# Biogeographical Classification of the Global Ocean From BGC-Argo Floats

Nicholas Bock, Marin Cornec, Hervé Claustre, Solange Duhamel

Global Biogeochemical Cycles®

RESEARCH ARTICLE
10.1029/2021GB007233

# Biogeographical Classification of the Global Ocean From BGC-Argo Floats

**Nicholas Bock**[1], **Marin Cornec**[2,3], **Hervé Claustre**[2], **and Solange Duhamel**[1,4]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA, [2]CNRS & Sorbonne Université, Laboratoire d'Océanographie de Villefranche, LOV, Villefranche-sur-Mer, France, [3]Now at School of Oceanography, University of Washington, Seattle, WA, USA, [4]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA

**Key Points:**

- Data from Biogeochemical-Argo provides a novel basis for defining six biogeographical provinces of the global ocean
- Light and nutrient fields explain most of the variability in phytoplankton distributions for each province
- Small particle concentrations in the euphotic zone are inversely related to their transfer efficiency toward the mesopelagic zone

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

N. Bock,
nab2161@columbia.edu

**Citation:**

Bock, N., Cornec, M., Claustre, H., & Duhamel, S. (2022). Biogeographical classification of the global ocean from BGC-Argo floats. *Global Biogeochemical Cycles*, *36*, e2021GB007233. https://doi.org/10.1029/2021GB007233

Received 20 OCT 2021
Accepted 16 MAY 2022

**Author Contributions:**

**Conceptualization:** Nicholas Bock, Marin Cornec, Hervé Claustre
**Data curation:** Nicholas Bock
**Formal analysis:** Nicholas Bock
**Investigation:** Nicholas Bock, Marin Cornec
**Methodology:** Nicholas Bock, Marin Cornec
**Project Administration:** Nicholas Bock, Hervé Claustre
**Software:** Nicholas Bock, Marin Cornec
**Supervision:** Hervé Claustre, Solange Duhamel
**Visualization:** Nicholas Bock
**Writing – original draft:** Nicholas Bock



**Abstract** Biogeographical classifications of the global ocean generalize spatiotemporal trends in species or biomass distributions across discrete ocean biomes or provinces. These classifications are generally based on a combination of remote-sensed proxies of phytoplankton biomass and global climatologies of biogeochemical or physical parameters. However, these approaches are limited in their capacity to account for subsurface variability in these parameters. The deployment of autonomous profiling floats in the Biogeochemical Argo network over the last decade has greatly increased global coverage of subsurface measurements of bio-optical proxies for phytoplankton biomass and physiology. In this study, we used empirical orthogonal function analysis to identify the main components of variability in a global data set of 422 annual time series of Chlorophyll *a* fluorescence and optical backscatter profiles. Applying cluster analysis to these results, we identified six biomes within the global ocean: two high-latitude biomes capturing summer bloom dynamics in the North Atlantic and Southern Ocean and four mid- and low-latitude biomes characterized by variability in the depth and frequency of deep chlorophyll maximum formation. We report the distribution of these biomes along with associated trends in biogeochemical and physicochemical environmental parameters. Our results demonstrate light and nutrients to explain most variability in phytoplankton distributions for all biomes, while highlighting a global inverse relationship between particle stocks in the euphotic zone and transfer efficiency into the mesopelagic zone. In addition to partitioning seasonal variability in vertical phytoplankton distributions at the global scale, our results provide a potentially novel biogeographical classification of the global ocean.

**Plain Language Summary** Due to the enormous spatiotemporal variability in environmental dynamics within the global ocean, understanding and predicting biological processes at broad scales requires identifying ecologically similar marine provinces. However, because such provinces are typically defined using data from satellites or from climatologies averaging many decades of ship-based measurements, they are limited in their ability to account for environmental variability beneath the sea surface. Over the last 10 years, the development of the global Biogeochemical Argo (BGC-Argo) networks of autonomous profiling floats has significantly relieved this constraint, providing measurements of physical, bio-optical, and biogeochemical parameters within the upper kilometer of the ocean. In this study, we explore the potential for using BGC-Argo time series data to develop a novel biogeographical classification based on the direct measurement of subsurface bio-optical parameters. We identify six phytoplankton biomes in the global BGC-Argo data set: two high-latitude biomes characterized by large phytoplankton blooms during summer months and four mid-to-low-latitude biomes characterized by complex subsurface dynamics reflecting basin-scale trends in environmental forcings. Our results capture the environmental drivers of phytoplankton distributions at the global scale, in addition to relationships between phytoplankton distributions in the upper sun-lit layer of the ocean and particle accumulation within the mesopelagic zone.

## 1. Introduction

Biogeographical partitions of marine environments aim to identify ecologically similar ocean regions at the basin and/or global scale. Such partitions allow for the extrapolation of sparse ship-based measurements, in turn making it possible to generalize species distributions (Brinton, 1962) and ecosystem processes (Longhurst, 1995); tune global bio-optical algorithms to specific regions or water types (Shi et al., 2009; T. S. Moore et al., 2009); and constrain expected variability in global models of biological (Behrenfeld et al., 2005; Platt & Sathyendranath, 1988) and biogeochemical (Sarmiento et al., 2004) processes. While these approaches primarily focus on the properties and characteristics of the upper ocean, classifications have been extended to the mesopelagic zone as well, based

**Writing – review & editing:** Nicholas Bock, Marin Cornec, Hervé Claustre, Solange Duhamel

on species distributions (Brinton, 1962; Steinberg et al., 2008), inferred subsurface biomass distributions (Proud et al., 2017), or global climatologies (Reygondeau et al., 2018; Sutton et al., 2017).

Because the majority of ocean primary production and biogeochemical cycling is carried out by microbial communities (Field et al., 1998), there is a particular interest in delineating regions characterized by similar trends in plankton dynamics. As such, biogeochemical classifications are often based on criteria presumed to reflect or at least correlate with the distribution or physiology of microbial communities. Platt and Sathyendranath (1988) identified nine ocean regions based on bathymetric elevation and latitude. Longhurst subsequently identified four general ocean biomes based on seasonal trends in vertical mixing, using these biomes as the basis for identifying 51 provinces based on metrics of stratification, estimated photic zone depths, and surface nutrient concentrations (Longhurst, 2007; Longhurst et al., 1995). Other strategies have leveraged global measurements of remote sensed parameters to establish classifications, including sea surface temperature, photosynthetically available radiation (PAR), and estimated concentration of Chlorophyll *a* ([Chl*a*]) at both regional (D'Ortenzio & Ribera d'Alcalà, 2008; Hardman-Mountford & McGlade, 2002) and global scales (Fay & McKinley, 2014; Hardman Mountford et al., 2008; Mustapha et al., 2014; T. S. Moore et al., 2001; Wilson & Coles, 2005).

While these and other approaches (see Krug et al., 2017 for a review) have been successful at defining marine provinces based on the criteria selected, they are limited in their ability to account for subsurface variability, due to poor global coverage of subsurface measurements. In the last 10 years, however, the widespread deployment of Biogeochemical Argo (BGC-Argo) floats has considerably relieved this constraint, providing high-frequency measurements of physical, biogeochemical, and bio-optical parameters in the horizontal, vertical, and temporal dimensions over long time periods (Claustre et al., 2020; Roemmich et al., 2019). These data have made it possible to characterize subsurface phytoplankton distributions and their drivers at both the basin (Barbieux et al., 2019; Lacour et al., 2017; Mignot et al., 2014) and global scale (Cornec et al., 2021). Because BGC-Argo floats typically complete transects to maximal depths of 1,000 m, they additionally provide an insight into linkages between ecological processes in the euphotic zone and carbon export in the mesopelagic zone (Briggs et al., 2011; Dall'Olmo & Mork, 2014).

In this study, we use empirical orthogonal function (EOF) analysis to characterize variability in annual time series of bio-optical parameters measured by BGC-Argo floats: in vivo chlorophyll fluorescence (FChl*a*) and particle backscattering coefficients ($b_{bp}$). At a first order, FChl*a* provides a coarse measure of algal biomass, although it can also be significantly influenced by phytoplankton physiology, especially across vertical gradients (Dubinsky & Stambler, 2009). Particle backscatter coefficients provide an overall proxy for particulate organic carbon (POC), especially within the 0.2–20 μm size range (Stramski et al., 2008). The incorporation of both FChl*a* and $b_{bp}$ in EOF analysis makes it possible to compare individual time series based on variability in phytoplankton biomass/physiology as well as particle concentrations within both the euphotic and mesopelagic zones.

Based on this analysis, we address the following objectives: (a) to describe the dominant spatiotemporal trends in vertical phytoplankton and suspended particle distributions within the global ocean; (b) to identify the environmental drivers associated with this variably; and (c) to evaluate linkages between biological variability in the upper ocean and biomass accumulation within the mesopelagic zone. Our results provide the first ever biogeographical classification of the global ocean that relies solely on bio-optical parameters measured by BGC-Argo floats and yield novel insights on the drivers of phytoplankton dynamics at the global scale.

## 2. Methods

### 2.1. Construction of Bio-Optical Time Series

This study relies on data acquisition from 205 BGC-Argo floats sampling physical and biogeochemical parameters across diverse regions of the global ocean (Argo, 2000; Claustre et al., 2020; Roemmich et al., 2019). In addition to FChl*a* and $b_{bp}$, these measurements include temperature, salinity, and up to four additional core BGC variables (PAR, dissolved oxygen, nitrate concentration, and pH). BGC-Argo float data are accessible through the Coriolis database (ftp://ftp.ifremer.fr/ifremer/argo). Monthly satellite [Chl*a*] and $b_{bp}$ measurements (hereafter $[Chla]_{sat}$ and $b_{bp\_sat}$, respectively) corresponding to each profile were obtained at a 4 km resolution from the Ocean Color-Climate-Change Initiative v4.0 database, which merges observations from MERIS, MODIS, VIIRS, and SeaWiFs (ftp://oc-cci-data:ELaiWai8ae@oceancolou.org/occci-v4.2/). Because $b_{bp}$ at 700 nm was not available

from the OCCCI database, 665 nm retrievals were downloaded and converted to 700 nm following the power law equation described in Sauzède et al. (2016).

Hydrological data were collected by the SBE 41 Seabird CTD sensors, which were processed and quality-controlled following Argo protocols as described by Wong et al. (2020). Floats were equipped with Seabird-Wetlabs sensors of three types (FLBB, ECO-Triplet, or MCOMS) that include a chlorophyll fluorometer (excitation at 470 nm; emission at 695 nm), and a backscattering sensor at 700 nm. Values for [Chl*a*] and $b_{bp}$ were retrieved from these measurements and quality-controlled (only data with good and probably good QC flags were used) following BGC-Argo protocols (Schmechtig et al., 2015, 2018a, 2018b). Non-photochemical quenching in near-surface measurements of [Chl*a*] was corrected for following Xing et al. (2012). Individual [Chl*a*] and $b_{bp}$ profiles were smoothed prior to EOF analysis to denoise data and remove spikes. Following Cornec et al. (2021), a 5-point rolling median filter was applied to [Chl*a*] profiles. For $b_{bp}$ profiles, a 7-point rolling median filter was applied, followed by a 5-point rolling mean filter. The 7-point median filter was used for $b_{bp}$ profiles to account for large spikes in individual profiles caused by aggregations of particles (Briggs et al., 2011).

[Chl*a*] and $b_{bp}$ profiles from each float were then grouped into individual time series based on the year of measurement. To account for the seasonal polarity of North and South Hemispheres, dates for floats in the Southern Hemisphere were shifted forward 6 months prior to grouping. Time series were only included in analyses if the first profile was completed before January 20, the last profile after December 10, and if there were no more than 20 days between two consecutive profiles between those dates.

To eliminate missing data in the final EOF matrix, profiles in each time series were interpolated vertically to 1-m intervals between 1 and 750 m. Each depth within the time series was then interpolated horizontally to 1-day intervals between 1 and 365 days. Interpolated [Chl*a*] and $b_{bp}$ time series were converted into vectors, creating $1 \times 273{,}750$ vectors for each parameter. [Chl*a*] and $b_{bp}$ vectors were then separately centered to mean 0, normalized to the standard deviation for all measurements in the vector, and joined end-on-end to create a $1 \times 547{,}500$ vector for each time series. This process was repeated for each complete time series in the data set, yielding a matrix consisting of 422 columns (corresponding to each time series in the data set) and 547,500 rows (corresponding to each day/depth in the time series). Including only depths up to 750 m in the time series matrix was found to maximize the proportion of overall variability captured in the first six singular vectors in EOF results. Note, however, that the estimation of the environmental parameters still takes into account measurements from the surface to 1,000 m for each profile.

### 2.2. Cluster Analysis to Characterize Similarities Between Time Series

EOF was performed on the matrix of combined [Chl*a*] and $b_{bp}$ time series, yielding (a) a matrix of singular vectors capturing varying fractions of overall variability in the original time series matrix (Figure 1) and (b) a matrix of weightings reflecting the importance of each singular vector to each of the original time series (Figure 2; see Text S1 in Supporting Information S1 for details). To identify groups of similar time series, the pairwise Euclidean distance was calculated between subsets of the weightings corresponding to the first six singular vectors (Pielou, 1984). Similar groups of weightings were then identified using divisive hierarchical clustering analysis—a clustering approach that has been indicated to provide more accurate clustering results than agglomerative methods, especially when a relatively small number of clusters is desired (Hastie et al., 2009).

In performing this analysis, several low latitude time series, primarily in the Arabian Sea (AR), were routinely grouped with high-latitude time series (Figure S1 in Supporting Information S1). Therefore, to prevent misclassification without removing data, two rounds of clustering were performed. In the first round, time series were clustered into two groups based on the EOF weightings corresponding to the first two singular vectors. Any time series with an absolute median latitude <25° that was clustered with the high-latitude time series was then manually reassigned to the low-latitude cluster. The threshold of 25° was selected to include all time series within the AR in the low-latitude cluster. This approach effectively forced the segregation of high-latitude and low-latitude floats at the first clustering step, albeit without requiring any assumption of the appropriate boundary between the low-latitude and high-latitude floats.

Divisive hierarchical cluster analysis was then performed separately for the time series in high-latitude and low-latitude clusters. Intracluster similarity was evaluated based on the average silhouette score for each cluster—a

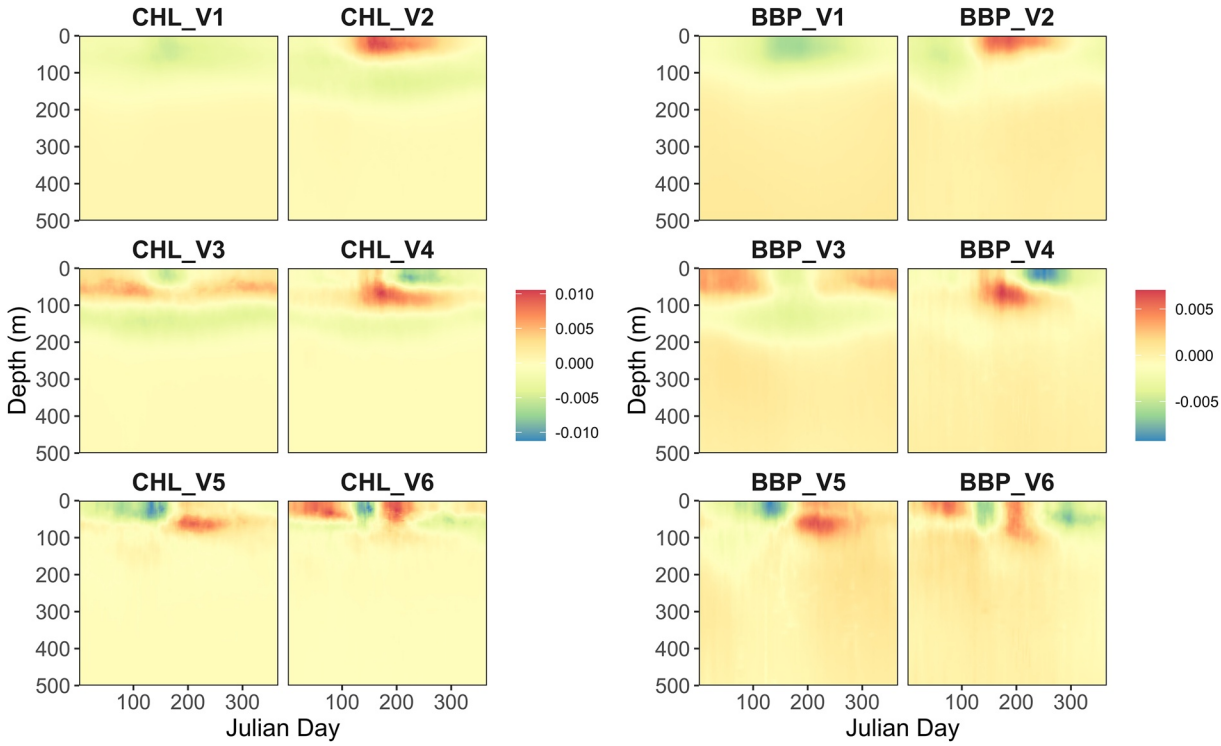


**Figure 1.** First six singular vectors obtained for rows of empirical orthogonal function (EOF) matrix corresponding to measurements of [Chl*a*] (CHL_V1—CHL_V6) and $b_{bp}$ (BBP_V1—BBP_V6). Although EOF was performed for measurements from 0 to 750 m, only depths up to 500 m are shown here. Units are arbitrary in either plot. Products of singular vectors and EOF weightings (Figure 2) yield scaled [Chl*a*] and $b_{bp}$ values in original time series matrix.

metric corresponding to the average distance between the weightings for a given time series and the weightings for all other time series within its assigned cluster (Rousseeuw, 1987; see Text S2 in Supporting Information S1 for details). The number of SVD coefficients included in cluster analysis and the number of clusters formed were selected independently for high-latitude and low-latitude floats to maximize intracluster similarity in either case. Maximal silhouette scores for low (high) latitude floats were then obtained when forming 4 (2) clusters based on the pairwise Euclidean distance of the first 6 (2) SVD coefficients (Table S1 in Supporting Information S1).

To provide a basis for comparing classifications based on SVD coefficients of vertical BGC-Argo data to those derived from surface measurements alone, time series were additionally classified by the standard deviation of annual $[\text{Chl}a]_{sat}$ measurements, following the variance bins used in the bioregionalization of Behrenfeld et al. (2005). Briefly, these bins were L0 (sd $[\text{Chl}a]_{sat} < 0.018$ mg m$^{-3}$), L1 ($0.018 <$ sd $[\text{Chl}a]_{sat} < 0.026$ mg m$^{-3}$), L2 ($0.026 <$ sd $[\text{Chl}a]_{sat} < 0.09$ mg m$^{-3}$), L3 ($0.09 <$ sd $[\text{Chl}a]_{sat} < 0.4$ mg m$^{-3}$), and L4 (sd $[\text{Chl}a]_{sat}$ 0.4 mg m$^{-3}$).

### 2.3. Determination of Environmental Parameters

Environmental parameters derived for each profile are described in Table 1. Season was determined based on the profile date after applying the adjustment for Southern Hemisphere floats. Winter, spring, summer, and fall seasons, respectively, correspond to December–February, March–May, June–August, and September–November. For time series from regions subjected to seasonal monsoon forcing (see below), the year was divided into four periods: Northeast Monsoon (December 1 to April 1); Spring Intermonsoon (April 1 to July 1); Southwest Monsoon (July 1 to September 15); and Fall Intermonsoon (September 15 to November 1). Intermonsoon date ranges were centered on seasonal maxima in surface temperature, while monsoon date ranges were centered on seasonal minima in surface temperature (Figure S2 in Supporting Information S1).

Mixed layer depth (MLD) was calculated as the depth where potential density increased above 0.03 kg m$^{-3}$ in comparison to a reference density at 5 m (de Boyer Montégut et al., 2004). The CANYON-B algorithm (Bittig

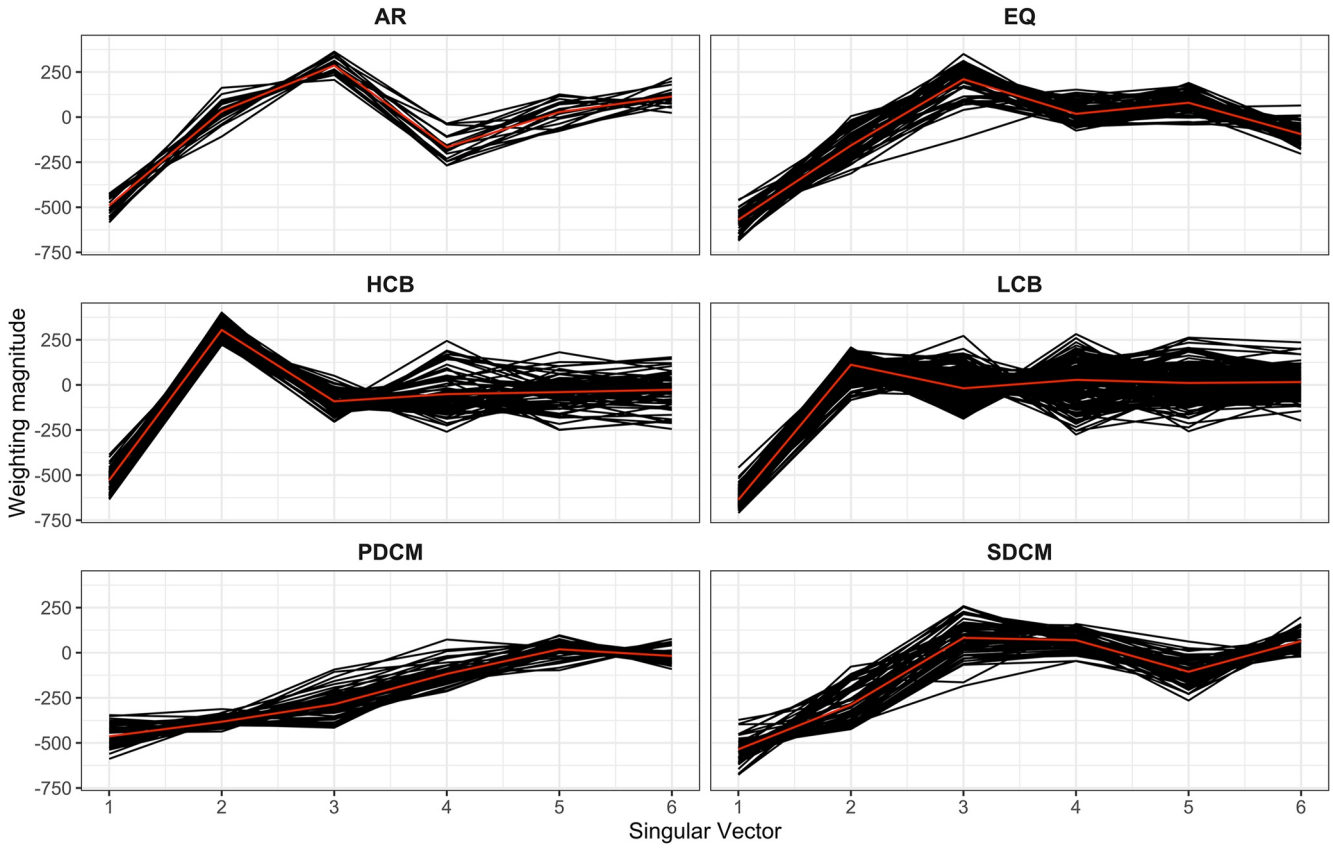


**Figure 2.** Weightings corresponding to the first six singular vectors obtained via singular value decomposition applied to the global Biogeochemical-Argo data set of [Chl*a*] and $b_{bp}$ profiles, grouped by biomes identified via cluster analysis. Red line corresponds to the mean weighting magnitude for each vector.

et al., 2018; Sauzède et al., 2017) was used to obtain estimates of nitrate, phosphate, and silicate concentrations for floats equipped with dissolved oxygen sensors. For floats deployed in the Mediterranean Sea, the CANYON-MED algorithm was used (Fourrier et al., 2020, 2021). Estimated concentrations were used to calculate the depth and slope for the nitracline, phosphocline, and silicacline. The nitracline depth was defined as the depth where nitrate concentrations increased above a threshold of 1 μM relative to surface concentrations, following Lavigne et al. (2015). Nitracline slope ($m_{NO3}$) was calculated by evaluating the rate of change in nutrients with depth between the nitracline depth and 1.25 times the nitracline depth (Cornec et al., 2021). The nitracline slope was interpreted as reflecting the upward vertical flux of nitrate at the depth of the nitracline (Gong et al., 2017). Appropriate thresholds for determining silicacline and phosphocline depths were determined for each profile as the ratio of maximal silicate or phosphate concentrations relative to the maximal nitrate concentrations. For example, if maximal silicate concentrations were five times greater than maximal nitrate concentrations, the silicacline would be defined as the depth where silicate concentrations increased above a threshold of 5 μM relative to surface concentrations.

The underwater light field for each profile was determined using float-based PAR measurements. For floats not equipped with radiometers (see Table S2 in Supporting Information S1), PAR profiles were modeled by first estimating surface PAR based on the date, latitude, and longitude of a given profile using the clear sky model of Gregg and Carder (1990). PAR profiles were then estimated as described in Cornec et al. (2021). Briefly, region-specific attenuation coefficients were then used to propagate surface PAR estimates based on [Chl*a*] profiles, following a modified implementation of Morel et al. (2007). Estimated PAR profiles were then corrected to account for cloud cover.

Measured and estimated PAR profiles were then used to estimate daily integrated PAR (iPAR) at each of the nutricline depths, following Lacour et al. (2017). Median iPAR within the mixed layer ($iPAR_{MLD}$) was derived for

**Table 1**
*Abbreviations Used in This Study*

| Symbol | Definition | Units |
|---|---|---|
| AR | Arabian Sea biome | |
| EQ | Equatorial biome | |
| HCB | High chlorophyll bloom biome | |
| LCB | Low chlorophyll bloom biome | |
| PDCM | Permanent deep chlorophyll maximum biome | |
| SDCM | Seasonal deep chlorophyll maximum biome | |
| STFZ | Subtropical Frontal Zone time series | |
| MED | Mediterranean Sea time series | |
| [Chl*a*] | Chl*a* concentration | mg $m^{-3}$ |
| $[Chla]_{mz}$ | Average Chl*a* concentration within the mesopelagic zone | mg $m^{-3}$ |
| $[Chla]_{pz}$ | Average Chl*a* concentration within the productive zone | mg $m^{-3}$ |
| $[Chla]_{sat}$ | Satellite [Chl*a*] | mg $m^{-3}$ |
| $b_{bp}$ | Particulate backscattering coefficient | $m^{-1}$ |
| $b_{bp\_pz}$ | Average $b_{bp}$ within the productive zone | $m^{-1}$ |
| $b_{bp\_mz}$ | Average $b_{bp}$ within the mesopelagic zone | $m^{-1}$ |
| iPAR | Daily integrated PAR | E $m^{-2}$ $d^{-1}$ |
| $iPAR_{MLD}$ | Median iPAR in the mixed layer | E $m^{-2}$ $d^{-1}$ |
| $iPAR_{NO3}$ | iPAR at the nitracline depth | E $m^{-2}$ $d^{-1}$ |
| MLD | Mixed layer depth | m |
| $mNO_3$ | Nitracline slope | µM $m^{-1}$ |
| $N_{MLD}$ | Average $NO_3$ concentration within the mixed layer | µM |
| PAR | Photosynthetically available radiation | E $m^{-2}$ |
| $P_{MLD}$ | Average $PO_4$ concentration within the mixed layer | µM |
| $S_{MLD}$ | Average $SiO_4$ concentration within the mixed layer | µM |
| $T_{effp}$ | Transfer efficiency of particle stocks from euphotic zone to mesopelagic zone | % |
| $Z_{ez}$ | Depth of euphotic zone; depth where PAR is reduced to 0.1% of its surface value | m |
| $Z_{NO3}$ | Depth of the nitracline | m |
| $Z_{pz}$ | Depth of productive zone; maximum of MLD or $z_{eu}$ for a given profile | m |

each profile by calculating iPAR at the median depth between the surface and MLD. The depth of the euphotic layer ($Z_{eu}$) was defined as the depth where PAR was reduced to 0.1% of its value at 1 m. Following Dall'Olmo and Mork (2014), the depth of the productive layer ($Z_p$) was defined as the greater value of the MLD or $Z_{eu}$ for a given profile. The depth of the mesopelagic layer ($Z_m$) was defined as the maximal depth of the float transit (typically around 1,000 m).

Average [Chl*a*] and $b_{bp}$ within the productive layer ($[Chla]_{pz}$ and $b_{bp\_pz}$, respectively) were calculated by integrating profiles from 1 m to the bottom of the productive layer and then normalizing to $Z_p$. Average $b_{bp}$ values within the mesopelagic zone ($b_{bp\_mz}$) were calculated by integrating from $Z_p$ to $Z_m$ and normalizing to the difference in depth between $Z_m$ and $Z_p$. The ratio of $b_{bp\_pz}$:$b_{bp\_mz}$ was interpreted as a proxy for transfer efficiency ($T_{effp}$), following Galí et al. (2021), albeit in our case based on comparative depth-integrated $b_{bp}$ values rather than POC estimates. Average $b_{bp}$ values were also calculated for each of the four equally sized depth ranges within the mesopelagic zone ($b_{bp\_m1} - b_{bp\_m4}$) by integrating $b_{bp}$ profiles within each depth range and then normalizing to the width of the depth range (Figure S3 in Supporting Information S1). The width of each depth range was defined as $(Z_m - Z_p)/4$. Because $Z_p$ was calculated as the greater value of the MLD or $Z_{eu}$ for each profile in the data set,

mesopelagic depth ranges were calculated for each profile. For ease of reporting, all $b_{bp}$ values are reported in units of $10^{-3}$ $m^{-1}$.

### 2.4. Statistics

All analyses were performed in R 4.0.2 using RStudio 1.3.1073 (R Core Team, 2016). Shapiro-Wilk tests were performed for each parameter to evaluate normality of data. Bartlett's tests were used to evaluate homogeneity of variance. Because distributions of parameter measurements were generally not normally distributed, the median is used as the measure of central tendency for all results. The interquartile range (IQR) is used as the measure of dispersion, calculated as the difference between the third and first quartiles. For non-normally distributed parameters, Kruskal-Wallis tests were used to evaluate significant differences in parameter measurements based on cluster, season, or month of sampling, and Dunn's post-hoc test was used to perform multiple comparisons. For normally distributed parameters, analysis of variance (ANOVA) tests was used, followed by Tukey's Honestly Significant Difference post-hoc test. To evaluate main sources of variability within the data set of environmental parameters, principal component analysis (PCA) was performed on log-transformed and standardized parameter values. To evaluate relationships between bio-optical and environmental parameters, linear regressions were performed on monthly parameter averages for each biome. Relationships between bio-optical parameters were evaluated by calculating Spearman's rank correlation coefficient ($\rho$) for pairs of depth-integrated parameters (e.g., $[Chl\mathit{a}]_{pz}$, $b_{bp\_pz}$, $b_{bp\_m1}$ — $b_{bp\_m4}$).

## 3. Results

### 3.1. Results of Factor Analysis

The first six singular vectors identified in EOF analysis captured 81% of the variability in the data set of 422 time series (Figure 1, Table S3 in Supporting Information S1). The contribution of each singular vector to the original time series depends on both the magnitude and sign of its corresponding coefficient (Figure 2). For example, the first singular vector alone, when multiplied by a negative coefficient, results in a time series characterized by large surface [Chl*a*] and $b_{bp}$ values during summer months. The second singular vector, when multiplied by a negative coefficient, counteracts the effect of the first singular vector and yields a time series characterized by a seasonal deep chlorophyll maximum (DCM). Conversely, if multiplied by a positive coefficient, the second singular vector would enhance the contribution of the first singular vector. In this way, each coefficient adds additional information about the vertical and seasonal distribution of [Chl*a*] and $b_{bp}$ in the original time series and can be used to classify time series via cluster analysis.

### 3.2. Results of Cluster Analysis

The initial clustering (Figure 3a) formed a low-latitude cluster with a unimodal latitudinal distribution centered on 7.4° ± 24.8°N ($N$ = 190) and a high-latitude cluster with a bimodal latitudinal distribution centered on 58.0° ± 8.2°N and −50.24° ± 6.85°N ($n$ = 232). High-latitude time series were further divided (Figure 3b, Figure S4 in Supporting Information S1) into a high-chlorophyll bloom (HCB) cluster centered at 60.4° ± 7.2°N ($n$ = 164) and a low-chlorophyll bloom (LCB) cluster centered at −49.9° ± −6.7°N ($n$ = 68). Low latitude time series were divided into AR, equatorial (EQ), permanent DCM (PDCM), and seasonal DCM (SDCM) clusters. AR time series were centered at 16.01° ± 2.57°N ($n$ = 18); EQ time series were centered at 4.18° ± 13.08°N ($n$ = 60); PDCM time series had a bimodal distribution centered at 23.68° ± 3.35°N and −22.06° ± 4.05°N ($n$ = 49); SDCM time series had a bimodal distribution centered at 36.01° ± 3.20°N and −34.98° ± 8.10°N ($n$ = 63). SDCM time series in northern latitudes were nearly entirely limited to the Mediterranean Sea, while those in southern latitudes were concentrated at latitudes separating the southern subtropical gyres from the Southern Ocean. To capture any differences in environmental drivers between these regions, Mediterranean (hereafter MED) and open ocean SDCM time series (hereafter Subtropical Frontal Zone, or STFZ) were analyzed separately in addition to those corresponding to the SDCM time series as a whole. Following Longhurst (2007), in the remainder of the text the HCB, LCB, SDCM, PDCM, AR, and EQ clusters are referred to as biomes, since each

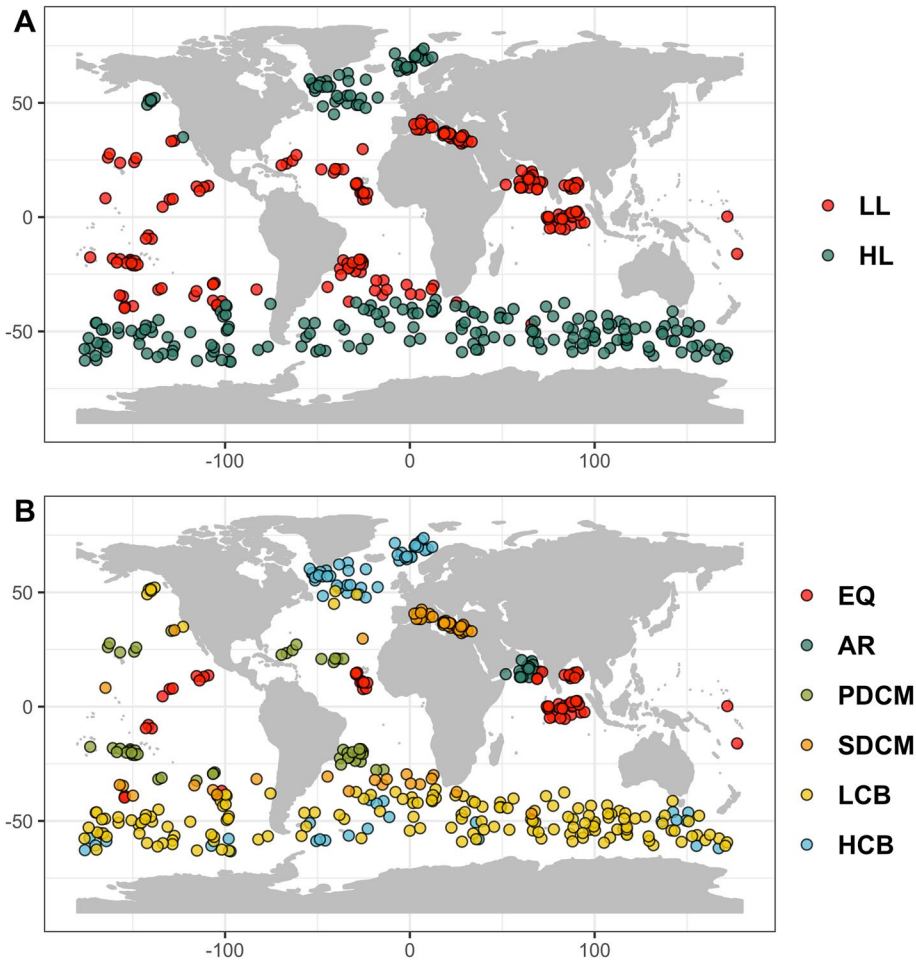


**Figure 3.** Geographic distributions of (a) High-latitude (HL) and low-latitude (LL) time series based on initial cluster analysis, after manual reassignment of Arabian Sea time series. (b) Equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low-chlorophyll bloom (LCB), and high-chlorophyll bloom (HCB) time series based on second round of cluster analysis, as described in text. Coordinates correspond to the median latitude and longitude for each time series.

has a characteristic seasonal phytoplankton cycle (see below). High-latitude clusters and low-latitude clusters are still identified as clusters since they are themselves composed of several biomes.

### 3.3. Trends in Bio-Optical Parameters for Each Biome

Averaging the vectors in the original time series matrix belonging to each biome revealed distinct seasonal trends in [Chl*a*] (Figure 4, Table 2) and $b_{bp}$ (Figure 5, Table 3). Time series in the EQ biome were characterized by a DCM at a median depth of 60.5 m (IQR = 20.5 m) that persisted throughout the year. Time series in the AR biome were characterized by the formation of a DCM during intermonsoon periods at median depths of 55.5 m (IQR = 18.4 m) during the Spring Intermonsoon and 34.5 m (IQR = 15.0 m) and during the Fall Intermonsoon. Time series in the SDCM biome were characterized by a spring phytoplankton bloom, followed by the formation of a DCM that reached a maximal depth of 91.0 m (IQR = 31.5 m) in summer months. Time series in the PDCM biome were characterized by a PDCM that reached a maximal depth of 136.0 m (IQR = 28.0 m) during summer months. Time series in the LCB and HCB biomes were both characterized by the formation of summer phytoplankton blooms, without the formation of SDCMs.

Values for $[\text{Chl}a]_{pz}$ showed marked seasonal and inter-biome variability (Table 2). HCB and LCB time series were characterized by globally maximal $[\text{Chl}a]_{pz}$ values during summer months. $[\text{Chl}a]_{pz}$ values were similarly

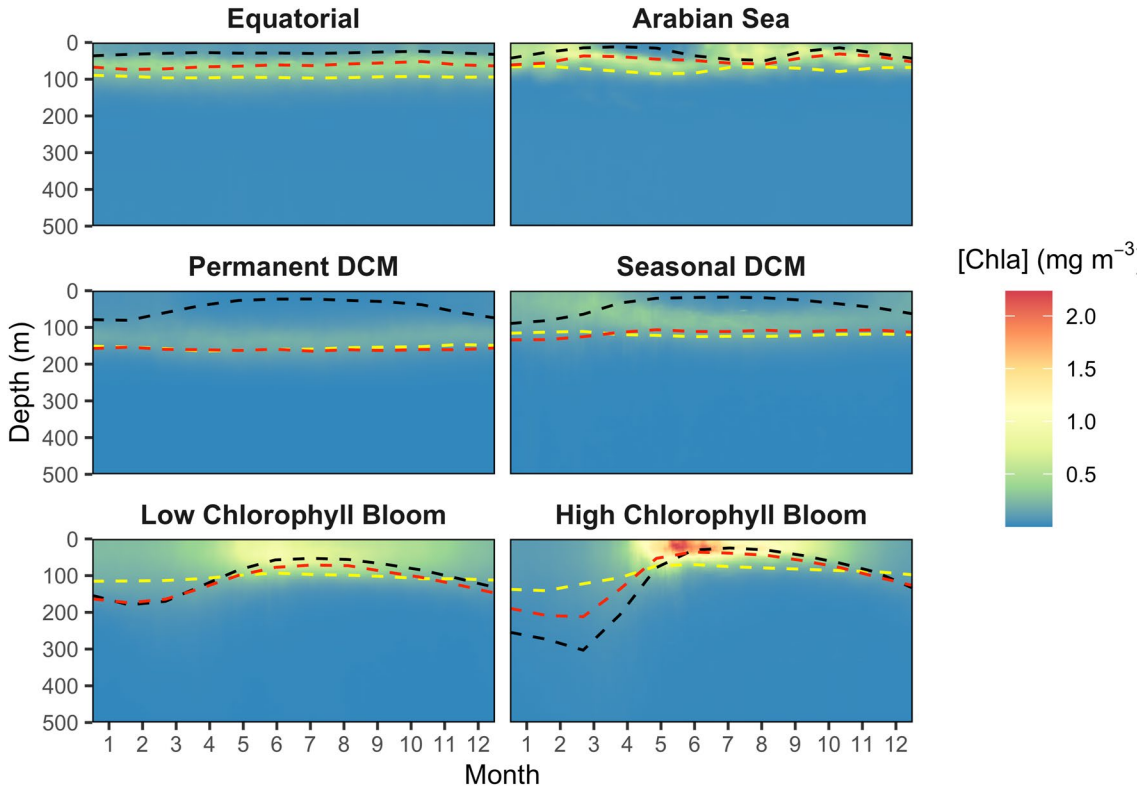


**Figure 4.** Mean [Chl*a*] (mg $m^{-3}$) for time series in the equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low chlorophyll bloom (LCB), and high chlorophyll bloom (HCB) biomes. Black dashed line corresponds to mean mixed layer depth for each biome. Red dashed line corresponds to median nitracline depth for each biome. Yellow line corresponds to the median euphotic zone depth (defined as the 0.1% PAR isolume) for each biome.

elevated in AR, although with bimodal seasonality. Time series in the EQ and SDCM biomes were both characterized by globally intermediate $[Chla]_{pz}$ values with limited seasonality compared to HCB, LCB, and AR. PDCM time series represented globally minimal values for both the magnitude and degree of annual variability in $[Chla]_{pz}$.

Monthly averaged $[Chla]_{pz}$ values (Table 2) were significantly correlated with $[Chla]_{sat}$ values (Table S4 in Supporting Information S1) in the AR ($\rho = 0.87$ and $p = 0.0003$), PDCM ($\rho = 0.92$ and $p < 1 \times 10^{-6}$), and HCB

**Table 2**
*Median $[Chla]_{pz}$ Values for Each Biome*

| | Winter | Spring | Summer | Fall | Annual |
|---|---|---|---|---|---|
| **EQ** | 0.20 (0.11) | 0.21 (0.10) | 0.20 (0.8) | 0.23 (0.8) | 0.20 (0.09) |
| **AR*** | 0.46 (0.27) | 0.24 (0.14) | 0.50 (0.53) | 0.46 (0.33) | 0.38 (0.32) |
| **SDCM** | 0.15 (0.11) | 0.17 (0.17) | 0.14 (0.10) | 0.11 (0.05) | 0.14 (0.10) |
| MED | 0.14 (0.07) | 0.15 (0.10) | 0.14 (0.10) | 0.11 (0.04) | 0.13 (0.07) |
| STFZ | 0.25 (0.19) | 0.31 (0.27) | 0.21 (0.16) | 0.12 (0.14) | 0.23 (0.20) |
| **PDCM** | 0.08 (0.03) | 0.08 (0.03) | 0.08 (0.03) | 0.08 (0.03) | 0.08 (0.03) |
| **LCB** | 0.32 (0.18) | 0.48 (0.34) | 0.62 (0.50) | 0.42 (0.29) | 0.39 (0.37) |
| **HCB** | 0.23 (0.20) | 0.48 (0.56) | 0.86 (0.93) | 0.45 (0.36) | 0.45 (0.71) |

*Note.* All values in units of mg $m^{-3}$. Interquartile range included in parentheses. Mediterranean (MED) and Subtropical Frontal Zone (STFZ) refer to subsets of seasonal deep chlorophyll maximum biome (SDCM) time series as described in the text. Asterisk indicates that in the Arabian Sea (AR) biome, winter, spring, summer, and fall correspond to Northeast Monsoon, Summer Intermonsoon, Southwest Monsoon, and Fall Intermonsoon periods, respectively.

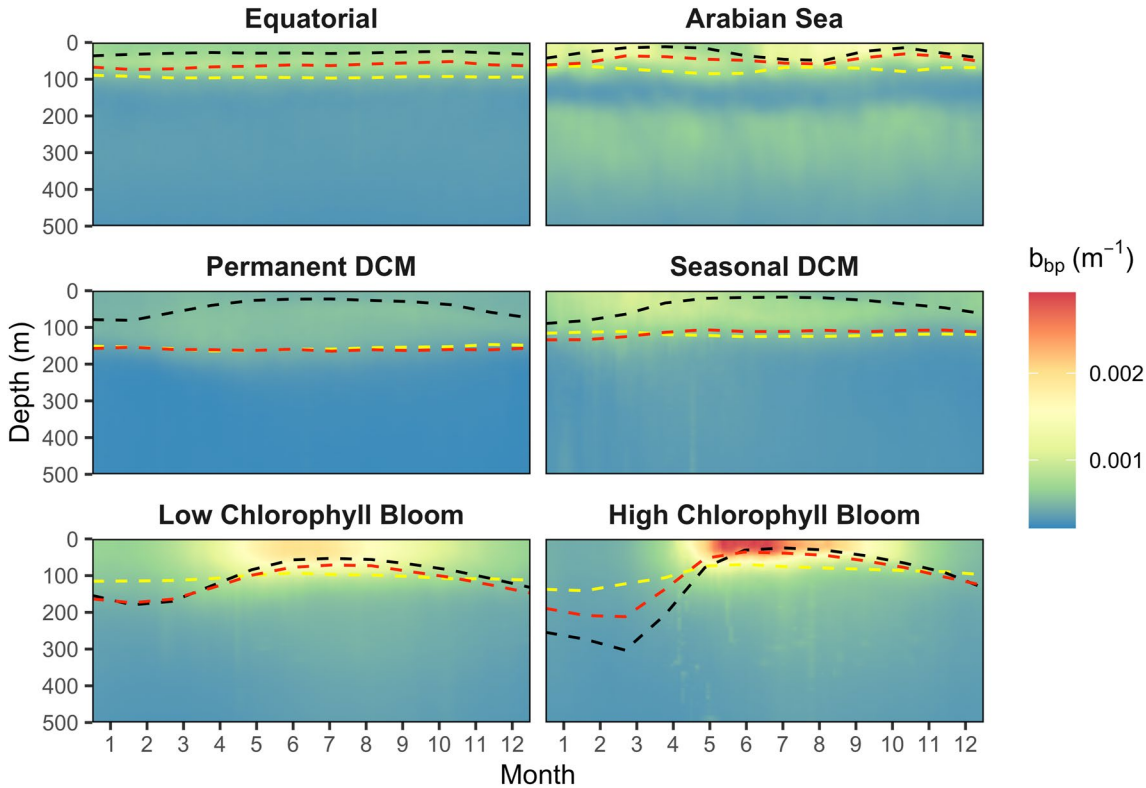


**Figure 5.** Mean $b_{bp}$ ($m^{-1}$) for time series in the equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low chlorophyll bloom (LCB), and high chlorophyll bloom (HCB) biome. Black dashed line corresponds to mean mixed layer depth for each biome. Red dashed line corresponds to median nitracline depth for each biome. Yellow line corresponds to the median euphotic zone depth (defined as the 0.1% PAR isolume) for each biome.

biomes ($\rho = 0.96$, $p = < 1 \times 10^{-6}$). However, correlations between satellite and float measurements of [Chl*a*] in the SDCM and EQ biomes were only statistically significant when comparing $[\text{Chl}a]_{sat}$ to float-measured [Chl*a*] integrated across the first optical depth alone (SDCM: $\rho = 0.96$ and $p < 1 \times 10^{-6}$, EQ: $\rho = 0.68$ and $p = 0.02$). No significant correlations were identified between satellite and float based measurements of [Chl*a*] in the LCB biome. $[\text{Chl}a]_{sat}$ variance bin classifications agreed with SVD coefficient classifications at very high and very low annual variability, with variance bins L0 and L4 primarily containing time series from the PDCM and HCB

**Table 3**
*Median $b_{bp\_pz}$ Values for Each Biome*

| | Winter | Spring | Summer | Fall | Annual |
|---|---|---|---|---|---|
| **EQ** | 0.65 (0.14) | 0.68 (0.13) | 0.68 (0.15) | 0.68 (0.16) | 0.67 (0.15) |
| **AR*** | 0.94 (0.37) | 0.78 (0.19) | 1.14 (0.63) | 0.94 (0.34) | 0.90 (0.38) |
| **SDCM** | 0.61 (0.38) | 0.75 (0.45) | 0.64 (0.31) | 0.52 (0.18) | 0.62 (0.31) |
| MED | 0.58 (0.26) | 0.69 (0.27) | 0.59 (0.18) | 0.50 (0.15) | 0.60 (0.24) |
| STFZ | 0.73 (0.40) | 1.06 (0.64) | 0.91 (0.49) | 0.65 (0.30) | 0.82 (0.50) |
| **PDCM** | 0.45 (0.09) | 0.48 (0.09) | 0.50 (0.08) | 0.47 (0.08) | 0.48 (0.08) |
| **LCB** | 0.85 (0.49) | 1.40 (0.78) | 1.57 (0.90) | 1.06 (0.58) | 1.12 (0.80) |
| **HCB** | 0.87 (0.59) | 1.38 (1.33) | 1.86 (1.25) | 1.10 (0.62) | 1.15 (1.30) |

*Note.* All values in units of $m^{-1}$. Interquartile range included in parentheses. Mediterranean (MED) and Subtropical Frontal Zone (STFZ) refer to subsets of seasonal deep chlorophyll maximum biome (SDCM) time series as described in the text. Asterisk indicates that in the Arabian Sea (AR) biome, winter, spring, summer, and fall correspond to Northeast Monsoon, Summer Intermonsoon, Southwest Monsoon, and Fall Intermonsoon periods, respectively.

biomes. However, intermediate variance bins grouped large numbers of time series from multiple biomes, as summarized in Table S5 in Supporting Information S1.

Trends in $b_{bp\_pz}$ generally mirrored those in [Chl*a*] in HCB, LCB, and AR biomes. Spearman's correlation coefficients for comparisons of $b_{bp\_pz}$ and $[Chla]_{pz}$ were 0.90 for the HCB biome, 0.80 for the LCB biome, and 0.72 for the AR biome. Relationships between $b_{bp\_pz}$ and $[Chla]_{pz}$ were weaker in the SDCM, EQ, and PDCM biomes, with correlation coefficients of 0.62, 0.61, and 0.17, respectively. Within the SDCM biome, correlation coefficients were 0.73 for STFZ time series and 0.51 for MED time series. All correlation coefficients were statistically significant, with *p*-values of less than $1 \times 10^{-24}$ for all comparisons. Monthly averaged values of $b_{bp\_pz}$ (Table 3) and $b_{bp\_sat}$ (Table S6 in Supporting Information S1) were significantly correlated in the SDCM ($\rho = 0.73$, $p = 0.009$), PDCM ($\rho = 0.92$ and $p \leq 1 \times 10^{-6}$), LCB ($\rho = 0.97$ and $p \leq 1 \times 10^{-6}$), and HCB ($\rho = 0.99$ and $p \leq 1 \times 10^{-6}$) biomes. Satellite $b_{bp}$ retrievals were not correlated with float-measured $b_{bp}$ in the AR or EQ biomes, even when only integrating float $b_{bp}$ across the first optical depth as described above.

Values for $b_{bp\_pz}$ were generally correlated with depth-integrated $b_{bp}$ values across different layers of the mesopelagic zone (Table S7 in Supporting Information S1). These relationships attenuated with depth, except in the case of AR, where correlations strengthened with depth. Values for $b_{bp\_m1}$ showed pronounced seasonality in the LCB and HCB biomes increasing to 0.50 $m^{-1}$ (IQR = 0.20 $m^{-1}$) and 0.55 $m^{-1}$ (IQR = 0.28 $m^{-1}$), respectively, during the summer bloom. In the SDCM biome $b_{bp\_m1}$ values reached maxima of 0.37 $m^{-1}$ (IQR = 0.13 $m^{-1}$) during the spring bloom. The EQ, AR, and PDCM biomes were characterized by comparatively limited seasonality in $b_{bp\_m1}$, with annual median values of 0.33 $m^{-1}$ (IQR = 0.10 $m^{-1}$), 0.51 $m^{-1}$ (IQR = 0.18 $m^{-1}$), and 0.26 $m^{-1}$ (IQR = 0.06 $m^{-1}$), respectively. Differences in $b_{bp\_m4}$ values between biomes were much less pronounced, with values in the EQ, SDCM, LCB, and HCB biomes being extremely similar (overall median = 0.25 $m^{-1}$, IQR = 0.09 $m^{-1}$). Values for $b_{bp\_m4}$ were comparatively elevated in the AR biome (median = 0.33 $m^{-1}$, IQR = 0.05 $m^{-1}$) and reduced in the PDCM biome (median = 0.20 $m^{-1}$, IQR = 0.06 $m^{-1}$). Values for $b_{bp\_pz}$ were significantly inversely correlated to $T_{effp}$ both for individual clusters (data not shown) and within the global data set ($\rho = -0.82$, $p < 1 \times 10^{-16}$; see Discussion in Section 4.3).

### 3.4. Global Variability in Environmental Parameters

Results of PCA performed on environmental parameters captured (a) the degree of correlation between individual environmental parameters and (b) main sources of variability in environmental parameters for each biome (Figure 6). The first two principal components captured 43.6% and 26.7% of variability in the data set, respectively. Based on the relative magnitude of the weightings associated with individual environmental parameters for each principal component (data not shown), the first principal component primarily captured variability related to surface temperature, $N_{MLD}$, and $iPAR_{MLD}$. The second principal component primarily captured variability related to $Z_{NO3}$, $iPAR_{NO3}$, $B_{bp\_pz}$, and $m_{NO3.}$

### 3.5. Environmental Setting of Biomes

#### 3.5.1. Equatorial Biome

Time series in the EQ biome were characterized by limited seasonal variability for all environmental parameters (Figure S5 in Supporting Information S1). Time series had a median surface temperature of 28.7°C (IQR = 2.2°C), and a median MLD of 25.0 m (IQR = 26.0 m). The nitracline was situated 29.0 m (IQR = 32 m) beneath the MLD throughout the year, and median nitrate concentrations within the mixed layer were perennially low (median = 0.2 μM, IQR = 0.9 μM). Values for nitracline slope in EQ time series represented global maxima, with an annual median value of 0.62 μM $m^{-1}$ (IQR = 0.69 μM $m^{-1}$). $iPAR_{NO3}$ were high throughout the year (median = 0.91 E $m^{-2}$ $d^{-1}$, IQR = 1.62 E $m^{-2}$ $d^{-1}$).

#### 3.5.2. Arabian Sea Biome

Time series in the AR biome were characterized by bimodal seasonality in virtually all parameters (Figure S6 in Supporting Information S1). Surface temperatures ranged from median values of 26.3°C (IQR = 1.2°C) during the Northeast Monsoon, to median values of 29.6°C (IQR = 1.7°C) during Summer Intermonsoon. MLDs reached minimal depths of 11.0 m (IQR = 6.00 m) and 12.0 m (IQR = 9.0 m) during the Spring Intermonsoon

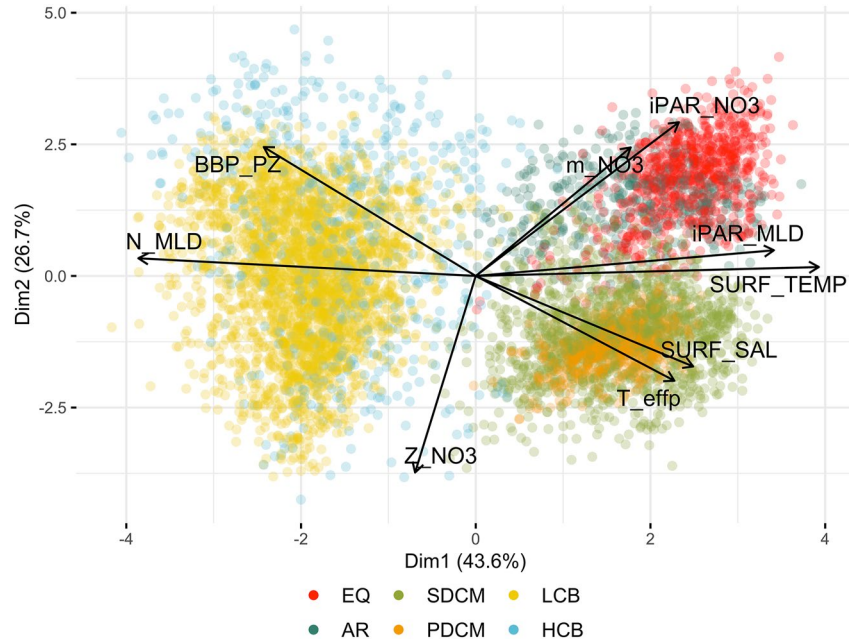


**Figure 6.** Principal component analysis results for environmental parameters corresponding to individual profiles in equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low chlorophyll bloom (LCB), and high chlorophyll bloom (HCB) biomes. Length of vectors for individual parameters corresponds to correlation between principal component analysis (PCA) loadings on each component. Angle between vectors corresponds to correlation between PCA loadings for individual parameters. Parameters are the same as those in Table 1, using an underscore to indicate subscript text.

and Fall Intermonsoon, respectively, and increased to 48.0 m (IQR = 41.0 m) and 45.0 m (IQR = 25.0 m) during the Northeast and Southwest Monsoon, respectively.

The nitracline was situated 26.0 m (IQR = 17.0 m) beneath the MLD during the Spring Intermonsoon and 14.0 m (IQR = 17.0 m) beneath the MLD during the Fall Intermonsoon. In contrast, the MLD and nitracline depths became coincident during the Northeast Monsoon and Southwest Monsoon. Nitrate concentrations in the mixed layer increased during monsoon periods, reaching annual maxima of 3.1 μM (IQR = 3.0 μM) and 2.4 μM (IQR = 3.1 μM) during the Northeast Monsoon and Southwest Monsoon, respectively. Nitrate concentrations in the mixed layer decreased to concentrations of 1.2 μM (IQR = 1.8 μM) and 1.7 μM (IQR = 3.0) during the Spring Intermonsoon and Fall Intermonsoon, respectively.

The nitracline slope was generally high throughout the year (median = 0.38 μM $m^{-1}$, IQR = 0.46 μM $m^{-1}$), although decreased sharply at the onset of Spring Intermonsoon period before increasing to annually maximal values during the Southwest Monsoon period. Values for $iPAR_{NO3}$ during intermonsoon periods represented global maxima, reaching 2.2 E $m^{-2}$ $d^{-1}$ (IQR = 3.2 E $m^{-2}$ $d^{-1}$) during the Spring Intermonsoon and 1.1 E $m^{-2}$ $d^{-1}$ (IQR = 2.0 E $m^{-2}$ $d^{-1}$) during the Fall Intermonsoon. Values for $iPAR_{NO3}$ were comparatively low during monsoon periods, decreasing to 0.1 E $m^{-2}$ $d^{-1}$ (IQR = 0.4 E $m^{-2}$ $d^{-1}$) during the Southwest Monsoon and to 0.14 E $m^{-2}$ $d^{-1}$ (IQR = 1.0 E $m^{-2}$ $d^{-1}$) during Northeast Monsoon.

### 3.5.3. Seasonal DCM Biome

Time series in the SDCM biome were characterized by strong seasonality in all environmental parameters (Figure S7 in Supporting Information S1). Surface temperatures in SDCM time series ranged from median values of 16.7°C during winter months to 23.7°C during summer months. The MLD reached a maximum of 72.0 m (IQR = 66.5 m) in January and shoaled to 13.0 m (IQR = 15.0 m) in summer months. The nitracline was located 38.0 m (IQR = 69.0 m) beneath the MLD during winter months, and 83.0 m (IQR = 64.0 m) during summer months. Nitrate concentrations within the mixed layer showed a small degree of seasonal variability, reaching maximal concentrations of 0.27 μM (IQR = 0.73 μM) during winter months, and decreasing to 0.07 μM

(IQR = 0.20 μM) during summer months. The nitracline slope was low compared to other biomes and showed limited seasonal variability (median = 0.05, IQR = 0.07). Median values of $iPAR_{NO3}$ showed limited seasonal variability, ranging from 0.008 E $m^{-2}$ $d^{-1}$ (IQR = 0.03 E $m^{-2}$ $d^{-1}$) during winter months to 0.19 E $m^{-2}$ $d^{-1}$ (IQR = 0.56 E $m^{-2}$ $d^{-1}$) during summer months.

Separately analyzing MED (Figure S8 in Supporting Information S1) and STFZ (Figure S9 in Supporting Information S1) time series revealed distinct hydrological and biogeochemical dynamics within the SDCM biome. The average MLD during wither months was significantly greater (ANOVA, $p < 2 \times 10^{-16}$) in STFZ time series (median = 106.0 m, IQR = 51.0 m) compared to MED time series (median 61.0 m, IQR = 57.0 m). In STFZ time series, average nitrate concentrations within the mixed layer ranged from 0.06 μM (IQR = 0.23 μM) to maxima of 0.70 μM (IQR = 1.12 μM) during winter months. Values for $N_{MLD}$ were comparatively reduced in MED time series, ranging from 0.07 μM (IQR = 0.15 μM) during summer months to 0.19 μM (IQR = 0.34 μM) during winter months. Differences in estimated phosphate concentrations within the mixed layer ($P_{MLD}$) were more pronounced, reaching maxima of 0.24 μM (IQR = 0.14 μM) during winter months in STFZ time series, compared to maxima of 0.02 μM (IQR = 0.01 μM) during winter months in MED time series.

#### 3.5.4. Permanent DCM Biome

Time series in the PDCM biome were characterized by intermediate seasonality in all environmental parameters and a permanent separation of the nitracline and MLD (Figure 4, Figure S10 in Supporting Information S1). Surface temperatures ranged from median values of 24.3°C (IQR = 2.3°C) in spring months to 27.3°C (IQR = 1.6°C) in fall months. MLD ranged from 21.0 m (IQR = 20.0 m) during summer months to 78.5 m (IQR = 51.0 m) in winter months. The nitracline was located 69.0 m (IQR = 56.25 m) beneath the MLD throughout the year, ranging from 69.0 m (IQR = 56.25 m) in winter months to 141.0 m (IQR = 52.0 m) in summer months. There were no PDCM profiles where the MLD was situated below the nitracline depth. Nitrate concentrations were depleted in the mixed layer throughout the year, representing global minima (median = 0.1 μM, IQR = 0.14). Median values for $iPAR_{NO3}$ ranged from 0.014 E $m^{-2}$ $d^{-1}$ (IQR = 0.04 E $m^{-2}$ $d^{-1}$) during fall months to 0.038 E $m^{-2}$ $d^{-1}$ (IQR = 0.08 E $m^{-2}$ $d^{-1}$) during spring months.

#### 3.5.5. Low Chlorophyll Bloom Biome

Time series in the LCB biome were characterized by deep winter mixing and relatively weak summer stratification (Figure S11 in Supporting Information S1). Surface temperatures showed comparatively little seasonal variability, ranging from 5.8°C (IQR = 6.6°C) during winter months to 8.1°C during summer months (IQR = 8.15°C). The MLD increased to 146 m (IQR = 87.5 m) during winter months and shoaled to 52 m (IQR = 47.0 m) during summer months. The nitracline depth was roughly coincident with the MLD throughout the year, located 6.9 m beneath the MLD on average (IQR = 7.4 m). Average nitrate concentrations showed limited annual variability, reaching maximal values of 21.6 μM (IQR = 13.3 μM) during spring months, and decreasing to 19.3 μM (IQR = 15.6 μM) during summer months. Average silicate concentrations were more variable, reaching values of 6.9 μM (IQR 17.5 μM) during winter months and decreasing to 2.6 μM (IQR = 6.9 μM) during summer months. $iPAR_{MLD}$ values increased from 0.04 E $m^{-2}$ $d^{-1}$ (IQR = 0.14 E $m^{-2}$ $d^{-1}$) during winter months to 2.8 E $m^{-2}$ $d^{-1}$ (IQR = 6.9 E $m^{-2}$ $d^{-1}$) during summer months.

#### 3.5.6. High Chlorophyll Bloom Biome

Time series in the HCB biome were characterized by deep winter mixing and a degree of summer stratification similar to that in low latitude floats (Figure S12 in Supporting Information S1). Surface temperatures represented global minima, ranging from 4.9°C (IQR = 3.3°C) during winter months to 8.6°C (IQR = 4.4°C) during summer months. Annual variability in MLD was greater than in any other biome, reaching maximal depths of 165 m (IQR = 139 m) during winter months and shoaling to 23 m (IQR = 23 m) during summer months. As in the LCB biome, the nitracline depth was generally coincident with the MLD throughout the year. Nitrate concentrations within the mixed layer were more variable than those in the LCB time series, decreasing from 15.9 μM (IQR = 15.2 μM) during winter months to 5.4 μM (IQR = 18.9 μM) during summer months. Silicate concentrations ranged from 7.9 μM (IQR = 18.4 μM) during winter months to minima of 1.8 μM during summer months (IQR = 3.2 μM). $iPAR_{MLD}$ values increased from 0.015 E $m^{-2}$ $d^{-1}$ (IQR = 0.05 E $m^{-2}$ $d^{-1}$) during winter months to 5.1 E $m^{-2}$ $d^{-1}$ (IQR = 7.5 E $m^{-2}$ $d^{-1}$) during summer months.

## 4. Discussion

### 4.1. Distribution of Biomes and Comparison With Existing Classifications

We identified six biomes within the global ocean based on time series of bio-optical measurements from BGC-Argo floats. Results of cluster analyses indicate that this grouping provides the greatest possible intracluster similarity in EOF weightings and, by extension, in the original [Chl*a*] and $b_{bp}$ time series. Clustering EOF weightings into smaller or larger numbers of groups reduced silhouette scores in either case. Dividing time series into a smaller number of clusters also resulted in less geographically or ecologically interpretable results, for example, leading to the fusion of EQ and PDCM biomes when grouping low-latitude time series into three clusters rather than four. Conversely, grouping low-latitude floats into five time series divided the SDCM biome into two clusters differentiated by the depth of the DCM during summer months, albeit with the new clusters having lower silhouette scores than the original SDCM biome. Although our results required the manual reclassification of AR time series and two EQ time series, doing so ultimately increased the silhouette scores of the final clusters, indicating that the manual reassignment leads to greater intracluster similarity than would otherwise be possible.

The distribution of the biomes in our results generally shows stronger regionalization than those derived from variability in surface measurements of bio-optical proxies alone (e.g., Behrenfeld et al., 2005; Hardman Mountford et al., 2008). As captured in Table S5 in Supporting Information S1, classifications based on $[Chl a]_{sat}$ variance bins reflect subsurface dynamics reasonably well in instances of very low $[Chl a]_{sat}$ variability (e.g., L0, which primarily included time series from the PDCM biome) or very high $[Chl a]_{sat}$ variability (e.g., L4, which primarily included time series from the HCB biome). However, accounting for subsurface variability in bio-optical measurements provides a major improvement in differentiating biomes characterized by intermediate variability in surface [Chl*a*]. The L2 variance bin, for instance (0.026 mg Chl $a$ $m^{-3}$ to 0.09 mg Chl $a$ $m^{-3}$) includes large numbers of time series from the EQ, SDCM, and LCB biomes, despite vastly different subsurface bio-optical dynamics and environmental conditions in each case. This is consistent with the strong correlations between $[Chl a]_{sat}$ and float-measured [Chl*a*] in the first optical depth but weak correlations between $[Chl a]_{sat}$ and $[Chl a]_{pz}$ in EQ, SDCM, and LCB time series, indicating that satellite measurements fail to adequately capture subsurface variability in [Chl*a*] in these biomes.

### 4.2. Environmental Drivers of Phytoplankton Distributions

Principal components' analysis of the environmental parameters associated with each time series highlighted three broad biogeochemical settings within the global data set: (a) Relatively low-light conditions in the upper layer with elevated mixed layer nutrient concentrations in the case of HCB and LCB biomes; (b) high-light conditions with strong vertical nutrient fluxes in the case of AR and EQ biomes; and (c) low-light conditions with weak vertical nutrient fluxes in the case of SDCM and PDCM biomes (Figure 6). These pairs of biomes are further differentiated based on seasonal variability in environmental and bio-optical parameters that is not readily captured by performing PCA on the overall data set.

#### 4.2.1. HCB and LCB Biomes

Both the HCB and LCB biomes were characterized by summer phytoplankton blooms that coincided with (a) an increase in average light levels within the mixed layer (Figure 7a, Table S8 in Supporting Information S1) and (b) a gradual drawdown of nitrate within the mixed layer, especially in the HCB biome (Figure 7b, Table S8 and Figures S11d and S12d in Supporting Information S1). The strength of these relationships in the LCB and HCB biomes highlights variability in $iPAR_{MLD}$ as being the primary driver of biomass accumulation in high latitudes, albeit against a background of perennially elevated nutrient concentrations (Obata et al., 1996; Sverdrup, 1953).

Despite these similarities, there were marked differences in the magnitude of the seasonal contrast in vertical [Chl*a*] distributions between the two biomes. $[Chl a]_{pz}$ values tripled between winter and summer months in the HCB time series, but barely doubled over the same time period in the LCB time series (Figure 4, Table 2). Similarly, $[Chl a]_{sat}$ tripled between winter and summer months in the HCB time series, while remaining comparatively constant throughout the year in the LCB time series (Table S4 in Supporting Information S1). Given that the majority of the LCB time series are located within the Southern Ocean, and given the perennially elevated nitrate concentrations within the mixed layer in the LCB time series (Figure S11d in Supporting Information S1), the reduced magnitude of summer $[Chl a]_{pz}$ and $[Chl a]_{sat}$ values in the LCB biome likely in part reflects

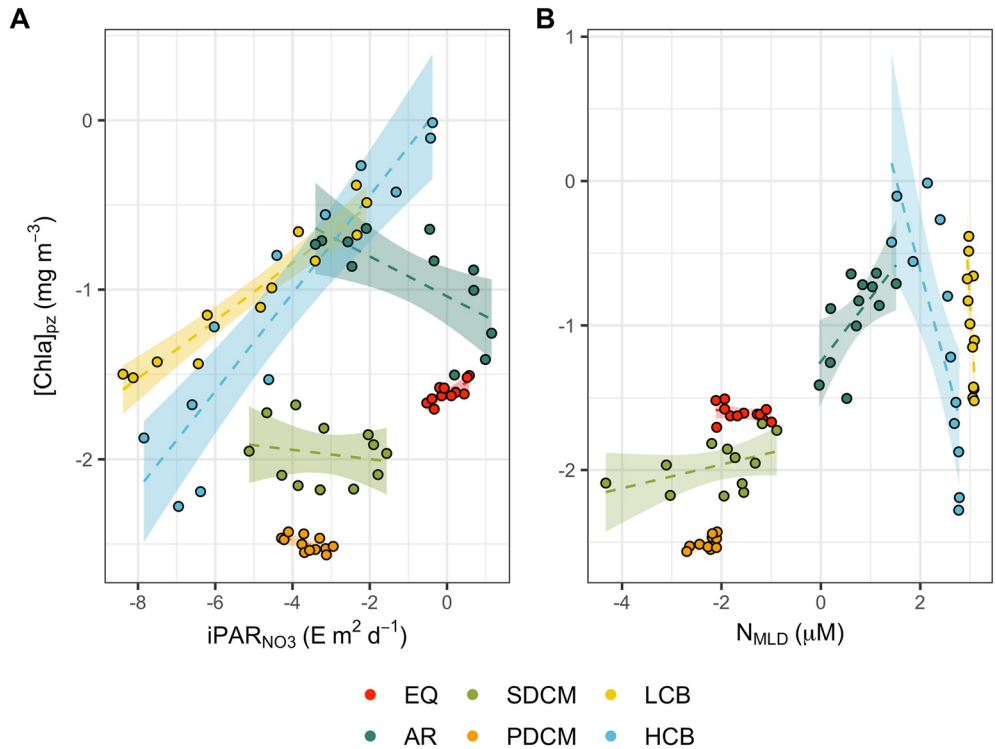


**Figure 7.** Relationships between log-transformed values for (a) $iPAR_{NO3}$ and $[Chla]_{pz}$ and (b) $MLD_N$ and $[Chla]_{pz}$ for equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low chlorophyll bloom (LCB), and high chlorophyll bloom (HCB) biomes. Individual points correspond to the average monthly value for each parameter. Dashed lines correspond to linear regressions (see regression results in Table S8 in Supporting Information S1). Ribbon corresponds to regression standard error.

limited iron availability within the productive zone (Boyd & Doney, 2002; C. M. Moore et al., 2013; Venables & Moore, 2010). The limited annual variability in $[Chla]_{sat}$ and the weak correlations between $[Chla]_{pz}$ and $[Chla]_{sat}$ within the LCB biome (Figures 8a and 8b) may be attributable to increased fluorescence per unit Chl*a* under iron limitation (Behrenfeld et al., 2006, 2009), which could increase float measurements of [Chl*a*] (based on chlorophyll fluorescence) relative to those made via satellite (based on ocean color).

Differences in [Chl*a*] distributions within HCB and LCB time series are also likely due to seasonal differences in the extent of vertical mixing between the two biomes. Comparatively deeper winter mixing in the HCB time series results in lower temperatures and reduced average irradiance for phytoplankton (Lacour et al., 2015), both of which are likely to reduce growth rates (Walter et al., 2015) and the concentration of algal cells in the mixed layer (Evans et al., 1985). Deeper summer mixing in the LCB time series presumably has a similar effect, resulting in lower average light levels within the mixed layer during maximal stratification compared to the HCB time series. Indeed, maximal monthly averaged iPAR values in the LCB biome are roughly half those in the HCB biome at the peak of the summer bloom and are accompanied by significantly lower values for all bio-optical parameters (Figures 8a–8c, Figures S11f and S12f in Supporting Information S1).

Finally, although the HCB and LCB biomes were clearly differentiated by the clustering algorithm (primarily due to the much larger weightings associated with the second singular vector in the HCB time series compared to the LCB time series [Figure 2]), there was a significant overlap in the global distribution of the LCB and HCB biomes (Figure 3, Figure S4 in Supporting Information S1). In particular, the occurrence of time series belonging to both biomes within the Southern Ocean is likely due to interannual or regional variability in bloom intensity, potentially in response to localized increases in iron concentrations along the sampling track of individual floats, hydrothermal vent activity (Ardyna et al., 2019; Schine et al., 2021), atmospheric deposition (Erickson et al., 2003), advection of iron-enriched water (Ellwood et al., 2008; Sedwick et al., 2008), or inputs from islands (Blain et al., 2007). Given that the HCB time series in the Southern Ocean were somewhat concentrated within

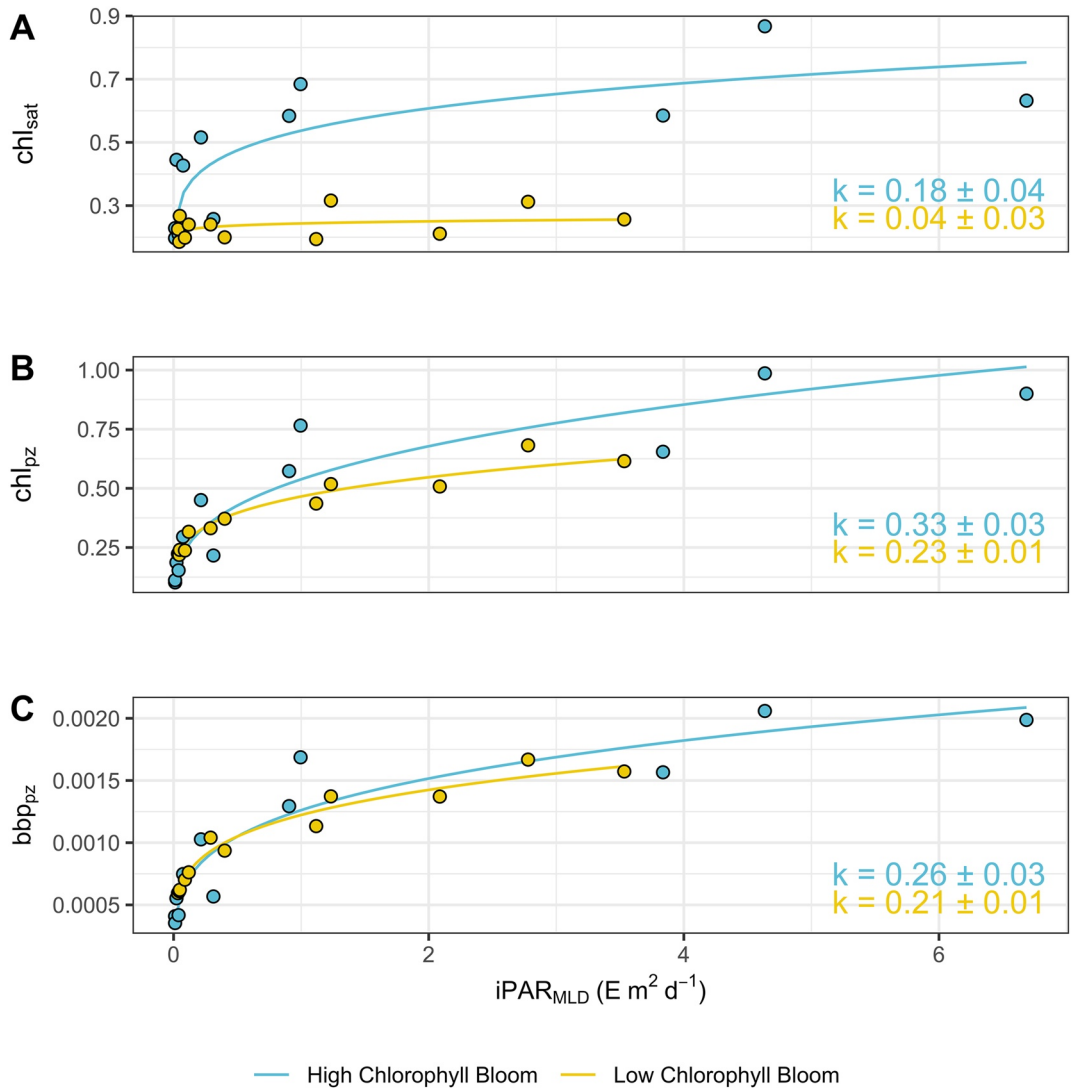


**Figure 8.** Relationships between monthly average $iPAR_{MLD}$ and monthly average values for (a) $[Chl a]_{sat}$, (b) $[Chl a]_{pz}$, and (c) $b_{bp\_pz}$ for HCB time series (blue) and LCB time series (yellow). Line corresponds to a power law function fit to the data, where $y = a \times x^k$. Model estimates for $k$ included for each comparison.

higher latitudes, their occurrence may also reflect local mitigation in nutrient limitation due to continental runoff or upwelling of bottom water (de Baar et al., 1995; Klunder et al., 2011), or locations with comparatively reduced MLDs during summer months (Dong et al., 2008; Kara et al., 2003).

#### 4.2.2. Seasonal DCM and Permanent DCM Biomes

Aside from the overall similarity in biogeochemical conditions highlighted by PCA results, the SDCM and PDCM biomes were both characterized by the formation of DCMs at depths considerably greater than those observed for other low-latitude biomes. However, SDCM time series were clearly differentiated from PDCM time series by the occurrence of a prominent spring bloom, as reflected by seasonal increases in $[Chl a]_{pz}$ (Figure 4, Table 2). It is worth highlighting, though, that the SDCM biome represents an admixture of time series from two geographically distinct systems. The STFZ time series captures a transition zone between the nutrient-limited waters of the central gyres and the nutrient-rich waters of the Southern Ocean, as reflected in part by globally intermediate nutrient concentrations within the mixed layer during winter months (Smythe-Wright et al., 1998; Spalding et al., 2012). In contrast, MED time series capture comparatively nutrient limited conditions, especially with

respect to $PO_4$, reflecting the oligotrophic status of the Mediterranean Sea (D'Ortenzio & Ribera d'Alcalà, 2008), especially in the Eastern basin. There were also a greater proportion of profiles in STFZ time series where the MLD exceeded the nitracline depth during winter mixing compared to MED time series (data not shown), highlighting the importance of advective nutrient transport to spring bloom dynamics in the Mediterranean (D'Ortenzio et al., 2014; Lavigne et al., 2013).

As a result of reduced nutrient availability, it is likely that seasonal [Chl*a*] variability in MED time series is strongly influenced by photoacclimation processes, whereby phytoplanktonic cells increase cellular concentrations of photopigments in response to light limitation (Fennel & Boss, 2003; Letelier et al., 2004; Mignot et al., 2014). This expectation is consistent with previous studies, which have demonstrated photoacclimation to play a major role in seasonal chlorophyll dynamics in the Mediterranean, both during the spring bloom (Bellacicco et al., 2016) and in the formation of the DCM (Cornec et al., 2021). Indeed, in our results, relationships between $[Chla]_{pz}$ and $b_{bp\_pz}$ were considerably weaker in MED time series compared to STFZ time series, suggesting changes in $[Chla]_{pz}$ in response to factors other than phytoplankton biomass alone. Saharan dust deposits have also been reported to enhance $b_{bp}$ relative to [Chl*a*] in the Mediterranean (Claustre et al., 2002), which may further weaken relationships between $[Chla]_{pz}$ and $b_{bp\_pz}$ within MED time series.

In contrast to SDCM time series, there were only minimal winter increases in $[Chla]_{pz}$ in PDCM time series, despite maximal MLDs in PDCM time series being similar or greater than those in SDCM time series. This reflects that phytoplankton biomass in oligotrophic regions (e.g., areas characterized by low levels of surface [Chl*a*] and dominated by the subtropical gyres) is not limited by weak vertical mixing so much as by a perennially deep nitracline inaccessible to mixing compared to other biomes (McClain et al., 2004). Globally minimal $iPAR_{NO3}$ values at the depth of the nitracline in the PDCM biome, along with weak correlations between $b_{bp\_pz}$ and $[Chla]_{pz}$, again suggest the critical importance of photoacclimation processes at and above the depth of the DCM in subtropical gyres as reported previously (Cornec et al., 2021; Mignot et al., 2014). Finally, the homogenous $b_{bp}$ distributions in PDCM profiles, contrasted to the unimodal distributions of [Chl*a*] (Figures 4, Figure 5), likely reflect the dominance of the $b_{bp}$ signal by non-algal biomass within the euphotic zone, as has been previously documented in oligotrophic regions for both attenuation coefficients (Claustre et al., 1999; Grob et al., 2007) and $b_{bp}$ (Bellacicco et al., 2019).

#### 4.2.3. Equatorial and Arabian Sea Biomes

The EQ and AR biomes both capture high-light and high-nutrient flux conditions compared to SDCM and PDCM biomes, resulting in a greater degree of biomass accumulation within the euphotic zone than might otherwise be possible (Beckmann & Hense, 2007; Cullen, 2015; Herbland & Voituriez, 1979). However, the two biomes are unambiguously distinguished from one another by the influence of the monsoon system in the AR biome, resulting in a marked bimodal seasonality in [Chl*a*] and $b_{bp}$ distributions. During monsoon periods, wind-induced upwelling results in elevated nitrate concentrations within the mixed layer and the coincidence of the MLD and nitracline (Bauer et al., 1991). As a result, and in contrast to the EQ biome, nutrient concentration within the mixed layer is the primary driver of biomass accumulation within the AR biome, as reflected by a positive relationship between $[Chla]_{pz}$ and $N_{MLD}$ and a negative relationship between $[Chla]_{pz}$ and $iPAR_{NO3}$ (Figure 7a, Table S8 in Supporting Information S1). Notably, this is the opposite of the trends observed in the LCB and HCB biomes, highlighting the distinct drivers of bloom dynamics across the three biomes: detrainment and subsequent depletion of nutrients in HCB and LCB, and sustained nutrient delivery via upwelling in AR (Chowdhury et al., 2021; Cullen et al., 2002). Shoaling of the MLD during the intermonsoon results in a rapid depletion of nitrate near the surface, accompanied by the formation of a DCM, and increases in $iPAR_{NO3}$ (Mccreary et al., 1996; Prasanth et al., 2021; Ravichandran et al., 2012).

In contrast, [Chl*a*] and $b_{bp}$ distributions in the EQ biome are representative of typical tropical structure, characterized by limited seasonal variability in all environmental parameters, resulting in a persistent DCM coincident with the depth of the nitracline (Herbland & Voituriez, 1979). It is also worth highlighting that time series in the Bay of Bengal were grouped with EQ time series instead of AR time series. This reflects the weakened influence of monsoon cycling on phytoplankton distributions in the Bay of Bengal compared to the AR, likely due to elevated stratification and reduced upwelling throughout the year in the Bay of Bengal (Gauns et al., 2005). Indeed, any monsoon influence for the Bay of Bengal floats is not readily discernable in the average [Chl*a*] or $b_{bp}$ time series for the EQ biome (Figures 4 and 5).

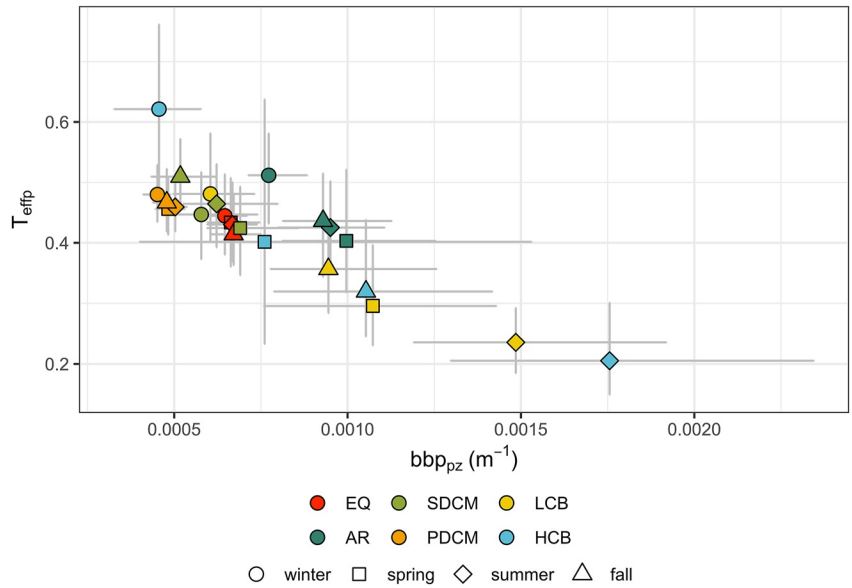


**Figure 9.** Relationship between transfer efficiency ($T_{effp}$) and euphotic zone particle concentration ($b_{bp\_pz}$) for equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low chlorophyll bloom (LCB), and high chlorophyll bloom (HCB) biomes. Individual points correspond to seasonal median values for each parameter. Point shapes correspond to season. Error bars correspond to the first and third quartiles for each parameter.

### 4.3. Seasonal Trends in Biomass Accumulation

The high vertical and temporal resolution of $b_{bp}$ measurements in our data allows us to investigate the variability of the organic particle stocks between the productive and the mesopelagic zone for each biome. It is important to note, though, that as a result of removing spikes in $b_{bp}$ profiles, our results primarily reflect contributions of the small and/or suspended fraction of overall particle stocks (Briggs et al., 2011; Galí et al., 2021). With this, our estimates of transfer efficiency and particle stocks across different depth horizons should be interpreted as representing only a portion of the particle stocks, especially for high-latitude biomes.

The observed inverse relationship between euphotic zone biomass ($b_{bp\_pz}$) and $T_{effp}$ (Figures 6 and 9) is consistent with previous studies reporting inverse relationships between primary production and transfer efficiency both within the Southern Ocean (Cavan et al., 2015) and at the global scale (Henson et al., 2019). This inverse relationship is primarily driven by seasonal variability in the $b_{bp\_pz}$ and $T_{effp}$, reflecting a seasonal decoupling of the particle stocks between the euphotic and mesopelagic layer, especially in the SDCM, LCB, and HCB biomes (Figure 10). This decoupling could result from various overlapping effects: (a) increased grazing rates by zooplankton over the course of the phytoplankton growing season, which would increase the recycling of biomass within the euphotic zone (Cavan et al., 2017); (b) time lags between production of organic particles in the productive layer and its export at depth (Laws & Maiti, 2019); (c) fragmentation of large particles formed in the euphotic zone within the upper mesopelagic zone (Briggs et al., 2020); (d) variability in the size or sinking speeds of particles in the euphotic zone (Henson et al., 2015); and (e) seasonal changes in remineralization rates in response to changes in temperature (Laws et al., 2000). However, further analysis is required to evaluate the relative importance of these processes across different biomes.

In contrast to the SDCM, LCB, and HCB biomes, the EQ and PDCM biomes show virtually no seasonal variability in transfer efficiency with respect to euphotic zone particle stocks (Figure 9). This is likely due to the more static environmental setting of these biomes, resulting in stable ecosystems dominated by smaller cells throughout the year (Landry et al., 1996; Rii et al., 2016). While some degree of seasonality is evident in the PDCM biome, as evidenced by the annual deepening of the DCM and the strong correlations between particle stocks in the euphotic zone and upper mesopelagic zone (Table S7 in Supporting Information S1), there is virtually no variability in particle distributions within the EQ biome, resulting in the nonsignificant correlations between integrated particle stocks across different depth horizons (Table S7 in Supporting Information S1). It is also worth highlighting that in the PDCM and EQ biomes that values for $T_{effp}$ would be least analogous to export ratios based

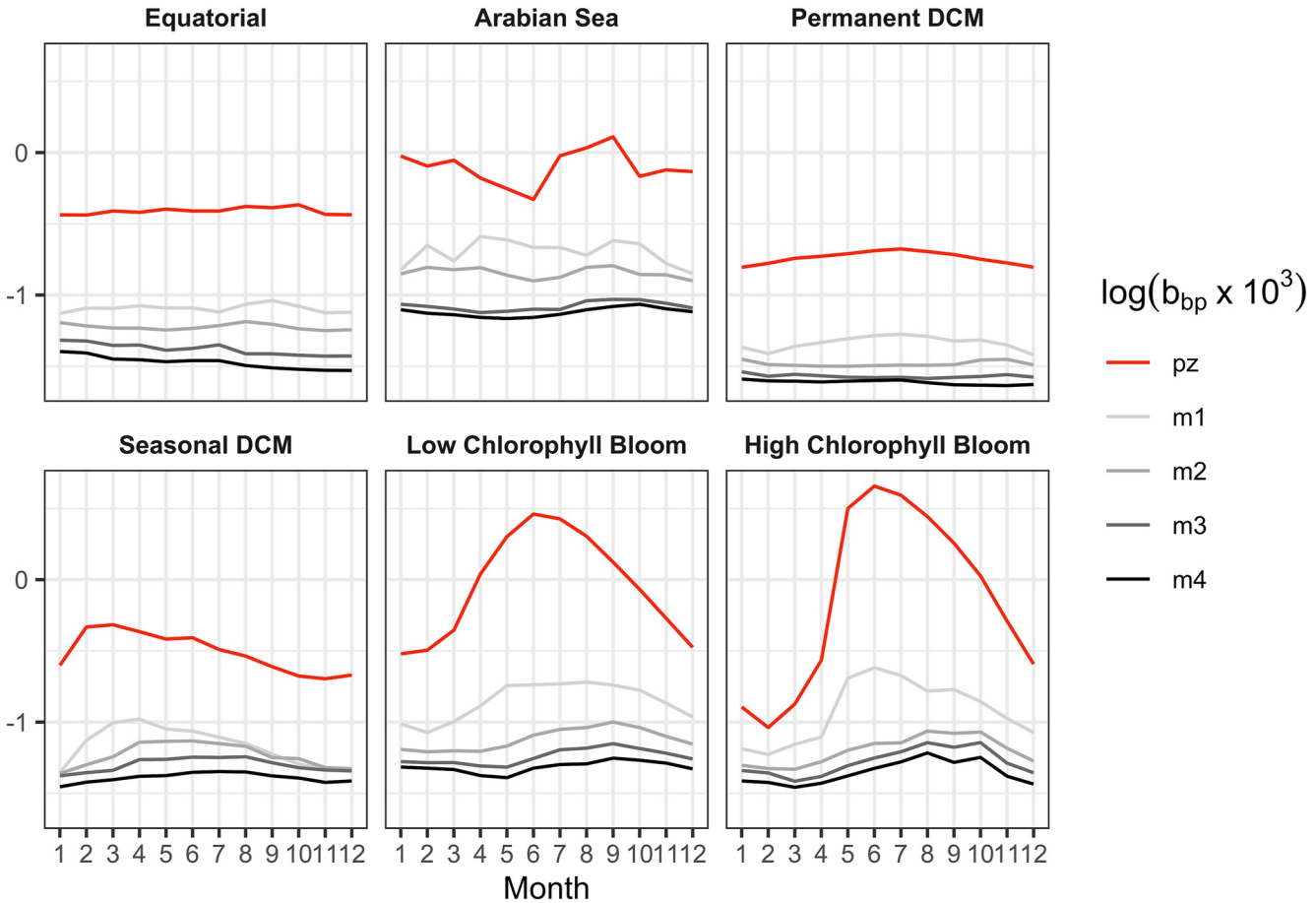


**Figure 10.** Average $b_{bp}$ values within the productive zone (pz) in addition to each of the four equally sized layers of the mesopelagic zone (m1–m4). All values are log-transformed and multiplied by $10^3$ for ease of comparison.

on primary production measurements (Henson et al., 2019), since $b_{bp\_pz}$ values in these biomes capture larger fractions of non-algal biomass than would be anticipated elsewhere (Bellacicco et al., 2019; Claustre et al., 1999).

In the AR biome, weak correlations between particle stocks in the euphotic and mesopelagic zones are presumably attributable to the presence of mesopelagic biomass maxima as evident in Figure 5 at depths of ~300 m. As previously reported, these maxima are likely caused by large populations of denitrifying bacteria at the depth of the oxygen minimum zone (Rasse et al., 2020; Roullier et al., 2014; Ward et al., 2009; Whitmire et al., 2009; Wojtasiewicz et al., 2020). This seems especially likely given the comparatively enhanced seasonality in the lower mesopelagic zone ($b_{bp\_m3}$ and $b_{bp\_m4}$ in Figure 10), and the significant correlation between $b_{bp\_pz}$ and $b_{bp\_m3}$ in the AR biome (Table S7 in Supporting Information S1).

## 5. Conclusions

To date, characterizing phytoplankton distributions at the global scale has been limited by the near invisibility of subsurface phytoplankton distributions to remote sensing platforms. As the issue was described by Platt and Sathyendranath, "to characterize the ocean water column using satellite-based observations of its surface layer is certainly not ideal. But is the only option available at these scales, now and into the foreseeable future" (Platt & Sathyendranath, 1988). However, with modern advances, our results demonstrate that the amount and extent of data collected by BGC-Argo profiling floats are now sufficient to resolve this constraint.

Beyond demonstrating the utility of data from BGC-Argo floats for partitioning biogeographical provinces within the global ocean, this analysis provides a new insight into the drivers of subsurface variability in phytoplankton distributions at the global scale, while also highlighting the role of subsurface processes in driving biomass

accumulation in the mesopelagic zone. Our results demonstrate the diversity of biogeochemical and hydrological interactions shaping phytoplankton distributions in the global ocean, especially highlighting (a) the importance of daily integrated irradiance within the mixed layer as a driver of biomass accumulation in high-latitude biomes; (b) the importance of light level at the depth of the nitracline and nutrient fluxes as drivers of biomass accumulation in low to midlatitude biomes; (c) patterns of co-limitation of light and nutrients in the onset of the summer bloom in the Southern Ocean; and (d) globally variable rates of biomass attenuation and accumulation for small particles in the mesopelagic zone.

The classifications in this study additionally extend beyond existing methods by fully accounting for vertical variability in phytoplankton distributions and particle concentrations within the upper kilometer of the ocean. In doing so, they provide a global template for contextualizing profile measurements of phytoplankton community structure and/or HPLC pigments from ship-based research programs (e.g., JGOFS, AMT, and HOT), or for constraining estimates of subsurface phytoplankton distributions for modeling primary production (Westberry et al., 2008) or phytoplankton community structure (Morel & Berthon, 1989; Uitz et al., 2006). Various seasonal patterns regarding the dynamics of small particles at depth also highlight the further potential for using subsurface classifications to constrain estimates of parameters related to the biological carbon pump (Boyd et al., 2019; Galí et al., 2021).

Any of these applications would benefit from extrapolating our current results to regions not directly sampled by floats, effectively constraining biogeographical classifications based on surface data by parameterizing geographic distributions of subsurface trends. Toward this end, we found a random forest classifier trained on latitude, longitude, and the standard deviation of $[\mathrm{Chl}a]_{\mathrm{sat}}$ for each time series to correctly predict biome membership for individual profiles in more than 96% of the cases (data not shown). And while our current data set is too sparse to apply this classifier to historical MODIS data with informative results, such a goal will surely become feasible in the near-future with the increasing deployment of additional BGC-Argo floats.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

All bio-optical and environmental data used in this study were collected and made freely available by the International Argo Program and the national programs that contribute to it (https://argo.ucsd.edu; https://www.ocean-ops.org). The Argo Program is part of the Global Ocean Observing System. All analyses were performed in R 4.0.2 using RStudio 1.3.107. RStudio is an open-source software freely available through the publisher website (https://www.rstudio.com/).

**Acknowledgments**
This study was funded by a European Research Council Advanced Grant to Hervé Claustre for the REFINE project (Grant Agreement 834177). Participation by N. Bock and S. Duhamel was additionally supported by National Science Foundation (NSF) OCE-1458070 awarded to S. Duhamel. The authors wish to thank the Argo Data Management team (ADMT) and the BGC-Argo Data Management task team (BGC ADMT).

## References

Ardyna, M., Lacour, L., Sergi, S., d'Ovidio, F., Sallée, J. B., Rembauville, M., et al. (2019). Hydrothermal vents trigger massive phytoplankton blooms in the Southern Ocean. *Nature Communications*, *10*(1), 1–8. https://doi.org/10.1038/s41467-019-09973-6

Argo. (2000). *Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)*. SEANOE. https://doi.org/10.17882/42182

Barbieux, M., Uitz, J., Gentili, B., Pasqueron De Fommervault, O., Mignot, A., Poteau, A., et al. (2019). Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database. *Biogeosciences*, *16*(6), 1321–1342. https://doi.org/10.5194/bg-16-1321-2019

Bauer, S., Hitchcock, G. L., & Olson, D. B. (1991). Influence of monsoonally-forced Ekman dynamics upon surface layer depth and plankton biomass distribution in the Arabian Sea. *Deep Sea Research Part A, Oceanographic Research Papers*, *38*(5), 531–553. https://doi.org/10.1016/0198-0149(91)90062-K

Beckmann, A., & Hense, I. (2007). Beneath the surface: Characteristics of oceanic ecosystems under weak mixing conditions – A theoretical investigation. *Progress in Oceanography*, *75*(4), 771–796. https://doi.org/10.1016/j.pocean.2007.09.002

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, *19*(1), 1–14. https://doi.org/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*(5), 779–794. https://doi.org/10.5194/bg-6-779-2009

Behrenfeld, M. J., Worthington, K., Sherrell, R. M., Chavez, F. P., Strutton, P., McPhaden, M., & Shea, D. M. (2006). Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. *Nature*, *442*(7106), 1025–1028. https://doi.org/10.1038/nature05083

Bellacicco, M., Cornec, M., Organelli, E., Brewin, R. J. W., Neukermans, G., Volpe, G., et al. (2019). Global variability of optical backscattering by non-algal particles from a biogeochemical-Argo data set. *Geophysical Research Letters*, *46*(16), 9767–9776. https://doi.org/10.1029/2019GL084078

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

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., et al. (2018). An alternative to static climatologies: Robust estimation of open ocean $CO_2$ variables and nutrient concentrations from T, S, and $O_2$ data using Bayesian neural networks. *Frontiers in Marine Science*, *5*(Sep), 1–30. https://doi.org/10.3389/fmars.2018.00328

Blain, S., Quéguiner, B., Armand, L., Belviso, S., Bombled, B., Bopp, L., et al. (2007). Effect of natural iron fertilization on carbon sequestration in the Southern Ocean. *Nature*, *446*(7139), 1070–1074. https://doi.org/10.1038/nature05700

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., & Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature*, *568*(7752), 327–335. https://doi.org/10.1038/s41586-019-1098-2

Boyd, P. W., & Doney, S. C. (2002). Modelling regional responses by marine pelagic ecosystems to global climate change. *Geophysical Research Letters*, *29*(16), 1806. https://doi.org/10.1029/2001gl014130

Briggs, N., Dall'Olmo, G., & Claustre, H. (2020). Major role of particle fragmentation in regulating biological sequestration of $CO_2$ by the oceans. *Science*, *367*(6479), 791–793. https://doi.org/10.1126/science.aay1790

Briggs, N., Perry, M. J., Cetinić, I., Lee, C., D'Asaro, E., Gray, A. M., & Rehm, E. (2011). High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom. *Deep-Sea Research Part I: Oceanographic Research Papers*, *58*(10), 1031–1039. https://doi.org/10.1016/j.dsr.2011.07.007

Brinton, E. (1962). UC San Diego Bulletin of the Scripps Institution of Oceanography. *Bulletin of the Scripps Institution of Oceanography of the University of California*, *8*(2), 51–270.

Cavan, E. L., Henson, S. A., Belcher, A., & Sanders, R. (2017). Role of zooplankton in determining the efficiency of the biological carbon pump. *Biogeosciences*, *14*(1), 177–186. https://doi.org/10.5194/bg-14-177-2017

Cavan, E. L., Le Moigne, F. A. C., Poulton, A. J., Tarling, G. A., Ward, P., Daniels, C. J., et al. (2015). Attenuation of particulate organic carbon flux in the Scotia Sea, Southern Ocean, is controlled by zooplankton fecal pellets. *Geophysical Research Letters*, *42*(3), 821–830. https://doi.org/10.1002/2014GL062744

Chowdhury, M., Biswas, H., Mitra, A., Silori, S., Sharma, D., Bandyopadhyay, D., et al. (2021). Southwest monsoon-driven changes in the phytoplankton community structure in the central Arabian Sea (2017–2018): After two decades of JGOFS. *Progress in Oceanography*, *197*(December 2020), 102654. https://doi.org/10.1016/j.pocean.2021.102654

Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the global ocean with biogeochemical-Argo. *Annual Review of Marine Science*, *12*, 23–48. https://doi.org/10.1146/annurev-marine-010419-010956

Claustre, H., Morel, A., Babin, M., Cailliau, C., Marie, D., Marty, J., et al. (1999). Variability in particle attenuation and chlorophyll fluorescence in the tropical Pacific: Scales, patterns, and biogeochemical implications. *Journal of Geophysical Research*, *104*(C2), 3401–3422. https://doi.org/10.1029/98JC01334

Claustre, H., Morel, A., Hooker, S. B., Babin, M., Antoine, D., Oubelkheir, K., et al. (2002). Is desert dust making oligotrophic waters greener? *Geophysical Research Letters*, *29*(10), 107-1–107-4. https://doi.org/10.1029/2001gl014056

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*(4), 1–30. https://doi.org/10.1029/2020gb006759

Cullen, J. (2015). Subsurface chlorophyll maximum layers: Enduring enigma or mystery solved? *Annual Review of Marine Science*, *7*, 207–239. https://doi.org/10.1146/annurev-marine-010213-135111

Cullen, J., Franks, P., Karl, D., & Longhurst, A. (2002). Physical influences on marine ecosystem dynamics. *Sea*, *12*, 297–336.

Dall'Olmo, G., & Mork, K. A. (2014). Carbon export by small particles in the Norwegian Sea. *Geophysical Research Letters*, *41*(8), 2921–2927. https://doi.org/10.1002/2014GL059244

de Baar, H. J. W., Bathmannt, U., Smetacek, V., Löscher, B. M., Veth, C., Bathmann, U., & Smetacek, V. (1995). Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean. *Nature*, *373*(6513), 412–415. https://doi.org/10.1038/373412a0

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*(12), 1–20. https://doi.org/10.1029/2004JC002378

Dong, S., Sprintall, J., Gille, S. T., & Talley, L. (2008). Southern ocean mixed-layer depth from Argo float profiles. *Journal of Geophysical Research*, *113*(6), 1–12. https://doi.org/10.1029/2006JC004051

D'Ortenzio, F., Lavigne, H., Besson, F., Claustre, H., Coppola, L., Garcia, N., et al. (2014). Observing mixed layer depth, nitrate and chlorophyll concentrations in the northwestern Mediterranean: A combined satellite and $NO_3$ profiling floats experiment. *Geophysical Research Letters*, *41*(18), 6443–6451. https://doi.org/10.1002/2014GL061020

D'Ortenzio, F., & Ribera d'Alcalà, M. (2008). On the trophic regimes of the Mediterranean Sea: A satellite analysis. *Biogeosciences Discussions*, *5*(4), 2959–2983. https://doi.org/10.5194/bgd-5-2959-2008

Dubinsky, Z., & Stambler, N. (2009). Photoacclimation processes in phytoplankton: Mechanisms, consequences, and applications. *Aquatic Microbial Ecology*, *56*(2–3), 163–176. https://doi.org/10.3354/ame01345

Ellwood, M. J., Boyd, P. W., & Sutton, P. (2008). Winter-time dissolved iron and nutrient distributions in the Subantarctic Zone from 40-52S; 155-160E. *Geophysical Research Letters*, *35*(11), 2–7. https://doi.org/10.1029/2008GL033699

Erickson, D. J., Hernandez, J. L., Ginoux, P., Gregg, W. W., McClain, C., & Christian, J. (2003). Atmospheric iron delivery and surface ocean biological activity in the southern ocean and Patagonian region. *Geophysical Research Letters*, *30*(12), 1–4. https://doi.org/10.1029/2003GL017241

Evans, G. T., Parslow, J. S., Evans, G. T., & Parslow, J. S. (1985). A model of annual plankton cycles. *Deep Sea Research – Part B: Oceanographic Literature Review*, *32*(9), 759. https://doi.org/10.1016/0198-0254(85)92902-4

Fay, A. R., & McKinley, G. A. (2014). Global open-ocean biomes: Mean and temporal variability. *Earth System Science Data*, *6*(2), 273–284. https://doi.org/10.5194/essd-6-273-2014

Fennel, K., & Boss, E. (2003). Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model. *Limnology & Oceanography*, *48*(4), 1521–1534. https://doi.org/10.4319/lo.2003.48.4.1521

Field, C. B., Behrenfeld, M. J., Randerson, J. T., Field, C. B., Behrenfeld, M. J., & Randerson, J. T. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

Fourrier, M., Coppola, L., Claustre, H., D'Ortenzio, F., Sauzède, R., & Gattuso, J.-P. (2020). A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean Sea: CANYON-MED. *Frontiers in Marine Science*, *7*, 620. https://doi.org/10.3389/fmars.2020.00620

Fourrier, M., Coppola, L., Claustre, H., D'Ortenzio, F., Sauzède, R., & Gattuso, J.-P. (2021). Corrigendum: A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean Sea: CANYON-MED. *Frontiers in Marine Science*, *8*, 620. https://doi.org/10.3389/fmars.2021.650509

Galí, M., Falls, M., Claustre, H., Aumont, O., & Bernardello, R. (2021). Bridging the gaps between particulate backscattering measurements and modeled particulate organic carbon in the ocean. *Biogeosciences Discussions*, *2021*, *19*, 1245–1275. Retrieved from https://bg.copernicus.org/preprints/bg-2021-201/

Gauns, M., Madhupratap, M., Ramaiah, N., Jyothibabu, R., Fernandes, V., Paul, J. T., & Prasanna Kumar, S. (2005). Comparative accounts of biological productivity characteristics and estimates of carbon fluxes in the Arabian Sea and the Bay of Bengal. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *52*(14–15), 2003–2017. https://doi.org/10.1016/j.dsr2.2005.05.009

Gong, X., Jiang, W., Wang, L., Gao, H., Boss, E., Yao, X., et al. (2017). Analytical solution of the nitracline with the evolution of subsurface chlorophyll maximum in stratified water columns. *Biogeosciences*, *14*(9), 2371–2386. https://doi.org/10.5194/bg-14-2371-2017

Gregg, W. W., & Carder, K. L. (1990). A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnology & Oceanography*, *35*(8), 1657–1675. https://doi.org/10.4319/lo.1990.35.8.1657

Grob, C., Ulloa, O., Li, W. K. W. W., Alarcon, G., Fukasawa, M., Watanabe, S., et al. (2007). Picoplankton abundance and biomass across the eastern South Pacific Ocean along latitude 32.5°S. *Marine Ecology Progress Series*, *332*, 53–62. https://doi.org/10.3354/Meps332053

Hardman Mountford, N. J., Hirata, T., Richardson, K. A., & Aiken, J. (2008). An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean. *Remote Sensing of Environment*, *112*(8), 3341–3352. https://doi.org/10.1016/j.rse.2008.02.016

Hardman-Mountford, N. J., & McGlade, J. M. (2002). Defining ecosystem structure from natural variability: Application of principal components analysis to remotely sensed SST. *Large Marine Ecosystems*, *11*(C), 67–82. https://doi.org/10.1016/S1570-0461(02)80028-3

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer New York. Retrieved from https://www.jstor.org/stable/2980421?origin=crossref

Henson, S., Le Moigne, F., & Giering, S. (2019). Drivers of carbon export efficiency in the global ocean. *Global Biogeochemical Cycles*, *33*(7), 891–903. https://doi.org/10.1029/2018GB006158

Henson, S., Yool, A., & Sanders, R. (2015). Global biogeochemical cycles carbon export: A model study. *Global Biogeochemical Cycles*, *29*, 33–45. https://doi.org/10.1002/2014GB004965.Received

Herbland, A., & Voituriez, B. (1979). Hydrological structure analysis for estimating the primary production in the tropical Atlantic Ocean. *Journal of Marine Research*, *37*(1), 87–101.

Kara, A. B., Rochford, P. A., & Hurlburt, H. E. (2003). Mixed layer depth variability over the global ocean. *Journal of Geophysical Research*, *108*(3), 1–15. https://doi.org/10.1029/2000jc000736

Klunder, M. B., Laan, P., Middag, R., De Baar, H. J., & van Ooijen, J. C. (2011). Dissolved iron in the Southern Ocean (Atlantic sector). *Deep-Sea Research Part II: Topical Studies in Oceanography*, *58*(25–26), 2678–2694. https://doi.org/10.1016/j.dsr2.2010.10.042

Krug, L. A., Platt, T., Sathyendranath, S., & Barbosa, A. B. (2017). Ocean surface partitioning strategies using ocean colour remote sensing: A review. *Progress in Oceanography*, *155*, 41–53. https://doi.org/10.1016/j.pocean.2017.05.013

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*(11), 836–839. https://doi.org/10.1038/NGEO3035

Lacour, L., Claustre, H., Prieur, L., & D'Ortenzio, F. (2015). Phytoplankton biomass cycles in the North Atlantic subpolar gyre: A similar mechanism for two different blooms in the Labrador Sea. *Geophysical Research Letters*, *42*(13), 5403–5410. https://doi.org/10.1002/2015GL064540

Landry, M. R., Kirshtein, J., & Constantinou, J. (1996). Abundances and distributions of picoplankton populations in the central equatorial Pacific from 12°N to 12°S, 140°W. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *43*(4–6), 871–890. https://doi.org/10.1016/0967-0645(96)00018-5

Lavigne, H., D'Ortenzio, F., Migon, C., Claustre, H., Testor, P., D'Alcalà, M. R., et al. (2013). Enhancing the comprehension of mixed layer depth control on the Mediterranean phytoplankton phenology. *Journal of Geophysical Research: Oceans*, *118*(7), 3416–3430. https://doi.org/10.1002/jgrc.20251

Lavigne, H., Ortenzio, F., Ribera D'Alcala, M., Claustre, H., Sauzede, R., & 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. https://doi.org/10.5194/bg-12-5021-2015

Laws, E. A., Falkowski, P. G., Smith, W. O., Ducklow, H., & McCarthy, J. J. (2000). Temperature effects on export production in the open ocean. *Global Biogeochemical Cycles*, *14*(4), 1231–1246. https://doi.org/10.1029/1999GB001229

Laws, E. A., & Maiti, K. (2019). The relationship between primary production and export production in the ocean: Effects of time lags and temporal variability. *Deep-Sea Research Part I: Oceanographic Research Papers*, *148*(October 2018), 100–107. https://doi.org/10.1016/j.dsr.2019.05.006

Letelier, R. M., Karl, D. M., Abbott, M. R., & Bidigare, R. R. (2004). Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. *Limnology & Oceanography*, *49*(2), 508–519. https://doi.org/10.4319/lo.2004.49.2.0508

Longhurst, A. R. (1995). Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, *36*(2), 77–167. https://doi.org/10.1016/0079-6611(95)00015-1

Longhurst, A. R. (2007). *Ecological geography of the sea* (2nd ed.). Elsevier. https://doi.org/10.1016/B978-0-12-455521-1.X5000-1

Longhurst, A. R., Sathyendranath, S., Platt, T., & Caverhill, C. (1995). An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research*, *17*(6), 1245–1271. https://doi.org/10.1093/plankt/17.6.1245

McClain, C. R., Signorini, S. R., & Christian, J. R. (2004). Subtropical gyre variability observed by ocean-color satellites. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *51*(1–3), 281–301. https://doi.org/10.1016/j.dsr2.2003.08.002

Mccreary, J. P., Kohler, K. E., Hood, R. R., & Olson, D. B. (1996). A four-component ecosystem model of biological activity in the Arabian Sea. *Progress in Oceanography*, *37*(3–4), 193–240. https://doi.org/10.1016/S0079-6611(96)00005-5

Mignot, A., Claustre, H., Uitz, J., Poteau, A., Ortenzio, F. D., & Xing, X. (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*(8), 856–876. https://doi.org/10.1002/2013GB004781.Received

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., et al. (2013). Processes and patterns of oceanic nutrient limitation. *Nature Geoscience*, *6*(9), 701–710. https://doi.org/10.1038/ngeo1765

Moore, T. S., Campbell, J. W., & Dowell, M. D. (2009). A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sensing of Environment*, *113*(11), 2424–2430. https://doi.org/10.1016/j.rse.2009.07.016

Moore, T. S., Campbell, J. W., & Feng, H. (2001). A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8), 1764–1776. https://doi.org/10.1109/36.942555

Morel, A., & Berthon, J.-F. (1989). Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. *Limnology & Oceanography*, *34*(8), 1545–1562. https://doi.org/10.4319/lo.1989.34.8.1545

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007). Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. *Remote Sensing of Environment*, *111*(1), 69–88. https://doi.org/10.1016/j.rse.2007.03.012

Mustapha, Z. B., Alvain, S., Jamet, C., Loisel, H., & Dessailly, D. (2014). Remote Sensing of Environment Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters. *Remote Sensing of Environment*, *146*, 97–112. https://doi.org/10.1016/j.rse.2013.08.046

Obata, A., Ishizaka, J., & Endoh, M. (1996). Global verification of critical depth theory for phytoplankton bloom with climatological in situ temperature and satellite ocean color data. *Journal of Geophysical Research*, *101*(C9), 20657–20667. https://doi.org/10.1029/96JC01734

Pielou, E. (1984). *The interpretation of ecological data: A primer on classification and ordination*. John Wiley & Sons.

Platt, T., & Sathyendranath, S. (1988). Oceanic primary production: Estimation by remote sensing at local and regional scales. *Science*, *241*(4873), 1613–1620. https://doi.org/10.1126/science.241.4873.1613

Prasanth, R., Vijith, V., Thushara, V., George, J. V., & Vinayachandran, P. N. (2021). Processes governing the seasonality of vertical chlorophyll-a distribution in the central Arabian Sea: Bio-Argo observations and ecosystem model simulation. *Deep Sea Research Part II: Topical Studies in Oceanography*, *183*, 104926. https://doi.org/10.1016/j.dsr2.2021.104926

Proud, R., Cox, M. J., & Brierley, A. S. (2017). Biogeography of the global ocean's mesopelagic zone. *Current Biology*, *27*(1), 113–119. https://doi.org/10.1016/j.cub.2016.11.003

Rasse, R., Claustre, H., & Poteau, A. (2020). The suspended small-particle layer in the oxygen-poor Black Sea: A proxy for delineating the effective $N_2$-yielding section. *Biogeosciences*, *17*(24), 6491–6505. https://doi.org/10.5194/bg-17-6491-2020

Ravichandran, M., Girishkumar, M. S., & Riser, S. (2012). Observed variability of chlorophyll-a using Argo profiling floats in the southeastern Arabian Sea. *Deep-Sea Research Part I: Oceanographic Research Papers*, *65*, 15–25. https://doi.org/10.1016/j.dsr.2012.03.003

R Core Team. (2016). R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reygondeau, G., Guidi, L., Beaugrand, G., Henson, S. A., Koubbi, P., MacKenzie, B. R., et al. (2018). Global biogeochemical provinces of the mesopelagic zone. *Journal of Biogeography*, *45*(2), 500–514. https://doi.org/10.1111/jbi.13149

Rii, Y., Karl, D., & Church, M. (2016). Temporal and vertical variability in picophytoplankton primary productivity in the North Pacific Subtropical Gyre. *Marine Ecology Progress Series*, *562*(4), 1–18. https://doi.org/10.3354/meps11954

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K. S., King, B., Moum, J., et al. (2019). On the future of Argo: A global, full-depth, multi-disciplinary array. *Frontiers in Marine Science*, *6*, 439. https://doi.org/10.3389/fmars.2019.00439

Roullier, F., Berline, L., Guidi, L., Durrieu De Madron, X., Picheral, M., Sciandra, A., et al. (2014). Particle size distribution and estimated carbon flux across the Arabian Sea oxygen minimum zone. *Biogeosciences*, *11*(16), 4541–4557. https://doi.org/10.5194/bg-11-4541-2014

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sarmiento, J. L., Slater, R., Barber, R., Bopp, L., Doney, S. C., Hirst, A. C., et al. (2004). Response of ocean ecosystems to climate warming. *Global Biogeochemical Cycles*, *18*(3). https://doi.org/10.1029/2003GB002134

Sauzède, R., Bittig, H. C., Claustre, H., de Fommervault, O. P., Gattuso, J. P., Legendre, L., & Johnson, K. S. (2017). Estimates of water-column nutrient concentrations and carbonate system parameters in the Global Ocean: A novel approach based on neural networks. *Frontiers in Marine Science*, *4*(May), 1–18. https://doi.org/10.3389/fmars.2017.00128

Sauzède, R., Claustre, H., Uitz, J., Jamet, C., Dall'Olmo, G., D'Ortenzio, F., et al. (2016). A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient. *Journal of Geophysical Research*, *121*(4), 2552–2571. https://doi.org/10.1002/2015JC011408

Schine, C. M. S., Alderkamp, A. C., van Dijken, G., Gerringa, L. J. A., Sergi, S., Laan, P., et al. (2021). Massive Southern Ocean phytoplankton bloom fed by iron of possible hydrothermal origin. *Nature Communications*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-21339-5

Schmechtig, C., Poteau, A., Claustre, H., D'Ortenzio, F., & Boss, E. (2015). Processing bio-Argo chlorophyll-a concentration at the DAC level. *Argo Data Management*, 1–22.

Schmechtig, C., Poteau, A., Claustre, H., D'Ortenzio, F., Dall'Olmo, G., & Boss, E. S. (2018a). Bio-Argo quality control manual for Chlorophyll-A concentration, 1–16. https://doi.org/10.13155/39459

Schmechtig, C., Poteau, A., Claustre, H., D'Ortenzio, F., Dall'Olmo, G., & Boss, E. S. (2018b). Processing BGC-Argo particle backscattering at the DAC level, 1–15. https://doi.org/10.13155/39459

Sedwick, P. N., Bowie, A. R., & Trull, T. W. (2008). Dissolved iron in the Australian sector of the Southern Ocean (CLIVAR SR3 section): Meridional and seasonal trends. *Deep-Sea Research Part I: Oceanographic Research Papers*, *55*(8), 911–925. https://doi.org/10.1016/j.dsr.2008.03.011

Shi, X. L., Marie, D., Jardillier, L., Scanlan, D. J., & Vaulot, D. (2009). Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS One*, *4*(10), e7657. https://doi.org/10.1371/journal.pone.0007657

Smythe-Wright, D., Chapman, P., Duncombe Rae, C., Shannon, L. V., & Boswell, S. M. (1998). Characteristics of the South Atlantic subtropical frontal zone between 15°W and 5°E. *Deep-Sea Research Part I: Oceanographic Research Papers*, *45*(1), 167–192. https://doi.org/10.1016/S0967-0637(97)00068-X

Spalding, M. D., Agostini, V. N., Rice, J., & Grant, S. M. (2012). Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. *Ocean & Coastal Management*, *60*, 19–30. https://doi.org/10.1016/j.ocecoaman.2011.12.016

Steinberg, D. K., Cope, J. S., Wilson, S. E., & Kobari, T. (2008). A comparison of mesopelagic mesozooplankton community structure in the subtropical and subarctic North Pacific Ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *55*(14–15), 1615–1635. https://doi.org/10.1016/j.dsr2.2008.04.025

Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R., et al. (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences*, *5*(1), 171–201. https://doi.org/10.5194/bg-5-171-2008

Sutton, T. T., Clark, M. R., Dunn, D. C., Halpin, P. N., Rogers, A. D., Guinotte, J., et al. (2017). A global biogeographic classification of the mesopelagic zone. *Deep-Sea Research Part I: Oceanographic Research Papers*, *126*(April), 85–102. https://doi.org/10.1016/j.dsr.2017.05.006

Sverdrup, H. U. (1953). On conditions for the vernal blooming of phytoplankton. *ICES Journal of Marine Science*, *18*(3), 287–295. https://doi.org/10.1093/icesjms/18.3.287

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, *111*(8). https://doi.org/10.1029/2005JC003207

Venables, H., & Moore, C. M. (2010). Phytoplankton and light limitation in the Southern Ocean: Learning from high-nutrient, high-chlorophyll areas. *Journal of Geophysical Research*, *115*(C2). https://doi.org/10.1029/2009jc005361

Walter, B., Peters, J., van Beusekom, J. E. E., & St. John, M. A. (2015). Interactive effects of temperature and light during deep convection: A case study on growth and condition of the diatom Thalassiosira weissflogii. *ICES Journal of Marine Science*, *72*(6), 2061–2071. https://doi.org/10.1093/icesjms/fsu218

Ward, B. B., Devol, A. H., Rich, J. J., Chang, B. X., Bulow, S. E., Naik, H., et al. (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature*, *461*(7260), 78–81. https://doi.org/10.1038/nature08276

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*, 1–18. https://doi.org/10.1029/2007GB003078

Whitmire, A., Ltelier, R., Villagrán, V., & Ulloa, O. (2009). Autonomous observations of in vivo fluorescence and particle backscattering in an oceanic oxygen minimum zone. *Optics Express*, *17*(24). https://doi.org/10.1364/oe.17.021992

Wilson, C., & Coles, V. J. (2005). Global climatological relationships between satellite biological and physical observations and upper ocean properties. *Journal of Geophysical Research*, *110*(10), 1–14. https://doi.org/10.1029/2004JC002724

Wojtasiewicz, B., Trull, T. W., Udaya Bhaskar, T. V. S., Gauns, M., Prakash, S., Ravichandran, M., et al. (2020). Autonomous profiling float observations reveal the dynamics of deep biomass distributions in the denitrifying oxygen minimum zone of the Arabian Sea. *Journal of Marine Systems*, *207*(March 2017), 103103. https://doi.org/10.1016/j.jmarsys.2018.07.002

Wong, A., Keeley, R., Carval, T., and the Argo Data Management Team (2020). Argo Quality Control Manual for CTD and Trajectory Data. http://dx.doi.org/10.13155/33951

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

## Reference From the Supporting Information

Menke, W., & Menke, J. (2016). *Environmental data analysis with MATLAB* (2nd ed.). Elsevier.