

LARS Technical Report 020179

Computer-aided Analysis Techniques
for Mapping Earth Surface Features

by R. M. Hoffer

Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
1979

This material was prepared, by invitation, for inclusion in a book entitled "Remote Sensing Applications in Agriculture and Hydrology", G. Fraysse, ed., to be published by A. A. Balkema, publishers, Rotterdam, Netherlands. Purdue Agricultural Experiment Station Journal No. 7471.

COMPUTER-AIDED ANALYSIS TECHNIQUES FOR MAPPING EARTH SURFACE FEATURES

Roger M. Hoffer
Purdue University, West Lafayette, Indiana, U.S.A.

1 INTRODUCTION

The need has long existed for a capability to obtain reliable information over large geographic areas in a timely manner. Such a need is present in many discipline areas. Because of the very rapid changes in the condition of agricultural crops and the influence of crop yield predictions on the world market, the need for accurate, timely information is particularly acute in agricultural information systems. For these reasons, as remote sensing technology has developed over the past few years, the potential for using this new technology has been receiving increased and widespread attention.

It is apparent that high flying aircraft or satellites are capable of collecting enormous quantities of data over vast geographic areas in a relatively short period of time. Such masses of data can be collected using a variety of sensor systems, each of which has its own particular advantages, as well as disadvantages. However, there is a major step between the collection of the data and the reduction of this data into useful information. A key factor in developing remote sensing technology, therefore, involves the data analysis techniques that can most effectively reduce the masses of data collected into the type of information which is required by the user.

One data analysis technique that has been developed over the past decade involves use of digital computers and the application of pattern recognition theory to multispectral scanner (MSS) data, obtained by either aircraft or satellites. This particular approach was conceived and developed by the Laboratory for Applica-

tions of Remote Sensing (LARS) at Purdue University (LARS Staff, 1968). This technique is directed at more effective utilization of the capabilities of computer systems in mapping and tabulating earth surface features over large geographic areas in a timely manner. This approach is not an attempt to develop a totally automatic data processing system but rather is directed at developing procedures for computer-aided analysis of remotely sensed data. Experience has shown that input by knowledgeable interpreters is a key and essential ingredient to the effective analysis of multispectral scanner data. One could therefore refer to these techniques as "automated", but it would be incorrect to call them "automatic", because the man-machine interaction is a definite requirement in using this type of analysis system.

In the development of computer-aided analysis techniques, multispectral scanners have been utilized as the primary data source because data from these sensor systems can be easily quantified and subsequently processed by a digital computer, and also because the multispectral data format is ideally suited for pattern recognition analysis. Many of the analysis techniques developed were based on the use of aircraft scanner data.

The launch of Landsat-1 (formerly ERTS-1) opened a new dimension in our capability to obtain data at any time of the year (depending on cloud cover). Since the primary data collection system on Landsat-1 is a 4-band multispectral scanner, there has recently been a considerably increased interest in computer-aided analysis techniques for processing this data. Several new capabilities for handling this type of data have been developed, offering con-

siderable potential for meeting agricultural inventory and land use mapping needs. It is the purpose of this paper to describe some of the basic as well as the newly developed techniques for working with multispectral scanner data obtained by aircraft or by satellite.

2 PROCESSING AND ANALYSIS OF MSS DATA BY DIGITAL COMPUTER

Current procedures for digital processing and analysis of data from multispectral scanner systems involve five primary areas of activity:

- Data Reformatting and Preprocessing,
- Definition of Training Statistics,
- Computer Classification of Data,
- Information Display and Tabulation, and
- Evaluation of Results

2.1 Data Reformatting and Preprocessing

Procedures for reformatting and preprocessing of the data do not involve any data analysis per se, but simply involve changing the characteristics of the raw data so that it is in a better format for the analysis sequence. With aircraft data, collected by a multispectral scanner system such as that shown in Figures 1 and 2, the major reformatting activity that is frequently encountered involves converting the data from an analog to a digital format. This is accomplished by a specially designed A/D (Analog-to-Digital) conversion system. The frequency of digitization, level of detail to which the data is digitized, and other factors involved in the A/D conversion can significantly influence the characteristics of the digital data.

In preprocessing the data, a number of things can be done to change (and hopefully improve) the data characteristics and quality. For aircraft data, these might include such changes as scan-line averaging, geometric corrections, sun-angle corrections, and data calibration. The rotational velocity of the scanner and the ground-speed of the aircraft must be correctly synchronized during collection of MSS data to avoid over-lap or under-lap of the individual scan lines. Very often when the aircraft is flying at low altitudes, the individual scan lines are overlapped to a large degree. When this occurs, only the n^{th} scan line of data is utilized in some cases, so that the over-lap does not cause serious geometric distortion in the data. In other instances, the scan lines can be averaged (using one of several

possible weighting formulas) in order to improve the signal-to-noise (S/N) characteristics of the data.

Geometric corrections of the data are often applied to aircraft-obtained MSS data because of the geometric distortions caused by the differences in look-angle of the scanner system as the Instantaneous Field-of-View changes from one side of the scan line to straight down to the other side of the scan line. The resultant data on both sides of the scan-lines are compressed in relation to the data obtained directly below the aircraft (See Figure 1). A common result of this geometric distortion caused by the rotational motion of the scanner mirror system and the data compression on the edges of the data is the so-called S-curve seen in roads or other linear features that cross the flight path of the aircraft at an angle. Figure 3 shows a good example of this type of distortion. Such distortions in the data can be corrected in the data reformatting phase of the data processing sequence.

Due to the relatively low altitudes of aircraft data collection missions, severe sun-angle effects have often been found in the data. These are similar to the back-scatter and hot spot (or forward-scatter) effects often seen on aerial photos. However, in collecting scanner data, if the aircraft is flying in a direction such that the individual scan lines are pointed toward or away from the sun, severe sun-angle effects can result. Averaging of the columns of scanner data along the entire flight line, and then adjusting the radiometric values of each data point in each scan line can largely correct these effects (Landgrebe, 1972).

Data calibration can be utilized to adjust the amplitude of the data from one channel (or wavelength band) to the next through the use of calibration pulses in the scanner data. One of the most useful data calibration procedures involves thermal infrared scanner data. If the scanner system being utilized is equipped with hot and cold calibration plates, the thermal infrared data can be adjusted in relation to the energy being emitted by these plates, and accurate values of the radiant energy being emitted by the ground can be obtained. This procedure is particularly useful for remotely obtaining temperatures of water bodies (Bartolucci, et al., 1973).

Thus far in this discussion of data

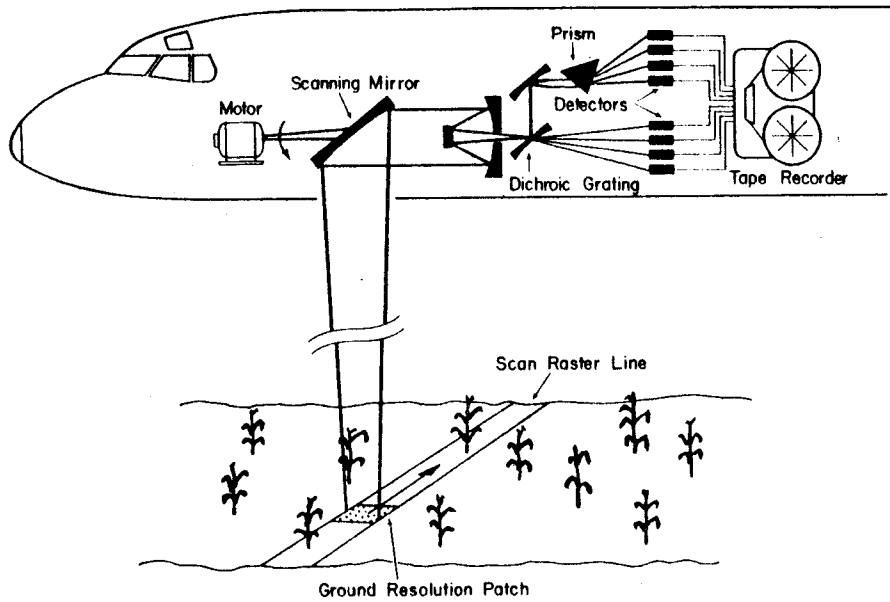


Figure 1. Schematic of a Multispectral Optical Mechanical Scanner.

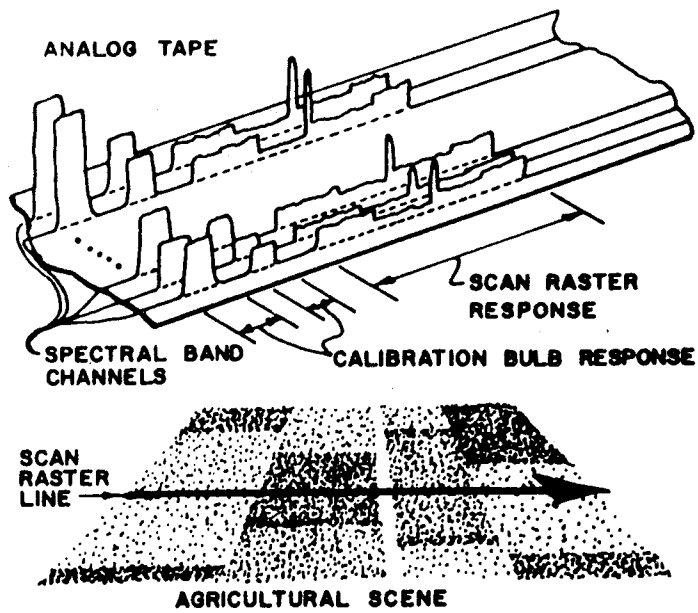
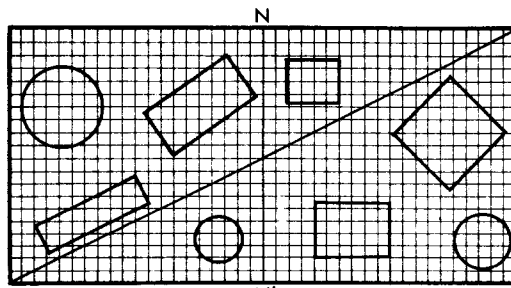
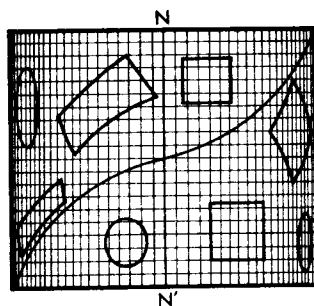


Figure 2. Data Collection Format, Using a Multispectral Scanner and Multiple-Track Recorder.

reformatting and preprocessing, we have focused our attention on aircraft data. However, since the launch of Landsat-1 in July 1972, several capabilities for reformatting and preprocessing this type of scanner data have been developed. Of particular importance are the following procedures: reformatting the scanner data to allow a full Landsat frame to be contained on a single data tape, geometrically correcting and scaling the data to a common map base, and overlaying multiple sets of digital data (including sets of Landsat data obtained on different dates, and overlaying Landsat data to other data bases).



Terrain assumed flat
N-N' = nadir line



Resultant geometry distortion

Figure 3. Geometric Distortion Due to Scale Compression on the Multispectral Imagery (from Tanguay, 1969).

The reformatting of the Landsat data is, in most cases, a fairly straightforward procedure that allows the subsequent data analysis activities to be carried out in a much more effective manner. Originally, the data is received on four, two, or one data tape(s) which contain, in total, a

single frame of Landsat data. In the case of four tapes, the data is in 7-track, 800 b.p.i. format, and each tape covers an area of approximately 46 kilometers (km) along a scan line (cross-track from the orbital path), and 185 km along the orbital path. The reformatting procedure allows the four data tapes to be merged onto a single 1600 b.p.i., 9-track data tape containing the entire 185 x 185 km area (approximately 115 by 115 statute miles, or an area of approximately 8,475,485 acres, or 3,429,896 hectares, or 13,243 square miles).

It could also be noted that each resolution element or "pixel" (as it is frequently called) of Landsat data represents an area on the ground of slightly over 0.46 hectares (1.1 acres) (approximately 56 by 79 meters, or 184 by 256 feet). An entire frame of Landsat data contains 2,340 scan lines of data and each scan line contains 3,560 resolution elements, resulting in a total of 7,553,520 individual data resolution elements. Since the data is collected in each of four wavelength bands, there are over 30,000,000 individual reflectance measurements contained on the data tape representing a single frame of Landsat data. The amplitude of each reflectance measurement ranges from 0 to 64, due to the method in which the signal is initially digitized (2^6 digitization accuracy). Since the satellite is capable of collecting a new frame of data every 25 seconds, one sees that the total data collection capability of this type of satellite scanner system is enormous! Growing appreciation for the large quantities of data being obtained and the ways in which digital computer systems can process this type of data has been a major factor in the recent interest in computer-aided analysis of multispectral scanner data.

The geometric rectification and scaling of Landsat MSS data has been a particularly significant data processing procedure. At LARS, this procedure was developed in 1973, and involves a five step sequence in which a frame of Landsat data is rotated, de-skewed, and rescaled to the specifications of the user (Anuta, 1973). The development of this procedure was necessitated by the geometric distortions found in the basic "system corrected" or "bulk" Landsat-1 data tapes. The geometric distortions were caused in part by the orbital path followed by the satellite and by the earth's rotation during the time in which the scan line of data is collected. Although NASA applies some 14 different op-

erations to the geometric rectification of Landsat imagery, the data tapes are not corrected for geometric distortions (Anuta, 1977). With such distortions present, the data analyst often could not be sure of his location in the data in relation to a particular location on the ground. The only input required for the basic geometric correction program is the reformatted data tape, and the latitude and longitude of the center point for the frame of data involved. The use of the "system corrected" or "bulk" data tapes allows these geometric correction procedures to be carried out without any loss in the radiometric quality of data. The usual output of the program is a geometrically corrected data tape which, if every resolution element is displayed on a standard computer line printer, results in a 1:24,000 logogrammatic printout, oriented with north at the top. Use of this scale allows the analyst to overlay the printout directly on 7½ minute U.S. Geological Survey topographic maps, or other 1:24,000 scale maps or images. This has proven to be extremely beneficial in helping the analyst locate various features of interest in the data (Hoffer, et.al., 1973).

Since the development of the initial geometric rectification program, a "precision" geometric correction procedure has been developed in which ground control points are located in the Landsat data, and a registration procedure is applied to the data to correct for the geometric distortions and provide an even more accurate output product from a cartographic standpoint. The mathematical details of these procedures have been fully described by Anuta (1973 and 1977).

A procedure for overlaying multiple sets of Landsat digital data has also been developed and offers considerable promise for many types of applications. In this procedure, the geometrically corrected data tape from one date is overlaid with a geometrically corrected data tape from a second, third, or more dates. A data set can thus be developed that allows the differences in spectral response between two different dates to be effectively utilized in identifying various cover types on the surface of the earth, or in delineating areas where changes have taken place in the earth surface cover. For example, in one study in the Rocky Mountains of Colorado, data from six different dates throughout the year were overlaid, and the increase and decrease of the area of the snowpack was determined as a function of date. Thus, in that particular

study, the overlay procedure resulted in a single data tape containing all 24 channels (4 wavelength bands x 6 dates), so that the analyst could easily determine the change in snow cover for any particular location on the ground from one date to another. This capability for overlaying multiple sets of satellite data has many potential applications in land use monitoring, forestry, agriculture and other disciplines where there is a need to determine changes in the characteristics of earth surface cover. It should be pointed out, however, that the ability to effectively utilize the overlay products has not yet been completely defined. Basically, the reason for this is that data on the tape is simply a set of reflectance measurements. Thus, just because there is a change in spectral response from one date to the next does not necessarily mean that a change in spectral response could be due to normal seasonal changes, such as trees leafing out, crops maturing, or even such things as water standing in fields after a heavy rain. It is therefore very important for the human interpreter to evaluate the changes in spectral reflectance recorded by the scanner system in order to differentiate those that are due to normal seasonal changes or weather conditions just prior to the time the data were obtained from those actual land use changes of importance.

Another major type of data overlay procedure that has been developed involves the use of Landsat data in conjunction with other ancillary data sources such as topographic data (elevation, slope, aspect), land ownership, political boundaries, watershed boundaries, soil type boundaries, etc. In overlays of these types, the result is a data base file which can be easily and effectively manipulated by the resource manager to combine various portions of the data base with the Landsat classification results, as needed. This particular capability for data manipulation has tremendous potential for the future use of digital computer systems in a wide variety of applications.

2.2 Definition of Training Statistics

The essence of computer-aided analysis of multispectral scanner data involves "training" the computer to recognize a particular combination of numbers representing the reflectance in each of several wavelength bands from a particular material or cover type of interest. After a good set of training statistics has been developed, the computer is programmed to classi-

fy the reflectance values for every resolution element for which reflectance measurements were obtained by the multispectral scanner system.

The real key to an accurate computer classification of the data lies in the definition of a set of training statistics which are truly representative of the spectral characteristics of the various earth surface features present in the multispectral scanner data. This leads to the need for the analyst to develop a thorough understanding of the spectral characteristics of earth surface features. Without an in-depth understanding of the spectral characteristics of the various materials with which he is involved, the analyst will not be able to be truly effective in developing an optimal set of training statistics to utilize in classifying the MSS data.

As indicated in Figure 4, different earth surface features have different amounts of reflectance in the various wavelength bands. One finds, however, that the spectral reflectance characteristics of a particular material are not always unique (Hoffer and Johannsen, 1969, and Sinclair, et.al., 1971). In many cases, different species of green vegetation have very similar reflectance patterns, and one also finds a certain degree of variation within any particular species. Therefore, it is not a straightforward procedure to define a particular spectral pattern that can be used to train the computer. Thus, one of the first

questions encountered in computer-aided analysis of multispectral data involves the identification of the type of categories or classes of material that the computer should be trained to recognize. Basically, there are two conditions which must be met by each class involved in an analysis of remote sensor data using pattern recognition techniques:

- The class must be spectrally separable from all other classes
- The class must be of interest to the user or have informational value.

In working with multispectral scanner data, one soon finds that often the classes of informational value cannot be spectrally separated at certain times of the year. One reason for this is that various species of green vegetation have very similar spectral characteristics, even though their morphological characteristics may be quite different. The need for a class to be both separable and have informational value therefore leads to two quite different approaches in training of the computer system.

The first approach is referred to as the "supervised technique" and involves use of a system of X-Y coordinates to designate to the computer system the locations in the data of known earth surface features that have informational value. For example, at a certain X-Y location in the data is a field of corn, at another is a field of soybeans, another contains wheat, pasture, etc. (See Figure 5). This super-

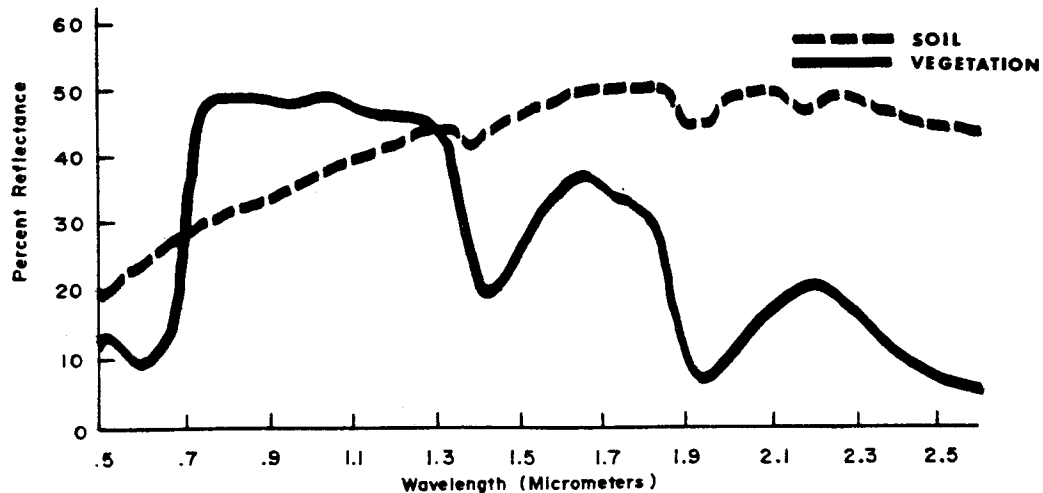


Figure 4. Spectral Reflectance Curves for Turgid Green Vegetation and Air Dry Soils. These curves represent averages of 240 spectra from vegetation and 154 spectra from air dry soils. This type of data is frequently referred to as representing the spectral signature for the materials of interest (from Hoffer, 1971).



Figure 5. Illustration of Digital Scanner Data Showing Locations of Training Areas.

vised technique or "training sample" approach has been used quite effectively for agricultural mapping (LARS Staff, 1968, Bauer and Cipra, 1973). One must be constantly alert, however, to recognize situations in which two classes may be of great informational value and interest, but which cannot be separated spectrally. In such situations these classes can be combined if there is a sound, logical reason for doing so (e.g., wheat and oats could be combined into a class defined as small grains, but wheat and soybeans could not be logically grouped).

A second approach to training the computer system involves the "clustering" technique (sometimes referred to as the "non-supervised" technique). In this approach the analyst simply designates the number of spectrally distinct classes into which the data to be classified should be divided. The computer is programmed to classify the data into the designated number of spectral classes and then print out a map indicating which resolution elements in the data belong to which spectral class (see Figure 6). The analyst then simply relates this classification output map to known surface observation data, and determines which materials actually are represented by each of the different spectral classes (e.g., spectral Class 1 is wheat, Class 2 is corn, Class 3 is bare soil, etc.). The difficulty with this approach lies in the inability of the analyst to knowledgeably define the number of spectral classes present. Also one finds that the classes of most interest often have subtle

spectral differences, whereas many of the other classes present in the data may be easily separated spectrally but are of little informational value or interest to the user. Experience at LARS has indicated that a combination of the two systems seems to be the most satisfactory and most effective procedure to follow. This is particularly true in wild land or other spectrally complex geographic areas.

A so-called "modified clustering approach" has recently been developed and has proven to be extremely effective in working with satellite multispectral scanner data, both from the Landsat and Skylab scanner systems (Fleming, et.al., 1975). In this method, several small blocks of data are defined, each of which contains several cover types. (See Figure 7). Each area or data block is first clustered separately, and the spectral classes for all cluster areas are subsequently combined. In essence, the modified cluster approach entails discovering the natural groupings present in the scanner data, and then correlating the resultant spectral classes with the desired informational classes (crop species, cover types, vegetative conditions, etc.). This technique is particularly useful in wildland areas, or where the fields are small, or where the cover types and spectral classes are complex. In most cases, less than one percent of the data involved in the analysis is used for the training phase.

Whichever method is utilized to develop the training statistics, one must always keep in mind that these data simply represent the spectral characteristics of the various cover types or earth surface features on the ground. Many of the data characteristics that a photo interpreter would utilize to identify a particular cover type of interest (such as shape, size, texture, shadow, association, etc.) are not used in the computer classification of the data.

2.3 Computer Classification of the Data

After the training statistics are defined, using the supervised, (non-supervised), modified-clustering or some other method, the data must be classified, using one of several classification algorithms. The essence of the theory behind most of the classification algorithms is illustrated in Figure 8. As this figure indicates, different earth surface cover types often have distinctly different spectral response

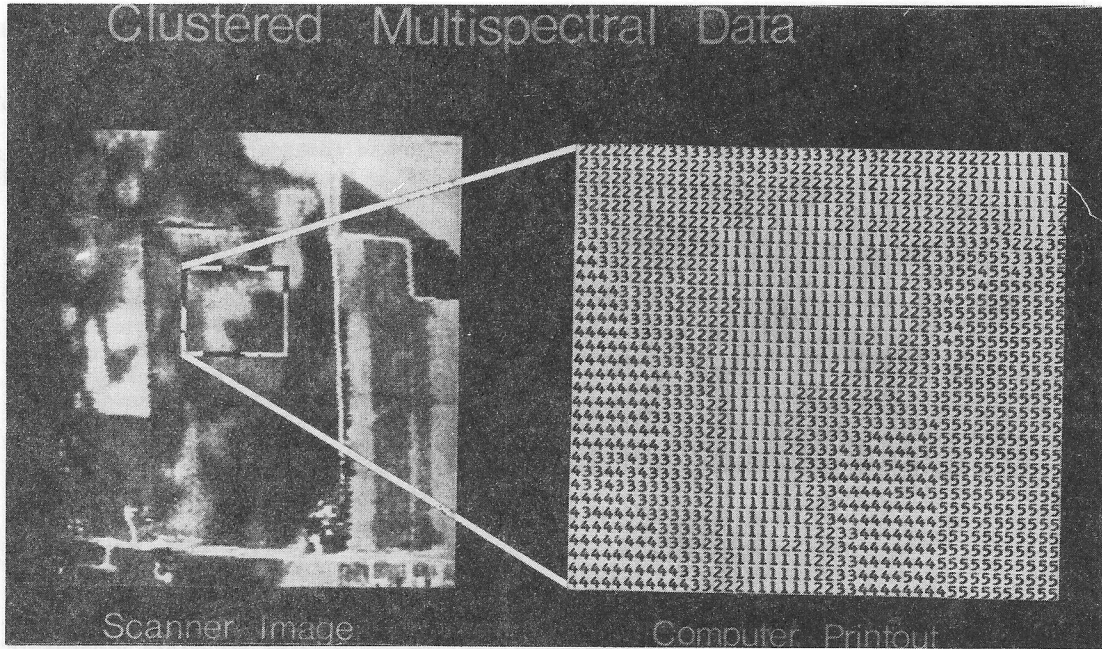


Figure 6. An Agricultural Area in Which the Clustering Algorithm Has been Used to Define Five Spectral Classes.

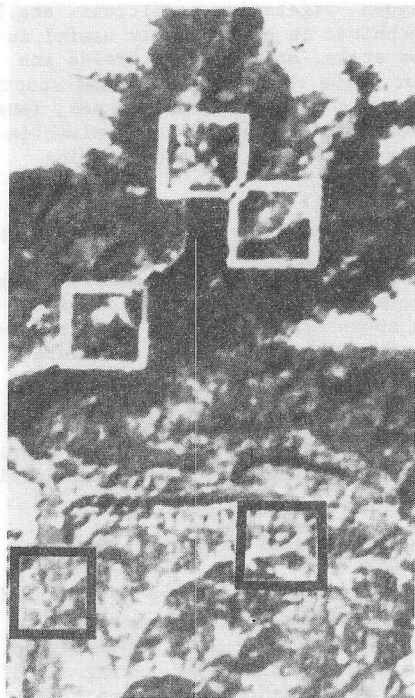


Figure 7. Digital Display Example of Landsat Data Illustrating the "Modified Cluster" Approach to Developing Training Statistics.

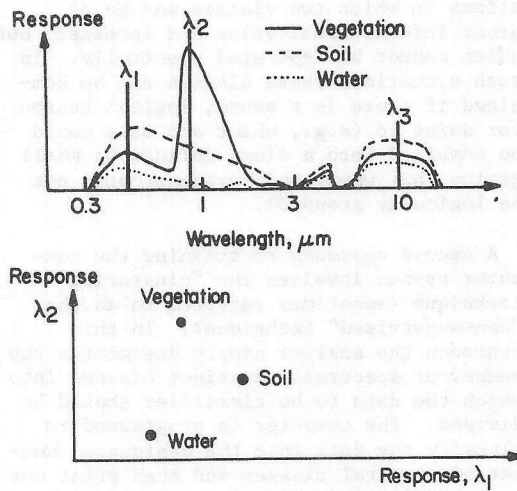


Figure 8. Spectral Data of Basic Cover Types Represented in Two-Dimensional Feature Space.

patterns. Multispectral scanners are designed to measure the relative reflectance or emittance in designated wavelength bands, as indicated by λ_1 , λ_2 , and λ_3 . By plotting the relative reflectance (response values) of vegetation, soil, and water for λ_1 versus λ_2 , one sees that these cover types occupy very different locations in two-dimensional space. Of course, in the real world, agricultural materials will have some degree of variability, and will therefore tend to occupy an area in such a two-dimensional space diagram, rather than a single point. A theoretical example of such a situation is illustrated in Figure 9. In this case, response values for data known to have come from soybeans,

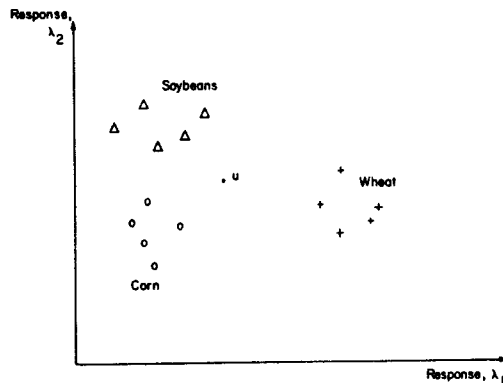


Figure 9. Agricultural Crop Species Defined as Training Samples in Two-Dimensional Feature Space.

corn, and wheat have been plotted in two-dimensional space. These data would be considered the training set. An unknown data point has been plotted at Point "u" near the center. The problem now is to define a classification algorithm that will divide this two-dimensional feature space into regions that can be used to "classify" any unknown data points. Any unknown data point would therefore be classified into one of the three categories for which the computer has been trained. One of the simplest classification algorithms is the "minimum distance to the means" in which the mean of each group of data is calculated, and then boundaries are defined which define the minimum distance between the means (as illustrated in Figure 10). In this case, the unknown Point "u" would be classified by the computer as soybeans, since it fell into that portion of the two-dimensional feature space.

Of course, in the actual analysis of multispectral scanner data, the computer is not limited to working in two-dimensional feature space, but works with numerical data vectors that can represent many wavelength bands of data as more and more wavelength bands of data are involved, and as the classification algorithm becomes more complicated, the classification time (and therefore cost) can increase significantly. A good example of the relationship between number of wavelength bands and classification time is shown in Figure 11. This figure also shows that the classification accuracy

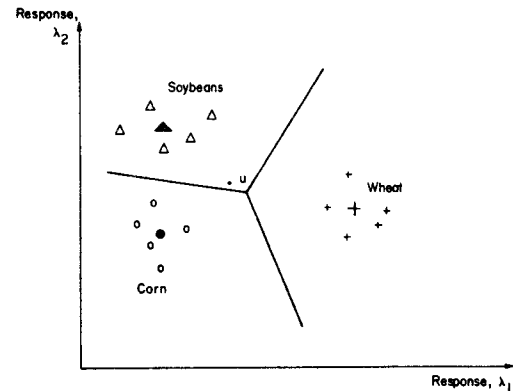


Figure 10. Minimum Distance to Means Classification.

does not increase linearly, or even increase at all, as the number of wavelength bands utilized is increased.

In the classification of multispectral scanner data, many different algorithms are available for use. However, one algorithm in particular--the maximum likelihood algorithm--has been used with considerable success by LARS researchers and others throughout the world. This algorithm has generally been found to be rather universally applicable to any cover type and produces classification results that usually have a relatively high degree of accuracy (as compared to other classification algorithms). Therefore, the maximum likelihood algorithm is often utilized as the "standard" against which the classification results from other algorithms are compared.

In the classification sequence involving the maximum likelihood algorithm, the spectral response data associated with each resolution element sensed by the scanner system is examined by the computer

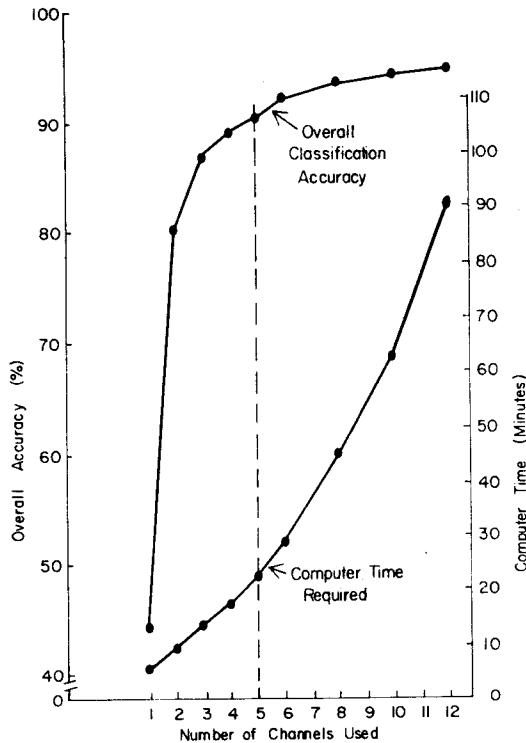


Figure 11. Overall Classification Accuracy and Computer Time Required as a Function of the Number of Channels Used. (From Coggeshall and Hoffer, 1973.)

and assigned to one of the spectral classes defined during the development of the training statistics. These classification results are then stored on magnetic tape, and the analyst can subsequently display these results in a variety of output formats. A key point in this classification procedure is that each data point or resolution element in the data is classified and displayed independently. There are many other algorithms that can be utilized (such as the parallel-piped approach used in the G.E. Image 100 system, the "layered" classification techniques, the "levels classifier", and others) to classify the data on a point-by-point or individual resolution element basis.

There are also several classification schemes which have been or are being developed which allow groups of spectrally similar resolution elements to be classified as a single unit. One such technique (developed at LARS) is referred to as the "ECHO" (Extraction and Classification of Homogeneous Objects) classifier. In this technique, the computer is programmed to

define the boundary around an area having similar spectral characteristics, and then the area within the boundary is classified as a single spectral class. This technique is somewhat similar to the so-called "Per-Field" classifier, except that with the Per-Field classifier, the analyst has to define the field boundaries, whereas with the ECHO classifier, the computer is programmed to define the boundaries. Other algorithms having somewhat similar characteristics have also been developed by other groups throughout the world.

The primary point that should be stressed in discussing computer classification of MSS data is that no matter which classification algorithm is utilized, the training statistics must be representative of the spectral characteristics of the various cover types present in that data set. If the training statistics are not representative, the classification results will not be satisfactory, no matter which classification algorithm is utilized (i.e., garbage in = garbage out)!

2.4 Information Display and Tabulation

Digital computer-aided analysis of remote sensor data allows a variety of output products to be obtained. These can basically be divided into the two following categories:

- Display of classification results in map format, and
- Display of classification results in tabular format.

Maplike displays of the computer classification results can be obtained in one of several different formats. One easily obtained type of output is the so-called "logogrammatic" display obtained from the computer line printer. For this output, the analyst selects various symbols to represent each of the different classes of interest such as C for corn, S for soybeans, D for deciduous forest, F for coniferous forest, W for water, etc. Such an output would be similar to that shown in Figure 12. When the symbols represent the various cover types that have been classified, such a map output would be referred to as a "thematic" map, since each symbol represents a different "theme" or cover type. One could also obtain a map showing the location of the various spectral classes defined in the training procedure, even if the cover type or characteristics of the various cover types were not yet known. In this case, the map would be referred to as a "spectral" map.

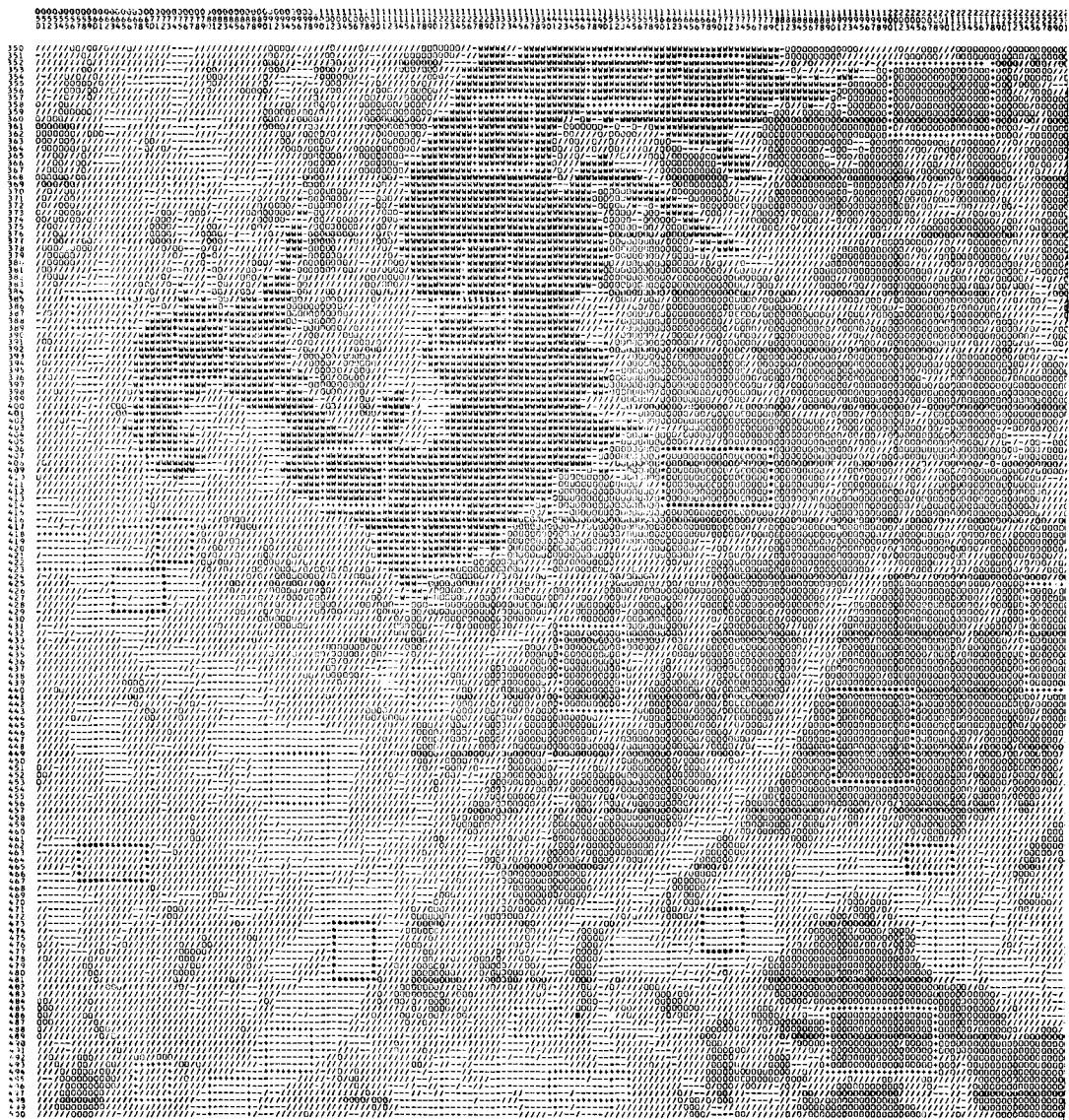


Figure 12. Line-Printer "Thematic" Map Showing Classification Results Based on a "Supervised" Approach Using Landsat Data. W = Water, O = Forest, - = Bare Soil, and / = Pasture. Training areas are outlined by * and Test areas are outlined by + symbols.



Figure 13. LARS Digital Display Unit with Light Pen for Interfacing with the Data.

Similar types of outputs, but having a very different appearance can be obtained through the use of a digital display device (as shown in Figure 13) in which each of the different classes of interest are displayed as a different color or possibly as a different tone of grey, depending on the type of display device available. This output provides a much more photographic-like output of the classification results and also has the advantage of displaying much larger geographic areas on a relatively small image. Many, many other types and formats of output devices are also available, including such things as the ink-jet plotter, digital film writers, Calcomp plotters, "varian" dot grid plotters, and others.

Tabular outputs of the classification results are particularly useful for acreage determinations for any particular area of interest. In this case, the analyst designates to the computer the X-Y coordinates representing the boundary of a test area (such as a quadrangle, county, watershed, etc.). The computer then summarizes the number of data points classified into each of the various cover type categories. Since each data point or resolution element of satellite data represents a certain area on the ground (approximately 0.46 hectares per resolution element for Landsat), this conversion factor can be applied to determine the number of hectares of each of the various cover types of interest. The percentage of the entire area covered by each of the cover types of interest can also be rapidly and easily calculated.

2.5 Evaluation of Results

Classification of large geographic areas can be accomplished very rapidly using computer analysis techniques. However, one must be able to verify the accuracy of such computer classification results. Are the resultant classification maps and tables reasonably accurate, and do they have a reasonable degree of reliability? Several different techniques have been developed and utilized to evaluate such computer classification results. Our experience at LARS has been that a combination of three different techniques provides the best overall indication of the classification accuracy. A qualitative evaluation of the classification results can be obtained by visually comparing the classification to an existing cover type map or to aerial photos of the region. Although the method is subjective, it does provide a quick, rough estimate of the accuracy of the classification. However, quantitative evaluation techniques allow more definitive evaluations of the computer classification results to be obtained.

One quantitative evaluation technique involves a sample of individual areas of known cover types which are designated as "test areas." Test areas are similar to training areas, in that they consist of blocks of data, designated by X-Y coordinates, for which the actual cover type has been determined using some form of reference data (e.g., visual observation on the ground, interpretation of large scale aerial photos, etc.). However, test areas are not utilized in developing the statistics on which the computer is trained. To avoid any possible bias on the part of the analyst, the test areas should be located prior to the classification and, if possible, should be located by means of a statistical sampling design. In many cases, however, the location of test areas tends to be governed by the availability of adequate reference data and also by an evaluation of the importance of statistically determining the classification performance in relation to the cost of obtaining this information. No matter which method is used in defining the location of test areas, the procedure followed is for the cover type classification obtained by the computer for the various test areas to be tabulated and compared to the actual cover type present in the test areas. This tabulation can involve the individual test areas or can be summarized for the entire set of test areas. An example of such a classification

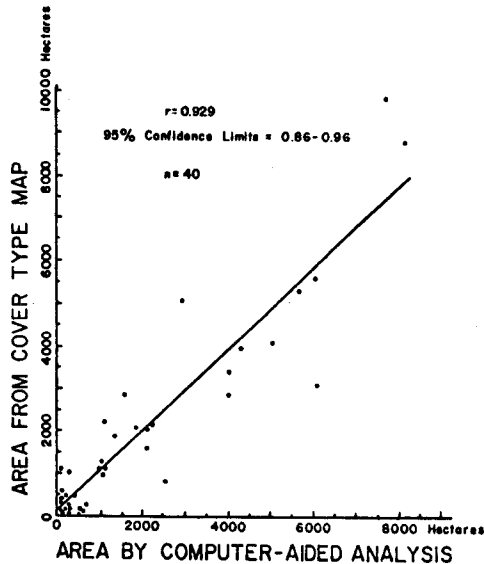


Figure 14. Acreage Estimates of Cover Types Obtained by Computer-Aided Analysis of Skylab MSS Data, Compared to Aerial Photo Interpretation Estimates.

One of the more common procedures for computer-aided enhancement of multispectral scanner data is to display individual wavelength bands of the original aircraft or satellite data through an appropriate color filter and obtain a "false color composite" of the imagery. With such procedures, one is combining three wavelength bands of scanner data onto a single image. Because the Landsat scanner system obtains two wavelength bands of data in the visible portion of the spectrum (green and red), as well as two in the reflective infrared wavelengths, one can obtain a combination of wavelength bands similar to that involved in normal color infrared film. Therefore, "false" color infrared composites of Landsat scanner data have been widely utilized. The reflected infrared portion of the spectrum allows many earth surface features to be effectively enhanced. Such "false" color infrared composites can be made with computer analysis and enhancement procedures through the use of a digital display device such as that shown previously in Figure 13. Of course such false color composites can also be made through the use of other optical enhancement devices, of which many types are on the market and are used very effectively.

Another type of enhancement that can be achieved with digital computer analysis techniques involves computing the ratios between individual wavelength bands, which are then displayed in various color com-

bination formats. In such instances, one can use a variety of ratio combinations. For example, you could obtain the ratio of a red band compared to an infrared band, or vice-versa; the ratio of two thermal infrared bands could be obtained; or one could add two visible bands together and two infrared bands together, then compute and display the ratio results.

In addition, there are many other more complex mathematical functions that can be applied to the data to obtain various types of enhancement effects. However, unless there is a theoretical basis for such functions, their value is questionable in terms of developing an operational procedure for processing multispectral scanner data in a reliable, predictable manner.

Along with the simple display and enhancement of the multispectral scanner data, one can also use digital computers very effectively to obtain an enlargement capability of small scale MSS satellite data. This enlargement of digital data has proven to be of particular importance in obtaining detailed images of the scanner data that have not been degraded by optical enlargement techniques. An example of such an enlargement sequence is shown in Figure 15. In essence, each frame of Landsat scanner data contains 3,560 sample points or resolution elements per line of data and there are 2,340 lines of data per frame. However, the digital display unit shown in Figure 13 has a capability of displaying only 800 resolution elements, or so-called pixels (picture elements) per line, and 500 lines of data. Therefore, to fit an entire Landsat frame of data on the screen requires that a subsample procedure be utilized in which only every fifth line and fifth column of data is initially displayed. Such an image is considerably degraded from the original but it does allow one to get a general look at the entire frame of scanner data. This has often proven useful when only the digital data tape was available. One could also display a portion of the entire frame using every fourth line and fourth column, or third line and third column or second line and second column, etc. If one displays every line and column over a small portion of the frame, a very high quality image can be obtained for that small portion of the entire Landsat frame as shown in Figure 15. A procedure has also been developed to display each resolution element measured by the satellite scanner system as four picture elements on the display screen (as shown in Fig. 15),

Table 1. Computer Classification Results Showing Test Field Performance for Aircraft Data Using Five Wavelength Bands (from Coggeshall and Hoffer, 1973).

Serial Number ----- 831219706 Classified - Sept. 6, 1972

Channels Used

Channel 4	Spectral Band 0.52 - 0.57 μm
Channel 6	Spectral Band 0.58 - 0.65 μm
Channel 9	Spectral Band 1.00 - 1.40 μm
Channel 10	Spectral Band 1.50 - 1.80 μm
Channel 12	Spectral Band 9.30 - 11.70 μm

Classes

	<u>Class</u>		<u>Class</u>
1	Deciduous	4	Forage
2	Conifer	5	Corn
3	Water	6	Soybean

Test Class Performance

<u>Group</u>	<u>No. of Samples</u>	<u>Percent Correct</u>	<u>Number of Samples Classified Into:</u>					
			<u>Deciduous</u>	<u>Conifer</u>	<u>Water</u>	<u>Forage</u>	<u>Corn</u>	<u>Soybean</u>
Deciduous	32,252	92.2	29,745	1,001	0	378	245	883
Conifer	88	96.6	3	85	0	0	0	0
Water	339	98.2	1	2	333	3	0	0
Forage	11,760	85.5	22	7	2	10,052	413	1,264
Corn	2,679	90.7	1	4	0	191	2,431	52
Soybean	2,676	95.7	10	0	0	91	15	2,560
Total	49,794		29,782	1,099	335	10,715	3,104	4,759

Overall performance: $45206/49794 = 90.8\%$

Average performance by class: $558.9/6 = 93.2\%$

performance evaluation tabulation is shown in Table 1.

A second quantitative method of evaluating the computer classification results is to compare acreage estimates obtained from the computer classification of satellite data to those obtained by some conventional method, such as manual interpretation of aerial photos. If an adequate number of relatively large areas are summarized, a statistical correlation can be obtained, thereby enabling a quantitative comparison between the computer-derived acreage estimates and the acreage estimates derived from conventional sources. Figure 14 is an illustration of such a comparison, in this case using a classification involving Skylab data.

There are many variations and refinements that can be incorporated into the various analysis and evaluation techniques. However, the general approaches described above have been found to be most effective

for computer-aided mapping of general agricultural and forest cover types, utilizing aircraft, Landsat, or Skylab multispectral scanner data.

3 COMPUTER-AIDED ENHANCEMENT OF THE DATA

Computer-aided enhancement of multispectral scanner data can assume many forms, and is sometimes utilized in lieu of computer classification of the data. Therefore, this aspect of data processing could be considered as an alternative to the definition of training statistics, classification, and display of results, as discussed above, although the reformatting and preprocessing, and the evaluation of the results obtained might still be involved. It is important to note that usually the enhancement of MSS data produces an image that must then be manually interpreted. In this regard, computer-aided enhancement is quite different from computer-aided classification.

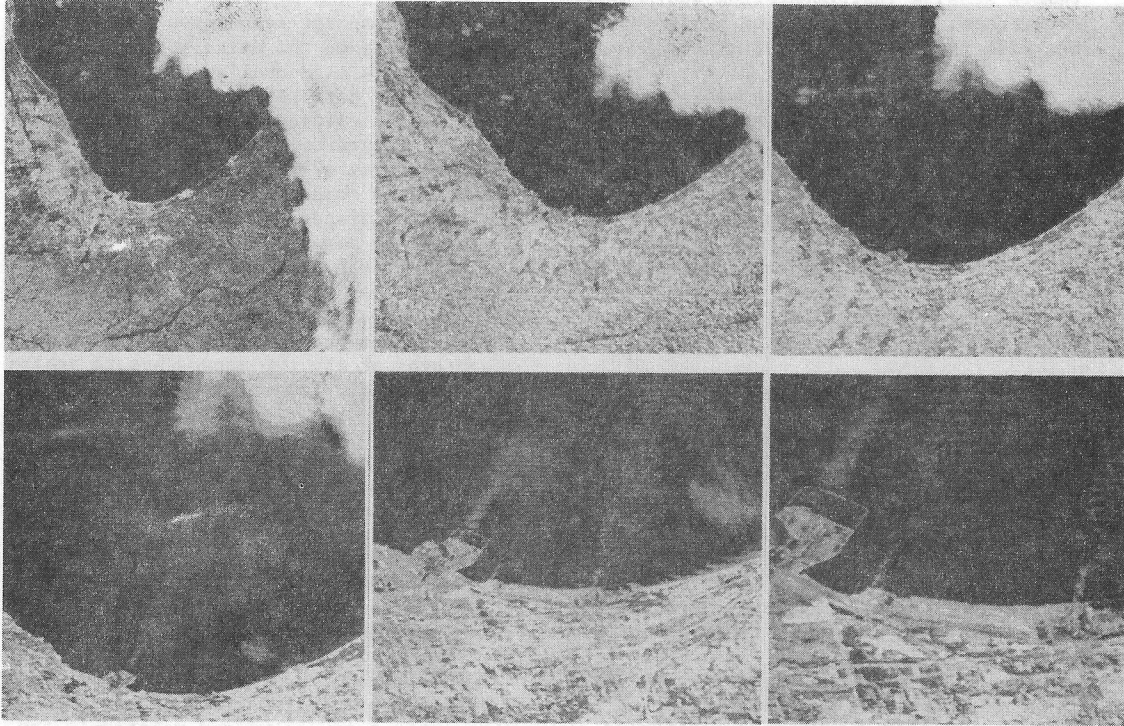


Figure 15. Examples of Computer Enhanced Landsat Data, Showing the Potential for Significantly Enlarging Small Sections of the Landsat Data. (Area shown is the southern tip of Lake Michigan and City of Chicago.)

or even 16 picture elements. Thus, even though an individual element might be too small to discern with the eye on the display screen because that resolution element is displayed as a single pixel, by displaying it as a cluster of four or sixteen pixels, a large enough area of the display screen is occupied by that resolution element so that it can be seen. In this way, very high quality images can be obtained for small geographic areas imaged by the Landsat scanner system. Details not easily seen on the original Landsat imagery (1:1,000,000 scale) are brought out effectively through this enlargement procedure. Many times there is also a tremendous amount of detail actually sensed by the Landsat scanner system that is not apparent until the data was put through such an enlargement procedure.

In addition to these general types of enhancement procedures, there are many other more specific types of enhancement routines to which the data can be subjected, depending on the particular application and type of data involved. The key point involved in the use of enhancement procedures as compared to classification procedures, however, is that the classification re-

quires that the computer be trained, whereas most enhancement procedures are fairly straightforward and can be applied with much less man/machine interaction being required.

4 ADDITIONAL CONSIDERATIONS IN COMPUTER-AIDED ANALYSIS OF MULTISPECTRAL SCANNER DATA

Thus far, we have considered the procedures for handling multispectral scanner data, but very often questions are raised concerning the applicability of one analysis technique versus another, or questions come up concerning the best time of the year to obtain remote sensor data for mapping and tabulating a particular cover type. Therefore, we need to examine a few of these questions, and compare the alternatives.

4.1 What Sensor System or Analysis Technique is "Best"?

In examining the possibilities for using remote sensing to help obtain accurate and timely information involving agricultural resources, one must review the advantages and limitations of the various

remote sensing instrumentation and analysis techniques, so that the user needs can be met with the most efficient system possible.

As an initial step in defining the type of remote sensor systems and analysis techniques which might be most effectively utilized, one needs to define his user requirements, or information needs. In doing so, it is often helpful to consider a series of questions related to the characteristics of remote sensor systems. These can frequently be grouped into spectral, spatial, and temporal considerations, as shown below.

What are your information needs???

- Spectral Considerations

- What are the earth surface features or cover types of interest?
 - Can they be spectrally separated from the other associated cover types?
 - Which wavelength regions or bands are most useful for spectrally differentiating and identifying the cover types of interest?
 - Are the spectral characteristics of these earth surface features unique or more distinctive during some particular time of the year?

- Spatial Considerations

- What size area is involved?
 - Do you need complete coverage of the area, or will a sample of the entire data set be adequate?
 - What format is required for the results -- maps and/or tables?
 - If maps are the final product, what scale and degree of accuracy is required?
 - What are the spatial characteristics of the cover types of interest, as compared to the characteristics of the instrumentation involved?

- Temporal Considerations

- How frequently is the information required?
 - What time of year is best (or required) for obtaining the information?
 - Are there special diurnal considerations involved in the data collection?

To be most effective, computer-aided analysis techniques require a quantitative data acquisition system. In most instances, multispectral scanner systems are the most suitable for providing quantitative data inputs to computer-aided analysis systems. As previously indicated, MSS data

can be classified by computer-aided techniques or computer techniques can be utilized to enhance the data and then the enhanced data is manually interpreted. Photographic data rather than MSS data can sometimes be obtained, in which case manual interpretation is usually the analysis technique utilized. Questions are often raised concerning the relative advantages or disadvantages of manual interpretation as compared to computer classification of the data. To develop the most useful operational system for analyzing multispectral scanner data; one must therefore consider the advantages and disadvantages of the two approaches to analyzing the data. Some of these considerations can be summarized as follows:

- A computer classification (using multispectral scanner data, probably from spacecraft altitudes) would potentially be suitable if:

- Geographic area of interest is very large (state, country)
 - Informational categories of interest are spectrally separable
 - Spectral characteristics of data are relatively simple (spectral classes are reasonably homogeneous and relatively few in number)
 - Spatial relationships are not required to achieve identification

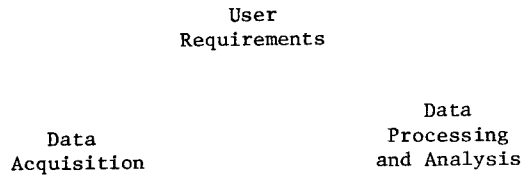
- Manual interpretation would probably be more suitable if:

- Geographic area is relatively small (town, county)
 - Requirement for detailed spatial information exceeds capabilities of multispectral scanner systems. (In this case, photographic data would have to be utilized).
 - "Convergence of evidence" principal is required to identify features of interest
 - Spectral characteristics of data are complex and difficult to characterize

From these types of considerations, it is clear that the "best" data collection system and analysis technique is largely a function of the particular problem with which the user of the information is involved. What may be most suitable in one situation may be quite unsuitable in another situation. Therefore, each situation must be evaluated independently, based on its own information requirements.

From this discussion, it becomes apparent that there are three basic components that

are involved in the use of any remote sensor system:



These three components are very closely interrelated, and the user requirements as well as the data acquisition system must be taken into account when deciding which data processing and analysis procedure to utilize. In every case, however, it is the User Requirements that are (or at least should be) the driving force behind the entire operation!

4.2 Temporal and Spatial Aspects of Spectral Response Patterns

Agricultural crops are particularly noteworthy in terms of very rapid changes in spectral characteristics at various times in the growing season (Sinclair, et.al., 1971). In late May, winter wheat presents a fairly solid canopy of lush green vegetation to the remote sensor, but by late June in central Indiana the same winter wheat will be golden brown and nearing maturity. Two weeks later, it has probably been harvested and one will see only the highly reflective yellow straw. In many cases, another two weeks' delay will allow many weeds and green understory vegetation to mix with the straw and an appearance very much like grazed pasture or perhaps hay is observed. Such rapid changes bring out the importance of understanding the seasonal changes of the crops or other earth surface features of interest.

Many times, people have discussed the usefulness of "crop calendars" which can be developed for any particular geographic area and will describe the general characteristics of the various cover types of interest as a function of the time of year. However, one must remember that such crop calendars can vary from one year to the next, depending on weather conditions of that particular year. This is particularly true early in the spring when excessive rains can cause significant delays in planting dates in any particular year. It should also be recognized that such crop calendars can be developed more effectively in areas of the world where seasonal changes are distinct. For example, in the United States there are relatively

narrow periods of time during which corn or soybeans are planted, grown, and subsequently harvested. However, in tropical regions, seasonal restrictions are not as apparent. For example, in Southeast Asia, rice is planted and harvested over a rather wide time span. Such a broad time frame can be even more complicated by second plantings of rice and by distinctly different agricultural practices from one region to the next.

Along with gross seasonal changes in spectral characteristics of agricultural crops and other earth surface features, one can also encounter more short-term temporal variations, such as differences in spectral response at different times of the day or night, and caused by a host of variables. Differences in sun angle certainly needs to be considered, as do variations in atmospheric attenuation. In some cases, vegetation that is not under moisture stress early in the morning will show severe symptoms of moisture stress later in the day. A good example of this is the way in which the top leaves on soybean plants will tend to turn bottom-side-up in the early part of the afternoon on many warm midsummer days. This apparently is a protective mechanism which occurs when the plant cannot obtain sufficient soil moisture. Since the lower surfaces of the soybean leaves are a much lighter green than the top of the leaves, the entire soybean field becomes a lighter green, more highly reflective vegetative surface. Thus, by reducing the amount of solar energy absorbed by the plants, the moisture stress is minimized to a certain extent.

In considering seasonal changes of agricultural crops, one also encounters the problem of definition of a particular cover type of interest. Suppose you are interested in utilizing remote sensing to map acreage of soybeans throughout Indiana. At what stage of development are you going to define a particular agricultural field as being "soybeans"? Do you call Field X a field of soybeans after the beans have been planted? Or after emergence? Or when the beans are four inches high? Ten inches? Or is it not until the soybeans are covering 25% of the ground surface? Or 50%? Etc., etc.??? Work with multispectral scanner data from aircraft has shown, for example, that the difference in reflective characteristics between corn and soybeans is very small. This has also been substantiated with laboratory spectral measurements (Sinclair, et.al., 1971; Hoffer and Johannsen, 1969). To put it another way, green vegetation tends to look very

much like other green vegetation. We found that the ability to spectrally differentiate between corn and soybeans during the time period near the end of June was highly correlated with the percentage of ground cover rather than any large distinctive difference in the spectral characteristics of the soybean and corn vegetation itself. Early planted corn fields could be reliably distinguished from soybean fields because of higher percentage of ground cover and smaller percentage of exposed soil involved in each resolution element measurement by the scanner system. However, corn planted later in the season was found to occupy about the same percentage of the ground surface as the soybeans did, and therefore, if one was facing the situation where 50% of the ground surface was bare soil and the other 50% was green vegetation, it did not seem to matter whether the green vegetation consisted of soybeans or corn. The very small spectral differences that do exist between soybean and corn canopies could not be distinguished because of the more powerful influence of the soil/vegetation mixture that was being measured by the scanner system! This type of problem has haunted many people involved in the analysis of multispectral scanner data of agricultural crops, particularly if the data was obtained early in the growing season. Careful selection of data sets to avoid such confusing problems whenever possible seems to be the most logical solution to these types of problems at the present time.

Another aspect of spatial variability of spectral signatures involves geographic variability of various categories or crop species of interest. One finds that the same crop species does not have the same spectral response pattern in all geographic locations on any one date. For example, wheat may be harvested in southern Indiana where it has reached maturity but has not yet been harvested in central Indiana and perhaps is still immature and green in northern Indiana and southern Michigan. Thus in discussing the spectral "signature" concept, it is impossible to define a single spectral response pattern that will be applicable for the same crop species in all geographic areas at any one time.

Another associated aspect of geographic variability of agricultural crops involves the idea that not all crop species are found in all geographic locations. Therefore, knowledge of location from which remote sensor data was obtained can prove quite useful in attempting to identify a

particular crop species, even though the spectral response pattern of that crop may not be well known at that time of the year because of lack of "ground truth" data. For example, if one analyzed a Landsat frame obtained on August 15th, and found that 40% of the total area of this Landsat frame consisted of small, rectangular fields of green vegetation, one would conclude that these were agricultural crops of the same species. If the interpreter then knew that the data came from north central Indiana and Illinois, and he had a reasonable knowledge of the agricultural characteristics of this area, he would immediately conclude that the crop species described by that particular spectral response pattern was corn rather than cotton or rice or some other crop species which does not grow in this geographic area to that extent. This points out the need for knowledge of the geographic location over which the remote sensor data was obtained and for knowledge of the agricultural characteristics of that particular geographic region, or in other words, there is a distinct need for a man-machine interaction in the process of analyzing remote sensor data.

4.3 Data Banks or the Extrapolation Approach?

Another subject that often arises in computer-aided analysis of remote sensor data involves two rather different concepts used in training the computer to recognize a given set of spectral response patterns (assuming for now that a supervised training mode is being utilized). These two concepts could be termed the "data bank" and the "extrapolation mode" for obtaining data to be used in the training procedure. In the "data bank" approach, the concept is to have many sets of spectral signatures available in computer storage. Such a "data bank" would include signatures for all different cover types of interest. When a new set of remote sensor data is obtained, one simply selects the signatures for the cover types of interest from the existing data bank, and uses these as training samples to classify the entire data set. The advantage is a great reduction in the amount of time required to train the classification processor as to the characteristics of the spectral signatures involved in each new set of data. Such a system would be highly desirable and, in theory, could work under certain selected sets of conditions. However, in the real world situation, such a concept does not appear to be practical in most cases, particularly for those cover types

or earth surface features that have distinct temporal variations in spectral response. The reasons for this involve the temporal effects upon the spectral characteristics of the earth surface features which are measured and which one is attempting to classify. Many studies have shown that the spectral characteristics of different types of vegetation change drastically as a function of time, both for a given growing season and from one year to the next. Plant maturity causes a nearly continuous change in spectral characteristics of agricultural crops. For instance, what is often measured early in the growing season is the amount of bare soil in proportion to the amount of green vegetation present in the instantaneous field of view of the scanner, rather than a unique spectral signature of the different crop species. Since the farmer must depend upon weather conditions to govern the date of planting for the various crop species, the height of the plants (and therefore the canopy coverage and the vegetation-soil relationship) will be quite different from one year to the next as of the same date on the calendar. The weather conditions governing the germination and rate of growth of the vegetation after planting are also important, and will cause considerable variation in spectral response throughout the growing season and from one growing season to the next. There are also difficulties in determining differences in atmospheric attenuation from one flight mission to another, and in being able to calibrate sensor data to a high degree of radiometric precision.

In view of such problems of variation from one flight mission to the next, it would seem that an extrapolation mode is much more logical for obtaining spectral data with which to train the computer for the classification task. In this case one simply uses the data collected under the existing atmospheric conditions and conditions of plant maturity, stress, growth, etc., abstracts the training samples from this same data set and then proceeds with the classification and analysis task. Of course, this approach assumes that a reasonable amount of ground truth or reference data is available for that particular flight mission.

5 SUMMARY AND CONCLUSIONS

When one considers the current situation of our agricultural and other natural resources, along with the predictions of continued population growth, a person cannot help but be concerned about the manner

in which we will be utilizing these resources during the next few decades and beyond. One of the major factors which will influence many decisions on use of these resources is the amount and detail of information which is available about the extent and condition of the resource base.

It becomes apparent that computer-aided analysis techniques are required if we are to take full advantage of our ability to collect data at frequent intervals over vast geographic areas. Further developments in the handling of spatial and temporal data, in addition to spectral data will bring significant improvements in computer-aided data analysis. It also seems clear that both satellites and aircraft will continue to be used in the collection of remote sensing data, and that ground observations will continue to be a necessary and integral requirement for accurate data analysis. However, even though major improvements in computer-aided analysis techniques will be achieved, man's contributions will always be required in the analysis of remote sensor data. A good understanding for instance, of the energy-matter inter-relationships and the spectral reflectance and emittance characteristics of the various cover types of interest will become increasingly important. Also, there will continue to be a major requirement for accurate and detailed manual photo interpretation in many application areas. All of these will be crucial in the maintenance of an effective man-machine interaction in the data analysis process.

Remote sensing has already proved to be a very useful tool in several areas of application in the management of our natural resources, and the potential applications seem almost unlimited. It is our goal to develop a technology that will aid in obtaining the best possible information concerning the location, extent, and condition of our resource base, and to make this information available in a timely manner to the people involved in the management of our agricultural, forest, water, and other resources.

6 REFERENCES

- Anuta, P. 1973, Geometric Correction of ERTS-1 Digital Multispectral Scanner Data, LARS Information Note 103073.
- Anuta, P.E. 1977, Computer-Assisted Analysis Techniques for Remote Sensing Data Interpretation, *Geophysics* 42(3): 468-481.
- Bartolucci, L.A., R.M. Hoffer, and T.R. West 1973, Computer-Aided Processing of Remotely Sensed Data for Temperature Mapping of Surface Water from Aircraft Altitudes, LARS Information Note 042373.
- Bauer, M.E. and J.E. Cipra 1973, Identification of Agricultural Crops by Computer Processing of ERTS MSS Data, Proceedings of the Symposium on Significant Results Obtained from ERTS-1, Goddard Space Flight Center, p. 205-212. Greenbelt, Maryland.
- Coggeshall, M.E. and R.M. Hoffer 1973, Basic Forest Cover Mapping Using Digitized Remote Sensor Data and ADP Techniques, LARS Information Note 030573.
- Fleming, M.D., J. Berkebile, and R. Hoffer 1975, Computer-Aided Analysis of LANDSAT-1 MSS Data: A Comparison of Three Approaches Including the Modified Clustering Approach, Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, p. 1B-54 to 1B-61. Purdue University, West Lafayette, Indiana. Also available as LARS Information Note 072475.
- Hoffer, R.M. and C.J. Johannsen 1969, Ecological Potentials in Spectral Signature Analysis, P.L. Johnson (ed.), Chapter 1 in *Remote Sensing in Ecology*, p. 1-16. University of Georgia Press, Athens, Georgia. Also available as LARS Information Note 011069.
- Hoffer, R.M. 1971, Remote Sensing Potentials for Resource Management, Proceedings of the Third International Seminar for Hydrology Professors, p. 211-227. Purdue University, West Lafayette, Indiana.
- Hoffer, R.M. 1972, ADP of Multispectral Scanner Data for Land Use Mapping, Invited paper presented at the 2nd UNESCO/IGU Symposium on Geographical Information Systems. Ottawa, Canada. Also LARS Information Note 080372.
- Hoffer, R.M. and LARS staff 1973, Techniques for Computer-Aided Analysis of ERTS-1 Data, Useful in Geologic, Forest and Water Resource Surveys, Proceedings of the Third ERTS-1 Symposium, 1(A): 1687-1708. Goddard Space Flight Center, Washington, D.C. Also LARS Information Note 121073.
- Landgrebe, D.A. and Staff 1972, Data Processing II: Advancements in Large Scale Data Processing Systems for Remote Sensing, Proceedings of the 4th Annual Earth Resources Program Review, II: 51-1 to 51-31, NASA MSC Publication No. 05937.
- LARS Staff 1968, Remote Multispectral Sensing in Agriculture, Laboratory for Applications of Remote Sensing, Vol. 3. Research Bulletin No. 844, Agricultural Experiment Station and Engineering Experiment Station, Purdue University, West Lafayette, Indiana.
- Sinclair, T.R., R.M. Hoffer and M.M. Schreiber 1971, Reflectance and Internal Structure of Leaves from Several Crops During a Growing Season, *Agronomy Journal* 63(6): 864-868. Also available as LARS Information Note 122571.
- Tanguay, M.G. 1969, Aerial Photography and Multispectral Remote Sensing for Engineering Soils Mapping, Joint Highway Research Project, Report No. 13, Purdue University, West Lafayette, Indiana.