# Performance Comparisons Between Information Extraction Techniques Using Variable Spatial Resolution Data

R. S. Latty, R. Nelson, B. Markham, D. Williams, D. Toll, and J. Irons
Earth Resources Branch, Code 623, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

ABSTRACT: The objective of the study was to determine to what extent performance levels achieved with per-field classifiers with Thematic Mapper (TM) resolution data would exceed that of the conventional per-point Gaussian maximum likelihood (GML) classifiers with Landsat Multispectral Scanner (MSS) resolution data. The analyses were performed with actual TM data (six reflective bands, acquired for the Washington, DC area on 2 November 1982). The TM data were degraded to approximately 90 m spatial resolution with a pixel period of 57 m to simulate the MSS data resolution and sampling properties. Results obtained over four replicates with the per-point GML classifier were compared to those achieved with the per-field supervised ECHO (Extraction and Classification of Homogeneous Objects) classifier. A nonparametric discriminant function was also examined. This discriminant employs the cumulative bin-wise differences between normalized univariate histograms. These histograms are defined over each individual surface feature.

The mean accuracy (78.1 percent) achieved with the per-point GML classifier using 57 m data exceeded the levels achieved with the 28.5 m data using supervised ECHO (60.1 percent) or the per-point GML classifier (72.0 percent). The accuracy achieved using 28.5 m data with the nonparametric approach was 94.5 percent (87.0 percent) for classifying polygons of minimum size of 90 (40) pixels or larger.

## INTRODUCTION

ONE OF THE PRINCIPAL advances noted for the Thematic Mapper (TM) sensor is the decreased instantaneous field of view (IFOV). The 42.5 microradian IFOV of TM and the 710 km nominal orbit altitude result in a 30 m nominal spatial resolution at the Earth surface. This is a considerable decrease in the projected pixel area when compared to the 79 m nominal spatial resolution of the Landsat Multispectral Scanner (MSS). This increased spatial resolution was established as a design specification by the TM Working Group (Harnage, 1975; Salomonson, 1978) with the expectation that this would greatly facilitate the extraction of information from the data, increase the level of information detail provided by the data, and improve spatial (or positional) accuracy.

A visual examination of the photographic imagery derived from TM data, compared to MSS data, confirms these earlier expectations. However, the human visual system exploits many properties of the scene which are not utilized by currently employed image-to-information computational procedures. The spatial properties of the scene are prominent among those which are not exploited. Since the spatial properties, in particular, appear to be strongly dependent on the spatial resolution of the imaging system, visual comparisons of the derived imagery are often misleading.

From an analytical standpoint, increased spatial resolution would be expected to influence the performance of conventional per-point classifying algorithms in different ways. Increasing the spatial resolution decreases the relative number of pixels which occur on boundaries between different cover types. The rate at which the relative number of boundary pixels decreases with an increase in spatial resolution is dependent on the size and shape of the surface features. In general, TM data should provide a smaller fraction of boundary pixels and, therefore, would be expected to provide higher classification accuracies. However, increased spatial resolution also resolves the variability in composition of many of the surface features (e.g., individual tree crowns and between-crown shadows in forests; streets, yards, and houses in residential areas). The increased spectral variability resulting from the increased spatial resolution tends to result in spectral similarity between individual pixels of different sur-

face cover types. Pixels on individual trees in large forests tend to have spectral values similar to pixels on individual trees in residential areas. While it may be argued that "forest" and "residential area" are abstractions unrealistic to be expected of automated interpretation of spectral data, it is precisely this type of information which is often required by many applications.

Spatial resolution not only affects the magnitude of spectral variability, it also influences the shape of the distribution of spectral values. The manner and extent to which the spectral distribution is influenced by the spatial resolution are dependent on the spatial distribution and spatial frequency of the areas of differing reflectance. Surface features with regularly spaced, high spatial frequency contrasts may be represented by spectral distributions which tend from unimodal at low spatial resolution (>1 times the periodicity of surface variation) to increasingly bimodal at high spatial resolution (<0.5 times the periodicity of surface variation).

While the influence of spatial resolution on various properties of the data is fairly clear, the resulting relationships between spatial resolution and the level of accuracy and precision attainable in extracting information from the data are less identifiable. Previous investigations conducted with per-point pattern recognition procedures concur that an increase in spatial resolution does not necessarily improve performance, and, in fact, often depresses performance (Sadowski and Sarno, 1976; Landgrebe et al., 1977; Latty and Hoffer, 1981; Markham and Townshend, 1981). Considerations regarding the fundamental operation of the per-pixel pattern recognition procedures relative to the high spatial resolution properties, suggest that improvements may be expected through the use of per-field processors (i.e., use of the spectral properties over some number of pixels of the surface feature in identifying the feature). Therefore, the objecive of the study was to determine to what extent the performance level observed for the per-point Gaussian maximum likelihood (GML) classifier using approximately 90 m spatial resolution data could be improved by employing per-field algorithms designed in accord with the data properties of the higher (e.g., 30 m), spatial resolution.

## BACKGROUND

Empirical studies have been conducted to assess the influence of spatial resolution on classification performance. In particular, some of these studies have examined the influence of spectral variability. Kan and Ball (1974) found that as the spatial resolution increased, the spectral variance increased, thus reducing the separability of classes in feature space. Sadowski and Sarno (1976) found spectral classes of differing forest cover types were often more statistically similar than spectral classes of the same cover class. This was observed to be more prevalent at higher spatial resolutions than at lower spatial resolutions. Latty and Hoffer (1981) found that the sharpest decrease in classification accuracy with increases in spatial resolution were observed for those classes with higher spectral variability (e.g., pine forest, second growth hardwood, and cut-over forest). No significant change in performance was observed for classes of lower spectral variability (e.g., soil, pasture, and crops). Markham and Townshend (1981) observed similar correlation between surface feature classes of high spectral variability and significant decreases in classification accuracy with increases in spatial resolution. They observed, further, that the influence of spectral variability was dependent on the relative positioning of the classes in multispectral feature space. Classes which were well separated from neighboring classes in multispectral feature space did not vary in classification accuracy with increasing spatial resolution even though the variance of these classes did increase with increased resolution. Landgrebe et al. (1977) attributed a slight decrease in pure pixel classification accuracy with increased spatial resolution observed in their study to the decrease in the signal/noise ratio accompanying the increased spatial resolution.

Similarly, empirical work has been conducted to determine the influence of boundary pixels on the relationship between classification performance and spatial resolution. Morgenstern et al. (1977) found that acreage estimation accuracy decreased in agricultural areas with decreasing spatial resolution and that this relationship was very much dependent on field size. They inferred that this was due to the higher relative frequency of boundary pixels in the lower spatial resolution data. Similar results were observed by Thomson and Erickson (1976). Markham and Townshend (1981) observed a decrease in classification accuracy for small surface features (i.e., roads and small water bodies) with decreasing spatial resolution due to boundary pixels. More recently Irons et al. (1984) and Williams et al. (1984) observed that spatial resolution was a nonsignificant factor as a source of variance in classification accuracy when the boundary pixels were included, but was a very significant factor when the boundary pixels were omitted. For a study site in the Washington, DC area, the counter influence of boundary pixels mitigated the adverse affect of increased spectral variability with increased spatial resolution. However, for a study site in Clarion County, PA, involving small surface features, much higher classification accuracies were observed for the higher spatial resolution data. A study by Markham (1984) which focused on detectability of surface water as a function of surface area with different spatial resolutions, demonstrated the strong dependence of

performance with variable spatial resolution on the size of the targets considered.

Based on the previous experience with 30 m (and higher) spatial resolution data and the more conventional per-point, or pixel-by-pixel, processing strategy, techniques have been sought to exploit the advantageous properties of high spatial resolution data. Some of the processing strategies attempt to exploit the increased spectral variability associated with increased spatial resolution in the discrimination process. This is the strategy of all efforts to use image texture, covariance differences, histogram shape, or piece-wise densities in the discrimination process.

The ECHO (Extraction and Classification of Homogeneous Objects) classification algorithm is one of the earlier integrated attempts to employ spectral measurements over a number of pixels to identify the surface feature. ECHO is an algorithm which was developed during the Large Area Crop Inventory Experiment (LACIE) era at Purdue University, Laboratory for Application of Remote Sensing. The various versions of the ECHO algorithm are discussed in Kettig (1975), Kettig and Landgrebe (1975), Kast and Davis (1977), and Landgrebe (1980). The experimental work and theoretical considerations which led to the strategy embodied in ECHO are found in Wacker (1971), Wacker and Landgrebe (1972), and Kettig and Landgrebe (1973).

Kettig (1975), Kettig and Landgrebe (1975), and Wiersma and Landgrebe (1976) observed a higher classification accuracy for Landsat MSS data when using the ECHO algorithm than when using the per-point Gaussian maximum likelihood (GML) classifier. Landgrebe et al. (1977) observed no significant difference between the acreage estimation accuracies achieved with ECHO and the per-point GML classifier at each simulated spatial resolution. ECHO performed slightly better at each resolution than the per-point GML classifier based on percent correctly classified field-center pixels. The ECHO algorithm performed increasingly better than the per-point GML classifier with decreasing signal/noise levels for 30 m spatial resolution data. Latty and Hoffer (1981) found that the ECHO algorithm performed better on the 30 m simulated spatial resolution data than the per-point GML classifier on the 30 m data, based on percent correctly classified field-center pixels. The difference was greatest for cover classes for which the highest spectral variability was observed.

## APPROACH

The overall approach and steps involved in the study are summarized in the schematic diagram in Figure 1.

### GROUND REFERENCE DATA

Aerial photography was collected over areas to the east of the Washington, DC metropolitan area on 13 July 1982. A subset of the 1:40,000 scale color infrared photographs served as the ground reference base. Each area on the selected photography was identified as: water, forest (pine and mixed hardwood, of varying canopy closure), pasture, agricultural crop lands (primarily corn, soybean, tobacco, bare soil, and crop stubble), and developed land (residential, commercial, multifamily dwellings, and transportation corridors). The boundaries of each cover type were digitized. The polygons generated from digitizing the cover type boundaries were used to generate raster images corresponding to each of the selected photographs. Two raster images were generated for each photograph: one with a pixel frequency every 28.5 m of ground scale, the second with a pixel frequency of every 57 m.

### SPECTRAL DATA

The first TM scene of the Washington, DC area which was adequately cloud-free with all six reflective bands properly functioning was collected on 2 November 1982 (path 15, row 33, scene number E-40109-151400). While the spatial resolution of TM is approximately 30 m, the geometric resampling of the data in the Scrounge system (Lyon, et al., 1983) resulted in a pixel period of 28.5 m. Subscenes corresponding to the areas covered by the selected aerial photographs were extracted from the TM scene. Only the six reflective bands were used. A second set of subscenes was generated, from the first, which had 90 m effective spatial resolution with a pixel period of 57 m (see Williams et al., 1984 for details of the procedure used). Each raster ground reference image was registered to the corresponding subscenes of spectral data.

### DATA SAMPLING

Each of the two (28.5 m and 57 m pixel frequency) images were sampled four times independently with two different sampling strategies. Each of the four independent samples provided a replicate (i.e., gave sensitivity to the statistical tests on the observed performance differences). The two sampling approaches provided two fundamentally different types of data on which to perform the experiment.

*Pixel-by-Pixel Sampling.* One set of replicates was generated by randomly sampling the ground reference image without replacement, by class, on a pixel-by-pixel basis. That is, each pixel of each class is treated as an independent, random event. Each selected pixel was then randomly assigned to the role of either training or testing (1200 training and 450 test pixels for each information class for the 28.5 m data; and 1000 training and 375 text pixels for each information class for the 57 m data).

*Polygon Sampling.* The alternate sampling approach generated a set of replicates where the individual polygons of the reference image were ran-
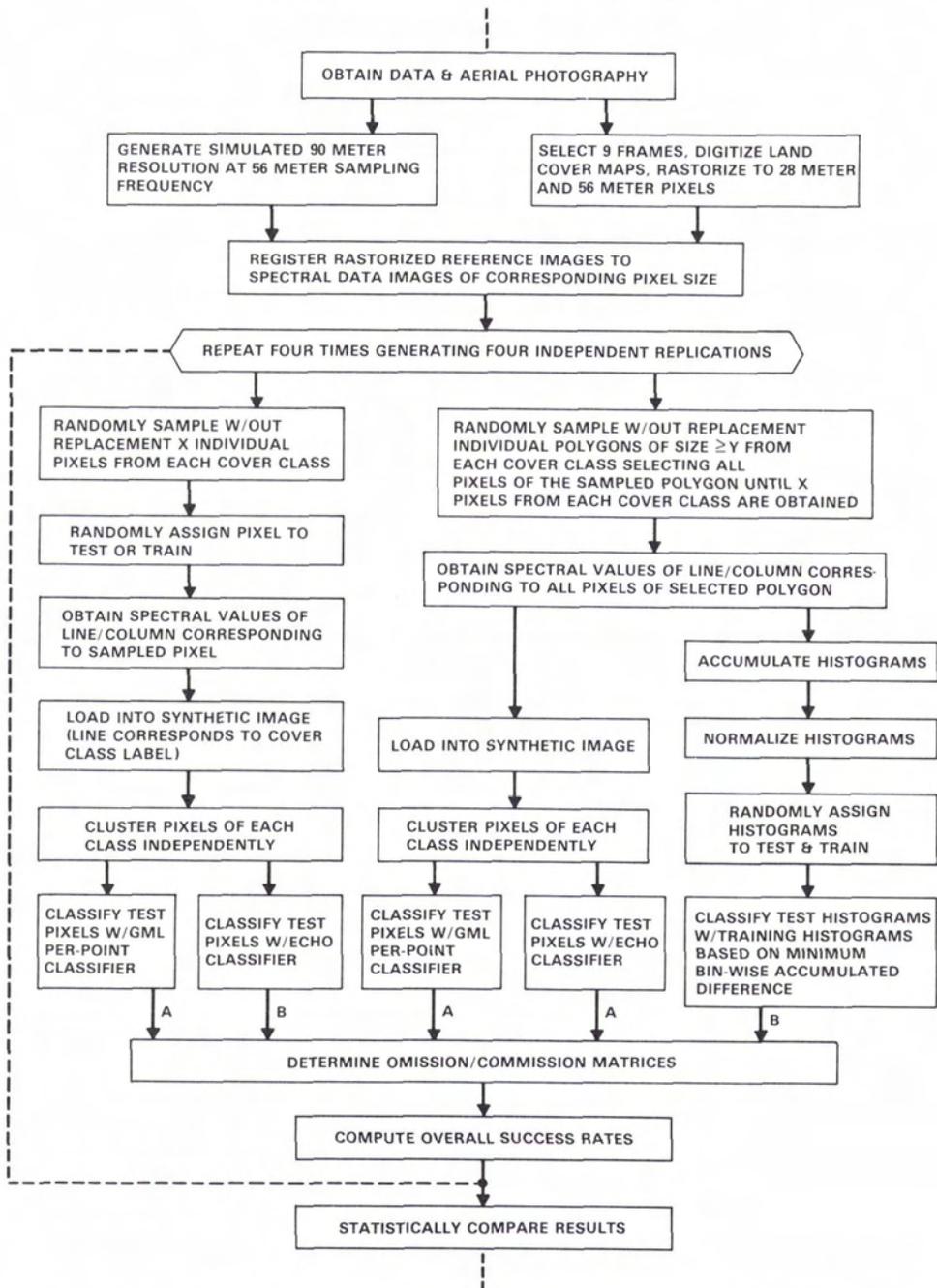
**EXPERIMENT CONCEPTUAL FLOW SCHEMATIC**



FIG. 1.   Design schematic summarizing the processing steps involved in conducting the study. Those paths labeled "A" were followed with both spatial resolutions, while those labeled "B" were followed with only the 28.5 m spatial resolution.

domly sampled without replacement by surface class. All pixels of each selected polygon were used to form either the training pixels or test pixels (1000 training and 600 test pixels for each information class for the 28.5 m data; and 750 training and 450 test pixels for each information class for the 57 m

data). The polygonal sampling approach was used to emulate more closely the form of training most commonly used in applied data analysis. The training pixels therefore provided a range of covariance more consistent with the design considerations of the supervised ECHO (see the discussion of the homoge-

neity and annexation criteria in Kettig and Land-grebe, 1975).

The line-column coordinate of each selected pixel in the reference image was used to direct the extraction of the spectral values from the coregistered spectral data image.

## DATA ANALYSIS

Each set of training pixels for each surface class in each replicate of each resolution was independently clustered into three spectral classes through the use of an iterative minimum Euclidean distance-to-the-means clustering algorithm (*CLUSTER of the LARSYS image processing system, see Phillips, 1973, Vol. 2). The mean vectors and covariance matrices of each spectral class of each surface cover class were used to classify the test pixels of the sampled spectral image data with a GML per-point classifier (*CLASSIFYPOINTS of the LARSYS image processing system; Phillips, 1973, Vol. 2; Swain and Davis, 1978, p. 150) and the original subscenes with the supervised ECHO classifier (*SECHO of LARSYS; Kettig and Landgrebe, 1975; Kast and Davis, 1977).

Similar steps were employed for the data obtained through polygon sampling, but practical considerations led to performance of this segment of the experiment on the Landsat Analysis System at NASA/GSFC. Clustering was, therefore, performed with the ISOCLS algorithm (Minter, 1972; Kan, 1973; Kan et al., 1973). The parameters employed in ISOCLS were set to emulate the algorithm used in LARSYS. The sampled spectral data were classified with a GML per-point classifier, and the subscene data were classified with the supervised ECHO algorithm.

An omission/commission frequency matrix was constructed for each classified image. The overall percent correct classification (PCC) accuracy observed for each replicate of each factor was arc sin square root transformed ($\theta = \sin^{-1}\sqrt{PCC}$) in order to satisfy the assumptions of analysis of variance (i.e., equal variances and normally distributed variables; see Neter and Wasserman, 1974, p. 426).

A Nonparametric Discriminant. Subsequent to examining the results achieved with the GML and ECHO classifiers a discriminant function was sought to:

- eliminate the dependence of the discriminant on restrictive assumptions regarding statistical properties of the data (i.e., data are unimodal, normally distributed, random variables),
- exploit the large number of spatially contiguous pixels comprising a single surface feature (e.g., for feature type definitions used in this study, 91 percent of the pixels in the DC images were contained in features of 40 or more pixels, 84 percent were of features of 90 or more pixels),
- avoid confusion between spectral subsets (clusters) of different surface features by providing a technique which does not require clusters,
- exploit the distributional properties characteristic

of spatially intact surface features in their discrimination,
- stabilize the estimate of the spectral measurement through an increased number of measurements used in representing each surface feature.

Nonparametric representations avoid the dependence on a priori assumptions regarding the distribution of data values, and therefore, tend to satisfy the first and third criteria identified above. Histograms represent the actual distribution of the data. Differences between histograms are therefore expected to exhibit distributional differences as well as differences in the mean and variance of the data. Stable estimates of actual histograms, however, require a larger number of observations than do stable estimates of parametric representations. How many pixels are required over a single feature in order to provide sufficiently stable histograms to serve in discrimination is not known. It was hypothesized that the number of pixels occurring in single features is, in general, sufficiently large to enable significant improvements in discrimination success rates. From these considerations a histogram-based discriminant function was developed. The discriminant can be considered as a multi-univariate cumulative bin-wise difference (CBD) between normalized histograms from a set of known surface features and a single unknown surface feature. That surface feature for which the CBD is minimum is the class to which the unknown is assigned. The difference between normalized histograms is given as:

$$D_j = \sum_{i=1}^{p} \sum_{k=1}^{q} |h_j(i,k) - H(i,k)| \tag{1}$$

where:

$D_j$ = the cumulative bin-wise difference between normalized histograms of the jth known surface feature and the unknown surface feature,

$p$ = the number of dimensions, or spectral bands,

$q$ = the number of quantization (bin) levels contained in the two histograms,

$h_j(i,k)$ = the normalized frequency of the kth bin level, of the ith band, for the jth surface feature of known identity,

$H(i,k)$ = the normalized frequency of the kth bin level, of the ith band, of the surface feature of unknown identity.

The normalized histogram is computed from the original histogram by dividing the frequency of each bin by the sum of the frequencies over all bins (i.e., the total number of observations over the surface feature).

The CBD approach presented here is only the discriminant function component of a complete classifier. In order to employ such a per-field approach,

the scene must first be segmented into individual, or subsets of individual, surface features. Although one of the authors (Latty, 1983; Latty, 1984) has been engaged in the development of a scene segmentation algorithm, the results presented here are based on the use of the polygons of the reference data as the basis for computing the feature histograms. Thus, the results indicate how well the features can be *discriminated* assuming the population of pixels belonging to each individual feature have previously been correctly defined.

Four independent sets of polygons were generated through the polygon sampling approach for each of two minimum sizes (40 and 90 pixels). The number of polygons selected for each surface feature type varied according to the minimum polygon size, due to the number of available polygons in the scene. These are presented with the results.

## RESULTS

Figure 2 illustrates the mean performance levels observed for each set of factor level combinations (spatial resolution and classifier) over all four replicates. Table 1 provides the computed $F$-statistic and α-level for all observed pair-wise differences between mean arc sin transformed performance levels. The four replicates in each treatment and the two treatments per test provide eight total observations ($n_t$) per test, resulting in seven ($n_t - 1$) degrees of freedom for the total sums-of-squares. One degree of freedom is associated with the treatment sums-of-squares ($r - 1$, where $r$ is the number of treatments), and six ($n_t - r$) degrees of freedom are associated with the sums-of-squares of error. Pair-wise

PERCENT



FIG. 2.   Mean percent of test pixels correctly classified over four replicates for each factor combination.

differences with a probability of falsely regarding equal means as being different (Type I error) not exceeding 5 percent are considered significant. However, while the significance of differences are discussed relative to the 5 percent probability, the actual probability associated with the computed $F$-value is presented in the following discussion.

### PIXEL-BY-PIXEL SAMPLING RESULTS

The percent correct classification (PCC) achieved with the 28.5 m and 57 m data, employing the per-point GML classifier for each replicate generated by the pixel-by-pixel sampling are presented in Table 2. The performance levels attained with the 57 m data were significantly greater (by 6.1 percent) than that achieved with the 28.5 m data at an α-level of 0.018 when using the per-point classifier.

Table 3 provides the PCC levels attained with the ECHO classifier using the 28.5 m data generated by the pixel-by-pixel sampling approach. The PCC levels achieved with the supervised ECHO on the 28.5 m data were an average of 11.8 percent lower than the mean PCC level achieved with the same data using the per-point GML classifier. This difference was significant below the 0.003 α-level. The PCC levels attained with the GML per-point classifier using 57 m data were, on the average, 17.9 percent higher than those achieved with the supervised ECHO using the 28.5 m data. This difference was significant at the 0.0001 α-level.

The lower PCC levels achieved with the supervised ECHO classifier are believed to be due in part to a magnitude of within-class variance inconsistent with the design properties of the supervised ECHO algorithm. The pixel-by-pixel sampling results in a high fraction of nonspatially contiguous pixels comprising the sample. Furthermore, the number of individual surface features from which the sample pixels are extracted tends to be higher for the pixel-by-pixel samples than for the polygon samples. These two properties of the sample pixels result in high within-class variances. Increasing within-class variances in the training statistics increases the apparent homogeneity of cells. This would be expected to increase the frequency of cells which pass the homogeneity test which are also comprised of different surface feature types. A similar relationship prevails for the influence of variance on the likelihood ratio computed as the test for annexing adjacent cells (cells previously determined to be homogeneous).

### POLYGON SAMPLING RESULTS

The selection of complete polygons, or some set of spatially contiguous pixels, in the sampling provides a closer approximation to what is actually done in computer-aided image interpretation applications. Analysts normally develop training statistics by circumscribing areas of identifiable surface features, or mapping known surface feature locations
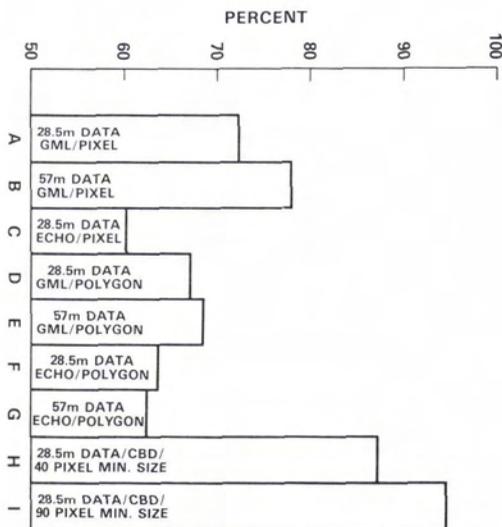
TABLE. 1. THE $F$-STATISTIC ($\alpha$-LEVEL) VALUES ASSOCIATED WITH EACH PAIR-WISE DIFFERENCE BETWEEN THE MEANS OF THE ARCSIN TRANSFORMED PERCENT CORRECTLY CLASSIFIED PIXELS OF EACH FACTOR LEVEL COMBINATION EXAMINED

|   | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|
| A | 10.45 (.018) | 23.27 (.003) | 1.05 (.345) | 2.50 (.165) | 6.42 (.044) | 7.54 (.033) | 44.49 (.0005) | 210.18 (.0001) |
| B | | 159.15 (.0001) | 7.89 (.031) | 81.01 (.0001) | 30.00 (.0015) | 30.28 (.0015) | 35.24 (.001) | 602.52 (<.0001) |
| C | | | 2.87 (.141) | 23.15 (.003) | 1.00 (.357) | 0.33 (.585) | 165.90 (.0001) | 724.97 (.0001) |
| D | | | | 0.06 (.813) | 0.70 (.434) | 1.12 (.331) | 27.23 (.002) | 79.23 (.0001) |
| E | | | | | 3.00 (.134) | 4.02 (.092) | 104.17 (.0001) | 702.22 (.0001) |
| F | | | | | | 0.09 (.779) | 65.57 (.0002) | 195.67 (.0001) |
| G | | | | | | | 64.64 (.0002) | 181.61 (.0001) |
| H | | | | | | | | 39.88 (.0007) |

A—28.5 meter/GML /pixel sampling
B— 57 meter/ GML /pixel sampling
C—28.5 meter/ECHO/pixel sampling
D—28.5 meter/GML /polygon sampling
E— 57 meter/ GML /polygon sampling
F—28.5 meter/ECHO/polygon sampling
G— 57 meter/ ECHO/polygon sampling
H—28.5 meter/CBD /40 pixel minimum
I —28.5 meter/CBD /90 pixel minimum

into the image coordinate system. Spectral class statistics computed from such sets of spatially contiguous pixels are more consistent with the design of the homogeneity and annexation tests of the supervised ECHO algorithm.

The PCC levels attained with the 28.5 m data and the 57 m data, using both the per-point and supervised ECHO classifiers are presented in Table 4. The PCC levels observed for the 28.5/GML combination were not found to be significantly different from those of the 57/GML combination ($\alpha = 0.813$). The 4.0 percent difference between the mean 28.5/ECHO combination and the mean 28.5/GML combination results was not significant ($\alpha = 0.434$). None

TABLE 2. PERCENTAGE OF TEST PIXELS CORRECTLY CLASSIFIED FOR EACH INFORMATION CLASS IN EACH OF THE FOUR REPLICATIONS, USING 28.5 METER AND 57 METER DATA WITH A PER-POINT GAUSSIAN MAXIMUM LIKELIHOOD CLASSIFIER, BASED ON PIXEL-BY-PIXEL SAMPLING

|  | 28.5 Meter | | | | 57 Meter | | | |
|---|---|---|---|---|---|---|---|---|
| Class | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Water | 90.4 | 90.9 | 99.1 | 90.2 | 97.1 | 94.6 | 94.1 | 94.1 |
| Forest | 82.2 | 86.4 | 78.9 | 82.9 | 88.3 | 81.0 | 80.0 | 90.7 |
| Pasture | 55.8 | 52.9 | 47.6 | 59.1 | 70.7 | 70.2 | 70.7 | 75.6 |
| Crop lands | 67.1 | 68.7 | 57.3 | 70.4 | 72.2 | 69.3 | 80.0 | 68.3 |
| Developed | 71.3 | 67.6 | 48.7 | 72.4 | 66.3 | 69.3 | 64.4 | 64.4 |
| [a]$x._i$ | 73.4 | 73.3 | 66.3 | 75.0 | 78.9 | 76.9 | 77.9 | 78.7 |
| *t$x._i$ | 58.7 | 58.9 | 54.5 | 60.0 | 62.6 | 61.3 | 62.0 | 62.5 |
|  | | [b]$x.. = 72.0$ | | | | [b]$x.. = 78.1$ | | |
|  | | *t$x.. = 58.1$ | | | | *t$x.. = 62.1$ | | |

* Arcsin square root transformation ($\sin^{-1}\sqrt{PCC}$, where $0 \leq PCC \leq 1$).
[a] Mean computed over all categories within each replicate.
[b] Mean computed over all replicate means within each treatment.

TABLE 3. PERCENTAGE OF TEST PIXELS CORRECTLY CLASSIFIED FOR EACH INFORMATION CLASS IN EACH OF THE FOUR REPLICATIONS, USING 28.5 METER WITH THE SUPERVISED ECHO ALGORITHM, BASED ON PIXEL-BY-PIXEL SAMPLING.

| Cover Class | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Water | 100.0 | 99.8 | 99.8 | 100.0 |
| Forest | 60.0 | 62.0 | 67.3 | 72.0 |
| Pasture | 48.9 | 43.8 | 41.6 | 52.2 |
| Crop lands | 50.7 | 48.0 | 51.3 | 60.2 |
| Developed | 39.8 | 29.8 | 40.7 | 35.1 |
| $^a x_{.j}$ | 59.9 | 56.7 | 60.1 | 63.9 |
| $^* t x_{.j}$ | 50.7 | 48.8 | 50.8 | 53.1 |

$$^b x.. = 60.1$$
$$^* t x.. = 50.9$$

\* Arcsin square root transformation ($\sin^{-1} \sqrt{PCC}$, where $0 \le PCC \le 1$).
$^a$ Mean computed over all categories within each replicate.
$^b$ Mean computed over all replicate means within each treatment.

of the pair wise differences between mean overall performance levels were found to be significant at $\alpha$-levels lower than 0.05. These results indicate that there are no real differences in performance between the supervised ECHO and the per-point GML classifiers (for either of the spatial resolutions examined). Nor are there any real differences in performance between the 28.5 m and 57 m data.

## NONPARAMETRIC PER-FIELD (CBD) APPROACH

The CBD approach was tested using only the 28.5 m data at two different minimum polygon sizes. The sizes were arbitrarily selected at 40 and 90 pixels. The performance level for each replicate for each minimum size are presented in Table 5. The overall performance level for the 90 and 40 pixel minimum size polygons are 94.5 and 87.0 percent, respectively. That is, of all pixels belonging to polygons of 90 (40) pixels or larger, an average of 94.5 (87.0) percent of those pixels were classified correctly using the CBD approach on the 28.5 m data. Pixels in polygons of 90 (40) pixels or more constitute 83.9 (90.8) percent of the total pixels in the subscenes examined given the surface feature definitions used in this study. Assuming all pixels in polygons of less than 90 (40) pixels were erroneously classified then the average performance over the entire subscenes would be 79.3 (79.0) percent. If the pixels of polygons less than 90 (40) pixels were assumed to have error rates equal to those achieved with the GML per-point approach then the average performance over the entire subscenes would be 90.0 (85.2) percent.

The observed performance levels for the 90 (40) pixel minimum size polygons were an average of 16.4 (8.9) percent higher than the highest performance levels observed for the 57 m data. These differences were significant below the 0.001 $\alpha$-level.

The differences are considerably larger between results obtained for the polygonal sampling approach.

## SUMMARY AND DISCUSSION

An experiment was conducted which allowed a rigorous test of the influence of classifier design, with data spatial resolution of TM (30 m) and approximately that of the Landsat MSS (90 m), on classification performance for a particular TM scene. The experiment involved evaluation of results for the per-point GML classifier and the supervised ECHO classifier with training and test data obtained in two fundamentally different ways: pixel-by-pixel sampling, and polygon sampling. The capability of a nonparametric per-field discriminant function was also examined. The results obtained were compared to those of the per-point GML and the supervised ECHO classifiers. To summarize, the results were:

- GML per-point results achieved with 57 m data were significantly ($\alpha = 0.018$) higher than those of the 28.5 m data based on pixel-by-pixel sampling,
- GML per-point results did not differ significantly ($\alpha = 0.813$) between the 57 m and 28.5 m data based on polygon sampling,
- supervised ECHO results achieved with 28.5 m data were significantly ($\alpha = 0.003$) less than the performance of the 28.5 m data, using the per-point GML classifier based on pixel-by-pixel sampling,
- supervised ECHO results achieved with 28.5 m data were significantly ($\alpha = 0.0001$) less than the performance achieved with the per-point GML classifier using 57 m spatial resolution data, based on pixel-by-pixel sampling,
- none of the performance levels observed for polygon sampling were significantly ($\alpha > 0.05$) different,
- the mean percentage of correctly classified pixels using the CBD approach was significantly ($\alpha = 0.0001$) greater than the highest performance level achieved with either classifier, with either resolution data (GML per-point, 57 m data).

It is clear that the performance of the per-point GML classifier was not enhanced by the use of the 30 m as compared to the 90 m resolution data, at least for surface features of this particular scene. For scenes of surface features which are sufficiently large to provide a large number of non-edge pixels relative to edge pixels, an increase in the spatial resolution of the data with the use of a per-point GML approach is not expected to improve performance. The use of the supervised ECHO classifier provided lower performance, for both resolutions, over the per-point GML classifier. While the ECHO may provide a potential for exceeding the performance levels attained with the per-point GML classifier, the performance is highly sensitive to a number of parameters, training class statistics, and the distributional properties of the data.

The specific reasons for the poor performance of the supervised ECHO are not clear. The supervised

TABLE 4.  PERCENTAGE OF TEST PIXELS CORRECTLY CLASSIFIED FOR EACH INFORMATION CLASS IN EACH OF THE FOUR REPLICATIONS, USING 28.5 METER AND 57 METER DATA WITH A PER-POINT GAUSSIAN MAXIMUM LIKELIHOOD AND THE SUPERVISED ECHO CLASSIFIERS, BASED ON POLYGON SAMPLING

| | Per-point GML | | | | | | | |
| | 28.5 Meter | | | | 57 Meter | | | |
| Class | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|
| Water | 100.0 | 99.8 | 100.0 | 100.0 | 94.0 | 90.4 | 86.0 | 91.8 |
| Forest | 55.0 | 82.5 | 55.3 | 88.0 | 77.3 | 66.4 | 73.1 | 84.9 |
| Pasture | 44.2 | 49.3 | 48.5 | 45.5 | 37.6 | 40.9 | 66.2 | 60.4 |
| Crop lands | 29.2 | 34.8 | 43.5 | 71.0 | 66.2 | 72.0 | 60.4 | 70.9 |
| Developed | 75.5 | 72.7 | 66.3 | 85.3 | 78.4 | 60.2 | 56.0 | 39.5 |
| [a]$x._i$ | 60.8 | 67.8 | 62.7 | 78.0 | 70.7 | 66.0 | 68.3 | 69.5 |
| *$tx._i$ | 51.2 | 55.4 | 52.4 | 62.0 | 57.2 | 54.3 | 55.8 | 56.5 |

[b]$x.. = 67.3$      [b]$x.. = 68.2$
*$tx.. = 55.2$      *$tx.. = 55.9$

| | Supervised ECHO | | | | | | | |
| | 28.5 Meter | | | | 57 Meter | | | |
| Class | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|
| Water | 99.7 | 99.8 | 99.7 | 99.8 | 94.2 | 89.6 | 82.4 | 90.2 |
| Forest | 63.2 | 16.3 | 63.0 | 86.0 | 62.9 | 46.7 | 32.2 | 78.4 |
| Pasture | 28.8 | 30.3 | 30.7 | 70.0 | 50.9 | 30.2 | 77.8 | 62.9 |
| Crop lands | 22.5 | 70.0 | 49.3 | 56.2 | 60.0 | 69.6 | 27.6 | 59.6 |
| Developed | 87.8 | 66.0 | 95.0 | 32.0 | 67.6 | 57.3 | 54.2 | 46.9 |
| [a]$x._i$ | 60.4 | 56.5 | 67.5 | 68.8 | 67.1 | 58.7 | 54.8 | 67.6 |
| *$tx._i$ | 51.0 | 48.7 | 55.3 | 56.0 | 55.0 | 50.0 | 47.8 | 55.3 |

[b]$x.. = 63.3$      [b]$x.. = 62.1$
*$tx.. = 52.8$      *$tx.. = 52.0$

* Arcsin square root transformation ($\sin^{-1}\sqrt{PCC}$, where $0 \le PCC \le 1$).
[a] Mean computed over all categories within each replicate.
[b] Mean computed over all replicate means within each treatment.

ECHO performs the per-point GML discriminant for all pixels which fail the homogeneity test. In performing the test for homogeneity, the algorithm evaluates the local variance over the cell relative to the variance of that spectral class to which the cell is most similar. High variances in the training statistics will result in a higher frequency of cells passing the homogeneity test. Where these cells encompass more than one surface feature type, the error frequency will be increased by the increased frequency of homogeneous cells. A similar relationship prevails for the annexation test. A suggested basis for the observed performance is that a larger number of spectral classes was needed to adequately characterize the spectral properties of contiguous pixels of a single surface feature. The lower variances would improve the performance of the segmentation component of supervised ECHO.

A lower spatial resolution represents a segmentation of sorts—the undirected averaging of spatially adjacent pixel values. The results indicate that this mode of segmentation is superior to that employed in supervised ECHO, given the properties of the class statistics and those of the scene. Additionally, it is suggested that the discriminant employed in the supervised ECHO is not entirely appropriate. The discriminant employed is the exponentiated component of the GML discriminant function summed over all pixels in the field (see Kettig and Landgrebe, 1975, p. 32). This is essentially a weighted vote rule—evaluated independently for each pixel, and summed over all pixels of the field. The performance relative to that of the per-point GML approach using 57 m data suggests that a simple averaging of the pixel values over all pixels of the field would have improved the performance of ECHO. Tests could be devised and conducted to test these hypotheses and determine where the problems exist. However, the problems associated with approaches which require:

● data clustering,
● knowledge of appropriate cluster numbers,
● specification of critical thresholds, and
● concern over the properties of class statistics relative to the spectral properties of discrete fields or cells

TABLE 5. PERCENTAGE OF TEST PIXELS CORRECTLY CLASSIFIED FOR EACH INFORMATION CLASS IN EACH OF THE FOUR REPLICATIONS, USING 28.5 METER DATA WITH A CUMULATIVE BIN-WISE DIFFERENCE DISCRIMINANT

| Class | Training Polygons | Testing Polygons | 40 Pixel Minimum | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| Water | 4 | 6 | 93.0 | 94.2 | 97.3 | 97.5 |
| Forest | 10 | 40 | 99.0 | 99.3 | 99.9 | 99.9 |
| Pasture | 10 | 40 | 81.4 | 81.2 | 83.8 | 85.9 |
| Crop land | 10 | 40 | 70.7 | 71.3 | 65.5 | 62.7 |
| Developed | 10 | 40 | 99.2 | 69.1 | 95.7 | 93.7 |
| $^{a}x_{\cdot j}$ | | | 88.6 | 83.0 | 88.4 | 87.9 |
| $^{*}tx_{\cdot j}$ | | | 70.3 | 65.7 | 70.1 | 69.7 |

$$^{b}x_{\cdot\cdot} = 87.0$$
$$^{*}tx_{\cdot\cdot} = 68.9$$

| Class | Training Polygons | Testing Polygons | 90 Pixel Minimum | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| Water | 2 | 2 | 100.0 | 100.0 | 100.0 | 100.0 |
| Forest | 10 | 40 | 100.0 | 100.0 | 99.8 | 99.7 |
| Pasture | 10 | 40 | 91.9 | 90.7 | 89.9 | 92.9 |
| Crop land | 10 | 20 | 87.2 | 85.3 | 82.0 | 83.8 |
| Developed | 10 | 20 | 98.5 | 92.5 | 99.2 | 95.6 |
| $^{a}x_{\cdot j}$ | | | 95.5 | 93.7 | 94.2 | 94.4 |
| $^{*}tx_{\cdot j}$ | | | 77.8 | 75.5 | 76.1 | 76.3 |

$$^{b}x_{\cdot\cdot} = 94.5$$
$$^{*}tx_{\cdot\cdot} = 76.4$$

* Arcsin square root transformation ($\sin^{-1}\sqrt{PCC}$, where $0 \leq PCC \leq 1$).
[a] Mean computed over all categories within each replicate.
[b] Mean computed over all replicate means within each treatment.

are sufficient in magnitude and number to warrant pursuing generically distinct alternatives.

A large improvement over the GML per-point and supervised ECHO, based on use with 57 m and 28.5 m data, was achieved with a nonparametric, per-field CBD discriminant. This discriminant assumes that the image has previously been successfully segmented. Successful segmentation results in an image in which each segment contains no more than one surface feature type, and that each segment is sufficiently large to provide enough of the feature distribution to render the feature most similar to other distributions of the same feature type. Some of the other shortcomings of the CBD approach include:

- the requirement for a fairly large number of pixels,
- an expected sensitivity to small shifts in bin location for sparse histograms,
- failure to characterize the multispectral shape of the distributions (color analog), and
- disregard for the spatial properties of the spectral distribution (texture analog).

Immediate improvements to such indices as percent correctly classified pixels are attainable by attending

to the above shortcomings. All but the first of the above listed shortcomings can be readily mitigated to some degree by approximating the distribution function, for example, with spline functions (Guseman and Schumaker, 1984), and computing local texture measures. The CBD discriminant was used to evaluate empirically the relative identification and discrimination performance attainable were the image to be segmented into component features prior to the identification and discrimination procedure. The results obtained with the CBD discriminant demonstrate that successful scene segmentation provides a significant opportunity for improving the information extraction performance.

In spite of the performance levels observed for CBD and the readily available mechanisms for improving the approach, the fundamental aspects of the approach are regarded as incipient, if not misguided. The CBD, like nearly all other discriminants, employs a one-step evaluation and irrevocable identification of each feature. Additional evidence for the identification of a surface feature should be pursued according to the level of certainty provided by the evidence currently evaluated. Operating on segmented scenes provides the oppor-

tunity to define not only the statistical distribution of spectral values for each surface feature (as in the CBD), but also the texture, size, shape, and relative position of each feature. The potential for extracting detailed and accurate information from digital imagery once we begin to define and operate on complete surface features, in lieu of pixels, is exemplified by the performance of our own visual/cognitive system. Once the resolution and data quality provided by the sensor affords the use of complete features, or meaningful facets, the nature of information extraction strategies which can be employed is markedly different from the conventional per-point techniques currently used. With such algorithms we can begin to realize the capabilities provided by high spatial resolution data.

## ACKNOWLEDGMENTS

## REFERENCES

Guseman, L. F., and Schumaker, L. L., 1984. Multivariate spline methods in surface fitting: *Symposium on Mathematical Pattern Recognition & Image Analysis, June 6-8*, Houston, Texas, pp. 137-163.

Harnage, J., 1975. Landsat-D Thematic Mapper technical working group—final report: JSC-09797, 82 pp.

Irons, J. R., Markham, B. L., Nelson, R. F., Toll, D. L., Williams, D. L., Latty, R. S., Kennard, R. L., and Stauffer, M. L., 1984. The effects of sensor advancements on Thematic Mapper data classification: *Proceedings of the 18th International Symposium on Remote Sensing of Environment*.

Kan, E. P. F., 1973. The JSC clustering program ISOCLS and its applications: Lockheed Electronics Co., Houston Aerospace Systems Division, 49 pp.

Kan, E. P. F., and Ball, D. A., 1974. Data resolution versus forestry classification: JSC-09478, 22 pp.

Kan, E. P. F., Holley, W. A., and Parker, H. D., Jr., 1973. The JSC clustering program ISOCLS and its applications: *Machine Processing of Remotely Sensed Data Symposium*, pp. 4B36-4B50.

Kast, J. L., and Davis, B. J., 1977. Test of spectral/spatial classifier: LARS Contract Report 112877, 159 pp.

Kettig, R. L., 1975. Computer classification of remotely sensed multispectral image data by extraction and classification of homogeneous objects: Ph.D. dissertation, School of Electrical Engineering, Purdue University, West Lafayette, IN.

Kettig, R. L., and Landgrebe, D. A., 1973. Automatic boundary and sample classification of remotely sensed multispectral data: LARS Information Note 041773, 13 pp.

Kettig, R. L., and Landgrebe, D. A., 1975. Computer classification of remotely sensed multispectral image data by extraction and classification of homogeneous objects: LARS Information Note 050975, 181 pp.

Landgrebe, D. A., 1980. The development of a spectral/spatial classifier: *Pattern Recognition*, v. 12, no. 3, pp. 165-175.

Landgrebe, D. A., Beihl, L., and Simmons, W., 1977. An empirical study of scanner system parameters: *IEEE Transactions on Geoscience and Electronics*, v. GE-15, no. 3, pp. 120-130.

Latty, R. S., 1983. Scene segmentation based on spatial cues: *1983 Harvard Computer Graphics Conference*, Cambridge, MA, Vol. II.

———, 1984. Scene segmentation through region growing: *1984 Machine Processing of Remotely Sensed Data Symposium*, pp. 305-313.

Latty, R. S., and Hoffer, R. M., 1981. Computer-based classification accuracy due to data spatial resolution using a per-point vs per-field classification techniques: *1981 Machine Processing of Remotely Sensed Data Symposium*, pp. 384-393.

Lyon, J. C., Fischel, D., and Beyer, E., 1983. Production and analyses of output data products for Landsat-4 in the engineering check-out phase: *Space Applications at the Crossroads*, 21st Goddard Memorial Symposium, (ed. J. H. McElroy and E. L. Heacock), Univelt, Inc., San Diego, CA, pp. 51-66.

Markham, B., 1984. Preliminary comparisons of the information content and utility of TM vs MSS data: *Landsat-4 Science Investigations Summary*, NASA Conference Publication 2326, pp. 135-136.

Markham, B. L., and Townshend, J. R. G., 1981. Land cover classification accuracy as a function of sensor spatial resolution: *Proceedings of the 15th International Symposium on Remote Sensing of Environment*, pp. 1075-1090.

Minter, R. T., 1972. Computer program documentation: ISOCLS—Iterative Self-Organizing Clustering Program: Lockheed Electronics Company, Houston Aerospace Division, Contract NAS 9-12200, 68 pp.

Morgenstern, J. P., Nalepka, R. F., and Erickson, J. D., 1977. Investigation of Thematic Mapper spatial, radiometric, and spectral resolution: *Proceedings of the 11th International Symposium on Remote Sensing of Environment*, pp. 693-701.

Neter J., and Wasserman, W., 1974. *Applied Linear Statistical Models:* Richard D. Irwin, Inc., Homeward, IL, 842 pp.

Phillips, T. L., 1973. *LARSYS Version 3 User's Manual:* Purdue University, Laboratory for Applications of Remote Sensing.

Sadowski, F., and Sarno, J., 1976. Forest classification accuracy as influenced by multispectral scanner spatial resolution: NASA Contract No. NAS9-14123, 130 pp.

Salomonson, V. V., 1978. Landsat-D systems overview: *Proceedings of the 12th International Symposium on Remote Sensing of Environment*, pp. 371-385.

Swain, P. H., and Davis, S. M., (eds.), 1978. *Remote*

*Sensing: The Quantitative Approach:* McGraw-Hill International Book Company, NY, 396 pp.

Thomson, F. J., and Erickson, J. D., 1976. A thematic mapper performance optimization study: *Proceedings of the 10th International Symposium on Remote Sensing of Environment,* pp. 85-98.

Wacker, A. G., 1971. The minimum distance approach to classification: Ph.D. dissertation, Dept. of Electrical Engineering, Purdue University, West Lafayette, IN., 345 pp.

Wacker, A. G., and Landgrebe, D. A., 1972. Minimum distance classification in remote sensing: LARS Print 030772, 25 pp.

Wiersma, D. J., and Landgrebe, D. A., 1976. The use of spatial characteristics for the improvement of multi-spectral classification of remotely sensed data: *1976 Machine Processing of Remotely Sensed Data Symposium,* pp. 2A18-2A25.

Williams, D. L, Irons, J. R., Markham, B. L., Nelson, R. F., Toll, D. L., Latty, R. S., and Stauffer, M. L., 1984. A statistical evaluation of the advantages of Landsat thematic mapper data in comparison to multispectral scanner data: *IEEE Transactions on Geoscience and Remote Sensing,* v. GE-22, no. 3, pp. 294-302.

---

# Symposium and Workshop on Remote Sensing Techniques and Applications in the Marine Environment

Sponsored by: Hydrographic Society of America, local chapter
ASPRS Mid-South Region, Gulf Coast Chapter
NASA/National Space Technology Laboratories

With the participation of NORDA and NAVOCEANO

Dates: October 30 and 31, 1985

Location: Univ. of Southern Mississippi Gulf Park Campus

Technical Sessions Followed by a Tour of NASA/NSTL

Send abstract including title, author(s), affiliations, addresses, phone number to:

Ramona Pelletier-Travis
Acting President, Gulf Coast Chapter
ASPRS Mid-South Region
Bldg 1100
NSTL, Mississippi 39529