UNIVERSITY OF BREMEN

Faculty of Biology/ Chemistry (FB02)

Estimating phytoplankton pigments in the changing Arctic Ocean

by

Yangyang Liu

A thesis for the degree of
Doctor of Philosophy in Natural Sciences (Dr. rer. Nat.)

Defended on: 20 May 2022

Evaluators: Prof. Dr. Dieter Wolf-Gladrow, Prof. Dr. Oliver Zielinski

Advisor: Prof. Dr. Dieter Wolf-Gladrow

When I let go of who I am, I become who I might be. (Lao Tsu)

Acknowledgements

Time flies. I still can't believe that by the time I finish this PhD thesis, 6.5 years has past since the beginning of my PhD study in September 2015. These years have been dramatic for me. A turning point of my PhD study - as well as of my life - happened in July 2020 when I changed my PhD supervisor, working group, and in part the thesis committee. While many people have been involved in my PhD life, here I acknowledge in particular those who have supported me scientifically and mentally the most to make this thesis possible.

First and foremost, I thank my teacher, mentor, and PhD supervisor, Dieter Wolf-Gladrow, for your tremendous patience and trust, for providing me with a role model of a scientist, a teacher, a supervisor, a person, and a being, for guiding me through the PhD amid life discomfort, and for the discussions on science, music, and life. I thank other members of my thesis committee in addition to Dieter, Kai Bishof, Christine Klaas, Eva-Maria Nöthig, and Benjamin Rabe, for always being helpful and guiding me through the PhD. I thank my guest advisor in the University of Maine, Emmanuel Boss, and your working group, for providing a free, open and friendly atmosphere of scientific exchange and discussions. I thank Rüdiger Röttgers for sharing your expertise with me and help with my work. I thank my previous working group, AWI Phytooptics group. During the time working in the group I developed skills necessary for my publications. I thank my current working group, AWI Polar Biological Oceanography section, especially Morten Iversen and Tanja Glawatty, for providing a welcoming and open working atmosphere. I thank Christine Klaas, Eva-Maria Nöthig, Jan Streffing, Laura Mathieu, and Francesca Doglioni, for reading my thesis and providing useful comments. I thank my friends, especially Fenina Buttler, Laura Mathieu, Jan Streffing, Julia Gelbing, Francesca Doglioni, Andrea Bleyer, Vasileios Pefanis, Yang Liu, and Victor Smetacek, and counselors/psychotherapists, Ute Seemann, Stefanie Lützen, and Rudolf Bühlmann, for your invaluable mental support for me. I thank my family, my father, mother, and sister, for providing me with the opportunity of personal development. Special thanks to my dearest niece for expressing to me your abundant unconditional love and creativity. Last but not the least, I thank myself, for being afraid and courageous, vulnerable and resilient, strong and soft, anxious and hopeful, and for keep questing for Truth and beauty.

Again, thank you all. Deep gratitude.

Contents

Appendix A Extra Publications

Chapter 1

Summary

Human-induced climate change is amplified in the Arctic. At the root of these amplifications are changes in air temperature and sea ice. The sea-ice cover is dramatically receding in the Arctic Ocean. In the study region of the thesis, the Fram Strait (the largest and only deep gateway to the Arctic Ocean) and its vicinity, changes have been observed in sea-ice conditions and water temperatures due to Arctic warming. This has consequences for phytoplankton. Phytoplankton are one of the main primary producers in the Arctic Ocean. Arctic warming induced alterations in light and nutrient regimes impact phytoplankton seasonality, biomass, community composition and distribution. Phytoplankton biomass and community composition are often indicated by their pigment composition and concentrations. To study the response of phytoplankton to the changing climate, this thesis aims to estimate phytoplankton pigments using observations from the shipboard underway flow-through AC-S spectrophotometer system and the Regulated Ecosystem model version 2 (REcoM2) (Hohn, 2008; Schartau et al., 2007) implemented with phytoplankton growth and photoinhibition models.

In the first part of the thesis, an underway flow-through AC-S system was set up onboard *R.V. Polarstern* during two Fram Strait cruises, PS93.2 and PS99.2. Hyperspectral particulate absorption coefficient ($a_p(\lambda)$) was derived from the underway AC-S measurements. Particulate absorption line height at 676 nm calculated from $a_p(\lambda)$ was empirically related to high performance liquid chromatography (HPLC) chlorophyll a (Chl a) concentrations for PS93.2 and PS99.2, respectively. Both relationships were applied to high frequency (4 Hz) AC-S data to estimate Chl a concentrations along the cruise tracks. In total, 24424 and 16110 Chl a data points were generated for PS93.2 and PS99.2, respectively. The reconstructed AC-S Chl a data sets were used to evaluate seven satellite Chl a algorithms. The number of AC-S-satellite match-ups is over one order of magnitude greater than HPLC-satellite match-ups. AC-S-satellite match-ups show that all algorithms were characterized by an overestimation of satellite Chl a. Two algorithms based on Polymer atmospheric correction processors (Steinmetz et al., 2011) generated data products with relatively high estimation precision and small error. The

Polymer atmospheric correction processors account for sun glint and thin clouds in their reflectance models to derive atmospheric corrected remote sensing reflectance, allowing a much larger spatial coverage of data than using standard atmospheric correction processors. In the Arctic Ocean where operational satellite ocean color data have relatively low space-time resolution, Polymer algorithms are promising candidates in enlarging satellite ocean color data sets, e.g., for Sentinel-3/OLCI satellite sensor, given more validation activities are performed in the future.

In the second part of the thesis, the underway flow-through AC-S system was set up onboard *R.V. Polarstern* during the Fram Strait cruise PS107, in addition to PS93.2 and PS99.2. AC-S derived hyperspectral $a_p(\lambda)$ were matched with HPLC pigments data. In total, 298 $a_p(\lambda)$-pigments match-ups were used as the pigment retrieval data set. Two pigment retrieval algorithms, Gaussian decomposition (Chase et al., 2013) and the singular value decomposition combined with non-negative least squares (SVD-NNLS) inversion method (Moisan et al., 2011) were compared and optimized for estimating various phytoplankton pigments or pigment groups from the $a_p(\lambda)$ data. The Gaussian decomposition method provides good estimates (median absolute percentage error, MPE 21-34%) of Chl a, chlorophyll b, chlorophyll c_1 and c_2, photosynthetic carotenoids and photoprotective carotenoids (PPC). This method outperformed the SVD-NNLS method in retrieving chlorophyll b, chlorophyll c_1 and c_2, photosynthetic carotenoids, and PPC. However, SVD-NNLS enables robust retrievals of specific carotenoids (MPE 37-65%), i.e., fucoxanthin, diadinoxanthin and $19^{'}$-hexanoyloxyfucoxanthin, which is currently not accomplished by Gaussian decomposition. More robust predictions are obtained using the Gaussian decomposition method when the observed spectral phytoplankton absorption coefficient ($a_{ph}(\lambda)$) is normalized by the package effect index at 675 nm. The latter is determined as a function of "packaged" $a_{ph}(675)$ and Chl a concentrations, which shows potential for improving pigment retrieval accuracy by the combined use of $a_{ph}(\lambda)$ and Chl a data. Both approaches provide useful information on pigment distribution, and hence phytoplankton community composition indicators, at a spatial resolution much finer than can be achieved with discrete HPLC samples.

Xanthophyll pigments provide one of the most important photoprotective mechanisms to dissipate the excess light energy and prevent photoinhibition. In the third part of the thesis, phytoplankton growth models of Geider et al. (1998), the Geider model, and Marshall et al. (2000), the Marshall model, were implemented into REcoM2 to predict the photoprotective needs of phytoplankton and their attributions from phytoplankton PPC, physiological state, and community composition. Assume that photoinhibition is negligible in phytoplankton communities acclimated to ambient light (Cullen et al., 1992). The difference between the photosynthesis–irradiance (P-E) curves with (Marshall) and without photoinhibition (Geider) is considered a measure of photoprotective needs in order to minimize such photoinhibition. The degree of phytoplankton

photoprotection is represented by 1-$\bar{\alpha}_{NP}/\bar{\alpha}$, where $\bar{\alpha}_{NP}$ and $\bar{\alpha}$ are model predicted initial slopes of the P-E curves of the Marshall and Geider models, respectively. It was then related to the HPLC PPC/Chl a data, producing a 4-D global map of PPC/Chl a estimates. These estimates were in agreement with field observations in most of the surface ocean, at depth and even across seasons, suggesting the role of PPC in photoprotective activities in the global ocean. However, at higher latitudes, discrepancies between predictions and observations suggested PPC content was insufficient to satisfy phytoplankton protective needs the community and thus other mechanisms of non-photochemical quenching were relevant. Furthermore, at higher latitudes, changes in PPC content can result from both physiological acclimation and shifts in community composition while in the rest of ocean taxonomic changes played a main role.

A comprehensive view of the phytoplankton community pigment signature is crucial for modeling the coupling of light absorption to carbon fixation in the ocean. Future validation of the above model can use the combined HPLC observations and pigment estimates from underway flow-through AC-S system. Furthermore, this work provides insights on how much of the variability in community PPC ratios is attributable to changes in community composition or changes in physiological state. This may allow an improvement of the match between satellite ocean color data and the underlying phytoplankton community. In addition, these insights may contribute to a better understanding of the effect of phytoplankton photoacclimation on the accuracy of satellite ocean color products in the Arctic Ocean.

Zusammenfassung

Der vom Menschen verursachte Klimawandel wird in der Arktis amplifiziert. Die Ursache für diese Amplifikation sind Veränderungen der Lufttemperatur und des Meereises. Die Meereisbedeckung im Arktischen Ozean geht dramatisch zurück. In der untersuchten Region dieser Arbeit, der Framstraße (dem größten und einzigen tiefen Zugang zum Arktischen Ozean) und ihrer Umgebung, wurden aufgrund der Erwärmung der Arktis Veränderungen der Meereisverhältnisse und der Wassertemperaturen beobachtet. Dies hat Folgen für die Phytoplankton Populationen. Phytoplankton ist einer der wichtigsten Primärproduzenten im Arktischen Ozean. Die durch die Erwärmung der Arktis verursachten Veränderungen des Licht- und Nährstoffregimes wirken sich auf die Saisonalität, die Biomasse, die Zusammensetzung und die Verteilung des Phytoplanktons aus. Die Biomasse und die Zusammensetzung der Phytoplankton-Artengemeinschaft lassen sich häufig anhand der Pigmentzusammensetzung und -konzentration bestimmen. Um die Reaktion des Phytoplanktons auf den Klimawandel zu untersuchen, sollen in dieser Arbeit die Phytoplanktonpigmente anhand von Beobachtungen eines AC-S-Durchfluss-Spektralphotometers an Bord des Schiffes und des Regulated Ecosystem Model Version 2 (REcoM2) (Hohn, 2008; Schartau et al., 2007), das mit Phytoplankton-Wachstums- und Photoinhibitionsmodellen implementiert wurde, geschätzt werden.

Im ersten Teil der Arbeit wurde ein Durchfluss-AC-S-System an Bord der *R.V. Polarstern* während zweier Fahrten in der Framstraße, PS93.2 und PS99.2, aufgebaut. Der hyperspektrale Partikelabsorptionskoeffizient ($a_p(\lambda)$) wurde aus den AC-S-Messungen an Bord abgeleitet. Die Höhe der partikulären Absorptionslinie bei 676 nm, die aus $a_p(\lambda)$ berechnet wurde, wurde empirisch mit den Chlorophyll a-Konzentrationen (Chl a) der Hochleistungsflüssigkeitschromatographie (HPLC) für PS93.2 bzw. PS99.2 in Beziehung gesetzt. Beide Beziehungen wurden auf hochfrequente (4 Hz) AC-S-Daten angewandt, um die Chl a-Konzentrationen entlang der Fahrtrouten abzuschätzen. Insgesamt wurden 24424 und 16110 Chl a-Datenpunkte für PS93.2 bzw. PS99.2 erzeugt. Die rekonstruierten AC-S-Chl a-Datensätze wurden zur Bewertung von sieben Satelliten-Chl a-Algorithmen verwendet. Die Anzahl der AC-S-Satelliten-Kolokalisierung ist um mehr als eine Größenordnung größer als die der HPLC-Satelliten-Kolokalisierung. AC-S-Satelliten-Abgleiche zeigen, dass alle Algorithmen durch eine Überschätzung des Satelliten-Chl a gekennzeichnet waren. Zwei Algorithmen, die auf Polymer-Atmosphärenkorrekturprozessoren basieren, erzeugten Datenprodukte mit relativ hoher Schätzgenauigkeit und geringem Fehler. Die Polymer-Atmosphärenkorrekturprozessoren berücksichtigen in ihren Reflexionsmodellen Sonnenreflex und dünne Wolken, um die atmosphärisch korrigierte Fernerkundungsreflexion abzuleiten, was eine viel größere räumliche Abdeckung der Daten ermöglicht als die Verwendung von Standard-Atmosphärenkorrekturprozessoren. Im Arktischen Ozean, wo die operationellen Satelliten-Ozeanfarbdaten eine relativ geringe räumliche und zeitliche Auflösung haben, sind Polymer-Algorithmen vielversprechende Kandidaten für die Erweiterung von Satelliten-Ozeanfarbdatensätzen, z.B. für den Sentinel-3/OLCI-Satellitensensor, vorausgesetzt, es werden in Zukunft mehr Validierungsaktivitäten durchgeführt.

Im zweiten Teil der Arbeit wurde das Durchfluss-AC-S-System an Bord der *R.V. Polarstern* während der Framstraßen Expedition PS107, zusätzlich zu PS93.2 und PS99.2, eingesetzt. Die via AC-S abgeleiteten hyperspektralen $a_p(\lambda)$ wurden mit HPLC-Pigmentdaten abgeglichen. Insgesamt wurden 298 $a_p(\lambda)$-Pigmentabgleiche als Datensatz für die Pigmentbestimmung verwendet. Zwei Algorithmen zum Pigmentabruf, die Gauß-Zerlegung (Chase et al., 2013) und die Methode der Singulärwertzerlegung in Kombination mit der Inversion der nicht-negativen kleinsten Quadrate (SVD-NNLS) (Moisan et al., 2011) wurden verglichen und für die Schätzung verschiedener Phytoplanktonpigmente oder Pigmentgruppen aus den $a_p(\lambda)$-Daten optimiert. Die Gauß-Zerlegungsmethode liefert gute Schätzungen (median absoluter prozentualer Fehler, MPE 21-34 %) für Chl a, Chlorophyll b, Chlorophyll c1 und c2, photosynthetische Carotinoide und photoprotektive Carotinoide (PPC). Diese Methode übertraf die SVD-NNLS-Methode bei der Ermittlung von Chlorophyll b, Chlorophyll c1 und c2, photosynthetischen Carotinoiden und PPC. SVD-NNLS ermöglicht jedoch robuste Vorhersagen für bestimmte

Carotinoide (MPE 37-65 %), d. h. Fucoxanthin, Diadinoxanthin und $19'$-Hexanoyloxyfucoxanthin, was derzeit mit der Gauß-Zerlegung nicht geleistet werden kann. Zuverlässigere Vorhersagen werden mit der Gauß-Zerlegungsmethode erzielt, wenn der beobachtete spektrale Phytoplankton-Absorptionskoeffizient ($a_{ph}(\lambda)$) durch den Paketeffekt-Index bei 675 nm normalisiert wird. Letzterer wird als Funktion der "verpackten" $a_{ph}(675)$- und Chl a-Konzentrationen bestimmt, was ein Potenzial zur Verbesserung der Genauigkeit der Pigmentbestimmung durch die kombinierte Verwendung von $a_{ph}(\lambda)$- und Chl a-Daten offenlegt. Beide Ansätze liefern nützliche Informationen über die Pigmentverteilung und damit Indikatoren für die Zusammensetzung der Phytoplankton-Gemeinschaft mit einer räumlichen Auflösung, die viel feiner ist als die, die mit diskreten HPLC-Proben erreicht werden kann.

Xanthophyll-Pigmente sind einer der wichtigsten photoprotektiven Mechanismen, um die überschüssige Lichtenergie abzuführen und Photoinhibition zu verhindern. Im dritten Teil der Arbeit These wurden die Phytoplankton-Wachstumsmodelle von Geider et al. (1998), das Geider-Modell, und Marshall et al. (2000), das Marshall-Modell, in REcoM2 implementiert, um den Photoprotektionsbedarf des Phytoplanktons und dessen Zuordnung zu Phytoplankton-PPC, physiologischem Zustand und Gemeinschaftszusammensetzung vorherzusagen. Es wird angenommen, dass die Photoinhibition in Phytoplanktonartengemeinschaften, die an das Umgebungslicht akklimatisiert sind, vernachlässigbar ist (Cullen et al., 1992). Die Differenz zwischen den Photosynthese-Strahlungsstärke-Kurven (P-E) mit (Marshall) und ohne Photoinhibition (Geider) gilt als Maß für den Photoprotektionsbedarf, um eine solche Photoinhibition zu minimieren. Der Grad der Photoprotektion des Phytoplanktons wird durch $1\text{-}\bar{\alpha}_{NP}/\bar{\alpha}$ dargestellt, wobei $\bar{\alpha}_{NP}$ und $\bar{\alpha}$ die modellmäßig vorhergesagten anfänglichen Steigungen der P-E-Kurven der Marshall- bzw. Geider-Modelle sind. Anschließend wurde sie mit den HPLC-PPC/Chl a-Daten in Beziehung gesetzt, wodurch eine 4-D-Global-Karte der PPC/Chl a-Schätzungen entstand. Diese Schätzungen stimmten mit den Feldbeobachtungen im größten Teil des Oberflächenozeans, in der Tiefe und sogar über die Jahreszeiten hinweg überein, was auf die Rolle von PPC bei den photoprotektiven Aktivitäten im globalen Ozean hindeutet. In höheren Breitengraden jedoch deutete die Diskrepanz zwischen Vorhersagen und Beobachtungen darauf hin, dass der PPC-Gehalt nicht ausreichte, um die Lichtschutzbedürfnisse der Phytoplanktonartengemeinschaft zu befriedigen, und dass daher andere Mechanismen der nicht-photochemischen Löschung von Bedeutung waren. Darüber hinaus können in höheren Breitengraden Veränderungen des PPC-Gehalts sowohl auf eine physiologische Akklimatisierung als auch auf Verschiebungen in der Zusammensetzung der Gemeinschaft zurückzuführen sein, während im übrigen Ozean taxonomische Veränderungen die Hauptrolle spielen. Ein umfassender Überblick über die Pigmentsignatur der Phytoplanktonartengemeinschaft ist für die Modellierung der Kopplung von Lichtabsorption und Kohlenstofffixierung im Ozean von entscheidender Bedeutung. Für die künftige Validierung des obigen Modells können die kombinierten HPLC-Beobachtungen und Pigmentschätzungen aus dem AC-S-Durchflusssystem verwendet werden. Darüber hinaus gibt diese Arbeit Aufschluss darüber, inwieweit die Variabilität der PPC-Verhältnisse in der Gemeinschaft auf Veränderungen in der Zusammensetzung der Gemeinschaft oder auf Veränderungen des physiologischen Zustands zurückzuführen ist. Dies könnte eine Verbesserung der Kolokalisierung zwischen den Farbdaten der Satelliten und der zugrunde liegenden Phytoplanktonartengemeinschaft ermöglichen. Darüber hinaus können diese Erkenntnisse zu einem besseren Verständnis der Auswirkungen der Photoakklimatisierung des Phytoplanktons auf die Genauigkeit von Satelliten-Ozeanfarbprodukten im Arktischen Ozean beitragen.

Chapter 2

Introduction

2.1 Phytoplankton

The term *phytoplankton* comes from the Greek words *phyton*, meaning 'plant', and *planktos*, meaning 'to wander'. In 1887, Victor Hensen, the first quantitative plankton ecologist, coined the term *plankton* (Hensen, 1887). Phytoplankton refer to the community of mostly single-celled, photosynthetic organisms that drift with water currents in aquatic ecosystems. Whereas most of the phytoplankton species are solo dwellers, i.e., it is only the single cell that interacts with its environment, some are capable of forming multicellular filaments, coenobia or mucilaginous colonies (Reynolds, 2006).

Phytoplankton are taxonomically diverse. In the contemporary ocean, $\sim 2 \times 10^4$ morphologically distinct species are distributed among at least eight taxonomic divisions or phyla (Falkowski et al., 2004). Many morphologically identical species are genetically distinct (e.g., Liu et al., 2009). This diverse group of organisms include both prokaryotes and eukaryotes. Cyanobacteria are the only extant prokaryotic phytoplankton that existed $\sim$ 2.5 billion years ago (Falkowski, 2014). These prokaryotes numerically dominate the phytoplankton community in marine ecosystems. All other groups are eukaryotic. They evolved more than 1.5 billion years ago and are thought to trace their lineages back to the endosymbiotic appropriation of a cyanobacterium into a heterotrophic host cell, where the appropriated cyanobacterium became a chloroplast (Delwiche, 1999; Falkowski et al., 2004). Subsequent evolution moved into different directions and gave rise to the green and red lineages. The green lineage distinguishingly features the synthesis of chlorophyll b as a secondary pigment and is dominated by green algae in the oceans. The red lineage comprises a diverse set of phytoplankton that synthesizes chlorophyll c as secondary pigments. Of the red-lineage phytoplankton groups, diatoms, dinoflagellates, haptophytes (including the coccolithophores), and chrysophytes are the most important (Falkowski, 2014).

Despite their deep diversity, phytoplankton share defining characteristics. Notably, they live in the upper illuminated layer of the ocean and carry out oxygenic photosynthesis. Oxygenic photosynthesis takes place in the chloroplasts of eukaryotic phytoplankton, while cyanobacteria employ specialized photosynthetic membranes. It is a redox reaction of the general form:

$$2H_2O + CO_2 + \text{light} \longrightarrow (CH_2O) + H_2O + O_2 \tag{2.1}$$

This equation can be seen as the simplified description of two reactions: the oxidation or 'splitting' of water into molecular, gaseous oxygen using light energy, and the reduction of CO_2 to carbohydrate (CH_2O) by addition of the intermediate hydrogen atoms from water. Chlorophyll a (Chl a), the photosynthetic pigment universally distributed among phytoplankton - except for prochlorophytes which contain divinyl Chl a - catalyzes the photochemical water splitting process using energy from sunlight. In the oceans, these reactions depend on light regime and nutrient availability (e.g., nitrogen, phosphorus, iron).

The impact of oxygenic photosynthesis by phytoplankton on Earth's climate, chemical composition and ecosystems is profound. Over 2 billion years ago, the rise of oxygen produced by cyanobacteria permanently oxidized Earth's atmosphere, which ultimately permitted eukaryotes and multicellular organisms to evolve (Falkowski, 2006). The organic carbon resulting from photosynthesis fuels the growth and respiratory demands of phytoplankton and all remaining organisms in the ocean through the food webs. Hence, the rate of photosynthesis places an upper bound on the overall biomass and productivity of oceanic ecosystems and constrains their overall net biological flow of energy (Falkowski and Raven, 2013). The global fisheries catches, for example, are constrained by the net primary productivity of marine phytoplankton (Chassot et al., 2010). The latter currently amounts to $\sim$ 50 petagrams of carbon per year, accounting for $\sim$ 45% of the global net primary productivity (Field et al., 1998). While most of the net primary production is locally remineralized to inorganic nutrients through grazing, bacterial consumption and respiration, approximately one third of the organic matter in the euphotic zone sinks to the ocean interior. The exported organic carbon is subsequently oxidized and remineralized by biological activity in deeper layers to inorganic carbon in the ocean interior, with a mere 0.2% of the organic carbon reaching the sediments. The sum of biological processes leading to carbon export out of the surface layers of the ocean, the so-called biological pump, is crucial to the maintenance of atmospheric CO_2 level and thus exerts an important control on global climate. In the absence of the biological pump, models predict that the partial pressure of atmospheric CO_2 would be nearly twice as high as the current value (Maier-Reimer et al., 1996). A very small fraction of the exported organic carbon escapes the biological pump and becomes incorporated into marine sediments. All above carbon exchange processes between the ocean and atmosphere are central to the global carbon cycle. Further, given that living organisms also contain important elements such as nitrogen, phosphorus, sulfur, and in some groups also silicon as well as a suite of metals such as iron, zinc, and cobalt. Alfred Redfield (1934) observed that carbon, nitrogen, and phosphorus in phytoplankton are commonly in the molar ratio C:N:P = 106:16:1. The biological pump also influences the cycling of these elements.

Based on their properties and distinct roles in biogeochemical cycles, phytoplankton are grouped into functional types, e.g., silicifiers, calcifiers, N_2-fixers, DMSP-producers, non N_2-fixing pico-phytoplankton in earth system and coupled General Circulation-Biogeochemical Models (Le Quéré et al., 2005). Silicifiers, mainly diatoms which are encased in shells (frustules) of amorphous silicon (opal), are responsible for most of the primary production and biomass during the spring blooms in temperate and polar regions. They contribute to the vertical carbon export into the ocean interior far more effectively than phytoplankton of smaller size. The frustules of diatoms are archived in sediments and form, among others, the opal belt at the Polar Front region around Antarctica (Lisitzin, 1971). Calcifiers, primarily coccolithophores, are responsible for more than half of the marine carbonate flux in sediments (Schiebel, 2002). They precipitate calcium carbonate, affecting ocean alkalinity and carbonate chemistry and thereby atmospheric CO_2 level on geological timescales. They also produce the densest ballasts that effectively facilitate the vertical carbon export into the ocean interior (Klaas and Archer, 2002). N_2-fixers comprise cyanobacteria, such as Trichodesmium, capable of reducing atmospheric N_2 to ammonium. N_2 fixation is the major nitrogen source supporting primary production throughout most of the world's oceans and influences CO_2 air-sea exchange (Falkowski, 1997; Gruber and Galloway, 2008). DMSP-producers, mainly haptophytes such as coccolithophores and *Phaeocystis*, produce dimethylsulfonium propionate (DMSP). DMSP released in seawater by phytoplankton is converted to the gas dimethyl sulfid that subsequently diffuses into the atmosphere, affecting the atmospheric sulfur cycle and possibly cloud formation, with direct impacts on the climate (Lovelock et al., 1972; McCoy et al., 2021). Non N_2-fixing pico-phytoplankton, in particular *Synechococcus* and *Prochlorococcus*, make a considerable contribution to primary production in oligotrophic regimes, but a negligible contribution to vertical carbon export. The functionality of phytoplankton is often related to their size. Therefore, phytoplankton functional types may be approximated using a size-based classification (e.g., Uitz et al., 2006).

Phytoplankton functional diversity respond to environmental changes. Currently, our planet is undergoing significant and accelerated climate change due to human activities. Human-induced warming reached $\sim$ 1°C above pre-industrial levels in 2017, increasing at a rate of $\sim$ 0.2°C per decade (IPCC, 2018). The global ocean has taken up more than 90% of the excess heat in the climate system (IPCC, 2019). The resulting warming of the oceans influence light (e.g., through sea ice melting and increased vertical stratification) and nutrient availability (e.g., by increasing seawater density stratification) and thereby phytoplankton seasonality, productivity and species composition. Continued uptake of excess atmospheric CO_2 by the oceans causes further ocean acidification, which may have important consequences for the growth and calcification rates of coccolithophores (Zeebe and Wolf-Gladrow, 2001). To better understand the responses of phytoplankton and feedbacks to a changing climate, it is of great importance to determine information on their biomass, net productivity and composition (taxonomic and functional types) globally and at relevant temporal and spatial scales.

Optical methods, such as ocean color radiometry and spectrophotometry, provide a useful tool for obtaining such information, both *in situ* and globally from space using Earth-observing satellites. The interactions between light and phytoplankton, including absorption of light by various pigments, form the basis of these methods.

2.2 Phytoplankton pigments

Phytoplankton have evolved a wide variety of cellular pigments. There are three chemically distinct types of pigments: chlorophylls, carotenoids and phycobiliproteins. All phytoplankton contain chlorophylls and carotenoids; cyanobacteria, rhodophytes, and cryptophytes also contain phycobiliproteins.

During oxygenic photosynthesis, phytoplankton pigments are responsible for the absorption of light energy and the primary steps in its conversion to chemical bond energy. They are bound in various combinations to specific proteins to form pigment-protein complexes and localized in the photosynthetic membranes. Depending on their roles in photosynthesis, phytoplankton pigments can be classified as photosynthetic reaction center pigments (e.g., Chl a, divinyl Chl a), light-harvesting pigments (e.g., chlorophylls b and c, phycobiliproteins, fucoxanthin, peridinin, 19'-hexanoyloxyfucoxanthin [Hex], 19'-butanoyloxyfucoxanthin), and photoprotective carotenoids (PPC, e.g., alloxanthin, diadinoxanthin, diatoxanthin, zeaxanthin, carotenes). Light-harvesting pigments act as antennas and capture solar radiation from different spectral wavelengths incident on phytoplankton cells. Upon absorption of light by a pigment molecule, light energy is transferred to chlorophyll molecules in the photosynthetic reaction centers where the oxidation of water (Equation 2.1) is induced. Under excess light conditions, PPC reduce the production of damaging reactive oxygen species, thus protecting pigments, proteins and lipids from oxidation and minimizing the ensuing reduction in photosynthesis rate. Despite the broad range in pigments produced by phytoplanktonic organisms, only Chl a - and divinyl Chl a in the case of prochlorophytes - are directly involved in photosynthesis. All other pigments, grouped under the term accessory pigments, are not essential for photosynthesis to occur: they assist in the acquisition of light energy and are part of a mechanism that allows photosynthesis to happen in less than optimum light conditions. Therefore, their composition and concentrations depend on environmental factors (e.g., light) and also on phytoplankton taxonomic affiliations.

Naturally, one application of pigment data is the study of phytoplankton photosynthetic response to changing light environments. For example, variations in certain pigment ratios, such as (diadinoxanthin + diatoxanthin)/Chl a, diadinoxanthin/Chl a, diatoxanthin/Chl a, diatoxanthin/(diatoxanthin + diadinoxanthin), have been used to infer photoacclimation at different time scales (Brunet et al., 2011). In my study, as published in Álvarez et al. (2019), the ratio PPC/Chl a was found to mediate the long-term photoprotective response of phytoplankton throughout most of the oceans. Pigment composition and concentrations also provide important information on phytoplankton biomass and broad taxonomic and functional diversity (Higgins et al., 2011). As a matter of empirical and operational convenience, Chl a concentrations are widely used as a proxy for phytoplankton biomass and hence also as an input parameter for many local and global primary productivity models (Behrenfeld and Falkowski, 1997). While most pigments are distributed across several taxonomic groups, certain key pigments are signatures for some phytoplankton groups. For example, fucoxanthin is abundant in diatoms, alloxanthin in cryptophytes, peridinin usually indicates the presence of dinoflagellates, Hex is associated with haptophytes and peridinin-lacking dinoflagellates, chlorophyll b is dominant in green algae, divinyl chlorophylls a and b are exclusively present in prochlorophytes, while phycobiliproteins and zeaxanthin are useful markers for cyanobacteria. Based on marker pigments, algorithms - such as CHEMTAX (Mackey et al., 1996) and Diagnostic Pigment Analysis (Vidussi et al., 2001) - have been developed to estimate the fractional biomass of different phytoplankton groups in water samples. In my study (Álvarez et al., 2019), the spatial and temporal distribution of the ratio PPC/Chl a was modelled for the global ocean and its variability due to the changes in phytoplankton community composition and physiological acclimation was analysed.

Phytoplankton pigments absorb light in mainly two regions of the visible spectrum (roughly 400-700 nm) (Figure 2.1): the blue and red bands. All pigments except phycobiliproteins absorb blue light. Red light, on the other hand, is absorbed only by chlorophylls, mainly Chl a. Chl a has two dominant peaks in an absorption spectrum: a primary peak at 440 nm and secondary one at 675 nm. Light from the middle regions of the visible spectrum is absorbed to a small extent by carotenoids. The absorption bands of carotenoids extend farther in the long wavelengths, while chlorophylls extend to the shortwave bands. The phycobiliproteins, however, absorb light more intensively in the middle regions (Figure 2.1). The composition and concentrations of pigments influence the shape and magnitude of the spectral absorption coefficients of phytoplankton ($a_{ph}(\lambda)$), one of the fundamental inherent optical properties (IOPs) of seawater, thus altering the appearance of ocean color. This allows the development of absorption spectral-based remote-sensing algorithms for distinguishing phytoplankton functional types from optical satellites.

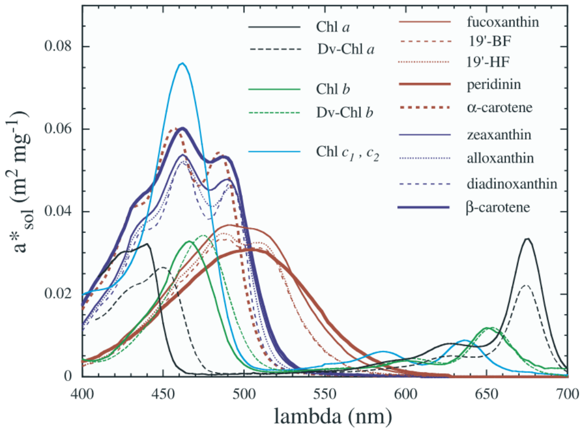


FIGURE 2.1: Mass-normalized absorption spectra of phytoplankton pigments based on the study by Bricaud et al. (2004).

2.3 Phytoplankton from space

Satellite ocean color remote sensing revolutionized ocean observing capabilities and has been in service for almost five decades. It can provide basin scale coverage of the upper mixed layer and investigate biogeochemical processes on monthly, inter-annual to annual timescales, which keeps stimulating interest among the ocean science community in deriving ocean biological and biogeochemical data products.

Ocean color remote sensing began with the launch of the proof-of-concept radiometer Coastal Zone Color Scanner (CZCS) aboard the Nimbus-7 satellite in 1978. Its initial objective was to quantify total phytoplankton pigment, defined as the sum of Chl a and phaeophytin, in coastal oceans (Gordon et al., 1983). Inadvertently, CZCS data revealed unprecedented information of open ocean mesoscale structures and the seasonal and inter-annual variability of total phytoplankton pigment (see the Journal of Geophysical Research special issue *Ocean Color From Space: A Coastal Zone Color Scanner Retrospective*), which greatly transcended the limits of *in situ* observations and opened up a new perspective in marine ecology and biogeochemistry. As a result, the objective of follow-on ocean color satellite missions, characterised by one-off multispectral radiometers such as SeaWiFS, MODIS, and MERIS, was expanded to routine global observations over an extended period of time. Additional spectral bands were incorporated to transition from total pigment mainly to Chl a and to a smaller extent to certain accessory pigments (e.g., phycobiliproteins). Data products have been used to address a wide range of scientific questions including the global decadal trends in primary productivity and the distributions of phytoplankton functional types (see McClain (2009) for a review). The new generation of satellite missions, such as Sentinel-3/OLCI (multispectral) and PACE/OCI (hyperspectral), are aimed for consistent observations of marine ecosystems to support long-time series climate data records and operational monitoring applications (see IOCCG (2012) and Groom et al. (2019) for reviews). Hyperspectral observations are more capable of discerning subtle differences in spectral absorption by various accessory pigments (e.g., chlorophylls b and c, carotenoids, phycobiliproteins) (e.g., Wang et al., 2016; Chase et al., 2017) and thereby phytoplankton functional types (see Chase (2020) for review). Table 2.1 summarizes selected polar-orbiting satellite ocean color sensor systems from the past, present and near future.

Sensor	**Satellite**	**Agency**	**Spatial resolution (km)**	**Spectral bands (400-890 nm)**	**Launch(end)**
CZCS	Nimbus-7	NASA	0.825	4	1978-(1986)
OCTS	ADEOS-1	NASDA	0.7	8	1996-(1997)
SeaWiFS	SeaStar	NASA	4; 1	8	1997-(2010)
MERIS	Envisat	ESA	1.2; 0.3	12	2002-(2012)
MODIS	Terra/Aqua	NASA	1	9	1999/2000
VIIRS	Suomi NPP	NOAA	0.75	7	2011
VIIRS	JPSS-1	NASA&NOAA	0.75	7	2017
OLCI	Sentinel 3A/B	ESA	0.3	21	2016/2018
OCI	PACE	NASA	1	hyperspectral (5 nm interval)	2022
OLCI	Sentinel 3C/D	ESA	0.3	21	>2022

TABLE 2.1: Selected past, current and planned polar-orbiting satellite sensor systems relevant for ocean color.

Ocean color radiometers mounted on satellites measure upwelling solar radiation at the top of the atmosphere. This signal can then be processed to obtain spectral water-leaving radiance (or reflectance), which is the sunlight backscattered out of the ocean after absorption and scattering by water and its constituents. The variable *ocean color* is commonly referred to as the spectral remote-sensing reflectance ($R_{rs}(\lambda)$), which is defined as the ratio of water-leaving radiance to downwelling irradiance just above the sea surface. The $R_{rs}(\lambda)$ is a function of the spectral absorption ($a(\lambda)$) and backscattering coefficients ($b_b(\lambda)$) of all constituents in seawater, including phytoplankton pigments. It can be approximated using the formula from Gordon et al. (1988).

$$R_{rs}(\lambda) = g_0\, u(\lambda) + g_1\, [u(\lambda)]^2\, ,\; u(\lambda) = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \tag{2.2}$$

where g_0 and g_1 are coefficients related to sun-sensor angular geometry (Morel and Gentili, 1993).

Phytoplankton pigments influence $R_{rs}(\lambda)$ mainly through their influence on $a(\lambda)$ and have different absorption properties from other water constituents (e.g., CDOM and non-agal particles). The algorithms to retrieve Chl a concentrations, the primary goal of satellite ocean color radiometry, make use of its distinctive absorption properties in seawater either implicitly (e.g., through empirical regression) or explicitly (e.g., through deriving $a_{ph}(\lambda)$). Standard Chl a algorithms, OCx, rely on empirical relationships that use blue-to-green band ratio of $R_{rs}(\lambda)$ (as summarized in `https://oceancolor.gsfc.nasa.gov/atbd/chlor_a/`). GSM, a semi-analytical algorithm, makes use of the known spectral shape of $a_{ph}(\lambda)$ (i.e., $a_{ph}(\lambda)$ normalised by Chl a concentrations) and inverts $R_{rs}(\lambda)$ to directly determine Chl a concentrations through optimization (Garver and Siegel, 1997). Other methods like QAA (Lee et al., 2002) and C2RCC (Doerffer and Schiller, 2007; Brockmann et al., 2016) first derive $a_{ph}(\lambda)$ from $R_{rs}(\lambda)$ in a semi-analytical way and then convert $a_{ph}(\lambda)$ to Chl a concentrations using regionally adapted bio-optical models (e.g., the power function from Bricaud et al. (1995)). In my study, different Chl a algorithms were evaluated and inter-compared for Aqua/MODIS and Sentinel-3/OLCI, as published in Liu et al. (2018).

Compared to Chl a, phytoplankton accessory pigments contribute much less to $R_{rs}(\lambda)$ signals, and their spectral differences in the contributions are subtle. Retrieving various accessory pigments, pigment groups, and single or multiple functional types indicated by their marker pigments from ocean color data is therefore more complex. Attempts have been made using spectral reflectance (e.g., $R_{rs}(\lambda)$) or absorption (e.g., $a_{ph}(\lambda)$) measurements alone or in combination. These algorithms make use of the distinctive absorption properties of different pigments (Figure 2.1) either implicitly (e.g., through empirical regression) or explicitly (e.g., through deriving the pigment-specific absorption). PHYSAT (Alvain et al., 2005, 2008), an algorithm implemented for multispectral satellite sensors SeaWiFS and MODIS, empirically relates the spectral changes of normalized water-leaving radiance, i.e., after removing the effect of Chl a, to dominant phytoplankton groups. Hyperspectral optical measurements and their derivatives have been explored to separate pigment absorption peaks and "shoulders" and quantify pigments or functional types from the magnitude of the derivative spectra or values at minimum or maximum peak wavelengths using, for example, regression analysis (e.g., Bidigare et al., 1989; Organelli et al., 2013), clustering analysis (e.g., Xi et al., 2015), similarity index (e.g., Millie et al.; Kirkpatrick et al., 2000; Craig et al., 2006; Xi et al., 2017). Spectral inversion methods explicitly use the shape of the component absorption spectra, e.g., specific absorption coefficients or Gaussian functions represented by phytoplankton pigments or functional types, as basis vectors. The $a_{ph}(\lambda)$ spectrum can be expressed as a linear combination of the products of the basis vectors and the concentrations of the components. The concentrations of phytoplankton pigments or functional types can then be solved using measured $a_{ph}(\lambda)$ or total particulate absorption spectra $a_p(\lambda)$ (e.g., Chase et al., 2013; Liu et al., 2019), or using $R_{rs}(\lambda)$ via the linkage between $a_{ph}(\lambda)$ and $R_{rs}(\lambda)$ (Equation 2.2) and by parameterizing other absorbing and scattering water constituents (e.g., Chase et al., 2017; Wang et al., 2016). In my study, two spectral inversion algorithms were compared and optimized for estimating various phytoplankton pigments, as published in Liu et al. (2019).

Algorithms for deriving phytoplankton Chl a, accessory pigments and functional types from ocean color data require *in situ* data for their development. In addition, they need a comprehensive database of colocated, concurrent satellite and *in situ* data for validation. The number of match-up data points puts a limit to the extent these algorithms can be validated and inter-compared. Therefore, efforts have been and continue to be made in improving techniques for field sampling and measurements.

2.4 Underway spectrophotometry

The hyperspectral AC-S spectrophotometer (or its former version, the 9-wavelength resolved AC-9) (Sea-Bird Scientific, Philomath, OR, USA) has been widely used for measuring *in situ* $a(\lambda)$ and beam attenuation coefficients ($c(\lambda)$) in natural waters. The AC-9 and AC-S were introduced in 1993 and 2002, respectively. Their introduction revolutionized the capabilities of measuring these two IOPs - $a(\lambda)$ and $c(\lambda)$ - *in situ* and remains industry standard today.

The AC-S concurrently measure $a(\lambda)$ and $c(\lambda)$ in two separate optical systems (Figure 2.2). They share a light source from a tungsten lamp and a rotating wheel embedded with a linear variable filter that samples over 80 wavelengths typically in the 398–740 nm spectral range (Full Width at Half Maximum: 10-18 nm, wavelength increment: $\sim$ 4 nm). A quasi-collimated light beam passes through the linear variable filter and enters the "a" and "c" flow tubes, respectively, through optical glass windows. The "a" flow tube has reflecting walls inside, with a diffuse collector at the end of the tube, in front of a large area detector. This design ensures that both forward scattered and transmitted light reach the detector, so that only absorbed and backscattered light are removed from the beam. The "c" flow tube has non-reflecting walls and a collimated, clear detector at its end. The light beam is scattered and absorbed in the tube, with transmitted and some scattered light reaching the detector's field of view. With this setting, both of the $a(\lambda)$ and $c(\lambda)$ measurements are subject to significant scattering error, in addition to the water temperature and salinity effects. Data correction methods for $a(\lambda)$ (as absorption measurements are the focus of this study) are summarized in IOCCG Protocol Series (2018).

In recent years, techniques have been developed to integrate the AC-S spectrophotometer into the flow-through system of moving ships (see IOCCG Protocol Series (2019) for review). In my study, the setup of such a system was adapted in the study region, the Fram Strait and its adjacent waters. The operational experience gained in my study

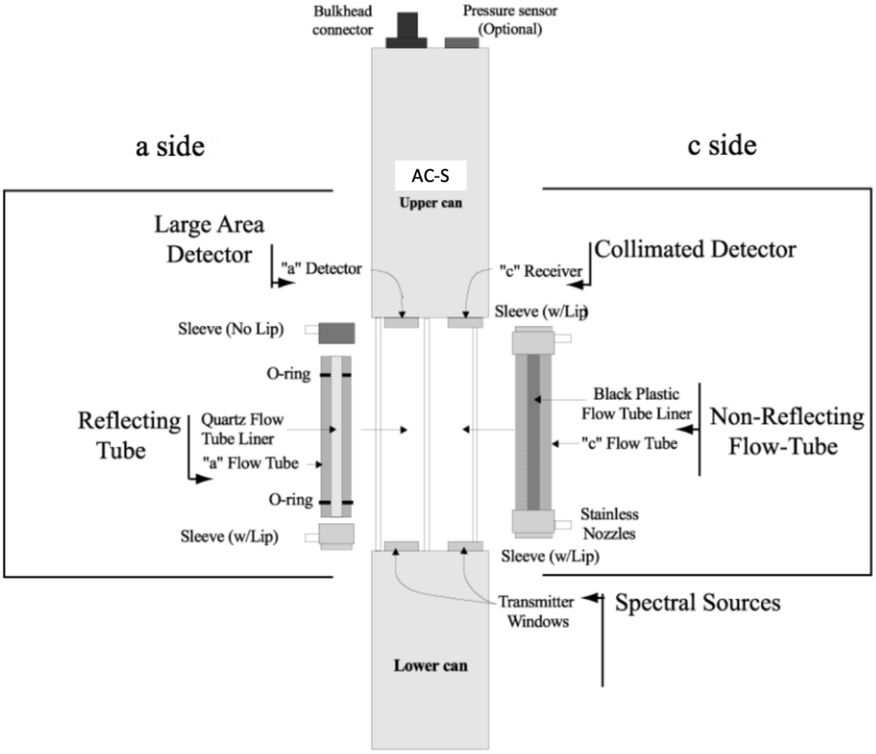


FIGURE 2.2: Schematic illustration of the AC-S spectrophotometer (adapted according to IOCCG Protocol Series (2018)).

contributed to the standard protocol of IOCCG Protocol Series (2019) as suggestions of operating such a system in polar seas. Figure 2.3 shows the setup of an underway AC-S flow-through system in my study (Liu et al., 2018). The system was operated such that within each hour, de-bubbled seawater directly entered the AC-S flow tubes for 50 minutes, and in the remaining 10 minutes it entered into a cartridge filter (pore size: 0.2 μm) beforehand and then to the AC-S flow tubes. The choice of the time slots is empirical and depends on how variable the dissolved optical properties are in the study region. A valve controller made the switch periodically. As a result, during the 50 minutes, the bulk $a(\lambda)$ and $c(\lambda)$ were measured (blue symbols in Figure 2.4), whereas during the 10 minutes, their dissolved fractions (particle size < 0.2 μm) were measured (red symbols in Figure 2.4). The $a(\lambda)$ and $c(\lambda)$ of dissolved matter in the 50 minutes were obtained through linear interpolation of two adjacent filtered seawater measurements. They were then subtracted from the bulk measurements to obtain the particulate fractions (particle size > 0.2 μm) in the time slot. After correcting for water temperature and salinity effects as well as scattering error, high quality $a_p(\lambda)$ data were output as the final data product of the underway AC-S flow-through system. From $a_p(\lambda)$ data, the concentrations of Chl a (Liu et al., 2018) and accessory pigments (Liu et al., 2019) were estimated.

A big advantage of operating AC-S in this underway setting is that the subtraction of the optical properties of 0.2-μm filtered seawater from those of the bulk greatly reduces

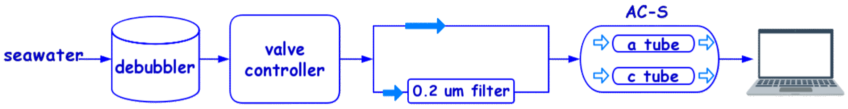


FIGURE 2.3: Schematic illustration of the underway AC-S flow-through system.

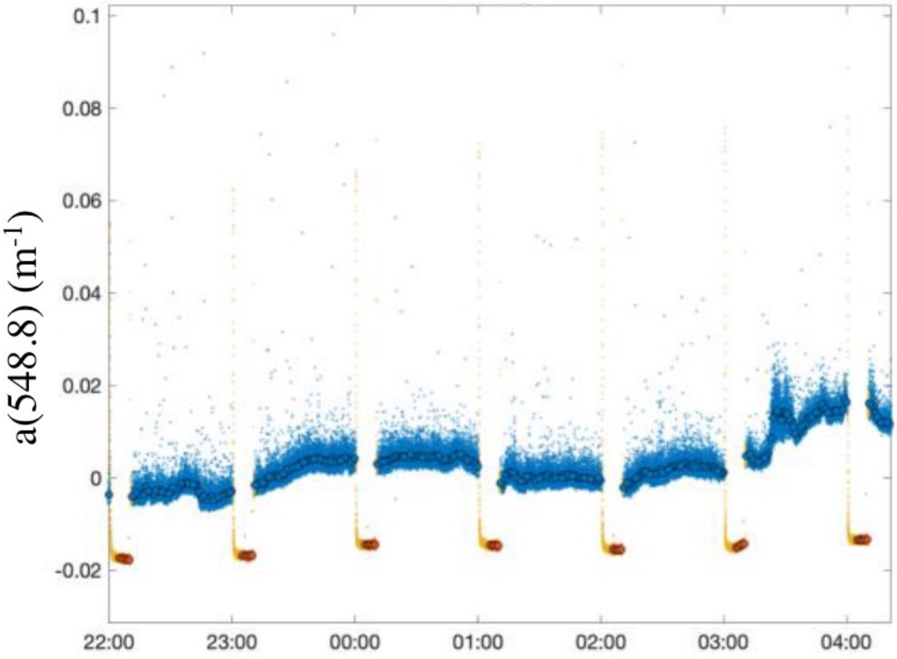


FIGURE 2.4: Example time series of absorption at 548.8 nm measured by an AC-S spectrophotometer in an underway flow-through system (IOCCG Protocol Series (2019)). Blue symbols are total seawater measurements, and red symbols are filtered seawater measurements.

the effect of instrumental drift, one of the major source of instrumental uncertainties of optical sensors like AC-S (Slade et al., 2010). The addition of a filter into the flow-through system operationally allows the separation of signals from the particulate and dissolved matters. These are much more useful than the bulk optical measurements when it comes to deriving biophysical parameters. In the traditional way of deploying AC-S without the filter, only the bulk $a(\lambda)$ and $c(\lambda)$ are measured. They can be mathematically decomposed into the contributions of particulate and dissolved matters (e.g., Bricaud and Stramski, 1990; Grunert et al., 2019), which adds methodological uncertainty.

Compared to traditional discrete water sampling technique, high data sampling frequency of AC-S (4Hz) greatly increases the number of *in situ* data points. In satellite data validation practices, many more match-up data points are produced and more than one *in situ* data points can match a satellite pixel. Given the mismatch in spatial scales between satellite and *in situ* data, averaged *in situ* data within a pixel is statistically more representative than a single data point. Moreover, the spatial variability within a satellite pixel can be addressed. In my study (Liu et al., 2018), *in situ* Chl a concentrations estimated from underway $a_p(\lambda)$ data and those measured from discrete samples were compared and used to validate Chl a products of Aqua/MODIS and Sentinel-3/OLCI generated from various algorithms.

2.5 Arctic Ocean

The Arctic is changing drastically in response to climate change and is transitioning to a warmer, less frozen region. This has consequences for terrestrial and marine ecosystems. At the root of the changes are changes in air temperatures and sea ice. The air temperatures of the Arctic are increasing more than twice as fast as the global average (Moon et al., 2021). The sea-ice cover is dramatically receding (IPCC, 2013). September sea ice extent declines at a rate of 13% per decade since 1979 (`https://climate.nasa.gov/vital-signs/arctic-sea-ice/`). The amount of multi-year ice reached the second lowest since 1985 in summer 2021 and the post-winter sea ice volume in April 2021 was the lowest since records began in 2010 (Moon et al., 2021). Sea ice decline has consequences for phytoplankton biomass (e.g., Moon et al., 2021), productivity (e.g., Moon et al., 2021), bloom timing and frequency (e.g., Kahru et al., 2011; Ardyna et al., 2014), and community composition (e.g., Ardyna et al., 2017). On the one hand, it allows more light transmission into the Arctic Ocean surface waters (Nicolaus et al., 2012). On the other hand, sea ice loss, larger river discharge, and increases in seawater temperatures enhance salinity and temperature stratification in the water column, which may limit nutrient replenishment into the sunlit surface waters. The impact of climate change on future changes in light and nutrient availability and their impact on phytoplankton are not well understood.

My study region is the Fram Strait and its adjacent waters. The Fram Strait is located between Svalbard and Greenland. It is the largest and only deep gateway to the Arctic Ocean. It is important for the mass and heat exchange between the Arctic Ocean and the rest of the world's ocean (e.g., Aagaard and Greisman, 1975; Rudels et al., 2013). Warm and saline Atlantic water flows northward in the West Spitsbergen Current while cold and fresh polar water and sea ice is transported southward in the East Greenland Current. The latter constitutes nearly all of the sea ice export from the Arctic Ocean (e.g., Kwok, 2009).

In recent decades, the Fram Strait has undergone changes in sea-ice conditions (e.g., Hansen et al., 2013; Krumpen et al., 2019) and a significant warming, high variability of Atlantic water inflow to the Arctic Ocean (Beszczynska-Möller et al., 2012). Mixing, eddies and recirculating water of the warm West Spitsbergen Current further complicate the hydrographic conditions (Walczowski, 2013; von Appen et al., 2015). As a result, alterations in light and nutrient regimes impact phytoplankton seasonality, biomass, community composition and distribution. The seasonal cycle of phytoplankton biomass has been significantly enhanced in the shallow upper water layers since 2008 (Nöthig et al., 2015). High Chl a distributions are found to be associated with different conditions of vorticity and mixed layer depth in eastern Fram Strait throughout the year 2009 (Schourup-Kristensen et al., 2021). A significant increase in summertime Chl a concentrations in the eastern Fram Strait was observed from 1991 to 2009 and a decrease since then until 2015, whereas on the western side there were minor changes (Nöthig et al., 2015, 2020). A shift of dominant phytoplankton assemblages from diatoms (mainly *Thalassiosira* spp., *Chaetoceros* spp. and *Fragilariopsis* spp.) towards coccolithophores (mainly *Emiliania huxleyi*) and more recently, *Phaeocystis* spp. (mainly *Phaeocystis pouchetii*) and other small pico- and nanoflagellates during summer months was reported (Nöthig et al., 2015; Hegseth and Sundfjord, 2008; Bauerfeind et al., 2009). Changes in phytoplankton biomass and composition impose consequences on higher trophic levels in the food web (e.g., Bauerfeind et al., 2009; Nöthig et al., 2015) and vertical particle fluxes, and thereby carbon sequestration by the biological carbon pump (von Appen et al., 2021).

2.6 Outline of the thesis

In this thesis, I aim to estimate phytoplankton pigments using observations from a shipboard flow-through AC-S system and a global biogeochemical model implemented with phytoplankton growth and photoinhibition models in the Arctic Ocean, more specifically, in the Fram Strait and its adjacent waters. Specific goals and main contents are listed as follows.

In **Paper I**, Chl a concentrations were estimated from underway AC-S data in the Fram Strait. These data were used to evaluate and inter-compare different Chl a algorithms for ocean color satellite sensors Aqua/MODIS and Sentinel-3/OLCI in the Fram Strait. This work confirms the great advantage of the underway spectrophotometry in enlarging *in situ* Chl a data sets for the Fram Strait and improving satellite Chl a validation and Chl a algorithm assessment over discrete water sample analysis.

In **Paper II**, two spectral inversion algorithms were compared and optimized for estimating various phytoplankton pigments from underway AC-S data in the Fram Strait. Both approaches provide useful information on pigment distributions, and hence, phytoplankton community composition indicators, at a spatial resolution much finer than that can be achieved with discrete samples.

In **Paper III**, the spatial and temporal distribution of the ratio of PPC to Chl a was modelled for the global ocean, with *in situ* data from the Fram Strait contributing to the evaluation of the model. Its variability due to the changes in phytoplankton community composition and physiological acclimation was analysed. This work shows that at high latitudes, both taxonomic changes and physiological acclimation determine phytoplankton pigment signature, whereas in the rest of the oceans, community composition plays a main role.

2.7 List of publications and declaration of own contribution

Paper I: **Liu, Y.**, Röttgers, R., Ramírez-Pérez, M., Dinter, T., Steinmetz, F., Nöthig, E. M., Hellmann, S., Wiegmann, S., Bracher, A. (2018). Underway spectrophotometry in the Fram Strait (European Arctic Ocean): a highly resolved Chl a data source for complementing satellite ocean color. Optics express, 26(14), A678-A696. `https://www.osapublishing.org/oe/abstract.cfm?uri=oe-26-14-A678`

I co-collected and analysed high performance liquid chromatography (HPLC) pigments data, QFT-ICAM absorption data, AC-S data and satellite Chl a data, evaluated satellite Chl a algorithms, and drafted the manuscript.

Paper II: **Liu, Y.**, Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., Pan, Y., Bracher, A. (2019). Retrieval of phytoplankton pigments from underway spectrophotometry in the Fram Strait. Remote Sensing, 11(3), 318. `https://www.mdpi.com/2072-4292/11/3/318`

I mainly designed the experiments. I co-collected and analysed pigments data, intercompared and developed pigment algorithms, and drafted the manuscript.

Paper III: Álvarez, E., Thoms, S., Bracher, A., **Liu, Y.**, Völker, C. (2019). Modeling photoprotection at global scale: The relative role of nonphotosynthetic pigments, physiological state, and species composition. Global Biogeochemical Cycles, 33(7), 904-926. `https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2018GB006101`

I co-collected and analysed HPLC pigments data, discussed with the first author about the quality control of *in situ* pigment data and the estimation of photoprotective ca-rotenoids from different models, and co-wrote the manuscript.

In addition, I published the following peer reviewed papers as a co-author within the duration of my PhD project, and listed them in the Appendix:

Xi, H., Losa, S. N., Mangin, A., Soppa, M. A., Garnesson, P., Demaria, J., **Liu, Y.**, d'Andon, O., Bracher, A. (2020). Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. Remote Sensing of Environment, 240, 111704. `https://www.sciencedirect.com/science/article/pii/S0034425720300730`

Sun, X., Shen, F., Liu, D., Bellerby, R. G., **Liu, Y.**, Tang, R. (2018). In situ and satellite observations of phytoplankton size classes in the entire continental shelf sea, China. Journal of Geophysical Research: Oceans, 123(5), 3523-3544. `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2017JC013651`

Yu, X., Shen, F., **Liu, Y.** (2016). Light absorption properties of CDOM in the Changjiang (Yangtze) estuarine and coastal waters: An alternative approach for DOC estimation. Estuarine, coastal and shelf science, 181, 302-311. `https://www.sciencedirect.com/science/article/abs/pii/S027277141630333X`

Chapter 3

Synthesis

Human-induced climate change is amplified in the Arctic. This has consequences for terrestrial and marine ecosystems. Phytoplankton are one of the main primary producers in the Arctic Ocean. Phytoplankton growth and photosynthesis and, therefore, phytoplankton biomass, productivity and functional diversity are limited by light and nutrients, which are altered by air temperature increase and sea ice decline. To study the response of phytoplankton to the changing climate, satellite ocean color data are often used. However, in the Arctic Ocean, these data have relatively lower space-time resolution than those in the low and middle latitudes due to the heavy clouds and fog, the prevailing low solar elevations and the presence of sea ice etc. (IOCCG, 2015). In addition, the validity of standard satellite ocean color algorithms is affected by, for example, terrigenous CDOM input from river discharge (e.g., Matsuoka et al., 2007, 2011) and phytoplankton photoacclimation to low irradiances that often leads to high intracellular pigment concentrations (e.g., Cota et al., 2003, 2004; Matsuoka et al., 2007, 2011).

The shipboard underway spectrophotometry technique provides a new perspective for evaluating satellite Chl a algorithms. Thanks to the underway setting, the inclusion of a 0.2-μm pore-size filter, and the high sampling frequency (4 Hz) of the AC-S spectrophotometer, this technique yields *in situ* surface Chl a data with high spatial resolution along the cruise tracks. This greatly increases the number of match-up data points between satellite and *in situ* observations compared to high performance liquid chromatography (HPLC) measurements of discrete water samples. This is addressed in **Paper I**.

In addition to Chl a concentrations, which are widely used indices of phytoplankton biomass and hence input parameters for many primary productivity models, output data of the underway AC-S system, hyperspectral $a_p(\lambda)$ data, provide estimates of surface accessory pigment concentrations, which are indicators of phytoplakton taxonomic and functional diversity. The pigment estimates are also obtained with high frequency along the cruise tracks. This is addressed in **Paper II**.

While *in situ* underway and satellite ocean color data only provides information of the surface ocean, numerical models can transcend this limit and yield data from the water column. In **Paper III**, mechanistic models for phytoplankton photoinhibition and growth are coupled with the global biogeochemical model RecoM2 to predict photoprotective needs of phytoplankton, which is in part mediated by pigment ratio PPC/Chl a in the Arctic Ocean.

The aims of this thesis are to estimate phytoplankton pigments using observations from the shipboard underway flow-through AC-S system and RecoM2 biogeochemical model implemented with phytoplankton growth and photoinhibition models in the Fram Strait and its adjacent waters.

3.1 Estimation of Chl a (Paper I)

An underway flow-through AC-S system was set up onboard *R.V. Polarstern* during two Fram Strait cruises, PS93.2 and PS99.2. Hyperspectral $a_p(\lambda)$ data was derived from the underway AC-S system with high spatial resolution ($\sim$ 300 m for one minute binned-averaged spectra when the ship is moving at $\sim$ 10 knots). Subsequently, particulate absorption line height at 676 nm calculated from $a_p(\lambda)$ was empirically related to HPLC Chl a concentrations for PS93.2 and PS99.2, respectively. Both relationships were applied to high frequency AC-S data to estimate Chl a concentrations. In total, 24424 and 16110 Chl a data points were generated for PS93.2 and PS99.2, ranging from 0.179 to 3.550 $mg\ m^{-3}$ (mean value: 1.042 $mg\ m^{-3}$) and from 0.003 to 2.701 $mg\ m^{-3}$ (mean value: 0.677 $mg\ m^{-3}$), respectively. These values are representative for the summertime Chl a variability observed in the Fram Strait (Nöthig et al., 2015).

The reconstructed *in situ* Chl a data sets from AC-S were used to evaluate the following seven satellite Chl a algorithms (Table 3.1). In addition, assessment using HPLC Chl a data were also performed for comparison. Notably, this work represented the earliest validation activity of Sentinel-3/OLCI (first OLCI sensor launched in 2016) Chl a products in the Arctic Ocean.

The number of AC-S-satellite match-ups is over one order of magnitude greater than HPLC-satellite match-ups. AC-S-satellite match-ups show that Chl a was overestimated by all satellite algorithms. Two Polymer algorithms of Sentinel-3/OLCI, Polymer-Standard and OCI-Polymer, generated data products with relatively high estimation precision and small error. The unbiased root mean square error and the mean Relative Percentage Difference are approximately 0.2 $mg\ m^{-3}$ and 140% (meaning on average, the estimated values are over twofold as the measured ones), respectively, for both algorithms when compared with AC-S derived Chl a data.

TABLE 3.1: Satellite Chl a algorithms being evaluated for the cruises PS93.2 and PS99.2.

Cruise	Satellite sensor	Atmospheric Correction	Chl a algorithm	Denotation
PS93.2	Aqua/MODIS	Standard	OCI	OCI
PS99.2	Aqua/MODIS	Standard	OCI	OCI
	Sentinel-3/OLCI	Standard	OC4	OC4-Operational
		C2RCC	C2RCC	C2RCC
		Polymer	Polymer	Polymer-Standard
		Polymer	OC4	OC4-Polymer
		Polymer	OCI	OCI-Polymer

Note: in the thesis **Polymer** is short for *POLYnomial based algorithm applied to MERIS* (Steinmetz et al., 2011).

The Polymer algorithms are of special interests in this work because they have atmospheric correction processors that incorporate the contributions of sun glint and thin clouds in their reflectance models to derive atmospheric corrected remote sensing reflectance (Steinmetz et al., 2011). This allows for a much larger spatial coverage of data than using standard atmospheric correction processors (Müller et al., 2015a,b). The number of AC-S matchups of the Polymer algorithms are two to three times greater than those by the operational OC4 algorithms of Aqua/MODIS and Sentinel-3/OLCI. In the Arctic Ocean where operational satellite ocean color data have relatively low space-time resolution, Polymer algorithms are promising candidates in enlarging satellite ocean color data sets for Sentinel-3/OLCI given more validation activities are performed in the future.

Further assessment of other types of satellite Chl a algorithms, e.g., semi-analytical methods such as QAA (Lee et al., 2002) and GSM (Doerffer and Schiller, 2007; Brockmann et al., 2016), or selection of neural network approaches based on water types (Hieronymi et al., 2017) can be exploited to obtain optimized Chl a data sets for the Fram Strait. The established bio-optical relationships between $a_p(\lambda)$ and Chl a in the Fram Strait facilitates the local application QAA for retrieving Chl a. When Polymer atmospheric correction processor is applied, high data coverage with reasonably low uncertainty can be expected.

This work also shows the capacity of the underway AC-S flow-through system in enlarging *in situ* Chl a data sets compared with HPLC discrete measurements. A more common approach of obtaining high frequency *in situ* Chl a data is to use *in situ* chlorophyll fluorometers. This approach causes data inaccuracies due to natural variations in the fluorescence to Chl a ratio (Roesler et al., 2017), the fluorescence of CDOM (e.g., Xing et al., 2017), and the non-photochemical quenching of phytoplankton (Huot and Babin, 2010). In the surface ocean where solar irradiance reaches its maximum before decreasing with depth, non-photochemical quenching can cause great underestimation of Chl a concentrations. The underway AC-S flow-through system provides a promising and even more accurate method of estimating surface Chl a data, as $a_p(\lambda)$ is one of the inherent optical properties of seawater that are not affected by the light field, in contrast to Chl a fluorescence.

3.2 Estimation of accessory pigments (Paper II)

The underway flow-through AC-S system was set up onboard *R.V. Polarstern* during the Fram Strait cruise PS107, in addition to PS93.2 and PS99.2. AC-S derived hyperspectral $a_p(\lambda)$ were matched with HPLC pigments data. In total, 298 $a_p(\lambda)$-pigments match-ups were used as the pigment retrieval data set. Two pigment retrieval algorithms, Gaussian decomposition (Chase et al., 2013) and the singular value decomposition combined with non-negative least squares (SVD-NNLS) inversion method (Moisan et al., 2011) were compared and optimized for estimating various phytoplankton pigments or pigment groups from the $a_p(\lambda)$ data. While both methods were first applied to the bio-optical data of the Fram Strait, the SVD-NNLS method is for the time applied to the underway spectrophotometry data.

The Gaussian decomposition method decomposes $a_{ph}(\lambda)$ into Gaussian functions ($a_{gaus}(\lambda)$) and correlates the amplitudes of the Gaussian functions with the concentrations of major pigment groups. The amplitude of each Gaussian function is assumed to represent the magnitude of the absorption coefficient of a specific pigment or pigment group at the Gaussian peak wavelength, based on known pigment absorption properties determined in laboratory analyses. In this work, each AC-S derived $a_p(\lambda)$ was decomposed to twelve Gaussian functions and one absorption spectrum by non-algal particles ($a_{NAP}(\lambda)$) approximated by an exponential function in the range of 400-700 nm. Eleven out of the twelve Gaussian amplitudes were correlated to HPLC pigment concentrations according to the absorption center wavelengths of the pigments (see Table 2, **Paper II**). The ($a_{gaus}(\lambda)$) that best correlates with a certain pigment was selected for predicting that certain pigment. As a result, $a_{gaus}(434)$, $a_{gaus}(660)$, $a_{gaus}(638)$, $a_{gaus}(523)$ and $a_{gaus}(492)$ were used to predict the concentrations of Chl a, chlorophyll b, chlorophyll c_1 and c_2, photosynthetic carotenoids and PPC, respectively (median absolute percentage error, MPE 20.8-34.0%). More robust predictions were obtained when $a_{ph}(\lambda)$, calculated as the difference between $a_p(\lambda)$ and $a_{NAP}(\lambda)$, was normalized by the package effect index at 675 nm. The latter is determined as a function of "packaged" $a_{ph}(675)$ and Chl a concentrations, which shows potential for improved pigment retrieval accuracy by the combined use of $a_{ph}(\lambda)$ and Chl a data.

The SVD-NNLS inversion method assumes that $a_{ph}(\lambda)$ can be reconstructed from the linear combination of pigment-specific absorption coefficients multiplied by corresponding pigment concentrations (Bidigare et al., 1987). In this work, AC-S $a_{ph}(\lambda)$ was calculated as mentioned above. This technique involves a first inversion of the observed pigment concentrations that derives pigment-specific absorption spectra using SVD and a second inversion of these derived pigment-specific absorption spectra that solves for pigment concentrations using NNLS. In the first inversion, to minimize the ill-conditioning of the matrix originated from the multicollinearity of phytoplankton pigment concentrations, thresholds of condition numbers and similarity indices together with sensitivity analyses with data perturbations were proposed to determine the number and types of pigments that should be inverted for. The SVD-NNLS method exhibited stable prediction accuracy (MPE 16-65%) for six types of pigments, i.e., Chl a, chlorophyll b, chlorophyll c_1 and c_2, diadinoxanthin, fucoxanthin, and Hex. When considering the package effect normalization, additional estimations of 19'-butanoyloxyfucoxanthin (MPE 67-70%) and peridinin (MPE 68-75%) were achieved.

Grouped photosynthetic carotenoids and PPC can also be retrieved using the SVD-NNLS method, but with less accuracy. On the other hand, SVD-NNLS enables robust retrievals of specific carotenoids (MPE 37–65%), i.e., fucoxanthin, diadinoxanthin and Hex, which is currently not accomplished by Gaussian decomposition.

This work improved the Gaussian decomposition algorithm from Chase et al. (2013) by considering pigment package effect and reconsidered the SVD-NNLS inversion technique from Moisan et al. (2011) by taking into account matrix conditioning.

When applying the two methods to all the AC-S data, accessory pigments can be estimated along the cruise tracks. This pigment information can be used to indicate phytoplankton community composition when verified by microscopic and flow cytometric techniques. It can support the evaluation of ocean color algorithms and coupled hydrodynamic-biological modelling. With the advancement of hyperspectral radiometers, these algorithms have great potential to be incorporated into the inversion of satellite ocean color measurements (e.g., PACE/OCI mission) for the synoptic detection of phytoplankton pigments and thus the monitoring of phytoplankton spatial and temporal dynamics. Thus, enhanced understanding in the responses of phytoplankton community composition and physiology to climate change can be expected.

3.3 Estimation of photoprotection (Paper III)

Phytoplankton growth models of Geider et al. (1998), the Geider model, and Marshall et al. (2000), the Marshall model, were implemented into the Regulated Ecosystem model version 2 (REcoM2) (Hohn, 2008; Schartau et al., 2007) to predict the photoprotective needs of phytoplankton and their attributions from phytoplankton PPC, physiological state, and community composition.

The Geider model describes phytoplankton photoacclimation via the changes in Chl a content. Within the model, the initial slope of the photosynthesis–irradiance (P-E) curve, α, is set constant and no other photoinhibitory parameter is considered. That is, photoinhibition does not take place in this model. By contrast, the Marshall model accounts for changes in Chl a-specific absorption cross-section and quantum yield of photosynthesis driven by the relative amount of active Photosystem II. Decreases in absorption cross-section decrease the initial slope of the P-E curve, α_{NP}. α_{NP} is generally described as the light limited slope of the P-E curve since under light saturated conditions photosynthesis is limited by dark reactions. The net result of the Marshall model is making α_{NP} variable along the P-E curve. The decrease in α_{NP} under high light conditions reverts the limit for photosynthesis from dark to light reactions, which eventually leads to photoinhibition. Assume that photoinhibition is negligible in phytoplankton communities acclimated to ambient light (Cullen et al., 1992). The difference between the P-E curves with (Marshall) and without photoinhibition (Geider) is considered a measure of photoprotective needs in order to minimize such photoinhibition. The difference is given by the difference between α and α_{NP}. This difference reflects a photoprotection gap that could be filled by a variable pool of photoprotective pigments or by other non-photochemical quenching mechanisms.

The REcoM2 model describes the dynamics of two phytoplankton types, diatoms and non-diatoms, that have group-specific photoprotective needs. The rest of the ecosystem is completed with zooplankton, detritus, and main nutrients compartments. Ocean circulation and mixing is derived from the MIT general circulation model (Hauck et al., 2013).

The coupled REcoM2 model was run in a nearly global model configuration from 80°S to 80°N on a horizontal 2° x 2° grid in the Northern Hemisphere and 2° x 2° times the cosine of the latitude in the Southern Hemisphere, with 30 depth layers (0 to 5,700 m). The model was spun up for 4 years and analyzed for the next fifth year in a 10-daily temporal resolution. Model output provided physical variables, temperature and average light, and biological variables, phytoplankton groups, nutrient limitation, Chl a, carbon, α (denoted as $\bar{\alpha}$) and α_{NP} (denoted as $\bar{\alpha}_{NP}$). Each output variable was averaged to a global 2° x 2° grid, within the 33 depth layers (0 to 5,750 m) and over 12 months. This resulted in a 4-D array per variable that had a common spatial and temporal resolution to be compared to observations (180 longitude x 90 latitude x 33 depth x 12 time).

The degree of phytoplankton photoprotection is represented by $1-\bar{\alpha}_{NP}/\bar{\alpha}$. It was then related to the HPLC PPC/Chl a data, producing a 4-D global map of PPC/Chl a estimates. These estimates were in agreement with field observations in most of the surface ocean, at depth and even across seasons, suggesting the role of PPC in photoprotective activities in the global ocean. However, at higher latitudes, discrepancies between predictions and observations suggested PPC content was insufficient to satisfy phytoplankton protective needs of the community and thus other mechanisms of non-photochemical quenching were relevant.

Variations in community PPC content can be driven by physiological acclimation as well as shifts in community composition. The coupled model output provided a full description of phytoplankton groups as well as group-sepcific PPC/Chl a, allowing the discernment of respective effects of intragroup photoacclimation and community composition. At higher latitudes, changes in PPC content can result from both effects while in the rest of ocean taxonomic changes played a main role.

This work potentially provides a comprehensive view of the phytoplankton community pigment signature, which is crucial for modeling the coupling of light absorption to carbon fixation in the ocean. Future validation of the model can use the combined HPLC observations and pigment estimates from underway flow-through AC-S system (results of **Paper I** and **II**).

In remote sensing applications, phytoplankton pigment databases have been extensively used to develop, validate, or refine bio-optical algorithms for estimating phytoplankton functional types via CHEMTAX (Mackey et al., 1996) and Diagnostic Pigment Analysis (Vidussi et al., 2001). This work provides insights on how much of the variability in community PPC ratios is attributable to changes in community composition or changes in physiological state. This may allow an improvement of the match between satellite ocean color data and the underlying phytoplankton community. In addition, these insights may contribute to a better understanding of the effect of phytoplankton photoacclimation on the accuracy of satellite ocean color products in the Arctic Ocean.

3.4 Discussion & Outlook

Phytoplankton pigments estimated from the shipboard underway flow-through AC-S system and phytoplankton groups estimated from the biogeochemical model RecoM2 provide information on phytoplankton diversity. The RecoM2 describes the dynamics of two phytoplankton types, diatoms and nondiatoms. The difference between groups is merely functional, as diatoms require silica and other phytoplankton do not. As for the pigments, there are around ten common marker pigments (see 2.2) that indicate the potential presence of certain phytoplankton groups. However, differentiating phytoplankton groups by marker pigments can be problematic, as there is substantial variability in pigment concentrations as a function of physiological responses to the environmental condition. In addition, a given marker pigment can be present in several phytoplankton groups (e.g., fucoxanthin in diatoms and haptophytes). Therefore, pigment derived phytoplankton diversity needs to be verified by microscopic data. With microscopy one is able to identify many phytoplankton to the species level, while to date, numerical modelling and pigment analysis can only derive grouped phytoplankton based on their biogeochemical functions. As mentioned in 2.1, in the ocean, there are $\sim 2 \times 10^4$ morphologically distinct phytoplankton species distributed among at least eight taxonomic divisions or phyla (Falkowski et al., 2004). Diatoms, for example, constitute one of the most diverse groups of phytoplankton. Recent estimations of the number of diatom species (including the ones from fresh water) range from 12,000 to 30,000 (Guiry, 2012; Mann and Vanormelingen, 2013). When estimating phytoplankton diversity using bio-optical or numerical models, one needs to bear in mind that there is great diversity even within a certain phytoplankton group.

Currently, the RecoM2 can resolve diatoms and nondiatoms. In the future, it is possible to expand the model capacity of simulating more phytoplankton functional types. The model of Le Quéré et al. (2005), for example, is able to predict seven functional types. An access to global marker pigment distributions estimated from optical data, either from *in situ* or satellite, may assist in further development and validation of the RecoM2. The pigment estimation algorithms developed in this thesis using underway spectrophotometry data have a great potential. The high spatial resolution underway AC-S data sets in the Fram Strait complement the global underway AC-S data set from the Tara Oceans expedition (Boss et al., 2013). The Gaussian decomposition method has been applied to the latter data set to obtain estimates for the same pigment groups (e.g., PPC) as my study (Chase et al., 2013). Furthermore, when the SVD-NNLS method applied to this global AC-S data set, the global distributions of several marker pigments, e.g., fucoxanthin, Hex, can be produced. Hex is a marker pigment of haptophytes. One type of haptophyte, *Phaeocystis*, also DMSP-producers in terms of their biogeochemical function, have a growing importance in the Fram Strait. Their growth data and data of Hex concentrations may be included in RecoM2 for their prediction in the Fram Strait, where coccolithophores, another haptophyte algae, are not prominent. Hex data may contribute to the model prediction of coccolithophores, the calcifiers, in waters where they are the predominant haptophytes, though the backscattering properties of coccoliths are more commonly explored, especially in satellite ocean color prediction for coccolithophores.

Though promising, the SVD-NNLS method has its inherent limitations. The main limitation is the ill-conditioning of the pigment matrix originated from the multicollinearity of phytoplankton pigment concentrations (**Paper II** Figure 4(b)). For example, Chl a shows high correlations (Spearman′s rank correlation coefficient $\geqslant 0.9$) with fucoxanthin, chlorophyll c_1 and c_2, chlorophyll c_3, and β-carotene in the study of **Paper II**. Fucoxanthin and chlorophyll c are the main light harvesting pigments of diatoms, while β-carotene is one of their PPC. This is in agreement with the fact that diatoms thrive in eutrophic waters where nutrients are abundant. Therefore, in eutrophic waters where diatoms are predominant, the SVD-NNLS method may not work. To minimize ill-conditioning and ensure stable pigment predictions in **Paper II**, nine pigments were involved in the pigment matrix, namely, fucoxanthin, diadinoxanthin, chlorophyll c_1 and c_2, chlorophyll b, peridinin, 19′-hexanoyloxyfucoxanthin, 19′-butanoyloxyfucoxanthin, and pheophytin a. This indicates a decent performance of SVD-NNLS in waters containing a mixture of several phytoplankton groups. In oligotrophic waters (e.g., the subtropical gyres) where phytoplankton of smaller size predominate, SVD-NNLS could perform well provided that pigment concentrations do not significantly co-vary (e.g., Chl a and zeaxanthin, the marker pigment for cyanobacteria) and that the pigment concentrations are sufficiently higher than zero.

The Gaussian decomposition and the SVD-NNLS methods can be incorporated into the reconstruction of hyperspectral $R_{rs}(\lambda)$ through Equation 2.2. A trial has been made that applied Gaussian decomposition to *in situ* hyperspectral radiometric data to retrieve phytoplankton pigment groups (Chase et al., 2017). This shows potential for their applications in deriving phytoplankton pigments from hyperspectral satellite ocean color radiometers (e.g., PACE/OCI). Hyperspectral observations are more capable of discerning subtle differences in spectral absorption by various accessory pigments than multi-spectral data. This is especially true for the SVD-NNLS method, as when it was applied to $a_p(\lambda)$ at ten MODIS bands, the number of pigments that can be estimated as well as the estimation accuracy was reduced (**Paper II**). In addition, extending the visible to UV light region could make a difference in retrieving phytoplankton pigments from hyperspectral ocean color data. In theory, it is more accurate to estimate pigments from $a_{ph}(\lambda)$ than from $R_{rs}(\lambda)$, since the former was only affected by pigment while the latter contain signals of all water constituents. Also, it is theoretically more accurate and physically explainable to derive spectral absorption coefficients than pigments from $R_{rs}(\lambda)$ (IOCCG, 2006). Grunert et al. (2019) has shown that when decomposing the spectral non-water absorption coefficients, a Level 2 satellite ocean color product, into the contribution of colored detrital matter (i.e., CDOM+non-algal particles) and phytoplankton, data at the UV wavelengths can accurately account for the effect of colored detrital matter, allowing accurate derivation of $a_{ph}(\lambda)$. Given that this method is applied to hyperspectral satellite non-water absorption products in the future, easier implementation of the Gaussian decomposition and the SVD-NNLS methods and more accurate pigment estimates can be expected.

References

Aagaard, K. and Greisman, P. Toward new mass and heat budgets for the Arctic Ocean. *Journal of Geophysical Research*, 80(27):3821–3827, 1975.

Alvain, S., Moulin, C., Dandonneau, Y., and Bréon, F.-M. 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, 2005.

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

Álvarez, E., Thoms, S., Bracher, A., Liu, Y., and Völker, C. Modeling photoprotection at global scale: The relative role of nonphotosynthetic pigments, physiological state, and species composition. *Global Biogeochemical Cycles*, 33(7):904–926, 2019.

Ardyna, M. and Arrigo, K. R. Phytoplankton dynamics in a changing Arctic Ocean. *Nature Climate Change*, 10(10):892–903, 2020.

Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., and Tremblay, J.-É. Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophysical Research Letters*, 41(17):6207–6212, 2014.

Ardyna, M., Babin, M., Devred, E., Forest, A., Gosselin, M., Raimbault, P., and Tremblay, J.-É. Shelf-basin gradients shape ecological phytoplankton niches and community composition in the coastal Arctic Ocean (Beaufort Sea). *Limnology and oceanography*, 62(5):2113–2132, 2017.

Bauerfeind, E., Nöthig, E.-M., Beszczynska, A., Fahl, K., Kaleschke, L., Kreker, K., Klages, M., Soltwedel, T., Lorenzen, C., and Wegner, J. Particle sedimentation patterns in the eastern Fram Strait during 2000–2005: Results from the Arctic long-term observatory HAUSGARTEN. *Deep Sea Research Part I: Oceanographic Research Papers*, 56(9):1471–1487, 2009.

Behrenfeld, M. J. and Falkowski, P. G. A consumer's guide to phytoplankton primary productivity models. *Limnology and oceanography*, 42(7):1479–1491, 1997.

Beszczynska-Möller, A., Fahrbach, E., Schauer, U., and Hansen, E. Variability in Atlantic water temperature and transport at the entrance to the Arctic Ocean, 1997–2010. *ICES Journal of Marine Science*, 69(5):852–863, 2012.

Bidigare, R., Smith, R., Baker, K., and Marra, J. Oceanic primary production estimates from measurements of spectral irradiance and pigment concentrations. *Global Biogeochemical Cycles*, 1(3):171–186, 1987.

Bidigare, R., Morrow, J., and Kiefer, D. Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea. *Journal of Marine Research*, 47 (2):323–341, 1989.

Boss, E., Picheral, M., Leeuw, T., Chase, A., Karsenti, E., Gorsky, G., Taylor, L., Slade, W., Ras, J., and Claustre, H. The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition. *Methods in Oceanography*, 7:52–62, 2013.

Bricaud, A. and Stramski, D. Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: A comparison between the Peru upwelling area and the Sargasso Sea. *Limnology and oceanography*, 35(3):562–582, 1990.

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

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

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., and Ruescas, A. Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In *Living Planet Symposium*, volume 740, page 54. 2016.

Brunet, C., Johnsen, G., Lavaud, J., and Roy, S. Pigments and photoacclimation processes. In *Phytoplankton Pigments – Characterization, Chemotaxonomy and Applications in Oceanography*, pages 445–454. 2011.

Chase, A. *Phytoplankton community composition in the surface ocean: methods for detection using optical measurements, pigment concentrations, and flow cytometry*. PhD thesis, University of Maine, 2020.

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Dall'Olmo, G., and Westberry, T. K. Decomposition of in situ particulate absorption spectra. *Methods in Oceanography*, 7:110–124, 2013.

Chase, A., Boss, E., Cetinić, I., and Slade, W. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: toward a global algorithm. *Journal of Geophysical Research: Oceans*, 122(12):9725–9743, 2017.

Chassot, E., Bonhommeau, S., Dulvy, N. K., Mélin, F., Watson, R., Gascuel, D., and Le Pape, O. Global marine primary production constrains fisheries catches. *Ecology letters*, 13(4):495–505, 2010.

Cota, G. F., Harrison, W. G., Platt, T., Sathyendranath, S., and Stuart, V. Bio-optical properties of the Labrador Sea. *Journal of Geophysical Research: Oceans*, 108(C7), 2003.

Cota, G. F., Wang, J., and Comiso, J. C. Transformation of global satellite chlorophyll retrievals with a regionally tuned algorithm. *Remote Sensing of Environment*, 90(3): 373–377, 2004.

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

Cullen, J. J., Yang, X., and MacIntyre, H. L. Nutrient limitation of marine photosynthesis. In *Primary productivity and biogeochemical cycles in the sea*, pages 69–88. Springer, 1992.

Delwiche, C. F. Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist*, 154(S4):S164–S177, 1999.

Doerffer, R. and Schiller, H. The MERIS Case 2 water algorithm. *International Journal of Remote Sensing*, 28(3-4):517–535, 2007.

Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature*, 387(6630):272–275, 1997.

Falkowski, P. G. Tracing oxygen's imprint on earth's metabolic evolution. *Science*, 311 (5768):1724–1725, 2006.

Falkowski, P. G. Biogeochemistry of primary production in the sea. In *Treatise on Geochemistry*, pages 163–187. Elsevier, 2014.

Falkowski, P. G. and Raven, J. A. *Aquatic photosynthesis*. Princeton University Press, 2013.

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

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 281(5374): 237–240, 1998.

Garver, S. A. and Siegel, D. A. 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, 1997.

Geider, R. J., MacIntyre, H. L., and Kana, T. M. A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, 43(4):679–694, 1998.

Gordon, H. R., Clark, D. K., Brown, J. W., Brown, O. B., Evans, R. H., and Broenkow, W. W. Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates. *Applied Optics*, 22(1):20–36, 1983.

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

Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C., Chauhan, P., Choi, J.-k., Chuprin, A., et al. Satellite ocean colour: current status and future perspective. *Frontiers in Marine Science*, 6:485, 2019.

Gruber, N. and Galloway, J. N. An Earth-system perspective of the global nitrogen cycle. *Nature*, 451(7176):293–296, 2008.

Grunert, B. K., Mouw, C. B., and Ciochetto, A. B. Deriving inherent optical properties from decomposition of hyperspectral non-water absorption. *Remote Sensing of Environment*, 225:193–206, 2019.

Guiry, M. D. How many species of algae are there? *Journal of Phycology*, 48(5):1057–1063, 2012.

Hansen, E., Gerland, S., Granskog, M., Pavlova, O., Renner, A., Haapala, J., Løyning, T., and Tschudi, M. Thinning of Arctic sea ice observed in Fram Strait: 1990–2011. *Journal of Geophysical Research: Oceans*, 118(10):5202–5221, 2013.

Hauck, J., Völker, C., Wang, T., Hoppema, M., Losch, M., and Wolf-Gladrow, D. A. Seasonally different carbon flux changes in the Southern Ocean in response to the southern annular mode. *Global Biogeochemical Cycles*, 27(4):1236–1245, 2013.

Hegseth, E. N. and Sundfjord, A. Intrusion and blooming of Atlantic phytoplankton species in the high Arctic. *Journal of Marine Systems*, 74(1-2):108–119, 2008.

Hensen, V. Über die Bestimmung des Planktons oder des im Meere treibenden Materials an Pflanzen und Thieren. *Bericht der Commission zur Wissenschaftlichen Untersuchung der Deutschen Meere*, 50(12-16):107, 1887.

Hieronymi, M., Müller, D., and Doerffer, R. The OLCI Neural Network Swarm (ONNS): a bio-geo-optical algorithm for open ocean and coastal waters. *Frontiers in Marine Science*, 4:140, 2017.

Higgins, H. W., Wright, S. W., and Schluter, L. Quantitative interpretation of chemotaxonomic pigment data. In *Phytoplankton Pigments – Characterization, Chemotaxonomy and Applications in Oceanography*, pages 257–313. Cambridge University Press, 2011.

Hohn, S. *Coupling and decoupling of biogeochemical cycles in marine ecosystems*. PhD thesis, Universität Bremen, 2008.

Huot, Y. and Babin, M. Overview of fluorescence protocols: theory, basic concepts, and practice. In *Chlorophyll a fluorescence in aquatic sciences: Methods and applications*, pages 31–74. Springer, 2010.

IOCCG. *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*. IOCCG Report Series, No. 5, International Ocean Colour Coordinating Group, Dartmouth, Canada, 2006.

IOCCG. *Mission requirements for future ocean-colour sensors*. IOCCG Report Series, No. 13, International Ocean Colour Coordinating Group, Dartmouth, Canada, 2012.

IOCCG. *Phytoplankton functional types from space*. IOCCG Report Series, No. 15, International Ocean Colour Coordinating Group, Dartmouth, Canada, 2014.

IOCCG. *Ocean colour remote sensing in polar seas*. IOCCG Report Series, No. 16, International Ocean Colour Coordinating Group, Dartmouth, Canada, 2015.

IOCCG Protocol Series. *Inherent Optical Property Measurements and Protocols: Absorption coefficient*. IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, International Ocean Colour Coordinating Group, Dartmouth, Canada, 2018.

IOCCG Protocol Series. *Inherent Optical Property Measurements and Protocols: Best practices for the collection and processing of ship-based underway flow-through optical data*. IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, International Ocean Colour Coordinating Group, Dartmouth, Canada, 2019.

IPCC. *Climate Change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of IPCC the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., 2013.

IPCC. Summary for policymakers. In *Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the* *threat of climate change, sustainable development, and efforts to eradicate poverty*. World Meteorological Organization, Geneva, Switzerland, 32 pp., 2018.

IPCC. *IPCC special report on the ocean and cryosphere in a changing climate*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 755 pp, 2019. `https://doi.org/10.1017/9781009157964`.

Kahru, M., Brotas, V., Manzano-Sarabia, M., and Mitchell, B. Are phytoplankton blooms occurring earlier in the Arctic? *Global Change Biology*, 17(4):1733–1739, 2011.

Kirkpatrick, G. J., Millie, D. F., Moline, M. A., and Schofield, O. Optical discrimination of a phytoplankton species in natural mixed populations. *Limnology and Oceanography*, 45(2):467–471, 2000.

Klaas, C. and Archer, D. E. Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochemical Cycles*, 16(4):63–1, 2002.

Krumpen, T., Belter, H. J., Boetius, A., Damm, E., Haas, C., Hendricks, S., Nicolaus, M., Nöthig, E.-M., Paul, S., Peeken, I., et al. Arctic warming interrupts the Transpolar Drift and affects long-range transport of sea ice and ice-rafted matter. *Scientific Reports*, 9(1):1–9, 2019.

Kwok, R. Outflow of Arctic Ocean sea ice into the Greenland and Barents Seas: 1979–2007. *Journal of Climate*, 22(9):2438–2457, 2009.

Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., et al. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11):2016–2040, 2005.

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

Lin, H., Kuzminov, F. I., Park, J., Lee, S., Falkowski, P. G., and Gorbunov, M. Y. The fate of photons absorbed by phytoplankton in the global ocean. *Science*, 351(6270): 264–267, 2016.

Lipps, J. H. *Fossil prokaryotes and protists*. Blackwell Scientific Publications, 1993.

Lisitzin, A. Distribution of siliceous microfossils in suspension and in bottom sediments. *The Micropaleontology of Oceans*, pages 173–195, 1971.

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., and de Vargas, C. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences*, 106(31):12803–12808, 2009.

Liu, Y., Röttgers, R., Ramírez-Pérez, M., Dinter, T., Steinmetz, F., Nöthig, E.-M., Hellmann, S., Wiegmann, S., and Bracher, A. Underway spectrophotometry in the Fram Strait (European Arctic Ocean): a highly resolved chlorophyll a data source for complementing satellite ocean color. *Optics Express*, 26(14):A678–A696, 2018.

Liu, Y., Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., Pan, Y., and Bracher, A. Retrieval of phytoplankton pigments from underway spectrophotometry in the Fram Strait. *Remote Sensing*, 11(3):318, 2019.

Lovelock, J. E., Maggs, R., and Rasmussen, R. Atmospheric dimethyl sulphide and the natural sulphur cycle. *Nature*, 237(5356):452–453, 1972.

Mackey, M., Mackey, D., Higgins, H., and Wright, S. CHEMTAX-a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series*, 144:265–283, 1996.

Maier-Reimer, E., Mikolajewicz, U., and Winguth, A. Future ocean uptake of CO_2: interaction between ocean circulation and biology. *Climate Dynamics*, 12(10):711–722, 1996.

Mann, D. G. and Vanormelingen, P. An inordinate fondness? The number, distributions, and origins of diatom species. *Journal of Eukaryotic Microbiology*, 60(4):414–420, 2013.

Marshall, H. L., Geider, R. J., and Flynn, K. J. A mechanistic model of photoinhibition. *The New Phytologist*, 145(2):347–359, 2000.

Matsuoka, A., Huot, Y., Shimada, K., Saitoh, S.-I., and Babin, M. Bio-optical characteristics of the western arctic ocean: implications for ocean color algorithms. *Canadian Journal of Remote Sensing*, 33(6):503–518, 2007.

Matsuoka, A., Hill, V., Huot, Y., Babin, M., and Bricaud, A. Seasonal variability in the light absorption properties of western Arctic waters: Parameterization of the individual components of absorption for ocean color applications. *Journal of Geophysical Research: Oceans*, 116(C2), 2011.

McClain, C. R. A decade of satellite ocean color observations. *Annual Review of Marine Science*, 1:19–42, 2009.

McCoy, I. L., Bretherton, C. S., Wood, R., Twohy, C. H., Gettelman, A., Bardeen, C. G., and Toohey, D. W. Influences of recent particle formation on Southern Ocean aerosol variability and low cloud properties. *Journal of Geophysical Research: Atmospheres*, 126 (8):e2020JD033529, 2021.

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

Moisan, J., Moisan, T., and Linkswiler, M. A. An inverse modeling approach to estimating phytoplankton pigment concentrations from phytoplankton absorption spectra. *Journal of Geophysical Research: Oceans*, 116(C9), 2011.

Moisan, T. A., Moisan, J. R., Linkswiler, M. A., and Steinhardt, R. A. Algorithm development for predicting biodiversity based on phytoplankton absorption. *Continental Shelf Research*, 55:17–28, 2013.

Moon, T. A., Druckenmiller, M. L., and Thoman, R. L., Eds. Arctic Report Card 2021, 2021. `https://doi.org/10.25923/5s0f-5163`.

Moore, C. C., Zaneveld, J. R. V., and Kitchen, J. C. Preliminary results from an in-situ spectral absorption meter. In *Ocean Optics XI*, volume 1750, pages 330–337. International Society for Optics and Photonics, 1992.

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

Müller, D., Krasemann, H., Brewin, R. J., Brockmann, C., Deschamps, P.-Y., Doerffer, R., Fomferra, N., Franz, B. A., Grant, M. G., Groom, S. B., et al. The Ocean Colour Climate Change Initiative: I. A methodology for assessing atmospheric correction processors based on in-situ measurements. *Remote Sensing of Environment*, 162:242–256, 2015a.

Müller, D., Krasemann, H., Brewin, R. J., Brockmann, C., Deschamps, P.-Y., Doerffer, R., Fomferra, N., Franz, B. A., Grant, M. G., Groom, S. B., et al. The Ocean Colour Climate Change Initiative: II. Spatial and temporal homogeneity of satellite data retrieval due to systematic effects in atmospheric correction processors. *Remote Sensing of Environment*, 162:257–270, 2015b.

Nicolaus, M., Katlein, C., Maslanik, J., and Hendricks, S. Changes in Arctic sea ice result in increasing light transmittance and absorption. *Geophysical Research Letters*, 39(24), 2012.

Nöthig, E.-M., Bracher, A., Engel, A., Metfies, K., Niehoff, B., Peeken, I., Bauerfeind, E., Cherkasheva, A., Gäbler-Schwarz, S., Hardge, K., et al. Summertime plankton ecology in Fram Strait — a compilation of long-and short-term observations. *Polar Research*, 34(1):23349, 2015.

Nöthig, E.-M., Ramondenc, S., Haas, A., Hehemann, L., Walter, A., Bracher, A., Lalande, C., Metfies, K., Peeken, I., Bauerfeind, E., et al. Summertime chlorophyll a and particulate organic carbon standing stocks in surface waters of the Fram Strait and the Arctic Ocean (1991–2015). *Frontiers in Marine Science*, 7:350, 2020.

Organelli, E., Bricaud, A., Antoine, D., and Uitz, J. Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site). *Applied Optics*, 52(11):2257–2273, 2013.

Reynolds, C. S. *The ecology of phytoplankton*. Cambridge University Press, 2006.

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmechtig, C., Poteau, A., et al. Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors. *Limnology and Oceanography: Methods*, 15(6):572–585, 2017.

Rudels, B., Schauer, U., Björk, G., Korhonen, M., Pisarev, S., Rabe, B., and Wisotzki, A. Observations of water masses and circulation with focus on the Eurasian Basin of the Arctic Ocean from the 1990s to the late 2000s. *Ocean Science*, 9(1):147–169, 2013.

Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., and Wolf-Gladrow, D. Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. *Biogeosciences*, 4(4):433–454, 2007.

Schiebel, R. Planktic foraminiferal sedimentation and the marine calcite budget. *Global Biogeochemical Cycles*, 16(4):3–1, 2002.

Schourup-Kristensen, V., Wekerle, C., Danilov, S., and Völker, C. Seasonality of mesoscale phytoplankton control in eastern Fram Strait. *Journal of Geophysical Research: Oceans*, 126(10):e2021JC017279, 2021.

Slade, W. H., Boss, E., Dall'Olmo, G., Langner, M. R., Loftin, J., Behrenfeld, M. J., Roesler, C., and Westberry, T. K. Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *Journal of Atmospheric and Oceanic Technology*, 27(10):1733–1746, 2010.

Steinmetz, F., Deschamps, P.-Y., and Ramon, D. Atmospheric correction in presence of sun glint: application to MERIS. *Optics Express*, 19(10):9783–9800, 2011.

Summons, R. E., Jahnke, L. L., Hope, J. M., and Logan, G. A. 2-methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature*, 400(6744):554–557, 1999.

Tappan, H. N. *The paleobiology of plant protists*. WH Freeman, 1980.

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

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., and Marty, J.-C. 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, 2001.

von Appen, W.-J., Schauer, U., Somavilla, R., Bauerfeind, E., and Beszczynska-Möller, A. Exchange of warming deep waters across Fram Strait. *Deep Sea Research Part I: Oceanographic Research Papers*, 103:86–100, 2015.

von Appen, W.-J., Waite, A. M., Bergmann, M., Bienhold, C., Boebel, O., Bracher, A., Cisewski, B., Hagemann, J., Hoppema, M., Iversen, M. H., et al. Sea-ice derived meltwater stratification slows the biological carbon pump: results from continuous observations. *Nature Communications*, 12(1):1–16, 2021.

Walczowski, W. Frontal structures in the West Spitsbergen Current margins. *Ocean Science*, 9(6):957–975, 2013.

Wang, G., Lee, Z., Mishra, D. R., and Ma, R. Retrieving absorption coefficients of multiple phytoplankton pigments from hyperspectral remote sensing reflectance measured over cyanobacteria bloom waters. *Limnology and Oceanography: Methods*, 14(7): 432–447, 2016.

Xi, H., Hieronymi, M., Röttgers, R., Krasemann, H., and Qiu, Z. Hyperspectral differentiation of phytoplankton taxonomic groups: a comparison between using remote sensing reflectance and absorption spectra. *Remote Sensing*, 7(11):14781–14805, 2015.

Xi, H., Hieronymi, M., Krasemann, H., and Röttgers, R. Phytoplankton group identification using simulated and in situ hyperspectral remote sensing reflectance. *Frontiers in Marine Science*, 4:272, 2017.

Xing, X., Claustre, H., Boss, E., Roesler, C., Organelli, E., Poteau, A., Barbieux, M., and d'Ortenzio, F. Correction of profiles of in-situ chlorophyll fluorometry for the contribution of fluorescence originating from non-algal matter. *Limnology and Oceanography: Methods*, 15(1):80–93, 2017.

Zaneveld, J. R. V., Kitchen, J. C., Bricaud, A., and Moore, C. C. Analysis of in-situ spectral absorption meter data. In *Ocean Optics XI*, volume 1750, pages 187–200. International Society for Optics and Photonics, 1992.

Zaneveld, J. R. V., Kitchen, J. C., and Moore, C. C. Scattering error correction of reflection-tube absorption meters. In *Ocean Optics XII*, volume 2258, pages 44–55. International Society for Optics and Photonics, 1994.

Zeebe, R. E. and Wolf-Gladrow, D. *CO_2 in seawater: equilibrium, kinetics, isotopes*. Gulf Professional Publishing, 2001.

Chapter 4

Publications

Paper I

Research Article Vol. 26, No. 14 | 9 Jul 2018 | OPTICS EXPRESS A678

Optics EXPRESS

Underway spectrophotometry in the Fram Strait (European Arctic Ocean): a highly resolved chlorophyll *a* data source for complementing satellite ocean color

YANGYANG LIU,[1,2,*] RÜDIGER RÖTTGERS,[3] MARTA RAMÍREZ-PÉREZ,[4] TILMAN DINTER,[1,5] FRANÇOIS STEINMETZ,[6] EVA-MARIA NÖTHIG,[1] SEBASTIAN HELLMANN,[1] SONJA WIEGMANN,[1] AND ASTRID BRACHER[1,5]

[1] *Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bussesstraße 24, Bremerhaven, 27570, Germany*
[2] *Faculty of Biology and Chemistry, University of Bremen, Leobener Straße NW 2, Bremen, 28359, Germany*
[3] *Institute for Coastal Research, Helmholtz-Zentrum Geesthacht, Center for Materials and Coastal Research, Max-Planck-Straße, Geesthacht, 21502, Germany*
[4] *Department of Physical and Technological Oceanography, Institute of Marine Sciences, Passeig Maritim de la Barceloneta 37-49, Barcelona, 08003, Spain*
[5] *Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, Bremen, 28359, Germany*
[6] *Hygeos, 165 Avenue de Bretagne, 59000 Lille, France*
[*] *yangyang.liu@awi.de*

Abstract: Satellite remote sensing of chlorophyll *a* concentration (Chl-a) in the Arctic Ocean is spatially and temporally limited and needs to be supplemented and validated with substantial volumes of *in situ* observations. Here, we evaluated the capability of obtaining highly resolved *in situ* surface Chl-a using underway spectrophotometry operated during two summer cruises in 2015 and 2016 in the Fram Strait. Results showed that Chl-a measured using high pressure liquid chromatography (HPLC) was well related (R^2 = 0.90) to the collocated particulate absorption line height at 676 nm obtained from the underway spectrophotometry system. This enabled continuous surface Chl-a estimation along the cruise tracks. When used to validate Chl-a operational products as well as to assess the Chl-a algorithms of the aqua moderate resolution imaging spectroradiometer (MODIS-A) and Sentinel-3 Ocean Land Color Imager (OLCI) Level 2 Chl-a operational products, and from OLCI Level 2 products processed with Polymer atmospheric correction algorithm (version 4.1), the underway spectrophotometry based Chl-a data sets proved to be a much more sufficient data source by generating over one order of magnitude more match-ups than those obtained from discrete water samples. Overall, the band ratio (OCI, OC4) Chl-a operational products from MODIS-A and OLCI as well as OLCI C2RCC products showed acceptable results. The OLCI Polymer standard output provided the most reliable Chl-a estimates, and nearly as good results were obtained from the OCI algorithm with Polymer atmospheric correction method. This work confirms the great advantage of the underway spectrophotometry in enlarging *in situ* Chl-a data sets for the Fram Strait and improving satellite Chl-a validation and Chl-a algorithm assessment over discrete water sample analysis in the laboratory.

OCIS codes: (010.4450) Oceanic optics; (010.1030) Absorption.

References and links

1. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, *Climate change 2013: The physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2013).

#313439 https://doi.org/10.1364/OE.26.00A678
Journal © 2018 Received 12 Apr 2018; revised 20 Jan 2018; accepted 28 Feb 2018; published 19 Jun 2018

2. J. Y. Park, J. S. Kug, J. Bader, R. Rolph, and M. Kwon, “Amplified Arctic warming by phytoplankton under greenhouse warming,” Proc. Natl. Acad. Sci. U.S.A. **112**(19), 5921–5926 (2015).
3. K. R. Arrigo and G. L. van Dijken, “Continued increases in Arctic Ocean primary production,” Prog. Oceanogr. **136**, 60–70 (2015).
4. A. J. Kettle and M. O. Andreae, “Flux of dimethylsulfide from the oceans: a comparison of updated data sets and flux models,” J. Geophys. Res. **105**(D22), 26793–26808 (2000).
5. M. Levasseur, “Impact of Arctic meltdown on the microbial cycling of Sulphur,” Nat. Geosci. **6**(9), 691–700 (2013).
6. R. Y. W. Chang, S. J. Sjostedt, J. R. Pierce, T. N. Papakyriakou, M. G. Scarratt, S. Michaud, M. Levasseur, W. R. Leaitch, and J. P. Abbatt, “Relating atmospheric and oceanic DMS levels to particle nucleation events in the Canadian Arctic,” J. Geophys. Res. Atmos. **116**(D17), D00S03 (2011).
7. P. Tunved, J. Ström, and R. Krejci, “Arctic aerosol life cycle: linking aerosol size distributions observed between 2000 and 2010 with air mass transport and precipitation at Zeppelin station, Ny-Ålesund, Svalbard,” Atmos. Chem. Phys. **13**(7), 3643–3660 (2013).
8. C. Roesler, J. Uitz, H. Claustre, E. Boss, X. Xing, E. Organelli, N. Briggs, A. Bricaud, C. Schmechtig, A. Poteau, F. D’Ortenzio, J. Ras, S. Drapeau, N. Haëntjens, and M. Barbieux, “Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors,” Limnol. Oceanogr. Methods **15**(6), 572–585 (2017).
9. X. Xing, H. Claustre, E. Boss, C. Roesler, E. Organelli, A. Poteau, M. Barbieux, and F. D’Ortenzio, “Correction of profiles of in-situ chlorophyll fluorometry for the contribution of fluorescence originating from non-algal matter,” Limnol. Oceanogr. Methods **15**(1), 80–93 (2017).
10. Y. Huot and M. Babin, “Overview of fluorescence protocols: Theory, basic concepts, and practice,” in *Chlorophyll a fluorescence in aquatic sciences: methods and applications, developments in applied phycology*, E. D. Suggett et al., ed. (Springer, 2011).
11. C. C. Trees, M. C. Kennicutt II, and J. M. Brooks, “Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments,” Mar. Chem. **17**(1), 1–12 (1985).
12. C. C. Trees, R. R. Bidigare, and J. M. Brooks, “Distribution of chlorophylls and phaeopigments in the Northwestern Atlantic Ocean,” J. Plankton Res. **8**(3), 447–458 (1986).
13. J. Neveux, D. Delmas, J. C. Romano, P. Algarra, L. Ignatiades, A. Herbland, P. Morand, A. Neori, D. Bonin, J. Barbe, and A. Sukenik, “Comparison of chlorophyll and phaeopigment determinations by spectrophotometric, fluorometric, spectrofluorometric and HPLCmethods,” Mar. Microb. Food Webs **4**(2), 217–238 (1990).
14. M. Babin, K. Arrigo, S. Bélanger, and M.-H. Forget, *Ocean Colour Remote Sensing in Polar Seas* (IOCCG, 2015).
15. K. M. Lewis, B. G. Mitchell, G. L. van Dijken, and K. R. Arrigo, “Regional chlorophyll a algorithms in the Arctic Ocean and their effect on satellite-derived primary production estimates,” Deep Sea Res. Part II Top. Stud. Oceanogr. **130**, 14–27 (2016).
16. B. G. Mitchell, “Predictive bio-optical relationships for polar oceans and marginal ice zones,” J. Mar. Syst. **3**(1-2), 91–105 (1992).
17. A. Matsuoka, Y. Huot, K. Shimada, S. Saitoh, and M. Babin, “Bio-optical characteristics of the western Arctic Ocean: implications for ocean color algorithms,” Can. J. Rem. Sens. **33**(6), 503–518 (2007).
18. S. Ben Mustapha, S. Bélanger, and P. Larouche, “Evaluation of ocean color algorithms in the southeastern Beaufort Sea, Canadian Arctic: new parameterization using SeaWiFS, MODIS, and MERIS spectral bands,” Can. J. Rem. Sens. **38**(5), 535–556 (2012).
19. A. Matsuoka, V. Hill, Y. Huot, M. Babin, and A. Bricaud, “Seasonal variability in the light absorption properties of western Arctic waters: parameterization of the individual components of absorption for ocean color applications,” J. Geophys. Res. **116**(C2), C02007 (2011).
20. J. Wang and G. F. Cota, “Remote-sensing reflectance in the Beaufort and Chukchi seas: observations and models,” Appl. Opt. **42**(15), 2754–2765 (2003).
21. M. C. Serreze, M. M. Holland, and J. Stroeve, “Perspectives on the Arctic’s shrinking sea-ice cover,” Science **315**(5818), 1533–1536 (2007).
22. E. Falck, G. Kattner, and G. Budéus, “Disappearance of Pacific water in the northwestern Fram Strait,” Geophys. Res. Lett. **32**(14), L14619 (2005).
23. R. R. Dickson, J. Meincke, and P. Rhines, *Arctic-Subarctic Ocean Fluxes: Defining the Role of the Northern Seas in Climate* (Springer, 2008), Chap. 3.
24. A. Beszczynska-Möller, E. Fahrbach, U. Schauer, and E. Hansen, “Variability in Atlantic water temperature and transport at the entrance to the Arctic Ocean, 1997–2010,” ICES J. Mar. Sci. **69**(5), 852–863 (2012).
25. L. H. Smedsrud, A. Sorteberg, and K. Kloster, “Recent and future changes of the Arctic sea-ice cover,” Geophys. Res. Lett. **35**(20), L20503 (2008).
26. B. Rabe, M. Karcher, U. Schauer, J. M. Toole, R. A. Krishfield, S. Pisarev, F. Kauker, R. Gerdes, and T. Kikuchi, “An assessment of Arctic Ocean freshwater content changes from the 1990s to the 2006–2008 period,” Deep Sea Res. Part I Oceanogr. Res. Pap. **58**(2), 173–185 (2011).
27. M. H. Halvorsen, L. H. Smedsrud, R. Zhang, and K. Kloster, “Fram Strait spring ice export and September Arctic sea ice,” Cryosphere Discuss. **9**(4), 4205–4235 (2015).
28. E. Bauerfeind, E. M. Nöthig, A. Beszczynska, K. Fahl, L. Kaleschke, K. Kreker, M. Klages, T. Soltwedel, C. Lorenzen, and J. Wegner, “Particle sedimentation patterns in the eastern Fram Strait during 2000–2005: Results

from the Arctic long-term observatory HAUSGARTEN," Deep Sea Res. Part I Oceanogr. Res. Pap. **56**(9), 1471–1487 (2009).

29. E. M. Nöthig, A. Bracher, A. Engel, K. Metfies, B. Niehoff, I. Peeken, E. Bauerfeind, A. Cherkasheva, S. Gäbler-Schwarz, K. Hardge, E. Kilias, A. Kraft, Y. Mebrahtom Kidane, C. Lalande, J. Piontek, K. Thomisch, and M. Wurst, "Summertime plankton ecology in Fram Strait—a compilation of long- and short-term observations," Polar Res. **34**(1), 23349 (2015).
30. A. Kraft, E. Bauerfeind, and E. M. Nöthig, "Amphipod abundance in sediment trap samples at the long-term observatory HAUSGARTEN (Fram Strait, ~ 79° N/4° E). Variability in species community patterns," Mar. Biodivers. **41**(3), 353–364 (2011).
31. P. Wassmann, D. Slagstad, and I. Ellingsen, "Primary production and climatic variability in the European sector of the Arctic Ocean prior to 2007: preliminary results," Polar Biol. **33**(12), 1641–1650 (2010).
32. P. Wassmann, "Arctic marine ecosystems in an era of rapid climate change," Prog. Oceanogr. **90**(1–4), 1–17 (2011).
33. M. Reigstad, J. Carroll, D. Slagstad, I. Ellingsen, and P. Wassmann, "Intra-regional comparison of productivity, carbon flux and ecosystem composition within the northern Barents Sea," Prog. Oceanogr. **90**(1–4), 33–46 (2011).
34. A. Forest, P. Wassmann, D. Slagstad, E. Bauerfeind, E. M. Nöthig, and M. Klages, "Relationships between primary production and vertical particle export at the Atlantic-Arctic boundary (Fram Strait, HAUSGARTEN)," Polar Biol. **33**(12), 1733–1746 (2010).
35. A. Cherkasheva, A. Bracher, C. Melsheimer, C. Köberle, R. Gerdes, E. M. Nöthig, E. Bauerfeind, and A. Boetius, "Influence of the physical environment on polar phytoplankton blooms: a case study in the Fram Strait," J. Mar. Syst. **132**, 196–207 (2014).
36. G. Dall'Olmo, T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade, "Significant contribution of large particles to optical backscattering in the open ocean," Biogeosciences **6**(6), 947–967 (2009).
37. G. Dall'Olmo, E. Boss, M. J. Behrenfeld, and T. K. Westberry, "Particulate optical scattering coefficients along an Atlantic Meridional Transect," Opt. Express **20**(19), 21532–21551 (2012).
38. T. K. Westberry, G. Dall'Olmo, E. Boss, M. J. Behrenfeld, and T. Moutin, "Coherence of particulate beam attenuation and backscattering coefficients in diverse open ocean environments," Opt. Express **18**(15), 15419–15425 (2010).
39. W. H. Slade, E. Boss, G. Dall'Olmo, M. R. Langner, J. Loftin, M. J. Behrenfeld, C. Roesler, and T. K. Westberry, "Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation," J. Atmos. Ocean. Technol. **27**(10), 1733–1746 (2010).
40. E. Boss, M. Picheral, T. Leeuw, A. Chase, E. Karsenti, G. Gorsky, L. Taylor, W. Slade, J. Ras, and H. Claustre, "The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition," Methods in Oceanography **7**, 52–62 (2013).
41. A. Chase, E. Boss, R. Zaneveld, A. Bricaud, H. Claustre, J. Ras, G. Dall'Olmo, and T. K. Westberry, "Decomposition of in situ particulate absorption spectra," Methods in Oceanography **7**, 110–124 (2013).
42. R. J. Brewin, G. Dall'Olmo, S. Pardo, V. van Dongen-Vogels, and E. S. Boss, "Underway spectrophotometry along the Atlantic Meridional Transect reveals high performance in satellite chlorophyll retrievals," Remote Sens. Environ. **183**, 82–97 (2016).
43. A. Lindfors, K. Rasmus, and N. Strömbeck, "Point or pointless - quality of ground data," Int. J. Remote Sens. **26**(2), 415–423 (2005).
44. S. Koponen, J. Attila, J. Pulliainen, K. Kallio, T. Pyhälahti, A. Lindfors, K. Rasmus, and M. Hallikainen, "A case study of airborne and satellite remote sensing of a spring bloom event in the Gulf of Finland," Cont. Shelf Res. **27**(2), 228–244 (2007).
45. G. Dall'Olmo, E. Boss, M. J. Behrenfeld, T. K. Westberry, C. Courties, L. Prieur, M. Pujo-Pay, N. Hardman-Mountford, and T. Moutin, "Inferring phytoplankton carbon and eco-physiological rates from diel cycles of spectral particulate beam-attenuation coefficient," Biogeosciences **8**(11), 3423–3439 (2011).
46. P. J. Werdell, C. W. Proctor, E. Boss, T. Leeuw, and M. Ouhssain, "Underway sampling of marine inherent optical properties on the Tara Oceans expedition as a novel resource for ocean color satellite data product validation," Methods in Oceanography **7**, 40–51 (2013).
47. G. Dall'Olmo, R. J. W. Brewin, F. Nencioli, E. Organelli, I. Lefering, D. McKee, R. Röttgers, C. Mitchell, E. Boss, A. Bricaud, and G. Tilstone, "Determination of the absorption coefficient of chromophoric dissolved organic matter from underway spectrophotometry," Opt. Express **25**(24), A1079–A1095 (2017).
48. F. Steinmetz, P. Y. Deschamps, and D. Ramon, "Atmospheric correction in presence of sun glint: application to MERIS," Opt. Express **19**(10), 9783–9800 (2011).
49. F. Steinmetz, D. Ramon, and P.-Y. Deschamps, "Algorithm Theoretical Base Document (ATBD) v1 - Polymer atmospheric correction algorithm (Issue 2.1)," (ESA Ocean Colour Climate Change Initiative (OC-CCI) Project Deliverable 2.1, 2016), *http://www.esa-oceancolour-cci.org/?q=documents#*.
50. J. M. Sullivan, M. S. Twardowski, J. R. V. Zaneveld, C. M. Moore, A. H. Barnard, P. L. Donaghay, and B. Rhoades, "Hyperspectral temperature and salt dependencies of absorption by water and heavy water in the 400-750 nm spectral range," Appl. Opt. **45**(21), 5294–5309 (2006).
51. J. R. V. Zaneveld, J. C. Kitchen, and C. C. Moore, "Scattering error correction of reflecting-tube absorption meters," Proc. SPIE **2258**, 44–55 (1994).

52. B. B. Taylor, E. Torrecilla Ribalta, A. Bernhardt, M. H. Taylor, I. Peeken, R. Röttgers, J. Piera, and A. Bracher, "Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance," Biogeosci. **8**(12), 3609–3629 (2011).
53. R. G. Barlow, D. G. Cummings, and S. W. Gibb, "Improved resolution of mono-and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC," Mar. Ecol. Prog. Ser. **161**, 303–307 (1997).
54. J. Aiken, Y. Pradhan, R. Barlow, S. Lavender, A. Poulton, P. Holligan, and N. Hardman-Mountford, "Phytoplankton pigments and functional types in the Atlantic Ocean: a decadal assessment, 1995–2005," Deep Sea Res. Part II Top. Stud. Oceanogr. **56**(15), 899–917 (2009).
55. F. Vidussi, H. Claustre, B. B. Manca, A. Luchetta, and J. C. Marty, "Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter," J. Geophys. Res. Oceans **106**(C9), 19939–19956 (2001).
56. J. Uitz, H. Claustre, A. Morel, and S. B. Hooker, "Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll," J. Geophys. Res. Oceans **111**(C8), C08005 (2006).
57. T. Hirata, N. J. Hardman-Mountford, R. J. W. Brewin, J. Aiken, R. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguchi-Aita, and Y. Yamanaka, "Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types," Biogeosciences **8**(2), 311–327 (2011).
58. R. J. Brewin, S. Sathyendranath, T. Jackson, R. Barlow, V. Brotas, R. Airs, and T. Lamont, "Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class," Remote Sens. Environ. **168**, 437–450 (2015).
59. S. N. Losa, M. A. Soppa, T. Dinter, A. Wolanin, R. J. Brewin, A. Bricaud, J. Oelker, I. Peeken, B. Gentili, V. Rozanov, and A. Bracher, "Synergistic Exploitation of Hyper-and Multi-Spectral Precursor Sentinel Measurements to Determine Phytoplankton Functional Types (SynSenPFT)," Front. Mater. Sci. **4**, 1–22 (2017).
60. S. G. Simis, M. Tijdens, H. L. Hoogveld, and H. J. Gons, "Optical changes associated with cyanobacterial bloom termination by viral lysis," J. Plankton Res. **27**(9), 937–949 (2005).
61. R. Röttgers and S. Gehnke, "Measurement of light absorption by aquatic particles: improvement of the quantitative filter technique by use of an integrating sphere approach," Appl. Opt. **51**(9), 1336–1351 (2012).
62. R. Röttgers, D. Doxaran, and C. Dupouy, "Quantitative filter technique measurements of spectral light absorption by aquatic particles using a portable integrating cavity absorption meter (QFT-ICAM)," Opt. Express **24**(2), A1–A20 (2016).
63. K. Barker, C. Mazeran, C. Lerebourg, M. Bouvet, D. Antoine, M. Ondrusek, and L. Zibordi, "MERMAID: The MEris MAtchup In-situ Database," in *Proceedings of the 2nd MERIS/(A) ATSR User Workshop*, H. Lacoste, L. Ouwehand, ed. (ESA Communication Production Office, 2008).
64. S. W. Bailey and P. J. Werdell, "A multi-sensor approach for the on-orbit validation of ocean color satellite data products," Remote Sens. Environ. **102**(1), 12–23 (2006).
65. R. J. Brewin, S. Sathyendranath, D. Müller, C. Brockmann, P. Y. Deschamps, E. Devred, R. Doerffer, N. Fomferra, B. Franz, M. Grant, S. Groom, A. Horseman, C. Hu, H. Krasemann, Z. P. Lee, S. Maritorena, F. Mélin, M. Peters, T. Platt, P. Regner, T. Smyth, F. Steinmetz, J. Swinton, J. Werdell, and G. N. White III, "The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms," Remote Sens. Environ. **162**, 271–294 (2015).
66. C. Sá, D. D'Alimonte, A. C. Brito, T. Kajiyama, C. R. Mendes, J. Vitorino, P. B. Oliveira, J. C. B. Da Silva, and V. Brotas, "Validation of standard and alternative satellite ocean-color chlorophyll products off Western Iberia," Remote Sens. Environ. **168**, 403–419 (2015).
67. D. M. Glover, W. J. Jenkins, and S. C. Doney, *Modeling methods for marine science* (Cambridge University Press 2011).
68. C. S. Roesler and A. H. Barnard, "Optical proxy for phytoplankton biomass in the absence of photophysiology: Rethinking the absorption line height," Methods in Oceanography **7**, 79–94 (2013).
69. S. C. Nardelli and M. S. Twardowski, "Assessing the link between chlorophyll concentration and absorption line height at 676 nm over a broad range of water types," Opt. Express **24**(22), A1374–A1389 (2016).
70. A. Bricaud, M. Babin, A. Morel, and H. Claustre, "Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization," J. Geophys. Res. Oceans **100**(C7), 13321–13332 (1995).
71. A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Claustre, "Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models," J. Geophys. Res. Oceans **103**(C13), 31033–31044 (1998).
72. D. Müller, H. Krasemann, R. J. Brewin, C. Brockmann, P. Y. Deschamps, R. Doerffer, N. Fomferra, B. A. Franz, M. G. Grant, S. B. Groom, F. Mélin, T. Platt, P. Regner, S. Sathyendranath, F. Steinmetz, and J. Swinton, "The Ocean Colour Climate Change Initiative: I. A methodology for assessing atmospheric correction processors based on in-situ measurements," Remote Sens. Environ. **162**, 242–256 (2015).
73. D. Müller, H. Krasemann, R. J. Brewin, C. Brockmann, P. Y. Deschamps, R. Doerffer, N. Fomferra, B. A. Franz, M. G. Grant, S. B. Groom, F. Mélin, T. Platt, P. Regner, S. Sathyendranath, F. Steinmetz, and J. Swinton, "The Ocean Colour Climate Change Initiative: II. Spatial and temporal homogeneity of satellite data retrieval due to systematic effects in atmospheric correction processors," Remote Sens. Environ. **162**, 257–270 (2015).

74. A. Bracher, M. Taylor, B. Taylor, T. Dinter, R. Roettgers, and F. Steinmetz, "Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of phytoplankton pigments concentrations," Ocean Sci. **11**(1), 139–158 (2015).
75. C. Brockmann, R. Doerffer, M. Peters, S. Kerstin, S. Embacher, and A. Ruescas, "Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters," in *Proceedings of Living Planet Symposium*, L. Ouwehand ed. (ESA, 2016).
76. R. Doerffer and H. Schiller, "The MERIS Case 2 water algorithm," Int. J. Remote Sens. **28**(3–4), 517–535 (2007).
77. G. F. Cota, J. Wang, and J. C. Comiso, "Transformation of global satellite chlorophyll retrievals with a regionally tuned algorithm," Remote Sens. Environ. **90**(3), 373–377 (2004).
78. Z. Lee, K. L. Carder, and R. A. Arnone, "Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters," Appl. Opt. **41**(27), 5755–5772 (2002).
79. S. Maritorena, D. A. Siegel, and A. R. Peterson, "Optimization of a semianalytical ocean color model for global-scale applications," Appl. Opt. **41**(15), 2705–2714 (2002).
80. M. Hieronymi, D. Müller, and R. Doerffer, "The OLCI Neural Network Swarm (ONNS): A Bio-geo-optical Algorithm for Open Ocean and Coastal Waters," Front. Mater. Sci. **4**, 1–18 (2017).

1. Introduction

Phytoplankton, the main primary producers at the base of marine food web, have distinctive impacts on the changes of Arctic climate system. The Arctic region is warming at rates double than the global average, coinciding with persistent sea ice decline [1]. Seasonal sea ice retreat favors phytoplankton bloom development and the extension of phytoplankton growing season, increasing annual mean phytoplankton biomass and production [2,3]. This increase, in turn, is expected to further warm the ocean surface layer by absorbing more solar radiation and triggering additional positive feedbacks, which could amplify Arctic warming by 20% [2]. Meanwhile, together with ice algae, the increasing phytoplankton stocks generate more dimethylsulphide, a trace gas that provides 80% of global biogenic atmospheric sulphur [4]. When released to the atmosphere, dimethylsulphide, via formation of sulphate aerosol, cool the Arctic atmospheric temperature by dispersing solar radiation [5–7].

Despite the climatic role of Arctic phytoplankton, ship-based *in situ* measurements of total chlorophyll *a* concentration (Chl-a), a universal proxy of phytoplankton biomass, are sparse in the harsh Arctic Ocean. *In situ* Chl-a can be measured either through *in situ* fluorometers or High-Performance Liquid Chromatography (HPLC). The former method is relatively simple and inexpensive, but can cause inaccuracies due to natural variations in the fluorescence to Chl-a ratio [8], the fluorescence of colored dissolved organic matter (CDOM) [9], and the non-photochemical quenching [10]. HPLC is more accurate, but requires more intensive labor, time, money and complex analysis [11–13]. In addition, HPLC Chl-a measurements are based on discrete water samples collected on board, therefore, greatly limited by low repeat frequency and spatial coverage.

Satellite remotely sensed Chl-a is one of the most appropriate data sets to study long-term phytoplanktonic variability. However, ocean color data of the Arctic Ocean suffer from poor spatio-temporal coverage and resolution because of the heavy clouds and fog, the prevailing low solar elevations in high latitudes and the presence of sea ice etc [14]. Furthermore, standard ocean color Chl-a algorithms may not be sufficient to account for the bio-optical heterogeneity within the Arctic Ocean [15]. In the areas subjected to freshwater inputs, terrigenous CDOM can cause an overestimation of Chl-a when using standard algorithms [16–18]. In addition, phytoplankton tend to increase Chl-a per cell to absorb more light under low light conditions. Thus, pigment packaging is enhanced and Chl-a specific absorption coefficient is lowered, causing underestimation of satellite retrieved Chl-a [16–20]. Therefore, Arctic satellite ocean color Chl-a data require to be validated, improved and supplemented with substantial volumes of *in situ* observations.

In 2016, the Sentinel-3 Ocean and Land Color Instrument (OLCI) was launched as the successor of the Medium Resolution Imaging Spectrometer (MERIS) (mission between 2002 and 2012). It has a full spatial resolution of approximately 300 m and a swath width of 1270 km. OLCI Level 2 ocean color products were first released in July 2017. A second OLCI sensor is launched in April 2018, followed by six more OLCI sensors operated until the late 2030s. To date, OLCI Chl-a retrievals have not yet been evaluated in the Arctic region. There is an urgent need for validation data for the new Sentienl-3 OLCI Chl-a product as well as other products.

The Fram Strait, the deepest gateway connecting the Arctic Ocean to the North Atlantic Ocean, is one of the most climate-relevant ocean passages. Water masses exchange here along with salt and heat, including southward transportation of Arctic cold fresher water and sea ice to the Atlantic by the East Greenland Current, which accounts for nearly all of the Arctic sea ice export to the Atlantic Ocean [21], and northward inflow of Atlantic warm saltier water to the Arctic Ocean carried by the West Spitsbergen Current [22,23]. Within the context of climate change, these processes have been enhanced over the past decades. Observations in the Fram Strait revealed increased warming [24] and a larger amount of Arctic freshwater and sea ice export [25–27], which further accelerated North Atlantic freshening and promoted Arctic ice melt. The consequent effects on planktonic community patterns might be inferred from the shift of dominated phytoplankton assemblages from diatoms to other smaller nano- or picoplankton species [28,29], together with an increase of warm-adapted zooplankton in the eastern Fram Strait [30]. Model simulations revealed that primary productivity differs in zones influenced by different water masses in the Fram Strait [31–34]. Recently, a time series study of summertime Chl-a in eastern Fram Strait showed increasing Chl-a trend over 20 years using combined analysis of satellite ocean color and discrete *in situ* measurements [29,35]. However, further long-term *in situ* investigations are needed as ground truth for the validation of new satellite sensors (e.g. OLCI) and evaluation of Chl-a algorithms in order to obtain consistent time series data by integrating different satellite sensor measurements.

The shipboard underway spectrophotometry represents a promising *in situ* Chl-a observation technique. It utilizes a WET Labs AC-S hyperspectral spectrophotometer (or its former alternative, the 9-wavelength resolved AC-9) that is operated in flow-through mode to derive particulate absorption coefficients (a_p) (for simplicity, the wavelength dependency of a_p is omitted in the context) [36–47]. Those a_p data are calculated by differencing measurements from temporally adjacent 0.2-µm filtered and whole water samples. Instrument drift is removed during the subtraction, which is difficult to be accounted for when routine calibrations are not possible [39]. Via empirical relationships between a_p and Chl-a, continuous surface Chl-a data along cruise tracks can be obtained. This technique considerably facilitates the collection of *in situ* Chl-a measurements with unprecedented temporal and spatial resolution due to high sampling frequency, low power consumption, and the cost-effectiveness of the AC-S or AC-9 instrument and other components of the flow-through system. While the applicability of this technique to derive highly resolved Chl-a has been evaluated on an equatorial transect [39], in Equatorial Pacific, North Atlantic, Mediterranean Sea and Subarctic Northeast Pacific [38], and during the cruises of Tara Oceans [40,46] and the Atlantic Meridional Transect [42], it has never been used in the Fram Strait.

Here, we evaluate the capability of obtaining highly resolved surface Chl-a using the shipboard underway AC-S flow-through system in the Fram Strait. To achieve a detailed understanding on the quality of the Chl-a data sets of several satellite sensors and various algorithms, the retrieved Chl-a data from AC-S a_p measurements are used to validate (1) OLCI Level 2 Chl-a operational products, (2) OLCI Level 2 Chl-a products processed with Polymer atmospheric correction algorithm (version 4.1) [48,49], and (3) Level 2 Chl-a operational products from NASA Aqua Moderate Resolution Imaging Spectroradiometer (MODIS-A).

2. Data and methods

Data were collected during two cruises onboard *R.V. Polarstern* to the Fram Strait: the PS93.2 cruise from 20 July to 14 August in 2015 and the PS99.2 cruise from 23 June to 16 July in 2016. Both cruises followed similar tracks, ranging from approximately latitudes 72°to 80°N and longitudes 10°W to 15°E (Fig. 1).

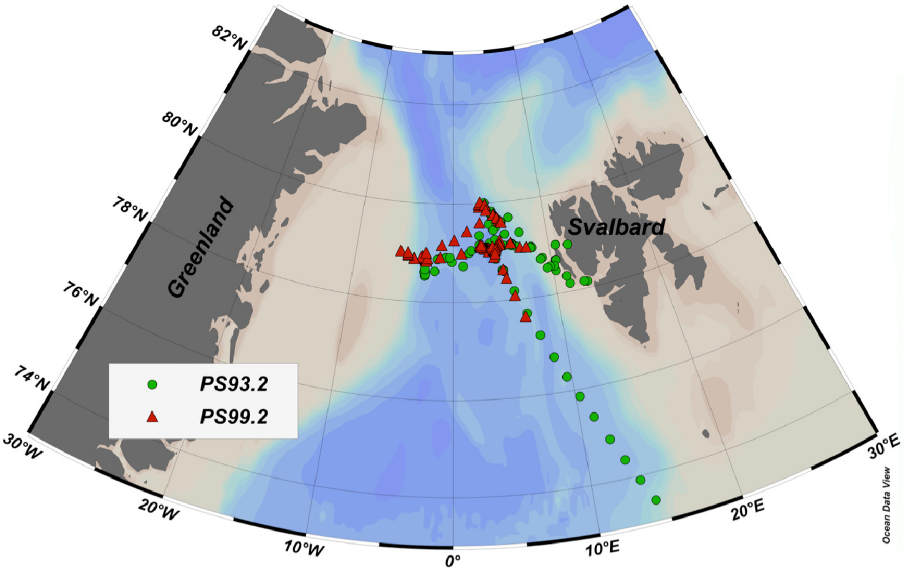


Fig. 1. Cruise tracks for PS93.2 (July-August 2015) and PS99.2 (June-July 2016) for underway HPLC data.

2.1 Underway AC-S measurements

A 25 cm-pathlength AC-S was integrated into the shipboard flow-through system to measure the hyperspectral absorption and beam attenuation coefficients over a spectral range of 400 – 740 nm. The effective spectral resolution (FWHM) is 10 nm and the sampling rate equals ~4 Hz with ~3.5 nm wavelength resolution.

The AC-S flow-through system was set up following Slade et al. [39]. The AC-S instrument was mounted to the seawater supply from the ship's membrane pump, with keel intake at roughly 11 m below the sea surface. The pumped seawater with a flow rate of 1 to 2 L min^{-1} first passed a de-bubbler (4H Jena, Germany) to reduce air bubbles. Via an electronically actuated valve controller (Isitec, Germany), the de-bubbled seawater was diverted either directly to the AC-S to measure the total non-water absorption and beam attenuation for 50 minutes per hour, or through a 0.2-um cartridge filter (Sartobran P, Germany) and then to the AC-S for the remaining 10 minutes each hour to provide a baseline for particulate absorption measurements. The AC-S instrument was cleaned every 1-2 days for blank measurements. The filter cartridges were replaced approximately once per week. The system was operated constantly except when anchoring the ship in ports, during daily cleaning of the AC-S tubes or while replacing the cartridge filter.

AC-S data were processed following the procedures adapted from Slade et al. [39]. Briefly, AC-S data were successively de-spiked, visually checked and 1-min median binned. Temperature and salinity dependency of pure water absorption were accounted for following Sullivan et al. [50]. The absorption coefficients measured over the filtering period was interpolated to the time of the total absorption measurements and finally subtracted to obtain a_p and particulate attenuation coefficient (c_p). Scatter and residual temperature correction for

a_p were performed following Slade et al. [39] to correct the overestimation of a_p due to the incomplete collection of the scattered light in the AC-S a-tube, and concurrently to account for the temperature differences between the filtered and unfiltered seawater and between the samples in the AC-S tubes and the thermosalinometer. The proportional scatter correction approach from Zaneveld et al. [51] was adopted.

2.2 Discrete water sampling

To validate the AC-S measurements, approximate 5 L of water samples were collected from the unfiltered AC-S outflow for later laboratory analysis of phytoplankton pigment composition and concentration and a_p. These samples were filtered on 25 mm (or 47 mm only for a_p measurements during PS93.2) diameter Whatman GF/F glass fiber filters (nominal pore size of 0.7 µm).

2.2.1 Phytoplankton pigment analysis

Filters for phytoplankton pigment composition and concentration analysis were immediately shock-frozen in liquid nitrogen and then stored at −80°C until further analysis at the laboratories of the Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research (AWI).

Pigment composition was analyzed by an HPLC system comprising a Waters 600 controller combined with a Waters 2998 photodiode array detector, a Waters 717plus auto-sampler and a LC Microsorb C8 HPLC column. A list of 23 pigments shown in Table 2 of Taylor et al. [52] were separated and quantified following an adjustment of the method described in Barlow et al. [53], as detailed in Taylor et al. [52]. The pigment data were quality assured according to Aiken et al. [54]. Chl-a was calculated as the summed contribution of monovinyl chlorophyll *a*, chlorophyllide *a* and divinyl chlorophyll *a* concentrations. Four phytoplankton groups, namely diatoms, green algae, prymnesiophytes (haptophytes) and prokaryotes, were identified and determined from their biomarker pigments using Diagnostic Pigment Analysis (DPA) [55–58] that is applied to a large global pigment data set by Losa et al. [59] to obtain the specific weights to convert the diagnostic pigment concentrations to Chl-a contributed by the aforementioned phytoplankton groups.

2.2.2 Particulate absorption measurements

The Quantitative Filter Technique (QFT) measurements of spectral light absorption were performed differently over the two cruises. Filters collected during PS93.2 were stored at −80°C after being immediately shocked-frozen in liquid nitrogen until analysis at AWI. Measurements were carried out on a dual-beam UV/VIS spectrophotometer (Cary 4000, Varian Inc.) equipped with a 150 mm integrating sphere (external DRA-900, Varian, Inc. and Labsphere Inc., made from SpectralonTM) following the method described in Simis et al. [60]. The filters were placed in the center of the integrating sphere and scanned in the range from 300 to 850 nm with the wavelength resolution of 1 nm. The baseline was recorded using a dry blank filter and a blank filter that was soaked in freshly produced Milli-Q water for more than 30 minutes. Optical density ($OD(\lambda)$) was measured and transformed to transmittance ($T(\lambda)$) following Eq. (1), which was finally used to calculate a_p following Eq. (2) using a path length amplification factor of 4.5 ($\beta = 1/4.5$) [61].

$$T(\lambda) = e^{-OD(\lambda)} \tag{1}$$

$$a_p(\lambda) = -\ln(T(\lambda) \times A \times \beta \times V^{-1}) \tag{2}$$

where V is the filtrated sample volume in m^3, A the filter clearance area in m^2 and λ wavelength in nm. The calculated a_p is in m^{-1}.

Filters collected during PS99.2 were immediately measured on board with QFT using a small portable integrating cavity absorption meter (QFT-ICAM) to avoid the artifacts and uncertainties from sample preservation and transport. Instrument setup and measurement procedure were detailed in Röttgers et al. [62]. Briefly, QFT-ICAM is made up of an 80 mm-diameter integrating spherical cavity with the highly reflective white PTFE walls, a CF-1000-HC lamp (Illumination Technology, USA) with an integrated filter wheel and a photodiode array detector (AVASPEC-ULS2048-RS-USB2, Avantes, the Netherlands). These three components are connected by two quartz-glass optical fibers. There are two thin Nylon strings in the middle plane across the whole cavity to hold the filters. Sample and reference filters were measured inside the integrating cavity by being moved in and outside of the light beam. A scan from 300 nm to 850 nm with the wavelength resolution of 0.3 nm was performed for each measurement. Dark currents were determined before filter measurements. a_p was calculated following Eqs. (1) and (2) using a path length amplification of 4.06 (β = 1/4.06) [62].

2.3 Satellite Chl-a validation

In situ Chl-a data derived from both AC-S and HPLC were used to validate satellite-derived Chl-a products. MODIS-A Level 2 operational products (R2014.0) retrieved from OCI algorithm were downloaded from the Ocean Color Website (http://oceancolor.gsfc.nasa.gov/) and analyzed for both *in situ* data sets of PS93.2 and PS99.2.

OLCI Level 2 Chl-a products were also validated with the PS99.2 data set, since OLCI started measuring in 2016. In this case, Chl-a products retrieved from Case 2 Regional CoastColour (C2RCC, version 0.15) and OC4 standard processor algorithms were used. OLCI Level 2 products (reprocessed data set REP NT_002 IPF-OL-2, version 06.08) were provided by EUMETSAT (https://www.eumetsat.int) in the context of Sentinel-3 Validation Team (S3VT) project.

In addition, three Chl-a data products produced from OLCI remote sensing reflectance data using Polymer atmospheric correction algorithm (version 4.1) were also evaluated because of their increased coverage as compared to the standard products. The Polymer algorithm provides a powerful atmospheric correction for ocean color data in the presence of contamination from sun glint, thin clouds heavy aerosol plumes or adjacency effect; these contaminated conditions are often not correctly treated by standard atmospheric correction schemes with extrapolation from the near infrared (e.g., used for MODIS-A OCI and OLCI OC4) [48,49]. We processed OLCI Level 1 products (reprocessed data set REP NT_002 IPF-OL-1-EO, version 06.06) from EUMETSAT with the Polymer algorithm (version 4.1) (provided at www.hygeos.com/Polymer). We used the Chl-a product retrieved from the Polymer atmospheric correction processing, which is an iterative spectral matching method using OLCI bands from 412 to 865 nm. This method relies on two simple models: a three-parameter model for the atmosphere and a two-parameter model for the ocean. The latter is based on a backscattering and an absorption term that is represented by Chl-a. Therefore, Polymer produces not only normalized spectral water reflectance but also Chl-a estimates. In this configuration, no adjustment of the radiometric calibration has been applied: the gain is set to 1 at all bands. Additionally, we applied the OCI and OC4 algorithms to the OLCI Polymer remote sensing reflectance to derive Chl-a products, which helps to evaluate specifically the effect of atmospheric correction methods on the different OLCI and MODIS-A products. Here we denote the OC4 algorithm with standard atmospheric correction as OC4-Operational, OCI and OC4 algorithm with Polymer atmospheric correction as OCI- and OC4-Polymer, respectively, and Polymer standard algorithm as Polymer-Standard.

When processing OLCI data, different bitmask settings were applied to approve only pixels measured under optimal conditions. S3VT recommends bitmask settings for the OC4 standard product as follows: [invalid, land, cloud, cloud_ambiguous, cloud_margin, snow_ice, suspect, hisolzen, saturated, highglint, whitecaps, ac_fail, oc4me_fail, annot_tau06, rwneg_o2, rwneg_o3, rwneg_o4, rwneg_o5, rwneg_o6, rwneg_o7, rwneg_o8] and for the C2RCC product as follows: [invalid, land, cloud, cloud_ambiguous, cloud_margin, snow_ice, suspect, hisolzen, saturated, highglint, whitecaps, ocnn_fail]. For the Polymer output, a [thick_aerosol, cloud_base] bitmask was applied.

In situ Chl-a data sets were then matched to satellite derived Chl-a values at the same day for MODIS-A 1x1 and OLCI 3x3 pixel windows following the procedure of MERMAID (MERIS Matchup In situ Database) [63]. Different pixel window sizes were chosen because of the relatively comparable spatial resolutions of 1 km for MODIS-A 1x1 and 900 m (3x300 m) for OLCI 3x3 pixel windows. Both MODIS-A and Sentinel-3 orbits are near-polar orbits. In polar regions, it is more possible to acquire multiple satellite retrievals for a given *in situ* record due to the increasing overlap in adjacent swaths. In this case, the multiple satellite data are considered as parallel measurements and included in the same pixel window. Satellite Chl-a data for which the median coefficients of variation with respect to log10 based Chl-a within one pixel window were greater than 0.15 were excluded to minimize the effect of mismatch in spatial scales of *in situ* and satellite data [64]. When more than one *in situ* measurements were matched to satellite data within the same pixel window, the *in situ* Chl-a values were averaged. Table 1 shows an overview over all the algorithms that are used to derive Chl-a products.

Table 1. Satellite Chl-a algorithms being evaluated for the cruises PS93.2 and PS99.2.

Cruises	Satellite Sensor	Atmospheric Correction	Chl-a algorithm	Denotation
PS93.2	MODIS-A	Standard	OCI	OCI
PS99.2	MODIS-A	Standard	OCI	OCI
	OLCI	Standard	OC4	OC4-Operational
		C2RCC	C2RCC	C2RCC
		Polymer	Polymer	Polymer-Standard
		Polymer	OC4	OC4-Polymer
		Polymer	OCI	OCI-Polymer

2.4 Statistics

We used a set of statistical tests following Brewin et al. [65] and Sá et al. [66] to evaluate the performance of (1) the AC-S data correction scheme, (2) AC-S based Chl-a retrieval models and (3) satellite Chl-a algorithms. These tests include the determination coefficient (R^2), the Pearson correlation coefficient (r), the root mean square error ($RMSE$), the mean absolute error (MAE), the bias error (δ), and the unbiased root mean square error (Δ), the mean Relative Percentage Difference (RPD) and the slope (S) and the intercept (I) of Type-2 regression. For all linear regressions, Type-2 regression analysis was used (MATLAB function lsqfitma.m) [67]. The equations for these statistical metrics are given below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_i^E - C_i^M)^2} \tag{3}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|C_i^E - C_i^M\right| \tag{4}$$

$$\delta = \frac{1}{N}\sum_{i=1}^{N}\left(C_i^E - C_i^M\right) \tag{5}$$

$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(C_i^E - \mu^E) - (C_i^M - \mu^M)]^2} \tag{6}$$

$$RPD = \frac{1}{N}\sum_{i=1}^{N}\frac{C_i^E - C_i^M}{C_i^M}\times 100 \tag{7}$$

where N is the total number of samples, i is the sample index, C is Chl-a (log-transformed except in (7)) and μ is the averaged C. The superscript E denotes the estimated value (e.g. from satellite data) and the superscript M the measured value (e.g. from *in situ* data).

3. Results and discussion

3.1 Chl-a estimation from AC-S

In this study, particulate absorption line height at 676 nm ($a_{LH}(676)$) is used to retrieve Chl-a. $a_{LH}(676)$ is chosen because it provides good estimates of Chl-a with lower contribution of accessory pigments and weaker packaging effect [68, 69]. $a_{LH}(676)$ was calculated from the particulate absorption coefficient at 676 nm above the baseline from 650 nm to 715 nm following Eq. (8) [68]. For both cruises, $a_{LH}(676)$ derived from the AC-S and filter-pad data are linearly related (for PS93.2, $r = 0.94$, $RMSE = 0.008$; for PS99.2, $r = 0.90$, $RMSE = 0.009$).

$$a_{LH}(676) = a_p(676) - \frac{a_p(715) - a_p(650)}{715 - 650}\times(676 - 650) - a_p(650). \tag{8}$$

AC-S derived $a_{LH}(676)$ calculated from Eq. (8) was related to HPLC Chl-a data for the two cruises in order to derive Chl-a from all the AC-S data. $a_{LH}(676)$ was log10 transformed and averaged within the period of 10 minutes before and after HPLC sampling time [42]. It was then back-transformed and related to HPLC Chl-a data by the robust fitting of the power function of Eq. (9) [70,71] using bisquare weighting scheme:

$$Chl-a = Aa_{LH}(676)^B. \tag{9}$$

Table 2 portrays the regression coefficients and statistics for Chl-a to $a_{LH}(676)$ relationships. $a_{LH}(676)$ and HPLC Chl-a data were both log10-transformed for the calculation of the statistical metrics. For PS93.2, the coefficients A and B are 86.1 ± 11.1 and 1.00 ± 0.03, respectively; for PS99.2, they are 36.2 ± 6.78 and 0.93 ± 0.05, respectively. The fits are shown in Fig. 2(a)-2(b). B value from PS93.2 is close to 1, indicating a linear relationship between Chl-a and $a_{LH}(676)$. A value from PS93.2 is over twice greater than that from PS99.2. For comparison, the relationships between Chl-a and filter-pad a_{LH} (676) were also derived (Table 2). The different fits obtained from AC-S and filter-pad data could be attributed to (1) varying degrees of sample-to-sample β variations for filter-pad data, (2) sensitivity of the calculated absorption coefficient to filter clearance area (1 mm measurement error for diameter causes 0.008 m^{-1} error in absorption coefficient), and (3) a wider spectral bandwidth of AC-S (10 nm) compared to Cary and QFT-ICAM (approximately 2 nm) resulting in flattening absorption peaks with AC-S data. Filter-pad data measured by Cary and QFT-ICAM can also differ from each other because samples for QFT-ICAM were immediately measured after filtration, whereas samples for Cary were stored at −80°C before finally measured. Nevertheless, the good R^2, $RMSE$ and MAE values support the use of AC-S flow-through system to obtain Chl-a with high quality.

Given the good HPLC Chl-a to AC-S $a_{LH}(676)$ relationships found for both cruises, coefficients A and B were then applied to all the AC-S derived $a_{LH}(676)$ data to retrieve continuous surface Chl-a data sets along the cruise tracks. Figure 3 shows the comparison between Chl-a derived from the underway AC-S flow-through system and HPLC measurements for both cruises. The two Chl-a data sets showed good consistency, with S and I being 0.97 ± 0.02 and 0.03 ± 0.02, and the log10 based r, $RMSE$ and MAE being 0.95, 0.097 and 0.064, respectively. In total, 24424 and 16110 Chl-a data were generated for PS93.2 and PS99.2, ranging from 0.179 to 3.550 mg m^{-3} (mean value: 1.042 mg m^{-3}) and from 0.003 to

2.701 mg m^{-3} (mean value: 0.677 mg m^{-3}), respectively. These values are representative for the summertime Chl-a variability observed in the Fram Strait [34]. The histograms of both *in situ* Chl-a distribution (Fig. 2(c) and 2(d)) and time series of Chl-a (Fig. 4(a) and 4(b)) show that Chl-a data derived from AC-S exhibit a general agreement with those from HPLC, however, yet much more highly resolved during both cruise periods.

Table 2. Regression coefficients and statistics for HPLC Chl-a and *in situ* a_{LH}(676) power function relationships for the cruises PS93.2 and PS99.2. The uncertainties of the regression coefficients *A* and *B* were calculated with 95% confidence bounds.

Cruise	Measuring method	*A*	*B*	R^2	*RMSE*	*MAE*	*N*
PS93.2	AC-S	86.1 ± 11.1	1.00 ± 0.03	0.90	0.079	0.051	134
	Cary	62.8 ± 9.85	1.05 ± 0.04	0.82	0.114	0.070	134
PS99.2	AC-S	36.2 ± 6.78	0.93 ± 0.05	0.90	0.120	0.083	84
	QFT-ICAM	57.5 ± 10.2	0.91 ± 0.04	0.90	0.133	0.089	104

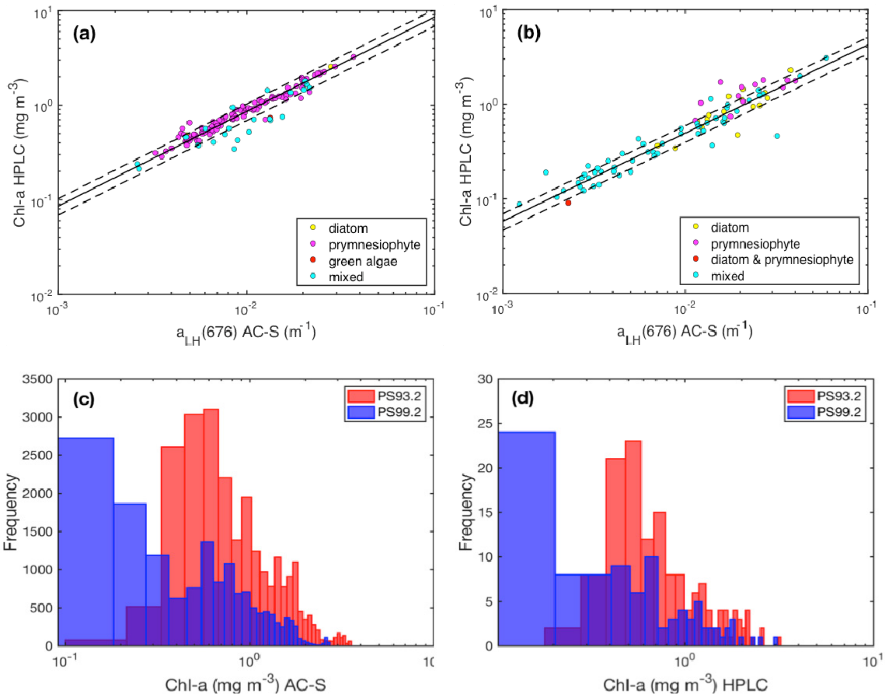


Fig. 2. (a) and (b) show the relationships between the collocated a_{LH}(676) derived from the underway spectrophotometry and HPLC Chl-a data for PS93.2 and PS99.2, respectively. The solid lines are power fits according to Table 2. The dash lines are ± 20% error lines. Different colors filled in the circles represent the dominating phytoplankton groups by diatom, prymnesiophyte, both diatom and prymnesiophyte or green algae, while the term “mixed” denotes a mixture of phytoplankton groups with no dominating group; (c) and (d) show frequency distribution of AC-S derived and HPLC Chl-a.

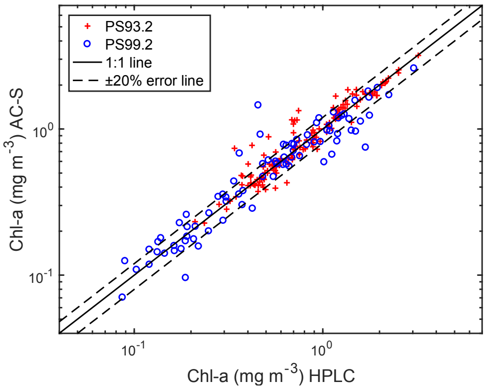


Fig. 3. Comparison between Chl-a data obtained from underway spectrophotometry and HPLC measurements.

The differences in the HPLC Chl-a to AC-S $a_{LH}(676)$ relationships as well as the scatter around the regression lines between the two cruises suggested an inter- and intra-cruise variability of chlorophyll-specific absorption line height. The relationship for PS99.2 exhibited larger variations than that for PS93.2 as inferred from greater *RMSE* and *MAE* values. This may originate from the different times of the years being sampled that favor different phytoplankton growth conditions: PS93.2 was conducted from late July to mid August and had a larger range and higher mean value of Chl-a data (see above); PS99.2 was conducted from late June to mid-July and also contained more very low Chl-a values (Fig. 2(c) and 2(d)). In addition, there were clear shifts of dominating phytoplankton groups (defined here as the fraction of a certain group calculated using DPA greater than 45%): Prymnesiophytes were dominant in 82.8% of the surface waters of PS93.2, while only 1.5% are diatom-dominated and 14.9% of the water samples were mixed without apparent dominating phytoplankton groups. In contrast, the fractions changed to 6.0%, 16.7% and 76.2% for PS99.2, respectively. Figure 2(b) shows that the data points outside of the ± 20% error lines can be diatom-, prymnesiophytes-dominated or phytoplankton-mixed samples. Further calculation showed that 73.1% of the "mixed" samples have a proportion of diatoms over 30%. According to Nöthig et al. [29], in the Fram Strait in summertime, prevailing diatoms are *Thalassiosira* spp and *Chaetoceros* spp, mainly *C. socialis*, *Fragilariopsis oceanica* and other pennate diatoms, whereas the main species of prymnesiophytes is in most cases *Phaeocystis pouchetii*. Therefore, we suspected that the more diverse phytoplankton composition observed in PS99.2 samples involved greater variations in cell size and pigment composition and hence, pigment packaging, contributing to the relatively larger discrepancy in the relationship of HPLC Chl-a data and AC-S derived $a_{LH}(676)$. Additional variability of these relationships is expected due to the absorption in the line height region by non-algal particles and CDOM [68], and to the assumption that CDOM absorption during non-filtering period of the underway AC-S flow-through system linearly varied in the AC-S data correction scheme. The amount of non-algal particles and CDOM may vary strongly between the period of phytoplankton bloom development and degradation.

3.2 Satellite Chl-a validation

The reconstructed *in situ* Chl-a data sets from the AC-S flow-through system along the cruise tracks were directly compared with the spatial and temporal coincident satellite Chl-a products. *In situ* Chl-a data were used to validate Chl-a operational products retrieved from MODIS-A OCI operational algorithm for PS93.2 and from MODIS-A OCI, OLCI OC4 and C2RCC operational algorithms for PS99.2. Furthermore, Polymer-Standard, OCI- and OC4-Polymer algorithms were compared and assessed. In addition, validation of the same Chl-a products and assessment of the above algorithms using HPLC Chl-a were also performed for comparison. Summary statistics are presented in Table 3. The *r*, *RMSE*, *δ* and Δ were calculated in log10 space, while *RPD* was calculated in linear space. Figure 5 shows the density plots of the validation results.

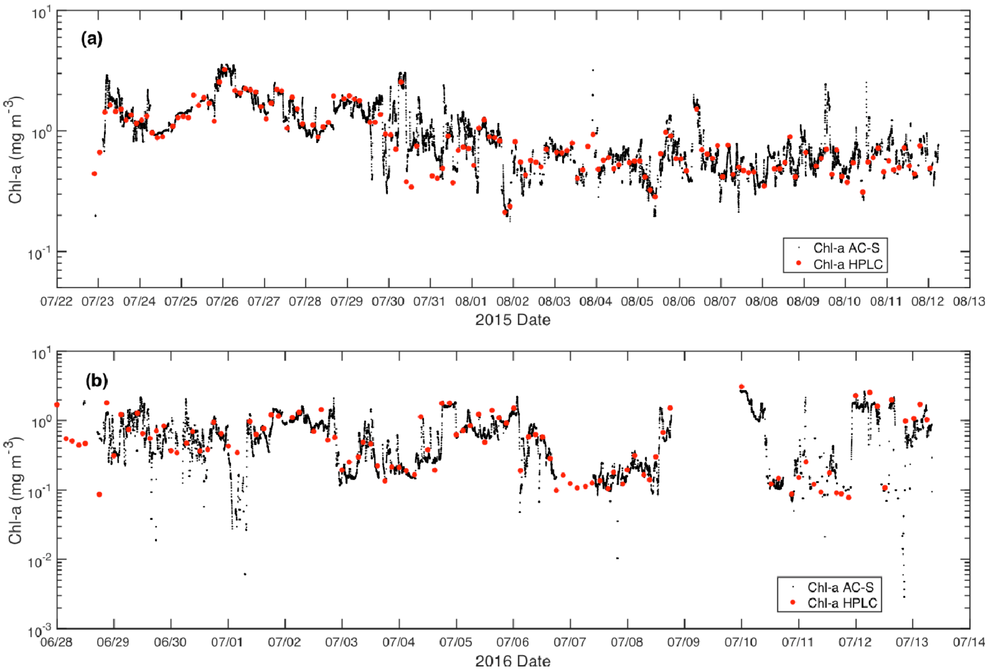


Fig. 4. Time series of AC-S derived and HPLC Chl-a data during the cruise period of (a) PS93.2, and (b) PS99.2.

As is shown in Table 3, the number of match-ups generated by AC-S derived to satellite Chl-a (hereinafter denoted as “AC-S matchups”) is over one order of magnitude greater than those from HPLC to satellite Chl-a (hereinafter denoted as “HPLC matchups”). For PS93.2, the number of AC-S matchups for MODIS-A is 80, in contrast with only 8 HPLC match-ups. For PS99.2, the numbers of AC-S matchups from the three Polymer atmospheric correction method based algorithms (i.e. OLCI Polymer-Standard, OCI- and OC4- Polymer algorithms) and OLCI C2RCC algorithm are two to three times greater than those by the operational MODIS-A OCI and OLCI OC4 algorithms. This is because OLCI C2RCC and Polymer are the atmospheric correction processors that incorporate the contributions of sun glint and thin clouds in their reflectance models to derive atmospheric corrected remote sensing reflectance, allowing for much larger coverage of data [72,73]. MERIS Polymer products have shown to improve the spatial coverage by almost a factor of two [74] and have proven successful for retrieving (MERIS) ocean color products: Polymer was selected as the MERIS processor for atmospheric correction for the Ocean Color Climate Change Initiative (OC-CCI) after an extensive validation and intercomparison with other atmospheric correction algorithms in which each algorithm’s uncertainty was assessed [73].

When considering the correlation between *in situ* and satellite data sets, for PS99.2, both AC-S derived and HPLC Chl-a data were reasonably correlated with the corresponding satellite Chl-a retrieved from the operational band ratio algorithms (i.e. OLCI OC4-Operational and MODIS-A OCI) ($r > 0.87$ and 0.79, respectively) and the three Polymer based algorithms ($r > 0.94$ and 0.96, respectively). However, the correlation coefficient significantly decreased when comparing Chl-a data from MODIS-A OCI algorithm to AC-S Chl-a for PS93.2 ($r = 0.25$). The decrease is even more significant for OLCI C2RCC algorithm ($r = 0.13$). Furthermore, the regression slopes for C2RCC algorithm ($S = 0.07$ and

0.11 with respect to AC-S and HPLC measurements, respectively) are far from the unity. A reason could be that those satellite Chl-a have less dynamic range than other satellite Chl-a products (Fig. 5(g) and 5(h)). Nevertheless, the density plots of OLCI C2RCC match-up analysis showed that a considerable proportion of match-up points were around the one-to-one line (Fig. 5(g) and 5(h)). For PS93.2, the regression slopes for MODIS-A OCI Chl-a to both AC-S and HPLC based data sets are 1.04 and 1.14, respectively. For PS99.2, except for OLCI C2RCC and OC4-Operational products, the slopes for all the other satellite algorithms for both AC-S and HPLC match-ups are close to or greater than the unity. For example, the slope for OLCI Polymer-Standard to HPLC Chl-a is 0.90 and to AC-S derived Chl-a 1.61, whereas for OC4-Operational they are 0.60 and 0.75, respectively.

Table 3. Summary statistics of linear regression analysis between *in situ* and coincident satellite Chl-a data.

Cruise	Satellite Sensor	Algorithm	*In situ*	*N*	*S*	*I*	*r*	*RMSE*	δ	Δ	*RPD (%)*
PS93.2	MODIS-A	OCI	AC-S	80	1.04	–0.08	0.25	0.302	–0.056	0.297	13
		OCI	HPLC	8	1.14	–0.41	0.92	0.380	–0.258	0.280	–34
PS99.2	MODIS-A	OCI	AC-S	512	3.33	–0.23	0.95	0.473	0.445	0.160	199
		OCI	HPLC	91	2.60	–0.09	0.79	0.449	0.402	0.200	47
	OLCI	Polymer-Standard	AC-S	2706	1.61	0.11	0.94	0.392	0.332	0.207	137
		Polymer-Standard	HPLC	257	0.90	0.16	0.96	0.229	0.118	0.196	43
		OCI-Polymer	AC-S	2077	1.98	–0.02	0.95	0.395	0.345	0.192	140
		OCI-Polymer	HPLC	213	1.12	–0.03	0.98	0.143	–0.061	0.129	19
		OC4-Polymer	AC-S	3396	3.38	–0.65	0.95	0.461	0.430	0.166	185
		OC4-Polymer	HPLC	212	1.71	–0.25	0.98	0.196	0.123	0.152	39
		OC4-Operational	AC-S	789	0.60	0.29	0.87	0.395	0.274	0.284	125
		OC4-Operational	HPLC	97	0.75	0.27	0.97	0.474	0.381	0.282	186
		C2RCC	AC-S	2144	0.07	0.16	0.13	0.282	0.048	0.278	34
		C2RCC	HPLC	271	0.11	0.13	0.47	0.230	–0.092	0.211	–10

Considering all four uncertainty measures of *RMSE*, Δ, δ and *RPD* with respect to the *in situ* AC-S measurements, OLCI Polymer-Standard algorithm yielded the most reliable estimation of Chl-a, closely followed by OCI-Polymer algorithm. Both OLCI Polymer-Standard and OCI-Polymer algorithms showed small values of Δ (0.207 and 0.192, respectively), which indicates a good estimating precision. Moreover, the bias quantified in terms of δ (0.332 and 0.345, respectively) and *RPD* (137% and 140%, respectively) are relatively small, ensuring the reasonable accuracy for these two algorithms. OLCI OC4-Polymer and MODIS-A OCI (PS99.2) algorithms performed similarly, providing the smallest values of Δ (0.166 and 0.160, respectively), and yet, higher uncertainty in terms of δ, *RPD* and *RMSE*. OC4-Operational algorithm performed better than OLCI OC4-Polymer and MODIS-A OCI (PS99.2) algorithms in terms of the smaller *RMSE, δ, and RPD*, and yet, a large Δ. OLCI C2RCC and MODIS-A OCI (PS93.2) showed much smaller δ (0.048 and –0.056, respectively) and *RPD* (34% and 13%, respectively) than other algorithms. However, the data dispersion indicated by Δ is much higher for these two algorithms, inferring a less precise estimation of Chl-a. The lower bias measures are probably because that the positive and negative errors cancelled each other, as shown in Fig. 5(b), and 5(h), whereas all the other algorithms mostly provided positive errors.

Overall, the Polymer-Standard and OCI-Polymer products gave the most reliable results among the six considered satellite products. The good performances of OCI-Polymer and Polymer-Standard algorithms may be attributed to the successful atmospheric correction of adjacent effects due to snow/ice and clouds in the study area [48,49]. However, when compared to AC-S data, their overestimation indicated by *RPD* was one order of magnitude higher than MODIS-A OCI for PS93.2. For PS99.2, MODIS-A OCI showed similar high values of uncertainty as OLCI OC4-Polymer products. The OCI and OC4 algorithms are both empirical algorithms that use a three-band and a four-band blue-green reflectance ratio, respectively, to directly retrieve Chl-a. These data products performed well when evaluated using a global data set, probably due to their immunity to scale errors or instrument noise in remote sensing reflectance data [65]. The OLCI C2RCC products showed relatively low uncertainty and low bias (small negative/positive bias for HPLC/AC-S match-ups), but had yet the least correlation to the *in situ* data sets among all operational products. Though C2RCC is used to generate the Case 2 water products of Sentinel-3 OLCI standard ESA products, it is based on an artificial neural network trained for a wide range of atmospheric and ocean optical conditions [75,76]. Moreover, its Chl-a products are estimated using particulate absorption at 440 nm derived from remote sensing reflectance [75,76], making the C2RCC algorithm work reasonably well in our study area.

MODIS-A OCI Chl-a products had relatively low uncertainty for PS93.2 and tended to underestimate Chl-a data at relatively low Chl-a values indicated by the negative value of δ. It may be attributed to the dominance of prymnesiophytes in the surface waters of PS93.2 (Fig. 2(a)). *Phaeocystis pouchetii*, the most likely main species of prymnesiophytes in our study area (mentioned in 3.1,) [29], is able to form large colonies that have strong self-shading, lowering the specific absorption coefficient and causing the underestimation of satellite Chl-a retrievals. However, overall, MODIS-A OCI (PS93.2) overestimated Chl-a, especially at relatively high Chl-a values (Fig. 5(b)). As pointed out above for PS99.2, MODIS-A OCI and OLCI OC4-Operational and three Polymer based products were characterized by an overestimation, as positive δ and *RPD* displayed. However, there is a proportion of underestimated Chl-a from OLCI C2RCC products when compared to HPLC data (*RPD* = –10%) and overestimation (*RPD* = 34%) when compared to AC-S data. For PS99.2, the Chl-a overestimation was significant for OLCI OC4-Polymer and MODIS-A OCI (*RPD* ≥ 185%) and slightly less significant for OLCI Polymer-Standard, OCI-Polymer and OC4-Operational (100% ≤*RPD* ≤ 140%). The overestimation of Chl-a by band ratio algorithms is consistent with previous match-up analysis that Chl-a is overestimated by OC4 algorithm in the Arctic Ocean [14,77]. This is because CDOM concentrations in the Arctic Ocean are much higher than the global average, challenging the global empirical relationships underlying OCI and OC4 algorithms. It also explains the overestimation by the Polymer standard Chl-a product which is retrieved assuming a variability of the ocean model depending only on two parameters, the Chl-a and a backscattering coefficient, not accounting for additional CDOM variability [48,49], which implies a (global) constant relationship between phytoplankton and other absorbing water constituents, i.e. CDOM and non-algal particles. As a consequence, an excess of CDOM can be interpreted as an additional concentration of Chl-a.

For a given satellite algorithm, AC-S and HPLC match-ups have different uncertainty measures and regression parameters for both cruises. This is especially true for MODIS-A OCI Chl-a products for PS93.2 and OLCI C2RCC Chl-a products for PS99.2, where different trends (i.e. overestimation or underestimation) of satellite Chl-a estimates were observed. This might be because of the insufficient coverage of HPLC data points. The density plots (Fig. 5) show that spatial and temporal variability of satellite data is accounted for to a greater extent when doing match-up analysis against AC-S data set.

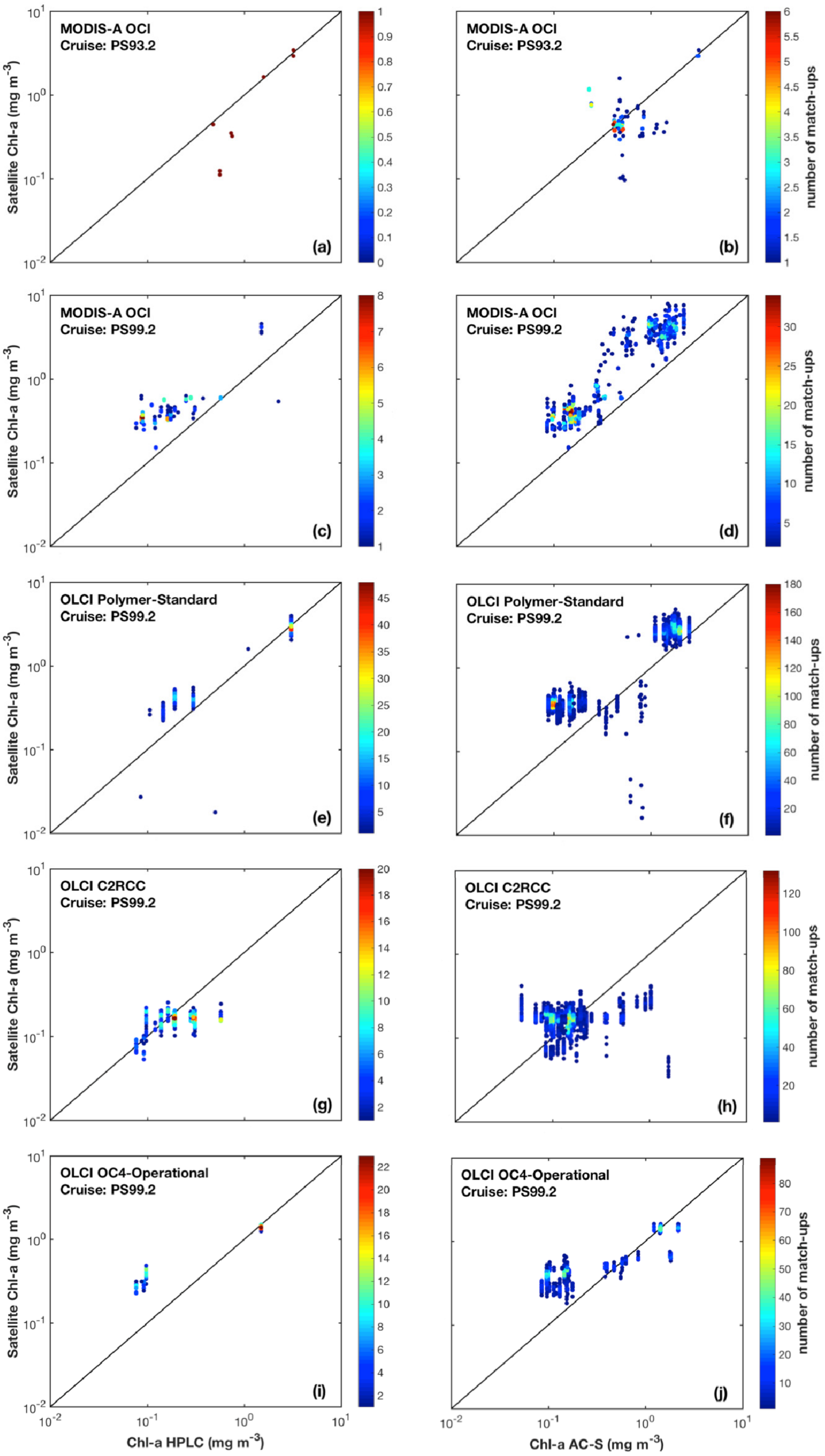


Fig. 5. Density plots of satellite to *in situ* match-up analysis.

In summary, validation results revealed acceptable performances of the operational band ratio algorithms (OCI, OC4) that are applied to the MODIS-A and OLCI sensors. Better results were achieved with the Polymer atmospheric correction algorithm applied to OLCI data. Overall, all the algorithms tended to overestimate Chl-a in the Fram Strait to various extent. The OLCI Polymer standard output provided the most reliable Chl-a estimates, and nearly as good results were obtained from the OCI-Polymer algorithm. The OC4-Operational algorithm performed slightly better than OC4-Polymer algorithm. The OLCI C2RCC Chl-a products showed relatively low uncertainty, but had yet the least correlation to the *in situ* data sets. MODIS-A OCI Chl-a products during the cruise period of PS93.2 did not correlate well to AC-S derived *in situ* AC-S derived Chl-a, but had yet relatively low uncertainty. For PS99.2, they showed similar high values of uncertainty as OLCI OC4-Polymer products. Further assessment of other types of Chl-a algorithms, e.g. semi-analytical methods (e.g., Quasi-analytical-algorithm (QAA) from Lee et al. [78] or Garver-Siegel-Maritorena (GSM) from Maritorena et al. [79]) or selection of neural network approaches based on water types [80] are necessary to be exploited to obtain optimized Chl-a data sets for the Fram Strait. Especially when Polymer atmospheric correction method is applied, high data coverage with low uncertainty can be expected.

4. Conclusion

We have shown the applicability of using the underway AC-S flow-through technique to continuously measure particulate absorption at high frequency in the Fram Strait. For the first time in this region, the relationships between AC-S derived particulate absorption and HPLC Chl-a were assessed during two summer cruises (PS93.2 in 2015 and PS99.2 in 2016). The good power law correlation enabled the estimation of continuous surface Chl-a along the entire cruise tracks. The continuous *in situ* data sets obtained from the underway AC-S flow-through system were used to validate satellite operational Chl-a products from MODIS-A OCI and OLCI OC4-Operational, C2RCC and Polymer-Standard algorithms, and to assess the performances of Polymer-Standard, OCI- and OC4-Polymer algorithms. For comparison, the validation of satellite operational products and assessment of different algorithms were also performed using HPLC measurements.

Statistics of linear regression analysis between *in situ* measurements and co-located satellite data for both cruises in the Fram Strait indicated reasonable performances of all algorithms. When considering AC-S match-ups, all algorithms were characterized by an overestimation of satellite Chl-a. However, there is underestimation of Chl-a for MODIS-A OCI (PS93.2) and OLCI C2RCC. This underestimation is also determined for HPLC match-ups. The OLCI Polymer-Standard and OCI-Polymer products had relatively high estimation precision and small bias with respect to both *in situ* AC-S and HPLC data sets, suggesting a successful atmospheric correction and the most reliable approximations of Chl-a data by OLCI Polymer-Standard and OCI-Polymer algorithm in the Fram Strait. When using HPLC data, the numbers of collocations are much lower, highlighting the capability of underway spectrophotometry to generate more sufficient surface Chl-a data sets for satellite Chl-a products validation and algorithms assessment in the Fram Strait. With the help of further automatic acquisition of AC-S data, the establishment of a long-term data record of satellite Chl-a data with determined uncertainties for this under-sampled remote area to be used in climate change research will be expected.

Funding

FRontiers in Arctic marine Monitoring (FRAM) project of Helmholtz Association Infrastructure Programme; Transregional Collaborative Research Center (TR 172) “ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)3” via the subproject C03, funded by the German Research Foundation

(DFG, Deutsche Forschungsgemeinschaft); China Scholarship Council (CSC); POLMAR Helmholtz Graduate School for Polar and Marine Research of the Alfred-Wegener-Institute.

Acknowledgments

We thank all the scientists, especially the chief scientist Thomas Soltwedel, the captains and crew on *R.V. Polarstern* for their support for collecting the *in situ* data used in this paper during the cruises PS93.2 and PS99.2. We acknowledge the discussions with Emmanuel Boss, Nils Haëntjens, Svetlana Loza and Hongyan Xi on AC-S data analysis and the interpretation of satellite validation. We thank Yashen Wang, Yanqun Pan, Jan Streffing, Xun Gong and Yuchen Sun for their suggestions in improving the codes for data processing. We thank Maryia Herasimava for helping quality control the HPLC data. We also thank Haozhuang Wang, Heng Wang, Saeid Bagheri Dastgerdi and Jan Streffing for their comments on English writing. We would also like to thank the Ocean Biology Processing Group of NASA for the processing and distribution of MODIS-A data and ESA/EUMETSAT for providing Sentinel-3 OLCI data used in the paper. We are grateful to the anonymous reviewers whose comments significantly improved the manuscript. All *in situ* data of our study are available at https://doi.org/10.1594/PANGAEA.885631.

Paper II

remote sensing

MDPI

Article

Retrieval of Phytoplankton Pigments from Underway Spectrophotometry in the Fram Strait

Yangyang Liu [1,2,*], Emmanuel Boss [3], Alison Chase [3], Hongyan Xi [1], Xiaodong Zhang [4,5], Rüdiger Röttgers [6], Yanqun Pan [7,8] and Astrid Bracher [1,9]

[1] Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, 27570 Bremerhaven, Germany; hongyan.xi@awi.de (H.X.); astrid.bracher@awi.de (A.B.)
[2] Faculty of Biology and Chemistry, University of Bremen, 28359 Bremen, Germany
[3] School of Marine Sciences, University of Maine, Orono, ME 04469, USA; emmanuel.boss@maine.edu (E.B.); alison.p.chase@maine.edu (A.C.)
[4] Division of Marine Science, University of Southern Mississippi, Stennis Space Center, MS 39529, USA; xiaodong.zhang@usm.edu
[5] Department of Earth System Science and Policy, University of North Dakota, Grand Forks, ND 58202, USA
[6] Helmholtz-Zentrum Geesthacht Center for Materials and Coastal Research, 21502 Geesthacht, Germany; ruediger.roettgers@hzg.de
[7] Department of Biology, Chemistry and Geography, Université du Québec à Rimouski, Rimouski, QC G5L 3A1, Canada; yanqun.pan@uqar.ca
[8] State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China
[9] Institute of Environmental Physics, University of Bremen, 28359 Bremen, Germany
* Correspondence: yangyang.liu@awi.de; Tel.: +49-471-4831-1785

Received: 21 December 2018; Accepted: 30 January 2019; Published: 6 February 2019

Abstract: Phytoplankton in the ocean are extremely diverse. The abundance of various intracellular pigments are often used to study phytoplankton physiology and ecology, and identify and quantify different phytoplankton groups. In this study, phytoplankton absorption spectra ($a_{ph}(\lambda)$) derived from underway flow-through AC-S measurements in the Fram Strait are combined with phytoplankton pigment measurements analyzed by high-performance liquid chromatography (HPLC) to evaluate the retrieval of various pigment concentrations at high spatial resolution. The performances of two approaches, Gaussian decomposition and the matrix inversion technique are investigated and compared. Our study is the first to apply the matrix inversion technique to underway spectrophotometry data. We find that Gaussian decomposition provides good estimates (median absolute percentage error, MPE 21–34%) of total chlorophyll-a (TChl-a), total chlorophyll-b (TChl-b), the combination of chlorophyll-c1 and -c2 (Chl-c1/2), photoprotective (PPC) and photosynthetic carotenoids (PSC). This method outperformed one of the matrix inversion algorithms, i.e., singular value decomposition combined with non-negative least squares (SVD-NNLS), in retrieving TChl-b, Chl-c1/2, PSC, and PPC. However, SVD-NNLS enables robust retrievals of specific carotenoids (MPE 37–65%), i.e., fucoxanthin, diadinoxanthin and 19′-hexanoyloxyfucoxanthin, which is currently not accomplished by Gaussian decomposition. More robust predictions are obtained using the Gaussian decomposition method when the observed $a_{ph}(\lambda)$ is normalized by the package effect index at 675 nm. The latter is determined as a function of "packaged" $a_{ph}(675)$ and TChl-a concentration, which shows potential for improving pigment retrieval accuracy by the combined use of $a_{ph}(\lambda)$ and TChl-a concentration data. To generate robust estimation statistics for the matrix inversion technique, we combine leave-one-out cross-validation with data perturbations. We find that both approaches provide useful information on pigment distributions, and hence, phytoplankton community composition indicators, at a spatial resolution much finer than that can be achieved with discrete samples.

Keywords: phytoplankton pigments; phytoplankton absorption; underway AC-S flow-through system; Gaussian decomposition; matrix inversion; singular value decomposition; non-negative least squares

---

1. Introduction

Phytoplankton account for approximately half of global primary production via photosynthesis [1] and form the base of the marine food web. Intracellular pigments of phytoplankton, composed of chlorophylls (a, b and c), carotenoids (carotenes and xanthophylls) and phycobiliproteins (phycoerythrin, phycocyanin and allophycocyanin) [2], play a vital role in photoprotection and the light-driven part of photosynthesis. Chlorophyll-b, -c and photosynthetic carotenoids (PSC), such as fucoxanthin (Fuco), act as antenna pigments that transfer the light energy to chlorophyll-a in the photosynthetic reaction centers of photosystems, assisting in light harvesting for photosynthesis. In cyanobacteria, red algae, and cryptophytes, phycobiliproteins are the major light-harvesting pigments [3]. Chlorophyll-a is crucial in converting the received light energy to chemically bonded energy. The carotenoids not involved in photosynthesis are photoprotective (PPC). In particular, some xanthophylls such as violaxanthin (Viola), zeaxanthin (Zea), diadinoxanthin (Diadino) and diatoxanthin (Diato) are involved in the xanthophyll cycle, one of the most important photoprotective mechanisms that drives the non-radiative dissipation of the excess light energy to prevent photoinhibition [4,5]. Therefore, their relative abundance can be used as a tracer of photoacclimation processes [5].

In the context of global climate change, knowledge of the distributions of phytoplankton pigments is useful to understand the impacts of the changing environment on primary productivity [6], phytoplankton diversity and community composition through appropriate analysis, for example, CHEMTAX [7] and diagnostic pigment analysis [8]. In remote sensing applications, phytoplankton pigment databases have been extensively used to develop, validate, or refine bio-optical algorithms for estimating phytoplankton biomass (often estimated using total chlorophyll-a (TChl-a) concentration) and functional types (via diagnostic pigment analysis) based on both cell size (micro-, nano- and pico-phytoplankton) and biogeochemical functions (e.g., calcification, silicification, dimethyl sulphide production and nitrogen fixation) [9] and references therein. These data sets are mainly based on high-performance liquid chromatography (HPLC) analysis of discrete water samples. This technique enables the accurate quantification of 25–50 pigments in a single analysis [10]. However, it requires highly trained personnel, intensive labor and time, expensive and complex analysis, and is limited by the sampling frequency, spatial coverage and additional issues related to discrete sampling such as sample handling, storage and transportation. While HPLC pigment analysis remains indispensable, it is necessary to explore methods that enable easier access to pigment data at higher spatial-temporal resolution.

Because optical measurements are currently the only means of collecting synoptic scale information on upper ocean particles (e.g., operational open-ocean satellite ocean color provides data daily with pixel size down to 300 m by 300 m), attempts have been made to quantify the concentrations of various phytoplankton pigments from these measurements (e.g., absorption or reflectance spectra). Optical methods take advantage of the distinctive absorption characteristics of different pigments and various approaches are applied, such as the decomposition of spectra into Gaussian functions, e.g., [11], spectral reconstruction, e.g., [12], derivative analysis, e.g., [13], partial least squares regression, e.g., [14], multiple linear regression [15], reflectance band ratio, e.g., [16,17], principal component analysis, e.g., [18] and artificial neural networks [19,20].

The Gaussian decomposition method decomposes phytoplankton absorption spectra ($a_{ph}(\lambda)$) into Gaussian functions and correlates the amplitudes of the Gaussian functions with the concentrations of major pigment groups. The amplitude of each Gaussian function is assumed to represent the magnitude of the absorption coefficient of a specific pigment or pigment group at the Gaussian peak wavelength, based on known pigment absorption properties determined in laboratory analyses. This method simultaneously retrieves the concentrations of chlorophyll-a, chlorophyll-b, chlorophyll-c and carotenoids [11,21–23] or of chlorophyll-a and phycocyanin [24,25]. However, the retrieval accuracy is generally limited by the variations in pigment package effect of field samples. Nevertheless, the Gaussian absorption coefficients of specific pigment groups were recently incorporated into the reconstruction of hyper- and multi-spectral remote sensing reflectance, allowing the robust estimation of the concentrations of TChl-a, total chlorophyll-b (TChl-b), the combination of chlorophyll-c1 and -c2 (Chl-c1/2) and PPC globally [23] as well as of phycocyanin in cyanobacteria bloom waters [24,25] from remote sensing reflectance data.

The spectral reconstruction method assumes that $a_{ph}(\lambda)$ can be reconstructed from the linear combination of pigment-specific absorption coefficients multiplied by corresponding pigment concentrations [26]. Moisan et al. [27,28] applied matrix inversion analysis to the reconstruction model and successfully estimated the concentrations of a series of pigments directly from $a_{ph}(\lambda)$. This technique involves a first inversion of the observed pigment concentrations that derives pigment-specific absorption spectra and a second inversion of these derived pigment-specific absorption spectra that solves for pigment concentrations. Four methods that solve least squares problems, i.e., singular value decomposition (SVD) [29], non-negative least squares (NNLS) [30] and two nonlinear least squares minimization schemes based on the Levenberg–Marquardt algorithm [31,32] were compared for the two inversions. They found that when the first inversion was carried out with SVD and the second one with NNLS, the inverse modeling technique yielded the most accurate pigment estimates. However, the retrieval accuracy is affected by the level of correlation between pigment concentrations, the contribution of a specific pigment to the spectral $a_{ph}(\lambda)$, pigment package effect, the missing absorption components by the pigments that exist in the samples but are not obtained by standard HPLC (e.g., mycosporine-like amino acids and phycobiliproteins) [27,28], and the number of spectral bands of $a_{ph}(\lambda)$ used in the inversion model [27,28,33]. Overall, the SVD-NNLS method achieved simultaneous statistically significant retrievals of TChl-a, total chlorophyll-c (TChl-c), β-carotene (β-Caro), Fuco, Viola, Diadino and peridinin (Peri) in U.S. east coast waters [27,28]. It was recently applied to $a_{ph}(\lambda)$ modeled from MODIS-Aqua TChl-a data for northeastern U.S. waters, yielding maps of the concentrations of ten pigments [34]. Similar approaches were successful in infering phytoplankton size classes globally [35,36] and taxonomic groups in the Chukchi and Bering Seas [33] from absorption data.

Derivative analysis of absorption spectra separates the secondary absorption peaks and shoulders contributed by phytoplankton pigments within the overlapping absorption regions [37]. Bidigare et al. [13] found that the fourth derivative maxima of particulate absorption spectra ($a_p(\lambda)$) provided strong linear relationships with chlorophylls (a, b and c) concentrations in Sargasso Sea. However, this method failed to estimate carotenoid concentrations because of the similarity of their spectral properties, the broad spectral absorption and relatively rounded absorption peaks that are less accessible to derivative analysis.

Principal component analysis (a.k.a. empirical orthogonal function analysis) derives several dominant modes (known as "principal components") of the spectra that mainly account for the variability in spectral shape and relates them to pigment concentrations. Bracher et al. [18] performed this analysis on both hyperspectral and multispectral remote sensing reflectance data and retrieved the concentrations of TChl-a, monovinyl-chlorophyll-a, PPC, PSC, Chl-c1/2, 19′-butanoyloxyfucoxanthin (But), 19′-hexanoyloxyfucoxanthin (Hex), Zea, phycoerythrin and the sum of α- and β-Caro from the linear combinations of the principal components in the Atlantic Ocean. This method is, however, only applicable to the pigments that have been identified in most collocated samples. It failed to retrieve the pigments that are mostly absent or below detection limit. Similarly, Soja–Woźniak et al. [38] applied this analysis on both hyperspectral and multispectral remote sensing reflectance data and successfully retrieved TChl-a, phycocyanin and phycoerythrin in the Gulf of Gdansk.

An artificial neural network relates spectra to pigment data with a nonlinear model that self-adjusts the model parameters (i.e., weight matrix) for the best fit. Bricaud et al. [19] developed a multilayer perceptron using a global data set and obtained estimations of the concentrations of TChl-a, TChl-b, TChl-c, PSC and PPC, with TChl-a and TChl-b being the most accurate and poorest estimates, respectively. The main limitation of this method lies in the biological variability embedded in the training data set.

More recently, there has been an increased use of in situ hyperspectral optical sensors to obtain pigment data from continuous optical measurements, e.g., [22]. In-line and autonomous measurements by new miniature sensors deployed on various platforms (e.g., profiling floats, autonomous surface water vehicles) have substantially increased the sampling frequency and spatial coverage of measurements. The shipboard underway spectrophotometry considerably facilitates the acquisition of $a_p(\lambda)$ with unprecedented spatial resolution. It utilizes an AC-S hyperspectral spectrophotometer (or the 9-wavelength resolved AC-9) (Sea-Bird Scientific, Philomath, OR, USA) operated in flow-through mode and derives $a_p(\lambda)$ by differencing the bulk seawater absorption measurements from temporally adjacent 0.2-μm filtered water sample measurements, e.g., [22,39–48]. It has provided surface TChl-a data along cruise tracks via the empirical relationships between the spectrophotometry derived $a_p(\lambda)$ and HPLC measured TChl-a concentrations [39,40,45–48]. Furthermore, Gaussian decomposition has been performed by Chase et al. [22] to retrieve major pigment groups from a globally extensive underway AC-S derived $a_p(\lambda)$ data set. Here we use a data set obtained with a similar underway system to compare and contrast two different methods to obtain information on the underlying pigments.

The Fram Strait, the region between Svalbard and Greenland, provides the only deep connection between the North Atlantic and Arctic Oceans (Figure 1). It is of great importance to the climate in the Arctic region, as it accounts for 75% of the mass exchange and 90% of the heat exchange between the Arctic Ocean and the rest of the world's ocean [49]. In recent decades, the Fram Strait has undergone a significant warming, high variability of Atlantic water inflow [50] and an overall increase of sea ice area export [51–55]. This impacts phytoplankton biomass, community composition and distribution by altering light and nutrient regimes. The seasonal cycle of phytoplankton biomass has been significantly enhanced in the shallow upper water layers since 2008 [56]. Phytoplankton distributions reflect the dominant local physical processes [56,57]. A significant increase in summertime chlorophyll-a concentration in the eastern Fram Strait was observed, whereas on the western side there were minor changes [56]. Furthermore, a shift of dominant phytoplankton assemblages from diatoms (mainly *Thalassiosira* spp., *Chaetoceros* spp. and *Fragilariopsis* spp.) towards coccolithophores (mainly *Emiliania huxleyi*) and more recently, *Phaeocystis* spp. (mainly *Phaeocystis pouchetii*) and other small pico- and nanoflagellates during summer months was suggested [56,58,59], which can strongly affect the functioning and stability of marine food webs [60,61]. The studies of phytoplankton community composition in this region are mainly based on discrete water samples or moored sediment traps. Because of the inherent limitations of these methods, the observations are scarce. Furthermore, it remains difficult to obtain information on phytoplankton community composition via satellite due to the poor spatial-temporal coverage of ocean color data in this region, e.g., [57] and the lack of assessment of the applicability of satellite algorithms determining the phytoplankton community structure for this region. Additionally, algorithms applicable to other waters for quantifying phytoplankton community structure or pigment composition from in situ optical measurements have not been assessed yet in this region.

The Fram Strait cruises PS93.2, PS99.2 and PS107 on R/V Polarstern collected a comprehensive in situ bio-optical data set and offer a unique opportunity for bio-optical modeling. In particular, underway spectrophotometry was applied during all three cruises. To obtain the information of individual phytoplankton pigments or pigment groups (e.g., PSC and PPC) from underway spectrophotometry, here, we compare and optimize the performances of two pigment retrieval approaches, Gaussian decomposition [22] and the matrix inversion technique [27,28], find the potential number and types of pigments that can be retrieved, and assess the applicability of the two approaches to the Fram Strait and its vicinity.

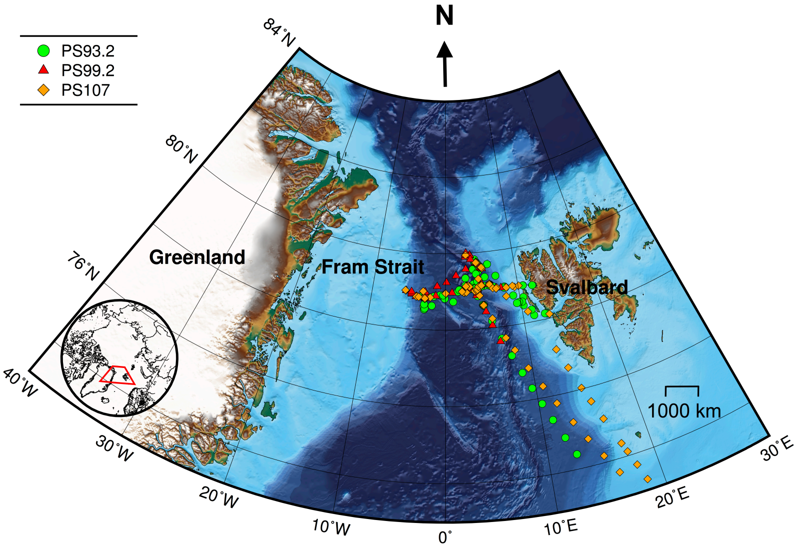


Figure 1. Cruise tracks for PS93.2 (July–August 2015), PS99.2 (June–July 2016) and PS107 (July–August 2017). Symbols denote locations where both AC-S and HPLC data were collected. Bathymetric grid data are extracted from the International Bathymetric Chart of the Arctic Ocean Version 3.0 [62]. Lambert azimuthal equal-area projection was used for mapping.

2. Data and Methods

2.1. Data Collection

Data were collected during three expeditions on R/V Polarstern: PS93.2 (July to August 2015), PS99.2 (June to July 2016) and PS107 (July to August 2017). These cruises have repeated survey design. Sampling sites were located in the Fram Strait and its vicinity, ranging from approximately latitudes 72° to 80°N and longitudes 10°W to 15°E (Figure 1).

The underway $a_p(\lambda)$ and discrete pigment concentration measurements of the surface water were collected for each expedition. The average velocity of the ship while moving is ~10 knots (5.1 m s^{-1}). Sampling methods and data analysis are detailed in Liu et al. [45]. Briefly, a 25-cm-pathlength AC-S spectrophotometer (spectral range: 400–740 nm, full width half maximum (FWHM): 10 nm, wavelength resolution: ~3.5 nm) was integrated into the shipboard flow-through system following the setup of Slade et al. [46]. Seawater was sampled at roughly 11 m below the sea surface from the ship's keel using a membrane pump. The flow rate of seawater is 1–2 L min^{-1}. The $a_p(\lambda)$ spectra were derived by subtracting the absorption coefficients of 0.2-μm filtered seawater from those of seawater materials measured by AC-S. Subsequently, they were corrected for temperature and salinity dependency of pure water absorption [63], scatter errors [64] and residual temperature effect [46]. Additionally, the effect of AC-S filter factors resulting in a smoothing of the measured $a_p(\lambda)$ spectra was corrected [22].

Discrete seawater samples were collected from the unfiltered AC-S outflow approximately every three hours. Seawater (1–3 L) was filtered with GF/F glass fiber filters (nominal pore size 0.7 μm) for HPLC phytoplankton pigment analysis (see Table 1 for the names and abbreviations of the pigments and pigment groups used in this study). Pigments were grouped following Hooker et al. [65]. Divinyl-chlorophyll-a and divinyl-chlorophyll-b were not found in our data set. For convenience, in the following context, the term "pigment" stands for either a specific type of pigment or a pigment group such as PSC and PPC. In addition, the spectral absorption coefficient of non-algal particles ($a_{NAP}(\lambda)$) in discrete water samples was measured for the determination of its spectral exponent. Seawater (0.2–1 L) was filtered to concentrate particulate materials on the GF/F filters. $a_p(\lambda)$ from discrete samples was determined using Quantitative Filter Technique [66–68]. Measurements for samples from PS93.2 were carried out on a dual-beam UV/VIS spectrophotometer (Cary 4000, Varian Inc., Palo Alto, CA, USA) (spectral range: 300–850 nm, FWHM: 2 nm, wavelength resolution: 1 nm) equipped with a 150 mm integrating sphere following Simis et al. [69], whereas filters collected during PS99.2 and PS107 were measured using a small portable integrating cavity absorption meter (spectral range: 300–850 nm, FWHM: 2 nm, wavelength resolution: 0.3 nm) [70], as detailed in Liu et al. [45]. $a_{NAP}(\lambda)$ was then obtained by measuring the sample filters bleached with 10% NaClO solution [69] following the same procedure as $a_p(\lambda)$ measurements. $a_{NAP}(\lambda)$ was approximated using an exponentially decaying function [71,72]:

$$a_{NAP}(\lambda) = a_{NAP}(400)\, e^{-S(\lambda-400)} \tag{1}$$

where S is the spectral exponent of $a_{NAP}(\lambda)$. Equation (1) was fit to $a_{NAP}(\lambda)$ for data between 380–620 nm excluding the 400–480 nm range (to eliminate residual chlorophyll-a absorption peak) using non-linear least squares method [72]. The median value of S for all three expeditions is 0.016 nm^{-1} (standard deviation with respect to the median value is 0.006 nm^{-1}), which was subsequently used in the decomposition of the AC-S derived $a_p(\lambda)$ to obtain $a_{ph}(\lambda)$ (see Section 2.2.1).

AC-S derived $a_p(\lambda)$ were averaged within the period of ten minutes before and after HPLC sampling time and were matched with HPLC pigments data. $a_{ph}(\lambda)$ (400–700 nm, wavelength resolution: ~3.5 nm) was obtained by numerical decomposition (see Section 2.2.1). In total, 298 $a_p(\lambda)$-pigments match-ups were obtained, which were subsequently used as the pigment retrieval data set. The link to the data used in this study is shared in the Supplementary Materials.

Table 1. Abbreviations of phytoplankton pigments and pigment groups analyzed in this study, and the minimum, maximum, mean and standard deviation of the pigment concentrations (mg m^{-3}).

Pigment/Pigment Group	Abbreviation	Minimum	Maximum	Mean	Standard Deviation
alloxanthin	Allo	0.00	0.16	0.01	0.01
chlorophyll-c1/2	Chl-c1/2	0.00	0.94	0.15	0.15
chlorophyll-c3	Chl-c3	0.00	0.83	0.08	0.11
α-carotene	α-Caro	0.00	0.04	0.00	0.01
β-carotene	β-Caro	0.00	0.07	0.02	0.01
diadinoxanthin	Diadino	0.00	0.49	0.10	0.08
diatoxanthin	Diato	0.00	0.05	0.01	0.01
fucoxanthin	Fuco	0.01	1.28	0.22	0.21
19′-hexanoyloxyfucoxanthin	Hex	0.00	1.63	0.23	0.24
19′-butanoyloxyfucoxanthin	But	0.00	0.51	0.04	0.05
neoxanthin	Neo	0.00	0.02	0.00	0.00
lutein	Lut	0.00	0.01	0.00	0.00
peridinin	Peri	0.00	0.45	0.03	0.06
prasinoxanthin	Prasino	0.00	0.05	0.00	0.01
pheophytin-a	Pheo-a	0.00	1.31	0.02	0.10
pheophorbide-a	Phide-a	0.00	0.17	0.01	0.02
violaxanthin	Viola	0.00	0.03	0.01	0.01
zeaxanthin	Zea	0.00	0.08	0.01	0.01
total chlorophyll-a	TChl-a	0.06	3.87	0.86	0.66
total chlorophyll-b	TChl-b	0.00	0.22	0.06	0.03
total chlorophyll-c	TChl-c	0.00	1.62	0.23	0.24
photosynthetic carotenoids	PSC	0.02	3.56	0.52	0.49
photoprotective carotenoids	PPC	0.01	0.64	0.17	0.11

Note: TChl-a = monovinyl-chlorophyll-a + chlorophyllide-a; TChl-b = monovinyl-chlorophyll-b; TChl-c = Chl-c1/2 + Chl-c3; PSC = Fuco + But + Hex + Peri [65], PPC = Allo + Diadino + Diato + Zea + α- + β-Caro [65].

2.2. *Retrieval of Phytoplankton Pigments*

Figures 2 and 3 illustrate the steps of applying Gaussian decomposition and the matrix inversion technique, respectively, to retrieve phytoplankton pigment concentrations, which are described in detail in the following subsections. The link to the codes for data processing is shared in the Supplementary Materials.

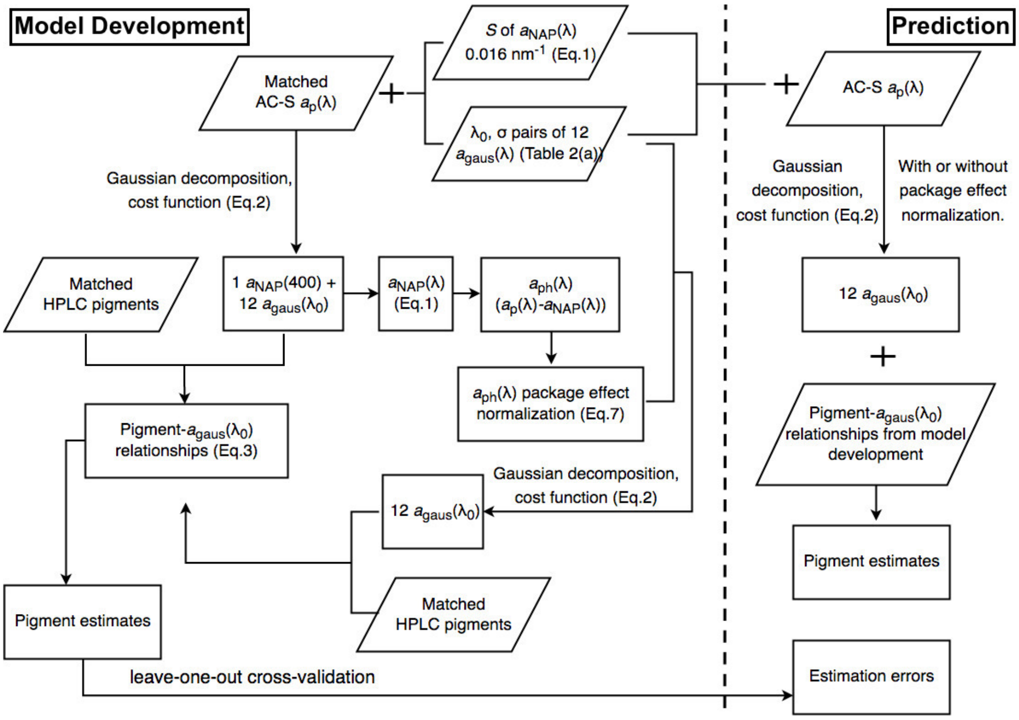


Figure 2. Schematic overview of the steps of applying Gaussian decomposition for phytoplankton pigment retrieval.

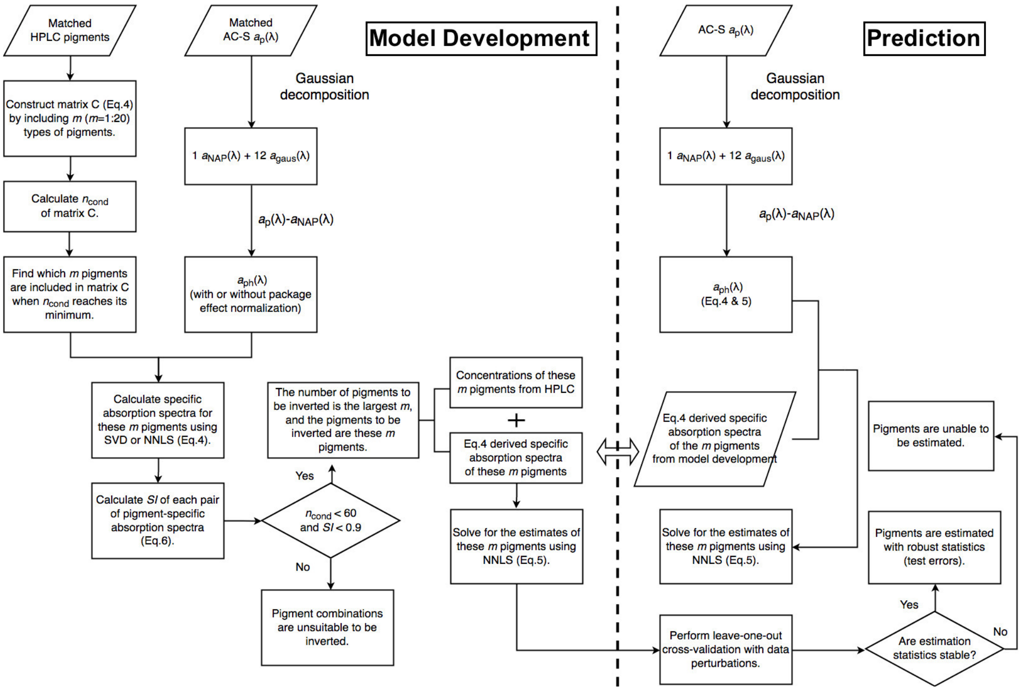


Figure 3. Schematic overview of the steps of applying the matrix inversion technique for phytoplankton pigment retrieval.

2.2.1. Gaussian Decomposition

Following Chase et al. [22], AC-S derived $a_p(\lambda)$ was decomposed to twelve Gaussian functions and one $a_{NAP}(\lambda)$ exponential function expressed by Equation (1) in the range of 400–700 nm. Each Gaussian function represents the absorption by a certain phytoplankton pigment. The absorption by the water-soluble photosynthetic pigment phycoerythrin was also represented as a Gaussian function though its concentration could not be validated with HPLC. The peak location and width of each Gaussian function shown in Table 2 were defined with fixed values based on known pigment absorption shapes [73]. The decomposition is optimized by minimizing the cost function using a weighted least squares method [22]:

$$\chi^2 = \sum_{\lambda=400}^{700} \frac{[a_p(\lambda) - \sum\limits_{i=1}^{12} a_{gaus,i}(\lambda) - a_{NAP}(\lambda)]^2}{\sigma_{SD}^2(\lambda)} \tag{2}$$

where $a_{gaus,i}(\lambda)$ denotes the *i*th Gaussian function, and $\sigma_{SD}(\lambda)$ is the standard deviation of the 20-min averaged matched $a_p(\lambda)$ spectra.

Table 2. The peak wavelengths (λ_0) and widths (σ) of the Gaussian functions for phytoplankton pigments, the statistics for the power function regression of $a_{gaus}(\lambda_0)$-pigment pairs in the training set (regression coefficients *A* and *B* of Equation (3) were calculated with 95% confidence bounds), and the statistics based on leave-one-out cross-validation. MAE is in mg m^{-3} (values outside the parentheses were calculated with linear-scale values, while inside the parentheses with log10-scale values), MPE in %, and N is the number of data points for the regressions.

(a) Decomposition of $a_{ph}(\lambda)$

λ_0 [nm]	σ [nm]	Pigment	*A*	*B*	R^2	MAE [b]	MPE [b]	N	MAE [c]	MPE [c]
406	16	TChl-a	17.60 ± 4.03	0.90 ± 0.08	0.75	0.28(0.18)	26.2	274	-	-
434	12	TChl-a	41.61 ± 6.71	1.12 ± 0.05	0.87	0.21(0.13)	20.7	297	0.22(0.13)	20.8
453	12	TChl-b & c [a]	1.18 ± 0.20	1.23 ± 0.05	0.92	0.00(0.14)	21.1	297	-	-
470	13	TChl-b	0.38 ± 0.11	0.50 ± 0.08	0.52	0.02(0.17)	29.5	296	-	-
492	16	PPC	1.23 ± 0.38	0.54 ± 0.09	0.50	0.06(0.18)	30.6	298	0.06(0.18)	31.5
523	14	PSC	25.25 ± 7.58	0.92 ± 0.08	0.76	0.20(0.21)	33.4	298	0.20(0.21)	34.0
550	14	phycoerythrin	-	-	-	-	-	-	-	-
584	16	Chl-c1/2	12.18 ± 5.18	0.85 ± 0.09	0.68	0.07(0.26)	44.9	297	-	-
617	13	TChl-a	21.00 ± 7.85	0.57 ± 0.07	0.66	0.33(0.20)	36.8	295	-	-
638	11	Chl-c1/2	49.89 ± 16.13	1.03 ± 0.06	0.81	0.05(0.20)	33.4	297	0.06(0.20)	33.5
660	11	TChl-b	0.66 ± 0.21	0.44 ± 0.06	0.57	0.02(0.16)	29.1	293	0.02(0.16)	29.3
675	10	TChl-a	19.70 ± 3.92	0.76 ± 0.05	0.82	0.24(0.14)	24.9	298	0.25(0.14)	25.3

(b) Decomposition of $\hat{a}_{ph}(\lambda)$.

λ_0 [nm]	Pigment	*A*	*B*	R^2	MAE [b]	MPE [b]	N	MAE [c]	MPE [c]
406	TChl-a	12.50 ± 1.37	0.96 ± 0.048	0.88	0.22(0.15)	19.5	274	-	-
434	TChl-a	19.23 ± 1.22	1.07 ± 0.026	0.96	0.13(0.08)	11.9	297	0.13(0.08)	12.2
453	TChl-b & c [a]	0.39 ± 0.05	1.10 ± 0.05	0.92	0.00(0.15)	25.8	297	-	-
470	TChl-b	0.30 ± 0.06	0.51 ± 0.07	0.60	0.02(0.15)	26.3	296	-	-
492	PPC	1.89 ± 0.32	0.77 ± 0.06	0.77	0.05(0.14)	21.4	298	0.05(0.14)	21.8
523	PSC	44.04 ± 5.31	1.19 ± 0.04	0.92	0.11(0.14)	20.4	298	0.11(0.14)	20.5
550	phycoerythrin	-	-	-	-	-	-	-	-
584	Chl-c1/2	16.73 ± 4.31	1.00 ± 0.06	0.82	0.06(0.22)	36.2	297	-	-
617	TChl-a	64.19 ± 13.15	0.83 ± 0.04	0.84	0.22(0.15)	24.1	295	-	-
638	Chl-c1/2	34.11 ± 7.20	1.06 ± 0.05	0.91	0.04(0.17)	27.2	297	0.04(0.18)	27.2
660	TChl-b	0.47 ± 0.12	0.41 ± 0.05	0.62	0.02(0.15)	27.9	293	0.02(0.15)	27.4
675	TChl-a	33.57 ± 0.72	1.00 ± 0.01	1.00	0.05(0.03)	3.6	298	0.05(0.03)	3.6

[a] 0.03(TChl-b) + 0.07(Chl-c1/2) [22,73]; [b] training errors; [c] test errors based on cross-validation.

The amplitude of $a_{NAP}(400)$ was derived by minimizing Equation (2) sample by sample and used to reconstruct $a_{NAP}(\lambda)$ for each sample according to Equation (1). $a_{ph}(\lambda)$ was obtained by differencing $a_p(\lambda)$ and $a_{NAP}(\lambda)$. The amplitude of each Gaussian function $a_{gaus}(\lambda_0)$ [m^{-1}] was derived by minimizing Equation (2) and related to the concentration of the corresponding pigment measured by HPLC (c [mg m^{-3}]) by fitting the following equation using Bisquare robust non-linear least squares method (data pairs with either $a_{gaus}(\lambda_0)$ or c being 0 were excluded) (Table 2):

$$c = A\ a_{gaus}(\lambda_0)^B \tag{3}$$

For convenience, we denote the five pigments that can be retrieved using Gaussian decomposition (Table 2), i.e., TChl-a, TChl-b, Chl-c1/2, PSC and PPC as "Gauss-5 pigments".

2.2.2. Matrix Inversion Technique

The $a_{ph}(\lambda)$ spectra can be reconstructed as the linear combination of the absorption spectra of individual pigments that equal to the pigment-specific absorption coefficients ($a_j^*(\lambda)$) multiplied by pigment concentrations (c_j) [26], i.e., $a_{ph}(\lambda) = \sum_{j=1}^{m} c_j a_j^*(\lambda)$. When there is more than one sample in the observed collocated pigment concentrations and $a_{ph}(\lambda)$ data set, the reconstruction model can be written in matrix multiplication form as:

$$\begin{bmatrix} c_{i=1,j=1} & \cdots & c_{i=1,j=m} \\ \vdots & \ddots & \vdots \\ c_{i=n,j=1} & \cdots & c_{i=n,j=m} \end{bmatrix} \begin{bmatrix} \widetilde{a}^*_{j=1}(\lambda) \\ \vdots \\ \widetilde{a}^*_{j=m}(\lambda) \end{bmatrix} = \begin{bmatrix} a_{ph,i=1}(\lambda) \\ \vdots \\ a_{ph,i=n}(\lambda) \end{bmatrix} \Longleftrightarrow \mathrm{C} \cdot \widetilde{\mathrm{A}} = \mathrm{A_{ph}} \tag{4}$$

where c is the observed pigment concentration (e.g., from HPLC), $\widetilde{a}^*(\lambda)$ is the derived pigment-specific absorption coefficient, n is the number of samples, i is the sample index, m is the number of pigments measured in each sample, and j is the pigment index. The c and $a_{ph}(\lambda)$ are known (the former from HPLC and the latter from spectrophotometry) while $\widetilde{a}^*(\lambda)$ is unknown. To solve for $\widetilde{a}^*(\lambda)$, the elements of matrix $\widetilde{\mathrm{A}}$, the inverse of matrix C is computed. Once this is done, the derived $\widetilde{a}^*(\lambda)$ is used with the observed $a_{ph}(\lambda)$ (l in Equation (5) is the number of wavelengths) to solve for the pigment concentrations ($\widetilde{c}$ in Equation (5)). Likewise, the computation of the inverse of matrix $\widetilde{\mathrm{A}}$ is necessary.

$$\begin{bmatrix} \widetilde{a}^*_{j=1}(\lambda_1) & \cdots & \widetilde{a}^*_{j=m}(\lambda_1) \\ \vdots & \ddots & \vdots \\ \widetilde{a}^*_{j=1}(\lambda_l) & \cdots & \widetilde{a}^*_{j=m}(\lambda_l) \end{bmatrix} \begin{bmatrix} \widetilde{c}_{i=1,j=1\ldots m} \\ \vdots \\ \widetilde{c}_{i=n,j=1\ldots m} \end{bmatrix} = \begin{bmatrix} a_{ph,i=1\ldots n}(\lambda_1) \\ \vdots \\ a_{ph,i=1\ldots n}(\lambda_l) \end{bmatrix} \Longleftrightarrow \widetilde{\mathrm{A}} \cdot \widetilde{\mathrm{C}} = \mathrm{A_{ph}} \tag{5}$$

Hence, this is a two step approach. First, where HPLC is available, Equation (4) is used to obtain the pigment-specific absorption spectra. Once those are available, Equation (5) is used to derive pigment concentrations directly from $a_{ph}(\lambda)$ (a step that does not require HPLC data). The SVD-NNLS approach proposed by Moisan et al. [27,28], i.e., solving Equation (4) with SVD least squares method on each wavelength and Equation (5) with NNLS method, was proved to give the best pigment estimates.

Singular Value Decomposition—Non-Negative Least Squares (SVD-NNLS)

The SVD-NNLS approach was adapted and tested using our data set. The matrix C in Equation (4) (the concentrations of the pigments listed in Table 1) was inverted using SVD. The least-squares solution of the overdetermined Equation (4) (in this study $n > m$) is derived by $\widetilde{\mathrm{A}} = \mathrm{C}^+ \cdot \mathrm{A_{ph}}$, where C^+ is the Moore-Penrose pseudoinverse of matrix C computed by SVD (MATLAB function *pinv*). $\widetilde{\mathrm{A}}$ provides the specific absorption spectra of each pigment and is then used in Equation (5) to solve for pigment concentrations via NNLS (MATLAB function *lsqnonneg*), i.e., by inverting Equation (5) using least squares method with the constraint $c_{i,j} \geq 0$.

To ensure robust solutions of the overdetermined systems, matrix C in Equation (4) and matrix $\widetilde{A}$ in Equation (5) should be constructed to avoid ill-conditioning, requiring that the columns and rows of matrix C have sufficient linear independence, and that the shapes of any two $\widetilde{a}^*(\lambda)$ be sufficiently different from each other [36]. Here, we used the condition number (n_{cond}) (MATLAB function *cond*) as a diagnostic for the degree of the well-conditioning of matrix C (a matrix with a high n_{cond} is ill-conditioned, and vice versa). In addition, a similarity index ($SI_{i,j}$) [74,75] was used to represent the similarity between the absolute values of two specific spectra $\widetilde{a}_i^*(\lambda)$ and $\widetilde{a}_j^*(\lambda)$ (denoted as $\widetilde{a}^{*+}(\lambda)$). The $SI_{i,j}$ is a number ranging from 0 (no similarity) to 1 (perfect similarity).

$$SI_{i,j} = 1 - \frac{2}{\pi} arccos(\frac{\widetilde{a}_i^{*+}(\lambda) \cdot \widetilde{a}_j^{*+}(\lambda)}{||\widetilde{a}_i^*(\lambda)|| \; ||\widetilde{a}_j^*(\lambda)||}) \tag{6}$$

where $||\widetilde{a}^*(\lambda)||$ is the norm of the vector $\widetilde{a}^*(\lambda)$ (MATLAB function *norm*). To maximize the number of pigment types to be determined (m) while reducing the degree of the ill-conditioning, we tested all the possibilities of combining pigments composed of matrix C (number of combinations $\frac{20!}{m!(20-m)!}$, 20 types of pigments measured in total) and examined n_{cond} and $SI_{i,j}$ for all cases. For example, when m = 20, matrix C includes all the pigments listed in Table 1 excluding TChl-c, PSC and PPC (denoted as "Fram-20 pigments") in 298 samples . When $1 \leq m < 20$, matrix C includes the concentrations of m types of pigments and a summed contribution of other pigments included in the Fram-20 pigments. m was then determined as the biggest number with n_{cond} smaller than 60 and $SI_{i,j}$ smaller than 0.9. For convenience, we denote the retrieval of these m pigments using SVD-NNLS as SVD-NNLS-m.

To compare with the results from Gaussian decomposition, the SVD-NNLS method was also applied to only retrieve Gauss-5 pigments (denoted as SVD-NNLS-5′), i.e., matrix C in Equation (4) only includes the concentrations of these five pigments and a summed contribution of other pigments.

The SVD derived specific absorption spectra are purely mathematical solutions. Therefore, they are expected to be different from those obtained through laboratory measurements performed on the extracted individual pigments in solution ($a^*(\lambda)$) [27]. For comparison, we tested the validity of NNLS in estimating pigment concentrations using $a^*(\lambda)$ from Bricaud et al. [73] in Equation (5). In this case, $a^*(\lambda)$ is available for 12 types of pigments (Divinyl-chlorophyll-a and divinyl-chlorophyll-b not considered), i.e., TChl-a, TChl-b, alloxanthin (Allo), But, Chl-c1/2, Diadino, Fuco, Hex, Peri, Zea, α- and β-Caro (denoted as "Bricaud-12 pigments"). The same criteria of n_{cond} and $SI_{i,j}$ were used for all pigment combinations (number of combinations $\frac{12!}{m!(12-m)!}$). For each combination, the $a^*(\lambda)$ for the pigments which are missing in the Fram-20 pigments was solved using SVD (Equation (4), $a_j^*(\lambda)$ replaces $\widetilde{a}_j^*(\lambda)$). We denote this case as Bricaud-SVD-NNLS, and the retrieval of m pigments as Bricaud-SVD-NNLS-m.

Non-Negative Least Squares—Non-Negative Least Squares (NNLS-NNLS)

Though SVD provides a powerful tool for matrix inversion, one concern is that the SVD derived specific absorption spectra can be negative, which are physically unsound and not intuitively understood. To cope with this issue, NNLS was used twice (denoted as NNLS-NNLS), i.e., to invert Equation (4) to solve for the non-negetive matrix $\widetilde{A}$, as well as to invert Equation (5) to derive the non-negative matrix $\widetilde{C}$. Similarly, NNLS-NNLS-5′ was tested for comparison with the results from Gaussian decomposition and SVD-NNLS-5′. Bricaud-NNLS-NNLS were also tested using the same way as Bricaud-SVD-NNLS with the exception that the $a^*(\lambda)$ for the missing pigments was solved using NNLS. The NNLS-NNLS approach was also performed by Moisan et al. [27,28]. The pigment estimation results from NNLS-NNLS are provided in Appendix A.

Sensitivity Analysis

The solution of matrix inversion can be sensitive to input errors depending on the degree of well-conditioning of the input matrices, which can affect the pigment estimation accuracy. To ensure the stability of the matrix inversion model and obtain reliable pigment estimation statistics, perturbations were introduced to the input data for 300 iterations, and the related parameters, i.e., n_{cond}, SI, $\widetilde{a}^*(\lambda)$ and cross-validation statistics (see Section 2.2.4) were calculated as the median values of the 300 sets. Assuming an uncertainty of 15% for HPLC pigment data, matrix C was perturbed with random values within ±15% of the measurements. Matrix $\mathrm{A_{ph}}$ was also perturbed by adding the random values within $\pm\sigma_{SD}(\lambda)$ (see Equation (2)) to the measurements. The results for three cases, i.e., with perturbed C, with perturbed $\mathrm{A_{ph}}$ and with both matrices perturbed were considered for both SVD-NNLS and NNLS-NNLS.

2.2.3. Normalization of $a_{ph}(\lambda)$ by Pigment Package Effect

The pigment package effect index $Q_a^*(\lambda)$ can be calculated as the ratio of the measured $a_{ph}(\lambda)$ to the absorption coefficient of the same pigments which would be dispersed into solution [73,76]. To partially account for the package effect, Moisan et al. [27,28,34] normalized the measured $a_{ph}(\lambda)$ by dividing it with $Q_a^*(675)$ and found improved capability of the matrix inversion technique in retrieving pigment concentrations. To test the performances of both Gaussian decomposition and the matrix inversion technique with the normalization strategy, $Q_a^*(675)$ was calculated, and $a_{ph}(\lambda)$ was normalized following

$$\hat{a}_{ph}(\lambda) = a_{ph}(\lambda)\,\frac{0.033\; c_{TChl\text{-}a}}{a_{ph}(675)} \tag{7}$$

where $c_{TChl\text{-}a}$ is TChl-a concentration (in mg $\mathrm{m^{-3}}$), 0.033 (in $\mathrm{m^2\ mg^{-1}}$) is the “unpackaged” Chl-a-specific absorption coefficient at 675 nm measured by Bricaud et al. [73], and the fraction is the inverse of $Q_a^*(675)$. Note that chlorophyll-a, divinyl-chlorophyll-a, chlorophyll-b, divinyl-chlorophyll-b and Chl-c1/2 absorb light at 675 nm, e.g., [12,73]. For simplicity, we assume that $Q_a^*(675)$ is only contributed by TChl-a, not only because TChl-a contributes most to $a_{ph}(675)$, but also due to the weaker dependence on pigment data for $Q_a^*(675)$ calculation. In theory, $Q_a^*(675)$ ranges from 0 (fully “packed”) to 1 (“unpackaged”). Due to uncertainties, samples with calculated $Q_a^*(675)$ greater than 1 are unavoidable in practise and included in the calculation of $\hat{a}_{ph}(\lambda)$.

The $\hat{a}_{ph}(\lambda)$ was subsequently used to retrieve pigment concentrations via Gaussian decomposition, SVD-NNLS and NNLS-NNLS methods. Results were compared with those using $a_{ph}(\lambda)$. To avoid confusion, in the following context, unless otherwise stated, the results are based on $a_{ph}(\lambda)$.

2.2.4. Statistics

For the development of pigment retrieval models, all the match-up points were used as training data, allowing the models to best account for the biological variations in the data set. Statistics for applying the model to the training data include the slope and the intercept of Model-1 Bisquare robust linear regression, the determination coefficient (R^2), the mean absolute error (MAE) and the median absolute percentage error (MPE). The equations for the statistical metrics are given below:

$$\mathrm{MAE} \;=\; \frac{1}{n}\sum_{i=1}^{n}|\widetilde{c}_{i,j} - c_{i,j}| \tag{8}$$

$$\mathrm{MPE} \;=\; median\ of\ \frac{|\widetilde{c}_{i,j} - c_{i,j}|}{c_{i,j}} \times 100\% \tag{9}$$

where n is the number of samples, $c_{i,j}$ is the concentration of the jth pigment in the ith sample measured by HPLC, and $\widetilde{c}_{i,j}$ is the estimated pigment concentration. Considering phytoplankton pigments are approximately log-normally distributed in the ocean, the slope, intercept and R^2 were computed in log10 space. Additionally, MAE is also calculated using log10($c_{i,j}$). To avoid confusion, in the following context, unless otherwise stated, MAE values are calculated using $c_{i,j}$.

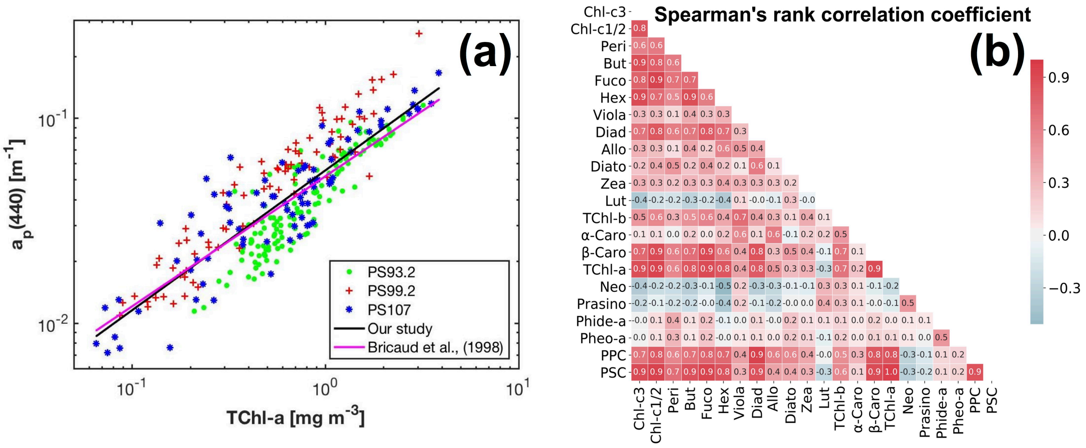
For model evaluation, leave-one-out cross-validation was performed (MATLAB function *crossvalind*) to estimate likely performance of each model on out-of-sample data. The pigment retrieval data set with N data points was split into two partitions: one partition is the testing set with one data point, and the other partition is the training set with the union of the other data points. Statistics were iteratively calculated N times, using a different data point as the testing set each time. The model prediction errors for out-of-sample data were defined as the average values of the statistics (mean value for MAE and median value for MPE, respectively) for the N sets. Cross-validation is an effective way of estimating model test errors when the number of data available is relatively small [18,77]. Compared to the random train/test split method and k-fold cross-validation, leave-one-out cross-validation has the advantage that the training set highly resembles the whole data set (the former has only one data point less than the latter), thus avoids the bias introduced to the estimation of the test errors due to less training data than data available.

For clarity, statistics (errors) obtained from running the trained model back on the training data are denoted as "training statistics (errors)", while those obtained when applying the trained model to the test data are called "test (or estimation or prediction) statistics (errors)".

3. Results

3.1. Characteristics of the Pigment Retrieval Data Set

The magnitudes of $a_p(440)$ and $a_p(675)$ vary in the range of 0.007–0.258 m^{-1} (median: 0.043 m^{-1}) and 0.001–0.086 m^{-1} (median: 0.013 m^{-1}), respectively. Overall, $a_p(440)$ varies as a power function of TChl-a concentration (Bisquare robust regression): $a_p(440) = 0.056\ (c_{TChl\text{-}a})^{0.682}$ ($R^2 = 0.82$, MAE = 0.01 m^{-1}). This relationship is close to the one derived by Bricaud et al. [71] for Case-1 waters (Figure 4a). TChl-a can also be related to $a_{ph}(675)$ via the cruise-specific power functions (Bisquare robust regression): $c_{TChl\text{-}a} = 56.46\ a_{ph}(675)^{0.984}$ ($R^2 = 0.96$, MAE = 0.12 mg m^{-3}) (PS93.2), $c_{TChl\text{-}a} = 26.3\ a_{ph}(675)^{0.935}$ ($R^2 = 0.96$, MAE = 0.14 mg m^{-3}) (PS99.2), and $c_{TChl\text{-}a} = 119.1\ a_{ph}(675)^{1.217}$ ($R^2 = 0.95$, MAE = 0.17 mg m^{-3}) (PS107).



	Chl-c3	Chl-c1/2	Peri	But	Fuco	Hex	Viola	Diad	Allo	Diato	Zea	Lut	TChl-b	α-Caro	β-Caro	TChl-a	Neo	Prasino	Phide-a	Pheo-a	PPC	PSC
Chl-c3																						
Chl-c1/2	0.8																					
Peri	0.6	0.6																				
But	0.9	0.8	0.6																			
Fuco	0.8	0.9	0.7	0.7																		
Hex	0.9	0.7	0.5	0.9	0.6																	
Viola	0.3	0.3	0.1	0.4	0.3	0.3																
Diad	0.7	0.8	0.6	0.7	0.8	0.7	0.3															
Allo	0.3	0.3	0.1	0.4	0.2	0.6	0.5	0.4														
Diato	0.2	0.4	0.5	0.2	0.4	0.2	0.1	0.6	0.1													
Zea	0.3	0.3	0.2	0.3	0.3	0.4	0.3	0.3	0.3	0.2												
Lut	-0.4	-0.2	-0.2	-0.3	-0.2	-0.4	0.1	-0.0	-0.1	0.3	-0.0											
TChl-b	0.5	0.6	0.3	0.5	0.6	0.4	0.7	0.4	0.3	0.1	0.4	0.1										
α-Caro	0.1	0.1	0.0	0.2	0.0	0.2	0.6	0.1	0.6	-0.1	0.2	0.2	0.5									
β-Caro	0.7	0.9	0.6	0.7	0.9	0.6	0.4	0.8	0.3	0.5	0.4	-0.1	0.7	0.1								
TChl-a	0.9	0.9	0.6	0.8	0.9	0.8	0.4	0.8	0.5	0.3	0.3	-0.3	0.7	0.2	0.9							
Neo	-0.4	-0.2	-0.2	-0.3	-0.1	-0.5	0.2	-0.3	-0.3	-0.1	-0.1	0.3	0.3	0.1	-0.1	-0.2						
Prasino	-0.2	-0.1	-0.2	-0.2	-0.0	-0.4	0.2	-0.1	-0.2	-0.0	0.0	0.4	0.3	0.1	-0.0	-0.1	0.5					
Phide-a	-0.0	0.1	0.4	0.1	0.2	-0.1	0.0	0.1	-0.0	0.2	0.1	0.1	0.1	0.0	0.2	0.0	0.1	0.1				
Pheo-a	0.0	0.1	0.3	0.1	0.2	-0.0	0.1	0.1	0.0	0.2	0.1	-0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.5			
PPC	0.7	0.8	0.6	0.7	0.8	0.7	0.4	0.9	0.6	0.6	0.4	-0.0	0.5	0.3	0.8	0.8	-0.3	-0.1	0.1	0.2		
PSC	0.9	0.9	0.7	0.9	0.9	0.8	0.3	0.9	0.4	0.4	0.3	-0.3	0.6	0.1	0.9	1.0	-0.3	-0.2	0.1	0.2	0.9	

Figure 4. (**a**) Variations of the AC-S derived $a_p(440)$ as a power function of TChl-a concentration; (**b**) the Spearman′s rank correlation coefficients between the concentrations of phytoplankton pigments in our data set (linear color bar scale).

The composition and data range of pigments (minimum, maximum, mean and standard deviation) are shown in Table 1. TChl-a concentration spans the range 0.06–3.87 mg m^{-3}. Only TChl-a, TChl-c, Fuco, Hex and PSC have a mean concentration greater than 0.2 mg m^{-3}. Other than that, the pigments with mean concentrations greater than 0.05 mg m^{-3} are Chl-c1/2, Diadino, TChl-b, and PPC. Large standard deviations were observed within individual pigments, and the ratios of standard deviation to mean value are in the range of 0.58–4.35. The correlation between the concentrations of phytoplankton pigments was represented by the Spearman′s rank correlation coefficient (Spearman’s ρ) (Figure 4b). High level of the correlation (e.g., Spearman’s ρ > 0.7 or <−0.7) is a concern as it can cause the ill-conditioning of Equation (4), influencing the accuracy of pigment estimation using the matrix inversion technique.

3.2. Gaussian Decomposition

Strong correlations were found between the Gaussian function amplitudes and the corresponding pigment concentrations ($R^2 > 0.5$) (Table 2 (a)). Among all the Gaussian functions representing TChl-a absorption, the amplitude at 434 nm has the strongest relationship to TChl-a concentration (R^2 = 0.87, MAE = 0.21 mg m^{-3}), closely followed by that at 675 nm. $a_{gaus}(638)$ is much better correlated with Chl-c1/2 than $a_{gaus}(584)$, while $a_{gaus}(660)$ provides slightly better correlation with TChl-b than $a_{gaus}(470)$. That the R^2 for TChl-b and PPC is smaller than 0.6 is likely due to the reduced dynamic range of TChl-b and PPC compared to TChl-a, Chl-c1/2 and PSC (Table 1). Considering the relationships for PSC and PPC as well as the strongest correlations for TChl-a, TChl-b and Chl-c1/2, the MPE ranges from 20.7% to 33.4%. The training errors for the five pigments increase in the order of: TChl-a, TChl-b, PPC, Chl-c1/2 and PSC. The strongest correlations for all five pigments are shown in Figure 5a,c,e,g,i. For comparison, TChl-a is also plotted against $a_{gaus}(675)$ in Figure 5k.

As a result of the above, $a_{gaus}(434)$, $a_{gaus}(660)$, $a_{gaus}(638)$, $a_{gaus}(523)$ and $a_{gaus}(492)$ were used to predict the concentrations of TChl-a, TChl-b, Chl-c1/2, PSC and PPC, respectively. Statistics based on leave-one-out cross-validation (Table 2 (a)) show that overall, all five pigments were reasonably well retrieved (MPE 20.8–34.0%). TChl-a and TChl-b have the least and the second least prediction errors, respectively, while PSC is most poorly estimated.

When using $\hat{a}_{ph}(\lambda)$, the performance of the Gaussian decomposition significantly improved. Strong contrast was observed in the relationships between the Gaussian function amplitudes and pigment concentrations before and after applying the package effect normalization according to Equation (7) (Figure 5, Table 2 (b)). All statistical parameters for the eleven relationships between Gaussian absorption and pigment concentrations are improved (Table 2 (b)). The R^2 for $a_{gaus}(434)$ to TChl-a correlation is 0.96 (0.87 before the normalization) and the regression coefficient B (Equation (3)) is close to one. Similarly, the R^2 for $a_{gaus}(638)$ to Chl-c1/2 correlation is 0.91 (0.81 before the normalization) and the regression coefficient B (Equation (3)) is also close to one. The relationships for PSC and PPC also improved in terms of increased R^2 and decreased MAE and MPE for both pigments, and of a much closer shift of B (Equation (3)) to one for PPC. In contrast, the relationship for TChl-b is least affected by the package effect normalization of $a_{ph}(\lambda)$, with still increased R^2, but only slightly decreased MAE and MPE, and a relatively large deviation of B (Equation (3)) to the unity (0.35–0.60 for both 470 nm and 660 nm regardless of the normalization). Cross-validation results (Table 2 (b)) further confirm the improved performance of Gaussian decompostion in estimating TChl-b, Chlc1/2, PPC and PSC after taking the package effect into account (MPE 29.3–34.0% before the normalization and 20.5–27.4% afterwards). The pigment with the lowest estimated accuracy is now Chl-c1/2, but still with an improved accuracy following normalization. Similarly, TChl-b retrievals have slightly reduced MAE and MPE than those before the normalization.

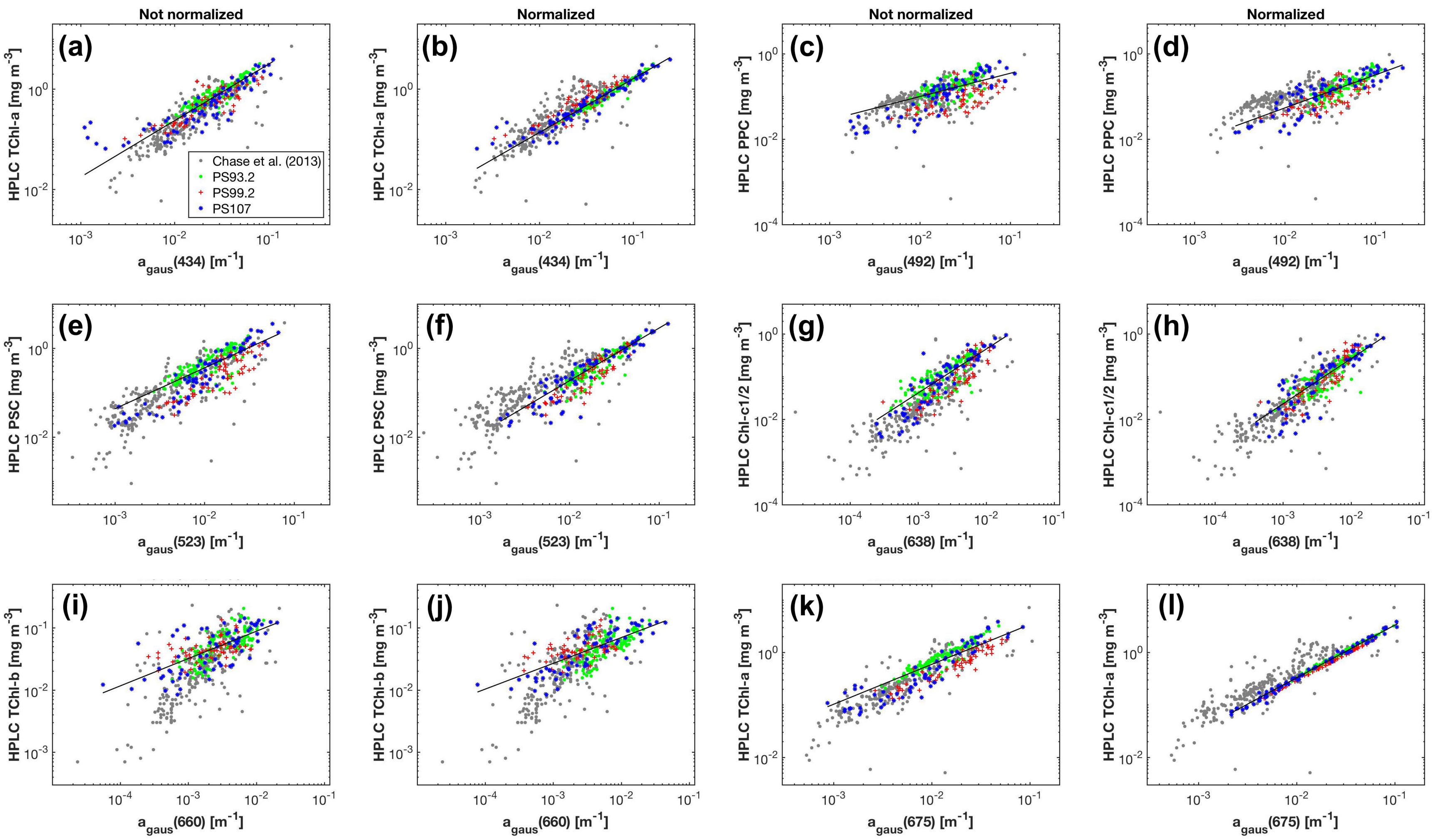


Figure 5. Concentrations of phytoplankton pigments measured by HPLC versus the magnitudes of the corresponding Gaussian functions obtained from the Gaussian decomposition of both $a_{ph}(\lambda)$ (**a**,**c**,**e**,**g**,**i**) and $\hat{a}_{ph}(\lambda)$ (**b**,**d**,**f**,**h**,**j**,**l**). The results of Chase et al. [22] are based on $a_{ph}(\lambda)$.

3.3. Matrix Inversion Technique

3.3.1. The Number of Pigment Types to Be Estimated

To apply the matrix inversion technique to estimate phytoplankton pigments, a key issue is to determine the number of pigments m included in matrix C in Equation (4). Table 3 summarizes the final selections of m that fulfilled the criteria mentioned in Section 2.2.2 and the corresponding n_{cond} and SI for all cases of the matrix inversion technique. The n_{cond} and maximum values of SI did not significantly change after data perturbations were introduced, regardless of the application of the package effect normalization.

As shown in Figure 6, the minimum values of n_{cond} for all pigment combinations increase with increasing m, indicating that the larger the number of pigment types to be estimated, the more sensitive the matrix inversion is to input errors. The largest value of m is nine and six for all combinations of Fram-20 and Bricaud-12 pigments, respectively, with the minimum value of n_{cond} smaller than the threshold 60. When composed of Gauss-5 pigments, matrix C has a n_{cond} of 47.8.

Based on the results fulfilling the n_{cond} criterion, when $\widetilde{a}^*(\lambda)$ is derived by SVD, the SI criterion allowed m to reach nine out of the Fram-20 pigments, and was satisfied by the Gauss-5 pigments. However, m for the NNLS-NNLS method decreased to six for of the Fram-20 pigments when package effect normalization was not performed, because the NNLS algorithm set some of the derived $\widetilde{a}^*(\lambda)$ in the pigment combinations ($m > 6$) to zero to avoid negative values [30]. When the package effect normalization was applied, the number of pigment combinations valid for SI calculation increased to nine, possibly because this normalization increases the inner differences of the set of the derived $\widetilde{a}^*(\lambda)$. However, the maximum SI value exceeded 0.9 for Gauss-5 pigments. Therefore, NNLS-NNLS-5′ method based on $\hat{a}_{ph}(\lambda)$ was not considered in the estimation of pigments. As for the Bricaud-12 pigments, the choice of SVD and NNLS influences the derived specific absorption of the missing pigments, as indicated by the different SI values. In this case, m reaches four. The specific pigments to be inverted for all cases are shown in Tables 4 and 5 and Table A1.

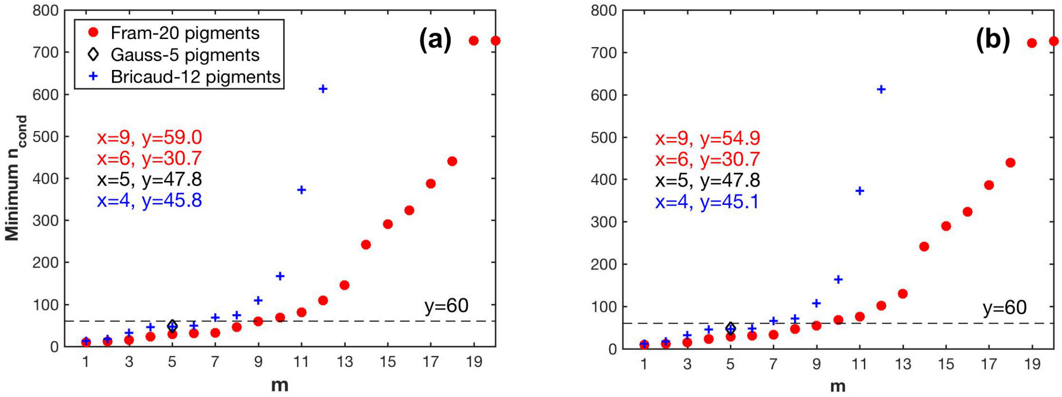


Figure 6. Variations in the minimum values of the condition number (n_{cond}) of matrix C in Equation (4) with different pigment combination (m pigment types to be estimated): (**a**) pigment data unperturbed; (**b**) pigment data perturbed.

Table 3. The m types of pigments to be estimated using the matrix inversion technique and the corresponding n_{cond} and maximum SI values.

(a) Pigment data perturbed

Method	**Pigments**	$a_{ph}(\lambda)$ **Based**			$\hat{a}_{ph}(\lambda)$ **Based**		
		n_{cond}	**Maximum** SI	m	n_{cond}	**Maximum** SI	m
SVD-NNLS	Fram-20	54.9	0.86	9	54.9	0.85	9
NNLS-NNLS	Fram-20	30.7	0.79	6	54.9	0.86	9
SVD-NNLS-5′	Gauss-5	47.8	0.81	5	47.8	0.84	5
NNLS-NNLS-5′	Gauss-5	47.8	0.79	5	47.8	-	5
Bricaud-SVD-NNLS	Bricaud-12	45.1	0.72	4	45.1	0.76	4
Bricaud-NNLS-NNLS	Bricaud-12	45.1	0.83	4	45.1	0.76	4

(b) $a_{ph}(\lambda)$ data perturbed

Method	**Pigments**	$a_{ph}(\lambda)$ **Based**			$\hat{a}_{ph}(\lambda)$ **Based**		
		n_{cond}	**Maximum** SI	m	n_{cond}	**Maximum** SI	m
SVD-NNLS	Fram-20	59.0	0.86	9	59.0	0.81	9
NNLS-NNLS	Fram-20	30.7	0.75	6	59.0	0.85	9
SVD-NNLS-5′	Gauss-5	47.8	0.74	5	47.8	0.82	5
NNLS-NNLS-5′	Gauss-5	47.8	0.78	5	47.8	-	5
Bricaud-SVD-NNLS	Bricaud-12	45.8	0.72	4	45.8	0.76	4
Bricaud-NNLS-NNLS	Bricaud-12	45.8	0.83	4	45.8	0.76	4

(c) Both pigment and $a_{ph}(\lambda)$ data perturbed

Method	**Pigments**	$a_{ph}(\lambda)$ **Based**			$\hat{a}_{ph}(\lambda)$ **Based**		
		n_{cond}	**Maximum** SI	m	n_{cond}	**Maximum** SI	m
SVD-NNLS	Fram-20	54.9	0.84	9	54.9	0.81	9
NNLS-NNLS	Fram-20	30.7	0.81	6	54.9	0.86	9
SVD-NNLS-5′	Gauss-5	47.8	0.79	5	47.8	0.82	5
NNLS-NNLS-5′	Gauss-5	47.8	0.80	5	47.8	-	5
Bricaud-SVD-NNLS	Bricaud-12	45.1	0.72	4	45.1	0.76	4
Bricaud-NNLS-NNLS	Bricaud-12	45.1	0.83	4	45.1	0.76	4

Table 4. Statistics for the Model-1 linear regressions between SVD-NNLS retrieved and measured pigment concentrations (regression coefficients were calculated with 95% confidence bounds). MAE is in mg m^{-3} (values outside the parentheses were calculated with linear-scale values, while inside the parentheses with log10-scale values), MPE in %, and N is the number of data points for the regressions. Unperturbed training data set was used.

(a) SVD-NNLS-9

Pigments	$a_{ph}(\lambda)$ **Based**						$\hat{a}_{ph}(\lambda)$ **Based**					
	Slope	**Intercept**	R^2	**MAE**	**MPE**	**N**	**Slope**	**Intercept**	R^2	**MAE**	**MPE**	**N**
TChl-a	0.97 ± 0.04	0.00 ± 0.01	0.92	0.15(0.10)	12.8	295	1.00 ± 0.00	−0.00 ± 0.00	1.00	0.01(0.01)	0.74	298
TChl-b	0.31 ± 0.15	−0.81 ± 0.20	0.31	0.04(0.32)	58.3	268	0.41 ± 0.15	−0.70 ± 0.20	0.25	0.04(0.30)	52.6	269
Chl-c1/2	0.65 ± 0.06	−0.21 ± 0.06	0.72	0.06(0.25)	42.6	286	0.60 ± 0.06	−0.27 ± 0.07	0.63	0.06(0.26)	37.1	275
But	0.43 ± 0.10	−0.53 ± 0.17	0.38	0.04(0.43)	100.0	209	0.54 ± 0.08	−0.45 ± 0.14	0.62	0.03(0.35)	62.6	206
Diadino	0.39 ± 0.07	−0.45 ± 0.08	0.40	0.06(0.30)	57.8	283	0.38 ± 0.08	−0.52 ± 0.09	0.46	0.06(0.30)	49.2	284
Fuco	0.84 ± 0.06	−0.03 ± 0.06	0.82	0.08(0.23)	33.7	276	0.85 ± 0.06	−0.07 ± 0.06	0.79	0.07(0.22)	31.6	286
Hex	0.62 ± 0.05	−0.16 ± 0.05	0.69	0.09(0.24)	37.3	266	0.72 ± 0.04	−0.13 ± 0.04	0.83	0.07(0.20)	32.0	270
Peri	0.74 ± 0.15	−0.30 ± 0.23	0.54	0.04(0.36)	57.5	134	0.67 ± 0.17	−0.37 ± 0.26	0.43	0.05(0.40)	66.6	128
Pheo-a	0.06 ± 0.87	−0.31 ± 0.65	-0.16	0.59(0.53)	166.0	22	−0.04 ± 1.02	−0.39 ± 0.77	−0.34	0.64(0.58)	132.0	22

Table 4. *Cont.*

(b) SVD-NNLS-5′

Pigments	$a_{ph}(\lambda)$ Based						$\hat{a}_{ph}(\lambda)$ Based					
	Slope	**Intercept**	R^2	**MAE**	**MPE**	**N**	**Slope**	**Intercept**	R^2	**MAE**	**MPE**	**N**
TChl-a	1.04 ± 0.04	−0.00 ± 0.02	0.91	0.16(0.12)	16.2	295	1.00 ± 0.00	−0.00 ± 0.00	1.00	0.02(0.01)	1.8	298
TChl-b	0.54 ± 0.21	−0.54 ± 0.29	0.12	0.05(0.38)	72.6	250	0.55 ± 0.19	−0.54 ± 0.26	0.18	0.05(0.34)	62.0	247
Chl-c1/2	0.55 ± 0.06	−0.25 ± 0.07	0.67	0.09(0.33)	62.4	268	0.52 ± 0.06	−0.31 ± 0.07	0.61	0.08(0.30)	59.1	269
PPC	0.50 ± 0.10	−0.26 ± 0.09	0.50	0.10(0.27)	55.2	278	0.51 ± 0.08	−0.25 ± 0.08	0.58	0.10(0.27)	55.9	288
PSC	0.66 ± 0.05	−0.02 ± 0.03	0.75	0.20(0.22)	37.8	288	0.70 ± 0.04	−0.02 ± 0.02	0.83	0.15(0.19)	32.1	292

(c) Bricaud-SVD-NNLS-4

Pigments	$a_{ph}(\lambda)$ Based						$\hat{a}_{ph}(\lambda)$ Based					
	Slope	**Intercept**	R^2	**MAE**	**MPE**	**N**	**Slope**	**Intercept**	R^2	**MAE**	**MPE**	**N**
TChl-a	0.85 ± 0.06	0.01 ± 0.02	0.75	0.26(0.15)	25.5	298	0.98 ± 0.01	0.01 ± 0.00	0.98	0.08(0.05)	5.3	298
Chl-c1/2	0.69 ± 0.04	−0.20 ± 0.05	0.80	0.05(0.19)	31.2	294	0.69 ± 0.06	−0.24 ± 0.07	0.65	0.06(0.22)	36.3	278
Diadino	0.60 ± 0.08	−0.22 ± 0.10	0.71	0.08(0.33)	75.6	245	0.68 ± 0.16	−0.04 ± 0.18	0.37	0.14(0.44)	111.8	206
Hex	0.10 ± 0.12	−0.66 ± 0.12	−0.01	0.32(0.60)	83.2	227	0.15 ± 0.13	−0.41 ± 0.14	0.10	0.40(0.62)	126.2	175

Table 5. Statistics of phytoplankton pigment retrieval using SVD-NNLS based on leave-one-out cross-validation. MAE is in mg m^{-3} (values outside the parentheses were calculated with linear-scale values, while inside the parentheses with log10-scale values) and MPE in %. "Perturb 1, 2 and 3" represent the input data with perturbations of pigment concentrations solely, $a_{ph}(\lambda)$ solely and both, respectively.

(a) SVD-NNLS-9

Pigments	Perturb 1 [a]		Perturb 2 [a]		Perturb 3 [a]		Perturb 1 [b]		Perturb 2 [b]		Perturb 3 [b]	
	MAE	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**
TChl-a	0.22(0.15)	22.3	0.17(0.12)	16.5	0.21(0.15)	21.0	0.01(0.01)	1.1	0.05(0.03)	4.3	0.05(0.03)	4.3
TChl-b	0.04(0.29)	60.2	0.03(0.27)	53.7	0.03(0.27)	53.6	0.04(0.28)	61.4	0.03(0.27)	52.9	0.04(0.27)	56.3
Chl-c1/2	0.07(0.26)	45.3	0.07(0.25)	41.1	0.08(0.26)	44.1	0.09(0.30)	52.9	0.08(0.27)	44.6	0.09(0.29)	50.4
But	0.03(0.32)	104.8	0.03(0.31)	80.7	0.03(0.31)	81.8	0.02(0.25)	68.7	0.02(0.26)	67.2	0.03(0.27)	69.8
Diadino	0.07(0.32)	64.4	0.08(0.31)	61.9	0.08(0.31)	65.2	0.07(0.32)	66.0	0.07(0.31)	59.7	0.08(0.32)	64.7
Fuco	0.09(0.22)	44.5	0.09(0.21)	36.9	0.09(0.22)	38.4	0.12(0.27)	53.1	0.10(0.23)	40.0	0.11(0.25)	44.5
Hex	0.09(0.24)	43.1	0.10(0.26)	42.6	0.11(0.26)	44.9	0.09(0.23)	42.4	0.08(0.22)	36.2	0.10(0.24)	42.1
Peri	0.02(0.17)	66.8	0.03(0.23)	90.4	0.03(0.23)	90.4	0.02(0.17)	68.3	0.02(0.21)	74.8	0.03(0.21)	76.3
Pheo-a	0.04(0.04)	123.7	0.02(0.03)	88.7	0.02(0.03)	90.0	0.04(0.04)	107.0	0.02(0.03)	95.2	0.02(0.03)	91.7

(b) SVD-NNLS-5′

Pigments	Perturb 1 [a]		Perturb 2 [a]		Perturb 3 [a]		Perturb 1 [b]		Perturb 2 [b]		Perturb 3 [b]	
	MAE	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**
TChl-a	0.22(0.17)	23.3	0.18(0.13)	17.3	0.22(0.16)	21.7	0.02(0.02)	2.2	0.05(0.03)	4.7	0.05(0.04)	4.9
TChl-b	0.04(0.32)	73.1	0.04(0.30)	58.5	0.04(0.30)	59.5	0.04(0.29)	66.2	0.04(0.28)	58.3	0.04(0.29)	61.2
Chl-c1/2	0.09(0.30)	68.5	0.07(0.26)	49.2	0.08(0.27)	50.4	0.09(0.31)	67.1	0.07(0.26)	47.9	0.09(0.29)	54.3
PPC	0.11(0.28)	62.1	0.09(0.24)	47.9	0.10(0.25)	51.1	0.11(0.28)	66.2	0.09(0.25)	53.2	0.10(0.26)	57.6
PSC	0.21(0.23)	42.8	0.20(0.22)	35.0	0.21(0.23)	38.0	0.20(0.24)	42.8	0.16(0.19)	28.7	0.20(0.23)	37.9

(c) Bricaud-SVD-NNLS-4

Pigments	Perturb 1 [a]		Perturb 2 [a]		Perturb 3 [a]		Perturb 1 [b]		Perturb 2 [b]		Perturb 3 [b]	
	MAE	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**
TChl-a	0.26(0.16)	26.3	0.26(0.15)	25.3	0.27(0.16)	26.6	0.08(0.05)	5.7	0.10(0.06)	8.3	0.10(0.06)	8.4
Chl-c1/2	0.05(0.19)	33.2	0.05(0.20)	32.6	0.05(0.20)	33.5	0.06(0.22)	39.0	0.06(0.23)	41.6	0.06(0.23)	43.1
Diadino	0.06(0.28)	77.2	0.07(0.28)	78.2	0.07(0.28)	80.1	0.10(0.31)	118.1	0.10(0.31)	117.6	0.10(0.32)	118.3
Hex	0.24(0.46)	83.6	0.24(0.44)	82.1	0.24(0.45)	82.7	0.23(0.37)	123.2	0.24(0.37)	112.2	0.24(0.38)	109.8

[a] $a_{ph}(\lambda)$ based; [b] $\hat{a}_{ph}(\lambda)$ based.

3.3.2. SVD-NNLS

The pigment-specific absorption coefficients obtained from SVD-NNLS-9, SVD-NNLS-5′ and Bricaud-SVD-NNLS-4 without input data perturbations are displayed in Figure 7. Each SVD derived specific spectra varies smoothly across the full bands and sufficiently differs from each other ($SI < 0.9$). Negative coefficients are permissible because these are mathematical constructs. Pigment concentration estimates were then solved via NNLS and compared to HPLC pigment concentrations. Training statistics (Table 4) and the scatter plots (Figure 8) show that the estimated and the corresponding measured pigments were correlated ($R^2 > 0.30$) except for Pheo-a (SVD-NNLS-9), TChl-b (SVD-NNLS-5′) and Hex (Bricaud-SVD-NNLS-4). For SVD-NNLS-9, the training errors for all pigments except for Peri and Pheo-a were reduced by the package effect normalization (Table 4 (a)). In contrast, this normalization increased training errors when using Bricaud-SVD-NNLS-4 (Table 4 (c)). For SVD-NNLS-5′, all Gauss-5 pigments except for PPC were observed to have smaller training errors with this normalization (Table 4 (b)).

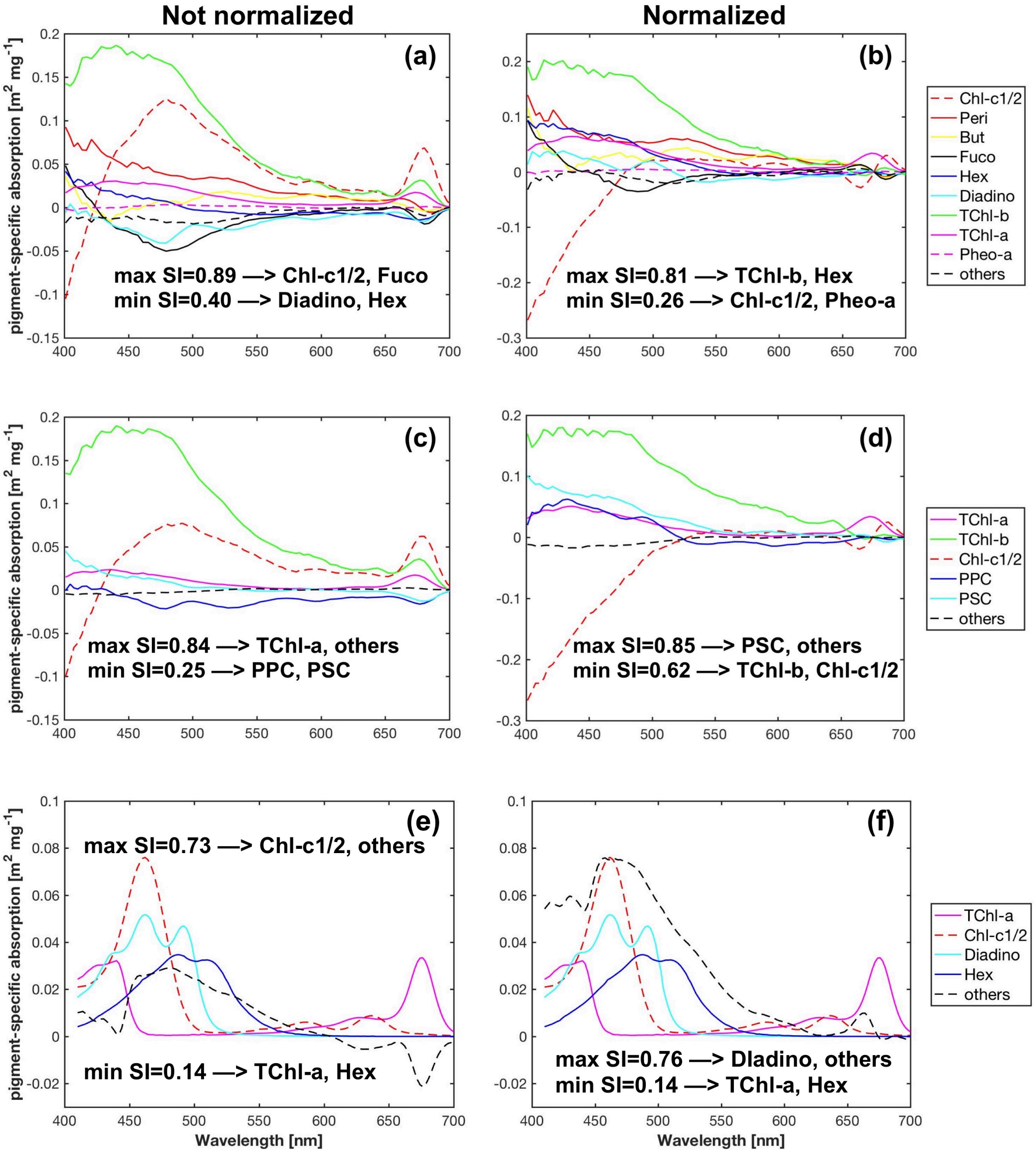


Figure 7. Pigment-specific absorption spectra obtained from SVD-NNLS-9 (**a**,**b**), SVD-NNLS-5′ (**c**,**d**) and Bricaud-SVD-NNLS-4 (**e**,**f**) without data perturbations, respectively. Cases with and without package effect normalization were compared.

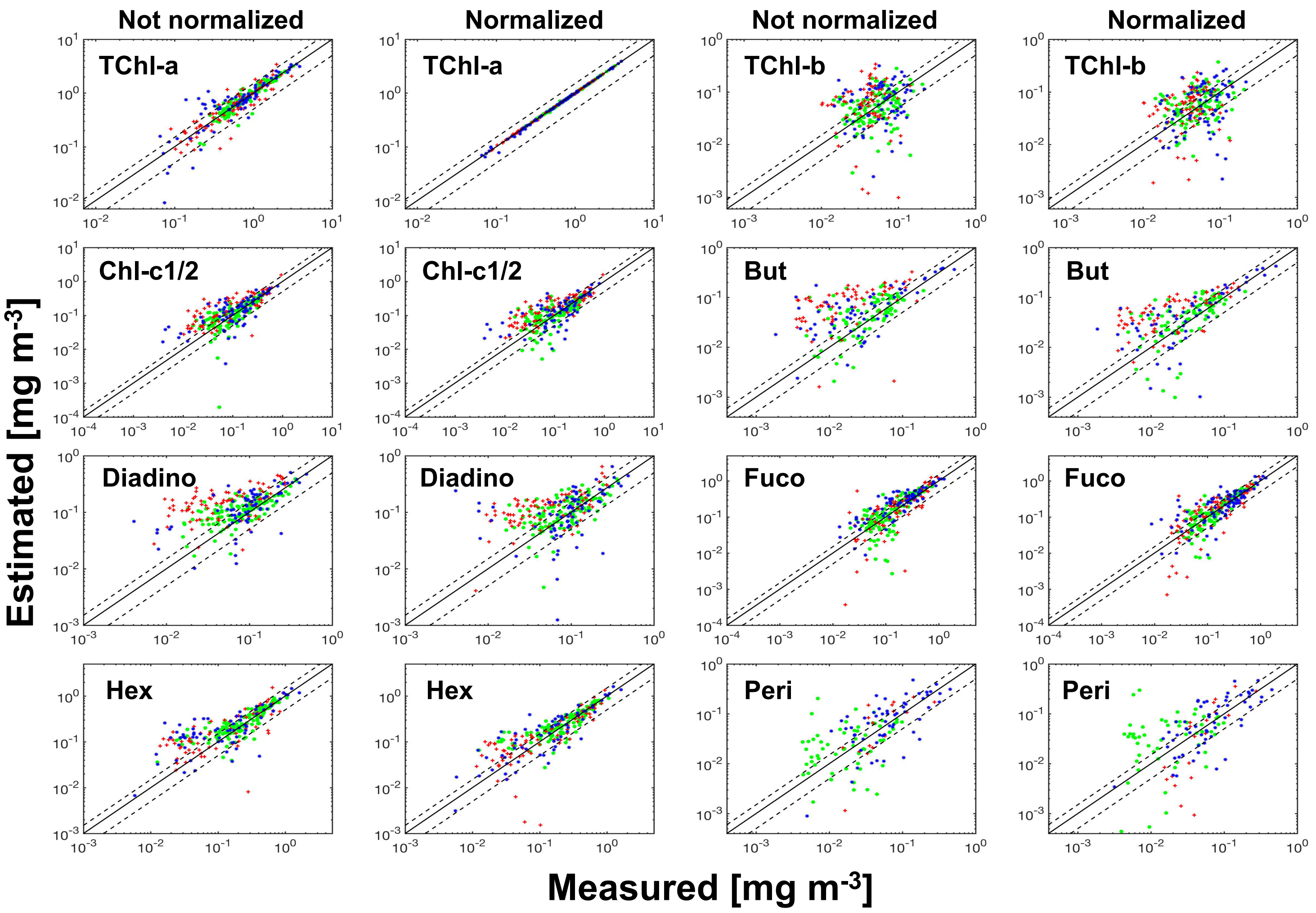


Figure 8. Scatter plots of SVD-NNLS-9 estimated pigment concentrations versus measured pigment concentrations for unperturbed training data set. The symbols of green circles, red crosses and blue stars represent the data from the cruises PS93.2, PS99.2 and PS107, respectively. Dash lines denote the 50% error lines, and solid lines are one-to-one lines.

To obtain robust pigment estimation statistics and evaluate the influences of the package effect normalization on them, data perturbations and cross-validation were combined to generate the test errors (Table 5). The SVD-NNLS-9 method (Table 5 (a)) exhibited stable prediction accuracy (MPE 16–65%) for six types of pigments, i.e., TChl-a, TChl-b, Chlc-1/2, Diadino, Fuco and Hex, in which MPE varied less than 10% and MAE less than a factor of 1.2 with input data perturbations. TChl-a had the least prediction error (MAE 0.17–0.22 mg m^{-3}, MPE 16–22%); Fuco, Hex and Chl-c1/2 shared the second best estimation (MPE 35–45%); Diadino was least accurately estimated (MAE 0.07–0.08 mg m^{-3}, MPE 61–65%), and TChl-b showed a slightly smaller MPE of 53–60% (MAE 0.03–0.04 mg m^{-3}). Though included in the calculation, But, Peri and Pheo-a exhibited relatively inconsistent prediction errors with the three cases of data perturbations, likely because of their relatively low concentrations (Table 1) and infrequent occurrence in the data set. When considering the package effect normalization, however, all nine pigments except for Pheo-a achieved stable prediction statistics (MPE 36–76%) with the different cases of data perturbations (MPE varied less than 13% and MAE less than a factor of 1.3, Table 5 (a)). The normalization obtained comparable prediction errors as those without normalization for TChl-b, Diadino and Hex, 8% higher of MPE for Chlc-1/2 and Fuco, and additional estimations of But (MPE 67–70%) and Peri (MPE 68–75%).

For SVD-NNLS-5′ (Table 5 (b)), all five pigments were stably retrieved (MPE 17–73%). TChl-a and PSC had the lowest prediction errors (MPE 17–23% and 35–43%, respectively), whereas TChl-b was the most poorly estimated (MPE 58–73%). Chl-c1/2 and PPC exhibited similar retrieval accuracy (MPE 48–69%). In comparison, the estimation errors with the application of the package effect normalization did not significantly change for TChl-b, Chlc-1/2 and PSC, and slightly increased for PPC.

The Bricaud-SVD-NNLS-4 method provided stable retrieval statistics for all four pigments (Table 5 (c)). The prediction errors increased in the order of TChl-a (MPE 25–27%), Chlc-1/2 (MPE 32–34%), Diadino (MPE 77–80%) and Hex (MPE 82–84%). When using $\hat{a}_{ph}(\lambda)$, the overall performance of this method was significantly hindered for Diadino and Hex (MPE increased ~40%) and slightly affected Chl-c1/2 (MPE increased ~10%).

3.3.3. Intercomparison between SVD-NNLS Applications

SVD-NNLS-9 and SVD-NNLS-5′ showed similar capabilities in retrieving TChl-a, and both methods outperformed Bricaud-SVD-NNLS-4 (Table 5). SVD-NNLS-9 surpassed SVD-NNLS-5′ in retrieving TChl-b and Bricaud-SVD-NNLS-4 in the retrieval of Diadino and Hex, respectively. For Chl-c1/2, Bricaud-SVD-NNLS-4 performed the best and SVD-NNLS-5′ the worst. Considering the overall pigment retrieval accuracy and the number of pigment types possible to retrieve, SVD-NNLS-9 performed best among the three SVD-NNLS methods.

3.3.4. Feasibility of SVD-NNLS-9 for Multispectral $a_{ph}(\lambda)$

The performance of the matrix inversion technique is sensitive to the number of wavebands and their locations on the $a_{ph}(\lambda)$ spectrum [27,33]. To test the feasibility of this technique using multispectral $a_{ph}(\lambda)$, SVD-NNLS-9 was performed with $a_{ph}(\lambda)$ at ten MODIS bands, i.e., 412, 443, 469, 488, 531, 547, 555, 645, 667 and 678 nm. In this case, only four types of pigments, i.e., TChl-a, TChl-b, Chlc-1/2 and Hex, were retrieved with stable and acceptable estimation statistics (Table 6) both with and without package effect normalization. The estimation errors of TChl-a and TChl-b slightly increased by 4 to 14% using multispectral $a_{ph}(\lambda)$, while for Chlc-1/2 and Hex, the MPE were approximately 30% and 20% higher than those with hyperspectral $a_{ph}(\lambda)$, respectively.

Table 6. Statistics of phytoplankton pigment retrieval using SVD-NNLS-9 with $a_{ph}(\lambda)$ at ten MODIS bands based on leave-one-out cross-validation. MAE is in mg m^{-3} (values outside the parentheses were calculated with linear-scale values, while inside the parentheses with log10-scale values) and MPE in %. "Perturb 1, 2 and 3" represent the input data with perturbations of pigment concentrations solely, $a_{ph}(\lambda)$ solely and both, respectively.

Pigments	Perturb 1 [a]		Perturb 2 [a]		Perturb 3 [a]		Perturb 1 [b]		Perturb 2 [b]		Perturb 3 [b]	
	MAE	MPE	MAE	MPE	MAE	MPE	MAE	MPE	MAE	MPE	MAE	MPE
TChl-a	0.25(0.18)	28.8	0.21(0.14)	19.8	0.28(0.18)	29.9	0.01(0.01)	0.3	0.03(0.02)	2.0	0.03(0.02)	1.9
TChl-b	0.04(0.28)	60.7	0.04(0.24)	68.2	0.03(0.23)	67.0	0.06(0.31)	74.7	0.04(0.24)	67.2	0.04(0.25)	73.3
Chl-c1/2	0.13(0.37)	85.3	0.16(0.38)	78.3	0.15(0.37)	80.9	0.18(0.43)	95.0	0.14(0.37)	72.2	0.18(0.40)	87.6
But	0.06(0.39)	185.5	0.08(0.41)	265.0	0.08(0.40)	266.3	0.03(0.30)	87.1	0.05(0.33)	162.1	0.05(0.34)	165.8
Diadino	0.15(0.49)	151.7	0.24(0.48)	188.5	0.23(0.45)	196.0	0.18(0.49)	168.5	0.20(0.45)	165.4	0.23(0.47)	193.4
Fuco	0.24(0.38)	114.7	0.23(0.34)	76.9	0.21(0.32)	76.0	0.35(0.47)	148.4	0.23(0.36)	78.8	0.31(0.39)	97.1
Hex	0.15(0.33)	64.7	0.18(0.34)	70.4	0.19(0.34)	73.4	0.18(0.34)	73.3	0.13(0.28)	58.4	0.18(0.31)	73.2
Peri	0.02(0.14)	89.7	0.04(0.26)	196.6	0.05(0.26)	211.8	0.02(0.12)	82.2	0.03(0.21)	141.0	0.04(0.22)	148.6
Pheo-a	0.07(0.05)	312.4	0.05(0.04)	296.1	0.06(0.04)	281.0	0.08(0.05)	341.4	0.05(0.04)	309.0	0.05(0.04)	285.9

[a] $a_{ph}(\lambda)$ based; [b] $\hat{a}_{ph}(\lambda)$ based.

3.4. Gaussian Decomposition versus SVD-NNLS

Considering both MAE and MPE, the Gaussian decomposition method revealed comparable capability in estimating TChl-a compared to SVD-NNLS-9 and SVD-NNLS-5′ (Tables 2 and 5 (a,b)). However, it outperformed SVD-NNLS-5′ in retrieving TChl-b (MPE ~ 40% lower), Chl-c1/2 (MPE ~ 35% lower), PSC (MPE ~ 8% lower) and PPC (MPE ~ 30% lower) and surpassed SVD-NNLS-9 in predicting TChl-b (MPE ~ 30% lower) and Chl-c1/2 (MPE ~ 12% lower). When the package effect normalization was applied, the results from Gaussian decomposition further improved the estimation of the Gauss-5 pigments (TChl-a excluded).

4. Discussion

The study of phytoplankton dynamics in relation to a changing climate requires access to high resolution phytoplankton pigment information in space and time. Unlike data obtained from HPLC analysis of discrete water samples, underway spectrophotometry is capable of providing nearly continuous in situ data records of spectral particulate absorption as well as phytoplankton absorption. The latter is dependent on the concentrations and composition of phytolankton pigments. Therefore, algorithms that link hyperspectral $a_{ph}(\lambda)$ obtained from underway spectrophotometry to the concentrations of various phytoplankton pigments provide pigment information with high spatial resolution (~300 m for one minute binned-averaged spectra when the ship is moving at ~10 knots). This pigment information can support the evaluation of ocean color algorithms and coupled hydrodynamic-biological modelling. With the advancement of hyperspectral radiometers, these algorithms have great potential to be incorporated into the inversion of satellite ocean color measurements for the synoptic detection of phytoplankton pigments and thus the monitoring of phytoplankton spatial and temporal dynamics.

Gaussian decomposition is an effective pigment retrieval algorithm that dates back to 1990s [11] and was recently modified for more extended applications, e.g., [22–25]. At the time this manuscript is drafted, it is the only method that has been applied to underway spectrophotometry data and successfully retrieved the concentrations of various pigments [22]. Compared with Gaussian decomposition and other methods from previous studies, e.g., [14,15,19,26] that are incapable of resolving PPC and PSC, the matrix inversion technique is capable of estimating these and various marker pigments indicative of phytoplankton composition [27,28,34]. Both Gaussian decomposition and the matrix inversion technique rely on the physical links between phytoplankton pigments and their distinct light absorption properties, while other methods' (e.g., principle component analysis) output is often physically uninterpretable.

To assess the utility of the two methods and improve their performances in our study area, we improved the Gaussian decomposition algorithm from Chase et al. [22] by considering pigment package effect and reconsidered the matrix inversion technique from Moisan et al. [27] by taking into account matrix conditioning. In the following discussion, we compare the results from this study to those from the literature, highlight the improvements we have made, and show applications to underway absorption data.

4.1. Gaussian Decomposition

Our $a_{gaus}(\lambda_0)$-pigment data falls into the range of the global data set from Tara expeditions [22] and has a large portion of overlap with the global data (Figure 5, Table 7) except for TChl-b (Figure 5i), likely due to the lack of low TChl-b concentrations in this study. Furthermore, our $a_{gaus}(\lambda)$-pigment data shows less scatter (Figure 5a,c,e,g,i,k), and the estimation errors for all Gauss-5 pigments are lower than those of the Tara data set (MPE 7.5–20% lower). This is probably because of a higher level of variability in pigment package effect of the global data set than that in the Fram Strait, as indicated by the greater quartile coefficient of dispersion of the pigment-specific absorption coefficient ($a_i^*(\lambda_0)$, defined as the ratio of $a_{gaus,i}(\lambda_0)$ to the corresponding pigment concentration) for the global data set (Table 7). In addition, it could also be due to the errors introduced by the extent of the Tara expeditions study (2.5 years) and resulting increased potential for methodological variability, while only four people were involved in discrete sampling in the current study.

After applying the package effect normalization, as expected, almost all the $a_{gaus}(675)$-TChl-a data points fall on the regression line (Figure 5l, Table 2), i.e., no additional package effect was found at $a_{gaus}(675)$, suggesting the successful simplification of $Q_a^*(675)$ calculation using only TChl-a concentration. Both the R^2 for $a_{gaus}(434)$ to TChl-a correlation and the regression coefficient B (Equation (3)) are close to one, indicating that the variations in the magnitude of $Q_a^*(675)$ account for a large proportion of the variations in the package effect due to TChl-a at 434 nm. Similar improvements have been observed for Chl-c1/2, PSC, PPC and to a less extent, TChl-b, which implicates the covariation between the package effect of these pigments at the corresponding wavelengths and that of TChl-a. This is also proved by the reduced data dispersion of $a_i^*(\lambda_0)$ except for TChl-b (Table 7) and the strong power law relationships (R^2 0.39–0.87) between $a^*_{TChl\text{-}a}(434)$, $a^*_{TChl\text{-}b}(660)$, $a^*_{Chl\text{-}c1/2}(638)$, $a^*_{PPC}(492)$, $a^*_{PSC}(523)$ and $a^*_{TChl\text{-}a}(675)$, respectively. The package effect of a specific pigment is a function of the concentration of this pigment as well as phytoplankton cell size [76]. The strong correlations between TChl-a concentration and the concentrations of Chl-c1/2 (Spearman's ρ = 0.9), PSC (Spearman's ρ = 1.0), PPC (Spearman's ρ = 0.8), and of TChl-b (Spearman's ρ = 0.7) (Figure 4b) explain the covariation of the package effect by the pigments.

In essence, the Gaussian decomposition method takes advantage of the absorption-pigment concentration correlation. The shape and magnitude of $a_{ph}(\lambda)$ is controlled by phytoplankton pigment composition and the level of the pigment package effect. When applying Gaussian decomposition, the absorption of individual pigments were separated and related to corresponding pigment concentrations. Therefore, the effects of pigment composition and package were separated, and the observed variations in absorption-pigment concentration correlation are mainly attributed to the package effect. The covarying absorption by more than one pigment that failed to be separated by this method is also a reason for these variations. Though simplified, the package effect is for the first time taken into account during Gaussian decomposition of $a_{ph}(\lambda)$ and improved the pigment estimation accuracy. It provides new insight into increasing pigment retrieval accuracy via the combined use of concurrent $a_{ph}(\lambda)$ and TChl-a concentration either obtained from field instrumental measurements/estimates or satellite data. To test the applicability of this normalization to underway spectrophotometry data when HPLC TChl-a data is not available, TChl-a was firstly calculated from the AC-S derived $a_{ph}(675)$ via the cruise-specific power functions (see Section 3.1). In this case, the improvement was also found with the normalization (Table 8). However, when TChl-a was derived either from $a_p(440)$ using the power law relationship described in Section 3.1 or the global relationship from Bricaud et al. [71], no improvement was observed (results not shown), possibly because of the more accurate TChl-a derived by the cruise-specific relationships. Therefore, the improved performance with this normalization relies on the accurate measurement or derivation of TChl-a data. An improved method for calculating pigment package effect is needed so that the calculation is independent of HPLC TChl-a data. Lin et al. [78] developed an optimization approach in estimating $Q_a^*(675)$ from $a_{ph}(\lambda)$ in the range of 650–700 nm based on pigment package effect theory. We tested this method and found that the calculated $Q_a^*(675)$ did not improve retrievals (results not shown). In summary, better results can be expected by further deciphering the influence of pigment package effect on the absorption-pigment concentration relationship.

Table 7. The range of values, median and quartile coefficient of dispersion (CD) for the Gaussian decomposition derived pigment-specific absorption coefficient at the corresponding wavelength $a_i^*(\lambda_0)$ (in m^2 mg^{-1}).

λ_0 (nm)	Pigment	Decomposition of $a_{ph}(\lambda)$			Decomposition of $\hat{a}_{ph}(\lambda)$			Chase et al. (2013) [22]		
		Range	Median	CD [%]	Range	Median	CD [%]	Range	Median	CD [%]
434	TChl-a	0.006-0.153	0.036	21.2	0.020-0.179	0.064	12.9	0.015-0.165	0.065	35.5
675	TChl-a	0.007-0.060	0.017	28.8	0.024-0.046	0.030	5.9	0.007-0.065	0.019	22.6
660	TChl-b	0.003-0.346	0.060	38.7	0.006-0.540	0.100	44.5	0-0.408	0.072	43.0
638	Chl-c1/2	0.004-0.163	0.024	40.0	0.010-0.333	0.039	31.6	0.010-0.247	0.051	41.2
492	PPC	0.059-0.827	0.142	31.9	0.077-1.012	0.253	21.5	0.049-0.797	0.097	40.7
523	PSC	0.011-0.192	0.029	41.5	0.024-0.483	0.048	28.3	0.010-0.243	0.035	47.2

Table 8. Statistics of phytoplankton pigments retrieval using Gaussian decomposition with package effect normalization based on leave-one-out cross-validation. Package effect normalization was performed with $c_{TChl\text{-}a}$ in Equation (7) calculated using cruise-specific $a_{ph}(675)$ (AC-S)-TChl-a (HPLC) relationships (see Section 3.1). MAE values outside the parentheses were calculated with linear-scale values, while inside the parentheses with log10-scale values.

λ_0 [nm]	Pigment	MAE [mg m^{-3}]	MPE [%]
434	TChl-a	0.19(0.13)	19.2
675	TChl-a	0.14(0.09)	12.4
660	TChl-b	0.02(0.16)	28.0
638	Chl-c1/2	0.06(0.18)	30.7
492	PPC	0.05(0.15)	24.5
523	PSC	0.16(0.16)	25.6

4.2. Matrix Inversion Technique

Compared to the seemingly physically interpretable NNLS derived $\widetilde{a}^*(\lambda)$ and the measured $a^*(\lambda)$ from extracted pigments in solution, the usage of the SVD derived $\widetilde{a}^*(\lambda)$ provided the most robust pigment estimates, though non-negative specific absorption cannot be guaranteed. This is consistent with the previous study from Moisan et al. [27]. It is worth noting that neither of the two types of $\widetilde{a}^*(\lambda)$ nor the measured $a^*(\lambda)$ are representative of the real in vivo "unpacked" pigment-specific absorption coefficient used in the reconstruction model. The former are pure mathematical optimization solutions based on least squares or constrained least squares minimization.

The different estimation errors of SVD-NNLS-9 and SVD-NNLS-5′ for TChl-b and Chlc-1/2 (Table 5 (a,b)) suggest the performance of SVD-NNLS in estimating pigments differs with the different way of grouping pigments for matrix C (Equation (4)). The estimation accuracy of a specific pigment using SVD-NNLS-9 and SVD-NNLS-5′ increases with the increase of the amount of the pigment in the samples, which was also observed by Moisan et al. [27]. The similar estimation statistics of all three SVD-NNLS methods for TChl-a (Table 5) is probably because TChl-a is the main pigment of phytoplankton and contributes the most to $a_{ph}(\lambda)$.

Moisan et al. [27] found that with the package effect normalization, the SVD-NNLS method yielded much better predictions of pigments. In this study, when considering the training accuracy of

SVD-NNLS-9 and SVD-NNLS-5′ (Table 4 (a,b)), most of the pigments showed improved MAE and MPE with this normalization. However, the cross-validation results calculated by taking into account the data perturbations (Table 5 (a,b)) showed randomly improved, reduced, or similar pigment prediction errors after the application of the normalization. The reason for this inconsistency is probably due to the fact that both Moisan et al. [27] and the training statistics in this study did not take into account the sensitivity of matrix inversion to input errors. Moisan et al. [27] also pointed out that the inverse model solutions are sensitive to the level of errors in the measured $a_{ph}(\lambda)$. The inconsistency between the results of the training (Table 4) and test errors (Table 5) confirms this sensitivity and that in this case, the training statistics are not appropriate for use to indicate pigment estimation errors. The results from cross-validation with data perturbations, on the other hand, encompassed the training statistics as their one special case and effectively reduced the sensitivity of the SVD-NNLS method to provide robust and stable pigment estimation statistics. Nevertheless, when performing package effect normalization on SVD-NNLS-9, two more pigments with lower concentrations also obtained robust statistics, possibly due to the enhancement of the differences between $\widetilde{a}^*(\lambda)$ of different pigments.

The sensitivity of the matrix inversion technique comes from the ill-conditioning of the linear systems (Equations (4) and (5)), which originates from the multicollinearity of phytoplankton pigments. In natural water samples, the multicollinearity of phytoplankton pigments (Figure 4b) are physiologically unavoidable. In other words, the ill-conditioning of Equations (4) and (5) is not completely avoidable. To reduce the degree of the ill-conditioning, the choice of the pigments included in matrix C (Equation (4)) is crucial. Our proposed n_{cond} and SI criteria to determine which pigments should be inverted for is empirically based on the understanding of the characteristics of the input data and many trials. The thresholds of n_{cond} and SI may change from case to case. The sensitivity analysis based on data perturbations (Table 3) shows the relative stable values of n_{cond} and SI. This is consistent with the singular value perturbation theorems [79], i.e., the singular values of a matrix are very stable with respect to changes in the elements of the matrix, because the n_{cond} is by definition the ratio of the largest to smallest singular values of a matrix. In contrast, though aware of the multicollinearity issue, Moisan et al. [27] did not pre-select the pigments to be determined. Instead, all types of pigments available from HPLC were included in the inverse modelling, which was also tested by this study (results not shown) and can lead to reduction of the pigment retrieval accuracy.

4.3. Applications

With the particulate absorption data collected by underway AC-S flow-through system, phytoplankton pigment concentrations along the cruise tracks are retrieved. Figure 9 shows an example of the underway Fuco and Hex estimated by SVD-NNLS-9. Fuco and Hex are dominant in diatoms and prymnesiophytes, respectively, which are the two common phytoplankton groups in the Fram Strait, e.g., [56]. However, we have to bear in mind that differentiating phytoplankton groups by marker pigments can be problematic, as there is substantial variability in pigment concentrations as a function of physiological responses to the environmental conditions. More importantly, a given marker pigment can be present in several phytoplankton groups (e.g., Fuco in diatoms and prymnesiophytes; more details in Wright and Jeffrey [80]). Data from the year 2015 indicates a co-prosperity of diatoms and prymnesiophytes. Overall, the concentrations of Fuco are relatively higher than those of Hex, which possibly reflects an overall higher biomass of diatoms (Figure 9a). In the year 2016, while this co-prosperity continued, the Hex concentrations exceeded the concentrations of Fuco in the western part of our study area (2 July 2016) (Figure 9b). In contrast, the year 2017 experienced an overall higher concentrations of Hex than Fuco (Figure 9c). These results are consistent with previous observations of the shift of phytoplankton assemblages from diatoms to *Phaeocystis spp.* (a type of prymnesiophyte) during the summer months in the Fram Strait [56,58,59]. In the future, access to similar high resolution phytoplankton pigment data verified by microscopic and flow cytometric techniques could support the studies on biogeophysical coupling in the Fram Strait and further enhance our understanding in the responses of phytoplankton community composition and physiology to climate change.

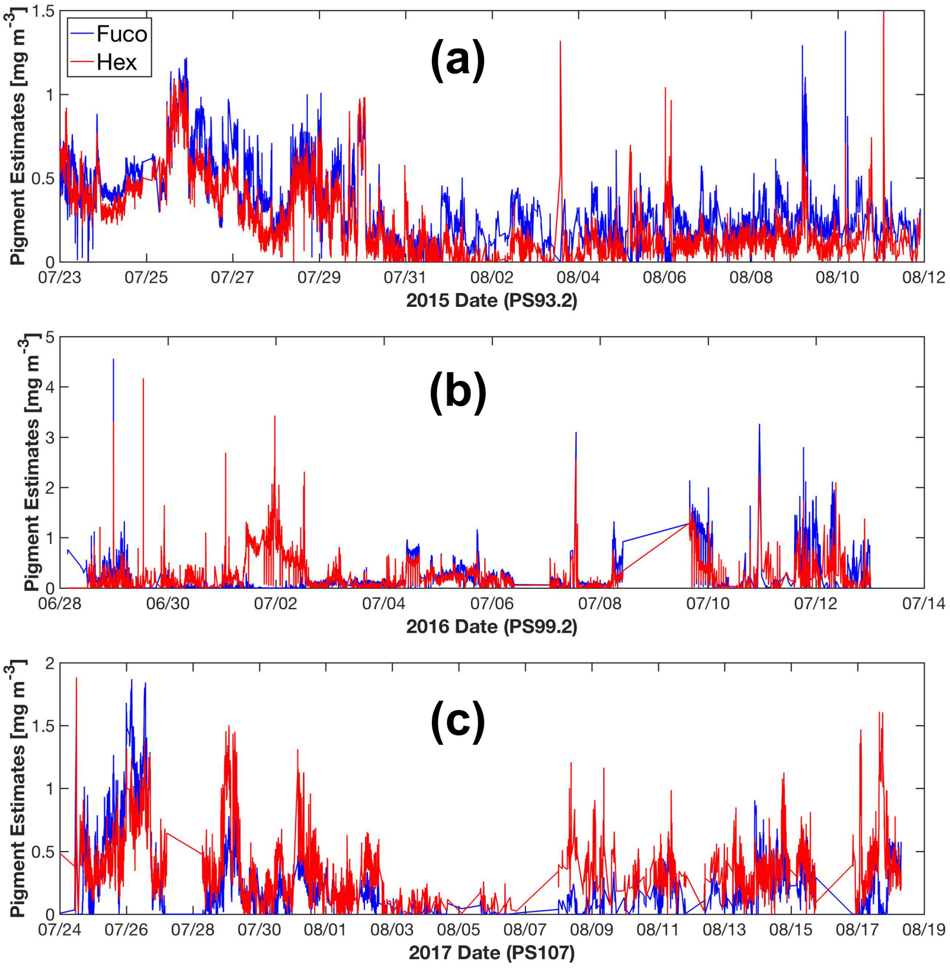


Figure 9. SVD-NNLS-9 estimated Fuco and Hex concentrations from underway spectrophotometry during the cruise periods of PS93.2 (**a**), PS99.2 (**b**) and PS107 (**c**).

5. Conclusions

We demonstrated the retrieval of high spatially resolved phytoplankton pigment concentrations in the Fram Strait and its vicinity from underway hyperspectral $a_{ph}(\lambda)$ (400–700 nm, ~3.5 nm wavelength resolution) by the application of Gaussian decomposition [22] and the matrix inversion technique [27]. Gaussian decomposition enables robust predictions of Gauss-5 pigments (MPE 21-34%). Improved retrieval accuracy was obtained by normalizing the $a_{ph}(\lambda)$ spectra with the pigment package effect factor at 675 nm. For the matrix inversion technique, although SVD cannot guarantee the derivation of non-negative pigment-specific absorption spectra, it generates more accurate pigment estimates compared to the NNLS derived spectra or the measured spectra from pigments in solution. To minimize the effect of the ill-conditioned matrices on pigment retrieval accuracy, we propose an innovative approach in selecting the pigments to be determined based on the combined use of data perturbations and leave-one-out cross-validation to generate robust pigment estimation statistics. Considering the overall pigment retrieval accuracy, SVD-NNLS-9 performed best among the three SVD-NNLS methods. The SVD-NNLS-9 method enables the robust estimations of six pigments (MPE 16-65%), i.e., TChl-a, TChl-b, Chl-c1/2, Diadino, Fuco and Hex, and two more being less accuraely estimated (MPE 67–76%), i.e., But and Peri, with the application of the package effect normalization. Gaussian decomposition outperforms SVD-NNLS-5′ in retrieving the TChl-b, Chl-c1/2, PPC and PSC, while both methods show similar capability in estimating TChl-a.

The matrix inversion technique has the advantage of retrieving the concentrations of several specific carotenoids, which is currently not accomplished by Gaussian decomposition, derivative analysis [13], partial least squares regression [14], and multiple linear regression [15]. However, its performance is sensitive to input errors when the input matrix is to some extent ill-conditioned. Therefore, sensitivity analysis such as the one based on data perturbations used in our study is always needed when assessing the performance of the matrix inversion technique in retrieving phytoplankton pigments or pigment related parameters. Future studies using methods such as principle component analysis and artificial neural network may show promise to obtain not only chlorophylls but also different types of carotenoids in our study area.

In addition to the number of pigments, the number of spectral bands used for pigment retrieval also significantly influence the performance of the matrix inversion technique. Compared with the results using hyperspectral $a_{ph}(\lambda)$, the number of pigments able to be retrieved by SVD-NNLS-9 was reduced to four, i.e., TChl-a, TChl-b, Chl-c1/2 and Hex, with increased estimation errors, especially for Chl-c1/2 and Hex, when using multispectral $a_{ph}(\lambda)$ (at ten MODIS bands). This suggests the advantage of using hyperspectral data for increasing the accuracy of phytoplankton pigment retrievals. It follows that $a_{ph}(\lambda)$ inverted from hyperspectral remote sensing reflectance measured by in situ or satellite radiometry has a greater potential for the application of Gaussian decomposition and the matrix inversion technique than multispectral radiometric measurements.

To apply Gaussian decomposition or the matrix inversion technique to a study area, prior knowledge of concurrent AC-S derived $a_{ph}(\lambda)$ and HPLC pigment concentrations in this region is necessary to derive either the regional $a_{gaus}(\lambda_0)$-pigment concentration relationship or the regional pigment-specific absorption spectra. With this knowledge, we apply both approaches to underway AC-S measurements in times when no HPLC data is available. Given that proxy-relation may change in the future, it is imperative to always collect some HPLC data to validate that derived relations or coefficients are still consistent.

The application of the two methods to our data obtain in three Fram Strait expeditions enables the derivation of pigment data sets along the cruise tracks. Future work could build upon these results, by deriving phytoplankton functional types based on retrieved marker pigments from hyperspectral phytoplankton absorption as well as hyperspectral remote sensing reflectance data. Such a high resolution data set will strengthen the study of phytoplankton dynamics in responses to environmental variables in the context of climate change.

Supplementary Materials: The quality controlled particulate absorption data from underway AC-S flow-through system, the HPLC phytoplankton pigments from discrete samples, and the estimated pigment concentrations along cruise tracks mentioned in this paper are available on PANGAE: https://doi.pangaea.de/10.1594/PANGAEA.894875. The MATLAB codes for data processing are available online at https://github.com/phytooptics.

Author Contributions: A.B. and Y.L. conceived and designed the experiments; Y.L. collected HPLC pigments data from the expeditions PS99.2 and PS107, and processed AC-S absorption data from all three expeditions; Y.L. led the data analysis and all coauthors assisted in it; E.B., A.C., A.B., H.X., X.Z. and R.R. helped with the interpretation of the data; E.B. and A.C. contributed the software of Gaussian decomposition and its application to this study; Y.P. contributed to programming and data visualization; Y.L. drafted the manuscript and all coauthors provided substantial comments and suggestions to improve it.

Funding: This research was funded by the Helmholtz Association Infrastructure Programme FRontiers in Arctic marine Monitoring (FRAM), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 268020496—TRR 172, within the Transregional Collaborative Research Center "ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)^3" project C03, and NASA Ocean biology and biogeochemistry grant NNX15AC08G. Y.L.'s Ph.D. fellowship is funded by the China Scholarship Council. Y.L.'s travel grants were funded by POLMAR Helmholtz Graduate School for Polar and Marine Research of the Alfred Wegener Institute.

Acknowledgments: We thank the captains, crew and scientists, espcially the chief scientists—Thomas Soltwedel and Ingo Schewe for supporting our work on R/V Polarstern during the expeditions PS93.2, PS99.2 and PS107. We thank Sonja Wiegmann for the HPLC analysis of pigments from the filtered water samples. We thank Sonja Wiegmann, Marta Ramírez-Pérez and Sebastian Hellmann for their help in collecting samples and runing the AC-S instrument during the expeditions. We acknowledge the use of Generic Mapping Tool (GMT) for geographical mapping. Thanks to Junfang Lin for the discussions of optical theories of the pigment package effect. Thanks to Annick Bricaud for providing pigment-specific absorption coefficients data used in Bricaud et al. [73]. Thanks to John Moisan for helping apply SVD-NNLS method to our data set. Thanks to Borut Umer and Karlis Mikelsons for the discussions of SVD related issues. Thanks to Jan Streffing and Nils Haëntjens for helping program using MATLAB. Thanks to Florian Riefstahl for helping with data visualization. Thanks to three anonymous reviewers whose comments significantly improved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following mathematical parameters are used in this manuscript:

Symbol	Description
$a_{gaus}(\lambda)$	Gaussian absorption coefficient (Equation (2))
$a_{NAP}(\lambda)$	spectral absorption coefficient by non-algal particles
$a_p(\lambda)$	spectral particulate absorption coefficient
$a_{ph}(\lambda)$	spectral phytoplankton absorption coefficient
$\hat{a}_{ph}(\lambda)$	pigment package effect normalized $a_{ph}(\lambda)$ (Equation (7))
A_{ph}	matrix of $a_{ph}(\lambda)$ (Equations (4) and (5))
$a^*(\lambda)$	real or measured pigment-specific absorption coefficient
$\widetilde{a}^*(\lambda)$	SVD or NNLS derived pigment-specific absorption coefficient (Equations (4) and (5))
$\widetilde{a}^{*+}(\lambda)$	absolute values of $\widetilde{a}^*(\lambda)$ (Equation (6))
$\|\|\widetilde{a}^*(\lambda)\|\|$	norm of $\widetilde{a}^*(\lambda)$ (Equation (6))
$\widetilde{\mathrm{A}}$	matrix of $\widetilde{a}^*(\lambda)$ (Equations (4) and (5))
A	regression coefficient of pigment concentration-$a_{gaus}(\lambda_0)$ power relationship (Equation (3))
B	regression coefficient (power) of pigment concentration-$a_{gaus}(\lambda_0)$ power relationship (Equation (3))
c	HPLC derived pigment concentration
C	matrix of c (Equation (4))
C^+	Moore–Penrose pseudoinverse of matrix C
$c_{TChl\text{-}a}$	HPLC derived TChl-a concentration
$\widetilde{c}^*$	estimated pigment concentration
$\widetilde{C}$	matrix of $\widetilde{c}^*$ (Equation (5))
CD	quartile coefficient of dispersion
m	number of pigment types
MAE	mean absolute error (Equation (8))
MPE	median absolute percentage error (Equation (9))
n	number of samples
n_{cond}	condition number of matrix C
R^2	determination coefficient
S	spectral exponent of $a_{NAP}(\lambda)$ (Equation (1))
SI	similarity index between two $\widetilde{a}^{*+}(\lambda)$ (Equation (6))
Spearman's ρ	Spearman′s rank correlation coefficient
$Q_a^*(\lambda)$	pigment package effect index
λ_0	peak wavelength of a Gaussian function
σ	width of a Gaussian function
$\sigma_{SD}(\lambda)$	standard deviation of the 20-minute averaged matched AC-S $a_p(\lambda)$ spectra (Equation (2))
χ^2	cost function of Gaussian decomposition (Equation (2))

Appendix A. Cross-Validation Results of NNLS-NNLS

As shown in Table 5 and Table A1, overall, pigment estimation errors from NNLS-NNLS were larger than the corresponding SVD-NNLS for all pigments excepte for TChl-a. The NNLS-NNLS-6 method exhibited robust predictions for TChl-a, Fuco and Hex (MPE 21–65%). With the package effect normalization, three more pigments, i.e., TChl-b, Chl-c1/2 and But were reasonably estimated (55–82%).

Table A1. Statistics of phytoplankton pigment retrieval using NNLS-NNLS based on leave-one-out cross-validation. MAE is in mg m^{-3} (values outside the parentheses were calculated with linear-scale values, while inside the parentheses with log10-scale values) and MPE in %. "Perturb 1, 2 and 3" represent the input data with perturbations of pigment concentrations solely, $a_{ph}(\lambda)$ solely and both, respectively.

(a) NNLS-NNLS-6 ($a_{ph}(\lambda)$ Based) and NNLS-NNLS-9 ($\hat{a}_{ph}(\lambda)$ Based)

Pigments	**Perturb 1 ^{a}**		**Perturb 2 ^{a}**		**Perturb 3 ^{a}**		**Perturb 1 ^{b}**		**Perturb 2 ^{b}**		**Perturb 3 ^{b}**	
	MAE	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**
TChl-a	0.22(0.16)	24.2	0.22(0.17)	21.5	0.22(0.16)	23.6	0.05(0.04)	4.2	0.07(0.05)	6.6	0.08(0.05)	6.6
TChl-b	-	-	-	-	-	-	0.03(0.24)	64.4	0.03(0.24)	55.2	0.03(0.24)	58.2
Chl-c1/2	-	-	-	-	-	-	0.11(0.32)	59.7	0.12(0.34)	66.6	0.12(0.33)	69.6
But	-	-	-	-	-	-	0.03(0.22)	82.2	0.03(0.22)	76.0	0.03(0.23)	79.6
Diadino	0.14(0.38)	267.8	0.45(0.43)	2496	0.15(0.37)	210.4	0.33(0.60)	238.2	0.21(0.48)	161.2	0.23(0.48)	222.3
Fuco	0.10(0.27)	46.5	0.10(0.27)	46.0	0.10(0.27)	46.4	0.13(0.22)	61.8	0.11(0.20)	50.1	0.13(0.21)	60.1
Hex	0.10(0.23)	57.4	0.13(0.24)	61.0	0.14(0.25)	65.3	0.09(0.20)	45.4	0.08(0.21)	42.1	0.09(0.21)	49.1
Peri	0.04(0.25)	185.5	0.04(0.25)	164.7	0.04(0.24)	162.2	0.02(0.13)	67.4	0.02(0.20)	104.0	0.02(0.19)	91.9
Pheo-a	0.01(0.02)	88.5	0.02(0.02)	94.3	0.02(0.02)	97.4	0.07(0.05)	322.9	0.05(0.05)	231.7	0.05(0.05)	234.1

(b) NNLS-NNLS-5′

Pigments	**Perturb 1 ^{a}**		**Perturb 2 ^{a}**		**Perturb 3 ^{a}**	
	MAE	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**
TChl-a	0.21(0.14)	22.0	0.18(0.12)	18.4	0.21(0.07)	21.5
TChl-b	0.04(0.31)	71.9	0.04(0.29)	62.5	0.04(0.27)	64.2
Chl-c1/2	0.06(0.22)	34.1	0.06(0.21)	32.4	0.06(0.30)	34.4
PPC	0.18(0.27)	96.1	0.18(0.28)	85.0	0.17(0.22)	81.5
PSC	0.21(0.20)	46.0	0.20(0.20)	41.0	0.22(0.22)	45.2

(c) Bricaud-NNLS-NNLS-4

Pigments	**Perturb 1 ^{a}**		**Perturb 2 ^{a}**		**Perturb 3 ^{a}**		**Perturb 1 ^{b}**		**Perturb 2 ^{b}**		**Perturb 3 ^{b}**	
	MAE	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**	**MAE**	**MPE**
TChl-a	0.25(0.15)	26.6	0.24(0.15)	26.0	0.25(0.15)	26.9	0.08(0.05)	5.54	0.10(0.06)	8.1	0.10(0.06)	8.1
Chl-c1/2	0.05(0.20)	35.0	0.05(0.20)	35.1	0.06(0.21)	35.7	0.06(0.22)	38.6	0.06(0.23)	41.5	0.06(0.23)	43.1
Diadino	0.14(0.44)	189.3	0.14(0.43)	188.8	0.14(0.43)	186.0	0.10(0.31)	118.1	0.10(0.31)	117.3	0.10(0.32)	117.4
Hex	0.17(0.35)	46.2	0.17(0.35)	47.3	0.17(0.35)	48.7	0.23(0.37)	123.1	0.24(0.37)	114.3	0.24(0.37)	110.8

^{a} $a_{ph}(\lambda)$ based; ^{b} $\hat{a}_{ph}(\lambda)$ based.

References

1. Field, C.B.; Behrenfeld, M.J.; Randerson, J.T.; Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **1998**, *281*, 237–240. [CrossRef] [PubMed]
2. Jeffrey, S.W.; Wright, S.W.; Zapata, M. Microalgal classes and their signature pigments. In *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*; Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., Eds.; Cambridge University Press: Cambridge, UK, 2011; pp. 3–77.
3. Zhao, K.; Porra, R.J.; Scheer, H. Phycobiliproteins. In *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*; Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., Eds.; Cambridge University Press: Cambridge, UK, 2011; pp. 375–411.
4. Müller, P.; Li, X.P.; Niyogi, K.K. Non-photochemical quenching. A response to excess light energy. *Plant Physiol.* **2001**, *125*, 1558–1566. [CrossRef] [PubMed]
5. Brunet, C.; Johnsen, G.; Lavaud, J.; Roy, S. Pigments and photoacclimation processes. In *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*; Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., Eds.; Cambridge University Press: Cambridge, UK, 2011; pp. 445–471.
6. Uitz, J.; Claustre, H.; Griffiths, F.B.; Ras, J.; Garcia, N.; Sandroni, V. A phytoplankton class-specific primary production model applied to the Kerguelen Islands region (Southern Ocean). *Deep Sea Res.* **2009**, *56*, 541–560. [CrossRef]

7. Mackey, M.D.; Mackey, D.J.; Higgins, H.W.; Wright, S.W. CHEMTAX-a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar. Ecol. Progr.* **1996**, *144*, 265–283. [CrossRef]
8. Vidussi, F.; Claustre, H.; Manca, B.B.; Luchetta, A.; Marty, J.C. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *J. Geophys. Res.* **1996**, *106*, 19939–19956. [CrossRef]
9. Bracher, A.; Bouman, H.A.; Brewin, R.J.; Bricaud, A.; Brotas, V.; Ciotti, A.M.; Clementson, L.; Devred, E.; Di Cicco, A.; Dutkiewicz, S.; et al. Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development. *Front. Mar. Sci.* **2017**, *4*, 55. [CrossRef]
10. Sosik, H.M.; Sathyendranath, S.; Uitz, J.; Bouman, H.; Nair, A. *In situ* methods of measuring phytoplankton functional types. In *Phytoplankton Functional Types from Space. Reports of the International Ocean-Colour Coordinating Group (IOCCG), No. 15*; Sathyendranath, S., Ed.; IOCCG: Dartmouth, NS, Canada, 2014; pp. 21–38.
11. Hoepffner, N.; Sathyendranath, S. Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. *J. Geophys. Res.* **1993**, *98*, 22789–22803. [CrossRef]
12. Bidigare, R.R.; Ondrusek, M.E.; Morrow, J.H.; Kiefer, D.A. In-vivo absorption properties of algal pigments. *Proc. SPIE* **1990**, *1302*, 290–303.
13. Bidigare, R.R.; Morrow, J.H.; Kiefer, D.A. Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea. *J. Mar. Res.* **1989**, *47*, 323–341. [CrossRef]
14. Organelli, E.; Bricaud, A.; Antoine, D.; Uitz, J. Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site). *Appl. Opt.* **2013**, *52*, 2257–2273. [CrossRef]
15. Sathyendranath, S.; Stuart, V.; Platt, T.; Bouman, H.; Ulloa, O.; Maass, H. Remote sensing of ocean colour: Towards algorithms for retrieval of pigment composition. *Indian J. Mar. Sci.* **2005**, *34*, 333–340.
16. Pan, X.; Mannino, A.; Russ, M.E.; Hooker, S.B.; Harding, L.W., Jr. Remote sensing of phytoplankton pigment distribution in the United States northeast coast. *Remote Sens. Environ.* **2010**, *114*, 2403–2416. [CrossRef]
17. Pan, X.; Wong, G.T.; Ho, T.Y.; Shiah, F.K.; Liu, H. Remote sensing of picophytoplankton distribution in the northern South China Sea. *Remote Sens. Environ.* **2013**, *128*, 162–175. [CrossRef]
18. Bracher, A.; Taylor, M.; Taylor, B.; Dinter, T.; Röttgers, R.; Steinmetz, F. Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of phytoplankton pigments concentrations. *Ocean Sci.* **2015**, *11*, 139–158. [CrossRef]
19. Bricaud, A.; Mejia, C.; Blondeau-Patissier, D.; Claustre, H.; Crepon, M.; Thiria, S. Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. *Appl. Opt.* **2007**, *46*, 1251–1260. [CrossRef] [PubMed]
20. Chazottes, A.; Crépon, M.; Bricaud, A.; Ras, J.; Thiria, S. Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method. *Appl. Opt.* **2007**, *46*, 3790–3799. [CrossRef] [PubMed]
21. Lohrenz, S.E.; Weidemann, A.D.; Tuel, M. Phytoplankton spectral absorption as influenced by community size structure and pigment composition. *J. Plankton Res.* **2003**, *25*, 35–61. [CrossRef]
22. Chase, A.; Boss, E.; Zaneveld, R.; Bricaud, A.; Claustre, H.; Ras, J.; Dall'Olmo, G.; Westberry, T.K. Decomposition of *in situ* particulate absorption spectra. *Methods Oceanogr.* **2013**, *7*, 110–124. [CrossRef]
23. Chase, A.; Boss, E.; Cetinić, I.; Slade, W. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: Toward a global algorithm. *J. Geophys. Res.* **2017**, *122*, 9725–9743. [CrossRef]
24. Wang, G.; Lee, Z.; Mishra, D.R.; Ma, R. Retrieving absorption coefficients of multiple phytoplankton pigments from hyperspectral remote sensing reflectance measured over cyanobacteria bloom waters. *Limnol. Oceanogr.-Methods* **2016**, *14*, 432–447. [CrossRef]
25. Wang, G.; Lee, Z.; Mouw, C. Multi-spectral remote sensing of phytoplankton pigment absorption properties in cyanobacteria bloom waters: A regional example in the western basin of Lake Erie. *Remote Sens.* **2017**, *9*, 1309. [CrossRef]
26. Bidigare, R.R.; Smith, R.C.; Baker, K.S.; Marra, J. Oceanic primary production estimates from measurements of spectral irradiance and pigment concentrations. *Glob. Biogeochem. Cycles* **1987**, *1*, 171–186. [CrossRef]
27. Moisan, J.R.; Moisan, T.A.; Linkswiler, M.A. An inverse modeling approach to estimating phytoplankton pigment concentrations from phytoplankton absorption spectra. *J. Geophys. Res.* **2011**, *116*. [CrossRef]

28. Moisan, T.A.; Moisan, J.R.; Linkswiler, M.A.; Steinhardt, R.A. Algorithm development for predicting biodiversity based on phytoplankton absorption. *Cont. Shelf Res.* **2013**, *55*, 17–28. [CrossRef]
29. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. Solution of Linear Algebraic Equations. In *Numerical Recipes: The Art of Scientific Computing*; Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 37–109.
30. Lawson, C.L.; Hanson, R.J. Linear least squares with linear inequality constraints. In *Solving Least Squares Problems*; Lawson, C.L., Hanson, R.J., Eds.; Prentice-Hall: Upper Saddle River, NJ, USA, 1974; pp. 158–173.
31. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [CrossRef]
32. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. App. Math.* **1963**, *11*, 431–441. [CrossRef]
33. Zhang, H.; Devred, E.; Fujiwara, A.; Qiu, Z.; Liu, X. Estimation of phytoplankton taxonomic groups in the Arctic Ocean using phytoplankton absorption properties: Implication for ocean-color remote sensing. *Opt. Express* **2018**, *26*, 32280–32301. [CrossRef] [PubMed]
34. Moisan, T.A.; Rufty, K.M.; Moisan, J.R.; Linkswiler, M.A. Satellite observations of phytoplankton functional type spatial distributions, phenology, diversity, and ecotones. *Front. Mar. Sci.* **2017**, *4*, 189. [CrossRef]
35. Devred, E.; Sathyendranath, S.; Stuart, V.; Platt, T. A three component classification of phytoplankton absorption spectra: Application to ocean-color data. *Remote Sens. Environ.* **2011**, *115*, 2255–2266. [CrossRef]
36. Zhang, X.; Huot, Y.; Bricaud, A.; Sosik, H.M. Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter. *Appl. Opt.* **2015**, *54*, 5805–5816. [CrossRef]
37. Butler, W.T.; Hopkins, D.W. Higher derivative analysis of complex absorption spectra. *Photochem. Photobiol.* **1970**, *12*, 439–450. [CrossRef]
38. Soja-Woźniak, M.; Craig, S.E.; Kratzer, S.; Wojtasiewicz, B.; Darecki, M.; Jones, C.T. A novel statistical approach for ocean colour estimation of inherent optical properties and cyanobacteria abundance in optically complex waters. *Remote Sens.* **2017**, *9*, 343. [CrossRef]
39. Boss, E.; Picheral, M.; Leeuw, T.; Chase, A.; Karsenti, E.; Gorsky, G.; Taylor, L.; Slade, W.; Ras, J.; Claustre, H. The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition. *Methods Oceanogr.* **2013**, *7*, 52–62. [CrossRef]
40. Brewin, R.J.; Dall'Olmo, G.; Pardo, S.; van Dongen-Vogels, V.; Boss, E.S. Underway spectrophotometry along the Atlantic Meridional Transect reveals high performance in satellite chlorophyll retrievals. *Remote Sens. Environ.* **2016**, *183*, 82–97. [CrossRef]
41. Dall'Olmo, G.; Boss, E.; Behrenfeld, M.J.; Westberry, T.K. Particulate optical scattering coefficients along an Atlantic Meridional Transect. *Opt. Express* **2012**, *20*, 21532–21551. [CrossRef] [PubMed]
42. Dall'Olmo, G.; Boss, E.; Behrenfeld, M.J.; Westberry, T.K.; Courties, C.; Prieur, L.; Pujo-Pay, M.; Hardman-Mountford, N.; Moutin, T. Inferring phytoplankton carbon and eco-physiological rates from diel cycles of spectral particulate beam-attenuation coefficient. *Biogeosciences* **2011**, *8*, 3423–3439. [CrossRef]
43. Dall'Olmo, G.; Brewin, R.J.; Nencioli, F.; Organelli, E.; Lefering, I.; McKee, D.; Röttgers, R.; Mitchell, C.; Boss, E.; Bricuad, A.; et al. Determination of the absorption coefficient of chromophoric dissolved organic matter from underway spectrophotometry. *Opt. Express* **2017**, *25*, A1079–A1095. [CrossRef] [PubMed]
44. Dall'Olmo, G.; Westberry, T.K.; Behrenfeld, M.J.; Boss, E.; Slade, W.H. Significant contribution of large particles to optical backscattering in the open ocean. *Biogeosciences* **2009**, *6*, 947. [CrossRef]
45. Liu, Y.; Röttgers, R.; Ramírez-Pérez, M.; Dinter, T.; Steinmetz, F.; Nöthig, E.M.; Hellmann, S.; Wiegmann, S.; Bracher, A. Underway spectrophotometry in the Fram Strait (European Arctic Ocean): A highly resolved chlorophyll a data source for complementing satellite ocean color. *Opt. Express* **2018**, *26*, A678–A696. [CrossRef]
46. Slade, W.H.; Boss, E.; Dall'Olmo, G.; Langner, M.R.; Loftin, J.; Behrenfeld, M.J.; Roesler, C. Westberry, T.K. Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *J. Atmos. Ocean. Technol.* **2010**, *27*, 1733–1746. [CrossRef]
47. Westberry, T.K.; Dall'Olmo, G.; Boss, E.; Behrenfeld, M.J.; Moutin, T. Coherence of particulate beam attenuation and backscattering coefficients in diverse open ocean environments. *Opt. Express* **2010**, *18*, 15419–15425. [CrossRef] [PubMed]

48. Werdell, P.J.; Proctor, C.W.; Boss, E.; Leeuw, T.; Ouhssain, M. Underway sampling of marine inherent optical properties on the Tara Oceans expedition as a novel resource for ocean color satellite data product validation. *Methods Oceanogr.* **2013**, *7*, 40–51. [CrossRef]
49. Wadhams, P. Sea ice thickness distribution in Fram Strait. *Nature* **1983**, *305*, 108. [CrossRef]
50. Beszczynska-Möller, A.; Fahrbach, E.; Schauer, U.; Hansen, E. Variability in Atlantic water temperature and transport at the entrance to the Arctic Ocean, 1997–2010. *ICES J. Mar. Sci.* **2012**, *69*, 852–863. [CrossRef]
51. Widell, K.; Østerhus, S.; Gammelsrød, T. Sea ice velocity in the Fram Strait monitored by moored instruments. *Geophys. Res. Lett.* **2003**, *30*. [CrossRef]
52. Smedsrud, L.H.; Sorteberg, A.; Kloster, K. Recent and future changes of the Arctic sea-ice cover. *Geophys. Res. Lett.* **2008**, *30*. [CrossRef]
53. Smedsrud, L.H.; Sirevaag, A.; Kloster, K.; Sorteberg, A.; Sandven, S. Recent wind driven high sea ice area export in the Fram Strait contributes to Arctic sea ice decline. *Cryosphere* **2011**, *5*, 821–829. [CrossRef]
54. Halvorsen, M.H.; Smedsrud, L.H.; Zhang, R.; Kloster, K. Fram Strait spring ice export and September Arctic sea ice. *Cryosphere Discuss.* **2015**, *9*, 4205–4235. [CrossRef]
55. Smedsrud, L.H.; Halvorsen, M.H.; Stroeve, J.C.; Zhang, R.; Kloster, K. Fram Strait sea ice export variability and September Arctic sea ice extent over the last 80 years. *Cryosphere* **2017**, *11*, 65–79. [CrossRef]
56. Nöthig, E.M.; Bracher, A.; Engel, A.; Metfies, K.; Niehoff, B.; Peeken, I.; Bauerfeind, E.; Cherkasheva, A.; Gäbler-Schwarz, S.; Hardge, K.; et al. Summertime plankton ecology in Fram Strait—A compilation of long-and short-term observations. *Polar Res.* **2015**, *34*, 23349. [CrossRef]
57. Cherkasheva, A.; Bracher, A.; Melsheimer, C.; Köberle, C.; Gerdes, R.; Nöthig, E.M.; Bauerfeind, E.; Boetius, A. Influence of the physical environment on polar phytoplankton blooms: A case study in the Fram Strait. *J. Mar. Syst.* **2014**, *132*, 196–207. [CrossRef]
58. Hegseth, E.N.; Sundfjord, A. Intrusion and blooming of Atlantic phytoplankton species in the high Arctic. *J. Mar. Syst.* **2008**, *74*, 108–119. [CrossRef]
59. Bauerfeind, E.; Nöthig, E.M.; Beszczynska, A.; Fahl, K.; Kaleschke, L.; Kreker, K.; Klages, M.; Soltwedel, T.; Lorenzen, C.; Wegner, J. Particle sedimentation patterns in the eastern Fram Strait during 2000–2005: Results from the Arctic long-term observatory HAUSGARTEN. *Deep Sea Res.* **2009**, *56*, 1471–1487. [CrossRef]
60. Downing, A.L.; Leibold, M.A. Ecosystem consequences of species richness and composition in pond food webs. *Nature* **2002**, *416*, 837–841. [CrossRef] [PubMed]
61. Narwani, A.; Mazumder, A. Bottom-up effects of species diversity on the functioning and stability of food webs. *J. Anim. Ecol.* **2012**, *81*, 701–713. [CrossRef] [PubMed]
62. Jakobsson, M.; Mayer, L.; Coakley, B.; Dowdeswell, J.A.; Forbes, S.; Fridman, B.; Hodnesdal, H.; Noormets, R.; Pedersen, R.; Rebesco, M.; et al. The international bathymetric chart of the Arctic Ocean (IBCAO) version 3.0. *Geophys. Res. Lett.* **2012**, *39*. [CrossRef]
63. Sullivan, J.M.; Twardowski, M.S.; Zaneveld, J.R.V.; Moore, C.M.; Barnard, A.H.; Donaghay, P.L.; Rhoades, B. Hyperspectral temperature and salt dependencies of absorption by water and heavy water in the 400–750 nm spectral range. *Appl. Opt.* **2006**, *45*, 5294–5309. [CrossRef]
64. Zaneveld, J.R.V.; Kitchen, J.C.; Moore, C.C. Scattering error correction of reflecting-tube absorption meters. *Proc. SPIE* **1994**, *2258*, 44–55.
65. Hooker, S.B.; Heukelem, L.V. A symbology and vocabulary for an HPLC lexicon. In *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*; Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., Eds.; Cambridge University Press: Cambridge, UK, 2011; pp. 243–256.
66. Yentsch, C.S. Measurement of visible light absorption by particulate matter in the ocean. *Limnol. Oceanogr.* **1962**, *7*, 207–217. [CrossRef]
67. Trüper, H.G.; Yentsch, C.S. Use of glass fiber filters for the rapid preparation of in vivo absorption spectra of photosynthetic bacteria. *J. Bacteriol.* **1967**, *94*, 1255–1256.
68. Yentsch, C.S.; Phinney, D.A. A bridge between ocean optics and microbial ecology. *Limnol. Oceanogr.* **1989**, *34*, 1694–1705. [CrossRef]
69. Simis, S.G.; Tijdens, M.; Hoogveld, H.L.; Gons, H.J. Optical changes associated with cyanobacterial bloom termination by viral lysis. *J. Plankton Res.* **2005**, *27*, 937–949. [CrossRef]
70. Röttgers, R.; Doxaran, D.; Dupouy, C. Quantitative filter technique measurements of spectral light absorption by aquatic particles using a portable integrating cavity absorption meter (QFT-ICAM). *Opt. Express* **2016**, *24*, A1–A20. [CrossRef] [PubMed]

71. Bricaud, A.; Morel, A.; Babin, M.; Allali, K.; Claustre, H. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *J. Geophys. Res.* **1998**, *103*, 31033–31044. [CrossRef]
72. Babin, M.; Stramski, D.; Ferrari, G.M.; Claustre, H.; Bricaud, A.; Obolensky, G.; Hoepffner, N. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *J. Geophys. Res.* **2003**, *108*, 3211–3231, [CrossRef]
73. Bricaud, A.; Claustre, H.; Ras, J.; Oubelkheir, K. Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *J. Geophys. Res.* **2004**, *109*. [CrossRef]
74. Millie, D.F.; Schofield, O.M.; Kirkpatrick, G.J.; Johnsen, G.; Tester, P.A.; Vinyard, B.T. Detection of harmful algal blooms using photopigments and absorption signatures: A case study of the Florida red tide dinoflagellate, Gymnodinium breve. *Limnol. Oceanogr.* **1997**, *42*, 1240–1251. [CrossRef]
75. Xi, H.; Hieronymi, M.; Krasemann, H.; Röttgers, R. Phytoplankton group identification using simulated and in situ hyperspectral remote sensing reflectance. *Front. Mar. Sci.* **2017**, *4*, 272. [CrossRef]
76. Morel, A.; Bricaud, A. Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep Sea Res.* **1981**, *28*, 1375–1393. [CrossRef]
77. Craig, S.E.; Jones, C.T.; Li, W.K.; Lazin, G.; Horne, E.; Caverhill, C.; Cullen, J.J. Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sens. Environ.* **2012**, *119*, 72–83. [CrossRef]
78. Lin, J.; Cao, W.; Zhou, W.; Sun, Z.; Xu, Z.; Wang, G.; Hu, S. Novel method for quantifying the cell size of marine phytoplankton based on optical measurements. *Opt. Express* **2014**, *22*, 10467–10476. [CrossRef] [PubMed]
79. Lawson, C.L.; Hanson, R.J. Perturbation theorems for singular values. In *Solving Least Squares Problems*; Lawson, C.L., Hanson, R.J., Eds.; Prentice-Hall: Upper Saddle River, NJ, USA, 1974; pp. 23–27.
80. Wright, S.W.; Jeffrey, S.W. Pigment markers for phytoplankton production. In *Marine Organic Matter: Biomarkers, Isotopes and DNA. The Handbook of Environmental Chemistry, Volume 2N*; Volkman, J.K., Ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 71–104.

Paper III

Global
Biogeochemical Cycles

RESEARCH ARTICLE
10.1029/2018GB006101

Key Points:

- Phytoplankton long-term photoprotective response is mediated in most parts of the ocean by a variable pool of xanthophyll pigments
- The variability in the pool of xanthophyll pigments is mainly driven by changes in phytoplankton community composition in the tropical and subtropical ocean
- At high latitudes other photoprotection mechanisms play a role, and both taxonomic shifts and photoacclimation determine community pigment signature

Supporting Information:

- Supporting Information S1

Correspondence to:
E. Álvarez,
eva.alvarez@awi.de

Citation:
Álvarez, E., Thoms, S., Bracher, A., Liu, Y., & Völker, C. (2019). Modeling photoprotection at global scale: The relative role of nonphotosynthetic pigments, physiological state, and species composition. *Global Biogeochemical Cycles*, *33*. https://doi.org/10.1029/2018GB006101

Received 6 OCT 2018
Accepted 16 JUN 2019
Accepted article online 22 JUN 2019

Modeling Photoprotection at Global Scale: The Relative Role of Nonphotosynthetic Pigments, Physiological State, and Species Composition

E. Álvarez[1], **S. Thoms**[1], **A. Bracher**[1,2], **Y. Liu**[1,3], **and C. Völker**[1]

[1]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany, [2]Institute of Environmental Physics, University of Bremen, Bremen, Germany, [3]Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

Abstract Microalgae are capable of acclimating to dynamic light environments, as they have developed mechanisms to optimize light harvesting and photosynthetic electron transport. When absorption of light exceeds photosynthetic capacity, various physiological protective mechanisms prevent damage of the photosynthetic apparatus. Xanthophyll pigments provide one of the most important photoprotective mechanisms to dissipate the excess light energy and prevent photoinhibition. In this study, we coupled a mechanistic model for phytoplankton photoinhibition with the global biogeochemical model Regulated Ecosystem Model version 2. The assumption that photoinhibition is small in phytoplankton communities acclimated to ambient light allowed us to predict the photoprotective needs of phytoplankton. When comparing the predicted photoprotective needs to observations of pigment content determined by high-performance liquid chromatography, our results showed that photoprotective response seems to be mediated in most parts of the ocean by a variable ratio of xanthophyll pigments to chlorophyll. The variability in the ratio appeared to be mainly driven by changes in phytoplankton community composition. Exceptions appeared at high latitudes where other energy dissipating mechanisms seem to play a role in photoprotection and both taxonomic changes and physiological acclimation determine community pigment signature. Understanding the variability of community pigment signature is crucial for modeling the coupling of light absorption to carbon fixation in the ocean. Insights about how much of this variability is attributable to changes in community composition may allow us to improve the match between remotely sensed optical data and the underlying phytoplankton community.

1. Introduction

Since phytoplankton organisms are living in a spatially and temporally dynamic light field, their cells have developed a variety of intracellular mechanisms to optimize light harvesting and utilization. Light fluctuations can be harmful to the photosynthetic machinery of microalgae when the harvesting of light energy exceeds photosynthetic carbon fixation capacity (Dubinsky & Schofield, 2010). Thus, to prevent damage of the photosynthetic apparatus caused by rapid light fluctuations, algal cells have evolved various physiological protective mechanisms for stress mitigation. These mechanisms are summarized under the term nonphotochemical quenching (NPQ; Lavaud, 2007). NPQ includes mechanistically distinct processes with likely independent evolutionary origins (Magdaong & Blankenship, 2018), but they share the nonradiative dissipation of excess energy within the photosynthetic apparatus as a common purpose.

Xanthophyll pigments (oxygenated carotenoids) are involved either directly or indirectly in the NPQ of excess light energy in the antenna of photosystem II complexes (PSII). They are ubiquitous in the global ocean (Bricaud et al., 2004; Demmig-Adams & Adams, 1996; Trees et al., 2000). In eukaryotes, one of the main mechanisms that xanthophylls use to perform their photoprotective function is the so-called xanthophyll cycle (XC; Lavaud, 2007). This mechanism involves the light-regulated switching of PSII from a light-harvesting state to an energy dissipating state (Brunet et al., 2011). But also, photosynthetic organisms that do not possess an active XC, such as cyanobacteria, possess xanthophylls with central functions in energy dissipation (Wilson et al., 2006) and locate some carotenoids in the cytoplasmic membrane for protection from high light (Masamoto et al., 1999). The role of xanthophylls is crucial in the modulation of the high light response via the kinetics and amplitude of NPQ that helps to protect the photosynthetic centers against the destructive influence of harmful radiation (Müller et al., 2001).

The amount of xanthophyll pigments relative to total chlorophyll (Chla) is a distinctive feature of different phytoplankton types. It reflects the selective pressure on variable pigment composition in local phytoplankton communities adapted to varying environmental conditions. Accessory pigments have been used as general diagnostic markers for specific phytoplankton groups, and changes in the community pigment ratios can be used to derive changes in taxonomic composition (Mackey et al., 1996). However, the ratios of individual accessory pigments to Chla can also vary as a function of physiological state. The plasticity in the response of phytoplankton to irradiance is species specific (MacIntyre et al., 2002). This plasticity depends on the photoacclimation ability and light history of the phytoplankton cells (Moore et al., 2006) and on the species-specific efficiency of the photoregulative mechanisms (Goss & Jakob, 2010; Lavaud, 2007). Whereas photoregulation involves fast photoprotective reactions that occur on a shorter time scale than pigment synthesis responses, photoacclimation includes longer-term mechanisms of photoresponse that involve changes in pigment content and composition (Demers et al., 1991). Ultimately, photoresponse is determined by the interaction between fast photoregulation and longer-term photoacclimation (Brunet & Lavaud, 2010) in the framework of a given genetic background.

Under acclimation to a prolonged light regime, from hours to seasons, phytoplankton cells change the size of the light harvesting apparatus to saturate dark reactions or to protect PSII. This means typically that the cellular content of photosynthetic pigments tends to increase under low light and to decrease under high light. Photoacclimation to low light includes not only chlorophyll (Chla) but also the accumulation of other photosynthetic accessory pigments, such as fucoxanthin in diatoms or peridinin in dinoflagellates that ensure an efficient utilization of the available light by absorbing photons outside the range of wavelengths accessible to Chla molecules. Photoprotective xanthophyll pigments accumulate in high-light-exposed cells that show higher xanthophyll content relative to Chla (MacIntyre et al., 2002). The ratios of photoprotective pigments to Chla concentrations can be used as quantitative markers for photophysiological state (Stolte et al., 2000). The variations in the pool size of xanthophyll pigments can provide information on the “average” light climate to which the cells have been exposed in the past hours/days.

Xanthophylls participate in the fast photoregulative reactions. Despite differences in their particular mechanisms of action (Fujiki & Taguchi, 2001; Horton et al., 2000; Jahns et al., 2009; Ting & Owens, 1993), it is clear that the function of the XC takes place in time scales of seconds to minutes and allows algae to accommodate to rapid changes in the light field without net change in pigments content. XC is activated when the incident light becomes excessive with respect to the optimum, which is necessary to maximize photosynthesis (Dubinsky & Stambler, 2009). Two different XCs have been described for eukaryotic algae. In the violaxanthin cycle, violaxanthin is reversibly converted into zeaxanthin by fast de-epoxidation using the intermediate antheraxanthin. The diadinoxanthin-diatoxanthin cycle involves the reversible conversion of diadinoxanthin into diatoxanthin. Violaxanthin cycle is active in all land plants, brown algae, and most of the green algae; diadinoxanthin-diatoxanthin cycle is active in a wide range of different algae including diatoms and haptophytes (for reviews see Goss & Jakob, 2010; Jahns et al., 2009). Since the time scale of xanthophyll cycling is much shorter than that of photoacclimation, both epoxidated and de-epoxidated xanthophylls act as photoprotection in the medium to long term and contribute to the photoprotection ability of phytoplankton communities.

Description of photoacclimation via the changes in Chla content is by now virtually standard in ocean biogeochemical models, mostly in terms of the parameterization given by Geider et al. (1998) and sometimes also by Geider et al. (1997). Despite the relevant role of xanthophylls in algae and its possibly important implications, less emphasis has been placed on the capacity of different phytoplankton species or groups to acclimate or regulate photophysiology through photoprotective pigments (Brunet & Lavaud, 2010). There are only a few models that explicitly represent the dynamics of reaction centers and the xanthophyll-mediated photoprotection. Polimene et al. (2012) proposed a model of DD and DT production through conversion of other pigments and de novo synthesis and hence predict the long-term photoprotective response in Bacilliarophyta and Haptophyta. Gustafsson et al. (2014) proposed a model that describes xanthophyll synthesis and cycling in zooxanthella and hence predicts the short-term oxidative stress that leads to coral bleaching. In both cases, the focus on a specific pigment type or mechanism of action permitted the authors to gain insight into particular situations, such as coastal areas dominated by diatoms (Polimene et al., 2014) or corals reefs of the Great Barrier Reef (Baird et al., 2018). Pigment- or mechanism-dependent models are challenging to extrapolate to diverse phytoplankton communities.

Models that do not parameterize short-term photoprotective mechanisms have the advantage that they can be generalized to diverse phytoplankton communities across the global ocean (Han, 2002; Han et al., 2000; Marshall et al., 2000; Ross et al., 2011; Zonneveld, 1998). These models predict the role of NPQ influencing the activity of PSII and consider the role of photoprotective pigments only implicitly. The mechanistic model of photoinhibition proposed by Marshall et al. (2000) accounts for changes in Chla-specific absorption cross-section and quantum yield of photosynthesis driven by the relative amount of active PSII. The decrease in both variables under high light conditions leads to the photoinhibition of the light harvesting apparatus and therefore to the decrease in photosynthetic rate. The fraction of PSII available for photochemistry depends on the protective effect of NPQ involving both closed reaction centers and photoprotective pigments. Rather than assuming any particular short-term mechanism of action for these latter, a variable xanthophyll pool size simulates the resilience of the phytoplankton community to photodamage.

In this study we used the Marshall model as a tool to evaluate the relevance of the different components of NPQ in the global ocean. Avoiding the representation of the full details of the species-specific short-term photoprotective mechanisms allowed us to represent long-term photoprotection in a global diverse ecosystem. We implemented both the phytoplankton growth models of Geider et al. (1998) and Marshall et al. (2000) into the Regulated Ecosystem model version 2 (REcoM2; Hohn, 2009; Schartau et al., 2007). Thereby, we present an ecosystem model that represents phytoplankton diversity with two phytoplankton groups that have group-specific photoprotective needs. Our approach was based on the major assumption that the photoinhibition predicted by the Marshall model should be negligible in phytoplankton communities fully acclimated to ambient light. The difference between the photosynthesis-irradiance curves predicted with and without photoinhibition, that is, by the Marshall and Geider models, respectively, can be considered a measure of the need for photoprotection in order to minimize such photoinhibition. We hypothesize that these photoprotection needs are covered mostly by a variable pool of photoprotective xanthophylls. By comparing the predicted photoprotection needs with global field observations of photoprotective carotenoid content, we evaluated the relevance of nonphotosynthetic xanthophylls for the photoprotection of the phytoplankton community at a global scale. Finally, we explored whether changes in the community aggregated xanthophyll pool were driven by intragroup physiological acclimation and/or by changes in community composition.

2. Model

2.1. Phytoplankton Bio-Optical Model

Within the mechanistic model of photoinhibition by Marshall et al. (2000), the Marshall model, the light harvesting apparatus is divided into two states, PSII with a functional D1 protein and hence active for light harvesting and PSII whose D1 protein has been damaged and hence can no longer participate in photosynthetic electron transfer. The relative amounts of these two types of PSII are represented by the state variables AD1 and DD1, which represent the relative amount of active and damaged PSII, respectively. Definitions of the variables and their units are summarized in Table 1. The model consists of two parts: The first part describes the damage-repair cycle, a set of functions that define the rates of damage to the D1 proteins of the PSII and the rate at which those damaged proteins are repaired; the second part predicts photoinhibition based on the statement that the amount of AD1 at a given time influences the light harvesting ability of the cell and hence the initial slope of the photosynthesis versus light curve ($P^{C}{}_{\text{phot}}-E$). Since the original paper by Marshall et al. (2000) contains some equations that do not balance, we included some modifications to reproduce the model behavior shown in the original publication (equations (1) to (7) in Table 2). For a detailed analysis of the modifications made to Marshall et al. (2000), see the supporting information (Flynn et al., 1999; Flynn & Flynn, 1998).

2.1.1. Damage and Repair of D1 Proteins

The rate of damage to AD1 (Gd) is modeled as a linear function of photon dose (equation 1). The slope of the relationship represents the target size for photodamage to PSII (square meter per Joule). The two NPQ mechanisms that prevent damage to AD1 are (i) the quenching provided by already damaged reaction centers (qRC) that protects the cell from further damage via multiplication with the term $(1 - \text{DD1})$ and (ii) the antenna-based NPQ (Q_e) that decreases Gd via multiplication with the term $(1 - Q_e)$. Whereas qRC is proportional to DD1 that is a state variable in the model, Q_e is not related to any explicit representation of

Table 1
Definitions of State, Intermediate, and Input Variables in the Phytoplankton Growth model

Variable	Definition	Units
State variables		
C	Carbon	mmolC m^{-3}
N	Nitrogen	mmolN m^{-3}
Si	Silica	mmolSi m^{-3}
Chla	Photosynthetic pigments	mgChla m^{-3}
AD1	Functioning D1 relative to total D1	Dimensionless
DD1	Damaged D1 relative to total D1	Dimensionless
Intermediate variables		
Gd	Damage rate	d^{-1}
Rep	Repair rate	d^{-1}
ϕ	Quantum yield of photosynthesis	mmolC J^{-1}
Q_e	Antenna-based non photochemical quenching	Dimensionless
a^*_{NP}	Cellular absorption cross section	m^2 $mgChla^{-1}$
α_{NP}	Initial slope of the photosynthesis light curve of damaged cells	m^2 molC $gChla^{-1}$ J^{-1}
Q	Nitrogen to carbon quota	mol mol^{-1}
Q_{Si}	Silica to carbon quota	mol mol^{-1}
Θ	Chla to carbon quota	g mol^{-1}
$LimQ_{min}$	Limitation term by approach to Q_{min}	Dimensionless
$LimQ^{Si}{}_{min}$	Limitation term by approach to $Q^{Si}{}_{min}$	Dimensionless
Nlimit	Nutrient growth-limitation term	Dimensionless
$LimQ_{max}$	limitation term by approach to Q_{max}	Dimensionless
$LimQ^{Si}{}_{max}$	Limitation term by approach to $Q^{Si}{}_{max}$	Dimensionless
Tfunc	Arrhenius function	Dimensionless
$P^C{}_{phot}$	Rate of photosynthesis	d^{-1}
$P^C{}_{max}$	Maximum rate of photosynthesis	d^{-1}
R	Phytoplankton respiration	d^{-1}
RChl	Loss rate of Chla	d^{-1}
$V^C{}_N$	Nitrogen uptake	molN $molC^{-1}$ d^{-1}
$V^N{}_{max}$	Maximum nitrogen uptake	molN $molC^{-1}$ d^{-1}
$V^C{}_{Si}$	Silica uptake	molSi $molC^{-1}$ d^{-1}
$V^{Si}{}_{max}$	Maximum silica uptake	molSi $molC^{-1}$ d^{-1}
Input variables		
T	Temperature	°K
E	Irradiance	J $m^{-2}d^{-1}$
Ni	Dissolved inorganic nitrogen (DIN)	mM
dSi	Dissolved silica	mM
Fe	Dissolved iron	µM

nonphotosynthetic pigments. In the original formulation Q_e is modified by a correction factor (*Ds*) that is intended to reflect the differential resilience of different species to photodamage via xanthophyll-based NPQ. By setting the parameter *Ds* to a constant value of 1, the Marshall model assumes a constant xanthophyll pigment pool size.

The repair rate of D1 proteins (*Rep*) is modeled as a function of DD1, the relative amount of damaged PSII that are ready to be repaired. Repair of DD1 consists on several steps. The apparent bottleneck of the whole process is the enzymatic removal of damaged proteins (Mellis, 1999). Hence, *Rep* must reach a maximum and has the form of a Michaelis-Menten function where *Y* is the maximum repair rate and *Z* the half saturation constant (equation (2)). As repair requires de novo synthesis of D1 proteins, *Rep* is limited by nutrient availability (*Nlimit*) in the original formulation (Marshall et al., 2000). We included a temperature dependency (*Tfunc*) given the temperature restrictions on D1-protein turnover (Ni et al., 2017). The change in AD1 is set to the difference between the damage and repair rates to D1 proteins (equation (3)) and the amount of DD1 is equal to 1−AD1 (equation (4)).

2.1.2. Effect of AD1 and NPQ on α

The photochemical efficiency of PSII (ϕ), defined as carbon fixed per unit of light absorbed, depends on the fraction of AD1 proteins available to photochemistry. The closure of reaction centers leads to a decrease in the value of ϕ. However, a loss of active photosystems of up to 25% has been found to have no impact on ϕ

Table 2
Phytoplankton Growth Model Equations

Eq.	Equations	Source
1	$Gd = (X \times E) \times AD1 \times (1 - Q_e)$	Modified from (Marshall et al., 2000)
2	$Rep = \left(\frac{Y \times DD1}{Z + DD1}\right) \times Nlimit \times Tfunc$	(Marshall et al., 2000)
3	$\frac{dAD1}{dt} = DD1 \times Rep - AD1 \times Gd$	Modified from (Marshall et al., 2000)
4	$\frac{dDD1}{dt} = AD1 \times Gd - DD1 \times Rep$	Modified from (Marshall et al., 2000)
5	$\phi = \min\left(\frac{\phi_{max}}{F} \times AD1, \phi_{max}\right)$	Modified from (Marshall et al., 2000)
6	$Q_e = 1 - {}^{\phi}/_{\phi_{max}}$	Modified from (Marshall et al., 2000)
7	$\alpha_{NP} = a^*_{NP} \times \phi \; ; \; a^*_{NP} = a^*_{PH} \times (1 - Q_e)$	(Marshall et al., 2000)
8	$LimQ_{min} = 1 - e^{-50 \times (abs(Q_{min} - Q) - (Q_{min} - Q))^2}$	(Hohn, 2009)
9	$LimQ^{Si}_{min} = 1 - e^{-1000 \times \left(abs\left(Q^{Si}_{min} - Q_{Si}\right) - \left(Q^{Si}_{min} - Q_{Si}\right)\right)^2}$	(Hohn, 2009)
10	$Nlimit_d = \min\left(LimQ_{min}, LimQ^{Si}_{min}, \frac{Fe}{Fe + k_{Fe}}\right); \; Nlimit_{nd} = \min\left(LimQ_{min}, \frac{Fe}{Fe + k_{Fe}}\right)$	Liebig's law
11	$LimQ_{max} = 1 - e^{-1000 \times (abs(Q - Q_{max}) - (Q - Q_{max}))^2}$	(Schartau et al., 2007)
12	$LimQ^{Si}_{max} = 1 - e^{-1000 \times \left(abs\left(Q_{Si} - Q^{Si}_{max}\right) - \left(Q_{Si} - Q^{Si}_{max}\right)\right)^2}$	(Hohn, 2009)
13	$Tfunc = e^{-Ae \times \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)}$	(Geider et al., 1998)
14	$P^C_{phot} = P^C_{max} \times \left(1 - e^{-\alpha\theta E / {}_{P}max^{C}}\right) \; ; \; P^C_{max} = P^C_{ref} \times Nlimit \times Tfunc$	(Webb et al., 1974)
15	$R = R_{ref} \times LimQ_{max} + \zeta \times V^C_N + \vartheta \times V^C_{Si}$	(Geider et al., 1998)
16	$\frac{dC}{dt} = C \times \left(P^C_{phot} - R - \eta_C \times LimQ_{max}\right)$	(Geider et al., 1998)
17	$\frac{dChla}{dt} = Chla \times \left(\frac{V^C_N \times Q^{Chl}_N}{\theta} \times \frac{P^C_{phot}}{\alpha \times \theta \times E} - RChl\right)$	(Geider et al., 1998)
18	$RChl = k \times \left(1 - e^{-\alpha\theta E / P^C_{max}}\right)$	(Álvarez et al., 2018)
19	$V^C_N = V^N_{max} \times \frac{Ni}{(Ni + k_N)} \; ; \; V^N_{max} = V^C_{ref} \times P^C_{max} \times Q_{max} \times LimQ_{max}$	(Geider et al., 1998)
20	$\frac{dN}{dt} = C \times V^C_N - N \times \eta_N \times LimQ_{max}$	(Geider et al., 1998)
21	$V^C_{Si} = V^{Si}_{max} \times \frac{dSi}{(dSi + k_{Si})} \; ; \; V^{Si}_{max} = V^C_{ref} \times P^C_{ref} \times Tfunc \times Q^{Si}_{max} \times LimQ_{max} \times LimQ^{Si}_{max}$	(Hohn, 2009)
22	$\frac{dSi}{dt} = C \times V^C_{Si} - Si \times \eta_{NSi} \times LimQ_{max}$	(Hohn, 2009)

Note. Details on the modifications made to Marshall et al. (2000) model in the Supporting Information (Flynn et al., 1999; Flynn & Flynn, 1998).

(Park et al., 1995). So for AD1 larger than 0.75, ϕ equals ϕ_{max}, and for AD1 smaller than 0.75, ϕ decreases linearly with D1 (equation (5)). The slope is set to $\phi_{max}/0.75$, so ϕ varies from ϕ_{max} to 0.

The buildup of a proton gradient across the thylakoid triggers the de-epoxidation of xanthophylls (Goss & Jakob, 2010) and hence an increase in Q_e. Since pH is not explicitly taken into account in the Marshall model, the latter assumes that ϕ is correlated with pH and uses ϕ as a proxy for Q_e (equation (6)). The slope $1/\phi_{\ max}$ makes Q_e a relative quantity that varies between 0 and 1.

Q_e has two roles in the model. It decreases the damage rate to D1 proteins (equation (1)) and decreases the Chla-specific optical absorption cross-section of photosynthetic pigments (a^*_{NP}) (equation (7)). Although

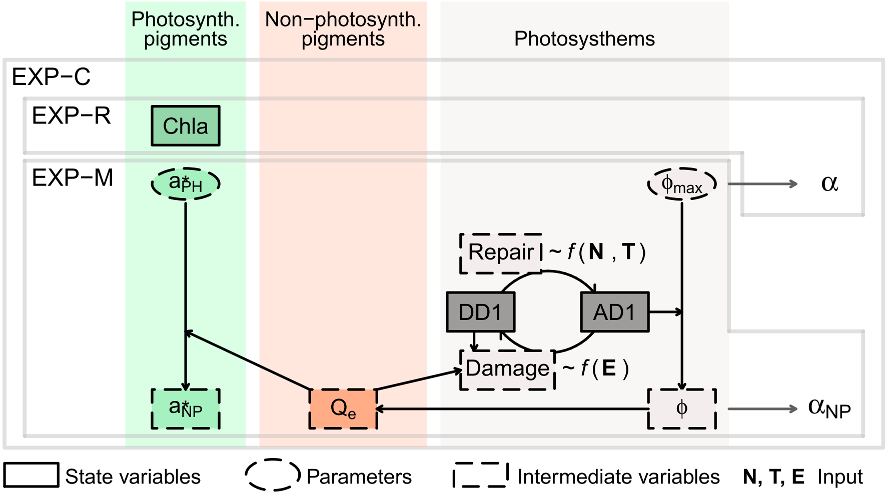


Figure 1. Schematic showing the parts of the model used in each experiment. The model is divided into photosynthetic and nonphotosynthetic pigments and reaction center dynamics. State, intermediate, input variables, and parameters are distinguished.

the latter is the simplest description of the effects of Q_e on $a*_{NP}$, it is sufficient to represent the decrease in $a*_{NP}$ as observed from saturating to inhibitory light levels (Kolber et al., 1988).

The decrease in the absorption cross-section acts alongside ϕ decrease to decrease the initial slope of the $P^C{}_{\text{phot}}-E$ curve (α_{NP}). α_{NP} is generally described as the light limited slope of the $P^C{}_{\text{phot}}-E$ curve since under light saturated conditions the photosynthesis is limited by dark reactions. However, the net result of the Marshall model is making α_{NP} variable along the curve. The decrease in α_{NP} under high light conditions reverts the limit for photosynthesis from dark to light reactions, which eventually leads to photoinhibition.

2.1.3. Combining the Models by Marshall and Geider et al. (1998)

The dynamic phytoplankton growth model by Geider et al. (1998) sets the initial slope of the $P^C{}_{\text{phot}}-E$ curve (α) as a constant parameter. With α being constant along the whole $P^C{}_{\text{phot}}-E$ curve and with no other photoinhibitory parameter being considered, photoinhibition does not take place. The differences between Geider and Marshall approximations are summarized in Figure 1. If we assume that photoinhibition is negligible in phytoplankton communities acclimated to ambient light (Cullen et al., 1992), the difference between the $P^C{}_{\text{phot}}-E$ curves with (Marshall) and without photoinhibition (Geider) reflects the need for photoprotection in order to limit such photoinhibition. The difference between Geider and Marshall $P^C{}_{\text{phot}}-E$ curves is given by the difference between α and α_{NP}. Hence, the difference between α and α_{NP} reflects a photoprotection gap that could be filled by the missing elements in the description of NPQ, such a variable pool of nonphotosynthetic pigments but also by any other mechanism not described explicitly such as constitutive heat dissipation. By comparing the predicted photoprotection gap to observations of photoprotective carotenoids under natural conditions, we tested whether the gap is filled by the presence of nonphotosynthetic pigments, or if further mechanisms are necessary to complete photoprotection.

We performed three experiments with different settings for our model:

EXP-R equivalent to the dynamic phytoplankton model by Geider, only α was computed and used in the $P^C{}_{\text{phot}}-E$ curve for production.
EXP-M equivalent to the original Marshall model, only α_{NP} was computed and used in the $P^C{}_{\text{phot}}-E$ curve for production.
EXP-C both α and α_{NP} were computed, α was used in the $P^C{}_{\text{phot}}-E$ curve for production, and the difference between α and α_{NP} was explored to evaluate photoprotection requirements.

2.2. Implementation Into REcoM2

The variability in pigment composition in a diverse phytoplankton community is not only dependent on the direct effect of environmental conditions on light harvesting traits for a particular phytoplankton group.

Table 3
Definitions of Parameters With Values for the Two Phytoplankton Groups

Parameter	Definition	Value diatoms	Value nondiatoms	Units
α	Initial slope of the photosynthesis irradiance curve	0.19	0.14	m^2 molC $gChla^{-1}$ J^{-1}
φ_{max}	Maximum quantum yield	$3.1 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	mmolC J^{-1}
a^*_{PH}	Chlorophyll absorption cross section	0.007	0.007	m^2 $mgChla^{-1}$
F	Minimum AD1 to keep $\varphi=\varphi_{max}$	0.75	0.75	relative (0-1)
X	Target size for photoinactivation	$1.5 \cdot 10^{-7}$	$7.5 \cdot 10^{-7}$	m^2 J^{-1}
Y	Maximum repair rate	10	10	d^{-1}
Z	Half saturation constant repair	0.3	0.3	relative (0-1)
$P^C{}_{ref}$	Maximum rate of photosynthesis	3.5	3	d^{-1}
R_{ref}	Maintenance respiration rate	0.01	0.01	d^{-1}
$V^C{}_{ref}$	Maximum nitrogen uptake	0.7	0.7	molN $molC^{-1}$ d^{-1}
k_N	Half-saturation constant nitrate uptake	1	0.55	mmolN m^{-3}
k_{Si}	Half-saturation constant silica uptake	4	—	mmolSi m^{-3}
k_{Fe}	Half-saturation constant iron uptake	0.12	0.02	µmolFe m^{-3}
Q_{min}	Minimum cell quota of nitrogen	0.04	0.04	molN $molC^{-1}$
Q_{max}	Maximum cell quota of nitrogen	0.2	0.2	molN $molC^{-1}$
$Q^{Si}{}_{min}$	Minimum cell quota of silica	0.04	—	molSi $molC^{-1}$
$Q^{Si}{}_{max}$	Maximum cell quota of silica	0.2	—	molSi $molC^{-1}$
$Q^{Chl}{}_N$	Maximum Chla to nitrogen ratio	4.2	3.78	gChl $molN^{-1}$
$Q^{SI}{}_N$	Minimum silica to nitrogen ratio	0.3	—	molSi $molN^{-1}$
ζ	Cost of N biosynthesis	2.33	2.33	molC $molN^{-1}$
?	Cost of Si biosynthesis	0	—	molC $molSi^{-1}$
k	Maximum loss rate of Chla	0.25	0.15	d^{-1}
η_C	Phytoplankton loss of C	0.1	0.1	d^{-1}
η_N	Phytoplankton loss of N	0.05	0.05	d^{-1}
η_{Si}	Phytoplankton loss of Si	0.05	—	d^{-1}
T_{ref}	Reference temperature	288.15		°K
Ae	Linear slope Arrhenius function	4500		°K
κ_W	Total light attenuation due to water	0.04		m^{-1}
a_{CHL}	Chla-specific attenuation coefficient	0.03		m^2 mg $Chla^{-1}$

Community composition influences the community aggregated pigment signatures. We chose the RecoM2, which provides the simplest framework to include phytoplankton diversity since it describes the dynamics of two phytoplankton groups with a detailed description of their elemental composition. The rest of the ecosystem is completed with zooplankton, detritus, and main nutrients compartments. Ocean circulation and mixing is derived from the MIT general circulation model. Details about the complete model setup can be consulted in the appendix to Hauck et al. (2013).

2.2.1. Phytoplankton Diversity

RecoM2 describes the dynamics of two phytoplankton types, diatoms and nondiatoms. The difference between groups is merely functional, as diatoms require silica and other phytoplankton do not. However, differences in parameter values between the two groups (Table 3) are size related, and the diatom group can be ascribable to large-sized phytoplankton while other phytoplankton represents smaller phytoplankton. We kept α values larger for diatoms compared to the values for small phytoplankton. This have provided realistic distribution of primary production in previous applications of REcoM2 (Álvarez et al., 2018; Schourup-Kristensen et al., 2014). With all PSII being active for photochemistry, α_{NP} should equal α. Since the Marshall model splits α_{NP} into absorption (a^*_{NP}) and photochemical (ϕ) components, we set the maximum values for those, a^*_{PH} and ϕ_{max}, in order to match the group specific α's (Table 3).

Literature values given for a^*_{PH} range from 0.005 to 0.025 $m^2 \cdot mg \cdot Chla^{-1}$ (Kromkamp et al., 2001; Megard et al., 1979; Oliver & Ganf, 1988) and given the package effect on pigment concentrations, larger cells tend to have smaller values of a^*_{PH} (Bricaud et al., 2004). However, we kept a^*_{PH} equal for the two groups in 0.007 $m^2 \cdot mg \cdot Chla^{-1}$, and hence, we have not considered a packaging effect. The light attenuation by phytoplankton (a_{CHL}) was also set constantly to 0.03 $m^2 \cdot mg \cdot Chla^{-1}$ (Table 3). Experimental values given for ϕ_{max} range from $2.1 \cdot 10^{-5}$ to $4.8 \cdot 10^{-4}$ mmolC/J (Du et al., 2018; Kiefer & Mitchell, 1983; MacIntyre et al., 2002; Raven &

Crawfurd, 2012). In the field, higher $\phi_{\max}$ have been documented in communities dominated by diatoms (Babin et al., 1996). So we gave $\phi_{\max}$ of $3.1 \cdot 10^{-4}$ mmolC/J to diatoms (which generated an α of 0.19; Table 3) and $2.3 \cdot 10^{-4}$ mmolC/J to other phytoplankton (which generated an α of 0.14; Table 3). There were also group differences regarding the damage-repair cycle. Picoplankton are generally reported to have damage coefficients that are higher (Nagy et al., 1995; Six et al., 2007) than for green algae (Oliver et al., 2003; Serôdio et al., 2017), but not very different from the values in diatoms (Campbell & Tyystjärvi, 2012; Lavaud et al., 2016). In this work we assigned smaller damage coefficients to diatoms than to other phytoplankton and kept equal repair parameters (Table 3).

2.2.2. Phytoplankton Growth Model

The remaining part of the phytoplankton growth model was the same for the three experiments described above and included the dynamics of the elemental pools of carbon (C), photosynthetic pigments (Chla), nitrogen (N), and silica (Si; equations (8) to (22) in Table 2). The nutrient limitation terms in REcoM2 include $LimQ_{\min}$ and $LimQ^{Si}{}_{\min}$ that limit processes dependent on the content of proteins for enzymatic reactions and thus decrease when protein content approaches the minimum cellular quota (equations (8) and (9)). *Nlimit* (equation (10)) is a combined nutrient-limitation term that computes the minimum of $LimQ_{\min}$, $LimQ^{Si}{}_{\min}$, and a limitation term for Fe in Michaelis-Menten form for diatoms ($Nlimit_d$) and only the minimum of $LimQ_{\min}$ and the term for Fe for other phytoplankton ($Nlimit_{\mathrm{nd}}$). $LimQ_{\max}$ and $LimQ^{Si}{}_{\max}$ limit processes that saturate when protein content approaches the maximum cellular quota (equations (11) and (12)). The temperature dependency is an Arrhenius function (Geider et al., 1998; equation (13)). All limitation terms multiply the rates they regulate, and hence, they are 1 for no limitation and approach 0 as limitation increases.

The $P^C{}_{\mathrm{phot}}-E$ curve is used as the exponential formulation in Geider et al. (1998) where maximum photosynthesis rate ($P^C{}_{\max}$) is limited by *Nlimit* and *Tfunc* (equation (14)). Phytoplankton respiration includes maintenance respiration and the cost of biosynthesis (equation (15)). The variation in C content is set to the rate of photosynthesis minus respiration and excretion (equation (16)).

The synthesis of photosynthetic pigments (represented by Chla) is equivalent to N assimilation regulated by the photochemical use of absorbed light, a term that allows photoacclimation (Geider et al., 1998; equation (17)). The loss of Chla is light dependent to account for photodamage in the light harvesting apparatus (Álvarez et al., 2018; equation (18)).

Although N and Si pools are not central in this work, they both shape the stoichiometric ratios that limit production and biosynthetic processes. N uptake ($V^C{}_N$) depends on available DIN, the half saturation constant k_N, and the maximal uptake rate ($V^C{}_{\max}$) that is proportional to $P^C{}_{\max}$ and declines when Q approaches $Q_{\max}$ (equation (19)). The rate of N assimilation is set to uptake minus excretion rates, and the latter is also limited by $LimQ_{\max}$ (equation (20)). Si uptake ($V^C{}_{Si}$) (equation (21)) and assimilation (equation (22)) are formulated in an equivalent way to those of N (Hohn, 2009).

2.2.3. The Rest of the Ecosystem and the Global Circulation Model

REcoM2 completes the ecological module with one zooplankton and one detritus compartment, and inorganic and organic forms of the main nutrients. Temperature (T) and dissolved nutrients (DIN, dSi, and Fe), like all other biogeochemical model variables, are advected and mixed by the ocean circulation derived from the MIT general circulation model. Average light (E) is computed in depth layers as an exponential decreasing function of depth with a depth-dependent light attenuation coefficient with two components, the attenuation coefficient due to water (κ_W) and the attenuation due to phytoplankton, proportional to total Chla concentration ($a_{CHL} \times TChla$) (Table 3).

For our study, in each experiment REcoM2 was run in a nearly global model configuration from 80°S to 80°N on a horizontal 2° × 2° grid in the Northern Hemisphere and 2° × 2° times the cosine of the latitude in the Southern Hemisphere, with 30 depth layers (0 to 5,700 m). The model was initialized with the January climatological fields of temperature, salinity, nitrate, and silicate from the World Ocean Atlas 2009 (Antonov et al., 2010; Garcia et al., 2010; Locarnini et al., 2010) and with mean alkalinity and preindustrial CO_2 fields from the Global Ocean Data Analysis Project (GLODAP; Key et al., 2004). The initial field for dissolved Fe was obtained from PISCES output (Aumont et al., 2003), with values south of 45°S set to average Southern Ocean vertical profiles from Tagliabue et al. (2012), to avoid a high-iron bias there. The model was spun up for 4 years and analyzed for the next fifth year in a 10-daily temporal resolution.

Model output provided physical variables, *T* and *E*, and biological variables by phytoplankton group, *Nlimit*, Chla, C, α_{NP} and α. The group-specific values were averaged considering the relative contribution of each group to total biomass to obtain the values for the whole phytoplankton community that we indicated with an overbar (see the supporting information for details on how we estimated community values for $\bar{\alpha}$ and $\bar{\alpha}_{NP}$ from several group-specific α's; Violle et al., 2007). Each output variable, including biological and physical, was averaged to a global 2° × 2° grid, within the 33 depth layers (0 to 5,750 m) used in Peloquin et al. (2013) and over 12 months. This resulted in a 4-D array per variable that had a common spatial and temporal resolution to be compared to observations (180 longitude × 90 latitude × 33 depth × 12 time).

Table 4
List of Data Sets With HPLC Measurements

Data set	Source
MAREDAT pigments	Peloquin et al. (2013)
ANT23.1 (PS69)	Bracher et al. (2015a, 2015b)
ANT24.1 (PS71.1)	Bracher (2015b, 2015e); Bracher et al. (2015b)
ANT24.4 (PS71.4)	Bracher (2015b, 2015c); Bracher et al. (2015b)
ANT25.1 (PS73)	Taylor et al. (2011a, 2011b)
ANT26.4 (PS75)	Bracher (2015c, 2015d); Bracher et al. (2015b)
ANT27.2 (PS77)	Bracher (2015d, 2015e); Trimborn et al. (2015)
ANT28.3 (PS79)	Bracher (2014a); Soppa et al. (2014)
MSM18	Bracher (2015a); Bracher et al. (2015b)
MSM09	Bracher and Taylor (2017)
SO202	Taylor and Bracher (2017); Zindler et al. (2013)
SO218	Bracher (2014b); Soppa et al. (2014)
SO234/235	Bracher et al. (2019); Booge et al. (2018)
SO243 (ASTRA)	Bracher (2019a); Booge et al. (2018)
PS93.2	Liu et al. (2018, 2019a)
PS99.2	Liu et al. (2018, 2019a)
PS107	Liu et al. (2019a, 2019b)
PS103	Bracher (2019b)
HE462	Bracher and Wiegmann (2019)
CLIVAR	*AEsOP*-CSIRO
BROKEWEST	*AEsOP*-CSIRO
Beagle (6 legs)	*AEsOP*-CSIRO
SOOP	*AEsOP*-CSIRO
SS (x6)	*AEsOP*-CSIRO
FR200001	*AEsOP*-CSIRO
TIP2000	*AEsOP*-CSIRO
NWS-jun03	*AEsOP*-CSIRO
Sniper (x4)	*AEsOP*-CSIRO
GBR (x5)	*AEsOP*-CSIRO

3. Data

3.1. Satellite Observations

To test the consistency between the modeled and observed phytoplankton biomass, we considered climatologies of Chla and Chla:C from satellite observations between 2006 and 2010. Monthly means of surface Chla were obtained from the Ocean Colour Climate Change Initiative data set v3.0 by the European Space Agency (http://www.esa-oceancolour-cci.org/). We considered Chla concentration generated by SeaDAS using a blended combination of OCI (OC4v6 + Hu's CI), OC3, and OC5 depending on water class memberships (4,320 by 8,640 pixels, monthly means). Monthly means of surface Chla:C data were obtained from the Ocean Productivity Dataset by Oregon State University (http://www.science.oregonstate.edu/ocean.productivity/index.php). We considered Chla:C from the carbon-based productivity model (updated CbPM; Westberry et al., 2008) from MODIS r2018 (GSM) data (1,080 by 2,160 pixels, monthly means).

3.2. Accessory Pigments

Pigment data were obtained from published high-performance liquid chromatography data sets (Booge et al., 2018; Bracher et al., 2015b; Liu et al., 2018, 2019b; Peloquin et al., 2013; Soppa et al., 2014; Taylor et al., 2011a; Trimborn et al., 2015; Zindler et al., 2013), from pigment data in the *AEsOP*-CSIRO database (http://aesop.csiro.au/) and from two not previously published HPLC data sets (Table 4). These new data sets (Bracher, 2019b; Bracher & Wiegmann, 2019) encompassed the following cruises: the R/V Heincke cruise HE462 in the North Sea from 30 April to 7 May 2016 and the R/V Polarstern cruise PS103 in the South Atlantic from 17 December 2016 to 28 January 2017. During the cruises, 0.25 to 2.5 L of seawater was filtered through Whatman GF/F filters. The sample filters were then shock-frozen in liquid N_2 and kept at −80 °C until analysis. HPLC pigment analysis was performed following the method of Barlow et al. (1997) that was adjusted to our temperature-controlled instruments (Liu et al., 2019a; Taylor et al., 2011a). We determined the concentrations of pigments listed in Table 2 of Taylor et al. (2011a).

In all cases, TChla (micrograms per liter) encompassed all the reported Chla derivatives, monovinyl Chla, divinyl Chla, and chlorophyllide a. Total accessory pigment concentration (AP; micrograms per liter) was calculated as the summed concentration of all carotenoids and chlorophyll b and c. Carotenoids were grouped into photosynthetic (PSC; micrograms per liter) and photoprotective (PPC; micrograms per liter) carotenoids according to Aiken et al. (2009). PSC consisted of fucoxanthin (Fuco), peridinin (Perid), prasinoxanthin (Pras), 19'-hexanoyloxyfucoxanthin (Hex), and 19'-butanoyloxyfucoxanthin (But). As we focused on the long-term photoprotective response, both epoxidated and de-epoxidated states of xanthophylls were considered to be photoprotective and hence PPC consisted of alloxanthin (Allo), lutein (Lut), violaxanthin (Viola), zeaxanthin (Zea), diadinoxanthin (DD), diatoxanthin (DT), and alpha/beta-carotenes (Caro).

Within the compilation of pigment observations, the MAREDAT data set (Peloquin et al., 2013) was the most extensive. It was quality controlled by flagging (i) samples in which TChla was zero or less; (ii) samples in which fewer than four nonzero accessory pigments were reported; (iii) samples that fell outside the range of two standard deviations of the regression line of the log linear relationship between TChla and AP; and (iv) the entire campaign's samples if more than 35 % of samples from a given field campaign was flagged during the third step. However, not all cruises in MAREDAT provided concentrations of the full set of PPC pigments. In this case, we limited the analysis to the samples that contained measurements for the seven PPC pigments. This reduced the MAREDAT data set to 8,574 samples, that is, 25% of the quality-controlled data.

All the other data sets formed the NEW database in which all cruises provided the seven PPC. This NEW database was quality controlled independently following Aiken et al. (2009) by flagging (i) samples in which TChla was zero or less; (ii) samples where the difference of TChla and AP was more than 30% of the total pigment concentration; and (iii) the entire campaign's samples if the regression between TChla and AP have a slope outside the range 0.7–1.4, explain less than 90% of total variance or less than 85% of the samples of that particular cruise passed the previous criteria. This reduced the data set to 5,831 samples, that is, 89% of the original merged data.

All field data were gridded to the same 4-D array as described for model output. We obtained 1,985 grid points (404 in surface) in MAREDAT, and we used this set to derive a purely empirical parameterization of the relative PPC content from the physical forcing, T, E, and *Nlimit*, in the three modeling experiments. With NEW we obtained 2,086 grid points (789 in surface) that we used to validate the empirical predictions, as they were independent from MAREDAT. The combination of the two data sets (ALL) that encompassed 3,982 grid points (1,122 in surface) was used to test the mechanistic predictions in the experiment EXP-C.

To estimate the contribution of diatoms to TChla, we calculated the fraction of diatoms in Chla (fdChla) as 1.41xFuco/diagnostic pigment (DP) with D1.41Fuco + 1.41Perid + 1.27Hex + 0.6Allo + 0.35But + 1.01Chlb + 0.86Zea (Uitz et al., 2006). As an estimate of the contribution of the diatoms to the total pool of PPC, we considered the sum of DD and DT (micrograms per liter) as the photoprotective carotenoids in diatoms (Aiken et al., 2009; Strain et al., 1944). All other PPCs were considered to belong to nondiatoms.

4. Results

The analysis of results comprised three stages: (i) the prediction of PPC/TChla from model output, both empirically (from E, T, and *Nlimit*) and mechanistically (from $\bar{\alpha}$ and $\bar{\alpha}_{NP}$), (ii) the exploration of the spatial and temporal variability of PPC/TChla and the match between predictions and observations, and (iii) the analysis on the relative contribution of changes in community composition and physiological acclimation to the predicted PPC/TChla variability.

4.1. Prediction of PPC/TChla: Empirical and Mechanistic Approaches

We proposed a mechanistic approach of exploring photoprotection by comparing $\bar{\alpha}$ and $\bar{\alpha}_{NP}$. One can wonder if a simple empirical prediction of PPC/TChla from the model input variables was comparable to the mechanistic prediction. The empirical approach relied on the direct prediction of pigments from the input variables to the phytoplankton growth model, temperature, light, and nutrient availability, and hence, it can be fitted to all the three experiments. EXP-C was, however, the only experiment where the two types of predictions can be performed simultaneously. Hence, we used it to test the ability of the empirical and mechanistic approaches to match observations.

4.1.1. Empirical Prediction

The in situ pigment content in the MAREDAT data set was plotted against the physical input to the phytoplankton growth model (Figure 2). Individual pigment ratios relative to TChla were compared against a general rule of photoacclimation: Decreasing ratios with light are typical for light-harvesting pigments, and increasing ratios with light characterize photoprotective pigments (Henriksen et al., 2002; Schlüter et al., 2000). This was the case for all PSC (Fuco, But, Hex, Perid, and Pras) and for six of our PPC (Zea, Viola, DD, DT, Caro, and Lut). Allo, generally reported as photoprotective (Aiken et al., 2009; Henriksen et al., 2002; Schlüter et al., 2000), showed a decreasing ratio with increasing E. However, its contribution to the total pool of PPC was minimal. The ratios of the aggregated photosynthetic accessory pigments (PSPSC + Chlb + Chlc) and PPC, both relative to TChla, showed a comparable pattern as a function of E

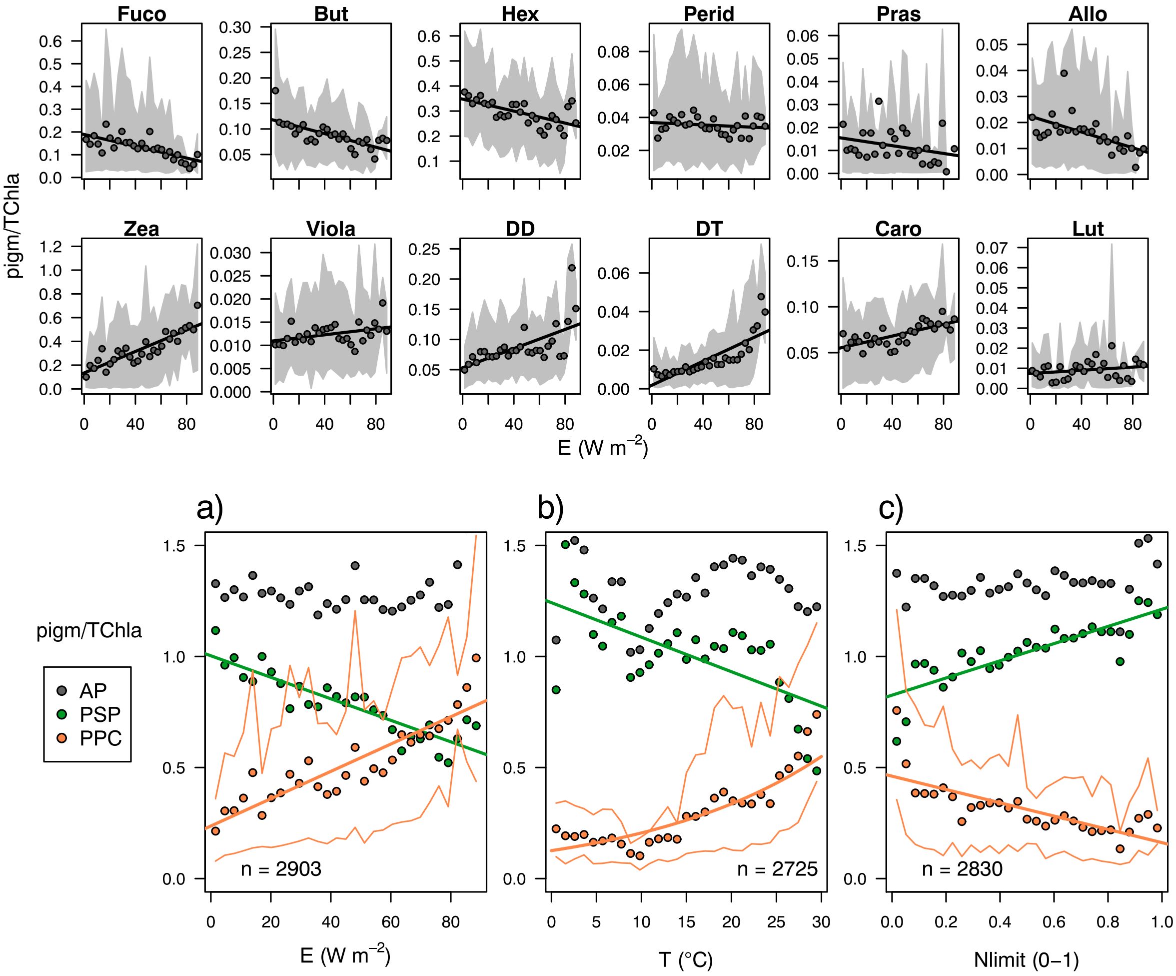


Figure 2. Variability of individual pigments to TChla ratios with light (small panels) and of total photoprotective (PPC), total photosynthetic accessory (PSP), and total accessory pigments (AP) with (a) light, (b) temperature, and (c) nutrient limitation in the EXP-C experiment. Dots indicate mean pigment ratios binned in *E*, *T*, and *Nlimit* classes, respectively. Solid lines show regression models fitted to binned data. Gray areas in the panels for individual pigments and narrow orange lines in the panels for aggregated pigments show percentiles 10 and 90 of the original range of pigment ratios.

(Figure 2a). This matched observed trends in the variation of PSP and xanthophyll's relative to Chla in numerous phytoplankton species (MacIntyre et al., 2002). PSP ad PPC ratios also followed inverse patterns with respect to *T* and *Nlimit* input variables. PSP content decreased with increasing *T* and nutrient limitation (limitation increases as Nlimit approaches 0) whereas PPC content responded opposed to that (Figures 2b and 2c).

The three input variables, *E*, *T*, *Nlimit*, and the interactions *E*:*T* and *E*:*Nlimit* had a significant effect on PPC/TChla (three-factor ANOVA, *p* value < 0.0001, $n = 2,830$), whereas the interactions *T*:*Nlimit* (*p* value $= 0.242$) and *E*:*T*:*Nlimit* (*p* value $= 0.447$) did not. Hence, we examined the ability to predict PPC/TChla through a simple empirical relationship that included all the three input variables as independent variables. To select the type of fit for each input variable, we binned pigments data in 30 *E*, *T*, and *Nlimit* classes, respectively. The best fit to the binned PPC/TChla data was linear for *E* (Figure 2a) but exponential for *T* (Figure 2b). Despite some deviation of the binned data at the extremes of *Nlimit*, PPC/TChla seemed to be reasonably well represented by a linear fit (Figure 2c). We fitted a multiple nonlinear model with nonlinear least squares method, function *nls* (Bates & Watts, 1988) in R (R Core Team, 2018). The resultant empirical model of PPC/TChla as a function of *E*, *T*, and *Nlimit* was PPC/TChla ~ $0.195 + 0.003 * E + (0.045 * \exp(0.070 * T)) - 0.193 * Nlimit$. We also fitted PPC/TChla to the input variables in the two other experiments, EXP-R and EXP-M. There were slight differences in the values of input variables among experiments, differences

Table 5
Skill Metrics per Experiment

Data set	Metric	EXP-R	EXP-M	EXP-C
$\log_{10}$(Chla) (µg L^{-1})	R	0.648	0.649	0.644
(n=9491)	Bias	0.117	0.219	0.158
Chl:C (g:g)	R	0.514	0.520	0.517
(n=9401)	Bias	0.000	0.002	0.001
PPC:TChla NEW (g:g)	R	0.649	0.651	0.649
(n=747)	Bias	-0.015	-0.014	-0.014
PPC:TChla ALL (g:g)	R	-	-	0.663
(n=1082)	Bias	-	-	-0.020
NPP	PgC $year^{-1}$	35.59	33.84	36.21
ExportP	PgC $year^{-1}$	6.82	6.71	6.87

Note. Annual average correlations for $\log_{10}$(Chla), Chla:C, and PPC/TChla at surface waters as compared to satellite OC-CCI Chla, satellite CbPM, and NEW (for empirical approach) or ALL (for mechanistic approach) HPLC data set, respectively. Also, total annual net primary production (NPP) and exported production (ExportP) integrated in the euphotic layer from the respective experiment are provided.

that were given by the effect of Chla on the attenuation of E and the effect of cellular quotas on *Nlimit*. Nevertheless, the differences in fitted coefficients and predictions were small, as shown by the similar metrics of the three empirical fits to observations of PPC/TChla (Table 5).

4.1.2. Mechanistic Prediction

Our approximation combining the models of Geider et al. (1998) and Marshall et al. (2000) provided both a maximum alpha ($\bar{\alpha}$) and the alpha that would appear under the effects of photodamage keeping the xanthophyll pigments pool constant ($\bar{\alpha}_{NP}$). $\bar{\alpha}$ was variable over the global oceans due to variable pigment composition in local phytoplankton communities adapted to varying nutrient and light limitation regimes. REcoM2 generated diversity in $\bar{\alpha}$ using constant group-specific α's and through changes in community composition (Figure 3a). This was a simplification but similar simple approximations have given good results when representing the variability of α in the global ocean (Arteaga et al., 2016). $\bar{\alpha}_{NP}$ was also variable over the global ocean and, overall, highlighted the areas where the risk of photoinhibition was higher (Figure 3b).

To further investigate how $\bar{\alpha}$ compared to $\bar{\alpha}_{NP}$ under the same conditions, we explored how $\bar{\alpha}$ and $\bar{\alpha}_{NP}$ varied as a function of light and nutrient limitation. With an increase in light and nutrient limitation, the community value of $\bar{\alpha}$ changed from a diatom-similar community to a small-phytoplankton-similar community, given a smaller prevalence of diatoms under such conditions (solid black line in Figure 4). $\bar{\alpha}_{NP}$ was close to $\bar{\alpha}$ under nutrient-replete and subsaturated light conditions, as represented in the central section of Figure 4, which indicates favorable conditions to keep a healthy light harvesting apparatus. When deviating from such favorable conditions, $\bar{\alpha}_{NP}$ decreased both with increasing light but also with increasing nutrient stress since repair mechanisms were nutrient limited (dotted lines in Figure 4).

The difference between $\bar{\alpha}$ and $\bar{\alpha}_{NP}$ under a particular set of environmental conditions reflected the degree of photoprotection necessary to keep $\bar{\alpha}_{NP}$ at the maximum value $\bar{\alpha}$. We expected that the difference between $\bar{\alpha}$ and $\bar{\alpha}_{NP}$ was somehow related to the amount of xanthophyll pigments in the case when photoprotection relied predominantly on photoprotective xanthophylls. The ratio between phytoplankton absorption computed without considering PPC and absorption computed considering all pigments has been used as a relative measure of the degree of photoprotection in phytoplankton (Lindley et al., 1995) and hence the amount of photoprotective pigments. We therefore proposed the ratio $\bar{\alpha}_{NP}/\bar{\alpha}$ as a proxy for the photoprotective gap in our model. When $\bar{\alpha}_{NP}/\bar{\alpha}$ was similar to one photodamage had a small impact on the light harvesting apparatus and hence the need for accumulating photoprotective pigments was small. Alternatively, an $\bar{\alpha}_{NP}/\bar{\alpha}$

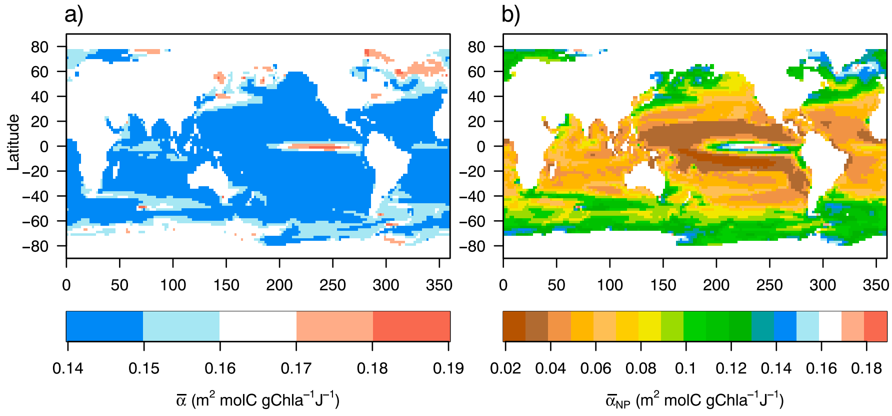


Figure 3. Variability of $\bar{\alpha}$ and $\bar{\alpha}_{NP}$ in the upper 15 m of the water column in the EXP-C experiment: (a) community $\bar{\alpha}$ as derived from standard REcoM2 based on Geider model and (b) $\bar{\alpha}_{NP}$ as derived from the Marshall model with constant xanthophyll pool.

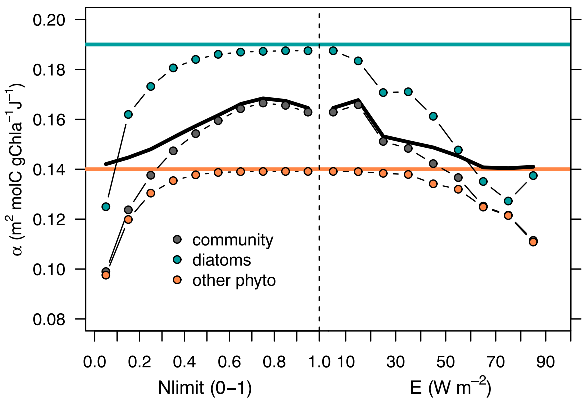


Figure 4. Variability of $\bar{\alpha}$ (lines) and $\bar{\alpha}_{NP}$ (dots) as a function of light (E) and nutrient limitation (*Nlimit*) in the EXP-C experiment. The orange and blue solid lines indicate α of undamaged PSII of diatoms and nondiatoms, respectively, and the black line indicates the community-aggregated value ($\bar{\alpha}$) that changes due to the relative contribution of diatoms and nondiatoms under different environmental conditions. Dotted lines indicate α_{NP} and $\bar{\alpha}_{NP}$ for the same groups.

smaller than one indicated that cells, without accumulating PPC or relying on other heat dissipating mechanisms, would be exposed to photodamage. We then used $1 - \bar{\alpha}_{NP}/\bar{\alpha}$ to predict the amount of accumulated xanthophylls that would avoid such photodamage and compared those predictions to observations of PPC/TChla.

4.1.3. Skill Metrics

To assess the predictions of the models against observations, we compared surface model results against log transformed Chla (micrograms per liter) from OC-CCI, Chl:C (in weight) from CbPM, and PPC/TChla (in weight) from the HPLC data sets, the latter sampled within 0- to 15-m depth. Correlation coefficients and average errors or bias (Stow et al., 2009) were computed for the three experiments EXP-R, EXP-M, and EXP-C, against the observational data sets. For each experiment, the empirical predictions of PPC/TChla were tested against the NEW data set, as this was independent from the MAREDAT data set used to fit the models. The mechanistic approach was only available for the experiment EXP-C, whose predictions were tested against the ALL data set. Results of these comparisons are presented in Table 5.

EXP-R predicted phytoplankton dynamics following Geider et al. (1998) with a photodamage-dependent loss term for Chla and provided Chla fields and Chl:C ratios well correlated to satellite derived observations (Álvarez et al., 2018). The inclusion of the original Marshall et al. (2000) model within REcoM2 (EXP-M), substituting $\bar{\alpha}$ with $\bar{\alpha}_{NP}$ in the P^{C}_{phot}-E curve, did not modify the correlation with surface log-transformed Chla or Chla:C significantly. Instead, it increased the bias of both estimates and decreased the total net primary production on an annual basis from 35.6 to 33.8 PgC/year. This showed that including details on the reversible regulation of PSII seems not to be relevant to predict Chla at global scale and may not be a high priority given the associated increase in computational costs.

The combined approach (EXP-C) provided comparable correlations and bias for Chla and Chl:C as EXP-R but allowed to predict PPC/TChla both empirically and mechanistically. The empirical predictions provided surface values of PPC/TChla highly correlated to observations, with a correlation coefficient of 0.649 and bias of −0.014. The mechanistic predictions that used $\bar{\alpha}$ for production and $\bar{\alpha}_{NP}/\bar{\alpha}$ as a proxy for photoprotection provided a correlation coefficient of 0.663 but a slightly larger bias of −0.020. Note, however, that the ALL data set against which we tested mechanistic predictions was larger than NEW.

4.2. Spatial and Temporal Variability of PPC/TChla

In this section, the empirical and mechanistic predictions of PPC/TChla for EXP-C were compared with HPLC field data. In surface waters (<15-m depth), field data (Figure 5a) showed high concentrations of PPC in tropical and subtropical areas with a decrease around the equator, which was very pronounced in the Pacific but also visible in the Atlantic (Lindley et al., 1995). Smaller values of PPC/TChla were obtained in temperate and polar waters. A comparable latitudinal pattern in PPC content has been reported by Bricaud et al. (2004). The empirical prediction of PPC/TChla showed the same latitudinal pattern with rather horizontal isolines (Figure 5b). The mechanistic prediction (Figure 5c) showed the latitudinal pattern, but with more longitudinal variability. The correlations in surface waters were almost identical for the two models when both were correlated to the NEW in situ data set, 0.649 for the empirical and 0.654 for the mechanistic. Mechanistic predictions showed a slightly larger bias (Table 5). Regardless of the latitudinal pattern, it was remarkable that in areas where the latitudinal pattern was not followed strictly, such as the upwellings of Morocco, Benguela, Peru, and Arabian sea, the mechanistic model matched the observations better than the empirical model, which suggested not-linear effects of nutrients, T and E on PPC/TChla ratios.

Observed and predicted PPC/TChla were compared for surface waters along a latitudinal gradient for the three major ocean basins (Figure 6). The reduced longitudinal variability of the empirical model was observed in the range of PPC/TChla values (orange areas in Figure 6) that did not cover the variability of field data at a given latitude. On the other hand, mechanistic predictions (blue areas in Figure 6) showed

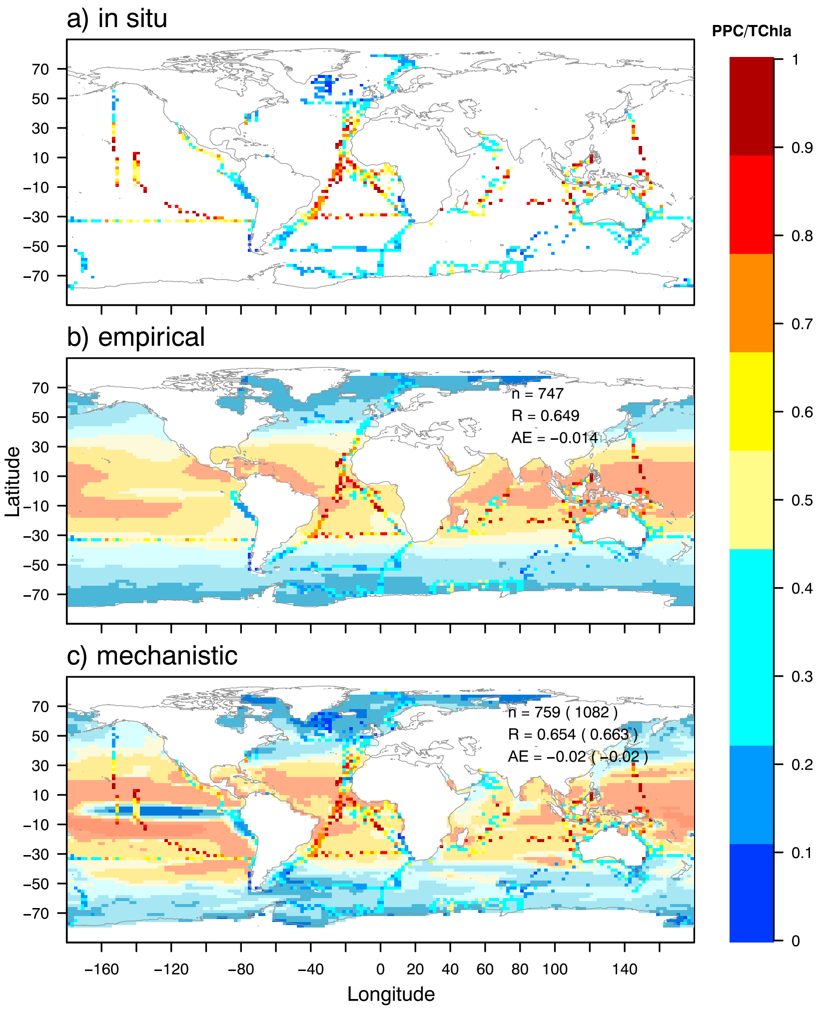


Figure 5. Photoprotective pigments in surface waters (averaged over the upper 15 m and scaled to 2° resolution): (a) PPC/TChla from HPLC data set ALL (also in c) and predicted (b) empirically and (c) mechanistically in the EXP-C experiment. Number of collocated observations (*n*), Person's correlation factor (*R*), and bias (*AE*) are shown for modeling predictions compared against NEW data set and against ALL data set within parenthesis in (c).

more variability and, hence, were able to predict PPC/TChla that did not simply follow a clear latitudinal gradient, as shown in the previous surface plots (Figure 5).

Zonal annual mean depth profiles of observed and predicted PPC/TChla showed that both types of predictions reproduced the in situ depth profiles (Figure 7). The field data showed a decrease of photoprotective pigments with depth in the entire tropical and subtropical ocean. In these areas, significant amounts of photoprotective pigments were observed close to the surface, but even at depths greater than 100 m, we found PPC/TChla ratios larger than 0.4. As expected due to the ambient light north and south of 40°, PPC presence was scarce, below 20% of TChla, but the gradient of decrease with depth was still visible (Figure 7a). The empirical model predicted this gradient quite precisely, although the values in surface tended to be smaller than the observations (Figure 7b). The gradient predicted by the mechanistic model was more abrupt, and values were much lower than observations at depths greater than 75 m. The values at the surface, as shown by previous figures, matched the observations well (Figure 7c). The occasional very high values at depths greater than 100 m in field samples at latitudes <40° were not reflected well in the predictions. While the empirical model predicted a low presence of pigments below 100 m ($0.3 > \mathrm{PPC/TChla} > 0.1$), the mechanistic approach predicted a near absence down to this depth ($\mathrm{PPC/TChla} < 0.1$). By observing

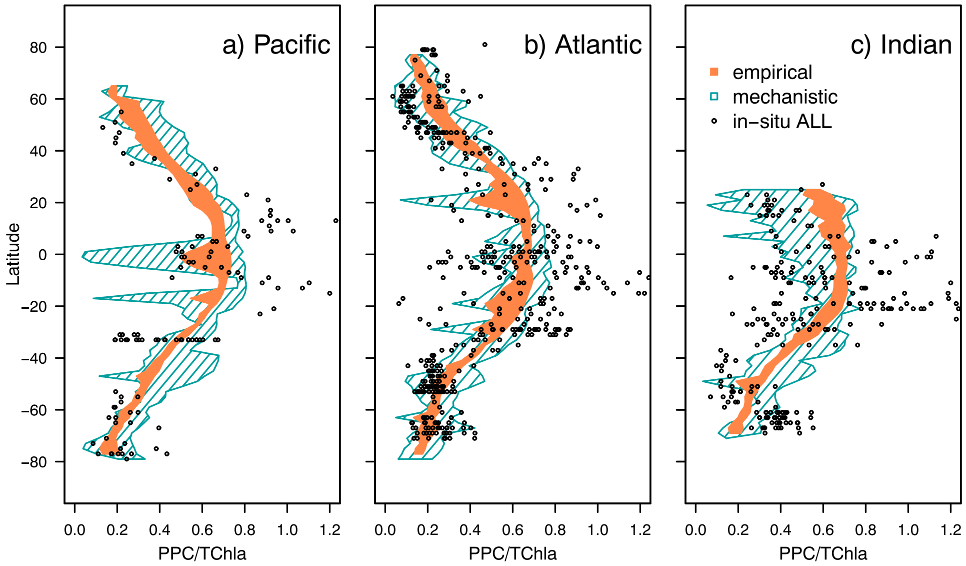


Figure 6. Latitudinal gradients of *in situ* (black dots) and predicted PPC/TChla values from the empirical and mechanistic approaches in the EXP-C experiment across surface waters of the (a) Pacific, (b) Atlantic, and (c) Indian Ocean.

the distribution of predicted TChla (white lines in Figure 7c), it was noticeable that high PPC/TChla values were predicted mainly for the phytoplankton living above the subsurface Chla maximum.

Averaged annual cycles of PPC/TChla predictions using the mechanistic approach (gray areas in Figure 8) showed a clear seasonality in temperate and high latitudes of both hemispheres (Figures 8a and 8c). In both cases, maximum values for the whole range of PPC/TChla occurred during summer and minimum values in

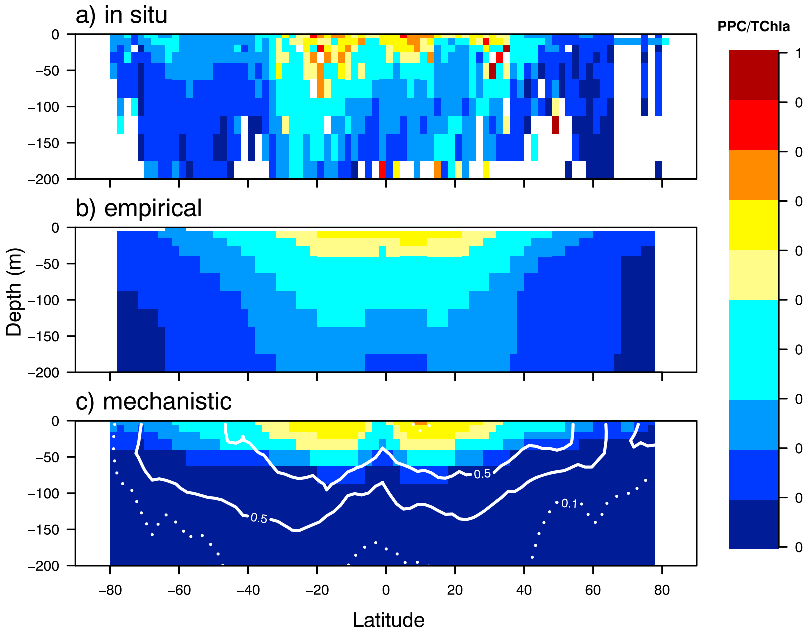


Figure 7. Zonal annual mean depth profiles of PPC/TChla as a function of latitude from (a) in situ HPLC ALL data set and predictions from (b) empirical and (c) mechanistic approaches in the EXP-C experiment. In panel (c) solid isolines show the predicted level of 0.5 μg Chla L^{-1} that enclose the subsurface Chla maximum and the dotted line shows the level of 0.1 μg Chla L^{-1}.

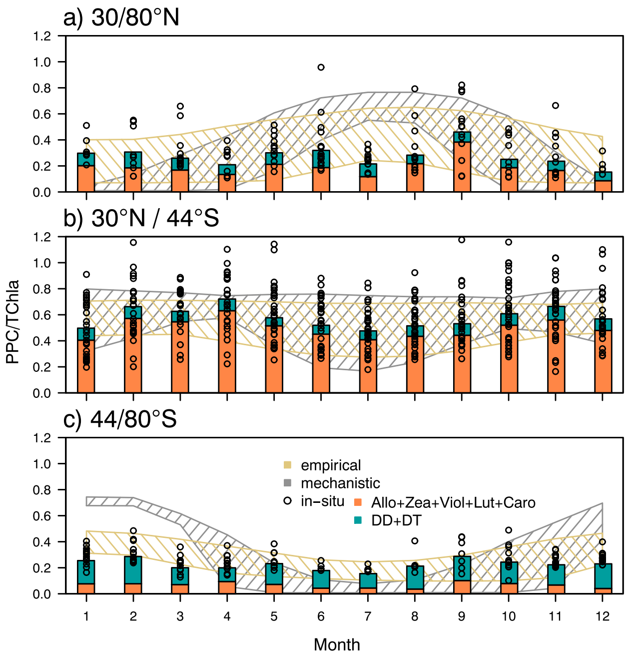


Figure 8. Averaged seasonal distribution of in situ (colored bars reflect the regional monthly average value of the specific PPC pigments and dots reflect the sum of all PPC pigments for each single data point in the respective region) and predicted (the range of values is given) PPC/TChla values in the EXP-C experiment for latitudes (a) north of 30°N, (b) between 30°N and 44°S, and (c) south of 44°S.

winter. In tropical and subtropical areas, the seasonality was only reflected in the range of the lowest values but constant for the maximum ones (Figure 8b). The empirical model predictions showed less seasonality in all regions, as they followed the observations where no seasonality for PPC/TChla was seen. The empirical predictions showed less variability than observations, which were better matched by the range of the mechanistic model predictions. Monthly mean values of in situ PPC/TChla (bars in Figure 8) were divided into the contribution of Allo, Viola, Lut, Zea, and Caro and the contribution of DD and DT. In high latitudes, where the contribution of diatoms was larger, the sum of DD+DT comprised an important proportion of PPC (Figure 8c).

4.3. Change in PPC/TChla: Effect of Community Composition and Physiological Acclimation

On an annual perspective, predicted PPC/TChla matched observations (Figures 5 and 6) or were lower (Figure 7), which indicated that the observed pool of PPC would be enough to cover the needs for photoprotection. Variations in community PPC content, however, can be driven by physiological acclimation as well as shifts in community composition. To explore which mechanism contributed more to the change in community PPC/TChla, both group-specific pigment content and a description of community composition were needed. This was challenging for observations because we did not have an independent estimate of community composition but the pigments themselves. The estimate of the contribution of diatoms to total chlorophyll relied on the use of DPs (Uitz et al., 2006; Figure 9a). Although the contribution of diatoms to total PPC pool can be illustrated with the sum DD+DT (Strain et al., 1944; Figure 9b), other groups can share the same pigments. This uncertainty could translate to the empirical model since it required a group-specific fit to observations to provide predictions for each individual group.

The advantage of the mechanistic approach in this context was that, in addition to predicting group specific pigment content, it provided a full description of community composition. Phytoplankton diversity represented by just two groups allowed us to build a description of the entire community composition by a single index of dominance that reflected the contribution of one of the groups to the total biomass. On an annual basis, the contribution of diatoms Chla to TChla (Figure 9c) and the contribution of diatom PPC to the total pool of PPC (Figure 9d) showed dominance of diatoms at the equator, high latitudes and near the eastern coasts of continents at midlatitudes.

In the absence of physiological acclimation that changes group-specific PPC/TChla ratios, dominance indexes (% of diatoms) in terms of TChla and in terms of PPC would vary in parallel. The variation over the year of the % of diatoms in terms of Chla, PPC, and also C were explored for nine areas of the world ocean (Figure 9). A noticeable difference between the time derivatives of the PPC- and Chla-based dominances (gray bars in Figure 9) indicated that the ratio PPC/TChla was variable within one or both phytoplankton groups. In the central areas of the Pacific, Atlantic, and Indian Oceans, dominance indexes varied in parallel, which indicated that changes in community PPC content were driven by changes in community composition. At higher latitudes, prior to the polar winter, PPC-based dominance changed faster than Chla-based dominance while prior to polar summer, PPC dominance changed slower than Chla-based dominance. Both phenomena implied intragroup acclimative changes in the PPC/TChla ratios.

5. Discussion

As in higher plants, most phytoplankton cells possess specialized carotenoid pigments that contribute to the rapid and harmless thermal dissipation of excess absorbed light energy in response to a sudden increase in irradiance (Demmig-Adams & Adams, 1992). This function reduces the excitation pressure on the reaction center of PSII and limit photoinhibition. Marshall et al. (2000) proposed a mechanistic model to describe

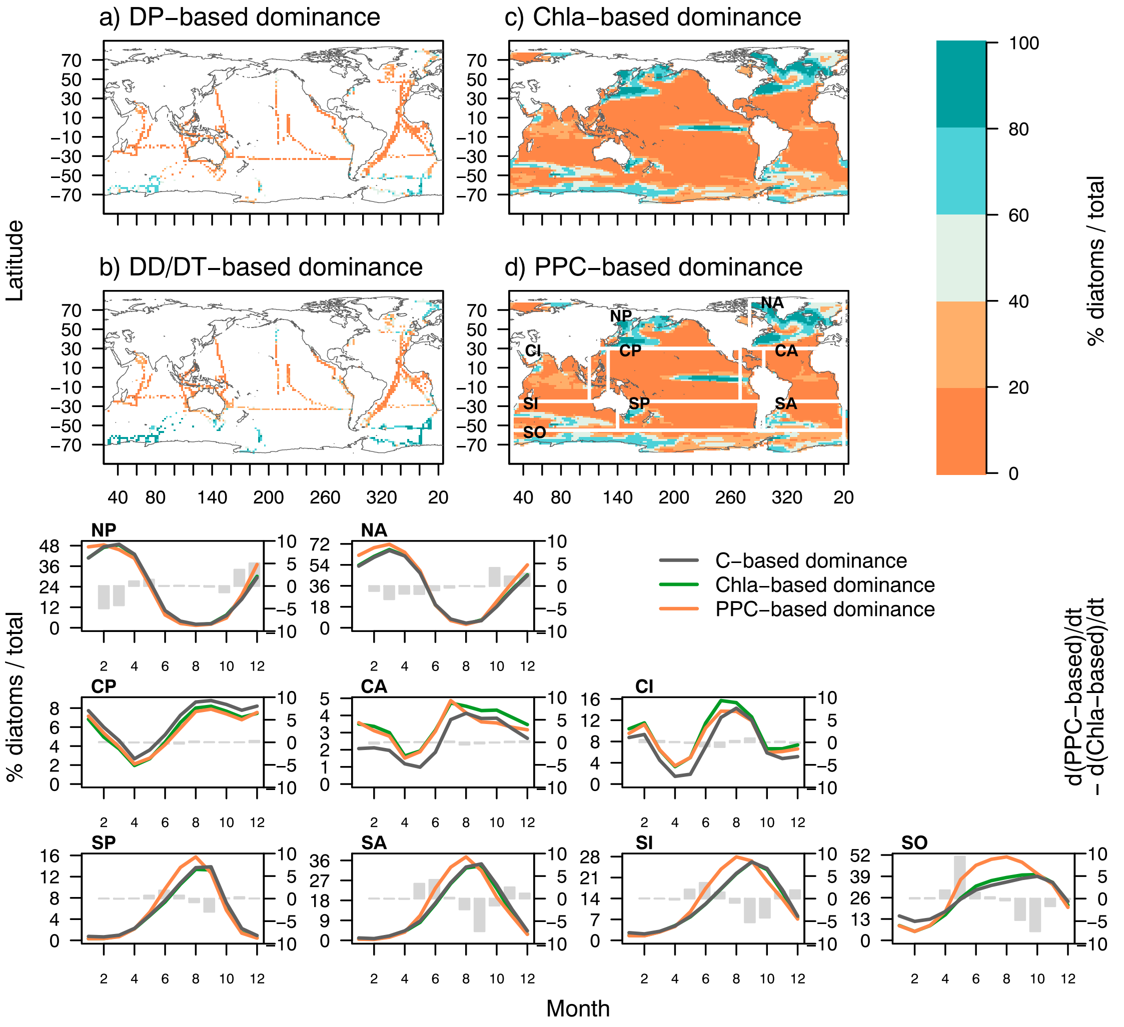


Figure 9. Observed contribution of diatoms (%) to total community in terms of (a) Chla derived from diagnostic pigments (from fd_{Chla}; Uitz et al., 2006) and (b) sum DD+DT to total PPC. Predicted contribution of diatoms (%) to total community in terms of (c) Chla and (d) PPC. Small panels show the variation along the year of C-, Chla-, and PPC-based dominance (% diatoms) in the nine ocean regions indicated in (d) left axis (note that the scale range is different for each panel). Gray bars show the difference between the time derivatives of the PPC-based and the Chla-based dominance (right axis, all using the same scale).

photoinhibition as a consequence of the decrease in the proportion of active PSII and the effect on the initial slope of the PE curve (α). Photoinhibition is expected to be small for natural phytoplankton communities acclimated to ambient light (Cullen et al., 1992), so we can assume that phytoplankton cells possess the mechanisms to keep α at a maximum value. Among these mechanisms, the presence of xanthophyll pigments is certainly expected to play a role since the thermal dissipation process is assisted by the nearly universal function of xanthophylls (Demmig-Adams & Adams, 1996). We therefore used the Marshall et al. (2000) model to estimate the extent of photoinhibition that phytoplankton communities would have to deal with. We explored whether the photoprotective role was performed by a variable pool of xanthophylls and whether the variability in the community xanthophyll pool was driven by shifts in community composition or implied intraspecific adjustments of pigment content.

5.1. Relevance of Photoprotective Pigments

Both this mechanistic and an empirical approach based on parallel in situ observations of PPC/TChla were able to predict the long-term photoprotective response and simulate the cellular accumulation of xanthophylls. The mechanistically determined need for photoprotection was in agreement with field observations of PPC/TChla in most of the surface ocean, at depth and even across seasons. This suggested a predominant role of PPC in photoprotective activities in the global ocean with implications for the fast regulation of photosynthetic productivity and carbon fluxes in the ocean.

The approach presented allowed us to predict PPC content via modeling and hence provided a comprehensive description of pigment content of the phytoplankton community at the global scale provided by models. The content in photosynthetic pigments (Chla) and nonphotosynthetic pigments (PPC) relative to biomass (C) relates to the contribution of photochemical and nonphotochemical pathways, respectively, to the fate of absorbed light (Lin et al., 2016). Knowledge of full pigment signatures is crucial for understanding the coupling of light absorption to carbon fixation in the ocean.

5.2. Relevance of Other Physiological Mechanisms

Discrepancies between the mechanistically determined need for photoprotection and observations of PPC/TChla highlighted scenarios where PPC content was apparently insufficient to protect the community and thus other mechanisms of NPQ became relevant for heat dissipation. Particularly during summer season at high latitudes (Figure 8), the amount of PPC fell below the needs for photoprotection, suggesting that in these areas alternative mechanisms may be relevant for photoprotection.

The difference between predictions and observations can be explained by simplifications of the model framework. A functional classification of carotenoids into two groups of pigments, photosynthetic and photoprotective, is commonly used in the literature (e.g., Bidigare et al., 1990). However, in some groups, the photosynthetic pigment fucoxanthin may be converted to DD relatively quickly (Harris et al., 2009; Polimene et al., 2012), thus giving fucoxanthin a role in long-term photoprotection. Also, the effectiveness of the XC varies across taxa (Goss & Jakob, 2010; Lavaud, 2007; Six et al., 2009), which can lead to smaller PPC content in communities dominated by groups containing more efficient XC. In addition to pigments, energy dissipation can be covered by other protective mechanisms not explicitly considered here, such as fluorescence or constitutive heat dissipation (Porcar-Castell et al., 2006). Autotrophic cells also possess some plasticity to reorganize their photochemical pathways when the photochemical capacity is exceeded, performing cyclic electron flow around PSII (Goss & Lepetit, 2015) or alternative electron transport (Wagner et al., 2006).

5.3. Change in PPC Content: Intragroup Acclimation or Shifts in Community Composition

Given that results showed that variable PPC/TChla ratios covered most of phytoplankton photoprotective needs, the question that arose next was whether this variability in pigment content was driven by shifts in community composition or implied physiological acclimation within a given group. The mechanistic approach proposed provided the two elements, which are necessary to tackle this question: group-specific PPC predictions and a complete description of community composition.

Group-specific photoprotective needs were generated by assigning smaller damage coefficients to diatoms than to other phytoplankton, larger photochemical efficiency to diatoms, and equal absorption cross-sections and repair rates to both groups. Although the variability in the predictions was high for both groups, the predicted needs for photoprotection tended to be higher in other phytoplankton than in diatoms. This corresponds to the observation that smaller phytoplankton can grow in a high but relatively constant light environment where they do not need to invest as heavily in photosynthetic machinery but require mechanisms to reduce the harmful effect of extensive periods of excess irradiance (Kropuenske et al., 2009). Smaller phytoplankton like prasinophytes cope with prolonged stress and subsequent recovery through a large induction and relaxation of XC-induced NPQ (Liefer et al., 2018). Diatoms on the other hand show lower susceptibility to photoinactivation of PSII (Key et al., 2010) and rely on constitutive dissipation of excitation energy more than on XC NPQ (Liefer et al., 2018).

Identification of microalgae species through diagnostic pigments implies the assumption that changes in community accessory pigments are solely driven by changes in community composition. This is reasonable when the pigment ratios are constant in time for a particular species. However, variations in diagnostic pigments to Chla ratios also arise due to factors such as nutrient status and light (Mackey et al., 1996). Our results showed that, for most of the tropical and subtropical ocean, photoprotective pigments dominated photoprotection and changes in total photoprotective pigments were caused mainly by taxonomic changes within the phytoplankton community. Our results also highlighted scenarios at high latitudes where changes in PPC content can also result both from shifts in community composition and from photoacclimation within the same community. During polar winter, in absence of nutrient limitation, physiological acclimation took place and contributed to changes in PPC/TChla ratios (Figure 9) that met photoprotective needs (Figures 8a and 8c).

5.4. The Role of Nutrients and Temperature on High-Latitude Photoprotection

During polar summer, the high requirement of photoprotection predicted by our mechanistic approach was not observed in field measured PPC/TChla (Figures 8a and 8c). At the same time, there was no sign of physiological acclimation (Figure 9). This suggested not only that other mechanisms of photoprotection may be more relevant than PPC-mediated mechanisms in these areas but also that PPC synthesis itself was limited. The observed physiological acclimation during winter when nutrients were replenished and temperatures were low suggested that nutrients and not temperature limited the use of PPC-based NPQ in those polar areas during the summer season. Under incomplete photoprotection provided by other mechanisms, net photodamage could eventually occur since the repair of damaged photosystems was both nutrient and temperature limited. In a previous application of the REcoM2 model, we showed that given the nutrient requirements of repair mechanisms, severe nutrient stress translates into photoinhibition of the light harvesting apparatus in the absence of complete photoprotection (Álvarez et al., 2018). Since it has been observed that photodamage can shape the phytoplankton community in the Southern Ocean (Alderkamp et al., 2010), our results show that nutrient limitation of PSII repair and/or PPC de novo synthesis may be the cause of these photoinhibitory responses at high latitudes.

6. Conclusions

This study suggests that a variable pool of xanthophyll pigments mediates the long-term photoprotective responses of phytoplankton throughout most of the ocean. Hence, the photoprotective pigments accumulated by phytoplankton can be accurately predictable by models. This potentially provides a comprehensive view of the phytoplankton community pigment signature in terms of photosynthetic and nonphotosynthetic pigments, which is crucial for modeling the coupling of light absorption to carbon fixation in the ocean. The variability in the pool of xanthophyll pigments seems to be driven mainly by changes in community composition in the tropical and subtropical ocean, and in these areas pigment composition may reflect taxonomic composition strongly. At higher latitudes, during the nutrient-limited summer season, photoprotection needs do not seem to be met by changes in xanthophyll pigments and must be met by other energy-dissipating mechanisms. When nutrients are resupplied and the nutrient limitation of pigment synthesis relaxes, both taxonomic shifts and intraspecific photoacclimation appear to shape community pigment signature. These insights about how much of the variability in community nonphotosynthetic pigments is attributable to changes in community composition or changes in physiological state may allow an improvement of the match between remotely sensed optical data and the underlying phytoplankton community.

Acknowledgments
This work is a contribution to the IPSO project “Improving the prediction of photophysiology in the Southern Ocean by accounting for iron limitation, optical properties and spectral satellite data information” funded by the Alfred Wegener Institute. A. B.’s contribution is partly supported by the Transregional Collaborative Research Center (TR 172) “ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)3” via the subproject C03. Y. L.’s contribution is supported with a PhD fellowship by the China Scholarship Council. For HPLC analysis of filters sampled at PS103 and HE462, we thank Sonja Wiegmann, and for supporting the water sampling and filtration at these cruises, we thank Rafael Goncalves Araujo and Julia Oelker (PS103), and Sebastian Hellman and Sonja Wiegmann (HE462). We also thank the captains and the crews, especially the chief scientists Olaf Boebel (PS103), Thomas Soltwedel (PS107), and Sören Kraegevsky (HE462), for their support. We thank two anonymous reviewers for their helpful and constructive comments that improved the manuscript. All field data used in this work and listed in Table 4 are publicly available in PANGAEA under the url's included in the references list and in the AEsOP database (under http://aesop.csiro.au/).

References

Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., & Hardman-Mountford, N. (2009). Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995-2005. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *56*(15), 899–917. https://doi.org/10.1016/j.dsr2.2008.09.017

Alderkamp, A.-C., de Baar, H. J. W., Visser, R. J. W., & Arrigo, K. R. (2010). Can photoinhibition control phytoplankton abundance in deeply mixed water columns of the Southern Ocean? *Limnology and Oceanography*, *55*(3), 1248–1264. https://doi.org/10.4319/lo.2010.55.3.1248

Álvarez, E., Thoms, S., & Völker, C. (2018). Chlorophyll to carbon ratio derived from a global ecosystem model with photodamage. *Global Biogeochemical Cycles*, *32*, 799–816. https://doi.org/10.1029/2017GB005850

Antonov, J. I., Seidov, D., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Garcia, H. E., et al. (2010). World Ocean Atlas 2009, volume 2: Salinity, 40 pp. In S. Levitus (Ed.), NOAA Atlas NESDIS 69 (pp. 184). Washington, DC: U.S. Gov. Printing Office.

Arteaga, L., Pahlow, M., & Oschlies, A. (2016). Modeled Chl:C ratio and derived estimates of phytoplankton carbon biomass and its contribution to total particulate organic carbon in the global surface ocean. *Global Biogeochemical Cycles*, *30*, 1791–1810. https://doi.org/10.1002/2016GB005458

Aumont, O., Maier-Reimer, E., Blain, S., & Monfray, P. (2003). An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles*, *17*(2), 1060. https://doi.org/10.1029/2001GB001745

Babin, M., Morel, A., Claustre, H., Bricaud, A., Kolber, Z., & Falkowski, P. G. (1996). Nitrogen- and irradiance-dependent variations of the maximum quantum yield of carbon fixation in eutrophic, mesotrophic and oligotrophic marine systems. *Deep-Sea Research Part I: Oceanographic Research Papers*, *43*(8), 1241–1272. https://doi.org/10.1016/0967-0637(96)00058-1

Baird, M., Mongin, M., Rizwi, F., Bay, L., Cantin, N., Skerratt, J., & Soja-Wozniak, M. (2018). A mechanistic model of coral bleaching due to temperature-mediated light-driven reactive oxygen build-up in zooxanthellae. *Ecological Modelling*, *386*, 20–37. https://doi.org/10.31230/osf.io/etjd6

Barlow, R. G., Cummings, D. G., & Gibb, S. W. (1997). Improved resolution of mono- and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC. *Marine Ecology Progress Series*, *161*, 303–307. https://doi.org/10.3354/meps161303

Bates, D. M., & Watts, D. G. (1988). Nonlinear Regression: Iterative Estimation and Linear Approximations. In D. M. Bates & D. G. Watts (Eds.), *Nonlinear Regression Analysis and Its Applications*. https://doi.org/10.1002/9780470316757.ch2

Bidigare, R. R., Ondrusek, M. E., Morrow, J. H., & Kiefer, D. A. (1990). In-vivo absorption properties of algal pigments. In *Proc. SPIE 1302, Ocean Optics X*, (pp. 556–565). Orlando, FL: 1990 Technical Symposium on Optics, Electro-Optics, and Sensors. https://doi.org/10.1117/12.21451

Booge, D., Schlundt, C., Bracher, A., Endres, S., Zäncker, B., & Marandino, C. A. (2018). Marine isoprene production and consumption in the mixed layer of the surface ocean—A field study over two oceanic regions. *Biogeosciences*, *15*(2), 649–667. https://doi.org/10.5194/bg-15-649-2018

Bracher, A. (2014a). Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXVIII/3. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848588

Bracher, A. (2014b). Phytoplankton pigments measured on water bottle samples during SONNE cruise SO218. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848589

Bracher, A. (2015a). Phytoplankton pigment concentrations during Maria S. Merian cruise MSM18/3. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848586

Bracher, A. (2015b). Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXIV/1. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848583

Bracher, A. (2015c). Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXIV/4. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848584

Bracher, A. (2015d). Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXVI/4. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848585

Bracher, A. (2015e). Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXVII/2. *PANGAEA*. https://doi.org/10.1594/PANGAEA.848590

Bracher, A. (2019a). Phytoplankton pigment concentrations during RV Sonne cruise SO243. *PANGAEA*. https://doi.org/10.1594/PANGAEA.898920

Bracher, A. (2019b). Phytoplankton pigment concentrations in the Southern Ocean during RV Polarstern cruise PS103 in Dec 2016 to Jan 2017. *PANGAEA*. https://doi.org/10.1594/PANGAEA.898941

Bracher, A., Cheah, W., & Wiegmann, S. (2019). Phytoplankton pigment concentrations in the tropical Indian Ocean in July and August 2014 during RV Sonne cruises SO234 and SO235. *PANGAEA*. https://doi.org/10.1594/PANGAEA.898929

Bracher, A., & Taylor, B. B. (2017). Phytoplankton pigment concentrations measured by HPLC during Maria S. Merian cruise MSM9/1. *PANGAEA*. https://doi.org/10.1594/PANGAEA.873070

Bracher, A., Taylor, M. H., Taylor, B. B., Dinter, T., Röttgers, R., & Steinmetz, F. (2015a). Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXIII/1. *PANGAEA*. https://doi.org/10.1594/PANGAEA.871713

Bracher, A., Taylor, M. H., Taylor, B. B., Dinter, T., Röttgers, R., & Steinmetz, F. (2015b). Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Ocean Science*, *11*(1), 139–158. https://doi.org/10.5194/os-11-139-2015

Bracher, A., & Wiegmann, S. (2019). Heincke Cruise HE462 phytoplankton pigments determined by HPLC. *PANGAEA*. https://doi.org/10.1594/PANGAEA.899043

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

Brunet, C., Johnsen, G., Lavaud, J., & Roy, S. (2011). Pigments and photoacclimation processes. In S. Roy, C. Llewellyn, E. Skarstad Egeland, & G. Johnsen (Eds.), *Phytoplankton Pigments, Characterization, Chemotaxonomy and Application in Oceanography*, (pp. 445–454). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511732263.017

Brunet, C., & Lavaud, J. (2010). Can the xanthophyll cycle help extract the essence of the microalgal functional response to a variable light environment? *Journal of Plankton Research*, *32*(12), 1609–1617. https://doi.org/10.1093/plankt/fbq104

Campbell, D. A., & Tyystjärvi, E. (2012). Parameterization of photosystem II photoinactivation and repair. *Biochimica et Biophysica Acta-Bioenergetics*, *1817*(1), 258–265. https://doi.org/10.1016/j.bbabio.2011.04.010

Cullen, J. J., Yang, X., & MacIntyre, H. L. (1992). Nutrient limitation of marine photosynthesis. In P. G. Falkowski, & A. D. Woodhead (Eds.), *Primary Productivity and Biogeochemical Cycles in the Sea*, (pp. 70–88). New York: Springer. https://doi.org/10.1007/978-1-4899-0762-2_5

Demers, S., Roy, S., Gagnon, R., & Vignault, C. (1991). Rapid light-induced changes in cell fluorescence and in xanthophyll-cycle pigments of Alexandrium excavatum (Dinophyceae) and Thalassiosira pseudonana (Bacillario-phyceae): A photo-protection mechanism. *Marine Ecology Progress Series*, *76*, 185–193. https://doi.org/10.3354/meps076185

Demmig-Adams, B., & Adams, W. W. (1992). Responses of plants to high light stress. *Annual Review of Plant Physiology and Plant Molecular Biology*, *43*(1), 599–626. https://doi.org/10.1146/annurev.pp.43.060192.003123

Demmig-Adams, B., & Adams, W. W. (1996). The role of xanthophyll cycle carotenoids in the protection of photosynthesis. *Trends in Plant Science*, *1*(1), 21–26. https://doi.org/10.1016/S1360-1385(96)80019-7

Du, N., Gholami, P., Kline, D. I., DuPont, C. L., Dickson, A. G., Mendola, D., et al. (2018). Simultaneous quantum yield measurements of carbon uptake and oxygen evolution in microalgal cultures. *PLoS ONE*, *13*(6), e0199125. https://doi.org/10.1371/journal.pone.0199125

Dubinsky, Z., & Schofield, O. (2010). From the light to the darkness: Thriving at the light extremes in the oceans. *Hydrobiologia*, *639*(1), 153–171. https://doi.org/10.1007/s10750-009-0026-0

Dubinsky, Z., & Stambler, N. (2009). Photoacclimation processes in phytoplankton: Mechanisms, consequences, and applications. *Aquatic Microbial Ecology*, *56*, 163–176. https://doi.org/10.3354/ame01345

Flynn, K. J., & Flynn, K. (1998). Release of nitrite by marine dinoagellates: Development of a mathematical simulation. *Marine Biology*, *130*(3), 455–470. https://doi.org/10.1007/s002270050266

Flynn, K. J., Page, S., Wood, G., & Hipkin, C. R. (1999). Variations in the maximum transport rates for ammonium and nitrate in the prymnesiophyte Emiliania huxleyi and the raphidophyte Heterosigma carterae. *Journal of Plankton Research*, *21*(2), 355–371. https://doi.org/10.1093/plankt/21.2.355

Fujiki, T., & Taguchi, S. (2001). Relationship between light absorption and the xanthophyll-cycle pigments in marine diatoms. *Plankton Biology and Ecology*, *48*(2), 96–103.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K., & Johnson, D. R. (2010). World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, and silicate), 26 pp. In *NOAA Atlas NESDIS 71*. Washington, DC: U.S. Government Printing Office. 398 pp.

Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1997). Dynamic model of phytoplankton growth and acclimation: Responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature. *Marine Ecology Progress Series*, *148*(1–3), 187–200. https://doi.org/10.3354/meps148187

Geider, R. J., Maclntyre, H. L., & Kana, T. M. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, *43*(4), 679–694. https://doi.org/10.4319/lo.1998.43.4.0679

Goss, R., & Jakob, T. (2010). Regulation and function of xanthophyll cycle-dependent photoprotection in algae. *Photosynthesis Research*, *106*(1-2), 103–122. https://doi.org/10.1007/s11120-010-9536-x

Goss, R., & Lepetit, B. (2015). Biodiversity of NPQ. *Journal of Plant Physiology*, *172*, 13–32. https://doi.org/10.1016/j.jplph.2014.03.004

Gustafsson, M. S. M., Baird, M. E., & Ralph, P. J. (2014). Modeling photoinhibition-driven bleaching in Scleractinian coral as a function of light, temperature, and heterotrophy. *Limnology and Oceanography*, *59*(2), 603–622. https://doi.org/10.4319/lo.2014.59.2.0603

Han, B.-P. (2002). A mechanistic model of algal photoinhibition induced by photodamage to photosystem-II. *Journal of Theoretical Biology*, *214*(4), 519–527. https://doi.org/10.1006/jtbi.2001.2468

Han, B.-P., Virtane, M., Koponen, J., & Straskraba, M. (2000). Effect of photoinhibition on algal photosynthesis: a dynamic model. *Journal of Plankton Research*, *22*(5), 865–885. https://doi.org/10.1093/plankt/22.5.865

Harris, G. N., Scanlan, D. J., & Geider, R. J. (2009). Responses of Emiliania huxleyi (Prymnesiophyceae) to step changes in photon flux density. *European Journal of Phycology*, *44*(1), 31–48. https://doi.org/10.1080/09670260802233460

Hauck, J., Völker, C., Wang, T., Hoppema, M., Losch, M., & Wolf-Gladrow, D. A. (2013). Seasonally different carbon flux changes in the Southern Ocean in response to the southern annular mode. *Global Biogeochem Cycles*, *27*, 1236–1245. https://doi.org/10.1002/2013GB004600

Henriksen, P., Riemann, B., Kaas, H., Sorensen, H. L. H. M., & Sørensen, H. M. H. L. (2002). Effects of nutrient-limitation and irradiance on marine phytoplankton pigments. *Journal of Plankton Research*, *24*(9), 835–858. https://doi.org/10.1093/plankt/24.9.835

Hohn, S. (2009). Coupling and decoupling of biogeochemical cycles in marine ecosystems (Ph.D Thesis). Universität Bremen.

Horton, P., Ruban, A. V., & Wentworth, M. (2000). Allosteric regulation of the light-harvesting system of photosystem II. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1402), 1361–1370. https://doi.org/10.1098/rstb.2000.0698

Jahns, P., Latowski, D., & Strzalka, K. (2009). Mechanism and regulation of the violaxanthin cycle: The role of antenna proteins and membrane lipids. *Biochimica et Biophysica Acta-Bioenergetics*, *1787*(1), 3–14. https://doi.org/10.1016/j.bbabio.2008.09.013

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., et al. (2004). A global ocean carbon climatology: Results from GLODAP. *Global Biogeochemical Cycles*, *18*, GB4031. https://doi.org/10.1029/2004GB002247

Key, T., McCarthy, A., Campbell, D. A., Six, C., Roy, S., & Finkel, Z. V. (2010). Cell size trade-offs govern light exploitation strategies in marine phytoplankton. *Environmental Microbiology*, *12*(1), 95–104. https://doi.org/10.1111/j.1462-2920.2009.02046.x

Kiefer, D. A., & Mitchell, B. G. (1983). A simple steady state description of phytoplankton growth based on absorption cross section and quantum efficiency. *Limnology and Oceanography*, *28*(4), 770–776. https://doi.org/10.4319/lo.1983.28.4.0770

Kolber, Z., Zehr, J., & Falkowski, P. (1988). Effects of growth irradiance and nitrogen limitation on photosynthetic energy conversion in photosystem II. *Plant Physiology*, *88*(3), 923–929. https://doi.org/10.1104/pp.88.3.923

Kromkamp, J. C., Domin, A., Dubinsky, Z., Lehmann, C., & Schanz, F. (2001). Changes in photosynthetic properties measured by oxygen evolution and variable chlorophyll fluorescence in a simulated entrainment experiment with the cyanobacterium Planktothrix rubescens. *Aquatic Sciences*, *63*(3), 363–382. https://doi.org/10.1007/PL00001360

Kropuenske, L. R., Mills, M. M., van Dijken, G. L., Bailey, S., Robinson, D. H., Welschmeyer, N. A., & Arrigoa, K. R. (2009). Photophysiology in two major Southern Ocean phytoplankton taxa: Photoprotection in *Phaeocystis antarctica* and *Fragilariopsis cylindrus*. *Limnology and Oceanography*, *54*(4), 1176–1196. https://doi.org/10.4319/lo.2009.54.4.1176

Lavaud, J. (2007). Fast regulation of photosynthesis in diatoms: Mechanisms, evolution and ecophysiology. *Functional Plant Science and Biotechnology*, *1*, 267–287.

Lavaud, J., Six, C., & Campbell, D. A. (2016). Photosystem II repair in marine diatoms with contrasting photophysiologies. *Photosynthesis Research*, *127*(2), 189–199. https://doi.org/10.1007/s11120-015-0172-3

Liefer, J. D., Garg, A., Campbell, D. A., Irwin, A. J., & Finkel, Z. V. (2018). Nitrogen starvation induces distinct photosynthetic responses and recovery dynamics in diatoms and prasinophytes. *PLoS ONE*, *13*(4), 1–24. https://doi.org/10.1371/journal.pone.0195705

Lin, H., Kuzminov, F. I., Park, J., Lee, S. H., Falkowski, P. G., & Gorbunov, M. Y. (2016). Phytoplankton: The fate of photons absorbed by phytoplankton in the global ocean. *Science*, *351*(6270), 264–267. https://doi.org/10.1126/science.aab2213

Lindley, S. T., Bidigare, R. R., & Barber, R. T. (1995). Phytoplankton photosynthesis parameters along 140 degrees W in the equatorial Pacific. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *42*(2–3), 441–463. https://doi.org/10.1016/0967-0645(95)00029-P

Liu, Y., Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., et al. (2019a). Phytoplankton pigment concentration measured by HPLC during POLARSTERN cruises PS93.2, PS99.2 and PS107. *PANGAEA*. https://doi.org/10.1594/PANGAEA.894875

Liu, Y., Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., et al. (2019b). Retrieval of phytoplankton pigments from underway spectrophotometry in the Fram Strait. *Remote Sensing*, *11*(3), 318. https://doi.org/10.3390/rs11030318

Liu, Y., Röttgers, R., Ramírez-Pérez, M., Dinter, T., Steinmetz, F., Nöthig, E. M., et al. (2018). Underway spectrophotometry in the Fram Strait (European Arctic Ocean): a highly resolved chlorophyll a data source for complementing satellite ocean color. *Optics Express*, *26*(14), A678–A696. https://doi.org/10.1364/OE.26.00A678

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., et al. (2010). World Ocean Atlas 2009, volume 1: Temperature, 40 pp. In S. Levitus (Ed.), NOAA Atlas NESDIS 68. Washington, DC: U.S. Government Printing Office 184 pp.

MacIntyre, H. L., Kana, T. M., Anning, J., Geider, R. J., Anning, T., & Geider, R. J. (2002). Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria. *Journal of Phycology*, *38*(1), 17–38. https://doi.org/10.1046/j.1529-8817.2002.00094.x

Mackey, M. D., Mackey, D. J., Higgins, H. W., & Wright, S. W. (1996). CHEMTAX - a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series*, *144*, 265–283. https://doi.org/10.3354/meps144265

Magdaong, N. C. M., & Blankenship, R. E. (2018). Photoprotective, excited-state quenching mechanisms in diverse photosynthetic organisms. *The Journal of Biological Chemistry*, *293*(14), 5018–5025. https://doi.org/10.1074/jbc.TM117.000233

Marshall, H. L., Geider, R. J., & Flynn, K. J. (2000). A mechanistic model of photoinhibition. *The New Phytologist*, *145*(2), 347–359. https://doi.org/10.1046/j.1469-8137.2000.00575.x

Masamoto, K., Zsiros, O., & Gombos, Z. (1999). Accumulation of zeaxanthin in cytoplasmic membranes of the cyanobacterium Synechococcus sp. strain PCC7942 grown under high light condition. *Journal of Plant Physiology*, *155*(1), 136–138. https://doi.org/10.1016/S0176-1617(99)80155-2

Megard, R. O., Combs, W. S., Smith, P. D., & Knoll, A. S. (1979). Attenuation of light and daily integral rates of photosynthesis attained by planktonic algae. *Limnology and Oceanography*, *24*(6), 1038–1050. https://doi.org/10.4319/lo.1979.24.6.1038

Mellis, A. (1999). Photosystem-II damage and repair cycle in chloroplasts: What modulates the rate of photodamage in vivo? *Trends in Plant Science*, *4*(4), 130–135. https://doi.org/10.1016/S1360-1385(99)01387-4

Moore, C. M., Suggett, D. J., Hickman, A. E., Kim, Y.-N., Tweddle, J. F., Sharples, J., et al. (2006). Phytoplankton photoacclimation and photoadaptation in response to environmental gradients in a shelf sea. *Limnology and Oceanography*, *51*(2), 936–949. https://doi.org/10.4319/lo.2006.51.2.0936

Müller, P., Li, X. P., & Niyogi, K. K. (2001). Non-photochemical quenching. A response to excess light energy. *Plant Physiology*, *125*(4), 1558–1566. https://doi.org/10.1104/pp.125.4.1558

Nagy, L., Bálint, E., Barber, J., Ringler, A., Cook, K. M., & Maroti, P. (1995). Photoinhibition and law of reciprocity in photosynthetic reactions of Synechocystis sp. PCC 6803. *Journal of Plant Physiology*, *145*(4), 410–415. https://doi.org/10.1016/S0176-1617(11)81763-3

Ni, G., Zimbalatti, G., Murphy, C. D., Barnett, A. B., Arsenault, C. M., Li, G., et al. (2017). Arctic Micromonas uses protein pools and non-photochemical quenching to cope with temperature restrictions on Photosystem II protein turnover. *Photosynthesis Research*, *131*(2), 203–220. https://doi.org/10.1007/s11120-016-0310-6

Oliver, R. L., & Ganf, G. G. (1988). The optical properties of a turbid reservoir and its phytoplankton in relation to photosynthesis and growth (mount bold reservoir, South Australia). *Journal of Plankton Research*, *10*(6), 1155–1177. https://doi.org/10.1093/plankt/10.6.1155

Oliver, R. L., Whittington, J., Lorenz, Z., & Webster, I. T. (2003). The influence of vertical mixing on the photoinhibition of variable chlorophyll a fluorescence and its inclusion in a model of phytoplankton photosynthesis. *Journal of Plankton Research*, *25*(9), 1107–1129. https://doi.org/10.1093/plankt/25.9.1107

Park, Y.-I., Chow, W. S., & Anderson, J. M. (1995). Light inactivation of functional photosystem II in leaves of peas grown in moderate light depends on photon exposure. *Planta*, *196*(3), 401–411. https://doi.org/10.1007/BF00203636

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., et al. (2013). The MAREDAT global database of high performance liquid chromatography marine pigment measurements. *Earth System Science Data*, *5*(1), 109–123. https://doi.org/10.5194/essd-5-109-2013

Porcar-Castell, A., Bäck, J., Juurola, E., & Hari, P. (2006). Dynamics of the energy flow through photosystem II underchanging light conditions: a model approach. *Functional Plant Biology*, *33*, 229–239. https://doi.org/10.1071/FP05133

Polimene, L., Brunet, C., Allen, J. I., Butenschön, M., White, D. A., & Llewellyn, C. A. (2012). Modelling xanthophyll photoprotective activity in phytoplankton. *Journal of Plankton Research*, *34*(3), 196–207. https://doi.org/10.1093/plankt/fbr102

Polimene, L., Brunet, C., Butenschön, M., Martinez-Vicente, V., Widdicombe, C., Torres, R., & Allen, J. I. (2014). Modelling a light-driven phytoplankton succession. *Journal of Plankton Research*, *36*(1), 214–229. https://doi.org/10.1093/plankt/fbt086

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raven, J. A., & Crawfurd, K. (2012). Environmental controls on coccolithophore calcification. *Marine Ecology Progress Series*, *470*, 137–166. https://doi.org/10.3354/meps09993

Ross, O. N., Geider, R. J., Berdalet, E., Artigas, M. L., & Piera, J. (2011). Modelling the effect of vertical mixing on bottle incubations for determining in situ phytoplankton dynamics. I. Growth rates. *Marine Ecology Progress Series*, *435*, 13–31. https://doi.org/10.3354/meps09193

Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., & Wolf-Gladrow, D. (2007). Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. *Biogeosciences*, *4*(4), 433–454. https://doi.org/10.5194/bg-4-433-2007

Schlüter, L., Møhlenberg, F., Havskum, H., & Larsen, S. (2000). The use of phytoplankton pigments for identifying and quantifying phytoplankton groups in coastal areas: testing the influence of light and nutrients on pigment/chlorophyll a ratios. *Marine Ecology Progress Series*, *192*, 49–63. https://doi.org/10.3354/meps192049

Schourup-Kristensen, V., Sidorenko, D., Wolf-Gladrow, D. A., & Völker, C. (2014). A skill assessment of the biogeochemical model REcoM2 coupled to the Finite Element Sea Ice – Ocean Model (FESOM1.3). *Geoscientific Model Development*, *7*(6), 2769–2802. https://doi.org/10.5194/gmd-7-2769-2014

Serôdio, J., Schmidt, W., & Frankenbach, S. (2017). A chlorophyll fluorescence-based method for the integrated characterization of the photophysiological response to light stress. *Journal of Experimental Botany*, *68*(5), erw492. https://doi.org/10.1093/jxb/erw492

Six, C., Finkel, Z. V., Irwin, A. J., & Campbell, D. A. (2007). Light variability illuminates niche-partitioning among marine picocyanobacteria. *PLoS ONE*, *2*(12), e1341. https://doi.org/10.1371/journal.pone.0001341

Six, C., Sherrard, R., Lionard, M., Roy, S., & Campbell, D. A. (2009). Photosystem II and pigment dynamics among ecotypes of the green alga Ostreococcus. *Plant Physiology*, *151*(1), 379–390. https://doi.org/10.1104/pp.109.140566

Soppa, M. A., Hirata, T., Silva, B., Dinter, T., Peeken, I., Wiegmann, S., & Bracher, A. (2014). Global retrieval of diatom abundance based on phytoplankton pigments and satellite data. *Remote Sensing*, *6*(10), 10,089–10,106. https://doi.org/10.3390/rs61010089

Stolte, W., Kraay, G. W., Noordeloos, A. A. M., & Riegman, R. (2000). Genetic and physiological variation in pigment composition of Emiliania huxleyi (Prymnesiophyceae) and the potential use of its pigment ratios as a quantitative physiological marker. *Journal of Phycology*, *36*(3), 529–539. https://doi.org/10.1046/j.1529-8817.2000.99158.x

Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A., et al. (2009). Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*, *76*(1–2), 4–15. https://doi.org/10.1016/j.jmarsys.2008.03.011

Strain, H. H., Manning, W. M., & Hardin, G. (1944). Xanthophylls and carotenes of diatoms, brown algae, dinoflagellates, and sea-anemones. *The Biological Bulletin*, *86*(3), 169–191. https://doi.org/10.2307/1538339

Tagliabue, A., Mtshali, T., Aumont, O., Bowie, A. R., Klunder, M. B., Roychoudhury, A. N., & Swart, S. (2012). A global compilation of dissolved iron measurements: Focus on distributions and processes in the Southern Ocean. *Biogeosciences*, *9*(6), 2333–2349. https://doi.org/10.5194/bg-9-2333-2012

Taylor, B. B., & Bracher, A. (2017). Pigment concentrations measured in surface water during SONNE cruise SO202/2 (TRANSBROM). *PANGAEA*. https://doi.org/10.1594/PANGAEA.880235

Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., et al. (2011a). Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance. *Biogeosciences*, *8*(12), 3609–3629. https://doi.org/10.5194/bg-8-3609-2011

Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., et al. (2011b). Pigments of phytoplankton during POLARSTERN cruise ANT-XXV/1. *PANGAEA*. https://doi.org/10.1594/PANGAEA.819070

Ting, C. S., & Owens, T. C. (1993). Photochemical and nonphotochemical fluorescence quenching processes in the diatom Phaeodacfyhm fricornutum. *Plant Physiology*, *101*(4), 1323–1330. https://doi.org/10.1104/pp.101.4.1323

Trees, C. C., Clark, D. K., Bidigare, R. R., Ondrusek, M. E., & Mueller, J. L. (2000). Accessory pigments versus chlorophyll a concentrations within the euphotic zone: A ubiquitous relationship. *Limnology and Oceanography*, *45*(5), 1130–1143. https://doi.org/10.4319/lo.2000.45.5.1130

Trimborn, S., Hoppe, C. J. M., Taylor, B. B., Bracher, A., & Hassler, C. S. (2015). Physiological characteristics of open ocean and coastal phytoplankton communities of Western Antarctic Peninsula and Drake Passage waters. *Deep Sea Research Part I: Oceanographic Research Papers*, *98*, 115–124. https://doi.org/10.1016/j.dsr.2014.12.010

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, *111*, C08005. https://doi.org/10.1029/2005JC003207

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, *116*(5), 882–892. https://doi.org/10.1111/j.2007.0030-1299.15559.x

Wagner, H., Jakob, T., & Wilhelm, C. (2006). Balancing the energy flow from captured light to biomass under fluctuating light conditions. *The New Phytologist*, *169*(1), 95–108. https://doi.org/10.1111/j.1469-8137.2005.01550.x

Webb, W. L., Newton, M., & Starr, D. (1974). Carbon dioxide exchange of Alnus rubra: a mathematical model. *Oecologia*, *17*(4), 281–291. https://doi.org/10.1007/BF00345747

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochem Cycles*, *22*, GB2024. https://doi.org/10.1029/2007GB003078

Wilson, A., Ajlani, G., Verbavatz, J.-M., Vass, I., Kerfeld, C. A., & Kirilovsky, D. (2006). A soluble carotenoid protein involved in phycobilisome-related energy dissipation in cyanobacteria. *Plant Cell*, *18*(April), 992–1007. https://doi.org/10.1105/tpc.105.040121.1981

Zindler, C., Bracher, A., Marandino, C. A., Taylor, B. J., Torrecilla, E., Kock, A., & Bange, H. W. (2013). Sulphur compounds, methane, and phytoplankton: Interactions along a north–south transit in the western Pacific Ocean. *Biogeosciences*, *10*(5), 3297–3311. https://doi.org/10.5194/bg-10-3297-2013

Zonneveld, C. (1998). Photoinhibition as affected by photoacclimation in phytoplankton: A model approach. *Journal of Theoretical Biology*, *193*(1), 115–123. https://doi.org/10.1006/jtbi.1998.0688

Appendix A

Extra Publications

Remote Sensing of Environment 240 (2020) 111704

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data

Hongyan Xi[a,*], Svetlana N. Losa[a,b], Antoine Mangin[c], Mariana A. Soppa[a], Philippe Garnesson[c], Julien Demaria[c], Yangyang Liu[a,d], Odile Hembise Fanton d'Andon[c], Astrid Bracher[a,e]

[a] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[b] Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
[c] ACRI-ST, 06904 Sophia Antipolis Cedex, France
[d] Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany
[e] Institute of Environmental Physics, University of Bremen, Bremen, Germany

ARTICLE INFO

Edited by Menghua Wang

Keywords:
Retrieval algorithm
Chlorophyll *a*
Phytoplankton functional types
Empirical orthogonal functions
Remote sensing reflectance
HPLC pigments
CMEMS GlobColour
OLCI

ABSTRACT

This study presents an algorithm for globally retrieving chlorophyll *a* (Chl-*a*) concentrations of phytoplankton functional types (PFTs) from multi-sensor merged ocean color (OC) products or Sentinel-3A (S3) Ocean and Land Color Instrument (OLCI) data from the GlobColour archive in the frame of the Copernicus Marine Environmental Monitoring Service (CMEMS). The retrieved PFTs include diatoms, haptophytes, dinoflagellates, green algae and prokaryotic phytoplankton. A previously proposed method to retrieve various phytoplankton pigments, based on empirical orthogonal functions (EOF), is investigated and adapted to retrieve Chl-a concentrations of multiple PFTs using extensive global data sets of in situ pigment measurements and matchups with satellite OC products. The performance of the EOF-based approach is assessed and cross-validated statistically. The retrieved PFTs are compared with those derived from diagnostic pigment analysis (DPA) based on in situ pigment measurements. Results show that the approach predicts well Chl-*a* concentrations of most of the mentioned PFTs. The performance of the approach is, however, less accurate for prokaryotes, possibly due to their general low variability and small concentration range resulting in a weak signal which is extracted from the reflectance data and corresponding EOF modes. As a demonstration of the approach utilization, the EOF-based fitted models based on satellite reflectance products at nine bands are applied to the monthly GlobColour merged products. Climatological characteristics of the PFTs are also evaluated based on ten years of merged products (2002 − 2012) through inter-comparisons with other existing satellite derived products on phytoplankton composition including phytoplankton size class (PSC), SynSenPFT, OC-PFT and PHYSAT. Inter-comparisons indicate that most PFTs retrieved by our study agree well with previous corresponding PFT/PSC products, except that prokaryotes show higher Chl-*a* concentration in low latitudes. PFT dominance derived from our products is in general well consistent with the PHYSAT product. A preliminary experiment of the retrieval algorithm using eleven OLCI bands is applied to monthly OLCI products, showing comparable PFT distributions with those from the merged products, though the matchup data for OLCI are limited both in number and coverage. This study is to ultimately deliver satellite global PFT products for long-term continuous observation, which will be updated timely with upcoming OC data, for a comprehensive understanding of the variability of phytoplankton composition structure at a global or regional scale.

1. Introduction

Over the past decades, satellite ocean color (OC) remote sensing has been widely used for estimating chlorophyll *a* (Chl-*a*) concentration, which is often used as an indicator of phytoplankton biomass. Beyond that, extracting information on phytoplankton community structure, e.g., phytoplankton functional types (PFTs), size classes (PSCs) and taxonomic composition, has become a research topic of priority, as it plays an important role in understanding the marine food web and aids the modelling associated with climate change impacts on

---

* Corresponding author.
E-mail address: Hongyan.Xi@awi.de (H. Xi).

https://doi.org/10.1016/j.rse.2020.111704
Received 4 July 2019; Received in revised form 7 January 2020; Accepted 3 February 2020

biogeochemical and ecological cycling of oceans (e.g., Falkowski et al., 1998; Le Quéré et al., 2005; IPCC, 2013; Bracher et al., 2017). In addition, accurate estimation on phytoplankton diversity and group distribution provides valuable information on identifying blooms caused by specific toxic algae, i.e., harmful algal blooms such as cyanobacterial blooms and red tides (e.g., Craig et al., 2006; Hu et al., 2010; Wang et al., 2017). A PFT is usually defined as a homologous set of "organisms related through common biogeochemical processes" such as silicification, calcification, nitrogen fixation, or dimethyl sulfide production, but are not necessarily phylogenetically affiliated (Falkowski et al., 2003; Litchman et al., 2006; IOCCG, 2014). However, as many phytoplankton groups which can be detected by remote sensing are also functional types, (e.g., diatoms are silicifiers, some cyanobacteria are nitrogen fixers, and coccolithophorids are calcifiers) (Bracher et al., 2017), these satellite proxies have been named PFTs for brevity (e.g., Losa et al., 2017).

Satellite OC remote sensing enables observation of phytoplankton over large areas or even at global scale. With previous (e.g., Sea-Viewing Wide Field-of-View Sensor – SeaWiFS and MEdium Resolution Imaging Spectrometer – MERIS) and current available OC satellites Moderate Resolution Imaging Spectroradiometer (MODIS), Visible Infrared Imaging Radiometer Suite (VIIRS), and especially the newly launched OLCI onboard Sentinel-3A (in February 2016) and 3B (in April 2018), a vast amount of quality controlled OC data are collected, allowing us to contribute to developing and/or improving methods and the corresponding applications to satellite data for estimating biogeochemical parameters in terms of global observation. There is a clear need to implement a sound PFT retrieval algorithm to the recent OLCI data, as well as to previous and current satellite OC time series data such as CMEMS GlobColour merged products (ACRI-ST GlobColour Team et al., 2017).

Different bio-optical and ecological algorithms have been developed to identify PFTs and phytoplankton taxonomic composition at the ocean surface, mainly based on phytoplankton abundance and inherent/apparent optical properties. Abundance-based approaches seek to establish empirical relationships between the PFTs and phytoplankton abundance or biomass, such as Chl-*a* concentration that can be retrieved from satellites (e.g., Uitz et al., 2006; Brewin et al., 2010, 2015; Hirata et al., 2011). Ecological-based approaches incorporate additional environmental parameters to identify ecological niches where particular phytoplankton communities may be found (Raitsos et al., 2008; Palacz et al., 2013). Efforts have also been made to combine abundance and ecological-based approaches (e.g. Brewin et al., 2015; Ward, 2015). Spectral-based approaches are more direct as they target known optical signatures and use satellite observed spectra to extract the signatures of specific PFT (e.g., Ciotti and Bricaud, 2006; Devred et al., 2006; Alvain et al., 2005, 2008; Hirata et al., 2008; Bracher et al., 2009; Kostadinov et al., 2009; Werdell et al., 2014; Brewin et al., 2015; Correa-Ramirez et al., 2018). These methods are mainly based on radiative transfer or bio-optical models and generally require high computation performance and adaptations for specific sensors. More complete reviews of these approaches are well detailed by the works of the IOCCG (2014), Bracher et al. (2017), and Mouw et al. (2017).

In this study, we seek to establish an approach that uses satellite reflectance data which inherit the information of various phytoplankton pigments and, therefore, allows retrieving the Chl-*a* concentrations of multiple PFTs. We choose the empirical orthogonal function (EOF) analysis, also known as principal component analysis, as it has been previously used for predicting ocean color metrics and various phytoplankton pigment concentrations by assessing variance of structures in spectral remote sensing reflectance (R_{rs}) or water leaving radiance (e.g., Lubac and Loisel, 2007; Craig et al., 2012; Taylor et al., 2013; Bracher et al., 2015; Soja-Woźniak et al., 2017). The spectral data are subject to EOF analysis to reduce the high dimensionality of the data and derive the dominant signals (EOF modes) that best describe the variance within the data set. Studies also proved that the EOF analysis could provide reliable retrievals even with limited number of data points (Craig et al., 2012; Bracher et al., 2015). Another advantage is that the models exhibited negligible loss of skill when applied to data sets with a reduced spectral resolution, which enables the applicability to the previous or currently existing multispectral OC sensors and future hyperspectral satellite missions such as PACE (Gregg and Rousseaux, 2017), HyspIRI (Lee et al., 2015) and EnMAP (Guanter et al., 2015).

Given that the EOFs derived from in situ or satellite hyper-/multispectral R_{rs} data have provided reliable retrievals of the concentrations of Chl-*a* and different pigments/pigment groups (Taylor et al., 2013; Bracher et al., 2015), we intend to present an implementation of the method proposed in Bracher et al. (2015) to retrieve PFTs instead of pigments, and to up-scale the application from regional to global scale by constructing large in situ data sets and multi-sensor OC products. Therefore, with the use of extensive in situ phytoplankton pigment data sets, satellite OC products, and matchups between in situ and satellite data, we propose an EOF-based global PFT retrieval approach by linking the variances in R_{rs} spectral structures to different PFTs. In the present study, we aim firstly to establish the EOF fitted model based on the nearly globally covered matchups between the satellite R_{rs} and the PFT Chl-*a* concentrations derived from diagnostic pigment analysis (DPA) of in situ HPLC pigment data, and cross-validate the performance of the EOF-based algorithm statistically; secondly, to set up the PFT retrieval scheme based on the EOF modes obtained from the matchups for the implementation to satellite OC products; thirdly, to investigate and evaluate the climatological characteristics of the PFTs retrieved from merged OC products (2002–2012) through inter-comparisons with other existing PFT/PSC products at the same period, and finally, to explore the potential of applying the approach to OLCI products based on a prediction scheme using a much more limited number of matchups.

2. Data and methods

2.1. Data sets

2.1.1. In situ databases of phytoplankton pigments

2.1.1.1. Pigment Database I (1997–2012). A large data set of the quality controlled near surface (first 12 m) HPLC phytoplankton pigments built for the ESA SynSenPFT Project (Bracher et al., 2016) was used for the extraction of the collocated R_{rs} spectra from satellite data. This HPLC pigment data set includes >15,000 sets of phytoplankton pigment data spanning 25 years from 1988 to 2012 covering the global ocean, collected from SEABASS, MAREDAT, LTER, BATS, AESOP-CSIRO, LOV and also from our own data published at PANGAEA (see Table 1 in Losa et al., 2017). Since SeaWiFS as an earlier OC sensor was launched in 1997, a subset for the period of 1997–2012 including 11,977 sets of pigment data was taken as Pigment Database I and used for the extraction of the R_{rs} matchups from GlobColour merged products. Yearly coverage of this matchup database spans from 3.2% (the least data points for 2012) to 9.3% (the most for 2004). 24.1%, 17.4%, 21.1%, and 37.4% of the data were collected during March–May, June–August, September–November, and December–February, respectively. Fig. 1(A) shows the spatial distribution of all the data points in this database in which all pigments are included, but only total chlorophyll *a* concentration (TChl-*a*, sum of monovinyl chlorophyll *a*, divinyl chlorophyll *a*, chlorophyll *a* allomers, chlorophyll *a* epimers, and chlorophyllide *a*) is present in the figure.

2.1.1.2. Pigment Database II (2016–2018). A relatively smaller ($n = 992$) phytoplankton pigment database of quality controlled near surface HPLC pigments was also built for the OLCI matchups from 2016 to 2018, involving our recently published data sets of HPLC based phytoplankton pigment concentrations collected mainly in late spring and summer from five cruises – Heincke462 in the North Sea

Table 1
Numbers of available R_{rs} matchups (1 × 1 pixel and averaged by 3 × 3 pixels) with different band combinations from the CMEMS GlobColour merged products. Bold highlights the matchups used in the EOF based algorithm. SeaW = SeaWiFS (1997–2010), MO = MODIS (2002–present), ME = MERIS (2002–2012), V = VIIRS/Suomi-NPP (2012–present). Note that with SeaWiFS included merged products, the bands from SeaWiFS only contributed until December 2010. Waveband centers for the four sensors were listed in Table S1 in the supplementary document.

Sensors	No. of matchups		Available wavebands (nm)												No. of bands
	1 × 1	3 × 3	412	443	490	510	531	547	551	555	560	620	670	678	
SeaW	1223	609	×	×	×	×				×			×		6
SeaW/ME	381	125	×	×	×	×				×	×	×	×		8
SeaW/MO/ME	766	516	×	×	×		×	×		×			×	×	8
SeaW/MO/ME	394	265	×	×	×	×	×	×		×			×	×	9
MO + V	25	27	×	×	×		×	×	×	×			×	×	9
SeaW/MO/ME	183	63	×	×	×	×	×	×		×	×	×	×	×	11
MO/ME/V	3	2	×	×	×	×	×	×	×	×	×	×	×	×	12

(April–May 2016): https://doi.pangaea.de/10.1594/PANGAEA.899043 (Bracher and Wiegmann, 2019), PS99 in the North Sea and the Fram Strait Arctic (June–July 2016): https://doi.pangaea.de/10.1594/PANGAEA.905502 (PS99.1) and https://doi.pangaea.de/10.1594/PANGAEA.898102 (PS99.2) (Liu et al., 2019a, 2019c), PS103 in the Southern Ocean: https://doi.pangaea.de/10.1594/PANGAEA.898941 (Bracher, 2019) (December 2016–January 2017), PS107 in the Fram Strait Arctic (July–August 2017): https://doi.pangaea.de/10.1594/PANGAEA.898100 (Liu et al., 2019b), and PS113 in the trans-Atlantic Ocean (May–June 2018): https://doi.pangaea.de/10.1594/PANGAEA.911061 (Bracher et al., 2020). Fig. 1(B) shows the locations of the data points from Pigment Database II (including all the pigments but with only TChl-*a* concentration present in the figure), which covers a large range of latitudes but focuses on the Atlantic Ocean only (60°W–20°E).

2.1.2. Satellite ocean color data

Satellite normalized remote sensing reflectance (R_{rs}) Level-3 (L3) products from multiple sensors were obtained from the CMEMS GlobColour data archive (http://www.globcolour.info/). The R_{rs} products used for matchup analysis included daily R_{rs} L3 products with 4-km resolution at the bands from either individual sensors (SeaWiFS, MODIS, MERIS, and VIIRS onboard Suomi-NPP) or the merged products

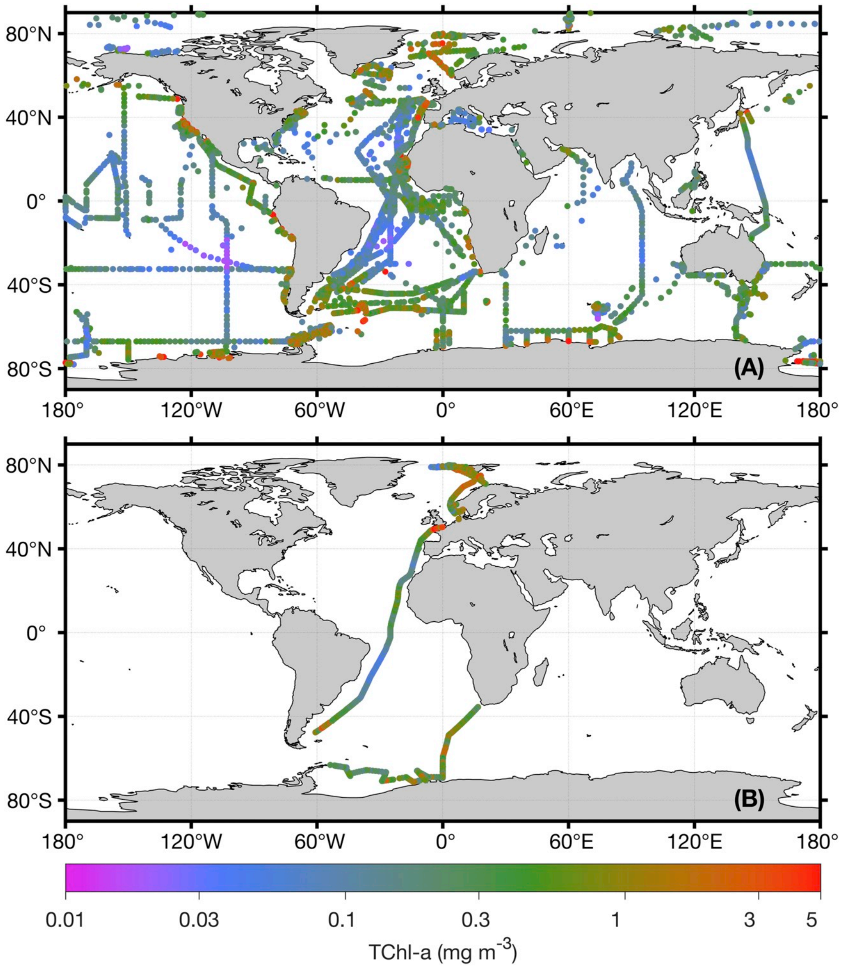


Fig. 1. Spatial distribution of the TChl-*a* concentration from the quality controlled in situ (A) Pigment Database I (1997–2012), and (B) Pigment Database II (2016–2018).

of two or more sensors. More details on the merged products are given in the GlobColour Product User Guide (ACRI-ST GlobColour Team et al., 2017). R_{rs} products from OLCI were not merged with any other sensor products and were therefore used separately for an OLCI only PFT retrieval scheme. Similar to the merged products, daily 4 km R_{rs} L3 products of OLCI were used for matchup extraction. In further application of the proposed approach to derive global long time series PFT products, monthly R_{rs} L3 products with 25 km spatial resolution from both, the merged products and OLCI data, were obtained for July 2002–April 2012 (time when SeaWiFS, MODIS and MERIS were in orbit, although SeaWiFS operation ended in late 2010 and then in late 2011 VIIRS was added), and April 2016–December 2018 (OLCI on Sentinel-3A in operation), respectively. In addition, the GlobColour merged ocean TChl-*a* monthly products with 25 km resolution in July 2002–April 2012 were also obtained for inter-comparison. The merged L3 TChl-*a* products were derived by a weighted average method (AVW) from single-sensor Level 2 chlorophyll products for case 1 waters (ACRI-ST GlobColour Team et al., 2017).

2.1.3. PFT retrieval input data

(A) PFT Chl-*a* concentrations derived from diagnostic pigment analysis (DPA)

Chl-*a* concentrations of PFTs were derived using an updated DPA method (Soppa et al., 2014; Losa et al., 2017). The DPA method was originally developed by Vidussi et al., 2001, adapted in Uitz et al. (2006) and further refined by Hirata et al. (2011) and Brewin et al. (2015). Basically, it relates the weighted sum of seven DPs (representative of individual PFTs) to TChl-*a* concentration, enabling us to determine the fraction of each PFT to the TChl-*a* thus to derive the PFT Chl-*a* concentrations. The partial coefficients of the DPs used in this study were derived from multiple linear regression using the data from the large global pigment data set as detailed in Table S1 of Supplementary Material in Losa et al. (2017) and were in good agreement with previous studies. The pigment concentrations of fucoxanthin, peridinin, 19′hexanoyloxy-fucoxanthin, 19′butanoyloxy-fucoxanthin, alloxanthin, chlorophyll *b*, zeaxanthin and divinyl chlorophyll *a* were used to derive the Chl-*a* concentrations of six PFTs in our study, that are, respectively, diatoms and dinoflagellates which are commonly considered as microphytoplankton, two types of nanophytoplankton – haptophytes and green algae (chlorophytes), and two picophytoplankton – prokaryotes, and *Prochlorococcus* which is a typical species of prokaryotes and commonly found in the subtropical region. PFT Chl-*a* concentrations <0.005 mg m^{-3} were excluded as such low values might contain much uncertainty. The rational for this threshold is that the surface Chl-*a* concentration encountered in the clearest ocean waters (South Pacific Gyre) was found to be in the range 0.01–0.02 mg m^{-3} (Morel et al., 2007). Therefore, values below 0.01 mg m^{-3} may be questionable. The corresponding PFT Chl-*a* concentration can be smaller. Considering the quality control on a large pigment data set as in Aiken et al. (2009), we chose the threshold of 0.005 mg m^{-3} for PFT Chl-*a* to minimize the influence of low accuracy in observations on the retrieval model, as it could bring much higher uncertainty to final prediction. The DPA derived PFT Chl-*a* concentrations for diatoms, haptophytes and prokaryotes from the pigment database I were published already in Losa et al. (2017) and are available from PANGAEA: https://doi.pangaea.de/10.1594/PANGAEA.875879 (Soppa et al., 2017).

(B) Matchups between in situ PFT and satellite R_{rs} data

Matchups to in situ PFT data were extracted from GlobColour global 4-km daily products for both merged and OLCI products. GlobColour "L3b" products with a sinusoidal projection were used so that each extracted pixel covers the same area. For each in situ measurement covered by a product, a matchup of 1 × 1 and 3 × 3 pixels around the in situ location was extracted. No specific quality filtering was applied at this stage because L3 products already exclude bad quality Level-2 pixels (ACRI-ST GlobColour Team et al., 2017). Averaged data based on 3 × 3 pixels were computed using the standard MERMAID tools (http://mermaid.acri.fr/) which follows the protocol from Bailey and Werdell (2006), in summary:

- only matchups containing at least 50% of valid pixels were kept;
- outlier pixels with (pixel value – median value) greater than $\pm 1.5 *$ standard deviation were removed;
- the matchups were removed if the coefficient of variation (CV) of the remaining pixels was higher than 0.15.

The same extraction and averaging protocol was used for merged and OLCI matchups. Based on the two HPLC pigment databases in Sect. 2.1.1, we have obtained the following matchups:

1) Matchups between daily merged Rrs products and in situ PFT data: the R_{rs} spectra at multispectral bands collocated with the PFT data derived from the Pigment Database I in Sect. 2.1.1 were extracted from the merged products (including SeaWiFS, MODIS, MERIS, VIIRS) from 1997 to 2012 archived in the GlobColour database. The extracted R_{rs} matchups included 1 × 1 pixel, and averaged R_{rs} by 3 × 3 pixels with the median and the standard deviation for each matchup. However, the same wavebands for R_{rs} data are not always available because different sensors have different spectral coverage at different periods (in addition to the exclusion of data with bad quality). Table 1 lists the numbers of matchups with different band combinations (from six to twelve bands) for R_{rs} matchups with 1 × 1 pixel and 3 × 3 pixels, respectively. Fig. 2 shows the corresponding geographical locations of 1 × 1 pixel matchups for R_{rs} at eight, nine and eleven bands, where the matchups were to some extent still globally distributed.
2) Matchups between daily OLCI R_{rs} and in situ PFT data: the Pigment Database II in Sect. 2.1.1 was used to derive the in situ PFT data and extract the corresponding OLCI R_{rs} matchups from 2016 to 2018. Table 2 lists the numbers of matchups with 10, 11 and 12 wavebands for R_{rs} data from S3A OLCI with 1 × 1 pixel and 3 × 3 pixels, respectively. Note that OLCI also includes the 709 nm and that OLCI itself does not have a band at 555 nm, but GlobColour database provides for MERIS and OLCI sensors the 555 nm through an inter-spectral conversion using:

$R_{rs}(555) = R_{rs}(560) * (1.02542–0.03757 * y - 0.00171 * y^2 + 0.0035 * y^3 + 0.00057 * y^4)$, where $y = \log10(\mathrm{CHL1})$ and CHL1 is the TChl-*a* concentration estimated by OC4 algorithm (ACRI-ST GlobColour Team et al., 2017). With this conversion, R_{rs} at 555 nm for OLCI were also included in our study.

2.2. Empirical orthogonal functions (EOF) based algorithm for PFT retrieval

2.2.1. EOF-based statistical approach

Following Bracher et al. (2015), each R_{rs} spectrum was firstly standardized by subtracting the mean spectral value and then divided by the spectral standard deviation (Taylor et al., 2013). The standardized data set of R_{rs}, denoted as matrix X (*M* observations × *N* wavelengths), was collocated to the respective DPA-based PFT data set C with *M* observations and 6 PFTs (*M* might be different for the six PFTs). As indicated in the model training box of Fig. 3, singular value decomposition (SVD) was applied to X for deriving the EOF modes:

$$X = U\Lambda V^T, \tag{1}$$

where matrix U ($M \times N$) contains column vectors of scores associated with EOF modes, matrix V ($N \times N$) contains the EOF loadings (spectral pattern), and Λ is an N × N matrix containing the singular values of X on the diagonal in decreasing order. For the PFT Chl-*a* prediction,

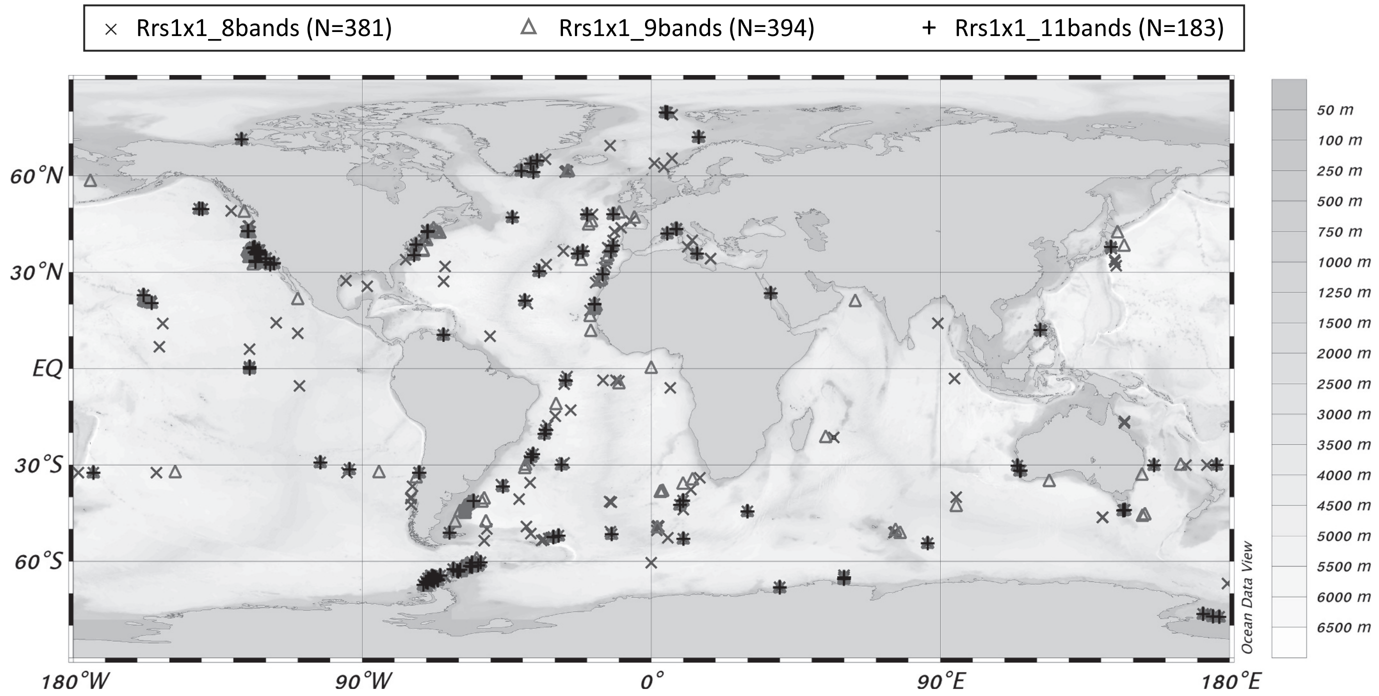


Fig. 2. Geographical locations of the single pixel matchups for merged R_{rs} at eight (in ×), nine (in △) and eleven bands (in +). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

generalized linear models (GLM) were created expressing the log-transformed Chl-*a* concentrations of each PFT, C_p, as a function of a subset of EOF scores (U). EOF modes with standard deviations (singular values from Λ) that are <0.0001 times the standard deviation of the first EOF mode were considered insignificant and thus omitted. The regression model for PFT prediction was expressed as:

$$\ln(C_p) = a_0 + a_1 u_1 + a_2 u_2 + \ldots a_n u_n, \tag{2}$$

where $u_{1,2,\ldots n}$ are the leading *n* EOFs from column vectors of U, a_0 is the intercept and $a_{1,2,\ldots n}$ are the regression coefficients. In addition, a stepwise routine was applied to search for smaller regression models, i.e., less *u* variables, through minimization of the Akaike information criterion (AIC). The significance of included terms was defined by the change in AIC (ΔAIC) with each term's removal.

2.2.2. Model assessment

We consider the coefficient of determination (R^2), the slope (*S*) and the intercept (*a*) of the GLM regression, which are based on the log-scaled predicted ($\ln(C_p)$) against the log-scaled observed ($\ln(C_o)$) PFT Chl-*a* concentration data, while the root-mean-square difference (RMSD), the median percent difference (MDPD), and the bias are based on the non-log-transformed data. Model performance statistics are expressed as:

$$R^2 = \frac{\sum_{i=1}^{M} (\ln(C_{pi}) - \ln(\overline{C_{oi}}))^2}{\sum_{i=1}^{M} (\ln(C_{oi}) - \ln(\overline{C_{oi}}))^2}, \tag{3}$$

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{M} (C_{pi} - C_{oi})^2}{M}}, \tag{4}$$

$$\text{MDPD} = \text{Median of}\left[\frac{|(C_{pi} - C_{oi})|}{C_{oi}} \times 100\right],\ i = 1, \ldots M, \tag{5}$$

$$\text{bias} = \frac{100}{M} \sum_{i=1}^{M} \frac{(C_{pi} - C_{oi})}{C_{oi}}, \tag{6}$$

where *M* is the number of observations in C_o, and $\overline{C_{oi}}$ is the mean of the observations, i.e., $\overline{C_{oi}} = \frac{1}{M}\sum_{i=1}^{M} C_{oi}$.

To test the robustness of the fitted model, cross-validation of the model fitting was carried out, similar to the procedure performed in Bracher et al. (2015). The collocated data were randomly split into two subsets, in which 80% of the data was used for model fitting/training, which included X^{train} (standardized R_{rs} spectra) and C^{train} (PFT Chl-*a* concentrations), and the rest 20% was used for prediction validation including X^{val} and C^{val}. The procedure was run for 500 permutations to eliminate the model uncertainty produced based on a spatially or temporally biased data set. For each permutation, with Eqs. (1)–(2) and the stepwise routine, a regression model was fitted between $\ln(C^{train})$ and U^{train}. The standardized validation set X^{val} was then projected onto the EOF loadings V^{train} and the inverse of singular values $\Lambda^{train-1}$ to derive their EOF scores U^{val}:

$$U^{val} = X^{val} \bullet V^{train} \bullet \Lambda^{train-1} \tag{7}$$

Table 2
Numbers of available OLCI R_{rs} matchups with 10, 11 and 12 wavebands.

Number of OLCI matchups			OLCI central bands (nm)												No. of bands
1 × 1	3 × 3	3 × 3 all[a]	400	412	443	490	510	555	560	620	665	674	681	709	
115	33	924		×	×	×	×	×	×	×	×	×	×		10
115	33	924	×	×	×	×	×	×	×	×	×	×	×		11
86	25	749	×	×	×	×	×	×	×	×	×	×	×	×	12

[a] 3 × 3 all: all available pixels in the 3 × 3 square were selected, but only matchup data with more than five out of nine pixels available were used.

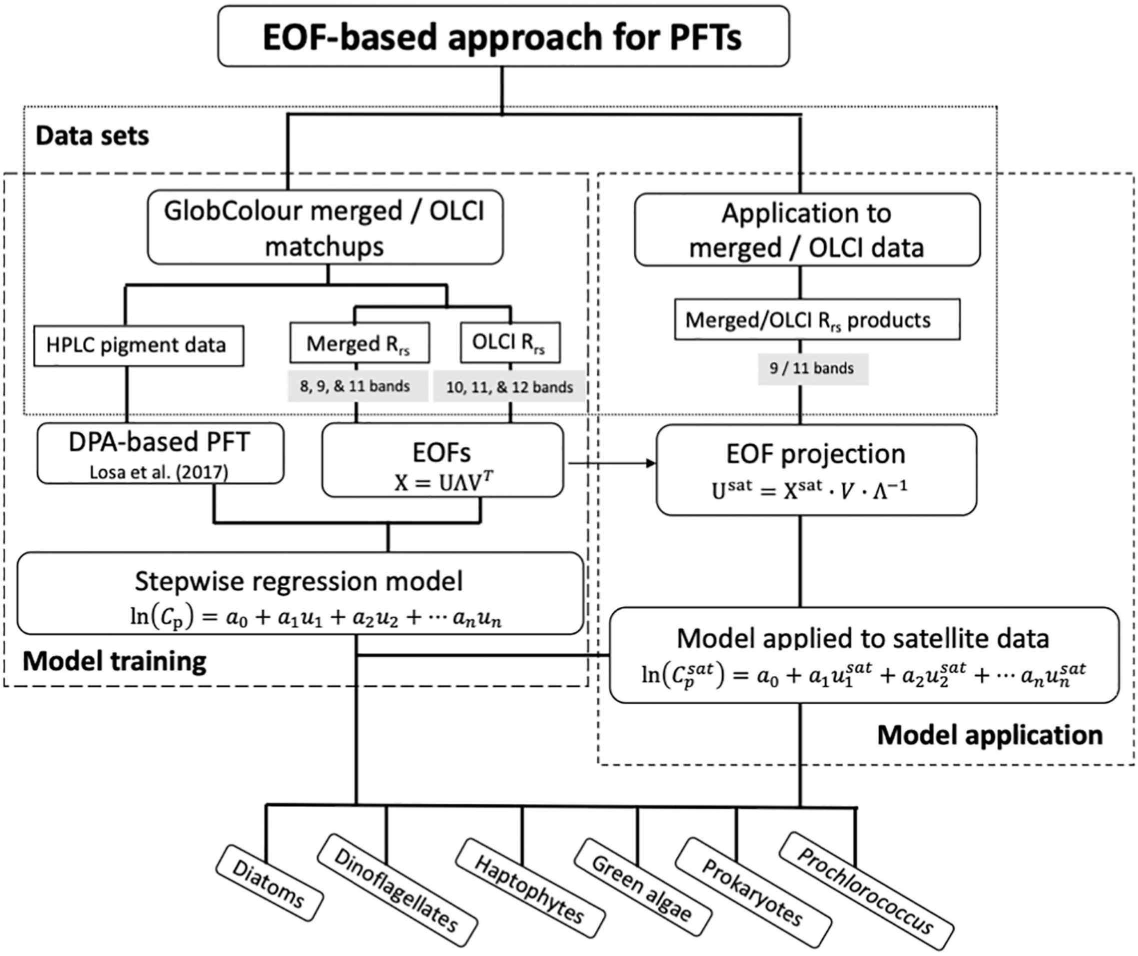


Fig. 3. Schematic flowchart of the EOF-based algorithm for predicting six PFTs with different input data sets. The left dashed-line box depicts the model training with the pigment-satellite matchup data and the right dashed-line box depicts the model application to satellite products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Lastly, the PFT Chl-*a* concentrations for the validation data set (C_p^{val}) were predicted using U^{val} of the selected EOF modes and the corresponding regression coefficients. The pairs of the observed and predicted PFT concentrations (C_o^{val} and C_p^{val}) of the 500 permutations were recorded for model assessment.

For each permutation, the R^2 for cross validation based on $\ln(C_p^{\text{val}})$ versus $\ln(C_o^{\text{val}})$ is determined, and the mean value of the R^2 from all permutations (R^2cv) is calculated. Similarly, other statistical parameters for cross validation are determined as follows by taking the mean values of the parameters from all permutations:

$$R^2cv = \frac{\sum_{i=1}^{M} (\ln(c_{pi}^{val}) - \ln(\overline{c_{oi}^{val}}))^2}{\sum_{i=1}^{M} (\ln(c_{oi}^{val}) - \ln(\overline{c_{oi}^{val}}))^2} \tag{8}$$

$$\text{RMSDcv} = \sqrt{\frac{\sum_{i=1}^{M} (C_{pi}^{val} - C_{oi}^{val})^2}{M}} \tag{9}$$

$$\text{MDPDcv} = \text{Median of}\left[\frac{|(C_{pi}^{val} - C_{oi}^{val})|}{C_{pi}^{val}} \times 100\right],\ i = 1, \ldots M \text{ (number of points for validation)} \tag{10}$$

2.2.3. PFT predictions from satellite data

As illustrated in Fig. 3 (model application part), we were able to apply the EOF analysis to satellite R_{rs} data listed in Sect. 2.1.2. Following Bracher et al. (2015), to predict PFTs globally using R_{rs} data from merged OC or OLCI products, for which we do not have corresponding pigment and PFT measurements, we projected standardized R_{rs} data from the satellite onto the EOF loadings (V) to derive a new set of EOF scores (U^{sat}), which was subsequently used for the prediction with the fitted model (see equations in model application of Fig. 3), where a_0 and $a_{1,2,\ldots n}$ were taken from the model developed with matchups from merged products or OLCI data as listed in Sect. 2.1.2.

2.3. PFT relative dominance

With the six retrieved PFTs in our study, we classified the relative PFT dominance in terms of Chl-*a* concentration on a global scale. The classification was performed simply based on the absolute values of the retrieved PFT Chl-*a* concentrations. For each set of the monthly PFT products, two steps were performed as follows. Step 1: the five PFTs – diatoms, dinoflagellates, haptophytes, green algae and prokaryotes – were compared pixelwise and the one with the highest Chl-*a* concentration was considered as the dominant PFT at this particular pixel. Since prokaryotes mainly contain *Prochlorococcus* and *Synechococcus*-like-cyanobacteria (SLC), Step 2 was performed to further assign the dominance of prokaryotes to either *Prochlorococcus*-dominated or SLC-dominated type. That is, for pixels where prokaryotes were the dominant group, we then compared the retrieved *Prochlorococcus* with prokaryotes – pixel with *Prochlorococcus* Chl-*a* concentration higher than 50% of that of the prokaryotes was defined as *Prochlorococcus* dominated, otherwise it was SLC dominated. With this straightforward classification we finally derived the dominance of diatoms,

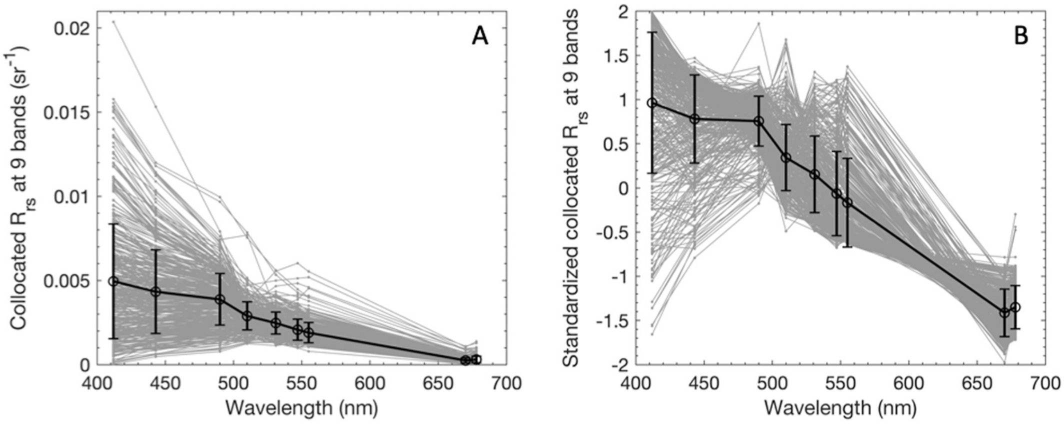


Fig. 4. (A) R_{rs} spectra at nine bands and (B) the corresponding standardized R_{rs} spectra from merged OC matchups at 1 × 1 pixel (in grey) with the mean spectra and standard deviation (black line with error bars).

dinoflagellates, haptophytes, green algae, *Prochlorococcus* and SLC from EOF-based PFTs.

3. Results and discussion

3.1. EOF analysis of Rrs data sets from GlobColour matchups

The matchups of satellite R_{rs} data sets highlighted in Table 1 with eight, nine and eleven bands (namely R_{rs_8}, R_{rs_9}, and R_{rs_11}) were taken as input data for the corresponding EOF analysis, respectively. The choice of the number of bands was based on previous positive experience with the eight MERIS bands (Bracher et al., 2015). In addition, it was tested if more spectral information would improve the retrieval results. As an example to illustrate satellite R_{rs} matchups, Fig. 4 shows the spectra of R_{rs_9} and the corresponding standardized spectra used in the EOF analysis. Most of the R_{rs} spectra presented quite typical spectral features of clean open ocean waters, i.e., high reflectance presented in blue band. However, our data set also contained cases of phytoplankton-rich waters with high reflectance in the green. With hyperspectral $R_{rs,}$ a few bio-optical features related to phytoplankton pigments and thus to PFTs can be caught only when they are prominent enough, such as phycocyanin (a marker pigment for cyanobacteria) which causes an obvious trough in 620–630 nm. While most spectral features in hyperspectral R_{rs} are often caused by a combined effect, e.g., both absorption and fluorescence peaks of phycoerythrin are located in green bands, where chlorophylls have the minimum absorption (Soja-Woźniak et al., 2017). With limited number of wavebands measured by multispectral sensors, it is even more challenging to identify directly the spectral features in terms of specific pigments of phytoplankton types.

As a statistical approach, EOF analysis on multispectral R_{rs} may not be able to catch the entire PFT absorption and scattering properties, but it provides information on to what extent the EOF modes (which have each their specific spectrum) are correlated to the PFTs. Following Sect. 2.2.1, the standardized R_{rs_8}, R_{rs_9}, and R_{rs_11} were decomposed by Eq. (1) into seven, eight, and ten EOF modes, respectively. As shown in Table 3, the first four modes already explain 99.51% to 99.71% of the total variance, with the first mode explaining 79.11%–82.51% of the total variance. Though previous studies (e.g., Craig et al., 2012; Bracher et al., 2015) have investigated the underlying bio-optical signature that the first several EOF modes may carry, it is still difficult to well define the distinct linkage between the EOF modes and the specific pigments or PFTs, as the significance level of the modes may change in different water types (Craig et al., 2012), and the PFT information cannot be the first-order reflected by the EOF modes derived from multispectral R_{rs} data. Nevertheless, a stepwise regression routine, via which the important modes to a certain PFT can be retained, was used to determine the PFT prediction models. Since the in situ PFT Chl-*a* concentrations derived from DPA are based on the marker pigments that were mostly identified in Bracher et al. (2015), we followed their study and included in the prediction model higher EOF modes. Though they contributed only a minute portion to the total R_{rs} variance, they might still inherit the optical signature by phytoplankton (partly group specific) pigments and therefore, be statistically significant for the prediction.

3.2. EOF-based algorithm for PFT retrievals

3.2.1. Stepwise regression procedure

As illustrated in Sect. 2.2.1, a stepwise routine was applied to determine the best EOF prediction model. The ΔAIC indicating the relative importance of the included terms (EOF modes) was presented in Table 4. For all three data sets, EOF-2 was the most important term in the respective models for TChl-*a* and Chl-*a* concentrations of most PFTs except for prokaryotes (also except for *Prochlorococcus* for R_{rs_11}). However, the second important EOF mode differed in PFT prediction models, and the total number of the EOF modes included in each model also varied. For instance, with data set R_{rs_9} only three EOFs were selected for *Prochlorococcus*, but all eight EOFs were included for haptophytes. It was also found that the most relevant EOF modes for prokaryotes and *Prochlorococcus* prediction were not fixed among the three R_{rs} data sets, indicating that the models are vulnerable and unstable, which was also reflected in their low performance (see Table 5 and Fig. 5). According to Bracher et al. (2015), EOF-2 is associated with Chl-*a*; the high importance of EOF-2 in the PFTs is likely due to the

Table 3
Percentage of total variance explained (%) by the decomposed EOF modes derived from three satellite matchup data sets R_{rs_8}, R_{rs_9}, and R_{rs_11} within the 1 × 1 pixel.

% of variance	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7	EOF-8	EOF-9	EOF-10
R_{rs_8} 1 × 1	82.51	14.78	2.14	0.28	0.18	0.08	0.02			
R_{rs_9} 1 × 1	79.11	17.75	2.03	0.79	0.22	0.06	0.03	0.01		
R_{rs_11} 1 × 1	79.28	17.60	1.76	0.87	0.25	0.13	0.05	0.05	0.01	0.01

Table 4
ΔAIC for the predictions of the TChl-*a* and six PFT Chl-*a* concentration by the EOF modes based on R_{rs} 1 × 1 matchups with eight, nine and eleven bands from merged OC products (R_{rs_8} 1 × 1, R_{rs_9} 1 × 1, and R_{rs_11} 1 × 1). Bold highlights the EOF mode with the highest ΔAIC for TChl-*a* and each derived PFT.

R_{rs_8} 1 × 1	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7
TChl-*a*	16.02	283.25	105.43	24.82			2.48
Diatom	8.16	130.24	90.83	10.53		0.89	
Haptophytes	42.34	214.50	4.57	24.04		1.45	
Prokaryotes	12.52						5.49
Dinoflagellates	5.69	122.46	54.56	0.41			
Green algae	1.14	92.25	8.05	1.49		9.25	
Prochlorococcus		7.29				6.87	0.73

R_{rs_9} 1 × 1	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7	EOF-8
TChl-*a*	38.27	416.17	109.26	58.11		3.13	10.07	
Diatom	20.05	217.09	80.52	30.17		9.43	7.07	1.14
Haptophytes	41.31	266.08	1.32	7.33	1.89	4.64	4.1	7.45
Prokaryotes	16.71		7.32	0.63	3.24	22.24	10.93	2.84
Dinoflagellates	4.85	177.95	27.59	24.62				7.14
Green algae		173.91	2.59		2.29	7.43	4.46	
Prochlorococcus		20.63				12.66	1.97	

R_{rs_11} 1 × 1	EOF-1	EOF-2	EOF-3	EOF-4	EOF-5	EOF-6	EOF-7	EOF-8	EOF-9
TChl-*a*	13.34	181.37	48.59	6.66	1.94				
Diatom	7.86	105.23	44.49		0.32	3.41			
Haptophytes	25.35	123.10		0.58	0.82	0.55		6.38	1.32
Prokaryotes	9.45				3.15	6.86	0.55	4.52	
Dinoflagellates	10.32	86.57	8.95	5.10					2.03
Green algae		102.48			1.73		8.36	1.82	0.39
Prochlorococcus	9.30			0.06	0.65			10.87	

elevation of Chl-*a* concentration in most of the PFTs when TChl-*a* increases. Since prokaryotes and *Prochlorococcus* mainly dominate in oligotrophic regions with very low biomass concentration, they do not have a collinearity in their Chl-*a* concentration with TChl-*a* as most other PFTs. A similar statement was also given in Bracher et al. (2015) for predicting pigments.

3.2.2. Performance of retrieval models based on matchups of merged R_{rs} data sets

Satellite PFT Chl-*a* and TChl-*a* concentrations were predicted with the regression models built based on the EOF scores derived from the R_{rs} data sets and the in situ PFT Chl-*a* concentrations. Matchups at different band settings and pixel level (1 × 1, 3 × 3 pixels) were taken as input for comparison between the results from different band

Table 5
Statistics of regression models for TChl-*a* and six PFT Chl-*a* concentrations using EOF modes based on R_{rs} matchups R_{rs_8}, R_{rs_9}, and R_{rs_11} within 1 × 1 pixel from merged products. Cross-validation (cv) results are presented with 500 permutations for data splitting into 80% of the data used for training and 20% for validation. N = number of valid matchups for each parameter.

	N	MDPD (%)	RMSD (mg m^{-3})	R^2	MDPDcv (%)	RMSDcv (mg m^{-3})	R^2cv
R_{rs_8} 1 × 1							
TChl-*a*	381	40.66	1.38	0.72	40.97	1.40	0.71
Diatoms	286	80.28	1.25	0.59	81.56	1.27	0.58
Haptophytes	366	57.16	0.30	0.58	57.97	0.30	0.54
Prokaryotes	348	62.32	0.15	0.05	62.95	0.14	0.04
Dinoflagellates	258	59.14	0.91	0.56	60.52	0.64	0.54
Green algae	239	60.51	0.12	0.50	61.81	0.12	0.47
Prochlorococcus	139	41.92	0.03	0.13	42.77	0.03	0.08
R_{rs_9} 1 × 1							
TChl-*a*	394	37.41	1.24	0.76	37.08	1.27	0.75
Diatoms	306	73.70	1.21	0.65	74.74	1.29	0.63
Haptophytes	387	47.16	0.22	0.64	48.62	0.24	0.61
Prokaryotes	367	53.70	0.13	0.15	55.08	0.13	0.11
Dinoflagellates	272	55.32	0.93	0.62	57.29	0.72	0.59
Green algae	262	55.81	0.11	0.51	56.26	0.11	0.48
Prochlorococcus	142	39.65	0.02	0.24	42.68	0.02	0.18
R_{rs_11} 1 × 1							
TChl-*a*	183	38.15	1.42	0.75	40.20	1.43	0.73
Diatoms	148	75.56	1.26	0.68	77.42	1.28	0.64
Haptophytes	179	53.04	0.28	0.61	55.84	0.29	0.54
Prokaryotes	171	61.41	0.17	0.13	62.61	0.16	0.08
Dinoflagellates	132	64.32	1.20	0.56	66.75	0.83	0.51
Green algae	116	54.52	0.12	0.60	58.60	0.13	0.48
Prochlorococcus	52	41.83	0.02	0.35	50.60	0.03	0.14

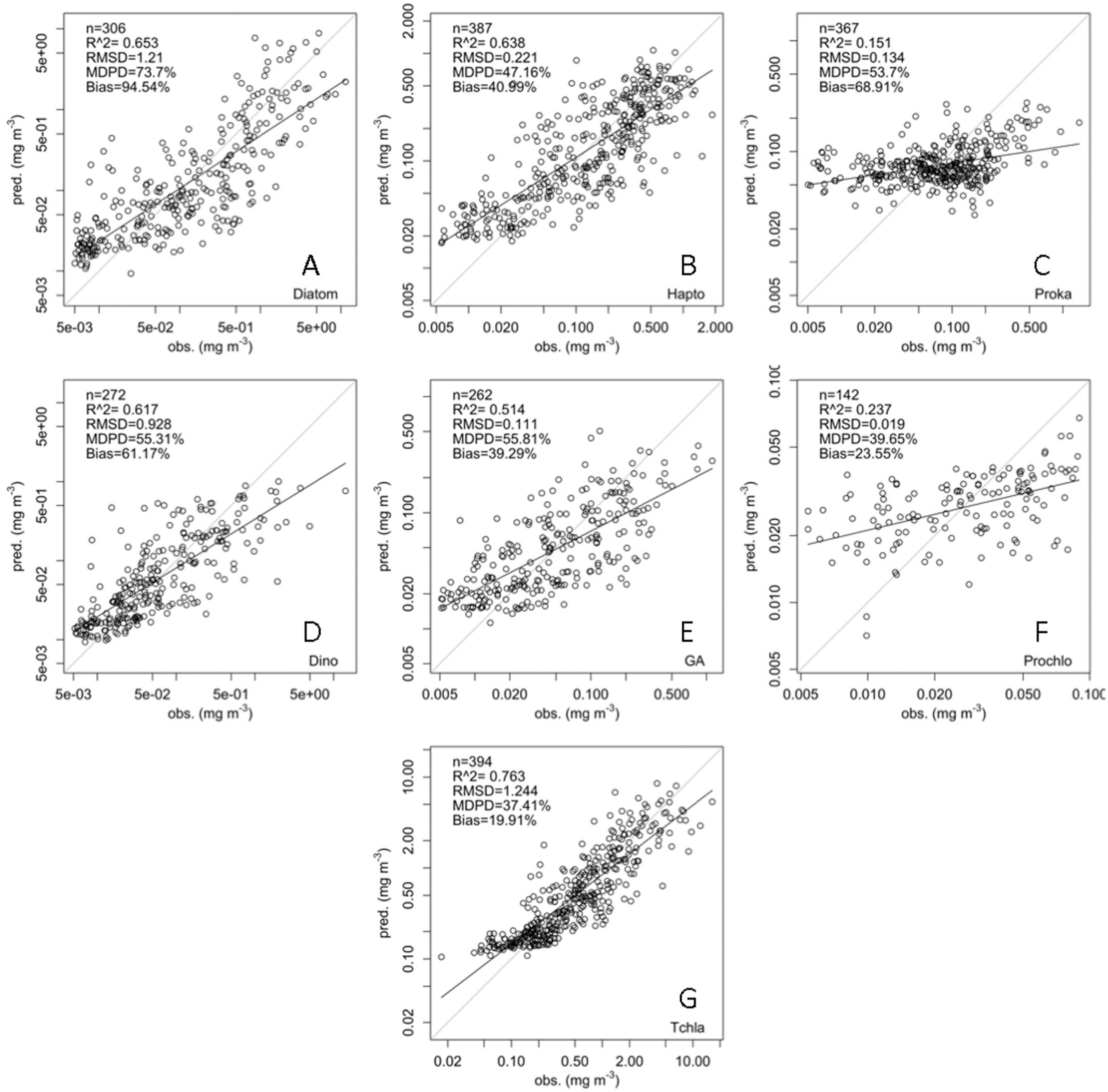


Fig. 5. Regressions between observed (x-axis, obs.) and predicted (y-axis, pred.) Chl-*a* concentrations of (A) diatoms, (B) haptophytes, (C) prokaryotes, (D) dinoflagellates, (E) green algae, (F) *Prochlorococcus*, and (G) TChl-*a* using EOF modes derived from merged R_{rs} products at 9 bands (1 × 1 pixel).

numbers, pixels and data points. Prediction model performances of using R_{rs} data sets with 1 × 1 and 3 × 3 matchups were statistically similar. Therefore, here we only presented and discussed in detail the results of the 1 × 1 pixel matchups, as there were more collocated data which should provide more robust predictions (statistics based on R_{rs} 3 × 3 data sets are presented in Table S2 in the supplementary document). The prediction models developed from the 1 × 1 collocated R_{rs} data sets were also later applied to the satellite products.

Statistics of the EOF-based regression models are listed in Table 5 for different R_{rs} data sets (R_{rs_8}, R_{rs_9} and R_{rs_11}). The predicted PFT Chl-*a* concentrations display slight differences between different band settings of the input R_{rs}. With all three data sets, the predicted and observed (based on in situ data) TChl-*a* and Chl-*a* concentrations for diatoms, haptophytes, dinoflagellates and green algae are well correlated, with $R^2 \geq 0.50$ and $R^2cv \geq 0.47$. TChl-*a* has the highest correlation ($R^2 \geq 0.72$), while Prokaryotes and *Prochlorococcus* have the weakest correlation between the predicted and observed concentrations but are generally better correlated using data set R_{rs_9} compared to the other two data sets. The MDPD are lowest for TChl-*a* and *Prochlorococcus* (< 42%) and low for haptophytes, dinoflagellates, green algae and prokaryotes (< 60% for data set R_{rs}_9). The highest MDPD was found for diatoms (< 80%). The MDPDcv of all cases are slightly higher but still comparable with the MDPD, indicating that the prediction models are stabilized. R_{rs_9} presents an overall lowest MDPD among the three data sets. RMSD values were calculated in non-log transformed manner, and thus vary depending on the corresponding

Chl-*a* concentration ranges of individual PFTs. TChl-*a* has the highest RMSD as it is the indicator of all phytoplankton biomass, whereas Chl-*a* of *Prochlorococcus* which is always low in concentration has the lowest RMSD. Among the three data sets, the lowest RMSD are found for R_{rs_9}. Hence, we conclude that the EOF-based models with R_{rs} at nine bands (see Table 1) perform best and slightly better than those with eleven bands, while the weakest are the models based on eight bands. This to some extent indicates that the performance of prediction models is not only subject to the number of bands (i.e., the more bands the better), but also to the number of matchups (with R_{rs_11} the least).

As a summary, Fig. 5 shows the observed against the predicted TChl-*a* and Chl-*a* concentrations for the six PFTs by the EOF-based method using R_{rs_9}. Corresponding to the statistics in Table 5, TChl-*a* and Chl-*a* of diatoms, haptophytes, dinoflagellates, and green algae which have relatively larger ranges in magnitude show relatively good predictions, with regression lines close to the 1:1 reference line and lower intercepts. Prokaryotes and *Prochlorococcus* are of weaker correlations with slopes much lower than 1 and higher intercepts, mainly due to their low concentrations, the narrow range of the variation, as well as the low variability in the concentrations especially for prokaryotes that could not be well interpreted by the EOF modes. Slopes of all regression lines <1 indicate that the models to some extent overestimate the variables in low concentrations and underestimate them in higher concentrations. Slopes of <1 were also shown in Bracher et al. (2015) for all the predictions of pigments and pigment composition, though in their study the prediction performance for some important pigments was statistically better compared to our prediction of PFT Chl-*a* concentration. Among the well predicted pigments in Bracher et al. (2015), zeaxanthin, typically used as a marker pigment for prokaryotes, showed the lowest correlation but reasonable MDPD, which corresponds to our lower R^2 values for prokaryotic phytoplankton. It is worth investigating further the prediction models and perform certain tuning procedure through mathematical methods to reduce these over- or underestimations, especially for picophytoplankton which are usually very low in concentration.

The cross-validation procedure effectively examined the robustness of the prediction models. The statistical parameters for cross-validation (averaged for all 500 permutations with 20% data for prediction) were nearly or as equivalently good as the statistics for the model trained with the whole data set (Table 5). This suggests that the number of data points (matchups) is adequate for a robust model establishment. In fact, in our study there were 52–394 data points for all matchups with different band settings, which is much higher than that was suggested to be necessary for robust model development by Craig et al. (2012) (15 points at a seasonal cycle) and Bracher et al. (2015) (50 points). However, since their studies were rather regional while we are focusing on the global scale, a higher number of points is expected in our study to enable a comprehensive coverage of the global ocean water types. From Table 5 one can see that the statistics of the cross validation are much worse than the original statistics for the green algae and *Prochlorococcus* Chl-*a* predictions using the data set R_{rs_11}, for which less available matchups were obtained. Therefore, though lower R^2 and higher MDPD were obtained with the data set R_{rs_9}, for these two PFTs, the cross-validation showed better results than that from the data set R_{rs_11}, convincing us the nine-band setting of the R_{rs} to be optimal for PFT model applications to satellite products without in situ matchups.

To better understand the performance of the EOF-based algorithm, Fig. S1 in the supplementary document shows the uncertainty for different ocean biomes in the algorithm derived Chl-*a* concentrations of the six PFTs using GlobColour merged R_{rs} at nine bands (global projection of the uncertainty is detailed in the supplementary document). Diatoms show underestimation in coastal regions (mean deviation of -0.11 mg m^{-3} in this biome), slight underestimation in high latitudes and near the equator (~ -0.02 mg m^{-3}), and very slight overestimation in the subtropical regions ($\sim$0.013 mg m^{-3}). Haptophytes, dinoflagellates, and green algae present similar uncertainty distributions, i.e., overestimation in higher than 40°N and subtropical regions and underestimation near the equator and in the Southern Ocean, but with different amplitudes. Both prokaryotes and *Prochlorococcus* show distinct overestimation in the central part of the oligotrophic gyres (0.026 and 0.014 mg m^{-3}, respectively) but underestimation in the surrounding areas of the gyres (-0.06 and $-$ 0.012 mg m^{-3}, respectively).

3.2.3. Application to merged products for global PFT retrieval

Given that the EOF-based PFT models based on the matchups of merged R_{rs} at nine bands show the best performance, we applied these models (based on the full data set fit) to the merged R_{rs} global products at the same nine bands for the period of 2002–2012. Selection criterion of the nine bands from merged R_{rs} products is detailed in Sect. 3.2.2 of the supplementary document. The numerical matrices and regression coefficients determined by Eqs. (1) and (2) used for the model implementation to the merged R_{rs} products at nine bands are also explained and provided in Tables S3 and S4 in the supplementary document.

Fig. 6 illustrates the global mean distribution Chl-*a* concentration of each PFT, based on the monthly PFT products derived from the merged R_{rs} products with 25 km resolution from 2002 to 2012. Diatom Chl-*a* concentrations are generally higher in high latitudes, marginal seas and coastal upwelling regions but are much lower in the tropical regions and extremely low in the subtropical gyres. The typical diatom abundant regions are higher than 40°N (North Atlantic, Bering Sea and Labrador Sea up to the Arctic Ocean), the Patagonian upwelling and most part of the Southern Ocean. The average Chl-*a* concentration of diatoms over the globe is $\sim$0.08 mg m^{-3}. Chl-*a* concentration of dinoflagellates is low nearly over the whole globe ($\sim$0.02 mg m^{-3}) but higher in the Arctic Ocean and Patagonian upwelling. Haptophytes with a global average Chl-*a* of 0.09 mg m^{-3} follow in distributions of the diatoms but have more spread regions of high Chl-*a* in the high latitudes, waters near the coasts, and equatorial regions (such as the west coast of Africa). Chl-*a* concentration of green algae (global average of 0.03 mg m^{-3}) is found typically higher in the Arctic and the near coast oceans around the southern part of South America. Prokaryotes and *Prochlorococcus* show distinctly different distribution features from the other four PFTs. Prokaryotes with a global average Chl-*a* concentration of 0.07 mg m^{-3} are much more abundant in the subtropical regions but also substantially contribute ($\sim$5–30% of TChl-*a*) in the Arctic Ocean. Waters such as the Baltic Sea, the east coast of China, and the west coast of Africa (around 5°S and 10–20°N) show very prominent abundance of prokaryotes. *Prochlorococcus* are generally very low on a global scale (global average 0.03 mg m^{-3}), especially in high latitude waters (not really detectable), slightly higher in subtropical regions and apparently abundant in some parts of the west coast of Africa similar to prokaryotes. Distribution of *Prochlorococcus* is supported by previous findings (Flombaum et al., 2013). Their quantitative model based on a large number of observations well defined the assessment of the *Prochlorococcus* abundance and the results match well our retrievals. In general the global average Chl-*a* concentrations of the PFTs retrieved from our study are consistent with those from Hirata et al. (2011), except that prokaryotes Chl-*a* is higher (0.07 mg m^{-3} in our study versus 0.04 mg m^{-3} from Hirata et al., 2011), mainly due to our elevated Chl-*a* prediction in the subtropics for prokaryotes. To illustrate the changes in the PFT Chl-*a* distribution with seasons, the monthly climatological products of each PFT are provided in Figs. S2-S7 in the supplementary document. For instance, diatom blooms are mainly detected during early summer in the Southern Ocean (December–January) and in the subarctic and Arctic waters (May–June). Haptophytes show similar seasonal changes in high latitudes as diatoms, but highly increase during the summer season in the equatorial Atlantic. A strong prokaryotic enhancement is also found during July–August at the west coast of South Africa.

Distribution of TChl-*a* retrieved by the EOF-based algorithm is

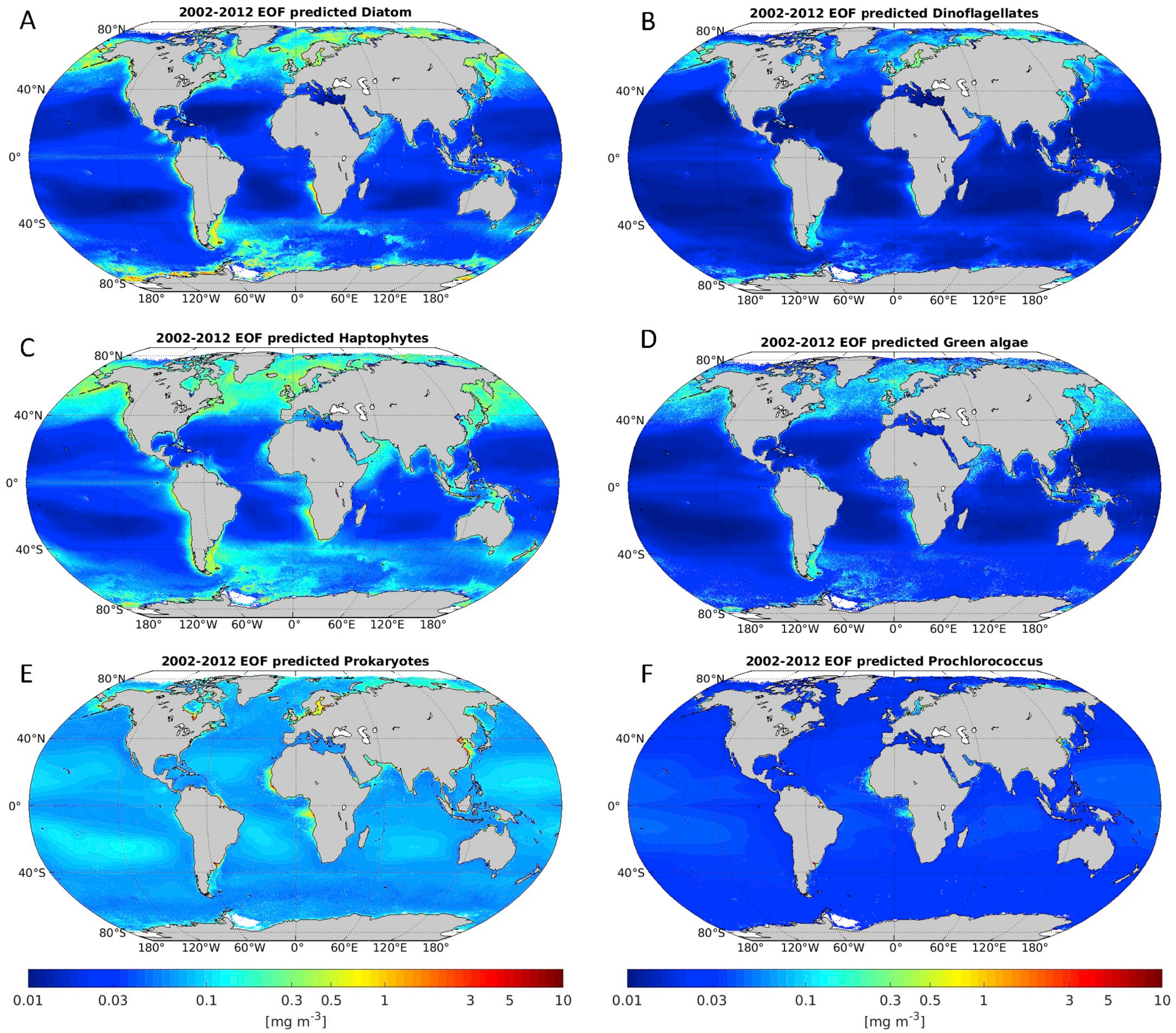


Fig. 6. Ten-year mean distribution (July 2002–April 2012) of the PFT Chl-*a* concentration for (A) diatoms, (B) dinoflagellates, (C) haptophytes, (D) green algae, (E) prokaryotes, and (F) *Prochlorococcus* retrieved by EOF-based algorithm from merged monthly R_{rs} products at nine bands.

presented in comparison to the GlobColour merged ocean chlorophyll products (mean over all years in Fig. 7, and climatological monthly mean in Figs. S8–S9). The ten-year mean of our EOF-based predicted TChl-*a* is generally in good agreement considering the distribution patterns with the standard products, though it is clearly seen that the EOF-based TChl-*a* shows higher/lower values in the subtropical gyres/coastal waters than the standard products. This was however expected, as the EOF-based retrieval models based on matchups already showed an over-/under-estimation for lower/higher values for all the retrieved variables/PFTs, as illustrated in Fig. 5. This flattening effect of the prediction is most prominent in prokaryotes and *Prochlorococcus*, of which the EOF-based models present the weakest correlation. An accurate retrieval of prokaryotic phytoplankton or its corresponding marker pigments (zeaxanthin, divinyl Chl-*a*) has always been a challenge so far (e.g., Bracher et al., 2015; Losa et al., 2017), as the picophytoplankton Chl-*a* concentrations are usually globally very low, even when dominating in oligotrophic oceans. This results in a narrow variation range and low variability in their concentrations compared to other PFTs, and also in a weak imprint on the spectral shape which are limited for the detection via the spectral analysis. An exception is that in the Baltic Sea prokaryotes can have high Chl-*a* concentrations especially during blooms. This is also reflected in our retrievals, though there are no matchups available included in the EOF analysis.

3.3. *Evaluation of the EOF-based PFT products*

3.3.1. *Inter-comparison with other PFT/PSC products*

To evaluate our retrieval algorithm, the derived Chl-*a* concentrations of diatoms, haptophytes and prokaryotes were compared with SynSenPFT Chl-*a* of diatoms, coccolithophores and cyanobacteria (Losa et al., 2017) and Chl-*a* of three PSCs (micro- > 20 µm, nano- 2–20 µm, and picophytoplankton < 2 µm, Sieburth et al., 1978) obtained with the PSC model of Brewin et al. (2010, 2015). Both SynSenPFT and PSC products developed within the frame of the SynSenPFT project (Losa et al., 2017) were available globally at 4 km daily resolution over the period from 2002 to 2012. Prior to the inter-comparison, both products

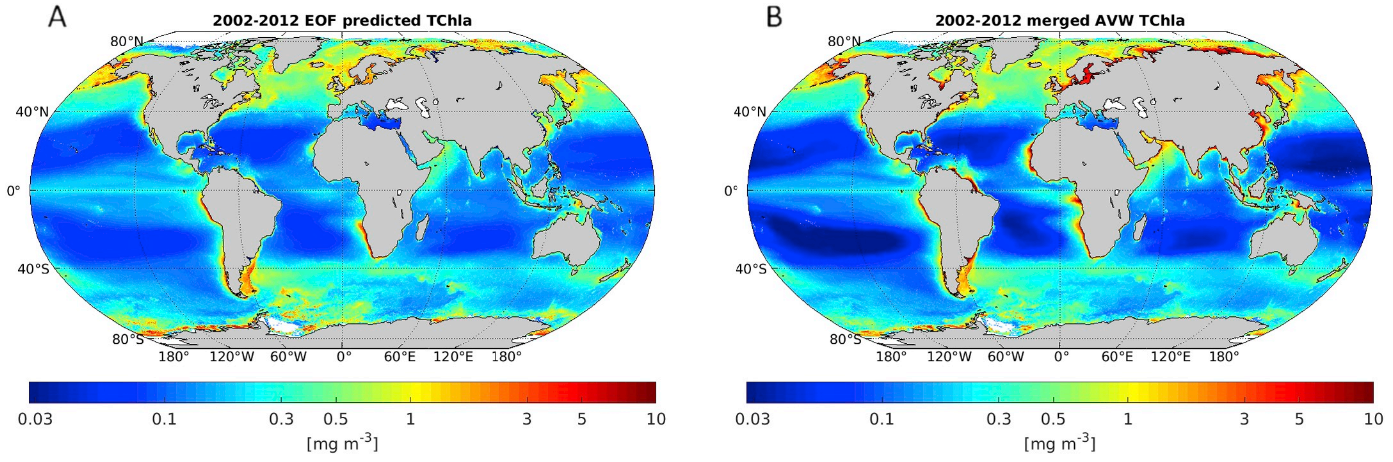


Fig. 7. Ten-year mean distribution (July 2002–April 2012) of (A) TChl-*a* concentration retrieved by EOF-based algorithm from merged monthly R_{rs} products at nine bands and (B) GlobColour AVW merged TChl-*a* concentration based on open ocean L2 chlorophyll products from SeaWIFS, MODIS and MERIS sensors.

were binned to monthly averages and re-gridded to 25 km resolution, to be consistent with our EOF-based PFT products. For simplification, in the following text the SynSenPFT derived Chl-*a* concentrations of diatoms, coccolithophores, and cyanobacteria are denoted as dia-SynSenPFT, coc-SynSenPFT, and cya-SynSenPFT, respectively; the Chl-*a* concentrations of micro-, nano- and picophytoplankton derived from PSC model are denoted as c-micro, c-nano, and c-pico, respectively. The EOF-based Chl-*a* products of the other two PFTs, green algae and *Prochlorococcus* were compared to those derived by OC-PFT method proposed by Hirata et al. (2011) using GlobColour AVW merged TChl-*a* monthly 25-km products as input for the same period (2002–2012). Dinoflagellates were not considered for comparison as the OC-PFT derived dinoflagellates showed very poor validation result (Hirata et al., 2011). It is noteworthy that OC-PFT also allows the retrieval of Chl-*a* concentrations of diatoms, haptophytes and prokaryotes, but as they are intrinsic in the SynSenPFT products (Losa et al., 2017) they were not included as separate products for the inter-comparison.

Following Losa et al. (2017), the time-latitude Hovmöller diagrams were generated covering the monthly means from 2002 to 2012 of the different PFT/PSC products. Since globally the Chl-*a* concentration is typically log-normally distributed (Campbell, 1995), all averaging was done in logarithmic space and then back-transformed to the original scale. The Hovmöller diagrams are presented in Figs. 8–11, where the left side of each subplot shows the monthly variation during the ten-year period (2002–2012), and the right side shows the climatological annual cycle. Since different studies tend to provide different retrieval information in terms of phytoplankton composition, the optimal way for the inter-comparison is to select the variables carrying the most similar PFT information, but one has to keep in mind that the products compared here are not always representing exactly the same quantities.

Diatoms derived from our study (Fig. 8A) and dia-SynSenPFT (Fig. 8B) show similar distributions with both lowest diatom Chl-*a* concentration in the subtropical regions especially in the gyres and higher concentration in high latitudes. Compared to dia-SynSenPFT, the EOF-based diatoms show generally lower Chl-*a* in the polar and tropical regions, however they indicate the same blooming periods for diatoms in May–June in the Arctic and December–January in the Southern Ocean. Dia-SynSenPFT presented distinct higher Chl-*a* from 10°S to 10°N during December to February 2005–2006, 2007–2008 and 2010–2011 than other years, whereas the change between the years is not evident in either our results or the c-micro products (Fig. 8C). Since microphytoplankton contain not only diatoms but also other micro-size phytoplankton such as dinoflagellates, the sum of EOF-based diatoms and dinoflagellates was also shown (Fig. 8D), presenting similar seasonal variation to c-micro but higher/lower Chl-*a* in the gyres/high latitudes.

Before comparing the EOF-based haptophytes to other products, it should be noted that coccolithophores are a main contributing PFT to haptophytes, while haptophytes are a part of nanophytoplankton, with the latter containing also *Phaeocystis*, cryptophytes, and a few other groups. Haptophytes derived from our study (Fig. 9A) are well consistent with coc-SynSenPFT (Fig. 9B), although again our retrievals show a relatively mild pattern with lower Chl-*a* in high latitudes and the 10°S-10°N equator belt. Chl-*a* concentration of coc-SynSenPFT from 10°N to 40°N during the summer time is significantly higher, but this pattern is not found in either our products or c-nano. Our haptophytes present similar distribution with c-nano (Fig. 9c) but lower Chl-*a* in the high latitudes and equatorial regions as expected. The climatological annual cycles of both are in very good agreement in the Southern Ocean, while in the Arctic c-nano shows much Chl-*a* enhancement in May–July. In addition, c-nano spreads more to the north until 25°N from the equator. However, caution should be taken since our DPA derived haptophytes contain only their nanophytoplankton fraction while their picophytoplankton fraction is neglected, whereas Brewin et al. (2015) consider part of the haptophytes in the picophytoplankton group when TChl-*a* is below 0.08 mg m^{-3}.

The overall Chl-*a* concentration of our EOF-based prokaryotes (Fig. 10A) is generally low (0.03–0.20 mg m^{-3}), but higher concentrations are found in the subtropical regions, only slightly lower than the maxima in the Arctic and in the Southern Ocean from 70°S to 80°S during the summer. On the contrary, both distributions of cya-SynSenPFT (Fig. 10B) and c-pico (Fig. 10C) show the lowest Chl-*a* in the gyres. Similar seasonality (with little changes) between the cya-SynSenPFT and c-pico is observed at the mid- to high latitudes, while the EOF-based prokaryotes show slightly lower Chl-*a* maxima as well as a different seasonal change in the Arctic, which have a clear elevation in Chl-*a* from spring to summer. It is noteworthy that the cyanobacteria derived from SynSenPFT include all the prokaryotic phytoplankton (Losa et al., 2017) which should thus be the same product as our EOF retrieved prokaryotes. The product of c-pico from Brewin et al. (2015) contains not only prokaryotes but also other picoeukaryotic phytoplankton (green algae and pico-sized haptophytes), therefore we also presented in Fig. 10D the sum of the prokaryotes and green algae Chl-*a* from our study, which shows much higher Chl-*a* concentration in general compared to c-pico, simply due to the predictions of high Chl-*a* of prokaryotes in the subtropical regions. Nevertheless, the high prokaryotes Chl-*a* concentrations in the subtropical regions are not only shown in our study, but are also found in the cyanobacteria simulated by NASA Ocean Biogeochemical Model (NOBM), which is a global biogeochemical model with coupled circulation and radiative models (Gregg, 2002; Gregg and Casey, 2007, figure not shown here but can be found in Losa et al., 2017). However, our prokaryotic phytoplankton

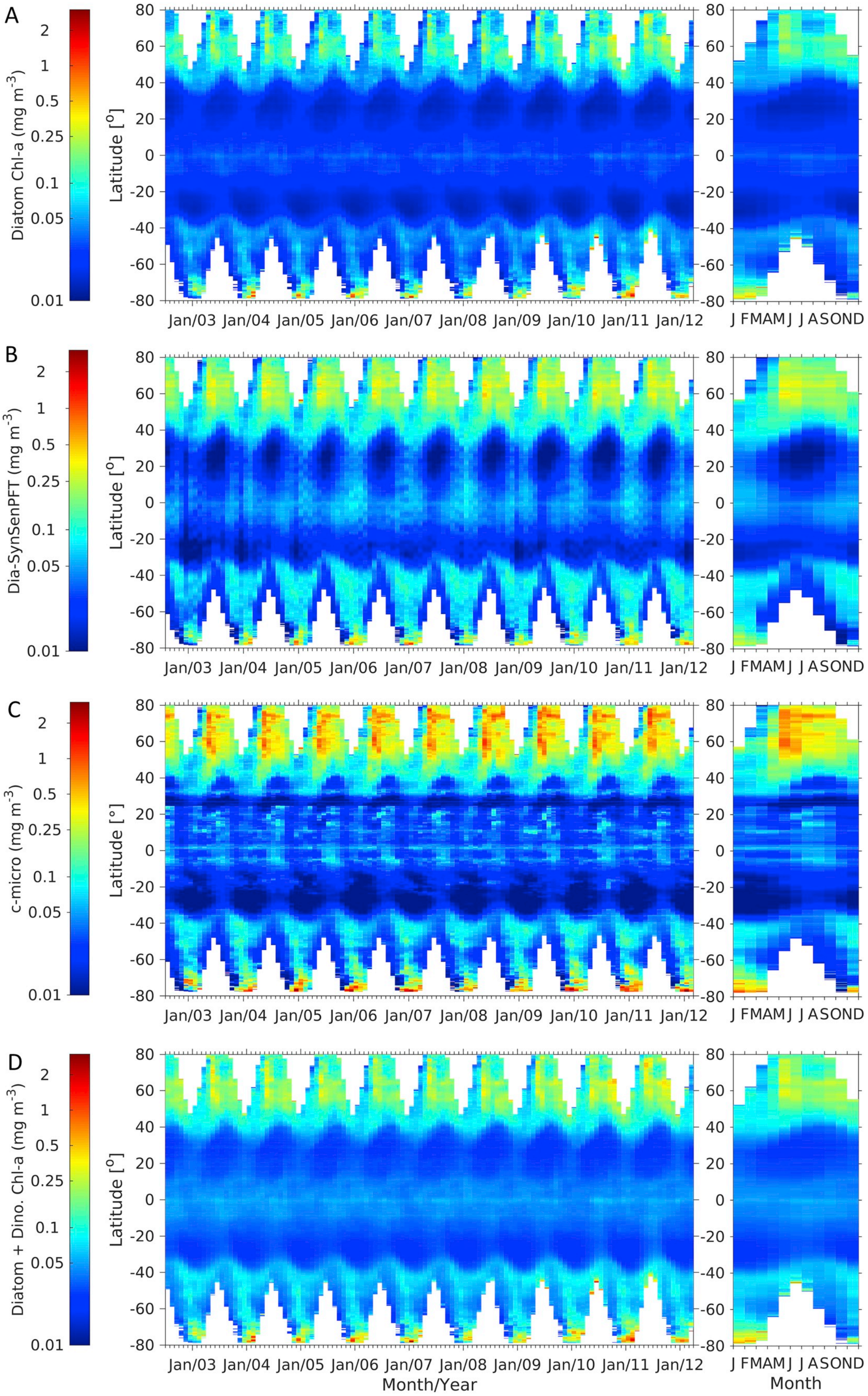


Fig. 8. Hovmöller diagrams of Chl-*a* concentrations of (A) diatoms derived from our study, (B) dia-SynSenPFT (Losa et al., 2017), (C) c-micro derived from PSC method (Brewin et al., 2015), and (D) sum of diatoms and dinoflagellates (Diatom + Dino.) from our study.

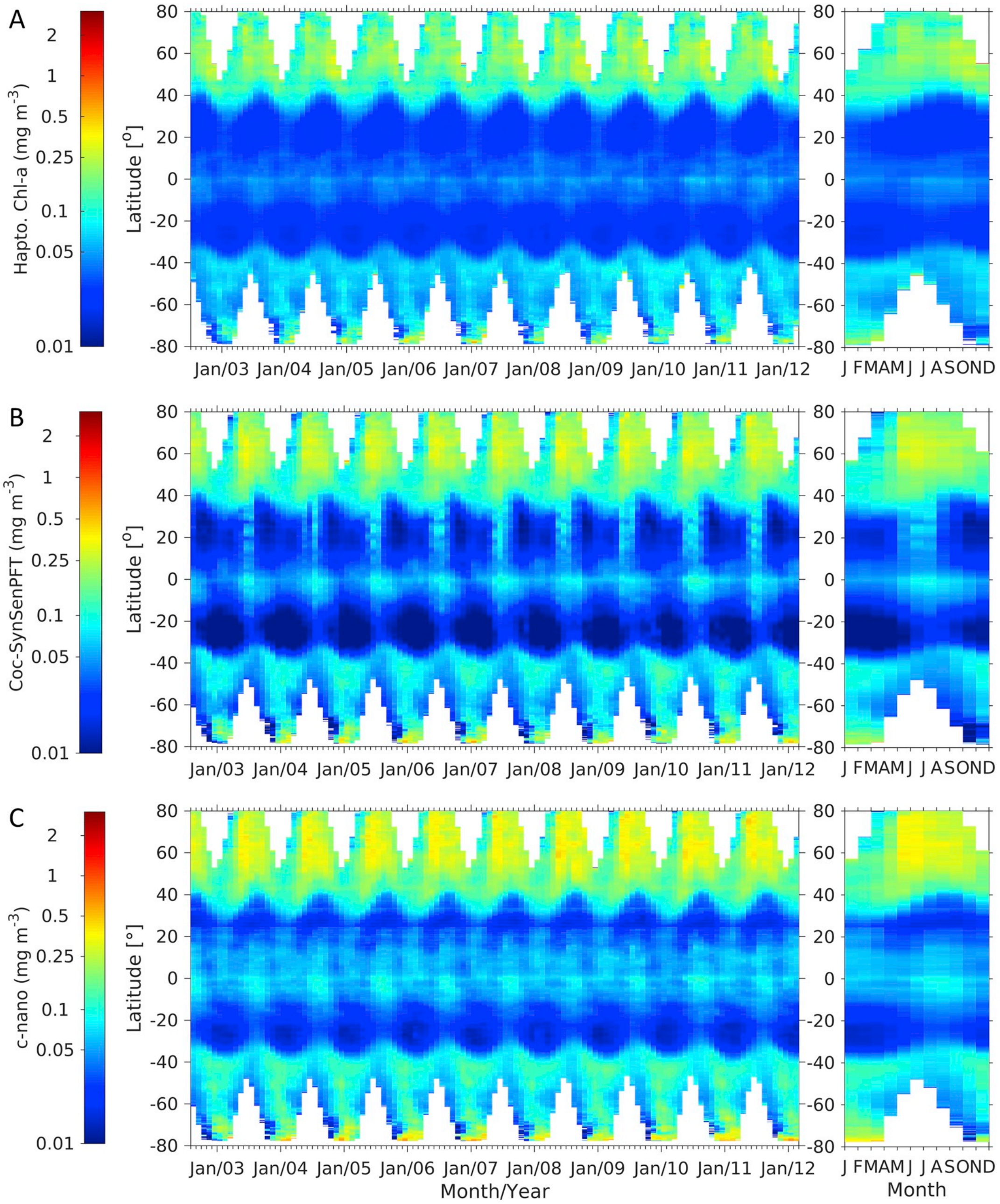


Fig. 9. Hovmöller diagrams of Chl-*a* concentrations of (A) haptophytes derived from our study, (B) coc-SynSenPFT (Losa et al., 2017), and (C) c-nano derived from PSC method (Brewin et al., 2015).

retrieval performance still needs to be further improved by potentially scaling the low concentration range or using non-linear prediction models.

Hovmöller diagrams of green algae and *Prochlorococcus* Chl-*a* concentrations derived by our study (Fig. 11A–B) are presented in comparison with those from OC-PFT (Hirata et al., 2011, Fig. 11D–E). Green algae in both products show distinct seasonality but Chl-*a* concentrations of green algae from our study are generally lower than those from OC-PFT (except for the subtropical regions), especially in the Arctic where OC-PFT shows enhanced green algae from late spring to early winter, whereas the EOF-based green algae show the lowest Chl-*a* during summer and increase in autumn to winter. *Prochlorococcus* Chl-*a* is generally very low (< 0.1 mg m^{-3}) for both products with quite different patterns presented. The EOF-based *Prochlorococcus* Chl-*a* concentrations are higher in mid- to low latitudes but lower in polar regions, corresponding to previous findings by Flombaum et al. (2013), while the OC-PFT *Prochlorococcus* shows higher Chl-*a* in the Southern Ocean which is outside the known distribution range and likely caused by undersampling of the in situ data (Hirata et al., 2011). Dinoflagellates show similar distribution with diatoms but with much lower Chl-*a* concentration, which is almost neglectable in subtropical regions and only higher than 0.05 mg m^{-3} in higher than 40°N with clear seasonality observed (Fig. 11C). However, an equivalent product is still necessary for dinoflagellates evaluation.

3.3.2. PFT Chl-a dominance comparison with PHYSAT products

We compared the PFT Chl-*a* dominance derived from our study for the period of 2002–2012 to the PHYSAT product from 1997 to 2006 (Alvain et al., 2008) which empirically relates the radiance anomaly to specific dominant phytoplankton groups. It is worth noting that the periods of the two compared products do not coincide, because we could only obtain the 12-month PHYSAT climatology data from 1997 to

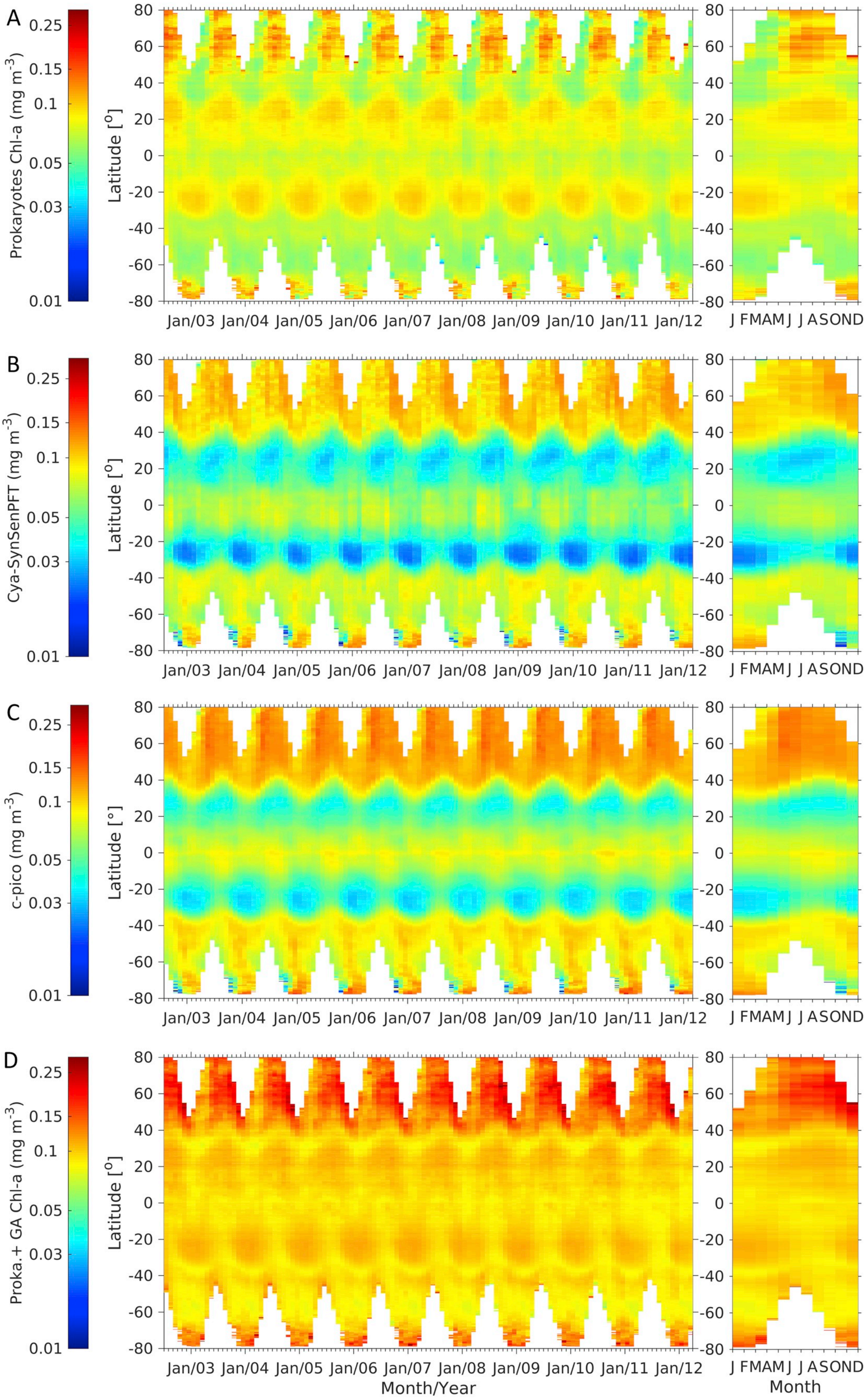


Fig. 10. Hovmöller diagrams of Chl-*a* concentrations of (A) prokaryotes derived from our study, (B) cya-SynSenPFT (Losa et al., 2017), (C) c-pico derived from PSC method (Brewin et al., 2015), and (D) sum of prokaryotes and green algae (Proka. + GA) from our study. Note that the color scale is different from Figs. 8–9.

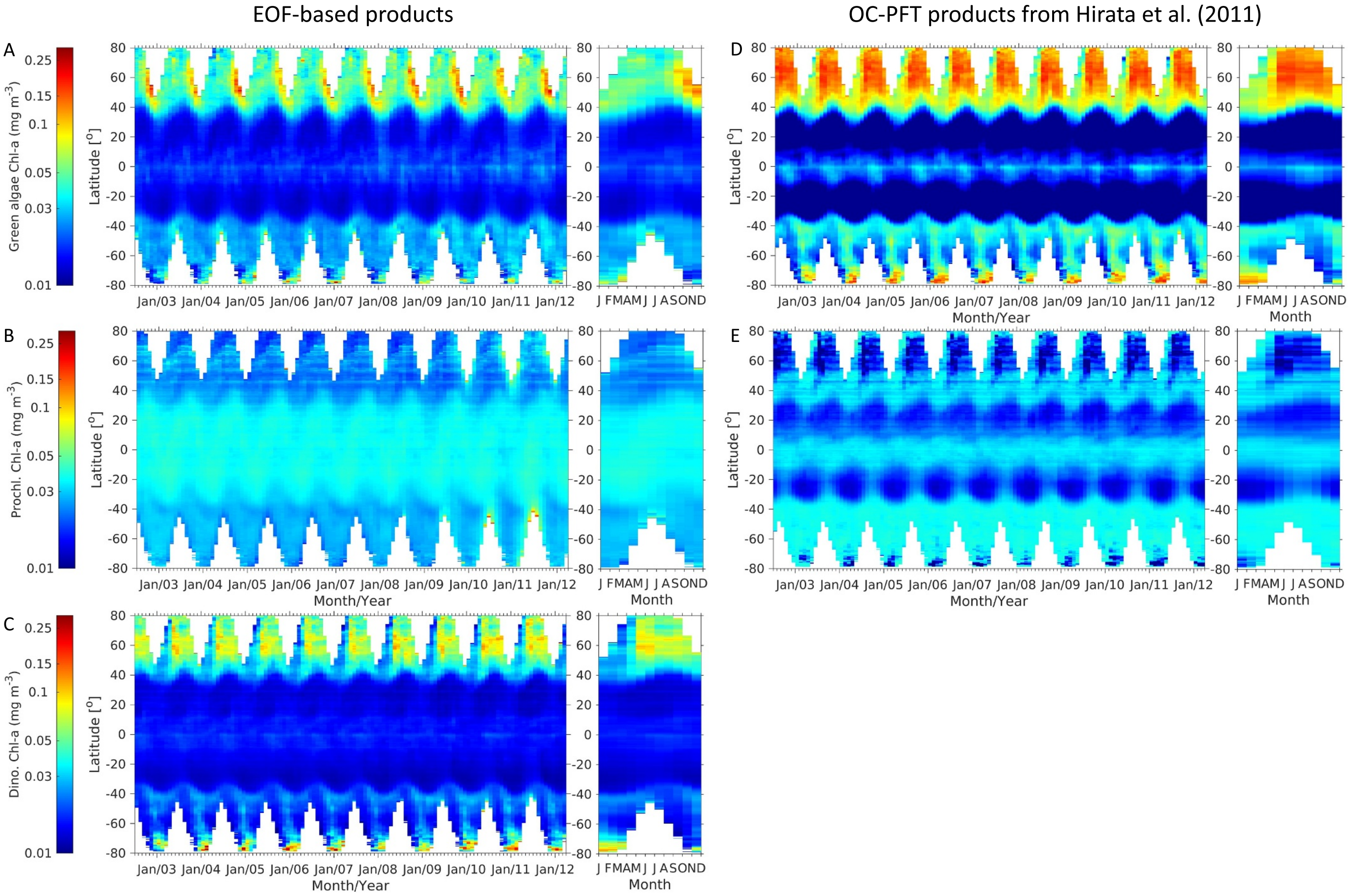


Fig. 11. Hovmöller diagrams of Chl-*a* concentrations of (A) green algae, (B) *Prochlorococcus*, and (C) dinoflagellates derived from our study, with the former two in comparison to (D) green algae and (E) *Prochlorococcus* from OC-PFT (Hirata et al., 2011). Note that the color scale is the same as Fig. 10.

2006 due to our limited access to the PHYSAT product. Diatoms, haptophytes, *Prochlorococcus* and SLC were included but the dominance of *Phaeocystis*-like group derived by PHYSAT was not available in our products. Distributions of dominant PFTs extracted from our products for four representative months (Fig. 12) generally present haptophytes and diatoms dominating in high latitudes and *Prochlorococcus* and SLC dominating in lower latitudes, which is well consistent with PHYSAT products. A more detailed comparison is described as follows.

In the high latitudes of the north hemisphere, our classified haptophytes dominance in January spreads a smaller range (30°N–50°N) than that from PHYSAT. In April, both products show haptophytes dominating above 30°N while diatoms are dominating in coastal areas only in our product. In July the two products show similar identification results with diatoms dominating in some parts of the North Sea, Norwegian Sea, Bering Sea and the Arctic waters, while haptophytes are still the major dominant PFT. Similar distribution is found for October as well despite of our product showing more diatoms in nearshore waters.

In the mid- to low latitudes, haptophytes and diatoms mainly dominate in coastal waters. Our product also shows dominance of haptophytes in the equatorial waters especially in the Pacific Ocean nearly for all seasons, which is barely presented in PHYSAT products. *Prochlorococcus* and SLC are two largely dominant groups in the mid- to low latitudes (Zubkhov et al., 1998). *Prochlorococcus* dominance is hardly found in 20°N–35°N from our product, whereas it is prominent in PHYSAT product especially in the north Pacific and north Atlantic gyres for all seasons. *Prochlorococcus* is found dominating in the low latitudes approximately between 15°S–15°N and SLC mainly dominates in the south Pacific gyre in both products. In the central to south Atlantic and Indian Ocean (equator to 40°S), our product shows SLC dominance in most of the regions in January and April, which decreases and is taken over by *Prochlorococcus* and haptophytes in July. However, PHYSAT products present dominance of both *Prochlorococcus* and SLC in this region in January and April, which is then gradually taken over by haptophytes in July with *Prochlorococcus* only dominating in the gyres. In the southern Pacific Ocean near 40°S both products show mainly *Prochlorococcus* dominating for nearly all seasons.

In the high latitudes of the south hemisphere (40°S–80°S), our product shows that *Prochlorococcus* and SLC spread more to the Southern Ocean especially from the south Pacific Ocean. In January, diatoms dominance of our product is found in Patagonian coastal waters and the south part of the Southern Ocean, while PHYSAT shows extensive diatoms dominance in 40°S–80°S with haptophytes and *Phaeocystis* also detected. For the other seasons, our product presents a smaller coverage of haptophytes dominance compared to PHYSAT products, and that diatoms are always dominant in Patagonian coastal waters and the west coast of South Africa.

Overall, besides some different distribution in diatoms dominance between PHYSAT and our products, the other main difference exists in the dominance of *Prochlorococcus* and SLC distributed in the central oceans, likely attributed to the low retrieval performance of prokaryotes. One should also keep in mind that PHYSAT product presents the climatology of 1997–2006 while ours for a more recent period (2002–2012) as explained in the beginning of this section. Recent long-term observations in the Arctic have shown a shift in phytoplankton composition from diatom-dominated to haptophyte-dominated and an

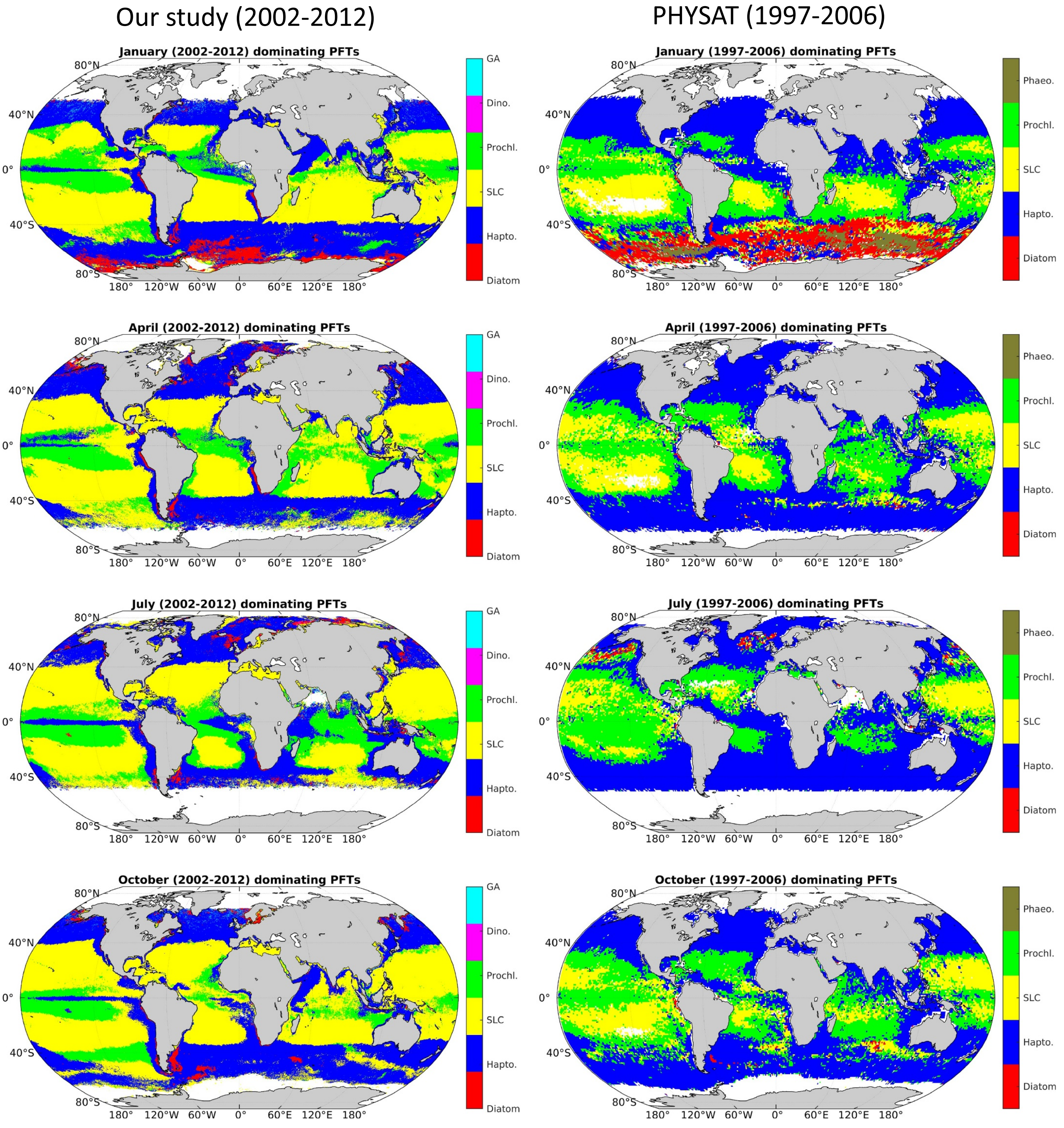


Fig. 12. PFT Chl-*a* dominance extracted from the EOF retrieved PFTs (2002–2012, left panel) versus the results derived by PHYSAT (1997–2006, Alvain et al., 2008, right panel) for representative months. Note that the classified dominant PFTs are not all the same as the PHYSAT product, i.e., dinoflagellates and green algae dominance were included in our product while *Phaeocystis* dominance was included in PHYSAT. Blank areas indicate no available data.

enhancement of prokaryotic phytoplankton indicating that smaller phytoplankton species appear more often in the high latitudes (Nöthig et al., 2015), which is also presented in our product. In general, the overall high consistency between the two products confirms that realistic information on the PFT dominance can be extracted from the EOF-based PFT products.

3.4. Potential application to Sentinel-3A OLCI products

3.4.1. EOF-based PFT algorithm based on collocated OLCI R_{rs} and in situ HPLC pigments

Based on the R_{rs} matchups extracted specifically from the OLCI products (listed in Table 2), the EOF-based algorithm for OLCI application was built using the matchups of R_{rs} at 10 bands, 11 bands and 12 bands (denoted as OLCI R_{rs_10}, OLCI R_{rs_11}, and OLCI R_{rs_12}, respectively) at 1 × 1 pixel and corresponding DPA-derived PFTs. The matchups for 3 × 3 averaged data were not used due to low number of points. As shown in Fig. 3, the same procedure was applied to OLCI matchups regarding EOF analysis and regression model establishment. Similar to Sects. 3.1 and 3.2.1, EOF analysis was performed and the contribution to total variance of each important EOF mode was provided in Table S5 (supplementary document), showing EOF-1 takes >85% and the first four EOF modes contribute >99.9% of the total spectral variance. Table S6 in supplementary document presents the stepwise routine generated ΔAIC showing the importance of the EOF modes. Different from that based on merged matchups (Table 4), EOF-3 and EOF-2 for most PFTs both have high ΔAIC scores, indicating both are relatively important. Statistical results of the prediction performance provided in Table S7 (supplementary document) shows little differences between using different band numbers for the input R_{rs} data sets. For all PFTs the predictions are comparable to that gained using merged matchups in Sect. 3.2.2, however low number of matchups led to weaker cross validation statistics. As an example, Fig. S10 (supplementary document) shows the comparison between the predicted and observed PFTs using OLCI R_{rs_11} at 1 × 1 pixel. In general, good predictions are achieved with OLCI R_{rs_11} for diatoms, haptophytes, dinoflagellates, green algae and *Prochlorococcus*, especially the predictions of the latter two PFTs were obviously better than those from merged matchups. However, performances for TChl-*a* and diatom Chl-*a* prediction are a bit downgraded with OLCI data compared to merged matchups, possibly due to the low quality of corrected R_{rs} at blue bands for OLCI. This needs to be further investigated as it does not apply to other PFTs. As the retrieval approach is also an empirical method based on regressions, other factors such as the lower number of matchup points andvariation range of input data do also have impacts on the OLCI model performance. Prokaryotes prediction still has the least good performance. Good performance for *Prochlorococcus* estimation is achieved but the robustness could be weak due to little number of matchups (only 17–22 points for 1 × 1 pixel).

3.4.2. Test output of global PFTs retrieved from S3A OLCI products

The EOF fitted models based on OLCI R_{rs_11} at 1 × 1 pixel were selected and applied to the OLCI R_{rs} L3 monthly products with 25 km spatial resolution. Fig. 13 shows the mean distribution of each PFT Chl-*a* concentration derived from OLCI products during April 2016–December 2018. Compared to the PFTs derived from merged products, diatoms derived from OLCI are also well represented for coastal regions and show similar distribution in polar regions, but have higher Chl-*a* in the gyres and lower at the Equator. Haptophytes, dinoflagellates and green algae show nearly identical distributions with those from merged products, suggesting that the fitted models of these PFTs are well defined for both satellite products. Prokaryotes present elevated Chl-*a* concentration in the oligotrophic regions, however, opposed to that derived from the merged products, low abundance of prokaryotes is detected in coastal waters and high latitudes especially in the Arctic. *Prochlorococcus* shows much spatial variation with higher Chl-*a* concentration in the Arctic Ocean, north and central Atlantic Ocean, central Pacific Ocean, most coastal waters, and scattered regions in the Southern Ocean. This is apparently inconsistent with the consensus that *Prochlorococcus* is hardly detectable in most of these regions. This misinterpretation might be attributed to the ill-defined prediction model due to low number of valid matchups for *Prochlorococcus*.

It is noteworthy that the PFTs are retrieved for different periods between OLCI (2016–2018) and merged products (2002–2012), and that the matchups extracted from OLCI data are not adequate for a global coverage. Relatively weak performance for prokaryotes and *Prochlorococcus* retrieval lies in both applications of OLCI and merged products, suggesting again that improvement on their models is necessary to achieve more reliable retrievals.

4. Conclusion and outlook

An EOF-based global retrieval algorithm for quantifying multiple PFTs was developed using collocated satellite R_{rs} data and DPA derived PFT Chl-*a* concentrations from in situ pigment data. R_{rs} matchups with different band numbers extracted from the GlobColour SeaWiFS/MODID/MERIS merged products were used to assess and compare the performance of corresponding EOF fitted models in predicting PFTs. The models developed using R_{rs} data set with nine bands slightly outperformed those using the other data sets. The retrieval skills for six PFTs (diatoms, dinoflagellates, haptophytes, green algae, prokaryotes and *Prochlorococcus*) were investigated and cross-validated via a bootstrapping method. Satisfactory retrievals were achieved for diatoms, dinoflagellates, haptophytes and green algae, while the correlation generated by the EOF-based models for prokaryotic phytoplankton was relatively weak, resulting in less accurate retrievals for prokaryotes and *Prochlorococcus*. Global PFT retrievals over a ten-year period (2002–2012) were obtained based on the EOF-based models using merged R_{rs} L3 products at nine bands, showing plausible distributions for most of the investigated PFTs in the open ocean.

Evaluations on the EOF-based PFT products were carried out through inter-comparisons with SynSenPFT, PSC and OC-PFT products. Time-latitude Hovmöller diagrams covering monthly means of 2002–2012 showed generally good agreement between our EOF-based PFTs and other PFT/PSC products, despite that prokaryotic phytoplankton showed higher Chl-*a* concentrations in the subtropical gyres, which needs to be further validated. Dominance of PFTs derived from the EOF-based PFT products was also in high agreement with PHYSAT-products. Implementation of EOF models to OLCI products showed potential for a continuous observation, though differences for certain PFTs appeared comparing to the PFT products derived from merged products, likely due to lower number and limited coverage of matchups.

Different from abundance-based PFT algorithms, the proposed retrieval algorithm directly uses the reflectance data from satellites, thus can avoid uncertainty generated in the chlorophyll products, and links the variation of satellite reflectance spectra via PFT specific EOF based regression models. In addition, the retrieval algorithm is still an empirical approach, which is subject to the input data sets for training with regard to the number of observations, range of data variation and homogeneity of the data distribution in space and time. The apparent over−/underestimation feature in the regression models should also be further investigated. Nevertheless, this study showed the high potential of the EOF-based algorithm for quantitatively retrieving PFTs globally using satellite reflectance products from different sensors, which was not adequately reported in previous studies. Future efforts will be put in improving the current algorithm especially for prokaryotes prediction, such as applying proper scaling to the data sets and using non-linear fitting models. Further work, which is ongoing, is focusing on updating and extending the global in situ pigment data sets (especially from 2012 to present). The updated data sets will be used for EOF re-training with hopefully globally distributed matchups for OLCI (2016–present) and MODIS/VIIRS (2012–present) merged products, to fill the gap between MERIS and OLCI (2012–2016), and ultimately enable a continuous PFT observation from multi-sensor data. The updated in situ data sets will also be used for a thorough validation of the satellite retrieved PFTs.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2020.111704.

Author contributions

Conceptualization: H.X. and A.B.; *Methodology*: A.B. and H.X.; *Writing – Original Draft*: H.X., *Writing – Review & Editing*: H.X., A.B., M.S., and Y.L.; *Validation*: H.X., A.B., S.L., and M.S.; *Formal analysis*: H.X.; *Investigation*: H.X., A.B., S.L., A.M., and P.G.; *Resources*: A.M., P.G., J.D., O.H.A., A.B., Y.L., M.S., and S.L.; *Visualization*: H.X. and S.L.; *Supervision*: A.B., M.A., and O.H.A.; *Project administration*: A.B. and H.X.

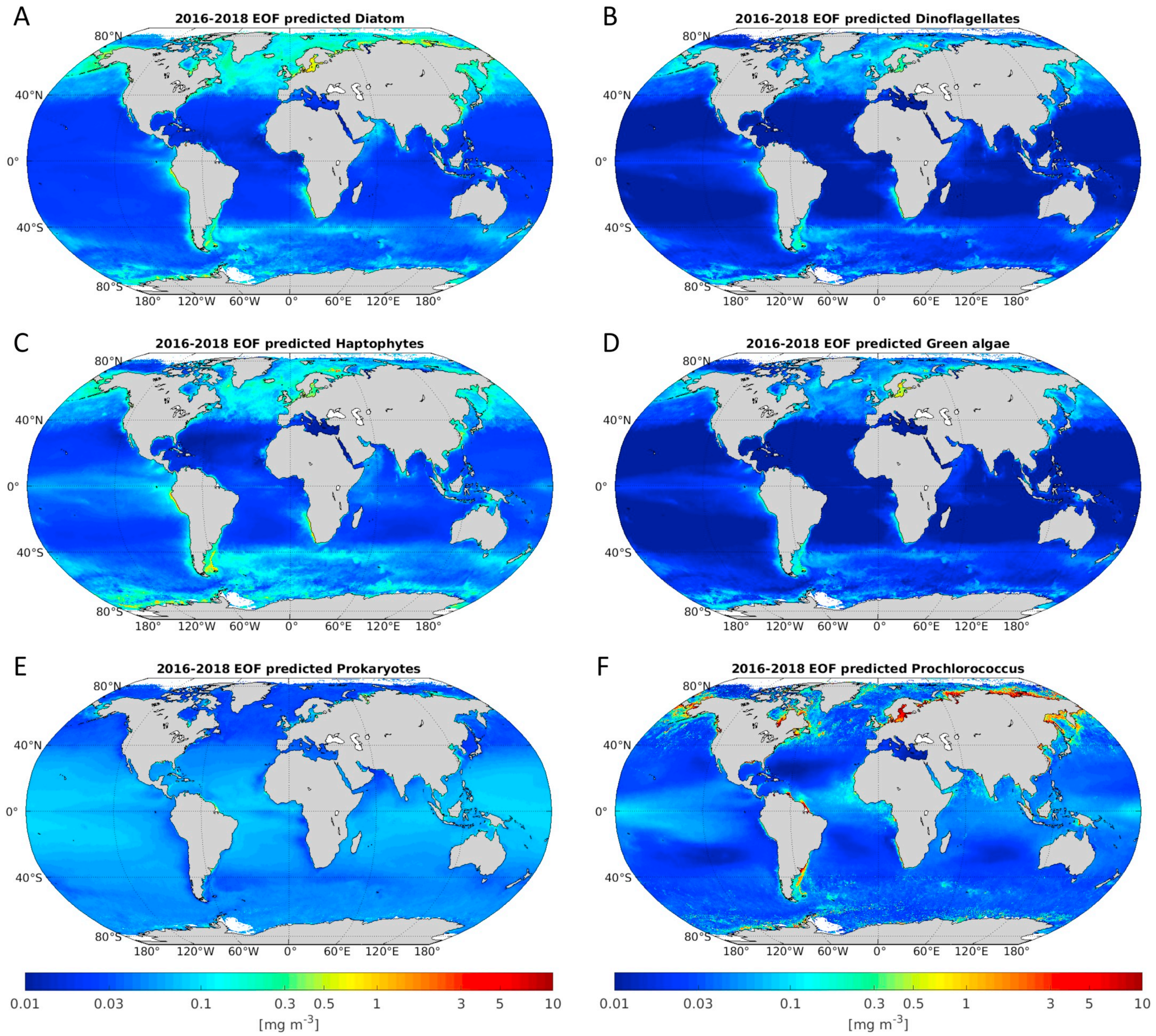


Fig. 13. Yearly mean distribution (April 2016–December 2018) of the PFT Chl-*a* concentrations for (A) diatoms, (B) dinoflagellates, (C) haptophytes, (D) green algae, (E) prokaryotes, and (F) *Prochlorococcus* retrieved by the EOF-based algorithm from OLCI monthly R_{rs} products at 11 bands.

Data availability

All data used in this study were obtained via the links provided in Sect. 2.1 and in Losa et al. (2017). Numerical matrices and regression coefficients regarding the EOF-based algorithm are provided in the supplementary document.

Declaration of competing interest

None.

Acknowledgement

This work is supported by a collaborative project between ACRI-ST and Phytooptics team at Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, OLCI-PFT (ACRI-AWI Offer #209-180104). The contribution of Svetlana N. Losa is partly made in the framework of the state assignment of the Federal Agency for Scientific Organizations (FASO) Russia (theme 0149-2019-0015). We thank Marc Taylor for the R script on the EOF method, Sonja Wiegmann for her dedication in measuring and processing the HPLC data for our own cruises, and all the previous and current Phytooptics team members who participated in the past cruises for data collection and analysis. We are thankful to all the scientists and crew involved in the global HPLC data collection and analyses for providing their pigment data. We thank Robert Brewin for the PSC products and Séverine Alvain for making the PHYSAT data available to us. Thanks also to NASA, ESA and EUMETSAT for the SeaWiFS, MODIS, VIIRS, MERIS, and OLCI data, and specially the CMEMS GlobColour program for providing the OLCI and merged ocean color L3 products. Marine Bretagnon is also acknowledged for providing useful suggestions during the OLCI-PFT project discussions. We finally thank the three reviewers for providing valuable comments that substantially improved the manuscript.

References

ACRI-ST GlobColour Team, Mangin, A., Fanton d'Andon, O., 2017. GlobColour Product User Guide, GC-UM-ACR-PUG-01, Version 4.1. (Sophia-Antipolis).

Aiken, J., Pradhan, Y., Barlowd, R., Lavender, S., Poulton, A., Patrick, H., Hardman-Mountford, N., 2009. Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995–2005. Deep-Sea Research II 56, 899–917. https://doi.org/10.1016/j.dsr2.2008.09.017.

Alvain, S., Moulin, C., Dandonneau, Y., Bréon, F.M., 2005. Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. Deep. Res. Part I Oceanogr. Res. Pap. 52, 1989–2004. https://doi.org/10.1016/j.dsr.2005.06.015.

Alvain, S., Moulin, C., Dandonneau, Y., Loisel, H., 2008. Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: a satellite view. Glob. Biogeochem. Cycles. https://doi.org/10.1029/2007GB003154.

Bailey, S.W., Werdell, P.J., 2006. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2006.01.015.

Bracher, A., 2019. Phytoplankton pigment concentrations in the Southern Ocean during RV POLARSTERN cruise PS103 in Dec 2016 to Jan 2017. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEAhttps://doi.pangaea.de/10.1594/PANGAEA.898941.

Bracher, A., Wiegmann, S., 2019. Phytoplankton pigment concentrations in the North Sea and Sogne Fjord from 29 April to 7 May 2016 during RV HEINCKE cruise HE462. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEAhttps://doi.pangaea.de/10.1594/PANGAEA.899043.

Bracher, A., Vountas, M., Dinter, T., Burrows, J.P., Röttgers, R., Peeken, I., 2009. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. Biogeosciences. https://doi.org/10.5194/bg-6-751-2009.

Bracher, A., Taylor, M.H., Taylor, B., Dinter, T., Röttgers, R., Steinmetz, F., 2015. Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Sci. https://doi.org/10.5194/os-11-139-2015.

Bracher, A., Soppa, M., Losa, S., Dinter, T., Wolanin, A., Brewin, R., Bricaud, A., 2016. Final report. version 1.2, 30 Nov 2016, electronic version: SEOM-SynSenPFT-FR-D4.3_v1.2.pdf.

Bracher, A., Bouman, H.A., Brewin, R.J.W., Bricaud, A., Brotas, V., Ciotti, A.M., Clementson, L., Devred, E., Di Cicco, A., Dutkiewicz, S., Hardman-Mountford, N.J., Hickman, A.E., Hieronymi, M., Hirata, T., Losa, S.N., Mouw, C.B., Organelli, E., Raitsos, D.E., Uitz, J., Vogt, M., Wolanin, A., 2017. Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. Front. Mar. Sci. 4, 1–15. https://doi.org/10.3389/fmars.2017.00055.

Bracher, Astrid, Wiegmann, Sonja, Xi, Hongyan, 2020. Phytoplankton pigment concentration and phytoplankton groups measured on water samples obtained during POLARSTERN cruise PS113 in the Atlantic Ocean. PANGAEA. https://doi.org/10.1594/PANGAEA.911061.

Brewin, R.J.W., Sathyendranath, S., Hirata, T., Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., 2010. A three-component model of phytoplankton size class for the Atlantic Ocean. Ecol. Model. https://doi.org/10.1016/j.ecolmodel.2010.02.014.

Brewin, R.J.W., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Airs, R., Lamont, T., 2015. Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2015.07.004.

Campbell, J.W., 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res. https://doi.org/10.1029/95JC00458.

Ciotti, A.M., Bricaud, A., 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. Limnol. Oceanogr. Methods 4, 237–253. https://doi.org/10.4319/lom.2006.4.237.

Correa-Ramirez, M., Morales, C.E., Letelier, R., Anabalon, V., Hormazabal, S., 2018. Improving the remote sensing of phytoplankton functional types (PFT) using empirical orthogonal functions: a case study in a coastal upwelling region. Remote Sens. 10 (4), 498. https://doi.org/10.3390/rs10040498.

Craig, S.E., Jones, C.T., Li, W.K.W., Lazin, G., Horne, E., Caverhill, C., Cullen, J.J., 2012. Deriving optical metrics of coastal phytoplankton biomass from ocean colour. Remote Sens. Environ. 119, 72–83. https://doi.org/10.1016/j.rse.2011.12.007.

Craig, S.E., Lohrenz, S.E., Lee, Z., Mahoney, K.L., Kirkpatrick, G.J., Schofield, O.M., Steward, R.G., 2006. Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, Karenia brevis. Appl. Opt. 45, 5414–5425. https://doi.org/10.1364/AO.45.005414.

Devred, E., Sathyendranath, S., Stuart, V., Maass, H., Ulloa, O., Platt, T., 2006. A two-component model of phytoplankton absorption in the open ocean: theory and applications. J. Geophys. Res. Ocean. https://doi.org/10.1029/2005JC002880.

Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. Science 281, 200–206. https://doi.org/10.1126/science.281.5374.200.

Falkowski, P.G., Laws, E.A., Barber, R.T., Murray, J.W., 2003. Phytoplankton and their role in primary, new, and export production. In: Fasham, M.J.R. (Ed.), Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change. Springer, pp. 99–121.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K.W., Lomas, M.W., Veneziano, D., Vera, C.S., Vrugt, J.A., Martiny, A.C., 2013. Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. Proc. Natl. Acad. Sci. 110, 9824–9829. https://doi.org/10.1073/pnas.1307701110.

Gregg, W.W., 2002. A coupled ocean-atmosphere radiative model for global ocean biogeochemical models. In: Suarez, M. (Ed.), NASA Global Modeling and Assimilation Series. 22. NASA Technical Memorandum 2002–104606, Greenbelt, MD, pp. 33.

Gregg, W.W., Casey, N.W., 2007. Modeling coccolithophores in the global oceans. Deep. Res. Part II Top. Stud. Oceanogr. https://doi.org/10.1016/j.dsr2.2006.12.007.

Gregg, W.W., Rousseaux, C.S., 2017. Simulating PACE Global Ocean Radiances. Front. Mar. Sci. 4, 1–19. https://doi.org/10.3389/fmars.2017.00060.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van Der Linden, S., Leitão, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote Sens. 7, 8830–8857. https://doi.org/10.3390/rs70708830.

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T.J., Barlow, R.G., 2008. An absorption model to determine phytoplankton size classes from satellite ocean colour. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2008.03.011.

Hirata, T., Hardman-Mountford, N.J., Brewin, R.J.W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., Yamanaka, Y., 2011. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences 8, 311–327. https://doi.org/10.5194/bg-8-311-2011.

Hu, C., Cannizzaro, J., Carder, K.L., Muller-Karger, F.E., Hardy, R., 2010. Remote detection of Trichodesmium blooms in optically complex coastal waters: examples with MODIS full-spectral data. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2010.04.011.

IOCCG, 2014. Phytoplankton functional types from space. In: Sathyendranath, S., Stuart, V. (Eds.), Reports of the International Ocean Color Coordinating Group No. 15. IOCCG, Dartmouth, NS.

IPCC, 2013. Climate change 2013: the physical science basis. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M.M.B., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK; New York, NY, USA, pp. 1535.

Kostadinov, T.S., Siegel, D.A., Maritorena, S., 2009. Retrieval of the particle size distribution from satellite ocean color observations. J. Geophys. Res. Ocean. https://doi.org/10.1029/2009JC005303.

Le Quéré, C., Harrison, S.P., Prentice, C.I., Buitenhuis, E.T., Aumonts, O., Bopp, L., et al., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Glob. Chang. Biol. 11, 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.1004.x.

Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J., Middleton, E.M., 2015. An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2015.06.012.

Litchman, E., Klausmeier, C.A., Miller, J.R., Schofield, O.M., Falkowski, P.G., 2006. Multi-nutrient, multi-group model of present and future oceanic phytoplankton communities. Biogeosciences 3, 585–606. https://doi.org/10.5194/bg-3-585-2006.

Liu, Y., Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., Pan, Y., Bracher, A., 2019a. Phytoplankton Pigment Concentration Estimated From Underway AC-S Particulate Absorption Data During POLARSTERN Cruise PS99.2. PANGAEAhttps://doi.org/10.1594/PANGAEA.898102.

Liu, Y., Boss, E., Chase, A., Xi, H., Zhang, X., Röttgers, R., Pan, Y., Bracher, A., 2019b. Phytoplankton Pigment Concentration Estimated From Underway AC-S Particulate Absorption Data During POLARSTERN Cruise PS107. PANGAEAhttps://doi.org/10.1594/PANGAEA.898100.

Liu, Y., Hellmann, S., Wiegmann, S., Bracher, A., 2019c. Phytoplankton Pigment Concentrations Measured by HPLC During POLARSTERN Cruise PS99.1. PANGAEA. https://doi.org/10.1594/PANGAEA.905502.

Losa, S.N., Soppa, M.A., Dinter, T., Wolanin, A., Brewin, R.J.W., Bricaud, A., Oelker, J., Peeken, I., Gentili, B., Rozanov, V., Bracher, A., 2017. Synergistic exploitation of hyper- and multi-spectral precursor sentinel measurements to determine phytoplankton functional types (SynSenPFT). Front. Mar. Sci. 4, 1–22. https://doi.org/10.3389/fmars.2017.00203.

Lubac, B., Loisel, H., 2007. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2007.02.012.

Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., Tieche, F., 2007. Optical properties of the "clearest" natural waters. Limnol. Oceanogr. 52 (1), 217–229.

Mouw, C.B., Hardman-Mountford, N.J., Alvain, S., Bracher, A., Brewin, R.J.W., Bricaud, A., Ciotti, A.M., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T.S., Roy, S., Uitz, J., 2017. A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the Global Ocean. Front. Mar. Sci. 4. https://doi.org/10.3389/fmars.2017.00041.

Nöthig, E.M., Bracher, A., Engel, A., Metfies, K., Niehoff, B., Peeken, I., Bauerfeind, E., Cherkasheva, A., Gäbler-Schwarz, S., Hardge, K., Kilias, E., Kraft, A., Kidane, Y.M., Lalande, C., Piontek, J., Thomisch, K., Wurst, M., 2015. Summertime plankton ecology in Fram Strait - a compilation of long-and short-term observations. Polar Res. https://doi.org/10.3402/polar.v34.23349.

Palacz, A.P., St. John, M.A., Brewin, R.J.W., Hirata, T., Gregg, W.W., 2013. Distribution of phytoplankton functional types in high-nitrate, low-chlorophyll waters in a new diagnostic ecological indicator model. Biogeosciences. https://doi.org/10.5194/bg-10-7553-2013.

Raitsos, D.E., Lavender, S.J., Maravelias, C.D., Haralabous, J., Richardson, A.J., Reid, P.C., 2008. Identifying four phytoplankton functional types from space: an ecological approach. Limnol. Oceanogr. https://doi.org/10.4319/lo.2008.53.2.0605.

Sieburth, J.M., Smetacek, V., Lenz, J., 1978. Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. Limnol. Oceanogr. 23, 1256–1263. https://doi.org/10.4319/lo.1978.23.6.1256.

Soja-Woźniak, M., Craig, S.E., Kratzer, S., Wojtasiewicz, B., Darecki, M., Jones, C.T., 2017. A novel statistical approach for ocean colour estimation of inherent optical properties and cyanobacteria abundance in optically complex waters. Remote Sens. 9, 1–22. https://doi.org/10.3390/rs9040343.

Soppa, M.A., Hirata, T., Silva, B., Dinter, T., Peeken, I., Wiegmann, S., Bracher, A., 2014. Global retrieval of diatom abundance based on phytoplankton pigments and satellite data. Remote Sens. 6, 10089–10106. https://doi.org/10.3390/rs61010089.

Soppa, M.A., Peeken, I., Bracher, A., 2017. Global Chlorophyll "a" Concentrations for Diatoms, Haptophytes and Prokaryotes Obtained with the Diagnostic Pigment Analysis of HPLC Data Compiled From Several Databases and Individual Cruises. PANGAEAhttps://doi.org/10.1594/PANGAEA.875879.

Taylor, B.B., Taylor, M.H., Dinter, T., Bracher, A., 2013. Estimation of relative phycoerythrin concentrations from hyperspectral underwater radiance measurements - a statistical approach. J. Geophys. Res. Ocean. 118, 2948–2960. https://doi.org/10.1002/jgrc.20201.

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. J. Geophys. Res. Ocean. 111. https://doi.org/10.1029/2005JC003207.

Vidussi, F., Claustre, H., Manca, B.B., Luchetta, A., Marty, J.-C., 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. J. Geophys. Res. Ocean. https://doi.org/10.1029/1999JC000308.

Wang, G., Lee, Z., Mouw, C., 2017. Multi-spectral remote sensing of phytoplankton pigment absorption properties in cyanobacteria bloom waters: a regional example in the Western Basin of Lake Erie. Remote Sens. https://doi.org/10.3390/rs9121309.

Ward, B.A., 2015. Temperature-correlated changes in phytoplankton community structure are restricted to polar waters. PLoS One. https://doi.org/10.1371/journal.pone.0135581.

Werdell, P.J., Roesler, C.S., Goes, J.I., 2014. Discrimination of phytoplankton functional groups using an ocean reflectance inversion model. Appl. Opt. 53, 4833. https://doi.org/10.1364/ao.53.004833.

Zubkhov, M.V., Sleigh, M.A., Tarran, G.A., Burkhill, P.H., Leakey, R.J., Raymond, J.G., 1998. Picoplanktonic community structure on an Atlantic transect from 50°N to 50°S. Deep-Sea Res. I 45, 1339–1355. https://doi.org/10.1016/S0967-0637(98)00015-6.

AGU100 ADVANCING EARTH AND SPACE SCIENCE
JGR
Journal of Geophysical Research: Oceans

RESEARCH ARTICLE
10.1029/2017JC013651

In Situ and Satellite Observations of Phytoplankton Size Classes in the Entire Continental Shelf Sea, China

Xuerong Sun[1], **Fang Shen**[1,2], **Dongyan Liu**[1], **Richard G. J. Bellerby**[1,3], **Yangyang Liu**[1], and **Rugang Tang**[1]

[1]State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai, China, [2]Institute of Eco-Chongming (IEC), Shanghai, China, [3]Norwegian Institute for Water Research, Bergen, Norway

Key Points:
- In situ PSCs obtained from HPLC method were investigated in the entire continental shelf sea of China, including the BS, the YS, and the ECS
- An improved algorithm was developed and validated to be applicable in the BS, YS, and the ECS
- Diurnal variations of PSCs were obtained when applying the improved algorithm to GOCI images

Correspondence to:
F. Shen,
fshen@sklec.ecnu.edu.cn

Citation:
Sun, X., Shen, F., Liu, D., Bellerby, R. G. J., Liu, Y., & Tang, R. (2018). In situ and satellite observations of phytoplankton size classes in the entire continental shelf sea, China. *Journal of Geophysical Research: Oceans*, *123*. https://doi.org/10.1029/2017JC013651

Received 29 NOV 2017
Accepted 12 APR 2018
Accepted article online 19 APR 2018

Abstract Phytoplankton size classes (PSCs) is of great significance for exploring marine ecological and biogeochemical processes. Remote sensing of PSCs has been successfully applied to open oceans; however, it is still quite limited for optically complex coastal oceans. In this study, the entire continental shelf sea of China including Bohai Sea (BS), Yellow Sea (YS), and East China Sea (ECS) characterized by distinctive turbid waters and impacted by plumes of large world-class river (the Changjiang River) was taken as an example of turbid coastal ocean for remotely sensed spatial-temporal distributions of PSCs. In situ data were collected from cruises during April to June in 2014 and an improved algorithm for PSCs retrieval was proposed. PSCs derived from GOCI (Geostationary Ocean Color Imager) images revealed that microplankton was dominant in the BS, the YS, and the nearshore ECS and nanoplankton distributed widely in the entire study area, while picoplankton mainly distributed in the offshore ECS in April, which was consistent with in situ investigation and related to environmental factors. Validation indicated that the improved algorithm provided a more accurate estimation of PSCs, with the root mean square error (RMSE) between estimated and measured size-fractionated concentrations been 0.774, 0.257, and 0.142 mg m^{-3} for micro, nano, and picoplankton, respectively. Diurnal variations of PSCs were mainly affected by tidal currents and light intensity depending on different water types. These illustrated that remote sensed spatial distributions as well as diurnal variations of PSCs are effective in turbid continental shelf seas of China.

1. Introduction

Phytoplankton are the fundamental component of the marine ecosystem. Complex biophysical controls (e.g., irradiance, temperature, salinity, nutrients, and grazing) and anthropogenic impact (e.g., runoff modification, nutrient fluxes) result in large spatial and temporal variations in phytoplankton biomass, community structure, and functionality (Behrenfeld & Boss, 2014; Cloern, 2001; Geider et al., 1998; Lindemann & John, 2014). According to Sieburth et al. (1978), phytoplankton can be operationally divided into three phytoplankton size classes (i.e., microplankton, >20 μm; nanoplankton, 2–20 μm; picoplankton, <2 μm). Phytoplankton size structure are recognized as prime physiological parameters that can influence many marine ecological and biogeochemical processes (Finkel et al, 2010). For example, physiology of phytoplankton, including metabolic rates, growth rates, nutrient uptake, and sinking rate are influenced by cell size (Geider et al., 1986; Waite et al., 1997). Several researches also suggested that size structure can result in different photosynthetic rates and maximum quantum yields, potentially leading to different carbon fixation attributes (Hirata et al., 2009; Uitz et al., 2008). Phytoplankton cell size also influence the optical properties of the ocean surface via light scattering and absorption (Devred et al., 2006; Yentsch & Phinney, 1989). Furthermore, shifts in PSCs can affect the relationship of phytoplankton grazers, and consequently change the function and structure of marine food webs (Legendre & Rassoulzadegan, 1995). Therefore, size structure of phytoplankton has been widely investigated in its role in primary production, carbon-specific photosynthesis, and export production (Cermeño et al., 2005; Hilligsøe et al., 2011; Teira et al., 2001) and utilized in establishing and improving biogeochemical models in marine system (e.g., Aumont et al., 2003; Ward et al., 2012).

A few tools can help to obtain the PSCs, e.g., microscopy, flow cytometry, size-fractionated filtration (SFF), and high performance liquid chromatography (HPLC), and each method has its own advantages and disadvantages (IOCCG, 2014, Table 2.3). Results of these approaches usually generate a “static” image, for the given environmental conditions and species distributions. However, oceanic environment and phytoplankton distributions are dynamic in space and time. Remote sensing has been regarded as the most important tool for acquiring the continuous observational data spatially and temporally. Hence, the demand to develop methods for identifying PSCs in ocean using remote sensing is urgent. To estimate size-fractionated chlorophyll-a concentration, statistical links between the total chlorophyll-a concentration and PSCs derived from HPLC should be developed. Vidussi et al. (2001) pioneered this approach by selecting seven diagnostic pigments to obtain the fractions of PSCs in the total chlorophyll concentration. Afterward, constant improvements of relationships between the total chlorophyll-a concentration and PSCs were made, such as determining weighting factors of diagnostic pigments and refining assignments of diagnostic pigments for classifying PSCs (Brewin et al., 2010; Hirata et al., 2011; Uitz et al., 2006). Through satellite images (i.e., SeaWiFs and MODIS), these statistical relationships have been successfully utilized in estimating PSCs in global ocean and different oceanic regions (e.g., Brewin et al., 2010, 2012, 2015; Brotas et al., 2013; Hirata et al., 2011; Lin et al., 2014; Uitz et al., 2006).

As mentioned above, relationships between PSCs and total chlorophyll-a concentration have been widely used in global oceanic waters. However, few relationships were carried out in studies for applying in coastal and estuarine systems, particularly in the region with large river plumes. The BS, the YS, and the ECS make up the largest marginal seas, and host one of the most turbid coastal and shelf seas environments globally, supporting high primary and fishery production. They are strongly impacted by the large Changjiang River plumes and adjoining ocean processes, and hydrodynamic features are predominantly driven by variations in tides, large seasonal climatic, and monsoon changes. Together these drivers result in complicated optical, hydrodynamic, and biogeochemical environments, which have a strong influence on the physiology of phytoplankton, leading to region-specific relationships between chlorophyll-a concentration and PSCs, compared to general oceanic conditions. Several studies in which pigment concentrations or size structures of phytoplankton based on SFF methods have been investigated in the BS, the YS, or the ECS (e.g., Deng et al., 2008; Fu et al., 2009; Huang et al, 2006; Sun et al., 2002, 2012; Zhu et al., 2009). However, to the best of our knowledge, there have not been published documents on the application of remote sensing derived PSCs in the entire continental shelf sea of China.

Therefore, we tried to develop an improved algorithm for the monitoring spatial distribution and diurnal variations of PSCs from remote sensing focusing on the BS, the YS, and the ECS. For this purpose, we presented comprehensive in situ measurements derived from HPLC method, aiming to seek relationships between proportions of PSCs and chlorophyll-a concentration. Based on these, existing models (Brewin et al., 2010; Hirata et al., 2011) in estimating PSCs from satellite images were tested for the applicability in the turbid seas affected by terrestrial inputs, of which three-component model (Brewin et al., 2010) was improved by parameterization and subsequently validated through match-ups. Spatial distributions and diurnal variations of PSCs from GOCI observations were analyzed and major controlling environmental factors were discussed.

2. Data and Method

2.1. In Situ Data

The BS is a shallow semienclosed marginal sea on the northern coast of China, of which the mean water depth is no more than 20 m (Figure 1). The Bohai Strait, which connecting the southern of Liaodong Peninsula and the northern of Shandong Peninsula, is defined as the boundary between the BS and the YS. The YS is surrounded by mainland China and the Korean Peninsula, including the north Yellow Sea (NYS) and the south Yellow Sea (SYS). The depth of YS increases gradually from nearshore area (continental shelf) to offshore area (central YS), approximately from 0–40 m to 60–80 m (Figure 1). To its south is the ECS, which is the one of the largest marginal seas in the world. The depth of ECS is shallow in the northwestern area and deep in the southeastern area, due to the impact of continental shelf. The ECS can be broadly divided into three parts: the continental shelf (0–60 m), the outer continental shelf (60–200 m), and Okinawa Trough (200–2,700 m). To the northwest of Taiwan Island, the depth is less than 200 m and it increases significantly toward southeastern area (Figure 1).

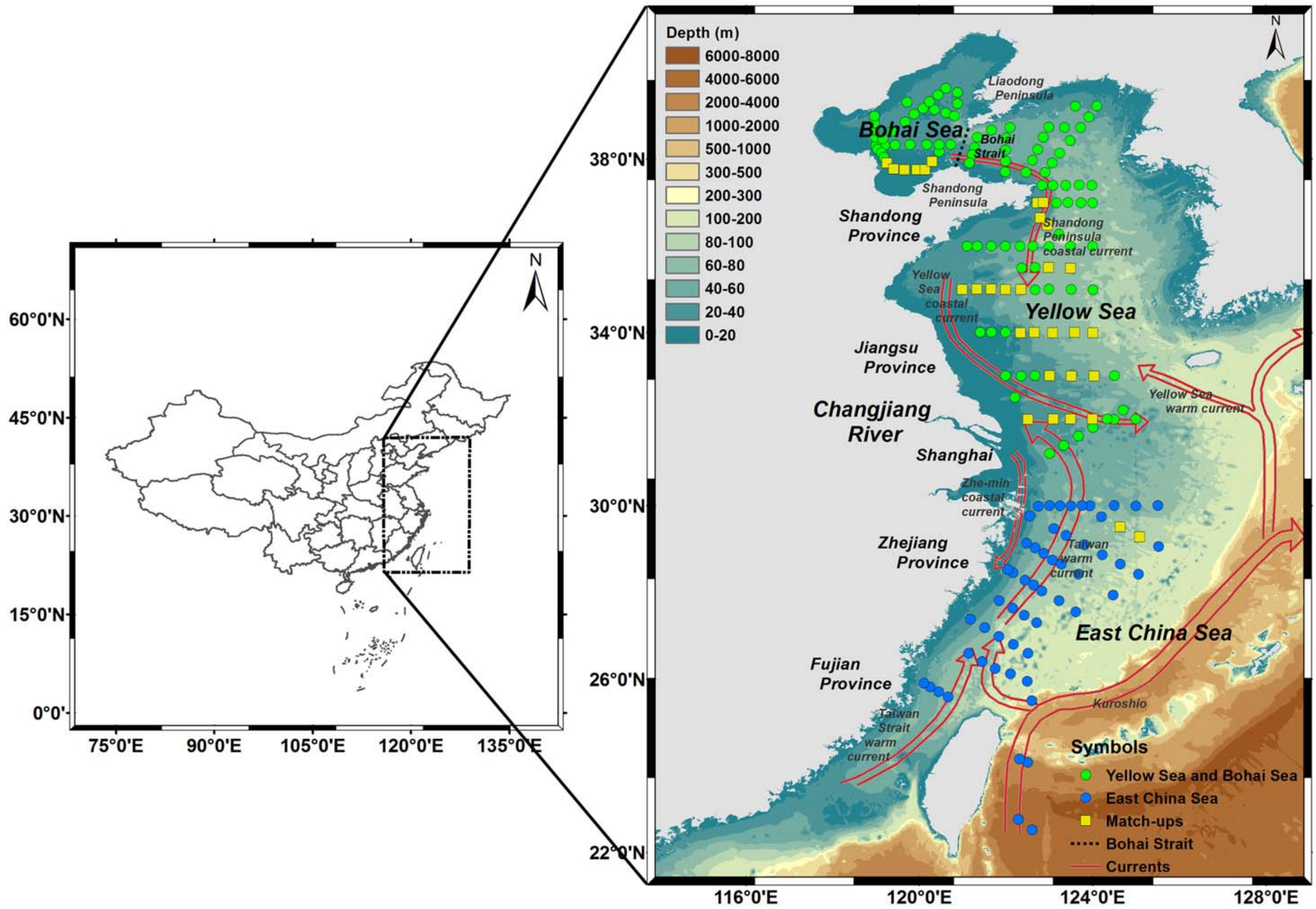


Figure 1. Locations of in situ data used in this study ($N = 180$), circle green symbols represent stations in the Bohai Sea and the Yellow Sea and circle blue ones are the stations in the East China Sea. Square yellow symbols stand for match-ups between in situ measurements and satellite images. Major water masses in summer-half-year are sketched with red lines (after Li et al., 2016). The background seawater depth (The GEBCO_2014 Grid, version 20150318) is obtained from GEBCO (http://www.gebco.net/).

Water masses and current systems are complex in the YS and the ECS during the sampling time, including coastal currents in the east, open ocean water in the west, and mixed water between them (Figure 1). Changjiang River carries large amount of fresh water into the sea, forming Changjiang diluted water (CDW) with low temperature, salinity, and more nutrients. The Shandong Peninsula coastal current comes out of the BS and flows into the YS bypassing the Shandong Peninsula. The Yellow Sea coastal current water flows southward along Jiangsu province and turns southeastward into the ECS, while the Zhe-min coastal current water flows through Zhejiang coastline. By contrast, the Kuroshio water is more powerful and characterized by higher temperature and salinity. After entering the ECS, it flows northeastward along the continental slope. Taiwan Strait warm current flows through Taiwan Strait and moves forward to the north, together with Kuroshio intrusion, forming Taiwan warm current (Chen et al., 1995; Li et al., 2006, 2016; Lie et al., 2001; Quan et al., 2013; Zhang et al., 2008).

The study is based on samples collected on two research cruises in the BS and the YS (31∼40°N, 118∼126°E, 28 April to 18 May) and the ECS (22∼30°N, 121∼126°E, 22 May to 11 June) in 2014 (Figure 1). Water samples for HPLC pigment concentration were collected with Niskin bottles attached to the conductivity-temperature-depth profiler (CTD, Seabird 911) rosette. During the analyses, five outliers of HPLC pigment concentrations were removed and 180 samples were acquired for analyzing surface distributions in the study. Among them, algae bloom was observed at station 4-0-0 (27.76°N, 122.55°E) and quite high chlorophyll concentration was obtained at station H40 (32°N, 124.99°E). To avoid the effect of bloom

Table 1
Symbols and Definitions

Symbol	Description	Units
C_E	The estimation of chlorophyll-a concentration using seven diagnostic pigments (equation (3))	mg m^{-3}
C_{HPLC}	Chlorophyll-a concentration derived from HPLC method (equation (5))	mg m^{-3}
$F_m/\ F_n/\ F_p$	Fraction of microplankton/ nanoplankton/ picoplankton (equations (4), (5), and (6))	Dimensionless
$C_m/\ C_n/\ C_p$	Concentration of microplankton/ nanoplankton/ picoplankton (equations (9), (10), and (11))	mg m^{-3}
$C_{n,p}$	Concentration of combined of nanoplankton and picoplankton (equation (8))	mg m^{-3}
$P_m/\ P_n/\ P_p$	Percentage of microplankton/ nanoplankton/ picoplankton (equations (12), (13), and (14))	%
$C_{n,p}^m$	Asymptotic maximum values for combined nanoplankton and picoplankton (equation (8))	mg m^{-3}
C_p^m	Asymptotic maximum values for picoplankton (equation (10))	mg m^{-3}
$D_{n,p}$	Fraction of total chlorophyll in combined nanoplankton and picoplankton as total chlorophyll tends to zero (equation (8))	Dimensionless
D_p	Fraction of total chlorophyll in picoplankton as total chlorophyll tends to zero (equation (10))	Dimensionless
r	Pearson linear correlation coefficient	Dimensionless
p	p Value	Dimensionless
δ	Bias between concentrations or percentages from measured and estimated data (equation (16))	mg m^{-3} or %
MAE	Mean absolute error between concentrations or percentages from measured and estimated data	mg m^{-3} or %
RMSE	Root mean squared error between concentrations or percentages from measured and estimated data	mg m^{-3} or %
MAE%	Relative mean absolute error between concentrations or percentages from measured and estimated data (equation (17))	%

water on the algorithm (details in section 4.4), these two samples were excluded and 178 samples were left in model reparameterization and validation processes. Underway data such as temperature and salinity were measured by CTD in the BS, the YS, and the ECS as well.

2.2. Satellite Data

The Geostationary Ocean Color Imager (GOCI) is a geostationary-orbiting ocean color sensor with the coverage including the YS and the ECS. The spatial resolution of GOCI image is 500 m and 8 images (one image per hour) are available during daytime (Ryu et al., 2011), not only increasing numbers of image over coastal ocean affected by cloud coverage, but also providing a capability to map short-time scale variation in the entire continental shelf sea of China. In this study, L1B data were obtained from the KOSC (Korea Ocean Satellite Center) website (http://kosc.kiost.ac/eng/). The surface chlorophyll-a concentration products were calculated through the GOCI Data Processing System (GDPS version 1.4.1), in which KOSC standard was the method for atmospheric correction and YOC (Yellow Sea Large Marine Ecosystem Ocean Color Work Group) was the method for chlorophyll-a concentration products. Based on the empirical algorithm applied for coastal waters from Tassan (1994), parameters of the algorithm YOC were optimized for the YS and ECS (Siswanto et al., 2011). The YOC algorithm is:

$$Chla_{yoc}=10^{\left(0.342-2.511*\log_{10}(R)-0.277*\log_{10}^{2}(R)\right)} \tag{1}$$

$$R=\left(\frac{R_{rs}(443)}{R_{rs}(555)}\right)\left(\frac{R_{rs}(412)}{R_{rs}(490)}\right)^{-1.012} \tag{2}$$

Since a whole GOCI image consists of 4-by-4 subimages taken by one camera, mosaic edge effects of top-of-atmosphere radiance from L1B products at some bands (e.g., Band 1 at 412nm) are the cause of spatial discontinuity of the chlorophyll-a concentration products. Therefore, in this study we utilized neighboring pixel interpolation method to reprocess the neighboring slots in order to avoid discontinuity. After top-of-atmosphere of radiance products which have slot margin effects are exported from the GDPS, pixels values on either sides of the slot margins are extracted to establish linear regression equations and calculate

Table 2
Major Diagnostic Pigments Used for Classification of PSCs From Brewin et al. (2010)*,* Hirata et al. (2011)*, and This Study, Along With the Taxonomic or Biogeochemical Significance* (Ras et al., 2007)

Diagnostic pigments	Designation	Brewin et al. (2010)	Hirata et al. (2011)	In this study
Fucoxanthin (Fuco)	Diatoms	Micro	Micro/Nano	Micro
Peridinin (Per)	Dinoflagellates	Micro	Micro	Micro
Alloxanthin (All)	Cryptophytes	Nano	Nano	Nano
19′-but-fucoxanthin (But)	Pelagophytes	Nano	Nano	Nano
19′- hex-fucoxanthin (Hex)	Prymnesiophytes	Nano/Pico	Nano/Pico	Nano/Pico
Chlorophyll-b (Chl-*b*)	Chlorophytes	Pico	Nano	Nano
Divinyl chlorohphyll-b	Prochlorophytes	Pico		
Zeaxanthin (Zea)	Cyanobacteria Prochlorophytes	Pico	Pico	Pico

slopes. Through the judgment of chlorophyll-a concentration, slots with outliers are calculated based on slopes. Then without mosaic edge effects, the processed top-of-atmosphere radiance L1B products are imported to the GOCI images.

Considering the frequency of the GOCI no-cloud overpassing and in situ survey time, the time window of match-ups was set to ±3.5 h. Under the condition of picking images with the shortest time interval, 31 sampling sites were matched to GOCI data in the study area (Figure 1). The matched images were acquired on 30 April, 1 May, 2 May, 3 May, 6 May, 9 May, 12 May, and on 9 June. Average values of 3-by-3 pixel box from the GOCI chlorophyll-a concentration products were regarded as the matched data. Eight images on 7 April 2013 were provided to show the GOCI-derived PSCs and the diurnal variation.

For comparison with physical variables on 7 April 2013, SNPP VIIRS (Suomi NPP Visible Infrared Imaging Radiometer Suite) daily composite sea surface temperature data at 4 km resolution was obtained from OceanColor website (https://oceancolor.gsfc.nasa.gov/).

2.3. Laboratorial Determination of Phytoplankton Size Classes

Laboratorial determination of PSCs was based on the high performance liquid chromatography (HPLC) method. Water samples (100∼2,000 mL) from Niskin bottles were filtered through the Whatman GF/F Glass Microfiber Filters (pore size 0.7 μm, diameter 25 mm), and the filters were kept in the aluminum foil and frozen in liquid nitrogen. Using a Shimadzu LC-20A high-performance liquid chromatography system (Kyoto, Japan), pigment concentrations in the ECS were processed using the method described by Wang et al. (2016). Detailed instrumentation and methodology of samples processing for HPLC pigments in the BS and the YS can be found in Zhang et al., (2016). Twenty phytoplankton pigments were measured, including chlorophyll-c3, chlorophyllide-a, chlorophyll-c2, peridinin, 19-but-fucoxanthin, fucoxanthin, neoxanthin, prasinoxanthin, 19-hex-fucoxanthin, violaxanthin, diadinoxanthin, alloxathin, diatoxanthin, zeaxanthin, lutein, chlorophyll-b, DV-chlorophyll-a, chlorophyll-a, α-carotene, and β-carotene. Pigments or pigment groups can be assigned to individual phytoplankton species, thus characterizing PSCs indirectly. We utilized seven diagnostic pigments (i.e., fucoxanthin, peridinin, alloxanthin, 19-but-fucoxanthin, 19-hex-fucoxanthin, chlorophyll-b and zeaxanthin) considering the differences in phytoplankton species of each study area (Brewin et al., 2010; Hirata et al., 2011) (Table 2).

In the global marine system, Devred et al. (2011) and Hirata et al. (2011) assigned part of the pigment fucoxanthin to the nanoplankton group by involving pigments 19′-but-fucoxanthin and 19′- hex-fucoxanthin in fucoxanthin adjustment when chlorophyll-a concentration was low (0.25 mg m^{-3}), because fucoxanthin

Table 3
Retrieved Parameter Values Derived From Fitting the Three-Component Model to In Situ Pigment Data From the YS and the ECS

Study	$C^m_{n,p}$	$D_{n,p}$	C^m_p	D_p	MAE[a]	MAE[b]
This study	0.329	1.000	0.052	0.914	0.127	0.045

[a] MAE was the mean absolute error between in situ and modeled $C_{n,p}$. [b] MAE was the mean absolute error between in situ and modeled C_p.

was also present in prymnesiophytes and chrysophytes. Considering most water samples were higher than 0.25 mg m^{-3} and diatoms had absolute advantage in the climatic spring in the BS, the YS, and the ECS (Gao et al., 2003; Guo et al., 2014), we assumed that fucoxanthin was a representative for microplankton. Brewin et al. (2010) combined pigments chlorophyll-b and divinyl chlorophyll-b as a whole and treated them as the diagnostic pigments of picoplankton. However, chlorophytes were one of the major composition of nanoplankton in the study area (Gao et al., 2013; Song et al., 2017). Therefore, in this study, chlorophyll-b was regarded as a biomarker of nanoplankton which has been utilized by Hirata et al. (2011). Zeaxanthin and alloxanthin were much lower in the study area, comparing to the other five diagnostic pigments. Even though values were small, these two pigments were detected in most stations through a more sensitive method developed for the purpose of decreasing the detection limit (Zhang et al., 2016).

The estimation of chlorophyll-a concentration (C_E) as proposed by Vidussi et al. (2001) and later refined by Uitz et al. (2006) can be inferred as

$$C_E = 1.41Fuco + 1.41Per + 1.27Hex + 0.6All + 0.35But + 1.01Chl - b + 0.86Zea \tag{3}$$

According to Brewin et al. (2010) and Hirata et al. (2011), the fractions (F) of the size-fractionated chlorophyll-a concentrations can be estimated as

$$F_m = \frac{1.41 \cdot (Fuco + Per)}{C_E} \tag{4}$$

$$F_n = \begin{cases} \dfrac{12.5 \cdot C_{HPLC} \cdot 1.27 \cdot Hex + 1.01 \cdot Chl - b + 0.35 \cdot But + 0.6 \cdot All}{C_E} & , C_{HPLC} < 0.08mg/m^3 \\ \dfrac{1.27 \cdot Hex + 1.01 \cdot Chl - b + 0.35 \cdot But + 0.6 \cdot All}{C_E} & , C_{HPLC} > 0.08mg/m^3 \end{cases} \tag{5}$$

and

$$F_p = \begin{cases} \dfrac{(-12.5 \cdot C_{HPLC} + 1) \cdot 1.27 \cdot Hex + 0.86 \cdot Zea}{C_E} & , C_{HPLC} < 0.08mg/m^3 \\ \dfrac{0.86 \cdot Zea}{C_E} & , C_{HPLC} > 0.08mg/m^3 \end{cases} \tag{6}$$

The subscripts m, n, and p refer to micro, nano, and picoplankton, respectively. C_{HPLC} represents chlorophyll-a concentration derived from in situ HPLC pigment data. C_{HPLC} and C_E were in good agreement, with a correlation coefficient of 0.843 and p-value of < 0.001. Size-fractionated percentages could be calculated by multiplying fractions by 100, and size-fractionated concentrations could be calculated by multiplying fractions by the C_{HPLC}.

2.4. Estimation of Phytoplankton Size Classes

2.4.1. Three-Component Model of PSCs

Brewin et al. (2010) developed a group of equations based on an underlying conceptual model (Sathyendranath et al., 2001) that was used to quantify the relationship between chlorophyll-a concentration and fractional contribution to chlorophyll-a for each size class. These equations were extended by Brewin et al., (2014, 2015), where total and size-fractionated concentrations are obtained from

$$C_{HPLC} = C_m + C_n + C_p \tag{7}$$

$$C_{n,p} = C^m_{n,p}\left[1 - \exp\left(-\frac{D_{n,p}}{C^m_{n,p}} C_{HPLC}\right)\right] \tag{8}$$

$$C_m = C_{HPLC} - C_{n,p}, \tag{9}$$

$$C_p = C^m_p\left[1 - \exp\left(-\frac{D_p}{C^m_p} C_{HPLC}\right)\right] \tag{10}$$

and

$$C_n = C_{n,p} - C_p \tag{11}$$

where C_{HPLC} is the sum of micro (C_m), nano (C_n), and picoplankton (C_p) chlorophyll-a concentration, and $C_{n,p}$ is the sum of nano and picoplankton. $C^m_{n,p}$ and C^m_p are the asymptotic maximum values for the classes whose sizes are smaller than 20 and 2 μm, respectively. Similarly, $D_{n,p}$ and D_p represent size-fractionated chlorophyll-a concentrations as total chlorophyll-a concentrations tends to zero. Therefore, $D_{n,p}$ and D_p are constrained to be less than or equal to one. The percentage of each PSCs (P_m, P_n, P_p, and $P_{n,p}$) to the C_{HPLC} can be calculated by dividing the size-fractionated chlorophyll-a concentration by the total and multiplying by 100, which are

$$P_m = \frac{C_m}{C_{HPLC}} \cdot 100 \tag{12}$$

$$P_n = \frac{C_n}{C_{HPLC}} \cdot 100 \tag{13}$$

$$P_p = \frac{C_p}{C_{HPLC}} \cdot 100 \tag{14}$$

and

$$P_{n,p} = \frac{C_{n,p}}{C_{HPLC}} \cdot 100 \tag{15}$$

2.4.2. Model Reparameterization and Algorithm Improvement

In this study, we reparameterized the three-component model of PSCs based on in situ pigment measurements from the BS, the YS, and the ECS. According to section 2.1 and 2.2, 31 satellite match-ups of 178 surface samples were removed, leaving 147 samples for parameter establishment. The unknown model parameters $C^m_{n,p}$, C^m_p, $D_{n,p}$, and D_p were obtained by fitting the equations (14) and (15) using nonlinear least-square regressions (MATLAB R2014b, Curve Fitting Tool). The newly obtained parameters are shown in Table 3.

2.5. Error Tests

Different parameters of error tests were used to compare (1) the estimation of size-fractionated concentration and percentage from the different algorithms and in situ measurements; and (2) the size-fractionated concentration derived from satellite images and in situ measurements. These parameters include Pearson linear correlation coefficient (*r*), *p*-value (*p*), and the bias (δ) which was calculated by

$$\delta = \frac{1}{N}\sum_{i=1}^{N}\left(X_{i,A} - X_{i,B}\right) \tag{16}$$

where N is the number of samples and X is variable derived from A and B methods, respectively. In section 3.3, estimation from satellite data were regarded as A, and B was in situ HPLC measurements. Besides, mean absolute error (MAE) measures the average magnitude of errors in a set of comparison, while the root mean squared error (RMSE) represents the sample standard deviation of the differences between predicted values and observed values. Relative mean absolute error (MAE%) between in situ measurements and observations from satellite images is computed according to

$$\mathrm{MAE\%} = \left[\frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_{i,A} - X_{i,B}}{X_{i,A}}\right|\right] \cdot 100 \tag{17}$$

where A is the in situ measurement and B is the estimation from satellite images.

3. Results

3.1. Total Chlorophyll-a Concentration and PSCs Derived From HPLC Pigments

Chlorophyll-a concentration (C_{HPLC}) ranged from 0.027 to 11.298 mg m^{-3} in the entire study area (Figure 2a). In the BS and the YS, the C_{HPLC} varied from 0.531 to 6.631 mg m^{-3}, and the average concentration was 1.849 mg m^{-3}. In the ECS, the C_{HPLC} had a wider range from 0.027 to 11.298 mg m^{-3}, with an average concentration of 0.454 mg m^{-3}. The C_{HPLC} was lower in the central BS, the northwest NYS, the central

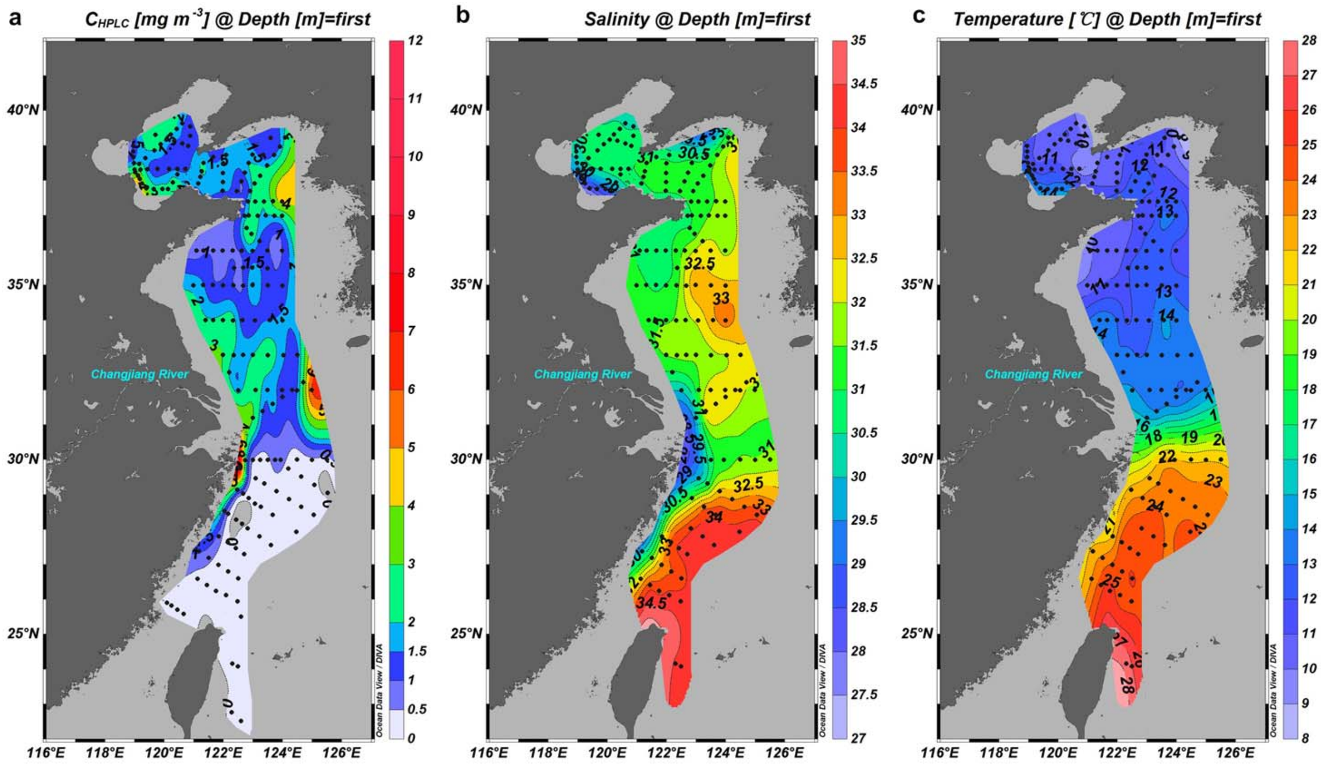


Figure 2. Distribution of surface (a) C_{HPLC}, (b) salinity, and (c) temperature in the Bohai Sea, the Yellow Sea, and the East China Sea during the 2014 cruise investigation.

SYS, and outer continental shelf area of the southeast ECS. By contrast, higher values were found in the coastal BS, the central NYS, coastal currents (i.e., Yellow Sea coastal current and Zhe-min coastal current) and the extension area of the Changjiang Diluted Water (CDW), which carries large quantities of freshwater from mainland and the salinity was less than 32.

Surface salinity and temperature ranged, respectively, from 27.95~33.34 and 8.74~14.79 (°C) in the BS and the YS (Figures 2b and 2c). Compared to temperature, salinity varied obviously from the nearshore area (i.e., northwest area of Shandong Peninsula and southeast area of Liaodong Peninsula) to the offshore area (i.e., central SYS). Higher salinities and temperatures were observed in the southeast of the ECS due to the influence of Kuroshio water and Taiwan warm current which transport high salinity and temperature surface waters to the eastern boundary of ECS, where the surface salinities and temperatures were 29.80~34.46 and 21.58~27.86 (°C), respectively. Salinities and temperatures were much lower in the nearshore area due to coastal currents and the CDW, compared to the offshore area.

The C_{HPLC}, salinity, and temperature had similar patterns of spatial distribution (Figure 2). In the nearshore area where salinity and temperature was lower, C_{HPLC} was higher, compared to the offshore area. With the increase of salinity and temperature, the C_{HPLC} tended to be lower. The C_{HPLC} had a negative relationship with salinity ($N = 35$, $r = -0.594$, $p < 0.001$) and temperature ($N = 35$, $r = -0.460$, $p < 0.01$) in the ECS.

Distribution of size-fractionated concentration of microplankton was similar to that of the C_{HPLC} (Figure 2a), having higher average value than nano and picoplankton, which were 1.116, 0.254, and 0.038 mg m^{-3}, respectively. In the nearshore area, both percentage and concentration of microplankton were higher than those of the other two groups (Figures 3a and 3d). Besides the central SYS, percentages of microplankton were higher than 70% in the BS and the YS. Nanoplankton was distributed extensively, especially in the ECS (Figures 3b and 3e), with the average percentage of 27.48% over the entire study area and 50.77% in the ECS. The percentage of picoplankton constituted the main background in the central SYS and offshore area

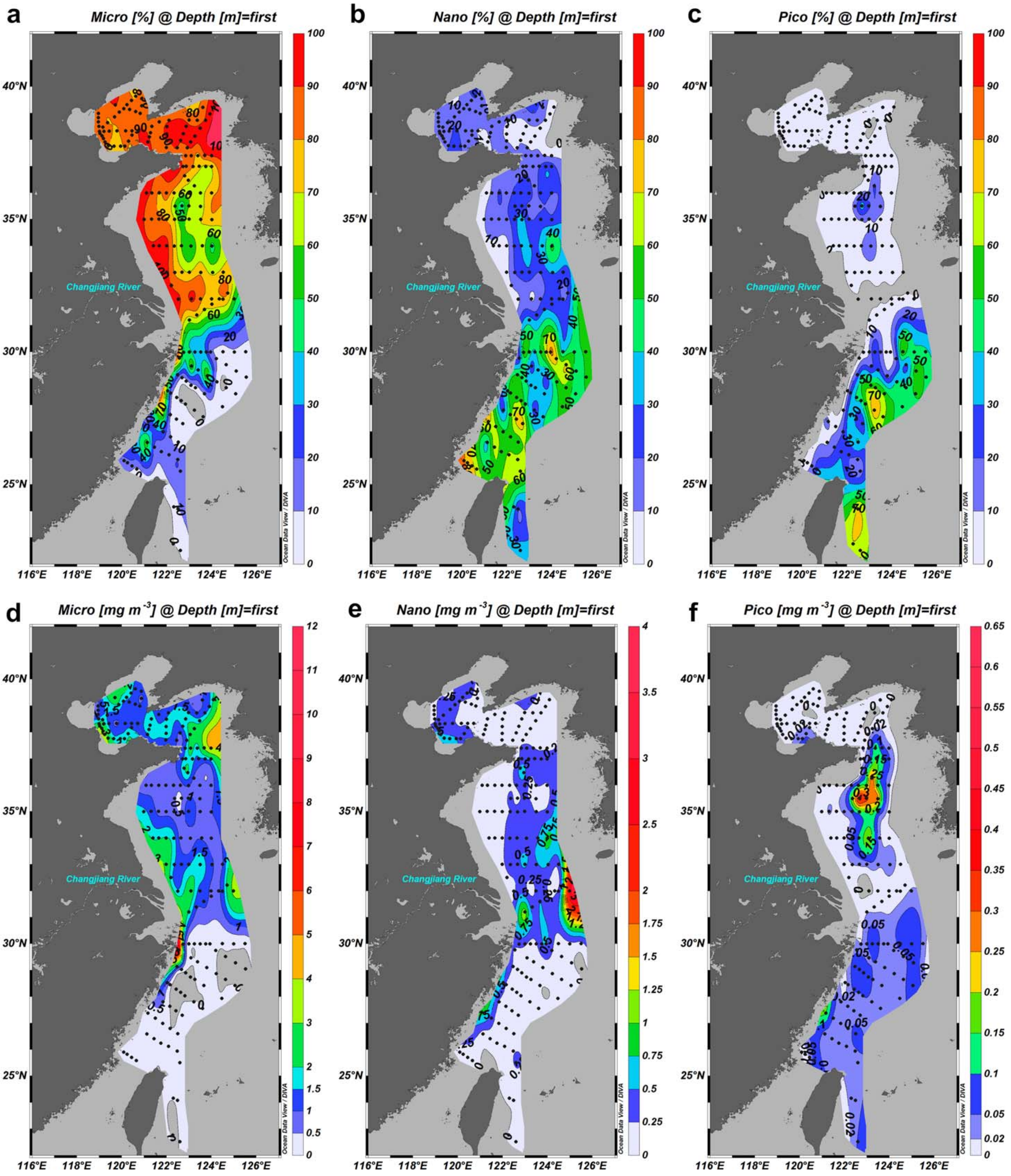


Figure 3. Spatial distribution of size-fractionated (a–c) percentages and (d–f) concentrations of micro, nano, and picoplankton in the Bohai Sea, the Yellow Sea, and the East China Sea during the 2014 cruise investigation.

in the ECS, where the C_{HPLC} was low (Figure 3c). For example, at station TW1–3-1 (24.08°N, 122.50°E) located east of Taiwan Island, the percentage of picoplankton was 78.54%. The concentrations of picoplankton were the lowest, which were no more than 0.1 mg m^{-3} in the most of the study area (Figure 3f).

When comparing size-fractionated percentages with the temperature and salinity, we found that in the BS and the YS, microplankton was negatively correlated with salinity ($N = 123$, $r = -0.361$, $p < 0.001$) and temperature ($N = 123$, $r = -0.379$, $p < 0.001$); nanoplankton was positively correlated with salinity ($N = 123$, $r = 0.275$, $p < 0.01$) and temperature ($N = 123$, $r = 0.362$, $p < 0.001$); and picoplankton had a positive correlation with temperature ($N = 123$, $r = 0.254$, $p < 0.01$). In the ECS, microplankton was negatively correlated with salinity ($N = 35$, $r = -0.414$, $p < 0.05$); picoplankton had positive correlation with salinity ($N = 35$, $r = 0.416$, $p < 0.05$) and temperature ($N = 35$, $r = 0.356$, $p < 0.05$).

3.2. PSCs Predicted by the Improved Algorithm

The improved algorithm (details in section 2.4.2) was applied to estimate concentrations and percentages of PSCs in the BS, the YS, and the ECS (Figure 4). Furthermore, estimates by other two algorithms from Brewin et al. (2010) and Hirata et al. (2011) are also shown in the plot, as a comparison. The size-fractionated percentages were smoothed with the 5-point running mean filer to improve the signal-to-noise ratio and make it easier to show changing regularities. The improved algorithm (red lines) fits the measurements well (Figure 4). The improved parameters were effective in predicting the trends in size-fractionated percentages of the PSCs in the BS, the YS, and the ECS, especially when chlorophyll-a concentration was high (Figures 4a–4d). As for concentrations, the improved algorithm had good agreements in estimating microplankton and nanoplankton. Because of the extremely low values of picoplankton in the BS and the YS (Figures 3f and 4h), the improved algorithm had lower precision in estimating picoplankton concentration when chlorophyll-a concentration was high, and its trend with the change of total concentration requires further study.

For the purpose of discussing the applicability of two existed methods (details in section 4.1), PSCs derived from three component model (Brewin et al., 2010) and from empirical equations (Hirata et al., 2011) were shown as well. Instead of estimating size-fractionated concentration (Brewin et al., 2010), Hirata et al. (2011) expressed the percentages of PSCs (P_m, P_n, and P_p) instead of concentration. The equations are

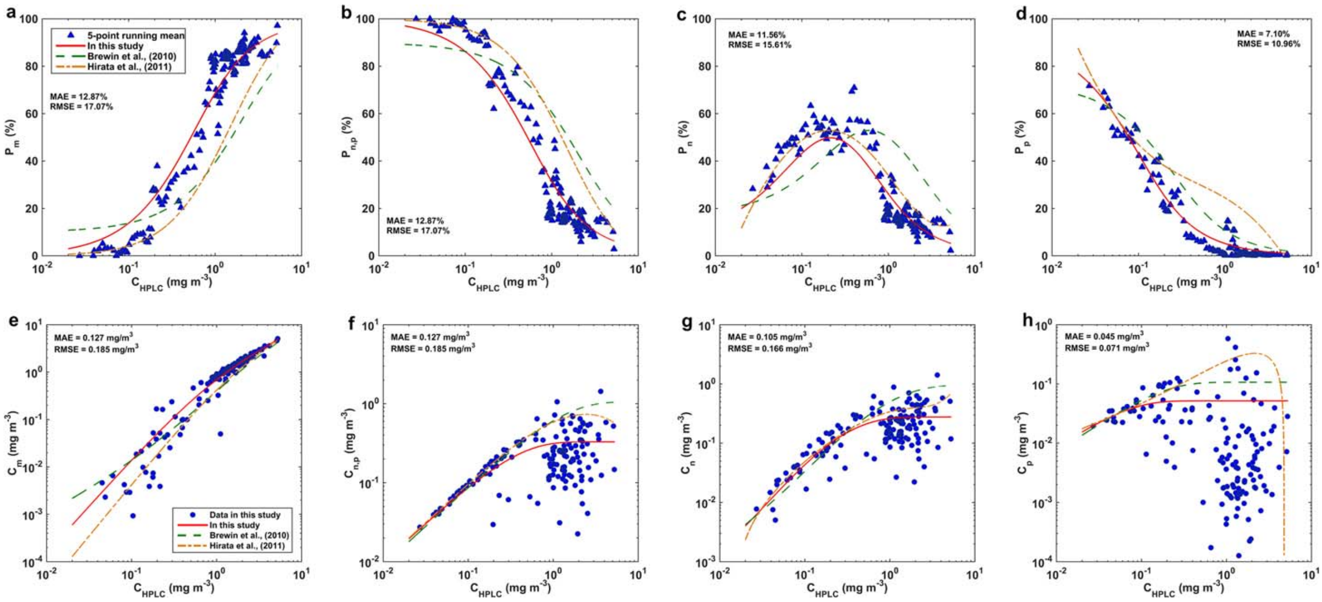


Figure 4. Relationships between size-fractionated (a–d) percentages and (e–h) concentrations, all as a function of C_{HPLC}. Measurements included in situ data (blue dots) and 5-point running mean data (blue triangles), $N = 149$. Estimations were predicted by the improved algorithm in this study (red lines), by Brewin et al. (2010) (green dotted lines) and by Hirata et al. (2011) (yellow dash dot lines). MAE and RMSE were calculated between in situ measurements and estimations from the improved algorithm.

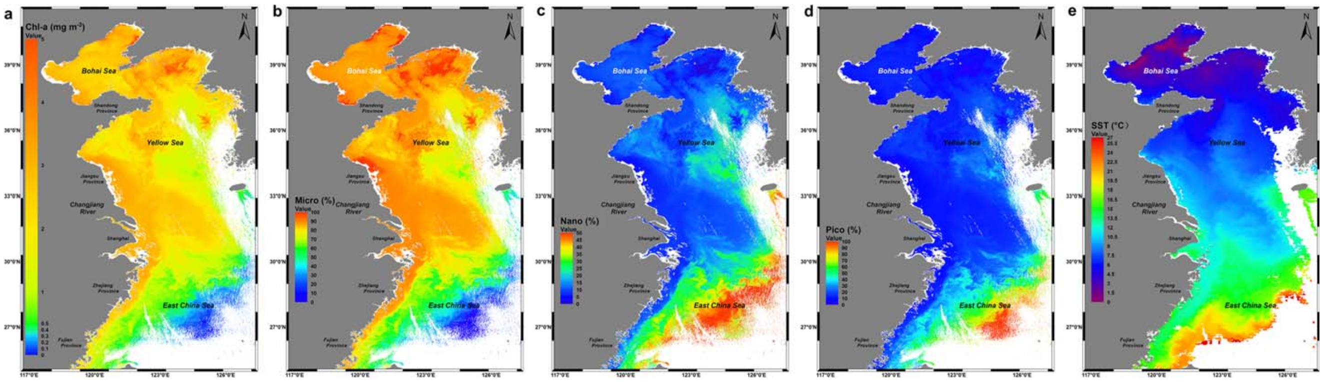


Figure 5. GOCI-derived chlorophyll-a concentration (a, Chl-a, in unit of mg m^{-3}) and percentage (%) of each PSCs (b–d) in the Bohai Sea, the Yellow Sea, and the East China Sea, which were estimated by the improved algorithm. The GOCI image was taken at 10:28 on 7 April 2013 (Beijing time). Daily sea surface temperature image of SNPP VIIRS (e, in unit of °C) on the same day.

$$P_m = (0.9117 + \exp(-2.7330 \cdot \log_{10}(C_{HPLC}) + 0.4003)^{-1} \cdot 100 \quad (18)$$

$$P_n = 100 - P_m - P_p \quad (19)$$

and

$$P_p = \left\{-[0.1529 + \exp(1.0306 \cdot \log_{10}(C_{HPLC}) - 1.5576)]^{-1} - 1.8597 \cdot \log_{10}(C_{HPLC}) + 2.9954\right\} \cdot 100 \quad (20)$$

the concentration of each PSCs can be calculated by multiplying its percentage by the total, which are C_m, C_n and $C_{p.}$

3.3. GOCI-Derived PSCs and Validation

All the GOCI images with less cloud coverage from April to June from 2011 to 2017 were processed, of which images on 7 April 2013 had the best image quality and the largest available areas. Figure 5 showed the spatial distributions of the chlorophyll-a concentration and the estimation of size-fractionated percentages by the improved algorithm, which were derived from the GOCI image at 10:28 (center time of the scene, Beijing time) on 7 April, 2013. Daily SNPP VIIRS sea surface temperature on the same day was shown as well.

Figure 5a illustrated that higher values of chlorophyll-a concentration distributed in the northern BS, northern NYS, central SYS, and along the nearshore area of the YS and the ECS. While in the offshore area, such as the eastern SYS and the southeastern ECS, concentrations were much lower. The distribution pattern of microplankton percentage was similar to the total concentration. Except for the offshore area of the ECS, microplankton was dominant in the study area where chlorophyll-a concentrations were higher than 1 mg m^{-3} (Figure 5b). Nanoplankton had lower percentage in the BS and nearshore area of the YS and the ECS, while in the offshore area of the SYS and the ECS, its percentage ranges from 20% to 50% (Figure 5c). By contrast, picoplankton only had higher distribution in the offshore area and percentage in these areas tended to be greater than 40% (Figure 5d). Figure 5e showed that temperature rose from north to south and from west to east in the study area, due to the influence of Taiwan warm current, the Kuroshio and its brunches. Temperature in the offshore area of the SYS and the ECS was higher, leading to lower percentage of microplankton and higher percentages of nano and picoplankton. In comparison, distribution of total chlorophyll-a concentration and microplankton percentage had clear negative correlation with temperature, while nanoplankton and picoplankton percentages were positively correlated to temperature. This estimated distribution of PSCs' derived from the GOCI image with the improved algorithm had a good consistency with the in situ results presented in section 3.1, especially for the BS and the YS, where both the image time and the investigation time were in April. However, the shaded area caused by the shape of clouds, leading to challenges in representing full knowledge of PSCs in the study area, especially for the edges. Additionally, there is a temporal limitation in observing continuous changes with one image.

Table 4
Comparison of Chlorophyll-a Concentration Between In Situ Measurements and Observations Derived From Satellite Data Using Different Chlorophyll-a Concentration Algorithms (N = 31)

	δ (mg m^{-3})	MAE (mg m^{-3})	RMSE (mg m^{-3})	MAE% (%)
OC2	0.368	0.720	0.928	69.00
YOC (preprocessing)	0.408	0.698	0.974	55.75
YOC (postprocessing)	−0.050	0.626	0.824	43.83

The improved algorithm was then applied to GOCI images to validate the estimation accuracy of PSCs through independent match-ups between satellite data and in situ measurements. During the investigation time, 31 in situ data were matched to the GOCI satellite images. In the GDPS (version 1.4.1), there are three algorithms for chlorophyll-a concentration retrieval, of which OC2 (ocean chlorophyll 2 algorithm) and YOC algorithms were selected for comparison. Based on the data set from oceanic waters, OC2 proposed by O'Reilly et al. (1998) was the default for chlorophyll analysis in the GDPS, whereas the YOC data set from Siswanto et al. (2011) covered the YS and the ECS. Chlorophyll-a concentrations derived directly from GDPS using YOC algorithm were named YOC (preprocessing). Considering the mosaic edge effects caused by Band 1 at 412 nm in L1B products, a neighboring pixel interpolation method was used (details in section 2.2) in processing chlorophyll-a concentration, thus using YOC (postprocessing) to distinguish from YOC (preprocessing). Statistically, the YOC algorithm was better than the OC2 algorithm in the study area, as it had lower bias, MAE, RMSE, and MAE% (Table 4).

Based on the images with neighboring pixel interpolation processing, independent satellite and in situ match-up data (N = 31) were compared, including chlorophyll-a concentration, size-fractionated concentration, and size fractionated percentages. When calculating errors, extremely low in situ size-fractionated concentrations which less than 0.01 mg m^{-3} were eliminated, leaving 29, 31, and 18 match-ups for micro, nano, and picoplankton. Figure 6a showed that satellite-derived total and microplankton chlorophyll-a concentration had good agreements with in situ measurements, with MAE% of 43.83% and 49.09%, respectively. Biases indicated that the improved algorithm overestimated microplankton concentration and underestimated nanoplankton concentration slightly. As for picoplankton concentration, poor MAE%

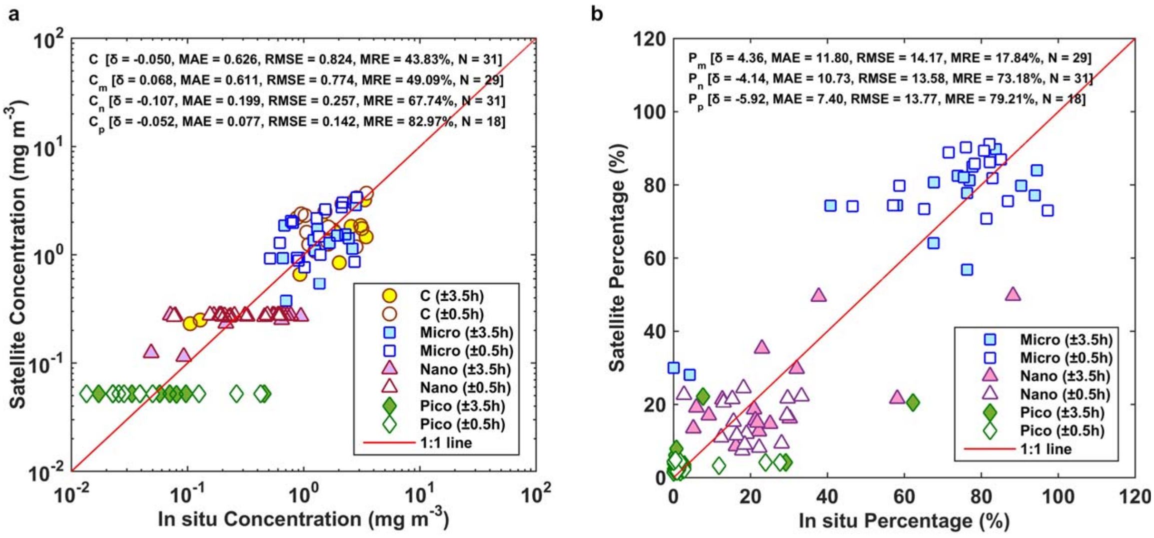


Figure 6. Validation of GOCI estimations using the improved algorithm through match-ups, including (a) total chlorophyll-a concentration (yellow circle) and size-fractionated chlorophyll-a concentration of micro (blue square), nano (purple triangle) and picoplankton (green diamond), and (b) size fractionated percentages of micro, nano, and picoplankton. Match-ups within ±3.5 h are shown in filled symbols, while match-ups within ±0.5 h are shown in empty symbols. Solid lines represent the 1:1 lines. The units of δ, MAE, RMSE are mg m^{3} in Figure 6a, and % in Figure 6b.

(82.97%) was obtained compared to the other groups, which was caused by the limited coverage of match-ups. Match-ups mostly distributed in the BS and the YS (Figure 1), where microplankton was dominant during the investigation and in situ picoplankton concentrations were low (Figures 3c and 3f). By contrast, picoplankton was abundant in the offshore ECS, however, cloudy weather was common during the ECS cruise and only two match-ups were found in the ECS. Figure 6b showed that the improved algorithm had a better accuracy in estimating the microplankton percentage (MAE% = 17.84%), while it underestimated percentages of nano and picoplankton, compared to the in situ measurements. Moreover, match-ups within ±0.5 h (empty symbols) were shown in the Figure 6 as well, and with this strict time window, the improved algorithm had a better accuracy in estimating PSCs, with the RMSE for size-fractionated chlorophyll-a concentrations and percentages are 0.848, 0.183, 0.139, and 14.48, 10.60, 10.14, respectively.

3.4. Diurnal Surface Variation of GOCI-Derived Chlorophyll-a and PSCs

Diurnal variation of chlorophyll-a and PSCs were derived from GOCI images on 7 April, 2013 from 8:28 to 15:28 (center time of the scene, Beijing time). Based on the available coverage of eight images, a transect and four regions of interest were analyzed, which are ROI-1 (center coordinate, 34.33°N, 123.15°E), ROI-2 (29.96°N, 122.56°E), ROI-3 (29.80°N, 122.94°E), and ROI-4 (29.62°N, 123.39°E), 11 pixels by 11 pixels (Figures 7a and 7b). ROI-1 is located far away from the coast, less affected by coast currents and contains clearer water. The transect in the southeastern area of Zhoushan Island is approximately parallel with the direction of tidal currents and vertical to isobaths. Due to the impacts of different water masses and current systems,

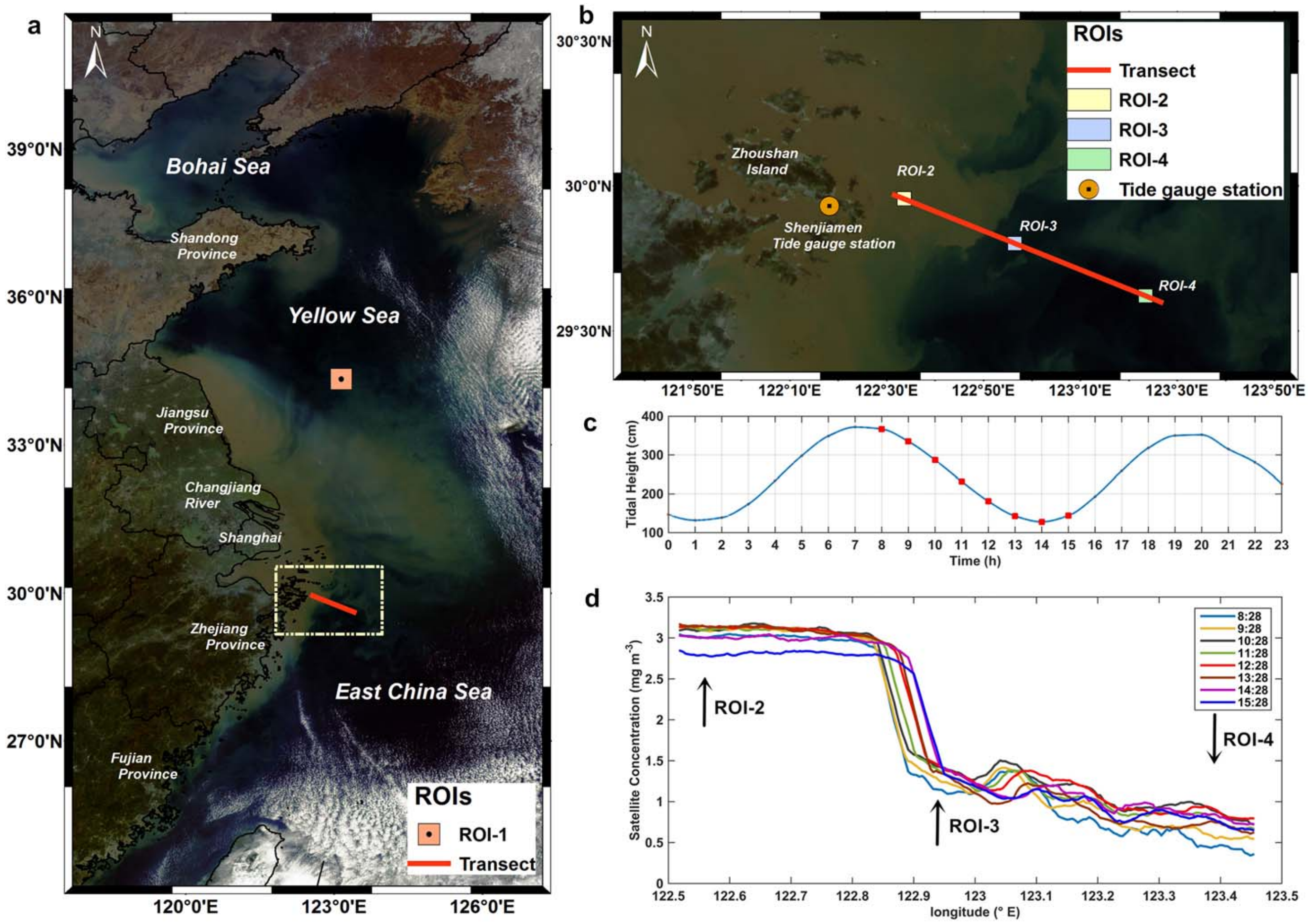


Figure 7. A transect and four regions of interest (ROIs) utilized in diurnal variation (a, b). The base map is a three-band composite true color image acquired at 10:28 (local time, red: Band 6, green: Band 4, blue: Band 2). (c) Tide height from Shenjiamen tide gauge station on 7 April 2013, the red squares represent hourly imaging time of GOCI images. (d) Diurnal variation of GOCI-derived chlorophyll-a concentration of the transect.

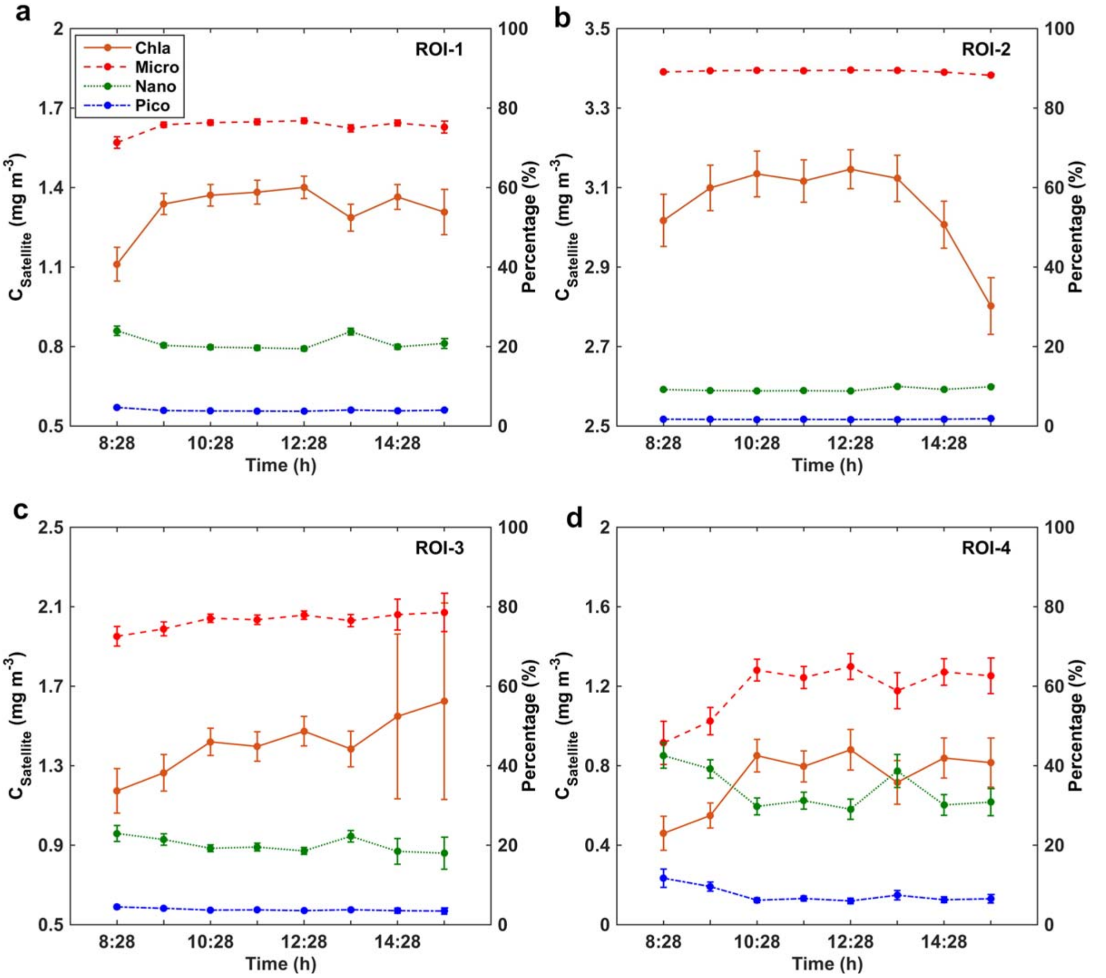


Figure 8. Diurnal surface variations of GOCI-derived chlorophyll-a concentration (mg m^{-3}) and percentage of PSCs (%) are included in (a) ROI-1, (b) ROI-2, (c) ROI-3, and (d) ROI-4. The axis on the left gives the concentration (mg m^{-3}) and the axis on the right gives the percentage (%) and deviation bar were included for chlorophyll-a concentration and percentages of PSCs.

chlorophyll-a concentration of the transect decreased obviously from nearshore area to offshore area (Figure 7d). Besides, characteristics of diurnal variation changed with the distance from the shore. Thus, ROI-2, ROI-3, and ROI-4 were extracted, located at nearshore area, the transition region with a steep drop in chlorophyll-a concentration and offshore area, respectively.

GOCI-derived chlorophyll-a values showed a regular pattern of variation during the daytime (Figures 8a, 8b, and 8d). In the morning, the concentration value was the lowest, after that it reached its highest value between 10:00 and 12:00. Then in the afternoon, the concentration tended to decline. However, due to the impact of tides, variation of chlorophyll-a concentration in the afternoon was different in the transition area (i.e., ROI-3). Diurnal variations in four ROIs were as follows.

1. In the ROI-1, chlorophyll-a value varied from 1.11 to 1.40 mg m^{-3} throughout the day and peaked at 12:28 (Figure 8a). Microplankton was the major component throughout the day, while nanoplankton had higher percentages in the morning and afternoon. Picoplankton was no more than 5% throughout the day.
2. Since the ROI-2 received more terrestrial inputs, chlorophyll-a concentration was the highest among four ROIs, ranging from 2.80 to 3.15 mg m^{-3} during the daytime (Figure 8b). Microplankton was dominant, while percentages of nanoplankton and picoplankton were less than 10% during the daytime.

Table 5

Comparison of Size-Fractionated Percentages and Concentrations Between In Situ Measurements and Estimations From Three Methods in Different Study Areas

			Percentage (%)			Concentration (mg m^{-3})		
Method	Study area	Error tests	Micro	Nano	Pico	Micro	Nano	Pico
Brewin et al. (2010)	Entire area	MAE	25.36	23.53	10.33	0.337	0.288	0.079
	N = 147	RMSE	29.87	26.03	13.33	0.434	0.367	0.096
Hirata et al. (2011)	Entire area	MAE	21.24	13.41	17.44	0.271	0.135	0.185
	N = 147	RMSE	27.07	16.48	19.13	0.352	0.187	0.221
In this study	Entire area	MAE	12.87	11.56	7.10	0.127	0.105	0.045
	N = 147	RMSE	17.06	15.61	10.96	0.185	0.166	0.071
In this study	BS and YS	MAE	10.63	7.83	4.22	0.162	0.129	0.058
	N = 93	RMSE	13.22	9.94	7.28	0.208	0.183	0.085
In this study	ECS-1	MAE	23.34	21.74	12.51	0.136	0.129	0.037
	N = 19	RMSE	29.43	26.86	17.04	0.225	0.211	0.047
In this study	ECS-2	MAE	13.13	15.94	11.82	0.027	0.029	0.018
	N = 35	RMSE	16.99	19.23	14.35	0.044	0.046	0.023

Presumably nutrient concentrations would provide enhanced growth conditions for microplankton. Size-fractionated percentages were stable within the day and insensitive to environmental changes.

3. As for the ROI-3, the biggest difference was that chlorophyll-a concentration increased at 14:28 and 15:28 (Figure 8c). As shown in the Figure 7c, all the eight images were acquired during the ebb tide phase when water flowed from nearshore area to offshore area. Under the influence of the tidal currents, high chlorophyll-a concentration nearshore move southeastward to the offshore area, as well as high percentages of microplankton. The increasing temperature and decreasing turbidity led to higher percentages of nanoplankton and picoplankton in this region, compared to the ROI-2.
4. The variation trend of chlorophyll-a concentration in the ROI-4 was similar to that in the ROI-1. However, chlorophyll-a concentrations were much lower, ranging from 0.46 to 0.88 mg m^{-3} (Figure 8d). Both microplankton and nanoplankton were dominant groups and picoplankton percentage was higher than 10% at 8:28. Percentages of nano and picoplankton decreased with the light intensity increased. The intrusion of Taiwan warm current with high temperature and salty water masses made it beneficial for the growth of small size phytoplankton.

In summary, characteristics of spatial distributions of PSCs were affected by environmental factors and diurnal variations were observed as well, which will be discussed later.

4. Discussion

4.1. Applicability of Methods for Deriving PSCs

By comparison, the improved algorithm had an advantage in estimating both percentages and concentrations of PSCs (Figure 4 and Table 5). When chlorophyll-a concentration was low, percentage and concentration of microplankton derived from Hirata et al. (2011) were more suitable to in situ measurements since there was an acceleration in the regression slopes (Figure 4a). However, when chlorophyll-a concentration was high, it underestimated. The three-component model (Brewin et al., 2010) had lower accuracy in estimating microplankton, with a little bit higher MAE and RMSE (Table 5). As for estimating nanoplankton and picoplankton, three methods had differences (Figures 4c, 4d, 4g, and 4h). Both algorithms of Brewin et al. (2010) and Hirata et al. (2011) seemed to overestimate nanoplankton and picoplankton when chlorophyll-a concentration was high, which made it less applicable in the study area. Since the three-component model is based on an underlying conceptual model (Sathyendranath et al., 2001), it was parameterized to fit the study area. As a result, the improved algorithm has a higher estimation accuracy.

Moreover, due to the effects of CDW and Zhe-min coastal current, the ECS was divided into two parts, one is the area where salinity was lower than 32 (ECS-1), the other is the outer continental area where salinity was higher than 32 (ECS-2). Table 5 showed that the improved algorithm had a better accuracy in estimating PSCs in the BS and the YS, since more measurements were utilized in parameterization and environmental factors were homogeneous relatively (Figures 2b and 2c). In the ECS, abundant fresh water and nutrients carried by the coastal current supported the growth of large size phytoplankton (Deng et al., 2008; Huang et al, 2006; Sun et al., 2012), which affected the distribution of PSCs, making the model less robust in the nearshore area (ECS-1). Without the influence of the coastal current, the improved algorithm provided a better estimation in the ECS-2.

4.2. Spatial Distribution of PSCs

GOCI-derived PSCs' distribution showed that microplankton and nanoplankton were generally the major contributors to coastal and transitional regions, and picoplankton was found to be abundant in the oligotrophic regions (Figure 5), which was consistent with in situ investigations in our study (Figure 2), previous studies in the BS, the YS, and ECS (e.g., Fu et al., 2009; Huang et al, 2006; Sun et al., 2002, 2012), and from other study areas (e.g., Arin et al., 2002; Chisholm et al., 1988; Madariaga & Orive, 1989; Marañón et al., 2001). Similar distribution of GOCI-derived PSCs was observed from our study (Figure 5) and the study of Sun et al. (2017). However, slight differences existed in the nearshore BS and SYS where microplankton instead of nano and picoplankton was dominant in our study, which might result from regional data sets for modeling and time scales utilized in the two studies.

Spatial distribution is closely related to the competitive abilities of different PSCs in response to environmental conditions. Figures 2b and 2c showed that salinity and temperature were higher in the central YS and the offshore ECS, due to the influence of Yellow Sea warm current, Taiwan warm current, and Kuroshio (Li et al., 2016; Quan et al., 2013; Yu et al., 2005). Results in section 3.1 revealed that temperature and salinity were correlated to the phytoplankton size structure, which might serve to explain why nano and picoplankton had higher concentrations and percentages offshore. Similar correlations were observed in a previous study in the same season and area (Sun et al., 2012). However, obvious relationships between PSCs and temperature or salinity in the YS and the ECS were not found from Deng et al. (2008), indicating that environmental factors that determined the PSCs were more than temperature and salinity. Some previous studies have suggested that nutrients have positive correlations with larger phytoplankton and negative correlations with smaller phytoplankton in the YS and the ECS (Deng et al., 2008; Sun et al., 2012). Indeed, higher dissolved inorganic nitrogen, phosphate, and silicate along coastlines and in the Changjiang Estuary and its adjacent areas were observed (Gong et al., 2003; Guo et al., 2014; Liu et al., 2015; Wang et al., 2003; Zhang et al., 2007), supporting our results that micro and nanoplankton were distributed with higher percentages in nearshore areas (Figure 3). With the enhanced solar radiation in April, sea surface temperature increased and the stratification occurred in the central YS (Li et al., 2016; Yu et al., 2005), preventing the supplementary of nutrients in the surface. Availability of nutrients became the major limiting factor for the growth of large-sized phytoplankton in more oligotrophic environments (Marañón, 2015). Due to the surface-area-to-volume ratio, small cells are considered to be more competitive under limiting nutrient conditions (Agawin et al., 2000; Marañón et al., 2013), which could help to explain why percentage and concentration of picoplankton were higher in the central YS (Figures 3c and 3f).

4.3. Diurnal Variation of PSCs

Unlike polar-orbiting ocean color satellites (e.g., SeaWiFs and MODIS), GOCI has higher temporal resolution which can capture diurnal changes. Plenty of researches showed applications in suspended particulate materials (Ge et al., 2015; He et al., 2013; Pan et al., 2018) and phytoplankton (Choi et al., 2014; Lee et al., 2012; Lou & Hu, 2014). The atmospheric correction utilized in this study (i.e., KOSC standard in GDPS system) is the same as that in the researches of diurnal variation of turbidity fronts (Hu et al., 2016a), tidal currents (Hu et al., 2016b), and salinity (Liu et al., 2017).

Hourly variations of chlorophyll-a concentration observed from space were obvious throughout the day in all the ROIs (Figure 8). Maximum chlorophyll-a concentration usually occur around noon, which was similar with the temporal pattern from Lorenzen (1963) and Maulood et al. (1978). Similar trends of variation between photosynthetically available radiation (PAR) and chlorophyll-a concentration among the four ROIs indicated that the diurnal variation of chlorophyll-a concentration might due to the ability of photosynthesis to light by phytoplankton (Figures 8 and 9a). Synchronous GOCI-derived PAR was calculated by

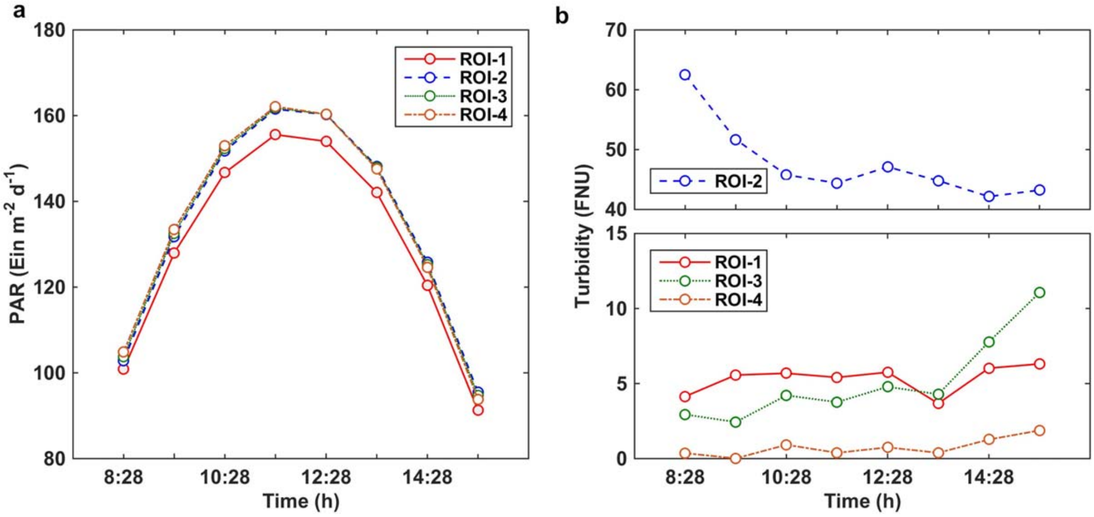


Figure 9. GOCI-derived diurnal variation of (a) photosynthetically available radiation (PAR) and (b) turbidity in the four ROIs.

$$\mathrm{PAR}=\sum_{i=1}^{5}\frac{\lambda_{i+1}-\lambda_i}{hc}\,\frac{\lambda_i E_d(\lambda_i)+\lambda_{i+1}E_d(\lambda_{i+1})}{2} \tag{21}$$

where $\lambda_i = 412$, 443, 490, 555, 660, and 680 nm, $h = 6.626*10^{-34}$ j s, $c = 2.996*10^8$ m s^{-1}, $E_d = L_w/R_{rs}$ and the unit of PAR was convert to Ein m^{-2} d^{-1}. The equation (21) is originated from Cao and Yang (2002).

As for PSCs, variations were much stronger in the offshore area. It is obvious that the percentages of microplankton increased with the increasing radiation, while the picoplankton was opposite (Figures 8a and 8d). Previous study found that micro and nanoplankton required more illuminance intensity than picoplankton for reaching the best growth condition (Sun et al., 2008). Dinoflagellates, such as *Ceratium tripos* and *Ceratium furca* were phototactic and had characteristics of diurnal vertical migration (i.e., upward migration to the surface in the morning and downward migration in the evening), leading higher abundance in the upper layers during the daytime (Blasco, 1978; Jephson & Carlsson, 2009). By contrast, small cells cope better with reduced light conditions, since they are less affected by the package effect (Finkel et al., 2004). Surface chlorophyll and abundance of *Synechococcus* were observed higher in the morning and after evening, while at noon the lower populations could be suffered from photoinhibition because of the high radiation (Mitbavkar & Saino, 2015; Vaulot & Marie, 1999). Since the phytoplankton mentioned above are common species in the ECS (Chen et al., 2006), their responses to the light can well explain the features of surface distribution in this study.

Furthermore, water turbidity is one of important factors for phytoplankton in turbid coastal oceans. Figure 9b showed GOCI-derived hourly turbidity of the four ROIs, which was computed following equations from Dogliotti et al. (2015). Compared to the ROI-1 which was located far from the coast, diurnal variations were more obvious in the ROI-2, ROI-3, and ROI-4. Turbidity was the highest in the ROI-2, where suspended particulate matters were mainly dominated by nonalgal particles. Under influences of tidal currents, coastal currents, and waves, the mechanism of the turbidity variation was complicated, and the relationship between turbidity and chlorophyll-a concentration was not significant. Diurnal variations of PSCs in the ROI-2 were stable and less affected by environmental factors (Figure 8b). The ROI-3 was located in turbid-clear transition zone. Diurnal variation of turbidity might be influenced by the tidal current. It was found in Figure 9b that an obvious increment of turbidity was found in the afternoon, which was corresponding to the ebb tidal phase of 7 April 2013 (Figure 7c). Horizontal transportation of the water mass was the cause of the variation in chlorophyll-a concentration and percentages of PSCs in the afternoon (Figure 8c). Diurnal variation of bloom surface distribution impacted by tidal situation in nearshore waters of ECS was also reported by Lou and Hu (2014). The ROI-4 was situated in the clearer water mostly dominated by high salinity and oligotrophic water mass of Taiwan warm current so that it had the lowest turbidity, where algal particles were the major component of suspended particulate matters. Thus, the fluctuation of turbidity suggested the variation of chlorophyll-a concentration (Figures 8d and 9b). Compared to other ROIs, percentages of PSCs had significant variations in the ROI-4 within a day due to the combined effects of both light intensity and turbidity. In addition to environmental factors mentioned before, intrinsic population processes caused by resuspension of benthic phytoplankton and grazing might also have influences on diurnal variation of PSCs. Consistent and more comprehensive monitoring in the whole water column is required in the future.

4.4. Potential and Limitation of Remote Sensing Estimation

In general, PSCs can be detected by remote sensing through three popular approaches—abundance-based, spectral-based, and ecological-based (Brewin et al., 2011b; IOCCG, 2014; Nair et al., 2008). The three-component model (Brewin et al., 2010) which is regarded as an abundance-based method can be used to estimate PSCs when chlorophyll-a concentration derived from space has been proved to be accurate. The accuracy of classification is completely relied on parameters in the models, thus model parameters are required to improve for different biogeochemical provinces (Devred et al., 2006). In this study, parameters of the model were tuned and validated according to our investigation. However, the abundance-based approach has its limitations that it might not work well in distinguishing algal blooms of different PSCs with the same chlorophyll concentration, since it assumes that larger cells dominate in higher concentrations and smaller cells in low concentrations. Many studies found that microplankton such as dinoflagellates were the most common dominances along coastal waters in the study area (e.g., Dai et al., 2013; Lou & Hu, 2014; Xia et al., 2007; Zhou et al., 2003) which confirmed the assumptions of the abundance-based model. Although there are few reports, there is the possibility that small size or mixed sizes phytoplankton blooms could exist. Therefore, bloom stations were removed in the reparameterization in the study. When analyzing PSCs in blooming conditions utilizing the abundance-based approach, additional environmental knowledge are required to improve the model reliability (Brewin et al., 2010).

In order to be more dependable, recent trend of detecting PSCs concerned more about the comparison and combination of different approaches. Spectral-based approach is a more direct way depending on features that spectrum shape of chlorophyll-specific absorption coefficient or particle backscattering coefficient varies with size structure (Brewin et al., 2011a; Ciotti & Bricaud, 2006; Ciotti et al., 2002; Kostadinov et al., 2009; Loisel et al., 2006; Uitz et al., 2008). Brewin et al. (2011b) has proved that both abundance-based and spectral-based approaches can have similar accuracy, however, comparison results remain to be tested in the BS, the YS, and the ECS. As discussed in the section 4.2, environmental factors are crucial to the distribution pattern of PSCs, and several studies have pointed that some of them, such as light availability and temperature are tightly related to parameters of abundance-based model (Brewin et al., 2015, 2017). The combination of empirical methods and the exploitation of additional environmental data will help to optimize models and obtain retrivals more accurately (Bracher et al., 2017).

5. Conclusions

Large numbers of in situ samples were collected in the entire continental shelf sea of China. Through the HPLC method, seven diagnostic pigments were obtained and assigned to corresponding size classes. In situ PSCs' distributions showed that microplankton was dominant in the BS, the YS, and nearshore ECS, while nanoplankton was the major contribution to the chlorophyll-a concentration in the ECS, and picoplankton had higher proportions in the offshore ECS. Temperature and salinity in the study area affected by continental shelf currents circulation had influences on the spatial distribution of PSCs. Microplankton was negatively correlated with the temperature and salinity, while nano and picoplankton had positive correlations.

We proposed an improvement on the parameterization of the three-component model (Brewin et al., 2010). Using strict match-ups, validation revealed that the improved algorithm had a higher accuracy in estimating PSCs in the BS, the YS, and the ECS, with RMSE of concentrations and percentages been 0.774, 0.257, 0.142, and 14.17, 13.58, 13.77 for micro, nano, and picoplankton, respectively.

GOCI-derived PSCs' spatial distribution was in good agreement with in situ measurements and previous studies, resulting from responses of PSCs to different environmental conditions. Characteristics of diurnal variations of PSCs' distributions in different water types were captured by this high temporal resolution satellite as well. Diurnal variation of PSCs in the offshore area was obvious and mainly affected by the light intensity, leading to maximum microplankton percentage around noon and higher percentages of nano and picoplankton in the morning and afternoon. However, in the nearshore area, diurnal variation of PSCs was relatively slight and water masses such as coastal currents and tides were major influencing factors. The study provides a beneficial approach of a consistently spatial-temporal observation of PSCs for better understanding marine ecological and biogeochemical systems in the entire continental shelf sea of China.

Acknowledgments
This work is supported by National Key R&D Program of China (2016YFE0103200), NSFC projects (41771378), and SKLEC-2016RCDW01. We would like to thank the KORDI/KOSC for providing GOCI data (http://kosc.kiost.ac/eng/) and NASA Ocean Color for providing SNPP VIIRS data (https://oceancolor.gsfc.nasa.gov/). In situ water sample data in the Bohai Sea, Yellow Sea, and East China Sea in 2014 were obtained from the "Dongfanghong 2" and "Kexue 1" cruise campaigns and all the scientists and crew who participated in the field surveys are sincerely appreciated. Data sets utilized in this study has been publicly available at https://figshare.com/s/65f0cbe1ed8932dafeb0. We are grateful to the team of Bangqin Huang from Xiamen University for helping with the laboratory analysis of HPLC pigment concentrations in the East China Sea and Jingjing Zhang from Yantai Institute of Coastal Zone Research for collection and laboratory analysis of HPLC pigment concentrations in the Bohai Sea and Yellow Sea in this work. The authors also greatly appreciate Michael A. St. John for instructive comments on the earlier versions of the manuscript and anonymous reviewers for their constructive comments and suggestions.

References

Agawin, N. S. R., Duarte, C. M., & Agustí, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnology and Oceanography*, *45*(3), 591–600. https://doi.org/10.4319/lo.2000.45.3.0591

Arin, L., Morán, X. A. G., & Estrada, M. (2002). Phytoplankton size distribution and growth rates in the Alboran Sea (SW Mediterranean): Short term variability related to mesoscale hydrodynamics. *Journal of Plankton Research*, *24*(10), 1019–1033. https://doi.org/10.1093/plankt/24.10.1019

Aumont, O., Maier-Reimer, E., Blain, S., & Monfray, P. (2003). An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles*, *17*(2), 1060. https://doi.org/10.1029/2001GB001745

Behrenfeld, M. J., & Boss, E. S. (2014). Resurrecting the ecological underpinnings of ocean plankton blooms. *Annual Review of Marine Science*, *6*(6), 167–194. https://doi.org/10.1146/annurev-marine-052913-021325

Blasco, D. (1978). Observations on the diel migration of marine dinoflagellates off the Baja California coast. *Marine Biology*, *46*(1), 41–47. https://doi.org/10.1007/BF00393819

Bracher, A., Bouman, H. A., Brewin, R. J. W., Bricaud, A., Brotas, V., Ciotti, A. M., et al. (2017). Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development. *Frontiers in Marine Science*, *4*, 55. https://doi.org/10.3389/fmars.2017.00055

Brewin, R. J. W., Ciavatta, S., Sathyendranath, S., Jackson, T., Tilstone, G., Curran, K., et al. (2017). Uncertainty in ocean-color estimates of chlorophyll for phytoplankton groups. *Frontiers in Marine Science*, *4*, 104. https://doi.org/10.3389/fmars.2017.00104

Brewin, R. J. W., Devred, E., Sathyendranath, S., Lavender, S. J., & Hardman-Mountford, N. J. (2011a). Model of phytoplankton absorption based on three size classes. *Applied Optics*, *50*(22), 4535–4549. https://doi.org/10.1364/AO.50.004535

Brewin, R. J. W., Hardman-Mountford, N. J., Lavender, S. J., Raitsos, D. E., Hirata, T., Uitz, J., et al. (2011b). An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. *Remote Sensing of Environment*, *115*(2), 325–339. https://doi.org/10.1016/j.rse.2010.09.004

Brewin, R. J. W., Hirata, T., Hardman-Mountford, N. J., Lavender, S. J., Sathyendranath, S., & Barlow, R. (2012). The influence of the Indian Ocean Dipole on interannual variations in phytoplankton size structure as revealed by Earth Observation. *Deep Sea Research Part II: Topical Studies in Oceanography*, *77–80*, 117–127. https://doi.org/10.1016/j.dsr2.2012.04.009

Brewin, R. J. W., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M., & Hardman-Mountford, N. J. (2010). A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecological Modeling*, *221*(11), 1472–1483. https://doi.org/10.1016/j.ecolmodel.2010.02.014

Brewin, R. J. W., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Airs, R., & Lamont, T. (2015). Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. *Remote Sensing of Environment*, *168*, 437–450. https://doi.org/10.1016/j.rse.2015.07.004

Brewin, R. J. W., Sathyendranath, S., Tilstone, G., Lange, P. K., & Platt, T. (2014). A multicomponent model of phytoplankton size structure. *Journal of Geophysical Research: Oceans*, *119*, 3478–3496. https://doi.org/10.1002/2014JC009859

Brotas, V., Brewin, R. J. W., Sá, C., Brito, A. C., Silva, A., Mendes, C. R., et al. (2013). Deriving phytoplankton size classes from satellite data: Validation along a trophic gradient in the eastern Atlantic Ocean. *Remote Sensing of Environment*, *134*, 66–77. https://doi.org/10.1016/j.rse.2013.02.013

Cao, W., & Yang, Y. (2002). A bio-optical model for ocean photosynthetic available radiation. *Tropic Oceanology*, *21*(3), 47–54.

Cermeño, P., Marañón, E., Rodríguez, J., & Fernández, E. (2005). Large-sized phytoplankton sustain higher carbon- specific photosynthesis than smaller cells in a coastal eutrophic ecosystem. *Marine Ecology Progress Series*, *297*, 51–60. https://doi.org/10.3354/meps297051

Chen, C. T. A., Ruo, R., Paid, S. C., Liu, C. T., & Wong, G. T. F. (1995). Exchange of water masses between the East China Sea and the Kuroshio off northeastern Taiwan. *Continental Shelf Research*, *15*(1), 19–39. https://doi.org/10.1016/0278-4343(93)E0001-O

Chen, J., Huang, B., Liu, Y., Cao, Z., & Hong, H. (2006). Phytoplankton community structure in the transects across east China Sea and Northern South China sea determined by analysis of HPLC photosynthetic pigment signatures. *Advances in Earth Science*, *21*(7), 738–746.

Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B., & Welschmeyer, N. A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature*, *334*, 340–343. https://doi.org/10.1038/334340a0

Choi, J., Min, J., Noh, J. H., Han, T., Yoon, S., Park, Y. J., et al. (2014). Harmful algal bloom (HAB) in the East Sea identified by the Geostationary Ocean Color Imager (GOCI). *Harmful Algae*, *39*, 295–302. https://doi.org/10.1016/j.hal.2014.08.010

Ciotti, A. M., & Bricaud, A. (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. https://doi.org/10.4319/lom.2006.4.237

Ciotti, A. M., Lewis, M. R., & Cullen, J. J. (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. https://doi.org/10.4319/lo.2002.47.2.0404

Cloern, J. E. (2001). Our evolving conceptual model of the coastal eutrophication problem. *Marine Ecology Progress Series*, *210*, 223–253. https://doi.org/10.3354/meps210223

Dai, X., Lu, D., Guan, W., Xia, P., Wang, H., He, P., et al. (2013). The Correlation between Prorocentrum donghaiense Blooms and the Taiwan Warm Current in the East China Sea - Evidence for the "Pelagic Seed Bank" Hypothesis. *Plos One*, *8*(5), e64188. https://doi.org/10.1371/journal.pone.0064188

Deng, C., Yu, Z., Yao, P., Chen, H., & Xue, C. (2008). Size-fractionated phytoplankton in the East China and Southern Yellow Seas and its environmental factors in autumn 2000. *Periodical of Ocean University of China*, *38*(5), 791–798

Devred, E., Sathyendranath, S., Stuart, V., Maass, H., Ulloa, O., & Platt, T. (2006). A two-component model of phytoplankton absorption in the open ocean: Theory and applications. *Journal of Geophysical Research*, *111*, C03011. https://doi.org/10.1029/2005JC002880

Devred, E., Sathyendranath, S., Stuart, V., & Platt, T. (2011). A three component classification of phytoplankton absorption spectra: Application to ocean-color data. *Remote Sensing of Environment*, *115*(9), 2255–2266. https://doi.org/10.1016/j.rse.2011.04.025

Dogliotti, A. I., Ruddick, K. G., Nechad, B., Doxaran, D., & Knaeps, E. (2015). A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment*, *156*, 157–168. https://doi.org/10.1016/j.rse.2014.09.020

Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., & Raven, J. A. (2010). Phytoplankton in a changing world: Cell size and elemental stoichiometry. *Journal of Plankton Research*, *32*(1), 119–137. https://doi.org/10.1093/plankt/fbp098

Finkel, Z. V., Irwin, A. J., & Schofield, O. (2004). Resource limitation alters the 3/4 size scaling of metabolic rates in phytoplankton. *Marine Ecology Progress*, *273*(1), 269–279. https://doi.org/10.3354/meps273269

Fu, M., Wang, Z., Li, Y., Li, R., Sun, P., Wei, X., et al. (2009). Phytoplankton biomass size structure and its regulation in the Southern Yellow Sea (China): Seasonal variability. *Continental Shelf Research*, *29*(18), 2178–2194. https://doi.org/10.1016/j.csr.2009.08.010

Gao, Y., Jiang, Z., Liu, J., Chen, Q., Zeng, J., & Huang, W. (2013). Seasonal variations of net-phytoplankton community structure in the southern Yellow Sea. *Journal of Ocean University of China*, *12*(4), 557–567. https://doi.org/10.1007/s11802-013-2258-x

Gao, Y., Yu, Q., Qi, Y., Zhou, J., Lu, D., Li, Y., et al. (2003). Species composition and ecological distribution of planktonic diatoms in the Changjiang River estuary during Spring. *The Chinese Journal of Applied Ecology*, *14*(7), 1044–1048.

Ge, J., Shen, F., Guo, W., Chen, C., & Ding, P. (2015). Estimation of critical shear stress for erosion in the Changjiang Estuary: A synergy research of observation, GOCI sensing and modeling. *Journal of Geophysical Research: Oceans*, *120*, 8439–8465. https://doi.org/10.1002/2015JC010992

Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, *43*(4), 679–694. https://doi.org/10.4319/lo.1998.43.4.0679

Geider, R. J., Platt, T., & Raven, J. A. (1986). Size dependence of growth and photosynthesis in diatoms: A synthesis. *Marine Ecology Progress*, *30*(1), 93–104

Gong, G. C., Wen, Y. H., Wang, B. W., & Liu, G. J. (2003). Seasonal variation of chlorophyll a concentration, primary production and environmental conditions in the subtropical East China Sea. *Deep Sea Res., Part II: Topical Studies in Oceanography*, *50*(6–7), 1219–1236. https://doi.org/10.1016/S0967-0645(03)00019-5

Guo, S., Feng, Y., Wang, L., Dai, M., Liu, Z., Bai, Y., et al. (2014). Seasonal variation in the phytoplankton community of a continental-shelf sea: The East China Sea. *Marine Ecology Progress Series*, *516*, 103–126. https://doi.org/10.3354/meps10952

He, X., Bai, Y., Pan, D., Huang, N., Dong, X., Chen, J., et al. (2013). Using geostationary satellite ocean color data to map the diurnal dynamics of suspended particulate matter in coastal waters. *Remote Sensing of Environment*, *133*, 225–239. https://doi.org/10.1016/j.rse.2013.01.023

Hilligsøe, K., Richardson, K., Bendtsen, J., Sorensen, L., Nielsen, T., & Lyngsgaard, M. (2011). Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO2 flux. *Deep Sea Res., Part I: Oceanographic Research Papers*, *58*(8), 826–838. https://doi.org/10.1016/j.dsr.2011.06.004

Hirata, T., Hardman-Mountford, N. J., Barlow, R., Lamont, T., Brewin, R., Smyth, T., et al. (2009). An inherent optical property approach to the estimation of size-specific photosynthetic rates in eastern boundary upwelling zones from satellite ocean colour: An initial assessment. *Progress in Oceanography*, *83*(1–4), 393–397. https://doi.org/10.1016/j.pocean.2009.07.019

Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., et al. (2011). Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, *8*(2), 311–327. https://doi.org/10.5194/bg-8-311-2011

Hu, Z., Pan, D., He, X., & Bai, Y. (2016a). Diurnal variability of turbidity fronts observed by geostationary satellite ocean color remote sensing. *Remote Sensing*, *8*(2), 147. https://doi.org/10.3390/rs8020147

Hu, Z., Wang, D., Pan, D., He, X., Miyazawa, Y., Bai, Y., et al. (2016b). Mapping surface tidal currents and Changjiang plume in the East China Sea from geostationary ocean color imager. *Journal of Geophysical Research: Oceans*, *121*, 1563–1572. https://doi.org/10.1002/2015JC011469

Huang, B., Liu, Y., Chen, J., Wang, D., Hong, H., Lu, R., et al. (2006). Temporal and spatial distribution of size-fractionized phytoplankton biomass in East China Sea and Huanghai Sea. *Acta Oceanologica Sinica*, *28*(2), 156–164.

IOCCG (2014). Phytoplankton functional types from space. In S. Sathyendranath (Ed.), *Reports of the international ocean-colour coordinating group* (15), Dartmouth, Canada: IOCCG.

Jephson, T., & Carlsson, P. (2009). Species and stratification-dependent diel vertical migration behaviour of three dinoflagellate species in a laboratory study. *Journal of Plankton Research*, *31*(11), 1353–1362. https://doi.org/10.1093/plankt/fbp078

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *Journal of Geophysical Research*, *114*, C09015. https://doi.org/10.1029/2009JC005303

Lee, Z., Jiang, M., Davis, C., Pahlevan, N., Ahn, Y., & Ma, R. (2012). Impact of multiple satellite ocean color samplings in a day on assessing phytoplankton dynamics. *Ocean Science Journal*, *47*(3), 323–329. https://doi.org/10.1007/s12601-012-0031-5

Legendre, L., & Rassoulzadegan, F. (1995). Plankton and nutrient dynamics in marine waters. *Ophelia*, *41*, 153–172. https://doi.org/10.1080/00785236.1995.10422042

Li, G., Han, X., Yue, S., Wen, G., Rongmin, Y., & Kusky, T. M. (2006). Monthly variations of water masses in the East China Seas. *Continental Shelf Research*, *26*(16), 1954–1970. https://doi.org/10.1016/j.csr.2006.06.008

Li, G., Qiao, L., Dong, P., Ma, Y., Xu, J., Liu, S., et al. (2016). Hydrodynamic condition and suspended sediment diffusion in the Yellow Sea and East China Sea. *Journal of Geophysical Research: Oceans*, *121*, 6204–6222. https://doi.org/10.1002/2015JC011442

Lie, H. J., Cho, C. H., Lee, J. H., Lee, S., & Tang, Y. (2001). Does the Yellow Sea Warm Current really exist as a persistent mean flow?. *Journal of Geophysical Research*, *106*(C10), 22199–22210. https://doi.org/10.1029/2000JC000629

Lindemann, C., & St. John, M. A. (2014). A seasonal diary of phytoplankton in the North Atlantic. *Frontiers in Marine Science*, *1*(37), 1–6. https://doi.org/10.3389/fmars.2014.00037

Lin, J., Cao, W., Wang, G., & Hu, S. (2014). Satellite-observed variability of phytoplankton size classes associated with a cold eddy in the South China Sea. *Marine Pollution Bulletin*, *83*(1), 190–197. https://doi.org/10.1016/j.marpolbul.2014.03.052

Liu, R., Zhang, J., Yao, H., Cui, T., Wang, N., Zhang, Y., et al. (2017). Hourly changes in sea surface salinity in coastal waters recorded by Geostationary Ocean Color Imager. *Estuarine, Coastal and Shelf Science*, *196*, 227–236. https://doi.org/10.1016/j.ecss.2017.07.004

Liu, X., Huang, B., Huang, Q., Wang, L., Ni, X., Tang, Q., et al. (2015). Seasonal phytoplankton response to physical processes in the southern Yellow Sea. *Journal of Sea Research*, *95*, 45–55. https://doi.org/10.1016/j.seares.2014.10.017

Loisel, H., Nicolas, J., Sciandra, A., Stramski, D., & Poteau, A. (2006). Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean. *Journal of Geophysical Research*, *111*, C09024. https://doi.org/10.1029/2005JC003367

Lorenzen, C. J. (1963). Diurnal variation in photosynthetic activity of natural phytoplankton populations. *Limnology and Oceanography*, *8*(1), 56–62. https://doi.org/10.4319/lo.1963.8.1.0056

Lou, X., & Hu, C. (2014). Diurnal changes of a harmful algal bloom in the East China Sea: Observations from GOCI. *Remote Sensing of Environment*, *140*, 562–572. https://doi.org/10.1016/j.rse.2013.09.031

Madariaga, I., & Orive, E. (1989). Spatiotemporal variations of size- fractionated primary production in the Gernika estuary. *Journal of Experimental Marine Biology and Ecology*, *127*(3), 273–288. https://doi.org/10.1016/0022-0981(89)90079-8

Marañón, E. (2015). Cell size as a key determinant of phytoplankton metabolism and community structure. *Annual Review of Marine Science*, *7*, 241–264. https://doi.org/10.1146/annurev-marine-010814-015955

Marañón, E., Cermeño, P., López-Sandoval, D. C., Rodríguez-Ramos, T., Sobrino, C., Huete-Ortega, M., et al. (2013). Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. *Ecological Letters*, *16*(3), 371–379. https://doi.org/10.1111/ele.12052

Marañón, E., Holligan, P. M., Barciela, R., Gonzalez, N., Mourino, B., Pazo, M. J., et al. (2001). Patterns of phytoplankton size structure and productivity in contrasting open-ocean environments. *Marine Ecology Progress Series*, *216*, 43–56. https://doi.org/10.3354/meps216043

Maulood, B. K., Hinton, G. C. F., & Boney, A. D. (1978). Diurnal variation of phytoplankton in Loch Lomond. *Hydrobiologia*, *58*(2), 99–117. https://doi.org/10.1007/BF00007992

Mitbavkar, S., & Saino, T. (2015). Diurnal variability of Synechococcus abundance in Sagami Bay, Japan. *Hydrobiologia*, *747*(1), 133–145. https://doi.org/10.1007/s10750-014-2125-9

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M., et al. (2008). Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, *112*(8), 3366–3375. https://doi.org/10.1016/j.rse.2008.01.021

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, *103*(C11), 24924–24937. https://doi.org/10.1029/98JC02160

Pan, Y., Shen, F., & Wei, X. (2018). Fusion of Landsat-8/OLI and GOCI Data for hourly mapping of suspended particulate matter at high spatial resolution: A case study in the Yangtze (Changjiang) estuary. *Remote Sensing*, *10*(2), 158. https://doi.org/10.3390/rs10020158

Quan, Q., Mao, X., Yang, X., Hu, Y., Zhang, H., & Jiang, W. (2013). Seasonal variations of several main water masses in the southern Yellow Sea and East China Sea in 2011. *Journal of Ocean University of China*, *12*(4), 524–536. https://doi.org/10.1007/s11802-013-2198-5

Ras, J., Claustre, H., & Uitz, J. (2007). Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: Comparison between in situ and predicted data. *Biogeosciences*, *5*(2), 353–369. https://doi.org/10.5194/bg-5-353-2008

Ryu, J. H., Choi, J. K., Eom, J., & Ahn, J. H. (2011). Temporal variation in Korean coastal waters using geostationary ocean color imager. *Journal of Coastal Research*, *SI64*, 1731–1735

Sathyendranath, S., Cota, G., Stuart, V., Maass, H., & Platt, T. (2001). Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, *22*(2), 249–273. https://doi.org/10.1080/014311601449925

Sieburth, J. M., Smetacek, V., & Lenz, J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnology and Oceanography*, *23*(6), 1256–1263. https://doi.org/10.4319/lo.1978.23.6.1256

Siswanto, E., Tang, J., Yamaguchi, H., Ahn, Y., Ishizaka, J., Yoo, S., et al. (2011). Empirical ocean-color algorithms to retrieve chlorophyll-a, total suspended matter, and colored dissolved organic matter absorption coefficient in the Yellow and East China Seas. *Journal of Oceanography*, *67*(5), 627–650. https://doi.org/10.1007/s10872-011-0062-z

Song, S., Li, Z., Li, C., & Yu, Z. (2017). The response of spring phytoplankton assemblage to diluted water and upwelling in the eutrophic Changjiang (Yangtze River) Estuary. *Acta Oceanologica Sinica*, *36*(12), 101–110. https://doi.org/10.1007/s13131-017-1094-z

Sun, D., Huan, Y., Qiu, Z., Hu, C., Wang, S., & He, Y. (2017). Remote-sensing estimation of phytoplankton size classes from GOCI satellite measurements in Bohai Sea and Yellow Sea. *Journal of Geophysical Research: Oceans*, *122*, 8309–8325. https://doi.org/10.1002/2017JC013099

Sun, J., Liu, D., Yang, S., Guo, J., & Qian, S. (2002). The preliminary study on phytoplankton community structure in the central Bohai sea and the Bohai strait and its adjacent area. *Oceanologia et Limnologia Sinica*, *33*(5), 461–471.

Sun, X., Dong, S., & Tang, Z. (2008). Influences of nutrients and illuminance on phytoplankton community structure. *South China Fisheries Science*, *4*(1), 1–9.

Sun, X., Ren, L., Zheng, S., Wen, F., Zhao, Y., & Sun, S. (2012). Phytoplankton size structure in the yellow sea and East China Sea in the spring and summer of 2011. *Oceanologia et Limnologia Sinica*, *43*(3), 419–428.

Tassan, S. (1994). Local algorithms using SeaWiFS data for the retrieval of phytoplankton, pigments, suspended sediment, and yellow substance in coastal waters. *Applied Optics*, *33*(12), 2369–2378. https://doi.org/10.1364/AO.33.002369

Teira, E., Serret, P., & Fernandez, E. (2001). Phytoplankton size-structure, particulate and dissolved organic carbon production and oxygen fluxes through microbial communities in the NW Iberian coastal transition zone. *Marine Ecology Progress Series*, *219*, 65–83. https://doi.org/10.3354/meps219065

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, *111*, C08005. https://doi.org/10.1029/2005JC003207

Uitz, J., Huot, Y., Bruyant, F., Babin, M., & Claustre, H. (2008). Relating phytoplankton photophysiological properties to community structure on large scale. *Limnology and Oceanography*, *53*(2), 614–630. https://doi.org/10.4319/lo.2008.53.2.0614

Vaulot, D., & Marie, D. (1999). Diel variability of photosynthetic picoplankton in the equatorial Pacific. *Journal of Geophysical Research*, *104*(C2), 3297–3310. https://doi.org/10.1029/98JC01333

Vidussi, F., Claustre, H., Manca, B., Luchetta, A., & Marty, J. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research*, *106*(C9), 19939–19956. https://doi.org/10.1029/1999JC000308

Waite, A. M., Fisher, A., Thompson, P. A., & Harrison, P. J. (1997). Sinking rate versus volume relationships illuminate sinking control mechanisms in marine diatoms. *Marine Ecology Progress Series*, *157*(8), 97–108

Wang, B., Wang, X., & Zhan, R. (2003). Nutrient conditions in the Yellow Sea and the East China Sea. *Estuarine, Coastal and Shelf Science*, *58*(1), 127–136. https://doi.org/10.1016/S0272-7714(03)00067-2

Wang, L., Huang, B., Chiang, K., Liu, X., Chen, B., Xie, Y., et al. (2016). Physical-biological coupling in the Western South China Sea: The response of phytoplankton community to a mesoscale cyclonic eddy. *Plos One*, *11*(4), e153735. https://doi.org/10.1371/journal.pone.0153735

Ward, B. A., Dutkiewicz, S., Jahn, O., & Follows, M. J. (2012). A size-structured food-web model for the global ocean. *Limnology and Oceanography*, *57*(6), 1877–1891. https://doi.org/10.4319/lo.2012.57.6.1877

Xia, P., Lu, D., Zhu, D., & Du, W. (2007). Trend and characteristics of harmful algal blooms in Zhejiang coastal waters. *Journal of Marine Sciences*, *25*(02), 47–56.

Yentsch, C. S., & Phinney, D. A. (1989). A bridge between ocean optics and microbial ecology. *Limnology and Oceanography*, *34*(8), 1694–1705. https://doi.org/10.4319/lo.1989.34.8.1694

Yu, F., Zhang, Z., Lan, J., Diao, X., Guo, J., & Ge, R. (2005). Analysis of Water Temperature Distribution Characteristics in the Southern Yellow Sea in Spring. *Advances in Marine Science*, *23*(3), 281–288.

Zhang, J., Liu, D., Cao, P., Wang, Y., Keesing, J. K., Li, J., & Chen, L. (2016). A highly sensitive method for analyzing marker phytoplankton pigments: Ultra-high-performance liquid chromatography-tandem triple quadrupole mass spectrometry. *Limnology and Oceanography: Methods*, *14*(10), 623–636. https://doi.org/10.1002/lom3.10117

Zhang, J., Liu, S. M., Ren, J. L., Wu, Y., & Zhang, G. L. (2007). Nutrient gradients from the eutrophic Changjiang (Yangtze River) Estuary to the oligotrophic Kuroshio waters and re-evaluation of budgets for the East China Sea Shelf. *Progress in Oceanography*, *74*(4), 449–478. https://doi.org/10.1016/j.pocean.2007.04.019

Zhang, S. W., Wang, Q. Y., Lü, Y., Cui, H., & Yuan, Y. L. (2008). Observation of the seasonal evolution of the Yellow Sea Cold Water Mass in 1996–1998. *Continental Shelf Research*, *28*(3), 442–457. https://doi.org/10.1016/j.csr.2007.10.002

Zhou, M., Yan, T., & Zou, J. (2003). Preliminary analysis of the characteristics of red tide areas in Changjiang River estuary and its adjacent sea. *Chinese Journal of Applied Ecology*, *14*(7), 1031–1038.

Zhu, Z., Ng, W., Liu, S., Zhang, J., Chen, J., & Wu, Y. (2009). Estuarine phytoplankton dynamics and shift of limiting factors: A study in the Changjiang (Yangtze River) Estuary and adjacent area. *Estuarine, Coastal and Shelf Science*, *84*(3), 393–401. https://doi.org/10.1016/j.ecss.2009.07.005

Estuarine, Coastal and Shelf Science 181 (2016) 302–311

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

journal homepage: www.elsevier.com/locate/ecss

Light absorption properties of CDOM in the Changjiang (Yangtze) estuarine and coastal waters: An alternative approach for DOC estimation

Xiaolong Yu [a, b], Fang Shen [a, *], Yangyang Liu [a]

[a] State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai, 200062, China
[b] Department of Water Resources, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500AE, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 18 February 2016
Received in revised form
31 August 2016
Accepted 5 September 2016
Available online 7 September 2016

Keywords:
Light absorption
CDOM
DOC
Salinity
The Changjiang (Yangtze) estuary

ABSTRACT

Field measurements of CDOM absorption properties and DOC concentrations were collected in the Changjiang estuarine and coastal waters from 2011 to 2013. CDOM absorption coefficient at 355 nm (a_g (355)) was found to be inversely correlated with salinity, with Pearson's coefficients r of −0.901 and −0.826 for summer and winter observations, respectively. Analysis results of the relationships between salinity and CDOM optical properties (*i.e.*, absorption coefficient and spectral slope) suggested that terrigenous inputs dominated CDOM sources in the Changjiang estuary, but the proportion of terrigenous CDOM declined with increasing salinity. The level of CDOM in the Changjiang estuary was lower compared to some of the major estuaries in the world, which could be attributed to several controlling factors such as vegetation cover in the drainage basin, the origin of recharged streams and high sediment load in the Changjiang estuary. We further evaluated the relationships between CDOM and DOC and their mixing behavior among world's major estuaries. An empirical model was finally developed to estimate DOC concentration from a_g (355) and spectral slope $S_{275-295}$ using a non-linear regression. This empirical relationship was calibrated using the Cal dataset, and was validated with the Val dataset, resulting in an acceptable error with the R^2 of 0.746, the RMSE of 20.99 μmol/L and the rMAD of 14.46%.

1. Introduction

Colored dissolved organic matter (CDOM) is an important component of coastal waters, controlling the functioning of ecological processes and biogeochemical cycles of marine ecosystems. Light absorption by CDOM mainly depends on the origin of CDOM. Terrestrial sources contain more humic acid and large molecules of dissolved organic matter (DOM), while aquatic sources contain more fulvic acid and smaller DOM molecules (Helms et al., 2008). Riverine discharge is considered as the main source of CDOM in most coastal waters, resulting in a robust correlation between salinity and the CDOM absorption coefficient (Fichot and Benner, 2011; Granskog, 2012; Xie et al., 2012). The aquatic CDOM is mainly produced locally by phytoplankton degradation and bacterial decomposition. The absorption spectrum of CDOM follows an exponential function, whereby the CDOM absorption decreases with increasing wavelength (Bricaud et al., 1981). The rate of this decrease, hereafter called the spectral slope (S, in nm $^{-1}$), can easily be computed from the absorption spectrum of CDOM. This spectral slope is related to the ratio of fulvic to humic acids and the molecular weight (MW) of fulvic acids and is, therefore, commonly used to identify the origins of CDOM (Carder et al., 1989; Keith et al., 2002). The slope S can be computed from the spectral range either using narrow waveband (*e.g.*, 275–295 nm) or broader waveband (*e.g.*, 250–700 nm). However, S computed from narrow waveband is known to be more sensitive to CDOM sources than the S determined from a broader waveband (Asmala et al., 2012; Fichot and Benner, 2011, 2012).

Dissolved organic carbon (DOC), on the other hand, represents 97% of the organic carbon in the ocean (Hansell and Carlson, 1998), and is therefore an essential part in the global carbon cycle.

---

* Corresponding author.
E-mail address: fshen@sklec.ecnu.edu.cn (F. Shen).

http://dx.doi.org/10.1016/j.ecss.2016.09.004

Phytoplankton converts inorganic carbon to organic carbon through photosynthesis (Longhurst and Harrison, 1989), contributing a substantial source of DOC in the oceans. Other sources of DOC in the oceans mostly come from riverine discharge and zooplankton activity (Kuliński and Pempkowiak, 2008). Terrigenous DOC from river discharges contributes a large portion of DOC in the coastal oceans and is further transported to the open ocean. To understand the migration of terrigenous DOC and estimate the DOC budget from rivers to the ocean, measurements of DOC concentrations are necessary in the coastal oceans. However, traditional DOC measurements are always limited due to discrete sampling. Therefore, a prompt and continuous estimation from earth observation data at large spatial scales would be strongly recommended (Liu et al., 2013; Mannino et al., 2008).

CDOM is the optically active component of DOM and therefore can be potentially linked to DOC concentration. Although CDOM only represents a portion of the entire DOC pool, optical properties of CDOM have been found to be empirically related to DOC concentration in some estuarine and coastal waters (Asmala et al., 2012; Ferrari, 2000; Fichot and Benner, 2011, 2012; Matsuoka et al., 2012a; Rochelle-Newall et al., 2014; Spencer et al., 2007). For example, DOC concentration can be linked to the CDOM absorption coefficient at a reference wavelength (*e.g.*, 250 nm or 350 nm) by single linear regression (Baker and Spencer, 2004; Matsuoka et al., 2012a; Spencer et al., 2009a). However, linear models rely heavily on the conservative mixing of DOC and CDOM and are not always practical in the coastal waters, such as in the Mississippi River estuary (Chen and Gardner, 2004), the Pearl River estuary (Chen et al., 2004) and West Florida Shelf (Del Castillo et al., 2000). The other approach is to estimate DOC from the spectral slope $S_{275\text{-}295}$ and the DOC-normalized CDOM absorption coefficient by non-linear regression (Fichot and Benner, 2012). The non-linear model is supposed to be more suitable for coastal waters, especially for river-influenced coastal waters with lignin as the important chromophore of the CDOM pool and the main terrigenous component of DOC (Fichot and Benner, 2012).

The hydrodynamic environment is unique in the Changjiang estuarine and coastal waters due to the joint influences of runoffs, tides and coastal circulations, which makes the mixing behavior between DOC and CDOM more complicated and variable at seasonal and regional scales. For example, conservative mixing behavior was observed in the Changjiang estuary with the DOC concentration linearly related to the CDOM absorption coefficient (Liu et al., 2013, 2014; Zhang et al., 2013). However, the linear correlation was not always practical in this region (Cauwet and Mackenzie, 1993; Liu et al., 2013), and could be easily collapsed in the zones with significant phytoplankton production (Liu et al., 2013, 2014). Therefore, field investigations are still necessary to better understand the variation of CDOM and DOC in the Changjiang estuarine and coastal waters. Moreover, an alternative model estimating the DOC concentration could be also useful especially when no correlation between CDOM and DOC concentration can be found.

In this study, we first investigate the seasonal and spatial variations of the CDOM optical properties, which were less well documented in the Changjiang estuarine and coastal waters. We further discuss the potential sources of CDOM and DOC, and employ a non-linear model to estimate the DOC concentration from CDOM optical properties. The developed relationship was found to be robust, and therefore could be an alternative approach to estimate the DOC concentration at large spatial scales from remotely retrieved CDOM optical properties in the Changjiang estuarine and coastal waters.

2. Materials and methods

2.1. Shipborne sampling and measurements

Water samples were collected during five cruises in the Changjiang estuarine and coastal waters from 2011 to 2013, and the sampling stations are presented in Fig. 1. The field campaigns were carried out in July 2011 (32 samples), July 2012 (29 samples), August 2013 (53 samples), February 2012 (37 samples) and March 2012 (24 samples). A total number of 175 water samples were collected from the surface layer and analyzed for their content of CDOM. Salinity (in PSU, practical salinity units) was recorded synchronously to CDOM sampling during the cruises using the CTD device (SeaBird Electronics INC). Note, however, that the salinities of 4 samples (out of 175) were not recorded during the CDOM sampling. DOC samples were collected simultaneously with surface CDOM samples in July of 2011 (31 samples) and August 2013 (53 samples). To investigate the vertical variation of CDOM, depth profile of CDOM was recorded during the cruise in July 2011, whereby CDOM samples were collected at two depths of 5 m (31 samples) and 10 m (29 samples).

Water samples were gravity-filtered on shipboard using a 0.22 μm polycarbonate membrane (Millipore, 47 mm diameter) under low vacuum immediately after the sampling. The membranes were soaked in 10% HCL for 15 min and then rinsed by Milli-Q water three times before the filtration. The filtered CDOM samples were collected in borosilicate glass vials, and then stored in the −40 °C refrigerator. All vials were pre-soaked in 10% HCL for 24 h, rinsed by Milli-Q water for three times, and pre-combusted at 450 °C for 5 h. DOC samples were filtered using a 0.45 μm nylon membrane (RephiLe RF-Jet Syringe Filter, 25 mm diameter) and collected in ampoule bottles, which were pre-combusted at 500 °C for 5 h. After the filtration, the ampoule bottles were sealed by fusing the bottleneck, and then stored in the - 40 °C refrigerator.

2.2. Laboratory measurements

Immediately prior to measurement, CDOM samples were unfrozen and warmed to room temperature under fully dark

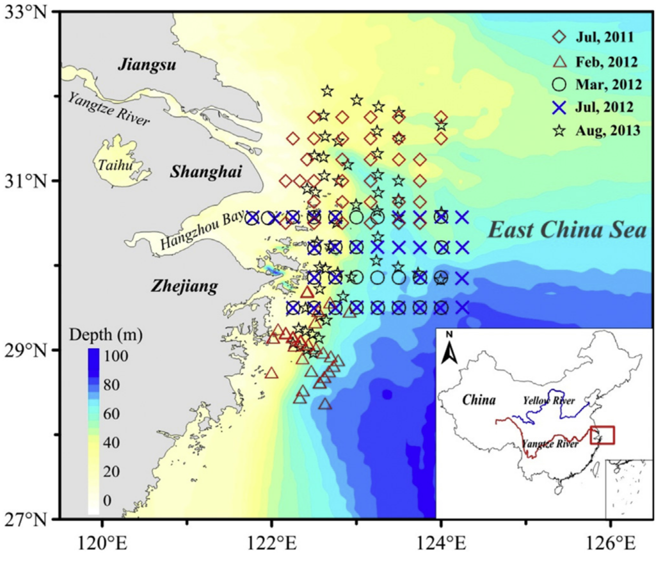


Fig. 1. Location of sampling stations in the Changjiang estuarine and coastal waters. Samples were collected from five cruises in summer (July 2011, July 2012 and August 2013) and winter (February and March of 2012).

conditions, which was used to minimize the freezing influence on the absorption spectra. The CDOM absorbance spectra were measured from 250 to 900 nm using a 10 cm quartz cuvette referenced to the Milli-Q water by the PerkinElmer Lambda 1050 UV/VIS spectrophotometer. The CDOM absorption coefficient $a'_g\ (\lambda)$ (in m^{-1}) was then computed from the absorptance $D\ (\lambda)$ as (Pegau et al., 1997):

$$a'_g\ (\lambda)\ =\ 2.303\ \times\ D\ (\lambda)/l \tag{1}$$

where, l is the cuvette path length. The effect of scattering was corrected by normalizing the spectra to zero at 700 nm (Bricaud et al., 1981),

$$a_g(\lambda)\ =\ a'_g(\lambda)\ -\ a'_g(\lambda_{700})\ \times\ \lambda/\lambda_{700} \tag{2}$$

where, λ_{700} is the 700 nm wavelength and $a_g\ (\lambda)$ is the CDOM absorption coefficient after the scattering correction. Fig. 2 shows the absorption spectra of all CDOM samples.

The spectral slope is derived from CDOM absorption spectra by fitting the absorption data to the exponential equation (*i.e.*, Eq. (3)) using the optimization method (Bricaud et al., 1981),

$$a_g(\lambda) = a_g(\lambda_0)e^{-S(\lambda-\lambda_0)} \tag{3}$$

where, λ is wavelength (nm) and λ_0 is the reference wavelength (nm). As shown in Fig. 2, S is apparently dependent on the wavelength interval over which it is calculated. We chose 440 nm as the reference wavelength to calculate the spectral slope for the broad range from 250 to 700 nm, which is hereafter denoted as S_{440}. $S_{275\text{-}295}$ and $S_{350\text{-}400}$ were also calculated for the narrow wavelength intervals from 275 to 295 nm and 350–400 nm, respectively. The spectral slope ratio S_R was then defined as the ratio of $S_{275\text{-}295}$ and $S_{350\text{-}400}$.

DOC concentration (in μmol/L) was determined in duplicate with a Shimadzu TOC-VCPH Total Organic Carbon analyzer using potassium hydrogen phthalate as the standard for the DOC calibration curve. The typical precision of this instrument is less than 2% in terms of the coefficient of variation. DOC samples were acidified by 100 μL 2 μmol/L phosphoric acid prior to the measurement to remove the inorganic carbon. To ensure the accuracy and consistency of measured DOC concentration, we analyzed the Florida Strait seawater reference, provided by Dr. Hansell's laboratory at the University of Miami, at the beginning and every fifth sample. The average measured DOC concentration of the Florida Strait seawater was 45.6 μmol/L, which fitted well to the expected range (44–46 μmol/L). DOC concentrations were estimated as the average value of the duplicate measurements. The DOC-normalized CDOM absorption coefficient ($a^*_g\ (\lambda_0)$, in L μmol^{-1} m^{-1}) was defined as the ratio of $a_g\ (\lambda_0)$ and the DOC concentration.

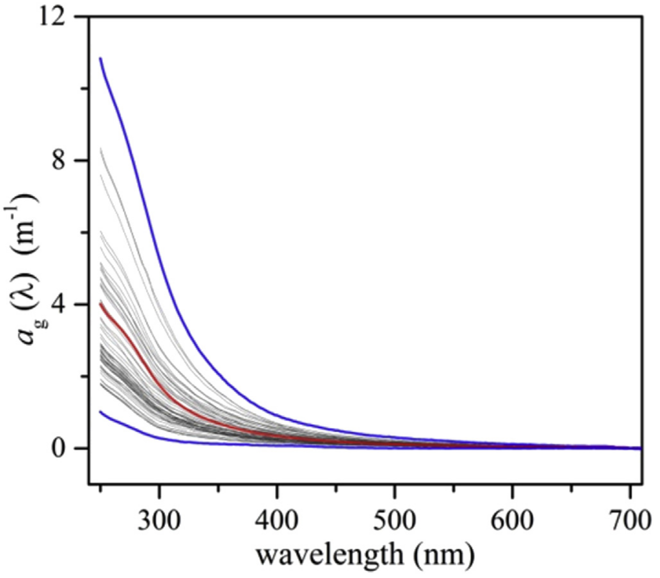


Fig. 2. CDOM absorption spectra of all measurements. The two blue spectra represent samples with maximum and minimum absorption coefficients, while the red spectrum represents the mean absorption coefficient of all samples.

2.3. Data analysis

To evaluate the relationship between salinity and CDOM absorption, we calculate the Pearson's coefficient r and the probability p of the linear regression. For the goodness-of-fit between derived and known DOC concentration, statistical parameters such as the determination coefficient R^2, the mean of absolute relative difference (rMAD) and the root mean square error (RMSE) are determined. The rMAD and the RMSE are defined as follows:

$$\text{rMAD} = \sum|1 - \text{derived/known}|/\text{N} \times 100\%, \tag{4}$$

$$\text{RMSE} = \sqrt{\sum(\text{derived} - \text{known})^2 \Big/ \text{N}}, \tag{5}$$

where, N is the number of observations.

3. Results

3.1. Spatial-temporal variation of CDOM

The distribution of the CDOM absorption coefficient (a_g (355)) at the surface layer in the Changjiang estuarine and coastal waters is presented in Fig. 3, where noticeable spatial-temporal variations of a_g (355) can be observed.

The a_g (355) values ranged from 0.42 to 1.91 m^{-1} in July 2011, from 0.11 to 0.84 m^{-1} in February 2012, from 0.13 to 1.21 m^{-1} in March 2012, from 0.26 to 1.36 m^{-1} in July 2012 and from 0.23 to 1.45 m^{-1} in August 2013, respectively. The a_g (355) values exhibited a gradual decrease from nearshore to offshore, forming a persistent northwest-southeast gradient, with low values (less than 0.5 m^{-1}) in regions far from the river mouth. Most samples in March and July 2012 were collected at the same sites (Fig. 3c and d), from which we can observe that CDOM absorption coefficients in summer were higher than in winter. The annual variation of a_g (355) over the study area in the summer from 2011 to 2013 is also observed according to Fig. 3a, d and 3e. The average values of a_g (355) are 0.89 m^{-1}, 0.52 m^{-1} and 0.60 m^{-1} in July 2011, July 2012 and August 2013, respectively.

As shown in Fig. 4, the vertical gradients of a_g (355) in July 2011 are random among the stations with no consistent depth profile. The largest variations of the a_g (355) depth-profile are found at stations adjacent to the Changjiang river mouth (*e.g.*, stations with longitudes below 122.5°E in Fig. 4c, e and 4f), while vertical variations in offshore are rather small, especially for stations with longitudes larger than 123°E.

3.2. CDOM and salinity

Salinity is a perfect indicator of the influence of river runoff in the estuarine regions. The variation of salinity is mainly attributed

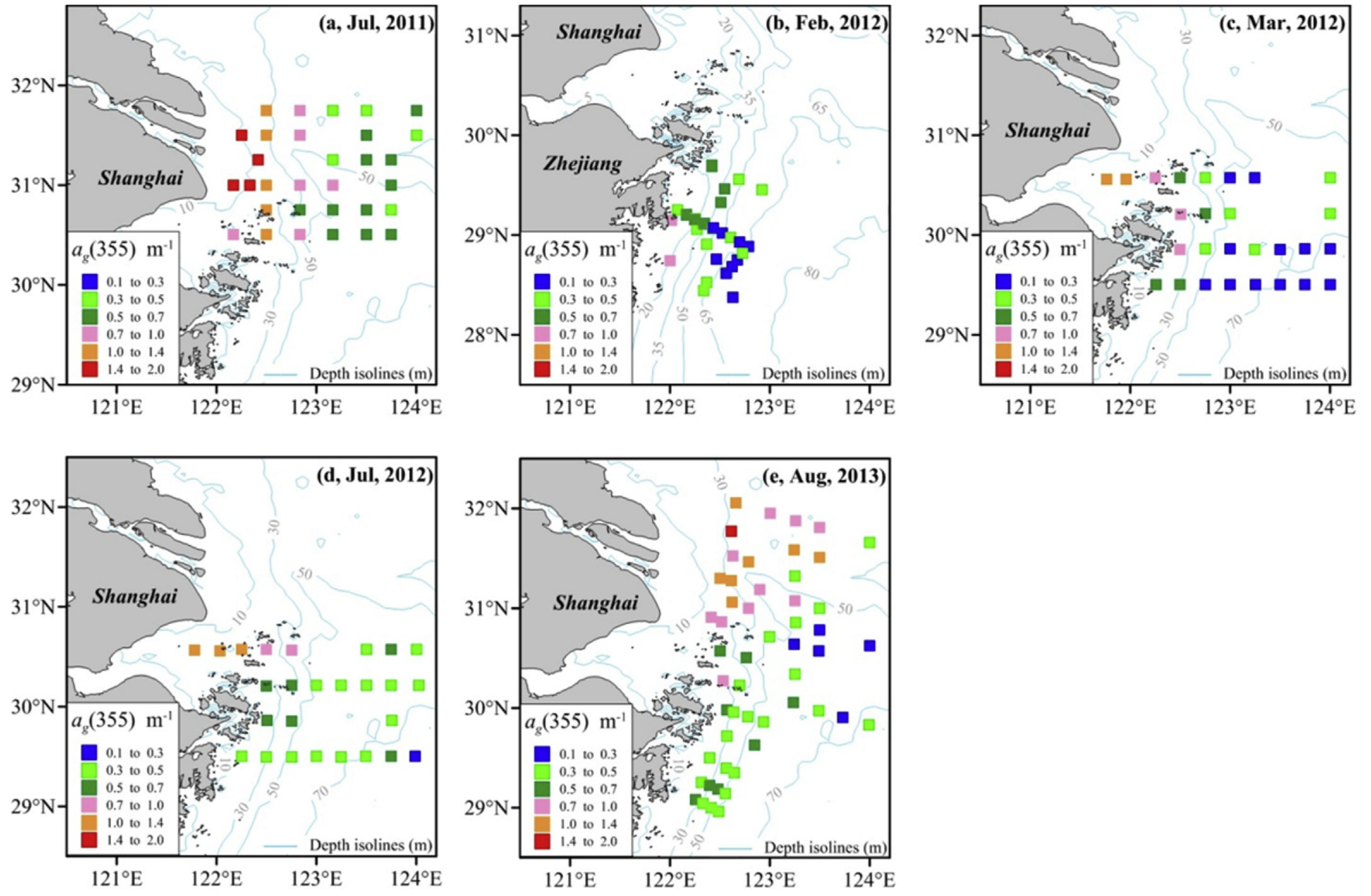


Fig. 3. Spatial distribution of surface a_g (355) over the Changjiang estuarine and coastal region. Data was collected from five cruises in July of 2011 (a), February (b), March (c), July (d) of 2012 and August of 2013 (e).

to the mixing between fresh river discharge and saline ocean water. Fresh water is almost free from salinity while the surface salinity in the open ocean is typically between 32 and 37 PSU (Lagerloef et al., 1995). To better explore the relationship between the CDOM absorption coefficient and salinity, CDOM measurements are divided into subsets based on two criteria. The first classification is based on the sampling seasons, *i.e.* winter and summer. The second is according to the salinity values, from which three classes are formed using two thresholds of 20 PSU and 32 PSU. We consider that samples with salinity lower than 20 PSU are riverine-dominant while marine-dominant samples have salinities greater than 32 PSU. Fig. 5 shows the relationships between salinities and a_g (355), as well as their linear fitting parameters.

From Fig. 5a we observe that a_g (355) is negatively correlated with salinity, with a more robust correlation in summer ($r = -0.901, p = 0$) than in winter ($r = -0.826, p < 0.001$). Moreover, as shown in Fig. 5b, the correlation is more robust for low salinity samples (<20 PSU, $r = -0.851, p = 0.002$) compared to high salinity samples (>32 PSU, $r = -0.572, p < 0.001$).

3.3. Spectral slope of CDOM absorption and salinity

The range of S_{440} for all samples was from 0.009 to 0.019 nm^{-1} with an average value of 0.015 nm^{-1}. Seasonal variations of S_{440} are also observed as the average value of S_{440} in July 2012 was 0.017 ± 0.001 nm^{-1}, while the average value in March 2012 was 0.0145 ± 0.0017 nm^{-1}. On the other hand, $S_{275\text{-}295}$ ranges from 0.015 to 0.032 nm^{-1} for all observations with an average value of 0.021 nm^{-1}. However, opposite to the seasonal variation of S_{440}, the average value of $S_{275\text{-}295}$ in March 2012 (0.023 ± 0.004 nm^{-1}) was higher than in July 2012 (0.018 ± 0.002 nm^{-1}).

No correlation between S_{440} and salinity can be found in this study (figure not shown), while both $S_{275\text{-}295}$ and S_R increased with the increasing salinity as shown in Fig. 6. The patterns between salinity and $S_{275\text{-}295}$ and S_R are in accordance with previous studies (Fichot and Benner, 2012; Helms et al., 2008). Moreover, the correlation with salinity is more robust in riverine-dominant samples for both $S_{275\text{-}295}$ and S_R.

3.4. CDOM properties and DOC

DOC concentrations ranged from 60.54 to 217.06 μmol/L in this study, with an average value of 115.41 μmol/L. DOC was not correlated to both salinity and a_g (355) in this study, presenting a non-conservative mixing behavior between CDOM and DOC. To estimate DOC from CDOM properties, we adopted the approach of Fichot and Benner (2012) and developed a non-linear relationship between $S_{275\text{-}295}$ and DOC-normalized CDOM absorption coefficient. All the DOC samples were divided into two optimal calibration and validation sets according to the GeoCal/Val model of Salama et al. (2012). The Cal and Val sets were selected such that they have the same moments (mean and standard deviation). An exponential function was used to fit a_g^* (355) and $S_{275\text{-}295}$ for the calibration dataset (Fig. 7a), resulting in a R^2 of 0.78. The fitting equation is expressed as,

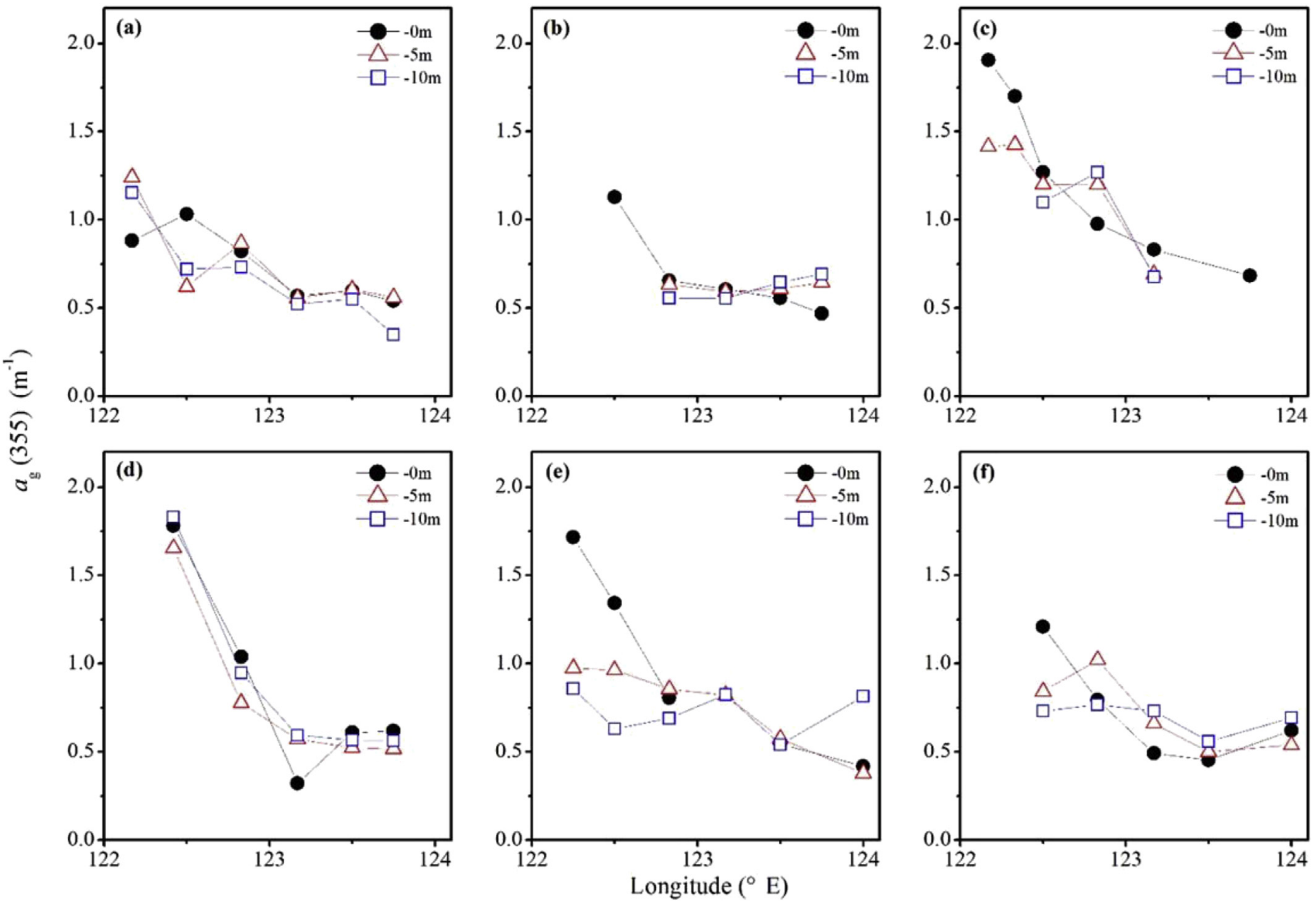


Fig. 4. Vertical distribution of a_g (355) along geographic longitude increasing in July 2011. Fig. 4a–f are corresponding to the six axial transects in Fig. 3a from the bottom up.

$$a_g^*(355) = e^{(-2.45 - 130.79 S_{275-295})} \tag{6}$$

The fitting equation from Fichot and Benner (2012) is also provided in Fig. 7a for reference. The developed model was applied to the Val dataset to estimate the error of retrieved DOC as shown in Fig. 7b. The accuracy of our model in Eq. (6) is expressed by the R^2 of 0.746, the RMSE of 20.99 μmol/L and the rMAD of 14.46%.

4. Discussion

4.1. Comparison of CDOM and DOC in several estuarine and coastal waters

Table 1 shows the comparison of several investigated parameters among the world's estuarine and coastal waters, including the

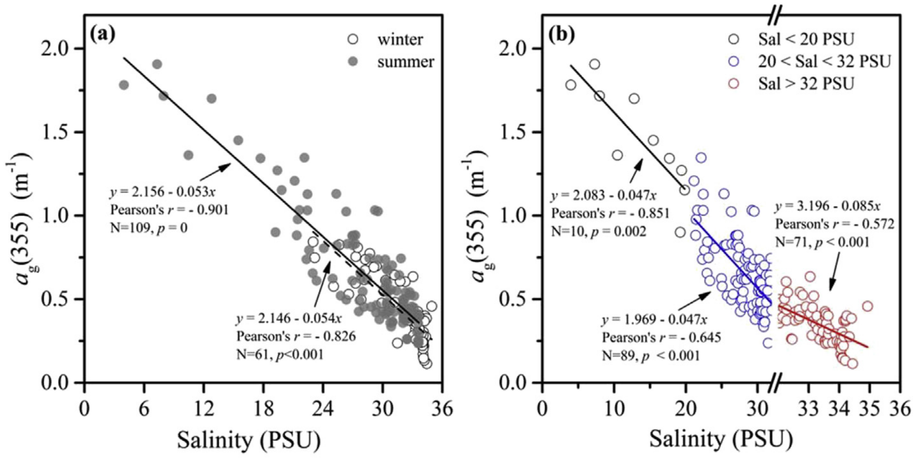


Fig. 5. Relationships between salinity and a_g (355). All CDOM measurements are divided into (a) two subsets for samples collected in winter and summer and (b) three subsets according to the salinity values with two thresholds at 20 and 32 PSU.

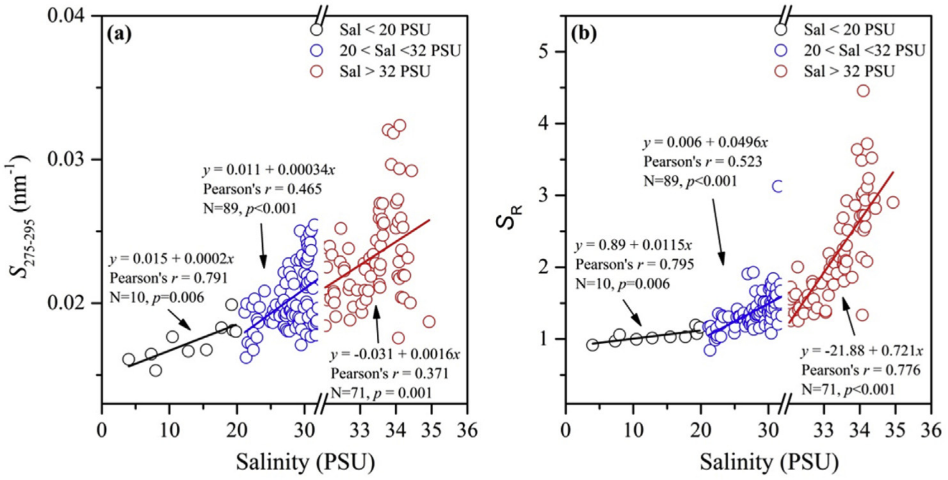


Fig. 6. Scatter plots between salinity and (a) $S_{275\text{-}295}$ and (b) S_R. All observations are divided into three subsets using thresholds of salinity at 20 PSU and 32 PSU.

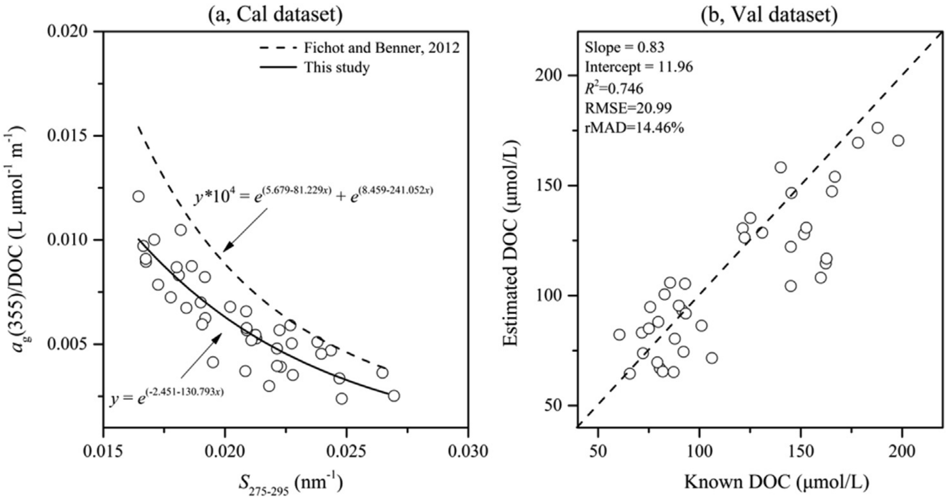


Fig. 7. Calibration and validation for DOC estimation model using the Cal and Val datasets. (a) A non-linear relationship (black line) is used to fit a_g^* (355) and $S_{275\text{-}295}$ using the Cal dataset. A referential relationship from Fichot and Benner (2012) is also provided (dash line). (b) Validation of the developed model using the Val dataset.

Table 1
Comparisons of CDOM optical properties, salinity, DOC concentration and the relationship between CDOM and DOC among different estuarine and coastal waters.

Region	ag (355) (m-1)	S440 (nm-1)	Salinity (PSU)	DOC (μmol/L)	CDOM vs. DOC	Season	References
Amazon River and Estuary	0.14–3.12	0.014–0.033	0–36.4	NA	NA	winter	(Green and Blough, 1994)
Orinoco River plume	0.93–9.02 a_g (300)	0.012–0.022	19.7–34.6	84–276	Correlated	fall	(Del Castillo et al., 1999)
St. Lawrence River Estuary	0.2–3.8 a_g (365)	0.015	0–35.5	NA	NA	summer	(Nieke et al., 1997)
Mackenzie River Estuary	0.6–2.02 a_g (320)	0.018–0.025 S_{320}	24.5–30.1	73.3–205.8	Correlated	fall	(Retamal et al., 2007)
Southern Beaufort Sea	0.018–1.08 a_g (440)	0.015–0.023	0–35.0	49–460	Correlated	summer	(Matsuoka et al., 2012a)
Scheldt Estuary	0.97–4.22 a_g (375)	0.017–0.019	0.7–29.6	NA	NA	winter	(Astoreca et al., 2009)
	2.34–4.29 a_g (375)	0.018–0.019	2.7–19.9	NA	NA	summer	
Northern Gulf of Mexico	3.96–17.52 a_g (350)	NA	0–37	232–611	Correlated	all	(Fichot and Benner, 2012)
Chesapeake Bay	0.4–4.5	0.015–0.024	0–33	60–290	Correlated	all	(Rochelle-Newall and Fisher, 2002)
Pearl River estuary	0.24–1.93	0.014–0.018	0–32.5	NA	NA	winter	(Hong et al., 2005)
Pearl River estuary	0.34–1.40	NA	0–35.0	86–250	Uncorrelated	summer	(Chen et al., 2004)
Changjiang estuary[a]	0.1–3.2	0.017–0.020	0–32.0	NA	NA	summer	(Guo et al., 2007)
Changjiang estuary[a]	0.10~2.82	0.017~0.020	0.12~29.4	41.7–126.2	Correlated	spring	(Sun et al., 2014)
Changjiang estuary[a]	0.43~3.74	0.018~0.022	0.2~26.5	139.4–220	Correlated	summer	(Zhang et al., 2013)
	0.05~ 1.15	0.012~0.023	19~34.5	63.7–129.5		winter[b]	
Changjiang estuary[c]	0.11–1.20	0.008–0.018	18.7–34.9	NA	NA	winter	This study
	0.23–1.91	0.012–0.025	4.0–33.6	60.5–217	Uncorrelated	summer	

[a] Included samples collected from the river mouth.
[b] Included data collected from the PN section in March of 2011 in their study.
[c] Only data from the surface layer is counted; NA means no data or information can be found in their study.

ranges of the CDOM absorption coefficient, spectral slope S_{440}, salinity, DOC concentration and the relationship between CDOM and DOC. In general, the level of a_g (355) in the Changjiang estuary was on par with the Pearl River estuary, but lower than most of the world's major estuaries, which is consistent with previous studies in the Changjiang estuary and other Chinese estuaries (Guo et al., 2014; Pan et al., 2012).

The abundance of DOM in the estuarine and coastal water is always influenced by vegetation, climate, and the types of rock and soil in the river basins (Ludwig et al., 1996; Pan et al., 2012). As shown by the Terra MODIS NDVI products (http://earthobservatory.nasa.gov/GlobalMaps/view.php?d1=MOD13A2_M_NDVI), the coverage of vegetation in the Changjiang River basin is lower compared to other drainage basins of the world's major rivers, which is probably due to rapid urbanization and massive deforestation in the past decades (Huang et al., 2015, 2016). The low degree of vegetative cover in the drainage basin could be an important factor in determining the relatively low content of soil organic matter, and therefore results in a low level of CDOM in the Changjiang estuary. The same factor was also reported to be responsible for the low level of CDOM in the Pearl River estuary (Chen et al., 2004; Hong et al., 2005).

The recharged streams could also have huge impact on the level of CDOM in the river. Compared to streams originated from marshland, peatland and rainforest, streams originated from glacier and mountain areas usually have a low degree of DOM and their discharges into the main river will dilute the concentration of DOM (Pan et al., 2012). The Changjiang River originates from the Tanggula Mountains that are underlain by glaciers in large expanses. Dilution by stream recharged from glacial melting could be an important factor in explaining the low concentration of CDOM in the estuary. On the other hand, the Changjiang River has a relatively high sediment load compared to other major rivers in the world with an average annual discharge of 0.39 billion of tons of sediments from 1951 to 2010, (Changjiang Sediment Bulletin of 2012, available in Chinese at http://www.cjh.com.cn/fileupload/2013-09-13/130913171010455.pdf). The high load of suspended matter prevents the light from penetrating into the water and limits the photosynthesis of phytoplankton in the upper layer. Therefore, the aquatic sources of CDOM from photosynthesis could be reduced significantly in the estuarine waters.

4.2. Spatial-temporal variation of CDOM in the Changjiang estuary

The northwest-southeast gradient of decreasing CDOM observed in this study could be attributed to the mixing of freshwater and seawater. The freshwater end-member contains much higher CDOM than seawater in the Changjiang estuary as shown in this study as well as in Guo et al. (2014). However, the influence of the Changjiang's runoff weakens with the increasing distance from the river mouth, resulting in the observed gradient. This kind of spatial distribution is in accordance with previous studies in this area (Lei et al., 2012; Sun et al., 2014), and was also observed in other river-influenced estuaries such as the St. Lawrence Estuary (Nieke et al., 1997), the Southern Bight of the North Sea (Warnock et al., 1999) and two estuaries of Canadian Arctic (Retamal et al., 2007). Note that the spatial distribution of CDOM in February 2012 (Fig. 3b) did not follow this gradient. A more likely explanation could be that the samples were collected during a tidal cycle. The tide type in the Hangzhou Bay and adjacent waters is semidiurnal and the sampling interval in the February cruise was less than 1 h. Therefore, CDOM variability could be affected by the resuspension of bottom sediments during the reverse phase of the tide. During the resuspension, bottom sediments with terrestrial matter are injected into overlying waters and thus increase the CDOM concentration (Bodineau et al., 1998).

The seasonal difference in CDOM, between March 2012 (Fig. 3c) and July 2012 (Fig. 3d), was mainly because of the changes in the Changjiang's runoff, which was almost doubled from March to July in 2012 (See also the Changjiang Sediment Bulletin of 2012), resulting in a higher level of CDOM in summer. In addition, due to the weak freshwater input in winter, tidal movement could be dominant in the coastal region, resulting in the low level of CDOM (Guo et al., 2014). However, the seasonal variation of CDOM was not obvious for the two stations in the Hangzhou bay (Fig. 3c and d), presenting a relatively stable level of CDOM in the Hangzhou Bay. In the turbid waters like the Hangzhou Bay, the observed seasonal variation of CDOM level could be the affected and balanced by several factors, such as CDOM could be removed by flocculation and particle sorption, added by particle desorption and degradation of particulate organic matter (POM), or released from interstitial water enriched with CDOM by resuspension (Guo et al., 2014). It is worthy to notice that sediment resuspension in the Hangzhou Bay is significant in winter due to the distinct structure of the bay and the mutual effects by wind and tide (Chen et al., 2003).

The vertical distribution of a_g (355) was more stratified in coastal waters and peaked in the surface layer, which might be due to hydrological dynamics and the mixing behavior between fresh river discharge and saline ocean water. Similar phenomena were also observed in other estuarine regions (Del Castillo and Miller, 2011; Gardner et al., 2005). The stratification was not significant in most offshore stations, which can be explained by the weak influence of the runoff.

Comparing with reported results from Guo et al. (2007) and Zhang et al. (2013), the interannual variation of a_g (355) in the summer are in general small in the estuarine and coastal waters. Note that the tabulated data in Table 1 for the results from Guo et al. (2007) and Zhang et al. (2013) included samples collected in the river's mouth with high absorption. The limited interannual variation of CDOM indicates that parameterization of CDOM absorption coefficient could adopt the locally defined parameters (*e.g.*, average a_g (355) and S_{400}) in reducing the uncertainties of the retrieval models on inherent optical properties (IOPs). This is mainly because CDOM contributes a large portion to the total absorption coefficient in most estuaries, and is therefore sensitive in IOPs-retrieval models (Garver and Siegel, 1997; Lee et al. 2002, 2007; Maritorena et al., 2002; Yang et al., 2013; Yu et al., 2016). However, CDOM parameterization (Eq. (3)) in these models always requires an empirical coefficient (*e.g.*, S_{440}) and a locally defined value from field measurements would meet the demand.

4.3. CDOM sources

Negative correlation between salinity and a_g (355) (Fig. 5a) suggests that terrestrial inputs were the main source of CDOM in the study area, especially for the riverine end observations with low salinity samples (Fig. 5b). The lower Pearson's coefficient between salinity and a_g (355) in high salinity water (>32 PSU) indicates a weak influence of runoff and a decreased proportion of terrigenous CDOM in the marine end observations. The same conclusion can be drawn from the relationship between salinity and $S_{275\text{-}295}$ on one hand, and the dimensionless ratio S_R on the other. The smaller $S_{275\text{-}295}$ and S_R are corresponding to the higher MW, which is mostly determined by humic acid from terrestrial sources (Helms et al., 2008). As shown in Fig. 6, the DOM shifted from high MW to low MW with increasing salinity, corresponding to a decreasing proportion of terrigenous CDOM. One important fact is that DOM molecular size was reduced due to photobleaching of terrigenous CDOM during downstream transit, resulting in larger $S_{275\text{-}295}$ with increasing salinity (Helms et al., 2008). Additionally, responses of CDOM absorption at 275 nm and 295 nm to the exposure to solar radiation are different, where a much greater fractional decrease in a_g (295) was found than in a_g (275), resulting in a smaller $S_{275\text{-}295}$ upstream (Del Vecchio and Blough, 2002, 2004; Fichot and Benner, 2012). Another factor that determined the high $S_{275\text{-}295}$ in marine end waters could be attributed to the aquatic CDOM, especially the protein-rich plankton CDOM production, which has been proven to have significant effect on the $S_{275\text{-}295}$ (Fichot and Benner, 2012).

Compared to the $S_{275\text{-}295}$, the S_R seems to be a more effective indicator of the proportion of terrigenous DOM in this study, since the negative relationships between S_R and salinity are more robust for the three types of waters (*i.e.*, riverine-dominant water, mixed water and marine-dominant water) as shown in Fig. 6b. This is mainly due to the fact that the values of the spectral slope $S_{350\text{-}400}$ generally decreased with increasing salinity (blue circles in Fig. 8), which is consistent with Helms et al. (2008). The increase of S_R is mainly due to the increase of $S_{275\text{-}295}$, while the decreasing trend of $S_{350\text{-}400}$ enhanced the sensitivity of S_R to terrigenous DOM.

4.4. DOC sources and estimation

The range of DOC concentration in this study (*i.e.*, 60.54 − 217.06 μmol/L) is comparable with the previous investigations in summer in the Changjiang estuarine and coastal waters with DOC concentration ranging from 139.4 to 220 μmol/L from Zhang et al. (2013). However, our results are higher than the investigations in spring from Sun et al. (2014) with DOC concentration ranging from 41.7 to 126.2 μmol/L. The difference between our results and results from Sun et al. (2014) could be firstly attributed to the difference in sampling sites, and secondly due to the larger Changjiang runoff in summer than in spring. On the other hand, DOC concentrations in this study, in agreement with CDOM absorption coefficient, are much lower than in other world's major estuaries. For instance, the DOC concentration in the Amazon River estuary was averaged at 350 ± 75 μmol/L (Ward et al., 2015), while it reached 446 μmol/L (Chen and Gardner, 2004) and 611 μmol/L (Fichot and Benner, 2012) in the Mississippi River estuary, respectively. The DOC concentration even reached as high as 800 μmol/L in the Congo River estuary (Spencer et al., 2009b). It is likely that the same factors driving the observed distribution of CDOM result in the relatively low DOC concentrations in the Changjiang estuary.

As shown in Table 1, the mixing behavior between DOC and CDOM in most estuarine and coastal waters is conservative. However, this mixing behavior between the two parameters could be highly variable at seasonal and regional scales. For example, non-conservative mixing between DOC and CDOM can be found in this study, as well as in previous study in the Changjiang estuary (Cauwet and Mackenzie, 1993), while a conservative mixing was also reported by Zhang et al. (2013) and Liu et al. (2014). The variation of DOC-CDOM relationship could be attributed to the asynchronous changes between DOC and CDOM (Vodacek et al., 1997). The discrepancy between DOC and CDOM in their stabilities and residence period is always significant in estuarine and coastal waters, because DOC are commonly unstable and easily decomposed compared to the CDOM in coastal waters (Spitzy and Ittekkot, 1986). Another factor resulting in the discrepancy could be attributed to phytoplankton production. It seems that DOC and CDOM produced by phytoplankton production is disproportionate, resulting in the collapse of DOC-CDOM correlation in significant phytoplankton production zones (Liu et al., 2013, 2014).

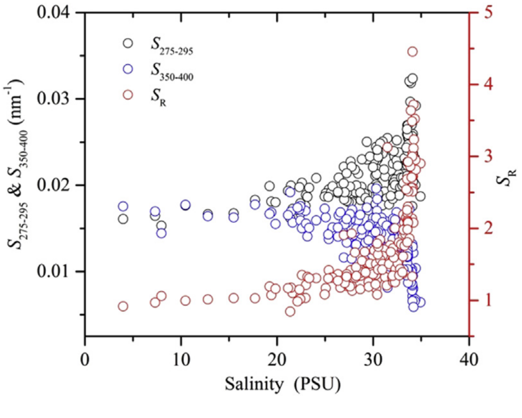


Fig. 8. The trends of the spectral slopes ($S_{275\text{-}295}$ and $S_{350\text{-}400}$) and the slope ratio S_R with increasing salinity.

Despite the discrepancy between DOC and CDOM, the developed non-linear model in this study presents a robust method to estimate DOC concentration from $S_{275\text{-}295}$ and a_g (355) as shown in Fig. 7. Note that despite S_R is a more indicative parameter to terrestrial DOM in this study, it seems to be independent to DOC, which is consistent with results from Helms et al. (2008). The developed model can be considered as satisfactory and the validation result is not biased since the adopted GeoCal/Val model (Salama et al., 2012) ensures that both calibration and validation datasets are independent to each other. Considering the phytoplankton production may result in the discrepancy between DOC and CDOM in this study, the non-linear model developed seems less vulnerable to the effect of phytoplankton production. This, on the other hand, shows that $S_{275\text{-}295}$ could be a reliable parameter to estimated DOC concentration in river-influenced estuarine and coastal waters.

Furthermore, considering the nature of this non-linear relationship is empirical, applying this relationship to different regions could require a fine-tuning on the empirical coefficients. For instance, adopting the empirical coefficients from Fichot and Benner (2012) will overestimate DOC concentration in the Changjiang estuarine and coastal waters as shown in Fig. 7a. On the other hand, as shown in this study, the developed relationship worked well for observations collected from both July 2011 and August 2013, indicating this relationship would be reliable and practical in the summer season in the Changjiang estuarine and coastal waters. Therefore, with the regionally defined DOC-CDOM relationship, DOC concentration can be potentially estimated from earth observation data since the CDOM absorption coefficient can be retrieved from remote sensing reflectance with acceptable accuracy (Lee et al., 2002; Matsuoka et al., 2012b; Siswanto et al., 2011).

5. Conclusion

Field measurements on CDOM optical properties and DOC concentration allow the investigation and discussion on the spatial-temporal variation of CDOM and DOC, the sources of CDOM and DOC and their correlations in the Changjiang estuarine and coastal waters. The following conclusions are drawn based on the results presented in this study:

1. CDOM absorption coefficient at 355 nm was negatively correlated with salinity, resulting in a decreasing gradient of a_g (355) from northwest to southeast. The spatial-temporal variation of a_g (355) could be highly dependent on the influence of Changjiang runoff. The vertical variation of CDOM was more pronounced in near-shore than that in offshore.
2. The sources of CDOM in the estuary were mainly from terrigenous inputs, while the level of CDOM could be determined by several factors in the drainage basin and the origin of recharged streams. Sediment resuspension could also be an important source of CDOM in winter, while aquatic sources could

contribute large proportions of CDOM in marine-dominant waters with high salinities.

3. A non-linear relationship was developed to estimate DOC concentration from a_g (355) and $S_{275\text{-}295}$, which was proven to be robust and stable for the Changjiang estuarine and coastal waters. Therefore, the developed relationship provides an alternative approach to estimate DOC at large spatial scales from remotely retrieved CDOM optical properties over the Changjiang estuarine and coastal waters.

Acknowledgements

This study was supported by the gs1:National Science Foundation of China (No.41271375) and the Doctoral Fund of the Ministry of Education of China (No.20120076110009). The authors would like to thank Cui Ying, Zhang Guosen, Zhang Jinfang, Chen Yu, Liu Meng and Peng Xiangyi for their helps with field and laboratory work. The authors are very grateful to Dr. Mhd. Suhyb Salama and Prof. Wouter Verhoef for their constructive comments and help in improving the quality of English writing.

References

Asmala, E., Stedmon, C.A., Thomas, D.N., 2012. Linking CDOM spectral absorption to dissolved organic carbon concentrations and loadings in boreal estuaries. Estuar. Coast. Shelf Sci. 111, 107–117.

Astoreca, R., Rousseau, V., Lancelot, C., 2009. Coloured dissolved organic matter (CDOM) in Southern North Sea waters: optical characterization and possible origin. Estuar. Coast. Shelf Sci. 85, 633–640.

Baker, A., Spencer, R.G., 2004. Characterization of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy. Sci. total Environ. 333, 217–232.

Bodineau, L., Thoumelin, G., Beghin, V., Wartel, M., 1998. Tidal time-scale changes in the composition of particulate organic matter within the estuarine turbidity maximum zone in the macrotidal Seine estuary, France: the use of fatty acid and sterol biomarkers. Estuar. Coast. Shelf Sci. 47, 37–49.

Bricaud, A., Morel, A., Prieur, L., 1981. Absorption by dissolved organic-matter of the sea (yellow substance) in the Uv and visible domains. Limnol. Oceanogr. 26, 43–53.

Carder, K.L., Steward, R.G., Harvey, G.R., Ortner, P.B., 1989. Marine humic and fulvic-acids – their effects on remote-sensing of ocean chlorophyll. Limnol. Oceanogr. 34, 68–81.

Cauwet, G., Mackenzie, F.T., 1993. Carbon inputs and distribution in estuaries of turbid rivers: the Yang Tze and Yellow rivers (China). Mar. Chem. 43, 235–246.

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

Chen, S.L., Zhang, G.A., Yang, S.L., 2003. Temporal and spatial changes of suspended sediment concentration and resuspension in the Yangtze River estuary. J. Geogr. Sci. 13, 498–506.

Chen, Z.Q., Li, Y., Pan, J.M., 2004. Distributions of colored dissolved organic matter and dissolved organic carbon in the Pearl River Estuary, China. Cont. Shelf Res. 24, 1845–1856.

Del Castillo, C.E., Coble, P.G., Morell, J.M., L O Pez, J.M., Corredor, J.E., 1999. Analysis of the optical properties of the Orinoco River plume by absorption and fluorescence spectroscopy. Mar. Chem. 66, 35–51.

Del Castillo, C.E., Gilbes, F., Coble, P.G., Müller-Karger, F.E., 2000. On the dispersal of riverine colored dissolved organic matter over the West Florida Shelf. Limnol. Oceanogr. 45, 1425–1432.

Del Castillo, C.E., Miller, R.L., 2011. Horizontal and vertical distributions of colored dissolved organic matter during the Southern Ocean Gas Exchange Experiment. J. Geophys. Res. 116, C00F07.

Del Vecchio, R., Blough, N.V., 2002. Photobleaching of chromophoric dissolved organic matter in natural waters: kinetics and modeling. Mar. Chem. 78, 231–253.

Del Vecchio, R., Blough, N.V., 2004. On the origin of the optical properties of humic substances. Environ. Sci. Technol. 38, 3885–3891.

Ferrari, G.M., 2000. The relationship between chromophoric dissolved organic matter and dissolved organic carbon in the European Atlantic coastal area and in the West Mediterranean Sea (Gulf of Lions). Mar. Chem. 70, 339–357.

Fichot, C.G., Benner, R., 2011. A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. Geophys. Res. Lett. 38, L03610.

Fichot, C.G., Benner, R., 2012. The spectral slope coefficient of chromophoric dissolved organic matter (S275-295) as a tracer of terrigenous dissolved organic carbon in river-influenced ocean margins. Limnol. Oceanogr. 57, 1453–1466.

Gardner, G.B., Chen, R.F., Berry, A., 2005. High-resolution measurements of chromophoric dissolved organic matter (CDOM) in the Neponset River estuary, boston Harbor, MA. Mar. Chem. 96, 137–154.

Garver, S.A., Siegel, D.A., 1997. Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation .1. Time series from the Sargasso Sea. J. Geophys. Res. Oceans 102, 18607–18625.

Granskog, M.A., 2012. Changes in spectral slopes of colored dissolved organic matter absorption with mixing and removal in a terrestrially dominated marine system (Hudson Bay, Canada). Mar. Chem. 134–135, 10–17.

Green, S.A., Blough, N.V., 1994. Optical-absorption and fluorescence properties of chromophoric dissolved organic-matter in natural-waters. Limnol. Oceanogr. 39, 1903–1916.

Guo, W.D., Stedmon, C.A., Han, Y.C., Wu, F., Yu, X.X., Hu, M.H., 2007. The conservative and non-conservative behavior of chromophoric dissolved organic matter in Chinese estuarine waters. Mar. Chem. 107, 357–366.

Guo, W.D., Yang, L.Y., Zhai, W.D., Chen, W.Z., Osburn, C.L., Huang, X., Li, Y., 2014. Runoff-mediated seasonal oscillation in the dynamics of dissolved organic matter in different branches of a large bifurcated estuary—The Changjiang Estuary. J. Geophys. Res. Biogeosci. 119, 776–793.

Hansell, D.A., Carlson, C.A., 1998. Deep-ocean gradients in the concentration of dissolved organic carbon. Nature 395, 263–266.

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol. Oceanogr. 53, 955–969.

Hong, H.S., Wu, J.Y., Shang, S.L., Hu, C.M., 2005. Absorption and fluorescence of chromophoric dissolved organic matter in the Pearl River Estuary, South China. Mar. Chem. 97, 78–89.

Huang, Y., Salama, M., Krol, M.S., Su, Z., Hoekstra, A.Y., Zeng, Y., Zhou, Y., 2015. Estimation of human-induced changes in terrestrial water storage through integration of GRACE satellite detection and hydrological modeling: a case study of the Yangtze River basin. Water Resour. Res. 51, 8494–8516.

Huang, Y., Salama, M.S., Su, Z., van der Velde, R., Zheng, D., Krol, M.S., Hoekstra, A.Y., Zhou, Y., 2016. Effects of roughness length parameterizations on regional-scale land surface modeling of alpine grasslands in the Yangtze River Basin. J. Hydrometeorol. 17, 1069–1085.

Keith, D.J., Yoder, J.A., Freeman, S.A., 2002. Spatial and temporal distribution of coloured dissolved organic matter (CDOM) in Narragansett Bay, Rhode Island: implications for phytoplankton in coastal waters. Estuar. Coast. Shelf Sci. 55, 705–717.

Kuliński, K., Pempkowiak, J., 2008. Dissolved organic carbon in the southern Baltic Sea: quantification of factors affecting its distribution. Estuar. Coast. Shelf Sci. 78, 38–44.

Lagerloef, G., Swift, C.T., Le Vine, D.M, 1995. Sea surface salinity: the next remote sensing challenge. Oceanography 8, 44–50.

Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K.L., Davis, C., 2007. Euphotic zone depth: its derivation and implication to ocean-color remote sensing. J. Geophys. Res. Oceans 112, C03009.

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

Lei, H., Pan, D.L., Bai, Y., Tao, B.Y., Sun, J., Zhang, L., Zhang, X., 2012. The proportions and variations of the light absorption coefficients of major ocean color components in the East China Sea. ACTA Oceanol. Sin. 31, 45–61.

Liu, Q., Pan, D., Bai, Y., Wu, K., Chen, C.-T.A., Sun, J., Zhang, L., 2013. The satellite reversion of dissolved organic carbon (DOC) based on the analysis of the mixing behavior of DOC and colored dissolved organic matter: the East China Sea as an example. ACTA Oceanol. Sin. 32, 1–11.

Liu, Q., Pan, D.L., Bai, Y., Wu, K., Chen, C.T.A., Liu, Z.L., Zhang, L., 2014. Estimating dissolved organic carbon inventories in the East China Sea using remote-sensing data. J. Geophys. Res. Oceans 119, 6557–6574.

Longhurst, A.R., Harrison, W.G., 1989. The biological pump - profiles of plankton production and consumption in the upper ocean. Prog. Oceanogr. 22, 47–123.

Ludwig, W., Probst, J.L., Kempe, S., 1996. Predicting the oceanic input of organic carbon by continental erosion. Glob. Biogeochem. Cycles 10, 23–41.

Mannino, A., Russ, M.E., Hooker, S.B., 2008. Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the US Middle Atlantic Bight. J. Geophys. Res. Oceans 113, C07051.

Maritorena, S., Siegel, D.A., Peterson, A.R., 2002. Optimization of a semianalytical ocean color model for global-scale applications. Appl. Opt. 41, 2705–2714.

Matsuoka, A., Bricaud, A., Benner, R., Para, J., Sempere, R., Prieur, L., Belanger, S., Babin, M., 2012a. Tracing the transport of colored dissolved organic matter in water masses of the Southern Beaufort Sea: relationship with hydrographic characteristics. Biogeosciences 9, 925–940.

Matsuoka, A., Hooker, S.B., Bricaud, A., Gentili, B., Babin, M., 2012b. Estimating absorption coefficients of colored dissolved organic matter (CDOM) using a semi-analytical algorithm for Southern Beaufort Sea (Canadian Arctic) waters: application to deriving concentrations of dissolved organic carbon from space. Biogeosci. Discuss. 9, 13743–13771.

Nieke, B., Reuter, R., Heuermann, R., Wang, H., Babin, M., Therriault, J.C., 1997. Light absorption and fluorescence properties of chromophoric dissolved organic matter (CDOM), in the St Lawrence Estuary (Case 2 waters). Cont. Shelf Res. 17, 235–252.

Pan, D., Liu, Q., Bai, Y., 2012. Progress in remote sensing of DOC:based on the analysis of conservative behaviors of DOC and CDOM in globla rivers estuaries. ACTA Oceanol. Sin. 34, 1–11 (In Chinese with English abstract).

Pegau, W.S., Gray, D., Zaneveld, J.R.V., 1997. Absorption and attenuation of visible and near-infrared light in water: dependence on temperature and salinity. Appl. Opt. 36, 6035–6046.

Retamal, L., Vincent, W.F., Martineau, C., Osburn, C.L., 2007. Comparison of the optical properties of dissolved organic matter in two river-influenced coastal regions of the Canadian Arctic. Estuar. Coast. Shelf Sci. 72, 261–272.

Rochelle-Newall, E., Hulot, F.D., Janeau, J.L., Merroune, A., 2014. CDOM fluorescence as a proxy of DOC concentration in natural waters: a comparison of four contrasting tropical systems. Environ. Monit. Assess. 186, 589–596.

Rochelle-Newall, E.J., Fisher, T.R., 2002. Chromophoric dissolved organic matter and dissolved organic carbon in Chesapeake Bay. Mar. Chem. 77, 23–41.

Salama, M.S., Van der Velde, R., van der Woerd, H.J., Kromkamp, J.C., Philippart, C.J.M., Joseph, A.T., O'Neill, P.E., Lang, R.H., Gish, T., Werdell, P.J., Su, Z., 2012. Technical Note: calibration and validation of geophysical observation models. Biogeosciences 9, 2195–2201.

Siswanto, E., Tang, J.W., Yamaguchi, H., Ahn, Y.H., Ishizaka, J., Yoo, S., Kim, S.W., Kiyomoto, Y., Yamada, K., Chiang, C., Kawamura, H., 2011. Empirical ocean-color algorithms to retrieve chlorophyll-a, total suspended matter, and colored dissolved organic matter absorption coefficient in the Yellow and East China Seas. J. Oceanogr. 67, 627–650.

Spencer, R.G.M., Ahad, J.M.E., Baker, A., Cowie, G.L., Ganeshram, R., Upstill-Goddard, R.C., Uher, G., 2007. The estuarine mixing behaviour of peatland derived dissolved organic carbon and its relationship to chromophoric dissolved organic matter in two North Sea estuaries (UK). Estuar. Coast. Shelf Sci. 74, 131–144.

Spencer, R.G.M., Aiken, G.R., Butler, K.D., Dornblaser, M.M., Striegl, R.G., Hernes, P.J., 2009a. Utilizing chromophoric dissolved organic matter measurements to derive export and reactivity of dissolved organic carbon exported to the Arctic Ocean: a case study of the Yukon River, Alaska. Geophys. Res. Lett. 36, L06401.

Spencer, R.G.M., Stubbins, A., Hernes, P.J., Baker, A., Mopper, K., Aufdenkampe, A.K., Dyda, R.Y., Mwamba, V.L., Mangangu, A.M., Wabakanghanzi, J.N., Six, J., 2009b. Photochemical degradation of dissolved organic matter and dissolved lignin phenols from the Congo River. J. Geophys. Res. Biogeosci. 114, G03010.

Spitzy, A., Ittekkot, V., 1986. Gelbstoff: an uncharacterized fraction of dissolved organic carbon. In: The Influence of Yellow Substances on Remote Sensing of Sea-water Constituents from Space, Vol. 2. GKSS, Geesthacht, Germany, p. 31. Appendix 1.

Sun, Q., Wang, C., Wang, P., Hou, J., Ao, Y., 2014. Absorption and fluorescence characteristics of chromophoric dissolved organic matter in the Yangtze Estuary. Environ. Sci. Pollut. Res. 21, 3460–3473.

Vodacek, A., Blough, N.V., DeGrandpre, M.D., Peltzer, E.T., Nelson, R.K., 1997. Seasonal variation of CDOM and DOC in the middle Atlantic Bight: terrestrial inputs and photooxidation. Limnol. Oceanogr. 42, 674–686.

Ward, N.D., Krusche, A.V., Sawakuchi, H.O., Brito, D.C., Cunha, A.C., Moura, J.M.S., da Silva, R., Yager, P.L., Keil, R.G., Richey, J.E., 2015. The compositional evolution of dissolved and particulate organic matter along the lower Amazon River—Óbidos to the ocean. Mar. Chem. 177, 244–256.

Warnock, R.E., Gieskes, W.W.C., van Laar, S., 1999. Regional and seasonal differences in light absorption by yellow substance in the Southern Bight of the North Sea. J. Sea Res. 42, 169–178.

Xie, H.X., Aubry, C., Belanger, S., Song, G.S., 2012. The dynamics of absorption coefficients of CDOM and particles in the St. Lawrence estuarine system: biogeochemical and physical implications. Mar. Chem. 128, 44–56.

Yang, W., Matsushita, B., Chen, J., Yoshimura, K., Fukushima, T., 2013. Retrieval of inherent optical properties for turbid inland waters from remote-sensing reflectance. Ieee Trans. Geosci. Remote Sens. 51, 3761–3773.

Yu, X., Salama, M.S., Shen, F., Verhoef, W., 2016. Retrieval of the diffuse attenuation coefficient from GOCI images using the 2SeaColor model: a case study in the Yangtze Estuary. Remote Sens. Environ. 175, 109–119.

Zhang, X.Y., Chen, X., Deng, H., Du, Y., Jin, H.Y., 2013. Absorption features of chromophoric dissolved organic matter (CDOM) and tracing implication for dissolved organic carbon (DOC) in Changjiang Estuary, China. Biogeosci. Discuss. 10, 12217–12250.

Ort, Datum: Bremerhaven, 17.7.2022

Versicherung an Eides Statt

Ich, Yangyang Liu, (Matr.-Nr. 3058650)

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.
Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.
Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Bremerhaven, 17.7.2022

Ort, Datum Unterschrift