OPTICAL CLASSIFICATION OF WATER TYPES BASED ON REMOTELY-SENSED OCEAN COLOUR

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

ABSTRACT

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. In contrast to previous studies that have focused primarily on pigment distributions, our goal is to extract as much information as possible from the full set of spectral bands. Our initial work includes a regional study of optical properties in the northwest Atlantic based on CZCS and SeaWiFS data, as well as the detection and identification of phytoplankton blooms around the globe. With this preliminary study, we have established a basis for the application of feature-based approaches for the optical classification of water types. Future work will involve integrating *in situ* data into classifier development, and exploring the utility of both model-based and feature-based classification approaches.

INTRODUCTION AND BACKGROUND

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. Remote sensing of ocean colour has significantly expanded our ability to study spatial and temporal variability in phytoplankton abundance and distribution, but 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. 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.

APPROACH, RESULTS, AND DISCUSSION

The work presented here is a preliminary analysis of CZCS and SeaWiFS data to explore the feasibility of using feature-based classification techniques to identify and delineate optical water types. The first portion of 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. This regional analysis was carried out in the absence of *in situ* data, with the goal of establishing whether there was a basis for separating, using ocean colour data, various water types found in the restricted geographic region on and around Georges Bank. The second part of our analysis focused on discrimination of globally-occurring waters in which the dominant source of optical variability is the presence of specific phytoplankton. This analysis was restricted to classification of ocean regions for which ground-truth information was available. Reports based on *in situ* observations of oceanic regions containing particular taxa of phytoplankton were identified from the literature, and the corresponding CZCS data were obtained.

Regional Water Types on and around Georges Bank

Our analysis of water types in the northwest Atlantic focused on the Georges Bank/Gulf of Maine area, but included water types farther to the south for context. For this initial analysis we subjectively selected geographic locations based on general knowledge of the hydrography and bathymetry, combined with examination of CZCS imagery from 7 July 1980 and 13 June 1980, and an 8 October 1997 SeaWiFS image.

For the CZCS image from 7 July 1980 (Fig. 1), we considered five locations: Gulf Stream waters (**GS**), Central Mid-Atlantic Bight waters (**CMAB**), waters located just south of Georges Bank (**S of GB**), waters on Georges Bank (**GB**), and Gulf of Maine waters (**GM**). Normalised water-leaving radiances (nL_w, Gordon *et al*. 1988) in each spectral band were extracted from the CZCS image for randomly selected pixels from within each of these regions. For this initial work, we examined only two types of features, the absolute nL_w signals for three individual bands (443, 520, and 550 nm) and ratios of these nL_w signals. Our simple approach was to project the data for all the regions into feature space and visually search for clusters using 3-dimensional graphics. In the single-band feature space, the five northwest Atlantic locations projected into five distinct clusters (Fig. 2A), suggesting that waters in these areas can be well-distinguished using a few spectral bands of ocean colour data.

For a preliminary look at temporal variability in this region we examined another clear CZCS image from one month earlier (13 June 1980, image not shown). The cloud-free area did not extend as far south, so we confined our analysis to 3 locations: the North Flank of Georges Bank (characterised on that day by striations in ocean colour that are thought to be biologically relevant, but caused by physical processes; Yentsch *et al*. 1994), the South Flank of Georges Bank, and the Gulf of Maine. In the restricted region on and around Georges Bank, 3 clusters emerge (Fig. 2B); these 3 water types are clearly distinguishable, and are different from waters in the same locations a month later (Fig. 2A). Comparison of the mean pigment concentrations for these waters (Fig. 3) emphasises that our approach takes advantage of more information inherent in ocean colour data than is contained in a derived pigment image. Projection in 3-dimensional band ratio feature space (443/520, 520/550, 443/550; not shown) revealed that water types with similar chlorophyll concentrations, such as the waters S. of Georges Bank and in the Gulf of Maine, clustered together. Importantly, these same waters formed distinctly separate clusters in the single-band feature space.

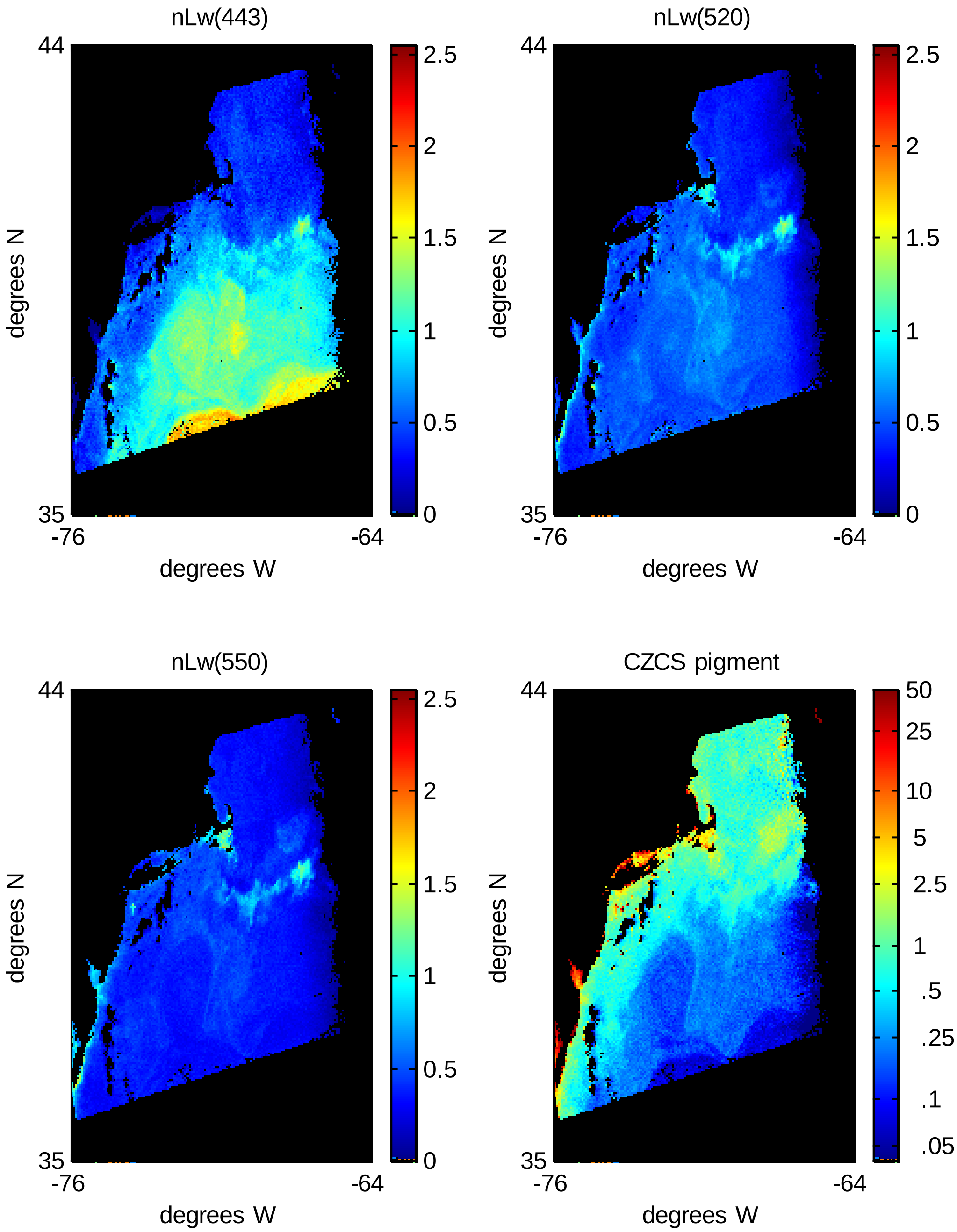


Figure 1. CZCS-derived normalized water-leaving radiances (mW/(sr·cm^2·µm)) in three spectral bands nLw(443), nLw(520), nLw(550), and chlorophyll (mg/m^3) for the northwest Atlantic, 7 July 1980. Georges Bank appears as a region of high pigment in the chlorophyll image. Land and clouds appear black.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

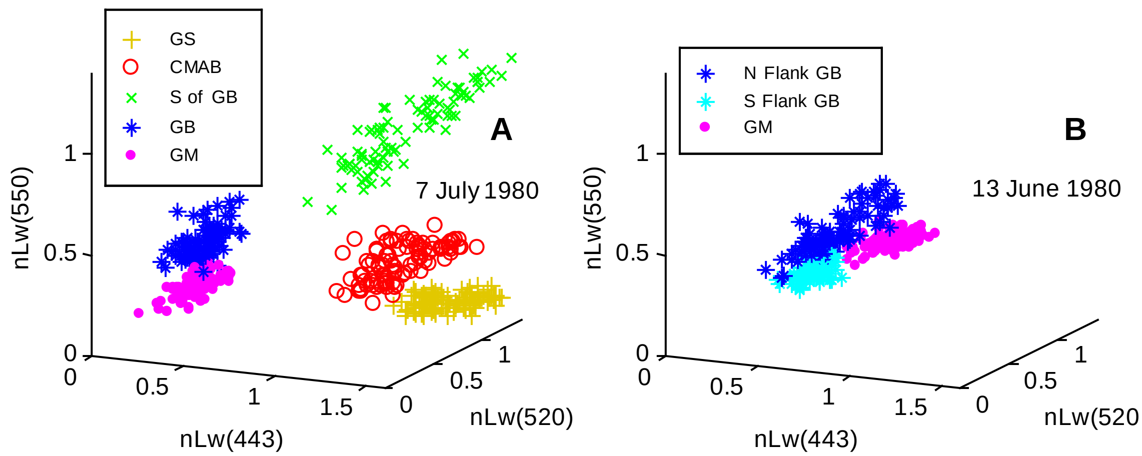


Figure 2. A) Projection of CZCS data in a single-band feature space for northwest Atlantic waters on 7 July 1980, including waters on Georges Bank and in the Gulf of Maine. The waters from five different locations form distinct clusters and are clearly distinguishable. B) Projection of CZCS data for waters on and around Georges Bank on 13 June 1980; these three water types are also easily separable in the single-band feature space and show some differences compared to the same locations sampled 1 month earlier. **GS** = Gulf Stream waters; **CMAB** = Central Mid-Atlantic Bight waters; **S of GB** = waters south of Georges Bank; **GB** = waters on Georges Bank; **GM** = Gulf of Maine waters; **N Flank GB** and **S Flank GB** are waters on the north and south flanks of Georges Bank respectively.

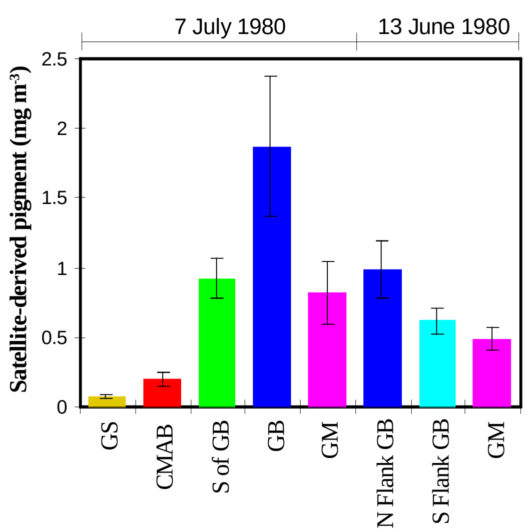


Figure 3. Chlorophyll-like pigment concentration derived from CZCS data for the clusters shown in Figs. 2A,B. Mean values (bar height) and standard deviations (black lines) were calculated for each of the different locations and dates from the same pixels used in Fig. 2. Interpretation based on derived pigment concentrations alone will fail to distinguish different water types which do not exhibit differences in pigment concentration. For example, although pigment concentrations for the **S of GB** waters (indicated by height of green bar) and **GM** waters (height of leftmost pink bar) on 7 July are the same, these water types are readily separable as distinct clusters in the single band projections shown in Fig. 2A.

In analysing the SeaWiFS image from 8 October 1997 (Fig. 4), pixels were selected from the same five locations (**GS**, **CMAB**, **S of GB**, **GB**, and **GM**) identified in the 7 July 1980 CZCS image. As a result of the increased spectral resolution of SeaWiFS data over CZCS data, the new sensor provides more degrees of freedom for classification (Fig. 5A), and the information contained in the additional spectral bands can be exploited to better discriminate water types. For comparison to the CZCS clustering analysis in this preliminary work, similar features (normalised water-leaving radiance at 443, 510 and 555 nm) were extracted from the SeaWiFS data from each of these five regions and

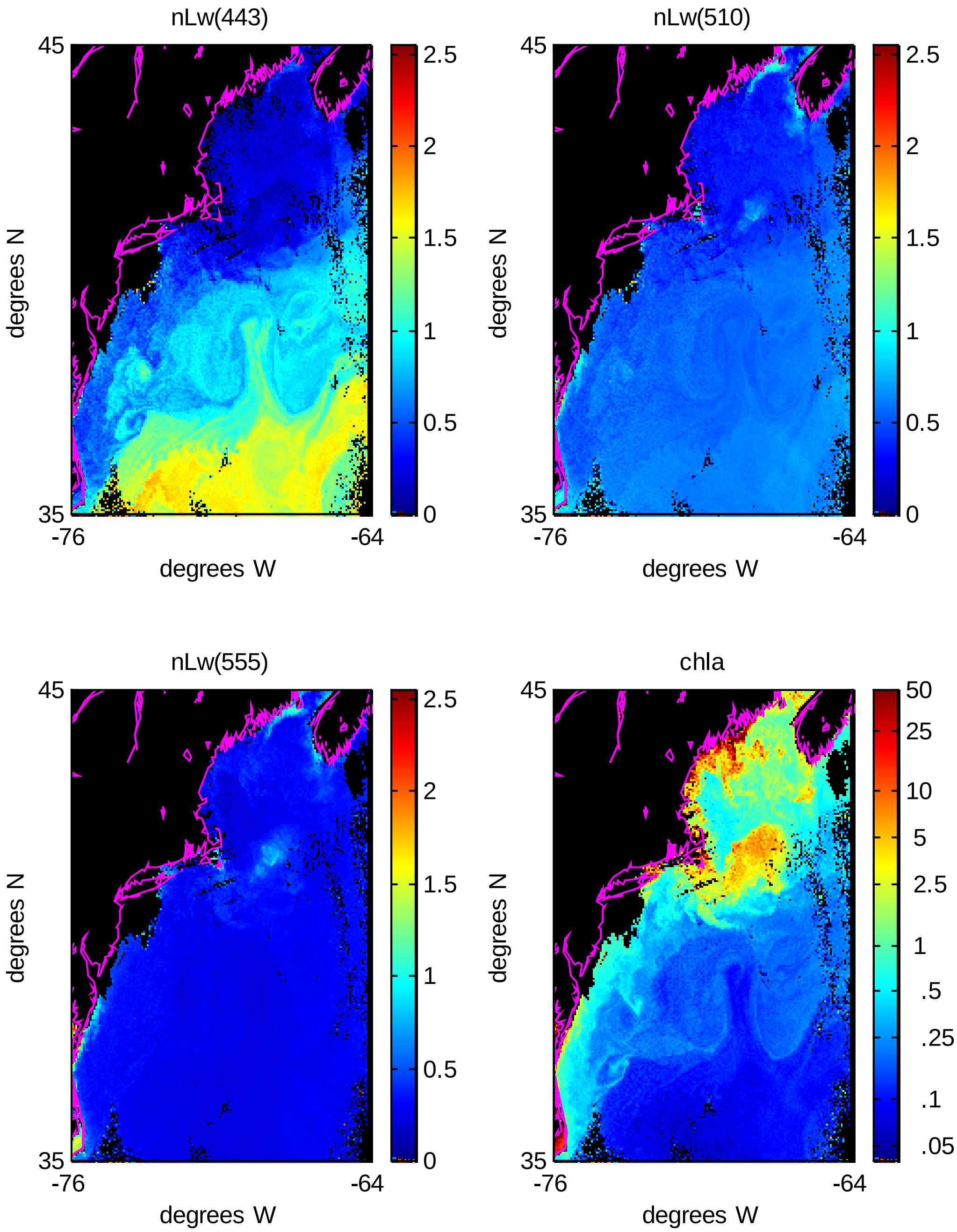


Figure 4. SeaWiFS-derived normalized water-leaving radiances (mW/(sr·cm^2·µm)) in three spectral bands nLw(443), nLw(510), nLw(555), and chlorophyll (mg/m^3) for the northwest Atlantic, 8 October 1997. Georges Bank appears as a region of high pigment in the chlorophyll image. Note that the pigment levels over the Bank are considerably higher on this day than they were on 7 July 1980 (see Fig. 1). Land and clouds appear black, coastline is shown in pink. This scene was processed using SeaDAS 3.1 from Level 1A HRPT data (ver. 1) acquired from the Goddard DAAC.

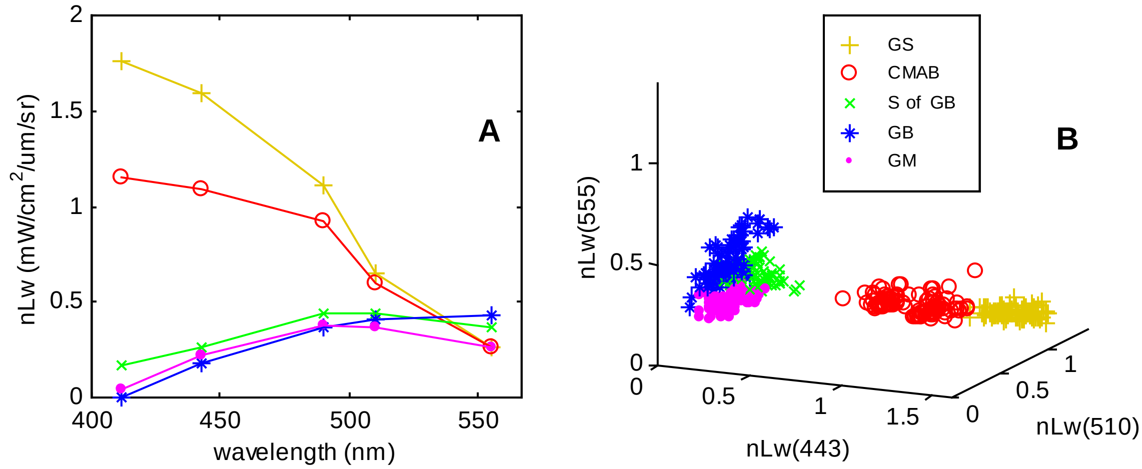


Figure 5. A) Mean spectra of normalised water-leaving radiance for five water types (see legend above) identified in the SeaWiFS imagery for the northwest Atlantic (including waters on Georges Bank and in the Gulf of Maine) on 8 October 1997. B) Projection of the SeaWiFS data for these same water types in a single-band feature space (similar to that shown in Fig. 2). The waters from these five different locations form distinct clusters, and project onto the same regions of feature space as did the CZCS-derived clusters from 7 July 1980 (Fig. 2A). **GS** = Gulf Stream waters; **CMAB** = Central Mid-Atlantic Bight waters; **S of GB** = waters south of Georges Bank; **GB** = waters on Georges Bank; **GM** = Gulf of Maine waters.

projected in feature space (Fig. 5B). Despite the fact that the SeaWiFS imagery was captured with a different sensor and during a different season several years after the CZCS image, the same water types are clearly distinguishable, occupying similar regions in this three-dimensional feature space. Notably, the waters over Georges Bank (**GB**) project onto the same location in this feature space for both the 8 October 1997 image and the 7 July 1980 image, even though the October 8 image revealed much higher derived pigment values for these waters (compare Figs. 1 and 4). Our simple clustering approach has the ability to correctly identify a water type over time despite large pigment differences, as well as to distinguish different water types that have the same pigment concentration (see Fig. 3), underscoring the fact that this technique is able to exploit additional information in ocean colour data not reflected in pigment distributions.

Some variability in these water types is evident from a comparison of Figs. 2A and 5B. For example, both the **CMAB** and **GS** waters in the autumn image (Fig. 5B) form tighter clusters compared to the summer image (Fig. 2A), which is likely attributable to a change in the optical properties of these water types as winter approaches. Perhaps the most notable difference between the 1980 and 1997 pictures is the cluster representing the waters located just south of Georges Bank (**S of GB**). This water type appeared as a distinct and quite separate cluster for 7 July 1980, and was easily distinguishable in the CZCS image as a bright feature in all three spectral bands, accompanied by high derived pigment values (Fig. 1). Such a bright region is not apparent in the 8 October 1997 SeaWiFS imagery (Fig. 4), and although this water type is clearly distinguishable from **GB** and **GM** waters (Fig. 5B), the water south of Georges Bank in autumn appears to be closely related to that on the Bank and in the Gulf of Maine, since the **S of GB** cluster projects near these other water types in feature space.

This variability in the optical properties of the **S of GB** water type between the two images could either be attributable to an episodic event (e.g., the occurrence of a coccolithophore bloom in July 1980), or it could be the result of systematic seasonal changes, which recur annually. Application of an automatic water type classification scheme to a long time series of imagery would allow characterisation of seasonal and annual temporal variability in the optical properties of northwest Atlantic water types.

Classification of Globally-Occurring Phytoplankton-Dominated Waters

Ocean regions considered in this initial cluster analysis of globally-occurring phytoplankton-dominated waters differed not only in the phytoplankton taxa present, but also in the abundance of phytoplankton cells present. The six data sets we considered were compiled based on observations reported in the literature (Table 1). Some represented bloom conditions of an identified phytoplankton taxon; for example, one water type (**cyantt**) was extracted from a *Trichodesmium* bloom (Subramaniam and Carpenter 1994) in the Gulf of Thailand (Fig. 6). Conditions with sub-bloom concentrations and mixed-taxa phytoplankton communities were also included in the analysis. For example, we used CZCS data for the Southern California Bight (**coccus**) corresponding to the first 13 stations of a ship-based transect (Fig. 7) conducted by Balch *et al*. (1989). The authors reported crossing a 25 km long patch containing the coccolithophore *Umbilicosphaera sibogae* (representing at most 60-75% of the phytoplankton biomass); abundances in the patch ranged between 2000 - 4000 cells/l with no detached coccoliths detected. For this same transect, Eppley *et al*. (1984) reported that *U. sibogae* was most abundant in the first 27 km of transect, but was present with other species including diatoms. This is in contrast to coccolithophore bloom conditions on the northwest European Shelf (**cocccb**) where Holligan *et al*. (1983) reported very high concentrations of *Emiliana huxleyi* (8.5×10^6 cells/l) and the presence of abundant detached coccoliths (up to 10^8 detached coccoliths/l).

For each of the phytoplankton data sets (Table 1), normalised water-leaving radiances in each spectral band were extracted from the corresponding CZCS data for

Table 1. Summary of data used in preliminary analysis of global ocean phytoplankton blooms.

Data set:	**coccus**[1]	**dinogp**[1]	**cocccb**[2]	**dinord**[2]	**cyantt**[3]	**cyanta**[3]
group:	coccolithophore	dinoflagellate	coccolithophore	dinoflagellate	cyanobacteria	cyanobacteria
species:	*U. sibogae*[a]	*G. polyedra*[a]	*E. huxleyi*[b]	no *in situ* obs.[c]	*Trichodesmium*[d]	*Trichodesmium*[d]
location:	E Pacific shelf (S. CA Bight) 117.5W 32.5N	E Pacific shelf (S. CA Bight) 117.5W 32.5N	NW Euro. Shelf (NW of France) 5.43E 50.04N	W Atlantic slope (NY Bight) 73.7W 39.3N	Gulf of Thailand (off Khanom R.) 100E 8N	NW Australia (Dampier Ar.) 119E 21S
date:	4 Oct. 1981	4 Oct. 1981	29 May 1982	3 Sept. 1985	21 Feb. 1982	1 Nov. 1980
CZCS file:	81278193901	81278193901	82149111148	n/a	82052051917	80306042051
resolution:	~1 km	~1 km	~1 km	~4 km	~1 km	~4 km
# pixels:	100 (st. 1-13)	40 (st. 31-36)	100	100	100	100

[a]Balch *et al*. 1989; Eppley *et al*. 1984. [b]Holligan *et al*.1983. [c]Brown & Yoder 1994b. [d]Subramaniam & Carpenter 1994.
[1]Barney Balch provided sample station positions; Curt Vandetta provided data from West Coast Time Series archives.
[2]Chris Brown provided data for 3 Sept. 1995 and the NW European Shelf (for which Steve Groom provided images).
[3]John Ryan processed Thailand, NW Australia data; Dan Ziskin and Frances at NASA/DAAC expedited data orders.

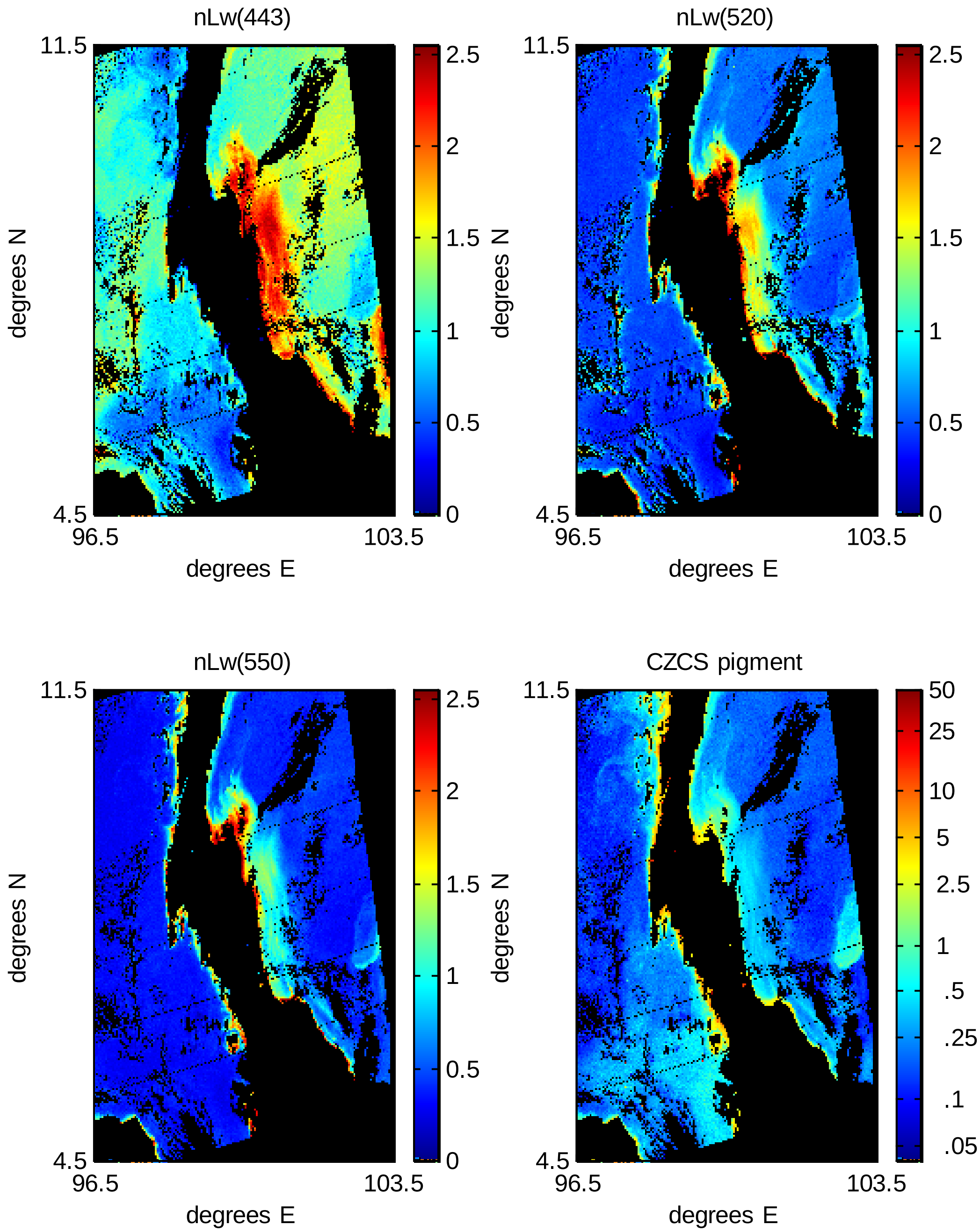


Figure 6. CZCS-derived normalized water-leaving radiances (mW/(sr·cm^2·µm)) in three spectral bands nLw(443), nLw(520), nLw(550), and chlorophyll (mg/m^3) for the *Trichodesmium* bloom in the Gulf of Thailand off the Khanom River east of the Thai peninsula (between ~6^o and 9^{o}N), 21 February 1982. Land and clouds appear black.

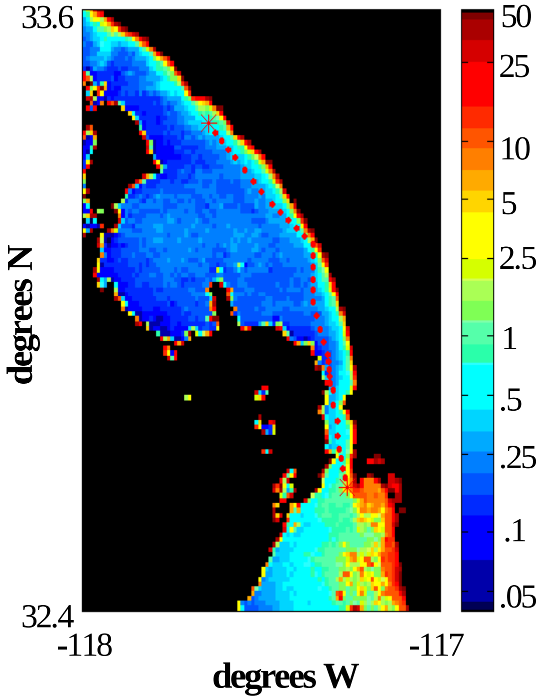
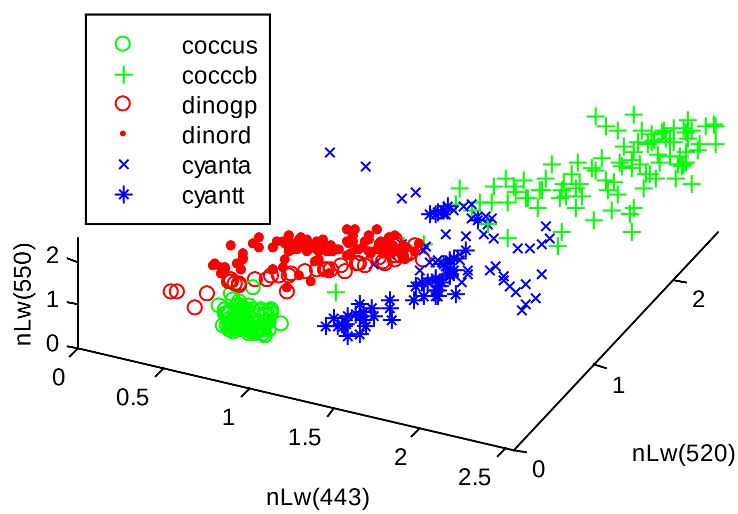


Figure 7. CZCS-derived chlorophyll-like pigments (mg m^{-3}) for the Southern California Bight, 4 October 1981. Transect stations occupied by Balch et al. (1989) are shown as red points. The coccolithophore *U. sibogae* dominated the phytoplankton biomass at the 13 northernmost stations and the dinoflagellate *G. polyedra* was highly abundant at the 6 southernmost stations. Land and clouds appear black.

selected pixels in the regions of interest. Single-band features and band-ratio features derived from the nL_w data were considered. Following feature extraction, the data were projected in several feature spaces and clusters identified. In a 3-dimensional single-band feature space (Fig. 8), the six data sets projected into 4 distinct clusters; the 4 different water types represented by these clusters are dinoflagellate-dominated waters, *Trichodesmium* bloom waters, coccolithophore bloom waters (with abundant detached coccoliths) and non-bloom coccolithophore waters (no detached coccoliths). Using this simple approach, it is possible to distinguish between oceanic regions containing distinct blooms of different major taxa, and also to separate coccolithophore waters into two optically different types. With this same approach we also found some evidence that we could distinguish "other waters", those outside the areas in each CZCS scene where phytoplankton blooms were documented. In single-band feature space, the four phytoplankton water types were clearly separable from the corresponding "other waters" (not shown). In addition, a preliminary analysis of uncorrected reflectance in the 750 nm band indicates that this feature may have discriminating power for cyanobacterial blooms; the *Trichodesmium* waters were associated with higher reflectance values in this band than the coccolithophore water types. This observation is consistent with reports of the

Figure 8. Projection of phytoplankton-dominated waters in single-band feature space. Data from six ocean regions cluster into four distinct water types. Dinoflagellate-dominated waters (shown in red), *Trichodesmium* bloom waters (shown in blue), and two types of coccolithophore waters (shown in green, bloom (+) and non-bloom (o) concentrations) are clearly distinguishable.



detection of *Amphanizomenon* and *Nodularia* blooms in the Baltic Sea using near-infrared signals from the LANDSAT MSS satellite (Horstmann *et al*. 1978, Ulbricht 1983a,b; all as cited in Borstad *et al*. 1992).

INDICATIONS FOR FUTURE RESEARCH

This initial work demonstrates the ability to use ocean colour data to distinguish between northwest Atlantic water types in summer and autumn, including those waters occurring within a spatially restricted region such as Georges Bank, where the interaction of tidal flow with complex bottom topography can result in the formation of fronts between different water types. Further development of classification techniques for these waters will be carried out in the context of *in situ* data collected during the ecological and hydrographic work in progress for the GLOBEC Georges Bank program, as well as our current research examining *in situ* optical variability in this region. We have also established the basis for developing a comprehensive approach to optically classify several different types of phytoplankton blooms simultaneously by automatically distinguishing them from each other. The next steps in the development of feature-based water type classifiers include the rigorous definition of water type classes, the formalisation of class boundaries, and the development and application of statistical decision rules for classification. Statistical decision theory can be used to derive an optimum classification rule if multivariate probability density functions are known for each class or if an empirical probability model can be obtained by statistical estimation. Alternatively, a decision rule may be derived directly from the distribution of samples in feature space. Feature-based classification is also possible where no *a priori* information is available, through the exploitation of naturally occurring groupings or clusters in the data. Features can then be empirically related to water type.

An immediate extension of this feature-based classification work is a hybrid approach, involving an integration of model-based inversion techniques and feature-based classification approaches. Combining model-based and feature-based techniques is a robust approach to the classification problem, since a hybrid technique can capitalise on the predictive power of the existing semi-analytic models as well as take advantage of the intrinsic features in the data that do not rely on assumptions inherent in the models. In this manner, the best attributes of both classification approaches can be exploited. Application of these classification techniques will contribute to the interpretation of the underlying properties that define water types, and will facilitate examination of spatial and temporal variability in water types which can be optically discriminated.

REFERENCES

Balch, W.M., R.W. Eppley, M.R. Abbot and F.M.H. Reid. 1989. Bias in satellite-derived pigment measurements due to coccolithophores and dinoflagellates. J. Plankton Res. 11: 575-581.

Borstad, G.A., J.F.R. Gower and E.J. Carpenter. 1992. Development of algorithms for remote sensing of Trichodesmium blooms. pp. 193-210 in Carpenter, E.J., D.G. Capone and J.G. Reuter (eds.) Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs. Kluwer.

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.

Eppley, R.W., F.M.H. Reid and E. Stewart. 1984. Length of phytoplankton species patches on the Southern California Shelf. Continental Shelf Research, 3: 259-266.

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.

Holligan, P.M., M. Voillier, M., D.S. Harbour, P. Camus and M. Champagne-Philippe. 1983. Satellite and ship studies of coccolithophore production along a continental shelf edge. Nature 304: 339-342.

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.

Yentsch, C.S., D.A. Phinney and J.W. Campbell. 1994. Color banding on Georges Bank as viewed by coastal zone color scanner. J. Geophys. Res. 99: 7401-7410.

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.