7.8 10084
CR-155762

INVESTIGATION OF ATMOSPHERE ON MULTISPECTRAL MAPPING OF ROCK TYPE BY COMPUTER, CRIPPLE CREEK-CANON CITY

(E78-10084) THE MIXTURE PROBLEM IN COMPUTER MAPPING OF TERRAIN: IMPROVED TECHNIQUES FOR ESTABLISHING SPECTRAL SIGNATURES, ATMOSPHERIC PATH RADIANCE, AND TRANSMITTANCE Final Report, 1 Jul. 1972 - 10 (Geological

N78-19563
HC A03/MF A01
Unclas
G3/43 00084

Harry W. Smedes
U.S. Geological Survey
Denver, Colorado 80225

Original photography may be purchased from
EROS Data Center
Sioux Falls, SD

March 15, 1978

Surrogate Type III Final Report for ERTS-1 Experiment SR-194
for
Period July 1, 1972 - May 10, 1974

ORIGINAL CONTAINS
COLOR ILLUSTRATIONS

Prepared for:
Goddard Space Flight Center
Greenbelt, Maryland 20771

Publication authorized by the Director, U.S. Geological Survey

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle SR 194
The Mixture Problem in Computer Mapping of Terrain: Improved Techniques for Establishing Spectral Signatures, Atmospheric Path Radiance, and Transmittance

5. Report Date
15 March 1978

6. Performing Organization Code

7. Author(s) Harry W. Smedes, Roland L. Hulstrom, and K. Jon Ranson

8. Performing Organization Report No.

9. Performing Organization Name and Address
U.S. Geological Survey
Federal Center
Denver, CO 80225

10. Work Unit No.
Task No. 31

11. Contract or Grant No.
S-70243-AG

13. Type of Report and Period Covered
Type III Progress Report
Final Report
1 Jul 72 - 10 May 74

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, MD

14. Sponsoring Agency Code

15. Supplementary Notes
Preprint from the Proceedings of the NASA Earth Resources Survey Symposium, Houston, Texas, June 8-13, 1975, v. 1-B, p. 1099-1159.

16. Abstract

Among the results of our LANDSAT and Skylab research programs on the effects of the atmosphere on computer mapping of terrain, we emphasize the following: (1) the concept of a ground "truth" map needs to be drastically revised; (2) the concept of training areas and test areas is not as simple as generally thought because of the problem of pixels that represent a mixture of terrain classes; (3) this mixture problem needs to be more widely recognized and dealt with by techniques of calculating spectral signatures of mixed classes, such as those we used, or by other methods; (4) atmospheric effects should be considered in computer mapping of terrain and in monitoring changes; and (5) terrain features may be used as calibration panels on the ground, from which atmospheric conditions can be determined and monitored.

17. Key Words Suggested by Author
LANDSAT
Skylab
Remote Sensing
Computer map
Wilderness map
Terrain map
Ground-cover map
Colorado
Ground truth

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	61	

Figure 2A. Technical Report Standard Title Page. This page provides the data elements required by DoD Form DD-1473, HEW Form OE-6000 (ERIC), and similar forms.

THE MIXTURE PROBLEM IN COMPUTER MAPPING OF TERRAIN: IMPROVED TECHNIQUES FOR ESTABLISHING SPECTRAL SIGNATURES, ATMOSPHERIC PATH RADIANCE, AND TRANSMITTANCE

By Harry W. Smedes, U.S. Geological Survey, Denver, Colorado,
Roland L. Hulstrom, Martin Marietta Aerospace Corporation, Denver, Colorado, and
K. Jon Ranson, Colorado State University, Fort Collins, Colorado

ABSTRACT

Among the results of our LANDSAT and Skylab research programs on the effects of the atmosphere on computer mapping of terrain, we emphasize the following: (1) The concept of a ground "truth" map needs to be drastically revised; (2) the concept of training areas and test areas is not as simple as generally thought because of the problem of pixels that represent a mixture of terrain classes; (3) this mixture problem needs to be more widely recognized and dealt with by techniques of calculating spectral signatures of mixed classes, such as those we used, or by other methods; (4) atmospheric effects should be considered in computer mapping of terrain and in monitoring changes; and (5) terrain features may be used as calibration panels on the ground, from which atmospheric conditions can be determined and monitored.

The natural terrain comprises a mixture of diverse classes, including combinations of such things as tree, grass, and other vegetative-cover species, different types of soil and rock, and water bodies of different size, depth, and clarity. Because the terrain features generally are small compared to the ground resolution element (= pixel) from multispectral scanner data, it is unusual for very many contiguous pixels to consist of a single terrain class. Commonly, each pixel is a mixture of two or more classes.

It has been shown that the spectral signature (radiance) of a mixture of terrain classes is not representative of any of the component classes composing the mixture. Automatic recognition processors may therefore misclassify mixture-containing pixels, which often leads to an underestimation of the amount of component classes present in the area scanned.

Even a single class may have a range in spectral signature depending on such things as size of pixel (sample size), solar aspect, orientation, and atmospheric conditions.

Because of this problem of mixtures in pixels, we at first attempted to compile ground control maps whose classes were designated by different increments of the more widely occurring mixtures and to train the computer to recognize them by selecting specific TRAINING AREAS of known proportions of classes. Not only was this a monumental task, loaded with subjective judgements and difficult to calibrate, but the TESTING of accuracy of the resulting computer-derived map proved to be equally monumental. After much effort we concluded that--if not impossible--it was certainly not at all feasible to prepare a truly accurate ground control map. In fact, while checking for errors in the computer map, the ground control or so-called ground "truth" map continually had to be upgraded. A serious problem that needs to be faced is: "what constitutes the ground truth?" Our first major conclusion was that, if properly trained on end-member and mixed classes, the computer made a more accurate map of terrain cover than we were able to compile as control data.

However, it is very difficult to measure precise proportions of mixtures and to locate those areas in terms of LANDSAT MSS, Skylab S-192, or other scanner pixels, especially in areas of mountainous terrain. This difficulty and the resulting uncertainty and imprecision led us to experiment with techniques to calculate mixes from data for end members and to use those calculated signatures in lieu of training areas for the mix classes. Previous research in this problem has largely been confined to classes of crops in flat agricultural fields. Our data and techniques will apply equally well to conditions ranging from croplands to wilderness.

Inasmuch as the radiance from a pixel is integrated over the entire area of the pixel (the instantaneous field of view of the scanner), mixture-class signatures can be determined from the known mean vectors and covariance matrices of the component classes and the proportional areas occupied by each component in the pixel. Signatures for mixed classes were calculated using spectral radiance from on-site measurements. In addition, several experiments were conducted with simulated LANDSAT MSS-type data to show the expected improvement in accuracy of automatic classification using simulated mixture-class signatures. We also studied methods for determining component class signatures when there were insufficient data points for conventional signature extraction.

Results are presented of a test area in mountainous terrain of south-central Colorado for which an initial classification was made using simulated mixture-class spectral signatures and actual LANDSAT-1 MSS data.

Not only are these techniques highly successful in terms of shorter preparation time and higher accuracy, but the calculated signatures also afford insights into what the appropriate increments of mixes and what the optimum wavelength bands are for the most accurate discrimination among specific classes.

Atmospheric effects must be known in order to combine spectral signatures derived directly from LANDSAT or Skylab scanner tapedata with those measured on-site and from aircraft. A few large homogeneous sites such as a large body of deep clear water, a large expanse of bare rock, dense forest, large nonvegetated dry lake bed or desert sand, and snow may serve as known natural calibration panels on the ground. These would be visible from spacecraft, from which the atmospheric path radiance and transmittance can be derived for each data set (mission or flight). This calculated path radiance and transmittance can then be applied to correct the tape data for true radiance values of the terrain classes. Used in another sense, this technique can serve as a useful means of monitoring atmospheric quality from spacecraft or aircraft, as atmospheric path radiance and transmittance are measures of air quality.

INTRODUCTION

Multispectral scanner data from satellites are used as input to computers for automatically mapping terrain classes of ground cover. Some major problems faced in this remote-sensing task include 1) the effect of mixtures of classes and, primarily because of mixtures, the problem of what constitutes ground "truth"; and 2) effects of the atmosphere on spectral signatures. This paper presents the fundamental principles of these problems and some of the results of our studies of them for a test site in Colorado, using LANDSAT-1 data.

The test site (figure 1) comprises about 2280 sq. km (880 sq. mi) of generally wildland terrain in south-central Colorado. It includes such landmarks as Pikes Peak, the Cripple Creek mining district, and Canon City. Altitudes range from 1525 to 4300 m. The terrain is highly varied and includes a diversity of rock types, soil, and vegetative cover, over a wide range of angle and aspect of slope.

The research was carried out as an integral part of LANDSAT and Skylab EREP projects funded by NASA, and by in-house research supported by the Martin Marietta Aerospace Corporation. Computer support was provided by Colorado State University and a computer-derived map was made by the Environmental Research Institute of Michigan (ERIM) as their part of a separate but coordinated LANDSAT project funded by NASA. This map provided further insights into the nature of the mixtures problem. This research effort was conceived, initiated and coordinated by Smedes. Hulstrom measured atmospheric properties and studied their effects, made on-site measurements of spectral signatures, and used these data in a computer program to calculate signatures of mixtures. Ranson studied the effects of mixtures using simulated LANDSAT data as part of a dissertation for an advanced degree at Colorado State University.

THE MIXTURE PROBLEM

The natural terrain is composed of mixtures of classes. This is true regardless of the size of the ground resolution element (= pixel) of the sensor system, and holds for satellite data with resolution on the order of 100 sq. meters, to that of the microscope with resolution of the order of a sq. micron or less. By tradition, each discipline of the natural sciences has accommodated this mixtures problem by means of graded orders or hierarchies of classifications. We thus have, in order of decreasing size of resolution element, such classes as galaxies; stars and planets; continents and oceans; mountains, ridges, valleys, plains and deserts; woodlands, meadows, bare rock, and soil; spruce, aspen, bunchgrass, granite, and quartz veins; bark, needles, leaves stems, quartz, feldspar, and mica grains; petioles, chloroplasts, mitochondria; perthite, albite with fluid inclusions , and magnetite inclusions in biotite. Because we are dealing with current satellite scanners, the resolution element or pixel size we are concerned with is 65-80 m square. This results in high likelihood that any given pixel will contain two or more of such classes as woodlands, meadow, bare rock, and soil. If a recognition processor (computer) is trained to recognize only the homogeneous component classes (trees, grass, rock, etc.) the overall consequences of a mixture of these classes occurring within a pixel is a misclassified or unclassified pixel, which tends to make estimates of the area covered by a terrain class lower than actually present (ref. 1). Errors in classifications due to the mixture problem alone will be 25 to 30 percent (ref. 2 and 3).

There are two basic kinds of mixtures. One is due to the boundary or edge effect, where a pixel encompasses the interface of two or more homogeneous end-member component classes, such as water and dense coniferous forest. The resulting spectral signature would not be representative of a pixel that contained only one of these classes. The other mixture is represented by pixels which contain homogeneous mixtures of such things as forest canopy and grass, rock, or soil understory. In the first case, only the pixels along the interface will be misclassified. In the second case, large clusters of pixels throughout the forest will be misclassified if only the component classes were used to train the computer (ref. 4, 5, 6).

Variations in the amount of vegetative cover will result in corresponding mixtures of the vegetation and underlying material. Consequently, if the computer is trained to recognize only one vegetation density, then misclassifications would occur for densities above or below that of the training class, within the constraint of some response threshold. These misclassifications occur because when a mixture of terrain classes is contained within the instantaneous field of view (IFOV) of a scanner, the spectral response obtained is unlike that of any of the component class signatures (ref. 8). To illustrate this concept, three terrain classes were identified on graymaps of August 20, 1972 (frame no. 1028-17135) LANDSAT-1 MSS data that appeared representative of grassland (A), dense forest assumed to be 100% cover (B) and a class representing an assumed 50-50 mixture (C) of classes A and B and located on the interface between them. Spectral signatures were extracted with 21 data points sampled for Class A, 72 for class B and 24 for class C. The mean spectral response curves for the three classes are shown in figure 2. Note that the spectral response for the mixture class C is uncharacteristic of either of its component classes (A and B) but that the response curves falls between the two components' curves. Researchers investigating this phenomenon have found that the relationship of a mixture and the signatures of the component classes may be approximated as a function of the area of the pixel occupied by each component terrain class and their respective spectral signatures.

Because of this problem of mixtures in pixels, we at first attempted to compile ground control maps whose classes were designated by different increments of the more widely occurring mixtures and to train the computer to recognize them by selecting specific training areas of known proportions of classes. This method, described in detail in ref. 6 and 7, is extremely time-consuming, loaded with subjective judgements, and very difficult to calibrate. Testing the accuracy of the resulting computer-derived map is even more time-consuming

Our conclusion, supported by other work (such as reported in ref. 6, 7, and 9), was that if the computer is properly trained on end-member components and on commonly-occurring mixtures, it could make a more accurate map of the terrain cover than we were able to compile as control data. This casts serious doubt upon and calls for reconsideration of what constitutes the ground "truth".

However, it is very difficult to measure precise proportions of mixtures and to locate those areas in terms of LANDSAT MSS, Skylab S-192, or other scanner pixels, especially in areas of mountainous terrain. This difficulty and the resulting uncertainty and imprecision led us to experiment with techniques to calculate mixes from data for end members and to use those calculated signatures in lieu of training areas for the mix classes. Previous research in this problem has largely been confined to classes of crops in flat agricultural fields. Our data and techniques will apply equally well to conditions ranging from croplands to wilderness.

Inasmuch as the radiance from a pixel is integrated over the entire area of the pixel (the instantaneous field of view of the scanner), mixture-class signatures can be determined from the known mean vectors and covariance matrices of the component classes and the proportional areas occupied by each component in the pixel.

Two approaches were taken. One was to measure the reflectivity of pure end-member components on site and then calculate signatures of various mixtures. The other was to use the satellite data itself to extract the radiance signatures of known end-member targets and then calculate radiance signatures for various mixtures. Simulated LANDSAT MSS-type data were used. They show the expected improvement in accuracy of automatic classification using simulated mixture-class signatures. In addition, methods have been developed for determining component class signatures from mixtures when there were insufficient data points for conventional signature extraction of the end-member components.

CALCULATED SPECTRAL SIGNATURES FOR MIXTURES

Basic Concept

The basic problem of correctly identifying ground resolution elements that contain mixtures of terrain classes with simulated mixture signatures is obtaining representative signatures for the component classes. A component class can be defined as a homogeneous or non-homogeneous group of materials that make up a discrete mapping class or terrain type. A discrete mapping class may or may not be found in combination with other discrete classes as mixtures. For example, in an area where there is a sparse covering of grass with bare soil showing through, the grass and soil may be considered as two discrete mapping classes. With the 0.4 ha resolution of LANDSAT, the scene would be viewed as a mixture of grass and bare soil so these two classes would most likely be combined into one component class for analysis. If another class were identified as dense forest, then any time enough trees existed on the grass-soil unit to affect the response of a pixel, then a mixture of these two classes (forest and the grass-soil association) would exist. Discrete mapping class determinations can usually be considered a function of the natural associations, the objectives of the user, spectral radiance, and the limitations of the MSS data.

When selecting component classes for mapping mixtures, care must be taken to avoid situations where the data for a class form a multimodal distribution. Multimodal distributions occur where the response data for a class are affected by variables such as slope and aspect, vegetation vigor, underlying soil spectra, and sensor scan angle effects. The standard method for dealing with this problem is to divide the multimodal class distribution into subclasses as a function of slope, etc., classify the data and then combine the results for each subclass (ref. 10, 11) for final display.

Calculations From On-site Measurements

<u>Mixtures</u>.- For on-site and near-surface measurements (negligible atmospheric transmission and path radiance) the radiance, N_t of a pure target end-member can be given as

$$N_t = \frac{H\rho}{\pi} \tag{1}$$

Where H is the total solar irradiance and ρ is the target reflectivity. The radiance from a mixture of targets, Nm, can be given as

$$N_m = \frac{H}{\pi}(\rho_1 A_1 + \rho_2 A_2 + \rho_3 A_3 + \cdots \rho_n A_n) \tag{2}$$

where ρ_1, ρ_2, etc. are the reflectivity signatures of each component target, and A_1, A_2, etc. are the fractional proportions of the pixel that each component target occupies. The assumption is made that none of the component targets have strong specular reflectivities.

Note that even for the near-surface conditions, there are atmospheric effects to be considered (H, in equations 1 and 2, above).

In order to simulate mixed target radiance signatures, a knowledge of the total incoming solar irradiance (H), the pure target reflectivity signatures, and the fractional area of each target need to be known. In this study the spectral distribution and magnitude of the incoming solar irradiance is measured for various sun angles and atmospheric conditions; and the pure target reflectivities are measured on site and from a helicopter. The fractional area occupied by each target is simply varied to simulate various combinations of mixtures.

The first mixture problem addressed was that of Pikes Peak Granite and Coniferous Forest. This represents a very simple and common mixture. The reflectivity for the Pikes Peak Granite and the total incident solar irradiance were measured on site (as discussed in the following sections) while the reflectivity signature for Coniferous Forest was taken from ref. 12. Figure 3 shows all three of the parameters. The resultant simulations of component and mixture signatures for the LANDSAT bands and for the continuous spectral region from 500 to 1100 nm are shown in figures 4 and 5. It is interesting to note that (as shown in fig. 4 and 5) for a 50-50 mixture of rock and trees, the resultant mixture radiance signature no longer contains the characteristic chloryphyll absorption feature. The resultant radiance signature would most likely be interpreted as a grayish, low-albedo rock.

Spectral contrast.- In addition to computer simulations of pure and mixed target radiance signatures, the spectral inherent contrast of various targets and their background can be calculated. The inherent contrast, C, of a target versus its background is

$$C = \frac{N_T - N_B}{N_B} \qquad (3)$$

where N_T is the radiance signature of the target and N_B is the radiance signature of the background. The target and/or background can be any pure target or mixture of targets.

Spectral contrast was calculated in order to determine if there existed any particular spectral region or band that gave a better contrast than other spectral regions or bands. The goal of these spectral contrast calculations is not to exclusively attempt to determine if "optimum" bands existed for the discrimination of rock types, vegetation types, etc. Rather, the goal is to determine "optimum" bands for maximum spectral contrast between a selected target(s) and its background; because, for the natural terrain existing in the Colorado test site, this goal is in line with the actual terrain conditions. For example, two rock types rarely exist side-by-side with good exposure; instead, a given rock type is exposed and surrounded by a background made up of a complex mixture of other terrain. If the rock types do exist side-by-side they usually do not fill the entire field of view of the sensor. This is due to the fact that they occur as small outcrops and/or they exist in layers exposed on extremely steep slopes or cliffs.

An example of the computer spectral contrast calculations is shown in figure 6. The pure target is Pikes Peak Granite, and the background consists of mixtures of granite and coniferous forest. It is obvious that there are "optimum" spectral regions; and, if the background comprises a mixture, the contrast is reduced tremendously. The same case, only applied to the broader bands of LANDSAT, is shown in figure 7. As can be seen, "optimum" bands still exist; however, the overall contrast is reduced because of the broader bands.

The computer calculations of spectral contrast are not limited to the LANDSAT spectral resolution, but can be used to determine spectral contrast for any selected wavelength resolution or band width. The resolution is only limited by the resolution of the instrument used to measure solar irradiance and reflectivity. Such computer simulations of radiance signatures and spectral contrast would be a very useful tool for planning aircraft multispectral remote sensing missions. A combination of a few selected measurements of solar irradiance and reflectivities, and the subsequent computer calculations of spectral contrast could be used to select film/filter combinations that would yield the highest contrast. This method would eliminate costly flights and subsequent data reduction performed in hopes of empirically determining optimum bands.

Calculations From LANDSAT Data

Modeling of spectral signatures of mixture.- Two approaches for modeling were used in this study. The first is a widely reported method that describes a mixture class signature in terms of weighted combinations of component class mean vectors and covariance matrices (ref. 1, 2, 8, and 13). The model describes the mean spectral response from a ground resolution element as:

$$M_{P_i} = \sum_{i=1}^{N} P_i M_i \tag{4}$$

where M_{P_i} = Mean response vector for a mixture of N component classes;

P = Relative amount (proportion) of class i;

M_i = Mean response vector for the ith component class.

The above relationship assumes statistical independence of normally distributed data points belonging to class i.

Assuming statistical independence for variables associated with elements from different object classes, the relationship for a mixture class covariance matrix can be written as:

$$C_P = \sum_{i}^{N} P_i C_i \tag{5}$$

where C_P = Covariance matrix for a mixture of N component classes;

P = Relative amount (proportion) of class i;

C_i = Covariance matrix describing the distribution of the ith component class.

Equations (4) and (5) represent the model used for automatic classification of mixtures of classes with a maximum likelihood processing algorithm. For supervised learning recognition processors, component class mean vectors and covariance matrices must be determined and proportions for each possible mixture must be specified. The approach presented here is based on modeling the spectral response within a single pixel and should be applicable to recognition processing with the RECOG (ref. 10, 14, 15) maximum likelihood algorithm (GLIKE) since classification is point by point.

Most of the work reported in the literature utilizing this modeling technique required sophisticated algorithms that calculate various mixture class signatures and select the one that gives the closest approximation of the response from a given pixel. In our study, however, we used the model to obtain a set of spectral signatures for specified mixtures of component terrain classes that are used by the RECOG pattern recognition routines to identify all MSS data points that exhibit that response. The value of this method lies in its straightforward applicability to the existing RECOG processing sequence, thus eliminating the need to develop new processing algorithms.

Equation 4 was used to simulate the spectral response expected from a mixture of the two component classes (A and B) whose extracted mean vectors were shown in Figure 2. Figure 8 shows the resulting simulated mean vector (D) for a 50% mixture of the two classes ie.

= P_B = 0.5. A likely explanation for the difference between curves C and D is the diffi-lty of accurately estimating cover densities from high altitude aerial photographs used ground control (in this case Mission 205, NASA RB-57 photographs with an approximate ale of 1:100,000) and the difficulty of locating non-homogeneous (mixture) training sets graymaps (ref. 16).

When a set of representative component terrain class signatures are obtained then the ignature of any mixture of these component classes may be determined using Equations 4 and . A computer program MIX was written to take two component class signatures and two propor-ion vectors and calculate the spectral response of any combination. The method involves caling the mean vector and covariance matrix of a component class by a proportion factor nd adding the result to the scaled mean vector and covariance matrix of another component lass. The proportion factors p_j must be within the set $0 \leq P \leq 1$ and $\sum P_j = 1$. The re-ulting spectral signature may then be used to classify all pixels in a set of LANDSAT MSS ata that exhibit a similar spectral response.

The second modeling technique used linear regression to predict the spectral response f a mixture-containing pixel based on the mean response vectors for known mixtures of errain classes. The regression model for a two-component mixture takes the form:

$$Y_i = C + B(X) \tag{6}$$

where Y = estimated mixture response in wavelength band i;

C = constant (Y intercept);

B = coefficient (slope of regression line);

X = proportion of one component of the mixture in the pixel scene.

Mead (ref. 5) indicated that this method may be useful in estimating the spectral response for varying densities of ponderosa pine, but also noted that the mean spectral response may be affected by the arrangement or distribution of ponderosa pine within a training set as described by the standard deviation.

The methods presented below, with the exception of signature extraction, were developed for mixtures of two component classes. The following is a synopsis of the techniques used to obtain component- and mixture-class spectral signatures for automatic analysis of MSS data:

METHOD	TECHNIQUE	COMPUTER PROGRAM
Signature Extraction	Statistical sampling of MSS data points	RECOG (Phase 2, Appendix A, B) (see ref. 14 & 15)
Component Signature Estimation	Solving simultaneous equations	SIGCALC
	Linear regression	STAT38R
Mixture Signature Simulation	Addition of weighted mean vectors and covariance matrices	MIX

Signature Extraction.- When areas on the ground have been satisfactorily identified as containing a known terrain class or mixture of terrain classes and the graymap cooridinates have been determined, then spectral signatures may be obtained by statistically sampling these points. This process is known as signature extraction and represents the conventional mode for obtaining signatures that are assumed to characterize a class of objects. The designated data points known to contain a terrain class are sampled and the overall mean is determined for each channel of MSS data. These means, four in the case of LANDSAT-1, are collectively known as the mean vector. Because of the inherent variability found in natural objects an additional set of statistics is obtained that characterizes the population distribution of the sampled data points. This set of statistics is the class covariance matrix and describes the within-channel and between-channels variation of the data. The technique used in this work is contained in the RECOG routines, specifically Phase 2 (Appendix A). The formulas used in these routines are described in Appendix B.

Estimating component-class spectral signatures.- One method for estimating component class signatures from known mixture class signatures involves solving two equations with two unknowns. The spectral signatures for unknown mixture classes are assumed to be related and can be described in equation form in the following manner:

$$P_a(MSR_a)_i + P_b(MSR_b)_i = Mp_i \tag{7}$$

where P_a, P_b = Proportions of classes A and B in the training set, respectively.

$(MSR_a)_i$,$(MSR_b)_i$ = Mean spectral responses of the component classes A and B, respectively in wavelength band i.

M_{p_i} = Mean spectral response recorded at the scanner for the mixture of classes A and B in wavelength band i.

The terms P_a, P_b, and M_p are known for each training set, leaving the mean spectral responses of the component classes A and B to be determined. If there exist two sets of spectral signatures describing training sets with similar components but of different proportions, then it is possible to calculate the two component classes by solving the two equations simultaneously. A general solution for the two-component class case takes the form:

$$MSR_{bi} = \frac{M_{p_{1i}} P_{a_2} - M_{p_{2i}} P_{a_1}}{P_{a_2} P_{b_1} - P_{b_2} P_{a_1}} \tag{8}$$

Equation 8 produces the mean spectral response for class B. To derive the mean spectral response for class A, the result can be substituted into equation 7 which can be rewritten as

$$MSR_{a_i} = \frac{MP_i - P_b MSR_{bi}}{P_a} \tag{9}$$

If the elements of the covariance matrix for a terrain class spectral signature behave in a similar fashion, then this method could be used to obtain a calculated covariance matrix for a component terrain class. The covariance matrix for a four-channel case such as LANDSAT-1 can be treated as a 10 element array to simplify the calculations. This can be done since the off-diagonal elements of the covariance matrix are mirror images of each other.

A computer program called SIGCALC was written to take two extracted mixture terrain-class spectral signatures and calculate the mean vector and covariance matrix for the component classes. In addition, the proportions of the component classes in the training set determined from ground truth must be specified. Situations where $P_{a_2} P_{b_1} = P_{b_2} P_{a_1}$ must be avoided since the denominator of equation 8 cannot equal zero.

The above method for determining component class spectral signatures should provide reliable results if the proportion estimates for the component classes are accurate. However, it is often difficult to measure the proportions of component classes with the accuracy needed to estimate representative component class signatures. Also, due to the variability of spectral responses for component classes found in nature it was decided to use another method that estimates the component-class spectral signatures from more than two mixture-class signatures. This method involves using a stepwise linear regression analysis.

The stepwise linear regression analysis used is an applications program available at Colorado State University called STAT38R (ref. 14, 15). This program was used to develop a regression model that estimates the spectral response of a terrain class, either a component or mixture, given a set of mixture-class spectral signatures and the proportions of the component classes.

Simulations of LANDSAT data.- In order to examine the applicability of modeling mixture-class spectral signatures for automatic classification of LANDSAT-1 MSS data, a series of experiments was performed with simulated data. They include: identifying the expected improvement of automatic classification using modeled mixture terrain-class signatures versus conventional component class analysis, comparing modeled and extracted mixture-class mean vectors and covariance matrices, and analyzing the signature calculation technique.

Creation of simulated data fields: The automatic classification procedure conducted for these experiments approximates that of conventional RECOG LANDSAT-1 MSS analysis (ref. 10, 14, 15) with the exception of using modeled mixture class signatures and simulated LANDSAT-1 data. The steps are: 1) Examine control data on site to establish locations of representative training sets, determining the line and column numbers of the training sets on graymaps and extracting the mean vectors and covariance matrices from the data. 2) Simulate mixture signatures with the component class signatures obtained in Step 1 using Program MIX. MIX requires the mean vectors and covariance matrices of the component classes as input, as well as the mixture proportions desired for each mixture class. The simulated mean vectors and covariance matrices are punched on computer cards by MIX for future use, 3) Generate a random-normal data field for each class in RECOG format with program DTAFILE. This program uses a random-number generator that selects points as belonging to a given class based on the mean spectral response in each LANDSAT-1 MSS band and the appropriate covariance matrix, all within a normal Gaussian distribution. The overlying assumption here is that wildland terrain classes are normally distributed, which conforms to the assumptions imposed on the mixture modeling method (ref. 8) and the maximum likelihood pattern recognition algorithym (GLIKE) of RECOG (ref. 10). DTAFILE was designed to create simulated data fields of specified size for each class described by a mean vector and covariance matrix. The generated data points are written on a permanent file and stored for later use by any routine that uses data in a RECOG format.

The simulated data fields for each of the component and mixture classes were generated to contain 1000 spectral response values in each wavelength band, with each point being described by four variables (spectral response in each wavelength band).

The data fiels as generated by DTAFILE represent LANDSAT-1 MSS data with the added advantage of "absolute" ground truth information. With these fields it is possible to analyze the accuracy of automatic classification using the conventional RECOG maximum likelihood routine and the mixture terrain-class modeling method. An example data field is shown in figure 9 as a graymap of a nine-class field. The class fields can be seen as horizontal bands 100 points across and 10 points down.

Classification of component classes: The initial experiment performed involved using a classification of simulated data that contained fields of two component classes and three mixture classes modeled from these components. Two component classes, Grassland (denoted by the symbol, G) and Forest (F), were identified on NASA RB-57 (scale 1:100,000) color IR aerial photographs and located on a graymap of August 20, 1972 LANDSAT-1 MSS (Frame no. 1028-17135) data over south-central Colorado. Signatures were extracted and then used to obtain mixture class signatures with program MIX. The mixture classes were simulated with a specified proportion increment of 0.25, producing three mixture classes of 75% Grassland - 25% Forest (C), 50% Grassland - 50% Forest (D), and 25% Grassland - 75% Forest (E). The mean spectral-response curves for the five classes are shown in figure 10.

The simulated data fields were then treated as actual LANDSAT-1 MSS data for purposes of classification. Component-class signatures were extracted from those fields created with component-class signatures (top 3 rows, fig. 9) and used to classify the entire set of simulated data. The classification display is shown in figure 11 with the Grassland field at the top and the Forest field at the bottom. The mixture class fields are located between the components with C nearest the Grassland field, D in the middle and E nearest the Forest field. Note the number of unclassified points (seen as blanks in the display) in each of the mixture-class fields. These points were thresholded out at the 11.100 percentile as not belonging to either of the component classes. Misclassifications of the mixture-class fields occur only in fields C and E, with points being misclassified as the component class which has the highest proportion in the mixture.

The absolute classification accuracies were determined for each clas . Since each data point was created as a specified class, a point was considered misclassified if it was classified as any class other than that specified. The results are summarized in the following classification confusion matrix (CCM) listing each class, the number of points that were classified correctly, and the number of points misclassified as other classes.

CLASS	F	E	D	C	G
FOREST (F)	1000				
25% GRASSLAND - 75% FOREST (E)	148				
50% GRASSLAND - 50% FOREST (D)					
75% GRASSLAND - 25% FOREST (C)					376
GRASSLAND (G)					1000

The table is read with the true class name and symbol on the left and the symbols of classes that a point was classified as across the top. The diagonal elements of the matrix represent accurate classifications. The off-diagonal elements represent either Type I or Type II errors, defined as follows.

Decision	True Classification	
	X belongs to A	X Does Not Belong to A
Classify AS A	Correct Decision	Type I Error
Do Not Classify AS A	Type II Error	Correct Decision

The overall accuracy of this classification was only 40.0% which in most LANDSAT-1 applications should be considered quite poor <u>even though the component classes were classified perfectly</u>.

Classification of two component and three mixture classes: The objectives of the study presented in this section were twofold. The first was to verify that the mixture-class spectral-signature modeling technique produces signatures that can be used by RECOG to classify pixels that contain mixtures. The second, assuming the first to be satisfied, was to determine the increase in information acquired using modeled mixture-class signatures over classification conducted with only component-class signatures.

The same data-field generated for the analysis of component classes was used for this experiment. Treating the data file as actual LANDSAT-1 MSS data, component-class mean vectors and covariance matrices were extracted. Signatures for the mixture-class fields were not extracted from the data, but instead were modeled from the component-class mean vectors and covariance matrices. This procedure follows that which could be used with actual LANDSAT-1 data since representative mixture-class training sets are usually difficult to locate in natural situations. The two extracted component and three modeled mixture-class signatures were used to classify the simulated data file.

At this point it was desirable to compare modeled and extracted mean vectors and covariance matrices. The signatures of the mixture classes were extracted from the simulated data and compared with those that were modeled from the two component classes, as follows:

	Extracted				Simulated			
CLASS	GRASS							
BAND	1	2	3	4				
MEAN	32.45	34.13	39.10	19.40				
STD. DEV.	1.07	1.82	1.54	.88				
CLASS	75% GRASS-25% FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	28.57	28.48	34.40	17.26	28.57	28.46	34.35	17.21
STD. DEV.	1.06	1.80	1.65	1.01	1.04	1.72	1.66	.99
CLASS	50% GRASS-50% FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	24.72	22.86	29.62	15.04	24.68	22.78	29.61	15.01
STD. DEV.	1.05	1.54	1.74	1.11	1.00	1.54	1.76	1.09
CLASS	25% GRASS-75% FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	20.84	17.14	24.93	12.91	20.80	17.11	24.86	12.82
STD. DEV.	.99	1.42	1.87	1.21	.97	1.35	1.86	1.18
CLASS	FOREST							
BAND	1	2	3	4				
MEAN	16.92	11.44	20.12	10.63				
STD. DEV.	.95	1.12	1.96	1.26				

The results indicate that the extracted and modeled mean and standard deviation vectors vary, at most, by less than 0.1 of one standard deviation unit, which can be considered as quite adequate for this analysis. Figure 12 shows the classification display of the results. A classification confusion matrix (CCM) was also compiled for this analysis. It is a comparison of extracted and modeled mean and standard deviation errors. The standard deviations shown represent the square root of diagonal covariance elements. The results are as follows:

CLASS	F	E	D	C	G
FOREST (F)	997	3			
25% GRASSLAND - 75% FOREST (E)	9	977	14		
50% GRASSLAND - 50% FOREST (D)		8	967	25	
75% GRASSLAND - 25% FOREST (C)			12	970	18
GRASSLAND (G)				8	992

The classification results indicate that simulated mixture-class spectral signatures can be used to classify MSS data using the RECOG program. The classification information and accuracy were also greatly increased by the use of mixture signatures. The overall classification accuracy using this technique increased to over 98%, with only a slight decrease in the accuracies of component classes. It should be noted here that automatic classification of generated data fields will normally produce high classification accuracies due to the lack of "alien objects" which are often found in real life situations (ref. 3). The accuracies produced here should, instead, be considered a measure of the effectiveness of the classification technique. These results do indicate that given a pixel containing a mixture, it should be possible to accurately identify it with this method.

Classification of three component and six mixture classes: The successful results obtained with the five-class analysis warranted further study, so an additional component class was added to the signature set. A mean vector and covariance matrix for Water (W) was extracted from August 20, 1972 LANDSAT-1 MSS data. Mixture signatures were modeled from the Grassland component class and the new Water class with a proportion increment of 0.25. The resulting three mixture classes and the component class combined with the original signature set bring the total number of signatures used to nine. Figure 13 shows the mean spectral response curves for the Grassland-Water mixture classes. The separation of these classes was adequate in all bands. The mean spectral response curves for all nine classes are illustrated in figure 14.

The classification procedure followed that of all previous experiments with a new data field being generated to include all nine classes. The classification display is shown in figure 15. A classification confusion matrix (CCM) for the 9-class analysis is as follows:

CLASS	F	E	D	C	G	Q	R	S	W
FOREST (F)	997	3							
25% GRASSLAND - 75% FOREST (E)	8	982	10						
50% GRASSLAND - 50% FOREST (D)		8	962	22		8			
25% GRASSLAND - 75% FOREST (C)			10	954	12	24			
GRASSLAND (G)				14	986				
75% GRASSLAND - 25% WATER (Q)			6	18		976			
50% GRASSLAND - 50% WATER (R)							999		
25% GRASSLAND - 75% WATER (S)								996	
WATER (W)									962

These results show a good classification accuracy for all the classes, with an average classification accuracy of 98%. Note that the greatest confusion exists between mixture classes that have the same proportion factor for a given component class, i.e., 75% Grassland-25% Forest and 75% Grassland-25% Water. This is probably the result of masking of the lesser components by the Grassland response.

The following comparison of the modeled and extracted mean and standard deviations was made as an additional check of the mixture modeling method:

CLASS GRASSLAND	Extracted				Simulated			
BAND	1	2	3	4				
MEAN	32.49	34.19	39.09	19.46				
STD. DEV.	1.11	1.91	1.49	.92				

CLASS FOREST	Extracted				Simulated			
BAND	1	2	3	4				
MEAN	16.95	11.48	20.19	10.67				
STD. DEV.	.92	1.15	1.95	1.26				

CLASS WATER	Extracted				Simulated			
BAND	1	2	3	4				
MEAN	16.72	8.13	3.98	.24				
STD. DEV.	.55	.74	.63	.44				

CLASS 75% GRASSLAND-25% FOREST	Extracted				Simulated			
BAND	1	2	3	4	1	2	3	4
MEAN	28.57	28.48	34.40	17.26	28.60	28.51	34.36	17.56
STD. DEV.	1.06	1.80	1.65	1.01	1.06	1.75	1.62	1.00

CLASS 50% GRASSLAND-50% FOREST	Extracted				Simulated			
BAND	1	2	3	4	1	2	3	4
MEAN	24.68	22.81	29.61	15.01	24.72	22.84	29.64	15.06
STD. DEV.	1.00	1.52	1.77	1.08	1.02	1.57	1.73	1.10

CLASS 25% GRASSLAND-75% FOREST	Extracted				Simulated			
BAND	1	2	3	4	1	2	3	4
MEAN	20.83	17.10	24.82	12.83	20.84	17.16	24.91	12.87
STD. DEV.	1.01	1.33	1.82	1.17	.98	1.38	1.84	1.18

CLASS 75% GRASSLAND-25% WATER	Extracted				Simulated			
BAND	1	2	3	4	1	2	3	4
MEAN	28.49	27.62	30.30	14.59	28.55	27.67	30.31	14.65
STD. DEV.	1.01	1.62	1.35	.83	1.00	1.69	1.33	.82

CLASS 50% GRASSLAND-50% WATER	Extracted				Simulated			
BAND	1	2	3	4	1	2	3	4
MEAN	24.55	21.04	21.47	9.77	24.61	21.16	21.53	9.85
STD. DEV.	.85	1.39	1.20	.76	.87	1.44	1.14	.71

CLASS 25% GRASSLAND-75% WATER	Extracted				Simulated			
BAND	1	2	3	4	1	2	3	4
MEAN	20.67	14.63	12.75	5.00	20.66	14.64	12.76	5.04
STD. DEV.	.71	1.11	.91	.68	.73	1.14	.92	.59

As in the case of the five-class classification, the simulated mixture signatures closely approximate those extracted from the generated datafields. These results imply that if a representative set of component signatures can be found, then mixtures of these classes may be successfully modeled and used to classify pixels containing those mixtures.

Test of signature estimation technique: In wildland areas where large areas of homogeneous terrain classes are difficult to locate, it becomes necessary to obtain representative spectral signatures from a relatively small number of pixels. In some areas it is even difficult to find any pixels that contain only one terrain class (ref. 16). Consequently, methods to obtain component terrain-class signatures from pixels containing mixtures of terrain classes in known proportions have been investigated (for example, see ref. 17). The method involves solving a set of linear equations assuming that the data is from a Gaussian distribution and each signature has a common covariance matrix. Our study used a similar, but much simplified technique to attempt to acquire component-class spectral signatures from training-set data of terrain classes of known mixtures. A test of the method to obtain component-class spectral signatures using equations 8 and 9 was conducted using simulated data fields created in the same manner as the preceding experiments. The specific objectives of this test were: 1) to verify that this method could be used to determine component class signatures suitable for automatic mapping of LANDSAT-1 type MSS data when only mixture terrain classes were available for signature extraction, and 2) find out the type of classification accuracies that can be expected when implementing these estimated signatures as input into the model and RECOG.

The initial test used the random-normal data fields for the nine-class analysis described above. Spectral signatures were extracted from the data for mixture classes 75% Grassland-25% Forest (C), 50% Grassland-50% Forest (D), 75% Grassland-25% Water (Q), and 50% Grassland-50% Water (R). The pairs of extracted signatures and their corresponding proportion factors were input into SIGCALC and the component signatures for Grassland, Forest and Water were calculated.

Comparison between component class mean and standard deviation vectors extracted from the data set and those calculated with SIGCALC is as follows:

	Extracted				Calculated			
CLASS	GRASS							
BAND	1	2	3	4	1	2	3	4
MEAN	31.93	33.64	38.58	18.89	32.05	33.70	38.62	19.05
STD. DEV.	1.14	1.91	1.53	.92	1.30	1.97	1.48	.97
CLASS	FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	16.43	10.93	19.69	10.13	16.41	10.94	19.58	9.95
STD. DEV.	.92	1.19	1.99	.96	1.05	1.21	1.98	1.31
CLASS	WATER							
BAND	1	2	3	4	1	2	3	4
MEAN	16.21	7.61	3.50	.04	16.17	7.70	3.64	.00
STD. DEV.	.63	.80	.66	.19	.61	.82	.76	.52

The calculated component class signatures closely approximate the extracted, with the average deviation being 0.08 for mean vectors and 0.15 for standard deviations.

In order to compare the effectiveness of estimated signatures with extracted signatures, the nine-class data field was classified by the computer. The class signatures used as input included the three estimated component signatures and six mixture-class signatures modeled from the estimated signatures. The classification accuracy obtained was approximately 2% less than that obtained when using extracted component signatures. The classification map is shown in figure 16. The classification confusion matrix (CCM) of this data set is as follows:

CLASS	F	E	D	C	G	Q	R	S	W
FOREST (F)	984	13							
25% GRASSLAND - 75% FOREST (E)	17	973	10						
50% GRASSLAND - 50% FOREST (D)		20	964	11		4	1		
25% GRASSLAND - 75% FOREST (C)			27	876	25	65			
GRASSLAND (G)				59	907				
75% GRASSLAND - 25% WATER (Q)			28	24	2				
50% GRASSLAND - 50% WATER (R)							1000		
25% GRASSLAND - 75% WATER (S)								1000	
WATER (W)									1000

An overall classification accuracy of 96.1% was obtained, indicating that good classification results can be expected using this method.

The simulated data fields that contained the mixture-class signatures used to calculate the component classes represent uniform data. That is, each data point (pixel) represents a similar class of objects. Because this situation is sometimes difficult to find in natural situations, the technique was tested for mixture signatures extracted from non-uniform data or training sets. Non-uniform training sets can be defined as areas that contain a mixture of component classes with single pixels containing various proportions of the components. If the non-uniform training set is considered in its entirety the mean spectral response should approximate that from a uniform training set with the same proportions of component classes. The concept of uniform and non-uniform data is illustrated in figure 17 as two 20-pixel training sets, one containing a uniform mixture of two component classes (white and shaded) (a) and one containing a non-uniform mixture (b), but both with identical overall proportions of component classes. In the uniform training set each pixel (outlined by heavy black lines) contains a 50% mixture of two component classes, producing a 50% mixture within the entire training set. Since the spectral response from a pixel is integrated at the scanner optics, the within-cell distribution of the components does not affect the between-cell variance of the training set. The non-uniform training set is composed of some pixels containing 100% of one class, some containing 100% of the other class and some that contain various mixtures of the two components. The overall mixture proportion averaged over the entire training set is 50% of each component class. Referring to the equation 28 (Appendix) it can be easily deduced that the mean vectors obtained from each training set should equal each other assuming constant illumination and no within-class variability. By equation 30 (Appendix) however, it can be seen that covariance matrices will most likely vary due to the between-pixel variability existing in the non-uniform training set.

To verify these statements, mixture-class signatures from simulated non-uniform data were extracted and compared with those extracted from uniform data created with the same input signatures. A comparison of extracted and calculated mean and standard deviation vectors of the component classes for uniform and non-uniform(*) data are as follows:

	Extracted				Calculated			
CLASS	GRASS							
BAND	1	2	3	4	1	2	3	4
MEAN	32.02	33.68	38.60	18.97	31.81	33.69	38.67	18.77
STD. DEV.	1.16	1.97	1.52	.96	1.08	1.86	1.64	1.00
CLASS	GRASS*							
BAND	1	2	3	4	1	2	3	4
MEAN	32.02	33.68	38.60	18.97	31.98	33.67	38.61	18.95
STD. DEV.	1.16	1.97	1.52	.96	5.17	8.59	11.15	6.02
CLASS	FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	16.43	10.93	19.69	10.13	16.53	10.93	19.55	10.13
STD. DEV.	.92	1.19	1.99	1.29	1.04	1.21	.88	1.31
CLASS	FOREST*							
BAND	1	2	3	4	1	2	3	4
MEAN	16.43	10.93	19.69	10.13	16.47	10.93	19.75	10.18
STD. DEV.	.92	1.19	1.99	1.29	5.02	7.32	6.33	4.35
CLASS	WATER							
BAND	1	2	3	4	1	2	3	4
MEAN	16.23	7.62	3.47	0.03	16.27	7.71	3.53	0.00
STD. DEV.	.63	.79	.67	.17	.63	.84	.56	.47
CLASS	WATER*							
BAND	1	2	3	4	1	2	3	4
MEAN	16.23	7.62	3.47	0.03	16.28	7.71	3.47	0.09
STD. DEV.	.63	.79	.67	.17	10.10	16.50	22.20	12.10

Note the close agreement of means and standard deviations from the uniform data and the close agreement of means but non-agreement of standard deviations calculated from the non-uniform data.

The consequence of using such large covariance matrices obtained from the non-uniform data is illustrated in Figure 18. The nine-class simulated data fields were classified with the calculated component signatures listed above. The first classification map (fig. 18a) was obtained using signatures calculated from uniform training sets and the second (fig. 18b) was obtained using signatures calculated from non-uniform training sets. The classification accuracy of component classes obtained for the uniform case was 99.1% whereas the accuracy obtained for the same classes in the non-uniform case was only 33.3%. The latter figure is misleading in that the nine-class data set was classified as 97.9% Grassland, 1.7% and .4% water. These figures indicate that the class distribution for Grassland was improperly defined by the covariance matrix.

The above study, coupled with the frequent lack of uniform mixture-class training sets in natural situations, lends itself to implementing an artificial covariance matrix for describing distributions of component classes. The use of a covariance matrix common to each class was discussed (ref. 8)for another, more sophisticated, signature calculation technique. Mean vectors for component classes can be successfully calculated from non-uniform mixture-class training sets. If a common covariance matrix that describes the proper distribution can be found, then with the calculated mean vectors, the signature should be representative of its respective class. To verify this, the component-class mean vectors calculated from non-uniform mixture class training sets and various common covariance matrices were used to classify the simulated nine-class data fields.

The first classification utilized a covariance matrix with diagonal elements (σ^2_{ii}) of (1,3,3,1) and off-diagonal elements of 0.5. The determinations for the covariance elements were made by a priori inspection of covariance matrices for component and mixture classes extracted from simulated uniform data. The standard deviations in each band represent an estimate of the average standard deviations for all classes. The off-diagonal elements were established to be 0.5 as an estimated average of the correlation values for all of the 9-class correlation matrices.

The classification procedure followed that of all earlier runs with the mixture class mean vectors being modeled from the component class mean vectors calculated from non-uniform data. The classification map is shown in figure 19; its corresponding CCM is as follows:

CLASS	F	E	D	C	G	Q	R	S	W
FOREST (F)	993	5							
25% GRASSLAND - 75% FOREST (E)	13	975	11				1		
50% GRASSLAND - 50% FOREST (D)		17	959	15		9	2		
25% GRASSLAND - 75% FOREST (C)			19	897	21	63			
GRASSLAND (G)				26	974				
75% GRASSLAND - 25% WATER (Q)			20	23		956			
50% GRASSLAND - 50% WATER (R)							1000		
25% GRASSLAND - 75% WATER (S)								1000	
WATER (W)									1000

The acceptable level of classification accuracy obtained (97.3%) indicated that utilizing a common covariance matrix may help solve the problem of establishing representative spectral signatures for component classes when there exist insufficient points for conventional signature extraction.

An additional classification was conducted with the same class mean vectors as the above analysis, but with a common covariance matrix with the same diagonal elements and off-diagonal elements set to 0.0. The purpose of this run was to establish the expected classification accuracy for 0.0 correlations between channels. The classification results were again promising,with the average accuracy obtained being 96.5%. The classification map is shown in fig. 20; its corresponding CCM is as follows:

CLASS	F	E	D	C	G	Q	R	S	W
FOREST (F)	970	25							
25% GRASSLAND - 75% FOREST (E)	26	939	35						
50% GRASSLAND - 50% FOREST (D)		20	942	13		20	5		
25% GRASSLAND - 75% FOREST (C)			13	896	25	66			
GRASSLAND (G)				15	983				
75% GRASSLAND - 25% WATER (Q)			20	22		958			
50% GRASSLAND - 50% WATER (R)							1000		
25% GRASSLAND - 75% WATER (S)								1000	
WATER (W)									1000

One additional common covariance matrix was calculated, and tested as suggested by Nalepka and others (ref. 8, p. 17). This matrix consisted of an average of the covariance matrices of their signature set. One immediately recognizable disadvantage of this method is that a set of representative covariance matrices must be obtained from the MSS data which have small values such as those expected from component class covariance matrices extracted from uniform data. The results of using this technique for calculating the diagonal elements, but setting the off-diagonal elements to 0.0 are shown in figure 21. The corresponding CCM is as follows:

CLASS	F	E	D	C	G	Q	R	S	W
FOREST (F)	960	23							
25% GRASSLAND - 75% FOREST (E)	25	939	35						
50% GRASSLAND - 50% FOREST (D)		20	941	13		22	4		
25% GRASSLAND - 75% FOREST (C)			10	893	26	70			
GRASSLAND (G)				18	977				
75% GRASSLAND - 25% WATER (Q)			23	25		952			
50% GRASSLAND - 50% WATER (R)							1000		
25% GRASSLAND - 75% WATER (S)								1000	
WATER (W)									1000

Note that the 9-class data classification results are slightly less (95.0%) than those obtained using an estimated average of the diagonal elements and setting the off-diagonal elements to 0.0.

The overall results obtained from using common covariances appear promising for classifying LANDSAT-1 type MSS data when signatures must be obtained from non-uniform mixture class training sets. This technique applied to automatic classification of actual wildland data, is described in the next chapter.

Tests using LANDSAT-1 MSS data.- Experiments similar to those performed with simulated MSS data were conducted with actual LANDSAT-1 MSS data. An initial experiment was performed using extracted assumed "pure" component-class spectral signatures and mixture-class spectral signatures simulated from these component classes to classify an area designated as the 11-Mile study area. An additional experiment using estimated component signatures and simulated mixture signatures was performed using several plots located in the Manitou study area where ground control data were more detailed. Both of these experiments utilize methods described above. The objective of these experiments was to relate the concepts just presented to a real-world application. Hopefully, this will provide insight as to how well these techniques work to improve classification accuracies and/or what other factors may need to be considered to obtain optimum results.

LANDSAT-1 MSS data acquisition and description: The set of LANDSAT-1 MSS data analyzed is from an August 15, 1973 (frame number 1388-17134) overpass of the test site. The data set was provided by Dr. Richard S. Driscoll, Remote Sensing Project Leader at the U.S. Forest Services Rocky Mountain Forest and Range Experiment Station located in Fort Collins, Colorado. The data were geometrically corrected (deskewed) by the Laboratory for Applications of Remote Sensing (LARS) located at Purdue University, West Lafayette, Indiana (ref. 5, 18). The advantage of the deskewed data is its better conformity to aerial photographs, especially when displayed as a microfilm graymap or recognition map. A large section of this LANDSAT-1 frame was also studied by Heller and others (ref. 4) and Fleming (ref. 19).

Two areas from this frame were selected for study. They are referred to below as the Elevenmile Canyon Reservoir Study Area and the Manitou Study Area.

Experiments were conducted in these study areas using extracted component class and simulated mixture-class spectral signatures to classify a portion of LANDSAT-1 MSS data. The steps involved in the classifications are: (1) selecting component classes, (2) extracting representative spectral signatures, (3) simulating specified mixture-class spectral signatures, (4) classifying the data with a maximum likelihood processor, and (5) analyzing the results.

Elevenmile Canyon Reservoir Study Area: This area was selected because it is representative of the diversity of topography, geology, and vegetative cover throughout much of the total test site; and because of the availability of LANDSAT-1 MSS tape data. The location of the site is shown in figure 1. Elevations range from 2500 to 3020 m. The site is at the southwest margin of South Park: it includes the eastern part of Elevenmile Canyon Reservoir. Mountainous terrain lies to the north and east; a wide rolling grassland area lies primarily to the south of the reservoir. Wet meadows occur primarily along stream courses.

Two principle types of rocks occur; volcanic rock and granodiorites. The most commonly found volcanics are upper and lower andesite members of the Thirty-nine Mile volcanic field. The upper member is generally associated with high flat-topped mountains and the lower with the low rolling hills characteristic of South Park.

The granodiorite found in the site is dark gray, medium to coarsely crystalline and contains gneiss. It is correlated with the precambrian Boulder Creek Granodiorite of the central front range of the Rocky Mountains. It is found mainly in the northern portion of the test area where it forms high steep mountains. Several outcrops can be found to the east of Elevenmile Canyon Reservoir and on both sides of Elevenmile Canyon, but their aerial extent is limited. Elsewhere, most of the bedrock is covered by some form of vegetation.

The vegetation in the area consists of a variety of forests and grasslands. The forests, which cover approximately one-third of the test site, include: (1) ponderosa pine, (2) Douglas-fir, (3) spruce fir, subalpine fir, and (4) Aspen. According to Heller and others (ref. 4), these forest species occur as "pure" types, but in many places there are various mixtures of species. There also exists a wide range of tree canopy densities, from very open to so dense that crowns nearly touch. In open stands of trees enough sunlight exists to allow the development of extensive herbaceous and/or shrubby understory, but under conditions of dense tree cover, little understory cover has developed.

The principle understory vegetation associated with Ponderosa pine forest includes Arizona fescue and mountain muhly. At higher elevations these species give way to Idaho fescue and Thurber fescue and oatgrass.

Approximately one-third of the study site consists of low rolling hills covered with vegetation characteristic of the short-grass prairie found throughout South Park. The principle species are blue grama and slim stem muhly. Mountain bunchgrass communities can also be found at the interface between grassland and forest.

Also occuring in the area are wet meadow and stream bank communities. These communities occur along the shore of Elevenmile Canyon Reservoir and along streams. Various species of sedges, rushes, and bulrush occur either as monospecific or mixed stands in the moist areas. In drier areas, blue grass and tufted hairgrass are found in mixed communities. The prominent shrubby communities consists of willow and shrubby cinquefoil. Although they occur sparsely throughout the study area, these communities are generally found associated with the wet meadows.

Five component classes were selected for use in this study area. They are Forest, Grassland, Water, Mountain Grassland and Wet Meadow. These classes were selected because of their importance in the area and also because mixtures of these various types occur. The Forest class is composed of several types of trees with a minimal amount of understory vegetation showing through the canopy. The Grassland class is characteristic of the short grass prairie community found throughout South Park and usually has much bare soil associated with it. The Water class is representative of the water of the reservoir. The Mountain Grassland class represents grassland communities found at higher elevations and at the interface between Forest and Grassland. The Wet Meadow class was selected to represent irrigated meadow areas and stream bank communities.

Homogeneous areas representative of each class were delineated on aerial photographs. These areas (training sets) were then transferred to LANDSAT graymaps and the line and point coordinates were determined. Figure 22 shows the microfilm graymap of MSS BAND 5 for the study site. Some modification of the photo locations were made on the graymaps as dictated by the graylevels associated with each training set. In this way, anomalous data points were avoided before signatures were extracted. A number of training sets were located for each class to provide statistically representative signatures. The number of points used to calculate each class signature were: Forest (75), Grassland (28), Mountain Grassland (20), Wet Meadow (20), and Water (96). The number of points for Grassland, Mountain Grassland and Wet Meadow was small due to difficulties encountered in locating training sets on computer graymaps, but are assumed representative for this study.

Mean vectors and covariance matrices were extracted for each training set for each class. The mean vectors for a class were compared with one another and the eigenvalues and eigenvectors of each class training set were determined and analyzed. Those training sets exhibiting dissimilar mean vectors and eigenvector plots were discarded so that final signatures obtained were as representative as possible.

The final component-class signatures were then used to classify the area. The resulting classification map is shown in Figure 23 as a microfilm display. As a check of how well the signatures identify their respective classes, the number of correct classifications in each training set was determined. All of the classes had training-set classification accuracies of 100% except Water which had 99%. A total of 60.5% of the test area was classified as one of the component classes, indicating that for a conventional analysis such as this, more classes need to be added in order to classify the entire area. To satisfy this need, the following mixture classes were selected:

1. Grassland-Forest--selected to identify pixels covering areas with trees growing over Grassland areas and interfaces between the two communities.
2. Grassland-Wet Meadow--selected to identify pixels falling on the interface of the grass covered areas and irrigated meadows and stream bank communities.
3. Forest-Mountain Grassland--selected to identify pixels covering the interfaces between the two communities and locate areas of Mountain Grassland that appear as openings in the Forest which are smaller than the resolution of the LANDSAT-1 MSS scanner.

A proportion increment of 0.25 was selected for each combination of classes, and mixture-class spectral signatures were simulated with program MIX. The final signature set used to classify the test area included the five extracted component classes and nine intermediate mixture classes. The microfilm classification map at the 11.100 Chi-Square threshold level is shown in Figure 24. An increase in the number of classified points resulted in 86.35% of the area being classified, rather than 60.5%. Most of the unclassified points are associated with the interface between Grassland and Water since no spectral signatures characteristic of these mixtures were used in the classification. Classification of Grassland-Wet Meadow and Forest-Mountain Grassland mixtures appear reasonable. Grassland-Forest mixtures of 50% Grassland-50% Forest and 25% Grassland-75% Forest also appear to be classified well. The 75% Grassland-25% Forest mixture class contained a large amount of obvious misclassifications in areas where no Forest was present.

To investigate the source of these misclassifications, spectral signatures were extracted from the data for areas classified as the 75% Grassland-25% Forest class. A comparison of the mean and standard deviation vectors for the extracted signature and the simulated mixture signature showed them to be almost identical, as shown below:

	Extracted				Simulated			
MSS BAND	1	2	3	4	1	2	3	4
Mean	29.42	28.79	37.34	19.21	28.97	28.87	37.19	19.09
Std. Dev.	.94	1.66	1.51	1.00	1.62	2.02	2.10	1.15

This indicates that the material that is being classified as a Grassland-Forest Mixture has a vegetation or soil type that produces a response similar to that calculated for the mixture. The consequences of such a misclassification are an overestimation of the amount of Forest in the scene and an underestimation of Grassland. Because of the difficulties encountered in establishing mixture densities from high altitude aerial photographs no quantitative accuracy analysis was obtained for this experiment.

Because of the gross misclassifications of the 75% Grassland-25% Forest class, the proportion increment used to simulate mixture class signatures was changed to 0.33 and new signatures were simulated. The resulting signature set included the five component classes and ten two-component mixture classes. Combinations of Grassland and Water and Water and Wet Meadow were made as an attempt to identify those pixels left unclassified around the reservoir in the previous experiment.

A classification was made, producing the microfilm display shown in figure 25. The results indicated that the areas previously misclassified as 75% Grassland-25% Forest were still being incorrectly identified. A number of points along the shoreline of the reservoir were identified as either 67% Grassland- 33% Water or 33% Grassland-67% Water. There appeared to be some confusion of 67% Wet Meadow-33% Water and 33% Wet Meadow-67% Forest as the component class with the greatest amount occuring in the mixture. The other mixture classes appear to be classified correctly since they appeared where mixtures would logically exist.

Manitou study area: The Manitou study area is located to the northeast of the Elevenmile Canyon Reservoir area in and around the U.S. Forest Service Manitou Experimental Forest. Only a few selected plots identified by Mead (ref. 5) were studied in this work.

The plots lie within an area located between 38° 52' 30" and 39° 22' 30" north latitude and 105° 00" and 105° 37' 30" west longitude, along and just north of the area shown in fig. 1. The area typifies the mountainous part of the east slope of the Rocky Mountains, with a large variation of slopes and aspects. The topography is dominated by mountains and valleys with a north-south orientation due to extensive faulting that occured during the Pliocene epoch and earlier.

For the most part, geologic materials in the area are covered with vegetation. The principle materials are pinkish to reddish soils developed from Pikes Peak granite. The underlying geologic unit is the Fountain Formation, but exposures are limited.

Variations in topography in the area have resulted in a complex mixing of plant communities generally aligned with the topography. For the most part, the plant communities are similar to those described for the forested hillslopes, mountains, and wet meadow areas of the Elevenmile Canyon Reservoir study area.

In this study area, experiments were performed to classify LANDSAT-1 MSS data with component class signatures estimated by solving simultaneous equations and linear regressions. Both of these techniques require data where the component classes and their respective proportions of the scene are known.

Several plots located in the Manitou study area and identified by Mead (ref. 5) were analyzed to obtain a number of plots suitable for signature estimation. Mead was specifically interested in the percent cover of ponderosa pine of the plots so it was a logical choice for one of the component classes. The second component class was identified as the other materials in the scene and denoted as Background. The Background class is actually a mixture in itself of soil, bare rock, and understory vegetation. Use of this Background class necessitated using only those plots which contained similar species occuring in the same relative amounts. It was also necessary to identify plots with similar slopes and aspects to avoid the signature variability associated with these factors. In summary, the criteria used for selecting the plots used for signature estimation were:

1. The plots must have known percent covers of ponderosa pine. (The Background percent cover was determined by subtracting the ponderosa percent cover from 100).

2. The plots must contain similar background materials such as vegetation, rocks and bare soil.

3. The plots must have similar orientation as determined by slope and aspect.

The first criterion was met by analysis of aerial photography by five photo interpreters from the U.S. Forest Service. The second was accomplished by field studies by Mead using a line-transect method to establish the materials present in each plot as well as their frequency of occurance. The third was met by locating the plots on topographic maps and determining slopes and aspects by a computer program called TOPOGO (Mead, ref. 5; Sharpnack and Akin, ref. 20).

Analyzing all of the plots with these criteria produced a small subset of only six plots representing 120 total LANDSAT-1 MSS data points. The plots are as follows:

PLOT NUMBER	MEAD'S PLOT NUMBER (ref. 5)	PONDEROSA PINE PERCENT COVER	SLOPE (%)	ASPECT	FREQUENTLY OCCURING BACKGROUND VEGETATION (see species list, below)
1	1	46	7.9	96	J, B, M, P, Q, D
2	2	74	17.0	53	B, J, M, P, D
3	3	76	6.9	108	J, B, M, Q, P, D
4	6	64	5.0	59	B, M, J, Q, F, D
5	14	22	8.1	77	B, M, F, P, B
6	17	16	9.9	134	M, AF, P, Q, BB

SPECIES LIST

ABBREVIATION	NAME
F	fringed sagebrush (*Artemisia frigida* Willd)
B	bear berry (*Arctostaphylos uva-ursi* (L.) Spreng.)
AF	Arizona fescue (*Festuca arizonica* Vasey)
J	common juniper (*Juniperus communis* (L.))
P	prairie junegrass (*Koelaria cristata* (L.) Pers.)
M	mountain muhly (*Muhlenbergia montana* (Nutt.) Hitch.)
Q	quaking aspen (*Populus tremuloides* Michx.)
D	Douglas-Fir (*Pseudotsuga menziesii* var. *glauca* (Beissn.) Franco)
BB	bottlebrush squirreltail (*Sitanion hystrix* (Nutt.) J.G.Sm.)

Spectral signatures for the six mixture plots were extracted from the LANDSAT-1 MSS data. The mean spectral response curves for the six plots are shown in Figure 26. Several pairs of signatures along with their respective mixture proportions of ponderosa pine and background were used as input for program SIGCALC. The results showed a wide variation in the estimated mean vectors for both of the component classes. In addition, the estimated covariance matrices contained very unrealistic values including negative variances. Consequently, this method was found to be unreliable for estimating component class spectral signatures for this case. Since results obtained with simulated data described above were acceptable, the error was probably due to inaccurate estimates of percent cover and variations of materials associated with the Background class. Since no representative spectral signatures could be obtained, no modeling of mixture class signatures or classification of MSS data was attempted.

Since the signature estimation technique used in the above study uses only data for two mixture training sets, slight errors in percent cover estimates will result in large errors for the calculated signature values. Also, the signatures, if successfully estimated, may only be representative of the material found in the specific training areas. To compensate for this all of the mixture training sets were used to determine component class signatures by means of linear regression techniques.

The method used for estimating component class signatures with linear regression was discussed above. The necessary input data were the mean spectral response recorded from each of the plots in each wavelength band and the proportions of the two component classes, ponderosa pine and Background. Models were developed for each wavelength band as follows:

MSS BAND	Regression Model
.5-.6 μm	$MSR=24.07-5.98(P_p)$
.6-.7 μm	$MSR=20.92-7.60(P_p)$
.7-.8 μm	$MSR=32.05-7.44(P_p)$
.8-1.1 μm	--

where MSR=Estimated mean spectral response
P_p=Proportion of ponderosa pine in the scene

No regression model could be developed for the .8-1.1 m wavelength band so a review of estimates of percent cover for the plots was made. It was found that for some plots, the cover estimates made by the five photo interpreters varied by as much as 30%. The standard deviations for the mean cover estimates ranged from 5.4 to 13.0 for the six plots. Additional cover estimates were then made using a dot grid superimposed on enlarged photos of the six plots. For these estimates no species determinations were made, but all trees were counted as belonging to a Forest class and all non-tree materials were counted as Background. New regression models were developed as follows:

MSS Band	Regression Model
.5-.6 μm	$MSR=25.70-10.22(F_p)$
.6-.7 μm	$MSR=23.46-14.03(F_p)$
.7-.8 μm	$MSR=32.24-13.23(F_p)$
.8-1.1 μm	$MSR=19.55-8.73\ (F_p)$

where MSR=Estimated mean spectral response
F_p=Proportion of Forest in the scene

High R^2 values indicate that the grid estimates for the plots are reasonable values for the amounts of Forest and Background. Since only six plots were used in developing the regression models they may only be applicable for analysis of these plots. Component-class signature estimates can be determined from the models by setting the Forest proportion (F_p) to 1.0 and calculating the MSR's in each band to determine the Forest signature and setting F_p to determine the Background signature. The mean spectral response curves for the estimated Forest and Background Signatures are shown along with the extracted mixture curves for each of the plots in Fig. 26.

The estimated mean vectors for the Forest and Background classes were input into program MIX and mixture class signatures were simulated. The proportion increment used was set at 0.25, producing three intermediate mixture signatures. No covariance matrices for the component-class signatures were estimated by the regression models due to the non-uniformity of the mixture plots, so a common covariance matrix with diagonal elements of 1.0, 3.0, 3.0, 1.0 and off-diagonal elements of 0.0 was used for each of the classes.

The six plots selected for the study were classified with a maximum likelihood pattern recognition processor. Initially, only the estimated signatures for Forest and Background were used to classify the plots. The total number of points classified as these classes was only five of 120, or approximately 4%, illustrating the need to account for pixels that contain mixtures. A second classification was conducted with the two component class signatures plus the three simulated mixture class signatures. The number of classified points increased to 112, or approximately 93%, with implementation of the mixture-class signatures.

The total amount of each component class in the plots was determined by summing the number of points classified as a class and multiplying by the proportion of each component class. A comparison of the percent cover estimates for Forest for each plot as determined by the dot-grid method and the classification map are as follows:

PLOT	GRID COVER ESTIMATE(%)	CLASSIFICATION COVER ESTIMATE (%)
1	55.0	63.75
2	65.0	71.25
3	60.0	71.25
4	40.0	48.75
5	20.0	28.75
6	30.0	37.50

If the dot-grid cover estimates can be assumed accurate then the amount of Forest occuring in the six plots was over-estimated by 11.5%.

EFFECT OF THE ATMOSPHERE ON SPECTRAL REFLECTANCE

Remote Sensing of Target Signatures

The signature of a target, as viewed from a satellite remote sensor (LANDSAT, Skylab EREP), is a radiance signature, N_s, given by

$$N_s = \frac{\rho H T}{\pi} + N_p \qquad (10)$$

where H is the total (180°, global) incoming solar irradiance, T is the beam transmittance (either the reflected target beam or the solar beam) of the atmosphere, N_p is path (between the satellite sensor and the target) radiance introduced by atmospheric scattering of sunlight, and ρ is the target reflectivity. This target reflectivity, ρ, is the true signature of the target. As can be seen in equation (10), the target's radiance signature is made up of three parameters other than the true target reflectivity signature. These parameters are the total incoming solar irradiance-H, the atmospheric beam transmittance-T, and the atmospheric path radiance-Np. They will be discussed in the following section, where all three will be referred to as atmospheric effects. The wavelength region considered in this paper is the solar reflective region from 400 to 1300 nanometers(nm). This region includes the LANDSAT bands (500-600, 600-700, 700-800, and 800-1100 nm) and most of the Skylab EREP S-192 scanner bands.

Because the target radiance signature is dependent upon atmospheric effects, one goal of this program is to correct the satellite data for these atmospheric effects. Comparisons can then be made concerning what improvements can be realized in automatic terrain mapping by computer by using algorithmus for correcting for atmospheric effects.

It is not within the scope of this paper to discuss each of these parameters in great detail in terms of atmospheric physics (radiative transfer); instead, the following subsections discuss them in sufficient detail to relate them to remote sensing.

Atmospheric Beam Transmittance

Atmospheric beam transmittance, T, can be given by the following expression (Beers, Lambert's, or Bouguer's law)

$$T = e^{-\tau_1 p} \qquad (11)$$

where lp is the path length, relative to the vertical path length of the beam, and τ is the atmospheric optical depth. The atmospheric path length is given by sec θ_o or sec θ, where θ_o is the solar zenith angle, sec θ_o is the path length of the solar beam, θ is the nadir angle of observation, and sec θ is the path length of the target reflected beam (see fig. 27). At low sun angles or oblique observation angles, θ_o or $\theta > 70°$, the path length needs to be corrected for atmospheric refraction (ref. 21). The atmospheric optical depth is wavelength - dependent (λ) and made up of three components: molecular (Rayleigh) scattering, τ_m; particle (Mie) scattering, τ_p; and, selective gaseous absorption, τ_a. Hence, the total optical depth is given as

$$\tau = \tau_m + \tau_p + \tau_a \tag{12}$$

The molecular-scattering optical-depth is proportional to λ^{-4}. This accounts for the blue sky, since the short wavelengths are scattered much more than the longer wavelengths. The wavelength-dependence of particle scattering is proportional to $\lambda^{-\alpha}$, where α is dependent upon particle size. For extremely small particles, α approaches 4.0 (as in Rayleigh scattering) and for large particles (clouds, fog, etc.) α approaches 0.0, which indicates gray scattering. The gaseous absorption optical depth is highly wavelength-selective, depending on the location of the various absorption bands of atmospheric constituents. Figure 28 illustrates the general nature of atmospheric beam transmittance of solar radiation.

Total Incoming Solar Irradiance

The total solar irradiance, H, is the sum of the direct-beam solar irradiance, on a horizontal surface, and the diffuse skylight, S; hence,

$$H = I \cos\theta_o + S \tag{13}$$

where I is the normal incident direct beam. This normal incident direct beam can be given, in terms of atmospheric effects, as

$$I = I_o e^{-\tau \sec\theta_o} \tag{14}$$

where I_o is the extraterrestrial solar irradiance, commonly referred to as the "solar constant". For "clear" days the diffuse skylight is normally blueish, which is indicative of the dominance of molecular (Rayleigh) scattering. However, the magnitude and spectral distribution of the diffuse skylight is a complex function of particle size, density, solar evelation and azimuth, optical depth, ground albedo, particle scattering phase function, particle albedo, and index of refraction. For representative measurements of total and diffuse irradiance see reference 22.

The diffuse skylight is also strongly influenced by clouds. Generally, clouds increase the skylight and change its spectral distribution from blue to white. Clouds can also strongly reflect solar radiation and cause "bright spots" (as opposed to shadows) on the ground, which can very significantly affect remote sensing of ground targets (ref. 23).

The total incident solar irradiance is very strongly influenced by the solar zenith angle cos θ_o (equation 13). Since the solar zenith angle is that angle between the target's normal and its planar surface, topography plays a significant role. For example mountainous terrane has very complex topography comprising north-, south-, east-, and west-facing slopes of various inclinations.

Atmospheric Path Radiance

Atmospheric path radiance results from molecules and large particles scattering sunlight in the field of view of the sensor. The magnitude and spectral distribution of path radiance is determined by all the variables that were given for diffuse skylight.

Atmospheric path radiance presents a special problem because it cannot be directly measured with ground-based instruments. Currently there are three techniques for deriving path radiance:

(1) Calculations using radiative transfer models
(2) indirect ground-based measurements
(3) combination of model calculations and ground-based measurements

Radiative-transfer models have been the subject of considerable research. A review and discussion of such models is given in reference 12. Models are useful for parametric studies of the influence of the atmosphere, given hypothetical situations. However, for specific conditions they are difficult to apply because, without measurements, one simply cannot confidently choose a "representative" atmospheric state. This is due to the fact that atmospheric aerosols and water vapor are known to be extremely variable as a function of geographical locations, altitude, and local conditions.

Indirect ground-based measurements of path radiance have been employed by Duntley (ref. 24) and Rogers, (ref. 25). This technique uses ground-based measurements of downward scattered sky radiance at a geometry (sun-sensor) similar to the satellite-sun-target geometry. The measured sky radiance is then extrapolated to yield a path radiance for the satellite sensor (see ref.'s 24 and 25).

The third technique was used recently by Hulstrom (ref. 22) to analyze the performance of the Skylab EREP sensors (S-190 A, S-191, and S-192). This technique uses ground-based measurements of optical depth/transmittance, target reflectivity, ground albedo, and total solar irradiance (see equation 10); then, the optical depth and ground albedo measurements are used as inputs for a radiative-transfer computer model that calculates path radiance. This technique eliminates uncertainties in atmospheric state, optical depth, and total (direct plus diffuse) irradiance. However, values for the particle scattering, phase function, and albedo have to be assumed.

A fourth technique of deriving path radiance is being attempted in this program. This technique employs the satellite data itself and helicopter measurements of selected ground targets. If one plots equation 10, as shown in Figure 29, it can be seen that the y intercept is the path radiance and the slope is equal to the product HT/π . The natural target reflectivities are measured from a helicopter platform. This is performed by having nearly identical spectral radiometers measure the target radiance, N_t , from the helicopter; and, the other radiometer measure the total incoming solar irradiance, H. The target reflectivity can then be determined by

$$\rho_t = \frac{\pi N_t}{H} \quad \text{(assumes Lambertian target)} \qquad (15)$$

Hence, one can determine the path radiance and the product HT/π . Therefore the reflectivity signatures of other (unknown) targets can be derived by

$$\rho = (N_s - N_p) \times \pi/HT \qquad (16)$$

The advantages of this technique are

(1) it measures path radiance as "seen" from the satellite sensor

(2) it measures the combined atmospheric effects of total solar irradiance, H, and atmospheric transmittance, T.

(3) if the selected natural ground target reflectivities remain nearly constant over periods of time (as was found to be valid for this test site) numerous simultaneous ground-based measurements are not required in order to calibrate atmospheric effects.

The possible short-comings of this technique are

(1) It assumes fairly uniform atmospheric conditions of the surface area within which the natural target "calibration" sites are located.

(2) The accuracy of the technique depends on the accuracy and repeatibility of the measurements of natural target reflectivity, and radiances measured by sensors in the satellite.

(3) It assumes that the sun angle with respect to the targets, both natural "calibration and unknown targets, are similar. This is apparent when one considers the sun angle dependence, I cos θ_o of the total solar irradiance and the product HT/π .

Initially, during the summer and fall of 1975, this technique will be employed for LANDSAT data.

Four techniques of deriving atmospheric path radiance and atmospheric effects have been presented. Each has it's advantages and short-comings. Unfortunately almost no data or experiments exist for checking and comparing the accuracy and validity of these techniques. All four techniques will be attempted during this program, and such comparisons subsequently made.

FIELD MEASUREMENTS OF ATMOSPHERIC EFFECTS AND TARGET REFLECTIVITY

Brief descriptions of the techniques and instrumentation used to measure atmospheric effects and target reflectivity are given. For a more detailed description see Hulstrom (ref. 22).

Atmospheric Effects

Atmospheric effects that are directly measureable by ground-based instruments are the total solar irradiance, H, and the beam transmittance, T.

The total solar irradiance was measured with an I.S.C.O. (Instrumentation Specialties Company) model SR (spectroradiometer) equipped with a SRR (spectroradiometer recorder) unit. This instrument consists of the SR and SRR units and a 6 ft fiber-optics bundle that has a flat teflon diffuser, 180° F.O.V., aperture at the end of the bundle. Light travels through the diffusser, through the fiber-optics bundle into the SR unit, where it is passed through a chopper, slit, and monochromator. The monochromator is a wedge interference filter. The wavelength of the instrument is varied by lengthwise movement of the wedge interference filter between the slit and the sensor. The wavelength region from 380 to 750 nm is measured with a silicon-junction photocell; and, the region from 750 to 1350 nm is measured with a germanium junction photocell. The 1/2 bandwidth is 15 nm from 380 to 750 nm, and 30 nm from 750 to 1350 nm. The total solar irradiance was measured by positioning the flat teflon diffuser so that it was horizontal/level. Hence, the 180°F.O.V. diffuser was exposed to the total irradiance (direct + diffuse). The diffuse skylight can be measured by simply shading the diffuser from the direct sunlight. A complete spectral scan, from 380 to 1350 nm, requires about 3 minutes. During this time interval, the total broadband irradiance (400-1100 nm) was continuously monitored with a Y.S.I. (Yellow Springs Instrument Co.) model 68 pyranometer. This was done in order to assure that stable conditions existed during the I.S.C.O. spectral scan.

The complete I.S.C.O. unit was calibrated with an I.S.C.O. S.R.C. (Spectroradiometer Calibrator) unit, which uses a ribbon-filament tungsten lamp, and a N.B.S., coil filament type lamp (1963 standard No. E.P.I. 1523). These standards were used to derive an absolute calibration factor, C.F., for the I.S.C.O. unit. This is derived by viewing the standard irradiance, I_{STD}, and recording a raw intensity I_r. The calibration factor is then

$$C.F. = I_{STD}/I_r \qquad (17)$$

The raw field measurement, I_r, is then multiplied by the calibration factor in order to obtain absolute quantities. However, this calibration factor is only valid for measurements of normal-incident (to the diffuser) irradiance. For measurements of total solar irradiance, where the sun is at various zenith angles (angles off of the normal to the diffuser), the cosine response of the diffuser has to be considered. This is evident from equation 13, i.e., I cos Θ_o. If a diffuser has a true cosine response, the intensity of a constant parallel source as measured at a 60° angle to the source, I (60°), would be 1/2 (cos 60°) that measured on normal incidence (0°). In more general terms,

$$I(\phi) = I(\phi = o)\cos\phi \tag{18}$$

where ϕ is the angle between the source and the normal to the diffuser. In the case of solar irradiance, $\phi = \Theta_o$. In order to determine the actual cosine response of the I.S.C.O. diffuser, extensive measurements of collimated light were made over a range of incident angles. This resulted in a cosine correction factor, C.C.F.,

$$C.C.F. = \frac{I(\phi=o)\cos\phi}{I(\phi)} \tag{19}$$

Examples of the cosine correction factors are shown in Figure 30. As can be seen the cosine response in the visible region is fairly true, while that in the infrared deviates by as much as 500%. The absolute total solar irradiance, H_{ABS}, is obtained by

$$H_{ABS} = (H_{raw})\,(C.F.)\,(C.C.F.) \tag{20}$$

Investigators have used the I.S.C.O. unit to measure target spectral albedo by pointing the diffuser up to measure the total incoming irradiance, $H\downarrow$, then pointing the diffuser down to measure the total outgoing irradiance, $H\uparrow$. The target's global albedo, p (180°),

$$p\,(180°) = \frac{H\uparrow}{H\downarrow} \tag{21}$$

If the diffuser's cosine response is not considered in this calculation, large errors (up to 500%) will result, especially in the infrared wavelengths, because the incoming irradiance has to be corrected for the cosine response (because the incoming irradiance is directional), while the outgoing irradiance is generally not directional, requiring no cosine correction.

In addition to the I.S.C.O. unit, Exotech Model 100 radiometers are also used for making total solar irradiance measurements in the specific LANDSAT wavelength bands. These units are calibrated in the same manner as discussed for the I.S.C.O. unit. At the time of this writing, no cosine response has been determined for these units.

The beam transmittance of the atmosphere was/is determined by using the Langley method. This method determines atmospheric optical depth by peforming several measurements of the solar direct-beam irradiance as it varies with relative air mass. Equation 14 can be re-written to give

$$\ln I = -\tau \sec\Theta_o + \ln I_o \tag{22}$$

In order to derive τ, a plot of the various measurements of I versus relative airmass-sec Θ_o is made. The y-intercept is I_o, hence τ can be calculated knowing I, I_o, and sec Θ_o. If the meter reading of the instrument, M, is linear with respect to I, then I can be replaced with M and I_o by M_o. M_o is the extraterrestrial meter reading of the instrument. For the measurements of direct solar irradiance, the ISCO unit was equipped with a 3° collimator. This resulted in beam transmittance for 27 wavelength measurements from 400 to 1300 nm. An example is shown in figure 31, where the four transmittance curves are for four sites located within and near the Colorado test site. Simultaneous measurements of beam transmittance reveal the uniformity of the atmosphere over the region. For the day shown in Figure 31, the atmosphere was fairly uniform.

Target Reflectivity

Both the I.S.C.O. and EXOTECH units were used to measure target reflectivity. The I.S.C.O. is modified with a 30° F.O.V. baffled collimator.

For ground measurements of target reflectivity, standard "gray" and standard "white" reflectance panels are used to determine target reflectivity. These panels are accurately calibrated with a Beckman DK-2A/Gier-Dunkle Integrating Sphere. Field measurements consist of the radiance of the "gray" card, N_{gc}, the radiance of the target, N_t, and the radiance of the "white" card, N_{wc}. The target reflectivity can then be derived by

$$\rho_t = \frac{N_t}{N_{gc}} \times \rho_{gc} \tag{23}$$

and

$$\rho_t = \frac{N_t}{N_{wc}} \times \rho_{wc} \tag{24}$$

where ρ_{gc} and ρ_{wc} are the reflectivities of the "gray" and "white" cards. The two derived target reflectivities are then averaged. This technique was verified by plotting radiance, N_{gc} and N_{wc}, versus reflectivity, ρ_{gc} and ρ_{wc}, and extrapolating this plot to zero radiance and reflectivity. If the card calibrations are accurate and if the field measurements are accurate, the plot should go through the origin. Several of these verifications revealed the technique and measurements to be consistent to within 1 to 2% reflectivity. Examples of absolute reflectivity for 15 targets and their backgrounds are listed below:

Target Description	Identification Number
Pikes Peak Granite	1
Lichen Rock Cover	2
Soil/Grass Background for Pikes Peak Granite	3
Cripple Creek Granite	4
Soil/Grass Background for Cripple Creek Granite	5
Volcanic Fine-Grain Mafic	6
Soil/Grass Background for Volcanic Mafic	7
Meadow Grass	8
Volcanic Crystal Tuff	9
Yellow Grass Background	10
Volcanic Andesite	11
Background Soil for Volcanic Andesite	12
Background Grass for Volcanic Andesite	13
Background Soil at Cameron Mountain	14
Background Soil/Grass at Cameron Mountain	15

Target No.	Percent Reflectance			
	0.5 to 0.6	0.6 to 0.7	0.7 to 0.8	0.8 to 1.1 m
1	14.0	18.4	22.6	25.6
2	23.6	28.5	36.1	35.2
3	12.7	14.8	20.7	23.4
4	17.3	22.6	24.5	25.6
5	11.1	13.6	19.3	22.6
6	11.7	14.5	19.1	19.4
7	10.8	16.8	19.5	26.0
8	5.8	5.4	30.8	42.5
9	27.5	31.8	35.8	40.4
10	11.8	13.6	19.1	18.9
11	14.5	17.7	19.8	20.9
12	9.3	13.3	18.0	18.7
13	6.5	8.5	12.7	17.3
14	10.8	16.3	19.2	23.2
15	9.6	12.8	19.8	24.1

In order to measure signatures from mixtures and for areas that are on the same order of size as the satellite sensors, helicopter measurements are employed. Helicopter measurements consist of using one EXOTECH radiometer, with a 1° FOV, in the helicopter to measure the target radiance, N_t and one EXOTECH radiometer located on the ground to measure total incoming solar irradiance, H. The reflectivity, ρ is then

$$\rho = \frac{\pi N_t}{H} \tag{25}$$

In addition, the helicopter measurements of N_t will be compared with the computer simulations of the radiance signature.

SUMMARY AND CONCLUSIONS

Methods have been discussed for simulating spectral signatures of mixtures of two terrain classes for automatic analysis of LANDSAT-1 MSS data using on-site measurements, simulated LANDSAT data, and actual LANDSAT-1 data. Tests of the methods for estimating component-class spectral signatures and simulating two component-class mixture signatures with simulated MSS data indicated that improvements in classification results over conventional component class analysis are possible using these techniques. Applying these techniques to actual LANDSAT-1 MSS data of wildland areas showed an increase in classification information over conventional analysis, but no quantitative accuracy analysis could be made due to difficulties of estimating mixture proportions from the control or "ground truth" data.

It was found that misclassification of pixels as mixtures can occur when the simulated spectral signatures approximate those of component classes or other mixtures in the scene. There was also evidence of misclassifications due to more than two component classes existing within a single pixel. Within-class variability, slope and aspect variability, and the sensitivity of the scanner in detecting changes in mixture proportions may also contribute to degrading the classification performance using these methods.

These techniques may prove valuable in areas where vegetation masks the characteristic spectral response of the underlying geologic material, and future work should include an examination of this application. Emphasis should also be placed on methods of acquiring better "ground truth" information upon which the spectral signature estimation and simulation techniques are based.

Properties of the atmosphere play an important role in affecting the spectral radiance received at the satellite by the scanner. Some atmospheric properties can be measured on site. Others must be determined by comparing satellite data with the on-site measurements. Once their signatures have been determined, it may be possible for some large natural terrain features to serve as calibration panels from which atmospheric properties can be calculated by the satellite data. When the desired remote sensor product is an image having only rough spectral information, atmospheric effects can be ignored. However, when quantitative spectral information is required, atmospheric effects cannot be neglected.

REFERENCES

Horwitz, H. M., R. F. Nalepka, P. D. Hyde and J. P. Morgenstern, 1971, Estimating the Proportions of Objects Within a Single Resolution Element of a Multispectral Scanner: Proceedings of the Seventh International Symposium on Remote Sensing of the Environment. Ann Arbor, Michigan. p. 1307-1320.

Pace, W. H. and D. M. Detchmendy, 1972, A Fast Algorithm for the Decomposing of Multispectral Data into Mixtures: University of Tennessee Space Institute, Volume II, Tullahoma, Tennessee, p. 831-848.

Horwitz, H. M., P. D. Hyde and W. Richardson, 1974, Improvements in Estimating Proportions of Objects from Multispectral Data: Environmental Research Institute of Michigan. ERIM 190100-25-T.

Heller, R. C., R. C. Aldrich, R. S. Driscoll, R. E. Francis and F. P. Weber, 1974, Evaluation of ERTS-1 Data for Inventory of Forest and Range Land and Detection of Forest Stress: Final Technical Report to NASA. Contract No. S-70251-AG. 218 p.

Mead, R. A., 1974, Factors Affecting Computer Recognition of Ponderosa Pine from ERTS-1 Imagery: Master's Thesis, Dept. of Earth Resources, Colorado State University, Fort Collins, Colorado, 85 p.

Root, R. R., 1974, Computerized Terrain Mapping of Yellowstone National Park: Ph.D. dissertation, Dept. of Earth Resources, Colorado State University, Fort Collins, Colorado 242 p.

Root, R. R., H. W. Smedes, N.E.G. Roller, and D. Despain, 1974, Color Terrain Map of Yellowstone National Park, Computer-derived from ERTS MSS Data: Proc. Ninth International Symposium on Remote Sensing of Environment, Ann Arbor, Mich., v. II, p. 1369-1398.

Nalepka, R. F., H. M. Horwitz, and P. D. Hyde, 1972, Estimating Proportions of Objects From Multispectral Data: Environmental Research Institute of Michigan. ERIM 31650-73-T.

Malila, W. A., D. P. Rice, and R. C. Cicone, 1975, Final Report on the CITARS Effort by the Environmental Research Institute of Michigan: Report No. NASA-CR-ERIM 109600-12-F, NASA Scientific and Technical Reports Facility, College Park, Md., 127 p.

Smith, J. A., L. D. Miller and T. Ells, 1972, Pattern Recognition Routines for Graduate Training in the Automatic Analysis of Remote Sensing Imagery-RECOG: Colorado State University Science Series No. 3A, Fort Collins, Colorado, 85 p.

Smedes, H. W., 1971, Automatic Computer Mapping of Terrain: International Workshop on Earth Resources Survey Systems, Government Printing Office, Wash., D.C., v. II, p. 345-406.

Malila, W. A., R. B. Crane, C. A. Omarzu, and R. E. Turner, 1971, Studies of Spectral Discriminators: Report No. NASA-CR-WRL 31650-22-T, NASA Manned Spacecraft Center, Houston, Texas.

Nalepka, R. F. and P. D. Hyde, 1973, Estimating Crop Acreages From Space Simulated Multispectral Scanner Data: Environmental Research Institute of Michigan, ERIM 31650-148-T.

Ells, T., L. D. Miller and J. A. Smith, 1972a, Programmer's Manual for RECOG (Pattern RECOGnition Programs): Colorado State University Science Series No. 3C, Fort Collins, Colorado, 216 p.

_____ 1972b. User's Manual for RECOG (Pattern RECOGnition Programs): Colorado State University Science Series 3B. Fort Collins, Colorado, 85 p.

Maxwell, E. L. and G. R. Johnson, 1974, A Remote Rangeland Analysis System: United States Geologic Survey. Report No. 1885-F, Colorado State University, Fort Collins, Colorado.

Crane, R. B. and P. D. Hyde, 1972, Signature Estimation from Satellite Multispectral Scanner Data: Proceedings of the Eighth International Symposium on Remote Sensing of Environment. Ann Arbor, Michigan, p. 765-773.

Anuta, P. E., 1973, Geometric Corrections of ERTS-1 Digital Multispectral Scanner Data: The Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette Ind., 23 p.

Fleming, M. D., 1974, Analysis of USFS Site 226C Colorado: The Laboratory for Application of Remote Sensing (LARS), Purdue University, West Lafayette, Ind., 53 p.

Sharphack, D. A. and G. Akin, 1969, An Algorithm for Computing Slope and Aspect from Elevations Photogramm. Eng., v. 35, p. 247-248.

Bemporad, A., 1907, Saggio di Una Nuova Formola Empirica per Rappresentare il Modo di Zoriare Della Radiazione Solare Col Variare Dello Spessore Atmosferico Attrazersato Dai Raggi: Fisica Terrestre, in Rend. Accademia Lincei, Roma, Ser. 5, vol. 16, p. 66-71.

Hulstrom, R. L., 1974, Spectral Measurements and Analyses of Atmospheric Effects on Remote Sensor Data: Proc. Soc. Photo-Optical Engineer, v. 51, p. 90-100.

_____1973, The Cloud Bright Spot: Jour. Photogrammetric Engineering, v. 39, p. 370-376.

Gordon, Jacqueline I., Harris, James L. Sr., and Duntley, Seibert Q., 1973, Measuring Earth-to-Space Contrast Transmittance from Ground Stations: Applied Optics, June, p. 1317-1324.

Rogers, R. and K. Peacock, 1973, A Technique for Correcting ERTS Data for Solar and Atmospheric Effects: Symposium on Significant Results of ERTS-1, NASA Spec. Paper 5P-327, p. 1115-1122.

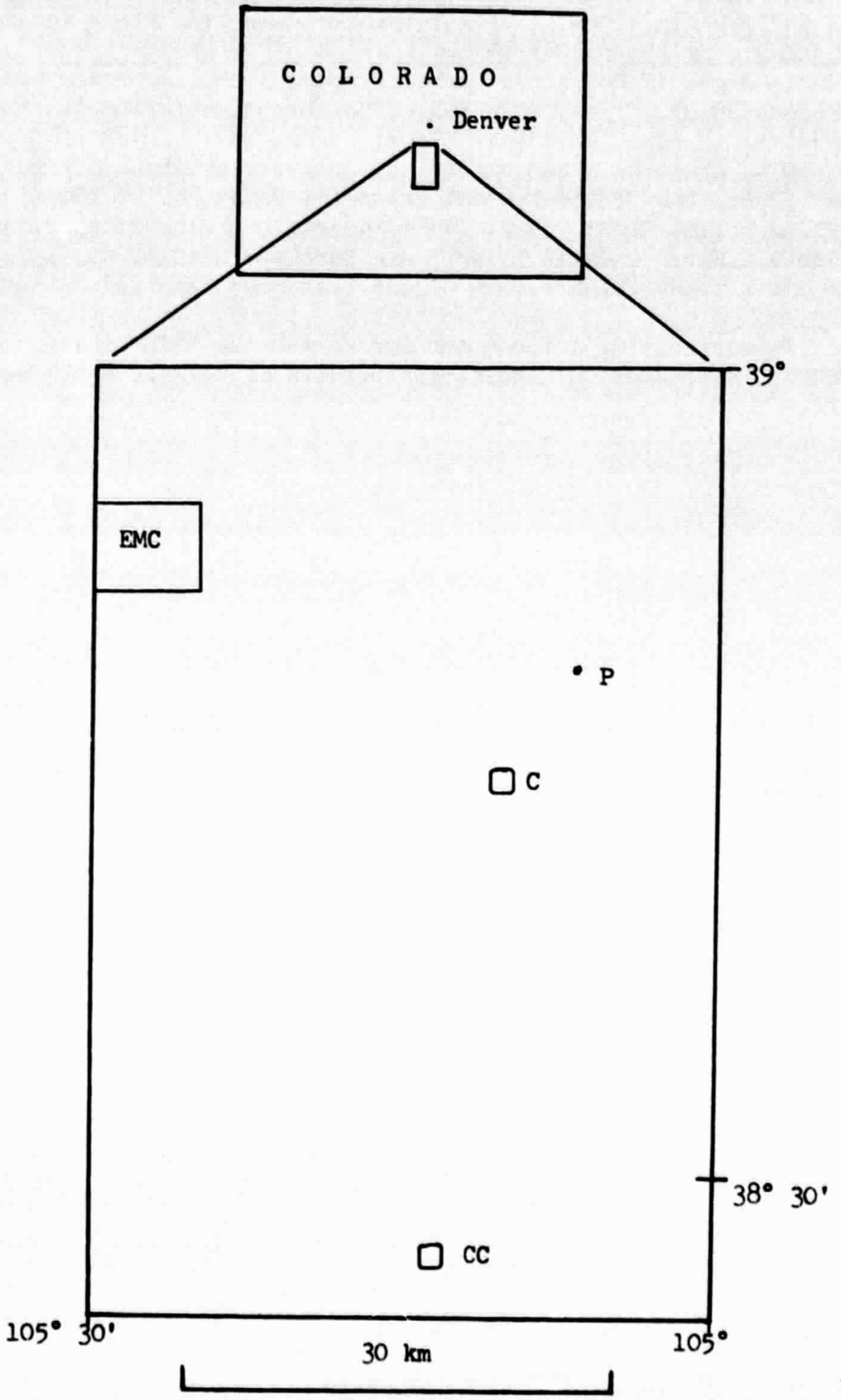


Figure 1.-Location of test site. EMC=Elevenmile Canyon Reservoir Study Area, P=Pikes Peak, C=Cripple Creek, CC=Canon City.

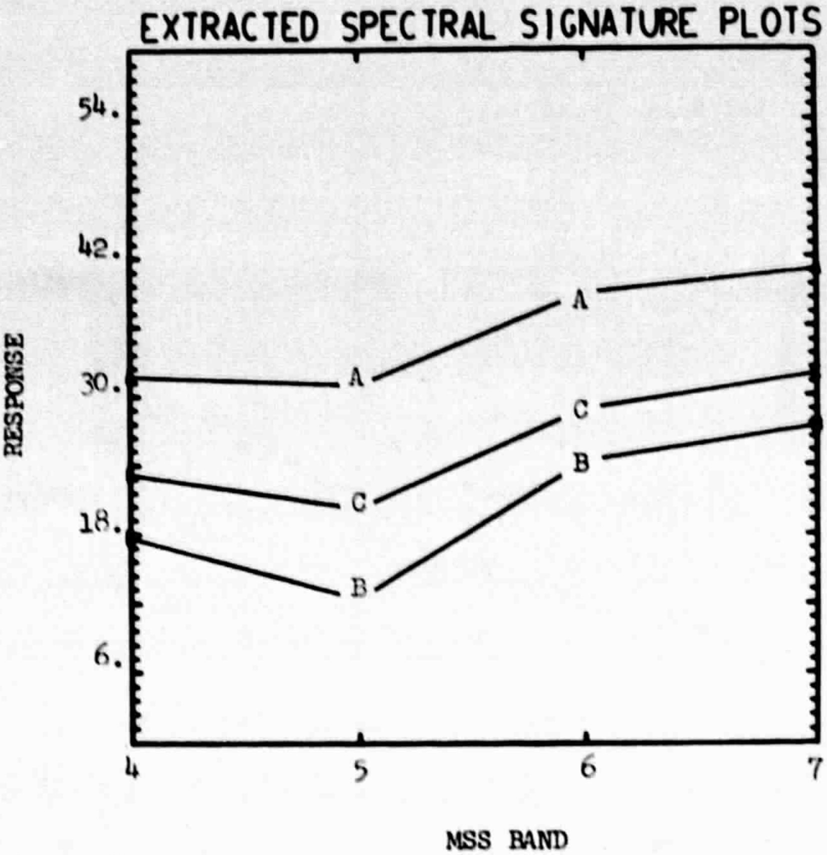


Figure 2.-Spectral response of two component terrain classes (A,B) and a mixture of them (C) extracted from LANDSAT-1 MSS data.

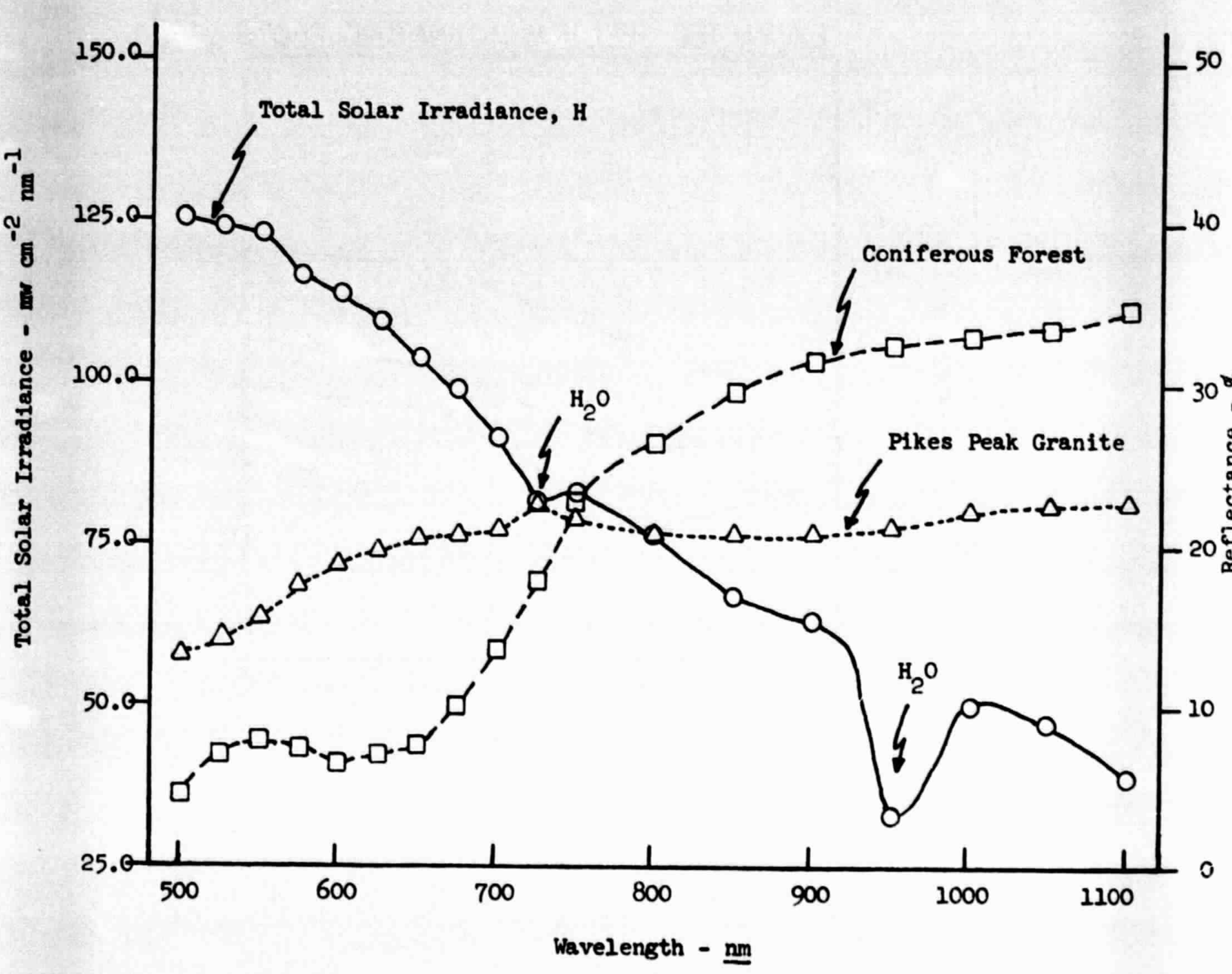


Figure 3.-Comparison of total solar irradiance and reflectivity of coniferous forest and Pikes Peak Granite.

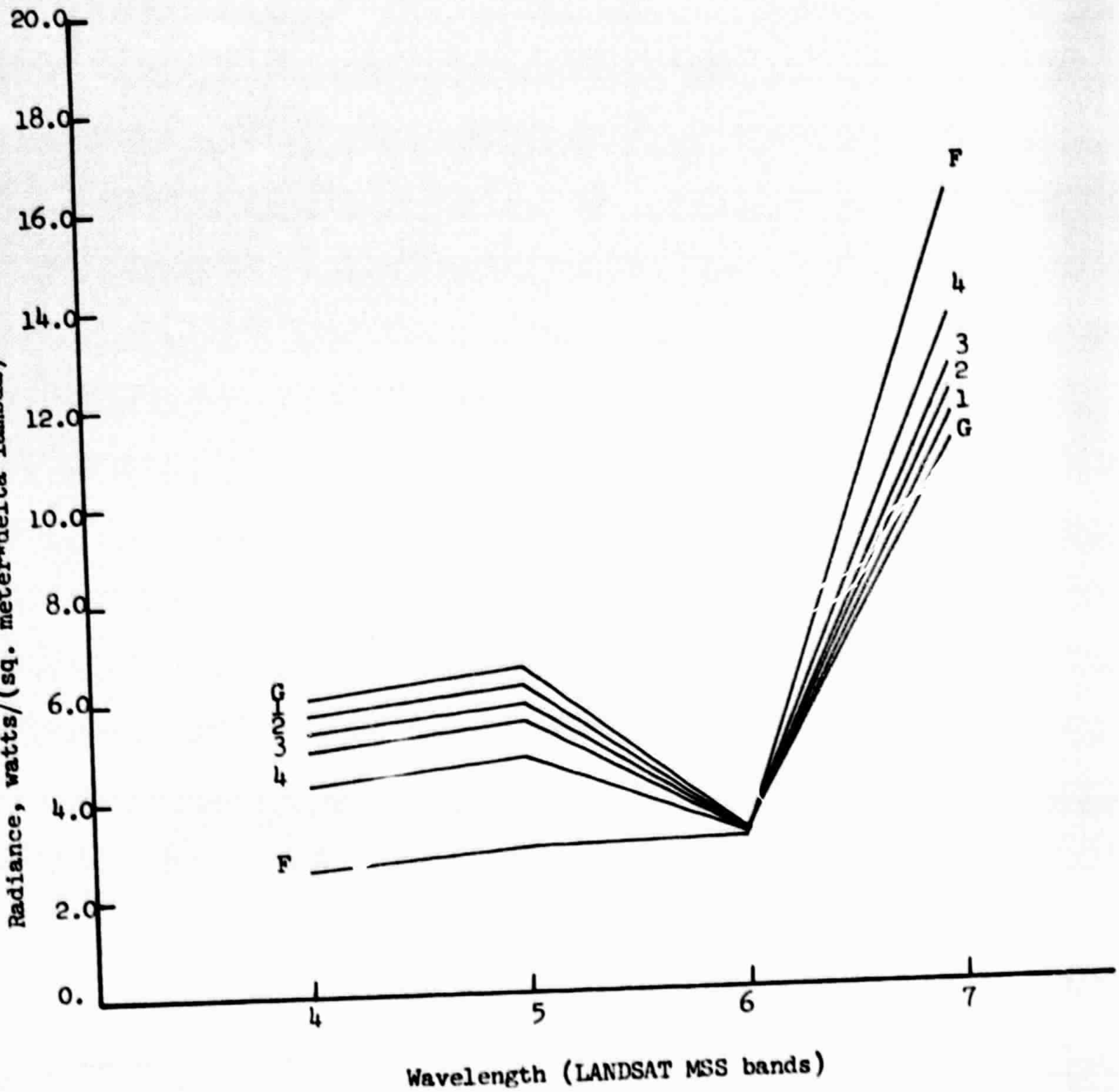


Figure 4.-Radiance signatures of Pikes Peak Granite, coniferous forest and mixtures of the two. Curves: G=100% Pikes Peak Granite, F=100% coniferous forest, 1=90% G 10% F, 2=80% G 20% F, 3=70% G 30% F, 4=50% G 50% F. Data from May 29, 1974; lat. 38.83° N, long. 105.26° W, solar elevation 64.51°, solar azimuth 126.62°, time 10:30 AM MST. Curves traced from computer printout.

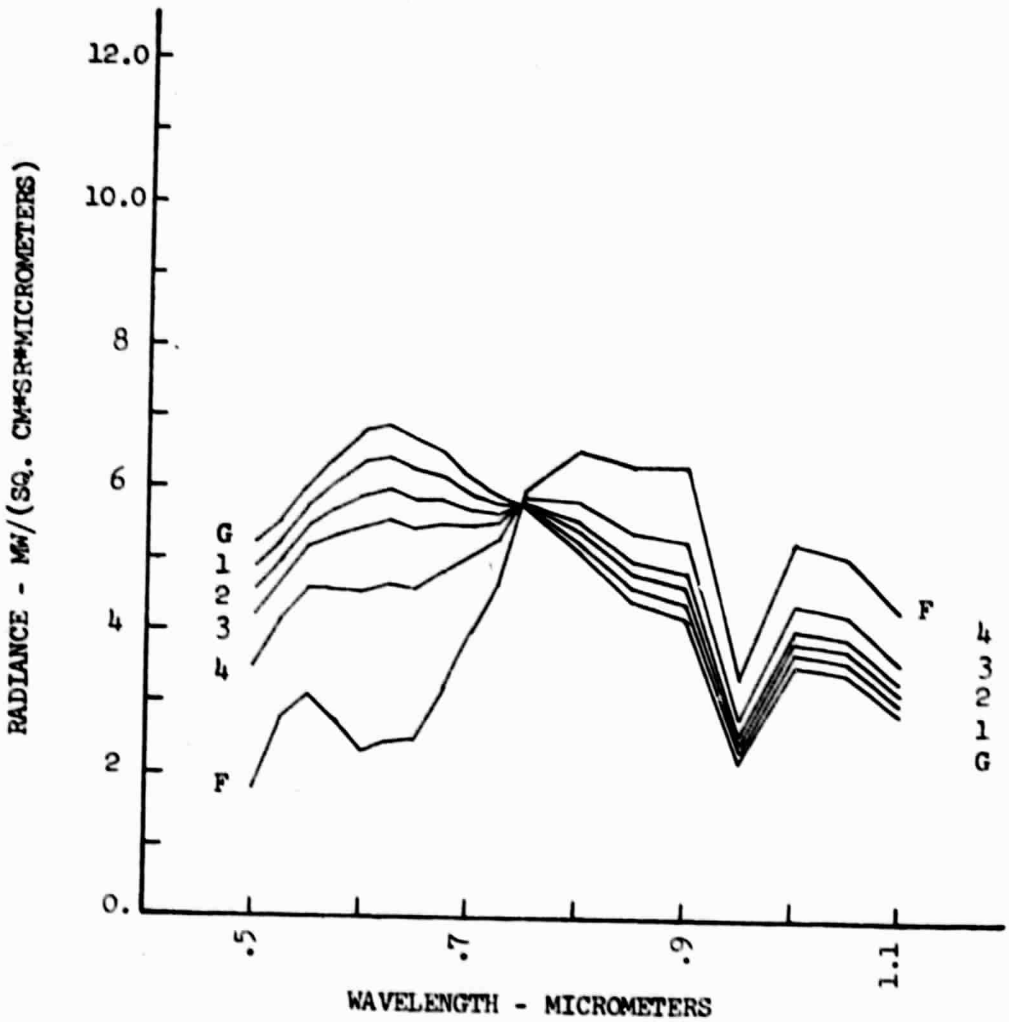
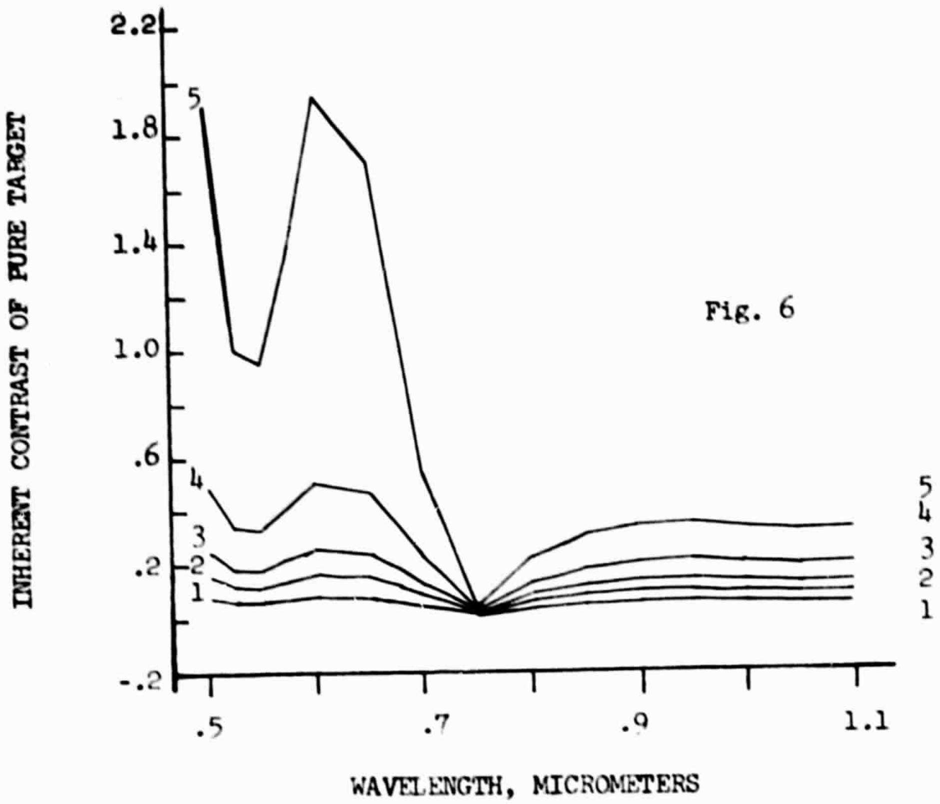
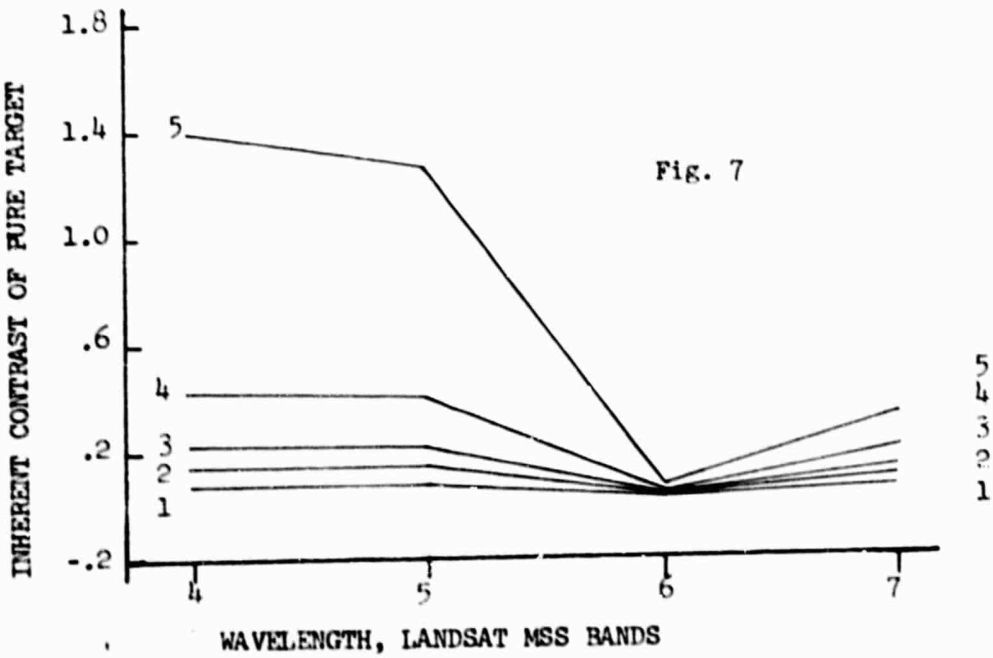


Figure 5.-Spectral radiance of Pikes Peak Granite, coniferous forest, and mixtures of the two. Curves, date, time, and other documentation same as in fig. 4.





Figures 6 and 7.-Spectral inherent contrast of Pikes Peak Granite (G) target with background of Coniferous Forest (F), and mixtures of granite and forest. Curves: 1=90% G 10% F, 2=80% G 20% F, 3=70% G 30% F, 4=50% G 50% F, 5=100% F.

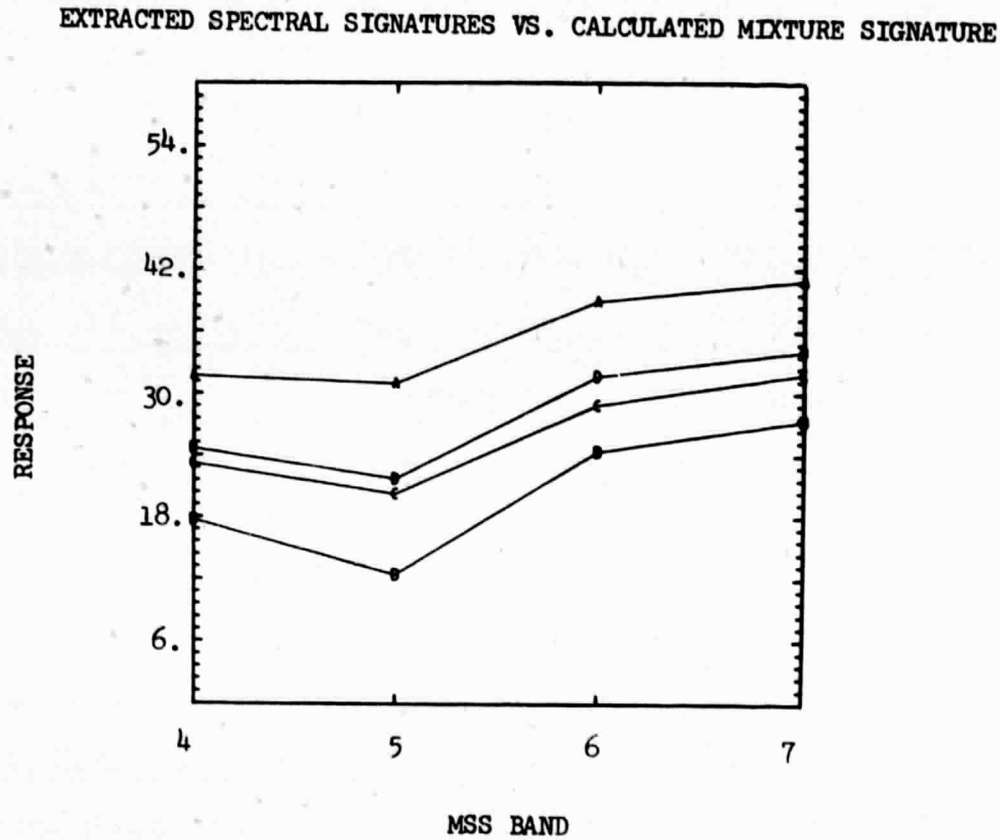


Figure 8.-Spectral response of two component terrain classes (A,B) and a mixture of them (C) extracted from LANDSAT-1 MSS data, and the mixture response (D) calculated using equation 4. A, B, and C, same data as in fig. 2.

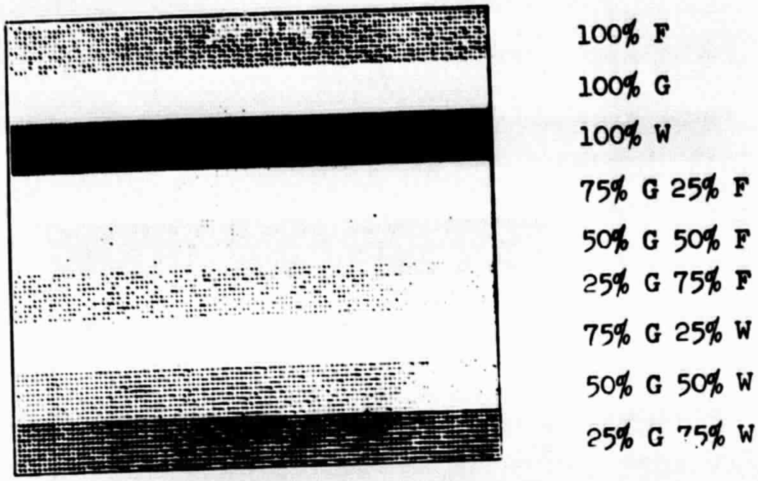


Figure 9.-Microfilm graymap of 9-class simulated data set. Each horizontal band represents a field of simulated data for a given class, as labelled. F=Coniferous forest, G=grassland, W=water.

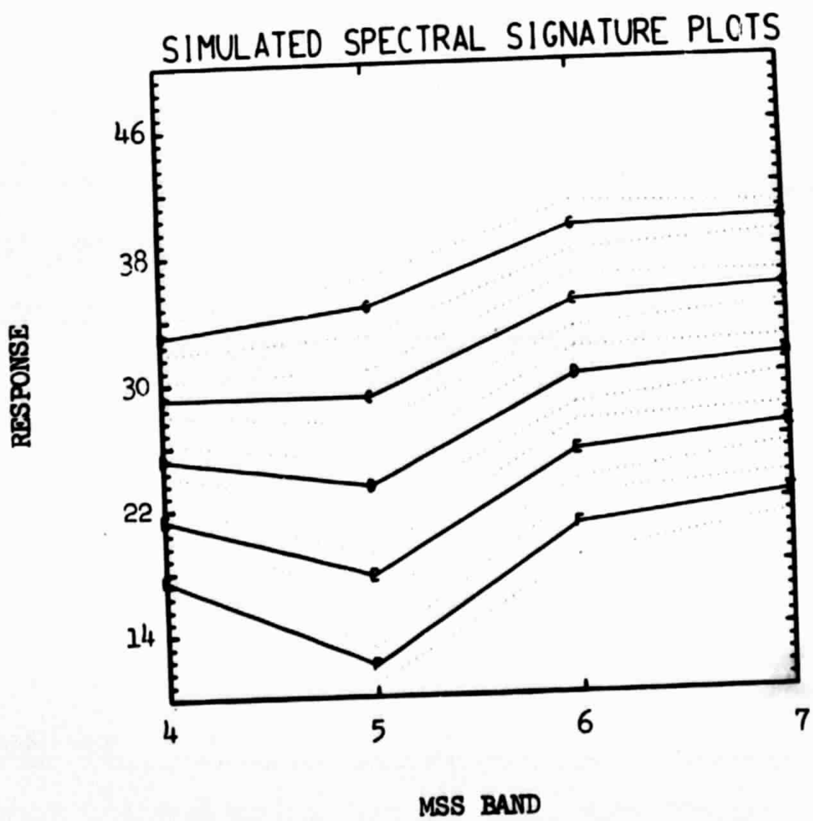


Figure 10.-Mean spectral response for two component classes and modeled mixture classes. Curves: G=100% grassland, F=100% forest, C=75% G 25% F, D=50% G 50% F, E=25% G 75% F. Dotted lines represent standard deviations.

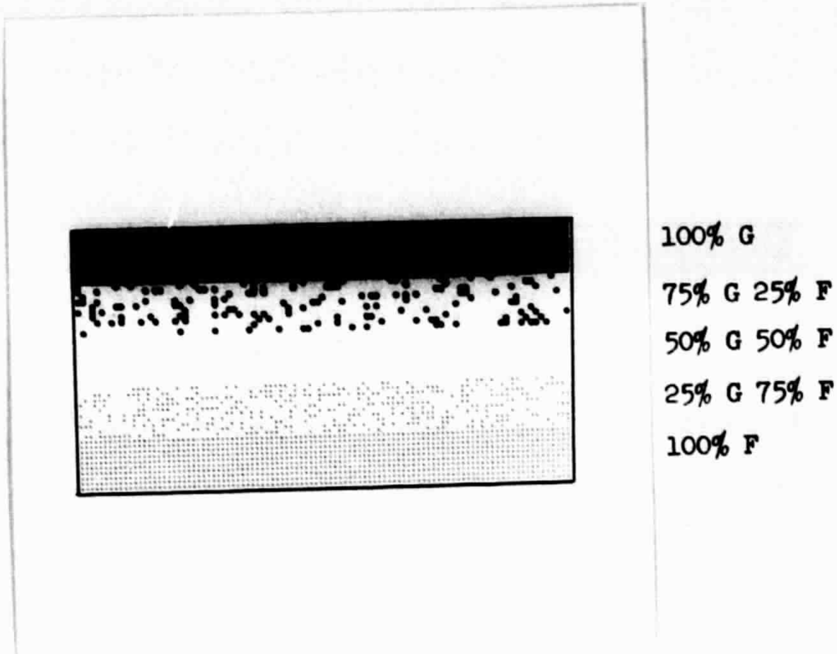


Figure 11.-Microfilm classification display for component-class analysis. Blanks denote unclassified points. Symbols are: • =grassland(G), .=forest(F). C,D,E same as in fig. 10.

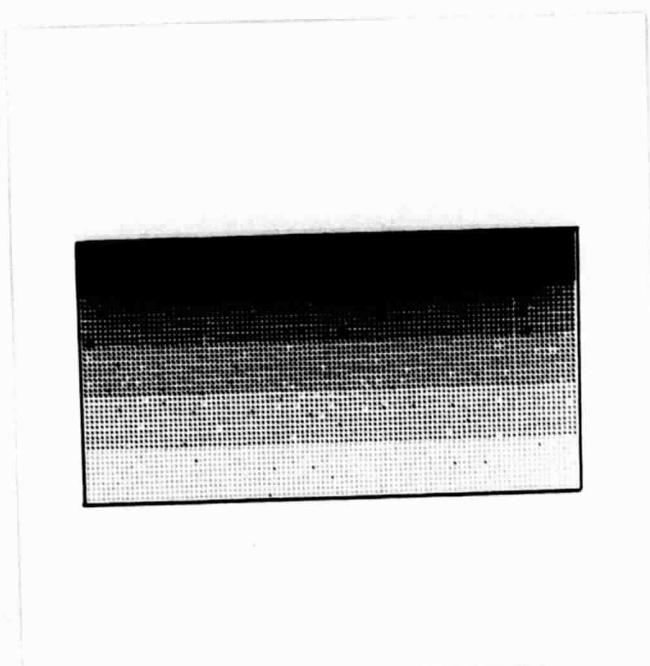


Figure 12.-Microfilm classification display for 5-class analysis. Symbols are: • =G, ▴=C, \=D, ⸜=E, .=F. Letters refer to classes and mixtures as in fig. 10.

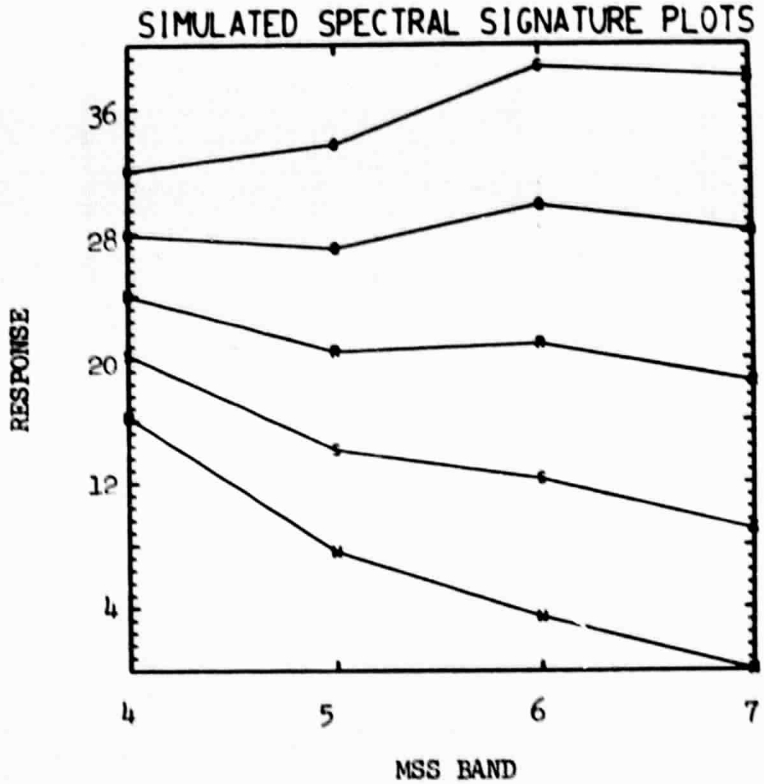


Figure 13.-Mean spectral response curves for simulated Grassland-Water mixtures. Dotted lines indicate standard deviations. Letter symbols are same as in matrix on p. 12.

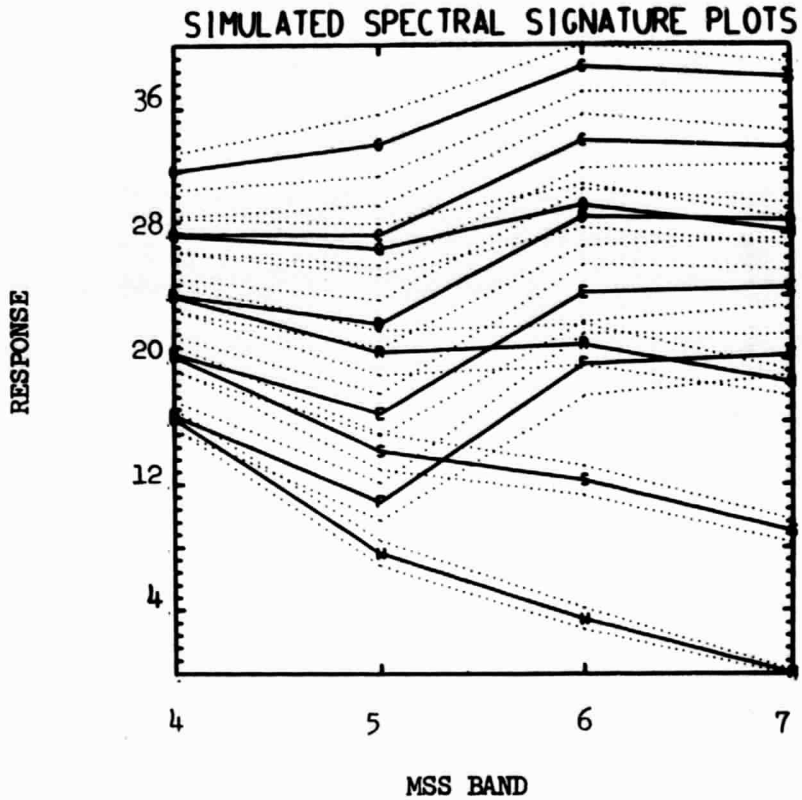


Figure 14.-Mean spectral response curves for component and mixture class signatures used for 9-class analysis. Dotted lines indicate standard deviations. Letter symbols are same as in matrix on p. 12.

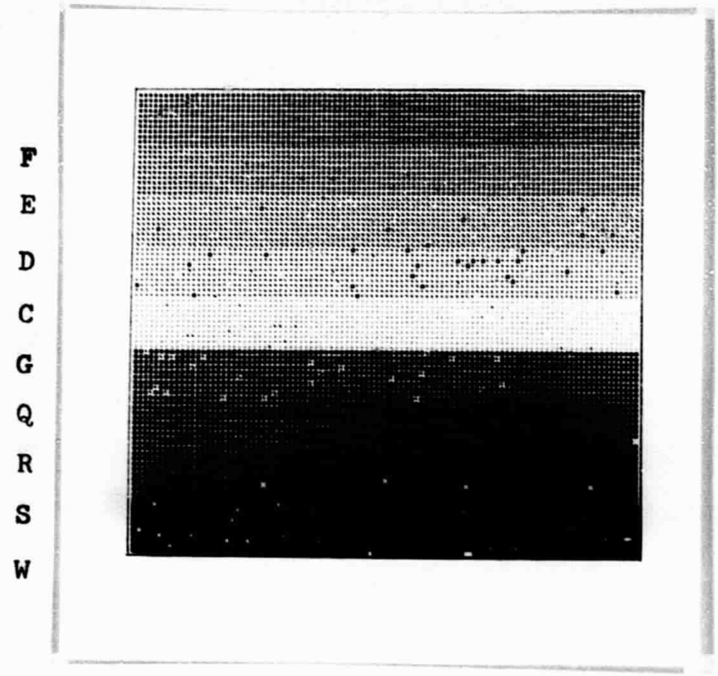


Figure 15.-Microfilm classification display of 9-class analysis. True class field symbols noted at left of display are same as in matrix on p. 12.

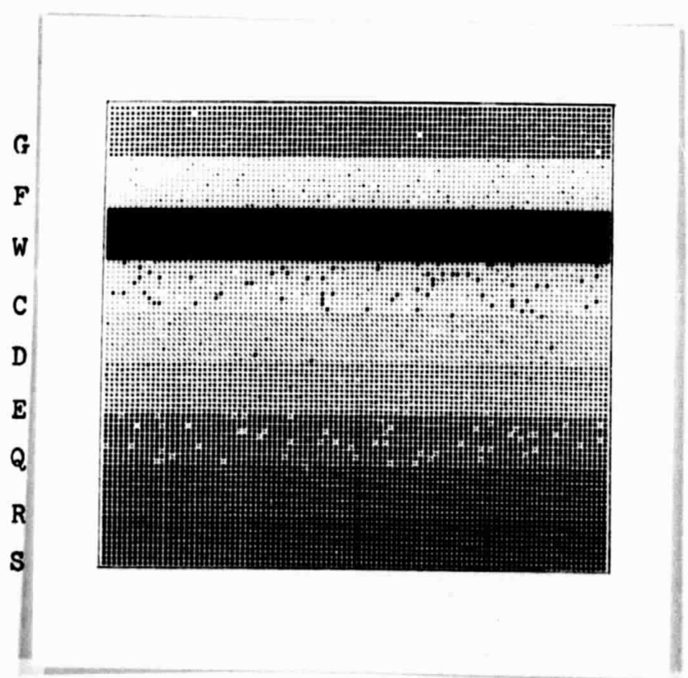
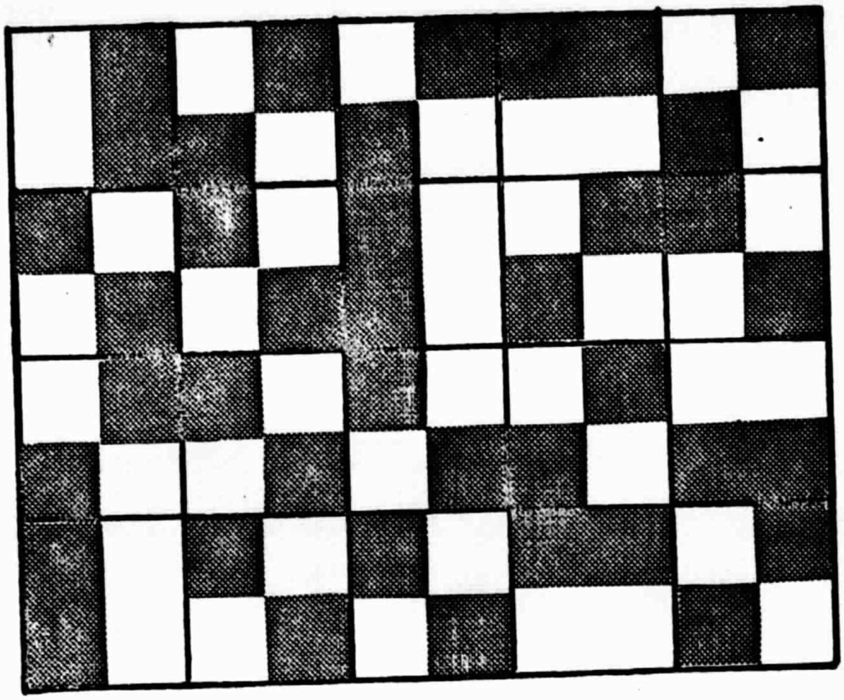


Figure 16.-Microfilm classification display for 9-class analysis with calculated component class signatures from uniform data. True class field symbols noted at left of display are same as in matrix on p. 12.

a.



b.

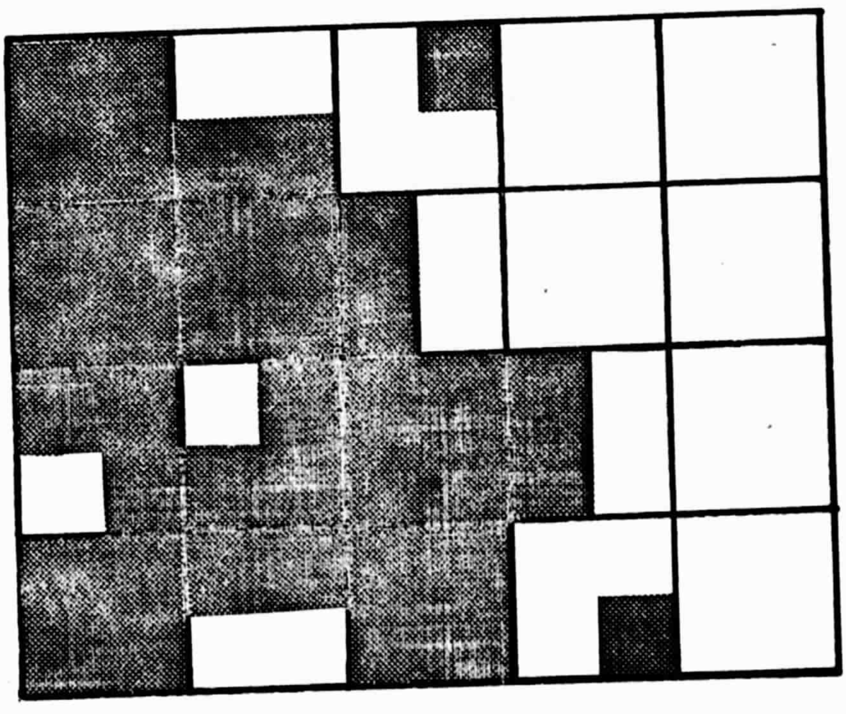
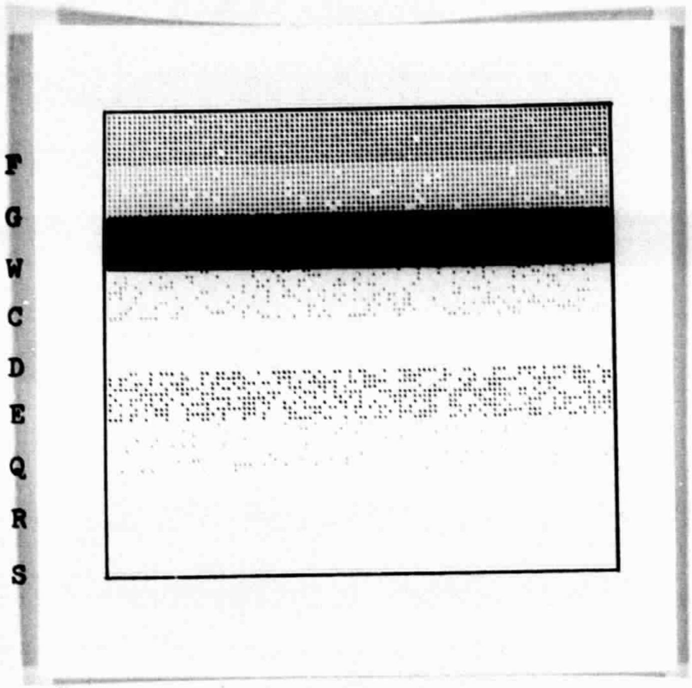


Figure 17.-Graphical representation of uniform and non-uniform training sets. a). Uniform mixture of 50% White - 50% Shaded. b). Non-uniform mixture of 50% White - 50% Shaded.

ORIGINAL PAGE IS
OF POOR QUALITY

a.



b.

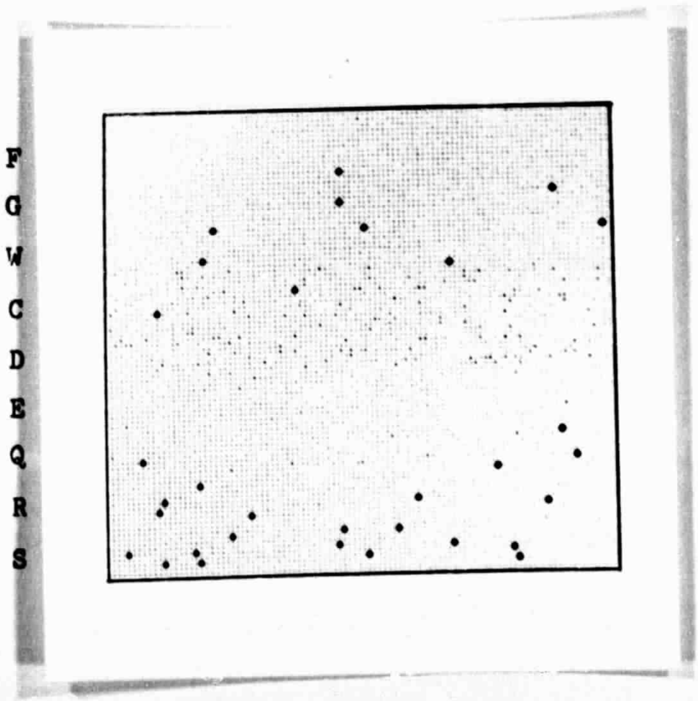


Figure 18.-Microfilm classification displays for component-class analysis using estimated mean vectors and covariance matrices. a) Spectral signatures estimated from uniform training sets. b) Spectral signatures estimated from non-uniform training sets. Symbols are: · =Grassland, \ =Forest, • =Water. True field classes noted at left of displays have same symbols as in matrix on p. 12.

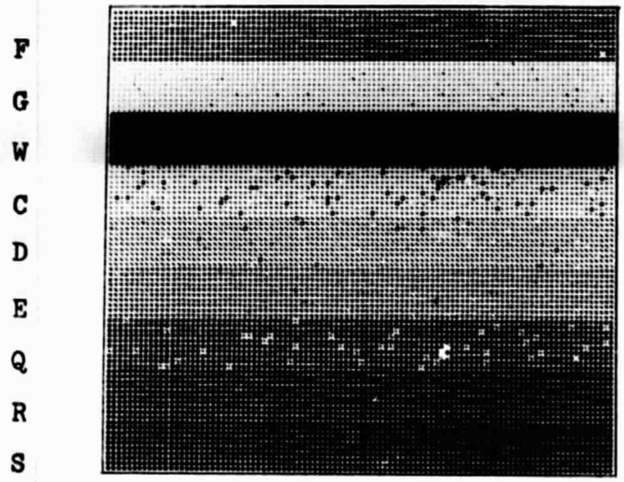


Figure 19.-Microfilm classification display of 9-class analysis using calculated component class mean vectors and common covariance matrix. Trueclass field symbols noted to left of display are same as in matrix on p. 12.

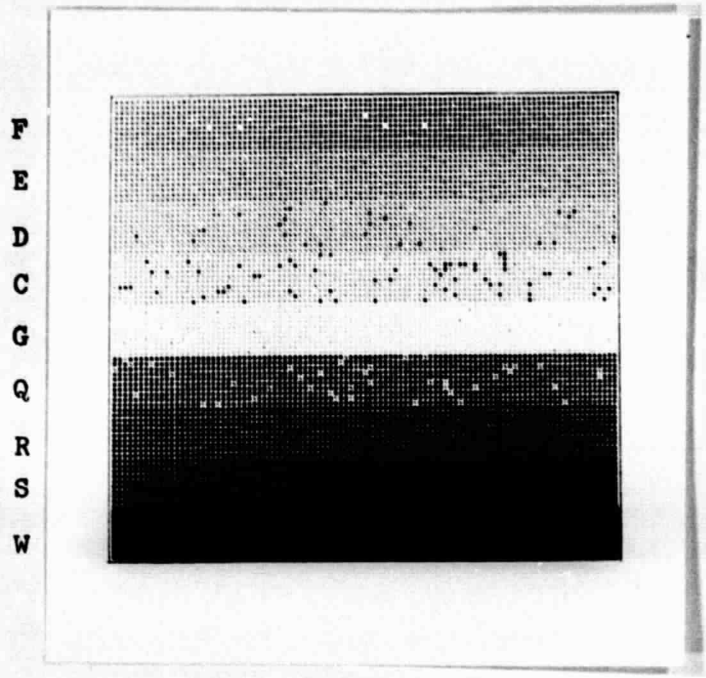


Figure 20.-Microfilm classification display of 9-class analysis using calculated mean vectors and common covariance matrix. True field class symbols noted at left of display are same as in matrix on p. 12.

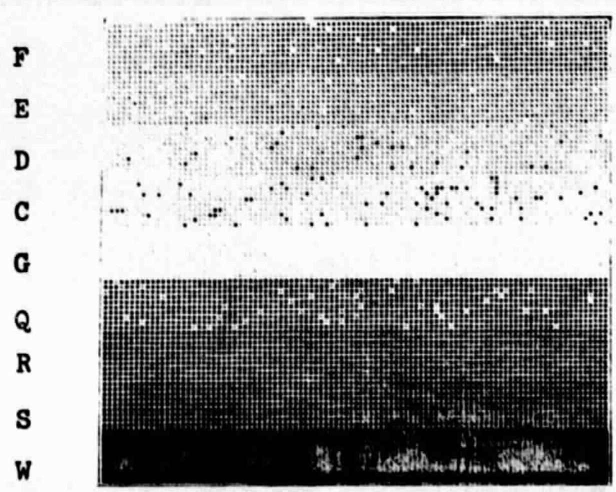


Figure 21.-Microfilm classification display for 9-class analysis using calculated mean vectors and averaged common covariance matrix. True class fields noted at left of display are same as in matrix on p. 12.

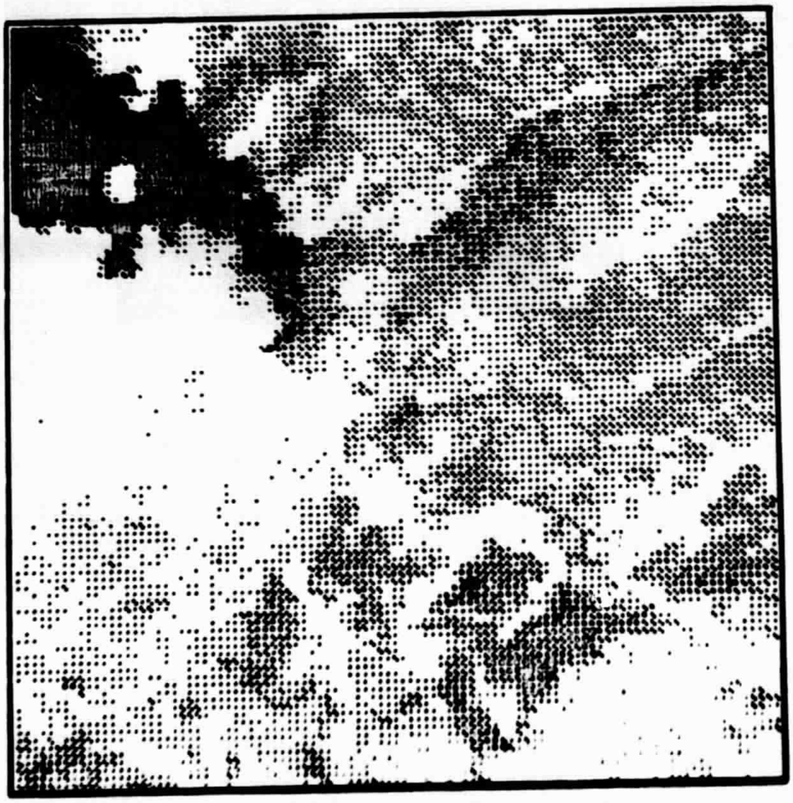


Figure 22.-Microfilm graymap of MSS Band 5 of the Elevenmile Canyon Reservoir Study Area.

ORIGINAL PAGE IS
OF POOR QUALITY

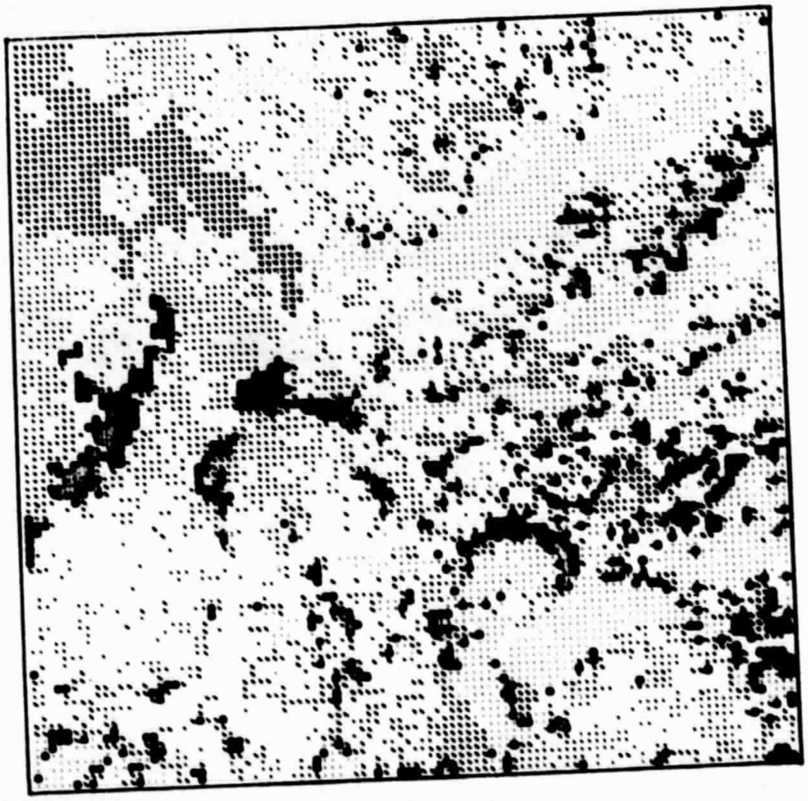


Figure 23.-Microfilm classification display of Elevenmile Canyon Reservoir Test Area with 5 Component Classes. Symbols are: . =Forest, ` =Grassland, ▼ =Water, \ =Mountain Grassland, • =Wet Meadow, Blank Unclassified.

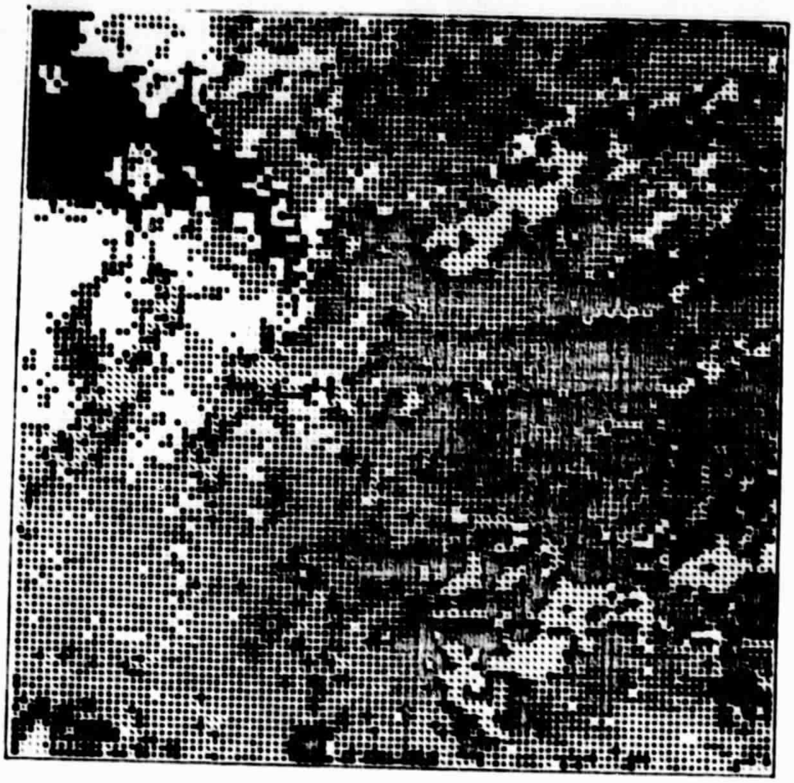


Figure 24.-Microfilm classification map of Elevenmile Canyon Reservoir study area with 5 component classes and 9 mixture classes. Because of limitations of the microfilm display, separate symbols could not be shown for the mixtures; however, for study purposes they were printed separately by a conventional line-printer.
Component classes:

' Grassland (G)
\ Forest (F)
⌉ Mountain Grassland (MG)
\ Wet Meadow (WM)
• Water (W)

Mixture classes:

X = 75% G 25% F, 50% G, 50% F, 25% G 75% F,
75% G 25% WM, 50% G 50% WM, 25% G 75% WM,
75% F 25% MG, 50% F 50% MG, 25% F 75% MG

Blank, unclassified

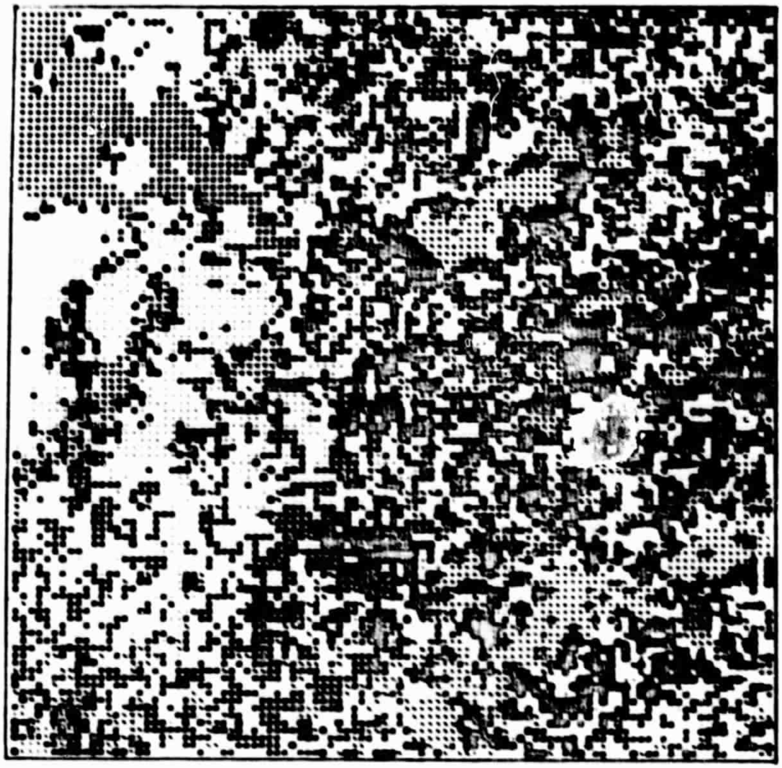


Figure 25.-Microfilm classification map of Elevenmile Canyon Reservoir study area with 5 component classes and 10 mixture classes. Because of limitations of the microfilm display, separate symbols could not be shown for the mixtures; however, for study purposes they were printed separately by a conventional line printer.
Component classes:

· Grassland (G)
\ Forest (F)
◣ Mountain Grassland (MG)
◥ Wet Meadow (WM)
x Water (W)

Mixture classes:

■ 67% G 33% F
● 67% G 33% W

X = 67% WM 33% W, 33% WM 67% W, 33% G 67% F,
33% G 67% W, 67% F 33% MG, 33% F 67% MG
67% G 67% WM, 33% G 67% WM

Blank, unclassified

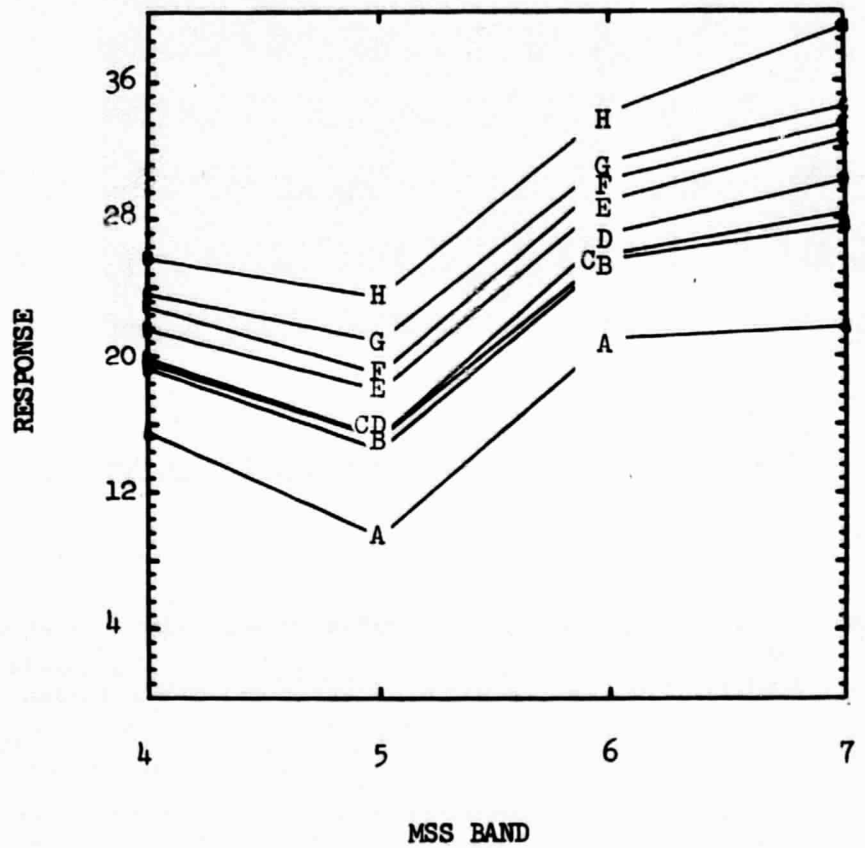


Figure 26.-Mean Spectral Response Curves for Manitou Study Plots and Estimated Component Classes. Symbols are: A=Estimated Forest Curve, B=Plot #2, C=Plot #3, D=Plot #1, E=Plot #4, F=Plot #6, G=Plot #5, H=Estimated Background Curve.

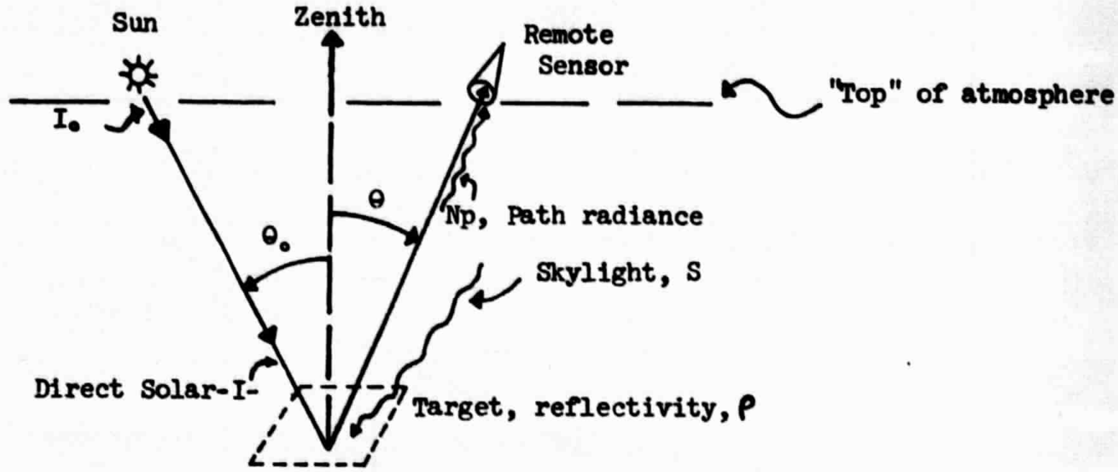


Figure 27.-Satellite remote sensing of target on the ground.

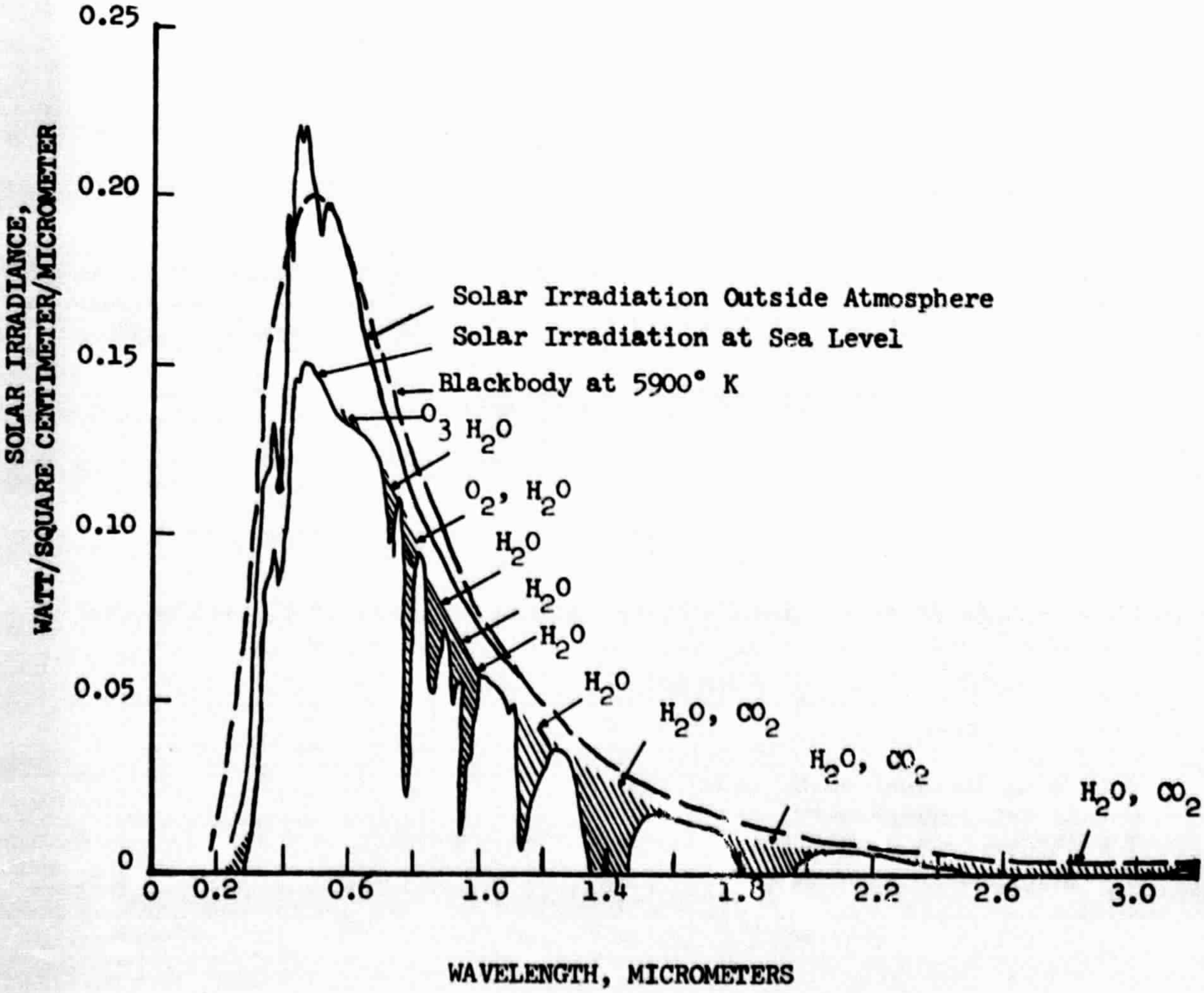


Figure 28.-Typical spectrum of solar energy reaching the surface of the earth.
From Handbook of Geophysics and Space Environment, S. L. Valley, Editor (New York: McGraw-Hill Book Company, Inc.), p. 16.2

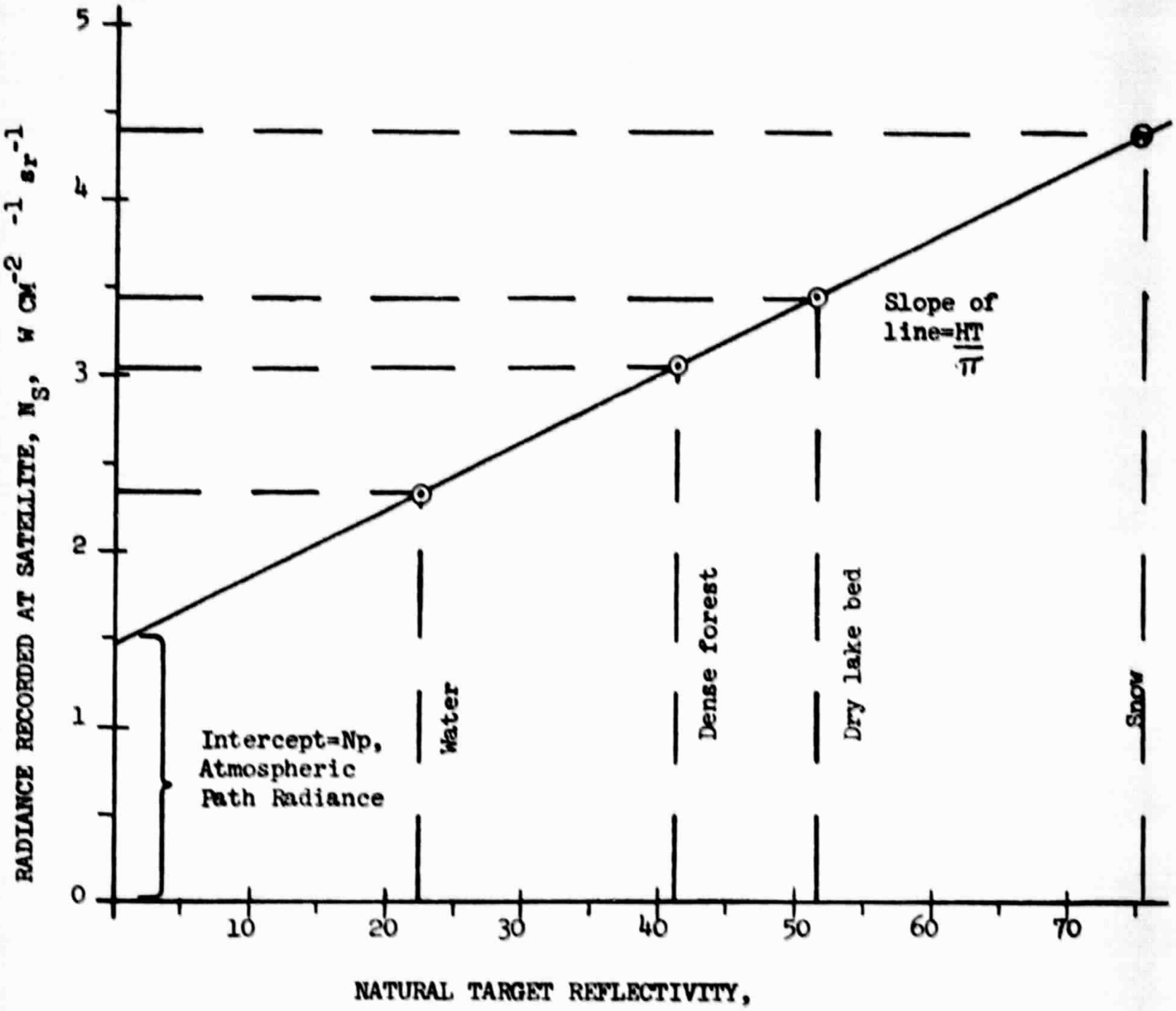
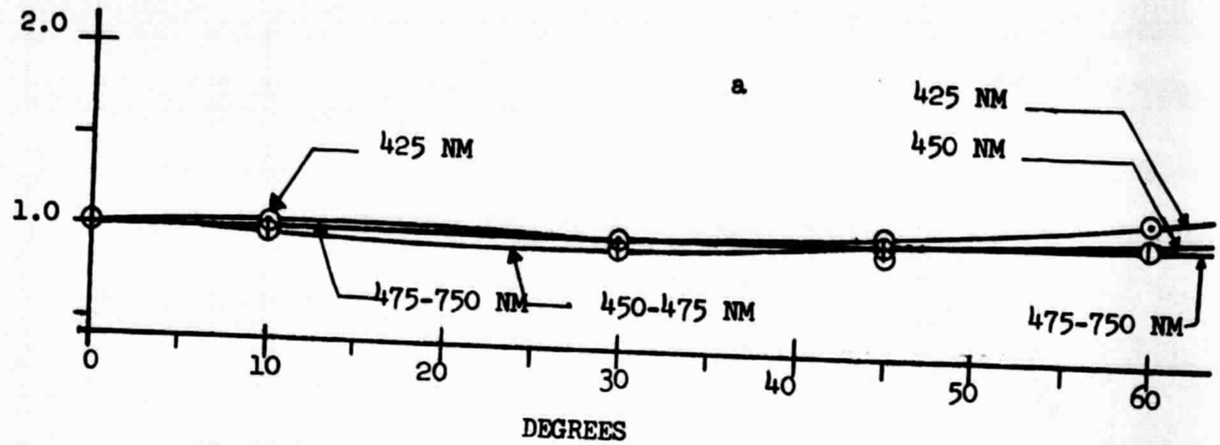


Figure 29.-Relation of atmospheric effects and the radiance of targets on the ground as recorded by satellite. Equation for this plot is:

$$N_s = \frac{H \rho T}{\pi} + N_p$$

H = Total incident solar irradiance
ρ = Target reflectivity
T = Atmospheric transmittance
N_p = Atmospheric path radiance



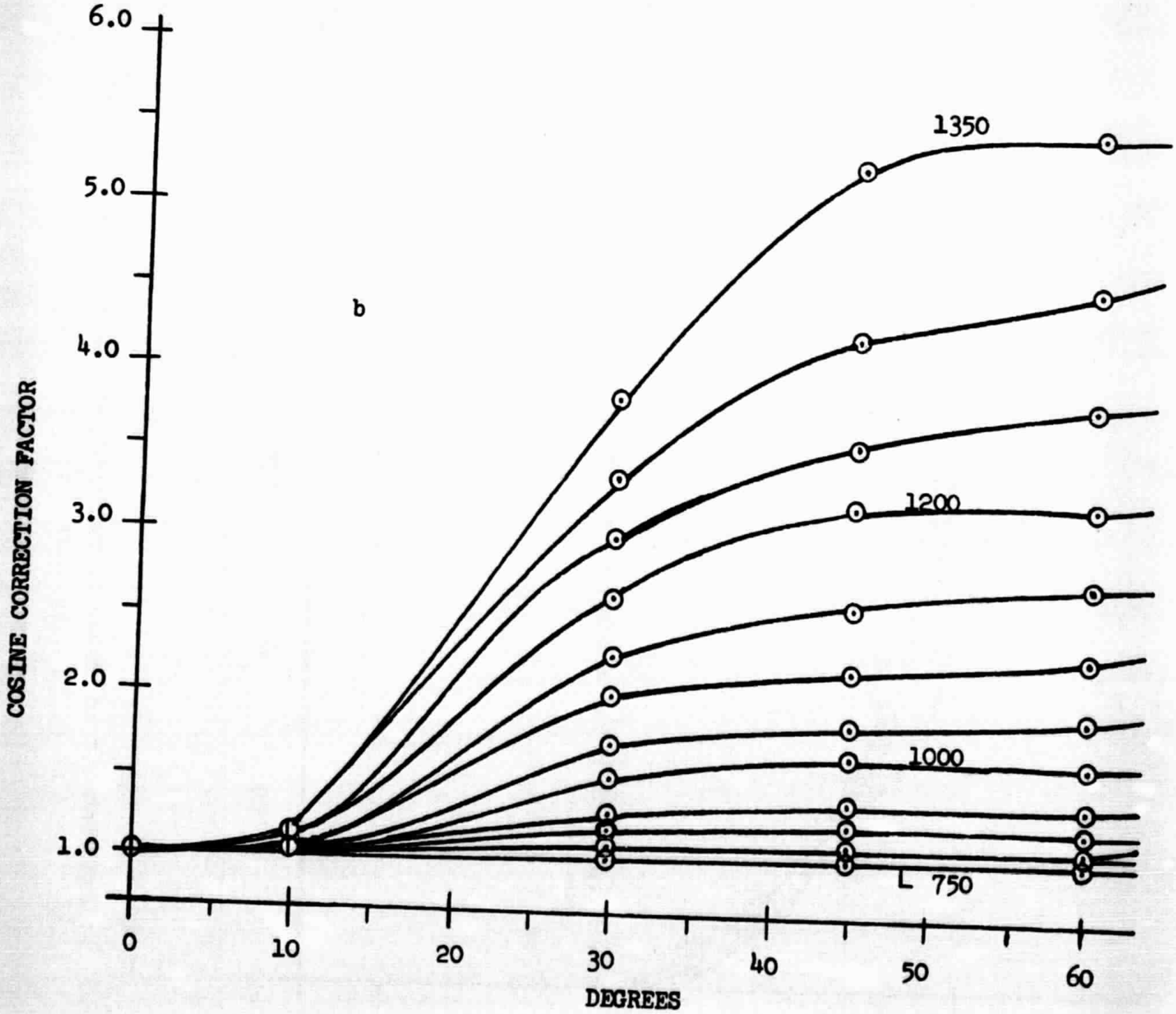


Figure 30.-Cosine correction factors for ISCO diffuser in the visible (a) and near-infrared (b) spectrum. I.S.C.O. 30109 using six-foot fiber optics probe. In b, each curve represents a wavelength from 1350 N.M. (top) to 750 N.M. (bottom).

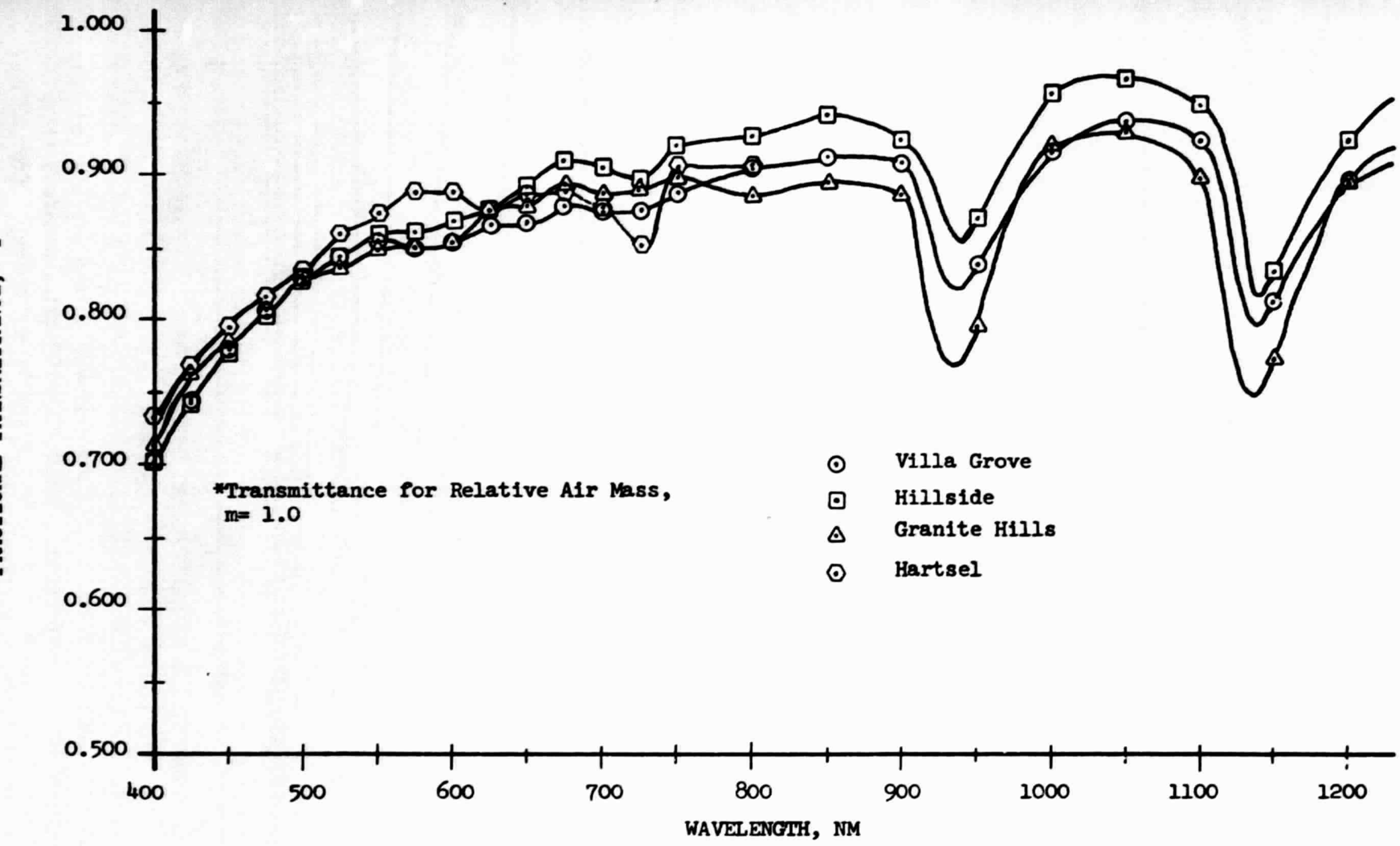


Figure 31.-Fractional transmittance for sites in south-central Colorado. All sites are separated by about 80 km. Granite Hills site lies within study area of figure 1. All data from October 2, 1974.

ORIGINAL PAGE IS
OF POOR QUALITY

0 10 20 30 40 50 KILOMETERS

FIGURE 32. Color composite LANDSAT-1 image of test site region. Image 1028-17135 of August 20, 1972. Annotations same as for figure 1: P=Pikes Peak, C=Cripple Creek, CC=Canon City, EMC=Elevenmile Canyon Reservoir. Because rectangular pixels are printed as squares by the computer the image is distorted, both in azimuths and distance. The scale shown is an average.

FIGURE 33 (ON FOLLOWING PAGE).--Computer-generated recognition map of test site. Compare with figure 32 for location of features. Made by Environmental Research Institute of Michigan (ERIM) from computer compatible tapes from LANDSAT-1 image 1028-17135 of August 20, 1972. Because rectangular pixels are printed as squares by the computer, the image is distorted both in azimuths and distances. The scale shown is an average.

EXPLANATION OF COLOR CODE

1. Niobrara Shale
2. Pierre Shale
3. Limestone
4. Canopy of Ponderosa Pine greater than 25 percent
5. Dakota Sandstone
6. Fountain Formation
7. Volcanic and plutonic rocks, undivided
8. Granite-Tundra composite
9. Granitic grus
10. Meadow
11. Water
12. Cloud shadow
13. Cloud

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX A

RECOG Computer Program Blocks

At Colorado State University, conventional automatic analysis of MSS data is accomplished through the use of a series of pattern recognition programs called RECOG. RECOG consists of six program blocks which are a modification of an original version, called LARSYS, developed at Purdue University (ref. 10). These programs provide a logical procedure for processing MSS data using supervised learning techniques. A brief description of each program block or phase is presented here.

Phase 1

Phase 1 is a display routine that provides the user with a computer line-printer or microfilm representation of the scene that is to be automatically analyzed. The MSS data are displayed as a graymap representing the radiation response from each pixel. A range of radiation responses (either specified or default) are coded as a symbol and displayed for a selected wavelength band. This display provides a pictorial representation of the MSS data from which the user may delineate field boundaries of terrain classes that he may wish to map.

Phase 2

The fields identified on the Phase 1 graymaps are used as training sets from which the mean spectral response and standard deviation vectors and correlation and covariance matrices are determined by Phase 2. The mean spectral response vector and covariance matrix provide a statistical spectral signature for a terrain class that is used in a later Phase to automatically classify each point in the MSS data set.

Phase 3

When the multispectral scanner has a large number of channels available, processing the data using all of the information becomes quite expensive. Phase 3 is designed to select a subset of optimum channels for identifying all of the terrain classes utilizing divergence criteria.

Phase 4

The spectral signatures obtained for each terrain class from Phase 2 can be analyzed as to how well they represent the class by selecting a subset of the MSS data and classifying it with Phase 4. Phase 4 is designed as an instructional mode and allows classification of the data set with three algorithms: LEVELS, a level-slicing routine; EUCLID, a Euclidian-distance routine; and GLIKE, a maximum-likelihood routine. GLIKE is the algorithm used to classify the data in the next Phase so it is valid to test the representativeness of the signature set with it. This allows refinement of each spectral signature by redefining the training set to discrad any point that would tend to make a signature unrepresentative of its respective terrain class.

Phase 5

Phase 5 is the actual classification mode for the RECOG pattern recognition sequence. The mean vector and covariance matrix for each terrain class are used with the maximum likelihood decision rule GLIKE to classify each pixel in the specified MSS data set. As the classification results are generated they are written onto a magnetic tape and stored as a permanent file.

Phase 6

The final step in the processing scheme, Phase 6, displays the results generated by Phase 5. The user is given the option to specify a threshold level which sets a confidence limit for the classified data points. This is designed to eliminate false classifications of data which do not fit any of the terrain classes. Current display modes available with Phase 6 include a thematic map produced on the computer line-printer and/or microfilm similar to those produced by Phase 1, except each pixel is identified by a symbol or intensity level (line printer and microfilm respectively) representing a terrain class. A more detailed description of RECOG can be found in ref. 4 and 15.

APPENDIX B

EQUATIONS

The mean vector and covariance matrix that describe the spectral response of a terrain class are determined from a set of MSS data points known to contain that class. The statistical equations used to compute them are discussed below.

Since the radiation response from a ground resolution element sensed by a multispectral scanner can be described by the column vector

$$X^A = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \tag{26}$$

where each x component represents the radiance recorded in a spectral channel, we can find expressions for M^A and C^A in terms of the radiation responses X^A. (ref. 10)

The mean vector M^A is given by the column vector

$$M^A = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} \tag{27}$$

where m_i is the mean spectral response in wavelength band i (for ERTS-1 i=1,....,4) given by

$$m_i = \frac{1}{n^A} \sum_{k=1}^{n^A} Xik \tag{28}$$

where n^A is the number of pixels (sample points) in the training set describing terrain class A, and k is the sample point index.

The covariance matrix which indicates how the radiation response in one MSS channel varies with the response in the other channels can be described as

$$C^A = \begin{pmatrix} \sigma^2_{11} & \sigma^2_{12} & \cdots & \sigma^2_{14} \\ \sigma^2_{21} & & & \vdots \\ \vdots & & & \\ \sigma^2_{41} & & & \sigma^2_{44} \end{pmatrix} \tag{29}$$

where $^2_{ij}$ is the covariance between channels i and j given by

$$\sigma^2_{ij} = \frac{1}{n^A} \sum_{k=1}^{n^A} (x_{ik} - m_i)\,(x_{jk} - m_j) \tag{30}$$

The standard deviation for channel i in class A is , and the correlation coefficient r_{ij} between channel i and channel j for class A is given by

$$r_{ij} = \frac{\sigma^2_{ij}}{\sigma_i \sigma_j}. \tag{31}$$