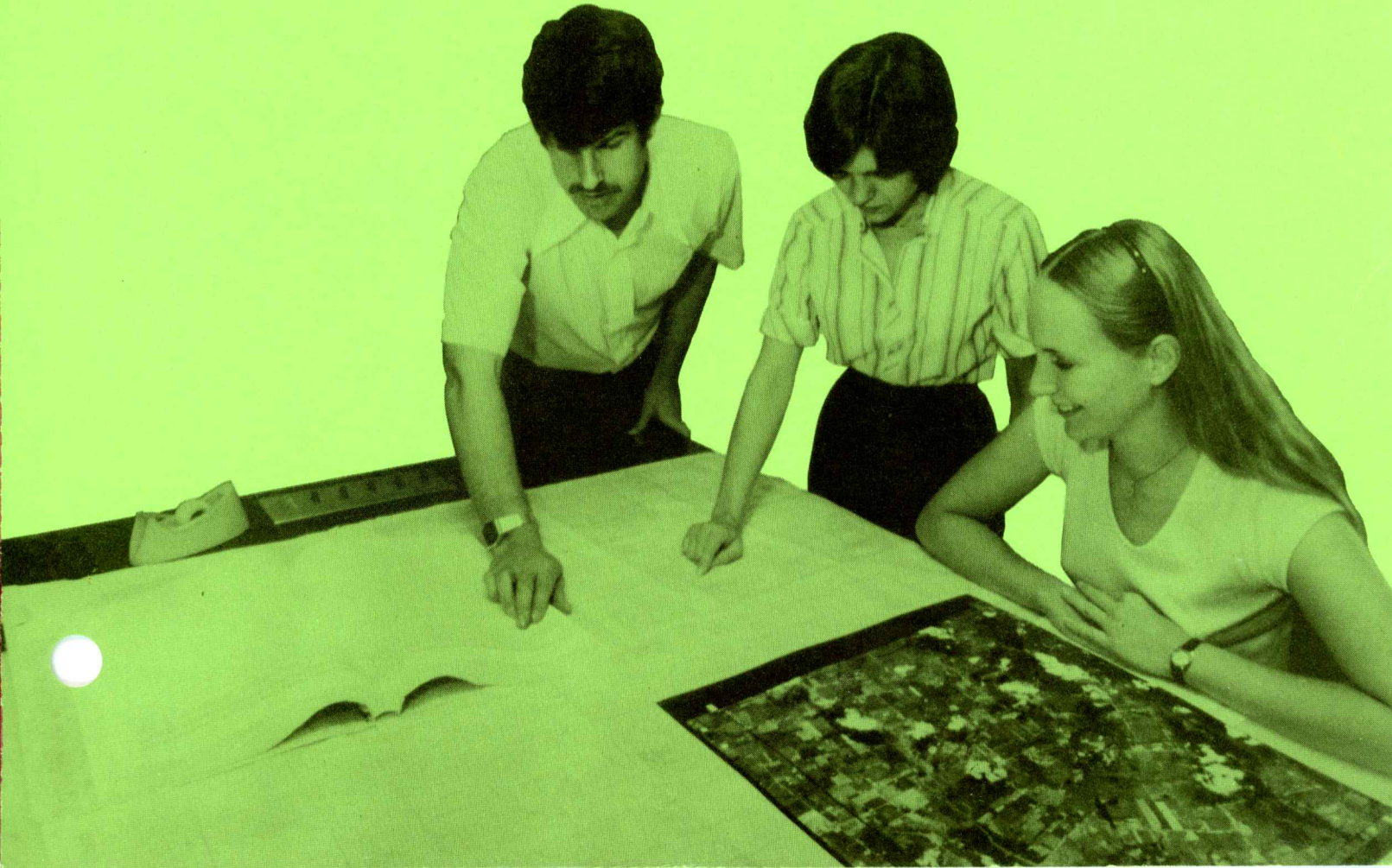
LARS Publication 010482

Flexible Workshop on Numerical Analysis of Multispectral Image Data

James C. Tilton
Luis A. Bartolucci

Purdue University
Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA



Flexible Workshop on Numerical Analysis of Multispectral Image Data

by

James C. Tilton
and
Luis A. Bartolucci

Page ii is blank

TABLE OF CONTENTS

PREFACE

For many decades, man has collected and analyzed remotely sensed data in order to learn more about the earth's surface. Since the development of the multispectral scanner in the 1960s, such remotely sensed data increasingly has been collected and stored in the digital format of the multispectral scanner. This digital format makes it convenient to enlist the computer as an analysis aid. In fact, with the large volume of data produced since 1972 by Landsat and other satellite-based multispectral scanners, the computer has become a necessary part of the analysis process simply because of the amount of data involved. We would still want to use the computer, though, even if the volume of data were not a consideration. A computer can help an analyst more effectively exploit the full information content of the data.

Several computer-assisted analysis techniques are based on statistical pattern recognition. This workshop introduces you to these statistical pattern recognition techniques through a series of exercises that simulate an analysis of multispectral scanner (MSS) data. The purpose of this workshop is NOT to train you to be analyst, but rather to give you an overview and understanding of how MSS data are analyzed. You will gain this understanding through learning about MSS data characteristics and statistical pattern recognition techniques and by working through the decision process used by an analyst.

Each chapter in this manual consists of a general discussion of a particular analysis step followed by a specialized case study which illustrates the processes described in the general discussion. The case study analysis problem is a computer-assisted analysis of MSS data selected from several available case studies. This case study is presented as an example of how numerical analysis might be performed on a particular data set. It should not be thought of as the only way the analysis could have been done.

While Landsat data is mentioned frequently in this manual, the general discussion is written so as to apply to multispectral scanner data from any source, though slanted towards satellite-based sources. The featured case study may be based on Landsat data or data from another type of multispectral scanner. Whenever there are special considerations associated with a non-Landsat data set, these are noted in the appropriate case study sections.

The case study featured in this manual was performed using the LARSYS numerical analysis system. Also, several LARSYS processors are described in the general discussion sections as illustrations of analysis procedures. We use LARSYS here only as

an example of a numerical analysis system. Similar analyses can be carried out using any of several other remote sensing analysis packages.

You will not actually use a computer during these workshops. Someone else has already run the required computer programs to analyze the data. But you will look at the computer inputs and outputs from an entire analysis sequence and learn to make the necessary analysis decisions. Your instructor will provide you with a binder containing computer printouts of the case study analysis.

The intended audience for this workshop are persons who have a basic background in remote sensing. This remote sensing background can be gained by means of the following educational materials or their equivalent:

"An Introduction to Quantitative Remote Sensing,"
by J. C. Lindenlaub and James D. Russell,
LARS Information Note 110474, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907.

Fundamentals of Remote Sensing Minicourse Series*. In particular, the following units:

-Remote Sensing: What Is It?
-The Physical Basis of Remote Sensing

-Spectral Reflectance Characteristics of Vegetation
-Spectral Reflectance Characteristics of Earth Surface Features

-Multispectral Scanners
-Interpretation of Multispectral Scanner Images

-Pattern Recognition in Remote Sensing
-Typical Steps in Numerical Analysis

Introduction to Quantative Analysis of Remote Sensing Data Videotape Series*. In particular, the following units:

-The Remote Sensing Information System
-The Role of Pattern Recognition in Remote Sensing

---

* The Minicourse Series and the Videotape Series may be obtained from Continuing Education Admininstration, 116 Stewart Center, Purdue University, West Lafayette, Indiana 47907.

A basic background may also be obtained by studying the introductory portions of one of the following texts:

Remote Sensing of Environment
J. Lintz, Jr. and D. S. Simonett, eds.
Addison and Wesley Publishing Company, Reading, MA, 1976.

Remote Sensing: The Quantitative Approach,
P. H. Swain and S. M. Davis, eds.,
McGraw-Hill Book Company, New York, NY, 1978.

Remote Sensing and Image Interpretation
T. M. Lillesand and R. W. Kiefer,
John Wiley and Sons, New York, NY, 1979.

These texts are also useful supporting references for this workshop manual.

This workshop manual is based upon materials presented in Workshop Series on Numerical Analysis of Remotely Sensed Data by Ronald K. Boyd and John C. Lindenlaub. The authors wish to thank Shirley M. Davis, Joan S. Buis and Roger M. Hoffer for their thoughtful editorial comments and content suggestions throughout the development of this manual.

CHAPTER I. INTRODUCTION

The numerical analysis of multispectral scanner (MSS) data is a dynamic process which requires interaction between man (analyst) and machine (computer). The process involves meshing the experience and insights of the analyst with appropriate computer programs to extract the maximum amount of information from the data. Numerical analysis techniques allow detailed study of digital data and have been shown to be cost effective in many cases. A typical analysis sequence is shown in Figure I-1. Even though it is shown here as a linear process, all of the steps are interrelated. At any step in the analysis, an interpretation of the results of that step can lead to a return to a previous step for revision of the procedure.

Numerical analysis of MSS data is to a large extent an art. While there are several general analysis guidelines, there is no single set procedure that must be followed for all data sets. Each analyst will analyze a particular data set in a slightly different way than another analyst would. In addition, different data sets will be analyzed somewhat differently even by the same analyst, depending on the characteristics of the data set and what information is desired as output from the analysis.

Chapters II through VI discuss each analysis step in turn. Before we can discuss these analysis steps, though, we should first review certain general characteristics of MSS data.

A multispectral scanner (MSS) gathers radiance data in various sections of the electromagnetic spectrum (wavelength bands). The number of these wavelength bands and the coverage of the electromagnetic spectrum depend on the particular scanner. For example, the Landsat MSS has four wavelength bands covering most of the visible and near infrared spectrum. The MSS aboard the Skylab space station had 13 wavelength bands covering most of the visible, near infrared and middle infrared spectrum, and a portion of the far (thermal) infrared spectrum.

Each MSS has a particular spatial resolution, i.e., for each scanner there is a limit to the smallest object that can be detected. This spatial resolution, which is determined by the scanner's optics and altitude, controls the resolution cell size or ground area coverage of each picture element (pixel) in the data. The Landsat MSS resolution cell is approximately 56m by 79m. The Skylab MSS had a spatial resolution close to that of the Landsat MSS. The Thematic Mapper aboard Landsat D has a resolution cell of 30m by 30m.

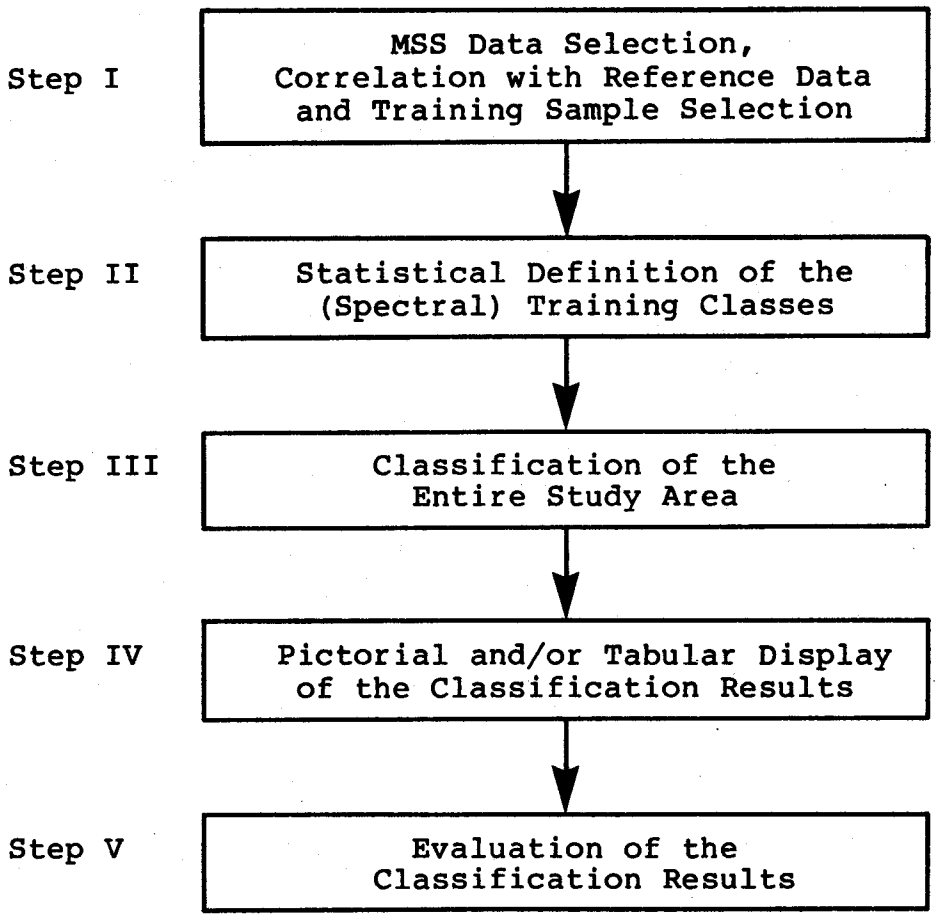


Figure I-1. Flowchart of analysis steps for classification of multispectral scanner data.

Calibration signals are recorded with each scan line collected by MSSs. A separate set of calibration signals is recorded for each MSS detector. These calibration signals are used for removing inconsistencies in the MSS electronics during data collection and for converting the data into radiometric units. Data values which are converted to radiometric units can be compared across wavelength bands and can be compared to data collected at a different time, whether or not it is collected by the same sensor. In the visible and near infrared spectral regions, calibration is often provided by the black background of the interior of the scanner, a reference lamp and sunlight. For thermal infrared wavelengths, calibration is provided by standard cold and hot thermal plates. The Landsat MSS uses only background and incandescent lamp references for calibration. The calibration of the Thematic Mapper is provided by three tungsten lamps which are switched on and off so as to provide eight calibration levels from background to full-scale illumination.

The above description of the spectral, spatial and radiometric characteristics of MSS data is intended only as a brief review of multispectral scanners. It is by no means comprehensive. For a more complete description of MSS data and systems see sections 2.6, 2.7 and 2.8 of Swain and Davis, Remote Sensing: The Quantitative Approach.

CHAPTER II. MSS DATA SELECTION, CORRELATION WITH REFERENCE DATA AND TRAINING SAMPLE SELECTION

Upon completion of this chapter, you should be able to:

1. State at least one reason why the quality of the data being considered for analysis must be evaluated.
2. Name at least two sources of data quality information.
3. Name at least three data quality problems that might hinder analysis.
4. Describe two types of geometric corrections that might aid in the analysis of Landsat data.
5. State one reason for correlating multispectral scanner data with reference data.
6. List at least four types of reference data.
7. Describe the correlation of ground features apparent in the multispectral scanner data with the same features in reference data.
8. State why a training sample is needed.
9. Contrast the supervised and nonsupervised approaches to obtaining a training sample.
10. Name at least two considerations that affect the selection of training areas in the hybrid approach to training sample selection.

The first step in multispectral classification begins with the selection of the MSS data set. The data set must be of sufficient quality and must be from an appropriate time of year so that the land cover classes of interest can be identified with the desired accuracy. Otherwise, no matter how carefully the ensuing analyses are carried out, we will not be able to produce the desired results. Next we must locate our study area in the MSS data and correlate our MSS data with the available reference data. Once this is done, we can use the reference data to help select a representative training sample from the MSS data. This training sample will be used later to statistically define the (spectral) training classes that will be used for training the classifier.

SELECTING THE DATA SET

Before we can select our MSS data set, we must consider characteristics of the land cover classes of interest that might limit our selection of data sets. Certain land cover classes may appear only during certain times of the year, and other land cover classes are easier to spectrally distinguish from other land cover classes in certain times of the year than other times. For example, agricultural crops can only be detected during the period from planting through harvest, coniferous trees can be identified most easily during the winter (when deciduous trees are leafless), and soils may be studied most easily during fall (after harvest and before the snows) and during spring (after the snow melt and before crop emergence).

After we have limited our MSS data selection by time of year, we must consider the general quality of the remaining data sets. This includes the actual radiometric quality of the data as well as possible problems with snow and cloud cover in the imagery.

A rough idea of data quality and cloud coverage for Landsat data sets from any specified geographic area can be obtained by scanning computer listings provided by the EROS Data Center (Sioux Falls, South Dakota). The EROS Data Center examines all the Landsat data received at the facility and rates the channel-by-channel quality of each data set on a scale of 0 to 9, with 9 indicating the highest quality. Rough estimates of the percentage of cloud coverage are also included. If you are using non-Landsat data, inquire about the availability of similar computer listings. With this type of information we can choose, for closer examination, data sets of acceptable quality and cloud coverage from the time of year most suitable for discriminating among the cover types of interest in our study area.

Once we have eliminated the most obviously unsuitable data sets from consideration by the above process, we can perform a more refined evaluation of the remaining data sets by directly examining the data in pictorial form. For Landsat data, we can obtain from the EROS Data Center black-and-white photographs made from each band of each data set to be examined. If you are using a non-Landsat data source, inquire about the availability of similar photographs made from the data sets of interest.

Gross data characteristics, including cloud cover, are obvious in a pictoral representation of the data. Figure II-1 shows an example of a photographic image made from a Landsat data set from the Chicago, Illinois area. Cloud cover shows up clearly over the eastern portion of this image. This image also points out a problem with the cloud cover percentage estimates given by the EROS Data Center computer listings. If we want only to study the immediate Chicago area, this image has an effective 0% cloud cover. The cloud cover percentage given by EROS for this scene is 20%.

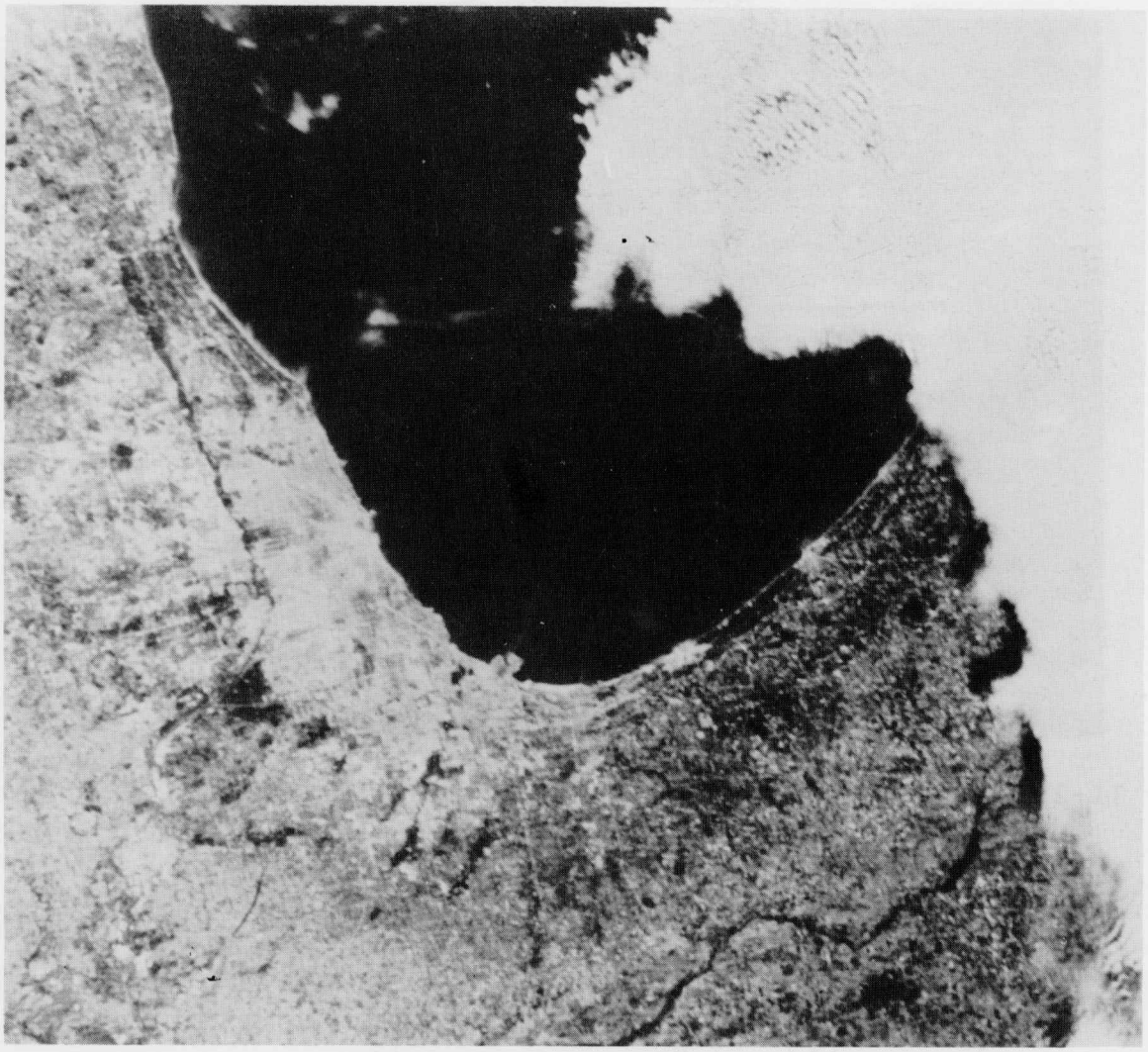


Figure II-1. Landsat scene over Chicago, Illinois and the surrounding area with clouds obscuring the east side of the scene.

A dramatic example of a data quality problem is shown in Figure II-2. In the Landsat scanner system, six lines of data are collected in each wavelength band each time the scanning mirror oscillates. A separate set of detectors is used for each of these scan lines. If these detectors and their associated electronics are not properly matched or calibrated, a striping effect such as shown in Figure II-2 may be noticeable in the imagery of one or more bands. The table below Figure II-2 shows the mean and standard deviation of the output of each of the six Band 4 detectors over the entire image shown in Figure II-2. Notice that the mean value for detector 3 is very low compared to that of the other detectors. Apparently a malfunction occurred in the detector electronics, resulting in the striping

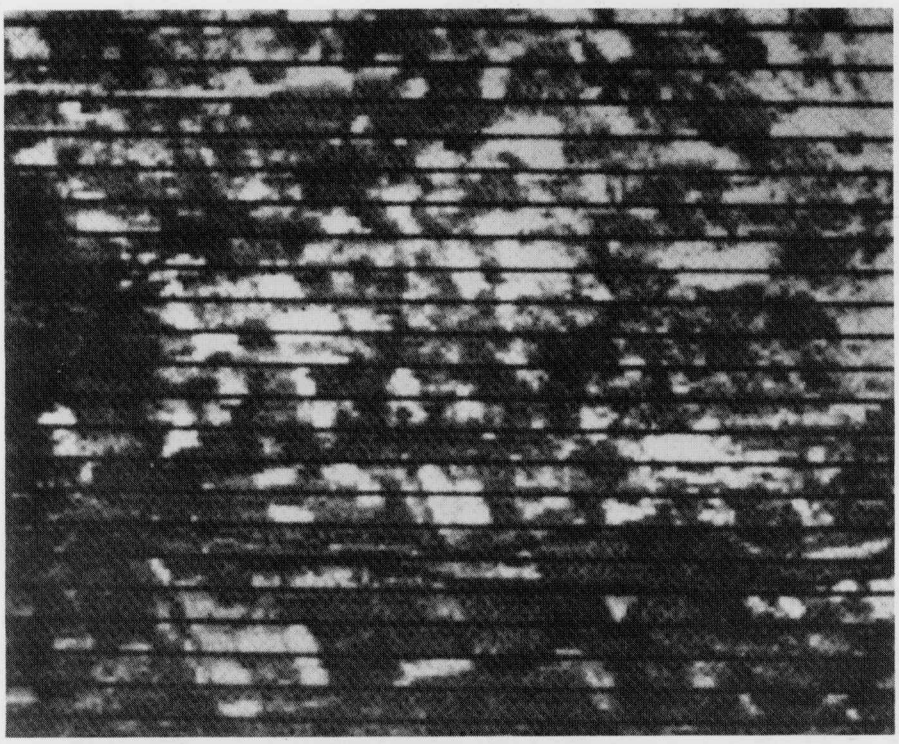


Figure II-2. Striping effect in imagery.

Detector	Mean	Standard Deviation
1	21.9	3.21
2	21.8	3.07
3	7.0	3.13
4	21.5	3.13
5	20.9	3.11
6	21.9	3.03

illustrated in Figure II-2. Other data quality problems sometimes encountered are dropped, bad or noisy data lines, also usually caused by malfunctions in the detector electronics. These data quality problems can be compensated for by various computer image restoration techniques. Presently the EROS Data Center processes most Landsat data to compensate for these radiometric and geometric problems.

After we have examined a pictorial representation of each candidate data set for data quality and cloud coverage, we can order computer compatible tapes (CCTs) for one or more of the data sets for computer analysis. Note that we do not have to obtain the relatively expensive CCTs until after we select the data set(s) which we will use in the analysis.

Now that we have our CCT for our MSS data set, we will want to display all or portions of the data in pictorial form. We do this to double-check data quality and to assist in the rest of the data analysis process. We can display our data by using a cathode ray tube device (CRT digital display), a line or dot-matrix printer/plotter, or an optical film writer.

LOCATING A STUDY AREA AND CORRELATING IT WITH REFERENCE DATA

Now that we have chosen the best available data set for analysis, we need to locate the study area within the data and correlate any reference data we may have (aerial photography, maps, etc.) with the multispectral scanner (MSS) data. This process is made easier by using MSS data which have been geometrically corrected and scaled to allow overlaying directly on any map-like reference data.

Imagery generated by uncorrected Landsat data exhibit a phenomenon where rectangular objects on the ground appear as skewed parallelograms. This skewing is caused by the rotation of the earth beneath the satellite. Since the Landsat orbit is not oriented exactly over the north pole, an uncorrected Landsat image also is rotated from a north and south orientation (about 12 degrees at 40 degrees north latitude). Both skew and rotation can be corrected by a geometric transformation of the data performed by a computer. An example of this is shown by Figure II-3 where Landsat data are displayed in image form both before and after geometric correction.

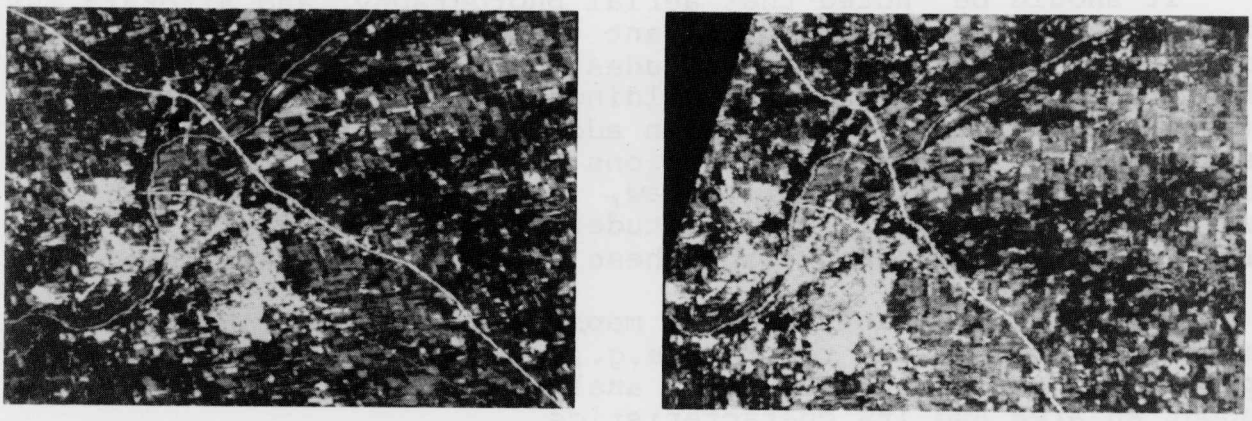


Figure II-3. Landsat image data before and after processing to remove the effects of the earth's rotation.

During geometric correction, the data may also be rescaled. The MSS data can be scaled to correspond to USGS topographic maps

(1:24,000) or any other scale desired to allow the MSS data to be overlaid on map reference data. While the uncorrected data are adequate for some analysis tasks, using geometrically corrected data simplifies the analyst's job of locating ground features since the data can be overlaid on map reference data and more easily compared with map-like reference data. Comparing map or map-like reference data with geometrically corrected MSS data is a convenient and effective way to locate the study area in the MSS data set.

Map-like reference data can come in several different forms. One such form of reference data is aerial photography. Photography can be collected at various altitudes, resulting in reference data at a range of mapping scales. In general, as a plane flies higher (or as the camera focal length decreases), each photograph will cover a larger area, but less detail will be discernible. Another variable in aerial photography is film type. Black-and-white panchromatic or infrared film, color and color infrared film all provide different kinds of information about a ground scene and can serve as reference data for an analyst who understands how to interpret photographic images.

Aircraft multispectral scanner data can also serve as map-like reference data for an analyst working with satellite data. Data collected by aircraft often provide more detailed information about the spectral characteristics of portions of a scene than the satellite data. For example, on aircraft-borne sensors there may be more spectral bands available and the data may be available at greater spatial and spectral resolution.

It should be noted that aerial photography and aircraft MSS image data may have significant distortions from strict map quality. At the lower altitudes these data are collected at features such as high-rise buildings and mountains distort the image away from map quality. In addition, aircraft multispectral scanner data will have distortions caused by unwanted aircraft motions such as pitch, roll, yaw, cross-path translations, and variations in velocity and altitude. We must be careful to allow for these problems when using these kinds of reference data.

Maps (e.g., county highway maps or U. S. Geological Survey maps) and historical records (e.g., past crop yields or weather patterns) can be useful to an analyst in providing information about an area and its characteristics.

Another source of information is observation at the scene by the analyst or other personnel. This is usually referred to as "field checking" or "ground truthing." Ground truthing may include collection of soil moisture samples, identification of crop varieties, determination of biomass, or other detailed observations. These observations can provide the key to successfully relating the spectral responses in the data to the cover types on the ground.

SELECTION OF THE TRAINING SAMPLE

The next step in this analysis of multispectral scanner data is the selection of the training sample. In order to understand what a training sample is and why it is needed, we need to review some pattern recognition concepts. The pattern recognition algorithms we will deal with require that examples of typical data from each class of interest be supplied to the computer programs. These data, called a training sample, are used to set certain parameters for the pattern recognition algorithms, in effect "training" the computer to recognize the classes of interest. When the classification operation is carried out by the pattern recognition algorithms, each data point to be classified is "compared" to the training sample for each class, and the pixel is assigned to the class it most closely resembles.

There are two major methods of obtaining a training sample. The first, referred to as the supervised approach, involves locating regions of pure pixels, each region representing a single cover type. We locate in the reference data several of these regions for each cover class, and find the corresponding regions in the multispectral data. The training sample for each class is formed by grouping together the data from all regions identified in the multispectral data as belonging to that ground cover class. The second, referred to as the nonsupervised approach, does not utilize reference data to select pixels to form the training sample. Rather a systematic or random sample of pixels is selected from the study area. This sample is then analyzed to identify groups of pixels that are spectrally similar. The cover type identity of each pixel group is established by comparing the pixels in each group to the reference data.

To summarize, in the supervised approach we specify the ground cover classes of interest and locate a sample of each of these classes in the multispectral data. In the unsupervised approach we analyze a systematic or random sample of the multispectral data to identify spectrally similar classes. We then assign ground cover class labels to these spectrally defined classes.

The supervised approach has as its strength the guarantee that ground covers and regions of most interest are represented in the training sample. We can assure this representation through strategic selection of the regions of pure pixels. A problem with this approach is that even a carefully selected training sample may not reflect the true spectral variability of each ground cover class. Classification accuracy may be reduced if our training samples aren't truly spectrally representative of each ground cover class. Another problem is that, for most data sets, many of the pixels are not from pure cover classes. When using the supervised approach, we may not be able to produce an accurate classification when a substantial number of mixed pixels

occur in a data set since the computer is not trained to identify mixed pixels with this approach.

The nonsupervised approach has as its strength the ability to group spectrally similar pixels regardless of their spatial positions. This can be advantageous when working with a heterogeneous scene in which the likelihood of observing several adjacent pixels of similar spectral response is low. An example of such a scene would be an area of mountainous terrain with many combinations of slope, aspect, elevation and vegetation. A problem with the nonsupervised approach is that important but infrequently occurring cover types could be missed by the systematic or random sample used in this approach.

A common mistake is to assume that because points are spatially adjacent and belong to the same cover type, they are spectrally similar. This is not necessarily the case. If we think of a corn field, experience tells us that within that field there may be low spots where the corn is greener and has greater ground cover and high spots where there is slight moisture stress and the corn is therefore thinner and more soil is "seen" by the remote sensing system. (In exceedingly wet years, the reverse effect may occur.) There are many other reasons for spectral variability within a cover type, and many cover types may be represented by two or more groups of pixels that are spectrally different. We can use the nonsupervised approach to identify this spectral variability in each ground cover class.

There is another approach to obtaining a training sample which combines aspects from both the supervised and nonsupervised approaches. As in the supervised approach, the training sample is defined by selecting regions of pixels as guided by reference data. However, in this hybrid approach these regions, called training areas, are selected to contain several cover types. Once the training sample is defined in this way, the sample is analyzed, as in the nonsupervised approach, to identify groups of pixels that are spectrally similar. The cover type identity of each pixel group is established by comparing the pixels in each group to the reference data.

The hybrid approach allows us to take advantage of the strengths of both the supervised and nonsupervised approaches while avoiding the problems inherent in these approaches. We can choose the training areas so that all cover types and regions of interest are included in the training sample, as in the supervised approach. But since the training areas are then processed to identify spectrally similar pixels within the area, as in the nonsupervised approach, spectral variability within particular cover types is also identified.

The first step in obtaining a training sample in the hybrid approach here is the selection of training areas. Experience has indicated that the task of identifying spectral groups of

pixels is made easier by choosing large training areas since this facilitates correlation with the reference data. The computer system being used, however, puts an upper limit on the size of the training areas. For LARSYS the limit is 25,000 pixels for one channel data or 40,000 pixels - divided by the number of channels - for multichannel data. For example, four channel data have a limit for a single training area of 10,000 pixels, or a square 100 lines by 100 columns.

A common procedure for selecting the training areas is to use in the available reference data to identify areas that contain the information classes of interest. These areas are then also identified in imagery formed from the multispectral scanner data. With this information in mind, we then select three to six training areas (five to ten in highly variable data sets) so that each area includes more than one cover type, and every cover type is included in at least one (preferably two or more) training areas. If only three training areas are selected, each area should be as large as possible given the constraints of your computer system. If more training areas are selected, each area can be a little smaller. An additional useful criterion is to include identifiable landmarks in each training area to facilitate correlation between the image data and the reference data. To help ensure that the training sample is representative, the training areas should be distributed somewhat uniformly throughout the area to be classified; however, this may not be possible if reference data is limited.

Self-Check

1. Explain in your own words why examination of data quality is necessary.

2. Name at least two ways in which the we can examine data quality.

3. Name at least two types of data quality problems we might find in Landsat data.

4. What two types of geometric correction aid in the analysis of Landsat data?

5. State one reason for correlating multispectral scanner data with reference data.

6. List at least four kinds of reference data.

7. State why a training sample is needed.

8. Name two considerations that should go into the selection of training areas in the hybrid approach.

CHAPTER III. STATISTICAL DEFINITION OF THE (SPECTRAL) TRAINING CLASSES

Upon completion of this chapter, you should be able to:

1. Describe at least two tasks a clustering algorithm can perform.
2. State the rule-of-thumb used to determine the number of clusters to request and give the reason behind it.
3. State why an analyst associates cluster classes with information classes.
4. Given printed cluster maps and available reference data, describe how to associate cluster classes with information classes.
5. Give an example of when we should augment the candidate training classes obtained from clustering the training sample.
6. Describe at least one method of visualizing the spectral characteristics of candidate training classes.
7. Describe the location of water, green vegetation, and bare soil on a bi-spectral plot.
8. Given two pairs of one-dimensional density functions, identify the pair which is separated by the larger statistical distance.
9. Name the two characteristics of Gaussian probability density functions which determine the statistical distance between the density functions.
10. Name two desirable characteristics of training classes.
11. Given a separability diagram and a list of the identities of each spectral class, describe how to select a set of training classes to use in the classification.
12. Discuss the utility and limitations of statistical distance measures for predicting the accuracy classifying with any set of training classes.

In the previous chapter we described how to select a training sample by using either the supervised, nonsupervised or hybrid approach. Now we will discuss how to use this training sample to define (spectral) training classes.

Under the supervised approach, the training classes are automatically defined through the process of selecting a training sample for each ground cover (information) class. All we need do before we classify the study area is characterize the training classes so that our classification algorithm can use them. Depending on the type of classifier used, this characterization is usually in terms of the mean vector and covariance matrix of each training class, or just the mean vector. The mean vector is composed of the averages of the data values in each wavelength band. The covariance matrix is a multivariate generalization of variance and is used to characterize the multidimensional "spread" or dispersion of the data in each training class sample.

Under the nonsupervised and hybrid approaches, we must analyze the training sample to find groups of pixels that are spectrally similar and that can be used as spectral training classes. First we cluster the training sample to define candidate training classes. (In the hybrid approach each training area is usually clustered separately.) Then we associate each candidate training class with an information class. If we find that there are information classes that are not represented by at least one candidate training class, we may augment our candidate training classes at this point by adding an appropriate training sample for each such information class using the supervised approach. Next we study the candidate training classes to see how they are related to each other spectrally. We can do this by plotting the statistical characteristics of the candidate training classes and by calculating statistical distances between the classes. Final spectral training classes are then defined by deleting and/or pooling the appropriate candidate training classes. All these steps are described in detail in the following sections.

CLUSTERING THE TRAINING SAMPLE

The first step in defining spectral training classes from the training sample is clustering, i.e., identifying natural groupings of pixels in the sample. Each pixel in the training sample has a location in the multispectral measurement space relative to all of the other pixels. Pixels naturally tend to occur in groups or clusters in this space as illustrated in Figure III-1.

Natural clusters can be visually detected only for data sets which are one-, two- or three-dimensional. To cluster higher dimensional data, and to save us from a possibly tedious task otherwise, we need to use a computer.

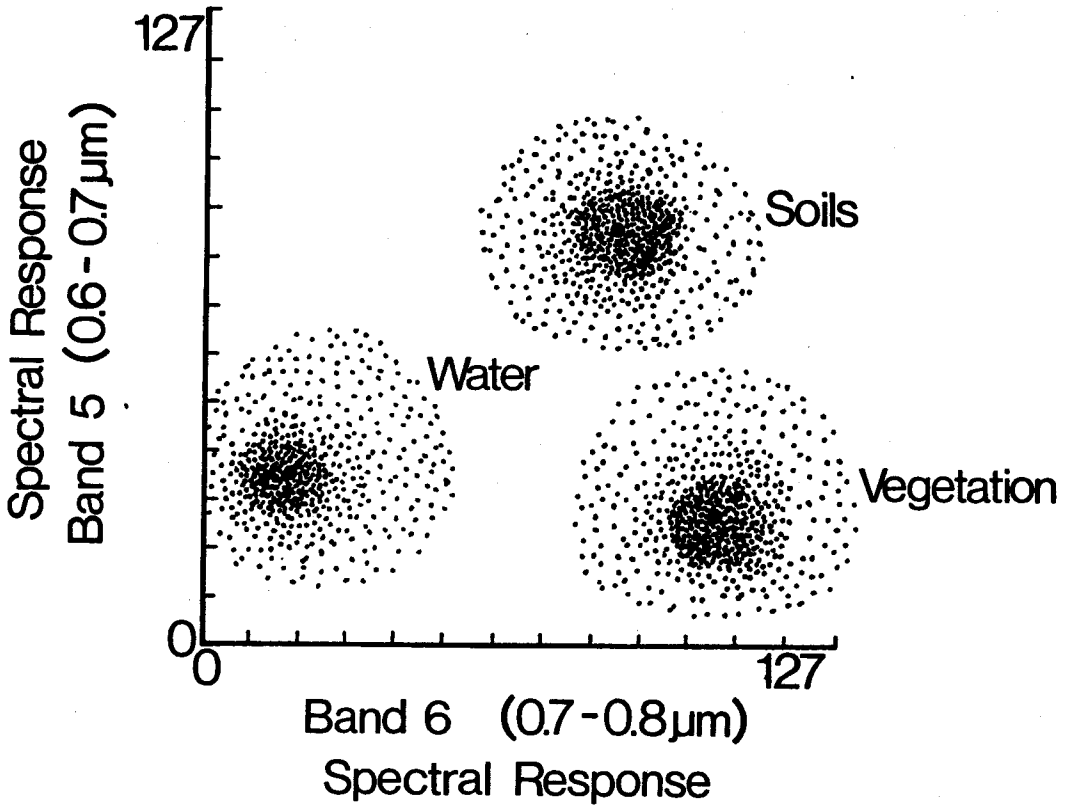


Figure III-1. Typical plot of the spectral responses of water, soils, and vegetation in Landsat bands 5 and 6. Note how the spectral responses of each one of these major ground cover types form "clusters" around a common center.

How can a computer cluster the training sample? In LARSYS, the CLUSTER processing function uses the following approach for clustering the training sample (or each training area): The analyst specifies the number of clusters desired and the processor designates the location of each initial cluster center in multidimensional space (see Figure III-2). Then the processor calulates the multidimensional Euclidean distance between each data point in the training sample and each cluster center, and assigns each pixel to the cluster with the nearest cluster center. New cluster centers are then determined by the processor by calculating the mean vector for the data points assigned to each original center. The computer then recalculates the multidimensional distance between each data point and the new cluster centers and reassigns each sample to the closest newly defined cluster center. The computer continues the cycle of calculating the cluster centers and reassigning data points until the percentage of data points that are not reassigned to a new cluster center reaches a value known as "convergence." Specifying a convergence value less than the default value of 100%, for example 98.5%, will result in a significant saving of

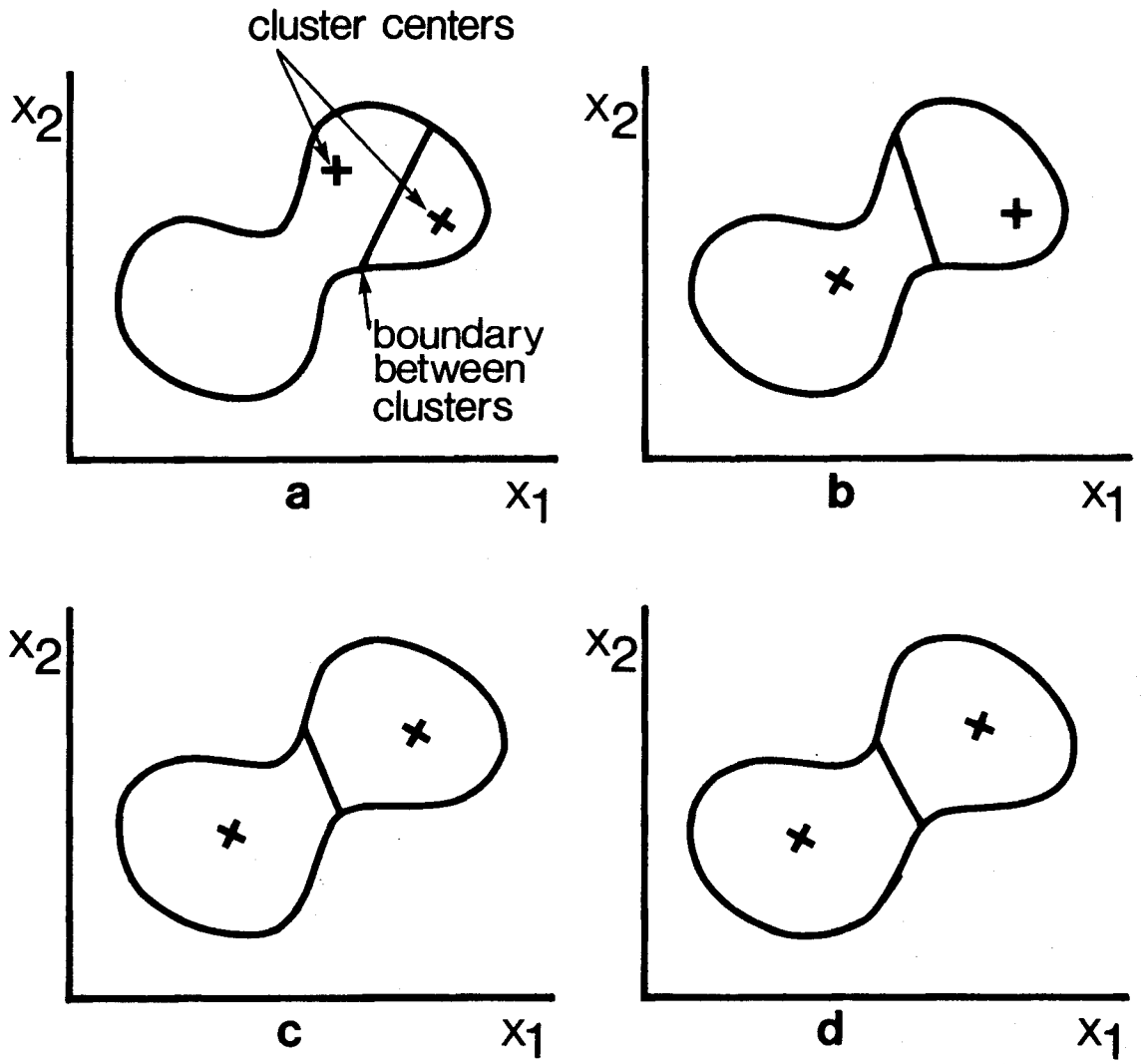


Figure III-2. Sequence of clustering iterations: (a) Initial cluster centers, (b) and (c) Intermediate steps, (d) Final center configuration.

computer time without seriously affecting the clustering results. See section 3.10 of Swain and Davis for more details.

A clustering algorithm such as CLUSTER is called a nonsupervised classifier because it groups pixels strictly on the basis of their multi-channel data values. Neither the location of the pixels relative to one another (spatial information) nor ground cover type is considered in determining the clusters. Rather, such an algorithm groups those pixels with similar spectral response values in the multiple channels. These natural groupings in the training sample are called cluster classes or spectral classes. In the nonsupervised and hybrid approaches, we use these cluster classes as our original candidate training classes.

When data from a ground scene are clustered, there is a tendency for the data points within each cluster class to be distributed in a Gaussian fashion. Figure III-3 shows a typical Gaussian function in one dimension -- commonly called a "bell-shaped" or "normal" curve. Figure III-4 shows a two-dimensional Gaussian density function. The fact that clusters in remotely sensed data tend to be Gaussian is important because several of the classification algorithms used are based upon a Gaussian assumption, i.e., that the distribution of the data values within each of the classes to be classified can be approximated by a Gaussian density function.

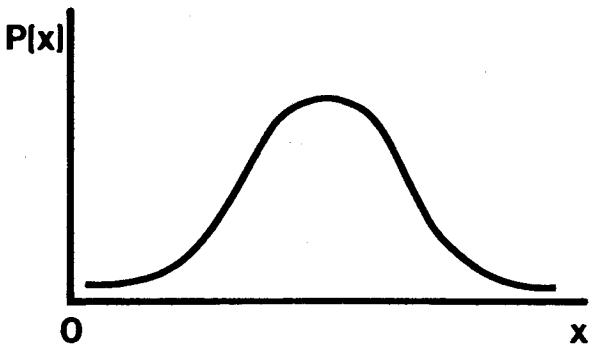


Figure III-3. Gaussian density function in one dimension.

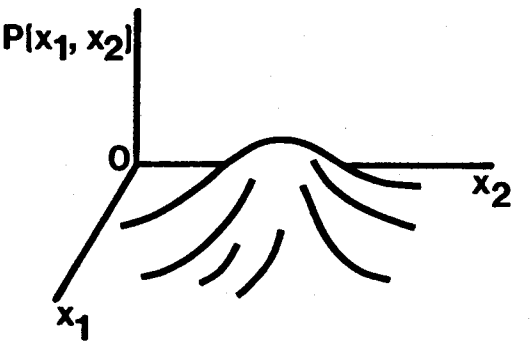


Figure III-4. Gaussian density function in two dimensions.

Often more than one Gaussianly distributed cluster is necessary to represent an information (ground cover) class. As an example, an agricultural crop might exhibit a multimodal distribution (more than one peak) due to different soils, moisture content, planting dates, crop density, seed varieties, or a combination of these factors. If necessary to satisfy the Gaussian assumption, the multimodal non-Gaussian density function in Figure III-5 could be decomposed into two Gaussian components by clustering, as shown in Figure III-6. These components are commonly referred to as "spectral subclasses." The subclass concept is an important one as it allows us to use a classification algorithm based upon a Gaussian assumption even though the information class distributions may be non-Gaussian. This is because the classification algorithm is based on the spectral classes rather than the information classes.

When using the CLUSTER processing function, we must specify the number of clusters that the data is to be grouped into. Experience has indicated that most cover types contain at least two subclasses. A rule-of-thumb is for us to request at least two times the number of information classes present. If our information classes are very general, we should request at least

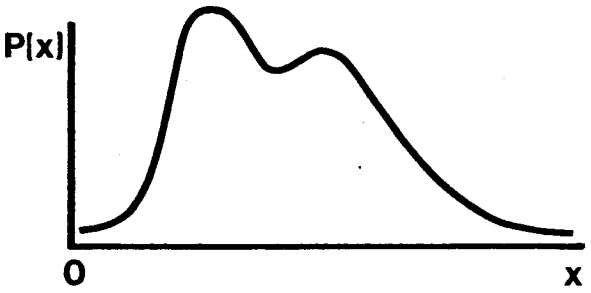


Figure III-5. Multimodal non-Gaussian density function.

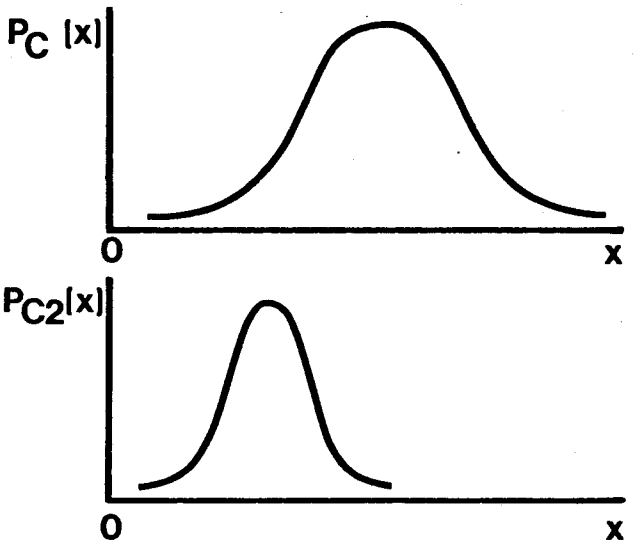


Figure III-6. Multimodal function decomposed into two Gaussian components.

three subclasses per information class. If we request an insufficient number of clusters, the cluster class variances will be very high and the resulting training classes will be difficult to separate. If too many clusters are requested, cluster class variances will be small, but the resulting training classes may be difficult to identify. However, the situation of having too many clusters can be remedied by deleting and/or pooling clusters, whereas when too few clusters are obtained, often the entire clustering process must be repeated. A "good" number of clusters, such as is suggested by the "2X" rule, will usually optimize these trade offs. However, in some cases it will be evident after examining the output that a different number of clusters is needed.

The CLUSTER processor prints the mean values and variances for each cluster class. The variances indicate the spread or dispersion of the data in each channel. If some classes have very high variances relative to other classes, it may be advisable to recluster the data requesting a larger number of clusters.

The CLUSTER processor provides another analysis tool for determining whether the resulting clusters are reasonable: a histogram. Each histogram shows the distribution of the data values of each channel for each cluster class. If the classification algorithm to be used is based upon a Gaussian assumptions, it is important to examine these histograms to make sure the Gaussian assumption is at least approximately satisfied

by the data. This can be done by noting whether or not each histogram has a roughly Gaussian shape.* If a substantial number of histograms significantly violate the Gaussian assumption, we should consider reclustering the data with more cluster classes.

When calculating statistical parameters to describe classes, we must take care to ensure that a sufficient number of training observations are available upon which to base the calculations. This means that we must be certain to have a sufficient number of pixels in each cluster class. For each cluster a minimum of one more pixel than the number of wavelength bands in the MSS data is required. However, to be sure that the pixels are statistically representative of the class being characterized, at least 10 times as many pixels as the number of wavelength bands is usually suggested as the minimum number to characterize each cluster.

ASSOCIATING THE CANDIDATE TRAINING CLASSES WITH INFORMATION CLASSES

A map can be printed showing the cluster class to which each pixel was assigned by the clustering process. Pixels assigned to the same cluster are represented by the same symbol. Thus, areas in the scene that are spectrally homogeneous (in the wavelength bands used) will be displayed with the same symbol. In this way the cluster processing function accomplishes boundary enhancement, allowing us to identify spectrally homogeneous areas from the clustered data more easily than from single-channel data. Cluster maps will be used in this section to help establish associations between the cluster classes (candidate training classes) and information classes.

There is not necessarily a one-to-one correspondence between information classes and cluster classes. Remember, an information class is a distinct cover type of interest as noted above, while a cluster class is a group of pixels which are spectrally similar. As shown in Figure III-7a, there may be a one-to-one correspondence between the two. It is more likely that several cluster classes will represent the same cover type (information class) as shown in Figure III-7b. Sometimes several information classes will be associated with the same cluster class (Figure III-7c). The latter situation indicates that the cover types are spectrally similar. It may be possible to separate the classes into different clusters by requesting a

---

* A subtle point here is that the density functions portrayed by the histograms are marginal density functions. There is no guarantee that, even if each marginal density is reasonably Gaussian, the joint density function is Gaussian. If a marginal density function is non-Gaussian, however, the joint density function is also non-Gaussian.

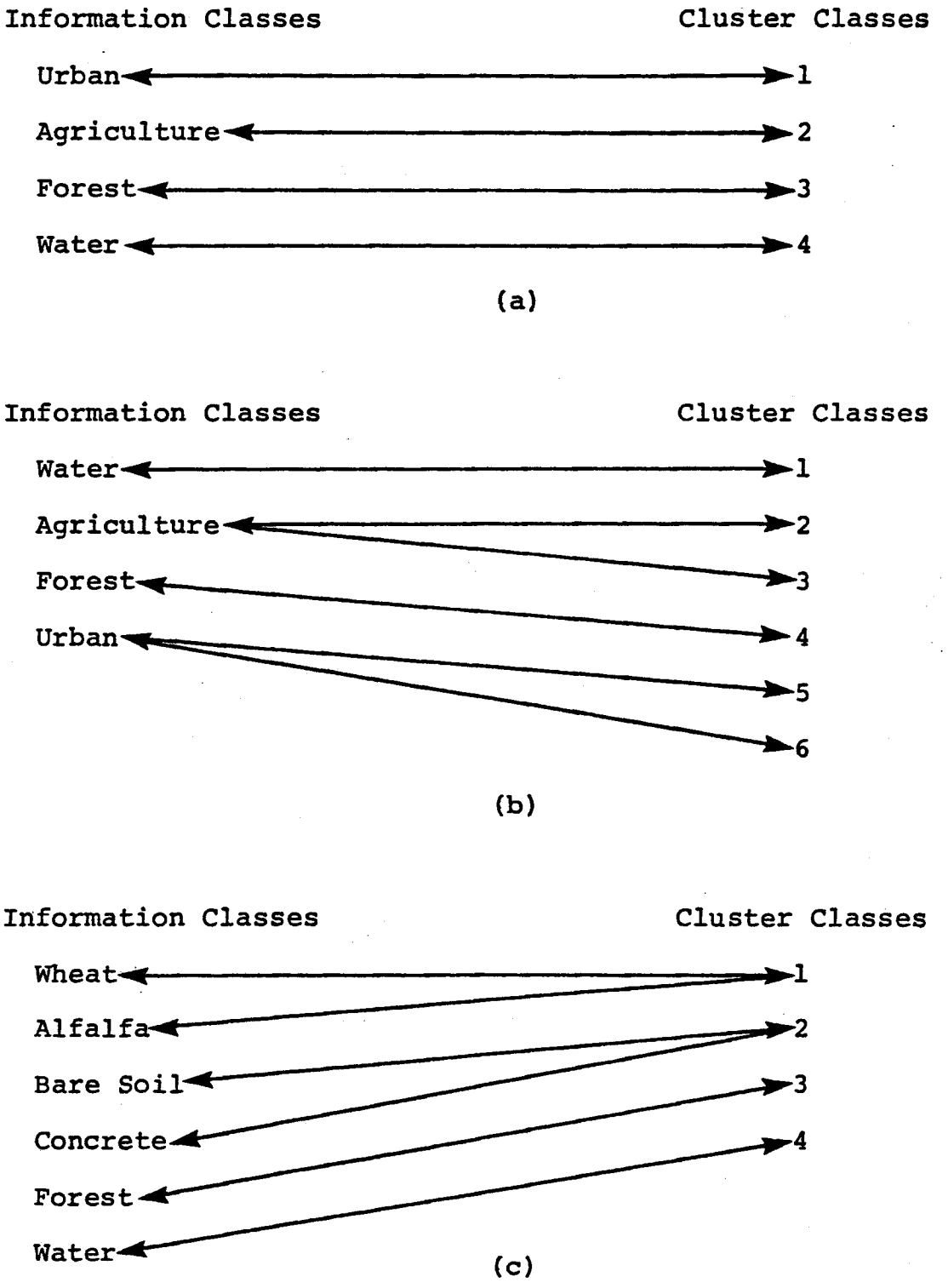


Figure III-7. Examples of relationships between different information classes and their cluster classes.

larger number of clusters.

To identify the clusters obtained, we make maximum use of all available reference data so that the cluster classes can be reliably identified. Note that if incorrect identifications are made in this step, they will be carried through to the classification step, resulting in incorrect maps and acreage estimates. The association of cluster classes and information classes is difficult and time consuming, but this step is most important for ensuring that the classifier is correctly trained.

Reference data often include aerial photography. An overhead projector can be used to superimpose a 9" x 9" transparency on the printed cluster maps. By varying the projector-to-wall distance, it is possible to project the transparency to the scale of the printout, and if the data have been geometrically corrected, a match can be obtained. A better match can be achieved with a 2" x 2" slide and slide projector.

U.S. Geological Survey quadrangle maps of the area are also useful for location purposes. This is especially true when working with lineprinter output of data that have been geometrically corrected and rescaled to the scale of the quadrangle maps. Lineprinter output can be directly overlaid onto the 7 1/2-minute maps and viewed on a light table.

An instrument that makes this alignment task even easier, a zoom transfer scope, uses a lens system to adjust the scale of two images or maps to match each other. We can view a cluster map and aerial photography superimposed on one another and identify each cluster reliably and quickly.

There are several points to remember. One is that if a single cluster appears to correspond to more than one information class, it should be identified that way. Another point is that, in the hybrid approach, each training area is clustered separately. Thus, the symbols in the cluster map for one training area do not necessarily correspond to the same information class in another training area.

AUGMENTING THE CANDIDATE TRAINING CLASSES

Upon completion of the cluster-class/information-class associations, it is important to check that all information classes known to be in the scene are represented. If we are using the nonsupervised approach, our systematic or random sample could have missed an infrequently occurring information class. Or, in the hybrid or supervised approach, we may have missed information classes, such as clouds and cloud shadows, which may occur frequently in our MSS data, but do not appear in our reference data.

Even if we aren't really interested in identifying pixels associated with such "other" classes, we must include training classes to represent them. If we don't, our classifier will be forced to err in classifying pixels belonging to these "other" classes. For example, a cloud shadow might be classified as water if we do not have a training class for cloud shadows. This could be a serious error if the classification results will be used to map bodies of water.

We can define training classes for missed information classes by using the supervised approach. In LARSYS, the statistical description of classes defined by the supervised approach can be obtained by the STATISTICS processor. This processor requires as input the line and column coordinates of the areas for which statistics are to be calculated. The algorthm reads the data values of the pixels within the coordinates specified and calculates the means and the covariances of those pixels without clustering.

In the hybrid approach, we could have selected our training areas to include all "other" classes as well as all information classes of interest. Then we would not have had to use the supervised approach to define training classes for these additional classes. However, when such an "other" class is spectrally very unique, which is usually the case with clouds, it is best to use the supervised approach to characterize the class. When cloud pixels are clustered with pixels of vegetation classes, the CLUSTER processor tends to identify several clusters spectrally between clouds and vegetation, each having very few pixels and not readily associated with any information class.

Another consideration may also lead us to augment our candidate training classes. We may find that certain clusters represent more than one information class. As mentioned earlier, this implicates that those information classes are spectrally similar. To remedy this we could recluster, requesting a larger number of clusters. In some cases, however, we need not recluster all of the training sample to split up the cluster in question. If pixels corresponding to this multi-information-class cluster are readily identified in the MSS data, we can select only pixels in that cluster for reclustering. For example, we may have noticed that deciduous forest and coniferous forest are spectrally similar and were clustered into one cluster class. If we can readily identify pixels of forests in the MSS data, we can cluster only those pixels to obtain separate cluster classes for deciduous and coniferous forest.

VISUAL REPRESENTATION OF CANDIDATE TRAINING CLASSES

We now should have a representative set of candidate training classes for our MSS data set. Several of these candidate training classes may be spectrally similar. If the spectrally similar classes represent the same information class, we usually will want to combine them together into one training class. If they represent different information classes we will usually want to judiciously delete some of the candidate training classes in order to avoid confusing our classifier. Generally, we want to reduce the number of training classes to a minimum, because this will save computer time and simplify interpretation of results. In doing so, therein, we must be careful that our final training classes are still representative of our MSS data set.

To begin reducing the number of candidate training classes, it is useful to visualize the spectral characteristics of all candidate training classes at one time. The MERGESTATISTICS processing function has as one of its output products a two-dimensional plot known as a coincident bi-spectral plot. Plotted on one axis are the class mean values in one channel for each candidate training class. On the other axis are plotted the class mean values for another channel. Alternatively, in order to use information from more than two channels in the bi-spectral plot, the class means for a set of channels can be averaged and plotted on one axis, and the class means for another set of channels can be averaged and plotted on the other axis. For example, with Landsat data the average class mean values in the infrared channels are usually plotted against one axis. The average class mean values in the visible channels are usually plotted against the other axis. The rationale for averaging the means in this way is based on the observation that responses in the two visible bands are highly correlated, as are responses in the two IR bands. (This may be observed by comparing the two visible gray scale images and the two IR gray scale Landsat images). Whatever combination of channels is used in the bi-spectral plot, information from a multi-dimensional measurement is displayed in two dimensions in this way. The final output is a plot providing a visual comparison of the means of all candidate training classes.

Another way to visually represent the candidate training classes is a plot of calibrated spectral mean curves. This plot can be used with equal effectiveness for both lower dimensional data such as Landsat MSS data (four bands) and higher dimensional data such as Skylab MSS data (thirteen bands). In contrast, it becomes difficult to effectively use a bi-spectral for data with higher dimensionality than Landsat data. Before a calibrated spectral mean curve can be generated, the class mean values in each wavelength band must be calibrated so that a particular radiance value in each band produces the same class mean value. Once this calibration is completed, the class means can be plotted against the wavelength bands as shown in Figure III-8.

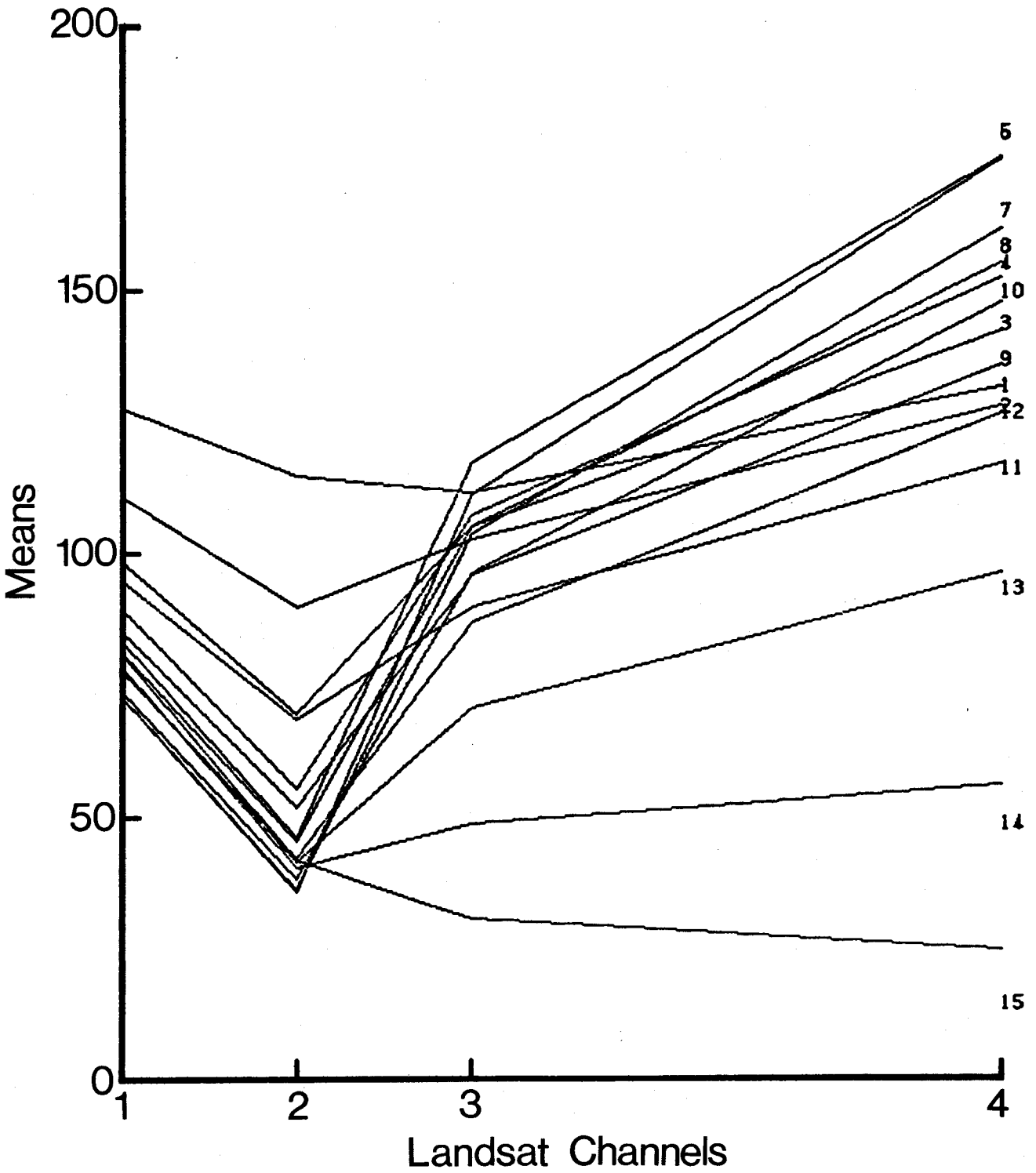


Figure III-8. Calibrated spectral mean curves for a Landsat training area located in Monroe County, Indiana.

CALCULATING STATISTICAL DISTANCES BETWEEN THE CANDIDATE TRAINING CLASSES

Although at this point it is possible to visualize and compare all candidate training classes, the comparison is on the basis of the means of the classes only. Before any decisions can be made about which classes to merge or delete, we will find it useful to consider the amount of dispersion or variability each class has.

The second major reason for considering the dispersion of each class is to give us some indication of the probability of correct classification in advance of doing the classification. If there appears to be considerable confusion among information classes, we may do more clustering on the areas already used, asking for a different number of clusters. Alternatively, additional training areas could be selected (in the hybrid approach) in an effort to get improved distinction among classes, instead of trying to combine the cluster classes we already have.

Calculating the "separability" of the cluster classes can help determine which cluster classes are similar and can serve as an indicator of probability of correct classification.

To explain how this is accomplished, we must first discuss the concept of "statistical distance." Figure III-9 shows two examples of one-dimensional density functions. Intuitively we know that the "distance" between the density functions is greater in case B than in case A. The distance between two Gaussian probability density functions depends not only on the distance between the mean values, but also on the "spread" of the data. Figure III-10 illustrates this point. The distances between the mean values are equal in both of the cases shown, but the smaller variances (smaller "spread") in case B of Figure III-10 result in a larger statistical distance between the two density functions.

In two dimensions the density functions may be represented by ellipses (Figure III-11). In three or more dimensions, as with four-channel Landsat data, the density functions are represented by ellipsoids (blimp-like surfaces of equal probability in the measurement space). As in the one-dimensional case, the statistical distance in two or more dimensions is an estimate of the overlap of the density functions. As shown in Figure III-11, the overlap is greater due to the slight shift in the orientation of one ellipse. Therefore, we would anticipate the classification accuracy in case B to also be less.

Transformed divergence (D_T) is a multivariate measure of statistical distance. There are other such measures of statistical distance, but transformed divergence performs performs well in estimating the probability of correct classification between pairs of classes. Experimental results of plotting probability of correct classification versus transformed

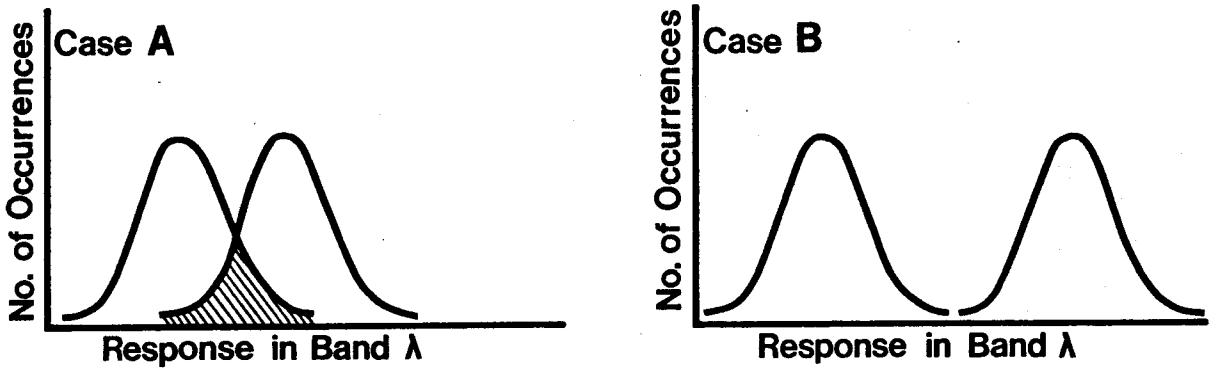


Figure III-9. Two pairs of one-dimensional density functions. The statistical distance between the density functions in case A is less than in case B.

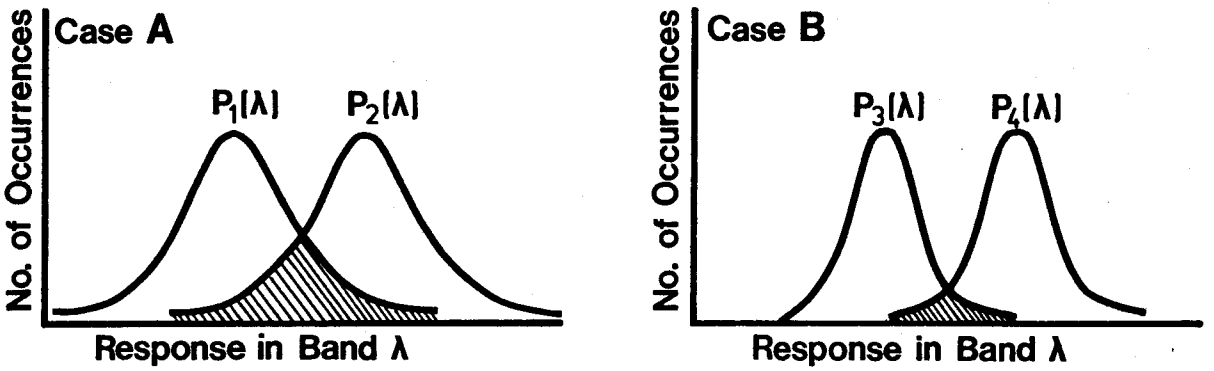


Figure III-10. Both pairs of distributions shown above have equidistant means, but the smaller variances of $P_3(\lambda)$ and $P_4(\lambda)$ cause this pair to have larger statistical distance.

divergence for training data are shown in the graph in Figure III-12. Notice that class pairs with larger transformed divergence values (D_T) also achieved a higher classification accuracy (P_C) although the relationship is not perfectly linear. This graph can help in determining what the minimum acceptable transformed divergence value between pairs of classes should be. According to Figure III-12, to achieve 85% accuracy, we see that the final training classes should have transformed divergence values between about 800 and 1800. Classes with transformed divergence of 800 would achieve 85% accuracy only infrequently. On the other hand at 1800, 85% accuracy could almost always be obtained. Although one might be tempted to require a very high threshold, as higher and higher transformed divergence values are required, only the more general (i.e. "agriculture" rather than "corn," "soybeans," "wheat," etc.) classes will remain distinct. Therefore a balance must always be struck between the level of detail desired and the minimum allowable transformed

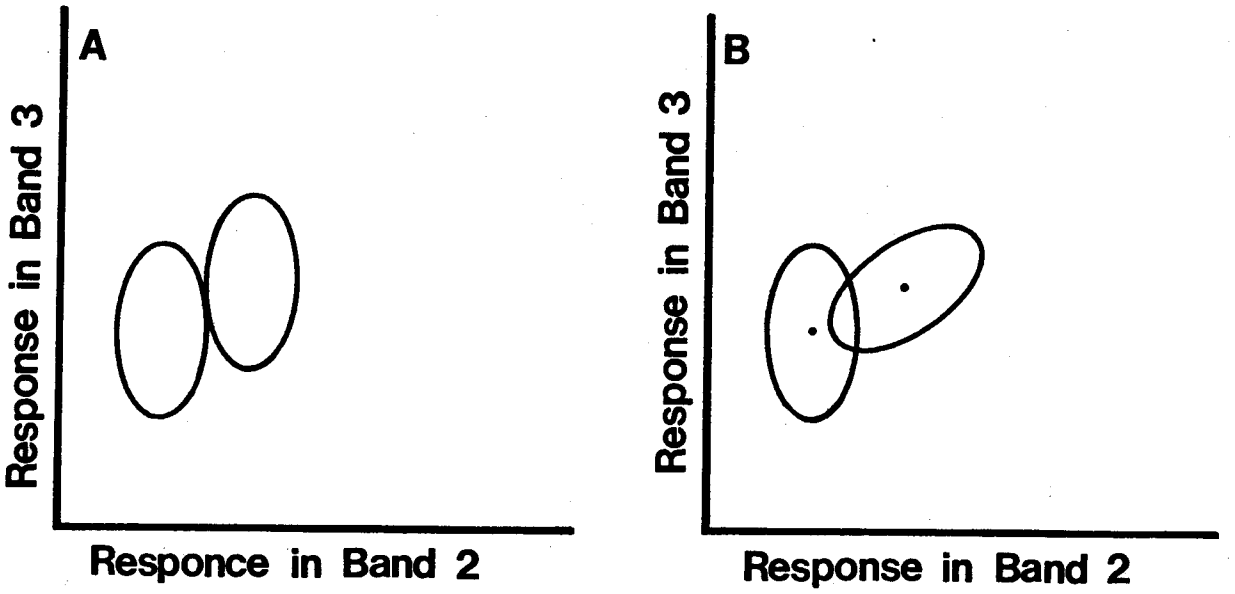


Figure III-11. Ellipses representative of training classes.

divergence value, one at which acceptable accuracy will be obtained most of the time for classes of the desired informational content. More information about statistical distance measures can be found in sections 3.7 and 3.8 of Swain and Davis.

In LARSYS, the SEPARABILITY processor is used to calculate the transformed divergence between pairs of training classes. An analyst typically evaluates information from both the SEPARABILITY processor and the bi-spectral plot by adding the separability information to the bi-spectral plot. This is done as illustrated in Figure III-13. A solid line is drawn between two classes on the bi-spectral plot if the transformed divergence between the pair of classes is less than 1000. A dashed line is drawn between the symbols if the transformed divergence is between 1000 and 1500. No line is drawn if the transformed divergence is greater than 1500.

Separability information can also be used with calibrated spectral mean curves in a similar fashion. In addition, the shape of the spectral mean curves give further information about the separability of the classes. Occasionally, the transformed divergence between two classes may be quite low even though the classes are actually well separated spectrally. When this happens, the two classes will have calibrated spectral mean curves with distinctly different shapes. For an example, see Figure III-14.

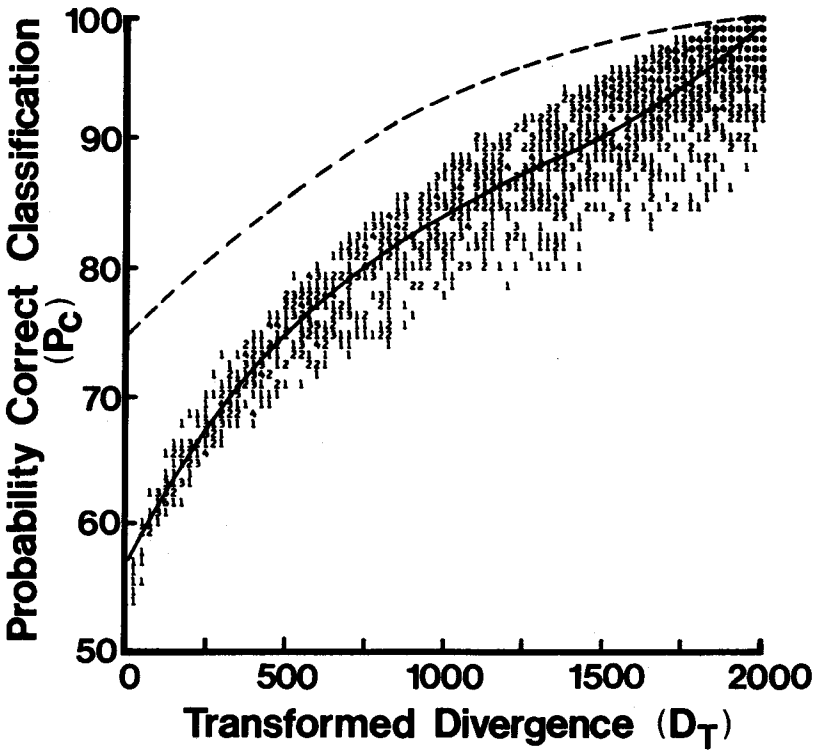


Figure III-12. Observed values of probability of correct classification versus transformed divergence.

REFINING THE SPECTRAL TRAINING CLASSES

The final step in the definition of the spectral training classes is to refine the candidate training classes into a set of final training classes. So far, the definition of spectral training classes has included clustering the training sample (or training areas), associating cluster classes with information classes (cover types), augmenting the cluster classes as necessary, calculating statistics of the cluster (candidate training) classes, calculating transformed divergence to get a measure of the distance between pairs of clusters, and summarizing these results on a bi-spectral plot. Next, we will make decisions as to which candidate training classes should be grouped to form spectral training classes, which candidate training classes can be used directly as spectral training classes and which should be deleted. In the example that follows, we will use a bi-spectral plot with cluster-class/information-class associations and transformed divergence values to help make these decisions. Later we discuss how the calibrated spectral mean curves, and observations of the normality of class histograms and the number of points in each candidate training class can be used to help resolve difficulties.

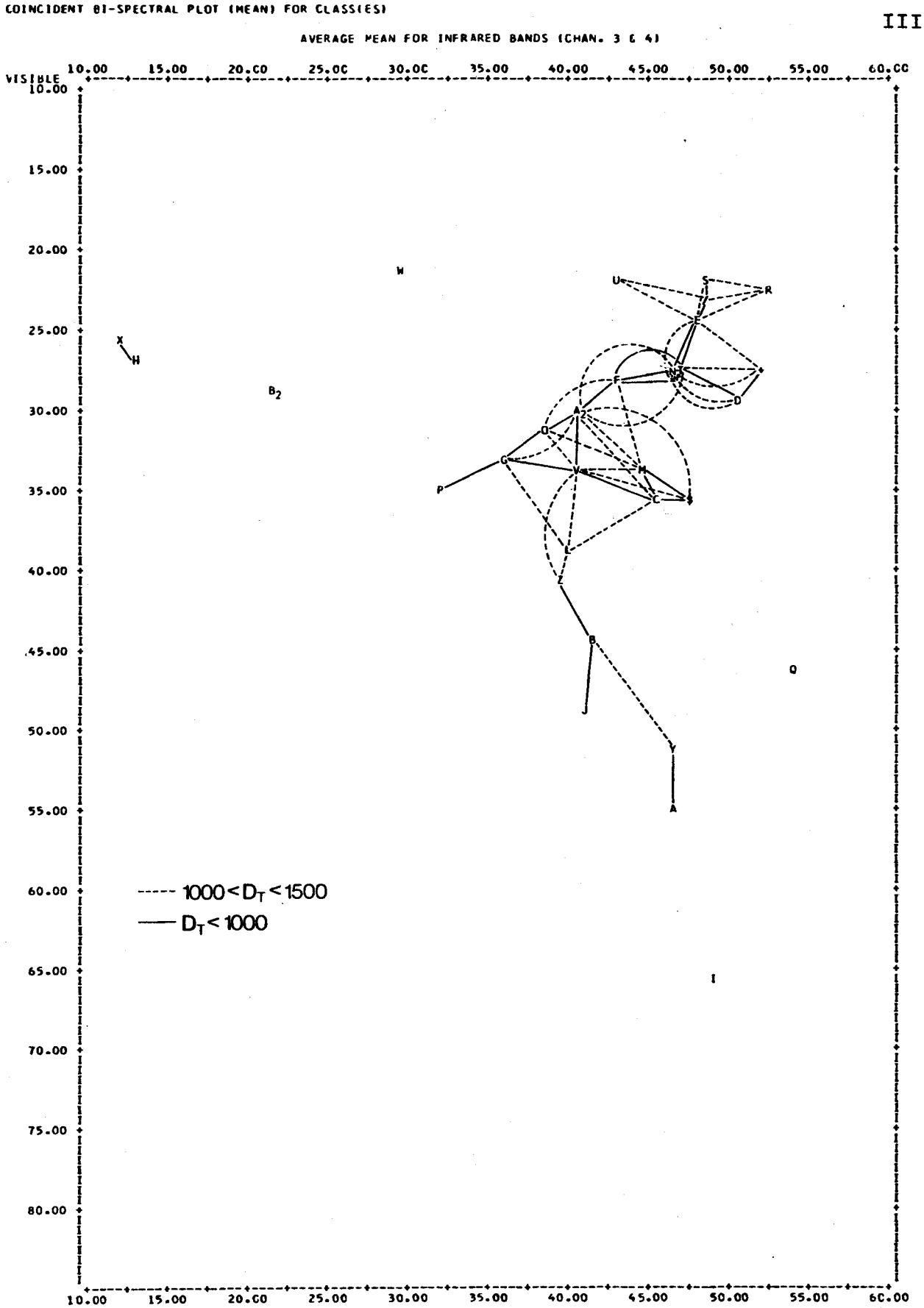


Figure III-13. Separability information added to a bi-spectral plot.

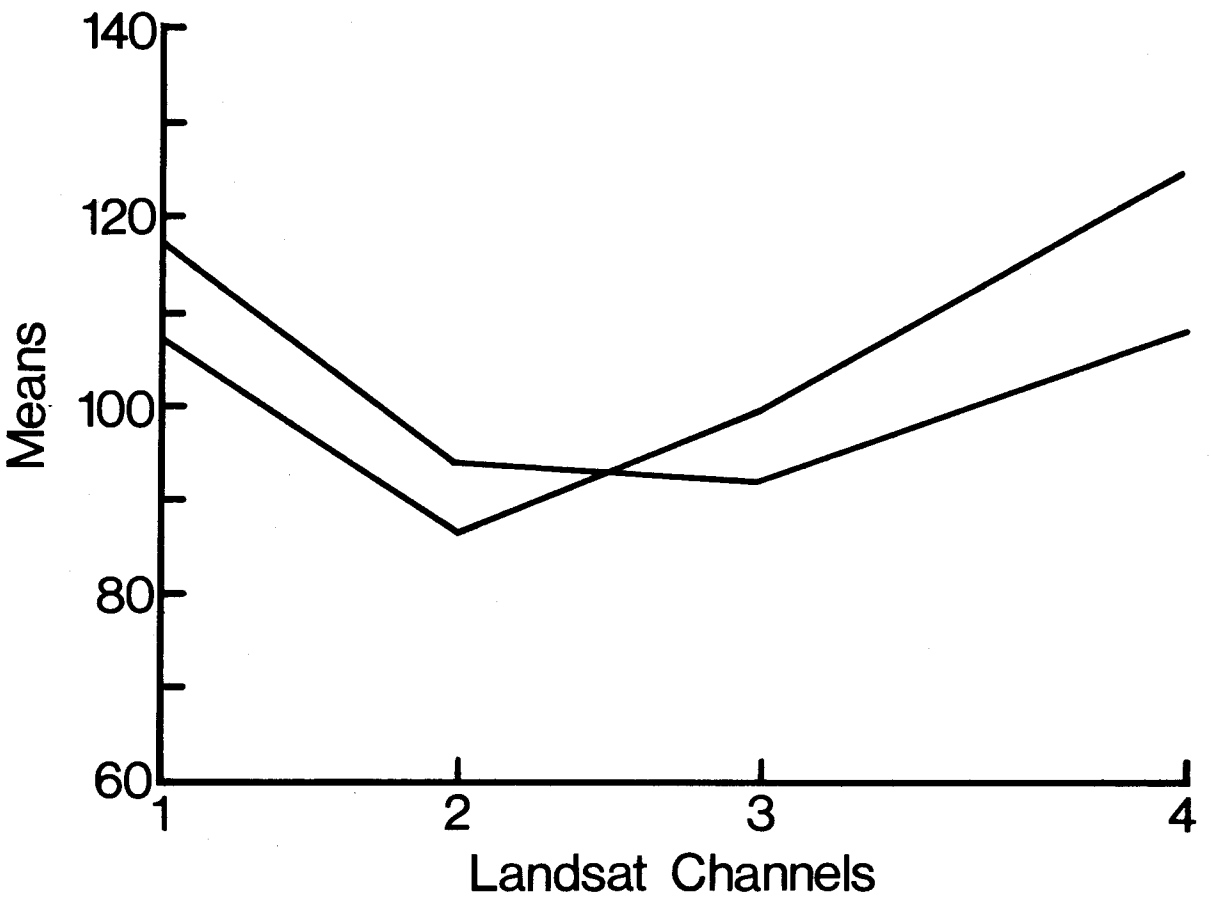


Figure III-14. Example of calibrated spectral mean curves of two well separated classes that have very low transformed divergence between them.

There are a number of ways to approach the task of selecting the final spectral training classes from the available candidate training classes. The overall goal is to achieve training classes that in total are representative of the information classes present and which are separable from one another. We will present two divergent philosophies for selecting the final training classes and will offer guidelines for handling difficulties.

The first philosophy or approach is the pooling approach. In this approach candidate training classes whose interclass statistical distances are less than a chosen minimum acceptable value are pooled. The pooling approach tries on the basis of transformed divergence alone to form all or nearly all of the candidate training classes into groups of classes connected to one another. The identities of the classes are used to indicate the portion of the plot (and thus the portion of multidimensional space) that belongs to each information class.

Although the identity of each candidate training class is considered when deciding which classes to pool, it is not required that the identities of the classes pooled be exactly the same. For example, in Figure III-15 those class pairs whose transformed divergence is less than 1000 have been connected by a solid line, and those between 1000 and 1500 by a dashed line. Two pools that might be chosen from that diagram are made up of classes assigned to the symbols R, S, /, E and U (representing forest), and +, D, T, =, N, F and A2 (representing agriculture). Notice that although symbol N is labeled forest, it might be pooled with agriculture since it is similar to and surrounded by classes labeled agriculture. It may be that N was erroneously labeled as forest; it is advisable to recheck the identity of such seemingly anomalous classes. In many cases the conclusion may be that such an "anomalous" class is really informationally more like the classes it was pooled with than its identity would suggest. However, if N is clearly a forest class, it would generally not be a good idea to pool it together with agriculture classes.

The second major philosophy or approach to selecting training classes is the deleting approach. In the deleting approach, groups of candidate training classes or individual candidate training classes are chosen which minimize the variance of each final training class. To accomplish this, classes on the borders between information classes are deleted. Returning to Figure III-15, the deleting approach might leave classes R, S, / and U to represent forest, and classes +, D, T, = and F to represent agriculture. Classes E , N and A2 would be deleted. As a result of using the deletion approach, the training classes for agriculture and forest are more separate and distinct from each other and other classes. Carried to extreme, the deleting approach leads to a larger proportion of more specific (but not necessarily more representative) training classes. As another example, classes R and S could be selected as a variety of forest and class U as another variety of forest. Classes / and E would be deleted.

The following logic can be drawn upon in deciding whether or not to delete a class: The spectral characteristics of any ground cover type are best described by a cloud of points rather than a single point. When that cloud gets broken into subclasses by clustering, one or more of the subclasses will be composed of points near the edge of the cloud. These edge points, which are not really very similar to the points in the center of the cloud that best represent the cover type, may poorly represent the information class in question. Further, it is likely that such edge subclasses will be confused with some subclass(es) of a different identity. When this is the case, deletion of such edge subclasses from the training classes is reasonable and valid.

It could be said that the pooling approach is more concerned with obtaining representative training classes while the deleting

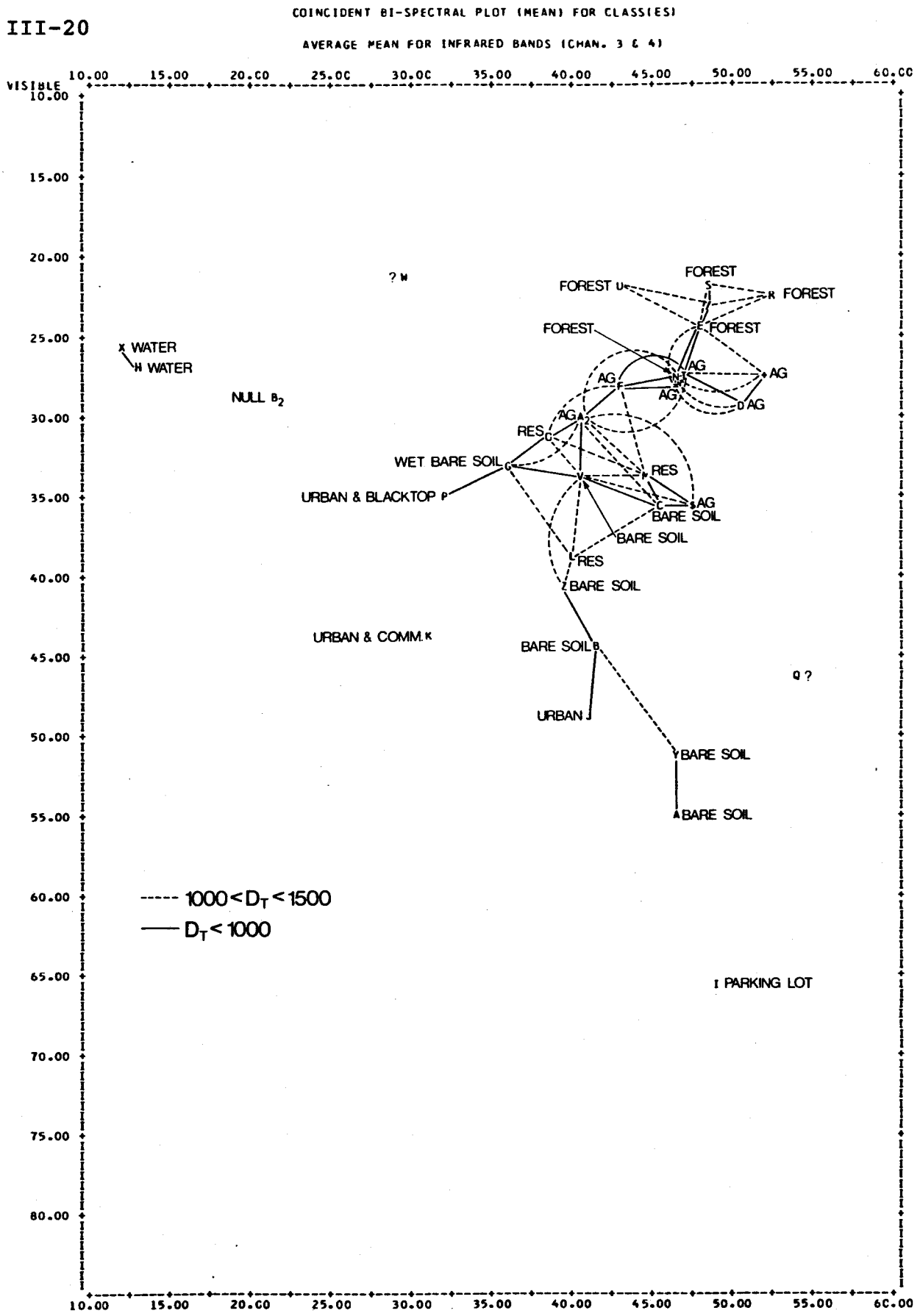


Figure III-15. Separability diagram with identities of classes added.

approach is more concerned with obtaining separable training classes. In practice it is necessary to consider both representation and separability. Often analysts will use pooling and deleting in combination to get the best trade off between representation and separability.

Up to this point we have considered relatively straightforward nonproblemmatic examples. Unfortunately, almost every analysis has some problems, and rarely is the decision process simple. In the previous example of selecting training classes for forest and agriculture, we needed to decide whether or not class E should be retained or deleted. Under the pooling approach we would retain class E and group it with classes R, S, / and U on the basis that deleting it would degrade the representativeness of forest. With the deleting approach, though, we would delete the class to enhance the separability of the proposed groups. On the basis of the bi-spectral plot and separability information it is difficult to decide which approach to use. Fortunately, we can often draw on additional information to help us decide whether to delete or pool a class. For example, how many pixels are in class E? How firmly is its identity established? How truly Gaussian is its distribution of data values as shown by its histograms? How large are the variances in each channel? The answers to such questions give a measure of the quality of the class. Whereas we could feel justified deleting a class with a small number of pixels, large variances and whose identity was not completely clear, we should feel uneasy about deleting a class with a large number of pixels, small variances and unmistakeable identity.

Calibrated spectral mean curves can also be used to help decide whether to pool or delete a class. For example, if the plot for the forest class E has a shape distinctly different from the nearby agricultural classes and is very similar in shape to the nearby forest classes, we would be more inclined to pool class E with the other forest classes than to delete it. If class E has a shape inbetween that of the nearby agricultural and forest classes, we would be inclined to delete the class.

Even more challenging situations may arise depending on the information classes we are interested in and the data set being analyzed. It may happen that informationally different classes are extensively confused with one another. In Figure III-15, classes $, M, C, V, O, L and G are all interconnected although O, M and L are identified as residential and $, C, V and G as bare soil and agriculture. The first step in such situations is to double check the identities of the classes. Presuming that they are correct, we could expend some effort toward refining the classes to enhance their separability. Such refinement would be especially appropriate if several of the classes have very non-Gaussian distributions, large variances or were difficult to identify. The refinement could involve reclustering the same training areas, requesting different numbers of clusters, or

selecting additional training areas and clustering those. The latter would be especially appropriate if the representativeness of the training classes is in question. When problems are encountered some refinement should be attempted. However, we must recognize that if the digital brightness values of pixels from different cover types are very similar, no amount of refinement will make them less similar, and therefore, we cannot expect the refinement process always to lead to greater separability of cover types.

When the refinement is not successful, we should consider what kind of classification errors might or might not be acceptable. For example, assume that we are interested in classifying an urban area and discover that a candidate training class identified as urban is similar to a candidate training class identified as agriculture. We could decide that, for our purposes, the error of classifying some agriculture data points as urban would not be too troublesome, while the error of classifying some urban points into agriculture would be disastrous. In that case, we could choose to eliminate the class identified as agriculture from any subsequent processing.

It is appropriate to say that the process of selecting training classes is the process of deciding how portions of the multidimensional space will be classified. In the previous example, we chose to label the portion of the multidimensional space in question as urban. Another criterion that we could use is to examine which information class is preponderant in that portion of multidimensional space. If the urban class being confused with agriculture in the previous example had been surrounded by other agriculture classes, a better decision might have been to reject the urban class and keep the agriculture class.

At times we might be tempted to retain a class that has less than the practical minimum number of pixels because we feel the class to be informationally important. However, we must think carefully about retaining such a class, for it is unlikely that the pixels in that class accurately represent the information class. In addition, there is a good chance that the covariance matrix for that class will be singular. If the covariance is singular, the class cannot be used by many classifiers, and deleting the class is unavoidable. A preferable alternative is to pick some additional training area(s) likely to contain more points which also represent this information class. If this is not possible, the results of the classification including this class should be closely scrutinized to be sure that the "ill-conditioned" class statistics have not resulted in peculiarities in the results. If we do note such peculiarities, we will have to delete the class.

We previously had available the statistical descriptions of candidate training classes which appear on the bi-spectral plot.

Before we can perform a classification based on the new spectral training classes obtained by pooling or deleting the candidate training classes, we must obtain statistical descriptions of these new classes. Since two of the classification algorithms we will consider are based on the assumption that the training classes can be represented by multivariate Gaussian probability density functions (defined by mean vectors and covariance matrices for groups of training samples), the same kind of statistical description must obtained by recalculating the mean vector and covariance matrix for each of these new groups of pixels. When the group of pixels making up the new spectral training class is a result of the pooling approach, we often call this process "pooling the statistics" of the constituent candidate training classes.

After obtaining the mean vectors and covariance matrices of the new spectral training classes, we should check their separability in order to get an indication of the probability of correct classification resulting from using these training classes. In LARSYS this is done by means of the SEPARABILITY processing function.

We should note that satisfying a certain transformed divergence threshold only ensures the degree to which the final training classes are spectrally different and that there is a certain probability of distinguishing among them. This does not ensure that they are representative -- as an analyst, we must do that. Representativeness is a function of the training sample selected, the accuracy of identification of cluster classes, and the number of cluster classes (and the ground cover associations of the cluster classes) which survive the selection step. Even when we feel confident in the representativeness, separability, and accuracy of identification of the classes, misclassification can occur. We can only attempt to optimize all three, make the classification and, based upon the results, determine the acceptability of the classification.

In cases of unacceptable results, previous steps may be repeated from the acquisition of scanner data through the selection of training classes. As noted earlier, although the analysis process appears to be a straight-line sequence of steps up to this point, in reality it is a very iterative sequence where the results of each step are closely scrutinized and the steps are often repeated before proceeding. The training class selection step is especially iterative in that the acceptability of the training class selections is tested by examining the separability of the newly formed training classes before proceeding to make the classification. Typically, we will have to make several attempts at selecting the training classes before we can arrive at an acceptable set of training classes.

One point which should be apparent by now is that there is no single correct way to progress through an analysis sequence. As

your understanding of the pattern recognition concepts increases and you gain experience in analysis, you may even develop new procedures yourself.

Self-Check

1. Describe two tasks the cluster processing function can accomplish.

2. State the rule-of-thumb used to determine the number of cluster classes to request and give the reasoning behind the rule.

3. State why cluster classes must be associated with information classes.

4. Name the two statistical parameters which define multivariate Gaussian distributions.

5. Explain why statistical descriptions of candidate training classes are needed at this point in the analysis.

6. Name one difference between the supervised and nonsupervised approach to the generation of training classes.

7. Describe one method of "seeing" the spectral characteristics of candidate training classes.

8. Identify the regions of a bi-spectral plot belonging to major cover types.

9. State two reasons why spectrally similar cluster classes are combined.

10. Look at the two pairs of one-dimensional density functions shown in Figure III-16. In which case is the statistical distance between density functions larger?

11. Name the two characteristics of Gaussian probability density functions which determine the statistical distance between the density functions.

12. What is the value of knowing the statistical distance between all possible pairs of classes?

13. How can the minimal allowable value of transformed divergence be determined?

14. What are two desirable characteristics for final training classes?

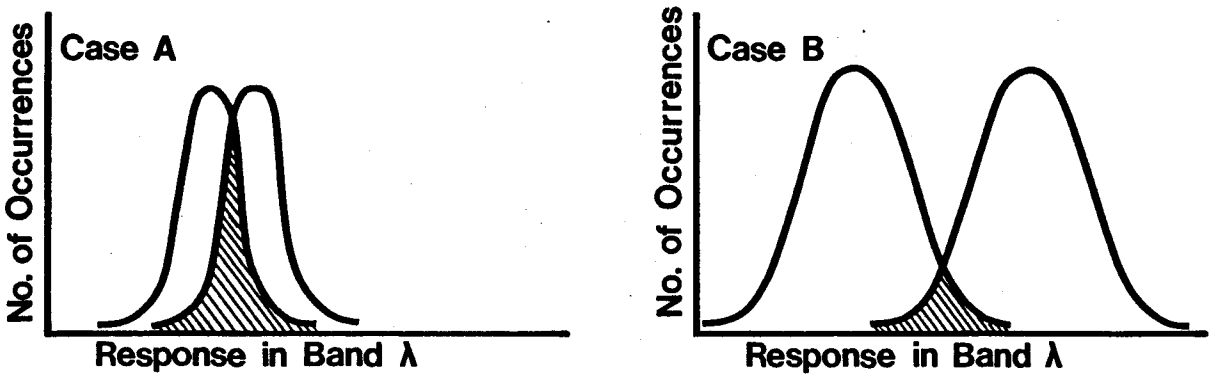


Figure III-16. Two pairs of one-dimensional density functions.

15. Is the pooling or deleting method of training class selection always preferred?

16. If transformed divergence values of less than 1500 (or any threshold) between the newly formed training classes exist, must previous steps be repeated, or can the analysis continue?

CHAPTER IV. CLASSIFICATION OF THE ENTIRE STUDY AREA

Upon completion of this chapter, you should be able to:

1. Name and briefly describe the decision rule implemented in the CLASSIFYPOINTS processing function.
2. Briefly describe the ECHO classifier algorithm.
3. Briefly describe the minimum distance classification algorithm.

Once a certain confidence in the training classes has been established, a step of special importance in the analysis sequence can be performed - classification. We may have only one classification algorithm available to us or we may be in a situation where we may select one from among several available. For purposes of illustration and comparison, we will discuss three different classification algorithms - maximum likelihood, ECHO, and minimum distance.

A classification algorthm must be defined in a quantitative way so that the computer can do the work. This can be accomplished by defining a set of functions (mathematical expressions) corresponding to the training classes. These functions, which are called "discriminant functions," are contructed so that when the data values belonging to a pixel being classified are substituted into them, the function having the largest (or, alternatively, the smallest) value corresponds to the class into which the pixel is to be classified.

The classification algorithm utilized by the LARSYS CLASSIFYPOINTS processor is based upon the maximum likelihood classification rule. Each pixel to be classified is "compared" to each training class and assigned to the class it most likely belongs to. The discriminant functions for the maximum likelihood classification rule are derived using statistical decision theory so as to minimize the probability of making an erroneous classification (see Swain and Davis, section 3.6 for details). When the classes are assumed to be characterized by multivariate normal density functions, the discriminant functions are defined in terms of the mean vectors and covariance matrices of the classes. The set of discriminant functions for the maximum likelihood classifier are defined so that when the data values belonging to a pixel are substituted into all of them, the function having the largest value determines the class that the pixel "most likely" belongs to.

The CLASSIFYPOINTS processor uses only spectral information in making classification decisions. The classification algorithm utilized by the ECHO processor uses spatial information as well as spectral information. ECHO is an acronym for Extraction and Classification of Homogenous Objects. When processing a data set, ECHO first divides the scene into rectangular cells. The cell size is chosen by the analyst usually on the basis of the average object size expected in the scene. Figure IV-1 illustrates a portion of a data set divided into 3 X 3 cells. The ECHO classifier examines the pixels within a cell and performs a statistical test to determine whether or not the

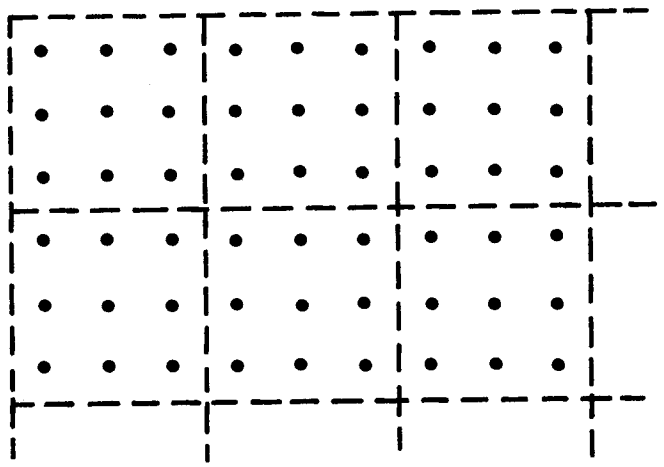


Figure IV-1. Data set divided into 3 x cells by ECHO. Each dot represents a pixel in the scene.

pixels within the cell are statistically similar. If they are, they are judged to belong to the same "object" in the scene. ECHO then examines a neighboring cell. If the pixels making up this cell are statistically similar, a test is performed on the two cells. If they both have similar statistical properties, the cells are combined to form an object. This process continues, cells are annexed to the object, until a cell is encountered which is not statistically similar to the cells comprising the object. This cell is then declared to belong to a different object. Once the objects have been identified, all of the pixels comprising the object are classified as a group by comparing the estimated probability distribution of the pixels in the object to the probability distributions of each of the training classes. In that way all of the pixels in each object are classified at one time as belonging to the most similar training class. If a non-homogeneous cell is encountered (the cell pixels fail the

statistical similarity test), each data vector in the cell is classified individually using the maximum likelihood (CLASSIFYPOINTS) decision rule. ECHO requires the same kind of training class statistics as does CLASSIFYPOINTS: the mean vector and covariance matrix of each training class.

The minimum distance classifier is a "point" classifier like CLASSIFYPOINTS in the sense that each pixel in the scene is classified individually. The classification algorithm is simpler in both concept and implementation than either of the previous two classifiers discussed. For each pixel in the scene, a multidimensional distance between the data values belonging to the pixel and the mean vector of each training class is computed. These distances are used as the discriminant functions for this classifier. The pixel is assigned to the nearest training class, i.e., the class with the smallest discriminant function. While computationally more efficient, the minimum distance classifier is inherently less powerful than the maximum likelihood classifier. However, the minimum distance classifier tends to yield approximately the same results as the maximum likelihood classifier when the training classes are generated in a way that leads to a large number of training classes representing the full range of spectral characteristics in the scene. The clustering algorithm described earlier yields classes with these characteristics. The maximum likelihood tends to perform better when the classes of interest have very similar or overlapping spectral characteristics.

Self-Check

1. Name and briefly describe the kind of decision rule implemented in the CLASSIFYPOINTS processing function.

2. With the aid of Figure IV-1, briefly describe how the ECHO classifier works.

3. Contrast the minimum distance classification algorithm with the maximum likelihood algorithm.

CHAPTER V. PICTORIAL AND/OR TABULAR DISPLAY OF THE CLASSIFICATION RESULTS

Upon completion of this chapter, you should be able to:

1. Name the two major formats for displaying classification results and give one reason for using each display format.

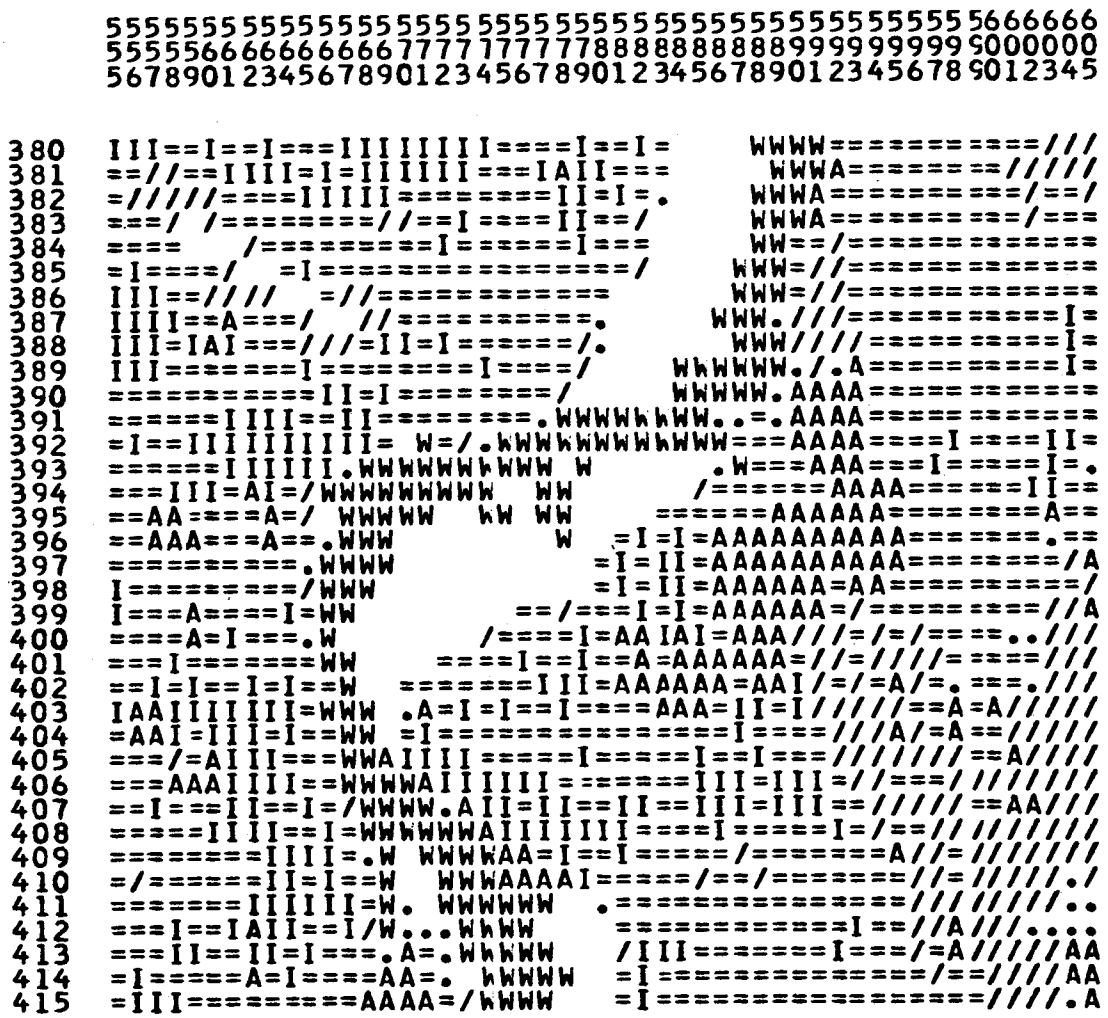
After the completion of the multispectral classification, the results can be displayed in several different formats according to the user needs and specifications. There are two major types of display formats: pictorial and tabular. For example, the classified area could be displayed as a map of a certain scale, projection, and minimum mapping unit. The different classes (ground cover types) can be represented by (1) alphanumeric symbols (Figure V-1), (2) graphic symbols (Figure V-2), (3) gray levels (Figure V-3), (4) boundary lines (Figure V-4), or (5) different colors. The classification results also could be displayed in a thematic map format in which only one class is represented.

The other major type of classification format, the tabular format, can be utilized when a user requires only information such as areal extent (acreage) or percentage of each one of the different cover types present in the study site. The output in this case is a table indicating the number of pixels classified into each cover type. The area represented by each data point multiplied by the number of data points per cover type will give the area per cover type. For geometrically corrected Landsat data scaled to 1:24,000 to match the U. S. G. S. 7 1/2 minute topographic quadrangle maps, the conversion factor is 1.145 acres per pixel. Thus if we find that 19,465 pixels were classified as deciduous forest in a classification of Landsat data, we can calculate that the study area contains 22,287 acres of deciduous forest.

Self-Check

1. What is the advantage of a pictorial display of classification results?

2. When should we use a tabular format for displaying classification results?



W - Water

/ - Grass

• - Soils

A - Chaparral

= - Forest-1

I - Forest-2

  - Sand

Figure V-1. Alphanumeric representation of a final multispectral classification.

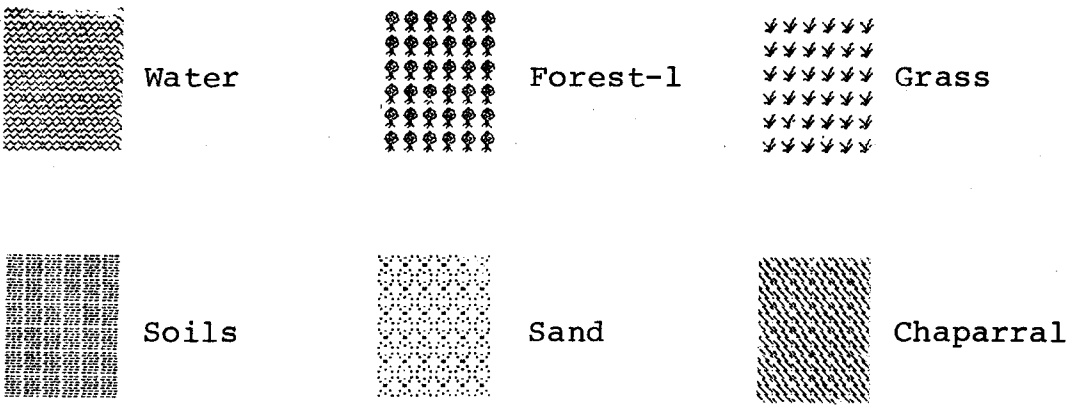


Figure V-2. Graphic symbols representation of a final multispectral classification.

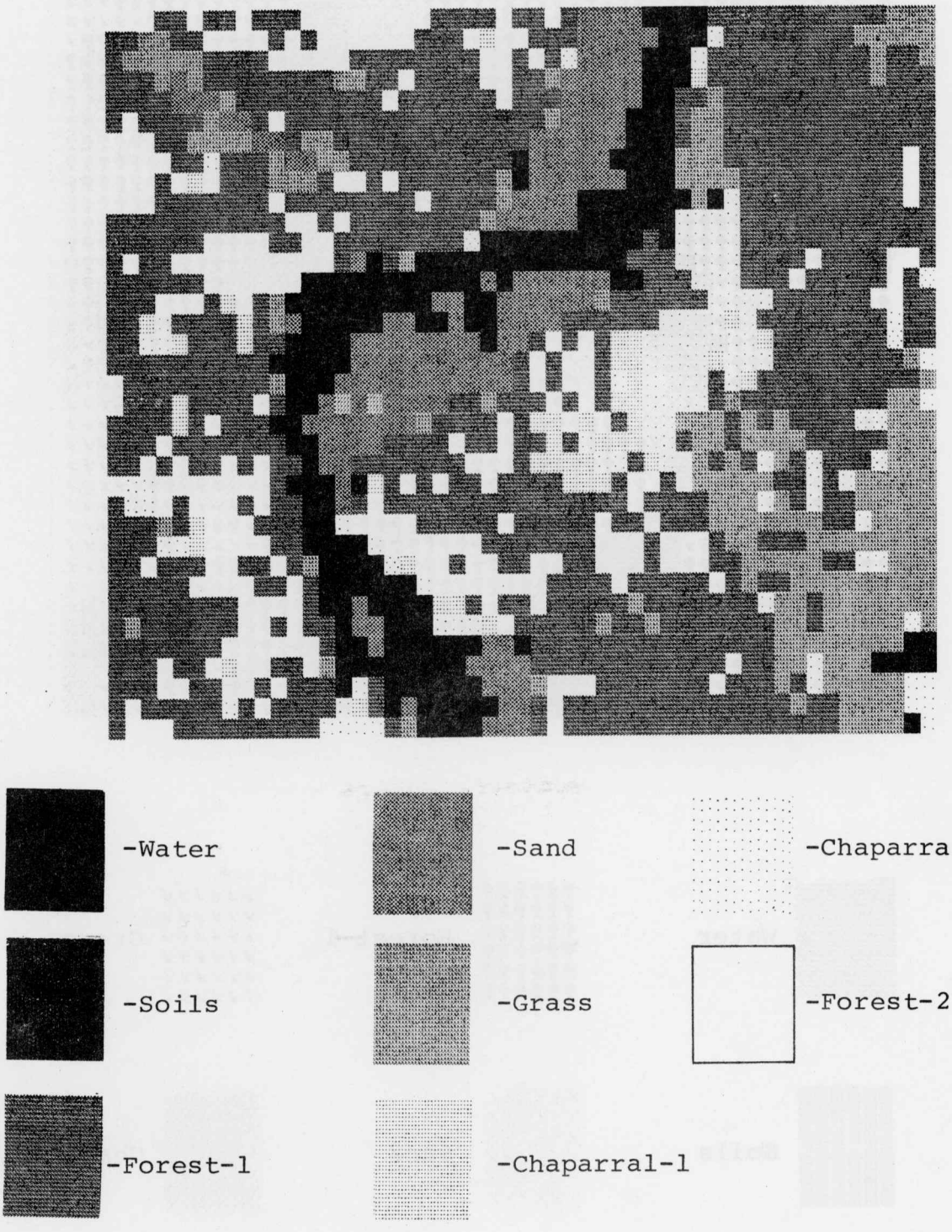
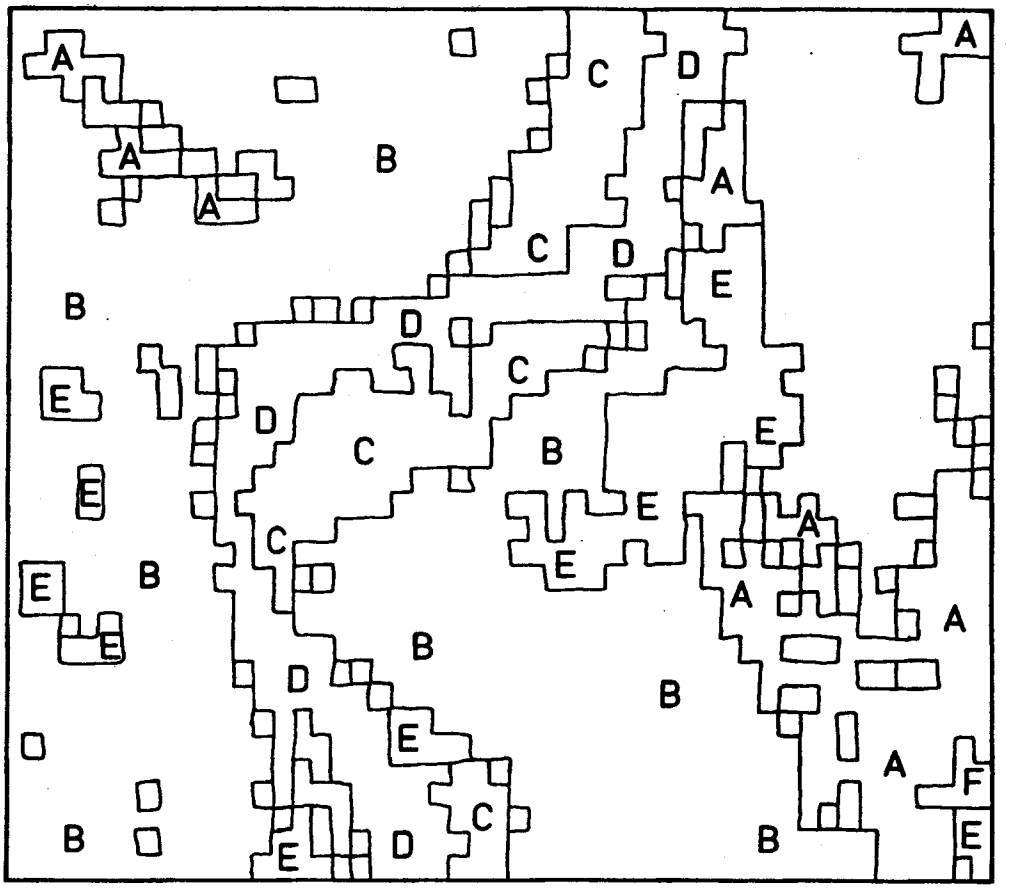


Figure V-3. Gray levels representation of a final multispectral classification.



A - Grass
B - Forest-1
C - Sand
D - Water
E - Chaparral
F - Soils

Figure V-4. Boundary line representation of a final multispectral classification.

CHAPTER VI. EVALUATION OF THE CLASSIFICATION RESULTS

Upon completion of this activity, you should be able to:

1. Given an example of a class performance matrix, indicate pixels correctly classified, errors of omission and errors of commission for a specified class.

For a multispectral classification to be of practical use, we must determine its accuracy and reliability. Using the numerical analysis approach, we can quantitatively assess the degree of accuracy of a multispectral classification. Experience has shown that the "test field performance" method is most effective in assessing classsification accuracy. Test fields of known cover types are selected and the computer determines the percentage of correctly classified pixels. The size of each test field is such that it contains only one cover type. Since the identity of the pixels in each test field is known along with their addresses in the data, a computer can locate these pixels in the classification result (stored on tape), observe the class into which each pixel was classified, and compare that result with the ground truth identity of each pixel to determine whether the pixel was correctly classified. Typically, several test fields are selected for evaluating the classification accuracy of each information class. The computer examines and tabulates the classification decision for each pixel in each test field and prints out a summary by field, by class (all test fields chosen for an information class), or both, as specified by the analyst. An example of tabular results for testing classifier performance is shown in Table VI-1. Such a table can be called a "test class performance matrix."

	No. of Samples	Percent Correct	Conif-erous	Decid-uous	Grass land	Barren	Water
Coniferous	9,634	94.6	9,110	22	53	21	428
Deciduous	1,475	87.5	113	1,286	76	0	0
Grassland	3,677	81.2	49	129	2,988	510	1
Barren	35	97.1	0	0	1	34	0
Water	1,349	98.9	15	0	0	0	1,334
Totals	16,170		9,287	1,437	3,118	565	1,763

Table VI-1. Test class performance matrix

What do the numbers in the performance matrix tell you about the classification? Look first at the 9,634 samples (pixels) of Coniferous forest. The table indicates that 9,110 of those pixels, or 94.6%, were correctly classified. Looking across that row, the table also indicates that 22 pixels which the analyst knows to be Coniferous forest were incorrectly classified as Deciduous forest, 53 were incorrectly classified as Grassland, 21 were incorrectly classified as Barren and 428 were incorrectly classified as Water. That is, there were 524 errors of omission for the 9,634 Confierous forest samples. Looking down for the column labelled Coniferous, there were 187 errors of commission for the Coniferous forest class. That is, 187 samples were called Coniferous that should not have been.

The five numbers on the major diagonal of the matrix can be summed and that total divided by the total number of samples is called overall performance. For Table VI-1, the overall perfomance is (9,110+1,286+2,988+34+1,334)/16,170 = 91.2%.

Often test fields are selected on the basis of what is readily visible in the gray scale images. This selection of the center of the large, more homogeneous areas as the basis for testing produces a bias in the classification accuracies obtained. A statistical method has been defined which practically eliminates this bias. Using this method we first divide the area classified into blocks of either two or three pixels on a side, such that a grid is formed over the data. Then blocks are selected out of the data set using a random number generator. The number of blocks selected usually corresponds to a percentage (commonly 5%) of the total data set. The cover type of each block is then identified from the reference data; mixed blocks are rejected. Blocks of the same cover type are grouped and used to evaluate the classification accuracy in the same way the "supervised" test fields were used. Although this method of selecting test fields requires more extensive reference data and personnel resources, it is more apt to provide a representative sample on which to base the evaluation.

Two additional guidelines for selecting test fields suggest that test fields should not fall inside of training areas and that they should together contain pixels for testing each cover type in roughly the same proportion as the cover types occur in the scene (the latter is relatively assured by the random selection method). For example, consider a scene that contains roughly equal amounts of forest and water. Suppose we decide to test with 100 pixels and choose 90 from water and 10 from forest and find that 80 of the water pixels are classified correctly while only one forest pixel is classified correctly. Is our classification then 81% correct overall? No - we have a <u>biased</u> estimate of classification accuracy. If we had chosen 50 water pixels and 50 forest pixels for testing and had the same per class accuracies for water and forest, we would have 45 water pixels classified correctly and 5 forest pixels classified

correctly. The unbiased estimate of classification accuracy is only 50%.

Self-Check

1. In Table VI-1, indicate the pixels correctly classified, errors of omission, and errors of commission for soybeans.

CHAPTER VII. CLOSING REMARKS

You have now completed step-by-step a numerical analysis of a MSS data set from the selection of the data set through the evaluation of the classification results. We have presented the analysis steps in a linear fashion. This linear presentation was adopted only because of convenience - in a "real life" analysis we would backtrack several times at various points in the analysis.

The analysis procedure used in the case study is a typical procedure based upon the experience of LARS researchers. However, it must be emphasized that the workshop has introduced you to an analysis procedure and not to the ultimate procedure. The analysis procedure used in the workshop was developed by making imaginative and intelligent use of available data processing algorithms. Different applications, perhaps your application, may well require a somewhat different approaches.

We now have classification results from our study area. The classification results themselves are not usually the real product of interest. Instead, an analysis is usually undertaken in order to gain information for use in such operations as forest management or land use planning. For instance, the analysis goals may involve learning where specific cover types are located or what proportion of the area belongs to each cover type, so that management and planning decisions can be made. To complete the analysis, the original analysis goals must be reviewed, and the desired information extracted.

Examples of results analysis and the extraction of useful information from multispectral data classifications may be found in several journals, including those listed here:

- -Pattern Recognition
- -Remote Sensing of the Environment
- -IEEE Transactions on Geoscience and Remote Sensing
- -Journal of Soil and Water Conservation
- -Photogrammetric Engineering and Remote Sensing
- -Agronomy Journal

Samples can also be found in a number of LARS publications, published proceedings of remote sensing conferences, etc.