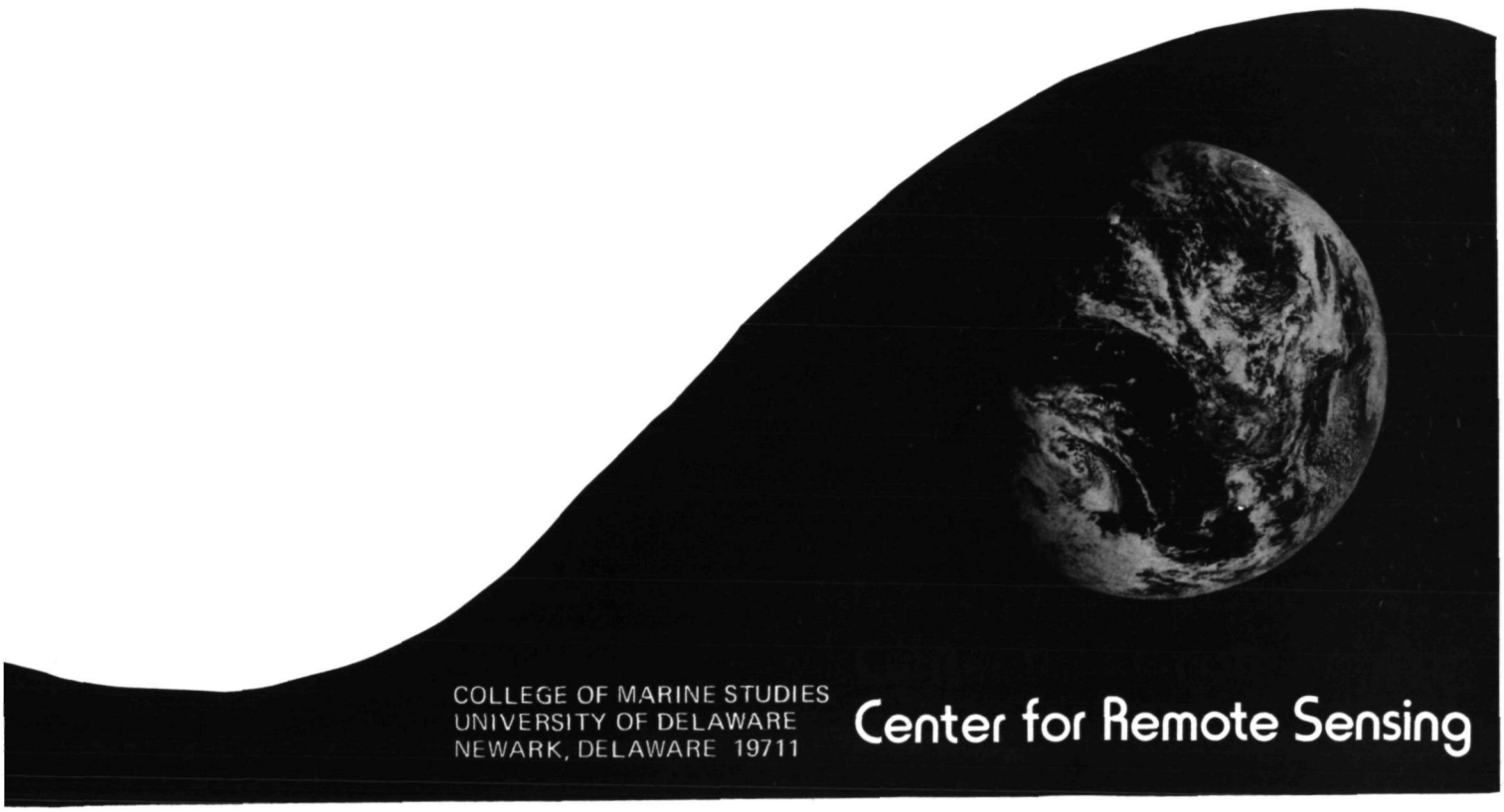
CRS-02-81

DETECTION OF OCEAN WASTE IN THE NEW YORK BIGHT

by W. Philpot and V. Klemas

Prepared for National Aeronautics
and Space Administration,
NASA Grant NSG 1398



COLLEGE OF MARINE STUDIES
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19711

Center for Remote Sensing

N79-21637

DETECTION OF OCEAN WASTE IN THE NEW YORK BIGHT

W. Philpot and V. Klemas
College of Marine Studies
University of Delaware
Newark, Delaware 19711

March 30, 1979
Final Report
NASA Grant NSG 1398

Prepared for
National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23665

I. INTRODUCTION

In order to control pollution in the ocean and near-shore waters it is important to have some effective means of monitoring the pollution in these waters. In most cases this will mean regular, if not frequent monitoring of large areas of open water. Thus, an effective monitoring system will probably have to include some form of remote sensing as part of the standard procedure since there is simply no other method of providing synoptic, large-area coverage at reasonable cost. The actual remote sensing techniques and technology used will depend on the particular application.

It is the purpose of this report to explore the application of remote sensing to detection and monitoring of ocean waste disposal in the New York Bight. This report will focus on the two major pollutants in this area--sewage sludge and iron-acid waste--and on detecting and identifying these pollutants. The emphasis is on the use of Landsat multispectral data in identifying these pollutants and distinguishing them from other substances.

The analysis technique applied to the Landsat data is the eigenvector (principal components) analysis which was described in an earlier report (Klemas et al., 1978). This approach proved to be quite successful in detecting iron-acid waste off the coast of Delaware and is applied here with relatively minor modifications. The results of the New York Bight work will be compared to the Delaware results.

Finally, other remote sensing systems (Nimbus G, aircraft photography and multispectral scanner systems) will be discussed as possible complements of or replacements for the Landsat observations.

II. LANDSAT AS A COASTAL WATER QUALITY MONITORING SYSTEM

As is apparent from its very name, Landsat was designed for use in studying land and was not intended for studying open waters at all. The dynamic range of the multispectral scanner (MSS) is such as to accommodate targets ranging from water (very low reflectivity) to dry sand (very high reflectivity). As a result, all the water features are generally restricted to about eight of the 128 gray levels available from the MSS imagery. In addition, two of the four spectral bands on the MSS are in the near infrared, a spectral region in which water is very highly absorbing making the effective depth penetration at these wavelengths very small. Nonetheless, there are several reasons for using Landsat data for observation of water properties. First of all, there is a surprising amount of structure apparent in the water in spite of the narrow dynamic range. Secondly, since Landsat has been operating for better than six years with repeat coverage of the same geographic locations every 18 days (every 9 days for the period that Landsat 1 and Landsat 2 were operating simultaneously), there is an enormous amount of data readily available for any coastal U.S. location during all seasons and for any tidal stage. There is simply no comparable set of spectral data available even for a single region. Finally, the relatively high resolution of the MSS ($\sim$80 m) is sufficient to allow identification of many features simply by their spatial patterns: sediment plumes, frontal zones, pollution dumps, circulation patterns, etc.

Thus, although Landsat may be less than ideal, it is probably the best system presently available for water quality monitoring in coastal waters. At the very least, the attempt to extract water quality information from Landsat data can provide some insight in the design and use of whatever system is ultimately used for water quality monitoring.

2.1 Eigenvector analysis of ocean color data

There are many ways to approach the analysis of the multispectral data from the Landsat MSS. The method chosen for this study is eigenvector analysis or principal components analysis. The use of principal components analysis in studies of ocean color was first undertaken by Mueller (1976), in his study of phytoplankton. Mueller's work was a classic application of the technique and demonstrated the effectiveness of eigenvector analysis in analyzing water color spectra. The derivation of eigenvector analysis will not be covered here, since it is covered in detail in several standard sources (see Morrison, 1976), and since a complete description of this technique for the ocean color application was provided in an earlier report (Klemas et al., 1978). Only the major points will be reviewed here.

The eigenvector analysis is not used here in the classical sense in which one of the main objectives is to reduce the number of significant variates. All four Landsat bands are useful for ocean color studies and any reduction in the number of bands would result in a significant loss of information. The eigenvector analysis is used here because the results can be described in geometric terms which aids considerably in understanding the system of variation, and because it is a statistical representation of the complete multivariate system unbiased by an assumption of dependent or independent variables. The latter is in direct contrast to multiple linear regression techniques which require an assumption of dependent and independent variables.

The method can be described using a two-dimensional (two-color) system. For illustration let us consider the problem of distinguishing between acid and sediment in a single Landsat scene. If the pixels in

this scene which corresponded to a sediment plume and an acid waste were plotted according to the intensities in bands 4 and 5, the green and red bands respectively, the result might appear something like that in Figure 1. In this figure the solid line outlines the region in two-color space in which the pixels corresponding to acid fall; the dashed line outlines the sediment pixels. At the low intensity end of these two regions is a third region outlined by a dotted line and corresponding to clear water. This schematic plot is fairly typical of scenes in which both sediment plumes and an acid waste appear. The position of the water region varies from one Landsat scene to the next, but the relative positions of the three regions remains fairly constant. The consistency of this pattern from scene to scene suggests that an automated classification scheme could be devised to identify these materials if the spectrum of clear water is known.

The eigenvector analysis is a method of providing a statistical description of these regions in multidimensional color space. For the purpose at hand, i.e. automated classification, it is convenient to assume that the clear water region defines the origin. The signature of each of the targets (sediment, acid waste, oil, sludge, etc.) is then described by its own set of eigenvectors (Klemas et al., 1978). The first eigenvector for each target corresponds to the direction in color space of the maximum variance for that target. In Figure 1 the first eigenvector for the acid waste, $\overline{A}$, lies along the major axis of the acid waste region. The second eigenvector defines the direction of maximum variance which is perpendicular to the first eigenvector. In the two-dimensional case of Figure 1, the second eigenvector for the acid waste would account for all of the remaining variance for this target and

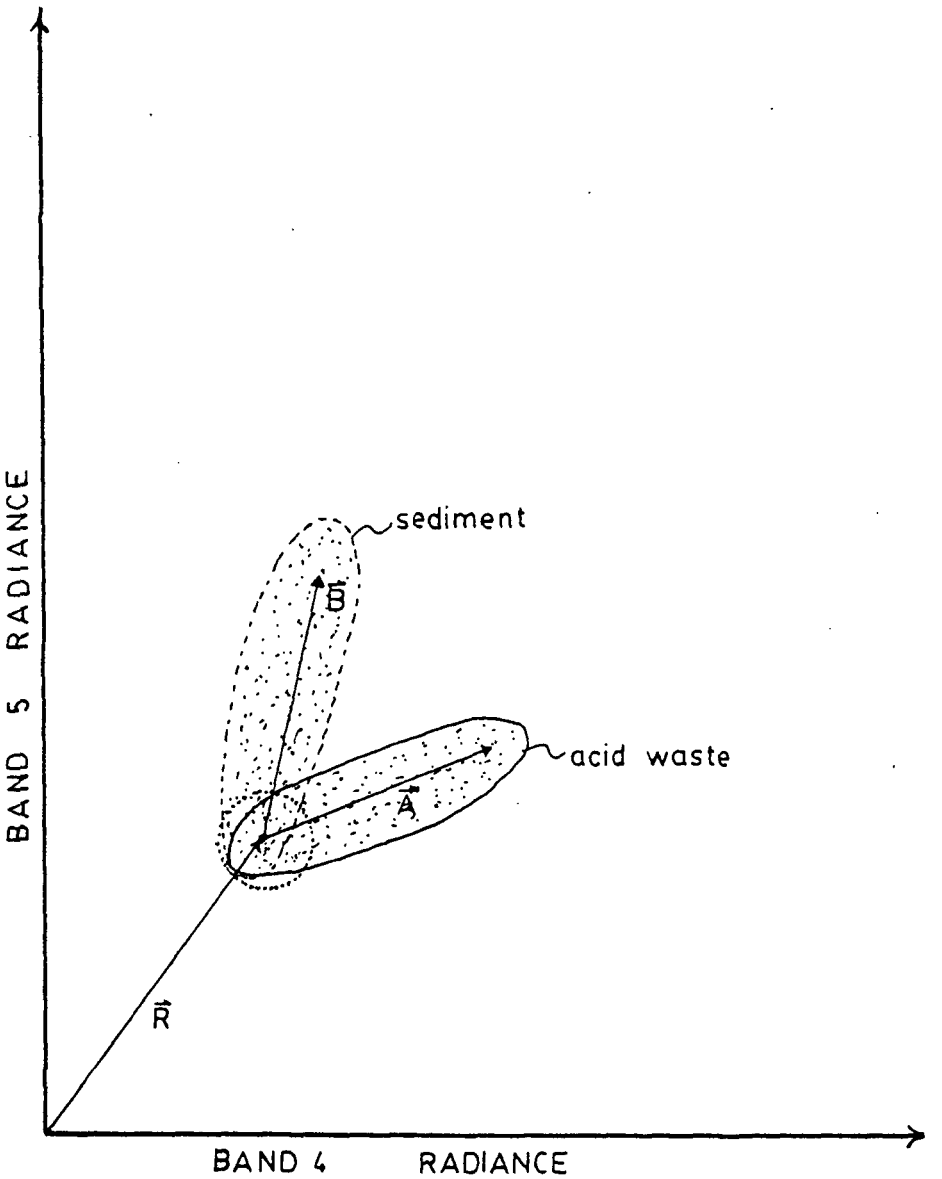


Figure 1. A two-band plot of Landsat data in which the regions corresponding to acid-waste, sediments and clear water.

would lie in the plane of the paper, perpendicular to the major axis of the acid region.

Once the targets have been described in terms of their eigenvectors it is then possible to use these eigenvectors as the basis of a classification scheme. We must be able to take a pixel from the Landsat data and decide which class, if either, it belongs to. Figure 2 illustrates the problem schematically for a two-band system. In this figure we have two classes whose first eigenvectors are $\hat{a}$ and $\hat{a}'$. The dotted lines parallel to the first eigenvectors represent the dispersion of data about the axis and can be characterized by the standard deviation (σ_2) of the training data along the direction of the second eigenvector ($\hat{a}_2$). For illustration we have chosen a distance of one standard deviation as a classification limit. The "clear" water mean is represented by o and $\vec{r}$ is the vector distance from the original color space origin to the clear water mean. A point at position A with position vector $\vec{p}_A$ relative to the "clear" water origin, clearly belongs to class 2 since it is within one standard deviation of axis of this class ($d' < \sigma_2'$) and well outside the same range of the other class ($d > \sigma_2$). The first step then is to find the distance of the test point from the axis of each of the classes and throwing out any point that is too far away. The distance d is given by:

$$d = |\vec{p} \times \hat{a}_1| = p\, a_1 \sin\theta_p \qquad (1)$$

A point at position B (Figure 2) is more difficult to assess since it is sufficiently close to the axes of both class 1 and class 2. The simplest criterion in this case is if the distance to one class relative to the one standard deviation limit of that class is less than the

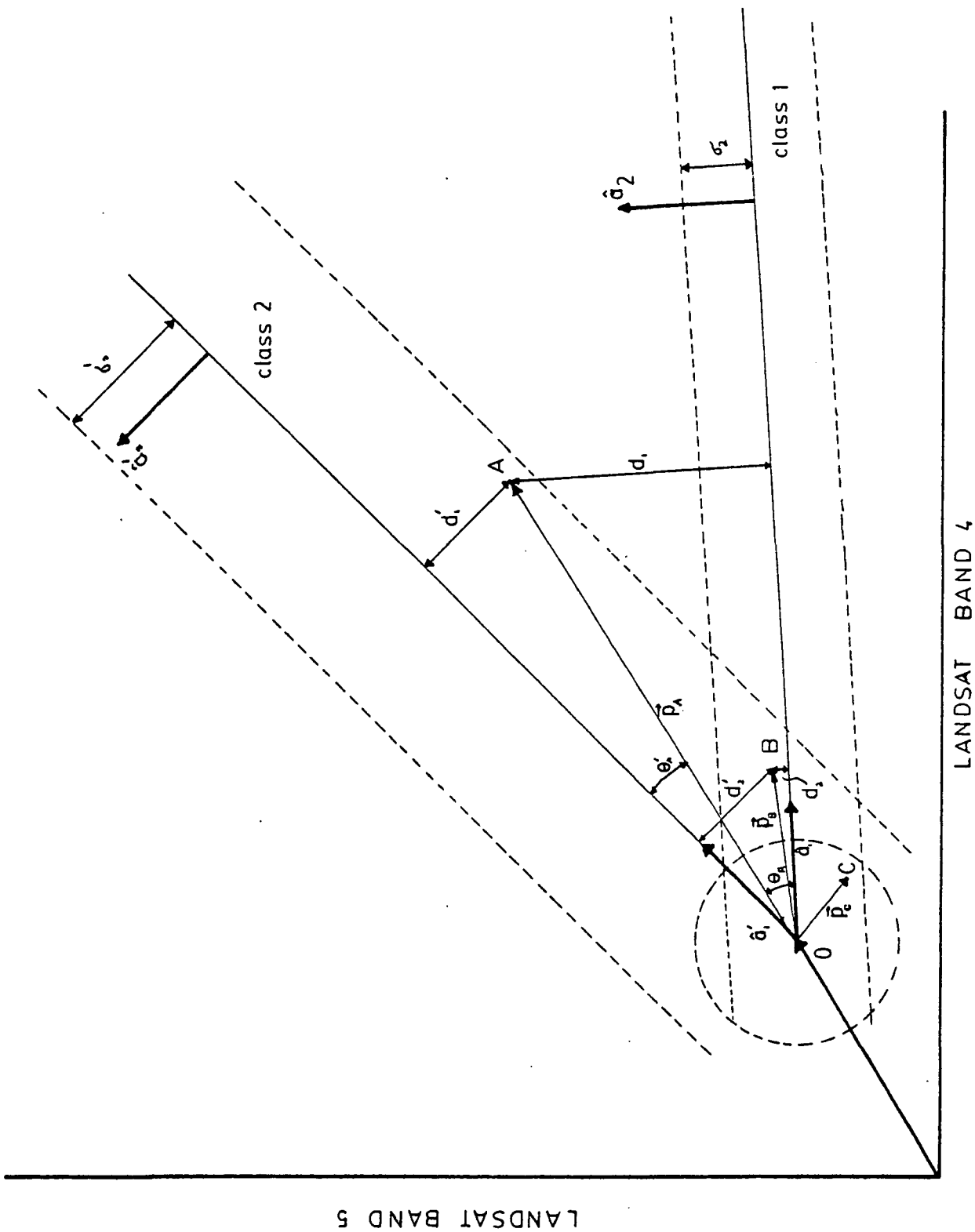


Figure 2. The eigenvector geometry applied to spectral classification.

distance to the second class relative to the one standard deviation limit of the second class. That is, if

$$\frac{d_1'}{\sigma_2'} < \frac{d_1}{\sigma_2} \qquad (2)$$

then the point is classified as belonging to class 2. The same sort of process could be used to place point C in one class or another. However, point C is within the range of the "clear water" region. Point C, therefore, would be classified as water.

There are several comments to be made at this point. First of all, with the Landsat data we are dealing with a four-dimensional system in which the first and second eigenvectors for different targets may not (and rarely do) fall in the same plane; the second eigenvector may not be in the direction which is significant for separating two classes. This can be understood conceptually for a three-dimensional case by realizing that the region in color space which is being used to identify the classification limit for a particular class is described by a circular cylinder whose major axis is the first eigenvector for that class and with a radius scaled by the standard deviation along the second eigenvector, σ_2. The actual distribution of the data would be better described by an ellipsoidal cylinder, the major and minor axes of which are scaled by the standard deviations along the second and third eigenvectors respectively.

The difference could lead to a misclassification. This is illustrated in Figure 3 which shows a cross-sectional view of two classes for the three-dimensional case. In this figure both of the second eigenvectors are in the direction of maximum variance about the major axis of the distribution but neither is along direction connecting the two

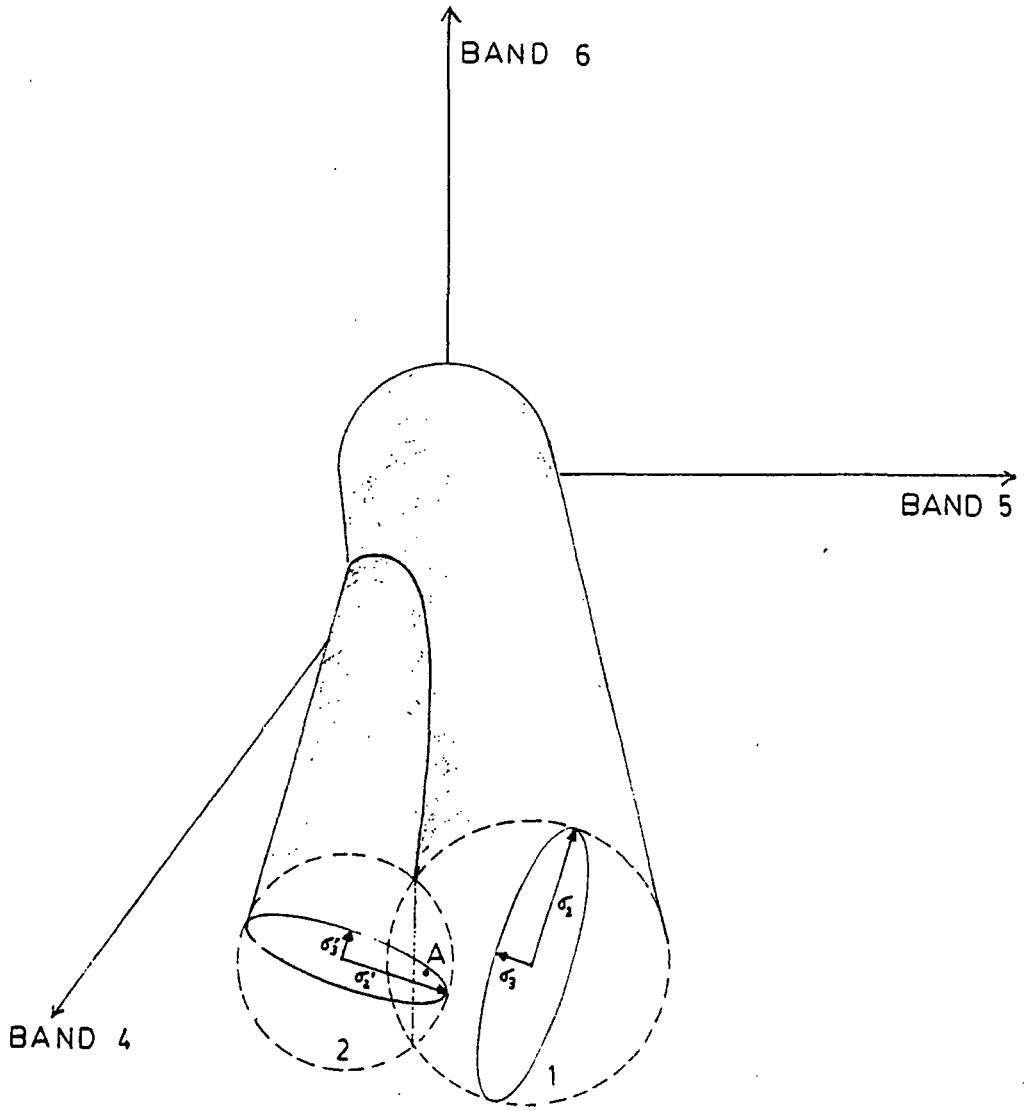


Figure 3. Cross-sectional view of two sample distributions in 3-dimensions.

distributions. Using the standard deviation along the second eigenvector as a charateristic distance for the classification is equivalent to assuming that the sample distribution filled the areas outlined by the circular cylinders rather than by ellipsoidal cylinders. In the case illustrated, a pixel which is located at point A would be classified as with class 1 rather than class 2. Accurate classification would require a more complex classification criterion than that of equation 2. In the name of calculational efficiency the results in this paper are based on the assumption that the distribution of the data about the major axis is essentially circular. This represents a worst case in that classification accuracy can only improve if this classification is refined.

A second comment on this eigenvector classification approach is that there is an implicit assumption that the variation along the direction of the first eigenvector is linear. Although this assumption appears to be adequate there is some evidence that it may not be true and that classification accuracy could be improved by taking the nonlinearities into account. This possibility will be covered in more detail below. For the present, however, we will continue to assume that the along axis variability is strictly linear.

Assuming that the eigenvectors for whatever substances are to be identified and that the "clear" water spectrum is known, the classification scheme is as follows (refer to Figure 2):

1) Find the vector $\bar{p}$ for the test pixel. If the pixel is described by vector $\bar{b} = (b_1, b_2, b_3, b_4)$ where b_1 is the intensity (in counts) in band 4, etc., and the "clear" water is given by $\bar{r} = (r_1, r_2, r_3, r_4)$, then $\bar{p}$ is given by

$$\bar{p} = \bar{b} - \bar{r} \qquad (3)$$

2a) If $|\vec{p}| < w$ where w is some minimum radius defining "clear" water, the A is classified as clear water.

2b) If $|\vec{p}| > w$ continue on to step three.

3) Find the perpendicular distance d from the test pixel at point A to the axis of each classification distribution.

$$d = |\bar{p} \times \hat{a}_1| = p \sin\theta_p \qquad (4)$$

where $\hat{a}_1$ is the primary eigenvector (a unit vector) for some class and θ_p is the angular separation of $\bar{p}$ and $\hat{a}_1$.

4) Test to see if this distance falls within the cutoff range ($d < \sigma_2$). This cutoff can be set as strictly or loosely as one might like. The scaling factor is the standard deviation σ_2 along the second eigenvector.

5a) If A is within σ_2 of only one class the classification is finished.

5b) If A is within σ_2 of more than one class then it should be placed in that class for which the ratio of d/σ_2 is smallest.

5c) If A is not within σ_2 of any class then it remains unclassified.

2.2 Pollution classification in the New York Bight

A major goal of this study is to apply the eigenvector classification procedure to waste plumes in the New York Bight. There are two major target pollutants in this region: an iron-acid waste and a sewage sludge waste. The iron-acid waste is similar to that which was discharged off the coast of Delaware and is expected to be spectrally quite similar to the Delaware waste. The Delaware iron-acid waste was the

subject of the earlier report by Klemas et al. (1978). The sewage sludge is secondary treated sewage. Both the acid and sludge are discharged within 20 miles of Sandy Hook, N.J.

Figure 4 is a Landsat image of the New York Bight on 19 August 1975. A fresh iron-acid waste discharge can be seen as a U-shaped pattern to the east and slightly south of Sandy Hook, N.J. An older acid discharge appears as a relatively bright smudge to the southwest of the fresh discharge. A sewage sludge discharge appears as a rather dull smudge to the northeast of the fresh acid discharge.

The eigenvectors for these two targets were found using training sets from the fresh acid discharge and the sewage sludge discharge and a "clear" water sample immediately north of the fresh acid discharge. A training set was also taken from an area just to the east of the fresh acid discharge which is covered by diffuse clouds. The older acid discharge was treated as an unknown material to be classified on the basis of the eigenvectors for the fresh acid waste.

Table 1 lists the eigenvectors for the 19 August 75 New York Bight targets and the angles between the first eigenvectors for all possible pairs. Also shown is the mean value of the "clear" water training set which was used as the origin for the eigenvector analysis. The clouds should be spectrally distinct from both the acid and the sludge. That this should be so can be seen from the angular separation of the first eigenvectors. The first eigenvector for the clouds is more than 24° from the first eigenvector for either sludge or acid. By the same criterion, the acid and sludge appear to be essentially indistinguishable. Their first eigenvectors are separated by less than 10° which means that, for the areas with weaker reflection, there will be much

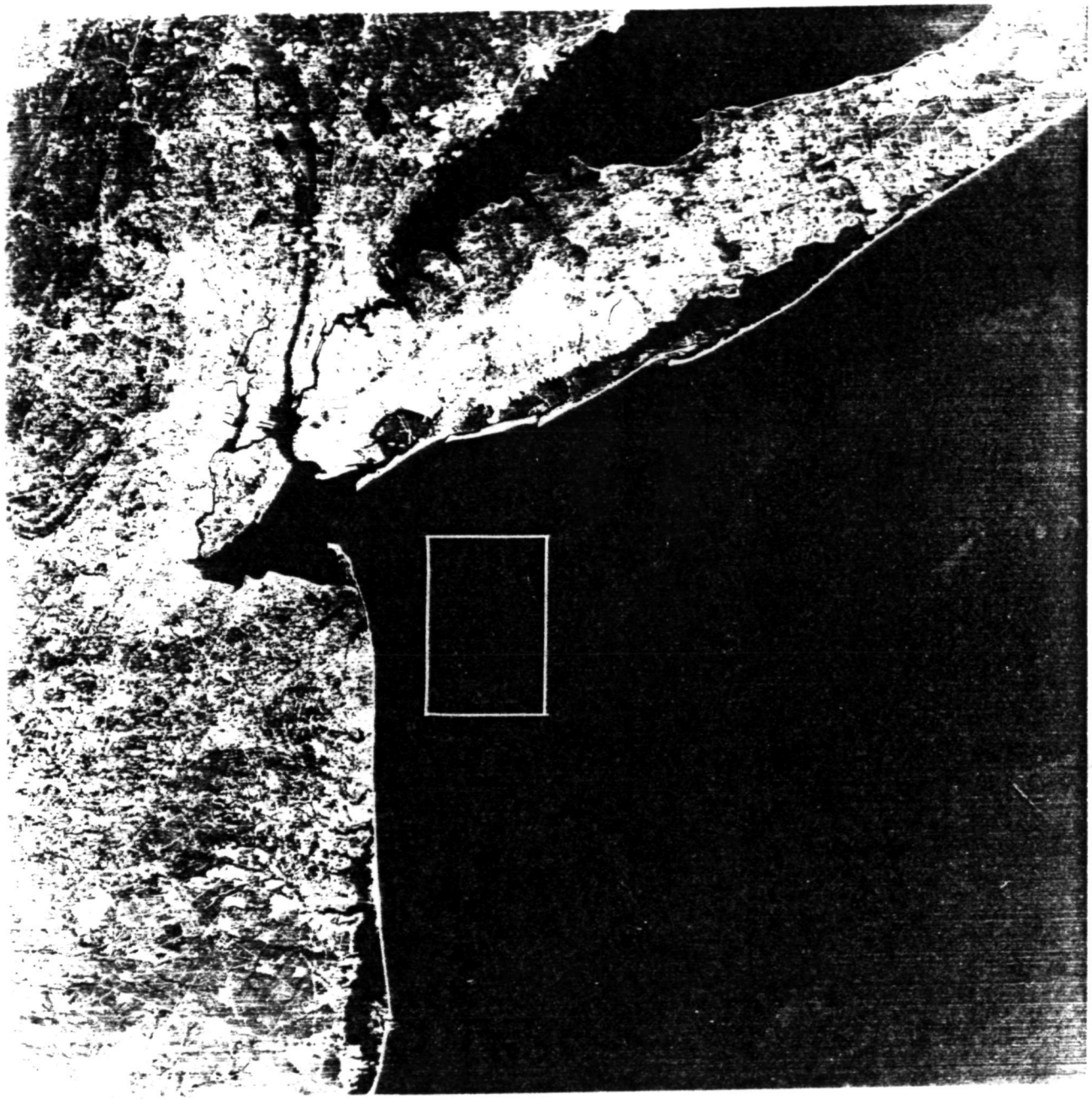


Figure 4. Landsat image of the New York Bight on 19 August 1975. The U-shaped pattern is the iron-acid waste. An older iron-acid waste appears to the lower left of the fresh discharge. A sewage sludge discharge appears to the upper left of the fresh acid discharge.

overlap in the spectral reflectance signatures and that there is unlikely to be any effective way of distinguishing between the two.

Figure 5 shows the results of the classification of the portion of the 19 August 75 New York Bight scene outlined in Figure 4. Figure 4A shows all the pixels classified as acid, sludge and clouds combined; 4B shows the pixels classified as sludge; 4C shows the pixels classified as acid. As was expected from the similarity between the eigenvectors for sludge and acid, the two pollutants are spectrally indistinguishable. In this case the classification would not be improved significantly by refining the classification procedure; not only are the first eigenvectors quite close, but the second eigenvectors are nearly parallel and along the direction of separation of the first eigenvectors.

There is at least a possibility that the acid and sludge may be distinguishable by the time rate of change of their reflectance spectra. Bowker (1978) has pointed out that the fresh sludge is brighter in Landsat band 6 than is the acid, and that the older sludge is darker than the acid in band 6. This trend is not clear from the eigenvector analysis--one would expect such a trend to cause the third component of the first or second sludge eigenvectors to be greater than their acid counterparts, exactly the opposite of what actually happens (Table 2). Nonetheless, this type information could be used to distinguish between acid and sludge.

2.3 Improving the classification

The distinction between the two pollutants and clouds is adequate but not as good as might have been expected from the angular separation of the eigenvectors. The eigenvectors for clouds are quite distinct

	Band	Acid	Sludge	Clouds	Unknown angles between the first eigenvectors of
Vector 1:	4	.7872	.8770	.6047	Acid and sludge = 5.9°
	5	.5755	.4371	.5238	Acid and clouds = 26.4°
	6	.2184	.1995	.5747	Sludge and clouds = 24.2°
	7	.0382	.0070	.1723	
Vector 2:	4	-.6076	-.3256	-.7934	
	5	.6718	.8334	.3625	
	6	.4232	-.4016	.4416	
	7	-.0203	.1952	.2098	
Vector 3:	4	.0703	-.3269	-.0302	
	5	-.4342	.2482	.7647	
	6	.8130	.8857	-.6365	
	7	.3809	.2171	-.0958	
Vector 4:	4	-.0762	.1342	.0620	
	5	.1700	-.2296	-.0971	
	6	-.3350	-.1206	-.2638	
	7	.9236	.9564	.9577	

Clear water mean in gray scale values (mw/cm^2)-ster-band)	Band 4	Band 5	Band 6	Band 7
	18.34 (0.358)	8.60 (0.135)	4.43 (0.061)	0.60 (0.038)

Table 1. A list of the four eigenvectors for pollutant targets in the New York Bight and the angular separation of the first eigenvectors.

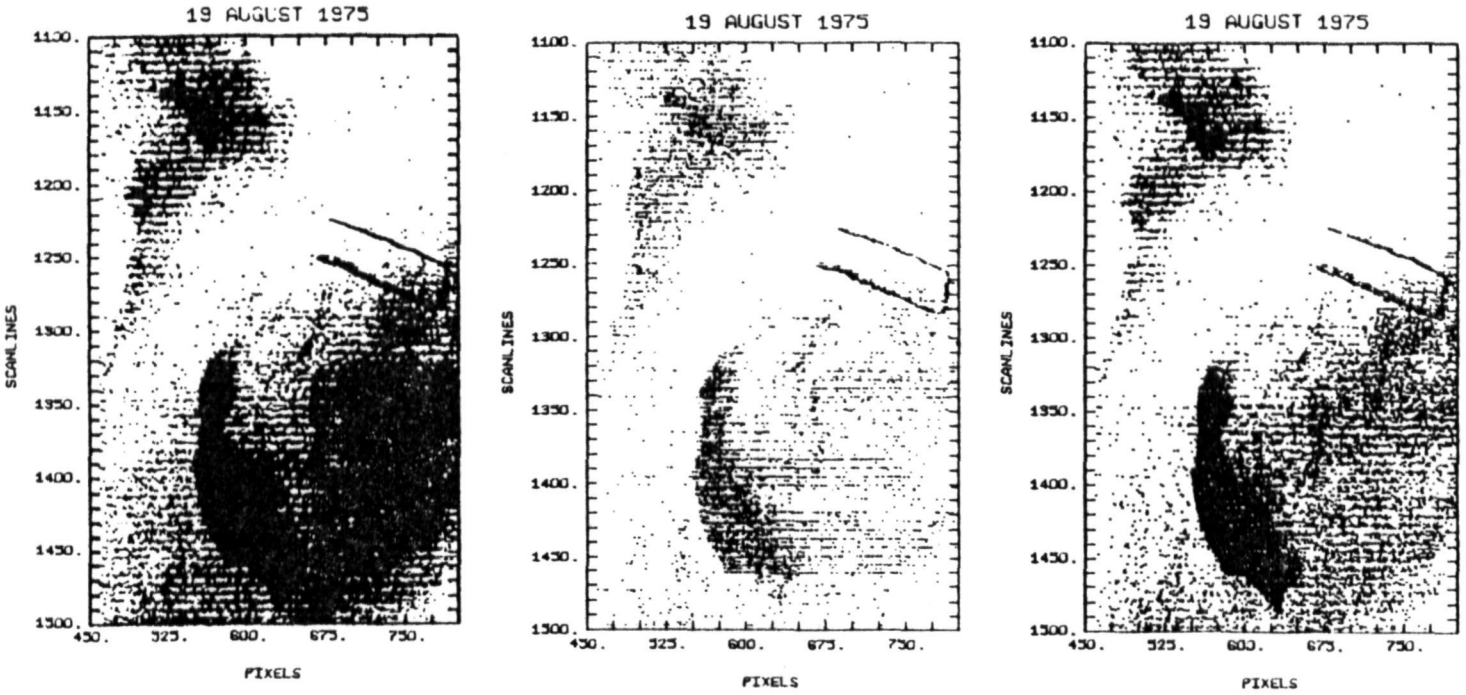


Figure 5. Sludge and acid waste plumes in the New York Bight, 19 August 1975 classified using a cylindrical limit. The scale is 1:400,000.

A. All pixels classified as either acid, sludge or clouds
B. Pixels classified as sludge
C. Pixels classified as acid

from those for the acid and sludge. The classification accuracy could be improved somewhat by more accurately representing the data distribution about the first eigenvector. This can be seen from the fact that the second eigenvectors are not parallel (they are separated by an angle of $\sim$27°), nor are the second eigenvectors along the direction of separation of the first eigenvectors. Thus, the situation illustrated by Figure 3 applies and classification could be improved. However, extending the analysis to account for the more accurate distribution would be considerably more expensive in computer time. For this case, it is probably unnecessary.

Another alternative exists for reducing the uncertainty in classification between clouds and acid. Most of the classification uncertainty exists for pixels near the origin, i.e., for pixels that are not much brighter than the clear water (see Figure 2). Thus, if the classification criteria are made more stringent in the vicinity of the origin the clarity of the results should improve. The simplest way of doing this is to reduce the classification limit. (This is equivalent to reducing the diameter of the circular cylinders in Figure 3.) However, this approach has the effect of reducing the number of pixels classified even in areas of relatively high radiance where there are few problems with classification. The results illustrated in Figure 5 already correspond to a limit of one standard deviation, σ_2, along the second eigenvector. Reducing this limit still further will clarify the results for areas of low radiance at the expense of significant losses in regions of higher radiance.

What is needed is a method of restricting the classification in the vicinity of the origin without seriously affecting the classification away from the origin. One way of doing this is to replace our cylindrical

classification limit with a conic limit. This "classification cone" would have the first eigenvector as its central axis and would open away from the origin--the point of the cone lying on the origin itself. This new geometry is illustrated in Figure 6. There are two classes shown in Figure 6. The first eigenvectors for each class, $\hat{e}_1$ and $\hat{e}_1'$, lie along the axes of the respective classification cones. The half angle of the cones are θ and θ', which are given by,

$$\tan \theta = a \frac{\sigma_2}{\sigma_1} \tag{5a}$$

$$\tan \theta' = a' \frac{\sigma_2'}{\sigma_1'} \tag{5b}$$

where σ_1 and σ_2 are the standard deviations of the training set data along the first and second eigenvectors respectively, and a and a' are weighting coefficients. Any pixel P falling within a given cone will then be grouped with the corresponding class, referring to Figure 6, since

$$\theta_p = \cos^{-1}(\vec{p} \cdot \hat{e}_1') \leq \tan^{-1}(a' \frac{\sigma_2'}{\sigma_1'}) = \theta' \tag{6}$$

Then the pixel P will be grouped with the primed class. If the pixel lies within the classification cone for more than one class then equation 2 may be used to choose the proper class.

Figure 7 shows the results of classifying acid, sludge and clouds using the conic limit. The coefficients used for these results were as follows: $a_{acid} = 0.7$, $a_{sludge} = 0.7$, $a_{clouds} = 3.0$. In descriptive terms this means that in order to be classified as either acid or sludge a pixel must lie quite close to the first eigenvector for the respective class, while the classification criterion for clouds is far less rigid.

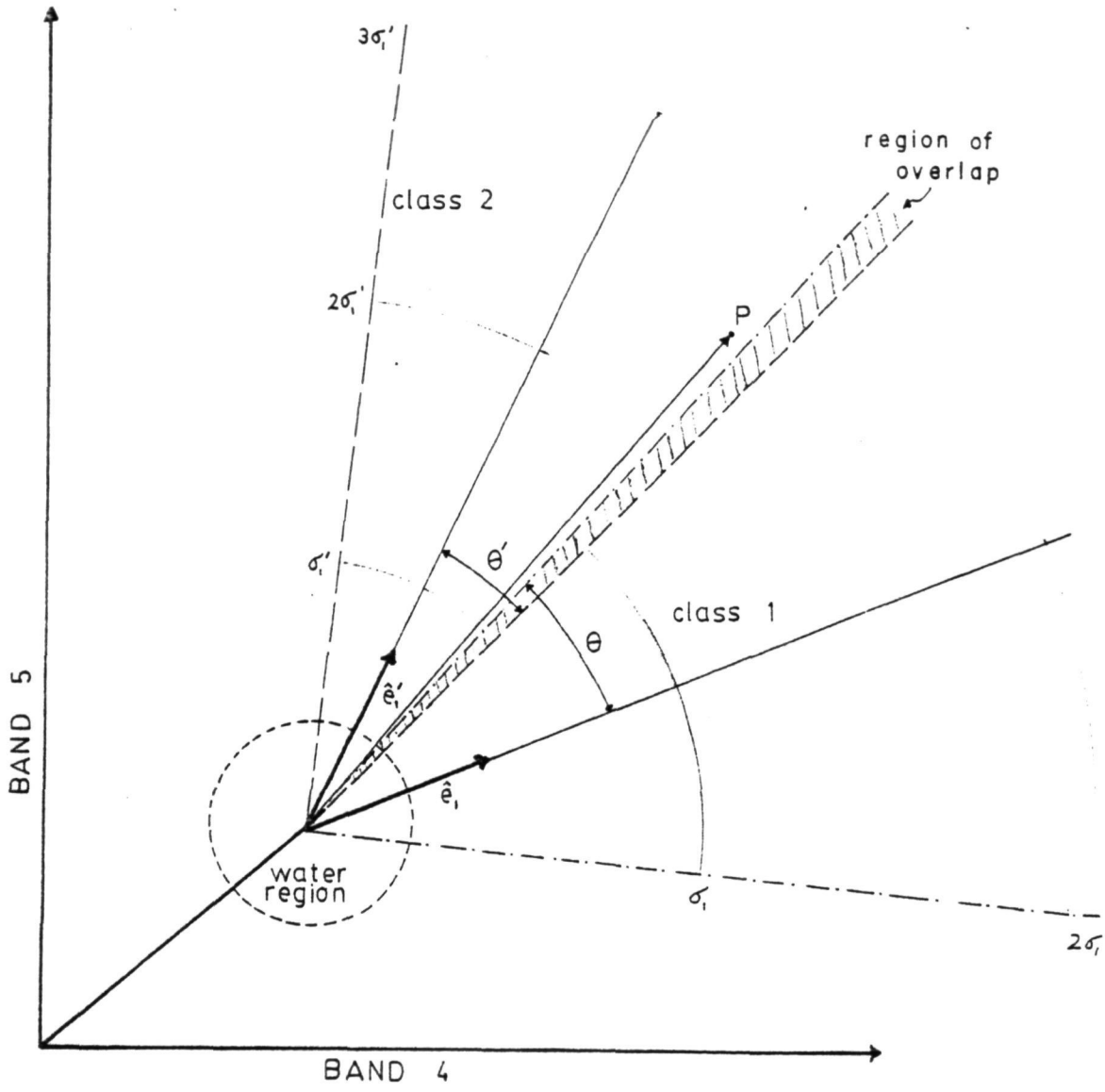


Figure 6. Two-dimensional geometry of the eigenvector classification scheme using a "classification cone."

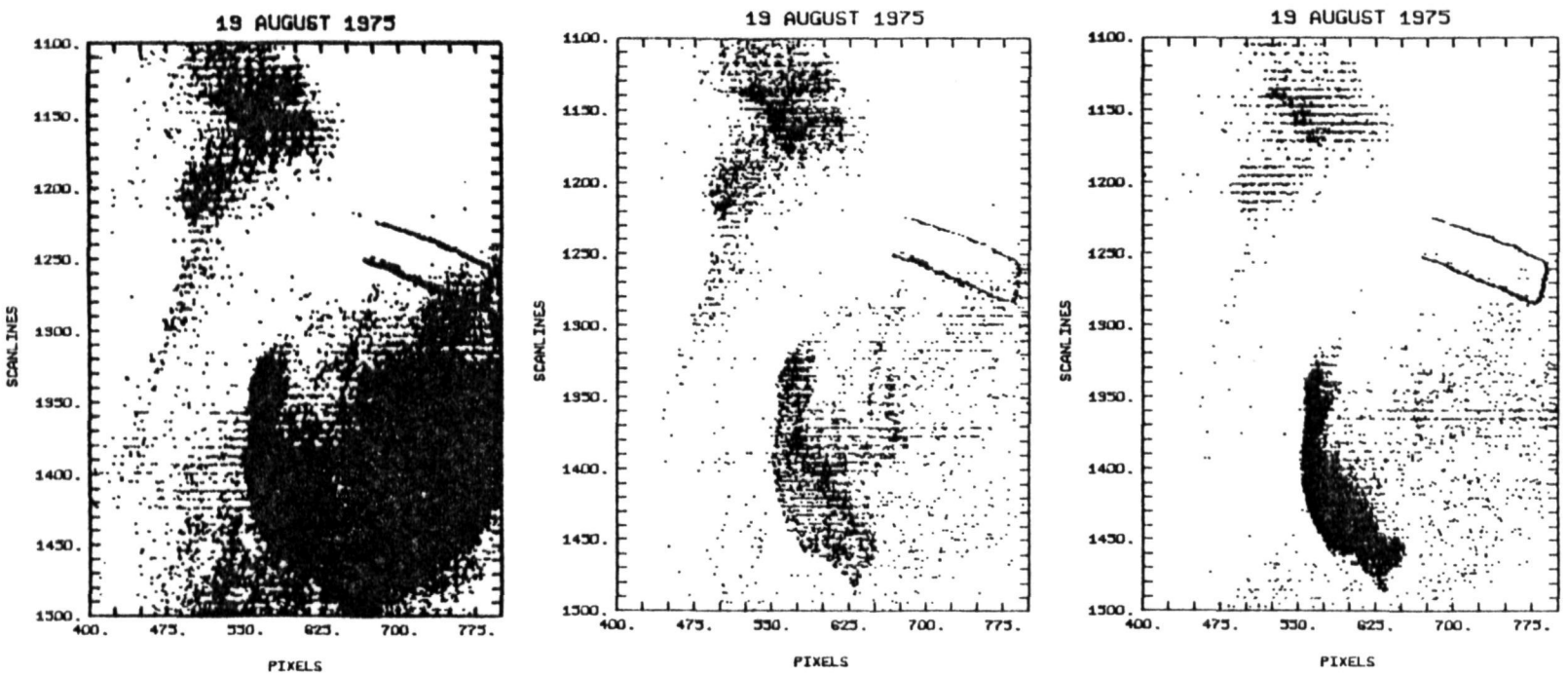


Figure 7. Sludge and acid waste plumes in the New York Bight, 19 August 1975, classified using a conic limit. The scale is 1:400,000.

A. All pixels classified as either acid, sludge or clouds
B. Pixels classified as sludge
C. Pixels classified as acid

As with Figure 5, Figure 7A shows all the pixels classified as acid, sludge and clouds; Figure 7B shows those pixels classified as sludge; and Figure 7C shows the pixels classified as clouds. The distinction between clouds and both the pollutants has changed significantly. This should not be too surprising since the main region of uncertainty was near the origin in the transformed color space. The misclassification of clouds as either acid or sludge has been reduced considerably. There are far fewer pixels classified as acid or sludge between the two acid dumps at the expense of an increased misclassification of acid and sludge as clouds. For the purpose of identifying pollutants, this sort of misclassification is preferable since there is greater certainty that those pixels classified as a pollutant are, in fact, the pollutant and not clouds. The distinction between the acid and sludge, on the other hand, is still rather poor although it is improved over the classification using a cylindrical limit. This time more of the sludge has been classified as sludge than as acid and vice versa, but there is still too much misclassification to consider the two classes to be separable.

If we consider the combined acid and sewage sludge as a pollutant class and look at these without the cloud overlay, a relatively dark area appears in the vicinity of pixel 675 and scan line 1350. The scatter of points between the two acid dumps is probably noise due to a misclassification of clouds. The dark feature, however, contains a much higher density of points than in the adjacent areas and is probably acid. This is an observation that cannot be verified by any other means. It is simply statistically unlikely to have such a dense cluster of points be misclassified.

2.4 Relating concentration and signal brightness

There is still more information available from this analysis which is not apparent from Figure 7. Klemas et al. (1978) suggested that it might be possible to relate the strength of signal (which corresponds to the magnitude of the position vector $\vec{p}$) for a given pixel to the concentration of the material in the water. The natural scaling of $|\vec{p}|$ is in terms of the standard deviation along the first eigenvector (see Figure 6). This is at least reasonable and is probably true in at least some situations. In any case we may easily separate the classified pixels by the strength of the signal. The acid class (the base plot in Figure 7) has been separated into two classes according to brightness. The results are presented in Figure 8. Figure 4A shows all the pixels whose position vectors fall within two standard deviations (σ_1) of the origin. Figure 4B shows all the pixels whose position vectors are greater than $2\sigma_1$.

The higher intensity pixels show essentially only acid, and show the features of the acid which which are most apparent to the eye in Figure 4. The seeming improvement in the classification, however, is a bit deceptive. As was mentioned earlier, the confusion between the clouds and the pollutants occurred primarily near the origin. The angle between the clouds and pollutants is large enough ($\sim$27°) for there to be little difficulty in making the distinction at higher intensities. In contrast, the uncertainty between the pollutants should exist at almost all intensities since the angle between their first eigenvectors is so small (<10°). Only about six pixels appear in the area of the sludge

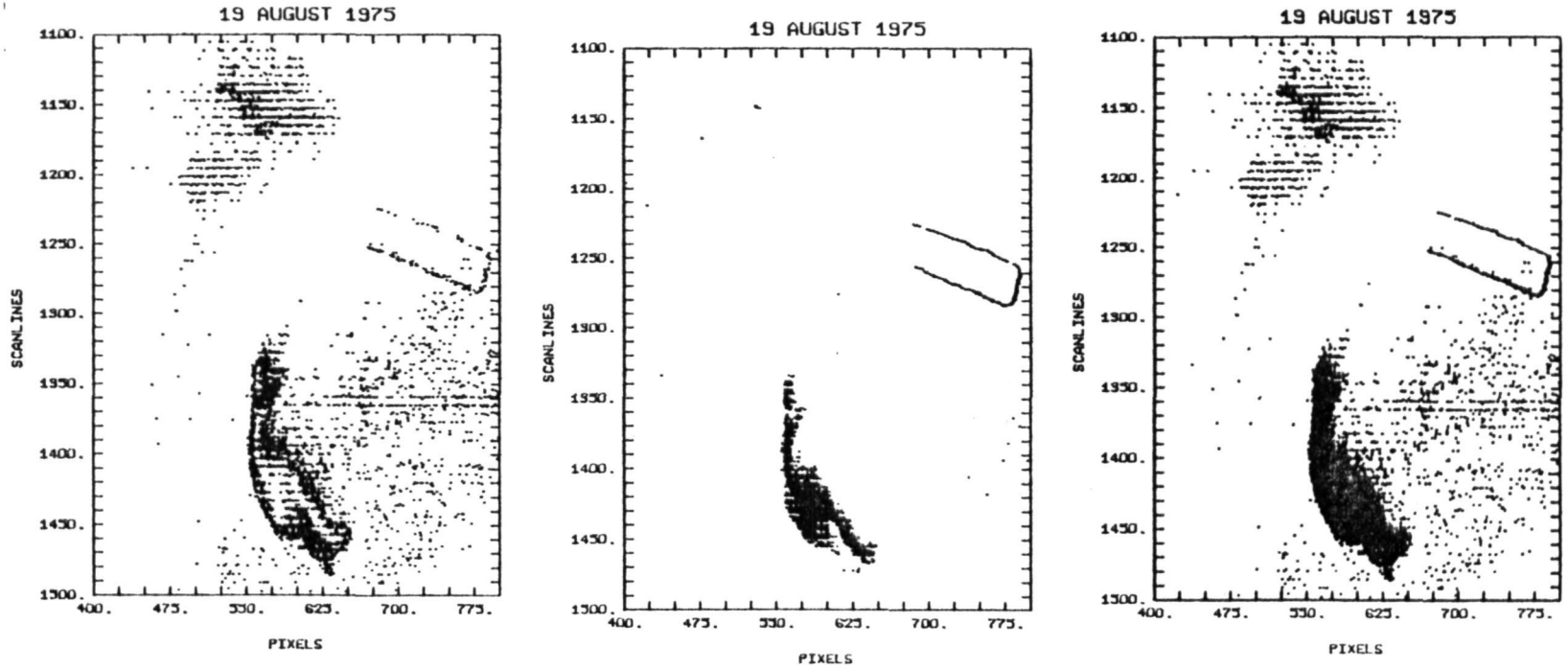


Figure 8. Separation of acid into "bright" and "dark" classes.

A. "Dark" acid
B. "Bright" acid
C. Total acid

dump, largely because very few of the pixels corresponding to sludge were bright enough. Had the sludge been more highly concentrated, implying a brighter signal, the uncertainty between sludge and acid would probably have been apparent in Figure 8. While it is far from certain that higher intensities are uniquely related to higher concentrations of the pollutant, the results in Figure 8 show the assumption to be at least qualitatively reasonable. The higher intensity pixels (base plot in Figure 8) form well defined patterns at the center of the acid waste sites; the patterns have rather sharp boundaries and there is little noise elsewhere in the scene. The lower intensity pixels form more diffuse patterns around the edges of the acid waste sites. The boundaries of these patterns are less well defined and there is much more noise in the plot. These are the same gross features one might expect to find in concentration distribution.

It must be emphasized that the higher intensity may also be related to the depth distribution of the material or chemical (or physical) changes in the material. It will be necessary to have far more information in order to speculate as to the physical meaning of the variation in intensity.

2.5 Comparison of New York and Delaware results: semi-automated classification

Thus far the eigenvector classification procedure has been applied only to single Landsat scenes. However, one of the advantages of the method is that if, as was suggested earlier, the spectra of the various targets is constant relative to the "clear" water standard, then it should be possible to use the method as a semi-automated system. In other words, the "clear" water standard must be chosen for each

Landsat Overpass	Region	ID Number	Time After Dump Completion	Pollutant
16 Aug 72[1]	N.Y.	1024-15071	--- ---	acid sludge
23 Oct 73	Del.	1457-15113	53 hrs. 36 min.	acid
15 Mar 74	Del.	1600-15031	6 hrs. 8 min.	acid
19 Aug 75	N.Y.	5122-14414	--- --- ---	fresh acid old acid sludge
19 Aug 75	Del.	5122-14420	during dump	acid
28 Aug 75	Del.	2218-14552	5 min.	acid
21 Oct 75	Del.	2272-15004	1 hr. 55 min.	acid
17 Nov 75	Del.	5212-14364	2 hrs. 41 min.	acid
19 Jan 76	Del.	2362-14540	39 min.	acid
24 Feb 76	Del.	2398-14531	3 hrs. 23 min.	acid
18 Apr 76	Del.	2452-14513	70 hrs. 19 min.	acid

Table 2. Landsat imagery used for comparison of eigenvectors.

different scene, but the eigenvectors describing a particular target should be essentially the same from scene to scene. This is actually the case to a remarkable extent.

To demonstrate this, eigenvectors for acid, sediment, sludge, and clouds were calculated for several Landsat scenes of the New York Bight and Delaware coast. Table 2 lists the scenes used for this work along with the age of the acid or sludge wastes, if known. The first eigenvectors for each target for every scene are listed in Table 3.

There is a qualitative similarity among the eigenvectors for each class and a dissimilarity between classes except for the sludge and acid. The comparison of the first eigenvectors can be quantified by considering the angular separation. The angles between each set of first eigenvectors were calculated and are presented in Table 4.[1] Earlier results (Klemas et al., 1978) suggested that the intraclass angles would generally be less than 10° while the interclass angles would generally be greater than 20°. At that time it seemed likely that there would be little difficulty in distinguishing among clouds, acid/waste and sediment using some mean eigenvectors. With this expanded data base the results allow for less optimism. Although clouds are still clearly distinguishable from everything else, it appears that the sludge is largely indistinguishable from acid and that, at least in some cases, acid and sediment will not be clearly separable using mean eigenvectors.

There is <u>still</u> reasonably good separation between acid and sediment in <u>individual scenes</u> (boxed values in Table 4). For instance, the separation between the first eigenvectors for acid and sediment is 23.8° for 19 January 1976 and 15.6° for 23 October 1973. These angles are

	Place	Band 4	Band 5	Band 6	Band 7
ACID					
16 Aug 72	N.Y.	0.6508	0.7347	0.1912	0.0060
23 Oct 73	Del.	0.8532	0.5088	0.1094	0.0348
15 Mar 74	Del.	0.8969	0.4158	0.1392	0.0334
19 Aug 75	N.Y.	0.7872	0.5755	0.2184	0.0382
19 Aug 75	Del.	0.7409	0.6491	0.1705	0.0246
28 Aug 75	Del.	0.7707	0.6081	0.1906	0.0032
21 Oct 75	Del.	0.8526	0.5127	0.1011	0.0000
17 Nov 75	Del.	0.9096	0.4026	0.1021	0.0142
19 Jan 76	Del.	0.8221	0.5248	0.2205	0.0147
24 Feb 76	Del.	0.8413	0.5264	0.1228	-0.0025
18 Apr 76	Del.	0.9802	0.1962	0.0026	-0.0269
SLUDGE					
6 Aug 72	N.Y.	0.4136	0.6967	0.5632	0.1616
19 Aug 75	N.Y.	0.8148	0.5021	0.2846	0.0551
CLOUDS					
23 Oct 73	Del.	0.5900	0.5837	0.5169	0.2096
15 Mar 74	Del.	0.5327	0.5871	0.5700	0.2157
19 Aug 75	N.Y.	0.5587	0.5579	0.5861	0.1820
19 Aug 75	Del.	0.6499	0.5435	0.4958	0.1906
17 Nov 75	Del.	0.5291	0.6000	0.5560	0.2258
19 Jan 76	Del.	0.4264	0.5879	0.6319	0.2707
SEDIMENT					
23 Oct 73	Del.	0.7198	0.6145	0.3203	0.0411
21 Oct 75	Del.	0.6613	0.6685	0.3404	0.0006
17 Nov 75	Del.	0.7021	0.6255	0.3397	0.0223
19 Jan 76	Del.	0.5268	0.7489	0.4021	0.0000
24 Feb 76	Del.	0.5129	0.7167	0.4725	0.0028

Table 3. First eigenvectors for several targets and several Landsat scenes.

		ACID										CLOUDS						SEDIMENT					SLUDGE
		23 Oct 73	15 Mar 74	19 Aug 75	19 Aug 75	28 Aug 75	21 Oct 75	17 Nov 75	19 Jan 76	24 Feb 76	18 Apr 76	23 Oct 73	15 Mar 74	19 Aug 75	19 Aug 75	17 Nov 75	19 Jan 76	23 Oct 73	21 Oct 75	17 Nov 75	19 Jan 76	24 Feb 76	19 Aug 75
ACID	23 Oct 73	0.2																					
	15 Mar 74	6.7	3.8																				
	19 Aug 75	8.2	(12.3)	0.5																			
	19 Aug 75	(10.9)	(16.5)	5.7	0.6																		
	28 Aug 75	8.9	(13.9)	3.2	3.3	0.8																	
	21 Oct 75	2.0	7.3	8.8	(11.0)	8.8	0.2																
	17 Nov 75	7.0	3.7	(13.9)	(17.6)	(15.1)	7.1	0.7															
	19 Jan 76	6.8	9.3	3.7	9.0	5.8	7.1	(11.0)	0.8														
	24 Feb 76	2.6	7.9	7.3	9.6	7.3	1.6	8.2	5.8	0.5													
	18 Apr 76	20.7	16.2	27.9	31.4	28.9	20.5	14.0	24.6	21.8	0.3												
CLOUD	23 Oct 73	30.2	31.7	22.9	24.6	24.7	31.3	34.2	24.7	29.9	46.5	0.7											
	15 Mar 74	34.6	36.1	27.1	28.6	28.7	35.6	38.5	29.0	34.2	50.7	4.6	0.9										
	19 Aug 75	33.8	34.8	26.4	28.3	28.1	34.8	37.3	27.9	33.4	49.2	4.9	3.1	0.5									
	19 Aug 75	26.9	27.8	20.0	22.6	22.1	28.0	30.3	21.3	26.7	42.3	4.5	8.5	7.4	0.8								
	17 Nov 75	34.4	36.0	26.9	28.1	28.4	35.4	38.5	28.9	34.0	50.8	4.4	1.3	4.2	8.6	0.7							
	19 Jan 76	42.1	43.6	34.7	35.7	36.2	43.2	46.1	36.6	41.8	58.2	12.6	7.7	9.7	15.9	7.8	0.4						
SED	23 Oct 73	[15.6]	18.7	(7.3)	(9.0)	(8.3)	16.0	20.7	(9.8)	14.5	34.3	16.8	20.7	19.9	14.4	20.5	28.2	0.2					
	21 Oct 75	19.7	23.3	(11.6)	(10.9)	(11.2)	[19.8]	25.1	(14.2)	18.2	38.7	17.0	20.1	19.6	15.8	19.9	27.1	5.2	0.7				
	17 Nov 75	17.2	20.3	(9.0)	(10.0)	(9.5)	17.5	[22.3]	(11.3)	15.9	35.8	16.3	20.0	19.2	14.3	19.8	27.4	1.9	3.5	0.8			
	19 Jan 76	29.0	32.9	21.0	19.1	20.3	29.1	34.7	[23.8]	27.5	48.4	17.1	18.3	18.6	18.4	17.9	23.2	14.5	9.6	12.9	0.6		
	24 Feb 76	31.3	34.5	23.1	22.2	22.9	31.5	36.5	25.6	[29.9]	49.9	15.0	15.4	15.5	16.7	15.3	20.1	16.0	11.7	14.3	4.5	0.3	
SLUDGE	19 Aug 75	10.4	11.2	5.9	11.6	9.0	11.2	13.3	4.5	10.1	26.3	21.2	25.5	24.2	17.4	25.4	33.1	8.7	13.7	10.2	23.1	24.2	0.5

Table 4. Angular separation (in degrees) between first eigenvectors. The intraclass angles are generally less than 10°; the interclass angles are generally greater than 20°. The diagonal elements are non-zero because of the uncertainty in the angle calculation. The results for 16 August 1972 have been omitted since the original data was suspect.

probably large enough to allow for good separation of the two classes for each scene separately.

The angular separation of acid eigenvectors for both days is 6.8° which is quite good for classification purposes. However, the angular separation of sediment eigenvectors for both days is 14.5°--practically the same as the separation between the acid and sediment eigenvectors for the 23 October 1973 data. It seems unlikely that a semi-automated classification procedure using mean eigenvectors would be useful in this case.

2.6 The possibility and implications of nonlinear spectral variations

The failure at the present time to find a set of typical or mean eigenvectors which could be used in a semi-automated classification system could have been due to the arbitrariness in choosing the "clear" water origin. Admittedly, this is a weak point in the present procedure; a less arbitrary method of choosing the origin might be expected to give better results.

Several attempts were made to improve the method of defining the "clear" water origin. All failed. However, the manner of the failure suggested that the variation in the first eigenvectors for a single class is not simply random. This is best illustrated with the acid. If we recalculate the eigenvectors for each date without reference to the predetermined clear water standard, the new vectors will describe the variability in that sample. If the variation of the spectral signatures were strictly linear then these new vectors should be very nearly identical to the old vectors. In fact, they are not at all similar either to the old vectors or to each other. Furthermore the variations

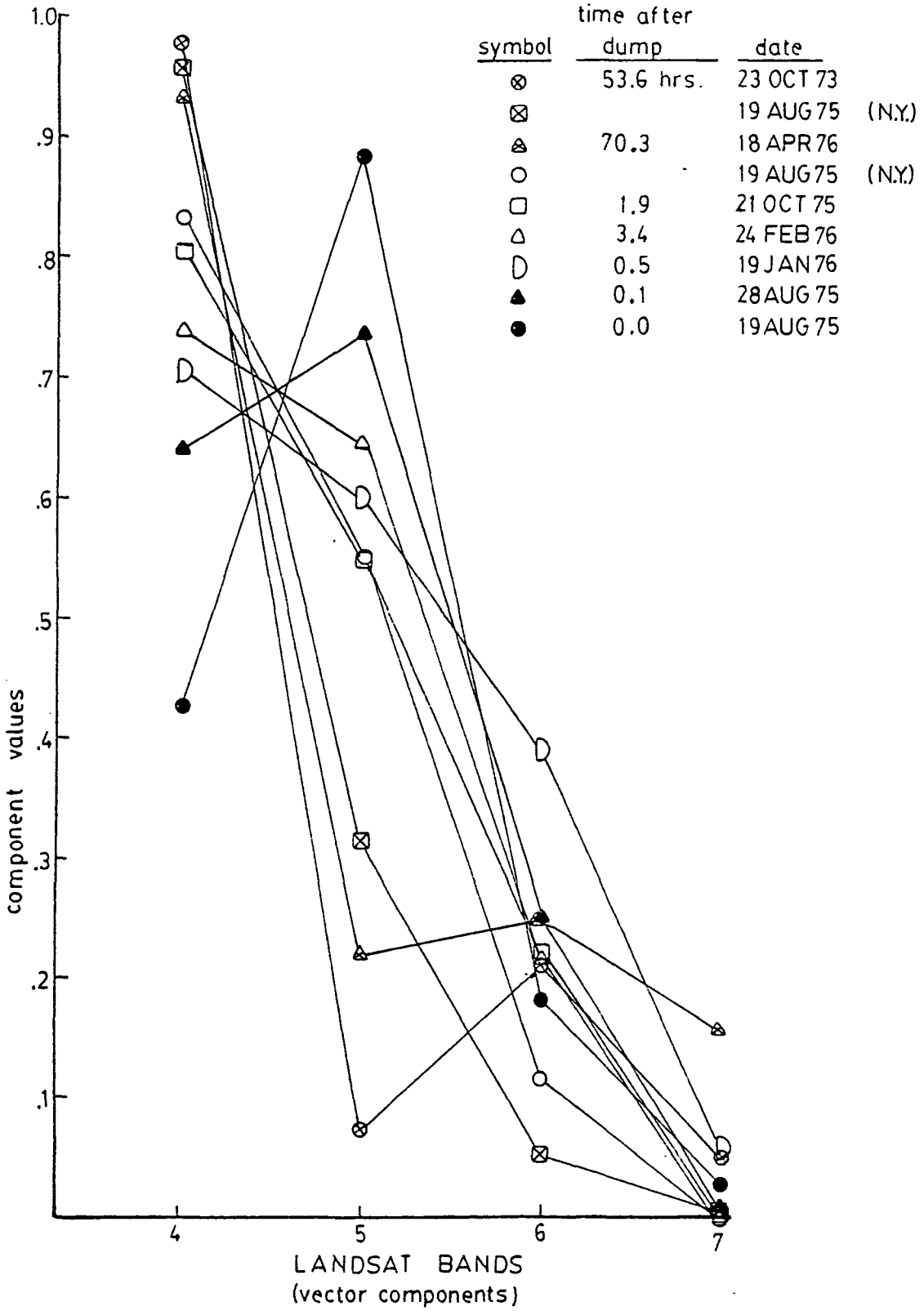


Figure 9. Acid eigenvectors calculated without reference to a clear water standard.

seem to be quite systematic. Figure 9 shows these new eigenvectors in graphic form. These vectors characterize the variance in the spectral characteristics within a given target for a particular day. On days when the acid waste is rather old (23 October 73, 19 August 75 and 18 April 73) almost all the variance is in the blue. For relatively fresh dumps the variance in band 4 is reduced and the variance in band 5 increases significantly. For the two freshest dumps (19 August 75 and 28 August 75) the variance in band 5 actually exceeds that in band 4. Thus there is some relationship between the age of the dump and its spectral characteristics. Indeed, Bowker (1977) has already shown this to be the case. The physical source of this effect is not yet clear. There are several possibilities: chemical weathering, physical changes (i.e., flocculation), settling, or some combination of the three.

Although it is not unlikely that all three mechanisms play some role, a reasonable argument can be made that settling is the dominant mechanism. The argument is quite simple. Regardless of the apparent color of an underwater object near the surface, the object will appear to be blue, blue-green, or green as it is moved deeper into the water column. The actual color will depend on the kind and the amount of scattering and absorbing material in the water. Nonetheless, the fact that water strongly absorbs red light means that the apparent color of an object will become more blue as it sinks.

This is nicely consistent with the vectors in Figure 9: only in the freshest dumps is there enough material near the surface to give a strong signal in the red and the older the acid, the less signal in the red, implying that most of the iron-acid floc is settling. Thus, it is possible that there is a direct relationship between the "blueness" of a

pixel and some average depth of the iron-acid floc.

If this supposition is correct, then the trend towards "blueness" with age should also be apparent in the original eigenvectors. This is, in fact, the case. Referring back to Table 3, the day for which the acid eigenvector is weakest in band 6 compared to other eigenvectors and strongest in band 4 is 18 April 1976. This acid dump was also the oldest (70.3 hrs.) and was the most marked anomaly in Table 4. The eigenvectors for the freshest acid dumps (19 August 1975 [Del], 28 August 1975, 19 January 1976, and 19 August 1975 [N.Y.]) are those with eigenvectors which are strongest of the acid eigenvectors in band 6 and weakest of all in band 4.[2] It is worthy of note that it is these same days which present the greatest obstacle to the use of mean eigenvectors: 1) the largest acid-acid angles (circled values in Table 4) are consistently between the acid eigenvector for one of these days and an acid eigenvector for one of the other days; 2) the smallest acid-sediment angles (also circled in Table 4) are consistently between the acid eigenvector for one of these days and a sediment eigenvector.

The most curious aspect of all this is that the fresher acid, which, presumably, is nearer the water surface, is harder to distinguish from sediment than the older acid (farther below the surface). To see how this might happen it is useful to combine the presumption of a relationship between the "blueness" of a signature and some averaged depth of the target substance with the premise of a direct relationship between intensity and concentration. If the variations of the spectral reflectance characteristics of a substance in the water depend on concentration and depth as described above then the eigenvectors which characterize the substance in water will vary from scene to scene as

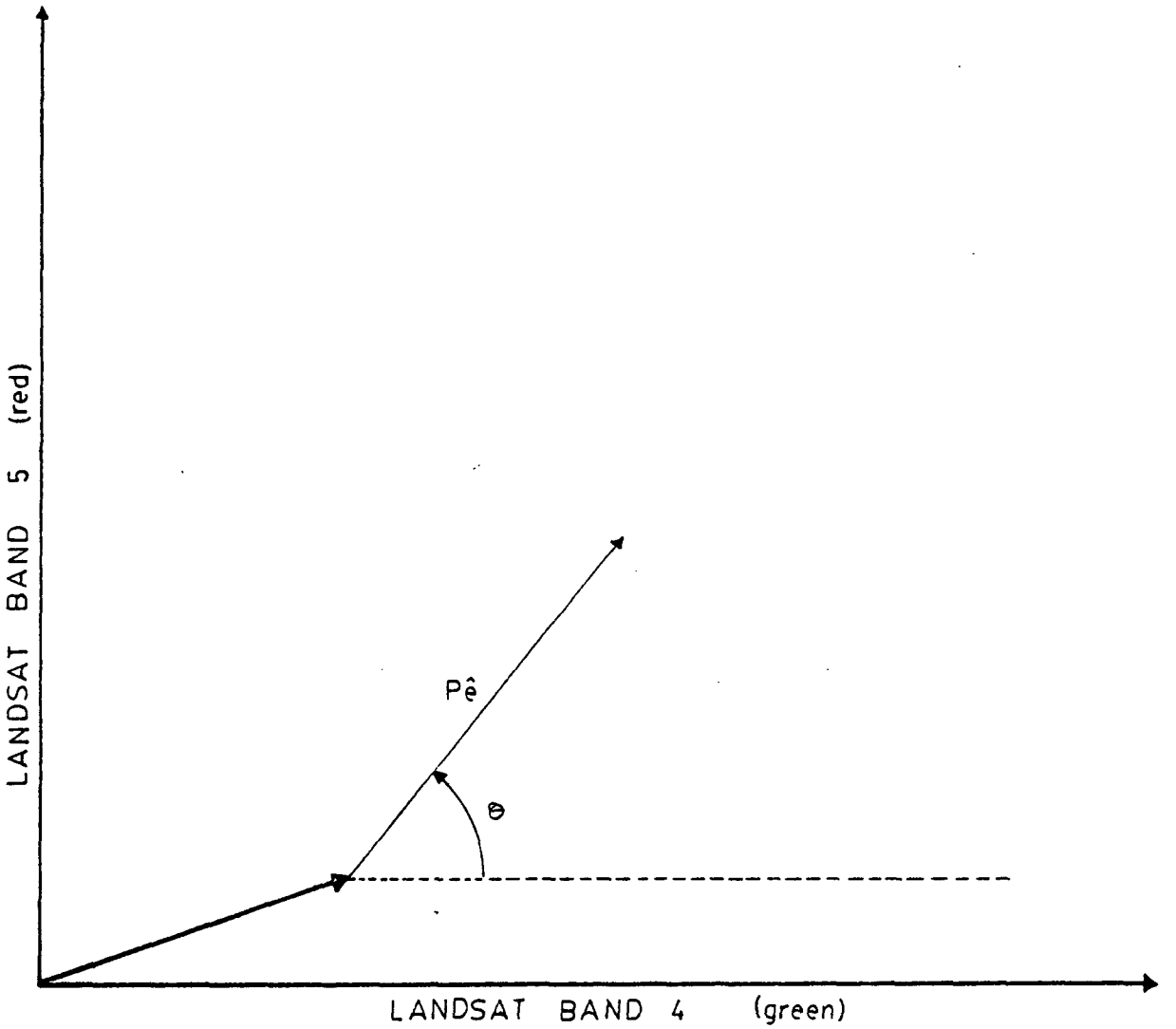


Figure 10. Suggested modes of variation of pollutant signatures in water. Variation in the radial direction relates to concentration; variation in the σ direction relates to "average" depth.

well. The manner of this variation is illustrated in Figure 10. Here the radial distance, p, of a pixel from the new origin, 0, along eigenvector $\hat{e}_1$ is related to concentration while a decrease in the angle θ is related to an increase in the "mean" depth.[3]

Assume for a moment that there existed a data set of noise-free spectra corresponding to all possible combinations of concentration and "mean" depth for a particular substance in water. For a given "mean" depth, a variation in concentration would shift the spectra toward or away from the "clear" water origin. For a given concentration, on the other hand, a change in the "mean" depth would cause a shift in the spectrum both in the θ and radial directions; as the "mean" depth increases, the signature becomes bluer and decreases in intensity due to increased attenuation by the water. The total distribution of points for two substances--acid waste and sediment--might appear as in Figure 11. Each region has been subdivided into three subregions corresponding to low, medium, and high concenttrations. In this schematic representation there is a region of overlap for the sediment and acid. The confusion is between all concentrations of sediment with large "mean" depths and low, near surface concentrations of acid--the sort of distribution that would most likely be found in a fresh acid dump.

Qualitatively, at least, the assumptions of a dependence of the spectra on concentration and "mean" depth lead to the same results observed in the eigenvector analysis results. These assumptions also lead to two major implications for ocean color analysis in general and pollution detection in particular:

1) There are at least two degrees of freedom in the systematic variability of the spectral reflectance characteristics.

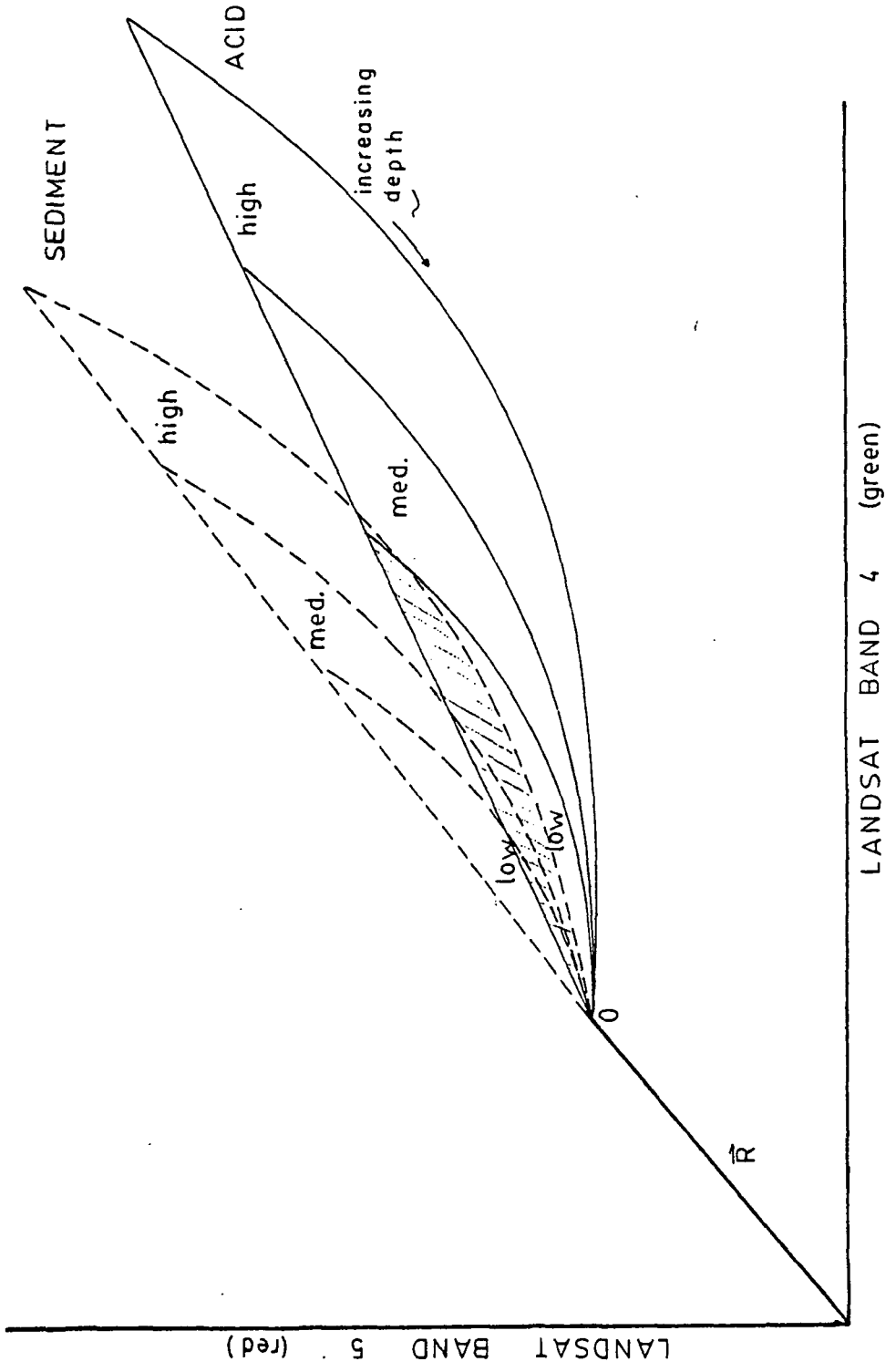


Figure 11. Hypothetical distribution patterns of points for two substances in water showing the relationship of position to concentration and depth of the substances. The shaded area shows the area of overlap, i.e. the region in which the two substances are indistinguishable.

Thus, the earlier assumption of essentially random variation about the first eigenvector is incorrect. Neither the cylindrical nor the conic limits are really appropriate for describing the variability in the data although, as has been shown above, they will both work well enough for classification in individual scenes.

2) There is a physical limit to the spectral separability of any two substances in the water--even classes as distinct as acid and sediment. Random noise may degrade the classification accuracy, but even if the data were noise-free some uncertainty in the classification would remain.

III. OTHER MEANS OF MONITORING WATER QUALITY IN THE NEW YORK BIGHT

Landsat was used for this study because of the large amount of data available, the good spatial resolution and the adequate spectral resolution. However, Landsat is not the ideal observational tool. The three major objections to the use of Landsat data for coastal water quality monitoring are that 1) the spectral response is not ideal for observations of water color; 2) the gain and dynamic range of the Landsat sensors are poorly suited to the purpose of water color observation; and 3) Landsat returns to cover the same area only once every 18 days making it difficult to use for waste dumps which will be apparent for a few days at most.

The recently launched Coastal Zone Color Scanner (CZCS) aboard Nimbus G will overcome some of these objections. The spectral response, gain and dynamic range of this instrument are intended specifically for chlorophyll observations, but should be quite useful for other ocean color observations. Indeed, CZCS should be nearly ideal for observing

large scale patterns of ocean color: distinguishing among large water masses, observing sediment plumes of major estuaries or detecting large plankton blooms.

The CZCS, however, lacks one feature available in Landsat imagery which may be essential for water quality monitoring in coastal areas: good spatial resolution. Although waste dumps in coastal waters can be rather large they are not usually large enough to provide a statistically significant sample for the course (800 m) resolution of the CZCS. For example, the fresh acid waste in Figure 4 is roughly 4 km x 8 km. The other two waste sites cover about the same area. This area would be covered completely by 50 pixels of CZCS. In the case of the fresh acid waste most of these pixels would be for the water. Even for the older waste sites where all 50 pixels might cover some of the waste, it would be difficult to distinguish systematic variations from random noise; one would need to use the entire sample as a training set. This is not to say that CZCS data should be ignored. The spectral information could be quite valuable. However, the CZCS will probably not be adequate alone for water quality monitoring. It would be much more worthwhile to use CZCS data in conjunction with Landsat data. The spectral and spatial resolution of the Landsat data is good for identifying and locating pollution dumps. The spectral characteristics could be refined using the CZCS data.

A major advantage to the CZCS is its relatively short return coverage. The satellite may cover the same scene during the same part of its orbit only about once every six days. However, any given ground point should be covered nearly once a day; the nadir point of the consecutive scenes will not be the same and distortion may be a problem,

but nearly daily coverage should be possible. This is an important factor for pollution monitoring in the coastal zone because of the dynamic character of coastal regions. If the CZCS can be used to identify pollutants by their spectral reflectance characteristics, then it should be possible to track the same pollutant dump over a period of several days. Of course, for the pollutant to be spectrally identifiable, the pollutant plume must be large enough to cover several pixels. The acid and sludge wastes discussed here cover large enough areas for tracking to be possible.

It would be difficult to improve upon satellite coverage using aircraft for monitoring of large coastal regions such as the New York Bight. An aircraft system would simply not be able to cover such an area as effectively. However, aircraft systems would be particularly useful for monitoring smaller areas such as legally defined dump sites, or for short-term tracking of a particular dump.

IV. SUMMARY

The major points of this study can be summarized as follows:

1) The eigenvector classification scheme was shown to be effective at distinguishing between pollutants (acid waste and sewage sludge) and clouds in the New York Bight.
2) The sewage sludge and acid waste were found to be spectrally indistinguishable in the test scene of the New York Bight.
3) Classification accuracy was improved by replacing the cylindrical limit with a conic limit.
4) Semi-automated classification of a Landsat scene using a predetermined set of mean eigenvectors does not appear feasible at this time.

5) It was demonstrated qualitatively that it is reasonable to relate higher intensity with higher concentration and that there is a systematic variability in the first eigenvector which could be related to a mean depth of the pollutant.

This last point is perhaps the most important finding of this study. Although the variability in the direction of the first eigenvector makes it difficult to use the algorithms presented here as a semi-automated classification technique, the fact that the variability is systematic suggests that it may be possible to identify a target as a pollutant and to make some estimate of the concentration and mean depth of the pollutant. Eigenvector analysis should still be useful for this although the criteria for classification would change.

The sense of the systematic variation in the spectral reflectance characteristics also implies a physical limit on the ability to distinguish between any two substances in the water. This is an encouraging result in that it is a consequence of the depth dependent variation and is precisely what should be expected.

The Landsat data shows surprisingly good definition of pollutants in the areas studied. It seems likely that further study along the lines of the present work could improve the results or at least better define the limits of the technique as used with the Landsat data. If the spectral resolution, dynamic range and return coverage could be improved upon there would be significant gains in terms of pollution monitoring. The CZCS provides all of these improvements and should be ideal for observation of ocean color; however, the loss of spatial resolution with CZCS will reduce its utility for pollution monitoring. Thus, Landsat should not be ignored as an instrument for pollution monitoring in the coastal zone.

References

Bowker, D. E. and W. G. Witte (1977) The use of Landsat for monitoring water parameters in the coastal zone. - Proceedings: AIAA 1977 Joint Conference on Satellite Applications to Marine Operations. New Orleans, La., Nov. 15-18.

Bowker, D. (1978) personal communication. - NASA Langley Research Center, Hampton, Va. 23365.

Deutsch, M., A. E. Strong and J. E. Estes (1977) Use of Landsat data for the detection of marine oil slicks. - Offshore Technology Conference, pp. 311-318, Houston, Tx.

Deutsch, M., A. E. Strong and G. Rabchevsky (1978) Detection of floating oil slicks by Landsat. - Photogrammetric Engineering and Remote Sensing.

Hammack, J. C. (1977) Landsat goes to sea. - Photogrammetric Engineering and Remote Sensing, 43: 683-691.

Johnson, R. W. (1978) Mapping of chlorophyll a disttributions in coastal zones. - Photograpmmetric Engineering and Remote Sensing, 44: 617-624.

Johnson, R. W., I. W. Duedall, R. M. Glasgow, J. R. Proni and T. A. Nelson (1977) Quantitative mapping of suspended solids in waste-water sludge plumes in the New York Bight Apex. - Journal Water Pollution Control Federation, October 1977, pp. 2063-2073.

Klemas, V., M. Otley, W. Philpot, and R. Rogers (1974a) Coastal and estuarine studies with ERTS-1 and Skylab. - Remote Sensing of Environment, 3: 153-174.

Klemas, V., D. Bartlett and R. Rogers (1976) Skylab and ERTS-1 investigations of coastal land-use and water properties in Delaware Bay. - Progress in Astronautics and Aeronautics, 48: 351-371.

Klemas, V., W. Philpot and G. Davis (1978) Determination of spectral signatures of substances in natural waters. - Final Report, NASA Grant, NSG 1149, NASA Langley Research Center, Hampton, Va.

Rouse, L. J. and J. M. Coleman (1976) Circulation observations in Louisiana bight using Landsat imagery. - Remote Sensing of Environment, 5: 55-66.

Strong, A. E. (1978) Chemical whitings and chlorophyll distributions in the Great Lakes as viewed by Landsat. - Remote Sensing of Environment, 7: 61-72.

Footnotes

1. The results for 16 August 1972 are included in Tables 2 and 3 for the sake of completeness. These results are suspect since the Landsat sensors were malfunctioning intermittently. Data which was obviously bad was excluded from the training sets; however, it is quite likely that some systematic error remains. For this reason the results from 16 August 1972 do not appear elsewhere in the paper.

2. It is important to keep in mind the distinction between the eigenvector components and the intensity of a pixel. The fact that an eigenvector for a particular class is dominated by the band 4 component does not imply that a pixel in that class has the greatest intensity in band 4. The eigenvector refers only to the variability in the intensity in band 4 and is unrelated to the absolute intensity. Thus, mean spectrum of a particular class might have the highest intensity in band 5 but very little variation about that mean while the mean for band 4, although having a lower intensity, might show much greater variability. The eigenvector for this class would then be dominated by the band 4 component rather than the band 5 component.

3. "Mean" depth--this term is used here very loosely. There is no intent to imply any particular vertical distribution of the acid.