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

Flavien Petit[1,2], Julia Uitz[1], Louison Dufour[3,4], Collin Roesler[5], Frédéric Partensky[3], Laurence Garczarek[3], Priscillia Gourvil[6], Céline Dimier[7], Melek Golbol[7,8], Vincenzo Vellucci[7,9], David Antoine[1,10] Christophe Penkerc'h[1,11], Vincent Taillandier[1], Hervé Claustre[1]

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

2National Oceanography Centre, Southampton, UK

3Sorbonne Université, CNRS, UMR7144 Adaptation and Diversity in the Marine Environment (AD2M),ECOMAP team, Station Biologique de Roscoff (SBR), 29680 Roscoff, France.

4Centre Algatech, Institute of Microbiology of the Czech Academy of Sciences, Novohradská 237, Třeboň1437901, Czech Republic

5Department of Earth and Oceanographic Science, Bowdoin College, Brunswick, ME, USA

6Sorbonne Université, CNRS FR2424, Roscoff Culture Collection (RCC), Station Biologique de Roscoff (SBR), 29680 Roscoff, France.

7 Sorbonne Université, CNRS, Institut de la Mer de Villefranche, IMEV, F-06230 Villefranche-sur-Mer, France

8 Sorbonne Université, MNHN, CNRS, IRD, Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques, LOCEAN, F-75005 Paris, France

9 Sorbonne Université, CNRS, OSU Stations Marines, STAMAR, 4 Place Jussieu, F-75252 Paris, France

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

11 INSU Division Technique (DT-INSU), UAR 855, CNRS, Plouzané, France

Correspondence to: Flavien Petit (flavien.petit@noc.ac.uk)

Abstract.

Phytoplankton community composition is a key determinant of ocean biogeochemical cycles, yet its observation from autonomous platforms remains challenging. In this study, we assessed the potential of *in situ* multispectral excitation fluorescence (MXF) to discriminate phytoplankton assemblages in the Northwestern Mediterranean Sea, with a view toward applications on Biogeochemical-Argo (BGC-Argo) profiling floats. Laboratory measurements on ten phytoplankton strains confirmed that MXF ratios at 440, 470, and 532 nm provide taxon-specific signatures, especially for picocyanobacteria and green algae. Field observations of phytoplankton pigments were clustered into four ecologically distinct phytoplankton communities along the seasonal cycle. A machine learning model was then trained to classify these clusters using MXF and additional bio-optical indices. Results show that existing BGC-Argo configurations (single-wavelength fluorescence, particulate backscattering and beam attenuation coefficients) reliably distinguish broad community structures such as pico- versus microphytoplankton dominance, but resolving finer pigment-based differences requires the added spectral information of MXF. The different excitation channels contributed unequally: 440 and 470 nm provided robust pigment sensitivity across communities, while 532 nm was particularly informative for detecting phycoerythrin- and chlorophyll *b*–rich taxa. Overall, combining MXF with bio-optical proxies improved classification performance by integrating pigment-specific and size-structure information, demonstrating the potential of MXF to enhance autonomous monitoring of phytoplankton community dynamics and their role in ocean biogeochemical cycles.

1 Introduction

Phytoplankton play a key role in global biogeochemical cycles, particularly in the carbon cycle. They fix dissolved inorganic carbon through photosynthesis and transfer a portion of it to higher trophic levels, initiating the biological carbon pump. This mechanism is pivotal in regulating the ocean's carbon storage. However, primary production, i.e., the rate at which phytoplankton produce organic carbon, varies significantly across different time and space scales. This variation is attributed to environmental changes that induce changes in phytoplankton community structure and biomass (Rousseau and Gregg, 2014). Therefore, in the current context of climate change, monitoring phytoplankton dynamics on a global scale is crucial.

The emergence of new observation platforms, such as BioGeoChemical-Argo (BGC-Argo) profiling floats equipped with miniaturized bio-optical sensors, offers the possibility to collect continuous vertical profiles of optical measurements that serve as proxies of biogeochemical variables (Biogeochemical-Argo Planning Group, 2016; Claustre et al. 2020). Fluorescence is a widely used proxy of chlorophyll-*a* (Chla) concentration, a ubiquitous pigment in phytoplankton organisms, which is in turn used as an indicator of phytoplankton biomass. Equipped with (single channel) fluorometers, autonomous platforms thus allow the observation of phytoplankton biomass variability across a wide range of spatial and temporal scales (e.g. Boss et al., 2008; Barbieux et al., 2019; Cornec et al., 2021; Bock et al., 2022).

Information on phytoplankton biomass only is however insufficient to understand the links between phytoplankton and the carbon cycle. Indeed, the composition of phytoplankton communities is known to be a critical determinant of the carbon cycle since several key processes largely vary between phytoplankton size classes or phylogenetic groups, such as CO_2 fixation through photosynthesis (Cermeño et al., 2005; Uitz et al., 2008), trophic interactions (Cushing, 1989; Finkel, 2007), elemental cycling (Morel 2008, Litchman et al., 2015), or carbon transfer to the deep ocean (Buesseler et al., 1998; Guidi et al., 2009; Henson et al., 2012; Bonnet et al., 2023).

However, the composition of phytoplankton communities cannot be measured directly from the sensors currently implemented on BGC-Argo floats. Only a few methods have been proposed so far to overcome this challenge and go beyond the mere estimation of Chla biomass from bio-optical measurements of BGC-Argo floats. Specifically, Sauzède et al. (2015) developed a neural network algorithm using the vertical shape of the *in-situ* fluorescence profile as input to retrieve the relative contribution to the Chla of the three phytoplankton size classes (pico-, nano- and microphytoplankton). Cetinić et al. (2015) proposed a simple community index based on the ratio of the fluorescence signal to the particulate backscattering coefficient. Similarly, Terrats et al. (2020) used this ratio to detect coccolithophore blooms. Finally, Rembauville et al. (2017) developed a regional approach to estimate the stock of particulate organic carbon (POC) of bacteria and three phytoplankton size classes. This approach has been developed for applications to BGC-Argo floats that measure not only the fluorescence and particulate backscattering coefficient, but also the beam attenuation coefficient. Yet, while those methods provide useful information about phytoplankton community composition, they mostly rely on regional empirical relationships between phytoplankton community composition and bio-optical indices.

Multispectral excitation fluorescence (MXF) is an alternative approach to retrieve information on the relative pigment composition of the phytoplankton assemblage, from which major taxa can be discriminated. MXF consists in measuring *in-situ* fluorescence signals in response to excitation at different wavelengths, corresponding to the absorption peaks of different accessory pigments used as biomarkers of specific taxa in the phytoplankton community (e.g. Yentsch and Phinney, 1985; Bricaud et al., 2004; Brewin et al., 2014). A combination of three wavebands centred around 440, 470 and 532 nm was previously investigated in freshwater environments (Proctor and Roesler, 2010) and in the Arabian Sea (Thibodeau et al., 2014), providing promising results and paving the way for use in open ocean waters. The present study aims at extending the approach of Proctor and Roesler (2010) and assessing the potential of *in-situ* MXF as a proxy of phytoplankton taxonomic composition in view of future applications to BGC-Argo floats.

Although standard BGC-Argo floats are all equipped with a single-channel fluorometer (with excitation at 470 nm) and a backscattering sensor (e.g. Bittig et al., 2019), part of the BGC-Argo fleet is also instrumented with additional sensors that may provide useful information on phytoplankton composition indicators, such as a beam transmissometer (Rembauville et al. 2017) or a dual-channel fluorometer with excitation at 440 and 470 nm (see https://vocab.nerc.ac.uk/collection/R27/current/). In this context, the present study aims to evaluate the potential of bio-optical measurements from the sensors currently implemented on BGC-Argo profiling floats, either alone or in combination with MXF, for retrieving information on phytoplankton community composition. Hence, we designed a predictive machine learning model to assess the ability of MXF combined with other bio-optical observations to infer taxonomic information.

For this purpose, we combined laboratory experiments and fieldwork conducted in the Northwestern (NW) Mediterranean Sea. The NW Mediterranean Sea provides a good case study because its pronounced seasonal phytoplankton biomass cycle (e.g. D'Ortenzio et al., 2005; Lavigne et al., 2015) and ecological succession (Vidussi et al. 2001; Marty et al. 2002) are comparable to those observed in the temperate regions of the open ocean.

For the laboratory work, we selected ten phytoplankton strains representative of the various taxa observed along the seasonal succession of the NW Mediterranean Sea. We measured the MXF response of each strain under controlled conditions, and measured the variability among taxa and strains. This allowed us to characterize the MXF sensor and validate the analysis of the field data.

During our fieldwork, we collected concomitant pigment concentrations and bio-optical parameters using a sensor package comprising a MXF sensor, a single-channel fluorometer, a backscatterometer and a transmissometer over an annual cycle in the NW Mediterranean Sea. This *in-situ* dataset was then used to develop and test a phytoplankton community composition discrimination model. Ultimately, we provide recommendations for the use of MXF, alone or in combination with other bio-optical indicators, to infer the taxonomic composition of phytoplankton communities from BGC-Argo profiling float observations.

2. Material & methods

2.1. Laboratory work

2.1.1. Phytoplankton strains and culture conditions

For laboratory experiments, we selected ten phytoplankton strains provided by the Roscoff Culture Collection (RCC; https://roscoff-culture-collection.org/). These strains are representative of the taxonomic diversity of the main eukaryotic and prokaryotic phytoplankton organisms encountered in open-ocean waters, and particularly in the NW Mediterranean Sea. The selected strains include three diatom species, one pelagophyte, one dinoflagellate and five photosynthetic prokaryotes (three *Synechococcus* and two *Prochlorococcus* strains; Table 1).

All strains were grown at a constant temperature of 21°C, under 50 µmol photons m^{-2} s^{-1} continuous white light provided by a white-blue-green LED system (Alpheus, France), and in either K+Si (Keller et al., 1987) or PCR-S11 culture medium (Rippka et al., 2000) for eukaryotes and prokaryotes, respectively. As fluorescence is significantly influenced by the physiology of phytoplankton cells, we used cultures in stable physiological status as assessed by a high PSII quantum yield (F_v/F_M) using a Phyto-PAM-II fluorometer (Walz, Effeltrich, Germany). The F_v/F_M parameter was calculated as $(F_M - F_0)/F_M$ (Pittera et al., 2014), where F_0 is the dark-adapted basal fluorescence, F_M is the maximal fluorescence associated with the closing of photosynthetic reaction centres, and F_v is the variable fluorescence. F_M was measured after exposure to saturating light pulses and addition of 100 µM of the photosystem II inhibitor 3′-(3,4-dichlorophenyl)-1′,1′-dimethylurea (DCMU; Parkhill et al., 2001). The F_v/F_M parameter was measured concomitantly to cell counts made using a Guava EasyCyte flow cytometer (Luminex Corporation, USA) all over the growth of each phytoplankton culture (Marie et al., 2001). The MXF protocol (see Section 2.1.2) was applied to each culture in the middle to late exponential growth phase,

https://doi.org/10.5194/egusphere-2025-6418
Preprint. Discussion started: 7 January 2026

just before the drop of the F_V/F_M index. The MXF protocol was repeated three times on distinct replicate culture vessels for
120 each strain (biological triplicates).

Table 1: Name, taxonomy, pigment composition as detected by HPLC pigment analysis, and size class of the ten phytoplankton strains used for the laboratory experiments. 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
125 for low-light adapted. The pigments measured are Chlorophyll a (Chla), Fucoxanthin (Fuco), Diadinoxanthin (Diad), Diatoxanthin (Diat), 19'-HF (19'-Hexanoyloxyfucoxanthin), 19'-BF (19'-Butanoyloxyfucoxanthin), Peridinin (Peri), Zeaxanthin (Zea), Divinyl-chlorophyll a and b (DV-Chla, DV-Chlb), Chlorophyll c1 and c2 or c3 (Chlc1 + c2, Chlc3).

Species name	**Class**	**RCC strain number**	**Other names**	**Main pigments**	**Size class**
Conticribra (Thalassiosira) weissflogii	Mediophyceae (diatom)	RCC76	CCMP1336	Fuco, Chlc1 + c2, Diad, Diat	Micro
Chaetoceros diadema	Mediophyceae (diatom)	RCC1717	RA080513-06	Fuco, Diad, Chlc1 + c2	Micro
Pelagomonas calceolata	Pelagophyceae	RCC100	CCMP1214	Chlc3, Chlc1 + c2, 19'-BF, Fuco, Diad, Diat	Nano
Scrippsiella sp.	Dinophyceae	RCC3006	VFAC24-3	Chlc1 + c2, Peri, Diad	Nano
Minidiscus sp.	Mediophyceae (diatom)	RCC4213	MACUMBA-SC18	Chlc1 + c2, Fuco, Diad, Diat	Nano
Prochlorococcus marinus (LL)	Cyanophyceae	RCC156	SS120-04/95	Zea, DV-Chlb, DV-Chla	Pico
Prochlorococcus marinus (HL)	Cyanophyceae	-	PCC 9511	Zea, DV-Chlb, DV-Chla	Pico
Synechococcus sp.	Cyanophyceae	RCC2319	MINOS11	Zea	Pico
Synechococcus sp.	Cyanophyceae	RCC2374	A15-62	Zea	Pico
Synechococcus sp.	Cyanophyceae	RCC2379	BOUM118	Zea	Pico

2.1.2. Multispectral fluorescence measurements

All MXF measurements were performed using an ECO 3X1M fluorometer (Sea-Bird electronics, USA), with three excitation wavebands centred onto 440, 470 and 532 nm, and emission onto 695 nm, with a 10-nm bandwidth. To determine the fluorescence to Chla slope factor (here expressed in fluorescence per Chla unit), the MXF measurements were collected for each culture over a 5-point dilution series ranging from 0.1 to 10 mg Chla m^{-3}. Each culture was dark acclimated for 2 h before dilution and MXF measurements. The MXF measurements were performed immediately after dilution to avoid any dilution-induced physiological stress. The ECO 3X1M sensor outputs were recorded with the TeraTerm® software. Each culture was diluted in a 1L glass beaker that was then placed under constant slow stirring. The multispectral fluorometer was placed at the centre of the beaker and the optical window was immersed 5 mm below the surface. Blank measurements were performed with culture media and were then subtracted from the culture measurements to remove any possible fluorescence signal from colour dissolved organic matter. Blank values were within a few counts of the dark reading, indicating that the measurements were not subject to optical interferences from the beaker edge or benchtop scattering. For each culture triplicate, on each dilution, we measured the fluorescence response during three series of one minute of continuous acquisition at 1 Hz, each separated by two minutes of darkness. The signal did not decrease significantly during acquisition, indicating that there was no quenching during the protocol application.

2.1.3. Laboratory data processing

For each of the ten selected phytoplankton strains grown in culture, the MXF measurements were processed as follows. First, for each dilution series, a blank value was recorded by measuring the average response of the culture medium alone and was then subtracted from the raw sensor output acquired for the culture as described above. The MXF measurements collected over the three consecutive acquisition periods were averaged to obtain a single fluorescence value, expressed in digital counts (DC). Finally, the dilution series was used to define a Chla-specific calibration value, expressed in units of DC (mg Chla m^{-3})$^{-1}$, for each of the three excitation wavelengths and each of the ten selected strains. This calibration value represents the coefficient of a linear regression performed between the fluorescence response at a given wavelength expressed in DC and the Chla concentration in mg m^{-3} for the entire dilution range and for each replicate of a given phytoplankton strain. For the sake of simplicity, the raw fluorescence signal, in DC, at an excitation wavelength λ, will be noted as F_{λ} (i.e., F_{440}, F_{470}, and F_{532} for the excitation wavelengths 440 nm, 470 nm, and 532 nm, respectively) and the Chla-specific calibration value will be noted as F^*_{λ} (i.e., F^*_{440}, F^*_{470}, and F^*_{532}).

2.2. Field measurements of bio-optical and biogeochemical variables

2.2.1. Sampling strategies

Concomitant phytoplankton pigment determinations and bio-optical measurements were performed at sea every month, from December 2020 to October 2021 at the BOUSSOLE station (Buoy for the acquisition of long-term optical time series), a long-term monitoring site located at 7°54′E, 43°22′N in the Ligurian (NW Mediterranean) Sea (Antoine et al., 2008). On each monthly cruise (GOLBOL Melek, VELLUCCI Vincenzo, ANTOINE David (2000) BOUSSOLE, https://doi.org/10.18142/1), a CTD-rosette equipped with an optical sensor package was used to perform casts from the surface down to 400 m depth. The optical package included an MXF sensor (the same ECO 3X1M as used for the laboratory experiments), an ECO FLBB sensor and a C-Rover beam transmissometer (both Sea-Bird Scientific). The ECO FLBB measures the Chla fluorescence at one excitation (470 nm) and one emission (695 nm) wavelengths, as well as the particulate backscattering coefficient at 700 nm ($b_{bp,}$ see section 2.2.2). We note that among the three excitation channels of the ECO 3X1M (440, 470, and 532 nm), the first two (440 and 470 nm) are shared with dual-channel fluorometers (Sea-Bird Scientific ECO FLBBFL, RBR Tridente) now implemented on some BGC-Argo floats. This correspondence allows us to also test the potential of dual-channel fluorometers for inferring phytoplankton composition indicators in our analysis. The C-Rover transmissometer measures the light beam transmitted between the emitter and receptor (at 650 nm and over an optical path length of 25 cm), allowing the calculation of the attenuation coefficient. From this, the particulate beam attenuation coefficient (c_p) is calculated by removing the attenuation of pure seawater. Both sensors have already been mounted on several BGC-Argo floats for different biogeochemical applications (e.g. Mignot et al., 2014; Rembauville et al., 2017; Barbieux et al., 2022). Concomitantly, seawater was sampled from the Niskin bottles attached to the CTD-rosette at ten discrete depths for pigment identification and quantification by High-Performance Liquid Chromatography (HPLC).

2.2.2 In-situ data processing

The factory-determined dark value of the ECO FLBB sensor was validated in the laboratory using black tape to cover the optical window, then subtracted from the raw DC following BGC-Argo data management recommendations (Schmechtig et al., 2018a). The optical backscattering coefficient was measured during the CTD-rosette upcast, which was used for seawater sampling. The angular scattering coefficient (β) was recorded every second at a central angle of 124° and a wavelength of 700 nm. To obtain the particulate angular scattering coefficient (β_p), the contribution of pure seawater, dependent on temperature and salinity (Zhang et al., 2009), was subtracted from β. The β_p coefficient was then converted into b_{bp} following standard conversion guidelines and applying a χ factor of 1.076 (Schmechtig et al., 2018b).

The ECO 3X1M sensor was mounted on the CTD-rosette frame, providing simultaneous MXF measurements. The same black tape procedure was used to subtract dark values. The raw fluorescence values, expressed as counts, were directly used as the fluorescence signal.

The particulate attenuation coefficient was corrected for sensor drift and calculated from total beam transmittance as in Barnes & Antoine 2014.

The outliers in fluorescence, c_p and b_{bp} datasets were detected and removed using a threshold of 1.5 simple moving average (Δdepth = 3 m). Each profile was then smoothed using a simple moving average (Δdepth = 3 m).

2.3. Determination of phytoplankton pigments

For both laboratory and field samples, Chla and accessory pigments were identified and quantified by HPLC analysis. Briefly, seawater from discrete field samples or cultures was filtered onto glass fibre filters (GF/F Whatman 25 mm), stored in liquid nitrogen during cruises and 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 by sonication in 100% methanol, clarified by filtration (GF/F Whatman 25 mm), and finally separated and quantified by HPLC. More details about the HPLC analytical protocol can be found in Ras et al. (2008). The total chlorophyll *a* concentration, [Chla], is defined as the sum of chlorophyll *a*, divinyl-chlorophyll *a* and chlorophyllid *a* concentrations.

For *in-situ* samples, we specifically investigated the distribution of seven diagnostic pigments (DP) : peridinin (Peri), 19'-butanoyloxyfucoxanthin (19-BF), fucoxanthin (Fuco), 19'-hexanoyloxyfucoxanthin (19-HF), alloxanthin (Allo), zeaxanthin (Zea), divinyl-chlorophyll *b* (DV-Chlb), and chlorophyll *b* (Chlb), with total chlorophyll *b* (TChlb) defined as the sum of DV-Chlb and Chlb, as well as divinyl-chlorophyll *a* (DV-Chla). These pigments are defined as biomarkers of major phytoplankton taxa and were further grouped into three phytoplankton size classes, i.e. micro- (>20 µm), nano- (2-20 µm) and picophytoplankton (<2 µm), according to the approach of Claustre (1994) and Vidussi et al. (2001). Following the equations given in Uitz et al. (2006), the DP-based method allowed the estimation of the relative contribution to the [Chla] of the 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 (Chase et al., 2020), this approach has been shown to provide reliable, quantitative information for analyses at large spatial and temporal scales (e.g., Vidussi et al., 2001; Bricaud et al., 2004; Uitz et al., 2006; Brewin et al., 2014).

2.4. Statistical analyses

To explore the potential of deriving phytoplankton community composition from MXF and other bio-optical measurements, we employed a multi-step analytical approach. First, phytoplankton pigment data were analyzed using Correspondence Analysis (CA) to identify similarities in pigment composition among samples. The first three dimensions of the CA were then subjected to a Hierarchical Ascending Classification (HAC) to define distinct clusters, corresponding to typical phytoplankton assemblages observed in the NW Mediterranean at the BOUSSOLE site (following Kramer and Siegel, 2019;

Uitz et al., 2023). These clusters were subsequently used as categorical targets to evaluate the ability of MXF and bio-optical descriptors to infer phytoplankton composition through a Histogram Gradient Boosting (HGB) classification model.

This clustering-based strategy offers two key advantages: (1) it transforms the prediction task from a regression to a classification problem, eliminating biomass effects and allowing the model to focus solely on taxonomic composition; and (2) it ensures that the model captures the most significant source of taxonomic variability in the dataset, thereby reducing the influence of minor variance components and mitigating the risk of overfitting.

2.4.1. Correspondence analysis

We used correspondence analysis (CA) to visualize the main similarities among samples based on their relative pigment concentrations. This method generates linear combinations of relative pigment contributions, creating a new multidimensional space where sample projections reflect their resemblance in pigment composition.

In this transformed space, samples with similar pigment signatures are located close to each other, while those with distinct compositions are further apart. The first dimensions of the analysis capture the most significant variance in the dataset, effectively summarizing the dominant patterns in pigment distribution. By reducing data complexity, this approach provides a clearer interpretation of how different phytoplankton communities are structured based on their pigment signatures.

We applied this statistical method for two key objectives. First, we used it to compare the pigment composition of cultured phytoplankton strains with that of field samples, assessing whether the selected strains accurately represent the seasonal phytoplankton succession at the sampled location in the NW Mediterranean Sea. This CA was performed using the seven DP concentrations. Field samples were projected in the transformed CA projection space, and the distance between field and laboratory samples will be discussed. Second, we conducted a CA exclusively on field samples, using the same set of pigments, and extracted the first three components as inputs for our clustering method described hereafter.

2.4.2. Clustering of phytoplankton pigment data

The pigment data from NW Mediterranean field samples, consisting of the set of pigment concentrations detailed in the previous section, were clustered to identify major phytoplankton assemblages along the seasonal cycle using the CA (cf. Section 2.4.1). The first three dimensions of the CA were used to quantify the resemblance in pigment composition across samples. A Hierarchical Ascending Classification (HAC) was then applied to these three dimensions, grouping samples based on their relative pigment composition rather than absolute pigment concentrations. The resulting cluster dendrogram was cut at a height of 20, minimising the intra-cluster variance, and yielding three initial clusters.

Given the distinct pigment composition of prokaryotic picophytoplankton versus micro-/nanophytoplankton communities, we repeated the clustering after excluding picophytoplankton-dominated samples to refine the classification within the micro-/nanophytoplankton group. This second clustering step divided the micro-/nanophytoplankton samples into two additional clusters. In total, four distinct phytoplankton communities were identified.

2.4.3. Classification of phytoplankton groups based on MXF and additional bio-optical proxies

Here we evaluated the possibility of using MXF measurements alone, or in combination with other bio-optical proxies measured by BGC-Argo floats to retrieve information on phytoplankton community composition. For this purpose, measurements of F_{440}, F_{470}, F_{532}, b_{bp} and c_p were used as inputs of a model aiming at predicting the four different clusters identified with the method described in the previous section (cf. Section 2.5.2). We tested its performance using different sets of inputs, corresponding to either already deployed or feasible BGC-Argo sensor combinations, and varying levels of prediction task complexity (i.e., different number of clusters to predict).

The classification of *in-situ* phytoplankton communities (i.e., the prediction of a categorical target variable) based on MXF and additional bio-optical measurements was performed using a Histogram Gradient Boosting (HGB) algorithm. This type of machine learning model is particularly well suited for tabular data, where each sample (row) is characterized by a consistent set of features (columns), and the dataset contains a relatively low number of observations (Chen and Guestrin, 2016; Shwartz-Ziv and Armon, 2022).

The model's performance depends not only on the total number of observations but also on the distribution of observations across target classes (here, phytoplankton community clusters). An imbalance in class representation can bias the model toward the dominant class, leading to an artificial overestimation of its performance for that group while reducing accuracy for underrepresented classes. To mitigate this imbalance, we applied the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002), which generates synthetic samples for minority classes to improve classification fairness and overall model performance. In the end, each cluster was represented by 32 samples.

Because the dataset consists of a time series of a single year of phytoplankton community succession, the phytoplankton biomass was strongly correlated with the community composition. Since the five different measured variables (i.e., F_{440}, F_{470}, F_{532}, b_{bp}, c_p) are significantly correlated with phytoplankton biomass, we used biomass-specific ratios to avoid overfitting and highlight intrinsic optical properties. First, F_{440} and F_{532} were normalized to F_{470}, which is the channel typically used to estimate Chla concentration from single-channel fluorometers. Second, each of the three fluorescence signals (F_{440}, F_{470} and F_{532}) was normalized to b_{bp} and c_p, used as particulate biomass proxies. The refractive index, representative of the composition of the particulate pool and estimated as a function of 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,l 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 random sampling. The classification results can be categorized into four different categories: True Positive (TP) corresponding to the accurate prediction of the presence of a class of a given phytoplankton assemblage, 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:

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

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

The precision can be interpreted as the fraction of positive predictions of the model that were accurate, while the recall can be interpreted as the fraction of positive samples that have been correctly predicted by the model. A precision or recall value of 1 indicates perfect classification, meaning no false positives (for precision) or false negatives (for recall). Conversely, a low precision value suggests a high number of false positives, whereas a low recall value indicates that many actual positive cases were missed by the model.

The performance of the HGB classification model was tested for six different combinations of optical properties (e.g. $F_{440,}$ $F_{470,}$ $F_{532,}$ $c_{p,}$ and b_{bp}) each consistent with potential future applications to BGC-Argo profiling floats (Table 2). Ultimately, we tested the performance of the model in predicting different numbers of clusters, ranging from 4 down to 2, each reflecting a different level of predictive complexity.

Table 2: List of sensor configurations tested in this study for retrieving phytoplankton community composition. For each configuration labelled A to F, we provide the measured variables along with examples of sensors that are available for integration, or already integrated, on BGC-Argo profiling floats.

Sensor configuration	**F440**	**F470**	**F532**	**b_{bp}**	**c_p**	**Combination of sensors**
A	X	X	X	X	X	MXF sensor Backscatterometer Transmissometer
B	X	X		X	X	Dual channel fluorometer Backscatterometer Transmissometer
C	X	X	X		X	MXF sensor Transmissometer
D	X	X		X		Dual channel fluorometer Backscatterometer

E	X	X	X			MXF sensor
F		X		X	X	Single channel fluorometer Backscatterometer Transmissometer

3. Results and discussion

3.1. MXF signal in laboratory-controlled conditions

After quantifying the calibrated fluorescence for each of the ten phytoplankton strains and for each excitation wavelength, we considered two Chla-specific fluorescence ratios F^*_{532} / F^*_{470} and F^*_{440} / F^*_{470}, as described in Proctor and Roesler (2010). Both ratios varied by a factor of approximately 2 when all taxa were considered (Fig. 1). The three *Synechococcus* strains consistently showed high F^*_{532} / F^*_{470} ratios (1.37 +/- 0.1) and low F^*_{440} / F^*_{470} ratios (0.91 +/- 0.06). By contrast, *Prochlorococcus* strains exhibited intermediate F^*_{532} / F^*_{440} and high F^*_{440} / F^*_{470} ratios (1.27 +/- 0.08 and 1.22 +/- 0.07 respectively). The diatom strains showed low to intermediate values (1 +/- 0.05 for both ratios). The dinoflagellates strain had similar average ratios as the diatoms, but replicates were a bit more variable. Finally, the pelagophyceae strains had the lowest F^*_{532} / F^*_{470} and F^*_{440} / F^*_{470} ratios (0.71 +/- 0.02 and 0.95 +/- 0.12, respectively) of all taxa.

The higher F^*_{532} / F^*_{470} values observed for the *Synechococcus* taxon may be explained by their higher fluorescence at 532 nm, induced by the presence of phycoerythrin. Indeed, this phycobiliprotein is systematically found in open ocean *Synechococcus* and binds two chromophores, phycourobilin (λ_{max} ~495 nm) and phycoerythrobilin (λ_{max} ~545 nm), the latter thus being the most excited at 532 nm (Six et al., 2007; Grébert et al. 2018). Consequently, the phycourobilin-rich strain RCC2379 expectedly exhibited a lower average F^*_{532} / F^*_{470} ratio than the two other *Synechococcus* strains both being chromatic acclimaters which, in white light, exhibit a low phycourobilin to phycoerythrobilin ratio (Six et al. 2004; Humily et al. 2013). Although the fairly high F^*_{532} / F^*_{470} ratio observed in *Prochlorococcus* are harder to explain, given the very low amounts of phycoerythrin present in these strains, the higher F^*_{440} / F^*_{470} ratio of the HL-adapted PCC 9511 compared to the LL-adapted RCC156 is consistent with the much higher DV-Chla (λ_{max} ~450 nm) to DV-Chlb (λ_{max} ~475 nm) ratio of the former (Moore et al. 1995). The differences in fluorescence responses among diatoms, dinoflagellates, pelagophyceae are also likely related to their distinct content in accessory chlorophylls, yet these differences

https://doi.org/10.5194/egusphere-2025-6418
Preprint. Discussion started: 7 January 2026

are somewhat subtle and harder to interpret in terms of fluorescence properties (Bidigare et al., 1989; Bricaud et al., 2004). Our results are in line with numerous previous studies that demonstrated that pigment composition influences both light absorption and fluorescence emission spectra in phytoplankton, leading to taxon-specific fluorescence signatures (Yentsch and Menzel, 1963; Johnsen and Sakshaug, 2007; Hu et al., 2010; MacIntyre et al., 2010; Proctor and Roesler, 2010). More specifically, some laboratory studies using monospecific cultures demonstrated that fluorescence spectra vary significantly across taxa when multiple excitation and emission wavelengths are used (Yentsch and Menzel, 1963; Johnsen and Sakshaug, 2007; Poryvkina et al., 1994). More recent work has expanded on these findings by incorporating mixed communities, showing potential for determination of natural assemblages of phytoplankton (Hu et al., 2010; Escoffier et al., 2015). While Hu et al. (2010) included both monospecific and mixed cultures, as well as coastal marine samples, most of these studies were conducted under controlled laboratory conditions rather than in open-ocean environments.

Only a few studies have investigated fluorescence responses in natural, mixed phytoplankton communities (Seppälä and Balode, 1998; Hu et al, 2010; Proctor and Roesler, 2010; Thibodeau et al. 2014). Some of these studies benefited from conditions that enhanced pigment-specific fluorescence signals, such as reversed filtration to increase pigment concentration (Seppälä and Balode, 1998), high resolution of fluorescence excitation/emission spectra (Seppälä and Balode, 1998; Hu et al, 2010) or naturally high chlorophyll concentrations in bloom conditions (Proctor and Roesler, 2010). These conditions contrast sharply with those of the open ocean, where phytoplankton communities are more diverse, pigment concentrations are lower, and taxonomic differentiation based on MXF alone is more challenging.

Despite this complexity, our MXF measurements on mono-specific cultures are promising and coherent with Proctor and Roesler (2010), who further demonstrated that intra-taxon variance in specific fluorescence ratios was lower than inter-taxon variance. Both studies suggest that a multispectral MXF sensor with three excitation channels (440, 470, and 532 nm) can provide sufficient sensitivity to distinguish taxa in controlled conditions even though natural communities consist of mixed assemblages of taxa with complex pigment signature, leading to less contrasted fluorescence response than for laboratory cultures. Therefore, in the following section, we investigate whether it is possible to resolve taxonomic composition from *in-situ* phytoplankton fluorescence signals by analysing a year-long dataset from the NW Mediterranean Sea. Additionally, we assess whether combining MXF measurements with other bio-optical proxies could improve classification performance, in anticipation of future deployment of multispectral fluorometers on BGC-Argo profiling floats.

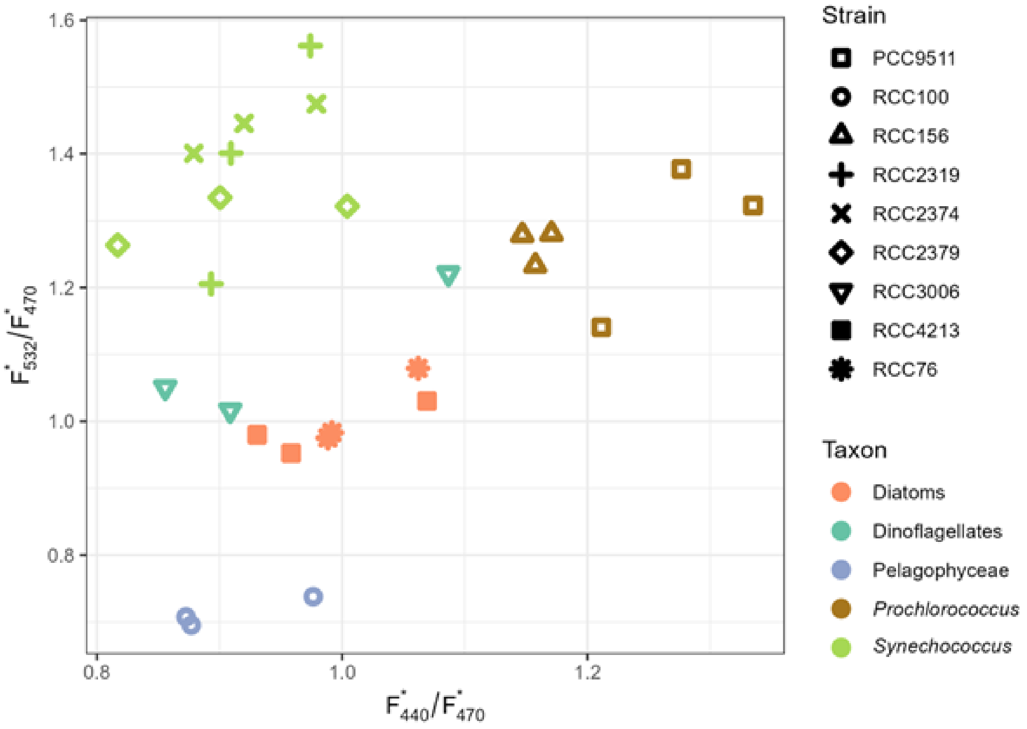

Figure 1: Scatterplot of F*440/F*470 vs. F*532/F*470 ratios for each phytoplankton strain grown in culture. The colour code indicates the taxon to which each strain belongs; the symbols indicate the strains, using the Roscoff Culture Collection code, when available (see Table 1).

3.2. Phytoplankton communities in the Northwestern Mediterranean Sea

Surface chlorophyll concentration at the BOUSSOLE site can reach values up to 5 mg Chla m^{-3} during the spring bloom and below 0.1 mg Chla m^{-3} in summer, along with seasonal changes of the phytoplankton community composition (Marty et al., 2002; Antoine et al. 2020). In winter, pigment indicators of prymnesiophyte, namely 19'-HF and 19'-BF, are observed alongside fucoxanthin, a marker of diatoms, and alloxanthin, a marker of cryptophytes (Marty et al., 2002; Mayot et al., 2017). During the spring phytoplankton bloom, fucoxanthin concentrations increase (Marty et al., 2002), which is likely related to an increase in diatoms. In summer, the mixed layer is phosphate-limited and phytoplankton communities are representative of stratified oligotrophic regions, with a prevalence of pigments specific to pico-phytoplankton, divinyl-chlorophyll-a and zeaxanthin (Marty et al., 2008). This phytoplankton diversity is comparable to what is observed in temperate open-ocean regions (Vidussi et al., 2001; Marty et al., 2002; Lavigne et al., 2015; Mayot et al., 2017).

The composition of the phytoplankton communities in the field samples was analysed using a pigment-based clustering approach (see Section 2.4.2). The clustering allowed grouping samples with similar pigment composition and led to the discrimination of four distinct phytoplankton assemblages (clusters) over the year (Fig. 2a-d). All clusters are dominated by nanophytoplankton but vary significantly in the partitioning between micro- and picophytoplankton (Fig. 2e-h). The first cluster corresponds to winter communities as well as deep autumn communities, with a large proportion of picophytoplankton and a significant contribution of Chlb, a pigment typical of green microalgae, mostly flagellates (Bustillos-Guzmán et al., 1995). The second cluster coincides with the bloom community with a shared contribution of micro- and nanophytoplankton. This assemblage is characterized by a high fucoxanthin contribution, typically associated with diatoms. The third cluster is associated with summer communities located at and below the level of the deep chlorophyll maximum (DCM) and also exhibits a mixed composition of micro- and nanophytoplankton. Finally, the fourth cluster is characteristic of picophytoplankton communities in surface waters from summer to autumn, typically dominated by picocyanobacteria (Barlow et al., 1997).

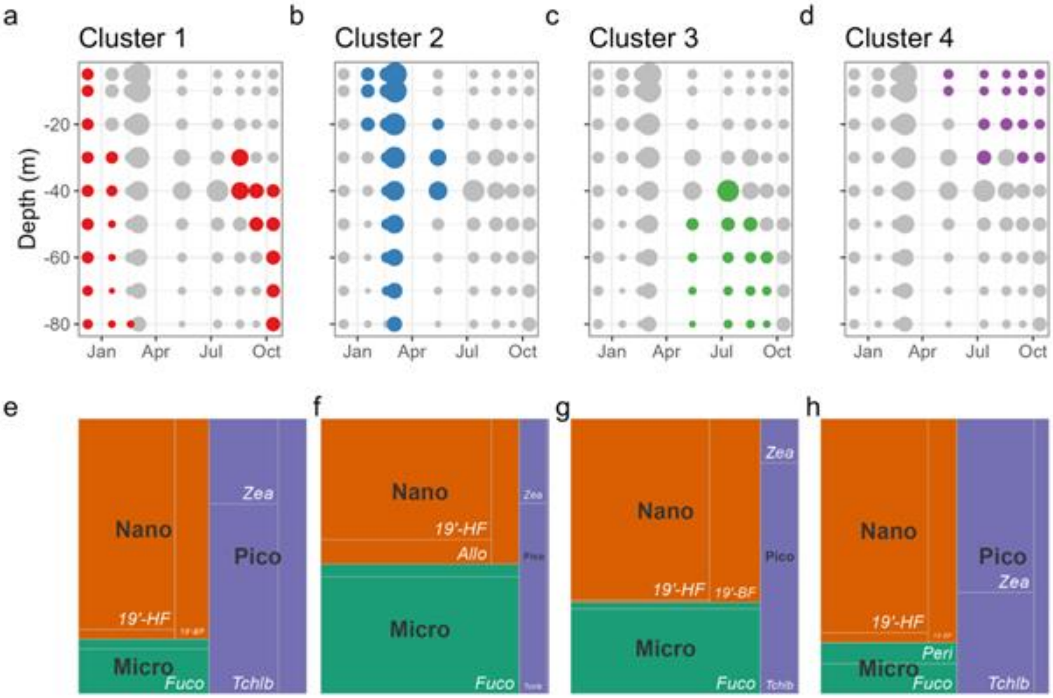

Figure 2: Distribution of the phytoplankton communities as determined from the cluster analysis applied to the field pigment data: (a-d) Vertical distribution of the four pigment-determined clusters indicative of the main phytoplankton communities encountered over an annual cycle in the northwestern Mediterranean Sea (BOUSSOLE site). The size of the dots indicates the Chla concentration, used as a proxy of the phytoplankton biomass. (e-h) Tree map of the relative pigment

concentration of each cluster (areas delimited by the grey lines), with the size class corresponding to the pigment taxa affiliation (colored).

The results from the culture experiments demonstrated that phytoplankton taxa can be distinguished by MXF at the taxonomic level of the Class for eukaryotic phytoplankton (e.g., Diatoms, Dinoflagellates, Pelagophyceae) and Genus for picocyanobacteria (*Synechococcus* and *Prochlorococcus*). We then compared the relative accessory pigment composition of the *in-situ* clusters to that of the five laboratory characterized taxa. This analysis seeks to determine whether the field samples fall in the range of variability of accessory pigment composition of the culture samples.

Similar to the field samples, a correspondence analysis (see Section 2.4.1) was applied to the pigment composition of 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 (Fig. 3). We observe three distinct groups corresponding to the different taxa represented by the ten selected strains. One is composed of Diatoms and Pelagophyceae, while the two others correspond to *Synechococcus* and *Prochlorococcus*, respectively. This highlights a strong contrast in pigment composition between *Synechococcus* which has Zea, *Prochlorococcus* which has DV-Chla, DV-Chlb and Zea, and the nano- and microphytoplankton taxa that share many different pigments (Jeffrey et al., 1997; Veldhuis et al., 2005).

The culture results provide a reference from which the projection space is computed, on which the field samples are projected (Fig. 3). The field data are evenly spread in the centre of the plan, indicating that the variability in the pigment composition of the field samples is similar to that observed in the laboratory cultures. Moreover, the four field-based clusters are fairly well distinguished in the CA. The only exception being a few samples of cluster 1 with high CA2 values, due to the presence of 19’-HF, a pigment that is not found in our cultured strains. However, this pigment has a very similar absorption signature to other carotenoids like 19’-BF or peridinin. We thus expect that the MXF sensor is sensitive enough to discriminate between different phytoplankton groups (clusters) characterized by distinct pigment composition in field samples.

In addition, we note that the *in-situ* samples have lower eigenvalues (i.e., absolute values on CA axes) than the laboratory samples indicating that the pigment variability is less contrasted in the field than in laboratory samples. This is not surprising because of the complexity of the pigment composition in natural samples associated with mixed phytoplankton assemblages, instead of single taxa in monospecific cultures. This characteristic of open-ocean samples could somewhat hamper the possibilities of inferring information on phytoplankton community composition from MXF measurements, a hypothesis that is tested in the next section.

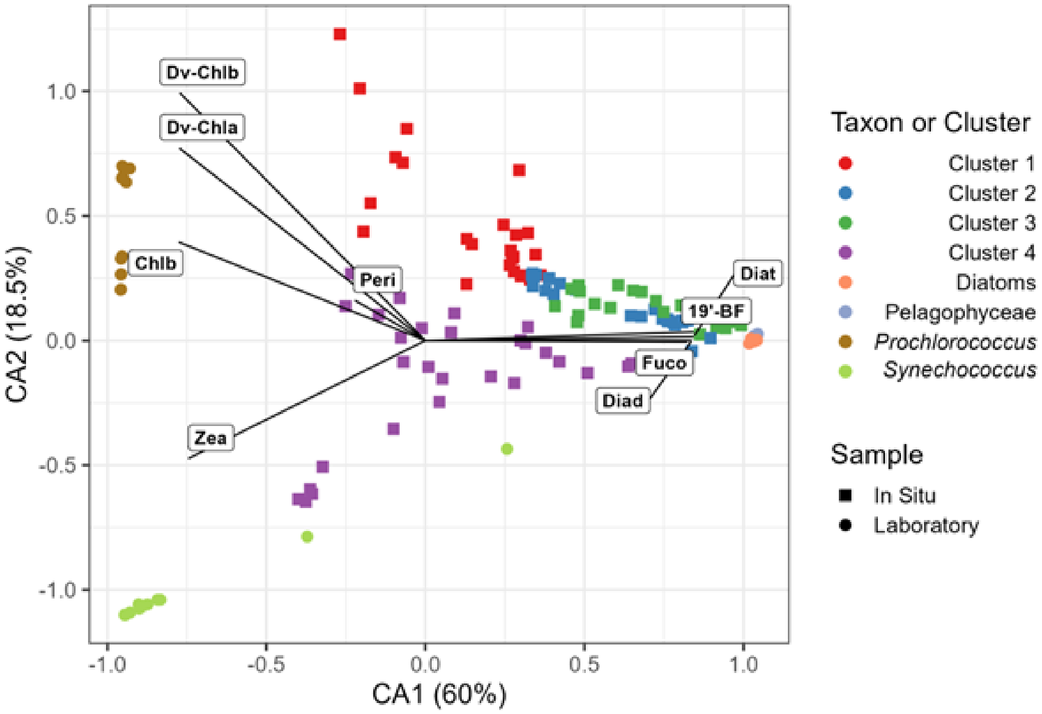

Figure 3: Correspondence analysis of the pigment concentrations of the strains grown in culture. The pigment concentrations 425 measured in the northwestern Mediterranean (BOUSSOLE site) seawater samples are represented using the same colour code as in Figure 2 and projected as supplementary observations

3.3. Discrimination of phytoplankton taxa from in-situ MXF and additional bio-optical variables

The predictive model (see Section 2.4.3) was tested using as inputs measurements from the MXF sensor alone (F_{440}, 430 F_{470}, F_{532}), or in combination with a backscatterometer (F_{440}, F_{470}, F_{532}, b_{bp}) and a transmissometer (F_{440}, F_{470}, F_{532}, b_{bp}, c_p). Additionally, the model was also tested using measurements from a dual excitation channel fluorometer and backscatterometer (F_{440}, F_{470}, b_{bp}), or combined with a transmissometer (F_{440}, F_{470}, b_{bp}, c_p) (Fig. 4 and Table A1). These two configurations are of particular interest as many BGC-Argo floats have been, and will continue to be, deployed with this specific set of sensors.

435 Considering the most comprehensive configuration, A (i.e., F_{440}, F_{470}, F_{532}, b_{bp}, c_p), the precision and recall scores of the HGB model are homogeneous among all four clusters with values between 60% and 85% (+/- 15%) (Fig. 4 and Table A1). These results based on annual sampling in the NW Mediterranean are quite robust, showing sufficient precision to demonstrate that MXF, when used in combination with a backscatterometer and a transmissometer, effectively allow four phytoplankton taxonomic groups to be distinguished. If we remove either the MXF sensor or the transmissometer, leading to measurements 440 of F_{440}, F_{470}, b_{bp} and c_p, or F_{440}, F_{470}, F_{532} and c_p (i.e., configurations B and C, respectively), we observe similar precision and

recall scores, and a marked variability between the clusters (Fig. 4 and Table A1). However, all four clusters displayed a mean precision of 67.25% (+/- 6.9%) and 69% (+/- 12%) for configurations B and C, respectively. When considering configuration D, which corresponds to a sensor load with a dual-channel fluorometer and a backscatterometer (i.e., F_{440}, F_{470} and b_{bp}), the scores are significantly lower with a mean precision of 62% (+/- 8.7%). The use of the MXF sensor only (configuration E, i.e. F_{440}, F_{470} and F_{532}) led to variable performances depending on the cluster. Thus, removing the transmissometer (c_p) or the 532-nm excitation fluorescence channel (F_{532}) seemingly induces a significant decrease in the general accuracy and recall scores of the model (Fig. 4). Finally, the configuration with only single-channel fluorescence measurements and two bio-optical indices (configuration F, i.e. F_{470}, b_{bp}, c_p) led to lower performance for every cluster, except for the picophytoplankton dominated one (i.e., cluster 4; see Fig. 2). This suggests that a configuration without MXF performs better when the phytoplankton communities have contrasted size structures.

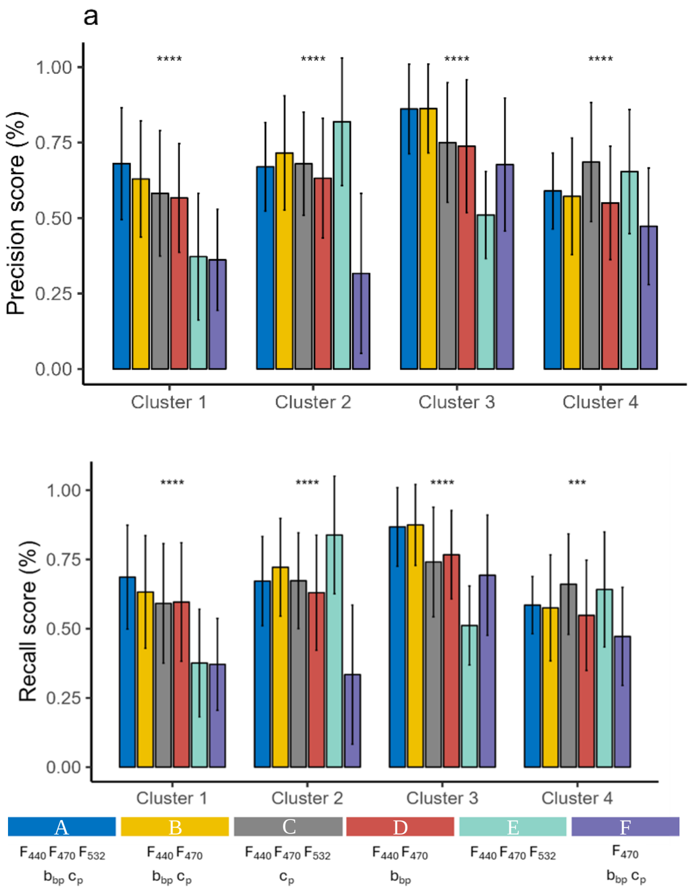

Figure 4. Precision and recall scores for phytoplankton community classification using different sensor package configurations. (A) Precision and (B) recall scores for predicting four phytoplankton clusters using various sensor combinations, either currently equipped on BGC-Argo profiling floats or suitable for future installations: full configuration (F440, F470, F532, bbp, cp); dual channel fluorescence with backscatterometer and transmissometer (F440, F470, bbp, cp); MXF with cp (F440, F470, F532, cp); dual channel fluorometer with backscatterometer (F440, F470, bbp); MXF only (F440, F470, F532) and single channel fluorescence with backscatterometer and transmissometer (F470, bbp, cp). Precision represents the fraction of positive predictions that were correct, while recall indicates the fraction of true positives correctly identified. Error bars reflect variability across cross-validation runs. Significance of paired t.test is indicated with stars (ns: p.value > 0.05 ; *: p.value <= 0.05 ; **: p.value <= 0.01 ; ***: p.value <= 0.001 ; ****: p.value <= 0.0001).

Based on configuration A, which combines MXF and all bio-optical indicators, we can evaluate the predictive power of each descriptor. The mean importance of the descriptors in the discrimination of distinct phytoplankton clusters, as indicated by the impurity metrics (Fig. 5), highlights a significant role of all selected descriptors (b_{bp}/c_p, F_{440}/F_{470}, F_{440}/c_p, F_{532}/F_{470}, F_{532}/b_{bp}, F_{470}/c_p, F_{470}/b_{bp}, F_{440}/b_{bp}), and particularly of b_{bp}/c_p, F_{440}/F_{470}, F_{440}/c_p and F_{532}/F_{440} ratios. The high importance of the b_{bp}/c_p ratio was expected as it has been already described as an indicator of the particle size distribution (Dall'Olmo et al., 2009; Slade and Boss, 2015; Organelli et al., 2020). Interestingly, the second and fourth strongest predictive descriptors correspond to the fluorescence ratios from the MXF sensor (F_{440}/F_{470} and F_{532}/F_{470}). This indicates the strong added value of the MXF descriptors in the predictive model, in line with our laboratory results. These results also support the findings of both Proctor and Roesler (2010) and Thibodeau et al. (2014) who showed that the F_{440}/F_{470} and F_{532}/F_{470} fluorescence ratios are related to the taxonomic composition of phytoplankton communities. Finally, we note that the commonly used phytoplankton community index, F_{470}/b_{bp} ratio (Cetinić et al., 2015; Lacour et al., 2019; Terrats et al., 2020), has a notably lower predictive robustness in our analysis. Therefore, the use of new sensor configurations, including a multiple excitation channel fluorometer and a beam attenuation transmissometer implemented on BGC-Argo floats, may have a stronger potential than previously described methods to detect seasonal succession of phytoplankton groups, with markedly different pigment composition or cell sizes (here picophytoplankton-dominated vs. microphytoplankton-dominated communities).

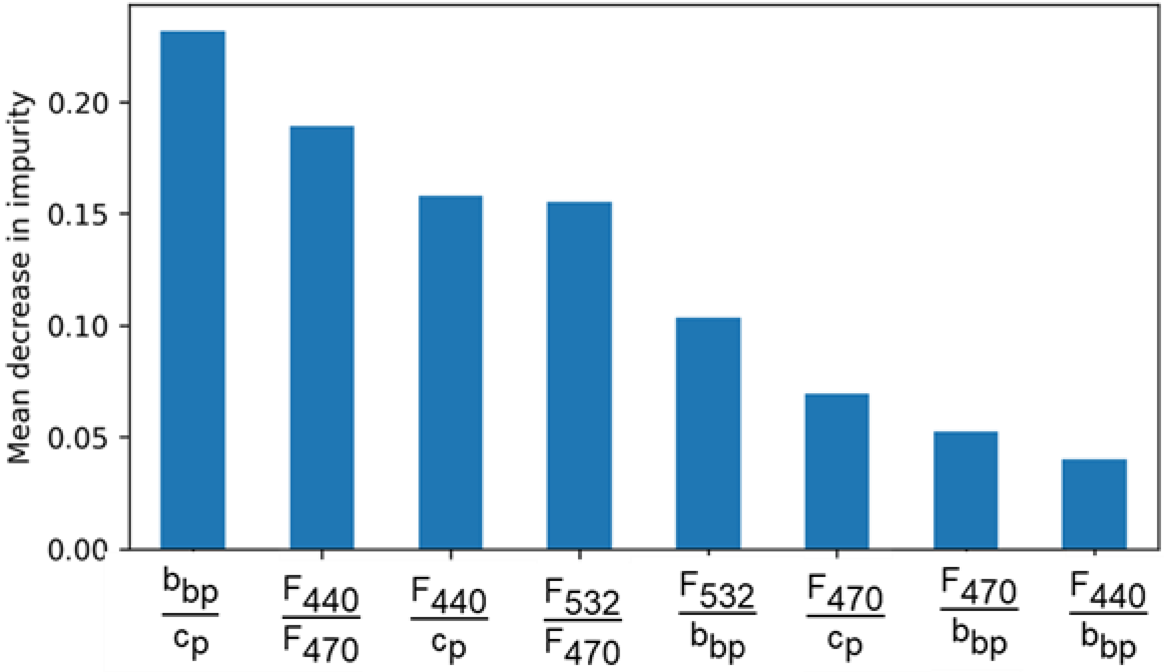

Figure 5: Importance of the different descriptors in the classification model, expressed as the mean decrease in impurity. The mean decrease in impurity reflects how much each descriptor contributes to improving the purity of the splits in the decision

tree. A higher value indicates that the descriptor plays a more significant role in distinguishing (i.e., pigment-based clusters) in the model.

3.4. Model tuning to predict phytoplankton communities

In Section 3.3, we considered four clusters found along the annual phytoplankton succession. We now evaluate the same model and hyperparameters for predicting two or three clusters (instead of four), to assess whether reducing the level of predictive complexity improves the model performances under different sensor configurations. First, we reduced the number of clusters from four to three by merging Clusters 2 and 3, the two clusters with the smallest distance, representing mixed nano- and microphytoplankton-dominated communities. The main distinction between these two clusters lies in their dominant pigments. Cluster 2 is indeed characterized by a higher relative contribution of alloxanthin, while Cluster 3 is dominated by 19’-HF. The three resulting clusters correspond to the surface summer picophytoplankton community associated with large concentrations of zeaxanthin (Cluster 4), the winter and deep summer picophytoplankton community with Chlb and DV-Chlb (Cluster 1), and the mixed micro- and nanophytoplankton community (Clusters 2 and 3 grouped together). Second, we reduced the initial number of four clusters to two, by merging Cluster 1 (dominated by Chlb-containing picophytoplankton) and Cluster 4 (dominated by zeaxanthin-containing picophytoplankton) to a picophytoplankton-dominated cluster essentially grouping *Synechococcus*, *Prochlorococcus* and chlorophytes. The second cluster consists of the previously merged cluster (initially Clusters 2 and 3) composed of micro- and nanophytoplankton.

Like the approach presented in Section 3.3, we evaluated the classification model performances for different sensor configurations using the mean balanced recall metrics, which reflects the percentage of correctly classified samples. The same cross-validation method was applied to ensure consistency (Fig. 6). We first observed that a decrease in prediction complexity (specifically, a reduction in the number of clusters) does not consistently lead to an improvement in the classification model performance, and that the effect of such a reduction depends on the sensor configuration. When all sensors are included (configuration A), the model performs best in predicting 2 or 3 clusters than 4 clusters, achieving above 75% recall. Similarly, the model using as inputs two fluorescence wavelengths, as well as the b_{bp} and c_p coefficients (i.e., configuration B) achieves above 75% recall when predicting 2 or 3 clusters.

It is worth noting that the model using as inputs bio-optical measurements from sensors already implemented on some BGC-Argo floats (i.e, configuration F with single-channel fluorescence, b_{bp} and c_p) demonstrates strong predictive capabilities for 2 or 3 clusters, achieving recall values of 72% and 74%, respectively. In the two-cluster prediction scenario, one cluster predominantly comprises picophytoplankton, such as *Synechococcus* and *Prochlorococcus*, while the other includes a mix of nano- and microphytoplankton. Interestingly, previous studies have shown that the combination of F_{470}, b_{bp}, and c_p effectively correlate with phytoplankton community size structure (Veldhuis et al., 2005; Brewin et al. 2011; Cetinić et al. 2015; Sauzède et al. 2015; Rembauville et al. 2017, Terrats et al. 2023) supporting our findings. In the three-cluster scenario, the model has to distinguish between two different types of picophytoplankton communities rather than one. These communities exhibit markedly different photoacclimation profiles, with deep-water communities displaying a F_{470}/b_{bp} ratio significantly distinct from that of surface communities (Bellacicco et al. 2016; Graff et al. 2016). Thus, the model successfully discriminates phytoplankton communities according to their average size and photoacclimation status. However, when the nano- and microphytoplankton-dominated cluster is further divided into two distinct communities based on varying carotenoid composition, the model performance declines markedly, resulting in a recall score of only 40%. This low score, compared to the high predictive performance of the other models, that all include multiple fluorescence with at least two wavelengths (i.e., configurations A to E), highlights the importance of MXF for pigment-based remote classification of phytoplankton communities.

When using MXF data only (i.e., configuration E), the model recall performance drops to approximately 50%, whatever the number of predicted clusters. In comparison, models conFig.d with the MXF sensor and a transmissometer (i.e., configurations C) reach around 65% recall, regardless of the amount of cluster to be predicted. Adding the particulate backscattering coefficient (b_{bp}, configuration A) slightly increases the recall score to around 75% when predicting 2 or 3 clusters. This highlights the importance of incorporating additional bio-optical indices with MXF, such as b_{bp} and c_p, to enhance the phytoplankton group classification model. Furthermore, when the backscatterometer and transmissiometer are present, the MXF 532-nm fluorescence channel does not improve the model performance (configuration A and B). This indicates that, overall, when bio-optical sensors measuring b_{bp} and c_p can be used as inputs, only two fluorescence wavelengths are needed to achieve an optimal classification of phytoplankton communities.

These findings demonstrate that meaningful phytoplankton taxonomic information can be retrieved from MXF signals. If the standard BGC-Argo configuration (i.e., configuration F) with one single-channel fluorescence and two optical indices (b_{bp} and c_p) shows good prediction performances for predicting two and three clusters, adding one more fluorescence wavelength (i.e., configurations A to E) significantly improves the performance when predicting four clusters. Thus, our results demonstrate that MXF is a promising avenue for remote classification of phytoplankton community composition.

Our results also highlight that the predictive skill of individual descriptors, particularly the 532 nm fluorescence channel, depends on both the way clusters are defined and on the ecological context in which communities occur. When clusters retain a strong contribution of Chlb–containing taxa (e.g., chlorophytes) or phycoerythrin-rich picocyanobacteria (*Synechococcus*), the F_{532}/F_{470} ratio provides valuable information on community composition. However, when such groups are merged into broader assemblages, the influence of F_{532} diminishes and the model relies more heavily on size- or biomass-sensitive proxies such as c_p and b_{bp}. Ecologically, this reflects the fact that pigment-based contrasts are strongest when communities differ in accessory pigment composition, while size-structure proxies dominate when communities are merged across pigment gradients. From a broader perspective, this sensitivity also explains why the usefulness of F_{532} will vary geographically: in regions where *Synechococcus* or green flagellates are recurrent and occasionally abundant (e.g., coastal upwelling systems), including 532 nm excitation is expected to significantly improve classification performance (Morel, 1997; Saito et al., 2005). In contrast, in persistently oligotrophic waters such as the subtropical gyres the added value of F_{532} is likely reduced.

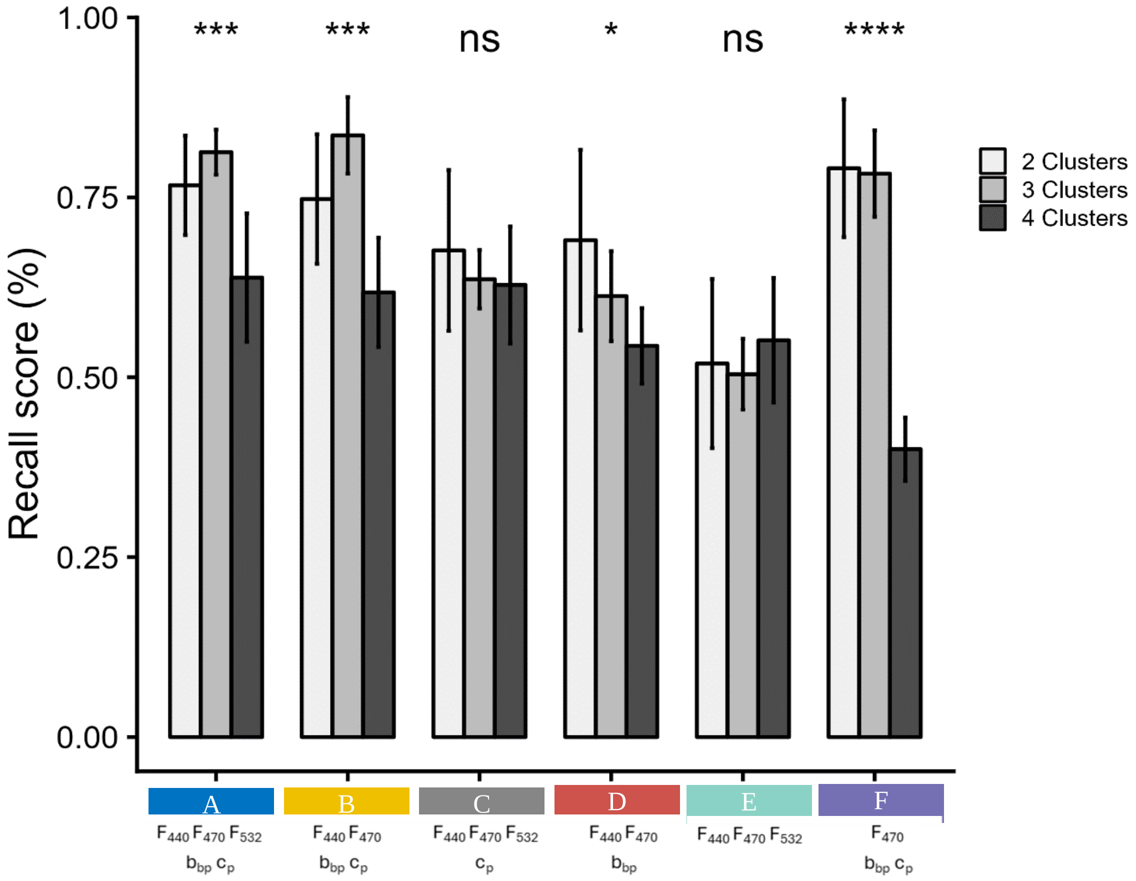

Figure 6: Values of the mean weighted recall resulting from cross-validation with different numbers of clusters and different sensor combinations, either currently equipped on BGC-Argo profiling floats or feasible for future deployments: full configuration - A (F440, F470, F532, bbp, cp); dual-channel fluorescence with backscatterometer and transmissometer - configuration B (F440, F470, bbp, cp); MXF with cp - configuration C (F440, F470, F532, cp); dual-channel fluorometer with backscatterometer - configuration D (F440, F470, bbp); MXF only - configuration E (F440, F470, F532); and single-channel fluorescence with backscatterometer and transmissometer - configuration F (F470, bbp,cp). Significance of paired t.test is indicated with stars (ns: p.value > 0.05; *: p.value <= 0.05 ; **: p.value <= 0.01 ; ***: p.value <= 0.001 ; ****: p.value <= 0.0001).

4. Conclusion and perspectives

Monitoring phytoplankton community composition is essential for understanding marine biogeochemical cycles, particularly those related to oceanic carbon dynamics. However, this remains challenging in open-ocean environments due to the limitations of current autonomous sensor technologies and the complex bio-optical signature of mixed phytoplankton assemblages. In this study, we developed and tested a machine learning (HGB) model to classify phytoplankton assemblages using *in-situ* MXF and additional bio-optical measurements. The goal was to assess the potential of MXF for taxonomic discrimination of phytoplankton communities from autonomous platforms, particularly BGC-Argo profiling floats. The HGB model uses MXF signals, specifically fluorescence excited at 440, 470, and 532 nm, alongside particulate backscattering (b_{bp}) and beam attenuation (c_p) coefficients as input features. Taxonomic information was assessed through pigment-based clustering of field samples collected over a full annual cycle in the NW Mediterranean Sea.

Our primary objective was to evaluate the predictive skill of MXF measurements. Under laboratory conditions, MXF provided sufficient sensitivity to distinguish five key phytoplankton groups. When applied to field data, MXF alone had lower performance due to the presence of complex mixed phytoplankton assemblages, compared to monospecific cultures. However, combining MXF with additional bio-optical indices improved discrimination of phytoplankton communities.

We assessed the model's performance across different sensor configurations, reflecting realistic BGC-Argo float payloads, and across varying levels of classification complexity, (i.e., number of predicted pigment-based clusters). While existing sensor configurations, such as single-wavelength fluorescence with b_{bp} and c_p, performed well for simple phytoplankton communities (two or three clusters), full separation of more complex community types (four clusters) required the richer spectral information provided by MXF. Notably, MXF enabled taxonomic discrimination within micro- and nanophytoplankton-dominated assemblages, even when using only two fluorescence channels, 440 and 470 nm. In contrast, bio-optical properties such as b_{bp} and c_p, which are robust proxies for particle size and biomass, were more effective at distinguishing between communities with pronounced size differences, such as pico- versus microphytoplankton.

Overall, our results suggest that integrating MXF into BGC-Argo platforms could significantly enhance our ability to monitor shifts in phytoplankton communities, particularly among groups with overlapping size distributions. This advancement would provide valuable insights into ecosystem dynamics and the role of phytoplankton taxonomic composition across broad spatial and temporal scales, enabled by the autonomous sampling capabilities of BGC-Argo floats, in open-ocean biogeochemical cycles.

Future work should aim to generalize these findings across diverse oceanic regions and refine the predictive model to account for a broader variety of phytoplankton taxa. Building a standardized dataset linking phytoplankton community composition, from pico- to micro-size classes, to MXF and bio-optical measurements will be crucial for improving our understanding of phytoplankton dynamics through autonomous observations.

https://doi.org/10.5194/egusphere-2025-6418
Preprint. Discussion started: 7 January 2026

Appendix A

Table A1 : Mean and standard deviation of precision and recall scores from the cross validation (repeated prediction of clusters from the machine learning model using randomly picked samples for training and testing). The standard deviation is indicated in parentheses. Each line corresponds to a specific combination of sensors, either currently deployed on BGC-Argo profiling floats or suitable for future deployments: full configuration (F_{440}, F_{470}, F_{532}, b_{bp}, c_p); dual channel fluorescence with backscatterometer and transmissometer (F_{440}, F_{470}, b_{bp}, c_p); MXF with backscatterometer (F_{440}, F_{470}, F_{532}, b_{bp}); dual channel fluorometer with backscatterometer (F_{440}, F_{470}, b_{bp}); MXF only (F_{440}, F_{470}, F_{532}) and single channel fluorescence with backscatterometer and transmissometer (F_{470}, b_{bp}, c_p).

	Configuration	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
		Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
$F_{440} + F_{470} + F_{532} + b_{bp} + c_p$	A	0.68 (0.15)	0.68 (0.18)	0.67 (0.15)	0.67 (0.16)	0.86 (0.14)	0.86 (0.14)	0.58 (0.12)	0.58 (0.1)
$F_{440} + F_{470} + b_{bp} + b_{bp}$	B	0.58 (0.2)	0.59 (0.21)	0.68 (0.17)	0.67 (0.17)	0.75 (0.19)	0.74 (0.19)	0.68 (0.2)	0.66 (0.18)
$F_{440} + F_{470} + F_{532} + c_p$	C	0.62 (0.19)	0.63 (0.2)	0.71 (0.19)	0.72 (0.17)	0.86 (0.14)	0.87 (0.14)	0.57 (0.19)	0.57 (0.19)
$F_{440} + F_{470} + b_{bp}$	D	0.56 (18)	0.59 (0.21)	0.63 (0.19)	0.63 (0.2)	0.74 (0.22)	0.76 (0.15)	0.55 (0.18)	0.54 (0.2)
$F_{440} + F_{470} + F_{532}$	E	0.37 (0.2)	0.37 (0.19)	0.81 (0.21)	0.83 (0.21)	0.51 (0.14)	0.51 (0.14)	0.65 (0.2)	0.64 (0.2)
$F_{470} + b_{bp} + c_p$	F	0.36 (0.26)	0.37 (0.16)	0.32 (0.26)	0.33 (0.25)	0.68 (0.22)	0.69 (0.21)	0.47 (0.19)	0.47 (0.17)

Code, data, or code and data availability

All the code used to process raw data, perform statistical analysis and make fgures shown in that paper are available on github https://github.com/Flavi1P/mf. Data are available upon request to the corresponding author.

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

Flavien Petit (corresponding author) designed the study, performed data acquisition and analysis, drafted and wrote the manuscript. Julia Uitz, Hervé Claustre, Colin Roesler, Frédéric Partensky and Laurence Garczarek were involved in the design of the study, data analysis and manuscript writing. Louison Dufour, Priscillia Gourvil, Céline Dimier, Vincenzo Vellucci, Christophe Penkerc'h and David Antoine were involved in the data acquisition and analysis and manuscript writing. All authors commented on the final manuscript.

Competing interests

This work does not present any conflict of interest.

Disclaimer

Copernicus Publications adds a standard disclaimer: "Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher."
Please feel free to add disclaimer text at your choice, if applicable.

Acknowledgements

This paper represents a contribution to the following projects: OBOO funded by CNRS INSU via the LEFE-CYBER program; EFFicACY funded by the French "Agence Nationale de la Recherche" (ANR- 19- ce02- 0019); REFINE funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N° 834177); Argo-2030 with support from the French government within the framework of the "Investissements d'Avenir" program (PIA EquipEx+) integrated within France 2030, and managed by ANR under the reference "ANR-21-ESRE-0019"; and BOUSSOLE funded by CNES and ESA (contract N° 4000119096/17/I-BG) with support of the French Oceanographic Fleet. FP was funded by a PhD grant from Sorbonne Université (Ecole Doctorale 129) and a research scientist position funded by the National Oceanography Centre.

Financial support

Support funds and grant agreement numbers are listed as specified upon manuscript registration and reported to FundRef upon publication.

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Antoine, D., P. Guevel, J.-F. Desté, G. Bécu, F. Louis, A. J. Scott, and P. Bardey. 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

Barbieux, M. and others. 2019. 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. and others. 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

Barlow, R. G., R. F. C. Mantoura, D. G. Cummings, and T. W. Fileman. 1997. Pigment chemotaxonomic distributions of phytoplankton during summer in the western Mediterranean. Deep Sea Research Part II: Topical Studies in Oceanography **44**: 833–850. doi:10.1016/S0967-0645(96)00089-6

Bellacicco, M., G. Volpe, S. Colella, J. Pitarch, and R. Santoleri. 2016. Influence of photoacclimation on the phytoplankton seasonal cycle in the Mediterranean Sea as seen by satellite. Remote Sensing of Environment **184**: 595–604. doi:10.1016/j.rse.2016.08.004

Bidigare, R. R., J. H. Morrow, and D. A. Kiefer. 1989. Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea. J Mar Res **47**: 323–341. doi:10.1357/002224089785076325

Biogeochemical-Argo Planning Group. 2016. The scientific rationale, design and implementation plan for a Biogeochemical-Argo float array, Ifremer.

Bittig, H. C. and others. 2019. A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. Front. Mar. Sci. **6**. doi:10.3389/fmars.2019.00502

Bock, N., M. Cornec, H. Claustre, and S. Duhamel. 2022. Biogeographical Classification of the Global Ocean From BGC-Argo Floats. Global Biogeochemical Cycles **36**: e2021GB007233. doi:10.1029/2021GB007233

Bonnet, S. and others. 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

Boss, E., W. S. Pegau, M. Lee, M. Twardowski, E. Shybanov, G. Korotaev, and F. Baratange. 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., D. Swift, L. Taylor, P. Brickley, R. Zaneveld, S. Riser, M. J. Perry, and P. G. Strutton. 2008. Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite. Limnology and Oceanography **53**: 2112–2122. doi:10.4319/lo.2008.53.5_part_2.2112

Brewin, R. J. W. and others. 2011. 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., S. Sathyendranath, P. K. Lange, and G. Tilstone. 2014. 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

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

Buesseler, K., L. Ball, J. Andrews, C. Benitez-Nelson, R. Belastock, F. Chai, and Y. Chao. 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

Bustillos-Guzmán, J., H. Claustre, and C. Marty. 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., E. Marañón, J. Rodríguez, and E. Fernández. 2005. 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

Cetinić, I., M. J. Perry, E. D'Asaro, N. Briggs, N. Poulton, M. E. Sieracki, and C. M. Lee. 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

Chase, A. P., S. J. Kramer, N. Haëntjens, E. S. Boss, L. Karp-Boss, M. Edmondson, and J. R. Graff. 2020. Evaluation of diagnostic pigments to estimate phytoplankton size classes. Limnology and Oceanography: Methods **18**: 570–584. doi:10.1002/lom3.10385

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. jair **16**: 321–357. doi:10.1613/jair.953

Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Proceedings of the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. 785–794.

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

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

Cornec, M. and others. 2021. Deep Chlorophyll Maxima in the Global Ocean: Occurrences, Drivers and Characteristics. Global Biogeochemical Cycles **35**: e2020GB006759. doi:10.1029/2020GB006759

Cremella, B., Y. Huot, and S. Bonilla. 2018. Interpretation of total phytoplankton and cyanobacteria fluorescence from cross-calibrated fluorometers, including sensitivity to turbidity and colored dissolved organic matter. Limnology & Ocean Methods **16**: 881–894. doi:10.1002/lom3.10290

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

Dall'Olmo, G., T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade. 2009. Direct contribution of phytoplankton-sized particles to optical backscattering in the open ocean.doi:10.5194/bgd-6-291-2009

D'Ortenzio, F., D. Iudicone, C. de Boyer Montegut, P. Testor, D. Antoine, S. Marullo, R. Santoleri, and G. Madec. 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

El Hourany, R., M. Abboud-Abi Saab, G. Faour, C. Mejia, M. Crépon, and S. Thiria. 2019. Phytoplankton Diversity in the Mediterranean Sea From Satellite Data Using Self-Organizing Maps. Journal of Geophysical Research: Oceans **124**: 5827–5843. doi:10.1029/2019JC015131

Escoffier, N., C. Bernard, S. Hamlaoui, A. Groleau, and A. Catherine. 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

Finkel, Z. V. 2007. CHAPTER 15 - Does Phytoplankton Cell Size Matter? The Evolution of Modern Marine Food Webs, p. 333–350. *In* P.G. Falkowski and A.H. Knoll [eds.], Evolution of Primary Producers in the Sea. Academic Press.

Frank, H. A., and R. J. Cogdell. 2012. 8.6 Light Capture in Photosynthesis, p. 94–114. *In* E.H. Egelman [ed.], Comprehensive Biophysics. Elsevier.

Graff, J. R., T. K. Westberry, A. J. Milligan, M. B. Brown, G. DallOlmo, K. M. Reifel, and M. J. Behrenfeld. 2016. Photoacclimation of natural phytoplankton communities. Marine Ecology Progress Series **542**: 51–62. doi:10.3354/meps11539

Grébert, T. and others. 2018. Light color acclimation is a key process in the global ocean distribution of Synechococcus cyanobacteria. Proceedings of the National Academy of Sciences **115**: E2010–E2019. doi:10.1073/pnas.1717069115

Guidi, L., L. Stemmann, G. A. Jackson, F. Ibanez, H. Claustre, L. Legendre, M. Picheral, and G. Gorskya. 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

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

Holtrop, T. and others. 2021. Vibrational modes of water predict spectral niches for photosynthesis in lakes and oceans. Nat Ecol Evol **5**: 55–66. doi:10.1038/s41559-020-01330-x

Hu, X., R. Su, F. Zhang, X. Wang, H. Wang, and Z. Zheng. 2010. Multiple excitation wavelength fluorescence emission spectra technique for discrimination of phytoplankton. J. Ocean Univ. China **9**: 16–24. doi:10.1007/s11802-010-0016-x

Humily, F., F. Partensky, C. Six, G. K. Farrant, M. Ratin, D. Marie, and L. Garczarek. 2013. A Gene Island with Two Possible Configurations Is Involved in Chromatic Acclimation in

Marine Synechococcus F. Rodriguez-Valera [ed.]. PLoS ONE **8**: e84459. doi:10.1371/journal.pone.0084459

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

Johnsen, G., and E. Sakshaug. 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. Journal of Phycology **43**: 1236–1251. doi:10.1111/j.1529-8817.2007.00422.x

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

Kramer, S. J., and D. A. Siegel. 2019. How Can Phytoplankton Pigments Be Best Used to Characterize Surface Ocean Phytoplankton Groups for Ocean Color Remote Sensing Algorithms? Journal of Geophysical Research: Oceans **124**: 7557–7574. doi:10.1029/2019JC015604

Lacour, L., N. Briggs, H. Claustre, M. Ardyna, and G. Dall'Olmo. 2019. The Intraseasonal Dynamics of the Mixed Layer Pump in the Subpolar North Atlantic Ocean: A Biogeochemical-Argo Float Approach. Global Biogeochemical Cycles **33**: 266–281. doi:10.1029/2018GB005997

Lavigne, H., F. D'Ortenzio, M. Ribera D'Alcalà, H. Claustre, R. Sauzède, and M. Gacic. 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., P. de Tezanos Pinto, K. F. Edwards, C. A. Klausmeier, C. T. Kremer, and M. K. Thomas. 2015. Global biogeochemical impacts of phytoplankton: a trait-based perspective. Journal of Ecology **103**: 1384–1396. doi:10.1111/1365-2745.12438

Litchman, E., P. de Tezanos Pinto, C. A. Klausmeier, M. K. Thomas, and K. Yoshiyama. 2010. Linking traits to species diversity and community structure in phytoplankton, p. 15–28. *In* L. Naselli-Flores and G. Rossetti [eds.], Fifty years after the "'Homage to Santa Rosalia'": Old and new paradigms on biodiversity in aquatic ecosystems. Springer Netherlands.

MacIntyre, H. L., E. Lawrenz, and T. L. Richardson. 2010. Taxonomic Discrimination of Phytoplankton by Spectral Fluorescence, p. 129–169. *In* D.J. Suggett, O. Prášil, and M.A. Borowitzka [eds.], Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications. Springer Netherlands.

Marie, D., F. Partensky, D. Vaulot, and C. Brussaard. 2001. Enumeration of Phytoplankton, Bacteria, and Viruses in Marine Samples. Current Protocols in Cytometry **10**: 11.11.1-11.11.15. doi:10.1002/0471142956.cy1111s10

https://doi.org/10.5194/egusphere-2025-6418
Preprint. Discussion started: 7 January 2026

Marty, J.-C., J. Chiavérini, M.-D. Pizay, and B. Avril. 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

Marty, J.-C., N. Garcia, and P. Raimbault. 2008. Phytoplankton dynamics and primary production under late summer conditions in the NW Mediterranean Sea. Deep Sea Research Part I: Oceanographic Research Papers **55**: 1131–1149. doi:10.1016/j.dsr.2008.05.001

Mayot, N. and others. 2017. Influence of the Phytoplankton Community Structure on the Spring and Annual Primary Production in the Northwestern Mediterranean Sea: PHYTOPLANKTON DYNAMICS IN THE NWM. Journal of Geophysical Research: Oceans **122**: 9918–9936. doi:10.1002/2016JC012668

Mena, C., P. Reglero, M. Hidalgo, E. Sintes, R. Santiago, M. Martín, G. Moyà, and R. Balbín. 2019. Phytoplankton Community Structure Is Driven by Stratification in the Oligotrophic Mediterranean Sea. Frontiers in Microbiology **10**.

Mignot, A., H. Claustre, J. Uitz, A. Poteau, F. D'Ortenzio, and X. Xing. 2014. Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation. Global Biogeochemical Cycles **28**: 856–876. doi:10.1002/2013GB004781

Moore, L., R. Goericke, and S. Chisholm. 1995. Comparative physiology of Synechococcus and Prochlorococcus: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. Mar. Ecol. Prog. Ser. **116**: 259–275. doi:10.3354/meps116259

Morel, A. 1997. Consequences of a *Synechococcus* bloom upon the optical properties of oceanic (case 1) waters. Limnology & Oceanography **42**: 1746–1754. doi:10.4319/lo.1997.42.8.1746

Morel, F. M. M. 2008. The co-evolution of phytoplankton and trace element cycles in the oceans. Geobiology **6**: 318–324. doi:10.1111/j.1472-4669.2008.00144.x

Organelli, E., G. Dall'Olmo, R. J. W. Brewin, F. Nencioli, and G. A. Tarran. 2020. Drivers of spectral optical scattering by particles in the upper 500 m of the Atlantic Ocean. Opt. Express, OE **28**: 34147–34166. doi:10.1364/OE.408439

Parkhill, J.-P., G. Maillet, and J. J. Cullen. 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

Pittera, J., F. Humily, M. Thorel, D. Grulois, L. Garczarek, and C. Six. 2014. Connecting thermal physiology and latitudinal niche partitioning in marine Synechococcus. ISME J **8**: 1221–1236. doi:10.1038/ismej.2013.228

Poryvkina, L., S. Babichenko, S. Kaitala, H. Kuosa, and A. Shalapjonok. 1994. Spectral fluorescence signatures in the characterization of phytoplankton community composition. Journal of Plankton Research **16**: 1315–1327. doi:10.1093/plankt/16.10.1315

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

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

Rembauville, M. and others. 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

Reynolds, C. S. 2006. The Ecology of Phytoplankton, Cambridge University Press.

Rippka, R. and others. 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

Roesler, C. S., and E. Boss. 5 In Situ Measurement of the Inherent Optical Properties (IOPs) and Potential for Harmful Algal Bloom Detection and Coastal Ecosystem Observations.

Rousseaux, C. S., and W. W. Gregg. 2014. Interannual Variation in Phytoplankton Primary Production at A Global Scale. Remote Sensing **6**: 1–19. doi:10.3390/rs6010001

Saito, M. A., G. Rocap, and J. W. Moffett. 2005. Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. Limnol. Oceanogr. **50**: 279–290. doi:10.4319/lo.2005.50.1.0279

Sauzède, R., H. Claustre, C. Jamet, J. Uitz, J. Ras, A. Mignot, and F. D'Ortenzio. 2015. 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

Schmechtig, C., H. Claustre, A. Poteau, and F. D'Ortenzio. 2018a. Bio-Argo quality control manual for the Chlorophyll-A concentration, Ifremer.

Schmechtig, C., A. Poteau, H. Claustre, F. D'Ortenzio, G. Dall'Olmo, and E. Boss. 2018b. Processing Bio-Argo particle backscattering at the DAC level, Ifremer.

Seppälä, J., and M. Balode. 1998. The use of spectral fluorescence methods to detect changes in the phytoplankton community, p. 207–217. *In* T. Tamminen and H. Kuosa [eds.], Eutrophication in Planktonic Ecosystems: Food Web Dynamics and Elemental Cycling: Proceedings of the Fourth International PELAG Symposium, held in Helsinki, Finland, 26–30 August 1996. Springer Netherlands.

Shima, S., R. P. Ilagan, N. Gillespie, B. J. Sommer, R. G. Hiller, F. P. Sharples, H. A. Frank, and R. R. Birge. 2003. Two-Photon and Fluorescence Spectroscopy and the Effect of Environment on the Photochemical Properties of Peridinin in Solution and in the Peridinin-Chlorophyll-Protein from Amphidinium carterae. J. Phys. Chem. A **107**: 8052–8066. doi:10.1021/jp022648z

Shwartz-Ziv, R., and A. Armon. 2022. Tabular data: Deep learning is not all you need. Information Fusion **81**: 84–90. doi:10.1016/j.inffus.2021.11.011

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

Six, C., J.-C. Thomas, L. Garczarek, M. Ostrowski, A. Dufresne, N. Blot, D. J. Scanlan, and F. Partensky. 2007. Diversity and evolution of phycobilisomes in marine Synechococcusspp.: a comparative genomics study. Genome Biology **8**: R259. doi:10.1186/gb-2007-8-12-r259

Slade, W. H., and E. Boss. 2015. Spectral attenuation and backscattering as indicators of average particle size. Appl. Opt., AO **54**: 7264–7277. doi:10.1364/AO.54.007264

Terrats, L. and others. 2023. BioGeoChemical-Argo Floats Reveal Stark Latitudinal Gradient in the Southern Ocean Deep Carbon Flux Driven by Phytoplankton Community Composition. Global Biogeochemical Cycles **37**: e2022GB007624. doi:10.1029/2022GB007624

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

Thibodeau, P. S., C. S. Roesler, S. L. Drapeau, S. G. Prabhu Matondkar, J. I. Goes, and P. J. Werdell. 2014. 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

Twardowski, M. S., E. Boss, J. B. Macdonald, W. S. Pegau, A. H. Barnard, and J. R. V. Zaneveld. 2001. A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. Journal of Geophysical Research: Oceans **106**: 14129–14142. doi:10.1029/2000JC000404

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

Uitz, J. U., Y. Huot, F. Bruyant, M. Babin, and H. Claustre. 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

Veldhuis, M. J. W., K. R. Timmermans, P. Croot, and B. van der Wagt. 2005. Picophytoplankton; a comparative study of their biochemical composition and photosynthetic properties. Journal of Sea Research **53**: 7–24. doi:10.1016/j.seares.2004.01.006

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

Yentsch, C. S., and D. W. Menzel. 1963. A method for the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. Deep Sea Research and Oceanographic Abstracts **10**: 221–231. doi:10.1016/0011-7471(63)90358-9

Yentsch, C. S., and D. A. Phinney. 1985. Spectral fluorescence: an ataxonomic tool for studying the structure of phytoplankton populations*. Journal of Plankton Research **7**: 617–632. doi:10.1093/plankt/7.5.617

Zhang, X., L. Hu, and M.-X. He. 2009. Scattering by pure seawater: Effect of salinity. Opt. Express **17**: 5698. doi:10.1364/OE.17.005698

Zhang, Y., X. He, Y. Bai, T. Li, X. Jin, D. Wang, and F. Gong. 2025. Satellite estimation of the spectral power exponent of particulate backscattering coefficient in the global ocean. Opt. Express, OE **33**: 5411–5428. doi:10.1364/OE.545222