MAPPING WATERMASS BOUNDARIES USING FLUOROSENSING LIDAR[1]

Charles C. Sarabun, Jr.
The Johns Hopkins University
Applied Physics Laboratory
Laurel, Maryland

SUMMARY

An initial application of multispectral LIDAR data from the NASA Airborne Oceanographic Lidar (AOL) to the mapping of watermass boundaries is presented. The approach uses the multispectral lidar data from the fluorosensing mode in a cluster analysis to define water types. Individual data points are classified as to parent water type(s) and then plotted in plan view to show the watermass boundaries and mixing regions. The methodology was applied to the AOL data from the 23 and 25 June SUPERFLUX overflights. The results are compared to salinity-mapping radar results from the same region.

INTRODUCTION

The regions where two or more different watermasses meet are usually characterized by a high degree of spatial and temporal variability. They are often the sites of locally intense mixing and interacting smaller-scale phenomena such as intrusions and interleavings. Field studies of such regions are difficult because of the multiplicity of length and time scales present, and conventional shipborne hydrographic techniques often cannot provide adequate spatial resolution or data of a sufficiently synoptic character. Remote sensing systems have the capability to survey large areas on a nearly synoptic basis and many of these systems are capable of providing the needed spatial resolution. Since investigators have shown that watermasses with distinct physical origins and histories often have a distinct biochemical makeup as well, (refs. 1,2,3,4), remote sensing systems which measure biochemical parameters could be employed to characterize water types present in a survey region, and to map their horizontal structure. One such system is the Airborne Oceanographic Lidar (AOL) operated by NASA/ Wallops Flight Center. This system actively irradiates the water column with light at a fixed wavelength, and measures the intensity of the return signal. Operated in the fluorosensing mode, the system measures a wideband spectrum of laser-stimulated fluorescence from the biochemical constituents of the water, such as chlorophyll and other light-absorbing pigments.

[1]This work was partially supported under Navy Contract N00024-81-C-5301.

To use the fluorosensing AOL data for classifying water types, one would ideally like to use all of the information available in the return spectra simultaneously. A convenient technique for dealing with data vectors consisting of many measured parameters is cluster analysis. Cluster analysis is a method of dividing a total data set into groups, or clusters, using all of the measured parameters. In this paper we describe an initial application of such a technique to AOL fluorosensing data. The data were obtained on 23 and 25 June as part of an examination of the application of aircraft remote sensing to the study of the Chesapeake Bay outflow. The AOL operation and data set are described elsewhere in these proceedings (ref. 5).

The data sets used in the analysis consisted of discrete spectral samples in twenty bands, plus simultaneously recorded data from a thermal infrared scanner. A sample AOL spectrum is shown in Figure 1. The 23 June data set consisted of 4053 sample spectra taken along the flightlines shown in Figure 2a. The 25 June data set consisted of 5410 sample spectra along the flightlines shown in Figure 2b. The data were smoothed along each flightline and rescaled to the interval [-1, +1] so that subsequent processing would not be dominated by any single band.

ANALYSIS METHODOLOGY

Analysis of the AOL data proceeded in three stages, (1) empirical orthogonal function (EOF) decomposition to reduce the dimensionality of the sample spectra, (2) cluster analysis to define basic water types and (3) projection of each data point on the characteristic vectors of the water types to determine the spatial distribution of each water type. Each of these processing stages is discussed below.

EOF Analysis

Because many of the spectral peaks seen in Figure 1 cover several adjacent spectral bands, the AOL data were subjected to an EOF decomposition to define a new orthogonal basis for the spectrum. This new basis is computed from the covariance matrix formed by using the entire set of spectral samples to compute the covariance between bands. The eigenvectors of this matrix form the new basis, and the eigenvalues represent the amount of the total variance in the data accounted for by the associated eigenvector (ref. 6). In practice, the first several eigenvalues accounted for almost all of the variance in the data. This fact allowed the dimensionality of the problem to be reduced in subsequent analysis by retaining only major contributions to the variance in the transformed spectra. The reduced, transformed spectra were then used in the cluster analysis (in what follows, sample spectrum means the transformed, reduced spectrum).

Cluster Analysis

The cluster analysis provides a means for dividing the total set of sample spectra into subsets, called clusters, where the sample spectra in each cluster are somehow similar. These clusters are then assumed to represent characteristic water types present in the surveyed region. There exists a variety of similarity (and dis-similarity) measures which could be used to subdivide the data (refs. 7,8). The similarity criterion used in the examples presented in this paper is essentially a distance measure in a space whose axes are the spectral bands of the sample spectra. A distance measure was selected to facilitate the assignment of percentages in the final stage of processing.

The distance measure used here is the ℓ_∞ norm where the distance, d_{ik}, between any two points $\vec{x}_i$ and $\vec{x}_k$ is defined as

$$d_{ik} = \underset{j}{\text{Max}} |x_{ij} - x_{kj}| \tag{1}$$

where j denotes a spectral band. The data are then arbitrarily divided into a given number of clusters, say L, and the centroid $\vec{Y}_k$ of the kth cluster is computed as

$$Y_{kj} = \frac{1}{m_k} \sum_{i=1}^{m_k} x_{ij} \tag{2}$$

where m_k is the number of sample spectral in the kth cluster and j is the spectral bands. The sum of the distances, E_k, of each element of the kth cluster from the cluster centroid is then computed as

$$E_k = \sum_{i=1}^{m_k} \left[\underset{j}{\text{Max}} |x_{ij} - Y_{kj}| \right] \tag{3}$$

The sum, D, of the E_k forms the objective function tested by the clustering algorithm to determine the locally optimal subdivision of the data into the prespecified number of clusters.

In application, each data point is experimentally transferred from its parent cluster to every other cluster until D reaches a minimum for that cluster level. Note that D_{min} is monotonically decreasing for increasing numbers of clusters, until $D_{min}=0$ when every point defines a separate cluster. The number of clusters, and hence water types, selected must depend in part upon the shape of the D versus cluster number curve, and the

physical significance of the number of clusters.

Projection of the Sample Spectra on the Cluster Centroids

The ultimate goal of the analysis is to classify each sample spectrum as to the parent water type(s) which makes up its spectral shape. We therefore wish to compute the scalar coefficients, A_k, such that

$$\max_j \left| x_{ij} - \sum_{k=1}^{L} A_k Y_{kj} \right| \tag{4}$$

is a minimum subject to the constraints that

$$\sum_{k=1}^{L} A_k = 1$$
$$0 \leq A_k \leq 1 \; . \tag{5}$$

This can be cast as a straightforward linear programming problem (ref. 9) which yields the desired A_k. Note that the A_k represent the proportion of each basic water type making up a particular $\vec{x}_j$, and that the criteria for best fitting the A_k has the same distance measure as the clustering algorithm.

To this point in the processing, no spatial information has been employed (except to assist in selecting an appropriate cluster level). The method classifies each data point based entirely upon its spectral characteristics. The results of the classification are then plotted in physical space to show the distributions of the different water types.

APPLICATIONS TO AOL FIELD DATA

The analysis technique described above was applied only to those flightlines outside the Bay mouth to attempt to define the boundaries of the Chesapeake Bay outflow plume. An L-band salinity mapping radar was flown simultaneously with the AOL and provides a basis for comparison with the AOL results reported here (ref. 10).

June 23, 1980 Data Set

The first data set considered was obtained during early ebb on 23 June 1980. The subset of flightlines used contained 1994 sample spectra. The EOF analysis was performed on the rescaled data and the sample spectra were transformed using the new basis. Since the first four eigenvectors accounted for 97 percent of the variance (Table I) only the spectral bands corresponding to the first four eigenvectors were retained in the transformed spectra.

The transformed sample spectra were then subdivided into one, two, three, four, and five clusters. Figure 3 shows a plot of D_{min} versus cluster number. D_{min} is monotonically decreasing and each increase in cluster number results in a decreasing reduction the value of D_{min}. Figure 4 shows the results of the clustering at the two and three cluster level. Note that the plume structure remains essentially unchanged but that the offshore region contains more structure at the higher cluster level. We thus have a well-defined baywater plume and an offshore region which can be further subdivided into at least two different water types; therefore the percentage distribution of the three water types, plume and two offshore water types, was computed for this data set.

Figure 5 shows the percentage distributions of the three water types. For comparison, the L-Band salinity map is shown in Figure 6. Our results show the Bay plume, Figure 5a, extending southward along the coast with two distinct bulges. The northward bulge is clearly the emerging plume for the current tidal cycle (the tide stage is early ebb), while the second bulge may well represent a remnant plume from the previous tidal cycle. The other two water types are shelf waters which have been subdivided into two sets, shelf water from north of the Bay mouth, Figure 5b, and shelf water from southeast of the Bay entrance, Figure 5c. Evidence that the second bulge of the plume is from a previous tidal cycle is seen in Figure 5b where an isolated pocket of northern shelf water lies between the southeast shelf water and the Bay water. A new influx of northern shelf water is apparent at the top of Figure 5b.

A comparison of the structure mapped by the analysis techniques used here and the L-band salinity map shows good agreement between the two within the license taken in contouring provided by the wide flightline spacing. Notice, however, that the clustering approach has been able to distinguish between two types of shelf water, especially east of the Bay entrance, thus providing potentially useful information about the complex circulation in this region.

June 25, 1980 AOL Data Set

The 25 June data set analyzed consisted of 3109 sample spectra. The results of the EOF analysis are given in Table I, where 97 percent of the variance is accounted for by the first four eigenvectors. The transformed spectra were clustered in the same way as the 23 June set, and the D_{min} values versus cluster number are plotted in Figure 3. The variance is more distributed over the eigenvectors than for the 23 June case, and there is a more evident difference between clustering at the two- and three-cluster level, Figure 7. For comparison with the 23 June results the analysis of this data set continued at the three-cluster level.

The results of mapping water type percentages are shown in Figure 8. These plots are considerably different from the results presented in Figure 6. Here we see that despite being very near slack water after flood in the tidal cycle, the Bay water type covers the whole northern and western portion of the region near the Bay entrance, with a band of roughly uniform width extending southward along the coast. Notice that the bulges seen on 23 June are not in evidence here. The other two water types defined by the technique are not as clearly distinguishable as in the previous example. Type 2, Figure 8b, could be interpreted either as northern shelf water trapped from the previous ebb cycle as in the 23 June case, or as an intermediate type consisting of a mixture of shelf, Figure 8c, and plume water.

The L-Band salinity map also shows a high degree of variability (Figure 9). Notice that the plume boundary of Figure 8a closely parallels the dotted boundary overlaid in Figure 9. The cluster analysis does not show the higher salinity tongue just south of Cape Charles in Figure 9. Also in Figure 9, the high salinity band (30-31 ppt) southeast of Cape Henry corresponds well with the type 2 water defined by the cluster analysis. The complex structure seen in Figure 9, especially the high salinity band, and the eastern extent of the plume in the northeast as defined by the cluster analysis could well be the result of offshore wind driving the surface waters eastward. Such a situation would spread the Bay water eastward of the Bay entrance, and could also result in local upwelling at the location of the high salinity band of Figure 9.

SUMMARY AND CONCLUSIONS

The results presented above are only preliminary; however, the methodology described here is shown to effectively define water types based solely on the AOL spectrum and the thermal infrared scanner data. It is noteworthy that despite the fact that no spatial information was employed in the analysis, the method divides the data into spatially contiguous, physically plausible clusters. A comparison of the results of the cluster analysis with a very limited alternate data set shows good agreement in general, although differences are apparent in detail. The complexity of the spatial structure developed for 25 June (both salinity and water type mapping) precludes detailed interpretation without additional supporting information such as wind conditions and exact tide stage.

The 25 June water type mapping results, in contrast, show a smooth, realistic structure. The clear delineation of three basic water types and the spatial plots of their distribution are suggestive of the circulation pattern in the region. On ebb, the Bay water emerges and flows south along the coast while shelf water from along the Delaware Peninsula is transported southward and lies between the plume water and shelf water from southeast of the Bay entrance. South of Cape Henry, the three water types interact and mix. During flood, the tidal currents off Virginia Beach are

directed roughly northwest which results in the trapping during flood of plume water and northern shelf water in the inshore region south of Cape Henry. Early in each ebb cycle these trapped remnants are still in evidence. For the plume, this results in the double bulge seen in Figure 6a, and the scalloping of the plume seen in SEASAT-SAR imagery of the coast south of Cape Henry. Thus, to the extent that the analysis defines realistic water types, the results provide useful information about the distribution of those water types and the circulation patterns which can produce such distributions.

With respect to the analysis methodology a number of areas bear further investigation. In the results presented here three analysis steps were performed, the EOF analysis, the clustering, and the assignment of water type percentages. For the EOF analysis the data were rescaled to [-1, +1] so that no single spectral band would dominate the results. One would certainly like to investigate other scalings such as unit variance scaling, no scaling, or some weighted rescaling. Further, one should investigate including the L-band results in the analysis as an additional dimension of the data vectors since these data were obtained simultaneously with the AOL data. In clustering the data the ℓ_∞ norm was used since that distance measure was easily employed in the later assignment of water type percentages. However, other norms do exist such as the euclidean or ℓ_2 norm and the ℓ_1 norm,

$$\sum_j |x_{ij} - x_{kj}| \qquad (6)$$

The second measure can easily be accommodated by the linear programming approach used in the third stage of the processing. The euclidean norm could also be accommodated by casting the assignment problem as a quadratic programming problem (ref. 11). Finally, the selection of final cluster level is presently subjective in that no absolute objective criterion exists for choosing an optimal cluster level. In practice it may not be possible to develop such a criterion in view of the monoticity of D_{min} with cluster level, however it may be possible to refine the selection process by also considering the distributions of number of spectra in each cluster and the mean and variance of the distance of sample spectra from their cluster centroids.

Despite the fact that none of the above variations was included in the preliminary analysis reported here, the results are physically realistic and compare favorably with a limited comparative data set. Further refinements in the approach may well improve the overall quality and confidence of the final results.

REFERENCES

1. Sarabun, C. C.: Structure and Formation of Delaware Bay Fronts. PhD. Dissertation, College of Marine Studies, University of Delaware, 1980.

2. Sick, L. V.; Johnson, C. C.; and Engel, A.: Trace Metal Enhancement in the Biotic and Abiotic Components of an Estuarine Tidal Front. Journal of Geophysical Research 83(C9), 1978, pp. 4659-4667.

3. Tyler, M. A.; and Seliger, H. H.: Annual Subsurface Transport of a Red Tide Dinoflagellate to Its Bloom Area: Water Circulation Patterns and Organism Distributions in the Chesapeake Bay. Limnology and Oceanography, 23(2), 1978, pp. 227-246.

4. Wiebe, P. H.; Hulbert, E. M.; Carpenter, E. J.; Jahn, A. E.; Knapp, G. P., III; Boyd, S. H.; Ortner, P. B.; and Cox, J. L.: Gulf Stream Cold Core Rings: Large-scale Interaction Sites for Open Ocean Plankton Communities. Deep-Sea Research, 23, 1976, pp. 695-710.

5. Hoge, F.; and Swift, R. N.: Application of the NASA Airborne Oceanographic Lidar to the Mapping of Chlorophyll and Other Organic Pigments. Chesapeake Bay Plume Study - Superflux 1980, NASA CP-2188, 1981 (Paper no. 26 of this compilation).

6. Wallace, J. M.; and Dickinson, R. E.: Empirical Orthogonal Representation of Time Series in the Frequency Domain. Part I: Theoretical Considerations. Journal of Applied Meteorology, 11(6), 1972, pp. 887-892.

7. Sneath, P. H. A.; and Sokal, R. R.: Numerical Taxonomy: The Principles and Practice of Numerical Classification. W. H. Freeman and Company, San Francisco, 1973.

8. Spath, H.: Cluster Analysis Algorithms for Data Reduction and Classification of Objects. John Wiley and Sons, New York, 1980.

9. Sarabun, C. C.: System Identification in Nonlinear Differential Equations by Quasi-linearization and Linear Programming. Masters Thesis, Univ. North Dakota, 1973.

10. Kendall, Bruce M.: Remote Sensing of the Chesapeake Bay Plume Salinity Via Microwave Radiometry. Chesapeake Bay Plume Study - Superflux 1980, NASA CP-2188, 1981 (Paper no. 10 of this compilation).

11. Hadley, G.: Linear Programming. Addison - Wesley Publishing Company, Inc., Reading, Mass., 1962.

TABLE I

PERCENT VARIANCE ACCOUNTED FOR BY
FIRST FOUR EMPIRICAL ORTHOGONAL FUNCTIONS.

EOF	23 June	25 June
1	89.5	73.8
2	3.6	12.4
3	2.5	8.6
4	1.4	1.7
TOTAL	97.0	96.5

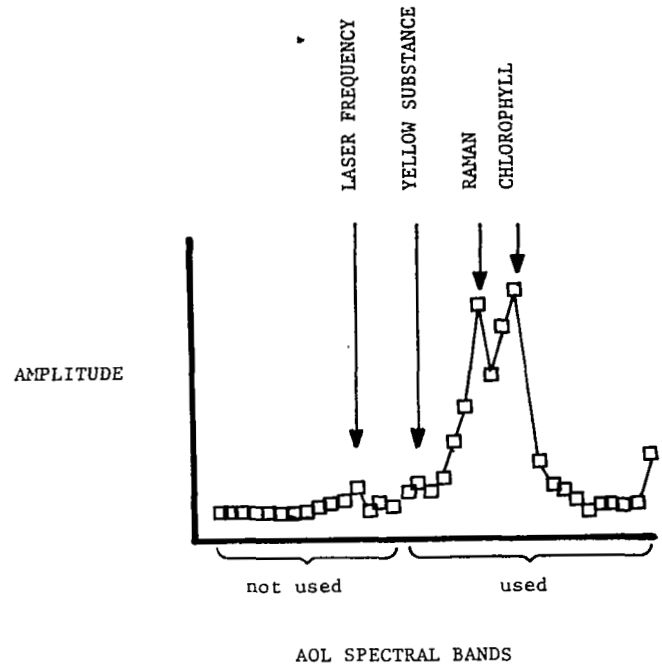
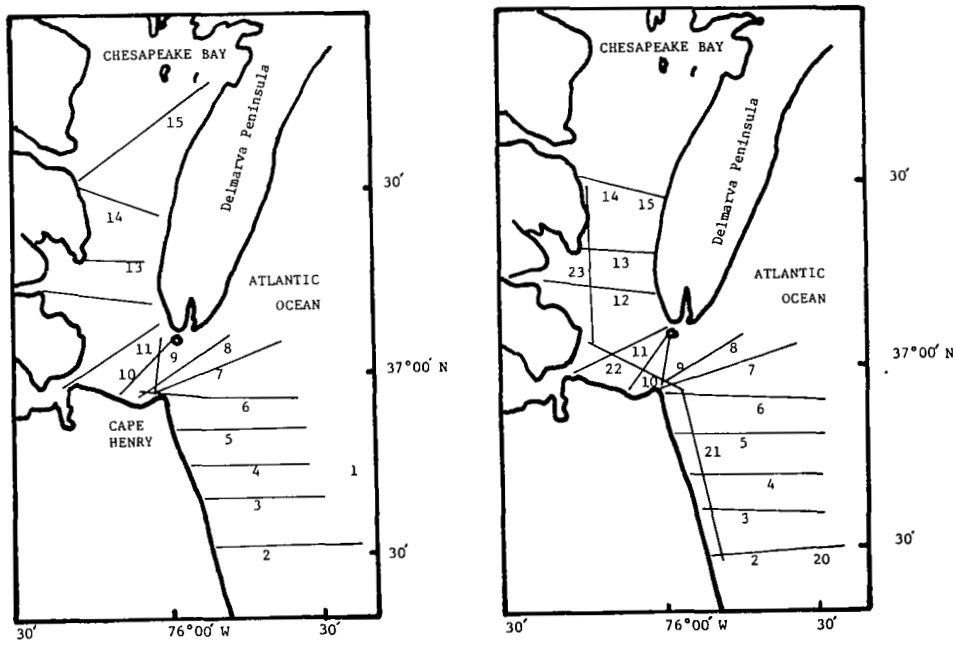


Figure 1.- AOL return spectrum. Only the leftmost 20 channels were used in the analysis. Major peaks are annotated. (Provided by F. Hoge and R. Swift, NASA/Wallops Flight Center.)



(a) June 23. (b) June 25.

Figure 2.- June 23 and 25 AOL flight lines.

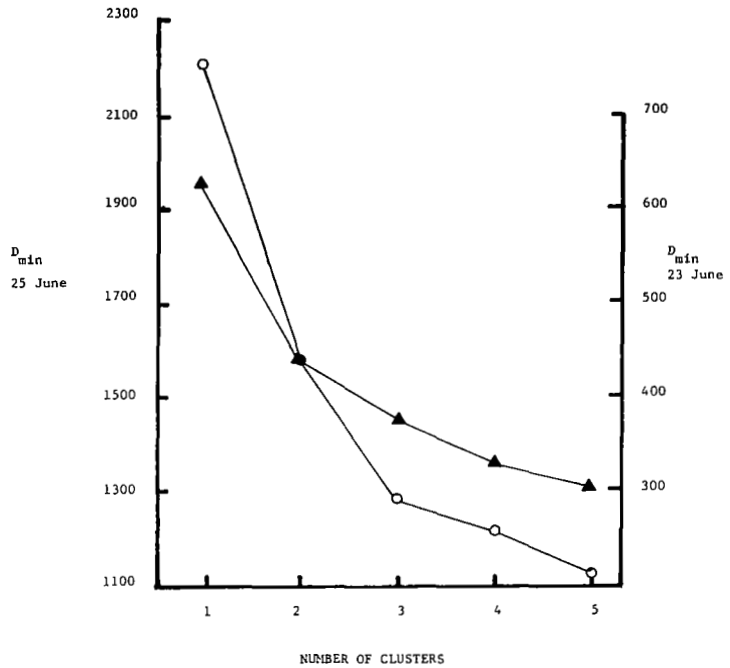
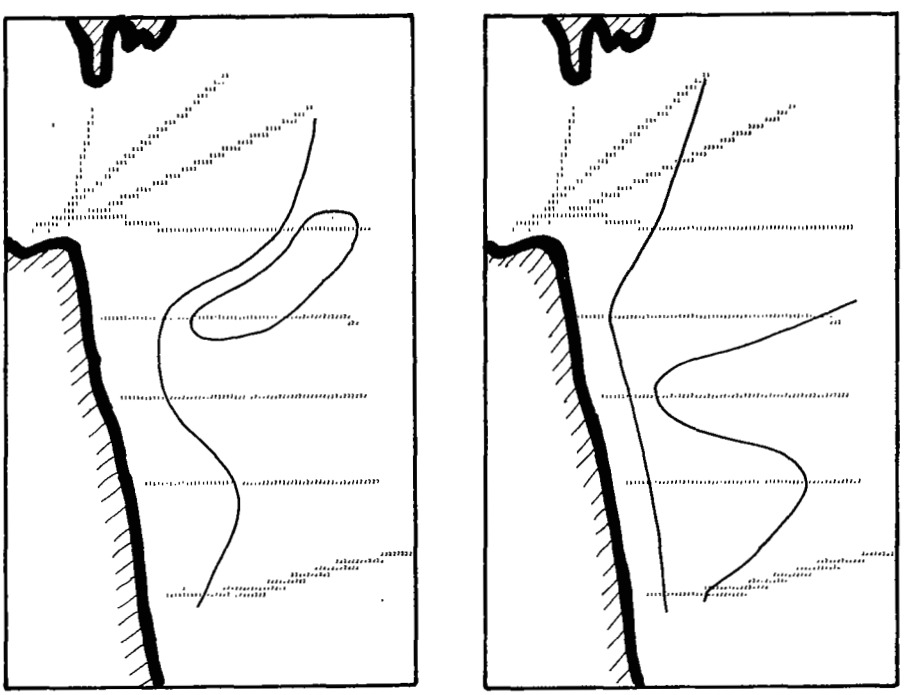
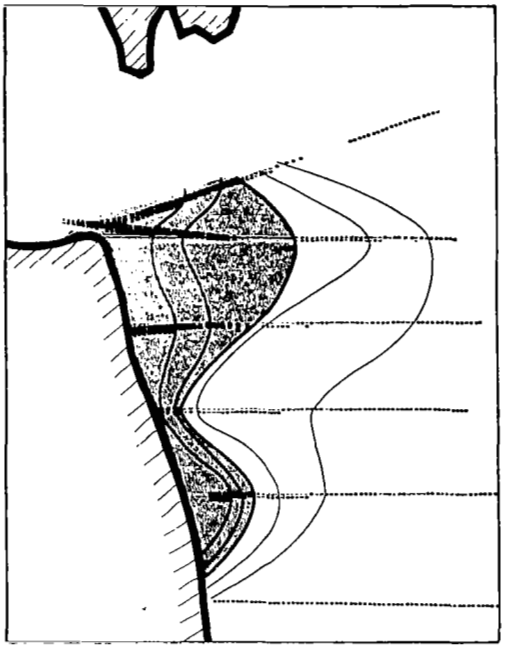


Figure 3.- D_{min} versus number of clusters for 23 June (▲) and 25 June (O).

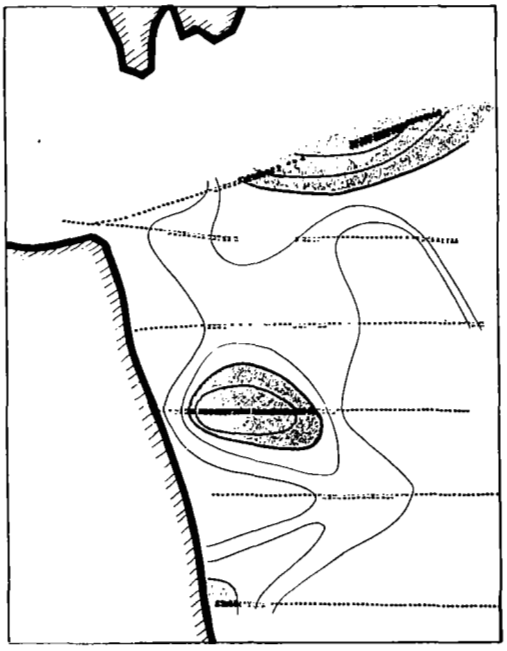


(a) Two clusters. (b) Three clusters.

Figure 4.- Spatial distribution of cluster assignments for June 23 for two and three clusters.

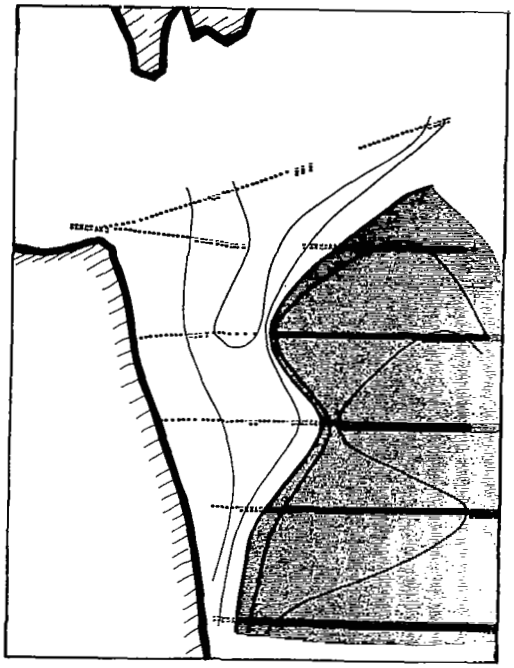


(a) Shaded areas indicate more than 50% water type 1.



(b) Shaded areas indicate more than 50% water type 2.

Figure 5.- Spatial distribution of water types 1, 2, and 3 for June 23.



(c) Shaded areas indicate more than 50% water type 3.

Figure 5.- Concluded.

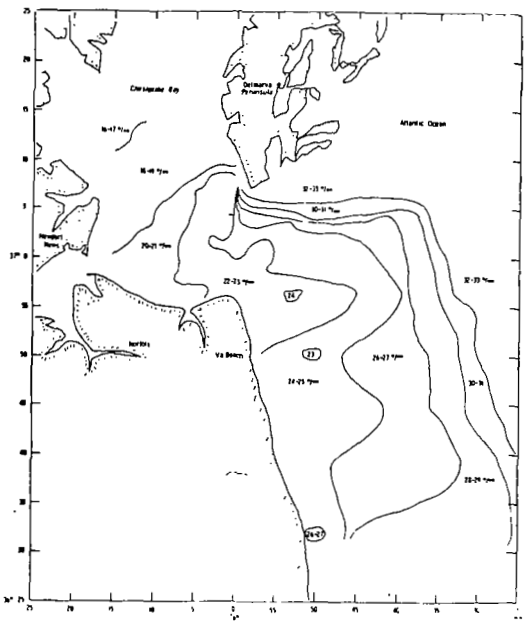
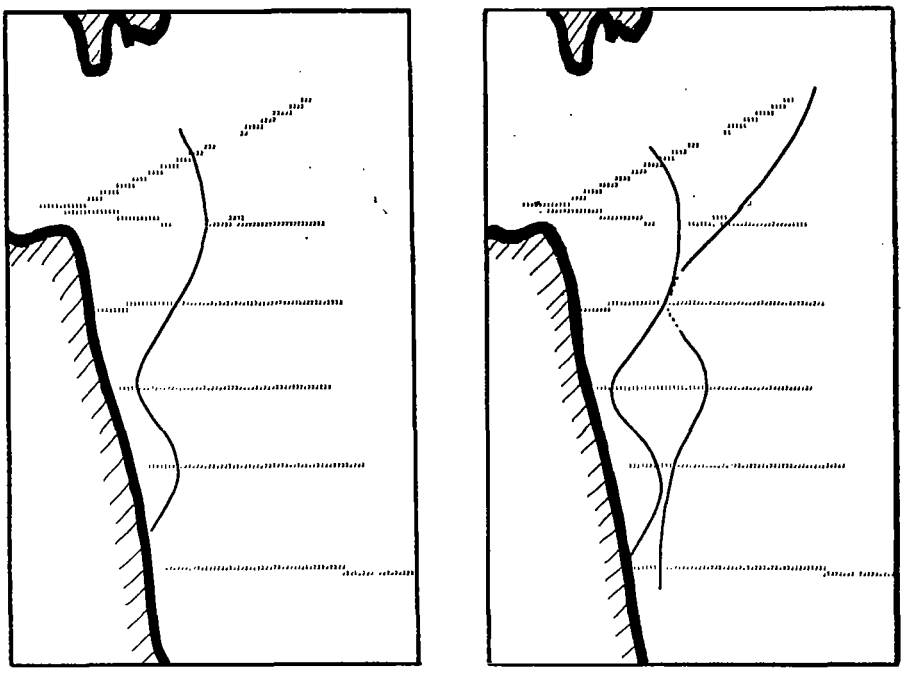
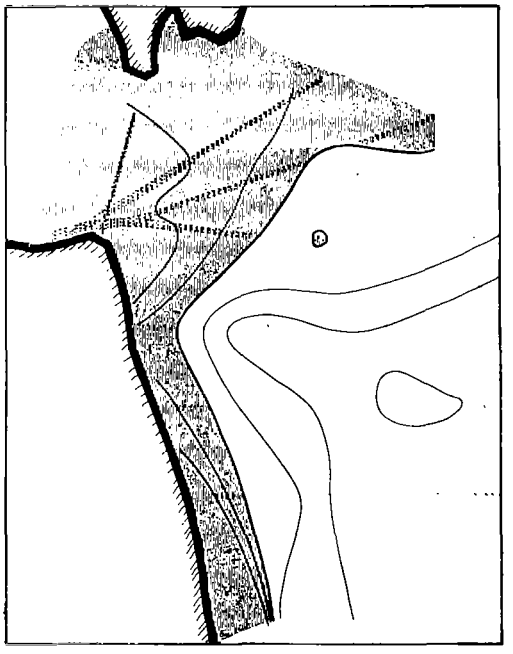


Figure 6.- L-band microwave radiometer salinity map (from ref. 10).



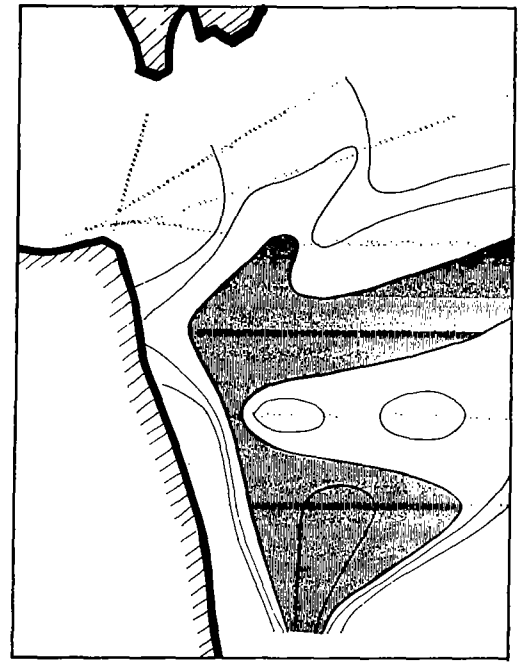
(a) Two clusters. (b) Three clusters.

Figure 7.- Spatial distribution of cluster assignments for June 25 for two and three clusters.

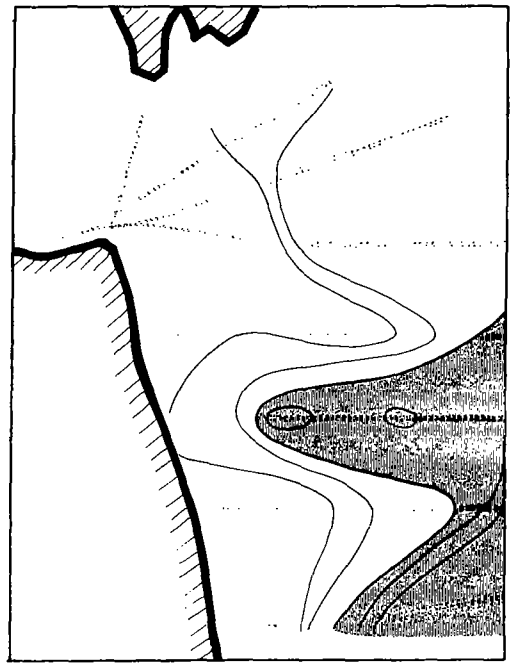


(a) Shaded areas indicate more than 50% water type 1.

Figure 8.- Spatial distribution of water types 1, 2, and 3 for June 25.



(b) Shaded areas indicate more than 50% water type 2.



(c) Shaded areas indicate more than 50% water type 3.

Figure 8.- Concluded.

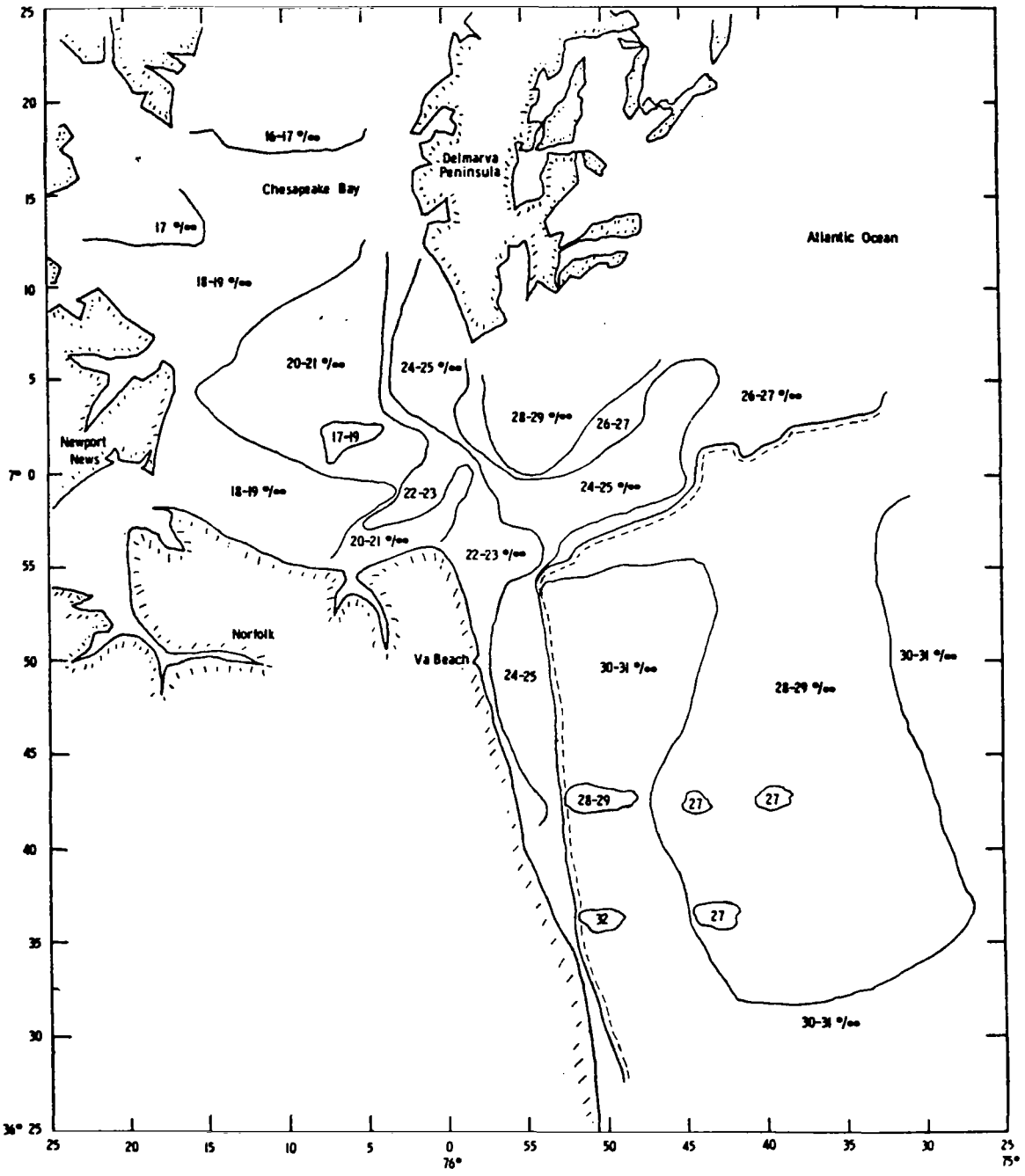


Figure 9.- L-band microwave radiometer salinity map (from ref. 10).