SAS® SYSTEM STATISTICAL PROCEDURES APPLIED TO THE ANALYSIS OF WATER QUALITY DATA

Paul D. Mowery, SCI Data Systems, Inc.

INTRODUCTION

This study deals with the multivariate statistical assessment of environmental change. Changes over time and space in Chesapeake Bay water quality. Water quality is defined as the ability of a body of water to support aquatic life and to provide a commercially and recreationally viable resource for everyone to utilize and enjoy. The objectives of the study are to develop a conceptual framework for analyzing historical water quality data, and to develop the statistical methodology for detecting future changes in water quality. The statistical questions addressed herein are:

How can we estimate, for some fixed time period, the way in which water quality varies from one area of the Chesapeake Bay to another?

How can we measure historical changes in water quality?

How can we detect, with some level of statistical power, future changes in water quality?

This paper demonstrates that Canonical Discriminant Analysis, applied to a carefully screened data set, can help answer these questions. Canonical Discriminant Analysis is related to the Multivariate Analysis of Variance (MANOVA) as well as to Principal Components Analysis (PCA). It provides a great deal more interpretation of factor level differences than MANOVA, especially if the analysis results are displayed graphically. Like PCA, Canonical Discriminant Analysis is a dimension reduction technique, but with the additional capability to handle grouped data. These groups (or factor levels in the parlance of experimental design) may be a product of a carefully designed experiment. However, in retrospective studies, they are often generated by the researcher.

The statistical results were calculated by SAS® PROC CANDISC. Many of the same results can be obtained via PROC GLM. In addition to these two procedures, PROC DISCRIM will test the hypothesis of homogeneous covariance matrices, and can further analyze the results of Canonical Discriminant Analysis. PROC CANDISC results can be most clearly understood (and explained to others) using SAS/GRAPH® procedures. This is demonstrated in the paper's last section.

To motivate the statistical model, the paper's first section provides a brief historical perspective of environmental change in Chesapeake Bay. Section 2 describes the water quality data base, which was developed over a six year period and includes field data collected by several research institutions and government agencies in the Chesapeake Bay region. Section 3 presents the conceptual environmental model and its statistical counterpart. Results obtained using PROC CANDISC are described in Section 4.

1. THE ENVIRONMENTAL PROBLEM

HISTORICAL PERSPECTIVE

The Chesapeake Bay is our nations's largest estuary. The Bay proper is approximately 200 miles long and varies in width from about four to 30 miles. The Bay's water surface is bordered by approximately 4600 miles of shoreline and thousands of acres of wetlands. The water surface of the Bay proper encompasses more than 2200 square miles. Including tributaries, that figure nearly doubles. The Bay holds about 18 trillion gallons of water.

In the last several years, there has been a great deal of public interest in Chesapeake Bay water and resource quality. This attention is well deserved. The Bay is currently in a dramatic environmental decline. Populations of freshwater spawning fish like shad and striped bass have decreased. Oyster bars which once produced bountiful harvests are now barren or covered with dead shell.

These declines in living resources have been accompanied by deteriorating water quality. Over enrichment of nutrients has caused decreased water clarity and excessive algae growth. Inorganic nutrients - such as nitrogen and phosphorus - are an essential component of the Bay's food chain. They are present in the water column naturally and are taken up by algae. These tiny plants then become food for a myriad of small aquatic animals, which are in turn eaten by larger animals, and so forth. However, natural levels of nutrients in the Bay have substantially increased due to human activities. Runoff from farms contains nutrients, expecially nitrogen. Sewage treatment plant effluent is very high in nutrients, especially phosphorus. These nutrients are easily utilized by the algae.

Excessive nutrients have resulted in an explosion in the algal population. Growing on or near the surface, these algae prevent light from reaching a very important Bay resource - submerged rooted vegetation. These grasses play a vital role in the Bay's ecosystem. They slow water velocities, causing particulates to settle, keeping the water clear. They are a direct food souce for waterfowl. They provide habitat for an intricately related group of animals. In recent years, the amount of submerged rooted vegetation has substantially declined.

Another detrimental effect of excessive nutrients is low dissolved oxygen, or hypoxia, near the bottom of the Bay. As algae die and sink to the bottom, oxygen is used up in the process of bacterial decay. In this way, high algae populations can cause a shortage of dissolved oxygen - making marine life impossible near the bottom of the Bay. Low dissolved oxygen is a natural phenomenon, but the condition has grown much worse in recent years.

Given the problems caused at least in part by excessive nutrients, there is currently a great deal of interest in reducing nutrient inputs from farms, sewage treatment plants, urban areas, and other sources. For example, the Maryland legislature recently enacted a ban on the sale of phosphate detergents. One consequence of all this attention is that the study of ambient nutrient concentrations (i.e., the dissolved levels in the water column) has become an important issue.

Environmental managers need to know which areas of the Bay are similar with respect to nutrients, and which are different. Since some areas of the Bay are healthier than others withrespect to resources, we would expect these areas to be associated with different ambient nutrient concentrations. Furthermore, the detection of statistically significant differences in nutrient levels between healthy and unhealthy areas is a prerequisite to understanding the factors most responsible for the environmental problems. A related issue is the detection of changes over time. If nutrient levels in one area of the Bay change, then we would hope to see parallel changes in the resources in this area.

2. THE DATA BASE

A number of universities, research institutions, and government agencies regularly sample for nutrients in the water column. The sampling designs and methodologies are not always the same, and these sampling programs were not all intended for Bay-wide assessments such as the present study. Therefore, we had to be selective in our choice of data sets, choosing only those using similar techniques. Fortunately, there is no lack of water quality studies or water quality data. For example, the U. S. Environmental Protection Agency recently built a large water quality data base by combining a large number of individual studies. This data base includes data on ambient nutrient conditions throughout Chesapeake Bay (Flemer et al 1983). It was used to estimate the model parameters described in the following sections.

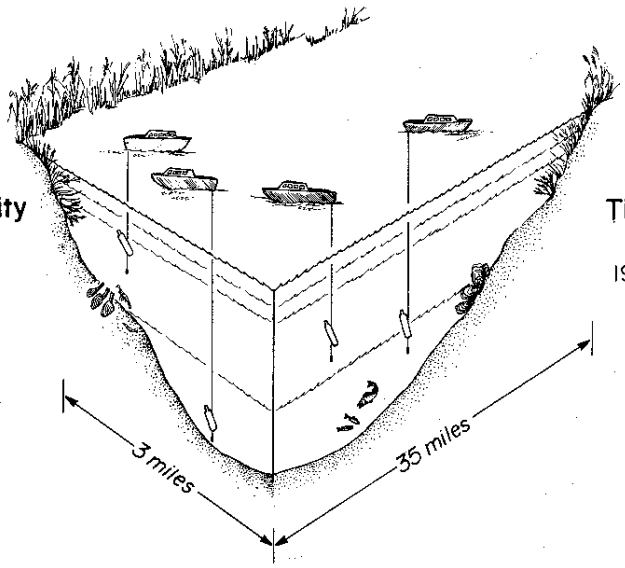
DATA POOLING

In this retrospective study, samples collected at the same time Bay-wide were rarely available. This required that we pool data collected on different dates in different locations into meaningful observational units prior to any statistical analysis. The way in which the data are pooled will be a major determinant of the outcome of the analysis.

The large volume of data associated with this study provides many possible data pooling strategies. The factors that potentially contribute to changes in ambient nutrient concentrations in the Bay include sampling location, sample depth, time of day, water temperature, tide stage, salinity, and storm events. One job of the statistician in this study was to assign each of these factors either to an experimental design factor, blocking factor, or replication factor. There are no true replicates due to the lack of strict randomization. We must be careful in choosing the factor(s) to act as replicates so that they reflect as accurately as possible the true random variability in the water column.

On a Bay-wide scale, the choice of spatial units (within which some sort of replication could be devised) was determined by reasons external to this study. The Bay had been divided into 43 large areas or segments by a previous study (Flemer et al 1983). This spatial delineation was maintained. Since the present study objective is to investigate spatial differences in nutrient concentrations, "Bay area" is clearly an experimental design factor. Nested within this factor is the effect due to differences among sampling stations. This was designated as the random error effect, and all data collected at the same station over some fixed time period were averaged.

Next, the question of the time period within which we would average data had to be resolved. Since nutrient concentrations change dramatically by season, season was selected as



that time period. A separate analysis was carried out for each season - spring, summer and fall (very few samples were collected during the winter). For each analysis, the data value associated with each sampling station was the average of all data collected at that station during the appropriate season.

Figure 1 illustrates the additional factors influencing ambient nutrient concentrations within one of the Bay areas. The area was sampled many times during the 1977 to 1980 baseline period. The 1977 to 1980 time period was chosen as the one against which future nutrient levels would be compared. On each visit, sampling station location varied. This is the replication factor. Also varying were the sampling date, the depth at which water quality samples were collected, salinity, time of day, water temperature, tidal stage, and weather conditions. The effects of these factors on nutrient concentrations were averaged into the station mean and were not estimated in the model.

The problem of pooling data points in a retrospective study has no perfect solution. In a designed experiment, some of the aforementioned factors influencing nutrient concentrations would be controlled by, e.g., blocking of the collection dates and stations. In the present study, the pooling strategy was driven by the need to avoid to the extent possible missing data, while choosing a strategy that is scientifically acceptable. The strategy chosen in this study - averaging within a season at each data collection station - presents the statistician with some problems. For example, each averaged nutrient concentration is based on a possibly different number of values, and, therefore, is associated with a different variance.

3. MULTIVARIATE MODEL

CONCEPTUAL/ENVIRONMENTAL BASIS

Our hypothesis is that Chesapeake Bay water quality, expressed as ambient nutrient concentrations, varies from one area of the Bay to another. The statistical objective is to accurately and precisely detect these differences. We would like to be able to group areas with similar water quality and use these groups as an environmental baseline. Future nutrient concentrations could then be compared with this baseline to evaluate changes in water quality.

Figure 2 shows a conceptual environmental baseline. Several different Bay areas are arranged by two indices of nutrient enrichment. The horizontal axis measures increasing nutrient enrichment from point sources such as sewage treatment plants. Patuxent and Back Rivers are highly enriched (or eutrophic) due to the large Baltimore - Washington population centers. The vertical axis measures increasing enrichment from nonpoint sources such as farm land. The Upper Main Bay is highly eutrophic due to this type of pollution.

FIGURE 2

NUTRIENT BASED CLASSIFICATION

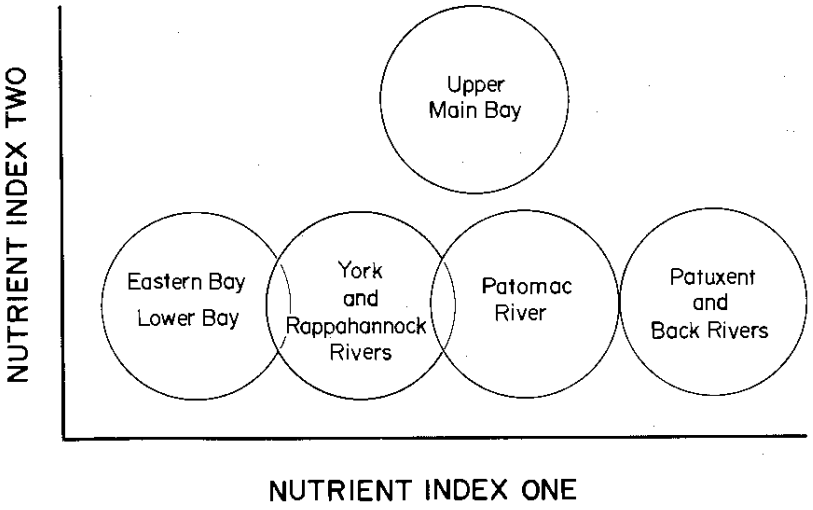
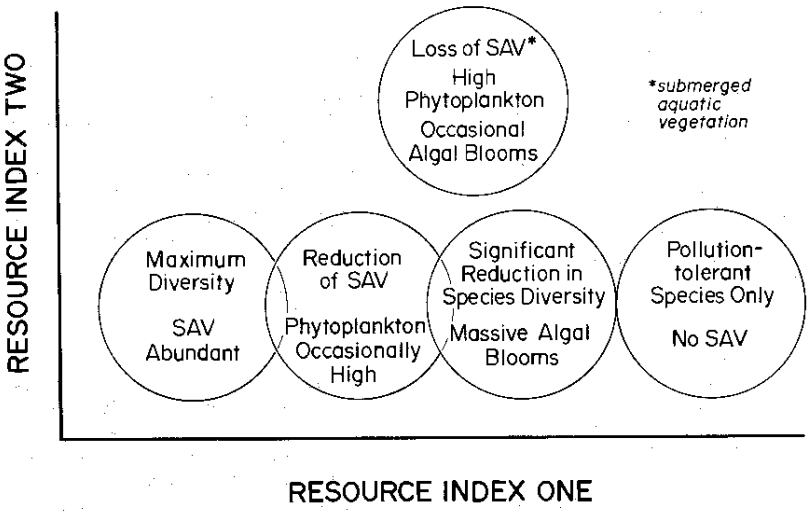


FIGURE 3

RESOURCE BASED CLASSIFICATION



After establishing the environmental baseline, we can compare it with future nutrient measurements collected from the same Bay areas. If a large data base (similar to the one described above) is available, we can reanalyze it in combination with the baseline data for evaluating temporal changes. For example, a shift in the Patuxent Group toward the Potomac Group may indicate an improvement in Patuxent River water quality. If only a few new data points are available, we would evaluate the statistical similarity among these and the existing groups using the technique of classification (see, e.g., Cooley and Lohnes 1971). The detection of water quality changes is an important statistical, environmental and political issue in the Bay region as government and industry begin the process of restoring water quality to a more healthy condition.

The evaluation of spatial and temporal changes in water quality, however, is not the "bottom line" from an environmental management perspective. If there are spatial differences in water quality, we would hope to see approximately parallel changes in biological populations such as submerged grasses, fin fish, and benthic animals. Unfortunately, the causal relationships between water quality and biota are difficult to quantify due to the complexity of the estuarine ecosystem. One way around this problem is to define a resource baseline similar to the nutrient/water quality baseline shown in Figure 2. This resource baseline (Figure 3) would group the same Bay areas by indices of resource viability, based on, e.g., the presence of submerged grasses, the abundance of fin fish, and the diversity of benthic species. If ambient nutrient concentrations and biological viability are causally linked, then changes in the nutrient baseline graph should ultimately be reflected by similar changes in the resource baseline graph.

ALGEBRAIC REPRESENTATION

The algebraic model used to build the environmental baseline is:

$$f(k,m) = U(0) + U(1)Y(1,k,m) + U(2)Y(2,k,m) + \ldots + U(p)Y(p,k,m); \quad (1)$$

where the Y's are the original dependent variables (ambient nutrient concentrations in this study), the U's are unknown coefficients to be estimated, and there are p dependent variables and k=1,2,...,g groups. The U's will be called canonical discriminant function coefficients. The values of the Equation (1) function will be called canonical discriminant scores. This is the standard Canonical Discriminant Analysis model.

It is possible to calculate the lessor of p and (g-1) canonical discriminant functions. The scores on any two functions will be uncorrelated. Similar to Principal Components Analysis, it is customary to designate the first function as the one which maximizes the differences among group mean canonical discriminant scores. These means will be called the group centroids. The coefficients of the second function are also derived to maximize differences among group centroids, but subject to the constraint that the two functions' scores be uncorrelated.

For parameter estimation and hypothesis testing, the usual MANOVA assumptions on the dependent variables are imposed:

- the total number of dependent variables must be less than (n-2);
- the variables must be measured at the interval level;
- no variable may be a linear combination of the other dependent variables;
- the covariance matrices for each group must be equal;
- each group must be drawn from a population with a multivariate normal distribution (Klecka 1980).

The assumption of a multivariate normal distribution is unnecessary for parameter estimation. This means that we can estimate the canonical discriminant model and evaluate group differences using graphical techniques (described in Section 4) in the absence of a multivariate normal distribution. Although MANOVA test statistics are robust under departures from normality and homogeneity of covariance matrices (Cooley and Lohnes 1971), the homogeneity of covariance matrices should always be investigated. The procedure proposed by Box (1949) is frequently used to test this assumption.

Equation (1) coefficients can be estimated without reference to a linear model. However, the canonical discriminant equation corresponds to the left side of the MANOVA model (in matrix notation):

$$Y = XB + E; \quad (2)$$

where X and E are the design matrix and error matrix, respectively. I prefer to think of the canonical discriminant model as a special case

of Equation (2) in which the design matrix corresponds to a one-way layout. In this case the MANOVA and Canonical Discriminant Analysis statistics for testing the hypothesis of no differences among group means are identical. Furthermore, Canonical Discriminant Analysis can be thought of as an extension to the MANOVA in which the function used to calculate this test is further examined for the strength of and reasons for any group differences. The advantage of thinking of Canonical Discriminant Analysis in this framework is that it is possible and often useful to apply the technique to more complicated linear models.

SAS PROCEDURES

The three main SAS procedures for analyzing multivariate grouped data are PROC GLM, PROC DISCRIM, AND PROC CANDISC. Each uses a different analytical framework. Each procedure's output provides a different interpretation of the model. PROC GLM can be used to test the hypothesis of no differences among group means. This is done via Wilks' Lambda as well as several other multivariate analogues to the univariate F-test. PROC GLM offers the advantage that data from complicated experimental designs can be analyzed. PROC GLM can optionally calculate the traditional Canonical Discriminant Analysis results. For a one-way layout, these results will be identical to those produced by PROC CANDISC (described below). However, Canonical Discriminant Analysis group means are always arithmetic averages, and within-group dispersion is always measured by the squared difference between each observation and its group mean. These are not always the least squares estimates produced by PROC GLM (in fact, they never are when the design is an unbalanced factorial). Therefore, with designs other than a one-way layout, you must carefully define the among-groups and error sums of squares with respect to PROC GLM model effects.

The emphasis of PROC DISCRIM is on classification. This is particularly useful when the group membership of some observations is unknown or in doubt. PROC DISCRIM computes a generalized squared distance function and uses it to measure the similarity between the observations and group centroids. An observation is classified into the group with which it has the smallest generalized squared distance. These classification results are not the same as those produced by PROC CANDISC. This later procedure computes generalized squared distances based on the canonical discriminant scores, not the original dependent variables. For evaluating group differences, PROC CANDISC produces more useful information and is recommended over PROC DISCRIM. However, PROC DISCRIM should be used to test the homogeneity of the covariance matrioes, a test that is not performed by PROC CANDISC.

PROC CANDISC carries out Canonical Discriminant Analysis. The technique is called "Canonical" Discriminant Analysis because of its resemblance to canonical correlation. By defining a series of dummy variables representing group membership, you can think of Canonical Discriminant Analysis as calculating a linear function of the dependent variables such that the correlation between this function and the group membership variables is maximized. Probably the most useful aspect of Canonical Discriminant Analysis is its ability to reduce a large number of dependent variables to two or three canonical discriminant functions that contain most of the information about group differences. (In this respect, Canonical Discriminant Analysis is very similar to Principal Components Analysis.) The canonical discriminant functions are used to evaluate group differences, and to classify new observations into one of the groups. In the process, PROC CANDISC generates a great deal of additional information that can be used to evaluate the strengths of group differences, and the reasons for these differences.

PARAMETER ESTIMATION AND TESTING

Equation (1) coefficients are estimated by calculating the eigenvalues and eigenvectors from the ratio of two quadratic forms:

$$L(i) = \frac{v(i)' A\, v(i)}{v(i)' W\, v(i)}; \qquad (3)$$

$$\text{subject to: } v(i)' v(i) = 1.0;$$

where the matrices A and W are the among-groups and within-groups sums of squares and cross-products, respectively, the vector v(i) represents the ith eigenvector, and L(i) is the ith eigenvalue, a scaler. The coefficients in Equation (1) are calculated from the eigenvectors by transforming each element so that the canonical discriminant scores will have zero mean and unit within-groups standard deviation. The standardization of the scores is extremely helpful in interpreting scatter plots.

The estimated coefficients of the first canonical discriminant function, when applied

to the original dependent variables, will produce scores with group centroids as different as possible. The second function is calculated to produce maximum group differences and be uncorrelated with the first. After calculating the first two or three functions, scores on the remaining functions usually provide little additional insight into group differences.

The total sums of sums of squares and cross-products matrix, T, is calculated as:

$$T = A + W; \qquad (4)$$

From this matrix, you can easily calculate the total sample correlation matrix, R. To calculate R, divide each element of T by the square root of the product of the two corresponding diagonal elements. The within-groups correlation matrix is calculated by applying the same calculations to the within-groups sums of squares and cross-products matrix, W. If there are real differences among the group, the within-groups correlations are better estimates of the relationships between the dependent variables (Klecka 1980). Correlations should be examined carefully since highly correlated dependent variables can make difficult the discovery of the variables most important to group differences.

The matrices A and T form the basis for Wilks' Lambda:

$$LAMBDA = \frac{|W|}{|T|}; \qquad (5)$$

This statistic tests the hypothesis of no differences among group means calculated on the original dependent variables. If there are differences, the determinant of W will be small relative to the determinant of T, and LAMBDA will be small. Another interesting interpretation of Wilks' Lambda is given by Cooley and Lohnes (1971), who note that the statistic can also be expressed as one minus the multivariate correlation ratio (often called Eta-Square). Thus as Wilks' Lambda approaches zero (i.e., becomes more statistically significant), Eta-Square approaches a perfect correlation ratio of unity.

Another result of Canonical Discriminant Analysis is the canonical correlation coefficient, R(i), which measures the correlation between each discriminant function and the group membership dummy variables. This can be calculated from the eigenvalues as:

$$R(i) = (L(i)/(1+L(i)))^{-1}; \qquad (6)$$

Another useful result, structure coefficients, are used to determine how closely each dependent variable and canonical discriminant function are related. The structure coefficient for each dependent variable and canonical discriminant function is calculated as the product-moment correlation coefficient between the variable and canonical discriminant score. When the absolute magnitude of this coefficient is near unity, the canonical discriminant function and the dependent variable are carrying nearly the same information about inter-group differences (Klecka 1980).

4. INTERPRETING THE ANALYSIS

The results in this section were calculated by PROC CANDISC. Graphical displays were generated using SAS/GRAPH. The analysis data set consists of ambient nutrient concentrations from the spring of 1977. For reasons independent of this study, 1977 was chosen as a baseline year against which data from subsequent years will be compared.

Four measured nutrient concentrations plus one calculated value were used as dependent variables:

SYMBOL	VARIABLE
N	Total Nitrogen (mg/l as N)
P	Total Phosphorus (mg/l as P)
NH3	Ammonia (mg/l as N)
NO3	Nitrate (mg/l as N)
NP	Nitrogen:Phosphorus Ratio

The Bay areas included in the analysis and associated plotting symbols are:

A	Upper Bay 1	H	Patuxent River
B	Upper Bay 2	I	Potomac River
C	Eastern Bay	J	Rhappahannock River
D	Elk River	K	York River
E	Sassafras River	L	James River
F	Chester River	M	Back River
G	Choptank River		

PLOTS OF CANONICAL DISCRIMINANT SCORES

The most illustrative way to interpret Canonical Discriminant Analysis is to plot the discriminant scores. A plot of canonical discriminant scores for the first two functions (Figure 4) shows one obvious atypical Bay area - Back River (plotting symbol "M"). This area was judged unlike any other area. Its data were removed from all subsequent analyses. Back River receives effluent from the Baltimore City sewage treatment plant, which causes very high ambient nutrient concentrations. For comparing 1977 with subsequent years, the data for Back River will be analyzed separately.

After removing Back River data and recalculating the scores, several distinct groups of Bay areas emerge on a plot of the scores for the first two functions (Figure 5). Each point on the Figure 5 plot represents the score for a unique data collection station. There are 67 points on the plot, and thus there are 67 stations at which nutrient levels were measured during the spring of 1977. Letters on the plot signify the eleven Bay areas. Each area measures approximately 150 square miles. For any single nutrient, the mean over all stations within a single area is an estimate of the average concentration in that area. Similarly, the mean canonical score is an estimate of the average score for that area. For maximum area differences, all station scores within the same Bay area should appear in a tight cluster of points, with little or no overlap between clusters.

In Figure 5, the areas appear to cluster into five groups. Two areas, the Potomac River (plotting symbol "I") and the Patuxent River (plotting symbol "H") comprise single area groups. The other nine areas seem to cluster into three additional groups. All eleven areas differ in the amount of intra-area dispersion, a hint that the covariance matrices may be unequal.

The Rhappahannock River - James River Group comprises by far the most compact cluster of station scores. This group is clearly distinct from the four other groups. Both of these rivers are located on Virginia's western shore.

The greatest dispersion occurs within the Eastern Bay - Sassafras River - Chester River Group. The Sassafras and Chester Rivers are adjacent rivers on Maryland's eastern shore. Both are moderately enriched due to farm runoff. Eastern Bay is a large embayment located directly south of the Chester River. The circulation patterns in Eastern Bay are different from those of the Chester and Sassafras Rivers. We would not necessarily expect water quality in Eastern Bay to be similar to that in the Chester and Sassafras Rivers. All three stations from Eastern Bay appear at the extreme top of the cluster. Given more stations from Eastern Bay, this area may become a separate group.

The Eastern Bay - Sassafras River - Chester River Cluster overlaps the cluster formed exclusively by Potomac River station scores. During 1977, the Potomac was a river in transition. Pollution abatement measures implemanted during the preceding decade had resulted in modest water quality improvements. However, serious and persistent problems remained. This transition may be responsible for the overlap shown in Figure 5.

Four scores depart from the groups. The most obvious is one Patuxent River station score. In addition, two Choptank River and one Potomac River Stations appear to be outliers. The individual nutrient measurements from these stations as well as the station locations and sampling methodologies will be evaluated.

CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

We can gain an understanding of the reasons underlying group differences by examining the standardized function coefficients. Standardized coefficients for the first two functions are:

$$f1 = -0.6N + 0.3P + 0.2NH3 + 0.9NO3 + 1.7NP \quad (7)$$

$$f2 = -1.5N + 0.1P + 0.7NH3 - 0.8NO3 + 1.8NP \quad (8)$$

FIGURE 4
FIRST TWO DISCRIMINANT SCORES

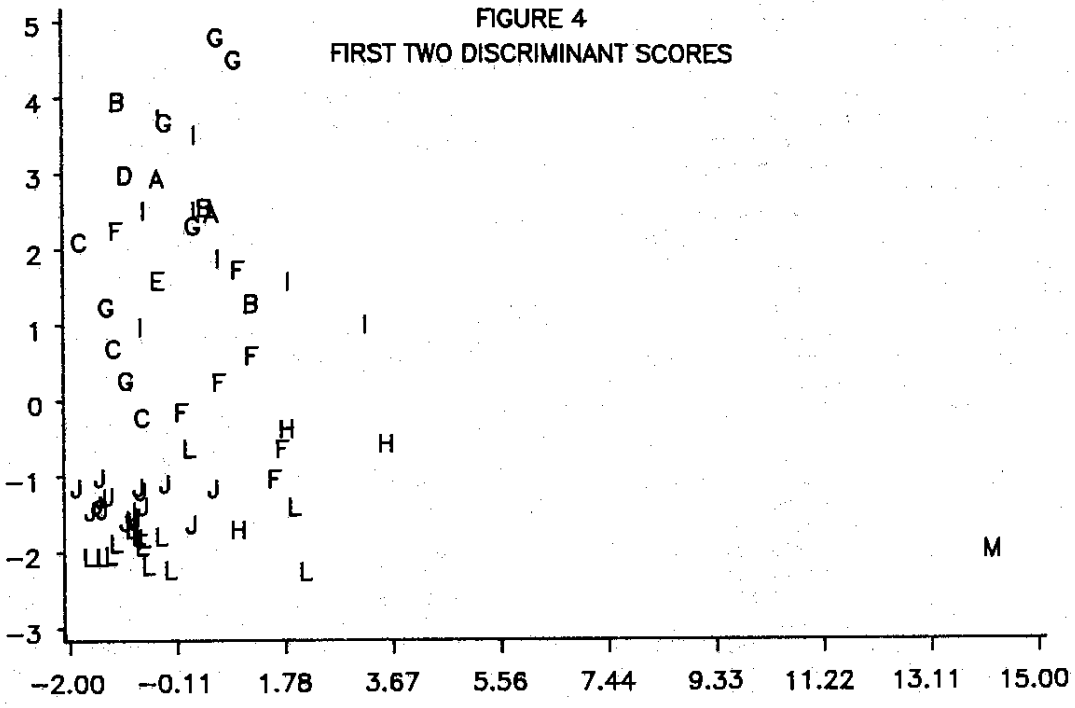
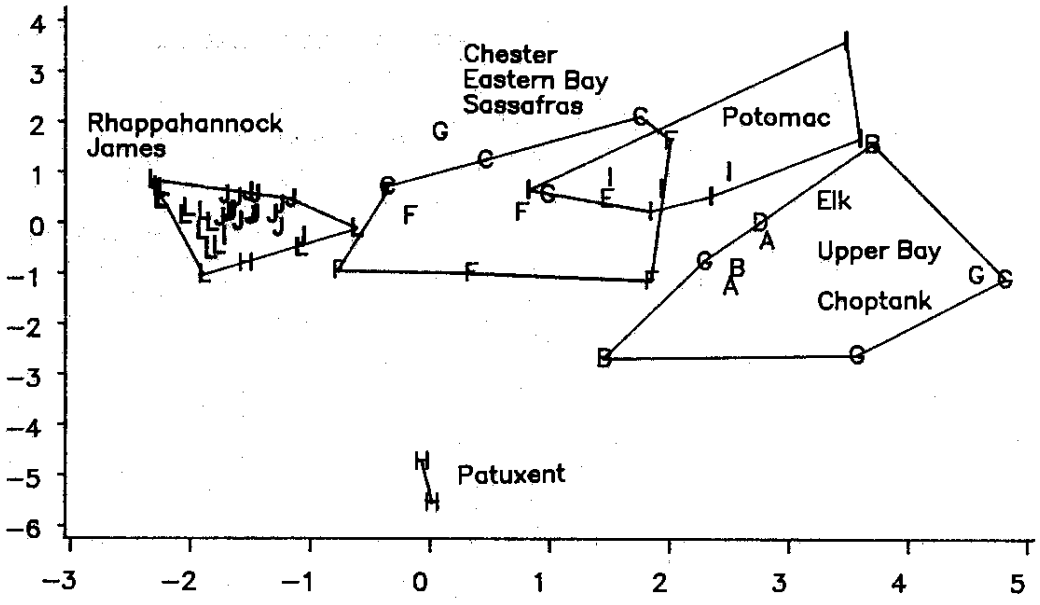


FIGURE 5
FIRST TWO DISCRIMINANT SCORES



In the first function, the nitrogen:phosphorus ratio and nitrate are the first and second most important dependent variables, respectively. High nitrate concentrations and/or high nitrogen:phosphorus ratios will locate stations to the extreme right on the horizontal axis. In the second function, total nitrogen and, again, the nitrogen:phosphorus ratio are the most important variables. High total nitrogen values will locate stations close to the bottom on the vertical axis (since the coefficient is negative). High nitrogen:phosphorus ratios will locate stations close to the top of the graph. The relative importance of nitrogen:phosphorus ratios in both functions may have a significant ecological interpretation, since this ratio is an index of the rate at which phytoplankton populations can potentially increase in density.

Figure 5 is a first step in establishing an environmental baseline for future comparisons of Bay water quality. The next step will be to develop similar graphs (using SAS multivariate procedures) for spring data collected in subsequent years (e.g., 1978 to 1980). Similarities and differences among these graphs will be evaluated. Stations not conforming to the general pattern will be investigated. The final discriminant score groups for the baseline period will be circulated among Chesapeake Bay water quality experts. The final groups will be based on their consensus opinion and on the statistical confidence intervals.

Classification analysis will be used to assign new observations to the most similar group. In this way, we hope to establish a continuing baseline against which progress in improving water quality can be measured.

ACKNOWLEDGMENTS

This study was funded in part by the U. S. Environmental Protection Agency, Chesapeake Bay Program. The views expressed are those of the author and not necessarily those of the Environmental Protection Agency.

REFERENCES

Box, G. E. P. 1949. A General Distribution Theory for a Class of Liklihood Criteria. Biometrika 36:317-346.

Cooley, W. W. and P. R. Lohnes. 1971. Multivariate Data Analysis. John Wiley and Sons, Inc. New York. 364 pp.

Flemer, D. A., G. B. Mackiernan et al. 1983. Chesapeake Bay: A Profile of Environmental Change. E. G. Macalaster, D. A. Barker, and M. E. Kasper, eds. U. S. Environmental Protection Agency's Chesapeake Bay Program. Washington, DC. 120 pp + Appendices.

Klecka, W. R. 1980. Discriminant Analysis. Sage Publications. Beverly Hills. 70 pp.

FOR ADDITIONAL INFORMATION

Please contact: Paul D. Mowery
SCI Data Systems, Inc.
530 College Parkway
Suite N
Annapolis, MD 21401
301 974-1340