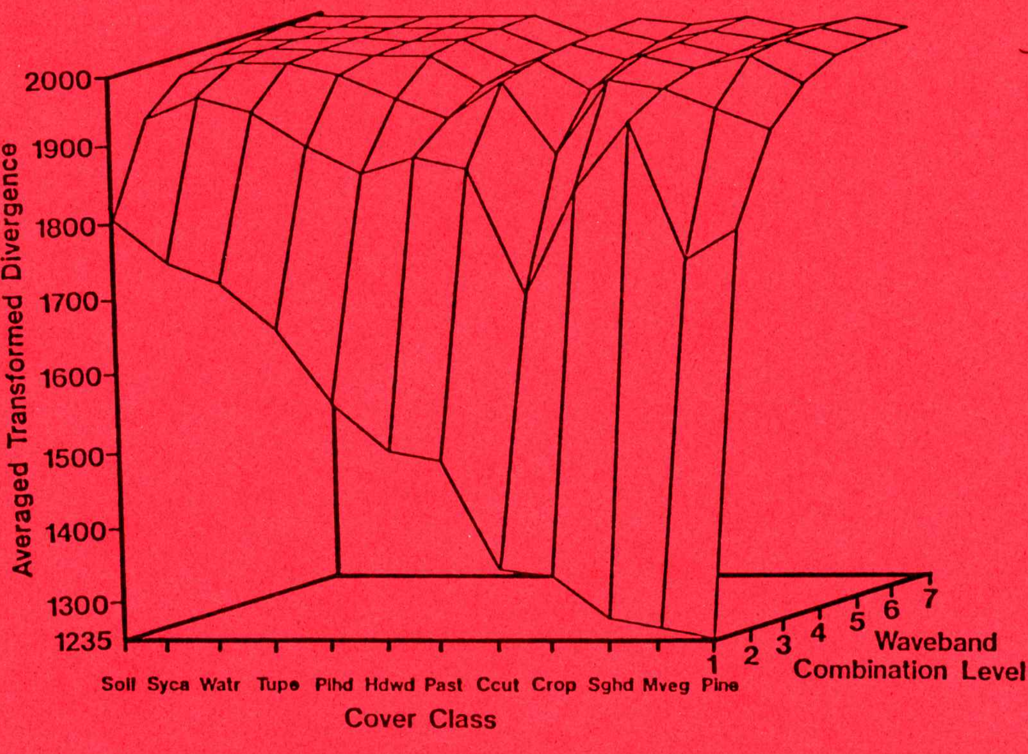
LARS Technical Report 041581

Waveband Evaluation of Proposed Thematic Mapper in Forest Cover Classification

R. S. Latty and R. M. Hoffer



Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA
1981

WAVEBAND EVALUATION OF PROPOSED THEMATIC MAPPER IN FOREST COVER CLASSIFICATION

Richard S. Latty and Roger M. Hoffer
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

This study involved the evaluation of the characteristics of multispectral scanner data relative to forest cover type mapping, using NASA's NS-001 multispectral scanner to simulate the proposed Thematic Mapper (TM). The objectives were to determine: (1) the optimum number of wavebands to utilize in computer classifications of TM data; (2) which channel combinations provide the highest expected classification accuracy; and (3) the relative merit of each channel in the context of the cover classes examined. Transformed divergence was used as a measure of statistical distance between spectral class densities associated with each of twelve cover classes. The maximum overall mean pair-wise transformed divergence was used as the basis for evaluating all possible waveband combinations available for use in computer-assisted forest cover classifications.

INTRODUCTION

Early work in leaf spectra analysis (Billings and Morris, 1951; Gates and Tantraporn, 1952; Gates, et al., 1965; Gausman, et al., 1969; Knipling, 1970; Wooley, 1971; Gausman, 1977) provided much of the initial understanding of the variations in the amount of radiant energy returned from vegetated surfaces. Colwell (1974) identified the value of hemispheric leaf reflectance as only one of several important parameters responsible for these variations, and cautioned against making inferences about scene reflectance from leaf spectra information alone. Plant canopy modeling efforts (Idso and De Wit, 1970; Nilson, 1971; Oliver and Smith, 1972; Suits, 1972; Colwell, 1973) have identified many of the parameters which account for variations in the amount of radiant energy returned from the scene. The selection of waveband combinations which will provide accurate classification of the various earth surface features requires an understanding of the reflective characteristics of those features relative to the various wavebands available. Properties of the data consequential to classification accuracy are not dependent solely on earth surface, atmospheric, and illumination conditions. They are also very dependent on the parameters of the sensor system to be employed (Silva, 1978). Therefore, the need exists to investigate these reflective properties employing data more closely simulating the data which will ultimately be employed for such classifications.

With parametric classifiers, the resulting classification accuracy is dependent on (1) the degree to which the training classes (i.e., spectral classes) represent the spectral variability of their respective cover classes, and (2) the level of statistical "separability" among the training classes (Swain, 1978). The first condition is difficult if not impossible to assess without conducting the actual classification - the expense of which precludes evaluating many different waveband combinations. One can justifiably assume that the first condition is satisfied if the points providing the data for establishing the training classes are randomly generated, and are "sufficient" in number for each class relative to the number of wavebands employed. The number of samples statistically sufficient for the development of training classes increases exponentially with an increase in the number of channels employed in classification (Duda and Hart; 1973). Duda and Hart (1973) pointed out that, "beyond a certain point, the inclusion of additional features leads to worse rather than better performance." They provide an excellent review of the problem. This problem has also been examined by Allais (1966), Dynkin (1961), Fukunaga and Kessell (1971), Kanal and Chandrasekaran (1971) and others. The level of statistical "separability" can be computed from the mean vectors and covariance matrices associated with each of the training classes employing one of several statistical distance measures (Kailath, 1967; Swain, Robertson and Wacker, 1971; Wacker and Landgrebe, 1972; King and Swain, 1973).

METHODS AND ANALYSIS

Data Acquisition

The data were obtained on May 2, 1979 from the NASA NC-130 aircraft flying at an altitude of 20,000 ft. (MGD) over an area immediately south of Camden, South Carolina. The multispectral scanner (MSS) data were obtained by the NASA NS-001 multispectral scanner. (Table 1 shows the NS-001 scanner specifications as compared to the Thematic Mapper). Color and color infrared photographs (1:40,000 scale transparencies) were obtained at the same time. Cloud coverage was minimal and atmospheric conditions were considered excellent.

Data Handling and Preprocessing

The across track change in scale of the imagery was adequately reduced by employing a geometric model which describes the ground resolution element dimensions as a function of aircraft altitude, IFOV (instantaneous field-of-view) of the scanner, and change in scan angle corresponding to the analog signal integration interval.

A study of the data quality revealed an apparent correlation between scan angle and response level (different for each channel). The relationships appeared to be sufficiently high to obscure sources of variation otherwise correlated with differences between cover classes. Therefore, an empirically derived function was generated which described the variation in response level by column (corresponding with scan angle). Data were employed from areas where no apparent stratification of cover class by column was present.* The shape of these functions were evaluated against both empirical (Anuta and Strahorn, 1973; Landgrebe, Beihl, and Simmons, 1977) and theoretical work (Kondratyev, 1969; Jurica and Murray, 1973) prior to actual response level adjustment. The final data product was considered appropriate for the analysis.

Table 1. Comparison of the NASA NS-001 multispectral scanner and the proposed Thematic Mapper (TM).

NS-001 Multispectral Scanner[1]				Proposed Thematic Mapper[2]			
Channel	Bandwidth (μm)	Low Level Input ($W \cdot CM^{-2} \cdot SR^{-1}$)	$NE\Delta\rho$	Channel	Bandwidth (μm)	Low Level Input ($W \cdot CM^{-2} \cdot SR^{-1}$)	$NE\Delta\rho$
1	0.45-0.52	8.7×10^{-6}	0.5%	1	0.45-0.52	2.8×10^{-4}	0.8%
2	0.52-0.60	6.8×10^{-6}	0.5%	2	0.52-0.60	2.4×10^{-4}	0.5%
3	0.63-0.69	5.0×10^{-6}	0.5%	3	0.63-0.69	1.3×10^{-4}	0.5%
4	0.76-0.90	4.4×10^{-6}	0.5%	4	0.76-0.90	1.6×10^{-4}	0.5%
5	1.00-1.30	6.0×10^{-6}	1.0%				
6	1.55-1.75	6.2×10^{-6}	1.0%	5	1.55-1.75	8.0×10^{-5}	1.0%
7[3]	2.08-2.35	4.7×10^{-5}	2.0%	6	2.08-2.35	5.0×10^{-5}	2.4%
8	10.4-12.5	NA	$NE\Delta T = 0.25^{\circ}K$	7	10.4-12.5	$300^{\circ}K$	$NE\Delta T = 0.5^{\circ}K$

(1) Data was obtained from the "Operations Manual, NS-001 Multispectral Scanner," NASA; JSC-12715, April 1977.

(2) Data was obtained from Salomonson, 1978.

(3) Channel 7 (2.08-2.35 μm) was not operational at the time of the mission; all subsequent references to "channel 7" refer to the 10.4-12.5 μm waveband.

Development of Spectral Classes

A COMTAL Vision One/20, displaying a composite of channels 3, 4, and 5, in conjunction with the aerial photography, was employed to ascribe cover class labels and ground condition descriptions to line-column coordinates in the imagery in a supervised fashion. This approach was considered more appropriate than the unsupervised clustering approach, since cover classes could be defined more nearly independent of their spectral characteristics in the wavebands to be evaluated. The method used to develop training classes was of particular concern since the affect of different within-class variances for each channel by cover class on cluster class composition is not currently well understood (Bartolucci, 1978; Anuta, 1979). Once the training fields had been identified, they were grouped according to cover class. The cover class groups of training fields were then individually clustered to resolve the cover classes into a set of spectral classes. This provided training class statistics corresponding to a set of spectral classes associated with each cover class. Clustering at this stage provided a means of

*The function was generated using data obtained outside of the area from which the data for this analysis was obtained.

establishing the spectral classes on the basis of spectral variability within each cover class, but did not completely avoid the problem mentioned above. Failure to provide training statistics representing the spectral variability within each cover class was considered more deleterious to the objective of the study than clustering to obtain those classes.

Data Analysis

The mean vector and covariance matrix computed for each of the spectral classes define the individual statistical density associated with each respective spectral class. A measure of statistical distance between all pair-wise combinations of the spectral classes provides information on the "separability" of these spectral classes. This "separability" represents an *a priori* estimate of the probability of correct classification (Swain, Robertson, and Wacker, 1971) for measurements provided by each channel or channel combination. Only pairs of spectral classes belonging to different cover classes are of interest, since low separability between different spectral classes of the same cover class does not affect classification accuracy.

Transformed divergence was used to compute the separability. Divergence is defined as:

$$D = \int [p_1(x) - p_2(x)] \ln \frac{p_1(x)}{p_2(x)} dx \qquad (1)$$

where: $p_1(x)$ = statistical density of spectral class 1

$p_2(x)$ = statistical density of spectral class 2

or computationally, for the Gausian multivariate case:

$$D = \frac{1}{2} \operatorname{tr} [(\Sigma_1 - \Sigma_2)(\Sigma_1^{-1} - \Sigma_2^{-1})] + \frac{1}{2} \operatorname{tr} [(\Sigma_1^{-1} + \Sigma_2^{-1})(m_1 - m_2)(m_1 - m_2)^T] \qquad (2)$$

where: Σ is the covariance matrix and m is the mean vector associated with the respective spectral class, and

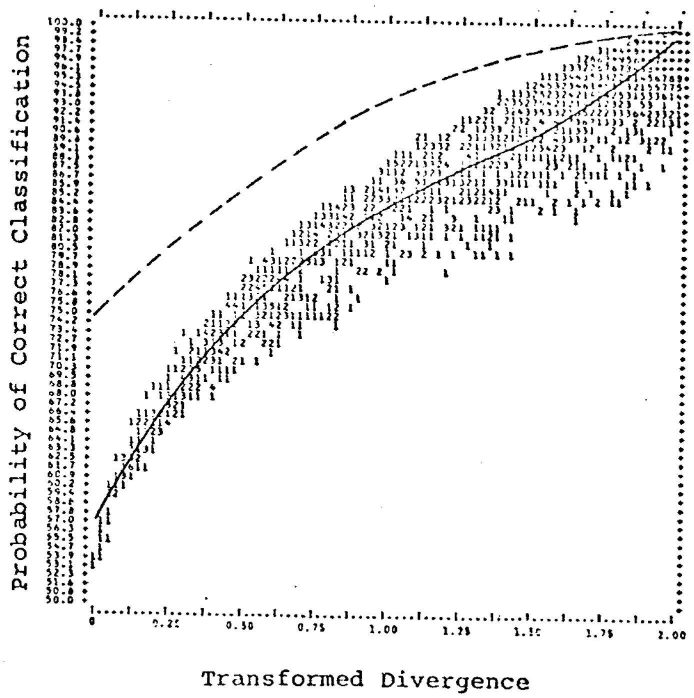
tr (trace) is the sum of the diagonal elements.

Since divergence increases without bound as the statistical distance between the two classes increases, a saturation transform is employed, resulting in a measure (i.e., transformed divergence) which corresponds more closely with percent correct classification (see Figure 1). After a certain level of statistical difference has been attained, virtually no confusion exists between the two class densities, and percent correct classification "saturates" toward 100%. The resulting transformed divergence is provided by:

$$TD = 2000 [1 - \exp(-D/8)] \qquad (3)$$

There are some disadvantages to the use of transformed divergence as a measure of statistical difference between class densities*, but because of relative computational efficiency it is used in lieu of the alternative measures.

Figure 1. Probability of correct classification regressed against transformed divergence. (Swain et al., 1971)



Transformed divergence (TD) values were computed for each pair of spectral classes representing different cover classes, for each channel and channel combination. These mean pair-wise TD-values were then sorted for each set of combinations involving the same number of channels. The seven channel combinations providing the highest mean pair-wise TD-values were obtained. Additional programs were written to generate summaries of the mean TD-values for each pair of cover classes (i.e., over all spectral classes representing the cover class pair) and each cover class

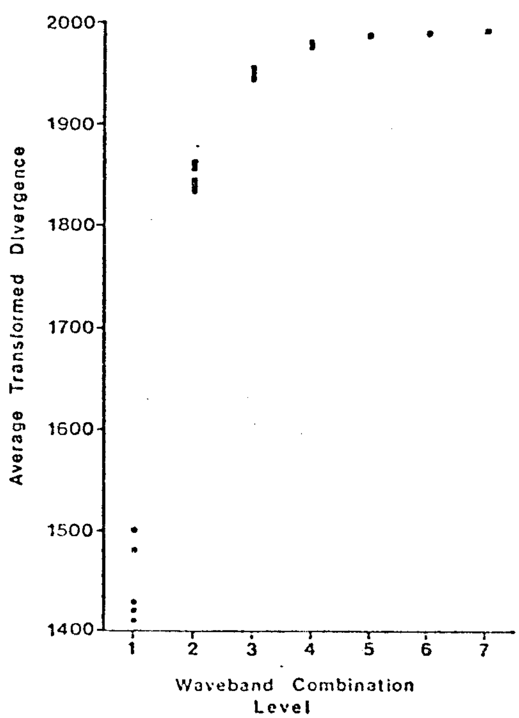
*It should be pointed out that transformed divergence is not "metric" in multivariate normal distribution functions of non-equivalent covariance matrices (Landgrebe and Wacker; 1972). That is, a pair of class densities having non-equivalent covariance matricies yet having equal mean vectors could have a transformed divergence value of zero. Also, there is no estimate for a lower confidence limit for the regression relation between transformed divergence and percent correct classification (Swain, Robertson, and Wacker; 1971).

(i.e., over all cover class pairs involving the jth cover class; j = 1,...,12) for these seven channel combinations.

RESULTS AND DISCUSSION

To define the optimum number of channels to use in a classification, the relationship between cost of misclassification and the probability of error must be determined. Otherwise there is no meaningful way to compare classification cost to classification accuracy. It can be observed from Figure 2 that the increase in transformed divergence (the correlate to probability of correct classification) drops off sharply after three channels, and very little is gained by using more than four channels. This result is similar to those obtained previously with the Michigan M-7, 12-channel scanner (Coggeshall and Hoffer, 1973), and the skylab 13-channel S-192 scanner (Hoffer et al., 1975). The shape of the relationship shown in Fig. 2 indicates that transformed divergence increases logarithmically as the combination level increases linearly*. The spread of the points representing the five highest ranked channel combinations for each combination level represents the difference between

Figure 2. Averaged transformed divergence for the best five waveband combinations for each combination level.



*To simplify the following discussions, "combination level" will refer to the number of channels involved in any particular set of channel combinations.

successively ranked averaged transformed divergence. As seen in Fig. 2, the mean difference between successively ranked mean separabilities decreases logarithmically as the combination level increases linearly. This implies that the rank of overall mean separability as a feature selection criterion decreases in value as the number of features comprising the selected feature subset increases.

The best combined sources of information for distinguishing between various cover classes need not have as a subset the best single source of information. This is indicated in Table 2, which shows, for example, that the single channel having the highest mean TD-value (i.e., channel 6) is not included in the 2, 3, and 4 channel combination levels having the highest mean TD-values. By comparing Table 2 with Table 3, it can be observed that the best channel or channel combination for each combination level, on the basis of mean overall separability, is not necessarily superior on a per cover class basis.

Table 2. Channel combinations, ranked by overall mean TD-value for combination levels one through six.

COMBINATION LEVEL

1	2	3	4	5	6
6	3,4	3,4,5	1,3,4,5	1,3,4,5,6	1,2,3,4,5,6
3	3,5	3,4,6	3,4,5,6	2,3,4,5,6	2,3,4,5,6,7
1	2,4	3,5,6	1,3,4,6	1,2,3,4,5	1,3,4,5,6,7
5	2,5	2,4,5	3,4,5,7	1,3,4,5,7	1,2,3,4,6,7
2	3,6	2,4,6	2,4,5,7	3,4,5,6,7	1,2,4,5,6,7
4	4,6	2,5,6	2,3,4,6	2,4,5,6,7	1,2,3,4,5,7
7	1,4	1,3,4	1,3,5,6	1,2,3,5,6	1,2,3,4,6,7

Table 3. Best channels and channel combinations by TD-value for each cover class. TD-value is in parentheses.

COMBINATION LEVEL

	1	2	3	4
soil	3(1820)	24(1941)	256(1987)	1346,2346,1356(1992)
past	6(1476)	35(1878)	345(1971)	3457(1987)
crop	3(1390)	34(1836)	345(1971)	1345(1991)
pine	2(1435)	34(1780)	346(1912)	3456(1960)
pihd	2(1580)	36(1883)	356(1982)	3456(1997)
hdwd	3(1688)	34(1881)	134(1933)	2346(1952)
sghd	3(1691)	35(1933)	346(1960)	1345,1346,2346(1972)
tupe	6(1658)	34(1896)	245,345(1979)	2457(1992)
syca	5(1753)	35(1979)	345(1994)	1345,1346,1356(1999)
ccut	6(1329)	46(1707)	356(1889)	3456(1947)
mveg	4(1270)	14(1739)	134(1941)	1345(1990)
watr	5(1853)	25(1988)	246,256(1999)	1345,1346,1356(2000)

SOIL, bare soil; PAST, pasture; CROP, row and cereal crops; PINE, pine forest; PIHD, pine-hardwood mix; HDWD, old age hardwood; SGHD, second growth hardwood; TUPE, water tupelo; SYCA, sycamore hardwood; CCUT, clearcut areas; MVEG, marsh vegetation; WATR, river water and quarry water.

Examination of the transformed divergence averaged for each cover class pair indicated that the proper selection of a single channel may provide greater separability between two cover classes than a combination of two or three channels. More specifically, the channel combination with the highest mean separability for a particular combination level does not necessarily provide a greater separability for all cover class pairs than channel combinations of a lower combination level, when the combination of the lower level is not a subset of the combination of the higher level. Examples of this relationship are: soil vs. water has a mean TD-value of 1942 in channel 6 and a mean TD-value of 1824 in channel combination 3,4; PIHD vs. CCUT has a mean TD-value of 1835 in channel 6 and a mean TD-value of 1641 in channel combination 3,4; PINE vs. MVEG has a mean TD-value of 1424 in channel 1 (the channel ranked third on the basis of mean overall TD-value) and the mean TD-value of 1182 in channel combination 3,4 (the number one ranked channel combination of all combinations involving two channels). The same relationship holds for many other cover class pairs. Such a relationship was not found when the lower level channel combination was a subset of the higher level channel combination (as would be expected).

The additional average separability achieved for each cover class, by increasing the combination level, varies greatly between cover classes and combination levels, but generally decreases logarithmically with increasing combination level. Figure 3 can be thought of as a "separability response surface." The apparent length of the lines connecting different combination levels of the same cover class is proportional to the added separability resulting from the information in the additional channel. Note that the greatest increase in separability due to the addition of the second channel occurs with second growth hardwood. As one would expect, the smallest increase in separability occurs with that cover class with the highest single channel separability (soil, in this case). It should be noted that the lines connecting the different cover classes are present merely to indicate relative differences of separability and in no way imply any functional relationship.

Figure 3 plots the maximum transformed divergence observed for each cover class in each combination level. This displays the maximum separability attainable for each cover class if the waveband combinations were selected on the basis of each cover class TD-value alone. As is clearly shown, the specific waveband combination resulting in each particular TD-value for any given waveband combination level is not constant over the different cover classes. In comparing Figures 3 and 4, it is apparent that the shapes of the curves increase in similarity with an increase in waveband combination level and are nearly identical in shape after combination level 4. This indicates that the separability by cover class provided by the best overall channel combination (Fig. 3) is nearly identical to the separability by cover class provided by the best channel combination for each individual cover class (Fig. 4) beyond waveband combination levels of 4. Thus, the best four waveband combination, based on overall transformed divergence, should provide very close to the maximum classification accuracy for each individual cover type. However, if one were interested only in a particular cover type, high classification accuracy could be achieved using less than four channels of data.

Figure 3. Averaged transformed divergence provided by the overall best waveband combination by waveband combination level and cover class.

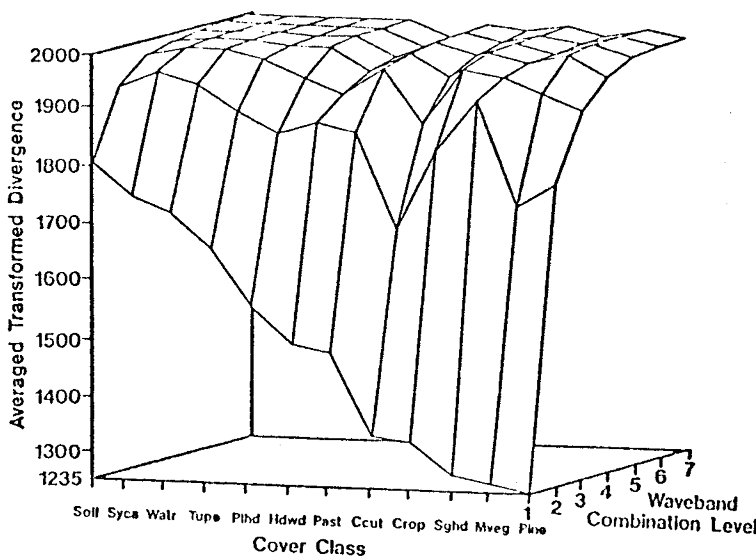
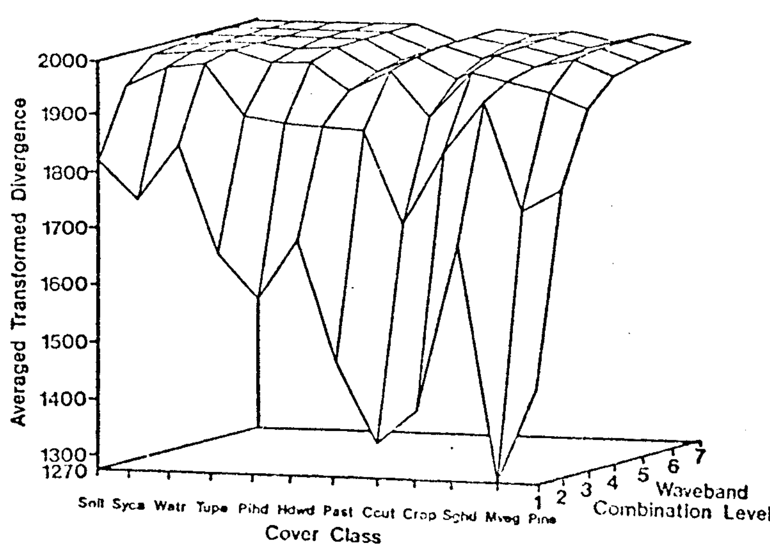


Figure 4. Averaged transformed divergence provided by the best waveband combination for each cover class by waveband combination level and cover class.



SUMMARY AND CONCLUSIONS

Based upon the results of this study, one would not expect a computer-based classification employing more than four channels to provide much improvement in classification accuracy. The highest overall mean separability was provided by channels 1, 3, 4, and 5 (0.45-0.52, 0.63-0.69, 0.76-0.90, and 1.0-1.3 μm). This channel combination did not always provide the highest mean separability by cover class nor by pairs of cover classes. A different set of cover classes, or even a subset of the cover classes considered in this work, could result in other channel combinations yielding higher predicted classification accuracies.

Results such as these are highly data and application dependent. The conclusions pertain to channel subsets selected for classification and in no way imply that scanner systems need only obtain data in those channels in order to adequately provide remote sensory data to the various disciplines. Similar studies involving different cover classes and different seasons need to be conducted along with follow-up studies involving actual classifications.

ACKNOWLEDGEMENTS

This work was supported by NASA Contract NAS9-15889, NASA Johnson Space Center, Houston, Texas.

REFERENCES

Anuta, P.E. and D.F. Strahorn. 1973. "Sun Angle Effect Preprocessing with Predicted Ramp Functions." LARS Technical Memorandum T-12.

Anuta, P.E. and Nim-Yau Chu. 1979. "Multidimensional Scaling for Clustering of Dissimilar Data Type." LARS Information Note, prepared for NSF, grant #ENG-7614400. 27 pp.

Bartolucci, L.A. and R.B. de Castro. 1979. "Clustering of Landsat MSS Data: Certain Limitations." LARS Technical Report.

Billings, W.D. and R.J. Morris. 1951. "Reflection of Visible and Infrared Radiation from Leaves of Different Ecological Groups." American Journal of Botany, Vol. 38, May 1951. pp. 327-331.

Colwell, J.E. 1973. "Bidirectional Reflectance of Grass Canopies for Determination of Above Ground Standing Biomass." Ph.D. dissertation. University of Michigan, Ann Arbor, Michigan.

Colwell, J.E. 1974. "Vegetation Canopy Reflectance." Remote Sensing of Environment, 3. pp. 175-183.

Dynkin, E.B. 1961. "Necessary and Sufficient Statistics for a Family of Probability Distributions." In Selected Translations in Mathematical Statistics and Probability, Vol. 1. pp. 17-39.

Duda, R.O. and P.E. Hart. 1973. "Pattern Classification and Scene Analysis." John Wiley & Sons, New York. 482 p.

Fukunaga, K. and D.L. Kessel. 1971. "Estimation of Classification Error." IEEE Trans. on Computers. Vol. 20:12. pp. 1521-1527.

Gates, D.M., H.J. Keegan, J.C. Schleter and V.R. Weidner. 1965. "Spectral Properties of Plants." Applied Optics, Vol. 4, No. 1. Jan. 1965. pp. 11-20.

Gates, D.M. and W. Tantraporn. 1952. "The Reflectivity of Deciduous Trees and Herbaceous Plants in the Infrared to 25 Microns." Science, 115. pp. 613-616.

Gausman, H.W., W.A. Allen and R. Cadenas. 1969. "Reflectance of Cotton Leaves and Their Structure." Remote Sensing of Environment, Vol. 1. pp. 19-22.

Gausman, H.W. 1977. "Reflectance of Leaf Components." Remote Sensing of Environment, Vol. 6. pp. 1-9.

Hoffer, R.M., R.E. Joosten, R.G. Davis and F.R. Brumbaugh. 1978. "Land Use and Cartography." Chapter 2 in: Skylab EREP Investigations Summary. NASA SP-399. National Aeronautics and Space Administration, Washington, D.C. pp. 7-77.

Idso, S.B. and C.T. de Wit. 1970. "Light Relations in Plant Canopies." Applied Optics, Vol. 9, No. 1. pp. 177-184.

Jurica, G.M. and W.L. Murray. 1973. "The Atmospheric Effect in Remote Sensing of Earth Surface Reflectivities." LARS Info ation Note.

Kailath, T. 1967. "The Divergence and Bhattochargga Distance Measures in Signal Selection." IEEE Trans. Comm. Tech., Vol. 15:7. pp. 52-60.

Kanal, L. and B. Chandrasekaran. 1971. "On Dimensionality and Sample Size in Statistical Pattern Classification." In Pattern Recognition, Pergamon Press, Vol. 3. pp. 225-234.

Knipling, E.B. 1970. "Physical and Physiological Basis for the Reflectance of Visible and Near-IR Radiation from Vegetation." Remote Sensing of Environment, Vol. 1. pp. 155-159.

Kondratyev, K.Y.A. 1969. "Radiation in the Atmosphere." Academic Press. 912 pp.

Landgrebe, D.A., L. Beihl and W. Simmons. 1977. "An Empirical Study of Scanner System Parameters." IEE Transactions on Geoscience Electronics, Vol. GE-15, No. 3. pp. 120-130.

Nilson, T. 1970. "A Theoretical Analysis of the Frequency of Gaps in Plant Stands." Agricultural Meteorology, Vol. 8. pp. 25-38.

Salomonson, V.V. 1978. "Landsat-D Systems Overview." Proc. of the 12th Inter. Symp. on Rem. Sens. of the Environ. pp. 371-385.

Silva, L.F. 1978. "Radiation and Instrumentation in Remote Sensing." In Remote Sensing: The Quantitative Approach, McGraw-Hill Inter. Book Co., N.Y. pp. 21-133.

Smith, J.A. and R.E. Oliver. 1972. "Plant Canopy Models for Simulating Composite Scene Spetroradiance in the 0.4 to 1.05 Micrometer Region." Proc. of the 8th Symp. on Remote Sensing of the Environ. pp. 1333-1353.

Suits, G.H. 1972. "The Calculation of the Directional Reflectance of a Vegetative Canopy." Remote Sensing of Environment, No. 2. pp. 117-125.

Swain, P.H. "Fundamentals of Pattern Recognition in Remote Sensing." In Remote Sensing: The Quantitative Approach, McGraw-Hill Inter. Book Co. pp. 136-185.

Swain, P.H., T.V. Robertson and A.G. Wacker. 1971. "Comparison of the Divergence and B-distance in Feature Selection." LARS Information Note. 12 pp.

Swain, P.H. and R.C. King. 1973. "Two Effective Feature Selection Criteria for Multispectral Remote Sensing." LARS Information Note. 5 pp.

Wacker, A.G. and D.A. Landgrebe. 1972. "Minimum Distance Classification in Remote Sensing." LARS Print. 25 pp.

Wooley, J.T. 1971. "Reflectance and Transmittance of Light by Leaves." Plant Physiology 47:656-662.