Programa de Doctorat en Enginyeria Òptica

Novel approach to improve the assessment of biodiversity of phytoplankton communities based on hyperspectral data analysis

Elena Torrecilla Ribalta

Supervisors

Dr. Jaume Piera Fernández, UTM/CSIC

Dr. Meritxell Vilaseca Ricart, CD6/UPC

A thesis submitted for the degree of

Philosophy Doctor (PhD)

July 2012

Abstract

Sustainable management of marine ecosystems requires a better knowledge about the space-time distribution and dynamics of ecological parameters such as phytoplankton communities, including critical bloom-forming algal groups. Better understanding of phytoplankton biodiversity and dynamics is essential in evaluating the role of each algal group in the global marine ecosystem and biogeochemical cycles. In attempting to address this question, in situ and remotely-sensed spectrometric optical observations have demonstrated to provide previously unavailable information regarding several optically active constituents in seawater at local and global scales, in particular, regarding phytoplankton community structure. In this sense, the advent of high spectral resolution (hyperspectral) optical sensors have raised new expectations about the possibilities of discriminating phytoplankton community composition in the ocean, beyond the estimation of only the primary pigment in phytoplankton, chlorophyll-a, a proxy for the phytoplankton biomass and primary production since it is common to all taxonomic groups. This PhD thesis has been carried out with the aim of improving our ability to extract information regarding phytoplankton community structure in the ocean by developing and evaluating a novel approach based on hyperspectral data analysis. In particular, a dissimilarity-based cluster technique, which accounts for complete spectral behaviour of hyperspectral data of each seawater sample, has been applied in combination with derivative spectroscopy, which exploits the spectral shape features of each analyzed spectrum. As a novelty, a validating tool has been proposed and proven useful to illustrate the effectiveness of the optical-based classification for discriminating different phytoplankton assemblages. This novel validation approach is based on the pigment composition analyzed in conjunction with concurrently obtained optical data, information which has been commonly used by the scientific community as a proxy for the phytoplankton composition. The feasibility of this methodology has initially been demonstrated using a simulation-based approach, i. e. using a radiative transfer modeling framework for open ocean and shallow coastal environments. In addition, different real open ocean environments corresponding to several stations in the eastern Atlantic Ocean have successfully been classified by applying the cluster analysis to different hyperspectral data sets including absorption and remote-sensing reflectance spectra and their second derivative spectra. This classification has served to identify a potential application of the proposed methodology: the establishment of different bio-optical provinces from the analysis of hyperspectral oceanographic observations, leading to examination of its biogeographical relevance by comparison to ecological provinces previously proposed in the literature. This thesis concludes by confirming the main hypothesis that discrimination of phytoplankton community structure and dynamics

in the ocean can be enhanced while using hyperspectral oceanographic observations. It is noteworthy that the proposed approach is generally applicable to different data sets, besides in-situ pigment or optical data data also to remotely-sensed, biogeochemical or hydrographic data sets.

Resum

La gestió sostenible dels ecosistemes marins requereix d'un millor coneixement de la distribució de certs paràmetres ecològics com les comunitats de fitoplàncton, inclosos aquells grups algals causants d'afloraments. La caracterització acurada dels patrons espacio-temporals de la biodiversitat del fitoplàncton és essencial per tal d'avaluar el paper de cada grup algal en l'ecosistema global marí i els cicles biogeoquímics. En l'intent d'abordar aquesta qüestió, observacions espectromètriques adquirides in situ i per teledetecció han demostrat proporcionar informació valuosa sobre la distribució de diversos components òpticament actius presents a l'aigua de mar, tant a escala local com global, i en particular, sobre l'estructura de les comunitats fitoplanctòniques. En aquest sentit, l'aparició de sensors òptics d'alta resolució espectral (hiperespectrals) ha augmentat les expectatives per discriminar la composició de les comunitats de fitoplàncton, permetent anar més enllà de l'estimació del pigment primari del fitoplàncton, la clorofil·la-a, utilitzat convencionalment com un indicador global de la biomassa i la producció primària de fitoplàncton ja que és un pigment comú a tots els grups taxonòmics. Aquesta tesi doctoral s'ha dut a terme amb l'objectiu de millorar la nostra capacitat d'extraure informació sobre l'estructura de les comunitats de fitoplàncton en l'oceà, mitjançant el desenvolupament i avaluació d'una nova aproximació basada en l'anàlisi de dades hiperespectrals. En particular, s'ha proposat una tècnica de classificació, on s'examinen les dissimilituds entre les signatures hiperespectrals de cada mostra d'aigua, i que ha estat aplicada en combinació amb l'espectroscopia derivativa, tècnica amb la qual s'exploren les característiques de la forma de l'espectre analitzat. Com a novetat, es proposa també una eina de validació per demostrar l'eficàcia d'aquesta nova tècnica de classificació. En el procés de validació es demostra que les classificacions obtingudes amb les dades òptiques i la metodologia proposada són molt semblants a les obtingudes amb dades basades en l'anàlisi de la composició pigmentària (utilitzant un cromatògraf líquid d'alta resolució, HPLC), tècnica habitualment utilitzada per la comunitat científica com a indicador de la composició de fitoplàncton. La viabilitat d'aquesta metodologia ha estat demostrada inicialment utilitzant una aproximació basada en simulacions, on la distribució del fitoplàncton està predeterminada i on s'han generat diferents escenaris lumínics d'aigües en mar obert i costaneres mitjançant un model de transferència radiativa. Per altra banda, escenaris reals de mar obert corresponents a diferents estacions en l'oceà Atlàntic han estat classificats satisfactòriament mitjançant les tècniques proposades, aplicades a dades hiperespectrals incloent espectres d'absorció i reflectància, així com els seus espectres derivats. Aquesta classificació ha servit per identificar una aplicació potencial de la metodologia proposada: l'establiment de diferents províncies bio-òptiques a partir de l'anàlisi de mesures hiperespectrals

oceanogràfiques, donant lloc a l'examen de la seva rellevància biogeogràfica en comparació amb províncies ecològiques proposades prèviament en la literatura. Aquesta tesi conclou amb la confirmació de la hipòtesi principal: una millor discriminació de l'estructura i dinàmica de les comunitats de fitoplàncton és possible mitjançant l'ús d'observacions hiperespectrals oceanogràfiques. Cal destacar que l'aproximació proposada és en general aplicable a diferents conjunts de dades, més enllà de la composició pigmentària o dades òptiques obtingudes in situ també a dades obtingudes per teledetecció, dades biogeoquímiques i altres paràmetres hidrogràfics.

Als meus pares,

l'Elena i el Sebastià,

perseverança i bondat

Agraïments/Acknowledgements

Durant tots els anys de desenvolupament d'aquesta tesi, he estat envoltada de persones que han estat fonamentals per poder fer-la realitat. Ara ha arribat el moment de que els agraeixi la seva contribució.

En primer lloc, el major agraïment és pels meus directors de tesi en Jaume Piera i la Meritxell Vilaseca. A en Jaume, creador inesgotable d'idees, li agraeixo l'haver-me proposat aquelles idees que van servir-me de llavor per anar desenvolupant el que ha esdevingut aquesta tesi. També l'haver-me introduït en aquest àmbit de recerca i la seva dedicació, així com la seva actitud motivadora i ment oberta en tot moment, valors que m'han permès anar conduint el meu treball amb llibertat i confiança. Ja als inicis de col·laborar al seu grup, em va dir que fer una tesi era com una "travessia del desert". Jo em sento afortunada, en aquesta travessia hi he trobat molta aigua. A la Meritxell, treballadora incansable, li estic molt agraïda pels coneixements i experiència que m'ha transmès al llarg d'aquests anys de supervisió. La seva actitud sempre activa m'ha servit en molts moments de referent per tirar endavant cada etapa de la tesi.

Aquesta tesi no hauria vist la llum sense el finançament per part del Ministeri d'Educació i Ciència, mitjançant els programes Junta d'Ampliació d'Estudis (JAE Predoc, I3P-BPD2005) i Estades Breus I3P de l'Agència Estatal Consell Superior d'Investigacions Científiques (CSIC), així com el programa de Menció Europea en el Títol de Doctor de la Secretaria General d'Universitats. La tesi ha estat realitzada en el marc dels projectes de recerca HIDRA (CSIC, PIE06-301102) i ANERIS (CSIC, PIF08-015).

També m'agradaria donar les gràcies als meus companys de grup i despatx, l'Isma i el Sergi. Tot i que cadascú ha hagut d'anar resolent la seva tesi des de punts de partida diferents, ha estat un plaer poder fer-ho en paral·lel amb ells. Crec que tots hem après molt amb l'experiència, no sempre ha estat senzill, però espero que aviat també puguin culminar la seva travessia doctoral. Tampoc vull deixar de mencionar el meu agraïment a la resta de companys de la UTM pel bon ambient i estones d'esbarjo o dinars que hem compartit. Així com a tot el personal directiu i administratiu del CMIMA per posar-m'ho fàcil en els múltiples tràmits que he hagut de resoldre durant aquests anys. Alguns investigadors de l'ICM també han contribuït a que pogués dur a terme aquesta tesi. Entre ells, a l'Elisa Berdalet li agraeixo que sempre estigués a disposició per resoldre'm tots aquells dubtes en relació a alguns conceptes de biologia marina que estaven fora dels meus coneixements. A l'Antonio Turiel, pel seu interès en la nostra recerca i per tot sovint aconseguir-me alguns articles als quals jo no tenia accés.

During my stay at Scripps (La Jolla, San Diego, US), I had the chance of being surrounded by an excellent group of people, and not only in scientific terms. I would like to express my deepest gratitude to Prof. Dariusz Stramski for introducing me to marine optics, being my scientific mentor and ensuring me an excellent stay in his lab. It was a pleasure to share many fruitful discussions with him and the rest of people from his lab (Rick Reynolds, Julia Uitz, Florence Lahet, Vanessa Wright, Jean and Guangming).

My thesis also involved a stay at AWI (Bremerhaven, Germany) and IUP (Bremen, Germany). I want to thank Prof. Astrid Bracher, who always welcomed me to her lab in a wonderful way, even when my proposals for visiting were just sent few days before the deadline. I appreciate her support and willingness to provide advice and help on any subject. I also want to thank the graduate students and post-docs (Bettina, Ali, Mariana, Anja, Mirjam) for making my stay and our research cooperation two much more enjoyable experiences than I could ever expect.

I would like to thank Marcel R. Wernand and Astrid Bracher for accepting to evaluate my work, as part of the external reviewing process before submitting this PhD thesis within the International Mention Program. Also thanks to Oliver Zielinski and Julia Busch for allowing me to participate during the field work performed at the Ebro Delta in 2011.

I am also grateful to the image artist Glynn Gorick for kindly giving me permission to use the image shown in the thesis cover.

La persona que m'ha anat seguint els passos de més aprop al llarg de la tesi és en Jordi. A ell li agraeixo la paciència que ha tingut en cada fase i les ganes de compartir amb mi els seus dots de "coaching". Amb ell he pogut posar final a aquesta tesi i iniciar la nova "tesi" que ara farem conjuntament amb els dos petits que venen de camí. Vull agrair a tots els meus amics (ja sabeu qui sou) per ser-hi des de fa tants anys i especialment per acompanyar-me en tot el que ha anat passant durant els anys de tesi. En aquest sentit, també vull agrair a la Laura que hagi estat la meva aliada en els moments més difícils.

En últim lloc però de forma molt especial, vull agrair als meus pares i germans el recolzament que sempre m'han donat, tot i que sovint se'ls hagi fet complicat entendre en el que he estat enfeinada. Sobretot amb ells vull compartir la gran satisfacció d'haver culminat aquesta tesi.

Contents

List of Figures

List of Tables

Introduction

Motivation

The oceans cover two-thirds of the Earth's surface and contain 97% of the planet's water, the most important substance for life on Earth. They have a profound influence on its environment since the oceans contain 99.5% of our planet's livable habitat. Monitoring and protecting water resources, including the oceans, is therefore a must and a challenge for all countries of the world. In this sense, the United Nations (UN) recommended the proclamation of an International Decade for Action from 2005 to 2015, "Water for Life", which provided an excellent opportunity for the international community to advance towards a truly integrated approach to the management of the world's water. In accordance, authorities and initiatives like the European Union Water Framework Directive (WFD, 2000/60/EC), the Global Ocean Observing System (GOOS) of UNESCO, or the European Network Global Monitoring for Environment and Security (GMES) were established as frameworks to assess the marine role in Earth system dynamics by promoting a dense and frequent monitoring of the ocean's composition. Of particular importance are climate research related monitoring applications as to record indicators of the ocean's response to climate change, and to calibrate and validate biogeochemical ocean-ecosystem models (LeQuéré *et al.*, 2005).

Sustainable management of oceanic ecosystems requires comprehensive information on phytoplankton community composition and distribution. Phytoplankton, as primary producers that convert light and CO_2 into biomass through the photosynthesis, stand at the base of the oceanic food web. They are important to monitor for being an indicator of the biological state of the oceans. Their importance as a food source for the pelagos and as a potential sink of atmospheric carbon have been widely recognised (Holligan, 1992). Some phytoplankton groups have also been identified to have an important role in the climate regulation and the atmospheric sulfur cycle (Fig. I.1a). These groups named DMS-producers are able to produce the cellular component dimethylsulfoniopropionate (DMSP), which once converted into sulfate molecules are emitted to the atmosphere and act as cloud condensation nuclei, increasing general cloud cover (Vallina and Simó, 2007). In addition, some phytoplankton species are able to cause harm through the production of toxins or by their accumulated biomass. Such outbreaks which typically occur in coastal areas are known as "harmful algal blooms" (HABs, Babin *et al.*, 2008). Social impact of environmental changes associated to algal blooms is significant (Fig. I.1b). When blooms are formed the toxins travel up the food chain and pose serious human-health threats and severely affect numerous industries and commercial fisheries, causing substantial economic losses in the form of reduced sales or diminished tourist activity.

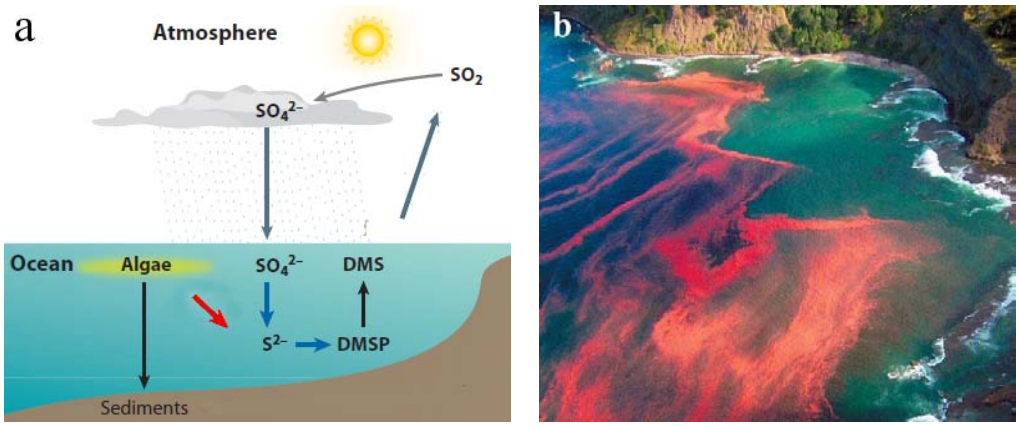


Figure I.1: (a) Diagram illustrating the role of phytoplankton in the sulfur cycle. *Source:* modified from Takahashi *et al.* (2011). (b) Episode of a "red tide" algal bloom (non-toxic) of *Noctiluca scintillans* in New Zealand. *Source:* M. Godfrey.

Phytoplankton have a pivotal role in driving and maintaining aquatic biodiversity and the marine food web as a whole. Causative links between phytoplankton growth, abundance, species success and succession and how they are affected by the driving environmental forces need to be established. Moreover, the understanding of the regulation of phytoplankton productivity and biomass in the ocean is crucial for the understanding on effects of global climate change by radiative forcing to the ocean's food web structures and carbon fluxes. In this sense, there is an important need for better information on the spatial distribution of biological activity in the upper ocean and its temporal variability, especially in the case of oceanic phytoplankton biomass. Further research is therefore needed to develop new techniques that enable to recognize phytoplankton-related ecological changes and perturbations.

For all the above-mentioned compelling reasons, increasing emphasis is being placed upon scientific research and monitoring of the phytoplankton biodiversity and dynamics. In attempting to monitor spatial and temporal variations of the phytoplankton community structure at a local and global scale, in situ and remote sensing optical observation represents the most suitable, fastest and less invasive method currently available for synoptically measuring wide-area properties of ocean ecosystems. Traditional description of global phytoplankton abundance and distribution involved the monitoring of the primary pigment in phytoplankton, chlorophyll-a, a proxy for the phytoplankton biomass. In fact, numerous studies over the past three decades have focused on the development of bio-optical algorithms linking different processes to this pigment (O'Reilly *et al.*, 2000; Reynolds *et al.*, 2001). However, biogeochemical cycling in marine systems is intimately linked to the activity of specific phytoplankton groups (see Table I.1), meaning that phenomena such as the production of sulfur components like DMSP or proliferation of algal blooms are taxon-dependent (Vila-Costa *et al.*, 2006; Stumpf *et al.*, 2003). Further efforts are thus needed to go beyond the estimation of only chlorophyll-a, which is common to all taxonomic groups.

Size Class	Taxonomic group	Phytoplankton Functional Group (biogeochemical role)	Diagnostic Pigment
Microplankton (> 20 µm)	Diatoms	Silicifiers	Fuco
	Dinoflagellates	Algal blooms (red tides, see Fig. 0.1)	Peri
Nanoplankton (2-20 µm)	Prymnesiophytes		Hex
	E. Huxleyi sp.	Calcifiers, DMS-producers	
	Phaeocystis sp.	DMS-producers	
	Chrysophytes		But
	Cryptophytes		Allo
	Chlorophytes		MVChl*b*
Picoplankton (<0.2 µm)	Pico-eukaryotes		Hex, Chl*b*
	Cyanobacteria		
	Synechococcus sp.		Zea, Phycoerythrin
	Prochlorococcus sp.		Zea, DVChla, DVChl*b*
	Trichodesmium sp.	Nitrogen-fixers	Phycoerythrin

Table I.1: Summary of properties of different phytoplankton groups, which play a specific role in the marine ecosystem and are characterized by specific bio-marker pigments. Source: generated from tables in Nair *et al.* (2008), Hirata *et al.* (2008) and Hirata *et al.* (2011).

Mapping marine phytoplankton community structure from a hyperspectral perspective

Today, in situ and remotely-sensed spectrometric optical observations of ocean waters have the potential to provide information regarding the concentrations of optically signicant constituents in seawater, in particular, regarding phytoplankton community structure. In fact, efforts to expand the use of spectrometric measurements for estimating some biogeochemically important ocean variables and phytoplankton-related phenomena have increased considerably. For example, in recent years, a variety of bio-optical methods have been established (Nair *et al.*, 2008; Brewin *et al.*, 2011) that use satellite spectrometric data to identify different phytoplankton communities through the detection of either Phytoplankton Functional Types (PFTs, i.e., phytoplankton that have a specific biogeochemical function such as calcification, silicification, DMS production or nitrogen fixation) or Phytoplankton Size Classes (PSCs, i.e., the autotrophic pool is partitioned according to cell size, which may be useful to better understand their contribution to global ocean primary production).

There are two main types of approaches to derive PFTs and PSCs from remote sensing ocean color data: (1) the abundance-based approach, in which the eutrophic status of the waters, indicated by the chlorophyll concentration or the magnitude of the absorption optical coefficient of phytoplankton, is related to community structure (Sathyendranath *et al.*, 2001; Uitz *et al.*, 2006; Devred *et al.*, 2006 and 2011; Aiken *et al.*, 2007; Hirata *et al.*, 2008 and 2011; Brewin *et al.*, 2010a; Mouw and Yoder, 2010) and (2) the spectral response approach, in which different PFTs and PSCs are detected providing they have contrasting optical signatures or traits (Sathyendranath *et al.*, 2004; Alvain *et al.*, 2005 and 2008; Ciotti and Bricaud, 2006; Bracher *et al.*,

2009; Brewin *et al.*, 2010b; Sadeghi *et al.*, 2012). All types of approaches have their limitations to describe phytoplankton biodiversity due to the fact that for instance the same taxonomic size class may contain phytoplankton of different functional types (see some examples in Table I.1). On the other hand, abundance-based algorithms may fail to distinguish between algal blooms of different PFTs or PSCs that have the same biomass. And a size-based classification alone may not be fully satisfactory from a biogeochemical perspective, since the functional role of phytoplankton in marine ecosystems also depends on the nutrient, iron, or light uptake. Furthermore, in the spectral-response approaches, accurately exploiting the spectral characteristics of different PFTs or PSCs to identify and distinguish among them may not always be successful. In this sense, further work is still required to assess the reliability of these bio-optical approaches. In order to confidently use these satellite products, as with any satellite-derived geophysical or biogeochemical product, validation exercises remain a key issue to ascertain accuracy and limitations (Brewin *et al.*, 2011). This is of particular importance given that the detection of PFTs from satellite data is a field of research in its early stages of development.

Although it is becoming a reality the global monitoring of certain dominant phytoplankton taxonomic and size groups from satellite data, the real challenge of understanding the phytoplankton biodiversity in the ocean lies in the identification of different groups of phytoplankton under non-bloom or dominant conditions (i.e., when multiple phytoplankton assemblages are present in the water) or in coastal waters (i.e., more optically complex environments). In order to overcome this challenge, technological advances and especially the advent of high spectral resolution (hyperspectral) sensors have raised new expectations about the possibilities of spectrally discriminating phytoplankton community composition in the ocean. Compared to multispectral sensors that collect optical data at a limited number of wavelengths, hyperspectral sensors give continuous spectral coverage over a broad wavelength range with better than 10 nm resolution and have opened the possibility for optical oceanographers to more accurately characterize complex oceanic environments (Chang *et al.*, 2004).

In all the above-described approaches, except in the method developed by Bracher *et al.* (2009), analyses of single band or single band-ratios obtained in discrete multispectral bands were employed to retrieve concentrations of the pigment chlorophyll-*a* and resolve the presence of specific PFTs or PSCs. The advantage of a hyperspectral-based over a these multispectral-based inversion methods for phytoplankton biodiversity identification stems from the fact that more accurate spectral information, such as spectral features related to characteristic phytoplankton pigment absorption peaks, is resolved. Hyperspectral observations are thus a key feature due to more subtle differences in targets are discriminated. This fact suggests that the effectiveness of hyperspectral optical information for assessing phytoplankton community structure in ocean water, even in non-bloom conditions, should be further explored.

Objectives

Better understanding of phytoplankton biodiversity and dynamics is essential in evaluating the contribution of each algal group in the global marine ecosystem and biogeochemical cycles. In order to overcome this challenge, the increasing availability of in situ and remotely sensed hyperspectral measurements of ocean waters raises new expectations and leads to a need for ongoing evaluation of high-spectral-resolution processing methods. In this thesis, it is attempted to attain one step further in monitoring of phytoplankton communities in the ocean by developing and evaluating a novel processing approach based on hyperspectral data.

The central question of this thesis is: given the capability to obtain hyperspectral oceanographic observations, how can our ability to extract information regarding phytoplankton community structure and dynamics in the ocean be improved? Other specific questions addressed in the framework of this thesis are:

- what role can the shape of hyperspectral data play in the assessment of phytoplankton composition of a seawater sample?,
- how effective can a hyperspectral-based approach be at discriminating different oceanic environments in which various phytoplankton groups co-exist under non-bloom or non-dominant conditions?,
- how feasible is the global identification of different types of water masses in terms of phytoplankton composition by assessing their hyperspectral optical signatures?
- what effect does a hyperspectral-based approach have on the analysis of phytoplankton community structure of both types of environments: open ocean or coastal waters?

Thesis outline

In order to answer all the above listed questions, the manuscript has been structured as follows[1]:

Chapter 1 first provides a general introduction to bio-optical oceanography, in which some key concepts are detailed regarding the radiative transfer in the ocean and biology of phytoplankton. In addition, the most relevant optical-based technologies to assess phytoplankton biodiversity are reviewed, including the advances achieved using the full potential of hyperspectral measurements. Finally, the role of radiative transfer modelling in this thesis is also emphasized.

In Chapter 2, a novel methodology which exploits the spectral shape features of hyperspectral data is proposed with the aim to automatically identify different phytoplankton communities in the ocean. As a first approach, a synthetic hyperspectral data set is generated using a oceanic radiative transfer model. A simulation-based framework, representing several simplified environmental scenarios in terms of phytoplankton composition, is used in this chapter to test the feasibility of the proposed methodology.

[*Torrecilla, E., J. Piera and M. Vilaseca (2009). Derivative analysis of oceanographic hyperspectral data. In G. Jedlovec (Ed.), Advances in Geoscience and Remote Sensing (pp. 597–619). Vienna: InTech.*]

Chapter 3 examines the potential of the methodology described in the previous chapter for classifying a small but carefully selected data set from real open ocean environments. The analysis is applied to hyperspectral absorption and remote-sensing reflectance measurements from the eastern Atlantic Ocean in 2005, corresponding to non-bloom conditions in terms of phytoplankton abundance and composition. Radiative transfer modelling is only considered in this chapter to reconstruct the hyperspectral remote-sensing reflectances at each station, since only multispectral measurements were carried out in the field. In order to illustrate the effectiveness of the optical-based classification, a new validating tool is proposed against in situ pigment information from each evaluated station (i.e., against a commonly used proxy for the phytoplankton composition). In addition, different analyses are conducted to demonstrate the important role in the performance of the selected spectral range and other parameters involved in the analysis of hyperspectral data.

[*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 Environment, 115, doi: 10.1016/j.rse.2011.05.014:25782593.*]

[1] articles directly arised from the elaboration of the thesis are referred between brackets in the corresponding chapter.

Chapter 4 is devoted to provide some results demonstrating the potential of the methodology described in the previous chapters by examining a larger database composed of 48 stations from the eastern Atlantic Ocean in 2008. In this case, the analysis is applied to both field hyperspectrally-resolved measurements of absorption and remote-sensing reflectance spectra. It is also discussed how the identification of several phytoplankton assemblages spatially distributed along the covered transect serve to demonstrate the feasibility of this approach to define different bio-optical provinces based on the phytoplankton community structure and their hyperspectral signatures. It is qualitatively examined the biogeographical relevance of these bio-optical provinces by comparison to ecological provinces previously proposed in the literature.

[*Taylor, B. B., E. Torrecilla, A. Bernhardt, M. H. Taylor, I. Peeken, R. Röttgers, J. Piera and A. Bracher (2011). Bio-optical provinces in the Eastern Atlantic Ocean and their biogeographical relevance. Biogeosciences, 8, 3609-3629, www.biogeosciences.net/8/3609/2011/, doi:10.5194/bg-8-3609-2011.*]

In Chapter 5, the role of hyperspectrally-resolved optical observations is also assessed but when considering optically more complex underwater scenarios. As a first approach to shallow estuarine environments and by means of model simulations, the sensitivity of remote-sensing reflectance spectra to the phytoplankton taxa, the bottom type, the abundance of suspended matter and the vertical structure of phytoplankton is examined. This study is focused on several underwater scenarios emulating similar conditions to those found in one of the most studied coastal sites in Catalonia: the Alfacs Bay (Ebro Delta, NW Mediterranean Sea).

Chapter 6 finally summarizes the results achieved in this thesis and points the direction of future work. Some guidelines for future developments and applications are proposed.

Additional material that is relevant to several chapters is provided in different Appendices. A detailed description of the oceanic RTE equation and its link with some of the optical properties is given in Appendix A. An independent corroboration for the pigment analysis performed in Chapter 3 is provided in Appendix B. Appendix C is devoted to describe a preliminary characterization and correction of the stray-light radiation on the response of the two in-water array spectroradiometers considered in one of the research projects in which this thesis has been developed.

1 Bio-optical oceanography: general background

This chapter, with the aim of underpinning the research work presented in the following chapters, provides a general overview of bio-optical oceanography. First, an introduction to fundamentals of the radiative transfer in the ocean and phytoplankton biology is supplied. Hence, the relevant optical-based technologies that have recently been developed to assess phytoplankton biodiversity are reviewed, with special attention to the hyperspectral technology. Finally, a brief overview of the role of radiative transfer models in this field of research is presented.

1.1 Oceanic radiative transfer and phytoplankton biology

It has long been recognized that transmission of light in sea water is essential to the productivity of the oceans. It provides the energy necessary for ocean currents, and the majority of marine life is supported by the thin layer of warm water near the ocean's surface. As it will be further described in greater detail, light plays a decisive role in the primary formation of biomass by oceanic phytoplankton (i.e., chlorophyll-bearing marine plants) through the process of photosynthesis, which is the basis of the entire marine food chain. Light transmission is therefore a key factor in the ecology of the upper ocean and biogeochemical cycling, since it has a strong influence on the dynamics of their chemical compounds.

Ecologists have traditionally classified ocean habitats and their organisms on the basis of light levels and depth. The major life zones in the ocean are shown in Fig. 1.1. The open sea is divided into several sections. The **euphotic zone** is the depth of the water that is exposed to sufficient sunlight for photosynthesis to occur and it can be affected greatly by seasonal turbidity. It extends from the surface downwards to a depth where light intensity falls down to 1% of that at the surface (i.e., the euphotic depth). Accordingly, its thickness varies widely on the extent of light attenuation in the water column. Typical euphotic depths vary from only a few centimetres in highly turbid waters to around 200 metres in the open ocean. About 90% of all marine life lives in the euphotic zone. The aphotic zone is the portion of the ocean where there is little or no sunlight. Consequently, bioluminescence is essentially the only light found in this zone. Most food comes from debris and dead organisms sinking to the bottom of the ocean from overlying waters. It comprises the **bathyal zone** (or middle zone) and the **abyssal zone**, where sunlight does not penetrate although there are many more nutrients than in the euphotic zone. The coastal ocean is the area of the ocean close to land, which tends to be warm and shallow with many nutrients available. The

higher sunlight penetration and the runoff of water from inland bringing nutrients from soil and plants, contribute to the fact that although coastal ocean comprise less than 10% of the ocean's total area, 90% of marine species inhabit this region. Coral reeefs, known for their great variety of organisms, can be found in this area.

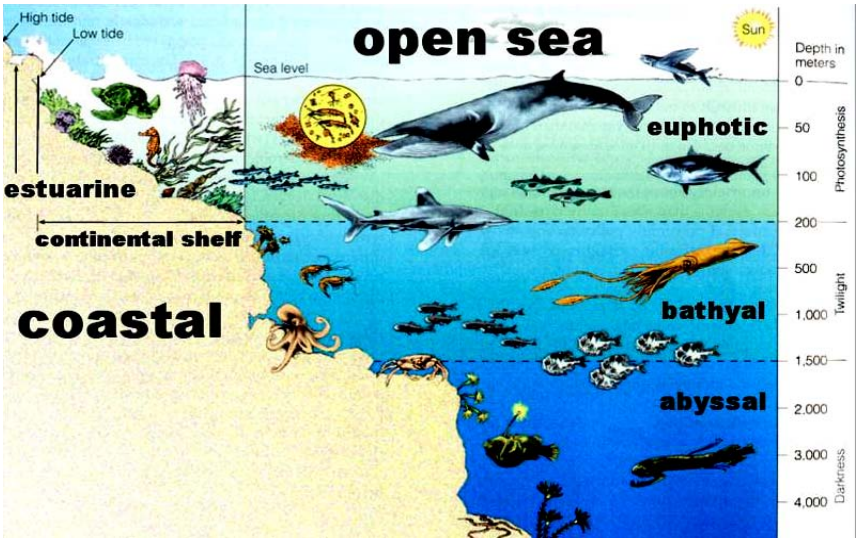


Figure 1.1: Diagram illustrating the major life zones and vertical zones in an ocean (not drawn to scale), based on the available light (*Source*: Miller, 2007).

Variability of light in the ocean is strongly influenced by the distribution of the components in the water column, which varies across horizontal and vertical space and time scales. In this sense, ocean waters vary in color from the deep blue of the open ocean to yellowish-brown in a turbid region.

The most optically significant water constituents are briefly described as follows[1]:

- **Pure water**

  Water itself limits the wavelength range of interest in optical oceanography from the near-ultraviolet to the near infrared because it is highly absorbing at wavelengths below 250 nm and above 700 nm.

- **Inorganic particulate matter (minerals)**

  These particles are created by weathering of terrestrial rocks and are entered the water as wind-blown dust settles on the sea surface, as rivers carry eroded soil to the sea or as currents resuspended bottom sediments. They can dominate water optical properties when present in sufficient concentrations.

[1] Air bubbles generated by breaking waves at sea surface can significantly contribute to scattering properites of a water body (Stramski *et al.*, 2004). Nevertheless, their effect is beyond the scope of the research presented in this thesis and therefore it is not described in detail.

- **Colored Dissolved Organic Matter (CDOM or gelbstoff)**

  These compounds are produced during the decay of plant matter. In high concentrations these compounds can color the water yellowish-brown. CDOM absorbs very little in the red, but absorption increases rapidly and exponentially with decreasing wavelength. It can be the dominant absorber at the blue end of the spectrum, especially in coastal waters influenced by river runoff. Scattering by CDOM is generally considered negligible.

- **Particulate organic matter**

  Phytoplankton are microscopic plants (see Fig. 1.2), which occur with incredible diversity of species, cell shapes, concentrations (from less than 0.03 $mg\,m^{-3}$ in oligotrophic waters up to about 30 $mg\,m^{-3}$ in eutrophic waters) and sizes (3 phytoplankton size classes (PSCs): microphytoplankton from 20 to 200 μm, nanophytoplankton from 2 to 20 μm and picophytoplankton < 2 μm).

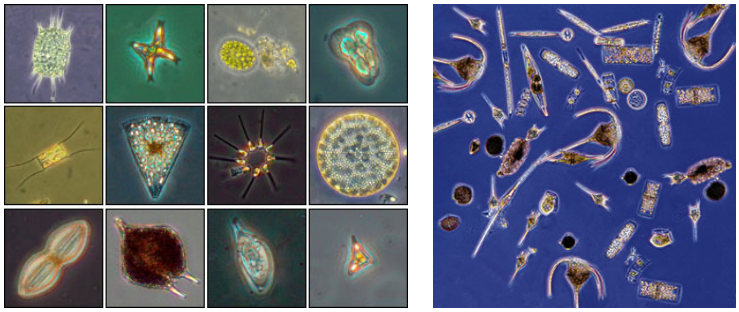


Figure 1.2: An astonishing diversity of phytoplankton live in Earth's oceans (*Source*: Smithsonian Environmental Research Center and U.K. Fisheries Research Service).

These algal organisms have an essential role in the ocean since, through the process of **photosynthesis**, they convert inorganic materials into organic compounds (Kirk, 1994). Phytoplankton, as primary producers of biomass, form the first link of the oceanic food chain that reaches through all trophic levels. In order to photosynthesise, phytoplankton require carbon dioxide (CO_2), water (H_2O) and light energy. The process itself occurs within the chloroplasts which contain the **photosynthetic pigments** that convert the radiant to chemical energy. The dominant pigment is chlorophyll-a which is common to all species, but there are also cholorphylls-b and -c plus several accessory pigments (i.e., carotenoids, xantophylls and the photoprotective phycobiliproteins: phycoerythrin and phycocyanin). Each of the above pigments is able to absorb light of wavelengths within the visible range with strong spectral selectivity (see Fig. 1.3). This range (approx. from 400 to 700 nm but occasionally slightly extended towards the near infrared and near ultraviolet regions) is referred to as photosynthetically available radiation (PAR) because it provides energy to marine

phytoplankton to perform photosynthesis. In fact, the visible and near infrared (approx. from 400 to 800 nm) constitute the only portion of the electromagnetic spectrum that penetrates water and directly probes the water colum (see Fig. 1.4). The solar spectrum peaks correspond to the maximum transparency of water and the peak in phytoplankton absorption. Thus, phytoplankton photosynthesis is tuned to the spectral range of maxium light.

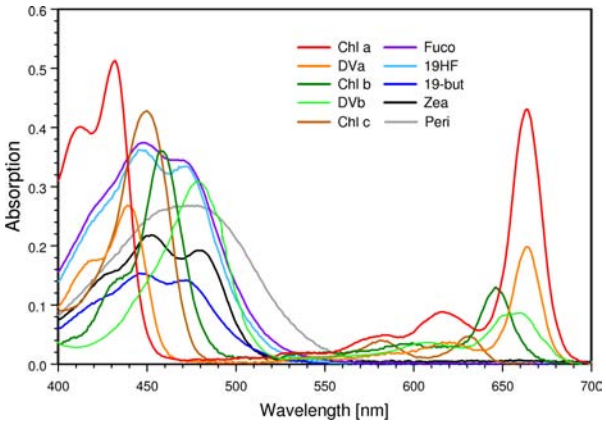


Figure 1.3: Absorption spectra of several individual phytoplanktonic pigments in solvents. Pigment abbreviations are: Chl a = monovinyl chlorophyll-a, DV a = divinyl chlorophyll-a, Chl b = monovinyl chlorophyll-b, DV b = divinyl chlorophyll-b, Chl c = chlorophyll-$c1$+$c2$, Fuco = fucoxanthin, 19HF = 19'-hexanoyloxyfucoxanthin, 19-But = 19'-butanoloxyfucoxanthin, Zea = zeaxanthin, Peri = peridinin (*Source*: Taylor *et al.*, 2011).

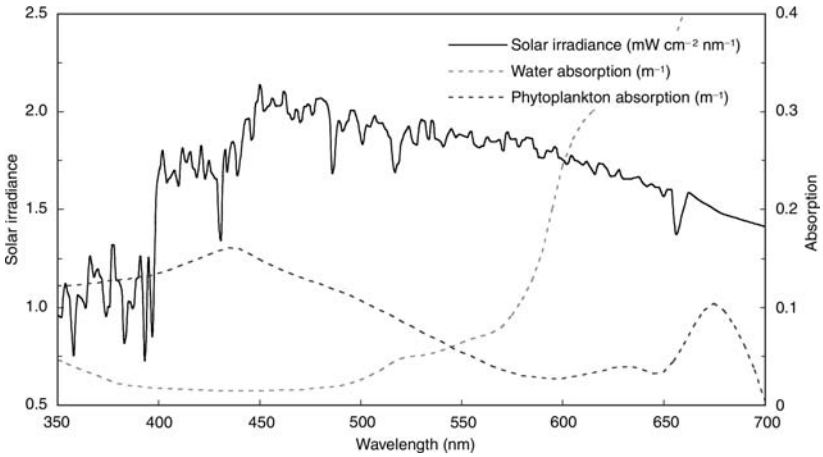


Figure 1.4: Solar irradiance, phytoplankton specific absorption and water absorption spectra (*Source*: McClain, 2001).

Each phytoplankton group differ significantly with respect to motility, cell-wall and pigment composition, likewise to nutrition and reproduction strategies. At the same time, all species are at the mercy of oceanic currents for transport to areas that are suitable for their survival and growth. Thus, physical processes and light conditions for phytoplankton photosynthesis play both a significant role in determining the composition and distribution of phytoplankton species. Phytoplankton, with a great diversity of species, are responsible for determining the optical properties of most oceanic waters.

Bacteria are the smallest living organisms on Earth (i.e., size range 0.2 - 1.0 μm). They can be significant scatterers and absorbers of light at blue wavelengths in oceanic waters, where the larger phytoplankton are relatively scarce.

Detritus are non living organic particles of various sizes that are produced when phytoplankton die and their cells break apart or when zooplankton graze on phytoplankton and leave cell fragments and fecal pellets. They have significat absorption only at the blue wavelengths. They can, however, contribute significantly to scattering in the open ocean.

Having identified the most optically significant water consituents, which are responsible for diversity in color of oceanic waters (see some examples in Fig. 1.5), a bipartite classification scheme for oceanic waters has traditionally been used in optical oceanography. It was introduced by Morel and Prieur (1977) and refined later by Gordon and Morel (1983) and Sathyendranath and Morel (1983). A pictorical representation of the two cases (i.e., Case 1 and Case 2) is presented in Fig. 1.5. Moreover, the component absorption and scattering coefficients for two hypothetical water bodies representing each of them are depicted in Fig. 1.6.

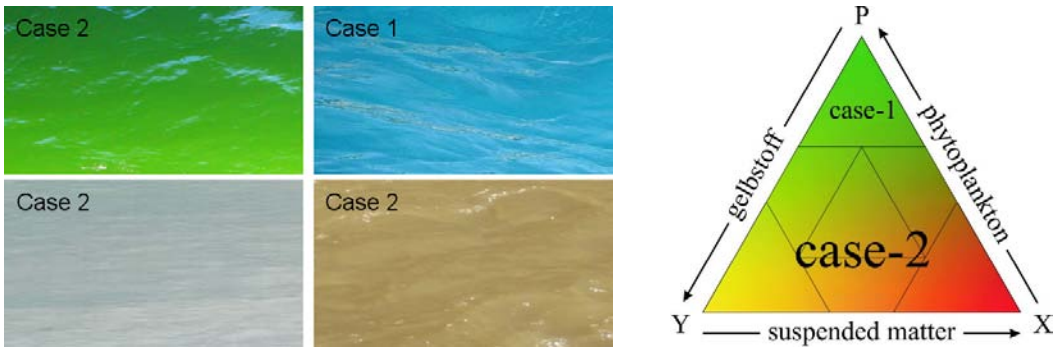


Figura 1.5: Examples of different colored sea areas and diagram representing the Case 1 and Case 2 waters. Case 1 waters represent the phytoplankton-dominant cases, whereas Case 2 waters represent all other possible cases (*Source*: Sathyendranath, 2000).

'Case 1 waters' are those in which the contribution by phytoplankton to the total absorption and scattering is high compared to that by other substances. Therefore, absorption by chlorophyll and related pigments plays the dominant role in determining the total absorption in such waters, although covarying detritus and dissolved organic matter derived from the phytoplankton can also contribute to absorption and scattering in these waters. Case 1 waters can range from very clear (i.e., oligotrophic, phytoplankton-poor) to very productive (i.e., eutrophic, phytoplankton-rich) water, depending on the phytoplankton concentration. On the other hand, **'Case 2 waters'** are 'everything else,' namely, waters where inorganic particles in suspension or colored dissolved organic matter (CDOM) from land drainage contribute significantly to the absorption and scattering. In this sense, the effect of pigments is relatively less important in determining the total absorption. Roughly 98% of the world's open ocean and coastal waters fall into the Case 1 category, but near-shore and estuarine Case 2 waters are disproportionately important to human interests such as recreation, fisheries and military operations. Even this classification scheme cannot be implemented in a very strict quantitative fashion, its value lies most in its use as a simple device to differentiate waters where phytoplankton-related features dominate the signal (see Chapters 2, 3 and 4) from more optically complex water bodies, where such simplifying assumptions would not hold (see Chapter 5).

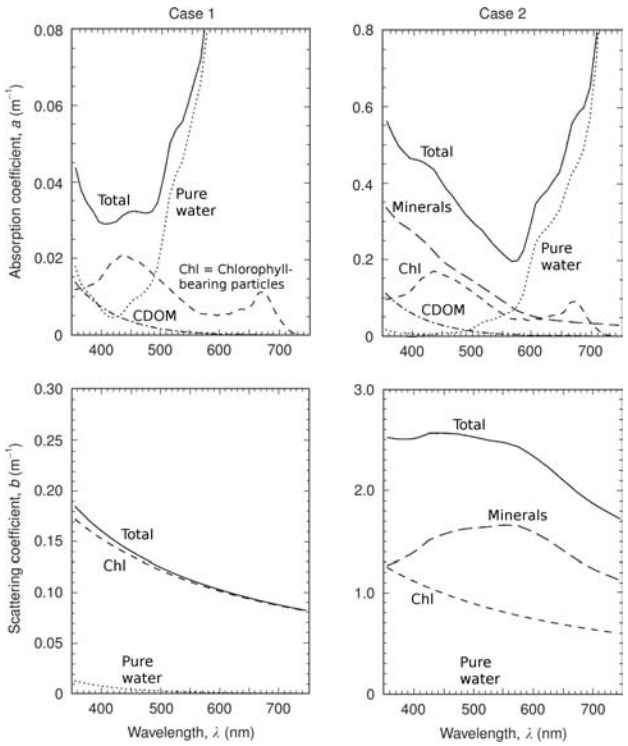


Figura 1.6: Absorption and scattering coefficients for the Case 1 and 2 water bodies and the contribution by each water consituent (*Source*: Mobley, 2001).

For the Case 1 scenario, the total absorption is dominated by phytoplankton (i.e., chlorophyll-bearing particles) at blue wavelengths and by the pure water at wavelengths greater than 500 nm. The total scattering is also dominated by phytoplankton. In the Case 2 scenario, mineral particles present an absorption comparable to or greater than that by the chlorophyll-bearing particles and are the primary scatterers.

In both Case 1 and Case 2 waters, oceanographic optical measurements can be important indicators of the health of the ocean in the form of changing turbidity together with changing diversity and distribution of phytoplankton species. In this sense, the assessment of optical properties of the ocean can help elucidating the link between the phytoplankton communities and several biogeochemical processes in which phytoplankton play a key role. For instance, as it has been described in the thesis introduction (see Fig. I.1), practical applications of knwoledge of plankton diversity and distribution in the oceans include ecosystem responses to climate change (including the carbon and sulfur cycle), detection of harmful algal blooms in coastal waters, food web modelling and management of marine resources. Knowledge of the variability of bio-optical properties of the ocean is essential for understanding all these processes.

The subdiscipline of **ocean optics**, which concerns the assessment of the propagation of light through the oceanic water column and surface, has become fundamental for understanding water dynamics and composition, including phytoplankton communities distribution. This subdiscipline is based on the radiative transfer (RT) theory (Kirk, 1994; Mobley, 1994), utilized as the theoretical framework and defined to better comprehend light propagation in the optically complex ocean. Radiative transfer is therefore a complex field that requires physical understanding of the absorption and scattering processes in the atmosphere and oceans, as well as sea surface and bottom reflection characteristics (see Fig. 1.7). In this sense, optical properties of sea water are divided into two classes:

- **Inherent Optical Properties (IOPs)** are those properties that are only affected by the constituents of the aquatic medium and are independent of the ambient illumination conditions. Examples of IOPs are the spectral absorption and scattering coefficients. IOPs are additive, which means that they result from additive contributions associated with all individual water components that absorb and scatter light (see Fig. 1.6). Therefore, their variation is generally attributed to variability in four constituents of the aquatic medium in natural oceanic waters: pure sea water, phytoplankton, non-algal particles and colored dissolved organic matter (CDOM).

- **Apparent Optical Properties (AOPs)** are not additive and depend both on the medium (i.e., on the IOPs) and on the directional structure of the ambient light field. Examples of AOPs are the spectral radiance and irradiance distributions.

Historically, the IOPs were easy to define and interpret in terms of water constituents, but hard to measure. This is, however, less true today because of advances in instrumentation. In contrast, the AOPs are much easier to measure and readily available, but are often more difficult to interpret because of the confounding environmental effects. For instance, a change in the state of the sea surface or in the sun's position modifies the spectral radiance distribution, and hence the AOPs, even though the IOPs are kept unchanged.

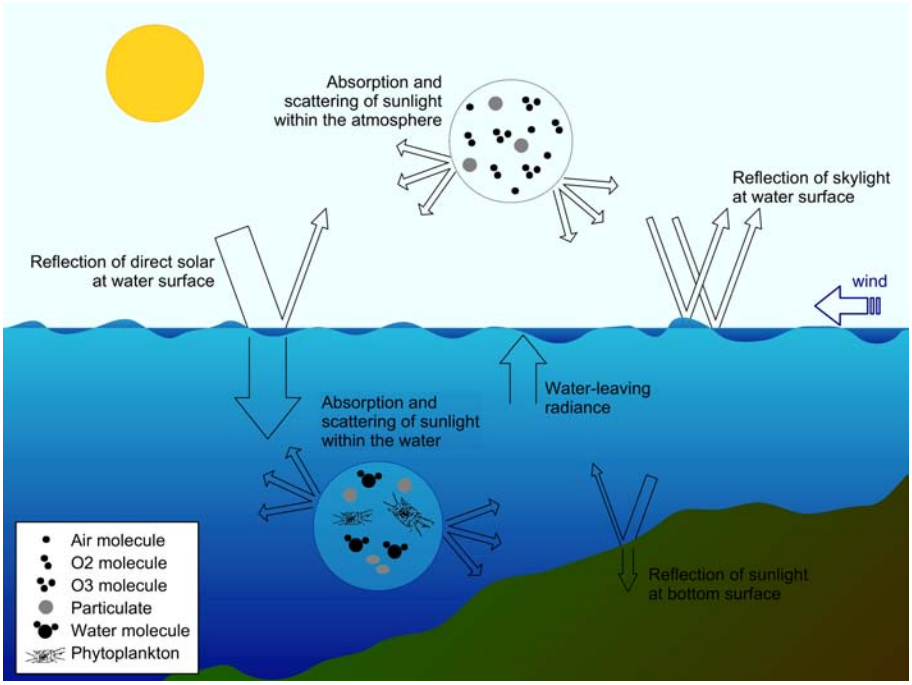


Figura 1.7: Depiction of the various optical pathways and interactions within the atmosphere-ocean system.

Radiative transfer theory provides the connection between the IOPs and the AOPs of a water body and is expressed as a mathematical structure by the **Radiative Transfer Equation (RTE)**. The physical environment of a water body – waves on its surface, the character of its bottom, the incident radiance from the sky – enters the theory via the boundary conditions necessary for solution of the equations arising from the theory. Figure 1.8 summarizes the relationships between the various radiometric quantities, including IOPs and AOPs. Note that all of them have wavelength dependence.

A detailed description of the oceanic RTE equation and its link with some of the optical properties is given in the Appendix A. Nevertheless, the definition of the most important optical properties necessary to understand the results presented in the following chapters (and also shown in Fig. 1.8) is provided below.

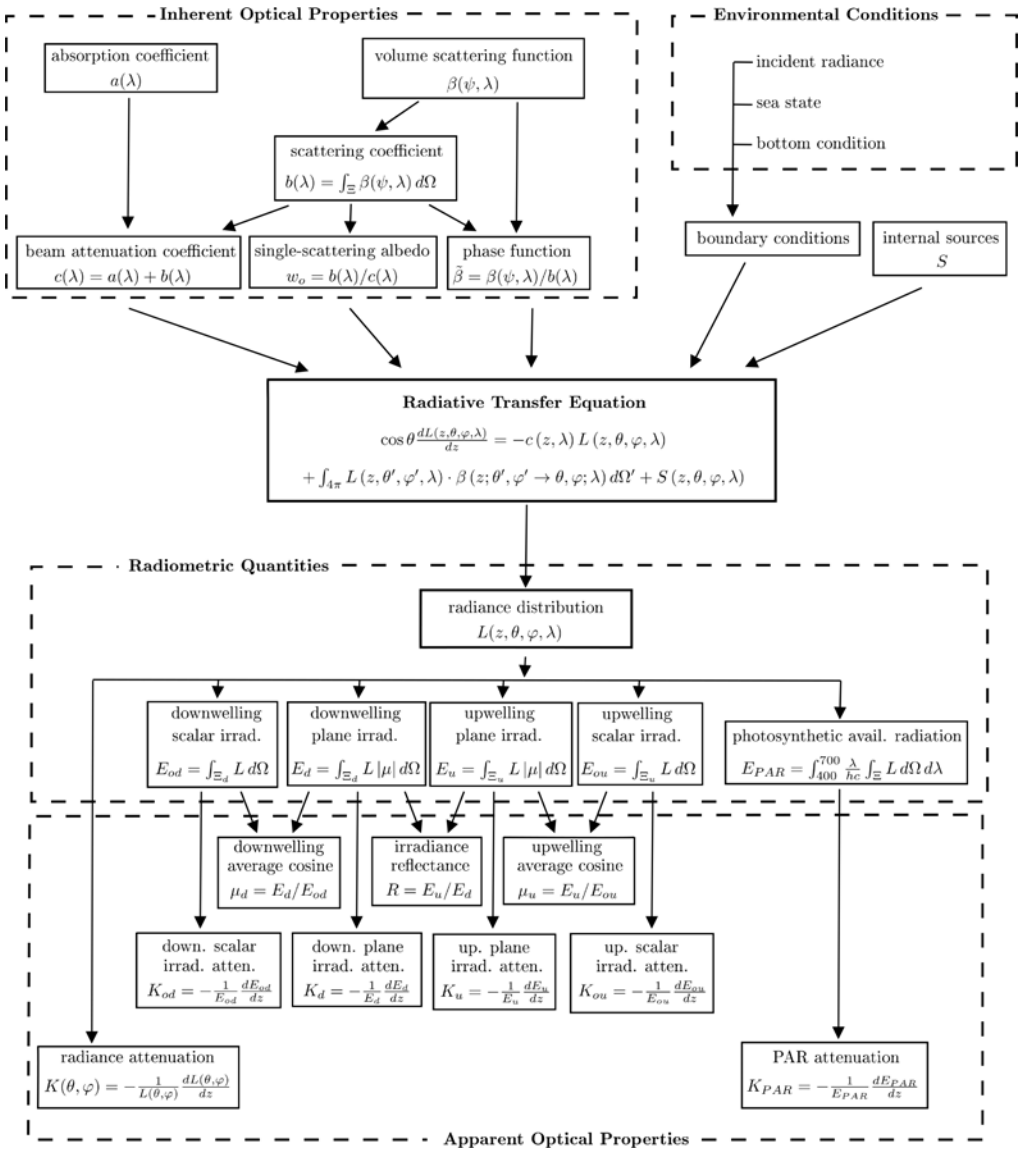


Figura 1.8: Relationships among the various optical quantities commonly used in hydrological optics, which are defined below (*Source*: Mobley, 1994).

Inherent Optical Properties (IOPs)

IOPs can be defined by using the diagram shown in Fig. 1.9, which illustrates a small volum ($\triangle V$) of water of thickness $\triangle r$, that is illuminated by a collimated beam of monochromatic light wavelength λ and spectral radiant power $\Phi_i(\lambda)\,[Wnm^{-1}]$. Part of the incoming beam of light is absorbed, $\Phi_a(\lambda)\,[Wnm^{-1}]$, and the other part is scattered, $\Phi_s(\lambda)\,[Wnm^{-1}]$, from the incident to other directions.

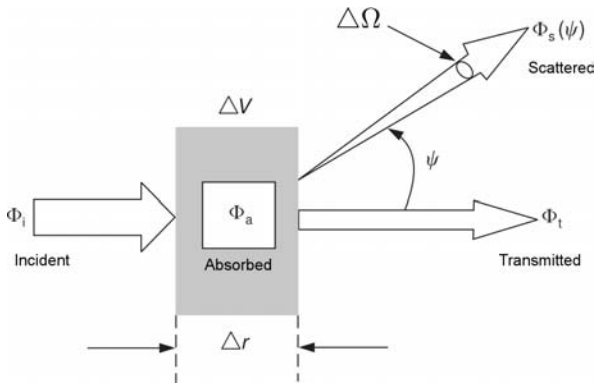


Figura 1.9: Geometry used to define the Inherent Optical Properties (IOPs) of a water body (*Source*: Mobley, 1994).

Absorption coefficient - $a(\lambda)$ is the absorbance (i.e., the ratio of the absorbed to incident spectral radiant flux) per unit distance:

$$a(\lambda) = lim_{\Delta r \to 0} \frac{1}{\Phi_i(\lambda)} \frac{\Phi_a(\lambda)}{\Delta r} \quad \left[m^{-1}\right] \tag{1.1}$$

Volume scattering function (VSF) - $\beta(\psi, \lambda)$ is the radiant flux per unit solid angle ($\Delta\Omega$) scattered in the ψ direction with respect to the direction of the incident light, per unit distance, expressed as a proportion of the incident flux:

$$\beta(\psi, \lambda) = lim_{\Delta r \to 0} lim_{\Delta\Omega \to 0} \frac{1}{\Phi_i(\lambda)} \frac{\Phi_s(\psi,\lambda)}{\Delta r \Delta\Omega} \quad \left[m^{-1} sr^{-1}\right] \tag{1.2}$$

Scattering coefficient - $b(\lambda)$ is obtained by integrating $\beta(\psi, \lambda)$ over all directions:

$$b(\lambda) = \int_{4\pi} \beta(\psi, \lambda)\, d\Omega = 2\pi \int_0^{\pi} \beta(\psi, \lambda) \sin\psi d\psi \quad \left[m^{-1}\right] \tag{1.3}$$

Beam attenuation coefficient - $c(\lambda)$ is the total loss of radiant energy from the original beam, which is the sum of absorption and scattering coefficients:

$$c(\lambda) = a(\lambda) + b(\lambda) \quad \left[m^{-1}\right] \tag{1.4}$$

Backscatter coefficient - $b_b(\lambda)$ is obtained when restricting the integration of $\beta(\psi, \lambda)$ to the interval $\pi/2 \leq \psi \leq \pi$:

$$b_b(\lambda) = 2\pi \int_{\pi/2}^{\pi} \beta(\psi, \lambda) \sin\psi d\psi \quad \left[m^{-1}\right] \tag{1.5}$$

Single-scattering albedo - $w_o(\lambda)$ is the ratio of scattering efficiency to total attenuation efficiency:

$$w_0(\lambda) = \frac{b(\lambda)}{c(\lambda)} \tag{1.6}$$

Volume scattering phase function - $\widetilde{\beta}(\psi,\lambda)$ provides the angular distribution of the scattered light and is obtained dividing $\beta(\psi,\lambda)$ by $b(\lambda)$**:**

$$\tilde{\beta}(\psi,\lambda) = \frac{\beta(\psi,\lambda)}{b(\lambda)} \quad \left[sr^{-1}\right] \tag{1.7}$$

It is important to note that in the preceding definitions, no inelastic scattering processes (e.g. fluorescence by phytoplankton or CDOM and Raman scattering by the water molecules) were considered. This factor appears as a source term ($S(z,\theta,\varphi,\lambda)$) in the radiative transfer equation (see Appendix A for further details). Moreover, due to the additive property, the total IOPs of a water body (i.e., $a(\lambda)$, $b(\lambda)$, $\beta(\psi,\lambda)$, $b_b(\lambda)$ and $c(\lambda)$) are the sums of the IOPs associated with each of the various components present in this water body. In this sense, for instance, $a_{total}(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{nap}(\lambda) + a_{cdom}(\lambda)$ where: w = pure water, ph = phytoplankton, nap = non-algal particles such as detritus or minerals and $cdom$ = colored dissolved organic matter (see the examples for absorption and scattering coefficients in Fig. 1.6).

Fundamental radiometric quantities in ocean optics

Spectral radiance - $L(z,\theta,\varphi,\lambda)$, as shown in Fig. 1.10, is the radiant energy incident in a time interval (Δt), onto a surface of area ΔA located at a position (x,y,z) and arriving through a set of directions contained in a solid angle ($\Delta\Omega$) about the direction (zenith angle θ and azimuthal angle φ) normal to the area ΔA and produced by photons in a wavelength interval $\Delta\lambda$:

$$L(x,y,z,\theta,\varphi,\lambda) = \frac{\Delta Q}{\Delta t \Delta A \Delta\Omega \Delta\lambda} \quad \left[Wm^{-2}sr^{-1}nm^{-1}\right] \tag{1.8}$$

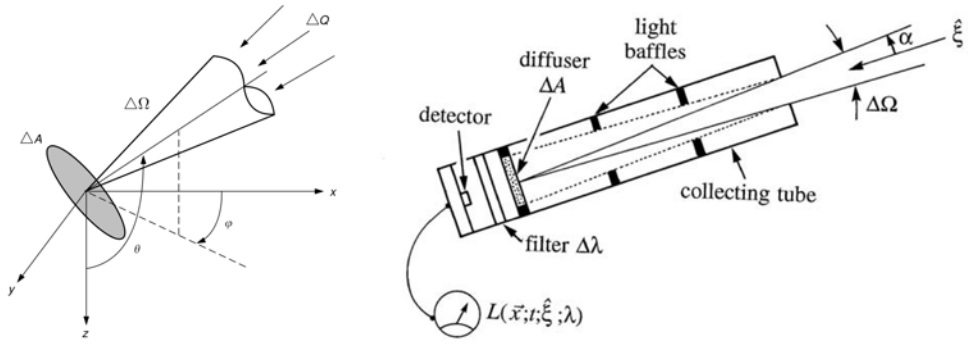


Figura 1.10: Geometries used to define the spectral radiance (*Source*: Mobley, 1994).

Scalar spectral irradiance - $E_o(z,\lambda)$ is obtained by integrating the radiance over all directions that reaches a certain surface. It is commonly measured using a spherical collector (see Fig. 1.11) and as the sum of the downwelling and upwelling components:

$$E_o(z,\lambda) = E_{ou}(z,\lambda) + E_{od}(z,\lambda) \quad \left[Wm^{-2}nm^{-1}\right]$$
$$E_o(z,\lambda) = \int_{\Xi_u} L(z,\theta,\varphi,\lambda)\, d\Omega + \int_{\Xi_d} L(z,\theta,\varphi,\lambda)\, d\Omega \tag{1.10}$$

Plane spectral irradiance - $E(z,\lambda)$ is obtained by using a plane collector (i.e., its sensitivity is proportional to $|\cos\theta|$, where θ is the angle between the photon direction and the normal to the surface of the detector, see Fig. 1.11) and as the sum of the downwelling and upwelling components:

$$E(z,\lambda) = E_u(z,\lambda) + E_d(z,\lambda) \quad \left[Wm^{-2}nm^{-1}\right]$$
$$E(z,\lambda) = \int_{\Xi_u} L(z,\theta,\varphi,\lambda)\,|\cos\theta| d\Omega + \int_{\Xi_d} L(z,\theta,\varphi,\lambda)\,|\cos\theta| d\Omega \tag{1.9}$$

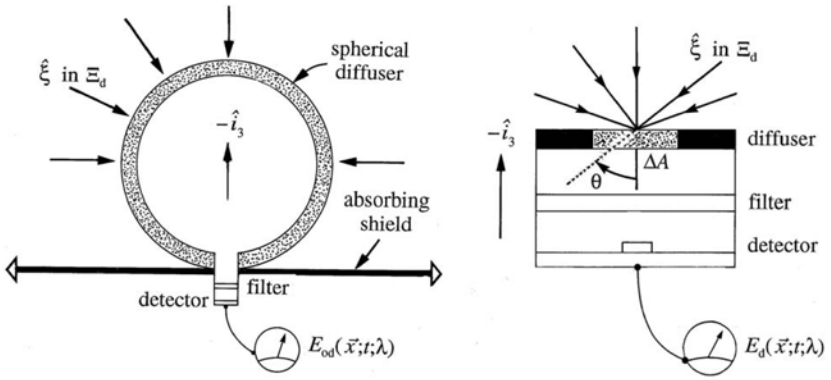


Figura 1.11: Geometries used to define the spectral scalar (left) and plane (right) irradiance. Note that Ξ_d denotes the hemisphere of downward directions, i.e., the set of directions (θ,φ) such that $0 \le \theta \le \pi/2$ and $0 \le \varphi \le 2\pi$, if θ is measured from the nadir direction (*Source*: Mobley, 1994).

The plane spectral irradiance, $E(z,\lambda)$, is the sum of the **plane spectral upwelling irradiance**, $E_u(z,\lambda)$, and the **plane spectral downwelling irradiance**, $E_d(z,\lambda)$, depending on the detector's orientation (in Fig. 1.11 the case for the downwelling irradiance is shown). The difference $E_d - E_u$ is called the net irradiance. On the other hand, the scalar spectral irradiance, $E_o(z,\lambda)$, is the sum of the **scalar spectral upwelling irradiance**, $E_{ou}(z,\lambda)$, and the **scalar spectral downwelling irradiance**, $E_{od}(z,\lambda)$. It is noted that in many oceanic environments horizontal variations of the IOPs and the radiance are much less than variation with depth. Therefore, it is assumed an horizontal homogeneity (i.e., these radiometric quantities vary only with depth, z). In addition, time-independent RT theory is adequate for most oceanographic studies due to the temporal variations of the IOPs or in the environment are much greater than the time required for the radiance to reach steady state after a change in IOPs or boundary conditions.

Another relevant measure of the light field in bio-optical oceanography is the **Photosynthetic Available Radiation (PAR)**. It is used in studies of phytoplankton biology to quantify the number of available photons, given an amount of radiant energy, that is relevant for photosynthesis to take place. PAR is defined by:

$$PAR(z) \equiv \int_{400nm}^{700nm} \frac{\lambda}{hc} E_o(z,\lambda)\, d\lambda \quad \left[fotons\, s^{-1} m^{-2}\right] \tag{1.11}$$

where $h = 6.6255 \cdot 10^{-34} J$ is the Planck constant and $c = 3.0 \cdot 10^{17} nm\, s^{-1}$ is the speed of light. The range approximately from 400 to 700 nm is used to calculate the PAR radiation because it provides the energy to marine phytoplankton to perform photosynthesis.

Radiative Transfer Equation (RTE)

The RTE equation connects the IOPs and the radiance, and for extension the rest of radiometric quantities and AOPs (see Fig. 1.8) and is defined as[2]:

$$\begin{aligned} \cos\theta \tfrac{dL(z,\theta,\varphi,\lambda)}{dz} &= -c(z,\lambda)\, L(z,\theta,\varphi,\lambda) \\ + \int_{4\pi} L(z,\theta',\varphi',\lambda) \cdot \beta(z;\theta',\varphi' \rightarrow \theta,\varphi;\lambda)\, d\Omega' &+ S(z,\theta,\varphi,\lambda) \end{aligned} \tag{1.12}$$

The source term, $S(z,\theta,\varphi,\lambda)$, can describe either an internal light source (i.e., bioluminiscence) or inelastically scattered light from other wavelengths. See Appendix A for further details on the RTE equation and these processes.

Apparent Optical Properties (AOPs)

Given the IOPs and boundary conditions of a water body (i.e., sea state, bottom type or incident radiance from the sky), the RTE is solved for the radiance distribution, which in turn, allows the computation of the AOPs. They are always obtained as a ratio of two radiometric variables.

Irradiance Reflectance - $R(z,\lambda)$ is the ratio of upward and downward irradiance:

$$R(z,\lambda) \equiv \frac{E_u(z,\lambda)}{E_d(z,\lambda)} \tag{1.13}$$

[2] The dependence on θ and φ is noted and combined by the symbol Ω.

Remote-sensing Reflectance - $R_{rs}(\theta, \varphi, \lambda)$ is defined as:

$$R_{rs}(\lambda) \equiv \frac{L_w(\lambda)}{E_d(\lambda)} \quad \left[sr^{-1}\right] \tag{1.14}$$

where $L_w(\lambda)$ is the water-leaving radiance or the upwelling radiance just above the sea surface (i.e., the total upward radiance minus the sky and solar radiance that was reflected upward by the sea surface) and $E_d(\lambda)$ the above water downward irradiance. $R_{rs}(\lambda)$ and $L_w(\lambda)$ are of particular interest because they can be retrieved with remote sensing techniques and represent what is generally called the **ocean color**.

Spectral diffuse attenuation coefficients - $K(z, \lambda)$ indicates the extinction of irradiance in the water column:

$$K(z, \lambda) = -\frac{1}{E(z,\lambda)} \frac{dE(z,\lambda)}{dz} \quad \left[m^{-1}\right] \tag{1.15}$$

Different diffuse attenuation coefficients can be defined (i.e., $K_d(z, \lambda), K_u(z, \lambda)$, $K_{od}(z, \lambda), K_{ou}(z, \lambda)$) using each corresponding irradiance (i.e., the plane irradiances: $E_d(z, \lambda), E_u(z, \lambda)$ or the scalar irradiances: $E_{od}(z, \lambda), E_{ou}(z, \lambda)$).

All the above described radiometric quantities and bio-optical properties permit the assessment of the biological state of the oceans and some of them have been used in this thesis with this purpose. As previously mentioned, however, in the case of AOPs, their interpretation in terms of water constituents and phytoplankton community composition becomes more complicated than in the case of IOPs since they are not additive. In fact, with the aim to better assess the ecosystem dynamics in marine waters from these bio-optical properties, several types of models have been reported over the last decades. On the one hand, the so-called "direct models" are used to predict AOPs given field measurements of IOPs and environmental conditions (see Section 1.4 and the results shown in the following chapters). On the other hand, "inverse models" are used to estimate IOPs given experimental measurements of AOPs (Lee, 2006). In addition, the additive property of IOPs permitted the development of differentiated models for absorption and scattering of each constituent of a water body.

1.2 From single to bulk particle measurements

Advances in core photonics and materials sciences as well embedded computing make a new realm of options available in applying optical technologies for identifying materials in the ocean (Moore *et al.*, 2009a; Zielinski *et al.*, 2009). Sensors incorporating optical techniques are filling increasingly significant roles in ocean monitoring and research. For instance, membranes and analyzers coupled with optical sensors have demonstrated to provide information on pH, nutrients, dissolved gases such oxygen, methane or carbon dioxide, and metal concentrations. In addition, optical techniques have demonstrated the ability to measure fundamental physical parameters such the

solar radiance distribution, absorption, scattering, stimulated fluorescence and flow cytometry, which play a key role in the identification of concentrations and dynamics of multiple phytoplankton and zooplankton species. In this sense, optical sensors have given researchers biological sampling capabilities on the same time and space scales as physical measurements.

With the aim to address the recent advances in monitoring of phytoplankton dynamics, this section reviews some of the existing and emerging optical-based technologies for in-water phytoplankton community composition assessment:

Single particle analysis

- **Microscopy** (including light and electron microscopy) is one of the earliest methods for identifying phytoplankton. Its ability to provide information on the phytoplankton composition up to the species level is second to none. However, it relies on the taxonomic skills of the observer and it is a very time consuming methodology, which makes it unsuitable for analysis of large numbers of samples. In addition, as a method entirely based on morphological characteristics, it can not identify some small picophytoplankton species that lack of distinct morphological features but have an essential role in the total marine primary production. In this sense, epifluorescence microscopy has been utilized to exploit the autofluorescence properties of chlorophyll and phycobiliproteins to differentiate different types of picophytoplankton species.

- **Flow cytometry** can resolve to a certain extend the limitations of microscopy (Sosik *et al.*, 2010). In this method, individual cells in liquid suspension (e.g., seawater for marine samples) are allowed to pass one by one through a focused light field (often a laser beam). Its fluorescence and light scattering properties are measured with detectors (usually photomultiplier tubes) as each cell passes this light field (see Fig. 1.12). Each phytoplankton group is identified given that it possesses different scattering properties (depending on the size, shape and refractive index of the cells) and fluorescence responses (associated with their photosynthetic pigments such as chlorophyll and phycobiliproteins). In general, flow cytometry measurements of phytoplankton permit enumeration, quantification of cell properties such as size and pigmentation, and some level of taxonomic or size-based discrimitation (e.g., microplankton diatoms, nanoplankton coccolithophores, picoplankton *Synechococcus* or *Prochlorococcus*). This technique has been used extensively by biological oceanographers to define the distributions and dynamics of marine phytoplankton.

  Despite the ability of flow cytometry to make rapid measurements of cells and to identify picoplankton, standard flow cytometers have a limited particle size range with an upper limit of only 15-20 μm. Moreover, since the accessory pigments carotenoids do not fluorescence directly, it makes difficult the identification of some phytoplankton groups. Nevertheless, the need of long-term, broad-space and higher-taxonomic measurement capabilities has motivated the

development of automated flow cytometers to be performed outside the laboratory. For instance, submersible instruments combined with cell-imaging capabilities have extended the use of flow cytometers to characterize and enumerate microplankton (Sieracki *et al.*, 1998; Olson and Sosik, 2007).

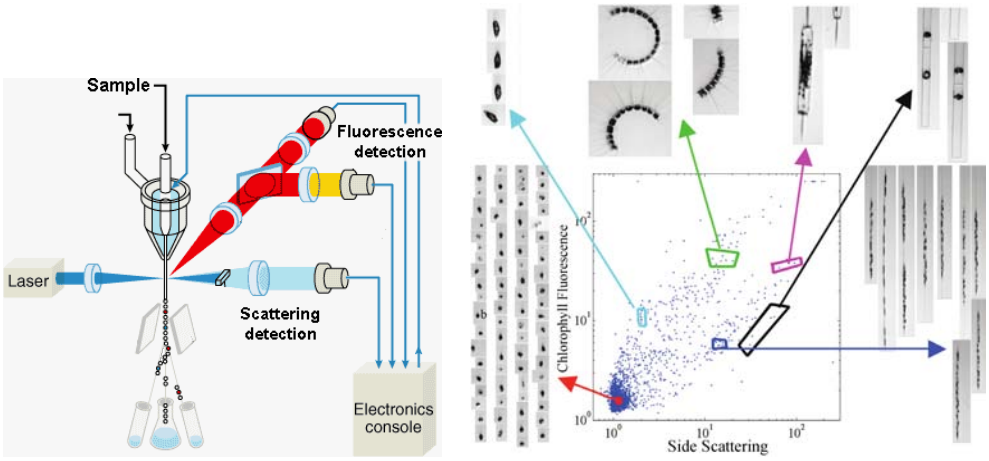


Figure 1.12: A typical flow cytometer configuration and an example of measurements of side scattering and chlorophyll fluorescence performed with the FlowCytobot instrument in 2004. Different regions in the scattering/fluorescence plot contain different species, which are identified by images of the corresponding phytoplankton cells collected by the imaging system (*Source*: Olson and Sosik, 2007).

Molecular analysis

- **High Performance Liquid Chromatography (HPLC)** facilitates the separation of phytoplankton communities on the basis of their pigment composition (Jeffrey and Vesk, 1997) and the estimation of the biomass of the different algal groups (see Fig. 1.13). Phytoplankton that cannot be separated by microscopic or flow cytometric techniques can be classified using this method. In HPLC, the sample to be separated is forced by a liquid at high pressure (the mobile phase) through a column that is packed with a stationary phase composed of irregularly or spherically shaped particles, a porous monolithic layer, or a porous membrane. The components of the sample move through the column at different velocities function of specific physical or chemical interactions with the stationary phase. The velocity of each component depends on the nature of each analyte, on the nature of the stationary phase and the composition of the mobile phase. The time at which a specific analyte emerges from the column is called the retention time and is provided by a detector. The retention time under particular conditions is considered an identifying characteristic of

a given analyte. The detector usually provides additional information related to the analyte such as its UV/VIS spectral signature, which helps to determine the phytoplankton community composition.

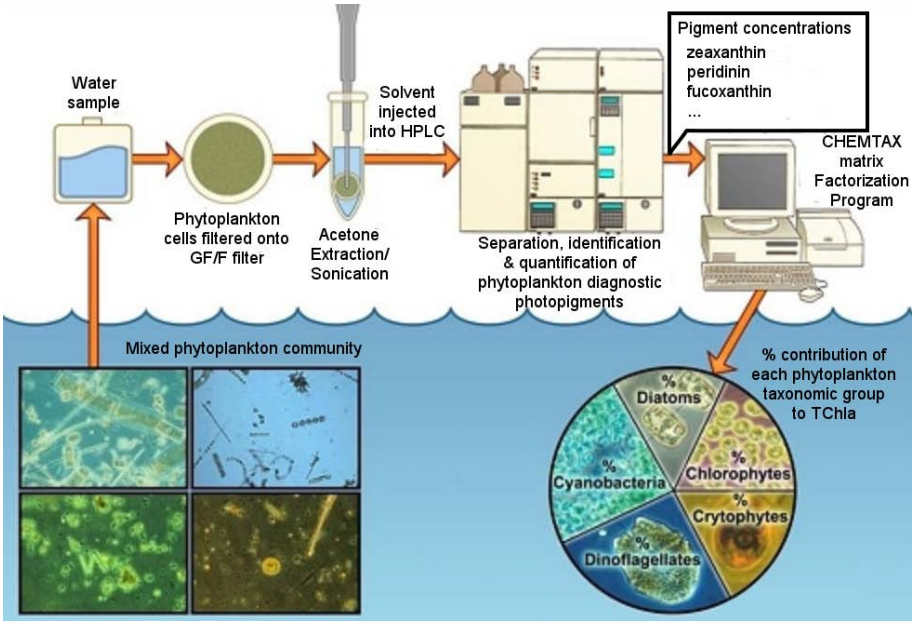


Figure 1.13: Diagram showing the methodology for determining algal community composition using diagnostic photopigment analyses by HPLC (*Source*: https://dcerp.rti.org).

Apart from chlorophyll-*a* which is common and present in all phytoplankton groups (in *Prochlorococcus* as divinyl chlorophyll-*a*), the inventory of all other pigments (i.e., chlorophylls-*b* and *c*, carotenoids and phycobiliproteins) varies in different taxa of phytoplankton. Many phytoplankton accessory pigments are taxonomically significant (see Table I.1). Therefore, detection of specific accessory pigments (i.e., marker pigments) in aquatic systems can reveal what types of phytoplankton are present there. The commercially available software CHEMTAX is a commonly used tool to reconstruct phytoplankton community composition from accessory pigment concentrations derived from HPLC analyses (Mackey *et al.*, 1996).

Pigment analyses are usually faster, require less taxonomic knowledge of the analyst and yield more precise measurements when compared to microscopic analyses. In part of the research presented in this thesis, HPLC pigment information has been used as the reference in terms of phytoplankton community composition (see Chapters 3 and 4). However, as an estimate of phytoplankton biomass and community composition, pigment analyses are not always necessarily accurate (see an example regarding this issue in Appendix B).

- **Other molecular techniques** that are beyond the scope of this thesis exploit genetic variations to distinguish between phytoplankton groups and provide a solution to the limitations encountered with HPLC. DNA sequencing techniques have opened avenues to distinguish organisms at all taxonomic levels, from the level of classes to ecotypes (Rocap *et al.*, 2003). Despite of the advances, probes are not available for all possible phytoplankton functional types (PFTs) and specificity of probes remains an area of ongoing research.

Bulk particle analysis

- **Secchi disc** and **Forel-Ule scale** are two simple and fast methods that were described at the end of the 19^{th} century to establish the ecological state of surface water (Wernand, 2011). The optical properties of seawater such as color and transparency in water have been measured for a long period: color through the Forel-Ule scale and transparency through the Secchi disc. The Forel-Ule Scale is a method to approximately determine the color of a water body by comparison with a standard color palette produced with mixtures of colored chemical solutions in a series of numerically designated vials (1-21). The Secchi disc is a circular white disk 30 cm in diameter, which is lowered over the side of a ship. The depth at which it is no longer visible is known as the Secchi depth and has been used as a rough measure of optical clarity or transparency. It is noted that these methods depend on the eyesight of the observer, thus, the major problem is the lack of absolute calibration. In addition, these methods are just helpful to classify gross phytoplanktonic activity of a water body.

- **Spectrometric techniques**

  Traditionally, as described in almost all previous methods, phytoplankton species discrimination has involved exhaustive and time consuming fieldwork, implying a ship-based approach to collect discrete water samples for later laboratory analysis to retrieve taxonomical information. These traditional methods often require filtering of discrete samples or even solvent extractions, which are time consuming and challenging to use when at sea. After the introduction of spectral radiometers to in situ and remote sensing oceanic observing platforms, the extrapolation of such information became easier, faster, much less invasive and usually non-destructive. In fact, by directly measuring the amount of light that is absorved, scattered or reflected from a bulk, which is a function of its pigment and chemical composition, one can possibly distinguish the phytoplankton composition of it. Scientists are now engaged in developing inversion methods to obtain biogeochemical products from these type of measurements. Next, a description of several in-water spectrometric-based technologies is provided:

  Measurement of Inherent Optical Properties

  Recent advances in optical instrumentation and methodologies now enable the in situ measurement of some IOPs (see Fig. 1.14). These instruments generally

employ an active light source given that IOP measurements are independent to the ambient radiance field. Therefore, they can be used at any time of day and are not subject to variable cloud or surface conditions and shadows from the observing platform or the instrument itself. The derivation of absolute coefficients of optical transfer properites is performed over relatively short path lengths and with a resolution of in-water variability at scales ranging down to a few centimeters.

Beam attenuation meters (also known as *transmissometer*s) operate upon the principle by which the attenuation (i.e., loss of light due to both absorption and scattering) of a narrow, well collimated beam of light through a path length of water is measured (Fig. 1.14a). Modern transmissometers for in situ ocean research operate at single or multiple wavelengths (Moore *et al.*, 2004a) and are used to estimate species composition, particulate organic carbon, suspended particles and visibility (Claustre *et al.*, 2008; Zaneveld and Pegau, 2003).

Absorption meters measure the light loss through a given path of water due only to the molecular absorption of the water and the components contained therein. The conceptual framework behind underwater absorption meters is based on minimizing measurement losses due to highly scattering media found in natural waters. In order to retrieve the absorption component, different designs have been developed: (1) Refractive tube absorption meters use a collimated beam propagating through a fixed path surrounded by a reflective tube and impinging upon a large area detector (Fig. 1.14b). Commercial units incorporating this design are available in multispectral and hyperspectral configurations (Zaneveld *et al.*, 2004). (2) The Point Source Integrating Cavity Absorption Meter (PSICAM) was designed to detect absorption spectra in an integrating cavity sphere filled with the water sample. The principle is based on multiple reflection and scattering, i.e. the light field is diffused inside the cavity (Röttgers *et al.*, 2005). In comparison to the refractive tube meters, it reduces the scattering effects on absorption measurements (i.e., the overestimation of the sample absorption resulting from backscattering). Submersible versions of this design have been developed by Dana and Maffione (2006) (Fig. 1.14c), likewise instruments built for operation on Autonomous Underwater Vehicles (Kirkpatrick *et al.*, 2006).

It is important to note that there exist similar approaches that measure the absorption coefficient by planktonic algae contained in a bulk but performed in the laboratory. These approaches deserve a mention since they also represent an essential source of information in terms of phytoplankton composition and have been used within the framework of this thesis (see Section 3.1.1.2 and 4.1.3). They are performed by combining light-transmission and light-reflection measurements of the algae previously concentrated onto a filter (i.e., the filter pad and T-R techniques, see Tassan and Ferrari, 1995) and using an integrating-sphere attached with a dual-beam spectrophotometer. In order to minimize the photon loss by backscattering, another solution has been developed in which

the sample is placed in the center of the integrating sphere (Simis *et al.*, 2005). In all these cases, as the filter pad technique is used to prepare the sample, it has to be considered that multiple scattering in the filter increases the optical path length. Therefore, a correction for the path length amplication factor has to be applied to convert the measured absorption of the algal particles on the filter to the absorption coefficient of the algal particles in suspension.

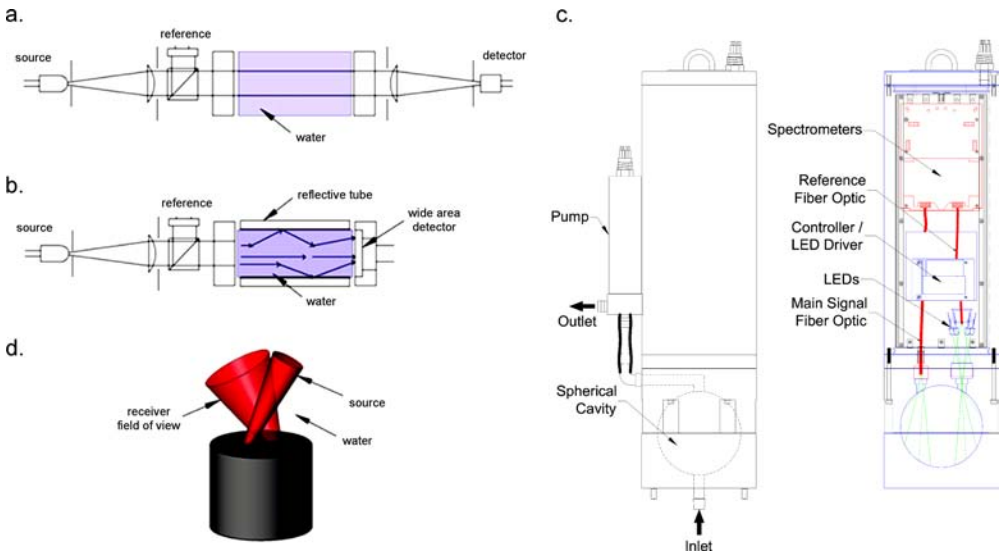


Figure 1.14: Schematic diagrams for IOP sensors. (a) Beam attenuation meter or transmissometer. (b) Reflective tube absorption meter. (c) Submersible spherical cavity absorption meter. (d) Backscattering sensor (*Source*: Moore *et al.*, 2009a; Dana and Maffione, 2006).

Backscattering meters are used to measure light scattered in the backward direction (i.e., between 90 and 180 degrees). The backscattering coefficient is important in understanding remote sensing signals seen for instance by satellites. These sensors are most widely used in applications focused upon validating ocean color measurements from satellites or also to retrieve information regarding the particle size distribution of a bulk, including phytoplankton. Optical backscattering sensors have been configured for measurement typically between 115 and 145 degrees (Fig. 1.14d) and multispectral configurations (Lee and Lewis, 2003; Moore *et al.*, 2000).

Fluorescence sensors are used to estimate concentrations and provide invaluable physiological information from phytoplankton materials. Due to its cellular pigment composition, phytoplankton species are characterized by specific excitation wavelengths triggering specific emission wavelengths. Fluorometers have been configured for in situ single channel detection, which provide reasonable proxies for phytoplankton biomass. Moreover, multichannel fluorometers have

been developed to detect spectral excitation and/or emission fluorescence signatures, which can be used to infer taxonomic composition and identify specific phytoplankton types (Moore *et al.*, 2004b; Chekalyuk *et al.*, 2006).

Measurement of Apparent Optical Properties

Sensors measuring AOPs are in general passive since these properties vary as the angular radiance distribution varies. It implies that they use as their light source the incident solar beam directly transmitted and scattered by the atmosphere, the sea surface and the ocean interior. In this sense, passive detection of sunlight results in relatively low energy consumption and implies that AOP sensors measure light levels experienced by the surrounding biota.

Radiance and irradiance sensors are used to determine and derive all the AOPs. They can be differentiated by the angular integration performed by the front-end optics, which is the component that captures the ambient light field. Narrow angular field-of-view (FOV) collectors are used to measure radiance, while plane collectors provide cosine weighting to yield plane irradiance, and spherical collectors, which weight all directions equally, measure scalar irradiance (see diagrams in Fig. 1.10 and 1.11). These sensors have been manufactured in multispectral and hyperspectral configurations to provide full spectral distribution over the ultraviolet, visible and near infrared bands. The spectral dispersion to separate the broadband field into narrow spectral intervals is accomplished by prism or grating elements. For some applications, spectral weighting filters have also been employed to return the integral energy over the entire visible spectrum and are known as irradiance PAR sensors.

Radiance and AOP sensors are now operationally for applications including the determination of phytoplankton biomass. For instance, it is estimated via the measurement of the in-water irradiance, the water leaving radiance or the reflectance at the sea surface. These signals are influenced by the absorption and scattering properties of the water column and hence, by the red fluorescence (approx. at 680 nm) resulting from phytoplankton activity. It is also under study the identification of different phytoplankton functional groups (PFTs) from the in-water and remote-sensing radiance characteristics and related AOPs (Nair *et al.*, 2008). These instruments are also used routinely for ocean color applications such remote sensing validation and calibration and for capturing optical changes associated with some phytoplankton-related episodic events (e.g., harmful algal blooms). As it is emphasized below, the increasing use of high spectral resolution (hyperspectral) sensors is being crucial in all these applications, in particular for phytoplankton species monitoring (see Section 1.3).

It is important to note that protocols for effective use of instruments measuring IOPs and AOPs have been developed during the last years (see Mueller *et al.*, 2003b; Mueller *et al.*, 2003a and references found therein). For instance, there exist international standards of radiance and irradiance measurements for ocean

color applications, which provide a reference to which instruments anywhere in the world (i.e., including onboard space-based platforms) and at any time can be calibrated and intercompared with a high degree of confidence.

- **Other approaches** that are beyond the scope of this thesis include the use of holographic instruments and confocal technologies for the 3D in-situ visualization of plankton (Malkiel *et al.*, 1999; Matsumoto, 1993).

1.3 From multi- to hyperspectral applications

It has been described how phytoplankton community composition in the ocean may be assessed using a wide range of optical-based techniques in the laboratory and in the field. Due to the faster and less invassive characteristics, however, a noticeable increase has been in the number of coastal and open ocean studies based on spectrometric measurements.

In situ and remotely-sensed spectrometric observations of ocean waters (i.e., IOPs and AOPs) provide information regarding the concentrations of optically significant constituents in seawater, and offer the ability to observe important biological and biogeochemical variables (e.g., Chang *et al.*, 2006). Numerous studies over the past three decades have focused on the development of bio-optical algorithms linking measurable spectral optical properties to the primary pigment in phytoplankton, chlorophyll-*a*, a proxy for the phytoplankton biomass and primary production (e.g., Bricaud *et al.*, 1998; Morel, 1988; O'Reilly *et al.*, 2000 and the illustrating example shown in Figure 1.15 from Reynolds *et al.*, 2001). Current global algorithms use simple wavelength-ratios of reflectances to retrieve chlorophyll concentrations, such as the NASA OC2 version 2 algorithm, also depicted in Fig. 1.15 and the NASA OC4 version 4 algorithm, both developed by O'Reilly *et al.* (2000). In particular, the OC4 algorithm is a four-band maximum band ratio formulation that uses a fourth order polynomial to estimate chlorophyll concentration and is given by the relation:

$$[Chl] = 10^{(0.366-3.067X+1.93X^2+0.649X^3-1.532X^4)} \tag{1.16}$$

where $X = log(R_{rs}(443)/R_{rs}(555) > R_{rs}(490)/R_{rs}(555) > R_{rs}(510)/R_{rs}(555))$.

More recently, efforts to expand the use of spectrometric measurements for estimating other biogeochemically important ocean variables and phytoplankton-related phenomena have increased considerably. For example, spectrometric measurements including satellite remote sensing have been used to detect harmful algal blooms (Cullen *et al.*, 1997; Stumpf *et al.*, 2003), surface concentrations of particulate inorganic and organic carbon (Balch *et al.*, 2005; Stramski *et al.*, 2008), particle size distribution (Kostadinov *et al.*, 2009), phytoplankton community composition and size structure (Aiken *et al.*, 2007, 2005; Ciotti and Bricaud, 2006; Nair *et al.*, 2008; Uitz *et al.*, 2006), and class-specific primary production (Uitz *et al.*, 2010).

It is noteworthy, however, and in accordance with the central topic of this thesis, that recent advances in measuring ocean spectrometric properties and light fields

within and leaving the ocean have included a progressive shift from using multispectral to high spectral resolution (hyperspectral) acquisition systems (Chang *et al.*, 2004). The capability to obtain measurements at hundreds of narrow and closely spaced wavelength bands from the ultraviolet to near-infrared, with a resolution better than 10 nm, has become one of the most powerful and fastest growing areas of technology in the field of ocean optics. The rapid maturing of optical instrumentation has led to hyperspectral technology, which has opened the possibility for optical oceanographers to more accurately characterize complex oceanic environments.

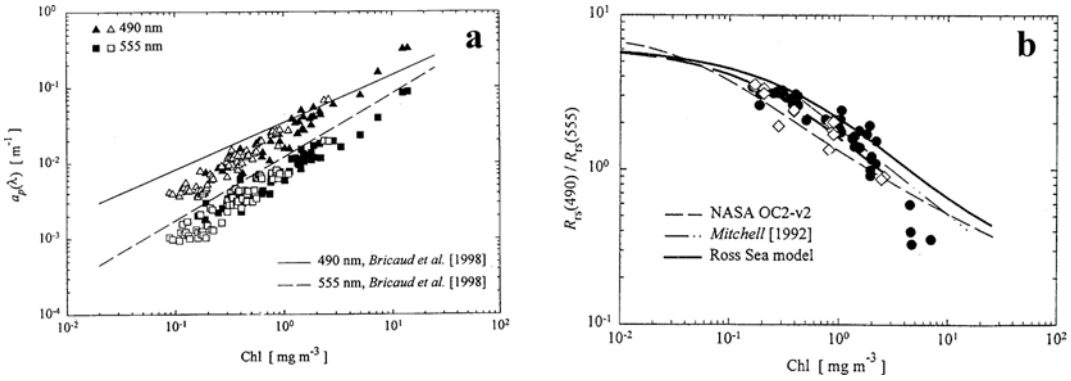


Figure 1.15: (a) Particulate absorption coefficients at particular wavelengths (i.e., 490 and 555 nm) as a function of chlorophyll concentration for different stations within the Southern Ocean. Solid and open symbols represent surface data from the Ross Sea and Antarctic Polar Front Zone, respectively. The lines illustrate the predicted relationship from the model developed by Bricaud *et al.* (1998). (b) Modeled relationships between the spectral band ratio of remote-sensing reflectance, $R_{rs}(490)/R_{rs}(555)$, and chlorophyll obtained in the same locations shown in (a). The lines inllustrate the predicted relationship from different models, among them the NASA OC2 version 2 algorithm developed by O'Reilly *et al.* (2000) (*Source*: Reynolds *et al.*, 2001).

The advantage of a hyperspectral over a multispectral inversion for phytoplankton species identification stems from the fact that more accurate spectral information, such as spectral features related to characteristic phytoplankton pigment absorption peaks, is resolved. This is a key factor in the identification of phytoplankton types because it relies on very small signal and changes in the shape of spectral optical characteristics. In the past, analyses of single band or single band-ratios obtained in discrete multispectral bands were employed to retrieve concentrations of the pigment chlorophyll-*a* (O'Reilly *et al.*, 2000, see Fig. 1.15), resolve the presence of specific phytoplankton functional groups (Sathyendranath *et al.*, 2004; Alvain *et al.*, 2005) or determine phytoplankton size classes (Hirata *et al.*, 2008). However, in most studies phytoplankton species mapping failed or was limited to the identification of some

phytoplankton groups given the limited number of spectral bands. The multispectral-based approach usually implied a lack of spectral resolution necessary to differentiate more species in the ocean. In this sense, the advent of hyperspectral sensors provides ample spectral information and facilitates to go beyond the estimation of the chlorophyll-*a* and identify more phytoplankton species, which are spectrally unique and complex. For instance, Craig *et al.* (2006) assessed the feasibility of remote detection and monitoring of the toxic dinoflagellate, *Karenia brevis*, accounting for complete spectral behavior of in situ hyperspectral measurements of remote-sensing reflectance, $R_{rs}(\lambda)$ and Lohrenz *et al.* (2008) directly explored the hyperspectral patterns of $R_{rs}(\lambda)$ to better characterize water mass properties in coastal areas (i.e., Case 2 waters). More recently, the composition and concentrations of two major phytoplankton functional groups (i.e., diatoms and cyanobacteria) have been derived from hyperspectral measurements of the satellite sensor SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Chartography) on ENVISAT and analysed with the PhytoDOAS method (Bracher *et al.*, 2009; Sadeghi *et al.*, 2012).

As for what this thesis concerns, all hyperspectral-based applications developed up to now and the increasing availability of hyperspectral observations of ocean waters clearly lead to a need for ongoing exploration of its potential to better map phytoplankton community composition and dynamics.

Expanding role of hyperspectral oceanography

New technologies and the miniaturization of electro-optical components have permitted the development of accurate, low-cost and energy-efficient hyperspectral sensors designed to measure oceanic high spectral resolution optical properties (i.e., including IOPs and AOPs). As a result, their use is no longer limited to bench-top applications but they are being deployed on fully fledged sea-going observational platforms. New interdisciplinary ocean observing platforms incorporating hyperspectral sensors are being deployed such as in-water vertical profiling systems, moorings, drifters, autonomous vehicles, air-borne and space-borne platforms (Dickey *et al.*, 2006; Twardowski *et al.*, 2005; Perry and Rudnick, 2003). Various satellite-based hyperspectral imagers are currently available: NASA's Hyperion sensor on the EO-1 satellite (Fig. 1.16a), ESA's CHRIS sensor on PROBA satellite or US Air Force Research Lab's FTHSI sensor on the MightySat II satellite. In addition, examples of existing hyperspectral airborne sensors providing high spatial resolution images of specific areas are the Ocean Portable Hyperspectral Imager for Low-Light Spectroscopy (PHILLS, Davis *et al.*, 2002), the Compact Airborne Spectrographic Imager (CASI) or the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). Other in situ ocean observing platforms for which hyperspectral sensors may be suitable are ships and moored buoys (Fig. 1.16b, Kuwahara *et al.*, 2007). These platforms offer continuous measurements at a high spatial and temporal resolution delivering data even under cloudy conditions, thereby complementing discrete observations by satellites and aircrafts and providing the data essential for calibration and validation purposes. Furthermore, Lagrangian platforms that follow a particular water mass (e.g. floats and drifters)

can provide hyperspectral measurements in oceanic regions not normally accessible by satellites or oceanographic research vessels. The goal of these monitoring missions is to determine the material composition of the water mass under investigation through the accurate analysis of simultaneous optical and hydrographic measurements.

The number of newly emerging ocean observing platforms capable of incorporating hyperspectral instrumentation is continuously growing. For instance, new autonomous underwater vehicles (AUVs) have been developed recently with integrated equipment for adaptive sampling and the ability to perform a wide-range of preprogrammed monitoring surveys over extended periods of time (Perry and Rudnick, 2003). Another member of this family of instruments are the special AUVs called gliders that propel themselves through the water by changing their buoyancy with a minimal expenditure of energy (Fig. 1.16c). The glider provides a wide spatial coverage measuring – among others - the optical properties during periods of three or four weeks (Wood, 2009). Another approach to monitor the variability of optical properties in the ocean is the use of vertical profiling systems (Fig. 1.16e, f) which can, for instance, be part of some of the emerging cabled observatories. Instrument platforms such as the Vertical Profiler System (VPS), part of Neptune Canada (Barnes *et al.*, 2008), the world's first regional-scale underwater ocean observatory plugged directly into the Internet, offer great power and bandwidth for gathering large quantities of hyperspectral data.

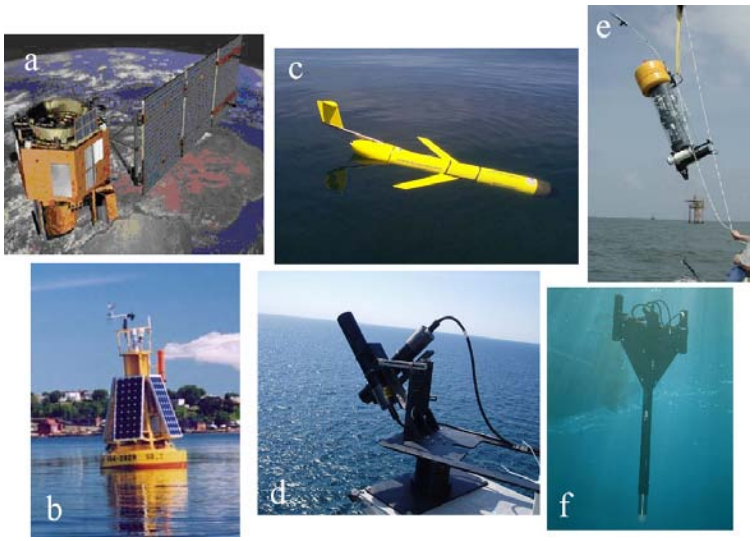


Figure 1.16: Examples of ocean observing platforms suitable to incorporate hyperspectral sensors.

The amount of hyperspectral data sets available collected using all these types of oceanographic observing platforms, covering a wide range of temporal and spatial scales, will increase in the near future. New interdisciplinary research initiatives are being successfully carried out such as the Hyperspectral Coastal Ocean Dynamics

Experiment (HYCODE, see Pegau, 2002) conceived to exploit the new capabilities of hyperspectral ocean color sensors. Within this framework, several collaborative short-term and long-term field experiments, involving the simultaneous use of hyperspectral devices deployed in situ and remotely, are being conducted for calibrating, "sea truthing" and relating subsurface optical properties to remote sensing ocean color measurements. These investigations are essential to develop and validate optical radiative models (see next Section) and to further our understanding of the processes that contribute to the temporal and spatial variability of IOPs and AOPs in the ocean. One of the most important challenges of HyCODE experiment is the standardization of the quality control protocols, including the definition of procedures to perform detailed analysis of the reliability of measurements and an evaluation of the impact of instrumentations inaccuracies.

1.4 Radiative Transfer (RT) models

In situ and remote sensing hyperspectral measurements in the ocean might be the two methods providing sufficient information on phytoplankton distribution in their spatial and temporal variability. However, collecting significant amount of hyperspectral data from such a high dynamic environment and under fully-controlled conditions is still not straightforward. In fact, there are currently no hyperspectral sensors that provide validated observations with a high temporal and spatial coverage of extended marine areas. In this sense, there are still several limitations in the applications where multispectral and hyperspectral data is used. For instance, despite current global remote sensing products (that rely on algorithms based on spectral ratios of the remote-sensing reflectance for the operational determination of chlorophyll-a, O'Reilly *et al.*, 2000) benefit from reduction in uncertainties partially cancelled out through ratios, residual uncertainties may still become the source of large uncertainties in derived products. These uncertainties are related to error estimates for atmospheric correction or to errors in the performance of the sensors themselves and are a limiting factor for the successfull development of the satellite-derived products.

One approach to finding a solution to these problems and exploring the feasibility of hyperspectral oceanographic observations to detect phytoplankton functional and toxic groups is using radiative transfer (RT) modelling. On the one hand, model-based approaches may be useful to show the feasibility of different optical measurements and their results for drawing the attention to the possibilities of how the problems that might arise could be solved. On the other hand, and as it is shown in this thesis, optical modelling can be an optimal and essential tool for developing and testing new methods for estimating variability and dynamics of phytoplankton communities. In any case, it is noteworthy that models are only useful for these applications when we have good knowledge on the phytoplankton properties and other marine components. The use of in-situ measurements of in-water properties, together with theoretical RT models and laboratory experiments is the most suitable combination to validate in situ and remote sensing observations and investigate biological activity, marine optical

properties, and changes in the concentration and composition of material in surface waters, over larger temporal and spatial scales (Tzortziou *et al.*, 2006).

In this thesis, the commercially available RT numerical model Hydrolight/Ecolight 5.0 (Sequoia Scientific, Inc., Mobley and Sundman, 2008) has been utilized. Time-independent radiance distributions within and leaving any plane-parallel water body (and derived quantities) have been computed as a function of depth, direction and wavelength given water column absorbing and scattering properties (i.e., the IOPs) and other oceanographic environmental conditions (see Fig. 1.8 for relationships between the various radiometric quantities). The Hydrolight/Ecolight code employs mathematically sophisticated invariant imbedding techniques to solve the radiative transfer equation (see Appendix A for details on this equation) and offers the possibility of performing numerical simulations in controlled environments (e.g., Mobley and Stramski, 1997 and Albert and Mobley, 2003). When computing the full radiance distribution, invariant imbedding is computationally extremely fast compared to other solution methods such as Monte Carlo simulations.

Hydrolight/Ecolight 5.0 has been used in a variety of studies ranging from bio-optical oceanography to remote sensing. In this thesis, the IOPs of different water types have been used to simulate different in-water light scenarios (each characterized by an hyperspectral optical signature) with the purpose of evaluating the feasibility of hyperspectral oceanographic observations to better map phytoplankton communities. As it is described in each chapter, modeled, laboratory and in situ data were collected and considered in the simulations performed for each of the presented studies.

2 Model-based assessment and novel approach to improve the use of hyperspectral data in monitoring of marine phytoplankton communities

The results from this chapter have mostly been published as:

Torrecilla, E., J. Piera and M. Vilaseca (2009). Derivative analysis of oceanographic hyperspectral data. In G. Jedlovec (Ed.), Advances in Geoscience and Remote Sensing (pp. 597–619). Vienna: InTech.

The increasing availability of in situ and remotely sensed hyperspectral measurements of ocean waters leads to a need for ongoing evaluation of high-resolution processing methods. In this sense, the main objective of this thesis is to attempt a verification of the hypothesis formulated: "Hyperspectral oceanographic observations play a key role in a better characterization of complex oceanic environments and mapping of phytoplankton communities", by developing new processing strategies.

Hyperspectral optical data provide the opportunity for improvements in spectral shape analysis and subsequent extraction of environmental information compared with low spectral resolution optical data (multispectral approaches). Derivative spectroscopy is a powerful technique of spectral shape analysis which enhances subtle features in hyperspectral data, and has been successfully used to obtain information about optically significant water constituents such as phytoplankton. For example, Craig *et al.* (2006) assessed the feasibility of detection of a toxic algal bloom of the dinoflagellate *Karenia brevis* from the analysis of the fourth derivative of phytoplankton absorption spectra, estimated from in situ hyperspectral measurements of remote-sensing reflectance, $R_{rs}(\lambda)$. The advantages offered by hyperspectral measurements of $R_{rs}(\lambda)$ in combination with derivative spectroscopy for identifying algal blooms were also demonstrated by Lubac *et al.* (2008), who based their analysis on the position of the maxima and minima of the second derivative of the spectral $R_{rs}(\lambda)$. Louchard *et al.* (2002) assessed major sediment pigments of benthic substrates from derivative spectra of hyperspectral $R_{rs}(\lambda)$ measured in shallow marine environments. In general, the optical detection of specific algal blooms appears feasible because certain accessory pigments with specific absorption features (see Fig. 1.3) are unique to individual phytoplankton taxa (e.g., Millie *et al.*, 1995) and can

be better differentiated in hyperspectral absorption data than in multispectral data with a limited number of spectral bands corresponding to certain wavelengths.

The advantages of high spectral resolution measurements suggest that the effectiveness of derivative spectroscopy applied to hyperspectral optical information for assessing phytoplankton diversity should be further explored. In particular, there is a need to test whether the hyperspectral approach, which has proven useful in inland and coastal waters (e.g., Hunter *et al.*, 2008; Lee and Carder, 2004; Lubac *et al.*, 2008), can also be effective for the identification of different phytoplankton assemblages at large spatial scales in the open ocean (i.e., Case 1 waters). These tests are also especially important for the common situation in which various phytoplankton groups co-exist at significant concentrations (i.e., a non-bloom condition). In such a case, as no single species dominates the assemblage, it seems also appropriate to consider some classification technique in combination with derivative spectroscopy to identify different phytoplankton assemblages, which accounts for complete spectral behavior of hyperspectral data (Duin *et al.*, 1997; Pekalska and Duin, 2000).

In this chapter, a new methodology applied to hyperspectral data is proposed (see diagram in Figure 2.1) with the aim to automatically identify different phytoplankton communities in the open ocean (i.e., non-bloom conditions). A dissimilarity-based cluster approach is evaluated and applied to a hyperspectral data set of $R_{rs}(\lambda)$, including ordinary spectra and its second derivative spectra, in which the global contribution over the entire spectral range of all pigments present in the water sample is considered.

In an ideal case, the effectiveness of any new method for estimating variability and dynamics of phytoplankton communities should be tested on the basis of field data. However, collecting significant amount of hyperspectral data from different environments under fully-controlled conditions is not straightforward. Moreover, the availability of satellite or in situ hyperspectral data sets for developing and validating new algorithms was still complicated at the beginning of this thesis. In fact, there are currently no hyperspectral sensors that provide validated observations with a high temporal and spatial coverage of extended marine areas.

The above mentioned reasons indicated the use of oceanic radiative transfer (RT) modeling as a primary approach in this thesis. In this chapter, a synthetic data set of hyperspectral $R_{rs}(\lambda)$ spectra has been generated and analyzed. As a first approach, several oceanic optical scenarios have been created using a simulation-based framework, to represent simplified environmental conditions in terms of phytoplankton composition. The main goal is to provide a first demonstration of the feasibility of hyperspectral oceanographic data to discriminate different phytoplankton communities in the open ocean by using the proposed methodology. In order to address this question, derivative spectroscopy and a cluster technique (HCA) are applied to the synthetic hyperspectral $R_{rs}(\lambda)$ spectra obtained with the commercially available RT model Hydrolight/Ecolight 5.0 (Mobley and Sundman, 2008).

The present chapter is organized in the following parts. In Section 2.1 a description is offered regarding some important issues that must be considered when derivative spectroscopy is applied to hyperspectral oceanographic data. In addition, the funda-

mentals of the applied HCA cluster-based approach are provided. In Section 2.2, some results demonstrating the feasibility of applying derivative spectroscopy and HCA to modeled hyperspectral $R_{rs}(\lambda)$ in open ocean environments are shown. Special attention is given in this section to the role of involved parameters during derivative computations of hyperspectral data, in order to better identify phytoplankton communities. Section 2.3 is devoted to summarize the main conclusions drawn from this first study.

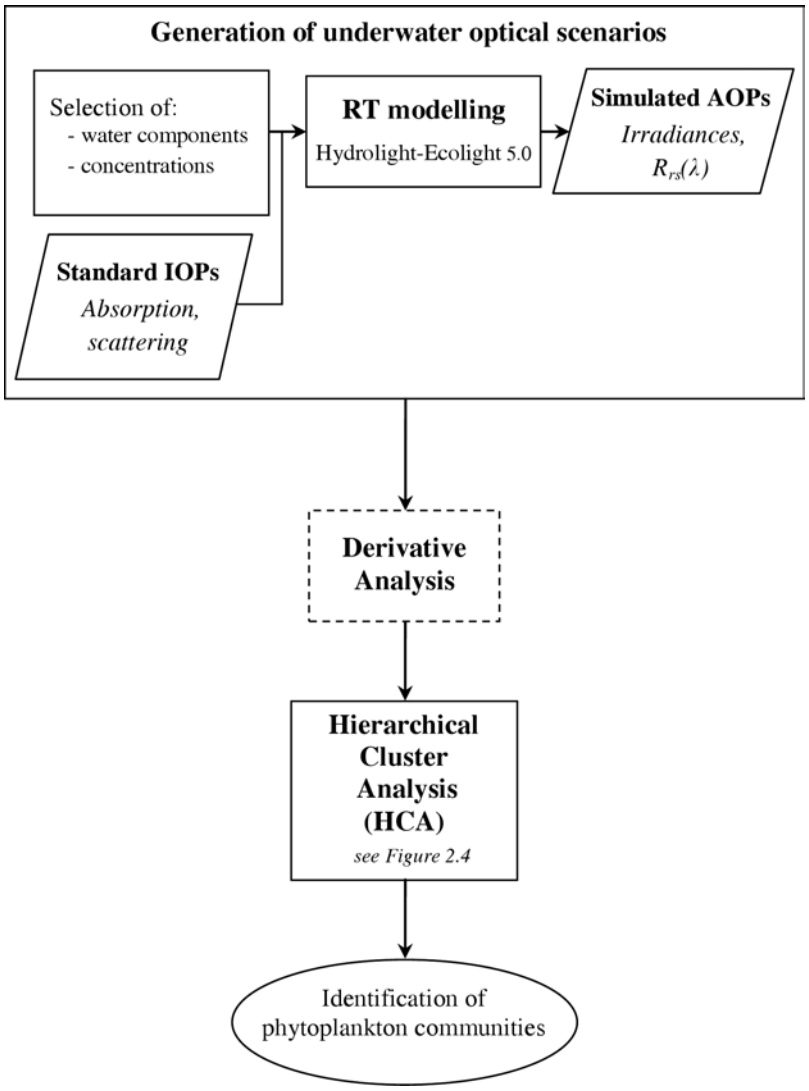


Figure 2.1: Flowchart showing the methodology proposed in this thesis. The automatic identification of phytoplankton communities is tested in this chapter by considering a synthetic data set of hyperspectral $R_{rs}(\lambda)$ spectra.

2.1 Hyperspectral data analysis

2.1.1 Derivative analysis

Derivative spectroscopy has been commonly used in the analysis of hyperspectral data using different computation algorithms such as finite divided differences or least square fitting (Tsai and Philpot, 1998; Ruffin *et al.*, 2008). In this thesis, the process of estimating derivative spectra has been addressed using a finite divided difference algorithm, named "finite approximation". As a less computationally expensive approach, it consists in computing the changes in curvature of a given spectrum ($s(\lambda)$) over a sampling interval ($\Delta\lambda$) or band separation (BS). In the discrete domain, this sampling interval is finite and small but not necessarily infinitesimally small and is defined as $\Delta\lambda = \lambda_j - \lambda_i$, where $\lambda_j > \lambda_i$. Therefore, the first and the subsequent nth derivative are obtained using Eqs. 1 and 2, respectively:

$$\left.\frac{ds}{d\lambda}\right|_i \approx \frac{s(\lambda_i) - s(\lambda_j)}{\Delta\lambda} \tag{2.1}$$

$$\left.\frac{d^n s}{d\lambda^n}\right|_j \approx \frac{d}{d\lambda}\left(\frac{d^{(n-1)}s}{d\lambda^{(n-1)}}\right) \tag{2.2}$$

Derivative analysis can be applied to hyperspectral measurements of both inherent and apparent oceanographic optical properties (e.g. spectral absorption coefficients, irradiance distributions or remote-sensing reflectances). For instance, as shown in Figure 2.2, it is a useful tool for enhancing spectral features, which are often relevant and related to absorption bands of pigments present in the considered water samples. In this sense, successful extraction of spectral details of interest through derivative spectroscopy depends on the band separation chosen. The importance of selecting a suitable band separation ($\Delta\lambda = BS$) in each case stems from the fact that spectral data features of interest with a smaller scale than the BS will not be preserved in the derivative results. In order to visualise this effect, Fig. 2.3 shows different derivative spectra corresponding to one of the spectral absorption coefficient shown in Fig. 2.2a, computed according to different finite band resolutions (BS).

Another important factor to consider is that noise level in hyperspectral data can be considerable, as the small amount of energy gathered by the narrow bandwidth may be exceeded by the intrinsic sensor's noise. In order to make an optimal application of derivative analysis, which is a technique clearly sensitive to noise, smoothing techniques must be applied to hyperspectral data prior to computation of derivative spectra (Vaiphasa, 2006). A number of smoothing algorithms have been developed within the last few decades (e.g. Savitzky-Golay, Kawata-Minami or mean-filter smoothing). In all approaches, the smoothing level applied depends on the size of the filter window (WS) used for averaging. It is therefore worth noting that an appropriate selection of the smoothing and derivative parameters (WS and BS) must be done according to the resolution of each type of hyperspectral data (Torrecilla *et al.*, 2007). An important effort must be made to determine the best trade-off between denoising and the ability to resolve fine spectral details of interest.

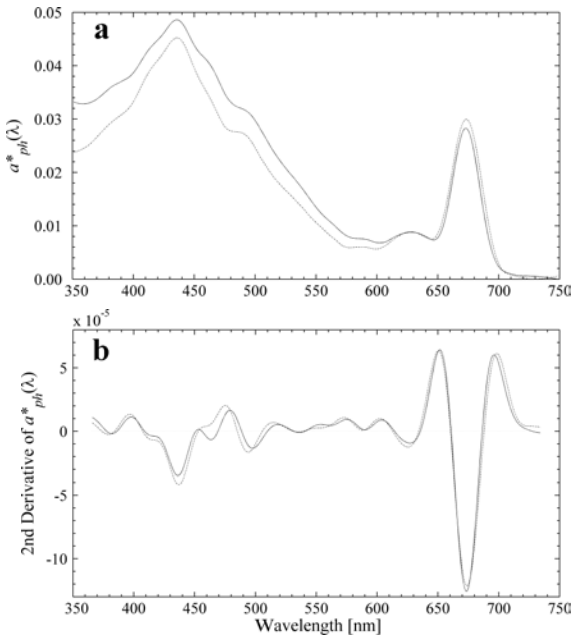


Figure 2.2: (a) Examples of specific absorption coefficient, $a^*_{ph}(\lambda)$, corresponding to two phytoplankton groups (from Kim and Philpot, 2006). (b) Results from second derivative analysis applied to the $a^*_{ph}(\lambda)$ spectra. Shape singularities in each $a^*_{ph}(\lambda)$ spectra are enhanced in the derivative domain.

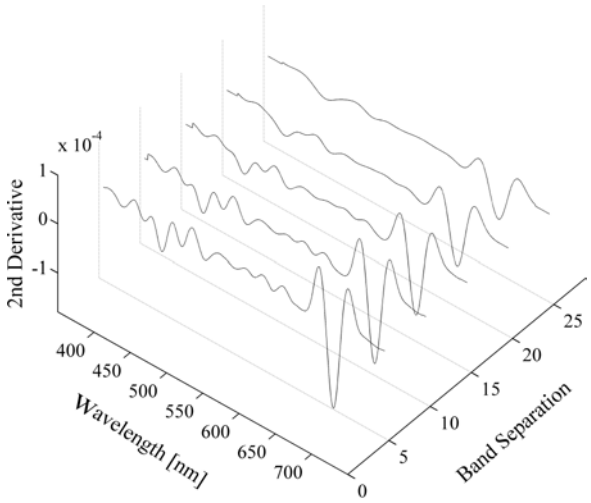


Figure 2.3: Second derivative spectra corresponding to one of the $a^*_{ph}(\lambda)$ spectra shown in Fig. 2.2a, computed for several values of band separation (BS), each of them leading to spectral features at different scales and resolution.

2.1.2 Hierarchical cluster analysis

As it has been mentioned, several of the studies dedicated to derivative analysis of hyperspectral remote-sensing reflectances have traditionally focused on the use of ratios of discrete values of derivative spectra to characterize or classify oceanic environments (Lubac *et al.*, 2008; Louchard *et al.*, 2002). However, more advanced and contemporary methods are beginning to take advantage of the information contained in the whole derivative spectrum. For example, Filippi (2007) proposes a combination of derivative spectroscopy and artificial neural network (ANN) algorithms. The derivative-neural approach has been proven to be effective for providing bathymetry, bottom type and constituent concentration estimations from $R_{rs}(\lambda)$ measurements.

In this chapter, the assessment of similarity between hyperspectral optical spectra has been made using a similarity-based cluster algorithm. This type of technique accounts for complete spectral behavior of optical data with no need to identify specific spectral features (Pekalska and Duin, 2000). This approach is particularly useful in optical oceanography when the global contribution over the entire spectral range of all pigments present in the water sample wants to be considered (Duin *et al.*, 1997).

A method based on a hierarchical cluster analysis (HCA) has been applied using hyperspectral optical data as input vectors (or objects). The HCA method utilizes an unsupervised classification algorithm which creates a hierarchical cluster tree (or dendrogram) by partitioning a given collection of unlabeled input data into clusters or groups of objects (Berkhin, 2006; Jain *et al.*, 1999). Each cluster includes objects that have a similar pattern between them and a dissimilar pattern from objects distributed in the other clusters. Figure 2.4 illustrates the process of creation of each cluster tree in which a linkage algorithm is utilized based on initial calculations of a pairwise distance between all objects included in the input data set.

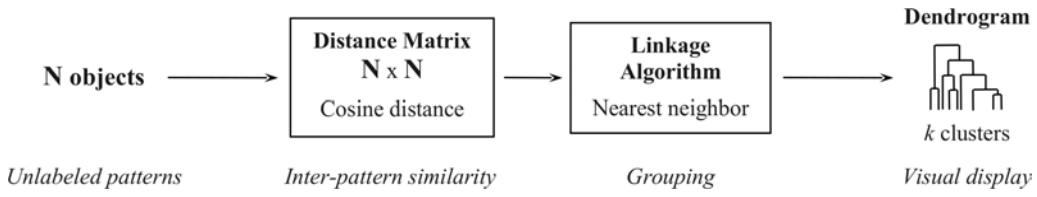


Figure 2.4: Diagram of stages in unsupervised hierarchical cluster analysis (HCA). Hyperspectral optical data (or derivative) are utilized as input.

Data in each cluster of a dendrogram share a proximity according to some defined distance measure. The selected distance measure determines how the similarity of two input objects is calculated. In this study, the partitioning of data into clusters

has been based on the determination of an angular distance between each pair of examined input data objects (i.e., each pair of reflectance or absorption spectra). In particular, the similarity between each pair of objects has been evaluated using the cosine distance, d, calculated as one minus the cosine of the angle θ between each pair of objects:

$$d(x_1, x_2) = 1 - cos\,\theta = 1 - \left(\frac{x_1 \cdot x_2}{||x_1|| \times ||x_2||}\right) \tag{2.3}$$

where x_1 and x_2 include the two considered input data objects and the cosine of the angle between the objects (θ) is obtained as the ratio of the dot product of the objects to the product of norms of the objects. Note that as the angle between the objects decreases the cosine approaches 1, resulting in a smaller distance between the input data objects and therefore higher similarity.

Other measures of similarity between the input objects have been tested, e.g., Euclidean distance using a similar approach to that proposed in Robila (2005). Finally, the cosine distance has been selected as the most appropriate measure for this study because it reflects mainly the differences in the spectral shape of optical data rather than magnitude. The cosine distance is also advantageous because it is scale invariant, i.e., insensitive to normalization of optical spectra at a specific wavelength.

As a linkage algorithm, the shortest distance D, also referred to as the nearest neighbor, has been computed to measure the distance between two clusters of objects in the tree:

$$D(a, b) = min\,[dist(x_{ai}, x_{bj})]\;\; i\,\epsilon\,(1, \ldots, n_a)\; and\; j\,\epsilon\,(1, \ldots, n_b) \tag{2.4}$$

where x_{ai} is the ith object in cluster a and x_{bj} is the jth object in cluster b.

In the traditional graphical representation of a dendrogram, the individual objects appear at one end and a single cluster containing all objects at the other end (e.g., Jain *et al.*, 1999). In this study, pairs of objects showing a small cosine distance between them (because of a similar pigment composition and spectral optical pattern) will provide a small linkage distance and therefore appear closer in the cluster tree.

2.2 Results and discussion

2.2.1 Experimental design

The potential offered by hyperspectral observations to identify phytoplankton communities in open ocean waters is investigated based on the analysis tools described in the previous section (i.e., derivative spectroscopy and HCA). A simulation-based framework is used to achieve this goal as a primary approach, which includes the use of the Hydrolight/Ecolight version 5.0 radiative transfer (RT) model (Mobley and Sundman, 2008). As one of the main interests of this study lies in a methodology

for estimating variability in phytoplankton communities from remote-sensing optical observations, this study is based on a set of remote-sensing reflectance, $R_{rs}(\lambda)$, spectra.

A brief description of the methodology proposed and carried out in this research work is provided here (and schematically presented in Figure 2.1). As a first stage, the generation of a set of simulated hyperspectral $R_{rs}(\lambda)$ spectra is performed using a RT oceanic model, which takes into account the constituents present in the water column, their inherent optical properties (i.e., absorption and scattering properties of each constituent) and their concentration profile. The next stage encompasses the computation of the second derivative spectrum of each modeled $R_{rs}(\lambda)$. Previous to derivative analysis, however, a mean filter smoothing type is also applied to $R_{rs}(\lambda)$ data. It consists of a simple average of points within the chosen filter window. As stated in Section 2.1.1, the efficient way to accomplish the smoothing and derivative processing is carefully adapting the filter size (WS) and the sampling interval (BS) in order to better match the scale of the spectral features of interest in each case. As a final stage, the goal of validating the potential of hyperspectral $R_{rs}(\lambda)$ measurements for identifying different phytoplankton communities is addressed through a similarity assessment between $R_{rs}(\lambda)$ spectra and also between second derivative spectra of $R_{rs}(\lambda)$. To make such a similarity assessment, an approach based on a hierarchical cluster analysis is used (see Section 2.1.2). Spectra corresponding to open ocean scenarios with a similar phytoplankton composition (i.e. with a similar spectral pattern) are expected to appear closer in the dendrogram than those having a different phytoplankton composition. The feasibility of considering hyperspectral measurements of $R_{rs}(\lambda)$ to identify phytoplankton communities is assessed by analyzing how close ordinary $R_{rs}(\lambda)$ spectra (or derivative spectra) dominated by the same phytoplankton group appear in the computed cluster tree.

2.2.2 Modeling of underwater optical scenarios

A synthetic data set of hyperspectral $R_{rs}(\lambda)$ has been created covering a range of environmental conditions in terms of phytoplankton composition of an hypothetical open ocean scenario. Despite natural phytoplankton assemblages usually consist of several species, the study presented in this chapter has focused on different simplified open ocean scenarios. As a first approach for testing the proposed methodology, different scenarios have been created under fully-controlled conditions and each dominated only by a single phytoplankton group.

The RT optical simulations have been carried out within the spectral region 350 to 750 nm with a spectral resolution of 1 nm. In order to emulate the real conditions for an open ocean scenario, the ocean has been assumed to be infinitely deep and the calculations included the inelastic Raman scattering and fluorescence by chlorophyll and CDOM within the ocean. The sea surface boundary conditions have been configured using the default settings in the Hydrolight/Ecolight RT model, being the same for all modeled scenarios (i.e., wind speed equal to 5 m/s, solar zenith angle equal to 30^{o} and null cloudiness). The inherent optical properties (IOPs) of the water column

required as input to the simulations have been derived from Pope and Fry (1997) for the pure water component and from Kim and Philpot (2006) for the phytoplankton component, which included specific absorption and scattering spectral coefficients corresponding to six different phytoplankton groups. Figure 2.5 shows the significant variability in spectral features of the specific absoprtion coefficients, $a^*_{ph}(\lambda)$, of these phytoplankton groups: cryptophyceae, cyanophyceae, diatoms, dinophyceae, prasinophyceae and prymnesiophyceae. The contribution of the inorganic component of particulate matter (i.e., minerals) to total absorption and scattering coefficients has been considered negligible for open ocean scenarios (see Fig. 1.6 for Case 1 waters). For the colored dissolved organic matter, CDOM, absorption has been configured to covary with particle absorption coefficient according to the commonly used model from Morel and Maritorena (2001).

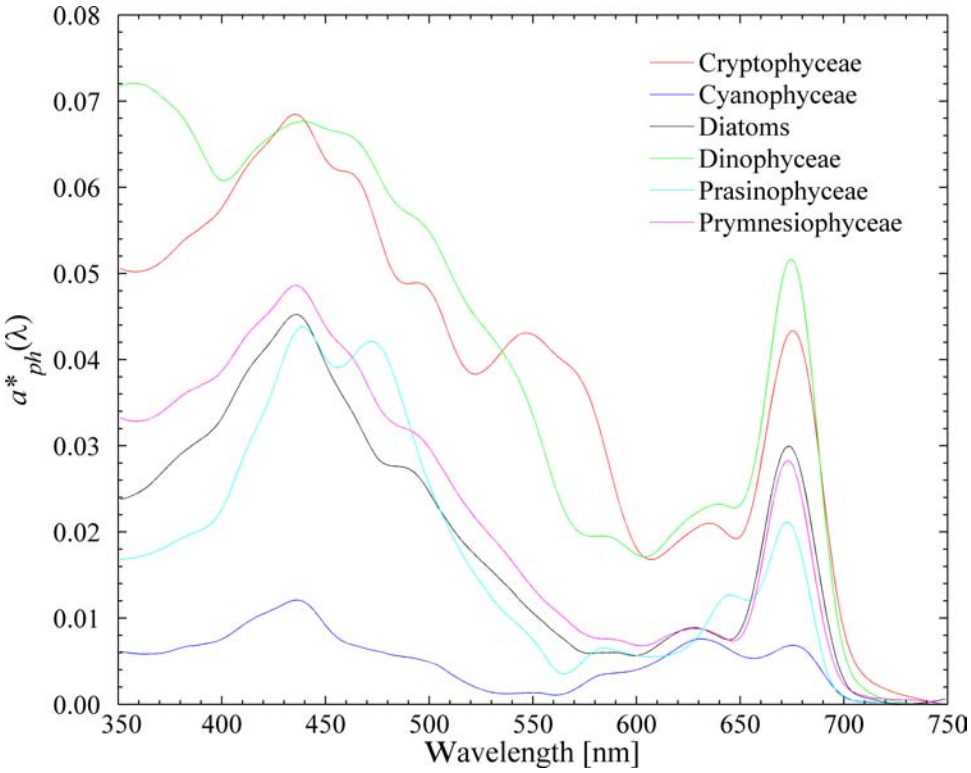


Figure 2.5: Specific absorption coefficients corresponding to six phytoplankton groups, each indicated in a different color.

A set of hyperspectral $R_{rs}(\lambda)$ spectra have been simulated, each of them characterized by the presence of only one phytoplankton group along the water column, which has been assumed to be optically homogeneous (i.e., constant concentration depth profile for all components). Therefore, a total of thirty $R_{rs}(\lambda)$ spectra have been generated from combining the six different dominating phytoplankton groups at five different concentration values (i.e., 0.01, 0.03, 0.05, 0.07 and 0.09 mg/m^3).

It is noted that these levels correspond to typical concentrations encountered under non-bloom sea conditions. Figure 2.6 illustrates how the modeled $R_{rs}(\lambda)$ spectra display a great variability in both magnitude and spectral shape, in correspondence with the variable phytoplankton composition at different concentration levels.

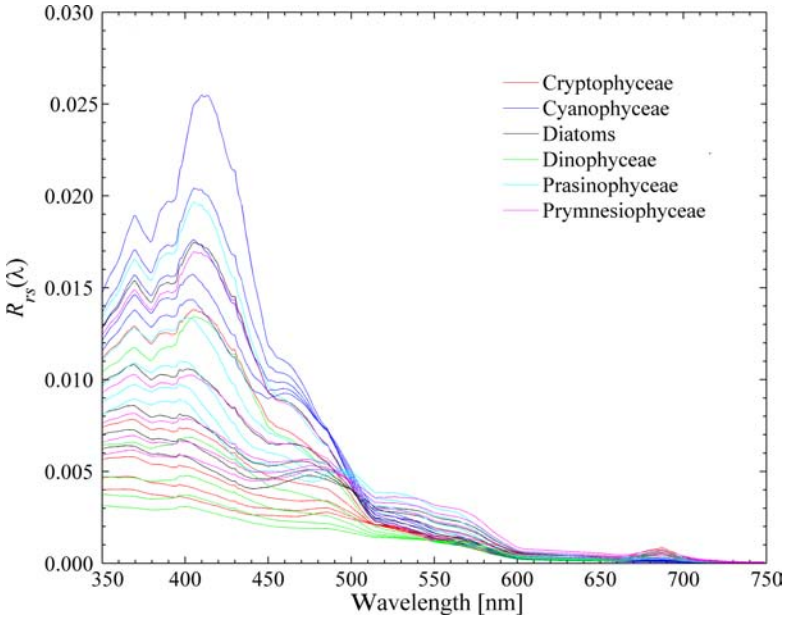


Figure 2.6: Results of RT simulations showing the differences in hyperspectral $R_{rs}(\lambda)$ spectra (1 nm resolution) corresponding to different dominating phytoplankton groups at different concentration levels (i.e., 0.01, 0.03, 0.05, 0.07 and 0.09 mg/m^3).

2.2.3 Automatic classification of phytoplankton communities

HCA clustering techniques have been used independently to identify and group similar phytoplankton communities from hyperspectral $R_{rs}(\lambda)$ spectra. The cluster tree obtained from the analysis applied to the ordinary hyperspectral $R_{rs}(\lambda)$ spectra is displayed in Fig. 2.7. Each scenario (i.e., each simulated $R_{rs}(\lambda)$) is identified with a specific label, consisting of the name of the dominating phytoplankton group and the concentration value. For instance, if diatoms are the dominating phytoplankton group, with a concentration of 0.05 mg/m^3 along the homogeneous water column, the label identifying that case will be Diat_0.05. In addition, a specific color is used to identify all cases corresponding to the same phytoplanktonic group.

It is worth noting that, when ordinary $R_{rs}(\lambda)$ spectra are used as input data for the cluster analysis, the dendrogram provides different clusters but most of them composed by cases corresponding to oceanic scenarios with a different phytoplanktonic dominance. The only phytoplankton group which is clustered satisfactorily and separately from the rest is the one corresponding to the group of prasinophyceae (identified

in color cyan), except for the lowest concentration (i.e., 0.01 mg/m^3). Whereas, the remaining cases dominated by the other algal groups (i.e., diatoms, cryptophyceae, cyanophyceae, dinophyceae and prymnesiophyceae) at different concentration levels are grouped all mixed. In addition, no improvements are obtained from testing the cluster analysis of ordinary $R_{rs}(\lambda)$ using an Euclidean distance measure instead of an angular distance (not shown).

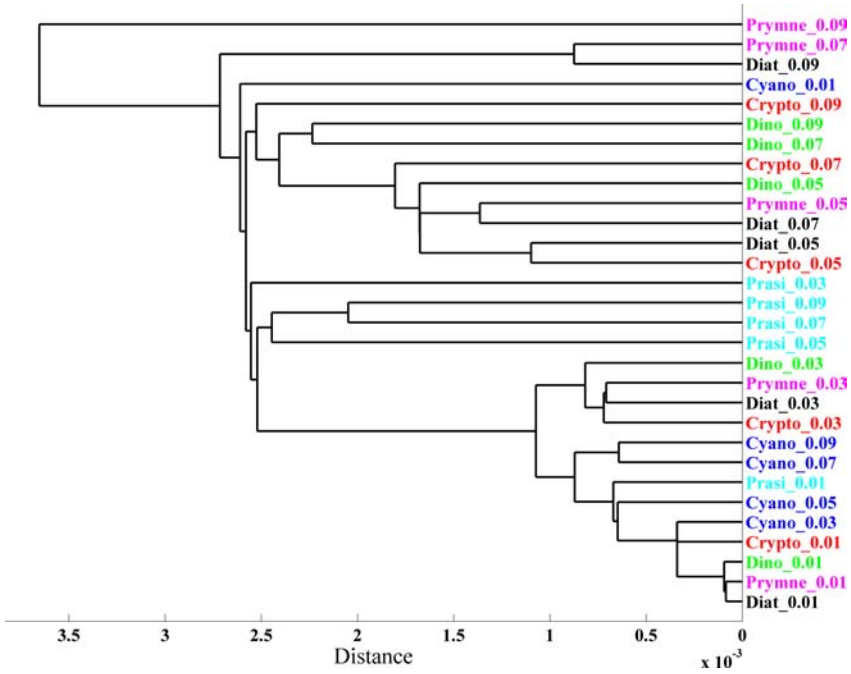


Figure 2.7: Cluster analysis based on modeled ordinary hyperspectral $R_{rs}(\lambda)$ spectra shown in Fig. 2.6.

These results point out the real challenge that implies the identification of phytoplankton community composition from the analysis of the apparent optical property: $R_{rs}(\lambda)$. The reason is that the spectral $R_{rs}(\lambda)$ of the ocean is highly influenced by the optical response of the phytoplankton pigment composition but also by many other optically significant non-phytoplankton constituents (e.g., CDOM).

The unsatisfactory results obtained from the cluster analysis based on the ordinary $R_{rs}(\lambda)$ spectra suggests the utilization of derivative spectra of $R_{rs}(\lambda)$ in order to improve our ability to classify different open ocean scenarios dominated by different phytoplankton groups. Figure 2.8 depicts the second derivative spectra of $R_{rs}(\lambda)$, obtained with the parameters WS and BS arbitrarily selected to be equal to 5 nm and 11 nm, respectively. These values provided a good trade-off between denoising and enhancement of spectral details of interest. Although derivative calculations have been made using data from the whole spectral range (i.e., 350 - 750 nm), the results from the derivative analysis of $R_{rs}(\lambda)$ spectra are reported between 361 nm ($\equiv$

$350+11$) and 739 nm ($\equiv 750-11$) according to the BS considered. As it is observed, derivative analysis of $R_{rs}(\lambda)$ data enhances shape singularities in hyperspectral data, which are significant since they are related to absorption features of phytoplankton pigments present in the samples, specially within the spectral region below 540 nm (see Fig. 1.3).

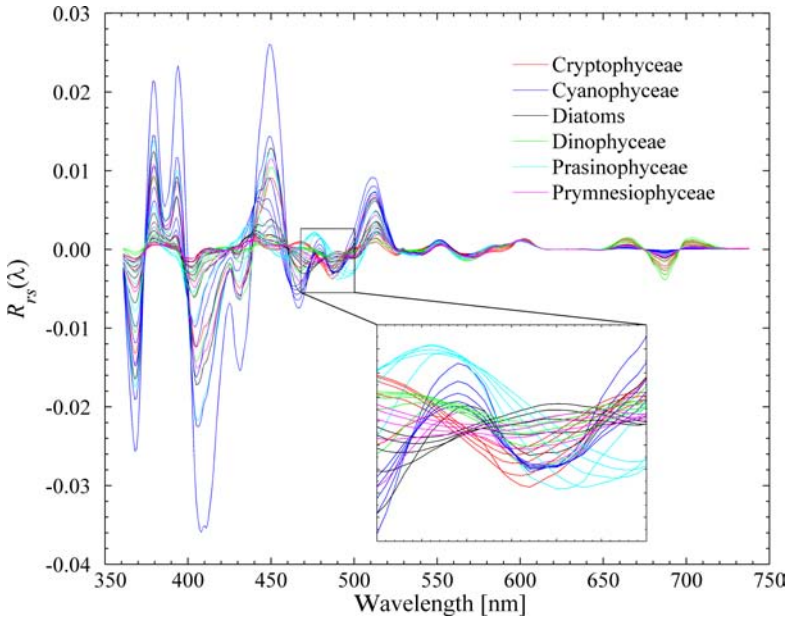


Figure 2.8: Second derivative spectra of the modeled $R_{rs}(\lambda)$ shown in Fig. 2.6. Derivative spectra for the range of 470-500 nm is enlarged with regard to the sensitivity analysis discussed at the end of this section (see Fig. 2.11). This is an optically-significant region in terms of detecting the phytoplankton community composition.

The cluster analysis based on the second derivative of $R_{rs}(\lambda)$ spectra and an angular distance measure provides better results (see Fig. 2.9) in comparison to the results obtained based on the ordinary $R_{rs}(\lambda)$ spectra. When second derivative spectra over the whole spectral range are considered for the cluster analysis, the majority of scenarios dominated by the same phytoplankton groups at different concentration levels are identified and connected by closer dendrites in the dendrogram. In this sense, derivatives of $R_{rs}(\lambda)$ spectra corresponding the same phytoplankton group (i.e., the same color) are located closer in the dendrogram and create clusters that are clearly separated one from each other, specially for the dinophyceae and prasinophyceae. The degree of separation between clusters which group scenarios of the same phytoplanktonic dominance is higher. However, there exist some cases which are not well grouped, corresponding to those with the lowest concentration levels (i.e. 0.01 mg/m^3 and in some case 0.03 mg/m^3). In fact, two "multi-color" clusters can be identified grouping all these scenarios due to $R_{rs}(\lambda)$ spectra corresponding to low concentra-

tions have a similar signature regardless the dominant phytoplankton group. This is something expected given that spectral features of absorption of phytoplankton are less evident and have a smaller effect in the $R_{rs}(\lambda)$ spectrum as the concentration decreases.

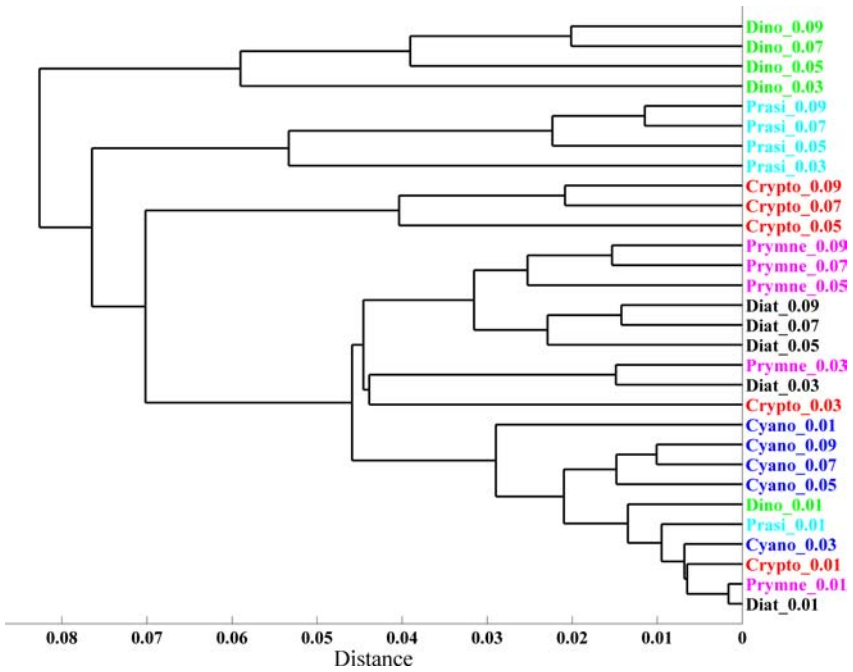


Figure 2.9: Cluster analysis based on the second derivative of hyperspectral $R_{rs}(\lambda)$ spectra for the total spectral range from 361 to 739 nm (see Fig. 2.8).

Despite the previous analysis does not allow to resolve different phytoplankton communities at low concentrations, the results provide the first detailed demonstration of the advantages and limitations of integrating hyperspectral $R_{rs}(\lambda)$ data, derivative spectroscopy and HCA cluster techniques for an automatic identification of phytoplankton communities in open ocean environments. The results confirm the potential of accounting for complete spectral behavior of derivative of hyperspectral $R_{rs}(\lambda)$ spectra in comparison with the use of multispectral observations or band ratios of discrete spectral values of derivatives (Lubac *et al.*, 2008; Louchard *et al.*, 2002). The HCA cluster analysis based on the hyperspectral data as input, with variable optical conditions, has been able to automatically bring together scenarios corresponding to the same phytoplanktonic dominance, going beyond the studies focused on the detection of the presence of a single phytoplankton group (Craig *et al.*, 2006).

It is important to note that the performance of cluster analysis applied to derivative spectra of $R_{rs}(\lambda)$ is clearly dependent on the examination of the spectral data and the optimal selection of involved parameters (i.e. smoothing filter size and band separation) for each particular data set and for each specific purpose of the analysis. For instance, if smoothing and derivative parameters are selected to be too coarse,

spectral features of interest might be lost, probably causing a worse performance. Figure 2.10 illustrates an example in which the cluster analysis is applied to second derivative spectra of the modeled hyperspectral $R_{rs}(\lambda)$ obtained using as parameters a $BS = 40$ nm and $WS = 40$ nm. In this case, scenarios dominated by the groups prymnesiophyceae and diatoms are clustered together and scenarios corresponding to the group of cyanophyceae (in color blue) appear also mixed with other phytoplankton groups. Therefore, the resulting analysis based on the derivative settings for spectra of $R_{rs}(\lambda)$ provides a worse partitioning, which prevents a satisfactory identification of different phytoplankton communities. It is important to note that a wide range of values for the BS and WS parameters were also tested to confirm that variability of our performance but results are not shown.

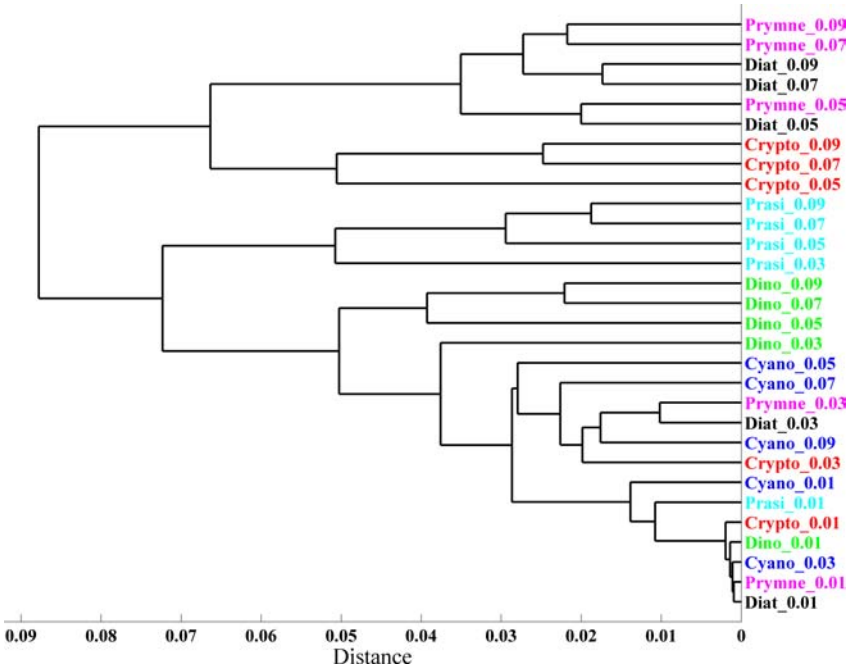


Figure 2.10: Cluster analysis based on the second derivative of $R_{rs}(\lambda)$ spectra using analyzing parameters that are too coarse (i.e. $BS = WS = 40$ nm).

In order to keep on investigating the importance of selecting the optimal parameters involved in the proposed analysis, it is also essential to take into account the spectral region considered for the analysis. Because the absorption characteristics of main accessory pigments occur at specific wavelength ranges, it is therefore useful to examine whether the cluster analysis of derivative spectra of $R_{rs}(\lambda)$ yields better results if a specific spectral range is considered. Following the examination of derivative spectra of $R_{rs}(\lambda)$ (see Fig. 2.8), a region in which the different scenarios present distinct remote-sensing optical patterns is detected. This spectral region ranges from 470 to 500 nm and coincides with a spectral region in which several accessory pigments have relevant spectral features (see Fig. 1.3). The dendrogram

obtained from cluster analysis as applied to derivative $R_{rs}(\lambda)$ spectra over this region is displayed in Fig. 2.11. The analysis based on this limited spectral information provide a very good performance, improving significantly the results obtained based on the complete spectral range (see Fig. 2.9). In this sense, scenarios corresponding to all groups of phytoplanktonic dominance are clearly identified and grouped in separate clusters, showing a much higher degree of separation between them. The methodology proposed, however, still presents some limitations when the analyzed underwater scenarios are characterized by a very low concentration level of phytoplankton (i.e., 0.01 mg/m^3). Regardless the dominant phytoplankton group, the optical patterns corresponding to these scenarios are too similar and therefore, are grouped in a unique cluster. This cluster also includes the scenarios dominated by cyanophyceae at different levels of concentration (in blue color), which in fact seems logical since it is the phytoplankton group that presents a weaker absorption pattern (see Fig. 2.5).

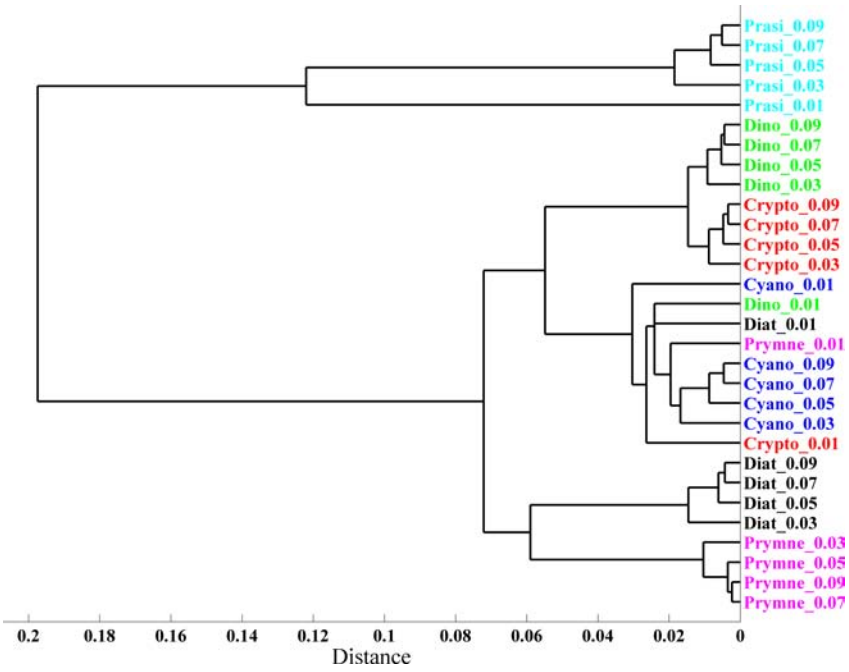


Figure 2.11: Cluster analysis based on second derivative of hyperspectral $R_{rs}(\lambda)$ spectra limited to the spectral range from 470 nm to 500 nm (see enlarged region in Fig. 2.8).

2.3 Summary and conclusions

The effectiveness of hyperspectral optical information, in particular remote sensed observations, for assessing phytoplankton diversity in the ocean has been explored in this chapter. As a primary approach, the feasibility of hyperspectral oceanographic

data to discriminate different phytoplankton communities in the open ocean (i.e., under non-bloom conditions) has been examined using a simulation-based framework. The utilization of the model Hydrolight/Ecolight version 5.0 has served to fully understand the fundamentals of the oceanic radiative transfer (RT) and to generate a synthetic data set of hyperspectral remote-sensing reflectance spectra, $R_{rs}(\lambda)$. In this chapter, several simplified scenarios have been utilized in order to base the analysis on hyperspectral data created under fully-controlled conditions and each dominated by a single phytoplankton group.

A new methodology has been proposed for estimating variability in marine phytoplankton communities from the modeled hyperspectral optical data set of $R_{rs}(\lambda)$, which is based on the assessment of differences in the shape of high spectral resolution data. Shape singularities in hyperspectral oceanographic data are relevant since they are related to absorption features of pigments present in the samples. In this sense, shape spectral features have been analyzed to obtain information about optically signicant water constituents, in particular about phytoplankton community composition. The availability of a hyperspectral data set corresponding to different modeled oceanic environments has provided the opportunity to test this new methodology, which consists in the application of derivative spectrospcopy and a HCA cluster technique. Derivative spectroscopy has proven useful to extract more information regarding phytoplankton communities by enhancing subtle features in hyperspectral data of $R_{rs}(\lambda)$. Whereas, the unsupervised hierarchical cluster analysis (HCA) applied to hyperspectral data and their derivative spectra has allowed to carry out an assessment of dissimilarity between high resolution optical data and to automatically distinguish those open ocean scenarios dominated by the same phytoplankton groups.

The HCA cluster analysis has been optimized by selecting an angular distance as a measure to determine the degree of similarity between each pair of examined input spectra, with the aim to recognize similar optical patterns from their shape features. In addition, this similarity-based cluster technique accounts for complete spectral behavior of optical data with no need to identify specific spectral features. This approach is particularly useful in cases when it is difficult or impossible to define explicitly a set of unambiguous diagnostic spectral features in the original optical data (Duin *et al.*, 1997; Pekalska and Duin, 2000), or in studies such as the ones presented in this chapter, in which the global contribution of all pigments over the entire spectral range is explored.

The application of derivative analysis to numerical simulations of hyperspectral $R_{rs}(\lambda)$ corresponding to different open ocean environments has yielded the most satisfactory results. A better ability for identifying different oceanic scenarios from derivative spectra of $R_{rs}(\lambda)$, in comparison with the use of ordinary $R_{rs}(\lambda)$ spectra, has been demonstrated using the HCA similarity-based cluster approach. In this sense, the cluster partitioning obtained from the analysis of derivative spectra of $R_{rs}(\lambda)$ as input data has provided a better separation of oceanic environments corresponding to the same phytoplanktonic dominance.

As a preliminary study, it has been demonstrated the potential of this methodology to provide a means for optical oceanographers to better map specific phytoplankton

communities from hyperspectral oceanographic information, including remote sensed observations. Nevertheless, the ability to automatically classify scenarios corresponding to the same phytoplanktonic dominance by analyzing complete spectral behavior of derivative of hyperspectral $R_{rs}(\lambda)$ data instead of discrete values at specific wavelengths, has shown to be dependent on the examination of the spectral data and on the optimal selection of parameters involved in the analysis. First, the achievement of a good performance from derivative analysis of hyperspectral $R_{rs}(\lambda)$ stems from the fact that a suitable selection of the smoothing and derivative parameters (i.e., filter size and band separation) must be done according to the resolution of each type of hyperspectral data (Torrecilla *et al.*, 2007). Otherwise, spectral features of interest are lost, preventing a satisfactory identification of different oceanic environments. Secondly, a significant improvement of our performance is achieved if the cluster analysis is limited to derivative spectra of $R_{rs}(\lambda)$ over a restricted spectral region. In particular, a better identification of phytoplankton communities has been obtained considering the spectral range from 470 to 500 nm, which overlaps with the wavelength range where absorption characteristics of main phytoplankton pigments occur. It is important to note, though, that this successful identification has only been possible for simulated scenarios whose phytoplankton concentration was larger than 0.01 mg/m^3. For low concentrations, the corresponding optical pattern of $R_{rs}(\lambda)$ spectrum has been very similar regarless the phytoplanktonic dominance of each modeled scenario and, therefore, not separated with the cluster analysis.

Following completion of the described analyses, the advantages of integrating RT-modeled hyperspectral data, derivative spectroscopy and the HCA cluster technique for an automatic identication of phytoplankton communities in open ocean environments have been identified, likewise its limitations. In this sense, future research described in chapters to follow has focused on addressing several questions. For instance, the exploration of the proposed methodology using a more realistic approach based on optical data collected in the field is described in Chapters 3 and 4. As a natural following step, the examination of the hyperspectral-based approach by considering larger data sets of optical observations, corresponding to real open ocean environments, becomes essential. In these chapters, it is also tackled the possibility to identify with the proposed methodology different real phytoplankton assemblages (i.e., scenarios in which different phytoplankton species co-exist). In order to do so, the methodology proposed in the present chapter (see diagram in Fig. 2.1) has been improved by developing an automatic tool of validation, aimed at evaluating how well different optical data sets perform when considered as input data in the cluster analysis. It is described how the effectiveness of the optical classification can be illustrated and quantified using two objective indices. This new validation criteria allows to go beyond the subjective evaluation used in the present chapter (i.e., the visual inspection of each cluster tree as a measure of validation of results).

Another assessment of the potential of the proposed methodology is described in Chapter 5 but in this case focused on more complex underwater scenarios corresponding to shallow estuarine environments (i.e., Case 2 waters). In this case, the RT modeling has been adapted to this type of coastal environments in which non-

phytoplankton optically significant water constituents play an important role besides phytoplankton (e.g. CDOM, suspended sediments). The proposed methodology has been tested for detecting different phytoplankton groups given different biomass distributions along the vertical structure of the water column and bottom types, based on the analysis of the remote-sensing reflectance.

2.4 Acknowledgements

This study was supported by the projects HIDRA (PIE06-301102) and ANERIS (PIF08-015) funded by the Spanish Ministry of Science and Innovation. E. Torrecilla was supported by a contract (Program I3P-BPD2005) from the Spanish National Research Council (CSIC).

3 Hyperspectral data for discriminating phytoplankton pigment assemblages in the open ocean

The results from this chapter have mostly been published as:

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 Environment, 115, doi: 10.1016/j.rse.2011.05.014:25782593.

The feasibility of hyperspectral oceanographic data to discriminate phytoplankton communities has been demonstrated in the previous chapter under a simulation-based framework and using a new approach. In this chapter, the analysis is extended to more realistic environmental scenarios based on field data collected as part of a research initiative carried out by the group from the Marine Physical Laboratory (Scripps Institution of Oceanography, San Diego, USA) led by Prof. Dariusz Stramski. In particular, phytoplankton pigment data are analyzed in conjunction with concurrently obtained optical data (i.e., absorption coefficients and remote-sensing reflectance), which were determined in the field along a north-to-south transect in the eastern Atlantic Ocean in 2005. The primary goal is to examine the feasibility of classifying different real open ocean environments under non-bloom conditions in terms of phytoplankton pigment assemblages from analysis of hyperspectral absorption and remote-sensing reflectance measurements. In order to address this question, as it is schematically presented in Fig. 3.1, the unsupervised hierarchical cluster analysis described in Section 2.1.2 is applied to the optical data sets including the spectra of absorption coefficients and remote-sensing reflectance and their second derivative spectra and to the pigment data set obtained from HPLC chromatographic analysis of seawater samples. In this chapter, a validating tool is proposed to evaluate how well different optical data sets perform. This novel approach is based on the pigment information, which has been commonly used by the scientific community as a proxy for the phytoplankton composition. Therefore, the pigment-based clusters provide a reference for partitioning the selected data sets into distinct subsets, each characterized by different phytoplankton pigment composition. In order to illustrate the effectiveness of the optical classification, the degree of similarity between the pigment-based clusters (the "sea truth") and optical-based clusters is quantified using two indices, cophenetic and Rand.

The present chapter is divided in three major parts:

- In Section 3.1, before the cluster analysis, since only multispectral measurements of remote-sensing reflectances were carried out in the field, a radiative transfer model is used to reconstruct the hyperspectral $R_{rs}(\lambda)$ spectrum at each station. A closure exercise is described to discuss the level of consistency between the measured multispectral $R_{rs}(\lambda)$ and the modeled hyperspectral $R_{rs}(\lambda)$. The successful simulations are based on a comprehensive suite of IOPs (also used in the further cluster analysis) and atmospheric measurements.

- In Section 3.2, as part of the cluster analysis, the new validation tool proposed to evaluate how well different optical data sets perform is described. The two indices, cophenetic and Rand, proposed as objective criteria for the evaluation of similarity between cluster partitions are described.

- In Section 3.3, the results from the cluster analysis of the pigment information and several optical data sets measured in the field and modeled as described in the first section are provided. Nine stations were selected in this study to represent distinct differences in major accessory pigments present in the samples. This analysis is basically presented as a proof-of-concept study in which the approach is to use a relatively small but carefully selected set of data, which exhibits significant contrasts in the composition of pigments, rather than to indiscriminately use a large data set. In addition, the degree of similarity between cluster trees is evaluated for calculations involving different spectral ranges of optical data. A sensitivity of cluster analysis to the choice of the derivative parameters, when required, is also included in this section.

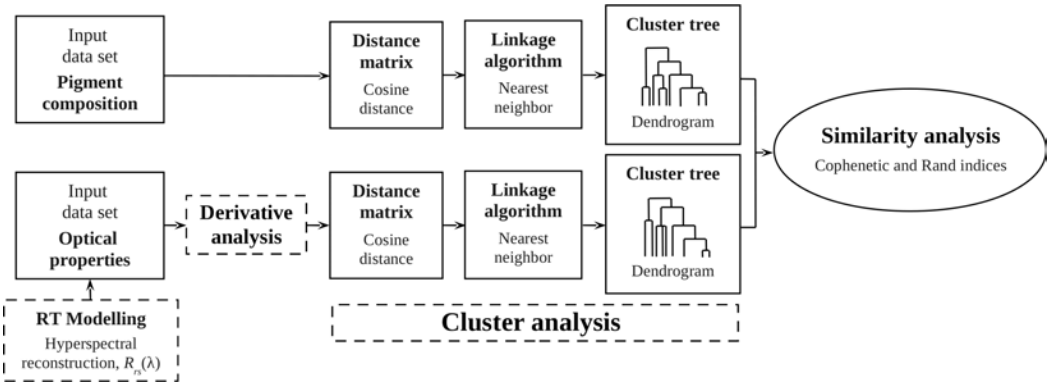


Figure 3.1: A schematic diagram illustrating the general approach to cluster analysis and similarity determination proposed in this chapter. The cluster tree obtained with pigment composition as the input (upper pathway) is used as the reference (the "sea truth") for comparison with results obtained using various optical data as input (i.e., different absorption coefficients and remote-sensing reflectances at different spectral resolutions).

3.1 Reconstruction of hyperspectral reflectances

The combination of detailed in situ and laboratory measurements is essential in gaining a better knowledge of the optical characteristics of the ocean waters. For instance, in situ measurements of upwelling and downwelling radiation within the water might permit the validation of the components of the radiation field resulting from radiative transfer model estimations. In this sense, the combination of measurements performed within the framework of the detailed field program described in this section, will be used to form a "closure experiment". This experiement will be possible since measured inherent optical properties will be used as input information to perform the model simulations of $R_{rs}(\lambda)$ spectra, whereas measured radiance and irradiance profiles will be used to compute related optical properties such as $R_{rs}(\lambda)$ and compared to the model's output. The radiative transfer (RT) model Hydrolight/Ecolight version 5.0 (Mobley and Sundman, 2008) will be used for this experiment which aim is two-fold. First, model calculations of hyperspectrally-resolved $R_{rs}(\lambda)$ corresponding to several stations will be obtained. Attention will be paid regarding possible errors related to the accuracy of instruments or the methodology of in situ measurements, as well as regarding uncertainties in some assumptions of underwater optical properties used in the model estimations. The results will serve to demonstrate if the conditions under which theoretical RT calculations of hyperspectral $R_{rs}(\lambda)$ were performed produce a close agreement with the experimental multispectral measurements of $R_{rs}(\lambda)$. The level of agreement between measured multispectral-$R_{rs}(\lambda)$ and modeled hyperspectral-$R_{rs}(\lambda)$ will be examined. Secondly, modeled hyperspectrally-$R_{rs}(\lambda)$ will be further analyzed for classification purposes (see Section 3.3).

3.1.1 Field measurements

Measurements of phytoplankton pigment composition and seawater optical properties were obtained during the ANT-XXIII/1 expedition of the R/V Polarstern, as part of a research initiative carried out by a group from the Marine Physical Laboratory (Scripps Institution of Oceanography) led by Prof. Dariusz Stramski. Measurements were collected along a north-to-south transect in the eastern Atlantic Ocean during October and November, 2005 (Fig. 3.2). The investigated area spanned a wide range of different oceanic environments between the English Channel and the waters off the African coast of Namibia. Typically, one full station was conducted daily near local noon throughout the cruise. These full stations consisted of in situ measurements of seawater inherent and apparent optical properties along with laboratory analyses of water samples collected from discrete depths with the ship's CTD/rosette system (Fig. 3.3). For the present proof-of-concept study, a subset of nine stations (see Fig. 3.2 for station locations) was selected for cluster analysis based on the observation of distinct differences in the ratios of dominant accessory pigments to total chlorophyll-*a* (see Section 3.3.1). The selected data from the nine stations are representative of surface waters within the top 5–10 m of the ocean, as the main interest lies in the methodology for estimating variability in phytoplankton communities from remote-

sensing reflectance. A brief description of the measurements is given in the following subsections. More methodological details which are beyond the scope of this study for the radiometric and backscattering measurements are in Stramski *et al.* (2008).

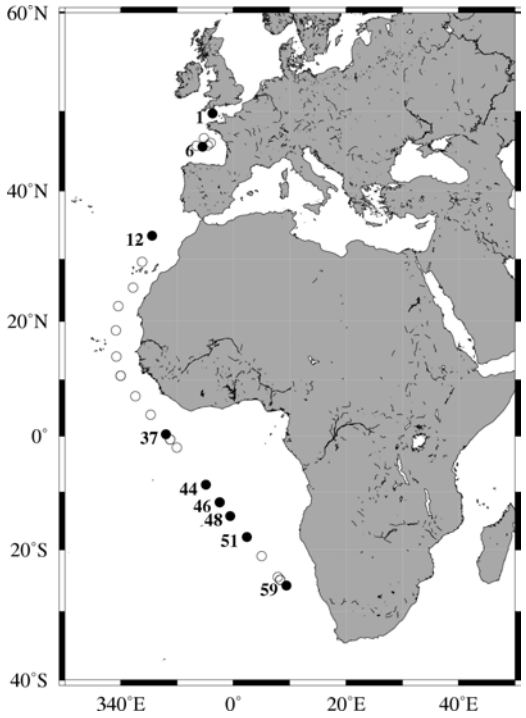


Figure 3.2: Map depicting the location of full stations sampled along the north-to-south ANT-XXIII/1 cruise track in the eastern Atlantic during October and November, 2005. Each full station consisted of in situ optical measurements accompanied by discrete water sample analyses. Stations chosen for use in this chapter are identified by filled circles and labeled.

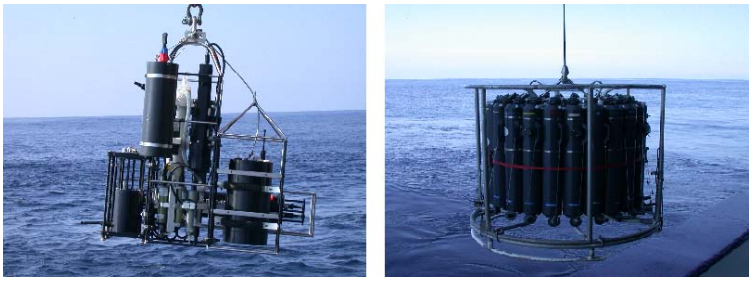


Figure 3.3: (left panel) Multisensor Datalogger System (MDS) used to measure in situ several IOPs of seawater. (right panel) CTD/rosette system used to collect water samples, which led to laboratory measurements of other IOPs. See Fig. 3.4 for details. *Source*: D. Stramski and R. Reynolds.

3.1.1.1 Pigment analysis

Concentrations of chlorophyll-*a* and accessory pigments in phytoplankton were measured on surface water samples from each station using HPLC techniques. As a standard security protocol, two sets of replicate samples were collected in the ship and analyzed at two laboratories, the Center for HydroOptics and Remote Sensing (CHORS) laboratory at San Diego State University (California, USA) and the GKSS Research Centre in Geesthacht (Germany). The CHORS analysis was based on a method described in Heukelem and Thomas (2001), and the GKSS samples were analyzed following the method of Zapata *et al.* (2000). However, the CHORS analysis included identification and quantification of more pigments (27) than the GKKS method (23) including alternative forms of chlorophyll-*a*, and therefore was chosen in this study as the primary pigment data set for identifying phytoplankton assemblages using cluster analysis (see Section 3.3.1). Throughout the rest of this chapter, it is used the CHORS values of the total chlorophyll-*a* (TChl*a)* as a measure of chlorophyll-*a*, which is defined as the summed contributions of concentrations of monovinyl chlorophyll-*a* (MVChl*a*), divinyl chlorophyll-*a* (DVChl*a*), chlorophyllide-*a* (Chlide), and the allomeric and epimeric forms of chlorophyll-*a*.

Following completion of these analyses, data quality problems in the CHORS analytical procedures were identified which suggest the potential of errors in the determination of some pigment concentrations in the samples (Hooker and Heukelem, 2009). Although it was applied an independently-derived correction factor to a few individual pigments, the nature of the methodological issues precludes development of a general correction scheme applicable for all pigments. In the Appendix B, these issues are discussed in more detail and the results of analyses are summarized, which indicate that use of the CHORS data provides nearly identical results with respect to station clustering and classification as the GKSS HPLC data. Because of this independent corroboration, it was continued to use the CHORS pigment data as the reference data set in this study. Because of a large range of pigment compositions across different phytoplankton classes, determination of phytoplankton composition from HPLC pigment data is not straightforward (e.g., Jeffrey *et al.* 1999). Whereas certain diagnostic pigments can serve as unambiguous markers for some phytoplankton classes (e.g., peridinin in dinoflagellates, alloxanthin in cryptophytes), many important pigments are shared by more than one algal taxa (e.g., fucoxanthin in diatoms, haptophytes, chrysophytes, and raphidophytes). Nevertheless, because many of the classes have distinctive suites of marker pigments, HPLC data can be useful for indicating their presence and abundance in a mixed phytoplankton population. Specifically, a useful indication of contributing phytoplankton classes can be obtained from the ratios of the concentrations of specific pigments to chlorophyll-*a* or the ratios of specific diagnostic pigments to the sum of these diagnostic pigments, because these ratios can differ between phytoplankton groups (Mackey *et al.* 1996; Vidussi *et al.* 2001; Wright *et al.* 1996).

For each of the nine stations selected in this study, it was calculated a set of pigment ratios for subsequent use in the cluster analysis (see Section 3.3.1). This set of pigment ratios consisted of ratios of the concentration of each individual pigment to the TChl$a$ concentration, as obtained from HPLC measurements at the CHORS laboratory. The following 24 pigments were included in these calculations: monovinyl chlorophyll-a, divinyl chlorophyll-a, chlorophyllide-a, chlorophyll-a allomer, chlorophyll-a epimer, monovinyl chlorophyll-b, divinyl chlorophyll-b, chlorophyll-$c2$, chlorophyll-$c3$, α - carotene, β - carotene, alloxanthin, diadinoxanthin, diatoxanthin, fucoxanthin, 19' - hexanoyloxyfucoxanthin, 19' - butanoyloxyfucoxanthin, neoxanthin, prasinoxanthin, violaxanthin, zeaxanthin, peridinin, pheophorbide-a, and lutein. The CHORS pigment data set also included a few additional pigments (chlorophyll-$c1$, gyroxanthin-diester, and pheophytin-a) which were below the detection level for the nine stations. They were not included in the analysis as they would have no effect on the results.

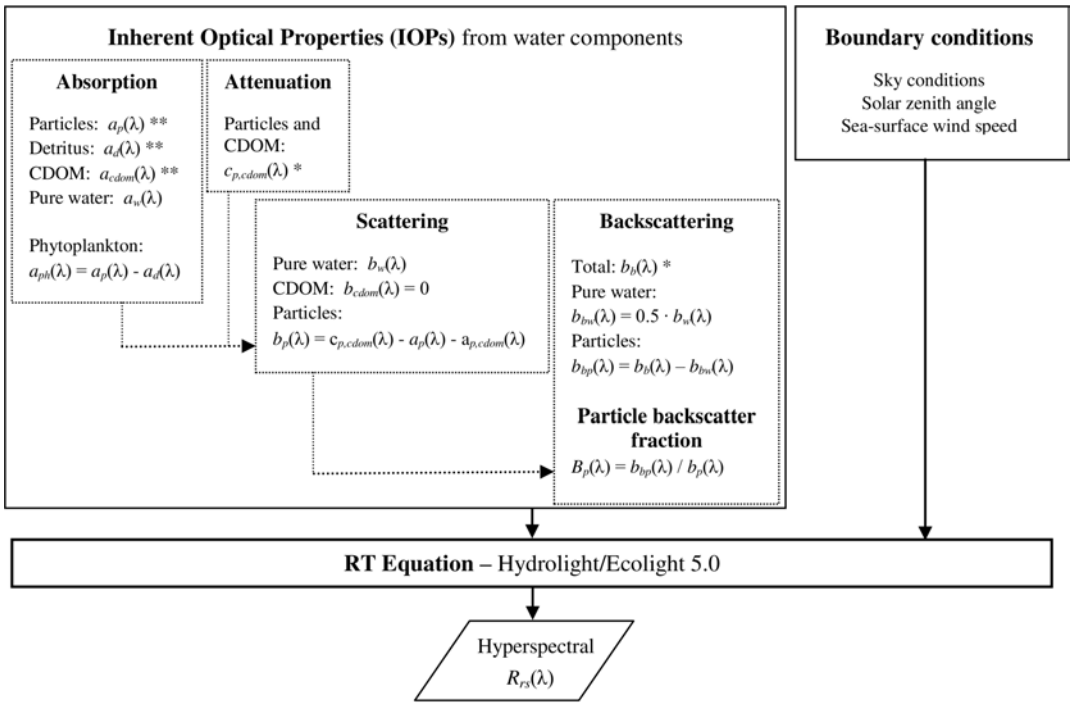
3.1.1.2 Inherent optical properties

The inherent optical properties (IOPs) of seawater were measured in situ along with laboratory analyses of water samples through the use of different instruments (Fig. 3.3). They have a two-fold application in this study. First, the absorption, scattering, and backscattering coefficients are used to define IOP inputs to radiative transfer simulations that generate hyperspectral data of remote-sensing reflectance (see Fig. 3.4 and Section 3.1.2). Secondly, the total and phytoplankton absorption coefficients, $a(\lambda)$ and $a_{ph}(\lambda)$, are utilized directly in the cluster analysis (Section 3.3.2).

As part of the work performed during the field campaign in 2005, the spectral absorption coefficients of particles, $a_p(\lambda)$, and colored dissolved organic matter (CDOM), $a_{cdom}(\lambda)$ in m^{-1}, were determined at 1-nm intervals from high spectral resolution measurements on freshly-collected discrete water samples with a point-source integrating cavity absorption meter (PSICAM) over the range 350–750 nm (Röttgers *et al.*, 2005; Röttgers and Doerffer, 2007). As the PSICAM did not provide data below 350 nm, the $a_p(\lambda)$ values within the 300 to 350 nm spectral range were obtained from filter pad measurements on discrete water samples collected on glass fiber filters (GF/F) and frozen in liquid nitrogen until analysis with a dual-beam spectrophotometer (Lambda 18, Perkin Elmer). The filter pad measurements were made with the transmittance-reflectance (T-R) technique of Tassan and Ferrari (1995, 2002) using a correction for the path length amplification factor from Stramska *et al.* (2006).

In this study, it has been chosen to use the PSICAM data of $a_p(\lambda)$ over the majority of the spectrum because the PSICAM technique involves a direct measurement of absorption on particle suspension with minimal scattering artifacts, which is expected to be generally superior to the filter pad measurements. The data of $a_{cdom}(\lambda)$ below 350 nm were obtained from an exponential fit to the PSICAM-measured $a_{cdom}(\lambda)$. A null point correction based on wavelengths in the far red or near-infrared was applied to all $a_p(\lambda)$ and $a_{cdom}(\lambda)$ spectra. The total spectral absorption coefficient, $a(\lambda)$, was determined as the sum of $a_p(\lambda)$, $a_{cdom}(\lambda)$, and the pure water component, $a_w(\lambda)$.

The latter was obtained from Pope and Fry (1997) for the spectral range 380–727 nm and from Fry *et al.* (2006) for the range 300–379 nm. It is important to note that the primary interest is in the spectral information contained at wavelengths longer than 350 nm extending throughout the visible part of the spectrum up to 725 nm where most phytoplankton pigments exhibit significant absorption features. However, data at wavelengths shorter than 350 nm are useful for this analysis, especially in the context of derivative spectra whose discrete values at specific wavelengths were calculated in this study using data covering a bandwidth on the order of 10–30 nm. The spectra of the phytoplankton absorption coefficient, $a_{ph}(\lambda)$, were determined as a difference between the absorption coefficient of particles, $a_p(\lambda)$, and the non-phytoplankton component of particulate absorption, $a_d(\lambda)$, which is commonly referred to as detrital absorption. These determinations were based on the T-R filter pad measurements, in which the $a_d(\lambda)$ spectra were measured on GF/F sample filters following treatment with sodium hypochlorite [NaOCl] (Ferrari and Tassan, 1999). In this treatment, the particles on the sample filter were exposed to a small amount of a 2% NaOCl solution for several minutes to bleach phytoplankton pigments.



* In situ measurements of IOPs collected with a Multisensor Datalogger System (Fig. 3.3, left panel).

** Measurements of IOPs obtained from laboratory analyses of water samples collected with the ship's CTD/rosette system (Fig. 3.3, right panel).

Figure 3.4: Diagram illustrating the undertaken process to determine each IOP and generate each modeled hyperspectral $R_{rs}(\lambda)$ spectra.

The spectral beam attenuation coefficient of particles and CDOM, $c_{p,cdom}(\lambda)$ in m^{-1}, was determined at each station from in situ measurements with two single-wavelength C-Star transmissometers (488 and 660 nm; WET Labs, Inc.). Note that the C-Star derived values represent the total beam attenuation, $c(\lambda)$, with pure seawater contribution, $c_w(\lambda)$, subtracted, i.e. $c_{p,cdom}(\lambda) = c(\lambda) - c_w(\lambda) = a_p(\lambda) + a_{cdom}(\lambda) + b_p(\lambda)$, where $b_p(\lambda)$ is the spectral scattering coefficient of particles and assuming that dissolved matter makes negligible contribution to scattering. In this study (see Fig. 3.4), the values of $b_p(\lambda)$ at 488 and 660 nm were thus calculated from C-Star attenuation and PSICAM absorption measurements as $b_p(\lambda) = c_{p,cdom}(\lambda) - a_p(\lambda) - a_{cdom}(\lambda)$. A power function fit was then applied to these values to produce the spectral data of $b_p(\lambda)$ over the range 300–750 nm with a 1 nm resolution. The pure seawater scattering coefficient, $b_w(\lambda)$, was calculated using the Buiteveld *et al.* (1994) equations with measured water temperature and salinity (see Twardowski *et al.*, 2007 and Stramski *et al.*, 2008). The total scattering, $b(\lambda)$, was obtained as a sum $b_w(\lambda) + b_p(\lambda)$.

The spectral backscattering coefficient, $b_b(\lambda)$ in m^{-1}, was determined in this study (see Fig. 3.4) by combining in situ measurements with three instruments, a Hydroscat-6 and two a-βeta sensors (HOBI Labs, Inc.), to yield a total of eight spectral bands: 420, 442, 470, 510, 550, 589, 620, and 671 nm. Because $b_b(\lambda)$ was generally observed to be a smooth monotonic function of wavelength within the spectral range of the measurements (Stramski *et al.*, 2008), the experimental data were fitted to a power function to obtain hyperspectral resolution over the 300–750 spectral range. The spectral backscattering coefficient of particles, $b_{bp}(\lambda)$ in m^{-1}, was determined as a difference between the total and pure seawater backscattering coefficients, $b_b(\lambda) - b_{bw}(\lambda)$, in which the pure seawater component, $b_{bw}(\lambda)$, was previously calculated as indicated: 0.5 $b_w(\lambda)$.

From the values of $b_p(\lambda)$ and $b_{bp}(\lambda)$, it was calculated the particle backscatter fraction $B_p(\lambda) = b_{bp}(\lambda)/b_p(\lambda)$ necessary for this study (see Fig. 3.4). These data were then fitted to a power function, $B_p(\lambda) = B_p(\lambda_0) \cdot (\lambda_0/\lambda)^m$, where λ_0 is the reference wavelength 550 nm. The backscattering fraction $B_p(\lambda_0)$ at the reference wavelength and the exponent m represent the best fit parameters of the linear regression analysis performed for the log-transformed data of $B_p(\lambda)$ vs. λ for each station. The parameters of the power function fit of $B_p(\lambda)$ were used as input to radiative transfer simulations (see Section 3.1.2).

3.1.1.3 Multispectral remote-sensing reflectance

Values of multispectral remote-sensing reflectance, $R_{rs}(\lambda)$ in sr^{-1}, were estimated at each station at 13 wavelengths from in situ measurements of underwater vertical profiles of spectral nadir upwelling radiance, $L_u(\lambda, z)$ in $Wm^{-2}\,sr^{-1}\,nm^{-1}$, and spectral downwelling plane irradiance, $E_d(\lambda, z)$ in $Wm^{-2}\,nm^{-1}$, where z is depth. These measurements were made with a freefall spectroradiometer (see Figure 3.5), the SeaWiFS Profiling Multichannel Radiometer (SPMR, Satlantic, Inc.). The wavelengths

for these measurements are 339, 380, 412, 443, 470, 490, 510, 532, 554, 589, 619, 666, and 683 nm. The radiometric measurements and data processing were consistent with methods recommended in NASA protocols (Mueller *et al.*, 2003b). The deployment of the spectroradiometer occurred simultaneously with water sample collection from the CTD/rosette, and immediately before or after deployment of the instrument package containing attenuation and backscattering sensors.

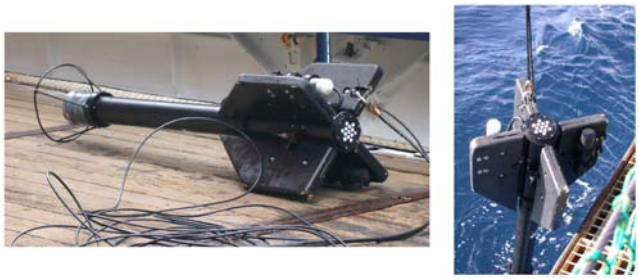


Figure 3.5: SeaWiFS Profiling Multichannel Radiometer (SPMR) utilized to measure upwelling radiances, $L_u(\lambda, z)$, and downwelling plane irradiances, $E_d(\lambda, z)$. *Source*: D. Stramski and R. Reynolds.

3.1.2 Modeled hyperspectral remote-sensing reflectance

Because hyperspectral radiometric measurements were not conducted during the ANT-XXIII/1 cruise and the primary interest is in the analysis of hyperspectral optical data, numerical simulations of radiative transfer (RT) were performed to estimate the hyperspectral remote-sensing reflectance, $R_{rs}(\lambda)$, for each of the nine selected stations (Fig. 3.4). The radiative transfer model Hydrolight/Ecolight version 5.0 was used (Mobley and Sundman, 2008). An important prerequisite for undertaking these RT simulations was the availability of a comprehensive suite of IOPs for each station for use as input to the simulations (see Section 3.1.1.2), and also the availability of the multispectral $R_{rs}(\lambda)$ derived from in situ measurements for use in validating the simulated hyperspectral $R_{rs}(\lambda)$.

The RT calculations were carried out within the spectral region from 300 nm to 725 nm with high spectral resolution (1 nm). Similarly to the absorption, the main interest is in the reflectance data at wavelengths longer than about 350 nm. However, in addition to the requirements associated with derivative calculations, the RT simulations below 350 nm are needed to account for Raman scattering contributions observed at $\lambda >$350 nm. The ocean was assumed to be infinitely deep and optically homogeneous, and the simulations included the Raman scattering and fluorescence by colored dissolved organic matter within the ocean. The sea surface boundary conditions were estimated from observations of wind speed and sky conditions (cloudiness) at each station site, and the solar zenith angle was calculated for the corresponding date, time, and geographic coordinates. The inherent optical properties of the

water column required as input to the simulations were derived from the IOP measurements in the surface waters and additional relevant determinations of $a(\lambda)$ and $b(\lambda)$ as described in Section 3.1.1.2. The selection of the particulate scattering phase function, which is also part of IOP inputs to the RT simulations, was based on the particle backscatter fraction $B_p(\lambda)$. The Fournier–Forand phase functions (Fournier and Forand, 1994; Fournier and Jonasz, 1999) were used, which are parameterized in terms of $B_p(\lambda)$ and are built into the Hydrolight/Ecolight model.

Fig. 3.6 compares the modeled hyperspectral $R_{rs}(\lambda)$ with the measured multispectral $R_{rs}(\lambda)$ for two representative stations. The model results compare reasonably well with measurements, which lends confidence to the use of hyperspectral $R_{rs}(\lambda)$ in the cluster analysis. This level of consistency between model and measurements suggests that the suite of parameters used as input to the RT simulations realistically represents the actual field conditions. The ability to define realistic inputs derives, in turn, from a comprehensive suite of IOP measurements carried out in the field.

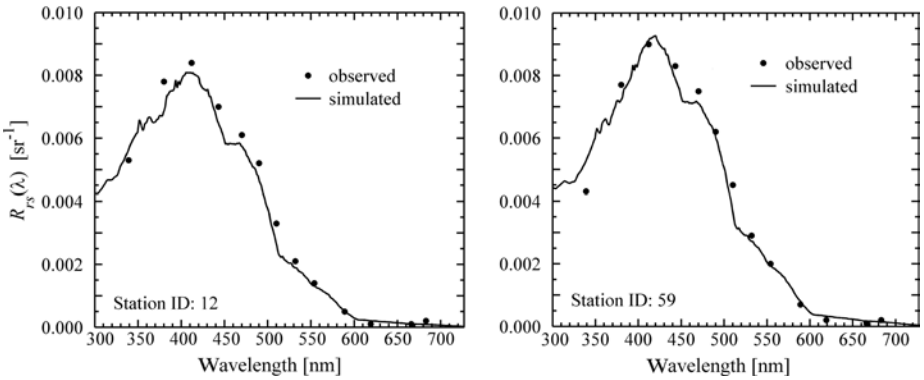


Figure 3.6: Hyperspectral (1 nm) determinations of the remote-sensing reflectance $R_{rs}(\lambda)$ obtained from radiative transfer simulations (solid line) compared with in situ multispectral measurements at 13 discrete bands (solid circles). Each panel illustrates a different station location.

3.2 Cluster analysis and similarity indices between dendrograms

A hierarchical cluster analysis (HCA) was used to classify the 9 selected stations into distinct groups on the basis of several types of input data vectors (or objects), which included the HPLC pigments and optical data derived from spectral absorption coefficients and remote-sensing reflectance. For a given type of data, the input to the cluster analysis consisted of 9 numerical data vectors, each representing one of the stations. For the input data representing the ratio of individual pigment concentrations to TChl$a$, an object for a given station is a data vector $\{p_1, p_2, p_3, \ldots, p_{24}\}$ where the consecutive elements p_i represent the ratio of each of the 24 individual pigment concentrations to TChl$a$ concentration.

Several types of optical data vectors were used as input to the HCA analysis, including objects consisting of hyperspectral data of the remote-sensing reflectance, $R_{rs}(\lambda)$, the phytoplankton absorption coefficient, $a_{ph}(\lambda)$, the sum of pure water and phytoplankton absorption coefficients, $a_w(\lambda)+a_{ph}(\lambda)$, and the total absorption coefficient, $a(\lambda)$. The input characterizing the hyperspectral remote-sensing reflectance for any given station was used in the form of the following data vector $\{R_{rs}(\lambda_1)/R_{rs}(555)$, $R_{rs}(\lambda_2)/R_{rs}(555)$, $R_{rs}(\lambda_3)/R_{rs}(555), ...,$ $R_{rs}(\lambda_n)/R_{rs}(555)\}$, where the consecutive elements represent the values of $R_{rs}(\lambda)$ at successive light wavelengths normalized to R_{rs} at 555 nm over the spectral range from λ_1 to λ_n. Similar input vectors were created for the different components of spectral absorption. Because the analysis is focused on the spectral shapes, all the optical spectra used in the cluster analysis were normalized by the value of the optical variable at 555 nm at which variations in R_{rs} within the open ocean are generally small. The spectra involving the absorption coefficients were additionally normalized by TChl*a* concentration to minimize variability in absorption associated with changes in phytoplankton biomass. The rationale for selecting the data of $a(\lambda)$, $a_w(\lambda)+a_{ph}(\lambda)$, and $a_{ph}(\lambda)$ to create input data vectors for the cluster analysis stems from the fact that the variation in the spectral shape of $a(\lambda)$ is typically a major determinant of the variation in the spectral shape of $R_{rs}(\lambda)$. In turn, the variations in the spectral shape of $a_w(\lambda)+a_{ph}(\lambda)$ or $a_{ph}(\lambda)$ can be viewed as an important or dominant source of variation in the spectral shape of $a(\lambda)$ in open ocean situations.

Vectors from the second derivative spectra of the hyperspectral reflectance and absorption objects were also created for input into the cluster analysis. The estimation of the second derivative spectra from these data was made with a finite divided difference algorithm (see Section 2.1.1), which computes the changes in curvature of a given spectrum over a sampling interval ($\triangle\lambda$) or band separation (BS) defined as $\triangle\lambda = \lambda_j - \lambda_i$, where $j > i$. As described in Section 2.1.1, because the identification of spectral details in the derivative spectra depends on the selection of the band separation, it was tested the sensitivity of cluster results to the choice of BS. It was also tested the sensitivity of cluster results to different values of the smoothing filter window (WS) used prior to computation of derivative spectra. The sensitivity analysis over a range of BS and WS values allowed to achieve the best compromise between the ability to resolve fine spectral details and the reduction of noise effects in the second derivative spectra. As discussed in Section 3.3.4, the optimal values of BS and WS chosen in this study for the derivative analysis of absorption data are 9 nm. For the derivative analysis of reflectance data, these values are 27 nm. Therefore, although the derivative calculations were made using data from the spectral range 300–725 nm, the results from the derivative analysis for absorption will be reported between the wavelengths of $\lambda_{min} = 309\,nm$ ($\equiv 300 + 9$) and $\lambda_{max} = 716\,nm$ ($\equiv 725 - 9$). For the reflectance derivative, the results will be reported between $\lambda_{min} = 327\,nm$ ($\equiv 300 + 27$) and $\lambda_{max} = 698\,nm$ ($\equiv 725 - 27$).

With regard to the analysis of remote-sensing reflectance, the cluster analysis was also applied to the multispectral reflectance data obtained from in situ SPMR measurements at several discrete wavelengths (and described in Section 3.1.1.3). The

3-element objects $\{R_{rs}(443)/R_{rs}(554), R_{rs}(490)/R_{rs}(554), R_{rs}(510)/R_{rs}(554)\}$ were examined, which consist of band ratios that are similar to those used in current research based on satellite ocean color observations such as the Sea-viewing Wide Field-of-View Sensor (SeaWiFS; O'Reilly *et al.*, 2000). The vectors consisting of 13 band ratios of remote-sensing reflectance with $R_{rs}(554)$ in the denominator, as determined from SPMR measurements at 13 wavebands, were also examined.

The HCA method, schematically presented in Fig. 3.1, was applied using the above described pigment and optical data vectors as input objects. This method utilizes the unsupervised classification algorithm defined in the previous chapter (see Section 2.1.2). Each cluster tree is obtained using a linkage algorithm (i.e., shortest distance) based on initial calculations of the pairwise distance (i.e., cosine angular distance) between all objects included in the input data set.

As a novelty in this study, in order to automatically evaluate the usefulness of optical data for discriminating phytoplankton pigment assemblages, the dendrograms obtained for the different spectral optical data were compared with a reference dendrogram obtained using the pigment composition data (see final step in Fig. 3.1). The pigment information has been commonly used by the scientic community as a proxy for the phytoplankton composition. For this evaluation, the use of two objective criteria of cluster similarity was proposed, which allowed to go beyond the subjective validation of performance made by visual inspection of the resulting cluster partitions. An example is shown in Fig. 3.7, illustrating how the similarity between two given dendrograms is computed using the cophenetic index and the Rand index:

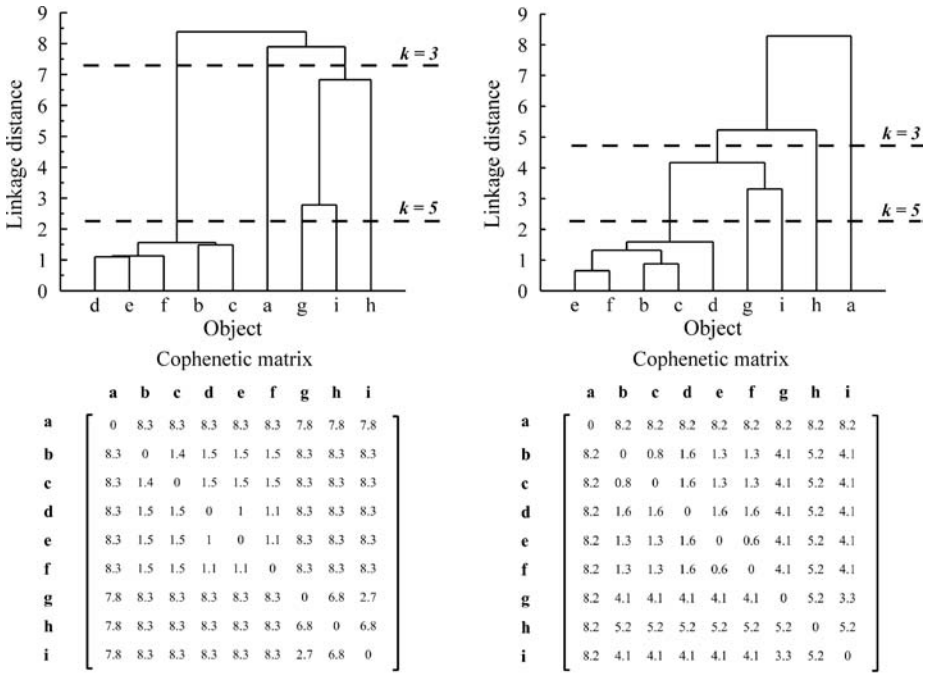
- The **cophenetic index** (Sokal and Rohlf, 1962), r_C, is a measure of how precisely two dendrograms preserve the pairwise distances between data objects. This index is computed from the cophenetic matrix (C) associated with each dendrogram. The elements of a cophenetic matrix ($c_{i,j}$) encode the distance between two objects (i, j), representing in the dendrogram the height of the link at which those two objects are first joined. This height is the distance between the two clusters that are merged by this link. The cophenetic index r_C represents the correlation between two cophenetic matrices (C_1 and C_2):

$$r_C = \frac{\sum_i \sum_j \left(c_{1_{i,j}} - \bar{c_1}\right)\left(c_{2_{i,j}} - \bar{c_2}\right)}{\sqrt{\left(\sum_i \sum_j \left(c_{1_{i,j}} - \bar{c_1}\right)^2\right)\left(\sum_i \sum_j \left(c_{2_{i,j}} - \bar{c_2}\right)^2\right)}} \tag{3.1}$$

  where $\bar{c_1}$ and $\bar{c_2}$ are the mean values of the elements of the matrices C_1 and C_2.

- The **Rand index** (Rand, 1971), r_R, provides a measure of the similarity between two hierarchical dendrograms in terms of the proportion of pairs of objects whose relationship is the same in both dendrograms. The r_R value of 1 means that all pairs of objects are clustered in the same way in both dendrograms. Note that this index has to be computed using all dendrograms cut horizontally at a level (i.e., at a specific linkage distance) which yields the optimal number of clusters (k). Otherwise, r_R would always provide a proportion of 100% because

a complete dendrogram always decomposes the input data all the way through the lowest level (i.e., until the branches consist only of single objects). Detecting natural groupings in the dendrogram and selecting the optimal number of clusters is performed by analyzing a diagram of the increasing linkage distances along the dendrogram. Based on the points at which the linkage distances between the objects change abruptly (which is associated with a steep increase of the within cluster variance), the optimal number of clusters k is determined and all objects located below the point where the hierarchical tree is cut off are assigned to a single cluster (Salvador and Chan, 2004).



Cophenetic matrix

	a	b	c	d	e	f	g	h	i
a	0	8.3	8.3	8.3	8.3	8.3	7.8	7.8	7.8
b	8.3	0	1.4	1.5	1.5	1.5	8.3	8.3	8.3
c	8.3	1.4	0	1.5	1.5	1.5	8.3	8.3	8.3
d	8.3	1.5	1.5	0	1	1.1	8.3	8.3	8.3
e	8.3	1.5	1.5	1	0	1.1	8.3	8.3	8.3
f	8.3	1.5	1.5	1.1	1.1	0	8.3	8.3	8.3
g	7.8	8.3	8.3	8.3	8.3	8.3	0	6.8	2.7
h	7.8	8.3	8.3	8.3	8.3	8.3	6.8	0	6.8
i	7.8	8.3	8.3	8.3	8.3	8.3	2.7	6.8	0

Cophenetic matrix

	a	b	c	d	e	f	g	h	i
a	0	8.2	8.2	8.2	8.2	8.2	8.2	8.2	8.2
b	8.2	0	0.8	1.6	1.3	1.3	4.1	5.2	4.1
c	8.2	0.8	0	1.6	1.3	1.3	4.1	5.2	4.1
d	8.2	1.6	1.6	0	1.6	1.6	4.1	5.2	4.1
e	8.2	1.3	1.3	1.6	0	0.6	4.1	5.2	4.1
f	8.2	1.3	1.3	1.6	0.6	0	4.1	5.2	4.1
g	8.2	4.1	4.1	4.1	4.1	4.1	0	5.2	3.3
h	8.2	5.2	5.2	5.2	5.2	5.2	5.2	0	5.2
i	8.2	4.1	4.1	4.1	4.1	4.1	3.3	5.2	0

Cophenetic index: $r_C = 0.844$

Rand index

Optimal number of cluster: $k = 3$

Object - pairs	Total
Together in both trees	12
Separated in both trees	12
Mixed	12

$r_R = 0.66$

Optimal number of cluster: $k = 5$

Object - pairs	Total
Together in both trees	18
Separated in both trees	18
Mixed	0

$r_R = 1$

Figure 3.7: Example of similarity between two hierarchical cluster trees, computed using the cophenetic and Rand indices. In the latter, an optimal number of clusters (k) needs to be pre-defined in the cluster tree.

3.3 Results and discussion

3.3.1 Classification of stations based on pigment composition

For the 9 stations selected in the study, the estimate of the TChl*a* concentration ranges from about 0.11 $mg\,m^{-3}$ at the southernmost station 59 in the open ocean off the coast of Namibia to 0.62 $mg\,m^{-3}$ at the northernmost station 1 in the English Channel (Fig. 3.2). The variability in pigment composition for the 9 stations is summarized in Table 3.1, which provides the ratios of the concentration of several dominant pigments, measured as described in Section 3.1.1.1, to TChl*a*. Apart from MVChl*a*, which is a principal pigment common to all phytoplankton, the second most important pigment at different stations was either DVChl*a*, zeaxanthin (Zea), 19' - hexanoyloxyfucoxanthin (Hex), or fucoxanthin (Fuco). Table 3.1 also identifies the two dominant pigments (excluding MVChl*a*) which yield the highest ratio to TChl*a* at each station. The values for the ratios of the two dominant pigments to TChl*a* were used as a basis for selecting the 9 stations. As these stations represent different pigment compositions, a class label A, B, C, D, E, or F was assigned to each station.

Station ID	2 dominant pigments	Class	TChl*a* [mg/m^3]	Ratio of concentrations of dominant pigments to TChl*a*					
1	Fuco ≈ MVChl*b*	A	0.62	MVChl*a*	Fuco	MVChl*b*	Chl*c2*	Pra	Hex
				0.95	0.22	0.21	0.11	0.07	0.07
6	DVChl*a* > Zea	B	0.28	DVChl*a*	MVChl*a*	Zea	Hex	*α*-caro	Chl*c2*
				0.50	0.47	0.31	0.18	0.09	0.09
12	DVChl*a* ≈ Zea	C1	0.14	MVChl*a*	DVChl*a*	Zea	Hex	Chl*c2*	But
				0.54	0.44	0.40	0.24	0.09	0.09
37	DVChl*a* ≈ Zea	C2	0.15	Zea	DVChl*a*	MVChl*a*	Hex	Dia	*α*-caro
				0.63	0.58	0.40	0.17	0.10	0.10
44	DVChl*a* ≈ Zea	C3	0.22	DVChl*a*	Zea	MVChl*a*	Hex	*α*-caro	Chl*c2*
				0.56	0.50	0.42	0.17	0.08	0.08
46	DVChl*a* ≈ Zea	C4	0.14	Zea	DVChl*a*	MVChl*a*	Hex	Dia	MVChl*b*
				0.52	0.50	0.49	0.20	0.09	0.08
48	Hex > Zea	D	0.21	MVChl*a*	Hex	Zea	DVChl*a*	But	Dia
				0.76	0.48	0.25	0.21	0.18	0.16
51	Hex > Fuco	E	0.26	MVChl*a*	Hex	Fuco	Dia	Chl*c2*	But
				0.86	0.35	0.24	0.23	0.18	0.13
59	Zea ≈ Hex	F	0.11	MVChl*a*	Zea	Hex	Dia	DVChl*a*	But
				0.80	0.39	0.34	0.21	0.18	0.12

Table 3.1: Stations sorted into different classes characterized by differing pigment assemblages based upon the ratios of the concentrations of two dominant accessory pigments to TChl*a* (see the 2nd and 3rd columns from the left). The ratios of six dominant pigments to TChl*a* are also displayed, with the two most dominant accessory pigments indicated within the shaded areas. Pigment abbreviations are: MVChl*a* = monovinyl chlorophyll-*a*, DVChl*a* = divinyl chlorophyll-*a*, Fuco = fucoxanthin, Hex = 19' - hexanoyloxyfucoxanthin, But = 19' - butanoloxyfucoxanthin, MVChl*b* = monovinyl chlorophyll-*b*, Chl*c*2 = chlorophyll-*c*2, Zea = zeaxanthin, Pra = prasinoxanthin, Dia = diadinoxanthin, and α - caro = α - carotene.

Most stations visited during the cruise within the tropical and subtropical regions of the Atlantic were dominated by Zea and DVChl*a*, but the relative predominance of these two pigments varied between the stations. These pigments are diagnostic of picophytoplankton that includes DVChl*a*- and Zea-containing prochlorophytes and Zea-containing cyanobacteria (mainly Synechococcus in the open ocean waters). Five stations (6, 12, 37, 44, and 46) were selected to represent this type of pigment assemblage. Note that 4 stations dominated by Zea and DVChl*a* with fairly similar ratios DVChl*a*/TChl*a* and Zea/TChl*a* are grouped within the same class C with labels C1, C2, C3, and C4. The station 6, where DVChl*a*/TChl*a* is significantly higher than Zea/TChl*a*, is considered as a separate class B. This station (or class B) is dominated by prochlorophytes as DVChl*a* is an unambiguous marker of this group. The class C stations also show significant role of prochlorophytes. However, this class exhibits a relatively higher contribution of Zea than class B, which is likely indicative of an increased role of cyanobacteria.

Fuco and MVChl*b* are the predominant accessory pigments at station 1 (class A). As these two pigments are not confined to one phytoplankton class, this station could have been dominated by Fuco-rich diatoms, haptophytes, and/or dinoflagellates, as well as MVChl*b*-rich prasinophytes and/or chlorophytes. The predominant accessory pigment at stations 48 (class D) and 51 (class E) is Hex, which suggests that haptophytes and/or chrysophytes are major phytoplankton groups at these locations. These stations are designated as different classes because they clearly differ in accessory pigments that follow Hex in ranking. Zea and Fuco are the second most important diagnostic pigments at stations 48 (class D) and 51 (class E), respectively. Finally, station 59 (class F) also shows a significant role of Hex-rich phytoplankton although Zea is the most important diagnostic pigment at this location, indicating potential significance of cyanobacteria and/or prochlorophytes.

Fig. 3.8a shows the hierarchical cluster tree obtained for the input data consisting of the ratios of concentrations of 24 individual pigments to TChl*a* at each station. Stations displaying a similar pigment composition (e.g., stations from class C and B) are linked by a shorter linkage distance (y-axis). In this sense, the optimal number of clusters (k) is derived from a diagram of linkage distances along the dendrogram (Fig. 3.8b). The first steep increase in the linkage distance observed in this diagram, which is associated with an increase of the within cluster variance, suggests an optimal partitioning of the pigment data into 5 clusters. The linkage distance of 0.023 can be selected to characterize this steep increase in variance (see dashed lines in Fig. 3.8a and b). For the dendrogram cut at a level of linkage distance of 0.023, all clusters are single object (i.e., single station) clusters, except for a multi-object cluster that includes stations C1, C2, C3, C4, and B. The results of this cluster analysis are quite consistent with the preliminary classification obtained by considering just two dominant diagnostic pigments (see Table 3.1). Note that stations C1, C2, C3, C4, and B are all characterized by relatively high ratios of DVChl*a* and Zea to TChl*a*. Some differences between these stations in terms of the relative roles of DVChl*a* and Zea do not, however, produce significant distances between the corresponding pigment data vectors and hence these 5 stations are grouped into a single cluster.

The dendrogram also indicates that the stations classified as A, D, E, and F display significant dissimilarities between each other and when compared to the stations classified as B and C. Note that the stations 48 (class D) and 59 (class F) have the Hex and Zea as dominant diagnostic pigments, albeit in reverse ranking (see Table 3.1), so these stations appear closer to one another in the dendrogram (Fig. 3.8a).

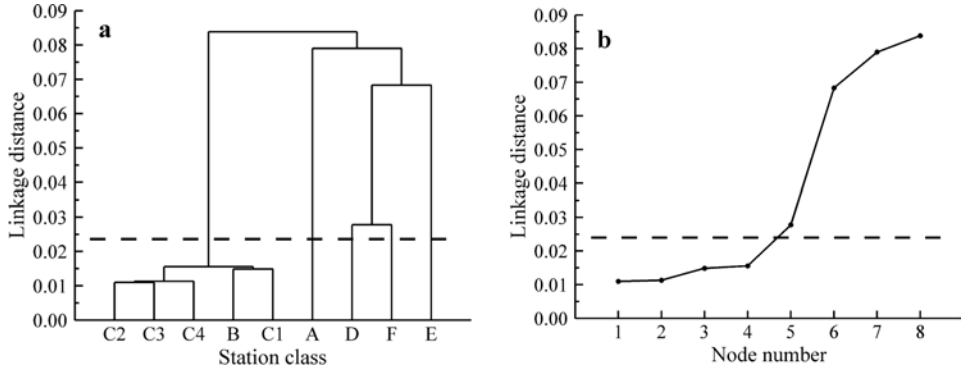


Figure 3.8: (a) Dendrogram obtained for the nine stations using 24 pigments to total chlorophyll-a (TChl$a$) ratios determined from the CHORS HPLC analysis. (b) Linkage distances obtained from the cluster analysis shown in (a) as a function of distance along the dendrogram.

In the following analysis of optical data as a means for assessing differences in pigment assemblages, the pigment-based cluster partitioning obtained with the 24 pigment ratios (as shown in Fig. 3.8a) is used as a reference.

3.3.2 Classification of stations based on absorption spectra

Fig. 3.9 shows hyperspectral data of absorption coefficients first normalized at 555 nm and then divided by the total chlorophyll-a concentration for the nine stations. These spectra are referred to as the spectral chlorophyll (Chl)-specific normalized absorption coefficients. Specifically, the Chl-specific normalized coefficients for the total absorption, $a_n^*(\lambda)$, the absorption of pure seawater plus phytoplankton, $a_{n,w+ph}^*(\lambda)$, and the absorption of phytoplankton alone, $a_{n,ph}^*(\lambda)$ were examined.

The differences in the shape of phytoplankton absorption in the UV and blue spectral regions are generally quite large between most stations (Fig. 3.9a). With the addition of the pure water contribution, differences in the spectral shape of $a_{n,w+ph}^*(\lambda)$ continue to be seen but are considerably smaller (Fig. 3.9b). Finally, upon further addition of the contributions associated with non-phytoplankton particles and CDOM, the spectral shape of total absorption again shows larger differences between the stations at wavelengths shorter than the normalization point at 555 nm (Fig. 3.9c). From the visual inspection of these plots it is, however, difficult to deduce to what extent the observed differences might be consistent with the classification of stations based on pigment composition.

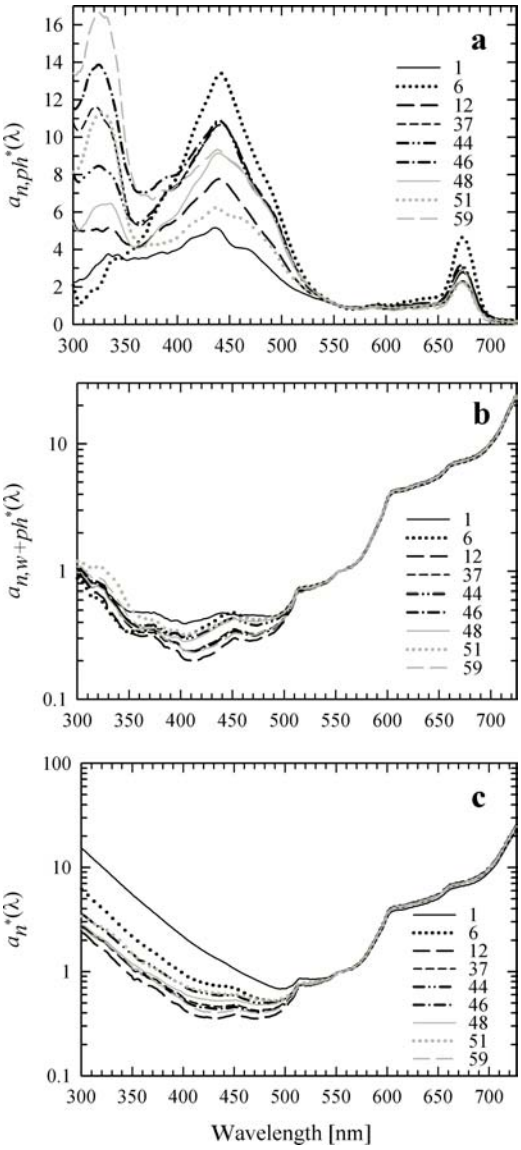


Figure 3.9: Chlorophyll-specific normalized absorption coefficients for the nine stations corresponding to: (a) phytoplankton, $a^*_{n,ph}(\lambda)$, (b) pure seawater plus phytoplankton, $a^*_{n,w+ph}(\lambda)$, and (c) total absorption, $a^*_n(\lambda)$.

Figure 3.10 illustrates the results from cluster analysis applied to the absorption spectra presented in Fig. 3.9 and the corresponding second derivative spectra over the entire spectral range from 300 nm to 725 nm. The derivative spectra of absorption were calculated with the parameters WS and BS equal to 9 nm as determined by the sensitivity analysis discussed in Section 3.3.4. In nearly all cases (Fig. 3.10a–e), the absorption-based cluster trees differ significantly from the reference pigment-based cluster tree shown in Fig. 3.8a. Thus, the full hyperspectral data of $a^*_n(\lambda)$, $a^*_{n,w+ph}(\lambda)$,

and $a^*_{n,ph}(\lambda)$ as well as the second derivative spectra of $a^*_n(\lambda)$ and $a^*_{n,w+ph}(\lambda)$ do not provide useful information for discriminating the differences in pigment assemblages at the examined stations. The only case in which stations are classified within the dendrogram in a similar way to the pigment-based cluster tree is when the second derivative of phytoplankton absorption spectra is considered (Fig. 3.10f). When this absorption-based dendrogram is cut horizontally at a level of linkage distance of 0.023 that yields 5 clusters, the same stations are grouped in separate clusters as in the pigment-based cluster tree. The result obtained from this study supports the potential usefulness of a method which accounts for the complete spectral behavior of the second derivative of $a^*_{n,ph}(\lambda)$ spectra for discriminating different pigment assemblages.

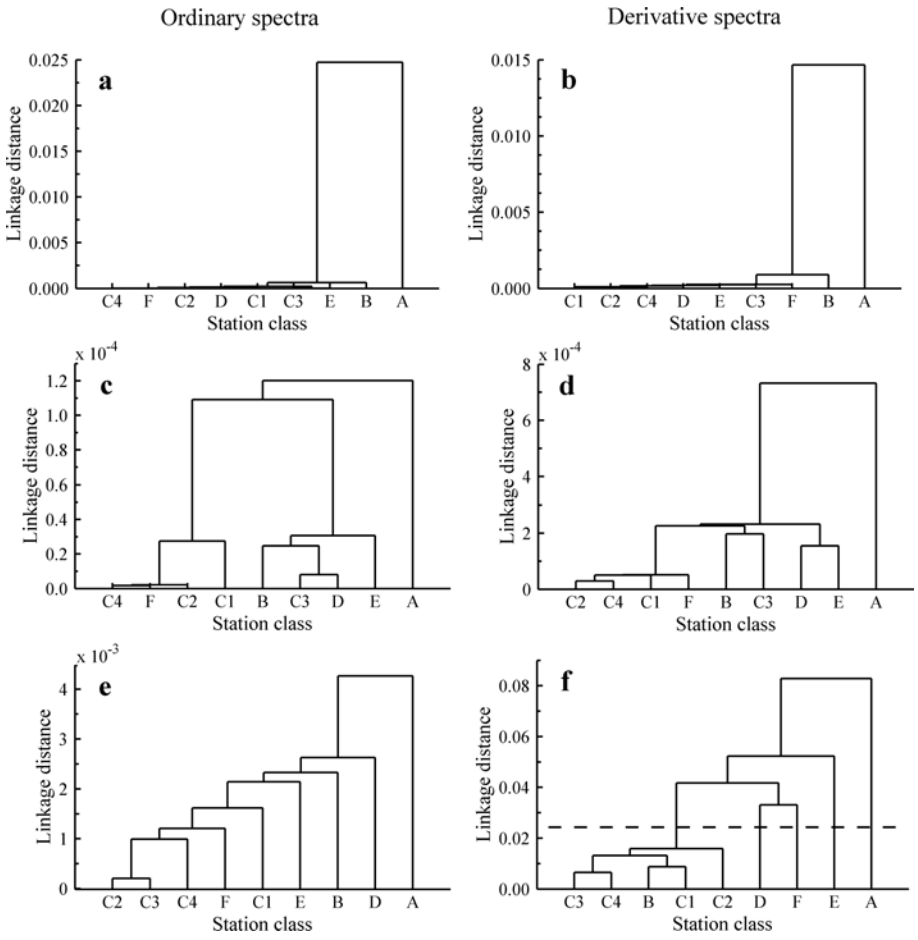


Figure 3.10: Results of cluster analysis applied to absorption data from the nine stations. The left panels represent dendrograms obtained using: (a) $a^*_n(\lambda)$, (c) $a^*_{n,w+ph}(\lambda)$, and (e) $a^*_{n,ph}(\lambda)$, and the right panels (b, d, and f) illustrate results obtained using each component's second derivative spectrum.

In the analysis above it was considered a spectral range from 300 nm to 725 nm, which is much broader than the spectral region where specific absorption imprints caused by accessory pigments occur. It was therefore considered useful to examine whether the cluster analysis of absorption data yielded similarity with pigment-based clusters if different, narrower spectral ranges were considered.

Within the framework of this thesis, a new graphical representation was created in order to display the results from the sensitivity of cluster analysis to the selection of the sepctral range. In each graph, the distribution of values for the cophenetic or Rand index is shown as a function of the spectral range considered, with the lower limit of the spectral range, λ_{min}, displayed along the y-axis (ordinate) and the upper limit, λ_{max}, along the x-axis (abscissa). The similarity indices are thus shown for many spectral ranges represented by many combinations of λ_{min} and λ_{max}. The higher values of indices, depicted by darker areas in the graphs, correspond to better similarity between a given optical-based cluster tree and pigment-based tree. The best degree of similarity is obtained when the indices are close to 1, indicated by the nearly black areas in the graphs.

Fig. 3.11 illustrates the degree of similarity between the absorption-based and pigment-based cluster trees for different spectral ranges of absorption data. The degree of similarity is shown in terms of both cophenetic and Rand indices. The three absorption spectra, $a_n^*(\lambda)$, $a_{n,w+ph}^*(\lambda)$, and $a_{n,ph}^*(\lambda)$, are considered in this analysis. According to the distributions of cophenetic index, a maximum degree of similarity (i.e., cophenetic index = 1) is obtained when the phytoplankton absorption spectrum $a_{n,ph}^*(\lambda)$ is analyzed over the spectral range approximately from λ_{min} =425 nm to λ_{max} =540 nm (Fig. 3.11e). The distribution of the Rand index indicates that the best similarity between the $a_{n,ph}^*(\lambda)$-based cluster tree and the pigment-based tree occurs within a broader spectral region, approximately between λ_{min} =390 nm and λ_{max} =610 nm (Fig. 3.11f). These optimal spectral regions generally overlap with the wavelength range where absorption characteristics of main accessory pigments appear (e.g., Fig. 1.3, Bricaud *et al.*, 2004). The remaining results in Fig. 3.11 (panels a, b, c, and d) show generally poor similarity between the $a_n^*(\lambda)$ or $a_{n,w+ph}^*(\lambda)$ data and pigment composition, regardless of the spectral range considered.

Fig. 3.12 depicts similar results, but for the similarity between pigment composition and the second derivative spectra of $a_n^*(\lambda)$, $a_{n,w+ph}^*(\lambda)$, and $a_{n,ph}^*(\lambda)$. The use of derivative spectra generally improves the similarity as indicated by the presence of darker areas or the larger extent of dark areas in the distributions of the cophenetic and Rand indices. For example, compared to the results for the ordinary spectra of $a_{n,w+ph}^*(\lambda)$ in Fig. 3.11d, a significant increase in the Rand index is observed for the second derivative spectra of $a_{n,w+ph}^*(\lambda)$ within the spectral range from λ_{min} = 440 nm to λ_{max} = 650 nm (Fig. 3.12d). The improvement is even more striking for the results involving the second derivative spectra of phytoplankton absorption (Fig. 3.12e, f). The spectral regions where the cophenetic and Rand indices assume high values near or equal to 1 are much larger compared with the analysis of ordinary (non-differentiated) spectra of $a_{n,ph}^*(\lambda)$. The darkest areas in the distributions of the indices in Fig. 3.12e, f cover a broad spectral range, approximately from λ_{min} = 370

nm to $\lambda_{max} = 716$ nm. This result indicates that a high degree of similarity between the second derivative spectra of $a^*_{n,ph}(\lambda)$ and pigment composition can be obtained for many different combinations of spectral ranges of absorption data (i.e., different combinations of λ_{min} and λ_{max}).

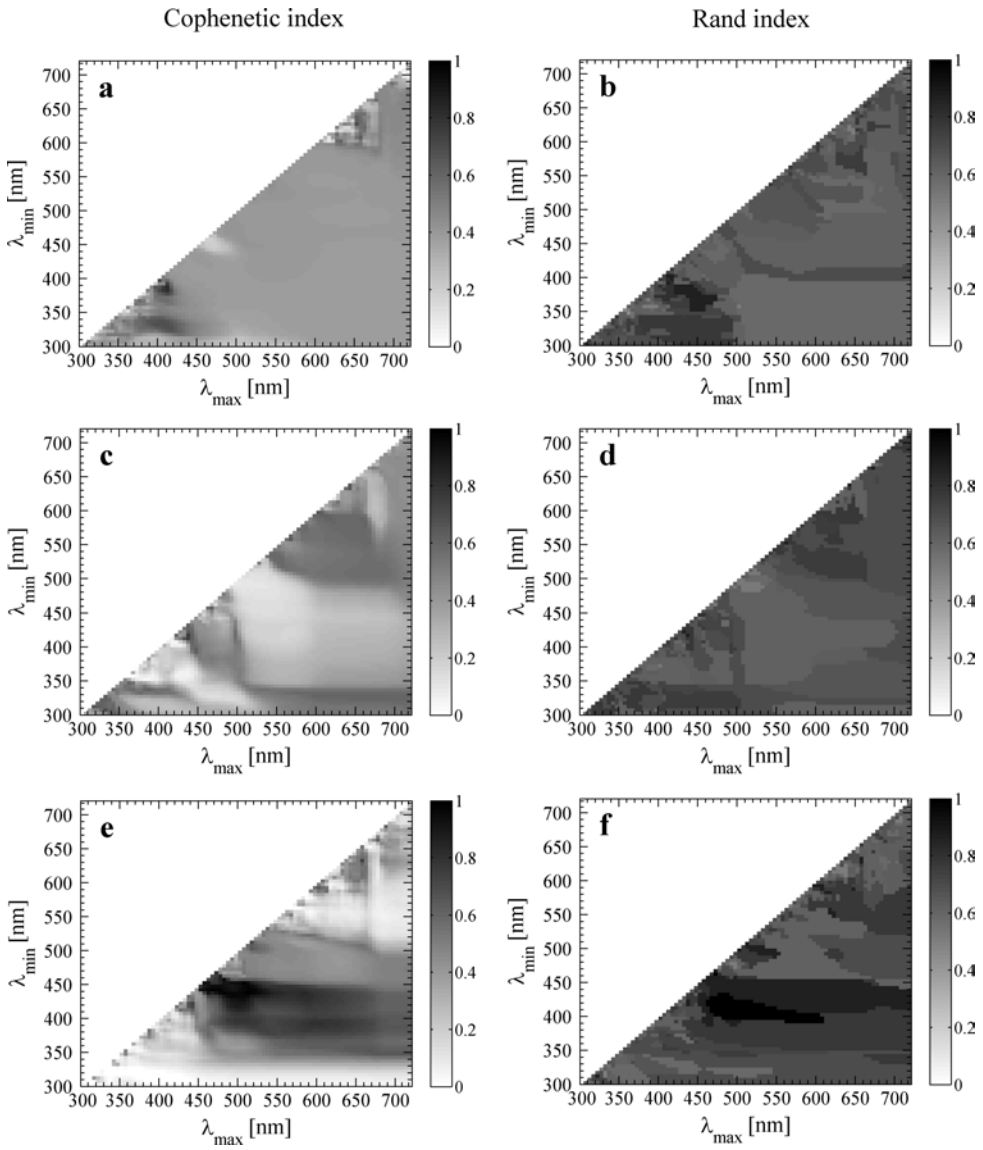


Figure 3.11: Similarity indices between absorption-based and pigment-based cluster trees obtained for the 9 stations using different combinations of spectral range for (a, b) $a^*_n(\lambda)$, (c, d) $a^*_{n,w+ph}(\lambda)$ and (e, f) $a^*_{n,ph}(\lambda)$. The y-axis indicates the lower limit of the spectral range (λ_{min}) and the x-axis the upper limit of the spectral range (λ_{max}) used in the cluster analysis. Left and right panels depict the cophenetic and Rand indices, respectively.

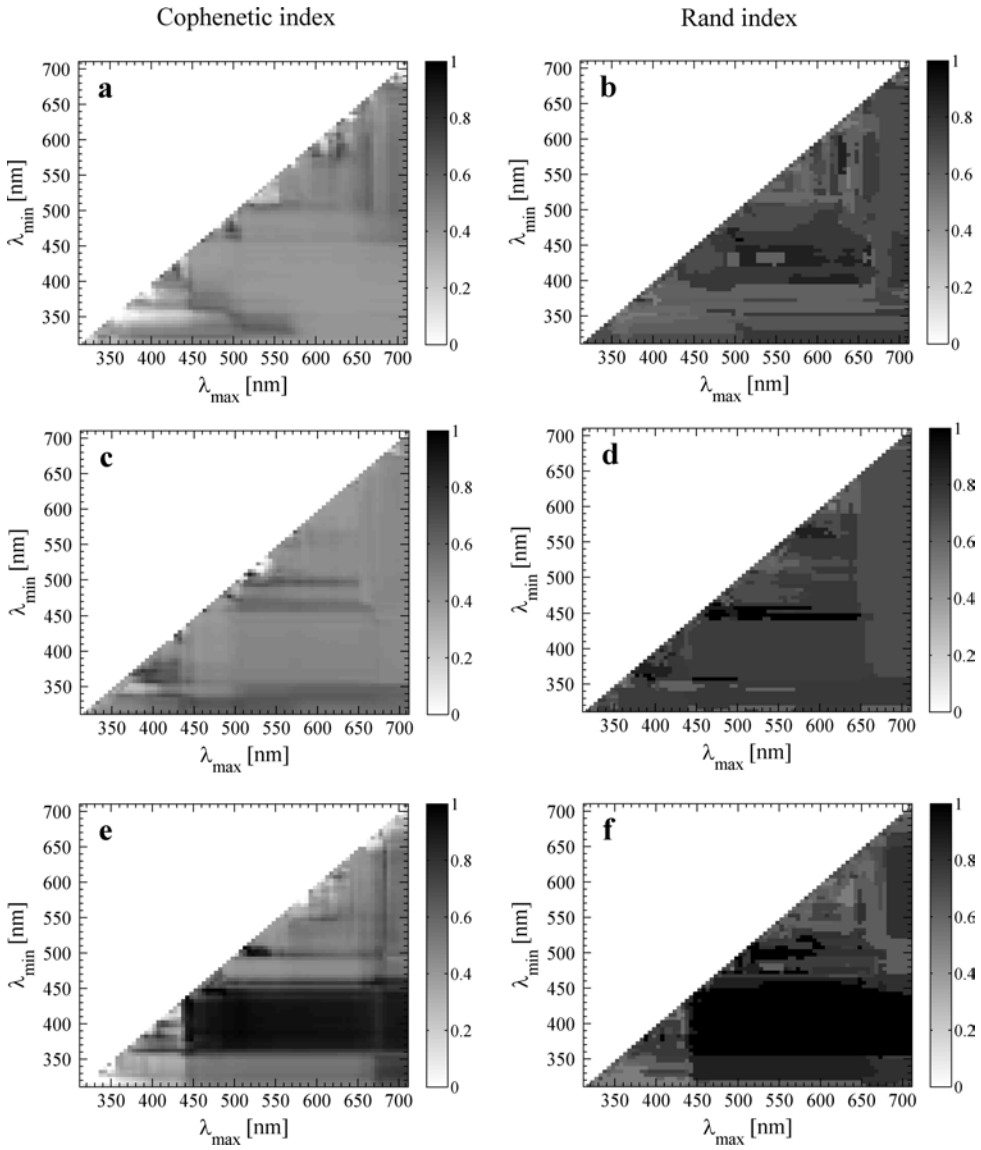


Figure 3.12: Similar to Fig. 3.11, but based on cluster trees obtained using different spectral range combinations for the second derivative spectra of (a, b) $a_n^*(\lambda)$, (c, d) $a_{n,w+ph}^*(\lambda)$ and (e, f) $a_{n,ph}^*(\lambda)$. Optimal values for band separation and window size determined from prior analyses (BS=WS=9 nm, see Section 3.3.4) were used in the calculation of derivative spectra.

Because the cophenetic index does not require a priori selection of the optimal number of clusters, this index facilitates comparison of the results for ordinary and derivative spectra with pigment-based cluster trees shown in Figs. 3.11e and 3.12e. The improved similarity obtained by utilization of the second derivative spectra of

$a^*_{n,ph}(\lambda)$ was evidenced as a marked increase in the values of the cophenetic index when computed for all possible combinations of spectral ranges. These increased values were associated with a broader overall spectral region, or equivalently a larger number of spectral ranges, for which the index exceeded 0.9 (see Fig. 3.13).

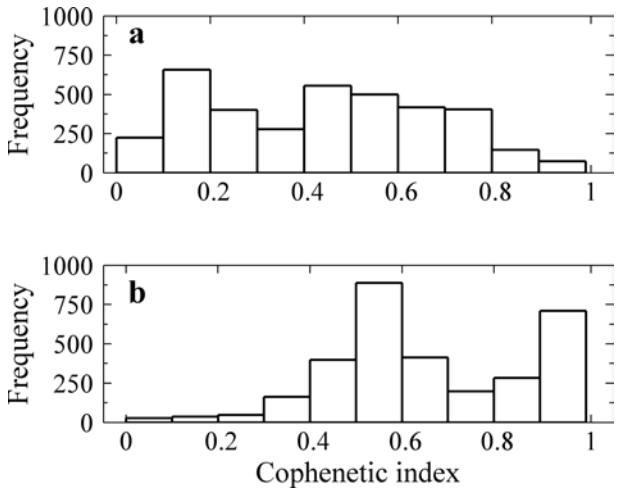


Figure 3.13: Histograms of cophenetic indices obtained for all combinations of spectral ranges shown in Fig. 3.11e and 3.12e based on (a) the absorption of phytoplankton, $a^*_{n,ph}(\lambda)$, and (b) its second derivative spectra.

3.3.3 Classification of stations based on remote-sensing reflectance

The relationship between the spectral remote-sensing reflectance, $R_{rs}(\lambda)$, of the ocean and phytoplankton pigment composition is less direct and far more complicated than that for the spectral phytoplankton absorption coefficient, mainly due to the presence of many optically significant non-phytoplankton constituents in seawater. The investigation of $R_{rs}(\lambda)$ is of particular interest, however, because information contained in this measurement provides a potential means for remote-sensing applications. Fig. 3.14 shows the Hydrolight-simulated $R_{rs}(\lambda)$ spectra normalized at 555 nm for the nine stations identified as classes from A (station 1) through F (station 59). In general, there are significant differences in the UV and blue spectral regions between these normalized spectra, with the largest contrast between the classes A (station 1) and C1 (station 12).

The dendrograms obtained from cluster analysis as applied to four different sets of input data vectors containing information about remote-sensing reflectance (as described in Section 3.2) are displayed in Fig. 3.15. The limited spectral information, i.e., the three reflectance band ratios used commonly in satellite ocean color applications (Fig. 3.15a) and the 13 band ratios corresponding to multispectral measurements with the SPMR instrument (Fig. 3.15b), provide a very dissimilar

classification of stations compared with the pigment-based cluster analysis (see Fig. 3.8a). All stations, with only the exception of class A, show very little separation in the cluster-tree based on multispectral reflectance data. The high spectral resolution (1 nm) normalized reflectance spectra over the entire spectral range 300– 725 nm also produce a dendrogram (Fig. 3.15c) that is very different from the pigment-based cluster tree. Although the stations belonging to class C are closer to one another compared with the multispectral-based cluster tree, they are grouped together with two other stations (classes E and D). In addition, station B forms a separate single-object cluster in Fig. 3.15c, whereas it is grouped in a single multi-object cluster together with the stations from class C in the pigment analysis. The only case when the cluster analysis of reflectance data provides a high degree of similarity with pigment analysis (i.e., Rand index of 0.78) is for the second derivative of hyperspectral normalized reflectance over the entire spectral range 327–698 nm (Fig. 3.15d). It is important to note that the derivative reflectance spectra were calculated with the parameters WS and BS of 27 nm (as supported by the sensitivity analysis discussed in Section 3.3.4). The stations A, E, and F in Fig. 3.15d form single-object clusters at a significant distance from the remaining stations. Similarly to pigment analysis, the stations C1, C2, C3, C4, and B are grouped relatively close to one another. However, station D also belongs to that group, which is not the case in the pigment-based cluster tree. This may be attributable to the fact that Zea and DVChl*a*, which are the two most dominant diagnostic pigments at stations C1, C2, C3, C4, and B, also play a significant role at station D where they are ranked as the second and third most important diagnostic pigments (see Table 3.1).

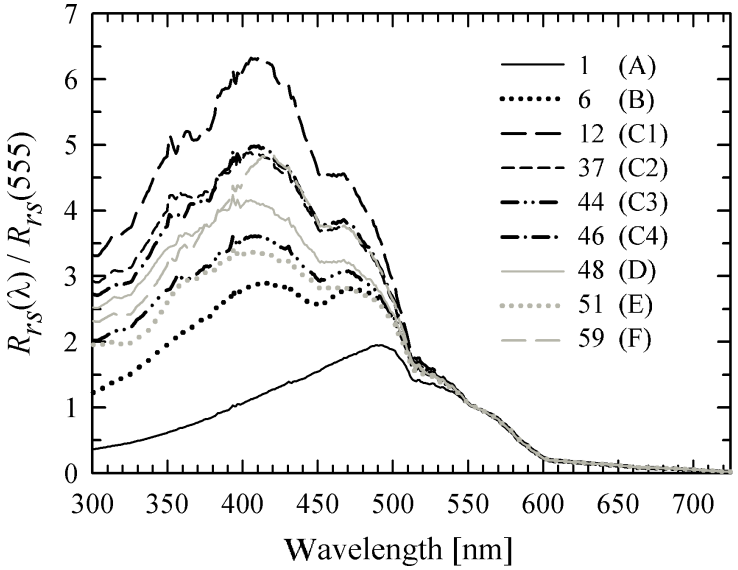


Figure 3.14: Hydrolight-simulated $R_{rs}(\lambda)$ spectra, normalized at 555 nm, computed for the nine stations using measured IOPs as input.

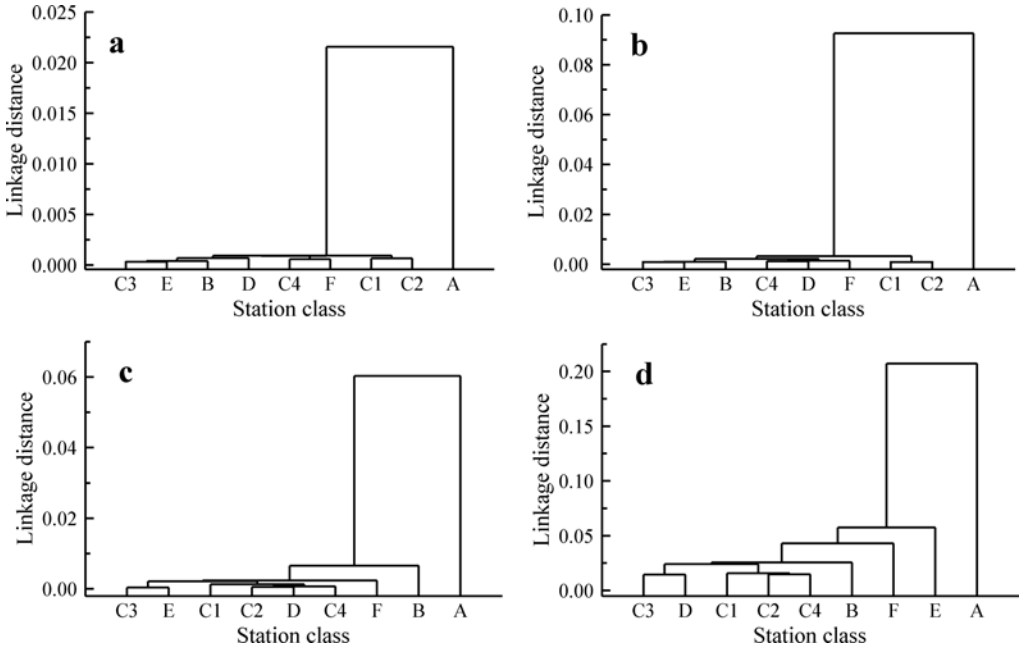


Figure 3.15: Dendrograms resulting from cluster analysis of the nine stations calculated using four different sets of input data vectors: (a) three reflectance band ratios of $R_{rs}(\lambda)$ based on 4 SeaWiFS wavebands obtained from measurements with SPMR instrument, (b) 13 band ratios corresponding to multispectral measurements of $R_{rs}(\lambda)$ with SPMR instrument, (c) hyperspectral (1 nm) ordinary (non-differentiated) normalized $R_{rs}(\lambda)$ spectra computed from the Hydrolight simulations, and (d) second derivative of hyperspectral normalized $R_{rs}(\lambda)$ spectra obtained using optimal values for band separation and smoothing filter window (i.e., $BS = WS =$ 27 nm, see Section 3.3.4).

The progression of linkage distances corresponding to the four dendrograms from Fig. 3.15 clearly illustrates the advantage of the second derivative spectra over the multispectral data or non-differentiated spectra of reflectance (see Fig. 3.16), and indicates that this approach enables better identification of the differences in the magnitude and shape between the high resolution spectra. In contrast, in the analysis of multispectral data and ordinary spectra stations are linked at a very small distance, suggesting that these types of reflectance data will be of little value for obtaining information about pigment assemblages from cluster analysis.

The improvement in the similarity between the pigment-based and reflectance-based classification achieved with the use of second derivative spectra as opposed to multispectral data or ordinary spectra of reflectance is presented in Table 3.2. The results for the derivative analysis of hyperspectral normalized $R_{rs}(\lambda)$ over the entire

spectral range 300–725 nm show a significant increase in both the cophenetic and Rand index when compared with multispectral and ordinary spectral data. However, the best performance is obtained when the derivative analysis is restricted to the spectral range from 435 to 495 nm (as supported by the sensitivity analysis discussed below). In this case, the similarity indices are highest.

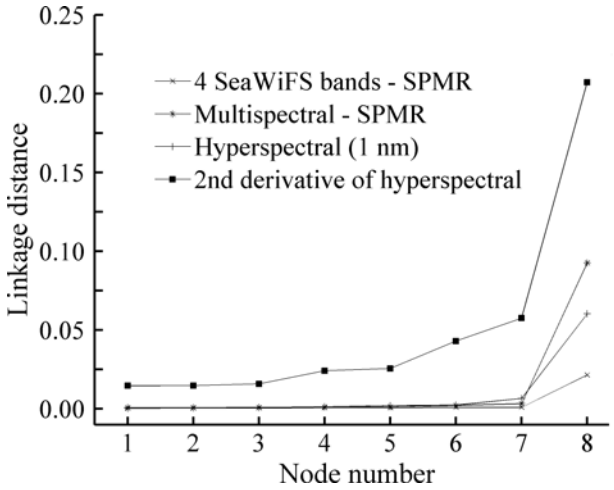


Figure 3.16: Linkage distances as a function of distance along the dendrogram for each cluster tree depicted in Fig.3.15.

Reflectance data	Rand index	Cophenetic index
3 band ratios based on 4 SeaWiFS bands	0.69	0.39
Multispectral (13 bands)	0.69	0.39
Hyperspectral (325 bands, range 300-725 nm)	0.69	0.39
Hyperspectral 2nd derivative (327-698 nm)	0.78	0.53
Hyperspectral 2nd derivative (435-495 nm)	1	0.65

Table 3.2: A comparison of similarity indices between pigment-based clusters and reflectance-based clusters for the different sources of reflectance data that are depicted in Fig. 3.15. For the case of the second derivative of hyperspectral reflectance, the result of computations for two different spectral regions is given.

Fig. 3.17 shows distributions of the cophenetic and Rand index which identify the optimal spectral ranges for the cluster analysis of the second derivative of remote-sensing reflectance. The cophenetic index is generally close to or slightly lower than 0.5 for most spectral ranges examined, that is for most combinations of λ_{min} and λ_{max} (Fig. 3.17a). In the spectral region from 435 nm to 510 nm, this index is about 0.62 and can be considered as an optimal spectral range for the application of derivative approach with potential for good similarity between the pigment-based and reflectance-based cluster trees. This result is also supported by very high value of the Rand index of 0.86 in that spectral range (Fig. 3.17b). The cophenetic and Rand indices attain even higher values of about 0.65 and 1, respectively, within somewhat narrower wavelength range from $\lambda_{min} = 435$ nm to $\lambda_{max} = 495$ nm, which defines an alternative optimal spectral range. In addition, the cophenetic and Rand indices suggest good performance of the derivative-based analysis over a broader spectral regions including shorter and longer wavelengths of λ_{min}. The Rand index is quite high for λ_{min} varying between 350 nm and 450 nm, for example as high as 1 when the spectral range is from $\lambda_{min} = 365$ nm to $\lambda_{max} = 480$ nm. The cophenetic index is also quite high for the spectral ranges from $\lambda_{min} = 327$ nm to $\lambda_{max} = 400$ nm and $\lambda_{min} = 610$ nm to $\lambda_{max} = 670$ nm. However, because the optical roles of different diagnostic pigments in these regions are insignificant or certainly less important than in the blue region, the use of the spectral range 435–510 nm or 435–495 nm, where both the cophenetic and Rand indices are relatively high, appears to be most reasonable.

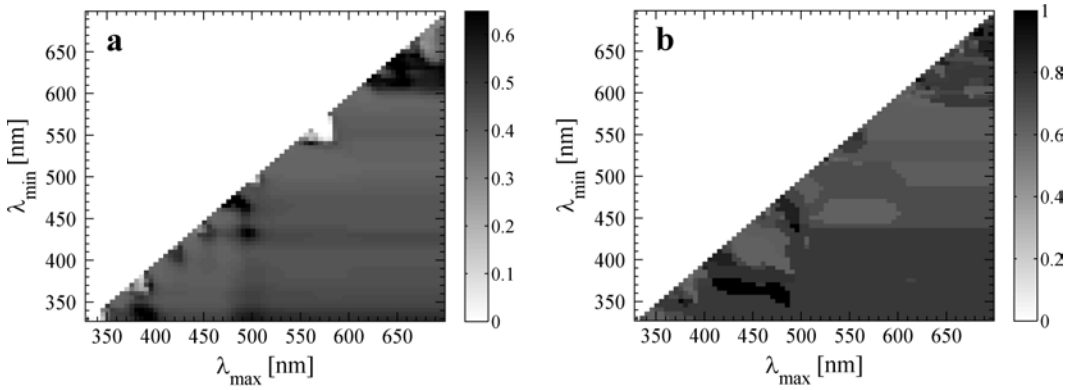


Figure 3.17: Similarity indices between reflectance-based and pigment-based cluster trees obtained for the nine stations using different combinations of spectral ranges for the second derivative of the hyperspectral normalized $R_{rs}(\lambda)$. Optimal values of $BS = WS = 27$ nm, determined from prior analyses, were used for the calculation of derivative spectra. The y-axis indicates the lower limit of the spectral range (λ_{min}) and the x-axis the upper limit of the spectral range (λ_{max}) utilized in the cluster analysis. Panels (a) and (b) depict the cophenetic and Rand indices, respectively.

3.3.4 Sensitivity analysis and determination of optimal parameters for derivative analysis

Although the previous results illustrate the advantages of derivative analysis, such analysis can be highly sensitive to parameters chosen for the calculation of derivative spectra, specifically the size of the filter window (WS) used in the spectral smoothing of the ordinary spectra and the band separation (BS) used in the calculation of derivatives. In order to display the results from the sensitivity of cluster analysis to the selection of these parameters, a new graphical representation was also used in this study. The distribution of values of the cophenetic and Rand index is shown as a function of the band separation (BS) and smoothing filter window (WS) considered. In particular, it was computed the distribution of cophenetic index between the pigment-based and absorption-based cluster trees using the second derivative spectra of $a^*_{n,ph}(\lambda)$ within the spectral range of 420–515 nm as input (Fig. 3.19a). This spectral range is adequate for this sensitivity analysis because it showed very high values of cophenetic and Rand indices in Figs. 3.11e,f and 3.12e,f. The distribution of cophenetic index in Fig. 3.19a is shown with the smoothing parameter WS varying from 1 to 29 consecutive samples (with a step of 2 samples) along the y-axis and the band separation parameter BS varying also from 1 to 29 samples with a step of 2 samples along the x-axis. Note that the number of consecutive samples is equivalent to the wavelength interval in nanometers because the spectral data (samples) have the resolution of 1 nm. The highest values of the cophenetic index are obtained for intermediate values of WS and BS around 9 nm–10 nm, and thus are the optimal values for the derivative analysis of the absorption spectra. The corresponding derivative spectra of $a^*_{n,ph}(\lambda)$ are presented in Fig. 3.18a. This result is consistent with the general expectation that if the values of WS and BS are too small, the derivative spectra are sensitive to noise and exhibit false spectral features, and on the other hand if the WS and BS are too large, the real significant spectral features are smoothed over and essentially removed from the analysis. As the best compromise, the cluster analyses of derivative spectra presented in Fig. 3.10 were obtained with WS and BS of 9 nm.

A similar analysis was performed to determine the appropriate sizes of WS and BS for utilization of the second derivative of the reflectance spectra (Fig. 3.19b). These results are shown for the derivative spectra calculated over one of the optimal spectral ranges, specifically 435– 495 nm (Fig. 3.18b), which showed high values for both cophenetic and Rand indices. The best similarity with cophenetic index of about 0.65 is obtained when the calculations of second derivative spectra are made with relatively large values of WS and BS. For example, a very good result is obtained if both WS and BS assume a value of 27 consecutive spectral samples (i.e., 27 nm as the resolution of the hyperspectral reflectance data is 1 nm). It is recalled that this value was used to compute the results pertinent to the derivative reflectance spectra presented in Fig. 3.15d. Similarly, good results are obtained with a smaller WS (~14 nm) and a larger BS (~37 nm), or vice versa. In contrast, if both WS and BS are small (less than about 10 nm) or large (above ~40 nm) the cophenetic

index is reduced significantly. This sensitivity analysis, in agreement with similar analysis for absorption spectra, supports the notion of a strong dependence of the derivative-based cluster trees on the selection of derivative parameters WS and BS.

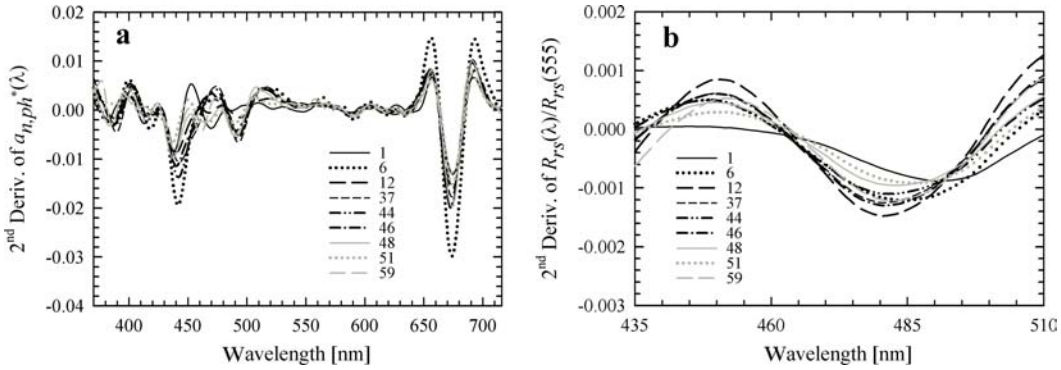


Figure 3.18: The second derivative spectra at each station of the (a) chlorophyll-specific normalized phytoplankton absorption coefficient, $a_{n,ph}^*(\lambda)$, and (b) normalized hyperspectral remote-sensing reflectance, $R_{rs}(\lambda)/R_{rs}(555)$. Optimal values were used for the derivative calculations, i.e., $BS = WS = 9$ nm for the absorption data and $BS = WS = 27$ nm for the reflectance data.

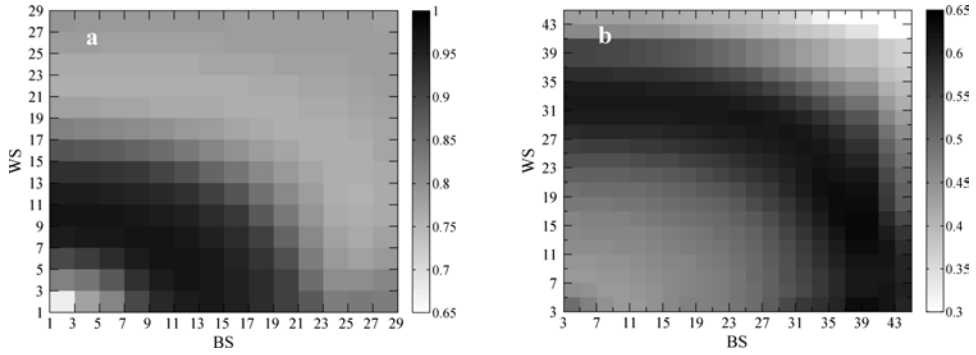


Figure 3.19: Cophenetic indices obtained from similarity analysis between the pigment-based cluster tree with trees obtained using the second derivative of hyperspectral normalized spectra for (a) phytoplankton absorption, $a_{n,ph}^*(\lambda)$, and (b) remote-sensing reflectance, $R_{rs}(\lambda)$, in which different parameter sets for the derivative analysis are considered. The y-axis indicates the size of the filter window used in smoothing of the spectra (WS), and the x-axis represents the band separation used in the calculation of the derivative (BS). The analysis was conducted using the optimal spectral regions of 420 to 515 nm for $a_{n,ph}^*(\lambda)$, and from 435 to 495 nm for $R_{rs}(\lambda)$.

3.4 Summary and conclusions

Through application of an unsupervised hierarchical cluster analysis to pigment and optical data from the eastern Atlantic Ocean in 2005, the potential usefulness of hyperspectral data of absorption coefficients and remote-sensing reflectances and their second derivative spectra for discriminating different phytoplankton pigment assemblages in the open ocean has been demonstrated. The ability to discriminate different phytoplankton pigment assemblages from the hyperspectral-based cluster approach has been optimized by selecting the optimal spectral range and the most suitable parameters used in the spectral derivative calculations (i.e., smoothing filter size and derivative band separation).

The assessment of similarity between the hyperspectral data and phytoplankton pigment composition has been made using a novel validation approach which is based on a cluster algorithm and two similarity indices, cophenetic and Rand. In this chapter, it has been demonstrated that application of these indices to quantify the similarity between different optical-based cluster trees and the pigment-based cluster tree provides a valuable methodology for identifying optical variables and the spectral ranges most suitable for characterizing the phytoplankton pigment assemblages. In this sense, the defined reference based on the pigment composition (the "sea truth") has proven valid as a proxy for phytoplankton community composition.

In order to address the main objectives of this chapter, the following methodology has been applied. First, a detailed description of the field measurements has been provided along with a model-data "closure exercise" for the reconstruction of hyperspectral $R_{rs}(\lambda)$ (Section 3.1). It has been demonstrated that the modeled data set of hyperspectral $R_{rs}(\lambda)$ is relatively consistent in comparison to measured multispectral $R_{rs}(\lambda)$. Regarding the pigment data, it has been shown by comparing two independent HPLC pigment data sets that the results and conclusions of this study were robust and independent of the choice of pigment data used in the creation of the reference pigment dendrogram for the cluster-based similarity analysis of pigment and optical data (see Appendix B). Secondly, a successful cluster analysis based on a small but carefully selected set of stations has been performed as a proof-of-concept study (Section 3.3).

The most promising results from the cluster analysis were obtained with the second derivative spectra of the phytoplankton absorption coefficient, $a^*_{n,ph}(\lambda)$, over the spectral range 370–716 nm (or narrower spectral regions from within that range), and the second derivative spectra of the remote-sensing reflectance, $R_{rs}(\lambda)$, over the spectral range from about 435 nm to 510 nm. In the course of this study, in addition to the second derivative spectra of $a^*_{n,ph}(\lambda)$ and $R_{rs}(\lambda)$, other absorption and reflectance data have been examined but they generally show either more limited value or no usefulness at all for discriminating phytoplankton pigment assemblages. For example, the cluster analysis of the ordinary (non-differentiated) reflectance spectra at 1 nm resolution or multispectal (13 wavebands) reflectance data exhibited very poor similarity with pigment-based clusters. Similar results were obtained for the ordinary spectra of the total absorption coefficient, $a^*_n(\lambda)$. However, the ordinary spectra of

$a^*_{n,ph}(\lambda)$ were observed to be useful, especially within the spectral range 425–540 nm.

As it will be assessed in the next chapter (Chapter 4), further work is needed with larger databases of measurements from various oceanic environments to determine the generality of the presented approach and the specific set of optimal parameters. It is, however, important to note that the proposed methodology is generally applicable to other data sets or other types of data and a similar optimization analysis can also be used to provide the best performance for a given data set. It is also essential to point out that this analysis is flexible and permits different optimal values to be utilized in the case of other optical data with differing spectral resolutions.

The results suggest that the minimum data necessary for application of the cluster analysis is information regarding either the hyperspectral phytoplankton absorption coefficient or the remote-sensing reflectance in the blue to green region of the spectrum, which will yield an initial classification of stations based on similarities and differences in the optical spectra. Without an independent reference as to what these optical classes represent in terms of the phytoplankton assemblage (e.g., pigment information), however, these stations cannot be assigned to a descriptive class. One possible scenario for the utilization of this approach is to apply it to optical data set that includes a reference subset for which both the optical and pigment data are available. First, the pairs of corresponding optical and pigment-based clusters can be created from the reference subset. Then, the remaining optical data can be classified in relation to pigment-based clusters on the basis of similarity with optical data from the reference subset. Another scenario is the application of this approach to optical data sets which are collected with no concurrent determinations of phytoplankton pigments. In this case, one can envision a similar procedure as the one described above but reference information would have to be derived in advance from other extensive concurrent pigment and optical data sets.

3.5 Acknowledgements

The research presented in this chapter was supported by the NASA Biodiversity and Ecological Forecasting Program (Grant NNX09AK17G), the NASA Ocean Biology and Biogeochemistry Program (Grant NNG04GO02G), and the Spanish National Research Council CSIC (projects ANERIS PIF08-015 and HIDRA 2006-301102). Part of this research was performed during a visit of E. Torrecilla at the Marine Physical Laboratory, Scripps Institution of Oceanography (San Diego, USA) supported also by CSIC (Program I3P). The Alfred Wegener Institute for Polar and Marine Research (AWI, Germany) kindly made it possible to participate in the cruise in the eastern Atlantic. We thank R. Röttgers for providing PSICAM data and HPLC data from analysis at GKSS Research Centre (Geesthacht, Germany). We are grateful to G. Chang (University of California, Santa Barbara) and an anonymous reviewer for providing valuable comments on the published manuscript (Torrecilla *et al.*, 2011b).

4 Hyperspectral data to define bio-optical provinces in the open ocean

The results from this chapter have mostly been published as:

Taylor, B. B., E. Torrecilla, A. Bernhardt, M. H. Taylor, I. Peeken, R. Röttgers, J. Piera and A. Bracher (2011). Bio-optical provinces in the Eastern Atlantic Ocean and their biogeographical relevance. Biogeosciences, 8, 3609-3629, www.biogeosciences.net/8/3609/2011/, doi:10.5194/bg-8-3609-2011.

In the previous chapter, a small but carefully selected data set from the eastern Atlantic Ocean was used to demonstrate the potential usefulness of hyperspectral data of absorption coefficient and remote-sensing reflectance for discriminating different phytoplankton pigment assemblages in the open ocean under non-bloom conditions. Moreover, different analyses were conducted to demonstrate the important role in the performance of the selected spectral range and parameters involved in the calculation of derivative spectra. Nevertheless, one of the main conclusions drawn from this study was that further work was needed with larger databases of measurements from various oceanic environments. In this sense, this chapter is devoted to provide some results demonstrating the potential of the methodology described in this thesis and already proven useful (see Chapters 2 and 3) by examining a database composed of 48 stations. In particular, the results from the application of the cluster-based approach are provided when considering a larger data set of hyperspectral observations collected in the eastern Atlantic Ocean in 2008 (see Fig. 4.1 for station locations). In this case, hyperspectrally-resolved measurements of remote-sensing reflectance spectra, $R_{rs}(\lambda)$, were part of the initial database to analyze, making the use of radiative transfer models not necessary in this study.

In this chapter, the pigment composition information from a total of 48 stations is analyzed in combination with simultaneous hyperspectral optical measurements, which were performed within the framework of a research initiative carried out by the Phytooptics group (University of Bremen and the Alfred-Wegener-Institute for Polar and Marine Research, AWI) led by Prof. Astrid Bracher. In Section 4.1, details on the sample collection and analysis of pigments and measurements of absorption and remote-sensing reflectance spectra are provided. Section 4.2 is devoted to describe

the results from the cluster analysis, which aim is to identify several phytoplankton pigment assemblages spatially distributed along a transect in the eastern Atlantic Ocean using hyperspectral observations. It is also discussed how such identification has served to demonstrate the feasibility of this approach to define different bio-optical provinces based on the phytoplankton community structure and their optical properties. The establishment of different bio-optical provinces, as one of the potential applications of the proposed methodology, has led to examination of its biogeographical relevance by comparison to ecological provinces previously proposed in the literature by Longhurst (2006). Finally, in Section 4.3 the conclusions of this study are presented by summarizing the achievements and pointing the direction of future work.

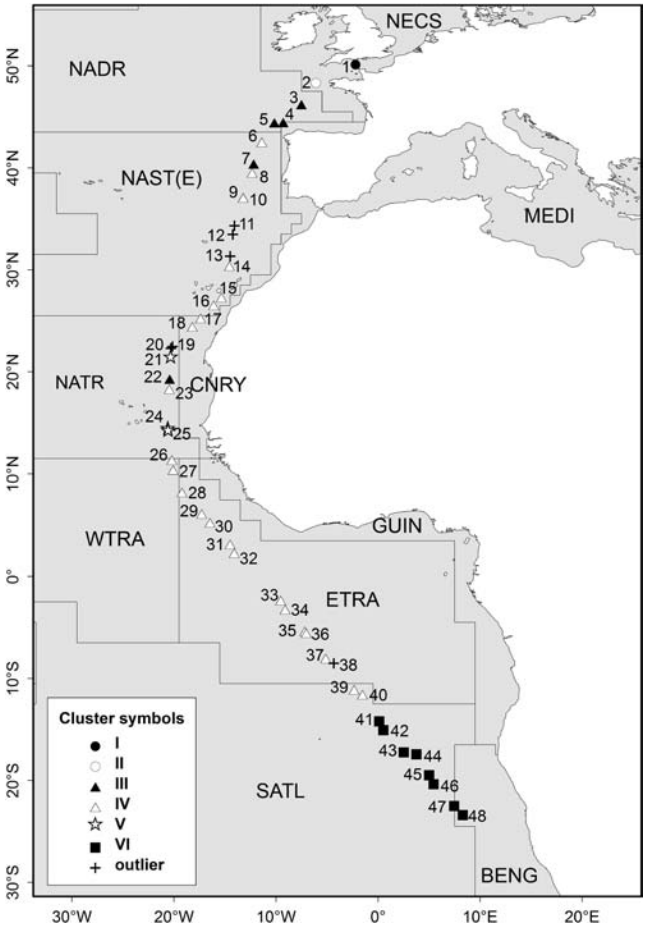


Figure 4.1: Map of the investigated area depicting the 48 stations sampled during the ANT XXV-1 cruise track during November 2008. Stations are identified by symbols corresponding to the groups defined based on the pigment composition (see Section 4.2 for details) and shown superimposed the oceanic provinces defined by Longhurst (2006) (see Section 4.2.1).

4.1 Field measurements and data analysis

4.1.1 Sample collection

Field measurements and data analysis presented in the following sections have been performed within the framework of a research initiative carried out by the Phytooptics group (University of Bremen and the Alfred-Wegener-Institute for Polar and Marine Research, AWI) led by Prof. Astrid Bracher. Water samples were collected at 48 stations during the ANT XXV-1 expedition of the R/V Polarstern along a north-to-south transect through the eastern Atlantic Ocean in November 2008 (Fig. 4.1). The transect spanned different oceanic environments between Bremerhaven (Germany) and Cape Town (South Africa). Typically, each station was performed daily in the morning and at noon local time and involved as described below surface water sampling and simultaneous radiometric measurements (i.e., remote-sensing reflectance). Nevertheless, when time constraints were found (i.e., for 21 stations), only surface waters were sampled through the moonpool of the ship with no radiometric measurements. Until analysis, water samples for pigment and absorption measurements were filtered on GF/F filters, shock-frozen in liquid nitrogen and stored at -80^{o}C. Additionally, measurements of surface salinity, temperature and fluorescence were conducted continuously throughout the cruise.

4.1.2 Pigment analysis

Concentration of 23 pigments in phytoplankton were measured on surface water samples from each station using HPLC, following a method described in Hoffmann *et al.* (2006) and adjusted to the instruments. Samples were measured using a Waters 600 controller combined with a photodiode array detector (PDA, Waters 2998) and an auto sampler (Waters 717plus). Identification and quantification of concentration of the different pigments were carried out using the commercially available program EMPOWER by Waters. The pigment data were quality controlled according to Aiken *et al.* (2009). HPLC data provided information about the presence and contribution of different phytoplankton groups which coexisted in the mixed populations under analysis.

4.1.3 Hyperspectral absorption measurements

The spectral absorption coefficient of particles, $a_p(\lambda)$, was determined from high spectral resolution measurements on discrete water samples with a dual-beam UV/VIS spectrophotometer (Cary 4000, Varian Inc.), equipped with a 150 mm integrating sphere (external DRA-900, Varian, Inc. and Labsphere Inc., made from Spectralon (TM)). A modified quantitative filterpad technique (see e.g., Simis *et al.*, 2005) was used in which the filters were placed in the center of the integrating sphere using a center-mount filter holder perpendicular to the light beam. A wavelength scan from 300 to 850 nm with a resolution of 1 nm (slit width 2 nm, scan rate 150 nm/min) was performed when the reflectance ports were covered with Spectralon(TM) reflectance

standards. The baseline was recorded beforehand with a clean, dry filter, and a filter which was soaked for more than 30 min in purified water which served as a reference. Each absorption coefficient was calculated from a optical density (OD) measurement using a path length amplification factor equal to 4.5 (i.e., $\beta = 1/4.5$, Röttgers, personal communication, 2011) as $a\,[m^{-1}] = -ln\,(T{\cdot}A{\cdot}\beta/V)$, where the transmittance is equal to $T = exp(-OD)$, $V\,[m^3]$ is the filtrated sample volume and $A\,[m^2]$ is the filter clearance area. Results from the original filter provided the $a_p(\lambda)$ measurements. In order to determine the spectral absorption coefficient of non-algal particles ($a_{nap}(\lambda)$), the algal pigments were bleached with sodium hypochlorite [NaOCl] as described in Tassan and Ferrari (1995) and Ferrari and Tassan (1999). The bleached samples were also measured as described above and therefore, each spectral absorption coefficient of phytoplankton ($a_{ph}(\lambda)$) was obtained by substracting $a_{nap}(\lambda) - a_p(\lambda)$.

4.1.4 Hyperspectral remote-sensing reflectance measurements

Similarly to multispectral measurements of remote-sensing reflectances described in Section 3.1.1.3, hyperspectrally-resolved $R_{rs}(\lambda)$ were determined at different stations along the cruise track by combining vertical profiles of irradiance and radiance underwater light fields. When no time constraints or technical problems were found, these profiles were measured with hyperspectral radiometers (RAMSES, TriOS GmbH, Germany), which cover a wavelength range from 320 nm to 950 nm with an optical resolution of 3.3 nm and a spectral accuracy of 0.3 nm. All measurements were collected with sensor-specific automatically adjusted integration times (i.e., between 4 ms and 8 s).

A total of 21 radiometric vertical profiles were collected simultaneously to the collection of water samples down to a maximum depth of 190 m. A radiance sensor measuring upwelling radiance ($L_u(\lambda)$) and a irradiance sensor measuring downwelling irradiance ($E_d(\lambda)$) were used. Furthermore, irradiance at the surface ($E_d^+(\lambda)$) was continously measured as a reference with a third sensor placed above-water, which allowed normalization of the in-water measurements according to Stramski *et al.* (2008). One of the in-water sensors was equipped with an inclination sensor and a pressure sensor. In order to avoid ship's shadow, the ship was oriented in such a way that the sun was illuminating the side where the radiometric measurements were taking place.

All radiometric data were checked for minimum incoming solar light or precipitation (Wernand, 2002). Regarding deck tilt of the ship, the pitch and roll data measured did not exceed values larger than 5^o. For the in-water data, the inclination in either dimension had to be smaller than 14^o (Matsuoka *et al.*, 2007). The profile data were averaged in discrete intervals of 2 m down to a depth of 48 m, intervals of 4 m for depths between 48 and 80 m and intervals of 10 m for all measurements below 80 m. As surface waves strongly affect measurements in the upper few meters, deeper measurements that are more reliable had to be used and extrapolated to the sea surface (Mueller *et al.*, 2003b). Similarly to Stramski *et al.* (2008) a depth interval was defined (i.e., $z' = 7$ to 21 m) to calculate the vertical attenuation coef-

ficients for downwelling irradiance and upwelling radiance, $K_d(\lambda, z')$ and $K_u(\lambda, z')$, respectively). With $K_d(\lambda, z')$ and $K_u(\lambda, z')$, the subsurface irradiance $E_d^-(\lambda, 0\,m)$ and radiance $L_u^-(\lambda, 0\,m)$ were extrapolated from the profiles of $E_d(\lambda, z)$ and $L_u(\lambda, z)$. For the calculation of the remote-sensing reflectance, the subsurface $L_u^-(\lambda, 0\,m)$ was propagated through the water-air interface by applying a transfer coefficient of 0.5425 as determined in Stramski *et al.* (2008) for the same geographical area. Each $R_{rs}(\lambda)$ in sr^{-1} was therefore calculated using the simultaneous above-water reference downwelling irradiance $E_d^+(\lambda)$ using the following expression:

$$R_{rs}(\lambda) = (0.5425 \cdot L_u^-(\lambda, 0\,m))/E_d^+(\lambda) \tag{4.1}$$

4.1.5 Derivative and cluster analysis

Similarly to results described in the previous chapter (see Section 3.3), the HCA cluster-based approach and one of the two defined objective criteria of cluster similarity (i.e., the cophenetic index) have been used in the analysis presented in this chapter. The selected stations have been automatically classified into different groups according to its phytoplankton pigment composition and optical properties, which include the use of ordinary spectra and its second derivative spectra.

It is noted, however, that an adjustment has been applied to that method as a result of the analysis of a larger database. In order to improve our performance, an Euclidean metric distance (d) has been utilized to generate the pigment-based cluster partition, instead of the angular metric distance used for the spectral data:

$$d(x_1, x_2) = \sqrt{(x_1 - x_2) \cdot (x_1 - x_2)'} \tag{4.2}$$

where x_1 and x_2 include the two considered input data objects. This type of distance has allowed to better point out differences between the 48 stations regarding the magnitude of ratios of concentrations of individual pigments to TChl*a*, and hence has provided better results. It is noted that this issue was not critical in the study described in the previous chapter given that the nine stations presented enough distinct differences in the ratios of dominant accessory pigments to TChl*a*.

4.2 Results and discussion

Through the use of an Euclidean metric distance in the cluster analysis of ratios of accessory pigments to TChl*a*, differences between the 48 stations in terms of the relative roles of individual pigments (i.e., according to their phytoplanktonic composition) have been better emphasized (see Fig. 4.2).

The cluster analysis based on the pigment information resulted in six major clusters as shown in Fig. 4.3. Each of the six clusters has been assigned a roman number and a different symbol, which will be also reproduced in the other cluster trees based on hyperspectral information with the aim to facilitate discussion of results. A clear

north-to-south structure can be distinguished when stations corresponding to each cluster are depicted on the map (see Fig. 4.1).

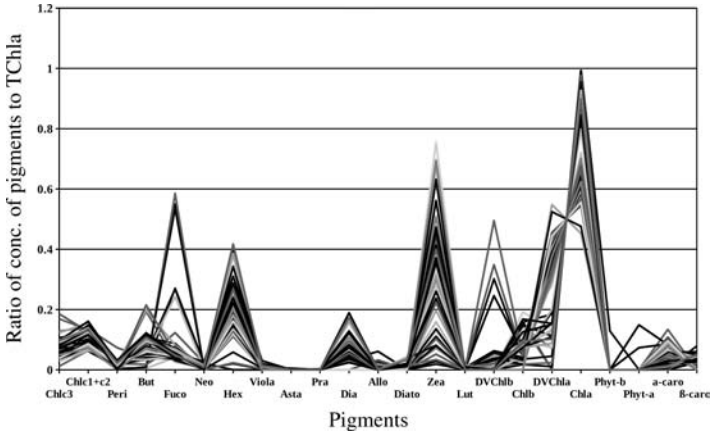


Figure 4.2: Ratios of 23 dominant pigments to TChl$a$ corresponding to the 48 stations. Pigment abbreviations are: Chl$c3$ = chlorophyll-$c3$, Chl$c1{+}c2$ = chlorophyll-$c1{+}c2$, Peri = peridinin, But = 19' - butanoloxyfucoxanthin, Fuco = fucoxanthin, Neo = neoxanthin, Hex = 19' - hexanoyloxyfucoxanthin, Viola = violaxanthin, Asta = astaxanthin, Pra = prasinoxanthin, Dia = diadinoxanthin, Allo = alloxanthin, Diato = diatoxanthin, Zea = zeaxanthin, Lut = lutein, DVChl$b$ = divinyl chlorophyll-b, Chl$b$ = monovinyl chlorophyll-b, DVChl$a$ = divinyl chlorophyll-a, Chl$a$ = monovinyl chlorophyll-a, Phyt-b = phaoephytin-b, Phyt-a = phaoephytin-a, α - caro = α - carotene and β - caro = β - carotene.

The results of this pigment-based cluster analysis are quite consistent with the classification obtained by just considering the dominant pigments in each cluster of stations, which main features are summarized in Table 4.1. For the 48 stations, there exists a larger variability in the estimate of TChl$a$ concentration in comparison with the previous study (see Section 3.3). It ranges from about 0.104 $mg\,m^{-3}$ at the most oligotrophic waters to 5.36 $mg\,m^{-3}$ at stations corresponding to some bloom events.

Cluster I comprises just station 1, which predominant pigments are Fuco and Chl$b$. It is clearly different from all other stations and this predominance suggests that it could be dominated by Fuco-rich diatoms, as well as Chl$b$-rich chlorophytes. It is noted that Fuco can also be a precursor of pigments Hex and But (Jeffrey and Vesk, 1997), which are specific for haptophytes and chrysophytes. Cluster II is also composed of a single station (station 2) and is associated with a local algal bloom given its high TChl$a$ concentration (2.06 $mg\,m^{-3}$). As Chl$b$ and Hex are the predominant pigments at this station, it could have been dominanted by chlorophytes and Hex-rich haptophytes. However, a very high concentration of the pigment Peri, up to 100 times larger than any other station, also suggests an important presence of dinoglagellates.

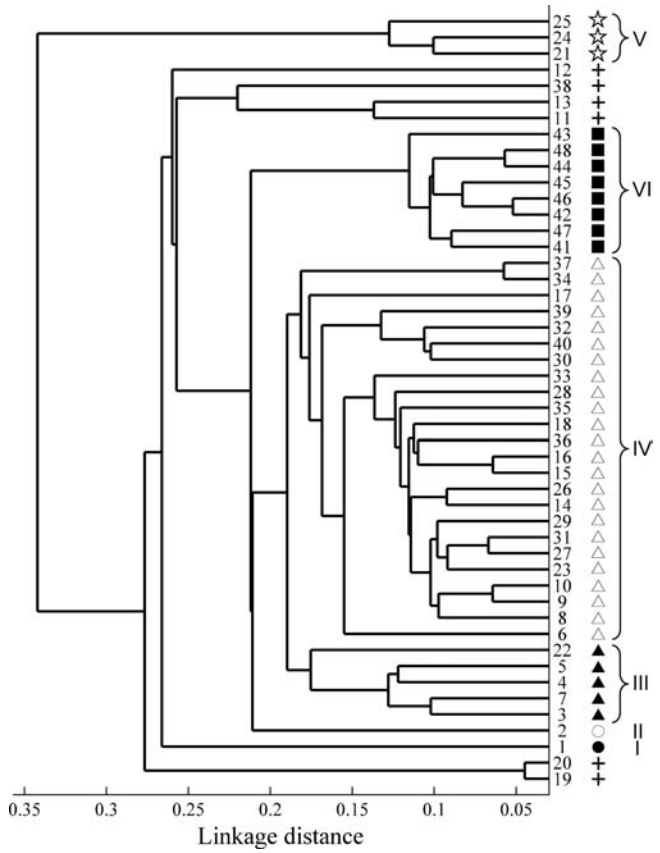


Figure 4.3: Cluster tree obtained for the 48 stations using pigment data determined by HPLC. Main clusters are identified with symbols and roman numbers.

Cluster III is geographically rather patchy but all stations are clearly dominated by the pigment Hex. Therefore, the main phytoplankton group at these stations is haptophytes, even a possible coexistence of other groups depending on the secondary pigments present in each station belonging to this cluster. Cluster IV is the biggest cluster and unifies all stations with DVChl*a*, Zea and Hex as dominant pigments. In accordance with the results shown in Section 3.3.1, DVChl*a* and Zea are diagnostic of picophytoplankton that includes DVChl*a*- and Zea-containing prochlorophytes and Zea-containing cyanobacteria (mainly Synechococcus in the open ocean waters). The abundance of picophytoplankton groups is also accompanied by the presence of haptophytes due to that pigment Hex follows in the ranking. In Table 4.1, stations from cluster IV are indicated separately into a northern and a southern sub-clusters. This geographic division can be explained by considering the levels of TChl*a* and other environmental factors (e.g., temperature and salinity profiles). Nevertheless, the reasons are beyond the scope of this study (further details can be found in Taylor *et al.*, 2011). Cluster V comprises the "bloom stations" (i.e., stations 21, 24 and 25) with Fuco as the dominant diagnostic pigment and very high TChl*a* concentrations, which

range from 1.3 to 5.36 $mg\,m^{-3}$. The blooms mainly consist of Fuco-rich diatoms, haptophytes and chrysophytes. The stations belonging to cluster VI (i.e., 41-48) are clearly separated from other stations due to the predominance of the marker pigments Zea and Hex. It suggests a dominance of Synechococcus-type cyanobacteria and haptophytes. Due to their distinct pigment composition, stations 11, 12, 13, 19, 20 and 38 have been clustered away from their geographical neighbours. Stations 19 and 20 are dominated by the pigment Fuco but in a less proportion than the "bloom stations" in cluster V. Stations 11, 12, 13 and 38 present a dominant role of picoplankton, in particular of prochlorophytes in comparison with their geographical neighbours. This is due to the dominance of DVChl*a* and DVChl*b*. Despite the differences in pigment composition, all these six stations have been defined as "outliers" given that, as described below, they do not show distinct optical features.

Cluster / Symbol	Stations	TChl*a* [mg/m^3]	Dominant pigments (sorted by dominance)	Temp. [°C]	Salinity
I ●	1	0.73	Fuco, Chl*b*	14.18	35.31
II ○	2	2.06	Chl*b*, Hex	13.32	35.42
III ▲	3-5, 7, 22	0.60 ± 0.26	Hex	17.39 ± 3.26	35.08 ± 0.29
IV-N △	6, 8-10, 14-16, 18	0.34 ± 0.16†	Zea, DVChl*a*, Hex	19.71 ± 2.10†	36.62 ± 0.35†
IV-S △	26-37, 39, 40	0.22 ± 0.06	Zea, DVChl*a*, Hex	26.82 ± 2.32	35.56 ± 0.77
V ☆	21, 24, 25	3.24 ± 1.66	Fuco, Dia, Chl*c1*,Chl*c2*	25.27 ± 2.65	36.01 ± 0.39
VI ■	41-48	0.32 ± 0.09	Zea, Hex	19.55 ± 0.48	35.99 ± 0.25
Outliers +	11	0.10	Zea, DVChl*a*, DVChl*b*	20.24	36.91
	13	0.16		20.73	36.96
	12	0.23	DVChl*b*, DVChl*a*, Hex	20.66	36.97
	19	0.64	Fuco, DVChl*a*, Zea	22.98	36.90
	20	0.82		23.11	37.00
	38	0.19	DVChl*b*, DVChl*a*, Hex	23.42	36.21

Table 4.1: Summary of cluster properties obtained for the 48 stations using the pigment information. Properties and pigments abbreviations are: TChl*a* = Chl*a* + DVChl*a* (Mean±SD); Temp. = Surface temperature (Mean±SD); Salinity = Surface salinity (Mean±SD); MVChl*a* = monovinyl chlorophyll-*a*, DVChl*a* = divinyl chlorophyll-*a*, Fuco = fucoxanthin, Hex = 19' - hexanoyloxyfucoxanthin, But = 19' - butanoloxyfucoxanthin, Chl*b* = monovinyl chlorophyll-*b*, Chl*c1* = chlorophyll-*c1*, Chl*c2* = chlorophyll-*c2*, Zea = zeaxanthin, and Dia = diadinoxanthin; † = significantly different from sub-cluster IV-S ($p<0.05$). (Table adapted from Taylor *et al.*, 2011).

The cluster analysis has also been performed based on $a_{ph}(\lambda)$ spectra measured for all 48 stations and shown in Fig. 4.4. Fig. 4.5 depicts the corresponding hierarchical cluster tree, including the symbols of the pigment-based cluster tree for each station. The partition provided by the cluster analysis based on the $a_{ph}(\lambda)$ data is quite similar

to the partition obtained using the pigment information. This is reflected in the high value of the cophenetic index (r_C) obtained, which is equal to 0.70 considering all 48 stations and equal to 0.84 when the "outliers" were taken out in the analysis (i.e., stations 11, 12, 13, 19, 20 and 38). In order to optimize the performance, a sensitivity test of the cluster analysis regarding the choice of the spectral range has been needed and is discussed below (see Fig. 4.8a). In accordance with the results shown in the previous chapter (Section 3.3), the best degree of similarity between these two cluster trees (i.e. the highest cophenetic index) is obtained when the $a_{ph}(\lambda)$ spectra are analyzed over a restricted spectral range from 435 to 520 nm. This is the range in which most relevant pigments show their main absorption characteristics. For $a_{ph}(\lambda)$ data, derivative computations do not improve the performance.

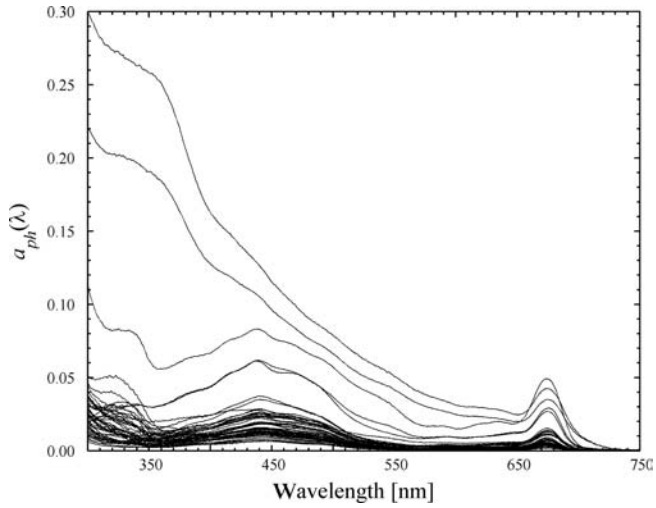


Figure 4.4: Spectral absorption of phytoplankton ($a_{ph}(\lambda)$) for all stations.

By taking into account the optimal spectral range, clusters I, II, V and VI are exactly reproduced by the $a_{ph}(\lambda)$ spectra, whereas clusters III and IV show a few differences. In particular, stations 6 and 8, which belong to cluster IV in the pigment-based tree, are classified with cluster III in the $a_{ph}(\lambda)$-based tree. This may be attributable to differences in the spectral shape of the $a_{ph}(\lambda)$ spectrum of these two stations around the wavelength 470 nm in comparison to other stations from group IV. Probably because its proximity in space, their shape shows a greater similarity with $a_{ph}(\lambda)$ spectra from stations in cluster III. In addition, most of the "outliers" described in the pigment-based cluster tree (i.e. stations 11, 12, 13 and 38) are grouped in the $a_{ph}(\lambda)$-based tree together with stations assigned to cluster IV (i.e., stations indicated with a triangular symbol in Fig. 4.5). Apart from DVChla, which is a principal pigment common to all stations in cluster IV and "outliers", the main reason for being singled out in the pigment-based tree is the abundance of pigment DVChl*b*. This pigment has its main absorption peak around 480 nm (see Fig. 1.3). However, since the pigments Zea and Hex (i.e., the dominant pigments in cluster IV)

also absorb strongly at that wavelength range, it would explain the similarity of the $a_{ph}(\lambda)$ spectra despite the different pigment composition. Stations 19 and 20 were also considered "outliers" and grouped separately in the pigment-based tree due to the dominance of the pigment Fuco in these two samples. Their location coincides with the outer rim of the diatom bloom detected in stations 21, 24 and 25 and characterized by a high concentration of the pigment Fuco. Nevertheless, stations 19 and 20 present a smaller percentage of Fuco compared to total pigment and are also characterized by the presence of DVChl*a* which delineated the stations before and after the bloom. Due to such differences in pigment composition, it could be expected that these stations also partition separately in the cluster tree based on the absorption information. However, both stations 19 and 20 clustered together with stations assigned to cluster IV, as the rest of "outliers".

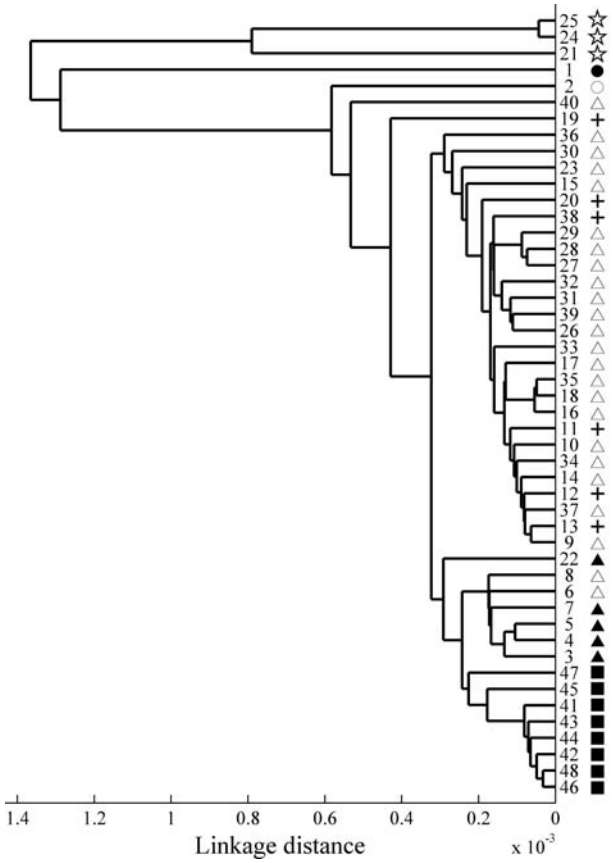


Figure 4.5: Results of cluster analysis applied to $a_{ph}(\lambda)$ from the 48 stations. Symbols correspond to pigment-based clusters (see Fig. 4.3).

To explore the potential of remote sensing information in distinguishing different phytoplankton assemblages, the partition of the stations has also been carried out based on available $R_{rs}(\lambda)$ spectra measured for 21 of the stations and shown in Fig. 4.6. As for the absorption data analysis, the cluster analysis based on $R_{rs}(\lambda)$ data also yields satisfactory results (Figure 4.7). A cophenetic index equal to 0.62 is obtained for all stations and equal to 0.90 when again the "outliers" stations have been ruled out in the analysis. As discussed in great detail in Section 3.3, a good performance has only been achieved when considering the second derivative of $R_{rs}(\lambda)$ spectra in this case over the spectral range of 435 to 580 nm. The ability to discriminate phytoplankton pigment assemblages from the derivative of $R_{rs}(\lambda)$ spectra has been optimized by selecting the most suitable parameters used in the spectral derivative computations. In particular, it has been determined that the optimal values for the smoothing filter window and band separation were 9 nm for remote-sensing reflectance (results from this sensitivity analysis are not shown). Due to time constraints at the start of the cruise, no $R_{rs}(\lambda)$ measurements were collected at the stations corresponding to pigment-based clusters I and II. In any case, clusters III, V and VI are well reproduced and cluster IV shows some discrepancies similar to the analysis based on $a_{ph}(\lambda)$ data. The "outliers" stations 12, 20 and 38 which clustered separately in the pigment-based tree (due to their high concentration of DVChl$b$) are associated with stations from cluster IV in the optical data, as happened in the $a_{ph}(\lambda)$-based cluster analysis. Unexpectedly, station 6 (from cluster IV in the pigment-based tree) has been singled out in the $R_{rs}(\lambda)$-based tree. Its distinct $R_{rs}(\lambda)$ spectrum, with a much lower in magnitude in comparison to other stations from cluster IV, might indicate that some problems or errors occured during the acquisition.

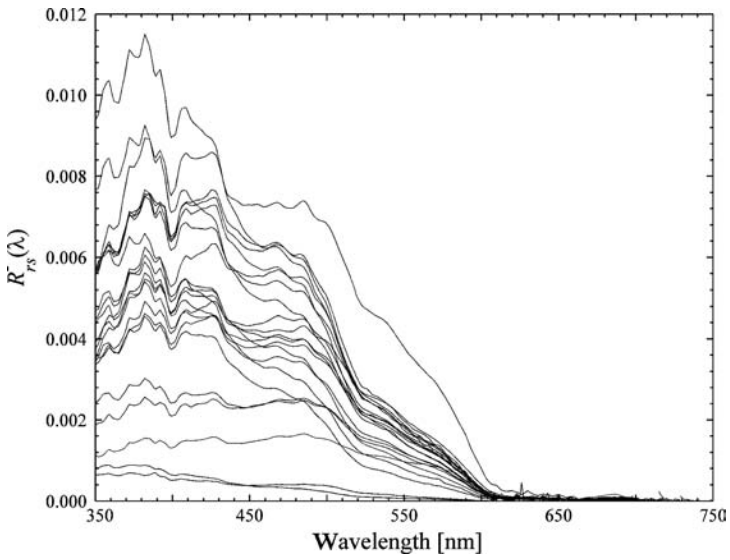


Figure 4.6: Remote-sensing reflectance spectra ($R_{rs}(\lambda)$) for the 21 stations measured.

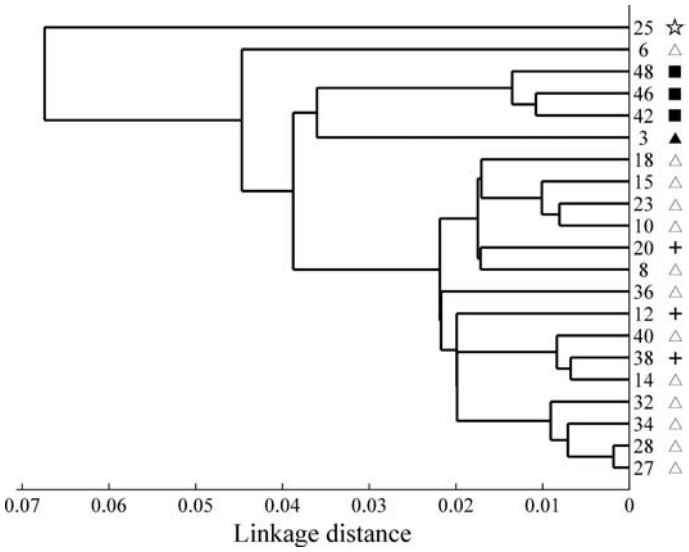


Figure 4.7: Cluster tree resulting from analysis of the 21 stations calculated using second derivative of hyperspectral $R_{rs}(\lambda)$ spectra (optimal values: $WS = BS = 9$ nm). Symbols correspond to pigment-based clusters (Fig. 3.8).

In order to optimize the results, the degree of similarity between cluster trees has been again evaluated for calculations involving different spectral ranges of optical data (Torrecilla *et al.*, 2011a). Fig. 4.8a illustrates the degree of similarity between the $a_{ph}(\lambda)$-based and pigment-based cluster trees for different spectral ranges of absorption data.

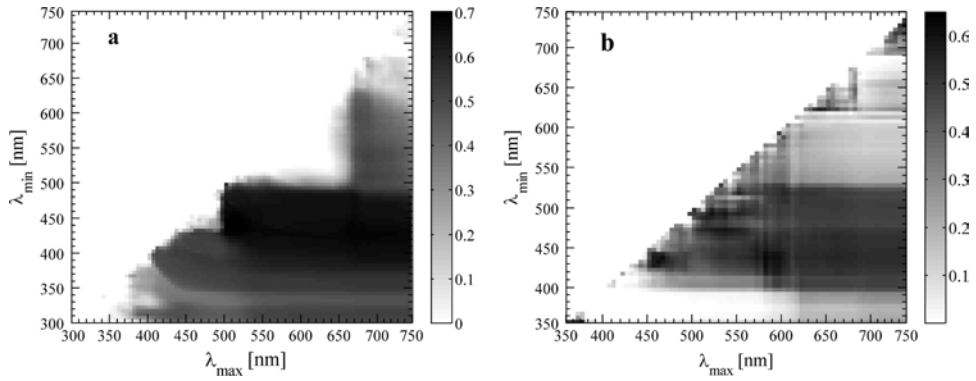


Figure 4.8: Cophenetic indices between pigment-based and optical-based cluster trees obtained using different combinations of spectral range for (a) $a_{ph}(\lambda)$ and (b) second derivative spectra of $R_{rs}(\lambda)$. Optimal values of $WS = BS =$ 9 nm, were used for the calculation of derivative spectra. The y-axis indicates the lower limit of the spectral range (λ_{min}) and the x-axis the upper limit of the spectral range (λ_{max}) utilized.

According to the results, the optimal region is clearly when the phytoplankton absorption spectrum is analyzed over the spectral range approximately from λ_{min} = 425 nm to λ_{max} = 540 nm. Fig. 4.8b shows distributions of the cophenetic index for the cluster analysis of the second derivative of $R_{rs}(\lambda)$. The spectral region from 435 nm to 580 nm, which includes the narrower region from 435 nm to 475 nm, can be considered as the optimal spectral range for the application of derivative approach with potential for good similarity between the pigment-based and reflectance-based cluster trees. Both optimal spectral regions identified, in particular the one based on the analysis of phytoplankton absorption data, overlap with the wavelength range where absorption characteristics of main accessory pigments appear (see Fig. 1.3). As expected, the results from this sensitivity analysis are consistent with the results shown in the previous study (Section 3.3). It is important to note that these results are also in agreement with the ones shown in the chapter 2, where the optimal range was from λ_{min} = 470 nm to λ_{max} = 500 nm. In that case, however, this range was narrower given that the analysis was performed considering simplified underwater optical scenarios, in which only a single phytoplankton group was present.

The results obtained with $a_{ph}(\lambda)$ and $R_{rs}(\lambda)$ spectra are summarized in Table 4.2, including the improvement achieved with the optimal ranges and derivative settings.

	Cophenetic index	
Phytoplankton absorption data – $a_{ph}(\lambda)$		
Hyperspectral (1 nm resolution), range 350-750 nm	0.45	(0.66)
Hyperspectral (1 nm resolution), range 435 – 520 nm	**0.70**	**(0.84)**
Remote-sensing reflectance data – $R_{rs}(\lambda)$		
Hyperspectral (1 nm resolution), range 350-750 nm	0.33	(0.53)
Hyperspectral (1 nm resolution), range 435 – 580 nm	0.49	(0.79)
Hyperspectral 2nd derivative, range 435 – 580 nm Optimal derivative parameters: BS = 9, WS = 9	**0.62**	**(0.90)**

Table 4.2: Summary of results including cophenetic indices between pigment-based and optical-based cluster trees obtained using different spectral ranges of $a_{ph}(\lambda)$, $R_{rs}(\lambda)$ and second derivative spectra of $R_{rs}(\lambda)$. In parenthesis, the cophenetic indices obtained once the "outliers" stations (i.e., 11, 12, 13, 19, 20 and 38) were ruled out from the analysis.

4.2.1 Bio-optical provinces and its biogeographical relevance

On land, the distinction between different ecosystems such as the edge of a forest or the extent of a desert is generally easier compared to the marine environment. Due to the highly dynamic nature of the oceans and the lack of knowledge about many of those dynamics, these boundaries between ecological systems are much harder to observe. They are, however, no less real (Platt and Sathyendranath, 1999; Hooker *et al.*, 2000; Devred *et al.*, 2007; Longhurst, 2006). Areas in the world's oceans with

similar physical and biological characteristics are generally referred to as ecological or biogeochemical provinces or units (e.g. Platt *et al.*, 2005).

Several studies have used TChl*a* as one of the descriptors to define marine ecological provinces (e.g. Devred *et al.*, 2007; Hardman-Mountford *et al.*, 2008). In this study, however, the global contribution of all pigment composition has been combined with hyperspectral optical data to establish a bio-optical geography of the eastern Atlantic Ocean. By applying an unsupervised HCA analysis to data from the eastern Atlantic Ocean, the potential applicability of pigment composition, phytoplankton absorption coefficients and remote sensing reflectances for identifying different bio-optical provinces in the ocean has been demonstrated (see a summary in Table 4.2).

The establishment of different bio-optical provinces has led to examination of its biogeographical relevance by comparison to ecological provinces previously proposed in the literature. In order to assess the bio-optical approach with an established biogeography, the bio-optical provinces/clusters have been compared to the widely used system defined by Longhurst (2006). Longhurst proposed a global partitioning into 4 biomes or basic vegetation types within the pelagic realm of the oceans: the Polar, Westerlies, Trades and Coastal biome. Within the biomes, 51 provinces were defined on the basis of a global data set including satellite images of surface chlorophyll fields, regional oceanography data and also considering many previous proposals for partitioning the oceans. Although the boundaries of the provinces were forced into a static grid as a matter of convenience, Longhurst pointed out that the position of the borders is dynamic and varies on annual, seasonal or even shorter time scales.

In Figure 4.1, the 48 stations with their assigned cluster symbols are plotted on a map, in which Longhurst's provinces are also depicted (shapefile from VLIZ, 2009). A wide discussion regarding the comparison between bio-optical provinces and Longhurst's provinces can be found in Taylor *et al.* (2011). Nevertheless, the main results with a summarized description of similarities observed are provided. Clusters I and II belong to the continental shelf of Western Europe, described by Longhurst as the Northeast Atlantic Shelves Province (NECS). This analysis separates the two stations into two clusters probably due to the fact that station 2 (cluster II) is situated in a area where tidal forces dominate the English Channel. Stations joined in cluster III do not all belong to the same geographical province. Stations 3-5 lie in the south-east corner of Longhurst's North Atlantic Drift province (NADR) and show the same bio-optical traits than stations 7 and 22. Nevertheless, oceanographic data and geography show that these two stations belong to different oceanic provinces. The sub-cluster IV-N lies within Longhurst's North Atlantic Subtropical Gyral Province, East (NAST(E)). However, this area is bio-optically not consistent and many outliers lie alongside the stations in this sub-cluster. These inconsistencies might be explained by typical physical processes in this region (e.g., eddy-driven vertical advection, internal waves, Mediterranean water eddies which flow out of the Strait of Gibraltar and move southwards once they have detached from the continental slope, etc.). Stations 19-25, including "bloom stations" (i.e., 21, 24 and 25), lie on the approximate boundary between the North Atlantic Tropical Gyral Province (NATR) and the Canary Current Coastal Province (CNRY), which comprises the southerly coastal flow of the

eastern boundary current of the North Atlantic from Cap Finisterre (Spain) to Cape Verde (Senegal). The sub-cluster IV-S including stations 26-40 is geographically a very consistent cluster and lies within Longhurst's Eastern Tropical Atlantic Province (ETRA). The boundaries of this province are clearly defined and in correspondence with the observed changes in the pigment composition and bio-optical traits from stations to the north and south. Perez *et al.* (2005) describe the typical phytoplankton community structure for this region as rather constant with the contribution of pico-phytoplankton to TChl*a* always exceeding 45%, a description which is corroborated by the results. Stations belonging to cluster VI are clearly separated through their bio-optical traits from the other stations. They present distinct pigment composition, $a_{ph}(\lambda)$ and $R_{rs}(\lambda)$ spectra and are clearly located in Longhurst's South Atlantic Gyral Province (SATL), which comprises the anticyclonic circulation of the South Atlantic, excluding the coastal boundary currents. The northern boundary of this province is clearly defined by the change in bio-optical data from stations 40 and 41. This coincides fairly well with the boundary set by Longhurst for the SATL province.

The bio-optical provinces have been shown to agree well with the provinces established by Longhurst and thus can be used to classify areas of similar biogeography. However, in areas with high variability, the Longhurst's provinces do not harbour consistent phytoplankton assemblages. In such areas, it would be advantageous to use a more flexible and dynamic approach to describe the ecosystem, in which physical and optical information could be combined (Devred *et al.*, 2007; Raitsos *et al.*, 2008). Figure 4.9 shows variability in the temperature and salinity regimes from stations selected for this study that combined with latitude and longitude information could be used to improve the ability to describe ecological complex regions.

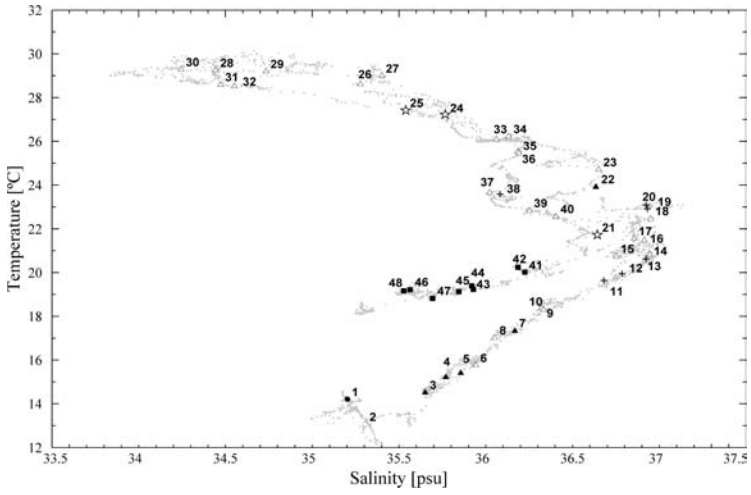


Figure 4.9: Temperature-salinity diagram for the north-south transect in the eastern Atlantic Ocean (see Fig. 4.1). Points corresponding to stations are labeled with their station number and their symbol according to the pigment-based cluster partition (see Fig. 4.3).

4.3 Summary and conclusions

In this chapter, a study based on a larger number of stations and representing a broader set of oceanic environments has successfully been described for discriminating different phytoplankton assemblages in the eastern Atlantic Ocean. The HCA cluster-based analysis has directly been applied to hyperspectral field measurements of both: phytoplankton absorption coefficient spectra and remote-sensing reflectance spectra (i.e., the use of radiative transfer models was not necessary). The 48 stations corresponded to open ocean waters and covered bloom and non-bloom conditions.

In this study, the pigment-based cluster partition has also been used as the reference in terms of the phytoplankton composition of the selected stations (the "sea truth"). However, an Euclidean metric distance has been utilized to generate this cluster partition, instead of the angular metric distance used for the analysis of the spectral data. Given that the 48 stations were representative of a broader set of ecological conditions, differences between the stations in terms of the relative roles of individual pigments (i.e., according to its phytoplanktonic community compostion) have been better emphasized using this type of distance. In accordance with the resulting cluster analysis described in the previous chapter (see Section 3.3), the most promising results have been obtained for phytoplankton absorption coefficient over the spectral range 435-520 nm and the second derivative spectra of the remote-sensing reflectance over the spectral range from 435-580 nm.

The proposed unsupervisded cluster algorithm applied to the measured optical parameters has allowed to define different bio-optical provinces. In this sense, since pigment composition corresponded well with the remotely observable hyperspectral $a_{ph}(\lambda)$ and $R_{rs}(\lambda)$ spectra, this method has potential to become an automated approach for the future to classify areas of similar biogeography. Partitioning the ocean into provinces can assist us in understanding complex patterns in the oceans and help us to extrapolate province-specific parameters over large spatial scales for a better estimation of global primary production and carbon budget. The approximation of ecological parameters (e.g. in modeling or in the interpretation of satellite data) can be more accurate if we know which areas exhibit similar dynamics. This novel approach, for instance, could be helpful to further assess the extension and dynamic of ecological provinces in the ocean on a large spatial and temporal scale. In this sense, hyperspectral sensors mounted on platforms such as buoys, gliders or satellites could provide essential data to identify shifting boundaries of established provinces or to track exceptions from the rule to improve our understanding of the biogeochemical cycles in the ocean. In terms of globally significant issues such as carbon export and primary production, the hyperspectral-based approach represents a possibility for a faster and detailed assessment of the state of a temporally and spatially variable marine environment.

Different approaches based on bio-optical data have been proposed to define different ecological environments or provinces, such as the classification by Hardman-Mountford *et al.* (2008) based on satellite-derived chlorophyll data or the cluster algorithm by Moore *et al.* (2009b) where optical water types are classified using

multispectral $R_{rs}(\lambda)$ observations. It is noted, however, that in light of the recent advances in hyperspectral technology on the remote sensing observing platforms and the results presented in this study, the possibility of characterizing marine ecological units and phytoplankton communities through hyperspectral measurements becomes increasingly interesting and important. For future applications, the use of a methodology based on hyperspectral resolution data sets is thus suggested. In fact, it seems suitable to use that type of data sets in combination with other physical parameters such as the sea-surface temperature (Devred *et al.*, 2007). Several approaches have already been proposed to detect Phytoplankton Functional Types (PFTs) from satellite sensors (e.g., Alvain *et al.*, 2008), but only Bracher *et al.* (2009) used hyperspectral data. However, dominant PFT identification is a slightly different concept than identifying different spatially distributed phytoplankton assemblages. In this sense, the novel methodology proposed in this thesis allowed to define different bio-optical provinces on the basis of phytoplankton community structure and their bio-optical traits, instead of a small set of functional groups.

Given significant interest in the development of the capabilities for large-scale characterization of phytoplankton biodiversity from optical measurements including remote-sensing observations, one may expect further expansion of comprehensive databases consisting of concurrent pigment and optical information in the near future. It is expected that the availability of such a larger data sets will support further work and optimization of this novel cluster-based approach and other techniques, such as neural networks (Aymerich *et al.*, 2009; Raitsos *et al.*, 2008), which exploit hyperspectral optical measurements as a source of information on phytoplankton community composition.

4.4 Acknowledgements

The research presented in this chapter was supported and funded by the Alfred Wegener Institute for Polar and Marine Research (AWI, Germany), the German Helmholtz Impulse Fond, the Spanish National Research Council (CSIC), and the OCEANET project. Part of this research was performed during a visit of E. Torrecilla at the PHYTOOPTICS group, University of Bremen and AWI (Germany), which was supported by the CSIC (project ANERIS PIF08-015) and the Spanish Ministry of Education (Ph.D. European Mentoring Program). We are grateful to NASA and ESA, particularly to the GlobColour Project, for processing and supplying satellite total chlorophyll-*a* concentrations. We thank E. Allhusen, C. Klein, M. Lunau, E. Nöthig and S. Wiegmann for helping with the field samples work and M. Soppa and G. Rohardt for further data analysis. We also thank the crew, principal investigators and other scientists on board the R/V Polarstern cruise ANTXXV-1 for support on board and fruitful discussions afterwards. We are grateful to two anonymous reviewers for providing valuable comments on the published manuscript (Taylor *et al.*, 2011).

5 Hyperspectral data for a remote sensing assessment of phytoplankton communities in estuarine waters by means of model simulations

In the previous chapters, a new approach based on the analysis of hyperspectral optical data has been demonstrated useful to successfully discriminate different phytoplankton communities in open ocean environments. The main goal of this chapter is to evaluate the feasibility of this approach when considering shallow estuarine waters.

Coastal and estuarine areas are among the most biologically productive, dynamic and vulnerable areas of the world oceans. Due to the complexity of physical-biological interactions along these regions, near shore areas have traditionally been intensely studied (Turner *et al.*, 1998; Crossland *et al.*, 2005). In comparison to land areas or open ocean waters, systematic monitoring of coastal environments involves specific challenges due to the common presence of suspended sediments and its high turbidity and dynamics. Improved understanding of coastal ocean physical processes and their effects on biology is especially important because the majority of the world's primary production occurs on continental shelves. In addition, improved management of interactions between human-induced and natural processes within these regions is needed. Several interdisciplinary coastal research and monitoring programs are being carried out conceived to improve our understanding of how coastal ecosystems function. For instance, the Global Ecology and Oceanography of Harmful Algal Blooms (GEOHAB) program of the Intergovernmental Oceanographic Commission of UNESCO (Gilbert and Pitcher, 2001) is focused on the study of harmful algal blooms (HABs) population dynamics. These mass occurrences typically attributable to single species of phytoplankton often appear in coastal waters under favorable conditions like when sunlight and nutrients become available. HABs cause problems due to their toxicity or in other ways (Kutser, 2009) and are a sensitive indicator for marine ecosystem changes. Another research initiative carried out by an interdisciplinar and multi-institutional team is the Layered Organization in the Coastal Ocean (LOCO) project (Sullivan *et al.*, 2010; Rines *et al.*, 2010), which aim is to investigate the highly-concentrated patches of plankton known as "thin layers". These structures range in thickness from few centimeters to few meters, can extend horizontally for many kilometers and persist for hours to weeks (Dekshenieks *et al.*, 2001; Cowles, 2003). Their understanding could be critical to HABs research.

The development of new optical-based observing platforms and algorithms to obtain previously unavalaible information on the phytoplankton communities in coastal waters has demonstrated to yield significant benefits for HABs monitoring (Dickey, 2004; Babin *et al.*, 2005; Schofield *et al.*, 2008). They represent a valuable complement to other more conventional approaches utilized by biologists working in coastal ecosystems. Several optical-based approaches have recently been developed to provide information on the abundance and dynamics of dissolved and particulate matter in coastal waters, some of them based on multispectral data or band-ratios obtained in discrete spectral bands (Doxaran *et al.*, 2006; Gitelson *et al.*, 2007; Astoreca *et al.*, 2009; Hommersom *et al.*, 2011). In the same direction, new approaches that fully exploit the hyperspectral capabilities of the new generation of in situ and spaceborne spectrometers are becoming an alternative in order to better characterize phytoplankton-dominated Case 2 waters (Lubac *et al.*, 2008; Hunter *et al.*, 2010). The contribution of hyperspectral sensors mounted on new autonomous, continuous and real-time ocean observing platforms to a better understanding of the biological productivity in coastal waters, even to the development of early-warning monitoring and prediction of HABs, needs further investigation, though.

In recent years, efforts to assess the effect of the vertical structure of seawater constituents on the optical property of remote-sensing reflectance have also increased. For example, numerical simulations of radiative transfer have been used to examine the effect of a nonuniform vertical profile of phytoplankton on the remote-sensing reflectance corresponding to open ocean environments (Frette *et al.*, 2001; Stramska and Stramski, 2005; Xiu *et al.*, 2008; Torrecilla *et al.*, 2010) or to more complex Case 2 waters dominated by cyanobacteria (Kutser *et al.*, 2008). In contrast to all these studies, the effect of a vertical inhomogeneity of phytoplankton in shallow estuarine waters on the remote-sensing signal still needs to be explored.

The aim of this chapter is to examine the feasibility of a hyperspectral-based approach for monitoring taxonomic composition of phytoplankton communities in shallow estuarine waters. As a proof-of-concept study and by means of model simulations, a set of hyperspectrally-resolved remote-sensing reflectances characterizing different shallow water scenarios is analyzed by using the dissimilarity-based cluster method described in Chapter 2. First, in Section 5.2.1, the sensitivity of remote-sensing reflectance spectra, $R_{rs}(\lambda)$, to phytoplankton taxa is assessed given different bottom type conditions and abundances of suspended matter. Second, in Section 5.2.2, a sensitivity analysis is also performed to evaluate if different shallow estuarine scenarios with distinct vertical structures in terms of phytoplankton distribution and abundance can be detected through the analysis of the $R_{rs}(\lambda)$ signal. The depth and amplitude of an idealized thin layer were varied in order to estimate their influence.

In order to acomplish the main goal of this chapter, a specific coastal field site has been taken as a case study area. The Alfacs Bay, an estuarine embayment of the Ebro Delta (NW Mediterranean Sea), is roughly 11 km long and 4 km wide, with a maximum depth of 6.5 m (Fig. 5.1a). It harbors an active finfish and shellfish aquaculture industry (mainly mussel and oister). Because of this, it is regularly and intensively monitored for environmental parameters, harmful microalgae and

marine toxins in molluscs, microbiology parameters, heavy metals and dioxins in molluscs. The monitoring programme, which is carried out by the Catalan Institute for Food and Agricultural Research and Technology (IRTA) since 1991, has reported the presence of several harmful algal species responsible for Diarrheic (DSP) and Paralytic Shellfish (PSP) Poisoning toxins in bivalves (Fig. 5.1b). Two toxic bloom-forming groups of phytoplankton reported in this area are dinoflagellates and diatoms (Fernández-Tejedor *et al.*, 2010), in which this study focuses. Specific absorption coefficient spectra of these two phytoplankton groups are depicted in Fig. 5.1c. This monitoring programme, which has evolved in parallel to international programmes on HABs and Spanish and European Union regulations in shellfish production areas, is strongly linked to ongoing research projects on harmful microalgae dynamics in which transport processes play a key role. Different studies have been developed in this area due to its scientific interest (Garcés *et al.*, 1999; Vila *et al.*, 2001; Llebot *et al.*, 2011), including the analysis of the growth and the spatio-temporal distribution of potentially harmful group of dinoflagellates.

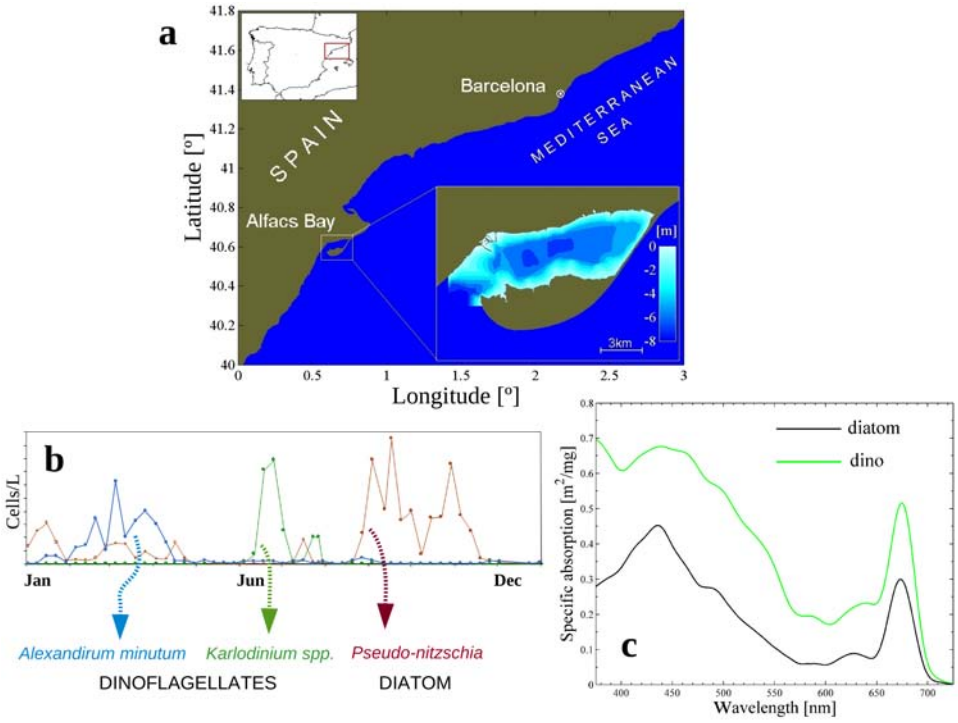


Figure 5.1: (a) Map depicting the location of the Alfacs Bay. (b) Different toxic algal blooms recorded during a weekly sampling in 2007 performed by the IRTA. (c) Chlorophyll-*a* specific absorption coefficient of two groups of phytoplankton: diatoms and dinoflagellates (Kim and Philpot, 2006), reported to be present in the Alfacs Bay. These groups show distinct optical features, in particular in the spectral region from 435 to 580 nm where absorption characteristics of main accessory pigments appear.

5.1 Optical modeling and data analysis

5.1.1 Modeled hyperspectral remote-sensing reflectance

In order to examine the sensitivity of the remote-sensing reflectance corresponding to shallow waters to the phytoplankton taxa, several secenarios have been simulated with the Hydrolight/Ecolight version 5.0 radiative transfer (RT) model (Mobley and Sundman, 2008). A Case 2 waters model representing a shallow estuarine environment have been parameterised by defining the number of optically active substances and providing their specific optical properties. The pure water absorption and scattering coefficients have been derived from Pope and Fry (1997). The specific absorption and scattering spectral coefficients corresponding to two phytoplankton groups representative for the considered area (i.e., diatoms and dinoflagellates, see Fig. 5.1c) have been taken from Kim and Philpot (2006). Due to the lack of knowledge about the colored dissolved organic matter (CDOM) in the Alfacs Bay, the absorption coefficient of CDOM has been calculated using a common exponential relationship included in the Hydrolight code, based on a model from Morel and Maritorena (2001):

$$a_{cdom}(z,\lambda) = 0.2 \cdot a_{ph}(z,440)\exp[-0.014(\lambda - 440)] \tag{5.1}$$

where $a_{ph}(z,440)$ is the absorption coefficient of the chlorophyll-bearing particles (i.e., phytoplankton) at 440 nm. The concentration of suspended sediments, SS, has been taken to vary from 0 to 6 mg/L based on field observations (M. L. Artigas, *unpublised data*). The mass-specific absorption and scattering coefficients of SS have been derived from a composite of data in Bukata *et al.* (1995) and Gallie *et al.* (1992), also included in the Hydrolight code. The modeling of the remote-sensing reflectances was carried out over the spectral region of 375-725 nm with 1 nm resolution and included the inelastic Raman scattering and fluorescence by chlorophyll and CDOM within the ocean. The sea surface boundary conditions have been configured using the default settings in Hydrolight (i.e., wind speed equal to 5 m/s in accordance to real conditions in the area, solar zenith angle equal to 30^o and clear sky condition). In order to emulate similar conditions to those found in the Alfacs Bay, the ocean has been assumed to be a finite-depth water body with a z_{max} equal to 6.5 m and examined at intervals $\triangle z$ equal to 0.05 m. Different bottom types have been considered corresponding to various sediments and benthic biota, among them those ones similar to the ones found in the studied area (de Pedro Puente, 2007). Each bottom type, characterized by a specific irradiance reflectance (see Fig. 5.2), was provided by the Hydrolight code.

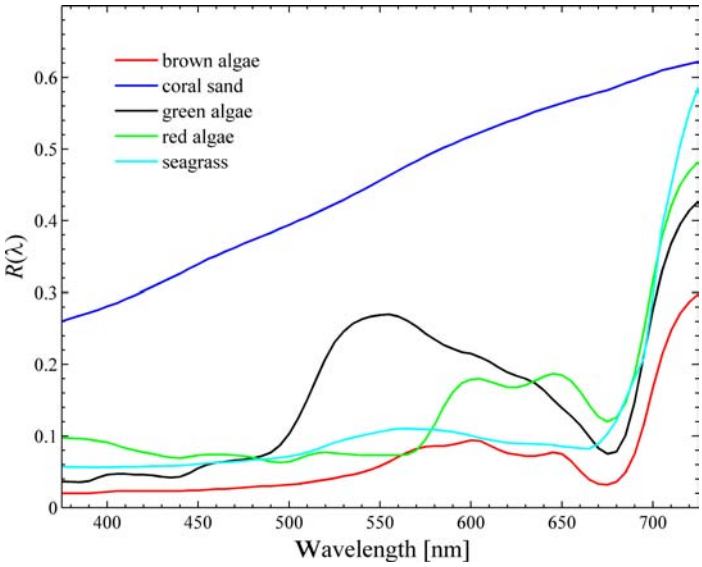


Figure 5.2: Bottom irradiance reflectances provided by the Hydrolight code and considered for the present study.

Two different vertical profiles of chlorophyll concentration have been used in the model simulations. On the one hand, several uniform vertical profiles of chorophyll concentration equal to 0.5, 1, 3, 5 and 7 mg/m^3 have been used to test the sensitivity of $R_{rs}(\lambda)$ to the phytoplankton taxa, bottom type and presence of suspended sediments (see Section 5.2.1). On the other hand, in order to examine their effect on the $R_{rs}(\lambda)$, different nonuniform vertical profiles of the chlorophyll concentration have been generated (see Section 5.2.2). These vertical profiles, which simulate the condition characterized by the presence of a subsurface chlorophyll maximum at a specific depth (i.e., a thin layer condition, see Fig. 5.3), have been approximated by a Gaussian function superimposed upon a constant background (Lewis *et al.*, 1983):

$$Chl(z) = Chl_0 + Chl_{max} \exp[-(z - z_{max})^2/2\sigma^2] \tag{5.2}$$

where Chl_0 is the background value of chlorophyll, z_{max} is the depth of the chlorophyll maximum, σ is the standard deviation that controls the thickness of the cholorphyll peak (i.e., 95% of the integrated biomass is located within a water layer of thickness equal to 4σ) and Chl_{max} is the amplitude of the chlorophyll maximum above the value of Chl_0. According to values derived from the analysis of water samples taken in the Alfacs Bay (Artigas *et al., submitted*), the calculations were made for two values of Chl_0 (1 and 3 mg/m^3), three values of z_{max} (3, 4, and 5 m), one value of σ (0.25 m) and three values of Chl_{max} (6, 9 and 12 mg/m^3).

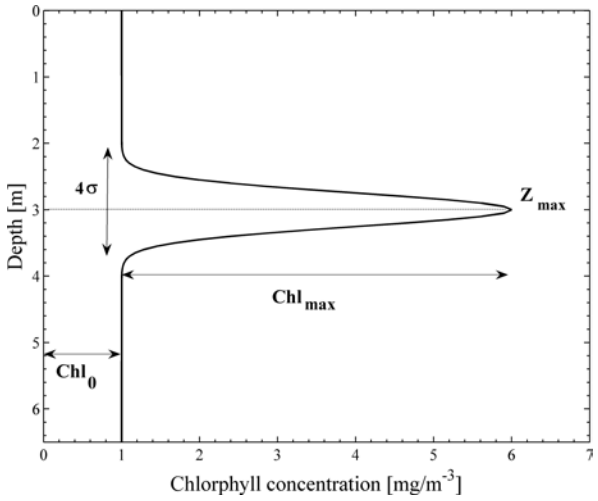


Figure 5.3: Diagram of the nonuniform vertical chlorophyll profile expressed in equation (5.2) and utilized in part of the RT simulations.

The role of each optically active component for the type of environments simulated in this chapter differs from the cases simulated in the previous chapters (i.e., Case 1 waters). Figure 5.4 shows an example of the most significant IOPs (i.e., absorption and backscattering spectral coefficients) considered as the input information to calculate one of the modeled remote-sensing reflectances. It is noted that the suspended matter has a higher influence on the total absorption and backscattering coefficient spectra, and hence also expected on the $R_{rs}(\lambda)$ signal, even becoming the most dominant component when increasing its concentration.

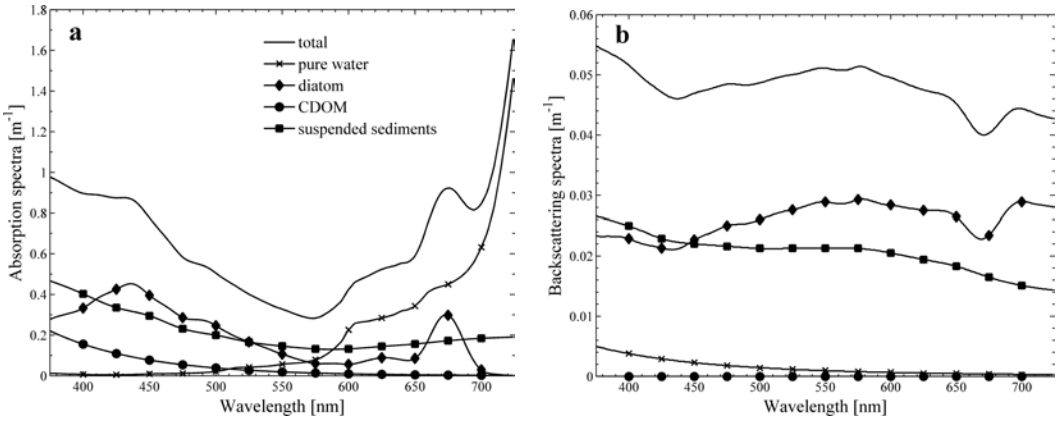


Figure 5.4: Examples of absorption (a) and backscattering (b) spectra utilized in one of the simulations for selected values of: Chl_0 (1 mg/m^3, uniform vertical profile) and SS (2 mg/L). In this case, the underwater scenario was characterized by the presence of the algal group of diatoms.

5.1.2 Cluster analysis and Multi-dimensional Scaling

A hierarchical cluster analysis (HCA) has been applied to the optical datasets of $R_{rs}(\lambda)$ generated for the study presented in this chapter. The assessment of the impact of having different underwater scenarios given different bottom types, vertical structures of phytoplankton and abundance of suspended sediments has been carried out by analyzing the similarity between each simulated hyperspectral $R_{rs}(\lambda)$. In this sense, similarly to previous chapters, the HCA analysis has allowed to examine our ability to distinguish $R_{rs}(\lambda)$ corresponding to different underwater conditions. The cosine distance has also been chosen as measure of similarity between spectra in order to give priority to differences in the spectral shape of optical data rather than magnitude.

It is noted in this study, however, that once a hierarchical cluster tree has been obtained corresponding to a set of underwater scenarios, a new graphical representation instead of a tree has been used to more easily examine our ability to distinguish one scenario from another. Multi-dimensional Scaling (MDS) is an ordination technique that, similar to cluster techniques, ranks objects according to their relative similarity (Shepard, 1980). In this study, it has been used as a complement of the cluster analysis for graphically representing dissimilarities in the $R_{rs}(\lambda)$ data set. It provides an alternative visualization of the complex relations among the $R_{rs}(\lambda)$ spectra obtained with the cluster analysis. MDS is based on the dissimilarity matrix obtained in the HCA analysis (using the cosine distance) but displays the inter-relations between each pair of objects (i.e., each pair of spectra) on a low dimensional plot (2D). The distance between objects on this biplot is an indicator of the similarity in the $R_{rs}(\lambda)$ spectra. The MDS representation provides a reliability measure, which indicates how well the 2D-ordination plot depicts the relationships in the similarity between objects. This is indicated by the minimized *stress* value.

5.2 Results and discussion

5.2.1 Effect of the bottom type on the remote sensing signal

The variability in remote-sensing reflectance spectra due to different characteristics of the bottom type is significant in optically shallow waters. In fact, it has been demonstrated useful to map bottom properties in coastal environments, in particular if observations at hyperspectral resolution are available. Figure 5.5 illustrates the influence of the different bottom types considered in this study (i.e., brown algae, coral sand, green algae, red algae and seagrass) on the $R_{rs}(\lambda)$ spectra when only pure water is assumed to be present in the modeled water body. Coral sand provides a higher $R_{rs}(\lambda)$ spectra than algae due to its higher reflectance over the visible part of the spectrum. $R_{rs}(\lambda)$ spectra corresponding to green algae is higher than those for other algae, in accordance with its reflectance spectrum (see Fig. 5.2). $R_{rs}(\lambda)$ spectra from red, brown algae and seagrass are more similar but show significant differences in shape.

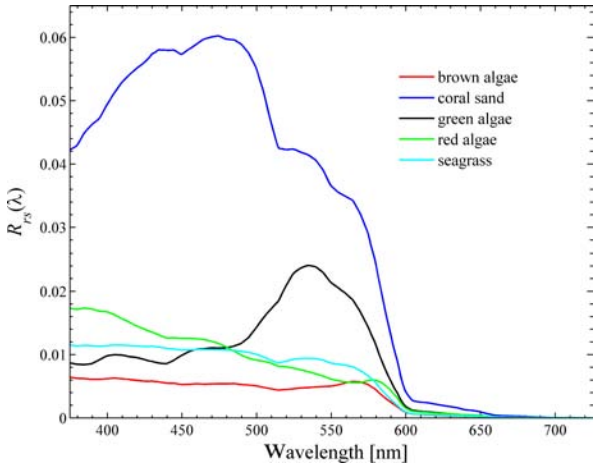


Figure 5.5: Remote-sensing reflectance, $R_{rs}(\lambda)$, of five different bottom types seen through clear water with a maximum depth equal to 6.5 m.

Our ability to discriminate different algal communities and bottom types from $R_{rs}(\lambda)$ information, however, can be complicated by the presence of high concentrations of phytoplankton biomass, for example in the case of existing bloom-forming or thin layer processes. The optical water clarity is degraded as the abundance of phytoplankton along the water column increases. Figure 5.6a shows the effects on $R_{rs}(\lambda)$ spectra caused by different parameters of the bottom type and the level of chlorophyll concentration (i.e., 0.5, 1, 3, 5 and 7 mg/m^3) corresponding to scenarios characterized by the presence of two algal groups: diatoms and dinoflagellates. In this case, it corresponds to uniform vertical profiles of chorophyll concentration. In general, the effect of increasing the chlorophyll concentration is to decrease $R_{rs}(\lambda)$ spectra in the blue spectral region and to increase it in the green and red spectral regions. The highest $R_{rs}(\lambda)$ spectra occur for scenarios characterized by the presence of diatoms, regardless the bottom type. Nevertheless, it is noted that bottom type has no impact on the $R_{rs}(\lambda)$ spectra for high concentrations of chlorophyll (> 3 mg/m^3). In this cases, the $R_{rs}(\lambda)$ spectra corresponding to a fixed chlorophyll concentration and all bottom types are seen in cyan color, since they have been plotted superimposed (i.e., it is just seen the $R_{rs}(\lambda)$ corresponding to the last bottom type).

In order to assess if different shallow water scenarios can still be dintinguished despite the presence of different algal groups and bottom type configurations (see Fig. 5.6a), the HCA cluster analysis has been applied to the dataset of hyperspectral $R_{rs}(\lambda)$ spectra. Figure 5.6b depicts by means of Multi-dimensional scaling (MDS) all cases according to their relative similarity found in the cluster analysis. It is noteworthy that scenarios dominated by either diatoms and dinoflagellates are clearly distinguished regardless the chlorophyll concentration of the uniform vertical profile (i.e., for concentrations from 0.5 to 7 mg/m^3). In addition, for chlorophyll concentrations up to 1 mg/m^3, the analysis of hyperspectral $R_{rs}(\lambda)$ spectra also allows to discriminate scenarios with different bottom characteristics. On the contrary, when

the chlorophyll increases, the information of the bottom type is reflected neither in the $R_{rs}(\lambda)$ spectra nor in the MDS plot.

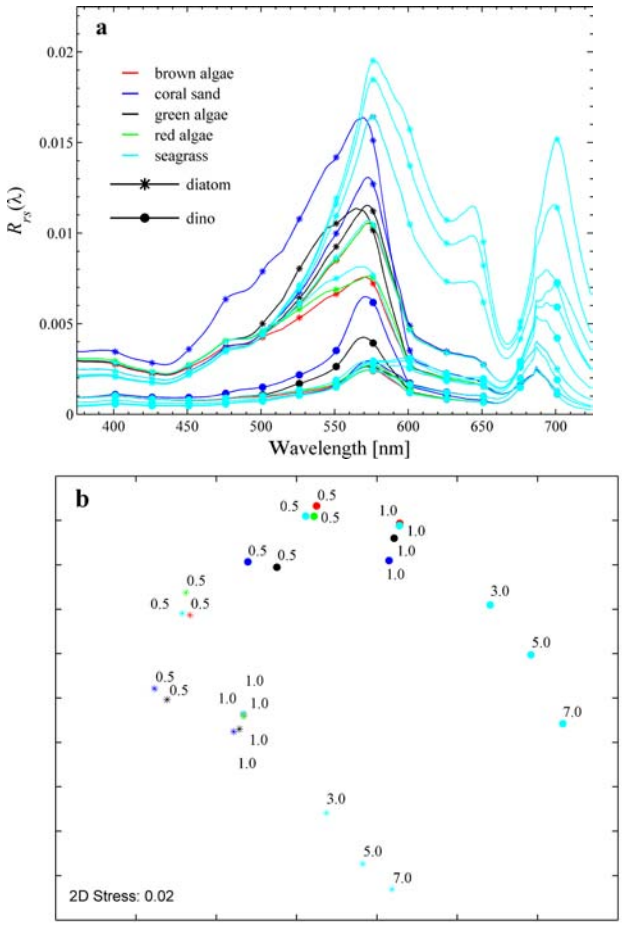


Figure 5.6: (a) Examples results of radiative transfer simulations of optically shallow waters, showing the $R_{rs}(\lambda)$ spectra corresponding to uniform vertical profiles of two phytoplankton groups at different concentrations and given different bottom types. (b) MDS plot visualizing similarities between the $R_{rs}(\lambda)$ spectra shown in (a). Symbols and colors represent each algal group and bottom type, respectively. Labels indicate the chlorophyll concentration considered in each modeled underwater scenario.

Monitoring of changes in phytoplankton taxa under different bottom conditions by remote sensing can also be complicated by the effect of another optically significant component found in estuarine waters: the suspended matter. Figure 5.7a shows the effects on $R_{rs}(\lambda)$ spectra caused by different parameters of the bottom type and the level of suspended sediment (i.e., 0, 2, 4 and 6 mg/L). The scenarios correspond to water bodies with diatoms or dinoflagellates at a fixed concentration equal to

0.5 mg/m^3 along the water column. The effect of increasing the concentration of the suspended matter is to increase absorption spectra in the blue spectral region and hence decrease $R_{rs}(\lambda)$ spectra in this region. Again, the highest $R_{rs}(\lambda)$ spectra occur for the group of diatoms regardless the bottom type and the concentration of suspended matter. Nevertheless, it is noted that bottom type has no impact on the $R_{rs}(\lambda)$ spectra for high levels of suspended sediments, cases in which $R_{rs}(\lambda)$ spectra are seen in cyan color as occured in spectra shown in Fig. 5.6a.

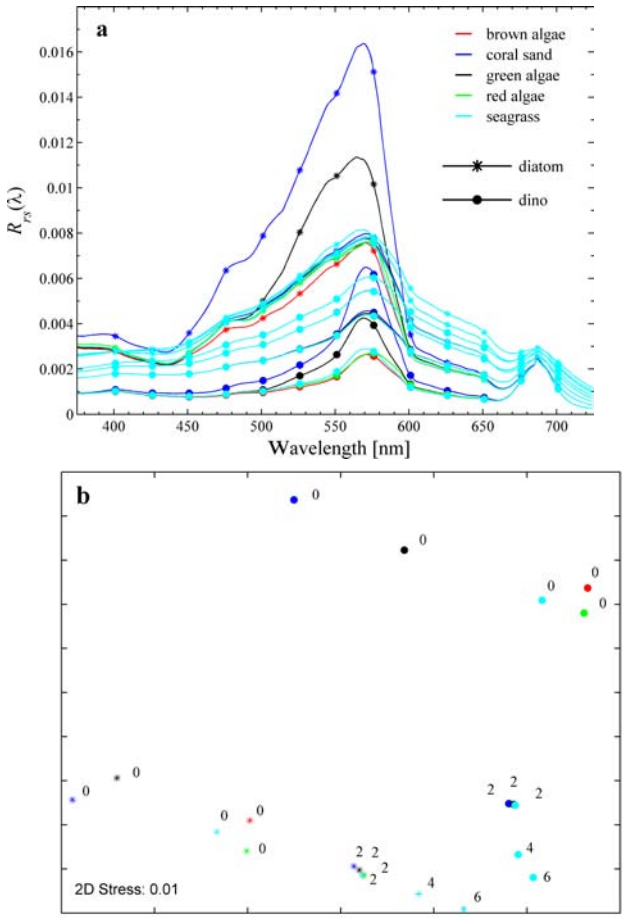


Figure 5.7: (a) Examples results of radiative transfer simulations of optically shallow waters, showing the $R_{rs}(\lambda)$ spectra corresponding to uniform vertical profiles of two phytoplankton groups at a concentration of 0.5 mg/m^3 and given different concentrations of suspended matter and bottom types. (b) MDS plot visualizing similarities between the $R_{rs}(\lambda)$ spectra shown in (a). Symbols and colors represent each algal group and bottom type, respectively. Labels indicate the concentration suspended sediment considered in each modeled underwater scenario.

The MDS plot shown in Fig. 5.7b better illustrates the dissimilarities between the $R_{rs}(\lambda)$ spectra of respective phytoplankton groups and bottom types, given different concentrations of suspended matter. By analyzing the dissimilarities in the whole spectral range, all scenarios corresponding to different phytoplankton taxa and different concentrations of suspended matter have been distinguished. In addition, it is noted that for concentrations of suspended matter smaller than 2 mg/L, scenarios characterized by different bottom types have also been clearly discriminated.

5.2.2 Effect of the vertical structure of phytoplankton on the remote sensing signal

After having assessed the influence on $R_{rs}(\lambda)$ spectra by the presence of different algal groups, levels of phytoplankton biomass, suspended matter and different bottom types, this section is aimed at evaluating how well different shallow estuarine scenarios with distinct vertical structures in terms of phytoplankton distribution and abundance can be detected through the analysis of the $R_{rs}(\lambda)$ signal.

Several underwater scenarios were generated simulating a thin layer condition taking place in a shallow estuarine area like the Alfacs Bay (see Fig. 5.3). Figure 5.8a shows the $R_{rs}(\lambda)$ spectra modeled by considering the presence of one algal group, diatoms or dinoflagellates and different combinations for the values of the background chlorophyll, Chl_0 (1 and 3 mg/m^3), depth of the chlorophyll maximum, z_{max} (3, 4, and 5 m), amplitude of the chlorophyll maximum, Chl_{max} (6, 9 and 12 mg/m^3) and a fixed value of the parameter related to the thickness of the chlorophyll peak, σ (0.25 m) . The level of suspended sediments was fixed (i.e., 2 mg/L for all scenarios), likewise the bottom type. It is also plotted the $R_{rs}(\lambda)$ spectra corresponding to uniform chlorophyll profiles when the concentration was equal to 1 and 3 mg/m^3. It can be easily seen that uniform and nonuniform vertical profiles of phytoplankton for each combination of parameters (indicated in black symbols and red lines, respectively) provide very similar, sometimes identical, $R_{rs}(\lambda)$ spectra. The corresponding MDS plot obtained after the application of the similarity-based cluster analysis (see Fig. 5.8b) also confirms that the contribution of the thin layer to the overall $R_{rs}(\lambda)$ spectra (red symbols) is nearly negligible when compared to the $R_{rs}(\lambda)$ obtained for uniform vertical profiles (black symbols). Only for two scenarios, corresponding to the case of having a background chlorophyll, Chl_0, equal to 1 mg/m^3 and a thin layer of Chl_{max} equal to 9 and 12 mg/m^3 at the upper level (i.e., z_{max} equal to 3 m), spectral discrepances allow to distinguish the nonuniform from the uniform situation.

The discrimination of different nonuniform vertical profiles of the chlorophyll concentration in shallow estuarine waters based on the analysis of hyperspectral $R_{rs}(\lambda)$ spectra has been unsuccessful. Scenarios characterized by two algal groups at different concentrations have been discriminated. Nevertheless, the analysis of the hyperspectrally-resolved $R_{rs}(\lambda)$ data set has not allowed to distinguish different vertical structures along the water column of these phytoplankton groups. Attempting to improve our results by applying derivative spectroscopy or limiting the analysis to specific spectral regions has not provided either a good performance (not shown).

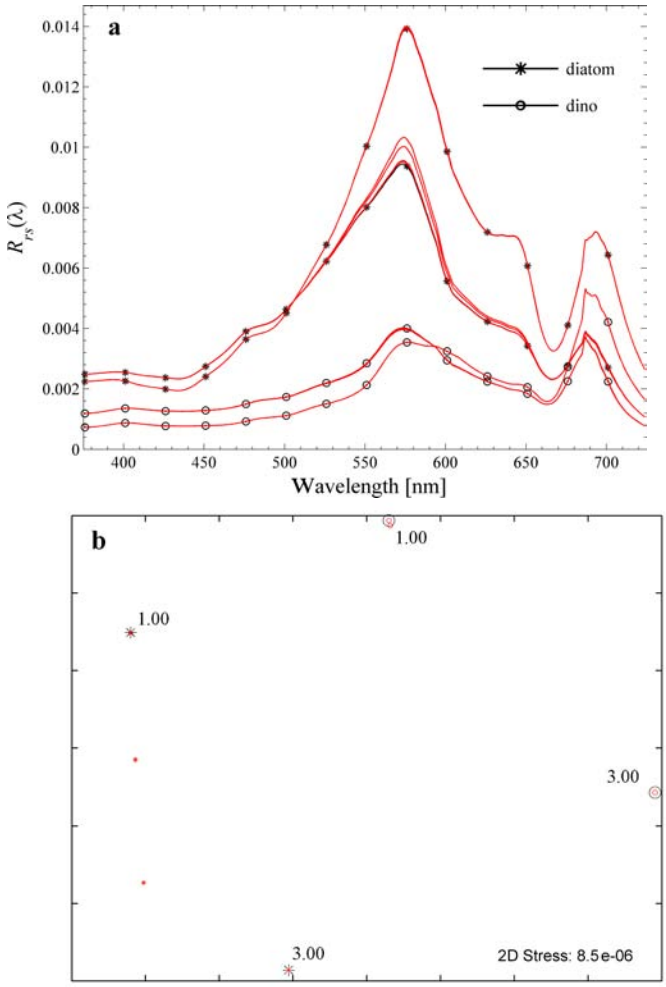


Figure 5.8: (a) Examples results of radiative transfer simulations of optically shallow waters, showing the $R_{rs}(\lambda)$ spectra corresponding to uniform (indicated in black symbols and lines) and nonuniform vertical profiles (in red) of two phytoplankton groups. (b) MDS plot visualizing similarities between the $R_{rs}(\lambda)$ spectra shown in (a).

The results of the here presented optical modeling exercise show that the vertical structure of the distribution of the phytoplankton biomass in the water column of shallow estuarine environments has little impact on the remote-sensing reflectance spectra. In this sense, as suggested by other authors (see several contributions in Babin *et al.*, 2008), a multidisciplinary approach based on highly resolved vertical profiling systems could provide fine-scale optical and hydrographic measurements so as to complement remote-sensing capabilities. A better monitoring of coastal blooms dynamics is among the advantages of obtaining a high resolution in the vertical distribution of phytoplankton. Ideally, these observation systems should resolve vertical

distributions of phytoplankton in relationship to temperature, salinity and currents. Additionally, nutrient availability should also be part of the monitoring strategy.

In case of using highly resolved vertical profilers including hyperspectral measurements, it is important to note that the proposed dissimilarity-based methodology could also be applicable. The underlying idea would be to investigate the suitability of this approach applied to periodical hyperspectral measurements along the water column in order to retrieve information about the temporal and spatial variability of the distribution of phytoplankton and other optically significant substances present in the area under study. Figure 5.9 illustrates an example in which the vertical structure of phytoplankton distribution corresponding to one modeled scenario is assessed by using the cluster-based approach. In particular, the dissimilarity analysis has been applied to a set of irradiance reflectance spectra obtained along the water column (i.e., every 0.05 m), allowing to detect the idealized thin layer of phytoplankton located at 4 m depth. In this example, the level of suspended sediments was equal to 2 mg/L and the bottom type was brown algae.

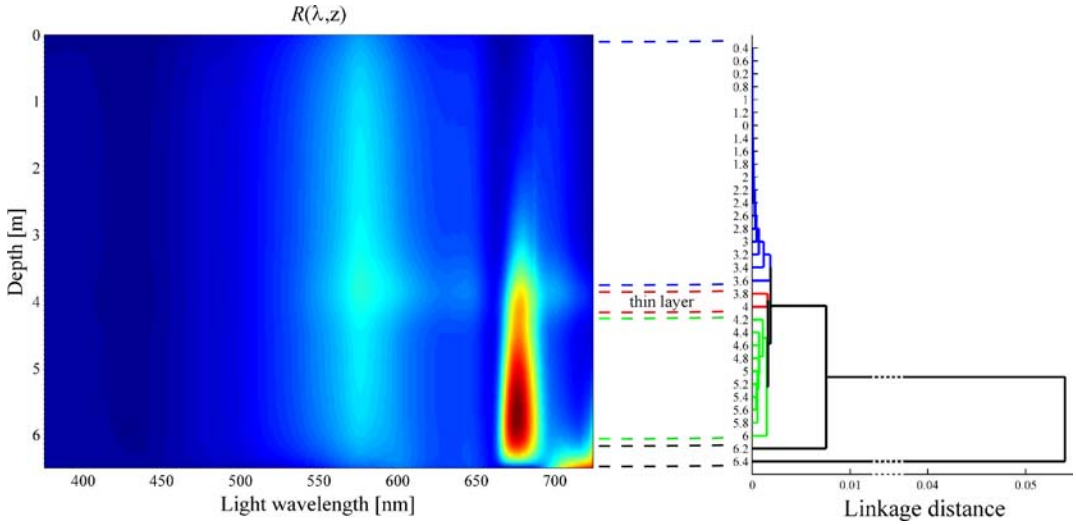


Figure 5.9: Results of cluster analysis based on the irradiance reflectance along the water column corresponding to a very shallow area characterized by the presence of a thin layer of phytoplankton at 4 m depth, obtained by means of radiative transfer simulation.

5.3 Summary and conclusions

In this chapter, hyperspectral optical signatures of shallow estuarine environments in relation to different phytoplankton comunities have been evaluated. In particular, the Alfacs Bay, an estuarine embayment of the Ebro Delta (NW Mediterranean Sea) with a maximum depth of 6.5 m, has been considered as a case study area. The presence of several harmful algal species responsible for different types of toxicity in

bivalves, mainly belonging to the groups of dinoflagellates and diatoms, have made it one of the most studied coastal sites in Catalonia.

As a proof-of-concept study and first approach to optically more complex underwater scenarios, the feasibility of a hyperspectral-based approach for monitoring different phytoplankton communities in shallow waters has been examined using an adapted simulation-based framework and the dissimilarity cluster-based approach described in Chapter 2. A set of remote-sensing reflectance spectra characterizing different shallow water scenarios have been simulated using the Hydrolight/Ecolight 5.0 radiative transfer model. The sensitivity of remote-sensing reflectance, $R_{rs}(\lambda)$, to different phytoplankton taxa has been assessed under different conditions in terms of bottom type and abundance of suspended matter (i.e., another optically significant water constituent in coastal waters besides phytoplankton). Additionally, a sensitivity analysis of remote-sensing reflectance spectra has also been performed under different conditions in terms of vertical structure of phytoplankton along the water colum. In this sense, a thin layer condition characterized by the presence of a subsurface chlorophyll maximum at a specific depth has been simulated based on field observations and theoretical estimates. The depth and amplitude of the idealized thin layer emulating similar conditions to those found in the Alfacs Bay have been varied in order to estimate their influence on the remote-sensing reflectance.

The results of these optical modeling exercises for shallow estuarine waters show that scenarios dominated by either diatoms and dinoflagellates have clearly been discriminated regardless the chlorophyll content, abundance of suspended matter, bottom type and vertical structure of the phytoplankton along the water column. In fact, the analysis of $R_{rs}(\lambda)$ spectra has also permitted to derive information about the bottom properties of each simulated underwater scenario as long as the chlorophyll concentration was not larger than 1 mg/m^3 or the concentration of suspended matter was smaller than 2 mg/L. On the contrary, the contribution of a idealized thin layer in the water column to the remote-sensing signal has been shown to be nearly negligible when compared to $R_{rs}(\lambda)$ spectra obtained for analogous uniform vertical profiles. Based on these results, the use of vertical profilers to observe fine-scale characteristics along the water column in terms of phytoplankton diversity is suggested in order to complement the remote-sensing approach, which has been proved unsuccessfull for this purpose. In this sense, the Marine Technology Unit (UTM, CSIC) is carrying out the project ANERIS[1], conceived to develop a new oceanographic profiler with high resolution autonomous sampling capabilities. In particular, it is being designed to gather fine-scale profiles of biological activity during the free-fall descent by using hyperspectral optical sensors, among others. As a preliminary calibration task and in parallel to the work presented in this thesis, the influence of the spectral stray-light radiation to the measured output signals of the integrated hyperspectral sensors has been characterized and corrected (see more details in the Appendix C). The

[1] Project ANERIS: ANálisis y desarrollo de una sonda ocEanogRáfica Inteligente con capacidad autónoma de obtención de muestraS, supported by the Spanish National Research Council (CSIC), PIF08-015.

spectral stray-light radiation is described as the background radiation that has been scattered due to imperfections in the dispersive element and other optical elements of the instrument.

Current uncertainties in the knowledge of marine optical characteristics and the ability to model the underwater radiation fields in a complex estuarine region like the Alfacs Bay indicate that much additional research is still needed. In this sense, despite the results indicate that hyperspectral observations might play a key role in the characterization of shallow estuarine environments, there is a strong need for more in situ data to have a better knowledge of the optical variability of phytoplankton and other significant constituents in this area. Several research projects (i.e., MESTRAL[2] and PHYTOSCOPE[3]) have recently been initiated within the framework of the Marine Technology Unit (UTM, CSIC), in order to address this issue. The characterization of the extremely high spatial and vertical variability in biomass during bloom events through the observation of its impact on the in situ and remote sensing optical properties will be one of the main goals. Additionally, new methods to maximize the potential of using hyperspectral data to effectively retrieve important marine properties in coastal environments will also be explored.

A combination of optical and hydrographic measurements will be essential for a better understanding of the factors regulating phytoplankton community composition in the Alfacs Bay. However, based on a previous field experience in this area in 2011[4] and other past initiatives performed by other research groups (Tzortziou *et al.*, 2006; Rhea *et al.*, 2007; Hommersom *et al.*, 2009) in other areas, it is expected that these new monitoring initiatives will involve many specific challenges in terms of instrument calibration and data processing, specially due to the shallowness, high dynamics and turbidity of this type of water environments.

5.4 Acknowledgements

We are grateful to M. L. Artigas and E. Berdalet for supplying chlorophyll and suspended matter concentrations characteristic for the Alfacs Bay (Ebro Delta, NW Mediterranean Sea). We also thank O. N. Ross, S. Ramirez and S. Pérez for helping with the field work performed in the Alfacs Bay in 2011, partially supported by the AQUALIGHT project funded by the EU Marie Curie Actions Program (FP7-PEOPLE-2009-IEF).

[2] Project MESTRAL: Modelling and advanced observational tEchnologies to link tranSport processes, opTically-active constituents, and wateR light-field vAriability in a coastaL ecosystem, supported by the Spanish Ministry of Science and Innovation, CTM2011-30489-C02-01.

[3] Project PHYTOSCOPE: PHYToplankton biOdiversity multiScale Characterization using advanced OPtical technologiEs, supported by the Total Fondation, BIO-2012-026 .

[4] Field experience resulted from a cooperation with the researchers J. Busch and O. Zielinski (Carl von Ossietzky Universität Oldenburg, UNIOL). Data still under analysis.

6 Summary and outlook

Over the last decades, the open and coastal ocean science community has expressed the urgent need to improve the observation of marine ecosystems for receiving a better knowledge about the space-time distribution and dynamics of ecological parameters such as phytoplankton communities, including critical bloom-forming algal groups. In attempting to address this question, in situ and remotely-sensed spectrometric optical observations of ocean waters have demonstrated to provide information regarding the concentrations of optically significant constituents in seawater, in particular, regarding phytoplankton community structure. The optical-based approach represents a suitable, fast and non-intrusive method currently available for measuring properties of ocean ecosystems at local and global scales. Moreover, it is noted that technological advances and especially the advent of high spectral resolution (hyperspectral) sensors have raised new expectations about the possibilities of discriminating phytoplankton community composition in the ocean. The advantage of hyperspectral data stems from the fact that more subtle differences in targets are captured, opening the possibility to achieve a species-level discrimination, beyond the estimation of only chlorophyll-*a* that is common to all taxonomic groups.

This PhD thesis has been carried out with the aim of improving our ability to extract information regarding phytoplankton community structure in the ocean by developing and evaluating a novel processing approach based on hyperspectral data. The work has been organized along several case studies, conducted to address several specific questions and elucidate the feasibility of hyperspectral oceanographic observations to discriminate different phytoplankton communities in open ocean and coastal environments:

- Answer to the question: *what role can the shape of hyperspectral data play in the assessment of phytoplankton composition of a seawater sample?* is provided in Chapter 2.

  A successful identification of different phytoplankton communities in the ocean has been achieved by applying a novel methodology to a hyperspectral data set of remote-sensing reflectance spectra corresponding to different underwater scenarios in terms of phytoplankton composition. In order to do so, a key issue has been the application of a dissimilarity-based cluster technique, which accounts for complete spectral behaviour of hyperspectral data, in combination with derivative spectroscopy, which exploits the spectral shape features of hyperspectral data. As a first approach, the suitability of this methodology has been demonstrated based on a simulation-based framework, implying the use of a synthetic hyperspectral data set obtained by radiative transfer modeling.

- Answer to the question: *how effective can a hyperspectral-based approach be at discriminating different oceanic environments in which various phytoplankton groups co-exist under non-bloom or non-dominant conditions?* is given in Chapter 3.

  In this case, the methodology described in the previous chapter has been applied to more realistic environmental scenarios based on field optical data determined at 9 stations along a north-to-south transect in the eastern Atlantic Ocean in 2005. Different real open ocean environments corresponding to non-bloom conditions have successfully been classified by applying the cluster analysis to second derivative spectra of different hyperspectral data sets (i.e., absorption and remote-sensing reflectance spectra). As part of this study, a new validating tool has been proposed and proven useful to illustrate the effectiveness of the optical-based classification. This novel approach is based on the pigment composition analyzed in conjuction with concurrently obtained optical data, information which has been commonly used by the scientific community as a proxy for the phytoplankton composition. In particular, two indices have proven valid to quantify the degree of similarity between the optical-based clusters and the pigment-based clusters (the "sea truth").

- Answer to the question: *how feasible is the global identification of different types of water masses in terms of phytoplankton composition by assessing their hyperspectral optical signatures?* is provided in Chapter 4.

  The potential of the methodology described in the previous chapters has been demonstrated by analyzing a larger database of hyperspectral observations collected at 48 stations in the eastern Atlantic Ocean in 2008. The identification of several phytoplankton assemblages has served to demonstrate the feasibility of this approach to classify oceanic areas of similar biogeography based on the phytoplankton community structure and their hyperspectral optical signatures (i.e., absorption and remote-sensing reflectance spectra). In this sense, a potential application of the proposed methodology has been identified: the establishment of different bio-optical provinces from the dissimilarity-based analysis applied to in situ and remotely-sensed hyperspectral oceanographic observations and their second derivative spectra.

- Answer to the question: *what effect does a hyperspectral-based approach have on the analysis of phytoplankton community structure of both types of environments: open ocean or coastal waters?* is provided in Chapter 5.

  The role of hyperspectrally-resolved optical observations and the dissimilarity-based methodology, successfully applied in open ocean environments, has also been examined in optically more complex underwater scenarios. As a proof-of-concept study and first approach to shallow estuarine environments, several underwater optical scenarios have been generated by means of model simulations, emulating similar conditions to those found in a selected case study area: the Alfacs Bay (Ebro Delta, NW Mediterranean Sea). Remote-sensing

reflectance spectra corresponding to different scenarios dominated by two phytoplankton groups (i.e., dinoflagellates and diatoms) at different concentrations have been discriminated given several assumptions, i.e., different bottom types, concentrations of suspended matter and vertical structures of phytoplankton along the water column.

The following paragraphs provide a summary of the key findings:

An optimal application of the proposed cluster-based methodology for estimating variability of marine phytoplankton communities from the analysis of hyperspectral data has implied the selection of an appropriate distance as a measure to determine the degree of similarity between each pair of input objects. For the pigment-based cluster partition, an Euclidean metric distance has been utilized. However, for the analysis of spectral data, an angular distance has been found more effective to recognize similar or dissimilar optical patterns from their shape features. Shape singularities in hyperspectral oceanographic data are relevant since they are related to absorption features of pigments present in the samples and hence, to their phytoplanktonic composition. Regarding this issue, the previous application of derivative spectroscopy has proven to be essential to enhance subtle features in hyperspectral data. A good performance has been possible after the suitable selection of the involved smoothing and derivative parameters (i.e., filter size and band separation).

As a novelty in this thesis, a validation approach has been proposed with the aim to automatically evaluate the usefulness of optical data for discriminating phytoplankton pigment assemblages. This validation approach is based on two similarity indices: cophenetic and Rand, which quantify the similarity between different optical-based cluster partitions and the pigment-based cluster partition. The defined reference based on the pigment composition (the "sea truth") has proven valid as a proxy for phytoplankton community composition. It is noted that the use of these indices has provided a valuable methodology for performing sensitivity tests in order to identify optical variables and the spectral ranges most suitable for characterizing different phytoplankton assemblages.

The most promising results from the cluster analysis applied to open ocean stations were obtained with the second derivative spectra of the phytoplankton absorption coefficient, $a^*_{n,ph}(\lambda)$, over the spectral range 370-725 nm (or narrower spectral regions from within that range, e.g., 435-520 nm), and the second derivative spectra of the remote-sensing reflectance, $R_{rs}(\lambda)$, over the spectral range from about 435 nm to 580 nm. In addition, other absorption and reflectance data were examined but they generally showed either more limited value or no usefulness at all for discriminating phytoplankton pigment assemblages (e.g., ordinary reflectance spectra at 1 nm resolution or multispectal reflectance data).

The results from the model-data "closure exercise" required for the reconstruction of hyperspectral $R_{rs}(\lambda)$ corresponding to stations from the eastern Atlantic Ocean in 2005 (Chapter 3) compared reasonably well with field multispectral $R_{rs}(\lambda)$ measurements. The level of consistency between model and measurements indicates that the suite of parameters used as input to the radiative transfer simulations realistically

represented the actual field conditions. It is noteworthy that the ability to define realistic inputs derived, in turn, from a comprehensive suite of IOP measurements carried out in situ and in the laboratory. These results serve to demonstrate that the conditions under which theoretical radiative transfer calculations of hyperspectral $R_{rs}(\lambda)$ were performed produced a close agreement with the experimental multispectral measurements of $R_{rs}(\lambda)$.

The identification of different bio-optical provinces from the analysis of 48 open ocean stations in the eastern Atlantic Ocean (Chapter 4), as one of the potential applications of the proposed methodology, has led to examination of its biogeographical relevance by comparison to ecological provinces previously proposed in the literature (Longhurst, 2006). The bio-optical provinces showed to agree well with these provinces and thus could be used to classify areas of similar biogeography. In fact, the use of hyperspectral observations in combination with other physical parameters (e.g., the sea-surface temperature) is suggested. In terms of globally significant issues such as carbon export and primary production, the hyperspectral-based approach represents a possibility for a faster and detailed assessment of the state of a temporally and spatially variable marine environment. At the moment, other approaches have already been proposed to detect simultaneously Phytoplankton Functional Types (PFTs) from satellite hyperspectral data (Bracher *et al.*, 2009). However, PFT identification is a slightly different concept than identifying different spatially distributed phytoplankton assemblages. The novel methodology proposed in this thesis allows to define different bio-optical provinces on the basis of phytoplankton community structure and their bio-optical traits, instead of a small set of functional groups.

The results from the optical modeling exercises for shallow estuarine waters have shown that scenarios dominated by either diatoms and dinoflagellates were clearly discriminated regardless of the chlorophyll content, abundance of suspended matter, bottom type and vertical structure of the phytoplankton along the water column. In fact, the analysis of $R_{rs}(\lambda)$ spectra also permitted to derive information about the bottom properties of each simulated underwater scenario as long as the chlorophyll concentration was not larger than 1 mg/m^3 or the concentration of suspended matter was smaller than 2 mg/L. On the contrary, the contribution of a idealized thin layer in the water column to the remote-sensing signal showed to be nearly negligible when compared to $R_{rs}(\lambda)$ spectra obtained for analogous uniform vertical profiles.

Overall, this thesis demonstrate how hyperspectral oceanographic observations can be used to better assess phytoplankton biodiversity and dynamics, furthering the understanding of the role of each algal group at a local scale or in the global marine ecosystem and biogeochemical cycles. It is noteworthy that the proposed approach is generally applicable to different data sets (i.e., in situ and remotely-sensed Inherent or Apparent Optical Properties) or other types of data (e.g., combination of optical and hydrographic data), gathered at the sea surface or along the water column. A similar optimization analysis can also be used to provide the best performance for a given data set. Moreover, this analysis is flexible and permits different optimal values to be utilized in the case of other optical data with differing spectral resolutions.

In terms of **future work**, the following points seem worthwhile to pursue as part of further investigations:

- The results from the assessment of phytoplankton communities in open ocean environments, even encouraging, are far from definitive. In fact, further investigation is needed with larger databases of measurements from various oceanic environments in order to determine the generality of the presented approach and the specific set of optimal parameters. The minimum data necessary for application of the dissimilarity-based proposed methodology is information regarding either the hyperspectral phytoplankton absorption coefficient or the remote-sensing reflectance in the blue to green region of the spectrum, which would yield an initial classification of stations based on similarities and differences in the optical spectra or their derivative spectra. Given significant interest in the development of the capabilities for large-scale characterization of phytoplankton biodiversity from optical measurements including remote-sensing observations, one may expect further expansion of comprehensive databases consisting of concurrent pigment and hyperspectral optical information in the near future. The availability of such a larger data sets will support validation and optimization of this dissimilarity-based approach and other classification techniques, such as unmixing or neural networks, which may exploit hyperspectral optical measurements as a source of information on phytoplankton community composition. In order to address these issues, the research project PHYTOSCOPE (PHYToplankton biOdiversity multiScale Characterization using advanced OPtical technologiEs) has recently been initiated within the framework of the Marine Technology Unit (UTM, CSIC) and in coordination with the Phytooptics Group (Alfred-Wegener Institute for Polar and Marine Research, AWI, and Institute for Environmental Physics of the University Bremen).

- Concerning the characterization of more complex shallow estuarine waters, there are some important aspects that may be interesting to explore. In particular to Alfacs Bay, current uncertainties in the knowledge of marine optical characteristics and the ability to model the underwater radiation fields in such a complex region indicate that more additional research is still needed. Despite the results from the performed modeling exercise indicate that hyperspectral observations might play an important role in the characterization of optically significant water constituents (i.e., phytoplankton, suspended matter and CDOM), there is a strong need for more in situ optical and hydrographic data. This observational effort will provide important insights into the factors regulating phytoplankton community composition in this area and will help to explore the full potential of hyperspectral observations. To overcome this challenge, the research projects: PHYTOSCOPE and MESTRAL (Modelling and advanced observational tEchnologies to link tranSport processes, opTically-active constituents, and wateR light-field vAriability in a coastaL ecosystem) have recently been initiated within the framework of the Marine Technology Unit (UTM, CSIC).

Appendix A

Radiative transfer equation in the ocean[1]

Radiative transfer (RT) theory aims to describe the interaction of light with matter and to quantify all the processes that affect the direction and the quantity of photons. In particular, the change of the light field under water is due to several processes and is described by the so called Radiative Transfer Equation (RTE). This appendix is devoted to explain in detail the derivation of this unique fundamental equation, which enables to link the two categories of optical properties (i.e., inherent and apparent optical properties) defined for a water body as it has been shown in Fig. 1.8. The description of these two categories (i.e., IOPs and AOPs) and the involved properties can be found in Section 1.1.

The path of the radiance $L(z, \theta, \varphi, \lambda)$ through a thin layer of water of thickness $dz = z_2 - z_1$ is shown schematically in Fig. A.1. The zenith and azimuthal angles θ', φ' and θ, φ denote the direction of the incoming and outgoing light beam, respectively. The processes responsible for the variation of $L(z, \theta, \varphi, \lambda)$ in a water body are primarily extinction processes (i.e., absorption or scattering of photons out of the light beam) and intensivation by scattering into the direction of light propagation or by emission. In this sense, it can be considered as a process of balance of light loss and gain.

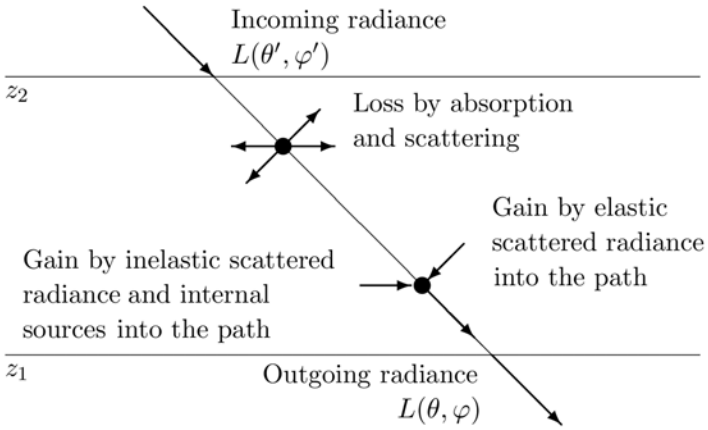


Figure A.1: Path of the radiance and influences of absorbing and scattering particles in a thin layer of water (*Source*: Albert, 2004).

[1] The description of the basic processes provided in this appendix are described in the following textbooks: Mobley, 1994 and Kirk, 1994.

The RTE is defined by the following integro-differential equation[2]:

$$\cos\theta \frac{dL(z,\theta,\varphi,\lambda)}{dz} = -c\,(z,\lambda)\,L\,(z,\theta,\varphi,\lambda)$$
$$+ \int_{4\pi} L\,(z,\theta',\varphi',\lambda)\cdot\beta\,(z;\theta',\varphi'\rightarrow\theta,\varphi;\lambda)\,d\Omega' + S\,(z,\theta,\varphi,\lambda) \tag{1.12}$$

This equation expresses that the change of a specific radiance $L(z,\theta,\varphi,\lambda)$ along an elementary distance $dz/\cos\theta$ results from:

- the attenuation (i.e., a loss) along the elementary slant path. It, thus, involves the beam attenuation coefficient $c(z,\lambda)$.
- the gain by elastic scattering into the direction of interest (θ,φ) from all other directions about the element. It depends on the scattering phase function $\beta\,(z;\theta',\varphi'\rightarrow\theta,\varphi;\lambda)$[3], which represents the scattering probability into a specific direction of the solid angle $d\Omega'$.
- the gain by inelastic scattering from other wavelengths and internal light sources into the direction of interest (θ,φ). It is indicated as a source term $S\,(z,\theta,\varphi,\lambda)$.

Among all the processes involved in the RTE equation, the process of scattering can be described as the deviation of a photon from a straight trajectory in the medium through which it passes. It takes place in natural water due to the interaction of photons with water molecules and water constituents. The influence of each constituent on the scattering process depends on wavelength, particle size, concentration and refractive index. When the total energy of the scattered photon is conserved, (i.e., it is scattered at the same wavelength, λ, as the incoming photon), it is called **elastic scattering**. However, when the energy is not conserved (i.e., it is scatterd at another wavelength, λ'), it is called **inelastic scattering**. In this last process, the scattered wavelength is longer than the absorbed ($\lambda' > \lambda$) and therefore some energy is transformed in contrast to elastic scattering, where only the direction of the incident radiation is changed.

Regarding the source term, $S\,(z,\theta,\varphi,\lambda)$, the first component is related to two types of inelastic processes: **Raman scattering** and **fluorescence**. The result of both processes is in essence the same: a photon with the wavelength (or frequency) different from that of the incident photon is produced and the molecule is brought to a higher or lower energy level. Nevertheless, the Raman effect differs from the process of fluorescence. For the latter, the incident light is completely absorbed and

[2] It is considered the situation of horizontally homogeneous water and time independence. The dependence on θ and φ is noted and combined by the symbol Ω.

[3] The angle between any direction from the surrounding space and the direction of interest is denoted $(\theta',\varphi'\rightarrow\theta.\varphi)$.

the system is transferred to an excited state from which it can go to various lower states only after a certain resonance lifetime. Another difference is that the Raman effect can take place for any wavelength of the incident light. In contrast to the fluorescence effect, the Raman effect is therefore not a resonant effect. In practice, this means that a fluorescence peak is anchored at a specific excitation wavelength, whereas a Raman peak maintains a constant separation from the excitation wavelength. Raman scattering in natural water is mainly caused by the water molecules. In contrast, fluorescence is the inelastic process ocurring in pigments contained in the water constituents such as phytoplankton and CDOM.

The second component of the source term $S(z, \theta, \varphi, \lambda)$ corresponds to the process of **bioluminiscence**. It is a internal light source where light is emitted by organisms in the water. The quantity of light emitted, and produced by chemical reactions of bacteria, phytoplankton and fish, is very small compared to the incident sun light. Therefore, its impact on radiance is neglected at the surface level but considered as the only natural source of light for deep waters.

It is important to note that despite the compactness of the RTE equation, its solution is not straightforward. Nevertheless, mathematical and numerical techniques have been developed to derive quasi-exact solutions, which in fact have been revealed extremely accurate. For instance, the RT model Hydrolight/Ecolight version 5.0, which is described in Section 1.4 and utilized within the framework of this thesis, employs invariant imbedding techniques to solve the RTE equation.

Appendix B

Independent corroboration for the pigment analysis

The Center for Hydro-Optics and Remote Sensing (CHORS) at San Diego State University in California was the designated HPLC analytical facility for NASA-sponsored investigations during the period encompassing the ANT-XXIII/1 cruise from 2005. It was subsequently determined that the measurements performed by CHORS during this time were plagued with a number of methodological problems, which potentially led to errors in the determinations of some pigments. Because these problems and associated uncertainties changed with time, the NASA team tasked with investigating these data quality problems at CHORS deemed that retrospective corrections for the entire span of several years is not feasible (Hooker and Heukelem, 2009). Their final recommendation to investigators was that use of the data be accompanied by a statement that "These data are not validated and should not be used as sole basis for scientific results or conclusions. Independent corroborating evidence is required".

Because all of the samples were analyzed as part of a single run, ad hoc corrections were developed for a subset of specific individual pigments using a combination of field data and laboratory standards measured during the SeaHARRE-3 intercalibration experiment (C. Trees, personal communication; Hooker *et al.*, 2009). These pigments included both the monovinyl and divinyl forms of chlorophyll-*a* (MVChl*a*, DVChl*a*) which are used in the calculation of total chlorophyll-*a* (TChl*a*), monovinyl chlorophyll-*b* (MVChl*b*), and the accessory pigments β - carotene and alloxanthin:

$$MVChla^{corr} = 0.598 \cdot MVChla^{uncorr} + 0.0073$$

$$DVChla^{corr} = 0.655 \cdot DVChla^{uncorr} + 0.0003$$

$$MVChlb^{corr} = 0.5682 \cdot MVChlb^{uncorr} + 0.001$$

$$\beta - carotene^{corr} = 0.7 \cdot \beta - carotene^{uncorr}$$

$$alloxanthin^{corr} = 0.8 \cdot alloxanthin^{uncorr}$$

where the superscripts corr and uncorr refer to corrected and uncorrected data, respectively, and concentrations are in units of $mg\,m^{-3}$. Other pigments did not show consistent relationships in the intercalibration results, and no corrections were attempted for them. It is cautioned that the corrections should not be used indiscriminately with other data sets affected by the CHORS data quality problems.

The values of TChl*a* calculated using the corrected concentrations of MVChl*a* and DVChl*a* exhibit reasonable agreement with fluorometrically-derived chlorophyll measurements, and provide realistic estimates of chlorophyll-specific phytoplankton absorption coefficients within the red peak of chlorophyll-*a*. In a comparison of the

corrected CHORS results with independent pigment determinations done by GKSS, some differences were observed in the concentrations of individual pigments between the two data sets. For example, the sum of MVChl*a* and DVChl*a* was generally higher (on average by 20% with a standard deviation of 23%, number of samples 25) for CHORS compared with GKSS. However, the CHORS data yielded more reasonable estimates of chlorophyll-specific phytoplankton absorption within the red peak of chlorophyll-*a*.

Despite such differences in the estimates of some individual pigment concentrations, with regard to the present application it is important to note that both laboratories provided similar characterization of samples in terms of the relative pigment composition as described by ratios of various individual pigments to TChl*a*. Both sets of HPLC analyses generally yielded the same dominant accessory pigments present at any given station, and similar trends in the pigment ratios among the stations. This is an essential result for the present study because in the cluster analysis only pigment ratios were utilized, and not individual pigment concentrations.

A series of analyses were performed to test for any differences in the results when either the CHORS or GKSS pigment data sets were used to generate the reference pigment dendrogram. For the nine stations selected in this part of the analysis, the cluster techniques described in Section 3.2 applied independently to the CHORS and GKSS sets of HPLC pigment ratios yielded a very similar partitioning of stations into clusters. This high degree of similarity between the two pigment-derived dendrograms exhibits a cophenetic index value of 0.92, indicating that overall pairwise distances between the data objects are well-preserved. This observation is further supported by additional cluster analyses comparing both pigment data sets with optical data such as the normalized phytoplankton spectral absorption coefficient (see Fig. B.1).

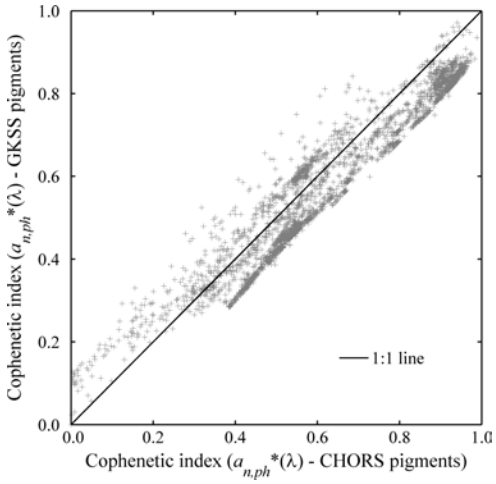


Figure B.1: Scatter plot of the cophenetic indices between the second derivative spectra of $a^*_{n,ph}(\lambda)$-based and the pigment-based cluster trees. The x- and y-axis indicate the indices obtained using the CHORS and GKSS HPLC analysis, respectively.

Comparisons of these absorption-based clusters with pigment-based clusters were found to be similar when either the CHORS or GKSS pigment data were used. In particular, Figure B.1 depicts how the values of the cophenetic index obtained from the analysis of similarity between the cluster trees of pigments and the second derivative spectra of normalized phytoplankton absorption are closely correlated between the two pigment data sets. The correlation coefficient is 0.97. The results include the cophenetic indices evaluated for all combinations of spectral ranges of the second derivative spectra of $a^*_{n,ph}(\lambda)$ as shown in Fig. 3.12.

In summary, the comparisons of two independent HPLC pigment data sets indicate that the results and conclusions of this study are robust and independent of the choice of pigment data used in the creation of the reference pigment dendrogram for the cluster-based similarity analysis of pigment and optical data. Therefore, it was chosen the one set of pigment results from CHORS for all subsequent analyses in the study.

Appendix C

Stray-light correction of in-water array spectroradiometers

In order to improve our understanding of phytoplankton biodiversity and dynamics in the ocean, the Marine Technology Unit (UTM, CSIC) is carrying out the multidisciplinary project ANERIS[4], in which this thesis has been developed. The project involves the development of a new observational device, i.e., an intelligent oceanographic profiler with high resolution autonomous sampling capabilities. In particular, it is being designed to gather fine-scale profiles of biological activity during the free-fall descent by using hyperspectral sensors, among others. This instrument is intended to be useful for studying different oceanic environments, covering a wide range of temporal and spatial scales where different events may happen and for validating remotely-sensed hyperspectral ocean data collected by airborne imagers.

Within the framework of the project ANERIS, when hyperspectral measurements will be used for further analysis, one of the most relevant issues to take into account is the uncertainty of the measurement system. Appropriate calibration strategies should be followed, since the accuracy of the measurements relies on the associated uncertainties of the measurement system. A well-calibrated spectroradiometer includes a characterization process, describing the instrument behavior in terms of the spectral sensitivity, the signal-to-noise ratio, the dark current, the wavelength calibration, the nonlinearity, the temperature dependence of the measurements and the spectral scattering, or what is called the spectral stray-light which will be described in detail below. In particular to ocean optics applications, in which light conditions can be extremely low, special attention should be paid to corrections for the inherent distortions of the hyperspectral sensors because the errors in the measured radiance distributions may be potentially significant and lead to inaccurate retrievals of water properties. This issue becomes even more important when the commonly derivative spectroscopy is used, which explores subtle features in shape of hyperspectral data (see Chapters 2, 3 and 4).

Another reason to conduct an accurate characterization of hyperspectral sensors is when a radiative transfer simulation-based approach is to be used as a basis for further validation of some processing technique (see the Hydrolight-based approach in Chapter 2). With the goal of exploring the potential of any processing technique, the response of the sensor with which the optical measurements would be acquired

[4] Project ANERIS: ANálisis y desarrollo de una sonda ocEanogRáfica Inteligente con capacidad autónoma de obtención de muestraS, supported by the Spanish National Research Council (CSIC), PIF08-015.

should be taken into account. More realistic retrievals could be carried out from hyperspectral oceanographic data if the effect of the sensor in terms of noise, sensitivity, spectral resolution or stray-light would be included in the simulation-based approach.

Hyperspectral sensors

Hyperspectral sensors generally use diffraction gratings or linear variable optical filters as a dispersive element to separate light into specific wave bands centered on desired wavelengths, and different scanning mechanisms to generate high resolution spectra or images. Mechanical-scanning spectrometers have been traditionally employed to obtain a spectrum by using a single photo-detector and rotating the dispersive element over a given specific spectral region after passing through the sample. However, more recent spectrometers known as spectrographs use a fixed grating and multi-element array detectors that allow an entire spectrum to be acquired over some finite spectral region simultaneously.

Array spectrometers are being widely used as a tool for rapid measurements of spectral distributions in oceanographic applications, in which acquisition speed is an important issue. Spectra can be measured using three types of arrays: photodiode, charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS). High performance is available in all technologies today when they are designed properly, and each have their own strengths and weaknesses. For instance, CCD arrays show a slightly higher dynamic range (sensitivity) and a lower system noise than CMOS ones, whereas CMOS ones offer more integration, smaller system sizes and higher speeds. The proper selection of the type of array depends on many parameters (e.g. pixel dimensions, sensitivity, spectral range coverage, dynamic range, saturation exposure, integration time) and on the specific application. Other important advantages of array spectrometers are the non moving parts, the robustness and the low production costs. A key challenge has also been to construct small size array spectrometers without compromising performance. However, they involve several drawbacks compared with mechanical-scanning spectrometers, such as the fixed wavelength resolution, the lower sensitivity and the higher stray-light radiation (described in more detail below), due to the lack of an output slit and the integral illumination over the full wavelength range.

New technologies and the miniaturization of electro-optical components have permitted the development of accurate, low-cost and energy-efficient hyperspectral sensors designed to measure oceanic optical properties. In the design and development of an instrument for hyperspectral measurements in the ocean, it is necessary to consider several optimal characteristics required of its detection system. The wavelength coverage should be broad, ideally from 400 to 700 nm, but occasionally slightly extended towards the near infrared and near ultraviolet regions. A low-light-level detector should be selected with a large dynamic range. Therefore, its responsivity should be high over the whole spectral region covered, especially in the blue and red spectral portions, and the dark signal should be low. Furthermore, in order to minimize a variety of external perturbations during measurement (e.g. changing sky condi-

tions, surface wave noise producing high-frequency fluctuations in irradiance, ship shadow effect), the scan time of the system should be fast. The fastest devices are multi-element array spectrometers with a fixed grating system (i.e., spectrographs).

Two commercially available miniature hyperspectral spectrographs were initially considered to become part of the instrument developed in the project ANERIS (see Fig. C.1). The first is the Ocean Optics USB4000 Spectrometer, a new device commonly used for precise measurements and calibration tasks. The USB4000 uses a Toshiba TCD1304AP 3648-element one-dimensional CCD-array detector, has a pixel-to-pixel spacing of approximately 0.24 nm and a FWHM bandwidth of approximately 6 nm. The second considered spectrograph is the Boehringer Ingelheim MicroParts GmbH UV/VIS Microspectrometer, a lower cost and energy-consuming device. The MicroParts uses a Hamamatsu S8378 256-element one-dimensional CMOS-array detector, has a pixel spacing of 3.5 nm and a FWHM bandwidth smaller than 10 nm. The analog-to-digital conversion resolution of both instruments is 16 bits.

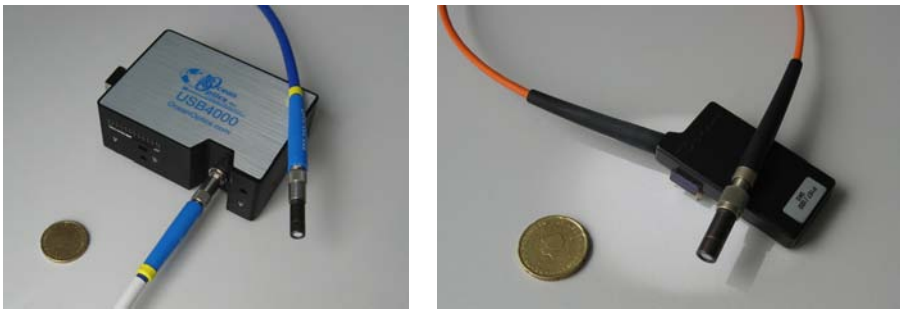


Figure C.1: Miniature hyperspectral sensors. (left) Ocean Optics USB4000 Spectrometer (CCD-array detector, dimensions $89.1 \times 63.3 \times 34.4$mm). (right) Boehringer Ingelheim MicroParts GmbH UV/VIS Microspectrometer (CMOS-array detector, dimensions $54 \times 32 \times 9.5$mm).

Data quality assurance and stray-light correction

When hyperspectral measurements must be performed and are to be used for further analysis, appropriate calibration and pre-processing strategies must be undertaken, since the accuracy of the measurements relies on the associated uncertainties of the measurement system. The clearest example in optical oceanography is perhaps the in situ radiometric measurements carried out during several field campaigns, which are currently essential to provide a proper calibration and validation of remote sensing measurements collected by ocean color satellites and airborne platforms (see more details from some existing initiatives in Zibordi *et al.*, 2002 and Clark *et al.*, 2003). A well-calibrated high spectral resolution array-based spectrometer, for both in situ and remote sensing applications, should include likewise its characterization and an evaluation of all meaningful sources of uncertainty (Voss *et al.*, 2008). It is therefore always necessary to describe the instrument's behavior in terms of responsivity, signal-to-noise ratio (SNR), dark current, nonlinearity, temperature dependence of measurements and spectral scattering, or what is called the spectral stray-light of an instrument.

With regard to the project ANERIS, the two considered spectrographs have multielement array detectors and its dispersive element is a single grating which is fixed (Palmer and Loewen, 2005). This fact makes the spectral selection to be determined by the image of the entrance slit onto a reference plane where the multi-element array detector is placed, as well as by the size of the individual elements in the detector array. Ideally, an image element on a pixel of the detector array for a particular wavelength is composed only of the spectral components of the source element within the instrument's bandpass at the particular wavelength. In practice, the image is modified by the presence of stray or scattered light. The **spectral stray-light** is described as the background radiation that has been scattered due to imperfections in the dispersive element and other optical elements of the instrument (surfaces, internal baffles, higher-order diffraction, etc.).

In ocean optics research, this unwanted radiation called stray-light can cause potentially significant errors in the measured radiances given that light conditions can be extremely low. These errors can be specially large when measuring spectral regions (e.g. blue or UV regions) where the spectrograph's response and the signal-to-noise ratio are small. In order to reduce measurement uncertainties when using the proposed spectrographs, correction of the instrument's response for measurement errors arising from the instrument's spectral stray-light must be considered. A key aspect when correcting the spectral stray-light is the spectral shape of the lamp-based calibration source used. For underwater light field measurements, differences between the spectral distribution of the calibration source and the test source (e.g. the Sun) should be avoided in spectral stray-light calibration.

With the aim to study and correct the effect of stray-light radiation on the response of the two array-based spectroradiometers considered in the project ANERIS for underwater radiometric measurement purposes, the spectroradiometer's response due to stray-light was assessed and characterized. In particular, a modified version of the method developed at the National Institute of Standards and Technology (NIST) by Zong *et al.* (2006) was performed to estimate the spectral stray-light on the response of the two spectrographs considered. The method is based on computing the ratio of the spectral stray-light signal to the total signal within the bandpass of a spectrograph and any response measured outside the instrument's bandpass is assumed to be spectral stray-light signal. In the original methodology, the ratios were computed using 80 laser lines with wavelengths spaced approximately 8 pixels apart and ratios at the intermediate pixels were obtained by interpolation. In our case, a set of monochromatic spectral line sources was used for calculating the ratio for each pixel of the array detectors, therefore, no interpolation was necessary. Furthermore, using a narrowband source covering all instrument's operational spectral range enabled to characterize the stray-light signal regardless of the source utilized during future measurements.

The relative response at every element i of the array detector to a fixed monochromatic excitation wavelength λ_j falling on the element j was called the spectral line-spread function ($f_{LSF\,i,j}$). Therefore, the spectrograph's response was divided into

two regions. The narrow peak region about element j was the instrument's band-pass containing the in-band response (IB) and the remaining broad region of low response that arose from spectral stray-light. The spectral stray-light signal distribution ($SDF_{i,j}$) was derived by normalizing the $f_{LSF\,i,j}$ to the IB area and setting values of array elements within the IB area to zero:

$$SDF_{i,j} = \frac{f_{LSF\,i,j}}{\sum\limits_{i \in IB} f_{LSF\,i,j}} \quad i \notin IB \;\; \text{(pixels outside } IB\text{)} \tag{1}$$

$$SDF_{i,j} = 0 \qquad i \in IB \;\; \text{(pixels inside } IB\text{)} \tag{2}$$

$$i = 1, 2, ..., n \quad \text{(total number of instrument's pixels)}$$

To obtain the spectral stray-light response function for each element i in the array, one computed $SDF_{i,j}$ for every excitation element j. Then a spectral stray-light distribution matrix, denoted D, was generated by filling the columns of an $n \times n$ matrix with each of the computed $SDF_{i,j}$. Once matrix D was obtained, the measurement equation had the general expression:

$$Y_{meas} = Y_{IB} + D\,Y_{IB} \tag{3}$$

where Y_{meas} was the total measured signal and Y_{IB} was the IB signals. Note that there was a Y_{meas} component at each pixel that arose from the spectral stray-light distribution matrix (D). That matrix measurement expression was rewritten as follows:

$$Y_{meas} = [I + D]\,Y_{IB} = A\,Y_{IB} \tag{4}$$

where I was the $n \times n$ identity matrix and $A = I + D$. Finally, one could obtain each unknown Y_{IB} by simply inverting matrix A:

$$Y_{IB} = A^{-1} Y_{meas} = C\,Y_{meas} \tag{5}$$

where C was called the spectral stray-light correction matrix.

The main advantages of applying that method are that the development of matrix C is only required once and real-time corrections of spectral stray-light are enabled by simply multiplying the measured spectra and the corresponding correction matrix. Note that light collection conditions determine each computed correction matrix C, which is inherent to each individual device. In that process, it is also important to consider the instrument's signal dynamic range and sensitivity within the total operational spectral range.

The magnitude of the spectral stray-light signal within both spectroradiometers considered was experimentally quantified measuring the response to a set of monochromatic line sources[5]. A Digikrom 240 Monochromator with a slit width of 25um was used. The system was also composed of an adjustable halogen lamp (Philips 15V 150W) attached to a stabilized DC power supply (Hewlett Packard 6642A) and a focusing lens, which allowed to light the monochromator's entrance slit with a rather uniform luminous field. As it has been defined, the stray-light radiation for each element of the array detector was estimated as the response measured outside the instrument's bandpass. Fig. C.2a shows several spectra obtained when the CCD-array detector instrument's input optics was illuminated by uniform monochromatic radiation at many wavelengths. By successive measurements of the response of both spectrographs to the wavelength tunable single monochromatic lines over the entire operational spectral range (370-725nm), each device's stray-light correction matrix C was implemented. Each measurement was collected as an average of 15 readings to reduce signal noise and dark current signal was also extracted before measuring.

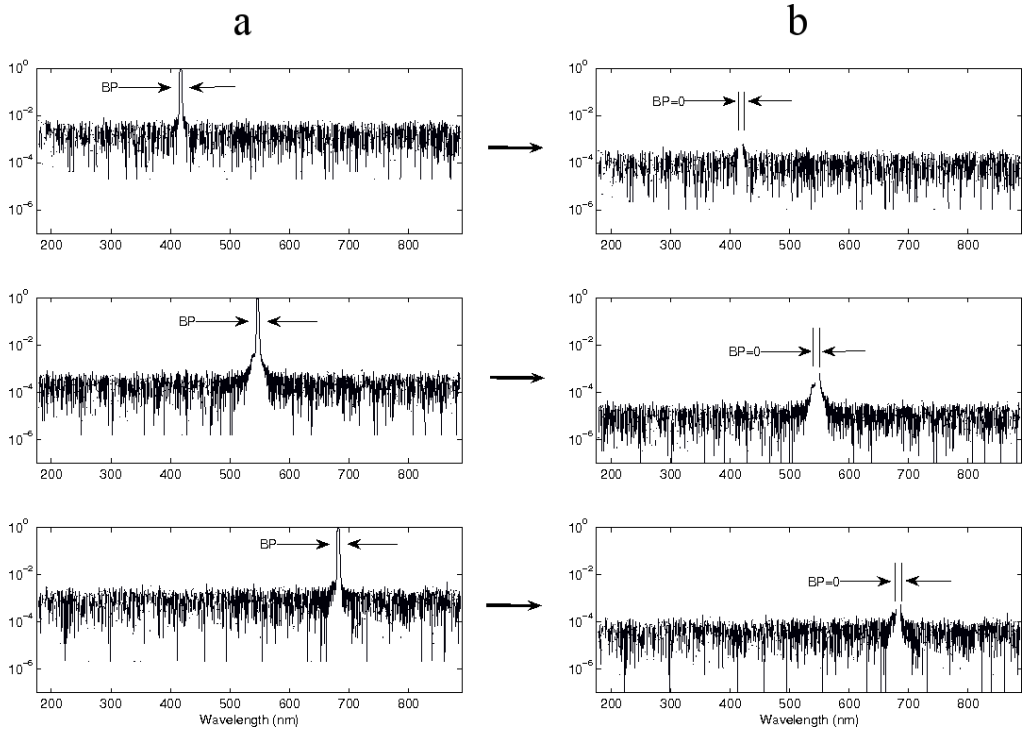


Figure C.2: (a) Spectral responses as the incident monochromatic wavelength was tuned over the spectral coverage of the CCD array-based spectrograph. (b) Corresponding ratio of spectral stray-light response to the total signal within the bandpass (BP) shown in (a). Y-axis are a logarithmic scale.

[5] All measurements were performed at the Centre for Sensors, Instruments and Systems Development (CD6), Universitat Politècnica de Catalunya.

For each spectrograph, each column of the correction matrix C was filled with the ratio, shown in Fig. C.2b, of the spectral stray-light signal to the total signal within the bandpass (BP) of the spectrograph to the corresponding fixed monochromatic excitation j (i.e. each spectral stray-light signal distribution, $SDF_{i,j}$). The diagonal elements of the matrix and surrounding elements within the instrument's bandpass were then all set to zero.

The effectiveness of the spectral stray-light correction method was validated using different entrance spectral stimuli, obtained with different sources and filters. Fig. C.3a depicts the result of applying the spectral stray-light correction method to the CCD-array spectrograph when a broadband source (an halogen lamp) and a green absorption bandpass filter were used. As it can be seen, the stray-light signals outside the filter's bandpass region was clearly reduced after applying the correction (dashed signal). Similar results were obtained using other bandpass filters with different spectral features (not shown).

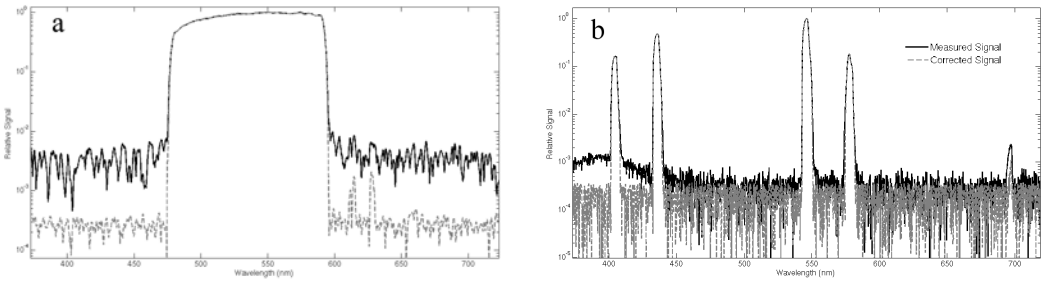


Figure C.3: (a) Spectral stray-light correction for a broadband source with a bandpass filter. (b) Spectral stray-light correction for a spectral wavelength calibration source. Normalized measured and corrected signals from the CCD-array spectrograph. Y-axis are a logarithmic scale.

A spectral wavelength calibration source, which produces low-pressure Mercury-Argon atomic emission lines from 253-1700nm, was also used to validate the effectiveness of applying the spectral stray-light correction method to the CCD-array spectrograph. Fig. C.3b shows the results obtained when using this other source. The stray-light errors were reduced as it can be noticed, specially in the spectral regions between the emission lines of the calibration source.

The spectral stray-light correction method was applied and validated as well for the CMOS-array spectrograph using a broadband source (an halogen lamp) and a green absorption bandpass filter. In that case, the relative spectral stray-light signals were reduced more than two orders of magnitude, to a level of 10^{-4}, as it can be seen in the right panel of Fig. C.4 where the reduction is shown in logarithmic scale.

An alternative method to correct the spectral stray-light radiation based on the use of cut-off filters (e.g. GG495) was proposed by the manufacturer of the CMOS-array spectrograph. A comparison between the results of spectral stray-light correction

obtained using that manufacturer's compensation algorithm and the method implemented in this study was made. Fig. C.5 shows how better stray-light corrections were achieved with the implemented method. The use of a set of monochromatic line sources allowed to accurately cover the entire instrument's operational spectral range and perform a better spectral stray-light correction.

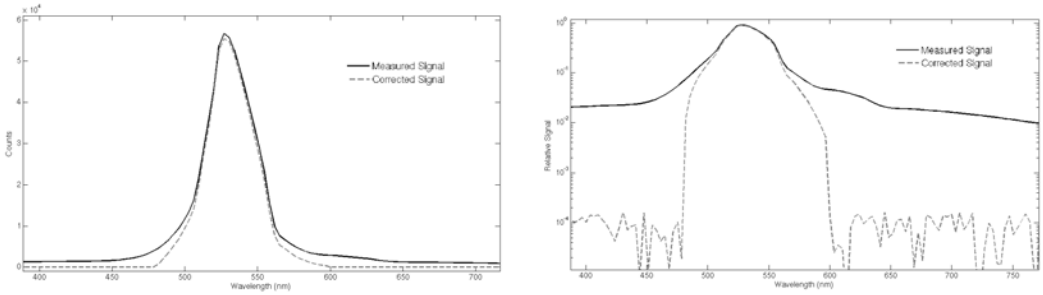


Figure C.4: Spectral stray-light correction for a broadband source with a bandpass filter. Normalized measured and corrected signals from the CMOS-array spectrograph. Y-axis are a linear scale (left panel) and a logarithmic scale (right panel).

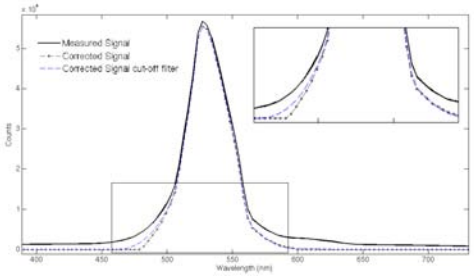


Figure C.5: Spectral stray-light correction using cut-off filters (dashed line) and monochromatic line sources with a correction matrix (dash-dot line). Measured and corrected signals from the CMOS-array spectrograph.

The implemented spectral stray-light correction method has been proved effective for both considered miniature hyperspectral sensors. The spectral stray-light contributions to the measured output signals of these spectrometers have been corrected using the spectral stray-light correction matrix and a simple matrix multiplication. Similarly to other recent initiatives (Feinholz *et al.*, 2009), application of this stray-light correction method will be essential to evaluate the influence of this unwanted scattered radiation on underwater hyperspectral measurements provided by the ANERIS instrument and on its later spectral analysis.

List of publications

Book chapter

E. Torrecilla, J. Piera and M. Vilaseca. Derivative analysis of oceanographic hyperspectral data. In G. Jedlovec (Ed.), *Advances in Geoscience and Remote Sensing*, pp. 597–619, Vienna: InTech, 2009.

Journal articles

E. Torrecilla, D. Stramski, R. A. Reynolds, E. Millan-Nunez and J. Piera. Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. *Remote Sensing Environment*, 115, pp. 2578–2593, doi: 10.1016/j.rse.2011.05.014:25782593, 2011.

B. B. Taylor, **E. Torrecilla**, A. Bernhardt, M. H. Taylor, I. Peeken, R. Röttgers, J. Piera and A. Bracher. Bio-optical provinces in the Eastern Atlantic Ocean and their biogeographical relevance. *Biogeosciences*, 8, pp. 3609-3629, doi:10.5194/bg-8-3609-2011, 2011.

E. Torrecilla, J. Piera and A. Bernhardt. Hyperspectral oceanographic data: potential application in open oceans, *Sea Technology*, 52 (2), pp. 17 – 22, 2011.

Peer-reviewed conference proceedings

E. Torrecilla, J. Piera, A. Bracher, B. B. Taylor and A. Bernhardt. Spectral range sensitivity analysis to improve hyperspectral remote sensing of phytoplankton biodiversity in the ocean. In *Proceedings of the IEEE Third Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHIPSERS*, 2011.

E. Torrecilla, S. Pons, A. Vilamala, I. F. Aymerich, J. Arcos, E. Plaza and J. Piera. Mapping marine phytoplankton assemblages from a hyperspectral and Artifitial Intelligence perspective. In *Proceedings of the IEEE/OEE Oceans Conference and Exhibition, OCEANS*, 2010.

E. Torrecilla, J. Piera, I.F. Aymerich, S. Pons, O. N. Ross and M. Vilaseca. Hyperspectral remote sensing of phytoplankton assemblages in the ocean: effects of the vertical distribution. In *Proceedings of the IEEE Second workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHIPSERS*, 2010.

I. F. Aymerich, S. Pons, J. Piera, **E. Torrecilla** and O. N. Ross. Comparing the use of hyperspectral irradiance reflectance and diffuse attenuation coefficient as indicators for algal presence in the water column. In *Proceedings of the IEEE Second workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHIPSERS*, 2010.

E. Torrecilla, M. Vilaseca, J. Piera and J. Pujol. Método correctivo de la radiación stray-light en espectroradiómetros para aplicaciones subacuáticas. *IX Reunión Nacional de Óptica*, 2009.

E. Torrecilla, S. Pons, M. Vilaseca, J. Piera and J. Pujol. Stray-light correction of in-water array spectroradiometers. Effects on underwater optical measurements. In *Proceedings of the IEEE/OEE Oceans Conference and Exhibition, OCEANS*, 2008.

S. Pons, I. F. Aymerich, **E. Torrecilla** and J. Piera. Monolithic spectrometer for environmental monitoring applications. In *Proceedings of the IEEE/OEE Oceans Conference and Exhibition, OCEANS*, 2007.

E. Torrecilla, I. F. Aymerich, S. Pons and J. Piera. Effect of spectral resolution in hyperspectral data analysis. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, 2007.

Non-peer-reviewed publications

B. B. Taylor, **E. Torrecilla**, A. Bernhardt, R. Röttgers and A. Bracher. Characterization of phytoplankton communities by optical properties in the Atlantic Ocean. *EGU General Assembly*, 2011.

A. Bracher, T. Dinter, A. Sadeghi, B. Taylor, **E. Torrecilla**, F. Wittke and K. Kirsten. Temporal and spatial dynamics of phytoplankton composition in the Western Pacific and it's link to halocarbon emissions. *Earth Observation for Ocean-Atmosphere Interactions Science, ESA-ESRIN, SOLAS, EGU joint Conference*, 2011.

J. Piera, O. N. Ross, **E. Torrecilla**, S. Pons, I. F. Aymerich, J. J. Dañobeitia, R. Quesada, M. L. Artigas and E. Berdalet. Sampling strategies for multi-scale characterisation of biological and hydrodynamic interactions: the role of lothir observatories. *XIV Workshop on Physical Processes in Natural Waters*, 2010.

O. N. Ross, J. Piera, **E. Torrecilla**, M. Lara and E. Berdalet. Using hyperspectral optical methods to examine the effect of turbulent mixing on ecosystem functioning in a coastal lagoon. *XIV Workshop on Physical Processes in Natural Waters*, 2010.

E. Torrecilla, J. Piera, D. Stramski, R. A. Reynolds and E. Millán-Núñez. Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. *XX Ocean Optics Conference*, 2010.

J. Piera, **E. Torrecilla**, S. Pons, I. F. Aymerich, O. N Ross, J. J. Dañobeitia, M. L. Artigas, E. Berdalet, R. Simó, T. Afinkiev, R. Fernández and M. Armada. ANERIS: An autonomous profiler for concurrent characterisation of physical and bio-optical processes at small scales. *XX Ocean Optics Conference*, 2010.

A. Theis, B. Schmitt, **E. Torrecilla** and A. Bracher. Bio-optical measurements in Atlantic open ocean waters for satellite validation. *XX Ocean Optics Conference*, 2010.

O. N. Ross, **E. Torrecilla**, J. Piera and E. Berdalet. Detection of Ocean Bio-Physical Interactions using Hyperspectral Optical Methods. *SCOR Workshop on Ocean Biology Observatories*, 2009.

J. Piera, O. N. Ross, **E. Torrecilla**, I. F. Aymerich, S. Pons, M. Fernández, J. Diogène and E. Berdalet. Observatoris submarins cablejats, el paper de les noves tecnologies en el seguiment del medi marí. *II Congrés d'aquacultura mediterrània*, 2009.

S. Pons, **E. Torrecilla**, M. Vilaseca and J. Piera. Low cost hyperspectral sensors: potential applications for characterization of multi-scale ocean processes. *"Blue" Photonics - Optics in the Sea, European Optical Society Topical Meeting*, 2009.

E. Torrecilla, D. Stramski, R. A. Reynolds, J. Piera and E. Millán Núñez. Identification of phytoplankton pigment assemblages using spectral shape analysis of hyperspectral remote-sensing reflectances. *XIX Ocean Optics Conference*, 2008.

E. Torrecilla, S. Pons, I. F. Aymerich and J. Piera. Comparative response of two different hyperspectral sensors. Application to derivative analysis of absorption spectra. In *Proceedings of the Second International Workshop on MARine TECHnology, MARTECH*, 2007.

S. Pons, I. F. Aymerich, **E. Torrecilla** and J. Piera. Low cost hyperspectral device suitable for monitoring sensor networks. In *Proceedings of the Second International Workshop on MARine TECHnology, MARTECH*, 2007.

E. Torrecilla, J. Piera, R. Quesada, I. Fernández and S. Pons. Characterization of the water optical properties using hyperspectral sensors. In *Proceedings of the First International Workshop on MARine TECHnology, MARTECH*, 2005.

J. Piera, R. Quesada, **E. Torrecilla**, I. Fernández and S. Pons. SAMPLER: An instrumentation Project for studying the effect of turbulence in aquatic systems. In *Proceedings of the First International Workshop on MARine TECHnology, MARTECH*, 2005.

Bibliography

Aiken, J., J.R. Fishwick, S.J. Lavender, R. Barlow, G. Moore, H. Sessions *et al.* (2007). Validation of meris reflectance and chlorophyll during the bencal cruise october, 2002: Preliminary validation of new products for phytoplankton functional types and photosynthetic parameters. *International Journal of Remote Sensing*, 28:497–516.

Aiken, J., Y. Pradhan, R. Barlow, S. Lavender, A. Poulton, P. Holligan and N. Hardman-Mountford (2009). Phytoplankton pigments and functional types in the atlantic ocean: A decadal assessment, 1995–2005. *Deep-Sea Research Part II – Topical Studies in Oceanography*, 56:899–917, doi:10.1016/j.dsr2.2008.09.017.

Albert, A. (2004). *Inversion technique for optical remote sensing in shallow water.* Ph.D. thesis, Fachbereich Geowissenschaften der Universität Hamburg.

Albert, A. and C.D. Mobley (2003). An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters. *Opt. Express*, 11 (22):2873–2890.

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

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, GB3001, doi:10.1029/2007gb003154.

Artigas, M.L., C. Llebot, O.N. Ross, N.Z. Neszi, J. Piera, M. Estrada and E. Berdalet (2012). Understanding the spatio-temporal variability of phytoplankton biomass distribution in a microtidal estuary. *Deep Sea Research II, submitted.*

Astoreca, R., V. Rousseau, K. Ruddick, C. Knechciak, B.V. Mol, J.Y. Parent and C. Lancelot (2009). Development and application of an algorithm for detecting phaeocystis globosa blooms in the case 2 southern north sea waters. *Journal of Plankton Research*, 31, No. 3:287–300.

Aymerich, I.F., J. Piera, A. Soria-Frisch and L. Cros (2009). A rapid technique for classifying phytoplankton fluorescence spectra based on self-organizing maps. *Applied Spectroscopy*, 63:716–726.

Babin, M., J.J. Cullen, C.S. Roesler, P.L. Donaghay, G.J. Doucette, M. Kahru, M.R. Lewis, C.A. Scholin, M.E. Sieracki and H.M. Sosik (2005). New approaches and technologies for observing harmful algal blooms. *Oceanography*, 18, No. 2:210–227.

Babin, M., C.S. Roesler and J. Cullen (eds.) (2008). *Real-time coastal observing systems for marine ecosystem dynamics and harmful algal blooms: Theory, instrumentation and modelling*. UNESCO Publishing.

Balch, W.M., H.R. Gordon, B.C. Bowler, D.T. Drapeau and E.S. Booth (2005). Calcium carbonate measurements in the surface global ocean based on moderate resolution imaging spectroradiometer data. *Journal of Geophysical Research*, 110, C07001, doi:10.1029/2004JC002560.

Barnes, C.R., M.M.R. Best and A. Zielinski (2008). The neptune canada regional cabled ocean observatory. *Sea Technology*, 49:10–14.

Berkhin, P. (2006). *Grouping multidimensional data: Recent advances in clustering*, chap. A survey of clustering data mining techniques, pp. 25–71. Berlin-Heidelberg: Springer.

Bracher, A., M. Vountas, T. Dinter, J.P. Burrows, R. Röttgers and I. Peeken (2009). Quantitative observation of cyanobacteria and diatoms from space using phytodoas on sciamachy data. *Biogeosciences*, 6:751–764, doi:10.5194/bg–6–751–2009.

Brewin, R.J.W., N.J. Hardman-Mountford, S.J. Lavender, D.E. Raitsos, T. Hirata, J. Uitz, E. Devred, A. Bricaud, A. Ciotti and B. Gentili (2011). An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. *Remote Sens. Environ.*, 115:325–339.

Brewin, R.J.W., S.J. Lavender, N.J. Hardman-Mountford and T. Hirata (2010a). A spectral response approach for detecting dominant phytoplankton size class from satellite remote sensing. *Acta Oceanologica Sinica*, 29:14–32.

Brewin, R.J.W., S. Sathyendranath, T. Hirata, S. Lavender, R.M. Barciela and N.J. Hardman-Mountford (2010b). A three-component model of phytoplankton size class for the atlantic ocean. *Ecological Modelling*, 221:1472–1483.

Bricaud, A., H. Claustre, J. Ras and K. Oubelkheir (2004). Natural variability of phytoplankton absorption in oceanic waters: Influence of the size structure of algal population. *Journal of Geophysical Research*, 109, C11010, doi:10.1029/2004JC002419.

Bricaud, A., A. Morel, M. Babin, K. Allali and H. Claustre (1998). Variations of light absorption by suspended particles with chlorophyll-a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *Journal of Geophysical Research*, 103:31033–31044.

Buiteveld, H., J.M.H. Hakvoort and M. Donze (1994). The optical properties of pure water. *Ocean Optics XII, J. S. Jaffe, ed., Proc. SPIE*, 2258:174–183.

Bukata, R.P., J.H. Jerome, K.Y. Kondratyev and D.V. Pozdnyakov (1995). *Optical Properties and Remote Sensing of Inland and Coastal Waters*. C.R.C. Press, Boca Raton, FL.

Chang, G.C., T. Dickey and M. Lewis (2006). *Remote sensing of the marine environment: Manual of remote sensing*, chap. Toward a global ocean system for measurements of optical properties using remote sensing and in situ observations, pp. 285–326. JohnWiley and Sons, New York.

Chang, G.C., K. Mahoney, A. Briggs-Whitmire, D. Kohler, C. Mobley, M. Moline *et al.* (2004). The new age of hyperspectral oceanography. *Oceanography*, 17 (2):22–29.

Chekalyuk, A.M., K.A. Moore, D.L. White and D.E. Porter (2006). Advaced laser fluorescence (alf) technology for estrurine and coastal environmental biomonitoring. Tech. rep., NOAA/UNH Cooperative Institute for Coastal and Estuaring Environmental Technology (CICEET).

Ciotti, A.M. and A. Bricaud (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at seawifs channels in a continental shelf region off brazil. *Limnology and Oceanography: Methods*, 4:237–253.

Clark, D.K., M.A. Yarbrough, M.E. Feinholz, S. Flora, W. Broenkow, Y.S. Kim, B.C. Johnson, S.W. Brown, M. Yuen and J.L. Mueller (2003). Moby, a radiometric buoy for performance monitoring and vicarious calibration of satellite ocean color sensors: measurement and data analysis protocols. Tech. rep., Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, NASA Technical Report Series, 2003 NASA/TM-2003-211621/Rev. 4-Vol. VI. Greenbelt, Maryland: NASA Goddard Space Flight Center.

Claustre, H., Y. Huot, I. Obernosterer, B. Gentili, D. Tailliez and M. Lewis (2008). Gross community production and metabolic balance in the south pacific gyre, using a non intrusive bio-optical method. *Biogeosciences*, 5:463–474, http://www.biogeosciences.net/5/463/2008/.

Cowles, T.J. (2003). *Handbook of scaling methods in aquatic ecology: measurements, analysis and simulation*, chap. Planktonic layers: physical and biological interactions on the small scale. CRC Press.

Craig, S.E., S.E. Lohrenz, Z. 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, no. 21:5414–5425.

Crossland, C., D. Baird, J.P. Ducrotoy, H. Lindeboom, R. Buddemeier, W. Dennison, B. Maxwell, S. Smith and D. Swaney (2005). *Coastal fluxes in the anthropocene*,

chap. The coastal zone: a domain of global interactions, pp. 1–37. International Geoshere Biosphere Programme Series, Springer.

Cullen, J.J., A.M. Ciotti, R.F. Davis and M.R. Lewis (1997). Optical detection and assessment of algal blooms. *Limnology and Oceanography*, 42:1223–1239.

Dana, D.R. and R.A. Maffione (2006). A new hyperspectral spherical absorption meter. In: *AGU Ocean Sciences meeting, Honolulu, Hawaii.*

Davis, C.O., J. Bowles, R. Leathers, D. Korwan, T.V. Downes, W. Snyder, W. Rhea, W. Chen, J. Fisher, P. BIssett and R.A. Reisse (2002). Ocean phills hyperspectral imager: design, characterization and calibration. *Optics Express*, 10 (4):210–221.

de Pedro Puente, X. (2007). *Situacions d'anòxia en zones estuàriques sense forçament mareal: una aproximació als balanços producció/consum d'oxigen.* Ph.D. thesis, Departament d'Ecologia, Universitat de Barcelona.

Dekshenieks, M.M., P.L. Donaghay, J.M. Sullivan, J.E.B. RInes, T.R. Osborn and M.S. Twardowski (2001). Temporal and spatial occurrence of thin phytoplankton layers in relation to physical processes. *Marine Ecology Progress Series*, 223:61–71.

Devred, E., S. Sathyendranath and T. Platt (2007). Delineation of ecological provinces using ocean colour radiometry. *Mar. Ecol.-Prog. Ser.*, 346:1–13, doi:10.3354/meps07149.

Devred, E., S. Sathyendranath, V. Stuart, H. Maas, O. Ulloa and T. Platt (2006). A two component model of phytoplankton absorption in the open ocean: Theory and applications. *Journal of Geophysical Research*, 111.

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

Dickey, T., M. Lewis and G. Chang (2006). Optical oceanography: recent advances and future directions using global remote sensing and in situ observations. *Reviews of Geophysics*, 44, RG1001:1–39.

Dickey, T.D. (2004). Studies of coastal ocean dynamics and processes using emerging optical technologies. *Oceanography*, 17, No. 2:9–13.

Doxaran, D., N. Cherukuru and S.J. Lavender (2006). Apparent and inherent optical properties of turbid estuarine waters: measurements, empirical quantification relationships, and modeling. *Applied Optics*, 45, No. 10:2310–2324.

Duin, R.P.W., D. deRidder and D.M.J. Tax (1997). Experiments with a featureless approach to pattern recognition. *Pattern Recognition Letters*, 18:1159–1166.

Feinholz, M.E., S.J. Flora, M.A. Yarbrough, K.R. Lykke, S.W. Brown, B.C. Johnson and D.K. Clark (2009). Stray light correction of the marine optical system. *Journal of Atmospheric and Oceanic Technology*, pp. 57–73.

Fernández-Tejedor, M., M. Delgado, E. Garcés, J. Camp and J. Diogène (2010). Phytoplankton response to warming in two mediterranean bays of the ebro delta. In: *CIESM Workshop 40 Phytoplankton Response to Mediterranean Environmental Change*.

Ferrari, G.M. and S. Tassan (1999). A method using chemical oxidation to remove light absorption by phytoplankton pigments. *Journal of Phycology*, 35:1090–1098.

Filippi, A.M. (2007). Derivative-neural spectroscopy for hyperspectral bathymetry inversion. *The Professional Geographer*, 59 (2):236–255.

Fournier, G. and J.L. Forand (1994). Analytic phase function for ocean water. In: *Proceedings of SPIE*, vol. 2258, pp. 194–201.

Fournier, G.R. and M. Jonasz (1999). Computer-based underwater imaging analysis. 3761:62–77.

Frette, O., S.R. Erga, J.J. Stamnes and K. Stamnes (2001). Optical remote sensing of waters with vertical structure. *Applied Optics*, 40, No. 9:1478–1487.

Fry, E.S., Z. Lu and X. Qu (2006). Optical absorption of pure water throughout the visible and near ultraviolet. In: *Ocean Optics XVIII, Montreal, Canada*.

Gallie, E.A., and P.A. Murtha (1992). Specific absorption and backscattering spectra for suspended minerals and chlorophyll a in chilko lake, british columbia. *Remote Sensing of Environment*, 39:103–118.

Garcés, E., M. Masó and J. Camp (1999). A recurrent and localized dinoflagellate bloom in mediterranean beach. *Journal of Plankton Research*, 21:2373–2391.

Gilbert, P.M. and G. Pitcher (eds.) (2001). *Global Ecology and Oceanography of Harmful Algal Blooms (GEOHAB): Science plan*. SCOR and UNESCO, Baltimore and Paris. URL `http://www.geohab.info`.

Gitelson, A.A., J.F. Schalles and C.M. Hladik (2007). Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake bay case study. *Remote Sens. Environ.*, 109:464–472.

Gordon, H.R. and A. Morel (1983). *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*. Springer-Verlag, New York.

Hardman-Mountford, N.J., T. Hirata, K.A. Richardson and J. Aiken (2008). An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean. *Remote Sens. Environ.*, 112:3341–3352, doi:10.1016/j.rse.2008.02.016.

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

Hirata, T., J. Aiken, N. Hardman-Mountford, T.J. Smyth and R. Barlow (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sens. Environ.*, 112:3153–3159.

Hirata, T., N.J. Hardman-Mountford, R.J.W. Brewin, J. Aiken, R. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguchi-Aita and Y. Yamanaka (2011). Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, 8:311–327.

Hoffmann, L.J., I. Peeken, K. Lochte, P. Assmy and M. Veldhuis (2006). Different reactions of southern ocean phytoplankton size classes to iron fertilization. *Limnology and Oceanography*, 51:1217–1229.

Holligan, P.M. (1992). *Primary productivity and biogeochemical cycles in the sea*, chap. Do marine phytoplankton influence global climate, pp. 487–501. Plenum Press, New York.

Hommersom, A., S. Peters, M.R. Wernand and J. deBoer (2009). Spatial and temporal variability in bio-optical properties of the wadden sea. *Estuarine, Coastal and Shelf Science*, 83:360–370.

Hommersom, A., M.R. Wernand, S. Peters, M.A. Eleveld, H.J.V. derWoerd and J. deBoer (2011). Spectra of a shallow sea - unmixing for class identification and monitoring of coastal waters. *Ocean Dynamics*, 61:463–480.

Hooker, S.B. and L.V. Heukelem (2009). *CHORS HPLC uncertainties: Final report.* Greenbelt, Maryland: NASA Goddard Space Flight Center World Wide Web page, from URL: http://oceancolor.gsfc.nasa.gov/DOCS/.

Hooker, S.B., L.V. Heukelem, C.S. Thomas, H. Claustre, J. Ras, L. Schlüter *et al.* (2009). The third seawifs hplc analysis round-robin experiment (seaharre-3). Tech. rep., NASA/TM–2009–215849. Greenbelt, Maryland: NASA Goddard Space Flight Center, 106 pp.

Hooker, S.B., N.W. Rees and J. Aiken (2000). An objective methodology for identifying oceanic provinces. *Prog. Oceanogr.*, 45:313–338.

Hunter, P.D., A.N. Tyler, L. Carvalho, G.A. Codd and S.C. Maberly (2010). Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sens. Environ.*, 114:2705–2718.

Hunter, P.D., A.N. Tyler, M. Présing, A.W. Kovács and T. Preston (2008). Spectral discrimination of phytoplankton colour groups: The effect of suspended particulate matter and sensor spectral resolution. *Remote Sensing of Environment*, 112:1527–1544.

Jain, A.K., M.N. Murty and P.J. Flynn (1999). Data clustering: A review. *ACM Computing Surveys*, 31:264–323.

Jeffrey, S.W. and M. Vesk (1997). *Phytoplankton pigments in oceanography: Guidelines to modern methods. Monographs on Oceanographic Methodology 10*, chap. Introduction to marine phytoplankton and their pigment signature, pp. 37–84. UNESCO Publishing, Paris.

Jeffrey, S.W., S.W. Wright and M. Zapata (1999). Recent advances in hplc pigment analysis of phytoplankton. *Marine and Freshwater Research*, 50:879–896.

Kim, M. and W. Philpot (2006). Ocean optical phytoplankton simulator. Available online at http://ceeserver.cee.cornell.edu/wdp2/OOPS/oops-manual.html. Consulted on 2009-06-12.

Kirk, J.T.O. (1994). *Light and Photosynthesis in Aquatic Ecosystems, 2nd Edition.* New York, Cambridge University Press.

Kirkpatrick, G.J., D.F. Millie, S.E. Lohrenz, M.A. Moline, I. Robbins and O. Schofield (2006). An in situ sensor for phytoplankton community structure based on light absorption. In: *AGU Ocean Sciences, Honolulu, Hawaii.*

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*, 114, C09015, doi:10.1029/2009JC005303.

Kutser, T. (2009). Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *International Journal of Remote Sensing*, 30, No. 17:4401–4425.

Kutser, T., L. Metsamaa and A.G. Dekker (2008). Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. *Estuarine, Coastal and Shelf Science*, 78:649–654.

Kuwahara, V.S., G. Chang and T.D. Dickey (2007). Innovations in ocean optics for coastal and open ocean mooring applications. In: *IEEE/OEE Oceans Conference and Exhibition, OCEANS'07, Europe, Aberdeen.*

Lee, M. and M.R. Lewis (2003). A new method for the measurement of the optical volume scattering function in the upper ocean. *Journal of Atmos. Ocean. Tech.*, 20:563–571.

Lee, Z. (ed.) (2006). *Remote sensing of inherent optical properties: fundamentals, tests of algorithms and applications*, vol. 5. Reports of the International Ocean-Colour Coordinating Group (IOCCG), Naval Research Laboratory, Stennis Space Center, USA.

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

LeQuéré, C., S.P. Harrison, C.I. Prentice, E.T. Buitenhuis, O. Aumonts, L. Bopp *et al.* (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11:2016–2040.

Lewis, M.R., J.J. Cullen and T. Platt (1983). Phytoplankton and thermal structure of the upper ocean: consequences of nonuniformity in the chlorophyll profile. *J. Geophys. Res.*, 88:2565–2570.

Llebot, C., J. Solé, M. Delgado, M. Fernández-Tejedor, J. Camp and M. Estrada (2011). Hydrographical forcing and phytoplankton variability in two semi-enclosed estuarine bays. *Journal of Marine Systems*, 86, iss. 3-4:69–86.

Lohrenz, S.E., W. Cai, X. Chen and M. Tuel (2008). Characterizing water mass properties in river dominated coastal waters using underway hyperspectral remote sensing reflectance. In: *Ocean Optics XIX, Barga, Italy.*

Longhurst, A. (2006). *Ecological Geography of the Sea, 2nd Ed.* Elsevier Academic press, USA.

Louchard, E.M., R. Reid, C. Stephens, C. Davis, R. Leathers, T. Downes and R. Maffione (2002). Derivative analysis of absorption features in hyperspectral remote sensing data of carbonate sediments. *Optics Express*, 10 (6):1573–1584.

Lubac, B., H. Loisel, N. Guiselin, R. Astoreca, L.F. Artigas and X. Mériaux (2008). Hyperspectral and multispectral ocean color inversions to detect phaeocystis globosa blooms in coastal waters. *J. Geophys. Res.*, 113, C06026.

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*, 144:265–283.

Malkiel, E., O. Alquaddoomi and J. Katz (1999). Measurements of plankton distribution in the ocean using submersible holography. *Measurement Science Technology*, 10:1142–1152.

Matsumoto, B. (ed.) (1993). *Cell biological applications of confocal microscopy.* Academic Press, San Diego.

Matsuoka, A., Y. Hout, K. Shimada, S. Saitoh and M. Babin (2007). Bio-optical characteristics of the western arctic ocean: implications for ocean color algorithms. *Can. J. Remote Sens.*, 33:503–518.

McClain, C. (2001). Ocean color from satellites. *Encyclopedia of Ocean Sciences, Academic Press*, pp. 1946–1959.

Miller, G.T. (2007). *Sustaining the Earth, 8th edition.* Environmental Science, Belmont, CA.

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*, 120:65–75.

Mobley, C. (2001). Radiative transfer in the ocean. *Encyclopedia of Ocean Sciences, Academic Press*, pp. 2321–2330.

Mobley, C.D. (1994). *Light and water: Radiative Transfer in Naturals Waters.* Academic Press, San Diego.

Mobley, C.D. and D. Stramski (1997). Effects of microbial particles on oceanic optics: Methodology for radiative transfer modeling and example simulations. *Limnology and Oceanography*, 42:550–560.

Mobley, C.D. and L. Sundman (2008). *Hydrolight Ecolight 5.0 User's Guide.* Mercer Island, Washington: Sequoia Scientific Inc.

Moore, C., A. Barnard, P. Fietzek, M.R. Lewis, H.M. Sosik, S. White and O. Zielinski (2009a). Optical tools for ocean monitoring and research. *Ocean Sci.*, 5:661–684.

Moore, C., B. Rhodes, A. Derr and R. Zaneveld (2004a). An instrument for hyperspectral characterization of inherent optical properties in natural waters. In: *ASLO/TOS Ocean Research Conference, Honolulu, Hawaii.*

Moore, C., M.S. Twardowski and J.R.V. Zaneveld (2000). The eco vsf – a multiangle scattering sensor for determination of the volume scattering function in the backward direction. In: *Ocean Optics XV, Monaco.*

Moore, C.C., J.D. Cunha, B. Rhoades, M. Twardowski and J.R.V. Zaneveld (2004b). A new in-situ measurement and analysis system for excitation-emission fluorescence in natural waters. In: *Ocean Optics XVII, Freemantle, Australia.*

Moore, T.S., J.W. Campbell and M.D. Dowell (2009b). A class-based approach to characterizing and mapping the uncertainty of the modis ocean chlorophyll product. *Remote Sensing Environment*, 113:2424–2430.

Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case i waters). *Journal of Geophysical Research*, 93:10749–10768.

Morel, A. and S. Maritorena (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106:7763–7780.

Morel, A. and L. Prieur (1977). Analysis of variations in ocean color. *Limnol. Oceanogr.*, 22:709–722.

Mouw, C.B. and J.A. Yoder (2010). Optical determination of phytoplankton size composition from global seawifs imagery. *J. Geophys. Res.*, 115, C12018, doi:10.1029/2010JC006337.

Mueller, J.L., G.S. Fargion and C.R.M. (eds.) (2003a). Ocean optics protocols for satellite ocean color sensor validation: Inherent optical properties: instruments, characterization, field measurements and data analysis protocols. Tech. rep., NASA/TM-2003-211621/Rev.4-Vol. IV. Greenbelt, Maryland: NASA Goddard Space Flight Center.

—— (2003b). Ocean optics protocols for satellite ocean color sensor validation: Radiometric measurements and data analysis protocols. Tech. rep., NASA/TM-2003-211621/Rev.4-Vol. III. Greenbelt, Maryland: NASA Goddard Space Flight Center.

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 Sens. Environ.*, 112:3366–3375.

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

O'Reilly, J.E., S. Maritorena, D.A. Siegel, M.C. O'Brien, D. Toole, B.G. Mitchell *et al.* (2000). Ocean color chlorophyll-a algorithms for seawifs, oc2 and oc4: version 4. Tech. rep., SeaWiFS postlaunch calibration and validation analyses, part 3, vol. 11. In S. B. Hooker, & E. R. Firestone (Eds.), Greenbelt, Maryland, NASA/TM-2000-206892, pp. 9–27.

Palmer, C. and E. Loewen (eds.) (2005). *Diffraction grating handbook, 6th edition.* Newport Corporation, New York.

Pegau, W.S. (2002). Influence of subsurface iop structure on the remote sensing reflectance during the 2001 hycode campaign. *Ocean Sciences Meeting.*

Pekalska, E. and R.P.W. Duin (2000). Classifier for dissimilarity-based pattern recognition. In: *15th International Conference on Pattern Recognition*, pp. 12–16.

Perez, V., E. Fernandez, E. Maranon, P. Serret, R. Varela, A. Bode, M. Varela, M.M. Varela, X.A.G. Moran, E.M.S. Woodward, V. Kitidis and C. Garcia-Soto (2005). Latitudinal distribution of microbial plankton abundance, production, and respiration in the equatorial atlantic in autumn 2000. *Deep-Sea Res. Pt. I*, 52:861–880, doi:10.1016/j.dsr.2005.01.002.

Perry, M.J. and D.L. Rudnick (2003). Observing the oceans with autonomous and lagrangian platforms and sensors: The role of alps in sustained ocean observing systems. *Oceanography*, 16 (4):31–36.

Platt, T., H. Bouman, E. Devred, C. Fuentes-Yaco and S. Sathyendranath (2005). Physical forcing and phytoplankton distributions. *Sci. Mar.*, 69:55–73.

Platt, T. and S. Sathyendranath (1999). Spatial structure of pelagic ecosystem processes in the global ocean. *Ecosystems*, 2:384–394.

Pope, R.M. and E.S. Fry (1997). Absorption spectrum (380–700 nm) of pure water. ii. integrating cavity measurements. *Applied Optics*, 36:8710–8723.

Raitsos, D.E., S.J. Lavender, C.D. Maravelias, J.A. Haralabous, J. Richardson and P. Reid (2008). Identifying four phytoplankton functional types from space: An ecological approach. *Limnology and Oceanography*, 53:605–613.

Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:846–850.

Reynolds, R.A., D. Stramski and B.G. Mitchell (2001). A chlorophyll-dependent semianalytical model derived from field measurements of absorption and backscattering coefficients within the southern ocean. *Journal of Geophysical Research*, 106:7125–7138.

Rhea, W.J., G.M. Lamela and C.O. Davis (2007). A profiling optics and water return system for validation and calibration of ocean color imagery. *Optics Express*, 15, No. 1.

Rines, J.E.B., M.N. McFarland, P.L. Donaghay and J.M. Sullivan (2010). Thin layers and species-specific characterization of the phytoplankton community in monterey bay, california, usa. *Continental Shelf Research*, 30:66–80.

Robila, S.A. (2005). Using spectral distances for speedup in hyperspectral image processing. *International Journal of Remote Sensing*, 26 (24):5629–5650.

Rocap, G., F.W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N.A. Ahlgren *et al.* (2003). Genome divergence in two prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*, 424:1042–1047.

Ruffin, C., R.L. King and N.H. Younan (2008). A combined derivative spectroscopy and savitzky-golay filtering method for the analysis of hyperspectral data. *GI-Science and Remote Sensing*, 45 (1):1–15.

Röttgers, R. and R. Doerffer (2007). Measurements of optical absorption by chromophoric dissolved organic matter using a point-source integrating-cavity absorption meter. *Limnology and Oceanography: Methods*, 5:126–135.

Röttgers, R., W. Schönfeld, P.R. Kipp and R. Doerffer (2005). Practical test of a point-source integrating cavity absorption meter: The performance of different collector assemblies. *Applied Optics*, 44:5549–5560.

Sadeghi, A., T. Dinter, M. Vountas, B. Taylor, M.A. Soppa and A. Bracher (2012). Remote sensing of coccolithophore blooms in selected oceanic regions using the phytodoas method applied to hyperspectral satellite data. *Biogeosciences, in press.*

Salvador, S. and P. Chan (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: *Proceedings of 16th IEEE International Conference on Tools with AI*, pp. 576–584.

Sathyendranath, S. (ed.) (2000). *Remote sensing of ocean colour in coastal and other optically-complex waters.*, vol. 3. Reports of the International Ocean-Colour Coordinating Group (IOCCG), Dartmouth, Canada.

Sathyendranath, S. and A. Morel (1983). *Remote Sensing Applications in Marine Science and Technology*, chap. Light emerging from the sea — interpretation and uses in remote sensing, pp. 323–357. D. Reidel Publishing Company, Dordrecht.

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

Sathyendranath, S., L. Watts, E. Devred, T. Platt, C.M. Caverhill and H. Maass (2004). Discrimination of diatom from other phytoplankton using ocean-colour data. *Mar. Ecol. Prog. Ser.*, 272:59–68.

Schofield, O., J. Bosch, S. Glenn, G. Kirkpatrick, J. Kerfoot, S. Lohrenz, M. Moline, M. Oliver and P. Bissett (2008). *Real-time coastal observing systems for marine ecosystem dynamics and harmful algal blooms: Theory, instrumentation and modelling*, chap. Bio-optics in integrated ocean observing networks: potential for studying harmful algal blooms, pp. 85–107. UNESCO Publishing.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–398.

Sieracki, C.K., M.E. Sieracki and C.S. Yentsch (1998). An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*, 168:285–296.

Simis, S.G.H., M. Tijdens, H.L. Hoogveld and H.J. Gons (2005). Optical changes associated with cyanobacterial bloom termination by viral lysis. *Journal of Plankton Research*, 27, doi:10.1093/plankt/fbi068:937–949.

Sokal, R.R. and F.J. Rohlf (1962). The comparison of dendrograms by objective methods. *Taxon*, 11 (2):33–40.

Sosik, H.M., R.J. Olson and E.V. Armbrust (2010). *Chlorophyll a fluorescence in aquatic sciences: methods and applications*, chap. Flow cytometry in phytoplankton research, pp. 171–186. Springer.

Stramska, M. and D. Stramski (2005). Effects of a nonuniform vertical profile of chlorophyll concentration on remote-sensing reflectance of the ocean. *Applied Optics*, 44, nº 9:1735–1747.

Stramska, M., D. Stramski, S. Kaczmarek, D.B. Allison and J. Schwarz (2006). Seasonal and regional differentiation of bio-optical properties within the north polar atlantic. *Journal of Geophysical Reserach*, 111:C08003. doi:10.1029/2005JC003293.

Stramski, D., E. Boss, D. Bogucki and K.J. Voss (2004). The role of seawater constituents in light backscattering in the ocean. *Progress in Oceanography*, 61:27–56.

Stramski, D., R. Reynolds, M. Babin, S. Kaczmarek, M. Lewis, R. Röttgers, A. Sciandra, M. Stramska, M.S. Twardowski, B.A. Franz and H. Claustre (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern south pacific and eastern atlantic oceans. *Biogeosciences*, 5:171–201.

Stumpf, R.P., M.E. Culver, P.A. Tester, M. Tomlinson, G.J. Kirkpatrick, B.A. Pederson *et al.* (2003). Monitoring karenia brevis blooms in the gulf of mexico using satellite ocean color imagery and other data. *Harmful Algae*, 2:147–160.

Sullivan, J.M., D.V. Holiday, M. McFarland, M.A. McManus, O.M. Cheriton, K.J. Benoit-Bird, L. Goodman, Z. Wang, M.T. Stacey, C. Greenlaw and M.A. Moline (2010). Layered organization in the coastal ocean: An introduction to planktonic thin layers and the loco project. *Continental Shelf Research*, 30:1–6.

Takahashi, H., S. Kopriva, M. Giordano, K. Saito and R. Hell (2011). Sulfur assimilation in photosynthetic organisms: Molecular functions and regulations of transporters and assimilatory enzymes. *Annu. Rev. Plant Biol.*, 62:157–184.

Tassan, S. and G.M. Ferrari (1995). An alternative approach to absorption measurements of aquatic particles retained on filters. *Limnology and Oceanography*, 40:1358–1368.

—— (2002). A sensitivity analysis of the 'transmittance-reflectance' method for measuring light absorption by aquatic particles. *Journal of Plankton Research*, 24:757–774.

Taylor, B.B., E. Torrecilla, A. Bernhardt, M.H. Taylor, I. Peeken, R. Röttgers, J. Piera and A. Bracher (2011). Bio-optical provinces in the eastern atlantic ocean and their biogeographical relevance. *Biogeosciences*, 8(12):3609–3629, www.biogeosciences.net/8/3609/2011/, doi:10.5194/bg–8–3609–2011.

Torrecilla, E., I.F. Aymerich, S. Pons and J. Piera (2007). Effect of spectral resolution in hyperspectral data analysis. In: *International Geoscience and Remote Sensing Symposium, Barcelona, Spain*, pp. 910–913.

Torrecilla, E., J. Piera, I. Aymerich, S. Pons, O. Ross and M. Vilaseca (2010). Hyperspectral remote sensing of phytoplankton assemblages in the ocean: effects of the vertical distribution. In: *IEEE Second Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHIPSERS'10, Iceland.*

Torrecilla, E., J. Piera, A. Bracher, B.B. Taylor and A. Bernhardt (2011a). Spectral range sensitivity analysis to improve hyperspectral remote sensing of phytoplankton biodiversity in the ocean. In: *IEEE Third Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHIPSERS'11, Lisbon, Portugal.*

Torrecilla, E., D. Stramski, R.A. Reynolds, E. Millan-Nunez and J. Piera (2011b). Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. *Remote Sensing Environment*, 115, doi: 10.1016/j.rse.2011.05.014:2578–2593.

Tsai, F. and W. Philpot (1998). Derivative analysis of hyperspectral data. *Remote Sensing Environment*, 66:41–51.

Turner, R.K., W.N. Adger and I. Lorenzoni (1998). Towards integrated modelling and analysis in coastal zones: Principles and practices. Tech. rep., Land-ocean interactions in the coastal zone (LOICZ) project, No. 11, 122 pp., Texel, The Netherlands.

Twardowski, M.S., H. Claustre, S.A. Freeman, D. Stramski and Y. Huot (2007). Optical backscattering properties of the "clearest" natural waters. *Biogeosciences*, 4:1041–1058.

Twardowski, M.S., M.R. Lewis, A.H. Barnard and J.R.V. Zaneveld (2005). *Remote Sensing of Coastal Aquatic Waters*, chap. In-water instrumentation and platforms for ocean color remote sensing applications, p. 69–93. Kluwer Publishing.

Tzortziou, M., J. Herman, C. Gallegos, P. Neale, A. Subramaniam, L. Harding and Z. Ahmad (2006). Bio-optics of the chesapeake bay from measurements and radiative transfer closure. *Estuarine, Coastal and Shelf Science*, 68:348–362.

Uitz, J., H. Claustre, B. Gentili and D. Stramski (2010). Phytoplankton class-specific primary production in the world's oceans: Seasonal and interannual variability from satellite observations. *Global Biogeochemical Cycles*, 24, GB3016, doi:10.1029/2009GB003680.

Uitz, J., H. Claustre, A. Morel and S. Hooker (2006). Vertical distribution of phytoplankton communities in open-ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, 111:CO8005. doi:10.1029/2005JC003207.

Vaiphasa, C. (2006). Consideration of smoothing techniques for hyperspectral remote sensing. *Journal of Photogrammetry and Remote Sensing*, 60, 2:91–99.

Vallina, S.M. and R. Simó (2007). Strong relationship between dms and the solar radiation dose over the global surface ocean. *Science*, 315:506–509.

Vidussi, F., H. Claustre, B.B. Manca, A. Luchetta and J.C. Marty (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern mediterranean sea during winter. *Journal of Geophysical Research*, 106:19939–19956.

Vila, M., J. Camp, E. Garcés, M. Masó and M. Delgado (2001). High resolution spatio-temporal detection of potentially harmful dinoflagellates in confined waters of the nw mediterranean. *Journal of Plankton Research*, 23 (5):497–514.

Vila-Costa, M., R. Simó, J.M. Gasol, B. Harada, D. Slezak and R.P. Kiene (2006). Dimethylsulfoniopropionate (dmsp) uptake by marine phytoplankton. *Science*, 314:652–654.

VLIZ (2009). Longhurst biogeographical provinces. Available online at http://www.vliz.be/vmdcdata/vlimar/downloads.php. Consulted on 2010-11-03.

Voss, K., H. Gordon, M. Lewis, C. Johnson, M. Yarbrough, S. Flora, M. Feinholz and C. Trees (2008). Radiometry and uncertainties from sortie: Spectral ocean radiance transfer investigation and experiment. In: *NASA Carbon Cycle and Ecosystems Joint Science Workshop*.

Wernand, M.R. (2002). Guidelines for (ship-borne) auto-monitoring of coastal and ocean color. In: *Ocean Optics XVI, Santa Fe, New Mexico, US*.

—— (2011). *Poseidon's paintbox – Historical archives of ocean colour in global-change perspective*. Ph.D. thesis, Universiteit Utrecht, Netherlands.

Wood, S. (2009). *Underwater Vehicles*, chap. Autonomous underwater gliders. Chapter 26, pp. 499–524. In-Tech, Vienna, Austria.

Wright, S.W., D.P. Thomas, H.J. Marchant, H.W. Higgins, M.D. Mackey and D.J. Mackey (1996). Analysis of phytoplankton of the australian sector of the southern ocean: comparisons of microscopy and size-frequency data with interpretations of pigment hplc data using the chemtax matrix factorisation program. *Marine Ecology Progress Series*, 144:285–298.

Xiu, P., Y. Liu and J. Tang (2008). Variations of ocean colour parameters with nonuniform vertical profiles of chlorophyll concentration. *International Journal of Remote Sensing*, 29, No. 3:831–849.

Zaneveld, J.R.V., C. Moore, A.H. Barnard, M. Twardowski and G.C. Chang (2004). Correction and analysis of spectral absorption data taken with the wet labs ac-s. In: *Ocean Optics XVII, Fremantle, Australia*.

Zaneveld, J.R.V. and W.S. Pegau (2003). Robust underwater visibility parameter. *Opt. Express*, 11:2997–3009.

Zapata, M., F. Rodriguez and J.L. Garrido (2000). Separation of chlorophylls and carotenoids from marine phytoplankton: A new hplc method using a reversed-phase c8 column and pyridine-containing mobile phases. *Marine Ecology Progress Series*, 195:29–45.

Zibordi, G., D. D'Alimonte, D.V. derLinde, J.F. Berthon, S.B. Hooker, J.L. Mueller, G. Lazin and S. McLean (2002). The eighth seawifs intercalibration round-robin experiment (sirrex-8), september–december 2001. Tech. rep., NASA Tech. Memo. 2002-206892/Vol. 21., edited by S. B. Hooker and E. R. Firestone, NASA Goddard Space Flight Center, Greenbelt, MD.

Zielinski, O., J.A. Busch, A.D.C. amdK. L. Daly, J. Engelbrektsson, A.K. Hannides and H. Schmidt (2009). Detecting marine hazardous substances and organisms: sensors for pollutants, toxins and pathogens. *Ocean Sci.*, 5:329–349.

Zong, Y., S.W. Brown, B.C. Johnson, K.R. Lykke and Y. Ohno (2006). Simple spectral stray-light correction method for array spectroradiometers. *Applied Optics*, 45 (No. 6):1111–1119.