1-1-1979

# Directed Canonical Analysis and the Performance of Classifiers under Its Associated Linear Transformations

Benjamin F. Merembeck

Brian J. Turner

Reprinted from

# Symposium on Machine Processing of Remotely Sensed Data

**June 27 - 29, 1979**

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana 47907 USA

IEEE Catalog No.
79CH1430-8 MPRSD

# DIRECTED CANONICAL ANALYSIS AND THE PERFORMANCE OF CLASSIFIERS UNDER ITS ASSOCIATED LINEAR TRANSFORMATIONS

BENJAMIN F. MEREMBECK, BRIAN J. TURNER
The Pennsylvania State University

## I. ABSTRACT

Directed canonical analysis is presented as an extension of the general form of canonical analysis, which is a method for reducing the dimensionality of multivariate data sets with minimum loss of discriminatory variance. The reduction takes the form of a linear transformation, $\underline{y} = C\underline{x}$, that condenses the discriminatory variance onto a relatively few, high-variance, orthogonal discriminant axes.

Canonical analysis is developed as an analog to the one-way MANOVA. The directed extension allows user-specified contrasts to define linear relationships that are known or suspected to exist within the data. The linear transformation, C, is defined by means of the symmetric canonical form of the matrix eigenproblem.

Canonical and principal components transformations and various distance classifiers were applied to 3 representative remotely sensed MSS data sets. Results indicate that use of a piecewise maximum likelihood classifier with the directed canonical discriminant axes will give the best overall combination of classification accuracy and computational efficiency if adequate sample sizes are available to estimate category statistics. For small sample sizes, piecewise Euclidean distance is recommended, which, in canonically transformed space, is equivalent to the Mahalanobis classifier.

## II. INTRODUCTION

The trend in remote sensing analysis technology is toward merged data sets of large dimensionality. The variables in the data sets may include single- and multi-date multispectral scanner (MSS) data, digitized aerial photography, environmental and topographic data, geophysical data, etc. Such large dimensionality presents machine processing problems in terms of storage, analysis, and data display. In the near future, the advent of the 7-channel LANDSAT thematic mapper (TM) with 30 m ground resolution, which is to supersede at least partially the 4-channel LANDSAT MSS with 80 m resolution, will further aggravate the machine processing problems.

The technique of canonical analysis (Seal 1964) has been used for some time to alleviate the problems associated with large dimensionality. In canonical analysis, the multivariate statistics associated with categories of interest within the data are used to find an orthogonal linear transformation, C, of the form

$$\underline{y} = C\underline{x} \qquad (1)$$

that defines a relatively small number of statistically independent canonical discriminant axes that explain the preponderance of the discriminatory variance associated with the categories. It is not uncommon to have in excess of 95% of the discriminatory variance explained on the first 2 or 3 axes regardless of the original number of variables (Gnanadesikan 1977).

## III. A DEFINITION OF CANONICAL ANALYSIS

It is convenient to define canonical analysis in terms of the multivariate analysis of variance (MANOVA) application of the general multivariate linear model. In MANOVA, a null hypothesis is set forth as

$$H_0: Q\mu = 0 \quad \text{versus} \quad H_1: Q\mu \neq 0 \qquad (2)$$

The matrix $\mu$ is an $h \times p$ matrix whose rows are the category mean vectors, where h is the number of categories and p is the original number of variables. In practice, $\mu$ is not known and must be estimated by the matrix M, whose rows are the category sample mean vectors. Q is a $q \times h$ hypothesis matrix, where $q \leq (h-1)$. The limitation on the number of rows of Q is set because only h-1 linearly independent combinations of the h categories can be defined and a maximum of h-1 axes are required to discriminate between the h categories (Scheffé 1959, Green and Carroll 1976). Each row of Q is a contrast that, by definition, must sum to 0 and defines a hypothetical linear relationship among the categories. Matrix 0 is a $q \times p$ null matrix.



CH1430-8/79/0000-0314$00.75 © 1979 IEEE

A. GENERAL CANONICAL ANALYSIS

The null hypothesis implicit in the general canonical analysis is given by

$$H_0\colon Q\mu = 0 \rightarrow \underline{\mu}_1' = \underline{\mu}_2' = \ldots = \underline{\mu}_h'$$

versus (3)

$$H_1\colon \text{at least 2 of the means differ}$$

where $\underline{\mu}_i'$, $i = 1, \ldots, h$, is the ith row of $\mu$. Therefore, the hypothesis matrix in the general canonical analysis is implicitly an $(h-1)\times h$ matrix of simple one-way contrasts over which there is no control. One of the many forms that the matrix of one-way contrasts can take is

$$Q = \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & -1 & \ldots & 0 & 0 \\ & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix} \tag{4}$$

The canonical axes associated with this hypothesis will attempt to separate the h categories on as few axes as possible.

B. DIRECTED CANONICAL ANALYSIS

The general canonical analysis has recently been extended to allow user-specified contrasts that define underlying relationships among categories or natural groupings of categories known a priori to exist within the data. The extension, called directed canonical analysis, was originally proposed by Dr. F. Yates Borden shortly before his death in 1977. It affords more control over the transformation than does the general form in that the directed canonical axes reflect relationships and groupings defined by the contrasts.

In directed canonical analysis, the rows of Q can be explicitly defined as contrasts that reflect known or suspected relationships among the categories. For example, consider the following 6 categories:

| Category | Known Type of Category |
|---|---|
| 1 | Forest vegetation with NW aspect |
| 2 | Forest vegetation with SE aspect |
| 3 | Non-vegetated land with NW aspect |
| 4 | Non-vegetated land with SE aspect |
| 5 | Lake water |
| 6 | River water |

These categories have a structure that can be investigated by proposing the orthogonal contrasts shown in Table 1. The Q matrix for the composite hypothesis in (2) is defined in Table 1.

Orthogonal contrasts, such as those in Table 1, are useful in hypothesis testing and data display. In hypothesis testing, the orthogonal structure leads to unambiguous, single degree of freedom tests for examining each orthogonal contrast independently of the others. For data display, the orthogonal structure implies that a discriminant axis can be found for each contrast that is independent of the other contrasts. The axes would then be used to photo-interpret a data set for a specific characteristic or set of characteristics. In pattern recognition procedures, such as classification, the orthogonal contrasts do not appear to have any particular advantage over non-orthogonal contrasts in defining the desired linear relationships among categories. A more detailed discussion of orthogonal contrasts can be found in Merembeck and Borden (1978).

Table 1. Orthogonal contrasts among categories (from Merembeck and Borden 1978).

| | Categories | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Coefficients | | | | | |
| Water vs non-water | -1 | -1 | -1 | -1 | +2 | +2 |
| Lake vs river water | 0 | 0 | 0 | 0 | +1 | -1 |
| Veg. vs non-veg. | +1 | +1 | -1 | -1 | 0 | 0 |
| NW vs SE aspect | +1 | -1 | +1 | -1 | 0 | 0 |
| Veg. vs non-veg.× NW vs SE aspect | +1 | -1 | -1 | +1 | 0 | 0 |

C. CONVERTING THE PROBLEM TO CANONICAL FORM

To test the null hypothesis, the total sample covariance matrix of the data, S, is partitioned into an among-categories sample covariance matrix, A, and a within-categories sample covariance matrix, W, that is assumed to be equal for, and common to, all categories. The relation is

$$S^* = A^* + W^* \tag{5}$$

where

$S^*$ = (total observations - 1)S,
$A^*$ = (among-categories degrees of freedom)A, and
$W^*$ = (within-categories degrees of freedom)W.

Matrices $S^*$, $A^*$, and $W^*$ are, respectively, the total, among-categories, and within-categories adjusted sums of squares and cross-products matrices.

W is found by

$$W = \left[\sum_{i=1}^{h} n_i - h\right]^{-1} \left[\sum_{i=1}^{h} (n_i - 1)S_i\right] \tag{6}$$

where $n_i$ is the number of observations for the ith category. $S_i$ is the ith category covariance matrix and is found by the conventional sums of squares and cross-products formulation,

$$S_i = (n_i - 1)^{-1} \sum_{i=1}^{n_i} (\underline{x}_i - \underline{m}_i)(\underline{x}_i - \underline{m}_i)'$$

where $\underline{m}_i'$ is the ith row of M.

The among-categories sample covariance matrix, A, is found from the general formulation adapted from Morrison (1976),

$$A = [M' Q'(Q N^{-1} Q')^{-1} Q M]/q \qquad (7)$$

where N is an h×h diagonal matrix with the numbers of observations for the categories in the diagonal positions. In the case where Q is a matrix of simple one-way contrasts associated with the general canonical analysis, (7) reduces to the sums of squares and cross-products form conventionally used to find A.

The null hypothesis (2) is tested by means of an "F" ratio that has several forms, one of which is

$$F = \frac{\underline{C}_i' A \underline{C}_i}{\underline{C}_i' W \underline{C}_i} \qquad (8)$$

Vector $\underline{C}_i'$ is the ith row of the linear transformation that will most strongly reject the null hypothesis. The solution for (8) is found by placing the matrices A and W into the canonical form defined by the characteristic equation

$$|A - dW| = 0 \qquad (9)$$

where d is the Lagrange multiplier. Equation (9) can be rewritten in two forms:

$$|W^{-1} A - dI| = 0 \qquad (10)$$

and

$$|W^{-\frac{1}{2}} A W^{-\frac{1}{2}/W} - dI| = 0 \qquad (11)$$

Both forms give the same eigenvalues but different eigenvectors. Since the MANOVA problem only requires the eigenvalues, the form in (10) is used in most MANOVA procedures. However, the use of the eigenvectors from the non-symmetric form of (10) will not yield the spherical, unit variance, within-categories transformed covariance matrix sought in canonical analysis. The normalized row eigenvectors from the symmetric form of (11) will spherize the within-categories covariance matrix (Green and Carroll 1976) and form the canonical linear transformation, C, such that

$$C A C' = D \qquad (12)$$

and

$$C W C' = I \qquad (13)$$

where D is the q×q diagonal matrix of eigenvalues of (9) and is the transformed among-categories sample covariance matrix and I is the q×q identity matrix.

The rows of C and the associated diagonal elements of D are ordered such that the transformation preserves as much of the among-categories discriminatory variance on the first canonical, or discriminant, axis as can be preserved on any 1 axis. The second canonical axis preserves as much of the remaining discriminatory variance that is orthogonal to, or uncorrelated with, the first axis as can be preserved on any 1 axis. The third canonical axis preserves as much of the remaining discriminatory variance that is orthogonal to the first 2 axes, and so on.

Seal (1964) gives an indirect method for solving (11) that was implemented in The Pennsylvania State University's Office for Remote Sensing of Earth Resources (ORSER) digital processing system (Turner et al. 1978) by Lachowski (1973). The program, CANAL, has since been extended to accommodate the directed canonical analysis.

## IV. CLASSIFIER PERFORMANCE UNDER TRANSFORMATION

A study was undertaken to investigate the effects of general and directed canonical analyses on thematic classification of remotely sensed MSS data. The techniques were assessed relative to each other and relative to classifications based on the original raw data and the orthogonal axes generated by principal components analysis.

In principal components analysis, the normalized row eigenvectors associated with the eigenvalues of the total covariance matrix, S, define an orthogonal linear transformation, C, such that

$$CSC' = D \qquad (14)$$

while

$$CC' = I \qquad (15)$$

The principal components transformation has many of the data reduction qualities of the canonical transformations but will not necessarily reflect the desired thematic relationships.

The primary classification criterion was that of maximum likelihood under the assumption of multivariate normality. The metric for this criterion, when used as a distance classifier, is

$$d_i^2 = \ln|S_i| + (\underline{x} - \underline{m}_i)' S_i^{-1} (\underline{x} - \underline{m}_i) \qquad (16)$$

(Gnanadesikan 1977).

The vector $\underline{x}$ is a p-variate raw data vector to be classified and is assumed to be distributed $MVN(\underline{\mu}_i, \Sigma_i)$. This classifier accounts for differences in the shape and orientation of the dispersions and for differences in total variance between categories.

Under the linear transformation $\underline{y} = C\underline{x}$, $\underline{y}$ is a q-variate ($q \leq p$) vector distributed $MVN(C \underline{\mu}_i, C \Sigma_i C')$ (Anderson 1958). In transformed space then, (16) becomes

$$d_i^2 = \ln|CS_iC'| + (\underline{y}-C\underline{m}_i)'(CS_iC')^{-1}(\underline{y}-C\underline{m}_i) \qquad (17)$$

Since the transformations do not specifically consider individual category dispersions, the transformed category covariance matrices, $CS_iC'$, i = 1, ..., h, will not necessarily be diagonal matrices in transformed space.

Since the category mean vectors and covariance matrices have to be transformed only once, the major computational cost in applying (17) is that of transforming the individual observations. It was hoped that only the first few canonical and principal components axes would be required to obtain, at a reduced cost, classification results equal to or better than results obtained by classifying raw data. Superior classification results in transformed space might be expected in the case of noisy data. The linear transformations have proven to be effective random noise filters (Merembeck et al. 1976).

To make the results more global, a number of other distance classifiers were included in the study:

1. The elliptical classifier, defined as

$$d_i^2 = \ln|S_i| + (\underline{x}-\underline{m}_i)' D_i^{-1} (\underline{x}-\underline{m}_i) \tag{18}$$

where $D_i$ is a diagonal matrix containing the diagonal entries of $S_i$, requires considerably fewer computations and less storage than does (16). The elliptical classifier does not account for differences in the spatial orientation of the category dispersions in that it assumes that the axes of the dispersions are parallel to the axes of the measurement space.

2. The Mahalanobis distance equivalent to the Wald-Anderson linear discriminant (Morrison 1976) has the form

$$d_i^2 = (\underline{x}-\underline{m}_i)' W^{-1} (\underline{x}-\underline{m}_i) \tag{19}$$

and assumes that the category covariance matrices are common to and equal for all categories; the same assumption made for canonical analysis. Therefore, by (13), this classifier reduces to simple Euclidean distance under the canonical linear transformation.

3. A classifier that is a compromise between maximum likelihood and the Mahalanobis classifier uses the pooled covariance of a family of categories whose probability dispersions are approximately proportional as determined by Mauchly's sphericity criterion as modified by Davis (1971). The classifier has the same form as (19) but uses the pooled matrix for the proportional family rather than the within-categories covariance matrix pooled over all categories.

4. The minimum Euclidean distance is of the form

$$d_i^2 = (\underline{x}-\underline{m}_i)' (\underline{x}-\underline{m}_i) \tag{20}$$

This is the fastest of the distance classifiers and requires the least storage.

## V. PROCEDURE

The classifiers were applied to a series of representative data sets consisting of from 4 to 22 original variables. The raw data and principal components and canonical transformations of the data were evaluated for each data set. The data consisted of 4-channel LANDSAT data, 8-channel multi-temporal LANDSAT data, and 22-channel SKYLAB data.

The measure of performance used for the comparisons was classification accuracy. Estimates of classification accuracy were determined both for classifications based on individual categories and for piecewise classifications. In piecewise classification, a number of individual categories are assigned the same mapping symbol or color. Classification of an element into any category having the same mapping symbol or color as the category to which the element belongs is considered to be correct classification in the piecewise procedure.

The contrasts for the directed canonical analyses of the test data sets were constructed to reflect the category groupings associated with piecewise classification. The directed canonical axes in this case attempt to separate the category groupings rather than the individual categories.

Two different estimates of classification accuracy were made:

1. Monte Carlo techniques were used to generate synthetic data that followed probability distributions defined by the categories in the observed data (Boullion et al. 1975). The synthetic data in both raw and linearly transformed form were presented to the classifiers and classification accuracy determined.

2. The observed data were resampled using a technique attributed to Lachenbruch (1967) that has been shown to give approximately unbiased estimates of classification accuracy. Lachenbruch's procedure, also known as a jackknife procedure, involves removing the effects of the element currently being classified from the sample mean and covariance matrix of the population to which the element belongs. The resampled data were presented to the adjusted classifiers in both raw and linearly transformed form.

Jensen's (1972) modification of Hotelling's $t^2$ was used as a global test for equality of the maximum likelihood performance profiles among the various linear transformations over the first few high-variance axes. The same differences between individual pairs of transformations were tested using Bonferroni multivariate multiple comparisons (Morrison 1976). The significance level of the tests was set at 0.05. Differences among classifiers were investigated by the same methods.

Further analyses were performed by generating comparisons of thematic maps that had been produced by using the maximum likelihood classifier on sets of the original raw data and on various combinations of transformed axes. The comparison maps were used to determine where, spatially, the differences in classification accuracy that were observed in the previous analysis were occurring. During this processing, the approximate number of machine cycles required to produce each of the thematic maps was determined. These results allowed a relative measure of cost versus performance to be determined.

## VI. RESULTS

In processing the data, it became evident that the Lachenbruch resampling method gave considerably more conservative error estimates than did the Monte Carlo techniques. Therefore, only the results obtained from the Lachenbruch resampling method are presented.

In evaluating the results after processing the data, it was found that graphing the performance of the classifiers for a given data set and transformation yielded plots that had virtually identical shapes for all classifiers. The plots varied only in height on the y (percentage error) axis. Therefore, only the graphs for the performance of the maximum likelihood classifier are presented with the results for each data set.

### A. FOUR-CHANNEL LANDSAT DATA

The 4-channel LANDSAT data were from scene 1028-15295, collected on August 20, 1972. The data set had been used previously in a small-area mapping study and its characteristics were well known (Merembeck 1978). The data set represents a predominantly forested area around the East Branch reservoir on the Clarion River in Elk County in northwestern Pennsylvania. A land-cover thematic map of the area had been produced using 16 categories and 9 mapping symbols.

Figure 1 is a graph of the performance of the maximum likelihood classifier operating both on individual categories and as a piecewise classifier. The directed canonical and principal components axes gave marginally better performance than did the general canonical axes. For all 3 transformations, only the first 2 axes, accounting for in excess of 99% of explained variance, were required for optimum classification.

Table 2 gives estimated error rates over all classifiers for the first 2 axes of the various transformations and for the 4 channels of raw data. There were no statistically significant differences at the 0.05 level among the performance profiles of the transformations over the first 3 axes or among classifiers operating by category on the first 2 directed canonical axes.

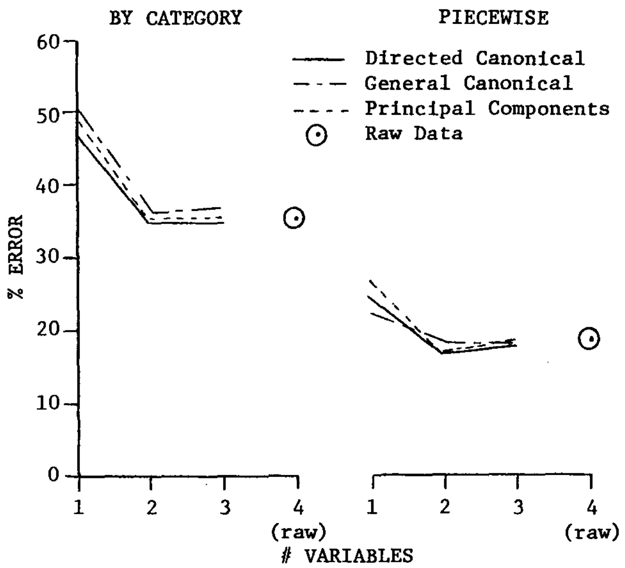


Figure 1. Maximum likelihood profiles for the 4-channel LANDSAT data set.

Table 2. Misclassification percentages for the East Branch data when using Lachenbruch's resampling method.

| | Directed Canonical | | General Canonical | |
|---|---|---|---|---|
| # Variables | 2 | | 2 | |
| % Variance | 99.39 | | 99.32 | |
| % Error | Cat. | PW | Cat. | PW |
| Max. likelihood | 34.8 | 17.2 | 35.7 | 17.8 |
| Elliptical | 35.4 | 17.5 | 36.6 | 18.4 |
| Pooled | 36.9 | 21.7 | 38.1 | 23.3 |
| Mahalanobis | 37.8 | 21.1 | 38.9 | 21.9 |
| Euclidean dist. | 37.8 | 21.1 | 38.9 | 21.8 |
| | **Principal Components** | | **Raw Data** | |
| # Variables | 2 | | 4 | |
| % Variance | 99.61 | | 100 | |
| % Error | Cat. | PW | Cat. | PW |
| Max. likelihood | 34.9 | 17.2 | 34.9 | 17.6 |
| Elliptical | 36.9 | 18.1 | 35.8 | 20.9 |
| Pooled | 37.4 | 21.9 | 38.4 | 23.4 |
| Mahalanobis | 36.6 | 21.9 | 38.2 | 21.2 |
| Euclidean dist. | 36.5 | 21.7 | 40.2 | 22.3 |

A comparison between the maximum likelihood results for the 4-channel raw data and for the first 2 directed canonical axes was generated for a block of 10,465 elements. The maps differed on 9,009, or 13.9%, of the elements. A visual comparison of the classification maps favored the transformed map; it exhibited better symbol grouping in classifying a number of the small openings in the forest canopy that had been the object of the original investigation using this data set (Merembeck 1978).

The classification loop required approximately 153,029,008 machine cycles to classify the 4-channel raw data. In transformed space, 73,115,504 machine cycles were required to transform and classify the data using the first 2 directed canonical axes; a 2.093 times reduction in favor of the transformation.

B. EIGHT-CHANNEL MULTI-TEMPORAL LANDSAT DATA

In this analysis, data from LANDSAT scenes 1583-15100, collected on February 26, 1974, and 1403-15134, collected on August 30, 1974, were geometrically corrected, registered, and merged onto an 8-channel multi-temporal data set. The data set had been used previously to map southern pine plantations on the Atlantic Coastal Plain in North Carolina near Albemarle Sound (Williams and Haver 1976). A land-cover map had been produced using 10 categories and 7 mapping symbols.

Figure 2 is a graph of the performance of the maximum likelihood classifier operating on individual categories. The shapes of the plots for piecewise maximum likelihood were virtually identical. By the fourth axis, results for the transformations had essentially duplicated those of the raw data. The canonical axes gave marginally better performance on the fourth axis than did principal components.

Table 3 gives estimated error rates over all classifiers using the first 4 axes of the various transformations and for the 8-channel raw data. There were no statistically significant differences at the 0.05 level among the performance profiles of the transformations over axes 2-4 or among classifiers operating by category on the first 4 directed canonical axes.

A comparison between the maximum likelihood results for the 8-channel raw data and for the first 4 directed canonical axes was generated for a block of 5,916 elements. The maps differed on 782, or 13.2%, of the elements. As with the 4-channel data, a visual comparison favored the transformed map due to more consistent symbol grouping.

The classification loop required approximately 110,283,888 machine cycles to classify the 8-channel raw data, whereas 58,801,984 machine cycles were required to transform and classify the data using the first 4 directed canonical axes; a 1.876 times reduction in favor of the transformation.

C. TWENTY-TWO CHANNEL SKYLAB DATA

The data set for the 22-channel SKYLAB S-192 scanner was a block of line-straightened data from a file entitled "SL3, Orbit 14, 5 August 1975, Tape 933847." The data represented an area around the town of Freeport on the Allegheny River in southwestern Pennsylvania. The data had been used previously for sensor evaluation and stripmine mapping (Barr 1978). A land-cover map of the area had been produced using 14 categories and 5 mapping symbols.

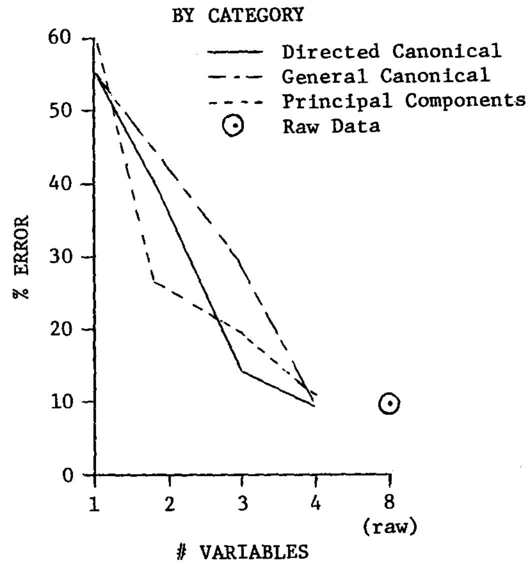


Figure 2. Maximum likelihood profiles for the 8-channel multitemporal LANDSAT data set.

Table 3. Misclassification percentages for Williams and Haver's (1976) data when using Lachenbruch's resampling method.

| | Directed Canonical | | General Canonical | |
|---|---|---|---|---|
| # Variables | 4 | | 4 | |
| % Variance | 99.69 | | 99.56 | |
| % Error | Cat. | PW | Cat. | PW |
| Max. likelihood | 9.7 | 6.7 | 9.6 | 6.6 |
| Elliptical | 14.6 | 10.0 | 13.7 | 9.3 |
| Pooled | 11.3 | 7.7 | 11.2 | 7.6 |
| Mahalanobis | 12.7 | 8.7 | 12.8 | 8.8 |
| Euclidean | 12.6 | 8.6 | 12.8 | 8.8 |
| | **Principal Components** | | **Raw Data** | |
| # Variables | 4 | | 8 | |
| % Variance | 97.04 | | 100 | |
| % Error | Cat. | PW | Cat. | PW |
| Max. likelihood | 10.7 | 7.5 | 9.7 | 6.6 |
| Elliptical | 12.5 | 7.8 | 10.4 | 6.9 |
| Pooled | 12.6 | 8.2 | 11.3 | 7.5 |
| Mahalanobis | 13.0 | 8.8 | 13.4 | 9.2 |
| Euclidean dist. | 15.2 | 10.6 | 15.2 | 10.6 |

Figure 3 is a graph of the performance of the maximum likelihood classifier operating both on individual categories and as a piecewise classifier. Both directed and general canonical analyses achieved performance on 3 axes equal to that of the 22-channel raw data. In classification by category, the general canonical analysis gave improved results over the raw data when using 4 axes. However, in piecewise classification, there was no advantage in going beyond 3 axes.

Table 4 gives estimated error rates over all classifiers for the first 3 axes of the directed canonical analysis, the first 4 axes of general canonical and principal components analyses, and for the 22 channels of raw data. The good performance of the Mahalanobis classifier relative to that of maximum likelihood indicates possible problems with sample size. Some of the urban and stripmine samples are quite small with 1 of the stripmine categories having only 30 observations.

Bonferroni multiple comparisons of the maximum likelihood performance profiles over transformed axes 2-4 indicate that the principal components axes gave significantly poorer performance than did either canonical transformation at the 0.05 level. There were no statistically significant differences at the 0.05 level among classifiers operating on the first 3 directed canonical axes.

The maximum likelihood results for the 22-channel raw data and for the first 3 directed canonical axes were compared by generating maps for a block of 3,696 elements. The maps differed on 1,644, or 44.5%, of the elements. Part of the discrepancy was due to unclassified elements. The program parameters had been set so that any element falling outside of the 95% confidence hyperellipsoid for a category could not be assigned to that category. A total of 1,071, or 29%, of the elements for the classification run on raw data fell outside of all of the 95% confidence hyperellipsoids and were not classified. This is probably a manifestation of the sample size problem noted earlier. Only 378, or 10%, of the elements in the transformed run were not classified.

A visual comparison of the classification maps clearly favored the transformed classification; symbol groupings, particularly in the urban and stripmine areas, were much more consistent than those in the classification map generated with raw data.

The classification loop required approximately 473,933,312 machine cycles to classify the 22-channel raw data. In transformed space, 47,665,984 machine cycles were required to transform and classify the data on the first 3 directed canonical axes; a 9.94 times reduction in favor of the transformation.

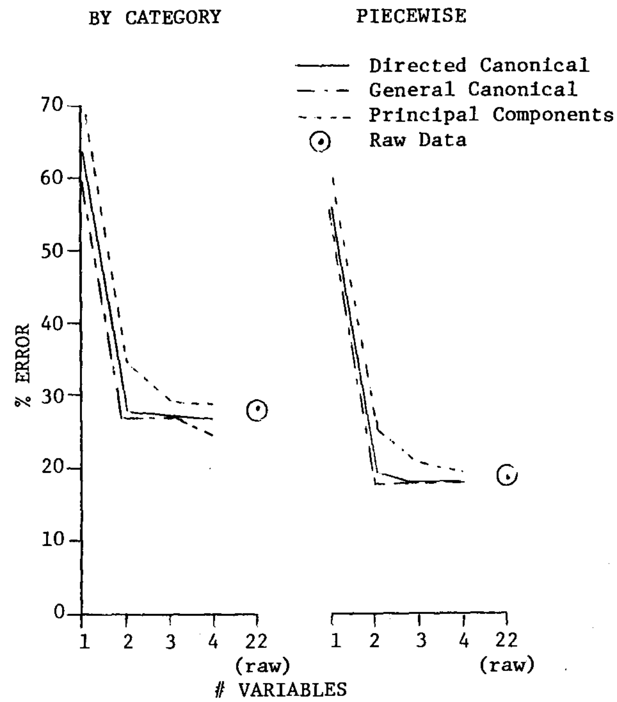


Figure 3. Maximum likelihood profiles for the 22-channel SKYLAB data set.

Table 4. Misclassification percentages for SKYLAB S-192 data using Lachenbruch's resampling method.

| | Directed Canonical | | General Canonical | |
|---|---|---|---|---|
| # Variables | 3 | | 4 | |
| % Variance | 99.25 | | 96.44 | |
| % Error | Cat. | PW | Cat. | PW |
| Max. likelihood | 26.7 | 18.1 | 24.8 | 18.1 |
| Elliptical | 27.8 | 19.6 | 25.2 | 18.3 |
| Pooled | 27.0 | 18.8 | 24.8 | 17.8 |
| Mahalanobis | 27.7 | 18.3 | 25.2 | 17.3 |
| Euclidean dist. | 27.6 | 18.2 | 25.1 | 17.3 |
| | **Principal Components** | | **Raw Data** | |
| # Variables | 4 | | 22 | |
| % Variance | 81.18 | | 100 | |
| % Error | Cat. | PW | Cat. | PW |
| Max. likelihood | 27.9 | 18.7 | 27.2 | 17.6 |
| Elliptical | 29.3 | 19.9 | 28.7 | 20.8 |
| Pooled | 28.2 | 19.0 | 31.1 | 20.2 |
| Mahalanobis | 29.6 | 18.9 | 24.7 | 16.6 |
| Euclidean dist. | 29.7 | 19.2 | 28.5 | 18.4 |

## VII. DISCUSSION

There were differences among the data sets as to the number of axes required for optimum classification, the estimated error rates, and the performance of the classifiers relative to each other. Some general observations, however, are possible.

All of the classifiers gave similar results; no statistically significant differences among the performances of the various classifiers were found. The results do suggest, however, that maximum likelihood will give somewhat better results than the others when adequate sample sizes are available. In the case of small sample sizes, the Mahalanobis classifier would seem to be a better choice.

Piecewise classification clearly improved the results for all combinations of classifiers and linear transformations. The diversity of spectral responses that can reasonably be grouped into the same cover type precludes mapping that type with just one set of category statistics. For that reason, virtually all land-cover maps generated at ORSER from MSS data have used the piecewise approach.

For the transformations, results favor the canonical analyses over principal components as dimensionality increases. The differences between the directed and general canonical analyses were quite small, and statistically non-significant, but seem to marginally favor the directed procedure. Another advantage to the directed procedure is that the contrasts can be defined to complement piecewise classification.

From the study results and previous experience with canonical analysis, a good rule of thumb for selecting axes appears to be: select the first $q \leq p$ axes such that the axes explain in excess of 95% of the discriminatory variance and none of the remaining axes explain more than 1% of the variance. This approach will typically at least halve the original number of variables while giving equivalent or improved classification and increased computational efficiency relative to the results from raw data.

A synthesis of the results for classifiers and for transformations indicates that use of a piecewise maximum likelihood classifier on the directed canonical axes will give the best overall combination of classification accuracy and computational efficiency if adequate sample sizes are available. If sample sizes are small, the use of piecewise Euclidean distance on the directed canonical axes is recommended since Euclidean distance is equivalent to the Mahalanobis classifier in canonically transformed space.

## VIII. LITERATURE CITED


Anderson, T. W. 1958. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York, N. Y.

Barr, D. M. 1978. Channel evaluation of SKYLAB's S-192 data. M.S. thesis, The Pennsylvania State Univ., University Park, Pa.

Boullion, T. L., P. L. Odell, and B. S. Duran. i1975. Estimating the probability of misclassification and variate selection. Pattern Recognition 7: 139-145.

Davis, A. W. 1971. Percentile approximations for a class of likelihood ratio criteria. Biometrika 58(2): 349-355.

Gnanadesikan, R. 1977. Methods for Statistical Data Analysis of Multivariate Observations. John Wiley & Sons, Inc., New York, N. Y.

Green, P. E. and J. D. Carroll. 1976. Mathematical Tools for Applied Multivariate Analysis. Academic Press, New York, N. Y.

Jensen, D. R. 1972. Some simultaneous multivariate procedures using Hotelling's $t^2$ statistics. Biometrics 28: 39-53.

Lachenbruch, P. A. 1967. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics 23: 639-645.

Lachowski, H. M. 1973. Canonical analysis applied to interpretation of multispectral scanner data. M.S. thesis, The Pennsylvania State Univ., University Park, Pa.

Merembeck, B. F. 1978. Small-area mapping and spectral signature extension in Pennsylvania hardwood forests using LANDSAT-I multispectral scanner data. M.S. thesis, The Pennsylvania State Univ., University Park, Pa.

________ and F. Y. Borden. 1978. Principal Components and Canonical Analysis for Reduction of the Dimensionality of Large Data Sets. ORSER Tech. Rept. 5-78, Office for Remote Sensing of Earth Resources, The Pennsylvania State Univ., University Park, Pa.

________, F. Y. Borden, M. H. Podwysocki, and D. N. Applegate. 1976. Application of canonical analysis to multispectral scanner data. Proc. 14th Annual Symp. on the Application of Computer Methods in the Mineral Industries, Oct. 4-8, University Park, Pa. p. 867-879.

Morrison, D. F. 1976. Multivariate Statistical Methods, 2d Ed. McGraw-Hill Book Co., New York, N. Y.

Scheffé, H. 1959. The Analysis of Variance. John Wiley & Sons, Inc., New York, N. Y.

Seal, H. L. 1964. Multivariate Statistical Analysis for Biologists. Methuen and Co., London, Eng.

Turner, B. J., D. N. Applegate, and B. F. Merembeck. 1978. Satellite and Aircraft Multispectral Digital Data User Manual. ORSER Tech. Rept. 9-78, Office for Remote Sensing of Earth Resources, The Pennsylvania State Univ., University Park, Pa.

Williams, D. L. and G. F. Haver. 1976. Forest Land Management by Satellite. Intralab Proj. No. 75-1, NASA, Goddard Space Flight Center, Greenbelt, Md.

Mr. Merembeck is a doctoral candidate in the School of Forest Resources at The Pennsylvania State University. He has been involved in the development and application of the Office for Remote Sensing of Earth Resources (ORSER) digital processing system since 1973. He anticipates taking a position with NASA at the Goddard Space Flight Center, Greenbelt, Maryland, in the spring of 1979.

Dr. Turner received a B.Sc. (Forestry) degree at the University of Sydney in Australia in 1959, an M.F. degree from Yale in 1962, and a Doctor of Forestry degree from Yale in 1965. He worked in governmental forestry in Australia until he went to Penn State in 1969. There he teaches in forest resource management and biometrics. Since its inception in 1970, he has been an active participant in the Office for Remote Sensing of Earth Resources at Penn State and has contributed to the ORSER computer processing system, which he now administrates. He has helped install the system and conducted workshops at several locations in the U.S. and Australia. He has been involved in research contracts with NASA, the National Park Service, and the Forest Service.