

SECTION 123

PATTERN RECOGNITION OF NATIVE PLANT COMMUNITIES--

MANITOU COLORADO TEST SITE

by

Richard S. Driscoll^{1/}--Rocky Mountain Forest and Range Experiment Station
U.S.D.A. Forest Service, Fort Collins, ColoradoABSTRACT

Optimum channel selection among 12 channels of multispectral scanner imagery identified six as providing the best information about 11 vegetation classes and two nonvegetation classes at the Manitou Experimental Forest (NASA Test Site 242). Intensive preprocessing of the scanner signals was required to eliminate a serious scan angle effect. Final processing of the normalized data provided acceptable recognition results of generalized plant community types. Serious errors occurred with attempts to classify specific community types within upland grassland areas. The consideration of the convex mixtures concept--effects of amounts of live plant cover, exposed soil, and plant litter cover on apparent scene radiances--significantly improved the classification of some of the grassland classes. The data processed was obtained as part of Mission 19 at 1000 hours on July 29, 1970, by the University of Michigan multispectral scanner flown at 915 meters (3,000 feet) above mean terrain elevation.

INTRODUCTION

Multispectral scanner imagery coupled with automation data processing may well be the future technique for classifying non-agricultural vegetation or especially monitoring changes in this vegetation. The scanner "looks" at a piece of landscape, the area depending on the resolution of the system, and records the

1/ Project Leader and Principal Plant Ecologist in charge of Remote Sensing Research.

data on magnetic tape for analyses. During the analysis process, the computer makes yes-no decisions to align the spectral radiances recorded for the resolution elements to predefined categories. Thu6::: and variances inherent to aerial photographs and photo-ix reter , including differences among interpreters and variances in in , may be circumvented.

For monitoring vegetation, plant species or groupings should maintain similar radiances provided physical and physiological factors are relatively constant (Knipling 1970). Physiological changes in the plant species will cause radiance changes of the species (Gausman, Allen and Cardenas 1969; Weber and Olson 1967). Structural changes in the plant community, which varies the geometry of the grouping, also changes the radiance of the community (Allen and Richardson 1968). For a given area of landscape therefore, changes in the scene caused by a disturbance of the vegetation should result in different scene radiances and these data would be used to assess change in vegetation over time. Frequently, these changes are subtle and important to land management decisions, but may not be read from aerial photographs or observed on the ground.

The problem of the effects of variable atmospheric paths on spectral signals obtained by remote sensors in the optical region of the spectrum and how to cope with them have been documented (Horvath, Braithwaite and Polcyn 1970). These data could be interfaced with raw multispectral Scanner data so that processing the latter deals with absolute and not relative values.

Before we can use these inferences, dependable evaluations must be made on the level of integrity with which multispectral scanning systems discriminate dissimilar plant communities. The research reported here are some of the most significant results of an experiment to test the hypothesis that multispectral imagery will identify specific plant communities determined by ecological analyses.

DATA ACQUISITION

THE STUDY AREA

The study area for this research is located approximately 25 airline miles northwest of Colorado Springs, Colorado, and is within the Manitou Experimental Forest. The Forest is a 16,000-acre area within the ponderosa pine/bunchgrass type where

scientists at the Rocky Mountain Forest and Range Experiment Station are involved in multidisciplinary wildland research. The vegetation is typical of much of the lower Montane Zone along the eastern slope of the Rockies; open to dense stands of ponderosa pine (Pinus ponderosa, Laws.) are interspersed with meadows and grassland parks.

The specific test site included an area approximately 7.28 square kilometers (2.81 square miles) with a mean ground datum of 2,350 meters (7,700 feet) (Fig. 1). Within the area were three general kinds of vegetation: (1) Ponderosa pine forest, (2) upland grasslands which included native and seeded grass stands, and (3) hydrophyllic communities.

GROUND DATA

Detailed mapping of plant communities was completed immediately prior to the scanner mission. This was based on current aspection of the plant communities and the relative composition of the communities regarding similarity of plant species components. Medium scale (1:8000) color infrared mapping photographs flown approximately 6 weeks prior to the scanner mission were used in conjunction with detailed ground search. At this photo scale, each mapping unit represented a specific community type.

Plant species abundance was determined for each community type within 2 weeks of the mission. Abundance was based on a 5-point rating scheme ordered from very abundant to rare (Oosting 1956). At the same time, percent plant foliar cover, percent bare soil surface, and percent plant litter cover of the soil surface (the sum of the three equalling 100 percent viewed vertically) was estimated. This was done by sampling with 9- X 9-foot sample plots (Fig. 2) located by restricted random fashion throughout each community type. This plot size related to the equivalent resolution element to be "seen" by the scanner at the planned flight altitude to represent scene radiance of the element . component mixtures.

Ground cover sampling in the forest communities was restricted to land surface visible from above without interference from the tree canopy. All ground data were obtained during late July and early August, the time when most plant species were actively growing. The mapping units, which corresponded to recognition categories of the scanner data, and a brief description of their characteristics are listed in Table 1. Also included are two nonvegetation categories

which were inclusions in the mapping units but which could possibly be discriminated in the multispectral analyses.

AERIAL DATA

All multispectral data collection was performed by the University of Michigan's multispectral scanning system. This system consisted of two *double-ended* optical mechanical scanners. Imagery included data records in 12 discrete spectrometer channels in the visible and near infrared regions of the spectrum (Table 2).

The scanner was flown at approximately 305 meters (1,000 feet) and 915 meters (3,000 feet) above ground datum on July 28 and 29, 1970. Four time periods were selected: 0830, 1200 and 1600 hours local sun time on July 28, and 1000 hours on July 29. Only the data obtained at the last time period at the 915 meter (3,000 feet) altitude was used for subsequent analyses. This time and altitude were selected for four primary reasons: (1) weather at the time of the overflight was clear with visibility in excess of 160 kilometers (100 miles), (2) video display of preprocessed data from channel 10 (.604 - .700 micrometers) indicated this information might provide the best opportunity for recognition processing, (3) environmental conditions during the 0830 flight, heavy predawn rain which left the target surfaces wet, and the 1200 and 1600 flights, during which cloud shadows were in the imaged areas, would produce false radiance signals, and (4) the "halo" effect in the imagery caused by the airplane shadow in the 1200 and 1600 flights which would cause data processing problems for removing the effect.

RESULTS AND DISCUSSION

PREPROCESSING

Conventional techniques for analyzing multispectral data for classification and mapping terrain features assume uniform apparent radiance in each class regardless of position in the scene. Previous studies have shown that this is not true because such things as atmospheric haze, variations in topography, scanner look angle, and variations in geometry of the scene to be classified all affect the apparent scene radiance (Solomonson and Mariatt 1971; Smedes et al. 1970) The scan angle effect should be expected since the look angle of the scanner mission system for a discrete bit of data

about a specified target varies through the arc of the scanner field of view. If this were the only factor affecting the apparent scene radiance, the data could be simply normalized since the strongest point of illumination would be at the nadir, tailing off to each end of the *scan line* to produce a normal distribution curve.

The previously described condition would seldom occur since the probability of scanner look angle being simultaneously perpendicular to the illumination source (the sun) and the target scene (a plant community of constant geometry) is nil. Therefore, the scan angle effect in relation to apparent scene radiance is influenced by atmospheric effects and bidirectional effects. Atmospheric effects were deemed negligible for this study since visibility was in excess of 160 kilometers and flight altitude was only 915 meters (3,000 feet) above the terrain.

Bidirectional reflectance effects are common and the data from this study were analyzed assuming such effects were present. These effects are the combined result of variation in sun angle, the bidirectional reflectance properties of the target scene caused by varying surface geometry, and scanner look angle. The fact that bidirectional/scan angle effects were present in the data was identified by a technique developed by Kreigler (1971). Figure 3 illustrates this effect from data plotted from channel 5 for three recognition categories. Each point on the curve represents the statistical mean value of a training set consisting of a block of approximately 100 resolution data cells.

This scan angle effect was eliminated from the data using transformation processes developed by Kreigler, et al. (1969). Figure 5 illustrates the results of normalizing the channel 5 data for the three example Categories.

RECOGNITION PROCESSING

Signature Selection.— Signature selection to represent recognition categories (Table 1) was relatively simple using the preprocessed data provided the selected training sets were relatively homogeneous. This was true for recognition categories 14, 16, and 17 (see Table 1 for explanation of categories). Since the location and spacial distribution of these units was quite limited and homogeneous, the spectral signature to represent each category was a single training set obtained directly from the data. More than one training set was used to determine the statistical spectral

signature for categories 2, 3, 4, 12, and 15 (see Table 1 for explanation of categories). However, the computed signature for each category was so similar to any one of the training sets for each category, it was arbitrary as to which one was used. Consequently, only one training area was chosen and its signature extracted directly from the data for all categories except category 2. Since this category was widely distributed throughout the area, information from two training sets were extracted from the data and used to determine a new recognition signature.

The process to obtain a representative signature for the upland grassland categories, units 5 through 9 (see Table 1 for explanation of categories), was much more tedious. Even after the data had been preprocessed, there was wide variation within and among category training set signature values. This was related to two factors:

(1) the inability to precisely locate a representative position for a specific training set area for a specific category due to ectonal variation among the categories, and (2) the varying amounts and kind of herbage cover, plant litter, and bare soil represented in the equivalent resolution element. It was difficult to discern boundaries among the grassland units in the gray map generated from preprocessed data in relation to ground control from the mapping photographs. In addition, there was considerable variation in the amounts of ground cover characteristics in ground sample plots, corresponding to the equivalent resolution element, within the originally mapped units although the data were adequate to describe the units on the ground.

Final signature selection for these upland grassland categories was based on a method which used the average probabilities of misclassification as criteria to determine similarity among units. The routine simultaneously considered statistical parameters (mean, variance, and covariance) for all channels of data to compute pairwise probabilities of misclassification. It involved testing distributions by a likelihood ratio test in which each distribution was tested pairwise with all other distributions. The recognized signature for that training set to be used for the spectral signature was that one with a probability density function greater than the probability density functions for all other training sets within the category. The relative location of training sets selected for final recognition processing are shown in Figure 5.

Optimum Channel Selection.— It was expected that not all channels of spectrometer data would be needed to classify the plant communities defined. Therefore, channel optimization was performed to determine the best set of channels to be used. This was based on the average

probability of misclassification in which a number of channels were chosen in an ordered selection scheme.

The process assumed that the spectral signatures represented a Gaussian distribution of random variables, an assumption which thus far has been good for similar data (Heller, et al. 1970). A pairwise classification scheme was used in which if there were M categories (recognition units), then there were MCM-1 probabilities of misclassification. For example, the probability of misclassification of each pair of categories was the probability of **misclassifying** category M2 as category M1. The end result, with weighted entries, provided an average pairwise probability of misclassification and identified the best channel for classifying the vegetation classes *previously* defined. This ordered selection next combined the first channel with each of the remaining n-1 channels and picked the best combination of two channels. The results of the ordered selection scheme to determine the spectral channels to use for final recognition processing is shown in Table 3.

Six channels were selected for final recognition processing. The channel selection scheme was stopped at this number because addition of data from other channels indicated minimum improvement in recognition accuracy. For example, adding channel 6 data indicated **an** increase in classification accuracy by only approximately 11 percent. This was deemed acceptable for this problem.

Final .Processing.- The decision rule to classify each data point using a likelihood ratio test was used for final processing. The test simultaneously compared the information content of each data point for category recognition and assigned the data point to a particular category when the following n-1 ratio tests were simultaneously satisfied:

$$f(M_i) > 1$$

$$f(M_i)$$

where:

$f(M_i)$ is the multivariate Gaussian probability density function for category M and

$f(M)$ is the multivariate **Gaussian** probability density function for category

The recognition processing produced a digital recognition map (Fig. 6). This area represents a unit of terrain approximately 7.28 square kilometers (2.81 square miles) in area. The color/symbol coding which was used is listed in Table 4. The exact coding is not recognizable in the photography due to scale reduction from the original digital color map. However, if the reader makes a correspondence between color shades and the location of training areas (Fig. 5), this will aid in interpreting the map.

Generalized plant communities were acceptably isolated. These included the forested areas (green), areas with hydrophyllic vegetation (black), and areas of upland herbaceous vegetation (red, blue, and purple). This was expected, however, since others have reported similar results (Smedee, et al. 1971; Heller, et al. 1970). Ground conditions of these units were very dissimilar which produced high contrast in apparent radiance signals.

However, problems existed for classifying the upland herbaceous plant communities (categories 5-9). Even though intense preprocessing of the data was done, a large amount of Misclassification Occurred. For example, the bluegrass seeding in the northwest portion of the area was synonymously identified with category 8 (abandoned fields with native vegetation) and with parts of Category 7 (native range). Likewise, the native range and abandoned fields categories were mixed, severely in some areas.

Due to severe mixing of the bluegrass seedings with other upland grassland categories in the northeast portion of the area, additional processing was done to improve recognition of this unit. The unit was separated into two additional categories. One, category 6-1, represented an old seeding contaminated with yellow sweetclover (Kelilotus officianalis (L.) Lam.) and the other, category 6-2, represented an old seeding contaminated with numerous native herbaceous species. Reprocessing by extracting spectral signatures for these categories from the data resulted in improved recognition of them. This verifies that community species composition as well as amounts of the three ground cover components are important characteristics to consider for analyses and recognition processing of multispectral scanner data for plant community classification.

It was not possible to follow this procedure with the other upland grassland communities where severe mixing occurred in the recognition processing. The ground control plan did not allow future location of specific areas either on the ground or in various kinds of imagery the exact position where the control data was obtained. Even with the severe mixing of these categories, it does not mean that multispectral scanner data cannot be used for identifying

and classifying specific plant community types within a generalized herbaceous system. Rather, it means that more discrete selection must be done of training sample areas to represent the community types relevant to specific ground control. For example, the native range (category 7) areas identified by the recognition processing were correctly identified, but not all native range areas were classified as native range. Those areas correctly identified consisted of plant communities with a composition primarily of vigorous stands of Arizona fescue and mountain muhly and little bare soil surface showing through the community canopy. Other areas not classified as native range but mixed with category 8 (abandoned fields) had considerably less herbaceous cover, less litter cover on the ground and more exposed bare soil surface. Similar conditions existed between the relatively pure big bluegrass seeding in the northwest corner of the area and the more sparsely vegetated native range areas.

It may be that the relative radiance of herbaceous vegetation is not sufficiently contrasting to provide discrete separation of community types within the general system. If this is the case, more consideration must be given to the combined relative amounts of vegetation cover, litter cover, and bare soil surface which may provide the information needed for acceptable multispectral recognition processing. Controlled experiments need to be initiated whereby the effective ground resolution elements can be isolated in the multispectral data for absolute information about the three ground cover characteristics. Also *more information* is needed about the ground scene radiance of various combinations of ground cover characteristics to determine the effects of these characteristics on effective scene radiance.

The normalized data from the six channels identified for digital processing was also processed through the Michigan Spectral Processing and Recognition Computer. The SPARC system accepts analog data and presents the results as a color map in analog form. Since the SPARC system can accept data about only eight recognition categories and we processed 14 recognition categories using six data channels, two separate operations were performed on the SPARC. The results of this processing are illustrated in Figure 7. The color codes are identified in Table 5.

There was some confusion interpreting the SPARC map as compared to the digital map. For example, the SPARC seeded crested wheatgrass is identified as yellow and the other seeded grassland (category 9) is identified as dark. green. Interpreting the digital map point by point shows some mixing of these two categories in the crested

wheatgrass area, but the major portion of the data points, in fact, represent the crested wheatgrase area. The SPARC map identifies this area primarily as other seeded grassland. The reason for this crossover is not known now, but available evidence indicates interpretation of the digital map may provide more discrete data about the individual plant community types. A similar relationship existed in SPARC classification of category 6-2 and category 7. Many small areas in the south of the SPARC map were classified as category 6-2 when they should have been category 7. The digital map again identified a mixing of these two categories in the area but the major portion of the data points identified as native range (category 7).

The utility of developing a multispectral processing technique for native vegetation should be obvious. An important concern about management of native vegetation relates to change in vegetation over time and determining what caused the change. Since the multispectral scanner and peripheral equipment records and stores bits of information about a small piece of landscape, depending on the resolution capabilities of the scanner system, any change in that piece of landscape theoretically would be identified by a change in apparent scene radiance in subsequent scanner data. This assumes that discrete limits of the information stored, that is the mean relative radiance levels with variance that are discrete for specific plant communities, can be identified. The percentage of data points representing a kind of plant community can be computed (Table 6) and any change in these relative values can be determined using sequential imagery. This would then represent a change in area comprising a specific community type. Theoretically, this technique would provide more accurate information than photointerpretation only since interpreter error would be minimized. The concept needs substantial additional research to identify the minimum level of integrity the recognition processing can classify plant communities and then tested over time to determine the repeatability of the technique.

SUMMARY AND CONCLUSIONS

1. Serious scan-angle effects were identified in multispectral scanner imagery taken at 1000 hours on July 29, 1970, at the Manitou Colorado Test Site, by the University of Michigan's multispectral scanner system. The effect was believed to have been caused by bidirectional reflectance which is the combined effect of variation in sun angle, scanner look angle, and the reflectance properties of the specific scene. Atmospheric attenuation was believed not to be a serious influence on the scanner data since visibility was in excess of 160 kilometers (100 miles) at the time of the data mission and the

flight altitude was only 915 meters (3,000 feet) above the terrain.

2. Preprocessing was performed on the data to remove the scan angle effect using techniques developed by University of Michigan IROL research engineers. Theoretically, with the scan angle effect removed, only one computer training sample per recognition category is required for further recognition processing.

Channel optimization for recognizing 11 vegetation and two nonvegetation categories representing mapping units was based on the average probability of misclassification in which the six best channels were chosen on an ordered selection scheme. These were:

Channel No.	Spectral Band (um)
10	0.604-0.700
12	0.725-0.920
5	0.478-0.508
9	0.566-0.638
7	0.514-0.558
6	0.492-0.536

4. The recognition processing results provided acceptable discrimination of generalized plant communities. These included: (1) ponderosa pine forested areas, (2) upland herbaceous vegetation, (3) hydrophyllic herbaceous vegetation. Two nonvegetation categories, asphalt roads and bare soil, were acceptably classified.

5. Serious problems exist in classification of upland herbaceous community systems. Seeded crested wheatgrass was satisfactorily classified. However, seeded big bluegrass was mixed seriously with native range which was in turn confused in the computer processing for abandoned fields with native vegetation significantly different from native range.

6. The classification of upland herbaceous plant communities could be surmounted by considering more carefully the convex mixtures concept. The concept relates to the influence of relative Amounts of plant foliar cover, bare soil, and dead plant litter on the soil surface on the apparent radiance levels of the mixtures. . In addition, species components of the plant communities, especially the amounts

of grass versus other herbaceous vegetation, need to be more seriously considered in relation to multispectral recognition processing for plant community classification.

ACKNOWLEDGEMENTS

Special appreciation is extended to the Willow Run Laboratories, University of Michigan, for collecting and processing the multispectral scanner imagery. Specifically, Research Engineers M. M. Spencer, F. J. Kneigler,, F. J. Thompson, and F. O. Hassell provided substantial contributions to this research.