8

NR

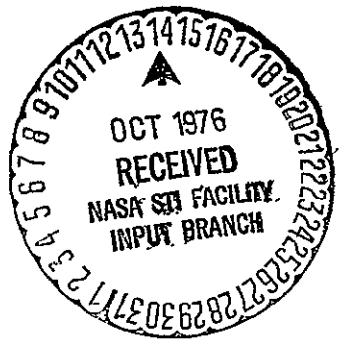
NASA CR-
147844

FINAL REPORT ON SKYLAB-EREP STUDIES IN COMPUTER MAPPING OF TERRAIN IN THE CRIPPLE-CREEK-CANON CITY AREA OF COLORADO

by

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

T-9612 B



(NASA-CR-147844) SKYLAB-EREP STUDIES IN COMPUTER MAPPING OF TERRAIN IN THE CRIPPLE CREEK-CANON CITY AREA OF COLORADO (Geological Survey) 77 p HC $5.00 CSCL 08B

N76-32622

Unclas
G3/43 05282

Contents

Illustrations

Table

FINAL REPORT ON SKYLAB-EREP STUDIES IN COMPUTER MAPPING OF TERRAIN IN THE CRIPPLE CREEK-CANON CITY AREA OF COLORADO

by

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

ABSTRACT

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 accurate control data, and 2) effects of the atmosphere on spectral responses. This paper presents the fundamental principles of these problems and some of the results of our studies of them for a test site of Colorado, using LANDSAT-1 data.

The natural terrain comprises a mixture of diverse classes, including combinations of such things as tree, grass, and other types of vegetative cover, 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) of multispectral-scanner data from spacecraft, 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 response of a mixture of terrain classes is not representative of any of the component classes composing the mixture. Automatic-recognition processors (computers) may therefore misclassify pixels which are mixtures of classes. This 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 response depending on such things as size of pixel (sample area), solar aspect, orientation of objects that make up the class, atmospheric conditions, and season.

Because of this problem of mixtures in pixels, we at first attempted to compile 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 judgments 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 map continually had to be upgraded. A serious problem that needs to be faced is: what constitutes an accurate control-data base? 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 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 spectral responses in lieu of training areas for the mix classes, extending our LANDSAT-1 research in the same area. Previous research in this problem has largely been confined to classes of crops in flat agricultural fields. Our data and techniques 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 spectral responses 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. Spectral responses for mixed classes were calculated using spectral radiance from on-site measurements. In addition, several experiments were conducted using simulated LANDSAT-1 multispectral (MSS)-type data to show the expected improvement in accuracy of automatic classification using simulated mixture-class spectral responses. We also studied methods for determining component-class spectral responses when there were insufficient data points for conventional extraction of spectral-response data.

Those techniques of calculating the mixes or the end members, either from on-site measurements or from the satellite data, are highly successful in terms of shorter preparation time and greater accuracy. In addition, the calculated spectral responses afford insights into what the appropriate increments of mixes and what the optimum wavelength bands are for the most accurate discrimination among specific classes.

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 response and actual LANDSAT-1 MSS data. A terrain map was also made from Skylab S-192 data without simulating mixture-class spectral response. Evaluations of both maps are presented.

Atmospheric effects must be known in order to combine spectral responses derived directly from LANDSAT-1 or Skylab scanner tape-data with those measured on site and from aircraft. A few large homogeneous sites such as large bodies of deep clear water, large expanses of bare rock, dense forests, large nonvegetated dry lakebeds, areas of desert sand, and snow may serve as known natural calibration panels on the ground. These would be visible from spacecraft, from whose sensor data the atmospheric path radiance and transmittance can be derived for each 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

Among the results of our Skylab research program on computer mapping of terrain, we emphasize the following: (1) 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; (2) 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; (3) the concept of a ground-control map needs to be revised; (4) atmospheric effects should be considered in computer mapping of terrain and in monitoring changes in the terrain; and (5) terrain features may be used as calibration panels on the ground, from which atmospheric conditions can be determined and monitored.

The test site (fig. 1) comprises about 2,280 sq km (880 sq mi) of generally wild-land terrain in south-central Colorado. It includes such landmarks as Pikes Peak, the Cripple Creek mining district, and Canon City. Altitudes range from 1,525 m to 4,300 m (5,000 ft to 14,100 ft). The terrain is highly varied and includes a diversity of rock types, soil, and vegetative cover, and a wide range of slope angle and aspect.

The research was carried out as an integral part of LANDSAT-1 and Skylab Earth Resources Experiment Package (EREP) projects funded by the National Aeronautics and Space Administration (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 separate but coordinated projects funded by NASA. The maps provide 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-1 data as part of a dissertation for an advanced degree at Colorado State University and made the principal evaluations of accuracy of the terrain map prepared from Skylab S-192 data.

We wish to express our appreciation for the helpful comments, discussion, and data furnished by Roy Mead, Gary Raines, Ralph Root, Frank Sadowski, James Smith, Fred Thomson, and Earl Verbeek. Special thanks go to Phyllis Adams and Martha Morris who did the typing.

REVISION OF WORK PLAN

The following two factors adversely affected the original work plan:

1) Although three Skylab EREP passes were made over the test area, data from the first and best one were not made available to us because the contract was not signed until shortly after the date of that mission. We were not notified of that first mission, but found out about it through colleagues at the Colorado School of Mines. We were not notified about the remaining two missions until after they had been completed. Therefore, although we had equipment and crews available, it was not possible to obtain measurements of atmospheric parameters and on-site spectral reflectance of terrain targets at the same time the Skylab data were acquired.

2) A computer-generated terrain map was not made available to us by ERIM until late October, 1975. This was well past the original termination date of the contract and close to the end of the extended-contract period. Along with that map was a copy of the ERIM final report (dated August, 1975). Because important terrain classes were deleted or combined, that map was not acceptable. Subsequently, at our project expense, a new map was prepared by ERIM and delivered on November 26, 1975.

Because it was impossible to obtain and compare on-site measurements of atmosphere and terrain targets simultaneously with the EREP data, or to experiment with the computer terrain-map data, the following research was conducted as deemed vital by previous and partly overlapping LANDSAT-1 studies in the same area:

A) Continuing research on the effects of mixtures of classes in computer mapping of terrain;

B) developing improved techniques for establishing spectral responses of mixtures;

C) evaluating effects of atmospheric properties on spectral responses;

D) determining which atmospheric properties are most important for terrain mapping via satellite remote-sensor data; and

E) preliminary testing of the feasibility of using natural terrain features for determing and/or monitoring atmospheric conditions.

THE MIXTURE PROBLEM

The natural terrain is composed of mixtures of classes of ground cover. This is true regardless of the size of the ground-resolution element (pixel) of the sensor system, and holds for satellite data having resolution on the order of 100 square meters, to that of the microscope having resolution on the order of a square micron or less. By tradition, each discipline of the natural sciences has accommodated this mixture 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, graphite, and quartz veins; bark, needles, leaves, stems, quartz, feldspar, and mica grains; petioles, chloroplasts, mitochondria; perthite, albite containing 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 to 80 m square (213 to 262 ft square). Any given pixel, therefore, very likely will contain two or more classes such 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 the true values (ref. 1). Errors in classification due to the mixture problem alone will be 25 to 30 percent (refs. 2 and 3).

There are two basic kinds of mixtures within pixels. One consists of two or more homogeneous end-member component classes, such as water and dense coniferous forest. This results in a boundary or edge effect. The resulting spectral response would not be representative of a pixel that contained only one of these classes. The other mixture consists of homogeneous mixtures of such things as forest canopy and a grass, rock, or soil understory. In the first kind of mixture, only those pixels along the interface will be misclassified. In the second kind of mixture, large clusters of pixels throughout the forest will be misclassified if only the component classes were used to train the computer (refs. 4, 5, 6, 7).

Variations in the amount of vegetative cover 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 classes (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-percent cover (B), and a class representing an assumed 50-50 mixture (C) of classes A and B, on the interface between them. Spectral responses 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 mixture-class C is uncharacteristic of either of its component classes (A and B) but falls between the two components' curves. Researchers investigating this phenomenon have found that the relationship of the spectral response of a mixture to that of the component classes is a function of the area of the pixel occupied by each component terrain class and the respective spectral responses of those classes.

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 references 6 and 7, is extremely time consuming, loaded with subjective judgments, 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 refs. 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 cast serious doubt upon and calls for reconsideration of what constitutes adequate control data.

However, it is very difficult to measure precise proportions of mixtures and to locate those areas in terms of LANDSAT-1 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 radiance data 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 radiances of various mixtures. The other was to use the satellite data itself to extract the radiance data for known end-member targets and then calculate radiance for various mixtures. Simulated LANDSAT-1 MSS-type data were used. These two approaches show the expected improvement in accuracy of automatic classification using simulated mixture-class radiances. In addition, methods have been developed for determining the component-class radiances from mixtures when there were insufficient data points for conventional radiance extraction of the end-member components.

CALCULATED SPECTRAL RESPONSES FOR MIXTURES

Basic Concept

There is a basic problem in using simulated mixture radiances to identify ground-resolution elements that contain mixtures. The problem is in obtaining representative spectral-response data for the end-member components of the mix. A component class is an arbitrarily selected group of materials that makes up a terrain type which is not readily subdivided and which is useful for land-management purposes. The component classes must be describable in quantitative terms such that map boundaries can be drawn around them. Because these component classes may be found in combination with others, as mixtures, they constitute end-members of the 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. Because of the 0.4-ha (1-acre) resolution of the LANDSAT-1 scanner, 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 scanner response from a pixel, a mixture of these two classes (forest and 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 (refs. 10, 11) for final display.

Calculations From On-Site Measurements

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

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

Where H is the total solar irradiance, ρ is the target reflectivity, and N is a noise or error term, the radiance from a mixture of n targets, N_m, can be approximated as

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

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 has strong specular reflectivities.

N_t is the total solar radiance. Because the target reflectivity ρ, changes with wavelength, N_t will also change as a function of wavelength. The noise or error term, C, includes re-reflected radiation such as the cloud-bright-spot phenomenon (ref. 12) and the effect of scattered light affecting the target. For example, a white object surrounded by green objects will appear slightly green. In terrain mapping, this phenomenon will be most pronounced at boundaries, and will disappear or become negligible away from those boundaries. The effect will appropriately be included in the solution of the mixture problem, dealt with below.

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

In order to simulate mixed-target spectral response, a knowledge of the total incoming solar irradiance (H), the pure target reflectivity, and the fractional area of each target need to be known. In this study the spectral distribution and magnitude of the incoming solar irradiance were measured for various sun angles and atmospheric conditions (refs. 13, 14); the pure target reflectivities were measured on site. The fractional area occupied by each target is simply varied to simulate various combinations of mixtures.

The first mixture problem addressed was that of Precambrian Pikes Peak Granite and Coniferous Forest. This is 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 arbitrarily taken from reference 12 because of the lack of opportunity to get above the forest for proper measurement in this test site. The data are assumed to be sufficiently close to be adequate for this test. Figure 3 shows all three of the parameters. The resultant simulations of component and mixture signatures for the LANDSAT-1 bands and for the continuous spectral region from 0.5 to 1.1 μm are shown in figures 4 and 5. It is interesting to note that in MSS-band 6, forest and granite are indistinguishable, as are all mixtures of the two (fig. 4).

Calculations from LANDSAT-1 Data

Modeling of spectral responses of mixtures.--Because we lacked Skylab S-192 tapes and the capability of working with such tape formats, we continued our studies using LANDSAT-1 data. This was appropriate inasmuch as the mixture problem is common to both LANDSAT-1 and Skylab scanner data, and the results are applicable to both. 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 (refs. 1, 2, 8, and 13). The model describes the mean spectral response from a pixel as:

$$M_{P_i} = \sum_{i=1}^{N} P_i M_i \tag{3}$$

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

P_i = 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:

$$C_P = \sum_{i}^{N} P_i C_i \tag{4}$$

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

P_i = relative amount (proportion) of class i;

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

Equations 3 and 4 represent the model used for automatic classification of mixtures of classes having 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 standard maximum-likelihood algorithms in such programs as RECOG (refs. 10, 15, 16), which we used because of its ready availability, LARSYS (ref. 16), and others that involve classification 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 to that 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 pattern-recognition routines to identify all MSS data points that show that

response. The value of this method lies in its straightforward applicability to the existing RECOG processing sequence which was available to us, thus eliminating the need to develop new processing algorithms.

Equation 3 was used to simulate the spectral response expected from a mixture of the two component classes (A and B) whose extracted mean vectors are shown in figure 2. Figure 8 shows the resulting simulated mean vector (D) for a 50-percent mixture of the two classes, i.e., $P_A = P_B = 0.5$. A likely explanation for the difference between curves C and D is the difficulty of accurately estimating cover densities from high-altitude aerial photographs used as ground control (in this case Mission 205, NASA RB-57 photographs having an approximate scale of 1:100,000) and the difficulty of locating nonhomogeneous (mixture) training sets on graymaps (ref. 17).

When a set of representative component terrain-class spectral responses are obtained, then the response of any mixture of these component classes may be determined using equations 3 and 4. A computer program MIX was written to take two component-class signatures and two proportion vectors and calculate the spectral response of any combination. The method involves scaling the mean vector and covariance matrix of a component class by a proportion factor and adding the result to the scaled mean vector and covariance matrix of another component class. The proportion factors p must be within the set $0 \leq p \leq 1$ and $\Sigma p = 1$. The resulting spectral signature may then be used to classify all pixels in a set of LANDSAT-1 MSS data that show a similar spectral response.

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

$$Y_i = C + B(X) \qquad (5)$$

where Y_i = 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 of 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 USED
Extraction of spectral response	Statistical sampling of MSS data points	RECOG (Phase 2, Appendix A, B) (see refs. 15 and 16)
Estimation of spectral response of components	Solving simultaneous equations	SIGCALC
	Linear regression	STAT38R
Simulation of spectral response of mixtures	Addition of weighted mean vectors and covariance matrices	MIX

Extraction of spectral response.--When areas on the ground have been satisfactorily identified as containing a known terrain class or mixture of terrain classes and the graymap coordinates have been determined, then spectral responses may be obtained by statistically sampling these points. This process is known as signature extraction and represents the conventional mode for obtaining spectral responses 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 for LANDSAT-1, are collectively known as the mean vector. Because of the inherent variability found in natural objects this set of statistics is obtained. 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. The same approach applies to Skylab S-192 data.

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

$$P_a(MSR_a)_i + P_b(MSR_b)_i = M_{P_i} \tag{6}$$

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 two sets of spectral responses exist describing training sets having 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:

$$(MSR_b)_i = \frac{M_{P_{1i}} P_{a_2} - M_{P_{2i}} P_{a_1}}{P_{a_2} P_{b_1} - P_{b_2} P_{a_1}} \tag{7}$$

Equation 7 produces the mean spectral response for class B. To derive the mean spectral response for class A, the result can be substituted into equation 6 which can be rewritten as:

$$(MSR_a)_i = \frac{MP_i - P_b(MSR_b)_i}{P_a} \tag{8}$$

If the elements of the covariance matrix for a terrain-class spectral response 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 treatment can be done because 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 responses 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 control data must be specified. Situations where

$$P_{a_2} P_{b_1} = P_{b_2} P_{a_1}$$

must be avoided because the denominator of equation 7 cannot equal zero.

The above method for determining component-class spectral signatures provides 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 spectral responses. Also, owing 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 responses for more than two mixture-class responses. 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 (refs. 15, 16). 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 responses and the proportions of the component classes.

Simulations of LANDSAT-1 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 using simulated data. These experiments included: identifying the expected improvement of automatic classification using modeled mixture terrain-class spectral responses versus conventional component-class analysis, comparing modeled and extracted mixture-class mean vectors and covariance matrices, and analyzing the spectral-responses calculation technique.

Creation of simulated data fields.--The automatic classification procedure conducted for these experiments approximates that of conventional LANDSAT-1 MSS maximum-likelihood analysis (refs. 10, 14, 15, 16) with the exception of using modeled mixture-class spectral responses and simulated LANDSAT-1 data. The

procedure is: 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 using the component-class signatures obtained in step 1 using a program which 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 this program for future use. 3) Generate a random-normal data field for each class. This is done by a program which 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 Gaussian distribution. The overlying assumption here is that wild-land terrain classes are normally distributed, which conforms to the assumptions imposed on the mixture modeling method (ref. 8) and maximum-likelihood pattern-recognition algorithms. A program 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 the same format.

The simulated data fields for each of the component and mixture classes were generated to contain 1,000 spectral-response values in each wavelength band, each point being described by four variables (spectral response in each wavelength band).

The data fields generated represent LANDSAT-1 MSS data with the added advantage of "absolute" ground-control information. Using these fields it is possible to analyze the accuracy of automatic classification using conventional maximum-likelihood routines and mixture terrain-class modeling method. An example data field is shown in figure 7 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. Spectral responses were extracted and then used to obtain mixture-class spectral responses with program MIX (ref. 10). The mixture classes were simulated by means of a specified proportion increment of 0.25, producing three mixture classes of 75 percent Grassland-25 percent Forest (C), 50 percent Grassland-50 percent Forest (D), and 25 percent Grassland-75 percent Forest (E). The mean spectral-response curves for the five classes are shown in figure 8.

The simulated data fields were then treated as actual LANDSAT-1 MSS data for purposes of classification. Component-class spectral responses were extracted from those fields created from component-class responses (top three rows, fig. 7) and used to classify the entire set of simulated data. The classification display is shown in figure 9, the Grassland field at the top and the Forest field at the bottom. The mixture-class fields are between the

components, 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, points being misclassified as the component class which has the highest proportion in the mixture.

The absolute classification accuracies were determined for each class. Because 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	UNCLASSIFIED
FOREST (F)	1,000					
25 PERCENT GRASSLAND-75 PERCENT FOREST (E)	148					852
50 PERCENT GRASSLAND-50 PERCENT FOREST (D)						1,000
75 PERCENT GRASSLAND-25 PERCENT FOREST (C)					376	624
GRASSLAND (G)					1,000	

The true class name and symbol are on the left side of the table and the symbols of classes to which a point was assigned are 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 percent, which in most LANDSAT-1 applications should be considered quite poor even though the component classes were classified perfectly.

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 technique of modeling mixture-class spectral responses produces signatures that can be used 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 spectral responses compared with classification conducted using only component-class spectral responses.

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. Spectral responses 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 because representative mixture-class training sets are usually difficult to locate in natural situations. The two extracted-component and three modeled mixture-class spectral responses 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 spectral responses 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				Modeled			
CLASS	GRASS							
BAND	1	2	3	4				
MEAN	32.45	34.13	39.10	19.40				
STANDARD DEVIATION	1.07	1.82	1.54	.88				
CLASS	75 PERCENT GRASS-25 PERCENT 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
STANDARD DEVIATION	1.06	1.80	1.65	1.01	1.04	1.72	1.66	.99
CLASS	50 PERCENT GRASS-50 PERCENT FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	24.71	22.86	29.62	15.04	24.68	22.78	29.61	15.01
STANDARD DEVIATION	1.05	1.54	1.74	1.11	1.00	1.54	1.76	1.09
CLASS	25 PERCENT GRASS-75 PERCENT 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
STANDARD DEVIATION	.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				
STANDARD DEVIATION	.95	1.12	1.96	1.26				

The standard deviations shown represent the square root of diagonal covariance elements. The results indicate that extracted and modeled mean and standard-deviation vectors vary by less than 0.1 of one standard deviation unit.

Figure 10 shows the classification results obtained by adding the modeled mixture-class responses to the data set. The classification-confusion matrix of the data of figure 10 is as follows:

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

The classification results indicate that simulated mixture-class spectral responses 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 responses. The overall classification accuracy using this technique increased to more than 98 percent, 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 owing 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 it should be possible to accurately identify a pixel containing a mixture, using this method.

Classification of three component and six mixture classes.--The successful results obtained in the five-class analysis warranted further study, so an additional component class was added to the data set. A mean vector and covariance matrix for Water (W) was extracted from August 20, 1972, LANDSAT-1 MSS data. Mixture spectral responses were modeled from the Grassland component class and the new Water class, using a proportion increment of 0.25. The resulting three mixture classes and the component class combined with the original spectral-response set, bringing the total number of spectral responses used to nine. Figure 11 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 12.

The classification procedure followed that of all previous experiments, a new data field being generated to include all nine classes. The classification display is shown in figure 13. A classification-confusion matrix for the nine-class analysis is as follows:

CLASS	F	E	D	C	G	Q	R	S	W	UNCLASSIFIED
FOREST (F)	997	3								
25 PERCENT GRASSLAND-75 PERCENT FOREST (E)	8	982	10							
50 PERCENT GRASSLAND-50 PERCENT FOREST (D)		8	962	22		8				
75 PERCENT GRASSLAND-25 PERCENT FOREST (C)			10	954	12	24				
GRASSLAND (G)				14	986					
75 PERCENT GRASSLAND-25 PERCENT WATER (Q)			6	18		976				
50 PERCENT GRASSLAND-50 PERCENT WATER (R)							999			1
25 PERCENT GRASSLAND-75 PERCENT WATER (S)								996		4
WATER (W)									962	38

These results show a good classification accuracy for all the classes and an average classification accuracy of 98 percent. Note that the greatest confusion exists between mixture classes that have the same proportion factor for a given component class, e.g., 75 percent Grassland-25 percent Forest and 75 percent Grassland-25 percent Water. This is probably the result of masking of the lesser components by the Grassland response. It suggests that increments smaller than 25 percent would not be accurately classified.

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

	Extracted				Simulated			
CLASS	GRASSLAND							
BAND	1	2	3	4				
MEAN	32.49	34.19	39.09	19.46				
STANDARD DEVIATION	1.11	1.91	1.49	.92				
CLASS	FOREST							
BAND	1	2	3	4				
MEAN	16.95	11.48	20.19	10.67				
STANDARD DEVIATION	.95	1.15	1.95	1.26				

	Extracted				Simulated			
CLASS	WATER							
BAND	1	2	3	4				
MEAN	16.72	8.13	3.98	0.24				
STANDARD DEVIATION	.55	.74	.63	.44				
CLASS	75 PERCENT GRASSLAND-25 PERCENT FOREST							
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
STANDARD DEVIATION	1.06	1.80	1.65	1.01	1.06	1.75	1.62	1.00
CLASS	50 PERCENT GRASSLAND-50 PERCENT FOREST							
BAND	1	2	3	4	1	2	3	4
MEAN	25.68	22.81	29.61	15.01	24.72	22.84	29.64	15.06
STANDARD DEVIATION	1.00	1.52	1.77	1.08	1.02	1.57	1.73	1.10
CLASS	25 PERCENT GRASSLAND-75 PERCENT FOREST							
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
STANDARD DEVIATION	1.01	1.33	1.82	1.17	.94	1.38	1.84	1.18
CLASS	75 PERCENT GRASSLAND-25 PERCENT WATER							
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
STANDARD DEVIATION	1.01	1.62	1.35	.83	1.00	1.69	1.33	.82
CLASS	50 PERCENT GRASSLAND-50 PERCENT WATER							
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
STANDARD DEVIATION	.85	1.39	1.20	.76	.87	1.44	1.14	.71
CLASS	25 PERCENT GRASSLAND-75 PERCENT WATER							
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
STANDARD DEVIATION	.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 data fields. 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.

Tests of technique of extracting spectral responses.--In wild-land areas where large areas of homogeneous terrain classes are difficult to locate, representative spectral response must be obtained 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. 18). Consequently, methods to obtain component terrain-class spectral responses from pixels containing mixtures of terrain classes in known proportions have been investigated (for example, ref. 19). The method assumes that the data are from a Gaussian distribution and each spectral response has a common covariance matrix, and involves solving a set of linear equations. Our study used a similar but much simplified technique to attempt to acquire component-class spectral responses from training-set data of terrain classes of known mixtures. A test of the method to obtain component-class spectral responses 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 spectral responses suitable for automatic mapping of LANDSAT-1 type MSS data when only mixed-terrain classes were available for signature extraction, and 2) to determine the degree of classification accuracies that can be expected when implementing these estimated spectral responses as input into the model and the computer classification program (RECOG in this study).

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

Comparison between component-class mean and standard-deviation vectors extracted from the data set and those calculated 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
STANDARD DEVIATION	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
STANDARD DEVIATION	.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	0.04	16.17	7.70	3.64	0.00
STANDARD DEVIATION	.63	.80	.66	.19	.61	.82	.76	.52

The calculated component-class spectral responses closely approximate the extracted values, the average deviation being 0.08 for mean vectors and 0.15 for standard deviations.

In order to compare the effectiveness of estimated versus extracted spectral responses, the nine-class data field was classified by the computer. The class spectral responses used as input included the three estimated component- and six mixture-class spectral responses modeled from the estimated spectral responses. The classification accuracy obtained was approximately 96 percent compared to 98 percent obtained when using extracted-component spectral responses. The classification map is shown in figure 14. The classification-confusion matrix of this data set is as follows:

CLASS	F	E	D	C	G	Q	R	S	W	UNCLASSIFIED
FOREST (F)	984	13								3
25 PERCENT GRASSLAND-75 PERCENT FOREST (E)	17	973	10							
50 PERCENT GRASSLAND-50 PERCENT FOREST (D)		20	964	11		4	1			
75 PERCENT GRASSLAND-25 PERCENT FOREST (C)			27	876	25	65				
GRASSLAND (G)				59	907					34
75 PERCENT GRASSLAND-25 PERCENT WATER (Q)			28	24	2	942				6
50 PERCENT GRASSLAND-50 PERCENT WATER (R)							1000			
25 PERCENT GRASSLAND-75 PERCENT WATER (S)								1000		
WATER (W)									1000	

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

The simulated data fields that contain the mixture-class spectral responses used to calculate the component classes represent uniform data. That is, each data point (pixel) represents a similar class of objects. Because this condition is sometimes difficult to find in natural situations, the technique was tested for mixture spectral responses extracted from nonuniform data or training sets. Nonuniform training sets can be defined as areas that contain a mixture of component classes having single pixels containing various proportions of the components. If the nonuniform training set is considered in its entirety, the mean spectral response should approximate that from a uniform training set having the same proportions of component classes. The concept of uniform and nonuniform data is illustrated in figure 15 as two 20-pixel training sets, one containing a uniform mixture of two component classes (white and shaded) (a) and one containing a nonuniform mixture (b), but both having identical overall proportions of component classes. In the uniform training set each pixel (outlined by heavy black lines) contains a 50-percent mixture of two component classes, producing

a 50-percent mixture within the entire training set. Because 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 nonuniform training set is composed of some pixels containing 100 percent of one class, some containing 100 percent 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 percent of each component class. Referring to equation 14 (Appendix C), it can be easily deduced that the mean vectors obtained from each training set should be equal, assuming constant illumination and no within-class variability. By equation 16 (Appendix C), however, it can be seen that covariance matrices will most likely vary owing to the between-pixel variability existing in the nonuniform training set.

To demonstrate these statements, mixture-class spectral responses from simulated nonuniform 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 nonuniform(*) 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
STANDARD DEVIATION	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
STANDARD DEVIATION	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
STANDARD DEVIATION	.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
STANDARD DEVIATION	.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
STANDARD DEVIATION	.63	.79	.67	.17	.63	.84	.56	.47

	Extracted				Calculated			
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
STANDARD DEVIATION	.63	.79	.67	.17	10.10	16.50	22.20	12.10

Note the close agreement of means and standard deviations derived from the uniform data and the close agreement of means but nonagreement of standard deviations derived from the nonuniform data.

The consequence of using such large covariance matrices obtained from the nonuniform data is illustrated in figure 16. The nine-class simulated data fields were classified with the calculated component spectral responses listed above. The first classification map (fig. 16a) was obtained using spectral responses calculated from uniform training sets and the second (fig. 16b) was obtained using spectral responses calculated from nonuniform training sets. The classification accuracy of component classes obtained for the uniform sets was 99.1 percent, whereas the accuracy obtained for the same classes in the nonuniform sets was only 33.3 percent. The latter figure is misleading in that the nine-class data set was classified as 97.9 percent Grassland, 1.7 percent Forest and 0.4 percent 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, technique of calculating spectral responses. Mean vectors for component classes can be successfully calculated from nonuniform mixture-class training sets, as shown above. If a common covariance matrix that describes the proper distribution can be found, then with the calculated mean vectors, the spectral response should be representative of its respective class. To verify this, the component-class mean vectors calculated from nonuniform 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 having 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 nine-class correlation matrices.

The classification procedure followed that of all earlier runs, the mixture-class mean vectors being modeled from the component-class mean vectors calculated from nonuniform data. The classification map is shown in figure 17; its corresponding classification-confusion matrix is as follows:

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

The acceptable level of classification accuracy obtained (97.2 percent) indicates that utilizing a common covariance matrix may help solve the problem of establishing representative spectral signatures for component classes when insufficient points for conventional signature extraction exist.

An additional classification was conducted using the same class mean vectors as the above analysis, but using a common covariance matrix having 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 among channels. The classification results were again promising, the average accuracy obtained being 96.5 percent. The classification map is shown in figure 18; its corresponding classification-confusion matrix is as follows:

CLASS	F	E	D	C	G	Q	R	S	W	UNCLASSIFIED
FOREST (F)	970	25								5
25 PERCENT GRASSLAND-75 PERCENT FOREST (E)	26	939	35							
50 PERCENT GRASSLAND-50 PERCENT FOREST (D)		20	942	13		20	5			
75 PERCENT GRASSLAND-25 PERCENT FOREST (C)			13	896	25	66				
GRASSLAND (G)				15	983					2

CLASS	F	E	D	C	G	Q	R	S	W	UNCLASSIFIED
75 PERCENT GRASSLAND-25 PERCENT WATER (Q)			20	22		958				
50 PERCENT GRASSLAND-50 PERCENT WATER (R)							1000			
25 PERCENT GRASSLAND-75 PERCENT 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 spectral response 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 19. The corresponding classification-confusion matrix is as follows:

CLASS	F	E	D	C	G	Q	R	S	W	UNCLASSIFIED
FOREST (F)	960	23								17
25 PERCENT GRASSLAND-75 PERCENT FOREST (E)	25	939	35							1
50 PERCENT GRASSLAND-50 PERCENT FOREST (D)		20	941	13		22	4			
75 PERCENT GRASSLAND-25 PERCENT FOREST (C)			10	893	26	70				1
GRASSLAND (G)				18	977					5
75 PERCENT GRASSLAND-25 PERCENT WATER (Q)			23	25		952				
50 PERCENT GRASSLAND-50 PERCENT WATER (R)							1000			
25 PERCENT GRASSLAND-75 PERCENT WATER (S)								1000		
WATER (W)									1000	

Note that the nine-class data-classification results using this method of average covariance matrices (95.0 percent) are only slightly less than those obtained using an estimated average of the diagonal elements and setting the off-diagonal elements to 0.0 (96.5 percent).

A summary of the results of each of the classification experiments using simulated LANDSAT-1 MSS data is as follows:

Experiment description	Classification accuracy (Percent correct)
Two component classes	40.0
Two component classes and three mixed classes	98.1
Two component classes having "close" spectral responses	39.8
Two component classes and three mixed classes having close spectral responses	58.2
Three component classes and six mixed classes	97.9
Three component classes and six mixed classes having estimated mean vectors and covariance matrices from uniform data	96.1
Three component classes and six mixed classes having estimated mean vectors and common covariance matrix (diagonal elements = 1,3,3,1; off-diagonal elements = 0.5)	97.2
Three component classes and six mixed classes having estimated mean vectors and common covariance matrix (diagonal elements = 1,3,3,1; off-diagonal elements = 0.0)	96.5
Three component classes and six mixed classes having estimated mean vectors and averaged covariance matrix (diagonal elements = 0.89, 2.12, 2.14, .87; off-diagonal elements = 0.0)	95.0

The overall results obtained from using common covariances appear promising for classifying LANDSAT-1-type MSS data when spectral responses must be obtained from nonuniform mixture-class training sets. Skylab S-192 data produced a basic terrain map only slightly more accurate than that from LANDSAT-1 MSS data. The advantage of narrower S-192 wavelength bands apparently is offset by the larger cell size and lower signal-to-noise ratio. It seems reasonable to assume that similar improvements could be obtained from S-192 data using all of the techniques described above. Those techniques, applied to automatic classification of actual wild-land data, are described in the next chapter.

Tests using LANDSAT-1 MSS data.--Experiments similar to those performed using simulated MSS data were conducted using actual LANDSAT-1 MSS data. An initial experiment was performed using extracted assumed "pure" component-class spectral responses and mixture-class spectral responses simulated from these component classes to classify an area designated as the Elevenmile Canyon Reservoir Study Area (fig. 1). An additional experiment using estimated-component

spectral responses and simulated-mixture spectral responses was performed using several plots located in the Manitou Study Area (fig. 1) where ground-control data were more detailed. Both of these experiments utilized methods described above. The objective of these experiments was to relate the concepts just presented to a real-world application. We hope they will provide insight as to how well the techniques work to improve classification accuracies and/or what other factors may need to be considered to obtain optimum results.

The set of LANDSAT-1 MSS data analyzed is from an August 15, 1973, (frame no. 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 in Fort Collins, Colorado. The data were geometrically corrected (deskewed) by the Laboratory for Applications of Remote Sensing (LARS) at Purdue University, West Lafayette, Indiana (refs. 5, 20). 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. 21).

Experiments were conducted in the study areas using extracted component-class and simulated mixture-class spectral responses to classify a part of LANDSAT-1 MSS data. The steps involved in the classifications are: (1) selecting component classes, (2) extracting representative spectral responses, (3) simulating specified mixture-class spectral responses, (4) classifying the data using 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 2,500 to 3,020 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 principal types of rocks occur: Tertiary volcanic rock and Precambrian granodiorite. The most commonly found volcanics are upper and lower members of the Oligocene Thirtynine Mile Andesite of the Thirtynine Mile volcanic field. The upper member is generally associated with high flat-topped mountains and the lower member with low rolling grass-covered hills characteristic of South Park which lies immediately west of this area.

The Precambrian X granodiorite of Boulder Creek type found in the site is dark gray, medium to coarsely crystalline, and locally gneissic. 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 part of the test area where it forms high steep mountains. Several outcrops can be found east of Elevenmile Canyon Reservoir and on both sides of Elevenmile Canyon, but their areal 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

ponderosa pine, Douglas-fir, spruce fir, subalpine fir, and aspen. According to Heller and others (ref. 4), these forest species occur as homogeneous stands and as various mixtures of species. A wide range of tree-canopy densities exists, from very open to so dense that crowns nearly touch. In open stands of trees enough sunlight enters to allow the development of an extensive herbaceous and/or shrubby understory, but where tree cover is dense, little understory cover has developed.

The principal understory vegetation associated with a ponderosa-pine forest includes Arizona fescue and mountain muhly. At higher elevations these species give way to Idaho fescue, 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 principal species are blue grama and slim-stem muhly. Mountain bunchgrass communities can also be found at the interface between grassland and forest.

Also occurring 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, bluegrass and tufted hairgrass are found in mixed communities. The prominent shrubby communities consist 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, Prairie 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 and a minimal amount of understody vegetation showing through the canopy. The Prairie Grassland class is characteristic of the short-grass prairie community and usually has much bare soil associated with it. The Water class is represented by the water of the reservoir. The Mountain Grassland class represents grassland communities found at higher elevations and at the interface between Forest and Prairie Grassland. The Wet Meadow class was selected to represent irrigated meadow areas and stream-bank communities.

Homogeneous uniform areas representative of each component class were delineated on aerial photographs. These areas (training sets) were then transferred to LANDSAT-1 graymaps and the line and point coordinates were determined. Figure 20 shows the microfilm graymap of MSS band 5 for the study site. Some modification of the photograph locations were made on the graymaps as dictated by the gray-levels associated with each training set. In this way, anomalous data points (that is, those that had high likelihood of representing some other class) were avoided before spectral responses were extracted. A number of training sets were located for each class to provide statistically representative spectral responses. 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 is 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 for each class training set were determined and analyzed. Those training sets showing dissimilar mean vectors and eigenvector plots were discarded so that final spectral responses obtained were as representative of pure end-member classes as possible.

The final component-class spectral responses were then used to classify the area. The resulting classification map is shown in figure 21 as a microfilm display. As a check of how well the spectral responses 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 percent except Water which had 99 percent. A total of 60.5 percent 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 having trees growing within Grassland areas and at 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 streambank communities.

3. Forest-Mountain Grassland--selected to identify pixels covering the interfaces between the two communities and to 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 responses were simulated using program MIX. The final spectral-response 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 22. An increase in the number of classified points resulted in 86.35 percent of the area being classified, rather than 60.5 percent. Most of the unclassified points are associated with the interface between Grassland and Water, as no spectral response characteristic of these mixtures was used in the classification. Classification of Grassland-Wet Meadow and Forest-Mountain Grassland mixtures appears reasonable on the basis of detailed study of aerial photographs. Grassland-Forest mixtures of 50-percent Grassland-50-percent Forest and 25-percent Grassland-75-percent Forest also appear to be classified well. The 75-percent Grassland-25-percent Forest mixture class contained a large number of obvious misclassifications in areas where no Forest was present.

To investigate the source of these misclassifications, spectral-response data were extracted from the data for areas classified as the 75-percent Grassland-25-percent Forest class. A comparison of the mean and standard-deviation vectors for the extracted spectral responses and the simulated mixture response 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
STANDARD DEVIATION	.94	1.66	1.51	1.00	1.62	2.02	2.10	1.15

This close correlation indicates that the material that is being classified as a Grassland-Forest Mixture probably has a vegetation and/or soil type that produces a response similar to that calculated for the mixture. The consequences of such misclassification are an overestimation of the amount of Forest in the scene and an underestimation of Grassland and/or soil. Because of the difficulties 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-percent Grassland-25-percent Forest class, the proportion increment used to simulate mixture-class spectral responses was changed to 0.33 and new spectral responses 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 in 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 23. The results indicated that the areas previously misclassified as 75-percent Grassland-25-percent Forest were still being incorrectly identified. A number of points along the shoreline of the reservoir were identified as either 67-percent Grassland-33-percent Water or 33-percent Grassland-67-percent Water. There appeared to be some confusion of 67-percent Wet Meadow-33-percent Water and 33-percent Wet Meadow-67-percent Forest as the component class having the greatest amount occurring in the mixture. The other mixture classes seem to be classified correctly as they appeared where mixtures would logically exist.

Manitou Study Area.--The Manitou Study Area (fig. 1) is northeast of the Elevenmile Canyon Reservoir Study Area in and around the U.S. Forest Service Manitou Experiment Forest. Only a few selected plots identified by Mead (ref. 5) were studied in this work.

The plots lie along and just north of the area shown in figure 1. The area typifies the mountainous part of the east slope of the Rocky Mountains, and has a large variation of slopes and aspects. The topography is dominated by mountains and valleys having a north-south orientation due to extensive faulting that occurred during the Pliocene Epoch and earlier. Elevations range from about 2,600 to 3,300 m (8,530 to 10,830 ft).

For the most part, geologic materials in the area are covered with vegetation. The underlying bedrock unit is the reddish-brown Permian and Pennsylvanian Fountain Formation, but exposures are limited. The principal materials are pinkish to reddish soils derived from the Precambrian Y Pikes Peak Granite and washed into this area.

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 the Manitou Study Area, experiments were performed to classify LANDSAT-1 MSS data using component-class spectral responses 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 in the Manitou Study Area and identified by Mead (ref. 5) were analyzed to obtain a number of plots suitable for estimation of spectral response. Mead was specifically interested in the percentage of ponderosa-pine cover 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 occurring in the same relative amounts. It was also necessary to identify plots having similar slopes and aspects to avoid the variability in spectral response associated with these factors. In summary, the criteria used for selecting the plots used for estimation of spectral response were:

1. The plots must have known percentage of ponderosa-pine cover. (The percentage of the Background class was determined by subtracting the percentage of ponderosa-pine 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 occurrence. The third was met by locating the plots on topographic maps and determining slopes and aspects by a computer program called TOPOGO (refs. 5, 22).

Analyzing all plots using 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)	PERCENTAGE OF PONDEROSA PINE COVER	SLOPE (Percent)	ASPECT	FREQUENTLY OCCURRING 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 responses 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 24. Several pairs of signatures along with their respective mixture proportions of Ponderosa pine and Background were used as input for signature-calculation program. 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. Consequently, this method was found to be unreliable for estimating component-class spectral responses for this case. Because results obtained using simulated data, described above, were acceptable, the error was probably due to inaccurate estimates of percentage of cover and variations of materials associated with the Background class. Since no representative spectral responses could be obtained, no modeling of mixture-class spectral responses or classification of MSS data was attempted.

Because the technique of estimating spectral response used in the above study uses data for only two mixture training sets, slight errors in estimates of percentage of cover will result in large errors for the calculated values of spectral responses. Also, the spectral responses, if successfully estimated, may be representative only of the material found in the specific training areas. To compensate for these errors, all of the mixture training sets were used to determine component-class spectral responses by means of linear-regression techniques. Additional cover estimates were made using a dot grid superimposed on enlarged aerial photographs of the six plots.

The method used for estimating component-class spectral responses using linear regression was a stepwise multiple-regression FORTRAN computer program. 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	R^2
4	(0.5-0.6 μm)	MSR = 28.21-9.41 (P_p)	0.90
5	(.6-.7 μm)	MSR = 26.59-13.11 (P_p)	.89
6	(.7-.8 μm)	MSR = 35.27-10.75 (P_p)	.91
7	(.8-1.1 μm)	MSR = 20.15-7.23 (P_p)	.99

where MSR = Estimated mean spectral response

P_p = Proportion of ponderosa pine in the scene

R^2 = Correlation coefficient.

High R^2 values indicate that the grid estimates for the plots are reasonable values for the amounts of Ponderosa pine and Background. Because only six plots were used in developing the regression models, they may be applicable only for analysis of these plots. Component-class signature estimates can be determined from the models by setting the proportion of Ponderosa pine (P_p) to 1.0 and calculating the MSRs in each band to determine the Ponderosa-pine signature, and setting Pp to 0.0 to determine the spectral response of the Background. The mean spectral-response curves for the estimated spectral responses of the Ponderosa pine and Background are shown along with the extracted mixture curves for each of the plots in figure 24.

The estimated mean vectors for the Ponderosa pine and Background classes were used as inputs into the computer program and mixture-class spectral responses were simulated. The proportion increment used was 0.25, producing three intermediate-mixture spectral responses. No covariance matrices for the component-class spectral responses were estimated by the regression models owing to the nonuniformity of the mixture plots, so a common covariance matrix having diagonal elements of 1,3,3,1, 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 signature responses for Ponderosa pine 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 percent, illustrating the need to account for pixels that contain mixtures. A second classification was conducted using the two-component-class spectral responses plus the three simulated mixture-class spectral responses. The number of classified points increased to 100, or 100 percent upon 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. The percentage of foliar cover of ponderosa pine determined from ground-control data, component-class classification, and component-class plus simulated mixed-class classification results are summarized on a plot-by-plot basis, as follows:

	PLOT					
	1	2	3	4	5	Average
Ground-control estimates (percent)	83	87	85	55	15	65
Component-classification estimates (percent)	75	95	100	15	5	58
Component-mixture classification estimates (percent)	85.00	90.00	86.25	63.75	42.50	73.5

Those plots having higher control-data estimates of Ponderosa pine (plots 1, 2, and 3) were overestimated by the component-class classification by an average of 10.33 percent. The component- and mixed-class classification results for the same plots were overestimated by only 2.10 percent. The remaining two plots having control-data estimates of 55 percent (plot 4) and 15 percent (plot 5) of Ponderosa-pine foliar cover yielded much higher deviations. Although it is difficult to validate a trend using the limited number of plots in this study, it appears that those plots having a greater proportion of the Background class showed poorer classification estimates of Ponderosa pine. These estimates may be due to varying mixtures of constituent vegetation types occurring within these plots (ref. 5), and/or to errors in the dot-grid method used for determining proportions from aerial photographs.

The results of this test show an improved estimate of the percentage of foliar cover of Ponderosa pine for plots having higher proportions of Ponderosa pine. In addition, the effects of an unrepresentative component class (Background) of degrading the classification estimates is indicated.

EFFECT OF THE ATMOSPHERE ON SPECTRAL RESPONSE

Remote Sensing of Target Spectral Response

The spectral response of a target, as viewed from a satellite remote sensor (LANDSAT-1, Skylab EREP), is a radiance value N_S, given by

$$N_S = \frac{\rho HT}{\pi} + N_P + C \qquad (9)$$

where H is the total (180^o, 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 radiance (between the satellite sensor and the target) introduced by atmospheric scattering of sunlight, ρ is the target reflectivity, and C is the error or noise term as in equation 1. This target reflectivity, ρ, is the true spectral response of the target. As can be seen in equation 9, the target's radiance response is made up of three parameters other than the true target reflectivity. These parameters are the total incoming solar irradiance (H), the atmospheric beam transmittance (T), and the atmospheric path radiance (N_P). These atmosphere parameters are discussed in detail in references 12, 13, and 14, and will not be repeated here inasmuch as it was not possible to conduct the planned studies of the effects of atmosphere using either LANDSAT-1 or Skylab data.

One new technique of deriving path radiance was formulated in this project, but was not tested. The technique is to employ the satellite data itself and measurements from helicopter of selected targets on the ground. The purpose of the helicopter is to enable measurements to be made of component classes of areal extent comparable to the satellite pixel size (0.4 ha or 1 acre) thereby including effects of variation in leaf orientation, range in particle size of materials, and other attributes, which need to be integrated over the full pixel size.

$$\rho_t = \frac{\pi Nt}{H} \text{ (assumes Lambertian target).} \qquad (10)$$

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

$$\rho = (N_s - N_p) \times \pi/HT. \qquad (11)$$

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) numerous ground-based measurements are not required in order to calibrate atmospheric effects. It is required that natural targets be selected whose reflectivities remain nearly constant over periods of time, not subject to rapid seasonal changes of growth such as in grasslands.

The possible shortcomings of this technique are:

(1) it assumes fairly uniform atmospheric conditions of the surface area within which all the natural target "calibration" sites are located. This shortcoming is minimized if the sites are all located near each other;

(2) it assumes that the sum angle with respect to the targets--both material "calibration" and unknown targets--is similar. This assumption is apparent from the sum-angle dependence, $I \cos \theta_o$, of the total solar irradiance and the product HT/π.

Although time and funds were insufficient to test this technique in the field, it is described here because it seems worthy of further study by someone.

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 (ref. 23). Once their spectral responses 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 (refs. 13, 14, 25).

Transmittance, optical depth, path radiance, and total incident solar irradiance determine contrast reduction and the satellite-measured target radiance. Our work and that cited demonstrates the techniques and instrumentation for making ground-based measurements of optical depth, diffuse sky irradiance, and total irradiance. These techniques and instrumentation will allow satellite data to be corrected for atmospheric effects. Our research has also demonstrated temporal and spatial variations in these parameters (ref. 13). Therefore, atmospheric effects should be derived for several locations within any given satellite frame.

Model calculations of LANDSAT-1 MSS radiance for several targets under different atmospheric conditions showed that satellite radiance measured over a target of deep water at wavelengths in the near-infrared region (greater than 0.750 μm) is essentially equal to the atmospheric path radiance because of the very low reflectivity of water at these wavelengths. It was then determined that for given solar zenith and observation angles, the satellite-measured path radiance could be related to atmospheric optical depth, ratio of diffuse to total irradiance, and visibility (refs. 13, 14). Hence, the measurement of satellite radiance over water targets is concluded to be a potential technique for defining atmospheric effects, contrast reduction of target-to-background, and visibility from a satellite or high-altitude aircraft. Additional work is required to expand these studies to include a greater range of wavelength and solar aspect and wider range of expected reflectivities from water surfaces, including turbid water.

Simultaneous measurements at two sites only 13 km (8 mi) apart revealed a 30-percent variation in optical depth for the 0.3- to 2.8-μm region, a 25-percent variation for the visible range, and a 36-percent variation in optical depth for

the near-IR region. Hence, even under "clear" conditions, spatial variations did occur that would significantly affect remote sensing. It was also discovered (ref. 12) that clouds can act as mirrors which cause as much as a 30-percent increase (compared to clear-sky conditions) in the total solar irradiance. This effect would result in a variation of 30 percent in the radiance recorded by the satellite-borne sensor.

Analyses of these data revealed that the spectral ratio of diffuse to total irradiance is extremely sensitive to and hence a good measure of atmospheric clarity (ref. 12). This ratio varied from about 0.45 for the shorter (blue) wavelengths to 0.05 for the near IR, and varied markedly from one supposedly "clear" day to the next.

It was shown that field measurements of target reflectivity, incident total irradiance, and atmospheric transmission/optical depth could be combined with computer calculations (based on field measurements of total irradiance and optical depth) to adequately predict the radiance value measured at the satellite (refs. 14 and 24).

TERRAIN MAP FROM S-192 DATA

A terrain-classification map was prepared for the western part of the area, shown in figure 25, using Skylab EREP S-192 data of August 4, 1973. The map area, about 32 x 40 km (20 x 25 mi), is centered near Cripple Creek and extends from Colorado Springs west to Elevenmile Canyon Reservoir. Elevations range from 1,500 to 4,270 m (4,900 to 14,000 ft). Cripple Creek lies at about 3,050 m (10,000 ft) elevation.

The terrain map, shown in figure 25, was prepared by ERIM by computer-implemented pattern-recognition processing of S-192 digital-tape data. Most of the processing was done on an IBM 7094 computer as described in the final report by Thomson (ref. 25). Although the terrain map used in Thomson's report was not used in this study, the processing described by him is the same as was used for the map of figure 25.

The following six channels of S-192 data were determined to be best suited for discriminating the units described in Table 1 and shown on the map of figure 25.

S-192 channel	Wavelength (μm)
4	0.56 - 0.61
5	.62 - .67
6	.68 - .76
9	1.09 - 1.19
11	1.55 - 1.75
12	2.05 - 2.35

These channels were selected on the basis of ordering of the channels according to signal/noise ratios for best overall mapping by rock type, as follows:

Order	Signal-to-noise ratio	Channel
1	1	12
2	2	4
3	3	11
4	4	5
5	6	6
6	8	9

The statistical probabilities of misclassification for each pair of signatures used in the recognition processing were the basis for combining volcanic and plutonic rocks except for the Pikes Peak Granite, and for designating rock units having more than 75 percent forest as a Forest category.

The average pairwise probability of misclassification for each of the channels selected is:

Channel	Average probability of misclassification
12	0.0171842
4	.0063140
11	.0914933
5	.0101393
6	.0076549
9	.0242315

The terrain classes mapped (fig. 26) are summarized in table 1. A more detailed description of the rock units can be found in reference 26.

Table 1.--Summary of terrain classes mapped

Map unit	Class	General location in test site	Description of class	Classes misclassified as this unit (in decreasing order)
1	Dakota Sandstone	Southern corner	Cretaceous gray to brown medium- to fine-grained sandstone. Forms moderate slopes that face south and are covered by a grass-park community containing sparse sagebrush. Deep gullies dissect this unit in many places, possibly creating edge-effect mixture problem in computer mapping.	3, 6, and nonvegetated alluvium and windblown sand.
2	Fountain Formation	Southern part	Permian and Pennsylvanian moderate reddish-brown arkosic conglomerate, coarse yellow-gray arkosic sandstone, and thin interbedded pale-green and dark-reddish-brown shale. Virtually devoid of vegetation. Forms wide valleys and gentle slopes.	12, 5, 1, 4, and nonvegetated windblown sand deposits in the southern part of the test site.
3	Niobrara Shale	Southeastern part	Cretaceous yellowish-brown, soft, thin-bedded calcareous shale interbedded with thin beds of limestone. This is the Smoky Hill Member of the Niobrara Shale. Sparse vegetative cover of grasses, forbs, and sagebrush which do not appreciably mark this underlying bedrock unit.	4, 1, and undivided limestone member of the Greenhorn Limestone which was considered to be of too limited areal extent to form a separate class.
4	Pikes Peak Granite	Central to extreme northeastern part	Precambrian pink to reddish-tan medium- to coarsely-crystalline biotite or hornblende biotite granite and quartz monzonite. Weathers readily to a coarse loose sand.	12, 6, 5, 3, nonvegetated alluvium, certain areas of Colorado Springs.
5	Pierre Shale	Southern corner and east edge	Cretaceous dark marine shale containing bentonite beds and well-preserved fossils.	1, 4, 2, 12, 3.
6	Dakota Sandstone and vegetation composite	Southern corner	Same as unit 1 except this unit here has approximately 50-percent cover of pinyon pine and juniper and understory of grass-forb community. Sparse sagebrush.	12, 1.

Table 1.--continued

Map unit	Class	General location in test site	Description of class	Classes misclassified as this unit (in decreasing order)
7	Forest	All but the eastern edge and southern corner	Greater than 75-percent canopy of trees. Includes pinyon-juniper forest in the arid southern part and forests of ponderosa pine, lodgepole pine, and spruce fir throughout most of the central and western part.	4, 12, 11.
8	Unclassified	Small scattered sites near east corner	Principally clouds. Scattered areas of pavement and other manmade structures in or near Colorado Springs.	
9	Cloud shadow	Central and northern part	Shadows of clouds.	11, 7.
10	Meadow	Throughout the site	Alluvium having dense cover of grasses, commonly associated with irrigated fields.	4 and an unmapped unit of granodiorite having vegetative cover of scrub oak-grass community east of largest cloud mass of fig. 25.
11	Water	Throughout the site	Lakes and reservoir of clear water. Most are manmade impoundments.	7, 9.
12	Volcanic and plutonic rocks, undivided	From the south-central to the western edge	Upper and lower members of the Oligocene Thirtynine Mile Andesite, Precambrian granodiorite, quartz diorite, and biotite gneiss.	1, 4, 2, 5, 3, 10, and local small areas of nonvegetated alluvium.

A large amount of Pikes Peak Granite was misclassified as the volcanic-plutonic unit. Numerous areas consist of nonvegetated arkosic alluvium composed of erosional debris from the Pikes Peak Granite and less abundantly from volcanic and undivided plutonic rocks, misclassified as Pikes Peak Granite. These misclassifications are caused by the similar chemical and spectral characteristics of the units. Although it is desirable to be able to sort out genetic units such as granite from those such as alluvium derived from the granite, such distinctions are inappropriate for multispectral sensors. They require other criteria for distinguishing them from one another.

Because of the large amount of cloud cover, cloud shadows were included as a map unit. However, because some 15 to 20 percent of the cloud shadow was misclassified as water, the result is a marked overestimation of the limited water resources in the test site. Away from cloud shadows, water bodies were mapped with an accuracy of about 95 percent, indicating the accuracy to be expected from S-192 data on a cloud-free day.

The recognition map (fig. 25) was evaluated quantitatively by two methods. The first consisted of identifying several points that were identifiable on the recognition map and the RB-57 aerial photography and constructing corresponding grid overlays for the recognition map and the photography. The resulting cell size encompassed a 5 x 5-pixel array. Several cells were then compared with ground truth to determine the accuracy of classification. The second method consisted of constructing overlays of 5 x 5- and 10 x 10-pixel arrays and comparing areas on the recognition map known to contain a specific terrain class as determined from ground-control data. Several areas were sampled using both methods and the results were compiled in the classification-confusion matrix shown below, data shown in percent. Diagonal elements are underlined.

Sample points	Class and abbreviation	CLASSIFIED AS											
		DSS	DSC	FTN	NS	PS	PPG	V-P	MDW	FOR	WAT	SC	Unclassified
346	Dakota Sandstone DSS	88.7	3.2		8.1								
200	Dakota Sandstone and vegetation composite DSC	1.5	90.5					8.0					
518	Fountain Formation FTN	6.6		61.2		11.2	3.8	17.2					
499	Niobrara Shale NS	4.6			90.2		5.2						
492	Pierre Shale PS	10.0		8.5	.5	69.7	8.7	2.6					
478	Pikes Peak Granite PPG		2.9		.4	1.3	61.7	33.7					
718	Volcanic-plutonic rocks V-P	1.4	21.2	6.7	1.8	2.4	13.9	51.9	0.7				
200	Meadow MDW						5.5		94.5				
364	Forest FOR						14.0	9.9		73.6	2.5		
131	Water WAT									4.6	94.7	0.7	
400	Cloud shadow SC									2.0	17.5	80.5	

For comparison, the diagonal elements of S-192 data (underlined in the preceding table) and of corresponding LANDSAT-1 data are as follows (in percent; dash leaders indicate not used):

Class	S-192	LANDSAT-1 MSS
Dakota Sandstone	88.7	--
Dakota Sandstone plus vegetation	90.5	--
Fountain Formation	61.2	67.7
Niobrara Shale	90.2	85.7
Pierre Shale	69.7	8.3
Pikes Peak Granite	61.7	71.2
Volcanic and plutonic rocks	51.9	--
Meadow	94.5	73.7
Forest	73.6	95.2
Water	94.7	70.6
Cloud shadow	80.5	100.0
Rocks, undivided	99.9	68.0
Vegetation, undivided	84.1	93.5

Rocks were generally classified better by S-192 than by LANDSAT-1 MSS data, in part because the training units were selected to detect or emphasize the rock understory rather than the vegetative canopy to a higher degree than was done using the LANDSAT-1 data. For example, grassland was a major category of the LANDSAT-1 map but is not a class for this study. Instead, it was classed largely as the undivided volcanic-plutonic class which underlies most of the grassland in the test site. Hence, part of the improved accuracy is due more to semantics than to performance of the scanner.

The technique of using simulated mixture signatures could be expected to materially improve the accuracy, as described in earlier sections of the report.

SUMMARY AND CONCLUSIONS

Methods have been discussed for simulating spectral response of mixtures of two terrain classes for automatic analysis of LANDSAT-1 MSS data using on-site measurements, simulated LANDSAT-1 data, and actual LANDSAT-1 data. Tests of the methods for estimating component-class spectral response and for simulating two component-class mixture responses using 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 wild-land areas showed an increase in classification information over conventional analysis, but no quantitative accuracy analysis could be made owing to difficulties of estimating mixture proportions from the control data. Similar improvements in accuracy should result if these techniques are applied to Skylab S-192 data.

It was found that misclassification of pixels as mixtures can occur where the simulated spectral response approximates that of component classes or other mixtures in the scene. There was also evidence of misclassifications due to the existence of more than two component classes 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-control information upon which the spectral-response estimation and simulation techniques are based.

The conclusion was reached that for computer mapping of terrain one has to choose between two options for acquiring spectral-response data to be used for establishing training sets:

1) Use on-site measurements, preferably from a helicopter, in order to ensure proper sample size compatible with the satellite sensor-cell size. These measurements must be corrected for atmospheric effects to agree with the radiance values recorded by the remote sensor. Measurements of target and of atmosphere must be made simultaneously with the satellite pass.

2) Extract spectral response data directly from the satellite remote-sensor data. In this case, the use of simulated mixture provides a real advantage as described above.

We conclude that the particular LANDSAT-1 MSS and Skylab S-192 data sets analyzed were acquired at times when the atmosphere was "clear," except for the clouds. We were able to map some terrain classes with accuracy greater than 95 percent without correcting for the atmosphere. Under these conditions it probably is not economical to attempt to improve the accuracy by correcting for atmosphere. The only real test of the atmosphere problem would be to analyze a data set acquired during a time of very hazy atmosphere.

We also conclude that, because of the variability of natural terrain classes, mixture increments smaller than about 25 percent are not feasible.

The technique of using natural terrain features as calibration panels may be a satisfactory substitute for on-site measurements, and has the advantage of serving to monitor air quality as well as to aid in calibrating sensor data for mapping of terrain.

Additional applications of some of the data and concepts developed during this LANDSAT-1-Skylab study apply to other problems of atmospheric visibility. For example, the Environmental Protection Agency (EPA) has requirements for defining clear-air visibility in order to establish clear-air baselines. Vertical and horizontal atmosphere visibility have obvious effects on aircraft operations and military surveillance.

REFERENCES

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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: Proceeding Ninth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, v. II, p. 1369-1398.

8. 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.

9. 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, Maryland, 127 p.

10. 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.

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

12. Hulstrom, R. L., 1973, The Cloud Bright Spot: Jour. Photogrammetric Engineering, v. 39, p. 370-376.

13. Smedes, H. W., R. L. Hulstrom and K. J. Ranson, 1975, The Mixture Problem in Computer Mapping of Terrain: Improved Techniques for Establishing Spectral Signatures, Atmospheric Path Radiance, and Transmittance: Proceedings of the NASA Earth Resources Survey Symposium, Houston, Texas, v. 1-B, p. 1098-1159.

14. Hulstrom, R. L., 1974, Spectral Measurements and Analyses of Atmospheric Effects on Remote Sensor Data: Proceedings Society Photo-Optical Engineers v. 51, p. 90-100.

15. 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.

16. _____1972b, User's Manual for RECOG (Pattern RECOGnition Programs): Colora State University Science Series 3B, Fort Collins, Colorado, 85 p.

17. Swain, P. H. and D. A. Germann, 1968, On the application of man-machine computing system to problems in remote sensing: Laboratory for Agricultura Remote Sensing (LARS) Information Note 051368, 10 p.

18. 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.

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

20. 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, Indiana, 23 p.

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

22. Sharphack, D. A. and G. Akin, 1969, An Algorithm for Computing Slope and Aspect from Elevations: Photogrammetric Engineering, v. 35, p. 247-248.

23. Thomson, F. and F. Sadowski, 1975, A Study of Atmospheric Effects on Patter Recognition Devices: Env. Research Inst. of Michigan (ERIM) Report No. NASA Cr-ERIM 193300-62-F, 66 p.

24. Adams, L., R. Hulstrom and L. Oldham, 1975, Inflight performance Evaluation of Satellite Remote Sensor Data by Ground-based Measurement: Proceedings of Sixth Aerospace Conference, Oct. 17, 1975, 14 p.

25. Thomson, F. J., 1975, S-192 Land Resource Maps of Cripple Creek Area: Fina Report to NASA, no. 101700-25-L, 21 p.

26. Grose, L. T., 1960, Geologic formations and structures of Colorado Springs area, Colorado: In The guide to the geology of Colorado Geol. Soc. America Rocky Mountain Assoc. of Geologists, and Colorado Scientific Society, p. 18

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. 17). 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 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 instruction mode and allows classification of the data set by means of 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 using it. This allows refinement of each spectral signature by redefining the training set to discard any point that has high likelihood of not belonging to the class.

Phase 5

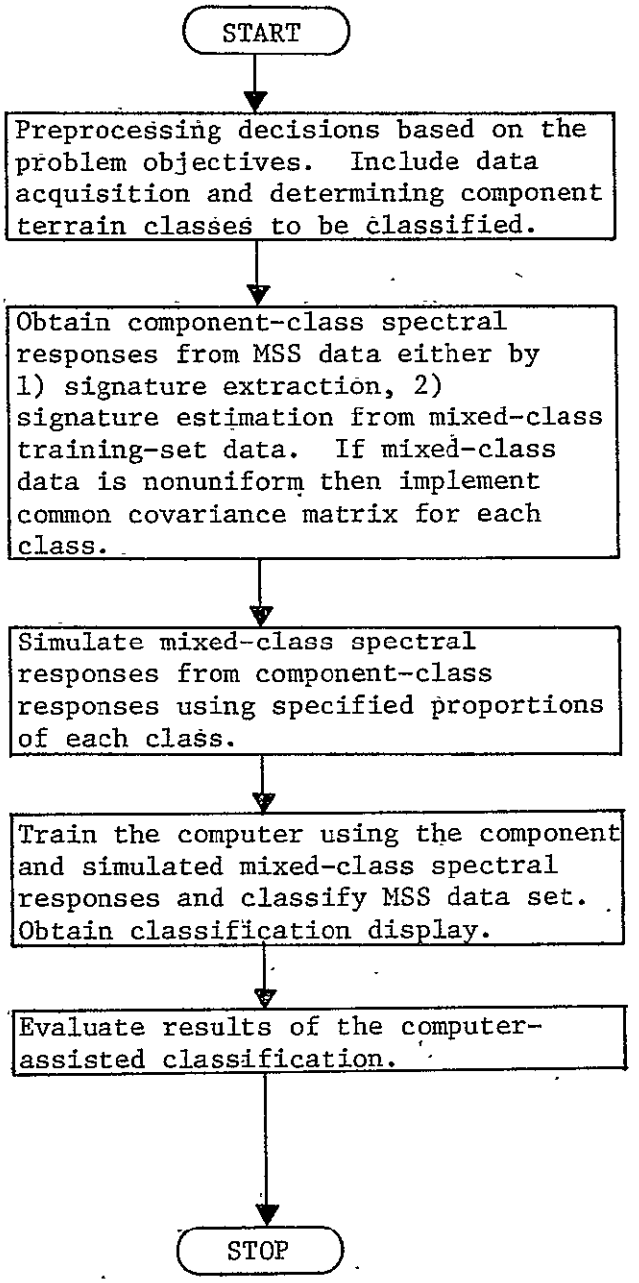
Phase 5 is the actual classification mode for the RECOG pattern-recognition sequence. The mean vector and covariance matrix for each terrain class are used with the maximum-likelihood-decision rule GLIKE to classify each pixel in the specified MSS data set. As the classification results are generated 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 that 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 references 15 and 16.

APPENDIX B

Flow chart of procedures for computer-assisted classification of mixtures in multispectral scanner data



TECHNIQUE

Decision making.

Ground-truth data analysis and 1) Phases 1 and 2 of RECOG, or 2) Phases 1 and 2 of RECOG and SIGCALC.

Program MIX.

Phases 5 and 6 of RECOG.

Comparison of computer-classified areas with corresponding areas on aerial photography or other ground truth.

APPENDIX C

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.

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

$$X^A = \text{for LANDSAT} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \qquad X^A = \text{for other scanners} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} \tag{12}$$

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 response X^A (ref. 1(

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

$$M^A = \text{for LANDSAT} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} \qquad M^A = \text{for other scanners} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_n \end{pmatrix} \tag{13}$$

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

$$m_i = \frac{1}{n^A} \sum_{k=1}^{n^A} x_{ik} \tag{14}$$

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

$$C^A = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \cdots & \sigma_{14}^2 \\ \sigma_{21}^2 & \vdots & \\ \vdots & \vdots & \\ \sigma_{41}^2 & & \sigma_{44}^2 \end{pmatrix} \tag{15}$$

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

$$\sigma_{ij}^2 = \frac{1}{n^A} \sum_{k=1}^{n^A} (x_{ik} - m_i)(x_{jk} - m_j). \tag{16}$$

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

$$r_{ij} = \frac{\sigma_{ij}^2}{\sigma_i \sigma_j} \tag{17}$$

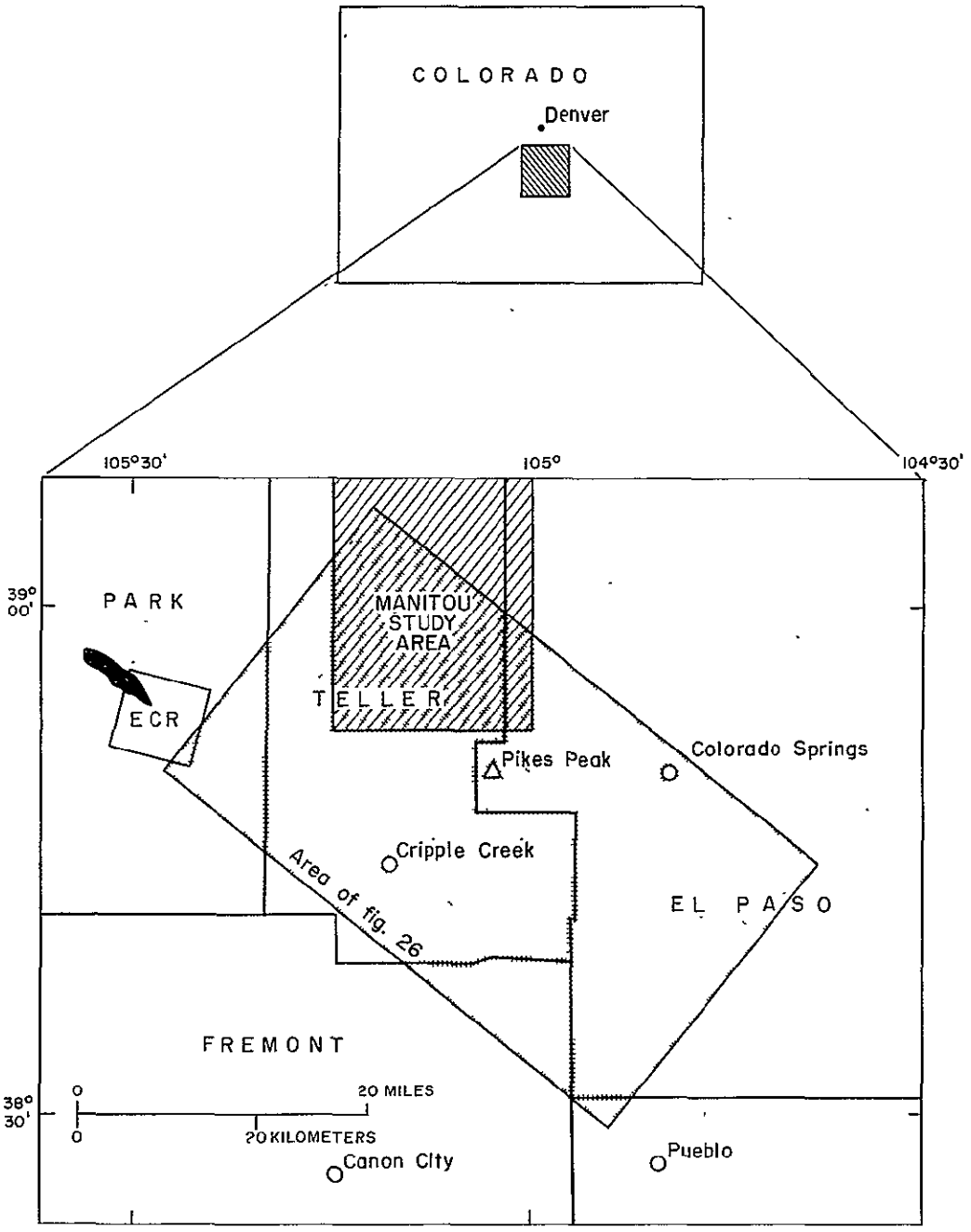


Figure 1. Location of test site in parts of Teller, El Paso, Park, Fremont, and Pueblo Counties, Colorado. Shaded area shows approximate limits of area covered by terrain map made from S-192 data (fig. 26). ECR, Elevenmile Canyon Reservoir Study Area; black area, Elevenmile Canyon Reservoir.

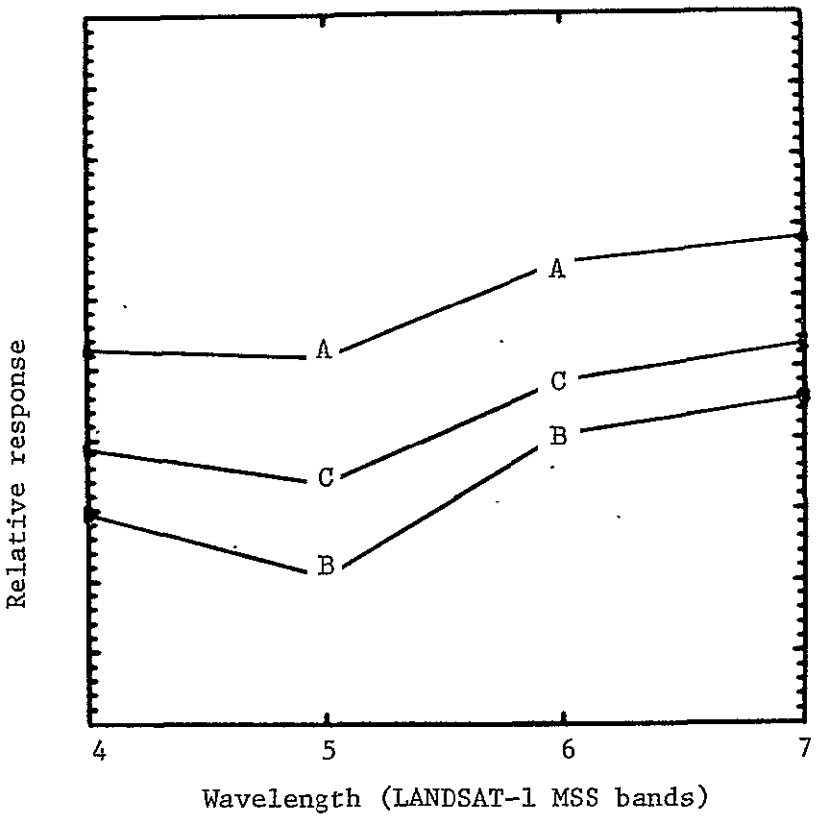


Figure 2. Spectral response of two-component terrain classes (A, B) and a mixture of them (C) extracted from LANDSAT-1 MSS data.

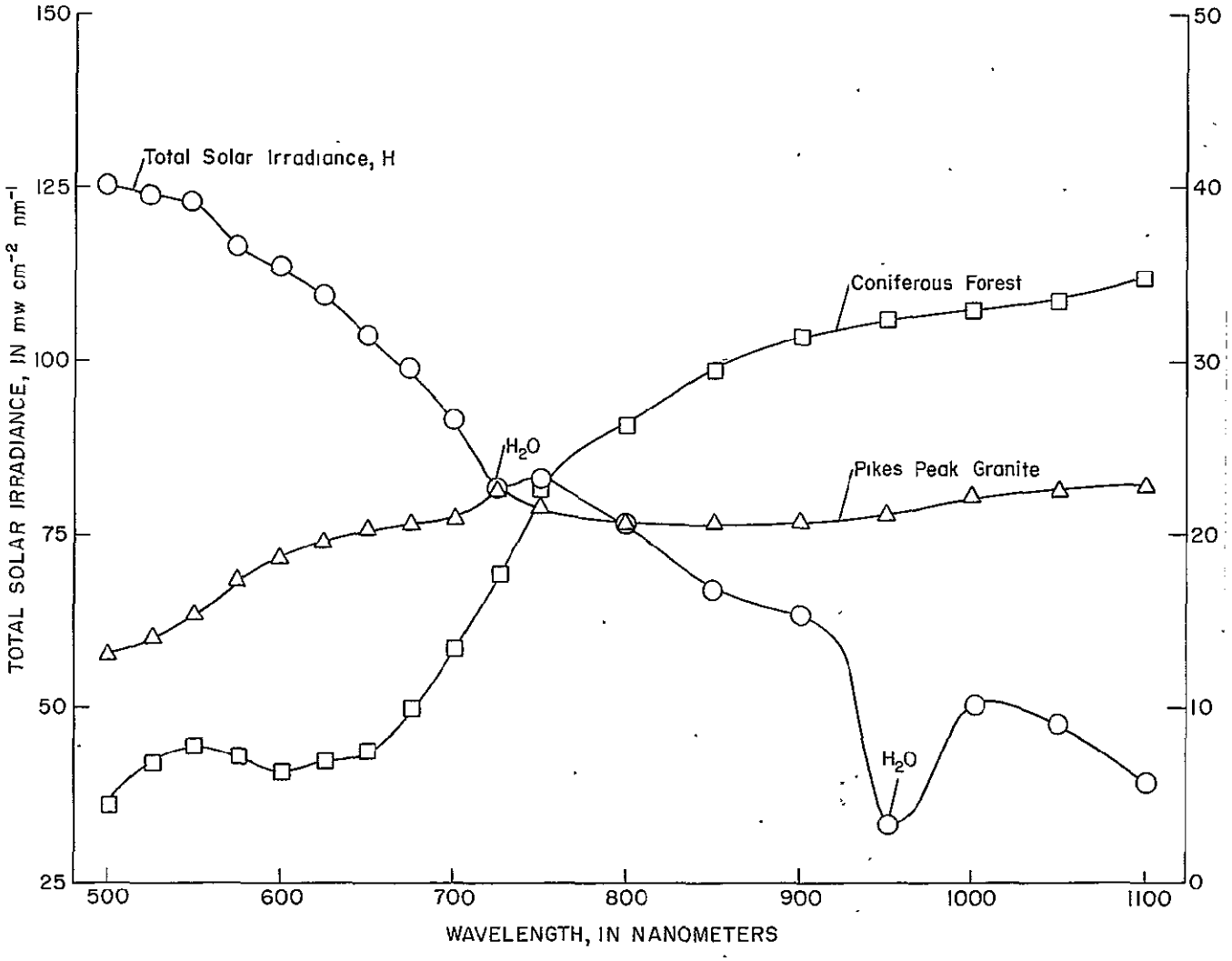


Figure 3. Comparison of total solar irradiance and reflectivity of Coniferous Forest and Pikes Peak Granite. Positions of absorption band shown for water vapor (H_2O).

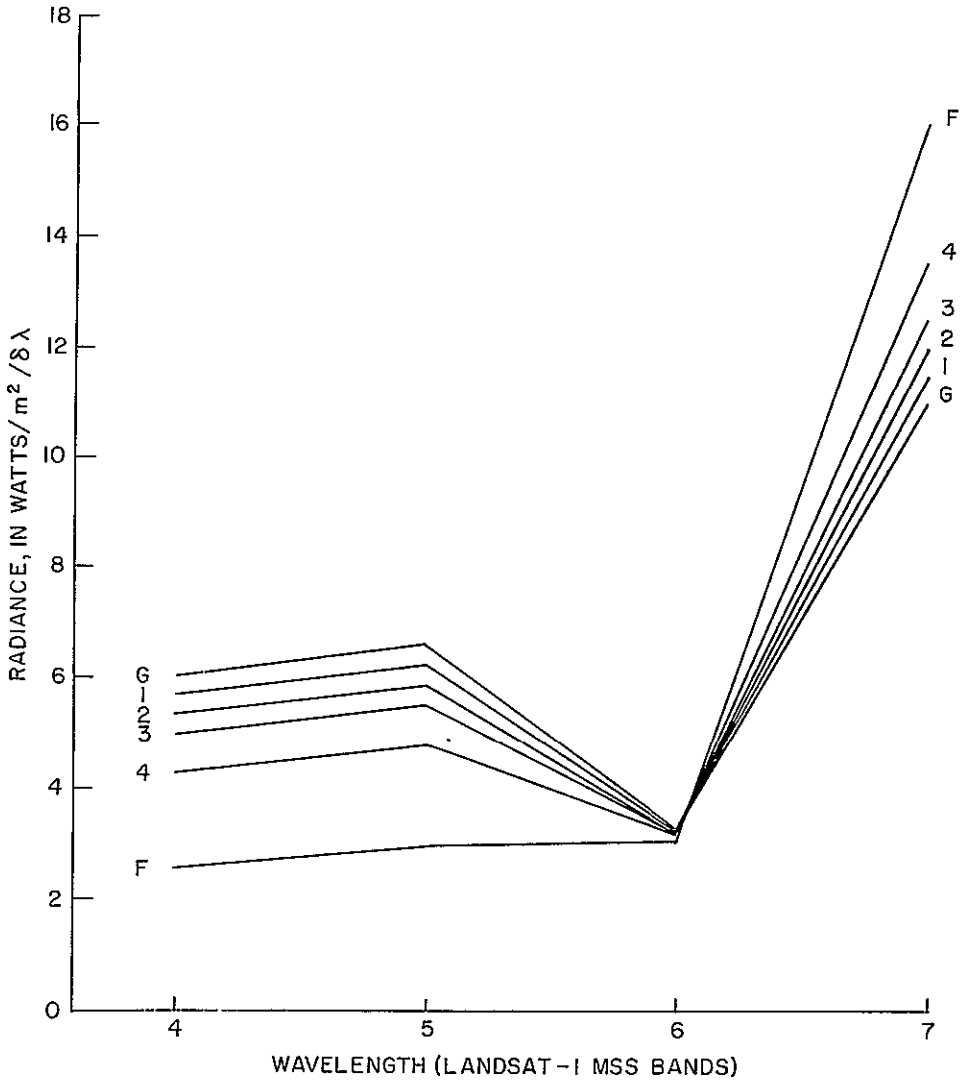


Figure 4. Spectral radiance of Pikes Peak Granite, Coniferous Forest, and mixtures of the two. Curves: G=100 percent Pikes Peak Granite, F=100 percent Coniferous Forest, 1=90 percent G-10 percent F, 2=80 percent G-20 percent F, 3=70 percent G-30 percent F, 4=50 percent G-50 percent F. Data from May 29, 1974; lat. 38.80° N., long. 105.26° W., solar elevation 64.51°, solar azimuth 126.62°, time 10:30 AM MST. Curves traced from computer printout.

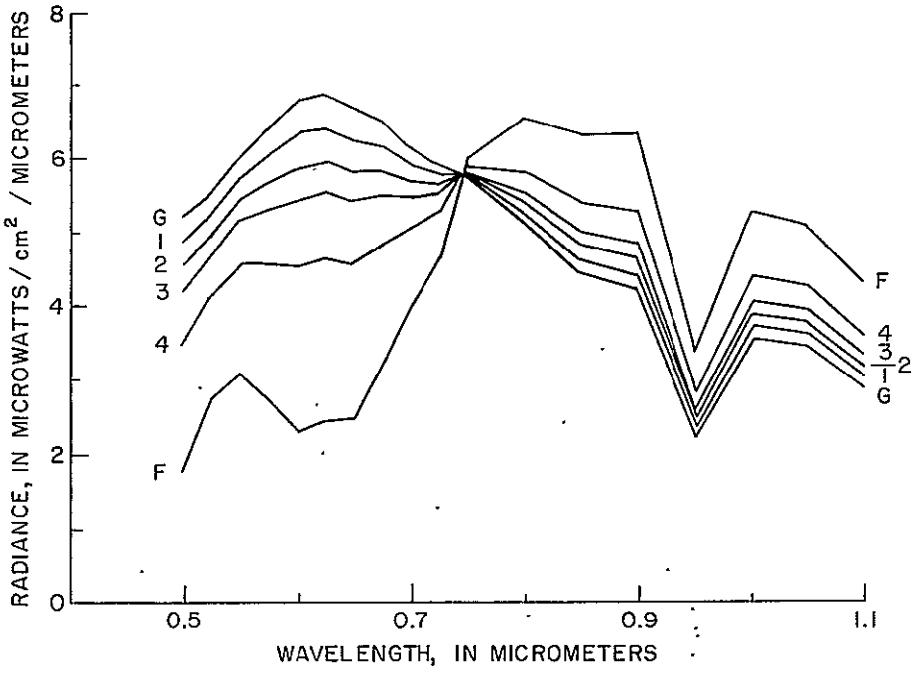


Figure 5. Spectral radiance of Pikes Peak Granite, Coniferous Forest, and mixtures of the two. Curves: G=100 percent Pikes Peak Granite, F=100 percent Coniferous Forest, 1=90 percent G-10 percent F, 2=80 percent G-20 percent F, 3=70 percent G-30 percent F, 4=50 percent G-50 percent F. Data from May 29, 1974; lat. 38.80° N., long. 105.26° W., solar elevation 64.51°, solar azimuth 126.62°, time 10:30 AM MST. Curves traced from computer printout.

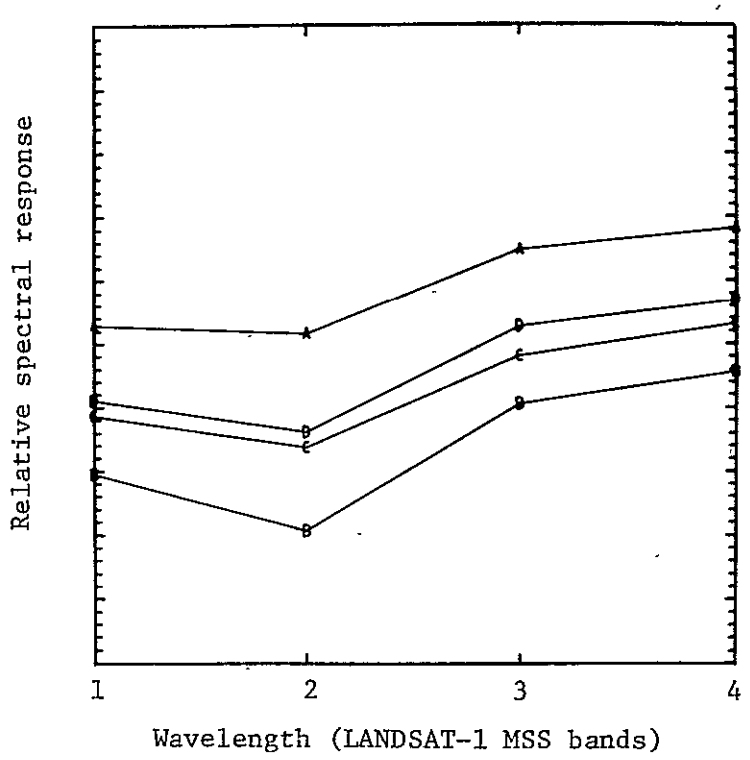


Figure 6. Spectral response of two component terrain classes (A,B), a mixture of them (C) extracted from LANDSAT-1 MSS data, and the mixture response (D) calculated using equation 3. A, B, and C are same data as in figure 2.

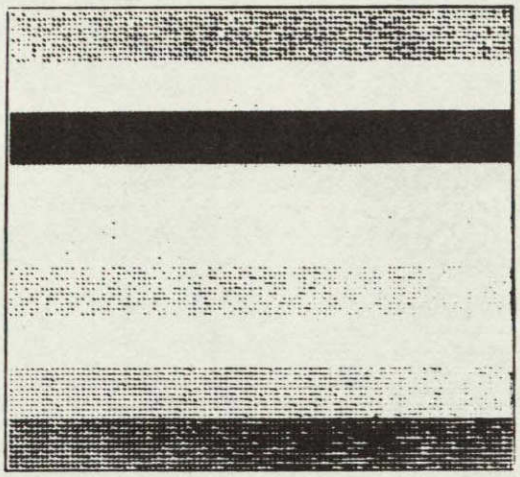


Figure 7. Microfilm graymap of nine-class simulated data set. Each horizontal band represents a field of simulated data for a given class, as labeled. F=Coniferous Forest, G-grassland, W=water.

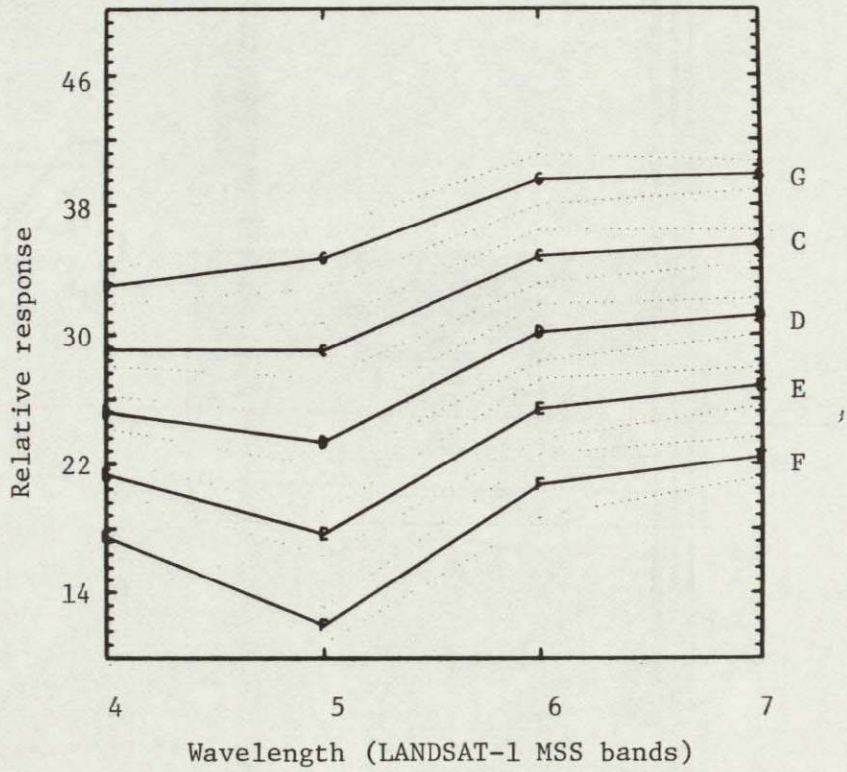


Figure 8. Mean spectral response for two component classes and modeled mixture classes. Curves: G=100 percent grassland, F=100 percent forest, C=75 percent G-25 percent F, D=50 percent G-50 percent F, E=25 percent G-75 percent F. Dotted lines represent standard deviations.

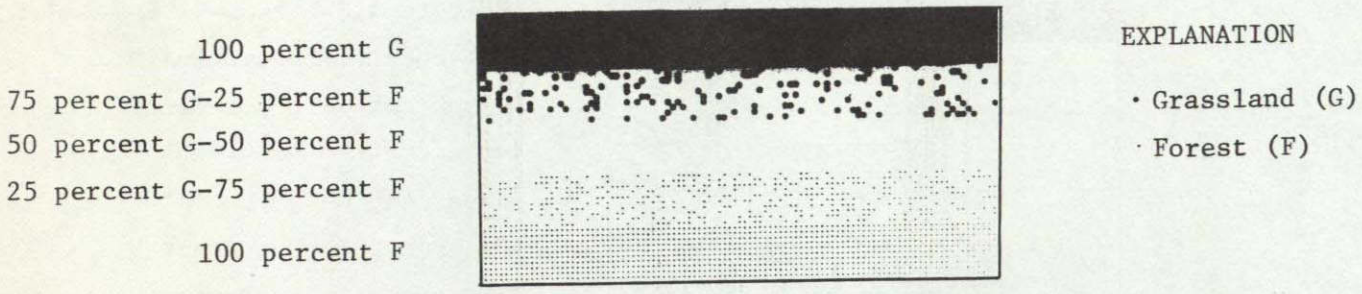


Figure 9. Microfilm classification display for component-class analysis. Blanks denote unclassified points.

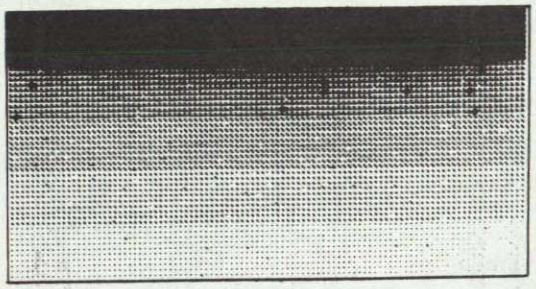


Figure 10. Microfilm classification display for five-class analysis.

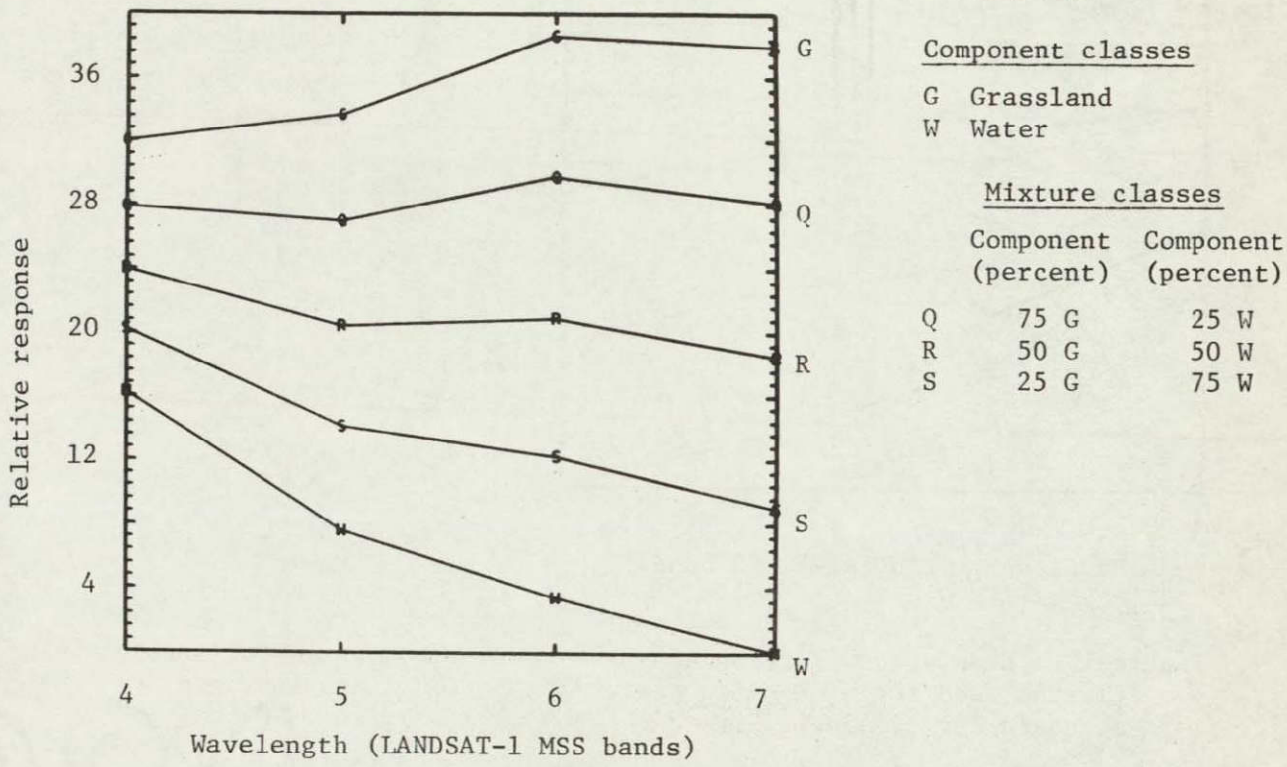
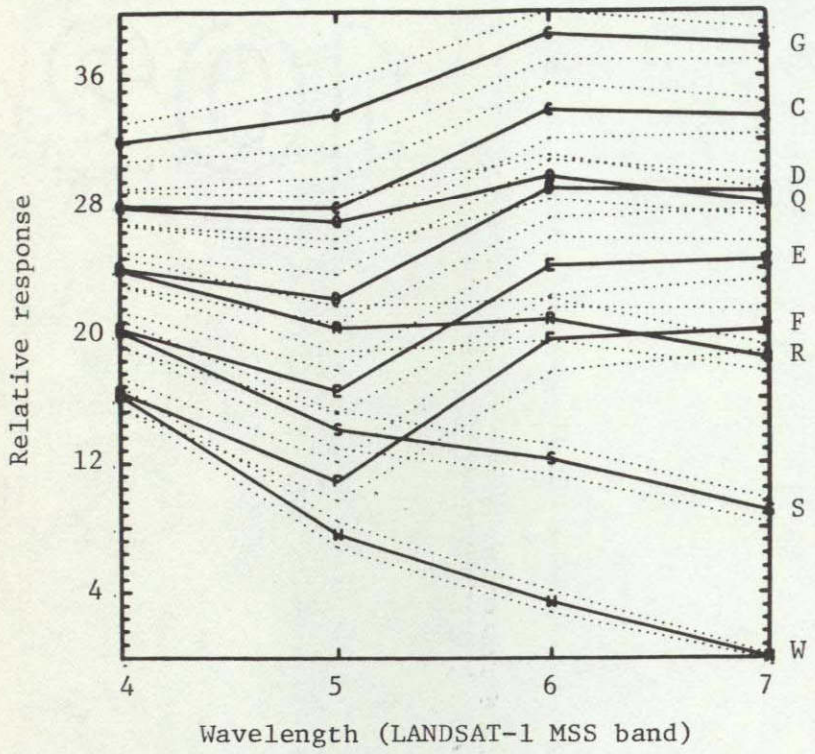


Figure 11. Mean spectral-response curves for simulated Grassland-Water mixtures. Dotted lines indicate standard deviations.



EXPLANATION

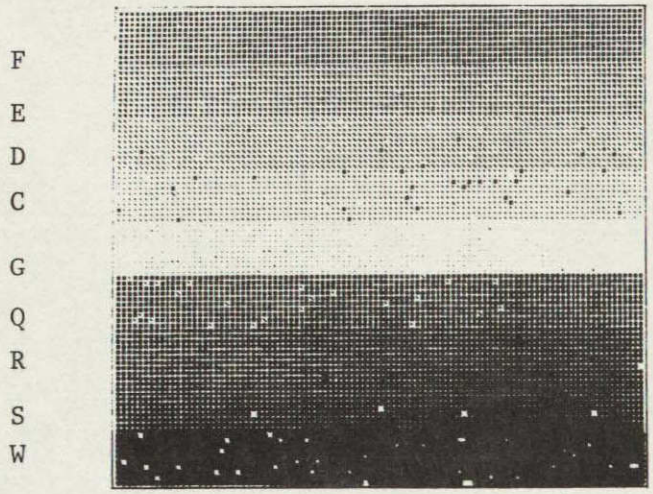
Component classes

G	Grassland
W	Water
F	Forest

Mixture classes

	Component (percent)	Component (percent)
C	75 G	25 F
D	50 G	50 F
E	25 G	75 F
Q	75 G	25 W
R	50 G	50 W
S	25 G	75 W

Figure 12. Mean spectral-response curves for component and mixture classes used for nine-class analysis. Dotted lines indicate standard deviations.



EXPLANATION

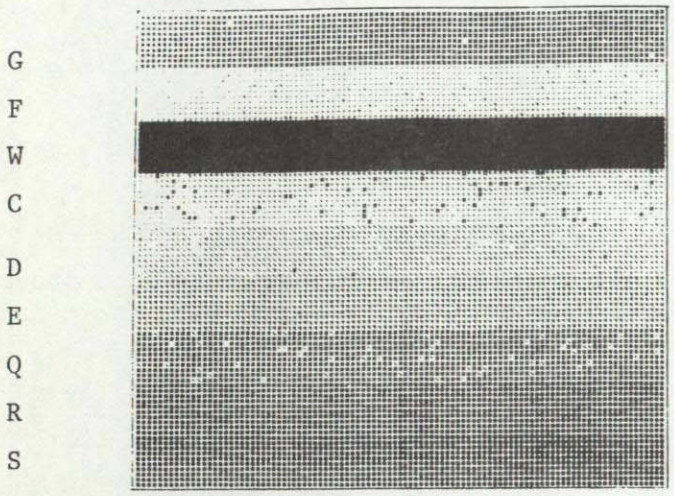
Component classes

G Grassland
W Water
F Forest

Mixture classes

	Component (percent)	Component (percent)
C	75 G	25 F
D	50 G	50 F
E	25 G	75 F
Q	75 G	25 W
R	50 G	50 W
S	25 G	75 W

Figure 13. Microfilm classification display of nine-class analysis. True class field symbols noted at left of display are same as in matrix on p. 14.



EXPLANATION

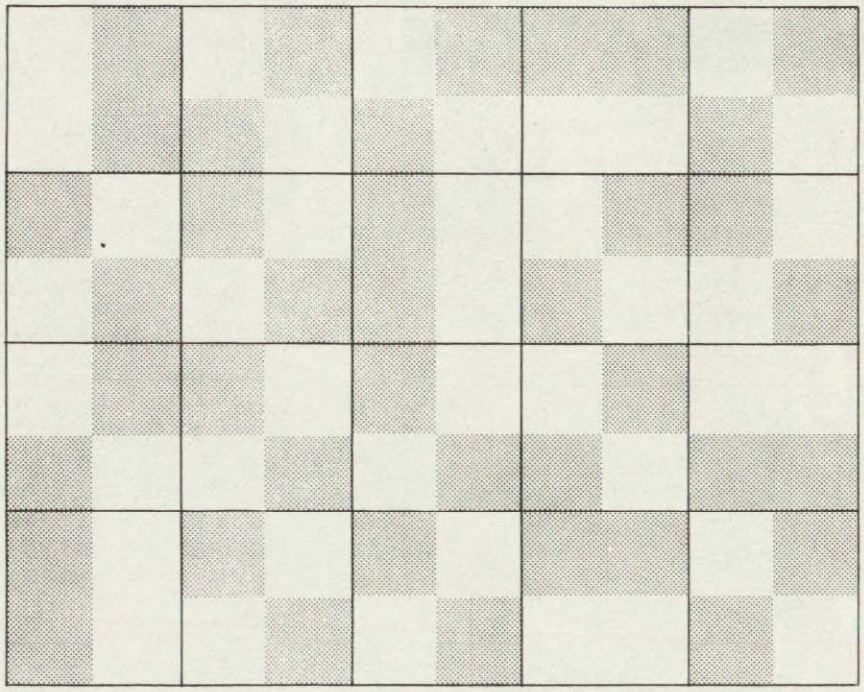
Component classes

G	Grassland
W	Water
F	Forest

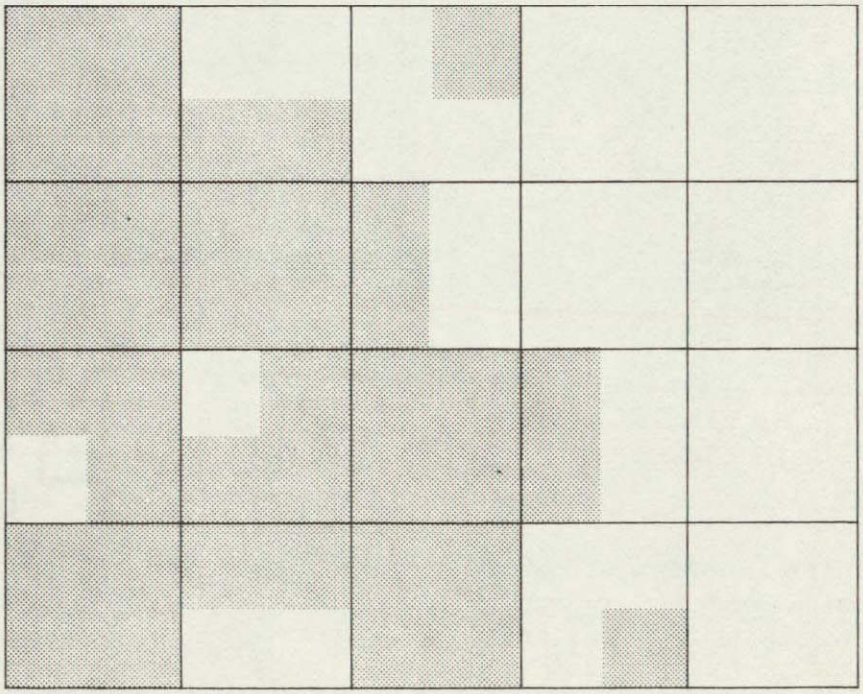
Mixture classes

	Component (percent)	Component (percent)
C	75 G	25 F
D	50 G	50 F
E	25 G	75 F
Q	75 G	25 W
R	50 G	50 W
S	25 G	75 W

Figure 14. Microfilm classification display for nine-class analysis using calculated component-class signatures from uniform data. True class field symbols noted on left of display are same as in matrix on p. 14.



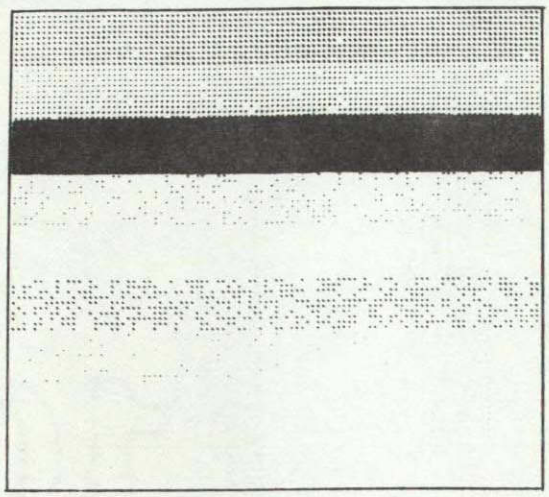
a



b

Figure 15. Graphical representation of uniform and nonuniform training sets. a) Uniform mixture of 50 percent White-50 percent Shaded. b) Nonuniform mixture of 50 percent White-50 percent Shaded.

a.



EXPLANATION

Component classes

G Grassland
W Water
F Forest

Blank areas are unclassified

Mixture classes

	Component (percent)	Component (percent)
C	75 G	25 F
D	50 G	50 F
E	25 G	75 F
Q	75 G	25 W
R	50 G	50 W
S	25 G	75 W

b.

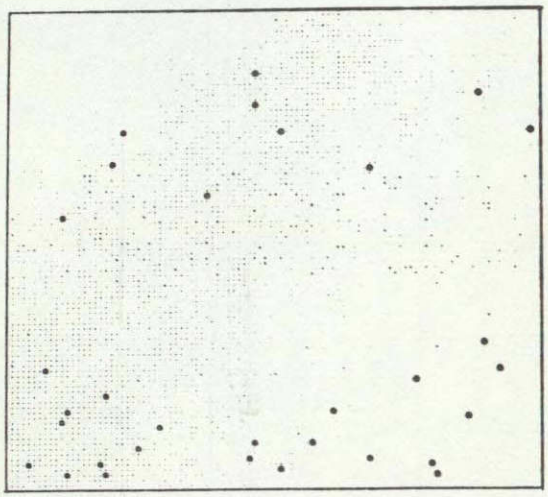
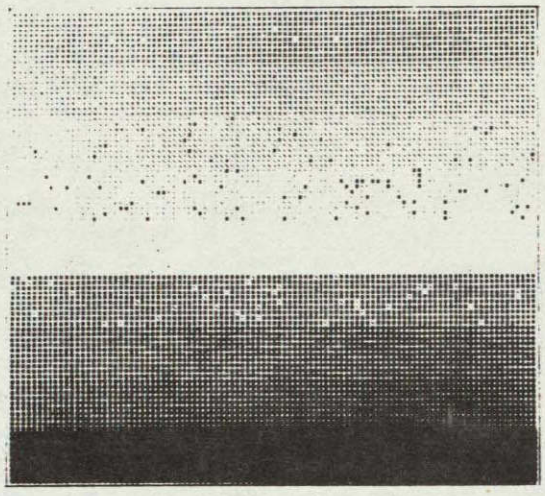


Figure 16. Microfilm classification displays for component-class analysis using estimated mean vectors and covariance matrices. a) Spectral signatures estimated from uniform training sets. b) Spectral signatures estimated from nonuniform training sets.

F
E
D
C
G
Q
R
S
W



EXPLANATION

Component classes

·	G	Grassland
·	W	Water
x	F	Forest

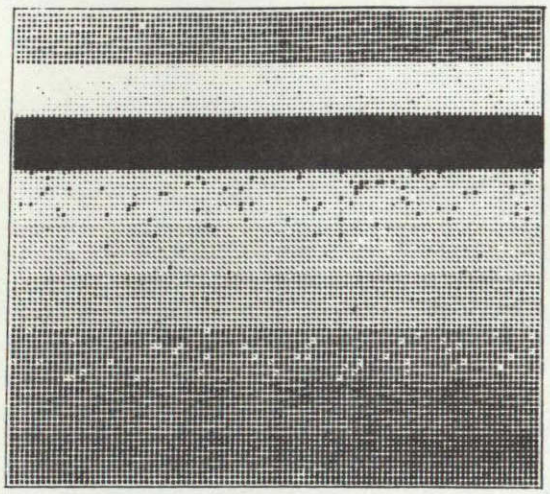
Blank areas are unclassified

Mixture classes

		Component (percent)	Component (percent)
`	C	75 G	25 F
`	D	50 G	50 F
·	E	25 G	75 F
·	Q	75 G	25 W
·	R	50 G	50 W
·	S	25 G	75 W

Figure 18. Microfilm classification display of nine-class analysis using calculated mean vectors and common covariance matrix. True field class symbols noted at left of display are same as in classification-confusion matrix on p. 14.

F
G
W
C
D
E
Q
R
S



EXPLANATION

Component classes

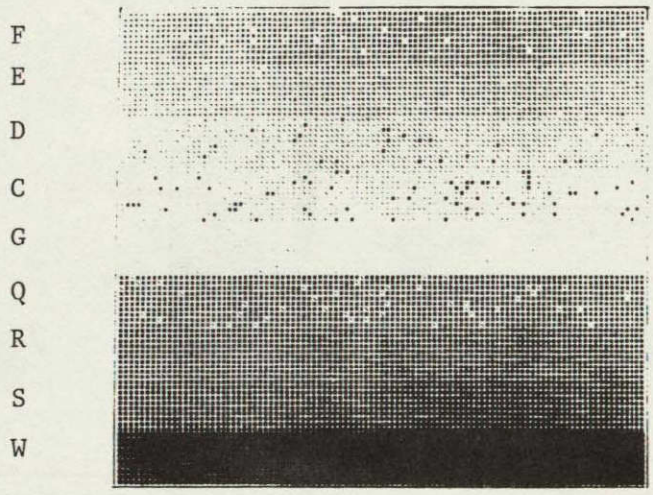
G Grassland
W Water
x F Forest

Blank areas are unclassified

Mixture classes

	Component (percent)	Component (percent)
C	75 G	25 F
D	50 G	50 F
E	25 G	75 F
Q	75 G	25 W
R	50 G	50 W
S	25 G	75 W

Figure 17. Microfilm classification display of nine-class analysis using calculated component-class mean vectors and common covariance matrix.



EXPLANATION

Component classes

. G Grassland
• W Water
x F Forest

Blank areas are unclassified

Mixture classes

	Component (percent)		Component (percent)	
`	C	75 G	25	F
'	D	50 G	50	F
~	E	25 G	75	F
^	Q	75 G	25	W
"	R	50 G	50	W
°	S	25 G	75	W

Figure 19. Microfilm classification display for nine-class analysis using calculated mean vectors and averaged common covariance matrix. True class field noted at left of display are same as in matrix on p. 14.

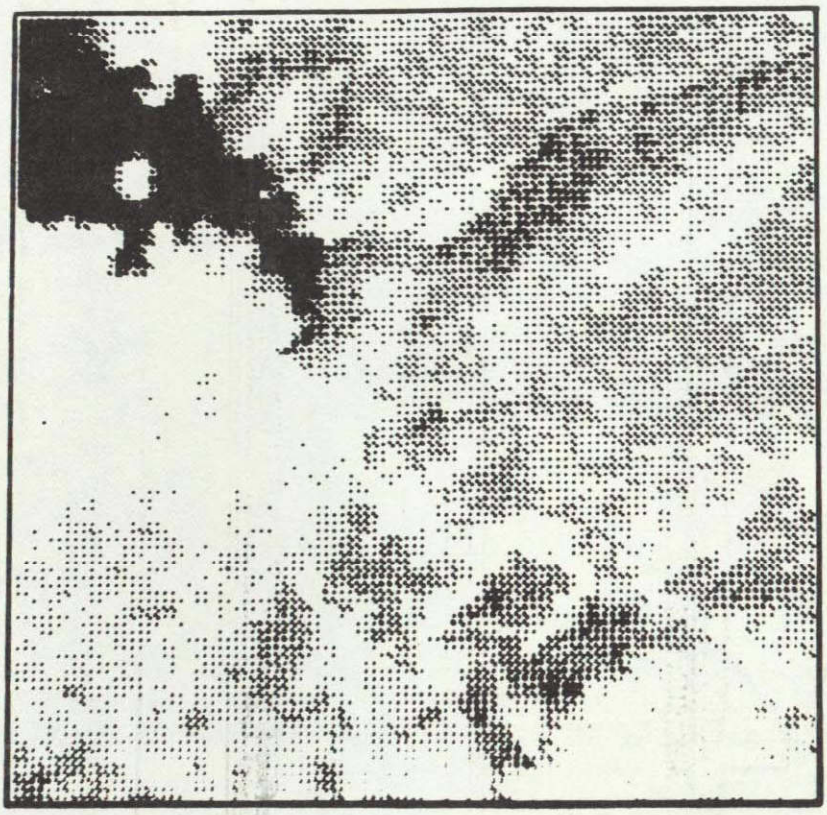


Figure 20. Microfilm graymap of LANDSAT-1 MSS Band 5 of the Elevenmile Canyon Reservoir Study Area. For location see figure 1. Dark area is east part of Elevenmile Canyon Reservoir, containing island.

EXPLANATION

· Forest
\ Grassland
· Water
\ Mountain Grassland
• Wet Meadow

Blank areas are unclassified.

Figure 21. Microfilm classification display of Elevenmile Canyon Reservoir Test Area with five component classes.

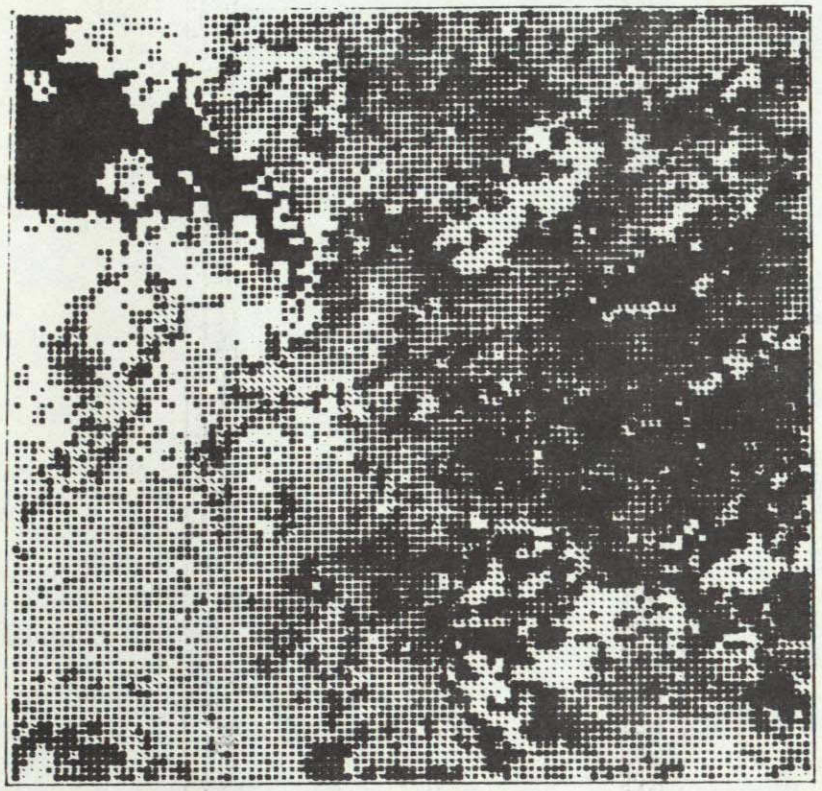


Figure 22. Microfilm classification map of Elevenmile Canyon Reservoir Study Area with five component classes and nine mixture classes. Because of limitations of the microfilm display, separate symbols could not be shown for the mixtures (x); however, for study purposes they were printed separately by a conventional line printer.

EXPLANATION

Component classes

- G Grassland
- W Water
- F Forest
- MG Mountain Grassland
- WM Wet Meadow

Mixture classes

Symbol	Component (percent)	Component (percent)
x	75 G	25 F
x	75 G	25 WM
x	75 F	25 MG
x	50 G	50 F
x	50 G	50 WM
x	50 F	50 MG
x	25 G	75 F
x	25 G	75 WM
x	25 F	75 MG

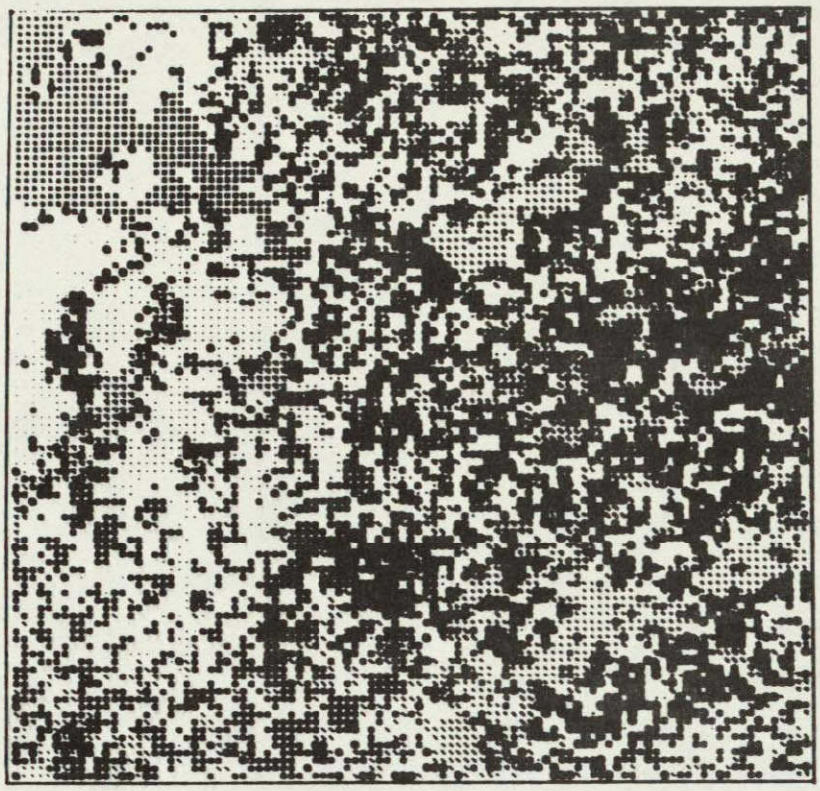


Figure 23. Microfilm classification map of Elevenmile Canyon Reservoir Study Area with five component classes and ten mixture classes. Because of limitations of microfilm display, separate symbols could not be shown for the mixtures (x); however, for study purposes they were printed separately by a conventional line printer.

EXPLANATION

Component classes

·	G	Grassland
x	W	Water
`	F	Forest
`	MG	Mountain Grassland
·	MW	Wet Meadow

blank areas are unclassified

Mixture classes

Symbol	Component (percent)	Component (percent)
·	67 G	33 F
·	67 G	33 W
x	67 WM	33 W
x	33 G	67 W
x	67 G	33 WM
x	33 WM	67 W
x	33 G	67 F
x	67 F	33 MG
x	33 F	67 MG
x	33 G	67 WM

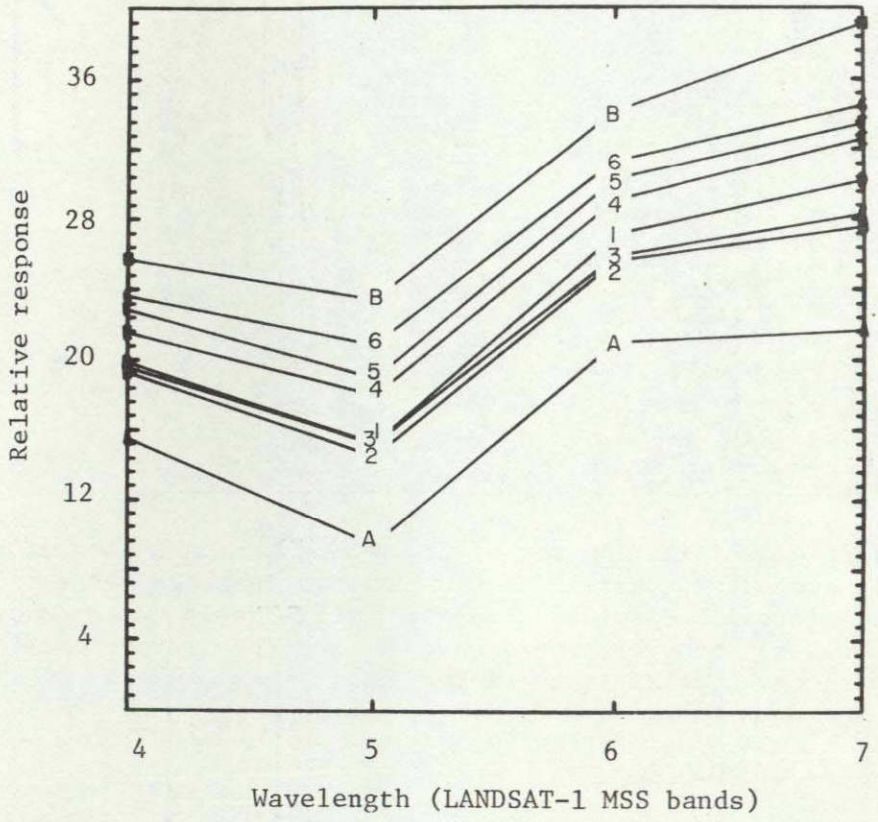
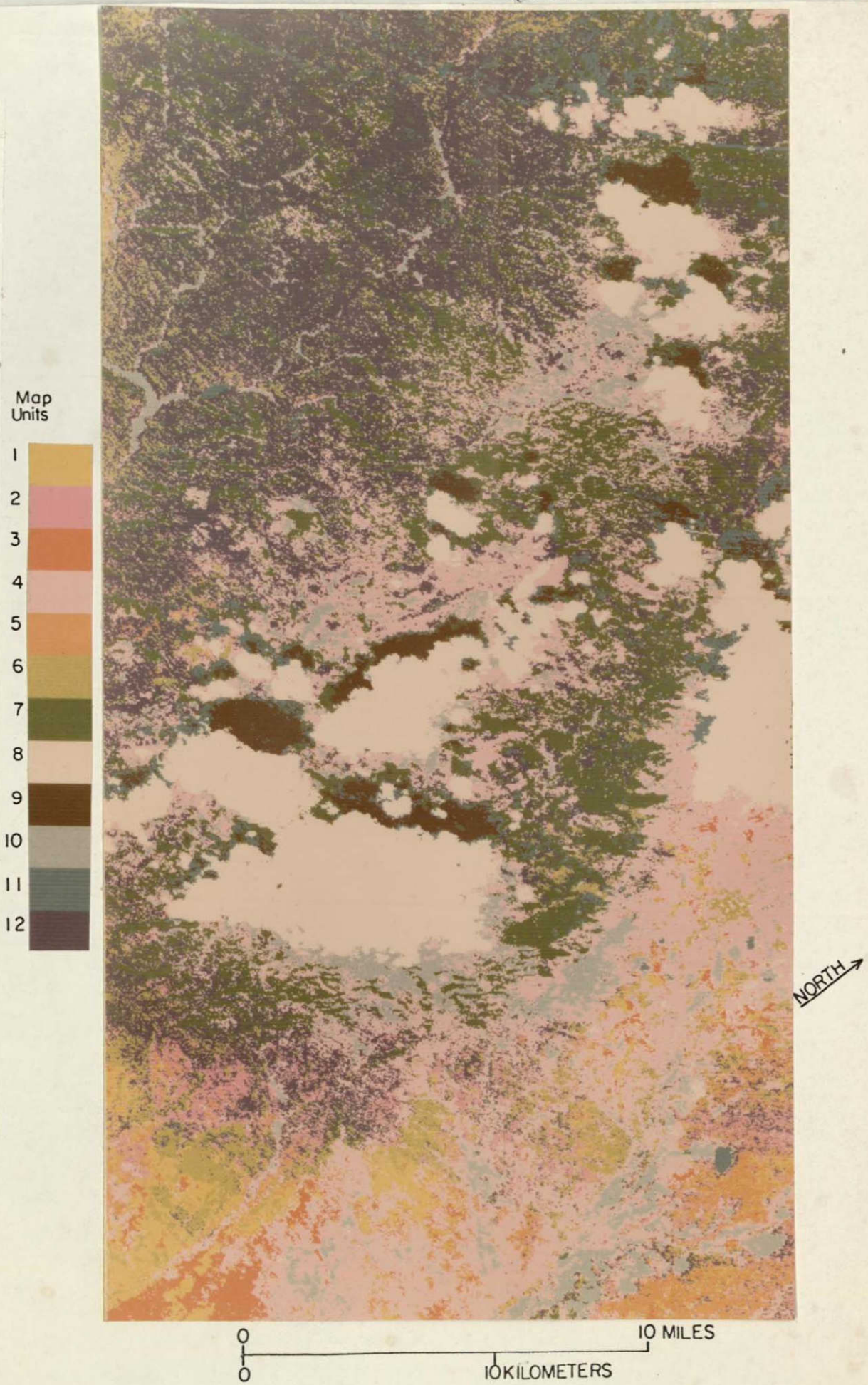


Figure 24. Mean spectral-response curves of the Manitou Study Area for estimated Ponderosa pine and Background (curves A and B, respectively) and extracted mixtures. Numbers refer to plots listed on p. 30.

Figure 25. (opposite) Color-coded terrain classification map of test site shown in figure 1. Map units (see numbered color patches on side of map) are:

1. Dakota Sandstone
2. Fountain Formation
3. Niobrara Shale
4. Pikes Peak Granite
5. Pierre Shale
6. Dakota Sandstone and vegetation composite
7. Forest
8. Unclassified (largely clouds)
9. Cloud shadow
10. Meadow
11. Water
12. Volcanic and plutonic rocks, undivided

For more detailed description of the map units, see table 1 and reference 26.


Map Units
1
2
3
4
5
6
7
8
9
10
11
12
NORTH
0
10 MILES
0
10 KILOMETERS