OPTICAL CLASSIFICATION OF NORTHWEST ATLANTIC WATER TYPES BASED ON SATELLITE OCEAN COLOUR DATA

Linda V. Martin Traykovski and *Heidi M. Sosik*
Biology Department, MS 32, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

ABSTRACT

Satellite ocean colour imagery provides a synoptic view of the optical properties of broad regions of the ocean, and sophisticated data analysis techniques are required for the interpretation of this data. We are developing optical water type classification approaches based on remotely-sensed water leaving radiance, with application to the study of spatial and temporal dynamics of ecologically and biogeochemically important properties of the upper ocean. For CZCS and SeaWiFS imagery of the northwest Atlantic region, pixels from several locations in the Georges Bank/Gulf of Maine area projected into distinct clusters in single-band feature space, suggesting that these waters can be easily distinguished using a few spectral bands of ocean colour data. Two different classification techniques have been developed. The Euclidean Classifier minimises the raw distance between each pixel and the centroid of the class to which it is assigned, whereas the Eigenvector Classifier is based on normalising the raw distances by the variance of each class, taking the shape of each class in feature space into account. These classifiers were applied to ocean colour images of the northwest Atlantic to elucidate the geographical location and extent of each water type.

INTRODUCTION AND BACKGROUND

The availability of remotely-sensed ocean colour data from satellite sensors such as SeaWiFS and CZCS has opened up new opportunities to study the spatial and temporal variability of phytoplankton distributions. Although remote sensing of ocean colour has significantly expanded our ability to study spatial and temporal variability in phytoplankton abundance and distribution, full exploitation of ocean colour imagery requires both developments in modelling of upper ocean optical properties (and their relationships with biological, physical, and chemical properties) and more sophisticated data analysis techniques. Optical water type classification approaches based on remotely-sensed water leaving radiance have great potential to contribute to the study of spatial and temporal dynamics of ecologically and biogeochemically important properties in the upper ocean. Particularly in coastal waters, both inherent and apparent optical properties are influenced by a wide array of physical, biological and chemical processes. These processes can lead to large sources of optical variability that may be independent of the abundance of phytoplankton pigments. In addition to these pigments, constituents such as dissolved organic matter (DOM) of both marine and terrigeneous origin, heterotrophic organisms, biological detritus, and inorganic particulate material can affect both the magnitude and spectral quality of reflected light. This complexity may interfere with accurate estimation of phytoplankton distributions based on optical signatures; however, it also presents the potential for deriving information about other water properties from space.

Because ocean colour signals vary in response to many processes, successful identification of optically different types of water is necessary for accurate retrieval of constituent concentrations. Satellite images of large geographic areas often reveal mesoscale reflectance features that are associated with physical, biogeochemical and biological processes in the upper ocean. Satellite data have been exploited to help identify the scales associated with these features; for example, Sathyendranath *et al*. (1991) used AVHRR imagery combined with local bathymetry to define water types in a study of productivity on Georges Bank. To date, efforts to identify mesoscale features or water type boundaries from remotely-sensed ocean colour data have generally relied only on pigment distributions or have involved relatively dramatic water type differences, such as those that occur near river plumes. The potential for using more information than is contained in pigment images and to discern more subtle differences in optical water types has not been fully explored.

There have been efforts to use CZCS data for water type identification using specialised algorithms designed to recognise the unique optical properties of a particular type of phytoplankton. A successful method was developed to detect coccolithophore blooms using CZCS remotely-sensed radiances based on a nonparametric parallelepiped supervised algorithm (Brown and Yoder 1994a,b), which was able to distinguish coccolithophore pixels from non-coccolithophore pixels. Subramaniam and Carpenter (1994) developed a protocol to identify *Trichodesmium* blooms from CZCS imagery based on high reflectivity from gas vacuoles and a phycoerythrin absorption feature at 550 nm, and were able to distinguish two *Trichodesmium* blooms from sediment whitings and from some portions of coccolithophore blooms. Attempts have been made to detect cyanobacterial blooms using a supervised classification technique (Zabicki 1995) using the observed ratio of (total radiance at 750 nm)/(total radiance at 670 nm); although this ratio was always lower for suspected cyanobacterial blooms than for sediment conditions, it was not possible to distinguish coccolithophore blooms from *Trichodesmium* blooms with this method. These taxon-specific algorithms can indicate the presence of near mono-specific blooms in the analysis of particular ocean regions at times when blooms of that type are thought to occur. The utility of these approaches may be limited, however, in the identification and classification of a broad range of water types that may span many scales of spatial and temporal variability. To fully exploit ocean colour data for the study of phytoplankton dynamics, it is necessary to develop a more universal scheme to optically classify many different types of phytoplankton blooms simultaneously by automatically distinguishing them from each other.

A promising approach to identifying optical water types based on remotely-sensed data is to develop a comprehensive framework within which different water types may be automatically and simultaneously distinguished from each other. Subsequently, additional information such as *in situ* observations can be used to categorise the water types in some ecologically relevant manner. The development of an automatic classification scheme essentially involves the inversion of observed data to retrieve a property of interest. Inversion schemes can be of two general types, those based on a forward model of the process, and those based only on intrinsic features in the data. Previous work on phytoplankton bloom identification (e.g., Brown and Yoder 1994a,b, Subramaniam and Carpenter 1994, Zabicki 1995) is a limited form of feature-based classification where a decision rule is applied to determine whether data fall inside or outside a single class boundary. Here we present a foundation for the development of a more comprehensive approach to the optical classification of water type.

CLASSIFIER DEVELOPMENT, RESULTS, AND DISCUSSION

The work presented here consists of an analysis of CZCS and SeaWiFS data to demonstrate the feasibility of applying feature-based classification techniques to identify and delineate optical water types. This work involved a regional study of CZCS and SeaWiFS imagery of waters in the northwest Atlantic, including Georges Bank and the Gulf of Maine. The goal of this regional analysis was to establish feature-based classification techniques for separating, using ocean colour data, various water types found in the restricted geographic region on and around Georges Bank. Two different classification approaches were developed and applied to imagery from this region.

Classifier Development

Our analysis of optical water types in the Northwest Atlantic is focused on the Georges Bank/Gulf of Maine area, but includes water types farther to the south for context (Figure 1). For our initial study, geographic locations were subjectively selected based on general knowledge of the hydrography and bathymetry, combined with examination of CZCS and SeaWiFS imagery. Six regions were considered: Gulf Stream waters (**GS**), Central Mid-Atlantic Bight waters (**CMAB**), CMAB waters near the coast and just south of Georges Bank (**cCMAB**), coccolithophore waters (**cocco**), waters on Georges Bank (**GB**), and Gulf of Maine waters (**GM**). Normalised water-leaving radiances (nL_w, Gordon *et al*. 1988) at 443 nm, 520 (or 510) nm and 550 (or 555) nm were extracted from the CZCS or SeaWiFS images for 100 randomly selected pixels from each of these six locations. These data were used as a training set for the classifiers. For an individual scene, the water types characteristic of these Northwest Atlantic locations are easily distinguishable as distinct clusters when projected in a three-dimensional feature space. In addition, these water types project onto the same regions in feature space for different scenes over time. For example, examining CZCS and SeaWiFS images spanning 17 years revealed that the waters over Georges Bank (**GB**) occupy the same region in feature space for a scene in October 1997 as they did in July 1980. This stationarity increases the robustness of the training set. Projection of the training set data in feature space clearly shows six well-delineated clusters (Figure 2). For each water type class i with n_j training data points, the class centroid $\mathbf{m}_i$ was computed as the arithmetic three-dimensional mean of the training data points $\mathbf{a}_{ij}$ for that class:

$$\mathbf{m}_i = \mathrm{sum}(\mathbf{a}_{ij})/n_j.$$

We have automated the water type classification process by applying statistical decision criteria to define class boundaries and assign pixels to a particular class. We have implemented two different feature-based classifiers, the Euclidean Distance Classifier, and the Eigenvector Classifier. The Euclidean Distance Classifier assigns each pixel $\mathbf{p}_j$ to a water type i based on the distance between that pixel and the centroid or mean of each class. The Euclidean distance d_{euc} between a pixel $\mathbf{p}_j$ and a class centroid $\mathbf{m}_i$ is defined as:

$$d_{euc}(\mathbf{p}_j\,,\,\mathbf{m}_i) = [(\mathbf{p}_j - \mathbf{m}_i)^{\mathbf{T}}(\mathbf{p}_j - \mathbf{m}_i)]^{1/2}$$

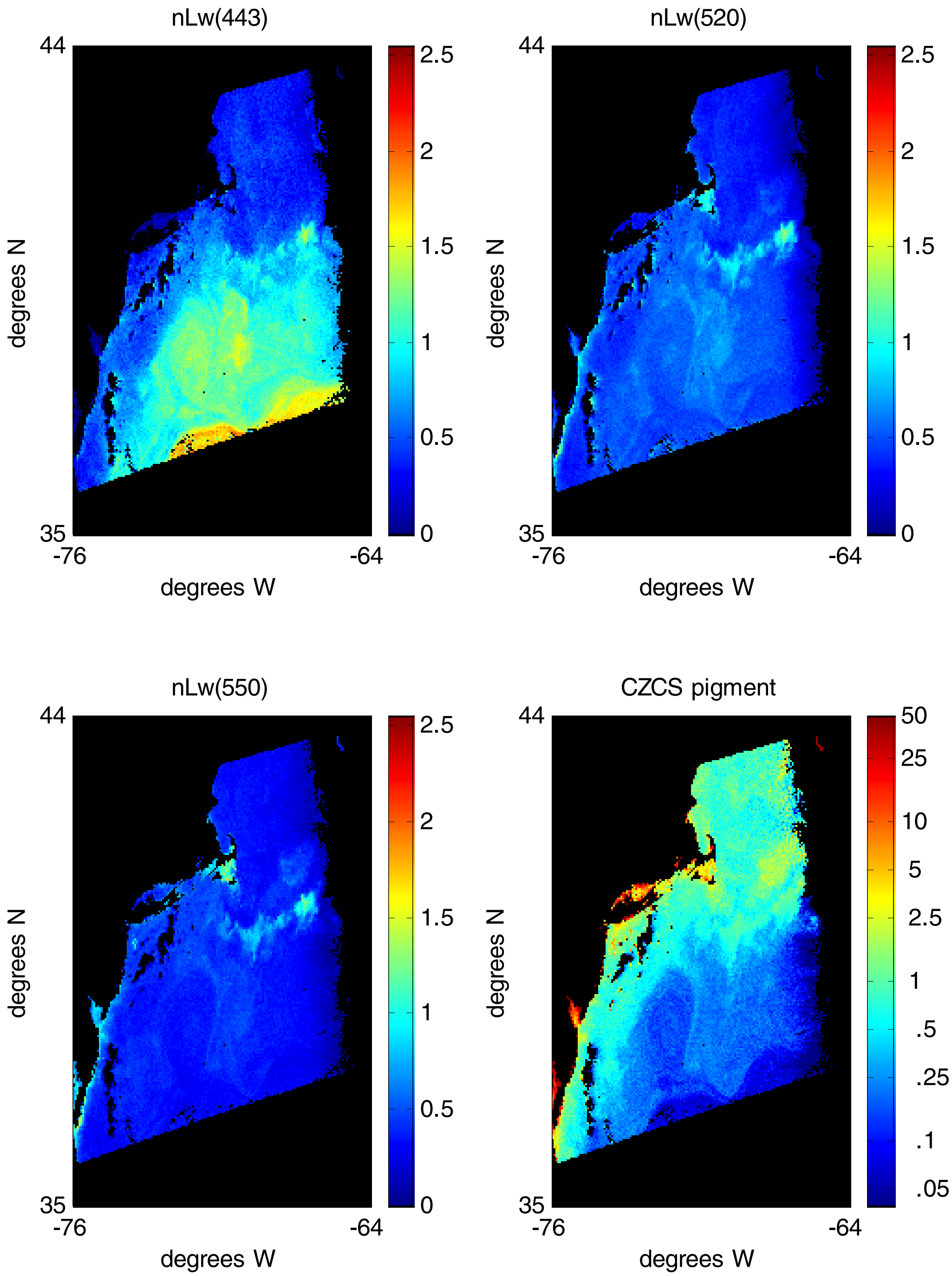

Figure 1. Normalised water-leaving radiances (nLw; mW/(sr·cm^2·µm)) at 443, 520, and 550 nm, and chlorophyll (mg/m^3) for the Northwest Atlantic, 7 July 1980. Georges Bank appears as a region of high pigment in the chlorophyll image. Land and clouds appear black. This scene was processed by John Ryan using east coast atmospheric correction algorithms. Jim Acker at NASA/DAAC helped acquire the CZCS data.

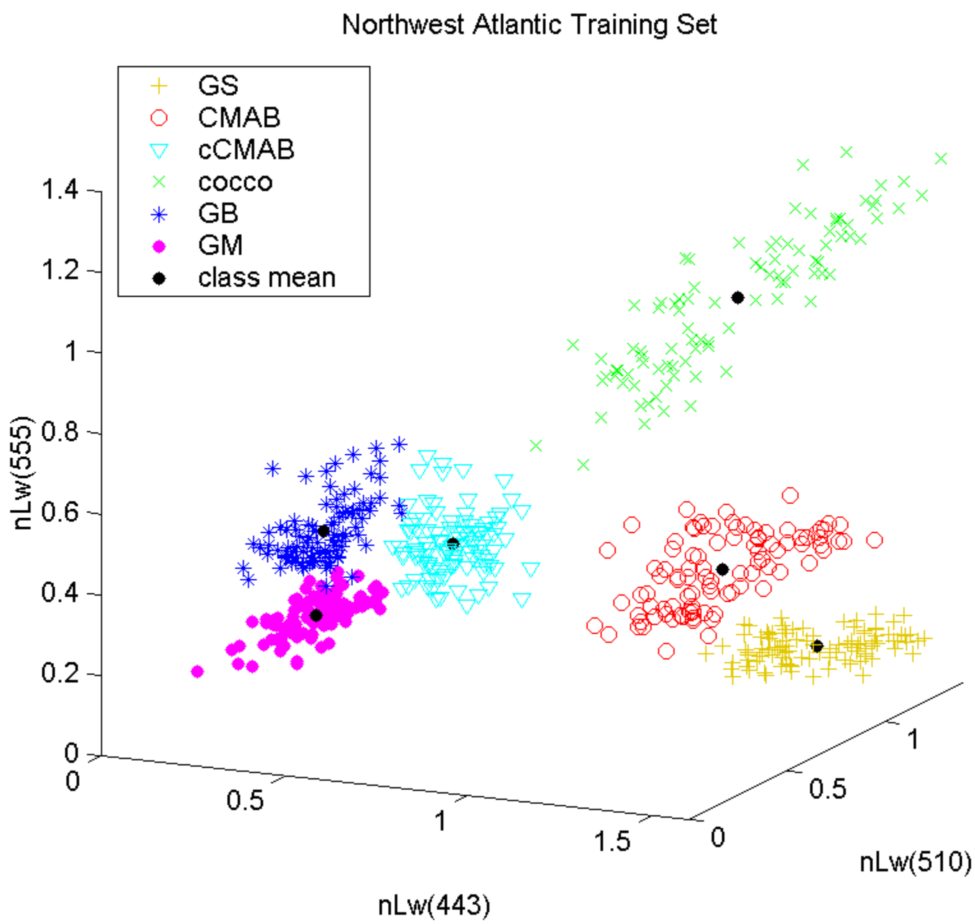

Figure 2. Training set for classification of Northwest Atlantic water types shown in a three-dimensional single band feature space; nLw in mW/(sr·cm^2·µm). The training set (derived from the 7 July 1980 CZCS image, see Figure 1), consists of 100 randomly chosen pixels from each of six different geographical regions. GS = Gulf Stream waters; CMAB = Central Mid-Atlantic Bight waters; cCMAB = coastal/northern CMAB waters; cocco = coccolithophore waters; GB = Georges Bank waters; GM = Gulf of Maine waters; class mean indicated as black dot at centroid of each cluster.

A pixel is assigned to the water type whose class centroid is the minimum euclidean distance away, so that the decision rule for the Euclidean Distance Classifier may be written as:

$$\mathbf{p}_j \; \varepsilon \; \mathbf{m}_k \;\; \text{iff} \;\; d_{euc}(\mathbf{p}_j\,, \mathbf{m}_k) < d_{euc}(\mathbf{p}_j\,, \mathbf{m}_l) \text{ for all } l \text{ not equal to } k.$$

The Euclidean Distance Classifier is well suited to spherical classes, *i.e.* classes whose boundaries in all directions are equidistant from the class centroid. For ellipsoidal classes, in which the region of feature space encompassed by the class is elongated in one or more dimensions relative to the other(s), classification success can be improved significantly by taking into account the ellipsoidal shape of the classes.

The Eigenvector Classifier was developed to include consideration of the three-dimensional shape of each class. With this classifier, each water type class is defined in terms of an ellipsoid in feature space, the principal axes of which are given by the three dominant eigenvectors (ψ_1, ψ_2, ψ_3) of the covariance matrix (**K**) of the training data in that class:

$\mathbf{K}_i = \mathbf{A}_i^{\mathbf{T}}\mathbf{A}_i$, where each row of $\mathbf{A}_i$ is a data point $\mathbf{a}_{ij}$ from the training set for class i.

The extent of each class in the three eigenvector directions is represented by the corresponding eigenvalues (λ_1, λ_2, λ_3). The result is that the classes occupy ellipsoidal regions in feature space, each oriented along their own eigenvector directions (Figure 3). With the Eigenvector Classifier, the distance of a pixel from the centroid of each class is computed in terms of components along the eigenvector directions for that class, and each component is normalised by the corresponding eigenvalue. The pixel is assigned to the class for which the normalised vector distance is smallest.

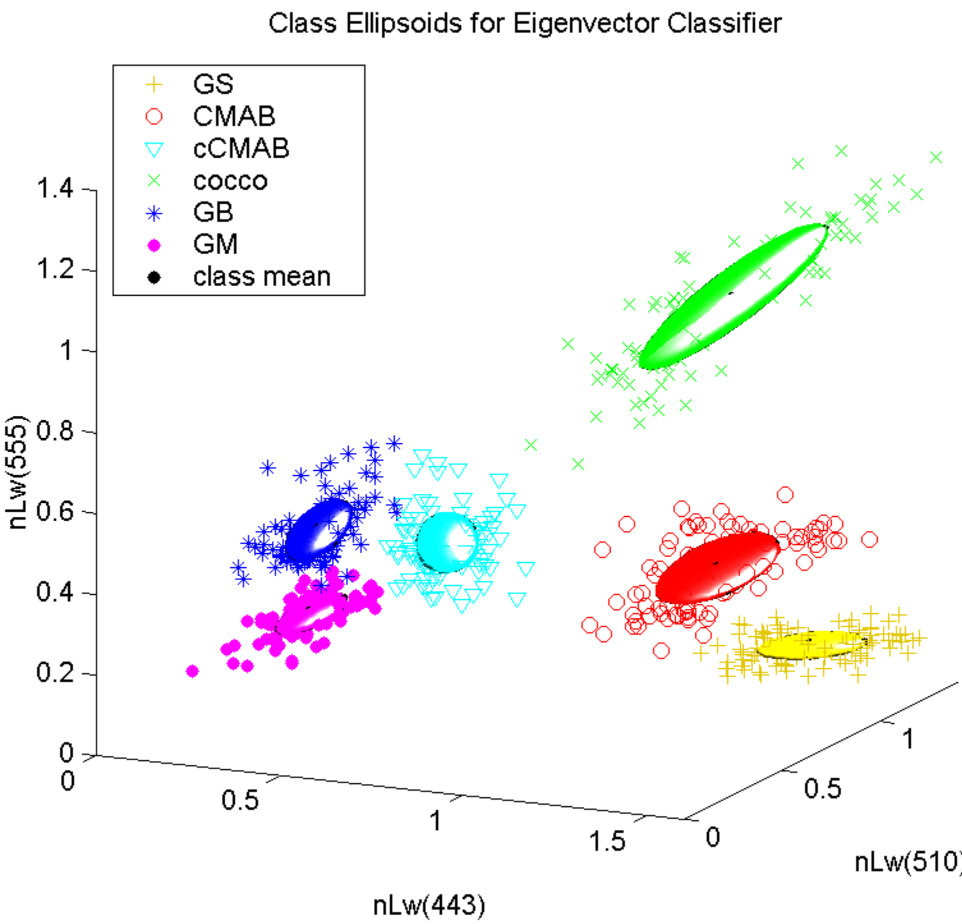

Figure 3. Ellipsoids defining the classes for the Eigenvector Classifier; ellipsoid orientation determined by the eigenvector directions for each class. The ellipsoid axis length in each eigenvector direction is equal to the square root of the corresponding eigenvalue, so that the ellipsoids shown encompass one standard deviation in each eigenvector direction of the training data for each class. This class definition accounts for the elongated shape of some of the classes. GS = Gulf Stream waters; CMAB = Central Mid-Atlantic Bight waters; cCMAB = coastal/northern CMAB waters; cocco = coccolithophore waters; GB = Georges Bank waters; GM = Gulf of Maine waters; class mean, where visible, indicated as black dot at centroid of each cluster.

Classification Results

Initial evaluation of these two classification techniques was carried out by constructing water type classes based on only half the pixels (randomly selected) of each water type in the Northwest Atlantic training set; classification was then carried out on the remaining half of the training set data (Figure 4). Applying the Euclidean Distance Classifier resulted in an average of 97.4% correctly classified pixels (sum(n_j) = 300), with a mean of 7.8 misclassified (std. dev. $\sigma = 2.285$) over 20 trials. Even higher success rates were achieved with the Eigenvector Classifier; consideration of the three-dimensional shape of each class reduced misclassification rates for the more elongated classes. Application of the Eigenvector Classifier gave an average of 99.1% correctly classified pixels (sum(n_j) = 300), with a mean of 2.85 misclassified ($\sigma = 1.565$) over 20 trials.

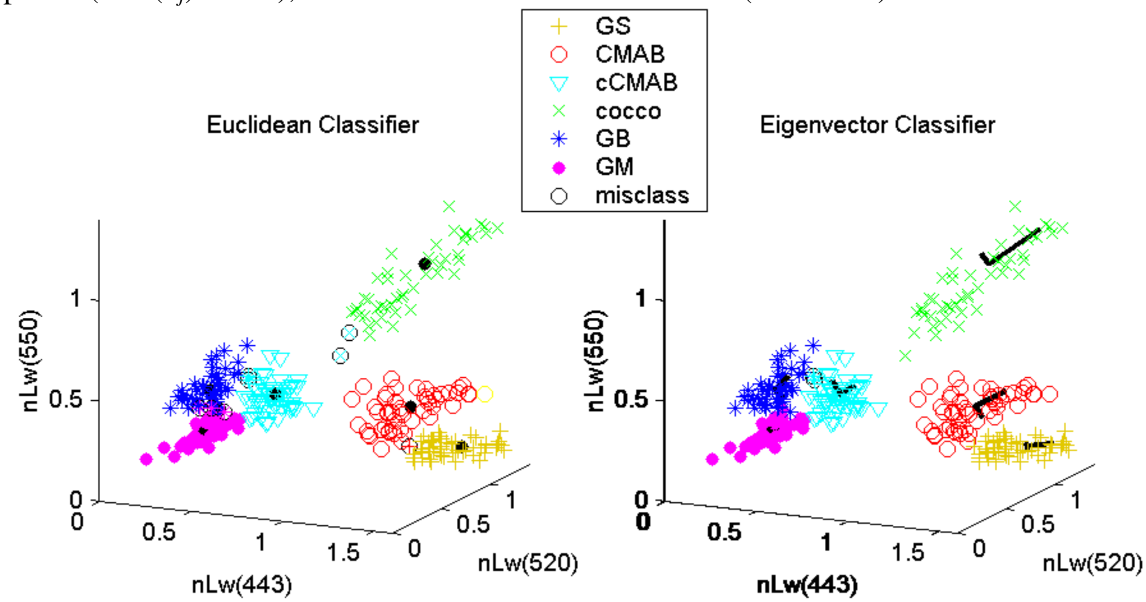

Figure 4. Initial classification of the Northwest Atlantic training set. Water type classes were constructed based on half the pixels (randomly chosen, 20 trials) of each water type in the training set; the remaining pixels (pictured above) were classified (misclassified pixels circled in black). Euclidean Distance Classification results at left (average 97.4% correct); class mean shown as black dots, misclassified pixels retain original symbol but take on colour of the class to which they were assigned. Eigenvector Classification results at right (average 99.1% correct); eigenvector directions indicated by black lines, length of lines indicates extent of class in each direction. GS = Gulf Stream waters; CMAB = Central Mid-Atlantic Bight waters; cCMAB = coastal/northern CMAB waters; cocco = coccolithophore waters; GB = waters on Georges Bank; GM = Gulf of Maine waters.

Both the Euclidean Distance and Eigenvector Classifiers were applied to several cloud-free ocean colour images of the Northwest Atlantic. Classification results reveal striking patterns of water type distribution throughout the region, as shown in the Euclidean Distance (Figure 5) and the Eigenvector (Figure 6) Classification results for both 7 July 1980 and 8 October 1997. The water types in each scene are clearly distinguishable, and classifier application reveals that waters of the same optical type

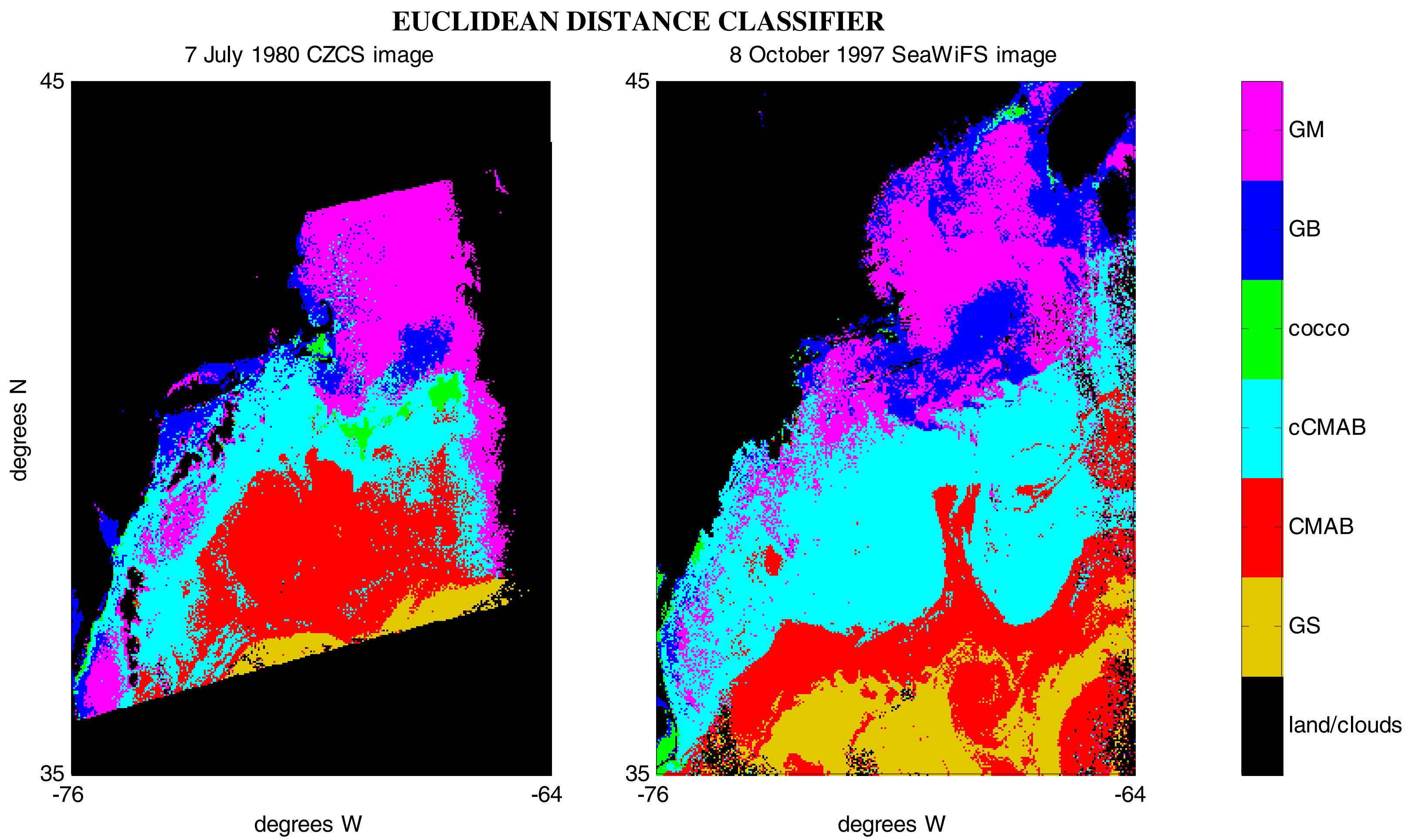

Figure 5. Classification results for the Euclidean Distance Classifier applied to the Northwest Atlantic on two different days: 7 July 1980 (CZCS image, left) and 8 October 1997 (SeaWiFS image, right). The water types are clearly distinguishable, and application of the classifier reveals that they form well-defined water masses. The distribution patterns of these water types are strikingly similar between the two scenes, even though they are 17 years apart. Mesoscale physical oceanographic features are apparent; differences may represent seasonal and/or inter-annual variability.

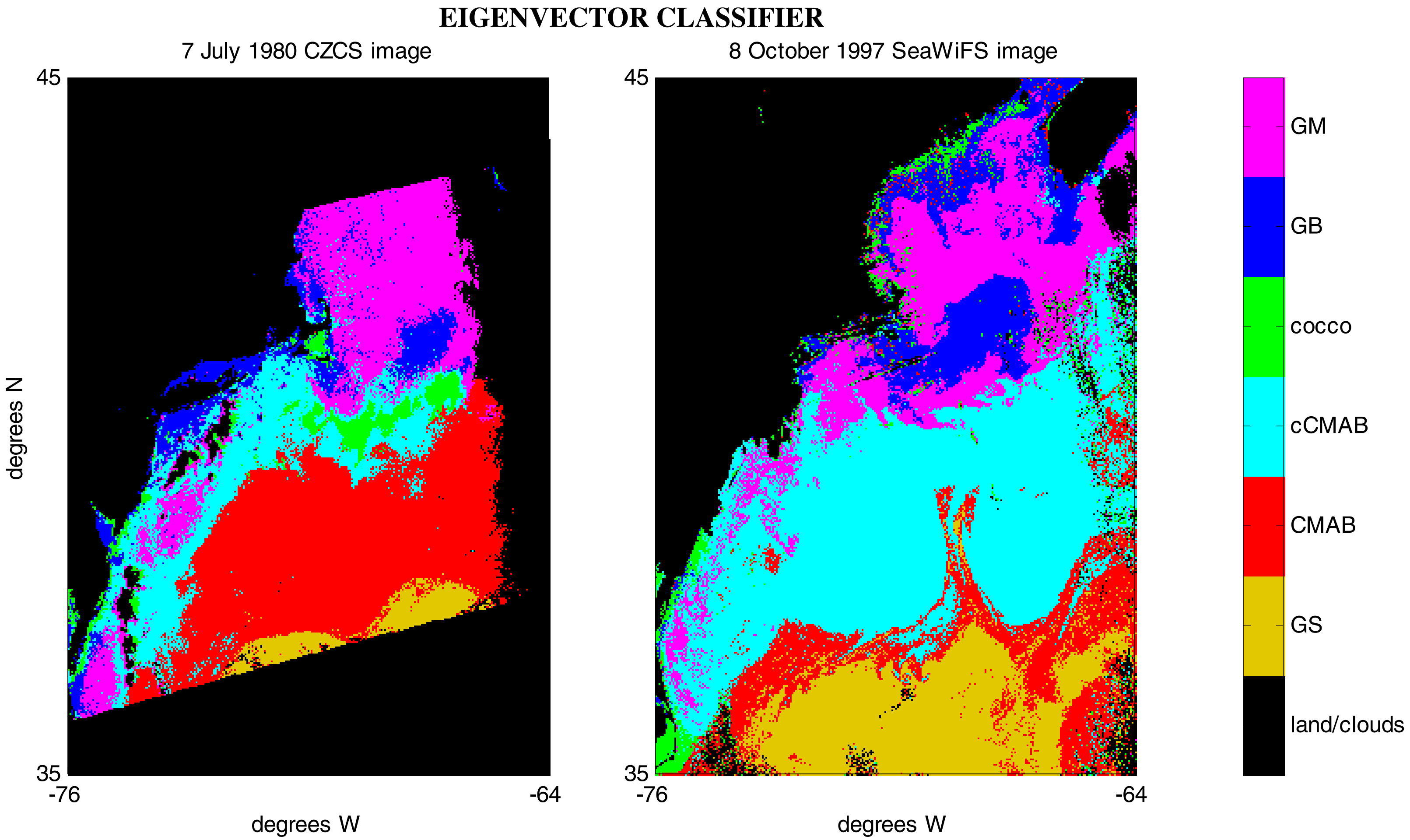

Figure 6. Classification results for the Eigenvector Classifier applied to the Northwest Atlantic on two different days: 7 July 1980 (CZCS image, left) and 8 October 1997 (SeaWiFS image, right). Results with this classifier are quite similar to the Euclidean Distance Classifier results shown in Figure 5. The Eigenvector Classifier seems to more correctly identify the pixels near the right cloud edge of the CZCS image as compared to the Euclidean Distance Classifier, but does not delineate the mesoscale eddy in the Gulf Stream revealed by the Euclidean Distance Classifier.

form well-defined water masses that remain in the same general geographical regions over time. *In situ* sampling of these optical water masses will allow more complete characterisation of their unique properties. Our feature-based classifiers are ideally suited to study temporal variability by tracking these optical water types through time. Applying the classifiers to a time series of images will allow changes in water type distributions to be followed easily. Mesoscale physical oceanographic features become apparent after classification, since they often result from interactions between water masses characterised by different optical water types. It is anticipated that our classification techniques will facilitate long-term studies by tracking optical water types through seasonal and inter-annual changes.

To better interpret the classifier results, a measure of Classification Goodness of Fit has been devised. This measure reveals the certainty with which each pixel is assigned to a given class. This is particularly important in assessing boundaries between water types as well as evaluating spurious pixels or groups of pixels that may not truly belong to any of the defined classes. To measure goodness of fit, a probability landscape is constructed, and the location of each pixel in the probability landscape determines its goodness of fit. Each class centroid represents the center of several concentric probability regions; in three-dimensional space, these regions may be thought of as probability shells. For example, the 5% probability shell for a given class is the region surrounding the class centroid within which the closest 5% of all the pixels in a scene fall. The metric for "closeness" is unique to each classifier; for the Euclidean Distance Classifier, the 5% probability shell includes pixels whose euclidean distance from the class centroid are in the smallest 5% of all euclidean distances for all pixels. The entire probability landscape consists of the combined probability shells for all the classes. In this fashion, the classifier not only assigns a pixel to a class, but it also determines how strongly that pixel belongs to that class. Pixels which fall in the lower order probability shells (e.g. 5% or 10%) for a given class are identified very strongly with that class, whereas pixels which fall in the higher order, outer shells (e.g. 90% or 95%) are very weakly associated with that class. Applying Classification Goodness of Fit measures to the Euclidean Distance Classifier results (Figure 7) reveals that the boundaries between water masses of different optical types are quite distinct, with pixels on either side of the boundaries being very strongly associated with their respective water type class. This goodness of fit measure also sheds some light on the classification results for some of the spurious groups of pixels. For example, for 7 July 1980, the longitudinally confined narrow band of pink pixels classed as **GM** to the extreme right of the CZCS image well south of the Gulf of Maine (see Figure 5, left) are shown to be only weakly associated (occurring in the 80% - 90% probability shells) with the **GM** optical type. It is likely that cloud-edge effects are confounding the optical signature of these pixels. Notably, the Eigenvector Classifier (see Figure 6, left) was better able to identify these pixels in spite of the noise in the data.

INDICATIONS FOR FUTURE RESEARCH

Our work thus far has demonstrated that the application of feature-based classification techniques to ocean colour data facilitates discrimination between Northwest Atlantic water types, including those waters occurring within a spatially

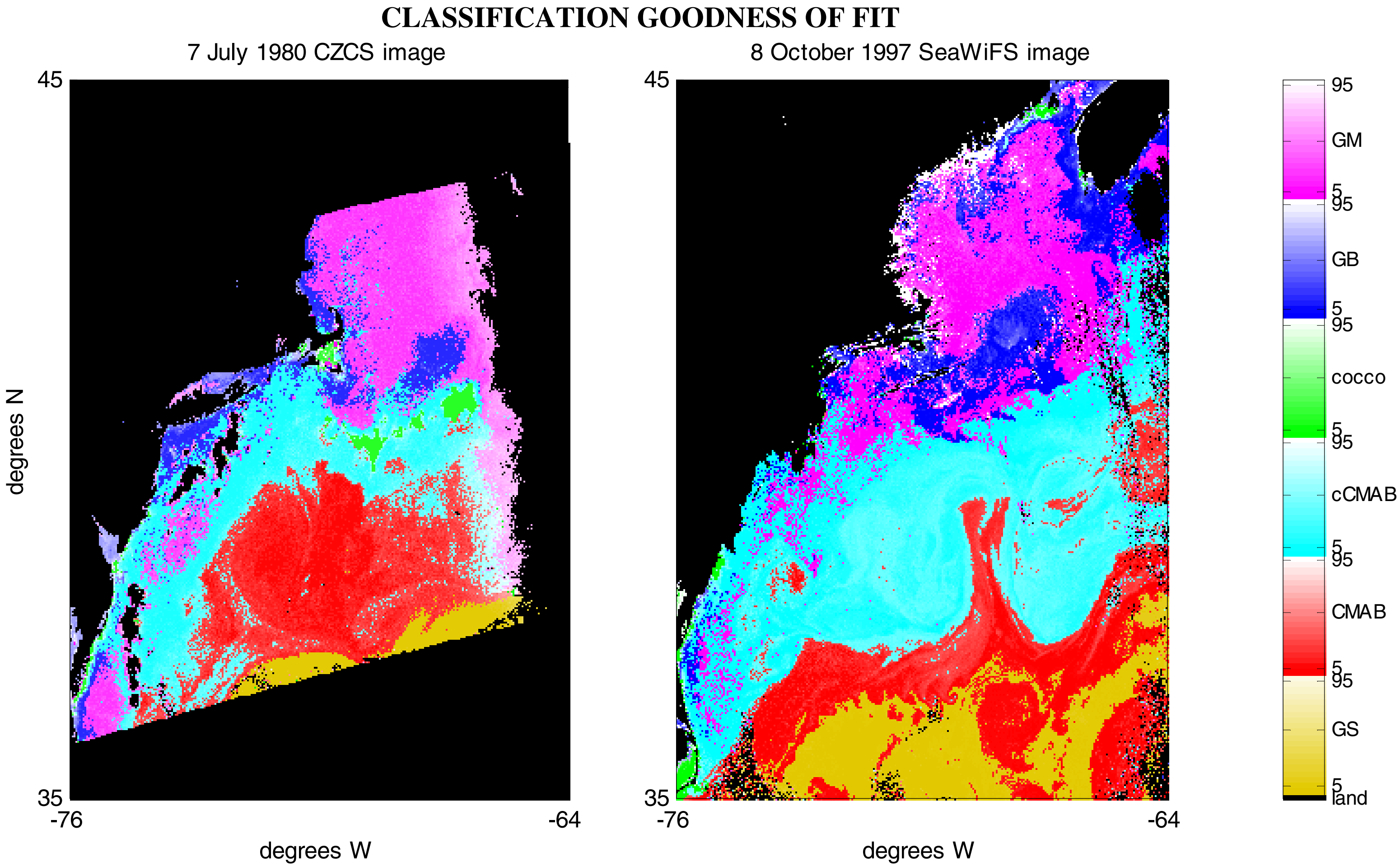

Figure 7. Goodness of fit for the Euclidean Distance Classifier (see Figure 5 for corresponding classifier results). Intensity of colour indicates goodness of fit to class, as shown in the legend at right. Goodness of fit for each pixel is measured by determining where that pixel falls in relation to the concentric probability regions surrounding each class center; a lower number indicates better fit. For pixels classified as GM, for example, the deepest pink colours indicate that those pixels were among the closest 5% of all pixels in the entire scene to the GM class center.

restricted region, where the interaction of tidal flow with complex bottom topography can result in the formation of fronts between different water types. The classifiers also show promise as a valuable tool for analysing patterns of seasonal and inter-annual variability in water type distributions over the region. Our next steps in the development of feature-based optical water type classification techniques include the development of an automated approach for the definition of water type classes. We plan to implement an adaptive scheme to search through feature space for local minima in pixel density; this approach will enable us to mathematically define class boundaries, facilitating the development and application of statistical decision rules for classification.

We are confident that once the SeaWiFS atmospheric correction issues are resolved and the accuracy of the water-leaving radiance retrievals in the Gulf of Maine improves, we will be able to expand and apply our feature-based classifiers to track water types over time. As an extension of our feature-based work, we will also explore a hybrid classification approach, which will involve integrating model-based inversion techniques and feature-based classifiers. The advantage of a hybrid technique is that it can capitalise on the predictive power of existing semi-analytic models, while taking advantage of the intrinsic features in the data, which are independent of assumptions inherent in the models. Development of classification techniques for Northwest Atlantic water types will be carried out in the context of *in situ* data collected during the ecological and hydrographic work for the GLOBEC Georges Bank program, as well as our research examining *in situ* optical variability in this region. Application of these classification techniques will contribute to the interpretation of the underlying properties that define optical water types, facilitating examination of spatial and temporal variability in water types using satellite ocean colour imagery.

REFERENCES

Brown, C.W. and J.A. Yoder. 1994a. Distribution pattern of coccolithophorid blooms in the western North Atlantic Ocean. Cont. Shelf Res. 14: 175-197.

Brown, C.W. and J.A. Yoder. 1994b. Coccolithophorid blooms in the global ocean. J. Geo. Res. 99: 7467-7482.

Gordon, H.R., O.R. Brown, R.H. Evans, J.W. Brown, R.C. Smith, K.S. Baker and D.K. Clark. 1988. A semi-analytic radiance model of ocean color. J. Geophys. Res. 93: 10909-10924.

Sathyendranath, S., T. Platt, E.P.W. Horne, W.G. Harrision, O. Ulloa, R. Outerbridge and N. Hoepfner. 1991. Estimation of new production in the ocean by compound remote sensing. Nature. 353: 129-133.

Subramaniam, A. and E.J. Carpenter. 1994. An empirically derived protocol for the detection of blooms of the marine cyanobacterium Trichodesmium using CZCS imagery. Int. J. Rem. Sens. 15: 1559-1569.

Zabicki, K.E. 1995. Determining the existence of cyanobacterial blooms using CZCS imagery. Student reports - NASA/University of Maryland Summer Fellowship Program in Remote Sensing of the Oceans.