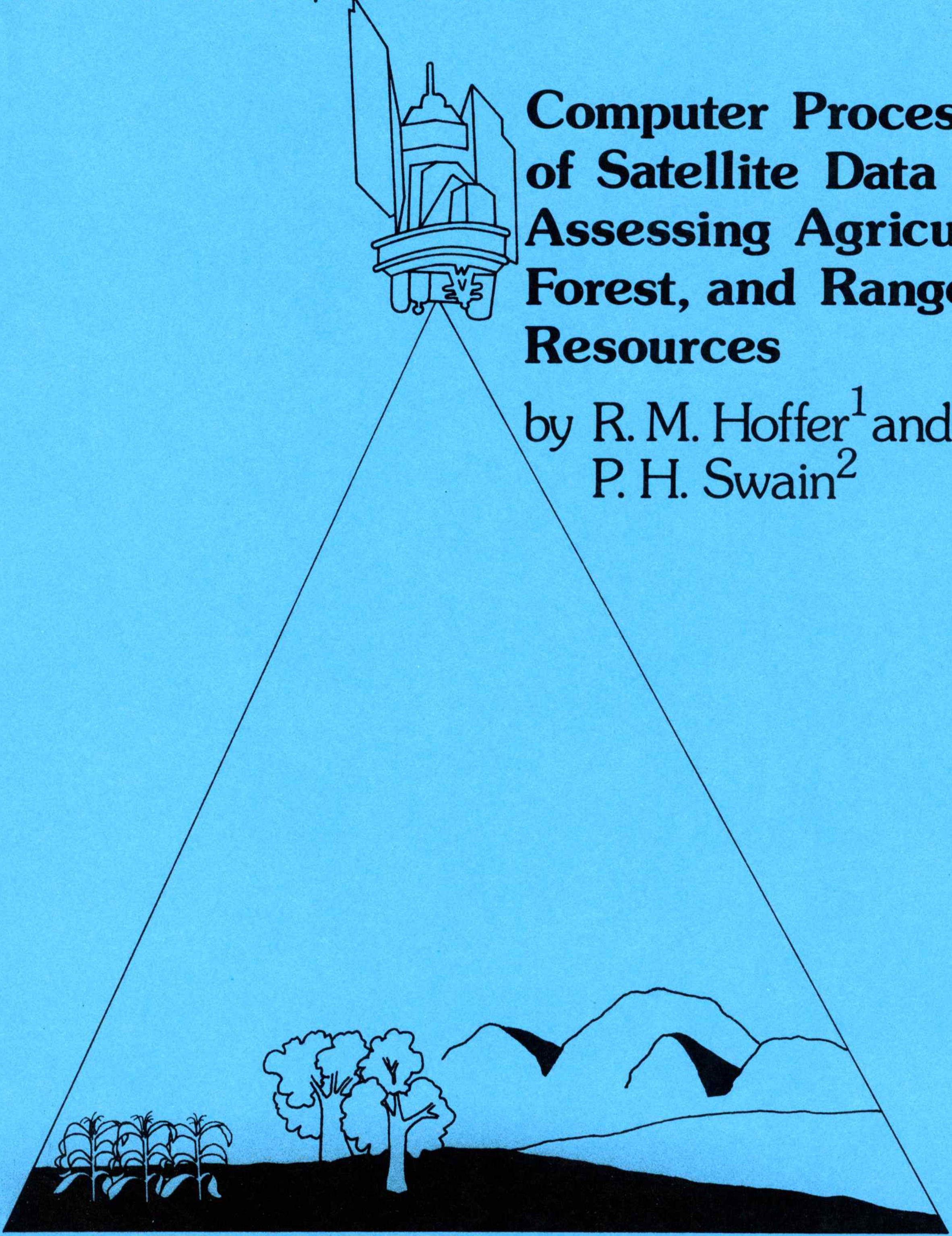
LARS Technical Report 072580

Computer Processing of Satellite Data for Assessing Agricultural, Forest, and Rangeland Resources

by R. M. Hoffer[1] and
P. H. Swain[2]



Department of Forestry and Natural Resources,[1]
School of Electrical Engineering,[2]
Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
1980

Reprint of paper from the International Archives of Photogrammetry, Vol. XXIII, Part B7, Commission VII, published by the XIV International Congress for Photogrammetry, Hamburg, Germany, 1980.

14TH CONGRESS OF THE INTERNATIONAL SOCIETY OF PHOTOGRAMMETRY
Hamburg, Germany -- 1980
Commision VII

Invited Paper

COMPUTER PROCESSING OF SATELLITE DATA FOR ASSESSING AGRICULTURAL, FOREST, AND RANGELAND RESOURCES

by

Roger M. Hoffer
Professor, Department of Forestry and Natural Resources,
Purdue University, West Lafayette, Indiana, U.S.A.

and

Philip H. Swain
Assoc. Professor, School of Electrical Engineering
Purdue University, West Lafayette, Indiana, U.S.A.

Abstract

The quantity of data produced by multispectral scanner systems from satellite (or even aircraft) altitudes clearly indicates the need for effective computer-aided analysis techniques. Since the launch of Landsat-1 in 1972, significant progress has been made in developing and refining a variety of techniques for processing and analyzing such data.

This paper first discusses three basic aspects of computer processing of MSS (multispectral scanner) data, including preprocessing, enhancement, and classification. Recent developments in classification techniques are then discussed. Consideration is given to the use of ancillary data as part of the classification process. The ECHO algorithm and the layered classification technique, which appear to be particularly important for effective analysis of earth resource features, are described. The paper concludes with a look to the future of remote sensing data collection systems and analysis techniques.

I. Introduction

Developments in remote sensing technology over the past fifteen years have produced entirely new dimensions in our capabilities to collect and analyze earth resource data. The advent of multispectral scanner (MSS) systems has been particularly significant, since data from MSS systems is in a quantitative format, ideally suited to analysis using pattern recognition techniques. Since 1972, when Landsat-1 was launched, a fantastic amount of digital format data has been collected--data covering nearly all areas of the world. In many regions, data has been obtained in different seasons of the year and for different years, thereby offering a tremendous potential for providing a wealth of resource information needed by a variety of discipline groups.

It is clear, however, that there is a major step between the collection of data and the reduction of such data into useful information! A critical factor in remote sensing technology, therefore, involves the various data analysis techniques that can most effectively reduce the volume of data collected into the type and format of information required by the user. The purpose of this paper is to examine key aspects of computer processing of multispectral scanner (MSS) data for meeting user needs, and to describe two recently developed analysis techniques that appear especially promising for assessing agricultural, forest, and rangeland resources.

II. Basic Concepts in Computer Analysis of MSS Data

During the past few years, many different computer processing and analysis techniques have been developed and tested for a variety of disciplines and conditions. Although specific techniques vary considerably, nearly all quantitative analysis of MSS data involves preprocessing of the data, followed by some type of enhancement and/or classification.

Preprocessing digital MSS data generally does not involve analysis of the data *per se*, but may simply involve changing the format of the tape to make it compatible with a particular computer system. Preprocessing could also involve complex corrections in the geometric and/or radiometric characteristics of the raw data in preparation for subsequent analysis. Geometric corrections of various types are often applied in order to provide the user with data having accurate and reliable cartographic characteristics. The amplitude of the signal may be adjusted in order to correct for variations in (a) solar angle due to differences in season, time of day, or topographic effects; (b) atmospheric attenuation; or (c) other factors causing unwanted variations in signal strength.

Once the data has been preprocessed, enhancement or classification techniques may be applied. Enhancement techniques can range from relatively simple to rather complex procedures for computer processing of the data, but in every case the final product is an image which is then given to the analyst or the resource manager to be interpreted in order to obtain the desired information. By using the appropriate procedure to process the scanner data, various vegetative, water, cultural, or other features can be effectively enhanced, thereby making the interpretation phase much faster, more accurate, and more reliable. Kourtz and Scott (1978) found that computer enhancement of Landsat data was the best method for obtaining forest-fire fuel maps for use by field personnel.

Classification techniques differ from enhancement techniques in that the data is actually categorized by the computer into some spectral or informational class of interest to the user. Such a process results in either maps or tables (or both) of the particular features or cover types of interest.

III. Computer Classification of Multispectral Scanner Data

There are two particularly important steps in computer classification of MSS data. The first involves the definition of the training statistics, and the second involves the actual classification of the data into the cover types or features of interest.

A. Developing Training Statistics

The process of developing training statistics is more of an art than a science in many respects, involving a man/machine interaction whereby the man "trains" the computer to recognize specific combinations of numbers consisting of radiance measurements in each of several wavelength bands. In theory, these combinations of spectral radiance measurements will characterize the various cover types of interest. The process of training the computer to identify specific cover types based upon their spectral characteristics generally involves fairly limited areas in the data for which accurate information exists concerning the type and condition of the ground cover. After a representative set of training statistics has been developed, the computer is programmed to classify the radiance values of each resolution element (or a statistical sample thereof) in the entire data set. In this way, the speed of the computer is used to advantage: a large geographic area can be mapped and acreage tabulations obtained at a much faster rate than would be possible using standard image interpretation techniques.

The development of training statistics and the actual classification of the data are often considered to be a single task. Much attention has been given to the various algorithms that can be utilized to classify the data, but relatively little emphasis has been given to the procedures used to train the classifier. It is our belief that the process involved in developing the training statistics is a very critical phase in the effective use of computers and satellite MSS data for mapping agricultural, forest, or rangeland cover types.

The two techniques for developing training statistics referred to most frequently in the literature are the so-called "supervised" and the "unsupervised" or "clustering" techniques (Swain and Davis, 1978). The supervised technique has proven satisfactory for identifying and mapping agricultural cover types (Bauer, 1975), but often does not provide an acceptable level of accuracy or reliability in wildland areas (Mead and Meyer, 1977). The primary reason for this is the difficulty the analyst has in defining locations in the data that represent all significant variations in spectral response for each cover type of interest. The unsupervised technique effectively overcomes this limitation of the supervised approach, but when working with large geographic areas, the amount of computer time involved in the iterative clustering sequence makes this technique prohibitively expensive.

Because of the difficulties experienced in using either the supervised or unsupervised technique for developing training statistics on forest and rangeland areas, several hybrid approaches have been developed which incorporate various aspects of both the supervised and the unsupervised techniques. One of these hybrid techniques, called the "Multi-Cluster Blocks" has been shown to be particularly effective in minimizing the amount of computer time as well as analyst time required, and at the same time enabling a higher classification accuracy to be obtained than can be achieved by any other technique.[1] This Multi-Cluster Blocks procedure for developing training statistics involves the selection of several relatively small but heterogeneous blocks of data, each of which contains several cover types. Each block is clustered individually, and then the spectral classes found in all cluster blocks are combined into a single set of spectral training statistics. Aerial photos of the cluster blocks are used to relate the spectral classes to the informational categories of interest. In essence, one could describe this Multi-Cluster Blocks technique as a logical method for using the computer to aid the analyst in discovering the natural spectral groupings present in the scanner data, and then correlating the resultant spectral classes with the desired informational classes (cover types, vegetative conditions, etc.). It is worth noting that typically less than one percent of the data to be classified is involved in this training phase of the analysis. We believe that the development of more effective training techniques such as this provides a significant improvement in the potential for effective utilization of these computer-aided analysis techniques and satellite data source.

B. Methods for Classifying MSS Data

1. General

After the classifier training process, the measurement values of each resolution element sensed by the scanner system are assigned, or classified, into one of the spectral classes defined by the training statistics. There is a wide range of classification algorithms available for this classification step, including such algorithms as the Nearest Neighbor, Minimum Distance to the Means, Maximum Likelihood, Parallel-piped, etc. It is beyond the scope of this paper to describe each of these algorithms, which vary considerably in computational complexity. The Maximum Likelihood algorithm has been the most frequently used, since it provides a sound theoretical basis and is one of the most generally applicable and also one of the most powerful algorithms available for this type of data analysis.

A key point concerning these algorithms is the fact that they all involve a procedure which entails an independent classification of each resolution element or pixel of the scanner data. This effectively limits the classification process to use of only the spectral data, or what a photo interpreter would call color or tone. Because texture is such an important attribute in differentiating and identifying many forest and rangeland cover types as well as many agricultural crop species, a classification algorithm that would enable the computer to utilize texture in addition to the spectral information would have considerable potential from an applications standpoint.

---

[1] In one test, the Multi-Cluster Blocks technique for developing training statistics resulted in an overall classification performance that was 14% higher than was obtained using the more common supervised technique (Fleming and Hoffer, 1977).

2. The ECHO Classifier

One classification algorithm that utilizes both the spectral and spatial (i.e., textural) characteristics of the data in the decision process is referred to as the "ECHO" (Extraction and Classification of Homogeneous Objects) classifier (Kettig and Landgrebe, 1976). In this technique, the computer is programmed to define the boundary around an area having generally similar spectral and textural characteristics, and then the entire area within the boundary is classified as a unit into one of the cover types or classes for which training statistics have been developed. Use of both the spectral and spatial characteristics in the data results in map outputs which have an appearance more like standard forest type maps, whereas classification algorithms that are based only on spectral data and involve a separate classification decision for each individual resolution element (i.e. "Per Point" classifiers) often result in output maps having a "salt and pepper" appearance. In a recent study, both a "Per Point" Maximum Likelihood classifier and the ECHO classifier were used to classify a wildland area containing coniferous and deciduous forest, rangeland, rock and soil, and water. The results of the classifications are shown in Figure 1 for a portion of the area involved. As can be seen, the ECHO algorithm produces a much different appearance than that obtained when using the standard "Per Point" classifer.

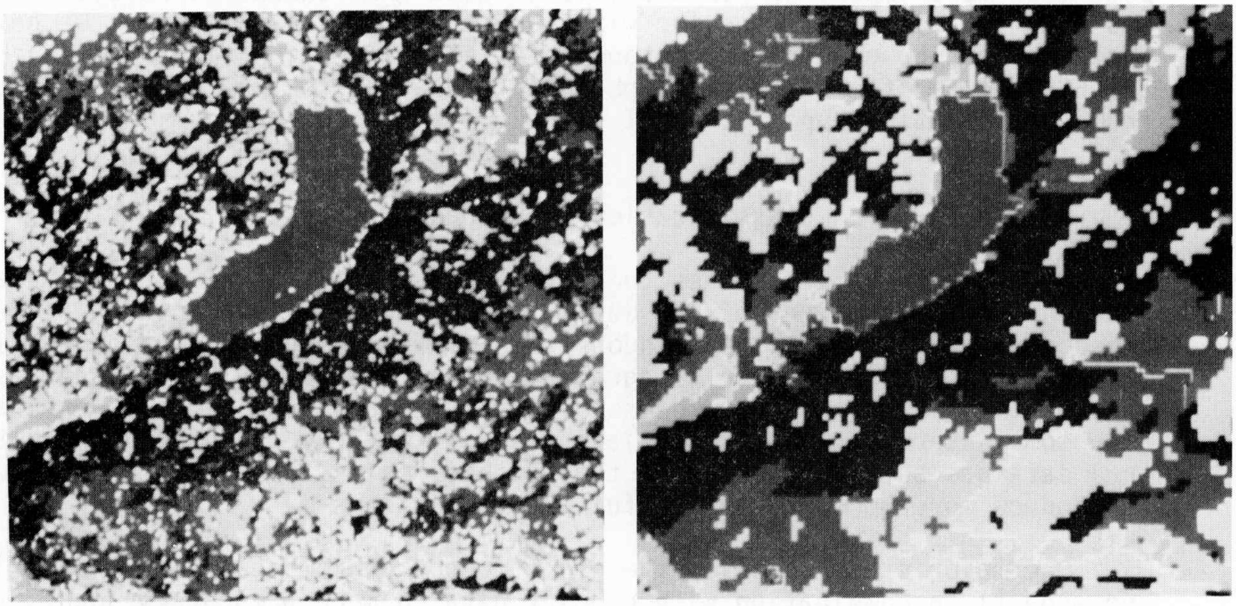


Figure 1. Classification maps obtained from the "Per Point" maximum likelihood classfier (left) and the ECHO classifier (right). A reservoir is seen near the center of the area.

In an attempt to evaluate the relative efficiency of the ECHO classifier, both analyst time and computer time were tabulated for a classification of Landsata data for a 15,300 hectare test site in an area of forest and rangeland cover types. The training statistics for both classifications were developed using the same technique (Multi-Cluster Blocks), so the only difference in the results was due to the classification algorithm used. Based upon a cost of $250/hour for CPU computer time and $10/hour for the analyst, the results indicated a total cost of $2.54 per 1,000 hectares to classify the data when using the Maximum Likelihood Per Point algorithm, but use of the ECHO algorithm resulted in a total cost of only $1.77 per 1,000 hectares. Thus, the ECHO classifier cost only 70% as much as the Per Point classifier, primarily due to the difference in the amount of computer time required (Hoffer, 1979). In another study, three agricultural test sites were classified and the results showed that the ECHO classifier required only 60% as much computer time as was needed for the Per Point classifier (Scholz et al, 1979).

In both of the above studies, the accuracies of the classifications obtained from the two classification algorithms were not significantly different. However, it is anticipated that use of scanner data having better spatial resolution than is presently obtained from Landsat will provide significant advantages for the ECHO algorithm over "Per Point" algorithms in terms of both efficiency and classification accuracy. Since the spatial resolution of Landsat data is relatively coarse (80 meters), the textural component in the data tends to be severely diminished or even lost. In the near future data will become available from satellite scanner systems having much better spatial resolution, such as the Thematic Mapper (30 meter resolution) to be launched by the United States in 1982, or SPOT (20 meter resolution) to be launched by France in 1984. The improved spatial resolution will result in an increased texture in the data, particularly for areas of forest cover. Therefore, the capability to effectively utilize the spatial (i.e, textural) as well as the spectral characteristics of the data will become much more important for analysis of data from these future systems than has been the situation thus far with Landsat or Skylab MSS data. We believe that the development and understanding of classification algorithms such as the ECHO classifier is a vital phase in preparing for the time when data from the Thematic Mapper and SPOT scanner systems becomes available.

3. The Layered classification technique

In addition to techniques to utilize the spatial as well as the spectral information content of MSS data, a second development in data processing capabilities that has tremendous potential involves the use of ancillary data (such as soils data or topographic data) along with the MSS data as part of the classification process. During the past few years, there have been some interesting and potentially significant experiments in which spectral data has been combined with topographic, soils, or land ownership data to produce more accurate and useful classifications.

One major study involved the use of topographic data (elevation, slope, and aspect) in combination with Landsat data to more accurately identify individual forest cover types (Hoffer et al, 1979). The test site involved a mountainous area in Colorado containing a complex mixture of forest types, rangeland, alpine tundra, water bodies, geological features and various man-made features. Previous studies had shown that deciduous and coniferous forest cover types could be identified and mapped with a reasonable degree of accuracy (i.e., 70-85%) using Landsat data, but that individual species of trees could not be effectively differentiated and identified using Landsat spectral data alone. However, as illustrated in Figure 2, in the San Juan Mountains in S.W. Colorado there is a rather well-defined distribution of cover types according to elevation. It was hypothesized, therefore, that if the areas of coniferous and deciduous forest cover could be identified and mapped using just the spectral information in Landsat data, topographic data could then be used to classify the data into individual forest species groups or cover types. Elevation data was available for the test site in a computer compatible format. A nearest neighbor algorithm was used to digitally overlay, pixel by pixel, the elevation data onto the Landsat data. Slope and aspect channels were generated from the elevation data, using an interpolation process, and added to the Landsat spectral-plus- elevation data set. This provided a data base in which every 80-meter X-Y coordinate contained four channels of Landsat spectral reflectance data, a channel of elevation data, a channel of slope data, and a channel of aspect data.

Analysis of this data set indicated that the most effective method for combining Landsat and topographic data was a "Layered" classification technique. This Layered Classifier allows the analyst to classify the data in a hierarchical sequence of steps or layers (Swain and Hauska, 1977). Figure 3 illustrates the concept of the Layered Classifier. Since different types of data could be used at the different layers or steps in the classification process, the Layered Classifier was particularly useful in this study in that Landsat satellite data could be used initially to identify coniferous and deciduous forest cover (as well as other major cover types), and then in the next layer, topographic data corresponding to the same X-Y coordinate on the ground could be used to identify the particular forest cover type. To accomplish this, two different sets of training statistics were required--one for the Landsat data and one for the topographic data. The Multi-Cluster Blocks technique was used to obtain the spectral training statistics, and a stratified random sample procedure was used to develop a "Topographic Distribution Model" which quantitatively described the distribution of the various cover types of interest as a function of elevation, slope, and aspect. The Layered Classification Tree shown in Figure 4 was then used to classify the combined spectral/topographic data. Only the Landsat spectral data was used in the first layer of the classification to identify the major cover types in the area (coniferous forest, deciduous forest, herbaceous, barren and water). In this study, fifteen spectral classes were used to identify these five major cover types. In the second

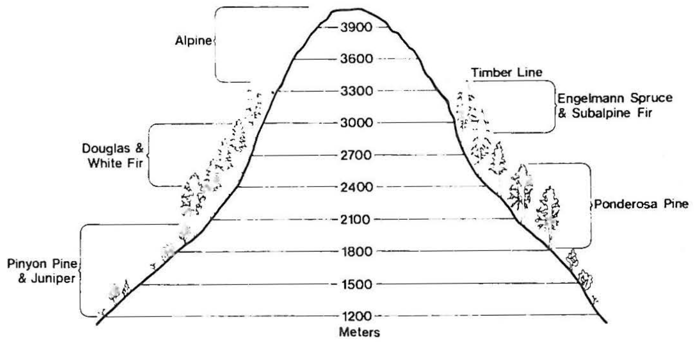


Figure 2. Relationship between elevation and distribution of vegetative cover types in the San Juan Mountains of Colorado.

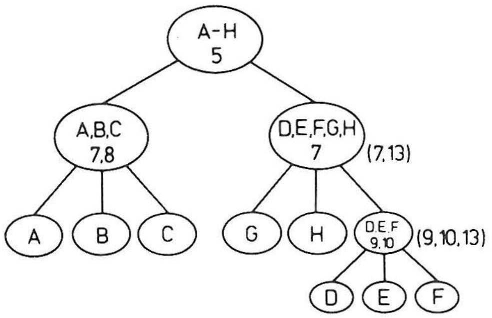


Figure 3. The Layered classification technique concept. Letters indicate the classes of cover type and numbers indicate wavelength bands used in the classification.

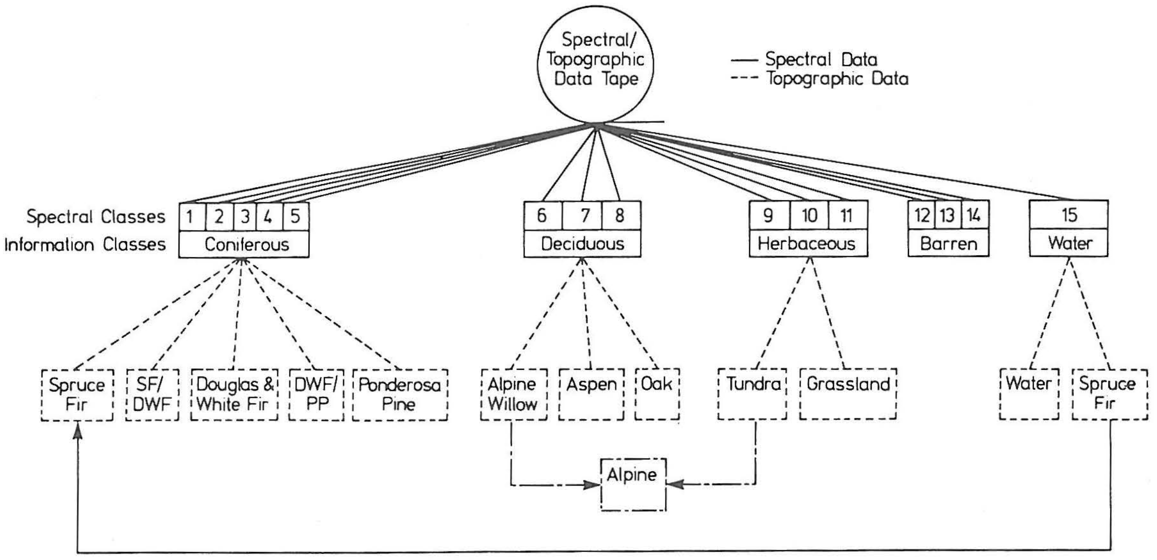


Figure. 4. Tree structure used with the Layered Classifier to combine spectral and topographic data.

layer, only the topographic data was used to classify the data into the individual cover type classes of interest. As indicated in Figure 4, after the classification was completed the alpine willow and tundra classes were combined into a single informational category called "alpine". It was also discovered during the analysis that some areas were being misclassified as water when only the Landsat data was used, because they had very low spectral response values, particularly in the reflective infrared wavelengths. In reality, these areas were generally spruce-fir forest cover in regions of topographic shadow. The topographic data was therefore used in the second layer of the classification process to re-examine each pixel that had initially been classified as water, and only those pixels having a zero slope were again classified as water. If the pixel had a slope of other than zero, it was put into the spruce-fir forest cover type.

The results of this study indicated that by using the layered classification technique and the combined spectral plus topographic data, classification accuracies were improved by 15% over those obtained using Landsat spectral data alone (Hoffer et al, 1979)! An additional test of this approach for combining spectral and topographic data is currently underway for a study area in the mountains of Washington state. Preliminary results again indicate a significant improvement in classification peformance over that obtained when using Landsat data alone.

Another study involving use of the Layered Classifier was directed at identifying wetland cover types in areas of freshwater wetlands. In this study, Landsat data and soils data were combined to create a digital data base. The soils data was then used in the first layer of the classification sequence to separate upland from bottomland soil types. The results indicated that use of the Layered Classifier and the combined spectral plus soils data enabled wetland areas to be differentiated from non-wetlands with an overall accuracy of 84%, whereas use of the spectral data alone resulted in only a 72% classification performance (Ernst and Hoffer, 1979). Hardwood swamp and upland hardwood forest areas could not be separated by using the

Landsat spectral data alone, but could be differentiated with the combined spectral plus soils data.

IV. Summary and Conclusions

The past decade has brought many significant and far-reaching changes in our ability to obtain earth resource data on a world-wide basis. Each year, there is a growing awareness of the value of satellite data and the potential applications of such data to meet a wide variety of information needs. There is also an increased understanding of both the capabilities and the limitations of Landsat MSS data and existing methods for analyzing such data. It is for these reasons that we believe that the 1980's will see major advances in our capabilities to monitor the earth's resources using satellites and computer processing techniques.

Research is needed in many areas, including the continued development of effective methods for handling and analyzing the increased amount of data that will be generated by the higher spatial resolution and increased number of spectral bands of the Thematic Mapper on Landsat D. More work is also needed on additional techniques for analyzing combined satellite and ancillary data. Many other research needs could be cited as well.

There are at least four key points which we believe will have major impacts on remote sensing in the coming decade. First, it seems clear that the improved spectral and spatial resolution that will be obtained by the Thematic Mapper and SPOT systems will enable significant improvements to be realized in mapping and assessing earth resources. Second, the capability to combine satellite spectral data with other types of data in digital data bases will enable significant advances to be achieved in our ability to use and apply remote sensing to a wide variety of user needs. Third, the recent designation of the Landsat Earth Resource Satellite as an operational program (as opposed to the previous research designation) will enable many potential users to proceed with plans to use such data, confident that data will continue to be obtained and made available. Fourth, the continued development in computer systems and their data storage and handling capabilities will play a major role in the future of remote sensing technology and the application of such technology to meet various information needs.

We believe that the coming decade will see remote sensing data collection systems and computer processing of this data play a major role in assessing agricultural, forest, rangeland and other earth resources in many countries throughout the world. The potentials are tremendous! We must meet the challenge of providing the users with accurate, reliable information in a timely manner!

References

Bauer, M. E. 1975. "The role of remote sensing in determining the distribution and yield of crops." Advances in Agronomy 27:271-304.

Ernst, C. L., and R. M. Hoffer. 1979. "Using Landsat MSS data with soils information to identify wetland habitats." Proceedings of the International Symposium on Satellite Hydrology, Sioux Falls, South Dakota. 18 pp.

Fleming, M. D., and R. M. Hoffer. 1977. "Computer-aided analysis techniques for an operational system to map forest land utilizing Landsat MSS data." *LARS Technical Report* 112277, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. 236 pp.

Hoffer, R. M., M. D. Fleming, L. A. Bartolucci, S. M. Davis and R. F. Nelson. 1979. "Digital processing of Landsat MSS and topographic data to improve capabilities for computerized mapping of forest cover types." *LARS Technical Report* 011579, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. 159 pp.

Hoffer, R. M. 1979. "Computer-aided analysis of remote sensor data -- magic, mystery, or myth?" Invited paper. *Proceedings of the International Conference on Remote Sensing in Natural Resources*, University of Idaho, Moscow, Idaho. 18 pp.

Kettig, R. L. and D. A. Landgrebe. 1976. "Classification of multispectral image data by extraction and classification of homogeneous objects." *IEEE Trans. Geos. Elect.* 14:19-25.

Kourtz, P. H. and A. J. Scott. 1978. "An improved image enhancement technique and its application to forest fire management." *Proceedings of the Fifth Canadian Symposium for Remote Sensing*, Victoria, B.C. 9 pp.

Mead, R., and M. Meyer. 1977. "Landsat digital data application to forest vegetation and land-use classification in Minnesota." *IAFHE RSL Research Report* 77-6, University of Minnesota, St. Paul, Minnesota. 57 pp.

Scholz, D., N. Fuhs, and M. Hixson. 1979. "An evaluation of several different classification schemes: their parameters and performance." *Proceedings of the Thirteenth International Symposium on Remote Sensing of the Environment*, Ann Arbor, Michigan. 7 pp.

Swain, P. H. and H. Hauska. 1977. "The decision tree classifier: design and potential." *IEEE Trans. Geos. Elect.* GE 15(3):142-147.

Swain, P. H. and S. M. Davis (eds.). 1978. *Remote Sensing: The Quantitative Approach*. McGraw Hill, New York, N.Y. 396 pp.