REMOTE SENSING OF WATER COLOR: MODEL SENSITIVITY ANALYSIS AND ESTIMATION OF PHYTOPLANKTON SIZE FRACTIONS

Zuchuan Li

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Earth Science,
Indiana University

July 2012

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Lin Li, Ph.D., Chair

Master's Thesis Committee

Meghna Babbar-Sebens, Ph.D.

Jeffrey S. Wilson, Ph.D.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents for their hard working to provide me with as many resources as they can to support my life and education. Without their sacrifices, I could not continue my education in China and be accepted by Wuhan University for my undergraduate study, and not mention to accomplish my master degrees in Chinese Academy of Sciences.

I appreciate the help from my graduate committees, especially my advisor and graduate chair, Dr. Lin Li. I have learned much from the three committees.

I thank Kaishan Song, Linhai Li, Kun Shi, Dawei Liu, Ying Sun, Shuai Li and Tingting Zhang for helping me in life and study between these two years.

I am also truly grateful to the faculty, staff and other students of the Department of Earth Science in IUPUI for easing my life and study.

At last, I would like to thank the financial support from NASA Energy and Water Cycle Study program (Grant No. NNX09AU87G) and others who I missed above and gave me help in my graduate study at IUPUI.

ABSTRACT

Zuchuan Li

REMOTE SENSING OF WATER COLOR: MODEL SENSITIVITY ANALYSIS AND ESTIMATION OF PHYTOPLANKTON SIZE FRACTIONS

Phytoplankton size classes (pico-plankton, nano-plankton, and micro-plankton) provide information about pelagic ocean ecosystem structure, and their spatiotemporal variation is crucial in understanding ocean ecosystem structure and global carbon cycling. Remote sensing provides an efficient approach to estimate phytoplankton size compositions on global scale. In the first part of this thesis, a global sensitivity analysis method was used to determine factors mainly controlling the variations of remote sensing reflectance and inherent optical properties inverse algorithms. To achieve these purposes, average mass-specific coefficients of particles were first calculated through Mie theory, using particle size distributions and refractive indices as input; and then a synthesis remote sensing reflectance dataset was created using Hydrolight. Based on sensitivity analysis results, an algorithm for estimating phytoplankton size composition was proposed in the second part. This algorithm uses five types of spectral features: original and normalized remote sensing reflectance, two-band ratios, continuum removed spectra, and spectral curvatures. With the spectral features, phytoplankton size compositions were regressed using support vector machine. According to validation results, this algorithm performs well with simulated and satellite Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer (MODIS), indicating

that it is possible to estimate phytoplankton size compositions through satellite data on global scale.

Lin Li, Ph.D, Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. GENERAL INTRODUCTION

Oceans play an important role in regulating atmospheric CO_2 concentration (Falkowski *et al.* 2000). Each day, more than a hundred million tons of CO_2 are transformed into organic materials by phytoplankton through primary production (Behrenfeld *et al.* 2006), and a large amount of CO_2 from the oceans is also released to the atmosphere due to remineralization through biota respiration (del Giorgio and Duarte 2002). For the first process, the downward export of biogenic particles, also known as biological pump, is responsible for transporting part of the produced organic materials into ocean interiors, decreasing the CO_2 concentration in the atmosphere, and as a result may influence climate (Falkowski *et al.* 1998; Boyd *et al.* 2000; Behrenfeld *et al.* 2006; Cassar *et al.* 2007; Kostadinov *et al.* 2010). This export of organic matter from the surface waters is regulated by various factors, including phytoplankton cell size (Buesseler *et al.* 2007; Burd and Jackson 2009; Guidi *et al.* 2009; Stemmann and Boss 2012), oceanic ecosystem structure (Boyd and Trull 2007), and mineral ballast (Francois *et al.* 2002; Boyd and Trull 2007; Lam *et al.* 2011). Particularly, phytoplankton cell size is an ecophysiological trait connecting to environmental conditions (Irwin *et al.* 2006; Finkel *et al.* 2010), and an important factor controlling particle aggregation and food web (Burd and Jackson 2009), which in turn regulate surface exports. Therefore, knowledge on the spatiotemporal distribution and variation of phytoplankton size fraction (PSF) is crucial for understanding the impact of climate changes on the biogeochemical cycling of elements, and the oceanic biogeochemical feedbacks in influencing climate in the future (Le Quere *et al.* 2005; Boyd and Trull 2007; Finkel *et al.* 2010).

Phytoplankton is divided into several groups according to biogeochemical roles. These groups mainly consist of nitrogen-fixers, calcifiers, silicifiers, and dimethylsulfide (DMS) producers (Anderson 2005; Le Quere *et al.* 2005; Hood *et al.* 2006; Forget *et al.* 2008). First, phytoplankton nitrogen-fixers utilize atmospheric nitrogen as a raw material for growth and have a direct impact on the nitrogen cycle and on other factors influencing climate change (Forget *et al.* 2008). Nitrogen-fixers offer nitrogen to support new production, which is especially important for nitrogen limitation of primary production (Dugdale and Goering 1967). According to (Gruber and Sarmiento 1997; Hood *et al.* 2006), the nitrogen-fixation rates are on the order of 20-40 Tg N y^{-1}, and 80-160 Tg N y^{-1} globally. According to (Le Quere *et al.* 2005), the cell size of phytoplankton nitrogen-fixers ranges from 0.5 to 2.0 μm. Second, phytoplankton calcifiers (coccolithophores) build external plates which are produced through calcification as follow (Hood *et al.* 2006):

$$Ca^{2+} + 2HCO_3^- \rightarrow CaCO_3 + H_2O + CO_2 \quad (1)$$

The produced calcium carbonate serves as a ballast mineral for transporting particulate organic carbon into deep ocean (Armstrong *et al.* 2002; Francois *et al.* 2002). Also, precipitation in calcium carbonate leads to deplete surface CO_3^{2-}, reduces alkalinity and tends to increase in the partial pressure of carbon dioxide in surface waters (Robertson *et al.* 1994). According to (Le Quere *et al.* 2005), the cell size of phytoplankton calcifiers ranges from 5 to 10 μm. Third, four taxonomic groups of phytoplankton silicifiers are identified, namely chrysophyta, silicoflagellates, xanthophyta and bacillariophyta (Forget *et al.* 2008). Among these taxonomic groups, diatoms (bacillariophyta) are the dominant silicifiers in the marine ecosystem. Diatoms

require Si to build cell wall, which functions as ballast mineral to accelerate sinking rate. This special requirement differs from other phytoplankton function groups. Diatoms have larger cell size (20-200 μm) and faster sinking rates than other phytoplankton groups (Le Quere *et al.* 2005; Guidi *et al.* 2009). Diatoms are major contributors to primary production, new production and organic matter export in the ocean. Fourth, DMS producers play an important role in regulating climate. In the ocean, DMS is produced by enzymatic cleavage of dimethylsulfoniopropionate (DMSP) found in phytoplankton. DMS has the potential to impact on climate through the following way. DMS is oxidized to various compounds, such as sulfate aerosols, which serve as cloud condensation nuclei (CCN). Furthermore, changing in cloud optical properties indirectly reduces the amount of incident solar radiation reaching the surface of the ocean, influencing photosynthesis available radiation and finally oceanic primary production. According to (Le Quere *et al.* 2005), the cell size of DMS producers is about 5 μm.

Existing methods for estimating phytoplankton functional groups, based on in-situ data, include microscope, flow cytometry and high performance liquid chromatography (HPLC) methods. First, microscope method is subdivided into light microscope and epifluorescence and electron microscope. Light microscope method relies on morphological characteristics, which cannot be used to discriminate picoplankton. While the discrimination abilities of epifluorescence and electron microscope methods were significantly improved, they are unsuitable for analyzing a large number of samples. Second, flow cytometry method is based on the fluorescence and light–scatter properties of phytoplankton cells. Although this method is superior to microscope method in time consuming, flow cytometry method has a limited particle size range (upper limit of only

15-20 μm) (Forget *et al.* 2008). Furthermore, the carotenoids in phytoplankton do not fluoresce directly, and as result some phytoplankton can be discriminated only through size (Forget *et al.* 2008). Third, HPLC provides a suit of pigment concentrations, which facilitate the separation of phytoplankton based on their marker pigments. As HPLC provides rapid method to discriminate phytoplankton groups with unique accessory pigments, it fails to discriminate the groups with similar marker pigments. Although in-situ methods offer high accurate results in phytoplankton functional types, they are unsuitable for deriving high spatiotemporal distribution of phytoplankton functional types on global scale.

Remote sensing offers an economic and efficient method for estimating high spatiotemporal distribution of phytoplankton functional types on global scale. Two major kinds of approaches have been applied to estimate phytoplankton functional types using remote sensing data. The first kind of approaches is based on the assumption that, according to observations and Mie theory modeling changing in phytoplankton cell size results in variation in IOPs (Morel and Bricaud 1981; Bricaud and Morel 1986; Stramski and Kiefer 1991; Boss *et al.* 2001; Stramski *et al.* 2001; Bricaud *et al.* 2004; Loisel *et al.* 2006; Zhou and Cao 2008). The second kind of approaches models PSF as a function of chlorophyll concentration based on the observation that large size phytoplankton dominates in eutrophic waters, and vice versa (Devred *et al.* 2006; Uitz *et al.* 2006; Brewin *et al.* 2011; Devred *et al.* 2011).

Although remote sensing offers a rapid and economic method to estimate high spatiotemporal results, including phytoplankton functional groups, it is a challenge to obtain high accurate parameters, in that both remote sensing signal (remote sensing

reflectance) and inverse algorithms are governed by various factors. For example, global sensitivity analysis (GSA) has been applied to the RTMs to determine factors mainly controlling the remote sensing reflectance (Duarte *et al.* 2003), indicating that component concentrations mainly regulate the variation of remote sensing reflectance. Similarly, sensitivity analysis methods were also applied to linear matrix inverse and quasi-analytical algorithm to determine their sensitivity to model parameters, such as exponential slope of color dissolved matter and power function slope of backscattering coefficient (Garver and Siegel 1997; Lee *et al.* 2010).

The objectives of this thesis are as follows:

(1) Determining the relative importance of factors controlling the variation of remote sensing reflectance;

(2) Determining the relative importance of factors controlling the performance of inherent optical properties inverse algorithms (quasi-analytical algorithm and linear matrix inverse);

(3) Developing a method for deriving phytoplankton size compositions using simulated and satellite datasets.

II. SENSITIVITY ANALYSIS FOR MODELING WATER OPTICAL PROPERTIES

ABSTRACT

Deriving inherent optical properties (IOPs) from measured reflectance spectra is a challenged problem because both reflectance spectra and inverse algorithms are influenced by various factors (e.g. ambient factors and concentrations of constituents). In this paper, sensitivity analyses were accomplished for: (1) reflectance spectra; and (2) algorithms for deriving IOPs. To analyze the sensitivity for reflectance spectra, the average mass-specific coefficients of particles are first calculated through Mie theory, using particle size distributions and refractive indices as inputs. Then, a synthesis dataset of reflectance spectra (i.e. remote sensing reflectance just above water surface) is created through solving radiative transfer model. A global sensitivity analysis (GSA) is applied to the simulated remote sensing reflectance just above water surface to understand the relative importance of corresponding driving factors. Also, GSA is applied to two popular IOPs inverse algorithms, quasi-analytical algorithm (QAA) and linear matrix inverse (LMI), to determine their sensitivity to algorithm parameters. The GSA results show that remote sensing reflectance is mainly governed by concentrations of constituents, and that the derived IOPs are mainly controlled by the exponential slope of colored dissolved organic matter and minerals/detritus absorption spectrum, and the slope of particulate backscattering spectrum.

Keywords: remote sensing reflectance, sensitivity analysis, quasi-analytical algorithm, linear matrix inverse

1. Introduction

Estimation of inherent optical properties (IOPs) (refer to table 2.1 for symbols and abbreviations) is a challenge in the studies of water color. The difficulty partially originates from the complicated dependence of remote sensing reflectance on various factors, including concentrations of constituents and light field factors (e.g. solar zenith angle and wind speed). Furthermore, due to variation in composition, the spectral shape of IOPs varies across regions, such as the exponential slope of colored detritus matter (CDM) absorption (Carder *et al.* 1989; Babin *et al.* 2003), and the power function slope of backscattering coefficient of particle matter (Loisel *et al.* 2006). While the variation in the spectral shape of IOPs leads to significant uncertainties to the outputs of IOPs inverse algorithms (Garver and Siegel 1997; Lee *et al.* 2010), changes both in concentrations of constituents and in particulate characteristics (e.g. size and refractive index) result in variation in remote sensing reflectance. Therefore, knowledge on factors governing the variation of remote sensing reflectance and the performances of IOPs inverse algorithms plays an important role in getting insight into the mechanisms of remote sensing reflectance and ultimately improving existing IOPs inverse algorithms.

Remotely deriving IOPs is commonly based on radiative transfer models (RTM), which rigorously model apparent optical properties (AOPs) as a function of IOPs (Gordon *et al.* 1988; Morel 1988; Kirk 1991; Zaneveld 1995; Haltrin 1998; Albert and Mobley 2003; Lee *et al.* 2004). Specifically, remote sensing reflectance just beneath water surface ($rrs(\lambda)$) can be written as (Gordon *et al.* 1975; Gordon *et al.* 1988):

$$rrs(\lambda) = \sum_{i=0}^{1} g_i \left(\frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \right)^{i+1} \quad (1)$$

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bph}(\lambda) + b_{bNAP}(\lambda) \quad (2)$$

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{CDOM}(\lambda) + a_{NAP}(\lambda) \quad (3)$$

where λ and i are wavelength and index, respectively; $a(\lambda)$ is the total absorption coefficient, $a_x(\lambda)$ is the absorption coefficient of a constituent x which could be w (water), ph (phytoplankton), CDOM, or NAP (non-algae particle); and $b_b(\lambda)$ is the total backscattering coefficient, and $b_{bx}(\lambda)$ is the backscattering coefficient of a constituent x. For the IOPs of constituents, excluding pure water and CDOM, they are defined as a function of concentrations and specific coefficients (e.g. $a_{ph}(\lambda) = Chl \times a^*_{ph}(\lambda)$), indicating that changes in particulate characteristics (e.g. particle size distribution (PSD) and refractive index) will result in variations in mass-specific coefficients (e.g. $a^*_{ph}(\lambda)$) and finally in $rrs(\lambda)$ (Bricaud and Morel 1986; Stramski and Kiefer 1991; Bricaud *et al.* 1995; Bricaud *et al.* 1998; Stramski *et al.* 2001; Doxaran *et al.* 2002; Binding *et al.* 2005). Actually, $rrs(\lambda)$ also depends on light field (i.e. g_0, g_1).

Based on equations (1)-(3), numerous inverse algorithms have been developed for deriving absorption and backscattering coefficients, including the quasi-analytical algorithm (QAA) (Lee *et al.* 2002) and the linear matrix inverse (LMI) (Hoge and Lyon 1996; Wang *et al.* 2005). The QAA derives IOPs with two consecutive steps. $a(\lambda)$ and $b_b(\lambda)$ are retrieved in the first step using remote sensing reflectance and backscattering spectral model. In the second step, $a(\lambda)$ is decomposed into its major constituents (i.e. phytoplankton and CDM). More details on QAA refer to appendix B. The LMI algebraically derives three major unknowns using remote sensing reflectance at three

wavelengths (refer to appendix C for details). The IOPs results from these algorithms are very encouraging for oceanic waters (Lee 2006); however the performances of these algorithms are subjected to the effects of various factors aforementioned. According to (Hoge and Lyon 1996), 10% uncertainty in the mass-specific absorption of phytoplankton will lead to 50% uncertainty in the estimation of IOPs. Additionally, the particulate backscattering coefficient slope (η) is a function of the PSD (Junge distribution) slope (ξ) (Loisel *et al.* 2006; Kostadinov *et al.* 2009; Kostadinov *et al.* 2010), rather than a constant as used by Garver-Siegel-Maritorena (GSM) (Garver and Siegel 1997; Maritorena *et al.* 2002). The variation of η also accounts for large proportions of derived-IOPs uncertainties (e.g. QAA) (Lee *et al.* 2010). Due to variation in composition, the colored dissolved matter (CDM) absorption spectrum slope (S) varies across regions (Carder *et al.* 1989; Babin *et al.* 2003), and also has significant impact on the performances of inverse algorithms, such as QAA (Lee *et al.* 2010) and GSM (Garver and Siegel 1997). Instead of being constants, light field (i.e. g_0, g_1) depends in complicated ways on ambient factors (e.g. solar zenith angle and wind speed) and IOPs of constituents (Morel and Gentili 1991; Morel and Gentili 1993; Zaneveld 1995; Morel and Gentili 1996; Loisel and Morel 2001; Morel *et al.* 2002; Albert and Mobley 2003; Lee *et al.* 2004; Chami *et al.* 2006). Therefore, to refine existing IOPs inverse algorithms, it is necessary to determine how the variations of these factors exert impacts on remote sensing reflectance and the performance of IOPs inverse algorithms.

Sensitivity analysis (SA) is to partition uncertainty in the output of a model to different sources of uncertainty in model inputs. It can be used to obtain the relative importance of individual model factor. SA provides the following knowledge on a model

(Saltelli 2000): (1) if a model reproduces studied systems or processes; (2) factors mostly attributed to output variability; (3) locating insignificant model parameters; and (4) determining factors interacting with each other. SA has been widely used in various research areas, including remote sensing applications. Ceccato *et al* (Ceccato *et al.* 2001) simulated a set of reflectance spectra through RTM, PROSPECT (Jacquemoud and Baret 1990), applied Extended Fourier Amplitude Test (EFAST) (Saltelli *et al.* 1999) to the simulated dataset, and found that to better derive equivalent water thickness (EWT) the shortwave infrared (SWIR) and NIR regions should be used together. Furthermore, Ceccato *et al* (Ceccato *et al.* 2002) used EFAST to address vegetation water content effects on the signal (reflectance spectra) measured at three levels: leaf, canopy and atmosphere. SA also has been applied to water-color remote sensing. The sensitivity of derived IOPs to model parameters was analyzed by varying one parameter and fixing the rest parameters each time (Garver and Siegel 1997). This method is simple in implementation, but the interactions among factors are omitted. Partial derivatives of QAA were deduced to analyze the uncertainties of derived IOPs, and the relative importance of analyzed parameters (Lee *et al.* 2010). However, this method is unsuitable for complex, nonlinear models. Global sensitivity analysis (GSA) was also applied to the remote sensing reflectance just above water surface ($Rrs(\lambda)$) simulated by Hydrolight (Mobley 1994) to determine the relative importance of various input parameters, such as the concentrations of constituents (Duarte *et al.* 2003).

Considering both remote sensing reflectance and IOPs inverse algorithms are nonlinear functions of various factors, we propose to use EFAST to determine key factors mainly controlling the variation of $Rrs(\lambda)$ and the performances of QAA and LMI, with

the aims of: (1) getting insight into the dependence of $Rrs(\lambda)$ on various factors, including particulate characteristics; and (2) improving existing IOPs inverse algorithms. To achieve the first goal, Mie theory was adopted to calculate mass-specific coefficients from PSD and reflective indices of particles. Using the calculated mass-specific coefficients, together with concentrations of constituents, volume scattering function (VSF) and ambient factors, a set of $Rrs(\lambda)$ was simulated with Hydrolight. EFAST was then applied to the simulated dataset to determine the sensitivity indices of input factors. In addition, EFAST was applied to two commonly used inverse algorithms in remote sensing of ocean color, QAA and LMI, to determine the relative importance of model parameters in the derivation of IOPs.

Table 2.1. Symbols and definitions

Symbol/abbreviation	Description	Unit
$Rrs(\lambda)$	Remote sensing reflectance above water surface	sr^{-1}
$rrs(\lambda)$	Remote sensing reflectance below water surface	sr^{-1}
λ	Wavelength	nm
g_i	Light distribution factor (i=0,1)	-
$a(\lambda)$	Total absorption coefficients of water column	m^{-1}
$c(\lambda)$	Beam attenuation coefficient	m^{-1}
$a^*(\lambda)$	Mass-specific absorption coefficient	$m^2\ g^{-1}$
$b^*(\lambda)$	Mass-specific scattering coefficient	$m^2\ g^{-1}$
$a_x(\lambda)$	Absorption coefficients of *x*. *x* could be *w*: water; *ph*: in vivo phytoplankton; *CDOM*: colored dissolved organic matter; *CDM*: colored detritus matter	m^{-1}
$b_b(\lambda)$	Total backscattering coefficients of water column	m^{-1}
$b_{bx}(\lambda)$	Backscattering coefficients of *x*. *x* could be *w*: water; *ph*: phytoplankton; *NAP*: non-algae particle	m^{-1}
η	Spectral absorption slope of particulate backscattering coefficient	-
ξ_x	PSD slope, x can be m (mineral) and ph (phytoplankton)	-

γ	Slope of the power-law particulate beam attenuation coefficient	-
S	Exponential slope of CDM absorption spectrum	nm^{-1}
S_{CDOM}	Exponential slope of CDOM absorption spectrum	nm^{-1}
S_{NAP}	Exponential slope of mineral absorption spectrum	nm^{-1}
D	Diameter of particle	μm
$D_{\max}$	The maximum of particulate diameter	μm
$D_{\min}$	The minimum of particulate diameter	μm
$N(D)$	Number of particles with diameter D	-
n_X	Real part of refractive index, x could be ph: phytoplankton, m: mineral	-
n'_X	Imaginary part of refractive index, x can be ph: phytoplankton, m: mineral	-
$\bar{Q}_a(\lambda)$	Average absorption efficiency	-
$\bar{Q}_b(\lambda)$	Average scattering efficiency	-
Chl	Chlorophyll concentration	$mg\ m^{-3}$
C	Mineral concentration	$g\ m^{-3}$
CDOM	Colored dissolved organic matter	-
NAP	Non-algal particles (concentration)	$g\ m^{-3}$
CDM	Colored detritus matter, i.e. CDOM+NAP	-
IOPs	Inherent optical properties	-
AOPs	Apparent optical properties	-
PSD	Particle size distribution	-

2. Datasets

Two datasets were used in this paper. One is for RTM sensitivity analysis, consisting of concentrations of constituents, particulate characteristics (i.e. PSD, refractive indices), and ambient factors. These parameters were used to simulate a set of $Rrs(\lambda)$. Another dataset, which is composed of model parameters and four $Rrs(\lambda)$ spectra extracted from the International Ocean Colour Coordinating Group (IOCCG) simulated dataset, was used for sensitivity analysis for QAA and LMI.

2.1 Dataset for RTMs sensitivity analysis

Parameters used for $Rrs(\lambda)$ sensitivity analysis are listed in table 2.2. Their variation ranges were determined with reference to values reported in literatures. First, the variation ranges of the refractive index, PSD, and VSF of phytoplankton were set. The real part of refractive index, n_{ph}, varies among species due to differences in water content (Stramski and Kiefer 1991; Ahn *et al.* 1992; Aas 1996; Stramski *et al.* 2001), with a range of 1.01-1.09 (relative to water) (Aas 1996), and being independent of wavelength. The imaginary part of refractive index, n'_{ph}, ranges from 0.002 to 0.02 at 440 nm (Ahn *et al.* 1992; Stramski *et al.* 2001; Zhou 2008). As n'_{ph} is mainly determined by the concentrations of contained pigments in one cell, ρ_{Chl}, the relationship between n'_{ph} and ρ_{Chl} could be determined (Fig. 2.1) from measured data (Bricaud and Morel 1986; Ahn *et al.* 1992; Stramski *et al.* 2001; Zhou 2008):

$$\rho_{Chl} = 880.12 n'_{ph} - 0.3382 \tag{4}$$

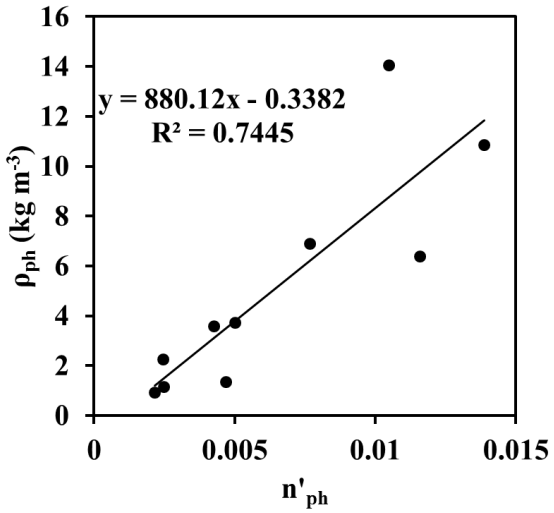


Fig. 2.1. Imaginary part of refractive index versus pigment concentration.

The diameter of phytoplankton cell was set between 0.05 and 200 μm. The slope of phytoplankton PSD, ξ_{ph}, was set between 3.5 and 4.5 (Stramski and Kiefer 1991). The Fournier-Forand phase function was selected to describe the distribution of scattered light by phytoplankton. This phase function was derived based on physically modeling the aquatic particulate populations and was found to be able to approximate to realistic values in the Hydrolight simulation (Fournier and Forand 1994; Fournier and Jonasz 1999; Mobley *et al.* 2002). The backscattering probability was calculated from the real part of refractive index and ξ_{ph}, and used to estimate VSF in simulation (Mobley 1994; Twardowski *et al.* 2001; Mobley *et al.* 2002).

Second, the variation ranges of refractive index, density, PSD, and VSF of NAP were determined. n_m ranges from 1.14 to 1.26 due to variation in sources (Twardowski *et al.* 2001). Based on a dataset including 29 mineral species, a linear relationship between density (ρ_m) and n_m was obtained (Wozniak and Stramski 2004):

$$n_m = 0.1475 \times 10^{-6} \rho_m + 0.7717 \tag{5}$$

For $n'_m(\lambda)$, the low limit was set as $n'_m(\lambda) = 0.007954 \exp(-0.007186\lambda)$ (Stramski *et al.* 2001; Wozniak and Stramski 2004), while the up limit at 440 nm was set as 0.007 (Patterson *et al.* 1977; Wozniak and Stramski 2004). As a result, $n'_m(\lambda)$ at 440 nm ranges from 0.0003 to 0.007. The PSD slope, ξ_m, was set between 3.5 and 4.5 as phytoplankton. The Fournier-Forand phase function was also selected as mineral VSF. The exponential slope of absorption spectrum (S_{NAP}) was set with a mean of 0.0123 nm^{-1} and a standard deviation of 0.0013 nm^{-1} (Babin *et al.* 2003).

Third, due to variation in composition (humic and fulvic acids) across different regions (Carder *et al.* 1989), the exponential slope of CDOM absorption spectrum (S_{CDOM}) has a mean value of 0.0176 nm^{-1} and a standard deviation of 0.002 nm^{-1} (Babin *et al.* 2003).

Fourth, the variation ranges of ambient factors were set as followed. Solar zenith angle was set between 0^0 and 60^0, because a large value of solar zenith angle causes the distribution of upwelling radiance to deviate much from uniform distribution and results in significant errors in atmospheric correction (Yang and Gordon 1997; Gordon and Franz 2008). Wind speed was set to range from 0 and 15 m s^{-1}.

Finally, with the above parameters, four datasets of $Rrs(\lambda)$ (2509 samples of each, in the wavelength region of 400-800 nm and with an interval of 5 nm) were simulated through Hydrolight. One is with chlorophyll concentration 0-3 mg m^{-3}, mineral concentration 0-3 g m^{-3}, and $a_{CDOM}(440)$ 0-0.2 m^{-1}; it is abbreviated as scenario A. Another is with chlorophyll concentration 0-100 mg m^{-3}, mineral concentration 0-100 g m^{-3}, and $a_{CDOM}(440)$ 0-6 m^{-1}; it is abbreviated as scenario B. For each scenario, it is further divided into two subgroups; one is for the case that the backscattering probabilities of phytoplankton and minerals are set as 1% and 1.83%, respectively; another is for the case that the backscattering probabilities are derived from PSD and real part of refractive index (Twardowski *et al.* 2001), which is abbreviated as Twardowski.

Table 2.2. Input parameters for RTMs SA; μ and δ represent mean and standard deviation, respectively

Type	Parameter	Range of variation	Probability distribution
Phytoplankton	Concentration (*Chl*)	0-100 mg m^{-3}	Uniform
	Real part of refractive index (n_{ph})	1.01-1.09	Uniform
	Imaginary part of refractive index (n'_{ph})	0.002-0.02	Uniform
	Particle size distribution (ξ_{ph})	3.5-4.5	Uniform
Minerals/detritus	Mineral concentration (C)	0-100 g m^{-1}	Uniform
	Real part of refractive index (n_m)	1.14-1.26	Uniform
	Imaginary part of refractive index (n'_m)	0.0003-0.007	Uniform
	Particle size distribution (ξ_m)	3.5-4.5	Uniform
	Exponential slope of spectral absorption (S_m)	μ: 0.0123 nm^{-1} δ: 0.0013 nm^{-1}	Gaussian
CDOM	$a_{cdom}(440)$	0-6.0	Uniform
	Exponential slope of spectral absorption (S_{cdom})	μ: 0.0176 nm^{-1} δ: 0.002 nm^{-1}	Gaussian
Ambient factor	Solar zenith angle	0-60 degree	Uniform
	Wind speed	0-15 m s^{-1}	Uniform

2.2 Dataset for QAA and LMI sensitivity analysis

SA parameters for QAA and LMI are listed in table 2.3. As light field factors (g_0 and g_1) are influenced by various factors, including solar zenith angle and IOPs, their values vary across different regions. In this study, they were assumed to be between the light field values for oceanic water provided by (Gordon *et al.* 1988), and for high scattering water provided by (Lee *et al.* 1999; Lee *et al.* 2002), as shown in table 2.3. The exponential slope of CDM absorption spectra (S) ranges from 0.005 to 0.03 nm^{-1}, the power-low function slope of backscattering spectra (η) ranges from 0 to 2, and the

phytoplankton absorption ratio between 410 and 440 nm ranges from 0.6 to 1 (Lee *et al.* 2009). Full-width at half max (FWHM) ranges from 85.0 to 93.5 (Hoge and Lyon 1996).

Table 2.3. Model parameters of QAA and LMI sensitivity analysis

Type	Parameters	Range of variation	Probability distribution
Light field	g_0	0.0840-0.095	Uniform
	g_1	0.0794-0.170	Uniform
IOP model parameters	Exponential slope of CDM spectral absorption (S)	0.005-0.03 nm^{-1}	Uniform
	Total particulate backscattering shape (η)	0-2	Uniform
	FWHM (δ) (LMI only)	85.0-93.5	Uniform
	ξ (QAA only)	0.6-1	Uniform
Water IOPs	Absorption (wavelength dependent)		Gaussian
	Backscattering (wavelength dependent)		Uniform

According to (Morel and Maritorena 2001), the results from bio-optical model, especially in the blue part (400-500 nm), are influenced by the difference in measured water absorption, when the concentration of chlorophyll is low. Therefore, uncertainties in the measurements of water absorption (Pope and Fry 1997) and backscattering (Smith and Baker 1981) was analyzed to determine their impacts on the performances of QAA and LMI.

Four $Rrs(\lambda)$ spectra, as shown in Fig. 2.2, were extracted from the IOCCG simulated dataset with a solar zenith angle of 30^0. The IOPs and concentrations of these four spectra are listed in table 2.4. Each of the four spectra was used as the input of QAA and LMI in the SA processes.

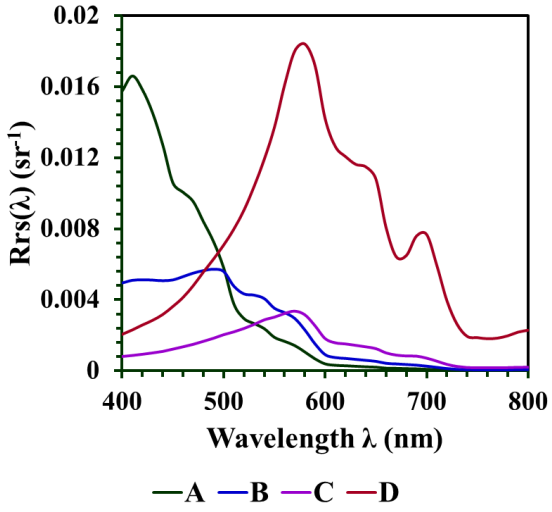


Fig. 2.2. $R_{rs}(\lambda)$ extracted from IOCCG simulated data with solar zenith angle of 30^0; A, B, C, and D: are the $R_{rs}(\lambda)$ with IOPs and concentrations of number 1, 2, 3, and 4 in table 2.4, respectively.

Table 2.4. Component IOPs and concentrations

Number	Chl (mg/m^3)	$a_{dm}(440)$ (m^{-1})	$a_{cdom}(440)$ (m^{-1})	S_{dm} (nm^{-1})	S_{cdom} (nm^{-1})
1	0.03	0.001937	0.002499	0.01332	0.0147
2	0.3	0.015283	0.025911	0.01436	0.0110
3	3.0	0.081522	0.416884	0.01332	0.0141
4	30.0	0.385726	0.606465	0.01484	0.0163

2.3 IOCCG simulated dataset

The IOCCG simulated dataset (Lee 2006) with a solar zenith angle of 30^0 contains IOPs and AOPs. The IOPs were generated with various available optical parameters, while AOPs were generated using Hydrolight (Mobley 1994) with the corresponding IOPs. This dataset has been widely used in testing and comparing algorithms of IOPs estimation (Lee 2006). This dataset was split into two parts to compare the performance of the algorithms for different ranges of concentrations. One contains all the samples with

chlorophyll concentration ranging from 0.03 to 30 mg m^{-3}. Another contains samples with chlorophyll concentration ranging from 0.03 to 0.3 mg m^{-3}.

3. Methods

3.1 Extended Fourier amplitude sensitivity test

EFAST belongs to variance-based GSA methods, whose main idea is to apportion the variation of model output to different sources of variation (Saltelli 2000). The main idea of EFAST is to convert a k-dimensional integral of variance into one-dimensional integral, calculating first order sensitivity index (FSI) and total sensitivity index (TSI), respectively. FSI measures only the main effect contribution of each input parameter on the output variance, while TSI takes into account the effects of interactions among factors (Ekström 2005). To demonstrate GSA more clearly, take reflectance spectrum Y_{rs} as an example. Assuming that Y_{rs} is only governed by the concentrations of phytoplankton X_{ph}, CDOM X_{CDOM}, and NAP X_{NAP}, the variance of Y_{rs} can be decomposed via EFAST as follow:

$$V(Y_{rs}) = V(X_{ph}) + V(X_{CDOM}) + V(X_{NAP}) + V(X_{ph}, X_{CDOM}) + V(X_{ph}, X_{NAP}) + V(X_{CDOM}, X_{NAP}) + V(X_{ph}, X_{CDOM}, X_{NAP}) \quad (6)$$

where $V(Y_{rs})$ is the variance of Y_{rs}, $V(X_{ph})$ is the variance of X_{ph}, $V(X_{CDOM})$ is the variance of X_{CDOM}, $V(X_{NAP})$ is the variance of X_{NAP}, $V(X_{ph}, X_{CDOM})$ is the variance of interaction between X_{ph} and X_{CDOM}, $V(X_{ph}, X_{NAP})$ is the variance of interaction between X_{ph} and X_{NAP}, $V(X_{CDOM}, X_{NAP})$ is the variance of interaction between X_{CDOM}

and X_{ph}, and $V(X_{ph}, X_{CDOM}, X_{NAP})$ is the variance of interaction among X_{ph}, X_{CDOM} and X_{NAP}. Ratio between $V(X_{ph})$ and $V(Y_{rs})$ is defined as the FSI of X_{ph}, S_{ph-1}:

$$S_{ph-1} = \frac{V(X_{ph})}{V(Y_{rs})} \tag{7}$$

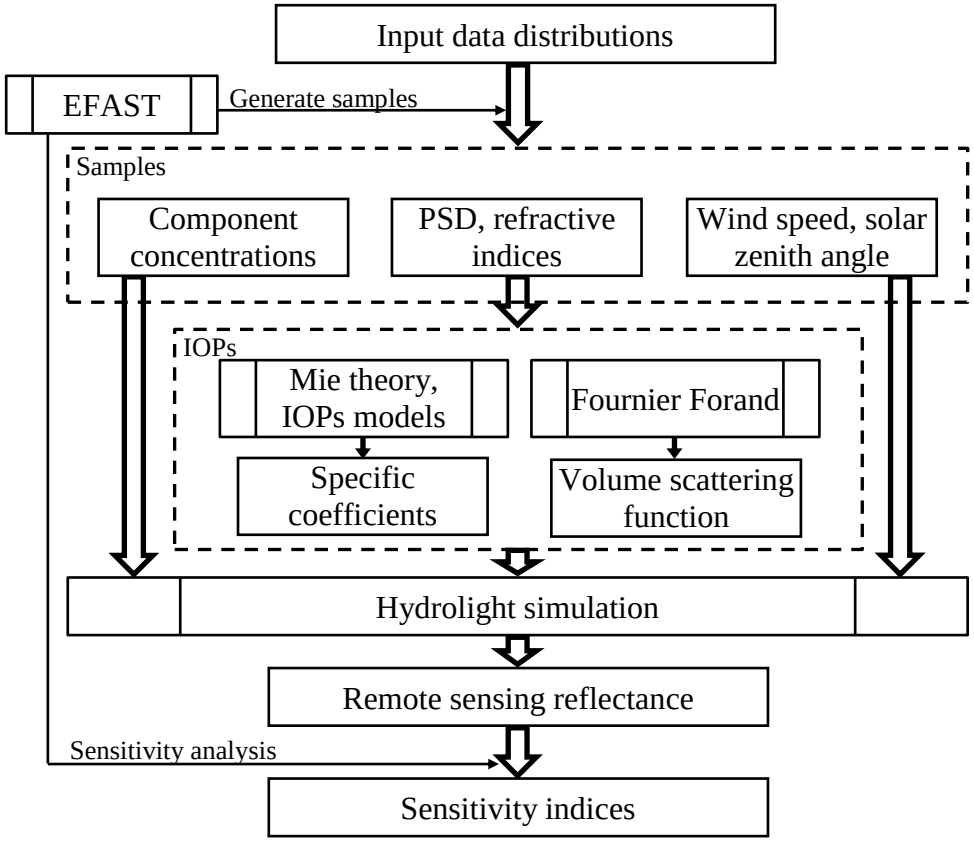
The TSI of X_{ph}, S_{ph-T}, is mathematically defined as:

$$S_{ph-T} = \frac{V(X_{ph}) + V(X_{ph}, X_{cdom}) + V(X_{ph}, X_{nap}) + V(X_{ph}, X_{cdom}, X_{nap})}{V(Y_{rs})} \tag{8}$$

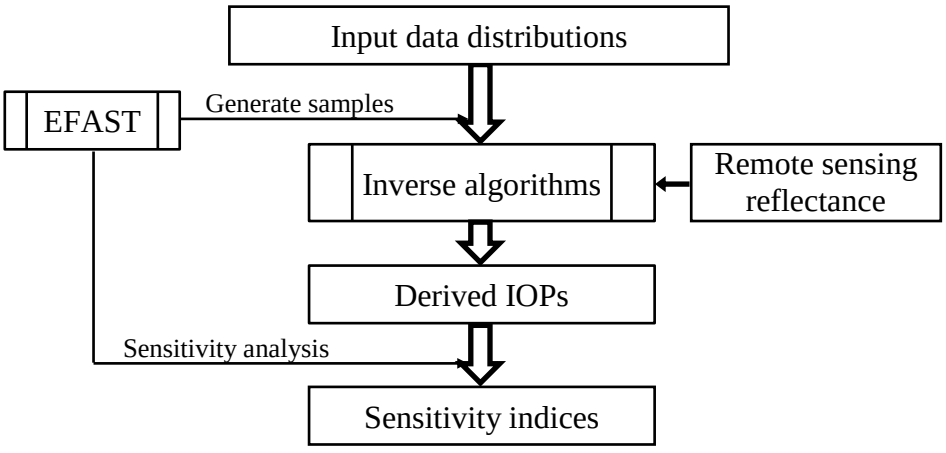
The FSI of a factor represents the percentage of output variance that it accounts for. For example, S_{ph-1} represents the percentage of Y_{rs} variance that is accounted for by X_{ph}. In other words, the significance of X_{ph} in regulating Y_{rs} increases with S_{ph-1}. A model without interactions between factors is said to be additive and the sum of FSI is one, meaning that the variances between factors are 0; however, a non-linear model with high interactions may have low value of FSI.

Ideally, GSA should be based on a broad range of measured $Rrs(\lambda)$, which represents all the cases of natural waters, coupling with various measured parameters, such as concentrations of constituents. However, it is time-consuming and expensive to measure such a dataset of $Rrs(\lambda)$. Therefore, RTM was used to simulate a dataset of $Rrs(\lambda)$ from basic parameters, including PSD and refractive indices of particles. Then, EFAST was applied to the simulated dataset to analyze the sensitivity of $Rrs(\lambda)$ to those parameters. As TSI takes into account the effects of interactions among factors (Ekström 2005), it is better than FSI in indicating the relative importance of a factor in regulating the output of a model, and thus TSI result will be shown in the following results.

3.2 Sensitivity analysis for RTMs



(A)



(B)

Fig. 2.3. Flow chart for SA; rectangle with four vertical lines represents processes or algorithms and regular rectangle denotes data; (A) for radiative transfer models; and (B) for QAA and LMI.

The SA approach for RTM is presented in Fig. 2.3(A). This approach consists of two major parts: data simulation and SA. In data simulation, the mass-specific coefficients of particles were calculated from PSD and refractive indices. This calculation is based three assumptions: (1) phytoplankton cell and NAP are spherical particles; (2) these particles are homogeneous with regard to their refractive indices; and (3) the assemblages of these particles follow Junge distribution. These assumptions allow the particulate average absorption, $\bar{Q}_a(\lambda)$, and scattering, $\bar{Q}_b(\lambda)$, efficiencies to be calculated through Mie theory (Bohren and Huffman 1983; Bricaud and Morel 1986). Using the calculated IOPs, together with concentrations of constituents and ambient factors, a set of $Rrs(\lambda)$ spectra was simulated with Hydrolight (Mobley 1994). Refer to appendix A for more details on data simulation. Finally, using EFAST, the sensitivity of RTM was determined with input parameters (e.g. PSD, concentrations of constituents, and ambient factors) and simulated $Rrs(\lambda)$.

3.3 Sensitivity analysis for QAA and LMI

A IOPs inverse algorithm is controlled by various factors excluding remote sensing reflectance, such as light field of reflectance model (Gordon *et al.* 1988; Garver and Siegel 1997). In order to improve or propose an algorithm for deriving IOPs, it is necessary to get insight into the dependence of derived IOPs on the uncertainties of model parameters. Here, EFAST was conducted to determine the sensitivity of the QAA- and LMI-derived IOPs to the uncertainties of measured pure absorption and scattering,

the light field, and the S and η values. The SA approach for QAA and LMI is presented in Fig. 2.3(B). In this part, the sensitivity of the derived IOPs by QAA and LMI to the model parameters (e.g. g_0 and g_1), as shown in table 2.3, was determined with four $Rrs(\lambda)$ spectra as input, respectively. For each $Rrs(\lambda)$ spectrum, sensitivity analysis for QAA and LMI was conducted through the following steps:

(1) Generating random samples of model parameters (table 2.3) according the predefined distribution through EFAST;

(2) Deriving IOPs through QAA and LMI using a spectrum of $Rrs(\lambda)$ in Fig. 2.2 and the generated samples of model parameters;

(3) Using EFAST to determine the sensitivity of IOPs derived from QAA and LMI to the analyzed model parameters.

4. Results

4.1 Sensitivity analysis for RTMs

Fig. 2.4 shows that the concentrations of constituents account for the variation of $Rrs(\lambda)$ in a large proportion. The TSI value for mineral concentration is the largest except for scenario A at wavelength shorter than about 450 nm (Fig. 2.4(A)). It has a trend of increasing with wavelength before reaching the maximum around 600 nm and 700 nm for scenarios A and B, respectively, and then flattens out. The TSI value for $a_{CDOM}(440)$ (indicator of CDOM concentration) decreases exponentially with wavelength and shows insignificant values beyond 600 nm, similar to the exponential decay characteristic of CDOM spectral absorption. The TSI values for chlorophyll concentration mimics its absorption characteristics in spectral shape, such as absorption

peak around 443 nm, yet does not show peak around 675 nm for scenario A, as shown in Fig. 2.4(A) and 2.4(C).

The second factor explaining a large proportion of $Rrs(\lambda)$ variation is the refractive indices of particulate matters (i.e. phytoplankton and minerals). The TSI value for the imaginary part n'_m mineral refractive index is among the largest values in the spectral regions of λ<500 nm for scenario A and λ<600 nm for scenario B. It decreases exponentially with wavelength, and is negligible (TSI<0.02) beyond 600 nm for scenario A and 700 nm for scenario B. The TSI variation for n'_m and the spectral shape of mineral absorption are similar. The TSI values for the real part of mineral refractive index (n_m) weakly depend on wavelength, and is much smaller for fixed backscattering probabilities (Fig. 2.4(A) and 2.4(B)) than those for backscattering probabilities calculated by Twardowski (Fig. 2.4(C) and 2.4(D)). The TSI value for the imaginary part of phytoplankton refractive index (n'_{ph}) is also similar to the variation of the phytoplankton absorption in shape, except for the absorption peak around 675 nm for scenario A, as shown in Fig. 2.4(A) and 2.4(C).

According to Fig. 2.4(A) and 2.4(B), the TSI value for the mineral PSD slope (ξ_m) is negligible (TSI<0.04) in the spectral region of <700 nm for the case of fixed backscattering probability (1.83%), but exceeds 0.04 in the spectral region of >700 nm. The TSI value for the phytoplankton PSD slope (ξ_{ph}) is negligible (TSI<0.02) for the case of fixed backscattering probability (1%). In contrast, for the case that backscattering probabilities are calculated from Twardowski, as shown in Fig. 2.4(C) and 2.4(D), the TSI value for the mineral PSD slope accounts for an important part of the variation of

$Rrs(\lambda)$ though the TSI value for the phytoplankton PSD slope is still negligible (<0.02). Fig. 2.4 also shows that the TSI values for model parameters, S_{NAP} and S_{CDOM}, and the TSI values for light field factors (i.e. wind speed and solar zenith angle) are negligible (TSI<0.02) as compared with the TSI values for other model parameters.

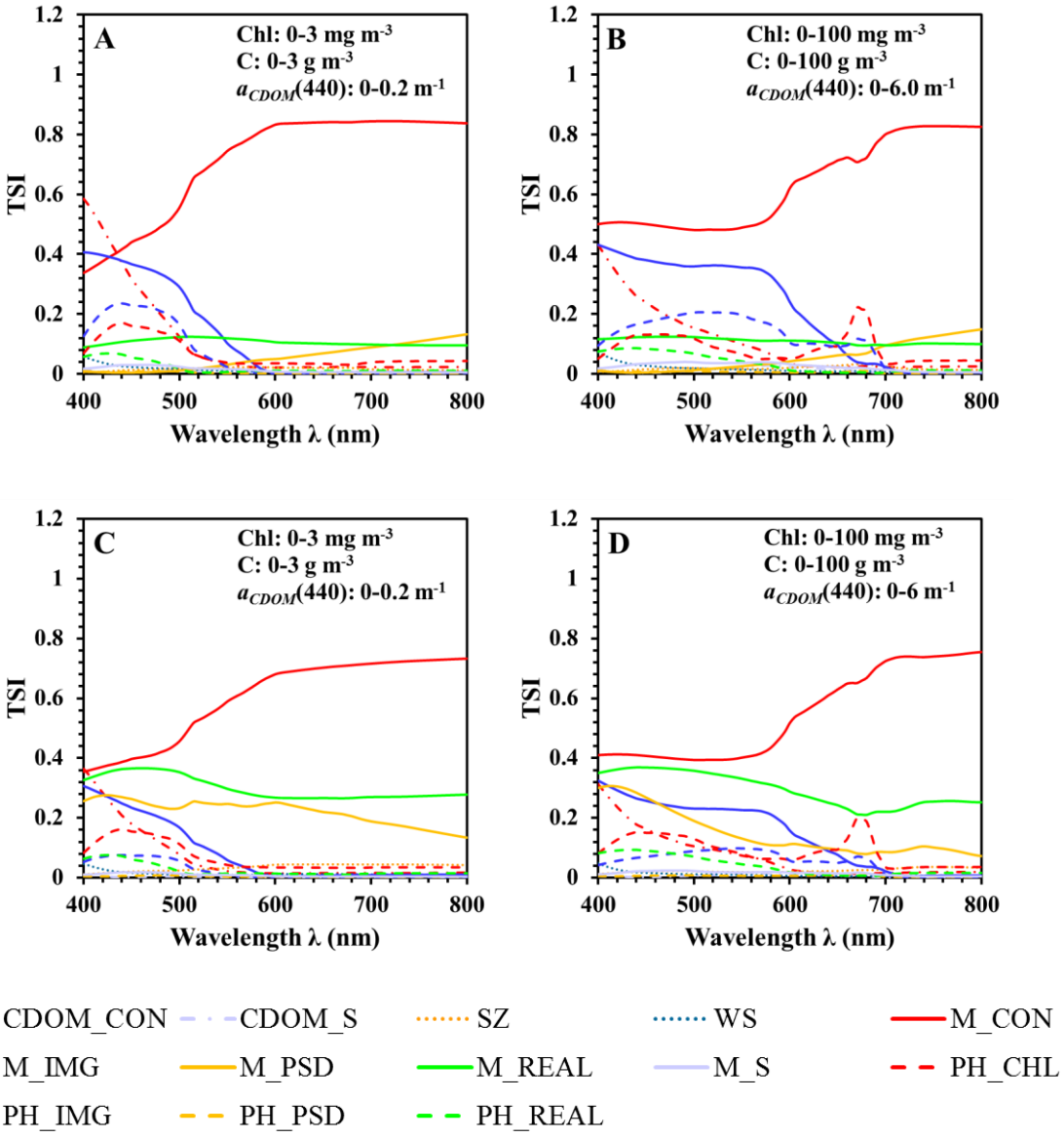
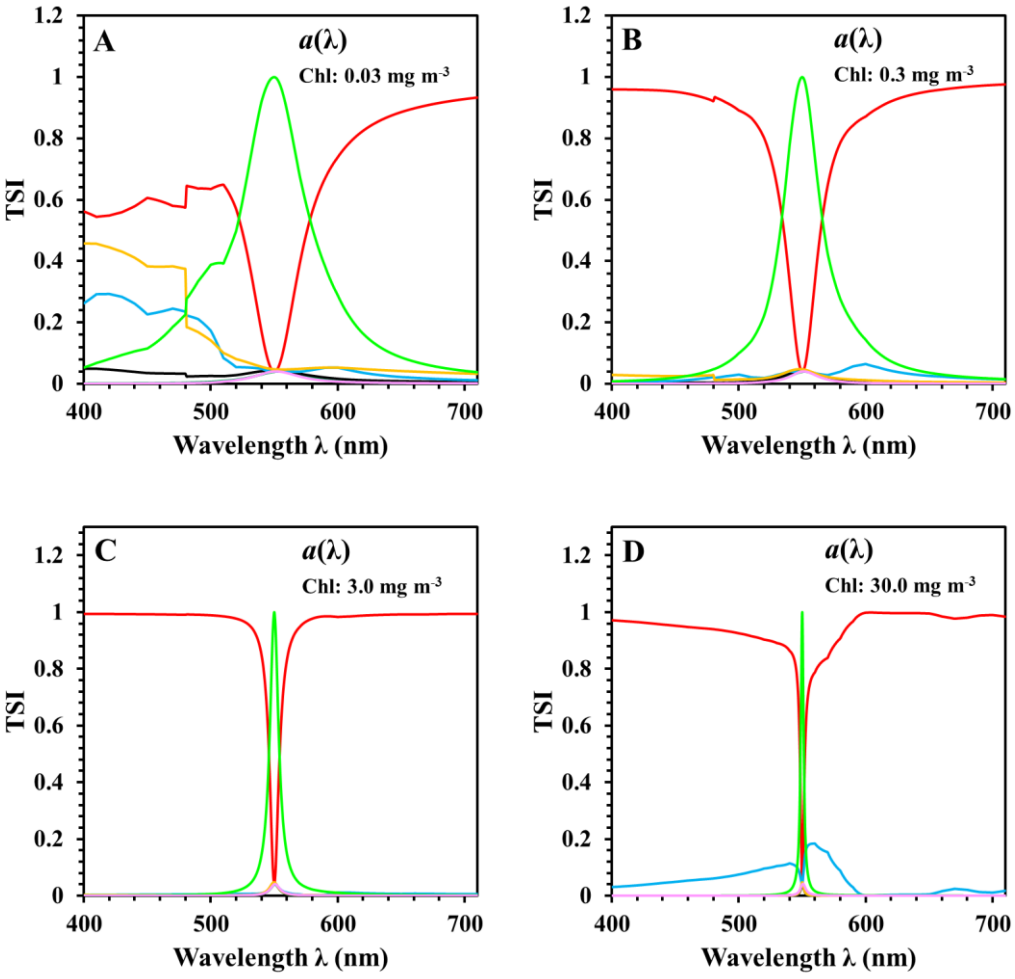


Fig. 2.4. TSI for the $R_{rs}(\lambda)$; A and B for the fixed backscattering probability (phytoplankton 1%, and minerals 1.83%), and C and D for backscattering probability model of Twardowski; CDOM_CON: CDOM absorption coefficient at 440 nm; CDOM

_S: exponential slope of spectral absorption; LF_SZ: solar zenith angle; LF_WS: wind speed; M_CON: mineral concentration; M_IMG: imaginary part of mineral refractive index; M_PSD: slope of mineral PSD; M_REAL: real part of mineral refractive index; M_S: exponential slope of mineral spectral absorption; CHL: chlorophyll concentration; PH_IMG: imaginary part of refractive index of phytoplankton; PH_PSD: slope of phytoplankton PSD; PH_REAL: real part of refractive index of phytoplankton.

4.2 Sensitivity analysis for QAA and LMI

4.2.1 Sensitivity analysis for QAA



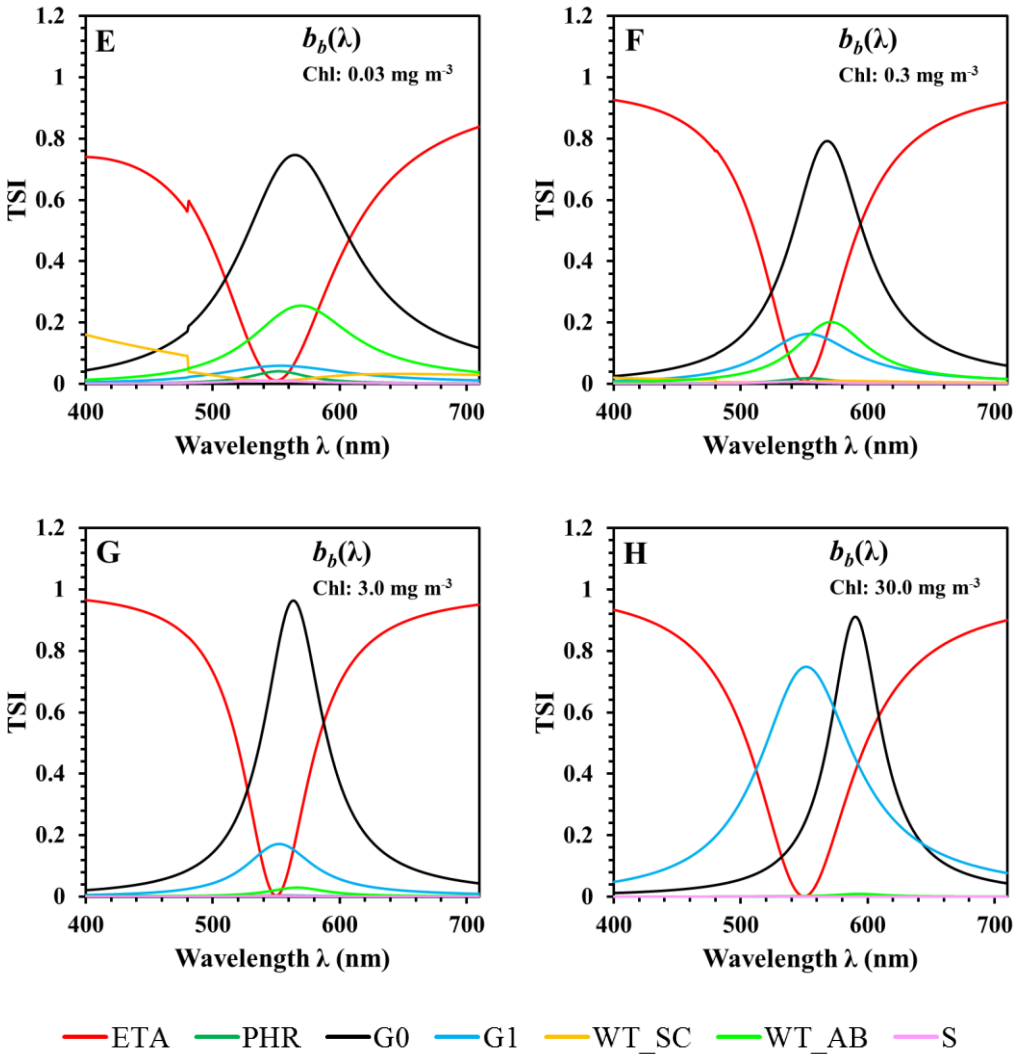


Fig. 2.5. The total sensitivity indices for total absorption and backscattering coefficients derived by QAA; ETA: power-law slope of particulate backscattering coefficient; PHR: ratio of phytoplankton absorption between 411 and 443 nm; G0 and G1: light field; WT_SC: scattering coefficient of water; WT_AB: absorption coefficient of water; S: exponential slope of CDM absorption coefficient.

The description of the SA results for QAA is focused how the QAA-derived $a(\lambda)$ and $b_b(\lambda)$ (Fig. 2.5) as well as $a_{ph}(\lambda)$ and $a_{CDM}(\lambda)$ (Fig. 2.6) are sensitive to the variation of the analyzed parameters in table 2.3. The impacts of model parameters on the

derived absorption and scattering spectra may be introduced in the four step calculations: (1) estimating $a(\lambda)$ at the reference wavelength (step 2 in table 2.7); (2) calculating the particulate backscattering coefficient at the reference wavelength by solving reflectance model (step 2 in table 2.7); (3) deriving particulate backscattering coefficients at other wavelengths (step 3 in table 2.7); and (4) obtaining $a(\lambda)$ (step 3 in table 2.5).

The first step is to estimate $a(\lambda)$ at the reference wavelength (λ_0=555 nm) (step 2 in table 2.7) and uncertainties were introduced by the measurement uncertainties of absorption coefficient of pure water. QAA directly calculates $a(\lambda_0)$ from $rrs(\lambda)$ without introducing other model parameters other than measurement uncertainties of absorption coefficient of pure water. Therefore the TSI value for the uncertainty of measured pure water absorption approximates to one at 555 nm for all the considered concentration combinations, as shown in Fig. 2.5(A)-(D).

The second step is to calculate the particulate backscattering coefficient at the reference wavelength (λ_0=555 nm) by solving reflectance model (step 2 in table 2.7) and uncertainties were introduced by the absorption and backscattering coefficients of pure water and the effects of ambient factors. The influences of light field parameters are introduced when calculating the backscattering coefficient at the reference wavelength (555 nm). Fig. 2.5(E)-(H) show that light field parameters mainly govern the variation of $b_b(\lambda)$ at the reference wavelength (555 nm), and the TSI value for g_0 approximates to 0.6. However, for the high combination (chlorophyll concentration is 30 mg m^{-3}), the importance of g_0 in regulating $b_b(\lambda)$ decreases, but the TSI value of g_1 increases to its maximum (~0.8).

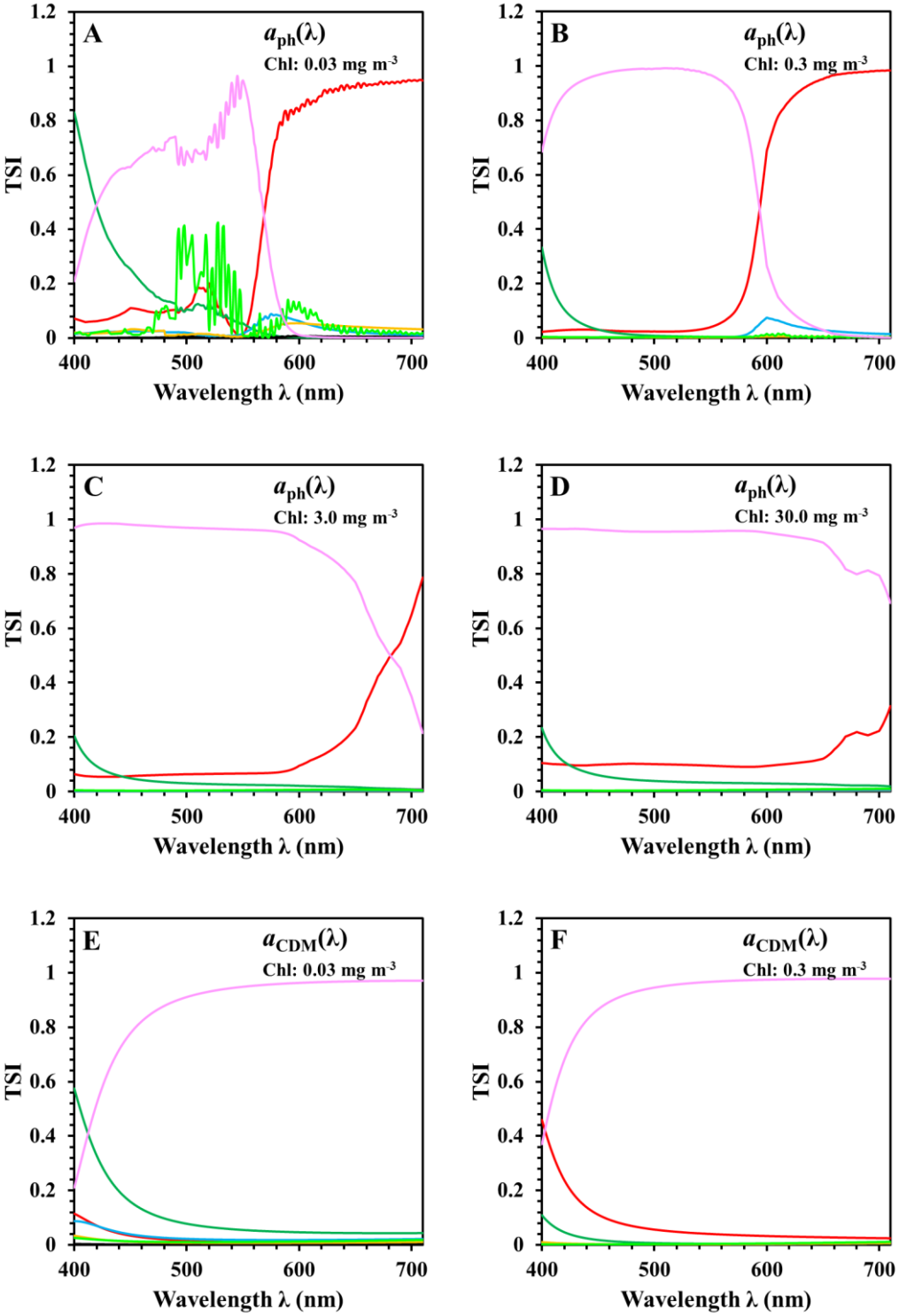
The third step is to derive particulate backscattering coefficients at other wavelength (step 3 in table 2.7) and uncertainties were introduced from $b_b(\lambda_0)$ and η. In this step, uncertainties of two new parameters may be introduced. According to Fig. 2.5(E)-(H), the TSI value for η increases with wavelength being far away from the reference wavelength (λ_0=555 nm). The contribution from the uncertainty of measured pure water scattering coefficients to the variation of estimated $b_b(\lambda)$ is apparent (0.05<TSI<0.08) for low combination in the spectral region of <450 nm, as shown in Fig. 2.5(E). Fig. 2.5(E)-(H) also show that the influences of light field parameters, which are introduced in calculating backscattering coefficient at the reference wavelength, are propagated to $b_b(\lambda)$ when a considered wavelength (λ) is close to the reference wavelength. Furthermore, while the TSI peak for g_0 centers at reference wavelength (λ_0=555 nm), the TSI peak for g_1 TSI moves to a wavelength longer than the reference wavelength. When the concentrations of constituents are low, the uncertainty in measured water absorption coefficients introduced in calculating absorption coefficient at the reference wavelength in step 2 is also propagated to the $b_b(\lambda)$ when a considered wavelength (λ) is close to the reference wavelength, as shown in Fig. 2.5(E) and 2.5(F).

The fourth step is to derive $a(\lambda)$ (step 3 in table 2.7) and uncertainties were from $b_b(\lambda)$ and ambient factors. In this step, uncertainties of five parameters may be introduced into $a(\lambda)$: (1) η for calculating particulate backscattering; (2) uncertainties in measured pure water scattering at wavelength λ, which affects $b_b(\lambda)$; (3) light field parameters for deriving $a(\lambda)$ from reflectance model; and (4) uncertainties in measured pure water absorption coefficients at wavelength λ. As shown in Fig. 2.5(A)-(D), the TSI

value for η increases with wavelength (λ) being far away from the reference wavelength, and it mainly controls the variation of $a(\lambda)$. The TSI value for the uncertainty of measured pure water scattering coefficient is around 0.4 in the spectral region of <480 nm (Fig. 2.5(A)). For light field parameters, their TSI values occupy an important portion in the sum of TSI in the spectral region of <500 nm for the case of low combination (Chl =0.03 mg m^{-3}). According to Fig. 2.2, this region show high $Rrs(\lambda)$ value, meaning that scattering is intensive and the second term of the reflectance model is important. Comparing Fig. 2.2 with Fig. 2.5(D), g_1 also controls the variation of $a(\lambda)$, but it is insignificant when compared with η for high concentrations of constituents. Uncertainty in measured pure water absorption is one of the major factors controlling $a(\lambda)$ for low concentrations of constituents (Fig. 2.5(A) and 2.5(B)), and its TSI value decreases with a considered wavelength being away from the reference wavelength (Fig. 2.5(A)-(D)).

From Fig. 2.6, it is evident that the phytoplankton absorption, $a_{ph}(\lambda)$, is mainly regulated by S, PHR ($a_{ph}(411)/a_{ph}(443)$), and η. The TSI value for S is the leading factor determining the variation of $a_{ph}(\lambda)$. The TSI value for PHR is important only for the case of low concentration combination (Fig. 2.6(A)). The TSI value for η is significant for the first three concentration combinations in the spectral regions >600 nm. In Fig. 2.6, it is also evident that the absorption of CDM ($a_{CDM}(\lambda)$) is mainly controlled by S, PHR, and η. The TSI value for S is the largest value in the spectral region of >440 nm, and account for over 90% of the sum of TSI. The TSI value PHR is significant for the low concentration combination at the region of 400-440 nm (Fig.

2.6(E)). The TSI value for η also occupies a large portion of the sum of TSI in the spectral region of <500 nm (Fig. 2.6(F)-(H)).



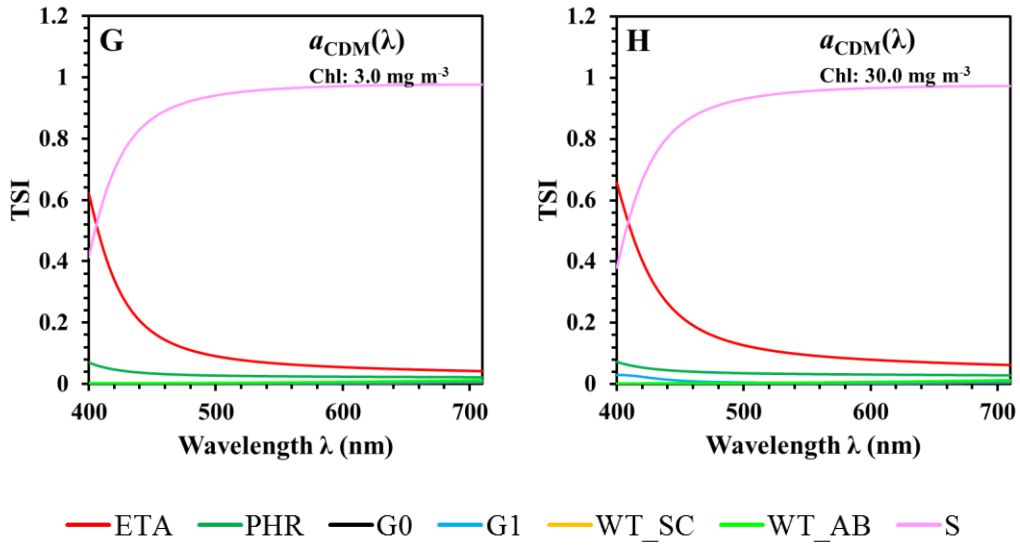


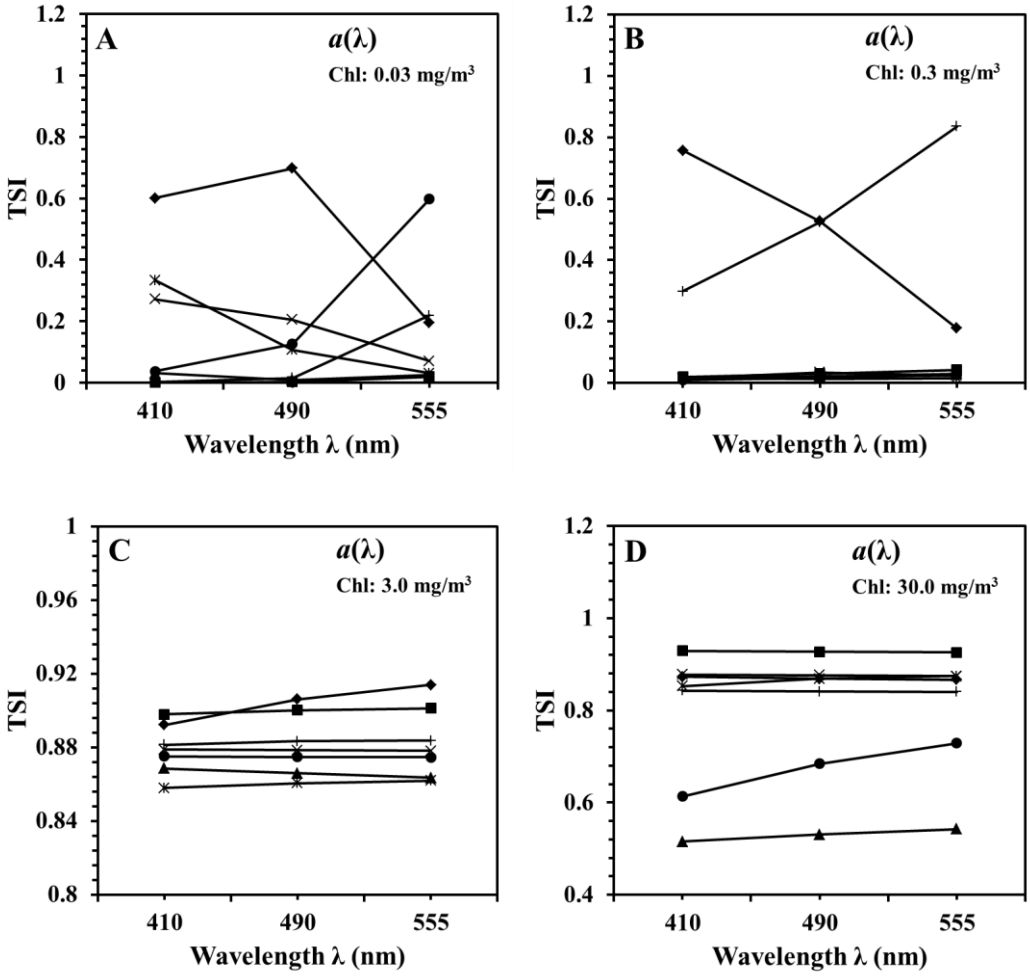
Fig. 2.6. The total sensitivity indices for absorption coefficients of phytoplankton and CDM derived by QAA; ETA: power-law slope of particulate backscattering coefficient; PHR: ratio of phytoplankton absorption between 411 and 443 nm; G0 and G1: light field; WT_SC: scattering coefficient of water; WT_AB: absorption coefficient of water; S: exponential slope of CDM absorption coefficient.

4.2.2 Sensitivity analysis for LMI

Fig. 2.7 and 2.8 display the SA results for the absorption and backscattering coefficients derived by LMI. $a(\lambda)$ is mainly controlled by η, g_1, and uncertainties of measured pure water absorption and scattering coefficients for low concentration combination (Chl =0.03 mg m^{-3}) (Fig. 2.7(A)). When chlorophyll concentration is 0.3 mg m^{-3}, the IOPs shape parameters, S and η, mainly control the variation of derived $a(\lambda)$ (Fig. 2.7(B)). However, $a(\lambda)$ is simultaneously regulated by all the parameters when chlorophyll concentrations are high (Fig. 2.7(C) and 2.7(D)). Similar to $a(\lambda)$, the variation of the derived $b_b(\lambda)$ is mainly determined by η and S, and simultaneously

governed by all the parameters when chlorophyll concentration higher than 3.0 mg m^{-3} (Fig. 2.7(G), and 2.7(H)).

Fig. 2.8(A) and 2.8(B) show that the variation of $a_{ph}(\lambda)$ is mainly governed by S. Quantitatively, the TSI/Σ(TSI) (Σ(TSI) is the sum of TSI) for S ranges from 69.8% (Chl =0.03 mg m^{-3}) for band 555 nm to 90.3% (Chl =0.3 mg m^{-3}) for band 410 nm. However, $a_{ph}(\lambda)$ are simultaneously governed by all the parameters (Fig. 2.8(C) and 2.8(D)), when the concentrations of constituents are high. Similar to $a_{ph}(\lambda)$, $a_{CDM}(\lambda)$ is also majorly governed by S, and simultaneously governed by all the parameters when chlorophyll concentration higher than 3.0 mg m^{-3} (Fig. 2.8(G), and 2.8(H)).



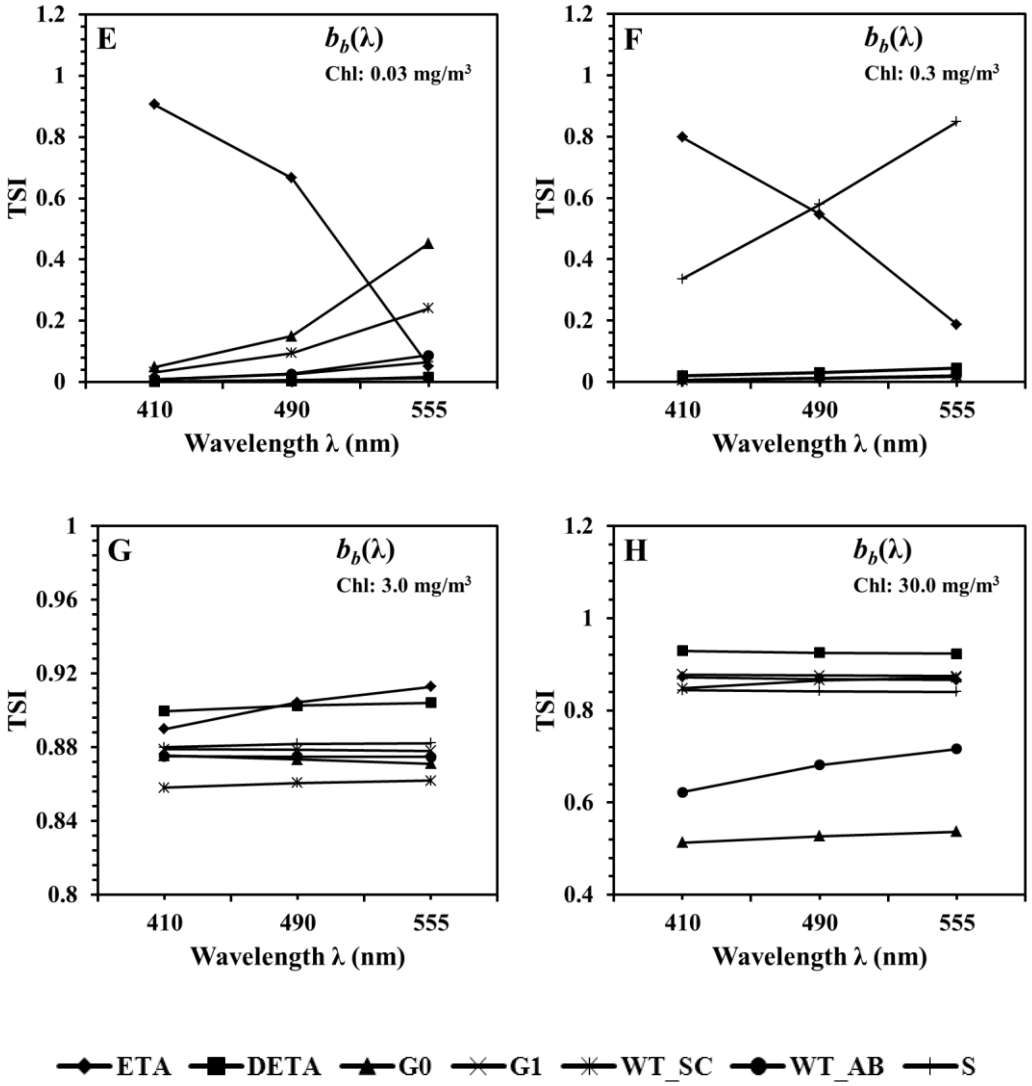
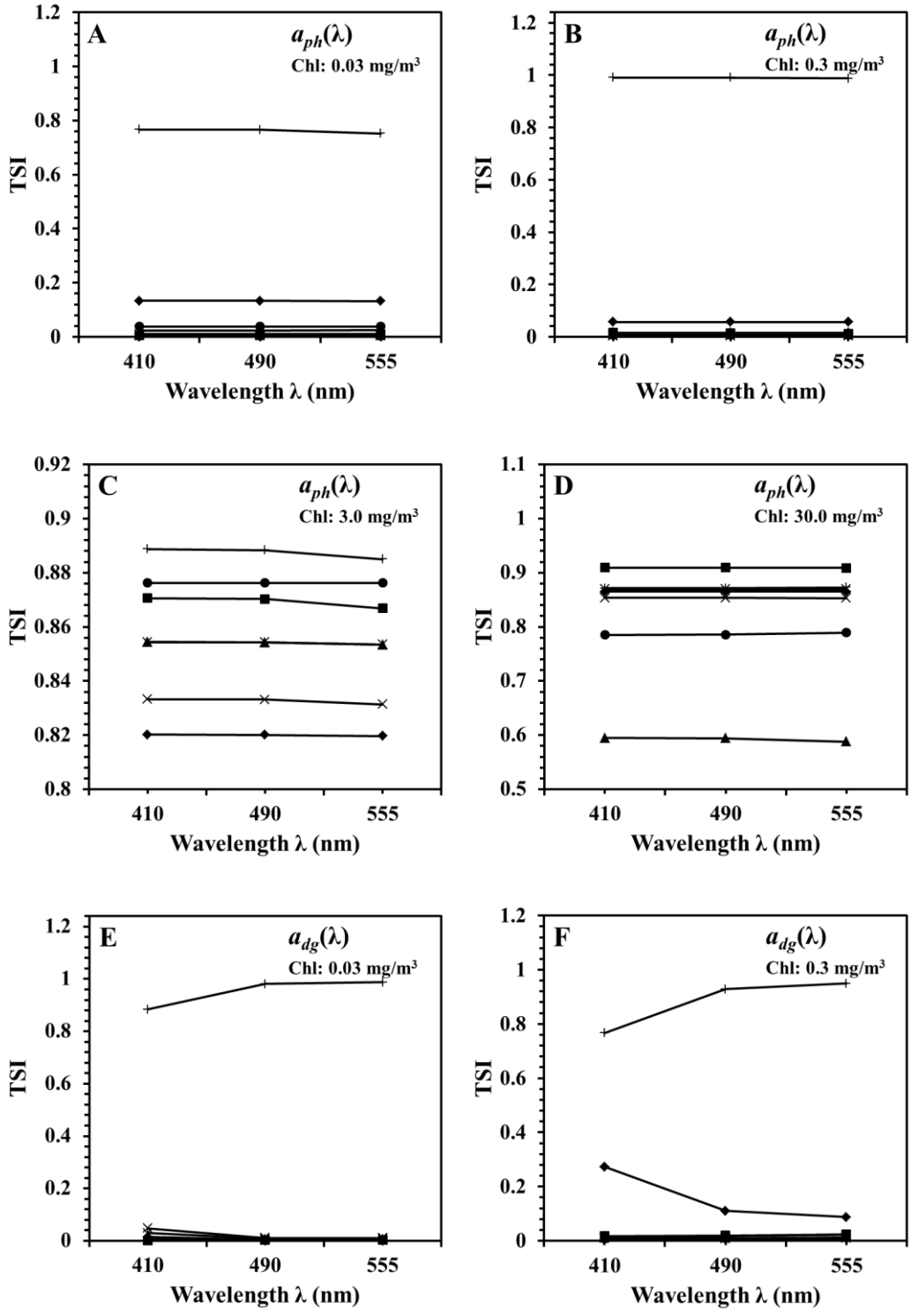


Fig. 2.7. The total sensitivity indices for total absorption and backscattering coefficients derived by LMI; ETA: power-law slope of particulate backscattering coefficient; PHR: ratio of phytoplankton absorption between 411 and 443 nm; G0 and G1: light field; WT_SC: scattering coefficient of water; WT_AB: absorption coefficient of water; S: exponential slope of CDM absorption coefficient.


A
$a_{ph}(\lambda)$
Chl: 0.03 mg/m³
B
$a_{ph}(\lambda)$
Chl: 0.3 mg/m³
C
$a_{ph}(\lambda)$
Chl: 3.0 mg/m³
D
$a_{ph}(\lambda)$
Chl: 30.0 mg/m³
E
$a_{dg}(\lambda)$
Chl: 0.03 mg/m³
F
$a_{dg}(\lambda)$
Chl: 0.3 mg/m³
TSI
Wavelength λ (nm)
410
490
555

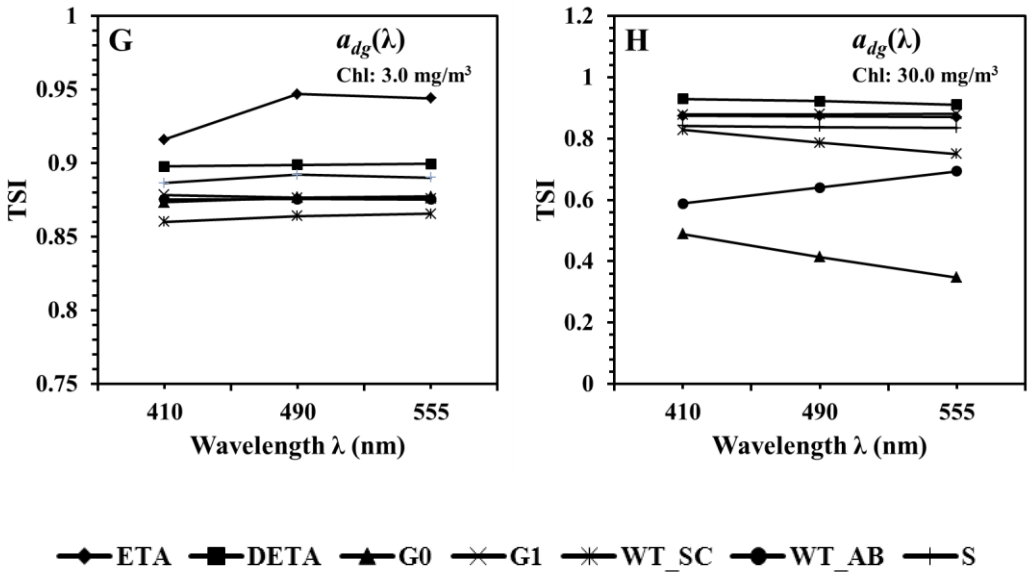


Fig. 2.8. The total sensitivity indices for absorption coefficients of phytoplankton and CDM derived by LMI; ETA: power-law slope of particulate backscattering coefficient; DETA: full-width at half max; G0 and G1: light field; WT_SC: scattering coefficient of water; WT_AB: absorption coefficient of water; S: exponential slope of CDM absorption coefficient.

5. Discussion

5.1 RTM

The results from SA for remote sensing reflectance indicate that the concentration, the mineral refractive index, and the mineral particle size primarily control the variation of remote sensing reflectance. The concentrations of constituents mainly regulate the variation of $Rrs(\lambda)$ for the following reasons. For NAP, it governs the variation of $Rrs(\lambda)$ mainly through scattering process (Stramski *et al.* 2004), which is directly related to $rrs(\lambda)$ according to equation 1. Therefore, $Rrs(\lambda)$ is extremely sensitive to the variation of minerals concentration, especially in the long-wavelength region (>700 nm). This is consistent with the conclusions of previous studies based on simulated

(Sathyendranath *et al.* 1989) and in-situ measured datasets (Toole and Siegel 2001; Lubac and Loisel 2007). The TSI value for $Rrs(\lambda)$ also indicates that the estimation of mineral concentration should use one or more near infra-red bands, where mineral concentration dominate in governing $Rrs(\lambda)$ variation. The influence of CDOM on $Rrs(\lambda)$ is mainly through absorption process (Toole and Siegel 2001; Lubac and Loisel 2007; Wang *et al.* 2010), and as a result its spectra for TSI and absorption are similar in shape (decreasing exponentially with wavelength). For chlorophyll concentration, the spectrum of its TSI, such as the peak around 443 nm and 675 nm for high combinations, shons similarity to that of absorption. This similarity in spectral shape can be attributed to its impact on $Rrs(\lambda)$ mainly through absorption process (Toole and Siegel 2001; Lubac and Loisel 2007; Wang *et al.* 2010).

The refractive index and PSD of constituents also make great contribution to the variation of $Rrs(\lambda)$. As the absorption efficiency of a constituent is linked to particle size and the imaginary part of refractive index through Mie theory (Morel and Bricaud 1981), the TSI spectra of n'_m and n'_{ph} are similar in shape to the absorption coefficient spectra. Small contribution of the phytoplankton PSD to the variation of $Rrs(\lambda)$ can be attributed to: (1) backscattering coefficient is more sensitive to PSD than absorption (Wozniak and Stramski 2004); and (2) phytoplankton regulates the variation of $Rrs(\lambda)$ mainly through absorption process. This result indicates that it is a challenge to retrieve phytoplankton size compositions through remote sensing (Brewin *et al.* 2011). In contrast, the PSD of minerals makes significant contribution to the variation of $Rrs(\lambda)$, because minerals regulate $Rrs(\lambda)$ mainly through backscattering coefficient, which is sensitive to PSD

(Stramski and Kiefer 1991; Wozniak and Stramski 2004). The real part of refractive index is linked to particle density, which determines mass-specific backscattering coefficient (Bricaud and Morel 1986; Wozniak and Stramski 2004) and backscattering probability (Twardowski *et al.* 2001; Loisel *et al.* 2007), which in turn regulate $Rrs(\lambda)$. Therefore, real part of refractive index for minerals is a key factor controlling the variation of mainly through scattering process.

However, several cautions should be pointed out when considering the SA result for remote sensing reflectance. This analysis is based on the following assumptions. For the calculation of averaged efficiencies of phytoplankton and minerals, particles are assumed to be spherical and homogeneous. However, this assumption may be violated, especially for larger particles (Clavano *et al.* 2007). The backscattering coefficient of non-spherical particles may be underestimated under this assumption (Clavano *et al.* 2007), and the corresponding backscattering efficiency may be significantly underestimated based on the spherical particle assumption in the relevant range (Kostadinov *et al.* 2009). Although aquatic particles follow Junge distribution in most cases, exceptions do exist (Peng and Effler 2007; Kostadinov *et al.* 2009). Second, particulate refractive index and PSD are randomly generated for given ranges extracted from previous studies. However, for a local region, these parameters may be correlated, and thus some combinations may be artificial. Third, the concentrations of constituents are assumed to be independent in order to simplify analysis. This assumption is expected to be violated for open oceanic waters.

5.2 QAA

First, uncertainties of measured pure water absorption and scattering will influence the bio-optical model for oligotrophic waters, which is consistent with result in previous

study (Morel and Maritorena 2001). Therefore, measurement uncertainties in pure water absorption and scattering should be taken into account for oligotrophic waters. Second, η is the major factor controlling the uncertainties of derived $a(\lambda)$ and $b_b(\lambda)$, and as a result methods to obtaining accurate value of η is required to further refine QAA model. This result is consistent with that obtained by (Lee *et al.* 2010) through different method. Third, the variation of light field factors has significant impact on the estimation of $b_b(\lambda)$. As light field factors are complicated functions of wind speed, solar zenith angle, and IOPs, further studies may be required to refine QAA algorithm, such as modeling molecular and particle scatterings with separate light field parameters (Lee *et al.* 2004), or employing a higher-order reflectance model (Albert and Mobley 2003). Furthermore, g_1 regulates the estimation of $b_b(\lambda)$ around reference wavelength and the estimation of $a(\lambda)$ for the wavelength domain when scattering is intensive, indicating the important contribution of multiple scattering to $Rrs(\lambda)$. Therefore, the squared term in $Rrs(\lambda)$ is not negligible for highly scattering waters in QAA model.

According to the estimations of CDM and phytoplankton absorption spectra, S is the key factor controlling the estimated values. S is more important than PHR in estimating absorption coefficients of phytoplankton and CDM, and should be improved first in future study. This finding is the same as previous study (Lee *et al.* 2010).

The SA for QAA is based on the following assumptions. First, the model of $a(\lambda)$ at the reference wavelength is accurate, and only the influences of light field factors and measurement uncertainty of pure water absorption coefficient are introduced into the model. However, this assumption is expected to be violated. Actually, an exponential

relationship between the QAA-derived $a(\lambda_0)$ and the corresponding error relative to lab measured $a(\lambda_0)$ exists though for the case of $a(550)$ <0.1 m^{-1}, this relative error is in general <16% (Lee et al. 2010). While η makes the greatest contribution to the variation of $a(\lambda)$ for high combinations according to results in this study, the uncertainty in $a(\lambda_0)$ accounts for larger proportions of uncertainty in the derived $a(\lambda)$ (Lee et al. 2010). Therefore, the uncertainty in $a(\lambda_0)$ model may be considered in the future study. Second, only four combinations of concentrations are analyzed. Therefore, the complicated interactions between constituents were not analyzed in this study.

5.3 LMI

The SA results for LMI show that, when the concentrations of constituents are low, the derived IOPs are mainly controlled by the model parameters IOPs, such as S and η, indicating that LMI may perform well when these parameters are appropriately modeled. However, when the concentrations of constituents are high, the derived IOPs values are sensitive to all the parameters, implying that all the model parameters highly interact with each other. This high degree interaction among model parameters may be due to the fact that unlike QAA algorithm, LMI is totally based on the spectral shape of IOPs models and derives the IOPs value through solving a linear system with the predefined spectral model, explaining high TSI values for all the analyzed parameters, especially for the case of high concentrations. It suggests that improving the performance of the model may require accurately modeling all these parameters.

5.4 Implication for ocean color algorithms

The SA results show that $Rrs(\lambda)$ is controlled by various factors, including those independent of wavelength. Therefore, sensitivity analysis was applied to the band ratios of $Rrs(\lambda)$ used by OC4v4 to determine its ability to avoid the influences of some uninterested factors. The OC4v4 algorithm is based on the band ratios between 443, 490, 510 and 555 nm, and is described in detail in (O'Reilly *et al.* 1998). Here, band 443 nm was replaced with 440 nm, in that 443 nm was not simulated and their difference is negligible. The sensitivities of band ratios to the analyzed factors are shown in Fig. 2.9. For the case of low combination (Fig. 2.5(A)), the TSI values for Chl , C and $a_{CDOM}(440)$ account for about 73% of total TSI. $TSI / \sum TSI_x$ ranges from 20.3% to 25.4% for Chl , from 21.8% to 25.7% for C , and from 23.8% to 39.4% for $a_{CDOM}(440)$. The percentage for n'_m TSI value is reduced to from 20.7% to 10%, while the TSI percentage of n'_{ph} is reduced to from 12.9% to about 6%. For the case of Fig. 2.5(B), the concentrations of constituents approximately account for 53% variation of the band ratios. This percentage is lower than that of Fig. 2.5(A). $TSI / \sum TSI_x$ ranges from 13.8% to 27.9% for Chl , from 9.5% to 16.1% for C ; and from 16.4% to 25.2% for $a_{CDOM}(440)$. The same TSI ratios for the refractive indices, n'_m and n'_{ph} account for about 9% and 6% of the variations of band ratios, respectively. However, this TSI ratio for the exponential slope of CDOM absorption spectrum increases significantly from 16.4% to 25.2%.

The SA results for band ratios indicate that the designed band ratios for the OC4v4 algorithm is effective in accounting for the component concentrations, and the effects of particulate characteristics, such as real part of mineral refractive index, can be

significantly reduced through band ratios, which is consistent with the results in previous studies (Moore *et al.* 1999; Doxaran *et al.* 2002). If $a_{CDOM}(440)$ and C are very low or highly correlated with Chl, OC4v4 should be robust. Otherwise, this algorithm may fail to work well, because other factors also account for a large proportion of the variations of band ratios. In addition, changes in particulate characteristics could also be reflected band-ratio variations. For example, the performance of the algorithm would be degraded by the variation of n'_m, which is consistent with the results of previous studies (Wozniak and Stramski 2004; Zhou and Cao 2008).

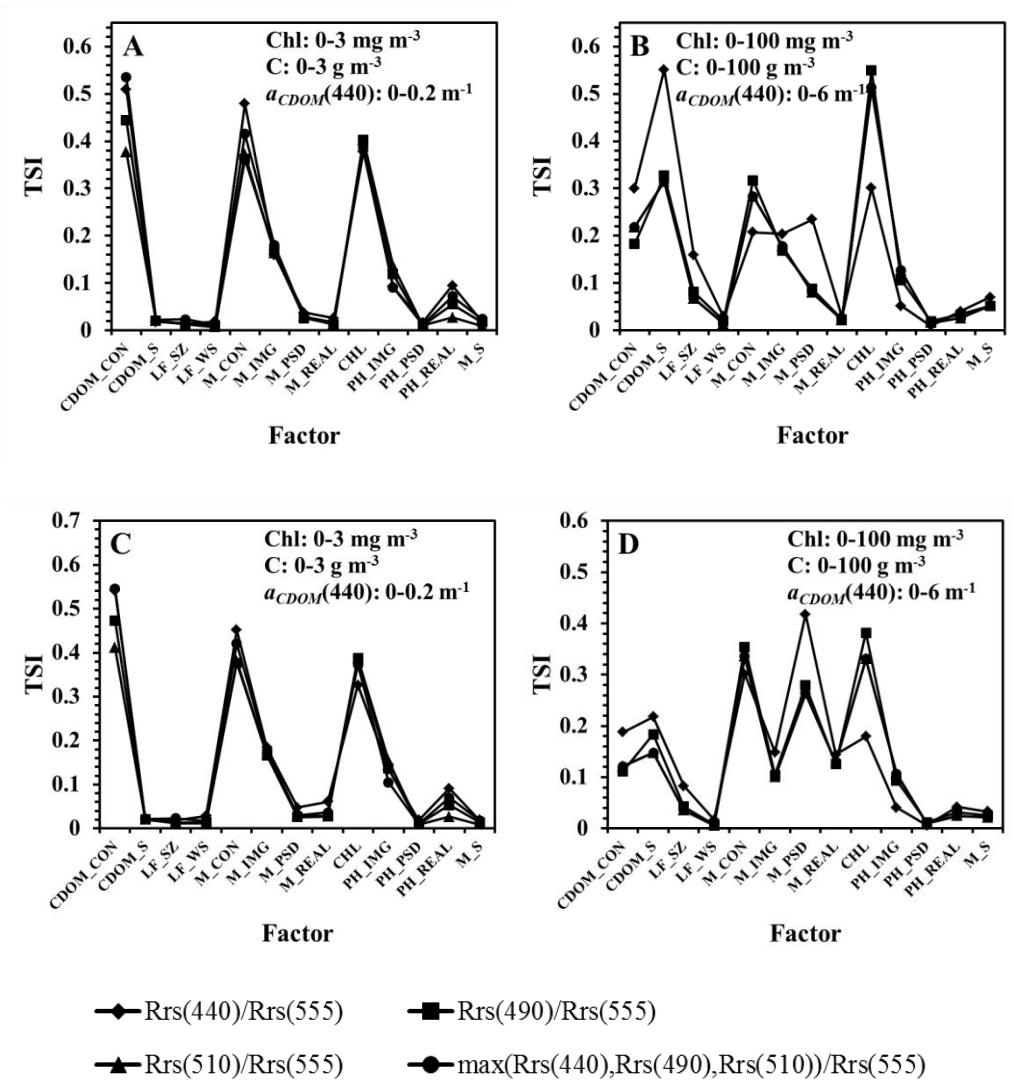


Fig. 2.9. TSI of band ratios; A and B for fixed backscattering probability (phytoplankton 1%, and minerals 1.83%), and C and D backscattering probability model of Twardowski; $Rrs(\lambda_1)/Rrs(\lambda_0)$ is the ratio of $R_{rs}(\lambda)$ between wavelengths λ_1 and λ_0 ,λ_1 can be 440, 490 and 510 nm, and λ_0 is 555 nm; max is an operation to obtain the maximum; CDOM_CON: CDOM absorption coefficient at 440 nm; CDOM _S: exponential slope of spectral absorption; LF_SZ: solar zenith angle; LF_WS: wind speed; M_CON: mineral concentration; M_IMG: imaginary part of mineral refractive index; M_PSD: slope of mineral PSD; M_REAL: real part of mineral refractive index; M_S: exponential slope of mineral spectral absorption; CHL: chlorophyll concentration; PH_IMG: imaginary part of refractive index of phytoplankton; PH_PSD: slope of phytoplankton PSD; PH_REAL: real part of refractive index of phytoplankton.

5.5 Improved linear matrix inverse (ILMI)

Based on the above SA results, an improved algorithm, improved linear matrix inverse (ILMI), for deriving $a(\lambda)$ and $b_b(\lambda)$ were developed. Two improvements were introduced into the ILMI algorithm: (1) band 710 nm was employed to accommodate the variation of mineral (Babin and Stramski 2002; Doron *et al.* 2007); and (2) QAA equations for η and S were employed. As a result, unlike LMI, an over constrained system was constructed with $Rrs(\lambda)$ bands of 410, 440, 490, 510, 555 and 710 nm. The IOCCG simulated dataset was used to test compare the algorithm with QAA and LMI.

For the case that Chl ranges from 0.03 to 30 mg m^{-3}, root-mean-square error (RMSE) and regression results between the derived and known IOPs are shown in table 2.5. Generally, the results of ILMI are better than those of QAA and LMI. For example, the RMSE of $a(410)$ derived by ILMI is 0.1241 which is 0.5031 and 0.2931 lower than

those of QAA and LMI, respectively. The RMSE values ILMI-derived IOPs at wavelengths other than 410 nm are also lower than those of QAA and LMI. Furthermore, the scatter plots between simulated and derived $a(490)$ and $b_b(490)$ are shown in figure 10 to compare the performances of ILMI, QAA, and LMI. The results for QAA and LMI show that: (1) $a(490)$ and $b_b(490)$ are overestimated or under estimated, especially when their values are high; and (2) the results are dispersed when their values are high. However, according to figure 10, the results of QAA and LMI seem to converge to the 1:1 line at the regions of low values. To further demonstrate these findings, the second group of experiments was conducted.

For the case that *Chl* ranges from 0.03 to 0.3 mg m^{-3}, the RMSE and regression results between the derived and known IOPs are shown in table 2.6. By comparing tables 2.5 with 2.6, the performance of the ILMI is almost the same; however, the performances of QAA and LMI for the low *Chl* are highly improved with respect to RMSE, R^2, and slope; and the performances of the ILMI, QAA, and LMI are competitive.

The experimental results indicate that ILMI is superior to the QAA and LMI when the total absorption and backscattering coefficients are high, however the three algorithms perform well when concentrations are low. These may be ascribed to the following reasons that: (1) ILMI incorporates a band with long wavelength (710 nm) to isolate the influence of the minerals, which is an important factor controlling the variation of $Rrs(\lambda)$ (Lee 2006); (2) QAA and LMI are based on $Rrs(\lambda)$ blue and green bands which are simultaneously influenced by various factors, which may degrade the performances of the algorithms when the values of $a(490)$ and $b_b(490)$ are high; and (3) compared with LMI, the improved algorithm has incorporated the S and η models, and their variations

significantly influence the performance of the LMI, especially for high concentration combinations.

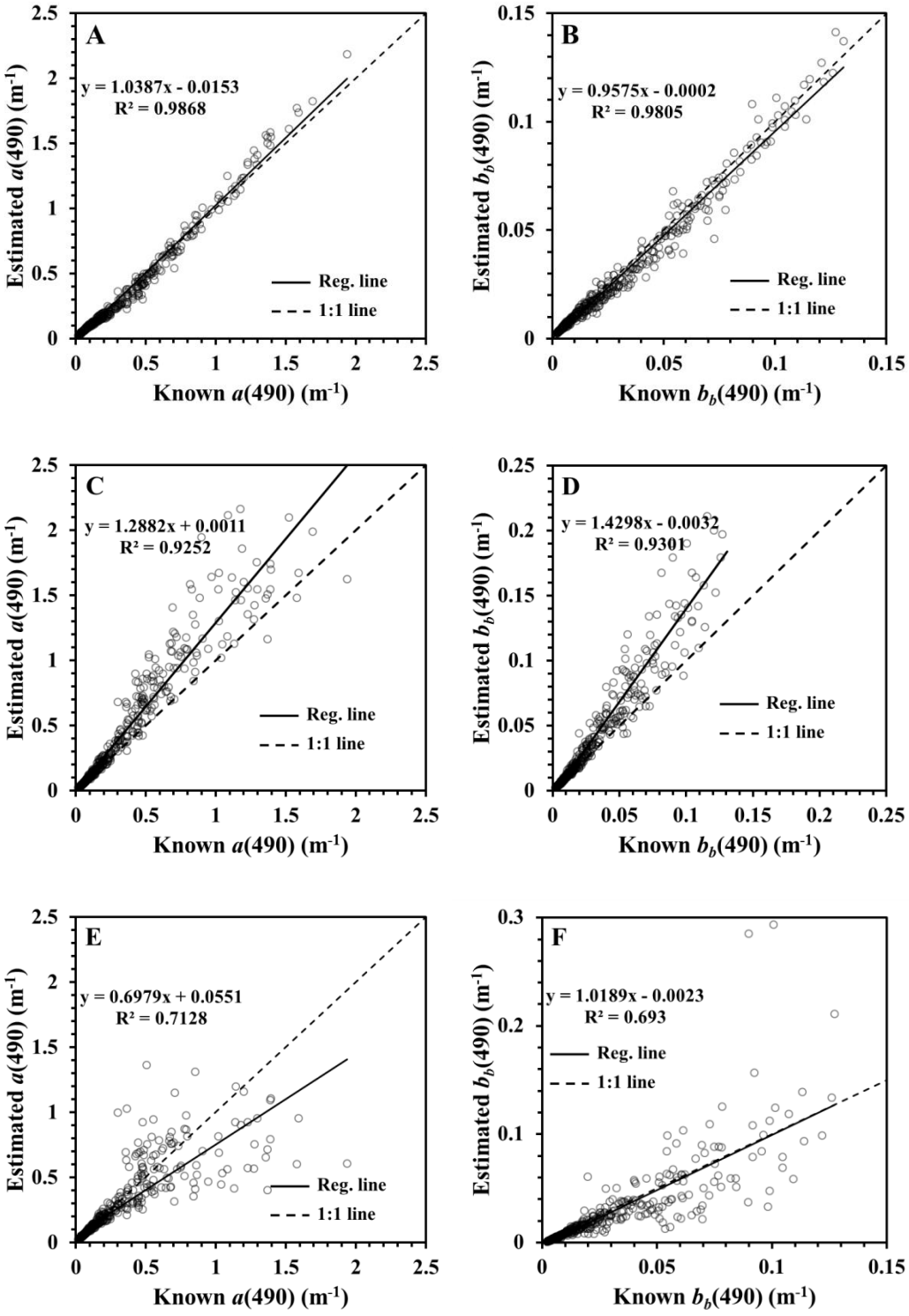


Fig. 2.10. The scatter plots between the derived and the known total absorption and backscattering coefficients at 490 nm; A and B for the ILMI; C and D for the QAA; and E and F for the LMI.

Table 2.5. RMSE and regression results between the derived and the known IOPs for the IOCCG simulated data set with a solar zenith angle of 30^0; chlorophyll concentration ranges from 0.03 to 30 mg m^{-3}; slope and intercept is regression line slope and intercept; N is the number of spectra of sample; n is the number of samples successfully derived; and slope, intercept, N, and n have the same meaning for the tables hereafter

Algorithm	IOPs	RMSE	R^2	Slope	Intercept	N	n
ILMI	a(410)	0.1241	0.9854	1.0436	- 0.0210	500	500
	a(440)	0.0703	0.9898	0.9933	- 0.0213	500	500
	a(490)	0.0440	0.9868	1.0387	- 0.0153	500	500
	a(510)	0.0339	0.9853	0.9982	- 0.0079	500	500
	a(555)	0.0263	0.9919	0.8597	+0.0091	500	500
	b_b(410)	0.0043	0.9768	0.9761	- 0.0003	500	500
	b_b(440)	0.0041	0.9788	0.9794	- 0.0003	500	500
	b_b(490)	0.0042	0.9805	0.9575	- 0.0002	500	500
	b_b(510)	0.0043	0.9812	0.9499	- 0.0003	500	500
	b_b(555)	0.0045	0.9819	0.9340	- 0.0003	500	500
QAA	a(410)	0.6272	0.9294	1.4610	- 0.0483	500	500
	a(440)	0.4296	0.9278	1.4263	- 0.0196	500	500
	a(490)	0.1818	0.9252	1.2882	+0.0011	500	500
	a(510)	0.1288	0.9233	1.2340	+0.0025	500	500
	a(555)	0.0650	0.9244	1.1991	+0.0013	500	500
	b_b(410)	0.0186	0.9338	1.4648	- 0.0048	500	500
	b_b(440)	0.0187	0.9332	1.4668	- 0.0040	500	500
	b_b(490)	0.0181	0.9301	1.4298	- 0.0032	500	500
	b_b(510)	0.0178	0.9293	1.4172	- 0.0031	500	500
	b_b(555)	0.0174	0.9267	1.3909	- 0.0027	500	500
LMI	a(410)	0.4172	0.7604	0.9178	+0.0957	500	478
	a(490)	0.1727	0.7128	0.6979	+0.0551	500	478
	a(555)	0.0840	0.6539	0.5208	+0.0515	500	478
	b_b(410)	0.0198	0.6893	1.1370	- 0.0049	500	478
	b_b(490)	0.0177	0.6930	1.0189	- 0.0023	500	478
	b_b(555)	0.0168	0.6941	0.9329	- 0.0014	500	478

Table 2.6. RMSE and regression results between the derived and the known IOPs for the IOCCG simulated data set with a solar zenith angle of 30^0; chlorophyll concentration ranges from 0.03 to 0.3 mg m^{-3}

Algorithm	IOPs	RMSE	R^2	Slope	Intercept	N	n
ILMI	a(410)	0.0077	0.9876	1.0278	+0.0052	175	175
	a(440)	0.0050	0.9893	0.9624	+0.0060	175	175
	a(490)	0.0053	0.9761	0.9982	+0.0050	175	175
	a(510)	0.0059	0.9551	1.0309	+0.0041	175	175
	a(555)	0.0027	0.9701	0.9313	+0.0071	175	175
	b_b(410)	0.0010	0.9472	1.0489	+0.0005	175	175
	b_b(440)	0.0008	0.9633	1.0508	+0.0004	175	175
	b_b(490)	0.0006	0.9803	1.0477	+0.0003	175	175
	b_b(510)	0.0005	0.9848	1.0377	+0.0003	175	175
	b_b(555)	0.0003	0.9902	1.0215	+0.0002	175	175
QAA	a(410)	0.0038	0.9933	1.0019	- 0.0028	175	175
	a(440)	0.0033	0.9926	1.0125	+0.0021	175	175
	a(490)	0.0016	0.9891	0.9775	+0.0018	175	175
	a(510)	0.0016	0.9757	0.9449	+0.0033	175	175
	a(555)	0.0017	0.9141	0.9324	+0.0035	175	175
	b_b(410)	0.0005	0.9688	0.9980	+0.0004	175	175
	b_b(440)	0.0004	0.9797	1.0019	+0.0003	175	175
	b_b(490)	0.0002	0.9901	0.9897	+0.0002	175	175
	b_b(510)	0.0002	0.9919	0.9800	+0.0001	175	175
	b_b(555)	0.0002	0.9921	0.9644	+0.0001	175	175
LMI	a(410)	0.0079	0.9924	1.0242	+0.0061	175	175
	a(490)	0.0031	0.9875	0.9461	+0.0046	175	175
	a(555)	0.0016	0.9791	0.8272	+0.0127	175	175
	b_b(410)	0.0023	0.9157	0.9987	- 0.0022	175	175
	b_b(490)	0.0013	0.9829	0.9837	- 0.0012	175	175
	b_b(555)	0.0010	0.9945	0.9491	- 0.0008	175	175

6. Conclusions

A GSA method, EFAST, has been applied to a dataset of simulated remote sensing reflectance and two commonly used IOPs inverse algorithms, QAA and LMI. For the sensitivity of remote sensing reflectance, the following conclusion may be drawn: (1) the variation of remote sensing reflectance is mainly regulated by concentrations of constituents, especially mineral concentration; and (2) particulate characteristics, such as

mineral particle size and real part of refractive index, play an important role in controlling the variation on remote sensing reflectance, however the influence of these characteristics can be avoided through band ratios.

Sensitivity analysis was also applied to QAA and LMI. Sensitivity analysis results demonstrated that the power-low slope of particulate backscattering coefficient plays an important role in governing the performances of QAA and LMI, and that all the considered parameters are important to LMI when the concentrations of constituents are high due to highly interact between factors.

Based on the sensitivity analysis results of remote sensing reflectance and QAA and LMI, an improved IOPs inverse algorithm was proposed. According to incorporate a near infra-red band of 710 nm and, and QAA S and η models, the ILMI model has some improvement when Chl is high, but the three algorithms are competitive for the case of low Chl .

Appendix A: Data simulation

The particulate average absorption, $\overline{Q}_a(\lambda)$, and scattering, $\overline{Q}_b(\lambda)$, efficiencies are calculated through (Bohren and Huffman 1983; Bricaud and Morel 1986):

$$\overline{Q}_a(\lambda) = \frac{\int_{D_{\min}}^{D_{\max}} N(D) D^2 Q_a(D,\lambda) dD}{\int_{D_{\min}}^{D_{\max}} N(D) D^2 dD} \tag{A1}$$

$$\overline{Q}_b(\lambda) = \frac{\int_{D_{\min}}^{D_{\max}} N(D) D^2 Q_b(D,\lambda) dD}{\int_{D_{\min}}^{D_{\max}} N(D) D^2 dD} \tag{A2}$$

If particulate density ρ is known, mass-specific coefficients for absorption, $a^*(\lambda)$, and scattering, $b^*(\lambda)$, can be calculated from the average efficiencies:

$$a^*(\lambda) = \frac{3\bar{Q}_a(\lambda)}{2\rho} \frac{\int_{D_{\min}}^{D_{\max}} N(D)D^2 dD}{\int_{D_{\min}}^{D_{\max}} N(D)D^3 dD} \tag{A3}$$

$$b^*(\lambda) = \frac{3\bar{Q}_b(\lambda)}{2\rho} \frac{\int_{D_{\min}}^{D_{\max}} N(D)D^2 dD}{\int_{D_{\min}}^{D_{\max}} N(D)D^3 dD} \tag{A4}$$

For NAP, the mass-specific coefficients were defined as the absorption or scattering coefficients per unit mass concentration of NAP. For phytoplankton, the mass-specific coefficients were defined as the absorption or scattering coefficients per unit mass concentration of chlorophyll. $N(D)$ is defined as (Bader 1970; Boss *et al.* 2001; Boss *et al.* 2001; Babin *et al.* 2003):

$$N(D) = KD^{-\xi} \tag{A5}$$

K can be any value, but only ξ matters in the calculations of $\bar{Q}_a(\lambda)$ and $\bar{Q}_b(\lambda)$. Thus K was set to be 1 in this study.

Total absorption and scattering coefficients are defined as:

$$a(\lambda) = a_w(\lambda) + a^*_{ph}(\lambda)Chl + a_{CDOM}(\lambda) + a^*_{NAP}(\lambda)C \tag{A6}$$

$$b(\lambda) = b_w(\lambda) + b^*_{ph}(\lambda)Chl + b^*_{NAP}(\lambda)C \tag{A7}$$

Particulate beam attenuation spectrum $c_p(\lambda)$ is modeled as (Voss 1992; Barnard *et al.* 1998; Roesler and Boss 2003):

$$c_p(\lambda) = c_p(550)\left(\frac{550}{\lambda}\right)^{\gamma} \tag{A8}$$

γ can be linked to ξ through the following equation (Boss *et al.* 2001; Boss *et al.* 2001; Twardowski *et al.* 2001):

$$\gamma = \xi - 3 \tag{A9}$$

For phytoplankton, $a_{ph}(\lambda)$ is calculated through multiplying the absorption coefficient at a specific wavelength (440 nm) by normalized absorption spectrum $a_{ph}^{N}(\lambda)$:

$$a_{ph}(\lambda) = a_{ph}(440)a_{ph}^{N}(\lambda) \tag{A10}$$

Based on absorption and attenuation spectra, $b_{ph}(\lambda)$ can be calculated:

$$b(\lambda) = c(\lambda) - a(\lambda) \tag{A11}$$

For NAP, $a_{NAP}(\lambda)$ is calculated through the following model:

$$a_{NAP}(\lambda) = a_{NAP}(440)\exp(-S_{NAP}(\lambda - 440)) \tag{A12}$$

Similarly, $b_{NAP}(\lambda)$ can be calculated with equation (A11).

For CDOM, $a_{CDOM}(\lambda)$ is modeled as:

$$a_{CDOM}(\lambda) = a_{CDOM}(440)\exp(-S_{CDOM}(\lambda - 440)) \tag{A13}$$

Based the calculated IOPs, together with ambient factors, a dataset of $Rrs(\lambda)$ could be simulated through Hydrolight (Mobley 1994).

Appendix B: QAA algorithm

QAA is an algorithm for deriving IOPs from $Rrs(\lambda)$ (Lee *et al.* 2002). Its main idea is first to empirically derive the absorption at a reference wavelength, which is selected according to the IOPs of water. The backscattering coefficient at the reference wavelength is estimated through the reflectance model (equation (1)). Then, backscattering coefficients of other wavelengths is determined through a particulate backscattering model. Finally, absorption coefficients spectrum is derived from reflectance model. Refer to table 2.7 for equations.

Table 2.7. The procedure of QAA

Step	Variable	Formula
1	$rrs(\lambda)$	$rrs(\lambda) = R_{rs}(\lambda) / (0.52 + 1.7R_{rs}(\lambda))$
	$\mu(\lambda) = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}$	$rrs(\lambda) = (g_0 + g_1\mu(\lambda))\mu(\lambda)$
		$\mu(\lambda) = \frac{-g_0 + \sqrt{(g_0)^2 + 4g_1 * rrs(\lambda)}}{2g_1}; g_0 = 0.089, g_1 = 0.125$
2	$a(\lambda_0)$	$\chi = \log\left(\frac{rrs(443) + rrs(490)}{rrs(\lambda_0) + 5\frac{rrs(667)}{rrs(490)}rrs(667)}\right)$
		$a(\lambda_0) = a_w(\lambda_0) + 10^{-1.146 - 1.366\chi - 0.469\chi^2}$
	$b_{bp}(\lambda_0)$	$b_{bp}(\lambda_0) = \frac{\mu(\lambda_0)a(\lambda_0)}{1 - \mu(\lambda_0)} - b_{bw}(\lambda_0)$
3	$b_{bp}(\lambda)$	$b_{bp}(\lambda) = b_{bp}(\lambda_0)\left(\frac{\lambda_0}{\lambda}\right)^{\eta}$
		$\eta = 2.0\left(1 - 1.2\exp\left(-0.9\frac{rrs(443)}{rrs(\lambda_0)}\right)\right)$
	$a(\lambda)$	$a(\lambda) = \frac{(1 - u(\lambda))(b_{bw}(\lambda) + b_{bp}(\lambda))}{u(\lambda)}$
4	$a_{CDM}(\lambda)$	$a_{CDM}(\lambda) = a_{CDM}(443)e^{-S(\lambda - 443)}$
	$a_{ph}(\lambda)$	$a_{ph}(\lambda) = a(\lambda) - a_{CDM}(\lambda) - a_w(\lambda)$
		$a_{CDM}(443) = \frac{(a(411) - \zeta a(443)) - (a_w(411) - \zeta a_w(443))}{\xi - \zeta}$
		$\zeta = 0.74 + \frac{0.06}{0.8 + rrs(443) / rrs(\lambda_0)}$
		$\xi = \exp(S(443 - 411))$
		$S = 0.015 + \frac{0.002}{0.6 + rrs(443) / rrs(\lambda_0)}$

Appendix C: LMI algorithm

LMI uses three bands (410, 490, and 555 nm) to derive the IOPs by solving a linear system:

$$\begin{bmatrix} 1 & 1 & \upsilon(\lambda_r) \\ a_{ph}^{\diamond}(\lambda_2) & a_{CDM}^{\diamond}(\lambda_2) & b_{bp}^{\diamond}(\lambda_2)\upsilon(\lambda_2) \\ a_{ph}^{\diamond}(\lambda_3) & a_{CDM}^{\diamond}(\lambda_3) & b_{bp}^{\diamond}(\lambda_3)\upsilon(\lambda_3) \end{bmatrix} \begin{bmatrix} a_{ph}(\lambda_r) \\ a_{CDM}(\lambda_r) \\ b_{bp}(\lambda_r) \end{bmatrix} = -\begin{bmatrix} a_w(\lambda_r) + b_{bw}(\lambda_r)\upsilon(\lambda_r) \\ a_w(\lambda_2) + b_{bw}(\lambda_2)\upsilon(\lambda_2) \\ a_w(\lambda_3) + b_{bw}(\lambda_3)\upsilon(\lambda_3) \end{bmatrix} \quad \text{(A14)}$$

Band 410 nm was set as the reference wavelength λ_r. $\upsilon(\lambda) \equiv 1 - \frac{1}{u}$, and $u = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}$ was solved from reflectance model. The absorption spectrum of phytoplankton, $a_{ph}^{\diamond}(\lambda)$, is defined as:

$$a_{ph}^{\diamond}(\lambda) = \exp\left(\frac{\lambda_r^2 + 886(\lambda - \lambda_r) - \lambda^2}{2\delta^2} \right) \quad \text{(A15)}$$

where δ is FWHM. The spectral model of CDM is defined as:

$$a_{CDM}^{\diamond}(\lambda) = \exp(-S(\lambda - \lambda_r)) \quad \text{(A16)}$$

The spectral model of total particulate backscattering coefficient is modeled as:

$$b_{bp}^{\diamond}(\lambda) = \left(\frac{\lambda_r}{\lambda} \right)^{\eta} \quad \text{(A17)}$$

III. ESTIMATION OF PHYTOPLANKTON SIZE COMPOSTION FROM REMOTE SENSING DATA

ABSTRACT

Phytoplankton size classes (pico-plankton, nano-plankton, and micro-plankton) provide information about pelagic ocean ecosystem structure, and their spatiotemporal variation plays a key role in understanding ocean ecosystem structure and global carbon cycling. In this paper, an algorithm based on spectral features was proposed to estimate phytoplankton size fractions from remote sensing data. First, five types of spectral features (original and normalized remote sensing reflectance, two-band ratios, continuum removed spectra, and spectral curvatures) were extracted from remote sensing reflectance. The spectral features sensitive to the phytoplankton size fractions were then determined through a band selection method. Finally, phytoplankton size fractions were regressed against the sensitive spectral features. The performance of the developed algorithm was validated with the simulated and the satellite Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer (MODIS) datasets, indicating that the algorithm is effective in estimating phytoplankton size fractions (R^2 is 0.9273, 0.9305 (f_m) for simulated data; R^2 is 0.6532 (f_m), 0.5201 (f_n), 0.5834 (f_p) for SeaWiFS; and R^2 is 0.8022 (f_m), 0.6726 (f_n), 0.7633 (f_p) for MODIS), and that the spectral features sensitive to phytoplankton size fractions mainly consist of continuum removed spectra as well as spectral curvatures and their wavelengths mainly center on the pigments-absorption domains. Global spatial distributions of pico-plankton, nano-plankton, and micro-plankton were also estimated from monthly SeaWiFS and MODIS images (June 2006). The patterns are consistent with current understanding that pico-

plankton accounts for large proportion of total phytoplankton in oligotrophic areas, that nano-plankton dominates in eutrophic and transitional areas, and that miro-plankton is dominant in high-productivity areas.

Keywords: phytoplankton size fraction, spectral feature, remote sensing, SeaWiFS, MODIS

1. Introduction

Oceans play an important role in regulating atmospheric CO_2 concentration (Falkowski *et al.* 2000). Each day, more than a hundred million tons of CO_2 are transformed into organic materials by phytoplankton through primary production (Behrenfeld *et al.* 2006), and a large amount of CO_2 from the oceans is also released to the atmosphere due to remineralization through biota respiration (del Giorgio and Duarte 2002). For the first process, the downward export of biogenic particles, also known as biological pump, is responsible for transporting part of the produced organic materials into ocean interiors, decreasing the CO_2 concentration in the atmosphere, and as a result may influence climate (Falkowski *et al.* 1998; Boyd *et al.* 2000; Behrenfeld *et al.* 2006; Cassar *et al.* 2007; Kostadinov *et al.* 2010). This export of organic matter from the surface waters is regulated by various factors, including phytoplankton cell size (Buesseler *et al.* 2007; Burd and Jackson 2009; Guidi *et al.* 2009; Stemmann and Boss 2012), oceanic ecosystem structure (Boyd and Trull 2007), and mineral ballast (Francois *et al.* 2002; Boyd and Trull 2007; Lam *et al.* 2011). Particularly, phytoplankton cell size is an ecophysiological trait connecting to environmental conditions (Irwin *et al.* 2006; Finkel *et al.* 2010), and an important factor controlling particle aggregation and food web

(Burd and Jackson 2009), which in turn regulate surface exports. Therefore, knowledge on the spatiotemporal distribution and variation of phytoplankton size fraction (PSF) (see table 3.1 for symbol/abbreviation) is crucial for understanding the impact of climate change on the biogeochemical cycling of elements, and the oceanic biogeochemical feedbacks in influencing climate in the future (Le Quere *et al.* 2005; Boyd and Trull 2007; Finkel *et al.* 2010).

Table 3.1. Symbols and descriptions

Symbol/abbreviation	Description	Unit
λ	Wavelength	nm
$R_{rs}(\lambda)$	Remote sensing reflectance above water surface	sr^{-1}
$N\_R_{rs}(\lambda)$	$R_{rs}(\lambda)$ normalized with its integration	-
$CR(\lambda_1,\lambda_2,\lambda_3)$	Continuum removed spectrum at λ_2 with beginning and ending wavelengths of λ_1 and λ_3, respectively	-
$CV(\lambda_1,\lambda_2,\lambda_3)$	Curvature at λ_2 with beginning and ending wavelengths of λ_1 and λ_3, respectively	-
$BR(\lambda_1,\lambda_2)$	Band ratio ($R_{rs}(\lambda_2)$/ $R_{rs}(\lambda_1)$)	-
PSF	Phytoplankton size fraction	-
CDOM	Colored dissolved organic matter	-
NAP	Non-algal particles	-
IOPs	Inherent optical properties	-
PSD	Particle size distribution	-
f_m	Micro-plankton fraction	-
f_n	Nano-plankton fraction	-
f_p	Pico-plankton fraction	-
[Chl]	Chlorophyll concentration	$mg\ m^{-3}$
$a(\lambda)$	Total absorption coefficients of water column	m^{-1}
$c_{ph}(\lambda)$	Beam attenuation coefficient of phytoplankton	m^{-1}
$a_x(\lambda)$	Absorption coefficients of *x*. *x* could be *w*: water; *ph*: phytoplankton; *g*: colored dissolved organic matter; *dm*: colored detritus matter; pico: pico-plankton; micro: micro-plankton	m^{-1}
$a_{ph}^{N440}(\lambda)$	$a_{ph}(\lambda)$ normalized with respect to 440nm	-
$b_b(\lambda)$	Total backscattering coefficients of water column	m^{-1}

$b_{bx}(\lambda)$	Backscattering coefficients of *x*. *x* could be *w*: water; *ph*: phytoplankton; *dm*: colored detritus matter	m^{-1}
$\tilde{b}_x$	Backscattering probability of of *x*. *x* could be *ph*: phytoplankton; *dm*: detritus/minerals	-
S_g	Exponential slope of CDOM absorption spectrum	nm^{-1}
S_{dm}	Exponential slope of detritus/minerals absorption spectrum	nm^{-1}

To achieve a better understanding of ocean biogeochemistry, phytoplankton is partitioned into groups according to their biogeochemical functions (e.g. calcifiers, silicifiers, nitrogen-fixers, and dimethyl sulphide producers) (Forget *et al.* 2008). The same purpose can also be achieved through partition of phytoplankton according to their cell sizes. Some specific accessory pigments are unique to individual phytoplankton taxa or groups, and thus always used as biomarkers of phytoplankton functional groups. Furthermore, as each functional group is roughly related to a size range, phytoplankton cell size can be estimated with some accessory pigments with some cautions (Vidussi *et al.* 2001; Uitz *et al.* 2006; Devred *et al.* 2011). In this way, phytoplankton is usually separated into pico-plankton (0.2-2 μm), nano-plankton (2-20 μm), and micro-plankton (20-200 μm) (Sieburth *et al.* 1978). High-performance liquid chromatography (HPLC) provides a suit of pigment concentrations which allows for calculating fractions of these phytoplankton size groups with respect to chlorophyll concentration (Vidussi *et al.* 2001; Uitz *et al.* 2006; Devred *et al.* 2011). However, this method is impractical to acquire high spatiotemporal resolution distribution of phytoplankton cell size on global scale.

Remote sensing has been proved to be an economic and efficient method for estimating the spatiotemporal distribution of phytoplankton size/species composition on global scale. Remote sensing methods for estimating phytoplankton size can be divided into two major categories. One is based on the characteristics of inherent optical

properties (IOPs) of phytoplankton, and another is based on the pigment concentration [Chl]. The IOPs-based approaches are based on observations and Mie theory modeling results, i.e. IOPs (including spectral shape and magnitude) vary with phytoplankton cell size (Morel and Bricaud 1981; Bricaud and Morel 1986; Stramski and Kiefer 1991; Boss *et al.* 2001; Stramski *et al.* 2001; Bricaud *et al.* 2004; Loisel *et al.* 2006; Zhou and Cao 2008). These approaches are further subdivided into three groups using the: particle size distribution (PSD) slope, unmixing, and "anomalies" methods. As the PSD of marine particles commonly follows the Junge distribution (Sheldon *et al.* 1972), the slope of Junge distribution is linked to that of total backscattering or beam attenuation coefficients, which can be further derived through Mie theory or an empirical polynomial, and finally the PSF can be modeled as a function of the slope of PSD (Boss *et al.* 2001; Twardowski *et al.* 2001; Kostadinov *et al.* 2009; Kostadinov *et al.* 2010). The unmixing approach assumes that absorption spectrum of phytoplankton ($a_{ph}(\lambda)$) can be linearly reconstructed by those of pico-plankton and micro-plankton (Ciotti *et al.* 2002), as shown in equation (A4). Based on this assumption, PSF can be directly derived from remote sensing reflectance ($Rrs(\lambda)$) or $a_{ph}(\lambda)$ (Ciotti and Bricaud 2006; Bricaud *et al.* 2012). Alternatively, with a look-up table based on $a_{ph}(\lambda)$ modeled with equation (A4), the PSF can be derived with satellite-observed $Rrs(\lambda)$ (Mouw and Yoder 2010). As phytoplankton cell size is a second-order factor driving the variation of $Rrs(\lambda)$, the third approach is to remove the spectral contribution of [Chl] by normalizing (dividing or subtracting) $Rrs(\lambda)$ with the remote sensing reflectance with corresponding [Chl], and then estimate the PSF from the normalized spectra (Moulin *et al.* 2005; Alvain *et al.* 2008; Brown *et al.* 2008).

The [Chl]-based approaches directly model PSF as a function of [Chl] based on the observation that large size phytoplankton dominates in eutrophic waters, and vice versa. Two categories of models have been adopted to model PSF based on [Chl]: two- and three-component models. The two-component model assumes that chlorophyll concentration contributed by the pico-plankton and micro-plankton is an exponential function of [Chl]. Through this model, the fractions of pico-plankton and micro-plankton were estimated from retrieved [Chl] (Devred *et al.* 2006). The three-component model assumes that the pico-plankton calculated from the two-component model can be further divided into pico-plankton and nano-plankton. Therefore, the fraction of nano-plankton could also be estimated from the estimated [Chl] (Brewin *et al.* 2011; Devred *et al.* 2011).

Both IOPs- and [Chl]-based approaches have disadvantages. While phytoplankton cell size results in second-order variability in $Rrs(\lambda)$, it is not easy to accurately estimate PSF or phytoplankton functional groups based on spectral shape information, especially through multiple spectral remote sensing data (Brewin *et al.* 2011). [Chl]-based approaches also fail to work well in some regions as [Chl] is not necessary to co-vary with phytoplankton size or functional types. Within the same phytoplankton group, changes in in pigment concentration may be due to photoacclimation, which generally reflects changes in environmental factors (Eppley 1972; Bricaud *et al.* 2004). To solve these problems, PSF has been tentatively derived from phytoplankton absorption spectra using machine learning methods, such as multilayer perceptions (Bricaud *et al.* 2007).

In this paper, a new algorithm is proposed for deriving PSF from remote sensing data. This algorithm is built upon five spectral features: original and normalized $Rrs(\lambda)$ (Lubac and Loisel 2007; Le *et al.* 2011), band ratio, spectral curvature (Campbell and Esaias

1982; Hoge and Swift 1986; Lee and Carder 2000), and continuum removed spectra (Huang *et al.* 2004), and consists of a band selection algorithm, support vector machine recursive feature elimination (SVM-RFE) (Guyon *et al.* 2002), and support vector machine (SVM) (Vapnik 1995) regression. With simulated and satellite datasets, SVM-RFE is used to determine the spectral features sensitive to PSF, and SVM regression to estimate PSF from the selected spectral features. Additionally, the model effectiveness for deriving PSF distribution is respectively examined with global monthly Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer (MODIS) images.

2. Datasets

The datasets consist of simulated and satellite-acquired image spectra as well as the corresponding PSF. The use of simulated datasets is to examine the performance of the developed algorithm with excluding the impacts of atmospheric correction for satellite-acquired image spectra, uncertainty in measured PSF, and time difference between satellite-acquired spectra and measured PSF.

2.1 Simulated datasets

A dataset with SeaWiFS (412, 443, 490, 510, 555 and 670 nm) spectral resolution was simulated using Hydrolight (Mobley 1994). The dataset has 1000 samples, and each sample includes $Rrs(\lambda)$, and the corresponding pico-plankton (f_p) and micro-plankton (f_m) fractions. For the simulated dataset, [Chl] was divided in to three intervals: [0.03, 0.3], [0.3, 3], and [3, 10]. Within each interval, the same number of samples was randomly generated. $a_{ph}(\lambda)$ was linearly mixed with pico-plankton and micro-plankton "basic vectors" (Ciotti *et al.* 2002), as shown in equations A3, A4 and A5. As the functional

relationship between [Chl] and f_m shows a sigmoid shape (Devred *et al.* 2006; Mouw and Yoder 2010), f_m (f_p=1-f_m) was randomly generated between the upper (f_m=1-*exp*(-15[Chl])) and lower (f_m=1-*exp*(-0.12[Chl])) boundaries for a specified [Chl] (Mouw and Yoder 2010). In this way, the correlation between PSF and [Chl] was modeled, and PSF was incorporated into $a_{ph}(\lambda)$ and finally into *Rrs*(λ). With the constructed IOPs as inputs, a dataset of *Rrs*(λ) was simulated using Hydrolight. See the appendix for more information on data simulation.

A dataset with MODIS (412, 443, 469, 488, 531, 547, 555, 645, 667 and 678 nm) spectral resolution was also simulated using Hydrolight (Mobley 1994). It also has 1000 samples, and each sample includes *Rrs*(λ) and the corresponding f_p and f_m.

2.2 Satellite datasets

SeaWiFS images were used because they have high spatiotemporal resolution and cover oceans of the world. Daily and monthly SeaWiFS *Rrs*(λ) images from September 1997 to December 2007 were downloaded from the NASA website (http://oceancolor.gsfc.nasa.gov/). The image data have a spatial resolution of 9×9 km. Daily images (level 3 mapped data) were used to extract the pixels matching in-situ observations which will be discussed in section 2.3. Monthly global area coverage (GAC, level 3 mapped data) images were used to derive global PSF distribution.

MODIS images were also used. The use of MODIS images is because MODIS images have more bands than SeaWiFS images and allow for comparing their mapping results. Daily and monthly MODIS *Rrs*(λ) images from July 2002 to December 2011 were downloaded from the NASA website (http://oceancolor.gsfc.nasa.gov/). The image data has a spatial resolution of 9×9 km, and includes 10 bands. Bands 469 and 555 nm

were derived from the corresponding bands with spatial resolution of 500×500 m, and band 645 nm from the corresponding band (250×250 m). These three bands were used together with seven MODIS-aqua bands to test whether PSF estimation could be improved with higher spectral resolution or not. Daily images (level 3 mapped data) are for extracting the pixels matching in-situ observations, while monthly GAC images (level 3 mapped data) are for deriving global PSF distribution.

2.3 Phytoplankton size fractions estimated from in-situ pigments

Several field-observed pigment datasets were employed to calculate PSF. These datasets including both case 1 (phytoplankton and associated material dominated) and case 2 (colored dissolved organic matter (CDOM) and non-algal particles (NAP) dominated) waters were downloaded from different websites (table 3.2). The geographic distribution of these extracted samples is shown in Fig. 3.1. Each sample contains seven pigments: fucoxanthin (Fuco), peridinin (Perid), 19'-hexanoyloxyfucoxanthin (Hex_fuco), 19'-butanoyloxyfucoxanthin (But_fuco), alloxanthin (Allo), chlorophyll *b* and divinyl chlorophyll *b* (TChl*b*), and zeaxanthin (Zea). With these pigments, the relative biomass proportions of micro-plankton, nano-plankton, and pico-plankton were calculated with equations (1)-(4) (Vidussi *et al.* 2001):

$$\sum DP_w = 1.41[Fuco] + 1.41[Perid] + 1.27[Hex\_fuco] + 0.35[But\_fuco] + 0.60[Allo] + 1.01[TChlb] + 0.86[Zea] \quad (1)$$

$$f_m = (1.41[Fuco] + 1.41[Perid]) / \sum DP_w \quad (2)$$

$$f_n = (1.27[Hex\_fuco] + 0.35[But\_fuco] + 0.60[Allo]) / \sum DP_w \quad (3)$$

$$f_p = (1.01[TChlb] + 0.86[Zea]) / \sum DP_w \quad (4)$$

where $\sum DP_w$ represents the [Chl] reconstructed from the seven accessory pigments.

Table 3.2. Used HPLC data sources

Project and Cruise	References	Year	Location	Source
Gep&COA-L	(Dandonneau *et al.* 2004)	1999-2002	Trans N Atlantic-S Pacific	http://www.lodyc.jussieu.fr/gepco
MOCE	--	1997-2002	N Pacific	http://www.seabass.gsfc.nasa.gov
NOMAD	(Werdell and Bailey 2005)	1997-2007	global	http://www.seabass.gsfc.nasa.gov
SeaBASS general search	(Werdell and Bailey 2005)	1997-2007	global	http://www.seabass.gsfc.nasa.gov

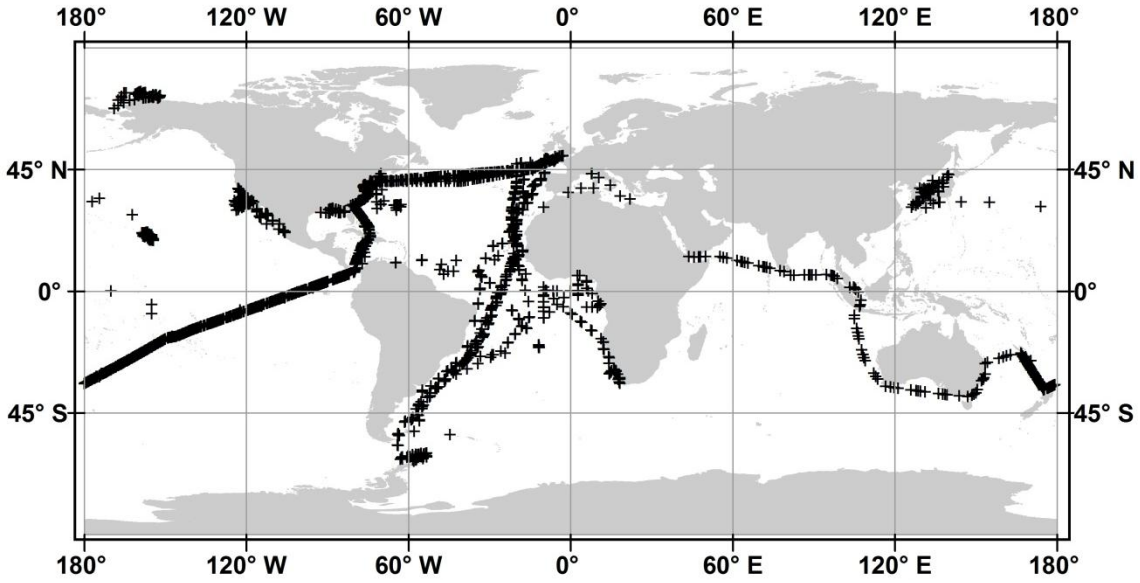


Fig. 3.1. Geographic distribution of extracted HPLC pigment observations.

The quality assurance procedure described in (Uitz *et al.* 2006) was carried out to assure that the field samples collected by different teams, and analyzed in different laboratories using a variety of HPLC instruments and protocols can be combined coherently. First, samples with incomplete biomarker pigments were rejected. Second, samples with [Chl] lower than 0.001 mg m^{-3} were removed, because values in this range

are below the sensitivity of HPLC. Third, accessory pigments concentrations below 0.001 mg m^{-3} were reset to zero. Fourth, samples collected within the top 5 m of a water column were retained, considering that signals acquired by remote sensing sensors are mainly from the first optical depth, where 37% irradiance at water surface remains. Finally, as [Chl] commonly co-varies in a quasi-linear manner with the sum of major accessory pigments (Trees *et al.* 2000), data that exceeded three standard deviations for the sum of the accessory pigments were identified and removed (Mouw and Yoder 2010). The removal of poor quality samples led to 616 and 484 matched samples for the SeaWiFS and MODIS satellite datasets, respectively.

3. Methods

The algorithm consists of three major steps: (1) constructing spectral features from $Rrs(\lambda)$, (2) determining the spectral features sensitive to PSF, and (3) estimating PSF with the selected features. The flow chart for this algorithm is shown in Fig. 3.2.

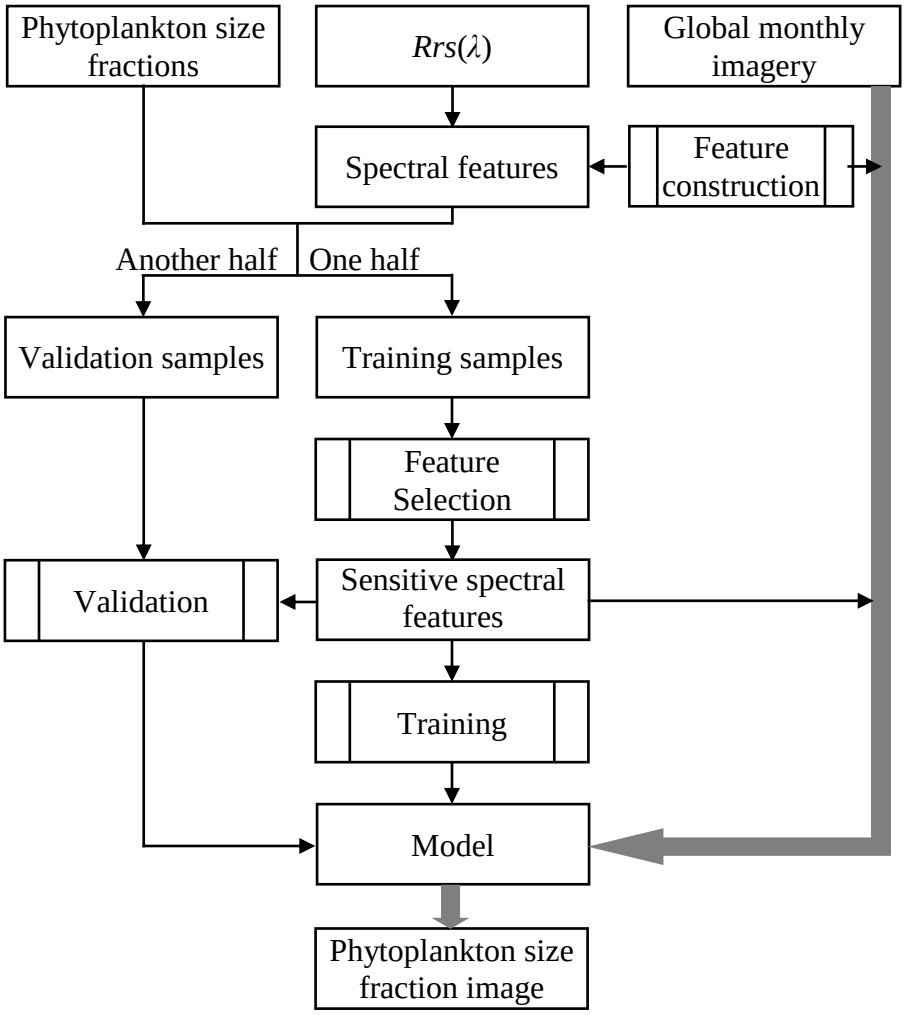


Fig. 3.2. The flow chart of the developed algorithm; rectangle represents data, and rectangle with two vertical lines represents function; narrow arrow represents data flow direction for training and validation; wide arrow represents data flow of estimating PSF from global monthly images.

3.1 Spectral feature construction

3.1.1 Remote sensing reflectance

Rrs(λ) contains both first-order variations related to concentrations of constituents and second-order variations tying to physical parameters, such as particle size. This justifies

the use of original $Rrs(\lambda)$ from SeaWiFS and MODIS in the proposed algorithm. SeaWiFS and MODIS satellite datasets contain 6 and 10 bands of $Rrs(\lambda)$, respectively.

3.1.2 Normalized remote sensing reflectance

Normalized remote sensing reflectance ($N\_Rrs(\lambda)$) is for removing magnitude difference (Lubac and Loisel 2007). This normalization process was implemented by dividing spectra with the reflectance integral calculated over the whole spectral range covered by a specified sensor (i.e. SeaWiFS and MODIS). $N\_Rrs(\lambda)$ was demonstrated to be effective in clustering spectra controlled by various factors (Lubac and Loisel 2007). Normalization yielded 6 and 10 $N\_Rrs(\lambda)$ features for the datasets with SeaWiFS and MODIS bands, respectively.

3.1.3 Band ratio

Band ratio is defined as:

$$BR(\lambda_1, \lambda_2) = Rrs(\lambda_2) / Rrs(\lambda_1), \quad \lambda_2 > \lambda_1 \tag{5}$$

As $Rrs(\lambda)$ is controlled by various factors (e.g. concentrations of constituents, wind speed, and solar zenith angle), band ratio reduces the influence of uninteresting factors, such as wind speed and solar zenith angle, and increases the correlation between a band ratio and the biomass or IOPs parameters. For each sample, 15 and 45 $BR(\lambda_1,\lambda_2)$ features were obtained for the datasets with SeaWiFS and the MODIS bands, respectively.

3.1.4 Spectral curvature

Spectral curvature is defined as (Campbell and Esaias 1982; Hoge and Swift 1986; Lee and Carder 2000):

$$CV(\lambda_1, \lambda_2, \lambda_3) = \frac{Rrs(\lambda_2)^2}{Rrs(\lambda_1) \times Rrs(\lambda_3)}, \ \lambda_3 > \lambda_2 > \lambda_1 \tag{6}$$

The logarithmic transformation of this feature is equivalent to applying difference to log(*Rrs*(λ)) twice. Equally, spectral curvature is a measure of the nonlinearity or curvature of *Rrs*(λ) in the logarithmic space. As a second-order difference operator, this feature has the merits of: (1) eliminating the linear variation of *Rrs*(λ); and (2) reducing the second-order variation to a constant. Consequently, spectral curvature is not sensitive to factors such as solar zenith angle and wind speed. For each sample, 20 and 120 *CV*(λ_1,λ_2,λ_3) features were obtained for the datasets with SeaWiFS and the MODIS bands, respectively.

3.1.5 Continuum removed spectra

Continuum removal is a procedure for isolating absorption features from a spectrum (Kokaly and Clark 1999; Huang *et al.* 2004). A continuum removed spectrum is defined as:

$$CR(\lambda_1,\lambda_2,\lambda_3)=\frac{Rrs(\lambda_2)}{R_c(\lambda_1,\lambda_2,\lambda_3)},\quad \lambda_3>\lambda_2>\lambda_1 \tag{7}$$

where $R_c(\lambda_1,\lambda_2,\lambda_3)$ is the minimal convex "hull" of straight-line segments fitted over the top of a spectrum, as shown in Fig. 3.3(A). Fig. 3.3(B) shows a continuum removed spectrum of *Rrs*(λ). Apparently, a continuum removal enhances depth of individual absorption feature. Additionally, continuum removal also serves to remove the influences of ambient factors (i.e. solar zenith angle and wind speed) (Yan *et al.* 2010). For each sample, 20 and 120 *CR*(λ_1,λ_2,λ_3) features were obtained for the datasets with SeaWiFS and MODIS bands, respectively.

Combining all the *Rrs*(λ), *N_Rrs*(λ), *BR*(λ_1,λ_2), *CR*(λ_1,λ_2,λ_3) and *CV*(λ_1,λ_2,λ_3) yields 67 features for the datasets with SeaWiFS bands, and 305 features for the datasets with MODIS bands.

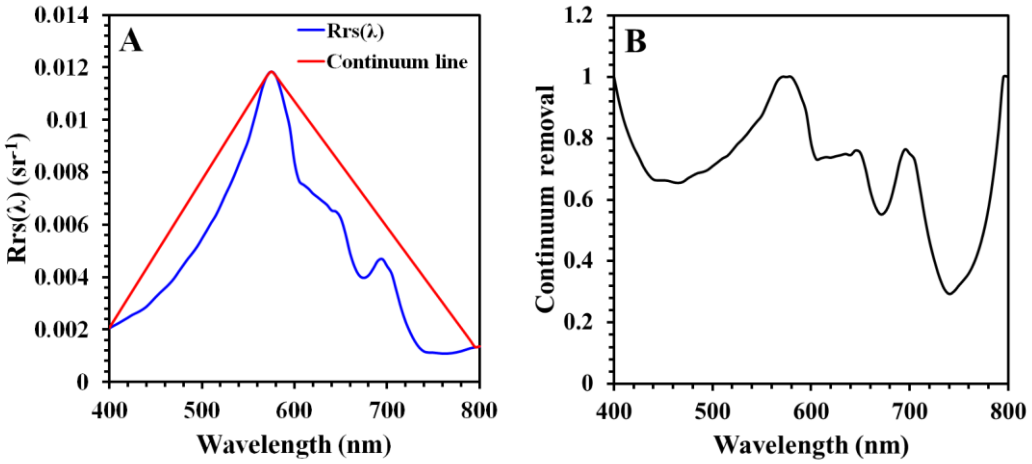


Fig. 3.3. *Rrs*(λ) and its continuum removed spectrum; (A) *Rrs*(λ) and its continuum line; (B) continuum removed spectrum of *Rrs*(λ).

3.2 Support vector machine

SVM is a method to locating an optimal discriminative boundary between binary classification problems. The use of kernel function enables SVM to deal with sophisticated classification problems with a low computational demand. SVM can also be extended to handle multiple classification problems by applying multiple SVMs simultaneously through a one-against-one or one-against-all schema. Additionally, SVM can also be used to regression problems. Detailed descriptions and explanations of SVM can be found in (Vapnik 1995). In this study, SVM is for feature selection and regression of PSF against spectral features. LIBSVM was used to implement SVM whose kernel function was selected as radial basis function (RBF).

3.3 Feature selection

Support vector machine recursive feature elimination (SVM-RFE) (Guyon *et al.* 2002) is a feature selection algorithm (Bazi and Melgani 2006; Archibald and Fann 2007;

Zhang and Ma 2009). SVM-RFE uses weight value ($W^2(\alpha^*)$) as the ranking criterion. The weight value is defined as:

$$W_{(-p)}^2(\alpha^*) = \sum_{i,j=1}^{N} \alpha_i^* \alpha_j^* y_i y_j K(x_i^{(-p)} x_j^{(-p)}) \quad (8)$$

where $x_i^{(-p)}$ stands for *ith* training sample without the *pth* feature; α^* denotes the support vectors, which are also the non-zero Lagrange multiplier in a dual optimization problem; *i* and *j* are training sample indices; *N* is sample number; y_i and y_j are labels of samples *i* and *j*, respectively; and $K(x, y)$ is kernel function for mapping feature space into a high dimensional space. The SVM-RFE workflow can be summarized as follows (Bazi and Melgani 2006):

Step 1) Set the feature subset *S* with the original spectral features;

Step 2) Train an SVM classifier on the feature subspace defined by *S*;

Step 3) For each feature $p \in S$, compute the ranking score

$$c_p = |W^2(\alpha^*) - W_{(-p)}^2(\alpha^*)|;$$

Step 4) Find the feature *q* in *S* with the smallest ranking score;

Step 5) Remove *q* from *S*;

Step 6) Go to Step 2) until the preset number of features is reached.

Discretization is required for PSF because SVM-RFE is based on discrete classes. As PSF is in the range of [0, 1], intervals were created by equally dividing [0, 1]. f_m is projected to a label related to an interval where f_m fall in. With constructed spectral features and discretized PSF, spectral features sensitive to PSF were determined by SVM-RFE.

3.4 Evaluation of results

Each dataset was randomly divided into two subsets with equal sample number. One subset was for selecting spectral features sensitive to PSF and training SVM. Another subset was used to validate the training model. The performance of the developed algorithm was evaluated by the coefficient of determination (R^2), relative error, and root mean square error (RMSE). Relative error and RMSE are defined as:

$$relative\_error = \frac{x_i - \hat{x}_i}{x_i} \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{10}$$

where N is sample number, and x_i and $\hat{x}_i$ are the reference and the estimated values for the *ith* sample, respectively.

4. Results

Validation results and global PSF distribution derived from monthly SeaWiFS and MODIS images were discussed in this study. Even though calibration and validation results could be simultaneously obtained, validation results were preferred because an algorithm with high accuracy in calibration is not necessary robust in performance when applied to a new dataset.

4.1 Results for simulated datasets

4.1.1 Simulated SeaWiFS dataset

Table 3.3. Validation results for the simulated SeaWiFS data set; feature number is the number of the spectral features; RMSE is root-mean-square error; R^2 is the coefficient of determination between the reference and estimated values; slope and intercept are for the regression between the estimated and reference values; and feature number, RMSE, R^2, slope, and intercept are used to represent the same meaning in other tables shown hereafter

Feature number	RMSE	R^2	Slope	Intercept
5	0.188	0.470	0.473	0.318
10	0.080	0.905	0.869	0.072
15	0.075	0.921	0.870	0.064
20	**0.070**	**0.927**	**0.898**	**0.054**
25	0.072	0.925	0.898	0.051
30	0.072	0.924	0.914	0.042
35	0.072	0.924	0.909	0.045
40	0.070	0.934	0.882	0.054
45	0.069	0.936	0.880	0.054
50	0.070	0.934	0.869	0.060
55	0.072	0.932	0.863	0.064
60	0.073	0.934	0.854	0.066
all(67)	0.073	0.932	0.856	0.066

The validation results of f_m derived from different numbers of features are listed in table 3.3. RMSE decreases sharply from 0.188 to 0.080 when the feature number increases from 5 to 10, then goes down slightly from 0.080 to 0.070 when the feature number further increases from 10 to 20, and finally flattens out with the use of additional features. In contrast, R^2 increases dramatically from 0.470 to 0.905 when the feature number increases from 5 to 10, then reaches 0.927 (20 features) which is close to maximum (0.936, 45 spectral features), and finally is not sensitive to additional spectral features. The slope of regression line between the estimated and reference values follows a similar variation pattern to that for R^2, except having a slightly decreasing trend with

the increasing feature number after reaching the maximum (0.914, 30 spectral features). In order to balance various factors, such as model simplicity and computation complexity, f_m derived with the top 20 spectral features is considered to be acceptable. The scatter plot between the reference and estimated f_m (Fig. 3.4(A)) shows a tight correlation and slight overestimation and dispersion for PSF lower than around 0.4. The selected top 20 spectral features (table 3.4) include 11 $CR(\lambda_1,\lambda_2,\lambda_3)$, 5 $CV(\lambda_1,\lambda_2,\lambda_3)$, 3 $Rrs(\lambda)$, and 1 $N\_Rrs(\lambda)$. Apparently, continuum removed spectra and spectral curvatures dominate the top 20 spectral features derived for the following bands: 443 nm (2), 490 nm (8), 510 nm (7), and 555 nm (3). Furthermore, among the top 20 features, spectral features derived from wavelengths of 490 and 510 nm are dominant.

Table 3.4. Top 20 out of 67 spectral features derived from the simulate SeaWiFS data set; $CR(\lambda_1,\lambda_2,\lambda_3)$ represents the continuum removed spectrum for wavelength λ_2, with λ_1 and λ_3 as the starting and the ending wavelengths, respectively; $BR(\lambda_1,\lambda_2)=Rrs(\lambda_2)/Rrs(\lambda_1)$ $(\lambda_2>\lambda_2)$; $CV(\lambda_1,\lambda_2,\lambda_3)$ represents spectral curvature with starting and ending wavelengths of λ_1 and λ_3, respectively; $Rrs(\lambda)$ represents remote sensing reflectance with wavelength λ; and $CR(\lambda_1,\lambda_2,\lambda_3)$, $BR(\lambda_1,\lambda_2)$, $CV(\lambda_1,\lambda_2,\lambda_3)$, and $Rrs(\lambda)$ represent the same meaning for the rest tables

Rank	Feature	Rank	Feature	Rank	Feature
1	CR(510,555,670)	8	CR(412,443,490)	15	N_Rrs(510)
2	CV(412,490,555)	9	CR(490,510,555)	16	CR(443,490,555)
3	Rrs(490)	10	CV(412,490,670)	17	CR(412,510,670)
4	CR(443,490,670)	11	CV(443,510,670)	18	Rrs(443)
5	CR(490,555,670)	12	CR(490,510,670)	19	CR(443,555,670)
6	CV(443,490,510)	13	Rrs(510)	20	CR(443,490,510)
7	CR(412,490,670)	14	CV(412,510,670)		

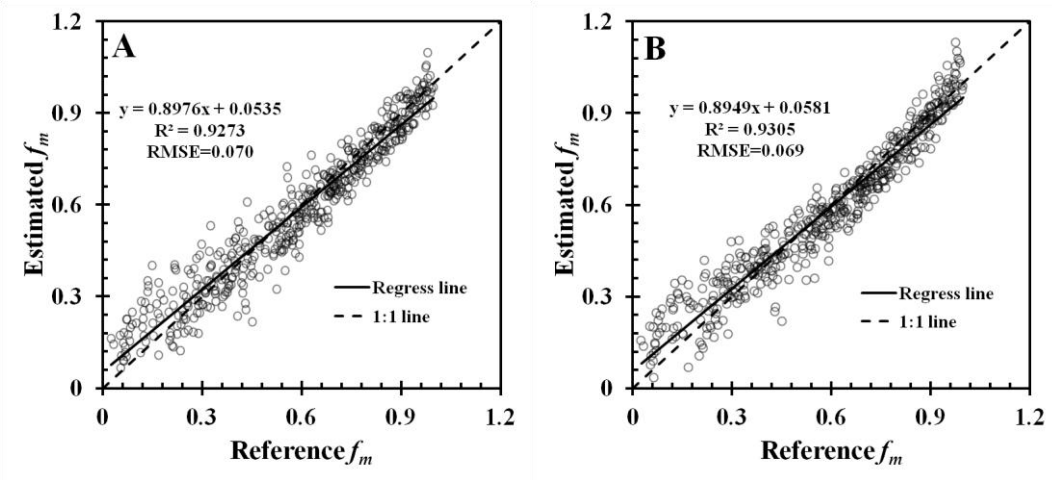


Fig. 3.4. Scatter plots between the reference and the estimated f_m derived from top 20 spectral features of simulated SeaWiFS and MODIS data sets; (A) and (B) are for the SeaWiFS and MODIS data sets, respectively.

4.1.2 Simulated MODIS dataset

Table 3.5. Validation results for the simulated MODIS data set

Feature number	RMSE	R^2	Slope	Intercept
5	0.194	0.437	0.421	0.351
10	0.089	0.882	0.870	0.083
15	0.073	0.924	0.874	0.070
20	**0.069**	**0.931**	**0.895**	**0.059**
25	0.072	0.922	0.913	0.048
30	0.072	0.922	0.915	0.047
50	0.067	0.938	0.897	0.043
100	0.071	0.936	0.865	0.058
150	0.075	0.930	0.848	0.068
200	0.080	0.925	0.816	0.083
250	0.083	0.920	0.800	0.093
all(305)	0.086	0.915	0.788	0.101

The validation results of f_m derived from different numbers of features are listed in table 3.5. The variations of RMSE, R^2, slope, and intercept with feature number behave similarly as those for the simulated SeaWiFS dataset. Again, the top 20 spectral features

(table 3.6) , which consist of 13 $CR(\lambda_1,\lambda_2,\lambda_3)$, 5 $CV(\lambda_1,\lambda_2,\lambda_3)$, and 2 $Rrs(\lambda)$, yielded acceptable f_m estimation (Fig. 3.4(B)). Continuum removed spectra and spectral curvatures still dominate the top 20 features derived for the following bands: 443 nm (1), 469 nm (2), 488 nm (10), 531 nm (1), 547 nm (5), and 555 nm (1). Comparisons between Fig. 3.4(A) and 3.4(B) indicate that the developed algorithm performs equivalently well for the simulated SeaWiFS and MODIS datasets.

Table 3.6. Top 20 out of 305 spectral features derived from the simulated MODIS dataset

Rank	Feature	Rank	Feature	Rank	Feature
1	CR(488,547,645)	8	CR(469,488,645)	15	CV(443,469,488)
2	CV(412,488,531)	9	CR(488,547,555)	16	CR(488,531,645)
3	Rrs(488)	10	CR(488,547,667)	17	CR(443,488,667)
4	CR(469,488,678)	11	CR(469,488,667)	18	CV(412,488,678)
5	CR(531,547,645)	12	CV(412,488,555)	19	CR(488,555,645)
6	CR(412,443,469)	13	Rrs(469)	20	CR(443,488,678)
7	CV(412,488,547)	14	CR(531,547,667)		

4.2 Results for satellite datasets

4.2.1 SeaWiFS satellite dataset

Correlations between f_m , f_n, and f_p calculated with in-situ HPLC pigments and f_m , f_n, and f_p derived using different numbers of spectral features are listed in table 3.7. The RMSE values of f_m, f_n, and f_p decrease with increasing feature number before reaching the minimum of 0.166, 0.115, and 0.159, respectively, and then have a slightly increasing trend with additional features. In contrast, R^2 increases with the feature number before reaching the maximum of 0.653, 0.520, and 0.583, respectively, and then has a trend of slightly decreasing with increasing feature number for f_m and f_n and flattening out for f_p. Similarly, the slope of the regression line for the correlation also has an increasing trend with feature number, and then decreases with increasing feature number. Taking into

account of model simplicity and computation complexity, f_m, f_n, and f_p derived with the top 15 spectral features are acceptable. Their scatter plots and relative error distributions (Fig. 3.5) show that the measured and estimated values are highly correlative, with R^2 of 0.6532, 0.5201, and 0.5834 for f_m, f_n, and f_p, respectively, and RMSE value of 0.166, 0.115, and 0.159, respectively. Overestimation and underestimation appear at the low and high value domains, respectively. According to Fig. 3.5(D), relative error ranging from -0.4 to 0.4 accounts for 51%, 60.1%, and 64.6% of the validation samples for f_m, f_n, and f_p, respectively.

The top 15 features (table 3.8) include: (1) 8 $CR(\lambda_1,\lambda_2,\lambda_3)$, 3 $CV(\lambda_1,\lambda_2,\lambda_3)$, 3 $Rrs(\lambda)$, and 1 $BR(\lambda)$ for f_m; (2) 9 $CR(\lambda_1,\lambda_2,\lambda_3)$, 2 $CV(\lambda_1,\lambda_2,\lambda_3)$, 3 $Rrs(\lambda)$, and 1 $BR(\lambda)$ for f_n; and (3) 9 $CR(\lambda_1,\lambda_2,\lambda_3)$, 2 $CV(\lambda_1,\lambda_2,\lambda_3)$, 3 $Rrs(\lambda)$, and 1 $BR(\lambda)$ for f_p. Among the 15 features, the number of features using wavelengths 412, 443, 490, 510, and 555 nm is as follows: 1, 2, 4, 6, and 2 for f_m, 1, 1, 2, 6, and 5 for f_n, and 1, 1, 2, 6, and 5 for f_p. Similar to the simulated datasets, continuum removed spectra and spectral curvature also dominate the 15 features, which have dominant wavelengths of 490, 510 and 555 nm. The top 15 spectral features intersect with the top 20 features of the simulated SeaWiFS dataset at: *Rrs*(490), *Rrs*(443), *CR*(510,555,670), *CR*(412,490,670), *CR*(443,490,670), *CR*(490,510,555), and *CR*(412,510,670).

Table 3.7. Validation results for f_m, f_n and f_p derived from the SeaWiFS satellite data set

Phytoplankton size type	Feature number	RMSE	R^2	Slope	Intercept
f_m	5	0.180	0.592	0.640	0.123
	10	0.177	0.608	0.657	0.119
	15	**0.166**	**0.653**	**0.678**	**0.112**
	20	0.167	0.650	0.677	0.112
	30	0.178	0.620	0.733	0.084
	40	0.177	0.625	0.737	0.084
	50	0.180	0.615	0.739	0.086
	60	0.176	0.621	0.698	0.086
	all(67)	0.178	0.624	0.740	0.078
f_n	5	0.129	0.401	0.430	0.162
	10	0.121	0.470	0.510	0.138
	15	**0.115**	**0.520**	**0.563**	**0.121**
	20	0.117	0.504	0.542	0.129
	30	0.124	0.460	0.553	0.123
	40	0.121	0.472	0.526	0.133
	50	0.123	0.463	0.552	0.125
	60	0.125	0.453	0.560	0.120
	all(67)	0.126	0.449	0.560	0.121
f_p	5	0.165	0.552	0.578	0.171
	10	0.161	0.576	0.601	0.161
	15	**0.159**	**0.583**	**0.608**	**0.162**
	20	0.160	0.583	0.617	0.161
	30	0.161	0.580	0.636	0.150
	40	0.161	0.578	0.631	0.157
	50	0.162	0.574	0.637	0.151
	60	0.160	0.582	0.618	0.158
	all(67)	0.159	0.583	0.615	0.161

Table 3.8. Top 15 out of 67 spectral features sensitive to f_m, f_n, and f_p for the SeaWiFS satellite data set

f_m		f_n		f_p	
Rank	**Feature**	**Rank**	**Feature**	**Rank**	**Feature**
1	Rrs(490)	1	CR(510,555,670)	1	CR(510,555,670)
2	CR(412,510,555)	2	CR(412,510,555)	2	CV(443,490,670)
3	CR(510,555,670)	3	CV(443,490,670)	3	Rrs(490)
4	CR(412,490,670)	4	CR(490,555,670)	4	CR(490,510,555)
5	BR(555,670)	5	Rrs(443)	5	BR(555,670)
6	CR(443,490,670)	6	CV(490,510,555)	6	Rrs(443)
7	CR(490,510,555)	7	CR(443,510,670)	7	CR(412,510,555)
8	Rrs(443)	8	Ratio(555,670)	8	CV(490,510,555)
9	CR(412,443,510)	9	CR(412,555,670)	9	CR(412,555,670)
10	CV(443,490,670)	10	CR(490,510,555)	10	Rrs(412)
11	CR(443,510,555)	11	Rrs(412)	11	CR(443,510,555)
12	CR(412,510,670)	12	CR(443,555,670)	12	CR(443,490,670)
13	CV(490,510,555)	13	Rrs(490)	13	CR(490,510,670)
14	CV(412,510,555)	14	CR(490,510,670)	14	CR(412,490,555)
15	Rrs(412)	15	CR(412,510,670)	15	CR(443,555,670)

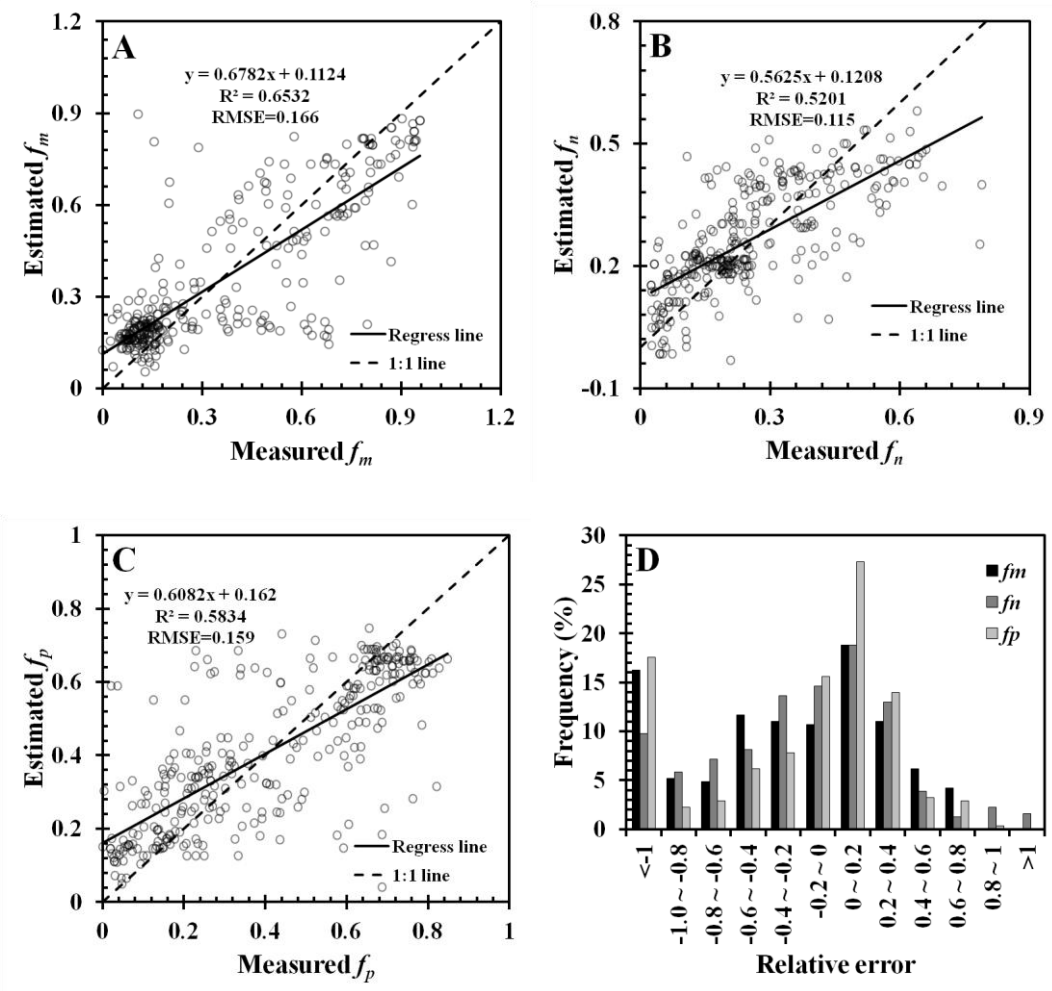


Fig. 3.5.Validation results of top 15 spectral features derived from the SeaWiFS satellite data set; (A), (B), and (C) are for f_m, f_n, and f_p, respectively; and (D) is distribution of relative error frequency (negative values represent overestimation and vice-versa).

4.2.2 MODIS satellite dataset

Correlations between in-situ and spectrally estimated f_m , f_n, and f_p are listed in table 3.9. The RMSE values for f_m, f_n, and f_p drop sharply with the feature number increasing from 5 to 30, and then slightly fluctuate with further increasing feature number. Conversely, R^2 increases quickly with feature number before slight fluctuation appears. The slopes of regression lines increase quickly before reaching a local maximum, and then show a decreasing trend of with further increasing the feature number. f_m, f_n, and f_p

derived with the top 30, 30, and 25 features, respectively, are acceptable. Their scatter plots (Fig. 3.6(A)-(C)) show that the measured values are highly correlated with the estimates, and that most samples converge to the 1:1 line. Samples with relative errors in (-0.4, 0.4) account for 74.0%, 57.0%, and 56.6% of validation samples for f_m, f_n, and f_p, respectively (Fig. 3.6(D)).

Table 3.9. Validation results for f_m, f_n and f_p derived from the MODIS satellite data set

Phytoplankton size type	Feature number	RMSE	R^2	Slope	Intercept
f_m	5	0.194	0.604	0.661	0.160
	15	0.151	0.760	0.814	0.098
	25	0.143	0.781	0.794	0.092
	30	**0.136**	**0.802**	**0.803**	**0.090**
	50	0.134	0.809	0.794	0.091
	100	0.134	0.811	0.805	0.081
	200	0.133	0.819	0.740	0.123
	all(305)	0.137	0.806	0.724	0.135
f_n	5	0.126	0.553	0.559	0.105
	15	0.114	0.633	0.685	0.083
	25	0.110	0.659	0.679	0.085
	30	**0.108**	**0.673**	**0.686**	**0.085**
	50	0.107	0.678	0.666	0.092
	100	0.103	0.698	0.679	0.083
	200	0.104	0.711	0.618	0.112
	all(305)	0.108	0.695	0.585	0.122
f_p	5	0.153	0.504	0.477	0.135
	15	0.109	0.749	0.761	0.073
	25	**0.107**	**0.763**	**0.843**	**0.048**
	30	0.107	0.761	0.834	0.051
	50	0.107	0.768	0.849	0.049
	100	0.108	0.753	0.761	0.076
	200	0.110	0.750	0.690	0.095
	all(305)	0.112	0.744	0.664	0.102

The top spectral features yielding acceptable estimates for f_m, f_n, and f_p are listed in table 10. These features consist of: (1) 19 $CR(\lambda_1,\lambda_2,\lambda_3)$, 5 $CV(\lambda_1,\lambda_2,\lambda_3)$, 5 $Rrs(\lambda)$, and 1 $N\_Rrs(\lambda)$ for f_m; (2) 22 $CR(\lambda_1,\lambda_2,\lambda_3)$, 5 $CV(\lambda_1,\lambda_2,\lambda_3)$, 2 $Rrs(\lambda)$, and 1 $N\_Rrs(\lambda)$ for f_n; and (3)

17 $CR(\lambda_1,\lambda_2,\lambda_3)$, 5 $CV(\lambda_1,\lambda_2,\lambda_3)$, 2 $Rrs(\lambda)$, and 1 $BR(\lambda)$ for f_p. Among the top 30 features for f_m and f_n, and the top 25 spectral features for f_p, the number of spectral features using wavelengths 443, 488, 547 and 645 nm is as follows: 4, 7, 4, and 10 for f_m; 4, 7, 4, and 11 for f_n; and 1, 6, 3, and 10 for f_p. Obviously, continuum removed spectra and spectral curvatures dominate the top spectral features, which have dominant bands of 443, 488, 547 and 645 nm. In addition, features for deriving f_m from simulated (top 20) and from satellite (top 30) datasets intersect at: $Rrs(488)$, $Rrs(469)$, $CR(469,488,678)$, $CR(412,443,469)$, $CR(469,488,645)$, $CR(469,488,667)$, and $CR(531,547,667)$.

Table 3.10. Top 30 out of 305 spectral features sensitive to f_m, f_n, and f_p for the MODIS satellite dataset

f_m		f_n		f_p	
Rank	**Feature**	**Rank**	**Feature**	**Rank**	**Feature**
1	CR(488,645,678)	1	CR(488,645,678)	1	CR(488,645,678)
2	CR(469,488,531)	2	CR(469,488,678)	2	CV(531,645,678)
3	Rrs(531)	3	CV(531,645,667)	3	CR(469,488,531)
4	CV(531,645,667)	4	CR(531,547,678)	4	CR(531,547,667)
5	CR(469,645,678)	5	CR(412,443,488)	5	CV(531,645,667)
6	CR(412,443,488)	6	CR(469,645,678)	6	N_Rrs(412)
7	CR(531,547,667)	7	CV(531,645,678)	7	CR(469,645,678)
8	CR(469,488,555)	8	CR(469,488,531)	8	CR(469,488,547)
9	CR(443,645,678)	9	CR(531,547,667)	9	CV(531,667,678)
10	CV(531,667,678)	10	CR(443,645,678)	10	CR(531,547,678)
11	Rrs(488)	11	Rrs(488)	11	CR(443,645,678)
12	CV(531,645,678)	12	CR(469,488,667)	12	Rrs(488)
13	CR(412,645,678)	13	N_Rrs(412)	13	CR(469,488,667)
14	CR(469,488,547)	14	CR(412,645,678)	14	CR(412,645,678)
15	CR(412,443,547)	15	CV(555,667,678)	15	CR(412,443,469)
16	CR(531,547,678)	16	CR(412,443,531)	16	CR(469,488,555)
17	CR(469,488,645)	17	CR(488,547,678)	17	CV(547,645,678)
18	Rrs(547)	18	CR(469,488,645)	18	CR(531,547,645)
19	CR(531,645,678)	19	Rrs(531)	19	CR(531,645,678)
20	CR(412,443,469)	20	CV(547,645,678)	20	CR(469,488,645)
21	CR(469,488,667)	21	CR(488,547,667)	21	CR(488,645,667)
22	N_Rrs(412)	22	CR(488,645,667)	22	CV(412,531,547)
23	CV(547,645,667)	23	CR(412,443,555)	23	Rrs(531)
24	CR(469,488,678)	24	CR(469,488,555)	24	CR(645,667,678)
25	CR(412,443,531)	25	CV(547,645,667)	25	CR(469,645,667)
26	CR(488,547,678)	26	CR(488,555,667)	26	CR(469,488,678)
27	CV(547,645,678)	27	CR(531,645,678)	27	CV(547,645,667)
28	Rrs(555)	28	CR(469,488,547)	28	Rrs(469)
29	Rrs(469)	29	CR(412,443,469)	29	CR(412,443,531)
30	CR(488,645,667)	30	CR(469,645,667)	30	CR(443,645,667)

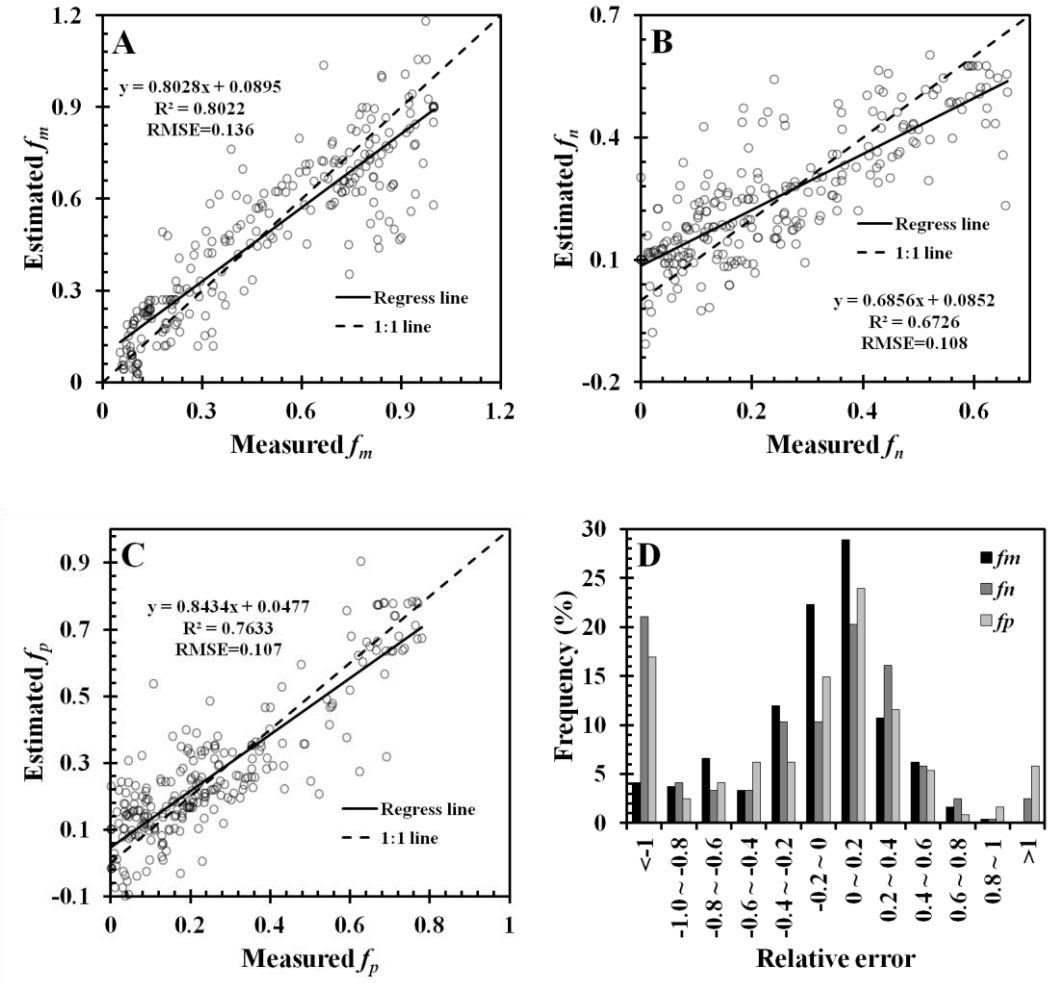
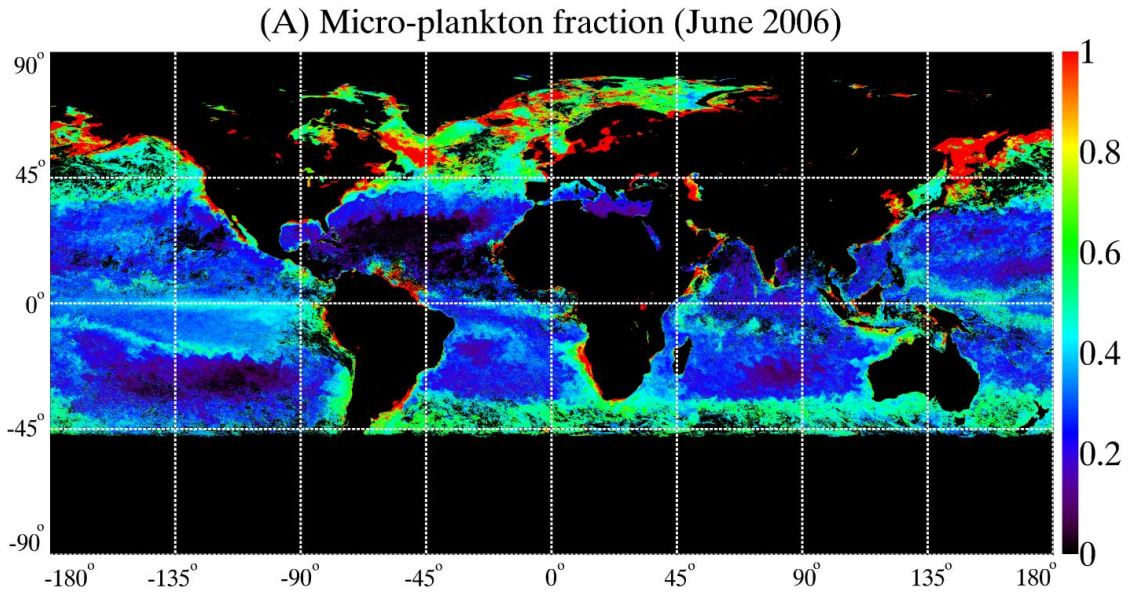


Fig. 3.6. Validation results of MODIS satellite data set; (A) is for f_m derived from top 30 spectral features, (B) is for f_n derived from top 30 spectral features, (C) is for f_p derived from top 25 spectral features, and (D) is distribution of relative error frequency (negative values represent overestimation and vice-versa).

4.3 Global distribution of phytoplankton size fractions

With the models regressed from the SeaWiFS and MODIS satellite datasets, respectively, global distributions of PSF were estimated using corresponding monthly satellite images. The global distribution maps of f_m, f_n, and f_p derived from SeaWiFS monthly images (June 2006) are shown in Fig. 3.7. Fig. 3.7(A) shows that micro-plankton dominates in high-productivity areas, such as the Eastern Boundary Currents, the North

Atlantic, the Equatorial Upwelling regions, and coastal and high latitude areas in general. Fig. 3.7(B) shows that f_n is dominant in transitional, upwelling, coastal and higher latitude regions, such as the Southern Ocean around 60^0 in latitude. Fig. 3.7(C) shows the spatial distribution of f_p, whose prominent feature is the five high-value areas which are approximately corresponding to the five oligotrophic subtropical gyres of the World Ocean – the North and the South Atlantic, the North and the South Pacific, and the South Indian gyres, respectively. These general patterns are consistent with the current understanding that low-productivity areas are dominated by small particles, whereas high-productivity areas are characterized by large particles (Falkowski *et al.* 1998; Boyd *et al.* 2000; Behrenfeld *et al.* 2006; Kostadinov *et al.* 2010).



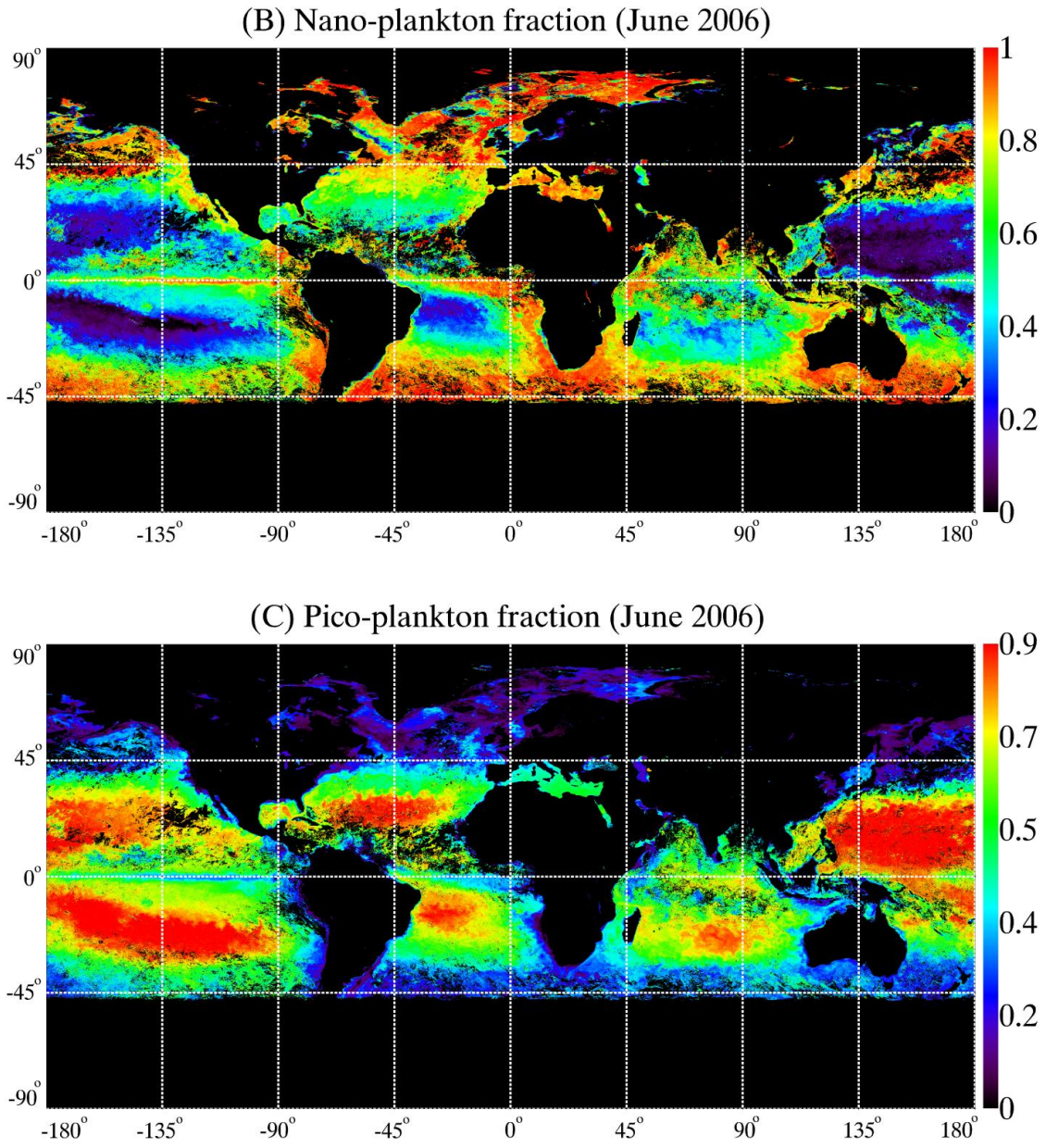
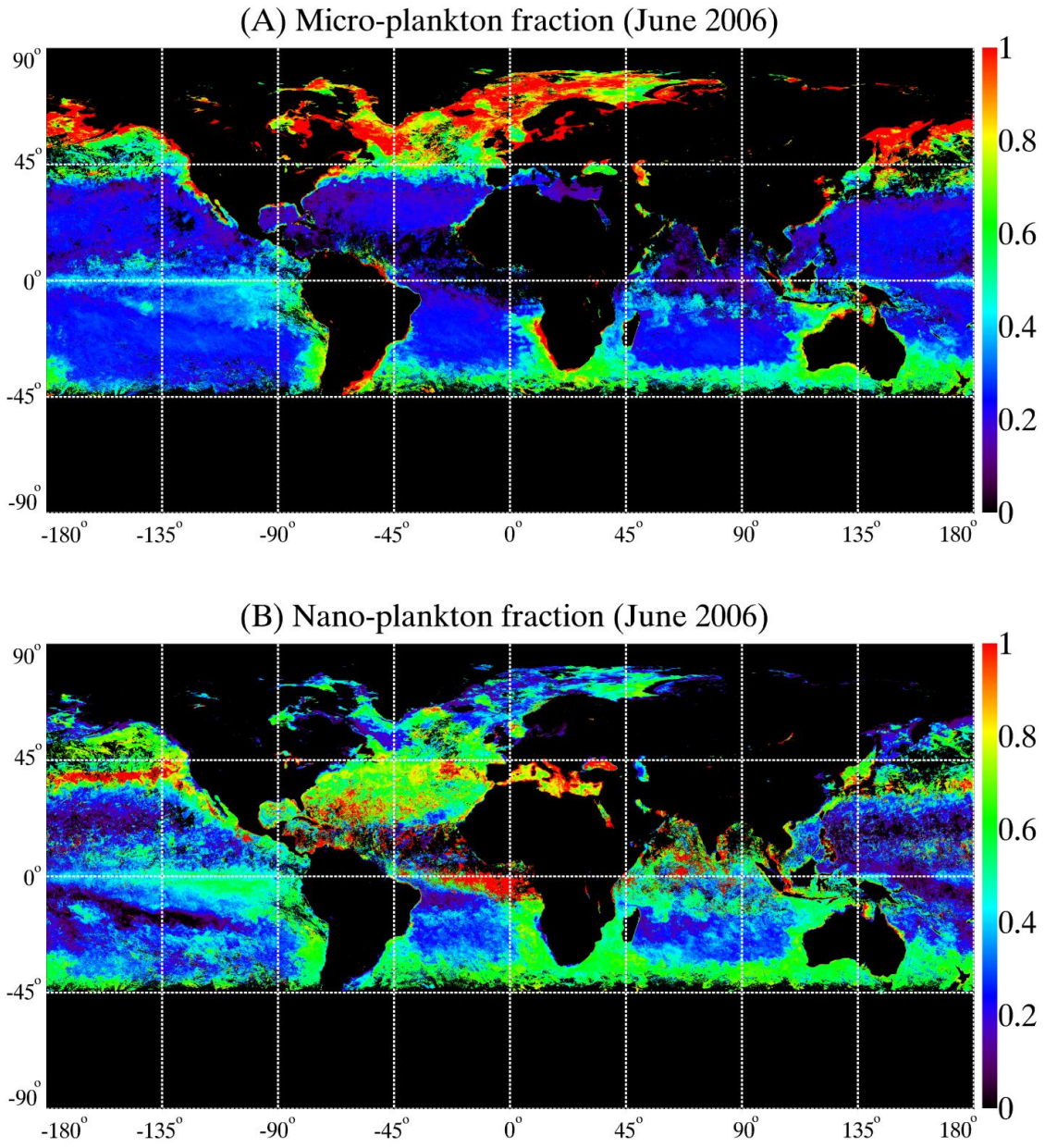


Fig. 3.7. Global phytoplankton size fractions derived from SeaWiFS monthly satellite image (June 2006); (A), (B), and (C) are for f_m, f_n, and f_p, respectively.

The global distribution maps of f_m, f_n, and f_p derived from MODIS monthly images (June 2006) are shown in Fig. 3.8. They show similar patterns to those derived from the SeaWiFS monthly images. However, results derived from the SeaWiFS images (Fig. 3.7) slightly differ from those from the MODIS images (Fig. 3.8). The results for the MODIS

images show higher f_m, than those for the SeaWiFS image in high-productivity regions, such as the North Atlantic and the higher latitude areas. In contrast, the results for the SeaWiFS images displays higher f_n than those for the MODIS images in some high-productivity areas, such as the Eastern Boundary Currents, the North Atlantic, the Equatorial Upwelling regions, and coastal and high latitude areas. Fig. 3.8(C) generally provides more details, especially in the five oligotrophic subtropical gyres.



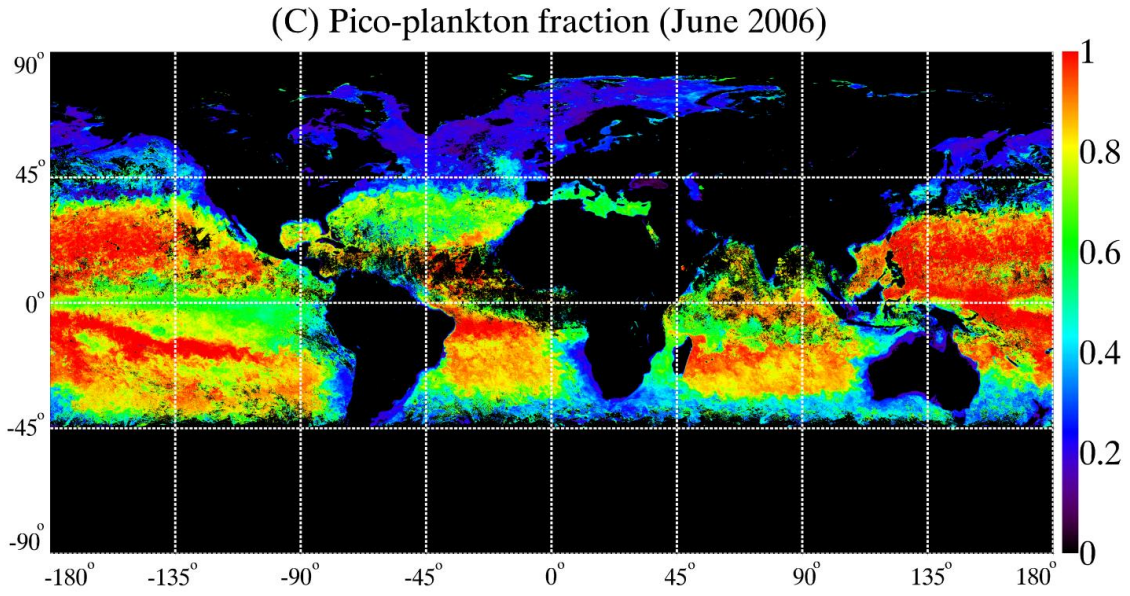


Fig. 3.8. Global phytoplankton size fractions derived from MODIS monthly satellite image (June 2006); (A), (B), and (C) are for f_m, f_n, and f_p, respectively.

5. Discussion

5.1 Spectral features sensitive to phytoplankton size fractions

Spectral features sensitive to PSF have been determined from the simulated and satellite SeaWiFS and MODIS datasets, indicating the importance of the spectral region from around 440 to 645 nm, and the effectiveness of continuum removed spectra and spectral curvature for deriving PSF.

First, selection of the important spectral region 440-645 nm can be attributed to the spectral characteristics of seven pigments (Fuco, Perid, Hex_fuco, But_fuco, Allo, TChl*b*, and Zea) (Bricaud *et al.* 2004; Aiken *et al.* 2008). These pigments have absorptions peaks in the spectral region from about 440 to 645 nm. Although overlaps among pigment absorptions occur, it is still possible to discriminate different pigments, such as discriminating peridinin from alloxanthin. Furthermore, changes in PSF are approximately reflected in variation of pigments, which lead to variation in

phytoplankton IOPs and thus in $Rrs(\lambda)$ in the spectral region from around 440 to 645 nm. Second, the dominance of continuum removed spectra in the spectral features selected to estimate PSF is ascribed to the effectiveness of this feature for: (1) reducing the influences of ambient factors (e.g. solar zenith angle), and (2) emphasizing pigments absorption, which is the major process phytoplankton controls $Rrs(\lambda)$ (Toole and Siegel 2001; Lubac and Loisel 2007; Wang *et al.* 2010). The spectral characteristics of the pigments plus the merits of continuum removed spectra mainly explain the selected features sensitive to PSF. The same explanation is applicable to the spectral curvature. Finally, overlapping (between the simulated and satellite datasets) in spectral features sensitive to f_m further suggests that continuum removed spectra are robust in isolating pigment absorption features and eliminating the influence of uninteresting factors, such as ambient factors. This explanation is based on the fact that the simulated SeaWiFS and MODIS datasets are free of the errors from atmospheric correction, measurement of pigment concentrations, and calculation of PSF.

However, some pigments are extremely similar in the shape and magnitude of absorption spectra, and as a result they are difficult to be discriminated through remote sensing data, especially multiple spectral remote sensing data. This spectral similarity can partially explain the scattering between the measured and remotely estimated PSF, particularly the PSF estimated with the 6 bands SeaWiFS images.

5.2 Validation of the algorithm

Results obtained with different numbers of spectral features indicate that with the simulated and satellite datasets reliable estimates can be obtained with about 20 spectral

features and slight degradation appears when incorporating more spectral features. This indicates that some spectral features are not effective for estimating PSF.

The results obtained from the top 20 spectral features for the simulated SeaWiFS and MODIS datasets indicate that the developed algorithm is robust for estimating PSF for both oligotrophic and eutrophic water, but shows slight overestimation and dispersion for the region lower than around 0.3 in f_m. This overestimation and dispersion may be due to the fact that, when f_m is very low, the water is commonly oligotrophic and $Rrs(\lambda)$ is mainly regulated by pure water scattering (Bricaud *et al.* 2012). Other factors, such as wind speed and bubble, may play the same important role in controlling the variation of $Rrs(\lambda)$ (Stramski *et al.* 2004).

Although the spectral features sensitive to PSF give rise to acceptable results for the satellite datasets, as shown in Fig. 3.5 and 3.6, correlations between the measured and estimated PSF show great dispersion, especially for the results of SeaWiFS satellite dataset. This low correlation can be attributed to several reasons. First, atmospheric signal accounts for a large proportion of $Rrs(\lambda)$, and it is challenging to correct its influence. Even though robust features such as continuum removed spectra and spectral curvatures are used, the impact of atmosphere on the estimated PSF is unavoidable. Second, there are spatiotemporal differences between the matched satellite image spectra and in-situ samples (Kostadinov *et al.* 2009; Kostadinov *et al.* 2010). The spatial resolution of satellite image spectra is 9×9 km, whereas the field observation is limited to one point. Apparently, they are comparable only when the water in a 9×9 km pixel is homogeneous. In order to get as many matched samples as possible, the field observations were matched with satellite image spectra in the same day, rather than

exactly the same observing time. Due to the dynamics of oceans, the time difference degrades the correlation of validation results. Third, the PSF assessment based on diagnostic pigments does not strictly reflect the actual size of phytoplankton communities, in that some pigments are not restricted to one phytoplankton group and also some phytoplankton groups have a wide size range (Vidussi *et al.* 2001; Uitz *et al.* 2006; Devred *et al.* 2011). In addition, the variation in phytoplankton absorption spectra could occur due to other factors, such as photoacclimation related to environmental factors (Eppley 1972; Bricaud *et al.* 2004), rather than variation in phytoplankton group or species (Brewin *et al.* 2011). Actually, phytoplankton cell size also plays an important role in regulating $a_{ph}(\lambda)$ (Lohrenz *et al.* 2003; Bricaud *et al.* 2004). Although correlation between pigment composition and cell size exists, scattering was observed based on in-situ observations (Bricaud *et al.* 2004). This ambiguousness is other possible reasons partially responsible for the scattering of the validation results for satellite datasets. Finally, the sensitivity limit of the spectral features may be partially responsible for overestimation and dispersion, such as f_m (SeaWiFS) lower than 0.4. The discrimination was improved with increased bands, as shown by the accuracy of the estimated f_m from the MODIS dataset.

The performance of the proposed algorithm is also compared with another phytoplankton size retrieval approach in (Mouw and Yoder 2010). The algorithm in (Mouw and Yoder 2010) is based on the assumption that normalized phytoplankton absorption coefficients can be reconstructed by the "basic vectors" of micro-plankton and pico-plankton (Ciotti *et al.* 2002). Thus, this algorithm derives f_p and f_m only. For estimation of f_m this algorithm results in R^2=0.6 and RMSE=0.1264 with a SeaWiFS

image dataset, but our algorithm gives rise to R^2=0.653 and RMSE=0.166 with the SeaWiFS image dataset, and R^2=0.802 and RMSE=0.136 with the MODIS image dataset. It is worth to mention that the satellite datasets used in this paper have wider range f_m than that used in (Mouw and Yoder 2010). The maximum of f_m is 0.956 and 0.999 for the SeaWiFS and the MODIS datasets, respectively, whereas most f_m reported in (Mouw and Yoder 2010) are lower than 0.70, except for 8 samples in the range between 0.7 and about 0.92. The developed algorithm was also compared with the algorithm described in (Ciotti and Bricaud 2006), which was applied to a local dataset. The proposed algorithm results in slightly less RMSE (0.166 for SeaWiFS, 0.136 for MODIS) than that (0.17) in (Ciotti and Bricaud 2006).

5.3 Mapping with satellite images

It should be emphasized that the estimation of global PSF is based on the assumption that the training samples represent typical characteristics of global ocean. According to the validation samples, as shown in Fig. 3.5 and 3.6, PSF includes a wide range of variation. Thus, the samples can represent the typical characteristics of global ocean. Otherwise, the global estimated PSF may be unreliable.

The spatial patterns of estimated global PSF with the developed algorithm are consistent with the current understanding of ocean ecosystem (Falkowski *et al.* 1998; Boyd *et al.* 2000; Behrenfeld *et al.* 2006; Kostadinov *et al.* 2010). Oligotrophic areas, such as the five typical subtropical gyres of the World Ocean (the North and the South Atlantic, the North and the South Pacific, and the South Indian gyres), have high fraction of pico-plankton, and as a result have minimal sinking/export (Boyd and Trull 2007). Nano-plankton accounts for large proportion of the phytoplankton in eutrophic and

transitional regions. Micro-plankton dominates in high-productivity areas, such as the Eastern Boundary Currents, the North Atlantic, the Equatorial Upwelling regions, and coastal and high latitude areas, and is considered to be responsible for new primary production, correlated with flux and to contribute to carbon flux significantly (Goldman 1993; Boyd and Trull 2007; Uitz *et al.* 2010).

6. Conclusions

Here, an algorithm based on spectral features was proposed to derive PSF through support vector machine. Spectral features sensitive to PSF were determined through a band selection method, and then they were used to estimate PSF through support vector machine. According to validation with the simulated and satellite datasets, the following conclusions may be drawn: (1) the developed algorithm is robust in deriving PSF (R^2 is 0.9273, 0.9305 (f_m) for the simulated data; R^2 is 0.6532 (f_m), 0.5201 (f_n), 0.5834 (f_p) for SeaWiFS; and R^2 is 0.8022 (f_m), 0.6726 (f_n), 0.7633 (f_p) for MODIS); and (2) the spectral features sensitive to PSF mainly consist of continuum removed spectra and spectral features, and the wavelengths of the features mainly center on domains of pigment absorption peaks. Multiple reasons attributable to the performances of the developed algorithm were also discussed. Global spatial distributions of derived PSF were consistent with current understanding of oceanic ecosystem, revealing that pico-plankton accounts for large proportion of total phytoplankton in oligotrophic areas, that nano-plankton dominates in eutrophic and transitional areas, and that miro-plankton is dominant in high-productivity areas. Other features, such as climatology data, mixed layer depth, and photosynthetic available radiation, can be incorporated into the feature space to further improve the performance of the algorithm in future work.

Data simulation

Hydrolight was adopted to simulate *Rrs*(λ) datasets with IOPs and other ambient factors (table 3.11) as input. $a(\lambda)$ and $b_b(\lambda)$ are modeled with four components (CDOM is non-scattering) as follows:

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{dm}(\lambda) + a_g(\lambda) \quad \text{(A1)}$$

$$b_b(\lambda) = b_{bw}(\lambda) + b_{b\,ph}(\lambda) + b_{bdm}(\lambda) \quad \text{(A2)}$$

$a_w(\lambda)$ and $b_{bw}(\lambda)$ are extracted from (Pope and Fry 1997) and (Smith and Baker 1981), respectively.

For phytoplankton, $a_{ph}(\lambda)$ is expressed as:

$$a_{ph}(\lambda) = a_{ph}(440)a_{ph}^{N440}(\lambda) \quad \text{(A3)}$$

$a_{ph}^{N440}(\lambda)$ is modeled as (Ciotti *et al.* 2002):

$$a_{ph}^{N440}(\lambda) = (1 - f_m)a_{pico}(\lambda) + f_m a_{micro}(\lambda) \quad \text{(A4)}$$

As [Chl] versus f_m results in a sigmoid shape (Devred *et al.* 2006; Mouw and Yoder 2010), given a [Chl], f_m is a random value between upper (f_m=1-*exp*(-15[Chl])) and lower (f_m=1-*exp*(-0.12[Chl])) boundaries (Mouw and Yoder 2010). According to (Bricaud *et al.* 1995), $a_{ph}(440)$ can be modeled with the following equation:

$$a_{ph}(440) = 0.04[Chl]^{0.668} \quad \text{(A5)}$$

$b_{bph}(\lambda)$ is modeled as (Voss 1992; Roesler and Boss 2003):

$$b_{b\,ph}(\lambda) = \tilde{b}_{ph} b_{ph}(\lambda) \quad \text{(A6)}$$

$$b_{ph}(\lambda) = c_{ph}(\lambda) - a_{ph}(\lambda) \quad \text{(A7)}$$

$$c_{ph}(\lambda) = c_{ph}(550)\left(\frac{550}{\lambda}\right)^{\eta_{ph}} \quad \text{(A8)}$$

$$c_{ph}(550) = cph\_coef[C]^{0.57} \tag{A9}$$

$\tilde{b}_{ph}$ is uniformly distributed between 0.005 and 0.01, and is used to derive the Fournier-Forand phase function for phytoplankton (Fournier and Forand 1994; Mobley *et al.* 2002). Similar to the simulated dataset of the International Ocean Colour Coordinating Group (IOCCG) (Lee 2006), cph_coef is a random number between 0.06 and 0.6, and η_{ph} is generated as follow:

$$\eta_{ph} = -0.4 + \frac{1.6 + 1.2\Re_1}{1 + [C]^{0.5}} \tag{A10}$$

where $\Re_1$ is a random value between 0 and 1.

For detritus/minerals, $a_{dm}(\lambda)$ is modeled as:

$$a_{dm}(\lambda) = a_{dm}(440)\exp(-S_{dm}(\lambda - 440)) \tag{A11}$$

$$a_{dm}(440) = \left(0.1 + \frac{0.5\Re_2 a_{ph}(440)}{0.05 + a_{ph}(440)}\right) a_{ph}(440) \tag{A12}$$

where $\Re_2$ is a random value between 0 and 1. S_{dm} is a random variable obeying Gaussian distribution (mean is 0.0123 nm^{-1}, and standard deviation is 0.0013 nm^{-1}) (Babin *et al.* 2003). $b_{bdm}(\lambda)$ is modeled as:

$$b_{bdm}(\lambda) = \tilde{b}_{dm} b_{dm}(\lambda) \tag{A13}$$

$$b_{dm}(\lambda) = b_{dm}(550)\left(\frac{550}{\lambda}\right)^{\eta_{dm}} \tag{A14}$$

$$b_{dm}(550) = bdm\_coef[M]^{0.63} \tag{A15}$$

$$\eta_{dm} = -0.5 + \frac{2.0 + 1.2\Re_3}{1 + [C]^{0.5}} \tag{A16}$$

where bdm_coef is a random value between 0.06 and 0.6, and $\Re_3$ is a random value between 0 and 1. $\tilde{b}_{dm}$ is uniformly distributed between 0.01 and 0.04, and is used to derive the Fourier-Forand phase function of minerals/detritus.

For CDOM, $a_{cdom}(\lambda)$ is modeled as:

$$a_g(\lambda) = a_g(440)\exp(-S_g(\lambda - 440)) \tag{A17}$$

$$a_g(440) = pa_{ph}(440) \tag{A18}$$

$$p = 0.3 + \frac{5.7\Re_4 a_{ph}(440)}{0.02 + a_{ph}(440)} \tag{A19}$$

$\Re_4$ is a random value between 0 and 1. S_g is a random variable obeying Gaussian distribution (mean is 0.0176 nm^{-1}, and standard deviation is 0.0020 nm^{-1}) (Babin *et al.* 2003). Other parameters for Hydrolight simulation are listed in table 3.11.

Table 3.11. Parameters used in the Hydrolight simulation

Parameter	Variability
Chlorophyll concentration	[0.03,0.3], [0.3,3], and [3,10] with unit of mg m^{-3}; each range randomly generates the same number of samples
Water depth	Infinity
Solar zenith angle	30 degree
Wind speed	5 m s^{-1}
Cloudiness	0

IV. CONCLUSIONS

First, the following conclusion can be drawn based on sensitivity analysis: (1) the variation of remote sensing reflectance is controlled by various factors, including particulate characteristics and mineral concentration; (2) the effects of the characteristics can be reduced by band ratios; and (3) the power-low slope of particulate backscattering coefficient plays an important role in governing the performances of algorithms, and that all the considered parameters are important when the concentration of components are high due to highly interact between factors.

Second, based on the sensitivity analysis results of remote sensing reflectance and QAA and LMI, an improved IOPs inverse algorithm was proposed. By comparing among ILMI, QAA, and LMI, the following conclusions can be drawn: (1) ILMI is superior to the QAA and the LMI for deriving the IOPs when their values are high; and (2) the three algorithms are competitive for the case of low IOPs values.

Third, an algorithm based on spectral features was proposed to derive PSF through support vector machine. Spectral features sensitive to PSF were determined through a band selection method, and then they were used to estimate PSF through support vector machine. According to validation with the simulated and the satellite datasets, the following conclusions can be drawn: (1) the developed algorithm is robust in deriving PSF (R^2 is 0.9273, 0.9305 (f_m) for the simulated data; R^2 is 0.6532 (f_m), 0.5201 (f_n), 0.5834 (f_p) for SeaWiFS; and R^2 is 0.8022 (f_m), 0.6726 (f_n), 0.7633 (f_p) for MODIS); and (2) the spectral features sensitive to the PSF are mainly composed of continuum removed spectra and spectral features, and the wavelengths of the features are mainly in the wavelength region where pigment absorption is strong. Multiple reasons attributable to

the performances of the developed algorithm were also discussed. Global spatial distributions of derived PSF were consistent with current understanding of oceanic ecosystem, revealing that pico-plankton accounts for large proportion of total phytoplankton in oligotrophic areas, that nano-plankton dominates in eutrophic and transitional areas, and that miro-plankton is dominant in high-productivity areas. Other features, such as climatology data, mixed layer depth, and photosynthetic available radiation, can be incorporated into the feature space to further improve the performance of the algorithm in future work.

REFERENCES

Aas, E. (1996). Refractive index of phytoplankton derived from its metabolite composition. Journal of Plankton Research **18**(12): 2223-2249.

Ahn, Y. H., A. Bricaud and A. Morel (1992). Light backscattering efficiency and related properties of some phytoplankters. Deep-Sea Research Part a-Oceanographic Research Papers **39**(11-12A): 1835-1855.

Aiken, J., N. J. Hardman-Mountford, R. Barlow, J. Fishwick, T. Hirata and T. Smyth (2008). Functional links between bioenergetics and bio-optical traits of phytoplankton taxonomic groups: an overarching hypothesis with applications for ocean colour remote sensing. Journal of Plankton Research **30**(2): 165-181.

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

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

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

Archibald, R. and G. Fann (2007). Feature selection and classification of hyperspectral images, with support vector machines. IEEE Geoscience and Remote Sensing Letters **4**(4): 674-677.

Armstrong, R. A., C. Lee, J. I. Hedges, S. Honjo and S. G. Wakeham (2002). A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. Deep-Sea Research Part Ii-Topical Studies in Oceanography **49**(1-3): 219-236.

Babin, M., A. Morel, V. Fournier-Sicre, F. Fell and D. Stramski (2003). Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. Limnology and Oceanography **48**(2): 843-859.

Babin, M. and D. Stramski (2002). Light absorption by aquatic particles in the near-infrared spectral region. Limnology and Oceanography **47**(3): 911-915.

Babin, M., D. Stramski, G. M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky and N. Hoepffner (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. Journal of Geophysical Research-Oceans **108**(C7).

Bader, H. (1970). The hyperbolic distribution of particle sizes. Journal of Geophysical Research **75**(15): 2822-2830.

Barnard, A. H., W. S. Pegau and J. R. V. Zaneveld (1998). Global relationships of the inherent optical properties of the oceans. Journal of Geophysical Research-Oceans **103**(C11): 24955-24968.

Bazi, Y. and F. Melgani (2006). Toward an optimal SVM classification system for hyperspectral remote sensing images. IEEE Transactions on Geoscience and Remote Sensing **44**(11): 3374-3385.

Behrenfeld, M. J., R. T. O'Malley, D. A. Siegel, C. R. McClain, J. L. Sarmiento, G. C. Feldman, A. J. Milligan, P. G. Falkowski, R. M. Letelier and E. S. Boss (2006). Climate-driven trends in contemporary ocean productivity. Nature **444**(7120): 752-755.

Binding, C. E., D. G. Bowers and E. G. Mitchelson-Jacob (2005). Estimating suspended sediment concentrations from ocean colour measurements in moderately turbid waters; the impact of variable particle scattering properties. Remote Sensing of Environment **94**(3): 373-383.

Bohren, C. F. and D. Huffman (1983). Absorption and scattering of light by samll particles. John Wiley, New York.

Boss, E., W. S. Pegau, W. D. Gardner, J. R. V. Zaneveld, A. H. Barnard, M. S. Twardowski, G. C. Chang and T. D. Dickey (2001). Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf. Journal of Geophysical Research-Oceans **106**(C5): 9509-9516.

Boss, E., M. S. Twardowski and S. Herring (2001). Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. Applied Optics **40**(27): 4885-4893.

Boyd, P. W. and T. W. Trull (2007). Understanding the export of biogenic particles in oceanic waters: Is there consensus? Progress in Oceanography **72**(4): 276-312.

Boyd, P. W., A. J. Watson, C. S. Law, E. R. Abraham, T. Trull, R. Murdoch, D. C. E. Bakker, A. R. Bowie, K. O. Buesseler, H. Chang, M. Charette, P. Croot, K. Downing, R. Frew, M. Gall, M. Hadfield, J. Hall, M. Harvey, G. Jameson, J. LaRoche, M. Liddicoat, R. Ling, M. T. Maldonado, R. M. McKay, S. Nodder, S. Pickmere, R. Pridmore, S. Rintoul, K. Safi, P. Sutton, R. Strzepek, K. Tanneberger, S. Turner, A. Waite and J. Zeldis (2000). A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. Nature **407**(6805): 695-702.

Brewin, R. J. W., E. Devred, S. Sathyendranath, S. J. Lavender and N. J. Hardman-Mountford (2011). Model of phytoplankton absorption based on three size classes. Applied Optics **50**(22): 4535-4549.

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 Sensing of Environment **115**(2): 325-339.

Bricaud, A., M. Babin, A. Morel and H. Claustre (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. Journal of Geophysical Research-Oceans **100**(C7): 13321-13332.

Bricaud, A., A. M. Ciotti and B. Gentili (2012). Spatial-temporal variations in phytoplankton size and colored detrital matter absorption at global and regional scales, as derived from twelve years of SeaWiFS data (1998-2009). Global Biogeochemical Cycles **26**: 1-17.

Bricaud, A., H. Claustre, J. Ras and K. Oubelkheir (2004). Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. Journal of Geophysical Research-Oceans **109**(C11).

Bricaud, A., C. Mejia, D. Biondeau-Patissier, H. Claustre, M. Crepon and S. Thiria (2007). Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. Applied Optics **46**(8): 1251-1260.

Bricaud, A. and A. Morel (1986). Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. Applied Optics **25**(4): 571-580.

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-Oceans **103**(C13): 31033-31044.

Brown, C. A., Y. Huot, P. J. Werdell, B. Gentili and H. Claustre (2008). The origin and global distribution of second order variability in satellite ocean color and its potential applications to algorithm development. Remote Sensing of Environment **112**(12): 4186-4203.

Buesseler, K. O., C. H. Lamborg, P. W. Boyd, P. J. Lam, T. W. Trull, R. R. Bidigare, J. K. B. Bishop, K. L. Casciotti, F. Dehairs, M. Elskens, M. Honda, D. M. Karl, D. A. Siegel, M. W. Silver, D. K. Steinberg, J. Valdes, B. Van Mooy and S. Wilson (2007). Revisiting carbon flux through the ocean's twilight zone. Science **316**(5824): 567-570.

Burd, A. B. and G. A. Jackson (2009). Particle aggregation. Annual Review of Marine Science **1**: 65-90.

Campbell, J. W. and W. E. Esaias (1982). Basis for spectral curvature algorithms in remote sensing of chlorophyll. Applied Optics **22**(7): 1094-1093.

Carder, K. L., R. G. Steward, G. R. Harvey and P. B. Ortner (1989). Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll. Limnology and Oceanography **34**(1): 68-81.

Cassar, N., M. L. Bender, B. A. Barnett, S. Fan, W. J. Moxim, H. Levy and B. Tilbrook (2007). The Southern Ocean biological response to Aeolian iron deposition. Science **317**(5841): 1067-1070.

Ceccato, P., S. Flasse, S. Tarantola, S. Jacquemoud and J. M. Gregoire (2001). Detecting vegetation leaf water content using reflectance in the optical domain. Remote Sensing of Environment **77**(1): 22-33.

Ceccato, P., N. Gobron, S. Flasse, B. Pinty and S. Tarantola (2002). Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1 - Theoretical approach. Remote Sensing of Environment **82**(2-3): 188-197.

Chami, M., D. Mckee, E. Leymarie and G. Khomenko (2006). Influence of the angular shape of the volume-scattering function and multiple scattering on remote sensing reflectance. Applied Optics **45**(36): 9210-9220.

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.

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

Clavano, W. R., E. Boss and L. Karp-Boss (2007). Inherent optical properties of non-spherical marine-like particles - From theory to observation. Oceanography and Marine Biology, Vol 45 **45**: 1-38.

Dandonneau, Y., P. Y. Deschamps, J. M. Nicolas, H. Loisel, J. Blanchot, Y. Montel, F. Thieuleux and G. Becu (2004). Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, equatorial Pacific and South Pacific. Deep-Sea Research Part Ii-Topical Studies in Oceanography **51**(1-3): 303-318.

del Giorgio, P. A. and C. M. Duarte (2002). Respiration in the open ocean. Nature **420**(6914): 379-384.

Devred, E., S. Sathyendranath, V. Stuart, H. Maass, O. Ulloa and T. Platt (2006). A two-component model of phytoplankton absorption in the open ocean: Theory and applications. Journal of Geophysical Research-Oceans **111**(C3).

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

Doron, M., M. Babin, A. Mangin and O. Hembise (2007). Estimation of light penetration, and horizontal and vertical visibility in oceanic and coastal waters from surface reflectance. Journal of Geophysical Research-Oceans **112**(C6)

Doxaran, D., J. M. Froidefond, S. Lavender and P. Castaing (2002). Spectral signature of highly turbid waters - Application with SPOT data to quantify suspended particulate matter concentrations. Remote Sensing of Environment **81**(1): 149-161.

Duarte, J., M. V. Reyes, S. Tarantola, F. Gilbes and R. Armstrong (2003). A probabilistic sensitivity analysis of water-leaving radiance to water constituents in coastal shallow waters. Proceedings of SPIE **5155**.

Dugdale, R. C. and J. J. Goering (1967). Uptake of new and regenerated forms of nitrogen in primary productivity. Limnology and Oceanography **12**(2): 196-206.

Ekström, P.-A. (2005). A simulation toolbox for sensitivity analysis. Masters Degree Project, Faculty of Science and Technology, Uppsala Universitet.

Eppley, R. W. (1972). Temperature and phytoplankton growth in the sea. Fishery Bulletin **70**(4): 1063-1085.

Falkowski, P., R. J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hogberg, S. Linder, F. T. Mackenzie, B. Moore, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek and W. Steffen (2000). The global carbon cycle: A test of our knowledge of earth as a system. Science **290**(5490): 291-296.

Falkowski, P. G., R. T. Barber and V. Smetacek (1998). Biogeochemical controls and feedbacks on ocean primary production. Science **281**(5374): 200-206.

Finkel, Z. V., J. Beardall, K. J. Flynn, A. Quigg, T. A. V. Rees and J. A. Raven (2010). Phytoplankton in a changing world: cell size and elemental stoichiometry. Journal of Plankton Research **32**(1): 119-137.

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

Fournier, G. and J. L. Forand (1994). Analytic phase function for ocean water. Proc. SPIE **2258**: 194-201.

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

Francois, R., S. Honjo, R. Krishfield and S. Manganini (2002). Factors controlling the flux of organic carbon to the bathypelagic zone of the ocean. Global Biogeochemical Cycles **16**(4).

Garver, S. A. and D. A. Siegel (1997). Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation .1. Time series from the Sargasso Sea. Journal of Geophysical Research-Oceans **102**(C8): 18607-18625.

Goldman, J. C. (1993). Potential role of large oceanic diatoms in new primary production. Deep-Sea Research Part I-Oceanographic Research Papers **40**(1): 159-168.

Gordon, H. R., O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker and D. K. Clark (1988). A semianalytic radiance model of ocean color. Journal of Geophysical Research-Atmospheres **93**(D9): 10909-10924.

Gordon, H. R., O. B. Brown and M. M. Jacobs (1975). Computed relationship between the inherent and apparent optical properties of a flat homogeneous ocean. Applied Optics **14**(2): 417-427.

Gordon, H. R. and B. A. Franz (2008). Remote sensing of ocean color: Assessment of the water-leaving radiance bidirectional effects on the atmospheric diffuse transmittance for SeaWiFS and MODIS intercomparisons. Remote Sensing of Environment **112**(5): 2677-2685.

Gruber, N. and J. L. Sarmiento (1997). Global patterns of marine nitrogen fixation and denitrification. Global Biogeochemical Cycles **11**(2): 235-266.

Guidi, L., L. Stemmann, G. A. Jackson, F. Ibanez, H. Claustre, L. Legendre, M. Picheral and G. Gorsky (2009). Effects of phytoplankton community on production, size and export of large aggregates: A world-ocean analysis. Limnology and Oceanography **54**(6): 1951-1963.

Guyon, I., J. Weston, S. Barnhill and V. Vapnik (2002). Gene selection for cancer classification using support vector machines. Machine Learning **46**(1-3): 389-422.

Haltrin, V. I. (1998). Self-consistent approach to the solution of the light transfer problem for irradiances in marine waters with arbitrary turbidity, depth, and surface illumination. I. Case of absorption and elastic scattering. Applied Optics **37**(18): 3773-3784.

Hoge, F. E. and P. E. Lyon (1996). Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors. Journal of Geophysical Research-Oceans **101**(C7): 16631-16648.

Hoge, F. E. and R. N. Swift (1986). Chlorophyll pigment concentration using spectral curvature algorithms: an evaluation of present and proposed satellite ocean color sensor bands. Applied Optics **25**(20): 3677-3682.

Hood, R. R., E. A. Laws, R. A. Armstrong, N. R. Bates, C. W. Brown, C. A. Carlson, F. Chai, S. C. Doney, P. G. Falkowski, R. A. Feely, M. A. M. Friedrichs, M. R. Landry, J. K. Moore, D. M. Nelson, T. L. Richardson, B. Salihoglu, M. Schartau, D. A. Toole and J. D. Wiggert (2006). Pelagic functional group modeling: Progress, challenges and prospects. Deep-Sea Research Part Ii-Topical Studies in Oceanography **53**(5-7): 459-512.

Huang, Z., B. J. Turner, S. J. Dury, I. R. Wallis and W. J. Foley (2004). Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. Remote Sensing of Environment **93**(1-2): 18-29.

Irwin, A. J., Z. V. Finkel, O. M. E. Schofield and P. G. Falkowski (2006). Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. Journal of Plankton Research **28**(5): 459-471.

Jacquemoud, S. and F. Baret (1990). Prospect - a model of leaf optical-properties spectra. Remote Sensing of Environment **34**(2): 75-91.

Kirk, J. T. O. (1991). Volume scattering function, average cosines, and the underwater light field. Limnology and Oceanography **36**(3): 455-467.

Kokaly, R. F. and R. N. Clark (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sensing of Environment **67**(3): 267-287.

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

Kostadinov, T. S., D. A. Siegel and S. Maritorena (2010). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. Biogeosciences **7**(10): 3239-3257.

Lam, P. J., S. C. Doney and J. K. B. Bishop (2011). The dynamic ocean biological pump: Insights from a global compilation of particulate organic carbon, $CaCO_3$, and opal concentration profiles from the mesopelagic. Global Biogeochemical Cycles **25**.

Le, C. F., Y. M. Li, Y. Zha, D. Y. Sun, C. C. Huang and H. Zhang (2011). Remote estimation of chlorophyll a in optically complex waters based on optical classification. Remote Sensing of Environment **115**(2): 725-737.

Le Quere, C., S. P. Harrison, I. C. Prentice, E. T. Buitenhuis, O. Aumont, L. Bopp, H. Claustre, L. C. Da Cunha, R. Geider, X. Giraud, C. Klaas, K. E. Kohfeld, L. Legendre, M. Manizza, T. Platt, R. B. Rivkin, S. Sathyendranath, J. Uitz, A. J. Watson and D. Wolf-Gladrow (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Global Change Biology **11**(11): 2016-2040.

Lee, Z. (2006). Remote sensing of inherent optical properties: fundamentals, tests of algorithms, and applications. Reports of the International Ocean-Colour Coordinating Group.

Lee, Z. and K. L. Carder (2000). Band-ratio or spectral-curvature algorithms for satellite remote sensing? Applied Optics **39**(24): 4377-4380.

Lee, Z., K. L. Carder and K. P. Du (2004). Effects of molecular and particle scatterings on the model parameter for remote-sensing reflectance. Applied Optics **43**(25): 4957-4964.

Lee, Z., B. Lubac, P. J. Werdell and R. Arnone (2009). An update of the quasi-analytical algorithm (QAA_v5). in International Ocean Color Group Software Report.

Lee, Z. P., R. Arnone, C. M. Hu, P. J. Werdell and B. Lubac (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. Applied Optics **49**(3): 369-381.

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

Lee, Z. P., K. L. Carder, C. D. Mobley, R. G. Steward and J. S. Patch (1999). Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization. Applied Optics **38**(18): 3831-3843.

Lohrenz, S. E., A. D. Weidemann and M. Tuel (2003). Phytoplankton spectral absorption as influenced by community size structure and pigment composition. Journal of Plankton Research **25**(1): 35-61.

Loisel, H., X. Meriaux, J. F. Berthon and A. Poteau (2007). Investigation of the optical backscattering to scattering ratio of marine particles in relation to their biogeochemical composition in the eastern English Channel and southern North Sea. Limnology and Oceanography **52**(2): 739-752.

Loisel, H. and A. Morel (2001). Non-isotropy of the upward radiance field in typical coastal (Case 2) waters. International Journal of Remote Sensing **22**(2-3): 275-295.

Loisel, H., J. M. Nicolas, A. Sciandra, D. Stramski and A. Poteau (2006). Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean. Journal of Geophysical Research-Oceans **111**(C9).

Lubac, B. and H. Loisel (2007). Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. Remote Sensing of Environment **110**(1): 45-58.

Maritorena, S., D. A. Siegel and A. R. Peterson (2002). Optimization of a semianalytical ocean color model for global-scale applications. Applied Optics **41**(15): 2705-2714.

Mobley, C. D. (1994). Light and water: radiative transfer in natural waters. Academic Press

Mobley, C. D., L. K. Sundman and E. Boss (2002). Phase function effects on oceanic light fields. Applied Optics **41**(6): 1035-1050.

Moore, G. F., J. Aiken and S. J. Lavender (1999). The atmospheric correction of water colour and the quantitative retrieval of suspended particulate matter in Case II waters: application to MERIS. International Journal of Remote Sensing **20**(9): 1713-1733.

Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case-I waters). Journal of Geophysical Research-Oceans **93**(C9): 10749-10768.

Morel, A., D. Antoine and B. Gentili (2002). Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. Applied Optics **41**(30): 6289-6306.

Morel, A. and A. Bricaud (1981). Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. Deep-Sea Research **28A**(11): 1375-1393.

Morel, A. and B. Gentili (1991). Diffuse reflectance of oceanic waters: its dependence on Sun angle as influenced by the molecular scattering contribution. Applied Optics **30**(30): 4427-4438.

Morel, A. and B. Gentili (1993). Diffuse reflectance of oceanic waters .II Bidirectional aspects. Applied Optics **32**(33): 6864-6879.

Morel, A. and B. Gentili (1996). Diffuse reflectance of oceanic waters .III. Implication of bidirectionality for the remote-sensing problem. Applied Optics **35**(24): 4850-4862.

Morel, A. and S. Maritorena (2001). Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research-Oceans **106**(C4): 7163-7180.

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

Mouw, C. B. and J. A. Yoder (2010). Optical determination of phytoplankton size composition from global SeaWiFS imagery. Journal of Geophysical Research-Oceans **115**.

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

Patterson, E. M., D. A. Gillette and B. H. Stockton (1977). Complex index of refraction between 300 and 700 nm for Saharan dust. Journal of Geophysical Research **82**(21): 3153-3160.

Peng, F. and S. W. Effler (2007). Suspended minerogenic particles in a reservoir: Light-scattering features from individual particle analysis. Limnology and Oceanography **52**(1): 204-216.

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

Robertson, J. E., C. Robinson, D. R. Turner, P. Holligan, A. J. Watson, P. Boyd, E. Fernandez and M. Finch (1994). The impact of a coccolithophore bloom on oceanic carbon uptake in the northeast Atlantic during summer 1991. Deep-Sea Research Part I-Oceanographic Research Papers **41**(2): 297-314.

Roesler, C. S. and E. Boss (2003). Spectral beam attenuation coefficient retrieved from ocean color inversion. Geophysical Research Letters **30**(9).

Saltelli, A. (2000). What is sensitivity analysis? In Sensitivity analysis, Wiley Ser. Probab. Stat.: 3-13.

Saltelli, A., S. Tarantola and K. P.-S. Chan (1999). A quantitative model independent method for global sensitivity analysis of model output. Technometrics **41**(1): 39-56.

Sathyendranath, S., L. Prieur and A. Morel (1989). A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. International Journal of Remote Sensing **10**(8): 1373-1394.

Sheldon, R. W., A. Prakash and W. H. Sutcliffe (1972). The size distribution of particles in the ocean. Limnology and Oceanography **XVII**(327-340).

Sieburth, J. M., V. Smetacek and J. Lenz (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. Limnology and Oceanography **23**(6): 1256-1263.

Smith, R. C. and K. S. Baker (1981). Optical properties of the clearest natural waters. Applied Optics **20**(2): 177-184.

Stemmann, L. and E. Boss (2012). Particle and plankton size and packaging: from determining optical properties to deriving the biological pump. Annual Review of Marine Science **4**: 1-27.

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**(1): 27-56.

Stramski, D., A. Bricaud and A. Morel (2001). Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. Applied Optics **40**(18): 2929-2945.

Stramski, D. and D. A. Kiefer (1991). Light scattering by microorganisms in the open ocean. Progress in Oceanography **28**(4): 343-383.

Toole, D. A. and D. A. Siegel (2001). Modes and mechanisms of ocean color variability in the Santa Barbara Channel. Journal of Geophysical Research-Oceans **106**(C11): 26985-27000.

Trees, C. C., D. K. Clark, R. R. Bidigare, M. E. Ondrusek and J. L. Mueller (2000). Accessory pigments versus chlorophyll a concentrations within the euphotic zone: A ubiquitous relationship. Limnology and Oceanography **45**(5): 1130-1143.

Twardowski, M. S., E. Boss, J. B. Macdonald, W. S. Pegau, A. H. Barnard and J. R. V. Zaneveld (2001). A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. Journal of Geophysical Research-Oceans **106**(C7): 14129-14142.

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**.

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

Vapnik, V. (1995). The nature of statistical learning theory. New York: Springer-Verlag.

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-Oceans **106**(C9): 19939-19956.

Voss, K. J. (1992). A spectral model of the beam attenuation coefficient in the ocean and coastal areas. Limnology and Oceanography **37**(3): 501-509.

Wang, P., E. S. Boss and C. Roesler (2005). Uncertainties of inherent optical properties obtained from semianalytical inversions of ocean color. Applied Optics **44**(19): 4074-4085.

Wang, Q., D. Y. Sun, Y. M. Li, C. F. Le and C. C. Huang (2010). Mechanisms of remote-sensing reflectance variability and its relation to bio-optical processes in a highly turbid eutrophic lake: Lake Taihu (China). IEEE Transactions on Geoscience and Remote Sensing **48**(1): 575-584.

Werdell, P. J. and S. W. Bailey (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sensing of Environment **98**(1): 122-140.

Wozniak, S. B. and D. Stramski (2004). Modeling the optical properties of mineral particles suspended in seawater and their influence on ocean reflectance and chlorophyll estimation from remote sensing algorithms. Applied Optics **43**(17): 3489-3503.

Yan, B. K., R. S. Wang, F. P. Gan and Z. C. Wang (2010). Minerals mapping of the lunar surface with Clementine UVVIS/NIR data based on spectra unmixing method and Hapke model. Icarus **208**(1): 11-19.

Yang, H. Y. and H. R. Gordon (1997). Remote sensing of ocean color: assessment of water-leaving radiance bidirectional effects on atmospheric diffuse transmittance. Applied Optics **36**(30): 7887-7897.

Zaneveld, J. R. V. (1995). A theoretical derivation of the dependence of the remotely sensed reflectance of the ocean on the inherent optical properties. Journal of Geophysical Research-Oceans **100**(C7): 13135-13142.

Zhang, R. and J. W. Ma (2009). Feature selection for hyperspectral data based on recursive support vector machines. International Journal of Remote Sensing **30**(14): 3669-3677.

Zhou, W. (2008). Theoretical modeling on light scattering properties of phytoplankton. Ph.D. dissertation.

Zhou, W. and W. Cao (2008). Modelling the influence of mineral particles and phytoplankton on chlorophyll estimation from ocean colour remote sensing. International Journal of Remote Sensing **29**(21): 6237-6248.

CURRICULUM VITAE

Zuchuan Li

Education:

Indiana University-Purdue University Indianapolis, Indianapolis (USA), Earth Science
Master of Science, July 2012

Chinese Academy of Sciences, Beijing (China), Geographic Information System
Master of Science, July 2010

Wuhan University, Wuhan (China), Geodesy and Geomatics Engineering
Bachelor of Science, July 2006

Publications and manuscripts:

Z. Li., L. Li, S. Song, & K. Shi. Sensitivity analysis for modeling optical properties. *Remote Sensing of Environment*, will be submitted.

Z. Li., L. Li, & S. Song. Estimation of phytoplankton size fractions based on spectral features of remote sensing data. *Journal of Geophysical Research - Oceans*, will be submitted.