

## Phytoplankton size distributions in the western North Atlantic and their seasonal variability

Nils Haëntjens <sup>1,\*</sup> Emmanuel S. Boss <sup>1</sup> Jason R. Graff <sup>2</sup> Alison P. Chase <sup>3,1</sup> Lee Karp-Boss <sup>1</sup>

<sup>1</sup>School of Marine Sciences, University of Maine, Orono, Maine

<sup>2</sup>Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon

<sup>3</sup>Applied Physics Laboratory, University of Washington, Seattle, Washington

### Abstract

Phytoplankton play a major role on Earth, impacting the global distribution and cycles of carbon, oxygen, nitrogen, sulfur, and other elements, and structuring marine food webs. One fundamental trait of phytoplankton with direct biogeochemical implications is their size, as it governs metabolic and sinking rates as well as prey–predator interactions. Phytoplankton size spans approximately 3.5 orders of magnitude (when expressed as an equivalent spherical diameter), and thus measuring the full range in size distribution of phytoplankton is challenging and rarely attempted. Here, we constructed phytoplankton size spectra by merging state-of-the-art cytometry and imaging cytometry measurements that were collected in the western North Atlantic Ocean, along a latitudinal gradient (36°N to 55°N) and during different phases of the annual cycle of phytoplankton. The derived spectra show a seasonal pattern that parallels changes in phytoplankton biomass, and do not always follow a commonly assumed power-law model. Shifts in size spectra were more pronounced in the sub-Arctic and temperate subregions, compared to the subtropical region of the study area. We evaluated the relationships between different size groups and environmental parameters to derive ecologically meaningful size groups. Finally, to simulate Ocean Color remote-sensing algorithms of phytoplankton size, we compared temporal variations in descriptors of the size spectra (median particle size, phytoplankton size distribution exponent) with optical size proxies derived from light absorption and attenuation; good agreement was observed in the northern sections of the study area where temporal changes in community size structure were more pronounced.

Functional traits provide an ecological framework for understanding and modeling phytoplankton communities and their distributions and responses to changes in the environment (Litchman and Klausmeier 2008; Barton et al. 2013; Weithoff and Beisner 2019). Among all traits, cell size is considered a primary trait because of its effects on a range of processes such as metabolic and sinking rates, prey–predator

interactions, and interactions with ambient flows (Kleiber 1932; Chisholm 1992; Karp-Boss et al. 1996; Brown et al. 2004). The distribution of phytoplankton biomass across a wide range of cell sizes is a fundamental property of marine ecosystems, with ecological and biogeochemical implications. For example, in most communities where small cells dominate, the microbial loop governs trophic pathways, while systems dominated by larger phytoplankton support a more “classic food chain” (i.e., direct energy transfer to mesozooplankton, Pomeroy 1974; Richardson 2019). Observed variations in the spatial distribution of phytoplankton size classes (PSCs) have been linked to spatial variations in resource supply (Huete-Ortega et al. 2011; Acevedo-Trejos et al. 2013; Marañón 2015). Temporal changes are also expected to be governed by similar processes, but few studies have examined temporal variations in phytoplankton size distribution (PSD). Using a trait based-model, Acevedo-Trejos et al. (2015) predicted that in the temperate ocean mean cell size will follow a seasonal pattern with a maximum in mean cell size in the spring (~ 25  $\mu\text{m}$ ) and a minimum (~ 9  $\mu\text{m}$ ) in the fall and winter. The model further predicts invariable mean cell size (~ 3.7  $\mu\text{m}$ ) throughout the year in tropical regions. Indirect

\*Correspondence: [nils.haentjens@maine.edu](mailto:nils.haentjens@maine.edu)

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

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

**Author Contribution Statement:** All authors contributed to data collection and processing on NAAMES campaigns. JRG: analyzed ICS data and provided critical observations regarding the merged PSD; AC and NH: analyzed IFCB data; NH: built the merged PSD, computed statistics (PCA, sPLS); LK-B and EB: supervised the research and provided critical knowledge to both analyze the data and frame the manuscript; NH: wrote the manuscript; All authors revised the manuscript.

observations from the North Atlantic Ocean, inferred from taxonomic composition (Benedetti et al. 2019) or remote sensing data (Kostadinov et al. 2010, 2017) are consistent with the model's predictions of seasonality in community size structure in the temperate ocean; however, direct size distribution measurements are still lacking.

PSDs are challenging to measure. Given the wide range of phytoplankton cell diameters, from the smallest cyanobacteria (with an equivalent spherical diameter [ESD] of  $\sim 0.5 \mu\text{m}$ ) to the largest dinoflagellates (ESD  $\sim 2000 \mu\text{m}$ , *Noctiluca*), a combination of instruments is needed to fully characterize PSDs of phytoplankton communities (Lombard et al. 2019). Each instrument has a defined cell or particle size detection range and measurement accuracy decreases at the lower and upper limits of that range. Another challenge in measuring PSDs resides in the ability to distinguish between phytoplankton and other cells and particles of the same size range (i.e., non-algal particles). Coulter counters and optically based particle sizers (e.g., LISST, Sequoia Scientific), for example, have been used to study size distributions of particles in the ocean. Particulate volume observed by these instruments is nearly uniformly distributed across log-spaced size bins according to a power law with an exponent ranging in value from three to four (Sheldon et al. 1972), but these measurements encompass all suspended particles and thus are not specific to phytoplankton. Early studies of phytoplankton size structure were based on chlorophyll *a* (Chl *a*) size fractionation techniques (Saijo 1964; Mullin 1965;

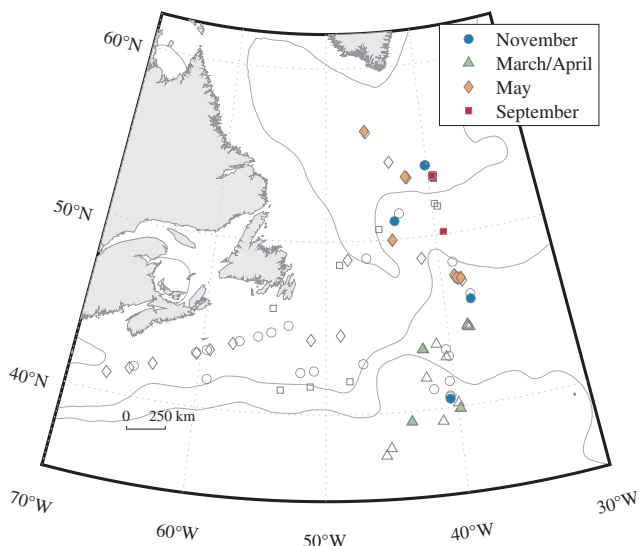
Sieburth et al. 1978), leading to the commonly used PSCs: picophytoplankton ( $0.2\text{--}2 \mu\text{m}$ ), nanophytoplankton ( $2\text{--}20 \mu\text{m}$ ), and microphytoplankton ( $20\text{--}200 \mu\text{m}$ ). These size classes are based on an operational definition rather than an ecological significance and Chl *a* size fractionation can suffer biases associated with the so-called filter effect (Runyan et al. 2020) and variability in the relation between Chl *a* and biomass (Sathyendranath et al. 2009). Analysis of pigments other than Chl *a* is also commonly used to retrieve major size classes based on empirical relationships among pigment composition, phytoplankton taxa, and their associated size (Mackey et al. 1996; Vidussi et al. 2001; Uitz et al. 2006). This approach has been widely used to obtain a global view of PSC through the development of remote sensing algorithms (Kostadinov et al. 2017; Mouw et al. 2017), but requires further validation as some key taxa exhibit a wide range of cell size (e.g., diatoms and dinoflagellates have representatives in both nano- and micro-size classes; Uitz et al. 2006; Kostadinov et al. 2010; Chase et al. 2020). Direct measurements of cell size have been done traditionally by microscopy for cells  $> 5 \mu\text{m}$  (Huete-Ortega et al. 2011; Marañón et al. 2012), but they are time consuming and therefore limited in temporal and spatial coverage. Traditional and imaging cytometers allow for automated and rapid sizing of phytoplankton cells, and their combined use shows great promise for advancing ecological knowledge on size structure of natural phytoplankton populations (Laney and Sosik 2014; Lombard et al. 2019) and for assessing other methods such as pigment-based size algorithms (Chase et al. 2020).

Here, we use cytometry, along with bio-optical proxies, to study variations in phytoplankton community size structure during different phases of the annual cycle of phytoplankton in the western North Atlantic Ocean. Temporal and spatial variations in PSD are examined in relation to variations in physical (e.g., mixed layer depth [MLD], temperature) and chemical (e.g., nutrients) descriptors of the environment, and the ecological merit of the traditional operational definition of PSCs (pico, nano, micro) is assessed in light of the finer resolution of the PSD available from the combined cytometry measurements.

## Materials and methods

### Data collection

Data presented in this study were collected on board the R/V Atlantis, as part of the North Atlantic Aerosol and Marine Ecosystems Study (NAAMES, Behrenfeld et al. 2019). The area occupied during the NAAMES campaigns encompassed four oceanographic subregions that were defined based on a smoothed mean dynamic topography analysis (Fig. 1; Della Penna and Gaube 2019) and correspond to biogeochemical provinces proposed by Longhurst (2007). Sampling took place in November 2015, May 2016, September 2017, and March–April 2018, representing different phases in the annual cycle of phytoplankton, namely the winter transition, climax transition, declining phase, and accumulation phase, respectively (Behrenfeld et al. 2019).



**Fig. 1.** Location of the NAAMES cytometric samples in the western North Atlantic. Colored symbols refer to stations in which both cytometry and contextual environmental data were collected. Gray markers denote PSDs with limited contextual data (e.g., no MLD, no nutrients). The contour lines delineate the different subregions averaged over all seasons. From south to north they are the Gulf Stream and Sargasso Sea province, the North Atlantic Subtropical Gyre province, the North Atlantic Temperate province, and the Atlantic sub-Arctic province. Note that subregions vary slightly from a season to the next (Della Penna and Gaube 2019).

Phytoplankton cells were sized using an Imaging FlowCytobot (McLane Research Laboratories, Inc), hereinafter referred to as cytobot, and an Influx Cell Sorter flow cytometer (ICS; Becton Dickinson), hereinafter flow cytometer. In this study, cell size is defined based on cell volume and expressed as the diameter of a sphere of equivalent volume, referred to as ESD. The cytobot was connected to a continuous flow-through system, with water pumped into the vessel from a depth of ~ 5 m, using a diaphragm pump (Cetinić et al. 2016; Boss et al. 2019) and a sample was automatically drawn from the main flow every ~ 23 min. The average volume analyzed per sample was 4.8 mL. In addition to the cytobot, the flow-through system was equipped with a thermosalinograph (Sea-Bird Scientific) and an ACS hyperspectral spectrophotometer (Sea-Bird Scientific), which provided additional size proxies as well as contextual environmental information. Discrete water samples for flow cytometry were collected regularly from the flow-through system and analyzed within a few minutes of collection; the average volume analyzed per sample was 1.2 mL. Samples were pre-filtered through a 64  $\mu\text{m}$  mesh, setting the upper size limit of the flow cytometer measurements. In addition to sample collection from the inline system, CTD Rosette casts were deployed on stations and surface samples (< 10 m) were taken for analysis by both the cytobot and the flow cytometer.

The cytobot best targets cells in the nano- and micro-size range while the flow cytometer targets cells in the pico-nano size range. Samples from the two instruments were merged to obtain a full size spectra. To increase statistical power for larger particles, cytobot samples collected within a 3-h window of a flow cytometer sample were combined. The average volume analyzed by the cytobot per sample group was  $29 \pm 9.7$  mL. We assumed that the samples were collected from the same water mass if no density or Chl *a* fronts were detected by the thermosalinograph and ACS. A front was defined as a region where Chl *a* changed by a factor greater than two within a 30-min interval (Chl *a* was estimated from ACS measurement, see below). If a front was detected over these 3-h windows, only samples on the same side of the front as the flow cytometer sample collection were used. This procedure resulted in 97 matching samples encompassing a wide range of physical-chemical conditions throughout the annual cycle and across the four provinces of the western North Atlantic.

Contextual environmental data from the NAAMES campaigns used here were obtained as follows: the MLD was computed for every CTD cast using the buoyancy frequency ( $N^2$ , Brunt-Väisälä frequency) according to Mojica and Gaube (unpublished) for consistency with another NAAMES manuscript (Baetge et al. 2020). Mean Brunt-Väisälä frequency ( $N^2$ ) between the surface and 100 m was also used as an indicator of the stratification of the upper water column. Inorganic nutrient concentrations (phosphate, nitrite + nitrate, and silicate) were measured from samples collected with Niskin bottles mounted on the CTD Rosette (Baetge et al. 2020). Chl *a* concentrations were derived from ACS measurements, using

the absorption line-height method (Boss et al. 2013) which is based on an empirical relationship between absorption line-heights at 676 nm ( $a_{\phi}(676)$ ) and Chl *a* concentrations ([Chl *a*]). The empirical relationship was tuned to the NAAMES data set using [Chl *a*] obtained from high-performance liquid chromatography (HPLC) pigment samples collected from the inline system or with Niskin bottles. HPLC samples analyzed at the Ocean Ecology Laboratory at NASA, following published protocols (Van Heukelem and Thomas 2001; Hooker et al. 2009). A type II least-squares regression was used to derive the following empirical relationship  $[\text{chl } a] = 95 \times a_{\phi}(676)^{1.06}$ , which was then applied to the entire ACS data set. Daily photosynthetically available solar irradiance ( $\text{PAR}_{\text{daily}}$ ) in the mixed layer ( $< I_{\text{ML}} >$ ) was estimated from the depth-averaged PAR from the surface to the MLD, using the exponential attenuation of PAR (Eq. 1).

$$< I_{\text{ML}} > = \frac{\text{PAR}_{\text{daily}}(z=0^-)}{K_{\text{dPAR}} \times \text{MLD}} (1 - e^{-K_{\text{dPAR}} \times \text{MLD}}). \quad (1)$$

The diffuse attenuation coefficient ( $K_{\text{dPAR}}$ ) was computed from HPLC [Chl *a*] following Xing and Boss (2021). More accurate models based on IOPs to compute  $K_{\text{dPAR}}$  (Lee et al. 2005) could not be used as dissolved absorption was not measured during all campaigns. Daily photosynthetically available radiation ( $\text{PAR}_{\text{daily}}$ ) above sea surface ( $z=0^+$ ) was integrated for each day from continuous PAR measurement from a sensor mounted at the top of the ship.  $\text{PAR}_{\text{daily}}(z=0^+)$  was then converted to  $\text{PAR}_{\text{daily}}$  just below the ocean's surface ( $z=0^-$ ) following Mobley and Boss (2012).

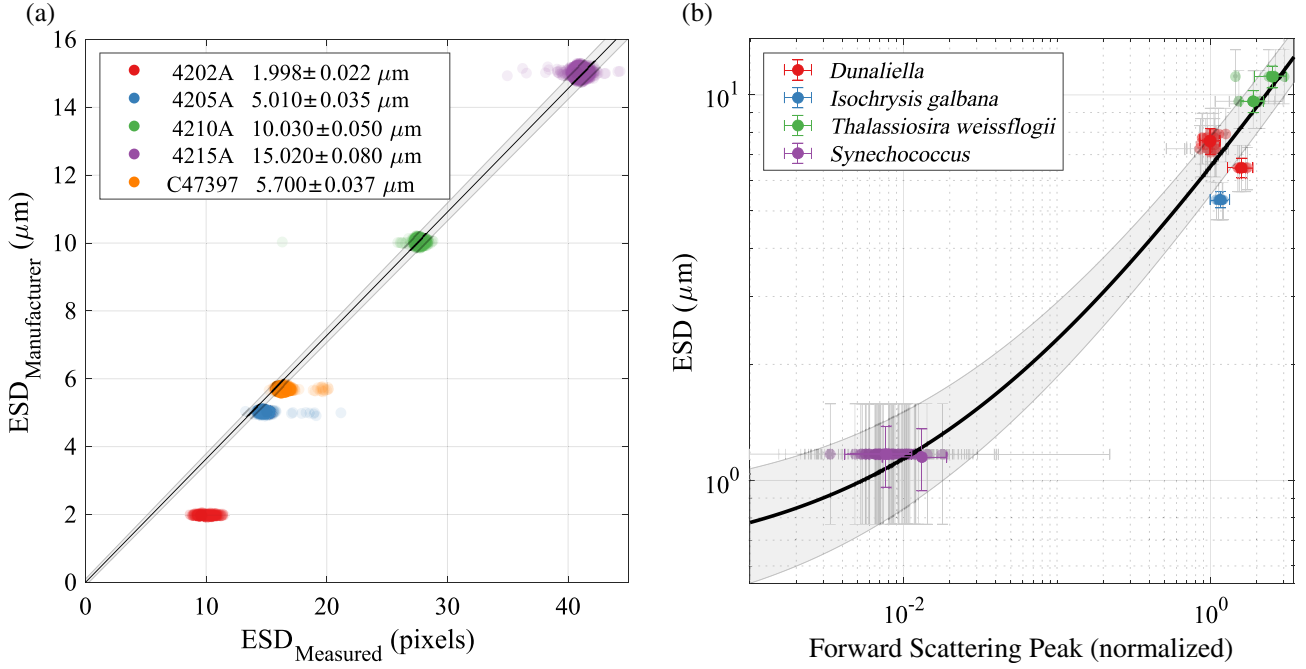
## Cytometers calibration

### Imaging FlowCytobot

The cytobot quantitatively imaged red fluorescing (ex/em 635/680 nm) particles—mainly chlorophyll containing particles—between ~ 3 and 150  $\mu\text{m}$ . All images were classified with EcoTaxa (Picheral et al. 2017) into taxonomic (or functional) categories as described in Chase et al. (2020). Nonphytoplankton and nonliving particles were removed from the analysis. For each cell, biovolume ( $V$ ), ESD, average cross-sectional area ( $\bar{G}$ ), and additional morphological features were computed in pixels following Moberg and Sosik (2012). The particles' nonrandom orientation in the cytobot was considered for computing the average cross-sectional area (Supporting Information Section S3). To convert the measurements to  $\mu\text{m}$ , we determined the instrument's resolution,  $2.71 \pm 0.04$  pixels  $\mu\text{m}^{-1}$ , using precision beads (Fig. 2a). Details about the size range imaged, the image curation, the volume sampled, and size uncertainties are provided in the Supporting Information Section S1.

### Influx cell sorter flow cytometer

The flow cytometer was set up during NAAMES campaigns to quantitatively measure forward and side scattering and fluorescence at 530 and 692 nm (ex 488 nm) of particles between



**Fig. 2.** (a) Calibration of the imaging cytometer. The ESD of beads specified by the manufacturer is plotted as a function of the diameter of beads measured by the cytometer. The beads used for the calibration were Duke Scientific 4202A (red), 4205A (blue), 4210A (green), 4215A (purple), and Life Technologies C47397 (orange) with average ESDs shown in the legend. The resultant regression equation (forced to zero) is  $\text{ESD}_{\mu\text{m}} = 0.364 (\pm 0.006) \text{ESD}_{\text{pixels}}$ . To derive uncertainties (shaded area), “noise” was added to the manufacturer’s size using a random number generator with a normal distribution set with the mean and the standard deviation specified by the bead manufacturer. (b) Calibration of the flow cytometer showing ESDs plotted as a function of the normalized forward scattering peak (FSC). *Dunaliella* (red), *Isochrysis galbana* (blue), and *Thalassiosira weissflogii* (green) cultures were sized with the imaging cytometer while the size of *Synechococcus* is based on a literature review (Supporting Information Table S1). *Synechococcus* cells were gated, looking at property–property plots of fluorescence peak at 530 and 692 nm, from surface samples of the four NAAMES campaigns. Data are fitted to a 1<sup>st</sup>-order polynomial by a weighted linear least-squares type II (MATLAB function `lsqcurvefit` from E. Peltzer, based on York 1966) after a Box–Cox transformation (Box and Cox 1964). The equation of the fit is  $\text{ESD}_T = 2.06 (\pm 0.15) \times \text{FSC}_T + 4.12 (\pm 0.26)$  with Box–Cox exponents  $\lambda_{\text{FSC}} = 0.453$  and  $\lambda_{\text{ESD}} = 0.754$ . Uncertainties in the conversion from FSC counts to ESD are represented by the shaded area and are estimated to be  $\pm 0.3 \mu\text{m}$  at  $1 \mu\text{m}$ ,  $\pm 0.5 \mu\text{m}$  at  $2 \mu\text{m}$ ,  $\pm 0.8 \mu\text{m}$  at  $4 \mu\text{m}$ , and  $\pm 1.5 \mu\text{m}$  at  $8 \mu\text{m}$ . Error bars correspond to the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the measured values on both x-axis and y-axis.

$\sim 0.6$  and  $\sim 10 \mu\text{m}$ . These measurements were used to gate particles into four groups: *Prochlorococcus*, *Synechococcus*, pico-eukaryote, and nano-eukaryote.

Forward scatter was used to size cells, as size has the most influence on the forward scattering signal (Jerlov 1976; Gin et al. 1999). However, one needs to realize that estimated sizes may be biased if cells have a significantly different shape or index of refraction (e.g., *Coccolithophores*) such that they scatter differently than the cells used for size calibration. To take into account changes in instrument gains and slight variations between runs, forward scattering was first normalized with reference beads and *Dunaliella* culture (Gin et al. 1999; Green et al. 2003). The calibration of the normalized forward scattering to ESD in  $\mu\text{m}$  was achieved by running a series of cultures in the flow cytometer (Fig. 2b). Cultures were also sized with the imaging cytometer, except for the *Synechococcus* for which the value was estimated from the literature (Supporting Information Table S1). Further details are given in the Supporting Information Section S2.

## Phytoplankton size distribution

### Definition

To derive PSD for the whole population of cells, we first derived PSDs from each cytometer and then merged the two distributions as described in the following sections. PSD is defined as the number of phytoplankton cells per size interval per volume sampled (also referred to as a size frequency or size spectrum; Jonasz and Fournier 2007):

$$n(D) = \frac{dN}{dD}, \quad (2)$$

where  $dN$  is the concentration of particles within the size interval  $[D, D+dD]$ . The units of  $n(D)$  are  $\# \text{ mL}^{-1} \mu\text{m}^{-1}$ . To obtain the abundance  $N$  in  $\# \text{ mL}^{-1}$ , Eq. 2 needs to be integrated across the size range of interest. Similarly, we obtained PSDs based on volume distribution ( $\mu\text{m}^3 \text{ mL}^{-1} \mu\text{m}^{-1}$ ) and cross-sectional area distribution ( $\mu\text{m}^2 \text{ mL}^{-1} \mu\text{m}^{-1}$ ) as follows:



$$v(D) = \frac{dV}{dD}, \quad (3)$$

$$g(D) = \frac{dG}{dD}, \quad (4)$$

with  $dV$  the average volume concentration and  $dG$  the average cross-sectional area concentration of particles within the size interval  $dD$ .

### Discretizing the PSD

**Boundaries.** For each cytometer, we computed PSDs using a standard binning algorithm (Sturges 1926) and determined the size thresholds for each instrument. The binning algorithm optimizes the bin width based on the abundance of cells of each size, providing the same bin width for each size bin.

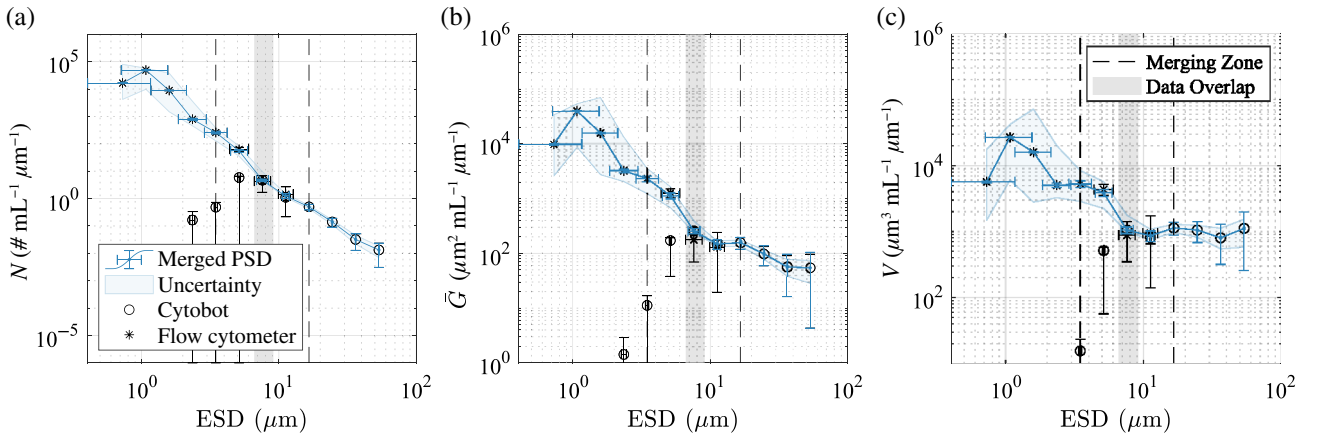
We defined the upper range of the flow cytometer size measurement to be the largest size bin that contains at least 100 cells. The lower range of the cytobot size measurement was defined as the size bin with the highest cell abundance among the  $< 10 \mu\text{m}$  size bins (Fig. 3). These size thresholds did not vary significantly within each campaign but did vary between the different campaigns due to differences in instrument settings and sensitivities. Therefore, a set of mean size thresholds was computed for each campaign (Supporting Information Table S2).

We set the lower boundary of the spectra ( $b_0$ ) to be the operational lower size measured by the flow cytometer ( $b_0 = 0.6 \mu\text{m}$  in ESD) and the upper boundary ( $b_n$ ) as the largest cytobot size bin across all NAAMES samples that contained

at least 10 particles ( $b_n = 65 \mu\text{m}$  in ESD). This choice of upper boundary led to the exclusion of cells greater than  $b_n$  (cells over the entire data set which account for 0.01% of the total number of cells).

**Bin resolution.** Traditional statistical binning methods, such as those of Sturges (1926), Scott (1979), and Freedman and Diaconis (1981), produce on average 17, 67, and 89 bins for the NAAMES data set, respectively. Uncertainties associated with these numbers of bin are still high (median  $\sigma$  for the last three bins is greater than 160%) and more than 40% of the bins have a width finer than the uncertainty in size. In this study, we aimed to achieve the highest PSD bin resolution (i.e., number of bins) while keeping the counting uncertainties low and limiting the number of bins with a width smaller than the size uncertainty (Supporting Information Fig. S9), resulting in a total of 12 bins for the NAAMES data set. The number of bins is lower than marine PSDs derived from other instruments (e.g., LISST PSDs are often reported binned to 32 bins). Note that the setting of bin size is fixed in commercial particle sizers (e.g., LISST, Coulter Counter) and one should consider whether to merge bins or samples for a given situation based on uncertainties as discussed here.

**Bin distribution.** To account for the fact that concentrations of particles in the ocean decrease with increasing cell size nearly following a power-law function (Sheldon et al. 1972),



**Fig. 3.** Examples of PSD spectra normalized by bin width and expressed in (a) abundance, (b) average cross-sectional area, and (c) biovolume as a function of the ESD. The PSD from the flow cytometer (black asterisk) and cytobot (open circles) are merged into one PSD (black line). The uncertainties on the x-axis correspond to the instrument uncertainty in size measurement, as defined in Table 1. The error bars on the y-axis represent the variability associated with counting statistics and the volume sampled uncertainties ( $\sqrt{\text{counting}_{\text{uncertainty}}^2 + \text{volume\_sampled}_{\text{uncertainty}}^2}$ ). Shaded blue areas correspond to the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the uncertainty propagation, using a Monte Carlo analysis (see “Inherent Uncertainties” section for details). The dashed vertical lines correspond to the merging area where both instruments contribute with variable weighting. The vertical gray areas correspond to the overlapping section between the two instruments defined (Supporting Information Table S2). All merged PSD from the four NAAMES campaigns are presented (Supporting Information Fig. S8).

size bins are typically distributed according to a geometric progression. Edges of size bins ( $b_i$ ) are defined by:

$$q = \left( \frac{b_n}{b_0} \right)^{1/n}, \quad (5)$$

$$b_i = b_0 q^i, \quad i = 0, 1, \dots, n, \quad (6)$$

with  $n$  being the number of bins and  $b_0$  and  $b_n$  being the lower and upper boundaries of the size distribution, respectively. The reported size for each bin is the geometric mean ( $\sqrt{b_i \times b_{i+1}}$ ) of the edges of the bin. The resultant size distribution bins are presented in Supporting Information Table S3.

### Merge size distributions

To merge the PSDs of the cytobot and flow cytometer, for each paired sample, we identified the overlapping area of the two PSDs ( $b \in [b_l, b_h]$ ) and calculated in that area a weighted mean of the flow cytometer and cytobot abundance ( $n(D)$ ), cross-sectional area ( $g(D)$ ), and volume ( $v(D)$ ) PSD.

$$\begin{cases} n(b) = w_{\text{flow\_cytometer}} n_{\text{flow\_cytometer}}(b) + w_{\text{cytobot}} n_{\text{cytobot}}(b) \\ w_{\text{flow\_cytometer}}(b) = \frac{b_h - b}{b_h - b_l}, \quad w_{\text{cytobot}}(b) = \frac{b - b_l}{b_h - b_l} \end{cases} \quad (7)$$

where  $b_l$  and  $b_h$  are two bin sizes smaller than the lower cytobot size bin ( $b_l$ ) and two bin sizes larger than the higher flow cytometer size bin ( $b_h$ ), respectively. For  $b < b_l$ , the flow cytometer PSD was exclusively used; likewise, for  $b_h < b$ , the cytobot PSD was exclusively used. In two samples of the first NAAMES campaign, the largest bin observed by the flow cytometer was lower than  $b_l$ . In these cases, we interpolated between the two last bins of the flow cytometer and the two 1<sup>st</sup> bins of the cytobot by fitting a robust linear regression.

### Inherent uncertainties

We propagated the uncertainties in the merged PSD using a Monte Carlo analysis. Input included uncertainties in cell concentration (Poisson's probability distribution function [PDF]),

instruments' size calibration ( $\pm 0.3$  to  $\pm 1.5 \mu\text{m}$  for the flow cytometer and  $\pm 0.1 \mu\text{m}$  for the cytobot), and the volume sampled by both cytometers (Table 1). All input parameters were varied independently, assuming that uncertainties are not correlated, and assuming a normal distribution with a variance defined by the standard deviation of each uncertainty. We ran the Monte Carlo analysis for each PSD with  $10^4$  realizations (doubling the number of realizations did not significantly change the output). Results of the simulations provided the upper and lower bounds of uncertainty for each PSD, which were defined as the 5<sup>th</sup> and 95<sup>th</sup> percentiles of all realizations (Fig. 3). Uncertainties from each instrument were weighted according to Eq. 7 in the merging zone of the PSD.

### Derived size indices

We calculated the median particle size,  $D50_V$  (the ESD at which 50% of the cumulative biovolume is reached), and the exponent of the power-law fit to the PSD ( $m$ ) (hereafter referred to as “slope”) as two descriptors of the PSD. A power law was fit to the PSD (Eq. 8) by transforming the original data into log space, thus transforming Eq. 8 into a simple linear equation (Eq. 9), hence the designation of “slope” for  $m$ .

$$n(D) = n(D_0) D^m, \quad (8)$$

$$\log \frac{n(D)}{n_0} = \log \frac{n(D_0)}{n_0} + m \log \frac{D}{D_0}, \quad (9)$$

where  $n_0 = 1 \text{ \# mL}^{-1} \mu\text{m}^{-1}$  and  $D_0 = 1 \mu\text{m}$ . Equation 9 was fitted to the data using a weighted robust type II least-squares bisector regression (MATLAB weighted robust fit function combined with MBARI *lsqbisec* function, <https://www.mbari.org/index-of-downloadable-files/>). The weights used for the fitting were log-transformed and were inversely proportional to the counting and volume sampled uncertainties of  $n(D)$  and the instruments' size calibration uncertainties for  $D$ . Unlike the power-law slope,  $D50_V$  is independent of the specific parametrization of the PSD and therefore serves as a more robust descriptor of the PSD (Runyan et al. 2020).

**Table 1.** Parameters used in the Monte Carlo uncertainties analysis and estimated uncertainties. The cytometers size calibration uncertainties are captured from the calibration equations. The counting uncertainties are considered based on the number of cells counted by the instruments in each size bin using Poisson's PDF (Supporting Information Section S4). “fsc” stands for forward scattering and correspond to the direct measurement made by the flow cytometer to retrieve size.

Parameter	Mean	Uncertainty	Reference
Flow cytometer calibration slope	1.8918 $t_{\text{fsc}}/t_{\mu\text{m}}^*$	0.2341 $t_{\text{fsc}}/t_{\mu\text{m}}^*$	Fig. 2b
Flow cytometer calibration intercept	3.2169 $t_{\mu\text{m}}^*$	0.6551 $t_{\mu\text{m}}^*$	Fig. 2b
Cytobot resolution	2.7488 pixels $\mu\text{m}^{-1}$	0.1 pixel $\mu\text{m}^{-1}$	Supporting Information Fig. S1
Flow cytometer volume sampled	Sample dependent	0.1%	Estimated
Cytobot volume sampled	Sample dependent	1%	Syringe manufacturer
Cell counts/size bin	—	Poisson's PDF	Supporting Information Fig. S7

\*Box-Cox transformed data, as indicated by the “t” preceding the units.

### Independent size indices

In addition to the cytometric observations, we used the underway measurements of inherent optical properties to compute phytoplankton size indices with published algorithms based on hyperspectral particulate attenuation ( $c_p(\lambda)$ ) and absorption ( $a_p(\lambda)$ ). The optical sensors were set-up and calibrated following the recommendations of Boss et al. (2019) and the calibration-independent technique for particulate absorption and attenuation coefficients of Dall’Olmo et al. (2009) and Slade et al. (2010).

The 1<sup>st</sup> size index is based on the spectral slope of particulate beam attenuation ( $\gamma_{cp}$ ) which has been shown to covary with the mean size of suspended particles (Boss et al. 2001). While this size index is not specific to phytoplankton, in the open ocean (far from terrestrial influences), it is expected to co-vary with changes in phytoplankton size (Smith and Baker 1978a,b). A 2<sup>nd</sup> optical size proxy ( $S_f$ ) is based on the relative contribution of picophytoplankton and microphytoplankton to the total phytoplankton absorption spectrum ( $a_p(\lambda)$ ), as determined from changes in the ratio of blue and red wavelengths in  $a_p(\lambda)$  (Ciotti and Bricaud 2006). We computed  $a_p(\lambda)$  by subtracting non-algal particle absorption ( $a_{NAP}(\lambda)$ ) spectra from the particulate absorption spectra ( $a_p(\lambda)$ ) following Chase et al. (2013). As a 3<sup>rd</sup> size index, we consider [Chl *a*] since it has been shown that phytoplankton pigment mass, which is typically dominated by [Chl *a*], covaries with mean phytoplankton size (Hirata et al. 2008).

### Multivariate analysis

To identify relations between the phytoplankton size “classes” and environmental parameters, we performed a sparse partial least-square (sPLS) analysis, a linear, multivariate, regression-based visualization technique that is typically used to explain relationship between continuous data sets. The

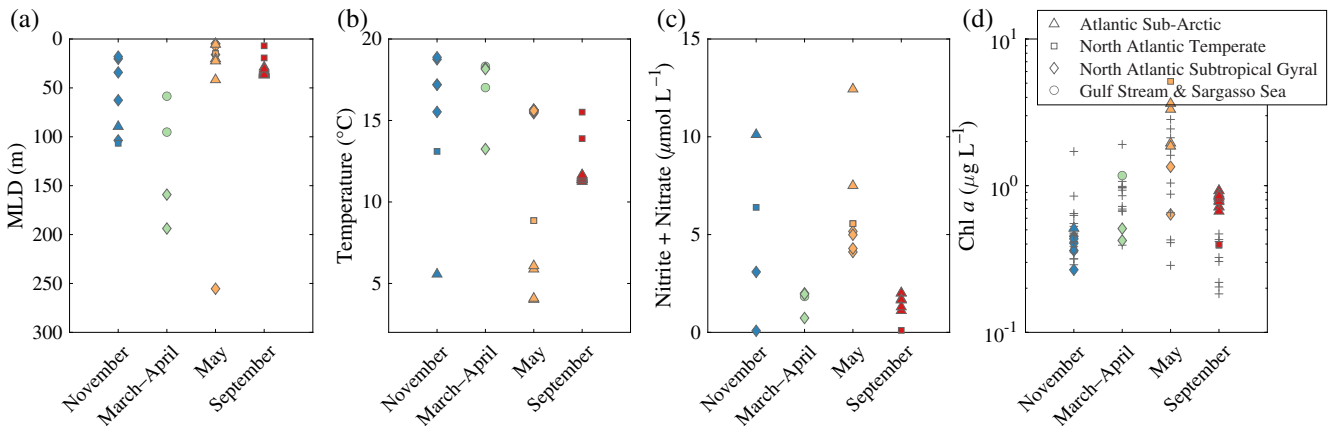
sPLS was computed with the R package mixOmics (Rohart et al. 2017) against the 12 size bins from the biovolume size distribution. The rows of the sPLS (corresponding to environmental parameters) were clustered using a complete linkage algorithm and Euclidean distance metric. Prior to computing the sPLS, the PSD was normalized by the total biovolume (NB: trends reported here were similar to those in the absence of normalization). We also conducted a principal component analysis, using the MATLAB function *pca* to examine how the PSD approximated by the single-value  $D50_v$  varies with respect to environmental parameters.

### Results and discussion

Along the  $42 \pm 2^\circ\text{W}$  meridional, the broad latitudinal range observed ( $36^\circ\text{N}$  to  $55^\circ\text{N}$ ) featured a strong seasonality expressed by the following factors: MLD ranging from 5 to 255 m, phytoplankton biomass represented by [Chl *a*], which varies by a factor of 100 ranging from 0.05 to  $5.14 \mu\text{g L}^{-1}$ , nitrite and nitrate surface concentrations ranging from 0.1 to  $10.1 \mu\text{mol N/L}$  (Fig. 4), and depth-integrated net primary production ranging from 78 to  $2296 \text{ mg C m}^{-2} \text{ d}^{-1}$  (Fox et al. 2020).  $\text{PAR}_{\text{daily}}$  over the mixed layer was highest in May and September and lowest in March–April. MLD and [Chl *a*] observations were consistent with a previous model and remote sensing reports (Behrenfeld 2010). Although inter-annual variability in the observed biogeochemical properties is expected (Benedetti et al. 2019), it is likely less significant than the general spatiotemporal pattern described.

### Temporal and spatial variability in phytoplankton size structure

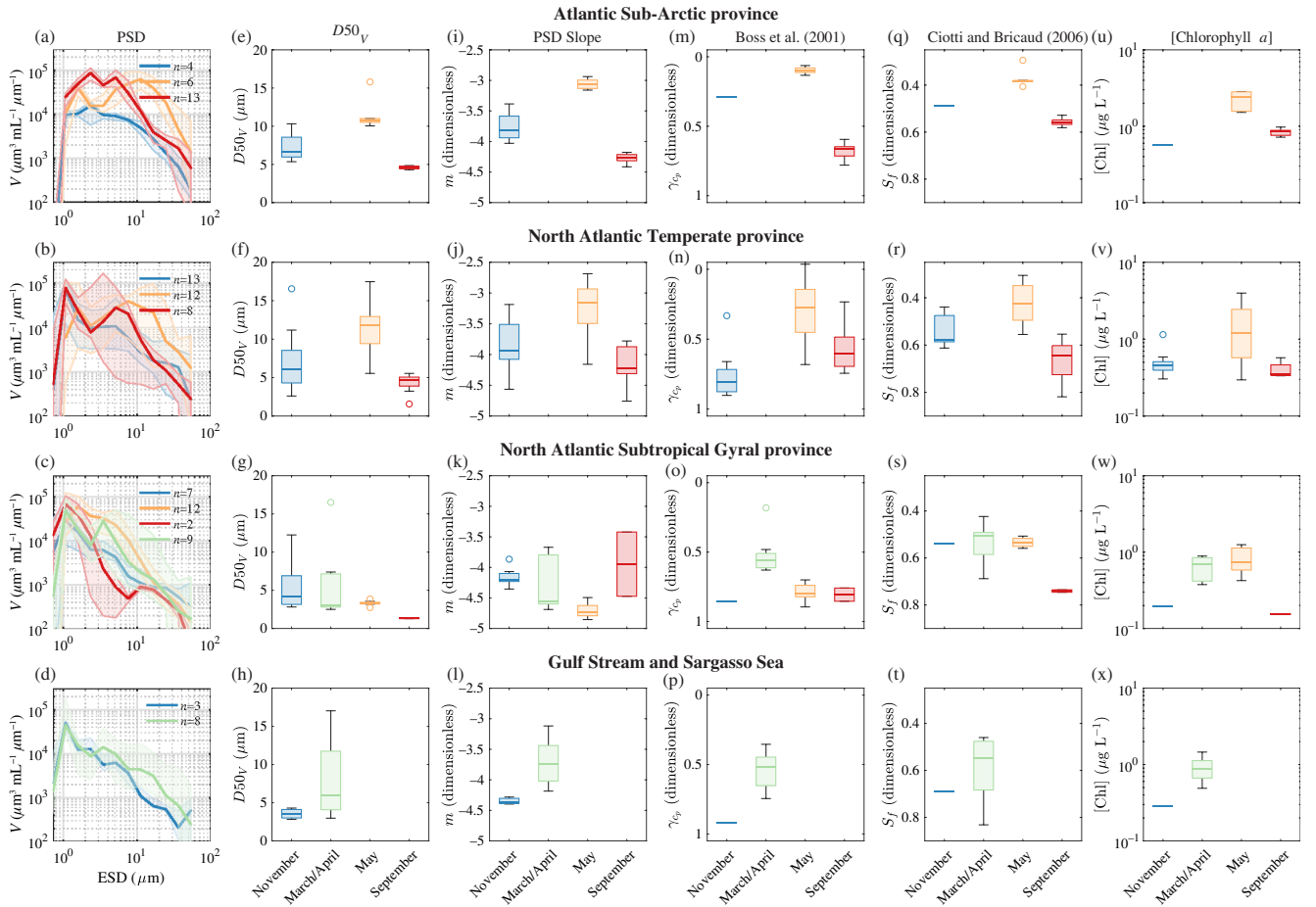
Seasonal variations in the size structure of phytoplankton were detected across the study area (Fig. 4). While the overall seasonal



**Fig. 4.** Contextual information for size distribution samples collected during the NAAMES campaigns. At every station visited, the MLD (a), surface temperature (b), surface nitrite and nitrate concentration (c), and surface chlorophyll concentration (d) are presented in the context of the annual cycle of the North Atlantic bloom: winter transition (November), accumulation (March–April), climax (May), and decline (September). Note that the deep MLD in May was observed in the core of an anticyclone eddy. All nutrient (ammonium, silicate, phosphate, nitrite, and nitrate) concentrations measured are available in Supporting Information Table S3.

pattern was similar in all subregions of the study area, there were latitudinal differences in observed dominant size class(es). In *late fall* (November 2015), when the mixed layer typically begins to deepen and phytoplankton biomass was relatively low (Figs. 4, 5), phytoplankton communities were dominated in biovolume by small cells (Fig. 5a–d). This was particularly evident in the temperate, subtropical gyre, and southern North Atlantic provinces in which “pico” phytoplankton dominated biovolume (ESDs  $\sim 0.9\text{--}1.3\ \mu\text{m}$ ). In the northern province of the study area, “pico” and “nano” phytoplankton ( $\text{ESD} < 9\ \mu\text{m}$ ) dominated the community, with a peak biovolume centered on  $\text{ESD} \sim 3\ \mu\text{m}$  (Fig. 5a). Sampling in *early spring* (March–April 2018) aimed at

capturing the accumulation phase (Behrenfeld et al. 2019). Due to weather conditions, only the two southern provinces were sampled on that cruise. The mixed layer was still relatively deep, but [Chl *a*] was, on average, higher than that observed during the winter transition period (Fig. 4). Size spectra shifted, with increased contribution of cells in size bins from  $3\text{--}25\ \mu\text{m}$ , though communities in both subregions were still highly dominated by “pico” size cells ( $\text{ESD } 0.9\text{--}1.3\ \mu\text{m}$ ; Fig. 5c,d). During the May 2016 campaign (*climax transition phase*; Behrenfeld et al. 2019), MLDs were shallow at most stations and average [Chl *a*] was the highest of all sampling times (Fig. 4), with an increased contribution of large cells (Fig. 5). Phytoplankton size



**Fig. 5.** (a–d) PSD and optical size indices (m–t) arranged by region and season. The subregions are from north to south: (1<sup>st</sup> row) the Atlantic sub-Arctic province, (2<sup>nd</sup> row) the North Atlantic temperate province, (3<sup>rd</sup> row) the North Atlantic subtropical Gyral province, and (4<sup>th</sup> row) the Gulf stream and Sargasso Sea province. Seasons are color coded: November (blue), March–April (green), May (orange), and September (red). (a–d) The transparent area surrounding each line correspond to the minimum and maximum PSD of each group. (e–h) Size index  $D50_V$  estimated from the PSD. (i–l) PSD slope ( $m$ ). (m–p) Size index  $\gamma_{cp}$  from Boss et al. 2001, higher values correspond to a small average size (the y-axis is reversed). (q–t) Size index  $S_f$  from Ciotti and Bricaud 2006, higher values correspond to smaller average size (the y-axis is reversed). (u–x) Chl *a* concentration [Chl *a*] as an indicator of size. A combination of instrument malfunction and bubble contamination resulted in a reduced data set to compute optical size indices ( $\gamma_{cp}$ ,  $S_f$ , Chl *a*) for November. In subplots (e–x), dots represent outliers (as defined in function boxchart from MATLAB); whiskers extend from the minimum to the maximum value excluding outliers; 25<sup>th</sup> and 75<sup>th</sup> percentiles are at the lower and upper edges of the colored bar; and the line in the middle of the bar correspond to the median value. Note that the lowest size bin ( $0.6\text{--}0.88\ \mu\text{m}$ ) corresponds to the size of the smallest phytoplankton: *Prochlorococcus*, as this taxon was not present in all samples the drop present in some of the PSD at the lowest size bin is real. However, the size uncertainties inherent to the empirical calibration and the variability in the index of refraction or shape of cells (see “Influx Cell Sorter Flow Cytometer” section) could contribute to a size bias at the lower end of the size distribution.



structure was similar in the two northern regions with cells  $\sim 5\text{--}30\ \mu\text{m}$  in ESD becoming dominant. In the North Atlantic subtropical gyre, the relative contribution of larger cells to phytoplankton biomass also increased, but the community was still dominated by cells  $< 10\ \mu\text{m}$ . The southern North Atlantic province was not sampled in May 2016. During the *decline phase* (September 2017), the relative contribution of cells ranging from  $\sim 5$  to  $30\ \mu\text{m}$  decreased in all regions. In the subtropical gyre, biovolume was centered on cells with ESDs  $\sim 0.9\text{--}1.3\ \mu\text{m}$ , similar to the peak biovolume observed during the winter transition phase (November 2015). The southern North Atlantic province was not sampled in September 2017.

The seasonal variation described here is based on samples collected in different years, but the pattern is consistent with the phenology of phytoplankton functional type (PFT) obtained from monthly climatology of remote sensing data. These models, however, do not estimate size per se but infer seasonality in size from changes in the fraction of microphytoplankton (or diatoms) to total phytoplankton, expressed in [Chl *a*], biovolume, or relative units (Kostadinov et al. 2017). An inter-comparison of the different size algorithms, however, revealed large qualitative and quantitative differences in relative contributions of microphytoplankton to total phytoplankton biomass. While validation of remote sensing data and algorithms is beyond the scope of this paper, we compared the median ESDs of the size spectra ( $D_{50V}$ ) and slopes of the size spectra to size proxies derived from bio-optical measurements (Fig. 5). Overall, the observed seasonality in size was captured by the different bio-optical proxies in all sub regions except for the North Atlantic subtropical gyral province. Poor agreement between PSD descriptors and size proxies in the latter province could be a result of the overall weaker seasonality in this province compared to the northern North Atlantic provinces, and therefore uncertainties and biases associated with the different size proxy measurements likely play a larger role.

In addition to temporal variations in PSD, our analysis also shows latitudinal differences. Most notably the strong negative correlation between the total biovolume of small cells ( $\sim 1\text{--}2\ \mu\text{m}$ ) and latitude ( $r > 0.69$ , Fig. 6), which is consistent with current understanding of phytoplankton ecology (Marañón 2015). The correlation becomes positive (albeit weak) for cells  $> \sim 7\ \mu\text{m}$  indicating greater contribution of larger cells at northern latitudes. These observed spatial and temporal trends in PSDs are consistent with the model predictions of Acevedo-Trejos et al. (2015), with one exception, which is that the median cell sizes are smaller than those predicted by the model (cf. Fig. 5; Supporting Information Table S4).

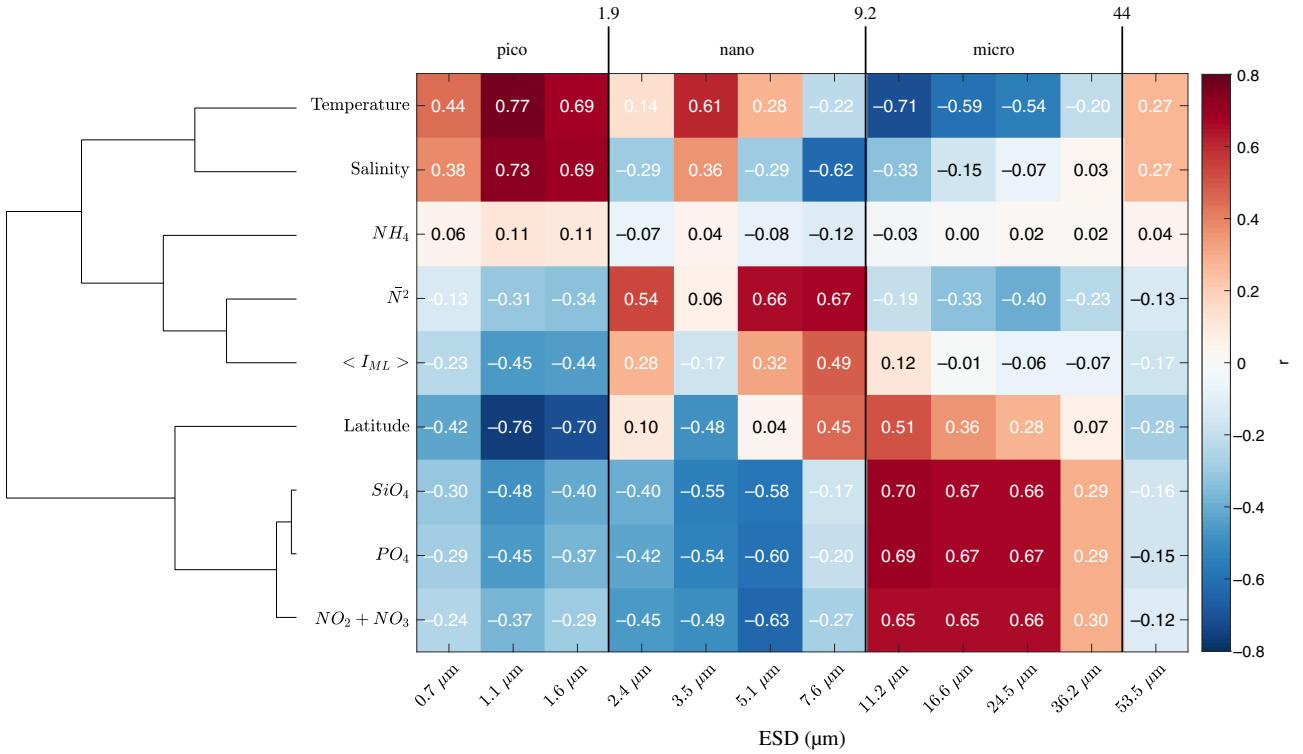
### Ecologically meaningful size groups

The size structure of the phytoplankton community in the surface of the ocean is the result of fundamental relationships among cell size, resource availability (nutrients and light), zooplankton grazing, and sinking (Litchman and Klausmeier 2008;

Acevedo-Trejos et al. 2015; Marañón 2015). Analyzing size bin variability with respect to environmental parameters distinguishes three ecologically distinct size clusters (Fig. 6; Supporting Information Fig. S7). The sPLS analysis reveals a sharp change in correlation of phytoplankton biovolume with temperature and stratification between  $1.6$  and  $2.4\ \mu\text{m}$  size bins. Similarly, a sharp change in correlation of phytoplankton biovolume with temperature, stratification, and nutrient concentrations between the size bins  $7.6$  and  $11.2\ \mu\text{m}$  is observed.

The 1<sup>st</sup> ecological size cluster consists of cells within  $0.88 < \text{ESD} < 1.9\ \mu\text{m}$  that were positively correlated with temperature and salinity and inversely correlated with latitude. The strong correlation with temperature, but weaker correlation with stratification suggests direct effects of temperature on the physiology of cells in this size cluster rather than indirect effects associated with water column stability (Li et al. 2009; Morán et al. 2010; Hilligsøe et al. 2011). On average, these cells show weaker correlation with nutrients, compared to cells in other size groups. The 2<sup>nd</sup> ecological size cluster includes cells within  $1.9 \leq \text{ESD} < 9.2\ \mu\text{m}$  that were positively correlated with stratification ( $\overline{N}^2$ ) and inversely correlated with nutrients. Given the strong correlation with stratification, we predict that this size group is most sensitive to variations in MLD and hence directly linked to bloom phenology (Marañón 2015; Behrenfeld and Boss 2018). The 3<sup>rd</sup> ecological size cluster encompasses cells within  $9 \leq \text{ESD} < 44\ \mu\text{m}$  that were positively correlated with nutrients and inversely correlated with temperature. The positive correlation of larger cells with surface nutrients (other group being inversely correlated or noncorrelated) is consistent with previous studies stating that larger cells predominate when nutrients are plentiful (Marañón 2015). The size bin centered at  $36.2\ \mu\text{m}$  exhibits lower correlations than the other bins of the 3<sup>rd</sup> cluster; this bin is likely a transition between the 3<sup>rd</sup> cluster and larger cells suggesting that the upper boundary of the 3<sup>rd</sup> cluster could be lower than  $44\ \mu\text{m}$ . The 1<sup>st</sup> and last bins of the PSD are less correlated with environmental parameters compared to other bins in their corresponding size groups which could result from relatively higher uncertainties in cell size ( $\pm 0.6\ \mu\text{m}$ ) and cell numbers ( $\sim 5$  cells sample<sup>-1</sup>) associated with the boundary bins  $< 0.88$  and  $> 44\ \mu\text{m}$ , respectively.

The sensitivity of each size class to distinct environmental parameters and interactions between them ultimately determine the shape of the PSD. It has been shown that particle size distributions in the ocean are well fit with a power-law model; that is, one would expect to see a unimodal distribution, typically with a slope of  $-4$  for  $n(d)$ . While such background slope is generally observed for particulate size distributions, especially for the Sargasso Sea, deviations have been reported at other locations such as the transition region between the Pacific and Arctic Oceans (Runyan et al. 2020) and in coastal waters (Reynolds et al. 2010). We observed a unimodal distribution in most



**Fig. 6.** Correlation (color) from sparse partial least squares analysis (sPLS, see “Methods” section) performed on environmental parameters and size classes of the PSD expressed in normalized biovolume. Correlation can be interpreted similarly to Person’s coefficients. Environmental variables are clustered with a complete Euclidean distance method; shorter branches indicate similarities between variables. Obtained pico-, nano-, micro-groups refer to the 1<sup>st</sup> size cluster ( $0.88 \leq \text{ESD} < 1.9 \mu\text{m}$ ), the 2<sup>nd</sup> size cluster ( $1.9 \leq \text{ESD} < 9.2 \mu\text{m}$ ), and the 3<sup>rd</sup> size cluster ( $9.2 \leq \text{ESD} < 44 \mu\text{m}$ ), respectively.

provinces over the period of observations (Fig. 5), with exceptions in May (the two northern North Atlantic provinces) and March (North Atlantic Subtropical Gyral province) where bi-modal size distributions were observed (Fig. 5). The deviation from the power-law model can be explained by the different sensitivity of each size group to environmental parameters (Acevedo-Trejos et al. 2015), favoring multiple taxa to bloom

simultaneously. Methods used in most previous studies (e.g., size fractionation or pigment analysis) are too coarse (resolving three size groups) to resolve such modulations in the shape of PSD.

Linking size ranges of the three ecological size clusters presented here with the commonly used definition of PSCs is a challenge. Size fractionation or pigment-based PSCs do not

**Table 2.** The 5<sup>th</sup> and 95<sup>th</sup> percentiles of the minor (ma) and major axis (MA) are reported for all cells measured by the cytotob within the size class expressed in ESD. The ma and MA are the length of the minor and major axis of the ellipse, respectively, that has the same normalized 2<sup>nd</sup> central moment as the blob region extracted from a picture of a cell. The total number of cells per size class is reported in the column *N*. Note that the cytotob is limited to observing cells  $> 6.5 \mu\text{m}$  (Supporting Information Table S2; Fig. 3), hence the minor axis of the size class  $[2, 9] \mu\text{m}$  is not taking into account smaller cells observed by the flow cytometer.

Size class ( $\mu\text{m}$ )	ESD ( $\mu\text{m}$ )	Minor axis ( $\mu\text{m}$ )	Major axis ( $\mu\text{m}$ )	<i>N</i>
ESD $\in [2, 9]$	2–9	3.7–8.6	6.1–19.8	537k
ESD $\in [9, 20]$	9–20	7.1–18.9	10.4–50.4	252k
ESD $\in [2, 20]$	2–20	4.1–15.3	6.4–30.7	789k
ESD $\in [20, 44]$	20–44	7.1–23.1	10.4–121	268k
ma $\in [2, 20]$	4.3–15.1	4.1–14	6.4–28.3	782k
ma $\in [20, 100]$	9.6–38.8	20.3–50	23.3–130	22k
MA $\in [2, 20]$	4.1–13.6	4.1–13.2	6.3–17.8	693k
MA $\in [20, 100]$	6.2–26	4.0–30.7	20.5–89.7	113k

have a clear size cutoff when compared to microscopy, cytometry, or resistive particle counter data. In an attempt to relate the ecological size clusters to the traditional size classes, we examined the ESDs for cells with minor axis length  $> 20 \mu\text{m}$  (i.e., cells that theoretically will be retained on  $20\text{-}\mu\text{m}$  filters) and cells with a major axis length  $< 20 \mu\text{m}$  (i.e., cells that theoretically will go through  $20\text{-}\mu\text{m}$  filters). Results suggests that a  $20\text{-}\mu\text{m}$  filter would retain cells with an  $\text{ESD} > 11.6 \pm 2 \mu\text{m}$  (Table 2) and therefore that the use of  $20\text{-}\mu\text{m}$  pore size filter has the potential to be consistent with the “ecological” size groups from this study. However, as pointed by the assessment of particle-size fractionation from Koestner et al. (2020), filtration efficiency for marine particle assemblage depends on many factors (e.g., filter type, particle shape) and an analysis must be performed before comparing results from studies that used different sizing methods, size metrics, or size cutoffs.

## Conclusion

We investigated spatial and temporal variations in phytoplankton community size structure within four subregions of the western North Atlantic, using direct and indirect measurements of PSDs. In the context of the annual cycle of phytoplankton, we show an increased contribution of large phytoplankton cells during the climax transition phase, with more pronounced shifts in size spectra in the sub-Arctic and temperate subregions. While seasonal variations in mean cell size or relative contributions of microphytoplankton to total biomass have been inferred from models (Acevedo-Trejos et al. 2015), and remote sensing observations (Kostadinov et al. 2017), respectively, to our knowledge, this study presents the 1<sup>st</sup> direct measurements of PSDs in relation to the annual cycle of phytoplankton in the North Atlantic Ocean. We further show that the temporal patterns in PSD are well captured by different optical size proxies when differences in size spectra are significant, but single descriptors ( $D_{50V}$ , PSD slope ( $m$ )) and proxies deviate when temporal differences in PSD are weaker (e.g., in the North Atlantic subtropical gyral province). Unlike single-size descriptors and proxies, the construction of a full PSD provides insights on relationships between different size groups and environmental parameters that can be useful for the parameterization of trait-based and size structure ecosystem models. Furthermore, deviations of the PSD from a power-law model suggest that multiple size groups can contribute equally to biomass at a given time.

The PSDs constructed here from measurements of individual cells enable us to define three size classes based on their relationships to key environmental factors: nutrient concentrations, temperature, and mixing. Future studies that measure phytoplankton ESDs (e.g., cytometry, LISST instruments, Coulter Counters) in other open ocean locations will reveal whether these relationships and size cutoffs in clustering are ubiquitous. Direct comparisons between size cutoffs for the ecological groups defined here and those traditionally used

with the filtration method are complex and require careful analysis, in part due to the nontrivial relation between size metrics (e.g., ESD and filter pore sizes).

Finally, remote sensing algorithms to derive PSCs or PFTs show promise in detecting changes in phytoplankton community structure on larger temporal and spatial scales than those covered by ships. The validation of these algorithms, and development of new ones, is currently limited by the lack of a global data set of direct in situ size measurements. The present study provides an example of how to construct phytoplankton PSDs from cytometry measurements, including estimates of uncertainties, as a first step toward establishing such a global database.

## References

- Acevedo-Trejos, E., G. Brandt, A. Merico, and S. L. Smith. 2013. Biogeographical patterns of phytoplankton community size structure in the oceans: Phytoplankton biogeographical patterns. *Glob. Ecol. Biogeogr.* **22**: 1060–1070. doi:10.1111/geb.12071
- Acevedo-Trejos, E., G. Brandt, J. Bruggeman, and A. Merico. 2015. Mechanisms shaping size structure and functional diversity of phytoplankton communities in the ocean. *Sci. Rep.* **5**: 8918. doi:10.1038/srep08918
- Baetge, N., J. R. Graff, M. J. Behrenfeld, and C. A. Carlson. 2020. Net community production, dissolved organic carbon accumulation, and vertical export in the western North Atlantic. *Front. Mar. Sci.* **7**: 227. doi:10.3389/fmars.2020.00227
- Barton, A. D., A. J. Pershing, E. Litchman, N. R. Record, K. F. Edwards, Z. V. Finkel, T. Kiørboe, and B. A. Ward. 2013. The biogeography of marine plankton traits. *Ecol. Lett.* **16**: 522–534. doi:10.1111/ele.12063
- Behrenfeld, M. J. 2010. Abandoning Sverdrup’s critical depth hypothesis on phytoplankton blooms. *Ecology* **91**: 977–989. doi:10.1890/09-1207.1
- Behrenfeld, M. J., and E. S. Boss. 2018. Student’s tutorial on bloom hypotheses in the context of phytoplankton annual cycles. *Global Change Biol.* **24**: 55–77. doi:10.1111/gcb.13858
- Behrenfeld, M. J., and others. 2019. The North Atlantic Aerosol and Marine Ecosystem Study (NAAMES): Science motive and mission overview. *Front. Mar. Sci.* **6**: 122. doi:10.3389/fmars.2019.00122
- Benedetti, F., and others. 2019. The seasonal and inter-annual fluctuations of plankton abundance and community structure in a North Atlantic marine protected area. *Front. Mar. Sci.* **6**: 214. doi:10.3389/fmars.2019.00214
- Boss, E., W. S. Pegau, W. D. Gardner, J. R. V. Zaneveld, A. H. Barnard, M. S. Twardowski, G. C. Chang, and T. D. Dickey. 2001. Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental

- shelf. *J. Geophys. Res.* **106**: 9509–9516. doi:[10.1029/2000JC900077](https://doi.org/10.1029/2000JC900077)
- Boss, E., and others. 2013. The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; contribution of the Tara Oceans expedition. *Methods Oceanogr.* **7**: 52–62. doi:[10.1016/j.mio.2013.11.002](https://doi.org/10.1016/j.mio.2013.11.002)
- Boss, E., and others. 2019. IOCCG ocean optics and biogeochemistry protocols for satellite ocean colour sensor validation inherent optical property measurements and protocols: Best practices for the collection and processing of ship-based underway flow-through optical data (v4. 0). In A. R. Neeley and A. Mannino [eds.], IOCCG protocol series, International Ocean Colour Coordinating Group (IOCCG) Dartmouth, NS, Canada.
- Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. *Ecology* **85**: 1771–1789. doi:[10.1890/03-9000](https://doi.org/10.1890/03-9000)
- Cetinić, I., N. Poulton, and W. H. Slade. 2016. Characterizing the phytoplankton soup: Pump and plumbing effects on the particle assemblage in underway optical seawater systems. *Opt. Express* **24**: 20703. doi:[10.1364/OE.24.020703](https://doi.org/10.1364/OE.24.020703)
- Chase, A., E. Boss, R. Zaneveld, A. Bricaud, H. Claustre, J. Ras, G. Dall’Olmo, and T. K. Westberry. 2013. Decomposition of in situ particulate absorption spectra. *Methods Oceanogr.* **7**: 110–124. doi:[10.1016/j.mio.2014.02.002](https://doi.org/10.1016/j.mio.2014.02.002)
- Chase, A. P., S. J. Kramer, N. Haëntjens, E. S. Boss, L. Karp-Boss, M. Edmondson, and J. R. Graff. 2020. Evaluation of diagnostic pigments to estimate phytoplankton size classes. *Limnol. Oceanogr. Methods* **18**: lom3.10385. doi:[10.1002/lom3.10385](https://doi.org/10.1002/lom3.10385)
- Chisholm, S. W. 1992. Phytoplankton size, p. 213–237. In P. G. Falkowski, A. D. Woodhead, and K. Vivirito [eds.], Primary productivity and biogeochemical cycles in the sea. Springer.
- Ciotti, Á. M., and A. Bricaud. 2006. Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil: Algal size and CDM from SeaWiFS data. *Limnol. Oceanogr. Methods* **4**: 237–253. doi:[10.4319/lom.2006.4.237](https://doi.org/10.4319/lom.2006.4.237)
- Dall’Olmo, G., T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade. 2009. Significant contribution of large particles to optical backscattering in the open ocean. *Biogeosciences* **6**: 947–967. doi:[10.5194/bg-6-947-2009](https://doi.org/10.5194/bg-6-947-2009)
- Della Penna, A., and P. Gaube. 2019. Overview of (sub)meso-scale ocean dynamics for the NAAMES field program. *Front. Mar. Sci.* **6**: 384. doi:[10.3389/fmars.2019.00384](https://doi.org/10.3389/fmars.2019.00384)
- Fox, J., and others. 2020. Phytoplankton growth and productivity in the western North Atlantic: Observations of regional variability from the NAAMES field campaigns. *Front. Mar. Sci.* **7**: 24. doi:[10.3389/fmars.2020.00024](https://doi.org/10.3389/fmars.2020.00024)
- Freedman, D., and P. Diaconis. 1981. On the histogram as a density estimator: L 2 theory. *Z. Wahrscheinlichkeit. Geb.* **57**: 453–476. doi:[10.1007/BF01025868](https://doi.org/10.1007/BF01025868)
- Gin, K. Y. H., S. W. Chisholm, and R. J. Olson. 1999. Seasonal and depth variation in microbial size spectra at the Bermuda Atlantic time series station. *Deep-Sea Res. I Oceanogr. Res. Pap.* **46**: 1221–1245. doi:[10.1016/S0967-0637\(99\)00004-7](https://doi.org/10.1016/S0967-0637(99)00004-7)
- Green, R. E., H. M. Sosik, R. J. Olson, and M. D. DuRand. 2003. Flow cytometric determination of size and complex refractive index for marine particles: Comparison with independent and bulk estimates. *Appl. Optics* **42**: 526. doi:[10.1364/AO.42.000526](https://doi.org/10.1364/AO.42.000526)
- Hilligsøe, K. M., K. Richardson, J. Bendtsen, L.-L. Sørensen, T. G. Nielsen, and M. M. Lyngsgaard. 2011. Linking phytoplankton community size composition with temperature, plankton food web structure and sea-air CO<sub>2</sub> flux. *Deep-Sea Res. I Oceanogr. Res. Pap.* **58**: 826–838. doi:[10.1016/j.dsr.2011.06.004](https://doi.org/10.1016/j.dsr.2011.06.004)
- Hirata, T., J. Aiken, N. Hardman-Mountford, T. J. Smyth, and R. G. Barlow. 2008. An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sens. Environ.* **112**: 3153–3159. doi:[10.1016/j.rse.2008.03.011](https://doi.org/10.1016/j.rse.2008.03.011)
- Hooker, S. B., and others. 2009. The third seaWiFS HPLC analysis round-robin experiment (seaHARRE-3). NASA.
- Huete-Ortega, M., A. Calvo-Díaz, R. Graña, B. Mouriño-Carballido, and E. Maraño. 2011. Effect of environmental forcing on the biomass, production and growth rate of size-fractionated phytoplankton in the Central Atlantic Ocean. *J. Mar. Syst.* **88**: 203–213. doi:[10.1016/j.jmarsys.2011.04.007](https://doi.org/10.1016/j.jmarsys.2011.04.007)
- Jerlov, N. G. 1976. Marine optics. Elsevier Science.
- Jonasz, M., and G. R. Fournier. 2007. Light scattering by particles in water: Theoretical and experimental foundations, 1st ed. Elsevier, Academic Press.
- Karp-Boss, L., E. Boss, and P. A. Jumars. 1996. Nutrient fluxes to planktonic osmotrophs in the presence of fluid motion. *Oceanogr. Mar. Biol.* **34**: 71–108.
- Kleiber, M. 1932. Body size and metabolism. *Hilgardia* **6**: 315–353. doi:[10.3733/hilg.v06n11p315](https://doi.org/10.3733/hilg.v06n11p315)
- Koestner, D., D. Stramski, and R. A. Reynolds. 2020. Assessing the effects of particle size and composition on light scattering through measurements of size-fractionated seawater samples. *Limnol. Oceanogr.* **65**: 173–190. doi:[10.1002/lno.11259](https://doi.org/10.1002/lno.11259)
- Kostadinov, T. S., and others. 2017. Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models. *Remote Sens. Environ.* **190**: 162–177. doi:[10.1016/j.rse.2016.11.014](https://doi.org/10.1016/j.rse.2016.11.014)
- Kostadinov, T. S., D. A. Siegel, and S. Maritorena. 2010. Global variability of phytoplankton functional types from space: Assessment via the particle size distribution. *Biogeosci. Discuss.* **7**: 4295–4340. doi:[10.5194/bgd-7-4295-2010](https://doi.org/10.5194/bgd-7-4295-2010)



- Laney, S. R., and H. M. Sosik. 2014. Phytoplankton assemblage structure in and around a massive under-ice bloom in the Chukchi Sea. *Deep-Sea Res. II Top. Stud. Oceanogr.* **105**: 30–41. doi:[10.1016/j.dsr2.2014.03.012](https://doi.org/10.1016/j.dsr2.2014.03.012)
- Lee, Z.-P., K.-P. Du, and R. Arnone. 2005. A model for the diffuse attenuation coefficient of downwelling irradiance. *J. Geophys. Res.* **110**: C02016. doi:[10.1029/2004JC002275](https://doi.org/10.1029/2004JC002275)
- Li, W. K. W., F. A. McLaughlin, C. Lovejoy, and E. C. Carmack. 2009. Smallest algae thrive as the Arctic Ocean freshens. *Science* **326**: 539. doi:[10.1126/science.1179798](https://doi.org/10.1126/science.1179798)
- Litchman, E., and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. *Annu. Rev. Ecol. Evol. Syst.* **39**: 615–639. doi:[10.1146/annurev.ecolsys.39.110707.173549](https://doi.org/10.1146/annurev.ecolsys.39.110707.173549)
- Lombard, F., and others. 2019. Globally consistent quantitative observations of planktonic ecosystems. *Front. Mar. Sci.* **6**: 196. doi:[10.3389/fmars.2019.00196](https://doi.org/10.3389/fmars.2019.00196)
- Longhurst, A. R. 2007. *Ecological geography of the sea*, 2nd ed. Academic Press.
- Mackey, M., D. Mackey, H. Higgins, and S. Wright. 1996. CHEMTAX—A program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* **144**: 265–283. doi:[10.3354/meps144265](https://doi.org/10.3354/meps144265)
- Marañón, E. 2015. Cell size as a key determinant of phytoplankton metabolism and community structure. *Annu. Rev. Mar. Sci.* **7**: 241–264. doi:[10.1146/annurev-marine-010814-015955](https://doi.org/10.1146/annurev-marine-010814-015955)
- Marañón, E., P. Cermeño, M. Latasa, and R. D. Tadonlélé. 2012. Temperature, resources, and phytoplankton size structure in the ocean. *Limnol. Oceanogr.* **57**: 1266–1278. doi:[10.4319/lo.2012.57.5.1266](https://doi.org/10.4319/lo.2012.57.5.1266)
- Moberg, E. A., and H. M. Sosik. 2012. Distance maps to estimate cell volume from two-dimensional plankton images: Distance map cell volume algorithm. *Limnol. Oceanogr. Methods* **10**: 278–288. doi:[10.4319/lom.2012.10.278](https://doi.org/10.4319/lom.2012.10.278)
- Mobley, C. D., and E. S. Boss. 2012. Improved irradiances for use in ocean heating, primary production, and photo-oxidation calculations. *Appl. Optics* **51**: 6549. doi:[10.1364/AO.51.006549](https://doi.org/10.1364/AO.51.006549)
- Morán, X. A. G., A. López-Urrutia, A. Calvo-Díaz, and W. K. W. Li. 2010. Increasing importance of small phytoplankton in a warmer ocean. *Glob. Chang. Biol.* **16**: 1137–1144. doi:[10.1111/j.1365-2486.2009.01960.x](https://doi.org/10.1111/j.1365-2486.2009.01960.x)
- Mouw, C. B., and others. 2017. A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. *Front. Mar. Sci.* **4**: 1–19. doi:[10.3389/fmars.2017.00041](https://doi.org/10.3389/fmars.2017.00041)
- Mullin, M. M. 1965. Size fractionation of particulate organic carbon in the surface waters of the western Indian Ocean. *Limnol. Oceanogr.* **10**: 459–462. doi:[10.4319/lo.1965.10.3.0459](https://doi.org/10.4319/lo.1965.10.3.0459)
- Picheral, M., S. Colin, and J.-O. Irisson. 2017. EcoTaxa, a tool for the taxonomic classification of images. <http://ecotaxa.obs-vlfr.fr>
- Pomeroy, L. R. 1974. The ocean's food web, a changing paradigm. *Bioscience* **24**: 499–504. doi:[10.2307/1296885](https://doi.org/10.2307/1296885)
- Reynolds, R. A., D. Stramski, V. M. Wright, and S. B. Woźniak. 2010. Measurements and characterization of particle size distributions in coastal waters. *J. Geophys. Res.* **115**: C08024. doi:[10.1029/2009JC005930](https://doi.org/10.1029/2009JC005930)
- Richardson, T. L. 2019. Mechanisms and pathways of small-phytoplankton export from the surface ocean. *Ann. Rev. Mar. Sci.* **11**: 57–74. doi:[10.1146/annurev-marine-121916-063627](https://doi.org/10.1146/annurev-marine-121916-063627)
- Rohart, F., B. Gautier, A. Singh, and K.-A. Lê Cao. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**: e1005752. doi:[10.1371/journal.pcbi.1005752](https://doi.org/10.1371/journal.pcbi.1005752)
- Runyan, H., R. A. Reynolds, and D. Stramski. 2020. Evaluation of particle size distribution metrics to estimate the relative contributions of different size fractions based on measurements in Arctic waters. *J. Geophys. Res. Oceans* **125**: 1–19. doi:[10.1029/2020JC016218](https://doi.org/10.1029/2020JC016218)
- Saijo, Y. 1964. Size distribution of photosynthesizing phytoplankton in the Indian Ocean. *J. Oceanogr. Soc. Jpn.* **19**: 187–189. doi:[10.5928/kaiyou1942.19.187](https://doi.org/10.5928/kaiyou1942.19.187)
- Sathyendranath, S., and others. 2009. Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. *Mar. Ecol. Prog. Ser.* **383**: 73–84. doi:[10.3354/meps07998](https://doi.org/10.3354/meps07998)
- Scott, D. W. 1979. On optimal and data-based histograms. *Biometrika* **66**: 605–610. doi:[10.1093/biomet/66.3.605](https://doi.org/10.1093/biomet/66.3.605)
- Sheldon, R. W., A. Prakash, and W. H. Sutcliffe. 1972. The size distribution of particles in the ocean. *Limnol. Oceanogr.* **17**: 327–340. doi:[10.4319/lo.1972.17.3.0327](https://doi.org/10.4319/lo.1972.17.3.0327)
- Sieburth, J. M. N., V. Smetacek, and J. Lenz. 1978. Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol. Oceanogr.* **23**: 1256–1263. doi:[10.4319/lo.1978.23.6.1256](https://doi.org/10.4319/lo.1978.23.6.1256)
- Slade, W. H., E. Boss, G. Dall'Olmo, M. R. Langner, J. Loftin, M. J. Behrenfeld, C. Roesler, and T. K. Westberry. 2010. Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *J. Atmos. Oceanic Tech.* **27**: 1733–1746. doi:[10.1175/2010JTECHO755.1](https://doi.org/10.1175/2010JTECHO755.1)
- Smith, R. C., and K. S. Baker. 1978a. Optical classification of natural waters 1: Optical classification. *Limnol. Oceanogr.* **23**: 260–267. doi:[10.4319/lo.1978.23.2.0260](https://doi.org/10.4319/lo.1978.23.2.0260)
- Smith, R. C., and K. S. Baker. 1978b. The bio-optical state of ocean waters and remote sensing 1: Remote sensing. *Limnol. Oceanogr.* **23**: 247–259. doi:[10.4319/lo.1978.23.2.0247](https://doi.org/10.4319/lo.1978.23.2.0247)
- Sturges, H. A. 1926. The choice of a class interval. *J. Am. Stat. Assoc.* **21**: 65–66.
- Uitz, J., H. Claustre, A. Morel, and S. B. Hooker. 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *J. Geophys. Res.* **111**: 1–23. doi:[10.1029/2005JC003207](https://doi.org/10.1029/2005JC003207)
- Van Heukelem, L., and C. S. Thomas. 2001. Computer-assisted high-performance liquid chromatography method

- development with applications to the isolation and analysis of phytoplankton pigments. *J. Chromatogr. A* **910**: 31–49. doi:[10.1016/S0378-4347\(00\)00603-4](https://doi.org/10.1016/S0378-4347(00)00603-4)
- Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and J.-C. Marty. 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *J. Geophys. Res.* **106**: 19939–19956. doi:[10.1029/1999JC000308](https://doi.org/10.1029/1999JC000308)
- Weithoff, G., and B. E. Beisner. 2019. Measures and approaches in trait-based phytoplankton community ecology—From freshwater to marine ecosystems. *Front. Mar. Sci.* **6**: 40. doi:[10.3389/fmars.2019.00040](https://doi.org/10.3389/fmars.2019.00040)
- Xing, X., and E. Boss. 2021. Chlorophyll-based model to estimate underwater photosynthetically available radiation for modeling, *in-situ*, and remote-sensing applications. *Geophys. Res. Lett.* **48**: 1–11. doi:[10.1029/2020GL092189](https://doi.org/10.1029/2020GL092189)
- York, D. 1966. Least-squares fitting of a straight line. *Can. J. Phys.* **44**: 1079–1086.

## Acknowledgments

The authors thank M. Behrenfeld and C. Hoestler for leading the NAAMES project, P. Gaube for support during the campaigns, and the crew of the R/V Atlantis for all their help at sea. The authors would like to thank all those who provided ancillary data and especially N. Huynh and C. Davie-Martin for organizing it. The authors appreciate the help from G. Bourdin and F. Lombard for advanced statistical analysis. This work was funded by the National Atmospheric and Space Administration (NASA) (Grant nos. NNX15AE67G and NNX15AAF30G). The authors are grateful to the two anonymous reviewers for their helpful comments that greatly improved the manuscript. Data from the NAAMES project is shared on the NASA SeaBASS archive (doi:[10.5067/SeaBASS/NAAMES/DATA001](https://doi.org/10.5067/SeaBASS/NAAMES/DATA001)).

## Conflict of Interest

There is no conflicts of interest.

Submitted 25 October 2021

Revised 28 March 2022

Accepted 13 May 2022

Associate editor: David Antoine