083177

INSTRUCTOR'S NOTES

FOR

A CASE STUDY USING ECHO FOR ANALYSIS OF MULTISPECTRAL SCANNER DATA

OCTOBER 15, 1977

BY
RONALD K. BOYD
JAMES D. RUSSELL

THE LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY, WEST LAFAYETTE, INDIANA 47906

This work was supported by the National Aeronautics and Space Administration under contract NAS9-14970.

T-1314/4

These notes are being provided to assist those designated as instructors for the ECHO case study, until the revised version of the LARSYS Educational Package instructors notes is available. The revised version will include these notes and will also be revised for use with Data-100 remote terminals and CRT screens.

A CASE STUDY USING ECHO
FOR ANALYSIS OF MULTISPECTRAL SCANNER DATA

Instructor's Notes

Materials Required

1. Student copy of A Case Study Using ECHO for Analysis of Multispectral Scanner Data by Donna Scholz, James Russell, John Lindenlaub and Philip Swain (LARS Information Note 909177)

2. Reference copy of LARSYS User's Manual edited by T.L. Phillips

3. Reference copy of ECHO User's Guide edited by James Kast

4. "An ECHO Analysis of Run 73001011" with field description cards for test fields.

5. Reference copies of LARS Information Notes:

   110474 - An Introduction to Quantitative Remote Sensing by J.C. Lindenlaub and J.D. Russell

   111572 - Pattern Recognition: A Basis for Remote Sensing Data Analysis by P.H. Swain

   062375 - Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects by R.L. Kettig and D.A. Landgrebe

   121275 - Computer-Aided Analysis of SKYLAB Multispectral Scanner Data in Mountainous Terrain for Land Use, Forestry, Water Resources and Geologic Applications by R. Hoffer

6. Reference Data: Two kinds of reference material are needed for this case study.

   a) Four U.S. Geological Survey 7.5 minute quadrangle maps covering the western side of Indianapolis, Indiana, for use in the Exercises. These are not provided as part of the site library. The site techniques specialist should be responsible for ordering these maps. Three or more sets of maps per site are recommended. The maps may be ordered from:

      Distribution Section
      U.S. Geological Survey
      1200 South Eads Street
      Arlington, Virginia 22202

The names of the maps required are:

Clermont, Indiana
Indianapolis West, Indiana
Bridgeport, Indiana
Maywood, Indiana

A set of the USGS maps should be loaned to the students during their study and collected from them afterward.

a) Six color infrared aerial photographs. The prints should be loaned to the students and collected from them upon completion of the case study. sets of prints are provided as part of the site library.

7. Multispectral Image Storage Tape: Check to see that one of the data tapes assigned to your terminal has a copy of run 73001011 on it. If it does enter the tape number and file number below for easy reference:

73001011 Tape TTT = _____; File F = _____

If it does not, see page 6 of A Survey of the LARSYS Educational Package, which is part of Unit I of the Educational Package.

8. Intermediate results tape - Tape 920. A copy may be made using the *COPYRESULTS function in LARSYS.

Instructor's Notes

Student/tutor interaction is an important part of this case study. It is recommended that you meet with your students after each step in the analysis. At the beginning of the case study you will want to make sure that each student knows about the various reference materials and where to secure them.

It should be pointed out that even though the Exercises which give the students "Hands-On" experience using the computer for an ECHO analysis come at the end of the case study, the corresponding Exercises can be done after each section of the case study and its associated example. In trying out this case study with students it was found that the students preferred going entirely through the discussion of ECHO and the examples before doing the Exercises, hence this format was used. Perhaps you will want to try both ways with your students and determine which they tend to prefer.

Two possible study sequences:

SEQUENCE A	SEQUENCE B
Section I	Section I
Section II	Exercise 1
Section III	Section II
Section IV	Exercise 2
Section V	Exercise 3
Exercise 1	Section III
Exercise 2	Exercise 4

Exercise 3
Exercise 4
Exercise 5
Exercise 6
Exercise 7
Exercise 8
Exercise 9
Exercise 10
Exercise 11
Exercise 5
Exercise 6
Exercise 7
Exercise 8
Section IV
Exercise 9
Exercise 10
Section V
Exercise 11

The following comments relate to the generalized steps in classifying earth resources data as shown in Figure 6 of the case study and reproduced on the next page.

Section I - Discuss with each student the objective that he wrote for Self-Check I-B. This is also a good time to determine if he has any questions about the case study thus far.

Section II - Discuss with the student his answers to the self check to determine if he has any questions on the material presented thus far.

Section III

·Cluster Analysis: At this point the student may be confused as to how many clusters should be requested from a training area. The number of clusters chosen typically ranges from 1.5 to 2.0 times the number of covertypes present. The rule could be stated as: use 2.0x when counting only major covertypes, e.g., agriculture, wooded, residential, commercial, etc.; use 1.5x when subcategories are counted, e.g., soybeans, corn, deciduous, coniferous, new residential, old residential, etc. However, situations will arise in which after doing the clustering, a more optimum number of clusters will be suggested, as is reviewed on page 25 of the text.

·Association of cluster classes with information classes: the student should be asked what "these data" at the last of the fourth paragraph on p.37 refers to. Although it may be true that the cover types cannot be separated using the current data set (i.e., Landsat frame) it may only be necessary to ask for a larger number of clusters to separate them.

·Calculation of statistical distance between clusters: in the example in this section the suggested groupings from the grouping table at the end of the cluster output were utilized. The student should understand that the advice of that grouping table should only be taken when it suggests grouping classes of similar informational identity. An additional reason for scrutinizing the suggested grouping is that it is

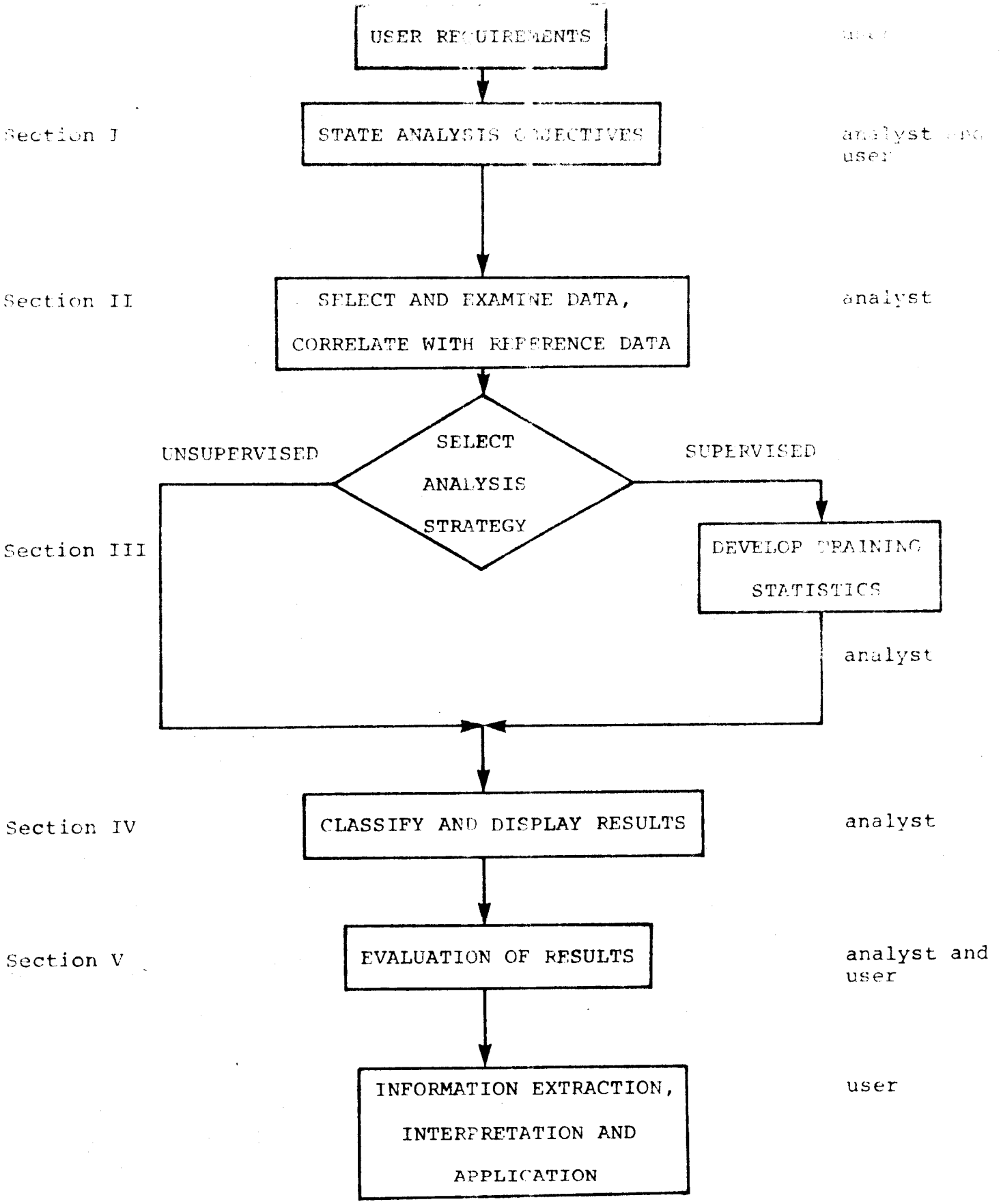

Figure 6: Generalized Steps in Classifying Earth Resources Data

entirely possible for another class from a different training area to be more like any of the classes the grouping table is suggesting grouping.

Also in the example a transformed divergence threshold of 1000 is used for the grouping table from SEPARABILITY. This is a rule of thumb suggested to prevent the student from getting suggested groupings of obviously different covertypes. Other larger values are possible and may be preferable.

Finally, in the example, the grouping table at the end of the SEPARABILITY output is accepted without question, the classes are grouped accordingly, and the resulting statistics deck is used to make the classification. This is a valid method, but not the only method, you may want the student to review pp.59-108 of the Landsat case study to get a different perspective.

Section IV - Discuss the ECHO classifier and the parameters necessary to run it with the student. It might be desirable to discuss how classifying a group of points instead of individual points could lead to higher classification accuracy.

Section V. - There are a few places in the data where test fields should not be picked, such as within training areas and over bad lines. Discuss this and the other criteria mentioned in the text with the student. In addition it might be useful to discuss the fact that training field accuracy is a measure of the separability of the training classes while test field accuracy indicates the representativeness of the training classes.

Exercise 1 - By now the student should not have too much difficulty accomplishing this. Discuss with him the user's requirements and insure that he understands the classes which are most important to separate. It may be beneficial to suggest 70% as the accuracy desired.

Exercise 2 - Let the student determine what a "representative set of channels" is and discuss his decision with him. Also he may not be aware that in order to obtain his grayscale maps at the same scale as the 7.5 minute topographic maps a line and column interval of one must be used. This necessitates getting a triple width printout, which he may need instruction on how to assemble.

Exercise 3 - It is important to emphasize that in this step the student's primary goal should be to gain familiarity with the study area and the location of landmarks on the grayscale maps. When he has completed this exercise he should be able to point out the general location of commercial, industrial, residential, and agricultural areas as well as the major landmarks.

It may be necessary to point out that although the text of the exercise begins by talking about means of referencing cluster maps and photos, that will be done in a later exercise. However, the techniques discussed are useful anytime two pieces of information must be compared with one another.

Exercise 4 - Analysis tasks vary greatly in the amount and quality of reference data that is available. This always has a significant impact on the objectives of the analysis. This case study is an example of an analysis task where the reference data is of good quality but limited to six aerial photographs and topographic maps which are 10 years old. It is suggested that you allow the student the chance to discover this and decide how it impacts the task before him on his own. He may go ahead and select his training areas or he may approach you with questions. In general it is best to have him pick the training areas on his own with a minimum of input from you. In this way students can learn from obvious mistakes, such as

a) Selecting training areas not covered in any aerial photography. In some cases this might be acceptable, however, the topographic maps are several years out of date and lack specificity, thus making correct identification of training classes later on difficult or impossible.

b) Selecting training areas that do not have distinctive landmarks, making exact location of the training area in the reference data much more difficult.

Although up to date reference data is limited to the six photographs, and it is strongly recommended that the training

areas be selected within them, the student still has some options:

a) It is not necessary to use all six photos as the training areas, a subset will suffice.

b) It is possible to use only parts of the photos as training areas. However, care should be taken to insure that the training areas selected are large enough to represent all the possible conditions of the covertypes present.

Exercise 5 - If this is the second case study your student has done he may be counting subcategories as covertypes, in which case he should request 1.5x clusters. If time and resources allow, permit the student to test his reasoning by using the numbers he arrives at and then going back and trying a different number of clusters if he is not satisfied with his first set.

In the example in the text statistics on the resulting clusters were punched for use in later steps. Ask the student to explain why it is necessary to punch the statistics before he knows if he will use them or not. His answer should include the concept that the machine 'forgets' what the clusters were when the job finishes.

Exercise 6 - The student should understand that when he is done associating his cluster classes with the reference data he should have a description of what each cluster represents on the ground. Some students become confused when two or more cluster maps have the same symbols, as if the classes with the same symbols should represent the same covertype. Therefore, it may be a good idea to underscore that each training area was clustered individually and that therefore clusters represented by the same symbol do not necessarily need to represent the same covertype. Encourage the student to do as much as possible on his own, but don't hesitate to demonstrate the process or assist him if he has difficulties.

Exercise 7 - Query the student to insure he understands that the grouping table is merely an aid and does not have to be exactly followed, especially when it suggests grouping clusters representing different covertypes.

In order to use the MERGESTATISTICS function and the SEPARABILITY function that produces a grouping table on the basis of transformed divergence it is necessary to either IPL LARSYSDV at the terminal or insert the I LARSYSDV card behind the normal batch job cards. For example:

```
ID        BATCH
BATCH MACHINE BATSHORT
BATCH ID TECTRA DONNA SCHOLZ
BATCH OUTPUT COMPUTER COMPUTER
I LARSYSDV
```

```
*MERGESTATISTICS
 .
 .
 .
 .
END
```

The MERGESTATISTICS control cards are available from the terminal by IPL'ing LARSYSDV and typing REFERENCE MERGESTATISTICS. The student may need help with the pool cards.

Exercise 8 - If time allows have the student experiment with different transformed divergence thresholds so he may see what is grouped in each case. In addition it may be desirable for him to draw a separability diagram. See pages 59-108 of the Landsat case study.

Exercise 9 - As ECHO is a developing processor, its control cards may change from time to time, current control card listing for SECHO is available from LARSYSDV in the same manner as for MERGESTATISTICS and SEPARABILITY. When generating intermediate results it is necessary to specify an intermediate tape and file number, not a results tape and file number. SECHO may be ran on LARSYSDV at the terminal or by batch in the same manner as MERGESTATISTICS and SEPARABILITY.

The following intermediate results for the Indianapolis area are available for the student to experiment with.

Tape 920

File 1	Cell size 2x2	Homogeneity Threshold	20.0
File 2	Cell size 2x2	Homogeneity Threshold	40.0
File 3	Cell size 2x2	Homogeneity Threshold	60.0
File 4	Cell size 2x2	Homogeneity Threshold	80.0

It is highly desireable for the student to experiment with these parameters and different annexation thresholds. If he uses the intermediate results already available this will not consume a large amount of computer time.

Exercise 10 and 11 - The student will probably want to try different symbol combinations. It would be most efficient to have him determine his classification accuracy at the same time he is making a display map.

Have the student select a set of test fields using the topographic maps and any of the aerial photographs he did not use as training areas. Have him use the ones he selects to evaluate the classification. It is possible to group training classes for testing. For example the hypothetical classes Urban1 and Urban2 could be tested against a single set of urban test fields if they were grouped for testing using the group option in PRINTRESULTS.

Discuss with the student reasons for differences in his

accuracy figures using his test fields and those supplied with the site library. A set is supplied with the site library to provide a benchmark against which the student's test fields can be compared and to allow student to student comparisons. Typical accuracies for a new student are in the neighborhood of 70%. It is not necessary for the student to repeat the analysis if his accuracy is lower than this. It is quite possible that a student achieving 80% accuracy would learn just as much as one achieving 60% accuracy. Nevertheless it would be good for the student to consider what he could do differently that would increase his accuracy.

Summary - As the main thrust of this case study is to broaden the students knowledge of analysis, the instructor should help the student to consider how and why some analysis steps may be done differently and where a given approach is most appropriate. People who understand those things will be the most valuable when it comes time to apply what they have learned.