2P (mix)



E7.4-10625
CR-138735



# ERTS TYPE II REPORT (MAY 3, 1974)

A. TITLE: Multispectral Signatures in Relation to Ground Control Signature Using Nested Sampling Approach.

B. PRINCIPAL INVESTIGATORS: R.J.P. LYON: F.R. HONEY
School of Earth Sciences
Stanford University
Stanford, California 94305

Phone: (415) 497-3262 or 497-4147

C. PROPOSAL #637; GSFC #UN 142: Contract # NAS 5-21884

D. TECHNICAL MONITOR: E.W. Crump
Code 430
Goddard Space Flight Center
Greenbelt, Maryland 20771

Phone: (301) 982-2857

E. PERIOD: March 3, 1974- May 3, 1974

---

Remote Sensing Laboratory
Stanford University
Stanford, California 94305





33

## F. ABSTRACT

Study of the serpentine areas of the San Francisco Peninsula has been extended, analysed and partially evaluated. Results from this study have been encouraging. Correlation between reflectances calculated from the satellite measurements and reflectances measured in the field have been high. The spectra of the serpentine species has been found sufficiently unique to enable discrimination of the areas from ERTS. A parallel study of an area of semi consolidated sandstones near Livermore was also carried out, with similar results to the serpentine study, but again with sufficiently unique signatures.

In order to enable the evaluation of studies of areas involving vegetation coverage to be made more rigorous, a botanist has been included in the group. Results of a study of the biomass, species composition and vigour of the Stanford grasslands area are presented. Correlations of their results with concurrent measurements of the reflectivity of the grass canopy are being performed.

The tape-reading and classification program, RIPPER, has been thoroughly tested and evaluated. Modifications to the program to increase efficiency of core storage and accelerate the clustering algorithm are being carried out.

G. PROBLEMS

None

H. ACCOMPLISHMENTS IN THE PAST REPORTING PERIOD

1. Field Data Collection

On the following dates field data was collected over the three mile transect crossing the four major rock and soil types of the Stanford Testsite.

| Date for Fieldwork | Grass Condition | ERTS Overpass (o/p) |
|---|---|---|
| March 18 | V. Green | 100% cld-March 12 - #1597 |
| March 21 | V. Green | March 30 - #1615 |
| April 21 | Green | April 17 - #1633 |
| May 5 | Green with seed heads | #1651 foggy |
| May 10 | Patchily green/dead | (none) |
| May 23 | Patchily dead/green | #1669 o/p |

2. Biomass Study

At 44 sites, equally spaced at ERTS cell distances ($\pm$ 0.05 mile), biomass studies have been concluded, in the period May 15-May 22 immediately prior to the final day-off of the grasses.

Reflectance (bi-directional FOV 15°) measurements, relative to a $BaSO_4$ plate, were made before the grass was completely removed from a 0.5 X 0.5 meter square, a color slide (K-II) film was made of each site in its original state. The reflectance was measured after cutting and the site re-photographed on color film.

The species proportions were estimated by our botanist, Roberta Sears and the complete grass sample bagged for determination of green-weight and weight loss after drying one night to 110°C. From those values a biomass/meter$^2$ can be found as well as an "index-of-vigor" from the weight loss ratio (see details elsewhere in this report).

3. Soils and Rocktype Maps

The geology and soils map were completed by J.Baker for the 8000 acres grassland of the Stanford Testsite. In particular a detailed topography, slope angle, geology and soils section was prepared at a 1:3600 scale, for the 3 mile transect so that these data can form the basis of discussion of the ground-reflectance data, taken over the past year along the same track.

4. Serpentine/Grass Cover Study

Saul Levine has completed the comparison between the CCT data from ERTS and the ground reflectance for 23 areas of serpentine and soil and for 2 areas on which a grass-fire burned off almost all of the grass cover. These are reported elsewhere in this report.

5. Computer Software Development

The CCT-read programs have now been fully implemented on the PDP-10 in a interactive mode. An atmospheric correction program is in use, with a self-clustering algorithm (see RSL Technical Report 74-4 attached).

An analysis of various classification programs has been completed by Alfredo Prelat and a new program to extract species proportions from mixed-pixel CCT data developed (see RSL Technical Report 74-4 attached).

6. Hardware Development

Development is commencing on the scanner system initially for low altitude operations. It is expected that this will take 1-2 months further.

# STUDY OF SERPENTINE OUTCROP AREAS ALONG INTERSTATE HIGHWAY I-280, STANFORD SITE, CALIFORNIA

BY SAUL LEVINE

## ABSTRACT

Ground reflectance measurements at the four ERTS bands, of serpentine exposures and soils on the San Francisco Peninsula show good correlation with distinctive four band spectra derived from ERTS-CCT data of 6 October, 1972. Excellent correlation is also shown, at the same date, between the ground and ERTS spectra at a test site in semi-consolidated sandstones located on the east side of the coastal range. It is believed both spectra are sufficiently unique from each other and their background to be discriminated by a computerized clustering program.

## I. PROCEDURES AND RESULTS

It had been reported previously, after a detailed seasonal study of the ERTS-CCT data (see March 3, 1974 report) that a strong liklihood existed that the reflectance spectra obtained during the 6 October overpass, due to the grass dieback, essentially represented that of the soil. To verify these tentative conclusions an extensive program of ground reflectance measurements of bare soil was instituted at study areas I and III (serpentine soils). Figure 1, and the Midway study area (marine sediments) Figure 2. The normalized ground spectra obtained are tabularized in Table I and compared with CCT deriven spectra in Figure 3. Study of this data reveals the following:

1. The variability of the data is lower at Farm Hill Road (Area III) and Midway than at Crystal Springs (Area I). It is believed this is due to the difference in the size of the test areas. The Farm Hill Road area is roughly 60 acres, Midway 960 acres, and Crystal Springs 15 acres. The Crystal Springs site is not only smaller but quite exposed and topographically lower, therefore subject to infiltration of material from adjacent soils derived from the nearby Franciscan formation exposures.

2. The ERTS-CCT normalized reflectance spectra for serpentine soils at 6 October are almost identical at Crystal Springs and Farm Hill road.

3. The ground reflectance spectra for serpentine soil at Farm Hill Road compares very favorably with the CCT data. The ground spectra for the roadcut-and outcrop-serpentines, while comparable to each other, are substantially different than that of the soil. It is believed that

because of the small area extent of the outcrops as compared to the soil and the limiting resolution of the ERTS sytem, the outcrops have little integrated effect and only the serpentine soil is detectable on the CCT data.

4. The correlation of the ground spectra obtained at Crystal Springs with the ERTS-CCT data is not quite as good for reasons treated in 1 above. The spectra seems more comparable to the sandstones soil at Midway. Infiltration of the Franciscan sandstones may explain the difference.

5. The variability of the serpentine outcrops at Farm Hill Road is higher than either the relatively fresh roadcuts or the serpentine soils. This is undoubtedly due to the variation in the weathering of the outcrop versus the freshness of the roadcut and the homogeniety of the serpentine soils.

6. The ground reflectance spectra obtained at Midway correlate very well with that derived from the 6 October ERTS-CCT data.

7. A significant difference is seen in both the ERTS-CCT spectra and the ground spectra for the serpentine soils of Crystal Springs and Farm Hill Road and the sandstone soils at Midway.

The Midway site, interestingly enough, turns out to be an explosives test area for the Lawrence/Livermore Radiation Laboratories at which yearly controlled burns are utilized for presenting fires as a result of the tests. It was noted, while at the site, that the heavy burn area at the north side seen in the ERTS imagery was adjacent to an extensive chemically-defoliated strip running across the entire north side of the site. Since it was believed possible that the block sampling utilized to derive the previous CCT spectra could have overlapped this area, it was decided to introduce a grid sampling system across the entire area and replot that data. The results are shown in Figure 4. It is noted that while the data obtained is essentially the same, the curves are more consistent leading to smoother seasonal trends.

## II. INTENDED ACTIVITY NEXT PERIOD

Since it is believed that this investigation has indicated that the serpentine soils at Crystal Springs and Farm Hill Road and the sediment derived soils at Midway are distinguishable and unique on the 6 October ERTS-CCT data, no further effort is contemplated other than applying the computerized clustering program described in SRSL Technical Progress Report No. 74-1. It is hoped that this program will demonstrate the practicality of utilizing ERTS-CCT data for mapping serpentine and sandstone-derived soils.

## TABLE I-GROUND MEASURED GROUP REFLECTANCE DATA, GROUP STATISTICS, AND NORMALIZED REFLECTANCES

### CRYSTAL SPRINGS (Group 1)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 7.22 | 8.83 | 11.00 | 12.15 |
| 5.33 | 7.32 | 8.81 | 10.70 |
| 12.67 | 14.16 | 16.98 | 18.15 |
| 6.99 | 8.69 | 10.92 | 13.84 |
| 8.05 | 9.72 | 12.67 | 14.55 |
| 7.39 | 9.22 | 10.23 | 11.30 |
| 7.15 | 8.40 | 10.00 | 11.20 |
| 9.22 | 11.62 | 12.46 | 16.20 |
| 10.15 | 11.08 | 15.10 | 17.10 |
| 10.82 | 13.68 | 15.43 | 18.20 |
| 7.86 | 10.10 | 13.13 | 14.85 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 8.44 | 10.26 | 12.43 | 14.36 |
| Std. Dev. | 2.08 | 2.17 | 2.56 | 2.80 |
| Coef. of Var. | 0.25 | 0.21 | 0.21 | 0.19 |
| Norm. Refl. | 1.00 | 1.22 | 1.47 | 1.70 |

### CRYSTAL SPRINGS (Group 2)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 10.30 | 13.33 | 15.45 | 16.40 |
| 12.12 | 16.30 | 16.67 | 17.36 |
| 11.25 | 12.78 | 14.26 | 15.67 |
| 11.35 | 12.91 | 15.12 | 16.33 |
| 11.31 | 11.46 | 18.69 | 23.44 |
| 9.40 | 13.27 | 13.46 | 14.31 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 10.96 | 13.36 | 15.61 | 17.25 |
| Std. Dev. | 0.96 | 1.60 | 1.86 | 3.20 |
| Coef. of Var. | 0.09 | 0.12 | 0.12 | 0.19 |
| Norm.Ref. | 1.00 | 1.22 | 1.42 | 1.57 |

### FARM HILL ROAD (OUTCROP-GROUP 1)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 16.10 | 16.71 | 17.72 | 18.02 |
| 14.29 | 14.87 | 14.92 | 16.67 |
| 12.47 | 11.69 | 13.65 | 13.20 |
| 15.50 | 15.64 | 15.69 | 17.43 |
| 15.86 | 15.86 | 19.50 | 20.30 |
| 13.85 | 13.15 | 15.49 | 15.88 |
| 14.39 | 16.02 | 18.43 | 19.63 |
| 14.82 | 18.53 | 19.44 | 21.86 |
| 15.66 | 19.46 | 21.63 | 22.71 |
| 16.92 | 18.28 | 17.65 | 18.16 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 14.99 | 16.02 | 17.41 | 18.37 |
| Std. Dev. | 1.29 | 2.41 | 2.46 | 2.85 |
| Coef.of Var. | 0.09 | 0.15 | 0.14 | 0.16 |
| Norm. Ref. | 1.00 | 1.07 | 1.16 | 1.23 |

### FARM HILL ROAD (OUTCROP-GROUP 2)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 15.30 | 15.01 | 14.70 | 14.90 |
| 16.74 | 17.41 | 17.97 | 18.82 |
| 13.81 | 15.92 | 16.03 | 17.74 |
| 14.62 | 15.49 | 17.59 | 18.24 |
| 16.53 | 19.28 | 20.00 | 21.56 |
| 12.86 | 13.01 | 12.88 | 12.81 |
| 14.59 | 15.52 | 17.83 | 16.73 |
| 14.05 | 14.41 | 15.28 | 15.46 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 14.81 | 15.75 | 16.53 | 19.69 |
| Std. Dev. | 1.33 | 1.89 | 2.24 | 2.53 |
| Coef. of Var. | 0.09 | 0.12 | 0.14 | 0.14 |
| Norm. Ref. | 1.00 | 1.06 | 1.12 | 1.19 |

### MIDWAY (SOIL-GROUP 1)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 15.21 | 19.55 | 21.67 | 24.50 |
| 14.85 | 17.83 | 24.08 | 26.40 |
| 14.75 | 18.10 | 22.50 | 25.30 |
| 15.85 | 19.72 | 24.70 | 26.47 |
| 15.73 | 20.32 | 25.63 | 27.90 |
| 15.79 | 19.43 | 21.33 | 24.46 |
| 13.50 | 17.90 | 20.40 | 22.27 |
| 15.05 | 18.20 | 23.30 | 25.35 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 15.09 | 18.83 | 22.95 | 25.33 |
| Std. Dev. | 0.77 | 0.98 | 1.79 | 1.68 |
| Coef. of Var. | 0.55 | 0.05 | 0.08 | 0.07 |
| Norm. Ref. | 1.00 | 1.25 | 1.52 | 1.68 |

### MIDWAY (SOIL-GROUP 2)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 10.48 | 12.68 | 14.85 | 17.16 |
| 9.62 | 12.30 | 14.37 | 16.23 |
| 9.45 | 11.46 | 13.40 | 16.10 |
| 9.53 | 10.92 | 12.32 | 15.84 |
| 10.13 | 11.98 | 14.47 | 16.75 |
| 10.39 | 11.92 | 14.47 | 16.97 |
| 10.75 | 12.82 | 15.25 | 17.53 |
| 10.42 | 12.17 | 14.62 | 17.05 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 10.09 | 12.03 | 14.22 | 16.70 |
| Std. Dev. | 0.50 | 0.62 | 0.93 | 0.59 |
| Coef. of Var. | 0.05 | 0.05 | 0.07 | 0.04 |
| Norm. Ref. | 1.00 | 1.19 | 1.41 | 1.66 |

### MIDWAY (SOIL-GROUP 3)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 18.60 | 21.30 | 25.20 | 30.90 |
| 15.33 | 19.34 | 21.80 | 23.27 |
| 13.40 | 16.72 | 21.10 | 21.50 |
| 17.20 | 19.23 | 23.61 | 25.80 |
| 14.53 | 17.10 | 19.10 | 22.40 |
| 14.63 | 18.96 | 21.67 | 21.50 |
| 16.43 | 18.46 | 21.93 | 22.43 |
| 16.22 | 19.40 | 21.80 | 24.30 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 15.79 | 18.81 | 22.02 | 23.97 |
| Std. Dev. | 1.66 | 1.43 | 1.78 | 3.38 |
| Coef. of Var. | 0.11 | 0.08 | 0.08 | 0.14 |
| Norm. Ref. | 1.00 | 1.19 | 1.39 | 1.52 |

### FARM HILL ROAD (I280 ROAD CUT)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 21.95 | 20.10 | 16.23 | 16.71 |
| 24.10 | 21.83 | 18.10 | 18.32 |
| 22.22 | 19.30 | 18.33 | 16.62 |
| 22.98 | 20.52 | 17.83 | 16.93 |
| 25.00 | 21.90 | 19.94 | 19.10 |
| 26.43 | 20.95 | 20.95 | 17.94 |
| 28.60 | 24.84 | 22.70 | 20.30 |
| 25.10 | 21.60 | 18.63 | 17.64 |
| 24.22 | 21.32 | 17.90 | 16.76 |
| 22.00 | 19.81 | 16.90 | 15.94 |
| 20.42 | 18.20 | 14.75 | 15.11 |
| 20.21 | 17.61 | 15.03 | 14.04 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 23.77 | 20.68 | 18.29 | 17.28 |
| Std. Dev. | 2.40 | 1.87 | 2.23 | 1.71 |
| Coef of Var. | 0.10 | 0.09 | 0.12 | 0.10 |
| Norm. Ref. | 1.00 | 0.87 | 0.77 | 0.73 |

### FARM HILL ROAD (SOIL)

Reflectance

| 4 | 5 | 6 | 7 |
|---|---|---|---|
| 7.48 | 11.07 | 14.91 | 16.60 |
| 7.09 | 10.71 | 14.40 | 16.20 |
| 6.97 | 10.78 | 14.68 | 17.05 |
| 7.25 | 10.93 | 14.43 | 16.39 |
| 7.27 | 12.04 | 14.68 | 17.81 |
| 7.33 | 11.25 | 15.15 | 17.05 |
| 7.66 | 11.29 | 15.19 | 16.81 |
| 8.42 | 12.58 | 16.22 | 17.81 |

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Mean | 7.43 | 11.33 | 14.96 | 16.97 |
| Std. Dev. | 0.45 | 0.65 | 0.58 | 0.59 |
| Coef. of Var. | 0.06 | 0.06 | 0.04 | 0.03 |
| Norm. Ref. | 1.00 | 1.52 | 2.01 | 2.28 |

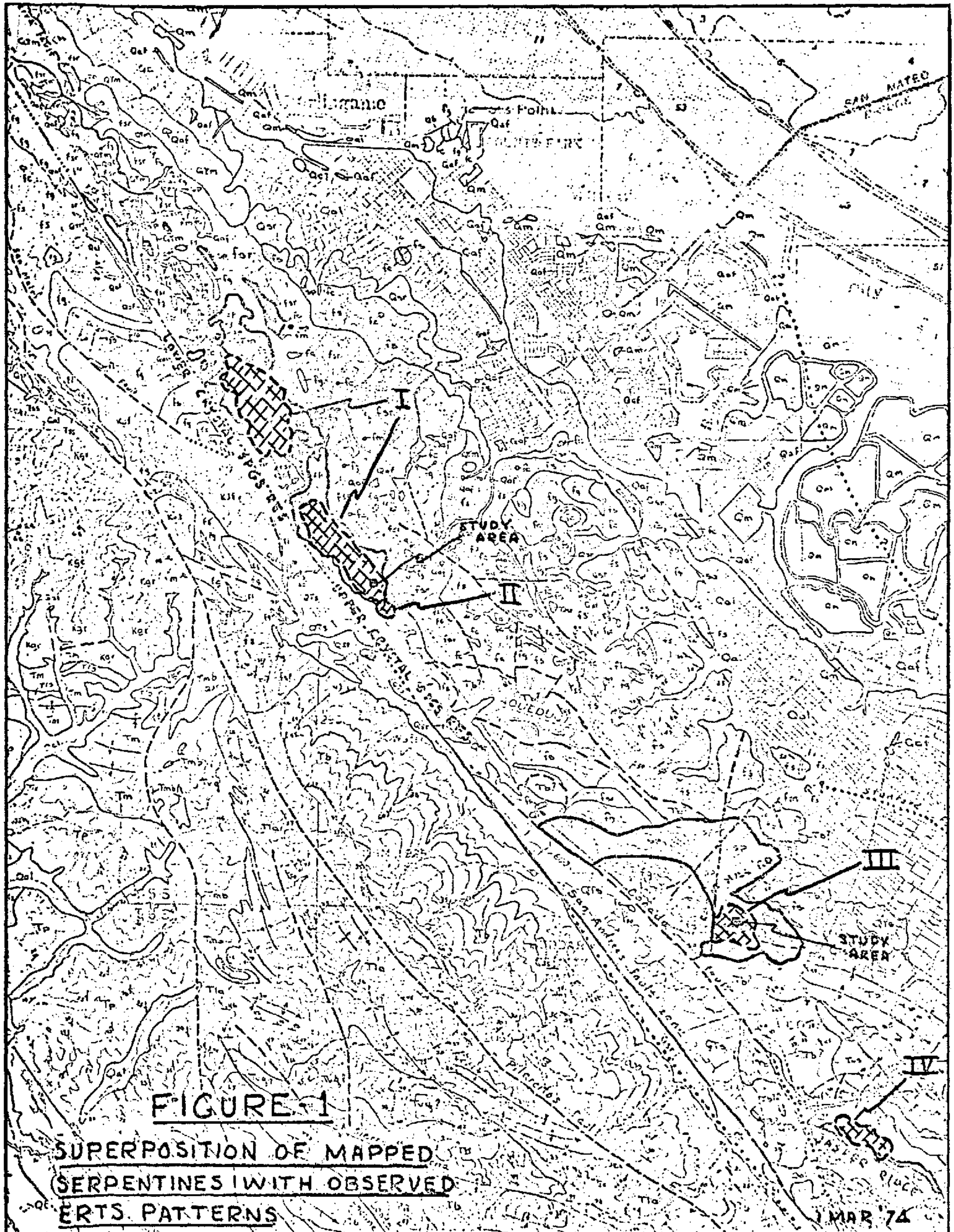



FIGURE 1

SUPERPOSITION OF MAPPED SERPENTINES WITH OBSERVED ERTS PATTERNS

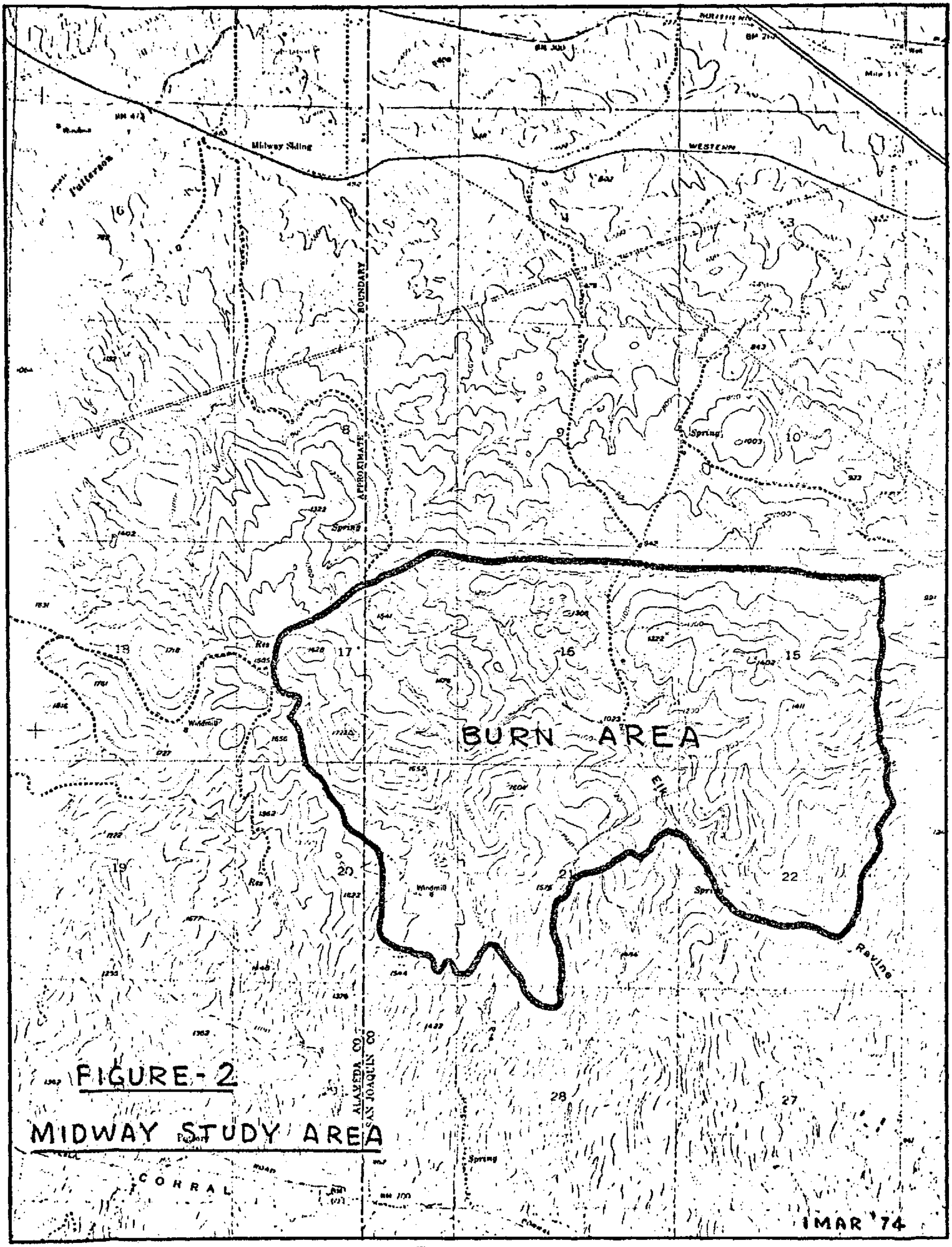



FIGURE-2

MIDWAY STUDY AREA

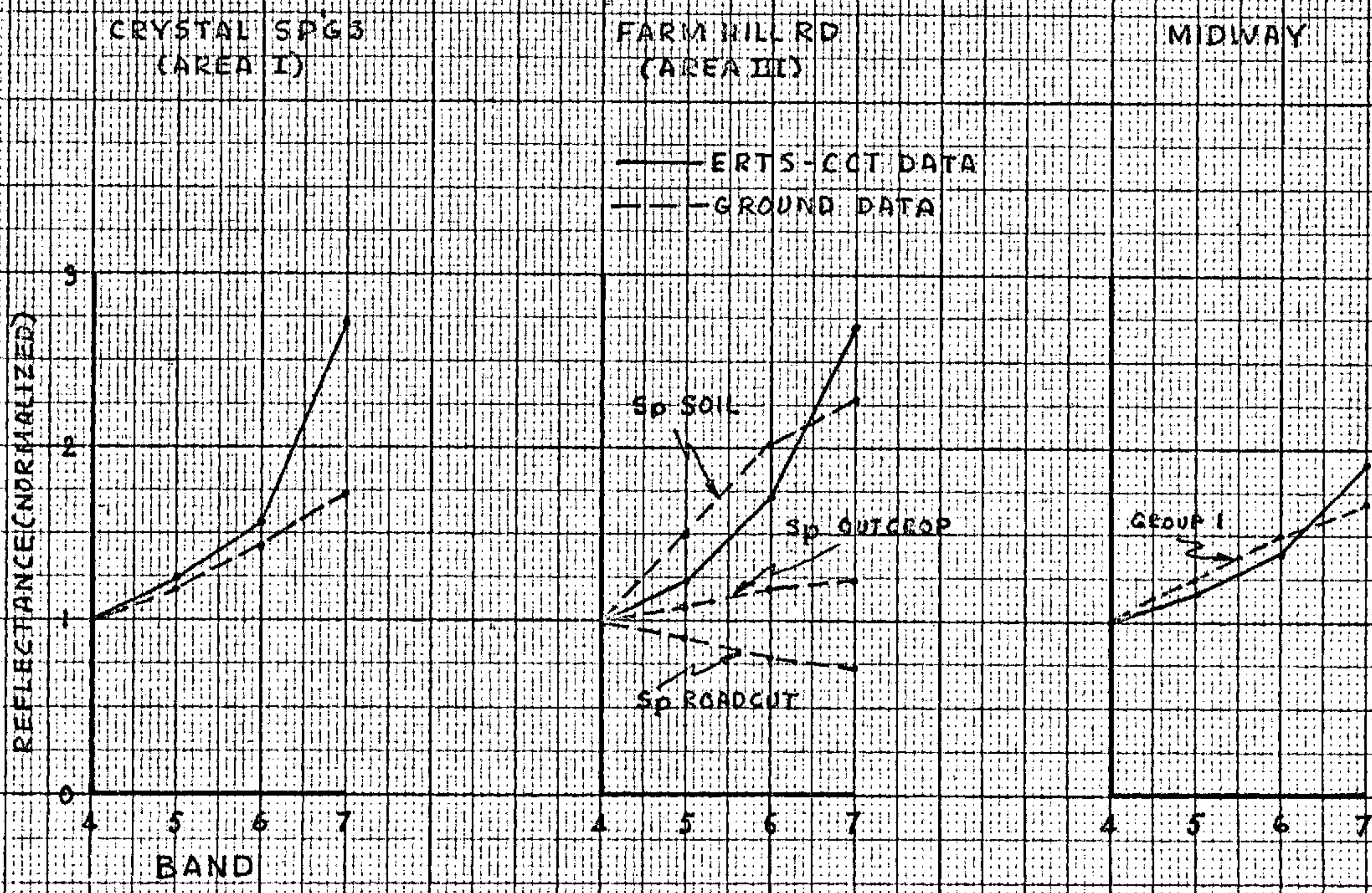



FIGURE 3 - COMPARISON OF NORMALIZED ERTS-CCT AND GROUND MEASURED REFLECTANCE

16 MAY '74


FIGURE 4 - RADIANCE AND NORMALIZED REFLECTANCE SPECTRA - MIDWAY

I. 2. APPLICATIONS RESULTS

## A. Biomass Study-The Vegetation of the Stanford Grassland (Preliminary Report)

by Roberta Sears

A study of the vegetation at selected sites in the Stanford grassland has been undertaken to aid in the interpretation of reflectance data from those sites.

The Stanford grassland is a typical representative of the California Valley Grassland plant community (1). It has been subjected to grazing by cattle for decades which has changed the species composition entirely. Few of the original native species remain. Most of the species of grasses and broadleaved plants found in the grassland today have been introduced from the Mediterranean region (2,3).

A preliminary study was done to determine the species of the grasses and broadleaved plants growing at the study sites. Plants were collected in early May in various stages of flower and seed formation. The plants were identified to the level of genus of species using Thomas's Flora of the Santa Cruz Mountains of California and Munz's A California Flora. Specimens of each species were dried and pressed for a permanent reference collection. The plant species found in the Stanford grassland study sites are:

| | Botanical Name | Common Name |
|---|---|---|
| Grasses: | Bromus mollis | Soft chess |
| | Avena fatua | Wild Oats |
| | Lolium multiflorum | Ryegrass |
| | Bromus rigidus | Ripgut grass |
| | Hordeum leporinum | Foxtail |
| | Hordeum glaucum | Wall barley |
| Broadleaved Plants: | Erodinm sp. | Filaree, needle plant |
| | Geranium sp. | Geranium |
| | Medicago sp. | Bur clover |
| | Convolvulus arvensis | Morning glory, bindweed |
| | Bellardia trixago | Bellardia |
| | Eschscholzia californica | California poppy |
| | Rumex sp. | Sorrel |

Initial observations also revealed that the vegetation at the study sites was variable in species composition, plant size, precent cover and time of onset of senescence and drying.

From May 15 to May 22 a detailed study of the vegetation at 44 sites was made to determine the species composition, biomass and stage of plant growth. At each site the vegetation was treated in the following manner:

1. A square, 0.5 m on a side, was marked off at a randomly selected site.
2. Reflectance relative to $BaSO_4$ was recorded.
3. The plant species were determined.
4. The percent contribution of each species to the total biomass was estimated by eye.
5. All vegetation within the square was cut off at ground level and put in an airtight plastic bag and taken to the lab.
6. Reflectance after cutting was recorded.
7. Total fresh weight of the vegetation at each site was measured.
8. All plant material was dried in ovens at 80° to 90°C for 48 hours.
9. Total dry weight was measured.

The reflectance at each site was measured before and after the removal of the vegetation cover, using the ERTS radiometer, 15°FOV bidirectional geometry (Table I). The sites vary considerably with regard to species composition, fresh weight, dry weight (biomass) and the ratio, dry weight/fresh weight. The dry weight/fresh weight measurement indicates the degree to which the plants have dried out. As the vegetation dries, the green color is lost and the leaves turn to yellow-green then tan. The ratio of dry weight to fresh weight is therefore, an indirect measure of the "greeness" of the vegetation. These data will be used to interpret the reflectance data taken at the same sites.

1. Munz, P.A. and D.D. Keck (1965) A California Flora, University of California Press, Berkeley and Los Angeles.
2. Thomas, J.H. (1961) Flora of the Santa Cruz Mountains of California, Stanford University Press, Stanford, California.
3. McNaughton, S.J. (1968) Ecology 49: 962-972 Structure and function of California grasslands.

TABLE I STANFORD GRASSLANDS

| Site Number | Total wet weight | Total dry weight | Dry weight / Wet weight |
|---|---|---|---|
| 946 | 158.3 | 78.9 | .495 |
| 942 | 117.3 | 52.8 | .450 |
| 941 | 105.0 | 62.8 | .598 |
| 947 (green) | 213.2 | 120.0 | .563 |
| 917 (dry) | 79.7 | 57.3 | .719 |
| 949 | 119.3 | 71.0 | .595 |
| 951 | 150.0 | 81.9 | .546 |
| 953 | 131.1 | 82.7 | .631 |
| 954 | 83.9 | 55.5 | .662 |
| 955 | 70.1 | 54.6 | .779 |
| 980 | 342.4 | 138.4 | .404 |
| 982 | 421.8 | 173.0 | .410 |
| | | | |
| 914 | 282.5 | 121.0 | .428 |
| 916 | 306.4 | 128.2 | .418 |
| 917 | 184.6 | 103.2 | .559 |
| 920 | 193.9 | 104.6 | .539 |
| 930 | 333.7 | 160.0 | .479 |
| 931 | 271.7 | 155.0 | .570 |
| 932 | 406.5 | 180.0 | .443 |
| 934 | 307.3 | 130.1 | .423 |
| 940 | 344.8 | 122.3 | .355 |
| 942 | 430.8 | 149.1 | .346 |
| 943 | 313.9 | 138.5 | .441 |
| 944 | 290.0 | 132.3 | .456 |
| | | | |
| 905 | 354.5 | 168.2 | .474 |
| 906 | 168.0 | 50.1 | .476 |
| 908 | 209.5 | 95.8 | .457 |
| 909 | 192.5 | 113.6 | .590 |
| 991 | 220.4 | 98.9 | .449 |
| 992 | 463.5 | 175.7 | .379 |
| 994 | 404.3 | 165.4 | .407 |
| 996 | 423.2 | 197.2 | .466 |
| 986 | 411.0 | 152.4 | .371 |
| 984 | 317.7 | 126.4 | .398 |
| | | | |
| 905 | 325.5 | 127.1 | .390 |
| 952 | 403.5 | 145.3 | .360 |
| 954 | 367.1 | 164.2 | .447 |
| 970 | 86.5 | 45.1 | .521 |
| 972 | 152.0 | 74.1 | .488 |
| 973 | 354.9 | 171.4 | .483 |
| 929 | 191.7 | 81.8 | .427 |
| 931 | 294.6 | 131.4 | .446 |
| 933 | 86.2 | 48.2 | .559 |
| 936 | 60.3 | 42.2 | .700 |

I. 3. SUITABILITY OF ERTS DATA

No comment

J. SIGNIFICANT RESULTS

None at this stage, results in active study and analysis.

K. NEXT PERIOD

1. Completion of Biomass/Reflectance Study.

2. Airborne system development.

L. PUBLISHED MATERIALS

1. Paper attached (SRSL #74-4)

2. Two sets of Abstracts submitted to IEEE for possible acceptance at September meeting in Philadelphia.

M. RECOMMENDATIONS

None

N. CHANGES IN STANDING ORDER FORMS

None

O. ACCESSION LIST FOR ERTS IMAGERY/TAPES OVER STANFORD

Enclosed

P. DATA REQUEST FORMS SUBMITTED

None

Q. MAILING LIST

At end of report.

STANFORD SRSL TECHNICAL REPORT #74-4

# STANSORT: STANFORD REMOTE SENSING LABORATORY PATTERN RECOGNITION AND CLASSIFICATION SYSTEM

F.R. Honey, Research Associate
A. Prelat , Research Associate
R.J.P. Lyon, Principal Investigator

School of Earth Sciences
Stanford University
Stanford, California 94305

## ABSTRACT

The principal barrier to routine use of the ERTS multispectral scanner computer compatible tapes, rather than photointerpretation examination of the images, has been the high computing costs involved due to the large quantity of information (4 X $10^6$ bytes) contained in a scene. STANSORT, the interactive program package developed at Stanford Remote Sensing Laboratories alleviates this problem, providing an extremely rapid, flexible and low cost tool for data reduction, scene classification, species searches and edge detection. The primary classification procedure, utilizing a search, with variable gate widths, for similarities in the normalized, digitized spectra is described in section 2, with associated procedures for data refinement and extraction of information. The more rigorous statistical classification procedures are described in section 3. The programs have been developed on an interactive computer (PDP-10) with the non-specialist operator in mind, requiring very little computing experience for their operation.

## 1. INTRODUCTION

When confronted with the overwhelming quantity of data available on magnetic tapes from the ERTS-1 multispectral scanner system, it may appear to an investigator that reduction, analysis and presentation of significant interpretations of the taped data using a digital computer would be an expensive and time consuming approach. In comparison, visual examination of the standard imagery product generated by NASA from the original data (or color-combinations of the data) is less expensive, though probably more time consuming. However, this photo-geologic approach can not be correlated readily with field-or laboratory-measured data (here referred to as "ground data").

Grouping data for a scene into distinguishable classes, for comparison with known ( or suspected) geologic, geobotanical, crop or urban features, may be accomplished with either a statistical or a non-statistical classification procedure. These procedures may be divided further into supervised procedures (requiring training groups consisting of either areas which are known to be uniform, or of digitized "ground data"), and unsupervised (self-training) procedures in which the data is split into its distinguishable groups with no prior knowledge of the number, or species, of these groups. The statistical classification procedures are reviewed and evaluated in section 3.

*This research report is based upon work performed under NASA Contract NAS 5-21884, the receipt of which is gratefully acknowledged.

## 2. NON-STATISTICAL PATTERN RECOGNITION, CLUSTERING PROCEDURE

### 2.1 UNSUPERVISED CLUSTERING PROCEDURE

The non-statistical gating procedure described below developed as a result of manual examination of digitized spectra plotted for a area using the four MSS bands. It was realized that, for an area of reasonable size (e.g. 30 X 30 pixels, where a pixel, 57 meters X 79 meters, covers approximately 1 acre) only a finite number of patterns appeared. Figure 1a illustrates the patterns appearing when traversing across a row of pixels for a scene in Mono Lake, California (Figure 2). Although overall levels vary it may be observed that similar patterns appear across this traverse. This variation in level is due mainly to a topographic effect: slopes facing the sun appear brighter than slopes facing away from the sun; and partially due to the texture of the surface within the pixel: smoother surfaces (generally composed of smaller, closely-packed particles) appear brighter than coarser, rough surfaces of the same material. This change in brightness level may be reduced considerably by normalizing to one of the channels. The effect of normalizing to channel 4 the patterns plotted in Figure 1a is illustrated in Figure 1b.

These normalized patterns could be grouped into classes now, just by their shape. This is a tedious approach when performed manually, yet the concept provides a very simple, rapid and economical technique when performed by computer. The computer is required to perform the minimum of operations, all arithmetical, merely comparing values within some preset range, to discriminate different classes.

#### 2.1.1 CONVERSION OF ERTS DIGITAL VOLTAGES TO REFLECTANCE VALUES

To compare the satellite results with ground data it is necessary to convert the ERTS digital values for each channel to some more absolute measurement which will be virtually independent of sun elevation and atmospheric effects. For this conversion, two (or more) "standard" targets are required in any ERTS scene. One of these should have as low a reflectance as possible (preferrably zero percent), so that it may be assumed that energy impinging on the detector from the direction of the low reflectance target arises only from the radiation back scattered into the sensor field of view from the atmosphere. The four ERTS channel values for this low reflectance target may then be subtracted from the corresponding values for all other pixels within the scene to give a measure of the radiation impinging on the detectors which arises specifically from the radiation reflected from each pixel. Obviously, not all scenes have zero-reflectance targets within them-in this case several targets having low reflectances must be chosen, and a linear extrapolation performed to give reasonable values for a zero-reflectance target.

To convert these corrected ERTS voltages to reflectance, a standard high reflectance target (or targets) within a scene must be chosen. By ratioing the corrected channel voltages for an unknown pixel to the corresponding corrected channel voltage for the high reflectance target, and then multiplying by the respective band pass reflectance (known from ground data) of the standard target(s) yields the band pass reflectance of the unknown pixels within the scene. This procedure may be represented explicitly by

$$\rho_{u,i} = \frac{Vu,i - Vz,i}{Vx,i - Vz,i} \times \rho_{s,i} \qquad (1)$$

where i is the channel number. $\rho_{u,i}$ is the bandpass reflectance of the pixel being examined. $\rho_{s,i}$ is the band pass reflectance of the standard target(s). Vu,i is the voltage in the ith channel of the unknown pixel. Vz,i is the voltage of the zero reflectance target Vs,i is the voltage of the high reflectance standard target.

The low and high reflectance standard targets must be chosen to cover an area at least three pixels square, preferrably larger, so that the center pixel (or pixels) of the standard target give voltages arising entirely from the standard targets, not affected by bordering species, particularly if the area is to be repetitively sampled by other ERTS overpasses.

Again, as would be expected, only a small, finite number of patterns appear when these reflectances are plotted. Levels of the reflectance patterns vary, due to the topographic and texture effects, but these variations may be removed by normalizing to one of the four channels (c.f. the ratio technique described by Vincent (1972)).

The technique described above has been developed into a rapid, inexpensive clustering program for an interactive computer system. With the man-machine interaction the investigator can rapidly choose his scene, display shadeprints as maps (for location) and optimize the gate used in the clustering to suit his particular requirement and the size or complexity of the area being examined.

The classification procedure searches through the array for the first unclassified pixel and a descriptor (alphabetic) assigned for this pattern. The remainder of the unclassified pixels are then compared with this "standard" pattern. If the pattern of an unclassified pixel agrees with the current "standard" pattern within the gate width, it is given the same descriptor as the current "standard". The program recycles until all pixels have been classified, or until the number of classes exceeds twenty six (arbitrary). The set of "standard" patterns generated during the search are stored in an array to be used, if required, to classify another scene in the same area. In this manner very large areas may have a cluster analysis performed on them.

### 2.2 SUPERVISED CLASSIFICATION PROCEDURE

Whilst the unsupervised clustering technique described above is useful for examination of unknown scenes, separating them into their spectrally distinguishable species, a similar but more powerful technique may be used to search the band pass reflectance patterns for known types, using the ground spectral data. The gate generally employed in this method is two or three times the largest standard deviation of the normalized bandpass reflectances of the measured ground target. Obviously for this approach the standard targets in a scene must be chosen carefully, according to statistical sampling technique, and their reflectances measured.

### 2.3 NOISE AND SMOOTHING

When radiance data for a large, uniform scene (eg. water) is examined, noise with a six row periodicity may be observed, resulting from the basic detector design in the MSS. This noise is in phase for channels 4,5 and 7, but out of phase by two scan lines for channel 6. The origin of this noise is not clear-it may be due to slight differences in detector responses or to a misalignment of the detector array. This noise must be removed as completely as possible for a reliable cluster analysis to be performed, not to do so makes each sixth line a separate class. Since it is in the form of one or two unit spikes every sixth line, it is ideally suited to treatment by a digital smoothing technique described by Savitsky and Golay (1964). As the convolution blurs the image slightly, it must be performed in two dimensions in order that the change in contrast will be uniform.

The result of this smoothing is evident in Figures 3a and 3b. Figure 3a is the result of a cluster analysis without smoothing on an island in Mono Lake. Figure 3b shows the analysis of the same scene (with the same tolerance) after smoothing of the data. The water surrounding the island appears non-uniform in the unsmoothed cluster analysis result, but becomes uniform after smoothing.

### 2.4 BOUNDARY SEARCHES-EDGE DETECTION

After smoothing it is possible to search for abrupt changes in contrast, such as occur at sharp boundaries, deep valleys or borders between materials with large differences in reflectivity. In this manner, a search for linear, curvilinear or elliptical features may be pursued and, hopefully, some correlation between the presence of these features, their intersections and changes in classification using the clustering algorithm may be observed. Figure 4 illustrates the result of "edge detection" for the same scene clustered in Figure 3b.

### 2.5 COST IN COMPUTATION

The clustering technique as outlined above has proven to be very rapid. No direct comparison with a statistical clustering program is available yet, although the BMD07M stepwise discriminant program has been employed and found to require approximately ten times the computing time, even when using training groups initially generated by our unsupervised clustering program. Obviously the time required for classification is a function of the number of pixels in the scene, and also the width of the gate. Figure 5 illustrated the times required as a function of the area of the scene in the vicinity of the island in Mono Lake for different gate widths.

At present the program is being extended to "defocus" the scene, so that large areas may be examined by averaging four, nine or sixteen pixels (in a square), clustering them to look for broad patterns, then examining sub-scenes of the large scene in 1 X 1 pixel detail. Statistical procedures are being inserted to provide means, standard deviations and histograms of areas classified by the clustering algorithm, or of areas selected by the operator.

The program has been developed with the non-specialist in mind. It is completely interactive and self explanatory so that a person with no computing experience is able to examine ERTS tapes. It is designed specifically for use with a limited budget, with fast turn around time.

## 3. SRSL NUMERICAL CLASSIFICATION TECHNIQUES

This section outlines the theoretical basis of numerical classification technqiues used in conjunction with the procedure described above. Together they constitute the software system for analysis of ERTS multispectral data in operation at the Stanford Remote Sensing Laboratory. Four numerical classification procedures are discussed, two of which are supervised and two of which are unsupervised.

### 3.1 SUPERVISED CLASSIFICATIONS

A classification is supervised when data points of unknown origin are assigned into a priori defined classes.

#### 3.1.1 NEAREST NEIGHBOR

Most of the classification techniques depend upon the assumption that samples have been drawn from a normal population. The Nearest Neighbor method makes inferences without any assumptions as to the form of distribution in the population. Such procedures are said to be non-parametric or distribution-free. The technique consists of classifying unknown data points into known categories through comparison with known data. Each unknown sample is allocated to that group to which it is nearest in terms of the $D^2$ generalized-distance statistic. Thus, the degree of similarity between two samples is provided by the distance that separates them; the shorter the distance the greater the degree of similarity, and vice-versa.

The nature of non-parametric statistical inferences usually requires testing with large amounts of data to achieve a respectable degree of accuracy (Swtizer el.al.,1968).

#### 3.1.2 MULTIVARIATE DISCRIMINANT ANALYSIS

Multiple discriminant analysis is a statistical method of assigning samples in a probabilistic manner to previously defined populations on the basis of a number of variables considered simultaneously.

The use of the discriminant function may be considered in terms of a population consisting of X variables, which forms a cluster of points in X-dimensional space. A second population, described by the same X variables, consists of a second cluster of points. The linear combination of variables, that defined a multi-dimensional plane efficiently separating the two clusters of points is the discriminant function. The degree of distinctness of the two clusters can be analyzed by measuring the "distance" between their multivariate means. Once this distinctness has been established and the separating plane computed, additional unknown samples can be assigned in one or the other of the groups depending on which side of the discriminant plane they fall.

The basic assumptions about the data are: (i) the observations in each group are randomly chosen; (ii) the probability of an unknown observation belonging to either group is equal; (iii) the frequency distributions of the groups are each multivariate Gaussian distributions with a common covariance matrix. This means that the distributions have identical bell-shapes and differ only in that they are centered at different points.

The BMD07M is a stepwise discriminant analysis program, and is part of a series of bio-medical statistical analysis programs compiled by the UCLA Health Services Computing Facility. The stepwise discriminant analysis indicates that the computation of discriminant functions is not simply based on the original variables considered as a whole, but rather that the variables are entered separately and consecutively by order of discriminatory power. The advantage of this procedure is to recognize the relative importance of each variable in classifying the samples into the different groups. Ranking the variables by predictive power permits a concentration of efforts on those factors which are important for classifying groups, and this can represent a highly effective means of reducing costs of data collection and processing.

The computational procedure of the stepwise discriminant technique is described in the user's guide of the BMD07M program (Dixon, W.J.,ed.,1972).

### 3.2 UNSUPERVISED CLASSIFICATIONS

Classification is unsupervised when similar data points are placed into an unknown number of distinct classes in which the data points of each class have a closer similarity to each other than to the data points in all other classes.

#### 3.2.1 CLUSTER ANALYSIS BASED ON DISTANCE-SIMILARITY MATRIX

A distance-similarity matrix is obtained to determine the relationship of the data. The use of distance is based on the concept that a quantitative measure of the degree of similarity between two variables or two samples is provided by the distance that separates them in a rectangular coordinate system. As the distance becomes shorter the degree of similarity increases and vice versa. The sample points are grouped or clustered in a hierarchical dendritic network (dendogram) in which their interrelationships, as contained in the distance-similarity matrix, are shown with greatest simplicity.

In a two-dimensional case, two samples are plotted according to the values of the two variables, X and Y. The distance between these two points is, by simple geometry, the square root of the sum of the squared differences between X and Y values of the two points; as in a right triangle the square of the hypotenuse is equal to the sum of the squares of the two sides of the triangle.

This calculation of the distance assumes that the input variables (or the axes from which they are measured) are uncorrelated, that is, orthogonal or at right angles to each other. However, most raw variables are correlated to different degrees so that the coordinate axes would not be at right angles and the simple Euclidean distance formula would be inaccurate. To overcome this difficulty, the raw variables are transformed to a new set of uncorrelated orthogonal variables by a series of linear transformations (for details see Sebestyen, 1962). In calculating the distance coefficients for the similarity matrix it is convenient to limit its value to the range 0.0 to 1.0. To satisfy this requirement, the original data is transformed, so that all the measurements are positive and range from zero to one.

Finally, a cluster analysis is performed to measure the degree of similarity between samples on the basis of the distance-similarity matrix. Distances close to 0.0 represent maximum similarity, distances close to 1.0 represent minimum similarity. A cluster diagram is printed out with the value of the distance coefficients. Groups of similar samples can be selected at any desired level of similarity, and each group can be plotted on a geometric matrix or map.

The present procedure accomplishes clustering by computing a matrix to measure all pairwise similarities between data points on the basis of the measurements corresponding to the channels of the scanner. The procedures cannot be used when large data sets are to be analyzed because the size of the distance-similarity matrix becomes too large for the core storage requirement of the computing equipment.

4.1.2 ISOMIX: AN ITERATIVE CLUSTERING PROGRAM

Similar cluster programs have been developed by Stanford Research Institute (Ball and Hall, "ISODATA", 1965), Purdue University (Wacker and Landgrebe, 1971) and Lockheed Electronic Company (Kan, Holley and Parker, "ISOCLS", 1973). ISOMIX (Stanford) essentially follows the iterative clustering procedure of ISOCLS; however, new statistical techniques have been added to help the analyst in the interpretation and evaluation of the final data points grouped into clusters. The following is an outline of the main steps: The program first computes the initial cluster centers and assigns them to regions of high sample-point density. Then the samples are sorted, one by one, on the basis of their distance from a set of initial cluster centers which create a cluster of data points or vectors X and Y is defined as

$$d(X,Y) = \sum_{i=1}^{n} |X_i - Y_i| \qquad (2)$$

After the samples have been sorted the mean and standard deviation for each subset in each dimension (variable) is computed.

Those clusters which contain only a few sample points are discarded. Splitting of the clusters takes place if the standard deviation in any dimension is greater than a suitable threshold specified by the analyst. When the cluster is divided two new centers are formed. These centers are $(\mu_1, \mu_2, \ldots, \mu_k + \sigma_k, \ldots\ldots \mu_n)$ and $(\mu_1, \mu_2, \ldots, \mu_k - \sigma_k, \ldots \mu_n)$ where $(\mu_1, \mu_2, \ldots, \mu_n)$ and $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ are the mean and standard deviation for the dimensions in the original cluster, and in the Kth dimension the original cluster contains the largest standard deviation.

The degree of distinctness of the clusters is measured by the similarity of "cluster centers" attached to regions of high density of data points. The distance or measure of similarity between two clusters $C_1$ and $C_2$, where $C_1$ is characterized by $\mu^{(1)} + (\mu_1^{(1)}, \ldots \mu_n^{(1)})$ and standard deviation $\sigma^{(1)} = (\sigma^{(1)}, \ldots \sigma_n^{(1)})$ and $C_2$ by $\mu^{(2)}$ and $\sigma^{(2)}$ (Kan, Holley and Parker, 1973), is defined as

$$D(C_1, C_2) = \left[ \sum_{i=1}^{n} \frac{\left(\mu_i^{(1)} - \mu_i^{(2)}\right)^2}{\sigma_i^{(1)} \cdot \sigma_i^{(2)}} \right]^{1/2} \qquad (3)$$

If the distance $D(.,.)$ between two clusters is less than a specified threshold, the two cluster centers are merged into one at a weighted mean of the two original clusters.

The progress is cycled repeatedly until the standard deviation in every channel of the generated clusters is less than the specified threshold, or the maximum number of clusters desired by the analyst is reached.

ISOCLS's chaining algorithm is used to link those subclusters which are close to at least one other subcluster. This linking process determines the subpopulations, the union of which constitutes the parent population.

In the last step the overall areal proportions of various clusters are obtained. For example, if $p_j$ is the areal proportion of a specified cluster j, $n_j$ is the number of sample points counted of the specified cluster j, and N the total number of sample points, then the usual estimator of the areal proportion $p_j$ is $\hat{p}_j = n_j/N$. Finally in the last step, the pattern complexity which gives the spatial scale of variation is also obtained. A pattern that has a cluster A with its samples in a contiguous body is less complex than another with the same proportion of cluster A distributed in many scattered smaller units. One index to express the pattern complexity (Switzer, 1973) is

$$X = \frac{\text{total length of boundaries between different sample points}}{(\text{area of region})\ 1/2} \qquad (4)$$

The value of X is invariant to the choice of the measurement unit. As the X value increases the pattern grows in complexity.

The output gives the statistics for each cluster and includes a map showing the final cluster assignments of all the points in the area analyzed. These maps are geographic matrices preserving the original position of the data points.

## 4. REFERENCES


Ball, G.H. and Hall, D.J.,1965, "ISODATA, a Novel Method of Data Analysis and Pattern Classification". Stanford Research Institute 1-16, Menlo Park, California.

BMD-Biomedical Computer Programs. Edited by W.J. Dixon, University of California, 1970, Publications in Automatic Computation, no.2, University of California Press.

Kan, E.P., Holley, W.A., and Parker, Jr.,H.D.,1973, "The JSC Clustering Program ISOCLS and its Applications", Proceedings of the IEEE on Machine Processing of Remotely Sensed Data, Catalog no.73 CHO 834 - 2ItE.,pp.4b36-4b50.

Savitsky,A. and Golay, M.J.E.,1964, "Analytical Chemistry", v.36, p.1627.

Sebestyen, George, 1962, Decision-Making Processes in Pattern Recognition. New York; The Macmillan Company.

Switzer, P.,et.al,1968, "Nearest Neighbor-A New Non-Parametric Test Used for Classifying Spectral Data". Stanford RSL Technical Report #68-3.

Switzer, P.,1972, "Applications of Random Process Models to Descriptions of Spatial Distributions of Qualitative Geologic Variables". 24th International Geological Congress. Montreal, Canada.

Vincent, R.K., 1972, "An ERTS Multispectral Scanner Experiment for Mapping Iron Compounds", Proc. 8th International Symposium on Remote Sensing of Environment, p.1239.

Wacker, A.G. and Landgrebe, D.A.,1970 "Boundaries in Multispectral Imagery by Clustering". IEEE Symposium on Adaptive Processes (9th) Decision and Control, pp.X14.1-X14.8

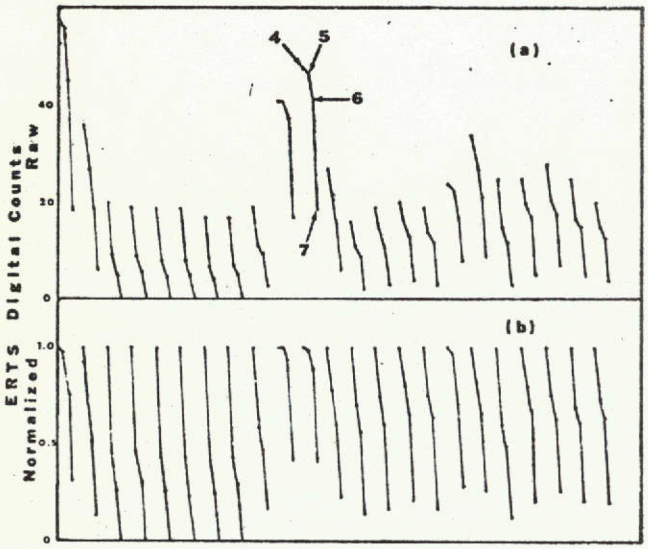



Figure 1. Digitized spectra resulting from plots of

(a) Raw ERTS digital voltages plotted against channel number.

(b) ERTS digital voltages normalized to channel 4, plotted against channel number,

for section of a scan line across Negit Island and portion of Mono Lake, California.

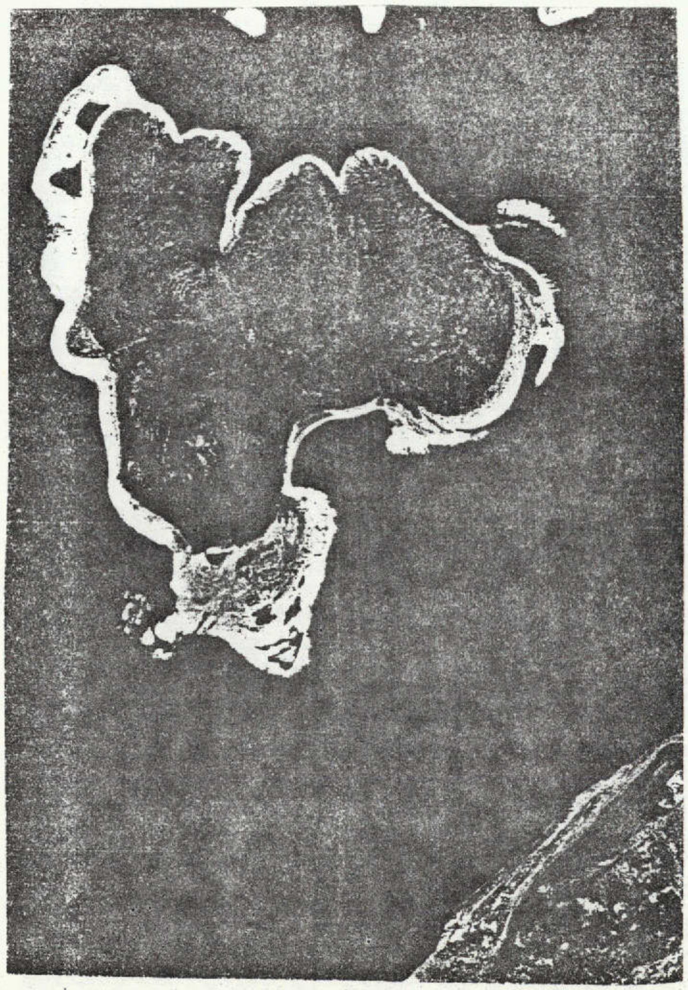


Figure 2. Negit Island, Mono Lake, California. Dark areas of island are basaltic lava flows and cones of varying texture, white "beaches" composed of calcareous tuffs.

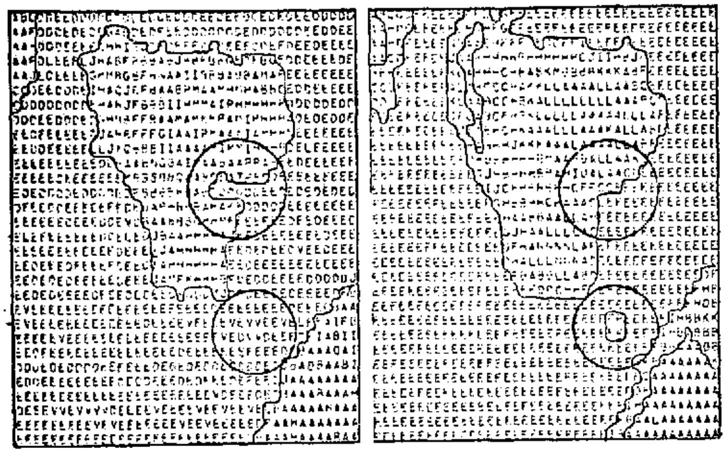


Figure 3. Cluster analysis result of the Mono Lake scene for

(a) unsmoothed data
(b) smoothed data

The water surrounding the island appears more uniform for the smoothed data. However, some structural detail of the island is removed due to the "defocusing" effect of the convolution.

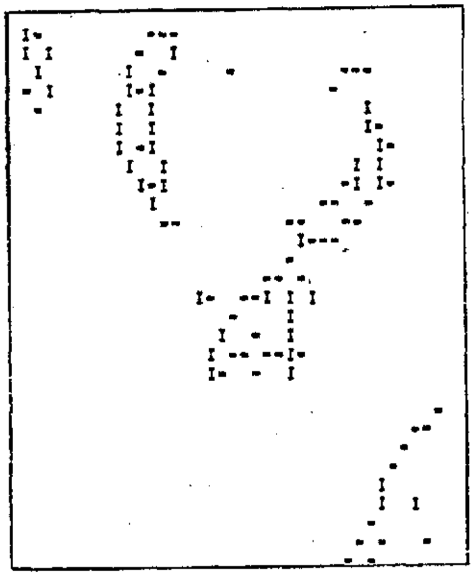


Figure 4. Result of "edge detection" on Negit Island scene. Only those borders with a large change in contrast are emphasized. Lower steps in the scanning technique would result in more detail.

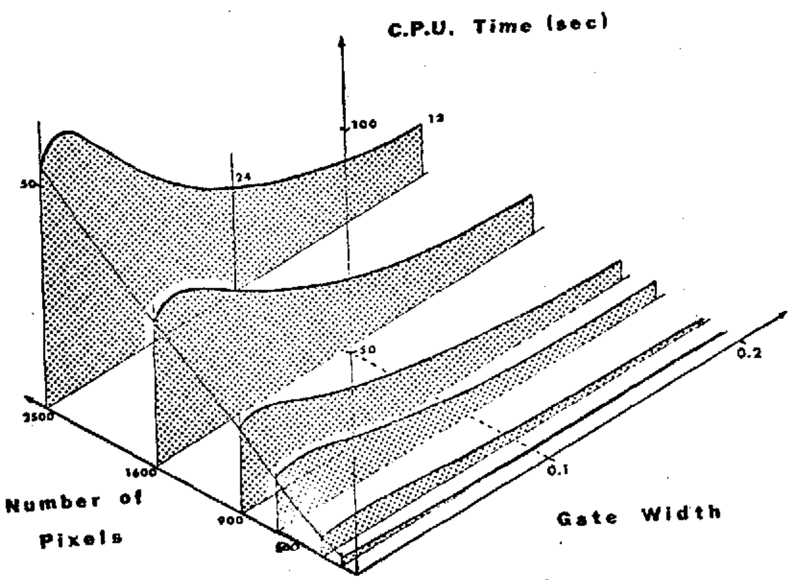
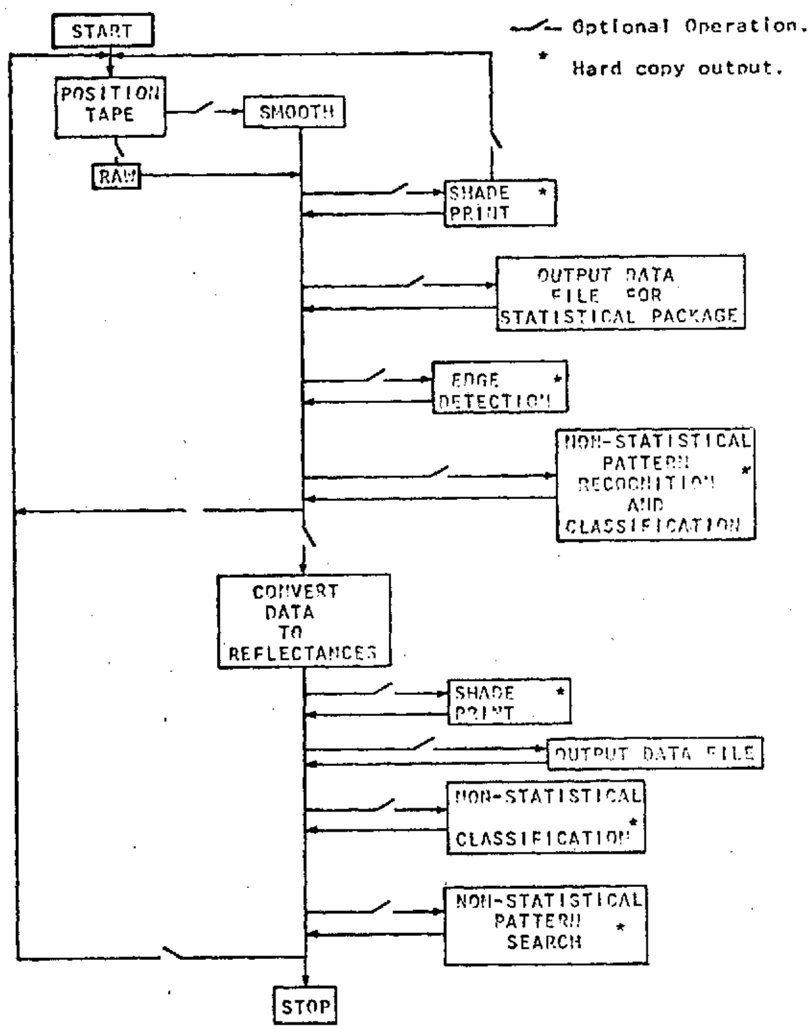



Figure 5. Computation times for unsupervised classification as a function of the number of pixels in the scene and of the gate width. In a first approach, gatewidths of approximately 0.1 normalized units are used.




Outline of program described in section 2. Most steps are optional, the operator being asked which procedures he wishes as the program proceeds. Results of each step may be displayed on the screen for examination.

24

# A COMPARISON OF OBSERVED AND MODEL-PREDICTED ATMOSPHERIC PERTURBATIONS ON TARGET RADIANCES MEASURED BY ERTS

by

F.R.Honey
R.J.P. Lyon

Stanford Remote Sensing Lab
Stanford University
Stanford, California 94305

## ABSTRACT

A study of the magnitude and effect of atmospheric backscattering and varying irradiance of a scene on the reflected radiance observed by ERTS. Radiance values of two standard reflectance targets are examined.

---

Dr. F.R. Honey
Department of Applied Earth Sciences
Stanford University
Stanford, California 94305
(415) 497-4147

Dr. R.J.P.Lyon
Department of Applied Earth Sciences
Stanford University
Stanford, California 94305
(415) 497-3262

# A COMPARISON OF OBSERVED AND MODEL-PREDICTED ATMOSPHERIC PERTURBATIONS ON TARGET RADIANCES MEASURED BY ERTS

by
F.R. Honey
R.J.P. Lyon

Stanford Remote Sensing Lab
Stanford University
Stanford, California 94305

SUMMARY

In order to be able to compare results from ERTS MSS data over a series of tapes, the perturbing effects of a variable contribution due to radiation scattered by the atmosphere into the detector field of view, and of the variation in the irradiance on a target with solar zenith angle, must be eliminated. These two effects may be compensated for, or entirely removed, by studying selected targets in a scene, one (or more) of low (zero) reflectance, one (or more) or high, known reflectance. In some scenes, however, suitable reflectance targets may not be obtained. When this occurs, atmospheric modelling must be employed to arrive at some values for the atmospheric scattering contribution, and for the irradiance on the scene.

Two targets of measured, constant reflectivity in the area of San Francisco, California are studied. The first standard, a waste products treatment pond at an oil refinery near Suisun Bay, having an area of approximately 0.3 square miles, and bandpass reflectances of <0.5% in all four bands, is assumed to have a zero contribution to the radiance recorded by ERTS. The radiance observed then arises entirely from atmospheric scattering. The variation in these radiance values as a function of solar zenith angle is compared with models for atmospheric scattering.

Dr. F.R.Honey
Department of Applied Earth Sciences
Stanford University
Stanford, California 94305
(415) 497-4147

Dr. R.J.P.Lyon
Department of Applied Earth Sciences
Stanford University
Stanford,California 94305
(415) 497-3262

26<

# A COMPARISON OF OBSERVED AND MODEL-PREDICTED ATMOSPHERIC PERTURBATIONS ON TARGET RADIANCES MEASURED BY ERTS

**F.R. Honey & R.J.P. Lyon**

The second target, a concrete parking apron for aircraft at Moffett Field, California, assuming that it remains dry during the period of study has constant reflectances of 27.8, 31.0, 30.0, and 32.3 percent bandpass reflectances in four MSS equivalent channels. Using these values, the radiance observed by ERTS may be corrected for the atmospheric contribution, and thus values for the irradiance on the target may be calculated. These values may be studied as a function of solar zenith angle and compared with results from models.

The technique of using standard targets within a scene is applied to a specific scene which contains an area of measured reflectivity.

27<

# A PROBABILISTIC MODEL TO DISCRIMINATE MULTISPECTRAL DATA MIXTURES

by

Alfredo Prelat

R.J.P. Lyon

Stanford Remote Sensing Lab

Stanford University

Stanford, California 94305

Multivariate discriminant analysis may be used to enhance the resolution capability of optical sensing device. Probabilistic functions are obtained to investigate the interrelationship among objects. In turn, the mixture of radiation received by the sensor from a ground resolution element containing different objects is quantified and defined.

Dr. A.E. Prelat
Department of Geology
Stanford University
Stanford, California 94305
(415) 497-4147

Dr. R.J.P. Lyon
Department of Applied Earth Sciences
Stanford University
Stanford, California 94305
(415) 497-3262

# A PROBABILISTIC MODEL TO DISCRIMINATE MULTISPECTRAL DATA MIXTURES

by

Alfredo Prelat

R.J.P. Lyon

Stanford Remote Sensing Lab
Stanford University
Stanford, California 94305

SUMMARY

The probabilistic model design attempts to solve the problem of variability of spectral signatures. This variability occurs when the instantaneous ground resolution elements contain different objects. As a consequence, the sensing devise receives a mixture of the radiation from each of the separate objects. These mixtures need to be recognized and defined when numerical classification techniques or pattern recognition algorithms are used to process the data.

The multivariate statistical technique known as discriminant analysis is used to classify the remote multispectral data into different groups. Discriminant scores are computed for each observation (pixel) by linear transformations, and the geometric planes separating different groups are established.

The basic assumptions about the data are (i) the observations in each group are randomly chosen; (ii) the probability of an unknown observation belonging to either group is equal; (iii) the frequency distributions of the groups are each multivariate Gaussian distributions with a common covariance matrix. This means that the distributions have identical bell-shapes and differ only in that they are centered at different points.

Dr. A.E. Prelat
Department of Geology
Stanford University
Stanford, California 94305
(415) 497-4147

Dr. R.J.P. Lyon
Department of Applied Earth Sciences
Stanford University
Stanford, California 94305
(415) 497-3262

A PROBABILISTIC MODEL TO DISCRMINATE MULTISPECTRAL DATA MIXTURES

by
A.E. Prelat & R.J.P. Lyon

The discriminant scores are arranged in a frequency table, grouping adjacent observations into class intervals. Outcome frequencies based on the frequency tables are obtained for each group.

In turn, the "score probability function" attached to different groups are defined. The score probability functions are used to differentiate between those observations which are mixtures of radiation from those which contain only one radiation. In this way, those areas under the score probability function which represent variability of spectral signatures are effectively defined. The number of observations in each group has to be at least equal to the number of variables measured multiplied by a factor of five. This is necessary in order that an adequate statistical base is provided to compute the "score probability function".

The objective of the model is to estimate the different score probability functions. Once the probability functions are available, they can be utilized to locate ground resolution elements containing different objects. The results of the procedure are used as input to clustering algorithms and numerical classification techniques for computing more accurate boundaries and separability among different groups.

TABLE 01. ERTS IMAGES ACQUIRED OVER STANFORD UNIVERSITY TEST AREA

(T)

| OBSERVATION ID | FIELD DATA | MICROFILM ROLL NO. | DATE ACQUIRED | CLOUD COVER | ORBIT NUMBER | PRINCIPAL POINT (C) OF IMAGE LAT. | LONG. | SUN AZIM | SUN ELEV | (R=REQUESTED) PRODUCTS RECD. AT STANFORD M | S | B7 | P | M9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● 1. 1003-18175 | - | 10001/0126/7 | 07/26/72 | 10 | 42 | 3805N | 12146W | 118.7 | 58.7 | 4 | 4 | - | 2 | R |
| 2. 1021-18172 | - | 10001/1226 | 08/13/72 | 0 | 293 | 3724N | 12145W | 124.5 | 55.8 | R | 8 | R | R | - |
| 3. 1039-18172 | - | 10002/0074 | 08/31/72 | 0 | 544 | 3725N | 12150W | 132.5 | 51.9 | 4 | 2 | R | R | - |
| 4. 1057-18172 | - | 10002/0598 | 09/18/72 | 20 | 795 | 3721N | 12149W | 140.2 | 47.1 | R | R | R | R | - |
| ● 5. 1075-18173 | - | 10004/0236 | 10/06/72 | 0 | 1046 | 3729N | 12144W | 146.8 | 41.6 | 4 | 8 | R | 2 | 4 |
| 6. 1093-NO FRAMES TAKEN | | | 10/21/72 | - | 1297 | | | 152. | 35. | - | - | - | - | - |
| 7. 1111-18181 | - | 10004/1570 | 11/11/72 | 60 | 1548 | 3715N | 12153W | 153.9 | 30.9 | 4 | 8 | - | 2 | - |
| 8. 1129-18181 | - | 10005/0498 | 11/29/72 | 20 | 1799 | 3725N | 12150W | 154.6 | 26.7 | 4 | 8 | - | 2 | - |
| 9. 1147-18181 | - | 10006/0333 | 12/17/72 | 90 | 2050 | 3718N | 12151W | 153.4 | 24.5 | - | - | - | - | - |
| 10. 1165-18175 | - | 10006/0898 | 01/04/73 | 10 | 2301 | 3724N | 12146W | 151.1 | 24.2 | 4 | 8 | - | 2 | R |
| ● 11. 1183-18175 | - | 10007/0170 | 01/22/73 | 20 | 2552 | 3732N | 12146W | 148.2 | 26.3 | 4 | 8 | R | 2 | 4 |
| 12. 1201-18181 | - | 10007/0782 | 02/09/73 | 80 | 2803 | 3725N | 12151W | 144.9 | 30.5 | - | - | - | - | - |
| 13. 1219-18182 | - | 10008/0440 | 02/27/73 | 100 | 3054 | 3726N | 12156W | 141.6 | 36.3 | - | - | - | - | - |
| 14. 1237-18183 | - | 10009/0470 | 03/17/73 | 40 | 3305 | 3727N | 12200W | 138.1 | 42.8 | 4 | 8 | - | 2 | - |
| ● 15. 1255-18183 | - | 10009/1329 | 04/04/73 | 0 | 3556 | 3730N | 12201W | 134.2 | 49.4 | 8 | 4 | - | 1 | 4 |
| ● 16. 1273-18183 | - | 10010/0613 | 04/22/73 | 0 | 3807 | 3726N | 12201W | 129.4 | 55.2 | 4 | 8 | - | 2 | 4 |
| ● 17. 1291-18182 | F | 10010/1539 | 05/10/73 | 0 | 4058 | 3731N | 12201W | 123.3 | 59.6 | 8 | 4 | - | 1 | 4 |
| 18. 1309-18181 | F | | 05/28/73 | | | 3735N | 12201W | 117.0 | 61.0 | 8 | 4 | - | 2 | R |
| 19. 1327-18180 | F | | 06/15/73 | | | 3730N | 12153W | 113.0 | 62.0 | 4 | 8 | - | 2 | R |
| ● 20. 1345-18174 | F | 10012/1181 | 07/03/73 | 30 | 4811 | 3725N | 12202W | 112.5 | 61.6 | 4 | 8 | - | 2 | R |
| 21. 1363-18173 | F | 10013/0135 | 07/21/73 | 30 | 5062 | 3725N | 12202W | 115.0 | 59.0 | 4 | 8 | - | 2 | R |
| 22. 1381-18172 | R | 10013/1276 | 08/08/73 | 50 | 5313 | 3721N | 12203W | 122.0 | 56.0 | - | - | - | - | - |
| 23. 1399-18170 | R | | 08/26/73 | | 5564 | 3726N | 12201W | 129.0 | 52.0 | - | 8 | - | 2 | 4 |
| 24. 1417-18164 | - | | 09/13/73 | | | 3725N | 12158W | 137.9 | 48.0 | 4 | 8 | - | 2 | - |
| 25. 1435 | - | | 10/01/73 | | | | | | | - | - | - | - | - |
| 26. 1453 | F | | 10/19/73 | | | | | | | - | - | - | - | - |
| 27. 1471- | - | | 11/06/73 | | | | | | | | | | | |
| ● 28. 1489-18152 | F | 10018/0397 | 11/24/73 | 30 | 6819 | 3727N | 12151W | 153.9 | 27.5 | | | | | |
| 29. 1507- | - | | 12/12/73 Rain | | | | | | | | | | | |
| ● 30. 1525-18145 | F | 10019/0697 | 12/30/73 Clear | 10 | 7321 | 3723N | 12155W | 151.0 | 23.0 | | | | | |
| 31. 1543-18141 | | 10020/0371 | 1/17/74 Cloud | 100 | 7572 | 3732N | 12150W | 148.0 | 25.0 | | | | | |
| 32. 1561-18133 | | | 2/04/74 Foggy | | | 3729N | 12145W | 144.0 | 28.0 | | | | | |
| 33. 1579-18131 | F | | 2/22/74 OK (SU) | | | 3733N | 12147W | 141.0 | 33.0 | | | | | |
| | | | 3/12/74 Cloudy | | | | | | | | | | | |

TABLE 02. TAPES IN STANFORD RSL DATA FILE

STANFORD

| | |
|---|---|
| 1003-18175 | 07/26/72 |
| (+1003-18175 RBV) | |
| 1075-18173 | 10/06/72 |
| 1183-18175 | 01/22/73 |
| 1255-18183 | 04/04/73 |
| 1273-18183 | 04/22/73 |
| 1291-18182 | 05/10/73 |
| 1345-18174 | 07/03/73 |
| 1489-18152 | 11/24/73 |

MONO LAKE

| | |
|---|---|
| 1055-18055 | 9/16/72 |
| 1091-18062 | 10/22/72 |
| 1063-18063 | 1/02/73 |
| 1235-18070 | 3/15/73 |
| 1307-18064 | 5/26/73 |
| 1397-18053 | 8/24/73 |
| 1361-18060 | 7/19/73 |

WALKER LAKE

| | |
|---|---|
| 1055-18053 | 09/16/72 |
| 1091-18055 | 10/22/72 |
| 1163-18060 | 01/02/73 |
| 1235-18064 | 03/15/73 |
| 1289-18063 | 05/08/73 |
| 1307-18062 | 05/26/73 |
| 1361-18054 | 07/19/73 |
| 1397-18051 | 08/24/73 |
| 1415-18045 | 09/11/73 |
| 1505-18032 | 12/10/73 |

SAN LUIS

| | |
|---|---|
| 1074-18114 | 10/05/72 |
| 1254-18125 | 4/03/73 |

S. SALINAS

| | |
|---|---|
| 1290-18130 | 5/09/73 |

BERRYESSA

| | |
|---|---|
| 1075-18170 | 10/06/72 |

ERTS MAILING LIST

| | Copies |
|---|---|
| Attn: J.D. Medwin<br>ERTS Contracting Officer, Code 245<br>Goddard Space Flight Center<br>Greenbelt, Maryland 20771 | 1 |
| Attn: Edward Crump<br>ERTS Technical Officer, Code 430<br>Goddard Space Flight Center<br>Greenbelt, Maryland 20771 | 1 |
| ERTS Project Scientist (UN 142), Code 650<br>Goddard Space Flight Center<br>Greenbelt, Maryland 20771 | 1 |
| ERTS Scientific Monitor (UN 142), Code 650<br>Goddard Space Flight Center<br>Greenbelt, Maryland 20771 | 1 |
| NASA Scientific and Technical Information Facility<br>Attention: (EARTH RESOURCES) -(UN142)<br>P.O. BOX 33<br>College Park, Maryland 20740 | 1 |
| Ernest I. Rich<br>Deans Office<br>Stanford University<br>Stanford, California 94305 | 1 |
| Jack Quade<br>MacKay School of Mines<br>University of Nevada<br>Reno, Nevada 85907 | 1 |
| John J. Ducey, ONR Representative<br>Rm. 167<br>Durand Bldg.<br>Stanford University<br>Stanford, California 94305 | 1 |
| Graduate Student Assistants: | |
| Gary Ballew | 1 |
| Bruce Salem | 1 |
| Aderbal C. Correa | 1 |
| Robert W. Campbell<br>Applied Earth Sciences<br>Stanford University<br>Stanford, California 94305 | 1 |
| Mr.J.H. Boeckel, Code 430<br>NASA/Goddard Space Flight Center<br>Greenbelt, Maryland 20771 | 2 |

| | |
|---|---|
| Lyon-lab | 1 |
| Honey | 1 |
| Lyon-files | 1 |
| Levine | 1 |

4

18