Spring 2008

# Dynamics of phytoplankton community composition in the western Gulf of Maine

Timothy S. Moore
*University of New Hampshire, Durham*

# DYNAMICS OF PHYTOPLANKTON COMMUNITY COMPOSITION IN THE WESTERN GULF OF MAINE

BY

TIMOTHY S. MOORE

B. Sc., Worcester Polytechnic Institute, Worcester, MA, 1988

M. Sc., University of Rhode Island, Kingston, RI 1995

DISSERTATION

Submitted to the University of New Hampshire
in partial fulfillment of
the requirements for the degree of

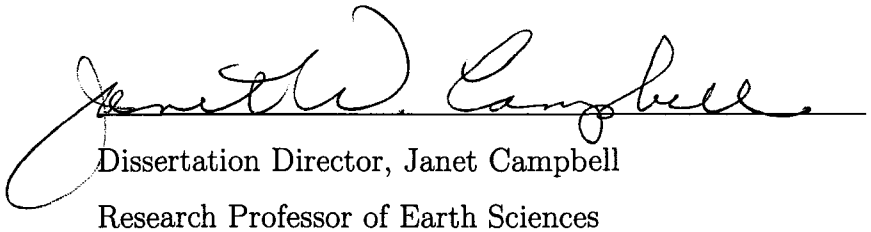
Doctor of Philosophy

in

Earth Sciences: Oceanography

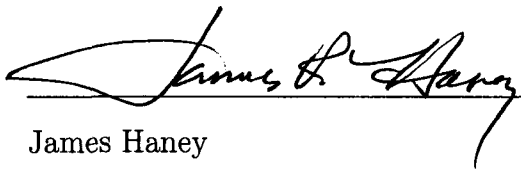
May 2008

This dissertation has been examined and approved.



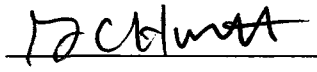
Dissertation Director, Janet Campbell

Research Professor of Earth Sciences



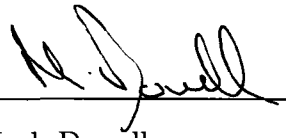
James Haney

Professor of Zoology



George Hurtt

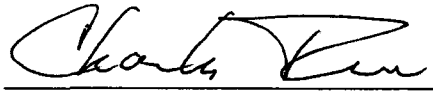
Associate Professor of Natural Resources



Mark Dowell

Systematic Observations of Land and Ocean

Joint Research Centre, Ispra Italy



Charles Trees

Head Remote Sensing and Data Fusion Branch

NATO Undersea Research Centre, La Spezia, Italy

April 21, 2008

Date

# DEDICATION

This work is dedicated to my parents, Ray and Peggy, and the cats.

# ACKNOWLEDGMENTS

There are many people to thank, and without their collective help and support this dissertation would not have been possible. First, I owe a great debt of gratitude and thanks to my main advisor, Janet Campbell. Janet provided many ideas, support, and belief in what I was doing. Its been a long ride, Janet. Also, I wish to thank the committee for their guidance and time. Specifically, thanks to Mark Dowell for supplying ideas, energy, and inspiration for this research. Thanks to George Hurtt for helpful suggestions and pushing me to raise the bar. Thanks to Jim Haney for his interest and helpful insight throughout the process, and to Chuck Trees for his comments and help, even from afar. This research also would not have been possible without the commitment of the crew of the R/V Gulf Challenger. Specifically, thanks to Paul Pelletier, Bryan Soares, and Deb Brewitt – a really great crew to go out to sea with. Thanks for taking us home when it got too rough. Thanks to all who participated on the cruises and formed the backbone of the data collection and processing team. This includes Chris Hunt, Mike Novak, Chris Manning, Rebecca Jones, Shawn Shellito, Deb Goodwin, Ken Fairchild and Amanda Plagge. You have been a great team to work with – many thanks. Thanks to Jason Perl from San Diego for HPLC guidance. Thanks to Erin Gordon and Megan Heidenreich for culture work, and to Gillian Fairchild for help with the HPLC. Special thanks to the OPAL PI's who have helped me over the years, including Doug Vandemark, Ru Morrison, Joe Salisbury, Jamie Pringle and Hui Feng. Special thanks to Amy Cline – you were a great neighbor to have and you helped keep me going. I hope I didn't slow you down with all my interruptions. Thanks to the OPAL staff over the years. Specifically, thanks to Theresa Hamer, Cheryl Moore, Donna Thibault, Laurinda Smith and Sheila Skane. I've really enjoyed interacting with you all. Thanks to my family for their belief that I could do this. Last but certainly not least, thanks to my wife Mary Maravic, who saw the good, the bad and the ugly and never wavered in her support.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

## DYNAMICS OF PHYTOPLANKTON COMMUNITY COMPOSITION IN THE WESTERN GULF OF MAINE

by

Timothy S. Moore
University of New Hampshire, May, 2008

This dissertation is founded on the importance of phytoplankton community composition to marine biogeochemistry and ecosystem processes and motivated by the need to understand their distributions on regional to global scales. The ultimate goal was to predict surface phytoplankton communities using satellite remote sensing by relating marine habitats – defined through a statistical description of environmental properties – to different phytoplankton communities. While phytoplankton community composition is governed by the interplay of abiotic and biotic interactions, the strategy adopted here was to focus on the physical abiotic factors. This allowed for the detection of habitats from ocean satellites based on abiotic factors that were linked to associated phytoplankton communities.

The research entailed three studies that addressed different aspects of the main goal using a dataset collected in the western Gulf of Maine over a 3-year period. The first study evaluated a chemotaxonomic method that quantified phytoplankton composition from pigment data. This enabled the characterization of three phytoplankton communities, which were defined by the relative abundance of diatoms and flagellates. The second study examined the cycles of these communities along with environmental variables, and the results revealed that the three phytoplankton communities exhibited an affinity to different hydrographic regimes. The third study focused on the implementation of a classifier that predicted phytoplankton communities from environmental variables. Its ability to differentiate communities dominated by diatoms versus flagellates was shown to be high. However, the in-

crease in data imprecision when using satellite data led to lowered performance and favored an approach that incorporated fuzzy logic. The fuzzy method is well suited to characterize the uncertainties in phytoplankton community prediction, and provides a measure of confidence on predicted communities. The final product of the overall dissertation was a time series of maps generated from satellite observations depicting the likelihood of three phytoplankton communities.

This dissertation reached the main goal and, moreover, demonstrated that improvements in the predictive power of the method can be achieved with increased precision and more advanced satellite-derived products. The results of this research can benefit present bio-optical and primary productivity models, and ecosystem-based models of the marine environment.

# CHAPTER 1

# PHYTOPLANKTON COMMUNITY COMPOSITION

## 1.1 Introduction

The composition of marine phytoplankton communities is a fundamental aspect of ocean ecology and biogeochemistry. Phytoplankton play a significant and dynamically active role in the global cycles of nutrients and elements, and form the base of marine food webs. Although they represent only 1% of the global standing stock of organic carbon, phytoplankton have short generation times (from days to weeks) and thus account for an estimated 45% of the global annual primary production (Falkowski and Raven, 1997). These roles – both from a biogeochemical and ecosystem point of view – depend on the community composition.

Certain algal groups perform different biogeochemical roles. These groups can be defined along taxonomic lines or categorized by the biogeochemical 'function' they perform. Diatoms, for example, are a group associated with high primary productivity and carbon export; coccolithophorids are important to carbon fluxes and the production of dimethyl sulfide which is linked to seeding cloud formation; cyanobacteria are numerically the most abundant phytoplankton and are significant components of open ocean communities.

The structure of the higher trophic levels and the general ecology of a particular marine ecosystem ultimately depend on the composition of the phytoplankton community. Fishery production, for example, has been linked with diatom communities (Cloern and Dufford, 2005; Barber and Hiscock, 2006). Thus, knowing the distribution of phytoplankton communities is central to further understanding

marine ecosystems and biogeochemical processes.

Phytoplankton species vary in their distributions over space and time. Taxonomically, there are more than 5,000 oceanic species distributed across at least 10 classes from several kingdoms, and range in cell diameter from less than 1 micron up to 1 millimeter (Falkowski and Raven, 1997). When it comes to describing the communities phytoplankton species form, there are many ways to view and define them. For example, communities can be defined by morphological traits (Reynolds et al., 2002), by size distribution (Sieburth et al., 1978), or along functional type (Moore et al., 2002). Despite their importance and regardless of the choice of community view, knowledge of the geographic distribution of phytoplankton communities remains elusive because of the high degree of difficulty in routinely measuring the community composition in a constantly changing ocean environment.

Microscopic identification remains the only direct method of quantifying phytoplankton composition, but this method has many drawbacks. It is an intensively time-consuming process and requires technical expertise in observing morphological differences that can be difficult to discern at microscopic levels even with training and experience. The recent application of high performance liquid chromatography (HPLC) for the detection and characterization of phytoplankton pigments from freshwater and marine water samples has promoted the development of new techniques for analyzing phytoplankton populations. The basic premise is built on the use of pigment markers as a means of separating phytoplankton classes apart from one another. Recent advances in this field have now permitted the rapid estimation of phytoplankton composition from water samples, and thus HPLC remains the most promising technique for assessing field populations of phytoplankton.

It is not practical to depend solely on laboratory analyses of water samples to characterize phytoplankton communities at the temporal and spatial scales desired. For this reason, satellites have become the prime source of monitoring the oceans as they provide global coverage at a spatial resolution of approximately 1 km and the additional benefit of highly repeated orbits which can provide daily regional coverage for several important ocean properties. Advances in satellite oceanography have revolutionized our understanding of phytoplankton biomass distributions on regional

and global scales. Phytoplankton biomass has been estimated from remote sensing instruments onboard earth-orbiting satellites since 1978 with the Coastal Zone Color Scanner (CZCS) instrument, and more recently with the Sea-viewing Wide Field of view Sensor (SeaWiFS) and the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. These instruments, commonly referred to as 'ocean color' sensors, measure the upwelling light field at key wavelengths in the visible part of the spectrum. Empirical algorithms have been developed to estimate total chlorophyll *a* concentration, which serves as a proxy for phytoplankton biomass. While biomass can be estimated using ocean color sensors, phytoplankton composition cannot with few exceptions (e.g., coccolithophorid blooms can be revealed by their unique optical signature).

### Statement of the problem

Phytoplankton community composition at regional and global spatial scales remains an elusive biological property of the oceans. Currently, there are few published methods available that address this issue. This is due in part to the scarcity of data and the complexity of the problem. Satellite observations are critical to obtaining synoptic measurements of the oceans. The central challenge lies in the ability to relate satellite measurements to different phytoplankton communities – however defined – and to spatially delineate these communities that have dynamic boundaries.

## 1.2 Background

Marine phytoplankton biogeography is a research area concerned with the spatial and temporal distribution of phytoplankton in the world's oceans. Early studies in the late 1800s (starting with the expeditions of the *Challenger* from 1872-1876) and early 1900s focused on the distribution of regional phytoplankton species, and led to the concept of *plankton elements*. Initially proposed by the researcher H.H. Gran (Semina, 1997), a plankton element is a phytoplankton assemblage characterized by an indicator species whose presence signified a distinct water mass upon

which the plankton types were dependent. As knowledge of local and regional distributions of phytoplankton taxa gradually increased during the twentieth century, global distributions began to emerge. Smayda (1958) was among the first researchers to compile and map the global distribution of diatom species, and speculate on the underlying factors controlling their distributions. Margalef (1961) proposed a method to classify the distribution of phytoplankton, also based on indicator species. The method was based on known relations between species and abiotic factors such as temperature, salinity and nutrients. Around the same time, researchers from the Soviet Union compiled phytoplankton distributions from data at over 2000 stations, and generated global maps of numerous species (Semina, 1997).

Longhurst (1995) recognized the importance of discontinuities between different marine ecosystems and advocated the partitioning of the global oceans into biogeographic provinces, much like terrestrial biogeographers do. However, Longhurst (1995) also realized that the paucity of accumulated knowledge of the distribution of marine organisms limited the degree to which marine biogeography could approach that of terrestrial environments. The knowledge of the distribution of oceanic phytoplankton species is hindered by the nature of the oceans, and the inability to adequately sample this environment. The oceans cover an area more than twice that of land, but the biology has been undersampled by orders of magnitude compared to terrestrial life, according to Longhurst (1995). The reasons for this stem in part from the nature of the marine environment – the sampling of the oceans has been limited to ship-board observation, which greatly undersamples hard-to-reach areas of the ocean. The horizontal movement of water transports species from one location to another, and boundaries between different water masses are leaky in that organisms can be exchanged across these dynamic interfaces. In addition, phytoplankton have a vertical distribution that is affected by their own ability to regulate water column position (e.g., buoyancy and flagellate-motility), as well as the vertical motion of the water column.

Satellite observations of the oceans – a relatively recent technological advancement – have overcome many problems associated with sampling and observation over the oceans. Satellites now offer almost-daily global coverage of the oceans, and rep-

resent a new source of information to be exploited. Contemporary oceanographic campaigns have augmented field collections with satellite observations (Sathyendranath et al., 2004; Kamykowski and Zentara, 2003; Platt et al., 2007). Satellite coverage extends to areas rarely sampled by ship, and has the ability to capture events with short turnover times that can be missed by infrequent repeat visits from ship. The approaches that these studies employ vary, but they share the use of satellite sea surface temperature or ocean color satellite data (as radiance fields or the derived chlorophyll-*a* product), or some combination thereof. What follows is not an exhaustive list, but leading examples that differ in their approach in using satellite data for identifying phytoplankton communities.

One of the first algorithms to use satellite data to identify a specific phytoplankton type was that of Brown and Yoder (1994) in detecting coccolithophore blooms in ocean color images. The principle behind this application is the observed effect on the light field from the release of the individual coccolith plates into the water column resulting from cellular death of these particular phytoplankton. The coccolith plates, made from calcium carbonate, are efficient at reflecting light, and elevate the levels of upwelled light as detected by ocean color satellites in all visible wavelengths. Brown and Yoder (1994) characterized this effect and demonstrated that ocean color satellites could detect this phenomenon. Iglesias-Rodriguez et al. (2002) associated coccolithophore blooms based on the Brown and Yoder (1994) model with the corresponding water characteristics in terms of temperature and photosynthetically available radiation (PAR) from co-located satellite images. The relationship served as the basis of a probability function to predict coccolithophore blooms based on these water characteristics, and was used in a forecast model to predict the effect of global climate change on the future distributions of coccolithophore blooms.

Subramaniam et al. (1999) developed an algorithm for detecting the cyanobacteria *Trichodesmium spp.* from ocean color radiance data. *Trichodesmium spp.* is a colonial forming algae that is responsible for most of the $N_2$ fixation in the open oceans, and can represent a significant source of new production in tropical and subtropical seas. *Trichodesmium spp.* contain gas vacuoles which produce a distinctive effect on the backscattering of light. Similar to the detection of coccolithophorids,

the backscattering signal can be detected in ocean color satellite radiances, and thus their distributions can be mapped from satellite data.

Kamykowski and Zentara (2003) presented a method to infer phytoplankton composition based on the upper ocean nutrient status (replete versus deplete) using the difference between sea surface temperature (SST) and the nutrient-depletion temperature (NDT) for nitrate. Basing phytoplankton community composition on HPLC data from cruises in the Atlantic Ocean and off the coast of California, the distribution of three different phytoplankton classes were organized along a progression of SST minus NDT. Phytoplankton community composition could then be derived from satellite SST imagery and a climatology of NDT. The success of this implementation depends on knowing the actual NDT values which vary from region to region (and are scarce in many regions). The use of a climatology as a substitute for actual contemporaneous conditions introduces uncertainty.

Sathyendranath et al. (2004) used ocean color data to differentiate diatom-dominated communities from other types of phytoplankton community (collectively referred to as 'mixed'). The approach they took was based on chlorophyll retrievals from different algorithms that used the radiance fields from the SeaWiFS sensor. Initially, HPLC data were used to segregate diatom-dominated samples from others, and specific bio-optical algorithms were developed for each data pool from co-measured radiance and other bio-optically relevant information. The rationale behind this was that the behavior of the spectral absorption coefficients for the two populations were distinct, which were then used to parameterize unique algorithms. Look-up tables were generated for each algorithm at two reflectance ratios (510:555 and 490:670 nm). To classify image pixels, satellite radiance fields were used to retrieve the chlorophyll concentration from the lookup tables using the diatom-specific algorithm for both ratios, resulting in two chlorophyll values. This was repeated using the tables based on the mixed populations. The community selected was the one that had the smaller differences between the two retrievals. The overall success rate was 72% for discriminating diatom-dominated populations from mixed populations based on the *in situ* data used in parameterizing the models.

Alvain et al. (2005) also used ocean color data to identify different phytoplankton

groups. In this application, satellite radiances were co-located with *in situ* HPLC measurements, and the measured chlorophyll concentration was used to predict radiance ratios (i.e., the inverse of the ratio algorithm). The assumption here is that the differences between the measured radiances and the expected radiances are attributed to pigments other than chlorophyll *a*. These pigments were used as bio-markers of four specific phytoplankton groups, which were characterized based on pigment ratios (but did not employ CHEMTAX or any other published method). These groups were assumed to be the only choices available, and also that the dominant type was representative of the radiance data. Using these relationships based on 41 measurements, phytoplankton distributions were mapped with ocean color satellite data. The overall success rate was 61% based on the *in situ* data used in parameterizing the model.

These studies share similar shortcomings and drawbacks. Three of the studies are directed at specific phytoplankton species (Brown and Yoder, 1994; Iglesias-Rodriguez et al., 2002; Subramaniam et al., 1999). These are limited in their use towards identifying phytoplankton communities. The other two methods that use ocean color data (Sathyendranath et al., 2004; Alvain et al., 2005) have assumptions that link phytoplankton groups to differences derived from the spectral reflectance signal, and therefore require very precise radiometric accuracy. Ocean color satellite data are known to have errors from a variety of sources (e.g., atmospheric correction uncertainties), and the derived radiometric differences at the spectral bands could be explained by phenomena other than phytoplankton pigments. In addition, different combinations of in-water properties (e.g., particle backscattering and phytoplankton absorption) can produce the same spectral reflectance signature. Thus, the in-water source of spectral variation is uncertain. Furthermore, many pigments which characterize different phytoplankton groups (i.e., the carotenoids) occupy the same range in their spectral absorption properties, and would exhibit similar effects on the resulting reflectance spectra. It is important to note that these studies represent the few approaches that are published, and the paucity of methods underscores the difficulties associated with the nature of the problem.

With this in mind, the approach presented here attempts to avoid the shortcom-

ings inherent in ocean color data by basing prediction of phytoplankton community composition on variables connected to the physical environment and not on bio-optical algorithms. This dissertation is also based on a unique data set collected from a field program in the western Gulf of Maine for over 3 years, which is still in operation. The data set is largely composed of measurements along two monthly transects, which enabled systematic sampling of the same sites under different environmental conditions (Figure 1-1).

## 1.3 Goal

The ultimate goal of this dissertation was to predict the composition of surface phytoplankton communities using satellite remote sensing by relating marine habitats – defined through a statistical description of environmental properties – to different phytoplankton communities. While phytoplankton community composition is governed by the interplay of abiotic and biotic interactions, the strategy adopted here was to focus on the physical abiotic factors and implicit correlated biotic factors. This allowed for the detection of habitats from ocean satellites based on abiotic factors that were linked to associated phytoplankton communities.

## 1.4 Approach

The distributions of phytoplankton biomass and community composition have been traditionally linked with physical forcing or 'bottom-up' control (Riley et al., 1949; Margelef, 1978; Smayda, 1980; Longhurst, 2007), and with the effects of grazing from higher trophic levels, or 'top-down' control (Banse, 1994). Both types of controls exert pressures, and their combined effects operate simultaneously on shaping the phytoplankton community. Seasonal changes in temperature, light, and nutrients affect both abiotic and biotic factors that control phytoplankton community composition. Temperate ecosystems exhibit regular patterns of community succession from spring through the fall in concert with these variables and are taken as paradigm (Smayda, 1980). Several recent studies have reported observations of

large-scale community change in the north Pacific (Karl et al., 1997) and north Atlantic (Leterme et al., 2005) as a response to a change in the environment over the last few decades. These studies highlight the degree to which phytoplankton communities can shift in response to longer term changes in the environment at large regional scales.

A data set from the western Gulf of Maine was used to test the feasibility of mapping phytoplankton communities based on physical properties of the upper ocean. This data set consisted of *in situ* measurements spanning a 3-year period, largely taken from monthly cruises (Figure 1-1). Data collected included water samples used for HPLC analysis, and a suite of co-measured bio-optical and environmental properties. By relating phytoplankton composition deduced from HPLC data with field measurements of the hydrographic environment, it was possible to link phytoplankton community distributions with habitat conditions. These relationships served as the basis for a mathematical algorithm which was applied to 8-day composites of satellite data to predict the phytoplankton community composition given remotely sensed physical properties. The result was a time series of maps of phytoplankton community distributions at the same space and time scales as the satellite data. In addition, fuzzy membership maps were produced that represent the confidence associated with the community maps.

To map distributions of phytoplankton community composition from satellite data required the development of a *classifier* that could predict the phytoplankton composition – defined below – based on properties amenable to remote sensing. This approach is consistent with the long-standing theory that physical processes determine the structure of the pelagic ecosystem from phytoplankton to higher trophic levels (Margelef, 1978; Cullen et al., 2002). This theory is well supported by field work in freshwater and marine systems (Smayda, 1980; Reynolds et al., 2000). The methodology was based on associations between the phytoplankton communities and their aquatic habitats as defined through a set of physical/chemical characteristics that had known ecological relevance. These relationships served as the basis of the *classifier*.

The distributions were restricted to surface populations only. In nature, there is

a vertical component to any phytoplankton community. The surface community can be very different from populations that reside deeper in the water column. These deeper populations can, at times, be significant to the processes previously described. However, the main goal was to infer phytoplankton community composition from satellite data, and these measured properties were in general restricted to the surface. Therefore, the vertical distribution was set aside for the present research.

Phytoplankton composition was defined at a broad level based on the relative abundance of diatoms and flagellates. This is a simplification of the composition of natural assemblages, which are typically heterogeneous communities comprised of species from multiple taxonomic classes. Diatoms are generally associated with high levels of primary production and carbon export, and are often singled out as a distinct phytoplankton group in marine models (Moore et al., 2002; Hood et al., 2006). Phytoplankton populations dominated by diatoms also have been shown to have distinctive optical characteristics, and as a consequence influence the relationships embedded in ocean color and primary productivity algorithms (Sathyendranath et al., 2004; Claustre et al., 2005). Distinguishing diatoms from other types of phytoplankton was central to the overall goal of this research.

Flagellates are representative of a diverse group of phytoplankton from several classes. These include dinoflagellates, prymnesiophytes (e.g., coccolithophorids), cryptophytes, chrysophytes, prasinophytes, and chlorophytes. This collection of phytoplankton are noted for their motile ability through their flagella. This group includes species that can form toxic blooms (e.g., some dinoflagellates), calcifying organisms that affect the alkalinity of the seawater (i.e., coccolithophores), nuisance species (e.g., the prymnesiophyte *Phaeocystis spp.*), and assorted small flagellates that are important to ecosystem structure (e.g., prasinophytes, cryptophytes, chrysophytes, and chlorophytes). While each of these flagellate species can be important at any given time, they rarely dominate the community composition by themselves. Thus, for the purposes of this study, they were grouped together as a community.

The diatom and flagellate phytoplankton communities were defined by their fraction of total phytoplankton biomass in terms of *Chla*, and were named *diatom-dominated*, *mixed*, and *flagellate-dominated*. At this level of composition, HPLC-

derived pigments can be used to quantify the phytoplankton composition. The pigments can yield quantitative information at the class level. If one were interested in species or genus level of composition, pigment-based methods would not be suitable, and one would need to go to microscopic methods.

Three studies were completed that addressed questions that led to the final goal. The main objective and summary of each study were:

1) objective: To evaluate the pigment-based method CHEMTAX as a means of quantifying phytoplankton composition to the class level. The main question examined was: how sensitive is the CHEMTAX program to input parameters that control the output of the algorithm? The outcome of this study was a characterization of the sensitivity of the resultant phytoplankton composition to different initialization schemes for CHEMTAX. These results were also compared with an independent quantitative assessment of phytoplankton composition based on microscopic analysis. A manuscript based on this study is presented here in chapter 2.

2) objective: To identify phytoplankton communities and their cycles in the western Gulf of Maine, and to evaluate the linkage with physical factors and correlated biotic factors. The main question addressed was: how do the phytoplankton communities evolve over seasonal cycles, and how do these relate to environmental variables? A principal component analysis was applied to the hydrographic data to discern the dominant modes and variables associated with environmental variability. The first three principal components, representing over 70% of the variability, were dominated by surface water temperature and covarying seasonal signals in light intensity, winds, and nutrients. When the environmental data were projected in the new coordinate system as defined by the first three principal components, data points associated with the different phytoplankton communities showed separation into different hydrographic domains. These results, presented here in chapter 3, led to the design of and implementation of the methods of the third study, which related phytoplankton community composition to hydrographic conditions.

3) objective: To develop a methodology to map the distributions of phytoplankton communities in the Western Gulf of Maine from satellite data. The main ques-

tion is: can physical variables and correlated biotic factors be used as a basis to map phytoplankton communities using satellite data, and what is the uncertainty associated with the resulting maps? This was addressed by developing a classifier that was based on statistical relationships between phytoplankton communities – defined in terms of the relative contributions of diatoms and flagellates – and key environmental variables (temperature, light intensity, wind speed, salinity, and light attenuation). This classifier was applied to MODIS and SeaWiFS satellite data from the Gulf of Maine, and the maps generated depict phytoplankton communities with dynamic boundaries in space and time. A fuzzy classification method permitted the communities to have graded transitions and uncertainty levels to be represented by fuzzy membership maps. These results are presented in chapter 4, and the overall dissertation results are put into the larger context of phytoplankton ecology in chapter 5.

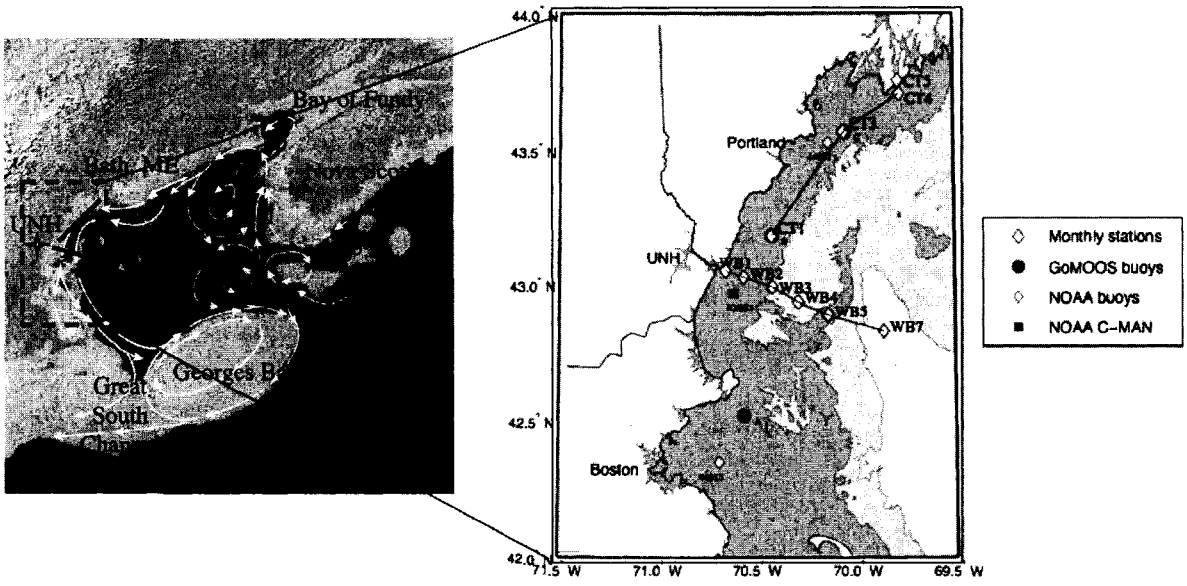



Figure 1-1: Study area and station locations associated with the two monthly transects - the Coastal Transect and the Wilkinson Basin Transect. Stations along these transects have been regularly visited every month since April 2004.

# CHAPTER 2

# AN EVALUATION OF METHODS FOR ESTIMATING PHYTOPLANKTON COMMUNITY COMPOSITION IN THE WESTERN GULF OF MAINE USING HPLC PIGMENT ANALYSIS

*This chapter has been submitted to the Journal of Plankton Research.*

**Abstract:** To understand seasonal patterns of phytoplankton communities in the western Gulf of Maine ecosystem, a pigment-based method was applied to *in situ* water samples to quantify the phytoplankton composition. This paper reports on the evaluation of a method based on HPLC pigments analyzed with CHEMTAX (Mackey et al., 1996). The method was applied to water samples and light measurements taken at stations along a cross-shelf transect from March through October 2005 under a variety of environmental conditions. CHEMTAX results were compared with phytoplankton community composition derived from microscopic cell counts. CHEMTAX estimates the fractional contribution of different algal groups to chlorophyll *a* whereas the microscopic technique estimates their carbon fraction. The comparisons between microscopy and CHEMTAX showed $r^2$ correlations that ranged from 0.35 to 0.82, but quantitative comparisons were problematic due to the nature of comparing class fractions of chlorophyll *a* (CHEMTAX) with class fractions of carbon (microscopy). Sensitivity to the initialization of CHEMTAX runs was also evaluated. The program requires an initial pigment ratio (IPR) table defining the ratio of accessory pigments to chlorophyll *a* for the algal classes present in the water samples. These ratios are known to vary among species acclimated

to different light levels. Sensitivity to the IPR table was assessed by comparing results from CHEMTAX using pigment ratios matched to each samples light level to those which used an average pigment ratio table. The relative differences were on the order of 10% across most algal classes, with dinoflagellates having the highest differences (18%) and chrysophytes the least (0.5%). We concluded that the use of microscopic cell counts was useful for revealing what species/groups are present in a sample, a function that is important in constructing the IPR. The use of light-dependent IPR tables must be considered as part of a larger set of parameters (i.e., the ratio limits and phytoplankton species) that operate in concert with each other. It is also important to have pigment ratios of locally observed species.

## 2.1 Introduction

The quantification of phytoplankton community composition has traditionally relied on microscopic cell counts. However, there are several major drawbacks with this method. Foremost, the counting of cells through a microscope is a time consuming process. Cells are individually identified and counted to species or genus level, and can require several hours to count one sample. This limits the quantity of samples that can be effectively analyzed. It also requires a trained and experienced individual to properly identify cells, particularly as cell size decreases. Many phytoplankton groups are too small to be identified with a traditional light microscope. This is especially true for many small flagellate species that occupy the same size range (2 to 10 $\mu$m) and have similar structural characteristics (e.g., flagella morphology), and picoplankton which have cell diameters less than 1 micron. Even with experienced analysts, microscopic counts are prone to subjective treatment and can vary by large amounts. Schluter et al. (2000) reported large differences in microscopic counts performed on the same samples from 2 different laboratories (up to a 10-fold difference). Wilhelm et al. (1991) quantified these kinds of differences and found that the coefficient of variation for cells counted using a microscope can vary between 15 and 50% between labs; others have shown similar results (Duarte et al., 2000).

An alternative method to microscopy is through the analysis of phytoplankton pigments. An important characteristic that differentiates algal groups is the composition of the pigments or light-harvesting system used in the capture of light energy for photosynthesis. Pigments have been used as phylogenic markers in determining the location of an algal class in the evolutionary tree (Rowan, 1989). The association between pigments and specific phytoplankton groups is called *chemotaxonomy*, and several methods have been developed that predict phytoplankton composition based on pigment levels.

Phytoplankton pigment concentrations can be routinely quantified using high performance liquid chromatography (HPLC) analysis. The underlying principle of pigment analysis is that most phytoplankton groups have unique pigments (or

pigment markers), such that the presence or absence of these groups can be identified based on the pigments found in a sample. There have been numerous methods built on this premise since the early 1990s, of which CHEMTAX has emerged as the most reliable (Llewellyn et al., 2005). Earlier methods allowed for marker pigments that had to be unique and could not be shared by multiple classes (e.g., Gieskes et al. (1988); Letelier et al. (1993)). This restriction has been shown to be a major drawback (Schluter et al., 2000). CHEMTAX is advantageous over these other methods, as it allows for different phytoplankton groups to share the same pigment, and a single group to contain multiple pigments. Since its introduction in 1996, CHEMTAX (Mackey et al., 1996) has been applied to freshwater lakes (Buchaca et al., 2005), estuarine systems (Lewitus et al., 2005; Ansotegui et al., 2003), and oceanic environments (Schluter et al., 2000; Riegman and Kraay, 2001; Llewellyn et al., 2005).

A flowchart schematic for CHEMTAX is shown in Figure 2-1. The program takes as input a matrix of sample phytoplankton pigments (measured from HPLC), and outputs the expected composition of phytoplankton at the class level. The algorithm is based on matrix factorization, and iterates to optimize the agreement between the observed pigment matrix (the input data) and the expected pigment matrix, which is derived from an initial pigment ratio (IPR) table – a user defined matrix – and the abundance of different algal populations (the output matrix). These latter 2 matrices are subject to modification by the algorithm to satisfy the optimization criteria. The IPR table is a $k$ by $m$ matrix with $k$ pigments and $m$ algal classes. Ideally, this table should represent all phytoplankton classes that are present in the samples and their expected pigment-to-chlorophyll $a$ ratios. This table is critical and should be set up with the utmost care. The degree to which this table can be modified is constrained by a user-defined matrix called the *ratio limit* table, which specifies the maximum percentage change during the iterative process.

Choosing the initial pigment ratios remains the biggest problem for CHEMTAX because there is insufficient knowledge of pigment ratios in the field. Consequently, this matrix is filled with values based on culture work. Since the algorithm is operating at the class level of taxonomy, these ratios should be representative of

species for that class. It is known, however, that pigment ratios can vary between species within a given class. The ratios also vary with environmental conditions and the physiologic state of the phytoplankton. In particular, the ratios are sensitive to light levels, as well as temperature and nutrient status. The recommendation for best performance of CHEMTAX remains to use pigment ratios from algal cultures under a range of conditions that mimic those in the field (Schluter et al., 2000; Llewellyn et al., 2005).

The present study examines the use of CHEMTAX for estimating the phytoplankton community composition using HPLC measurements from the western Gulf of Maine. The sensitivity to the configuration of the IPR table and other user-defined parameters is statistically quantified. The CHEMTAX results are also compared to microscopic cell counts. The microscopic assessments are in the form of carbon fractions, whereas the CHEMTAX results are in the form of chlorophyll-*a* fractions. Since these quantities represent different properties of biomass, they are not directly equivalent. The limitations and uses of using microscopic results in assessing CHEMTAX success is discussed.

## 2.2 Methods

### 2.2.1 Study site and data collection

The study site is located in the western Gulf of Maine (Figure 1-1). Data used in this study were obtained from March through October 2005 at 2m and 10m depth along the Wilkinson Basin Transect (stations labeled WB1 through WB7). Three types of phytoplankton analysis were performed – HPLC, cell counts using a light microscope, and cell counts from flow cytometry. All water samples were collected using Nisken bottles on station.

### 2.2.2 Cell counting methods

Cells greater than 10 $\mu$m were counted manually using a light microscope, and cells less than 10 $\mu$m, including picoplankton, were counted with flow cytometric

methods. Results from both analyses were combined to yield a total cell count for each phytoplankton group for each sample. Phytoplankton groups were partitioned into diatoms, dinoflagellates, cyanobacteria, cryptophytes, and small flagellates which included prymnesiophytes, chlorophytes, prasinophytes, and chrysophytes. Prymnesiophytes (mainly coccolithophorids) were observed and identified in the light microscope and grouped into the small flagellate category. Other small flagellates were not individually differentiated by flow cytometry nor by the light microscope. Cell counts were converted to carbon concentrations using biovolume-carbon relationships as described below.

**Microscopy: netplankton**

Whole water samples (250 - 1000 mL) were collected at surface (0-2 m) and at 10 meters and preserved with formaldehyde or Lugol's Solution for microscopic enumeration. Samples were concentrated into 20 mL vials using a settling method as described in Rowan (1978); whole water samples were transferred to a graduated cylinder and remained undisturbed for at least 48 hours allowing particulates to settle to the bottom. From this cylinder, the upper volume was siphoned away with mesh-covered tubing (to minimize unwanted particle removal), leaving behind a concentrated water sample, which was transferred into 20 mL vials. These samples were stored until counted (one to six months after collection). Prior to counting, the sample was mixed thoroughly (gently as to minimize disturbance of cellular and chain integrity), and 1 mL was pipetted into a 1 mL Sedgewick Rafter chamber. A Leica DM light microscope was used to count and identify phytoplankton cells greater than 10 $\mu$m diameter to the genus level. The number of cells counted ranged from 25 to 400 per sample. Cell dimensions were recorded for at least 20 cells of each genus and averaged. In some cases, genus groups were split into 2 size categories when there was numerically significant numbers of different sizes. Cell counts were included in the carbon estimation if at least 25 cells for each genus were observed.

#### Flow cytometry: pico- and nanoplankton

Approximately 1 mL of water was transferred from each whole water sample (at the time of collection) to a cryovial and combined with a small amount of formaldehyde (approximately 50 $\mu$L). These samples were refrigerated for 1-4 hours, and then placed in a liquid nitrogen dewar where they remained until analysis. All cryovial samples were analyzed at the J.J. MacIsaac Facility for Individual Particle Analysis at the Bigelow Laboratory in West Boothbay Harbor, Maine. The cytometric analysis partitioned the data into counts and concentrations for the following compartments: cyanobacteria, cryptophytes, eukaryotes less than 2 $\mu$m, eukaryotes between 2 and 5 $\mu$m, and eukaryotes between 5 and 10 $\mu$m. The eukaryotic community was assumed to comprise prasinophytes, cryptophytes, chlorophytes, chrysophytes, and/or prymnesiophytes. Cryptophytes could be identified from their size and fluorescent characteristics during the flow cytometric analysis, and their concentrations were subtracted from the '2-5' $\mu$m category. The remaining eukaryote size fractions were assumed to be a mixture of prasinophytes, chlorophytes, chrysophyes, and/or prymnesiophytes.

### 2.2.3 Phytoplankton carbon estimation

Cell concentrations were converted to biovolumes using average cell dimension measurements and species- and/or genus-specific geometric formulae (Sun and Liu, 2003). Diatoms, which may have had several species in a given sample with varying cell shapes, were assigned a formula that best matched their shape. For the flow cytometric results, all shapes were assumed to be spherical and an average size was given to each group. Biovolumes were converted to carbon units using carbon-biovolume relationships found in Menden-Deuer and Lessard (2000) for all groups. Each phytoplankton group was given its own carbon-biovolume relationship. For the small flagellate group (both size fractions), the carbon-biovolume relationship for prymnesiophytes was used.

### 2.2.4 HPLC method and analysis

Water volumes ranging between 500-1000 mL were filtered onto 25 mm Whatman G/FF filters (nominal pore size of 0.7 microns) for HPLC analysis. Filters were folded in half and wrapped in foil and immersed in liquid nitrogen, where they were kept until removed for analysis (typically 3-6 months). Upon removal, filters were placed in 15 mL test tubes containing 3.5 ml of 95% acetone which also contained a quantity of beta-apo-8'-caratenol (from Fluka) as an internal standard. These were kept refrigerated at -18C for 24 hours, and then were sonicated on ice for 1 minute and returned to refrigeration for another 24 hours. The test tubes were then centrifuged for 6 minutes at 4400 RPM, and filtered through a 0.2 micron disposable filter attached to a glass syringe with a Luer-Lock tip. Approximately 1 mL of the filtered extract was transferred to an amber 2 mL HPLC vial, which was placed in the autosampler for analysis.

The HPLC method used for all samples was based on Van Heukelem and Thomas (2001). The main hardware components of this system were all Series 200 Perkin Elmer products, and consisted of an autosampler, a pump capable of delivering multiple (maximum 4) mobile phase solvents, a photodiode array detector, and a column oven. The column type used for all samples was a ZORBAX Eclipse XDB C8 column (3.5$\mu$m, 4.6 x 150 mm). Absorbance was recorded at 436 and 450 nm. The autosampler was maintained at 4C during operation. The column temperature was maintained at 60C.

Pigments were identified and peak areas were converted into concentrations based on standards obtained from Sigma (chlorophyll *a* and chlorophyll *b*) and DHI. A sample chromatogram is shown in Figure 2-2. The 436 nm channel was used to calculate total chlorophyll *a* – defined as the sum of chlorophyll *a*, chlorophyllide *a*, and chlorophyll *a* epimer. The 450 nm channel was used for all remaining pigments.

### 2.2.5 CHEMTAX experimental design

The execution of CHEMTAX is sensitive to two key tables initialized at the start of the program – the IPR table and the ratio limits table. CHEMTAX was

run using two different configurations for the IPR tables. Scenario A sorted samples into three groups based on light-level ranges and used variable IPR tables that were matched to light levels. Scenario B used an IPR table averaged for all light levels. Sensitivity to the ratio limits, ranging from 25 to 500%, was also investigated.

Samples were initially sorted by month, and were separately prepared as input matrices into CHEMTAX. IPR tables were constructed for each month using values from the literature that corresponded closest to the phytoplankton species found in the microscopic samples (Table 2.1). For each month, 4 different IPR tables were constructed – one based on average light levels (AVG) and three for different light levels (as a function of photosynthetically available radiation – henceforth PAR). The high light table (HL) corresponded to PAR levels exceeding 300 $\mu$mol photons $m^{-2}$ $s^{-1}$; a medium light table (ML) to values between 100 and 300 $\mu$mol photons $m^{-2}$ $s^{-1}$; a low light table (LL) corresponded to PAR levels below 100 $\mu$mol photons $m^{-2}$ $s^{-1}$. Values in the IPR tables were derived from two sources which listed pigment ratios for different phytoplankton species at three different light levels – Schluter et al. (2000) and Henriksen et al. (2002). The boundaries between LL, ML, and HL tables were based the light levels presented in these studies.

In scenario A, samples were segregated by light level and run through CHEMTAX using IPR tables that were matched to the measured field PAR levels adjusted to depth. In scenario B, all monthly samples were run through CHEMTAX as a group using the AVG table.

In addition, each scenario was run at 7 different ratio limits. These were 25, 50, 100, 200, 300, 400 and 500 which correspond to the maximum percentage change by which the IPR table was allowed to adjust itself during the iterative fitting. The schematic for this experimental design is shown in Figure 2-3.

### 2.2.6 PAR calculation

Surface PAR measurements were recorded on station using a Satlantic OCR-ES Radiometer at intervals of every second, and the PAR attenuation coefficient in the

**IPR Table: Initial ratios of pigments to chlorophyll *a***

| *Class* | *PAR* | *Peri* | *19'but* | *Fuco* | *19'hex* | *Prasin* | *Allo* | *Zea* | *ChlB* | *Viola* |
|---|---|---|---|---|---|---|---|---|---|---|
| | H | 0 | 0 | 0.485 | 0 | 0 | 0 | 0 | 0 | 0 |
| Diatoms[1] | M | 0 | 0 | 0.585 | 0 | 0 | 0 | 0 | 0 | 0 |
| | L | 0 | 0 | 0.485 | 0 | 0 | 0 | 0 | 0 | 0 |
| Dino- | H | 0.343 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| flagellates[2] | M | 0.636 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | L | 0.711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cyano- | H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.939 | 0 |
| bacteria[2] | M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.673 | 0 |
| | L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.457 | 0 |
| Prymnes- | H | 0 | 0 | 0.073 | 0.685 | 0 | 0 | 0 | 0 | 0 |
| iophyte[1] | M | 0 | 0 | 0.089 | 0.735 | 0 | 0 | 0 | 0 | 0 |
| (*E.hux.*) | L | 0 | 0 | 0.121 | 0.811 | 0 | 0 | 0 | 0 | 0 |
| Chloro- | H | 0 | 0 | 0 | 0 | 0 | 0 | 0.100 | 0.188 | 0.011 |
| phytes[1] | M | 0 | 0 | 0 | 0 | 0 | 0 | 0.060 | 0.178 | 0.031 |
| | L | 0 | 0 | 0 | 0 | 0 | 0 | 0.01 | 0.187 | 0.044 |
| Prasino- | H | 0 | 0 | 0 | 0 | 0.317 | 0 | 0.087 | 0.527 | 0.035 |
| phytes[2] | M | 0 | 0 | 0 | 0 | 0.346 | 0 | 0.054 | 0.663 | 0.080 |
| | L | 0 | 0 | 0 | 0 | 0.322 | 0 | 0.011 | 0.790 | 0.072 |
| Crypto- | H | 0 | 0 | 0 | 0 | 0 | 0.339 | 0 | 0 | 0 |
| phytes[1] | M | 0 | 0 | 0 | 0 | 0 | 0.271 | 0 | 0 | 0 |
| | L | 0 | 0 | 0 | 0 | 0 | 0.172 | 0 | 0 | 0 |
| Chryso- | H | 0 | 0.54 | 0.337 | 0 | 0 | 0 | 0 | 0 | 0 |
| phytes[2] | M | 0 | 0.93 | 0.620 | 0 | 0 | 0 | 0 | 0 | 0 |
| | L | 0 | 1.56 | 0.974 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 2.1: Initial pigment ratios for the 3 different PAR levels: H (high light), M (medium light), and L (low light). The PAR-specific values were used for scenario A, while average pigment ratios were used for scenario B. Values were obtained from: [1]Schluter et al. (2000) and [2]Henriksen et al. (2002).

water column was measured using a HyperPro radiometer profiler (Satlantic, Inc.). The surface PAR measurements were averaged for the station duration (typically 0.5 - 1 hour), and used to obtain baseline surface PAR values. These were then propagated to the sample depths to get depth-resolved PAR values using the diffuse attenuation coefficient for PAR derived from the HyperPro.

## 2.3 Results

### 2.3.1 Comparison of CHEMTAX and microscope counts

The proportions of the different algal groups resulting from the microscopic-based carbon estimates compared to the CHEMTAX estimates showed a range of quantitative agreement for both scenarios (Figure 2-4, Table 2.2). The highest correlation was using the PAR-matched IPR tables in scenario A at a ratio limit = 100. For this scenario and ratio limit, the highest correlations were for diatoms with $r^2 = 0.82$ and the lowest correlations were for cyanobacteria with $r^2 = 0.35$. The largest differences between scenarios were seen in the dinoflagellates, with $r^2 = 0.56$ to 0.18 for scenarios A and B, respectively. These values are comparable to correlation values reported by others (Schluter et al., 2000; Garibotti et al., 2003; Llewellyn et al., 2005).

| *Algal Group* | *A* | *B* |
|---|---|---|
| Diatoms | 0.82 | 0.65 |
| Dinoflagellates | 0.56 | 0.18 |
| Cryptophytes | 0.48 | 0.36 |
| Small Flagellates | 0.49 | 0.34 |
| Cyanobacteria | 0.35 | 0.40 |
| Average | 0.54 | 0.39 |

Table 2.2: $r^2$ values between CHEMTAX scenarios and cell counts (converted to carbon) for a ratio limit=100.

Overall, diatoms had the highest correlations, but high $r^2$ is not a sufficient condition for agreement. The large non-zero intercept indicates that diatoms were attributed 30-40% of the biomass using CHEMTAX when no diatoms were found

in the microscopic samples. This may be a result of the presence of fucoxanthin-containing prymnesiophytes. The marker pigment for diatoms is fucoxanthin, which is also found in prymnesiophytes. There were prymnesiophytes identified in the microscopic samples during summer and fall months when few diatoms were seen. The pigment ratios in the IPR tables for prymnesiophytes for these months were based on *Emiliania huxleyi*, but the prymnesiophyte species identified in the summer and fall samples were *Calyptrolithina spp.* and *Calyptrosphaera spp.* Schluter et al. (2000) reported significant sensitivity with prymnesiophyte IPR values in CHEMTAX.

The relationship between carbon and CHEMTAX-derived fractions for dinoflagellates is less clear. Other studies report a variety of problems relating dinoflagellate fractions to those obtained with CHEMTAX (Llewellyn et al., 2005). The problems are several. Dinoflagellates can be heterotrophic, and heterotrophic dinoflagellates are not readily distinguishable from autotrophic dinoflagellates in a microscope, whereas HPLC is measuring only autotrophic dinoflagellates (Llewellyn et al., 2005). Another complication is the ratio of carbon to chlorophyll *a*, which is generally higher for dinoflagellates than diatoms owing to their celluloid cell exteriors. Both of these factors could explain points below the 1:1 line.

The biomass fraction estimates for the small flagellates in CHEMTAX (prasinophytes, chlorophytes, prymnesiophytes, and chrysophytes) follow the general trend with carbon-based estimates. CHEMTAX underestimated the fraction in March, May and July compared to carbon estimates. This coincided with the higher proportions of biomass allocated to diatoms by CHEMTAX for these months. Cryptophytes were attributed higher biomass fractions by CHEMTAX than carbon for June

and July samples. Cyanobacteria were in general underestimated by CHEMTAX compared to carbon estimates. This is a similar result as reported in Havskum et al. (2004). Zeaxanthin is the marker pigment for cyanobacteria, but this pigment is also shared by prasinophytes and chlorophytes, and all three had zeaxanthin entries in the IPR table. Schluter et al. (2000) and Henriksen et al. (2002) report different zeaxanthin:chla ratios for *Synnechococcus spp.*, with Schluter et al. (2000) reporting values around 50 percent higher for similar PAR levels. It is unknown which ratio set is representative of the Gulf of Maine species. Pigment ratios determined from culture experiments with Gulf of Maine strains would have been preferable, but were lacking for this analysis.

### 2.3.2 Effects of variable IPR tables on CHEMTAX

The choice of the IPR tables between the two scenarios can be seen in Figure 2-5, which compares the class fractions for the 2 scenarios with the ratio limit=100. The average absolute differences between scenario A and B for all 8 classes are shown in Table 2.3, along with the range and the relative differences. The greatest dispersion is seen in the dinoflagellates which had the highest relative difference of 18%. The smallest differences were seen in the chrysophytes with a relative difference of less than 1%. The average relative difference was 9%. This is the percent by which CHEMTAX output differs when using an average IPR table versus PAR-specific IPR table(s).

These results demonstrate the sensitivity of CHEMTAX to the initial pigment ratios. Overall, the average differences were surprisingly small given the range of

| *Algal Group* | *Average Absolute Difference* | *Range* | Relative Difference (%) | *Average Standard Deviation* |
|---|---|---|---|---|
| Diatoms | 0.052 | 0-0.85 | 7.0 | 0.051 |
| Dinoflagellates | 0.064 | 0-0.35 | 18.3 | 0.024 |
| Cryptophytes | 0.021 | 0-0.30 | 7.0 | 0.013 |
| Prasinophytes | 0.036 | 0-0.45 | 8.0 | 0.029 |
| Prymnesiophytes | 0.051 | 0-0.41 | 12.4 | 0.040 |
| Chlorophytes | 0.023 | 0-0.22 | 10.4 | 0.024 |
| Chrysophytes | 0.001 | 0-0.2 | 0.5 | 0.009 |
| Cyanobacteria | 0.008 | 0-0.11 | 8.0 | 0.005 |
| Average | 0.032 | | 9.0 | 0.025 |

Table 2.3: Average absolute differences between class fractions for scenario A and B at a ratio limit=100. Range and absolute differences are expressed as fractions of 1. The relative difference was computed from the average absolute difference divided by the range. The last column was derived by calculating the average of the standard deviation for each class over the range of ratio limits for Scenario A.

the IPR values for each scenario (Table 2.1), which generally vary by a factor of 2 or more, and in the case of zeaxanthin by a factor of 10. In fact, magnitudes of the relative differences were unrelated to the algal classes' IPR ranges. For example, the IPR values for dinoflagellates varied by a factor of 2 and had the highest relative difference (18%), whereas cyanobacteria had a similar range in IPR values, yet had a relative difference of only 8%.

### 2.3.3 Effects of varying ratio limits on CHEMTAX

The effect on the outcome of CHEMTAX from varying the ratio limits is illustrated in Figure 2-6, which shows the diatom class outputs for scenario A for the selected ratio limits. The average standard deviation over the range of ratio limits for scenario A was 0.051 for diatoms. Results for other groups were better and are shown as the last column in Table 2.3. The results for scenario B were similar (not

shown).

The overall effect of the ratio limits is reflected in the final pigment ratios, which are the values to which the pigment ratios were adjusted at the completion of a CHEMTAX run. As the ratio limits were increased, the IPR tables were allowed to change over larger ranges. The average fractional differences between starting IPR tables and final pigment ratios for both scenarios across the 7 ratio limits are summarized in Table 2.4.

| | *Scenario* | |
|---|---|---|
| *Ratio Limit* | *A* | *B* |
| 25 | 24.4 | 22.0 |
| 50 | 29.2 | 24.6 |
| 100 | 34.5 | 27.3 |
| 200 | 42.3 | 36.6 |
| 300 | 47.6 | 42.0 |
| 400 | 51.6 | 47.4 |
| 500 | 54.6 | 51.6 |

Table 2.4: Final pigment ratio average change (%).

The final ratio values for scenario A changed slightly more than scenarios B for all ratio limits. As expected, pigment ratios changed by larger amounts as the ratio limit was raised. However, the average change to IPR values approach the maximum (i.e., the ratio limit) only at a ratio limit = 25. At higher ratio limits, the average change in the IPR tables did not reach their respective limits. Scenario B contained an input matrix of phytoplankton samples at different photo-adaptive states (i.e., the physiological adjustments to different light levels), whereas scenario A segregated data into PAR-based matrices, and used PAR-matched IPR tables. Having an input matrix composed of samples from mixed conditions may limit the

degree to which the IPR table can change inside the fitting procedure of CHEMTAX because adjustments to IPR values are being forced onto samples that may not be in the same light-adapted state. In this case, the input matrix would contain different pigment ratio characteristics for different samples. Thus, the composition of the input pigment matrix, in terms of light adapted states, has an apparent effect on the performance of CHEMTAX.

## 2.4 Discussion

### 2.4.1 CHEMTAX performance and assumptions

The different algal groups showed variable agreement between microscopic carbon and the CHEMTAX-based assessment of phytoplankton composition under the different scenarios. Scenario A had the highest $r^2$ values, which ranged from 0.82 for diatoms to 0.35 for cyanobacteria. These $r^2$ values are in the same range as those reported for similar studies (e.g., Llewellyn et al. (2005); Schluter et al. (2000); Garibotti et al. (2003)). The same $r^2$ range has been described as 'good' by some investigators and 'poor' by others.

There is uncertainty inherent in both methods of community assessment due to a variety of factors. However, the general performance of CHEMTAX is difficult to evaluate, since the comparisons inevitably rely on carbon-based estimates from microscopy. Ideally, one would want to compare calculations based on the same unit (i.e., chlorophyll *a*), but this is not practical with microscopic samples as one would need to know intracellular chlorophyll concentration for each cell and

species. Therefore, most correlations in the literature compare carbon fractions to CHEMTAX chlorophyll-*a* fractions.

The nature of the comparisons – one an estimate of the biomass fraction of chlorophyll *a* and the other an estimate of the carbon-based fraction – are related through the carbon-to-chlorophyll ratio (*C:Chl*). *C:Chl* in bulk water samples ranges from 10 to over 200, and depends on the physiological state of a given phytoplankton community. It also varies among species and algal groups (Geider, 1987). Diatoms, for example, tend to have *C:Chl* in the range from 10-50 (Cloern et al., 1995), while dinoflagellate and cyanobacteria *C:Chl* values can be much higher (Geider et al., 1998). These differences are partly due to their compositional makeup and partly due to their differential responses to light, temperature, and nutrient conditions.

In order for there to be perfect agreement between the cell counts and CHEMTAX results, species-specific *C:Chl* must be identical to the community *C:Chl*, but this is generally not the case. A plot of total carbon versus chlorophyll *a* for this data set is shown in Figure 2-7. Although *C:Chl* values are within reported ranges, overall they are quite variable. The average *C:Chl* calculated for the entire phytoplankton population for the data shown in Figure 2-7 was 52, which is consistent with the traditional global mean of 50 (Eppley, 1972). When the class specific *C:Chl* is higher than the community *C:Chl*, points will be pushed to right of the 1:1 line in Figure 2-4. This situation would most likely occur for dinoflagellates and cyanobacteria in a community dominated by diatoms, since diatoms have lower *C:Chl* than either group. Conversely, points will be pushed to the left of the 1:1 line for a class whose *C:Chl* is lower than the community *C:Chl*. This situation is likely to

occur in a phytoplankton community dominated by cells that have a higher *C:Chl* than diatoms. The plots in Figure 2-4 do reveal these basic trends. Diatoms are generally to the left of the 1:1 line at low to middle levels of biomass fraction, and converge on the 1:1 line at high fractions; dinoflagellate and cyanobacteria points are to the right of the 1:1 line. However, as described below, there are other uncertainties found in the various conversions and assumptions, which are cumulative and therefore potentially very high.

The main value of using microscopy in conjunction with CHEMTAX is to know which species should be used to populate the IPR tables, and to approximate taxonomic composition with comparisons. For example, it can provide supporting evidence for the presence or absence of an algal class, such as the absence of diatoms in summer samples found in this research. This can be used as a simple pass or fail check on CHEMTAX results, but a more quantitative performance evaluation is limited, if not impossible.

### 2.4.2 Error sources in microscopy

There are a number of inherent error sources associated with microscopic counting, as well as the conversion from cell counts to biomass and from biomass to carbon totals. The carbon estimates for the different algal groups combined light microscopic enumeration with flow cytometric techniques. The light microscopic counting was restricted to cells $> 10$ $\mu$m. Cells of smaller diameter are simply too small to identify under the magnification of the light microscope (400X for the Leica). Other techniques are needed to identify and quantify smaller cells. One method is based

on epifluorescent techniques which require filtration onto a membrane and subsequent illumination with directed light beams. Based on the color of the fluorescent emission, cells can be assigned to different algal groups. Flow cytometry offers an alternative method (Yentsch and Campbell, 1991). This technique is able to count and determine size characteristics of cells smaller the 10 $\mu$m, including picoplankton $< 1$ $\mu$m. In addition, it performs very well in distinguishing cyanobacteria and cryptophytes from other groups because of the presence of phycobiliproteins which have distinctive fluorescence characteristics. Both of these methods have been used in previous phytoplankton community assessment studies (Duarte et al., 2000; Schluter et al., 2000; Gin and Lee, 2003).

Flow cytometry can differentiate living cells from non-living particles and can assign the cells to different size ranges. However, flow cytometry cannot differentiate algal groups other than cyanobacteria and cryptophytes. 'Other' living cells were all assumed to be flagellates representing prymnesiophytes, chlorophytes, chrysophytes and/or prasinophytes. The HPLC pigment data confirm that one or more of these groups were present in most samples. The carbon-volume relationships for these flagellate groups are similar, although this is from limited published data (Menden-Deuer and Lessard, 2000). The carbon estimates for this size fraction would not have changed if this group had been segregated into separate algal classes, assuming there were no diatoms or dinoflagellates in this size range. There are a few pennate diatoms that approach this size, but none were seen or identified in the light microscope.

There can be large errors associated with the counting of phytoplankton in the

light microscope. Reports of counts on the same sample show a wide range of variability, reaching up to a factor of 10 difference (Schluter et al., 2000). On several samples in this study, replicate counts were made by different individuals and compared. The differences were not as large as previously reported, but did exhibit an average difference of 30% for species counts. Fixatives can distort original cell diameter by shrinkage or expansion, but no adjustments were made to cell sizes in this study.

There are a number of relationships for converting biovolume to carbon units for different algal groups, with little consensus among them. Garibotti et al. (2003) found a 3-fold difference between diatom carbon estimates (at intermediate to high biomass regions) using the Strathmann (1967) and the Montagnes and Franklin (2001) equations, whereas Llewellyn et al. (2005) found a factor of 2 between the same equations. The differences are related to how vacuoles are treated in the cells and in the carbon quotas. Other factors which can effect the carbon content are the nutritional state and the light level. Cells in general will begin to lower overall carbon content in high light environments (Cullen and Lewis, 1988), or when nutritionally deplete (Menden-Deuer and Lessard, 2000). In this study, the relationships given in Menden-Deuer and Lessard (2000) were used as they represented diverse collections of algae across physiologic states.

### 2.4.3 Error sources in CHEMTAX

CHEMTAX requires the initialization of two user-defined tables prior to program execution, and these are the IPR and ratio limits tables. Both tables have an

influence on the outcome of CHEMTAX, and each was systematically varied in order to assess the effects on the output matrix. Two scenarios were set up for CHEMTAX using different implementation schemes for IPR tables. The first – scenario A – initialized IPR tables with values that were matched to the light level of the sample matrix. Scenario B initialized the IPR table with an average pigment ratio and was used for all samples, regardless of light level. In addition, each scenario was executed at different ratio limits.

In the case of scenario A, three PAR regimes were defined and IPR tables were constructed for each regime. The two prime sources for the IPR values were Henriksen et al. (2002) and Schluter et al. (2000). These two studies used similar PAR ranges for their culture experiments, and thus set the boundaries between light regimes for scenario A. The choice of which IPR value to use for a given class was based on a match as close as possible to the species observed in the water samples. However, the IPR values used in this analysis may be based on species that are either not found or not representative of the algae that inhabit the Gulf of Maine waters. For example, the dinoflagellates *Ceratium longipes* and *Ceratium fucus* were observed in the samples from June through October. Yet, there were no pigment ratios for either of these species found in the literature. The dinoflagellate IPR values were based on culture experiments with *Scrippsiella sp.* (Henriksen et al., 2002), which were observed in the Gulf of Maine samples but in fewer concentrations than other dinoflagellates. The peridinin:chlorophyll *a* values for two different dinoflagellates in Henriksen et al. (2002) varied from 0.375 to 0.746 (a factor of 2) at the same light level. Thus, choosing which pigment ratio to use can be difficult when multiple species for a given class are present in the sample, as was often revealed

through the microscopic analysis.

In some cases, one species for a given class would only be present during restricted times of year. For example, this problem occurred during summer months when diatoms were reported in the CHEMTAX results, yet no diatoms were identified in the microscopic samples. This is a result of assigning fucoxanthin to diatoms, when the fucoxanthin in these samples was most likely derived from prymnesiophytes, which were identified in the microscopic samples (e.g., *Calyptrolithina* spp.). The IPR tables were populated with values based on *Emiliania huxleyi* which was not observed in the water samples, and may have different pigment ratio characteristics than species that were identified in the samples (e.g., *Calyptrolithina* spp.). The consequences of this are not known, since values for *Calyptrolithina* sp. were not found in the literature.

There is similar uncertainty with the cyanobacteria class. There are reports of two strains of *Synechococcus spp.* in studies of the north Atlantic and the Pacific Oceans (Olson et al., 1990) – a bright light strain and a dim light strain. The PAR-dependent ratios of zeaxanthin:chla vary by 50% for *Synechococcus spp.*, and it is not known which ratio set is more applicable to the Gulf of Maine. This highlights pigment variability within genus and species, and the implications for CHEMTAX are that it is imperative to know the pigment characteristics of all phytoplankton types in a given area.

The ratio limits can also affect the outcome of CHEMTAX. The purpose of the ratio limit is to constrain the degree to which CHEMTAX can adjust ratio values during the iterative fitting process. The ratios can be tightly constrained (i.e., small

ratio limits) or relaxed (i.e., large ratio limits). The final pigment ratios will always vary from the original IPR values, and for this study the average differences between initial and final pigment ratios are summarized in Table 2.4. As ratio limits were increased, the IPR values changed by larger amounts. However, only at a ratio limit of 25 (the lowest) did the IPR values actually reach the ratio limit. Scenario A, which was based on PAR-grouped samples, had higher overall changes than scenario B. Considering scenario A, the phytoplankton were assumed to be in a similar light-acclimated state. If the overall species assemblage in a given class had similar pigment composition and ratios which matched the IPR values, it would be expected that the final ratio values would not change significantly from the initial values. If the IPR values did not represent the true class pigment ratios, it would be expected that they would adjust during the iterations until a more representative value were reached. This could be a large change if the true ratios and IPR used in the run were far apart. This is a possibility, since pigment ratios for species found in the Gulf of Maine were not available from literature. Scenario B used input matrices with samples from a wide range of PAR levels, the IPR tables were thus fitted to samples that have diverse actual pigment ratios. Since CHEMTAX fit the IPR to the whole set, any gain in changing the IPR tables to values representative of one PAR regime would be lost on samples from another PAR regime. This is perhaps why the IPR tables changed less for scenario B. Once the IPR values were changed to a certain point, further change did not improve the residuals between iterations. In other words, samples (from the same algal class) in the same matrix conditioned to different light levels may not share the pigment ratios, and improvements in some samples during the iteration could cause other samples to increase their residual

error. This point was achieved more quickly in scenario B than A, and the mixed sample matrix may be the reason.

Other factors not included in this study for affecting IPR values are nutrient conditions and growth phase. These are both important to pigment ratios as shown by Henriksen et al. (2002) and Schluter et al. (2000). Ratios can vary by as much as a factor of 4 for cells in stationary versus exponential growth (Henriksen et al., 2002); such states were not known for the samples taken. The IPR values taken from Henriksen et al. (2002) were for the exponential growth phase. During at least one cruise (June), diatoms were suspected of being in stationary growth phase. Samples taken on June 17 showed extremely high levels of diatoms (counts over 100,000 cells/liter). A cruise one week later to the same stations did not detect any diatoms from water samples. The June 17 samples were likely at the end of the diatom bloom and subsequently disappeared as a result of nutrient exhaustion (although it is unclear whether it was related to silicate or nitrate). Since IPR values change during CHEMTAX iterations, it is possible that the ratios self-adjusted to stationary phase values for scenario A.

### 2.4.4 Time scales of photoacclimation – consequences for IPR

The rationale for using PAR-dependent IPR tables to CHEMTAX is valid only if the phytoplankton are photoacclimated to the PAR used in the tables. Photoacclimation is the process of modifying the photosynthetic apparatus to the external light conditions. This process includes modifications to both the intracellular pigment concentration and composition. Phytoplankton will experience fluctuations

in light intensity if there is active vertical mixing. The extent to which the light level changes depends on the intensity of the mixing rate and the depth of the mixed layer. If cells are rapidly circulating through different light conditions, the rates of the photoacclimative processes may not be fast enough to adapt to the depth-dependent light level.

There is considerable variability in the response rates of photoacclimation for different species under different conditions (e.g., Staehr et al. (2002)). Cullen and Lewis (1988) observed a faster response in photosynthetic parameters in phytoplankton cultures when going from high to low light, compared to the response from low to high light. Phytoplankton respond within minutes to changes in the light environment, as Oliver et al. (2003) have demonstrated. Variable fluorescence, a measure of the photosynthetic efficiency, can quickly respond (minutes to hours) to light shifts in a vertically circulating environment. However, full acclimation is a longer process and may take hours to days (Geider et al., 1998).

An additional consideration is the ability of phytoplankton to regulate their vertical position by either buoyancy mechanisms or vertical swimming in the case of cells with flagella. For example, dinoflagellates have been known to be able to maintain position in the water column, and can achieve daily migration distances of 20 meters (McGillicuddy et al., 2003). Similarly, cyanobacteria are known to change vertical position through internal buoyancy, and at least some species of diatoms have this ability as well (Geider et al., 1998). The ability to move confounds the effects of vertical mixing by the water column, and the process of photoacclimation.

## 2.5 Conclusions

The CHEMTAX program quantifies the phytoplankton class composition based on the ratios of marker pigments to chlorophyll *a*. An important first step is the configuration of the initial pigment ratios for the expected algal classes. In practice, these are based on values obtained from cultured species representative of the various classes. However, these ratios are dependent on algal growth stage, light levels and quality, nutrients, and temperature. Pigment ratios can also vary considerably between species within a class. Previous recommendations for improving CHEMTAX results are to know the phytoplankton species of the given water mass, and the environmental conditions (e.g., light) beforehand. Initial pigment ratios that adhere to these recommendations should give the best results for CHEMTAX.

This study examined the sensitivity of CHEMTAX to different treatments of the initial pigment ratio table and the ratio limits table. The use of initial pigment ratios matched to the light level of the samples was compared to the use of pigment ratios averaged for all light conditions. The average relative difference between CHEMTAX results for the scenarios was 9%. The most affected algal class were the dinoflagellates, followed by the prymnesiophytes. The effects of different ratio limits were on the same order. Assuming these effects are additive and independent, combined differences could reach plus or minus 15-20%, dependent on the algal class.

It is not possible to evaluate different IPR tables or ratio limits by comparisons between CHEMTAX and microscopy. This is because the chemotaxonomic method and microscopic assessments inherently measure different aspects of community biomass. Assumptions about the conversion between one form and the

other can be easily violated. Thus, quantitative comparison between microscopy and CHEMTAX is limited. Microscopy will provide information on species present in a sample, which is important for basing IPR entries for a given algal class. At a minimum level, it provides a measure of the presence/absense for different algal classes in comparison with CHEMTAX results. This in itself was found to be valuable.

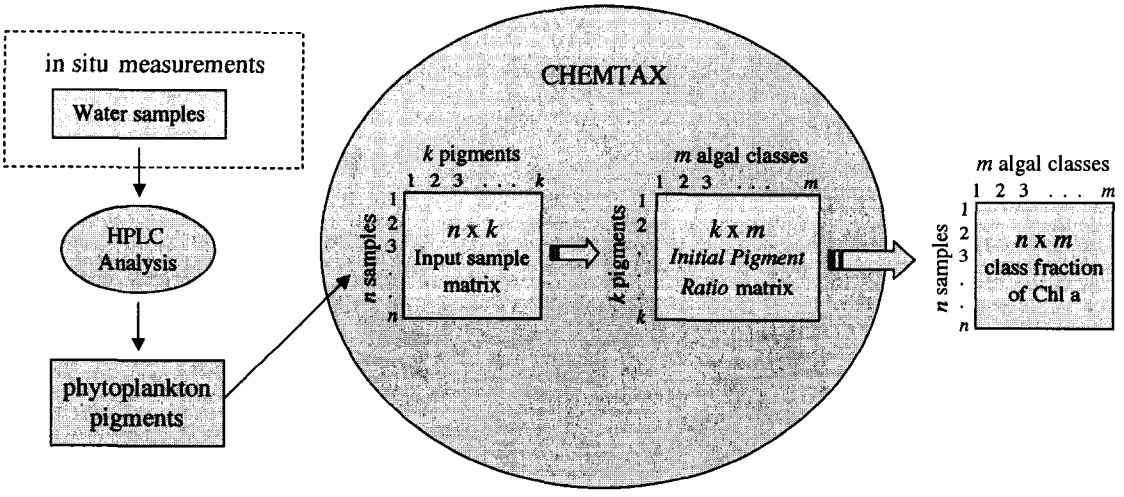



Figure 2-1: Flowchart of the CHEMTAX program to calculate phytoplankton community composition from HPLC data.

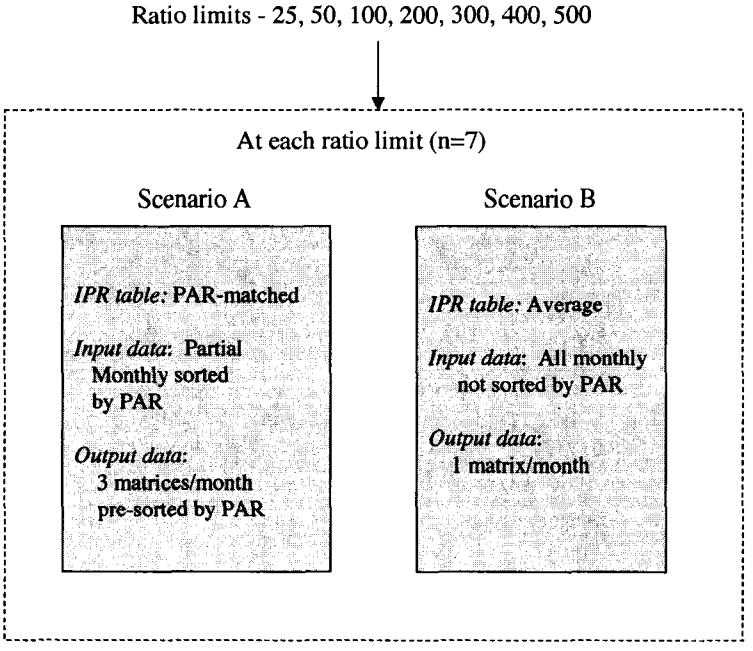



Figure 2-3: Schematic of the CHEMTAX experimental design. Scenario A was based on matching IPR tables to the sample PAR, while scenario B used an average IPR table. Each scenario was run with 7 different ratio limit values.

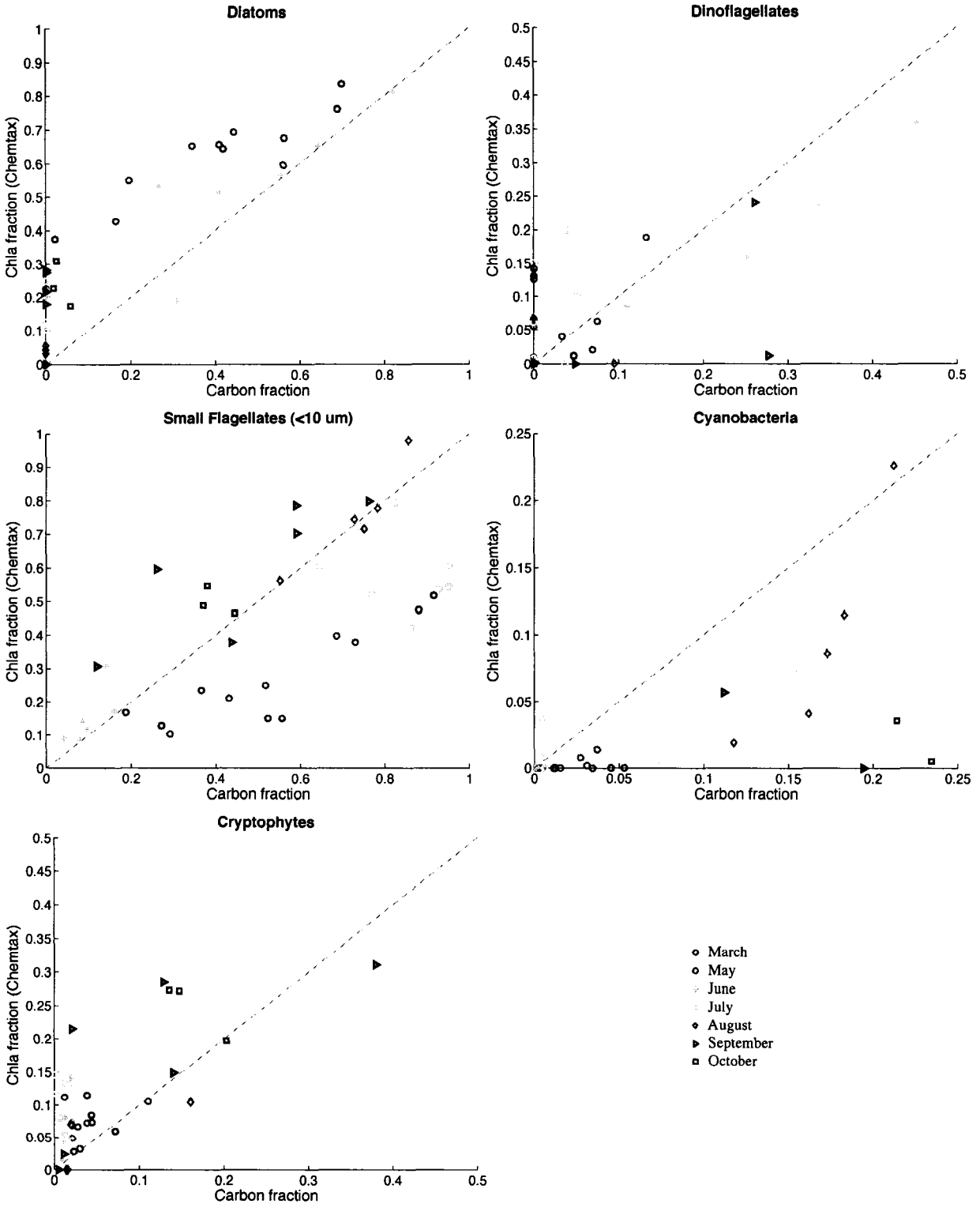



Figure 2-4: CHEMTAX class fractions versus carbon-estimated class fractions at a ratio limit = 100. Red dashed line is 1:1.

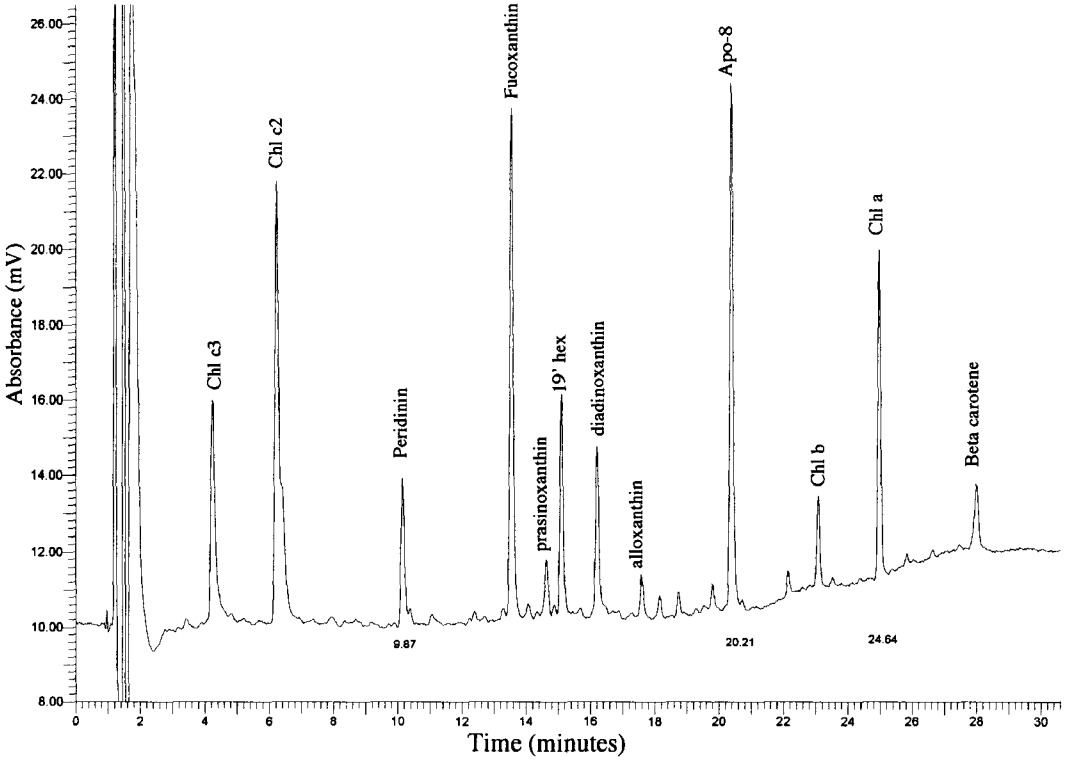



Figure 2-2: HPLC chromatogram from a June sample at 450 nm.

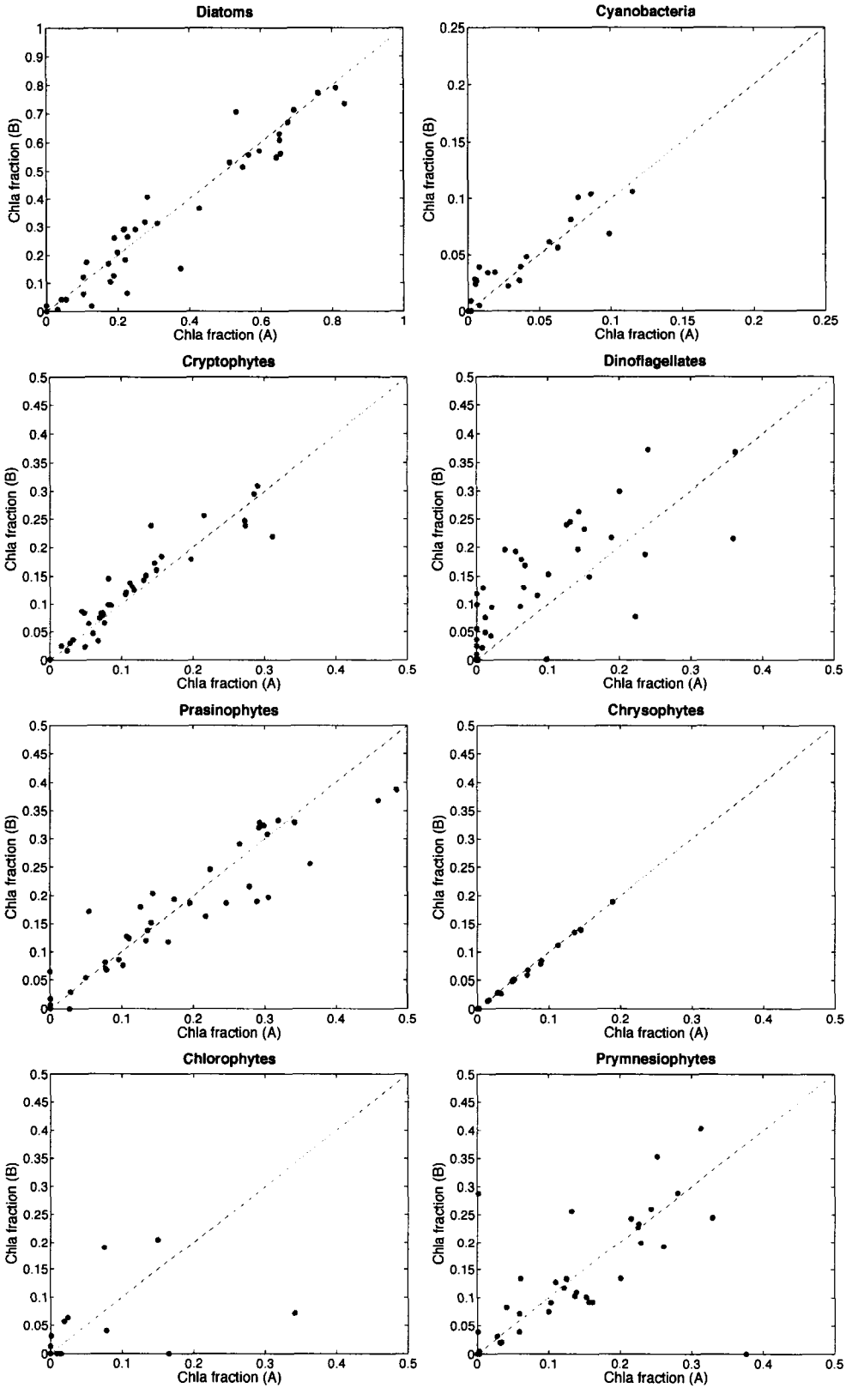



Figure 2-5: CHEMTAX class fractions for scenario A versus scenario B for the eight different algal classes (N=40). Red dashed line is 1:1.

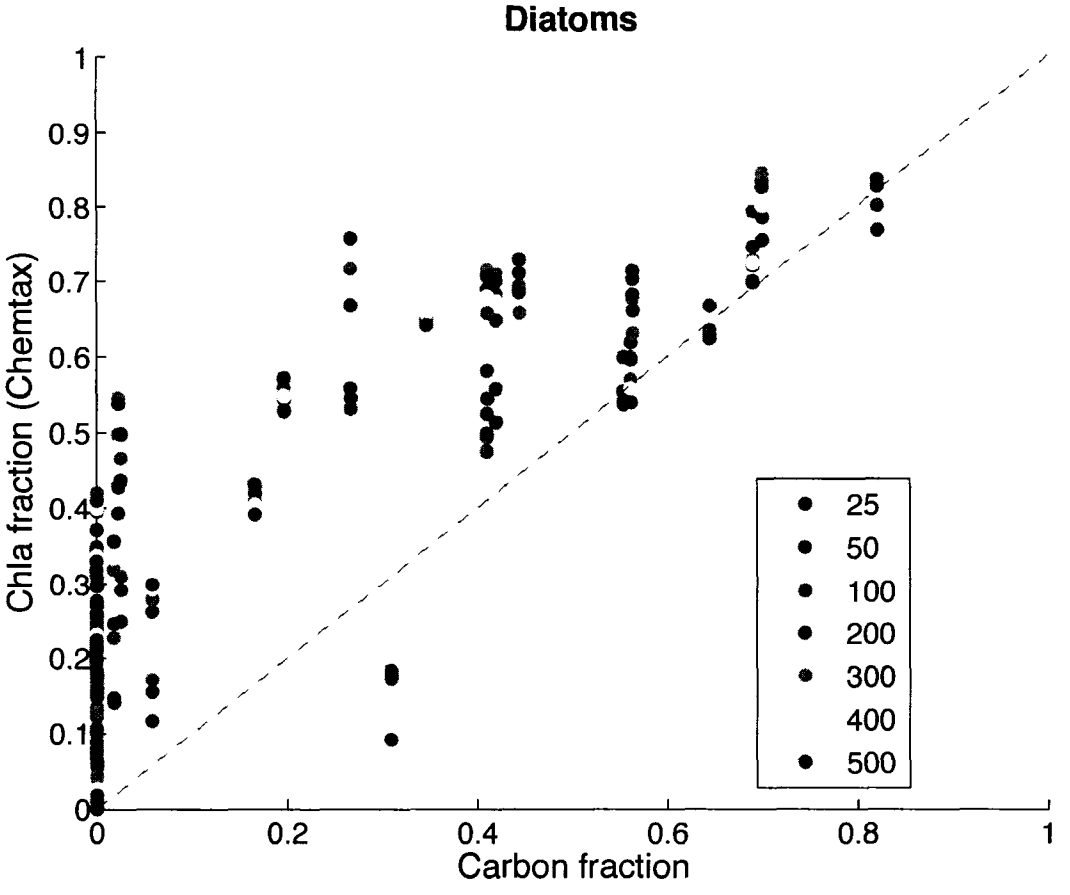



Figure 2-6: CHEMTAX vs. carbon-based diatom fractions using variable IPR tables (scenario A): ratio limits at 25, 50, 100, 200, 300, 400 and 500. The ratio limit values are the maximum percentage the IPR tables can change by in the CHEMTAX program.

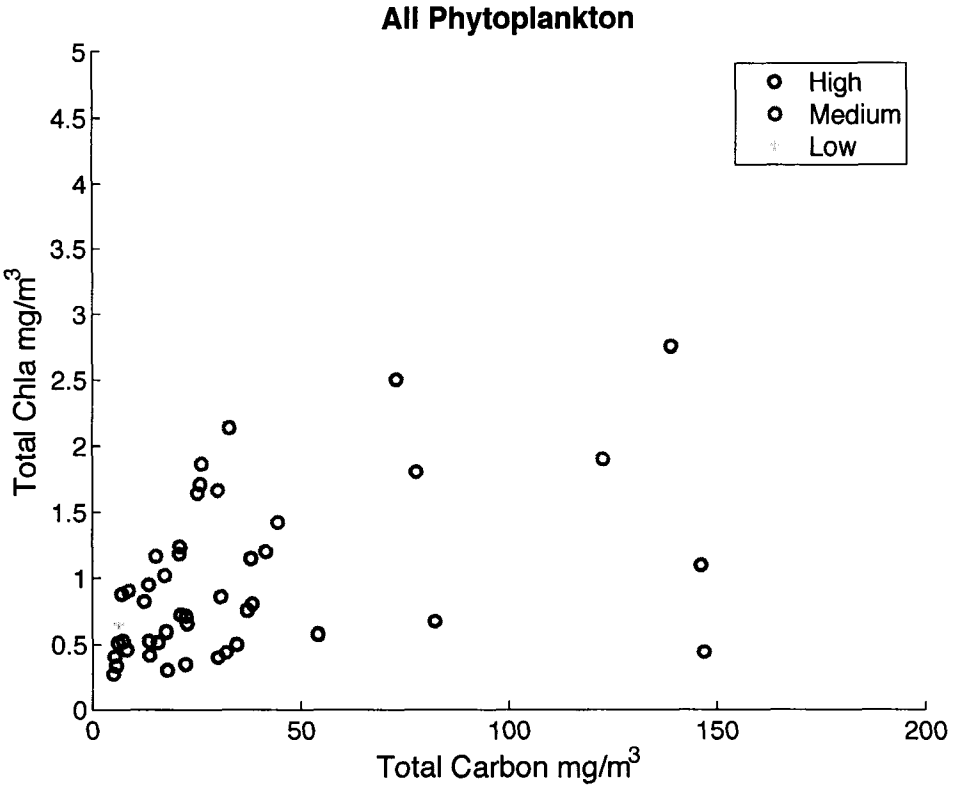



Figure 2-7: Total chlorophyll a vs. total phytoplankton carbon - points are color-coded by light level: green (low light), red (medium) and blue (high light)

# CHAPTER 3

# SEASONAL DYNAMICS OF PHYTOPLANKTON COMMUNITIES IN THE WESTERN GULF OF MAINE: LINKAGES BETWEEN PHYSICAL FORCING AND POPULATION COMPOSITION

*The following chapter has been submitted to the Marine Ecology Progress Series.*

**Abstract:** From June 2004 through November 2007, routine measurements of physical, chemical and biological variables were made along 2 transects every month in the western Gulf of Maine. The phytoplankton community was characterized from HPLC pigment data and exhibited seasonal variability in its composition. A sparse winter population of diatoms and flagellates was succeeded by a spring bloom dominated by diatoms, but also accompanied by increases in dinoflagellates, cryptophytes, prymnesiophytes, and other flagellates. A summer transition followed with the disappearance of diatoms and a remaining community dominated by flag-

ellates with minor contributions from cyanobacteria. A second bloom occurring in the fall was composed of a mixed community of diatoms, dinoflagellates, and other flagellates. The relationship between surface phytoplankton communities and environmental factors was examined using principal component analysis. The first 3 principal components accounted for more than 71% of the variance in a set of nine environmental variables. The first principal component (47% of the variance) was associated with the seasonal variation in temperature, light, wind speed, the ratio of the euphotic depth to the mixed layer depth, salinity, and nutrient concentrations. When data points were projected in the subspace defined by the first three principal components, similar phytoplankton communities tended to group together. The environmental factors were seen as key drivers for this pattern, and phytoplankton composition was consistent with habitat preferences as inferred from an adaptation of the data to the Reynolds Intaglio.

## 3.1 Introduction

The influence of environmental factors on phytoplankton cell physiology and ecology has been a focus of marine research for over 100 years. It is generally understood that the rise and fall of phytoplankton species within a community can be caused or triggered by a combination of factors (e.g., light, temperature, nutrients, grazing, competition) that are working simultaneously. These factors affect life stage evolution and overall community composition, and are constantly varying in the marine environment (Smayda, 1980). Results of past research efforts have led to the establishment of basic relationships between the environmental state and the composition of the phytoplankton community (Sommer, 1989; Goericke, 1998; Kruk et al., 2002) or the life stage of a particular species (McGillicuddy et al., 2003). These studies found positive correlations between the environmental state (as a combination of factors) and the phytoplankton community composition or species condition. However, knowledge of the specific mechanisms responsible for phytoplankton succession within marine systems is still incomplete and unresolved.

Cycles of community succession at seasonal scales are governed by changes in the physical forcing on the habitat suitability for both phytoplankton and grazers (Margelef, 1978; Smayda, 1980; Cullen and Lewis, 1988; Banse, 1994). Margalef's Mandala (Margelef, 1978) is a 2-dimensional plot that predicts the expected change in phytoplankton composition forced by nutrient availability and turbulence, which were considered to be the principal determinants shaping the population composition. In this depiction, phytoplankton composition moves along a trajectory from high nutrient, turbulent conditions favorable to diatoms to low nutrient, stable con-

ditions favorable to flagellates. This concept was further explored by Reynolds et al. (2000, 2002) and subsequently evolved into the Reynold's Intaglio – a template for describing phytoplankton communities and predicting where to find them along nutrient/light gradients.

Changes in environmental conditions occur on different time scales, from short-term (e.g., days) to longer-term (seasonal) to even much longer (decadal, millenial), and thus bring about change in phytoplankton populations in freshwater and marine environments along these same time scales (Grover and Chrzanowski, 2005; Barber and Hiscock, 2006; Karl et al., 1997). In both short- and long-term cases, the use of physical/environmental factors in describing phytoplankton communities is attractive because some of the key variables that influence the plankton community (e.g., temperature and light) can be remotely sensed from satellite platforms. This could provide the means to depict phytoplankton communities on the same spatial and temporal scales as the satellite data. We hypothesize that there is a predictive capability in this approach. Physical aspects of the environment (e.g., temperature, light, wind strength) can be measured, as can the phytoplankton structure through microscopic and pigment analysis. If correlations can be made between a set of physical variables and the composition of the phytoplankton community, then detection of the environmental conditions could be used to predict the phytoplankton community composition. This concept has been applied to freshwater (Grover and Chrzanowski, 2005; Kruk et al., 2002; Reynolds, 1984) and oceanic systems (Kamykowski and Zentara, 2003). These research approaches have used statistical methods (e.g., principal component analysis) to relate the physical variables to different species or algal groups. The work presented here extends this idea by

analysing the relationship between the physical environment and the phytoplankton community in the western Gulf of Maine. We also acknowledge that included in any relationship based on physical factors includes correlated biotic factors (e.g., grazing and competition). In this sense, the question we are seeking to answer is to what degree does the physical environment, along with correlated biotic process, influence the phytoplankton community composition?

The objectives of this study were to 1) identify and characterize the evolution of phytoplankton populations in the western Gulf of Maine over the study period, and 2) examine the associations between hydrographic forcing and phytoplankton communities.

## 3.2 Methods

### 3.2.1 Study Area

The study site is located in the western Gulf of Maine (Figure 3-1), and encompasses a region that stretches from the Merrimac River in the south to the Kennebec River to the north. The western boundary is fixed along the New England coastline, and the eastern edge is taken to be roughly 75 km offshore (at the longitude of the farthest offshore station). The measurements in this study were obtained from June 2004 through November 2007. A common set of measurements were made at each station, although the number of depths (for discrete bottle measurements) and sampling frequency varied among the stations. The set of core measurements included radiometric and CTD profiles, and discrete samples for HPLC, nutrients,

and absorption analysis. Nutrient samples were analyzed at UNH by the McDowell lab using a nutrient analyzer and included nitrate, nitrite, ammonium, silicate, and phosphate. A Wetlabs ACS instrument capable of measuring vertical profiles of particle backscattering, total absorption, total scattering, total attenuation, and fluorometric-excited emission was also routinely deployed.

The sampling program involved two transects that were visited every month – one that extended from Portsmouth, NH to Bath, ME (the Coastal Transect) and another that extended from Portsmouth, NH to an offshore location 70 kilometers away in Wilkinson Basin (the Wilkinson Basin Transect). Whole water samples were collected for microscopic and flow cytometric analysis from stations along these transects from March 2005 through March 2006 at surface and near-surface (10 meters depth) locations. These cruises were all conducted aboard the UNH vessel Gulf Challenger based in Portsmouth, NH.

Results presented in this chapter will focus on data from the Wilkinson Basin Transect, and will generally be restricted to surface values, unless specifically noted. The Wilkinson Basin Transect (henceforth, WB) – a cross-shelf transect covering a distance of approximately 75 kilometers – was repeated once a month, although weather or ship maintenance schedules sometimes prevented or interfered with a designated month's run. In all, a total of 38 out of 41 possible months had some WB data. [Weather occasionally prevented all stations from being visited during a cruise.]

The transect itself was composed of 6 stations, with the nearest inshore station (WB1) being a few kilometers offshore at a depth of 20 meters and the last station

(WB7) over 70 kilometers offshore (WB6 was discontinued early in the program and is omitted). Water depth initially increased with distance offshore from WB1 to WB4, and then shoaled as the transect crossed Jeffrey's Ledge where WB5 is at a depth of 60 meters. The farthest station from shore, WB7, is located in Wilkinson Basin, one of the deepest parts of the Gulf of Maine with a depth greater than 270 meters.

### 3.2.2 Environmental variables

Environmental variables were selected based on factors put forth by Longhurst (1995) for predicting the sequence of the seasonal phytoplankton growth cycle. The list assembled by Longhurst was intended for global application, but is also relevant on regional spatial scales such as the Gulf of Maine. The fundamental point of view adopted by Longhurst (1995) and others (e.g., Cullen and Lewis (1988); Smayda (1980)) is that growth is governed by the interactions between light, nutrients, mixing and stability of the upper water column. Based on the overlap of Longhurst's factors and the *in situ* data set, nine variables were selected. These were temperature, surface PAR, vertical attenuation coefficient, wind speed, nitrate concentration, phosphate concentration, dissolved silicate concentration, the ratio of the euphotic depth to the mixed layer depth, and salinity. The number of stations with all of the above measurements totaled 85. The software program Matlab (www.mathworks.com) was used to perform the data analysis.

### Temperature and salinity data

Surface values of temperature and salinity were extracted from station CTD profiles using a Seabird instrument. Occasionally, the profiler was not functional on some cruises and stations. In these instances, temperature and salinity values were taken from a flow-through system, which measured along-track surface values of temperature and salinity (and other variables). Mixed layer depth ($Z_m$) was calculated based on temperature and salinity profiles using a routine in Matlab (www.mathworks.com). The program searches a density profile over a moving depth window for a density change that exceeds a defined threshold. The default threshold for mixed layer depth calculation was set at a difference of 0.5 sigma units. When this threshold was not reached, it was apparent the entire water column was well mixed, and $Z_m$ was set to the water column depth.

### PAR data

Two data sets for photosynthetic active radiation (PAR) were available – the instantaneous PAR as measured onboard cruises, and daily average PAR fields from satellite data. The latter was chosen for the prime reason that it is the light field experienced over a period of days that is significant to phytoplankton community selection, and not instanteneous PAR which is more directly affecting shorter-term physiological mechanisms on the cellular level. Thus, derived PAR fields were extracted from SeaWiFS data. Daily SeaWiFS data during the study period were processed from level 1 to level 2 using SeaDAS, a software package specifically designed to process ocean color satellite data that was developed and distributed freely

by NASA (http://seadas.gsfc.nasa.gov). The output data produced during the level processing, in addition to water-leaving radiance and chlorophyll, included a PAR field based on the Frouin algorithm (Frouin et al., 2003). These daily PAR fields were then averaged for the preceeding 8 days of each cruise. *In situ* stations were co-located with the 8-day averages, and data were extracted from the imagery.

### Light attenuation coefficient

The vertical diffuse attenuation coefficient of downwelling light at 490 nm, $K_{d490}$, was obtained in a number of ways. It was calculated directly from downwelling light profiles measured with a Satlantic hyperspectral profiler, but this was not available for every station. In such cases, $K_{d490}$ was modeled using optical properties measured with a Wetlabs ACS profile system and the QAA model of Lee et al. (2002). When the two methods were compared, they agreed with an $r^2 = 0.71$.

### Winds and turbulent mixing

Wind data were downloaded off the GoMOOS website (www.gomoos.org) from fixed buoy measurements within and around the study area. Wind speeds were extracted from 20 different NOAA buoys and C-MAN towers, and GoMOOS buoys in and around the Gulf of Maine (Figure 3-1). The nearest buoy to the ship position was then selected as representative of the station wind field. Surface wind speeds (U) were converted to $U^3$ as a measure of wind mixing strength on the surface waters, and were then averaged over a period of 8 days preceding the date of the cruise. This is a proxy for the friction velocity associated with the turbulent mixing

energy (Archer, 1995).

### 3.2.3 Phytoplankton community composition

Phytoplankton pigment concentrations were measured from HPLC analysis, and phytoplankton class fractions were estimated using the CHEMTAX algorithm as described in the previous chapter. Pigment data were loaded into CHEMTAX, and the result was an output matrix which consisted of the percentage class contribution to *Chla* for eight phytoplankton classes. The sum of class percentages for each sample equaled 100%.

The eight classes were subsequently grouped into diatoms, flagellates, and cyanobacteria. Phytoplankton communities were formed from these three groups and described as *diatom-dominated* (70% or more of *Chla* attributed to diatoms), *flagellate-dominated* (70% or more of *Chla* attributed to flagellates), and *mixed* (all remaining combinations). This view of the phytoplankton community follows previous studies differentiating phytoplankton communities (Sathyendranath et al., 2004), and is comparable to the subdivisions of phytoplankton in marine ecosystem models (Moore et al., 2002).

### 3.2.4 Principal component analysis

Principal component analysis (PCA) is a statistical method used to reduce the dimensionality of multivariate data by linearly transforming them into a new data projection with minimal loss of information. The power of PCA is in data reduction by revealing the significant modes of variance within the data, and altering the axes

(the eigenvectors) to orient along these modes. The new orientation is such that the first principal component (PC) is a directed axis along the vector that accounts for most of the variance. The axes of the following PCs are orthogonal to one another, and account for sequentially decreasing amounts of variance. A $p$-dimension data set can have a maximum of $p$ principal components.

PCA was applied to the matrix of the nine environmental variables. Since these variables are of different units and range over different scales, all data were normalized to standardized variables by subtracting the mean from each value and dividing by the standard deviation. The PCA analysis supplied two important pieces of information for the present study. First, the main sources of variation within the physical system were characterized. Secondly, a qualitative assessment of the association of phytoplankton composition and physical factors could be made. This was done by plotting the environmental data points in their new PC orientation (whose distribution is strictly governed by their physical characteristics). These points were then color-coded by the phytoplankton community to which they belong. This revealed tendencies of phytoplankton communities to separate from one another and cluster into groups in the physical space.

## 3.3 Results

### 3.3.1 Hydrographic variability along the Wilkinson Basin transect

The general hydrographic conditions along the WB transect in the upper 10 meters showed strong seasonality in temperature, surface PAR, wind-driven turbu-

lence, and salinity (Figure 3-2). Along-transect variations from same-day cruises showed inshore-to-offshore gradients in these variables as well. The surface temperatures reached annual minima of less than 5C during February/March with the coldest waters occurring nearshore (WB1 and WB2) and getting progressively warmer heading offshore to WB7. This pattern was observed during the spring but occasional reverses occurred in the late spring/summer, when the warmest waters were found inshore and cooler waters offshore. Late summer maxima reached levels between 15C and 20C. The winter/spring station temperatures of 2006 and 2007 were warmer than those of 2005.

Salinity ranged from a low of 28.5 PSU at WB1 and WB2 in the spring to a high of near 33 PSU at WB7 during winter, with a strong seasonal component visible at all stations. The cross-transect gradient shows higher salinities offshore – away from freshwater sources – during all months of the year. This gradient was weakest during winter and strongest during spring. There were extreme gradients during May of each year when, as a result of heavy rains and river runoff, nearshore salinities were less than 29 PSU, and offshore salinities were greater than 32 PSU.

Satellite-derived daily PAR values exhibited a seasonal cycle, but minima/maxima were offset compared to the temperature cycle. The highest values occurred in late June ($>$ 50 mole photons $m^{-2}d^{-1}$) and lowest values occurred in December (10 mole photons $m^{-2}d^{-1}$), whereas temperature minima were in February and maxima in August. Extended periods of cloud cover in May 2006, July 2006, and June 2007 lowered PAR values compared to the same periods in other years. Values for the wind strength ($U^3$) were highest in winter and lowest in the summer, but the

seasonal patterns exhibited higher variability within seasons compared to the other variables.

Mixed layer depth ($Z_m$) showed a wide degree of variability between stations and seasons, but followed a common progression (Figure 3-3). Maximal $Z_m$ values occurred during fall and winter for all stations, with the deepest $Z_m$ at WB7, the station with the greatest water depth. Minimal $Z_m$ occurred during summer months, as well as at inshore stations in the spring. The factors governing $Z_m$ are different for each season. Wind and convective mixing deepen $Z_m$ in fall and winter, while freshwater runoff and/or upper layer warming due to increased solar radiation shoal $Z_m$ in spring/summer. The relative vertical positions of $Z_m$ and the euphotic depth ($Z_{eu}$) – the depth at which available light is 1% of the surface value – determine the amount of light experienced by cells in the mixed layer. This is indicated by their ratio, $Z_{eu}/Z_m$, in Figure 3-3. Values less than 1 indicate potential light limitation, while values greater than 1 are situations where phytoplankton cells are exposed to suitable light for growth throughout the mixed layer. The latter occurs in the summer, when surface PAR is greatest and $Z_m$ is minimal. During winter months, this ratio decreases and can reach values below one, particularly at deeper stations with WB7 having ratios as low as 0.2.

Surface nutrient concentration cycles followed inverse patterns compared to temperature (Figure 3-4). Dissolved inorganic nitrogen (DIN defined as the sum of nitrate plus nitrite) had its highest concentrations (5-10 $\mu$M) in the winter months and lowest concentrations ($< 1$ $\mu$M) during summer months of 2005 and 2006 (2007 data are incomplete for nutrients). Silicate (SiO2) exhibited this same pattern, with

highest concentrations reaching up 10 $\mu$M, and lowest concentrations in the summer less than 1 $\mu$M. Phosphate levels also followed the trends of SiO2 and DIN exhibiting an inverse relationship with temperature; annual highs occurred in the winter months and lows occurred in summer months. Cross-transect gradients were exhibited in all 3 nutrient species, with the largest spread in winter months and weakest in summer when nutrient concentrations were at their annual minima. Higher nutrient concentrations were typically found near the coast and declined offshore out to WB7, although this trend was at times reversed in the winter and spring months.

### 3.3.2 Near-surface total biomass and phytoplankton community cycles

Total chlorophyll *a* concentration (*Chla*) is the main photosynthetic pigment found in all phytoplankton and is generally taken as a proxy for the biomass of the whole phytoplankton community. The trends in surface chlorophyll at all WB stations followed a similar pattern – the highest values occurred during the spring and fall, and the lowest values during the summer and winter, and are displayed as the overall magnitude of each plot in Figure 3-5. The highest spring *Chla* values (>3 mg/$m^3$) were observed at WB1 and WB2 during 2006 and 2007. Although the spring bloom of 2005 was of smaller magnitude, *Chla* levels stayed elevated for a longer period compared to 2006 and 2007. The fall *Chla* values of 2006 and 2007 reached levels comparable to spring blooms of those years, and exceeded values of 3 mg/$m^3$ at WB3 and WB5 in 2006. There was a general decrease in surface *Chla* values from inshore to offshore in all seasons. WB7, the farthest station from shore, routinely had the lowest levels, whereas WB5 – situated on Jeffrey's Ledge – usually

had higher values than either WB4 or WB7.

The surface phytoplankton community off the coast of New Hampshire alternated between *diatom-dominated* communities in the winter and spring to *flagellate-dominated* communities in the summer and fall (Figure 3-5). [Note: The flagellate group in these plots has been split into two sub-groups – dinoflagellates and 'small flagellates' (comprising cryptophytes, prymnesiophytes, prasinophytes, chrysophytes, and chlorophytes). The dinoflagellates occupied a cell size range of greater than 20 microns, while the latter group occupied the size range from 2 to 20 microns.] A similar pattern generally occurred at all stations. Diatoms generally dominated the community during the spring bloom (when total biomass reached its highest levels), with a secondary bloom occurring in June of 2005 and 2007. Diatoms were generally present in low numbers or absent in the surface waters during the summer at all stations. Diatoms were also significant during brief periods in the fall (i.e., the fall bloom), and would occasionally dominate at inshore stations. However, the fall blooms were typically mixtures of flagellates and diatoms, or dominated by flagellates as in 2007.

Small flagellates were present year-round at all stations, whereas dinoflagellates were only present in spring, summer, and fall. The small flagellates were significant in the spring and would generally dominate the community after the spring bloom. The combined flagellates dominated summer months at all stations. Cyanobacteria represented a small but significant fraction of the community in the late summer, where they reached their highest levels. Although cyanobacteria never dominated the community in terms of biomass, they probably were numerically superior to

other phytoplankton in that season. The flagellates maintained their dominance through the fall. The fall blooms of 2004 and 2007 were dominated by flagellates and not diatoms at most stations, while the fall blooms of 2005 and 2006 varied in terms of dominance between these groups. There was a general decline in all phytoplankton groups during the winter, with the dinoflagellates having diminished to undetectable levels in each year.

### 3.3.3 Principal component analysis

The variables used in the PCA were temperature, PAR, $K_{d490}$, salinity, wind stress, nitrate concentration, phosphate concentration, silicate concentration, and the ratio of the euphotic depth to the mixed layer depth ($Z_{eu}/Z_m$). The first three principal components accounted for more than 71% of the variance (Table 3.1). The eigenvalues of the first 2 principal components exceed 1, indicating these are significant (Legendre and Legendre, 1998). The first principal component (*PC1*) is positively correlated with wind speed and nutrient concentrations, and negatively correlated with temperature, PAR and $Z_{eu}/Z_m$ (Figure 3-6). *PC2* is positively correlated with $K_{d490}$ and to a lesser degree salinity, whereas *PC3* is positively correlated with silicate. The variables were previously shown to have seasonal signals, and most are strongly associated with *PC1*. This indicates that all nine variables were significant, and exhibited an even distribution of sample points over seasons and conditions.

Figure 3-7 depicts the three phytoplankton communities (defined previously) relative to the physical domains in biplots of the first three principal component axes.

Table 3.1: PCA eigenvalues for a matrix of 9 environmental variables

**PCA eigenvalues**

| *PCA* | *Latent* | *Percent Variance* | *Cumulative Percent* |
|---|---|---|---|
| 1 | 4.2627 | 47.36 | 47.36 |
| 2 | 1.2071 | 13.41 | 60.77 |
| 3 | 0.9261 | 10.29 | 71.06 |
| 4 | 0.7162 | 7.96 | 79.02 |
| 5 | 0.6231 | 6.92 | 85.94 |
| 6 | 0.5397 | 6.00 | 91.94 |
| 7 | 0.3814 | 4.24 | 96.18 |
| 8 | 0.2292 | 2.55 | 98.73 |
| 9 | 0.1145 | 1.27 | 100.0 |

Biplots of eigenvectors with scores show that phytoplankton communities tend to organize into different regions of PC space. *Diatom-dominated* communities tended to occur on the right-hand side of each plot, and were associated with higher nutrient concentrations and wind speeds. In contrast, *flagellate-dominated* communities occurred more frequently on the left side of each plot, and were associated with warmer temperatures, higher PAR, and higher $Z_{eu}/Z_m$. *Mixed* communities are distributed on both sides of the axis. These plots show an affiliation of phytoplankton communities to different types of habitats, as defined by the nine environmental factors. It is important to note that these distributions also include correlated biotic effects, such as the effects from grazing and competition that correlate with temperature for example.

## 3.4 Discussion

### 3.4.1 Description of the phytoplankton community - caveats and considerations

Phytoplankton communities in coastal areas are a heterogeneous mix of species from different taxonomic classes. It is not unusual, and possibly frequent, to find a few species that numerically dominate the population mixed with a diverse set of less abundant species, especially in bloom conditions. In the western Gulf of Maine, blooms can develop at any time dependent on the local conditions. As Barber and Hiscock (2006) pointed out, when positive growth conditions occur, they can apply to many different species and an apparent bloom of a few species often obscures increased biomass in less conspicuous phytoplankton. Consequently, when dealing with the problem of using descriptive words to represent a mixed phytoplankton community, simply labeling a water mass as composed of 'diatoms' is often overlooking the contributions of other phytoplankton present in the population. Conversely, describing a community by listing every species present would not only be extremely challenging but would result in as many descriptions of communities as there are combinations of species – a task that would add an unnecessary level of complexity to an already burdensome task. This problem has been and is still under debate (see Reynolds et al. (2000) and Reynolds et al. (2002) and references therein). Since diatoms are a class of algae taxonomically different from other phytoplankton and unique in terms of their importance to food webs and marine biogeochemistry, it was essential to have these algae defined as their own group. The approach taken here to differentiate diatoms from all other phytoplankton is consistent with the

functional group view.

The issue of how to define the remaining phytoplankton was a problem of whether to differentiate among flagellates and picoplankton, and how to differentiate the flagellates and the picoplankton communities from diatoms. To satisfy these considerations, phytoplankton at the taxonomic level were viewed as belonging to one of three classes – diatoms, flagellates, or cyanobacteria – and at the community level were assigned to one of three possible groups based on the relative proportions of these different phytoplankton classes. The three groups – *diatom-dominated*, *mixed*, and *flagellate-dominated* – were in some sense arbitrarily defined. However, since the goal was to characterize phytoplankton *communities*, criteria had to exist to distinguish one group from another, and this inevitably involved subjective criteria. Given the uncertainties in defining community composition using pigment (or any other) methods, the distinction between a *diatom-dominated* community, a *mixed* community, and a *flagellate-dominated* community is not discrete, but instead is a continuous transition. While this sacrifice in complexity of group composition may overlook some important facets of phytoplankton ecology – that of species succession – it does serve to represent the general characteristics of phytoplankton at a community level.

The term 'dominant' in the naming convention does not necessarily imply numerical superiority nor high biomass. Wintertime phytoplankton biomass levels and cell numbers were relatively low compared with other time periods, and contained stations classified as diatom-dominated communities. These are relative terms, and signify the relative community composition, which in some seasons can be very low

in terms of overall phytoplankton numbers and biomass levels. This instantaneous composition also does not give information of the direction of community succession. A dominant phytoplankton species can decline within a few days to levels of insignificance to the population. As a consequence of the sampling frequency of a site (typically once per month), important stages in the development of a community (a continuous process) may be missed. Thus, the overall progression of community succession was broadly revealed at seasonal time scales, with the implicit foreknowledge that community change is continuous, and this change can occur rapidly at finer time scales than the sampling carried out here.

The vertical distribution was also not considered here. Subsurface phytoplankton communities were evident at different times of the year (not shown). This was most pronounced in the summer months, when a subsurface chlorophyll maximum was present, which could have a different composition than the overlying community. This aspect was not the focus of this study. The restriction to analysis of surface water was deliberate, but it is acknowledged that the vertical distribution of community composition is of consequence to phytoplankton community organization as it influences successional composition of the surface waters as a source of species which later can become dominant. This can occur through flagellate migration – which can span distances of 20 meters/day – or mixing events which can re-distribute vertically partitioned communities.

### 3.4.2 Phytoplankton community composition and distributions

Seasonal variation in phytoplankton community composition in temperate oceanic environments has been widely observed (Smayda, 1980; Longhurst, 2007). The phytoplankton populations in the western Gulf of Maine have shown both temporal (seasonal) and spatial (inshore vs. offshore) variability. In this area diatoms dominated the community in winter and spring samples, and flagellates dominated in summer and fall samples. This trend was observed to some degree at all WB stations, although a gradient was exhibited from inshore to offshore. The farthest offshore station (WB7) exhibited a much smaller diatom presence compared to the inshore stations. Flagellates dominated the community in summer and fall, and at offshore stations in winter and spring. Conversely, diatoms were occasionally dominant in the fall at inshore stations.

The general patterns observed here are consistent with previous studies from the Gulf of Maine and other nearby coastal regions. Gran (1932) and Lillick (1938) conducted Gulf-wide studies of phytoplankton distributions in the Gulf of Maine. Both studies observed successional patterns throughout the Gulf of Maine. Low winter biomass, from October through January, was characterized by a low abundance of diatoms with fewer numbers of cold-water dinoflagellates. [Small flagellates were not enumerated.] Spring bloom communities replaced winter populations typically in March/April, with diatoms being dominant. The summer community was composed of dinoflagellates, coccolithophores, and to a lesser extent diatoms. Late summer/early fall communities were dominated by localized diatom blooms with a dinoflagellate/coccolithophore background community. More recently, Tamigneaux

et al. (1997) reported similar phytoplankton succession in the Gulf of St. Lawrence, just to the north of the Gulf of Maine. The spring bloom community was dominated by chain-forming diatoms, which gave way to a small (<5 $\mu$m) flagellate-dominated community in the summer, and cyanobacteria also reached their peak in the late summer. The fall phytoplankton populations were found to consist of dinoflagellates, small flagellates, and diatoms.

### 3.4.3 Response of phytoplankton to environmental forcing

Coastal marine phytoplankton live in a dynamic environment with fluctuating conditions. The success of any species depends on its ability to maintain growth under these unstable conditions. There are many factors – both abiotic and biotic – which influence the growth of phytoplankton, and these at times can be differentially favorable for some phytoplankton and simultaneously disadvantageous for others. The set of abiotic environmental variables used in this study was based on lists assembled by Longhurst (2007) and Smayda (1980) of factors controlling the growth of phytoplankton and influencing species succession, and from other studies looking at environmental factors and the response of phytoplankton (Harris, 1986; Grover and Chrzanowski, 2005; Kruk et al., 2002; Reynolds, 1984). The primary variables in this list are temperature, light availability, turbulence, salinity, nutrients, mixed layer depth, and the ratio of the euphotic depth to the mixed layer depth. The cube of the wind speed was used as a proxy for turbulence. Other influential factors not considered here are predation and sinking rate (Smayda, 1980). There is no single factor that determines the community structure, but rather a collective combination of these variables. Multivariate analysis can reveal which

environmental factors are significant. Grover and Chrzanowski (2005) reported that environmental variables accounted for less than half of the variance of phytoplankton composition using multivariate analyses, while Kruk et al. (2002) reported that environmental variables accounted for 75% of the variability in phytoplankton community composition. Both of these studies were based in freshwater lakes, and were examining the composition of phytoplankton to the genus and/or species level. In these aforementioned studies, temperature, light and nutrients were the dominant environmental factors. The objective of multivariate analysis in this study was not to develop regression models between environmental factors and phytoplankton response, but to determine the significant environmental variables and if there is a connection between environmental variables and phytoplankton communities defined at broad levels, not on species. There was strong qualitative evidence of this as exhibited by the groupings of phytoplankton communities within the PCA biplots.

Broad associations between data points of similar community composition correlated to environmental factors. The community composition, while dominated by changes at the seasonal level, can also vary over shorter time and space scales. The environmental factors have a strong seasonal component, but are also at any one time governed by the local meteorological conditions. This includes wind events and coastal river runoff from storms, both of which can impact water column stability. The spring and fall are subject to more variations in the local weather, which consequently leads to more variable, mixed communities than summer. The general pattern observed in the series of PCA biplots showed an oscillation between summertime and wintertime conditions, both of which contained different phytoplankton populations of *flagellate-dominated* and *diatom-dominated* communities,

respectively.

The first principal component (*PC1*) contained about half of the variance of the environmental variables, of which temperature, nitrate, and wind speeds were the dominant factors. This has also been shown to be the case in previous studies (Grover and Chrzanowski, 2005; Kamykowski and Zentara, 2003). Temperature has been shown to influence the species dominance of cultured phytoplankton populations (Goldman, 1977), as temperature directly impacts growth rates and shapes potential fundamental niche zones for species. It is also strongly correlated with nutrient concentration and other factors with strong seasonal variability (PAR, wind speed, $Z_m$). The robust nutrient-temperature relationship has been used to create nutrient-depletion temperatures – the temperature at which a nutrient concentration reaches a level deemed 'deplete' (Kamykowski and Zentara, 2003). In many cases, this is defined as the temperature when the nutrient concentration reaches zero from a line fitted to nutrients versus temperature data. In actuality, the nutrient level at which diatoms disappear may be non-zero. The evidence gathered here indicates that the summer absence of diatoms is due to silicate and/or nitrate limitation. The demand for silicate by diatoms is unique, and without it there cannot be any diatom growth. Previous studies show a marked decline in diatom biomass with silicate levels less than $2\mu$M (Egge and Aksnes, 1992). Silicate levels appeared to diminish faster than nitrate during the spring to summer transition in 2005 at most stations. Nitrate continued to be removed after the silicate reached its minimum. The re-appearance of diatoms in the surface did not occur until September, and then in only low numbers. Silicate and nitrate continued to rise through the fall as a result of physical processes, and was not apparently being significantly removed by

diatoms (which were perhaps being controlled through selective grazing processes). The diatom community did start to replenish itself in the late fall, but by this time of the year light availability became a potential limiting factor for phytoplankton growth.

Light throughout the mixed layer becomes limiting due to both shorter daily photoperiods and an increase in $Z_m$, which mixes cells over greater depth ranges and results in lower overall light exposure. $Z_m$ is dependent on the mixing energy provided by winds, and the thermal heating or cooling of the water column. The euphotic depth – $Z_{eu}$ – is independent from $Z_m$, and depends on the surface light intensity and the light attenuation coefficient. The ratio $Z_{eu}/Z_m$, used here as an index for the potential of light limitation, has important consequences for phytoplankton growth and species selection. Decreasing values of $Z_{eu}/Z_m$ can result from diminishing PAR and/or a deepening of $Z_m$, both of which occurred in winter and caused $Z_{eu}/Z_m$ to decrease below one. Data presented in Harris (1986) showed that variability in $Z_{eu}/Z_m$ drove changes in species composition, and in turn caused changes in the community composition which occurred on the scale of days.

The role of nutrient limitation and light limitation each impose a different challenge for phytoplankton, and thus each will affect the community in different ways. Low-light tolerant species will be able to survive the wintertime conditions better than species that have high light requirements. This physiological constraint is operating at the species level, and less obviously at the group level. Diatoms may be able to maintain dominance, but within that group species composition is changing as a result of the conditions. This was seen in coastal stations where chain-forming

diatoms disappeared and were replaced by the large diatom *Coscinodiscus spp.* This phenomena was also observed by Smayda (1980) in winter in Narragansett Bay.

### 3.4.4 Grazing impacts

Grazing effects were not explicitly considered in this study, and it is not known to what extent the grazing community influenced the composition of the phytoplankton community. To consider how grazing may be impacting the phytoplankton community, the study by Tamigneaux et al. (1997) provides insight into zooplankton/phytoplankton trophic interactions. Tamigneaux et al. (1997) studied the impacts of grazing on phytoplankton in the Gulf of St. Lawrence, which has phytoplankton community successional patterns similar to those of the Gulf of Maine. Tamigneaux et al. (1997) examined the grazing impacts on two different phytoplankton size classes, and observed that the size structure of grazers and phytoplankton had similar seasonal changes. Despite the high grazing efficiency of large ciliates on diatoms in the spring (by high consumption rates), the grazers did not succeed in controlling phytoplankton growth, and diatoms were still able to be the dominant phytoplankton during the spring bloom. Nano-phytoplankton (consisting of small flagellates) and cyanobacteria showed tighter biomass control by their protozooplankton grazers (consisting of smaller ciliates). However, during the summer, copepods had high concentrations and were preying on protozooplankton, allowing a relaxation of the biomass controls on nanoflagellates and cyanobacteria which permitted increases in the biomass of these phytoplankton.

Manning and Bucklin (2005) studied the copepod community off the coast of New Hampshire during 2002 and 2003, and reported the highest biomass of copepods were found in July/August (Figure 3-8). The summertime periods of the highest zooplankton densities in this figure coincide with the highest biomass of cyanobacteria from this study. The conclusions of Tamigneaux et al. (1997) could apply to the Gulf of Maine, and are supported by the data. The effects of grazing are most notably expressed in the phytoplankton community by increases in the cyanobacterial population in the summer, when protozooplankton are being consumed by copepods. The net effect was that the overall composition of the summer phytoplankton community had an increase in the relative cyanobacterial biomass compared to flagellates, but the relative contribution of cyanobacteria was never greater than 20 percent. This did not change the summer community from being flagellate-dominated. The impact of grazing on the spring and fall phytoplankton communities is not known.

### 3.4.5 Ecological significance

Margelef (1978) suggested that variability in phytoplankton composition was driven by 2 main factors: turbulence and nutrient availability. Under this view, phytoplankton 'life forms' are adapted to different habitats along a continuum between two extremes – a nutrient-replete high turbulent state favoring diatoms and a nutrient-deplete stratified state favoring flagellates. The findings of this research conform with this conceptualized model; that is, the distribution of phytoplankton populations follow a pattern closely associated with the seasonal progression of environmental factors that affect turbulence and nutrients. In the western Gulf

of Maine, *diatom-dominated* communities tend to occur in the winter and spring when nutrients and turbulence are high, and *flagellate-dominated* communities during summer when nutrients and turbulence are minimal. Reynolds (1984) addressed some shortcomings with the Margalef Mandala and extended it to freshwater phytoplankton. This basic template was later adapted to marine environments as a means of describing specific marine habitats associated with dinoflagellate taxonomic preferences (Smayda and Reynolds, 2001) and was called the 'Intaglio' - in essence a modification of the Margalef Mandala with phytoplankton survival 'strategies' and habitats superimposed. This plot is portrayed with data from this study coded by phytoplankton group in Figure 3-9. [Note: nitrate concentration was used as a proxy for nutrient supply for the y axis and the x axis has $Z_{eu}/Z_m$ inverted to $Z_m/Z_{eu}$]. In general, the 2 main phytoplankton groups (diatom and flagellate groups) tend to occupy different areas in Figure 3-9 and conform to the habitat preferences as partitioned by Smayda and Reynolds (2001). *Flagellate-dominated* points occupy areas described as stratified and post-upwelling, while *diatom-dominated* points tend to lie in the temperate ocean and shallow shelf water regions. The increase in nutrients is associated with 'mixing events' (e.g., wind or convective mixing) which break down vertically stratified layers and replenish the surface waters with nutrients.

The marine environmental state passes through different habitats along a trajectory governed by its seasonal progression, with a corresponding change in the phytoplankton community. The physical and environmental factors of a habitat impose constraints upon the existing population, which is exploited by species that can best tolerate or take advantage of these conditions. Smayda and Reynolds (2001) suggest that the ability of a species to achieve success is based on its cellular phys-

iological requirements and its morphological characteristics in the context of the environmental condition. From this perspective, diatoms dominate in the nutrient-replete, high turbulent waters found in the winter and spring (and occasionally fall). Turbulence is suited to diatoms since they do not have motile capabilities, and without some level of turbulence non-motile particles denser than water will quickly sink (Smayda, 1980). Conversely, turbulence can damage the flagella of motile phytoplankton, and cause structural damage to flagellated cells. These environments may not be suited for these types of cells. The results in this study adhere to the Mandala/Intaglio template, but do not definitively confirm the concept of the 'life form' selection process behind the distributions. What is evident from this analysis is that phytoplankton communities in the Gulf of Maine exhibited a preference for given habitats. These habitats can be defined in terms of physical/envirionmental factors, which have different phytoplankton communities in terms of the relative abundance of diatoms to flagellates.

It should be noted that Smayda and Reynolds (2001) assigned different dinoflagellate species (not phytoplankton groups) to different regions in Figure 3-9, and that dinoflagellates as a whole are distributed across many habitats on this template. This is also true for other classes. For example, diatoms can be found in habitats ranging the whole spectrum. Species are not excluded from any one area of the template which depicts areas where certain species might be favored. This applies to the phytoplankton community grouping defined in this study. As exhibited in Figure 3-9, *flagellate-dominated* communities were found in the high-nutrient, high-turbulent zones of the Intaglio favorable to diatoms, and *diatom-dominated* communites were found in low-turbulence, low-nutrient zones favorable to flagellates.

Smayda and Reynolds (2001) suggest that such occurrences are the consequence of the stochastic aspects of phytoplankton distributions; that is, a species that is 'on site' with the largest number of inocula and most appropriate life-form morphologies/physiologies are likely to gain the greatest advantage. In this sense, what the template suggests is the most probable habitat for a given species. This applies to the phytoplankton community distributions as well, suggesting that this approach could be used to indicate the most probable phytoplankton community given a set of nutrient/light/mixing conditions.

## 3.5 Conclusions

This study characterized the surface phytoplankton community and the physical/chemical environment in the western Gulf of Maine during a three year period from 2004 through 2007. Annual patterns recurred in phytoplankton composition over this period, and the variations in phytoplankton composition at a broad level of classification showed strong correlation with environmental variables. The dominant mode of environmental variability (the first principal component) was associated with variables that had a strong seasonal signal (e.g., temperature and PAR).

The oscillation of community composition from diatoms to flagellates and back occurred from winter/spring to summer/fall in every year. However, the magnitude and exact timing of the events varied from year to year. The spring blooms of 2006 and 2007 were of greater magnitude compared to 2005, but were of shorter duration. The fall bloom of 2006 was the largest of the 3 fall periods and reached a magnitude comparable to the spring bloom.

The environmental factors that were linked to phytoplankton community composition in this study include variables that can be measured from space which includes temperature, light availability, and wind fields. If the link between phytoplankton composition and environmental factors is robust, then it may be possible to use information measured from satellites to predict phytoplankton populations along with a statistically defined measure of confidence or uncertainty. This will be explored in the next chapter.

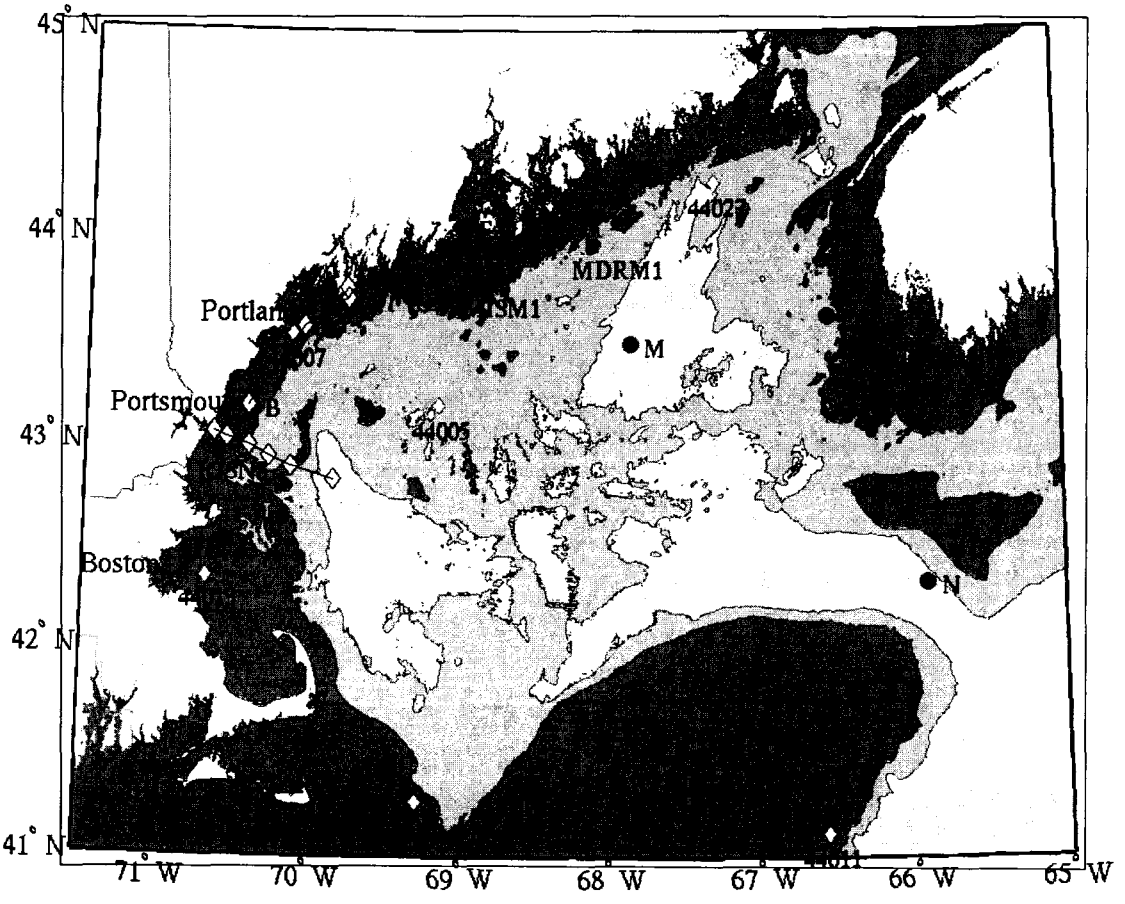



Figure 3-1: Study area showing station and buoy locations; stations (white diamonds) are shown for the Coastal and Wilkinson Basin transects. Wind fields were extracted from the nearest buoys – GoMOOS (red circles), NOAA (yellow diamonds) or C-MAN (blue squares) locations.

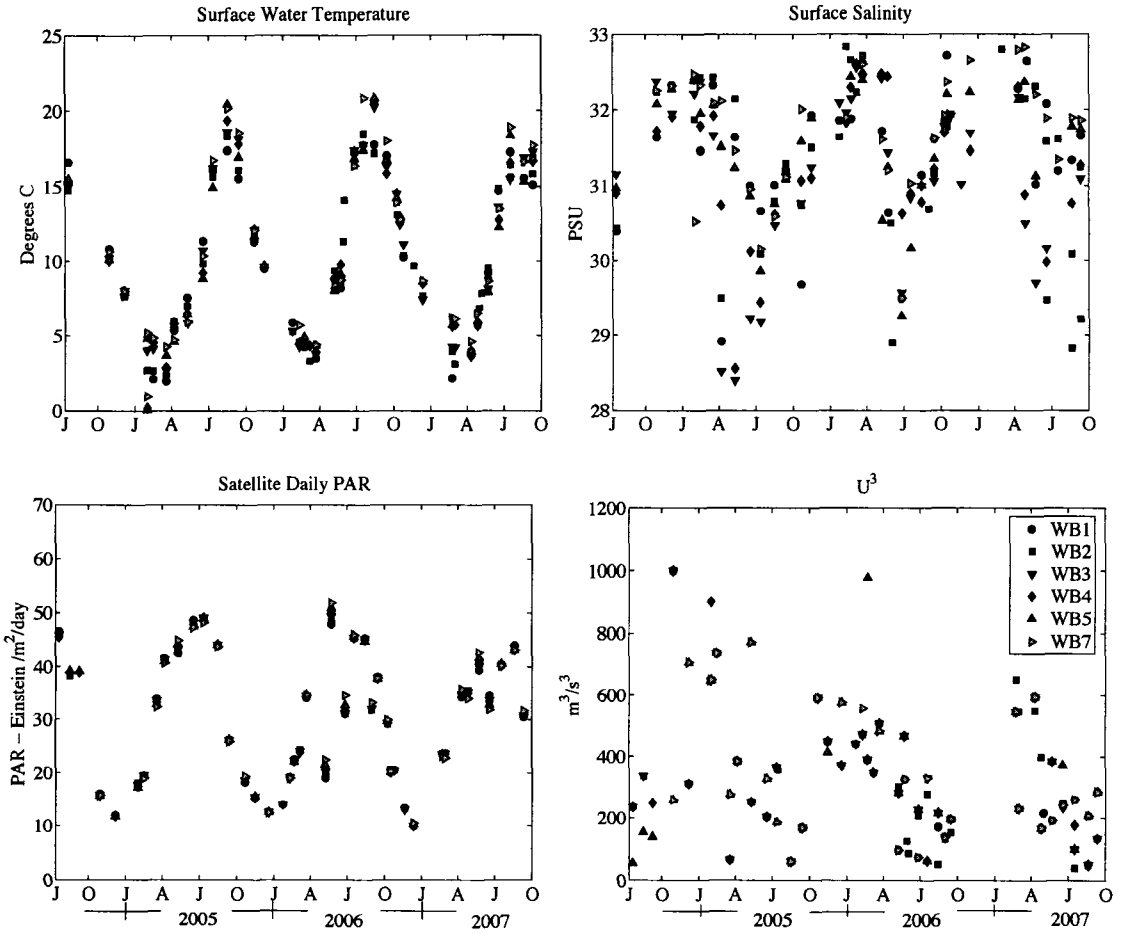



Figure 3-2: Time series of surface water temperature (upper left), salinity (upper right), satellite daily PAR (lower left), and the wind field $U^3$ (lower right). Temperature and salinity are from *in situ* measurements. PAR fields were obtained from satellite, and wind fields were obtained from the nearest fixed buoy. Both were averaged for the preceding 8 days from the each station date.

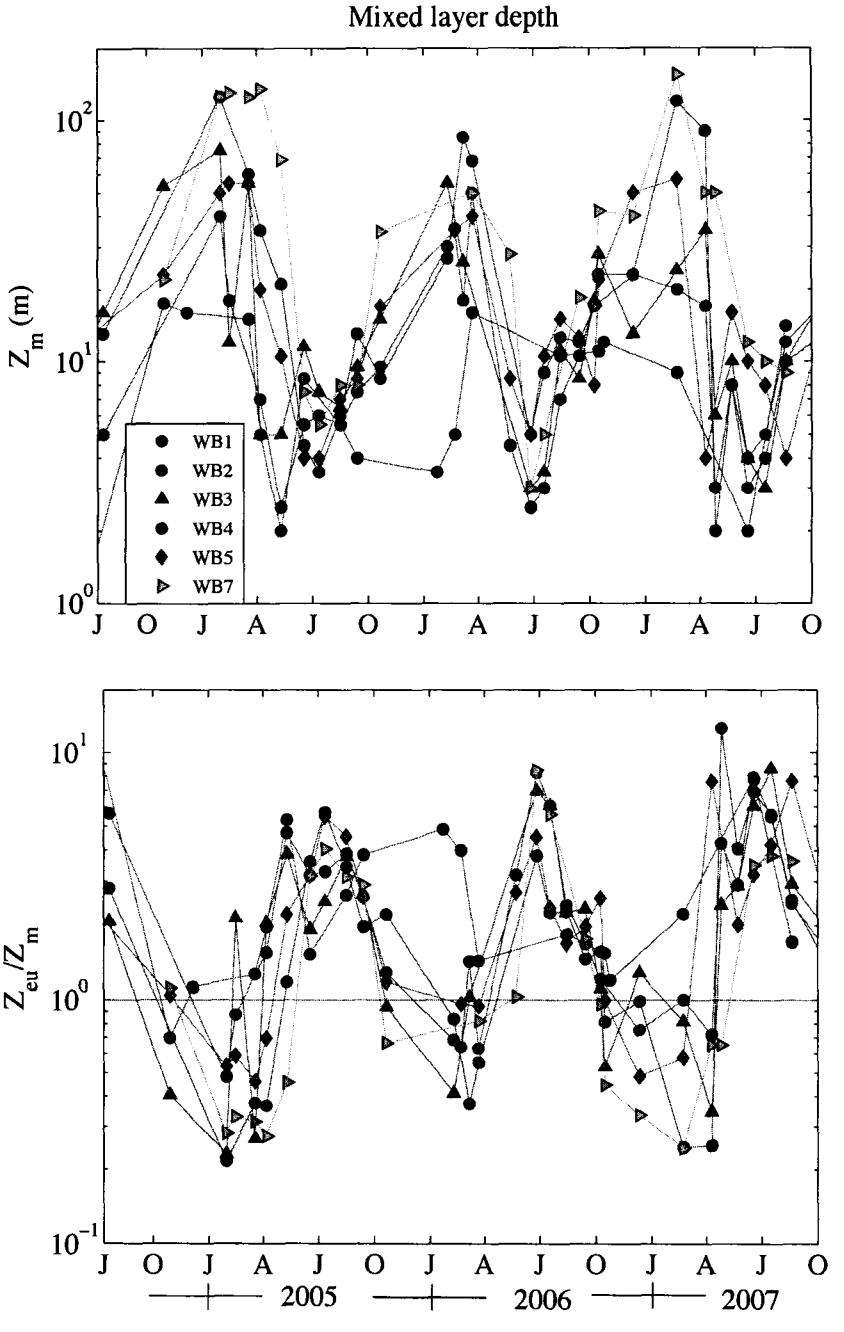



Figure 3-3: Top: Time series of mixed layer depth along Wilkinson Basin. Bottom: time series of the ratio of the euphotic depth ($Z_{eu}$) to mixed layer depth ($Z_m$) along Wilkinson Basin. Red line (1:1) separates potential light-limited conditions (below) from light-saturated conditions (above) in the mixed layer.

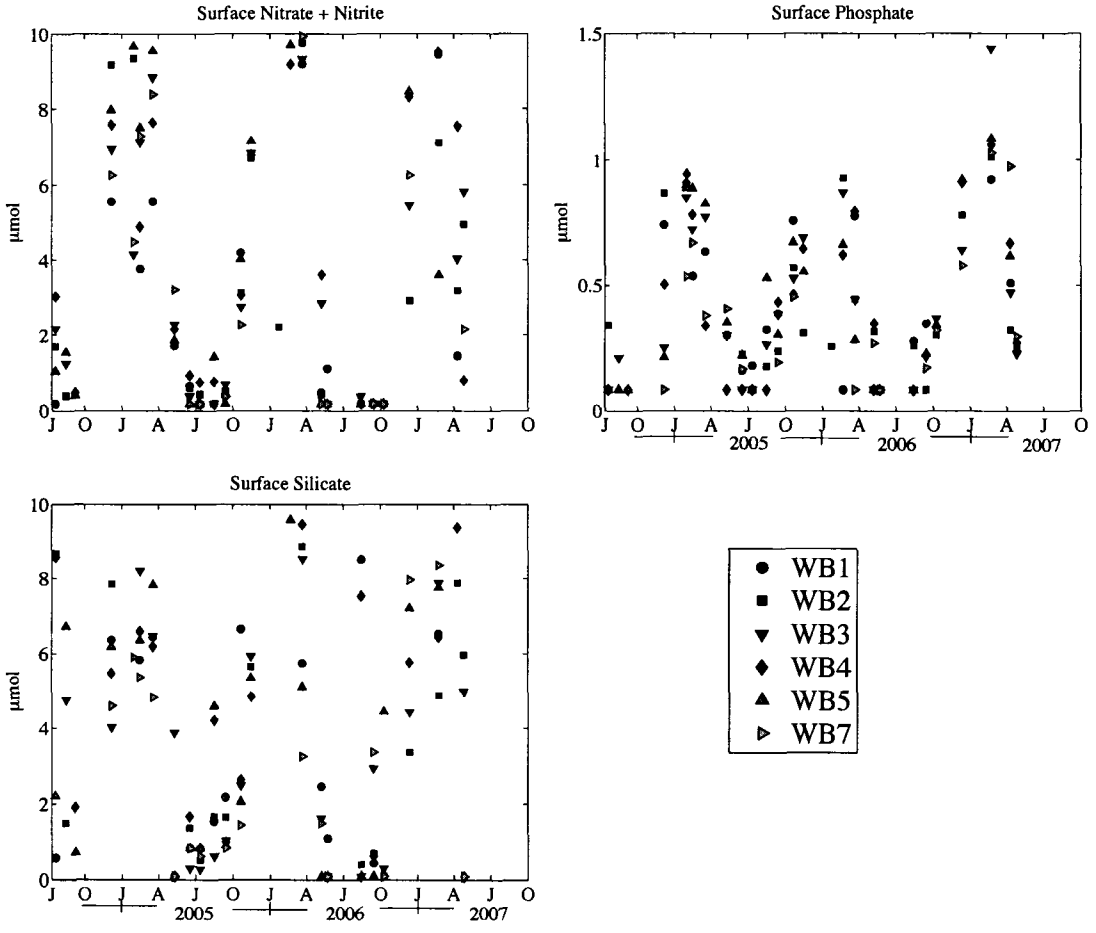



Figure 3-4: Time series of surface nutrients: nitrate + nitrite (upper left), phosphate (upper right) and silicate (lower left) from Wilkinson Basin stations.

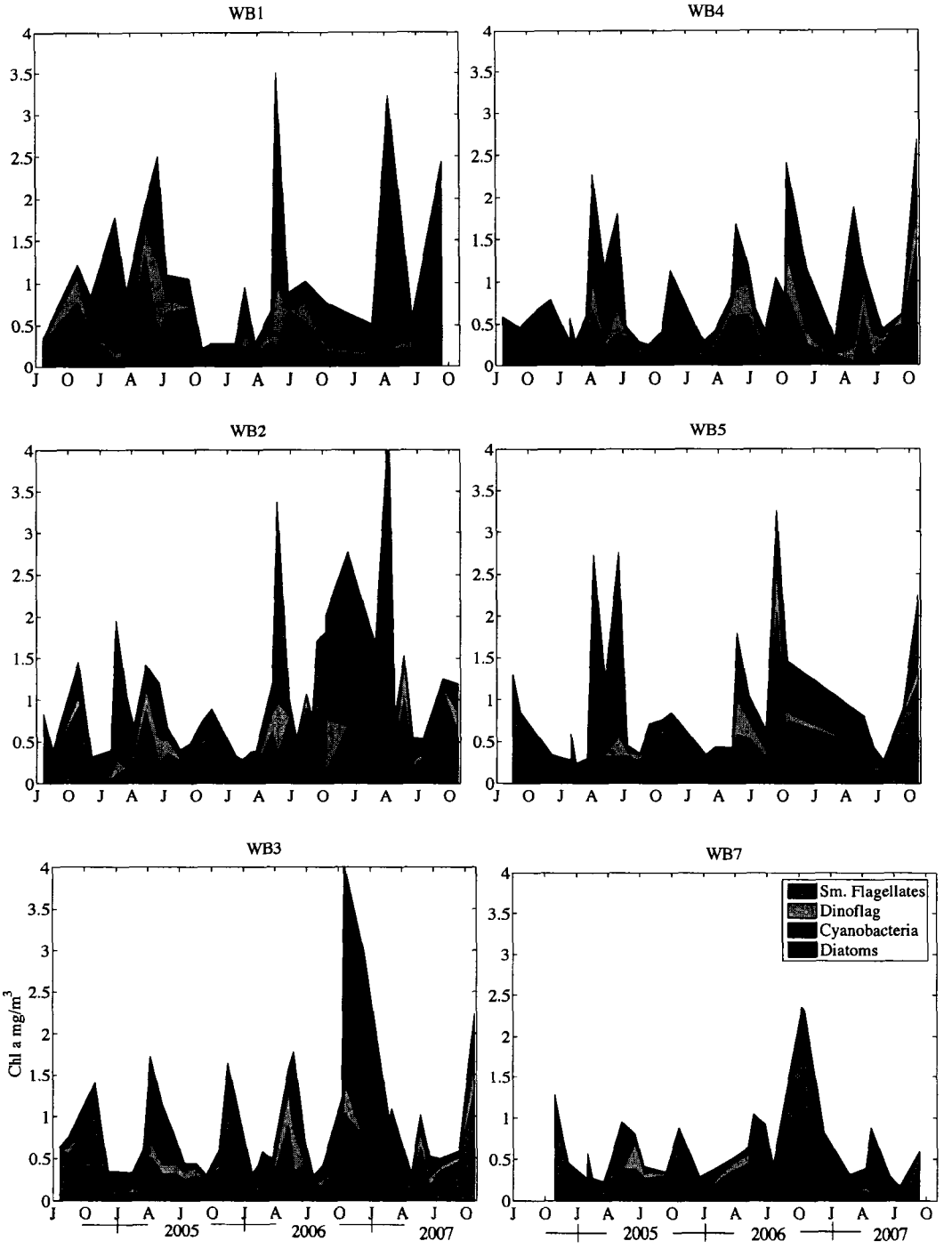



Figure 3-5: Phytoplankton composition time series along the Wilkinson Basin Transect derived from pigment data via CHEMTAX. Four groups are displayed: diatoms, cyanobacteria, dinoflagellates and small flagellates. The small flagellates are the sum of cryptophytes, prasinophytes, chlorophytes, chrysophytes and prymnesiophytes.

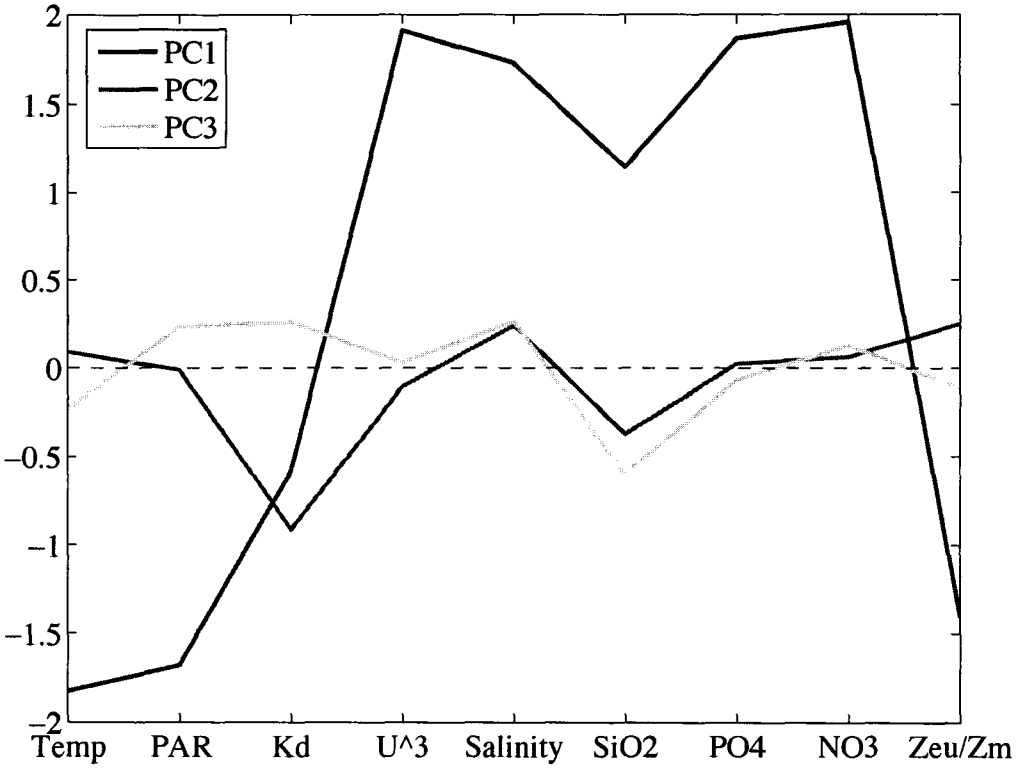



Figure 3-6: PCA eigenvectors for the first 3 principle components (accounting for 71% of the variance).

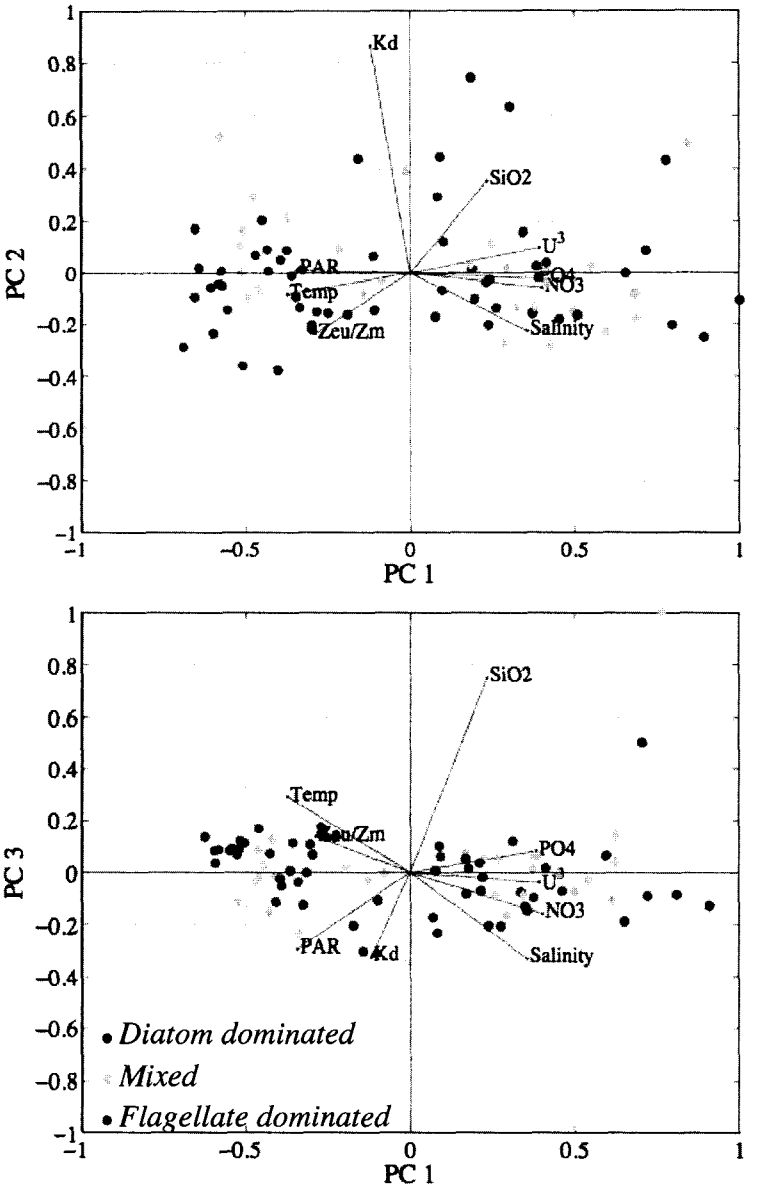



Figure 3-7: PCA biplots for PC 1 and PC 2 (top), and PC 1 and PC 3 (bottom). Data points are color-coded by phytoplankton composition: blue - *diatom-domintated*; green - *mixed*; red - *flagellate-dominated*. Note: signs are different from previous plot as the variable with the greatest magnitude (NO2) was assigned a positive value, which reversed the signs.

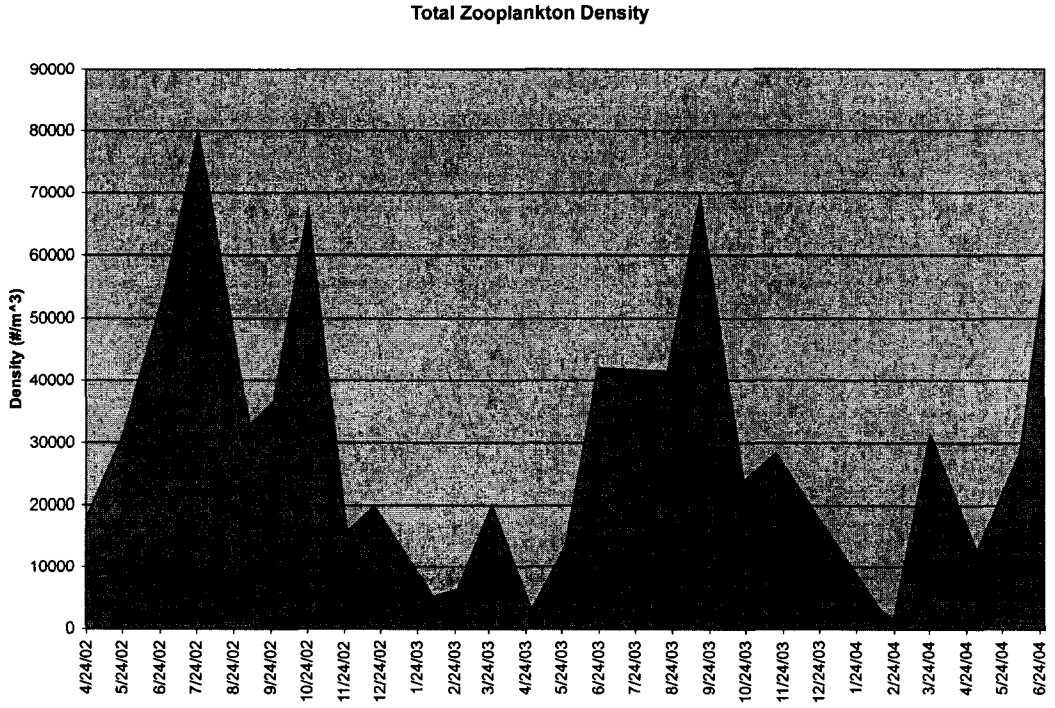



Figure 3-8: Time series of total zooplankton density at WB2 between April 2002 and June 2004. (Figure provided by Chris Manning.)

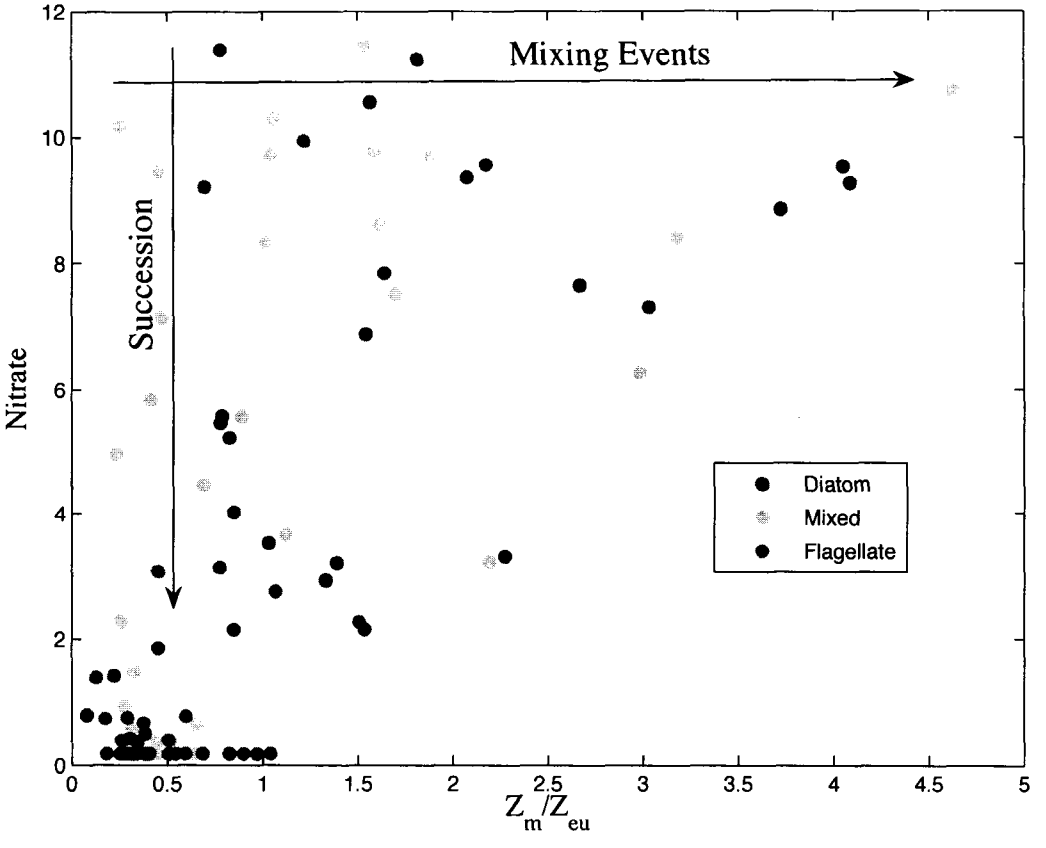



Figure 3-9: Depiction of study data on a Reynold's Intaglio plot. Data are coded by phytoplankton category. Mixing events - measured by increasing $Z_m/Z_{eu}$ - tend to disrupt the successional cycle from becoming flagellate-dominated nutrient limited and low $Z_m/Z_{eu}$) and spurs communities that favor diatoms.

# CHAPTER 4

# MAPPING THE DISTRIBUTIONS OF PHYTOPLANKTON COMMUNITIES IN THE WESTERN GULF OF MAINE USING SATELLITE DATA

*The following chapter has been submitted to the Journal of Geophysical Research - Oceans.*

**Abstract** A classifier that uses hydrographic information as input was implemented to predict phytoplankton community composition in the western Gulf of Maine. Three phytoplankton communities were designated – *diatom-dominant*, *flagellate-dominant*, and a diatom-flagellate mixture – based on HPLC pigment methods and the resulting proportions of diatoms versus flagellates. Co-measured environmental data were then sorted by phytoplankton community, and the statistical properties of the hydrographic variables associated with each community were characterized. These properties were then used to create a classifier that was applied to satellite data for computing the likelihood of each phytoplankton community existing at each pixel. Before applying the classifier to satellite data, its performance was

evaluated using *in situ* data. Based on a set of five environmental variables – water temperature, PAR, salinity, wind speed, and the light attenuation coefficient – the classifier achieved an average success rate of 82%. When applied to satellite data, the result were maps of the most likely phytoplankton communities and associated membership maps expressing the likelihood of each community. These patterns are consistent with previous ship-board observations of phytoplankton community composition. This new method enables phytoplankton communities to be mapped from satellite data at scales that are required for further understanding of marine phytoplankton ecology. These maps have the potential to be used in constraining variable parameters in primary productivity models, and as a source of comparison for marine ecosystem/biogeochemical models of phytoplankton distributions.

## 4.1 Introduction

The prediction of phytoplankton community composition remains a challenging problem in aquatic ecology. The 'rules' governing assembly are uncertain (Cloern and Dufford, 2005) and frequently attributed to species 'being in the right place at the right time' (Smayda and Reynolds, 2001). This is partly attributable to the environment that phytoplankton inhabit, which can experience dramatic changes in conditions. At any given time, it is the combined influence of abiotic and biotic factors that govern the composition of the community, and the influence of any one specific factor is dependent on a given species' adaptive capabilities to that factor (Smayda, 1980). Thus, phytoplankton are subject to selective pressures which influence the community composition and which can range over ecological and geological time scales. Turbulence – and its effect on water column stability – has been seen as a key environmental factor in selecting for diatoms or coccolithophores (Tozzi et al., 2004). Similarly, temperature was correlated with the decline of diatoms and increase in dinoflagellates in the North Atlantic from a 50-year data set of continuous plankton recorder data between 1950 and the present (Leterme et al., 2005). These studies support the long-standing paradigm illustrated in Margalef's Mandala (Margelef, 1978) that certain phytoplankton types are favored over others when conditions of their habitat preference occur.

The composition of the phytoplankton community is important because different species influence biogeochemical cycles and marine food webs in selective ways. Diatoms, for example, are often viewed as distinct from other phytoplankton for several reasons. They have a unique requirement for silica and are adapted to grow

at high rates under turbulent conditions, which often leads to bloom domination following turbulent mixing events. This quality makes them major contributors to the export of organic matter to the deep ocean (Falkowski et al., 2003), as many can sink before grazers have a chance to consume them. The nutritional qualities of diatoms and other phytoplankton also affect production at higher trophic levels (Ban et al., 1997; Cloern and Dufford, 2005). As a result, marine ecosystem and biogeochemical models explicitly include diatoms as a distinct phytoplankton group (Doney et al., 2003). Diatoms have also been singled out in bio-optical and primary productivity models (Sathyendranath et al., 2004; Claustre et al., 2005). Thus, their distributions and seasonal cycles are critical to understanding and improving the modeling of these processes and interactions.

Much has been learned about phytoplankton biomass distributions and cycles since ocean color remote sensing platforms became orbital in the 1970's. Global fields of chlorophyll *a* concentration (*Chla*), a proxy for phytoplankton biomass, are now routinely measured from satellites. While this has enabled researchers to observe and monitor changes in algal biomass for the world's marine environment over the last decade, the ability to identify phytoplankton species is limited. It is possible to detect a few species based on their distinctive optical signatures contained in the detected light signals (e.g., Brown and Yoder (1994); Subramaniam et al. (1999)), but in general it is not possible to distinguish the species composition of the phytoplankton community in remote sensing images. Studies that have used remote sensing data for phytoplankton composition have been previously reviewed in section 1.2.

The objectives of this study were to 1) develop a classifier based on environmental (physical) variables for distinguishing diatom from non-diatom populations; 2) evaluate the performance of the classifier using a validation data set, and 3) apply the classifier to satellite imagery. The end result of this analysis was the creation of weekly spatial maps depicting phytoplankton communities within the Gulf of Maine, and accompanying membership maps depicting the certainty or likelihood of each community.

## 4.2 Methods

### 4.2.1 General approach

The general approach was centered on the development and implementation of a classifier that takes as input a set of hydrographic variables and predicts the phytoplankton community composition in terms of the relative dominance of diatoms and flagellates. This approach involved two major steps. The first step used *in situ* data to train the classifier, which is then evaluated using the original data as well as 'unseen' subsets of the data. In its simplest form, the classifier predicts the phytoplankton community based on observed physical/hydrographic variables. It calculates the distance of an observed environment (a vector of the hydrographic measurements) to the mean conditions of known *habitats*, and selects the closest *habitat*. Based on the training set, each *habitat* is associated with one of three phytoplankton communities.

The second step involved the application of the classifier to satellite data using

the same basic form of classifier, which was then adapted with a fuzzy logic approach. The reasons for using a fuzzy approach in the application to satellite data were based on the increase in uncertainties inherent in satellite data and the fact that fewer variables are accessible from satellites. An example of the former is the imprecision in satellite-derived measurements of a variable compared to an *in situ* measurement. As a consequence of the latter, habitats became less defined as a variable(s) was removed from its characterization and there was an increase in the overlap of communities.

Fuzzy logic was first introduced by Zadeh (1965) as a mathematical way to represent vagueness and imprecision inherent in data. The idea behind fuzzy sets simply states that an object can have partial membership to more than one set. This concept does not preclude the classical view that an object must belong exclusively to only one set. It is in fact a superset of classical set theory, as full membership to exclusively one set is still permitted either directly from the memberships themselves, or through the 'hardening' of fuzzy memberships (i.e, assigning the object to the class with the highest membership).

In the application to satellite data, the likelihood that a phytoplankton community exists at each location in the image is estimated using fuzzy membership functions. The fuzzy memberships are calculated using the same distance measurement as in step one, but a chi-square probability distribution function is applied to this measurement. The result is a set of numbers between 0 and 1 representing the likelihood that the observation 'belongs' to the known habitats. At each pixel, the memberships for all habitats belonging to the same phytoplankton community were summed, and normalized to the sum of all memberships. The end result were

'membership maps' depicting a probability of occurrence for each of the three communities at each pixel. Both fuzzy and 'hard' memberships will be shown with the satellite data.

### 4.2.2 Training the classifier

The classifier was trained on a data set that was collected in the western Gulf of Maine (Figure 4-1), and processed as described in section 3.2. There were two components of the data set: 1) biological data pertaining to the phytoplankton communities derived from HPLC analysis, and 2) the co-measured hydrographic data set comprising water temperature, photosynthetic available radiation (PAR), salinity, the diffuse light attenuation coefficient at 490 nm, and the strength of mixing from the wind (proportional to the cube of the wind speed). The total number of data points that had these variables was 255. The five variables selected were based on their current (or future) availability from remote sensing.

A schematic illustrating the sequence of the classifier training is shown in Figure 4-2. The initial step was to sort the *in situ* physical data into three subsets belonging to the three phytoplankton communities. The number of points in each phytoplankton community were 51, 94 and 110 for *diatom-dominated*, *mixed*, and *flagellate-dominated*, respectively. Within each community, the physical data were distributed across an unknown number of clusters, representing the *habitats* where the phytoplankton communities were found. The *habitat* centers were determined by applying a clustering algorithm to each community subset of physical data. There are many clustering algorithms available, but one that has been used in a wide vari-

ety of applications is the fuzzy $c$-means (FCM) clustering algorithm (Bezdek, 1981). The FCM was selected for this work based on its commercial availability (as a Matlab toolbox), and the broad array of *validity functions* that have been developed to assess the performance of the FCM clustering process.

The exact number of *habitat* centers ($c$) was selected with the aid of the validity functions. These functions measure different aspects of the relations between clusters and the input data set for a given $c$. These aspects include the separation of cluster centers, the compactness of clusters, and the overlap of data points shared between clusters. Six validity measures were used to derive the optimal $c$ for each community; these were the compactness and separation index (Xie and Beni, 1991), the Davies-Bouldin index (Davies and Bouldin, 1979), and a family of four related functions known as the Generalized Dunn Indices (Bezdek and Pal, 1993). Since there is no single validity function that is perfect for all conditions, the preferred strategy is the use of multiple validity functions and then 'polling' each validity function to determine which value of $c$ worked best for that measure. The highest number of votes for a given $c$ was a key factor in deciding the best $c$ for each phytoplankton community.

Once the optimal $c$ was determined, each *habitat* was characterized by computing the mean and covariance matrix of the environmental variables. These statistics were used to compute the distance of any point to a *habitat* center. The distance measure used was the Mahalanobis distance (Rencher, 1995) , defined as:

$$Z^2 = (\vec{V_{rs}} - \vec{\mu_{ji}})^t \Sigma_{ji}^{-1} (\vec{V_{rs}} - \vec{\mu_{ji}}) \tag{4.1}$$

where $\vec{V_{rs}}$ is an observation represented as a vector of environmental variables, $\vec{y_{ji}}$ is the $i^{th}$ class mean of the $j^{th}$ phytoplankton community, and $\Sigma_{ji}^{-1}$ is the covariance matrix for class $j$ of the $i_{th}$ phytoplankton community. The Mahalanobis distance is the multivariate equivalent of the standardized random variable $Z = (X - M)/S$, which is the distance of the univariate random variable $X$ from its mean $M$ normalized by the standard deviation $S$.

The habitat with the smallest $Z^2$ is assigned to the data point. A schematic of the sequence of the flow is shown in Figure 4-3. In this depiction, the input data could be *in situ* environmental data or a vector of satellite data. The source of the data is not relevant, only that it follows the form of a vector that matches that of the *habitat* centers. This is the basic form of the classifier, and predicts the phytoplankton community associated with the nearest habitat (cluster). The classifier was then evaluated by comparing the predicted community with its original designation as determined by pigment analysis.

### 4.2.3 Evaluating the classifier

The main aspect of classifier performance that was assessed was *discriminability* and its associated error rate. Discriminability is defined by how well a rule (or classifier) can classify unseen data. To evaluate a classifier, a test data set independent of the training data is required. However, due to the limited amount of data available, all data were used in the training set for the classifier. To compensate for the lack of an independent test set, a number of simulations were performed by randomly removing 10% of the data points from each phytoplankton community to be used as

an evaluation set (Table 4.1). The remaining 90% of the data were used to re-train the classifier following the process described previously and illustrated in Figure 4-2. Performance was evaluated by assessing the predicted phytoplankton community with its designation based on the minimum $Z^2$ distance (following Figure 4-3). This process was repeated 100 times. Performance was evaluated on every training and evaluation paired set.

| Community | #points | Training Set | Test Set |
|---|---|---|---|
| *Diatom* | 51 | 46 | 5 |
| *Mixed* | 94 | 85 | 9 |
| *Flagellate* | 110 | 99 | 11 |
| Totals | 255 | 230 | 25 |

Table 4.1: Sample size of the training and test data sets. Test sets were created from a random selection of 10% of the points for each phytoplankton community. 100 training and test sets were created. Each training set was used to reconfigure the classifier (new class means and covariance matrices), and then evaluated with the associated test set.

It was also of interest to evaluate the performance of the classifier when using progressively fewer input fields. The motivation stems from the state of readiness of certain remote sensing technologies. For example, sea surface salinity will be a future satellite measurement but does not exist presently. Likewise, the technology of high-resolution sea surface winds is not yet fully realized.

To examine this quantitatively, the classifier was configured (i.e., means and covariance matrices generated) with progressively fewer variables as input, henceforth referred to as scenarios (Table 4.2). For each scenario, performance was evaluated on the training set. While this does not cover every possible combination of the

Table 4.2: Scenario configuration of variable composition

| *Scenario* | *Variables* |
|---|---|
| 1 | *Temp, PAR,* $K_{d490}$*,* $U^3$*, Salinity* |
| 2 | *Temp, PAR,* $K_{d490}$*,* $U^3$ |
| 3 | *Temp, PAR,* $K_{d490}$ |
| 4 | *Temp, PAR* |

variables, it does serve to gauge the impact of different variables in the classifier and provide a measure of performance at different levels of knowledge of the system.

### 4.2.4 Applying the classifier to satellite data

The classifier was applied to satellite data that were mapped to a common projection over the Gulf of Maine. The pixel resolution was set to 1.25 $km^2$. All satellite data were obtained from NASA (http://oceancolor.gsfc.nasa.gov/). SST data were from the MODIS Aqua satellite and processed at UNH. Photosynthically Available Radiation (PAR) data were generated during the processing of daily SeaWiFS data from level 1 to level 2. This product was derived from the Frouin algorithm (Frouin et al., 2003), and 8-day composites were generated from the daily images at the same map projection and scale as the SST imagery. Wind fields were obtained from fixed buoys situated in and around the Gulf of Maine. Gridded maps of wind fields were generated by assigning each map pixel to its nearest buoy. This was determined to be of better quality than modeled NCEP winds, which have poor performance in coastal regions. Hourly buoy wind speeds were converted to $U^3$ (wind speed cubed) and adjusted to an anometer height of 10 meters. Both PAR and $U^3$ were averaged over 8 day intervals beginning with each calendar year.

The uncertainties increase for remote sensing variables compared to the same measurement made *in situ*. To address this increase in uncertainty, fuzzy memberships were produced from the Mahalanobis distance (described previously) for the satellite data. To convert the Mahalanobis distance to a fuzzy membership, a chi-square probability function was used. In mathematical terms, if the probability distribution of points belonging to the cluster centered at $\vec{\mu_{ji}}$ is normal and $\vec{V_{rs}}$ is a member of that population, then $Z^2$ as defined by equation 4.1 has a chi-squared distribution with $n$ degrees of freedom where $n$ is the dimensionality of $\vec{V_{rs}}$. The likelihood that $\vec{V_{rs}}$ is drawn from that population can be defined as:

$$f_{ji} = 1 - F_n(Z^2) \tag{4.2}$$

where $F_n(Z^2)$ is the cumulative chi-square distribution function with $n$ degrees of freedom. The result is a number between 0 and 1, which was defined as the fuzzy membership of $\vec{V_{rs}}$ to the $i^{th}$ cluster of community $j$.

This calculation was made to each *habitat* center for any given input vector. To obtain the probability that an observation belongs to a given phytoplankton community, the fuzzy memberships ($f_{ji}$) were summed over all clusters associated with the $j^{th}$ phytoplankton community, and then divided by the sum of fuzzy memberships to all the clusters. This resulted in a number between 0 and 1 representing the probability that the observation belonged to that phytoplankton community. This was performed for each phytoplankton community, with the end result being a probability to each phytoplankton community for a given observation, and the sum of the probabilities being equal to one. The community with the highest probability was then assigned as the most likely phytoplankton community (i.e., by 'hardening' the

fuzzy memberships). The results were mapped to the same projection as the satellite data and ultimately represented spatial distributions of expected phytoplankton communities. Satellite performance was evaluated by comparing match-ups between satellite predictions and the *in situ* data.

## 4.3 Results

### 4.3.1 Classifier Evaluation

#### Cluster Analysis

A total of twenty-seven clusters (i.e., *habitat* centers) were identified by applying the FCM algorithm to the data set. The number of clusters were 8, 10, and 9 for *diatom-dominated*, *mixed*, and *flagellate-dominated* communities, respectively. The numbers were selected with the aid of the validity functions (previously described). The data distribution of all 255 points along with the cluster means are shown in Figure 4-4 in a series of plots in two dimensions with temperature as the horizontal axis on each plot. The ensemble of plots presents the data in their five dimensional space, projected onto two dimensions at a time.

In each two-dimensional plot, the separation of *diatom-dominated* (blue) from *flagellate-dominated* (red) points is evident. The distribution of points in the temperature-PAR plot (upper left) has a circular pattern with the time trajectory going clockwise (e.g., sample points from January are located in the lower left corner of the plot; July points are located in the upper right). The plot of temperature-wind speed (upper left) shows similarity to Figure 3-9, which was the depiction of the western Gulf of

Maine data in the Reynolds' Intaglio form. In this plot, diatoms show an association with higher wind speeds. The distribution of *habitat* centers is in accordance with the conceptual models of Margelef (1978) and Smayda and Reynolds (2001). These two plots show that a time-of-year element and a physical mixing/turbulence component are captured in the distribution of the data and their *habitat* centers.

For each *habitat* identified by the FCM algorithm, a mean and covariance matrix was computed. Large symbols in Figure 4-4 are the means associated with the clusters. The covariance matrix expresses the dispersion of the cluster points about the mean. In most cases, the dispersion is low, and the points are concentrated about the mean. However, there are several clusters made up of more disperse points. For example, the *diatom-dominated* cluster center located near the June samples (in the upper-left temperature-PAR plot) is the result of several points dispersed between May and July; it is not physically close to any of these points. Other clusters with dispersed points were evident for the other two phytoplankton communities (e.g., a *flagellate-dominated* cluster in the May region and a *mixed* cluster in April).

These plots show that *diatom-dominated* points are segregated from *flagellate-dominated* points fairly well. Points from the *mixed* community are interspersed through these other two, and exhibit overlap in each of the paired plots. Based on these plots, it is evident that it will be more difficult to discern *diatom-dominated* or *flagellate-dominated* points from *mixed*, than *diatom-dominated* from *flagellate-dominated* communities.

### Performance evaluation

The classifier performance was evaluated by comparing the assigned community of each data point (as determined by the minimum Mahalanobis distance to a cluster) to its 'actual' phytoplankton community based on the pigment analysis. Performance statistics were the percentage of correct classifications. In addition, an 'egregious' error statistic was calculated and defined as when a *diatom-dominated* community was assigned to the *flagellate-dominated* community, and vice-versa.

The performance statistics were calculated on the full data set (100% of the data), and on the training (90% of the data) and corresponding test data (10%) subsets. These results are shown in Table 4.3.

| Data Set | *Diatom-dominant* | *Mixed* | *Flagellate-dominant* | Overall | Egregious Error |
|---|---|---|---|---|---|
| Full (100%) | 81.1 | 77.7 | 87.3 | 82.1 | 3.7 |
| Training (90%) | 82.2 | 74.3 | 80.7 | 78.6 | 7.1 |
| Test (10%) | 54.2 | 49.0 | 56.8 | 55.4 | 16.3 |

Table 4.3: Performance of the classifier on the different data sets based on five variables (temperature, PAR, Kd, salinity, wind speed). The numbers shown are the percentage of correct classifications, and percentage of egregious errors.

Overall, the classifier performance with the full data set was 82% successful, and egregious errors were $< 4\%$. In some sense, this is a measure of the goodness of fit of the classifier to the data. However, success was not 100% because the classifier was trained on subsets of the data, whereas the 82% is reflective of results after re-combining the subsets back into one pool. Thus, the overall performance includes data that were 'unseen' by each community, and reflects the extent to

which clusters overlap and are subject to mistaken classification. If the clusters did not overlap and were well separated, then there would be no misclassification. The performance numbers do indicate that clusters, for the most part, were well separated. Furthermore, the low percentage of egregious errors indicates that the *flagellate-dominated* and *diatom-dominated* communities are more distinct from each other than either is with the *mixed* community.

The average performance values for the 100 simulation training and test sets assess the performance on 'unseen' data. The training sets have performance values that are approximately the same as the results using all the data. However, the performance values for the test data sets were significantly lower in every community. Success values averaged 55%, and egregious errors were 16.3% compared to 7.1% for the training sets.

This decrease in performance is due to several factors. The random removal of 10% of the data points from the training pool at times caused clusters to disappear, significantly impacting the cluster statistics. The test data in these simulations were therefore not able to affiliate with the cluster they were previously part of because it no longer existed. The consequence of this was to assign the test point to another nearby cluster that often belonged to another phytoplankton community, thus lowering the success rate. This suggests two limitations of the present method. First, this classification method is sensitive to the size of the training data set. The clusters are at the low end of a threshold in terms of the number of data points that define a cluster, and thus are not robust. An increase in the number of training data points should lead to greater cluster stability, assuming the data

points join an existing cluster and do not form new clusters in a different region of the environmental hyperspace. This leads to the second point – the number of clusters could change as a result of adding data that lie apart from existing clusters. Thus, adding or subtracting points could modify cluster statistics (means and covariances) and also the optimal number of clusters.

### Classifier performance with fewer input variables

Performance statistics were calculated for the classifier when using progressively fewer variables, reflecting the loss of information due to unavailable data from satellite (Table 4.4). The overall performance values decreased from a high of 82% when using all five variables (scenario 1) to a low of 65% when using only temperature and PAR. The decrease in performance was greater in the *diatom-dominated* and *mixed* communities, while the performance in the *flagellate-dominated* showed a smaller change, declining from 87% to 76%. The greatest increase in egregious errors, jumping from 4% to 8%, occurred when the wind field was removed when going from scenario 2 to 3.

| Scen | Vars | *Diatom-dominant* | *Mixed* | *Flagellate-dominant* | Overall | Egregious Error |
|---|---|---|---|---|---|---|
| 1 | T, P, K, $U^3$, S | 81.1 | 77.7 | 87.3 | 82.1 | 3.7 |
| 2 | T, P, K, $U^3$ | 75.5 | 65.9 | 84.6 | 75.5 | 3.7 |
| 3 | T, P, K | 64.7 | 63.8 | 76.6 | 69.3 | 7.6 |
| 4 | T, P | 56.8 | 58.8 | 76.4 | 65.8 | 6.7 |

Table 4.4: Success (as a percentage) of classifying the full training data (255 points) using different combinations of physical variables (scenarios). Scenario 1 was based on all 5 physical variables, while scenario 4 was based on using only 2 – temperature and PAR. Egregious errors occur when a *diatom-dominated* sample is classified as *flagellate-dominated*, and vice versa. Variables are T=temperature, P=PAR, K=$K_{d490}$, $U^3$ = wind mixing strength, and S=salinity.

These results establish the best case performance for application to satellite data. Since salinity is not yet available from remote sensing, the best achievement for success (scenario 2) is 75%, a decline of 7% from the full suite of variables. When wind was removed (scenario 3) the performance dropped significantly for the *diatom-dominated* community, dropping by 10%, while the performance for the *flagellate-dominated* and *mixed* communities decreased by smaller amounts. The further removal of $K_{d490}$ (scenario 4) caused additional declines, but not as drastic as between scenario 2 and 3.

Wind speed and $K_{d490}$ appeared to be important in differentiating *diatom-dominated* from *mixed* communities. When either variable was removed, data points that were *mixed* became more susceptible to misclassification as *diatom-dominated*, and vice-versa. Since the performance for the *flagellate-dominated* community remains relatively constant across scenarios, the overall performance decrease must be caused by the other two communities. Furthermore, the additional misclassifications must then occur between *diatom-dominated* and *mixed* communities as the variables are progressively removed, with large declines having occurred when wind and $K_{d490}$ were removed.

### 4.3.2 Application to satellite imagery

#### Uncertainty characterization

Satellite data are inherently noisy, and are subject to errors as a result of algorithm and/or instrument uncertainties. To evaluate performance sensitivity to satellite data, the classifier was applied to the *in situ* data with one input variable

replaced with satellite values – one at a time – for temperature, PAR, Kd, and winds. Salinity remained as *in situ* values. These results are shown in Table 4.5. The satellite values were obtained by extracting co-located pixel data from the imagery (8-day averages). Out of the 255 points in the full data set, 233 had valid corresponding satellite match-ups (Figure 4-5).

| Satellite Sub. | *Diatom-dominant* | *Mixed* | *Flagellate-dominant* | Overall | Egregious Error |
|---|---|---|---|---|---|
| No subs | 88.1 | 75.0 | 85.7 | 82.4 | 2.7 |
| SST | 88.1 | 65.5 | 86.7 | 79.2 | 2.7 |
| PAR | 78.6 | 61.9 | 77.2 | 71.9 | 3.4 |
| Kd | 78.6 | 57.1 | 85.7 | 74.0 | 6.7 |
| $U^3$ | 54.8 | 52.4 | 75.3 | 63.2 | 8.8 |

Table 4.5: Classifier performance after substitution of input variables with satellite matchup data (N=233 points). In each case only one data set was replaced by satellite data.

The variable that incurred the smallest overall change was SST, which decreased 3% from the full original data set. [The slightly higher values for the first row in the table compared to the first row in Table 4.3 are due to the omission of data points where there were no valid co-located satellite fields. This was due to either cloud cover or some stations being located inside estuaries and screened out by the satellite land mask.] The variable substitution that caused the greatest change was the wind field, decreasing overall performance from 83% to 61%. The *diatom-dominated* community performance decreased by 34% with this substitution. In general, this community showed the largest decreases when variables were substituted. These results indicate that classifier performance declines when applied to satellite data.

The values displayed in Table 4.5 reflect the sensitivity to 'individual' elements,

and are not the collective effect since only one input variable was replaced each time. The decrease in performance can be attributable to satellite uncertainties that originate from two primary sources. The first source is that the measurements themselves are not identical. Both SST and Kd from satellite are derived from algorithms, and thus have algorithm uncertainty. The second, and more problematic, source is related to the difference in the time period that the *in situ* and satellite measurements were averaged over. This point can be illustrated by examining the relations between *in situ* measurements and the satellite extractions in Figure 4-5. The PAR data set used for the *in situ* and satellite fields are from the same source – the SeaWiFS imagery. Likewise, the wind speed fields used in both data sets are also from the same source – fixed buoy measurements. In both cases, the only difference is the 8-day window over which they were averaged. For the *in situ* fields, PAR and winds were averaged over the eight days preceding the station date. For the satellite fields, the 8-day averages were organized according fixed calendar dates, and thus are from a slightly different time period. The spread in the data points in all plots of Figure 4-5 contain this effect to some degree, as well as algorithm uncertainties for SST and Kd.

To assess total performance using satellite data as input, the classifier was applied to satellite match-ups as described in section 4.3.2 for scenarios 2-4 (as defined in Table 4.2). Classifier performance based on a comparison of match-ups between the satellite hard memberships (to the most likely phytoplankton community) and the original phytoplankton community assignment is shown in Table 4.6.

The overall percentage of correct assignments from the imagery for scenarios 2

| Scen | Var | *Diatom-dominant* | *Mixed* | *Flagellate-dominant* | Overall | Egregious Error |
|---|---|---|---|---|---|---|
| 2 | T, P, K, U | 50.0 | 39.7 | 74.3 | 57.2 | 6.8 |
| 3 | T, P, K | 66.7 | 40.4 | 71.4 | 60.9 | 5.4 |
| 4 | T, P | 40.5 | 45.2 | 72.4 | 56.7 | 9.5 |

Table 4.6: Success (as a percentage) of correct classifications from satellite data (N=233 points) for the scenarios 2-4. Scenario 1 was not included because of the absence of satellite salinity data. Variables are T=temperature, P=PAR, K=$K_{d490}$, and U = wind stress.

- 4 ranged from 57% to 61% and egregious errors from 6-10%. The best results were produced for scenario 3, which was based on temperature, PAR, and $K_{d490}$. Performance for scenario 3 is significantly better than for scenario 4 for the *diatom-dominated* community (66.7% compared to 40.5%) and for egregious errors (5.4% compared to 9.5%). Consistent with the results in Tables 4.4, including $K_{d490}$ as a factor is important for discerning *diatom-dominated* from *mixed* communities. The performance for scenario 2 is lower than scenario 3 (even though scenario 2 uses more variables) because there is a compounding effect from noise in the satellite variables. In the case of scenario 2, wind 'noise' combined with $K_{d490}$ noise worsens performance compared to when the wind is removed (scenario 3). As was shown in Table 4.5, the substitution of 'satellite' wind fields resulted in the greatest drop in performance compared to other variables. There is a trade-off when applying the classifier to satellite data; a reduction in the number of input fields generally results in lower performance, but the inclusion of such fields could lower performance due to the measurement 'noise' which compounds with the uncertainties from the other factors.

### Generation of spatial phytoplankton community maps

The preceding analyses were based on using the minimum Mahalanobis distance as the criteria to assign an observation to a phytoplankton community. The increase in uncertainties and the decline in performance when applying the classifier to satellite data provide the impetus to use fuzzy logic for the satellite application. In addition to providing the most likely phytoplankton community to be expected from a vector of satellite observations, the probabilities of the observation 'belonging' to each of the three communities are produced.

Using the classifier based on scenario 3 (temperature, PAR, and $K_{d490}$), satellite data were used as an input to the membership function, and spatially-resolved maps of the fuzzy memberships were produced. Figure 4-6 shows a set of 8-day composites for SST (from MODIS-Aqua), PAR (from SeaWiFS), and $K_{d490}$ (from SeaWiFS) from September 22 to September 30, 2005. After using these fields as input to the classifier, the resulting fuzzy membership maps are shown in Figure 4-7. These fuzzy maps display the probability of a particular community occurring for each pixel, whereas the hard membership map (lower right) shows the spatial distributions of phytoplankton communities with the highest probabilities.

The fuzzy logic approach allows the possibility of a given pixel having membership to more than one community. In reality, of course, it is not possible for more than one community to exist in a given location at the same time. The fuzzy memberships allow for ambiguity and reflect the fact that there is never 100% certainty that a given community will exist in a given place based on knowledge of the environment. The fuzzy maps when stacked together fit like pieces of a jigsaw puzzle,

with each phytoplankton fuzzy map being a piece. Transitions from one community to another are captured with grading memberships, as the memberships to one community fade into rising memberships to another. While these transitional regions are seen in the hard class map as abrupt discontinuities, the fuzzy maps provide information on the degree of overlap.

The hard membership map, while showing the three classes in one image, obscures the added information from the fuzzy maps – the degree of certainty that exists at each pixel. For example, the fuzzy membership distribution of the *mixed* community is shown to be high not only in the eastern Gulf of Maine, but also near Cape Cod and over Georges Bank. The *flagellate-dominated* community has higher memberships to the latter areas, and thus it appears in the hard classification map that a *mixed* community exists only in the eastern Gulf. However, the fuzzy memberships suggest that there is a reasonable likelihood that Georges Bank is a *mixed* community. In other words, it is less certain that Georges Bank is a *flagellate-dominated* community compared to the central Gulf of Maine, where the fuzzy memberships are very high to the *flagellate-dominated* community and very low to the other two. Thus, the fuzzy memberships allow for the possibility of another community existing at a certain location, whereas the hard membership maps depict the communities as either present or absent.

The distribution of phytoplankton communities and the change in their distributions as governed by changes in hydrographic conditions during September 2005 are shown in the fuzzy and hard membership maps in Figure 4-8 and Figure 4-9, respectively. In this time series, the Gulf of Maine is shown to be dominated by flag-

ellates as depicted in the hard maps. During that one month period, diatoms began to appear in the western Gulf off of Nova Scotia. The apparent evolution of the diatoms can be seen as diatoms progressively stretch southward in the coastal waters off Maine, eventually reaching Boston and Cape Cod. The community distribution patterns are much more discernible in the fuzzy maps. The sequence of the fuzzy maps show the evolution of each community in 'isolation', whereas the hard maps obscure the development of communities that are not the dominant community.

The change in community composition is a relative phenomenon, which could have several causes. For example, the increase in the *mixed* community extent in the coastal areas – displacing the *flagellate-dominated* community present in early September – could be interpreted as a decline in flagellates. Conversely, it could be attributed to a rise in diatoms populations or some combination of both. The chlorophyll images for the same time sequence (Figure 4-10) show increases in chlorophyll levels in the coastal area, indicating it may be that diatoms were increasing in abundance. The pairing of chlorophyll and the membership images are complimentary, and enhance the information about phytoplankton dynamics.

The fuzzy and hard phytoplankton community maps present information that is different from the standard chlorophyll product derived from ocean color satellites. Chlorophyll images display overall biomass levels, while the fuzzy membership maps display the relative phytoplankton community composition from one of three broadly defined categories. When both types of maps are compared, different patterns are revealed. These image pairs highlight the complementary nature of these phytoplankton maps.

## 4.4 Discussion

### 4.4.1 General observations

The sequence of change in phytoplankton communities (referred to as *succession*) is driven by the differential response of phytoplankton species to environmental factors. Typically, the response time to environmental change is the order of 2-10 days (Harris, 1986), although it may be longer. This 'lag' time depends on the species and their physiological adaptive response (e.g., growth rate) to the environmental factor(s) involved, and is generally unknown. This introduces an uncertainty into the prediction of phytoplankton community composition based on coincident environmental conditions. While the lag response time comprises a significant source of uncertainty, there are other sources. The very definition of a 'community' is subject to vagueness, as the criteria which separates different communities depends on subjective thresholds which are not certain. There are also instrument and algorithm errors, which influence data precision and the relationships dependent upon those data.

Uncertainties in input variables propagate imprecision to the output, as the algorithm's 'decisions' are based on imprecise information. A methodological approach for handling data ambiguity is fuzzy logic. This was used as the basis for generating probability maps for phytoplankton community composition from satellite data. The main objective of this study was to construct a classifier to predict the phytoplankton community composition based on environmental variables (temperature, PAR, $K_{d490}$, salinity and wind speed), and to apply the classifier to satellite data. The classifier worked well on the original *in situ* data, with an overall pre-

diction success of 82%. When satellite variables were used, the success was much lower. The uncertainties associated with different satellite data sets were shown to have differential impacts on the performance of the classifier. The overall decline in performance led to the use of a fuzzy logic approach, which was used to generate probabilities for phytoplankton distributions. This approach allowed multiple outcomes to occur, and thus it is difficult to evaluate fuzzy performance using the same criteria as the non-fuzzy approach. The fuzzy approach allows any outcome to occur from given input data and thus always produces a correct result, but the correct result may have a lower probability than another outcome. The performance measures of the non-fuzzy approach, however, provided insight into the performance of the fuzzy classifier.

### 4.4.2 Classifier performance and error analysis

The performance of the classifier in successfully predicting the phytoplankton community was on the order of 63-82%, depending on the variables included, when the data used to 'calibrate' the function was used to evaluate it. Based on the simulation tests using 'unseen' data, the classifier could discriminate the three communities only 50-60% of the time and thus showed an overall decrease in performance. The performance numbers were between 60-65% when satellite matchups were extracted and compared to the training data. These numbers would be closer to the *in situ* results if it were not for errors (or noise) in the fields of the satellite data themselves.

There are few studies to compare these performance numbers against. Sathyendranath et al. (2004) reported a success rate of 72% for a classifier used to distinguish

diatoms from mixed phytoplankton populations, and Alvain et al. (2005) achieved a success rate of 61% for a classifier used to identify four different phytoplankton communities. Iglesias-Rodriguez et al. (2002) gives probabilities of coccolithophore blooms, but does not give performance results for an independent data set. [See section 1.2 for further review.] The classifier designed in this study had success rates that were comparable to or higher than the studies of Sathyendranath et al. (2004) and Alvain et al. (2005). With so few other results to compare, the results of this study and Sathyendranath et al. (2004) set the benchmark for phytoplankton classifier performance.

The satellite performance evaluations were based on the 'hard' class memberships, but this ignores the additional information provided using the fuzzy classifier. The fuzzy values provide a measure of the probability or likelihood of occurrence for a given phytoplankton community based on the observed physical environment. They give the relative likelihood, given the conditions, but almost any outcome has some chance of occurring based on many factors not directly considered in the design of the classifier (e.g., phytoplankton nutrient storage, grazing). When considering these probabilities and the expected outcome, it is important to understand the mathematical behavior of the classifier and the potential sources of error involved in mis-classifications.

The fuzzy memberships are ultimately dictated by the statistical distributions of the *in situ* measurements used in the training set. These distributions are dependent on the number of overall input data points, and the number of clusters per group. Performance diagnostics from the simulations in which 10% of the points

were removed indicate the severity of this sensitivity. A major reason was having too few points in the classifier design. A total of 255 points were available for the training and 26 clusters were identified. For the simulations, the number of points was reduced to 230 (25 being set aside for independent evaluation). The location of clusters varied depending on which points were removed and thus exhibited instability from simulation to simulation. For example, there were test runs when entire clusters were removed from the training pool, and this lowered the performance results when trying to classify the test data for that run (producing values of 0% success for the test data set). As a result, the average performance scores for the simulations were lower compared to those that used the training sets. This is an indication that there too few points in the training set to establish stability in the cluster statistics. In other words, there are *habitats* that are not represented in the data set. The classification of data points from these unaccounted *habitats* generally lead to higher errors. The location and number of these unaccounted *habitats* remains unknown.

An additional indication of data sparseness was seen in the mathematical behavior of the covariance matrices. For several of the clusters represented with too few (e.g., 5 or less) points, the covariance matrices were statistically unstable. As a result singularities occurred during matrix inversion, which caused the calculation of the Mahalanobis distance to blow up. To remedy this, a common covariance matrix (i.e., the average covariance matrix for all classes) was used for these clusters. This is routinely used when the number of data points is low (Hoffbeck and Landgrebe, 1996) and the subsequent singular value problems arise. By assuming that all classes covary the same way, the true statistical spread of points for a given class

is compromised which will affect the membership outcome. Despite this, Hoffbeck and Landgrebe (1996) have shown that the use of a common covariance matrix still leads to higher classification accuracy when the training sample sizes are small.

### 4.4.3 Sources of uncertainties

Most mis-classifications occurred between 'adjacent' classes, that is, between either *diatom-dominated* or *flagellate-dominated* and *mixed.* A much smaller fraction of the errors (2-10%) occurred between *diatom-dominated* and *flagellate-dominated* classes. These were the so-called egregious errors. Thus, if a pixel is classified as *diatom-dominated*, it is highly unlikely that it is actually *flagellate-dominated* and vice-versa. The *mixed* community had the lowest performance, and this is attributed to having mis-classifications with both *diatom-dominated* and *flagellate-dominated.*

The number of error sources is large and has a cascade effect that propagates through the various stages of the analysis. To begin with, the assignment of the data points to phytoplankton groups has some uncertainty. The phytoplankton community assignment was based on results from the pigment analysis, and arbitrary divisions defining 'dominance' by diatoms or flagellates. While the CHEMTAX results for diatom fractions had the best agreement with cell counts compared with other phytoplankton groups ($r^2$ of 0.8 - see 2.2), there was still sufficient uncertainty that could lead to an incorrect assignment. As the cluster analysis was shown to be sensitive to the number of data points, this could have a significant impact on the statistical characteristics for the entire classifier design.

Another significant source of uncertainty lies in the time lag response to en-

vironmental changes for phytoplankton. A typical lag response is between 1-10 days, although some can be longer (Harris, 1986). The causes of time lag responses depend on the environmental factors involved and the species. Species that are stimulated for growth show a lag that is on the order of their cellular growth rate. Other lags can occur under unfavorable conditions, when cells are no longer in a favorable growth environment. This could be caused by nutrient scarcity or low light, for example. Under these conditions, cells can alter physiologic mechanisms to compensate for adverse conditions, thus enabling a population to maintain itself, at least temporarily. For example, diatoms have a storage capacity which enables them to maintain growth despite nutrient-deplete conditions. This could explain the June 2005 phytoplankton community dynamics along the Wilkinson Basin Transect. Diatoms were abundant at nearly all surface samples along the transect on June 17, 2005. Nutrients (silicate and nitrate) at this time were depleted, yet diatoms were numerically dominant. Samples from a cruise 11 days later on June 28 along the same transect showed no trace of any diatoms in the surface waters. The diatom population was either advected out of the area, sank to the bottom, and/or was consumed by grazers.

The practical effect of this has implications on the statistical distributions and cluster properties for data points that are in this mode. In the preceding example, the sample points may have been assigned to a completely different phytoplankton community had the sampling occurred a week later. This situation poses a problem for the classifier, as it obscures the influence of the environment on the composition of the community by assuming that cells are acclimated to their environment. It is not known for how many or which of the samples this was an issue.

This raises the conceptual problem: do such time lags render correlations between species composition and ecological conditions at points in time inappropriate? It probably depends on the environmental situation, and whether species composition has already re-aligned itself to the environmental conditions. The state of the phytoplankton community (i.e., the degree of time-lag adjustments) in relation to the environment is difficult to gauge. The phytoplankton community at the time the *in situ* measurements were taken may have adjusted to the environment, or it may have been in the process of re-adjustment. This source of uncertainty is not associated with any instrument error, but is an inherent aspect of the stochastic nature of phytoplankton community response.

### 4.4.4 Ecological significance

The habitat preference for different phytoplankton groups is generally understood on broad levels, but remains an elusive property at large. Many species for example exhibit a wide tolerance for environmental factors such as pH and salinity, and can be found at any time throughout the year. Other species, such as *Phaeocystic spp.*, appear only during brief periods of the year. In many cases, the factors which trigger a phytoplankton response are not clearly understood. Still, the environment in which phytoplankton live can be described by a variety of variables, and these define the habitat.

Several environmental factors shown to be useful predictors of different phytoplankton communities are also amenable to remote sensing. Several key variables that define habitat preference – turbulence, mixed layer depth, and water column

stability – are more difficult to measure and are not available from remote sensing. These were not included in characterizing the habitats. However, water temperature – one of the most widely measured properties – is related to these other properties, and thus can be viewed as an index of environmental condition (Bouman et al., 2003). Temperature and nutrients in this study were inversely related (figure not shown) in a manner similar to other findings where SST has been used in conjunction with nutrient depletion temperatures (Carder et al., 1999; Kamykowski and Zentara, 2003; Iglesias-Rodriguez et al., 2002). Temperature has also been related to turbulence and mixed layer depth (Rodriguez et al., 2001; Bouman et al., 2003), and used directly to account for phytoplankton growth rates (Eppley, 1972) and indirectly for phytoplankton optical variability (Bouman et al., 2003) and primary production parameter estimation (Platt et al., 2007). The practical consequence is that temperature tends to serve as a proxy for other physical/chemical properties.

The three different phytoplankton communities exhibited a strong separation in five-dimensional space based on the physical variables selected. *Habitat* centers were identified and served as the basis for predicting phytoplankton community composition from satellite measurements of the same variables. However, as variables were removed, the separation of habitats decreased and habitats began increasingly to overlap. Wind strength was important in distinguishing *diatom-dominated* from *mixed* communities, for example. Without this variable present, the chances increased that one would be mis-classified as the other.

The overall uncertainty associated with the classifier was the underlying motivation to use fuzzy logic, which was used to generate probabilities for finding a

community at a given location. These probabilities were supplied by the classifier in the form of fuzzy memberships. From a mathematical perspective, the fuzzy memberships can be regarded as the likelihood of a observation (pixel or measurement) belonging to that community. From an ecological perspective, this does not mean that one would find the community with the highest membership at that location, but gives the probability of that community occurring. A community with a lower probability, however, could also occur.

When the fuzzy memberships or probabilities are mapped, as in Figure 4-8, they typically form spatially coherent patterns even when memberships are not high. For example, the *mixed* community maps in the figure show memberships over Georges Bank and other parts of the Gulf that resemble ocean features. While the *flagellate-dominated* community had higher probabilities in these same areas, a *mixed* community could have existed there instead. The maps show how likely and where potential communities overlap. In the same figure, the *mixed* and *diatom-dominated* communities did not show membership to the water mass in the center of the Gulf of Maine. It is reasonable, therefore, to expect that a *flagellate-dominated* community would be found there, but along the coast and over Geores Bank it is less certain.

Based on the current state of available *in situ* data, buoy technology, and satellite capability, the fuzzy approach is advantageous over a purely classical approach in classifying phytoplankton communities. The amount of uncertainties infiltrate the classical approach leading to mis-classification problems. The fuzzy approach deals with uncertainty by assigning probabilities to different outcomes, thus permitting

more than one outcome to occur. One utility of this approach is that the classifier has the capability of including more variables as the satellite and buoy technology evolve. For these reasons, phytoplankton community prediction is better served with a probabilistic approach.

## 4.5 Conclusions

The results provided here show a new approach for estimating phytoplankton community composition that can be applied to satellite imagery. Based on physical variables measured *in situ*, three different phytoplankton communities showed separation and were successfully identified 82% of the time. However, when the classifier was applied to satellite data, performance declined as noise from measurement imprecision and uncertainties were introduced. To handle this, the classifier was adapted with fuzzy logic that permitted the classifier to produce probabilities of occurrence for each community. It was possible to map these probabilities to represent potential community distributions. This allowed ambiguity in predicting the community existing at any location. The observed communities were also shown to be tracked over time, thus permitting the ability to observe community change (succession) over space and time. The current application considers the phytoplankton community as a whole, but could also be applied to individual species such as *Alexandrium spp.* and others that have significant impacts at the ecological and/or socio-economic level. The prediction of such a species could have enormous benefits.

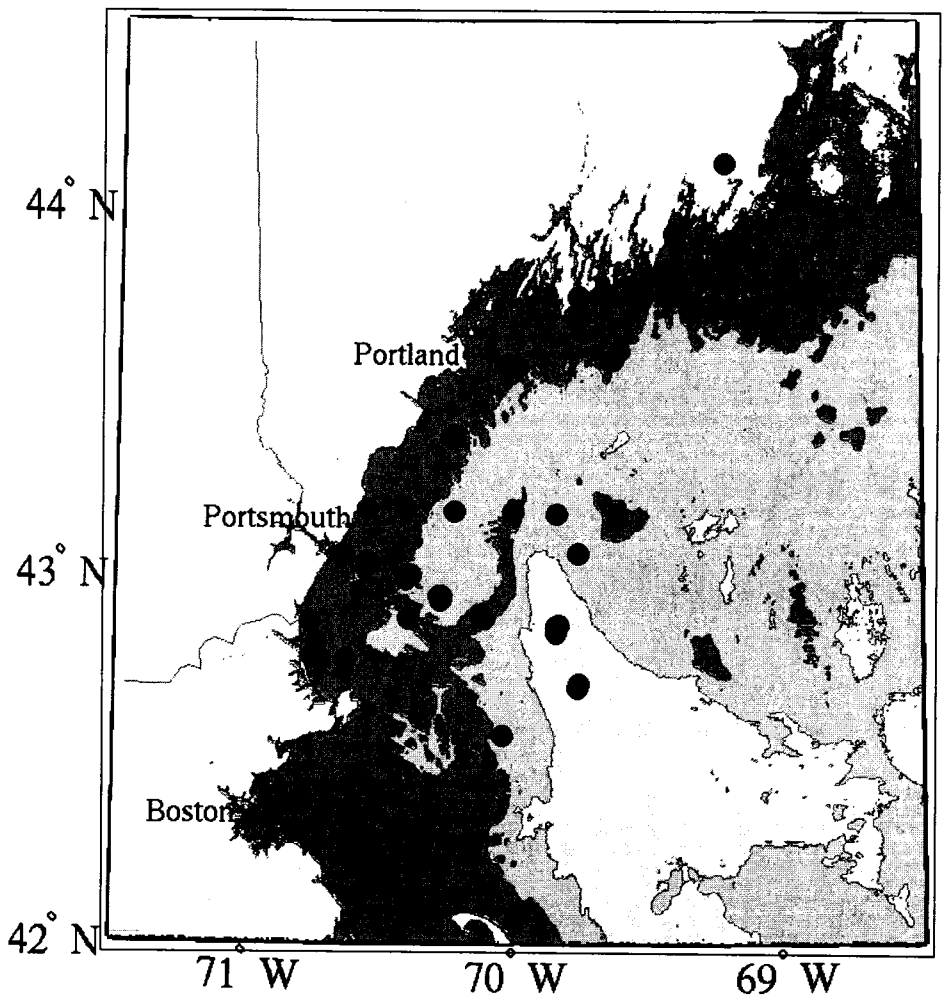



Figure 4-1: Study area with sample locations (N=255).

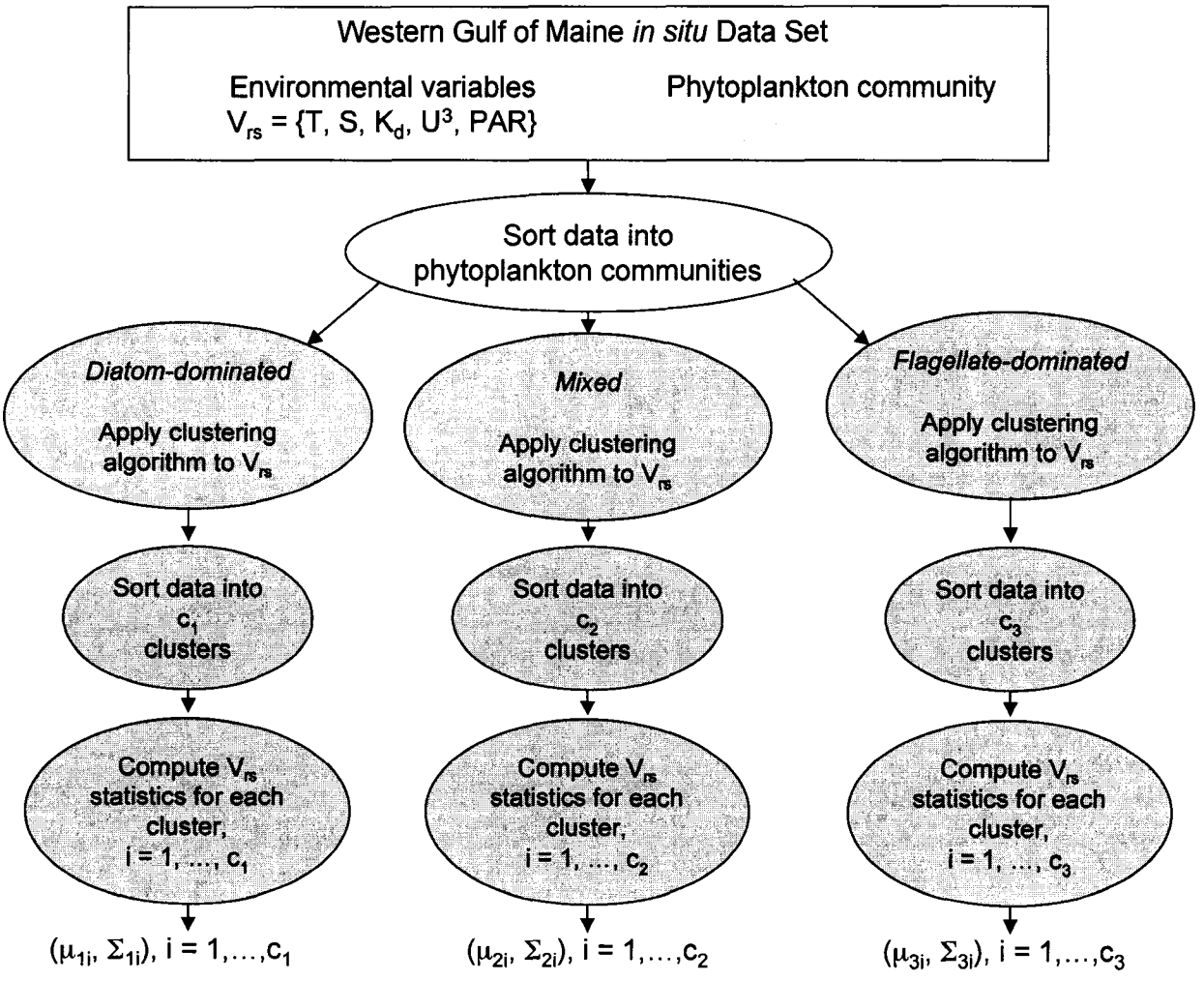



Figure 4-2: Schematic of classifier training.

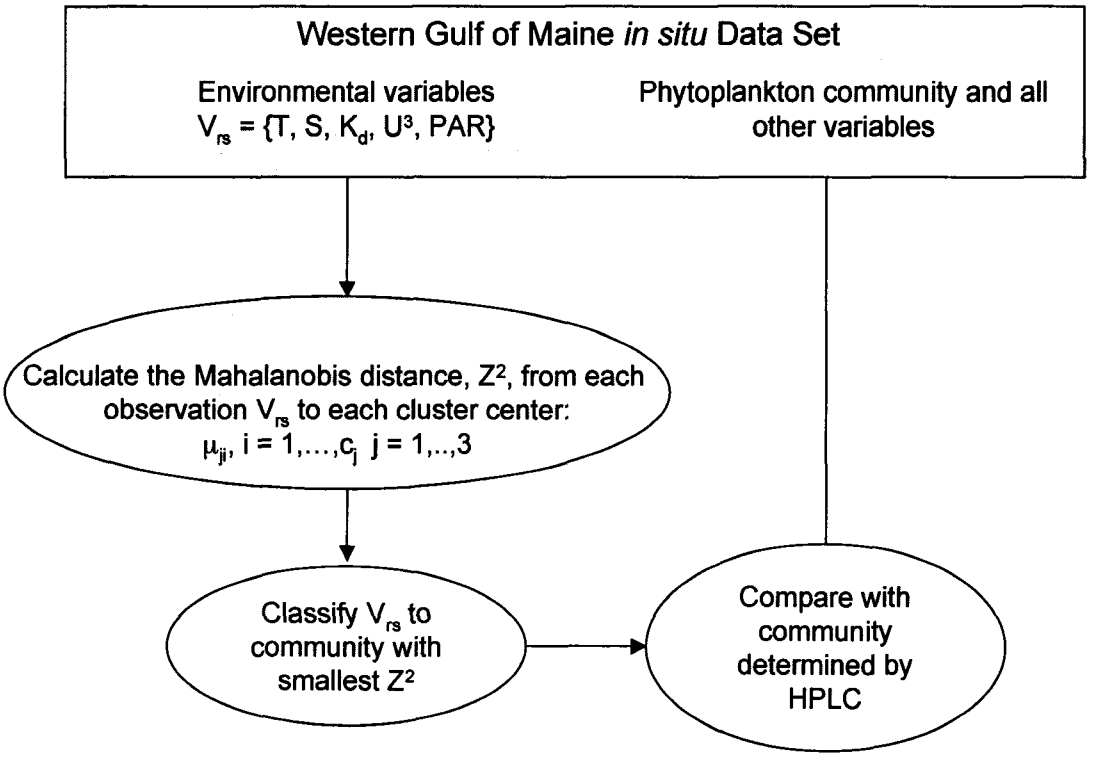



Figure 4-3: Schematic of the operation of the classifier as applied to *in situ* data.

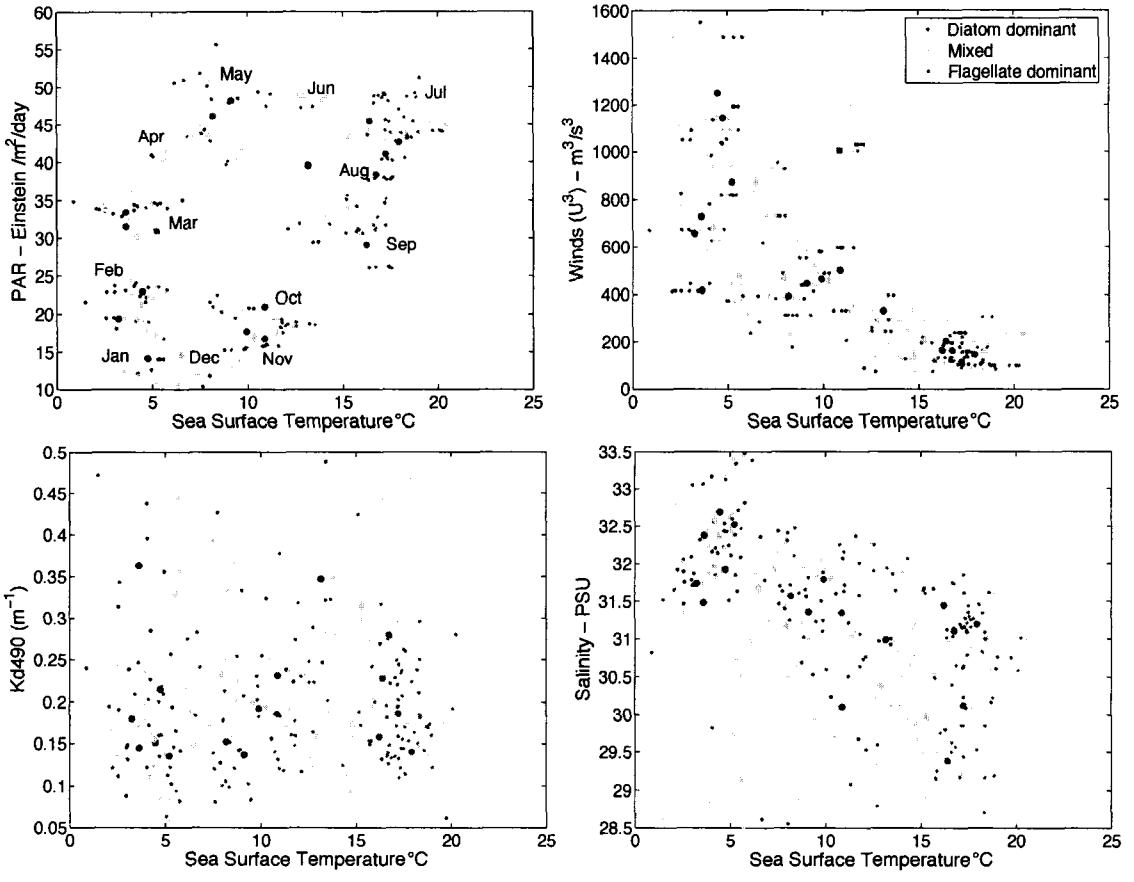



Figure 4-4: Distribution of physical data by environmental pairs color-coded by phytoplankton community: blue – *diatom-dominated*; green – *mixed*; red – *flagellate-dominated*. Larger points indicate location of the cluster centers ($c$). Top left: temperature-PAR; top right: temperature-wind strength; bottom left: temperature-$K_{d490}$; bottom right: temperature-salinity. Total N=255, total $c$ = 27; *diatom-dominated* N=51, $c$=8; *mixed* N = 94, $c$=10; *flagellate-dominated* N=110, $c$=9.

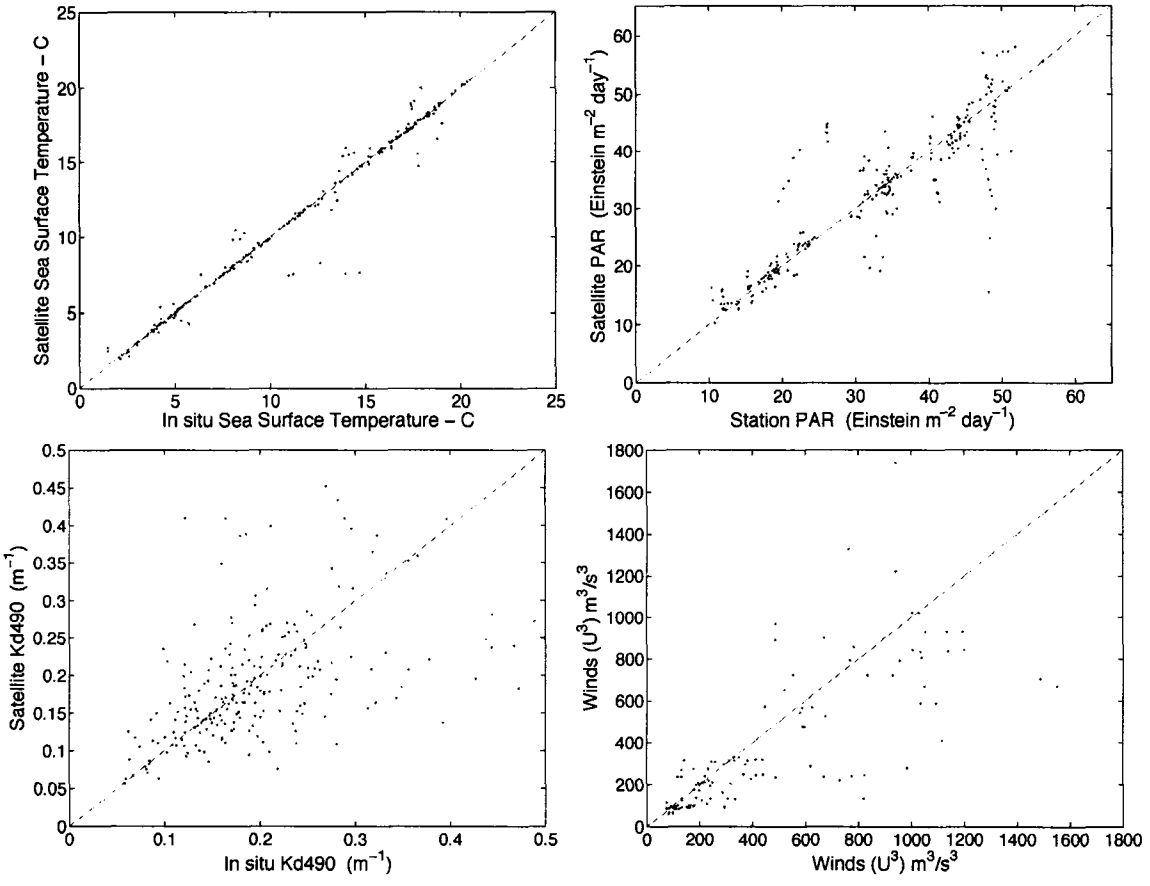



Figure 4-5: Satellite extractions versus *in situ* data. All satellite fields were obtained from images that were based on 8-day averages. The *in situ* fields were instantaneous measurements for temperature and $K_{d490}$, while PAR and winds were derived by averaging the preceding 8 days from the date of the measurement. This is a different 8-day window compared to the satellite 8-day averages, which were fixed according the calendar beginning on January 1.

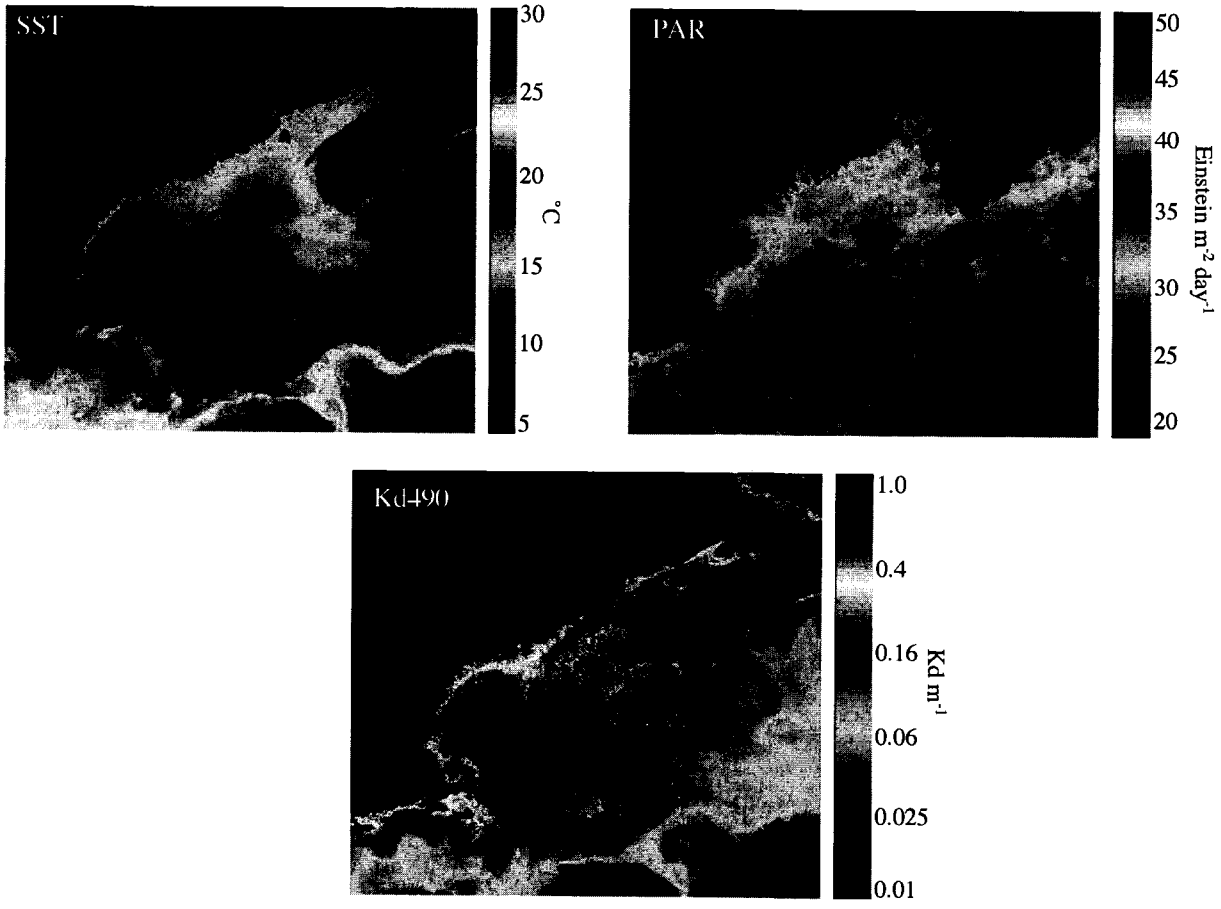



Figure 4-6: A set of SST (MODIS-Aqua), PAR (SeaWiFS), and Kd490 (SeaWiFS) 8-day composites for September 6 - 13, 2005 as inputs into the fuzzy classifier.

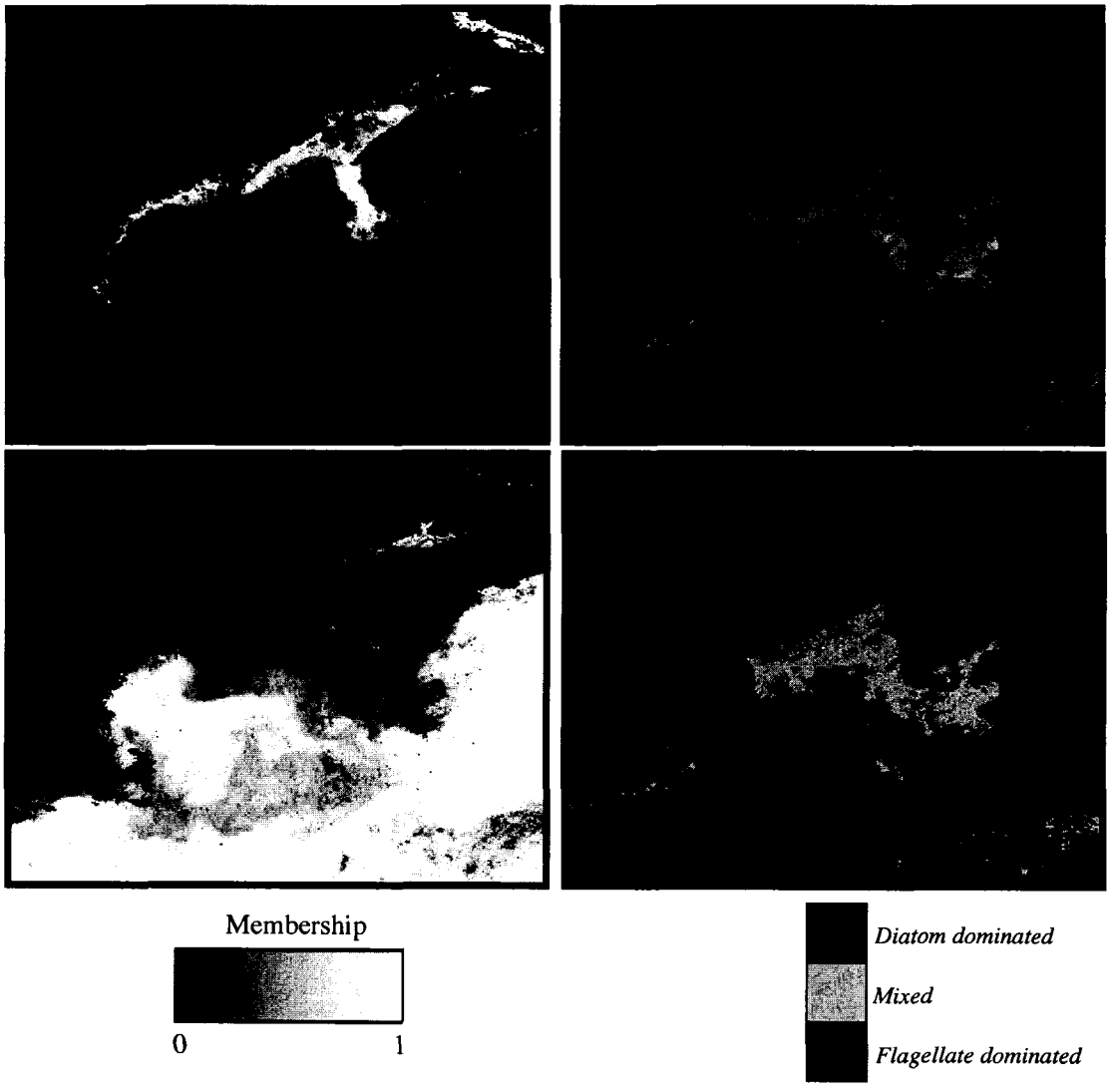



Figure 4-7: Fuzzy and hard membership maps derived from the set of SST, PAR, and Kd (scenario 3) shown in figure 4-6. The fuzzy memberships are probabilities of occurence, and sum to 1 for each pixel. The hard membership map was produced by assigning the pixel to the community with the highest fuzzy membership.

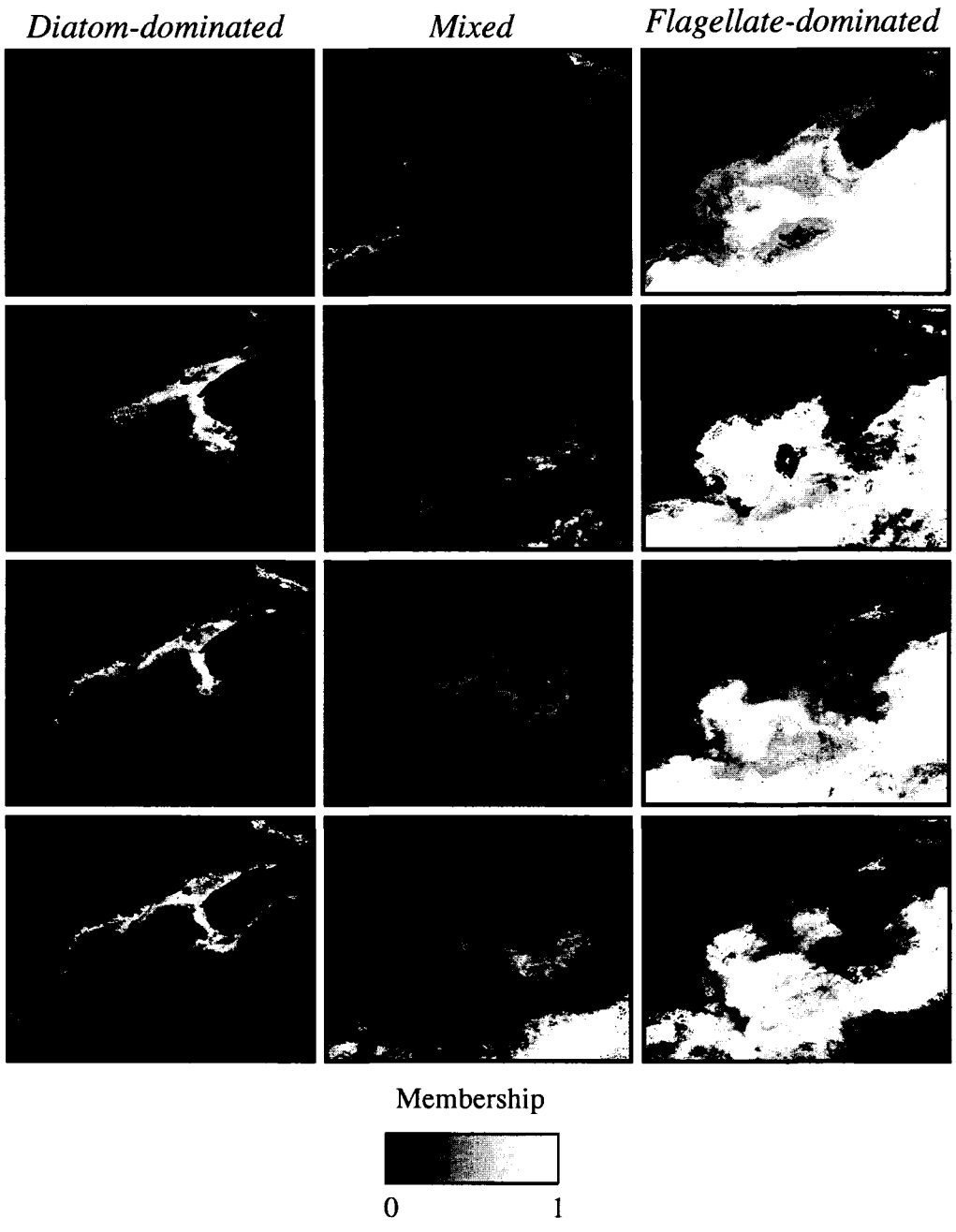



Figure 4-8: Fuzzy membership map sequence from September 6 through October 8, 2005. Left column –*diatom-dominated* memberships; middle column – *mixed*; right column – *flagellate-dominated* memberships. Each row represents the same time period as indicated in the dates shown in the *diatom-dominated* plots.

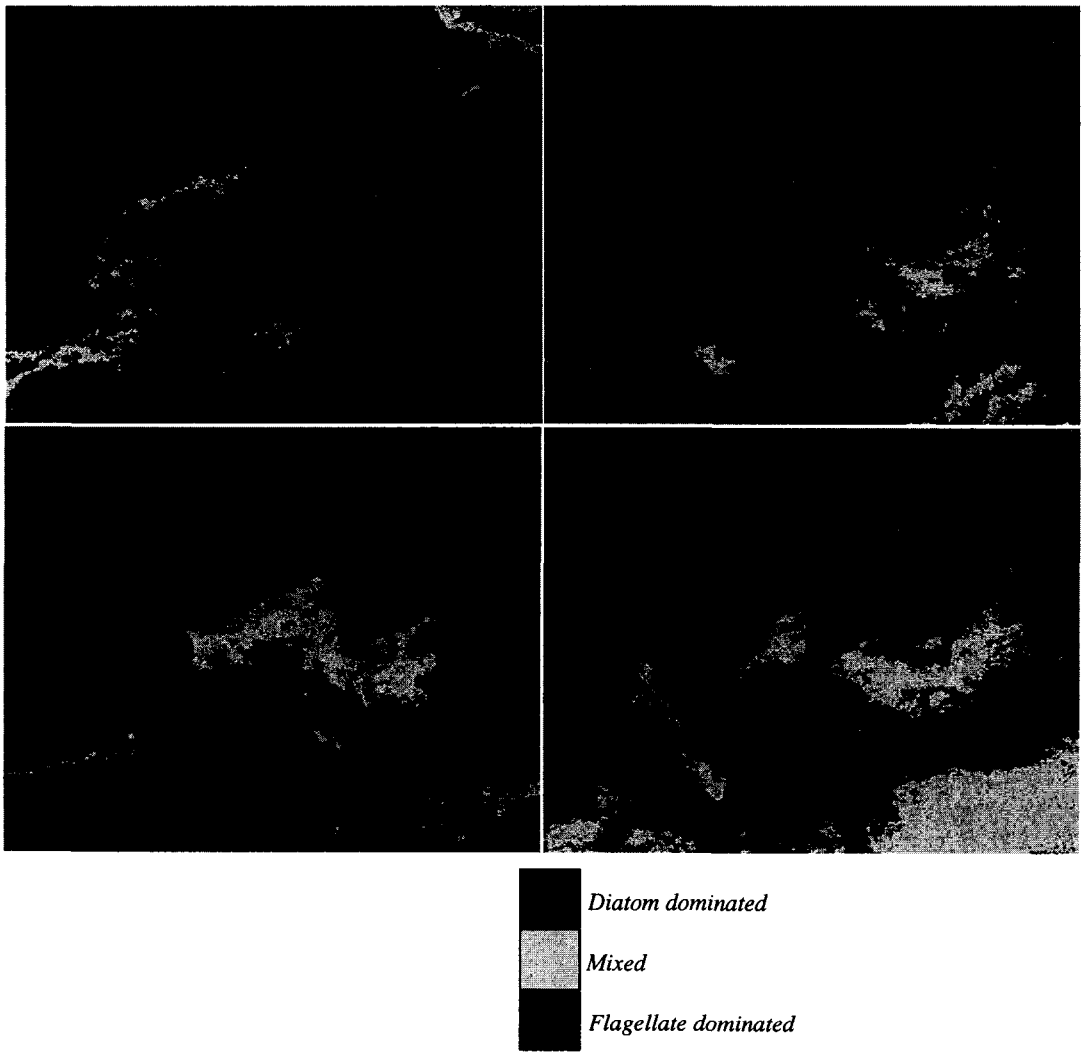



Figure 4-9: Phytoplankton community progression as depicted from the hard classification maps from September 6 - 30, 2005 based on scenario 3. Black area in upper right image is missing data (also seen in figure 4-8).

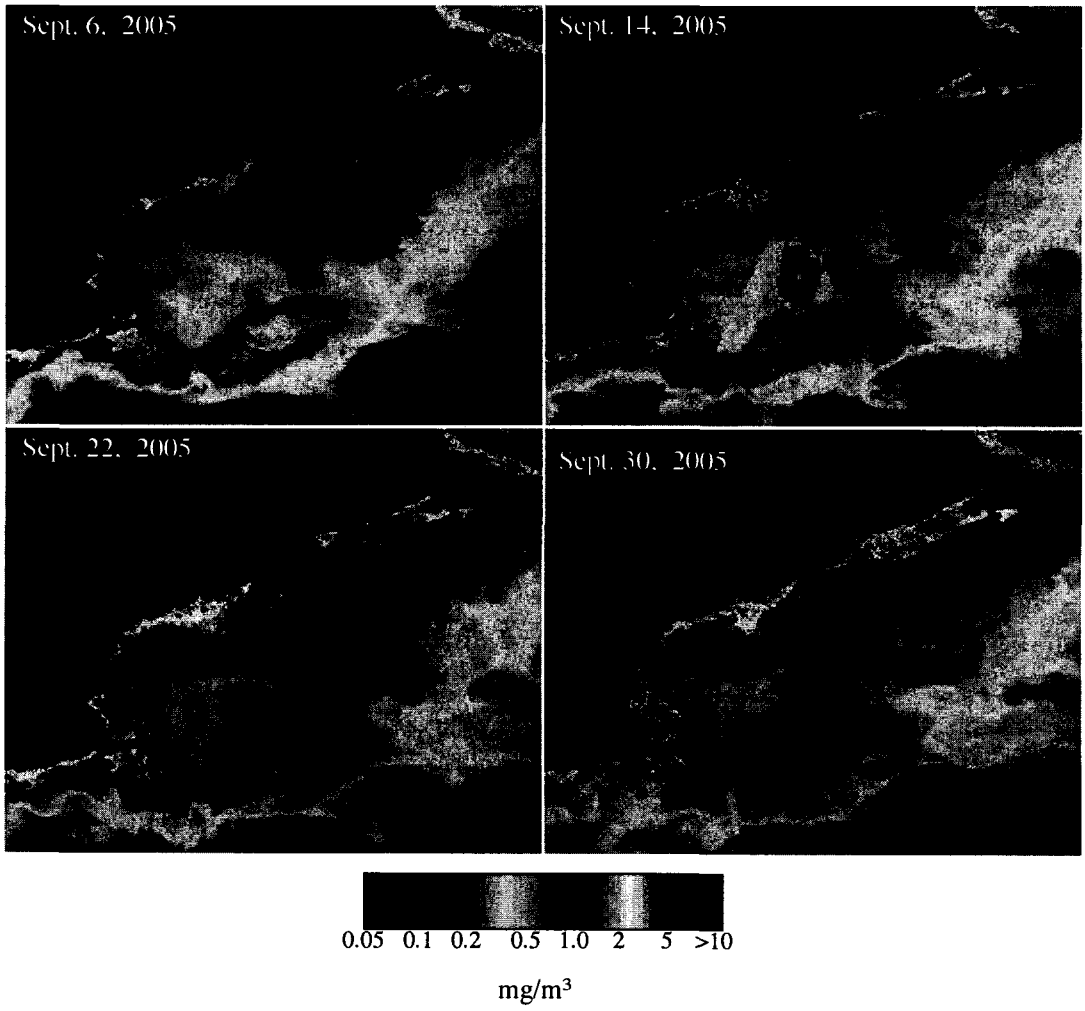



Figure 4-10: 8-day composites of chlorophyll a concentration from the SeaWiFS satellite.

# CHAPTER 5

# SUMMARY

## 5.1 Motivations

The fundamental question addressed in this dissertation was: can phytoplankton communities be mapped from satellite data? It is a scientific question that has implications for climate change and coastal water quality, and thus is also a societal question. Although the specific reasons that make this question relevant to different people vary according to the contextual application (e.g., marine ecosystem, biogeochemistry or marine monitoring), the central motivation is the scientific need to understand the diversity and geographical distribution of phytoplankton communities in marine habitats. While satellite remote sensing technologies have enabled the observation of phytoplankton biomass levels and their changes across regional and global spatial scales, the types of phytoplankton and the communities they form remain unknown and largely indistinguishable from space.

It is extremely difficult to map phytoplankton communities without satellite observation. Phytoplankton are largely invisible to the naked eye, and their turnover rates are on the order of 30 days compared to 30 years for terrestrial plants. In addition, phytoplankton inhabit a medium which has horizontal and vertical movement

that influences their distributions. The time and space scales needed to monitor phytoplankton changes pose different challenges than terrestrial systems. While direct analysis of water samples from field programs provides the most conclusive method of identifying the types of phytoplankton present in the water, the practical limitations of human and financial resources combined with accessibility of offshore sites (e.g., weather, remoteness) often inhibit the ability to collect samples at the necessary time and space scales required to capture the succession of phytoplankton communities. Satellites can overcome the problems of time and space coverage, but the problem of how to best use the data to detect different phytoplankton communities remains the pivotal challenge.

The approach of this research was to use satellite information to detect habitats that were linked to different phytoplankton communities. The basis of this approach was an empirical study focused on establishing statistical relationships between phytoplankton communities and the environment in which they were observed. By using variables that could be measured from space, it was then possible to apply the same phytoplankton-habitat relationships to the satellite data to create geographic maps of phytoplankton communities.

The three studies of the dissertation addressed different aspects of the main problem using a data set collected in the Gulf of Maine over a 3-year period. The first study examined the use of a chemotaxonomic method to quantify phytoplankton composition from pigment data. The level of composition was to the class level, and was supported by microscopic observations. The second study examined the cycles of environmental variables and the phytoplankton communities through the

study period. Using a principal component analysis on nine different environmental factors, the dominant mode of variance was identified as seasonal, and phytoplankton communities exhibited an affinity to different hydrographic regimes. The third study extended this finding to the construction of a classifier to predict phytoplankton communities from environmental variables. The major findings and conclusions of each study will be reviewed. These will be discussed in conjunction with the relevant ecological issues surrounding the central themes of the dissertation.

## 5.2 General considerations and conclusions

One of the basic issues relating to the central theme was how to define a phytoplankton community. There are multiple ways to partition phytoplankton – from a biogeochemical perspective (Hood et al., 2006) to a size-based fractionation (Vidussi et al., 2001). Hutchinson (1967) defined a phytoplankton *community* as *'a collection of species living together and usually linked to a particular habitat.'* This definition is closely related to the *association*, defined as *an assemblage of species that recurs under comparable ecological conditions in different places* (Hutchinson, 1967). Each definition is based at a species level of taxonomy and links them to their environment. Kruk et al. (2002) used a variant of these definitions to organize phytoplankton communities into 17 different groups in terms of morphological similarities which made them suited for different environments. In contrast, the biogeochemical view breaks down phytoplankton along functional lines, that is, according to their biogeochemical function. The number of functional groups is usually on the order of 4-6 different groups. A typical functional group listing includes diatoms, coccolithophores, flag-

ellates, and cyanobacteria. In this regard, all diatoms (which are taxonomically at the class level) are viewed as a single entity, as are coccolithophores.

With regards to the data collected in the present dissertation, taxonomic composition to the species level was available for a limited number of samples. However, HPLC samples were routinely collected, and composition to the class level was possible. The focus of study 1 was to characterize the phytoplankton community to the class level using HPLC pigments and the CHEMTAX method. The outcome of study 1 was the subdivision of phytoplankton samples into eight different classes. However, the validation of the method was problematic. Although cell counting techniques are the only alternative to quantifying relative community composition, comparisons between the two methods do not produce reliable metrics. This is due to the nature of the methods (carbon fractions versus chlorophyll fractions) and variability of the intracellular carbon:chlorophyll ratio needed to translate one quantity to the other. Results from this study focused instead on output differences between CHEMTAX simulations that were initialized with different parameters.

The initial pigment ratio table, which is used as the starting point in the iterative process of the CHEMTAX method, is frequently mentioned in the literature. These ratios should be based on species found in the study area, and also should reflect the appropriate physiological state of each class. Both require some prior knowledge of each sample. Microscopic analysis can identify the dominant species representing each class, but the physiological state is harder to ascertain. In study 1, light levels were used as a proxy for physiological state. The use of light-specific pigment ratios versus averaged ratios over different light regimes did not significantly alter

the outcome of CHEMTAX. Since species from the same class can have different pigments and different ratios, knowing the species is critical. Mis-representations that occurred in CHEMTAX output in different seasons could be attributed to inaccurate pigment ratios for certain classes. This was due to the unavailability of pigment ratios for the species present and the substitution of ratios for other species in the same class.

*Conclusion 1 – CHEMTAX is a robust method for determining taxonomic composition to the class level, but pigment ratios from species in the study area are needed.*

The eight classes generated from CHEMTAX were the basis of forming the eventual phytoplankton community definitions in studies 2 and 3. Thus, each community had to be based on a combination of these eight classes. However, the number of class combinations was large compared to the number of samples, and these had to be reduced by class re-combination to limit the myriad set of possible class combinations. Since many flagellate species share the same size range and are morphologically similar, all flagellate classes were combined. This reduced the eight classes to three groups – diatoms, flagellates, and cyanobacteria – that are aligned along functional group divisions. It was from these groups that the eventual phytoplankton 'communities' were defined.

The problem still remained of how to define communities from these three phytoplankton groups; boundaries needed to be specified. The question as to where does one community begin and another end is difficult to establish, as the real world is a continuum of graded transitions rather than readily defined compartments with discrete edges. Sathyendranath et al. (2004) based diatom communities on a fu-

coxanthin:chlorophyll *a* ratio of 0.4 or greater. This was based on cultured diatom fucoxanthin:chlorophyll *a* ratios, which generally are between 0.4 and 0.5. In this research, a diatom-dominated community was defined by a threshold of 70% or more of the chlorophyll *a* attributed to diatoms. The corresponding fucoxanthin:chlorophyll *a* ratio for this community ranged between 0.3 and 0.5. This highlights the difficulties with establishing thresholds for defining a diatom or any other type of community, since different criteria will produce different results. It is reasonable to say, however, that a phytoplankton community that has 70% or more of the biomass composed as diatoms is *diatom-dominated.* Likewise, a community in which 70% of the biomass is flagellates is conservatively *flagellate-dominated.* A so-called *mixed* community falls in between these, but these thresholds were arbitrary. Ultimately, the communities were subjectively defined by setting these thresholds.

The relationship between these communities and environmental factors was explored in study 2, and quantified into a predictor in study 3. A key question is: in a multivariate environment, how many variables is enough and which variables are significant? The choice of environmental factors was guided by variables suggested in Longhurst (2007) and others (Smayda, 1980; Harris, 1986). Study 2 examined nine different variables. The set of variables selected described various aspects of the environment, and included physical data (temperature, mixed layer depth, diffuse attenuation coefficient), meteorological data (wind speed and surface light intensity), chemical data (nutrient concentrations), and implicitly included correlated biotic factors, if any, such as grazing and competition. The focus of study 2 was to examine the cycles of these variables in conjunction with phytoplankton community cycles. A principal component analysis of the physical data qualitatively revealed

that *diatom-* and *flagellate-dominated* communities were associated with different physical environments.

*Conclusion 2 – Phytoplankton communities, defined at a broad level, exhibit affinities to different environmental habitats.*

The adaptation of the data set to the Reynolds Intaglio shows that diatom communities preferred environments with high nutrients and well-mixed water columns, whereas flagellate communities were prevalent in low-nutrient, highly stratified waters. These findings – although not unexpected – shed light on and document the phytoplankton community cycles in the western Gulf of Maine, which have not been addressed in any recent publication. The overall results of study 2 provided the qualitative evidence to seek the phytoplankton community-physical environment relationships in a quantified way, which was the focus of study 3.

In study 3, the relationships were established with a statistical model based on empirical distributions of the data. A classifier was developed, which took as input a set of physical data and predicted one of three phytoplankton communities. In this study, the set of physical variables was reduced from nine to five based on the preference for variables that are amenable to remote sensing.

The success rate (i.e., performance) of the classifier was determined by comparing the predicted phytoplankton community to the 'actual' community for each data point. This was done with the training data set, and a set of 100 test data sets randomly selected (10% of the data) and not used for training. Not surprisingly, the performance of the classifier was higher with the training data set (82%) than the test data sets (~55%). This decrease in performance is attributed to the change

in the statistical characteristics of the data set when a random subset (10%) was removed, indicating that the data set probably contained too few points. Most of the errors were between 'adjacent' phytoplankton groups; that is, it was more frequent for a *diatom-dominated* or a *flagellate-dominated* community to be predicted as a *mixed* community or vice versa. The mis-identification of a *diatom-dominated* community with a *flagellate-dominated* community (or vice versa) occurred less than 4% of the time with the training data set, and these were the so-called 'egregious' errors. Compared to other studies (Sathyendranath et al., 2004; Alvain et al., 2005), this method achieved higher success rates based on the same criteria (evaluating the training set).

*Conclusion 3a – Phytoplankton communities can be predicted based on environmental factors and success can range from 55% to 82%, depending on the data set. These represent potential upper and lower limits for community prediction.*

Errors progressively increased when environmental variables were removed, that is, when fewer than 5 variables were used as input. In addition, errors were higher when the input fields were satellite data, presumably as a result of 'noisy' satellite data. As a consequence, uncertainty increased in both input and output fields when the classifier was applied to satellite data. The best performance results when applied to satellite occurred when three fields were used as input – sea surface temperature, daily PAR, and $K_{d490}$ – and attained an overall success of 61% (*diatom-dominated* and *flagellate-dominated* community successes were 67% and 71%, respectively). Despite the overall decrease in successful classification, egregious errors remained low (~5%). Although the classifier had more trouble predicting the

true class with satellite data, it was still highly successful at differentiating *diatom-dominated* communities from *flagellate-dominated* communities.

The overall decline in performance when using satellite data reflects the increase in uncertainty in predicting phytoplankton communities from space. A method that has been designed to accommodate data imprecision and uncertainty – in this case errors associated with satellite data – is fuzzy logic. This method was adapted to the classifier and assigned probabilities to each community of its likelihood of occurrence. The confidence level in predicting the correct phytoplankton community now can be quantified and represented as maps (matching the satellite input data) depicting the probability of each community occurring. In this way, phytoplankton communities can be mapped along with the uncertainty imposed by the imprecision in the satellite data.

*Conclusion 3b – Satellite data can be used as input to a classifier to predict phytoplankton community distributions, but performance declines due to increased noise in the satellite data fields. A fuzzy logic approach is well suited to deal with data imprecision, and can supply confidence levels for the predicted communities.*

## 5.3 Problems in ecological prediction

There is a cascade of ecologically relevant factors which vary across different scales and influence the response of species on short-term physiological and longer-term ecological levels (Harris, 1986). Species are differentially affected by these factors, and it is their cumulative effect that decides the composition of the phytoplankton population at any given moment. The responses have inherent time

lags, and while physiological responses occur on the order of minutes to hours, the influence on the community composition is on the order of days. It is difficult to determine the state of a phytoplankton community vis-a-vis its response to its environment. The question arises as to whether the community is acclimated to its ambient environment, or is it in the process of changing?

Theoretically, a closed system in which the environment is constant will be in equilibrium (Harris, 1986). Time and history need not be considered since there are no time lags associated with the population. This situation might be comparable to a tropical environment that has limited variance in its environment and a relatively stable phytoplankton community with a constant overall biomass level. The other extreme is a non-equilibrium system which is dominated by environmental disturbances and disruption, and is subject to outside species immigration. The species in this type of community will respond differently according to their growth rates and the frequency of disturbances. This situation might be represented by the passage of a storm that mixes the water column, resets nutrient levels, and introduces new species. In the marine environment, both types of system occur at different times, although it is debatable whether equilibrium states ever truly exist at all. A governing factor which dictates which system prevails at any given moment is the frequency of disturbances and the community response time to such disturbances. Within a community, species will respond differently (i.e., in terms of their growth rate response) to the frequency of disturbance. Thus, characterizing such temporal fluctuations and the variance in properties such as $Z_{eu}/Z_m$ are important for community prediction (Harris, 1986).

Data interpretation is affected by the state of the phytoplankton community, its overall status in regard to the time lag response to its environment, and the frequency and timing of the sampling. A location sampled on a given day might have a different phytoplankton community just a few days later, even though the overall environmental conditions have not changed. Depending then on when the site is sampled, a different result may occur. This type of uncertainty is not related to measurement precision but to the frequency of sampling and the state of the community. Thus, the ability to accurately predict the phytoplankton community given knowledge of the physical environment is subject to these considerations.

*Conclusion 4 – Due to the inexactness of the knowledge inherent to the marine system, the prediction of phytoplankton communities will never be 100%. In this regard, a method based in fuzzy logic which deals with probability and not certainty is best suited to the problem of predicting phytoplankton communities, whether using environmental or some other factors as the basis of prediction.*

## 5.4 Final Conclusions

Predicting the phytoplankton community composition given the knowledge of the ocean environment will never be 100% accurate. Precision is compromised because there were many uncertainties and shortcomings associated with the measurements, ambiguity in defining communities, and with the general nature of the differential response of phytoplankton species to environmental change. The latter includes the time lags of response by phytoplankton to their environment on physiological (minutes/hours/days) and ecological (days/weeks/seasons) scales. It has been shown

that a classifier can be applied to oceanographic environmental data to predict the phytoplankton community composition with reasonable accuracy. There is greater success as more environmental information is included in the function.

The ultimate goal of this research was to predict phytoplankton community distributions using satellite remote sensing. The variability in the spatial and temporal distributions of phytoplankton communities require observational power only afforded through satellite data. The fundamental need for satellite data is based on the dual capabilities of their inherent spatial coverage and repeat cycles which cannot be matched by *in situ* observations. The classifier could be applied to satellite data, provided that the environmental variables used in the classifier match data available from satellite. Some of the key environmental variables are not mature or available yet as satellite products (e.g., salinity, high resolution winds). The classifier omitted these data in its application to satellites. Additionally, the satellite products generally have less precision than their *in situ* counterparts, and the overall effect is an increase in system noise and more uncertainty in the data. The combined effects of input data reduction and increased noise resulted in lower performance and greater uncertainty.

The fuzzy methodology is well suited for dealing with these types of uncertainties; it also has the capability to resolve transient and changing conditions over space and time natural to the marine environment. This dissertation reached the main goal and, moreover, demonstrated that improvements in the predictive power of the method can be achieved with increased precision and more advanced satellite-derived products. It also has highlighted the difficulties and challenges in resolving

phytoplankton community composition. To achieve better predictive accuracy in the future, it is important to maintain a strong field program in conjunction with satellite data to calibrate and validate evolving models. This includes the use of fixed buoys and drifters (e.g., ARGO floats) to augment field programs. These resources are now forming an expanding and increasingly important aspect of ocean observing. As new satellite sensors come on-line, the real-time information collected about the ocean will provide new possibilities to describe and understand marine environments. Regardless of which method is ultimately employed, any progress or future success in understanding phytoplankton community dynamics will require the integration of the complementary information from all these data sources.

# Bibliography

Alvain, S., Moulin, C., Dandonneau, Y., and Breon, F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep-Sea Research II*, 52:1989–2004.

Ansotegui, A., Sarobe, A., Trigueros, J. M., Urrutxurtu, I., and Orive, E. (2003). Size distribution of algal pigments and phytoplankton assemblages in a coastal-estuarine environment: contribution of small eukaryotic algae. *Journal of Plankton Research*, 25(4):341–355.

Archer, D. (1995). Upper ocean physics as relevant to ecosystem dynamics: a tutorial. *Ecological Applications*, 5:724–739.

Ban, S., Castel, J., Chaudron, Y., Christou, E., Escribano, R., Fonda Umani, S., Gasparini, S., Guerrero Ruiz, F., Hoffmeyer, M., Ianora, A., Kang, H., Laabir, M., Lacost, A., Miralto, A., Ning, X., Poulet, S., Rodriguez, V., Runge, J., Shi, J., Starr, M., Uye, S., and Wang, Y. (1997). The paradox of diatom-copepod interactions. *Marine Ecology Progress Series*, 157:287–293.

Banse, K. (1994). Grazing and zooplankton production as key controls of phytoplankton production in the open ocean. *Oceanography*, 7:13–20.

Barber, R. T. and Hiscock, M. R. (2006). A rising tide lifts all phytoplankton: Growth response of other phytoplankton taxa in diatom-dominated blooms. *Global Biogeochemical Cycles*, 20:10.1029/2006GB002726.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum, NY.

Bezdek, J. C. and Pal, N. R. (1993). Some new indexes of cluster validity. *IEEE Trans. Sys., Man, and Cybernet. (B - Cybernet.)*, 28:301–315.

Bouman, H. A., Platt, T., Sathyendrenath, S., Li, W. K. W., Stuart, V., Fuentes-Yaco, C., Maass, H., Horne, E. P. W., Ulloa, O., Lutz, V., and Kyewalyanga, M. (2003). Temperature as indicator of optical properties and community structure of marine phytoplankton: implications for remote sensing. *Marine Ecology Progress Series*, 258:19–30.

Brown, C. and Yoder, J. (1994). Distributions of coccolithophorid blooms in the western North Atlantic Ocean. *Continental Shelf Research*, 14:175–197.

Buchaca, T., Felip, M., and Catalan, J. (2005). A comparison of HPLC pigment analyses and biovolume estimates of phytoplankton groups in an oligotrophic lake. *Journal of Plankton Research*, 27(1):91–101.

Carder, K. L., Chen, F. R., Lee, Z. P., Hawes, S. K., and Kamykowski, D. (1999). Semianalytic moderate-resolution imaging spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. *Journal of Geophysical Research*, 104:5403–5421.

Claustre, H., Babin, M., Merien, D., Ras, J., and Prieur, L. (2005). Towards a taxon-specific parameterization of bio-optical models of primary production: A case study in the north atlantic. *Journal of Geophysical Research*, 110:10.1029/2004JC002634.

Cloern, J., Grenz, C., and Vidergar-Lucas, L. (1995). An empirical model of the phytoplankton chlorophyll:carbon ratio the conversion factor between productivity and growth rate. *Limnol. Oceanogr.*, 40:1313–1321.

Cloern, J. E. and Dufford, R. (2005). Phytoplankton community ecology: principles applied in San Francisco Bay. *Marine Ecology Progress Series*, 285:11–28.

Cullen, J. J., Franks, P. J. S., Karl, D. M., and Longhurst, A. (2002). *The sea*, chapter Physical influences on marine ecosystem dynamics, pages 297–336. John Wiley and Sons, Ltd.

Cullen, J. J. and Lewis, M. R. (1988). The kinetics of algal photoadaptation in the context of vertical mixing. *Journal of Plankton Research*, 10(5):1039–1063.

Davies, D. L. and Bouldin, D. W. (1979). A clustering separation measure. *IEEE Trans. Patt. Anal. Mach. Intel.*, 1:224–227.

Doney, S. C., Lindsay, K., and Moore, J. K. (2003). *Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change*, chapter Global ocean carbon cycle modeling. Springer-Verlag, Berlin.

Duarte, C., Agusti, S., Gasol, J., Vaquez, D., and Vazquez-Dominguez, E. (2000). Effect of nutrient supply on the biomass structure of planktonic communities: an experimental test on a Mediterranean coastal community. *Marine Ecology Progress Series*, 206:87–95.

Egge, J. K. and Aksnes, D. L. (1992). Silicate as regulating nutrient in phytoplankton competition. *Marine Ecology Progress Series*, 83:281–289.

Eppley, R. W. (1972). Temperature and phytoplankton growth in the sea. *"Fisheries Bulletin"*, 70:1063–1085.

Falkowski, P. F., Laws, E. A., Barber, R. T., and Murray, J. (2003). *Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change.*, chapter Phytoplankton and their role in primary, new, and export production., pages 99–121. Springer-Verlag, Berlin.

Falkowski, P. G. and Raven, J. A. (1997). *Aquatic Photosynthesis.* Blackwell Publications, Inc.

Frouin, R., Franz, B. A., and Werdell, P. J. (2003). *SeaWiFS Postlaunch Technical Report Series: Volume 22, Algorithm Updates for the Fourth SeaWiFS Data Reprocessing*, chapter Chapter 8: The SeaWiFS PAR Product. NASA.

Garibotti, I. A., Vernet, M., Kozlowski, W. A., and Ferrario, M. E. (2003). Composition and biomass of phytoplankton assemblages in coastal Antarctic waters: a comparison of chemotaxonomic and microscopic analyses. *Marine Ecology Progress Series*, 247:27–42.

Geider, R. (1987). Light and temperature dependence of the carbon to chlorphyll *a* ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton. *New Phytol.*, 106:1–34.

Geider, R., MacIntyre, H., and Kana, T. (1998). A dynamic regulatory model of phytoplanktonic acclimation of phytoplankton to light, nutrients and temperature. *Limnol. Oceanogr.*, 43:679–694.

Gieskes, W., Kraay, G., Nontji, A., and Sumoto, D. (1988). Monsoonal alternation

of a mixed and layered structure in the phytoplankton of the euphotic zone of the Banda Sea: mathematical analysis of algal pigment fingerprints. *Neth. J. Sea Res.*, 22:23–137.

Gin, K. Y. H. and Lee, Y. K. (2003). Phytoplankton community structure in singapore's coastal waters using hplc pigment analysis and flow cytometry. *Journal of Plankton Research*, 25(12):1507–1519.

Goericke, R. (1998). Response of phytoplankton community structure and taxon-specific growth rates to seasonally varying physical forcing in the Sargasso Sea off Bermuda. *Limnol. Oceanogr.*, 43:921–935.

Goldman, J. C. (1977). Temperature effects on phytoplankton growth in continuous culture. *LO*, 22:932–936.

Gran, H. H. (1932). Studies on the biology and chemistry of the gulf of maine. *Marine Ecology Progress Series*, 144:265–283.

Grover, J. P. and Chrzanowski, T. H. (2005). Seasonal dynamics of phytoplankton in two warm temperate reservoirs: association of taxonomic composition with temperature. *Journal of Plankton Research*, 27:1–17.

Harris, G. (1986). *Phytoplankton Ecology.* Chapman and Hall.

Havskum, H., Schluter, L., Scharek, R., Berdalet, E., and Jacquet, S. (2004). Routine quantification of phytoplankton groups - microscopy or pigment analyses? *Marine Ecology Progress Series*, 273:31–42.

Henriksen, P., Riemann, B., Kaas, H., Sorensen, H. M., and Sorensen, H. L. (2002).

Effects of nutrient-limitation and irradiance on marine phytoplankton pigments. *Journal of Plankton Research*, 24(9):835–858.

Hoffbeck, J. P. and Landgrebe, D. A. (1996). Covariance matrix estimation and classification with limited training data. *IEEE Trans. Patt. Anal. Mach. Intel.*, 18:763–767.

Hood, R. R., Laws, E. A., Armstrong, R. A., Bates, N. R., Brown, C. W., Carlson, C. A., Chei, F., Doney, S. C., Falkowski, P. G., Feeley, R. A., Fiedrichs, M. A. M., Landry, M. R., Moore, J. K., Nelson, D. M., Richardson, T. L., Salihoglu, B., Schartau, M., Toole, D. A., and Wiggert, J. D. (2006). Pelagic functional group modeling: Progress, challenges and prospects. *Deep-Sea Research II*, 53:459–512.

Hutchinson, G. (1967). *A Treatise on Limnology, Vol. II.* Wiley, New York.

Iglesias-Rodriguez, M. D., Brown, C. W., Doney, S. C., Kleypas, J., Kolber, D., Kolber, Z., Hayes, P. K., and Falkowski, P. G. (2002). Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids. *Global Biogeochemical Cycles*, 16:1100.

Kamykowski, D. and Zentara, S. J. (2003). Can phytoplankton community structure be inferred from satellite-derived sea surface temperature anomalies calculated relative to nitrate depletion temperatures? *Remote Sensing of Environment*, 86:444–457.

Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J., and Hebel, D. (1997). The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature*, 388:533–538.

Kruk, C., Mazzeo, N., Lacerot, G., and Reynolds, C. S. (2002). Classification schemes for phytoplankton: a local validation of a functional approach to the analysis of species temporal replacement. *Journal of Plankton Research*, 24(9):901–912.

Lee, Z., Carder, K. L., and Arnone, R. A. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, 41:5755–5772.

Legendre, L. and Legendre, P. (1998). *Numerical Ecology*. Elsevier Science Ltd.

Letelier, R. M., Bidigare, R. R., Hebel, D. V., Ondrusek, M., Winn, C. D., and Karl, D. M. (1993). Temporal variability of phytoplankton community structure based on pigment analysis. *Limnol. Oceanogr.*, 38(7):1420–1437.

Leterme, S. C., Edwards, M., Seuront, L., Attrill, M. J., Reid, P. C., and John, A. W. G. (2005). Decadal basin-scale changes in diatoms, dinoflagellates, and phytoplankton color across the North Atlantic. *Limnol. Oceanogr.*, 50:1244–1253.

Lewitus, A. J., White, D. L., Tymowski, R. G., Geesey, M. E., Hymel, S. N., and Noble, P. A. (2005). Adapting the chemtax method for assessing phytoplankton taxonomic composition in southeastern u.s. estuaries. *Estuaries*, 28(1):160–172.

Lillick, L. C. (1938). Preliminary report of the phytoplankton in the gulf of maine. *American Midland Naturalist*, 20(3):624–640.

Llewellyn, C. A., Fishwick, J. R., and Blackford, J. C. (2005). Phytoplankton community assemblage in the English Channel: a comparison using chlorphyll a

derived from HPLC-CHEMTAX and carbon derived from microscopy cell counts. *Journal of Plankton Research*, 27(1):103–119.

Longhurst, A. (1995). Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.*, 36:77–167.

Longhurst, A. (2007). *Ecological Geography of the Sea.* Academic Press, San Diego.

Mackey, M. D., Mackey, D. J., Higgins, H. W., and Wright, S. W. (1996). Chemtax - a program for estimating class abundance from chemical markers: application to hplc measurements of phytoplankton. *Marine Ecology Progress Series*, 144:265–283.

Manning, C. A. and Bucklin, A. (2005). Multivariate analysis of the copepod community of near-shore waters in the western Gulf of Maine. *MEPS*, 292:233–249.

Margalef, R. (1961). Distribucion ecologica y goegrafica de las especies del fitoplankton marino. *Investgacion Pesquera*, 19:81–101.

Margelef, R. (1978). Life forms of phytoplankton as survival alternatives in an unstable environment. *Oceanol. Acta*, 1:493–509.

McGillicuddy, D. J., Signell, R. P., Stock, C. A., Keafer, B. A., Keller, M. D., Hetland, R. D., and Anderson, D. M. (2003). A mechanism for offshore initiation of harmful algal blooms in the coastal Gulf of Maine. *Journal of Plankton Research*, 25:1131–1138.

Menden-Deuer, S. and Lessard, E. J. (2000). Carbon to volume relationships for dinoflagellates, diatoms and other protist plankton. *Limnol. Oceanogr.*, 45(3):569–579.

Montagnes, D. J. S. and Franklin, D. J. (2001). Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: reconsidering some paradigms. *Limnol. Oceanogr.*, 46(8):2008–2018.

Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., and Fung, I. Y. (2002). An intermediate complexity marine ecosystem model for the global domain. *Deep-Sea Research II*, 49:403–462.

Oliver, R. L., Whittington, J., Lorenz, Z., and Webster, I. T. (2003). The influence of vertical mixing on the photoinhibition of variable chlorophyll a fluorescence and its inclusion in a model of phytoplankton photosynthesis. *Journal of Plankton Research*, 25(9):1107–1129.

Olson, R., Chisholm, S., Zettler, E., and Armbrust, E. (1990). Pigments, size, and distribution of Synechococcus in the North Atlantic and Pacific Oceans. *Limnol. Oceanogr.*, 35:45–58.

Platt, T., Sathyendranath, S., Forget, M. H., White, G., Caverhill, C., Bouman, H., Devred, E., and Son, S. (2007). Operational mode estimation of primary production at large geographical scales. *Remote Sensing of Environment.*

Rencher, A. C. (1995). *Methods of Multivariate Statistics.* John Wiley and Sons, Ltd.

Reynolds, C. S. (1984). Phytoplankton periodicity: the interactions of form, function and environmental variability. *Freshwater Biology*, 14:111–142.

Reynolds, C. S., Dokulil, M., and Padisak, J. (2000). Understanding the assembly

of phytoplankton to the trophic spectrum: where are we now? *Hydrobiologia*, 424:147–152.

Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L., and Melo, S. (2002). Towards a functional classification of the freshwater phytoplankton. *Journal of Plankton Research*, 24(5):417–428.

Riegman, R. and Kraay, W. (2001). Phytoplankton community structure derived from HPLC analysis of pigments in the Faroe-Shetland Channel during summer 1999: the distribution of taxaonomic groups in relation to physical/chemical conditions in the photic zone. *Journal of Plankton Research*, 23(2):191–205.

Riley, G. A., Stommel, H., and Bumpus, D. F. (1949). Quantitative ecology of the plankton of the western North Atlantic. *Bull. Bing. Oceanogr. Coll.*, 12:1–169.

Rodriguez, J. T., Tintore, J., Allen, J. T., Blanco, J. M., Gomis, D., Reul, A., Ruiz, J., Rodriguez, V., Echevarria, F., and Jimenez-Gomez, F. (2001). Mesoscale vertical motion and the size structure of phytoplankton in the ocean. *Nature*, 410:360–363.

Rowan, K. S. (1978). *Photosynthetic manual.* UNESCO.

Rowan, K. S. (1989). *Photosynthetic pigments in algae.* Cambridge University Press.

Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C., and Maass, H. (2004). Discrimination of diatoms from other phytoplankton using ocean-colour data. *Marine Ecology Progress Series*, 272:59–68.

Schluter, L., Mohlenberg, F., Havskum, H., and Larsen, S. (2000). The use of phytoplankton pigments for identifying and quantifying phytoplankton groups in

coastal areas: testing the influence of light and nutrients on pigment/chlorophyll *a* ratios. *Marine Ecology Progress Series*, 201:107–120.

Semina, H. J. (1997). *Advances in Marine Biology: The Biogeography of the Oceans*, chapter An Outlline of the Geographical Distribution of Oceanic Phytoplankton, pages 528–560. Academic Press.

Sieburth, J., Smetacek, V., and Lenz, J. (1978). Pelagic ecosystem structure. heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol. Oceanogr.*, 23:1256–1263.

Smayda, T. J. (1958). Biogeographical studies of marine phytoplankton. *Oikos*, 9:158–191.

Smayda, T. J. (1980). *The physiological ecology of phytoplankton*, chapter Phytoplankton species succession, pages 493–570. Blackwell Scientific Publications.

Smayda, T. J. and Reynolds, C. S. (2001). Community assembly in marine phytoplankton: application of recent models to harmful dinoflagellate blooms. *Journal of Plankton Research*, 23:447–462.

Sommer, U. (1989). *Plankton ecology: succession in plankton communities*, chapter The role of competition for resources in phytoplankton succession, pages 57–106. Springer-Verlag, Berlin.

Staehr, P. A., Henriksen, P., and Markager, S. (2002). Photoacclimation of four marine phytoplankton species to irradiance and nutrient availability. *Marine Ecology Progress Series*, 238:47–59.

Strathmann, R. R. (1967). Estimating the organic content of phytoplankton from cell volume or plasma volume. *Limnol. Oceanogr.*, 12:411–418.

Subramaniam, A., Carpenter, E., and Falkowski, P. G. (1999). Bio-optical properties of the marine diazotrophic cyanobacteria *Trichodesmium spp.* II. A reflectance model for remote sensing. *Limnol. Oceanogr.*, 44:618–627.

Sun, J. and Liu, D. (2003). Geometric models for calculating cell biovolume and surface area for phytoplankton. *Journal of Plankton Research*, 25(11):47–59.

Tamigneaux, E., Mingelbier, M., and Legendre, L. (1997). Grazing by protists and seasonal changes in the size structure of protozooplankton and phytoplankton in a temperate nearshore environment (western Gulf of St. Lawrence, Canada. *Marine Ecology Progress Series*, 146:231–247.

Tillman, D. (1976). Ecological competition between algae: experimental confirmation of resource-based competition theory. *Science*, 192:463–465.

Tozzi, S., Schofield, O., and Falkowski, P. G. (2004). Historical climate change and ocean turbulence as selective agents for two key phytoplankton functional groups. *Marine Ecology Progress Series*, 274:123–132.

Van Heukelem, L. and Thomas, C. S. (2001). Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *Journal of Chromatography*, 910:31–49.

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., and Marty, J. C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation

in the eastern mediterranean sea during winter. *Journal of Geophysical Research*, 106(7):19,939–19,956.

Wilhelm, C., Rudolph, I., and Renner, W. (1991). A quantitative method based on HPLC-adided pigment analysis to monitor structure and dynamics of the phytoplankton asssemblages - a study from Lake Meerfelder Maar (Eifel, Germany). *Arch Hydrobiol*, 123:21–35.

Xie, X. L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Trans. Patt. Anal. Mach. Intel.*, 13:841–847.

Yentsch, C. and Campbell, J. (1991). Phytoplankton growth: perspectives gained by flow cytometry. *Journal of Plankton Research*, 13:83–108.

Zadeh, L. (1965). Fuzzy sets. *Inform. Cont.*, 8:338–353.