical unit number assigned to the NHSTUN file.
OPTION	CLASS Default: Field bases	Fields will be histogrammed on basis of classes.
OPTION	SUBCLS Default: Field bases	Fields will be histogrammed on basis of subclasses.
OPTION	FIELD Default: Field bases	Fields will be histogrammed on per-field basis.
OPTION	MEANS	The means of each field will be computed for the COLOR channels on the CHANNELS card and output on the NHSTUN file.

TABLE 15-1.– Concluded.

Keyword	Parameter and default values	Function
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
*END	Blank	Signals the end of the control cards.
$END	Blank	Signals the end of all card input for the processing function.

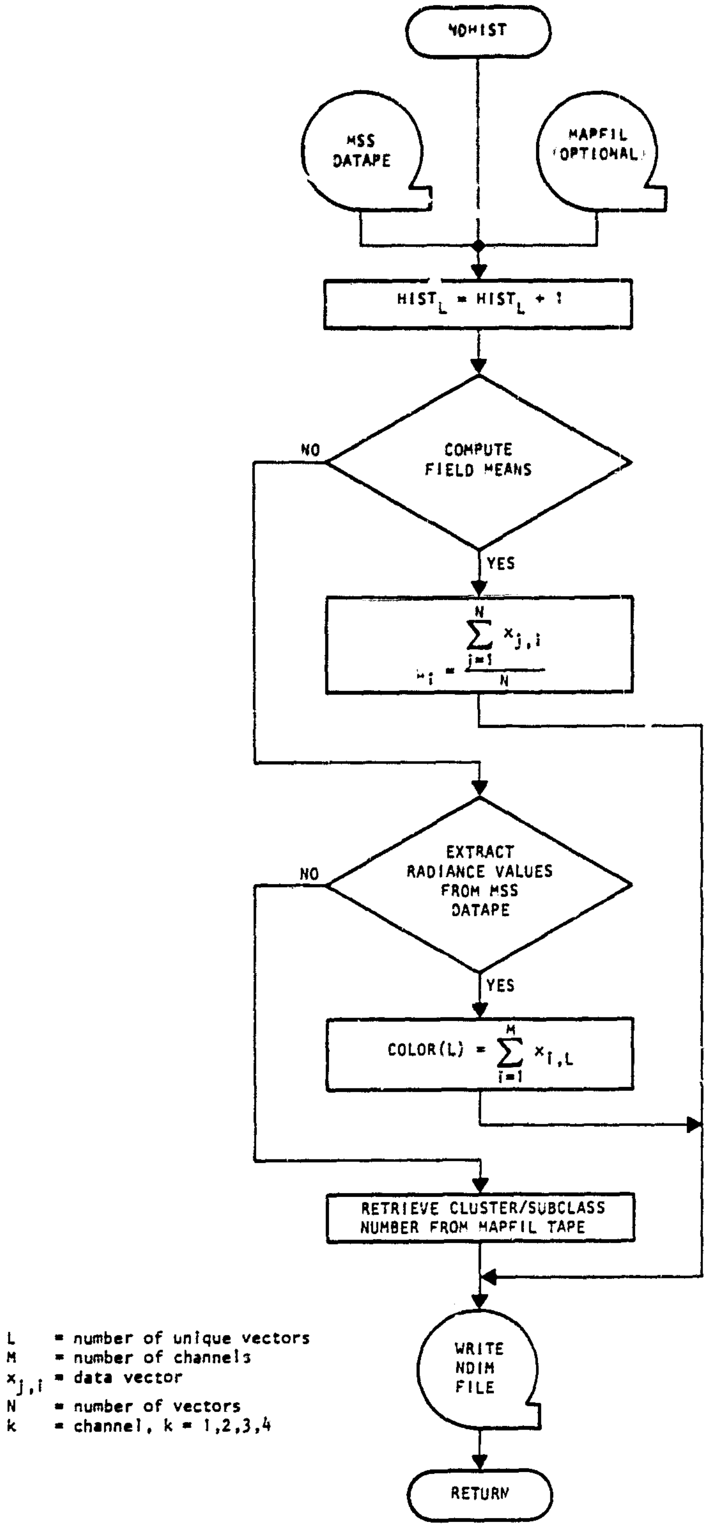


Figure 15-1.— Functional flow chart for the NDHIST processor.

16. SCATTER PLOT PROCESSOR – SCTRPL

The SCTRPL processor reads the NHSTUN file written by the NDHIST processor, determines the location of each unique data vector on the scatter plot, and outputs the spectral plot in the Universal format. A scatter plot is created and output for each file stored on the NHSTUN unit.

The location (line and sample intersection) of each pixel on the two-axis scatter plot will be computed using either the radiance values or two linear combinations of radiance values. (This option is controlled in section 15 by the CHANNELS control card, table 15-1.)

If the data vector is to be transformed (see B-MATRIX control card, table 16-1), the following equation and conditions will be applied:

$$y = Bx + c \tag{16-1}$$

where

y = a 2-by-1 vector

B = a 2-by-n vector, $n \leq 16$

x = an n-by-1 vector, $n \leq 16$

c = a 2-by-1 vector

If the transformed data are to be rescaled (see SCALE control card, table 16-1), the following equation will be applied:

$$y_i = \left(\frac{HI_i - LO_i}{R_i}\right) \times |MIN_i - Z_i| \tag{16-2}$$

where

HI_i = an input parameter for the upper rescale limit for channel i

LO_i = an input parameter for the lower rescale limit for channel i

MIN_i = minimum value for channel i

Z_i = transformed data point for channel i

R_i = range (MAX_i - MIN_i) for channel i

Y_i = rescaled transformed data value for channel i

MAX_i = maximum value for channel i

The scatter plot is created and output line by line. All the pixels belonging to a line, as determined by the second coordinate of the pixel, are collected; and, in the sample location determined by the first coordinate, the color assignment and frequency of occurrence of each pixel are output as channels 1 through n, respectively. [See procedures for NDHIST processor (section 15.1) for definition of color assignments.]

The dimensions of the output file are user controlled by the following input control cards, which are defined in greater detail in table 16-1 (the control cards for the SCTRPL processor) and table 15-1 (the control cards for the NDHIST processor).

Keyword	Parameters
SIZE	XLOW=0,XHIGH=128,XSIZ=129,YLOW=0, YHIGH=64,YSIZ=65
CHANNELS	PLOT=3,4,COLOR=5,6,7,8

The output file will contain:

a. 129 samples per scan line with a maximum data resolution of 128 on channel 3.

b. 65 lines per file with a maximum data resolution of 64 on channel 4.

c. 5 channels with channels 1 through 4 containing the color pixel (determined by channels 5, 6, 7, and 8 on CHANNELS control card) and channel 5 the frequency.

d. A color pixel for each positional pixel determined by channels 3 and 4 on the CHANNELS control card.

If a MAPUNT file containing the subclass or cluster numbers is input to the NDHIST processor, either a SAVTAP file related to the MAPUNT file must be input (see STATFILE control card) or the user must input the color codes on cards (see COLOR control card).

Optionally, a line printer pixel frequency scatter plot will be output (see PIXPLT control card). The frequency of occurrence or log of frequency of occurrence will be represented by a symbol (see SYMBOL control card). The location of the symbol on the plot will be determined by the radiance value of the pixel. If the data have been transformed, then the data must be rescaled to exercise this option.

A functional flow diagram of the SCTRPL processor is given in figure 16-1.

16.1 INPUT FILES

The NHSTUN file created by NDHIST must be input. (See the HISFIL control card and appendix E for format of the NHSTUN tape.)

The SAVTAP file created by the STAT or ISOCLS processor optionally may be input. (See the STATFILE control card and section 4.1 for a description of the file.)

16.2 OUTPUT FILES

A multifile Universal-formatted tape containing the spectral plots and color keys, when applicable, will be output. (See the SCTRUN control card and appendix G for tape format.)

16.3 SCRATCH FILES

The program dynamically assigns random access disk storage for scratch files.

16.4 CARD INPUT

16.4.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1. For example,

$SCTRPL

This card directs the system monitor routine to select the SCTRPL processor and initiates loading of all the SCTRPL routines into the system.

16.4.2 SYSTEM CARD FILES

The module STAT and B-matrix card files may be input. See section 3 for formats.

16.4.3 CONTROL CARDS

Table 16-1 lists the options and control cards for the SCTRPL processor.

16.4.4 FIELD DEFINITIONS

Field definitions do not apply to this processor.

16.5 CARD OUTPUT

The SCTRPL processor does not provide punched card output.

16.6 RESTRICTIONS

The system-related restrictions in section 19 apply to this processor.

If the color codes for the scatter plot tape SCTRUN are to be principal component (PC) colors, the user must ensure that the values are positive.

The maximum dimension of the B-matrix is 2 by 16; the maximum number of elements in additive vector b is 16.

The maximum number of channels on the output tape SCTRUN is 5. Color codes are the first n-1 channels; the frequency is the *nth* channel.

The maximum number of channels selected from the SAVTAP file is 4.

The maximum size of the output tape SCTRUN is 200 samples per scan line and 200 lines.

16.7 DIAGNOSTIC MESSAGES

16.7.1 SUBROUTINE LINPLT

Message	Explanation
A TOTAL OF ______ POINTS WERE NOT DISPLAYED ON THE LINE PRINTER GRAPH. THE POINTS WERE OUT OF RANGE IN EITHER THE X DIRECTION OR Y DIRECTION.	Data may be rescaled to a resolution of 100.

16.7.2 SUBROUTINE SETADR

Message	Explanation
NOT ENOUGH DISK SPACE.	Adjust parameters.
TOTAL WORDS OF DISK SPACE = XXXXXXXXXXXX.	
TOTAL WORDS OF DISK SPACE = XXXXXXXXXXXX.	

16.7.3 SUBROUTINE SET11

	Message	Explanation
a.	INVALID CONTROL CARD -- IGNORED.	Check spelling of keyword.
b.	ERROR ON CHANNELS CARD.	Check parameter field of CHANNELS card.
c.	ERROR ON STATFILE CARD.	Check parameter field of STATFILE card.
d.	ERROR ON NHSTUN HISTOGRAM FILE CARD.	Check parameter field of HISFIL card.
e.	ERROR ON OPTION CARD.	Check parameter field of OPTION card.
f.	ERROR ON OPTION CARD.	Check format and parameter.
g.	ERROR ON TAPE SIZE CARD.	Check parameter field of SIZE card.
h.	ERROR ON SCATTER PLOT TAPE CARD.	Check parameter field of SCTRUN card.
i.	ERROR ON B-MATRIX CARD.	Check spelling and parameter.
j.	ERROR ON SCALING CARD.	Check parameter field of SCALE card.

k. DATA MUST BE RESCALED BEFORE PIXEL FREQUENCEY PLOT OPTION MAY BE SELECTED.	Transformed data must be rescaled for line printer plot.
l. NO. OF PLOTTING CHANNELS NO. OF B-MATRIX CHANNELS MUST BE EQUAL. CHANNELS ARE ________, RESPECTIVELY.	Number of channels to be transformed must equal the number of channels in transformed matrix.

Message	Explanation
m. ERROR IN POSITIONING NHSTUN FILE TO FILE ________.	Physical tape error occurred. Resubmit run.

16.7.4 SUBROUTINE SORTVC

Message	Explanation
ERROR IN SORTING VECTORS.	Self-explanatory.

16.7.5 SUBROUTINE VECSCN

Message	Explanation
ERROR OCCURRED SCANNING VECTOR CARD.	Check keypunching error on cards.

TABLE 16-1.— SCTRPL PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNELS	$n_1, n_2, n_3, \cdots, n_i$ $i \leq$ number of channels on SAVTAP ≤ 30 Default: First 4 channels from NDIM file	Statistics for these channels will be extracted from the SAVTAP file; n_i must be a subset of channels on the SAVTAP file.
STATFILE	UNIT=N,FILE=M Default: None	N is the logical unit number assigned to the SAVTAP file; M is the number of the file to be processed.
HISFIL	UNIT=N Default: N=13	N is the logical unit number assigned to the NHSTUN file.
PIXPLT	LOG	Line printer pixel scatter plot of the log of frequency of occurrence will be printed.
PIXPLT	FREQ	Line printer pixel scatter plot of the frequency of occurrence will be printed.
PIXPLT	RESCALE Default: No rescaling. XSIZ=101, YSIZ=101; the range for x-axis is XLOW+XSIZ-1; the range for y-axis is YLOW+YSIZ-1.	The frequency of occurrence of the pixel for the line printer scatter plot will be rescaled to ranges XLOW, XHIGH, YLOW, and YHIGH. XSIZ will determine the number of bins on the x-axis; YSIZ, the

[a]The keyword must be left justified in card columns 1 through 10.

[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 16-1.– Continued.

Keyword	Parameter and default values	Function
		number of bins on the y-axis. (See SIZE control cards.)
COLOR	$(m_1),(m_2),\cdots,(m_p)$ or $L^*(m_1)$, $K^*(m_{L+1})$ L and K are integer repetition factors. Default: No user input of color codes	$m_1=n_1,n_2,\cdots,n_\ell$ is the color assignment for cluster 1; $m_2=n_1,n_2,\cdots,n_\ell$ is the color assignment for cluster 2; $m_p=n_1,n_2,\cdots,n_\ell$ is the color assignment for cluster n. $p\leq 60$ and $\ell\leq 4$; $0\leq n_i\leq 255$.
SIZE	XSIZ=N Default: XSIZ=101	The number of samples per line to output on the scatter plot tape; $N\leq 200$.
SIZE	YSIZ=N Default: YSIZ=101	The number of lines to output on the scatter plot tape; $N\leq 200$.
SIZE	XHIGH=N Default: XHIGH=100	The upper limit of the radiance value for the sample axis (x-axis); $N\leq 255$.
SIZE	XLOW=N Default: XLOW=0	The lower limit of the radiance value for the sample axis (x-axis); $0\leq N\leq$ XHIGH.
SIZE	YHIGH=N Default: YHIGH=100	The upper limit of the radiance value for the line axis (y-axis); $N\leq 255$.
SIZE	YLOW=N Default: YLOW=0	The lower limit of the radiance value for the line axis (y-axis); $0\leq N\leq 255$.

TABLE 16-1.– Continued.

Keyword	Parameter and default values	Function
PLOTAP	UNIT=N Default: N=12	N is the logical unit number assigned to the spectral plot tape.
B-MATRIX	CARDS Default: None	The B-matrix is being input by cards.
B-MATRIX	FILE Default: None	The B-matrix is being input by file.
BVEC	$T_1, T_2, \cdots, T_n$ n=number of linear combinations in B-matrix$\leq$2 Default: T_n=0.0	Elements of the additive vector to be used in the transformation; T is a floating-point number.
BCKGND	N Default: N=255	If N=0, background will be black; if N=255, background will be white.
SCALE	FILE	The scale factors will be computed from the NHSTUN file.
SCALE	XMAX=T Default: XMAX will be computed from the NDIM file.[c]	The upper range for the transformation of the sample values (x-axis); T is a floating-point number.
SCALE	XMIN=T Default: XMIN will be computed from the NDIM file.[c]	The lower range for the transformation of the sample values (x-axis); T is a floating-point number.

[c]If one of the parameters XMIN, XMAX, YMIN, or YMAX is input, all four parameters must be input.

TABLE 16-1.– Continued.

Keyword	Parameter and default values	Function
SCALE	YMAX=T Default: YMAX will be computed from the NHSTUN file.[d]	The upper range for the transformation of the line values (y-axis); T is a floating-point number.
SCALE	YMIN=T Default: YMIN will be computed from the NHSTUN file.[d]	The lower range for the transformation of the line values (y-axis); T is a floating-point number.
SCALE	RESCALE Default: No rescaling of the transformed data.	The transformed data will be rescaled to the range of XLOW, XHIGH, YLOW, and YHIGH. (See SIZE control card.)
HED1	Any 60 characters beginning in column 11 Default: LYNDON B. JOHNSON SPACE CENTER	Replaces first header line with the indicated characters in the parameter field.
HED2	Any 60 characters beginning in column 11 Default: HOUSTON, TEXAS	Replaces second header line with the indicated characters in the parameter field.
DATE	Any 12 characters beginning in column 11 Default: Current date	Prints the indicated 12 characters in the right corner of the heading in place of the current date.

[d] If one of the parameters XMIN, XMAX, YMIN, or YMAX is input, all four parameters must be input.

TABLE 16-1.- Concluded.

Keyword	Parameter and default values	Function
SYMBOLS	$S_1, S_2, S_3, \cdots, S_k$ $k \leq 32$ Default: Two sets of 10 symbols overprinted	Character set separated by commas, with a maximum of 32 characters. The number of symbols/2 determines the number of bin levels. The first set of symbols is overprinted by the second set. A blank is not a legitimate character.
MODULE	Blank	Initiates the input of the module STAT card file which immediately follows this card (see section 3.1.4.1 for format).
COMMENT	Any 60 characters beginning in column 11 Default: No comments printed	Prints a comment line using the 60 characters found in the parameter field.
END	Blank	Signals the end of the control cards.
$END*	Blank	Signals the end of all card input for the processing function.

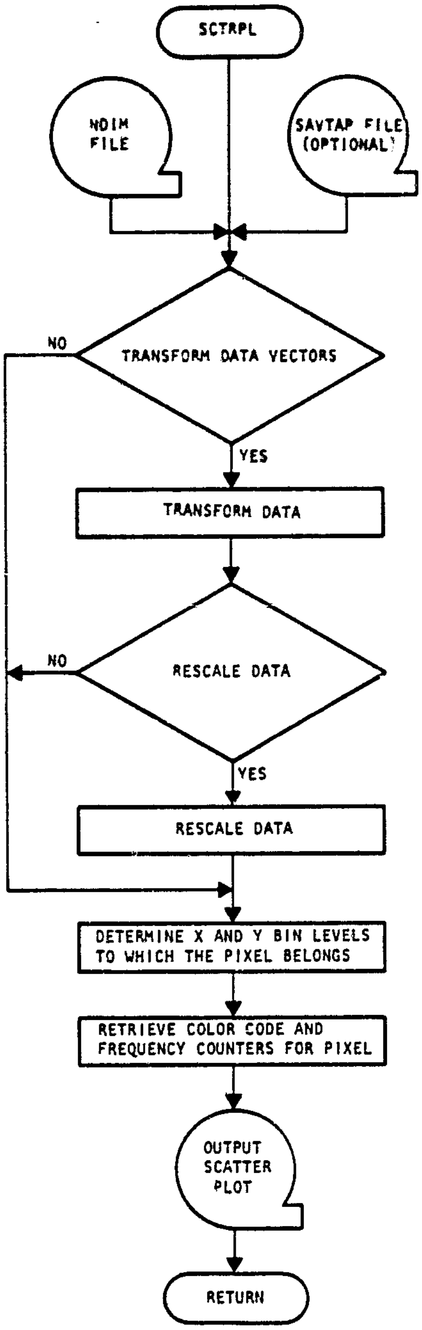


Figure 16-1.– Functional flow chart for the SCTRPL processor.

17. DOT DATA PROCESSOR — DOTDATA

17.1 PROCEDURES

In implementing Procedure 1, a method which allows the user the capability to label certain MSS data points corresponding to pixels (known as dots) was added to the system by the processor DOTDATA. The main function of this processor is to output a file containing the dots of interest. This file is an interface for three processors, ISOCLS, LABEL, and DISPLAY.

The dots are defined by field cards. Any subset of the 209 possible grid points may be selected by the user. Optionally, the dots may be labeled at the initial creation of the file or in the LABEL processor. All the categories of interest do not have to be defined in this processor.

If a dot is to be retyped, the file must be re-created by executing this processor again.

The files contain both type 1 and type 2 dots. Type 1 dots (starting and labeling dots) are written on one file; type 2 dots (bias correction dots) are written on a second file.

By an OPTION control card, the user may request that the spatial and spectral information relating to each dot on the file be printed on the line printer.

17.2 INPUT FILES

This processor requires an MSS data tape (DATAPE). The tape assignment defaults to logical unit 11; but by input of the DATAFILE control card, the user may assign any available logical unit. (See section 3.2 for further information on format.)

17.3 OUTPUT FILES

DOTUNT, a multifile unformatted FORTRAN written tape, is output as an interface to the processors ISOCLS, LABEL, and DISPLAY. The default unit for DOTUNT is logical unit 19. (See appendix H for format of the tape.) The logical unit and file number can be controlled by the DOTUNT control card.

17.4 SCRATCH FILES

The DOTDATA processor does not require an additional scratch file.

17.5 CARD INPUT

The processor keyword is left justified, beginning in column 1: $DOTDATA. This card directs the system monitor volume to select the DOTDATA processor and causes all the routines used by it to be loaded into the system.

17.5.1 SYSTEM CARD FILES

The DOTDATA processor does not use any special input files.

17.5.2 CONTROL CARDS

Control cards allow the user to input various options. These cards are identified by a keyword left justified in columns 1 through 10 of the card, with the parameter values or additional keywords in columns 11 through 72 (beginning in any column after column 10). These control cards may be in any order, but they must be the first cards after the processor card $HIST. Table 17-1 lists all available options, along with their default values.

17.5.3 FIELD DEFINITIONS

The user defines the grid points to extract from the MSS data tape by the field card. The definition and order of the field card(s) determines the position of the dots on the dot data file, DOTUNT. The analyst will need to know the position of the dots for defining starting vectors in ISOCLS and for labeling or relabeling the dots in LABEL.

At the time of defining the fields, the type for each dot is defined by the TYPE card. By option, the analyst may label each dot by a CLASSNAME card. If this card is omitted, the unlabeled dots must be labeled by the control card DOTLABEL or excluded from the set by the control card EXCLUDE in the labeling processor, LABEL.

An example of a field data set which is expected by this processor follows.

```
*END
TYPE        1
CLASSNAME   WHEA (optional)
LAB1        (10,10),(10,20),(196,10)
LAB2        (10,10),(10,20),(196,20)
CLASSNAME   NONW (optional)
LAB3        (10,10),(10,50),(100,50)
TYPE        2
CLASSNAME   WHEA (optional)
BIA1        (10,10),(10,40),(196,40)
CLASSNAME   NONW (optional)
BIA2        (10,10),(10,70),(196,70)
$END
```

Two files are written. File 1 contains 38 WHEA dots, followed by 10 NONW dots; all of which are type 1 dots. File 2 contains 19 WHEA dots, followed by 19 NONW dots; all of which are type 2 dots.

Note: All names on CLASSNAME cards are read from col. 11 through 15.

If the CLASSNAME cards were omitted, file 1 would contain 48 unlabeled type 1 dots. File 2 would contain 19 unlabeled type 2 dots.

In both cases, the reference number for the dots in file 1 defined by LAB1 field card is 1 through 19, the LAB2 field card reference number is 20 through 38, and LAB3 field card is 39 through 48. The reference number for the dots in file 2, defined by BIA1 field card, is 1 through 19, and for BIA2 field card, 20 through 38.

If the LACIE dot input option is invoked, there are no TYPE or CLASSNAME card images. Each input card image has the form

$$\text{DOT} \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} \begin{Bmatrix} A \\ B \\ \vdots \\ Z \end{Bmatrix} n_1, n_2, \ldots, n_N$$

DOT begins in column 1. The dot type (1 or 2) appears in column 5. A one-character category name appears in column 7.

17.6 CARD OUTPUT

DOTDATA does not produce any card decks.

17.7 RESTRICTIONS

System restrictions apply to DOTDATA.

17.8 DIAGNOSTIC MESSAGES

17.8.1 SUBROUTINE DOTS

Message	Explanation
**** NOTE - TOTVEC WAS GREATER THAN 250, THEREFORE TOTVEC WAS SET TO 250 ****.	Total number of dots allowable is 250 of Type 1 and 250 of Type 2.

17.8.2 SUBROUTINE SET13

	Message	Explanation
a.	INVALID CONTROL CARD — IGNORED.	Check spelling of keyword.
b.	ERROR ON DATA CARD.	
c.	ERROR ON DATAFILE CARD.	Check parameter field.
d.	ERROR ON DOTFILE CARD.	Check parameters.
e.	ERROR ON OPTION CARD.	Check format and parameters.

TABLE 17-1.— DOTDATA PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNEL	DATA=$n_1, n_2, \cdots, n_{30}$ Default: None	Integer numbers separated by commas referring to the channels on the MSS data tape.
DATAFILE	UNIT=n,FILE=m Default: n=11,m=1	n is the Fortran unit number assigned to the MSS data tape; m is the file number of the data to process.
DOTFIL	OUTPUT/UNIT=n,FILE=m Default: n=19,m=1	n is the Fortran unit number assigned to the dot data file output by this processor; m is the number of the file to output.
OPTION	PRINT Default: no line printer output.	Initiates the printing of the dot data file information.
OPTION	LACIE	Enables user to input dots according to the LACIE format.
HED1	Any 60 characters Default: LYNDON B. JOHNSON SPACE CENTER	First line of the heading on line printer output.
HED2	Any 60 characters Default: HOUSTON, TEXAS	Second line of the heading on line printer output.

[a]The keyword must begin in column 1.

[b]The parameter data can extend from column 11 to column 72.

TABLE 17-1.– Concluded.

Keyword (a)	Parameter and default values (b)	Function
DATE	Any 12 characters Default: Present date	Date in the heading on line printer output.
COMMENT	Any 60 characters Default: None	Comment printed with heading on line printer output.
*END	Blank	Indicates the end of the control card inputs.
$END	Blank	Indicates the end of all the card inputs.

[a]The keyword must begin in column 1.

[b]The parameter data can extend from column 11 to column 72.

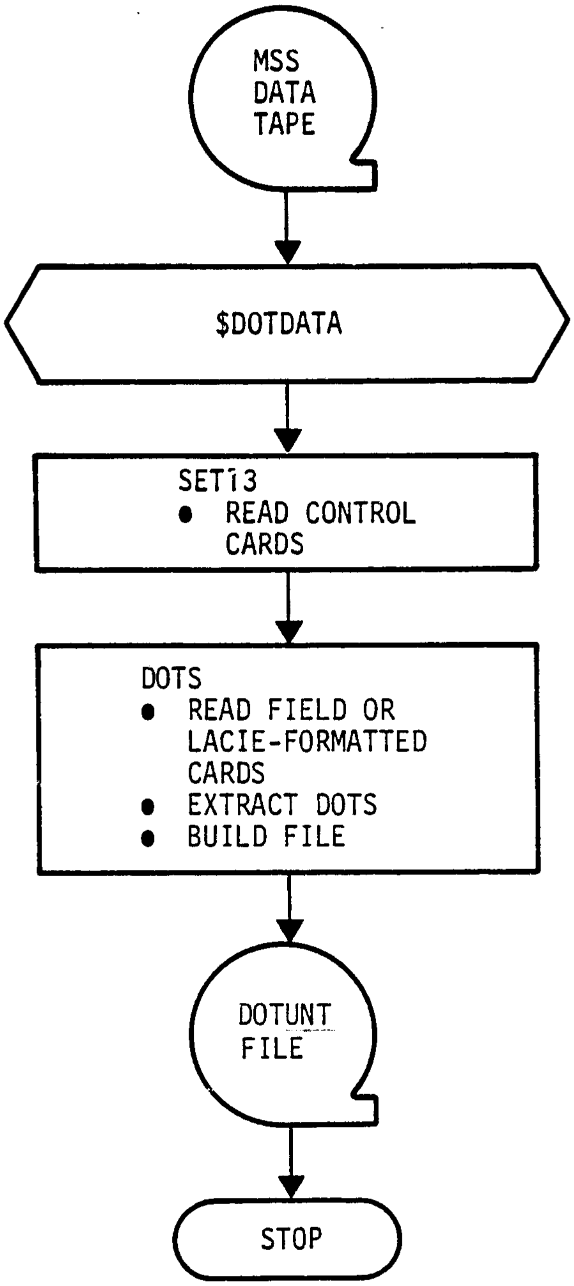


Figure 17-1.— Functional flow chart for the DOTDATA processor.

18. AUTOMATIC CLUSTER LABELING PROCESSOR — LABEL

To aid the analyst in supervising the labeling of the statistics obtained from the clustering processor ISOCLS, a new technique for labeling the statistics was implemented.

Two procedures for labeling the statistics are provided. Optionally, the analyst may select either the k-nearest-neighbor procedure or the all-of-a-kind procedure (fig. 18-1).

The labels in the dot data file, DOTUNT, and/or previously labeled statistics file, SAVTAP, may be relabeled by control card input (table 18-1) and the updated file(s) output to tape. Optionally, (1) a conditional or mixed cluster map may be output to tape; (2) an unconditional cluster map may be output in the format acceptable by the DISPLAY processor; (3) the spatial and spectral information of the relabeled DOTUNT may be output to the line printer; and (4) the statistics of the SAVTAP file may be output to the line printer.

18.1 PROCEDURES

A distance table containing the L_1 or L_2 distance of all the type 1 dots and the mean of the clusters is computed.

$$L_1 = \sum_{i=1}^{n} \left| x_i - u_i \right| \tag{18-1}$$

$$L_2 = \sqrt{\sum_{i=1}^{n} (x_i - u_i)^2} \tag{18-2}$$

where

x_i = *ith* element of the dot vector

u_i = *ith* element of the mean vector

n = number of channels

Using the table as input to the k-nearest-neighbor procedure or all-of-a-kind procedure, the statistics generated during clustering are labeled.

For the k-nearest-neighbor procedure, all the labels of the k-nearest labeling dots to a given cluster are polled. The label with the majority of the dots labels the cluster. If a tie occurs, then k-1 dots are considered.

For the all-of-kind procedure, all of the labeling dots within a cluster are polled. If all the dots are of one category, the cluster labels that category. If the cluster contains dots for more than one category, the label with the majority of the dots labels the cluster. If there are no labeling dots within a cluster, the labeling defaults to the k-nearest-neighbor procedure.

Optionally, a conditional cluster map may be output. A cluster is tagged as conditional if the distance between the nearest identically labeled labeling dot and mean of the cluster is greater than the analyst input threshold value t.

Optionally, a mixed cluster map may be output. A cluster is tagged as mixed if the labeling dots within a cluster are of more than one category.

Optionally, an unconditional cluster map may be output in the format acceptable to the DISPLAY processor. Information used in the thresholding procedure in DISPLAY is dummied. If thresholding of the clustered data is desired, it can be performed by exercising the conditional map option in this processor.

18.2 INPUT FILES

Optionally, the three files listed below may be input.

a. Statistics file (SAVTAP) from ISOCLS or STAT processor

b. Cluster map file (MAPUNT) from ISOCLS processor

c. Dot data file (DOTUNT) from DOTDATA processor

For format of the files, see sections 4.1 and 5.1.

18.3 OUTPUT FILES

Optionally, the following files are output.

a. Labeled statistics file (SAVTAP) by using a procedure (See section 4.1 for format of file.)

b. Labeled statistics file (SAVTAP) by control card input

c. Relabeled dot data file (DOTUNT)

d. A conditional cluster map (See section 5.1 for format of file.)

e. A mixed cluster map

f. An unconditional cluster map in the format acceptable by the DISPLAY processor (See appendix C for format of file.)

A line printer summary of the following is output.

a. Summary of selected options

b. Table of L_1 or L_2 distances

c. Summary of the labeling dots within a cluster for the all-of-a-kind procedure

d. Summary of the labeling dots for the k-nearest dots to a cluster for the k-nearest-neighbor procedure

e. Spatial and spectral information of relabeled DOTUNT file

f. Mean and covariances of labeled or relabeled statistics

18.4 SCRATCH FILES

LABEL does not require an additional scratch file.

18.5 CARD INPUT

Formats for all system card input are defined in section 3.1.

18.5.1 PROCESSOR CARD

The processor keyword is left justified starting in column 1: $LABEL (Loads into the system all the routines needed for executing this processor.)

18.5.2 SYSTEM FILES

The LABEL processor does not use any special input decks.

18.5.3 CONTROL CARDS

Control cards allow the user to input various options. These cards are identified by a keyword, left justified in columns 1 through 10 of the card, with the parameter values or additional keywords in columns 11 through 72 (beginning in any column after column 10). These control cards may be in any order, but they must be the first cards after the processor card LABEL. Table 18-1 lists all available options, along with their default values.

18.5.4 FIELD DEFINITIONS

If the all-of-a-kind procedure is selected, a conditional or mixed cluster map is output, or a DISPLAY interface tape is output, a MAPUNT tape must be input and a field card defining the area of the unconditional cluster map (input MAPUNT) must also be input. The vertices must reflect the sample number and line number of the MSS data tape used to create the unconditional cluster map. That is, the field card must be identical to the field card input to ISOCLS.

If a procedure is not to be executed, a MAPUNT file is not output, or a MAPTAP file is not output, a field card is not input.

Example of the data set:

- MAPUNT tape is being input.

```
*END
FIELD          (1,1),(1,1),(196,1),(196,117),(1,117)
$END
```

- MAPUNT tape is not being input.

```
*END
$END
```

18.6 CARD OUTPUT

This processor does not produce cards.

18.7 RESTRICTIONS

General system restrictions apply to LABEL.

18.8 DIAGNOSTIC MESSAGES

18.8.1 SUBROUTINE ALLKIN

Message	
a. LABELING BY ALL-OF-A-KIND PROCEDURE.	Supervisory Messages.
b. ** DEFAULTING TO K-NEAREST NEIGHBOR PROCEDURE ** .	
c. A TIE OCCURRED. THE FOLLOWING DOTS WERE DISCARDED.	

18.8.2 SUBROUTINE FILERD

	Message	Explanation
a.	NOT ENOUGH CORE TO STORE DOTFILE.	Self-explanatory.
b.	NOT ENOUGH CORE OR DISK SPACE OF CLUSTER MAP INFO.	Self-explanatory.

18.8.3 SUBROUTINE KNEAR

	Message	Explanation
a.	LABELING BY XXX NEAREST NEIGHBOR PROCEDURE.	Supervisory Messages.
b.	A TIE OCCURRED. THE FOLLOWING DOT(S) WERE DISCARDED.	

18.8.4 SUBROUTINE LABLR

Message	Explanation
CATEGORIES HAVE NOT BEEN DEFINED.	Check spelling and keywords on CATEGORY control card.

18.8.5 SUBROUTINE MANORD

	Message	Explanation
a.	ERROR IN INPUT OF CLASS NAMES. NAMES ON STAT FILE ARE:	Self-explanatory.
b.	NAMES INPUT ARE:	Self-explanatory.

18.8.6 SUBROUTINE SET14

	Message	Explanation
a.	INVALID CONTROL CARD - IGNORED.	Check spelling of keyword.
b.	ERROR ON CHANNELS CARD.	Check format and parameters.
c.	ERROR ON DATAFILE CARD.	Check format and parameters.
d.	ERROR ON MAPFIL CARD.	Check format and parameters.
e.	ERROR ON MAPFIL CARD.	Check format and parameters.
f.	ERROR ON OPTION CARD.	Check parameter.
g.	ERROR ON STATFILE CARD.	Check format and parameters.
h.	ERROR ON PROCEDURE CARD.	Check spelling and parameter.
i.	ERROR ON MAPTAP CARD.	Check format and parameters.
j.	NO. OF STAT CHANNEL AND DOT DATA CHANNELS MUST BE EQUAL.	Self-explanatory.
k.	A LABELING PROCEDURE MAY NOT BE CHOSEN WHEN UPDATING THE DOTUNT OR SAVTAP FILES.	You may want to go through $LABEL again after FILES have been updated.
l.	USER HAS NOT INPUT ONE OF THE REQUIRED FILES: SAVTAP MAPUNT OR DOTUNT.	Self-explanatory.

18.8.7 SUBROUTINE STOMAP

Message	Explanation
NOT ENOUGH DISK SPACE TO STORE DAS TAPE DATA.	A DAS tape output unit should be assigned.

TABLE 18-1.— LABEL PROCESSOR OPTIONS AND CONTROL CARDS

Keyword (a)	Parameter and default values (b)	Function
CHANNEL	STAT=$n_1,n_2,\cdots,n_{30}$ DATA=$m_1,m_2,\cdots,m_{30}$ Default: n_i = all channels on SAVTAP file m_i = all channels on DOTUNT file	n_i and m_i are integer numbers separated by commas referring to the channels on the SAVTAP file and the DOTUNT file, respectively.
DOTFILE	INPUT/FILE=m,UNIT=n Default: None	Defines the unit and file for the input dot data file, DOTUNT. n is the Fortran unit number assigned to the input DOTUNT file. m is the number of the input file to process.
DOTFILE	OUTPUT/FILE=m,UNIT=n Default: None	Defines the unit and file to which the relabeled DOTUNT file is output. n is the Fortran unit number assigned to the output DOTUNT file. m is the number of the file to output.
STATFILE	INPUT/FILE=m,UNIT=n Default: None	Defines the unit and file for the input SAVTAP file.

[a]The keyword must be left justified in columns 1 through 10.
[b]The parameter and default values are in card columns 11 through 72 (beginning in any column past 10).

TABLE 18-1.– Continued.

Keyword	Parameter and default values	Function
		n is the Fortran unit number assigned to the input SAVTAP file. m is the number of the file to process.
STATFILE	OUTPUT/FILE=m,UNIT=n Default: None	Defines the unit and file to which the labeled/relabeled SAVTAP file is output. n is the Fortran unit number assigned to the output SAVTAP file. m is the number of the file to output.
MAPTAP	OUTPUT/FILE=m,UNIT=n Default: No DISPLAY interface tape will be output.	Defines the unit and file to which the unconditional cluster map, MAPTAP, is output. n is the Fortran unit number assigned to the output MAPTAP file. If executing back to back with DISPLAY, n must be assigned to unit 2. m is the number of the file to output.
MAPFILE	INPUT/FILE=m,UNIT=n Default: m=1,n=16	Defines the unit and file for the input unconditional cluster map, MAPUNT, that is output by ISOCLS during the clustering process.

TABLE 18-1.— Continued.

Keyword	Parameter and default values	Function
		n=16 is the Fortran unit number assigned to the input MAPUNT file. (n must be assigned to unit 16 if executing back to back with ISOCLS). m is the number of the file to process.
MAPFILE	OUTPUT/FILE=m,UNIT=n Default: n=16,m=1	Defines the unit and file to which the conditional and/or mixed cluster map is output. n is the Fortran unit number assigned to the output MAPUNT file. m is the number of the file to output.[c]
DOTLABEL	Category name, $n_1, n_2, \cdots, n_{250}$ Default: None	The DOTUNT file is labeled or relabeled by this card. Category name is the label the analyst is assigning to the dots n_j, $j=1,\cdots,i$. The category name may be composed of a maximum of 6 characters. n_j are integer numbers separated by commas referring to the position of the dot on the DOTUNT file.

[c] If both type of maps are output, the conditional map is output on file m; the mixed map, on file m+1.

TABLE 18-1.- Continued.

Keyword	Parameter and default values	Function
STALABEL	Class name, n_i, $n_2, \cdots, n_{250}$ Default: None	The SAVTAP file may be manually relabeled by this card. n_j $(j=1,2,\cdots,i)$ are the number of the subclasses on the SAVTAP that are to be regrouped into another class. Class name is the name of the class to which the subclasses n_j are to be reassigned. Class name must match a name on the SAVTAP file.
DISTANCE	L1 Default: L_1 distance	The L_1 distance is used in computing the distance between the means of the cluster and the labeling dots.
DISTANCE	L2 Default: L_1 distance	The L_2 distance is used in computing the distance between the means of the cluster and the labeling dots.
OPTION	COND Default: None	A conditional cluster map is output.
OPTION	MIXED Default: None	A mixed cluster map is output.
THRESHOLD	T Default: T=25.0	T is the threshold parameter used in creating the conditional cluster map. T is a floating-point number.

TABLE 18-1.– Continued.

Keyword	Parameter and default values	Function
NEAREST	K Default: K=1	K is the number of dots to be used in the k-nearest-neighbor procedure. K is an integer number. $k \leq 250$.
PROCED	NAME Default: N=K-NEAREST	NAME is an alpha word. NAME=K-NEAREST (Use the k-nearest-neighbor procedure.) NAME=ALL (Use the all-of-a-kind procedure.) NAME=MANUAL (Use the manual procedure of relabeling the DOTUNT or SAVTAP file.
MODULE	Blank	Initiates the input of the module STAT card file. The deck must immediately follow this card.
EXCLUDE	$n_1, n_2, \cdots, n_{250}$ Default: All dots on the DOTUNT are used.	n_i are integers numbers referring to the dots on the DOTFIL that are to be excluded in all calculations (i.e., dots within a DO/DU area).
SUNANG	$m_1, m_2, \cdots, m_i$ Default: No Sun angle correction is to be applied.	m_j are integer Sun angle numbers used in computing the L_1 or L_2 distances. A Sun angle must be input for each acquisition of interest. An acquisition is assumed to be a 4 channel pass.

TABLE 18-1.— Continued.

Keyword	Parameter and default values	Function
		Example: If the distance is computed using 16 channels, 4 Sun angles (m_1, m_2, m_3, m_4) must be input.
SUNANG	FILE Default: No Sun angle correction is applied.	Sun angles will be extracted from the DOTUNT file.
OPTION	STATS Default: Statistics not printed.	Means and covariances for labeled or relabeled statistics on SAVTAP file are printed.
OPTION	DOTS Default: Relabeled DOTFIL is not printed.	Spatial and spectral information of relabeled DOTUNT file is printed.
HED1	Any 60 characters Default: LYNDON B. JOHNSON SPACE CENTER	First line of the heading on line printer output.
HED2	Any 60 characters Default: HOUSTON, TEXAS	Second line of the heading on line printer output.
DATE	Any 12 characters Default: Present date	Date in the heading on line printer output.
COMMENT	Any 60 characters Default: None	Comment printed with heading on line printer output.

TABLE 18-1.– Concluded.

Keyword	Parameter and default values	Function
*END	Blank	Indicates the end of the control card inputs.
$END	Blank	Indicates the end of all the card inputs.

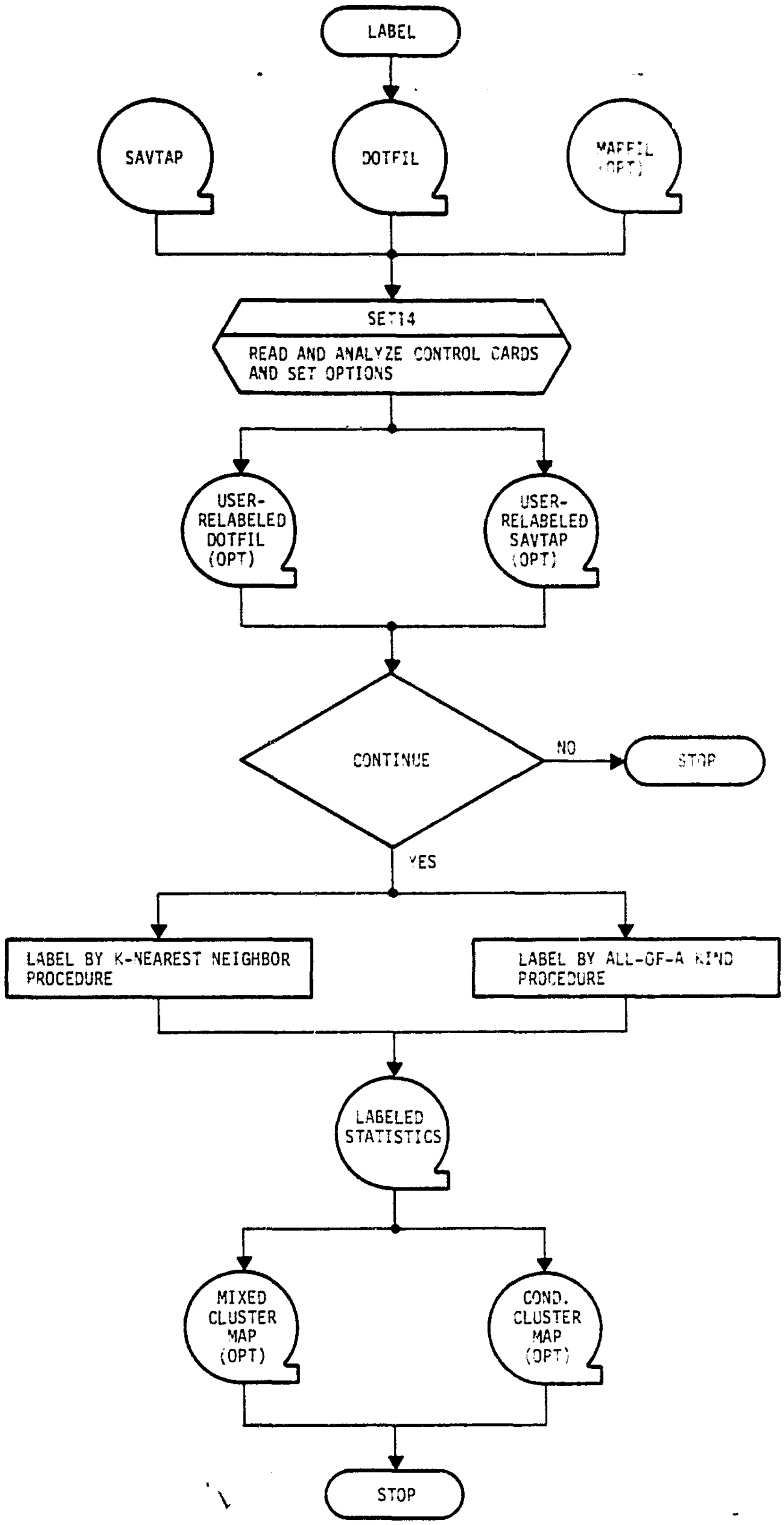


Figure 18-1.- Functional flow chart for the LABEL processor.

19. SYSTEM RESTRICTIONS

The system is limited in every processor to processing no more than 30 channels of data. The MSS data tape (DATAPE) may have more than 30 channels, but for processing purposes a subset of those channels must be selected via the CHANNELS control card.

A maximum of 60 categories, classes, and subclasses may be processed. However, the maximum number of channels and subclasses may not be processed at one time. The arrays within the system are dimensioned variably according to user requests. The amount of storage available will not accommodate the arrays which are dimensioned number of subclasses by number of channels, if both maximums are used. Restrictions under the STAT, SELECT, and CLASSIFY processors allow the user to compute approximately whether or not the numbers of channels and subclasses selected are acceptable. When core storage requirements are exceeded, a diagnostic message is printed, and the user must reduce his requirements in order to get a successful execution. Restrictions specific to individual processors are noted in the documentation for each processor.

Owing to the virtual storage characteristics of VM/370, arrays can be enlarged by the System Maintenance Group at user request.

APPENDIX A

SYSTEM DIAGNOSTIC MESSAGES

APPENDIX A

UTILITY SUBPROGRAMS — DIAGNOSTIC MESSAGES

The diagnostic messages listed in this section are printed by utility subprograms used by more than one processor. Diagnostics which occur only in a given processor are listed in the sections for the individual processor.

SUBPROGRAM: Message	Explanation
BUFILL: XXXX BYTES EXPECTED. XXXX BYTES ON RECORD.	Self-explanatory.
CLSCHK:	
a. ** CLSCHK** - REQUESTED SUBCLASS NO. XXX IS NOT AVAILABLE IN INPUT STATISTICS -- REQUEST IGNORED.	Either by the SUBCLASS control card or by default, a subclass number has been requested to be used in classification, which is greater than the largest subclass number available in the input training subclasses. The CLASSIFY processor ignores the requested subclass number and deletes it as a possibility for use in classification.
b. ** CLSCHK** - REQUESTED SUBCLASS NO. XXX FOR GROUP NO. XXX IS NOT AVAILABLE IN INPUT STATISTICS FILE.	A subclass number input by either the SUBCLASS or the GROUP control card is greater than the largest training subclass number available. The requested subclass will be

SUBPROGRAM: Message	Explanation
	deleted and ignored in classification.
c. ***** CLSFY/FETCHK --- CHANNEL XXX NOT IN TRAINING DATA --- TRAINING DATA CHANNELS ARE $C_1, C_2, C_3, \cdots, C_N$.	A channel requested on the CHANNELS control card or in the B-matrix input for use in classification is not available in the set of channels which was used to obtain training subclass statistics. The available set of channels in the training statistics is printed out as part of the diagnostic message. Messages d and e are added to this when the B-matrix is not involved (d) and when B-matrix channels are input (e).
d. ***** CHANNEL XX IGNORED (NOT USED) IN CLASSIFICATION.	If the B-matrix is not being input to the CLASSIFY processor (i.e., B-matrix channels are not involved), the requested channel in the previous message will be deleted from the list of channels and ignored by the CLASSIFY processor. In this case, this message is added to the previous diagnostic message.
e. ***** B-MATRIX CHANNELS MUST BE EQUAL TO OR A SUBSET OF AVAILABLE TRAINING DATA CHANNELS --- THE INPUT B-MATRIX CHANNEL SET IS $B_1, B_2, \cdots, B_M$:	If the B-matrix is input to the CLASSIFY processor, the B-matrix channels become the set which is to be used in classification; and if one of the B-matrix channels is not available in the

SUBPROGRAM: Message	Explanation
	training subclass statistics, the processor cannot continue. In this case, this will be the remainder of the error message.
CMERR:	
ERROR HAS OCCURRED.	Self-explanatory.
CRDSTA:	
EXCEEDED CORE LIMITS. REDUCE NO. OF TRAINING CLASSES OR FEATURES. EXITING FROM CRDSTA.	The combination of total number of channels, subclasses, and training fields must be reduced to fit in the internal core storage available to the processor. Total storage is 10 600 locations.
FLDINT:	
a. FEATURE NUMBERS XXXXX AND ABOVE ARE NOT ON DATA TAPE.	User has requested a channel not on MSS DATAPE.
b. FIRST SCAN ON THIS TAPE IS NUMBERED XXXXXX. FIELD DEFINITION IN ERROR.	Self-explanatory.
c. NUMBER OF SAMPLES OF PER SCAN ON THIS TAPE IS XXXXXX. FIELD DEFINITION IN ERROR.	Self-explanatory.
d. THIS TAPE CONTAINS ONLY XXXXXX CHANNELS.	Self-explanatory.

SUBPROGRAM: Message	Explanation
FSBSFL:	
FSBSFL ONLY SKIPS FORWARD.	Self-explanatory.
FSFMFL:	
FSFMFL ONLY SKIPS FORWARD.	Self-explanatory.
GETST:	
a. ERROR IN POSITIONING UNIT XXX TO FILE XXX.	Self-explanatory.
b. REQUESTED SUBCLASS IS NOT ON STAT FILE. STAT FILE CONTAINS XXX SUBCLASSES.	Self-explanatory.
c. CHANNEL NO. XX IS NOT ON TRAINING STAT FILE. CHANNELS ARE $C_1, C_2, C_3, \cdots, C_N$.	Self-explanatory.
GRPSCN:	
///// FROM SUBR. GRPSCN --- CLASS XXXXX INCORRECT --- CLASS XXXXX IGNORED. CARD BEING SCANNED IS XXXX•••XXXX.	One of the class numbers listed on the GROUP control card has one or more of the following three conditions: (1) not in ascending order, (2) greater than the largest class number allowable (30), or (3) has already been used in another GROUP card. The erroneous GROUP control card is printed as part of the message. The processor will delete the erroneous class number from the list and proceed grouping all other listed classes.

SUBPROGRAM: Message	Explanation
HISTGM:	
TOO MUCH DATA REQUESTED -- SAMPLE END WAS RESET TO XXXXX.	Self-explanatory.
I4A1BN:	
EBCDIC TO BINARY INTEGER CONVERSION ERROR AT CHARACTER XXXXX OF XXXXX CHARACTER FIELD: XXXX•••XXXX.	Self-explanatory.
LABMAN:	
a. ERROR IN POSITIONING SIG. EXTENSION TAPE TO FILE XXX. OUTPUT FILE NOT WRITTEN.	Self-explanatory.
b. THE STATISTICS FILE FOR XXXX CLASSES AND XXXX SUBCLASSES HAS BEEN WRITTEN.	Self-explanatory.
c. THE STATS WERE WRITTEN ON FILE XXX.	Self-explanatory.
d. THE STATS FOR A PARTICULAR CLASS OR SUBCLASS SHOULD BE REFERRED TO IN LATER RUNS BY THE FOLLOWING NUMBERS (WHICHEVER APPLICABLE).	Self-explanatory.
LAREAD:	
a. ERROR IN FIELD CARD TERMINATING RUN.	A field description card has an incorrect format. All vertices must be separated by commas and enclosed in parentheses, and

SUBPROGRAM: Message	Explanation
	sample and line numbers must be integers. The card which caused the error is printed out with this message.
b. INCORRECT FIELD CARD, TERMINATING RUN.	Same.
LINERD:	
a. FIELD BOUNDARY FOR THIS FIELD DEFINED BEYOND SCOPE OF DATA. THIS FLIGHT LINE CONTAINS XXXXXX SCAN LINES.	User has requested scan line not on MSS DATAPE.
b. FLDINT MUST BE CALLED TO INITIALIZE PARAMETERS FOR A NEW FIELD.	For every field input there must be a call to FLDINT to reset parameters for positioning the MSS DATAPE.
RDDOTS:	
a. CHANNEL XX IS NOT ON DOTFIL. CHANNELS ARE $C_1, C_2, \cdots, C_N$.	Self-explanatory.
b. DOT NO. XXX IS NOT ON DOTFIL. FILE CONTAINS XXX DOTS.	Self-explanatory.
RDMODK:	
ERROR IN TRYING TO POSITION STAT FILE TO FILE XXX IN CRDSTA.	An error occurred in positioning the SAVTAP file, and no statistics were written. Resubmit the run.

SUBPROGRAM: Message	Explanation
REDDAT:	
CHANNEL NO. XX IS NOT A TRAINING CHANNEL. XX TRAINING CHANNELS ARE $C_1, C_2, \cdots, C_N$.	Self-explanatory.
REDSAV:	
a. STAT FILE WAS NOT CREATED. EXITING FROM **REDSAV**.	An error occurred in positioning the SAVTAP file, and no statistics were written. Resubmit run.
b. ERROR IN POSITIONING STAT FILE TO FILE XXX. EXITING FROM REDSAV.	Same.
c. USER HAS REQUESTED XX CHANNELS, XX SUBCLASSES, AND XX CLASSES. THIS COMBINATION OF STATS WILL NOT FIT IN CORE. PLEASE REDUCE REQUEST.	The fixed amount of internal core storage available to the processor for storing class descriptions, number of subclasses in each class, subclass descriptions, field information, vertices, covariances, means, and working area has been exceeded. The total amount of storage available for the above information is 10 600 locations.[1] Reduce the requested combination.

[1] The equation for computing the required storage is: STORAGE = 2(number of classes) + (number of subclasses) + 4(number of fields) + 2(total number of vertices for all the fields) + (number of subclasses + 1)[(number of channels)(number of channels + 1)/2] + (number of subclasses)(number of channels).

SUBPROGRAM: Message	Explanation
SEARCH:	
a. SEARCHING FOR LINE.	Self-explanatory.
b. RECORDS PER SCAN, XXXXX. SCANS PER RECORD, XXXXX.	Self-explanatory.
c. FOUND IT AFTER XXX TRIES.	Self-explanatory.
d. FAILED AFTER XXXXX TRIES -- ABORTING.	Self-explanatory.
e. SCAN XXXXX IS MISSING -- USING PREVIOUS SCAN INSTEAD.	Self-explanatory.
TAPHDR:	
a. UNRECOVERABLE ERROR READING HEADER RECORD.	Error occurred while trying to read header record.
b. A LINE NO. IS LESS THAN OR EQUAL ZERO.	The first line number on the data tape is less than or equal to zero.
c. LAST SCAN LINE READ XXXXX. ISTAT = XXXXX.	Self-explanatory.
d. INTERNAL DIMENSIONS TOO SMALL FOR DATA. NUMBER OF CHANNELS ON DATA TAPE = XXXXXXX. NUMBER OF POINTS/CHANNEL = XXXXXXX.	The maximum record size of the data record exceeded 6800 words.
e. CHECK THE FOLLOWING POSSIBLE ERRORS. 1. DATA TAPE IS NOT IN UNIVERSAL OR LARSYS FORMAT.	The MSS DATAPE must be in LARSYS III or Universal format.

SUBPROGRAM: Message	Explanation
f. ONLY ONE OR LESS RECORDS PER CHANNEL ACCEPTABLE AT THIS TIME.	All of the samples of one channel must be contained within one record.
g. NO. OF RECORDS PER DATA SET = XXXXX. MUST BE LESS THAN OR EQUAL 15.	One data set cannot contain more than 15 records.
h. NO. OF BITS/PIXEL = XXXXX. ONLY 8 BITS ACCEPTABLE AT THIS TIME.	According to the header record, the samples on the MSS DATAPE do not equal eight bits. It is assumed that the header record is in error, and execution continues.
i. DATA ORDER INDICATOR = XXXXX. DATA MUST BE ORDERED BY PIXEL.	Information from header records indicates data are not ordered samples by channels.
WRTHED:	
NUMBER OF SAMPLES WAS RESET TO 2998.	Number of samples per line cannot be more than 2998.

APPENDIX B

LARSYS III FORMAT FOR AN MSS DATA STORAGE TAPE

APPENDIX B

LARSYS III FORMAT FOR AN MSS DATA
STORAGE TAPE

This is the third version of the format used in Purdue's LARSYS. The only difference in the second and third version of the format is one word in the header record. That difference is transparent to this system.

There are four types of (physical) records on the Multispectral Scanner Data Storage Tapes. They are:

1. ID record - 200 full words fixed length
2. Data record - variable length
3. End-of-Tape records - 200 full words fixed length
4. End-of-File records - IBM Standard

A Multispectral Scanner Data Storage Tape contains one or more data runs consisting of an ID record, several data records and an End-of-File record. After the last data record on the tape, an End-of-Tape record and two End-of-File records are written on the tape.

For the purposes of th[illegible] presentation, a 'word' is defined to be 32 bits and a byte to be 8 bits. Further details regarding the physical records follows:

1. <u>ID record</u> (200 full words fixed length)

	FORMAT	DESCRIPTION
ID(1)	Integer	LARS Tape Number (e.g., 1, 17, 102, etc.)
ID(2)	Integer	File number on this tape
ID(3)	Integer	Run number (8 digits aabbbbcc) aa - last 2 digits of the Year data was acquired bbbb - running serial number for the year data was taken cc - uniqueness digits for runs which would otherwise have the same run number
ID(4)	Integer	Continuation Code ID(4) = 0 means the first line of data follows this ID record ID(4) = X means that the data following this ID record is a continuation of a flight line started on tape X
ID(5)	Integer	Number of Data Channels (Spectral bands) on tape (30 maximum)
ID(6)	Integer	Number of Data Samples per channel per scan line
ID(7-10)	Alpha-numeric (4A4)	Flightline Identification (16 characters)

	FORMAT	DESCRIPTION
ID(11)	Integer	Month data was taken
ID(12)	Integer	Day data was taken
ID(13)	Integer	Year data was taken
ID(14)	Alpha-numeric (1A4)	Time data was taken
ID(15)	Integer	Altitude of aircraft
ID(16)	Integer	Ground heading of aircraft
ID(17-19)	Alpha-numeric (3A4)	Date data run was generated on this tape (12 characters)
ID(20-50)	Integer	All zero (to be defined later)
ID(51)	Real	Lower limits in micrometers of first spectral band on tape
ID(52)	Real	Upper limits in micrometers of first spectral band on tape
ID(53)	Real	The suggested value of "C0" calibration pulse
ID(54)	Real	The suggested value of "C1" calibration pulse.
ID(55)	Real	The suggested value of "C2" calibration pulse.
ID(56-200)	Real	Repeat of ID(51-55) for ID(5) channels in order of appearance in Data Records.
ID(51-200)	Real	= 0.0 if Data Channels do not exist.

Data Record:

Each data record will contain one scan line of data from ID(5) (see ID Record) channels. The first half word (2 bytes) of the record will be the record number. The second half word (2 bytes) will be the roll paramater which is a number indicating relative roll of the aircraft for this scan line of data. If the roll parameter is -32,767, the data for the given line does not exist. If the roll parameter has not been calculated, it will be set to 32,767. The fifth byte will be the first data sample from the first channel. The data samples are ordered: $Channel_1$, $Sample_1$ - $Sample_N$; $Channel_2$, $Sample_1$-$Sample_N$; and so on through ID (5), Channels and ID (6) data samples per channel. A data record (scanline) will be ID(5)* ID(6)+4 bytes long.

The data from each channel will be from the field of view of the scanner except the last six bytes. The last six are calibration data in the order of appearance:

1.	C_0	"0" or dark level
2.	VC_0	Variance of C_0
3.	C_1	Calibration source C_1
4.	VC_1	Variance of C_1
5.	C_2	Calibration source C_2
6.	VC_2	Variance of C_2

where C_i = Calibration value i and VC_i = calculated variance of calibration value i.

On good data records all 8-bit data and calibration values will be integers in the range of 0 to 255 with no sign included in the eight bits. A sample data value of 0 to 255 is the result of the 8-bit analog-to-digital conversion which produces the multispectral scanner data tape. With 8-bit A/D conversion, data values range between 0 to 255 with 0 usually indicating low relative irradiance and 255 usually indicating high relative irradiance.

End-of-Tape Record:

The End-of-Tape Record is very similar to the ID Record with 200 full words in the following format.

WORD	FORMAT	DESCRIPTION
ID(1)	Integer	LARS tape number
ID(2)	Integer	File number on this tape
ID(3)	Integer	Set equal to zero
ID(4)	Integer	Continuation Code ID(4) = 0 means end of data ID(4) = X means data in previous file is continued on tape X
ID(5-50)	Integer	All zero (may be defined later)
ID(51-200)	Real	0.0 (may be defined later)

APPENDIX C

UNIVERSAL FORMAT FOR AN MSS DATA TAPE

APPENDIX C

UNIVERSAL FORMAT FOR AN MSS DATA TAPE

This is an adaptation of the Universal Data Tape Format as defined in the Earth Resources Data Format Control Book, Vol. 1 (ref. 12, section 7).

Ground Rules

The ground rules for the UNIVERSAL format as accepted by all the processors within this system are as follows:

- 8 bits = 1 byte
- The header record is the first record on a tape.
- The header record is fixed length equal to 3060 bytes.
- Data following the header will be arranged by data sets, where a data set is defined as the ancillary data and all of the video data for one scan line for all active channels.
- Data sets will be recorded in variable length physical records, not to exceed 3000 bytes of information per record. Note, since 3000 bytes is not compatible with the word lengths of all computers, the computer generating the tape will add a sufficient number of fill zeros to the end of the data to make the record length divisible by 32, 36, 48, and 60 bits (180 bytes). Therefore, it is possible to have a max physical record length of 3060 bytes, but under no condition will data exceed 3000 bytes.

- Data sets will be packed into consecutive physical records of equal length. Under no condition will a data set begin in the middle of a physical record unless the data set can be completed in that record. If two or more records are needed for the data set, the data set will be divided but under no condition will the data for a video channel begin in the middle of a physical record unless the data for that video channel can be completed in that record. Consequently, data sets which are lengthy will be divided so that the ancillary block and video data from an integral number of channels will be in one record and remaining video data will follow in succeeding records with an integral number of channels per record. Fill zeros will be supplied at the ends of the records as required to satisfy the equal length contraint noted above.
- All data in the header record and ancillary blocks will be in binary.
- The tape format will be as follows:

```
Header Record
    IRG*

Ancillary Block     Data Set
Video Block

    IRG**

Ancillary Block     Data Set
Video Block

    IRG
     .
     .
     .
    EOF
```

* IRG = Inter Record Gap - always follows the header record.
** An IRG may appear between the ancillary block and the video block so that the recording of a data set requires more than one physical record; or a physical record may contain two or more data sets, not separated by any IRG. See ground rules above and data set description following for criteria determining the placement of IRG's.

Header Record

Although the header record is 3060 bytes in length, only a portion of the information is pertinent to the system at this time. A general description of the data that is unpacked by the TAPHDR routine is as follows:

BYTE NO.	DESCRIPTION	NO. OF BYTES
89	Processing flag 0 = Raw Data 1 = Processed data from computing system	1
90	No. of channels in this job	1
91	No. of bits in a picture element (Must be 8 at this time)	1
92-93	Address of start of video data gives location of start of video within scan.	2
96-97	No. of video elements per scan within a single channel.	2
100-101	Physical record size in bytes This number must be a multiple of 180 bytes.	2
102	No. of channels per physical record This field refers to the second and subsequent records within the recording of a data set. Bytes 1785-1786 give the number of channels of data in the first record of a data set. If no. elements per channel greater than 3K, this field will equal 0.	1

BYTE NO.	DESCRIPTION	NO. OF BYTES
103	No. of physical records per scan per channel. This field is used only when the no. of elements per channel is greater than 3K. Otherwise it is equal to 0.	1
104	No. of records to make a complete data set	1
105-106	Length of ancillary block in bytes	2
107	Data Order Indicator 0 = Video ordered by channel. 1 = Video ordered by pixel	1
108-109	Start Pixel No. Number of the first pixel per scan on this tape referenced to original image. The first pixel in the original image is pixel number one.	2
110-111	Stop Pixel No. Number of the last pixel per scan on this tape referenced to original image.	2
1778	Number of Data Sets per Physical Record	1
1785-1786	Number of channels in the first physical record of the data set	2
1787-1788	Total number of bytes per scan per channel	2

Data Sets

Ancillary Block

- The first block of a data set is the ancillary block.
- The length of the ancillary block is variable, with the number of bytes given in the header record.

- The first word (2 bytes) of every record is a counter giving the number of the physical record within the video data set. This is primarily intended for use in data sets that are greater than 3000 bytes long and therefore require more than one physical record for recording. This word will always equal "1" for the first record of a data set.
- Bytes 3 through 6 will contain the current GMT at the start of this data set recorded in tenths of milliseconds.
- Bytes 7 through 70 will indicate channel status for this scan, one byte per channel, where the LSB = 0 indicates the channel is sync, and the LSB = 1 indicates the channel not in sync.
- 71-72 contain the scan line number. This will be an arbitrary but sequential count for each scan line that appears in the data run.
- Bytes 73 through N will be dependent on whether this job contains raw or processed data. (See byte 89 in the header record.) The value of N will be given in bytes 105 and 106 in the header record and will always be equal to or greater than 70.
- If this job contains raw data, bytes 73 through N will contain the housekeeping data channel from the sensor, if one is available.
- A job containing processed data will, in addition to the 70 bytes of ancillary data already described, contain, at a minimum, the following pieces of information:
  - Latitude of the aircraft or of the center of the image from EREP or satellite in binary.
  - Longitude of the aircraft or of the center of the image from EREP or satellite in binary.

- Altitude in meters recorded in binary.
- Heading in tenths of a degree.
- Ground speed in meters per second.
- Roll - Defined in specific formats, following.
- Pitch - Defined in specific formats, following.
- Yaw - Defined in specific formats, following.
- Sun angle.

The specific formats for each sensor (following in this section) shall provide where this data will appear in the format.

Other parameters may be added, if required, with the length of the ancillary block given in the header.

Video Data

- Following the ancillary block in each data set will be the video data from all of the active channels for one scan. The video data from all of the active channels for one scan will comprise a video block.
- Video blocks within a data run will always contain the same number of video channels.
- Each video block will be the same number of bytes in length. If video data is not available to fill a block, fill zeros will be added to make it the same length as preceding video blocks.
- Video data less than 8 bits per pixel will be packed, right justified, in an 8-bit byte with zeros added to the left.

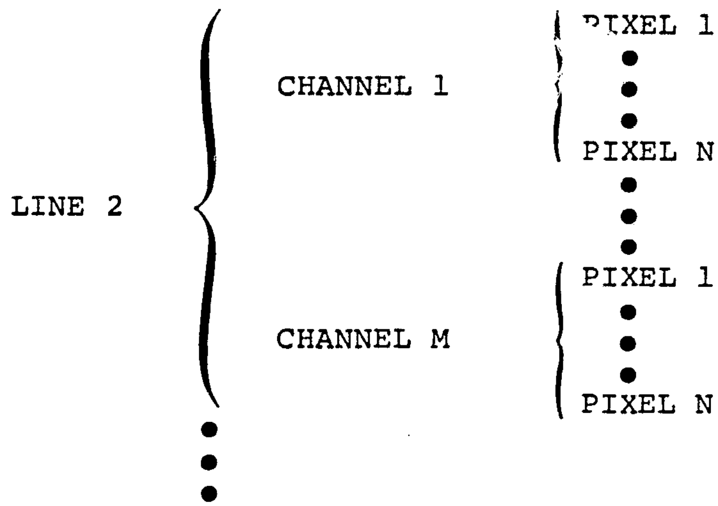
- Video data greater than 8 bits per pixel will be packed, right justified, in as many 8-bit bytes as necessary to hold the pixel, with zeros added to the left.
- If this tape contains raw data, the PCM sync words associated with the video data, if any, will be included with the video data on this tape. If this tape contains processed data, no sync words will be present.
- Calibration data that is associated with each scan within each channel will be included, if this tape contains raw imagery data, in the same sequence as it appears in the data stream on the flight tape. If this tape contains processed imagery data, the appearance of the calibration data will depend on the specific sensor requirements and will be specified in the respective format following in this document.
- The combined length of the ancillary block and the video block will determine the relationship between data sets and physical records. Some data runs may contain data sets which are so small more than one can be packed into one physical record. Others may contain data sets which will require a whole physical record for each. Still others may contain data sets which are so long that each data set will require two or more physical records.
- Data sets will be packed in physical records depending on the length of the data set. The ancillary block will always appear in the first physical record per data set. Following the ancillary block, as many complete channels in this data set will be recorded as will fit in up to 3000 bytes. If the data set is too long to be recorded in one physical record, the second and subsequent records will begin with the next active channel in the data set.

- If a video block is divided between more than one record for recording, the number of data channels in the first record may vary from the number of channels in the second and successive records; however, the number of channels in all records following the first per data set will always be the same. The number of channels in the first record and the number in successive records will be given in the header record. In records following the first, if video data is not available so as to allow all records to contain the same number of channels, fill zeros will be added in lieu of the video data in order to make all records the same length. In addition, fill zeros will be added to either the first record or all of the successive records, depending on which is shorter, so as to make all of the records the same length.

- The only arrangement of pixels within a scan of data will be by <u>channel</u>. The Universal format* will be as follows:

- LINE 1
  - CHANNEL 1
    - PIXEL 1
    - PIXEL 2
    - •
    - •
    - •
    - PIXEL N
  - CHANNEL 2
    - PIXEL 1
    - PIXEL 2
    - •
    - •
    - •
    - PIXEL N
  - •
  - •
  - •
  - CHANNEL M
    - PIXEL 1
    - •
    - •
    - •
    - PIXEL N

* If this tape contains raw imagery data, the PCM sync words, if any, that are associated with the data on the flight tape will be included with the data.


LINE 2
CHANNEL 1
PIXEL 1
PIXEL N
CHANNEL M
PIXEL 1
PIXEL N

APPENDIX D

MAPTAP FILE FORMAT

APPENDIX D

MAPTAP FILE FORMAT

The file, MAPTAP, is output by the processor CLASSIFY. It contains the statistics actually used in classification, the training field, category, class, and subclass information, and the classified data.

Each file consisting of the following types of records:

	4	-	run header records
Repeated for each classified field	1	-	field header record
	N	-	data records
	1	-	end-of-field record
	1	-	end-of-run record
	end-of-file mark		

All records are written with a nonformatted FORTRAN write statement.

RUN Header Record 1

```
WRITE(MAPTAP)(DATE(I),I=1,2),BMFLG,BMCOMB,BMFEAT,NOCLS2,
              NOFLD2,NOSUB2,NOFET2,TOTVT2,NOCAT,VARSZ2,
              (FETVC2(I),I=1,NOFET2)
```

FORTRAN NAME AND DIMENSION	DESCRIPTION
DATE(2)	Date classification was performed
BMFLG	Flag indicating B-MATRIX was used in classification
BMCOMB	No. of linear combinations in B-MATRIX

FORTRAN NAME AND DIMENSION	DESCRIPTION
BMFEAT	No. of channels used in computing the B-MATRIX
NOCLS2	No. of classes
NOFLD2	No. of training fields
NOSUB2	No. of subclasses
NOFET2	No. of channels used in classification
TOTVT2	No. of vertices in training fields
NOCAT	No. of categories
VARSZ2	Size of covariance for each subclass
FETVC2(NOFET2)	Actual channels used in classification

RUN Header Record 2

```
WRITE(MAPTAP)(CATNAM(I),I=1,NOCAT1),(CLSMTX(I),I=1,NOCLS2),
             (SUBNO(I),I=1,NOCLS2),(SUBDES(I),I=1,NOSUB2),
             ((FLDMTX(I,J),I=1,4),J=1,NOFLD2),((VERTEX(I,J),
             I=1,2),J=1,TOTVT2),(SUBCAT(I),I=1,NOSUB2),
             (CLSVC2(I),I=1,NOSUB2),(KATNO (I),I=1,NOCLS2),
             (KEPPTS(I),I=1,NOSUB2)
```

FORTRAN NAME AND DIMENSION	DESCRIPTION
CATNAM(NOCAT1)	Category names (if available) NOCAT1 = no. of categories if CATEGORY classifier was applied NOCAT1 = no. of classes if STANDARD classifier was applied.
CLSMTX(NOCLS2)	Class names

FORTRAN NAME AND DIMENSION	DESCRIPTION
SUBNO(NOCLS2)	No. of subclasses in each class
SUBDES(NOSUB2)	Subclass names
FLDMTX(4,NOFLD2)	Training field information 1 - field name 2 - Class number field belongs to 3 - Subclass number field belongs to 4 - No. of vertices in this field
VERTEX(2,TOTVT2)	Vertices for all the fields; ordered $(sample,line)_1,(sample,line)_2,\ldots (sample,line)_{TOTVT2}$
SUBCAT(NOSUB2)	Contains the category number to which each corresponding subclass belongs
CLSVC2(NOSUB2)	Contains the class number to which each corresponding subclass belongs
KATNO(NOCLS2)	Contains the category number to which each class belongs
KEPPTS(NOSUB2)	Contains the total number of training field pixels in each subclass

RUN Header Record 3

```
WRITE(MAPTAP)((COVMTX(I,J),I=1,VARSZ2),J=1,NOSUB2)
            ((AVEMTX(I,J),I=1,NOFET2),J=1,NOSUB2)
```

FORTRAN NAME AND DIMENSION	DESCRIPTION
COVMTX(VARSZ2,NOSUB2)	Original or B-transformed covariance matrix for each subclass
AVEMTX(NOFET2,NOSUB2)	Mean vector for each subclass

Run Header Record 4

```
WRITE(MAPTAP)((COVMTX(I,J),I=1,VARSZ2),J=1,NOSUB2),
         (CON(I),I=1,NOSUB2),(DET(I),I=1,NOSUB2)
```

FORTRAN NAME AND DIMENSION	DESCRIPTION
COVMTX(VARSZ2,NOSUB2)	'Modified' Cholesky Factorization of the covariance matrix for each subclass
CON(NOSUB2)	Subclass constants
DET(NOSUB2)	Determinant of covariance matrix for each subclass

Field Header Record

```
WRITE(MAPTAP)(FLDINF(I),I=1,6),PTS,LINES,FLDESC,NC,
         (VERTCS(I),I=1,NC),(VERTCS(I+NC),NC=1,NC)
```

FORTRAN NAME AND DIMENSION	DESCRIPTION
FLDINF(6)	Rectangular coordinates surrounding the field classified. 1 - line start 2 - line stop 3 - line increment 4 - sample start 5 - sample stop 6 - sample increment
PTS	No. of points in the rectangular field defined in FLDINF

FORTRAN NAMES AND DIMENSION	DESCRIPTION
LINES	No. of lines in the rectangular field defined in FLDINF
FLDESC	Name of the classified field
NC	No. of vertices in the classified field
VERTCS(2,NC)	Vertices for the classified field. Vertices are ordered $(\text{sample, line})_1$, $(\text{sample, line})_2$,... $(\text{sample, line})_{NC}$

Data Record

```
WRITE(MAPTAP)ILINE,(IR(I),I=1,PTS),(VR(I),I=1,PTS)
```

FORTRAN NAME AND DIMENSION	DESCRIPTION
ILINE	Line number in reference to the multispectral scanner data tape
IR(PTS)	Subclass number to which each classified data point belongs
VR(PTS)	Likelihood that the point belongs to that subclass

End-of-Field Record

An end-of-field record has the same format as a data record with ILINE=0.

End-of-Run Record

An end-of-run record has the same format as field header record with PTS=0.

APPENDIX E

NHSTUN FILE FORMAT

APPENDIX E

NHSTUN FILE FORMAT

The interface file written to the NHSTUN unit is output by the NDHIST processor and read by the SCTRPL processor.

All records are written with a nonformatted Fortran WRITE statement. The header record is always the first file on the tape.

The format of the tape is as follows:

```
              File 1   HEADER RECORD
                       END OF FILE (EOF)

                      ⎧RECORD 1
                      ⎪RECORD 2
                      ⎪RECORD 3  (optional)
         Data file 1  ⎨RECORD 4
                      ⎪RECORD 5
                      ⎪RECORD 6
                      ⎪RECORD 7  (optional)
                      ⎩EOF
              :          :        :
         Data file N   EOF
```

The contents of each record are as follows:

<u>Header Record</u>

TOTMNS	— Total number of means computed
SIZE	— NOFET2/4
NOFET2	— Number of channels to histogram
(FETVC2(I),I=1,NOFET2)	— Actual channels to histogram
NCLRCH	— Number of color code channels
(CLRVEC(I),I=1,NCLRCH)	— Actual color code channels

C-4

Record 1

NOFLD2	— Number of fields histogrammed
NOSUB2	— Number of subclasses histogrammed
TOTVT2	— Number of vertices
NOVEC	— Number of unique vectors histogrammed

Record 2

CLSVC2	— Class name
(SUBVC2(I),I=1,NOSUB2)	— Subclass names
((FIELDS(I,J),I=1,4),J=1,NOFLD2)	— Field information
((VERTEX(I,J),I=1,2),J=1,TOTVT2)	— Field vertices

Record 3 (optional)

(MEANS(I),I=1,TOTMNS)	— Mean stats for input fields

Record 4

((PLOT(I,J),I=1,SIZE),J=1,NOVEC)	— Data vectors

Record 5

(ID(I),I=1,NOVEC)	— Class/subclass/field the data vectors belong to

Record 6

(COUNTR(I),I=1,NOVEC)	— Number of occurrences of the data vectors

Record 7 (optional)

(COLOR(I),I=1,NOVEC)	— Color codes extracted from MSS data tape

APPENDIX F

DESCRIPTION OF CLUSTER IMAGE DISPLAY
WITH COLOR KEYS

APPENDIX F

DESCRIPTION OF CLUSTER IMAGE DISPLAY WITH COLOR KEYS

The cluster image data tape output by the ISOCLS processor contains the mean vector to which each corresponding pixel was assigned during clustering and a color key. The color key is an n number of square images of 10 samples by 10 lines in dimensions. A color code square is composed of the mean vector for a given cluster. The color codes are ordered according to cluster number or greenness. The greenness ordering (G) is a function of the four Landsat channels:

$$\left.\begin{aligned} G_{i,N} &= -0.29\mu_{1,N} - 0.56\mu_{2,N} + 0.60\mu_{3,N} + 0.49\mu_{4,N} \\ G_N &= \sum_{i=1}^{M} G_{i,N} \end{aligned}\right\} \quad \text{(E-1)}$$

where

M = number of passes for multiregistered Landsat data

μ_1 = first channel of pass i

μ_2 = second channel of pass i

μ_3 = third channel of pass i

μ_4 = fourth channel of pass i

i = number of pass

N = cluster number

The number of color codes per scan line is computed by

$$K = \frac{\text{number of samples per scan line}}{11} \qquad \text{(E-2)}$$

The number of lines required to display the color codes is computed by

$$L = \left[\frac{(\text{number of clusters} - 1)}{K} + 1\right] \times 11 \qquad \text{(E-3)}$$

The clustered field and color key are separated by a scan line of zeros. Each color code square is separated by a 1-by-10 vertical line of zeros.

The data are output in LARSYS III or Universal format (see appendixes A and B, respectively).

The structure of the file is as follows.

HEADER RECORD

IRG[1]

N records — Mean vector for each corresponding pixel

IRG

(N + 1) records — Scan line of zeros

IRG

(N + 2) records — 10 lines { color code 1 (10 samples) 0 0 0 ⋮ 0 | color code 2 0 0 0 ⋮ 0 | ... 0 0 0 ⋮ 0 | color code K }

IRG

⋮

(N + L) records — color code (last cluster)

E-O-F[2]

---

[1]Inter-record gap.

[2]End of file.

APPENDIX G

SCTRUN FILE FORMAT

APPENDIX G

SCTRUN FILE FORMAT

The scatter plot image, written to the SCTRUN unit, contains two-axis colorcoded spectral plot(s) and is output in the Universal format (appendix B) by the SCTRPL processor. Each file of the multifile tape contains (1) a single scatter plot image, of which N - 1 channels are color assignments and the N*th* channel is the frequency channel, and (2) a color key, unless the color assignment is the radiance values of the output pixel.

The color key is an N number of square images dimensioned 10 samples by 10 lines. A color code square is composed of the colors assigned to a given pixel. Each color code is ordered according to its cluster association; i.e., the color code associated with cluster 1 is output first, followed by the color code associated with cluster 2, etc. The number of color codes per scan line is computed by

$$K = \frac{\text{number of samples per scan line}}{11} \qquad \text{(F-1)}$$

The number of lines required to display the color codes is computed by

$$L = \left[\frac{(\text{number of clusters} - 1)}{K} + 1\right] \times 11 \qquad \text{(F-2)}$$

The scatter plot image and color key are separated by a scan line of zeros. Each color code square is separated by a 1-by-10 vertical line of zeros.

The dimensions of the output tape are user controlled via the control cards SIZE and CHANNELS (see section 16, table 16-1, and section 15, table 15-1, respectively), where

(1) number of samples per scan lines = XSIZ

(2) number of channels = dimensions of color pixel plus the frequency channel

(3) number of scan lines = YSIZ + L

The structure of the file is as follows:

```
                              HEADER RECORD
                                  IRG[1]
YSIZ records        - scatter plot image
                                  IRG
                                   or
                                  EOF (if color key is omitted)
(YSIZ + 1) records - scan line of zeros
                                  IRG
                                  ⎧          0          0       0
                                  ⎪          0          0       0
(YXIZ + 2) records - 10 lines     ⎨ color    0  color   0  ...  0  color
                                  ⎪ code 1   :  code 2  :       :  code K
                                  ⎩          0          0       0
                                  \_________/ \________/      \________/
                                  10 samples
                                  IRG
                                   :
(YXIZ + L) records - color code (last cluster)
                                  E-O-F[2]
```

---

[1]Inter-record gap.

[2]End of file.

APPENDIX H

DOT DATA FILE FORMAT (DOTUNT)

APPENDIX H

DOT DATA FILE FORMAT (DOTUNT)

The tape, written on the DOTUNT unit, is output by the DOTDATA processor. The records are written with an unformatted FORTRAN write.

A file is output for each TYPE of field(s). The file consists of the following records:

Repeat for each TYPE	Rec. No. 1, Rec. No. 2	field information
	Rec. No. 3	data record
	E-O-F	

TYPE 1 consists of labeling dots; TYPE 2 consists of bias correction dots.

Rec. No. 1

```
WRITE (DOTUNT) NOCAT, NOFEAT, NOFLD, TOTVRT, TOTDOT, NOSUN,
               (CATNAM(I), I=1, NOCAT), SIZE
```

Parameter	Dimension	Definition
NOCAT	1	Number of category names.
NOFEAT	1	Number of channels.
NOFLD	1	Number of fields.
TOTVRT	1	Number of vertices.
TOTDOT	1	Number of dots.
NOSUN	1	Number of sun angles.

Parameter	Dimension	Definition
CATNAM	NOCAT	Array containing the category names.
SIZE	1	4 + NOFEAT

Rec. No. 2

```
WRITE (DOTUNT) (FETVEC(I), I=1, NOFEAT)
               ((FLDSAV(I,J), I=1,4), J=1, NOFLD),
               ((VERTEX(I,J), I=1,2), J=1,TOTVRT),
               (ANGLE(I), I=1, NOSUN)
```

Parameter	Dimension	Definition
FETVEC	NOFEAT	Array containing the channel numbers.
FLDSAV	(4,NOFLD)	Array containing the field description.
VERTEX	(2,NOFLD)	Array containing the field vertices.
ANGLE	NOSUN	Array containing the sun angles.

Rec. No. 3

```
WRITE (DOTUNT) ((DOTS(I,J), I=1, SIZE), J=1, TOTDOT)
```

Parameter	Dimension	Definition
DOTS	(TOTDOT, SIZE)	Array containing the dot information. DOTS(1,I)=sample number for dot i. DOTS(2,I)=line number for dot i.

Parameter	Dimension	Definition
		DOTS(3,I)=type number for dot i. DOTS(4,I)=category number for dots i (optional). DOTS(5,I) ⋮ =dot vector i. DOTS(4+NOFEAT,I)

APPENDIX I

BASELINE RUN

APPENDIX I

BASELINE RUN

This run is published in volume I, section 20, and is not repeated here.

APPENDIX J

REFERENCES

APPENDIX J

REFERENCES

1. Ratcliff, M. L.: Description and User's Guide for a Processing System for Airborne Multispectral Scanner Data. LEC/Houston Aerospace Systems Division (HASD), NASA/JSC Internal Note 0C0010 (Houston), Oct. 1970.

2. Baker, K.: Algorithm Documentation for CLASSIFY. NASA/JSC informal memorandum (Houston), June 1974.

3. Minter, R. T.: Computer Program Documentation, ISOCLS, Iterative Self-Organizing Clustering Program. LEC/HASD, NASA/JSC program MSC-C094 (Houston), Oct. 1972.

4. Baker, K.: User Guide to University of Houston Linear Feature Selection Program. NASA/JSC informal memorandum (Houston), Oct. 1973.

5. Quirein, J. A.: Requirements Document to Incorporate Linear Feature Selection Techniques Into the Earth Resources Interactive Processing System. LEC-1403, LEC/HASD, NASA/JSC (Houston), Nov. 1973.

6. Fletcher, R.; and Powell, J.: A Rapidly Converging Descent Method for Minimization. British Computer J., 1963, pp. 163-168.

7. Quirein, J. A.: An Iterative Approach to the Feature Selection Problem. Proc. Conf. on Machine Processing of Remotely Sensed Data, Purdue Univ. (W. Lafayette, Ind.), Oct. 16-17, 1973.

8. Hallum, C. R.; and Minter, T. C.: Maximum Likelihood Classification by Thresholding. LEC/HASD, NASA/JSC Technical Report 640-TR-114 (Houston), June 1972.

9. Feiveson, A. H.: Classification by Thresholding. NASA/JSC Internal Note 08447 (Houston), Sept. 1973.

10. Eppler, W. G.: Empirical Distribution of Quadratic Form Used for Thresholding. LEC/HASD FSD-0001, NASA/JSC (Houston), Nov. 1972.

11. Van Rooy, D. L.; and Lynn, M. S.: Use of Spatial Information in Classification of Remotely Sensed Data. Rice Univ. (Houston), May 1973.

12. Earth Resources Data Format Control Book. NASA/JSC Technical Report PHO-TR543 (Houston), Mar. 1975.