UNIVERSITY OF CALIFORNIA

SANTA CRUZ

# IDENTIFYING AND TRACKING EVOLVING WATER MASSES IN OPTICALLY COMPLEX AQUATIC ENVIRONMENTS

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

OCEAN SCIENCES

by

**Sherry L. Palacios**

March 2012

The Dissertation of Sherry L. Palacios
is approved:

_____
Professor Raphael M. Kudela, Chair

_____
Professor Christopher A. Edwards

_____
Professor G. Jason Smith

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Sherry L. Palacios

Identifying and tracking evolving water masses

in optically complex aquatic environments

Earth's climate is intimately associated with biogeochemical processes of the sea. Biological Oceanography explores mechanisms controlling carbon uptake by phytoplankton, carbon transfer through biogeochemical processes, and energy flow through ecosystems. Satellite Oceanography affords a synoptic view of the sea surface and reveals underlying physical, chemical, and biological processes. Since the advent of ocean color satellites in 1978, ocean color algorithms evolved from quantifying phytoplankton biomass to addressing more complex bio-optical and oceanographic problems: characterizing inherent optical properties of the water column, estimating primary productivity, and detecting water masses. Locating a water mass, tracking its changes, and discriminating its constituents using bio-optical algorithms are the three objectives of this dissertation. The first objective identifies the location of the Columbia River Plume (CRP) by using light absorption by chromophoric dissolved organic matter ($a_{CDOM}$) as an optical proxy for salinity. It relates *in situ* measurements of $a_{CDOM}$ to salinity using linear regression analysis, then computes "synthetic" salinity using MODIS-Aqua satellite imagery. The algorithm is robust at predicting salinity of the CRP on the Oregon and Washington shelf. The second objective identifies sub-mesoscale features within the CRP and tracks their

changes in space and time. It employs k-means clustering and discriminant function analysis to identify water types from bio-optical and environmental input variables using *in situ* and MODIS-Aqua satellite observations. The algorithm is robust at identifying features in satellite and mooring data, consistent with measured and modeled water masses in previous work. The third objective involves development of an optical model (PHYDOTax) that discriminates phytoplankton taxa contained within an algal bloom. A hyperspectral ocean color signature-library for known phytoplankton (dinoflagellates, diatoms, haptophytes, cryptophytes, chlorophytes, cyanophytes, and phycocyanin-containing eukaryotes) was developed and then PHDYOTax decomposed ocean color spectra for culture mixtures and field samples into constituent taxa. PHYDOTax is robust at discriminating phytoplankton taxa and is one of the first algorithms to distinguish dinoflagellates from diatoms in ocean color data. These algorithms are new tools for the oceanographic community to constrain the location of carbon uptake and transfer through space and time in the CRP, and to partition energy flow through different phytoplankton-taxon dominated ecosystems.

# DEDICATION

To my husband, Richard S. Pasetto, who helped make this possible.

ACKNOWLEDGEMENTS

A very special thanks is due to my advisor, Raphael M. Kudela, for his generosity of ideas, research opportunities, funding, and time. I appreciate that he entrusted me with a great deal of autonomy as a scientist. Dr. Kudela returned edits faster than anyone I have ever met, sometimes entire manuscripts in less than a day. I am grateful for his support and encouragement to expand my scope as a scientist to include unique teaching opportunities and technology development.

I am thankful for my committee members, Christopher A. Edwards and G. Jason Smith for their support, wise counsel, and patience. I thank Kenneth Bruland for serving on my qualifying exam committee.

I thank my colleagues and friends for fruitful scientific conversations, moral support, review of papers, and afternoon walks: Tawnya D. Peterson, Kendra Hayashi, Clarissa Anderson, Andrea Joy VanderWoude, Shelly Benoit, Meredith Armstrong Howard, Petra Dekens, Olivia Cheriton, Rozalind Jester, Tyler Johnson, Kristen Buck, Maeve Lohan, Ana Aguilar-Islas, Leslie Roland Germain, Jason Hilton, Ryan Paerl, Melissa “Misty” Blakely Peacock, Anna McGaraghan, and Emily Schaller.

I appreciate the assistance I received from the administrative staff of the Ocean Sciences Department (UCSC). Diana Austin and Meyo Lopez helped guide me through the system on numerous occasions and I am thankful for their help.

Special thanks goes to the captains and crew of the three research vessels used for this dissertation: *R/V Point Sur, R/V Wecoma*, and *R/V John H. Martin*. The marine technicians, Stewart Lamerdin and Ben Jokinen of the *R/V Point Sur* and Daryl Swensen and Dave O'Gorman of the *R/V Wecoma* frequently helped me diagnose and solve myriad instrument troubles.

The text of this dissertation includes a reprint from the following published material:

> Palacios, S. L., T. D. Peterson, and R. M. Kudela (2009), Development of synthetic salinity from remote sensing for the Columbia River plume, J. Geophys. Res., 114, C00B05, doi:10.1029/2008JC004895.

Raphael M. Kudela directed and supervised the research, which forms the basis for the dissertation. As a result, he is listed as co-author for this reprint and will be co-author for the second and third chapters when they are published. The second co-author, Tawnya D. Peterson, assisted with method development and data collection for this published material, but otherwise the body of work was substantiatively my own.

The second and third chapters are being submitted for publication in the first half of 2012. I am very grateful to my four co-authors on the third chapter: Heidi M. Sosik, Tawnya D. Peterson, Kendra Hayashi, and Mike Jacox. Heidi Sosik and Kendra Hayashi provided phytoplankton cell-count data used to validate the algorithm in

Chapter 3. Mike Jacox provided helpful insight on the subtleties of the matrix algebra used in the algorithm for Chapter 3. Tawnya D. Peterson assisted with data collection for chapters 2 and 3. While the co-authors provided valuable assistance, the work was substantiatively my own.

This research was funded by grants from NSF (OCE-0238347), NASA (NNX09AT01G-01-B), NOAA GOES-R COAST project, NESDIS-NASA UARC Award (TO.030.6.MD.D-Kudela), UC-Santa Cruz Chancellor's Dissertation Year Fellowship, Dr. Earl H. Myers and Ethel M. Myers Oceanographic and Marine Biology Trust, and a travel grant from the Graduate Student Association (UCSC).

Finally, I thank my family for their love, support, and patience: Richard S. Pasetto, Gloria Palacios, Nancy Palacios Brooks, Monica Palacios-Boyce, Richard A. Pasetto, Marcella Pasetto, Christopher Pasetto, Juli Pasetto, and my sixteen nieces and nephews.

# INTRODUCTION

*No water, no life. No blue, no green.* --Sylvia Earle

## Ocean Biogeochemistry

The bio-optical properties of the aquatic environment are a reflection of the biogeochemical processes occurring within the medium. Biogeochemistry is the interdependence of the physical, chemical, geological, and biological realms that drive carbon fixation by phytoplankton, cycling of nutrients, flow of fixed carbon through ecosystems, and ultimate delivery of that carbon as sediment deposited on the sea floor. An over-arching theme in oceanography is to attempt to quantify and describe biogeochemical processes in order to understand carbon flow through the ocean biosphere. Ocean bio-optics attempts to relate the color of the sea to the constituents and processes in the upper water column using bio-optical algorithms so that observations of ocean biogeochemistry can be made over the large spatial scales and fine temporal scales afforded by satellites. Observations of these properties or processes can be used to understand the uptake of carbon through photosynthesis, the location of water masses, and the residence time of water masses. These types of observations can also be used to validate geophysical fluid dynamic models and ecosystem models. An understanding of the location of a water mass, its changing structure in space and time, and the constituents contained within it contribute to estimates of primary productivity in the ocean, transfer of carbon through ecosystems,

and transport of carbon through biogeochemical processes – all of which affect climate.

## Satellite and Airborne Sensors

The ocean is a vast, dynamic, and difficult to observe environment. As a result it is traditionally under-sampled in both space and time using mooring or shipboard measurements alone. Satellite and airborne sensors complement the incomplete record collected by more traditional means. The polar-orbiting satellite, MODIS-Aqua, collects near-daily observations of the surface of the Earth at a nominal resolution of one kilometer per pixel at nadir. Of the 36 spectral bands of information collected by the sensor, eight are used for ocean color study and can be further refined to 250 meter per pixel resolution through a cubic-spline interpolation scheme [*Franz et al.*, 2006] of three 250 m resolution bands. At either 1 km or 250 m resolution, and with near-daily overpasses of wide geographic swaths, MODIS is an improvement in temporal and spatial resolution over traditional sampling techniques. Airplane mounted sensors are limited temporally by research expeditions and weather conditions, but have greater spatial resolution than most civilian satellites at scales of tens of meters per pixel. Airborne sensors also often collect observations at higher spectral resolution than the standard ocean color satellites. The greater degrees of freedom afforded by more spectral bands allows for increased computational and statistical complexity in ocean color algorithms. As a result, airborne imaging can be used to validate existing satellite sensors, develop algorithms for them, and as models to develop satellites with higher spectral resolution.

Since the launch of the Nimbus-7 satellite in 1978, with the first ocean color sensor – the Coastal Zone Color Scanner (CZCS) [*Gordon et al.*, 1983], several generations of satellites have been used by the ocean color community. Some commonly used satellite sensors include the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), the Moderate Resolution Imaging Spectroradiometer (MODIS), the MEdium Resolution Imaging Spectrometer (MERIS), Visible Infrared Imager/Radiometer Suite (VIIRS), and the Hyperion Imaging Spectrometer. These include multi-spectral sensors, those collecting imaging data from a few discrete wavebands in the visible/near infrared range (e.g. eight for MODIS), and hyperspectral sensors collecting imagery in the visible/near infrared range at much higher spectral resolution (e.g. seventy for Hyperion). The MODIS-Aqua satellite sensor is used for the first two chapters of this dissertation. The third chapter uses an airplane-mounted sensor named the Spectroscopic Aerial Mapper with On-board Navigation (SAMSON), which measures ocean color at 256 wavebands in the visible/near infrared [*Davis and Bissett*, 2007].

Ocean Bio-optics

Make a casual observation of the sea and the surface appears blue for most of the world's ocean. The magnitude of light incident on the Earth's surface is greatest in the visible range of the electromagnetic spectrum (wavelengths: 400 – 700 nm). As light travels from the sun, through the atmosphere, down into the water column, and then is backscattered upward and out of the sea surface, it interacts with scattering and absorbing constituents [*Kirk*, 1994]. Those constituents change the magnitude of

light distributed across the spectrum so that the fraction re-emitted from the surface has a different color from the light that was incident on the surface. The eye absorbs the light emitted from the sea and the visual perception system interprets it as a blue or blue-green color.

The human eye evolved to absorb light most effectively at 555 nm, and like this biological model, other biological systems have evolved to capture light in the visible part of the spectrum. The system with the greatest biological impact is the chloroplast, where photosynthesis occurs in eukaryotic phytoplankton and vascular plants. Photosynthesis harnesses light energy and converts it to chemical energy. The key photo-reactive pigment for photosynthesis is chlorophyll-a, which absorbs light strongly at 440 nm, 680 nm, and 700 nm. Energy yielded during the light reactions fuels the biosynthesis of organic carbon in the Calvin cycle of the dark reactions. Uptake of carbon dioxide during photosynthesis, and conversion to biologically available, organic forms is the primary production that fuels global ecosystems. This process of carbon uptake by phytoplankton is a necessary step in the carbon cycle that helps to balance Earth's atmospheric carbon concentrations.

Almost all phytoplankton contain chlorophyll-a. The pigment has a direct relationship to phytoplankton biomass, and so the presence of the pigment can be used as a proxy for phytoplankton biomass in remote sensing imagery. Satellite remote sensing of chlorophyll-a can therefore be used to derive global estimates of phytoplankton standing stocks, and combined with sea surface temperature and incident solar irradiance can be used to estimate primary productivity, and thus

carbon export to sediments, carbon transfer within ecosystems, and other biogeochemical processes. By measuring this one optically active constituent of the sea, enormous progress has been made in quantifying and understanding the Earth's carbon cycle.

Like the human eye, an imaging spectrometer on an airplane or satellite detects the water leaving radiance, or light emitted from the sea surface. Ocean color retrievals from these sensors can be used to estimate chlorophyll-a concentrations. Chlorophyll-a differentially absorbs light across the visible spectrum, which results in characteristic peaks and troughs in the water leaving radiance. The most commonly used chlorophyll-a algorithms use the ratio of two magnitudes from this spectrum as an optical proxy for chlorophyll-a concentration (e.g. [*O'Reilly et al.*, 1998]). In empirical algorithms, chlorophyll-a estimates from satellite retrievals are matched in space and time to *in situ* measurements of chlorophyll-a and a relationship is computed using a linear or polynomial best fit. The equation from this fit is used to estimate chlorophyll-a concentration from satellite observations.

Sensing of chlorophyll-a using an empirical algorithm is just one example of how a biogeochemically important constituent can be measured using remote sensing imagery. Other constituents such as sediment and chromophoric dissolved organic matter (CDOM) can be measured using existing bio-optical algorithms. Semi-analytical and analytical models that rely on first principals of bio-optics have been developed since the launch of the first satellites to estimate a variety of oceanographically meaningful properties such as the inherent optical properties, light

absorption and backscattering [*Lee et al.*, 1994; *Maritorena et al.*, 2002]. An advantage of semi-analytical and analytical algorithms over empirical models is that they may describe a wider range of possible outcomes than may be available in the observed environment used to build the empirical model. This dissertation employs both an empirical (Chapter 1) and semi-analytical (Chapters 2 and 3) approach to algorithm development [*Palacios et al.*, 2009; *Palacios et al.*, in prep-a; *Palacios et al.*, in prep-b].

Unifying Bio-optical Equation

One fundamental, unifying equation ties together the bio-optical algorithms developed in this dissertation. It is the equation for remote sensing reflectance, which describes the relationship of inherent optical properties (IOP) of the constituents within the water column with the color of the light emitted from the sea surface [*Mobley*, 1994]. Using radiative transfer theory, remote sensing reflectance, $R_{rs}$, can be defined in terms of the inherent and apparent optical properties (AOP) of an optically deep and vertically homogeneous water column (Equation 1):

$$R_{rs}(\lambda) = \frac{t^2}{n^2} \frac{\mathrm{f}}{\mathrm{Q}(\lambda)} \frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)} \quad (1)$$

where λ is wavelength, t is the transmittance across the air-seawater interface, n is the index of refraction of seawater, f is a function of the solar zenith angle, Q(λ) is the upwelling irradiance-to-radiance ratio, $b_b(\lambda)$ is the spectral backscattering coefficient,

and $a(\lambda)$ is the total spectral absorption coefficient [*Lee et al.*, 1994]. Equation 1 can be simplified to (Equation 2):

$$R_{rs}(\lambda) = C\,\frac{b_b(\lambda)}{a\,(\lambda)+b_b(\lambda)} \tag{2}$$

where C is an approximation to account for transmittance across the air-sea interface, the index of refraction, the solar zenith angle, and the upwelling irradiance-to-radiance ratio.

The spectral light absorption coefficient can be further decomposed to (Equation 3):

$$a\,(\lambda) = a_{\mathrm{w}}\,(\lambda) + a_{\mathrm{NAP}}\,(\lambda) + a_{\mathrm{ph}}\,(\lambda) + a_{\mathrm{CDOM}}\,(\lambda) \tag{3}$$

where the subscripts w, NAP, ph, and CDOM refer to: water, non-algal particles (detritus), phytoplankton, and chromophoric dissolved organic matter.

The optical scattering coefficient can be described by (Equation 4):

$$b\,(\lambda) = b_{\mathrm{fw}}(\lambda) + b_{\mathrm{fp}}(\lambda) + b_{\mathrm{bw}}(\lambda) + b_{\mathrm{bp}}(\lambda) \tag{4}$$

where the subscripts f and b refer to light scattered in the forward, f, and backward, b, directions, w and p refer to light scattered by water and particles. For this dissertation, the inherent optical properties of interest include $a_{\mathrm{CDOM}}$, $a$, and $b_{\mathrm{bp}}$. These IOP's are the dominant variables in defining river plumes as a water mass ($a_{\mathrm{CDOM}}$),

characterizing the sub-mesoscale features of the river plume ($a_{CDOM}$), and for discriminating among phytoplankton taxa (IOP's: $a$, $b_{bp}$; AOP: $R_{rs}$).

## Dissertation Objectives

The goal of this dissertation is to develop bio-optical algorithms that address three general questions related to biogeochemistry: What is the location of the water mass? How is the water mass changing in space and time? What constituents are contained within the water mass? Specifically, the first objective is to develop an optical proxy for the low salinity water of the Columbia River Plume (CRP) in order to detect the plume water mass in satellite imagery of the Oregon and Washington shelf. Salinity cannot presently be sensed at the fine spatial and temporal scales needed to understand the CRP's dynamic processes using direct satellite measurements, but ocean color satellites do have that resolution. The second objective is to develop a statistical clustering technique to identify sub-mesoscale features within the plume in space and over a time series of days in order to track changes within the CRP related to changes in biogeochemical processes such as mixing and phytoplankton growth. The third objective is to develop a phytoplankton taxon discriminator in order to partition the chlorophyll pool of an algal bloom into its resident algal taxa.

These objectives are addressed in the three chapters of this dissertation:

Chapter 1: Development of synthetic salinity from remote sensing for the Columbia River Plume

Chapter 2: Optical characterization of water masses within the Columbia River Plume

Chapter 3: Discrimination of phytoplankton taxa in an optically complex aquatic environment

Study Locations

1. Columbia River Plume

The Columbia River drains a watershed of approximately 600,000 $km^2$ of the northwest United States and southwest Canada. The river flows into the Pacific Ocean (Fig. 1) and forms a buoyant plume, which remains a coherent structure, or water mass, until mixing incorporates it with shelf water. The plume delivers fresh water, nutrients, phytoplankton, and pollutants to the Oregon and Washington shelf. It is an important source of silicic acid to this highly productive region that supports an extensive salmonid fishery. The CRP is a dynamic feature that is governed by physical processes nearshore, including wind forcing [*Hickey*, 1989]. Depending on the direction of wind flow, the plume tends southward and offshore during upwelling winds and northward during relaxation or downwelling-favorable winds. The plume waters can be observed in both the northward and southward direction at the same time because of the presence of aged plume waters on the shelf. Little of this dynamic nature of the plume is captured with the limited number of moorings in the area, which includes just three moorings in the vicinity of the plume off Oregon and

Washington. Recently, more moorings inside the Columbia River Estuary and near the mouth have been installed, but the shelf continues to remain undersampled. The motivation for the first chapter was to develop an algorithm that could identify the location of the CRP over a wide spatial scale not possible with the existing moorings.

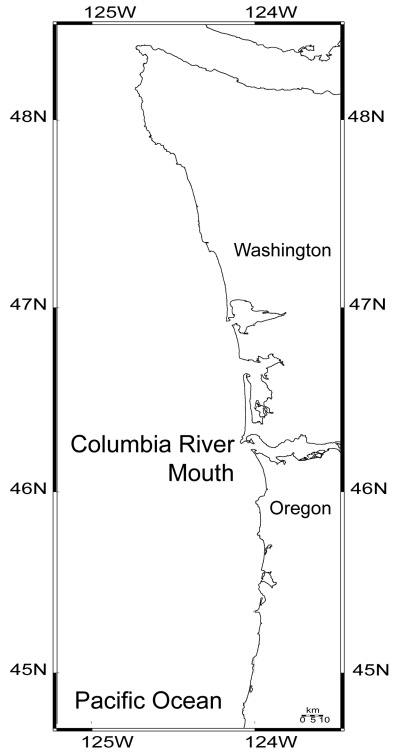



Fig. 1. Study Location – Chapters 1 and 2. The Columbia River Plume is located on the Oregon and Washington shelf and can be found northward or southward from the mouth of the river.

Physical processes, such as mixing and winds, dominate the behavior and structure of the plume near to shore (< 100 km). Smaller, or sub-mesoscale, features within the plume water mass have been observed and modeled for the CRP. These features affect the retention of water on the shelf and play an important role in

primary productivity of the shelf and offshore regions. This includes a bulge feature at the river mouth that acts as a retentive bio-reactor and also as a barrier to southward flowing water from the Washington shelf [*Horner-Devine*, 2009; *Kudela et al.*, 2010]. In addition, dipole eddy features have been observed and modeled for the shelf region and these are likely the result of bottom topography [*Banas et al.*, 2009]. These eddies influence retention and cross-shelf transport of material to the California Current System. The ability to identify and characterize these sub-mesoscale features within the larger CRP water mass was the motivation for chapter 2 of the dissertation. The approach taken was a statistical classifier of water types that has been applied in other regions of the world and for non-river water masses [*Martin Traykovski and Sosik*, 2003; *Oliver et al.*, 2004]. The types of possible observations from this method could be used to validate geophysical models of the plume that predict the presence and location of sub-mesoscale features in both space and time.

2. Monterey Bay – Algal Bloom

The Monterey Bay is an open bay on the central California coast (Fig. 2). It lies at the margin of the California Current System, an eastern boundary current. Typical of an eastern boundary current region, Monterey Bay experiences wind-driven coastal upwelling, which fuels high levels of primary productivity. The upwelling season (approximately March – August) tends to be dominated by the diatom phytoplankton group. During extended periods of wind relaxation, the dinoflagellates tend to proliferate where conditions are calm and favorable for growth. Though there is a seasonal component to the dominance of diatoms or

dinoflagellates in the bay, both taxa can co-occur at the same time and location. These two phytoplankton groups are represented most in phytoplankton samples in the bay, but other taxa (e.g. haptophytes, cryptophytes, cyanophytes, and chlorophytes) can be present.

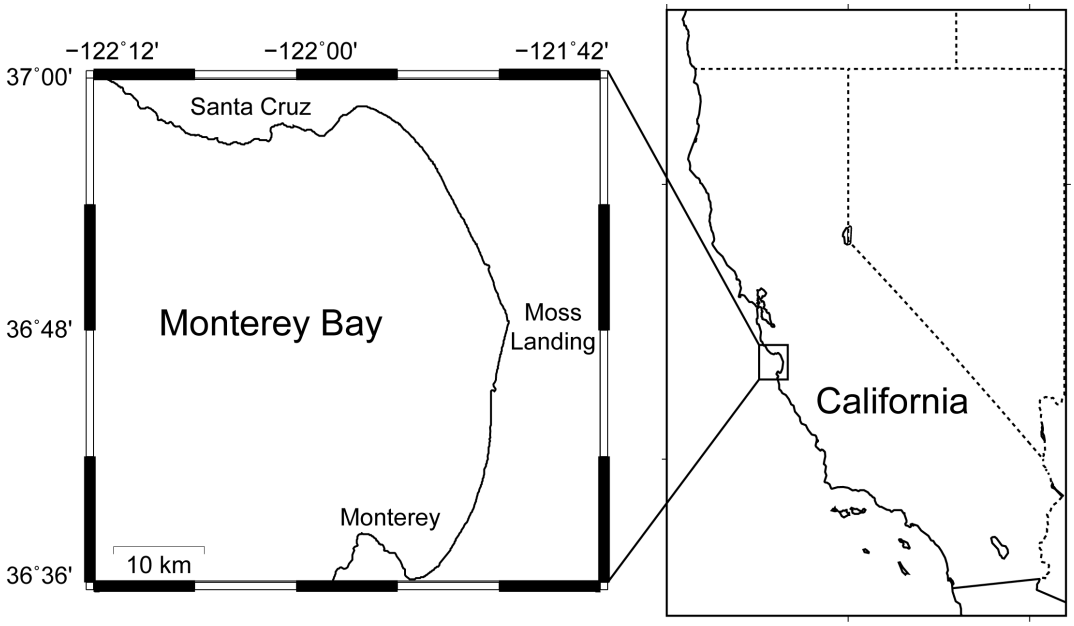



Fig. 2. Study Location – Chapter 3. The Monterey Bay.

Like grasses supporting a bison-dominated ecosystem and forbs supporting elk, the phytoplankton groups of the ocean support a variety of different ecosystems with their representative dominant taxa (e.g. salmon, mackerel, cnidarians). Identifying which phytoplankton taxa are present can be useful to partition carbon flows to different ecosystems. Collecting whole water samples and enumerating cells using microscopes or imaging flow-through systems is historically the method used to identify phytoplankton taxa. Until the present study, it was not possible to

differentiate between diatoms and dinoflagellates using an ocean color algorithm with remotely sensed data. One motivation for this study was to develop a bio-optical algorithm to discriminate phytoplankton taxa using ocean color data for use in ecosystem models and to validate plankton functional type models. The second motivation was to develop an algorithm to monitor for harmful algae using airborne or satellite imagery. Traditional sampling techniques (e.g. collecting whole water at fixed stations) are episodic and often costly. Remote detection can be used to improve spatial sampling and guide adaptive sampling decisions during a bloom event. The phytoplankton discrimination algorithm developed for this dissertation may improve: sampling of harmful algal blooms, estimates of carbon flow through different phytoplankton-dominated ecosystems, and validation of existing plankton functional types with observations of taxon-specific biomass estimated from ocean color imagery.

## 5. Literature Cited


Banas, N. S., P. MacCready, and B. M. Hickey (2009), The Columbia River plume as cross-shelf exporter and along-coast barrier, *Continental Shelf Research*, *29*(1), 292-301.

Davis, C., and W. P. Bissett (2007), Characterization of a harmful algal bloom in Monterey Bay, CA using airborne hyperspectral imagery, in *Hyperspectral Imaging and Sounding of the Environment (HISE)*, edited, Optical Society of America (OSA), Santa Fe, NM.

Franz, B. A., P. J. Werdell, G. Meister, E. J. Kwiatkowska, S. W. Bailey, Z. Ahmad, and C. R. McClain (2006), MODIS land bands for ocean remote sensing applications, paper presented at Proc. Ocean Optics: XVIII, Montreal, Canada, 9 - 13 October 2006.

Gordon, H. R., J. W. Brown, O. B. Brown, R. H. Evans, and D. K. Clark (1983), Nimbus-7 Czcs - Reduction of Its Radiometric Sensitivity with Time, *Applied Optics*, *22*(24), 3929-3931.

Hickey, B. M. (1989), Patterns and processes of circulation over the Washington continental shelf and slope, in *Coastal oceanography of Washington and Oregon*, edited by M. R. Landry and B. M. Hickey, pp. 41-109, Elsevier, New York.

Horner-Devine, A. R. (2009), The bulge circulation in the Columbia River plume, *Continental Shelf Research*, *29*(1), 234-251.

Kirk, J. T. O. (1994), *Light and photosynthesis in aquatic ecosystems*, 509 pp., Cambridge University Press, New York.

Kudela, R. M., et al. (2010), Multiple trophic levels fueled by recirculation in the Columbia River plume, *Geophysical Research Letters*, *37*, 7.

Lee, Z. P., K. L. Carder, S. K. Hawes, R. G. Steward, T. G. Peacock, and C. O. Davis (1994), Model for the Interpretation of Hyperspectral Remote-Sensing Reflectance, *Applied Optics*, *33*(24), 5721-5732.

Maritorena, S., D. A. Siegel, and A. R. Peterson (2002), Optimization of a semi-analytical ocean color model for global-scale applications, *Applied Optics*, *41*, 2705 - 2714.

Martin Traykovski, L. V., and H. M. Sosik (2003), Feature-based classification of optical water types in the Northwest Atlantic based on satellite ocean color data, *J. Geophys. Res.*, *108*(C5), -.

Mobley, C. D. (1994), *Light and Water. Radiative Transfer in Natural Waters*, Academic, New York.

O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain (1998), Ocean color chlorophyll algorithms for SeaWiFS, *Journal of Geophysical Research-Oceans*, *103*(C11), 24937-24953.

Oliver, M. J., S. Glenn, J. T. Kohut, A. J. Irwin, O. M. Schofield, M. A. Moline, and W. P. Bissett (2004), Bioinformatic approaches for objective detection of water masses on continental shelves, *J. Geophys. Res.*, *109(C07S04)*, doi:10.1029/2003JC002072.

Palacios, S. L., T. D. Peterson, and R. M. Kudela (2009), Development of synthetic salinity from remote sensing for the Columbia River Plume, *Journal of Geophysical Research-Oceans*, *114*(C00B05).
Palacios, S. L., T. D. Peterson, and R. M. Kudela (in prep-a), Optical characterization of water masses within the Columbia River Plume.
Palacios, S. L., H. M. Sosik, K. K. Hayashi, M. Jacox, T. D. Peterson, and R. M. Kudela (in prep-b), Discrimination of phytoplankton taxa in an optically complex aquatic environment.

# CHAPTER 1

## Development of Synthetic Salinity from Remote Sensing for the Columbia River Plume


### Abstract

The Columbia River Plume (CRP) is an ecologically important source of nutrients, pollutants, and fresh water to the Oregon/Washington shelf. It is traditionally under-sampled, with observations constrained to ships or moorings. High spatial and temporal resolution observations afforded by satellites would increase sampling if the plume could be quantitatively detected in the imagery. Two empirical algorithms are presented using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) to estimate sea surface salinity in the region of CRP. Salinity cannot be detected directly, so a proxy for fresh water is employed. Light absorption by chromophoric dissolved organic matter ($a_{CDOM}$) is inversely proportional to salinity and linear due to conservative mixing of CDOM-rich terrestrial runoff with surrounding ocean water. To estimate synthetic salinity, simple linear (salinity vs. $a_{CDOM}$) and multiple linear (salinity & temperature vs. $a_{CDOM}$) algorithms were developed from in situ measurements of $a_{CDOM}$ collected on the Coastal Ocean Processes-River Influences on Shelf Ecosystems cruises. These algorithms were applied to MODIS 250m-resolution data layers of sea surface temperature and absorption by colored dissolved and detrital matter ($a_{CDM}$) estimated at 350nm and 412nm from the GSM01 algorithm. Validation of MODIS-derived synthetic salinity

with coincident in situ measurements revealed significant correlation during both downwelling (simple: $b_1$=0.95, $r^2$=0.89; multiple: $b_1$=0.92, $r^2$=0.89) and upwelling periods (simple: $b_1$=1.26, $r^2$=0.85; multiple: $b_1$=1.10, $r^2$=0.87) using the 412nm data layer. Synthetic salinity estimated using the 350 nm data layer consistently over-estimated salinity. These algorithms, when applied to $a_{CDM}$ at 412nm, enable synoptic observations of CRP not permitted by ships or moorings alone.

Key Terms: Columbia River Plume, chromophoric dissolved organic matter, CDOM, optics, remote sensing, MODIS

## 1.0 Introduction

The Columbia River watershed spans 674,000 $km^2$ and includes parts of the northwestern US and southwestern Canada. The mouth of the river lies at the border of the US states of Washington and Oregon (46° N, 124° W). Approximately 77% of the fresh water flow to the NE Pacific Ocean, from San Francisco to the Strait of Juan de Fuca, comes from the Columbia River (Hickey 1989). Discharge varies from 3,000 – 17,000 $m^3$ $s^{-1}$ (averaging 7,000 $m^3$ $s^{-1}$ annually), and reaches a maximum freshet in the spring due to snowmelt (Hickey 1998). Flow of the Columbia River pulses tidally to the ocean where it becomes a buoyant plume (Hickey 1989). The forces of wind stress, Coriolis, and inertia influence the flow of the Columbia River Plume (CRP) as it exits the mouth. With southerly winds, downwelling conditions prevail and the plume flows northward along the Washington shelf. Northerly winds induce

upwelling and the plume flows southward and offshore (Hickey 1989). It was widely accepted that the northward flow dominates in the winter and southward flow dominates in the summer. However, this canonical view of flow has recently been challenged (Garcia Berdeal *et al.* 2002). Even short term (hours to days) oscillations in wind forcing can cause transient reversals in the typical seasonal flow patterns (Garcia Berdeal *et al.* 2002; Hickey 1998), with reversals more probable in summer than winter. Thomas and Weatherbee (2006) analyzed the variability of the CRP along the Washington and Oregon coastal margin over a period of six years (1998 – 2003) using Sea-viewing Wide Field-of-view Sensor (SeaWiFS) imagery. The analysis included a supervised classification scheme using 5 channels of normalized water-leaving radiance ($nL_w$ at 412, 443, 490, 510 and 555 nm). Their observations confirmed the bi-modal flow of the plume, but they lacked in situ measurements for a validation of ocean color products and could relate statistical patterns in the imagery but not geochemically describe those patterns.

The CRP supplies silicate, nitrate, trace metals, pollution, fresh water, and organic matter (both dissolved and particulate) to the Washington and Oregon shelves (Aguilar-Islas & Bruland 2006; Carpenter & Peterson 1989; Hill & Wheeler 2002; Kachel & Smith 1989; Klinkhammer *et al.* 2000; Landry *et al.* 1989; McCarthy & Gale 1999). Stratification caused by the fresh water lens and retention of nutrient-rich waters on the shelf promote phytoplankton growth in this ecologically important habitat for juvenile salmon (De Robertis et al. 2005; Morgan et al. 2005). Therefore, identifying and tracking the plume is important for understanding the physical

processes affecting the biology and chemistry of the region. Some constituents in the water clearly act as optical tracers of the plume as it flows away from the river mouth (Thomas and Weatherbee, 2006). Plume-stimulated phytoplankton biomass, suspended inorganic and organic material (e.g. sediments), and chromophoric dissolved organic matter (CDOM) all have optical signals that may be useful for detecting and tracking the plume as it migrates along the Washington shelf or is absorbed into the California Current.

Salinity presently cannot be directly detected using satellites at small regional scales (10's of km), but sensors are in development to estimate global salinity from space at spatial scales of 10's to 100's of km and temporal scales of 30 days or more between estimates (Lagerloef *et al.* 2008). Despite improvements in remotely detecting salinity, most of the world's river plumes are too small and temporally dynamic to discern in the coastal environment using these new satellites. Existing ocean color sensors (e.g. Moderate Resolution Imaging Spectroradiometer: MODIS) can detect CDOM, which can be used as an optical tracer of river plumes in coastal margins (Binding & Bowers 2003; Callahan *et al.* 2004; Hu *et al.* 2004; Johnson *et al.* 2003; Vasilkov *et al.* 1999) at spatial and temporal scales relevant to their dynamic processes.

The unique optical character of CDOM in river water (Chen *et al.* 2004; Hernes & Benner 2003) can be employed as a proxy for low-salinity water near shore (DelVecchio & Subramaniam 2004; Johnson *et al.* 2003). CDOM is operationally defined as the substance that passes through a 0.2 μm filter and absorbs light strongly

from the ultraviolet to ~450 nm (Kirk 1994; Babin *et al.* 2003). It is a fraction of the total DOM pool and can be important in biogeochemical processes (Mopper & Kieber 2002) as well as an optical tracer. CDOM contributes to the quality and quantity of light leaving the sea surface and its variability in natural waters can be used to distinguish water types (Bricaud *et al.* 1981; Carder *et al.* 1989). The bulk CDOM pool in river plumes is enriched in tannins and lignins which contain highly absorbing aromatic rings (Blough & Del Vecchio 2002). This terrestrially derived CDOM absorbs light strongly at 350 nm and has a different optical character than marine CDOM which has few, if any, aromatic rings and absorbs weakly at 350 nm. The sources of CDOM in river plumes can be degraded vegetative material in the watershed; the product of phytoplankton and bacterial metabolism in the river, estuary, or plume after it exits the river mouth; or leachate from soils and estuarine sediments (Blough & Del Vecchio 2002; Klinkhammer *et al.* 2000). CDOM at the land-sea interface is generally terrestrially derived and loses its distinguishing character as it mixes conservatively with the old, "dirty laundry" CDOM of the open ocean (Siegel *et al.* 2002; Stedmon & Markager 2003). It is the difference in optical character between terrestrial and marine derived CDOM that can be exploited to optically detect low salinity plumes in near shore environments (Vasilkov *et al.* 1999; Johnson *et al.* 2003; Conmy *et al*. 2004; DelVecchio & Subramaniam 2004).

Remote sensing in the coastal environment affords a synoptic view that would be costly or not possible using shipboard measurements alone. However ground-truth measurements are needed, particularly in an environment as dynamic as the Columbia

River Plume. This study measures those in situ optical properties and is the first to relate salinity to optical properties to identify this water mass on the Washington and Oregon shelf. The objectives of this study were (a) to develop empirical algorithms to estimate low salinity water in the region of the Columbia River Plume; and (b) to apply the algorithms to 250m resolution MODIS data products. The study was conducted during the CoOP-RISE (Coastal Ocean Processes – River Influences on Shelf Ecosystems) cruises in June 2004, June 2005, August 2005, and June 2006 at the mouth and surrounding areas of the Columbia River.

## 2.0 Methods

### 2.1 In Situ Sample Collection

Four research cruises were conducted aboard two vessels, the *R/V Wecoma* and *R/V Point Sur*, on the Washington and Oregon shelf in the vicinity of the CRP during June to July 2004, May to June 2005, August 2005, and May to June 2006. Hereafter the cruises will be referred to as June 2004, June 2005, August 2005, and June 2006. The study area was contained within a region bounded by 44.7°N and 48.5°N latitude and 123.5°W and 125.2°W longitude (Fig. 1.1). Three of the cruises were scheduled to occur during maximum river discharge in the spring and one during reduced flow in the late summer (Fig. 1.2) (USGS, discharge at the Beaver Army Terminal, Quincy, OR).

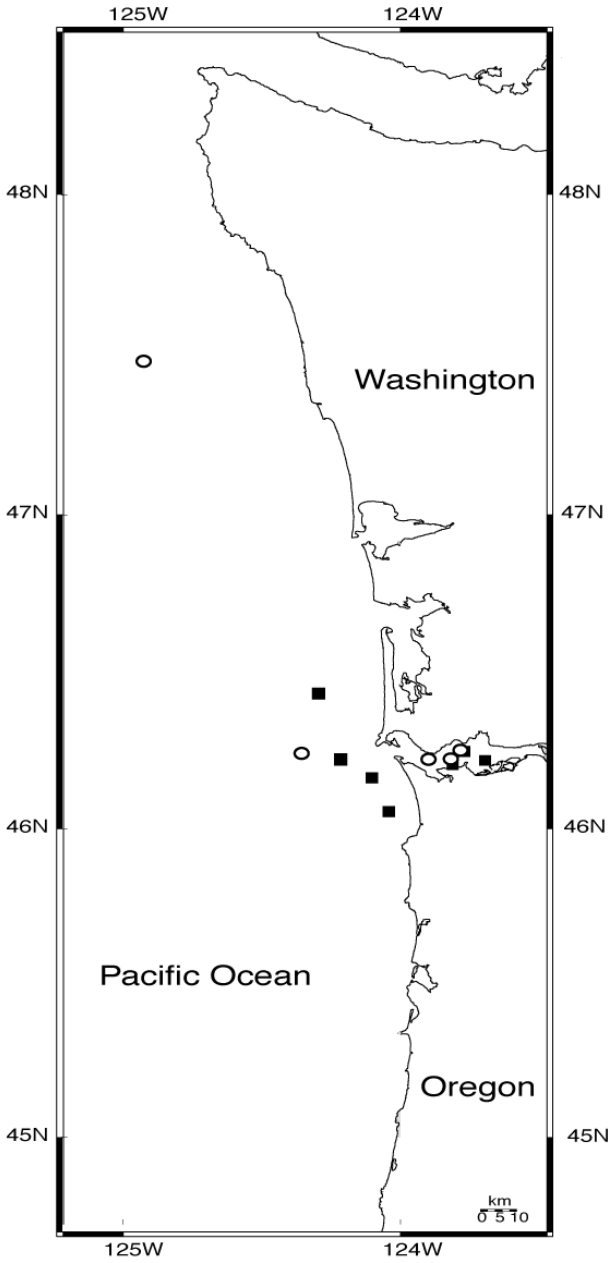



Figure 1.1. Study area. Washington and Oregon, United States shelf. Validation stations for 2004 (squares) and 2005 (circles).

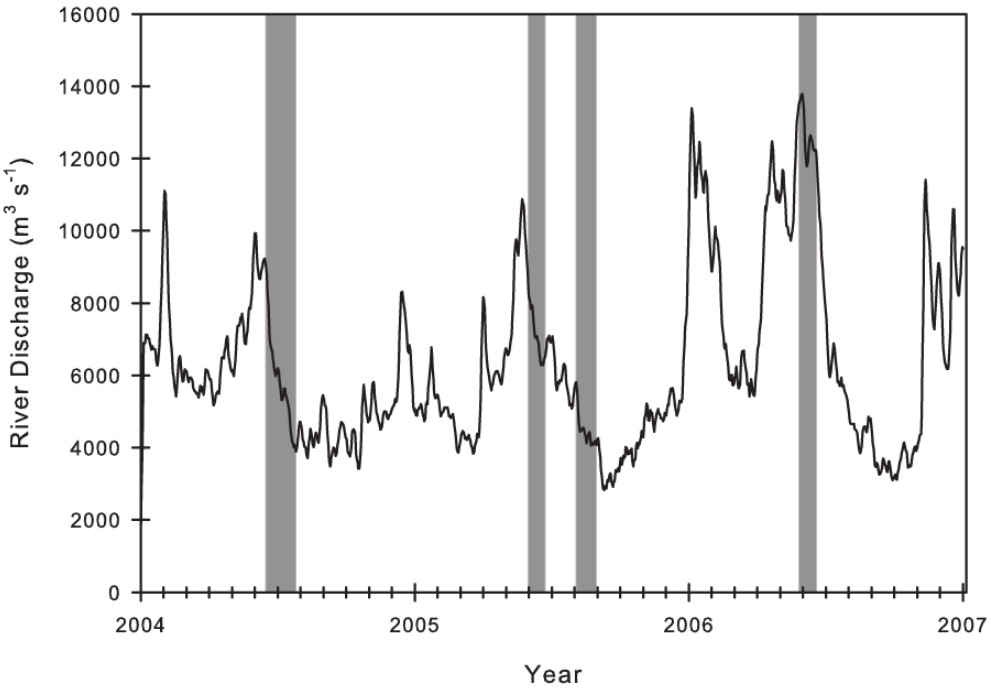



Figure 1.2. River discharge. Discharge of the Columbia River during the study period (5 day running averages). Measured at the Beaver Army Terminal, Quincy, Oregon, upstream of the Columbia River mouth. Volume reported as $m^3 s^{-1}$. Cruise dates noted by shaded area.

Light absorption by CDOM ($a_{CDOM}$) was measured from discrete water samples collected near the surface from the CTD rosette and the underway water stream. The underway system collected water at 4 m in 2004 and at 2 m in 2005 and 2006. Temperature and salinity were recorded at the time of water collection. Each sample was filtered through a 0.2 μm Nuclepore® polycarbonate filter, sealed in a polypropylene vial, and stored in the dark at 4°C until it was processed (within six weeks) in the lab at UC-Santa Cruz. CDOM optical properties are stable for up to four months if stored this way (Johannessen *et al.* 2003). Absorbance (A) was measured on a Cary UV-VIS spectrophotometer (300-800 nm with 0.5 nm resolution) using a 0.1 m pathlength quartz cuvette. A Millipore Q-water blank was subtracted from these values and spectral absorption ($a$) was calculated using equation (1) (Kirk 1994)

$$a_{CDOM}(\lambda) = \frac{2.303 * \left(A_{(\lambda)} - A_{(750)}\right)}{0.1} \quad (1)$$

The value 2.303 is the correction factor converting log (10) to natural log, $A_{(750)}$ is subtracted from all values of A to account for scattering by small particles, and 0.1 m is the cuvette pathlength to give units $m^{-1}$ for absorption. Light absorption at wavelength 350 nm was used in the analysis because terrestrially-derived CDOM absorbs strongly at this wavelength due to tannins and lignins (Hernes & Benner 2003). This terrestrially-derived CDOM acts as a tracer for fresh water on the shelf because CDOM concentration is inversely proportional to salinity and its conservative decrease is a result of mixing.

## 2.2 Statistical Models

Two empirical algorithms were developed for each cruise from the $a_{CDOM}$ measurements collected in situ. The first algorithm ("simple model") was a simple linear regression testing salinity vs. $a_{CDOM}$ at 350 nm. The second algorithm ("multiple model") was a multiple linear regression testing salinity and temperature vs. $a_{CDOM}$ at 350 nm. Temperature was included in the multiple model to account for the possible contrast in temperature in river water from oceanic water, particularly because upwelling (represented by cool temperatures) is a dominant process in this region. An additional reason was to tune the salinity estimate to the CRP and away from the "false plumes" of nearby estuaries that may have had different native temperatures. A critical p-value of 0.05 was used to determine significance. These tests were conducted for each cruise, and then a multivariate analysis of co-variance (MANCOVA) was used to determine differences among cruises. Statistical tests were computed using the MATLAB Statistics Toolkit (The MathWorks, Inc.).

## 2.3 Satellite Analysis

### 2.3.1 Data Collection

During the three year study period, only five days out of a total of 95 cruise days were completely clear of clouds over the entire Oregon and Washington shelf. Of these five clear days, two days were selected that had sufficient shipboard and mooring data to compare in situ measurements of salinity to satellite derived estimates of salinity. These two days represented a period of wind driven oceanographic downwelling (21 July 2004) and oceanographic upwelling (25 August

2005) (Fig. 1.3). In addition to these two clear days used in the salinity analysis, a sub-set of images from partly cloudy days was used to evaluate how well the satellite-derived estimate of light absorption by dissolved and detrital matter ($a_{CDM}$) fit to in situ, shipboard measurements of $a_{CDOM}$. Partly cloudy scenes were used when the region in the image matching the in situ collection location was clear of clouds or edge effects of clouds. Of fourteen, partially clear days in June 2004 and August 2005, only four in 2004 and three in 2005 met the requirements for comparison (2004: 11, 12, 17 & 21 July; 2005: 22, 23 & 25 August).

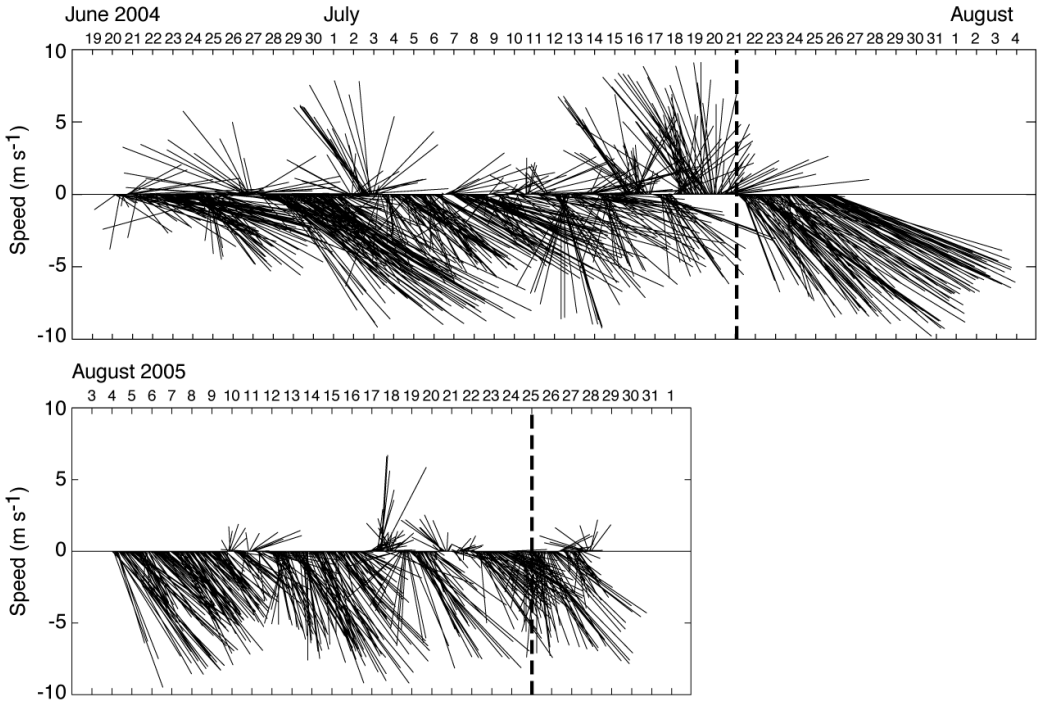



Figure 1.3. Wind vectors. Speed and direction of wind flow at the Columbia River mouth during the study periods in (a) 2004 and (b) 2005. Dashed lines represent date of satellite image capture and subsequent analysis. Note that 21 July 2004 was a period of downwelling-favorable winds and 25 August 2005 a period of upwelling-favorable winds.

Level 0 MODIS swaths occupying the region 40° to 50° N and 128° to 123° W for the study days were downloaded from LAADS Web (Goddard Space Flight Center, Level 1 and Atmosphere Archive and Distribution System). These data were processed to Level 2 geophysical products using SeaDAS (version 5.1) with the default settings of the NASA/Goddard Ocean Biology Processing Group (OBPG) using a two wavelength (1240 and 2130 nm) atmospheric correction (Franz *et al.* 2006). Data were projected using a sinusoidal projection at 250 m resolution and sub-setted to the region of the Oregon and Washington shelf bounded by 44.7° to 48.5° N and 123.4° to 125.2° W. The data layers acquired and processed in this study included sea surface temperature (SST) and light absorption by colored dissolved and detrital matter at 412 nm ($a_{dg}$ (l)) using the GSM01 (Maritorena *et al.* 2002), QAA (Lee *et al.* 2002), and Carder (Carder *et al.* 1998) algorithms. Estimates of $a_{dg}$ 350 nm were calculated using equation 2 (Twardowski *et al.* 2004):

$$a_\lambda = a_\lambda * \exp(s(\lambda_0 - \lambda)) \qquad (2)$$

where $l_0$ equals 412; l equals 350; and the spectral slope parameter, *s*, was 0.0206 for GSM01, 0.015 for QAA, and 0.022 for the Carder algorithm-- the standard parameter values used by SeaDAS. Although the data used for these analyses (SST and $a_{dg}$ (l)) are at a native resolution of 1 km (nadir), SeaDAS applies a bilinear cubic interpolation to create pseudo-250 m resolution data; the algorithms then use the true 250 m resolution data at 645 nm and 500 m resolution data at 469 nm and 555 nm wavelengths to interpolate the 250 m resolution for the other wavelengths (Franz *et al.* 2006).

2.3.2 $a_{CDOM}$ vs. $a_{CDM}$ and Algorithm Comparison

The three satellite algorithms were carefully evaluated to determine which one had the best fit of in situ $a_{CDOM}$ at 412 nm to satellite-derived $a_{dg}$ 412 nm. Sediment and other colored detrital material influence the $a_{dg}$ measurement; the goal was to quantify the error due to the presence of this material and to choose the algorithm with the best fit for $a_{CDOM}$ for use in the salinity estimate. Shipboard measurements of $a_{CDOM}$ were collected within one hour of the satellite over-pass on each of the seven days in 2004 and 2005 identified above (locations not shown). Diverse water types were sampled in 2004, though less so in 2005 due to the limits of available satellite imagery. The imagery was sub-setted to a 4 by 4 pixel box ($10^6$ $m^2$) centered on the in situ sample location. The mean $a_{dg}$ 412 nm value of this 4 by 4 pixel box was calculated for each scene for each of the three satellite algorithms. The in situ $a_{CDOM}$ at 412 nm was the independent variable and these $a_{dg}$ 412 nm estimates for each algorithm were the dependent variables in tests of simple linear regression with ANOVA. The slopes from the different regression tests were compared using a MANCOVA. A critical p-value of 0.05 was used to determine significance.

Careful evaluation of the GSM01, QAA, and Carder algorithms revealed that in June 2004 in the region of the CRP, GSM01 had greater fidelity between in situ $a_{CDOM}$ and satellite derived $a_{CDM}$ with respect to fit, slope, and intercept (Table 1.1). The slopes for GSM01 and Carder were different in 2004 ($F = 2.35$, $DFn = 1$, $DFd = 4$, $p = 0.04$). In August, there was the expected trend between $a_{CDOM}$ and $a_{dg}$, but no significant fit for any of the algorithms and so they could not be compared to each

other. This lack of fit may have been the result of high turbidity, but more likely a result of low sample size exacerbated by the narrow range of $a_{CDOM}$ values used in the comparison. We decided to restrict the salinity analysis of this study to the MODIS-derived data layers from the GSM01 algorithm for both years with the understanding that the over-estimate observed was likely due to the presence of detrital and colored particles contributing to the $a_{dg}$ measurement. As a result of this off-set, synthetic salinity estimates have the potential to be negative if the $a_{dg}$ estimate exceeds the parameters in the synthetic salinity model.

Table 1.1. Validation of in situ $a_{CDOM}$ (412 nm) surface measurements with MODIS-derived $a_{dg}$ 412 nm estimates for the three algorithms: GSM01, QAA, and Carder. ($\beta_1$ = regression coefficient ; $\beta_0$ = intercept; $p \leq 0.05$ is significant).

| | Results of Regression Analysis | | | | |
|---|---|---|---|---|---|
| Estimate | $r^2$ | n | β1 | β0 | p |
| 2004 | | | | | |
| GSM01 | 0.98 | 4 | 1.4 | -0.23 | < 0.05 |
| QAA | 0.59 | 4 | 0.67 | -0.05 | n.s. |
| Carder | 0.92 | 4 | 0.75 | -0.09 | < 0.05 |
| 2005 | | | | | |
| GSM01 | 0.95 | 3 | 2.5 | -0.59 | n.s. |
| QAA | 0.95 | 3 | 2.4 | -0.59 | n.s. |
| Carder | 0.27 | 3 | 1.4 | -0.21 | n.s. |

2.3.3 Synthetic Salinity Estimates

Synthetic salinity was computed by inverting the in situ linear regression models to solve for salinity. The MODIS-derived data layers (250 m resolution GSM01 $a_{dg}$ at 350 nm and 412 nm) were used as the inputs for $a_{CDOM}$ and the SST data layer as the input for temperature (for the multiple algorithm). Both GSM01 $a_{dg}$

412 nm and the estimated values for GSM01 $a_{dg}$ 350 nm were used to calculate synthetic salinity in order to determine if the $a_{dg}$ 412 nm data layer was sufficient to accurately predict salinity, or if the derived data layer was needed since the underlying models were developed using $a_{CDOM}$ at 350 nm. The "simple" model computed synthetic salinity using equation (3):

$$\text{synthetic salinity} = \frac{a_{CDOM} - \beta_0}{\beta_1} \tag{3}$$

where $b_0$ and $b_1$ are the intercept and regression coefficients, respectively. The standard equation (4) for the multiple model was inverted to solve for synthetic salinity:

$$\text{synthetic salinity} = \frac{a_{CDOM} - \beta_0 - \beta_2(\text{temperature})}{\beta_1} \tag{4}$$

where $b_2$ is the second regression coefficient. The empirical algorithms developed for each cruise were only applied to imagery collected during the respective cruise. Four estimates of synthetic salinity were computed for each date: simple synthetic salinity using $a_{dg}$ 350 nm and 412 nm and multiple synthetic salinity using $a_{dg}$ 350 nm and 412 nm.

2.4 Validation

In situ measurements of salinity were compared to estimates of synthetic salinity to validate the robustness of the underlying empirical models. Several independent sources of in situ salinity measurements were used in the validation: underway shipboard measurements from the *R/V Wecoma* and *R/V Point Sur*, *M/V Forerunner*, CoOP-RISE moorings located on the shelf, and CORIE moorings

immediately within the river mouth (Table 1.2). For comparison of in situ and remotely sensed data, the mean of four minutes of salinity data (two-minute intervals centered at the time of the satellite over-pass) was used as salinity data for the mooring and shipboard sources. Synthetic salinity from the remote sensing layers was selected using a 4 by 4 pixel box ($10^6$ $m^2$) centered on the in situ site. Pixels influenced by land were flagged using the SST and $a_{dg}$ 412 nm bands and removed. The remaining pixels were used to compute the mean synthetic salinity value in this $10^6$ $m^2$ box. These mean values were used in a test of simple linear regression (in situ salinity vs. synthetic salinity) to determine a fit between measured and modeled salinity and to estimate error. A critical p-value of 0.05 was used to determine significance.

Table 1.2. Date, data source, and location for in situ, ground-truth salinity measurements. CoOP-RISE temporary deployed moorings include RISO, RINO, and RICE (courtesy, E. Dever). CORIE measurements include moorings: Grays Point and Cathlamet Bay North Channel 3 (CBNC3); and shipboard measurements on the *M/V Forerunner* (courtesy, A. Baptista).

| Data Source | Latitude (°N) | Longitude (°W) | Depth |
|---|---|---|---|
| | | 21 July 2004 | |
| *R/V Wecoma* | 46.200 N | 123.806 W | 4 m |
| *R/V Point Sur* | 46.242 N | 124.260 W | 4 m |
| RISO | 46.053 N | 124.101 W | 1 m |
| RINO | 46.437 N | 124.301 W | 1 m |
| RICE | 46.167 N | 124.195 W | 1 m |
| Grays Point | 46.262 N | 123.767 W | 6.4 m |
| CBNC3 | 46.210 N | 123.714 W | 6.5 m |
| | | 25 August 2005 | |
| *R/V Wecoma* | 47.493 N | 124.912 W | 2 m |
| *R/V Point Sur* | 46.251 N | 124.360 W | 4 m |
| *M/V Forerunner* | 46.233 N | 123.872 W | 1 m |
| Grays Point | 46.262 N | 123.767 W | 6.4 m |
| CBNC3 | 46.210 N | 123.714 W | 6.5 m |

## 3.0 Results

### 3.1 Field Setting

River discharge varied seasonally, with a trend of relatively high flow during the spring and reduced flow in the late summer and autumn (Fig. 1.2). This seasonal pattern is typical for the Columbia River. High flows are due to runoff of snowmelt in the watershed and can vary in timing annually. Maximum discharge preceded the June 2004 and June 2005 cruises, and was concurrent with the June 2006 cruise. Peak discharge in June 2006 was 50% greater than peak discharge in 2004 and 2005, which were similar to each other. A delay in the onset of upwelling favorable winds in 2005 resulted in a period in June with no upwelling and anomalously warm waters throughout the northern California Current system; however, by our August cruise (Fig. 1.3), strong upwelling-favorable conditions were present (Hickey *et al.* 2006; Kudela *et al.* 2006). Anomalously warm SST in June 2005 decreased the gradient in temperatures between plume and offshore water, which may have influenced mixing of the plume with surrounding marine waters.

Light absorption by CDOM was linear and inversely proportional to salinity (Fig. 1.4). The relationship of salinity to $a_{CDOM}$ at 350 nm in the simple model was significant for discrete water samples measured for each of the cruises (Table 1.3). When the cruises were compared to each other using MANCOVA, the slopes were found to be statistically different ($F = 22.41$, $DFn = 3$, $DFd = 171$, $p < 0.001$), and therefore intercepts could not be compared among all of the cruises. When the June cruises from all years were compared to each other, the relationship between salinity

and $a_{CDOM}$ at 350 nm was similar between 2005 and 2006, but not 2004 (F = 3.64, DFn = 2, DFd = 140, p = 0.03). Nevertheless, June values were pooled to produce a "universal" statistical relationship of salinity to light absorption by CDOM for the spring freshet, with the understanding that natural variability exists among years. This universal synthetic salinity model followed the same trend as the statistically significant models from individual cruises and may be useful to estimate synthetic salinity on the shelf in future work. The relationship of salinity and temperature to $a_{CDOM}$ at 350 nm in the multiple model was significant for discrete water samples measured for each of the cruises (Table 1.4).

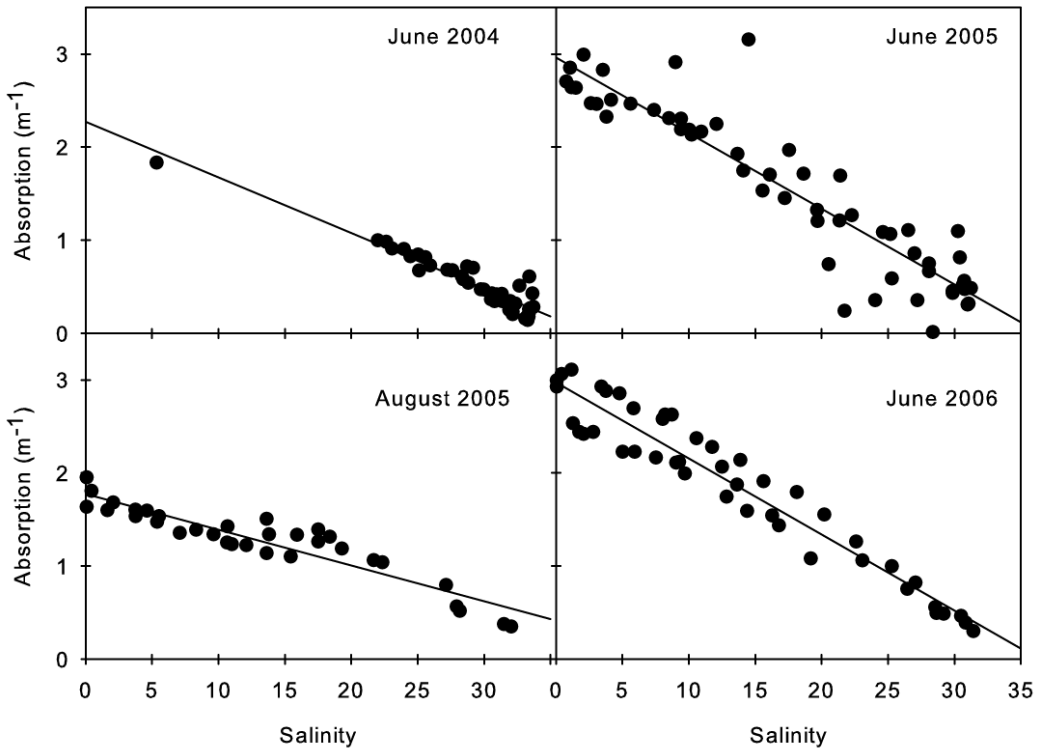



Figure 1.4. In situ salinity model. Relationship of salinity to light absorption at 350 nm for discrete water samples collected during the June 2004, June 2005, August 2005, and June 2006 Coastal Ocean Processes–River Influences on Shelf Ecosystems (CoOP-RISE) cruises. Light absorption at 350 nm was inversely related to salinity at each sample period. See Table 3 for statistics.

Table 1.3. The results of simple linear regression to test the relationship of salinity to $a_{CDOM}$ at 350 nm for discrete water samples collected within and outside of the Columbia River Plume in June 2004, June 2005, August 2005, June 2006, and all June cruises pooled ($\beta_1$ = regression coefficient/slope; $\beta_0$ = intercept; $p \leq 0.05$ is significant).

| | | | Results of Regression Analysis | | |
|---|---|---|---|---|---|
| Cruise | $r^2$ | n | β1 | β0 | p |
| Jun 2004 | 0.9 | 46 | -0.06 | 2.27 | < 0.05 |
| June 2005 | 0.85 | 55 | -0.08 | 2.96 | < 0.05 |
| August 2005 | 0.87 | 33 | -0.04 | 1.77 | < 0.05 |
| June 2006 | 0.93 | 45 | -0.08 | 2.97 | < 0.05 |
| Universal June | 0.92 | 146 | -0.08 | 2.95 | < 0.05 |

Table 1.4. The results of multiple linear regression to test the relationship of salinity and temperature to $a_{CDOM}$ at 350 nm for discrete water samples collected within and outside of the Columbia River Plume in June 2004, June 2005, August 2005, June 2006, and all June cruises pooled ($\beta_1$ = regression coefficient for salinity; $\beta_2$ = regression coefficient for temperature; $\beta_0$ = intercept; $p \leq 0.05$ is significant).

| | | | Results of Regression Analysis | | | |
|---|---|---|---|---|---|---|
| Cruise | $r^2$ | n | β1 | β2 | β0 | p |
| Jun 2004 | 0.9 | 46 | -0.06 | 0.01 | 2.09 | < 0.05 |
| June 2005 | 0.85 | 55 | -0.08 | -0.13 | 5 | < 0.05 |
| August 2005 | 0.87 | 33 | -0.07 | -0.12 | 4.42 | < 0.05 |
| June 2006 | 0.96 | 45 | -0.08 | 0.2 | -0.17 | < 0.05 |
| Universal June | 0.92 | 146 | -0.08 | 0.02 | 2.67 | < 0.05 |

3.2 Observations During a Downwelling Period (21 July 2004)

The CRP formed a coherent feature on the Washington shelf, north of the river mouth in the image dated 21 July 2004 (Fig. 1.5 and 1.6). There was a distinct bolus of what appeared to be fresh water west of the river mouth. The satellite overpass occurred 4.5 hours after maximum ebb tide during a period of very low spring tides. River discharge averaged 3816 $m^3$ $s^{-1}$ this day. The river plume appears to extend northward along the shelf, which may be remnant plume water from previous days during this downwelling period.

The plume can be clearly distinguished in both the simple and multiple model estimates for synthetic salinity using GSM01 $a_{dg}$ 350 nm (Fig. 1.5) and GSM01 $a_{dg}$ 412 nm (Fig. 1.6). The simple and multiple models corresponded closely to each other for each data layer ($a_{dg}$ 350 nm: $b_1 = 1.006$; $r^2 = 0.99$; $p \leq 0.05$; $a_{dg}$ 412 nm: $b_1 = 1.015$; $r^2 = 0.99$; $p \leq 0.05$, where $b_1$ is the slope of the regression and $b_0$ is the intercept). Synthetic salinity estimates of the simple algorithm were approximately 0.6 % less than estimates of the multiple algorithm using $a_{dg}$ 350 nm and 1.5% less using $a_{dg}$ 412 nm. In the simple algorithm and using $a_{dg}$ 350 nm (Fig. 1.5A), synthetic salinity estimates ranged from -47.0 in the estuary to 36.2 offshore. The nascent plume at the mouth of the river had synthetic salinity that ranged from 5.2 to 13.0 and plume axis salinities ranging from -5.6 to 10.5. The edges of the plume were defined by a steep gradient in synthetic salinity from 6.3 to 22.9 over a distance less than 4 km. In the simple algorithm and using $a_{dg}$ 412 nm (Fig. 1.6A), synthetic salinity

estimates ranged from 0.6 in the estuary to 37.4 offshore. Synthetic salinity at the river mouth ranged from 21.8 to 31.9 and plume axis salinities ranged from 21.8 to 27.9. The edges of the plume were defined by a fairly steep gradient in synthetic salinity over less than 4 km from 28.6 to 31.3. In the multiple algorithm and using $a_{dg}$ 350 nm (Fig. 1.5B), synthetic salinity ranged from -47.3 in the estuary to 38.0 offshore. The mouth of the plume ranged from -7.9 to 14.1 and along the plume axis had a synthetic salinity of -7.3 to 9.8. The edges of the plume had a steep gradient from 9.9 to 22.7. In the multiple algorithm and using $a_{dg}$ 412 nm (Fig. 1.6B), synthetic salinity estimates ranged from 0.7 in the estuary to 37.4 offshore. The mouth of the plume ranged from 20.5 to 31.0 and along the plume axis had a synthetic salinity of 20.7 to 27.5. The edges of the plume ranged from 24.7 to 31.9.

3.3 Observations During an Upwelling Period (25 August 2005)

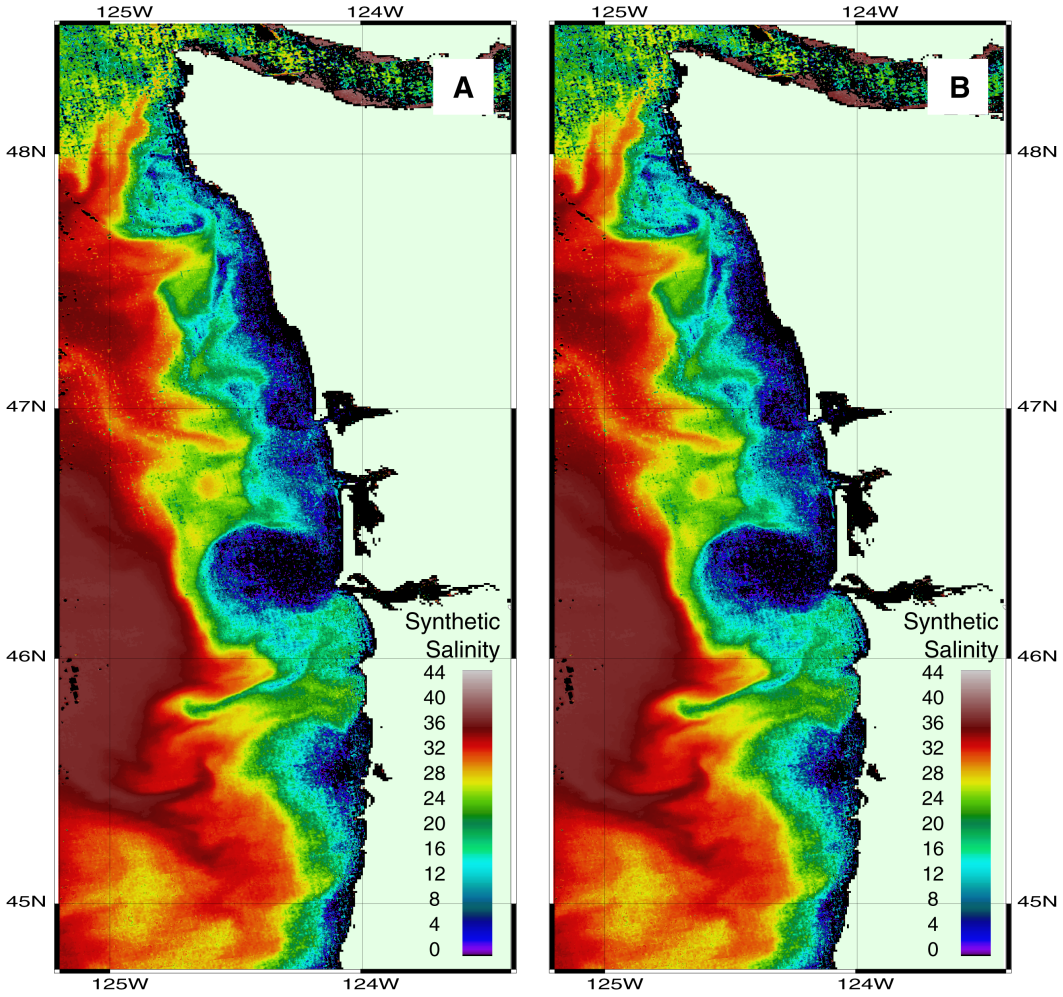



Figure 1.5. Synthetic salinity. Estimates of sea surface salinity for 21 July 2004, a period of downwelling oceanographic conditions. (a) Simple algorithm showing salinity versus $a_{CDOM}$ at 350 nm and (b) multiple algorithm showing salinity and temperature versus $a_{CDOM}$ at 350 nm. Satellite products used in analysis included the 250 m MODIS data layers for sea surface temperature and GSM01 $a_{dg}$ 350 nm.

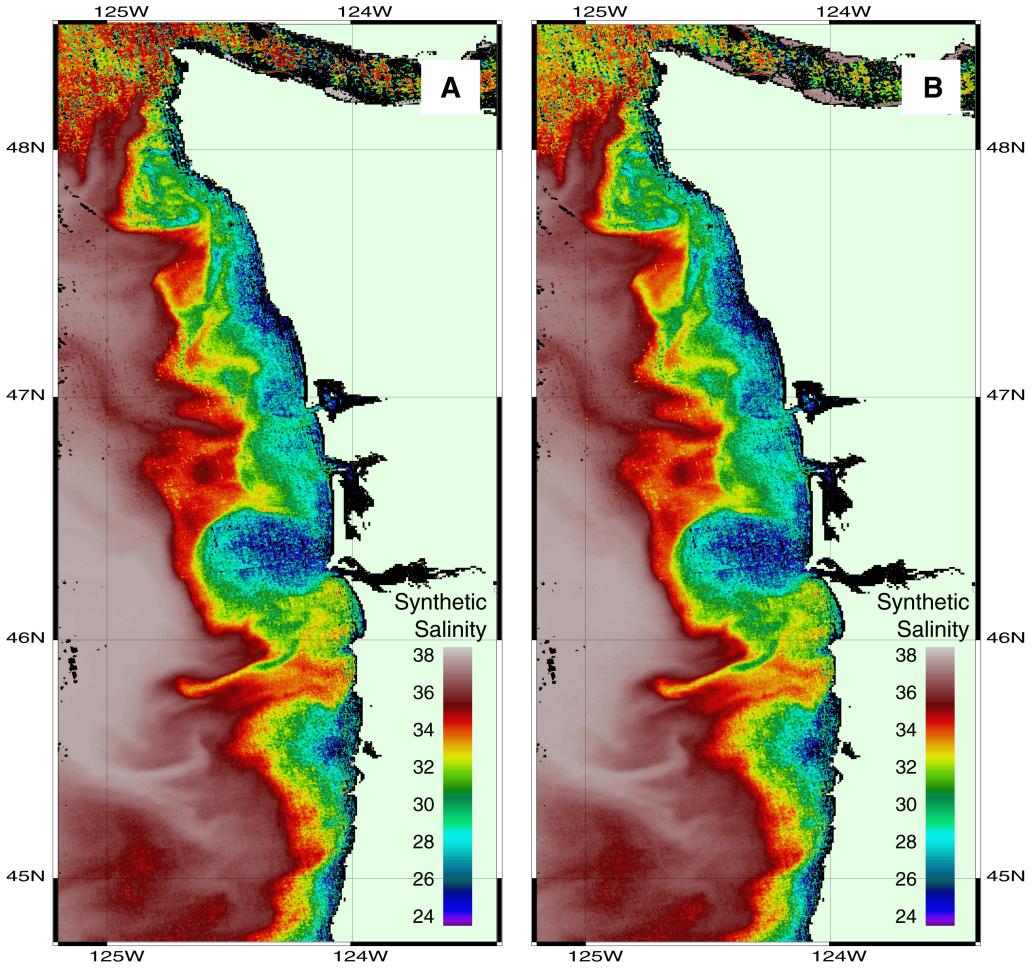



Figure 1.6. Synthetic salinity. Estimates of sea surface salinity for 21 July 2004, a period of downwelling oceanographic conditions. (a) Simple algorithm showing salinity versus $a_{CDOM}$ at 412 nm and (b) multiple algorithm showing salinity and temperature versus $a_{CDOM}$ at 412 nm. Satellite products used in analysis included the 250 m MODIS data layers for sea surface temperature and GSM01 $a_{dg}$ 412 nm.

The CRP was a diffuse feature on the Oregon and Washington shelf in the image dated 25 August 2005 (Fig. 1.7 and 1.8). The CRP was only faintly present west and south of the river mouth. The satellite over-pass occurred 3.5 hours following maximum ebb tide during the neap tide, so the tidal range was very small. River discharge averaged 3852 $m^3$ $s^{-1}$ this day, similar to the flow on 21 July 2004. This was a period of upwelling favorable winds and the plume appeared to form a disjointed fan of water near and southward of the river mouth, where it was absorbed into what appeared to be the California Current.

The simple and multiple algorithm estimates of synthetic salinity on the shelf did not closely correspond to each other in either the simple or multiple model estimates for synthetic salinity using GSM01 $a_{dg}$ 350 nm (Fig. 1.7) and GSM01 $a_{dg}$ 412 nm (Fig. 1.8). Estimates for synthetic salinity using the simple model overestimated values compared to the multiple model ($a_{dg}$ 350 nm: $b_1 = 0.53$; $r^2 = 0.89$; $p \leq 0.05$; $a_{dg}$ 412 nm: $b_1 = 0.48$; $r^2 = 0.54$; $p \leq 0.05$). Synthetic salinity estimates of the simple algorithm were approximately 47 % more than estimates of the multiple algorithm using $a_{dg}$ 350 nm and 52% more using $a_{dg}$ 412 nm. In the simple algorithm and using $a_{dg}$ 350 nm (Fig. 1.7A), synthetic salinity estimates ranged from -52.0 in the estuary to 39.5 offshore. The river mouth had synthetic salinity that ranged from 9.3 to 21.6 and plume axis salinities ranging from 16.2 to 21.3. The edges of the plume were not well defined and ranged in synthetic salinity from 17.7 to 24.0 over a distance greater than 10 km. In the simple algorithm and using $a_{dg}$ 412 nm (Fig. 1.8A), synthetic salinity estimates ranged from 0.6 in the estuary to 42.8 offshore.

Synthetic salinity at the river mouth ranged from 31.2 to 36.2 and plume axis salinities ranging from 29.9 to 37.5. The edges of the plume gradually transitioned from 36.4 to 38.5. In the multiple algorithm and using $a_{dg}$ 350 nm (Fig. 1.7B), synthetic salinity estimates ranged from -53.9 in the estuary to 31.3 offshore. The mouth of the plume ranged from 5.22 to 15.9 and along the plume axis had a synthetic salinity of 14.5 to 24.8. The edges of the plume gradually decreased from 20.6 to 26.9. In the multiple algorithm and using $a_{dg}$ 412 nm (Fig. 1.8B), synthetic salinity estimates ranged from 2.9 in the estuary to 32.7 offshore. The mouth of the plume ranged from 24.5 to 28.6 and along the plume axis had a synthetic salinity of 27.1 to 33.4. The edges of the plume ranged from 31.9 to 38.5 over a span of approximately 10 km.

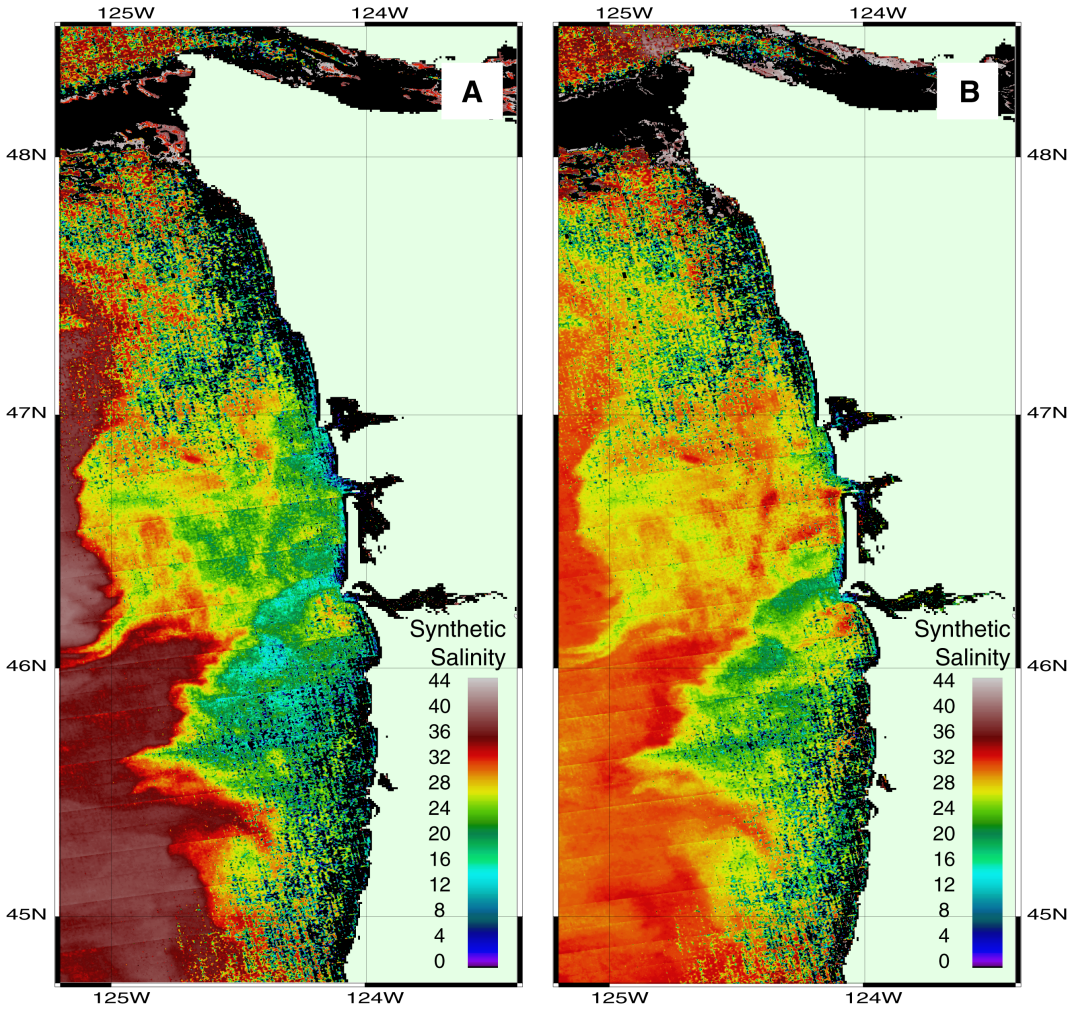



Figure 1.7. Synthetic salinity. Estimates of sea surface salinity for 25 August 2005, a period of upwelling conditions. (a) Simple algorithm showing salinity versus $a_{CDOM}$ at 350 nm and (b) multiple algorithm showing salinity and temperature versus $a_{CDOM}$ at 350 nm. Satellite products used in analysis included the 250 m MODIS data layers for sea surface temperature and GSM01 $a_{dg}$ 350 nm.

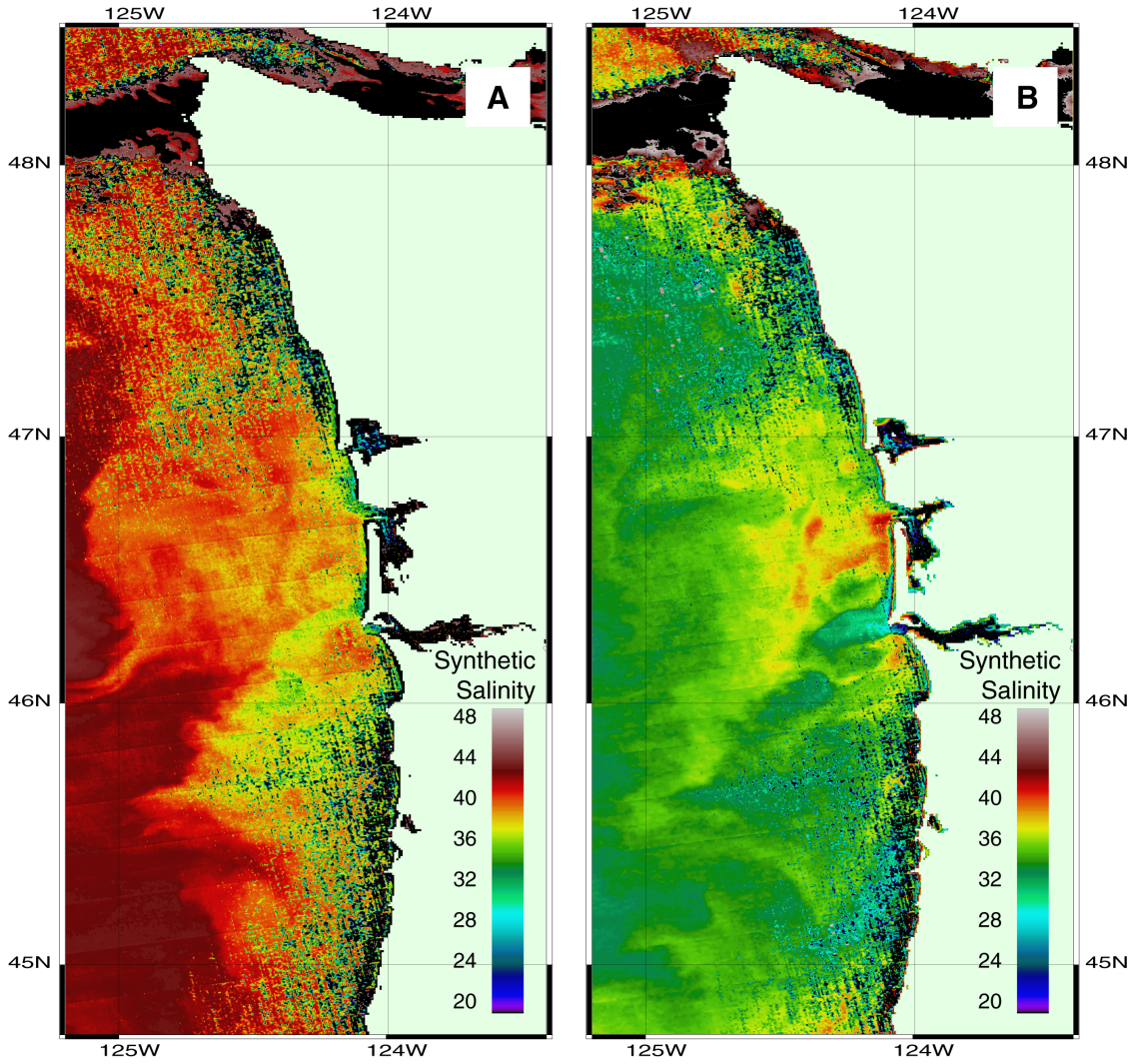



Figure 1.8. Synthetic salinity. Estimates of sea surface salinity for 25 August 2005, a period of upwelling conditions. (a) Simple algorithm showing salinity versus $a_{CDOM}$ at 412 nm and (b) multiple algorithm showing salinity and temperature versus $a_{CDOM}$ at 412 nm. Satellite products used in analysis included the 250 m MODIS data layers for sea surface temperature and GSM01 $a_{dg}$ 412 nm.

3.4 In Situ Salinity vs. Synthetic Salinity Validation

Both the simple and multiple synthetic salinity estimates that used the GSM01 $a_{dg}$ 412 nm data layer predicted salinity in the region of the CRP with a significant fit of in situ salinity to synthetic salinity measurements and the slopes only moderately deviated from the 1:1 relationship (Table 1.5; Fig. 1.9A and 1.9B). The mean square prediction errors (MSPE) were less than the mean square error (MSE) and therefore the models were effective predictors of salinity. The model estimates using the GSM01 $a_{dg}$ 350 nm data layers produced spurious salinity values for both the 2004 and 2005 data sets (Table 1.5; Fig. 1.9C and 1.9D). Salinity estimates ranged from negative to unrealistically high values; the slopes and intercepts for the validation reveal these extremes (Table 1.5). Synthetic salinity computed using the GSM01 $a_{dg}$ 412 nm data layer closely approximated in situ salinity, even though the in situ models were developed using light absorption at 350 nm. This suggests that the GSM01 $a_{dg}$ retrievals are offset (correct spatial pattern, wrong magnitude), reflecting some bias in the GSM01 algorithm. This offset, in addition to limitations in the range of magnitude of underlying data used to develop the synthetic salinity model may explain negative salinity values found in this study. This discrepancy was somewhat present in the model estimates using $a_{dg}$ 412 nm, and extreme for estimates using $a_{dg}$ 350 nm. Therefore, applying the synthetic salinity models to the GSM01 $a_{dg}$ 412 nm data layer instead of $a_{dg}$ 350 nm was the preferred method, and only the validations using the $a_{dg}$ 412 nm method are described.

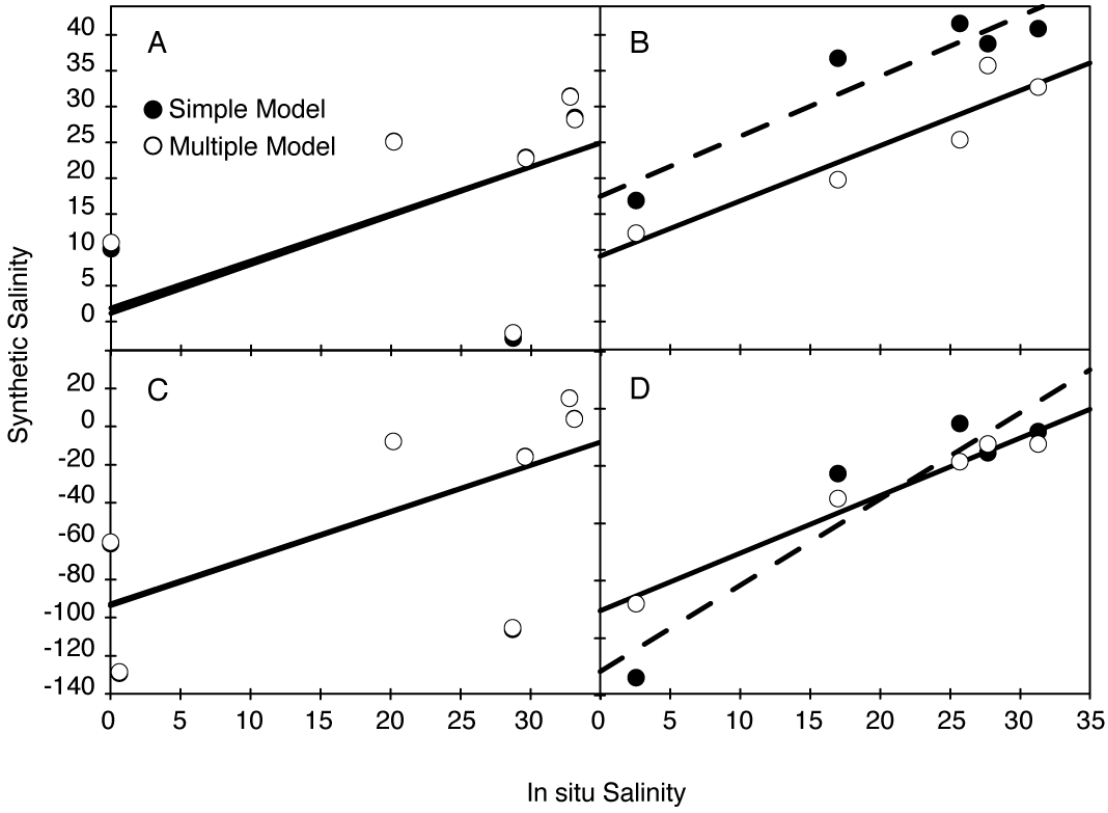



Figure 1.9. Model validation. Comparison of near-simultaneous (±2 min, 106 $m^2$ grid) MODIS-derived estimates of synthetic salinity and in situ salinity. In situ measurements collected aboard the *R/V Wecoma, R/V Point Sur*, and *M/V Forerunner* and via the CoOP-RISE and Columbia River Estuary mooring arrays for (a) 21 July 2004, synthetic salinity from GSM01 $a_{dg}$ 412 nm; (b) 25 August 2005, synthetic salinity from GSM01 adg 412 nm; (c) 21 July 2004, synthetic salinity from GSM01 $a_{dg}$ 350 nm; and (d) 25 August 2005, synthetic salinity from GSM01 $a_{dg}$ 350 nm. Simple algorithm (solid circles and dashed lines) and multiple algorithm (open circles and dashed lines). See Table 1.5 for statistics.

Table 1.5. Validation of MODIS-derived synthetic salinity estimates with in situ measurements of salinity for 21 July 2004 and 25 August 2005. Synthetic salinity estimates derived from MODIS data layers GSM01 $a_{dg}$ 412 nm and GSM01 $a_{dg}$ 350 nm ($\beta_1$ = regression coefficient ; $\beta_0$ = intercept; $p \leq 0.05$ is significant).

| | Results of Regression Analysis | | | | | | |
|---|---|---|---|---|---|---|---|
| Estimate | $r^2$ | n | β1 | β0 | MSE | MSPE | p |
| 21 July 2004 | | | | | | | |
| GSM01 adg 412 nm | | | | | | | |
| simple model | 0.89 | 7 | 0.95 | -0.77 | 33.28 | 23.77 | < 0.05 |
| multiple model | 0.89 | 7 | 0.92 | -0.07 | 32.37 | 23.12 | < 0.05 |
| GSM01 adg 350 nm | | | | | | | |
| simple model | 0.89 | 7 | 3.42 | -100.6 | 428.1 | 305.8 | < 0.05 |
| multiple model | 0.89 | 7 | 3.39 | -99.9 | 424.8 | 303.4 | < 0.05 |
| 25 August 2005 | | | | | | | |
| GSM01 adg 412 nm | | | | | | | |
| simple model | 0.85 | 5 | 1.26 | -0.71 | 75.86 | 45.52 | < 0.05 |
| multiple model | 0.87 | 5 | 1.1 | -1.4 | 49.1 | 29.45 | < 0.05 |
| GSM01 adg 350 nm | | | | | | | |
| simple model | 0.85 | 5 | 4.5 | -117 | 975.6 | 585.3 | < 0.05 |
| multiple model | 0.86 | 5 | 2.95 | -67.84 | 386.8 | 232.1 | < 0.05 |

On 21 July 2004, the simple and multiple models underestimated salinity by approximately 5% and 8%, respectively (Table 1.5). The effect of temperature on the multiple algorithm estimate from 2004 was limited because the regression coefficient for temperature was very small ($b_2 = 0.01$); for the same reason, estimates from both models are similar. The 25 August 2005 results show wider variability between the two models with the multiple algorithm more closely approximating the in situ salinity measurements (Table 1.5). The simple algorithm overestimated salinity by 26%. The multiple model overestimated salinity by 10%. The effect of temperature on the multiple algorithm improved the accuracy of the salinity estimates, because the regression coefficient for temperature ($b_2 = - 0.12$) in the 2005 multiple algorithm was relatively large compared to the 2004 model. Thus, temperature had a larger impact on the fit of modeled to measured salinity in 2005, an upwelling period. For both the downwelling and upwelling conditions, and despite the uncertainties inherent in this sort of analysis, both algorithms were robust predictors of salinity (MSPE < MSE in all instances).

## 4.0 Discussion

The synthetic salinity algorithms developed in this study will permit quantitative observations of the CRP at a higher spatial and temporal resolution than what is currently available by shipboard and mooring observations alone. No previous study has validated satellite imagery of the CRP with in situ, biogeochemically-relevant optical properties. These validated algorithms will enable the objective detection of the plume so that future analyses can describe the geophysical parameters

of the plume and not just the location of the putative plume (Thomas & Weatherbee 2006).

The CRP is a dynamic mesoscale feature of the Oregon and Washington shelf. Applying the algorithms of this study to the relatively new 250 m resolution data products from MODIS (Franz *et al.* 2006) was a deliberate choice to spatially resolve fine structure within and near the plume. The gradient in salinity along fronts and the filament extending southwestward from the plume bolus in 2004 (Fig. 1.6) would have been poorly resolved in traditional 1 km resolution MODIS data. Data richness increases 16-fold with the new 250 m resolution data products. Although the native resolution is still 1 km for the SST and CDM products, future detailed studies of the fine structure of the plume may be possible; keeping in mind that these products are interpolations of lower-resolution data.

Both empirical algorithms developed in this study generally underestimated salinity during downwelling and overestimated salinity during upwelling, with the multiple algorithm having a closer fit during upwelling. Including temperature in the multiple algorithm conferred an advantage to predicting salinity with more accuracy in 2005 and reduced the detection of "false plumes". Existing mooring arrays in the region could be used for near real-time calibration of synthetic salinity computed using daily collections of MODIS-derived $a_{dg}$ 412 nm and SST data layers.

Potential sources of error to explain why synthetic salinity diverges from in situ salinity may be associated with natural variability of the CRP; the default spectral slope parameter used in the GSM01 algorithm to estimate CDM; the off-set between

$a_{CDOM}$ and $a_{dg}$ estimates; and the quality of the underlying in situ data used to develop the synthetic salinity models. The CRP is a dynamic feature. In situ measurements used in the salinity validation were limited to a few shipboard and mooring locations. So, all of the variability of the plume may not be captured by these ground-truth measurements. Divergence from unity in slopes for both the $a_{CDM}$ vs. $a_{CDOM}$ analysis and the salinity comparisons may be partly explained by the decision to use the globally generalized spectral slope parameter, *s*, employed by GSM01. The global parameter (the default in SeaDAS) over-estimates light absorption by CDM in the southern California Current (Kudela & Chavez 2004). Modifying the algorithm with a spectral slope parameter tuned to the CRP may improve estimates of GSM01 $a_{dg}$ 350 nm, resulting in statistically similar slopes between modeled and measured salinity. However, even regional tuning of *s* may not substantially improve estimates of CDM (Kostadinov *et al.* 2007). Negative synthetic salinity is especially acute in regions where colored detrital material off-sets $a_{dg}$ from $a_{CDOM}$ estimates, and is particularly apparent at the mouth of the Columbia River which is a region of sediment re-suspension (Aguilar-Islas & Bruland 2006). CDM includes colloids and detrital material in addition to CDOM. These additional materials could contribute to higher absorption values at 350 nm and 412 nm resulting in specious negative salinity values. Derived satellite estimates for $a_{CDOM}$ relative to total absorption (using the method of Belanger *et al.* 2008) indicate that detrital material made up to 40% of the $a_{dg}$ 412 nm estimate at the mouth, but only 10% of the estimate in the plume and even less off-shore in 2004 (data not shown). Finally, another source of error in the

validation may be the quality of data used to develop the synthetic salinity models. A greater number of in situ measurements across the full range of salinity and CDOM concentrations would have improved the models by capturing more of the natural variability of the region. While the relationships deviate from unity both when comparing the simple to multiple models and when validating the model estimates to in situ measurements; all of the salinity comparisons are statistically significant and are useful to describe salinity from ocean color on the Oregon and Washington shelf.

Optical oceanographers have recognized the utility of exploiting the unique optical character of rivers to distinguish them from offshore water (Binding & Bowers 2003; Callahan *et al.* 2004; Carder *et al.* 1989; Hu *et al.* 2004; Johnson *et al.* 2003; Conmy *et al*. 2004; DelVecchio & Subramaniam 2004) and other nearby river plumes (Chen & Gardner 2004). In the U.S., the Mississippi River Plume and the Hudson River Plume have been well characterized optically (Chang *et al.* 2002; Johnson *et al.* 2003). Early work in these regions has evolved into more mature studies that integrate physical modeling with optical measurements (Oliver *et al.* 2004). Along the New Jersey shore, these models have revealed the varied and sometimes predictable movement of coastal jets (Chang *et al.* 2002) and upwelling (Johnson *et al.* 2003). Predictions of these and other coastal features have improved as a result of the integration of high resolution physical measurements with in situ and remote optical measurements (Johnson *et al.* 2003; Vasilkov *et al.* 1999).

The CRP supplies nitrogen, silica, and trace elements to the coastal margin (Carpenter & Peterson 1989; Hill & Wheeler 2002). It is an important controller of

biomass and primary productivity in the region (Small & Menzies 1981). Plume-stimulated phytoplankton biomass in this region supports extensive zooplankton populations and juvenile salmon stocks (De Robertis *et al.* 2005; Morgan *et al.* 2005). Phytoplankton productivity is tightly coupled with the physical forcing of upwelling and plume position in this region. Even brief reversals in wind direction can alter the direction of river flow (Garcia Berdeal *et al.* 2002; Hickey 1998) and this can have an impact on the distribution of standing stocks of biomass. The mechanisms that control the direction of flow and the ultimate fate of the fresh water from the plume continue to be refined (Garcia Berdeal *et al.* 2002). The canonical flow hypothesis (Hickey 1979) with northward flow during downwelling favorable winds, and southwest-ward flow during upwelling favorable winds may be an incomplete description of flow. A bifurcated plume or a condition of repeated return into the estuary may be a reality of the CRP that cannot be easily observed and verified from shipboard or drifter measurements alone

The CRP has historically been described based on salinity (Hill & Wheeler 2002; Small & Menzies 1981). Previous studies (e.g. Thomas & Weatherbee (2006)) have successfully characterized the physical characteristics of the CRP using supervised classification of ocean color data, but were unable to directly link these processes to biogeochemically relevant parameters. Our approach provides a link between the purely statistical methods of Thomas and Weatherbee (2006) with the biogeochemically significant parameter: salinity. Increased frequency of observations over larger spatial scales utilizing remote sensing will improve our understanding of

CRP flow, delivery of nutrients to the shelf, and residence time of this biological incubator so important to ecosystem health.

## 5.0 Acknowledgements

We wish to thank the captain and crew of the R/V *Wecoma* and R/V *Point Sur* for assistance with the logistics of this project; K. Bruland for donating ship space and time in June 2004 (to SP); R. Pasetto, M. Howard, A. VanderWoude, A. Roberts, M. Blakely, D. Swenson, and D. O'Gorman for field and instrument support; A. Baptista, E. Dever, and B. Hickey for mooring and oceanographic data. Two anonymous reviewers provided insightful comments, and we thank them for their time and effort. Funding for this project was provided by National Science Foundation grant #OCE-0238347. This is RISE contribution #27.

## 6. References


Aguilar-Islas A. M. & Bruland K. W. (2006) Dissolved manganese and silicic acid in the Columbia River plume: A major source to the California current and coastal waters off Washington and Oregon. *Mar Chem* 101, 233 - 247.

Babin, M. D. Stramski, G.M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner (2003) Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *J. Geophys. Res.* 108(C7), 3211, doi:10.1029/2001JC000882.

Belanger, S Babin, M., & Larouche, P. (2008) An empirical ocean color algorithm for estimating the contribution of chromophoric dissolved organic matter to total light absorption in optically complex waters. *J. Geophys. Res.* 113, C04027, doi: 10.1029/2007JC004436.

Binding C. E. & Bowers D. G. (2003) Measuring the salinity of the Clyde Sea from remotely sensed ocean colour. *Estuar Coast Shelf S* 57, 605 - 611.

Blough N. V. & Del Vecchio R. (2002) Chromophoric DOM in the Coastal Environment. In: *Biogeochemistry of marine dissolved organic matter* (eds. D. A. Hansell & C. A. Carlson) pp. 509-546. Academic Press, San Francisco.

Bricaud A., Morel A. & Prieur L. (1981) Absorption by Dissolved Organic-Matter of the Sea (Yellow Substance) in the Uv and Visible Domains. *Limnology and Oceanography* 26, 43-53.

Callahan J., Dai M., Chen R. F., Li X., Lu Z. & Huang W. (2004) Distribution of dissolved organic matter in the Pearl River Estuary, China. *Mar Chem* 89, 211 - 224.

Carder K. L., Steward R. G., Harvey G. R. & Ortner P. B. (1989) Marine Humic and Fulvic-Acids - Their Effects on Remote-Sensing of Ocean Chlorophyll. *Limnology and Oceanography* 34, 68-81.

Carder K. L., Chen, F. R., Lee, Z. P., Hawes, S. K. & Kamykowski, D. (1998) Semianalytic MODIS algorithms for chlorophyll *a* and absorption with bio-optical domains based on nitrate-depletion temperatures. *J. Geophys. Res.* 104(C3), 5403 - 5421.

Carpenter R. & Peterson M. L. (1989) Chemical cycling in Washington's coastal zone. In: *Coastal oceanography of Washington and Oregon* (eds. M. R. Landry & B. M. Hickey). Elsevier, New York.

Chang G. C., Dickey T. D., Schofield O. M., Weidemann A. D., Boss E., Pegau W. S., Moline M. A. & Glenn S. M. (2002) Nearshore physical processes and bio-optical properties in the New York Bight. *J. Geophys. Res.* 107(C9), 3133, doi:10.1029/2001JC001018.

Chen R. F., Bissett P., Coble P., Conmy R., Gardner G. B., Moran M. A., Wang X., Wells M. L., Whelan P. & Zepp R. G. (2004) Chromophoric dissolved organic matter (CDOM) source characterization in the Louisiana Bight. *Mar Chem* 89, 257 - 272.

Chen R. F. & Gardner G. B. (2004) High-resolution measurements of chromophoric dissolved organic matter in the Mississippi and Atchafalaya River plume regions. *Mar Chem* 89, 103 - 125.

Conmy R. N., Coble, P.G., Chen, R.F., Gardner, G.B. (2004) Optical properties of colored dissolved organic matter in the Norther Gulf of Mexico. *Mar Chem* 89, 127 - 144.

DelVecchio R. & Subramaniam A. (2004) Influence of the Amazon River on the surface optical properties of the western tropical North Atlantic Ocean. *J. Geophys. Res.* 109, C11001, doi:10.1029/2004JC002503.

De Robertis A., Morgan C. A., Schabetsberger R. A., Zabel R. W., Brodeur R. D., Emmett R. L., Knight C. M., Krutzikowsky G. K. & Casillas E. (2005) Columbia River plume fronts. II. Distribution, abundance, and feeding ecology of juvenile salmon. *Mar Ecol-Prog Ser* 299, 33 - 44.

Franz B. A., Werdell P. J., Meister G., Kwiatkowska E. J., Bailey S. W., Ahmad Z. & McClain C. R. (2006) MODIS land bands for ocean remote sensing applications. In: *Proc. Ocean Optics: XVIII*, Montreal, Canada.

Garcia Berdeal I., Hickey B. M. & Kawase M. (2002) Influence of wind stress and ambient flow on a high discharge river plume. *J Geophys Res-Oceans* 107(C9), 3130, doi:10.1029/2001JC000932.

Hernes P. J. & Benner R. (2003) Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. *J. Geophys. Res.* 108(C9), 3291, doi:10.1029/2002JC001421.

Hickey B., MacFadyen A., Cochlan W., Kudela R., Bruland K. & Trick C. (2006) Evolution of chemical, biological, and physical water properties in the northern California Current in 2005: Remote or local wind forcing? *Geophys Res Lett* 33 ,L22S02, doi:10.1029/2006GL026782.

Hickey B. M. (1979) The California Current System -- hypotheses and facts. *Prog Oceanogr* 8, 191 - 279.

Hickey B. M. (1989) Patterns and processes of circulation over the Washington continental shelf and slope. In: *Coastal oceanography of Washington and Oregon* (eds. M. R. Landry & B. M. Hickey) pp. 41-109. Elsevier, New York.

Hickey B. M. ed. (1998) *Coastal oceanography of Western North America from the tip of Baja California to Vancouver Island*. Wiley, New York.

Hill J. K. & Wheeler P. A. (2002) Organic carbon and nitrogen in the northern California current system: comparison of offshore, river plume, and coastally upwelled waters. *Prog Oceanogr* 53, 369 - 387.

Hu C., Montgomery E. T., Schmitt R. W. & Muller-Karger F. E. (2004) The dispersal of the Amazon and Orinoco River water in the tropical Atlantic and Caribbean Sea: Observations from space and S-PALACE floats. *Deep-Sea Research II* 51, 1151 - 1171.

Johannessen S. C., Miller W. L. & Cullen J. J. (2003) Calculation of UV attenuation and colored dissolved organic matter absorption spectra from measurements of ocean color. *J. Geophys. Res.* 108(C9), 3301, doi:10.1029/2000JC000514.

Johnson D. R., Miller J. & Schofield O. (2003) Dynamics and optics of the Hudson River outflow plume. *J Geophys Res-Oceans* 108(C10), 3323, doi:10.1029/2002JC001485.

Kachel N. B. & Smith J. D. (1989) Sediment transport and deposition on the Washington continental shelf. In: *Coastal oceanography of Washington and Oregon* (eds. M. R. Landry & B. M. Hickey) pp. 287 - 365. Elsevier, New York.

Kirk J. T. O. (1994) *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press, New York.

Klinkhammer G. P., McManus J., Colbert D. & Rudnicki M. D. (2000) Behavior of terrestrial dissolved organic matter at the continent-ocean boundary from high-resolution distributions. *Geochimica et Cosmochimica Acta* 64, 2765 - 2774.

Kostadinov, T. S. Siegel, D.A., Maritorena, S. & Guillocheau, N. (2007) Ocean color observations and modeling for an optically complex site: Santa Barbara Channel, California, USA. *J. Geophys. Res.* 112, C07011, doi:10.1029/2006JC003526.

Kudela R. M. & Chavez F. P. (2004) The impact of coastal runoff on ocean color during an El Nino year in Central California. *Deep-Sea Research Part I-Topical Studies in Oceanography* 51, 1173-1185.

Kudela R. M., Cochlan W. P., Peterson T. D. & Trick C. G. (2006) Impacts on phytoplankton biomass and productivity in the Pacific Northwest during the warm ocean conditions of 2005. *Geophys Res Lett* 33, L22S06, doi:10.1029/2006GL026772.

Lagerloef G., Colomb F. R., Le Vine D., Wentz F., Yueh S., Ruf C., Lilly J., Gunn J., Chao Y., Decharon A., Feldman G. & Swift C. (2008) The Aquarius/SAC-D Mission: Designed to meet the salinity remote-sensing challenge. *Oceanography* 21, 68 - 81.

Landry M. R., Postel J. R., Peterson W. K. & Newman J. (1989) Broad-scale distributional patterns of hydrographic variables on the Washington/Oregon shelf. In: *Coastal oceanography of Washington and Oregon* (eds. M. R. Landry & B. M. Hickey) pp. 1 - 40. Elsevier, New York.

Lee Z.P., Carder, K.L., & Arnone, R.A. (2002) Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl Optics 41(27), 5755 - 5772.

Maritorena S., Siegel D. A. & Peterson A. R. (2002) Optimization of a semi-analytical ocean color model for global-scale applications. *Appl Optics* 41, 2705 - 2714.

McCarthy K. A. & Gale R. W. (1999) Investigation of the distribution of organochlorine and polycyclic aromatic hydrocarbon compounds in the lower Columbia River using semipermeable membrane devices. In: *Water-Resources Investigations Report*. US Geological Survey, Portland, OR.

Mopper K. & Kieber, D.J. (2002) Photochemistry and the cycling of carbon, sulfur, nitrogen and phosphorus.. In: *Biogeochemistry of marine dissolved organic matter* (eds. D. A. Hansell & C. A. Carlson) pp. 455-489. Academic Press, San Francisco.

Morgan C. A., De Robertis A. & Zabel R. W. (2005) Columbia River plume fronts. I. Hydrography, zooplankton distribution, and community composition. *Mar Ecol-Prog Ser* 299, 19 - 31.

Oliver M. J., Glenn S., Kohut J. T., Irwin A. J., Schofield O. M., Moline M. A. & Bissett W. P. (2004) Bioinformatic approaches for objective detection of water masses on continental shelves. *J. Geophys. Res.* 109(C07S04), doi:10.1029/2003JC002072.

Siegel D. A., Maritorena S., Nelson N. B., Hansell D. A. & Lorenzi-Kayser M. (2002) Global distribution and dynamics of colored dissolved and detrital organic materials. *J Geophys Res-Oceans* 107(C12), 3228, doi:10.1029/2001JC000965.

Small L. F. & Menzies D. W. (1981) Patterns of primary productivity and biomass in a coastal upwelling region. *Deep-Sea Research* 28A, 123 - 149.

Stedmon C.A. & Markager, S. (2003) Behaviour of the optical properties of coloured dissolved organic matter (CDOM) under conservative mixing. *Estuar Coast Shelf* 57, 973 - 979.

Thomas A. C. & Weatherbee R. A. (2006) Satellite-measured temporal variability of the Columbia River plume. *Remote Sens Environ* 100, 167 - 178.

Twardowski M. S., Boss E., Sullivan J. M. & Donaghay P. L. (2004) Modeling the spectral shape of absorption by chromophoric dissolved organic matter. *Mar Chem* 89, 69-88.

Vasilkov A. P., Burenkov V. I. & Ruddick K. G. (1999) The spectral reflectance and transparency of river plume waters. *Int J Remote Sens* 20, 2497-2508.

# CHAPTER 2

## Optical characterization of water masses within the Columbia River Plume

Abstract:

The Columbia River Plume (CRP) is a buoyant plume that influences the Oregon and Washington shelf with the delivery of freshwater, silicic acid, trace metals, and particulate and dissolved organic matter. The highly dynamic plume contains sub-mesoscale features that have an impact on the chemistry, biology, and transport of water and material to the offshore environment. Bio-optical classification of the larger plume water mass has confirmed seasonal and annual flow patterns, but has not described the internal structure of the plume in a biogeochemically relevant way, as there were no *in situ* data to validate classification. The objectives of this study were to 1) statistically define water types within the CRP using *in situ* measurements of biogeochemically and bio-optically relevant variables, 2) to build a training data set from these water types, and 3) to apply this training dataset to 250 m resolution MODIS Aqua imagery from an oceanographically downwelling (2004) and upwelling (2006) period to predictively discriminate water masses within the plume. This study's classification technique was effective at predicting water types in the CRP. The three-variable input matrix (temperature, salinity, and chlorophyll-a fluorescence) performed better than the two-variable input matrix (temperature and salinity) at distinguishing fine-scale structure within the plume at the river mouth. Retentive features such as the plume bulge and eddies were observed at the river

mouth and on the Washington shelf. This classification approach was limited to the available continuous variables measured by ship-board, mooring, and satellite sensors. Two new classification methods are proposed that build on the framework of the classifier described here.

Key Terms: Columbia River Plume, water mass, statistical classifier, MODIS

## 1.0 Introduction

Understanding the regional variability in river plume dynamics historically has been confined to ship-board surveys and mooring observations, which are limited in time and space. The advent of new technologies, such as ocean color satellites, has fundamentally altered the spatial scale and temporal resolution at which we sample. Polar-orbiting satellites (e.g. MODIS Aqua) offer nearly daily synoptic views of entire coastal regions over periods of years. Ocean color data are readily available and can be used for a variety of purposes including identifying mesoscale features from space [*Martin Traykovski and Sosik*, 2003], understanding upper ocean dynamics [*Oliver et al.*, 2004], and river plume identification and tracking [*Chen et al.*, 2004; *Conmy et al.*, 2004; *Del Castillo et al.*, 2001; *Del Vecchio and Subramaniam*, 2004; *Oliver et al.*, 2004; *Thomas and Weatherbee*, 2006]. Development of new data products and the application of multivariate statistical techniques to those data layers have extended these observations beyond predominantly physical characteristics to

biogeochemically relevant parameters for river plumes [*Martin Traykovski and Sosik*, 2003; *Oliver et al.*, 2004].

## 1.1 Resolving River Plumes from Ocean Color Observations

The objective detection of water masses using optics signals an increasingly important area of oceanography where high resolution, *in situ* studies are costly, impractical, or not possible [*Chang et al.*, 2002; *Oliver et al.*, 2004; *Vasilkov et al.*, 1999]. A water mass can be defined as a coherent body of water with a defined set of physical or biogeochemical properties (e.g. salinity and temperature). For conservative properties such as temperature and salinity, the initial state of the property is retained until mixed with another water mass or altered through interactions with the atmosphere [*Tomczak*, 1999]. The life cycle of a water mass includes formation, evolution, and decay [*Tomczak*, 1999]. A river plume is one example of a water mass with highly variable and constantly changing physical and bio-optical properties. Changes in the plume are due to interactions with the environment and the biologically active communities contained within it. The low salinity water of river plumes can be optically distinguished from nearby marine waters due to the contrast in light absorption, $a$, by sediment particles and chromophoric dissolved organic matter (CDOM) present in the river water. The optical properties of these terrigenously sourced constituents are different from marine sources [*Blough and Del Vecchio*, 2002; *Hernes and Benner*, 2003]. The conservative relationship between salinity and $a_{\mathrm{CDOM}}$ has been used as an optical

proxy for low salinity water in the Hudson River outflow plume [*Johnson et al.*, 2003], Amazon and Orinoco River system [*Del Vecchio and Subramaniam*, 2004], Columbia River Plume [*Palacios et al.*, 2009], and the Mississippi River system [*D'Sa and Miller*, 2003].

## 1.2 The Columbia River Plume

The Columbia River drains a watershed approximately 660,480 $km^2$ that includes parts of the northwestern US and southwestern Canada. It is the largest river on the North American west coast contributing 77% of the fresh water flow to the NE Pacific Ocean from San Francisco to the Strait of Juan de Fuca [*Naik and Jay*, 2005]. Discharge averages 6,970 $m^3\ s^{-1}$ annually, and ranges from 3,000 – 17,000 $m^3\ s^{-1}$. The spring freshet is a period of maximum flow due to snowmelt [*B M Hickey*, 1998]. Flow of the Columbia River pulses tidally to the ocean where it becomes a buoyant plume [*B M Hickey*, 1989]. The river conveys fresh water, dissolved and particulate organic matter, nutrients (e.g. silica), trace metals, and pollutants to the coastal shelf region [*Aguilar-Islas and Bruland*, 2006; *Bruland et al.*, 2008; *B M Hickey*, 1989]. Delivery of nitrate to the shelf is typically minimal compared to nitrate supply from wind-driven upwelling (20 - 25 μM) [*B M Hickey and Banas*, 2008], but can be an important contributor when upwelling is weak (5 – 18 μM) [*Bruland et al.*, 2008]. Wind stress, Coriolis, and inertia influence the flow of the Columbia River Plume (CRP) as it exits the river mouth. During periods of southerly winds, downwelling conditions prevail and the plume flows northward along the Washington shelf.

Northerly winds induce upwelling and the plume flows southward and offshore [*B M Hickey*, 1989]. Even short-term (hours to days) oscillations in wind forcing can cause transient reversals in the typical seasonal flow patterns [*Garcia Berdeal et al.*, 2002; *B Hickey et al.*, 2005; *B M Hickey*, 1998; *Thomas and Weatherbee*, 2006], with reversals more common in summer than winter.

The CRP influences the shelf ecosystem by several important mechanisms. Stratification caused by the fresh water lens and retention of nutrient-rich waters on the shelf [*B M Hickey and Banas*, 2008] promote phytoplankton growth within- and at the margins of the plume [*Kudela and Peterson*, 2009]. The CRP impedes southward flow of recently upwelled water from the Washington shelf resulting in retention on the shelf and enhanced productivity [*Banas et al.*, 2009a]. The CRP is not a homogeneous water mass: a bloom formed within a decaying eddy can persist on the shelf and may drive productivity [*Banas et al.*, 2009a]. During favorable conditions, a bulge feature is established at the river mouth [*Horner-Devine*, 2009], south of the shelf retention area. This retention feature near the mouth acts as a bio-reactor for phytoplankton and can influence productivity on the shelf [*Kudela et al.*, 2010] and to off-shore waters by enhanced dispersion by the plume after the bulge decays [*Banas et al.*, 2009a]. The CRP is a structurally complex and dynamic system that is strongly influenced by wind-driven upwelling or downwelling. Within-plume water mass identification from satellite observations may therefore support the validation of models that predict the sub-mesoscale features of the plume.

## 1.3 Classification Techniques

Classification of water types using ocean color imagery has far-reaching applications for ecological and biogeochemical modeling [*Schofield et al.*, 2004]. Statistical, feature-based classification techniques have been successful in characterizing river plumes [*Oliver et al.*, 2004; *Thomas and Weatherbee*, 2006] and phytoplankton blooms of the North Atlantic [*Martin Traykovski and Sosik*, 2003; *T S Moore et al.*, 2001]. Feature-based classification techniques use the inherent characteristics of the input data, for example ship-board observations of ocean color or environmental data, to derive clusters within a multivariate statistical space. Clusters can be related to water types with their respective biogeochemical and bio-optical properties. These water type definitions can then be employed as training data to discriminate water types in satellite retrievals of ocean color or derived biogeochemical data products. Decision rules can influence the classification success. Objective methods, such as k-means clustering, are less biased [*Chang et al.*, 2002; *Oliver et al.*, 2004] as they are based on patterns naturally occurring in the data [*Martin Traykovski and Sosik*, 2003]. Similarly, other computational intelligence methods such as a neural-network approach or fuzzy-c means classification have been used to describe a migrating plume front [*Oliver et al.*, 2004] and other water masses in remote sensing imagery [*T S Moore et al.*, 2001; *Ressom et al.*, 2005].

A six year survey of SeaWiFS multispectral satellite imagery was used to characterize the seasonal and interannual variability of the Columbia River Plume [*Thomas and Weatherbee*, 2006]. The study used a maximum likelihood classifier

based on the spectral properties of normalized water leaving radiance to identify the plume and to describe its behavior. The authors identified the location of four spectrally defined water masses that were subjectively labeled “plume core”, “plume and inner-shelf water”, “offshore water”, and “other shelf water.” The classifier depended exclusively on satellite data with no *in situ* measurements to relate the remotely sensed signatures to *in situ* environmental parameters such as temperature, salinity, or chlorophyll-a biomass. Because the study lacked *in situ* validation, these labels could not be confirmed. Use of biogeochemical data in classification extends the validity of labels applied to the water masses. Despite these limitations, the study described well the canonical flow patterns of the CRP that are consistent with model predictions [*B Hickey et al.*, 2005; *B M Hickey*, 1998] and ship-board and mooring observations. Trends found in Thomas and Weatherbee [2006] were also observed in the present study and guided the decision to observe the system separately during downwelling or upwelling conditions.

Water mass classification within the larger mesoscale feature of the plume may elucidate finer-scale physical, chemical, and biological processes occurring in what may otherwise appear to be a homogeneous feature. Development of 250 m resolution algorithms [*Franz et al.*, 2006] for MODIS ocean color has allowed the analysis of fine structures within mesoscale features. The objectives of this study were 1) to statistically define water types within the Columbia River Plume using *in situ* measurements of biogeochemically and bio-optically relevant variables, 2) to build a training data set from these water types, and 3) to apply this training data set

to 250 m resolution MODIS imagery from oceanographically downwelling (17 – 22 July 2004) and upwelling (22 – 26 June 2006) periods to predictively discriminate water masses within the plume.

## 2.0 Methods

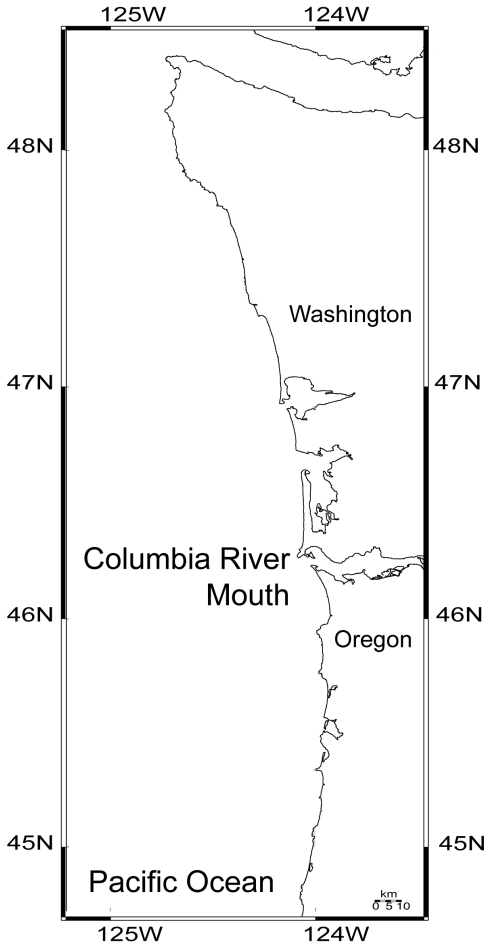



Fig. 2.1. Study Location. Coastal-shelf region offshore of Oregon and Washington, USA and under the influence of the Columbia River Plume.

### 2.1 Field Study

Field data collection occurred within the Columbia River, its estuary, and along the coastal margin of Washington and Oregon, USA as part of the Coastal Ocean

Program – River Influences on Shelf Ecosystems (CoOP RISE) project [*B M Hickey et al.*, 2010] (Fig. 2.1). The RISE project was designed to observe the Columbia River Plume at three high river discharge periods during the spring and early summer of 2004, 2005, and 2006, and one low discharge period in late summer 2005. A suite of physical, chemical, biological, and bio-optical measurements was collected during cruises aboard the *R/V Wecoma* and *R/V Point Sur*. The measurements of interest for this study included temperature, salinity, and chlorophyll-a fluorescence from 4 m depth. Only data from the *R/V Wecoma* were used in the present study. Three moorings were also deployed at locations bounding and contained within the putative flow regions of the CRP during downwelling, transition, and upwelling conditions. These locations varied between the years of this study – 2004: RISO 46.053°N 124.100°W, RICE 46.167°N 124.195°W, RINO 46.437°N 124.301°W and 2006: RISO 45.500°N 124.102°W, RICE 46.167°N 124.195°W, RINO 47.016°N 124.492°W. The moorings collected temperature, salinity, and chlorophyll-a fluorescence (in 2004 only) at three depths – surface, 5 m, and 20 m. Only the surface and 5 m measurements were used in this study. Observations of temperature, salinity, and chlorophyll-a fluorescence from the underway flow through system on the *R/V Wecoma* and from the moorings were used to build the training and validation data sets. Environmental measurements of winds, tides, and river discharge during the study period were obtained from the National Oceanographic and Atmospheric Administration (NOAA) website. Offshore winds were collected from NOAA National Data Buoy Center mooring # 46029. Tides were collected from NOAA

Tides and Currents station # 9439040 from Astoria, OR. River discharge was collected from the United States Geological Survey (USGS) river gauge station located in the Columbia River at the Beaver Army Terminal station # 14246900 near Quincy, OR. Periods of oceanographic downwelling and upwelling for the region were inferred from the wind vectors (Fig. 2.2A & B) and confirmed with the upwelling index obtained from NOAA Pacific Fisheries Environmental Laboratory. The period 17 to 21 July 2004 was used to represent an oceanographic downwelling period as underway, mooring, and imagery could be merged into one data set for the analysis. The period 22 to 26 June 2006 was used to represent an oceanographic upwelling period.

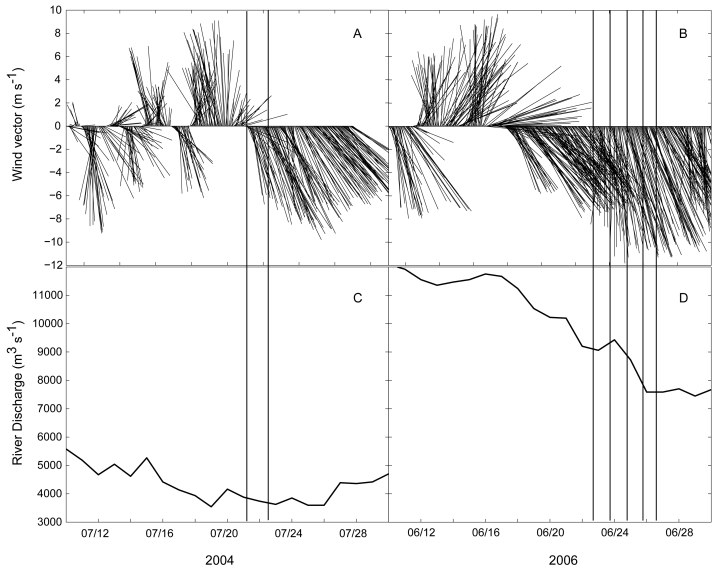



Fig. 2.2. Environmental Conditions – Winds and River Discharge. Prevailing wind direction and magnitude for 2004 (A) and 2006 (B) at National Data Buoy 46029 (46.144°N 124.51°W) located just offshore of the Columbia River mouth. River discharge for 2004 (C) and 2006 (D) measured at the Beaver substation. Vertical lines correspond with the satellite imagery capture dates used in this study (Table 2.1). Dates are Greenwich Mean Time (GMT).

## 2.2 Satellite Observations

Of the ninety-five cumulative research cruise days from 2004 to 2006, only five were completely clear of clouds in the region bounded by 44.7°N to 48°N and 123.8°W to 125.2°W. Of those five completely clear days, four were during periods of oceanographic upwelling. The one downwelling day (21 July 2004) was selected together with an additional, nearly-clear day on 22 July 2004 to determine if the water mass predictions could be tracked in time (Table 2.1). This latter date represented a period of transition from downwelling to upwelling and was characterized by poor ocean color retrieval near shore. The upwelling period used in this study did not exactly coincide with the cruise dates, as the region was completely covered in clouds during the cruise, but it did occur immediately after the end of the final research cruise. This period of relatively clear days spanned 22 to 26 June 2006 (Table 2.1).

Table 2.1. MODIS Satellite overpasses corresponding to the dates of this study.

| Year | Date | Time (GMT) | Satellite | Oceanographic Condition |
|---|---|---|---|---|
| 2004 | 21-Jul | 21:25 | MODIS-Aqua | downwelling |
| | 22-Jul | 20:30 | MODIS-Aqua | transition |
| 2006 | 22-Jun | 19:10 | MODIS-Terra | upwelling |
| | 23-Jun | 21:35 | MODIS-Aqua | upwelling |
| | 24-Jun | 20:40 | MODIS-Aqua | upwelling |
| | 25-Jun | 21:25 | MODIS-Aqua | upwelling |
| | 26-Jun | 20:30 | MODIS-Aqua | upwelling |

### 2.2.1 Image Collection and Processing

Level 0 MODIS-Aqua and –Terra swaths contained within the region of 40°N to 50°N and 123°W to 128°W were downloaded from the LAADS Web (Goddard Space Flight Center, Level 1 and Atmospheric Archive and Distribution System) for 21 and 22 July 2004, and 22, 23, 24, 25, 26 June 2006 (Table 2.1). The data were processed to Level 2 geophysical data products using SeaDAS (v. 5.2). Data were interpolated to 250 m resolution using a two wavelength atmospheric correction (1240 nm and 2130 nm) [*Franz et al.*, 2006]. Data were projected using a cylindrical projection at 250 m resolution. This higher resolution imagery was preferred as it allows for detection of finer spatial-scale, geophysical features not otherwise evident in the original 1 km resolution imagery. The data layers included sea surface temperature (SST), light absorption by detritus and CDOM at 412 nm ($a_{dg}412$) computed using GSM01 [*Maritorena et al.*, 2002], and fluorescence line height (FLH) [*Abbott and Letelier*, 2006] which provides an estimate of chlorophyll-a biomass without the optically contaminating influence of CDOM inherent in the standard chlorophyll data products. The FLH and $a_{dg}412$ estimates also use separate wavelengths, making the two data sets independent (in contrast to most chlorophyll algorithms). Data were exported from SeaDAS and then imported into MATLAB (The Mathworks, Inc.) for analysis. Synthetic salinity was computed from $a_{dg}412$ using the previously determined relationship between salinity and $a_{CDOM}$ for the CRP [*Palacios et al.*, 2009]. The relationship for July 2004 is shown in equation 1:

$$S'_{2004} = \frac{adg412 - 2.27}{-0.06} \quad (1)$$

where $S'_{2004}$ is synthetic salinity for 2004. The relationship for June 2006 is shown in equation 2:

$$S'_{2006} = \frac{adg412 - 2.97}{-0.08} \quad (2)$$

The SST, $S'_{year}$, and FLH data layers were used in the analysis.

2.3 Classification of Water Types

This study used a multivariate statistical approach to classify water types within the Columbia River Plume. It was a three-part approach that used *in situ* data to objectively create a training data set of water types from biogeochemical measurements using k-means clustering, then applied the training data set to satellite imagery to predict water type using discriminant function analysis, and finally validated the projected water masses in the satellite imagery with pixel match-ups of water masses predicted from the mooring measurements. The term "water type" refers to the classification of water in statistical space. The term "water mass" refers to the projection of that water type into geospatial space.

2.3.1 Quality Control and Data Standardization

Quality control and data standardization were the first set of three procedures executed as a part of the water mass classification technique. This was to account for outliers in the data, to mask for particular time periods or conditions (e.g. cloud cover), to re-range the input data from zero to one to account for differences in sensor type (e.g. ship-board chlorophyll-a fluorescence vs. satellite derived FLH) and differences in magnitude among variables (e.g. 25 to 30 range for salinity vs. 0 to 0.07 range for FLH) which would bias the clustering algorithm to the variable with the higher magnitude, and to account for temporal autocorrelation in the ship-board measurements used to build the training data set with the k-means classifier.

Ship-board, mooring, and satellite data layers were inspected for outliers, with only extreme instances removed (determined subjectively). These were generally attributed to bubbles in the flow-through system. Data were filtered to remove the estuaries from both underway and satellite data. The underway data were then filtered to include only a downwelling time period (15 to 16 and 17 to 21 July 2004) or an upwelling time period (12 to 17, 18 to 22 May 2006, and 5 to 12 June 2006). The data were also filtered to include only plume waters (salinity ≤ 32 [*Huyer et al.*, 2005]) in the underway, mooring, and satellite imagery, the exception being the satellite imagery from 23, 25, and 26 June 2006 when unrealistically high synthetic salinity magnitudes were predicted, presumably due to poor atmospheric correction of the remote sensing data. The plume was masked at salinity of 33 for those images. This salinity was selected after careful consideration of in-water salinity measured by the moorings at the time of satellite overpass. Data were standardized to a zero to one

range by dividing the value of an element in a variable by the maximum magnitude for that variable in a particular data set for underway, mooring, or satellite data. Assuming a linear trend in the relationship between chlorophyll-a concentration and fluorescence, re-ranging the fluorescence signal from the ship, mooring, or satellite observations obviated the need to first convert fluorescence to chlorophyll-a prior to re-ranging. Standardization to the range was used instead of normalization to the mean for two reasons: 1) while normalization would have accounted for the variance, the orders-of-magnitude difference between ship-board fluorescence and satellite FLH prevented one data set (ship-board fluorescence) from being used as a training data set for the other data set (FLH) in the discriminant function analysis step of the algorithm, 2) the range-standardized method produced the same results as the mean-normalized method in those cases in which chlorophyll fluorescence was not used (data not shown). Finally, temporal autocorrelation in the underway ship-board data was determined, using the temperature observations collected each minute, to eliminate the effect of persistence and estimate the effective sample size, n', of data to be used in building the training data set and grouping variables with the k-means classifier [*Wilks*, 1995]. The original, plume-masked data from 2004 (n = 3609) were reduced to n' = 46 or a lag of 1.3 hours. The underway data from 2006 (n= 5904) were reduced to n' = 28 for 2006 for a time lag of 3.5 hours. These smaller data sets were used to build the training data and grouping variables with the k-means classifier.

### 2.3.2 Building the Training Data

Following data quality control and standardization, a k-means clustering analysis was applied to the underway data to identify water types. For each year, two input matrices were analyzed and included either a two-variable matrix (temperature and salinity) or a three-variable matrix (temperature, salinity, and fluorescence). The k-means test was performed iteratively on each input matrix to generate three, four, or five cluster grouping outputs (hereafter labeled "cluster groupings"). As a result, twelve training groups were generated from the data (2 years x 2 input matrices x 3 cluster groupings).

The k-means clustering method is based on principal components analysis (PCA). PCA is a data reduction technique that performs an orthogonal transformation of the data and re-assigns the sample data into a new set of uncorrelated variables. Each principal component describes some portion of the variance, with the first PC having the highest variance, the second having the next highest, etc. For the k-means test, the number of cluster groupings is selected *a priori*. The number of clusters defines the number of centroids. The position of the centroids within PC-space is assigned by a weighting factor. The distance of the point's position relative to the nearest centroid defines membership of a sample point in a cluster, plotted in PC-space. Since selection of the number of clusters can be subjective, it is necessary to evaluate if the choice is appropriate for the amount of variability in the data. The k-means analysis in this study was performed in MATLAB using the MATLAB "kmeans.m" function. The random seed was re-set to 1 prior to each iteration of the

k-means test to assure continuity in clustering. The k-means test was followed by the MATLAB "silhouette.m" function to determine if an appropriate number of clusters was chosen (critical s-value ≥ 0.6). These k-means classifications became the grouping variables for the training data set used in the discriminant function analysis (DFA) to predict water types in the satellite and mooring data.

### 2.3.3 Predicting Water Types

Discriminant function analysis was used to classify water types in the satellite and mooring data using the training data derived from ship-board measurements. DFA is a classification technique that discriminates categories (or groups) from a set of independent and continuous predictor variables [*Quinn and Keough*, 2002]. The DFA was performed using the MATLAB "classify.m" function applied to the satellite and mooring data after data quality control and standardization as described above. The satellite data were grouped into two-variable (SST and synthetic salinity) and three-variable (SST, synthetic salinity, and FLH) input matrices. The mooring data for the year 2004 were grouped into two-variable (surface temperature and salinity) and three-variable (surface temperature, salinity, and fluorescence) matrices. Because no fluorescence data were available in 2006, only the two-variable input matrix was used that year for the moorings. The predicted classifications for the satellite imagery were projected geospatially to identify and map water masses of the Columbia River Plume. The mooring predictions were used in validation.

## 2.4 Validation

*In situ* classification of water mass using the mooring data was compared to predictions for the satellite imagery to evaluate whether the statistical classification was robust with two independent data sets. Comparisons were made by matching mooring and satellite data in space and time. For the mooring data, a time window of 1.5 hours before and after the satellite overpass (Table 2.1) was collected and the median of the water masses from that time period was used for validation. For the satellite match-ups, one 4 by 4-pixel box ($10^6$ $m^2$), co-located with the moorings, was obtained from the imagery. The median of the water mass for each 1 km box was used in the validation. Occasionally, clouds or atmospheric effects resulted in loss of satellite matchup data, or filtering of non-plume water resulted in the loss of both mooring and satellite data. Those match-ups were excluded from the analysis resulting in n = 36 match-ups for 2004 and n = 33 match-ups for 2006. These match-ups were used in the validation. To test the null hypothesis that there was no difference in water type prediction in satellite imagery or mooring data, a Wilcoxon Rank-Sum test was used (critical p-value = 0.05).

# 3.0 Results

## 3.1 Environmental Conditions

This study focused on two periods during the CoOP-RISE study that corresponded to oceanographic downwelling (2004) and upwelling (2006). Downwelling favorable conditions occurred 15-16 and 17-21 July 2004 (Fig. 2.2A). Upwelling favorable

conditions occurred between 12-17 and 18-22 May 2006, and 5-12 June 2006 (Fig. 2.2B). Columbia River discharge in 2006 was double the discharge in 2004 (Fig. 2.2C & D) during the study period. The annual discharge pattern (data not shown) also differed between the years. Mean daily river discharge in 2004 exhibited two distinct peak periods of flow, one in early February at 11,330 $m^3 s^{-1}$ which was followed by a period of very low flow until another peak in early June at 9710 $m^3 s^{-1}$. Discharge during the study period was approximately 3900 $m^3 s^{-1}$ (Fig. 2.2C). River discharge in 2006 reflected a more typical annual flow pattern, with mid-winter peak flow at 12,000 – 14,000 $m^3 s^{-1}$, gradually declining to 4190 $m^3 s^{-1}$ in late March. The spring freshet commenced on 10 April 2006 and ranged from 7580 to 13,220 $m^3 s^{-1}$ through the end of June 2006 (Fig. 2.2D).

The tides during the period of satellite data acquisition are presented in Fig. 2.3. The tidal cycle was at maximum flood on 21 July 2004, and at the end of ebb on 22 July 2004. This was a period of neap tides; the semi-diurnal tides were approximately equal in magnitude during this time of the month. So, while the satellite images were captured two hours apart in tidal phase, the difference in tide height was less than one meter. Image capture in 2006 occurred during spring tide and was coincident with the slack flood tide in four of the five images – 22 to 25 June 2006. The exception was 26 June 2006, which was captured during maximum flood. The tidal height was within 0.12 m for 22 to 25 June, ranging from 1.80 to 1.92 m MLLW, and measured 1.25 m above MLLW on 26 June 2006.

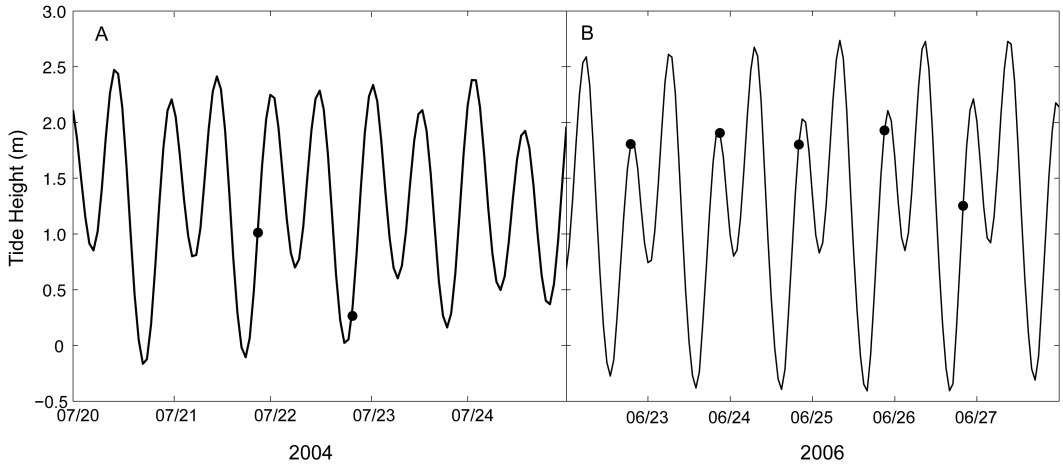



Fig. 2.3. Environmental Conditions - Tides. Tide height measurements at the Astoria tide station (ID #9439040) for 2004 (A) and 2006 (B). Data markers (•) correspond with the satellite imagery capture dates used in this study (Table 1). Dates are GMT.

3.2 *In situ* Measurements and Model Building

The results of the k-means cluster analysis for the *in situ* measurements are presented in Table 2.2 and Table 2.3. MATLAB arbitrarily assigns a cluster number to the groups. To preserve continuity across clustering methods, the cluster numbers were re-ordered by ascending salinity. Re-ordering resulted in assignment of cluster 1, 2, and 5 for the three cluster grouping k-means analysis, 1, 2, 3, 5 for the four cluster grouping analysis, and 1, 2, 3, 4, 5 for the five cluster grouping analysis.

Mean temperature, salinity, and chlorophyll-a fluorescence for each of the water type assignments in 2004 are presented in Table 2.2. For the two-variable (temperature and salinity) k-means clustering, the three, four, and five cluster groupings exhibited a similar pattern. The water type with the lowest mean salinity (22.48, 18.67, and 17.94) was also the one with the highest mean temperature. The location of water type 1 was near the river mouth. Ascending levels of salinity and

decreasing water temperature were the pattern for subsequent water types in the three, four, and five cluster groupings, with the exception of water type 5 which had elevated salinity and elevated temperature. For the three-variable input matrix (temperature, salinity, and fluorescence) k-means clustering, the pattern was less distinct. Salinity increased with increasing water type number (after ranking), however trends in temperature were less well-defined because fluorescence contributed to the division of the water types. Chlorophyll-a fluorescence was higher in the mid-salinity region of the plume than in the lowest and highest salinity regions in all of the cluster groupings. Mean temperature, salinity, and chlorophyll-a fluorescence for each of the water type assignments in 2006 are presented in Table 2.3. Despite higher river discharge, the salinity range and water type patterns in 2006 were consistent with those in 2004. Overall, temperature was lower in 2006 than in 2004. In the three-variable classification, chlorophyll-a fluorescence was dominant in the mid-salinity water types. The three, four, and five cluster groupings for both years were all valid ($s \geq 0.6$) and did not artificially subdivide the statistical space.

Table 2.2. Mean temperature, salinity, and fluorescence for the input matrices used to create the training data-set from k-means cluster analysis. Only the plume (salinity <= 32; n = 46) was used to create the water types. Measurements were collected aboard the *R/V Wecoma* during a period of oceanographic downwelling (July 17 to 21, 2004). Standard deviation in parenthesis.

| **2004 - Temperature and Salinity Input Matrix** | | | | **2004 - Temperature, Salinity, and Fluorescence Input Matrix** | | | | |
|---|---|---|---|---|---|---|---|---|
| **Cluster** | **Water Type** | **T (°C)** | **Salinity** | **Cluster** | **Water Type** | **T (°C)** | **Salinity** | **Fluor (0 - 1 range)** |
| 3 | 1 | 17.08 (0.64) | 22.48 (2.30) | 3 | 1 | 17.43 (0.55) | 19.22 (2.01) | 0.13 (0.01) |
| 3 | 2 | 15.75 (0.80) | 27.31 (1.25) | 3 | 2 | 16.28 (0.97) | 25.33 (2.09) | 0.41 (0.05) |
| 3 | 3 | n/a | n/a | 3 | 3 | n/a | n/a | n/a |
| 3 | 4 | n/a | n/a | 3 | 4 | n/a | n/a | n/a |
| 3 | 5 | 17.01 (0.63) | 29.67 (0.76) | 3 | 5 | 16.88 (0.71) | 29.05 (1.45) | 0.19 (0.06) |
| 4 | 1 | 17.53 (0.57) | 18.67 (1.84) | 4 | 1 | 17.43 (0.55) | 19.22 (2.01) | 0.13 (0.01) |
| 4 | 2 | 17.00 (0.62) | 23.64 (1.06) | 4 | 2 | 16.96 (0.65) | 23.64 (0.74) | 0.40 (0.06) |
| 4 | 3 | 15.63 (0.68) | 27.41 (1.23) | 4 | 3 | 15.48 (0.59) | 27.29 (1.19) | 0.41 (0.04) |
| 4 | 4 | n/a | n/a | 4 | 4 | n/a | n/a | n/a |
| 4 | 5 | 17.01 (0.63) | 29.67 (0.76) | 4 | 5 | 16.88 (0.71) | 29.05 (1.45) | 0.19 (0.06) |
| 5 | 1 | 17.62 (0.66) | 17.94 (1.36) | 5 | 1 | 17.43 (0.55) | 19.22 (2.01) | 0.13 (0.01) |
| 5 | 2 | 16.98 (0.61) | 23.33 (1.10) | 5 | 2 | 17.21 (0.69) | 23.45 (0.85) | 0.36 (0.05) |
| 5 | 3 | 16.07 (0.81) | 26.57 (0.67) | 5 | 3 | 16.06 (0.67) | 24.77 (1.14) | 0.46 (0.02) |
| 5 | 4 | 15.20 (0.39) | 28.53 (1.00) | 5 | 4 | 15.54 (0.65) | 27.70 (1.08) | 0.40 (0.03) |
| 5 | 5 | 17.01 (0.63) | 29.67 (0.76) | 5 | 5 | 16.88 (0.71) | 29.05 (1.45) | 0.19 (0.06) |

Table 2.3. Mean temperature, salinity, and fluorescence for the input matrices used to create the training data-set from k-means cluster analysis. Only the plume (salinity <= 32; n = 28) was used to create the water types. Measurements were collected aboard the *R/V Wecoma* during a period of oceanographic upwelling (May 18 to 22 and Jun 7 to 11, 2006). Standard deviation in parenthesis.

| **2006 - Temperature and Salinity Input Matrix** | | | | **2006 - Temperature, Salinity, and Fluorescence Input Matrix** | | | | |
|---|---|---|---|---|---|---|---|---|
| **Cluster** | **Water Type** | **T (°C)** | **Salinity** | **Cluster** | **Water Type** | **T (°C)** | **Salinity** | **Fluor (0 - 1 range)** |
| 3 | 1 | 15.43 (0.36) | 21.11 (4.23) | 3 | 1 | 15.43 (0.36) | 21.11 (4.23) | 0.25 (0.12) |
| 3 | 2 | 15.27 (0.36) | 27.36 (3.70) | 3 | 2 | 15.27 (0.36) | 27.36 (3.70) | 0.15 (0.07) |
| 3 | 3 | n/a | n/a | 3 | 3 | n/a | n/a | n/a |
| 3 | 4 | n/a | n/a | 3 | 4 | n/a | n/a | n/a |
| 3 | 5 | 13.66 (1.50) | 28.82 (4.37) | 3 | 5 | 13.66 (1.50) | 28.82 (4.37) | 0.10 (0.06) |
| 4 | 1 | 15.42 (0.36) | 21.00 (4.74) | 4 | 1 | 15.47 (0.38) | 20.68 (4.60) | 0.21 (0.07) |
| 4 | 2 | 15.45 (0.31) | 22.20 (2.15) | 4 | 2 | 15.26 (0.24) | 22.95 (0.75) | 0.44 (0.11) |
| 4 | 3 | 13.66 (1.50) | 28.82 (4.37) | 4 | 3 | 15.27 (0.36) | 27.36 (3.70) | 0.15 (0.07) |
| 4 | 4 | n/a | n/a | 4 | 4 | n/a | n/a | n/a |
| 4 | 5 | 15.19 (0.40) | 29.14 (2.81) | 4 | 5 | 13.66 (1.50) | 28.82 (4.37) | 0.10 (0.06) |
| 5 | 1 | 15.37 (0.44) | 18.38 (3.26) | 5 | 1 | 15.41 (0.49) | 19.35 (6.06) | 0.14 (0.03) |
| 5 | 2 | 15.44 (0.36) | 21.95 (4.43) | 5 | 2 | 15.45 (0.29) | 22.65 (2.05) | 0.33 (0.10) |
| 5 | 3 | 15.40 (0.19) | 23.90 (1.94) | 5 | 3 | 15.37 (0.18) | 22.92 (3.39) | 0.22 (0.04) |
| 5 | 4 | 13.66 (1.50) | 28.82 (4.37) | 5 | 4 | 13.66 (1.50) | 28.82 (4.37) | 0.10 (0.06) |
| 5 | 5 | 15.20 (0.44) | 29.48 (2.92) | 5 | 5 | 15.20 (0.44) | 29.48 (2.92) | 0.10 (0.02) |

Mean temperature, salinity, and chlorophyll-a fluorescence for each of the water type predictions in the mooring data are presented for 2004 in Table 2.4 and for 2006 in Table 2.5. These mean values fall within the training data ranges as expected from the DFA for the three and four cluster groupings in 2004 and all of the cluster groupings for 2006. As with the training data, chlorophyll-a fluorescence is higher in the mid-salinity plume waters (water types 2, 3, and 4 in the 5 cluster grouping). Overall, the range of salinity, temperature, and chlorophyll-a fluorescence values measured by the moorings over the downwelling or upwelling periods was almost as wide as those measurements collected from the ship, which covered a larger spatial scale.

Table 2.4. Mean temperature, salinity, and fluorescence for the surface and 5 m measurements collected from the RISO, RICE, and RINO moorings deployed for the CoOP-RISE project. Water type was assigned using discriminant function analysis and the training data-set from the underway data. Water type was used to validate the satellite predictions. Only the plume (salinity <= 32) was used to generate the cluster predictions from the training data. Measurements were collected from the RISE moorings during a period of oceanographic downwelling and the onset of upwelling (July 17 to 22, 2004). Standard deviation in parenthesis.

| **2004 - Temperature and Salinity Input Matrix** | | | | **2004 - Temperature, Salinity, and Fluorescence Input Matrix** | | | | |
|---|---|---|---|---|---|---|---|---|
| **Cluster** | **Water Type** | **T (°C)** | **Salinity** | **Cluster** | **Water Type** | **T (°C)** | **Salinity** | **Fluor (0 - 1 range)** |
| 3 | 1 | 16.93 (0.77) | 24.09 (1.55) | 3 | 1 | 17.37 (1.50) | 21.91 (2.20) | 0.06 (0.08) |
| 3 | 2 | 14.77 (1.09) | 28.76 (1.69) | 3 | 2 | 15.62 (1.21) | 27.79 (2.26) | 0.52 (0.15) |
| 3 | 3 | n/a | n/a | 3 | 3 | n/a | n/a | n/a |
| 3 | 4 | n/a | n/a | 3 | 4 | n/a | n/a | n/a |
| 3 | 5 | 16.50 (0.69) | 29.28 (1.40) | 3 | 5 | 16.42 (0.95) | 29.33 (1.57) | 0.19 (0.09) |
| 4 | 1 | 18.69 (0.68) | 19.78 (0.91) | 4 | 1 | 17.60 (1.16) | 21.76 (1.96) | 0.06 (0.07) |
| 4 | 2 | 16.79 (0.64) | 25.01 (1.27) | 4 | 2 | 16.60 (0.64) | 25.59 (1.37) | 0.53 (0.18) |
| 4 | 3 | 14.81 (1.14) | 28.71 (1.54) | 4 | 3 | 14.93 (1.12) | 29.39 (1.32) | 0.47 (0.13) |
| 4 | 4 | NaN | NaN | 4 | 4 | n/a | n/a | n/a |
| 4 | 5 | 16.41 (0.74) | 29.65 (1.10) | 4 | 5 | 16.47 (0.83) | 29.51 (1.24) | 0.18 (0.08) |
| 5 | 1 | 18.76 (0.66) | 19.64 (0.80) | 5 | 1 | 17.61 (1.17) | 21.72 (1.95) | 0.05 (0.07) |
| 5 | 2 | 16.93 (0.63) | 24.45 (1.13) | 5 | 2 | 17.08 (0.57) | 24.91 (1.54) | 0.32 (0.07) |
| 5 | 3 | 16.03 (0.61) | 27.23 (0.76) | 5 | 3 | 15.96 (0.80) | 26.85 (1.70) | 0.63 (0.11) |
| 5 | 4 | 14.02 (0.94) | 30.10 (0.96) | 5 | 4 | 14.85 (1.21) | 29.73 (1.25) | 0.41 (0.09) |
| 5 | 5 | 16.46 (0.72) | 29.88 (0.86) | 5 | 5 | 16.54 (0.77) | 29.47 (1.24) | 0.17 (0.08) |

Table 2.5. Mean temperature and salinity for the surface measurements collected from the RISO, RICE, and RINO moorings deployed for the CoOP-RISE project. Water type was assigned using discriminant function analysis and the training data-set from the underway data. Water type was used to validate the satellite predictions. Only the plume (salinity <= 32) was used to generate the water type predictions from the training data. Measurements were collected from the RISE moorings during a period of oceanographic upwelling that overlapped with the satellite image capture (Jun 22 to 26, 2006). Standard deviation in parenthesis. n/a = water type number not computed, n.d. = water type not present in data-set.

| **2006 - Temperature and Salinity Input Matrix** | | | |
|---|---|---|---|
| **Cluster** | **Water Type** | **T (°C)** | **Salinity** |
| 3 | 1 | 16.19 (0.16) | 11.65 (0.55) |
| 3 | 2 | 13.87 (1.40) | 19.66 (3.14) |
| 3 | 3 | n/a | n/a |
| 3 | 4 | n/a | n/a |
| 3 | 5 | 13.19 (1.65) | 28.65 (2.64) |
| 4 | 1 | 16.23 (0.09) | 11.31 (0.45) |
| 4 | 2 | 14.68 (1.17) | 17.35 (2.73) |
| 4 | 3 | 14.19 (1.32) | 24.96 (2.74) |
| 4 | 4 | n/a | n/a |
| 4 | 5 | 12.80 (1.53) | 28.90 (2.78) |
| 5 | 1 | n.d. | n.d. |
| 5 | 2 | 13.92 (1.70) | 17.43 (3.06) |
| 5 | 3 | 13.29 (1.35) | 22.13 (1.51) |
| 5 | 4 | 13.29 (1.73) | 26.19 (1.37) |
| 5 | 5 | 13.27 (1.60) | 30.53 (1.05) |

3.3 Satellite Observations

Satellite derived variables for sea surface temperature, synthetic salinity, and fluorescence line height (FLH) from 21 July 2004 are presented in Fig. 2.4. The imagery in Fig. 2.4 includes all of the pixels in the scene, but only those belonging to the plume (salinity ≤ 32) were used in the water mass classification. The sea surface temperature imagery (Fig. 2.4A) reveals warm flow from the Columbia River mouth, which extended westward as a warm core (17 to 18º C) that bifurcated along the boundary of the bulge-like feature evident in the synthetic salinity (Fig. 2.4B). Non-plume sea surface temperature located near shore, south of the mouth was cooler (15 to 16º C) relative to the warmer plume and the offshore surface waters. The northward trending plume, emergent from the Columbia River mouth (Fig. 2.4B), was consistent with the inertial forcing of the plume under downwelling conditions. Estimates of low salinity water near shore and south of the Columbia River mouth (45. 530º N, 124.0º W) or north of Grey's Harbor (46.930º N, 124.20º W) were likely due to the influence of CDOM from nearby estuaries on the synthetic salinity proxy. Fluorescence line height was elevated at the forward edge of the bulge-like feature (Fig. 2.4C). FLH was also elevated on the Washington shelf in the remnant plume. Prior to 21 July 2004, when this image was captured, downwelling favorable conditions had prevailed for four days (Fig. 2.2), resulting in the transport of low salinity water as far north as 47.90º N. The retention of this plume feature on the shelf likely resulted in the elevated FLH signal.

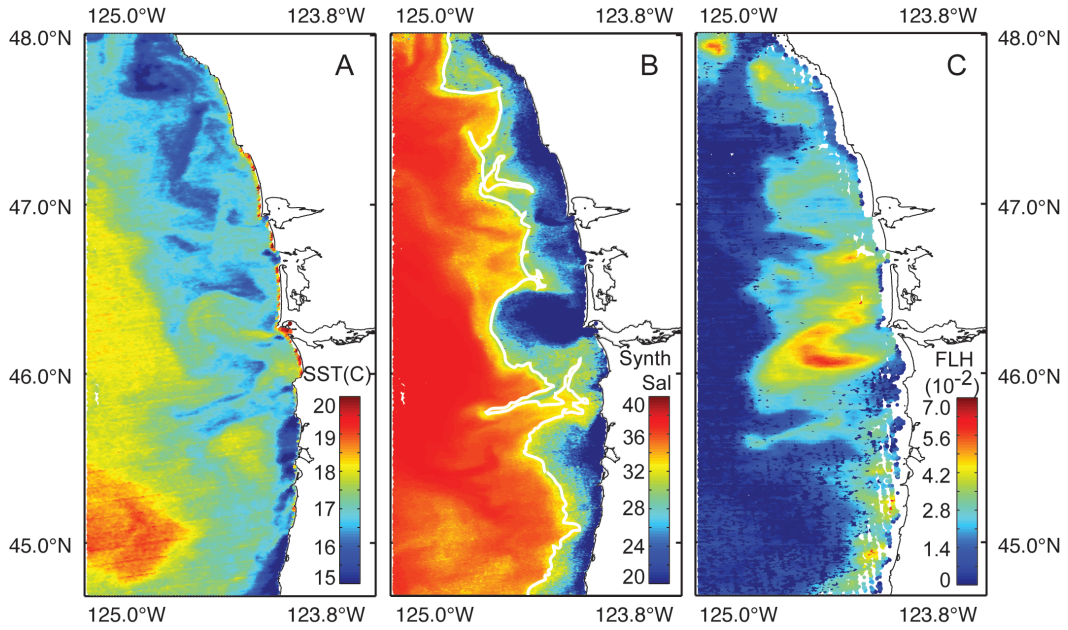



Fig. 2.4. Satellite Imagery – 21 July 2004. MODIS-Aqua satellite data layers from a period of oceanographic downwelling. Sea-surface temperature (A), synthetic salinity computed from Palacios et al. (2009) (B), and Fluorescence Line Height (C). Salinity of 32 demarcated by solid white line (—).

Satellite observations of the region of the CRP during oceanographic upwelling (24 June 2006) are presented in Fig. 2.5. Sea surface temperature was low near the Oregon coast, north of the river mouth, and in the far north of the Washington shelf during this period of intense upwelling (Fig. 2.5A). A warm core of water was extruded from the river mouth, as in 2004, and joined the warmer offshore waters. The magnitude of synthetic salinity was similar to 2004, but the distribution of low salinity water on the shelf was more widespread. The synthetic salinity field appears to resemble the bifurcated flow described elsewhere [*Garcia Berdeal et al.*, 2002; *B Hickey et al.*, 2005], and included features typically found during both downwelling and upwelling periods, such as the dipole eddy that forms just offshore and south of the Columbia River mouth [*Banas et al.*, 2009a], low salinity water on the Washington shelf near to the coast, and the south-west trending low salinity plume [*B Hickey et al.*, 2005]. The onset of upwelling winds occurred 2.5 days prior to image capture (Fig. 2.2) and so it is possible the decaying plume on the Washington shelf was still evident in the imagery.

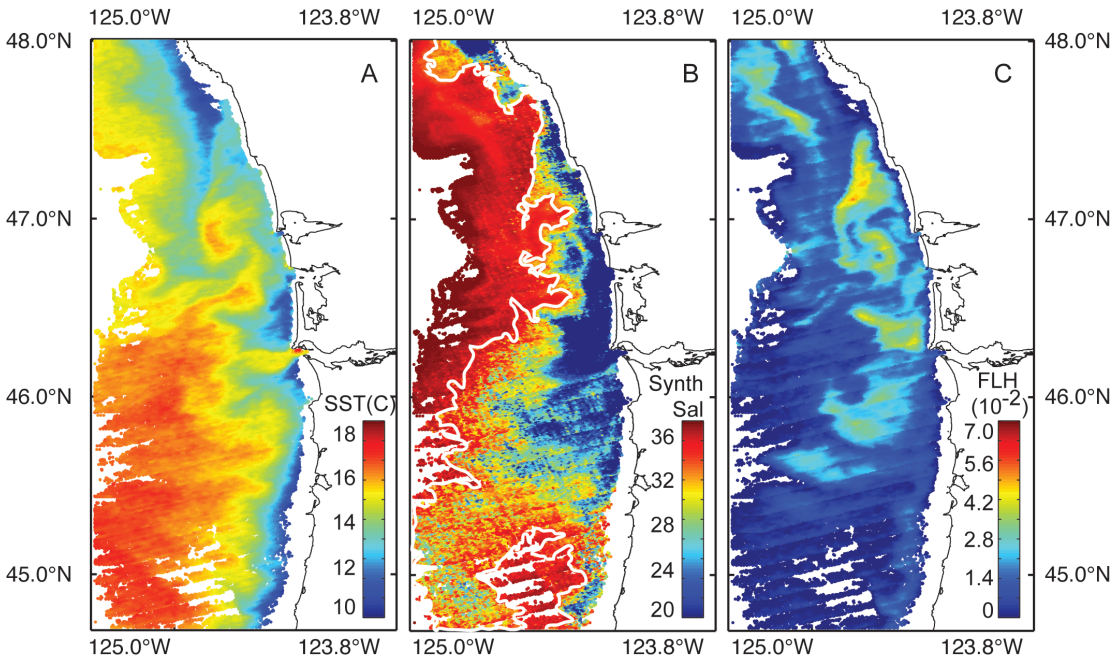
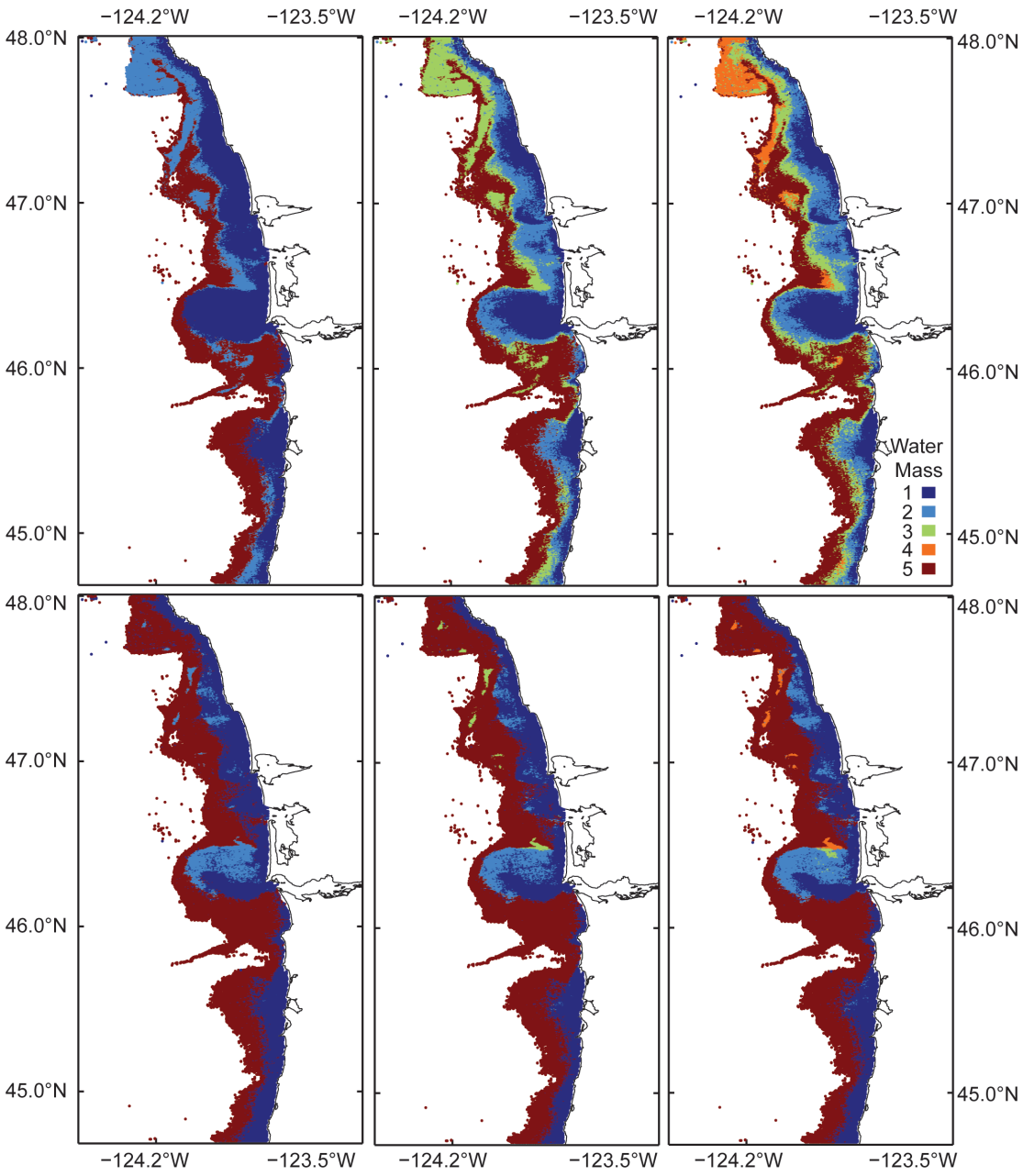



Fig. 2.5. Satellite Imagery – 24 June 2006. MODIS-Aqua satellite data layers from a period of oceanographic upwelling. Sea-surface temperature (A), synthetic salinity computed from Palacios et al. (2009) (B), and Fluorescence Line Height (C). Salinity of 32 demarcated by solid white line (—). Missing data appear as white voids in the figure, primarily located near 47ºN, 125.0ºW and scattered at 45.0ºN, 124.8ºW. These are due to clouds or atmospheric effects on image processing.

Fig 2.6. Plume Water-mass Classification – 21 July 2004. Statistically defined water masses in the Columbia River Plume determined using discriminant function analysis and a training data set from underway, ship-board measurements collected from the *R/V Wecoma* in 2004. Panels A – C represent classifications based on only temperature and synthetic salinity as the input matrix, computed for three (A), four (B), and five (C), possible cluster goupings. Panels D - F represent classifications based on temperature, synthetic salinity, and fluorescence line height as the input matrix, computed for three (D), four (E), and five (F), possible cluster groupings. Voids in the imagery are from regions of salinity greater than 32 or due to clouds.


−124.2°W
−123.5°W
48.0°N
47.0°N
46.0°N
45.0°N
Water
Mass
1
2
3
4
5

3.4 Statistical Classification of Plume Water Masses

Statistical classification of plume water-masses from 21 July 2004 is presented in Fig. 2.6. The plume was present near the coast along the Washington shelf. A tidal lens near the river mouth was clearly evident in all iterations of the DFA using the two input matrices and all cluster groupings. Patterns in the location of the two lowest salinity water masses were consistent for all cluster groupings in the two-variable input matrix analysis (Fig. 2.6A – C). Additional water masses (Fig. 2.6B & C) subdivided the salinity field between intermediate (water mass 3) and more marine (water mass 5) salinity. In the three-variable input matrix analysis, the lowest salinity water mass was confined to a smaller tongue of water within the larger tidal lens feature (Fig. 2.6D-F). The tidal lens feature was well defined in these estimates as the second freshest salinity- water mass 2 (Fig. 2.6D-F). Including FLH in the three-variable matrix had two effects on the classification of the plume in 2004: the refinement of this tongue feature in the tidal lens at the river mouth and a constraint in the number of water masses that characterized the shelf water.

Statistical classification of plume water-masses from 24 June 2006 is presented in Fig. 2.7. Because only the two-variable input matrix could be validated, only the imagery results from the two-variable input matrix are shown (Fig. 2.7). The image was captured during a period of upwelling, which followed an extended period of oceanographic downwelling (Fig. 2.2). As a result, both the remnant northward trending plume and the southward trending plume are present in the imagery (Fig. 2.5B). The lowest salinity water mass is found at and just north of the river mouth

(Fig. 2.7A-C) in all cluster-grouping predictions. The southward trending plume is evident in all three cluster-groupings (Fig. 2.7A – C), with an incremental division of the intermediate salinities among the different grouping methods. Of note for all cluster-grouping methods are the fine scale features within the plume. One of these features, located south of the tidal lens of the river mouth, may be the remnant eddy structure from the dipole eddy that is periodically established at this location [*Banas et al.*, 2009a].

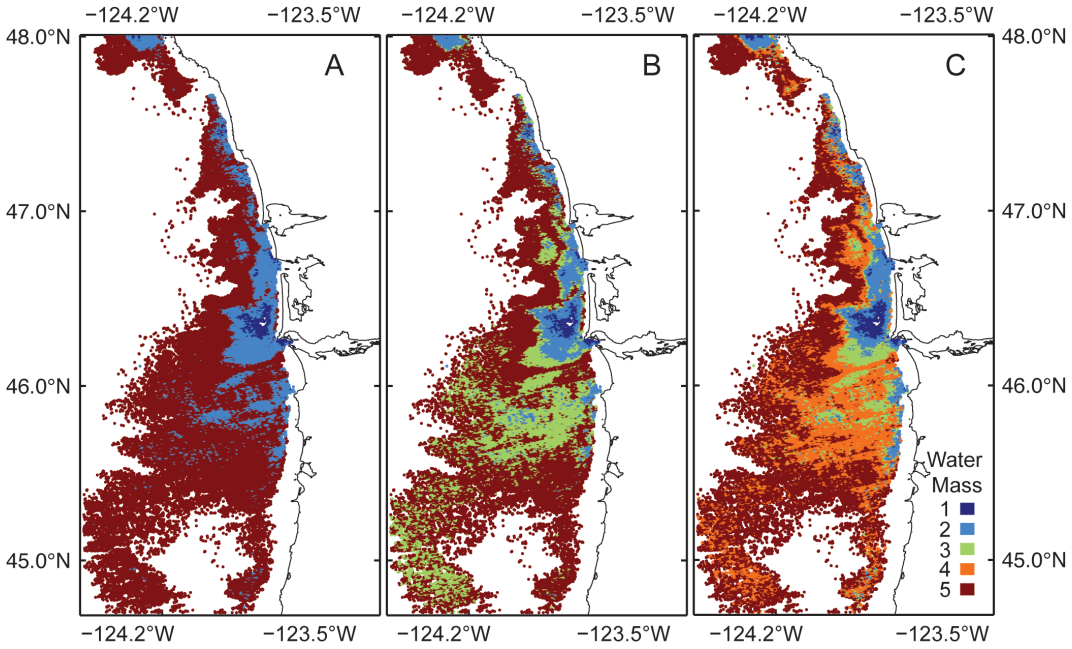



Fig 2.7. Plume Water-mass Classification – 24 June 2006. Statistically defined water masses in the Columbia River Plume determined using discriminant function analysis and a training data set from underway, ship-board measurements collected from the *R/V Wecoma* in 2006. Panels A – C represent classifications based on temperature and synthetic salinity as the input matrix, computed for three (A), four (B), and five (C), possible cluster groupings. Voids in the imagery are from regions of salinity greater than 32 or due to clouds.

The results from the two-variable input matrix with predictions for the 4 cluster groupings for the time series from 22 to 26 June 2006 are presented in Fig. 2.8. Patterns in the estimated water masses indicate a southward flowing plume on 22 June 2006 (Fig. 2.8A) with a relatively low salinity water mass (water mass 3) found offshore of the putative plume. Water mass 5 is a persistent feature along the outer Washington shelf from 22 – 25 June. In 2006, this water mass was characterized by high salinity and low temperature and could be indicative of either a well-mixed, older, plume, or recently upwelled water. If due to wind-driven, coastal upwelling, the enhanced FLH signal in Fig. 2.5C may indicate injection of nutrients to the surface layer with subsequent biological response. Water mass patterns shift on 23 June 2006 to a less defined plume (Fig. 2.8B). On 24 June 2006 (Fig. 2.8C), the low-salinity bolus of plume water is evident near the river mouth and remains a relatively intact and coherent structure within the larger plume. Intermediate salinity waters persist on the shelf. On 25 and 26 June 2006 (Fig. 2.8D&E), the fine-scale features of the river plume are less evident on the shelf. These dissipated plume water mass predictions may be partly due to masking the plume at unrealistically high salinity levels. Despite this, the shelf waters exhibit a constantly changing suite of water masses during a period of upwelling.

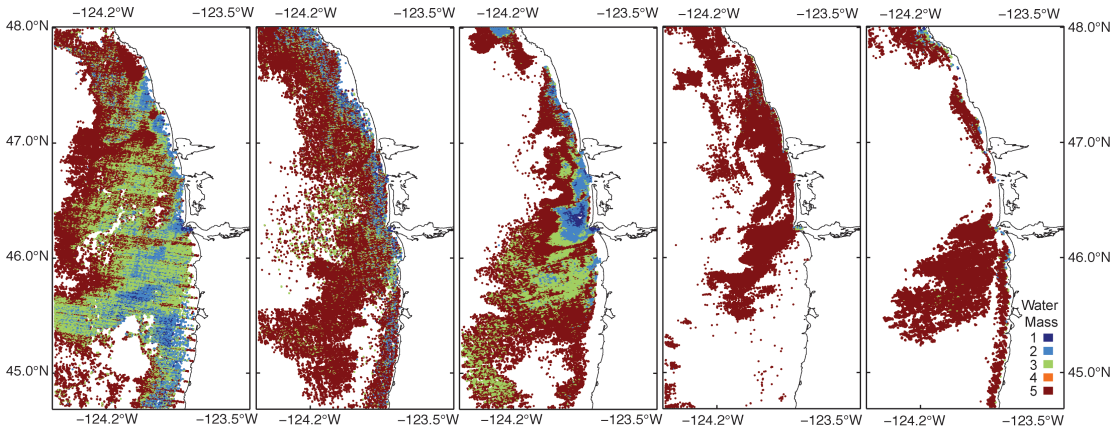



Fig 2.8. Plume Water-mass Classification – Time Series. Statistically defined water masses in the Columbia River Plume from 22 to 26 June 2006. Input matrix included temparature and salinity, and the four cluster grouping result is shown: 22 June (A), 23 June (B), 24 June (C), 25 June (D), 26 June 2006 (E).

## 3.5 Validation

The simultaneous water mass predictions for both the moorings and satellite images are presented in Table 2.6. These results were used for the validation to test the null hypothesis that there was no difference between the DFA predictions for the moorings or the satellite imagery. The results of the validation are presented in Table 2.7. There was no difference between mooring and satellite predictions for the three cluster grouping method in 2004. In 2004, there was a difference between mooring and satellite for the 4- and 5-cluster grouping method using the two-variable input matrix. The difference for the 4- and 5-cluster grouping method was solely due to the influence of the results from 22 July 2004. When the analysis was performed on only 21 July 2004, there was no difference between mooring and satellite for any of the cluster grouping methods in 2004 for either of the input matrices ($p > 0.05$ for both input matrices and all cluster groupings). Validation confirmed that there was no difference between mooring and satellite predictions for all cluster grouping methods in 2006 (Table 2.7). The fit between mooring and satellite predictions was better in 2006 than in 2004.

Table 2.6. Validation. Water mass predictions determined from DFA for the moorings and the 4 by 4-pixel box match-up from the satellite imagery.

| Date | Latitude | Longitude | Input Matrix | Mooring ID | Cluster Group | Water Mass Prediction | |
|---|---|---|---|---|---|---|---|
| | | | | | | Mooring | Satellite |
| 21-Jul-04 | 46.053 | -124.1005 | [Temp, Sal] | RISO | 3 | 5 | 5 |
| 21-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal] | RICE | 3 | 5 | 1 |
| 21-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal] | RINO | 3 | 1 | 1 |
| 21-Jul-04 | 46.053 | -124.1005 | [Temp, Sal] | RISO | 4 | 5 | 5 |
| 21-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal] | RICE | 4 | 5 | 2 |
| 21-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal] | RINO | 4 | 2 | 1 |
| 21-Jul-04 | 46.053 | -124.1005 | [Temp, Sal] | RISO | 5 | 5 | 5 |
| 21-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal] | RICE | 5 | 3 | 2 |
| 21-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal] | RINO | 5 | 2 | 1 |
| 21-Jul-04 | 46.053 | -124.1005 | [Temp, Sal, Fluor] | RISO | 3 | 5 | 5 |
| 21-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal, Fluor] | RICE | 3 | 5 | 1 |
| 21-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal, Fluor] | RINO | 3 | 5 | 2 |
| 21-Jul-04 | 46.053 | -124.1005 | [Temp, Sal, Fluor] | RISO | 4 | 5 | 5 |
| 21-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal, Fluor] | RICE | 4 | 5 | 1 |
| 21-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal, Fluor] | RINO | 4 | 2 | 2 |
| 21-Jul-04 | 46.053 | -124.1005 | [Temp, Sal, Fluor] | RISO | 5 | 5 | 5 |
| 21-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal, Fluor] | RICE | 5 | 5 | 1 |
| 21-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal, Fluor] | RINO | 5 | 2 | 2 |

Table 2.6. Continued

| Date | Latitude | Longitude | Input Matrix | Mooring ID | Cluster Group | Water Mass Prediction Mooring | Water Mass Prediction Satellite |
|---|---|---|---|---|---|---|---|
| 22-Jul-04 | 46.053 | -124.1005 | [Temp, Sal] | RISO | 3 | 5 | 1 |
| 22-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal] | RICE | 3 | 1 | 1 |
| 22-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal] | RINO | 3 | 5 | 1 |
| 22-Jul-04 | 46.053 | -124.1005 | [Temp, Sal] | RISO | 4 | 5 | 1 |
| 22-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal] | RICE | 4 | 2 | 1 |
| 22-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal] | RINO | 4 | 5 | 1 |
| 22-Jul-04 | 46.053 | -124.1005 | [Temp, Sal] | RISO | 5 | 5 | 1 |
| 22-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal] | RICE | 5 | 2 | 1 |
| 22-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal] | RINO | 5 | 5 | 1 |
| 22-Jul-04 | 46.053 | -124.1005 | [Temp, Sal, Fluor] | RISO | 3 | 5 | 1 |
| 22-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal, Fluor] | RICE | 3 | 1 | 1 |
| 22-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal, Fluor] | RINO | 3 | 5 | 1 |
| 22-Jul-04 | 46.053 | -124.1005 | [Temp, Sal, Fluor] | RISO | 4 | 5 | 1 |
| 22-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal, Fluor] | RICE | 4 | 1 | 1 |
| 22-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal, Fluor] | RINO | 4 | 5 | 1 |
| 22-Jul-04 | 46.053 | -124.1005 | [Temp, Sal, Fluor] | RISO | 5 | 5 | 1 |
| 22-Jul-04 | 46.1671 | -124.1954 | [Temp, Sal, Fluor] | RICE | 5 | 1 | 1 |
| 22-Jul-04 | 46.4374 | -124.3013 | [Temp, Sal, Fluor] | RINO | 5 | 5 | 1 |

Table 2.6. Continued

| Date | Latitude | Longitude | Input Matrix | Mooring ID | Cluster Group | Water Mass Prediction Mooring | Water Mass Prediction Satellite |
|---|---|---|---|---|---|---|---|
| 22-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 3 | 5 | 2 |
| 22-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 3 | 2 | 5 |
| 22-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 3 | 5 | 5 |
| 22-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 4 | 5 | 3 |
| 22-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 4 | 3 | 3 |
| 22-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 4 | 5 | 4 |
| 22-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 5 | 5 | 3 |
| 22-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 5 | 3 | 4 |
| 22-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 5 | 5 | 4 |
| 23-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 3 | 5 | 5 |
| 23-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 3 | 2 | 5 |
| 23-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 3 | 5 | 5 |
| 23-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 4 | 5 | 5 |
| 23-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 4 | 2 | 5 |
| 23-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 4 | 5 | 5 |
| 23-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 5 | 5 | 4 |
| 23-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 5 | 3 | 5 |
| 23-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 5 | 5 | 5 |
| 24-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 3 | 5 | 5 |
| 24-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 3 | 2 | 2 |
| 24-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 3 | 5 | n.d. |
| 24-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 4 | 5 | 5 |
| 24-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 4 | 2 | 2 |
| 24-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 4 | 5 | n.d. |
| 24-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 5 | 5 | 4 |
| 24-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 5 | 3 | 3 |
| 24-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 5 | 5 | n.d. |

Table 2.6. Continued

| Date | Latitude | Longitude | Input Matrix | Mooring ID | Cluster Group | Water Mass Prediction | |
|---|---|---|---|---|---|---|---|
| | | | | | | Mooring | Satellite |
| 25-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 3 | n.d. | n.d. |
| 25-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 3 | 2 | 5 |
| 25-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 3 | 5 | 5 |
| 25-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 4 | n.d. | n.d. |
| 25-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 4 | 3 | 5 |
| 25-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 4 | 5 | 5 |
| 25-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 5 | n.d. | n.d. |
| 25-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 5 | 3 | 5 |
| 25-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 5 | 5 | 5 |
| 26-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 3 | n.d. | n.d. |
| 26-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 3 | 5 | 5 |
| 26-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 3 | 5 | n.d. |
| 26-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 4 | n.d. | n.d. |
| 26-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 4 | 5 | 5 |
| 26-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 4 | 5 | n.d. |
| 26-Jun-06 | 45.5003 | -124.1029 | [Temp, Sal] | RISO | 5 | n.d. | n.d. |
| 26-Jun-06 | 46.1668 | -124.1955 | [Temp, Sal] | RICE | 5 | 4 | 5 |
| 26-Jun-06 | 47.0166 | -124.4921 | [Temp, Sal] | RINO | 5 | 5 | n.d. |

Table 2.7. Validation of water-mass identification between MODIS-derived estimates and in situ, mooring estimates for July 2004 and Jun 2006. Two- and three- variable input matrices tested separately for each of three, four, or five cluster grouping. H0 rejected at critical p-value less than 0.05.

| Date | Input Matrix | Cluster Grouping | Rank Sum | p-value | H0 |
|---|---|---|---|---|---|
| 2004 | [Temp, Sal] | 3 | 48.0 | 0.24 | 0 |
| | [Temp, Sal] | 4 | 52.0 | 0.05 | 1 |
| | [Temp, Sal] | 5 | 51.5 | 0.05 | 1 |
| | [Temp, Sal, Fluor] | 3 | 50.5 | 0.08 | 0 |
| | [Temp, Sal, Fluor] | 4 | 49.5 | 0.13 | 0 |
| | [Temp, Sal, Fluor] | 5 | 49.5 | 0.13 | 0 |
| 2006 | [Temp, Sal] | 3 | 115.5 | 0.37 | 0 |
| | [Temp, Sal] | 4 | 123.5 | 0.85 | 0 |
| | [Temp, Sal] | 5 | 125.0 | 0.94 | 0 |

## 4.0 Discussion

### 4.1 Classifier Reveals Plume Structure

The Columbia River Plume water mass identified in this analysis forms a coherent structure on the Oregon and Washington shelf during both downwelling (Fig. 2.4B) and upwelling (Fig. 2.5B), consistent with published observations or model predictions [*Banas et al.*, 2009a; *Garcia Berdeal et al.*, 2002; *B M Hickey*, 1989; *B M Hickey and Banas*, 2008; *B M Hickey et al.*, 2010; *Thomas and Weatherbee*, 2006]. Patterns in distribution of low salinity water differ during downwelling and upwelling. The downwelling plume forms a compressed water mass near shore with a bulge-like feature at the river mouth [*Horner-Devine*, 2009]. The enhanced chlorophyll-a signal observed at the southwest edge of the plume was likely fueled by non-linear mixing of nutrient rich waters to the surface [*Jay et al.*, 2009]. During upwelling the plume is more diffuse and extends offshore and to the south, with some low salinity water near shore.

Water mass classification revealed a complex structure of surface water masses not readily apparent in the satellite-derived salinity or temperature images. During 2006 a bi-directional plume was evident [*B Hickey et al.*, 2005], which was likely acting as a retentive feature for the water upwelled along the Washington shelf [*Banas et al.*, 2009a]. Thus our water mass classification approach has revealed physically and biogeochemically relevant patterns within the plume that are consistent with previous observations. If salinity is a proxy for time since the plume departed the river mouth [*W S Moore and Krest*, 2004], then water masses

representing the putative "new" plume water and intermediate-age plume are evident, as well as recently upwelled water and other shelf water. For example, the low-salinity, moderate chlorophyll-a signal water mass, identified as water mass 2, was repeatedly present in the near-field region of the plume (Fig. 2.7) or in the tidal lens when that feature was present (Fig. 2.6). This water mass is ecologically important as a region of enhanced primary productivity [*Kudela and Peterson*, 2009] but lacks the grazing pressure [*Peterson and Peterson*, 2009] associated with older (higher salinity) regions of the plume. While water mass 2 could be distinguished with temperature and salinity alone, the addition of FLH helped to refine the boundaries of the water masses (Fig. 2.6) and contributed to a better fit of the cluster predictions in the validation. This demonstrates that the classifier is not simply defining gradients in salinity.

Increased spatial resolution afforded by 250 m MODIS ocean data products enables the analysis of fine scale structure in mesoscale features. Because this study examined ephemeral features of a dynamic system, daily images were needed; compositing over several days was not an option. Therefore, cloud cover impeded the analysis and limited the number of available days that could be examined. Nonetheless, n-dimensional analysis on 250 m MODIS data products is possible for the Columbia River Plume and smaller plume systems of the US west coast with the understanding that imagery may be scant due to atmospheric effects, and may thus be biased to those periods when clear satellite imagery is available. Water mass classification techniques have been applied to describe patterns in space [*Martin*

*Traykovski and Sosik*, 2003; *T S Moore et al.*, 2001], and have been used to describe the migration of salinity fronts in time [*Oliver et al.*, 2004]. Daily satellite overpasses make this a possibility in many regions of the world, especially locations where cloud cover is less persistent, and could be applied to time-averaged imagery for regions that evolve over longer time periods. The Pacific Northwest of the US west coast is not the ideal location for a satellite-derived time series that depends on daily satellite imagery to track the formation, evolution, and decay of a river plume. Problems related to atmospheric correction interfered with estimates of light absorption by colored dissolved and particulate matter at 412 nm ($a_{dg}412$), resulting in unrealistically high salinity. This problem was partially alleviated during data standardization, but may have contributed to the ambiguous results of the time series, especially on 25 and 26 June 2006. Further work is needed to perform this sort of time series analysis using daily images. A region with episodic storms and periods of clear sky between storms may be better suited for testing a time series approach using water mass classification based on observations. One suitable region on the US west coast would be the Santa Barbara Channel during the winter storm season.

One limitation of the statistical approach used in this study was related to the low number of continuous variables used to develop the training data set from underway measurements, and hence the satellite and mooring variables used in the discriminant function analysis. This choice was deliberate as the measurements used to create the training data set needed to match the mooring observations used for validation. Both ship-board measurements and derived satellite products had a greater

diversity of independent variables than did the moorings. Possible additional classifiers could include backscattering, attenuation, and particulate light absorption. Despite this limitation in building the training data, the DFA approach effectively predicted water type for all cluster groupings. An alternative validation approach would involve randomly sub-setting the ship-board observations prior to model building, and then validating using those data. This approach was rejected because the test and validation data sets could not be considered independent of each other. While the moorings provide truly independent data from the satellite observations, the number of moorings and physical spacing among those moorings provides another hindrance. The three moorings were located near the river mouth. Though they encountered all of the water mass predictions in the three, four, and five cluster grouping analyses, the lowest, intermediate, and highest salinity predictions (water masses 1, 2, and 5) were most common due to the proximity to the river mouth. Wider geographic distribution of validation data from moorings may have increased the water masses encountered.

## 4.2 Performance of Classification Predictions

The discriminant function analysis approach to water type classification was successful for both the two-variable and three-variable input matrices and for all cluster groupings for both the satellite and mooring predictions (Table 2.6), when excluding 22 July 2004 from the analysis. When evaluating the performance for the two-variable relative to the three-variable approach in 2004, the addition of

chlorophyll-a fluorescence doubled the number of correct match-ups between satellite and mooring predictions. There was no bias for a particular cluster grouping in either 2004 or 2006. More widely spaced moorings or even measurements from an autonomous underwater vehicle would improve statistical power of the validation.

## 4.3 Future Directions

This study demonstrates the feasibility of using an unsupervised, statistical approach to classifying water types within the CRP with only a cursory regard to biogeochemical differences among the water types. As a first approximation, salinity can be qualitatively used as a proxy for time since exiting the river mouth. Assigning labels to the water types (e.g. “new plume”, “old plume”, etc.) is risky as that information is not present in the data used to build the training set. Using “low-salinity” is more appropriate as no time estimate was independently measured [*W S Moore and Krest*, 2004].

We propose two directions for future development of feature-based water type classification for the Columbia River Plume – an empirical, spatial approach incorporating more biogeochemical variables and a theoretical, model approach to estimate residence time. The new approaches would modify the development of the training data set from a purely statistical, unsupervised approach to one that includes *a priori* knowledge of the environment and biology of the system. The k-means approach used in this study was an unbiased method to distinguish water types using natural tendencies in the data to aggregate around the centroids in statistical space; as

such, the clusters do not necessarily have any direct correspondence to biogeochemical properties of the ocean. By adding a limited degree of subjectivity to the decision rules governing the training set, the new techniques may produce more ecologically and biogeochemically meaningful results. Specifically, we propose development of modified spatial and residence-time statistical models.

The proposed spatial approach differs from the one in the present study as it would incorporate more data variables into the training data set and DFA. Group membership and ranges in magnitude of the training data would be established subjectively based on published data [*Bruland et al.*, 2008]. The chemical properties of the river, estuary, near-field plume, non-plume coastal water, and offshore end-member water masses are well defined for the Columbia River region [*Aguilar-Islas and Bruland*, 2006; *Bruland et al.*, 2008]. The nutrients nitrate, silicic acid, dissolved iron, and dissolved manganese are related to temperature and salinity in these regions [*Bruland et al.*, 2008; *B Hickey et al.*, 2005; *B M Hickey and Banas*, 2008] and therefore could be modeled using multiple linear regression. This proposed classification approach would build a training data set from the ranges of *in situ*, measured salinity, temperature, and nutrients expected for each water mass during upwelling or downwelling conditions. These best-fit equations from the linear regression could be used with the satellite estimates of sea surface temperature and salinity to compute data layers representing nutrient concentrations. These image layers then would be used in the DFA to predict the location of the water masses spatially using the ecologically and biogeochemically relevant training parameters.

Validation could be performed using ship-board observations. While the propagation of error would magnify through each step of the proposed technique, this method could provide useful information to better understand the physical and chemical mechanisms leading to changes in the distribution of phytoplankton biomass and productivity. This method would increase the statistical power of the DFA, as the number of input variables (salinity, temperature, nitrate, silicic acid, dissolved iron, etc.) would exceed the number of groups being classified.

The proposed residence time approach differs from the classifier in the present study as it would use model estimates for a suite of biogeochemical variables evolving through time as the training data set. Instead of water types being defined statistically based on similar patterns in the observations, these groupings would be defined by the range of properties modeled for each time-step in the model output. Predicted properties at a time increment, instead of similarity by water type, would describe these new "time types". The data types generated from the model would need to be physical or bio-optical with analogues in the suite of data products available from ocean color imagery. These could include FLH for a chlorophyll-a biomass estimate, bio-volume to partition the chlorophyll pool into size classes [*Kostadinov et al.*, 2009], synthetic salinity, and modeled nutrients. For example, the method could be applied to imagery of the plume region using model estimates for phytoplankton growth and grazing for the CRP [*Banas et al.*, 2009b]. The DFA, performed on an image from a single day, could produce age estimates for water masses within the plume. This hybrid, model-observation approach creates the

possibility of estimating water age from even a single static image of the plume region. A limitation of this approach would be the quality of the model used to generate the age-clusters, the problems inherent in error propagation, and the difficulty of validating the predictions.

## 5.0 Conclusions

The water mass classifier developed in this study is the first to statistically differentiate the sub-mesoscale features of the CRP in a biogeochmically relevant way using 250 m MODIS ocean color imagery. In contrast to previous work [*Thomas and Weatherbee*, 2006] this classifier describes the water mass features in the plume and on the shelf in the same units used by observational and theoretical studies. These results are therefore viable as a data source to validate circulation and productivity models. Limitations included few data variables available from the various sensor platforms to use in building and testing the classifier, the sensitivity of the classifier to atmospheric correction artifacts, and the persistent cloud cover in the region where the classifier was developed. Despite the limitations, the classifier was effective at distinguishing sub-mesoscale features within the plume and the shelf waters including remnants of eddies, bulge-like structure, and recently upwelled filaments of water along the coast. These features were distinguished solely based on unsupervised, multivariate statistics with no manipulation of decision rules, *a priori*. Two alternate approaches were proposed that do subjectively group the training data in a classifier based on prior knowledge of the environment. These two techniques may better

inform our understanding of the mechanisms driving productivity within the plume or contribute to our ability to accurately estimate the age of the plume using ocean color satellite imagery.

Acknowledgements:

We wish to thank the captains and crews of the *R/V Wecoma* and *R/V Point Sur* for assistance with the logistics of this project; K. Bruland for donating ship space and time in June 2004 (to S.P.); R. Pasetto, M. Howard, A. VanderWoude, A. Roberts, M. Blakely, D. Swenson, and D. O'Gorman for field and technical support; E. Dever, and B. Hickey for mooring and oceanographic data, and G.J. Smith and P. Raimondi for guidance on the statistics. Funding for this project was provided by National Science Foundation grant OCE-0238347 and NASA grant NNX09AT01G-01-B.

6. Literature Cited

Abbott, M. R., and R. M. Letelier (2006), Algorithm theoretical basis document: Chlorophyll fluorescence (MODIS product number 20)*Rep.*

Aguilar-Islas, A. M., and K. W. Bruland (2006), Dissolved manganese and silicic acid in the Columbia River plume: A major source to the California current and coastal waters off Washington and Oregon, *Marine Chemistry*, *101*, 233 - 247.

Banas, N. S., P. MacCready, and B. M. Hickey (2009a), The Columbia River plume as cross-shelf exporter and along-coast barrier, *Cont. Shelf Res.*, *29*(1), 292-301.

Banas, N. S., E. J. Lessard, R. M. Kudela, P. MacCready, T. D. Peterson, B. M. Hickey, and E. Frame (2009b), Planktonic growth and grazing in the Columbia River plume region: A biophysical model study, *J Geophys Res-Oceans*, *114*, 21.

Blough, N. V., and R. Del Vecchio (2002), Chromophoric DOM in the Coastal Environment, in *Biogeochemistry of marine dissolved organic matter*, edited by D. A. Hansell and C. A. Carlson, pp. 509-546, Academic Press, San Francisco.

Bruland, K. W., M. C. Lohan, A. M. Aguilar-Islas, G. J. Smith, B. Sohst, and A. Baptista (2008), Factors influencing the chemistry of the near-field Columbia River plume: Nitrate, silicic acid, dissolved Fe, and dissolved Mn, *J Geophys Res-Oceans*, *113*, 23.

Chang, G. C., T. D. Dickey, O. M. Schofield, A. D. Weidemann, E. Boss, W. S. Pegau, M. A. Moline, and S. M. Glenn (2002), Nearshore physical processes and bio-optical properties in the New York Bight, *J. Geophys. Res.*, *107(C9), 3133*, doi:10.1029/2001JC001018.

Chen, R. F., P. Bissett, P. Coble, R. Conmy, G. B. Gardner, M. A. Moran, X. Wang, M. L. Wells, P. Whelan, and R. G. Zepp (2004), Chromophoric dissolved organic matter (CDOM) source characterization in the Louisiana Bight, *Marine Chemistry*, *89*, 257 - 272.

Conmy, R. N., P. G. Coble, R. F. Chen, and G. B. Gardner (2004), Optical properties of colored dissolved organic matter in the Northern Gulf of Mexico, *Marine Chemistry*, *89*, 127 - 144.

D'Sa, E. J., and R. L. Miller (2003), Bio-optical properties in waters influenced by the Mississippi River during low flow conditions, *Remote Sensing of Environment*, *84*, 538 - 549.

Del Castillo, C. E., P. G. Coble, R. N. Conmy, F. E. Muller-Karger, L. Vanderbloemen, and G. A. Vargo (2001), Multispectral in situ measurements of organic matter and chlorophyll fluorescence in seawater: Documenting the intrusion of the Mississippi River plume in the West Florida Shelf, *Limnology and Oceanography*, *46*(7), 1836-1843.

Del Vecchio, R., and A. Subramaniam (2004), Influence of the Amazon River on the surface optical properties of the western tropical North Atlantic Ocean, *J Geophys Res-Oceans*, *109*(C11), 13.

Franz, B. A., P. J. Werdell, G. Meister, E. J. Kwiatkowska, S. W. Bailey, Z. Ahmad, and C. R. McClain (2006), MODIS land bands for ocean remote sensing applications, paper presented at Proc. Ocean Optics: XVIII, Montreal, Canada, 9 - 13 October 2006.

Garcia Berdeal, I., B. M. Hickey, and M. Kawase (2002), Influence of wind stress and ambient flow on a high discharge river plume, *J Geophys Res-Oceans*, *107(C9), 3130*, doi:10.1029/2001JC000932.

Hernes, P. J., and R. Benner (2003), Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments, *J. Geophys. Res.*, *108(C9), 3291*, doi:10.1029/2002JC001421.

Hickey, B., S. Geier, N. Kachel, and A. F. MacFadyen (2005), A bi-directional river plume: The Columbia in summer, *Cont. Shelf Res.*, *25*(14), 1631-1656.

Hickey, B. M. (1989), Patterns and processes of circulation over the Washington continental shelf and slope, in *Coastal oceanography of Washington and Oregon*, edited by M. R. Landry and B. M. Hickey, pp. 41-109, Elsevier, New York.

Hickey, B. M. (Ed.) (1998), *Coastal oceanography of Western North America from the tip of Baja California to Vancouver Island*, 345 - 393 pp., Wiley, New York.

Hickey, B. M., and N. S. Banas (2008), Why is the Northern End of the California Current System So Productive?, *Oceanography*, *21*(4), 90-107.

Hickey, B. M., et al. (2010), River Influences on Shelf Ecosystems: Introduction and synthesis, *J Geophys Res-Oceans*, *115*, 26.

Horner-Devine, A. R. (2009), The bulge circulation in the Columbia River plume, *Cont. Shelf Res.*, *29*(1), 234-251.

Huyer, A., J. H. Fleischbein, J. Keister, P. M. Kosro, N. Perlin, R. L. Smith, and P. A. Wheeler (2005), Two coastal upwelling domains in the northern California Current system, *J. Mar. Res.*, *63*(5), 901-929.

Jay, D. A., J. Pan, P. M. Orton, and A. R. Horner-Devine (2009), Asymmetry of Columbia River tidal plume fronts, *J. Mar. Syst.*, *78*(3), 442-459.

Johnson, D. R., J. Miller, and O. Schofield (2003), Dynamics and optics of the Hudson River outflow plume, *J Geophys Res-Oceans*, *108(C10), 3323*, doi:10.1029/2002JC001485.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2009), Retrieval of the particle size distribution from satellite ocean color observations, *J Geophys Res-Oceans*, *114*.

Kudela, R. M., and T. D. Peterson (2009), Influence of a buoyant river plume on phytoplankton nutrient dynamics: What controls standing stocks and productivity?, *J Geophys Res-Oceans*, *114*, 15.

Kudela, R. M., et al. (2010), Multiple trophic levels fueled by recirculation in the Columbia River plume, *Geophys. Res. Lett.*, *37*, 7.

Maritorena, S., D. A. Siegel, and A. R. Peterson (2002), Optimization of a semi-analytical ocean color model for global-scale applications, *Applied Optics*, *41*, 2705 - 2714.

Martin Traykovski, L. V., and H. M. Sosik (2003), Feature-based classification of optical water types in the Northwest Atlantic based on satellite ocean color data, *J. Geophys. Res.*, *108*(C5), -.

Moore, T. S., J. W. Campbell, and H. Feng (2001), A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms, *Ieee Transactions on Geoscience and Remote Sensing*, *39*(8), 1764-1776.

Moore, W. S., and J. Krest (2004), Distribution of Ra-223 and Ra-224 in the plumes of the Mississippi and Atchafalaya Rivers and the Gulf of Mexico, *Marine Chemistry*, *86*(3-4), 105-119.

Naik, P. K., and D. A. Jay (2005), Estimation of Columbia River virgin flow: 1879 to 1928, *Hydrol. Process.*, *19*(9), 1807-1824.

Oliver, M. J., S. Glenn, J. T. Kohut, A. J. Irwin, O. M. Schofield, M. A. Moline, and W. P. Bissett (2004), Bioinformatic approaches for objective detection of water masses on continental shelves, *J. Geophys. Res.*, *109(C07S04)*, doi:10.1029/2003JC002072.

Palacios, S. L., T. D. Peterson, and R. M. Kudela (2009), Development of synthetic salinity from remote sensing for the Columbia River plume, *J Geophys Res-Oceans*, *114*, 14.

Peterson, J. O., and W. T. Peterson (2009), Influence of the Columbia River plume on cross-shelf transport of zooplankton, *J Geophys Res-Oceans*, *114*, 11.

Quinn, G. P., and M. J. Keough (2002), *Experimental design and data analysis for biologists*, xvii, 537 p. pp., Cambridge University Press, Cambridge, UK ; New York.

Ressom, H., R. L. Miller, P. Natarajan, and W. H. Slade (2005), Computational intelligence and its application in remote sensing, in *Remote Sensing of Coastal Aquatic Environments*, edited by R. L. Miller, pp. 205 - 227, Springer, Netherlands.

Schofield, O., T. Bergmann, M. J. Oliver, A. Irwin, G. Kirkpatrick, W. P. Bissett, M. A. Moline, and C. Orrico (2004), Inversion of spectral absorption in the optically complex coastal waters of the Mid-Atlantic Bight, *J Geophys Res-Oceans*, *109(C12), C12S04, doi*, 10.1029/2003JC002071.

Thomas, A. C., and R. A. Weatherbee (2006), Satellite-measured temporal variability of the Columbia River plume, *Remote Sensing of Environment*, *100*, 167 - 178.

Tomczak, M. (1999), Some historical, theoretical and applied aspects of quantitative water mass analysis, *J. Mar. Res.*, *57*(2), 275-303.

Vasilkov, A. P., V. I. Burenkov, and K. G. Ruddick (1999), The spectral reflectance and transparency of river plume waters, *Int J Remote Sens*, *20*(13), 2497-2508.

Wilks, D. S. (1995), *Statistical Methods in Atmospheric Sciences: An Introduction*, 467 pp., Academic Press, San Diego.

# CHAPTER 3

## Discrimination of phytoplankton taxa in an optically complex aquatic environment


### Abstract

A new hyperspectral bio-optical algorithm has been developed to discriminate phytoplankton taxa in optically complex, case 2 waters. The semi-analytical, phytoplankton detection with optics (PHYDOTax) algorithm is based on first principles of bio-optics with possible applications to biogeochemical modeling, testing of plankton functional type (PFT) models, and detection and monitoring of harmful algal blooms. A signature library of remote sensing reflectance ($R_{rs}$) spectra for seven major phytoplankton groups (diatoms, dinoflagellates, haptophytes, chlorophytes, cryptophytes, cyanophytes, and unspecified phycocyanin-containing picoeukaryotes – UPCE) was developed using measured and modeled inherent optical properties as inputs to the radiative transfer equations. Normalized $R_{rs}$ spectra were sub-setted to10 nm resolution from 455 nm to 675 nm to create the signature library. This library and the inverse-matrix-based decomposition algorithm, PHYDOTax, were used to discriminate taxon-specific biomass in both synthetic phytoplankton mixtures and field samples from Monterey Bay, CA in 2006, 2008, 2009, and 2010. Validation with the synthetic mixtures showed strong correlation between algorithm predictions and known mixture proportions for all taxa but one, *Emiliana huxleyi*. Field validation demonstrated a strong correlation between measured and modeled taxon-specific biomass for diatoms, dinoflagellates, haptophytes, chlorophytes, and

UPCE (but not cryptophytes). Cyanophytes could not be field validated. PHYDOTax was applied to hyperspectral imagery for the Monterey Bay in 2006 and the algorithm predicted a dominant dinoflagellate bloom (> 60% of chlorophyll-*a* biomass) with relatively high diatom biomass within the bloom (~ 20% of chlorophyll-*a* biomass) and at the periphery of the bloom; a pattern confirmed with *in situ* cell counts. PHYDOTax is unique in that it can discriminate between dinoflagellates and diatoms, a distinction historically considered challenging using chlorophyll-*a*, pigments, or light absorption spectra alone. With increased availability of hyperspectral remote sensing imagery on existing satellites, and the launch of new satellites, PHYDOTax holds promise for validating plankton functional type models, modeling biogeochemical cycles, and monitoring harmful algae in optically complex coastal waters.

Key Terms: PHYDOTax, hyperspectral, phytoplankton discriminator, harmful algal bloom (HAB), plankton functional type (PFT)

## 1. Introduction

### 1.1 Ocean Color Remote Sensing

The age of ocean color remote sensing began in 1978 with the deployment of the Nimbus-7 satellite and Coastal Zone Color Scanner (CZCS) ocean color imager [*Gordon et al.*, 1983]. The initial goal of ocean color remote sensing was to estimate global chlorophyll- *a* biomass [*O'Reilly et al.*, 1998]. Early work focused on open-

ocean, case 1 waters where phytoplankton dominates ocean color compared with the contribution by inorganic particles [*Morel and Prieur*, 1977]. More sophisticated algorithms were developed to estimate chlorophyll-*a* in optically complex, case 2 waters [*Carder et al.*, 1989], where chlorophyll-*a* contributes less than inorganic particles to ocean color. Chromophoric dissolved organic matter (CDOM) is present in both case 1 and case 2 water, but generally does not co-vary with chlorophyll-*a* in case 2 waters. With credible chlorophyll-*a* biomass estimates for both case 1 and case 2 waters, we have a more thorough understanding of global ecosystem dynamics, climate processes, and ocean circulation. Ocean color algorithms have expanded in scope and number, beyond just bulk chlorophyll-*a* estimates, to include the determination of inherent optical properties (IOPs) [*Lee et al.*, 2002], primary productivity [*Behrenfeld and Falkowski*, 1997], water mass detection [*Martin Traykovski and Sosik*, 2003], cell bio-volume [*Kostadinov et al.*, 2009], and red-tide indices [*Ahn and Shanmugam*, 2006]. As the study of ocean color moves to the next generation, even more complex questions are being asked. We know the phytoplankton community is there, now we wish to know who the occupants are. Ocean color algal discriminators can be used to resolve this question.

## 1.2 Algal Discriminators

Algal discrimination has a rich and varied history that pre-dates ocean color satellite observations and has expanded substantially in recent years [*Nair et al.*, 2008]. Many algorithms exist and this is a product of their development – usually

empirically, for a specific geographic region in order to answer specific questions. Algal discriminator algorithms can be grouped into two major types: Those that distinguish size-class-- pico-, nano-, and micro-plankton, and those that distinguish taxon abundance [*Nair et al.*, 2008]. The reason for the dichotomy stems from the application of the algorithms. Phytoplankton size-class corresponds to ecological functional type, and these algorithms are applied primarily to biogeochemical modeling questions [*Nair et al.*, 2008]. Algorithms that discriminate taxon abundance are used for ecological modeling, but also in harmful algal bloom detection and monitoring. Some examples of both types of algal discriminator algorithms include: patterns in chlorophyll-*a* anomaly over time [*Hu et al.*, 2005], pigment ratios in whole water samples (i.e. CHEMTAX) [*Mackey et al.*, 1996], light absorption spectral shape [*Ciotti et al.*, 2002; *Sathyendranath et al.*, 2004a; *Sathyendranath et al.*, 2004b], the relationship of backscattering to chlorophyll-*a* to detect the toxic dinoflagellate *Karenia brevis* [*Cannizzaro et al.*, 2009], the relationship of remote sensing reflectance ($R_{rs}$) to chlorophyll-*a* to detect red-tides in optically complex waters [*Ahn and Shanmugam*, 2006], the spectral shape of $R_{rs}$ [*Craig et al.*, 2006], and the relationship between the spectral shape of water leaving radiance ($L_w$) to diagnostic pigment bio-markers [*Alvain et al.*, 2005]. In addition to size-class or taxon algorithms, other algorithms, not detailed in this study, classify water masses with characteristics of particular algal taxa as well as other optical properties (e.g. high turbidity or CDOM) [*Hommersom et al.*, 2011; *Martin Traykovski and Sosik*, 2003; *Moore et al.*, 2001].

The types of questions that can be asked with algal discriminators are varied. Can observations of phytoplankton taxa from remote sensing data validate plankton functional (PFT) type models? Can carbon flow through different ecological pathways be quantified starting with observations of phytoplankton taxa in remote sensing imagery? Can carbon export estimates be better constrained, and how might this influence climate models? Can remote observations of phytoplankton taxa be used to identify and track harmful algal blooms (HABs)?

To illustrate, size-class can be modeled from ocean color remote sensing data. These size-class algorithms generally are based upon the spectral shape of a particular inherent optical property (IOP), such as light absorption or backscattering, or on the concentration of chlorophyll-*a* and other bio-marker pigments of particular phytoplankton taxa [*Claustre*, 1994; *Devred et al.*, 2011; *Uitz et al.*, 2006]. The spectral shape of light absorption [*Ciotti et al.*, 2002] or the composition of pigment bio-markers [*Ciotti et al.*, 1999; *Sathyendranath et al.*, 2001; *Uitz et al.*, 2006] are used to infer phytoplankton size-class based on trends found in empirical data. Though there are limitations (e.g. package effect or overlap of pigments among taxa), these methods do differentiate pico-, nano-, and micro- plankton in natural samples collected either *in situ* or from ocean color imagery. These algorithms are suitable for biogeochemical modeling, or for validating plankton functional type models that require no more than three phytoplankton size-classes. However, for more complex treatment of the phytoplankton, observations of more than these three size-classes are needed.

In comparison, taxon-specific biomass can be discriminated from ocean color remote sensing data. These estimates are useful for biogeochemical modeling, validation of plankton functional type models, and to monitor for harmful algae. Similar to size-class algorithms, taxon-specific composition of natural waters can be inferred from ocean color in several ways: by the spectral shape of IOPs such as light absorption [*Sathyendranath et al.*, 2004b; *Subramaniam et al.*, 2002] and backscattering [*Cannizzaro et al.*, 2008] and by the relationship of spectral shape of the surface $L_w$ or $R_{rs}$ to accessory pigment concentration [*Alvain et al.*, 2005]. This last classification algorithm, PHYSAT, is based on empirical methods and can discriminate among five major phytoplankton groups: haptophytes, *Prochlorococcus*, *Synechococcus*-like, diatoms, and *Phaeocystis*-like [*Alvain et al.*, 2005; *Alvain et al.*, 2008]. Algorithms such as PHYSAT attempt to refute the accusation that ecological modelers are "running before they can walk" [*T R Anderson*, 2005] by identifying plankton functional types through satellite observations that can be used to validate those models. PHYSAT is limited to case 1 waters, was built using empirical relationships, and at present only distinguishes between two major taxa in the larger size-class (haptophytes and diatoms). These limitations prevent it from being applied convincingly to case 2 waters of the coastal ocean where higher nutrient concentrations support diverse populations of phytoplankton in the microplankton size-class. These phytoplankton include diatoms, dinoflagellates, some haptophytes and cryptophytes. These larger, coastal taxa are inherently difficult to discriminate from each other because of similarities in light absorption spectra or in the overlap of

pigments among groups due to shared evolutionary ancestors [*Dierssen et al.*, 2006; *Falkowski et al.*, 2004; *Keeling*, 2004; *Lewitus et al.*, 2005].

Taxon-specific algal discriminators are promising tools to synoptically monitor for harmful algae with remote sensing imagery. Harmful algal blooms (HABs) are algal blooms that have deleterious effects on human or commercial activities [*D M Anderson et al.*, 2000] and are monitored to mitigate negative impacts. Because of costs and time delays, new methods have evolved to monitor HABs at larger spatial scales or higher temporal resolution in near real-time as a complement to existing agency and volunteer monitoring networks. Some of these new methods include instrumented mooring arrays that collect data relevant to HAB detection [*C Scholin et al.*, 2009], ocean color remote sensing imagery [*Stumpf*, 2001], and data assimilation techniques to aid HAB prediction [*Stumpf et al.*, 2003].

*In situ* moorings and satellite imagery provide high-resolution temporal and spatial data useful for monitoring HABs [*C Scholin et al.*, 2009; *Stumpf et al.*, 2003]. Several satellite algorithms have been developed to detect and monitor for the neuro-toxic dinoflagellate *Karenia brevis*, which forms dense blooms on the West Florida Shelf (WFS), a case 1 body of water. These algorithms include a chlorophyll-*a* anomaly [*Hu et al.*, 2005], spectral light absorption similarity index [*Millie et al.*, 1997], backscattering relationship to chlorophyll-*a* [*Cannizzaro et al.*, 2008; *Cannizzaro et al.*, 2009], and a red-tide index from an ocean color band-ratio method [*Shanmugam*, 2011]. The *K. brevis* work on the WFS has borne a group of bio-optical models that provide a robust test of algal discrimination in a relatively simple

optical environment. In more optically complex, case 2 waters, with multiple phytoplankton taxa, some of these methods would be less successful. One study [*Ahn and Shanmugam*, 2006] found good agreement with the red-tide index developed for the eutrophic waters in the region of the Yellow Sea of northeast Asia. This algorithm successfully discriminates a red-tide upon a background of detritus, CDOM, and chlorophyll-*a*. A second iteration of this algorithm was tested in both eutrophic and oligotrophic waters and it also successfully identified the red-tide despite some limitations due to errors in satellite-derived chl-*a* estimates [*Shanmugam*, 2011].

These red-tide indices are effective at defining one dominant bloom taxon, but they are insensitive to distinguishing more than one taxon that may compose the HAB at one time. A real need exists for an algorithm that simultaneous solves for multiple phytoplankton taxa in coastal waters. A semi-analytical phytoplankton discriminator algorithm may be a better approach to answering a wider range of questions in a flexible and credible way because it combines the generality afforded by using first principles of bio-optics, with the specificity of taxon-specific measurements of inherent optical properties.

## 1.3 Optics

In order to build an algorithm to detect phytoplankton taxa using first principles of bio-optics, it is important to first understand some basic concepts. Remote sensing reflectance, $R_{rs}(\lambda)$, is the quantity of ocean color detected at the sea surface by the imaging sensor aboard a ship, an airborne, or satellite platform. It is

defined as the ratio of water leaving radiance, $L_w(\lambda,0^+)$ to downwelling irradiance, $E_d(\lambda,0^+)$, just above the air-sea interface (Equation 1):

$$R_{rs}(\lambda) = \frac{L_w(\lambda,0^+)}{E_d(\lambda,0^+)} \quad (1)$$

Remote sensing reflectance is also defined in terms of the inherent and apparent optical properties of an optically deep and vertically homogeneous water column using radiative transfer theory (Equation 2):

$$R_{rs}(\lambda) = \frac{t^2}{n^2} \frac{\mathrm{f}}{\mathrm{Q}(\lambda)} \frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)} \quad (2)$$

where t is the transmittance across the air-seawater interface, n is the index of refraction of seawater, f is a function of the solar zenith angle, Q(λ) is the upwelling irradiance-to-radiance ratio, $b_b(\lambda)$ is the spectral backscattering coefficient, and $a(\lambda)$ is the total spectral absorption coefficient (Lee et al. 1994). Equation 2 can be simplified to (Equation 3):

$$R_{rs}(\lambda) = C \frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)} \quad (3)$$

where C is a constant to account for transmittance across the air-sea interface, the index of refraction, the solar zenith angle, and the upwelling irradiance-to-radiance ratio. The total spectral absorption coefficient can be further decomposed to (Equation 4):

$$a(\lambda) = a_w(\lambda) + a_{NAP}(\lambda) + a_{ph}(\lambda) + a_{CDOM}(\lambda) \quad (4)$$

where the subscripts w, NAP, ph, and CDOM refer to: water, non-algal particles (detritus), phytoplankton, and chromophoric dissolved organic matter. The backscattering coefficient can be further defined (Equation 5):

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda) \quad (5)$$

where the subscripts w and p refer to water and particles. The radiative transfer equations (RTE) can be solved using measured inputs for the absorbing and scattering properties of the water body (e.g. $a_{NAP}(\lambda)$, $a_{ph}$, $a_{CDOM}(\lambda)$,$b_{bp}(\lambda)$, and the scattering phase function; with $b_{bw}(\lambda)$ and $a_w(\lambda)$ supplied from published values [*Pope and Fry*, 1997]), the nature of the wind-blown sea surface, the reflectance properties of the bottom of the water column, and the incident sun- and sky-radiance [*C.D. Mobley*, 1994; *C. D. Mobley and Sundman*, 2008].

The radiative transfer equations are used to solve for the spectral radiance distribution within some defined column of water. The magnitude of $a(\lambda)$ is generally much larger than $b_b(\lambda)$ in natural, open-ocean and coastal waters. Early algal discriminators rely only on variability in $a(\lambda)$ [*Craig et al.*, 2006; *Millie et al.*, 1997; *Sathyendranath et al.*, 2004a]. However, despite its relatively low magnitude, $b_b(\lambda)$ is an important variable in conferring brightness and quality to the remote sensing reflectance, and should not be ignored [*Dierssen et al.*, 2006]. Phytoplankton groups containing similar pigments and therefore having similar absorption properties could differ greatly in backscattering due to differences in size (bio-volume) or composition of the cell wall [*Kirk*, 1994]. Remote sensing reflectance, or alternatively normalized water leaving radiance, incorporates both the $a(\lambda)$ and $b_b(\lambda)$ components of light and therefore may be a better variable to use instead of light absorption when discriminating among similar algal taxa. This is especially true when differentiating among members of the "brown" algal taxa defined by Beutler et al. [*Beutler et al.*, 2002]. These taxa coexist in mixed assemblages in the case 2 waters of Monterey Bay, CA where the present study is located.

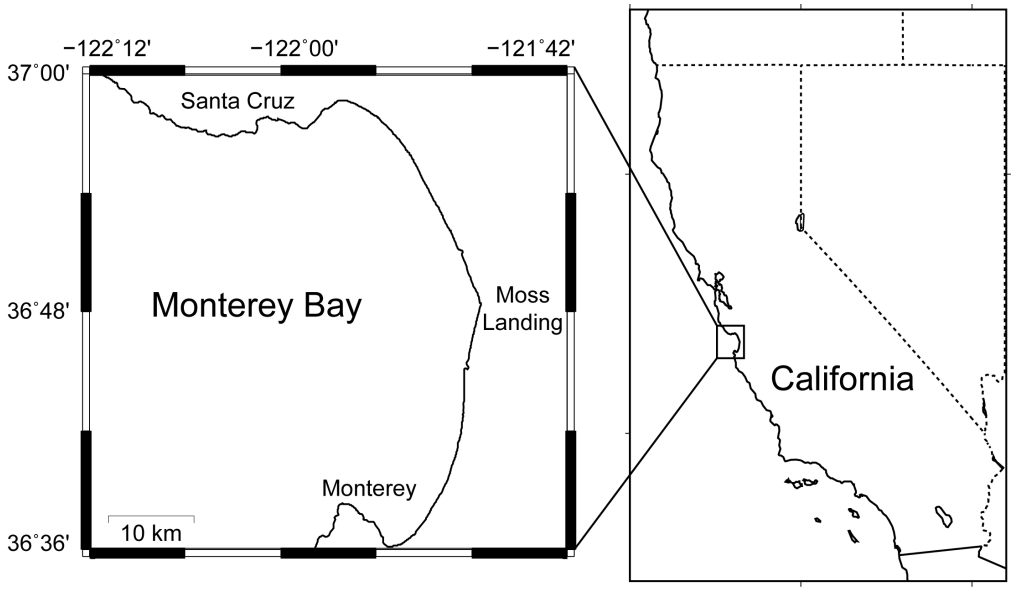



Fig. 3.1. Study Location. The Monterey Bay is a semi-enclosed bay along the central coast of California, USA.

## 1.4 Monterey Bay

The Monterey Bay, CA is an open bay located on the west coast of the USA, at the eastern fringe of the Northern California Current System (CCS), an eastern boundary current (Fig. 3.1). Its wind-driven circulation has been extensively described elsewhere [*Breaker and Broenkow*, 1994; *Pennington and Chavez*, 2000], but generally follows three oceanic seasons: upwelling (April – August), oceanic (September – October), and Davidson (November – March). This seasonal cycle is the climatological trend, though dates can vary and seasons can be interrupted with circulation consistent with other seasons. Typically during the upwelling season, recently upwelled water from just north of the bay is entrained into the bay where it bifurcates into a languid, northward flowing, cyclonic surface current and a southward flowing, anti-cyclonic surface current. The oceanic period is characterized by a relaxation of upwelling favorable winds, warming, and stratification of the surface ocean. The northern part of the bay, near Santa Cruz, lies in the wind shadow of the Santa Cruz Mountains [*J. P. Ryan et al.*, 2009] and water temperatures here tend to be warmer than other regions of the bay. The Davidson period generally experiences downwelling favorable winds with episodic winter storm systems.

Patterns in phytoplankton ecology correspond to changes in physical and chemical conditions [*Smayda and Reynolds*, 2001] during these seasons. Diatoms tend to dominate during upwelling with its high nutrient flux and strong sheer-stress associated with vigorous mixing [*Smayda*, 1997]. During this period, some toxic

species of the diatom genus *Pseudonitszchia* produce domoic acid-- a neurotoxin that causes amnesiac shellfish poisoning and has caused extensive die-offs of marine mammals [*C A Scholin et al.*, 2000]. Dinoflagellates are sensitive to vigorous mixing and generally are not as abundant in Monterey Bay during the spring and summer. This slow growing group thrives in the nutrient-poor, stratified warm waters [*Smayda*, 1997] of the oceanic period. Dinoflagellates form extensive blooms in the northern Monterey Bay in a region termed the "red-tide incubator" [*J. P. Ryan et al.*, 2008]. Winds and prevailing water entrainment patterns concentrate dinoflagellates into this incubator [*J. P. Ryan et al.*, 2009]. Episodic wind reversals can result in mixed assemblages of diatoms and dinoflagellates [*Fawcett et al.*, 2007]. In the Monterey Bay, it is not uncommon to find two co-dominant taxa either comingled within a surface bloom or in vertical layers with dinoflagellates in the nutrient-poor surface and a subsurface diatom bloom (personal observation).

The frequency and duration of these red-tide events have increased in recent years [*Jester et al.*, 2009]. The red-tide incubator of the northern part of the bay provides an excellent opportunity to study the physics supporting the blooms [*J.P. Ryan et al.*, 2005], the physiology of the various species that compose the blooms each year [*Kudela et al.*, 2008], the ecological succession of those species and their parasites [*Mazzillo et al.*, 2011], and the optics of the blooms. Some of these red-tide events have proved to be harmful to wildlife [*Jessup et al.*, 2009] and human health [*Honner et al.*, 2010] and are closely monitored.

One approach to monitoring is to identify and track algal taxa using optical signatures of the phytoplankton. Monitoring for HABs in Monterey Bay was one of the motivations for the algorithm developed in this study, though the algorithm has wider possible applications for modeling carbon flow in ecosystems and validating plankton functional type models. The objectives of this study were 1) to develop an optical signature library of phytoplankton taxa found in Monterey Bay, CA, 2) to develop a phytoplankton discrimination algorithm using the signature library, 3) to apply the algorithm to hyperspectral remote sensing reflectance spectra collected from a ship-board spectroradiometer and an airborne imager, and 4) to validate the algorithm using "synthetic" phytoplankton mixtures and field measurements of natural waters.

## 2.0 Methods

Nine large-scale cultures (Table 3.1) were grown for this study. The inherent optical properties of these cultures were measured and modeled. Because only six taxonomic groups were represented by the measured cultures, IOPs from other studies were also used [*Dierssen et al.*, 2006; *Stramski and Kiefer*, 1991]. The radiative transfer equations (RTE) (HydroLight™ V. 4.2; Sequoia Scientific, Inc.) computed remote sensing reflectance using the phytoplankton culture IOPs. These culture $R_{rs}$ spectra formed the signature library. Unknown, natural water $R_{rs}$ spectra from Monterey Bay were decomposed into constituent library components using the phytoplankton discriminator algorithm developed in this study. Biomass for each constituent was

computed and the model was validated using "synthetic" phytoplankton mixtures and independent species quantification measurements from the field. The study design is depicted schematically in Fig. 3.2.

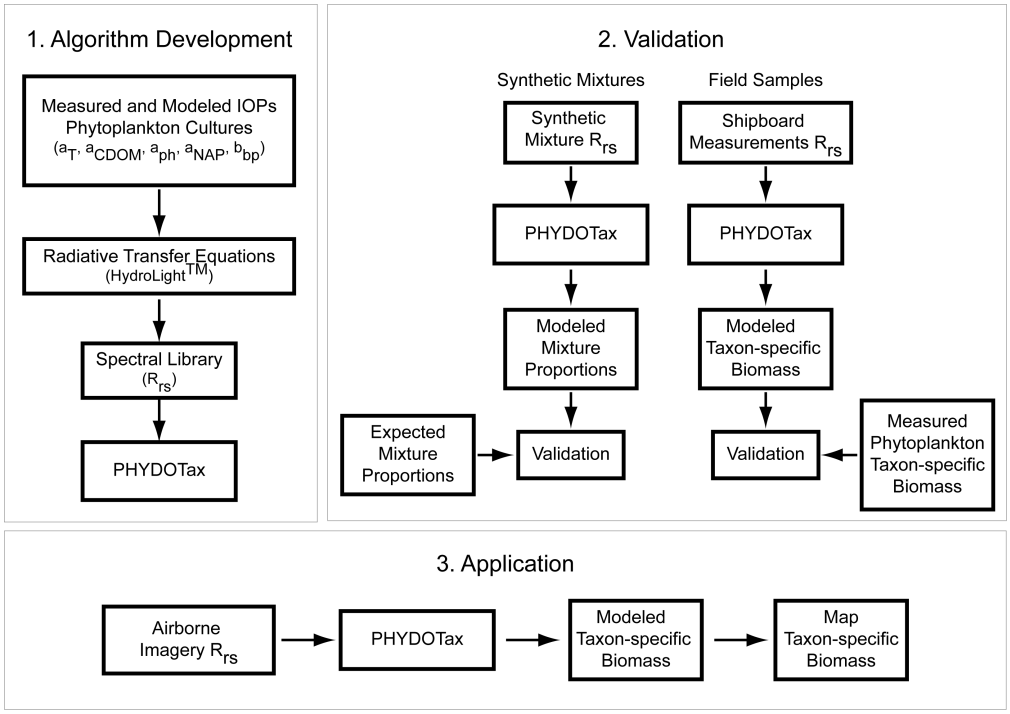



Fig. 3.2. Study Design. Schematic representation of the development, application, and validation of the phytoplankton discriminator, PHYDOTax.

## 2.1 Large-Scale Cultures

Nine phytoplankton cultures were grown to either 20 L or 200 L volumes. Genera selected included ones representative of the major color groups and those that could be cultured to a large volume. The nine phytoplankton genera included *Akashiwo, Amphidinium, Dunaliella, Isochrysis, Pseudonitszchia, Heterosigma, Skeletonema, Synechococcus, and Thalassiosira* (Table 3.1).

Table 3.1. Phytoplankton cultures used in the study.

| Culture | Family | Group | Sample Date |
|---|---|---|---|
| *Akashiwo sanguinea* | Gymnodiniaceae | dinoflagellate | June 2, 2009 |
| *Amphidinium carterae* | Gymnodiniaceae | dinoflagellate | February 16, 2007 |
| *Dunaliella tertiolecta* | Chlorophyceae | chlorophyte | May 1, 2009 |
| *Isochrysis galbana* | Isochrysidaceae | haptophyte | April 23, 2009 |
| *Pseudo-nitzschia sp.* | Bacillariaceae | diatom | May 1, 2007 |
| *Heterosigma* | Heterosigmataceae | raphidophyte | October 24, 2007 |
| *Skeletonema sp.* | Skeletonemaceae | diatom | December 3, 2008 |
| *Synnechococcus* | Synechococcaeceae | cyanobacteria | May 1, 2007 |
| *Thalassiosira pseudonana* | Thalassiosiraceae | diatom | March 24, 2009 |
| *Phycocyanin-rich Pico-eukaryotes* | | UPCE | Stramski et al. 2001 |
| *Emiliana huxleyi* | Noëlaerhabdaceae | haptophyte | Stramski et al. 2001 |
| *Rhodomonas salina* | Pyrenomonadaceae | cryptophyte | unpub. data |
| *Prymnesium parvum* | Prymnesiophyceae | haptophyte | Stramski et al. 2001 |
| *Pavlova lutheri* | Pavlovaceae | haptophyte | Stramski et al. 2001 |
| *Porphyridium cruentum* | Porphyridiaceae | rhodophyte | Stramski et al. 2001 |
| *Chroomonas fragarioides* | Chroomonadaceae | cryptophyte | Stramski et al. 2001 |
| *Alexandrium catenella* | Goniodomataceae | dinoflagellate | unpub. data |
| *Ceratium sp.* | Ceratiaceae | dinoflagellate | unpub. data |

The 200 L volume cultures included *Amphidinium, Pseudonitszchia, Heterosigma, Skeletonema,* and *Synechococcus*. The 20 L volume cultures consisted of *Akashiwo, Dunaliella, Isochrysis, and Thalassiosira*, grown in polycarbonate carboys. The 200 L culture chamber was a hard-bodied, translucent, cylindrical vessel measuring approximately 0.60 m in diameter and 1.0 m in height, lined with a translucent culture bag, aerated constantly with an aquarium bubbler, and loosely covered with a translucent lid to limit viral or bacterial contamination from the air circulation system in the environmental growth room. The seawater matrix for both the 20 L and 200 L

cultures was composed of InstantOcean™ artificial seawater amended with Guillard's f/2 culture media (Sigma-Aldrich Co.). The environmental growth room was maintained at 15°C ambient temperature and a 12:12 light cycle at ~ 100 μmol photons $m^{-2}$ $s^{-1}$ (Sylvania "soft white" fluorescent lights). Growth rate was monitored with daily chlorophyll-*a* biomass measurements.

## 2.2 Sampling of Algal Cultures

Optical sampling commenced during late log-phase of algal growth. Discrete water samples were collected and filtered to measure chlorophyll-*a*, particulate light absorption ($a_p$), and light absorption by CDOM. A two-liter volume of the unfiltered culture sample was passed through a WETLabs, Inc. Spectral Absorption and Attenuation (ACS) light meter. The cultures grown to 200L were also sampled with a HOBILabs HydroScat-6 backscattering sensor suspended directly into the growth chamber.

### 2.2.1 Filtered Culture Samples

Chlorophyll-*a* was measured fluorometrically following the EPA Method 445.0 [*Arar and Collins*, 1997] on a Turner Designs TD-700 fluorometer and concentration was calculated using the calibration coefficient for the particular fluorometer (calibrated annually). Optical density (OD) of particles and CDOM was measured on a Cary UV-Vis Spectrophotometer (300 – 800 nm, with 0.5 nm resolution). Light absorption by algal and non-algal particles was measured by filtering an aliquot of sample (in

triplicate) onto a glass fiber filter (GF/F; Whatman®) using a low pressure vacuum. Sample and MilliQ-water blank filters were immediately scanned on the spectrophotometer for OD of total particles, then bleached with a 1.52 M NaClO solution, and re-run on the spectrophotometer to derive the algal- and non-algal particle contributions to total particulate absorption. The MilliQ blank was subtracted from the sample OD. The particulate absorption coefficient was calculated using Equation 6:

$$a_p(\lambda) = \frac{2.303 * \mathrm{OD}_{(\lambda - \lambda 750)}}{\mathrm{V} / \mathrm{A}_{\mathrm{pad}}} \qquad (6)$$

Where 2.303 is the correction factor for the natural log transform, V is volume filtered, and $A_{pad}$ is the area of the filter-pad. The multiple scattering effects of the glass fiber filters were subtracted following the protocol of Cleveland and Weidemann [*Cleveland and Weidemann*, 1993]. The absorption coefficient of non-algal particles, $a_{NAP}(\lambda)$, the bleached scan, was subtracted from the total absorption coefficient, $a_p(\lambda)$, to obtain the absorption coefficient for phytoplankton, $a_{ph}(\lambda)$. Biomass normalized light absorption was calculated by dividing $a_{ph}(\lambda)$ by the chlorophyll-*a* concentration of the culture sample. The absorption coefficient of CDOM was measured by filtering a volume of culture through a 0.2 μm Nuclepore® polycarbonate filter and then scanning the filtrate immediately on the spectrophotometer using a 0.1m pathlength quartz cuvette. A MilliQ-water blank was

subtracted from these values and the absorption coefficient was calculated using Equation 7:

$$a_{CDOM} = \frac{2.303 * \mathrm{OD}_{(\lambda - \lambda 750)}}{0.1} \quad (7)$$

The total absorption coefficient was determined by adding the absorption coefficients for particles, CDOM, and published values for pure water (Pope and Fry 1997).

2.2.2 Optical Instrument Sampling

2.2.2.1 Spectral absorption and attenuation data collection and processing

The total absorption and attenuation coefficients of the cultures were measured using a hyper-spectral absorption and attenuation (ACS) meter manufactured by WETLabs, Inc. The ACS features two rigid flow-through tubes through which water is pulled, with the pump upstream of the sensor. Light emitting and detecting sensors measure total absorption and attenuation. For the benchtop set-up for this study, the flow configuration was altered from the standard field set-up to assure bubble-free sampling. Instead of the field deployed Seabird T5 pump, a large volume peristaltic pump (ColeParmer, Inc.) was placed upstream of the flow tubes and water was drawn successively through each rigid tube via flexible tubing for approximately one minute. Data were captured every 250 ms on a nearby PC-laptop running WetView 7.1 (WETLabs, Inc.). Care was taken to collect at least 10 consecutive seconds of bubble-free spectra. Clean-water calibrations were collected prior to each culture

sampling to evaluate instrument drift over the course of the study. Drift was negligible.

The culture biomass for *Amphidinium, Synechococcus, and Thalassiosira* exceeded the sensitivity of the attenuation sensor. Attenuation values for these cultures were modeled by fitting a sixth order polynomial to the known attenuation spectra for cultures that were within the sensor range. A least-squares fit was applied to the roots of the polynomial to absorption at 673 nm (the chlorophyll-*a* peak). New attenuation spectra were reconstructed for the three cultures. Attenuation was estimated for these cultures by modeling chlorophyll-specific attenuation for the other cultures and deriving spectral attenuation for the unknown samples by using the linear relationship of the known samples at each sensor wavelength (C. Mobley, pers. comm.).

2.2.2.2 Backscattering data collection and processing

The HydroScat-6, HS6, (HOBILabs, Inc.) is a field deployable instrument that measures optical backscattering at six independent wavelengths and one acceptance angle (140º). It also measures fluorescence at two wavelengths, one of which is for chlorophyll-*a* biomass. The measurement is converted to volume scattering function at 140º and that value is converted to backscattering. A sigma correction is applied to account for light lost due to attenuation between the sample material and the sensor (HS6 Manual). These sigma-corrected values were used in this study to obtain optical backscattering over the visible range.

The particular instrument used during this study measured backscattering at the wavelengths; 442 nm, 488 nm, 532 nm, 589 nm, 620 nm, and 671 nm prior to 2008, and then was refurbished and new diodes and sensors were installed which measured backscattering at 420 nm, 442 nm, 470 nm, 510 nm, 590 nm, and 700 nm. The magnitude and spectral shape for each culture were of interest, not a single measurement at a particular wavelength, so this change in detection wavelengths was not problematic.

Because of the geometry of the optical diodes and sensors, sampling small culture volumes (< 200 L) was not recommended due to optical contamination by the sides and bottom of the culture container, and backscattering was measured directly in the culture chamber for the 200 L cultures only. The aquarium bubbler was turned off at least 10 minutes before sampling. The lid was removed and the bag liner adjusted so it was directly against the container walls. The culture was gently "stirred" with a MilliQ-rinsed plastic paddle. The instrument was powered and the face was carefully lowered into the culture at a 45º angle and then positioned horizontally to face downward in the culture; this reduced the likelihood of bubbles adhering to the sensor windows. The instrument face was positioned 15 cm below the surface of the culture and 45 cm above the bottom of the culture chamber. Care was taken to keep the instrument centered in the culture chamber to prevent optical contamination from the sides of the container. Data were collected each second for a minimum of three minutes, stored to internal memory and then uploaded, with calibrations and corrections applied, using HydroSoft (HOBILabs, Inc.). Suspect data

were flagged and deleted. Reliable spectra were loaded into MATLAB (The Mathworks, Inc) and converted to text files formatted for HydroLight™. These data, and the total absorption and attenuation data from the whole water samples and ACS, were used as inputs to compute remote sensing reflectance. Backscattering for the other cultures grown only to 20 L was modeled using HydroLight™.

Particulate backscattering was modeled for the published phytoplankton cultures and the smaller volume cultures grown for this study. Backscattering values were modeled with HydroLight™ using the case 2 method where IOPs were obtained from a 4-component model for case 2 waters. Minerals and CDOM were held at zero and only the contribution to backscattering due to phytoplankton was estimated. Output wavelengths corresponded with the HS6 wavelengths. Initially, this method was performed on the absorption and attenuation properties of the five large volume cultures as a test to see if HydroLight™ could approximate the correct $b_{bp}$. The method was repeated six times for each culture to test six scattering phase functions in order to find the best fit between modeled and measured $b_{bp}$. The FFbb016 discretized phase function had the best fit (data not shown). So $b_{bp}$ was modeled for the other library taxa using this input parameter in HydroLight™.

## 2.3 Modeling $R_{rs}(\lambda)$

Remote sensing reflectance for the phytoplankton cultures was computed using the radiative transfer equations (RTE) to estimate the radiance distribution within an idealized, mono-specific water column. HydroLight™ solves the RTE with user-

supplied data and sub-routines. Table 3.2 summarizes the HydroLight™ simulation input parameters used for each culture. Generally, user supplied inputs included light absorption by phytoplankton and CDOM; total absorption and attenuation measured with the ACS; published values for absorption and backscattering by water [*Pope and Fry*, 1997]; and particulate backscattering. In some instances the inputs for absorption, attenuation, or $b_{bp}$ were modeled from complimentary measurements and then applied to the RTE as 'user-supplied' inputs to HydroLight™. Seven combinations of sun angle, wind speed, and cloud cover were modeled for each culture to approximate the conditions most likely encountered in the field. The air-water surface boundary conditions were set to 0 or 5 m $s^{-1}$ wind speed and a semi-empirical sky model based on RADTRAN. Sky conditions were set to 0, 20, 30, or 45° sun angle and either 0 or 20% cloud cover. The angular pattern for sky radiance was modeled with the hcnrad function within HydroLight™. An infinitely deep bottom-boundary condition was established and the RTE were solved at 1 nm resolution for the upper 1.5 m of the idealized water column. Biomass-normalized Rrs ($R_{rs\ norm}$) was computed by dividing Rrs by its value at 673 nm. Rrs norm was binned to 5 nm resolution for use in the signature library. Only the surface $R_{rs\ norm}$ spectra were used for the signature library.

Table 3.2.Input parameters used for radiative transfer equations to simulate remote sensing reflectance spectra from pure phytoplankton cultures. 'X' denotes measured parameter.

| Culture | $a_{ph}$ | $a_{CDOM}$ | ACS - a | ACS - c | HS6 - $b_b$ | Sun Angle (degrees) | Wind Speed (m $s^{-1}$) | Cloud Cover (%) |
|---|---|---|---|---|---|---|---|---|
| *Akashiwo sanguinea* | X | X | X | X | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Alexandrium* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Ceratium* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Dunaliella tertiolecta* | X | X | X | X | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Emyliana huxleyii* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Pavlova lutheri* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Prymnesium parvum* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Pseudo-nitzschia sp.* | X | X | X | X | X | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Porphyridium cruentum* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Heterosigma* | X | X | X | X | X | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Rhodomonas salina* | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Skeletonema sp.* | X | X | X | X | X | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Synnechococcus* | X | X | X | modeled | X | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| *Thalassiosira pseudonana* | X | X | X | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |
| UPCE | X | X | modeled | modeled | modeled | 0, 20, 30, 45 | 0, 5 | 0, 20 |

## 2.4 Field Study

The field study was conducted in the Monterey Bay (Fig. 3.1), an open bay along the central California, USA coast. Data collection occurred during four cruises aboard the *R/V John H. Martin* in September 2006, October 2008, October 2009, and October 2010. *In situ* optical and whole water sampling occurred inside and outside of red-tide patches, primarily in the central and north-east sections of the bay. The underway data acquisition system on the boat recorded temperature, salinity, and chlorophyll-*a* fluorescence. For the 2006, 2008, and 2010 cruises, a mini-rosette, equipped with a Seabird SBE 19plus CTD, WetStar fluorometer (WETLabs, Inc), and ten 1.5 L Niskin bottles sampled from the water column at each station. A 5 L Niskin bottle deployed to 5 m and a bucket sample from the surface were used to collect water samples in 2009. Discrete water samples were collected to measure chlorophyll-*a*, particulate light absorption, light absorption by CDOM, pigments (2006), and a suite of other biologically relevant data not used in this study. Following water collection, hyper-spectral absorption and attenuation (ACS, WETLabs, Inc) and optical backscattering (HS6, HOBILabs, Inc.) were collected. At the same time, hyper-spectral downwelling irradiance ($E_d$) and upwelling radiance ($L_u$) (350 – 800 nm; 0.3 nm resolution) were measured with the Satlantic HyperPro II profiling spectral radiometer in "floater" mode (HyperTSRB-Profiler II, Satlantic, Inc). Surface and profiling measurements were collected with the instrument in both floater and profiling mode. The HyperTSRB-Profiler II data were processed using ProSoft 7.7.12 to Level 3 and then remote sensing reflectance was computed from $L_u$ and $E_d$

using MATLAB. Surface $R_{rs}$ was normalized to the passive-fluorescence peak at a wavelength higher than 673 nm to derive a biomass normalized, remote sensing reflectance spectrum. Only surface $R_{rs\ norm}$ was used in this study.

In 2006, phytoplankton identification and cell enumeration were performed on water samples using the Flow CytoBot [*Olson and Sosik*, 2007; *Sosik and Olson*, 2007]. Cell area per milliliter ($\mu m\ mL^{-1}$) was measured for every chlorophyll-*a* containing cell and then grouped into genus or photosynthetic group. A total of nine major groups were observed: dinoflagellates, diatoms, haptophytes, chlorophytes, unspecified chlorophyll-containing, cryptophytes, chrysophytes, silicoflagellates, and "round cell" which was a mixture of chlorophytes, cryptophytes, and haptophytes. The "round cell" group was divided into thirds and each third was applied to chlorophytes, cryptophytes, and haptophytes. The total area of cells was computed for the taxonomic groups (dinoflagellates, diatoms, haptophytes, chlorophytes, cryptophytes, unspecified picoeukaryotes, chrysophytes, and silicoflagellates) and the proportion of each was computed by dividing the group area by the total area. The proportion was then multiplied times the total chlorophyll-*a* concentration to derive the measured taxon-specific biomass used in validation (Equation 8).

$$B_{taxon} = \frac{A_{taxon}}{\sum A_{taxon}} * [chl - a] \qquad (8)$$

Where $B_{taxon}$ is taxon-specific biomass from measured samples, $A_{taxon}$ is the area per mL measured for a particular taxon, and [chl-a] is measured chlorophyll-*a* concentration. Chrysophytes and silicoflagellates were not used in the validation because the algal discriminator did not include those taxa in the signature library. Cyanophytes were not detected with the Flow CytoBot. Instead, the ratio of the pigment zeaxanthin to chlorophyll-*a* was used as a marker for the cyanobacterium *Synechococcus [Kana et al., 1988]*.

2.5 Hyperspectral Imagery

Details of image collection can be found elsewhere [*Davis and Bissett*, 2007], but the salient points are summarized here. Hyper-spectral overflight imagery was collected as a part of the Coastal Ocean Applications and Science Team (COAST) cruise from September 3 – 15, 2006 in Monterey Bay, CA. COAST was formed by the National Oceanographic and Atmospheric Administration (NOAA) to develop ocean color algorithms for the next Geo-stationary Operational Environmental Satellite (GOES-R). Imagery was collected on four days during this period using the Spectroscopic Aerial Mapper with On-board Navigation (SAMSON, Florida Environmental Research Institute). SAMSON was mounted inside of a Twin Otter aircraft and collected hyper-spectral imagery covering 256 bands in the UV to NIR range (3.5 nm resolution over 380 to 970 nm) with 5 m spatial resolution. Overflights covered a grid in the north-east part of the bay in a region that could be completed in 30 minutes, and then re-sampled for a five-hour duration. The time series was intended to capture

the surface expression of vertically migrating dinoflagellates present in an extensive red-tide that coincided with the research cruise. Imagery data were calibrated, geolocated, and atmospherically corrected. The best image available (10 am local time; September 12, 2006) was downloaded from the server, imported into ENVI (ITT, Inc.) where a land and kelp forest mask was applied to the scene and a correction factor (1 x 10 -6) was applied to obtain the accurate scale for remote sensing reflectance (D. Kohler, pers. comm.). The imagery had a spectral resolution of 5 nm at 10 m spatial resolution. Data were then exported to ASCII and imported into MATLAB to estimate chlorophyll-*a* biomass and for further analysis with the phytoplankton discriminator algorithm.

Chlorophyll-*a* biomass was estimated from fluorescence line height (FLH) [*Abbott and Letelier*, 2006]. A linear regression of ship-board measured FLH was fit to chlorophyll-*a* measurements from water samples for the 2006 research cruise. This relationship ($y = 6.65 * 10^4$ (FLH) -37.7; $r^2 = 0.94$, $p < 0.05$) was then used to compute chlorophyll-*a* concentration from imagery FLH.

## 2.6 Phytoplankton Discriminator Algorithm

The phytoplankton taxon discrimination algorithm employs similar techniques as CHEMTAX [*Mackey et al.*, 1996], an algal discriminator based on phytoplankton pigment absorption. The algorithm uses phytoplankton detection with optics to discriminate to algal taxon, hence its name: PHYDOTax. The algorithm is composed of three steps. The first step uses matrix algebra to decompose a measured, or

unknown, $R_{rs\ norm}$ spectrum into its constituent parts represented in a signature library, developed from phytoplankton culture measurements. The signature library matrix, S, is a subset of the culture library $R_{rs\ norm}$ at 10 nm increments from 455 nm to 675 nm. The unknown vector, u, is a subset of $R_{rs\ norm}$ at 10nm increments from 455 nm to 675 nm for each unknown sample (i.e. culture mixtures, *in situ* $R_{rs}$ measurements from shipboard instruments, or $R_{rs\ norm}$ spectra from imagery). The algorithm can be described as the solution for a vector of coefficients, m, in the following equation (Equation 9):

$$u = S \bullet m \tag{9}$$

Which, when re-arranged to solve for m is the solution to the dot product of the inverse-matrix of the signature library, S, and the unknown spectrum, u:

$$m = S^{-1} \bullet u \tag{10}$$

Because S is a non-square matrix (i.e. more $R_{rs\ norm}$ measurements than taxon categories being solved), the pseudo-inverse of S is used instead. The solution is a best-fit approach using a least squares minimization technique. Using the pseudo-inverse of S, a more accurate representation of Equation 10 is (Equation 11):

$$m = S^{+} \bullet u \tag{11}$$

where + represents the pseudo-inverse (note: + is not to be mistaken for 'T'). As this is an over-determined solution, the number of wavelengths used in the signature library must exceed the number of taxa being resolved.

For the second step, the result of the dot product is used to compute the proportion of each signature taxon within the unknown (Equation 12):

$$\text{Proportion} = \frac{X_{unknown}}{\sum X_{unknown}} \qquad (12)$$

In the third step, this proportion is then multiplied times the total chlorophyll-*a* biomass to arrive at taxon-specific biomass present within the unknown sample.

## 2.7 Validation

PHYDOTax was first validated using "synthetic" mixtures of the culture library. These synthetic mixtures were simply the mathematical addition of library $R_{rs\ norm}$ spectra that were then re-normalized by the biomass peak. Two-, three-, five-, six-, and seven- taxon combinations were computed using this method for a total of 158 possible synthetic mixtures. A Wilcoxon rank-sum test was used to evaluate if the model predictions for phytoplankton proportion computed from PHYDOTax fit the expected proportions. The non-parametric Wilcoxon rank sum test was selected instead of the Chi-square goodness-of-fit because proportions were being compared, and because the expected proportions were would not have met the assumption of a

normal distribution of values [*Zar*, 1996]. A critical p-value of 0.05 was used to evaluate significance.

The second validation was conducted on field measurements collected in 2006 and 2010. Validation tested the null hypothesis that there was no difference in taxon-specific biomass between measured and modeled estimates. A paired t-test (critical p-value = 0.05) was used to test the hypothesis. Thirteen surface stations from 2006 and five from 2010 had both model predictions and cell enumeration data. Additionally, a test of linear regression was performed on these biomass estimates to determine the goodness-of-fit. Of these thirteen validation stations in 2006, only eleven had pigment measurements. Zeaxanthin concentrations ranged between 0 and 0.26 mg m-3 for all stations. Zeaxanthin to chlorophyll-*a* ratios were essentially 0 mg $m^{-3}$ for seven of those stations, 0.01 mg $m^{-3}$ for three stations, and 0.03 mg $m^{-3}$ for one station. These ratios were so low as to be considered evidence of no cyanophytes in the samples [*Kana et al.*, 1988]. PHYDOTax predicted zero biomass for cyanophytes in all but one station (M006 at 1.18 mg m-3). Because of these numerous zero-values, it was not possible to perform statistics comparing measured and modeled taxon-specific biomass, other than to note a 91% agreement in zero biomass.

## 3.0 Results

### 3.1 Library Development

The total light absorption coefficient of algal cultures used in this study, both measured and from published values, is presented in Figure 3.3. The absorption

spectra of dinoflagellates, diatoms, raphidophyte, cryptophytes, and haptophytes share similar shape due to pigment composition (Fig. 3.3A & 3.3B). Cyanophytes and chlorophytes share a similar shape with cyanophytes having a distinct peak at 620 nm (Fig. 3.3C). Rhodophytes and UPCE (Fig. 3.3D) share similar pigments and also spectral shape. Of the algal cultures, five reside in the "brown" color group defined by Beutler et al. 2002 [*Beutler et al.*, 2002]. Therefore, distinguishing among them solely on the basis of absorption properties is inconclusive. Of note are the absorption spectra of the cyanophyte, rhodophyte, and unidentified pycocyanin-rich picoeukaryotes (UPCE). These taxa have distinguishing absorption peaks at 550 nm and at 625 nm (Fig 3.1B) related to the phycobili-protein pigments present. These taxa can be resolved from the brown group based on absorption properties (analysis not shown).

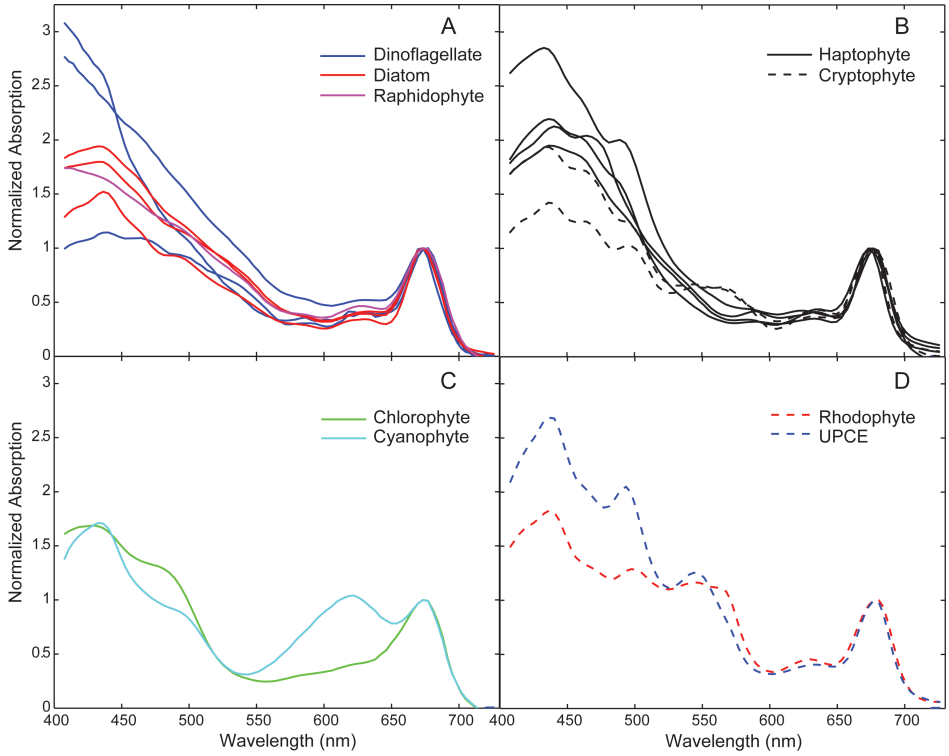



Fig. 3.3. Absorption Coefficient – cultures. Total absorption coefficient of phytoplankton cultures. Spectra measured using the WetLabs, Inc. ACS and processed according to Methods, spectra normalized to peak value at 673 nm. (A) dinoflagellates, diatoms, and raphidophytes, (B) haptophytes and cryptophytes, (C) chlorophytes and cyanophytes, (D) rhodophytes and unspecified phycocyanin containing pico-eukaroyotes – UPCE.

Figure 3.4 presents the backscattering coefficient of some representative algal cultures used in this study. Variation in bio-volume, cell wall material, and cell concentration can result in differences in backscattering magnitude and shape. For example, chlorophytes (green) and UPCE (blue dashed) had higher backscattering than the diatoms (red) and dinoflagellates (blue solid), which have a larger bio-volume. This is a pattern consistent with observations and optical principles.

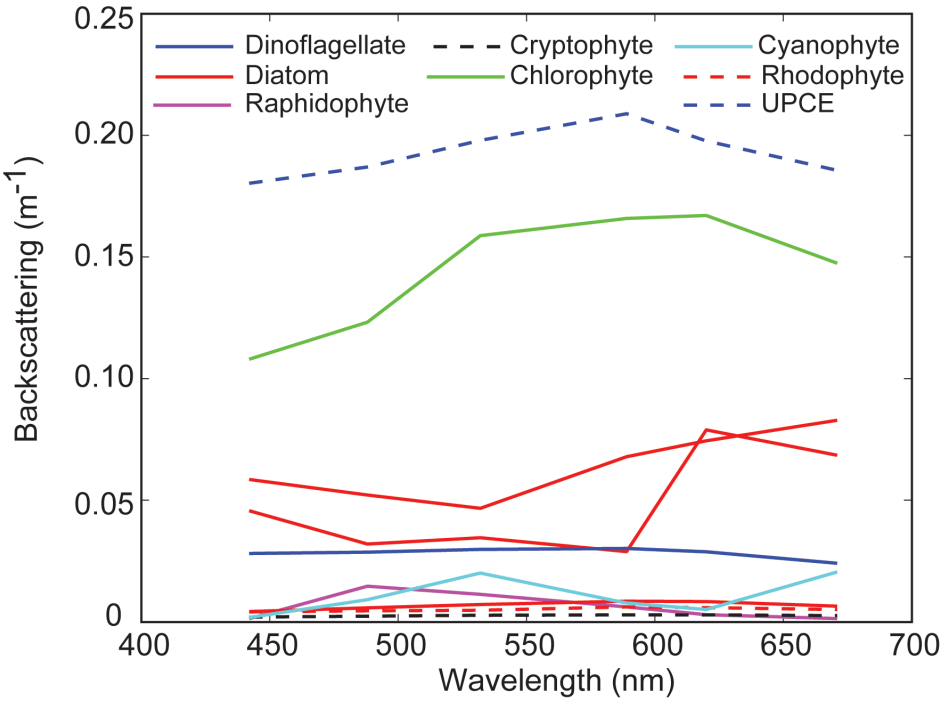



Fig. 3.4. Backscattering Coefficient – cultures. Total backscattering coefficient, measured and modeled, for some representative algal taxa.

The mean normalized $R_{rs}$ spectra for the library taxa are presented in Figure 3.5. Dinoflagellates, diatoms, raphidophytes, and haptophytes shared similar spectral shape and magnitude (Fig. 3.5A). Dinoflagellates and diatom spectra varied slightly in the peak wavelength and complexity in spectral shape at the shoulder from 550 nm to 650 nm. A large spectral peak between 500 nm and 600 nm, relative to biomass, characterized haptophytes (Fig. 3.5B). Cryptophytes had no dominant peak, but instead had multiple smaller peaks (Fig. 3.5B). The spectral shape of chlorophytes followed a relatively simple shape, as expected from its pigment composition, and had a peak between 525 nm and 600 nm that was approximately equal in magnitude to the biomass peak (Fig. 3.5C). Cyanophytes had a distinctively low peak height in the 525 nm to 575 nm range relative to the biomass peak (Fig. 3.5C). Rhodophytes differed from all other taxa with a prominent peak shifted further into the red (600 nm) with several lesser peaks between 490 nm and 540 nm (Fig. 3.5D). Because rhodophytes and raphidophytes compose such a small portion of the species assemblage in the Monterey Bay, they were eliminated from the spectral library for this study.

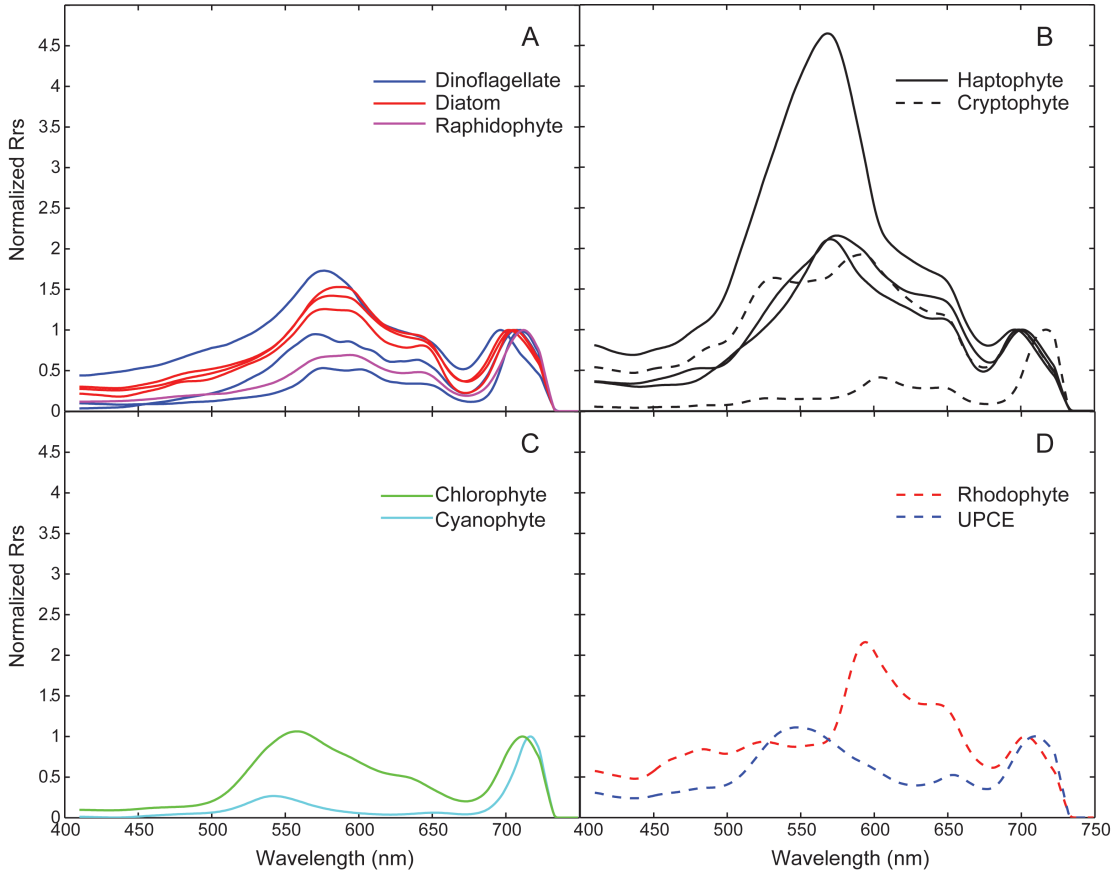



Fig. 3.5. Modeled Remote Sensing Reflectance (Rrs) - cultures. Rrs spectra of algal cultures, computed using radiative transfer equations, according to Methods, spectra normalized to peak value greater than 680 nm to derive 'biomass' normalized Rrs. (A) dinoflagellates, diatoms, and raphidophytes, (B) haptophytes and cryptophytes, (C) chlorophytes and cyanophytes, (D) rhodophytes and UPCE.

3.2 *In situ* Field Predictions

The field stations from each of the four research cruises are presented in Fig. 3.6. Station location was selected to sample both red-tide bloom and non-bloom waters in all years. Stations were distributed throughout the northern part of Monterey Bay in 2006. Sampling in 2008 was mostly in the northern part of the bay with some sampling just south of Moss Landing, CA in order to collect spectra from a filament of red-tide water in that region. The cluster of stations near Moss Landing is from an eight-hour time series within a dinoflagellate –dominated red-tide (~ 30 mg chl $m^{-3}$). Sampling in 2009 and 2010 was also distributed throughout the northern part of the bay. Only a subset of stations are represented in the figure in which both whole water and optical measurements were available.

Normalized remote sensing reflectance spectra for each station represented in Fig. 3.6, is presented in Fig. 3.7. The spectral shape and magnitude vary among stations within and among years as would be expected from the diverse water masses sampled (red-tide patch, river plume, non-bloom waters). Generally, there is a prominent peak between 550 nm and 620 nm. In some spectra, the magnitude of $R_{rs\ norm}$ was high from 400 nm to 550 nm, relative to the peak. This was likely due to the contribution of small scattering particles. These spectra in Fig. 3.7 were the ship-board measurements of hyperspectral $R_{rsnorm}$ used to model taxon-specific biomass with PHYDOTax.

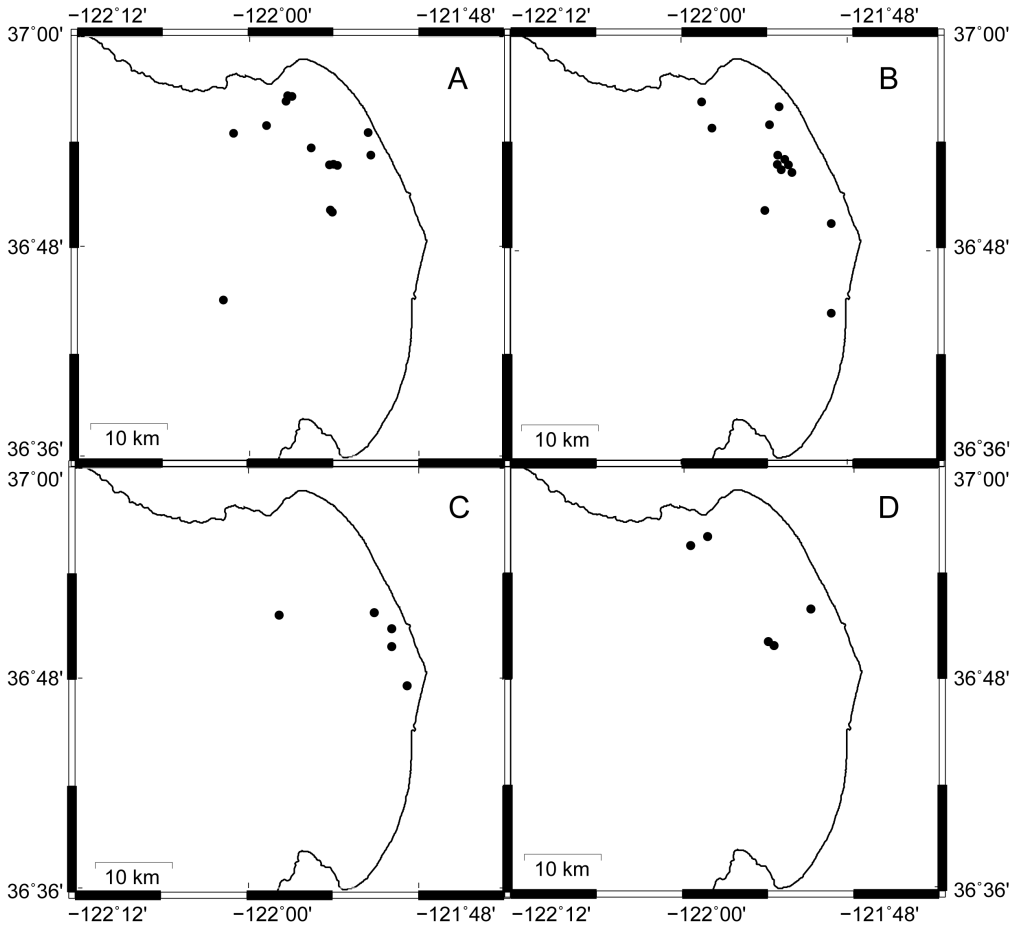



Fig. 3.6. Study Stations. Study stations for each research cruise during the four-year program. (A) COAST - 2006, (B) MB08 – 2008, (C) MB09 – 2009, (D) MB10 – 2010.

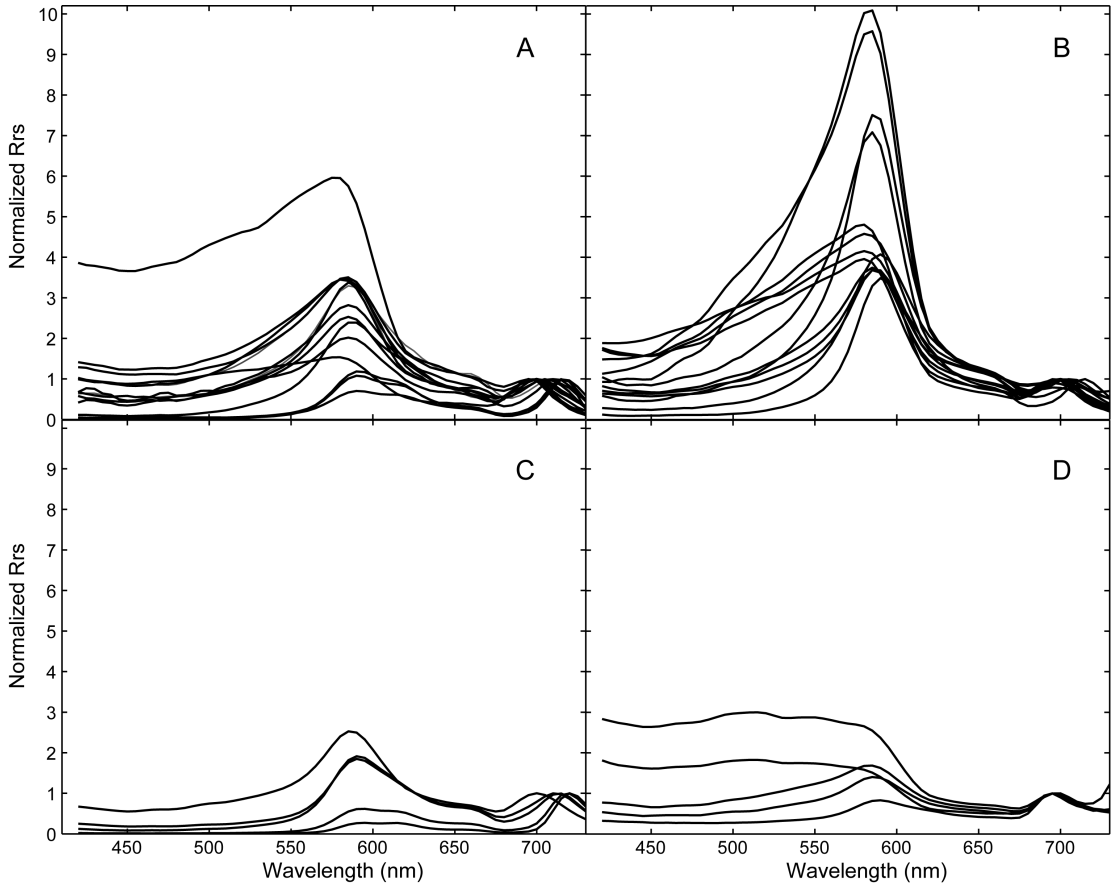



Fig. 3.7. In situ Rrs – field samples. Measured surface remote sensing reflectance, normalized by the maximum peak value greater than 673 nm, from the four research cruises (A) COAST - 2006, (B) MB08 - 2008, (C) MB09 - 2009, and (D) MB10 - 2010. Rrs was measured *in situ* using a Satlantic HyperPro Profiler II. Only surface samples were used in this study.

PHYDOTax predictions for taxon-specific biomass in 2006 are presented in Table 3.3. Measured chlorophyll-*a* concentration was higher than the typical concentration (5 mg $m^{-3}$) found during September in the Monterey Bay [*Pennington and Chavez*, 2000]. In fact, chlorophyll-*a* concentration was above typical levels at all stations in the bay ranging from 5.7 to 332 mg $m^{-3}$. PHYDOTax predicted that dinoflagellates dominated the biomass at several stations, sometimes reaching 70% of the total in 2006. At other stations, however, PHYDOTax predicted fairly evenly mixed assemblages of dinoflagellates and diatoms. Within the red-tide, PHYDOTax predicted relatively high concentrations of diatom biomass ranging from 0 to 101 mg $m^{-3}$. Haptophytes were predicted to have relatively low biomass, possibly due to the season. Chlorophytes, UPCE, and cryptophytes were predicted to be present in low concentrations. Trace levels of cyanophytes were predicted for all shipboard measurements, with the exception of the offshore M1 station where cyanophytes were predicted to occupy 15% of the taxon assemblage. Dinoflagellates were predicted to have had high chlorophyll-*a* concentrations in the northern part of the bay, intermingled with diatoms. The stations with the predicted concentrations of high diatom biomass were isolated to one part of the bay, inshore of the usual location of the red tide incubator, near Aptos, CA.

Table 3.3. COAST 2006. Taxon-specific biomass estimates from PHYDOTax for Rrs collected from the ship in September 2006.

| | | | | Measured | Predicted Chlorophyll Concentration by Taxonomic Group | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Station** | **Date** | **Lat** | **Lon** | **[chl-a] (mg m-3)** | **Dino** | **Diatom** | **Hapto** | **Chloro** | **Cyano** | **UPCE** | **Crypto** |
| M002 | 9/5/06 | 36.8776 | -121.8971 | 7.8 | 5.64 | 1.28 | 0.55 | 0.34 | 0 | 0 | 0 |
| M003 | 9/5/06 | 36.9385 | -121.9543 | 40.0 | 17.50 | 5.64 | 2.34 | 7.85 | 0 | 6.67 | 0 |
| M004 | 9/5/06 | 36.8782 | -121.9023 | 14.0 | 7.52 | 2.61 | 1.00 | 1.18 | 0 | 0.86 | 0.83 |
| M005 | 9/6/06 | 36.8785 | -121.8980 | 8.0 | 5.72 | 1.24 | 0.36 | 0.61 | 0 | 0 | 0.08 |
| M006 | 9/6/06 | 36.8778 | -121.8948 | 8.0 | 2.35 | 4.30 | 0.13 | 0.05 | 1.18 | 0 | 0 |
| M007 | 9/7/06 | 36.8876 | -121.8550 | 300.0 | 140.47 | 101.53 | 6.21 | 26.91 | 0 | 24.88 | 0 |
| M008 | 9/8/06 | 36.8341 | -121.9009 | 25.0 | 14.26 | 5.71 | 1.02 | 0.40 | 0 | 1.50 | 2.09 |
| M009 | 9/8/06 | 36.7497 | -122.0291 | 8.0 | 6.31 | 0 | 0.17 | 0.09 | 0 | 0 | 1.43 |
| M011 | 9/11/06 | 36.8951 | -121.9248 | 50.0 | 32.80 | 7.48 | 1.88 | 3.14 | 0 | 4.71 | 0 |
| M012 | 9/11/06 | 36.9087 | -121.8582 | 332.0 | 170.15 | 89.75 | 9.19 | 22.01 | 0 | 40.90 | 0 |
| M013 | 9/12/06 | 36.9082 | -122.0161 | 5.7 | 3.99 | 0.84 | 0.19 | 0.68 | 0 | 0.00 | 0 |
| M014 | 9/12/06 | 36.8354 | -121.9036 | 39.0 | 18.16 | 6.84 | 1.60 | 5.07 | 0 | 7.32 | 0 |
| M015 | 9/12/06 | 36.9154 | -121.9770 | 119.0 | 63.65 | 34.65 | 3.63 | 6.69 | 0 | 10.37 | 0 |
| M016 | 9/15/06 | 36.9430 | -121.9489 | 32.0 | 12.38 | 11.38 | 3.24 | 2.74 | 0 | 2.26 | 0 |
| M017 | 9/15/06 | 36.9437 | -121.9519 | 35.0 | 18.87 | 7.72 | 1.04 | 3.46 | 0 | 3.91 | 0 |

Monterey Bay experienced a less intense and patchier red-tide in 2008 than in 2006 (Table 3.4). Measured chlorophyll-*a* concentrations ranged from 3.8 to 30 mg $m^{-3}$. PHYDOTax predicted that the bloom was dominated by dinoflagellates, but only slightly more than diatoms in most cases at 30 to 50% of the biomass. One exception was a station at the Salinas River mouth south of Moss Landing where diatoms represented 52% of the taxon assemblage. Predicted concentrations of all other phytoplankton taxa were very low for 2008. The red-tide existed in sparse filaments oriented north to south from the northern part of the bay off of Aptos, CA towards Moss Landing, CA and then southward to Marina, CA. One of these patches located just offshore of Moss Landing, CA was intensively sampled for an 8-hour time series beginning at 10 am and ending at 6 pm local time on October 17, 2008. Only five hours of data were used in which sun angle produced reliable $R_{rs}$ measurements. The sample "station" was defined by the location of a drifter that was deployed to the densest part of the red-tide. The station migrated slightly due to tides. The percentage of dinoflagellates gradually increased for the first two hours from 70% to 75% of the total chlorophyll-*a* pool, decreased at noon to 31%, and then rose again in the early afternoon to 67%. This pattern may represent vertical migration of these motile organisms. Diatoms were predicted to be present in the bay, near the time series and at the mouth of the Salinas River, south of Moss Landing, CA. Cyanophytes were also predicted for the mouth of the Salinas River. Taxon-specific biomass predictions for the other library taxa were negligible.

Table 3.4. MB08. Taxon-specific biomass estimates from PHYDOTax for Rrs collected from the ship in October 2008.

| | | | | Measured [chl-a] | Predicted Chlorophyll Concentration by Taxonomic Group | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Station** | **Date** | **Lat** | **Lon** | **(mg $m^{-3}$)** | **Dino** | **Diatom** | **Hapto** | **Chloro** | **Cyano** | **UPCE** | **Crypto** |
| C01 | 10/13/08 | 36.9000 | -121.9600 | 5.4 | 1.27 | 2.64 | 0.17 | 0 | 1.25 | 0 | 0.08 |
| C06 | 10/15/08 | 36.7416 | -121.8200 | 19.5 | 4.31 | 10.26 | 0.68 | 0 | 4.25 | 0 | 0 |
| C07 | 10/17/08 | 36.8577 | -121.8668 | 13.3 | 9.31 | 2.47 | 0.82 | 0.06 | 0 | 0.64 | 0 |
| C08 | 10/17/08 | 36.8681 | -121.8702 | 5.7 | 4.28 | 0.19 | 0.31 | 0.35 | 0 | 0.57 | 0 |
| C09 | 10/17/08 | 36.6267 | -121.8795 | 8.1 | 2.51 | 3.39 | 0.24 | 0 | 1.55 | 0 | 0.40 |
| C10 | 10/17/08 | 36.8640 | -121.8787 | 6.6 | 4.45 | 1.73 | 0.42 | 0 | 0 | 0 | 0 |
| C12 | 10/17/08 | 36.8717 | -121.8849 | 11.4 | 5.56 | 3.49 | 0.34 | 0.79 | 0 | 1.22 | 0 |
| C13 | 10/20/08 | 36.8793 | -121.8745 | 29.3 | 17.02 | 9.40 | 1.42 | 0.61 | 0 | 0.85 | 0 |
| C14 | 10/20/08 | 36.8826 | -121.8827 | 30.0 | 13.84 | 14.85 | 1.20 | 0 | 0.11 | 0 | 0 |
| C16 | 10/23/08 | 36.9182 | -121.8933 | 18.0 | 8.11 | 5.36 | 0.74 | 1.76 | 0 | 2.03 | 0 |
| C17 | 10/23/08 | 36.8256 | -121.8193 | 7.6 | 1.85 | 4.10 | 0.17 | 0 | 1.49 | 0 | 0 |
| C18 | 10/29/08 | 36.9396 | -121.9751 | 3.8 | 0.75 | 1.92 | 0.09 | 0 | 1.04 | 0 | 0 |
| C19 | 10/29/08 | 36.9349 | -121.8821 | 7.5 | 2.18 | 3.50 | 0.47 | 0 | 1.35 | 0 | 0 |

Measured chlorophyll-*a* concentrations were the highest of the four years of the study in 2009 (Table 3.5). Concentrations ranged from 31 to 525 mg $m^{-3}$. One station not reported in the table exceeded 2000 mg $m^{-3}$ chlorophyll-*a*. PHYDOTax predicted that dinoflagellates and diatoms were approximately evenly represented in the bloom, the exception being a station at the Old Salinas River mouth immediately south of Moss Landing. Chlorophyll-*a* concentrations were relatively lower here than at other stations in 2009, but dinoflagellates made up 74% of the assemblage according to PHYDOTax. Unfortunately, radiometric measurements for the northern part of the bay were corrupted and could not be used in the analysis, hence the lack of stations in the red-tide incubator.

Measured chlorophyll-*a* concentrations were much lower in 2010 than in 2009, ranging from 1.2 to 74.2 mg $m^{-3}$ (Table 3.6). Higher concentrations were found in the northern part of the bay, just offshore of Santa Cruz. PHYDOTax predicted that dinoflagellates dominated the assemblage only slightly more than diatoms (~ 40 vs. 30 %) at all but one station in the bay where diatoms were predicted to occupy 45% of the taxon assemblage. This station was located just south of the Santa Cruz Wharf next to the Environmental Sample Processor deployed by researchers from the Monterey Bay Aquarium Research Institute to study the red-tide. Coincidentally, this location was also where a distinctive subsurface diatom bloom was observed just beneath the surface aggregation of dinoflagellates (M. Peacock, pers. comm.).

Table 3.5. MB09. Taxon-specific biomass estimates from PHYDOTax for Rrs collected from the ship in October 2009.

| Station | Date | Lat | Lon | Measured [chl-a] ($mg\ m^{-3}$) | Predicted Chlorophyll Concentration by Taxonomic Group | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Dino | Diatom | Hapto | Chloro | Cyano | UPCE | Crypto |
| C06 | 10/6/09 | 36.7943 | -121.8084 | 31.0 | 22.79 | 3.03 | 0.94 | 1.46 | 0 | 2.78 | 0 |
| C08 | 10/7/09 | 36.8602 | -121.9585 | 400.0 | 210.25 | 129.59 | 10.88 | 15.01 | 0 | 34.27 | 0 |
| C09 | 10/7/09 | 36.8310 | -121.8233 | 65.0 | 27.94 | 25.84 | 1.33 | 3.41 | 0 | 6.49 | 0 |
| C11 | 10/8/09 | 36.8632 | -121.8470 | 525.0 | 239.77 | 202.96 | 9.22 | 28.45 | 0 | 44.59 | 0 |
| C12 | 10/8/09 | 36.8478 | -121.8263 | 230.0 | 96.08 | 89.73 | 6.84 | 13.64 | 0 | 23.70 | 0 |



Table 3.6. MB10. Taxon-specific biomass estimates from PHYDOTax for Rrs collected from the ship in October 2010.

| Station | Date | Lat | Lon | Measured [chl-a] ($mg\ m^{-3}$) | Predicted Chlorophyll Concentration by Taxonomic Group | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Dino | Diatom | Hapto | Chloro | Cyano | UPCE | Crypto |
| E3St1 | 10/11/10 | 36.8328 | -121.8938 | 1.2 | 0.42 | 0.49 | 0.04 | 0 | 0.14 | 0 | 0.06 |
| E12St4 | 10/12/10 | 36.9240 | -121.9886 | 74.2 | 25.67 | 32.44 | 1.65 | 3.66 | 0 | 5.62 | 5.17 |
| E25St11 | 10/13/10 | 36.8661 | -121.8518 | 8.5 | 5.50 | 1.24 | 0.18 | 0.15 | 0 | 0.27 | 1.16 |
| E36St16 | 10/14/10 | 36.8361 | -121.9003 | 2.1 | 1.10 | 0.51 | 0.07 | 0.02 | 0.14 | 0 | 0.26 |
| E42St18 | 10/14/10 | 36.9322 | -121.9693 | 41.1 | 20.31 | 13.11 | 0.52 | 1.39 | 0.00 | 2.51 | 3.25 |

3.3 Validation

Validation of the model using the synthetic mixtures demonstrated that PHYDOTax accurately predicted taxonomic composition for the mixtures (Table 3.7). The null hypothesis: no difference between predicted and expected proportion, was accepted for all of the taxa ($p > 0.05$).

Table 3.7. Validation of Model - Synthetic Mixtures. Mann-Whitney test results comparing expected phytoplankton proportions to PHYDOTax predictions. A critical p-value of 0.05 was used. > 0.05 indicates PHYDOTax prediction and hypothetical proportion were not statistically different.

| Taxon | rank sum | z-value | p-value | $H_0$ |
|---|---|---|---|---|
| Dinoflagellate | 25187 | 0.19 | 0.85 | 0 |
| Diatom | 24682 | -0.47 | 0.64 | 0 |
| Haptophyte | 25188 | 0.20 | 0.84 | 0 |
| Chlorophyte | 24774 | -0.43 | 0.67 | 0 |
| Cyanophyte | 25197 | 0.26 | 0.80 | 0 |
| UPCE | 24779 | -0.42 | 0.68 | 0 |
| Cryptophyte | 24623 | -0.66 | 0.51 | 0 |

In the two-taxa mixtures, PHYDOTax accurately predicted dinoflagellates, diatoms, chlorophytes, cyanophytes, cryptophytes, and UPCE in all instances. When haptophytes were a part of the mixture, PHYDOTax erroneously applied between 1 and 3 % of the mixture to diatoms, applied the remaining 47 - 49% of the mixture to haptophytes, and then 50% to whichever taxon was a part of the mixture. A similar pattern occurred in the three-, five-, and six- taxon mixtures, with negligible mis-apportioning to diatoms when one of the haptophyte taxa was present in the mixture. The predictions for the seven-taxon mixture were an exact match to the expected

proportions. Despite this small failure, the modeled predictions were still statistically the same as the expected proportions. The exception was when the mixture included the spectra for *Emiliana huxleyi*. When *E. huxleyi* was included, PHYDOTax erroneously attributed 50% of the population to diatoms, 30% to cryptophytes, and 15% to haptophytes. This only occurred with *E. huxleyi* in the three-taxon mixtures and not in the two-taxon mixtures and could be due to the unique bio-optics of *E. huxleyi* as a coccolithophorid. Future iterations of PHYDOTax may need to separate *E. huxleyi* from the main haptophyte group in the spectral library.

Field measurements of taxon-specific biomass are reported in Table 3.8 for the cell counts that were measured using the Flow CytoBot in 2006 and by microscope enumeration in 2010. Chlorophyll-*a* concentrations in Table 3.8 match those from Table 3.3 (2006) and Table 3.6 (2010) as both measurements were collected from the same water mass at the same time. Measured dinoflagellate chlorophyll-*a* concentration ranged from 0.32 to 103.4 mg $m^{-3}$. Dinoflagellates represented between 21 and 90% of the taxon assemblage. Diatoms had chlorophyll-*a* concentrations ranging from 1.65 to 160.4 mg $m^{-3}$, with representation between 2 and 75% of the community. Other taxa were present, but generally no greater than 5 to 10% of the total. Curiously, what was considered to be an overwhelmingly massive dinoflagellate-dominated bloom, was actually fairly evenly mixed between dinoflagellates and diatoms at many stations. One exception was the extremely high concentration of dinoflagellates found at a station just south of the Santa Cruz Wharf. This station did not have the highest total chlorophyll-*a* concentration (35 mg $m^{-3}$),

but it did have the highest concentration of dinoflagellates at 90% of the assemblage. The other exception was the intense diatom bloom located near the beach in Aptos, CA. Diatoms were nearly 53% of the assemblage at 160.4 mg $m^{-3}$ chlorophyll-*a*. The patterns of mixed diatom-dinoflagellate assemblages and of the intense diatom bloom at the fringe of the red-tide were found in the PHYDOTax predictions for the $R_{rs}$ measurements from both the ship and imagery.

Table 3.8. Measured taxon-specific chlorophyll concentration. Computed in 2006 using cell counts collected with the FlowCytobot from water collected at each station during the 2006 cruise (Sosik, unpubl. data). Computed in 2010 for diatoms and dinoflagellates from cell counts. Stations and taxa shown include only those that matched PHYDOTax output. (n.d. = no data).

| | | | | Measured | Measured Chlorophyll Concentration by Taxonomic Group | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Station** | **Date** | **Lat** | **Lon** | **[chl-a] (mg m-3)** | **Dino** | **Diatom** | **Hapto** | **Chloro** | **Cyano** | **UPCE** | **Crypto** |
| M002 | 9/5/06 | 36.8776 | -121.8971 | 7.8 | 3.39 | 2.42 | 0.15 | 0.30 | n.d. | 0.67 | 0.15 |
| M003 | 9/5/06 | 36.9385 | -121.9543 | 40.0 | 24.23 | 5.12 | 1.39 | 1.46 | n.d. | 5.20 | 1.39 |
| M006 | 9/6/06 | 36.8778 | -121.8948 | 8.0 | 1.69 | 3.69 | 0.39 | 0.56 | n.d. | 0.78 | 0.39 |
| M007 | 9/7/06 | 36.8876 | -121.8550 | 300.0 | 26.58 | 160.40 | 16.49 | 16.49 | n.d. | 43.80 | 16.49 |
| M008 | 9/8/06 | 36.8341 | -121.9009 | 25.0 | 18.36 | 1.65 | 0.56 | 0.60 | n.d. | 1.46 | 0.58 |
| M009 | 9/8/06 | 36.7497 | -122.0291 | 8.0 | 4.19 | 2.43 | 0.14 | 0.20 | n.d. | 0.48 | 0.13 |
| M011 | 9/11/06 | 36.8951 | -121.9248 | 50.0 | 18.49 | 18.66 | 1.63 | 1.89 | n.d. | 4.44 | 1.43 |
| M012 | 9/11/06 | 36.9087 | -121.8582 | 332.0 | 103.42 | 144.13 | 12.35 | 12.76 | n.d. | 31.74 | 10.50 |
| M013 | 9/12/06 | 36.9082 | -122.0161 | 5.7 | 0.32 | 4.32 | 0.20 | 0.20 | n.d. | 0.39 | 0.20 |
| M014 | 9/12/06 | 36.8354 | -121.9036 | 39.0 | 11.71 | 15.47 | 1.37 | 1.66 | n.d. | 4.12 | 1.27 |
| M015 | 9/12/06 | 36.9154 | -121.9770 | 119.0 | 40.11 | 48.85 | 4.34 | 5.22 | n.d. | 10.46 | 4.34 |
| M016 | 9/15/06 | 36.9430 | -121.9489 | 32.0 | 6.72 | 17.51 | 1.97 | 1.51 | n.d. | 2.64 | 1.51 |
| M017 | 9/15/06 | 36.9437 | -121.9519 | 35.0 | 31.59 | 0.82 | 0.33 | 0.39 | n.d. | 0.91 | 0.32 |
| E3St1 | 10/11/10 | 36.8328 | -121.8938 | 1.2 | 0.07 | 1.08 | n.d. | n.d. | n.d. | n.d. | n.d. |
| E12St4 | 10/12/10 | 36.9240 | -121.9886 | 74.2 | 72.70 | 1.50 | n.d. | n.d. | n.d. | n.d. | n.d. |
| E25St11 | 10/13/10 | 36.8661 | -121.8518 | 8.5 | 7.44 | 1.06 | n.d. | n.d. | n.d. | n.d. | n.d. |
| E36St16 | 10/14/10 | 36.8361 | -121.9003 | 2.1 | 0.31 | 1.79 | n.d. | n.d. | n.d. | n.d. | n.d. |
| E42St18 | 10/14/10 | 36.9322 | -121.9693 | 41.1 | 38.20 | 2.90 | n.d. | n.d. | n.d. | n.d. | n.d. |

Results of the test of linear regression are plotted in Fig. 3.8 where a trend between measured and modeled taxon-specific biomass is evident for dinoflagellates, diatoms, haptophytes, chlorophytes, and UPCE. The results of the paired t-test indicate that the model successfully predicted dinoflagellates, diatoms, haptophytes, UPCE, and cryptophytes, but not chlorophytes (Table 3.9). The low biomass of chlorophytes and cryptophytes during the period of the cruise may have contributed to the failure to see a trend (cryptophytes) or for the null hypothesis to be rejected (chlorophytes). Cyanophytes could not be validated with field measurements, as there were too few non-zero data-points to make the comparison.

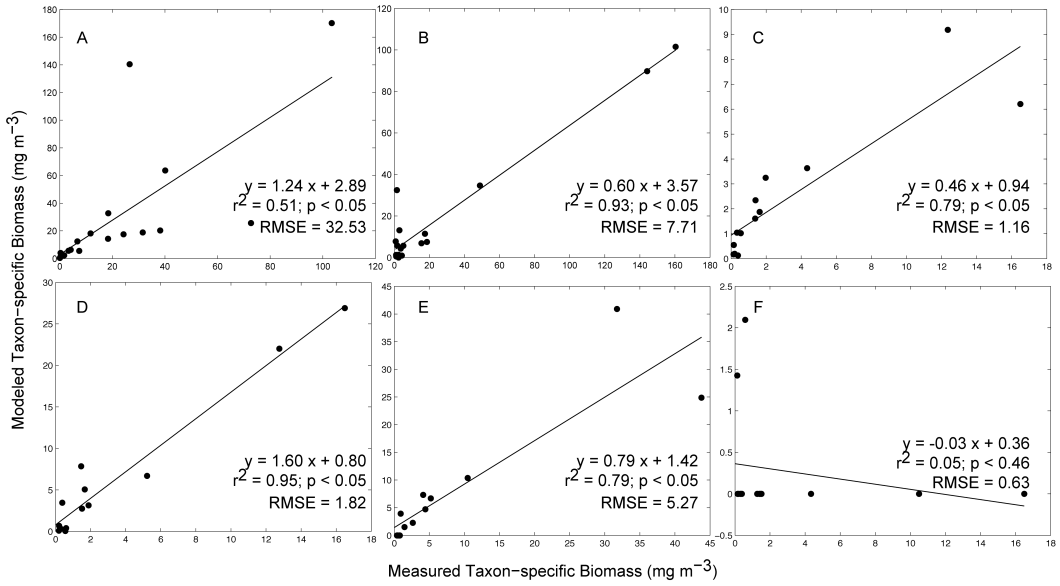



Fig. 3.8. Validation – field samples. Comparison of measured taxon-specific biomass and model estimates from PHYDOTax. Dinoflagellates (2006 and 2010) (A), diatoms (2006 and 2010) (B), haptophytes (C), chlorophytes (D), UPCE (E), and cryptophytes (F) are shown.

Table 3.9. Validation of Model - Field Samples. Regression and paired t-test results comparing taxon-specific chlorophyll biomass from measurements vs. model predictions for COAST 2006 and MB10 (diatoms and dinoflagellates). A critical p-value of 0.05 was used. Paired t-test: > 0.05 indicates measured and modeled values were not statistically different. (n.t. = not testable)

| | Regression | | | | | | Paired t-test | | |
|---|---|---|---|---|---|---|---|---|---|
| **Taxon** | **$r^2$** | **slope** | **intercept** | **MSE** | **MSPE** | **p-value** | **t-statistic** | **p-value** | **$H_0$** |
| Dinoflagellate | 0.51 | 1.24 | 2.89 | 1190.53 | 1058.25 | < 0.05 | -1.04 | 0.31 | 0 |
| Diatom | 0.93 | 0.60 | 3.57 | 66.93 | 59.50 | < 0.05 | 1.23 | 0.23 | 0 |
| Haptophyte | 0.79 | 0.46 | 0.94 | 1.60 | 1.36 | < 0.05 | 0.92 | 0.38 | 0 |
| Chlorophyte | 0.95 | 1.60 | 0.80 | 3.92 | 3.32 | < 0.05 | -2.74 | 0.02 | 1 |
| Cyanophyte | n.t. | n.t. | n.t. | n.t. | n.t. | n.t. | n.t. | n.t. | n.t. |
| UPCE | 0.79 | 0.79 | 1.42 | 32.78 | 27.73 | < 0.05 | 0.20 | 0.84 | 0 |
| Cryptophyte | 0.05 | -0.03 | 0.36 | 0.47 | 0.40 | 0.46 | 1.90 | 0.08 | 0 |

3.4 Imagery

Chlorophyll-*a* biomass was estimated to be very high in the red-tide incubator of the northern Monterey Bay in the 2006 imagery (Fig. 3.9). The bloom was concentrated in two patches, one close to shore near Aptos, CA and another just offshore, parallel to the shore and the prevailing internal wave fronts that are commonly present at that location [*J. P. Ryan et al.*, 2009]. Surface concentrations of chlorophyll-*a* ranged from near zero offshore to more than 400 mg $m^{-3}$ in the densest part of the bloom, as confirmed by water samples collected during the research cruise.

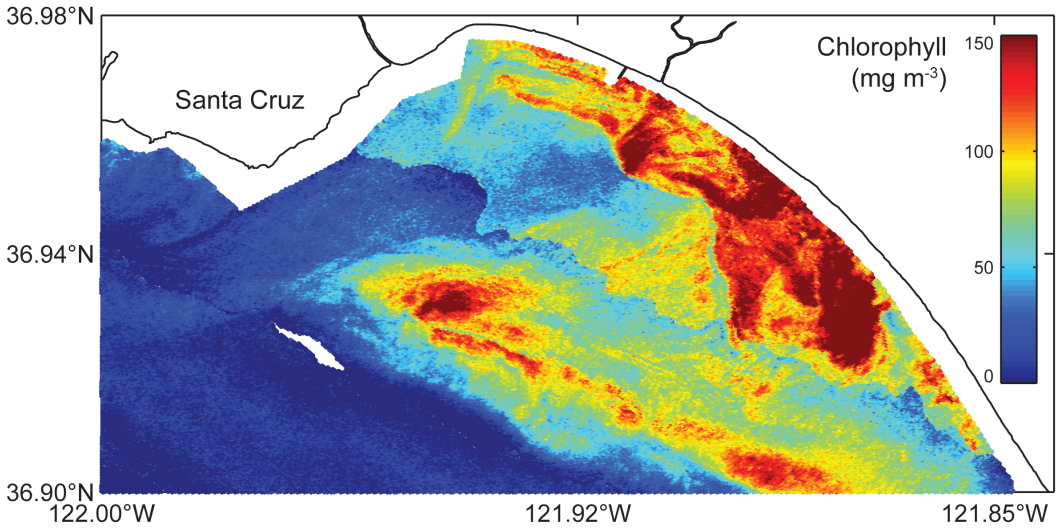



Fig. 3.9. Derived Chlorophyll-*a* Biomass – imagery. Chlorophyll-*a* was derived from fluorescence line height (FLH) using the relationship between *in situ* chorophyll-*a* measurements and FLH computed from *in situ* Rrs measurements. Relationship was applied to imagery FLH to derive chorophyll-*a* biomass. Image collected September 12, 2006 using SAMSOM Hyper-spectral airborne sensor.

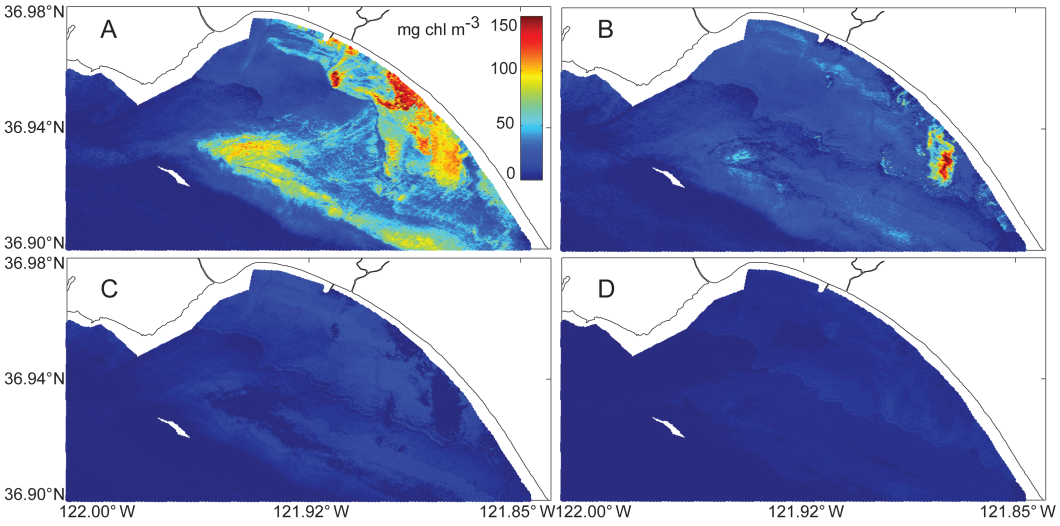



Fig. 3.10. Predicted taxon-specific biomass – 2006 Imagery. Estimates of the chlorophyll-*a* concentration represented by phytoplankton taxon; (A) dinoflagellates, (B) diatoms, (C) cyanobacteria, (D) haptophytes. Image collected September 12, 2006 using SAMSOM Hyper-spectral airborne sensor.

Water measurements indicate that the bloom was dominated by the dinoflagellate *Akashiwo sanguinea*. PHYDOTax predicted that dinoflagellates were the primary taxon present in the bloom region with 40 to 90% of the taxon-assemblage composed of this group (Fig. 3.10A). Dinoflagellate-specific chlorophyll-*a* biomass reached 182 mg $m^{-3}$ in some regions and ranged between 15 and 117 mg $m^{-3}$ in most parts of the bloom. PHYDOTax predicted the presence of diatoms within the dinoflagellate bloom (~20%), but at lower concentrations (3.9 to 61 mg $m^{-3}$), and at the periphery of the bloom at concentrations as high as 179 mg $m^{-3}$ (Fig. 3.10B). This pattern is borne out in the water measurements collected at stations within the bloom during the research cruise, though only two stations coincided with the time of image capture and at locations outside of the image boundary. The presence of all other taxa in the library was negligible (Fig. 3.10C & 3.15D). Haptophytes, chlorophytes, UPCE, and cryptophytes were predicted to have median chlorophyll-*a* concentrations of 6.6, 0, 0, and 8.3 mg $m^{-3}$ respectively. Cyanophytes were predicted to have a median chlorophyll-*a* concentration of 9.7 mg $m^{-3}$, mostly co-located with the intense dinoflagellate bloom.

## 4.0 Discussion

In this paper we have presented a bio-optical model to simultaneously discriminate multiple phytoplankton taxa from hyperspectral remote sensing reflectance spectra. This phytoplankton discriminator algorithm is a robust tool for quantifying taxon-specific biomass in optically complex, case 2 waters. Validation confirms that

PHYDOTax can distinguish among diatoms, dinoflagellates, haptophytes, cryptophytes, chlorophytes, cyanophytes, and UPCE. PHYDOTax is the first ocean color algorithm to discriminate between diatoms and dinoflagellates. The generally accepted paradigm has been that these two taxa cannot be differentiated from each other because of the overlap in pigment composition or similarities in light absorption spectra [*Dierssen et al.*, 2006]. This novel algorithm is a promising tool to partition the phytoplankton biomass in the coastal environment where diatoms and dinoflagellates occur together and may support different ecosystems or occupy different niches within a harmful algal bloom.

In addition to developing a new algorithm to distinguish algae from ocean color data, this is the first published study demonstrating the feasibility of measuring backscattering directly from large-volume phytoplankton cultures. Backscattering measurements were collected to more accurately model $R_{rs}$ of the pure cultures. These measurements are important for the ocean color community to better characterize IOPs of taxonomically pure end-members [*Nair et al.*, 2008].

## 4.1 Comparison with existing algorithms

When compared to existing algorithms, PHYDOTax has several advantages over its predecessors. PHYDOTax's conceptual ancestor, CHEMTAX, is limited in spatial resolution as it is implemented on pigment data collected *in situ*. CHEMTAX credibly estimates taxon-assemblage. However, its input requirements are onerous: water must first be collected *in situ*; the pigments extracted and measured using high-

performance liquid chromatography by a trained technician; then the expected library taxa and pigment ratios tuned to a particular habitat or region; followed by execution of the algorithm [*Mackey et al.*, 1996]. Assignment to an incorrect algal taxon is possible simply by choosing pigment ratios tuned to a different geographic region or the wrong suite of expected algal taxa [*Lewitus et al.*, 2005; *Mackey et al.*, 1996]. Compared to PHYDOTax, application of this model is limited in spatial and temporal scale as it was designed to validate taxon-composition from whole water samples, and not from remote observations from moorings or satellites.

An active and successful research campaign has been the detection of the toxic dinoflagellate *Karenia brevis* on the west Florida shelf using remote ocean color observations from moorings [*Stumpf et al.*, 2003], autonomous underwater vehicles [*Robbins et al.*, 2006], airborne sensors [*Cannizzaro et al.*, 2008], and satellites [*Hu et al.*, 2005; *Shanmugam*, 2011; *Stumpf et al.*, 2003; *Tomlinson et al.*, 2009]. Bio-optical algorithm development has been prolific. An early model was a two-dimensional spectral analysis technique to determine similarity [*Millie et al.*, 1997]. *Karenia brevis* contains the pigment bio-marker gyroxanthin-diester, which has distinctive absorption properties. At only 5% of the magnitude of the spectrum [*Millie et al.*, 1995], it is a relatively weak signal and is overwhelmed by chlorophyll-*c* and fucoxanthin. Historically, this algorithm has been used with *in situ* water samples, but with recent advances in decomposing IOPs from hyperspectral ocean color imagery [*Lee and Carder*, 2004] these comparisons can be made over much larger spatial scales. Another algorithm that exploits the contrast of the unique properties of *K.*

*brevis* against an optically different background is that developed by Cannizzaro and collaborators [*Cannizzaro et al.*, 2008; *Cannizzaro et al.*, 2009]. This model distinguishes *K. brevis* by its distinctively low chlorophyll-specific particulate backscattering coefficient at 550 nm ($\leq 0.0045$ $m^2$ $mg^{-1}$) [*Cannizzaro et al.*, 2009]. This differs dramatically from the type of phytoplankton that may be present in the case 1 waters where *K. brevis* is typically found. Using a large hyperspectral dataset for the WFS as a test, this model accurately discriminated *K. brevis* blooms from non-blooms 99% of the time [*Cannizzaro et al.*, 2009]. Both of these *K. brevis*-specific algorithms were developed with hyperspectral data to discriminate just one phytoplankton species, albeit a noxious one, from a background of non-bloom waters. These, and other remote sensing algorithms for the WFS, have been successful and a necessary "proof-of-concept" that ocean color can be used to identify harmful algae in case 1 waters. A logical next step addressed in the present is to detect algae in more optically complex, case 2 waters, or in waters with a plurality of algal taxa influencing the ecosystem and bio-optics.

Two bio-optical algorithms that distinguish red-tides in optically complex waters have been applied to northeast-Asia coastal waters [*Ahn and Shanmugam*, 2006], the Arabian Sea and Gulf of Oman, and the West Florida Shelf as a test of case 1 waters [*Shanmugam*, 2011]. These two algorithms, the red-tide index (RI) and the algal bloom index (ABI) use an empirical relationship between chlorophyll-*a* concentration and band-ratios of ocean color to detect red-tides. They were developed to differentiate dinoflagellate blooms upon a background of turbid water and other

phytoplankton species using existing multispectral satellite imagery from SeaWiFS or MODIS [*Ahn and Shanmugam*, 2006; *Shanmugam*, 2011]. These algorithms successfully identify the presence of a red-tide in both the case 1 and case 2 water types. Both band-ratio methods fail with incorrect chlorophyll-*a* estimates. So, the choice of chlorophyll-*a* algorithm or the atmospheric correction parameters used in processing the imagery can affect algorithm performance [*Shanmugam*, 2011]. This is particularly a concern in near-shore environments where the atmospheric composition can differ greatly from near-shore to offshore. Despite these caveats, the algorithms are effective at distinguishing red-tides in optically complex waters and can do so with existing satellites. However, they are limited to a bulk estimate of the red-tide and do not distinguish among other coincident phytoplankton taxa that may contribute to the deleterious effects of a bloom.

The phytoplankton classification algorithm PHYSAT does simultaneously solve for multiple phytoplankton taxa using multispectral ocean color imagery [*Alvain et al.*, 2005]. It is effective at characterizing the taxonomic composition in case 1 waters of the global ocean [*Alvain et al.*, 2008; *d'Ovidio et al.*, 2010] and could be a valuable algorithm for describing the partitioning of carbon flow [*Masotti et al.*, 2011] in several biogeochemical provinces including the north Atlantic and the Southern Ocean [*Alvain et al.*, 2008]. PHYSAT and PHYDOTax are similar in that they solve for multiple taxa in one iteration of the algorithm. They differ in the taxonomic composition of the signature library; the approach to development – PHYSAT is empirical and PHYDOTax is semi-analytical; and the water types where

they can be applied. PHYSAT will be technically challenged to distinguish among some taxa traditionally grouped in the brown color group [*Beutler et al.*, 2002] which includes diatoms and dinoflagellates. PHYSAT uses the relationship between diagnostic pigment bio-markers and water-leaving radiance to classify groups: Pigment overlap in dinoflagellates and diatoms discourages delimiting the two groups [*Dierssen et al.*, 2006]. Because PHYDOTax uses $R_{rs}$, which incorporates backscattering, it does not suffer from the pigment overlap problem. Pigment composition must be unequivocal among the phytoplankton taxa in order to clearly characterize the different groups using PHYSAT. Further increases in algal groups may be limited by vague pigment distinctions in the remaining taxa not yet characterized. Another limitation is extending PHYSAT to optically complex waters where higher turbidity, CDOM, and phytoplankton are conflated in one water-leaving radiance spectrum. A semi-analytical approach that includes a range of concentrations of non-algal particles or CDOM in the library building spectra may more closely approximate the bio-optics of a phytoplankton bloom in case 2 waters. In its current state, PHYDOTax has non-varying $a_{NAP}$ and $a_{CDOM}$ in its signature library; however, it has the potential to include such variability because of its semi-analytical approach to modeling library spectra. Even without including variability in non-algal particles and CDOM, PHYDOTax can predict the composition of phytoplankton taxa in the optically complex waters of Monterey Bay, CA.

## 4.2 Implications for PFT Modeling and HAB Monitoring

PHYDOTax successfully discriminates phytoplankton taxa in coastal waters. It may be possible to apply the algorithm to case 1 waters with future enhancements to the signature library to include *Prochlorococcus*, *Trichodesmium*, and other phytoplankton more typical of the open ocean. Plankton functional type models vary in complexity from relatively simple NPZD models with one to three phytoplankton size-classes to a more complex 78 phytoplankton-type model (i.e. DARWIN) [*Follows et al.*, 2007]. High-complexity models are alluring as they promise to explain the ecosystem more thoroughly. Validating these models is problematic without an equal number of independent validation observations as there are input variables [*T R Anderson*, 2005]. Satellite observations have historically provided relatively few of the independent observations needed to validate models. Recent work by Alvain et al. [*Alvain et al.*, 2005; *Alvain et al.*, 2008; *Alvain et al.*, 2006] has made it possible to observe some plankton functional types in the open ocean. PHYDOTax may do the same for the coastal ocean. While PHYDOTax will never meet the 78 input parameter threshold for independent validation, it does predict more phytoplankton groups than existing algorithms, and more signature library taxa could be added to the model in the future. For implementations of DARWIN, or simplifications of DARWIN, in the CCS [*Goebel et al.*, 2010], PHYDOTax could be a viable tool to produce validation observations. Ultimately, one goal of PFTs is to constrain estimates of carbon flow. It may soon be possible to validate those predictions in the CCS with increased availability of hyperspectral imagery along the

CA coast using the Hyperspectral Imager for the Coastal Ocean (HICO™) imaging spectrometer on the International Space Station.

Taxonomically, PHYDOTax discriminates to the family level or higher. Because of shared pigments, morphology, and cell wall material within taxa, it is unlikely that the algorithm will ever differentiate within the existing taxa-- the exception being haptophytes and dinoflagellates. These two phylogenetically and bio-optically diverse groups [*Falkowski et al.*, 2004; *Keeling*, 2004; *Lewitus et al.*, 2005] may be separated to improve performance of the algorithm. PHYDOTax does not resolve to the species level, so it cannot be used to directly detect and monitor a HAB. However, it can be used to identify the dominant taxon within a bloom. As evidenced in this study (Fig. 3.10), PHYDOTax described a mixed dinoflagellate-diatom bloom in 2006. The ability of the model to simultaneously resolve dinoflagellates and diatoms could aid in describing taxon patchiness within a larger bloom. *In situ* measurements confirmed the small diatom bloom found at the periphery of the larger red-tide (Fig. 3.10B). Prior to PHYDOTax; chlorophyll-*a*, light absorption, and particulate backscattering could be discerned from the hyperspectral $R_{rs}$ imagery [*Lee and Carder*, 2004]. Conclusions could be drawn as to the likelihood of taxa within the bloom, but no further predictions could be made. Using PHYDOTax, knowledge of the taxonomic composition of the HAB could improve decision making and adaptive sampling of the bloom.

## 4.3 Limitations and Future Work

PHYDOTax has its limitations, and like its conceptual predecessor, CHEMTAX, should not be treated like a "black box." Like CHEMTAX, PHYDOTax is an over determined solution and therefore the number of independent variables retrieved (i.e. expected library taxa) cannot exceed the number of wavelengths used in the algorithm (twenty-three wavebands). In addition, some *a priori* knowledge of the expected taxonomic composition is needed to select the signature library taxa to use in the model-runs. In the future, the hope is to add more library taxa to the signature library to extend the utility of PHYDOTax beyond the seven taxa currently in it. The existing library is composed of taxa found mostly in the coastal ocean of Monterey Bay. A next step is to include *Prochlorococcus* and *Trichodesmium,* and a wider diversity of eukaryotes including chrysophytes and silicoflagellates. Additionally, to include more species within the haptophyte and dinoflagellate groups (beyond the three in each) to more adequately represent the wide range of bio-optical subtypes in those two groups [*Falkowski et al.*, 2004; *Keeling*, 2004; *Lewitus et al.*, 2005]. A third limitation is that the algorithm uses hyperspectral data. One sensor, the Hyperion Imager aboard NASA's EO-1 satellite, has the spectral and spatial resolution needed to accommodate PHYDOTax. Imagery from this sensor is captured by on-demand requests from approved research projects for both terrestrial and aquatic systems. Demand by the many stakeholders limits the availability of imagery from the coastal ocean. Until recently, Hyperion images were very costly. With the launch of HICO™, repeated satellite retrievals over the coastal ocean are now available to ask relevant

scientific questions using PHYDOTax. The limited availability of hyperspectral imagery does not preclude the implementation of PHYDOTax on data-sets from mooring arrays that measure surface remote sensing reflectance. So, while HICO data may be limited in its coverage or temporal resolution, taxon-specific discrimination is currently possible with existing instrumented moorings.

The future of ocean color remote sensing is gradually moving towards hyperspectral imagers with the temporal and spatial resolution found on MODIS. Hyperspectral sensors are expensive to develop and the anticipated computing and data-storage needs are very costly. Not surprisingly, government entities equivocate in embracing missions to launch hyperspectral imagers into space. With little commitment to launch hyperspectral imagers, the research community is reluctant to develop and test hyperspectrally-based algorithms. However, within the last five years the impasse has broken; the number of algorithms has increased [*Cannizzaro et al.*, 2008; *Craig et al.*, 2006; *Torrecilla et al.*, 2011]. PHYDOTax is an example of one, and also provides a valuable lesson: The need for "hyperspectral" may not require 1 nm resolution with sub-nanometer bandwidth. PHYDOTax operates at 10 nm resolution with a ~2 nm bandwidth and can discriminate among taxa previously thought indivisible. If other bio-optical algorithms can perform at similar spectral resolution, then it may be possible to launch a lower resolution sensor that still affords the computational skill needed for spectral analysis, but at a resolution that is affordable to develop and maintain.

## 5.0 Acknowledgements

This work was funded by the NOAA GOES-R COAST project, NESDIS-NASA UARC Award (TO.030.6.MD.D-Kudela), UC-Santa Cruz Chancellor's Dissertation Year Fellowship (SP), Dr. Earl H. Myers and Ethel M. Myers Oceanographic and Marine Biology Trust (SP). We thank the captain and crew of the *R/V John H. Martin* for services rendered throughout the four years of field-work. SP wishes to especially thank Clarissa Anderson, Misty Blakely, Taylor Crockford, Meredith Howard, Maria Kavanaugh, Rick Pasetto, G. Jason Smith, and Rick Stumpf for assistance. We thank Christopher Edwards and the anonymous reviewers for their time and attention given to this manuscript.

Literature Cited

Abbott, M. R., and R. M. Letelier (2006), Algorithm theoretical basis document: Chlorophyll fluorescence (MODIS product number 20)*Rep.*

Ahn, Y. H., and P. Shanmugam (2006), Detecting the red tide algal blooms from satellite ocean color observations in optically complex Northeast-Asia Coastal waters, *Remote Sensing of Environment*, *103*(4), 419-437.

Alvain, S., C. Moulin, Y. Dandonneau, and F. M. Breon (2005), Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery, *Deep-Sea Research Part I-Oceanographic Research Papers*, *52*(11), 1989-2004.

Alvain, S., C. Moulin, Y. Dandonneau, and H. Loisel (2008), Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view, *Global Biogeochemical Cycles*, *22*(3).

Alvain, S., C. Moulin, Y. Dandonneau, H. Loisel, and F. M. Breon (2006), A species-dependent bio-optical model of case I waters for global ocean color processing, *Deep-Sea Research Part I-Oceanographic Research Papers*, *53*(5), 917-925.

Anderson, D. M., P. Hoagland, Y. Kaoru, and A. W. White (2000), Economic impacts from harmful algal blooms (HABs) in the United Stats*Rep.*, Woods Hole Oceanographic Institution, Woods Hole, MA.

Anderson, T. R. (2005), Plankton functional type modelling: running before we can walk?, *Journal of Plankton Research*, *27*(11), 1073-1081.

Arar, E. J., and G. B. Collins (1997), Method 445.0: In vitro determination of chlorophyll a and pheophytin a in marine and freshwater algae by fluorescence, edited by U. S. E. P. Agency, Cincinnati, Ohio.

Behrenfeld, M. J., and P. G. Falkowski (1997), Photosynthetic rates derived from satellite-based chlorophyll concentration, *Limnology and Oceanography*, *42*(1), 1-20.

Beutler, M., K. H. Wiltshire, B. Meyer, C. Moldaenke, C. Luring, M. Meyerhofer, U. P. Hansen, and H. Dau (2002), A fluorometric method for the differentiation of algal populations in vivo and in situ, *Photosynthesis Research*, *72*(1), 39-53.

Breaker, L. C., and W. W. Broenkow (1994), The circulation of Monterey Bay and related processes, *Oceanography and Marine Biology*, *32*, 1 - 64.

Cannizzaro, J. P., K. L. Carder, F. R. Chen, C. A. Heil, and G. A. Vargo (2008), A novel technique for detection of the toxic dinoflagellate, Karenia brevis, in the Gulf

of Mexico from remotely sensed ocean color data, *Continental Shelf Research*, *28*(1), 137-158.

Cannizzaro, J. P., C. M. Hu, D. C. English, K. L. Carder, C. A. Heil, and F. E. Muller-Karger (2009), Detection of Karenia brevis blooms on the west Florida shelf using in situ backscattering and fluorescence data, *Harmful Algae*, *8*(6), 898-909.

Carder, K. L., R. G. Steward, G. R. Harvey, and P. B. Ortner (1989), Marine Humic and Fulvic-Acids - Their Effects on Remote-Sensing of Ocean Chlorophyll, *Limnology and Oceanography*, *34*(1), 68-81.

Ciotti, A. M., J. J. Cullen, and M. R. Lewis (1999), A semi-analytical model of the influence of phytoplankton community structure on the relationship between light attenuation and ocean color, *Journal of Geophysical Research-Oceans*, *104*(C1), 1559-1578.

Ciotti, A. M., M. R. Lewis, and J. J. Cullen (2002), Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient, *Limnology and Oceanography*, *47*(2), 404-417.

Claustre, H. (1994), THE TROPHIC STATUS OF VARIOUS OCEANIC PROVINCES AS REVEALED BY PHYTOPLANKTON PIGMENT SIGNATURES, *Limnology and Oceanography*, *39*(5), 1206-1210.

Cleveland, J. S., and A. D. Weidemann (1993), QUANTIFYING ABSORPTION BY AQUATIC PARTICLES - A MULTIPLE-SCATTERING CORRECTION FOR GLASS-FIBER FILTERS, *Limnology and Oceanography*, *38*(6), 1321-1327.

Craig, S. E., S. E. Lohrenz, Z. P. Lee, K. L. Mahoney, G. J. Kirkpatrick, O. M. Schofield, and R. G. Steward (2006), Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, Karenia brevis, *Applied Optics*, *45*(21), 5414-5425.

d'Ovidio, F., S. De Montec, S. Alvain, Y. Dandonneau, and M. Levy (2010), Fluid dynamical niches of phytoplankton types, *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18366-18370.

Davis, C., and W. P. Bissett (2007), Characterization of a harmful algal bloom in Monterey Bay, CA using airborne hyperspectral imagery, in *Hyperspectral Imaging and Sounding of the Environment (HISE)*, edited, Optical Society of America (OSA), Santa Fe, NM.

Devred, E., S. Sathyendranath, V. Stuart, and T. Platt (2011), A three component classification of phytoplankton absorption spectra: Application to ocean-color data, *Remote Sensing of Environment*, *115*(9), 2255-2266.

Dierssen, H. M., R. M. Kudela, J. P. Ryan, and R. C. Zimmerman (2006), Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments, *Limnology and Oceanography*, *51*(6), 2646 - 2659.

Falkowski, P. G., M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield, and F. J. R. Taylor (2004), The evolution of modern eukaryotic phytoplankton, *Science*, *305*(5682), 354-360.

Fawcett, A., G. C. Pitcher, S. Bernard, A. D. Cembella, and R. M. Kudela (2007), Contrasting wind patterns and toxigenic phytoplankton in the southern Benguela upwelling system, *Marine Ecology-Progress Series*, *348*, 19-31.

Follows, M. J., S. Dutkiewicz, S. Grant, and S. W. Chisholm (2007), Emergent biogeography of microbial communities in a model ocean, *Science*, *315*(5820), 1843-1846.

Goebel, N. L., C. A. Edwards, J. P. Zehr, and M. J. Follows (2010), An emergent community ecosystem model applied to the California Current System, *Journal of Marine Systems*, *83*(3-4), 221-241.

Gordon, H. R., J. W. Brown, O. B. Brown, R. H. Evans, and D. K. Clark (1983), Nimbus-7 Czcs - Reduction of Its Radiometric Sensitivity with Time, *Applied Optics*, *22*(24), 3929-3931.

Hommersom, A., M. R. Wernand, S. Peters, M. A. Eleveld, H. J. van der Woerd, and J. de Boer (2011), Spectra of a shallow sea-unmixing for class identification and monitoring of coastal waters, *Ocean Dynamics*, *61*(4), 463-480.

Honner, S., R. M. Kudela, and E. Handler (2010), Bilateral mastoiditis from red tide exposure, *Journal of Emergency Medicine*.

Hu, C. M., F. E. Muller-Karger, C. Taylor, K. L. Carder, C. Kelble, E. Johns, and C. A. Heil (2005), Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters, *Remote Sensing of Environment*, *97*(3), 311-321.

Jessup, D. A., M. A. Miller, J. P. Ryan, H. M. Nevins, H. A. Kerkering, A. Mekebri, D. B. Crane, T. A. Johnson, and R. M. Kudela (2009), Mass Stranding of Marine Birds Caused by a Surfactant-Producing Red Tide, *Plos One*, *4*(2).

Jester, R., K. Lefebvre, G. Langlois, V. Vigilant, K. Baugh, and M. W. Silver (2009), A shift in the dominant toxin-producing algal species in central California alters phycotoxins in food webs, *Harmful Algae*, *8*(2), 291-298.

Kana, T. M., P. M. Glibert, R. Goericke, and N. A. Welschmeyer (1988), ZEAXANTHIN AND BETA-CAROTENE IN SYNECHOCOCCUS WH7803 RESPOND DIFFERENTLY TO IRRADIANCE, *Limnology and Oceanography*, *33*(6), 1623-1627.

Keeling, P. J. (2004), Diversity and evolutionary history of plastids and their hosts, *American Journal of Botany*, *91*(10), 1481-1493.

Kirk, J. T. O. (1994), *Light and photosynthesis in aquatic ecosystems*, 509 pp., Cambridge University Press, New York.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2009), Retrieval of the particle size distribution from satellite ocean color observations, *Journal of Geophysical Research-Oceans*, *114*.

Kudela, R. M., J. P. Ryan, M. D. Blakely, J. Q. Lane, and T. D. Peterson (2008), Linking the physiology and ecology of Cochlodinium to better understand harmful algal bloom events: A comparative approach, *Harmful Algae*, *7*(3), 278-292.

Lee, Z. P., and K. L. Carder (2004), Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance, *Remote Sensing of Environment*, *89*(3), 361-368.

Lee, Z. P., K. L. Carder, and R. A. Arnone (2002), Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters, *Applied Optics*, *41*(27), 5755-5772.

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

Mackey, M. D., D. J. Mackey, H. W. Higgins, and S. W. Wright (1996), CHEMTAX - A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton, *Marine Ecology-Progress Series*, *144*(1-3), 265-283.

Martin Traykovski, L. V., and H. M. Sosik (2003), Feature-based classification of optical water types in the Northwest Atlantic based on satellite ocean color data, *J. Geophys. Res.*, *108*(C5), -.

Masotti, I., C. Moulin, S. Alvain, L. Bopp, A. Tagliabue, and D. Antoine (2011), Large-scale shifts in phytoplankton groups in the Equatorial Pacific during ENSO cycles, *Biogeosciences*, *8*(3), 539-550.

Mazzillo, F., J. Ryan, and M. Silver (2011), Parasitism as a biological control agent of dinoflagellate blooms in the
California Current System, *Harmful Algae*, *10*, 763 - 773.

Millie, D. F., G. J. Kirkpatrick, and B. T. Vinyard (1995), RELATING PHOTOSYNTHETIC PIGMENTS AND IN-VIVO OPTICAL-DENSITY SPECTRA TO IRRADIANCE FOR THE FLORIDA RED-TIDE DINOFLAGELLATE GYMNODINIUM BREVE, *Marine Ecology-Progress Series*, *120*(1-3), 65-75.

Millie, D. F., O. M. Schofield, G. J. Kirkpatrick, G. Johnsen, P. A. Tester, and B. T. Vinyard (1997), Detection of harmful algal blooms using photopigments and absorption signatures: A case study of the Florida red tide dinoflagellate, *Gymnodinium breve*, *Limnology and Oceanography*, *42*(5), 1240-1251.

Mobley, C. D. (1994), *Light and Water. Radiative Transfer in Natural Waters*, Academic, New York.

Mobley, C. D., and L. K. Sundman (2008), Hydrolight 5 Ecolight 5 Users' Guide*Rep.*, Sequoia Scientific, Inc., Bellevue, WA.

Moore, T. S., J. W. Campbell, and H. Feng (2001), A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms, *Ieee Transactions on Geoscience and Remote Sensing*, *39*(8), 1764-1776.

Morel, A., and L. Prieur (1977), Analysis of Variations in Ocean Color, *Limnology and Oceanography*, *22*(4), 709-722.

Nair, A., S. Sathyendranath, T. Platt, J. Morales, V. Stuart, M. H. Forget, E. Devred, and H. Bouman (2008), Remote sensing of phytoplankton functional types, *Remote Sensing of Environment*, *112*(8), 3366-3375.

O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain (1998), Ocean color chlorophyll algorithms for SeaWiFS, *Journal of Geophysical Research-Oceans*, *103*(C11), 24937-24953.

Olson, R. J., and H. M. Sosik (2007), A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot, *Limnology and Oceanography-Methods*, *5*, 195-203.

Pennington, J. T., and F. P. Chavez (2000), Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989 - 1996 in Monterey Bay, California, *Deep-Sea Research II*, *47*, 947 - 973.

Pope, R. M., and E. S. Fry (1997), Absorption spectrum (380-700 nm) of pure water .2. Integrating cavity measurements, *Applied Optics*, *36*(33), 8710-8723.

Robbins, I. C., G. J. Kirkpatrick, S. M. Blackwell, J. Hillier, C. A. Knight, and M. A. Moline (2006), Improved monitoring of HABs using autonomous underwater vehicles (AUV), *Harmful Algae*, *5*(6), 749-761.

Ryan, J. P., A. M. Fischer, R. M. Kudela, J. F. R. Gower, S. A. King, R. Marin, and F. P. Chavez (2009), Influences of upwelling and downwelling winds on red tide bloom dynamics in Monterey Bay, California, *Continental Shelf Research*, *29*(5-6), 785-795.

Ryan, J. P., H. M. Dierssen, R. M. Kudela, C. A. Scholin, K. S. Johnson, J. M. Sullivan, A. M. Fischer, E. V. Rienecker, P. R. McEnaney, and F. P. Chavez (2005), Coastal ocean physics and red tides: an example from Monterey Bay, California, *Oceanography*, *18*(2), 246 - 255.

Ryan, J. P., J. F. R. Gower, S. A. King, W. P. Bissett, A. M. Fischer, R. M. Kudela, Z. Kolber, F. Mazzillo, E. V. Rienecker, and F. P. Chavez (2008), A coastal ocean extreme bloom incubator, *Geophysical Research Letters*, *35*(12).

Sathyendranath, S., G. Cota, V. Stuart, H. Maass, and T. Platt (2001), Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches, *International Journal of Remote Sensing*, *22*(2-3), 249-273.

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

Sathyendranath, S., et al. (2004b), A multispectral remote sensing study of coastal waters off Vancouver Island, *International Journal of Remote Sensing*, *25*(5), 893-919.

Scholin, C., et al. (2009), Remote Detection of Marine Microbes, Small Invertebrates, Harmful Algae, and Biotoxins Using the Environmental Sample Processor (Esp), *Oceanography*, *22*(2), 158-167.

Scholin, C. A., et al. (2000), Mortality of sea lions along the central California coast linked to a toxic diatom bloom, *Nature*, *403*(6765), 80-84.

Shanmugam, P. (2011), A new bio-optical algorithm for the remote sensing of algal blooms in complex ocean waters, *Journal of Geophysical Research-Oceans*, *116*.

Smayda, T. J. (1997), Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea, *Limnology and Oceanography*, *42*(5), 1137-1153.

Smayda, T. J., and C. S. Reynolds (2001), Community assembly in marine phytoplankton: application of recent models to harmful dinoflagellate blooms, *Journal of Plankton Research*, *23*(5), 447-461.

Sosik, H. M., and R. J. Olson (2007), Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry, *Limnology and Oceanography-Methods*, *5*, 204-216.

Stramski, D., and D. A. Kiefer (1991), LIGHT-SCATTERING BY MICROORGANISMS IN THE OPEN OCEAN, *Progress in Oceanography*, *28*(4), 343-383.

Stumpf, R. P. (2001), Applications of satellite ocean color sensors for monitoring and predicting harmful algal blooms, *Human and Ecological Risk Assessment*, *7*(5), 1363-U1315.

Stumpf, R. P., M. E. Culver, P. A. Tester, M. Tomlinson, G. J. Kirkpatrick, B. A. Pederson, E. Truby, V. Ransibrahmanakul, and M. Soracco (2003), Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data, *Harmful Algae*, *2*(2), 147-160.

Subramaniam, A., C. W. Brown, R. R. Hood, E. J. Carpenter, and D. G. Capone (2002), Detecting Trichodesmium blooms in SeaWiFS imagery, *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, *49*(1-3), 107-121.

Tomlinson, M. C., T. T. Wynne, and R. P. Stumpf (2009), An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, Karenia brevis, *Remote Sensing of Environment*, *113*(3), 598-609.

Torrecilla, E., D. Stramski, R. A. Reynolds, E. Millan-Nunez, and J. Piera (2011), Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean, *Remote Sensing of Environment*, *115*(10), 2578-2593.

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *Journal of Geophysical Research-Oceans*, *111*(C8).

Zar, J. H. (1996), *Biostatistical Analysis*, 3rd ed., Prentice Hall College Div, Upper Saddle River, NJ.

# CONCLUSIONS

The bio-optical algorithms developed in this dissertation addressed three oceanographic objectives: to identify where the water mass is, how the water mass is changing in space and time, and what is contained within the water mass. The first chapter demonstrated that light absorption by chromophoric dissolved organic matter, $a_{CDOM}$, is a reliable proxy for salinity in the brackish environment of the Columbia River Plume (CRP) on the Oregon and Washington shelf. The second chapter described how to identify sub-mesoscale water masses in the larger CRP, consistent with other observational and modeling studies. The third chapter introduced an algal discriminator algorithm that was able to differentiate several phytoplankton taxa (including diatoms, dinoflagellates, haptophytes, cryptophytes, chlorophytes, cyanophytes, and unspecified phycocyanin-containing picoeukaryotes) in case 2 waters. The remainder of this section summarizes each chapter's results, impact and implications, and future work.

## Chapter 1

An optical proxy for salinity was developed for the CRP by relating salinity to $a_{CDOM}$ for 2004, 2005, and 2006 [*Palacios et al.*, 2009]. A generalized algorithm was also developed that can be applied to imagery in other years- though only during spring-summer periods. The algorithm was robust at estimating the salinity of the CRP and is the first to relate ocean color to *in situ* measurements of salinity for the

CRP. As a result, it is now possible to estimate salinity for the Oregon and Washington shelf where the plume resides – a vast improvement in spatial resolution over the limited observations possible with existing mooring instrumentation. In addition to a dramatic increase in the spatial scale over which salinity now can be observed, these observations can be made at near-daily frequency, so long as weather conditions permit cloud-free observations of the sea. These types of satellite observations of "synthetic" salinity will be useful to the scientific community to validate physical models describing the location of the plume on the shelf.

One limitation of the algorithm is that it was developed primarily during the spring and summer flow periods, and only for a few years. The bio-optics of the river water may vary according to season as well as inter-annually, but because of limitations in the research schedule, it was not possible to develop a generalized proxy for year-round flow or for longer term changes in bio-optical signatures. Future work could improve on the model by including year-round measurements of $a_{CDOM}$ and salinity to extend the proxy to other seasons.

Chapter 2

A statistical clustering technique was developed to classify water masses based on characteristic bio-optical and environmental properties [*Palacios et al.*, in prep-a]. The algorithm relied on *in situ* measurements of salinity, temperature, and chlorophyll-a magnitude to group water masses into biogeochemically-based clusters. This is the first study to statistically define water masses within the plume based on

the biogeochemical properties of the water and not solely on the optical properties. This study also validated the water mass estimates with *in situ* measurements. This new, biogeochemically-based algorithm captured sub-mesoscale features within the plume that may be useful for validating other geophysical models of the CRP.

The time series analysis component of this study was less conclusive in tracking a water mass in time. The method can define these features spatially; it may be necessary to perform this type of analysis in a region with more consistent cloud-free days to adequately address temporal dynamics. This was not the case for this study: the region of the CRP experienced five completely cloud-free days during the cumulative 95-days of research cruise time used for model building and validation. Conducting this type of study in a region with more clear days would prove a better test of the time series analysis component of the study. Santa Barbara, CA channel would be an excellent candidate with its episodic winter storms interspersed with nearly cloud-free days for image capture.

Chapter 3

A semi-analytical, phytoplankton taxon discrimination algorithm (named "PHYDOTax") was developed for this study [*Palacios et al.*, in prep-b]. The algorithm was based on first principles of bio-optics and was applied to synthetic culture mixtures and to field observations of remote sensing reflectance. Validation demonstrated that PHYDOTax was a robust tool for discriminating algal taxa using hyperspectral ocean color data. Unique to PHYDOTax is its ability to discriminate

between dinoflagellates and diatoms using ocean color data. As a result, PHYDOTax may be a valuable tool for the Oceanography community to partition the chlorophyll pool into its constituent parts—especially in the coastal zone of the California Current System where diatoms alternate with dinoflagellates as the dominant taxon present. In addition, PHYDOTax is a new and important tool for the validation of ecosystem models, tracking carbon flow through different phytoplankton taxon-dominated systems, and the detection of harmful algae.

One of the biggest limitations of PHYDOTax may also be considered a feature- the representative phytoplankton taxa in the signature library. The efficacy of PHYDOTax is limited by which taxa are present in the signature library. Like its conceptual predecessor, CHEMTAX [*Mackey et al.*, 1996], PHYDOTax can only predict estimates of phytoplankton taxa in a natural sample from the member taxa in the library. Consequently, if PHYDOTax is applied to a region with considerably different taxa than in the existing library, the estimates will not be credible. However, what may appear to be a flaw—the limitation of the signature library—is also the algorithm's strength. PHYDOTax is flexible. The library can be changed with user-supplied pure culture estimates of remote sensing reflectance, $R_{rs}$. These $R_{rs}$ estimates for phytoplankton can be derived from the radiative transfer equations and taxon-specific inputs for light absorption, backscattering, and attenuation. With the appropriate signature library, PHYDOTax can be applied to many different regions of the surface ocean.

The second limitation of PHYDOTax is that it does not presently allow for an 'other' classification for estimates that do not belong to any library taxa. If the library is not representative of the environment, it is possible to mis-classify a natural sample into a library taxon that is not correct. A solution to this problem could be achieved in future iterations of the algorithm by including a null set in the computation.

Another limitation of the algorithm is that it can only be used for surface samples and does not account for the range of spectral shapes that may be expected for algae growing under varying light and nutrient concentrations. Both of these limitations are related to the effect of depth, biomass concentration, and biomass composition on the flattening of the peaks in the $R_{rs}$ spectrum. Flattening of the $R_{rs}$ spectrum has a direct effect on the results of the algorithm. One approach to address this limitation is to increase the number of $R_{rs}$ spectra for each taxon in the signature library. These additional spectra would represent the range of $R_{rs}$ spectra expected from varied light and nutrient growth conditions, the "package effect" due to chloroplast arrangement under high biomass conditions, and the effect of depth on the ambient light field. If it is possible to resolve the issue of the flattened $R_{rs}$ spectrum, the utility of PHYDOTax can be expanded to estimate taxon-specific biomass in depth profiles collected during research cruises or instrumented moorings such as MBARI's M1 mooring. The algorithm could also be applied to glider-based observations of the underwater light field to estimate taxon-specific biomass in three-dimensions, a data-set that could be used to validate three-dimensional models of ecosystems.

It is worth noting that this algorithm may address a number of questions important to Biological Oceanography. It may be a valuable tool in observing the total phytoplankton biomass pool partitioned into its constituent taxa. This has applications for the sub-disciplines of ecosystem modeling and harmful bloom detection. Another area where it could have an important impact is on validating Plankton Functional Type (PFT) models. PHYDOTax was developed to discriminate among actual taxa, not just phytoplankton size-class, which may afford the specificity needed to validate models with a large number of phytoplankton groups. Finally, the ability of PHYDOTax to discriminate between diatoms and dinoflagellates using ocean color data is especially useful in the coastal environment where these two taxa dominate and frequently co-occur.

Interestingly, two unexpected potential impacts of PHYDOTax were observed after the model was developed. The first is related to future satellite development. PHYDOTax uses hyperspectral data but at a resolution broader than what may be considered "hyperspectral" by the Ocean Optics community (10 nm vs. 1 – 2 nm). PHYDOTax demonstrates that it is possible to make robust algorithm predictions with lower-resolution optical data. When agencies weigh the advantages of developing hyperspectral remote sensing satellites, the cost of developing a high resolution sensor with its incumbent data acquisition and storage requirements can halt projects early in the development cycle. PHYDOTax, and algorithms with similarly less demanding spectral resolution requirements, are evidence that lower resolution satellites may be sufficient to ask Oceanographically relevant questions.

The second impact is related to applying the conceptual framework of PHYDOTax to existing multi-spectral satellite data. If it is possible to build an algorithm like PHYDOTax for multi-spectral data, then there is great potential to mine the historical data of multi-spectral sensors to characterize phytoplankton taxon distribution and abundance over time. This could provide insight into ecological succession over time, comparisons of El Niño and La Niña periods, anomalously large storm events, and the incidence of harmful algal bloom events. An algorithm such as this (revised for a multi-spectral sensor, tuned with a spectral library of phytoplankton appropriate for a region, with the number of taxa not exceeding the number of sensor wavebands) could be a valuable forensic tool for shedding new light on existing long-term data sets, possibly revealing hidden patterns with revelations of historical ecosystem development and climate change.

Literature Cited

Mackey, M. D., D. J. Mackey, H. W. Higgins, and S. W. Wright (1996), CHEMTAX - A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton, *Marine Ecology-Progress Series*, *144*(1-3), 265-283.

Palacios, S. L., T. D. Peterson, and R. M. Kudela (2009), Development of synthetic salinity from remote sensing for the Columbia River Plume, *Journal of Geophysical Research-Oceans*, *114*(C00B05).

Palacios, S. L., T. D. Peterson, and R. M. Kudela (in prep-a), Optical characterization of water masses within the Columbia River Plume.

Palacios, S. L., H. M. Sosik, K. K. Hayashi, M. Jacox, T. D. Peterson, and R. M. Kudela (in prep-b), Discrimination of phytoplankton taxa in an optically complex aquatic environment.