Developement and exploitation of new approaches for observation of phytoplankton community composition from BGC-Argo floats in open ocean

Flavien Petit

Sorbonne Université

Ecole doctorale des sciences de l'environnement d'Ile de France (ED129)

Unité de recherche Laboratoire d'Océanographie de Villefranche-sur-Mer

Thèse présentée par **Flavien Petit**

Soutenue le 17 mars 2023

En vue de l'obtention du grade de docteur de Sorbonne Université

Discipline Océan Atmosphère, Climat et Observations spatiales

Spécialité Océanographie

Development and exploitation of new methods for observing phytoplankton community composition from BGC-Argo profiling floats in the open ocean

---

Thèse dirigée par	Hervé Claustre	directeur
	Julia Uitz	co-directrice

Composition du jury

Président	Damien Cardinal	Professeur à Sorbonne Université
Rapporteurs	Astrid Bracher	Professeure à l'Alfred Wegener Institut
	Yannick Huot	Professeur à l'Université de Sherbrooke
Examinateurs	Christina Schallenberg	Chargée de recherche à l'Université de Tasmanie
	Bob Brewin	Maître de conférence à l'Université de Exeter
	Anne Petrenko	Maître de conférence au Meditteranean Institut of Oceanography
Directeurs de thèse	Hervé Claustre	Directeur de recherche au CNRS
	Julia Uitz	Chargée de recherche au CNRS

Mots clés : Phytoplancton, fluorescence, observations *in-situ*, pigments, optique marine, composition des communautées phytoplanctoniques

Keywords : Phytoplankton, fluorescence, in-situ observations, pigments, marine optics, phytoplankton community composition

Cette thèse a été préparé au Laboratoire d'Océanographie de Villefranche-sur-mer

Laboratoire d'Océanographie de Villefranche-sur-Mer
Sorbonne Université
181 Chemin du Lazaret
06230 Villefranche-sur-Mer
France
Site : https://lov.imev-mer.fr

Remerciements

Je tiens tout d'abord à remercier Julia et Hervé, pour la confiance qu'ils m'ont accordée en me proposant cette thèse. Votre encadrement m'a permis de m'émanciper dans mon travail au cours de ces années. Vos précieux conseils ont pu guider ma réflexion et m'ont beaucoup appris, tant sur le plan scientifique que méthodologique.

Merci Julia d'avoir toujours été là quand il le fallait, bénéficier de ton expertise scientifique sur chaque sujet abordé dans cette thèse a été un atout majeur dans la réussite de cette thèse. Merci aussi pour les corrections pleines de conseils et d'explications qui m'ont été d'une aide précieuse pour mieux appréhender l'écriture scientifique. Et surtout merci pour ton enthousiasme qui a été une source de motivation durant ces trois années.

Merci Hervé d'avoir toujours écouté mes idées, échanger ensemble sur les différents points clés de cette thèse a toujours été un vrai plaisir. Merci de m'avoir toujours poussé et fait confiance pour produire des méthodes et analyses innovantes.

Merci aux membres du jury, Astrid, Yannick, Damien, Robert, Christina et Anne d'avoir pris le temps de lire mon travail et de l'évaluer.

Merci à toutes les personnes qui m'ont aidé dans la collecte et l'analyse de données.

Merci à Melek, Céline, Emilie et Enzo de m'avoir accueillis à bord des missions BOUSSOLE, même s'il a fallut bien serrer le planning pour que toutes ces manips puissent tenir. Grâce à vous, embarquer a toujours été un réel plaisir, et j'ai beaucoup appris à vos cotés. Melek je te remercie aussi pour toutes ces heures passées à identifier le phyto.

Merci à toute la team roscovite, Louison, Fred, Laurence pour m'avoir accueilli et aider dans mes mesures de labo. Evidemment énorme merci à toi Louison, tu as fait un travail colossal, sans toi aucune manip labo n'aurait abouti. On a quand même bien rigolé dans notre petite salle fluorescence.

Merci Céline et Joséphine pour toutes ces données HPLC, et pour avoir toujours trouvé un petit créneau pour analyser mes échantillons.

Merci aussi à tous les collègues de l'équipe OMTAB, pour toutes ces discussions scientifiques ou non, au coin café ou au détour d'un couloir. Et surtout merci pour les apérOMTAB !

Merci spécial à la Mezzanine. Marine, Loulou, Marin, Elsa, Loïc, on a vraiment les meilleurs thésards sur ce bureau. Merci pour la bonne humeur, les craquages, les pauses thés, les bières et les ambiances Noël 4 mois dans l'année.

Merci aussi aux amis thésards de la station zoo, Ophélie pour les sorties grimpes et autres joyeusetés sportives, Flo pour les sorties vélo, Thelma pour nos discussions autour d'un thé.

Merci à Thomas et Manon, le bureau de la joie et de la bonne humeur, j'ai une revanche à prendre sur Mario Kart par contre.

Merci à tous les nouveaux copains que j'ai rencontrés à Villefranche, Raph, Steve, Camille, Laurie, Alex, Théo, Noémie. La meilleure équipe pour trinquer, en terrasse ou sur la plage.

Merci Thomas de m'avoir révélé tous les secrets du surf en Méditerranée, toutes ces histoires de vagues si rares et ses heures à débattre du meilleur modèle météo, de la meilleure orientation.

Merci à la team du Mont Gros, pour nos soirées voisins, les soirées Arkose, les sorties falaise, les trails, les pan bagna sans thon ou même les fifa tard le soir. Issa MGTT !

Merci aux amis parisiens, le sang, comme on dit. PA, Hanaé, Nico, merci pour les nuits de fêtes mémorables à Biarritz, Paris ou Roscoff. Vivement les prochaines ! PA il faut que tu viennes découvrir le surf en Med ! Tu verras c'est pas pire que les break de Saint Gilles.

Merci Matthieu, mon bro, grâce à nos conversations je crois que je suis devenu plus intelligent et j'ai pu mener à bien cette thèse.

Merci aussi à la famille niçoise, Domi, Bertrand et Gilles pour les repas dominicaux et tous les bons moments partagés.

Merci à toi Amanda. Merci pour tous ces petits moments que l'on a vécu à deux qui ont embelli ma vie. Merci aussi de m'avoir soutenu et encouragé (notamment pendant les trails). Tu as toujours su me faire voir la vie : «le verre à moitié plein». Nos parenthèses à deux en Italie ou dans l'arrière-pays sont mes meilleurs souvenirs de ces trois années de thèse.

Enfin, merci à ma famille. Merci à mes parents qui m'ont toujours poussé à faire ce que j'aimais et qui ont tout fait pour que j'y arrive. Merci à ma grand-mère pour tous ses encouragements. Merci à toi Marjo, quel plaisir quand je recevais un message qui me disait que tu descends sur Nice, et avec PM c'est encore mieux. Merci à vous tous d'être là pour moi.

Table of content

Abstract

Phytoplankton play a key role in the regulation of many biogeochemical cycles. It is responsible for half of the world's primary production, contributing to the marine food chain and regulating carbon fluxes between the ocean and the atmosphere. These processes vary with both phytoplankton biomass and community composition. It is therefore a critical challenge to monitor phytoplankton biomass and community composition on a global scale. The BioGeoChemical-Argo (BGC-Argo) program aims to monitor and understand key biogeochemical processes on a global scale by developing an array of profiled Argo floats equipped with an array of biogeochemical sensors. The floats measure *in-situ* fluorescence, an indicator of chlorophyll-a concentration, used as a proxy of phytoplankton biomass. However, this measurement has been shown to be highly variable on a global scale. Moreover, it remains a challenge to estimate phytoplankton community composition from BGC-Argo floats. Some methods have been previously published but are either restricted to specific oceanic region or provide a limited information. Thus, there is a crucial need to i) better assess phytoplankton biomass from BGC-Argo float and ii) develop methods to estimate phytoplankton community composition.

First, this thesis presents the role of the phytoplankton community in the variability of the in-situ fluorescence response. A database of fluorescence and chlorophyll-a concentration measurements was studied to demonstrate the key role of phytoplankton community composition on the relationship between fluorescence and chlorophyll-a concentration at different spatial and temporal scales. It revealed a significant impact of the phytoplankton community composition on the *in-situ* fluorescence signal, with an impact of the size structure of the community but also of the presence of non-photosynthetic pigments. Then, we developed two different methods to evaluate the phytoplankton community composition. The first one is based on the use of multispectral fluorescence, to estimate a taxonomic index of phytoplankton composition. We collected laboratory and field data to better understand the fluorescence response at different wavelengths as a function of phytoplankton community composition. Ultimately we propose a method that allows predictions of four different phytoplankton community composition clusters from multispectral fluorescence. The second method uses a combination of optical and hydrographic measurements to estimate the concentration of four different plankton groups in organic carbon and total particulate organic carbon concentration. This method has been validated on three different dataset covering contrasted trophic environments. These two methods allowed us to highlight the possibility of estimating the composition of the phytoplankton community from the BGC-Argo profiling floats.

Résumé

Le phytoplancton joue un rôle clé dans la régulation de nombreux cycles biogéochimiques. Il est responsable de la moitié de la production primaire mondiale, contribue à la chaîne alimentaire marine et régule les flux de carbone entre l'océan et l'atmosphère. Ces processus varient en fonction de la biomasse phytoplanctonique et de la composition de la communauté. La surveillance de la biomasse phytoplanctonique et de la composition des communautés à l'échelle mondiale constitue donc un défi majeur. Le programme BioGeoChemical-Argo (BGC-Argo) vise à surveiller et à comprendre les processus biogéochimiques clés à l'échelle mondiale en développant une série de flotteurs Argo profilés équipés d'un ensemble de capteurs biogéochimiques. Les flotteurs mesurent la fluorescence in-situ, un indicateur de la concentration en chlorophylle-a, utilisé comme indicateur de la biomasse du phytoplancton. Cependant, il a été démontré que cette mesure est très variable à l'échelle mondiale. En outre, l'estimation de la composition de la communauté phytoplanctonique à partir des flotteurs BGC-Argo reste un défi. Certaines méthodes ont déjà été publiées, mais elles sont soit limitées à une région océanique spécifique, soit ne fournissent qu'une information limitée. Il existe donc un besoin crucial de i) mieux évaluer la biomasse phytoplanctonique des flotteurs BGC-Argo et ii) de développer des méthodes pour estimer la composition de la communauté phytoplanctonique.

Tout d'abord, cette thèse présente le rôle de la communauté phytoplanctonique dans la variabilité de la réponse de fluorescence in-situ. Une base de données de mesures de fluorescence et de concentration en chlorophylle-a a été étudiée pour démontrer le rôle clé de la composition de la communauté phytoplanctonique sur la relation entre la fluorescence et la concentration en chlorophylle-a à différentes échelles spatiales et temporelles. Elle a révélé un impact significatif de la composition de la communauté phytoplanctonique sur le signal de fluorescence in-situ, avec un impact de la structure de taille de la communauté mais aussi de la présence de pigments non-photosynthétiques. Ensuite, nous avons développé deux méthodes différentes pour évaluer la composition de la communauté phytoplanctonique. La première est basée sur l'utilisation de la fluorescence multispectrale, pour estimer un indice taxonomique de la composition du phytoplancton. Nous avons collecté des données en laboratoire et sur le terrain pour mieux comprendre la réponse de la fluorescence à différentes longueurs d'onde en fonction de la composition de la communauté phytoplanctonique. Finalement, nous proposons une méthode qui permet de prédire quatre groupes différents de composition de la communauté phytoplanctonique à partir de la fluorescence multispectrale. La deuxième méthode utilise une combinaison de mesures optiques et hydrographiques pour estimer la concentration de quatre groupes différents de plancton en carbone organique et la concentration totale de carbone organique particulaire. Cette méthode a été validée sur trois jeux de données différents couvrant des environnements trophiques contrastés. Ces deux méthodes nous ont permis de mettre en évidence la possibilité d'estimer la composition de la communauté phytoplanctonique à partir des flotteurs profileurs BGC-Argo.

Table of abbreviations		
Abbreviations	Significance	Units
a	Absorption	m^{-2}
a*	Specific absorption	m^{-2} (mg Chla)$^{-1}$
b_{bp}	Particulate backscattering	m^{-1}
c_p	Particulate attenuation	m^{-1}
Chla	Chlorophyll-*a*	
[Chla]	Chlorophyll-*a* concentration	mg m^{-3}
f_{micro}	The fraction of Chlorophyll-*a* concentration attributed to microphytoplankton	%
f_{nano}	The fraction of Chlorophyll-*a* concentration attributed to nanophytoplankton	%
f_{pico}	The fraction of Chlorophyll-*a* concentration attributed to picophytoplankton	%
F	Raw fluorescence signal	Digital Counts
F_{Chla} or $[Chla]_{Fluo}$	Chlorophyll-*a* concentration estimated from *in-situ* fluorescence	mg m^{-3}
F_{440}	Raw fluorescence signal from 440 nm excitation	Digital Counts
F_{470}	Raw fluorescence signal from 470 nm excitation	Digital Counts

F_{532}	Raw fluorescence signal from 532 nm excitation	Digital Counts
z_{eu}	Euphotic depth	m

I Introduction

I.1. A world ocean in a context of global changes

Since the industrial revolution, the climate and ecosystems are modified on a global scale by anthropogenic activities. The emission of greenhouse gases resulting from the combustion of fossil fuels have modified the radiative equilibrium of the Earth system leading to a global warming of the atmosphere (Trenberth et al., 2014). As a result, the mean temperature between 2013 and 2022 is 1.14°C (+/- 0.12°C) higher than the mean temperature between 1850 and 1900 (Bindoff et al., 2019). In oceanography, this warming and anthropogenic forcing are of a major concern because of their multiple consequences. The combination of these two factors have been shown to induce global warming, acidification and deoxygenation of the ocean (Bindoff et al., 2013). All of these consequences impact the functioning of marine ecosystems and biogeochemical cycles (e.g. Doney et al., 2009; Hoegh-Guldberg and Bruno, 2010; Finzi et al., 2011; Harada, 2016).

The carbon biogeochemical cycle is of a particular interest because of its major role in the regulation of the Earth's climate. The ocean plays a key role in the carbon cycle due to its size and interactions with the atmosphere (Broecker and Peng, 1974; DeVries, 2022). The total amount of carbon stored in the ocean is regulated not only by the solubility of CO_2 in cold waters followed by subduction to the deep ocean, but also by the marine biosphere and the process of photosynthesis that induce the so-called biological carbon pump (e.g. BCP, Sigman and Boyle, 2000; Berner and Kothavala, 2001). The BCP is initiated in the surface waters through net primary production (NPP) by phytoplankton organisms (Falkowski et al., 1998; Field et al., 1998; Boyd et al., 2019). Autotrophs organisms contribute to reduce DIC concentration in surface waters by the conversion of DIC in organic carbon (Ito and Follows, 2005). A part of this organic carbon is then exported to deeper water masses.

I.2. Biogeochemical significance of phytoplankton

Phytoplankton are unicellular phototrophs or mixotrophs organisms. They live in the upper layer of the water column, where the photosynthetic active radiation (PAR), *i.e.* irradiance within the spectral domain from 400 to 700 nm, is sufficient to sustain photosynthesis. This process allows phytoplankton to use the energy of solar radiation to transform dissolved CO_2 and nutrients into organic carbon in the form of sugar. This production of organic matter from dissolved CO_2 is the first step of the trophic food web and is thus called primary production. Phytoplankton are ubiquitous in all marine and freshwater environments and contribute ~50% of the world's (ocean + land) primary production (Field et al., 1998). They are the keystone of the BCP; their role is essential in the oceanic carbon cycle and, thus, in the regulation of the Earth's climate.

A fraction of the organic carbon produced by photosynthesis is exported to the deep water masses, while the other fraction is transferred to mixotrophic and heterotrophic organisms of the marine food web (Berglund et al., 2007; Herndl and Reinthaler, 2013). The mechanisms by which biogenic carbon is exported towards the deep ocean range from sedimentation of organic particles (gravitational pump), to physical mixing (mixed layer pump) and vertical migrations of organisms between the euphotic and the mesopelagic zones (Siegel et al., 2016; Boyd et al., 2019). This carbon flux toward deep ocean is attenuated by variable microbial remineralization through the whole water column (Herndl and Reinthaler, 2013).

The rate of carbon export through the BCP is then constrained by the NPP and by these exports mechanisms. While phytoplankton biomass is the first parameter that regulates the intensity of primary production in the environment, the composition of the phytoplankton communities also significantly influences the NPP (e.g. Cermeño et al., 2005; Uitz et al., 2008). Phytoplankton community composition plays also a determining role in the pathways towards which carbon is directed (Michaels and Silver, 1988; Cushing, 1989; Finkel, 2007) and, thus, in the carbon export mechanisms (Buesseler et al., 1998; Guidi et al., 2009; Briggs et al., 2011). Phytoplankton size has been shown as a particularly important aspect of community structure in the regulation of the carbon cycle in the ocean (Michaels and Silver, 1988; Hilligsøe et al., 2011). Phytoplankton organisms are typically sorted into three distinct size classes; microphytoplankton (>200 µm),

nanophytoplankton (between 2 and 20 µm) and picophytoplankton (<2 µm). In this context, understanding and assessing the dynamics of phytoplankton biomass and community composition on the global scale is an essential need to better understand and predict the carbon cycle.

I.3. Seasonal variability in phytoplankton biomass and community composition

Phytoplankton biomass and community structure vary depending on environmental conditions, following nutrient availability and light intensity. In the case of a typical seasonal dynamics, those two parameters follow an seasonal pattern that can, on average, be described as follows (Marty et al., 2002). In winter, the water column is mixed due to physical forcing, which results in a replenishment of the ocean upper layer with nutrients, but also to a decrease in the average light intensity received by phytoplankton cells in the mixed layer (Figure 1-1). The phytoplankton biomass is relatively low and dominated by nano- and picophytoplankton. In the spring, warmer weather allows an increase in solar radiative energy. The concurrent presence of nutrients due to winter mixing and increase in solar energy leads to a phytoplankton bloom (Sverdrup, 1953; Fischer et al., 2014) typically dominated by large diatoms and associated with a strong increase in primary production. But the warmer weather also induces a rise of the mixed layer depth which results in a nutrient limitation. This, combined with an increase of grazing pressure from zooplankton, induces a decay of phytoplankton biomass along all the summer season. In such low-nutrients high-light conditions, the phytoplankton assemblage is usually dominated by small organisms, prokaryotes in particular (Marty et al., 2002). Finally, in fall, the mixed layer depth deepens, sometimes leading to a second bloom, yet of moderate amplitude.

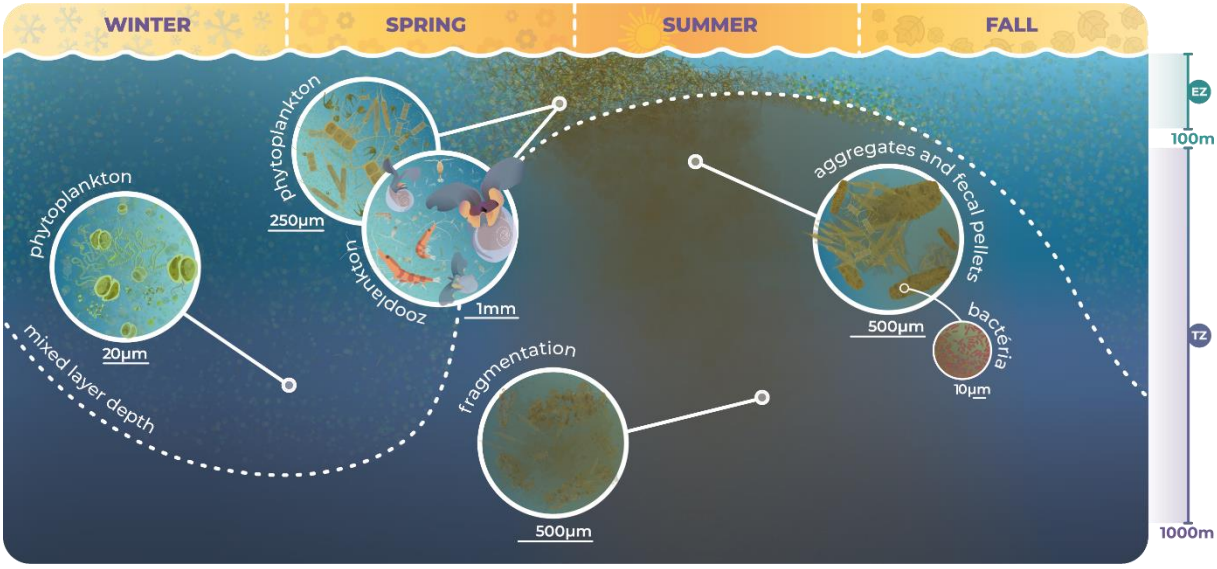

Figure 1-1 : Seasonal dynamic of the mixed layer depth associated to the nutrient availability and its consequence on the phytoplankton succession in seasonal regions. Illustration by Thomas Boniface (REFINE project; https://erc-refine.eu/).

This contrast between high biomass waters dominated by nano- and microphytoplankton and low biomass waters dominated by small phytoplankton can also be observed on a global spatial scale (Figure 1 -2). Indeed, we observe that phytoplankton biomass is higher in near-coastal areas where upwelling of nutrient-rich deep waters favour primary production, as well as at high latitudes where meteorological conditions favour convective mixing and, thus, the supply of nutrients from deep waters (Dale et al., 1999; Lacour et al., 2017). In contrast, low latitude environments, with warm conditions and strong incident solar irradiance, are characterized by shallow mixed layers and strong stratification that usually prevent nutrient injection from the deep layer to the upper ocean. In such conditions low phytoplankton biomass dominated by picophytoplankton prevails (Raven, 1998; Agawin et al., 2000). Such a strong variability in the phytoplankton biomass and community structure over a broad range of temporal and spatial scales induces equally large variability in phytoplankton-dependent processes, the carbon cycle in particular. Therefore, it appears crucial to develop a capability to assess phytoplankton biomass and community composition on the global and seasonal scales. This still represents a critical challenge in oceanography.

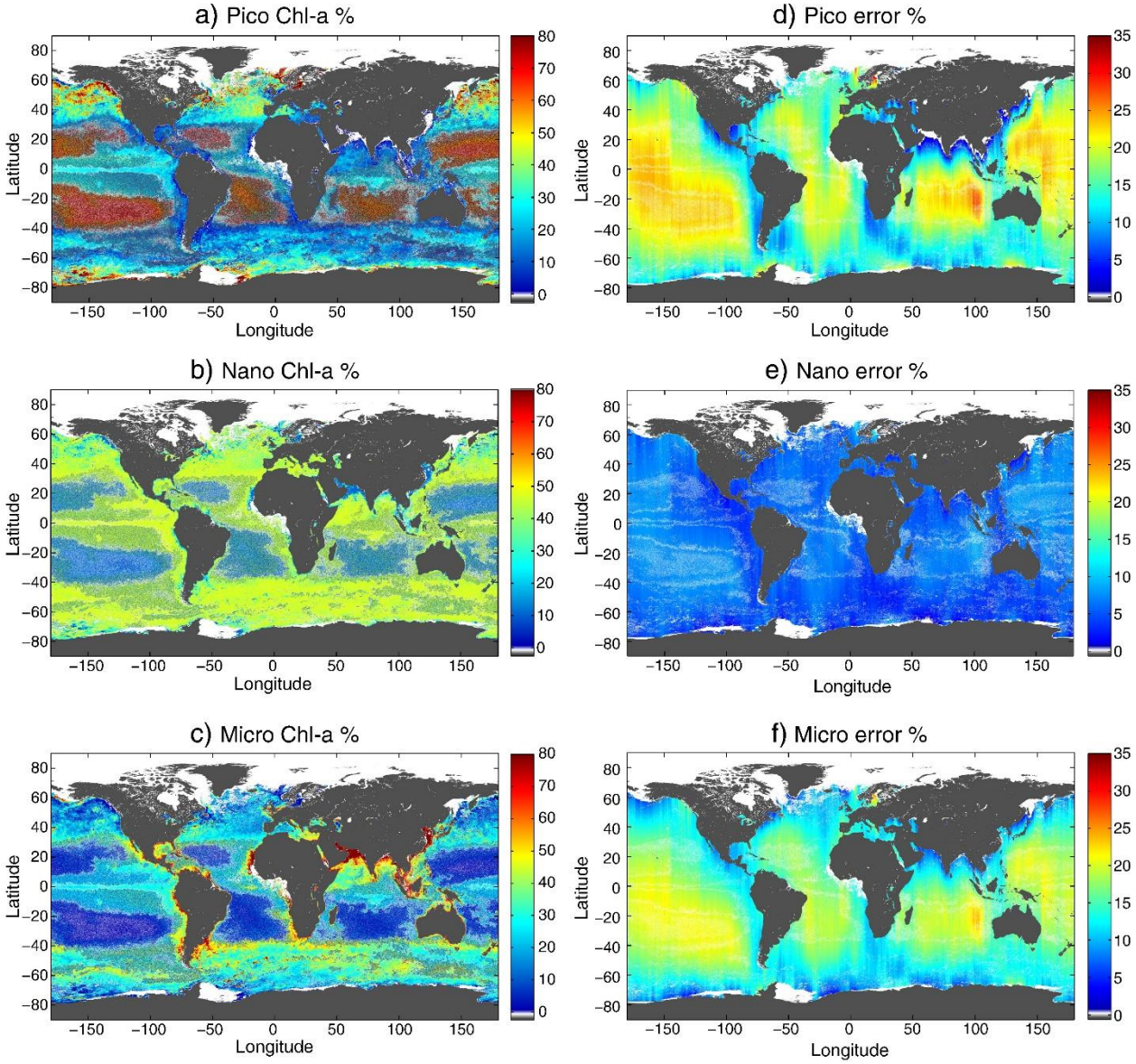

Figure 1 -2 : Global maps of (a) percentage of picoplankton, (b) percentage of nanoplankton, and (c) percentage of microplankton, estimated for the period of winter 2010–2011 (averaged over day 355, 2010 to day 79, 2011), by the method described in the text. Error budgets (as described in the text) expressed in percentages, in estimating the fractional contributions of (d) picoplankton, (e) nanoplankton, and (f) microplankton to total chlorophyll using the method presented here. Pixels with missing data are white. From Roy et al. (2013)

I.4. Determining phytoplankton biomass and composition, state of the art and challenges

Observing phytoplankton has always been a complex task. Although small and not observable to the naked eye, phytoplankton organisms have a size that spans three

different classes from picophytoplankton (< 2 µm) to nanophytoplankton (2 to 20 µm) and microphytoplankton (20 to 200 µm), which represents nine order of magnitudes of volume (Figure 1.3, Finkel et al., 2010). This often leads to the need for combining several methods to observe the entire phytoplankton assemblage, rather than just a fraction.

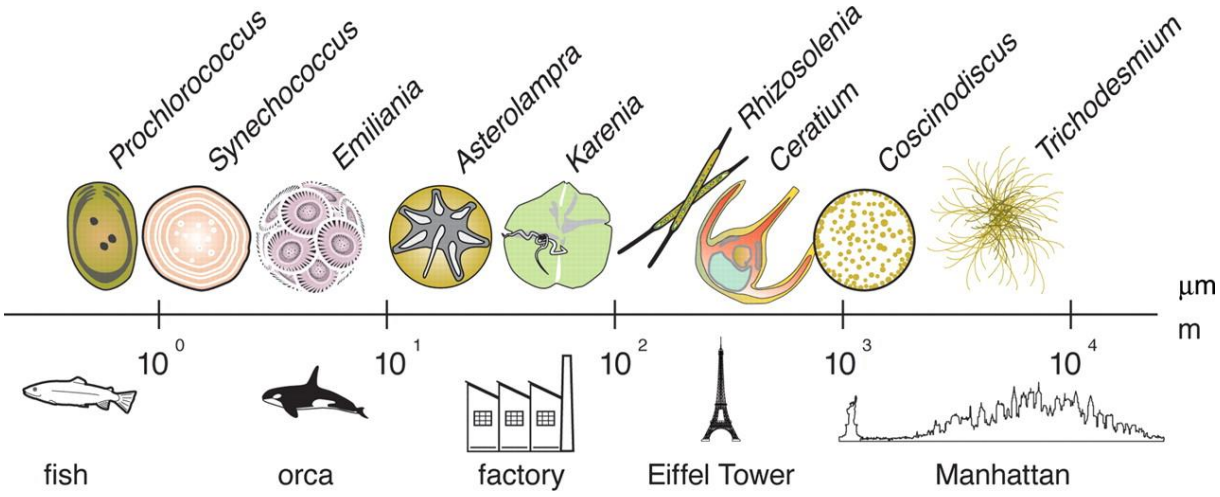

Figure 1 -3 : A comparison of the size range (maximum linear dimension) of phytoplankton relative to macroscopic objects. (Adapted from Finkel et al., 2010)

We can differentiate several methods for assessing phytoplankton biomass and/or community composition in two different categories; i) methods based on discrete seawater sampling and subsequent laboratory analysis; and ii) methods based on automatic in-situ high frequency observations.

I.4.1 Discrete measurements

Microscopy

Historically, the observation of phytoplankton cells has been done using an optical microscope. The most widely used technique for enumerating phytoplankton cells is inverted microscopy, where an inverted microscope enables to look at the bottom of a sedimented water sample (Utermöhl, 1958). This technique allows to identify, at best, species (sometimes the genus), estimate the abundance and measure the size of the cells ranging from tens of micrometres to millimetres. More recently, electron microscopes have made it possible to accurately observe smaller phytoplankton cells at higher

magnification, thus increasing the possibility of accurate taxonomic identification (Booth, 1993a). However, these methods remain very expensive and time consuming, and require a high level of expertise, which makes it difficult to implement them routinely for a large number of samples.

Flow cytometry

In order to complete the observation of the lower end of the phytoplankton size spectrum, flow cytometry is commonly used (Chisholm et al., 1988). In a flow cytometer, the phytoplankton cells individually pass through a stream of water and in front of a laser beam. The optical properties provide information about the size and structure of the cells (Dubelaar and Jonker, 2000). This method enables to identify phytoplankton taxa and estimate cellular abundance of small phytoplankton ranging from 0.3 to 20 µm. The technic has evolved and different types of flow cytometry methods are used to answer different observations needs. We can notable think of submersible flow cytometer (Sosik et al., 2003; Thyssen et al., 2014) or flow cytometer combined with imaging systems (Sosik and Olson, 2007). Unlike microscopy, this method allows rapid analysis of samples in batches, but provides information biased towards small phytoplankton taxa.

HPLC pigment analysis

In order to observe the phytoplankton community as a whole, other methods have been developed, such as pigment analysis. The different phytoplankton taxa are characterized by their pigment assemblage. The pigments present in a seawater sample can be separated, identified and quantified by High Performance Liquid Chromatography (HPLC). Chlorophyll-a (Chla) is ubiquitous among all phytoplankton taxa, excepted for Prochlorococcus which has divinyl Chlorophyll-a (DVChla). Hence the total Chla, determined as the sum between Chla and DVChla, can be used as a proxy of the total phytoplankton biomass. In contrast, several accessory pigments are biomarker of certain phytoplankton taxa and, thus, may be used to determine the community structure (Claustre, 1994; Jeffrey et al., 1997; Roy et al., 2011). Nevertheless, the chemotaxonomic biomarker approach is not without weaknesses. While a few pigments can unambiguously be assigned to a single phytoplankton taxon, other pigments can be shared between different taxa and each taxon contains different accessory pigments (Table 1.1). An

approach to mitigate this bias is to select pigments that predominantly covary with biomass at the global scale, and weight their contribution to biomass by an empirically determined factor (Mackey et al., 1996; Vidussi et al., 2001; Uitz et al., 2006).

Table 1.1 : Major pigments used for classification of phytoplankton groups. The most commonly used pigment algal-class associations are in bold.

Pigments	Taxonomic significance
Zeaxanthin	**Cyanobacteria**, Chlorophytes, Chrysophytes, Dinoflagellates
Divinyl-chlorophyll-b	**Prochlorophytes**
Chlorophyll-b	**Chlorophytes**, Prasinophytes
19'hexanoyloxyfucoxanthin	**Prymnesiophytes**, Chrysophytes, Dinoflagellates
19'butanoyloxyfucoxanthin	**Pelagophytes**, Prymnesiophytes
Alloxanthin	**Cryptophytes**
Fucoxanthin	**Diatoms**, Prymnesiophytes, Chrysophytes, Dinoflagellates
Peridinin	**Dinoflagellates**

In addition to being comprehensive (it covers the full spectrum of the phytoplankton assemblage) and quantitative, the pigment-based approach to observing phytoplankton communities is less time consuming and more practical than traditional microscopic analysis, and can therefore be applied to a large number of seawater samples. However, it only provides a proxy of the phytoplankton biomass as the pigment to carbon ratio (or the intracellular pigment content) varies with the physiological status of the phytoplankton cells and/or with environmental conditions. Moreover, the pigment-based approach requires to sample seawater and cannot provide continuous measurements of phytoplankton biomass or community composition.

I.4.2. High frequency measurements

As discussed above, phytoplankton community structure varies in time and space. Thus, in order to monitor the dynamics of the many biogeochemical processes that depend on phytoplankton, it has been necessary to develop methods that allow a holistic view of phytoplankton dynamics. For this purpose, non-intrusive methods that allow high frequency observations of the biomass and composition of phytoplankton communities have been developed.

Fluorescence

The most commonly used technic to estimate phytoplankton biomass at high frequency consists of quantifying chlorophyll-a concentration based on in vivo fluorescence. Fluorescence is the emission of energy, in the form of photons, by a molecule that returns from a single excited state to its ground state. The first use of fluorescence as a tool to assess Chla in oceanography has been described by Lorenzen (1966). Chla absorbs light in the blue and red part of the visible light spectrum. This energy can be dissipated for charge separation (photosynthesis), as heat or fluorescence (Huot and Babin, 2010). The source of excitation energy can be either solar light in the case of satellite sensors, or an actinic light of a calibrated wavelength (usually between 430 and 470 nm) in the case of in-situ sensors. The light which is reemitted by fluorescence is redshifted at 685 nm. The fluorescence intensity depends on the excitation energy wavelength, the light absorption coefficient and the quantum yield of fluorescence (*i.e.* the fraction of absorbed light energy that will ultimately be reemitted as fluorescence) of the Chla. It can be formulated as follows:

$$F = [Chla]\, a^*(\lambda)\, \phi\, E(\lambda) \qquad (1.1)$$

where the fluorescence signal, F (mol quanta), is function of the spectral Chla-specific absorption coefficient of phytoplankton at a wavelength λ, a*(λ) (m2 (mg Chla)-1); the quantum yield of fluorescence, Φ (relative unit); and the light energy emitted by the fluorometer at the considered wavelength, E(λ) (mol quanta m^{-2}) (Cosgrove and Borowitzka, 2010). While F is defined as a quantity of energy (i.e. mol quanta), the output of the fluorometer is expressed in digital counts or in relative fluorescence units (RFU). As Chla is ubiquitous among all phytoplankton organisms we can use the fluorescence

intensity as a proxy of the Chla concentration. The transformation of the electric signal received by the sensors to a Chla concentration value can be expressed as follows:

$$[Chla]_{fluo} = (F - Dark\ signal) * Calibration\ slope \quad (1.2)$$

with "Dark signal" as the number of fluorescence relative units in the dark and "Calibration slope" as the factory determined coefficient that permits to convert the number of counts into $[Chla]_{fluo}$ (mg m^{-3} $count^{-1}$). The Calibration slope is derived from a calibration performed by the manufacturer, based on a series of measurements of $[Chla]_{fluo}$ for a range of concentrations of a monospecific culture. However, the relation between Chla and biomass is not straightforward and varies depending on several factor such as photoacclimation, physiological state or taxonomic belonging.

The advantage of *in-situ* fluorescence goes beyond the estimation of Chla concentration and might provide information about the phytoplankton community composition. More recently, the use of *in-situ* multispectral fluorescence to estimate phytoplankton community structure has emerged (Proctor and Roesler, 2010; Thibodeau et al., 2014). This method relies on the specific absorption properties of biomarker pigments that occur in the phytoplankton community. The principle of multispectral fluorescence is to excite different parts of the visible light spectrum, characteristic of key accessory pigments, biomarker of phytoplankton taxa, allowing the identification of phytoplankton community composition. Although the results of recent studies are promising, this method remains experimental and requires further investigation before it can be safely deployed on a large scale.

Optical measurements

Optical measurements of ocean inherent optical properties (IOPs) to retrieve information on particulate composition and assemblage is another opportunity to assess phytoplankton biomass and community structure. Specifically, the measurements of particulate backscattering (b_{bp}) and particulate attenuation (c_p) have been studied to estimate particulate concentration and composition in the ocean (Bricaud and Morel, 1986; Babin et al., 2003; Huot et al., 2007). These measurements have the advantage of not depending on Chla and therefore not fluctuating with the physiological state of the phytoplankton community. However b_{bp} and c_p alone don't provide phytoplankton specific

information, because they reflect the whole particulate assemblage, including zooplankton, detritus, bacteria or mineral particles. The interpretation must then be done in term of particulate organic carbon (POC) (Stramski et al., 1999; Gardner et al., 2006; Cetinić et al., 2012) which encompass phytoplankton biomass but also bacteria and other organic particles. However, the analysis of IOPs in parallel with *in-situ* fluorescence measurements provide an interesting possibility of assessing phytoplankton biomass (Behrenfeld and Boss, 2006; Westberry et al., 2010; Graff et al., 2015) .

A major advantage of fluorescence and optical properties is that they can be easily measured by miniaturized sensors implemented onto *in-situ* autonomous platforms. These platforms have filled a critical observational gap in oceanography by providing data in areas where navigation is not possible because of remoteness or rough meteorological conditions, and by greatly increasing the global coverage of oceanographic measurements. Although there are many types of autonomous platforms, we here focus on BioGeoChemical Argo profiling floats, which are dedicated to the observation of biogeochemical cycles in the global ocean.

I.5. Satellite observation of the phytoplankton community composition

In the last two decades numerous algorithms have been developed to retrieve phytoplankton community composition from space (Bracher et al., 2017). They rely on different optical properties resulting from the variety of pigment composition and cell structure to detect the presence of specific phytoplankton groups (e.g., Le Quéré et al., 2005). It allowed a great improvement in our understanding of the phytoplankton community composition dynamics on the global scale, and the field of application of such methods is still very wide. However, the satellite detection is still restricted to the surface of the ocean and does not provide a satisfying vertical resolution. To address such issue, different *in-situ* autonomous platform appear as an opportunity to complement satellite observations (Figure 1.4). In this thesis work, we will focus on the BioGeoChemical Argo program.

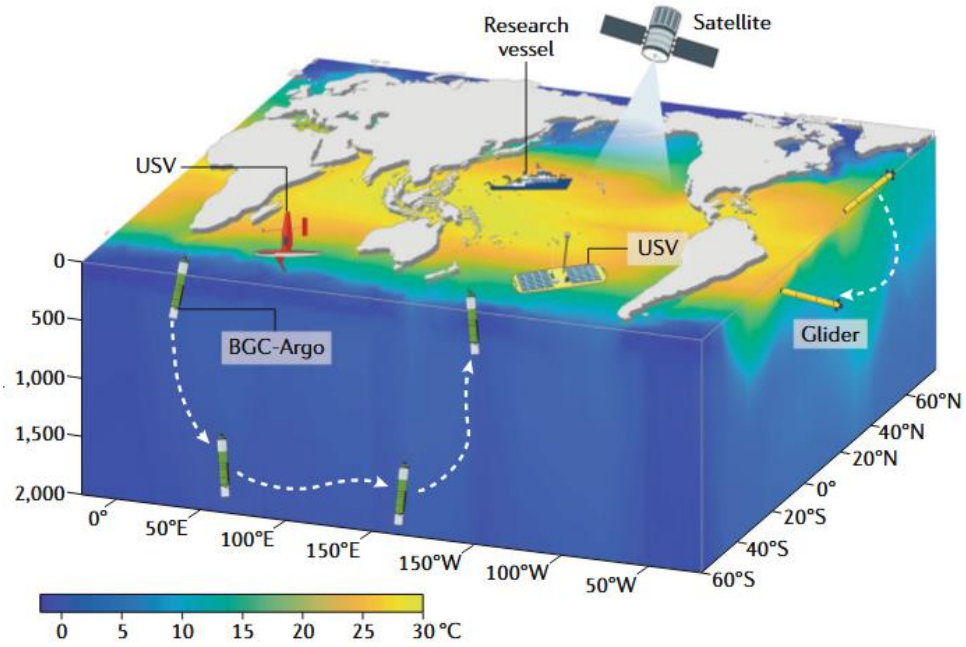

Figure 1-4 : The different oceanographic observation platforms. (Chai et al., 2020)

I.6. BioGeoChemical Argo profiling floats

The BioGeoChemical-Argo (BGC-Argo) program relies on a fleet of profiling floats deployed in the global open ocean (Biogeochemical-Argo Planning Group, 2016). BGC-Argo floats are equipped with a suite of miniaturized sensors that allow to measure the hydrological, optical and biogeochemical properties of the ocean (Claustre et al., 2020). Once deployed, the float performs vertical profiles from 0 to 2000 m depth following a programmed cycle. The float starts its descent to 2000 m depth where it drifts for a defined period of time (usually ~10 days), and then returns to the surface while acquiring measurements. The data are then transmitted by satellites during the short time the float is at the surface. BGC-Argo floats allow the observation of ocean properties throughout the water column and on a quasi-global scale, which make them a unique tool for biogeochemical studies.

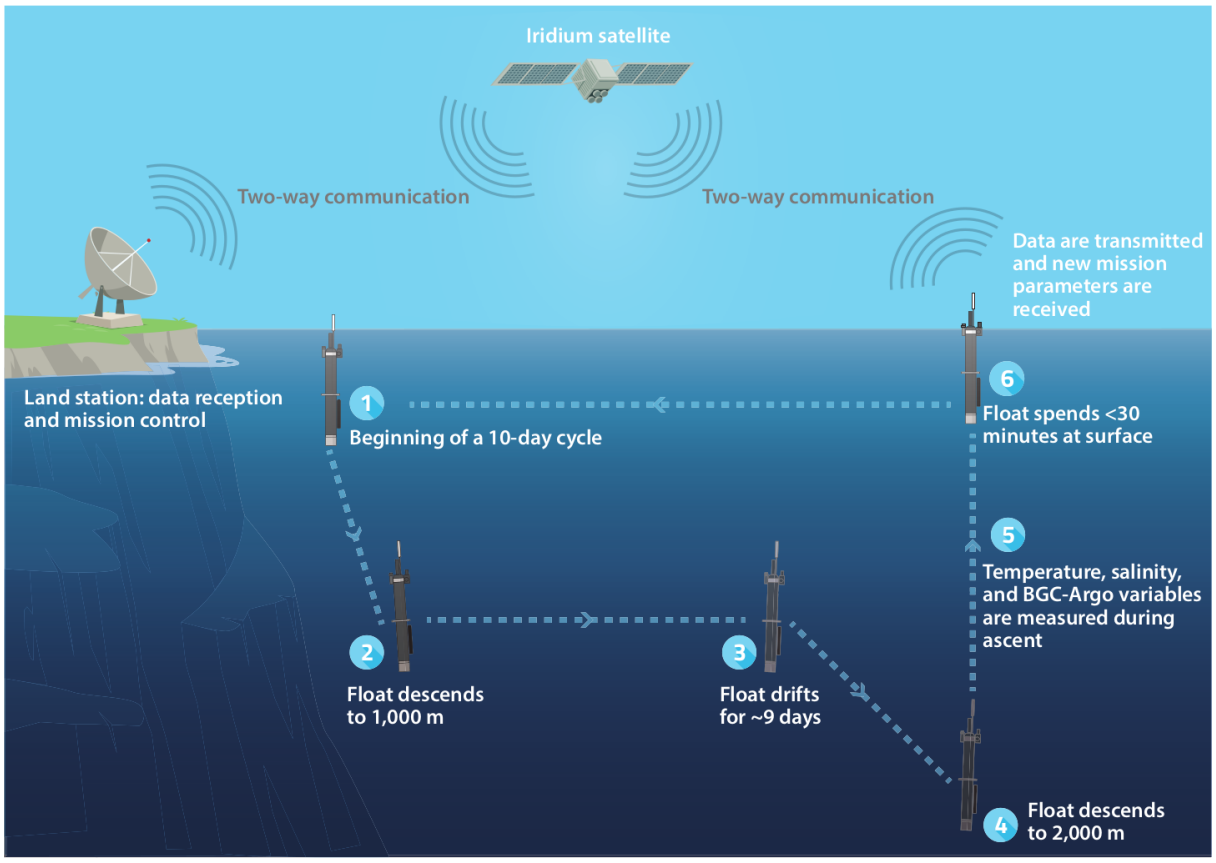

Figure 1 -5 : Example of a 10 days cycle of a BGC-Argo float. Illustration from Claustre et al. (2020).

The acquisition of information related to phytoplankton communities by BGC-Argo floats is currently limited to estimating the Chla concentration from fluorescence, a proxy for algal biomass, and the particulate backscattering and attenuation coefficients, both indicators of POC. Different studies have provided methods to go beyond the biomass information and characterize the structure of phytoplankton communities (e.g. Briggs et al., 2013; Sauzède et al., 2015; Rembauville et al., 2017; Terrats et al., 2020) but they remain limited either in their regions of application, in the information provided, or in their dependence on satellite measurements. Considering the key role of phytoplankton in global biogeochemical cycles, particularly the carbon cycle, it is critical not only to improve the estimation of phytoplankton biomass based on BGC-Argo floats, but also to develop methods that yield information on community composition.

I.7. Research objectives and work plan

The objective of this thesis work are twofold: (1) to improve our understanding of the variability in the fluorescence-to-chlorophyll a ratio in order to better constrain the estimation of Chla concentration from BGC-Argo floats; and (ii) to develop novel approaches to estimate both the biomass and composition of phytoplankton communities in view of applications to BGC-Argo floats.

As discussed above, the estimation of phytoplankton biomass from BGC-Argo floats essentially relies on the assessment of the Chla concentration based on *in-situ* fluorescence measurements. However, it has been shown that the ratio between the fluorescence signal and the Chla concentration varies on the global and regional scales depending (Roesler et al., 2017; Schallenberg et al., 2022). The first goal of the present work is thus to investigate the sources of variability in the relationship between *in-situ* fluorescence and Chla concentration on regional and seasonal scales. We especially focus this analysis on the influence of phytoplankton community composition on the *in-situ* fluorescence signal. This work has been published as a research article in the journal *Frontiers in Marine Science*.

Petit, F., Uitz, J., Schmechtig, C., Dimier, C., Ras, J., Poteau, A., et al. (2022). Influence of the phytoplankton community composition on the in situ fluorescence signal: Implication for an improved estimation of the chlorophyll-a concentration from BioGeoChemical-Argo profiling floats. Frontiers in Marine Science 9 https://doi.org/10.3389/fmars.2022.959131

The second objective of this work is to evaluate the possibility of using *in-situ* multispectral fluorescence to estimate the composition of phytoplankton communities for future applications to BGC-Argo floats. We based our work on the use of a multispectral fluorometer, that emits exciting light at three different wavelengths, and attempt to retrieve information on phytoplankton community (taxonomic) composition based on the fluorescence response. This approach relies on the hypothesis that the fluorescence response to light excitation in the three different channels varies depending on light absorption by the various accessory pigments used as biomarkers of taxonomic groups. This work combined laboratory and field work to evaluate the sensor response to phytoplankton monospecific cultures and *in-situ* natural communities.

Finally, we investigate the possibility to extend the method proposed by Rembauville et al. (2017) for the Southern Ocean to other open ocean regions characterized by different biogeochemical regimes and phytoplankton assemblages. This method enables to retrieve the relative (%) contrbution to the C of four different plankton groups based on optical measurements as acquired by BGC-Argo floats. For this purpose we used a larger dataset that comprises data of three different cruises located in the northwestern and the eastern Mediterranean sea and the Southern Ocean. Ultimately, we improved the method of Rembauville by introducing a method to predict POC concentration from standard BGC-Argo floats measurements. Thus the new proposed method enables to retrieve quantiative information on plankton community composition expressed in terms of POC concentrations.

This thesis is organized in six chapters. The first two chapters are dedicated to present the introduction and an overview of the data and methodological aspects, respectively. Chapters 3 to 5 present the research work. Finally the last chapter discusses conclusions and perspectives.

Although not yet published, Chapters 4 and 5 were written in the format of a scientific article and can be read independently to the rest of the manuscript. This also explains some redundancy in the introduction and methodological sections.

II Data and methods

This thesis heavily relies on data analysis and the development and evaluation of machine learning models. The first objective is achieved by analyzing a database of concurrent *in-situ* fluorescence and Chla concentration measurements. The two other parts of this PhD work focus primarily on the development and evaluation of machine learning models. The challenge behind the development of a machine learning method is to gather concurrent data of the explained variables, called “targets”, and the explanatory variables, called “descriptors”. In order to evaluate the performances of machine learning models applied to BGC-Argo sensor measurements for estimating phytoplankton community composition, we built dedicated database. For this purpose we designed and implemented two different sampling protocols. The general philosophy of the sampling protocol is presented in the following sections and further detailed in the corresponding chapters.

II.1. Fluorescence and HPLC database

The first research objective aims to investigate the regional, seasonal and vertical variability in the relationship between the *in-situ* fluorescence signal and the Chla concentration in the global open ocean. We thus gathered a dataset of concomitant measurements of *in-situ* fluorescence and HPLC-determined Chla concentration. For this purpose, we used the first fluorescence profiles acquired by BGC-Argo floats during their time series in various open ocean regions and HPLC Chla vertical profiles obtained from shipborne seawater sampling coinciding in space and time with the float deployments. A volume of 2.7 L of seawater is collected and filtered through GF/F filters. The filters are then stored in liquid nitrogen for laboratory analysis at the SAPIGH analytical service at IMEV following the protocol of Ras et al. (2008). The resulting database covers different contrasted bioregions considered to be representative of the variability encountered in the global ocean.

In order to account for the seasonal dimension, we also consider data from a fix observation station in the Northwestern Mediterranean Sea. The BOUSSOLE (BOUée pour l'acquiSition d'une Série Optique à Long termE) project collects monthly vertical profiles of *in-situ* fluorescence concomitantly to discrete seawater samples.

In this thesis, we specifically investigate the presence of seven diagnostic pigments (DP) identified among the full suite of accessory pigments as biomarkers of major phytoplankton taxa, further grouped into three size classes (Claustre, 1994; Vidussi et al., 2001). The seven DP are presented in Table 3.2 along with their abbreviations which, for the sake of simplicity, will be used hereafter. The DP-based method allows the estimation of the relative contribution to the Chla concentration of three phytoplankton size classes: micro- (>20 µm), nano- (2-20 µm) and picophytoplankton (<2 µm) following equations given in Uitz et. al (2006) :

$$f_{micro} = \frac{1.41[Fuco] + 1.41[Peri]}{\sum DP_w} \quad (1.3)$$

$$f_{nano} = \frac{1.27[19' - HF] + 0.6[Allo] + 0.35[19' - BF]}{\sum DP_w} \quad (1.4)$$

$$f_{pico} = \frac{0.86[Zea] + 1.01[Tchlb]}{\sum DP_w} \quad (1.5)$$

where $\sum DP_w$ is the weighted sum of the concentrations of the seven diagnostic pigments computed as:

$$\sum DP_w = 1.41\,[Fuco] + 1.41\,[Peri] + 1.27\,[19' - HF] + 0.6\,[Allo] + 0.35\,[19' - BF] + 0.86\,[Zea] + 1.01\,[Tchlb] \quad (1.6)$$

with f_{micro}, f_{nano} and f_{pico} the respective contribution to the total algal biomass of the micro-, nano- and picophytoplankton classes. We note that, because it relies on biomarker pigment concentrations, this approach yields an average, synthetic estimate of both the taxonomic and size composition of the phytoplankton communities. Although we recognize that it has limits because some phytoplankton taxa may occasionally span several size classes and some DP may be found in several taxa, this approach has been

shown to provide reliable, quantitative information for use on large spatial and temporal scales (*e.g.* Bricaud, 2004; Uitz et al., 2006; Brewin et al., 2014).

II.2. Laboratory fluorescence measurements

Our second research objective investigates the possible use of a multispectral fluorometer as a tool to assess phytoplankton community taxonomic composition. The first step of this work is to understand the response of the fluorometer to distinct phytoplankton taxa. To this end, we selected ten phytoplankton strains representative of major phytoplankton taxa found in temperate open ocean regions and used the multispectral fluorometer to evaluate their response in different conditions of concentrations (dilutions). The fluorescence response of each strain was measured on a dilution series in order to obtain a linear regression model between the fluorescence response and the Chla concentration of the strain culture. The slope of the regression model is defined as the Chla biomass calibration value, *i.e.* the fluorescence response increase per units of Chla concentration. Finally we could analyse the variability of the multispectral fluorescence values between different taxa to estimate the possibility to use the multispectrale fluorometer to estimate *in-situ* taxonomic information on the phytoplankton community.

II.3. Field work

A key objective of this thesis work is to design and evaluate the performance of two different models to retrieve information on phytoplankton community composition. The first one (Section IV) is based on multispectral fluorescence while the second (Section V) relies on optical coefficient measurements. In order to collect the data needed to develop these models, we carried out monthly sampling at sea for one year at the BOUSSOLE site, in the Northwestern Mediterranean Sea. The sampling protocol was designed to collect

concomitant measurements of multispectral fluorescence, optical coefficients and phytoplankton community compositon. A CTD-rosette device equipped with a package of dedicated sensors was deployed in order to acquire the fluorescence and optical data concomitantly to seawater samples. For each CTD cast, we measured vertical profiles of fluorescence in response to the three different excitation channels and of the particulate backscattering (bbp) and attenuation (cp) coefficients. Concomitantly seawater samples were collected at 10 different depths for laboratory analysis of HPLC pigments, POC and taxonomic determination (microscopy and flow cytometry).

The Northwestern Mediterranean Sea: A case study

The developed sampling protocol was implemented at the long-term fixed station BOUSSOLE (BOUée pour l'acquiSition d'une Série Optique à Long termE) located in the Ligurian (Northwestern Mediterranean) Sea at 7°54'E, 43°22'N (Antoine et al., 2008). This site is particularly relevant to this study which aims to develop methods for applications the global scale. Indeed, the Northwestern Mediterranean is characterized by a marked seasonality (Marty et al., 2002) where phytoplankton communities encountered throughout the year represent, to some extent, the diversity that can be observed on the global scale. The summer is characterized by stratified waters with a shallow mixed layer, leading to oligotrophic communities dominated by picophytoplankton such as *Synechococcus* or *Prochlorococcus*. Other seasons are characterized by a deeper mixed layer. In spring, the increase of sun light associated with nutrient-replenished upper layer resulting from a deep winter mixed layer depth (MLD) leads to the emergence of a seasonal bloom progressively consuming the surface nutrients and eventually leading to the formation of a deep chlorophyll maximum (DCM) (e.g. Lavigne et al., 2015; Barbieux et al., 2019).

Assessment of the phytoplankton community composition

The phytoplankton community composition was assessed in two different ways, first based on pigment composition, second based on combined flow cytometry and inverted microscopy cell identification and enumeration. First, the pigment-based method consists of HPLC analysis of seawater samples at nine depths (5, 10, 20, 30, 40, 50, 60, 70, and 80

meters). We used the diagnostic pigment approach to investigate the phytoplankton community structure.

The second approach to characterize the phytoplankton community is based on a measurement of the carbon content of planktonic groups. Microphytoplankton organisms were counted and measured by optical microscopy. 500mL of seawater were sampled and fixed with a lugol solution. The samples were then sedimented in a Utermöhl chamber and analyzed by microscopy (Karlson and al, 2010). The bacteria, pico- and nanophytoplankton abundances were determined by flow cytometry analysis. A volume of 1.5 mL of seawater was sampled, fixed with 0.2 mL of glutaraldehyde (0.5% final concentration), flash-frozen in liquid nitrogen, and stored at −80°C until analysis in the laboratory. As this approach is more time consuming than HPLC analysis, we performed these analyses at three depths (*i.e.* 5 m, DCM and below the DCM). The specific depths were selected on board by looking at the shape of the *in-situ* fluorescence profile. In the end, this fieldwork allowed the creation of a database gathering information on the taxonomic and pigment composition of the phytoplankton community associated with corresponding b_{bp} and c_p coefficients and multispectral fluorescence signal.

III Influence of the phytoplankton community composition on the in situ fluorescence signal: Implication for an improved estimation of the chlorophyll a concentration from BioGeoChemical-Argo profiling floats

F. Petit[1], J. Uitz[1], C. Schmechtig[2], C. Dimier[3], J. Ras[1], A. Poteau[1], M. Golbol[3], V. Vellucci[3], H. Claustre[1]

[1]CNRS and Sorbonne Université, Laboratoire d'Océanographie de Villefranche, LOV, 06230 Villefranche-sur-Mer, France

[2] OSU Ecce Terra, UMS 3455, CNRS and Sorbonne Université, Paris 6, 4 place Jussieu, 75252 Paris CEDEX 05, France

[3] Sorbonne Université, CNRS, Institut de la Mer de Villefranche, IMEV, F-06230 Villefranche-sur-Mer, France

III.1. Introduction

The ongoing global changes result in significant changes of oceanic biogeochemical cycles (Bindoff et al., 2019). The drivers of these modifications have to be better characterized and understood, a prerequisite for the predictive modeling of the evolution of the ocean biogeochemistry and its feedback to climate. As phytoplankton play a critical role in oceanic biogeochemical cycles (Falkowski, 1994), mapping their biomass at appropriate spatial and temporal scales appears essential. However, the wide spatial distribution and variable cell sizes (Chisholm, 1992; Roy et al., 2013) of phytoplankton communities make a global assessment of their biomass challenging. Several techniques are used that cover distinct ranges of the size spectrum of phytoplankton organisms, from flow cytometry analysis (Dubelaar and Jonker, 2000) to microscopic observation (Booth, 1993b; Karlson et al., 2010), or High Performance Liquid Chromatography (HPLC) pigment analysis (Jeffrey et al., 1997). Nevertheless, these techniques rely on discrete sampling, which restricts the spatial and temporal coverages of the observations. Satellite-based methods permit to estimate phytoplankton biomass on a global scale but are limited to the ocean surface layer (i.e. the light first penetration depth) and hence do not encompass the vertical distribution of phytoplankton within the whole water column.

Initially introduced to oceanography in 1966 (Lorenzen, 1966), *in-situ* fluorescence relies on the fluorescence property of the chlorophyll-a (Chla) molecule. A fluorometer emits blue light exciting the chlorophyll-a (Chla) molecule and detects the red shifted light that is subsequently reemitted in the environment; the increase in the fluorescence signal is then interpreted as an increase in Chla concentration. Thanks to the possibility to equip with fluorometers autonomous *in-situ* platforms (profiling floats, gliders), *in-situ* fluorescence has become the most widely used technique to assess the phytoplankton biomass at large space and time scales with a fine vertical resolution. In particular, the BioGeoChemical-Argo (BGC-Argo) program aims to monitor and understand key biogeochemical processes on the global scale by developing a network of profiling Argo

floats equipped with a suite of biogeochemical sensors (Roemmich et al 2019; Claustre et al., 2020). BGC-Argo floats are all equipped with an ECO-series Chla fluorometer (SeaBird Electronics), providing time series of vertical profiles of Chla concentration for a broad range of oceanic regimes, resulting in a powerful dataset for investigating the distribution and dynamics of the phytoplankton biomass. Nevertheless, a recent study by Roesler et al. (2017) pointed to the large regional (natural) variability in the relationship between fluorescence and reference Chla concentration measurements, in addition to a global overestimation bias of the Chla concentration by fluorometers of the ECO-series. These results stress out the necessity to better understand the sources of variability in the fluorescence-to-Chla concentration relationship, with an aim to improve the calibration of fluorescence into Chla equivalent for all types of *in-situ* platforms, BGC-Argo profiling floats in particular.

The variability in the ratio between the fluorescence signal and the Chla concentration has been shown to depend not only on the physiological state of phytoplankton cells (Behrenfeld et al., 2009; Escoffier et al., 2015; Schuback et al., 2021; Gorbunov and Falkowski, 2022), but also on the accessory pigment composition that varies with the phytoplankton taxonomic composition (Johnsen and Sakshaug, 2007; Proctor and Roesler, 2010; Roy et al., 2011). Two main factors influence the fluorescence process; the amount of (blue) light absorbed by cells which depends on both the intensity of the light emitted by the sensor and the absorption capacity of the cells (Bricaud et al., 1995, 2004; Roy et al., 2013), and the efficiency of the cells in reemitting part of the absorbed blue light as red light, an efficiency known as the fluorescence quantum yield (Alpine and Cloern, 1985; Olaizola and Yamamoto, 1994; Falkowski and Kolber, 1995).

Despite a significant variability in phytoplankton communities and environmental conditions encountered by the BGC-Argo floats and their attached fluorometers, an identical standard calibration equation is used to convert the electric signal associated to red light emission into Chla concentration (Bittig et al., 2019).

While phytoplankton communities and, hence, pigment composition (e.g. Claustre, 1994; Mackey et al., 1996; Jeffrey et al., 1997; Zapata et al., 2004; Ras et al., 2008) vary

tremendously at the global scale, the effect of such variations on the estimation of Chla concentration based on *in-situ* fluorescence measurements is still poorly understood and quantified. This knowledge is however crucial to better assess phytoplankton biomass in the world ocean and, thus, to improve our understanding of biogeochemical cycles.

In this study, we examine the role of phytoplankton community structure in driving the fluorescence signal through its effects on the phytoplankton absorption coefficient and fluorescence quantum yield. To this end, we analyze data of BGC-Argo *in-situ* chlorophyll fluorescence, HPLC pigment concentrations, and phytoplankton absorption spectra, from various regions of the global open ocean. To investigate further the effect of seasonal succession and vertical distribution of phytoplankton communities on the fluorescence signal, we also consider a time series of *in-situ* fluorescence, HPLC pigments and phytoplankton absorption spectra acquired in the North Western Mediterranean (Ligurian) Sea. This region is characterized by a strong seasonality and contrasted oceanographic and biogeochemical conditions (Marty et al., 2002; Lavigne et al., 2015; Mayot et al., 2017) which will possibly permit to generalize the regional results to other temperate areas. Ultimately, we discuss possibilities to improve the calibration of the in vivo fluorescence signal to Chla concentrations by taking into account information on the phytoplankton community composition.

III.2. Data and Methods

III.2.1. Fluorescence principle

The determination of the Chla concentration, [Chla] (mg m^{-3}); using an *in-situ* fluorometer is based on the fluorescence principle expressed as follows:

$$F = [Chla]\ a^*(\lambda)\ \Phi\ E(\lambda) \qquad (1)$$

where the fluorescence signal, F (mol quanta), depends on the spectral Chla-specific absorption coefficient of phytoplankton at a wavelength λ, a*(λ) (m^2 (mg Chla)$^{-1}$); the quantum yield of fluorescence, Φ (relative unit); and the light energy emitted by the fluorometer at the considered wavelength, E(λ) (mol.quanta m^{-2}) (Cosgrove and Borowitzka, 2010). While F is defined as a quantity of energy (i.e. mol quanta), the output of the fluorometer is expressed in digital counts or in relative fluorescence units (RFU).

BGC-Argo floats are equipped with a Seabird Electronics SBE (previously WET Labs) ECO-series fluorometer. This sensor emits exciting energy on a range of wavelength going from 454 to 480 nm with a peak at 470 nm (Schmechtig et al., 2018). We note that this wavelength does not coincide with the *in vivo* Chla absorption peak in the blue region of the spectrum (Bricaud et al., 2004), but rather excites accessory pigments that transfer the energy to the reaction centers (RC) of the photosynthetic apparatus. The energy reemitted by phytoplankton cells is measured at a nominal wavelength of 690 nm and expressed as counts that correspond to F. This signal is then converted into Chla concentration ($[Chla]_{fluo}$) following Equation (2):

$$[Chla]_{fluo} = (F - \text{Dark signal})\ \text{Calibration slope} \qquad (2)$$

with “Dark signal” the number of fluorescence relative units in the dark and “Calibration slope” the factory determined coefficient that permits to convert the number of counts into $[Chla]_{fluo}$ (mg.m^{-3} count^{-1}). The Calibration slope is derived from a calibration performed by the manufacturer, based on a series of measurements of $[Chla]_{fluo}$ for a range of concentrations of a monospecific culture of the diatom Contricriba *weissfglogii* (also known as Thalassiosira weissflogii)in controlled conditions (SBE ECO Chlorophyll Fluorometer Characterization sheet).

The bias in the estimate of $[Chla]_{fluo}$ is quantified in reference to [Chla] determined from HPLC analysis ($[Chla]_{HPLC}$), assuming that $[Chla]_{HPLC}$ is equal to [Chla] in Equation (1). As in Roesler et al. (2017), the bias introduced in the conversion of F into $[Chla]_{fluo}$ is assessed through the coefficient of a linear regression between concurrent measurements of $[Chla]_{fluo}$ and $[Chla]_{HPLC}$, and is referred to as the "slope factor". Combining and rearranging Equations (1) and (2), the slope factor (dimensionless) can be defined as follows:

$$\text{Slope factor} = ((a^*(\lambda)\ \Phi\ E) - \text{Dark Signal})\ \text{Calibration slope} \quad (3)$$

with two constant values, E and Calibration slope, and two variable quantities, $a^*(\lambda)$ and Φ. The only two parameters that can explain the variability of the slope factor are therefore $a^*(\lambda)$ and Φ.

In order to grasp the variability of these important quantities, we analyzed different datasets that are presented in the next sections. The combination of this different datasets allows to investigate their spatial variability on the global scale as well as their seasonality and depth variability.

III.2.2. Global scale databases

In order to examine the sources of spatial variability of the fluorescence signal on the global scale, we analyze the "Glo-Argo" database. This database comprises concurrent measurements of $[Chla]_{fluo}$, $[Chla]_{HPLC}$, and concentrations of accessory pigments, along with indirect determinations of the $a^*(470)$, which is the phytoplankton absorption coefficient at 470 nm, the excitation wavelength of the ECO fluorometers mounted on the BGC-Argo floats. The $[Chla]_{fluo}$ data arise from BGC-Argo float fluorescence measurements, and the HPLC-determined $[Chla]_{HPLC}$ and accessory pigments

concentrations from discrete seawater samples collected simultaneously to the BGC-Argo float deployments. The phytoplankton absorption coefficient was generally not measured at float deployment and was thus derived from $[Chla]_{HPLC}$ using the empirical relationship of Bricaud et al. (1995). This relationship is based on a global scale database of contrasted environment and has proven robust for application to large scale dataset and does not show substantial regional varibility – see Figure 3.1 in Bricaud et al. (1995).

Overall, the Glo-Argo dataset includes *in*-situ measurements associated with 46 BGC-Argo floats deployed in 8 oceanic provinces defined by Longhurst (2006), also called "bioregions". We will further refer to these bioregions by the following abbreviations (codes): ANTA, Antarctic Province; ARCH, Archipelagic Deep Basin Province; ARCT, Atlantic arctic Province; BPLR, Boreal Polar Province; EMED, Eastern Mediterranean Sea; SANT, Subantarctic Province; SPSG, South Pacific Subtropical Gyre Province; and WMED, Western Mediterranean Sea. They range from polar to subtropical biomes and are thus representative of most of the environmental conditions observed in the global ocean (Table 3.1). From the entire BGC-Argo fleet, we selected the floats that have been deployed with concomitant high-resolution vertical sampling for HPLC analysis. The float data were extracted from the official Coriolis database (ftp://ftp.ifremer.fr/ifremer/argo). The first measurement profile of each float has been matched with its associated HPLC profile at deployment. A maximum delay of 24h between the float fluorescence profile and the seawater sampling was set as a condition for the matchup the final matchups have a time lag of 9 h ± 6 h (mean ± standard deviation).

All of the BGC-Argo floats are equipped with identical SBE ECO Puck Triplet fluorometers, to which the same standard calibration procedure is applied (Schmechtig et al., 2018). This ensures that the observed variability in the fluorescence signal can be attributed to natural variability rather than to inter-sensor variability. The $[Chla]_{fluo}$ is calculated from F, expressed as counts, a relative unit, following the manufacturer calibration coefficient (cf. Equation 2). The data are corrected for non-photochemical quenching (NPQ) following the method of Xing et al. (2012). In brief, this method

extrapolates the deep fluorescence value toward surface. It has been validated using data from contrasted conditions in the Southern Ocean and Mediterranean Sea and been successfully used in various open oceans (Barbieux et al., 2018; Mignot et al., 2018; Taillandier et al., 2018). The $[Chla]_{fluo}$ values greater than four times the Cook distance of the regression between $[Chla]_{fluo}$ and $[Chla]_{HPLC}$ in each bioregion are identified as outliers and removed. Roesler et al. (2017) recommends a factor of 2 to be applied to the BGC-Argo $[Chla]_{fluo}$ data so as to account for the mean overestimation of [Chla] by the ECO-series fluorometers on the global scale. Nevertheless, the correction factor of 2 is not applied here as our goal is to compare regional values of the slope factor computed for the Glo-Argo database with those of Roesler et al. (2017). Finally, the euphotic depth (Z_{eu}) was calculated from the $[Chla]_{FLUO}$ vertical profile adjusted with $[Chla]_{HPLC}$ following Morel and Maritorena (2001), which allows to compute Zeu for all vertical profiles, including the 13 night profiles; the data below Zeu were discarded. The final dataset comprises 335 samples distributed over 46 different profiles in 8 oceanic bioregions (Table 3.1).

Our global-scale investigation is completed by the analysis of the "Glo-aphy" dataset, which comprises concurrent HPLC pigment determinations and on-filter measurements of the phytoplankton absorption spectrum for seawater samples collected over 13 cruises from 1990 to 2016, most of them being part of the databases presented in Bricaud et al. (1995, 2004, 2010). This dataset includes 3340 *in-situ* data spanning 14 different bioregions that cover contrasted ocean environments from polar to subtropical biomes and, thus, may be considered as representative of the global open ocean. As for the Glo-Argo dataset, Zeu was computed following Morel and Maritorena (2001) and the data located below Zeu were discarded.

Table 3.1. Summary of the three different datasets used in the present study, the name of the sampled bioregions with the correspond abbreviations, the available measurements and the number of samples in each dataset. The bioregions correspond to the oceanic provinces defined by Longhurst (2007).

Name of the dataset	Bioregions (code)	Available measurements	Number of samples
Glo-Argo	Antarctic (ANTA) Archipelagic Deep Basins (ARCH) Atlantic Arctic (ARCT) Boreal Polar (BPLR) Eastern Mediterranean (EMED) Subantarctic (SANT) S. Pacific Subtropical (SPSG) Western Mediterranean (WMED)	. Fluorescence (SBE ECO-series; 454 - 480-nm excitation channel) . HPLC pigments . Chla-based absorption	335
Glo-aphy	Antarctic (ANTA) Indian S. Subtropical Gyre (ISSG) Mediterranean (MEDI) N. Atlantic Drift (NADR) N. Atlantic Subtropical Gyre (NASE) N. Atlantic Tropical Gyre (NATR) Pacific Equatorial Divergence (PEQD) Subantarctic (SANT) S. Pacific Subtropical (SPSG) S. Subtropical Convergence (SSTC) W. Pacific Warm Pool (WARM)	. HPLC pigments . On-filter absorption	3340
Med-Bouss	Mediterranean (MED)	. Fluorescence (Chelsea Mini Aquatracka; 378 - 483-nm excitation channel) . HPLC pigments . On-filter absorption	843

III.2.3. North Western Mediterranean Sea fluorescence database

In order to understand the seasonal and vertical variability of the fluorescence signal, we completed our analysis using the "Med-Bouss" database, a time series acquired at the long-term fixed station BOUSSOLE (BOUée pour l'acquiSition d'une Série Optique à Long termE) located in the Ligurian (Northwestern Mediterranean) Sea at 7°54'E, 43°22'N (Antoine et al., 2008). Seawater sampling was carried out at the BOUSSOLE station every month from 2013 to 2015, at seven discrete depths (5, 10, 20, 30, 40, 50 and 60 m) using a CTD-rosette system equipped with 12-L Niskin bottles (Golbol et al., 2000). A Chelsea Aquatracka III fluorometer was mounted on a CTD-rosette device and acquired fluorescence profiles concomitantly with seawater sampling. The Aquatracka fluorometer has an excitation waveband ranging from 378 to 483 nm with a peak at 430 nm, and an emission waveband ranging from 670 to 700 centered on 685 nm. The fluorometer calibration did not change over the whole time series. The fluorescence output is expressed as RFU. The total dataset comprises 843 samples for which are available concurrent data of $[Chla]_{fluo}$, $[Chla]_{HPLC}$, HPLC-determined accessory pigments, and on-filter phytoplankton absorption spectra.

III.2.4. Phytoplankton absorption

For both the Glo-aphy and Med-Bouss databases, the spectral phytoplankton absorption coefficient, $a_{ph}(\lambda)$ (m^{-1}), was measured following the same analytical protocol and using the same filters as those used afterwards for HPLC analyses (Bricaud et al. 1995, 2004 and 2010). In brief, absorption measurements were performed with a Perkin Elmer 850 spectrophotometer, equipped with a 150 mm diameter integrating sphere, using a blank filter as reference. Spectra were shifted to zero in the near infrared by subtracting the average optical density between 750 and 800 nm (Röttgers and Gehnke, 2012). Optical densities were then corrected for the amplification effect (Bricaud and Stramski, 1990) and converted into particulate absorption (in m^{-1}) (Allali et al., 1997). The contribution of phytoplankton to the particulate absorption was then determined following the numerical

decomposition of Bricaud et al. (2010). In this study we focus on the specific absorption wavelength of 470 nm, which corresponds to the excitation wavelength of the SBE ECO-series fluorometer mounted on the BGC-Argo floats. We recall that phytoplankton absorption was not systematically available for the Glo-Argo dataset and that a_{ph} was derived from $[Chla]_{HPLC}$ following the general relationship of Bricaud et al. (1995) regardless of the bioregion.

III.2.5. Phytoplankton pigments and community composition

For the global (Glo-Argo and Glo-aphy) and Mediterranean (Med-Bouss) datasets, the Chla concentration and the composition of phytoplankton communities are estimated using HPLC pigment measurements. In brief, seawater from discrete sampling is filtered onto glass fiber filters (GF/F Whatman 25 mm), that are stored in liquid nitrogen during cruises then transferred at −80°C in the laboratory until further analysis at the SAPIGH HPLC analytical facility at the Institut de la Mer de Villefranche (IMEV). Phytoplankton pigments are extracted from the cells by sonication in 100% methanol, clarified by filtration (GF/F Whatman 0.7 µm), and finally separated and quantified by HPLC. More details about the HPLC analytical protocol may be found in Ras et al. (2008). The concentration of total chlorophyll-a (TChla) is defined as the sum of Chla, divinyl-chlorophyll-a and chlorophyllid-a concentrations. In this thesis, we will refer to [Tchla] as [Chla]. We specifically investigate the distribution of seven diagnostic pigments (DP) identified, among the full suite of accessory pigments, as biomarkers of major phytoplankton taxa, further grouped into three size classes (Claustre, 1994; Vidussi et al., 2001). The seven DP are presented in Table 3.2 along with their abbreviations which, for the sake of simplicity, will be used hereafter.

The DP-based method allows the estimation of the relative contribution to the [Chla] of three phytoplankton size classes : micro- (>20 µm), nano- (2-20 µm) and pico-phytoplankton (<2 µm) following equations given in Uitz et al. (2006). We note that,

because it relies on biomarker pigment concentrations, this approach yields an average, synthetic estimate of both the taxonomic and size composition of the phytoplankton communities. Although we recognize that it has limits because some phytoplankton taxa may occasionally span several size classes and some DP may be found in several taxa, this approach has been shown to provide reliable, quantitative information for use on large spatial and temporal scales (e.g. Bricaud et al., 2004; Uitz et al., 2006; Brewin et al., 2014).

Table 3.2. Major biomarker pigments used in this study with their abbreviation, taxonomic significance and associated size class (Vidussi et al. 2001). Pico stands for picophytoplankton (0.2-2 µm), nano for nanophytoplankton (2-20 µm) and Micro for microphytoplankton (20-200 µm).

Diagnostic Pigments	Abbreviations	Taxonomic significance	Size class
Zeaxanthin	Zea	Cyanobacteria	Pico
Chlorophyll b+Divinyl-chlorophyll b	Tchlb	Green Flagellates and Prochlorophytes	Pico
19'hexanoyloxyfucoxanthin	19'-HF	Prymnesiophytes	Nano
19'butanoyloxyfucoxanthin	19'-BF	Pelagophytes	Nano
Alloxanthin	Allo	Cryptophytes	Nano
Fucoxanthin	Fuco	Diatoms	Micro
Peridinin	Peri	Dinoflagellates	Micro

III.2.6. Data analysis

We determined the three variables of interest, i.e. slope factor, Chla-specific absorption at 470 nm (a*(470)) and quantum yield of fluorescence (Φ), using a type I linear regression model applied at the regional scale for the Glo-Argo and Glo-aphy datasets and at the seasonal and vertical scale for the Med-Bouss dataset.

The slope factor is computed as a regression of type I between $[Chla]_{fluo}$ and $[Chla]_{HPLC}$. A type I regression is selected in regard of the robustness of the HPLC estimation of [Chla], considered as the reference (Claustre et al., 2004). The intercept is set to 0 because, after subtraction of the dark signal, an absence of Chla in the environment should be associated with a null fluorescence signal. Thus, the intercept correspond the dark calibration, which improves the robustness of the statistics.

The Chla specific absorption coefficient of phytoplankton at 470 nm, a*(470), is calculated as the coefficient of the regression between the phytoplanktonic absorption at 470 nm, $a_{ph}(470)$ (m^{-1}), and $[Chla]_{HPLC}$ (mg m^{-3}). The retrieved a*(470) coefficient is then expressed in m^{-2} mg $Chla^{-1}$.

The fluorescence quantum yield, Φ, is calculated as the coefficient of the regression between F and aph(λ) (m^{-1}), λ being the wavelength of the excitation peak of the considered fluorometer. For the Glo-Argo dataset, F is expressed as counts and λ equals 470 nm; and for Med-Bouss dataset, F is expressed as RFU and λ equals 430 nm. The obtained Φ is then expressed either as counts.m^{-1} or RFU.m^{-1} for the Glo-Argo or Med-Bouss dataset, respectively. It represents the raw output value of the fluorometer subsequently to light absorption by phytoplankton cells.

Each regression is performed on the bioregion scale for the Glo-Argo and Glo-aphy dataset. For the Med-Bouss dataset, we merged the data acquired during the three consecutive years in order to optimize the number of data per regression and thus the robustness of the statistics. Statistics from the regressions are provided in the supplementary material.

A principal component analysis (PCA), performed with the FactoMineR package version 2.4 (Lê et al., 2008), is used to investigate the succession of phytoplankton community at the Mediterranean Sea site (Med-Bouss dataset). All analyses were performed with the R software, version 4.1.2.

III.3. Results

III.3.1. Global variability of the slope factor

In order to investigate the influence of phytoplankton community composition on the estimation of $[Chla]_{fluo}$, concomitant measurements of $[Chla]_{fluo}$ from BGC-Argo floats and $[Chla]_{HPLC}$ are merged into a global-scale dataset (Glo-Argo). This dataset covers contrasted bioregions from high to low latitudes (Figure 3.1a; Table 3.1). Thus, the variability of the slope factor in this dataset may be considered as representative of its variability on the global scale. The mean value of the slope factor is 2.2, which indicates that the SBE ECO fluorometer overestimates [Chla] on the global scale by a factor 2.2. Additionally, the slope factor shows different values depending on the bioregions (Figure 3.1b). There is a large variability between high and low latitude environments, with a maximum value of 2.8 observed in the Boreal Polar Province (BPLR) in the North Atlantic Ocean, and a minimum value of 1.1 in the Archipelagic Deep Basin Province (ARCH) in the subtropical Pacific Ocean. In the Mediterranean Sea, the slope factor ranges from 1.7 to 1.9 in the eastern (EMED) and western (WMED) basins, respectively. Our results are consistent with those of Roesler et al. (2017), who reported a mean global value of 2, and a regional variability characterized by higher values in polar regions and lower values in subtropical regions.

We now consider the distribution of the slope factor on a ternary diagram representing the pigment-based composition of phytoplankton communities in the Glo-

Argo dataset (Figure 3.2). The data follow a gradient from microphytoplankton-dominated communities typical of polar and sub-polar regions, to picophytoplankton-dominated communities in the subtropical regions. In between, there are mixed and nanophytoplankton-dominated communities associated with the temperate waters of the Mediterranean Sea. The regression between the values of the slope factor and the Chla relative contribution of the three phytoplankton size classes is displayed as the background of the ternary diagram. It shows a clear pattern where the communities dominated by larger phytoplankton groups are associated with high values of the slope factor. Thus, our analysis indicates that the observed regional patterns of the slope factor (Figure 3.1 and Roesler et al., 2017) are consistent with the patterns associated with the composition of the phytoplankton communities (Figure 3.2). This suggests that the regional variability in the slope factor may in fact be influenced by phytoplankton community composition.

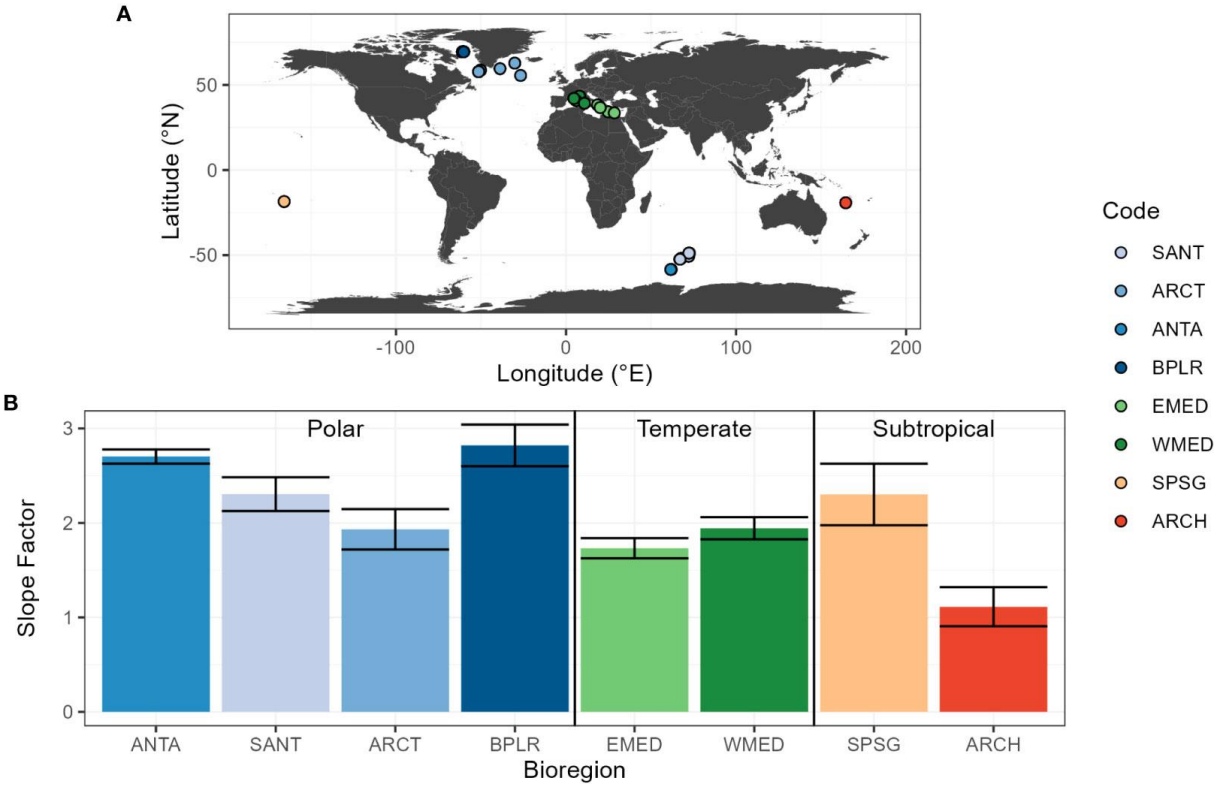

Figure 3.1: (A) Geographic location of the sampling stations in the Glo-Argo database and (B) mean value of the slope factor, defined as the coefficient of the regression between [Chla]Fluo and [Chla]HPLC, for each bioregion of the Glo-Argo database (ANTA, Antarctic Province; ARCH, Archipelagic Deep Basin Province; ARCT, Atlantic arctic Province; BPLR, Boreal Polar Province; EMED, Eastern Mediterranean Sea; SANT, Subantarctic Province; SPSG, South Pacific Subtropical Gyre Province; WMED, Western Mediterranean Sea).

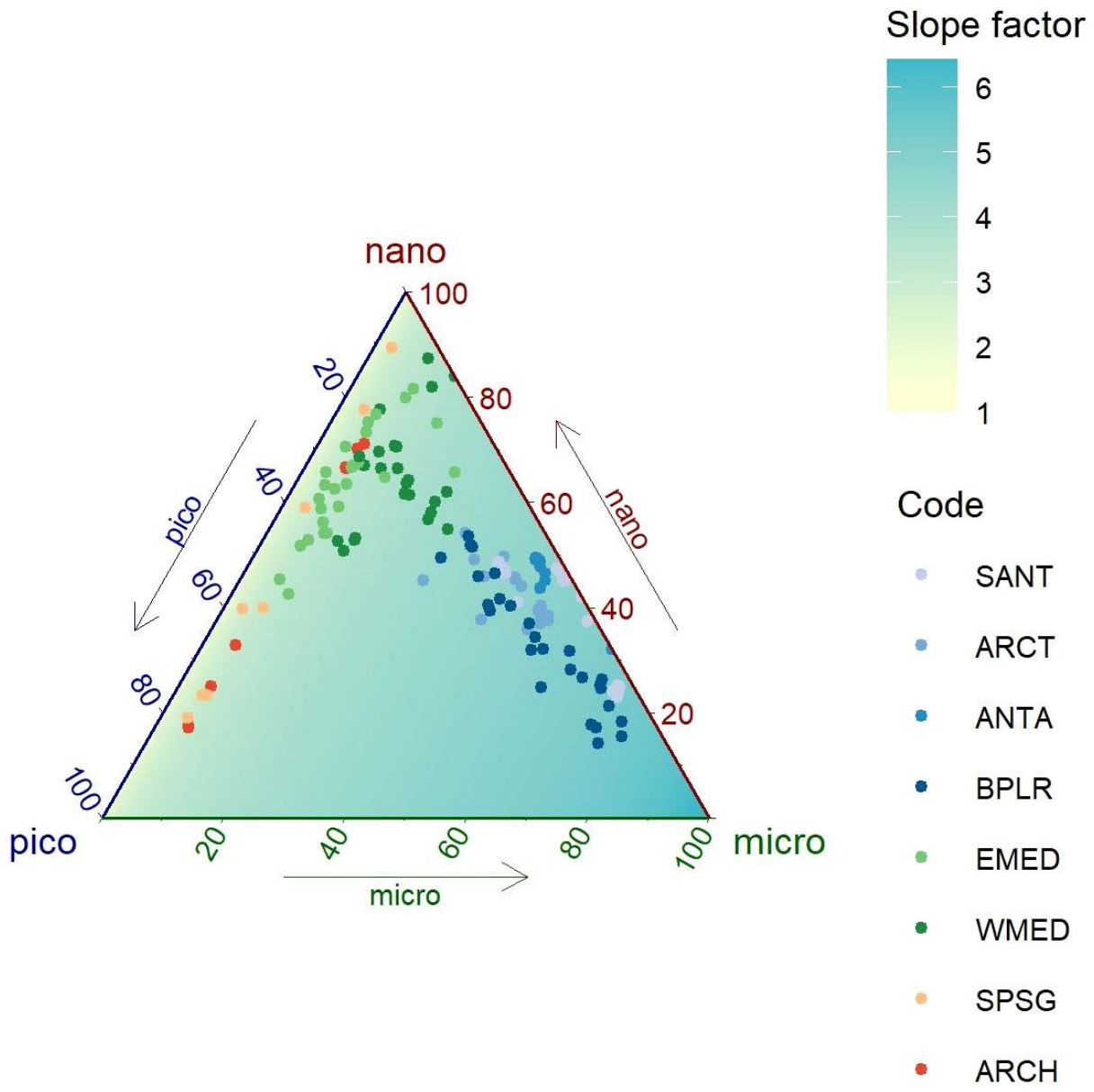

Figure 3.2: Ternary diagram of the pigment-based relative contribution to the Chla concentration of three phytoplankton size classes (micro-, nano, and picophytoplankton) in the global Glo-Argo dataset. The color of the background indicates the value of the [Chla]fluo / [Chla]HPLC ratio extrapolated with a linear regression model. The color of the datapoints indicates the corresponding bioregions (ANTA, Antarctic Province; ARCH, Archipelagic Deep Basin Province; ARCT, Atlantic arctic Province; BPLR, Boreal Polar Province; EMED, Eastern Mediterranean Sea; SANT, Subantarctic Province; SPSG, South Pacific Subtropical Gyre Province; WMED, Western Mediterranean Sea).

In order to understand the mechanisms underpinning the correlation between the slope factor and phytoplankton community composition, we examine the variability of the two main photophysiological properties influencing the fluorescence signal, i.e. the Chla-specific absorption coefficient and the fluorescence quantum yield.

III.3.2. Variability of the phytoplankton absorption coefficient

The influence of the phytoplankton community composition on the phytoplankton Chla-specific absorption coefficient is first investigated using the Glo-aphy dataset. Similarly to the Glo-Argo database, Glo-aphy encompasses a wide variety of oceanic regimes, including polar, temperate and subtropical bioregions (Figure 3.3a; Table 3.1). The value of a*(470) follows a regional distribution (Figure 3.3b) and ranges from 0.024 to 0.071 m^2 (mg Chla)$^{-1}$, with a magnitude of variation of 2.95. On the one hand, the North Atlantic Subtropical gyre (NASE) and North Atlantic Drift (NADR) bioregions display the highest a*(470) coefficient of the dataset with values of ~0.07 m^2 (mg Chla)$^{-1}$. These bioregions are characterized by oligotrophic waters, dominated by picophytoplankton with large concentrations of zeaxanthin. On the other hand, polar bioregions like SANT and ANTA, dominated by larger cells, show the lowest a*(470) values of ~0.02 m^2 (mg Chla)$^{-1}$.

In addition to the global-scale analysis, the vertical and seasonal variability of a*(470) induced by changes in the phytoplankton communities is investigated using the Mediterranean (Med-Bouss) dataset. This dataset covers three successive annual cycles and seven different depths, and presents contrasted environmental conditions due to the pronounced seasonality of this bioregion (Marty et al., 2002; Durrieu de Madron et al., 2011) with $[Chla]_{HPLC}$ values ranging from ~0.03 to 2.50 mg m^{-3}. In the next paragraphs, we first describe the seasonal and vertical distribution of phytoplankton communities

based on a Principle Component Analysis (PCA) and then present the distribution of the a*(470) coefficient.

A PCA projection of the pigment data from the Med-Bouss dataset is performed to visualize the distribution of the different phytoplankton communities over the seasonal cycle and the vertical dimension (Figure 3.4). On the first principal component (PC1), we observe a discrimination of the different sampling depths, indicating that PC1 represents the vertical distribution of the

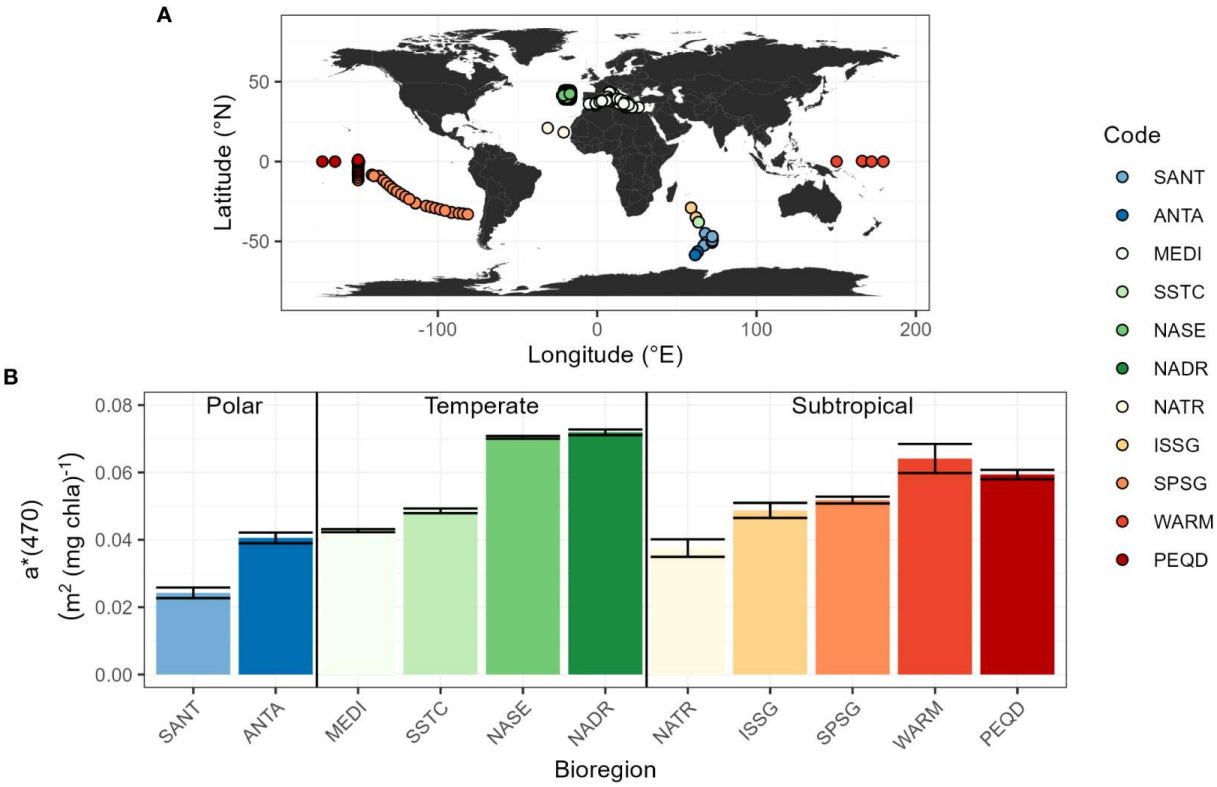

Figure 3.3: Distribution of the Chla-specific absorption coefficient at 470 nm, a(470) (m2(mg Chla)-1), calculated from a linear regression between aph(470) and the Chla concentration for each bioregion of the global Glo-aphy dataset (see Table 3.1 for code significance). The error bars indicate the 95% confidence limits.*

phytoplankton communities. On the second principal component (PC2), we observe a seasonal discrimination, spring and winter in the lower part of the axis and summer and fall in the upper part. Indeed, summer is characterized by stratified waters with a shallow mixed layer, whereas other seasons are characterized by a deeper mixed layer. In spring, the increase of sun light associated with nutrient-replenished upper layer resulting from a deep winter mixed layer depth (MLD) leads to the emergence of a seasonal bloom, progressively consuming the surface nutrients and eventually leading to the formation of a deep chlorophyll maximum (DCM) (Lavigne et al., 2015; Barbieux et al., 2019b). In summer and fall, surface waters are typically dominated by zeaxanthin, indicative of the presence of pico-prokaryotes (Sammartino et al., 2015; Trombetta et al., 2020) whereas the concentration of 19'-BF and 19'-HF increase in deeper waters, indicating an enhanced contribution of nanophytoplankton cells such as Prymnesiophytes and Chrysophytes (Bustillos-Guzmán et al., 1995) to the phytoplankton assemblage at depth. In winter, we observe an increased concentration of fucoxanthin, mostly indicative of the presence of diatoms (generally microphytoplankton). Those observations are consistent with the phytoplankton succession described by Marty et al. (2002), which gives us good confidence that the PCA representation efficiently grasps the vertical as well as seasonal variations of phytoplankton communities.

In the Med-Bouss dataset, the a*(470) coefficient ranges from 0.017 to 0.068 m^2 $(mg\ Chla)^{-1}$ (Figure 3.5), which is coherent with observations on the global scale in comparable regimes (Figure 3.3). It shows a relatively low vertical variability in winter and fall, when the water column is mixed (e.g. D'Ortenzio et al., 2005), with values comprised between 0.03 and 0.04 m^2 $(mg\ Chla)^{-1}$ in winter, and 0.04 to 0.05 m^2 $(mg\ Chla)^{-1}$ in fall (Figure 3.5). The a*(470) coefficient reaches a maximum of 0.07 m^2 $(mg\ Chla)^{-1}$ in summer at surface, in picophytoplankton dominated communities, and slowly decreases from fall to winter when phytoplankton communities are dominated by larger cells. Lower values are observed in winter and below 20 m in spring and summer where it decreases to less than 0.04 m^2 $(mg\ Chla)^{-1}$. This is consistent with our observations on the global-scale (Glo-

aphy) database where higher values of a*(470) are encountered in picophytoplankton-dominated communities.

In order to determine the influence of phytoplankton pigments and, thereby, community composition on phytoplankton absorption, we consider the distribution of a*(470) in the orthogonal projection of PC1 and PC2 of the pigment based PCA (Figure 3.5). We observe, in particular, that the increase in a*(470) in the surface layer during the summer season coincides with the occurrence of picophytoplankton-dominated communities, with high concentrations of zeaxanthin. The highest values of a*(470) (>0.04 m^2 (mg Chla)$^{-1}$) are exclusively observed in such communities, while moderate to low values of a*(470) are distributed amongst nano- and micro-phytoplankton-dominated communities with no clear pattern in terms of pigments composition.

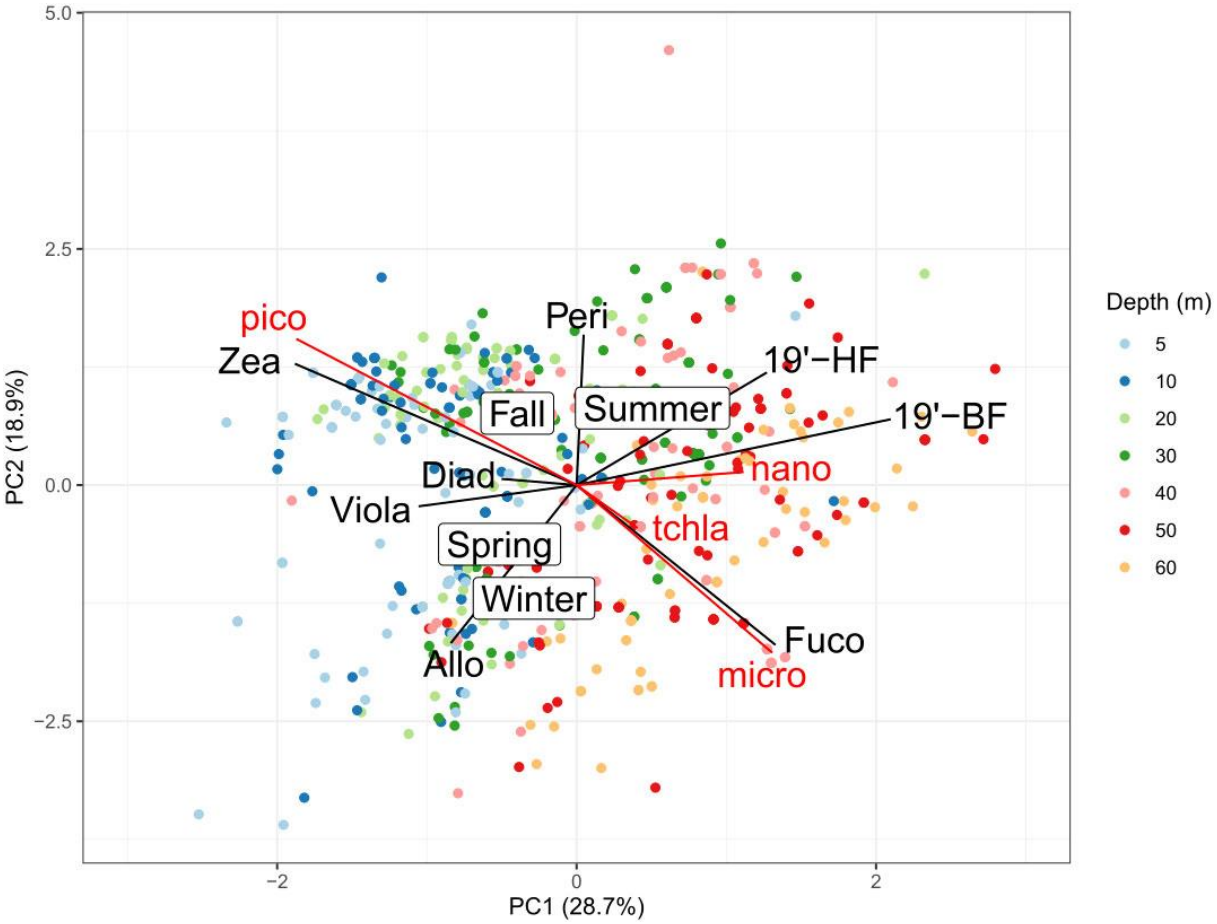

Figure 3.4: Orthogonal projection of the two first principal components of the PCA analysis of the concentrations of the seven diagnostic pigments in the Northwestern Mediterranean Med-Bouss database, between 5 and 60 m depth over 3 years (2013–2015). The color of each point represents the sampling depth as indicated. The relative contribution to the Chla concentration of each of the three phytoplankton size classes and the seasons are superimposed as supplementary variables onto the PCA projection space.

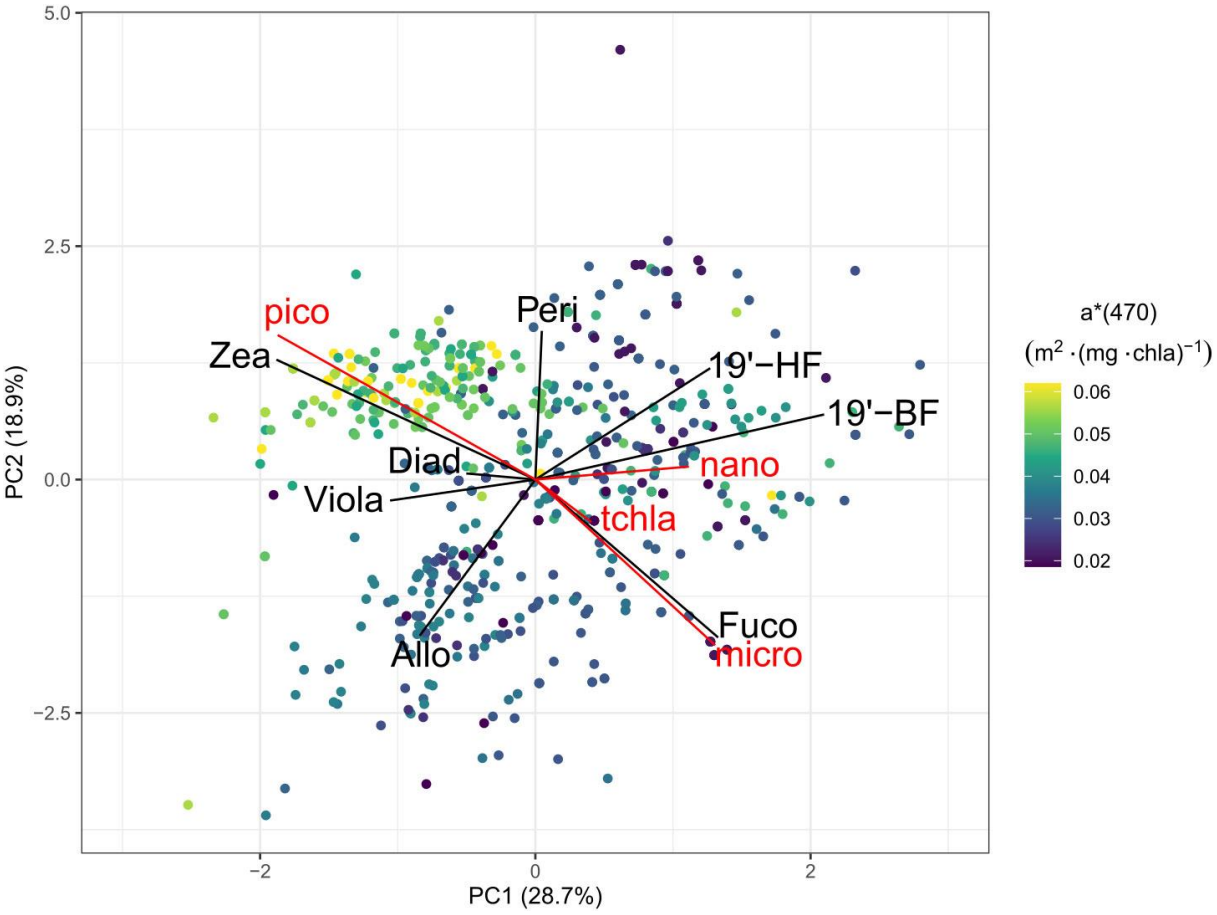

Figure 3.5: Orthogonal projection of the two first principal components of the PCA analysis of the concentrations of the seven diagnostic pigments in Northwestern Mediterranean Med-Bouss dataset, between 5 and 60 m depth over 3 years (2013–2015). The color of each point represents the Chla-specific absorption coefficient at 470 nm, a(470), which did not serve in the construction of the projection. The relative contribution to Chla of each of the three phytoplankton size classes and the seasons are superimposed as supplementary variables onto the PCA projection space.*

Our results indicate a significant variability of the a*(470) coefficient, not only on the global scale among the considered bioregions (Figure 3.3b), but also within a given bioregion, vertically within the water column as well as seasonally (Figures 3.5-3.6). Importantly, the global-scale variability in the a*(470) coefficient (Figure 3.3) does not follow the same trend as the slope factor (Figure 3.1), which shows higher values in high latitude regions and lower values in low latitude regions. Therefore, the a*(470) coefficient is unlikely to be the only driver of the variability of the ratio between F and [Chla], i.e. the slope factor. This led us to examine the scales of variability of the fluorescence quantum yield.

III.3.3 Variability of the fluorescence quantum yield

We investigate the role of the second potential driver of the slope factor, the quantum yield of fluorescence, using the Med-Bouss and Glo-Argo datasets. In the Mediterranean dataset, the fluorescence quantum yield appears to vary between 5 and 12 RFU.m^{-1} depending on the season and depth (Figure 3.7). The maximum values are observed near the DCM in spring (12 RFU.m^{-1}) and at depth in winter (10 RFU.m^{-1}), whereas the minimum values are observed below 20 m in summer (~5 RFU.m^{-1}). The summer season is characterized by a relatively low vertical variability. In contrast, a strong vertical pattern is found in spring, when the fluorescence quantum yield shows significantly higher values at 30-40 m than at surface (5 m). We also note that the fluorescence quantum yield shows larger values in spring than in summer. In fall, the standard deviation at most depths is larger than the vertical and seasonal variability.

On a global scale, the regional variations determined from the Glo-Argo dataset for the fluorescence quantum yield are presented in Figure 3.8. The fluorescence quantum yield varies between 6800 counts.m^{-1} and 13900 counts.m^{-1} for the subtropical ARCH and polar BPLR bioregions, respectively. The maximum values are found in high

latitude regions in contrast to subtropical regions characterized by lower values of the fluorescence quantum yield. Temperate regions have a similar value of Φ as subtropical regions with around 1000 counts m^{-1}.

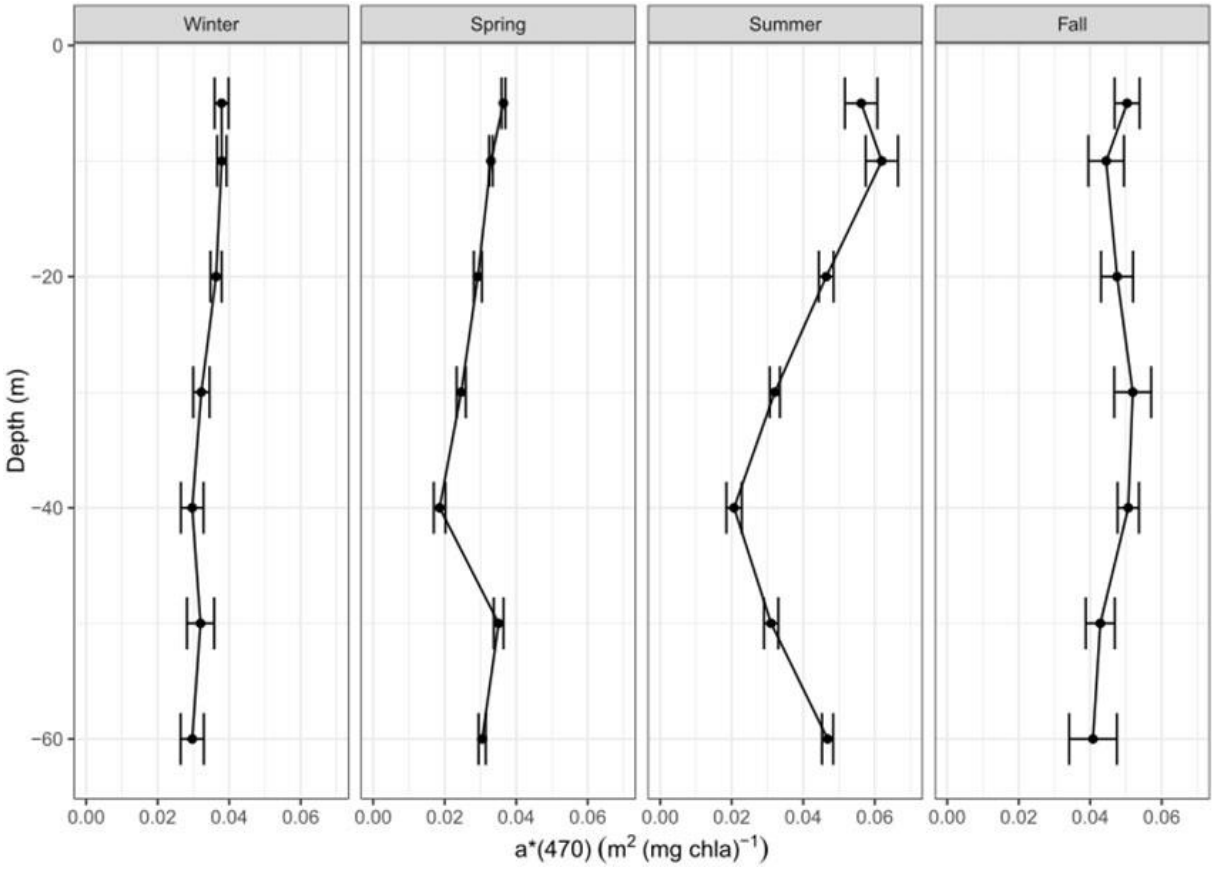

Figure 3.6: Vertical distribution of the phytoplankton chlorophyll-specific absorption coefficient at 470 nm, a(470) in m2 (mg Chla)-1, for each season and sampling depth of the Med-Bouss dataset.*

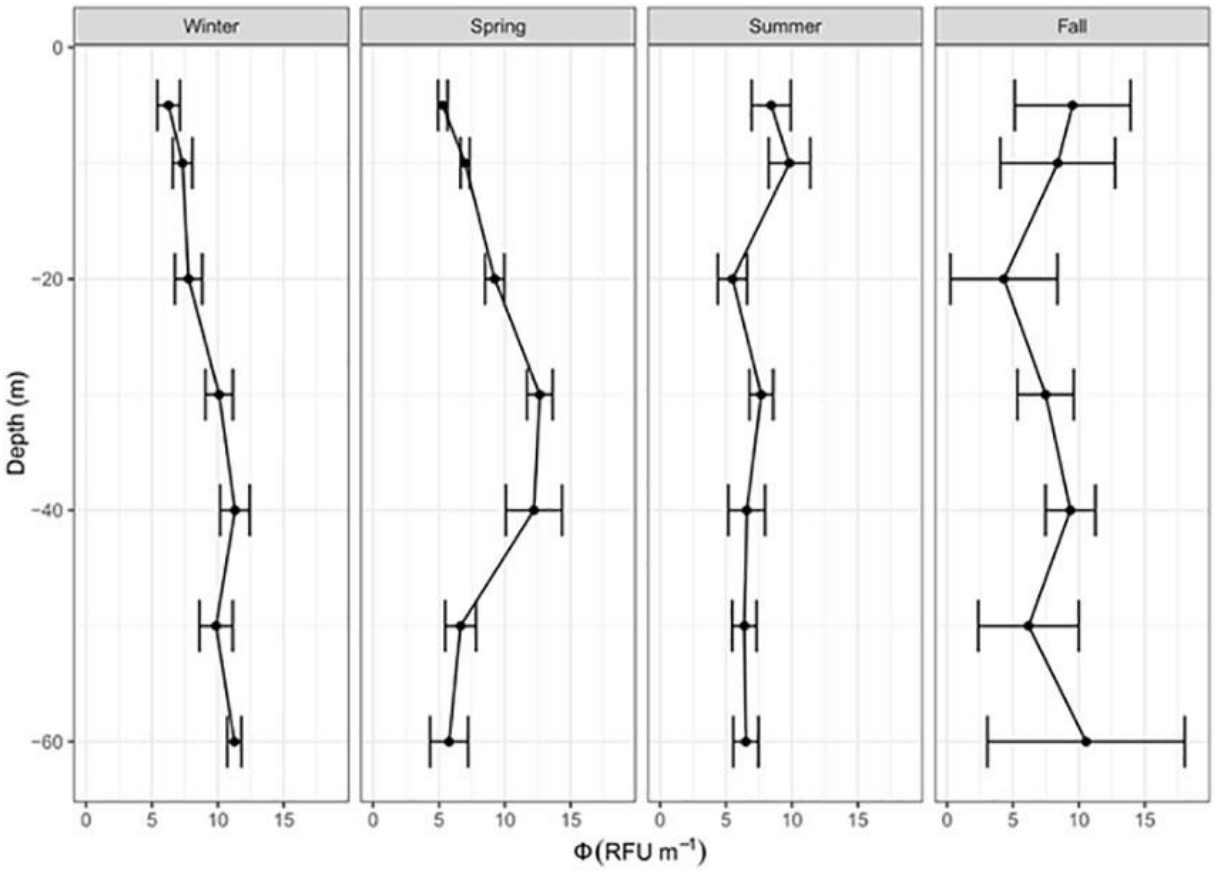

Figure 3.7: Vertical distribution of the quantum yield of fluorescence, Φ (RFU m-1), for the Med-Bouss dataset for each season and each sampling depth.

For comparison purposes, we present a summary of the regional distribution of slope factor (Figure 3.8a, same as Figure 3.1b), the Chla-specific absorption coefficient (Figure 3.8b) and fluorescence quantum yield (Figure 3.8c) based on the analysis of the Glo-Argo database. On the one hand, some regions located in the subtropical biome (SPSG and ARCH) show a high value of the a*(470) coefficient, yet a relatively low value of Φ. This suggests that, in such systems, the absorbed light energy is not dissipated as fluorescence, which would result in a decrease of the F-to-[Chla] ratio and, thus, of the slope factor (close to 1 in the case of ARCH). On the other hand, high-latitude regions (SANT and ARCT) show moderate to low and moderate to high values of a*(470) and Φ, respectively, suggesting that the little energy that is received by phytoplankton cells is mostly reemitted as fluorescence, which leads to an increase in the F-to-[Chla] ratio and high values of the slope factor (~3).

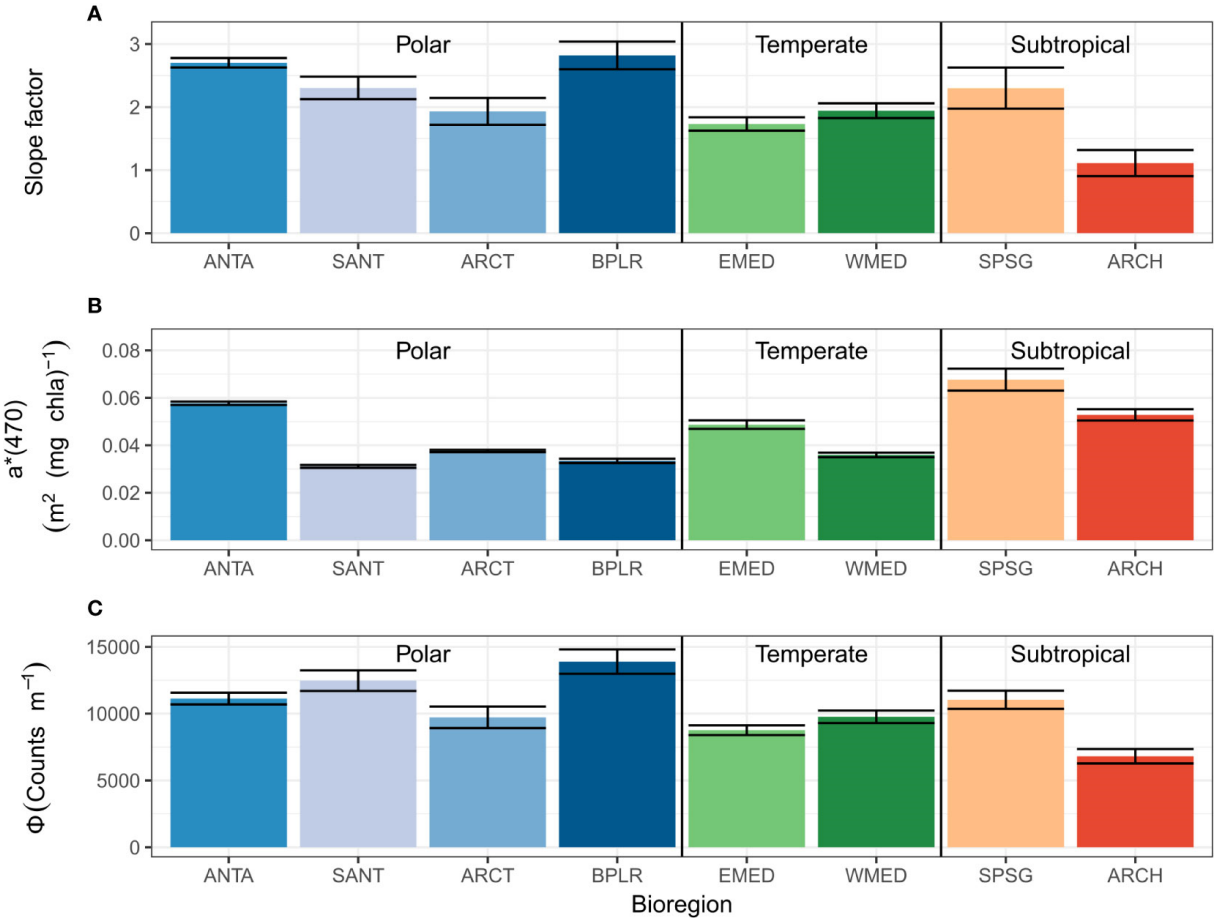

Figure 3.8: (A) Distribution of the slope factor, defined as the coefficient of the regression between [Chla]Fluo and [Chla]HPLC, (B) Chla-specific absorption coefficient at 470 nm, a(470) (m2(mg Chla)-1), and (C) fluorescence quantum yield, Φ (counts m-1), for the eight bioregions of the Glo-Argo dataset (ANTA, Antarctic Province; ARCH, Archipelagic Deep Basin Province; ARCT, Atlantic arctic Province; BPLR, Boreal Polar Province; EMED, Eastern Mediterranean Sea; SANT, Subantarctic Province; SPSG, South Pacific Subtropical Gyre Province; WMED, Western Mediterranean Sea). The error bars indicate 95% confidence limits.*

III.4. Discussion

The estimation of [Chla] based on *in-situ* fluorometers relies on the correlation between F and [Chla]. This correlation has been shown to be significantly variable at the global scale between different regions of the open ocean (Roesler et al., 2017). We studied here the role of the composition of phytoplankton communities on this variability through its influence on the Chla-specific absorption coefficient and fluorescence quantum yield.

As expected from numerous previous studies (e.g. Bricaud et al., 2004; Brewin et al., 2011; Uitz et al., 2015), we observe maximum values of a*(470) in surface oligotrophic waters, *i.e.* in seasonally or permanently stratified open ocean conditions where picophytoplankton generally dominate the phytoplankton communities (Figure 3.2). The shape of the phytoplankton absorption spectrum is known to be mainly influenced by the package effect occurring in phytoplankton cells and the accessory pigment composition (Greg Mitchell and Kiefer, 1988; Bricaud et al., 1995, 1999; Cleveland, 1995). Picophytoplankton-dominated communities are characterized by a limited package effect because of a small average cell size, and large concentrations of non-photosynthetic carotenoids such as zeaxanthin, both of which lead to enhanced values of a*(λ) in the blue spectral region (Allali et al., 1997; Ciotti et al., 2002; Bricaud et al., 2004; Uitz et al., 2008). In such systems, the effect of photoacclimation combines with that of community composition and influences the vertical distribution of the a*(470) coefficient (e.g. Bricaud et al., 2004; Uitz et al., 2008; 2015). First, acclimation to low light levels leads to lower a* values at depth due to an increase in the degree of pigment packaging associated with enhanced intracellular Chla content. Second, the zeaxanthin-to-Chla ratio is known to increase in high light regime (Six et al., 2004; Dubinsky and Stambler, 2009), resulting in larger a* values at surface.

Moreover, zeaxanthin is thought to be involved in photoprotection mechanisms (Demmig-Adams, 1990). This pigment enhances the energy dissipation through heat, hence diminishing the photosynthetic quantum yield (Bidigare et al., 1989; Olaizola and

Yamamoto, 1994) and the fluorescence quantum yield (Kiefer and Reynolds, 1992). This is consistent with our observation that bioregions with high concentrations of zeaxanthin, such as SPSG and ARCH, are also associated with relatively low values of the fluorescence quantum yield (Figure 3.8). The effect of the decrease in the fluorescence quantum yield due to zeaxanthin appears to be greater than that of the increase in a*(470), resulting in a lower slope factor in picophytoplankton-dominated communities.

In the Mediterranean Sea in summer, the surface waters show high values of the Chla-specific absorption coefficient associated with a dominant contribution of picophytoplankton to the algal community and large concentrations of zeaxanthin. Yet they do not present a significantly lower fluorescence quantum yield (Figure 3.7), compared to the other depths and seasons, as could be expected from our results based on the global dataset (Figure 3.8). This result may be explained by the optical specifications of the Chelsea Aquatracka III fluorometer used during the BOUSSOLE cruises. Indeed, this fluorometer has an excitation wavelength of 430 nm that does not coincide with the absorption peak of zeaxanthin, unlike the ECO fluorometers mounted on the BGC-Argo floats that have an excitation wavelength of 470 nm. The Chelsea Aquatracka III fluorometer excites a region of the phytoplankton absorption spectrum closer to the absorption peak of Chla in solution compared to the spectral region targeted by the ECO-series fluorometer. Thus, the influence of zeaxanthin on the fluorescence response is diminished. Our results, therefore, suggest that using a fluorometer with an excitation wavelength of 430 nm may reduce the effect of zeaxanthin on the fluorescence quantum yield and, thereby, the influence of the phytoplankton community composition on the slope factor and subsequent [Chla] retrieval.

The fluorescence quantum yield has been reported to be dependent on the physiological state of phytoplankton cells (Falkowski and Kolber, 1995; Behrenfeld et al., 2009). In this study, we found a significant influence of phytoplankton community composition. Consistent trends are observed between the datasets from the Northwestern Mediterranean Sea and global ocean. In particular, higher values of Φ are

found in communities characterized by a dominance of large phytoplankton cells and high [Chla], as is the case in the Northwestern Mediterranean Sea in spring or in high latitude bioregions. In contrast, lower Φ values are found in oligotrophic regions, which we attribute to the presence of photoprotective pigments, specifically zeaxanthin.

Explaining the high values of Φ in the high latitude regions remains a challenge. In the Southern Ocean, iron limitation plays an important role in the modulation of the fluorescence signal (Schallenberg, 2022). Southern Ocean phytoplankton communities often cope with iron and light colimitation (e.g. Boyd et al., 2000). In such conditions, they adapt their photosynthetic apparatus by increasing the size of the light-harvesting antenna in the photosynthetic unit, rather than by increasing the number of reaction centers, which would be iron-consuming (Strzepek et al., 2012, 2019; Schallenberg et al., 2020). An increase in the size of the antenna leads to a modification of the light transfer efficiency of the photosynthetic unit (Strzepek et al., 2012; Schallenberg et al., 2020), with a probable influence on the F-to-[Chla] ratio. Nevertheless, in this study, we observe high values of Φ in three different polar regions (SANT, ARCT, BPLR; Figure 3.8), not only in the Southern Ocean but also in the Arctic Ocean. This suggests that the large Φ values do not necessarily solely result from regional environmental conditions, such as iron depletion that is more specific to the Southern Ocean and may lead to atypical photophysiological properties (Behrenfeld et al., 2009). The large Φ values may also result from a common factor in these polar regions, such as a predominance of large phytoplankton cells (Figure 3.2). Due to the relatively small size of the Glo-Argo dataset, we could not investigate the temporal dynamics of the F-to-[Chla] ratio in relation to iron depletion or input events, which would help to disentangle the likely combined effects of iron and community composition on the fluorescence quantum yield.

In addition, the limited size of the Glo-Argo dataset hinders our ability to account for the seasonal and/or vertical variability of a*(470) and Φ in each bioregion that may lead to discrepancies between bioregions from the same biome. For instance, the SPSG bioregion presents a relatively high Φ despite belonging to the subtropical biome. The

SPSG bioregion is characterized by a relatively large diversity of phytoplankton community size structure, being dominated by picophytoplankton or nanophytoplankton (Figure 3.2). This diversity may lead to unexpectedly high values of Φ, compared to the other subtropical region. This result is consistent with our conclusion that, beyond the physiological state of phytoplankton cells, the fluorescence quantum yield may be largely influenced by the composition of phytoplankton communities.

The present study establishes a correlation between the phytoplankton community composition and the two major drivers of the *in-vivo* fluorescence process, a*(λ) and Φ. The effects of those two parameters are integrated in the slope factor index, which reflects the variability of the F-to-[Chla] ratio. Thus, the $[Chla]_{Fluo}$ estimation shows spatial, seasonal and vertical variability that reflects changes in phytoplankton community composition. Therefore, we suggest that integrating indicators of the phytoplankton community composition into the conversion of F into $[Chla]_{Fluo}$ would contribute to improving the estimation of $[Chla]_{fluo}$.

III.4. Conclusion and perspectives

In-situ fluorescence used in combination with autonomous platforms, such as BGC-Argo profiling floats, provides a powerful means for obtaining a global assessment of the Chla concentration, a major proxy of the phytoplankton biomass (e.g. Lombard et al., 2019; Claustre et al., 2020). However, the natural variability in the phytoplankton fluorescence response may considerably impede the estimation of the Chla concentration and, hence, needs to be apprehended (e.g. Proctor and Roesler 2010; Roesler et al., 2017). The present study aims to better understand the influence of phytoplankton community composition on the natural variability of the *in-situ* fluorescence signal. In this end, we analyzed concurrent measurements of fluorescence, pigment concentrations determined by HPLC, and phytoplankton absorption spectra from various open ocean regions, and

addressed the global variability in addition to the seasonal and vertical variability within a given region. Our results suggest a significant influence of phytoplankton community composition on the fluorescence signal, resulting from different photophysiological properties, i.e. the absorption coefficient and the fluorescence quantum yield, between phytoplankton groups.

Our results indicate that accounting for the composition of phytoplankton communities may help to better constrain the estimation of the Chla concentration based on *in-situ* fluorescence measurements. Hence, we suggest that incorporating an index of phytoplankton community composition into the conversion of F into $[Chla]_{fluo}$ may be a relatively simple, effective way to improve the quantification of the Chla concentration based on BGC-Argo float observations. Different methods have been introduced to derive information on the taxonomic or size structure of phytoplankton communities from BGC-Argo float measurements. In particular, Briggs et al. (2013) proposed a method to infer the mean size of the particle pool from high frequency time series of the particulate backscattering coefficient. Sauzède et al. (2015) developed a neural network-based approach to retrieve the Chla concentration attributed to three distinct phytoplankton size classes from the shape of the fluorescence profile. More recently, Rembauville et al. (2017) introduced a method that retrieves the relative contribution to the stock of particulate organic carbon of three phytoplankton size classes. While the methods of Sauzède et al. (2015) or Rembauville et al. (2017) provide comprehensive information on the community size structure from machine learning, some methods rely on the particulate backscattering-to-Chla ratio and can be used to estimate a simple optical index of the phytoplankton community composition (Cetinić et al., 2015; Lacour et al., 2017). Therefore, the global scale analysis of the covariance between such optical index of the community composition, a*(470) and Φ could allow to optimize the conversion of F into $[Chla]_{fluo}$ for application to BGC-Argo float fluorometers. Today, the challenge is to collect and merge the appropriate dataset. It should be large enough to cover different contrasted oceanic systems encompassing a continuum of phytoplankton community taxonomic and size structure and generate statistically robust results. The resulting

relationship could allow the emergence of a real time correction of the $[Chla]_{Fluo}$. Ultimately, the correction factor would reflect phytoplankton community composition variability and thus, would vary on the regional and seasonal scale. Thus, we strongly support Bittig et al. (2019) recommendation to collect seawater samples to perform phytoplankton HPLC pigments and spectral absorption analyses at BGC-Argo float deployment.

Our study shows that non-photosynthetic pigments may play an important role in the reduction of the ratio between fluorescence and Chla concentration in the surface layer of stratified oligotrophic waters. The current 470 nm excitation wavelength of the ECO-series fluorometer equipping BGC-Argo floats does not coincide with the spectral band of the Chla absorption maximum. The use of a fluorometer with an excitation closer to 430 nm would ensure to target the Chla in vivo absorption peak and, hence, likely reduce the influence of accessory pigments on the fluorescence signal (Bricaud et al., 2004; Proctor and Roesler, 2010). We recommend to investigate such possibility in the future, by deploying in the field a fluorometer with two excitation channels, one at 470 nm and one closer to 430 nm. This should be done in association with HPLC pigment measurements in order to compare the variability of the slope factors obtained from the two different excitation channels.

Eventually, taking into account the composition of phytoplankton communities for the retrieval of the Chla concentration from current in-situ fluorometers, or using new multi-channel sensors implemented on observation platforms, BGC-Argo floats in particular, will lead to substantially more robust estimates of the phytoplankton biomass on a broad range of spatial and temporal scales. Such quantitative information is critical to understand and model biogeochemical cycles in the global ocean. In a context of climate change, there is an urgent need to improve biogeochemical models whose initialization and validation will, in the future, rely more and more on BGC-Argo float observations, as this global observation system develops (Cossarini et al., 2019; Claustre et al., 2020).

III.5) Supplementary material

Table S1 : Statistics of the linear regression model applied to the data of each bioregion of the Glo-Argo dataset in order to estimate the regional values of the slope factor, a(470) (m-2 mg Chla-1) and Φ (counts.m-1).*

	Slope Factor				a*(470)				Φ			
Code	Slope Factor	Std. Error	p. value	Adj. R^2	a*(470)	Std. Error	p.value	Adj. R^2	Φ	Std. Error	p. value	Adj. R^2
ANTA	2.7	0.075	<0.001	0.98	0.06	0.0007	<0.001	1	586	142	<0.001	0.34
ARCH	1.11	0.207	<0.001	0.78	0.05	0.0024	<0.001	0.98	1484	244	<0.001	0.82
ARCT	1.93	0.213	<0.001	0.52	0.04	0.0005	<0.001	0.99	1331	305	<0.001	0.19
BPLR	2.82	0.22	<0.001	0.84	0.03	0.0009	<0.001	0.98	3265	899	<0.001	0.29
EMED	1.73	0.106	<0.001	0.91	0.05	0.0018	<0.001	0.96	1060	185	<0.001	0.54
SANT	2.3	0.17	<0.001	0.66	0.03	0.0006	<0.001	0.97	2455	522	<0.001	0.2
SPSG	2.3	0.326	<0.001	0.86	0.07	0.0046	<0.001	0.96	852	189	<0.001	0.71
WMED	1.94	0.11	<0.001	0.82	0.04	0.001	<0.001	0.96	1136	268	<0.001	0.22

Table S2 : Statistics of the linear regression model applied to each bioregion of the Glo-aphy dataset in order to estimate the regional values of a(470) (m-2 mg Chla-1).*

Code	a*(470)	std. Error	p.value	Adj. R^2
PEQD	0.06	0.001	<0.001	0.95
WARM	0.06	0.004	<0.001	0.97
SPSG	0.05	0.001	<0.001	0.93
ISSG	0.05	0.002	<0.001	0.97
NATR	0.04	0.002	<0.001	0.92
NADR	0.07	0.0008	<0.001	0.98
NASE	0.07	0.0004	<0.001	0.97
SSTC	0.05	0.0007	<0.001	0.99
MEDI	0.04	0.0004	<0.001	0.92
ANTA	0.04	0.001	<0.001	0.96
SANT	0.02	0.002	<0.001	0.98

Table S3 : Statistics of the linear regression model applied to each depth (in m) and seasons of the Bouss-Med dataset to estimates the values of a(470) (m^{-2} mg $Chla^{-1}$) and Φ ($RFU.m^{-1}$).*

		a*(470)				Φ			
Depth	Season	a*(470)	Std. Error	p.value	Adj. R^2	Φ	Std. Error	p.value	Adj. R^2
5	Winter	0.01	0.002	<0.001	0.8	12.94	2.7	<0.001	0.6
5	Spring	0.01	0.001	<0.001	0.94	14.76	0.9	<0.001	0.91
5	Summer	0.03	0.004	<0.001	0.52	12.68	2.5	<0.001	0.41
5	Fall	0.03	0.004	<0.001	0.7	18.28	6.6	<0.001	0.23
10	Winter	0.01	0.001	<0.001	0.89	17.15	2.3	<0.001	0.76
10	Spring	0.01	0.001	<0.001	0.93	20.30	1.2	<0.001	0.92
10	Summer	0.03	0.004	<0.001	0.61	15.34	2.6	<0.001	0.49
10	Fall	0.02	0.005	<0.001	0.55	15.65	6.9	<0.001	0.2
20	Winter	0.01	0.002	<0.001	0.84	16.32	3.1	<0.001	0.62
20	Spring	0.01	0.001	<0.001	0.76	27.27	3.3	<0.001	0.71
20	Summer	0.02	0.002	<0.001	0.71	9.41	2.5	<0.001	0.54
20	Fall	0.03	0.005	<0.001	0.6	6.64	6.5	<0.001	0.43
30	Winter	0.01	0.002	<0.001	0.68	22.75	4.1	<0.001	0.28
30	Spring	0.01	0.001	<0.001	0.57	35.99	6.3	<0.001	0.8
30	Summer	0.01	0.001	<0.001	0.55	14.93	3.4	<0.001	0.84
30	Fall	0.03	0.005	<0.001	0.6	10.83	3.1	<0.001	0.35
40	Winter	0.01	0.002	<0.001	0.59	29.23	4.6	<0.001	0.52
40	Spring	0.00	0.001	<0.001	0.7	0.80	9.1	<0.001	0.15
40	Summer	0.00	0.001	<0.001	0.76	4.76	4.5	<0.001	0.92
40	Fall	0.02	0.003	<0.001	0.46	18.21	3.6	<0.001	0.75
50	Winter	0.01	0.003	<0.001	0.78	19.19	5.1	<0.001	0.43
50	Spring	0.01	0.001	<0.001	0.96	17.25	3.6	<0.001	0.31
50	Summer	0.01	0.001	<0.001	0.8	14.54	3.3	<0.001	0.83
50	Fall	0.02	0.004	<0.001	0.7	12.64	7.9	<0.001	0.92
60	Winter	0.01	0.002	<0.001	0.82	29.84	4.2	<0.001	0.74
60	Spring	0.01	0.001	<0.001	0.74	18.40	4.7	<0.001	0.69
60	Summer	0.02	0.001	<0.001	0.71	13.38	2.4	<0.001	0.77
60	Fall	0.01	0.005	<0.001	0.66	28.89	14.1	<0.001	0.62

IV Assessing phytoplankton community composition using *in-situ* multispectral fluorescence and potential for application to BGC-Argo profiling floats

Flavien Petit[1], Julia Uitz[1], Louison Dufour[2], Frédéric Partensky[2], Laurence Garczarek[2], Priscillia Gourvil[3], Collin Roesler[4], Céline Dimier[5], Mélek Golbol[5], Vincenzo Vellucci[5], David Antoine[1,6], Christophe Penkerc'h[5], Vincent Taillandier[5], Hervé Claustre[1]

[1]Centre national de la Recherche Scientifique (CNRS) and Sorbonne Université, Laboratoire d'Océanographie de Villefranche (LOV), Villefranche-sur-Mer, France

[2]Sorbonne Université, CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment (AD2M), Station Biologique de Roscoff (SBR), Roscoff, France.

[3]Roscoff Culture Collection, Sorbonne Université, CNRS FR2424, Station Biologique de Roscoff (SBR), Roscoff, France.

[4]Department of Earth and Oceanographic Science, Bowdoin College, Brunswick, ME, USA

[5]Observatoire des Sciences de l'Univers (OSU) and Sorbonne Université, Institut de la Mer de Villefranche (IMEV), Villefranche-sur-Mer, France

[6]Remote Sensing and Satellite Research Group, School of Earth and Planetary Sciences, Curtin University, Perth, WA 6845, Australia

IV.1. Introduction

Phytoplankton plays a key role in the global biogeochemical cycles (Field et al., 1998). In the current context of climate change, it is therefore crucial to monitor phytoplankton dynamics on the global scale. The emergence of new observation tools such as Biogeochemical Argo (BGC-Argo) profiling floats opens the possibility to embark sensors to collect continuous vertical profiles of optical and biogeochemical variables on a quasi-global scale (Claustre et al., 2020). Among those, fluorescence enables to estimate the chlorophyll-a concentration ([Chla]), a widely used proxy for phytoplankton biomass. However, the composition of phytoplankton communities is known to be a critical determinant of the carbon cycle, since some processes can largely vary between phytoplankton size classes and/or phylogenetic groups, such as the CO_2 assimimation through photosynthesis (Cermeño et al., 2005; Uitz et al., 2008), trophic interactions (Cushing, 1989; Finkel, 2007) and carbon transfer dynamics to the deep ocean (Buesseler et al., 1998; Guidi et al., 2009; Henson et al., 2012; Bonnet et al., 2023). In spite of this key role, the composition of phytoplankton communities cannot be measured directly from BGC-Argo floats. Only a few methods have been proposed so far to overcome this lack and go beyond the mere estimation of the Chla biomass from BGC-Argo floats. Sauzède et al. (2015) developed a neural network based on the shape of *in-situ* fluorescence profiles to retrieve the relative contribution to the [Chla] of the three phytoplankton size classes (pico-, nano- and microphytoplankton). Rembauville et al. (2017) developed a regional scale approach to estimate the particulate organic carbon stock of these three size classes from a combination of optical measurements and hydrographic parameters. Cetinic et al. (2015) proposed a simple index using [Chla] and particulate backscattering coefficient to determine changes in phytoplankton community composition. Similarly, Terrats et al. (2020) used the ratio between *in-situ* fluorescence and particulate backscattering coefficient to detect coccolithophore blooms from BGC-Argo floats. Yet, while those methods provide useful information about the phytoplankton community

composition, neither of them can be used to retrieve information on their taxonomic composition at the global scale, as possible using BGC-Argo floats.

An alternative to these approaches is to use multispectral fluorescence to retrieve information about the relative pigment composition, from which major taxa can be discriminated. Multispectral fluorescence consists in measuring *in-situ* fluorescence signal from different excitation wavelengths, corresponding to the absorption peaks of accessory pigments used as biomarkers of specific taxa in the phytoplankton community (e.g. Bricaud et al., 2004; Uitz et al., 2006; Brewin et al., 2014). A combination of three wavebands centred around 435, 470 and 532 nm was previously investigated in coastal and freshwater environments, providing promising results and paving the way for its use in open ocean waters (Proctor and Roesler, 2010; Thibodeau et al., 2014).

The present study aims at assessing the potential of *in-situ* multispectral fluorescence as a proxy of the marine phytoplankton taxonomic community composition, as this could enable oceanographers to retrieve this key parameter from BGC-Argo floats which are deployed at many sites in the open ocean. Our working hypothesis was that the pigment composition of phytoplankton communities from different sites or seasons differs sufficiently to induce a significant variability of the fluorescence response between the three excitation wavelengths of the multispectral fluorometer, so that this variability can be used to define a classification method from the fluorometer output. To test this hypothesis, we combined laboratory and field work to investigate the possibility to retrieve taxonomic information from multispectral fluorescence (MSF). First, we selected ten different phytoplankton strains that correspond to the typical diversity observed during the seasonal succession of the North Western (NW) Mediterranean Sea, where the field measurements were conducted. We measured the fluorescence response of each strain in controlled conditions and tested the multispectral fluorometer sensitivity to the variations in the phytoplankton taxa. Second, we collected concomitant measurements of MSF and pigment concentrations to test the possibility of phytoplankton community composition discrimination from MSF during a time series of one year in the NW Mediterranean Sea. The NW Mediterranean Sea was chosen as a study site because of its marked seasonality, with spring blooms followed by summer oligotrophic periods and

winter mesotrophic periods (D'Ortenzio et al., 2005; Lavigne et al., 2015). Thus, we assume that the various contrasted environments encountered seasonally in the NW Mediterranean Sea is representative of the variability of trophic conditions encountered in the global open ocean. Based on those observations, we designed a classification method to assess the different phytoplankton communities based on a year of observations of phytoplankton community composition and multispectral fluorescence values in combination with bio-optical sensors similar to those fitted to standard BGC-Argo floats. Ultimately, we provide recommendations for the use of multispectral fluorescence to discriminate the taxonomic composition of the phytoplankton community from BGC-Argo profiling floats.

IV.2. Materials & Methods

IV.2.1. Laboratory work

IV.2.1.1. Phytoplankton strains and growth conditions

Ten strains were selected for the laboratory experiment and provided by the Roscoff Culture Collection (RCC; https://roscoff-culture-collection.org/). They are representative of the taxonomic diversity of the major components of eukaryotic and prokaryotic phytoplankton communities in open ocean waters, such as the Mediterranean Sea. The selection includes three diatoms belonging to different size classes, one pelagophyte, one dinoflagellate and five picocyanobacteria: (three *Synechococcus* and two *Prochlorococcus;* Table 4.1). All strains were grown under the same controlled conditions at a constant temperature of 21°C and 50 µE m^{-2} s^{-1} of continuous white light, in K+Si (Keller et al., 1987) or PCR-S11 medium (Rippka et al., 2000) for eukaryotes and prokaryotes, respectively. As fluorescence is significantly influenced by the physiology of phytoplankton cells, it is essential to use cultures in good physiological status, as assessed by a high fluorescence yield (F_v/F_M) value using a PhytoPAM-II fluorometer (Walz, Effeltrich, Germany). The F_v/F_M

parameter was calculated as $(F_M - F_0)/F_M$, where F_0 is the dark-adapted minimal fluorescence and F_M the maximal fluorescence, associated with the closing of photosynthetic reaction centers. F_M was measured after exposure to a saturating light and addition of 100 µM of the photosystem II inhibitor 3'-(3,4-dichlorophenyl)-1',1'-dimethyl urea (DCMU; Parkhill et al., 2001), The F_v/F_M parameter was measured concomitantly to cell counts using a Guava EasyCyte flow cytometer (Luminex Corporation, USA) all over the growth of each phytoplankton culture. The MSF protocol (see section 2.1.2) was performed on each culture in the late exponential growth phase, just prior to the drop of the F_v/F_M index. The MSF protocol was repeated three times for each strain (biological triplicates).

IV.2.1.2) Multispectral fluorescence measurements

MSF was measured for different [Chla]. After a dark acclimation for 2 h, each culture was diluted in different volumes of fresh medium to produce a dilution series ranging from 0.1 to 10 mg Chla m^{-3}. Fluorescence measurements were performed as soon as possible after the dilution to avoid any physiological stress of the diluted culture using a ECO 3X1M (SeaBird electronics, USA), with three channels corresponding to the following excitation wavebands: 435, 470 and 532 nm. The 3X1M sensor outputs were recorded with the TeraTerm software. Each experimental culture was diluted in a 1L beaker that was then placed under constant slow steering. The multispectral fluorometer was placed at the center of the beaker and immersed by 5 mm. Control measurements were performed in fresh medium to ensure strains MSF measurements were not subjected to optical interferences from the beaker edge scattering. For each culture and each dilution, we measured the fluorescence response during three series of one minute of continuous acquisition, each separated by two minutes of darkness.

IV.2.2. Time series acquisition in the field

The same multispectral fluorometer was deployed at sea every month from December 2020 to October 2021, at the BOUSSOLE station, a long-term monitoring site located at 7°54'E, 43°22'N in the Ligurian current, NW Mediterranean Sea (Antoine et al., 2008). On each monthly cruise, a CTD-rosette device equipped with a dedicated optical package was used to perform casts from the surface down to 400 m depth. The optical package includes our multispectral fluorometer, a standard ECO-series FLBBCD (SeaBird) fluorometer, a backscatter meter as well as a C-Rover transmissiometer (Seabird). The ECO FLBBCD measured *in-situ* fluorescence, with an excitation peak at 470 nm and emission detection at 695 nm, and particulate backscattering (b_{bp}) at 700 nm. The C-Rover transmissometer measured beam attenuation, from what the particulate attenuation (c_p) is derived by removing the attenuation of the dissolved fraction. Both sensors have been mounted on several BGC-Argo floats for different biogeochemical applications (Rembauville et al., 2017; Barbieux et al., 2022). Concomitantly, seawater was sampled at 10 discrete depths for pigment identification and quantification by High Performance Liquid Chromatography (HPLC).

Table 4.1 : Name, taxonomy, pigment composition as detected by HPLC and size class of the ten strains used in the laboratory protocol.

Pico stands for picophytoplankton (0.2-2 µm), Nano for nanophytoplankton (2-20 µm) and Micro for microphytoplankton (20-200 µm); HL stands for high-light adapted; LL stands for low-light adapted.

Species name	**Class**	**RCC strain number**	**Other name**	**HPLC measured pigments**	**Size class**
Conticribra (Thalassiosira) weissflogii	Mediophyceae (diatom)	RCC76	CCMP1336	Fuco, Chl C1 + c2, Diad, Diat,	Micro
Chaetoceros diadema	Mediophyceae (diatom)	RCC1717	RA080513-06	Fuco, Diad, Chl c1 + c2	Micro
Pelagomonas calceolata	Pelagophyceae	RCC100	CCMP1214	Chl c3, Chl c1 +c2, But, Fuco, Diad, Diat	Nano
Scrippsiella sp.	Dinophyceae	RCC3006	VFAC24-3	Chl c1 + c2, Peri, Diad	Nano
Minidiscus sp.	Mediophyceae (diatom)	RCC4213	MACUMBA-SC18	Chl c1 + c2, Fuco, Diad, Diat	Nano
Prochlorococcus marinus (LL)	Cyanophyceae	RCC156	SS120-04/95	Zea, Dv Chlb, Dv Chla	Pico
Prochlorococcus marinus (HL)	Cyanophyceae	-	PCC9511	Zea, Dv Chl b, Dv Chl a	Pico
Synechococcus sp.	Cyanophyceae	RCC2319	MINOS11	Zea	Pico
Synechococcus sp.	Cyanophyceae	RCC2374	A15-62	Zea	Pico
Synechococcus sp.	Cyanophyceae	RCC2379	BOUM118	Zea	Pico
Prochlorococcus (High light)	Cyanophyceae	-	PCC9511	Zea, Dv Chl b, Dv Chl a	Pico

IV.2.3. Processing of fluorescence and optical measurements

The output of the fluorometer was expressed as digital counts (DC). For the culture measurements, a blank was performed on each dilution series by measuring the response of the culture medium alone. This value was then subtracted from the raw output of the sensor. The three consecutive series of acquisition of each dilution were then averaged. The dilutions series allowed us to define a Chla biomass calibration value, expressed as DC $(mg\ m^{-3})^{-1}$, for each of the three excitation wavelengths and for the ten selected strains. This calibration value is the regression coefficient of a linear regression performed between the fluorescence response expressed in DC and the Chla biomass in $mg\ m^{-3}$ for the entire dilution range and for each replicate of a given phytoplankton strain.

For the *in-situ* field time series, the factory-determined dark value was validated in the lab with black tape on the sensor and was subtracted from the raw DC, following the BGC-Argo data management recommendations (Schmechtig et al., 2018a). The optical backscattering coefficient was acquired during the downcast of the rosette used for water sampling. The angular scattering coefficients β was measured every second at a central angle of 124° and at a wavelength of 700 nm. The β coefficient was converted into the particulate angular scattering coefficient βp by removing the contribution of pure seawater, which in turn depends on temperature and salinity. Then, βp was converted into b_{bp} following the guidelines to convert raw signal to b_{bp} and applying a χ factor equal to 1.076 (Schmechtig et al., 2018b).

For fluorescence, C_p and b_{bp} outliers were detected and removed using a threshold of 1.5 · simple moving average (Δ depth = 3 m). Each profile was then smoothed using a simple moving average (Δ depth = 3 m).

IV.2.4. Phytoplankton pigments

For both the BOUSSOLE *in-situ* and laboratory samples, [Chla] measurements and composition of phytoplankton communities were made using HPLC. In brief, 2.7L of

seawater from discrete sampling was filtered onto glass fiber filters (GF/F Whatman 25 mm), stored in liquid nitrogen during cruises, then transferred at −80°C in the laboratory until further analysis at the SAPIGH HPLC analytical facility of the 'Institut de la Mer de Villefranche' (IMEV; https://lov.imev-mer.fr/web/facilities/sapigh/). Phytoplankton pigments were extracted from the cells by sonication in 100% methanol, clarified by filtration (GF/F Whatman 0.7 µm), and finally separated and quantified by HPLC. More details about the HPLC analytical protocol may be found in Ras et al. (2008). The total [Chla] was defined as the sum of Chla, divinyl-chlorophyll-a and chlorophyllid-a concentrations.

As concerns the *in-situ* data, we specifically investigated the distribution of seven diagnostic pigments (DP) identified as biomarkers of major phytoplankton taxa, which were further grouped into three phytoplankton size classes : micro- (>20µm), nano- (2-20µm) and picophytoplankton (<2µm; Claustre, 1994; Vidussi et al., 2001). Following equations given in Uitz et al. (2006), the DP-based method allowed the estimation of the relative contribution to the [Chla] of these three size classes. Because it relies on biomarker pigment concentrations, this approach yields an average, synthetic estimate of both the taxonomic and size composition of the phytoplankton communities. Although it has limits because some phytoplankton taxa may occasionally span over several size classes and some DP may be found in several taxa, this approach has been shown to provide reliable, quantitative information for use on large spatial and temporal scales (e.g. Vidussi et al., 2001; Bricaud, 2004; Uitz et al., 2006; Brewin et al., 2014).

IV.2.5. Statistical analyses

IV.2.5.1) Clustering of phytoplankton pigment data

The *in-situ* HPLC data were clustered in order to define phytoplankton communities that are meant to be discriminated by a classification method. A suite of pigment concentrations, combining DP and pigment observed on the selected strains (i.e.,

peridinin; 19'-butanoyloxyfucoxanthin; fucoxanthin; 19'-hexanoyloxyfucoxanthin; diadinoxanthin; alloxanthin; zeaxanthin; divinyl-chlorophyll *b*; chlorophyll *b* and divinyl-chlorophyll *a*) were used as descriptors of a correspondence analysis (CA). The first two dimensions of the CA were used as numerical estimation of the pigment composition resemblance of each sample. A clustering was then applied onto the first 2 dimensions of the CA, using a Hierarchical Ascending Classification (HAC). The method allows one to define clusters based on the relative pigment composition of the community, not on the absolute concentration value of each pigment. The resulting cluster dendrogram was cut at a height of 20, leading to three clusters. As there is a strong difference in the pigment composition of prokaryotic picophytoplankton and micro-/nanophytoplankton communities, the same method was applied after excluding picophytoplankton-dominated samples to better assess the variability of micro-/nanophytoplankton dominated communities. The cluster excluding picophytoplankton was then divided into two different clusters. In the end, four different phytoplankton communities were defined.

IV.2.5.2) Classification of phytoplankton groups

The classification of the *in-situ* phytoplankton communities derived from the MSF and optical measurements was performed with a Histogram Gradient Boosting algorithm. This type of machine learning model is particularly well suited for tabular data, comprising samples (rows) with the same set of features (columns) and with a low number of observations (Chen and Guestrin, 2016; Shwartz-Ziv and Armon, 2022). The imbalance in the number of samples per cluster was counterbalanced by a Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). Each cluster was oversampled in order to get as many observations as in the largest cluster. In the end, each cluster was represented by 32 samples. As the dataset consists of a time series of a single year of

phytoplankton community succession, the phytoplankton biomass was strongly correlated to phytoplankton community composition. In order to avoid overfitting due to the strong correlation between the different measured variables (*i.e.* F_{440}, F_{470}, F_{532}, b_{bp}, C_p) and phytoplankton biomass, we used different ratios. F_{440} and F_{532} were divided by F_{470} and the three fluorescence signals were divided by b_{bp} and C_p. The refractive index, representative of particulate composition and calculated as b_{bp}/C_p, was also used (Twardowski et al., 2001; Boss et al., 2004). The hyperparameters of the model, i.e. the parameters influencing the learning process, were defined using a cross validation grid search. In brief, the model has a learning rate of 0.05, a number of 400 estimators and a maximum depth of 8. The influence of the different descriptors was inspected through the mean impurity index, which reflects the importance of each descriptor in the succession of the decision trees. The model was validated with 20 cross validations, using a stratified shuffle split method with a test size of 20%, which allows one to obtain the same proportion of the four clusters in each learning and testing dataset with a random sampling. The classification results can be categorized in four different categories: True Positive (TP) corresponding to the accurate prediction of the presence of a class, True Negative (TN) to the accurate prediction of the absence of a class, False positive (FP) to the wrong prediction of the presence of a class and False negative (FN) to the wrong prediction of the absence of a class. The performance of the classification method was assessed through two different parameters, precision and recall, defined as follows:

$$\text{Precision} = TP/TP+FP \quad (1)$$

$$\text{Recall} = TP/TP+FN \quad (2)$$

The precision can be interpreted as the percentage of accurate prediction of the model, while the recall can be interpreted as the percentage of samples that have been correctly predicted by the model. Ultimately, the performance of the HGB classification model was tested regarding four different sensor combinations in view of potential future applications to BGC-Argo profiling floats.

IV.3. Results and discussion

IV.3.1. Qualification of the MSF using laboratory phytoplankton cultures

Once the calibrated fluorescence at each excitation wavelength was quantified for each strain, the two fluorescence ratios were computed, $F_{chl}(440):F_{chl}(470)$ and $F_{chl}(532):F_{chl}(470)$. These ratios vary by a factor of 5 depending on the strain (Figure 4.1). *Synechococcus* strains consistently showed a high $F_{Chl}532/F_{Chl}440$ and $F_{Chl}470/F_{Chl}440$ ratios around 1 and 3, respectively. The other studied taxa have substantially lower $F_{Chl}532/F_{Chl}440$ and $F_{Chl}470/F_{Chl}440$ ratios. *Prochlorococcus* has the lowest $F_{Chl}532/F_{Chl}440$ and $F_{Chl}470/F_{Chl}440$ ratios of all taxa, with values of 0.02 and 1, respectively. Diatoms show similar $F_{Chl}470/F_{Chl}440$ values to *Prochlorococcus* with higher $F_{Chl}532/F_{Chl}440$ reaching 0.14. The pelagophyte and dinoflagellate strains exhibited the highest $F_{Chl}470/F_{Chl}440$ ratio of all tested eukaryotic taxa with a mean value of 1.35. The pelagophyte distinguished itself from the dinoflagellate by its lower $F_{Chl}532/F_{Chl}440$ ratio (0.08 vs. 0.18).

Higher values of $F_{Chl}532/F_{Chl}440$ for the *Synechococcus* taxon may be explained by a higher fluorescence response to the 532 nm excitation channel, induced by the presence of phycoerythrin, a phycobiliprotein systematically found in open ocean *Synechococcus* (Six et al., 2007; Grébert et al. 2018). More generally, phycobiliproteins extend the absorption properties of cyanobacteria to the green/red part of the visible light spectrum with regard to Chla (Frank and Cogdell, 2012; Holtrop et al., 2021). The difference in the $F_{Chl}470/F_{Chl}440$ ratio values between pelagophytes and diatoms may be explained by a difference in the carotenoid compositions.

The intra-taxa variance of the two ratios being lower than the inter-taxa variance, the MSF at 440, 470 and 532 nm appears to be sensitive enough to discriminate the four different taxa in controlled laboratory conditions. Thus, we hypothesized that it can provide information on the taxonomic composition of phytoplankton communities in the field, at least when in-situ phytoplankton community pigment composition is similar to the one tested in the laboratory. To validate this assumption, we evaluated the possibility to

discriminate different phytoplankton groups in the NW Mediterranean Sea based on *in-situ* MSF and additional optical measurements.

IV.3.2. Phytoplankton communities in the Northwestern Mediterranean Sea

Vertical profiles of MSF, b_{bp} and C_p were acquired simultaneously with seawater pigment analyses at the BOUSSOLE site once a month for one year. This site is characterized by a strong seasonality that leads to the presence of contrasted phytoplankton communities throughout the year (marty et al., 2002). The diversity of phytoplankton communities may be compared to the trophic gradient observed on the global scale going from microphytoplankton-dominated communities during the seasonal bloom in temperate region, to picophytoplankton-dominated communities in stratified oligotrophic conditions, observed seasonally in temperate region or permanently in subtropical region (Lavigne et al., 2015; Mayot et al., 2017).

The composition of the phytoplankton communities was inferred using a pigment-based clustering approach, allowing to group samples with similar pigment composition, which led to the discrimination of four distinct phytoplankton clusters over the year (Figure 4.2). The first cluster corresponds to deep, late summer and winter communities with a large proportion of picophytoplankton and a significant contribution to pigment composition of chlorophyll-b. The second cluster coincides with the bloom community with a shared contribution of microphytoplankton and nanophytoplankton, and with a high fucoxanthin contribution, typically associated with diatoms. The third cluster is associated with summer communities below the deep chlorophyll maximum (DCM), also exhibiting a mixed composition of micro- and nanophytoplankton. Finally, the fourth cluster is characteristic of surface summer picophytoplankton-dominated communities, typically associated with *Synechococcus*.

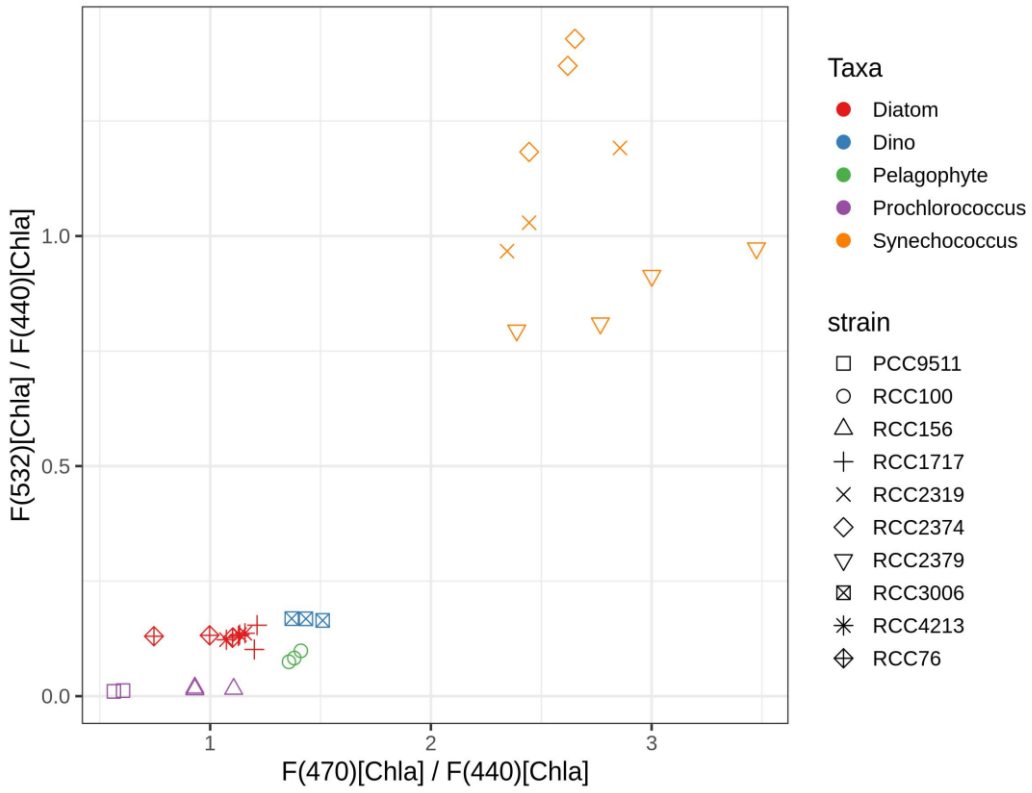

Figure 4.1 : Scatterplot of the ratios of F440/F470 vs. F532/F470 for each phytoplankton strain grown in culture. The color code indicates the taxon to which each strain belongs; the symbols indicate the strain code in the RCC.

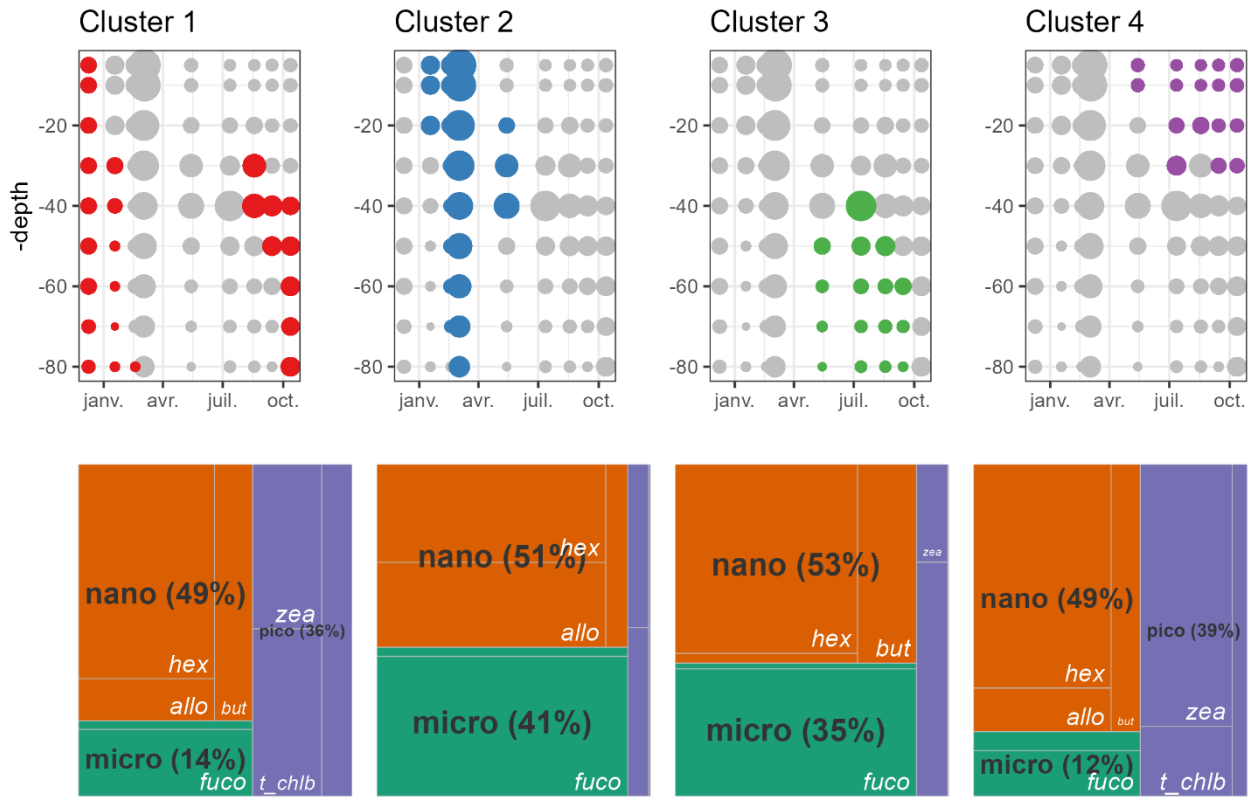

Figure 4.2 : Vertical distribution of four pigment-determined clusters indicative of different phyto-plankton communities over an annual cycle in the Northwestern Mediterranean Sea (BOUSSOLE site). The size of the dots indicates the chlorophyll a concentration, a proxy of the phytoplankton biomass. Treemap of the relative pigment concentration of each cluster, with the size class corresponding to the pigment taxa affiliation as color.

After demonstrating that the MSF allows discrimination of five different taxa, we compared the relative pigment composition of our *in-situ* clusters to that of the five laboratory characterized taxa. As the main driver of MSF variability between phytoplankton communities is the pigment composition, this analysis will determine if we are in the range of variability where we know that the MSF can discriminate different phytoplankton taxa. A correspondence analysis was applied to the pigment composition determined for the ten phytoplankton strains grown in the laboratory. This method allows a visualization of the different strains in a space where the distance between two samples reflects their relative pigment composition similarity (Figure 4.3). We observe three distinct poles corresponding to the different taxa represented by the ten selected strains. One is composed of diatoms and pelagophytes, while the two others correspond to *Synechococcus* and *Prochlorococcus*, respectively. The data from the *in-situ* seawater samples are projected on the same graph as supplementary observations. These data are evenly spread in the center of the plan, indicating that the variability in the pigment composition in the field is similar to that observed in the laboratory cultures. Moreover, the four field-based clusters are well distinguished in the CA projection plan. As we already showed that the MSF can be used to discriminate the different laboratory characterized taxa, we can expect that it will be sensitive enough to discriminate the *in-situ* clusters that have a comparable pigment composition.

However, the *in-situ* samples have lower eigenvalues, *i.e.* absolute values on CA axes, indicating that the pigment variability is less contrasted than in laboratory strains. This result is not surprising because of the complexity of the pigment composition in natural samples associated with mixed phytoplankton assemblages, instead of a single taxon in monospecific cultures. This characteristic of open ocean samples could somewhat hamper the possibilities of inferring information on phytoplankton community composition from MSF measurements. This will be tested in the next section.

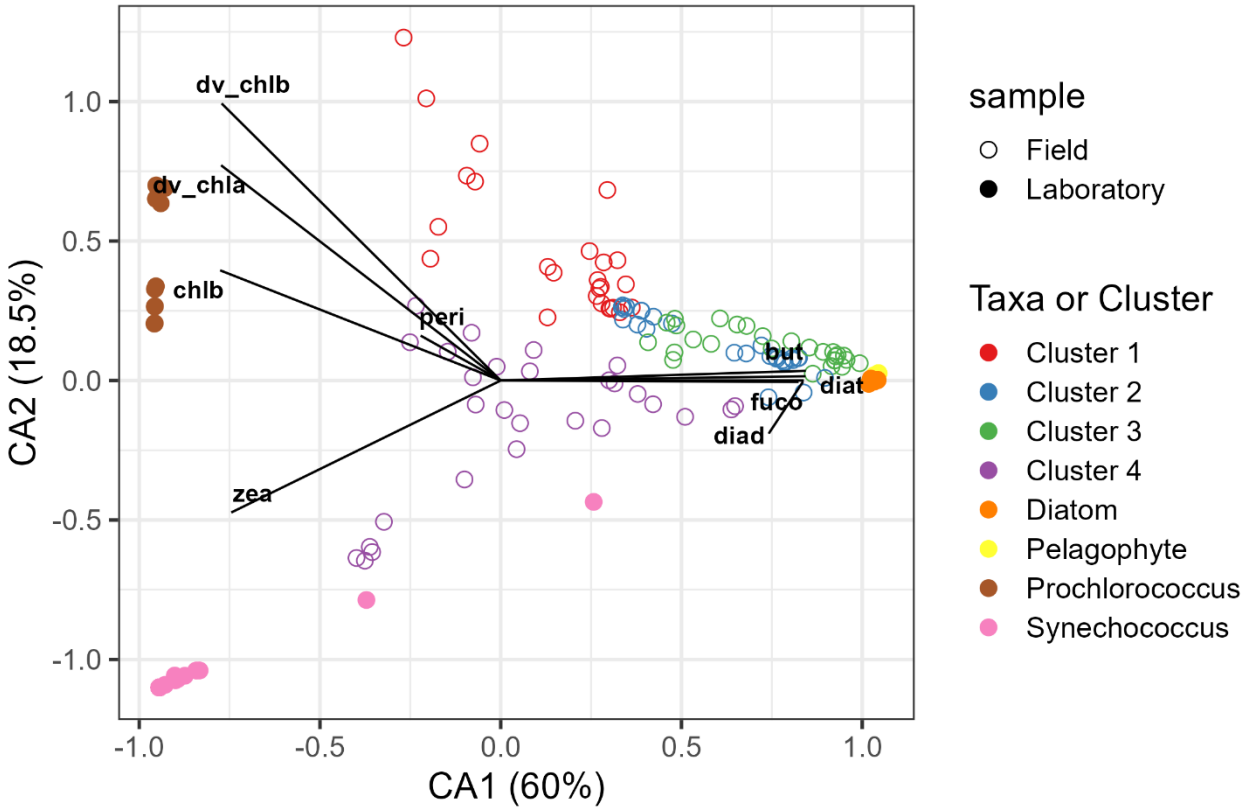

Figure 4.3 : Correspondence analysis of the pigment concentrations of the strains grown in culture. The in-situ pigment concentrations measured in the BOUSSOLE seawater sample are represented using the same color code as in Figure 4.2 and projected as supplementary observations.

IV.3.3. Discrimination of phytoplankton taxa from *in*-situ fluorescence and additional optical measurements

In view of application of MSF to BGC-Argo profiling floats, we evaluated the benefit of using additional optical sensors in combination with MSF for the discrimination of phytoplankton taxa. Several sensor configurations were tested. We trained a Histogram Gradient Boosting (HGB) to predict four different phytoplankton clusters from a combination of MSF, b_{bp} and C_p.

The mean global accuracy of the HGB classification using all the descriptors is 72% (+/- 10%) (Table 4.2). This demonstrates the possibility to use MSF in combination with a backscatterometer and a transmissometer to discriminate phytoplankton taxonomical groups from sensors embarked on BGC-Argo floats.

The mean importance of descriptors in the discrimination of clusters (*i.e.* impurity) (Figure 4.4, Table 4.2) indicates a significant role of all the descriptors with a particular importance of the bb_p/C_p, F_{470}/F_{440} and F_{532}/F_{440} ratios.

BGC-Argo is a global network of profiling floats dedicated to the observation of biogeochemical variables (Claustre et al., 2020). Currently, two types of sensor configuration, including at least two excitation wavelengths for fluorescence, have been deployed. The first one is composed of F_{440}, F_{470}, b_{bp} and C_p and the second one is the same without C_p measurements. In the following section, we will investigate how the model perform using these two configurations, and estimate how it would perform considering a potential integration of the ECO 3X1M on the profiling float. To do so, we tested the prediction of the four clusters with a different combination of descriptors, corresponding to different BGC-Argo float sensor packages (Figure 4.5). In the case of a sensor package with MSF (F_{440}, F_{470} F_{532}), backscatter meter ($bb_{p)}$ and transmissiometer (c_p), the precision and recall scores are homogeneous among all four clusters with values varying between 65% and 75%. A configuration with F_{440}, F_{470}, bb_p and c_p or with F_{440}, F_{470}, F_{532} and bb_p, lead to an overall slightly lower precision and recall scores and with more

variability, although all four clusters display scores above 60%. When only taking into account F_{440}, F_{470} and bb_p, the scores are significantly lower. Finally, the use of MSF measurements only, F_{440} F_{470} and F_{532}, led to variable performances depending on the cluster. Thus, removing transmissometer or the 532 nm excitation fluorescence seemingly trigger a significant decrease in the global accuracy and recall of the model (Figure 4.5), which may lead to a lower performance of some BGC-Argo sensors packages.

Table 4.2 : Description and impurity of the different descriptors of the four environmental phytoplankton clusters defined in Fig. 4.2.

	F440/F470	**F532/F470**	**BBP/Cp**	**F440/BBP**	**F470/BBP**	**F532/BBP**	**F440/Cp**	**F470/Cp**
mean	0.9	0.6	894	0.73	0.84	0.45	609	685
std	0.09	0.25	424	0.28	0.4	0.16	276	314
min	0.63	0.19	0.1	0.37	0.35	0.3	0.05	0.05
max	1.18	1.1	1983	1.68	2.22	1.46	1369	1527
Impurity	0.19	0.19	0.24	0.07	0.08	0.08	0.07	0.05

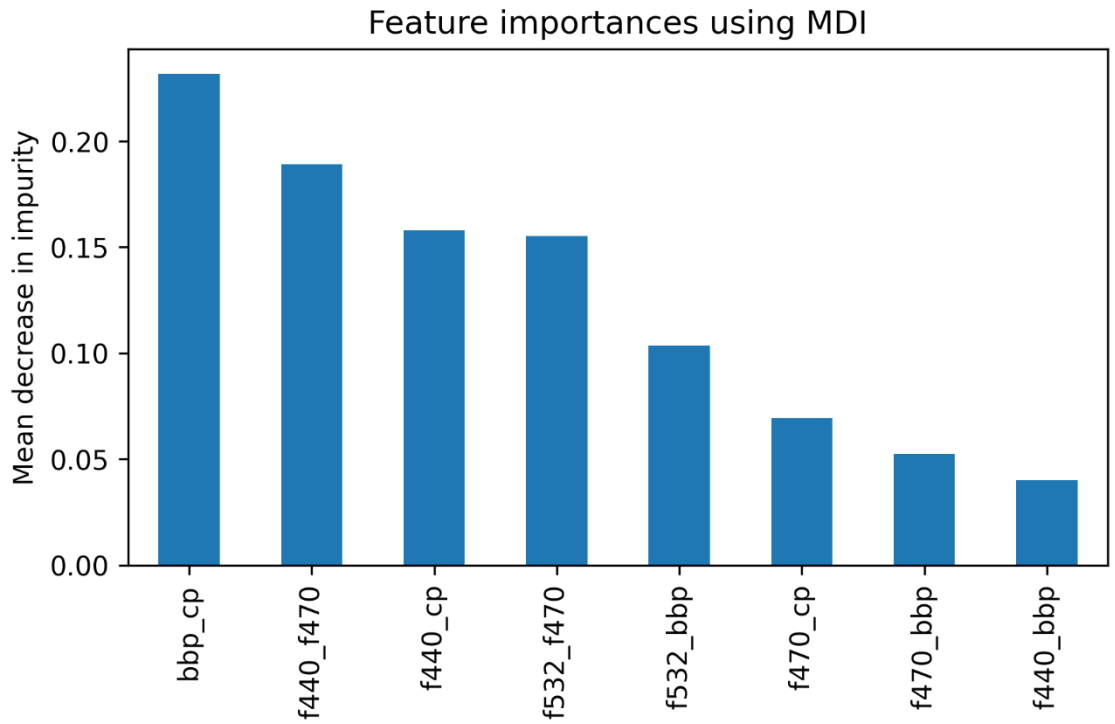

Figure 4.4 : Importance of the different descriptors in the classification model, expressed as the mean decrease impurity.

In order to compensate for the decrease of performance of the classification, we may reduce the level of discrimination of the model by reducing the number of clusters. In doing so, we were able to evaluate the performances on two other clustering choices (Figure 4.6). On the first one, we reduced the number of clusters to three clusters by grouping the cluster 2, made of mixed communities dominated by nano and microphytoplankton with a domination of alloxanthin, and cluster 3, gathering mixed communities dominated by 19'-HF. This led to a discrimination between surface, summer picophytoplankton with zeaxanthin, deep and winter picophytoplankton with chlorophyll and divinyl chlorophyll b and mixed communities of micro and nanophytoplankton. On the second clustering choice, we predicted only two clusters corresponding to picophytoplankton with zeaxanthin or chlorophyll b, and mixed communities of microphytoplankton and nanophytoplankton. The performance of the model, for each BGC-Argo sensors configuration (Figure 4.5), is evaluated by looking at the mean balanced recall on the same cross validation method as used before (Figure 4.6). In the configuration where all sensors are available, the number of predicted clusters does not influence the performance with a consistent balanced recall of around 75%. On the contrary, with fewer sensors, reducing the number of clusters leads to an increase in performances. When there is only two fluorescence excitation wavelengths, *i.e.* F_{440} and F_{470}, the absence of c_p measurement leads to a decrease of recall from 71.1% to 57.7% when four clusters are predicted. However, if the number of clusters is reduced to only two, the recall of configuration with F_{440}, F_{470} and b_{bp} increase from 57.7% to 78% (Figure 4.6 and Table 4.2). Thus, these results enlighten the possibility of retrieving taxonomic information with fewer sensors, which could be of particular interest in view of the integration of a new ECO-puck sensor measuring F_{440}, F_{F470} and BBP parameters from BGC-Argo floats.

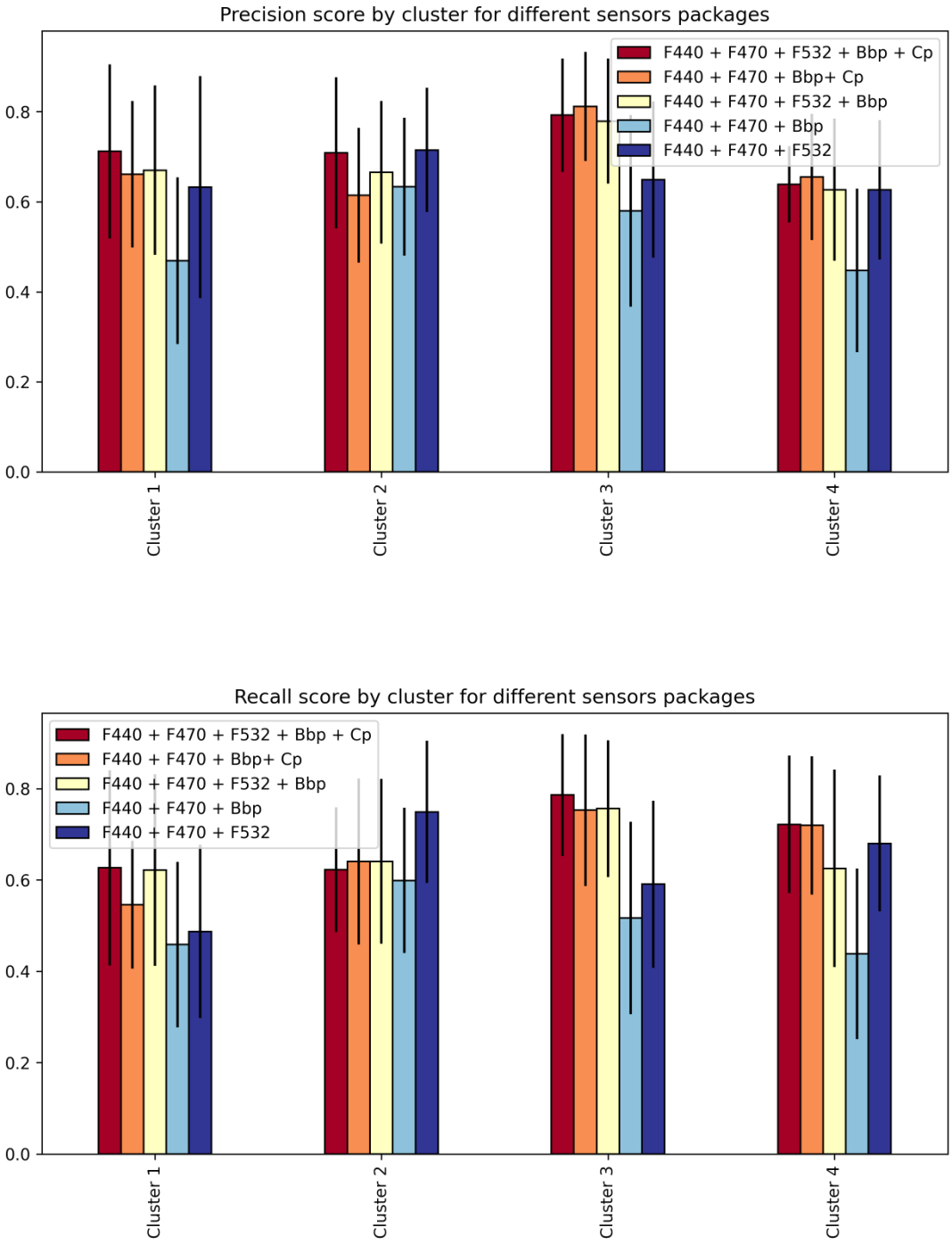

Figure 4.5 : Accuracy and recall of the HGB classification model for each cluster on each BGC-Argo sensors package

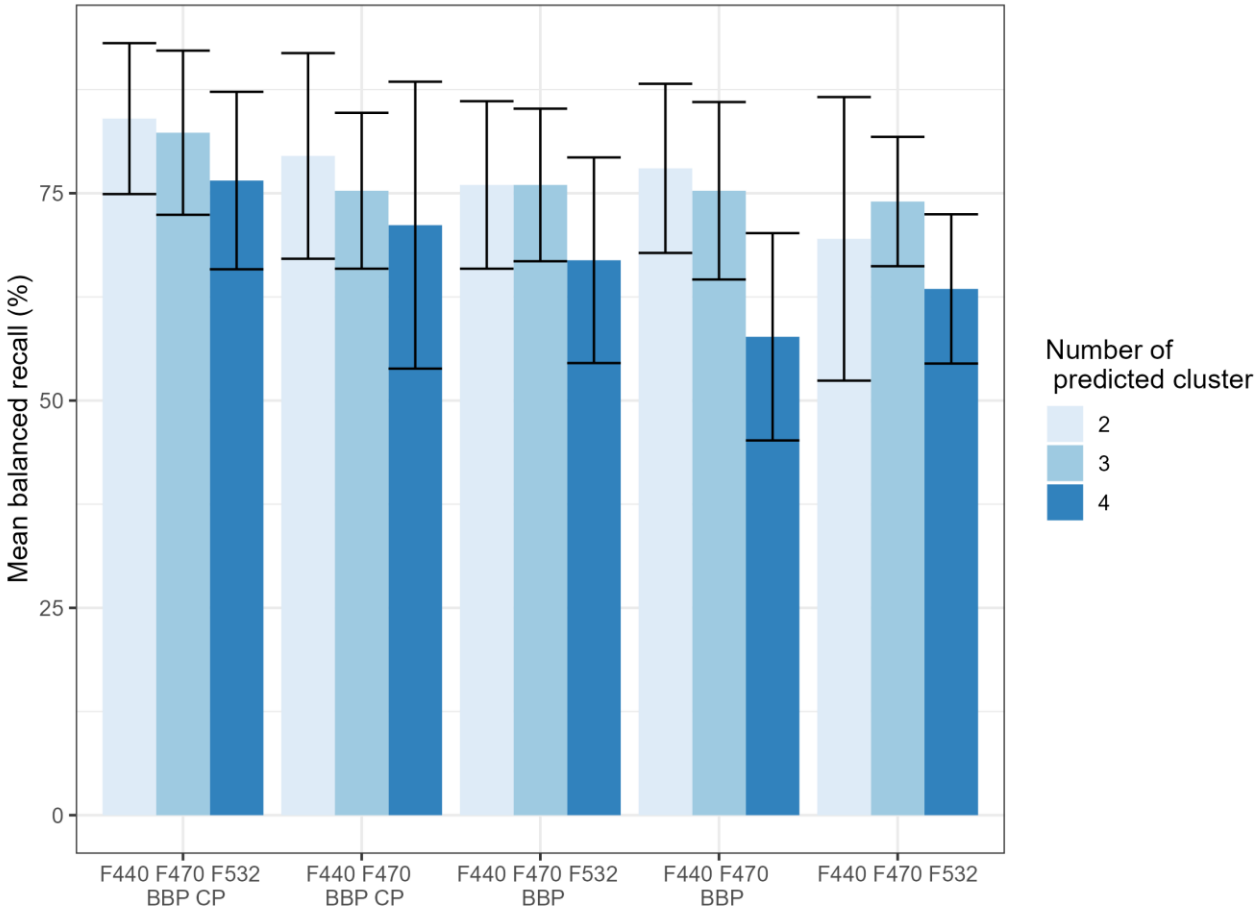

Figure 4.6 : Values of the mean weighted recall resulting from a cross validation with different number of clusters and different sensor configurations.

IV.4. Conclusion and perspectives

Phytoplankton community structure plays a key role in numerous processes that regulate the carbon cycle (i.e. nutrient uptake, CO_2 fixation through photosynthesis efficiency, sinking speed). However, the *in-situ* high frequency measurement tools mostly take account of only Chla concentration, overlooking the community composition. We studied here the possibility to use multispectral fluorometer as a new sensor to enable the estimation of the phytoplankton community taxonomic composition.

Our results from laboratory experimentations show that the multispectral fluorometer signal depend on the observed strain, suggesting that it can be used *in-situ* to estimate the phytoplankton community composition. The one year time series of multispectral fluorescence revealed that the ratio between the three excitation channels depends on phytoplankton community composition. The clustering method reveal four contrasted phytoplankton communities. Those cluster were fairly predicted from our machine learning model based on MSF and optical measurements. We highlighted the possibility to decrease the taxonomic information to establish predictive models with fewer sensors.

V Estimation of particulate organic carbon and its size partition between four plankton groups from BGC-Argo floats

Flavien Petit[1], Julia Uitz[1], Melek Golbol[2], Vincenzo Vellucci[2], Céline Dimier[2], Joséphine Ras[2], David Antoine[1,3], Antoine Poteau[1], Christophe Penkerc'h[2], Hervé Claustre[1].

[1] Centre national de la Recherche Scientifique (CNRS) and Sorbonne Université, Laboratoire d'Océanographie de Villefranche (LOV), Villefranche-sur-Mer, France

[2] Observatoire des Sciences de l'Univers (OSU) and Sorbonne Université, Institut de la Mer de Villefranche (IMEV), Villefranche-sur-Mer, France

[3] Remote Sensing and Satellite Research Group, School of Earth and Planetary Sciences, Curtin University, Perth, WA 6845, Australia

V.1. Introduction

Phytoplankton play a key role in the regulation of the carbon biogeochemical cycle (Falkowski, 1994b). By transforming dissolved inorganic carbon into particulate organic carbon through photosynthesis, they are the keystone of the biological carbon pump (BCP, Buesseler et al., 2007). The BCP involves a series of processes, starting with primary production and eventually leading to the export and storage of carbon in the deep ocean. Many of these processes depend not only on phytoplankton biomass but also on community structure, from the intensity of CO_2 fixation through photosynthesis (e.g. Cermeño et al., 2005; Uitz et al., 2008), to trophic interactions (e.g. Cushing, 1989; Finkel, 2007) and the transfer of carbon to depth (Michaels and Silver, 1988; Buesseler et al., 2007; Guidi et al., 2009; Henson et al., 2012). In a context of global changes where anthropogenic activities modify the carbon cycle (Friedlingstein et al., 2022) and the functioning of the ecosystems (Hoegh-Guldberg and Bruno, 2010; Mayersohn et al., 2022), it is a critical challenge to develop an ability to monitor phytoplankton biomass and community structure on the global scale. Global ocean biogeochemical models (GOBMs) aim to take into account the phytoplankton influence on carbon cycling, but rely on sparse observations of biomass and diversity for their validation. Assessing the biomass and composition of phytoplankton communities at large scales is technically challenging due to the broad size range covered by the organisms (e.g. Finkel et al., 2010) and their strong space-time variability.

While shipborne measurements provide detailed, yet but spatially sparse information on phytoplankton, ocean color satellites and autonomous in-situ bio-optical platforms provide large-scale observations. In particular, the BioGeoChemical-Argo (BGC-Argo) program aims to monitor and understand key biogeochemical processes in the global open ocean based on a network of profiling floats equipped with a suite of physical and biogeochemical sensors (Biogeochemical-Argo Planning Group, 2016; Roemmich et al., 2019; Claustre et al., 2020).

The capacity to extract information on phytoplankton biomass and community composition from BGC-Argo floats is of a great interest to refine estimations of carbon fluxes in the global ocean. BGC-Argo floats estimate phytoplankton biomass through a proxy, the chlorophyll-a (Chla) concentration, derived from *in-situ* fluorescence measurements. However, not only the Chla-to-carbon ratio may vary depending on several factors, light and nutrient availability in particular (Geider, 1987; Dubinsky and Stambler, 2009), but also the fluorescence-to-Chla varies significantly on a broad range of scales (Roesler et al., 2017; Petit et al., 2022; Schallenberg et al., 2022).

A promising alternative is to use optical properties, such as the particulate backscattering (b_{bp}) or beam attenuation (c_p) coefficients, to retrieve the carbon biomass of phytoplankton (e.g. Oubelkheir et al., 2005; Loisel et al., 2011; Koestner et al., 2022), or a combination of b_{bp} and c_p to retrieve information about the plankton community composition (Rembauville et al., 2017; Terrats et al., 2020). While simple models have difficulty accounting for the complex, variable bio-optical relationships, machine learning methods may significantly improve our ability to derive biomass from optical measurements. However, precedent methods are often based on spatially and temporally limited database or retrieve only a limited information on the phytoplankton community composition. Thus, there remains a need to develop methods that quantify the biomass and structure of phytoplankton communities based on the variables measured by BGC-Argo floats.

In the current study we investigate the possibility to use two machine learning regression models to retrieve i) the particulate organic carbon (POC) concentration and ii) the relative contribution of four plankton groups to the POC from BGC-Argo float measurements. All BGC-Argo floats measure temperature, salinity, fluorescence and the b_{bp} coefficient, which are then used as predictors of the POC concentration. In order to go beyond the biomass information and to predict the composition of the plankton assemblage, we not only use the b_{bp} coefficient but also the c_p coefficient. The c_p coefficient is not one of the standards variables (EOVs) of the global BGC-Argo program, as is b_{bp}, but it can be (and frequently is) measured by BGC-Argo floats. This allows to use the b_{bp}-to-c_p ratio which is

known to vary with the composition of the particle pool and phytoplankton assemblage (e.g. Twardowski et al., 2001; Boss et al., 2004; Cetinić et al., 2012).

We defined a standard protocol to construct a dataset of concomitant measurements of POC, relative contribution to POC of four plankton groups and hydrographical as well optical measurements, measured from the same suite of sensors as those mounted on BGC-Argo floats. This protocol has been implemented in the Southern Ocean, North Western and Eastern Mediterranean Sea. The resulting dataset covers contrasted plankton assemblages and tropic regimes, which allows us to develop methods potentially applicable on large to global scales. Ultimately, we apply the proposed ML based methods to a time series acquired by a BGC-Argo float deployed in the Northwestern Mediterranean Sea as a case study. We then discuss the potential of such an approach to infer quantitative information on the biomass and composition of plankton and phytoplankton communities, across a wide range of trophic regimes, through application to BGC-Argo floats deployed in the global open ocean.

V.2. Material and methods

V.2.1. Standard sampling protocol

Based on Rembauville et al. (2017), we defined a standardized sampling protocol for implementation on different oceanographic cruises in order to obtain harmonized interoperable datasets in different regions of the global open ocean. This protocol includes, first, discrete seawater sampling from Niskin bottles mounted on a CTD-rosette device for the determination of the concentration of the (total) particulate organic carbon as well as its partitioning into four different plankton groups, *i.e.* bacteria, pico-, nano- and microphytoplankton (see section II.1.2). Second, we use hydrographic and optical sensors attached to the frame of the CTD-rosette for concomitant high frequency measurements of the particulate backscattering coefficient at 700 nm and particulate beam attenuation coefficient at 660 nm (see section II.1.3), in addition to basic hydrographical properties

(pressure, temperature, and salinity) and Chla concentration determined from *in-vivo* fluorescence. Specifically, we used an SeaBird ECO FLBBCD sensor for measurement the particulate backscattering coefficient and Chla fluorescence, and a SeaBird C-Rover beam transmissometer for measurement the particulate attenuation coeffcient. Temperature and salinity are determined using SBE sensors.

V.2.1.1) Oceanographic field cruises

The sampling protocol was implemented in the course of three different oceanographic missions (Figure 5.1), i.e. the Southern Ocean and Climate (SOCLIM) cruise conducted in the Southern Ocean (SO), the Pelagic Ecosystem Response to dense water formation in the Levant Experiment (PERLE) cruise, and the recurrent monthly cruises at the BOUée pour l'acquiSition d'une Série Optique à Long termE (BOUSSOLE) fixed station, in the Mediterranean Sea.

The SOCLIM cruise took place in the Indian sector of the SO in October 2016 on board of the R/V Marion Dufresne II. The sampling protocol was performed at 11 stations between 35°S and 58.5°S providing 36 data points. The PERLE mission is composed of three cruises; PERLE0, PERLE1 and PERLE2, in the eastern Mediterranean basin between 2018 and 2019, providing 33 data points. The BOUSSOLE mission consists of monthly cruises in the Northwestern Mediterranean Sea at the BOUSSOLE long-term observation site situated at 7°54'E, 43°22'N (Antoine et al., 2008).The protocol was performed over a full seasonal cycle from November 2020 to October 2021, providing a total of 27 samples. The final database is thus composed of 99 samples distributed in contrasted oceanic regimes of the SO and Mediterranean Sea.

V.2.1.2) Plankton organic carbon assessment

For determining the concentration of POC, 2.7 L of seawater were sampled at 3 depths (at the surface, at the level of the deep chlorophyll maximum or DCM, and just below the DCM) and filtered onto pre-calcinated (24 h, 450°C) GF/F filters. Blanks were measured at each station by filtering 2 L of milli-Q water. Filters were then stored into pre-calcinated glass vials and dried in an oven (24 h, 50°C). Back in the laboratory,

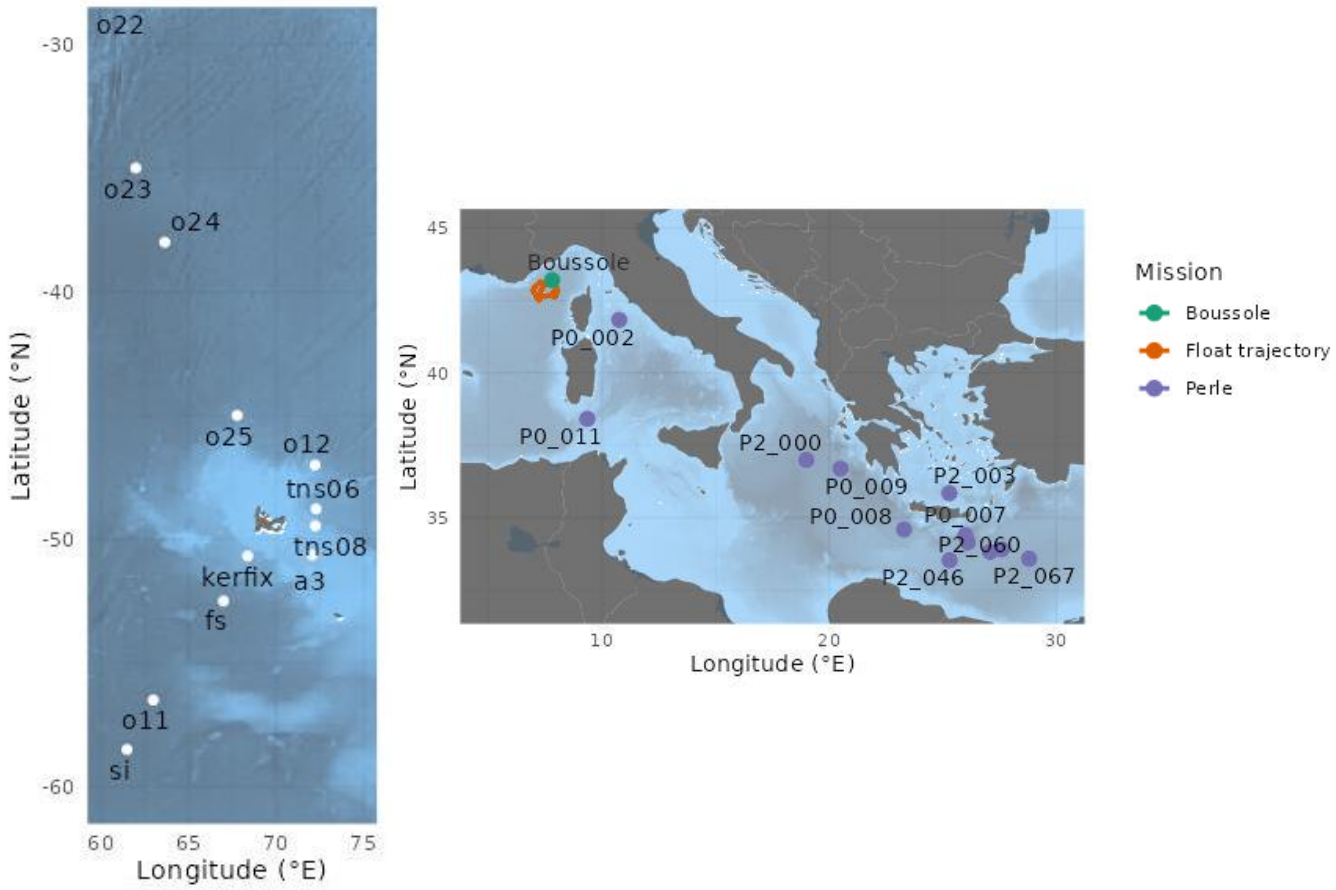

Figure 5.1 : Geographic location of the stations sampled during the three different missions: (a) SOCLIM and (b) BOUSSOLE and PERLE. The trajectory of the BGC-Argo profiling float lovbio064b (WMO 6901496) deployed at the BOUSSOLE site is also shown in (a). The sampling station names are indicated on the map.

samples were fumigated with pure HCl (24 h) to dissolve the carbonate fraction. The analysis of carbon was performed with a CHN analyser (Perkin-Elmer 2400) calibrated with acetanilide.

For flow cytometry analyses, 1.5 mL of seawater were sampled from the Niskin bottles, fixed with 0.2mL of glutaraldehyde (0.5% final concentration), flash-frozen in liquid nitrogen, and stored at −80°C until analysis in the laboratory. The abundance of heterotrophic microbes (including Bacteria and Archaea), pico- and nanoplankton was measured by flow cytometry using a FACSCalibur instrument The biovolume and carbon content of bacteria, picoplankton and nanoplankton were estimated using constant cell volume and cell carbon content by size class from the literature (Table 5.1).

For microplankton abundance and diversity, 100 mL of seawater was sampled in opaque bottles, fixed with acidic lugol (1% final concentration) and stored at 4°C. Back in the laboratory, samples were sedimented into an Utermöhl counting chamber (24 h, dark). Microplankton cells were enumerated and identified to the most precise taxonomic level possible using an inverted microscope with phase contrast, Olympus IX70 at X400 magnification for SOCLIM and PERLE data and NIKON Eclipse TS2 at X200 for BOUSSOLE data. For each sample, 500 to 1,500 cells were enumerated and species were identified following recommendations by Hasle and Syvertsen (1997). Morphometric measurements were made from high-resolution images (Olympus DP71 camera) using the Fiji image processing package (available at http://fiji.sc/Fiji). Biovolume of each microplankton species/group was estimated from morphometric measurements (performed on 20 individuals randomly selected for each species/group) and shape-specific equations (Hillebrand et al., 1999). Carbon content of each microplankton group was calculated using group-specific equations from the literature (Table 5.1). The full list of microplankton groups/species and the associated biovolume is given in the supporting information dataset. The total plankton carbon (Ctot) was defined as the sum of bacteria, pico-, nano- and microplankton carbon biomass:

$$Ctot = Cbact + Cpico + Cnano + Cmicro$$

Table 5.1 : Plankton Groups Considered in This Study and Their Associated Characteristics

Plankton group	Corresponding taxa	Analytical method	Biovolume (μm^3)	Carbon content (pgC)
Bact	Heterotrophic bacteria	Cytometry	0.25[a]	0.015[a]
Pico	Prochlorococcus		0.68[b]	0.029[b]
	Synechococcus		0.86[b]	0.080[b]
	Picoeukaryotes		2.76[b]	0.73[b]
Nano	Nanoplankton		284[c]	15[c]
Micro	Diatom (55 groups)	Optical microscopy	Shape-specific	$C=0.117V^{0.881}$[d]
	Dinoflagellate (14 groups)		Shape-specific	$C=0.760V^{0.819}$[d]
	Ciliate (4 groups)		Shape-specific	$C=0.216V^{0.939}$[d]
	Silicoflagellate (1 group)		3288	$C=0.261V^{0.860}$[d]

[a] Bratbak (1985)

[b] Grob et al. (2007)

[c] Verity et al. (1992)

[d] Menden-Deuer and Lessard (2000)

V.2.1.3) Optical data

Vertical profiles of the chlorophyll a fluorescence (F), particulate backscattering coefficient at 700 nm (bbp) and particulate beam attenuation coefficient at 650 nm (c_p) were acquired using sensors similar to those mounted onto BGC-Argo floats. A SeaBird ECO FLBBCD sensor was mounted on the CTD-rosette frame to measure chlorophyll fluorescence and

b_{bp} coefficient. Similarly, the c_p coefficient was measured using a SeaBird C-Rover transmissometer (25 cm path length).

The fluorescence data were corrected for non-photochemical quenching (NPQ) on daytime profiles following the method of Xing et al. (2012) and converted to chlorophyll a concentration (F_{Chl}, mg m^{-3}) by first applying the factory calibration (dark value and slope) (Schmechtig et al., 2018) and then multiplying by a factor of 0.5 following the recommendation of Roesler et al. (2017). The b_{bp} coefficient (m^{-1}) was calculated as described in Schmechtig et al. (2015) and references therein. Briefly, raw instrument measurements (counts) were transformed into the total volume scattering function (β) at an angle of 124° and wavelength of 700 nm by applying the manufacturer-provided scaling factor and dark count. The b_{bp} coefficient was then calculated as follows :

$$b_{bp} = 2\pi\chi(\beta-\beta sw) \quad (1)$$

where χ is a wavelength-dependent conversion factor (here 1.142) and βsw is the contribution of pure seawater to scattering that depends on temperature and salinity. The c_p coefficient was retrieved from the transmittance measured by the transmissometer. The transmittance data (T) were transformed into c_p as:

$$c_p = 1/l*\ln(T) \quad (2)$$

where l is the sensor pathlength (25 cm) and T derived as:

$$T=(Vs-VD)/(Vw-Vd) \quad (3)$$

where Vs in the instrument signal, Vd is the dark value and Vw is the value in the pure water used for calibration.

The windows of the optical instruments were carefully cleaned prior to each deployment and no drift correction needed to be applied to the shipborne measurements (see section 2.2). Spikes were removed from the Chla fluorescence and b_{bp} signals by applying a low-pass filter that consists of a 5-point running median followed by a 7-point running mean

(Briggs et al., 2011). Finally, the CTD and optical data were averaged at the depth corresponding to the discrete seawater sampling for plankton diversity determinations (± 1 m).

V.2.2. BGC-Argo timeseries

A BGC-Argo profiling float (World Meteorological Organisation identifier: 6901496) has been selected for prediction of POC and partitioning into plankton group-specific POC based on the machine learning methods presented below (Section II.3). This float was equipped with temperature, salinity, b_{bp}, c_p and fluorescence sensors (Seabird Electronic ECO-Puk triplet, Seabird C-Rover). This float cycled from March 2014 to April 2015 in the Northwestern Mediterranean Sea (BOUSSOLE region). The float data were downloaded from the Argo database accessible at ftp://ftp.ifremer.fr/ifremer/argo/. The data were processed following the standard BGC-Argo protocol, as described in Section II.1.3. In addition, c_p data were corrected from sensor drift by subtracting a median c_p value, used as an "offset", computed from the c_p values acquired between 300 m and the maximum sampled depth (Barbieux et al., 2022).

V.2.3. Machine learning-based models

Two different machine learning models have been used in order to address the two main objectives of this work. First, the estimation of the total POC concentration from the standard set of BGC-Argo sensors has been investigated through the use of an extreme gradient boosting regressor (XGBoost) algorithm. Second, the prediction of the partition of the total POC into four plankton classes based on the standard BGC-Argo sensor

package in addition to a transmissometer, has been investigated using a Partial Least Square analysis (PLS) method in a manner analogous to Rembauville et al. (2017).

V.2.3.1) XGBoost model for predicting the total POC stock

XGBoost (XGB), for Extreme Gradient Boosting, is an effective tree based ensemble learning algorithm (Chen and Guestrin, 2016). It builds several models sequentially where each new model attempts to correct errors from the previous one. XGBoost uses the gradient descent algorithm to minimize the loss function of the model. The descriptors are temperature, salinity, pressure, b_{bp}, Chla concentration estimated from fluorescence (F_{Chl}) and the b_{bp}/F_{Chl} ratio. The different hyperparameters, which are the parameters that define the model architecture, are optimized through a gridsearch procedure. The gridsearch algorithm calculates the model performances for a combination of hyperparameters values. The resulting model is composed of 700 trees of a depth of 4, with a learning rate of 0.2. The XGBoost model is performed with the XGBoost Python library version 1.6.0.

V.2.3.2) Partial Least Square regression for predicting the POC content of the plankton groups

The Partial Least Square regression (PLS) is a regression method that is particularly well suited to predict multiple outputs from an ensemble of predictors that are significantly correlated. Here we predicte the relative carbon content associated with each of four plankton groups ($C_{group,}$ %) based on temperature, salinity, pressure, b_{bp}, FChl, c_p, and the ratios of FChl/b_{bp}, FChl/c_p and b_{bp}/c_p. This method has been developed initially for the SO by Rembauville et al. (2017), providing promising results and potential for extension to other oceanic regions. The covariance between hydrographical and optical data that are used as predictors and phytoplankton taxa abundances has been studied with a redundance analysis (RDA). This analysis allows to explore the multiple correlations between a matrix Y of response variables and a matrix X of explanatory variables. We have

used here the taxa cells counts from microscopy and flow cytometry analysis as the explanatory variables and the predictors used in the PLS as the response variables. The PLS model is performed with the Scikit-learn Python library version 1.2.0 whilst RDA is performed with the Vegan R package version 2.6.

V.3.3. Evaluation of the performances of the models

The performances of the machine learning models are evaluated by measuring a set of metrics through a procedure of cross validations. A cross validation consists of randomly subsampling the dataset into a learning subset and a testing subset, in order to train the model on the learning set and test the model on the testing set. This procedure enables to assess the capacity of the model to predict outcomes from data that have not been used for the training even on a relatively small dataset. This aims to avoid overfitting. A classical performance metrics in regression models is the determination coefficient (R^2) defined as follows:

$$R^2 = \frac{\sum_{i=1}^{n} (yi - \hat{y}i)^2}{\sum_{i=1}^{n} (yi - \bar{y})^2} \quad (4)$$

where n is the total number of observations, yi is the observation number i, ŷi is the prediction corresponding to observation number I and $\bar{y}$ is the mean of the measures. This metrics allows to determine the strength of the correlation between the predictions and the actual observations. We further use the Root Mean Square Error (RMSE) as an indicator of the accuracy of the predictions. The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}i - \bar{y})^2}{n}} \quad (5)$$

This metric is expressed in the same units as the predicted variables. The mean absolute percentage error (MAPE) is also used and defined as follows:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}i - \bar{y}}{yi} \right| \quad (6)$$

The MAPE expresses the accuracy of the regression as a percentage of the absolute value of the predicted variables.

All data processing is performed with the R software version 4.2.2. All machine learning models are executed with Python software version 3.9.

V.3. Results and discussion

V.3.1. Prediction of the POC stock

We investigated the possibility to predict the stock of the particulate organic carbon from the standard BGC-Argo variables.The XGBoost algorithm applied to this dataset yields a POC predictive model with a R^2 of 0.72 (+/- 0.27), a RMSE of 15 (+/- 7) mg m^{-3}, and a MAPE of 34 (+/- 12)% (Figure 5.2). The MAPE score indicates that, in similar oceanic conditions as those represented in the learning dataset, the POC concentration can be predicted with an expected accuracy of 66%. The learning dataset covers a large trophic gradient, from the seasonal bloom encountered in the naturally iron-enriched waters of the Kerguelen Plateau or in the Northwestern Mediterranean Sea, to the oligotrophic conditions of the subtropical Indian Ocean or of the Eastern Mediterranean Sea. Hence we can expect that the performance of the model would be similar in a wide range of trophic conditions on

the global scale. Moreover, the diversity of the conditions found in the Mediterranean Sea is particularly well represented in the dataset. We should thus be able to use the proposed predictive model to derive a timeseries of POC concentration based on measurements from a BGC-Argo float deployed in the Northwestern Mediterranean Sea.

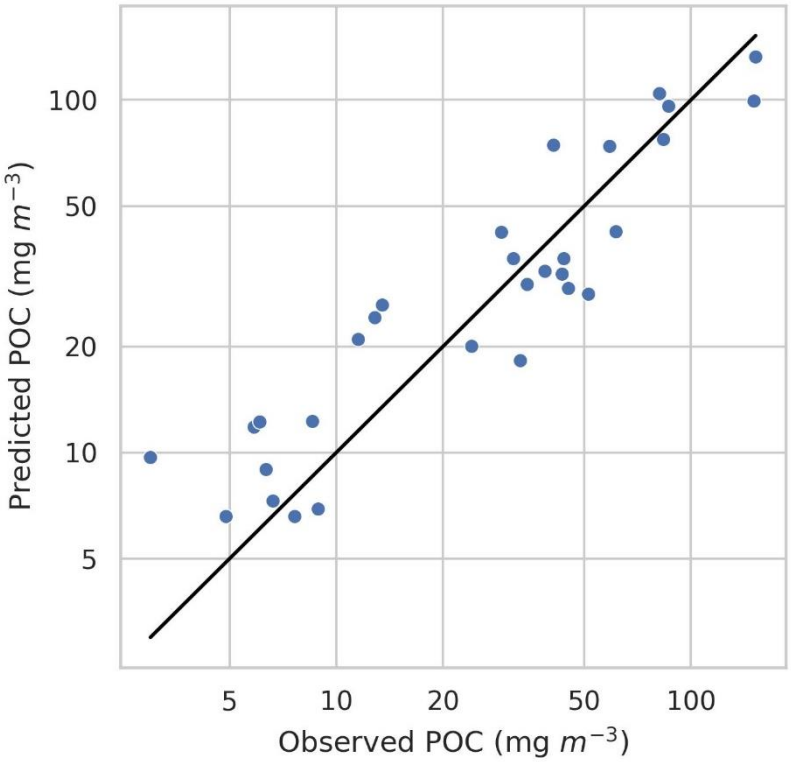

Figure 5.2 : Comparison of predicted versus measured POC concentrations. The POC predictions are obtained using the XGBoost algorithm applied to 30% of the dataset not used for training.

V.3.2. Partition of the particulate organic carbon in the planktonic community

We investigate the possibility to use a PLS regression model to retrieve the relative contribution of four plankton group to the total particulate organic carbon content, based on optical and hydrological data similar to those measured by BGC-Argo floats. We define the four plankton groups as follows: heterotrophic bacteria, picophytoplankton (0.2-2 µm), nanophytoplankton (2-20 µm) and microphytoplankton (20-200 µm). The same database as the one used to develop the XGBoost model, in addition to data from the C-Rover transmissometer (c_p), is used to test this model.

The three regions show a distinct distribution of the carbon among the three plankton groups (Figure 5.3). The SOCLIM dataset is characterized by a dominant contribution of microphytoplankton to POC that reaches more than 50% at the a32 and a31 stations. In contrast, the sampling stations O23 and O24, that are situated north of the Polar Front, display a dominant contribution of nanophytoplankton to Ctot. In the BOUSSOLE dataset, the POC content appears to be strongly dominated by nanophytoplankton, whose contribution reaches up to 80%, especially between the monthly cruises b225 and b228, that took place in late winter and spring. The cruises b229 to b235 also show a dominance of nanophytoplankton (~75% of Ctot) yet with a significant contribution of bacteria (~20%). The PERLE dataset shows a mixed plankton assemblage, with a significant contribution of bacteria and nanophytoplankton to Ctot at most of the sampling stations. The contribution of microphytoplankton to Ctot in the PERLE 1 cruise data is equally important as that of nanophytoplankton. The resulting global dataset is thus composed of contrasted plankton assemblages that may represent, to a certain extent, the diversity of plankton assemblage observed on the global scale.

We use this dataset to evaluate the possibility to extend the method developed by Rembauville et al. (2017) for the SO to other oceanic provinces. We performed a cross-validation of the PLS model using the descriptors presented in Section V.II.3.2 to predict the contribution of the four plankton groups to Ctot. The cross-validation process was applied to each of the individual cruise datasets as well as to the combined datasets. The resulting RMSE of the predictions of the Ctot contribution of each plankton group for each

dataset is presented in Figure 5.4. Higher RMSE values are observed for bacteria and nanophytoplankton, with values comprised between 10 and 12.5%, whereas picophytoplankton shows the lowest RMSE (<5%). The microphytoplankton group is characterized by a significant variability in the RMSE among the three cruise datasets, which reflects the variability in the relative contribution to Ctot of this group as seen in Figure 5.3. Indeed, the lower RMSE associated with the BOUSSOLE dataset results from a lower average contribution of microphytoplankton to Ctot, while it is larger in the other datasets.

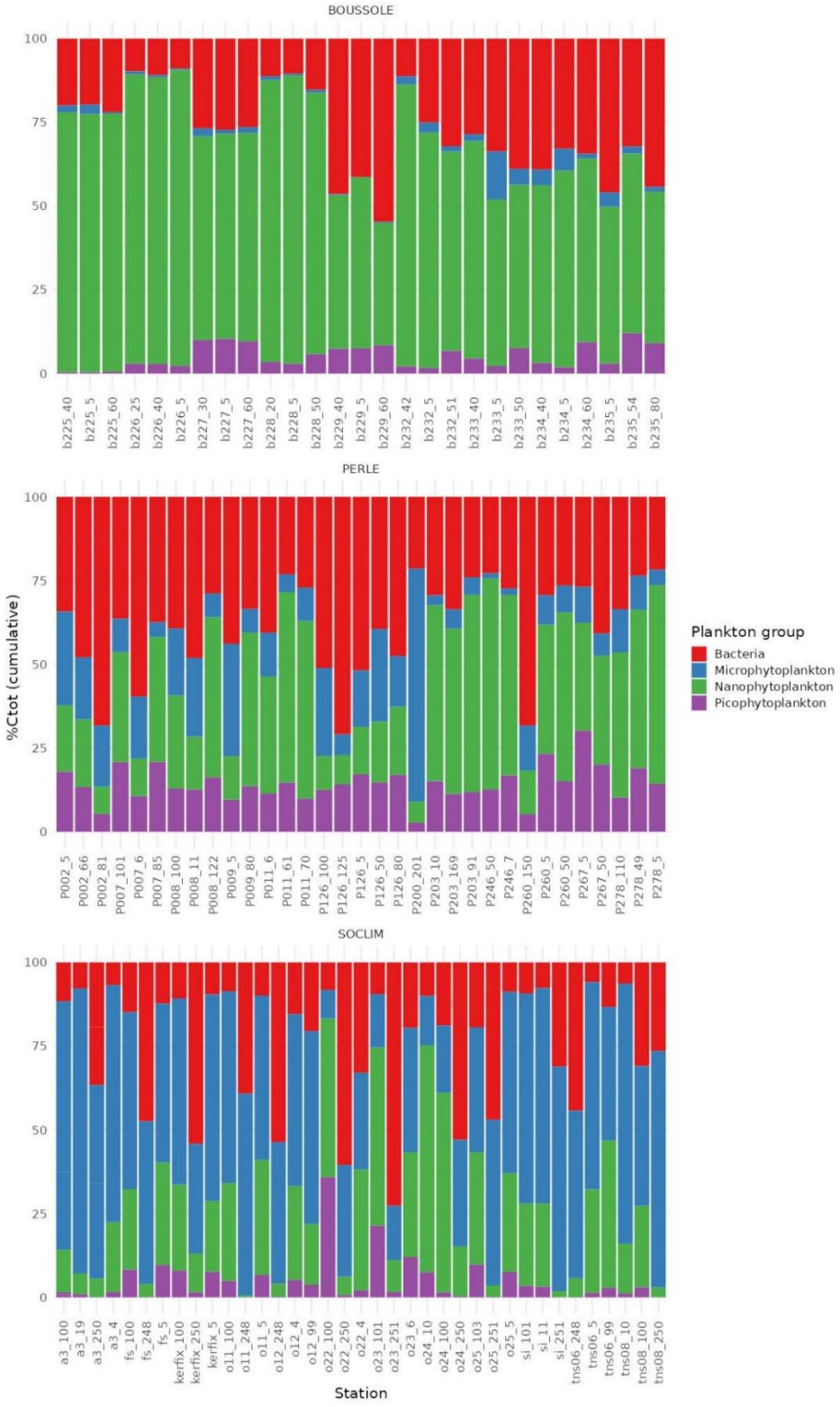

Figure 5.3 : Relative contribution of each plankton group to the total particulate organic carbon, represented as cumulative percentage, for the three missions represented in the dataset and each sampling station. The sampling station is indicated on the X axis with the station name followed by the depth of sampling separated by underscore symbole.

In order to understand the variability in the model performances in predicting the relative carbon contribution of the four plankton groups, we consider the covariations between the cellular abundance of the different plankton taxa and the selected predictors using an RDA (Figure 5.5). The first Principal Component (PC1) of the RDA accounts for 29% of the dataset variance, and opposes the cellular abundance of phototrophic organisms to that of heterotrophic bacteria. The abundance of bacteria is anti-correlated, on the one hand, to temperature and salinity because they occur in a deeper layer than phytoplankton organisms and, on the other hand, to the F_{Chla}/b_{bp} ratio because they do not synthetise chlorophyll.

The second Principal Component (PC2) represents 20% of the variance of the dataset. While all picophytoplankton organisms, *Synechococcus* and *Prochlorococcus*, are associated with negative eigenvalues on the PC1, microphytoplankton taxa are associated with both negative and positive eigenvalues of PC2, depending on the taxa. The cellular abundance of diatoms and silicoflagellates is anti-correlated with that of ciliates and dinoflagellates, which likely reflects distinct microphytoplankton assemblages sampled in the PERLE and BOUSSOLE regions. The PERLE dataset is indeed characterized by a large abundance of dinoflagellates while, in the BOUSSOLE and SOCLIM dataset, the microphytoplankton community shows a dominance of diatoms and silicoflagellates. Interestingly, the optical variables are associated with large PC2 eigenvalues, with the b_{bp} and c_p values strongly correlated with the abundance of diatoms and silicoflagellates. This may be explained by the presence of a silica frustule that modifies the light scattering properties of these organisms (Sun et al., 2016). Previous studies have used the influence of the POC-to-Chla ratio and external skeleton composition that both affect scattering properties to define an index of diversity based on the b_{bp}-to-fluorescence ratio (Cetinić et al., 2015; Terrats et al., 2020). Dinoflagellates and ciliates do not have the same optical properties as diatoms and coccolithophores, leading to discrepancies within the microphytoplankton group and to lower performances of the model to predict the Ctot contribution of microphytoplankton when the full (combined) dataset is used for the algorithm training (Figure 5.4).

The global fit of the model is presented in Figure 5.6 which also shows the R^2 values associated with the cross validations. The best R^2 are obtained when the BOUSSOLE or SOCLIM dataset are used for the algorithm training, while the model evaluation on the PERLE dataset is characterized by a lower R^2 value. These low values may be due to a much more limited variability of the plankton assemblage in the PERLE data. The PERLE dataset is overall characterized by a large abundance of dinoflagellates (Figure 5.5) which have different optical properties than other taxa belonging with the microphytoplankton group. Moreover, the cross-validation procedure leads to a random sub-sampling of the dataset, which may, in the case of a small number of observations, lead to an under-representation of a plankton assemblage from the learning or testing subset. This type of methods may be subject to overfitting, and learning dataset must be large and contrasted enough to make an objective prediction. However, the robust performances of the model for predicting the plankton group-specific contribution to Ctot in the full dataset indicate that this method could be used with a large dataset to predict plankton assemblage in a large variety of oceanic regimes. Given the overall robust performances of the model when based on the Northwestern Mediterranean (BOUSSOLE) dataset, we are confident in its application to the timeseries collected by float # 6901496 and the derived predictions of the plankton assemblage.

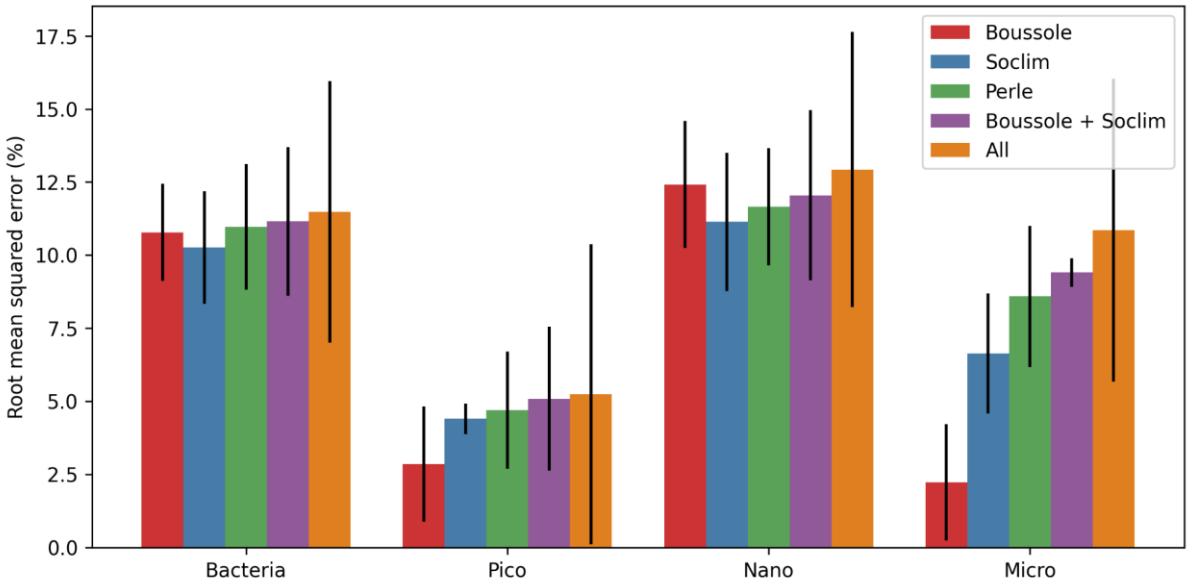

Figure 5.4 : Root mean square error (in %) of the predicted contribution of each plankton group to POC for the different dataset cross-validation. The vertical black lines represent the standard deviation of the different folds of the cross-validation procedure.

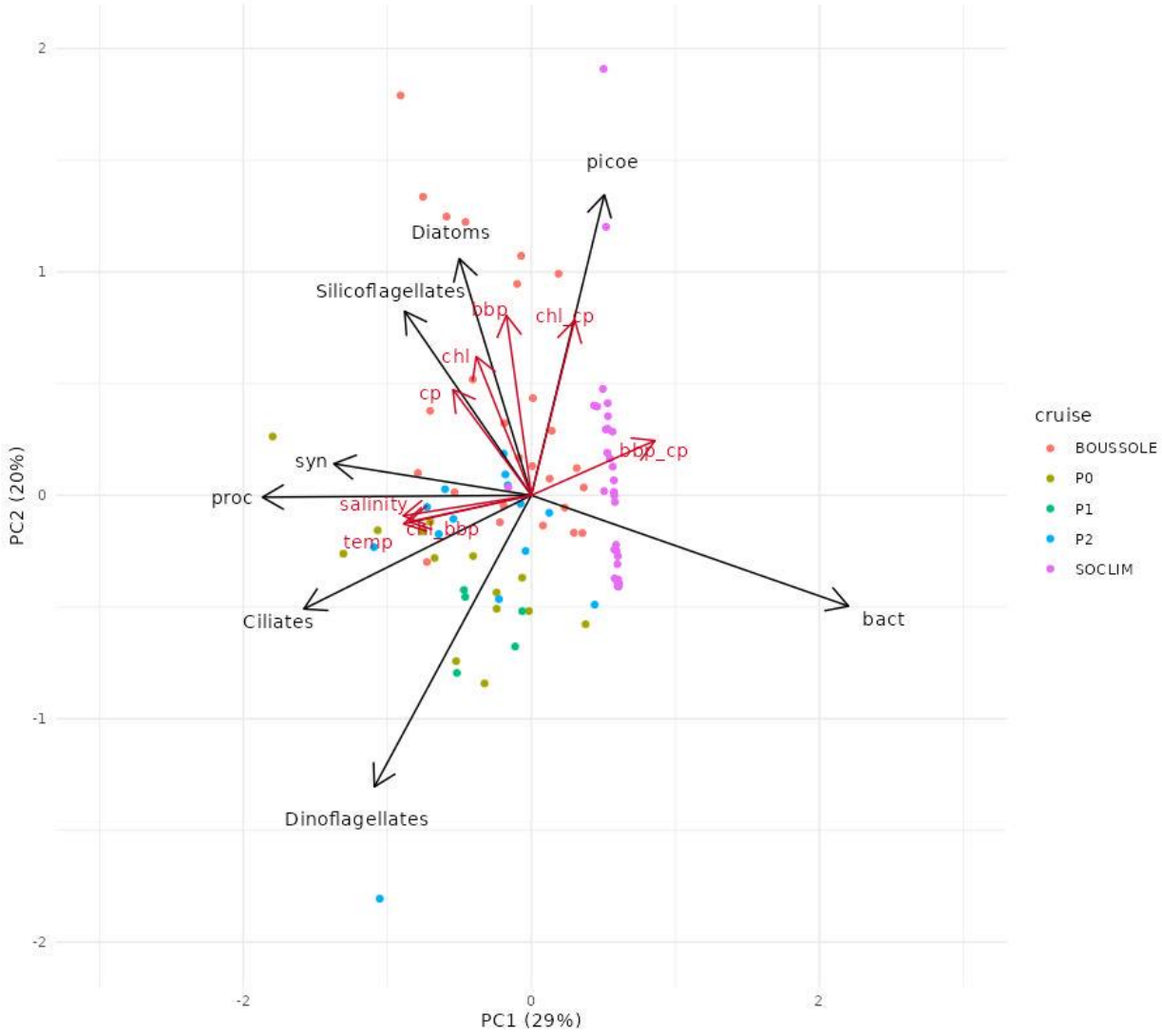

Figure 5.5 : RDA analysis of the plankton diversity based on the microscopy and flow cytometry cell counts for the global dataset (i.e. including the BOUSSOLE, PERLE and SOCLIM data). The hydrographical and optical data are projected as environmental data. P0, P1 and P2 respectively stand for PERLE 0, PERLE 1 and PERLE 2 cruises. Syn stands for Synechococcus, picoe for picoeukaryote, proc for Prochlorococcus and bact for bacteria and to lower performances of the model to predict the Ctot contribution of microphytoplankton when the full (combined) dataset is used for the algorithm training (Figure 5.).

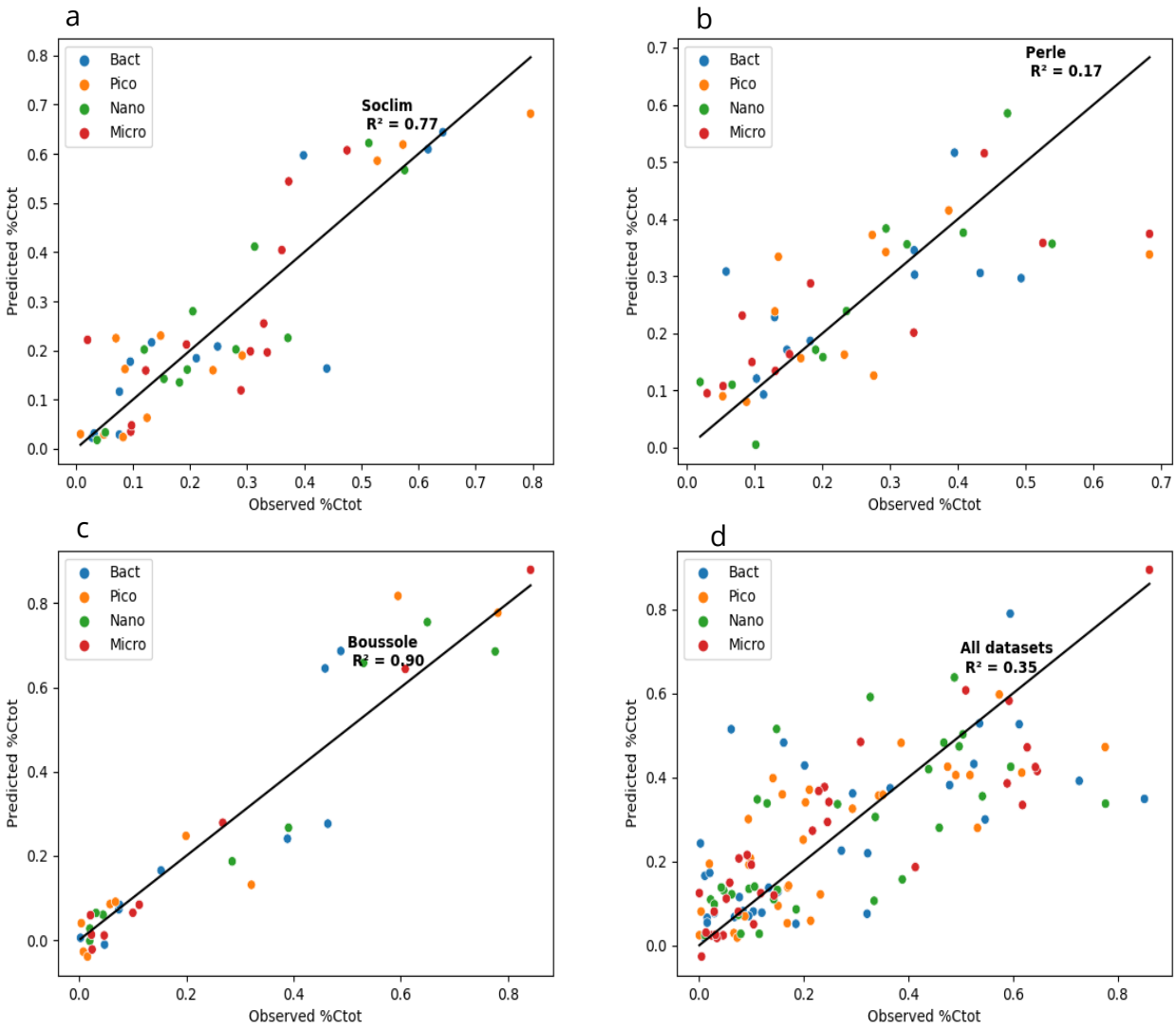

Figure 5.6 : Correlation between the predicted and the observed relative contribution to Ctot/POC/Carbon of the four plankton groups obtained using the partial least square regression model applied to the different dataset : (a) SOCLIM, (b) PERLE, (c) BOUSSOLE, (d) Combination of the three campagne. Bact stand for bacteria, pico for picophytoplankton, nano for nanophytoplankton and micro for microphytoplankton. The black line represents a 1:1 regression.

V.3.3. Time series of the stock of particulate organic carbon and its partition between four plankton groups derived from BGC-Argo float measurements in the Mediterranean Sea

In this section we make use of the XGBoost model to predict the POC concentration based on measurements from the BGC-Argo float (WMO #6901496) deployed in the BOUSSOLE area (Figure 5.7). Over the float timeseries, we observe two different events of mixed layer deepening followed by an increase in F_{Chla}. The deepening of the mixed layer may have led to an injection of nutrients into the upper oceanic layer, to the benefit of phytoplankton. This hypothesis would explain the observed increase in the Chla concentration. The rapid consumption of nutrients in surface waters associated with a rapid restratification of the water column is followed by the formation of a DCM that that tends to deepen until July, after which it disappears. This Chla dynamics is coherent with the seasonality documented in the Northwestern Mediterranean Sea (D'Ortenzio et al., 2005; Lavigne et al., 2015; Barbieux et al., 2019, 2022). Overall the predicted POC concentration follows the temporal Chla dynamics, which suggests that, on a seasonal scale, the F_{Chla} can be used as a proxy of phytoplankton biomass. Interestingly, we also note that the DCM is accompanied by a deep POC maximum, indicating that the DCM is actually a deep biomass maximum (DBM; Barbieux et al., 2019; Mignot et al. 2014). However the relationship between the Chla and POC concentrations also presents small variations depending on the photoacclimation status and composition of phytoplankton communities (Dubinsky and Stambler, 2009; Brunet et al., 2011). The Chla-to-POC ratio increases from summer to fall below the mixed layer, reaching the maximum value of the timeseries, 0.05. This may be associated with the presence of a phytoplankton assemblage adapted to lower light conditions but close to the nutricline (Staehr et al., 2002; Graff et al., 2016). Conversely, in surface waters, the Chla-to-POC ratio is low (~0.01), which indicates a lower intracellular Chla content consistently with more favorable light conditions. Between May and June, at a depth of ~50 m, the Chla/POC ratio is low despite a high POC concentration, thus indicative of high carbon biomass but low Chla concentrations or a change of phytoplankton community composition.

We now use the PLS regression method to infer from the BGC-argo timeseries changes in the plankton assemblage associated with the dynamics in POC (Figure 5.8). The resulting timeseries shows a dominance of nanophytoplankton in the upper layer of the water column and of bacteria at depth. The maximum values of the POC concentrations, observed in May at ~40 m, coincide with a contribution to Ctot of nanophytoplankton close to 100%. This reflects the variability observed in the training Northwestern Mediterranean dataset. Interestingly, the increase in the F_{Chla}/POC ratio below the MLD in summer is associated with an increase in the Ctot contribution of picophytoplankton, close to 15%. One caveat of the method is that it can predict values of the relative contribution to Ctot that are negative or above 100%. This may happen when the optical and hydrographical conditions of the predicted data are not well represented in the training dataset. Here, 11% of the predicted values were either inferior to 0% or superior to 100%.

The observed patterns suggest that the proposed algorithms applied to BGC-Argo float data allows, not only to predict the POC concentration, but also the contribution of four different plankton groups to the POC stock.

The combination of both the XGBoost and PLS methods would enable to estimate plankton group specific carbon stock and fluxes (Boyd and Newton, 1999; Mouw et al., 2016). Such information derived from BGC-Argo floats measurements would be of a great interest to better assess the carbon cycle in the global ocean. The precision of the predicted information appears to be dependent on the number of data as well as the range of optical and hydrographical environments associated to plankton assemblage represented in the learning dataset. In view of developing such method for future applications to the global BGC-Argo network, we recommend the acquisition of an in-situ database of interoperable optical, hydrographical and plankton diversity measurements similar to the datasets collected and utilized in the present study. Field sampling should include, at a minimum, one sample a at the surface, one at the depth of the chlorophyll maximum and one below the euphotic zone in order to account for the strong vertical variability often encountered in the distribution of open ocean phytoplankton and bacteria communities. We also recommend the optical and hydrographical data be

processed following the BGC-Argo guide for good practices (Bittig et al., 2019) and the plankton group-specific carbon biomass be assessed as described in Section II.1.2, following Rembauville et al. (2017).

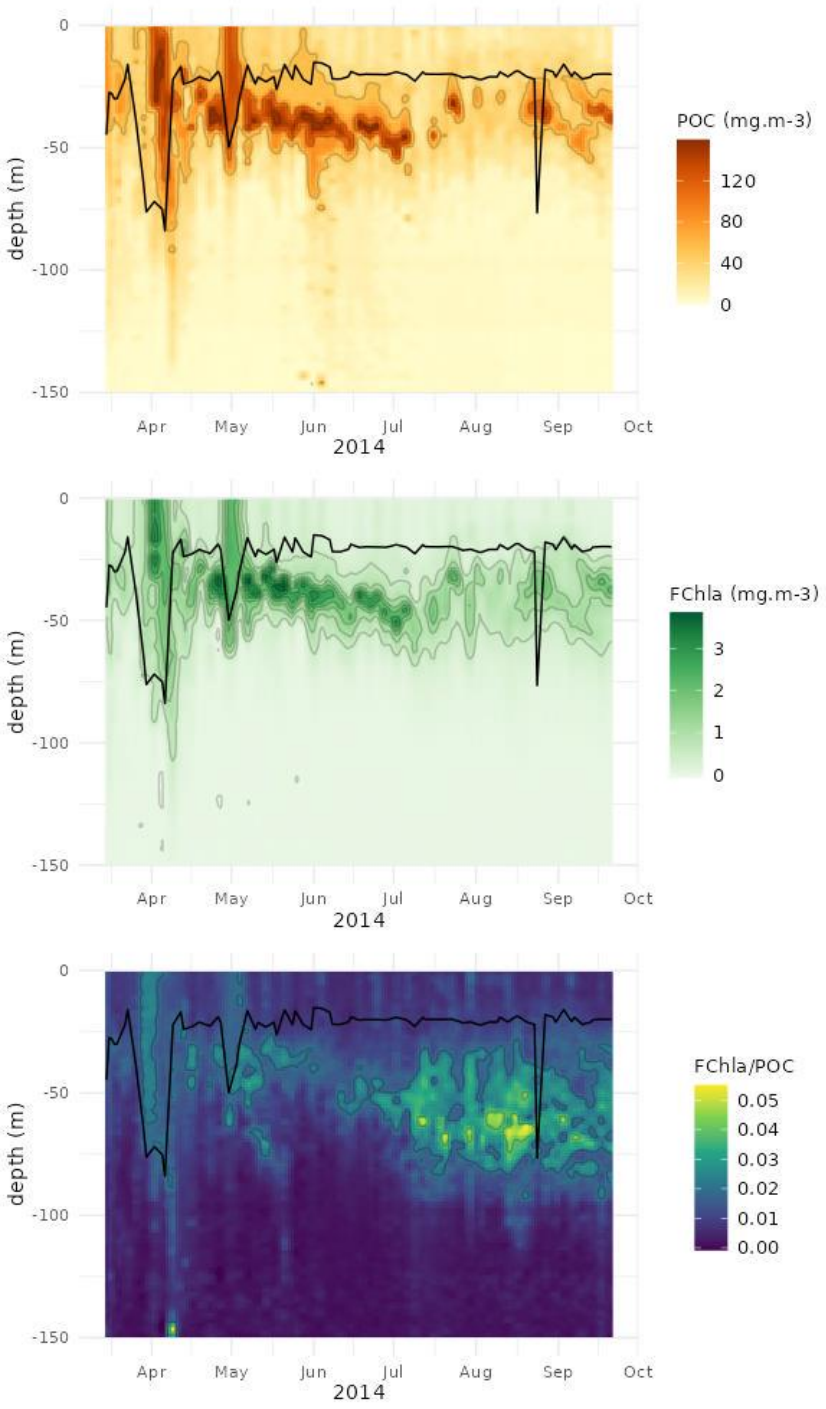

Figure 5.7 : Time series of (a) the POC concentration (mg m^{-3}) predicted with the XGBoost model applied to the BGC-Argo profiling float lovbio064b data, (b) the FChl measured by the BGC-Argo float, and (c) the ratio F_{Chla}/POC calculated from predicted POC and measured FChl.) The bold black line corresponds to the mixed layer depth.

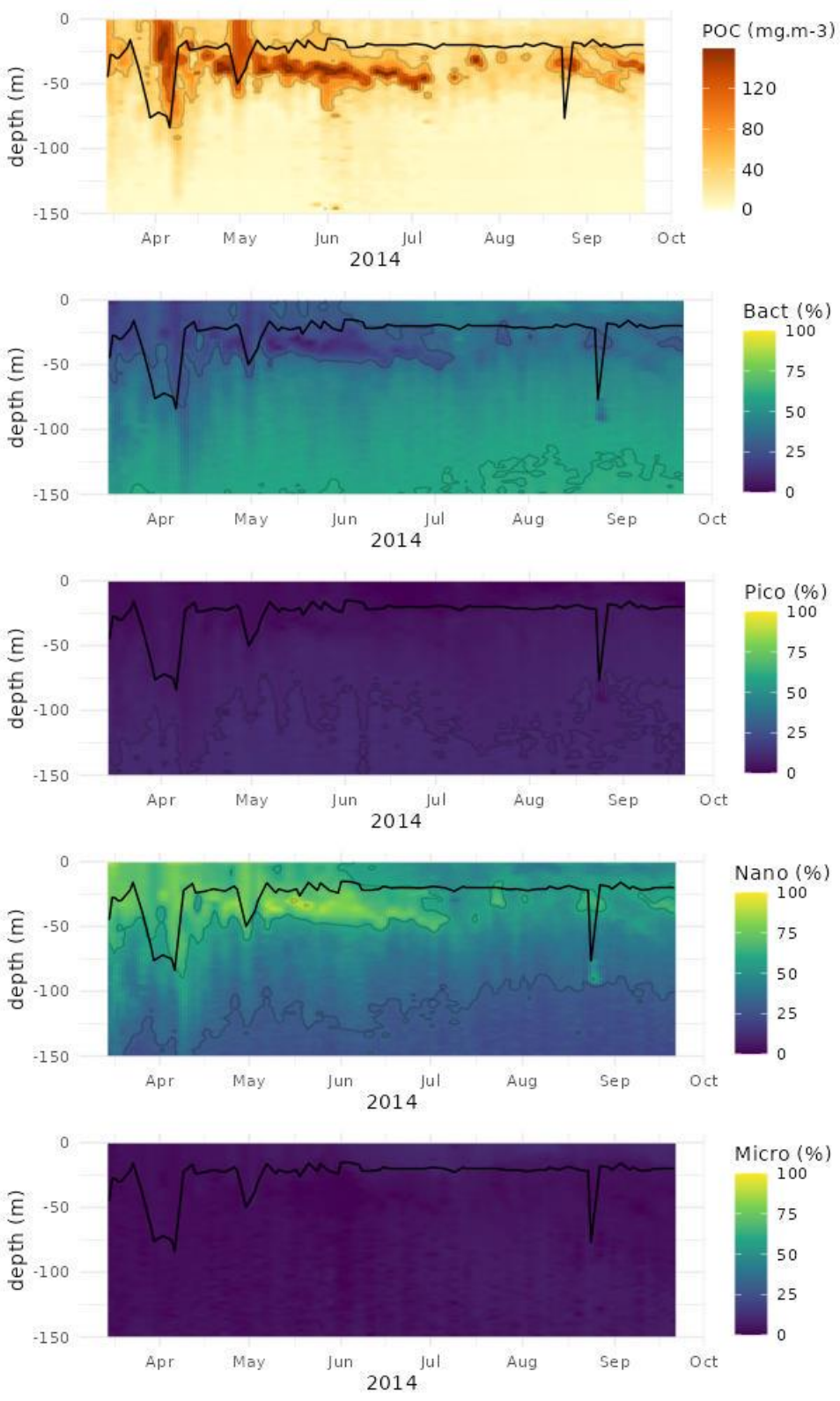

Figure 5.8 : Time series of (a) the POC concentrations predicted with XGBoost model, and (b–d the)relative contribution of the four plankton groups to Ctot predicted with the PLS model, applied to the BGC-Argo float lovbio064b data. Here the PLS model was trained using the whole database, i.e. comprising the data from the BOUSSOLE, PERLE and SOCLIM missions. The black bold line represents the mixed layer depth.

V.4. Conclusion

The present study uses two different machine learning based methods to estimate POC concentration and its partition between four plankton groups. The POC prediction uses temperature, salinity, in-situ fluorescence and bbp data as predictive variables, the standard BGC-Argo measurements. The prediction of plankton assemblage uses the same variable plus the cp to enable particle structure determination. Both methods were evaluated on three different cruises, located in Southern Ocean, northwestern Meditteranean sea and eastern Mediterranean sea. The POC concentration prediction showed good results on the whole dataset, indicating a possible use on large scale with the help of BGC-Argo data. The plankton assemblage prediction also presented promising results on the different dataset. We note a sensitivity to the representativity of optical environment and plankton assemblages represented in the dataset.

The predictive methods were used to assess POC concentration and plankton assemblage of a BGC-Argo float timeseries situated in the northwestern Mediterranean sea. We were able to retrieve the temporal and vertical POC dynamics as well as plankton assemblage succession. The combination of these two methods presented in these paper appear as promising tools to observe phytoplankton community from BGC-Argo floats on the global scale.

VI General conclusion and perspectives

VI.1. Main results

Decades of oceanographic studies have demonstrated the importance of phytoplankton biomass and community composition on marine biogeochemical cycles (Falkowski, 1994b; Morel and Price, 2003; Litchman et al., 2015) and functioning of ecosystems (Cushing, 1989; Field et al., 1998; Finkel, 2007). Yet, observing phytoplankton at global, temporal, and vertical scales remains a challenge. In this thesis, we developed new methods to estimate phytoplankton biomass and community composition from BGC-Argo profiling floats. The main results of this work are summarized in the following section.

Estimation of the phytoplankton biomass

In-situ fluorescence is a very convenient tool for assessing the Chla concentration on large scales. It is widely used for its ease of application and low deployment cost. It is a key variable in the BGC-Argo program, in which each profiling float is equipped with a fluorometer (Claustre et al., 2020). However, the relationship between the *in-situ* fluorescence signal and the Chla concentration has been shown to vary considerably in the global ocean, depending on the bioregion (Roesler et al., 2017; Schallenberg et al., 2022a). Because oceanic bioregions are characterized by specific environmental conditions (the light-nutrient regime in particular), with phytoplankton communities responding to these conditions, we hypothesized that the composition of the phytoplankton communities influences the fluorescence signal on a regional scale. Three different databases of concurrent *in-situ* fluorescence and HPLC pigment measurements were analyzed to assess the influence of changes in phytoplankton community composition on the fluorescence signal regionally as well on the vertical and seasonal scales (section III).

Phytoplankton community composition appears to significantly influence the *in-situ* fluorescence signal, and ultimately its conversion to Chla concentration at all considered

scales. We find a key influence of the size structure, where communities with a dominance of microphytoplankton appear to have a higher fluorescence-to-Chla ratio. Conversely, picophytoplankton-dominated communities have a lower fluorescence-to-Chla ratio, in part due to their pigment composition. Indeed, the presence of zeaxanthin, a non-photosynthetic pigment is often present in large concentration in picophytoplankton-dominated communities, and it is assumed that the energy absorbed by this pigment is directly re-emitted as heat so that the fluorescence de-excitation pathway is less important in such communities (Bidigare et al., 1989). These results could largely explain the overestimation of the Chla concentration from *in-situ* fluorescence in meso- to eutrophic environments, such as the iron-enriched waters of the Kerguelen Plateau or the Northwestern Mediterranean spring bloom which show large contributions of microphytoplankton (diatoms) to the algal assemblage, and the Chla underestimation in oligotrophic conditions, such as the Mediterranean summer surface waters, typically dominated by picophytoplankton.

We investigated in section V the use of machine learning algorithms to estimate the POC stock and the contribution of phytoplankton to this stock, in order to possibly complement the fluorescence-based determination of the phytoplankton biomass. The combination of the XGBoost and PLS regression methods provided promising results on the estimation of phytoplankton biomass for applications to BGC-Argo float sensors. The results showed robust performances of the machine learning based models in different contrasted open ocean environments, indicating the possibility of applying the methods on the global scale. The application of the proposed methods to a BGC-Argo float deployed in the Northwestern Mediterranean Sea clearly demonstrates the feasibility to monitor the POC dynamics within the entire water column and throughout the whole timeseries. Moreover, the combined use of modelled POC and Chla estimated from fluorescence allowed us to observe different mechanisms that influence the carbon-to-Chla ratio such as photoacclimation of phytoplankton cells to changing light conditions over the seasonal cycle.

Estimation of phytoplankton community composition

The estimation of phytoplankton community composition from BGC-Argo float is a critical challenge that has led to several studies (e.g. Briggs et al., 2013; Cetinić et al., 2015; Sauzède et al., 2015; Rembauville et al., 2017). All the methods proposed so far rely on the measurements of seawater optical properties; some of them allow a quantification of the size structure of the phytoplankton community (Sauzède et al. 2015; Rembauville et al. 2017). Although these methods provide promising results for the estimation of phytoplankton community structure from BGC-Argo floats, they are either restricted to a specific oceanic region or provide information limited to the phytoplankton taxonomic composition. In this thesis we evaluated two different approaches to estimate phytoplankton community structure.

The first approach makes use of *in-situ* multispectral fluorescence (section IV). It is based on the difference in light absorption by different accessory pigments, used as biomarkers of specific phytoplanktonic taxa (Escoffier et al., 2015). It has already been used in freshwater environments and also showed promising results in the open ocean (Proctor and Roesler, 2010; Thibodeau et al., 2014b). We therefore studied the response of a 3-channel multispectral fluorometer based on laboratory experiments with ten phytoplankton strains grow in culture. The same multispectral fluorometer has also been deployed at sea over one year to study the response of multispectral fluorescence to phytoplankton seasonal succession. This work was conducted with the aim to propose a method for predicting the composition of phytoplankton communities based on a fluorometer that could easily be mounted on a BGC-Argo proflíling float.

In the laboratory, the multichannel fluorometer responded very different to the different selected phytoplankton taxa. It thus appears possible to distinguish different phytoplanktonic groups according to the ratios of the fluorescence signal emitted in response to excitation in the three different channels. This indicates that the fluorometer is quite sensitive to phytoplankton community composition. The results of the field deployment of the multichannel fluorometer indicate a less marked differentiation of the different communities. This can be explained by the composition of natural phytoplankton assemblages characterized my mixed populations, in comparison with the mono-specific cultures used for laboratory experiments. However, by combining *in-situ*

multispectral fluorescence with optical and hydrographic measurements, the use of a machine learning model, in the present work a histogram gradient boosting, allowed us to predict four different mixed phytoplankton communities. Each community was dominated by a certain taxonomic groups but still is composed of several groups. Two of the four communities were dominated by picophytoplankton organisms, one by *Synechococcus*, the other by *Prochlorococcus*. The two remaining communities were dominated by nanophytoplankton but with varying relative contributions of micro- and picophytoplankton. In the present work, we also determined the degree of information (i.e. the number of phytoplankton communities) that can be predicted from the method applied to different configurations of BGC-Argo sensors.

The second approach to assessing phytoplankton composition from BGC-Argo data is based on an empirical relationship between bio-optical properties and plankton community size structure (section V). It consists of a PLS regression model that predicts the relative contribution to POC of four plankton groups (i.e. the three phytoplankton size classes plus bacteria) from hydrographical and optical data. This method has been first developed by Rembauville et al. (2017) based on a dataset collected in the Indian sector of the Southern Ocean. One of the goals of this thesis work was to evaluate the possibility of extending this method to different open ocean environments and trophic regimes.

The results showed the robust performance of the method in various oceanic environments. However, such method needs a learning dataset that covers a wide range of optical regimes associated with different plankton assemblages. The work of data collection carried out during this thesis allowed to construct a database comprising measurements from contrasted optical and biogeochemical conditions encountered over a seasonal cycle in the Northwestern Mediterranean Sea. The developed method, applied to observations from a BGC-Argo float deployed in the same region as the training dataset, revealed consistent annual and vertical dynamics of the plankton assemblage. Such of machine learning-based models hence offer a promising way for retrieving information related to phytoplankton community structure on large spatial and temporal scales based on BGC-Argo floats.

VI.2. Biogeochemical implications

Improving phytoplankton biomass estimates from BGC-Argo floats

Our results highlight the need to develop alternative methods to the current linear conversion of the *in-situ* fluorescence signal in terms of Chla concentration. As phytoplankton community composition significantly influences the fluorescence-to-Chla ratio, it is essential to introduce a correction coefficient that accounts for the variability in the phytoplankton communities encountered by the BGC-Argo floats over their lifetime. The estimation of Chla concentration based on fluorescence could be complemented, using additional BGC-Argo sensors, by other methods for estimating the composition of the phytoplanktonic communities, such as those presented in this thesis. Coupling these approaches could both help to better constrain the estimation of the Chla concentration and provide information on the taxonomic composition or size structure of phytoplankton communities. Another possibility to better constrain the fluorescence-based Chla concentration would be to use fluorometers with different excitation wavelengths. This would enable to use the fluorescence signal emitted in response to excitation in the wavelength channel corresponding with light absorption by Chla and avoid the absorption of non-photosynthetic pigments (Bricaud et al., 2004). This could be especially efficient in stratified oligotrophic regimes, where the vertical variability of such pigment concentration is pronounced and thus likely to induce errors in the estimation of the Chla concentration based on BGC-Argo fluorometers.

Accurate estimation of the Chla concentration from fluorometers mounted on BGC-Argo floats is valuable information, especially when complemented by other bio-optical variables. We considered, in this thesis, the possibility of assessing POC concentration from BGC-Argo float sensors. While the POC had previously been inferred from bio-optical empirical relationships using the b_{bp} or c_p coefficient (e.g. Gardner et al., 2006; Loisel et al., 2011; Cetinić et al., 2012), machine learning methods help to better account for the variety of factors affecting the bio-optical relationships. Ultimately, improving the estimation of the Chla and POC concentrations will increase the accuracy of the retrieved Chla-to-POC ratio, a key index of the physiological status of phytoplankton communities

(Behrenfeld and Boss, 2003; Westberry et al., 2008; Graff et al., 2016; Barbieux et al., 2022) and input variable of satellite carbon-based primary production models (Behrenfeld et al., 2005; Westberry et al., 2008).

Observing phytoplankton community composition from BGC-Argo floats

In the present thesis, we introduced two different machine learning methods based on either multispectral fluorescence or optical measurements, both dedicated to estimating phytoplankton community compositon from BGC-Argo floats. The multispectral fluorescence-based approach enables to predict taxonomic information (*i.e.* 4 different types of mixed phytoplankton assemblages) and represents a significant opportunity for the global BGC-Argo program. Indeed, the four different types of predicted communities are associated with distinct biogeochemical regimes encountered in the open ocean. The method is particularly efficient to distinguish between two picophytoplankton communities, dominated by either *Synechococcus* or *Prochlorococcus*, based on *in-situ* sensors, which, to our knowledge, had never been done in the past. This is of a particular interest as those two taxa are estimated to contribute 10–40% of global net primary production (Raven, 1998; Agawin et al., 2000; Uitz et al., 2010), a significant fraction of which being potentially exported (Stukel et al., 2013; Visintini et al., 2021). Phytoplankton is often represented in biogeochemical modelling by few size classes and this may lead to inaccuracy in biogeochemical cycles modelling (Gruber and Doney, 2009). Such method deployed on BGC-Argo floats would allow to precise the community composition in such models and reduce the bias.

The second method permits to infer the partitioning of the total POC stock among four plankton groups from bio-optical and hydrological BGC-Argo measurements. Multiple algorithms have been developed for determining phytoplankton community composition on the global scale based on satellite ocean color observations (e.g. Brewin et al., 2011; Bracher et al., 2017; Mouw et al., 2019). Although extremely valuable because of their wide scope, most of these products are restricted to the surface layer of the ocean. In contrast, the proposed BGC-Argo-based method allows a vertical description of the size structure of the phytoplankton community. Our results suggest that this method can be applied to wide range of open ocean environments. This method is of a particular interest to quantify

carbon stocks in regions characterized by high seasonality where the phytoplankton community succession and vertical distribution vary significantly throughout the year (Henson et al., 2012; Cornec et al., 2021).

The two methods presented in this work therefore appear to be complementary in terms of the information they provide. The multispectral fluorescence-based approach appears particularly relevant for stratified oligotrophic environments, where limited phytoplankton biomass and community gradients preclude the use of empirical relationships that typically require datasets with a large range of variability to perform in a robust manner. In contrast, the method based on optical and hydrographic measurements would be more powerful for regions with marked seasonal variability in the phytoplankton biomass and size structure.

VI.3. Perspectives

Data collection

The implementation of predictive methods from *in-situ* sensors relies, above all, on the possibility of collecting accurate reference data, representative of a wide range of environments. Moreover, the variables of interest and the predictive variables must be measured concomitantly. The collection of baseline data on phytoplankton community composition is based on seawater sampling, which reduces the amount of data collected per cruise. During this thesis work, we have developed different protocols to build a reference database for phytoplankton community composition for developing predictive methods. These protocols resulted in a one-year sampling effort in the Northwestern Mediterranean. The promising results of the methods based on these data encourages the extension of this protocol to different oceanic regions. Increasing the amount of data will not only improve the performance of the models, but also provide a better understanding of the mechanisms underlying the relationships between phytoplankton community composition and the different predictor variables. It is also particularly

important to perform these samples during float deployments, in order to improve the performances of these methods on BGC-Argo floats.

Future studies

The studies carried out within the framework of this thesis focus on the composition of phytoplankton communities. In future, it would be interesting to test the covariance between the fluorescence-to-Chla ratio and the phytoplanktonic groups predicted by multispectral fluorescence or optical measurements. This could result in new methods to converte *in-situ* fluorescence into Chla concentration, taking account of the community composition. In the longer term, as more data become available for training, we hope to be able to test the predictive method presented in Chapter 5 to other ocean regions. Ultimately, this method could be applied to the BGC-Argo floats equipped with the relevant sensor package in order to generate large-scale 3D estimates of phytoplankton group-specific POC concentrations. These POC estimates would be valuable to improve our understanding of the responses and biogeochemical impact of phytoplankton communities to changes in environmental drivers also measured by the BGC-Argo floats (light regime and mixed layer depth, in particular). Such estimates could also serve as a reference for the validation of global biogeochemical models that integrate different classes of plankton and lack *in-situ* validation data (e.g. Le Quéré et al., 2005; Stocker et al., 2013).

The methods presented here, along with recent works (Rembauville et al. 2017; Terrrats et al. 2020), pave the way for taking into account phytoplankton community composition in biogeochemical cycle studies using BGC-Argo floats. This is in line with the evolution of the BGC-Argo program that aims not only to increase the size of the fleet, but also to extend the range of variables observed by the floats (Claustre et al., 2020; Bishop et al., 2022; Picheral et al., 2022), an evolution consistent with recommendations of the international community to enhance biological observations from existing programs (e.g. Miloslavich et al., 2018; Boss et al., 2022). The parallel development of new predictive methods and integration of novel sensors onto BGC-Argo floats will eventually allow a finer understanding and quantification of global biogeochemical processes along with integrative analyses of ecological processes.

References

Agawin, N. S., Duarte, C., and Agustí, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production (Errata). *Limnology and Oceanography* 45, 591–600. doi: 10.2307/2670836.

Allali, K., Bricaud, A., and Claustre, H. (1997). Spatial variations in the chlorophyll-specific absorption coefficients of phytoplankton and photosynthetically active pigments in the equatorial Pacific. *Journal of Geophysical Research: Oceans* 102, 12413–12423. doi: 10.1029/97JC00380.

Alpine, A. E., and Cloern, J. E. (1985). Differences in in vivo fluorescence yield between three phytoplankton size classes. *Journal of Plankton Research* 7, 381–390. doi: 10.1093/plankt/7.3.381.

Antoine, D., Guevel, P., Desté, J.-F., Bécu, G., Louis, F., Scott, A. J., et al. (2008). The “BOUSSOLE” Buoy—A New Transparent-to-Swell Taut Mooring Dedicated to Marine Optics: Design, Tests, and Performance at Sea. *Journal of Atmospheric and Oceanic Technology* 25, 968–989. doi: 10.1175/2007JTECHO563.1.

Babin, M., Morel, A., Fournier-Sicre, V., Fell, F., and Stramski, D. (2003). Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. *Limnology and Oceanography* 48, 843–859. doi: 10.4319/lo.2003.48.2.0843.

Barbieux, M., Uitz, J., Bricaud, A., Organelli, E., Poteau, A., Schmechtig, C., et al. (2018). Assessing the Variability in the Relationship Between the Particulate Backscattering Coefficient and the Chlorophyll *a* Concentration From a Global Biogeochemical-Argo Database. *J. Geophys. Res. Oceans* 123, 1229–1250. doi: 10.1002/2017JC013030.

Barbieux, M., Uitz, J., Gentili, B., Pasqueron de Fommervault, O., Mignot, A., Poteau, A., et al. (2019a). Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database. *Biogeosciences* 16, 1321–1342. doi: 10.5194/bg-16-1321-2019.

Barbieux, M., Uitz, J., Gentili, B., Pasqueron de Fommervault, O., Mignot, A., Poteau, A., et al. (2019b). Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database. *Biogeosciences* 16, 1321–1342. doi: 10.5194/bg-16-1321-2019.

Barbieux, M., Uitz, J., Mignot, A., Roesler, C., Claustre, H., Gentili, B., et al. (2022). Biological production in two contrasted regions of the Mediterranean Sea during the oligotrophic period: an estimate based on the diel cycle of optical properties measured by BioGeoChemical-Argo profiling floats. *Biogeosciences* 19, 1165–1194. doi: 10.5194/bg-19-1165-2022.

Behrenfeld, M. J., and Boss, E. (2003). The beam attenuation to chlorophyll ratio: an optical index of phytoplankton physiology in the surface ocean? *Deep Sea Research Part I: Oceanographic Research Papers* 50, 1537–1549. doi: 10.1016/j.dsr.2003.09.002.

Behrenfeld, M. J., and Boss, E. (2006). Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass. *J Mar Res* 64, 431–451. doi: 10.1357/002224006778189563.

Behrenfeld, M. J., Boss, E., Siegel, D. A., and Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles* 19. doi: 10.1029/2004GB002299.

Behrenfeld, M. J., Westberry, T. K., Boss, E. S., O'Malley, R. T., Siegel, D. A., Wiggert, J. D., et al. (2009). Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences* 6, 779–794. doi: 10.5194/bg-6-779-2009.

Berglund, J., Müren, U., Båmstedt, U., and Andersson, A. (2007). Efficiency of a phytoplankton-based and a bacterial-based food web in a pelagic marine system. *Limnology and Oceanography* 52, 121–131. doi: 10.4319/lo.2007.52.1.0121.

Berner, R. A., and Kothavala, Z. (2001). Geocarb III: A Revised Model of Atmospheric CO2 over Phanerozoic Time. *American Journal of Science* 301, 182–204. doi: 10.2475/ajs.301.2.182.

Bidigare, R. R., Schofield, O., and Prézelin, B. B. (1989). Influence of zeaxanthin on quantum yield of photosynthesis of Synechococcus clone WH7803 (DC2). *Marine Ecology Progress Series* 56, 177–188.

Bindoff, N. L., Cheung, W. W. L., Kairo, J. G., Arístegui, J., Guinder, V. A., Hallberg, R., et al. (2019). Changing Ocean, Marine Ecosystems, and Dependent Communities. *Marine Ecosystems*, 142.

Biogeochemical-Argo Planning Group (2016). *The scientific rationale, design and implementation plan for a Biogeochemical-Argo float array*. Ifremer doi: 10.13155/46601.

Bishop, J. K. B., Amaral, V. J., Lam, P. J., Wood, T. J., Lee, J.-M., Laubach, A., et al. (2022). Transmitted Cross-Polarized Light Detection of Particulate Inorganic Carbon Concentrations and Fluxes in the Ocean Water Column: Ships to ARGO Floats. *Frontiers in Remote Sensing* 3. Available at: https://www.frontiersin.org/articles/10.3389/frsen.2022.837938 [Accessed February 2, 2023].

Bittig, H. C., Maurer, T. L., Plant, J. N., Schmechtig, C., Wong, A. P. S., Claustre, H., et al. (2019). A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. *Frontiers in Marine Science* 6. Available at: https://www.frontiersin.org/article/10.3389/fmars.2019.00502 [Accessed January 28, 2022].

Bonnet, S., Benavides, M., Le Moigne, F. A. C., Camps, M., Torremocha, A., Grosso, O., et al. (2023). Diazotrophs are overlooked contributors to carbon and nitrogen export to the deep ocean. *ISME J* 17, 47–58. doi: 10.1038/s41396-022-01319-3.

Booth, B. (1993a). Estimating cell concentration and biomass of autotrophic plankton using microscopy. *Handbook of Methods in Aquatic Microbial Ecology*, 199–206.

Booth, B. C. (1993b). "Estimating Cell Concentration and Biomass of Autotrophic Plankton Using Microscopy," in *Handbook of Methods in Aquatic Microbial Ecology* (Paul F. Kemp), 199–205.

Boss, E., Pegau, W. S., Lee, M., Twardowski, M., Shybanov, E., Korotaev, G., et al. (2004). Particulate backscattering ratio at LEO 15 and its use to study particle composition and

distribution. *Journal of Geophysical Research: Oceans* 109. doi: 10.1029/2002JC001514.

Boss, E., Waite, A. M., Karstensen, J., Trull, T., Muller-Karger, F., Sosik, H. M., et al. (2022). Recommendations for Plankton Measurements on OceanSITES Moorings With Relevance to Other Observing Sites. *Frontiers in Marine Science* 9. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2022.929436 [Accessed February 3, 2023].

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., and Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature* 568, 327–335. doi: 10.1038/s41586-019-1098-2.

Boyd, P. W., and Newton, P. P. (1999). Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces? *Deep Sea Research Part I: Oceanographic Research Papers* 46, 63–91. doi: 10.1016/S0967-0637(98)00066-1.

Boyd, P. W., Watson, A. J., Law, C. S., Abraham, E. R., Trull, T., Murdoch, R., et al. (2000). A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* 407, 695–702. doi: 10.1038/35037500.

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. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2017.00055 [Accessed January 20, 2023].

Bratbak, G. (1985). Bacterial Biovolume and Biomass Estimations. *Applied and Environmental Microbiology* 49, 1488–1493. doi: 10.1128/aem.49.6.1488-1493.1985.

Brewin, R. J. W., Devred, E., Sathyendranath, S., Lavender, S. J., and Hardman-Mountford, N. J. (2011a). Model of phytoplankton absorption based on three size classes. *Appl. Opt., AO* 50, 4535–4549. doi: 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, 325–339. doi: 10.1016/j.rse.2010.09.004.

Brewin, R. J. W., Sathyendranath, S., Lange, P. K., and Tilstone, G. (2014a). Comparison of two methods to derive the size-structure of natural populations of phytoplankton. *Deep Sea Research Part I: Oceanographic Research Papers* 85, 72–79. doi: 10.1016/j.dsr.2013.11.007.

Brewin, R. J. W., Sathyendranath, S., Tilstone, G., Lange, P. K., and Platt, T. (2014b). A multicomponent model of phytoplankton size structure. *Journal of Geophysical Research: Oceans* 119, 3478–3496. doi: 10.1002/2014JC009859.

Bricaud, A. (2004). Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *J. Geophys. Res.* 109, C11010. doi: 10.1029/2004JC002419.

Bricaud, A., Allali, K., Morel, A., Marie, D., Veldhuis, M. J. W., Partensky, F., et al. (1999). Divinyl chlorophyll a-specific absorption coefficients and absorption efficiency factors

for Prochlorococcus marinus: kinetics of photoacclimation. *Marine Ecology Progress Series* 188, 21–32. doi: 10.3354/meps188021.

Bricaud, A., Babin, M., Claustre, H., Ras, J., and Tièche, F. (2010). Light absorption properties and absorption budget of Southeast Pacific waters. *Journal of Geophysical Research: Oceans* 115. doi: 10.1029/2009JC005517.

Bricaud, A., Babin, M., Morel, A., and Claustre, H. (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. *Journal of Geophysical Research* 100, 13321. doi: 10.1029/95JC00463.

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

Bricaud, A., and Morel, A. (1986). Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. *Appl. Opt., AO* 25, 571–580. doi: 10.1364/AO.25.000571.

Bricaud, A., and Stramski, D. (1990). Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: A comparison between the Peru upwelling areaand the Sargasso Sea. *Limnology and Oceanography* 35, 562–582. doi: 10.4319/lo.1990.35.3.0562.

Briggs, N., Perry, M. J., Cetinić, I., Lee, C., D'Asaro, E., Gray, A. M., et al. (2011). High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom. *Deep Sea Research Part I: Oceanographic Research Papers* 58, 1031–1039.

Briggs, N. T., Slade, W. H., Boss, E., and Perry, M. J. (2013). Method for estimating mean particle size from high-frequency fluctuations in beam attenuation or scattering measurements. *Appl. Opt., AO* 52, 6710–6725. doi: 10.1364/AO.52.006710.

Broecker, W. S., and Peng, T.-H. (1974). Gas exchange rates between air and sea. *Tellus* 26, 21–35. doi: 10.3402/tellusa.v26i1-2.9733.

Brunet, C., Johnsen, G., Lavaud, J., and Roy, S. (2011). "Pigments and photoacclimation processes," in *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography* Available at: https://hal.archives-ouvertes.fr/hal-01101814 [Accessed July 20, 2022].

Buesseler, K., Ball, L., Andrews, J., Benitez-Nelson, C., Belastock, R., Chai, F., et al. (1998). Upper ocean export of particulate organic carbon in the Arabian Sea derived from thorium-234. *Deep Sea Research Part II: Topical Studies in Oceanography* 45, 2461–2487. doi: 10.1016/S0967-0645(98)80022-2.

Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Lam, P. J., Trull, T. W., Bidigare, R. R., et al. (2007). Revisiting Carbon Flux Through the Ocean's Twilight Zone. *Science* 316, 567–570. doi: 10.1126/science.1137959.

Bustillos-Guzmán, J., Claustre, H., and Marty, C. (1995). Specific phytoplankton signatures and their relationship to hydrographic conditions in the coastal northwestern Mediterranean Sea. *Marine Ecology Progress Series* 124, 247–258. doi: 10.3354/meps124247.

Cermeño, P., Marañón, E., Rodríguez, J., and Fernández, E. (2005a). Large-sized phytoplankton sustain higher carbon-specific photosynthesis than smaller cells in a

coastal eutrophic ecosystem. *Marine Ecology Progress Series* 297, 51–60. doi: 10.3354/meps297051.

Cermeño, P., Marañón, E., Rodríguez, J., and Fernández, E. (2005b). Size dependence of coastal phytoplankton photosynthesis under vertical mixing conditions. *Journal of Plankton Research* 27, 473–483. doi: 10.1093/plankt/fbi021.

Cetinić, I., Perry, M. J., Briggs, N. T., Kallin, E., D'Asaro, E. A., and Lee, C. M. (2012). Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research: Oceans* 117. doi: 10.1029/2011JC007771.

Cetinić, I., Perry, M. J., D'Asaro, E., Briggs, N., Poulton, N., Sieracki, M. E., et al. (2015). A simple optical index shows spatial and temporal heterogeneity in phytoplankton community composition during the 2008 North Atlantic Bloom Experiment. *Biogeosciences* 12, 2179–2194. doi: 10.5194/bg-12-2179-2015.

Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco California USA: ACM), 785–794. doi: 10.1145/2939672.2939785.

Chisholm, S. W. (1992). "Phytoplankton Size," in *Primary Productivity and Biogeochemical Cycles in the Sea*, eds. P. G. Falkowski, A. D. Woodhead, and K. Vivirito (Boston, MA: Springer US), 213–237. doi: 10.1007/978-1-4899-0762-2_12.

Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B., and Welschmeyer, N. A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334, 340–343. doi: 10.1038/334340a0.

Ciotti, Á. M., Lewis, M. R., and Cullen, J. J. (2002). Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnol. Oceanogr.* 47, 404–417. doi: 10.4319/lo.2002.47.2.0404.

Claustre, H. (1994). The trophic status of various oceanic provinces as revealed by phytoplankton pigment signatures. *Limnology and Oceanography* 39, 1206–1210. doi: 10.4319/lo.1994.39.5.1206.

Claustre, H., Hooker, S. B., Van Heukelem, L., Berthon, J.-F., Barlow, R., Ras, J., et al. (2004). An intercomparison of HPLC phytoplankton pigment methods using in situ samples: application to remote sensing and database activities. *Marine Chemistry* 85, 41–61. doi: 10.1016/j.marchem.2003.09.002.

Claustre, H., Johnson, K. S., and Takeshita, Y. (2020). Observing the Global Ocean with Biogeochemical-Argo. *Annual Review of Marine Science* 12, 23–48. doi: 10.1146/annurev-marine-010419-010956.

Cleveland, J. S. (1995). Regional models for phytoplankton absorption as a function of chlorophyll a concentration. *Journal of Geophysical Research: Oceans* 100, 13333–13344. doi: 10.1029/95JC00532.

Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., et al. (2021). Deep Chlorophyll Maxima in the Global Ocean: Occurrences, Drivers and Characteristics. *Global Biogeochemical Cycles* 35, e2020GB006759. doi: 10.1029/2020GB006759.

Cosgrove, J., and Borowitzka, M. A. (2010). “Chlorophyll Fluorescence Terminology: An Introduction,” in *Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications*, eds. D. J. Suggett, O. Prášil, and M. A. Borowitzka (Dordrecht: Springer Netherlands), 1–17. doi: 10.1007/978-90-481-9268-7_1.

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., et al. (2019). Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea. *Ocean Modelling* 133, 112–128. doi: 10.1016/j.ocemod.2018.11.005.

Cushing, D. H. (1989). A difference in structure between ecosystems in strongly stratified waters and in those that are only weakly stratified. *Journal of Plankton Research* 11, 1–13. doi: 10.1093/plankt/11.1.1.

Dale, T., Rey, F., and Heimdal, B. R. (1999). Seasonal development of phytoplankton at a high latitude oceanic site. *Sarsia* 84, 419–435. doi: 10.1080/00364827.1999.10807347.

Demmig-Adams, B. (1990). Carotenoids and photoprotection in plants: A role for the xanthophyll zeaxanthin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1020, 1–24. doi: 10.1016/0005-2728(90)90088-L.

DeVries, T. (2022). The Ocean Carbon Cycle. *Annual Review of Environment and Resources* 47, 317–341. doi: 10.1146/annurev-environ-120920-111307.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A. (2009). Ocean Acidification: The Other CO2 Problem. *Annual Review of Marine Science* 1, 169–192. doi: 10.1146/annurev.marine.010908.163834.

D’Ortenzio, F., Iudicone, D., de Boyer Montegut, C., Testor, P., Antoine, D., Marullo, S., et al. (2005). Seasonal variability of the mixed layer depth in the Mediterranean Sea as derived from in situ profiles. *Geophysical Research Letters* 32. doi: 10.1029/2005GL022463.

Dubelaar, G. B. J., and Jonker, R. R. (2000). Flow cytometry as a tool for the study of phytoplankton. *Scientia Marina* 64, 135–156. doi: 10.3989/scimar.2000.64n2135.

Dubinsky, Z., and Stambler, N. (2009). Photoacclimation processes in phytoplankton: mechanisms, consequences, and applications. *Aquat. Microb. Ecol.* 56, 163–176. doi: 10.3354/ame01345.

Durrieu de Madron, X., Guieu, C., Sempéré, R., Conan, P., Cossa, D., D’Ortenzio, F., et al. (2011). Marine ecosystems’ responses to climatic and anthropogenic forcings in the Mediterranean. *Progress in Oceanography* 91, 97–166. doi: 10.1016/j.pocean.2011.02.003.

Escoffier, N., Bernard, C., Hamlaoui, S., Groleau, A., and Catherine, A. (2015). Quantifying phytoplankton communities using spectral fluorescence: the effects of species composition and physiological state. *Journal of Plankton Research* 37, 233–247. doi: 10.1093/plankt/fbu085.

Falkowski, P. G. (1994a). The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynthesis Research* 39, 235–258. doi: 10.1007/BF00014586.

Falkowski, P. G. (1994b). The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynth Res* 39, 235–258. doi: 10.1007/BF00014586.

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* 281, 200–206. doi: 10.1126/science.281.5374.200.

Falkowski, P., and Kolber, Z. (1995). Variations in Chlorophyll Fluorescence Yields in Phytoplankton in the World Oceans. *Functional Plant Biology* 22, 341. doi: 10.1071/PP9950341.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237.

Finkel, Z., Beardall, J., Flynn, K., Quigg, A., Rees, A., and Raven, J. (2010). Pytoplankton in a changing world: Cell size and elemental stoichiometry. *Journal of Plankton Research* 32, 119–137. doi: 10.1093/plankt/fbp098.

Finkel, Z. V. (2007). “CHAPTER 15 - Does Phytoplankton Cell Size Matter? The Evolution of Modern Marine Food Webs,” in *Evolution of Primary Producers in the Sea*, eds. P. G. Falkowski and A. H. Knoll (Burlington: Academic Press), 333–350. doi: 10.1016/B978-012370518-1/50016-3.

Finzi, A. C., Austin, A. T., Cleland, E. E., Frey, S. D., Houlton, B. Z., and Wallenstein, M. D. (2011). Responses and feedbacks of coupled biogeochemical cycles to climate change: examples from terrestrial ecosystems. *Frontiers in Ecology and the Environment* 9, 61–67. doi: 10.1890/100001.

Fischer, A. D., Moberg, E. A., Alexander, H., Brownlee, E. F., Hunter-Cevera, K. R., Pitz, K. J., et al. (2014). Sixty Years of Sverdrup: A Retrospective of Progress in the Study of Phytoplankton Blooms. *Oceanography* 27, 222–235.

Friedlingstein, P., O’Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global Carbon Budget 2022. *Earth System Science Data* 14, 4811–4900. doi: 10.5194/essd-14-4811-2022.

Gardner, W. D., Mishonov, A. V., and Richardson, M. J. (2006). Global POC concentrations from in-situ and satellite data. *Deep Sea Research Part II: Topical Studies in Oceanography* 53, 718–740. doi: 10.1016/j.dsr2.2006.01.029.

Geider, R. J. (1987). LIGHT AND TEMPERATURE DEPENDENCE OF THE CARBON TO CHLOROPHYLL a RATIO IN MICROALGAE AND CYANOBACTERIA: IMPLICATIONS FOR PHYSIOLOGY AND GROWTH OF PHYTOPLANKTON. *New Phytol* 106, 1–34. doi: 10.1111/j.1469-8137.1987.tb04788.x.

Golbol, M., Vellucci, V., and Antoine, D. (2000). BOUSSOLE. doi: 10.18142/1.

Gorbunov, M. Y., and Falkowski, P. G. (2022). Using Chlorophyll Fluorescence to Determine the Fate of Photons Absorbed by Phytoplankton in the World’s Oceans. *Annual Review of Marine Science* 14, 213–238. doi: 10.1146/annurev-marine-032621-122346.

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall’Olmo, G., Dongen-Vogels, V. van, et al. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers* 102, 16–25. doi: 10.1016/j.dsr.2015.04.006.

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall' Olmo, G., Reifel, K. M., et al. (2016). Photoacclimation of natural phytoplankton communities. *Marine Ecology Progress Series* 542, 51–62. doi: 10.3354/meps11539.

Greg Mitchell, B., and Kiefer, D. A. (1988). Chlorophyll α specific absorption and fluorescence excitation spectra for light-limited phytoplankton. *Deep Sea Research Part A. Oceanographic Research Papers* 35, 639–663. doi: 10.1016/0198-0149(88)90024-6.

Grob, C., Ulloa, O., Claustre, H., Huot, Y., Alarcón, G., and Marie, D. (2007). Contribution of picoplankton to the total particulate organic carbon concentration in the eastern South Pacific. *Biogeosciences* 4, 837–852. doi: 10.5194/bg-4-837-2007.

Gruber, N., and Doney, S. C. (2009). "Ocean Biogeochemistry and Ecology, Modeling of," in (Boston: Academic Press), 89–104. doi: 10.1016/B978-012374473-9.00741-4.

Guidi, L., Stemmann, L., Jackson, G. A., Ibanez, F., Claustre, H., Legendre, L., et al. (2009). Effects of phytoplankton community on production, size, and export of large aggregates: A world-ocean analysis. *Limnology and Oceanography* 54, 1951–1963. doi: 10.4319/lo.2009.54.6.1951.

Harada, N. (2016). Review: Potential catastrophic reduction of sea ice in the western Arctic Ocean: Its impact on biogeochemical cycles and marine ecosystems. *Global and Planetary Change* 136, 1–17. doi: 10.1016/j.gloplacha.2015.11.005.

Henson, S. A., Sanders, R., and Madsen, E. (2012). Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochemical Cycles* 26. doi: 10.1029/2011GB004099.

Herndl, G. J., and Reinthaler, T. (2013). Microbial control of the dark end of the biological pump. *Nature Geosci* 6, 718–724. doi: 10.1038/ngeo1921.

Hilligsøe, K. M., Richardson, K., Bendtsen, J., Sørensen, L.-L., Nielsen, T. G., and Lyngsgaard, M. M. (2011). Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO2 flux. *Deep Sea Research Part I: Oceanographic Research Papers* 58, 826–838. doi: 10.1016/j.dsr.2011.06.004.

Hoegh-Guldberg, O., and Bruno, J. F. (2010). The Impact of Climate Change on the World's Marine Ecosystems. *Science* 328, 1523–1528. doi: 10.1126/science.1189930.

Huot, Y., and Babin, M. (2010). "Overview of Fluorescence Protocols: Theory, Basic Concepts, and Practice," in *Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications*, eds. D. J. Suggett, O. Prášil, and M. A. Borowitzka (Dordrecht: Springer Netherlands), 31–74. doi: 10.1007/978-90-481-9268-7_3.

Huot, Y., Brown, C. A., and Cullen, J. J. (2007). Retrieval of phytoplankton biomass from simultaneous inversion of reflectance, the diffuse attenuation coefficient, and Sun-induced fluorescence in coastal waters. *Journal of Geophysical Research: Oceans* 112. doi: 10.1029/2006JC003794.

Intergovernmental Panel on Climate Change ed. (2014). "Detection and Attribution of Climate Change: from Global to Regional," in *Climate Change 2013 – The Physical Science Basis* (Cambridge University Press), 867–952. doi: 10.1017/CBO9781107415324.022.

Ito, T., and Follows, M. J. (2005). Preformed phosphate, soft tissue pump and atmospheric CO2. *Journal of Marine Research* 63, 813–839. doi: 10.1357/0022240054663231.

Jeffrey, S. W., Mantoura, R. F. C., Wright, S. W., International Council of Scientific Unions, and Unesco eds. (1997). *Phytoplankton pigments in oceanography: guidelines to modern methods*. Paris: UNESCO Pub.

Johnsen, G., and Sakshaug, E. (2007). Biooptical characteristics of PSII and PSI in 33 species (13 pigment groups) of marine phytoplankton, and the relevance for pulse-amplitude-modulated and fast-repetition-rate fluorometry1 - Johnsen - 2007 - Journal of Phycology - Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1529-8817.2007.00422.x?casa_token=jWEUfuxxVLQAAAAA%3AT_jXMtekOkVqD6xMOl2RGqu0mzOrF1HELQKF6t2XuSLPkPKkvXB2WvRo007g2weS17GYbbhuSX2DfAgZ [Accessed January 27, 2022].

Karlson, B., and al, et (2010). Microscopic and molecular methods for quantitative phytoplankton analysis. Unesco doi: 10.25607/OBP-1371.

Keller, M. D., Selvin, R. C., Claus, W., and Guillard, R. R. L. (1987). Media for the Culture of Oceanic Ultraphytoplankton1,2. *Journal of Phycology* 23, 633–638. doi: 10.1111/j.1529-8817.1987.tb04217.x.

Kiefer, D. A., and Reynolds, R. A. (1992). “Advances in Understanding Phytoplankton Fluorescence and Photosynthesis,” in *Primary Productivity and Biogeochemical Cycles in the Sea* Environmental Science Research., eds. P. G. Falkowski, A. D. Woodhead, and K. Vivirito (Boston, MA: Springer US), 155–174. doi: 10.1007/978-1-4899-0762-2_10.

Koestner, D., Stramski, D., and Reynolds, R. A. (2022). A Multivariable Empirical Algorithm for Estimating Particulate Organic Carbon Concentration in Marine Environments From Optical Backscattering and Chlorophyll-a Measurements. *Frontiers in Marine Science* 9. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2022.941950 [Accessed December 23, 2022].

Lacour, L., Ardyna, M., Stec, K. F., Claustre, H., Prieur, L., Poteau, A., et al. (2017). Unexpected winter phytoplankton blooms in the North Atlantic subpolar gyre. *Nature Geoscience* 10, 836–839. doi: 10.1038/ngeo3035.

Lavigne, H., D’Ortenzio, F., Ribera D’Alcalà, M., Claustre, H., Sauzède, R., and Gacic, M. (2015). On the vertical distribution of the chlorophyll *a* concentration in the Mediterranean Sea: a basin-scale and seasonal approach. *Biogeosciences* 12, 5021–5039. doi: 10.5194/bg-12-5021-2015.

Litchman, E., de Tezanos Pinto, P., Edwards, K. F., Klausmeier, C. A., Kremer, C. T., and Thomas, M. K. (2015). Global biogeochemical impacts of phytoplankton: a trait-based perspective. *Journal of Ecology* 103, 1384–1396. doi: 10.1111/1365-2745.12438.

Loisel, H., Vantrepotte, V., Norkvist, K., Mériaux, X., Kheireddine, M., Ras, J., et al. (2011). Characterization of the bio-optical anomaly and diurnal variability of particulate matter, as seen from scattering and backscattering coefficients, in ultra-oligotrophic eddies of the Mediterranean Sea. *Biogeosciences* 8, 3295–3317. doi: 10.5194/bg-8-3295-2011.

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., et al. (2019). Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Frontiers in Marine Science* 6. Available at: https://www.frontiersin.org/article/10.3389/fmars.2019.00196 [Accessed May 19, 2022].

Longhurst, A. R. (2006). *Ecological geography of the sea*. 2nd ed. Burlington, MA: Academic Press.

Lorenzen, C. J. (1966). A method for the continuous measurement of in vivo chlorophyll concentration. *Deep Sea Research and Oceanographic Abstracts* 13, 223–227. doi: 10.1016/0011-7471(66)91102-8.

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

Marty, J.-C., Chiavérini, J., Pizay, M.-D., and Avril, B. (2002). Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999). *Deep Sea Research Part II: Topical Studies in Oceanography* 49, 1965–1985. doi: 10.1016/S0967-0645(02)00022-X.

Mayersohn, B., Lévy, M., Mangolte, I., and Smith, K. S. (2022). Emergence of broadband variability in a marine plankton model under external forcing. *Journal of Geophysical Research: Biogeosciences* n/a, e2022JG007011. doi: 10.1029/2022JG007011.

Mayot, N., D'Ortenzio, F., Uitz, J., Gentili, B., Ras, J., Vellucci, V., et al. (2017). Influence of the Phytoplankton Community Structure on the Spring and Annual Primary Production in the Northwestern Mediterranean Sea. *Journal of Geophysical Research: Oceans* 122, 9918–9936. doi: 10.1002/2016JC012668.

Menden-Deuer, S., and Lessard, E. J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnology and Oceanography* 45, 569–579. doi: 10.4319/lo.2000.45.3.0569.

Michaels, A. F., and Silver, M. W. (1988). Primary production, sinking fluxes and the microbial food web. *Deep Sea Research Part A. Oceanographic Research Papers* 35, 473–490. doi: 10.1016/0198-0149(88)90126-4.

Mignot, A., Ferrari, R., and Claustre, H. (2018). Floats with bio-optical sensors reveal what processes trigger the North Atlantic bloom. *Nat Commun* 9, 190. doi: 10.1038/s41467-017-02143-6.

Miloslavich, P., Bax, N. J., Simmons, S. E., Klein, E., Appeltans, W., Aburto-Oropeza, O., et al. (2018). Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Global Change Biology* 24, 2416–2433. doi: 10.1111/gcb.14108.

Morel, A., and Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research: Oceans* 106, 7163–7180. doi: 10.1029/2000JC000319.

Morel, F. M. M., and Price, N. M. (2003). The Biogeochemical Cycles of Trace Metals in the Oceans. *Science* 300, 944–947. doi: 10.1126/science.1083545.

Mouw, C. B., Barnett, A., McKinley, G. A., Gloege, L., and Pilcher, D. (2016). Phytoplankton size impact on export flux in the global ocean. *Global Biogeochemical Cycles* 30, 1542–1562. doi: 10.1002/2015GB005355.

Mouw, C. B., Ciochetto, A. B., and Yoder, J. A. (2019). A Satellite Assessment of Environmental Controls of Phytoplankton Community Size Structure. *Global Biogeochemical Cycles* 33, 540–558. doi: 10.1029/2018GB006118.

Olaizola, M., and Yamamoto, H. Y. (1994). Short-Term Response of the Diadinoxanthin Cycle and Fluorescence Yield to High Irradiance in Chaetoceros Muelleri (bacillariophyceae)1. *Journal of Phycology* 30, 606–612. doi: 10.1111/j.0022-3646.1994.00606.x.

Oubelkheir, K., Claustre, H., Sciandra, A., and Babin, M. (2005). Bio-optical and biogeochemical properties of different trophic regimes in oceanic waters. *Limnology and Oceanography* 50, 1795–1809. doi: 10.4319/lo.2005.50.6.1795.

Parkhill, J.-P., Maillet, G., and Cullen, J. J. (2001). FLUORESCENCE-BASED MAXIMAL QUANTUM YIELD FOR PSII AS A DIAGNOSTIC OF NUTRIENT STRESS. *Journal of Phycology* 37, 517–529. doi: 10.1046/j.1529-8817.2001.037004517.x.

Petit, F., Uitz, J., Schmechtig, C., Dimier, C., Ras, J., Poteau, A., et al. (2022). Influence of the phytoplankton community composition on the in situ fluorescence signal: Implication for an improved estimation of the chlorophyll-a concentration from BioGeoChemical-Argo profiling floats. *Frontiers in Marine Science* 9. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2022.959131 [Accessed October 3, 2022].

Picheral, M., Catalano, C., Brousseau, D., Claustre, H., Coppola, L., Leymarie, E., et al. (2022). The Underwater Vision Profiler 6: an imaging sensor of particle size spectra and plankton, for autonomous and cabled platforms. *Limnology and Oceanography: Methods* 20, 115–129. doi: 10.1002/lom3.10475.

Proctor, C. W., and Roesler, C. S. (2010). New insights on obtaining phytoplankton concentration and composition from in situ multispectral Chlorophyll fluorescence: In situ phytoplankton composition. *Limnology and Oceanography: Methods* 8, 695–708. doi: 10.4319/lom.2010.8.0695.

Quéré, C. L., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* 11, 2016–2040. doi: 10.1111/j.1365-2486.2005.1004.x.

Ras, J., Claustre, H., and Uitz, J. (2008). Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: comparison between in situ and predicted data. 17.

Raven, J. A. (1998). The twelfth Tansley Lecture. Small is beautiful: the picophytoplankton. *Functional Ecology* 12, 503–513. doi: 10.1046/j.1365-2435.1998.00233.x.

Rembauville, M., Briggs, N., Ardyna, M., Uitz, J., Catala, P., Penkerc'h, C., et al. (2017). Plankton Assemblage Estimated with BGC-Argo Floats in the Southern Ocean: Implications for Seasonal Successions and Particle Export: PLANKTON ASSEMBLAGE BGC-ARGO. *Journal of Geophysical Research: Oceans* 122, 8278–8292. doi: 10.1002/2017JC013067.

Rippka, R., Coursin, T., Hess, W., Lichtlé, C., Scanlan, D. J., Palinska, K. A., et al. (2000). Prochlorococcus marinus Chisholm et al. 1992 subsp. pastoris subsp. nov. strain PCC 9511, the first axenic chlorophyll a2/b2-containing cyanobacterium (Oxyphotobacteria).

International Journal of Systematic and Evolutionary Microbiology 50, 1833–1847. doi: 10.1099/00207713-50-5-1833.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., et al. (2019). On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Frontiers in Marine Science* 6. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2019.00439 [Accessed September 8, 2022].

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., et al. (2017). Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors: Unbiased chlorophyll from in situ fluorometers. *Limnology and Oceanography: Methods* 15, 572–585. doi: 10.1002/lom3.10185.

Röttgers, R., and Gehnke, S. (2012). Measurement of light absorption by aquatic particles: improvement of the quantitative filter technique by use of an integrating sphere approach. *Appl. Opt., AO* 51, 1336–1351. doi: 10.1364/AO.51.001336.

Roy, S., Llewellyn, C., Egeland, E. S., and Johnsen, G. eds. (2011). *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*. Cambridge: Cambridge University Press doi: 10.1017/CBO9780511732263.

Roy, S., Sathyendranath, S., Bouman, H., and Platt, T. (2013). The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data. *Remote Sensing of Environment* 139, 185–197. doi: 10.1016/j.rse.2013.08.004.

Sammartino, M., Di Cicco, A., Marullo, S., and Santoleri, R. (2015). Spatio-temporal variability of micro-, nano- and pico-phytoplankton in the Mediterranean Sea from satellite ocean colour data of SeaWiFS. *Ocean Sci.* 11, 759–778. doi: 10.5194/os-11-759-2015.

Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., et al. (2015a). Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications. *Journal of Geophysical Research: Oceans* 120, 451–470. doi: 10.1002/2014JC010355.

Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., et al. (2015b). Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications. *Journal of Geophysical Research: Oceans* 120, 451–470. doi: 10.1002/2014JC010355.

Schallenberg, C. (2022). Iron limitation drives high fluorescence/chlorophyll ratios in the southern ocean and points the way to improved phytoplankton biomass estimates from BGC-Argo floats.

Schallenberg, C., Strzepek, R. F., Bestley, S., Wojtasiewicz, B., and Trull, T. W. (2022a). Iron Limitation Drives the Globally Extreme Fluorescence/Chlorophyll Ratios of the Southern Ocean. *Geophysical Research Letters* 49, e2021GL097616. doi: 10.1029/2021GL097616.

Schallenberg, C., Strzepek, R. F., Bestley, S., Wojtasiewicz, B., and Trull, T. W. (2022b). Iron Limitation Drives the Globally Extreme Fluorescence/Chlorophyll Ratios of the

Southern Ocean. *Geophysical Research Letters* 49, e2021GL097616. doi: 10.1029/2021GL097616.

Schallenberg, C., Strzepek, R. F., Schuback, N., Clementson, L. A., Boyd, P. W., and Trull, T. W. (2020). Diel quenching of Southern Ocean phytoplankton fluorescence is related to iron limitation. *Biogeosciences* 17, 793–812. doi: 10.5194/bg-17-793-2020.

Schmechtig, C., Claustre, H., Poteau, A., and D'Ortenzio, F. (2018). Bio-Argo quality control manual for the Chlorophyll-A concentration. doi: 10.13155/35385.

Schuback, N., Tortell, P. D., Berman-Frank, I., Campbell, D. A., Ciotti, A., Courtecuisse, E., et al. (2021). Single-Turnover Variable Chlorophyll Fluorescence as a Tool for Assessing Phytoplankton Photosynthesis and Primary Productivity: Opportunities, Caveats and Recommendations. *Frontiers in Marine Science* 8. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2021.690607 [Accessed July 20, 2022].

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., et al. (2016). Prediction of the Export and Fate of Global Ocean Net Primary Production: The EXPORTS Science Plan. *Frontiers in Marine Science* 3. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2016.00022 [Accessed January 30, 2023].

Sigman, D. M., and Boyle, E. A. (2000). Glacial/interglacial variations in atmospheric carbon dioxide. *Nature* 407, 859–869. doi: 10.1038/35038000.

Six, C., Thomas, J. C., Brahamsha, B., Lemoine, Y., and Partensky, F. (2004). Photophysiology of the marine cyanobacterium Synechococcus sp. WH8102, a new model organism. *Aquatic Microbial Ecology* 35, 17–29. doi: 10.3354/ame035017.

Sosik, H. M., and Olson, R. J. (2007). Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography: Methods* 5, 204–216. doi: 10.4319/lom.2007.5.204.

Sosik, H. M., Olson, R. J., Neubert, M. G., Shalapyonok, A., and Solow, A. R. (2003). Growth rates of coastal phytoplankton from time-series measurements with a submersible flow cytometer. *Limnology and Oceanography* 48, 1756–1765. doi: 10.4319/lo.2003.48.5.1756.

Staehr, P. A., Henriksen, P., and Markager, S. (2002). Photoacclimation of four marine phytoplankton species to irradiance and nutrient availability. *Marine Ecology Progress Series* 238, 47–59. doi: 10.3354/meps238047.

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., et al. (2013). "Technical summary," in *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press), 33–115.

Stramski, D., Reynolds, R. A., Kahru, M., and Mitchell, B. G. (1999). Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing. *Science* 285, 239–242. doi: 10.1126/science.285.5425.239.

Strzepek, R. F., Boyd, P. W., and Sunda, W. G. (2019). Photosynthetic adaptation to low iron, light, and temperature in Southern Ocean phytoplankton. *Proceedings of the National Academy of Sciences* 116, 4388–4393. doi: 10.1073/pnas.1810886116.

Strzepek, R. F., Hunter, K. A., Frew, R. D., Harrison, P. J., and Boyd, P. W. (2012). Iron-light interactions differ in Southern Ocean phytoplankton. *Limnology and Oceanography* 57, 1182–1200. doi: 10.4319/lo.2012.57.4.1182.

Stukel, M. R., Décima, M., Selph, K. E., Taniguchi, D. A. A., and Landry, M. R. (2013). The role of Synechococcus in vertical flux in the Costa Rica upwelling dome. *Progress in Oceanography* 112–113, 49–59. doi: 10.1016/j.pocean.2013.04.003.

Sun, B., Kattawar, G. W., Yang, P., Twardowski, M. S., and Sullivan, J. M. (2016). Simulation of the scattering properties of a chain-forming triangular prism oceanic diatom. *Journal of Quantitative Spectroscopy and Radiative Transfer* 178, 390–399. doi: 10.1016/j.jqsrt.2016.02.035.

Sverdrup, H. U. (1953). On Conditions for the Vernal Blooming of Phytoplankton. *ICES Journal of Marine Science* 18, 287–295. doi: 10.1093/icesjms/18.3.287.

Taillandier, V., Wagener, T., D'Ortenzio, F., Mayot, N., Legoff, H., Ras, J., et al. (2018). Hydrography and biogeochemistry dedicated to the Mediterranean BGC-Argo network during a cruise with RV <i>Tethys 2</i> in May 2015. *Earth Syst. Sci. Data* 10, 627–641. doi: 10.5194/essd-10-627-2018.

Terrats, L., Claustre, H., Cornec, M., Mangin, A., and Neukermans, G. (2020). Detection of Coccolithophore Blooms With BioGeoChemical-Argo Floats. *Geophysical Research Letters* 47, e2020GL090559. doi: 10.1029/2020GL090559.

Thibodeau, P. S., Roesler, C. S., Drapeau, S. L., Prabhu Matondkar, S. G., Goes, J. I., and Werdell, P. J. (2014a). Locating Noctiluca miliaris in the Arabian Sea: An optical proxy approach. *Limnology and Oceanography* 59, 2042–2056. doi: 10.4319/lo.2014.59.6.2042.

Thibodeau, P. S., Roesler, C. S., Drapeau, S. L., Prabhu Matondkar, S. G., Goes, J. I., and Werdell, P. J. (2014b). Locating *Noctiluca miliaris* in the Arabian Sea: An optical proxy approach. *Limnology and Oceanography* 59, 2042–2056. doi: 10.4319/lo.2014.59.6.2042.

Thyssen, M., Grégori, G. J., Grisoni, J.-M., Pedrotti, M. L., Mousseau, L., Artigas, L. F., et al. (2014). Onset of the spring bloom in the northwestern Mediterranean Sea: influence of environmental pulse events on the in situ hourly-scale dynamics of the phytoplankton community structure. *Frontiers in Microbiology* 5. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2014.00387 [Accessed February 3, 2023].

Trenberth, K. E., Fasullo, J. T., and Balmaseda, M. A. (2014). Earth's Energy Imbalance. *Journal of Climate* 27, 3129–3144. doi: 10.1175/JCLI-D-13-00294.1.

Trombetta, T., Vidussi, F., Roques, C., Scotti, M., and Mostajir, B. (2020). Marine Microbial Food Web Networks During Phytoplankton Bloom and Non-bloom Periods: Warming Favors Smaller Organism Interactions and Intensifies Trophic Cascade. *Frontiers in Microbiology* 11. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2020.502336 [Accessed May 19, 2022].

Twardowski, M. S., Boss, E., Macdonald, J. B., Pegau, W. S., Barnard, A. H., and Zaneveld, J. R. V. (2001). A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case

I and case II waters. *Journal of Geophysical Research: Oceans* 106, 14129–14142. doi: 10.1029/2000JC000404.

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

Uitz, J., Stramski, D., Baudoux, A.-C., Reynolds, R. A., Wright, V. M., Dubranna, J., et al. (2010). Variations in the optical properties of a particle suspension associated with viral infection of marine bacteria. *Limnology and oceanography* 55, 2317–2330.

Uitz, J., Stramski, D., Reynolds, R. A., and Dubranna, J. (2015). Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. *Remote Sensing of Environment* 171, 58–74. doi: 10.1016/j.rse.2015.09.027.

Uitz, J. U., Huot, Y., Bruyant, F., Babin, M., and Claustre, H. (2008). Relating phytoplankton photophysiological properties to community structure on large scales. *Limnology and Oceanography* 53, 614–630. doi: 10.4319/lo.2008.53.2.0614.

Utermöhl, H. (1958). Methods of collecting plankton for various purposes are discussed. *SIL Communications, 1953-1996* 9, 1–38. doi: 10.1080/05384680.1958.11904091.

Verity, P. G., Robertson, C. Y., Tronzo, C. R., Andrews, M. G., Nelson, J. R., and Sieracki, M. E. (1992). Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton. *Limnology and Oceanography* 37, 1434–1446. doi: 10.4319/lo.1992.37.7.1434.

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., and Marty, J.-C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research: Oceans* 106, 19939–19956. doi: 10.1029/1999JC000308.

Visintini, N., Martiny, A. C., and Flombaum, P. (2021). Prochlorococcus, Synechococcus, and picoeukaryotic phytoplankton abundances in the global ocean. *Limnology and Oceanography Letters* 6, 207–215. doi: 10.1002/lol2.10188.

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

Westberry, T. K., Dall'Olmo, G., Boss, E., Behrenfeld, M. J., and Moutin, T. (2010). Coherence of particulate beam attenuation and backscattering coefficients in diverse open ocean environments. *Optics Express* 18, 15419–15425.

Xing, X., Claustre, H., Blain, S., D'Ortenzio, F., Antoine, D., and Guinet, C. (2012a). Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the Kerguelen region (Southern Ocean). 13.

Xing, X., Claustre, H., Blain, S., D'Ortenzio, F., Antoine, D., Ras, J., et al. (2012b). Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the Kerguelen region (Southern Ocean): Quenching correction for chlorophyll fluorescence. *Limnology and Oceanography: Methods* 10, 483–495. doi: 10.4319/lom.2012.10.483.