NASA CONTRACTOR REPORT 166569

(NASA-CR-166569) AN EVALUATION OF THEMATIC MAPPER SIMULATOR DATA FOR THE GEOBOTANICAL DISCRIMINATION OF ROCK TYPES IN SOUTHWEST OREGON (Technicolor Government Services, Inc.) 63 p HC A04/MF A01 CSCL 08B G3/42
N84-21915
Unclas
13000

An Evaluation of Thematic Mapper Simulator Data for the Geobotanical Discrimination of Rock Types in Southwest Oregon

K.J. Weinstock
L.A. Morrissey

CONTRACT NAS2- 11101
January 1984

NASA CONTRACTOR REPORT 166569

ORIGINAL CONTAINS
COLOR ILLUSTRATIONS

An Evaluation of Thematic Mapper Simulator Data for the Geobotanical Discrimination of Rock Types in Southwest Oregon

K.J. Weinstock
L.A. Morrissey
Technicolor Government Services, Inc.
Ames Research Center
Moffett Field, California

Prepared for
Ames Research Center
under Contract NAS2-11101

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

The Geobotany project benefitted substantially from the participation and support of the following individuals and agencies: David A. Mouat, Principal Investigator; Don H. Card, Statistician and Elizabeth Horn, Applications Scientist, NASA/Ames Technology Applications Branch; and Robert E. Frenkel and Chris Kiilsgaard, Oregon State University. We would also like to thank Mark Settle, NASA Non-Renewable Resources Program Manager; Len Ramp, Oregon Department of Geology and Mineral Industries; Don L. McClennan, and Dave Galloway, U.S. Forest Service.

I. INTRODUCTION

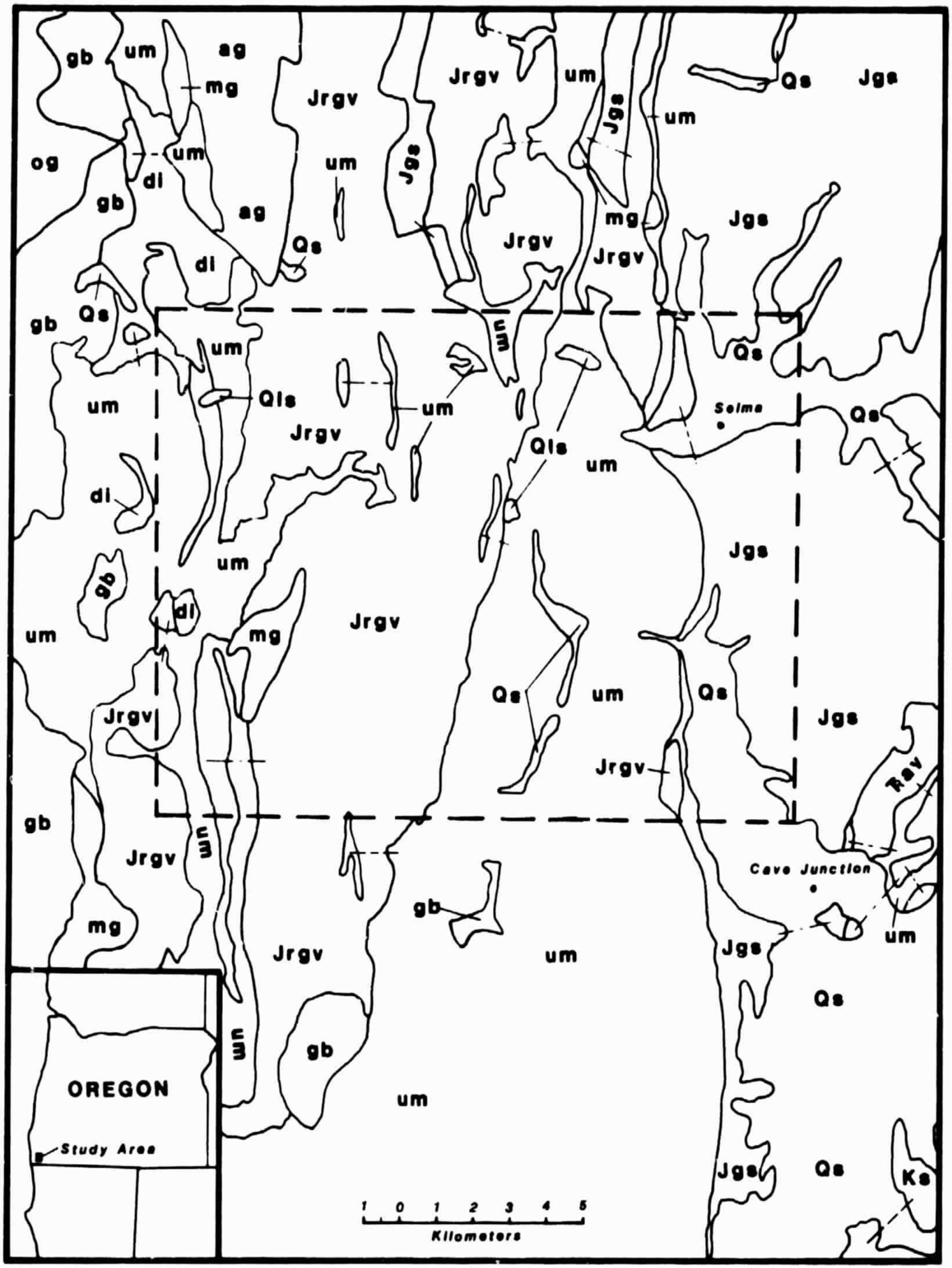
The utility of Thematic Mapper Simulator (TMS) data in the discrimination of rock types, including those associated with known chromite and nickel deposits, has been investigated by scientists at NASA/Ames Research Center. The identification of rock types using current mineral exploration technology often involves destructive and time-consuming techniques. Rock-type identification may be assisted by the use of remote sensing of associated vegetation, particularly in areas of dense vegetative cover where surface materials are not imaged directly by the sensor. The present research involves the geobotanical discrimination of ultramafic parent materials and is part of a larger effort to develop and test analytical techniques for lithologic mapping and mineral exploration.

The goal of the research being reported here was to evaluate the utility of remotely sensed data to discriminate vegetation types associated with ultramafic parent materials in a study area in southwest Oregon. To accomplish this, a number of specific objectives were identified. These included: 1) establishment of the association between vegetation and rock types, 2) examination of the spectral separability of vegetation types associated with rock types, 3) determination of the contribution of each TMS band for discriminating vegetation associated with rock types and 4) comparison of analytical techniques for spectrally classifying vegetation [7] [8].

STUDY AREA

A study area of approximately 200 square kilometers, located within the Siskiyou Mountains of southwestern Oregon was selected because it is characterized by a wide diversity of rock types and vegetation. It also contains known deposits of chromium and nickel associated with ultramafic rock types.

The geology of the area is one of pre-Tertiary sediments and volcanics that have been folded, faulted and intruded by serpentinized masses of ultramafic and granitic rocks (see Figure 1). The sedimentary rocks have been partially metamorphosed and consist of slaty shales and siltstones. The volcanics are also partially metamorphosed and consist of thick breccias and tuffs as well as andesitic flow rocks.



SOUTHWEST OREGON STUDY AREA

Qs	QUATERNARY SEDIMENTS	ag	GNEISSIC AMPHIBOLITE
Qls	LANDSLIDE DEPOSITS	di	QUARTZ DIORITE
Ks	MARINE SEDIMENTARY ROCK	gb	GABBRO
Jgs	METASEDIMENTARY ROCK	mg	METAGABBRO
Jrgv	METAVOLCANIC ROCK	og	OLIVINE GABBRO
T̄Rav	METASEDIMENTARY ROCK	um	ULTRAMAFICS

Figure 1. Simplified geology of the study area and test site.

]

The climate of the study area is strongly affected by maritime air from the Pacific Ocean. Summers are mild and dry with mean temperatures of approximately 20 degrees Celsius; winters are cool and wet with mean temperatures of approximately 5 degrees Celsius. Precipitation often exceeds 2,000 millimeters. Snowfall is dependent on elevation with heavy accumulations above 1000 meters. Temperature and snow accumulation are affected by aspect as well as elevation, and the effect on the vegetation is often dramatic[4].

The vegetation of the study area is extremely diverse partly as a result of a myriad of microclimates and partly as a result of the diverse lithology, which can be characterized generally as ultramafic and non-ultramafic. The geochemistry of ultramafic parent materials has profound implications on plant distribution and composition. The resulting soils, largely derived from weathered olivine, are rich in iron and magnesium silicates but low in calcium. Typically, these soils have extremely low calcium to magnesium ratios which may affect the vegetation through minimal calcium nutrient uptake and/or magnesium toxicity. The presence of heavy metals (chromium and nickel) in these soils may also have a toxic influence on the vegetation[3]. At the lowest elevations, vegetation on ultramafic rocks (hereafter referred to as ultramafic vegetation) differs from the surrounding vegetation in that virtually no broadleaf tree species are found in the overstory. On more intermediate sites, the ultramafic vegetation consists of a very distinctive open forest of several conifer species with a shrub and grass understory. The non-ultramafic vegetation consists of a denser conifer forest with a considerable component of broadleafed trees. A report being prepared by Frenkel and Kiilsgaard [2] contains a detailed discussion of the vegetation of the study area.

II. DATA COLLECTION

The acquisition of several types of data was required for the digital analysis of the study area. Included among these were the following:

- o Thematic Mapper Simulator (TMS) digital data
- o Landsat digital data (in the form of CCT's)
- o Field data
- o Color infrared (CIR) U-2 and C-130 photography
- o Topographic maps and orthophotoquads
- o Geologic map and other ancillary map information

DIGITAL DATA

Digital data from several sensors was acquired for use in the analysis of geobotanical relationships within the study area. A Daedalus DEI-1260 Multispectral Scanner acquired Thematic Mapper Simulator (TMS) data for the study area during three consecutive summers from 1981 to 1983. The TMS data acquired during July, 1982 provided the foundation for the bulk of the digital analysis work completed during the project (Figure 2). The 1981 flight (81-171) was a pre-TMS 10-channel configuration which was only

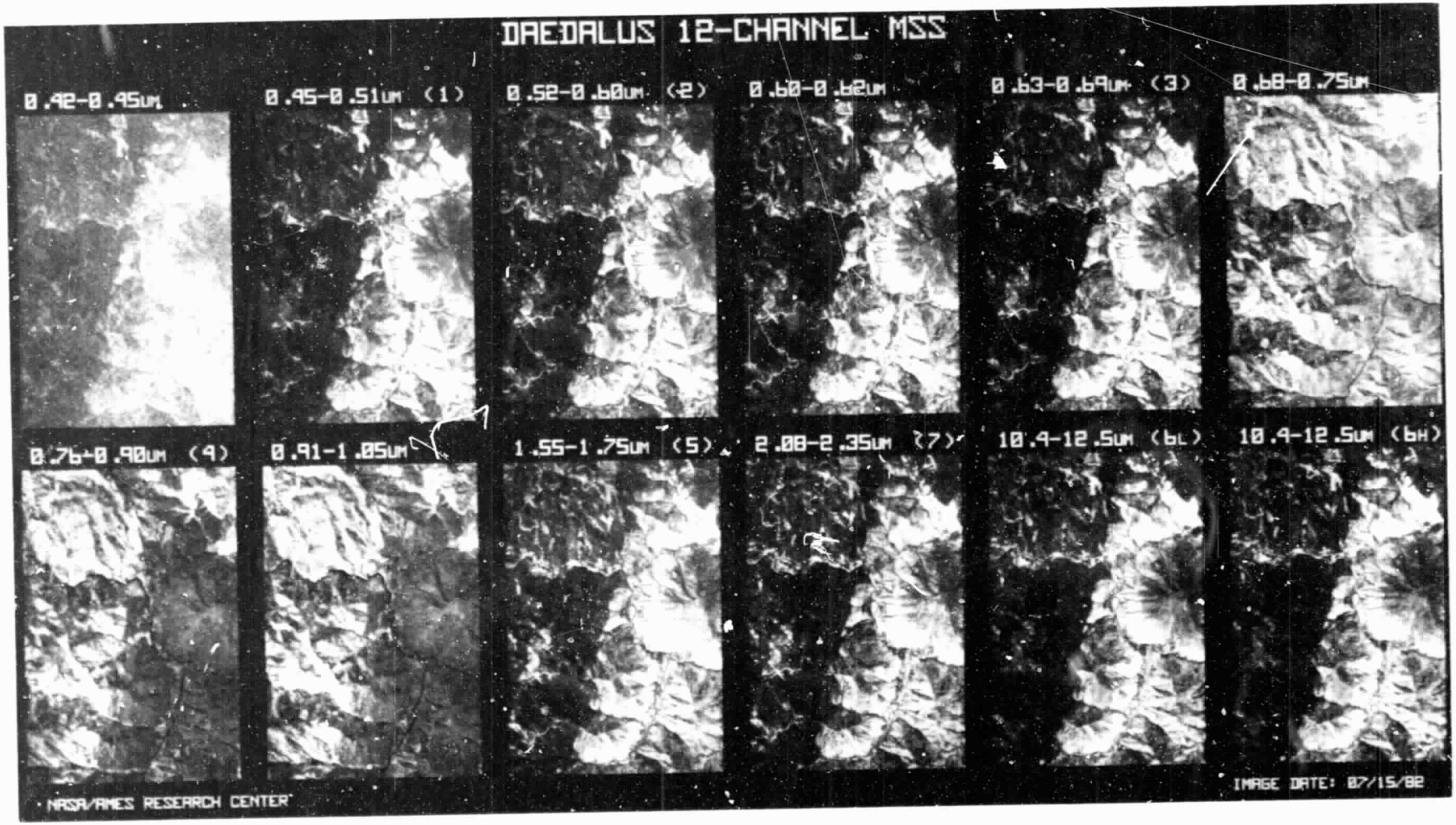


Figure 2. Twelve channel Daedalus scanner data. Note: TMS bands are indicated in parentheses.

marginally useful because it did not adequately simulate the spectral bandwidths on the 7-channel Thematic Mapper. The 1983 flight (83-143) was acquired after the majority of the data analysis had been completed. Furthermore, the data quality from this flight was poor due to morning haze which resulted in an apparent shift to increased DN (digital number) values at the left margin of the image. In addition to the Daedalus scanner data, TMS data was also acquired by the NS001 scanner in May, 1983. This data, although of higher spatial resolution, was not usable due to severe relief displacement related to the lower altitude of the C-130 platform.

The Thematic Mapper (TM) sensor aboard Landsat-4 was originally intended to be the primary digital data source for the project. Multiple failures of the satellite communications hardware and the subsequent delay in the Tracking Data and Relay Satellite (TDRS) achieving operational status precluded the use of TM digital data during the project. One Landsat-2 MSS scene (ID 82241318120X0) in Computer Compatible Tape (CCT) format was acquired for use in reconnaissance of the pilot study area.

FIELD DATA

The collection of ground data is an essential part of the identification of spectral classes generated by a multispectral classifier. In an unsupervised clustering approach, ground data provides the basis from which spectral classes can be assigned to information classes. As an unsupervised clustering approach was utilized for all classifications generated during the project, the acquisition of ground data that was complete and as accurate as possible was of utmost importance.

The responsibility for primary field data collection was given to R. E. Frenkel and C. Kiilsgaard of Oregon State University (OSU) under a NASA Consortium Agreement (final report in press). The OSU investigators were joined by TGS and NASA/Ames personnel in the field during three consecutive summers from 1981 to 1983. Field data collected in 1981 and 1982 was limited due to the inaccessibility, steep terrain and large areal extent of the study area. After completion of the 1982 field work, a generalized vegetation map of the study area and a vegetation classification system (Table 1) were completed by OSU.

Table 1. A systematic classification of vegetation for the Central Siskiyou Mountains.

- 120 FOREST
  - 122 SIERRAN-CASCADE MONTANE CONIFER FOREST
    - 122.5 CONIFER FOREST
      - 122.51 MIXED CONIFER SERIES
        - 122.511 TRUE FIRS (WHITE, SHASTA RED, GRAND)
        - 122.512 JEFFREY PINE, INCENSE CEDAR, DOUGLAS FIR
        - 122.513 PORT ORFORD CEDAR, DOUGLAS FIR, WESTERN HEMLOCK
        - 122.514 DOUGLAS FIR, PONDEROSA PINE, SUGAR PINE
        - 122.515 KNOBCONE PINE, LODGEPOLE PINE
    - 122.7 BROADLEAF EVERGREEN FOREST
      - 122.71 MIXED BROADLEAF EVERGREEN SERIES
        - 122.711 CHINKAPIN, TAN OAK, PACIFIC MADRONE, CANYON LIVE OAK
    - 122.8 BROADLEAF DECIDUOUS FOREST
      - 122.81 MIXED BROADLEAF-DECIDUOUS SERIES
        - 122.811 OREGON WHITE OAK, CALIFORNIA BLACK OAK
        - 122.812 BIGLEAF MAPLE, RED ALDER, OREGON ASH
    - 122.9 CONIFER-BROADLEAF FOREST
      - 122.91 MIXED CONIFER-BROADLEAF SERIES
        - 122.911 DOUGLAS FIR, CANYON LIVE OAK, CALIFORNIA BLACK OAK, PACIFIC MADRONE
        - 122.912 DOUGLAS FIR, CHINKAPIN, TAN OAK, PACIFIC MADRONE
        - 122.913 DOUGLAS FIR, TAN OAK, PACIFIC MADRONE
        - 122.914 DOUGLAS FIR, CANYON LIVE OAK
- 130 SHRUB
  - 132 SIERRAN-CASCADE MONTANE SHRUBLAND
    - 132.21 MANZANITA SERIES
      - 132.211 GREEN MANZANITA, WHITE LEAVED MANZANITA, HUCKLEBERRY OAK, CEANOTHUS
      - 132.212 CALIFORNIA LAUREL, CALIFORNIA COFFEEBERRY, BOX-LEAVED GARRYA
      - 132.222 TAN OAK, BLACKBERRY, VINE MAPLE, OCEAN SPRAY
- 140 GRAMINOID
  - 142 SIERRAN-CASCADE MONTANE GRASSLAND
    - 142.31 MIXED BUNCHGRASS SERIES
      - 142.311 GRASSES (FESCUE, CHEAT, QUACK, WILD RYE), SEDGE

In order to maximize the efficiency and accuracy of field data collection, the study area and field data collection methodology were modified after the 1982 summer field session. The study area was reduced from approximately 200 square kilometers to approximately 15 square kilometers centered upon the Sourdough Flat area in the western part of the original test site. The area surrounding Sourdough Flat contains a good sample of most vegetation/geologic associations that had been identified within the overall study area. The increased accessibility and reduction in areal extent permitted a more thorough and systematic sampling procedure to be utilized during the 1983 field work.

The positional accuracy of sampling units mapped by the OSU investigators in 1981 and 1982 was limited by the lack of large-scale topographic maps and suitable large-scale aerial photography for the area. In late May, 1983, a C-130 aircraft from NASA/Ames acquired high resolution color infrared (CIR) photography (flight 83-093A) for the entire study area utilizing an IRIS 2 panoramic camera system. The use of this photography led to an increase in accuracy in plotting the location of the ground sampling units, subsequently verified during the 1983 summer field work.

The type of data collected in the field and the methodology of data collection were also revised prior to the 1983 field season. Previously, data collection had been limited to a traditional floristic vegetation sampling scheme centering upon species occurrence and dominance. For the 1983 work, in addition to the traditional vegetation sampling scheme, NASA/Ames investigators instituted the collection of surficial geologic and soil samples, and other soil variables including color, texture, pH and profile development. An estimation of total coverage by cover type was included in the field data collection process in order to record the composite interaction of the landcover components to achieve a better understanding of the effect of these components on the spectral reflectance at a particular site. This helped to avoid the problems of confusion and/or poorly-defined spectral classes in the post-field spectral class labeling process. A sample field form is shown in Appendix A.

ANCILLARY MAP DATA

Several types of published map data were acquired to support research efforts. Standard USGS 15-minute topographic quadrangles were used for project planning and reconnaisance, as well as providing generalized elevation information. Standard black-and-white diazo copies of 7.5-minute orthophotoquads served as base maps for the generation of the vegetation maps completed by OSU. Generalized geologic information was obtained at scales ranging from 1:48,000 to 1:100,000 (State of Oregon-Department of Geology and Mineral Industries). More detailed geologic information was available for very limited sections of the study area.

III. DATA ANALYSIS

COMPUTER SYSTEMS

Digital analysis of the Eight Dollar Mountain TMS scene utilized computer systems at NASA/Ames Research Center. The availability of a variety of such systems allowed use of the most efficient machine for each particular processing step. Initial entry of the TMS digital data from Computer Compatible Tapes, subsequent extraction of the Eight Dollar Mountain subscene, preprocessing and unsupervised clustering was accomplished using the IDIMS (Interactive Digital Image Manipulation System) software implemented on an HP 3000 Series III minicomputer. The IDIMS software (developed by ESL, Inc.) was also resident on an HP 3000 Series 33 minicomputer which was utilized for most functions requiring the use of a color display, and for the generation of spectral statistics for sample areas within the scene. Generation of the reference data set utilized specially written modules within the ELAS software package which were implemented on a SEL 32/77 minicomputer. Classification of the scene was completed using a maximum-likelihood classifier on the CDC-7600 mainframe at the Ames Central Computing

Facility. Principal components transformations and band-ratios were generated using the VICAR (Video Image Communication and Retrieval) software developed at the Jet Propulsion Laboratory. VICAR software is resident on an IBM 4341 mainframe at the ARC Central Computing Facility.

PREPROCESSING

Preprocessing steps were undertaken before any Daedalus data was used for digital or statistical analysis. Preprocessing consisted of evaluating the quality of the data, both visually and digitally.

Visual evaluation of the data was accomplished using the IDIMS system to display each data channel individually on a monitor to inspect for faulty data lines, band-to-band misregistration and large-scale shifts in DN values across the scene. With the aid of topographic maps and image coordinates obtained from the display screen, an approximate per-pixel resolution was calculated. Image scales were computed in both the along-track and across-track directions to determine if distortions related to scanning rate and scanner geometry were similar in both directions.

Digital evaluation of the data using the IDIMS function HISTOG was performed to generate histograms for each data channel. Histograms provided the following information: (1) the dynamic range of the DN values for each channel, including the distribution of the data and any possible point of data saturation (usually 0 or 255), and (2) the occurrence of abnormalities inherent within the data due to telemetry error or reformating conversion to Computer Compatible Tape. Drop-outs or zeroing of some of the least significant bits for one or more channels was documented for the pre-TMS configuration Daedalus flight of September 11, 1981.

GENERATION OF THE REFERENCE DATA SET

A specialized data set was utilized in the statistical analysis of the relationships of geology and vegetation to spectral response. Although the information content of this data set was generalized in comparison to site data collected in the field, it provided an unbiased and random sample of the vegetation and geology characteristics found in the test site. The ELAS software package was used to generate this reference data set. Approximately 300 sample sites were defined randomly within the limits of the TMS image, delimiting the test site (592 lines, 750 samples). A 3 x 3 pixel polygon was chosen to represent each sample site in order to reduce the effect of single-pixel positional errors when locating the sampling sites on related maps and aerial photography. A second ELAS module was used to extract and print DN values for each site (9 pixels) for all seven TMS channels. A subjective criteria was employed to eliminate those sites with high inter-polygon pixel variability. High variability among pixels within a polygon suggests non-homogeneity, generally an indication that the polygon includes portions of two or more landcover types. This editing process reduced the number of sample sites to 254.

The image was transferred to the IDIMS system where the functions TSDEFINE and TSSELECT were utilized to calculate statistics (means, standard deviation, covariance and correlation matrices) for each sample site.

A false-color composite image containing the sample sites was generated with a DICOMED film recorder. Vegetation and geologic information were recorded for each sample site by optically combining (with a Zoom Transfer Scope) vegetation and geology maps with color transparencies obtained from the DICOMED. A data file containing spectral information (means for all 7 channels), vegetation type, and geologic type for each sampling site was created in a format usable by the BMDP statistical software package.

The IDIMS function PIXSERT was employed to generate a single-band image containing only the sample sites and a background DN value of 255. Each sample site was represented as a homogeneous 3 x 3 area with a DN corresponding to the vegetation class identified for that site. This resulting image was utilized in later contingency table analysis during evaluation of several multispectral classifications.

OUTPUT PRODUCTS

Digital analysis output products for the project included statistical results and summaries in tabular form and color and black and white photographic products. The tabular output products were generated during intermediate and final statistical analyses utilizing several programs within the BMDP statistical software package.

Positive and negative 4" x 5" transparencies were generated utilizing the IDIMS software and a DICOMED D47 film recorder. Photographic output products include: (1) a black and white band-by-band image illustrating the full 12-channel datset; (2) a black and white band-by-band 7-channel image of the simulated TM channels; (3) a maximum likelihood classification utilizing the seven TM channels; (4) a maximum likelihood classification utilizing the first four prinpipal components derived from the 7-channel dataset; (5) black and white images of each of the first four principal components; (6) a false-color composite utilizing Daedalus channels 7, 5,and 3 and (7)a color split-image of Eight Dollar Mountain illustrating differences between ultramafic and non-ultramafic areas utilizing a false-color composite and color ratio image (Figure 3). These products can be found in Appendix B.

ORIGINAL PAGE
COLOR PHOTOGRAPH

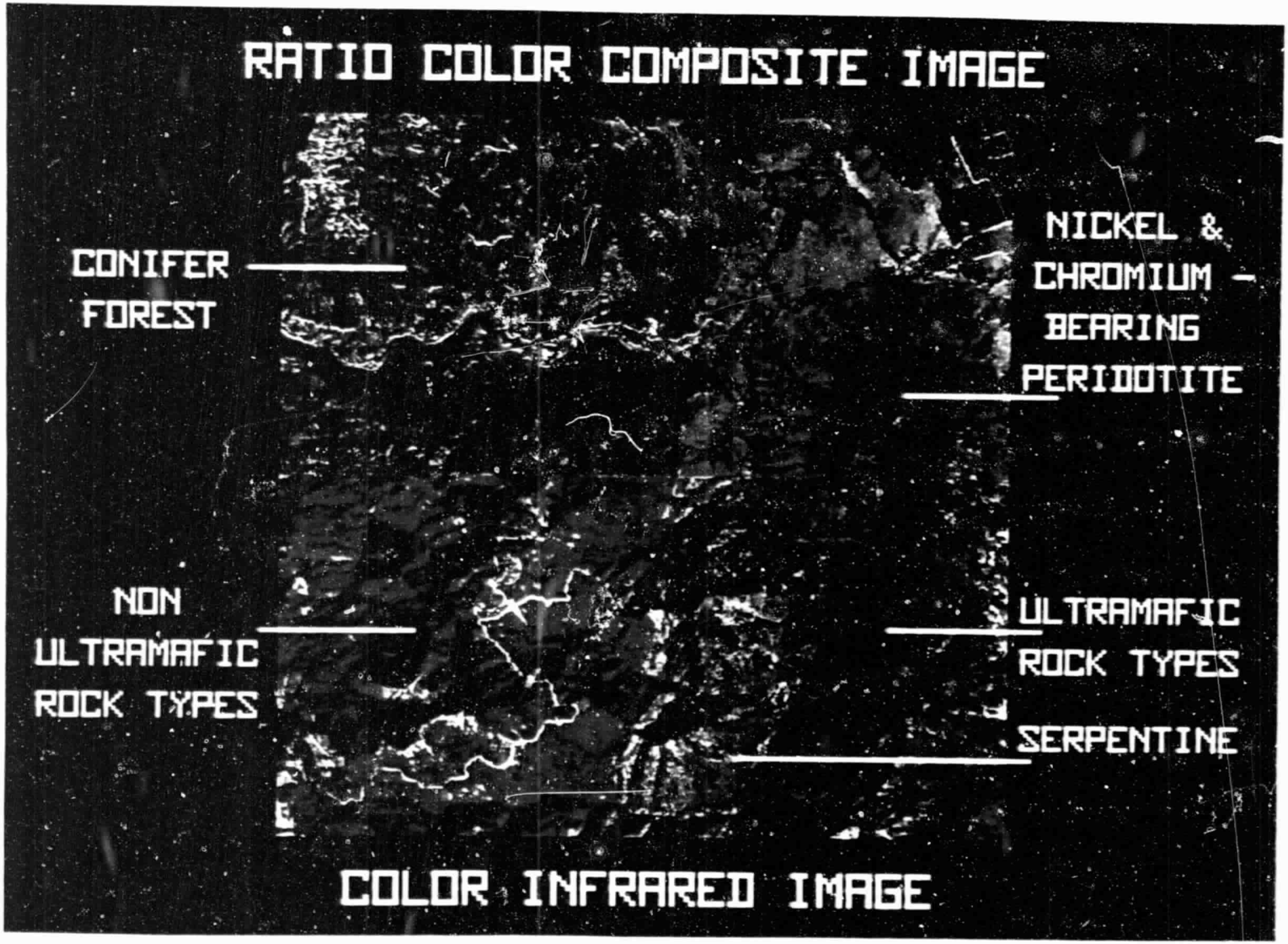


Figure 3. False-color composite and ratio image of the test site.

IV. STATISTICAL ANALYSIS

TMS data was integrated with vegetation and geologic information for subsequent statistical analyses which included a chi-square test, an analysis of variance, stepwise discriminant analysis and Duncan's Multiple Range Test [1]. Results indicate that ultramafic rock types are separable from non-ultramafics based on vegetation type and cover.

METHODOLOGY

Twelve channels of spectral information with a 30 meter spatial resolution were acquired. Of these twelve channels, seven which simulate the Thematic Mapper (TM) were selected for utilization in subsequent statistical analyses. Field data collection during the summers of 1981 and 1982 was supplemented with data from approximately 208 sites randomly distributed throughout the study area to sample the full range of vegetative conditions and rock types. Each of the 208 sites was characterized by rock type[5] and vegetation type[2] . Mean reflectance values were calculated for each of the 208 random sites and integrated with the vegetative and geologic factors for subsequent statistical analyses.

The ability to discriminate ultramafic rock types based on their vegetative association with remotely sensed data requires an understanding of these relationships. The association between vegetation and rock types was analyzed by constructing a co-occurrence matrix (contingency table) of vegetation and geologic classes for the 208 random sample sites, and computing a chi-square statistic for each combination of vegetative class and geologic class. The chi-square statistic was based on the co-occurrence and the mutual non-occurrence for each class pair; a large value indicating statistical significance of the association for each pair. The strength of the association was evaluated by a "correlation-like" measure called the Yule coefficient of association[6]. In conjunction with visual inspection of the co-occurrence matrix, the chi-square tests and the Yule coefficients allowed selection of those class combinations having close associations.

The second series of statistical tests involved a one-way analysis of variance (ANOVA) and a stepwise discriminant analysis of the reflectance data to determine the ability of the spectral bands to discriminate among vegetation classes. Each band was analyzed independently using the ANOVA, and all TM bands were analyzed simultaneously using discriminant analysis. F-ratios, which are essentially the between-class variation divided by the within-class variation, served as measures of separability for the vegetation types. For ANOVA, the F-ratios allowed ranking of the spectral bands as to their relative ability to discriminate among vegetation classes. The band with the highest F-ratio was given the top ranking of one; subsequent bands were ranked according to their decreasing F-ratios. However, this ranking procedure applies to individual bands and is not appropriate for the selection of optimal band combinations. Therefore, discriminant analysis was performed to identify the best combination of spectral bands for discriminating vegetation associated with specific rock types. This analysis complemented the ANOVA which was performed to analyze the contribution of each individual spectral band. The best subset of bands was determined for vegetation associated with ultramafics and for all vegetation classes.

Finally, Duncan's Multiple Range Test was performed on data having significant F-ratios to determine which subgroupings of classes were statistically non-distinguishable. Relating these homogeneous subgroupings to specific vegetation types contributed further to the evaluation of the use of spectral reflectance data in vegetation discrimination. Duncan's Multiple Range Test is a procedure for performing multiple comparisons between class means subsequent to an analysis of variance. A significant F-ratio in the ANOVA indicates at least one significant pair-wise mean difference, but does not specify which pairs differ. The advantage of Duncan's test over the usual method of performing multiple t-tests is that it maintains a fixed level of significance for all statements regarding mean differences.

RESULTS

Based on an analysis of the chi-square test, several vegetation types were found to occur exclusively on specific parent materials (Table 2). The most highly correlated association (.93) involved the Jeffrey pine-dominated coniferous forest (class 122.512) and ultramafic rock units. The Manzanita (class 132.211) and California laurel (132.212) shrublands were also found to be solely associated with ultramafic materials. These three vegetation types can

Table 2. Chi-square statistics for vegetation and lithologic associations.

VEGETATION TYPE	GEOLOGIC UNIT	COEFFICIENT OF ASSOCIATION	LEVEL OF SIGNIFICANCE	CHI SQUARE VALUE ($\chi 2$)
122.512 JEFFREY PINE	ULTRAMAFIC	0.93	< 0.001	46.2
122.512 JEFFREY PINE	METAVOLCANIC	-0.88		19.9
122.514 DOUGLAS FIR	METAVOLCANIC	0.68		12.6
122.911 DOUGLAS FIR MIXED	ULTRAMAFIC	-0.67		16.9
122.911 DOUGLAS FIR MIXED	METASEDIMENTARY	0.79		16.4
132.212 CALIFORNIA LAUREL	ULTRAMAFIC	0.87		16.8
122.514 DOUGLAS FIR	ULTRAMAFIC	-0.74	< 0.01	9.7
132.211 MANZANITA	ULTRAMAFIC	0.67		7.6
122.912 DOUGLAS FIR MIXED	ULTRAMAFIC	-0.50	< 0.025	5.3
122.913 DOUGLAS FIR MIXED	ULTRAMAFIC	-0.81		5.2

be considered as indicative of underlying ultramafic parent materials in the study area. Conversely, Douglas-fir dominated coniferous forest (class 122.514) and mixed coniferous broadleaf (classes 122.911, 122.912 and 122.913) forests were conspicuously absent from ultramafic sites as indicated by a significant negative correlation. The Douglas fir-dominated forest type (class 122.514) occurs mainly on metavolcanic parent materials, while Douglas fir mixed forest (class 122.911) was found predominantly on metasedimentary materials. The remaining vegetation types were associated with several rock types.

Using ANOVA, spectral reflectance bands were ranked (by means of their F-ratios) according to their usefulness in separating the vegetation types (Table 3). Results indicate that, on a per-channel basis, TM band 3, (with an F-ratio of 16.51) best differentiates vegetation followed by the thermal band (TM 6); while TM band 4 (near infrared) was the least useful (F-ratio of 6.59) for separating the vegetation types. This ranking of spectral bands based on the F-ratio applies to all vegetation classes.

Stepwise discriminant analysis was performed to identify the optimal combination of spectral bands for

Table 3. One-way analysis of variance for vegetation discrimination based on spectral data.

TM BAND	F-RATIO*	RANK
1	14.37	4
2	14.63	3
3	16.51	1
4	6.59	7
5	11.15	6
6	16.13	2
7	13.39	5

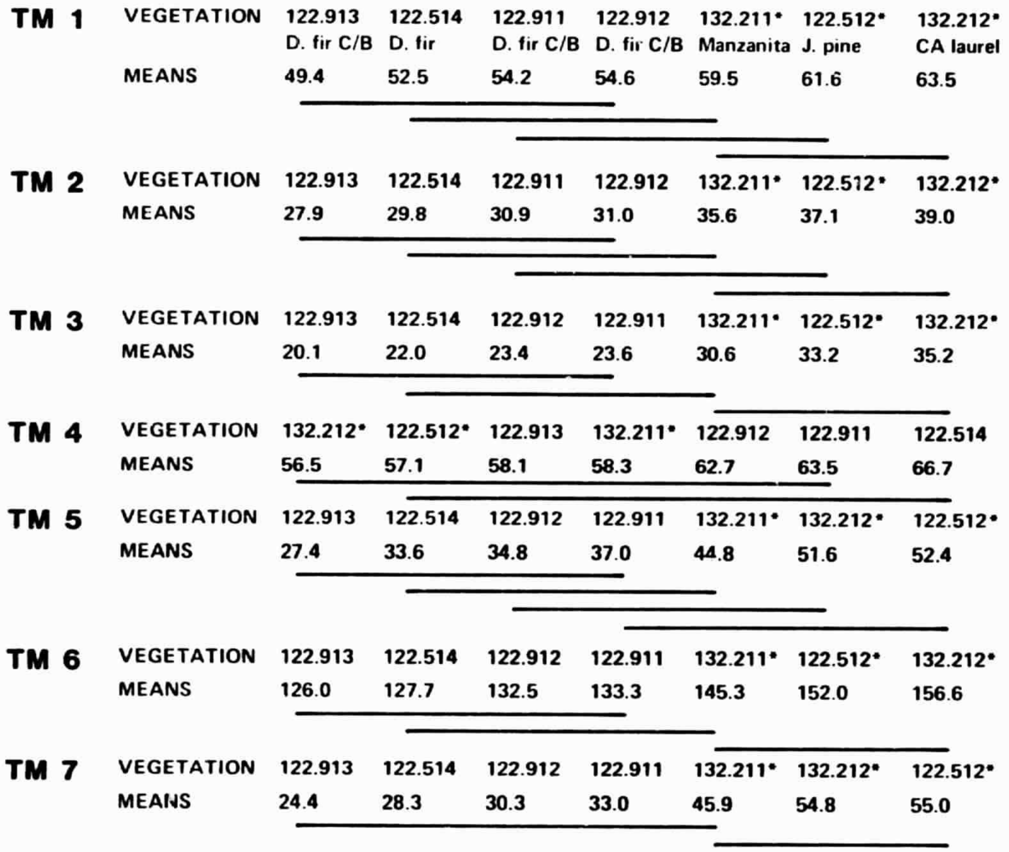
***ALL VALUES ARE SIGNIFICANT AT THE 0.001 LEVEL**

discriminating vegetation classes. The results (Table 4) indicate that TM band 6 (thermal band) best discriminated ultramafic vegetation with other TM bands not contributing significantly to the discrimination. TM bands 3, 4, and 2 (in order of importance) significantly contributed to the discrimination of all vegetation classes. The selection of TM band 3 as the best single band for all vegetation is consistent with the ANOVA. The apparent inconsistency of rankings for ANOVA and discriminant analysis concerning band 4 is a result of partial correlations between the TM bands. ANOVA, a univariate technique, ignores partial correlations between bands and, therefore, the ANOVA ranking is not appropriate for the selection of optimal bands.

Duncan's Multiple Range Test was performed next to determine whether ultramafic vegetation (determined in the chi-square test) was separable from vegetation occurring on other rock types (Figure 4). Based on the Multiple Range Test, TM bands 3 and 6 were found to be the best bands for differentiating those vegetation types associated with ultramafic rock types as indicated in the initial chi-square analysis. The spectral means for the Jeffrey pine-dominated forest

Table 4. Optimal band combinations for discriminating vegetation based on stepwise discriminant analysis.

CLASSES	BANDS	F-RATIOS	SIGNIFICANCE
ULTRMAFIC VEGETATION	TM 6	3.17	$p < .05$
ALL VEGETATION	TM 3	18.21	$p < .001$
	TM 4	5.87	$p < .001$
	TM 2	2.69	$p < .025$



TM 1	VEGETATION	122.913 D. fir C/B	122.514 D. fir	122.911 D. fir C/B	122.912 D. fir C/B	132.211* Manzanita	122.512* J. pine	132.212* CA laurel
	MEANS	49.4	52.5	54.2	54.6	59.5	61.6	63.5
TM 2	VEGETATION	122.913	122.514	122.911	122.912	132.211*	122.512*	132.212*
	MEANS	27.9	29.8	30.9	31.0	35.6	37.1	39.0
TM 3	VEGETATION	122.913	122.514	122.912	122.911	132.211*	122.512*	132.212*
	MEANS	20.1	22.0	23.4	23.6	30.6	33.2	35.2
TM 4	VEGETATION	132.212*	122.512*	122.913	132.211*	122.912	122.911	122.514
	MEANS	56.5	57.1	58.1	58.3	62.7	63.5	66.7
TM 5	VEGETATION	122.913	122.514	122.912	122.911	132.211*	132.212*	122.512*
	MEANS	27.4	33.6	34.8	37.0	44.8	51.6	52.4
TM 6	VEGETATION	122.913	122.514	122.912	122.911	132.211*	122.512*	132.212*
	MEANS	126.0	127.7	132.5	133.3	145.3	152.0	156.6
TM 7	VEGETATION	122.913	122.514	122.912	122.911	132.211*	132.212*	122.512*
	MEANS	24.4	28.3	30.3	33.0	45.9	54.8	55.0

*VEGETATION TYPES ASSOCIATED WITH ULTRAMAFICS

Figure 4. Discrimination between vegetation classes for each TMS band. Note: underscored class means are not significantly different.

(class 122.512) and the California laurel shrubland (class 132.212) were similar to the Manzanita shrubland (class 132.211), yet separable from all other vegetation associated with non-ultramafics. An analysis of spectral means indicates that ultramafic vegetation is separable from vegetation found on other rock types. In addition, TM band 7 successfully separated vegetation associated with ultramafics, although this band is less successful in the separation of all vegetation types. However, for all three bands (TM 3, TM 6, and TM 7), the Manzanita shrubland was spectrally similar to vegetation types associated with both ultramafics and non-ultramafic rock types.

V. EVALUATION OF ANALYTICAL TECHNIQUES

To assess and compare the performance of a number of analytical techniques for deriving detailed vegetation information from TMS data, accuracy assessments were performed for a number of data sets. The four analytical tehcniques evaluated the following data sets: 1) all seven TMS bands, 2) four MSS equivalent bands, 3) near infrared(IR)/Red band ratio, and 4) principal components analysis. The verification of classification performance utilized 254 randomly located sites for each of the classifications.

DATA SET PREPARATION

MULTISPECTRAL CLUSTERING AND CLASSIFICATION

Clustering and classification was performed on two subsets of the Daedalus multispectral scanner data set. One such subset utilized 7 Daedalus channels (2,3,5,7,9,10 and 11) to simulate the equivalent spectral channels of the Landsat 4 Thematic Mapper, while a second subset consisted of 4 Daedalus channels (3,5,6 and 8) to approximate the spectral response of the multispectral scanner (MSS) aboard Landsats 1-4.

The IDIMS function MAGNIFY was used to resample the subset images by a factor of two in both the line and sample directions (resulting in a sampling of 25% of the data) to reduce the processing time required by the clustering algorithm. An unsupervised clustering approach was utilized because of the paucity of ground data available, high spectral confusion related to the increase in sensor resolution, and the lack of a refined classification scheme linking the various elements of the vegetation cover. Clustering of both the 4-channel and 7-channel data sets was completed utilizing the IDIMS function ISOCLS.

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis (PCA) is a data reduction technique based upon a variance- maximizing transformation. A significant amount of redundancy exists in TMS data due to the narrow spectral bandwidth and spectral adjacency of several sequential channels. The PCA transformation can reduce the number of dimensions (channels) by producing orthogonal linear combinations of the original variables, while preserving the important information content.

Principal components analysis was performed on the 7-channel TMS data set using the VICAR applications program EIGEN. Results were computed for both the covariance and correlation matrices, but an apparent software error provided conflicting eigenvalues from the correlation matrix. Results from the principal components analysis (covariance matrix eigenvalue weightings) are summarized in Table 5. The first component accounted for 87% of the total variance with the shortwave IR bands (5 and 7) and the thermal band (6) making the greatest contributions. Components 1 through 4 were united into a multi-band image and classified utilizing an unsupervised approach similar to that described for the 4-channel and 7-channel data sets.

RATIO DATA SET

Ratioing is a technique in which two co-registered images (channels) are divided pixel-by-pixel. The resultant image, which may be rescaled to the maximum dynamic range of the display device (usually 0 - 255), tends to enhance subtle differences among classes (e.g. lithology). Ratio techniques are generally utilized for the differentiation and enhancement of geologic

Table 5. Eigenvalue weightings for each principal component.

PCA #	% OF TOTAL VARIANCE	WEIGHTING OF BANDS (HIGH → LOW)
1	86.87	6, 7, 5, 3, 1, 2, 4
2	8.95	4, 5, 7, 3, 6, 2, 1
3	2.17	7, 6, 1, 5, 4, 2, 3
4	1.15	1, 3, 2, 7, 5, 4, 6
5	0.68	5, 6, 4, 7, 1, 2, 3
6	0.16	1, 3, 2, 6, 5, 7, 4
7	0.02	2, 3, 1, 4, 5, 7, 6

materials and have been applied to a lesser extent to the enhancement of natural vegetation.

The VICAR applications function F2 and the IDIMS function DIVIDE were utilized to perform ratioing on several channels of the TMS data. Resultant values were linearly rescaled to a range of 0 to 255 to retain continuity in dynamic data range with the other data sets utilized in the project. The following ratios were computed: 7/5, 4/3, 3/2, 5/7, 3/6, 3/4, and 2/3. After an evaluation of the ratio data sets had been completed, the 4/3 ratio (Near IR/Visible) was selected for further study because of its sensitivity to differences in native vegetation.

METHODOLOGY

To estimate the relative performance of the derived classifications, a number of cluster samples randomly located throughout the test site were compared with the vegetative map units. Although accuracy is referred to throughout this assessment, it is only valid to the extent that the verification data base (the vegetation map produced by OSU) is an accurate representation of land cover. Two types of comparison were made between the TMS derived classifications and the vegetation map: 1) the creation of contingency tables, and 2) calculation of percentage correct for each vegetation class.

A random sampling cluster approach was utilized in the performance assessment. Cluster sampling was chosen in order to minimize the effect of single-pixel location errors. A cluster size of three pixels by three pixels was chosen. Following identification of the random samples , those containing more than one vegetation class were deleted from the file.

The reference data set was utilized initially to identify the vegetation designation of each spectral class within the classification. A contingency table of the spectral classes and the vegetation designations for the random sites was generated to assign a vegetation label to each class based upon the highest number of occurrences. A second contingency table, with spectral classes grouped by vegetation designation, was generated to assess the relative (not absolute) performance of the various analytical techniques.

Evaluation of classification accuracy necessitates sampling each of the cover classes. Based on the random sampling strategy utilized in this procedure, only 13 of the 16 OSU vegetation classes were sampled. Of the 13 classes sampled, 5 classes received less than 2% of the total sample sites, reflecting their rare occurrence in the test site (Table 6).

RESULTS

The comparative performance of the four analytical techniques is illustrated in Figures 5 and 6. These figures employ bar charts to depict classification accuracies, both by vegetation class and overall, for each of the analytical techniques. The proportion correct represents the correspondence between the vegetation map and the TMS classifications. For a given category, the proportion correct is simply the sum of the classified pixels which matched the sample designation divided by the total number of samples from the reference data set. The probability correct is based on the estimation of the number of classified pixels which match the designation given by the OSU researchers. The second estimate involves the overall proportion correct for all vegetation classes together.

Individual class accuracies for the grouped classes are shown in Figure 5. Classification

Table 6

Areal Average of Vegetation Classes
Based on Random Sampling

Class	Series	Dominant	Areal Extent
122.511	Mixed Conifer	True Fir	15.0%
122.512	Mixed Conifer	Jeffrey Pine	2.0
122.513	Mixed Conifer	Port Orford Cedar	14.0
122.514	Mixed Conifer	Douglas Fir	.8
122.711	Broadleaf Evergreen	Chinkapin	9.0
122.811	Broadleaf Deciduous	Oregon White Oak	.8
122.812	Broadleaf Deciduous	Bigleaf Maple	1.6
122.911	Mixed Conifer-Broadleaf	Douglas Fir	.8
122.912	Mixed Conifer-Broadleaf	Douglas Fir	23.2
122.913	Mixed Conifer-Broadleaf	Douglas Fir	12.6
132.211	Shrubland	Green Manzanita	5.9
132.212	Shrubland	California Laurel	6.7
142.311	Graminoid	Grasses	7.5

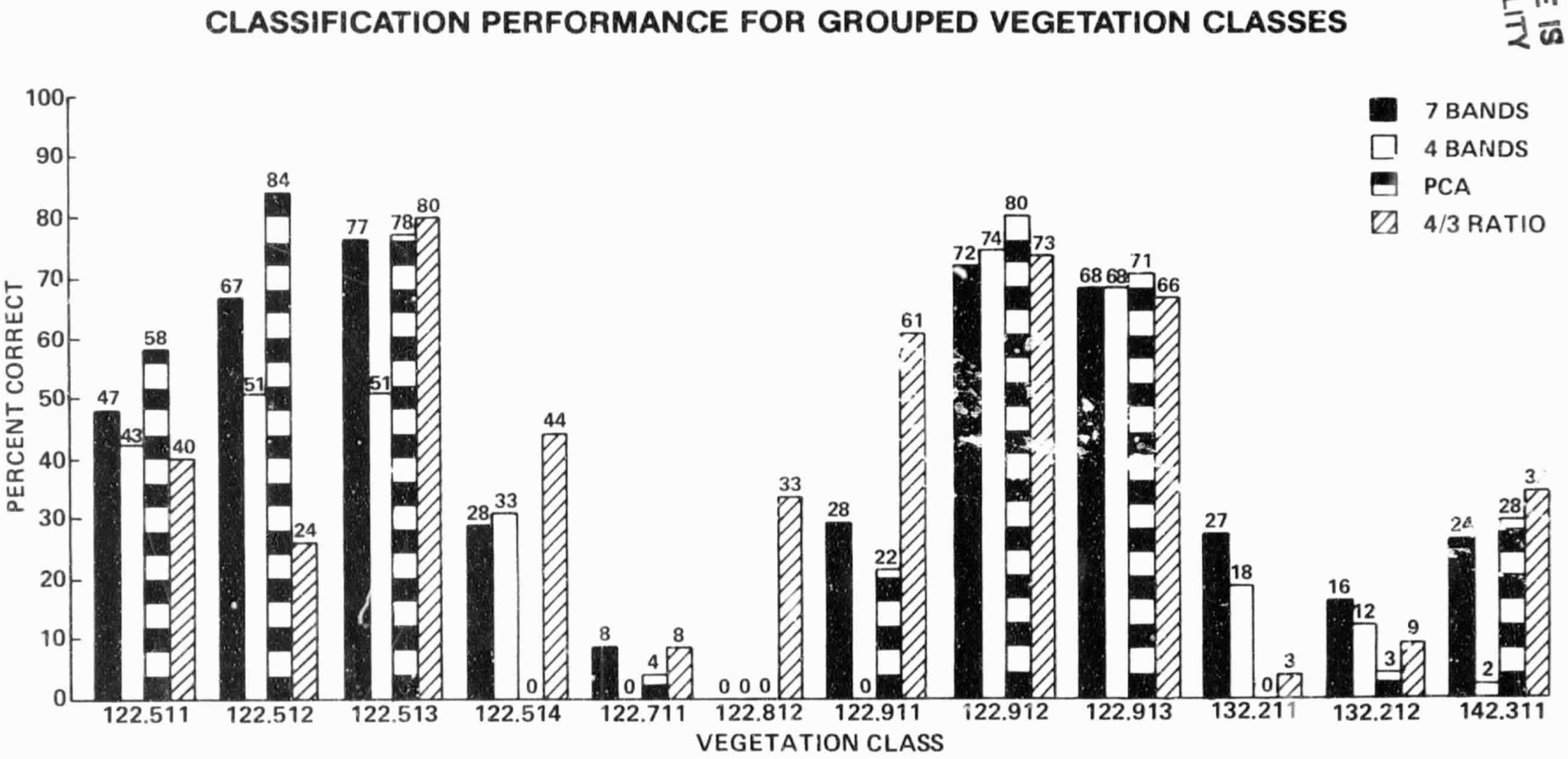


Figure 5. Classification performance for each technique for all vegetation classes. Note: 7 bands correspond to TMS band configuration; 4 bands correspond to MSS band configuration.

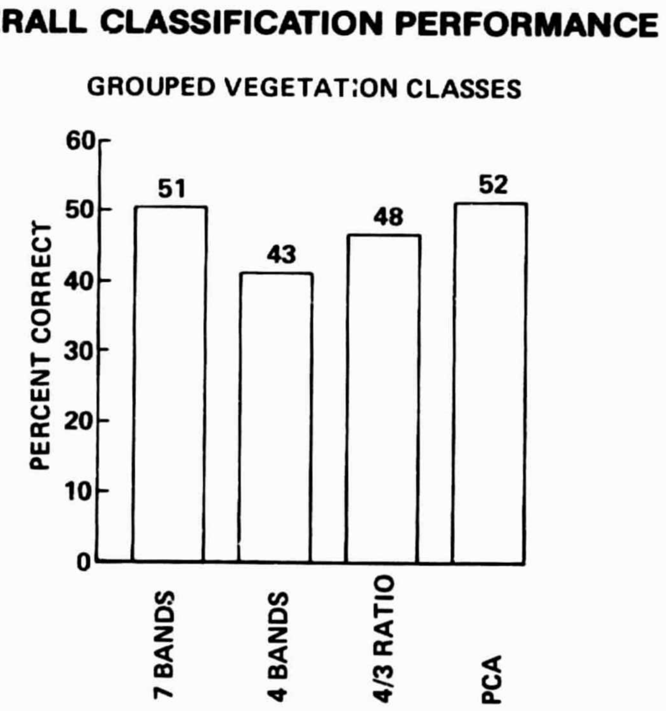


Figure 6. Overall classification performance for each technique. Note: 7 bands correspond to TMS band configuration; 4 bands correspond to the MSS band configuration.

performance for ultramafic associated vegetation was 84% for mixed conifer forests (122.512) based on the PCA and lower for the montane shrub classes (132.211 and 132.212) with accuracies of 27% and 16%, respectively. High mapping performance was noted for all techniques for those vegetation classes with large areal coverage (122.512, 122.912 and 122.913). The mapping performance for the ratioed data set was highest for those vegetation classes with a large proportion of herbaceous and deciduous vegetation (142.311, 122.911, 122.812 and 122.513). The principal components analysis yielded higher classification accuracies for the forest classes than the other techniques, followed closely by the 7-banded classification. This same trend is reflected in overall performance, where the percent correct for the 7-banded and the principal components analysis were 51% and 52%, respectively (Figure 6). Therefore, the use of a data compression technique (such as principal components analysis) results in accuracies as high as that obtained using all seven TMS bands while reducing computation costs and complexity. The PCA classification is shown in Figure 7. Although the overall accuracies were low, reflecting to a large degree the use of an unsupervised classification approach, the intent is to observe the relative (not absolute) accuracy of the various techniques for deriving detailed vegetation information. The 4-banded classification (MSS simulation) had consistently lower mapping accuracies than the other techniques.

Utilization of the vegetation map as a base for determining the relative accuracy of the TMS-derived classifications assumes that the vegetation map is 100% correct. As a result, any error that may exist within the vegetation map may adversely affect the overall probability correct for each of the classifications. Therefore, to evaluate the vegetation map, approximately 10% of the random sites were field checked in the summer of 1983. A verification of the vegetation designations derived from the vegetation map revealed that 16 out of 20, or 80% were correctly identified. This 20% error factor may well account for the low overall performance of the classifications, regardless of the technique utilized in the analysis. Other factors which may have resulted in lower classification accuracies include: 1) the attempted classification of 13 detailed vegetation classes, 2) overlapping class definitions and 3) spectral confusion

Figure 7. Classification for the principal components data set.

between various land cover classes. However, the relative performance of the four analytical techniques provides insight into the utility of these techniques for extracting detailed vegetation information.

VI. CONCLUSIONS

Results indicate that, in many cases, TMS data was able to successfully discriminate vegetation associated with specific rock types. In particular, there was a close correspondence between several vegetation classes and ultramafic rock types, three vegetation classes occurring solely on ultramafics while another three types were totally absent. Two other classes occur mainly on metasedimentary and metavolcanic parent materials. Vegetation associated with ultramafics was clearly separable from other vegetation types with the TMS data. Based on Duncan's Multiple Range Test, three bands (TM 3, TM 6 and TM 7) were particularly well suited for separating vegetation classes associated with ultramafic rock types. This is consistent with the results of the discriminant analysis which identified TM band 6 as the best band for discriminating ultramafic vegetation, while bands 2, 3, and 4 were selected as optimal for differentiating all vegetation within the test site.

A comparison of analytical techniques based on the classification of the four data sets (which included four MSS equivalent bands, all 7 TMS bands, Near IR/Visible ratio, and a principal components analysis) provided insight into the ability of these techniques to differentiate detailed vegetation classes. Classification accuracies for the principal components data set was highest, followed closely by the seven band classification (full TM configuration). The classification of the first four principal components provided the highest mapping performance while reducing computation requirements and complexity. The four band classification (MSS simulation) had consistently lower mapping performance. Low overall accuracies were the result of the unsupervised clustering and classification techniques utilized in the comparison which did not optimize analyst interaction and information extraction.

REFERENCES

[1] D.B. Duncan, Multiple Range and Multiple F Tests, Biometrics, Vol. 11, pp. 1-42, 1955.

[2] R. Frenkel and C. Kiilsgaard, Vegetation of the Southwest Oregon Study Area, Corvallis, Department of Geography, Oregon State University, unpublished report and maps, 1984.

[3] D.A. Mouat, The Response of Vegetation to Geochemical Conditions. Proceedings of the International Symposium on Remote Sensing of Environment, Second Thematic Conference, Remote Sensing for Exploration Geology, Ft. Worth, pp75-83, 1982.

[4] D.A. Mouat, L.A. Morrissey, and E.M. Horn, Geobotanical Discrimination of Ultramafic Parent Materials: An Evaluation of Remote Sensing Techniques, Proceedings of the Sixteenth International Symposium of Remote Sensing of Environment, Ann Arbor, Environmental Research Institute of Michigan, 10 pp., 1982.

[5] L. Ramp, and N.V. Peterson, Geology and Mineral Resources of Josephine County, Oregon, State of Oregon, Department of Geology and Mineral Industries Bulletin 100, 45 pp.

[6] G.U. Yule and M.G. Kendall, Introduction to the Theory of Statistics, Hafner Publishing Company, New York, 701p., 1950.

[7] Morrissey, L.A., K.J. Weinstock and D.A. Mouat, Statistical Analysis of Thematic Mapper Simulator Data for the Geobotanical Discrimination of Rock Types in Southwest Oregon, 1983 International IEEE Geoscience and Remote Sensing Symposium, San Francisco, California, volume 1, 1983.

[8] Morrissey, L.A., K.J. Weinstock, D.A. Mouat and D.H. Card, Statistical Analysis of Thematic Mapper Simulator Data for the Geobotanical Discrimination of Rock Types in Southwest Oregon, IEEE Transactions on Geoscience and Remote Sensing, (under submission), 1984.

APPENDIX 1

APPENDIX 1

Data Form Used in 1982 Field Reconnaissance

VEGETATION MAP OF SISKIYOU MOUNTAINS, PLOT DATA

Plot No. ____________ Landform Code ____________

Date ____________ Bedrock (%) ____________

Photo ____________ Loose Rock (>7cm %) ____________

Elevation (ft) ____________ Bare soil (%) ____________

Aspect (°Mag) ____________ Moss & Lichens (%) ____________

Slope (%) ____________ Litter (%) ____________

- -

Classification ____________

Total Coverage % Conif. ________ Brd Evgrn ________

Brd Decid. ________ Shrub ________ Herb/Grass ________

Canopy Over. ____________ Understory ____________

TREE SHRUB HERB

M

R

NASA

Overstory ______ Understory ______ Soil ______

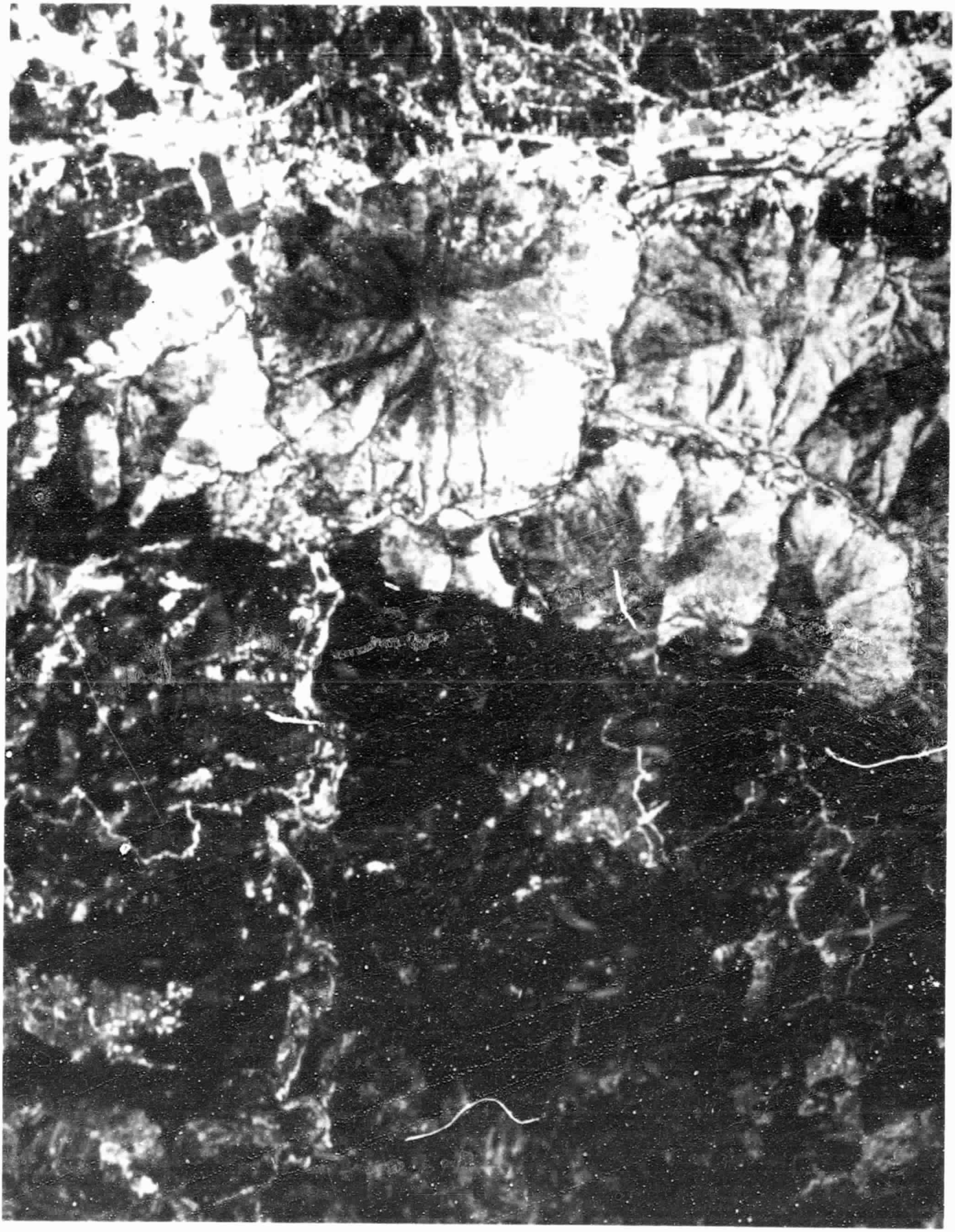
Herb ______ Rock ______ Litter ______

APPENDIX 2

ORIGINAL PAGE
COLOR PHOTOGRAPH

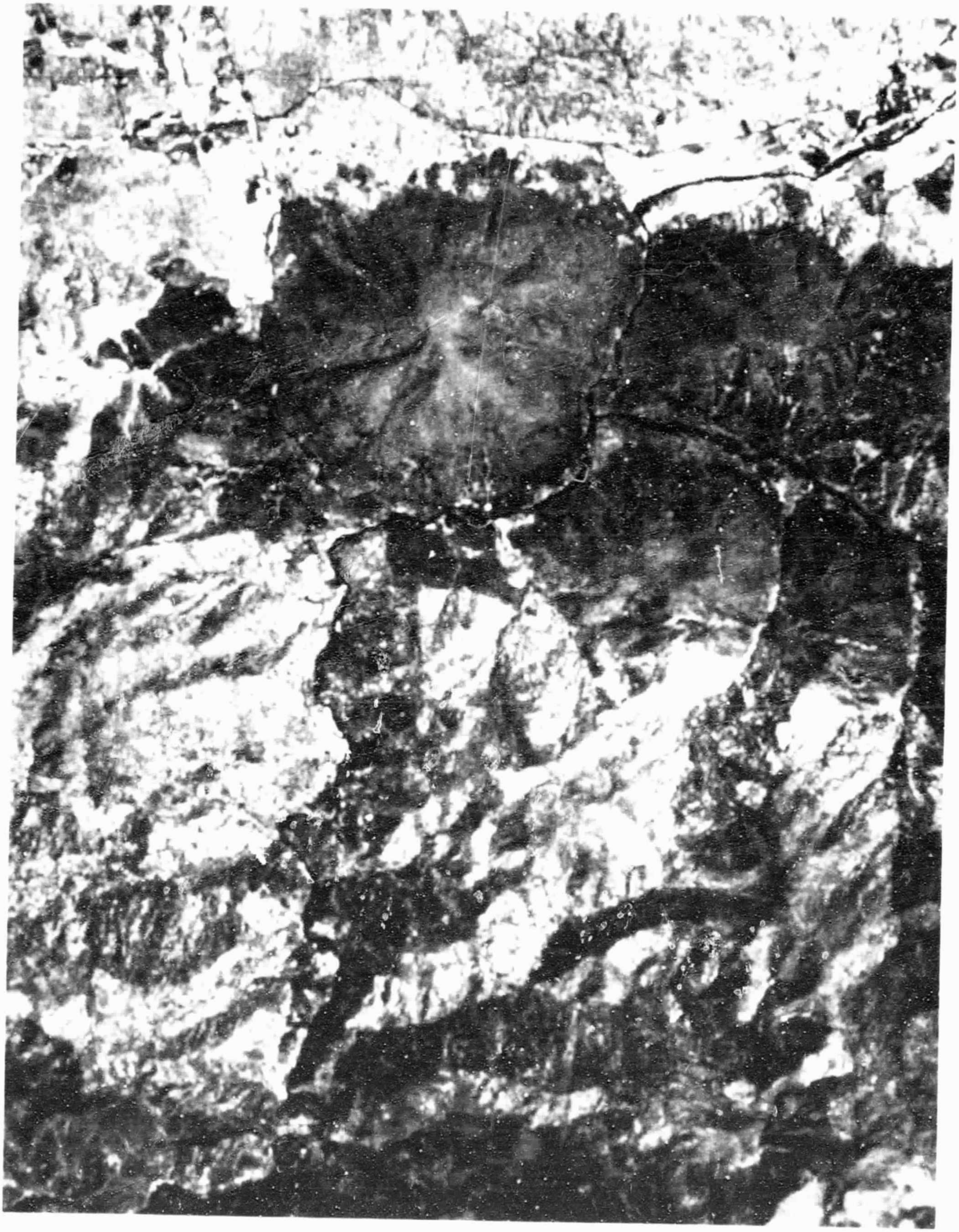
~~ORIGINAL PAGE~~
~~COLO~~ ~~OGRAPH~~

1. Seven channel classification.



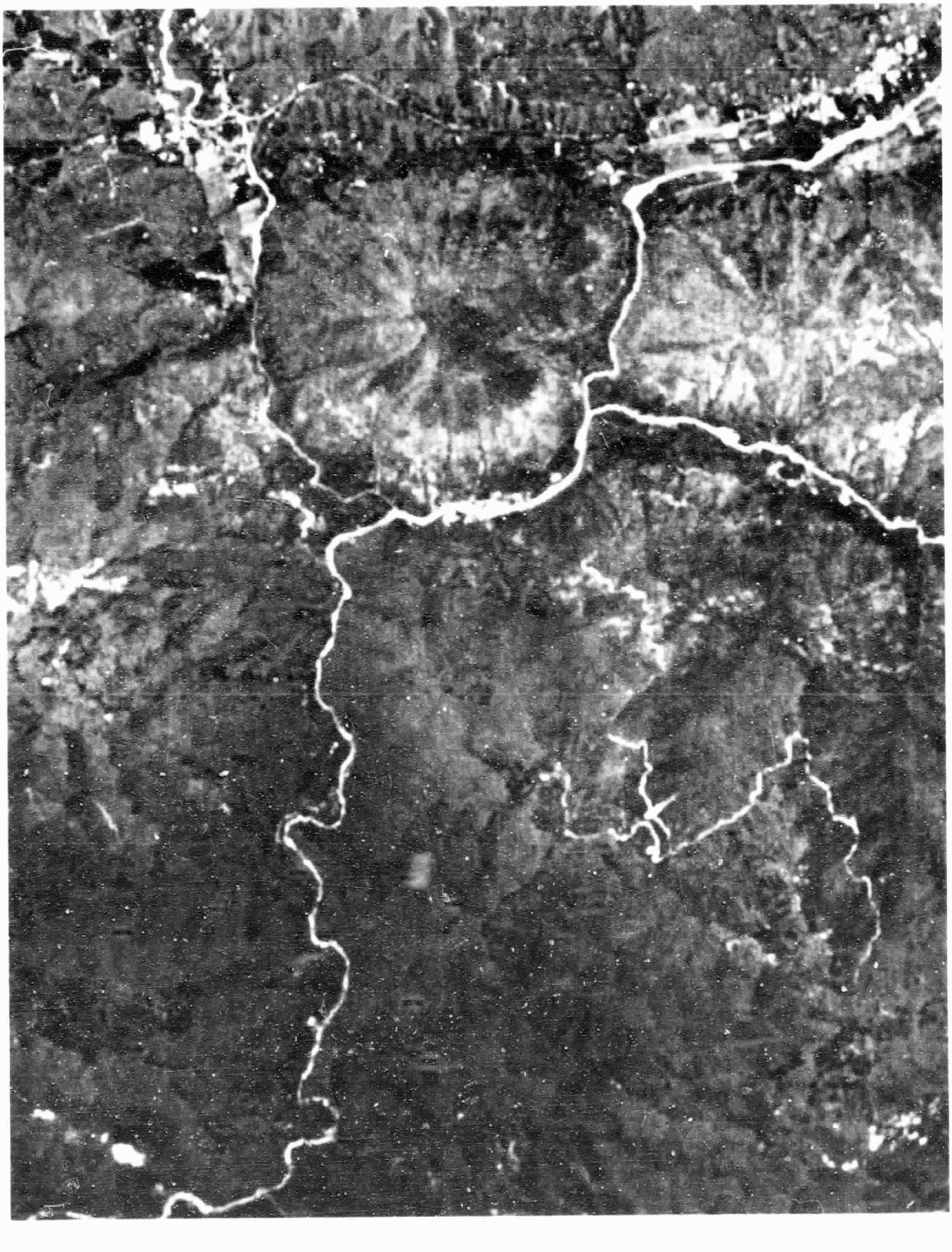
2. First Principal Component.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

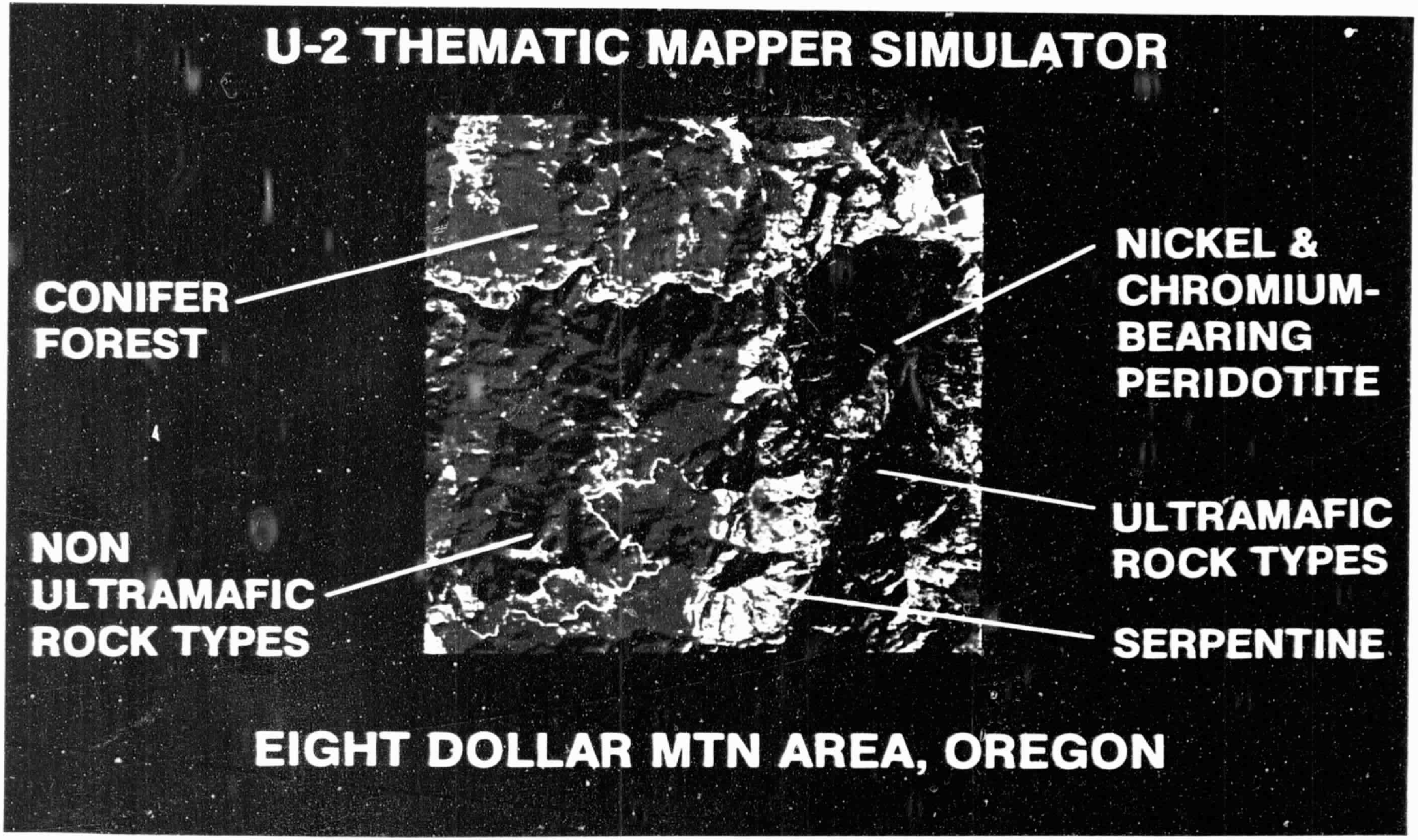


3. Second Principal Component.

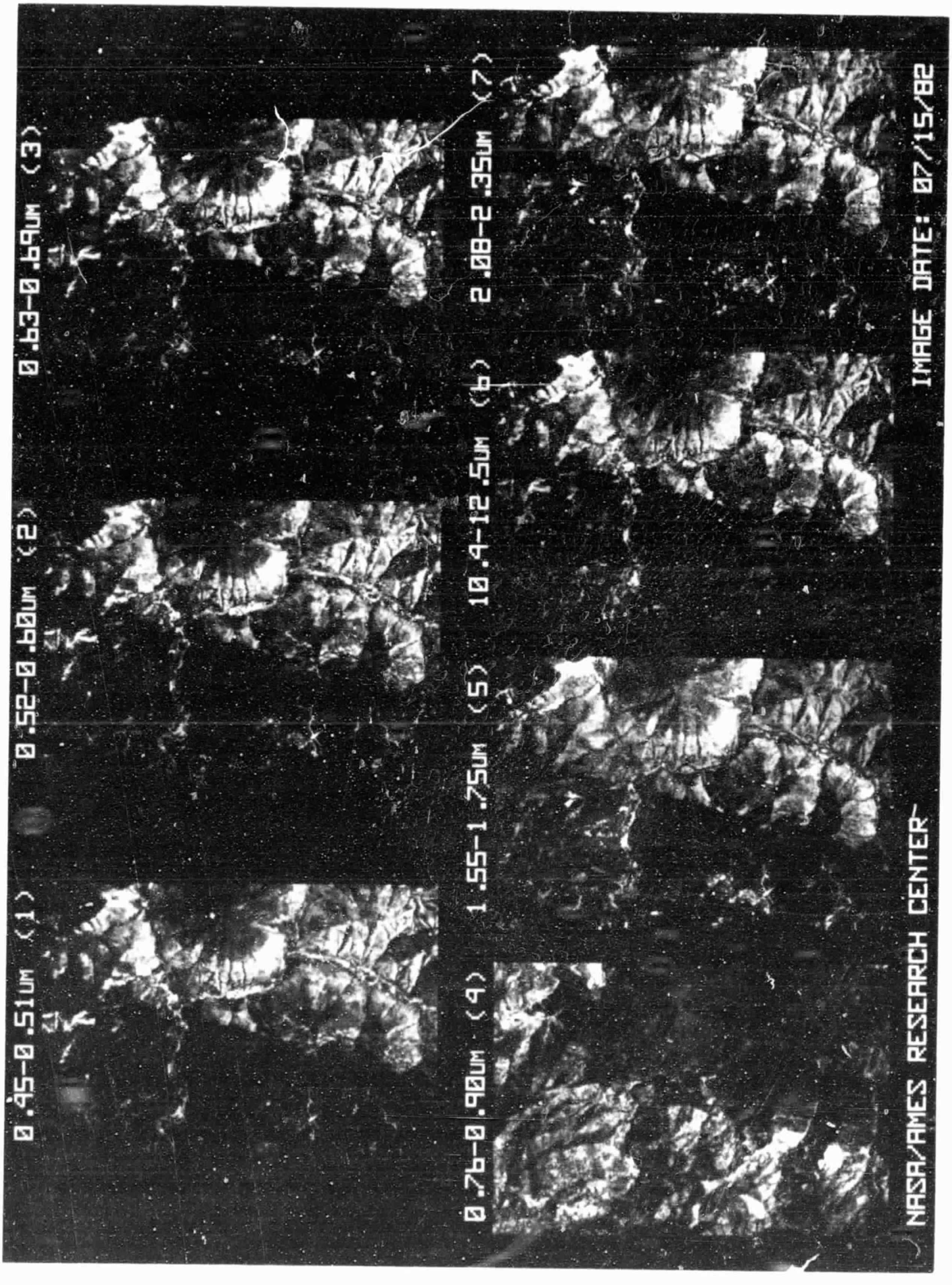
4. Third Principal Component.



5. Fourth Principal Component.



6. False-color Composite.



7. Band-by-band TMS channels.